Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze della Terra, della Vita e dell'Ambiente

Ciclo 29

Settore Concorsuale di afferenza: 05/B2

Settore Scientifico disciplinare: BIO/06

TITOLO TESI

Molecular and comparative genomics analysis of proteins and peptides involved in epidermal differentiation of reptiles and birds (Sauropsida)

Presentata da: **Karin Brigit Holthaus**

Coordinatore Dottorato **Relatore**

Prof.ssa Barbara Mantovani **Prof. Lorenzo Alibardi**

Esame finale anno 2017

Dedicated to

all the people that believed in me

and

to all true researchers

that seek enrichment in scientific knowledge more than financial gain

that put personal gratification in the research before the "'ranking'' deriving from it

Table of contents

List of publications that have originated from this PhD thesis

This PhD thesis is based on the following publications, which will be referred to by their Roman numbers (I-V) in the text:

I. Holthaus K.B.[†], Strasser B.[†], Sipos W., Schmidt H.A., Mlitz V., Sukseree S., Weissenbacher A., Tschachler E., Alibardi L., Eckhart L. 2016. Comparative genomics identifies epidermal proteins associated with the evolution of the turtle shell. Mol Biol Evol 33(3):726-37.

II. Holthaus K.B., Spisni E., Alibardi L. 2016. Microbicide activity of two reptilian antimicrobial peptides on Gram positive and Gram negative bacteria. J Immuno Biol 1:1.

III. Alibardi L., Holthaus K.B., Sukseree S., Hermann M., Tschachler E., Eckhart L. 2016. Immunolocalization of a histidine-rich Epidermal Differentiation Protein in the chicken supports the hypothesis of an evolutionary developmental link between the embryonic subperiderm and feather barbs and barbules. PLoS One 11(12): e0167789.

IV. Holthaus K.B., Mlitz V., Strasser B., Tschachler E., Alibardi L., Eckhart L. 2017. Identification and comparative analysis of the epidermal differentiation complex in snakes. Sci. Rep. 7, 45338; doi: 10.1038/srep45338.

V. Holthaus K.B. & Alibardi L. 2017. Disulfide-bond-mediated cross-linking of corneous beta-proteins in lepidosaurian epidermis. Zoology. Under review.

[†] These authors have contributed equally to the work.

Abstract

The epidermis protects the body of vertebrates against many biological, chemical, physical and mechanical hazards present in the environment. The adaptation of the epidermis to a permanent life outside of the water was an essential step in the evolution of reptiles, birds, and mammals. The common ancestral origin of these terrestrial vertebrates is reflected in the shared presence of a gene cluster called the Epidermal Differentiation Complex (EDC), which controls the formation of the cornified layers of the epidermis. The aim of this study was to characterize novel factors and processes that contribute to the protective functions of the epidermis in reptiles and birds (Sauropsida).

Comparative studies of genome and transcriptome sequences led to the first description of the EDC and its protein products in turtles, crocodilians and snakes. In each of these reptilian clades 50-100 epidermal differentiation genes were characterized at the sequence level. By reverse transcription-polymerase chain reaction experiments, turtle EDC genes were confirmed to be expressed in the skin or specifically in the scutes of the turtle shell. A newly generated antibody was used to demonstrate the expression of an avian EDC protein in feathers.

Furthermore, specific features of the epidermal barrier, such as the defense mechanisms against microbes and the molecular cross-linking of corneous beta proteins (beta-keratins), were investigated *in vitro*. Reptilian skin-specific peptides were tested for their antimicrobial activities on Gram-negative and Gram-positive bacteria. Disulfide bonds of corneous beta proteins were studied by exposure to reduction and alkylation or oxidation followed by western blot analysis.

Together, the results of these bioinformatics studies, analyses of gene expression, and mechanistic assays provided important new insights into the evolution and functions of epidermal proteins in sauropsids and the epidermal barrier to the environment in general.

Acknowledgements

First of all I thank Prof. Lorenzo Alibardi of the Department of BiGeA, University of Bologna for having offered this opportunity to study the reptile and bird epidermis. I have immensely appreciated the autonomy he has given me without ever lacking guidance though. Whenever I have needed advice he has shared without reservations his broad experience in the research field of Herpetology. His huge knowledge of sauropsids especially on their epidermal and regeneration processes combined to his passion for his work have been a true inspiration. Whilst his intense, continuous and broadly orientated scientific production has been a motivation to me.

I would also like to express my gratitude to Prof. Spisni of the Department of BiGeA, University of Bologna for letting me use his lab and lab facilities during my PhD study. On numerous occasions he has been ready with help and advice and this has been of particular importance for the part concerning the antimicrobial assays. A warm thanks to all the lab members as well for having adopted this "stray" researcher and accepted her presence.

Another person I absolutely need to thank is Alessio Papi who has taught me all the lab work basics as well as cell culture and has been essential during the initial phase of the antimicrobial assays. He has not only offered to share his office with me, but also has supported my presence there for the last 3 years. Last but not least I owe all I know about electrophoresis and western blotting to Francesca Borsetti of the Proteomics lab. Her door has always been open for help, advice and solutions to "strange" problems that happen to still inexperienced researchers.

The *S. aureus* strain implied in the microbial assays has been kindly donated by Dr. Sandra Turroni of the Department of Pharmacy and Biotechnology, University of Bologna.

Most of the work presented here was done in collaboration with the Research Division of Biology and Pathobiology of the Skin, Department of Dermatology of the Medical University of Vienna. I thank Prof. Erwin Tschachler, Head of the Division, for offering me hospitality in his lab on more than one occasion. My immense gratitude goes to Prof. Leopold Eckhart who has been my supervisor when staying in Vienna. He has been indispensable for the comparative study with his

help, advice and teaching. Furthermore his synthetic, precise and scientifically critical approach has been essential during data elaboration and manuscript revisions. He has been able to extract scientifically sound results from hypothetical ones and prevented me to get lost in the how and why.

Many thanks go to my Viennese colleague Betty Strasser who has put the basis for the comparative study and with whom I shared the first year of my PhD. She has patiently introduced me to both the bioinformatics methods and the experimental ones used for the comparative analysis of the EDC. A special thanks goes to Veronika Mlitz who apart from contributing to the results has never refused her help at any time. Her broad experience in the lab and with bioinformatics analysis has been extremely useful. Benz (Supawadee Sukseree) has contributed to the very nice histological data and has been a valuable help during immunohistochemical experiments. Further thanks go to Maria Buchberger and Bahar Golabi for their help in the lab.

The analysis of the turtle corneous beta proteins (beta keratins) has been done in collaboration with Heiko A. Schmidt of the Center for Integrative Bioinformatics Vienna (CIBIV), Max F. Perutz Laboratories, University of Vienna.

Anton Weissenbacher (Vienna Zoo) and Wolfgang Sipos (University of Veterinary Medicine, Vienna, Austria) have made essential contributions in the preparation of tissue from *E. orbicularis* embryos that was used for gene expression analyses.

Part of the work in Vienna was supported by the Austrian Science Fund (FWF): P23801.

In the end I thank Prof. Luisa Dalla Valle of the Department of Biology, University of Padua, for her advice and practical help in some of the lab work.

Of course I cannot do without thanking my parents who have always encouraged and supported my academic career. Thanks to Massimo, who although never fully understanding where I have been talking about these last years, has listened to me all the same. I’m grateful to my partner, family and friends for their support and faith in me.

List of abbreviations

α	Alpha
Ab	Antibody
AMP	antimicrobial peptide
β	Beta
BSA	Bovine serum albumin
C	Cysteine
°C	Degrees Celsius
CBP	Corneous beta protein (known also as beta-keratin)
cDNA	Complementary deoxyribonucleic acid
CE	Cornified cell envelope
CRNN	Cornulin
C-terminus	Carboxy terminus
DTT	Dithiothreitol
EDC	Epidermal differentiation complex
EDMTFH	Epidermal Differentiation protein starting with MTF motif and rich in Histidine
g	Gram
G	Glycine
H_2O_2	Hydrogen peroxide
IC_{50}	Half maximum inhibitory concentrion
K	Lysine
kDa	Kilodalton
KIF	Keratin intermediate filament (before known as alpha keratin)
KRTAP	Keratin associated proteins of mammalians
l	liter
LB	Lennox broth
µg	Microgram
µl	Microliter
M	Molar
mg	Milligram
MIC	Minimal inhibitory concentration
min	Minute
ml	Milliliter

mM	Millimolar
µM	Micromolar
MW	Molecular weight
Mya	Million years ago
nm	Nanometer
N-terminus	Amino terminus
P	Proline
pH	Potential of Hydrogen (measure of acidity or alkalinity)
PGLYRP	Peptidoglycan recognition protein
Q	Glutamine
RNA	Ribonucleic acid
RT-PCR	Reverse transcriptase polymerase chain reaction
S	Serine
SCFN	Scaffoldin
SEDC	Simple epidermal differentiation protein
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
TBS	Tris-buffered saline (buffer)
SFTP	S100 fused-type protein
SPRR	Small proline rich protein
TCHH	Trichohyalin
Tris-HCl	Tris-Hydrochloride (buffer)
UV	Ultra violet
WB	Western blot

Chapter 1 General introduction

1. Introduction

1.1 Evolutionary history of the Sauropsida clade (birds and reptiles)

The very first traces of vertebrate organisms date back to the Cambrian explosion around 525 million years ago (Mya), a time at which vertebrate life was confined to the water (Shu et al., 1999). According to the fossil record vertebrates started to colonize the land much later about 360-370 Mya during the Upper Devonian (Pough et al., 1999). The conquest of land has been one of the major sources for increased biodiversity and speciation in vertebrates. This radical change in habitat has forced vertebrates to adapt to a variety of new environmental challenges and has given origin to the first amphibian-like Tetrapods during the Paleozoic era. All amphibians, sauropsids (reptiles and birds), mammals and their relative ancestors belong to the Tetrapods (from Greek four footed). The epidermis, subject of the present study, as the outermost layer exposed to the environment had to undergo major changes to avoid dehydration and offer protection against UV radiation as well as both mechanical and physical threats, all of which were no longer tampered by the surrounding water. Another important adaptation in the transition to complete terrestrial life has been the formation of a series of protective layers and membranes around the embryo resolving the question of reproduction outside an aquatic environment. One of these membranes, the amnios, has giving the name to the clade of the amniotes (mammals, sauropsids and extinct relatives). The stem amniotes, ancestors of both mammals and sauropsids can be traced down to approximately 330-340 Mya during the Lower Carboniferous period of the Paleozoic era (van Tuinen & Hadley, 2004; Pough et al., 2015). During the Late-Carboniferous this stem clade diverged into the lineages leading to the Synapsida (from greek fused arch) to which existing Mammalia and their extinct ancestors belong, and to the Diapsida (from Greek two arches) to which Sauropsida or extant reptiles, birds and their extinct ancestors belong. Extant Sauropsida are divided into Archosauria (birds and crocodilians), Testudines (turtles and tortoises) and Lepidosauria (tuatara, lizards and snakes) as shown in Figure 1. Still during the Late Paleozoic era Sauropsida underwent their first

division into the ancestral lineages of Archosauromorphs and Lepidosauromorphs that later led, on one hand, to the Archosauria and Testudines and on the other hand to the Lepidosauria. A lot of uncertainty exists about when the stem Lepidosauria and the divergence into Sphenodontia and Squamata originated. Estimates for the stem lepidosaurs range as widely as 226 to 289 Mya spanning the Late Paleozoic to the Early Mesozoic era (Jones et al., 2013). Further branching of the Archosauria into Avemeta-tarsalia and Crurotarsi, that after ulterior radiations gave rise to nowadays birds and crocodilians respectively, is estimated to have occurred around 219-255 Mya (Chiari et al. 2012; Shen et al., 2011; Janke & Arnason, 1997). For the turtle-archosaur divergence the molecular clock is more precise and set around 250-257 Mya during the Late Paleozoic (Wang et al., 2013; Chiari et al., 2012). The following era, the Mesozoic, also known as the Age of Reptiles because of their wide radiation, has given origin to the major ancestral lineages of all of today's living reptiles. It was not until after the mass extinction of the Permian-Triassic that the first mammals appeared on the scene (Kielan-Jaworowska et al., 2004).

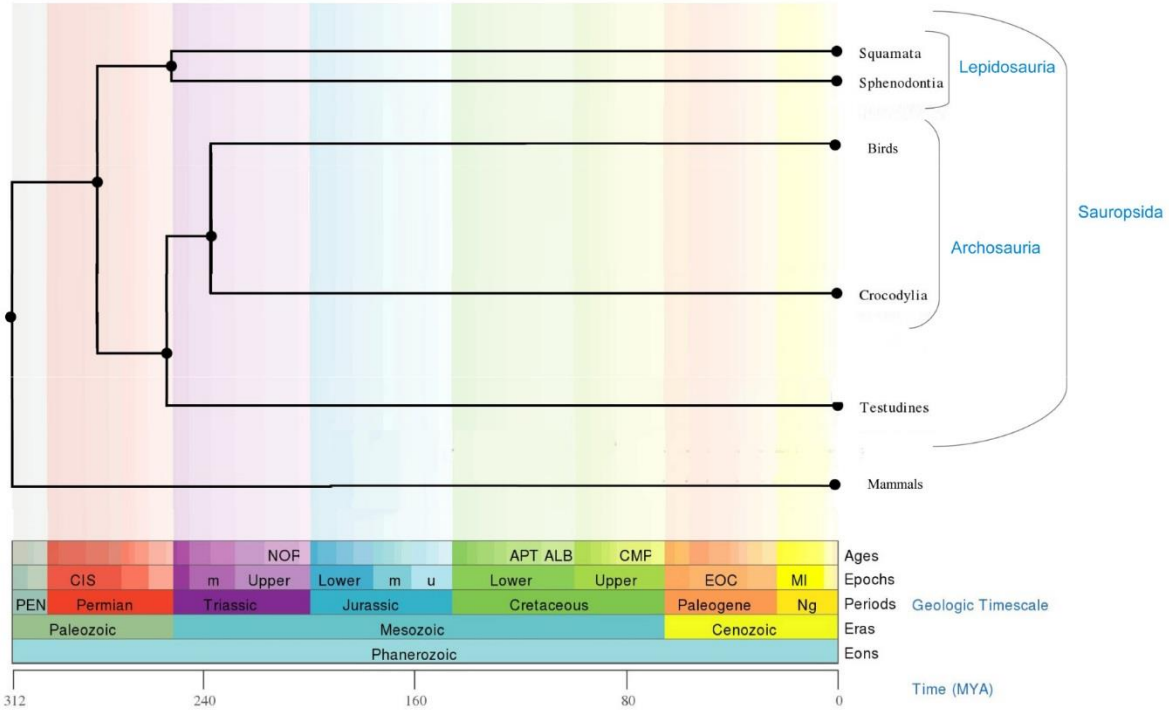


Figure 1. Amniote phylogenetic tree with main divisions on the right. Created with timetree (www.timetree.org) and subsequently modified.

1.2 Phylogeny of extant Sauropsida clades

The first fossils of stem turtles date back to the Late Triassic around 210 Mya (Gaffeney 1986; Rougier et al. 1995), however most recent common ancestors of living turtles (crown turtles) are estimated to have arisen around 157-160 Mya (Chiari et al., 2012). This time point more or less coincides with the split of the turtle lineage into Pleurodira and Cryptodira, respectively side-necked and hidden-necked turtles. Molecular and paleontological studies agree on the timing of origin of crown turtles (Shen et al., 2011; Hugall et al., 2007). Although the ancestral position of the Testudines is still under debate, the general accepted view supported by various molecular studies considers Testudines a sister clade to archosaurs (Fong et al. 2012; Field et al. 2014; Hedges, 2012; Chiari et al, 2012; Tzika et al., 2011; Iwabe et al., 2005). For a long time Testudines were thought due to the lack of an orbital opening (fenestra) in the skull to belong to the Anapsids (from greek no arch), an out group to the sauropsids that are all diapsids (from greek two arches) characterized by two orbital openings (fenestrae). This view purely based on morphological evidence has been now abandoned by most researchers and the anapsid skull is believed to be a secondarily derived character.

Crocodilians similar to modern ones (Eusuchia) were first found during the Triassic, while the radiation of extant families did not start until the upper Cretaceous and Paleocene according to the fossil records (Brochu C.A., 2003). Of the Crocodylia clade that was widely spread and diversified during most of the Cenozoic era, now only three families remain; the Alligatoridae, the Gavialidae and the Crocodylidae. During the Middle-Late Jurassic approximately 160 Mya the first stem birds that gave rise to present-day living species evolved (Godefroit et al., 2013; Xu et al., 2011). In birds as for most extant sauropsid clades, crown members were not encountered in the fossil record until the Cretaceous. Due to the lack of apparent morphological similarity with reptiles, birds are often put into a separate class, but this clade actually nests within the monophyletic sauropsids, as part of archosaurs. Through recent phylogenetic analysis it has become clear that their closest living relatives are the crocodilians (Chiari et al. 2012; Fong et al. 2012;

Field et al. 2014). A main subdivision in the classification of extant birds is given by Paleognata (from greek old jaws) and Neognata (from greek new jaws) and the latter is further divided into Galloanserea and Neoaves. Modern birds comprehend 40 orders according to one of the latest and most extensive studies (Jarvis et al., 2014), however some controversy in bird systematics still exists.

Considering that the earliest Sphenodontia fossil records found are of the Middle-Triassic the division of Lepidosauria into the lineages of Sphenodontia and Squamata is thought to have originated at that time, even though squamate fossils that testify their existence are not found until later during the Late Jurassic (Jones et al., 2013; Fraser & Benton, 1989; Evans, 1995). One of the most recent analysis puts the molecular clock for the crown squamate group at 193 Mya (Jones et al. 2013). The monophyletic order of the Squamata includes all extant snakes, true lizards and legless lizards, while the sister taxa of the Sphenodontia comprises only one extant species. Based on the latest molecular studies five main suborders are generally recognized (Zheng & Wiens, 2016; Pyron et al., 2013; Vidal & Hedges, 2005). These suborders are Dibamidae, Gekkota, Toxifera, Scincoidea and Lacertoidea, but the division in higher-level relationships within suborders is still controversial. Serpentes are now placed within Toxifera together with Iguania and Anguimorpha which are considered sister groups to snakes (Pyron et al., 2013; Wiens et al., 2012; Vidal & Hedges, 2005).

The comparative genomics analysis of this study includes members of all orders of Sauropsida except the Tuatara (Sphenodontia) for which no genome is available yet (see Figure 2).

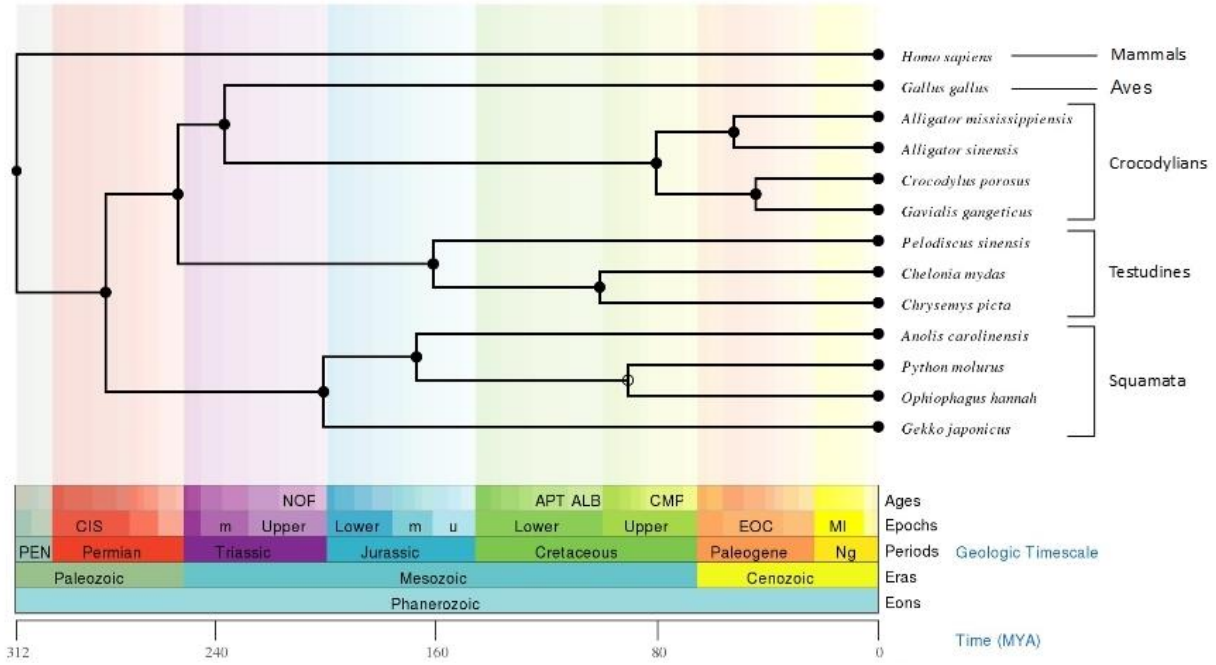


Figure 2. Phylogenetic tree of analyzed sauropsid species. On the right are indicated the main clades to which the species belong. Created by timetree (www.timetree.org) and subsequently modified.

1.3 Diversity of extant Sauropsida clades

Testudines are probably the best recognizable clade due to the evolutionary novelty of the turtle shell composed of a dorsal carapace and a ventral plastron connected by lateral bridges (Zangerl 1969). The shell is composed of dermal plates which are covered externally by epidermal scutes, that represent a flat and not overlapping form of reptilian scales. Vertebrae, ribs and the pelvic and pectoral girdle have been fused with the shell and are an integral part of it. However, soft shelled and leatherback sea turtles have a modified shell that in both cases has lost the external coverage of scutes and has leather-like skin instead. Furthermore the leatherback turtles have also lost the bony layer of the carapace, while the soft shelled forms have reduced it. Other specific anatomical traits of turtles are the already mentioned anapsid skull and the loss of teeth replaced by what is called rhamphotheca, a hard cornified beak-like structure. The 346 members (Uetz & Hošek, 2017) of Testudines are divided into 13 families that are land dwelling (tortoises) as well as marine and fresh water dwelling (turtles) with a pretty much cosmopolitan distribution. Turtles are

of particular interest not only because their modified and unique body plan and consequently modified developmental regulation, but also because their longevity and sex determination mechanisms.

No more than 24 species of crocodylians remain today and most belong to the family of the Crocodylidae. Merely two species, compose the Gavialidae family (the true and false gharial) and the alligator genus (American and Chinese alligator). The remaining members of the Alligatoridae family are caimans represented by 6 species. Crocodilians are mainly confined to tropical and subtropical fresh waters, but some of these predators can also be found in brackish and marine waters and at higher temperate latitudes. A characteristic feature of crocodilians is their heavily armored body. This thick armor consists not only in epidermal scutes, but also in underlying dermal plates (osteoderms). In general crocodilians have retained many primitive traits in their morphology and ecology, although showing a slow rate of genomic change they should not be considered living fossils (Castoe & Pollock, 2013). The different families are mainly distinguished by the snout shape and characteristics. A feature for which this order it is much studied is the sex determination of offspring by temperature. Crocodilians are largely exploited for the commercial use of their skin in the leather industry. At first sight crocodilians do not seem to have much in common with their sister taxa, Aves, but they share oviparity, parental care for their offspring, well developed social interactions and vocal communications.

Although looking very different from the other reptiles, Aves (birds) are the only extant clade that is derived from the extinct dinosaurs. With over 18.000 species (Barrowclough et al., 2016) birds are the most species-rich clade within Sauropsida and they have been able to adapt to most diverse environments. They are the only truly cosmopolitan sauropsids inhabiting even the Antarctic. The most distinct features of Aves are the ability to fly, their unique integument with epidermal derived feathers, hard shelled eggs and loss of teeth that were replaced by a beak. It is also the only endothermic or “warm blooded” member of the Sauropsida. Even if best known for the dote of flight, not all members are capable of it, some birds like kiwi, ostrich, emu, cassowary

and rhea are runners, while others like penguins specialized as swimmers. Adaptation to particular feeding habits is reflected in extremely specialized beaks and is one of the main sources of diversity. Another source of bird specific diversity in function of their habitat and ecology is their plumage. Most diversity resides in the order of the Passeriformes that covers over half of the bird species.

Lepidosauria are distinguished for having a transversal cloacal slit as opposed to longitudinal, loss of a single penis, regular shedding of epidermal layers (ecdysis) and caudal autotomy or self-amputation of their tail (Pough et al. 2001).This order is also the only one where viviparity is part of the reproduction strategy. What mainly distinguishes the lizard-like Sphenodontia is their primitive skull which has retained the lower temporal bar. Nowadays the Sphenodontida (Tuatara) survive with only one species that lives in New Zealand. With over 10.000 species (Uetz & Hošek, 2017) the Squamata are a large clade of reptiles that occupies many different habitats and have a global distribution. These reptiles are a model for studying ecology, behaviour, evolution and origins of asexuality, viviparity, body form and venom. The loss of limbs is the most distinctive trait of snakes, but these specialized animals are also well known for the venom production of many members and their capacity to enlarge body and jaws to eat prey which outsizes them. About 70% of the snakes belong to the family of the Colubridae. Even though snakes are often seen as a distinct group they actually nest within the lizard clade and the reduction and loss of limbs has also happened various times in lizards of the Dibamidae, Anguidae and Amphisbaenia. There is no specific trait that distinguishes true lizards, they form a very versatile and heterogeneous group of squamates. Squamate ecology is the most diverse of all sauropsids and comprehends terrestrial, aquatic, arboreal, burrowing and gliding species. These animals exhibit body forms that can be either generic or very specialized and body size varies immensely from 1,6 cm in the dwarf gecko to over 5 m in the green anaconda. Just to illustrate some of the enormous squamate diversity encountered I mention here geckos with their climbing capacity given by modified digits with adhesive setae, chameleons with their camouflaging capacity by changing the

colour of the skin, the enormous hunting Komodo monitor and the marine algae eating iguana of the Galapagos.

1.4 The epidermis

The most superficial layer of vertebrate integument is the epidermis which composes together with the underlying dermis the skin. Between the epidermis of embryonic ectodermal origin and the dermis of mesenchymal origin is present the basement membrane. Although separated by this physical barrier exchange of information and interaction takes place between the epidermis and dermis especially during formation of epidermal appendages (Chuong, 1998; Douailly, 1977). The vertebrate epidermis is pluristratified forming a squamous stratified cornified epithelium covering the external part of the body and comprises various compartments. These compartments are: a basal layer or stratum germinativum with a proliferating cell type, various layers (suprabasal) of differentiating epidermal cells also known as keratinocytes and in tetrapods a superficial layer (stratum corneum) of dead terminally differentiated cells called corneocytes (Figure 3).

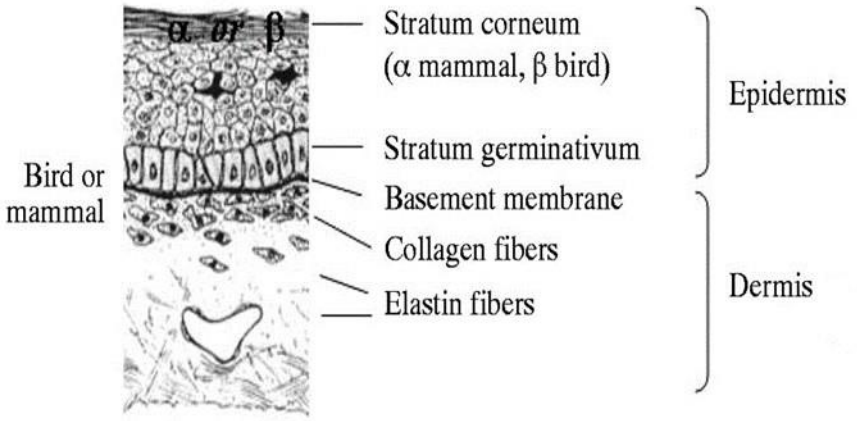


Figure 3. Generalized drawing of the structure of mammalian/bird skin. α or β indicate a layer characterized by respectively an alpha or beta X-ray diffraction pattern. Modified from: Lillywhite H.B., 2006. J. Exp. Biol. 209:202–226.

Among tetrapods the type, number and characteristics of the suprabasal keratinocyte layers can differ and will be described in the various sections dedicated to the analyzed clades. In mammals the structural composition of keratinocytes is known and it consist for 80-90 % in keratin intermediate filaments (KIFs) and filaggrin, and around 7-10% in cornified cell envelope proteins like involucrin, small proline-rich proteins (SPRRs), trichohyalin and loricrin (Steven & Steinert, 1994).

The main function of the epidermis is to protect the vertebrate organism by forming a barrier against biological pathogens and aggressions of chemical, physical and mechanical nature. In all vertebrates the epidermis is keratinized, meaning that the cytoplasma of their keratinocytes becomes packed with KIFs that replace intracellular organelles and become the major structural component. Cornification, often confused with keratinization, is a process that involves the linking of proteins to the cytoskeleton of keratin intermediate filaments forming a resistant and amorphous structure that can be more or less hard according to the type of proteins associated and bonds implied (Alibardi, 2006, 2016a). In some aquatic vertebrates (fish) local cornification takes place, but as an adaptation to life out of the aqueous environment in terrestrial vertebrates (Tetrapods) the whole epidermis undergoes cornification forming the cornified layer. This cornified layer (stratum corneum) consists of dead flat cells (corneocytes) which have undergone a programmed cell death forming one layer in amphibians and several in amniotes. In amniotes the process of cornification comprises not only the formation of a cornified layer like in other tetrapods (amphibians), but also of a cornified cell envelope surrounding the corneocytes and the deposition of lipid sheets between cells (Eckhart et al., 2013; Henry et al., 2012; Candi et al., 2005; Kalinin et al., 2002). In other vertebrates that are aquatic or semi-aquatic, like fish and amphibians, this is not the case. They have a mucus-rich epidermis with numerous glands that allows gas and liquid exchanges with the environment. But in order to adapt to a fully terrestrial (dry) environment, amniotes needed an epidermal barrier which was first of all efficient against cutaneous water loss and secondly more efficient against environmental hazards, like UV radiation and mechanical forces, that had without the buffering

effect of the water a higher impact. For this purpose the cornified cell envelope evolved, further strengthening the corneocytes of the cornified layer and reducing the risk of dehydration by adding lipids both to the surface of the cornified cell envelope and in the inter cellular spaces (Eckhart et al., 2013; Henry et al., 2012; Kalinin et al., 2002; Candi et al, 2005).

A significant difference between mammals and some reptiles is found in the mechanism of desquamation or sloughing of dead cells (corneocytes) that are continually replaced by inner cells that move outwards giving a constant state of dynamic equilibrium in mammals. Instead in sauropsids sloughing is generally periodical and with pieces or even whole epidermal generations (squamates) and not more or less continuous and through single corneocytes as in mammals (Alibardi 2005b; Maderson et al., 1998; Landmann, 1986).

A key innovation of the amniote integument following full cornification was the creation of a rich diversity of hard epidermal derived appendages some of which are unique, like hair in mammals and feathers in birds while others are shared between clades like nails/claws, beaks and spines (Wu et al, 2004, Chang et al., 2009). Similarly to the epidermis also skin appendages are composed of a filament-matrix structure where the filaments are formed by KIFs and the matrix by keratin associated proteins (KRTAPs) that are chemically interacting or even cross-linked by covalent bonds to the KIFs (Matsunaga et al., 2013; Fuijkawa et al., 2012; Gillespie, 1991). The properties of KIFs are to increase tensile strength, but also flexibility or mechanical resistance of the epidermis, while keratin associated proteins confer physical and biochemical resistance, inflexibility and extreme mechanical resilience to the corneous material and the corneous cell envelope of the corneous layer (Alibardi, 2006, 2013a; Resing & Dale, 1991; Kalinin et al, 2002).

The differences between the hard appendages and the epidermis have two causes. In appendages the intense cross-linking is more extensive and obtained through the formation of numerous disulfide bonds. Secondly, the structural proteins forming the corneous material (or matrix component) are in mammalian appendages of a special kind, namely keratin associated proteins (KRTAPs)(Eckhart et al., 2013; Rogers et al., 2006; Gillespie, 1991; Powell & Rogers,

1994; Matoltsy, 1987), while in sauropsid appendages corneous beta proteins (CBPs), traditionally known as beta-keratins, increase in proportion with respect to the epidermis (see next section) (Gregg & Rogers, 1986; Brush, 1993; Sawyer et al.; 2000, Sawyer & Knapp, 2003; Alibardi et al., 2009). The association of KIFs with other structural proteins results in the harder and more resistant structure of skin appendages such as horns, hairs, claws or scales.

1.4.1 The epidermis of Sauropsida

The principal differences between mammalian and sauropsid epidermis are given by the epidermal component of scales and scutes, the existence of corneous beta proteins (beta-keratins), the presence of epidermal layers defined alpha- and beta-layers in Sauropsida (Maderson, 1965; Baden & Maderson, 1970; Landmann, 1986; Alibardi & Toni, 2006). The corneous beta proteins (CBPs) accumulate in the so called beta cells forming a hard and stiff layer, called beta layer where cell boundaries are partial or absent (Maderson et al., 1972, 1998; Landmann, 1979; Alibardi & Sawyer, 2002; Maderson & Alibardi, 2003a; Alibardi, 2012). The name beta layer derives from the beta sheet pattern of the CBPs seen under X-ray diffraction, while the alpha helix pattern of keratin intermediate filaments (previously also known as alpha keratins), gives rise to the term alpha layer (Rudall, 1947; Baden & Maderson, 1970; Fraser et al., 1972).

In general the presence of the softer and more flexible alpha layers is related to regions that require higher elasticity and pliability such as the interscale and interscute regions (hinge regions) as well as body regions involved in movement like the turtle neck and the lizard's dewlap. Another important role of the alpha-layer, and the squamate mesos-layer, is to prevent water loss to the external environment (Maderson et al., 1978; Lillywhite & Maderson, 1982; Menon et al., 1996; Lillywhite, 2006). Where the need for protection from the environment is prevalent as in scales and scutes, the beta layer is predominant (Spearman, 1969; Spearman & Riley, 1969; Alibardi, 2016a). In squamates the disposition of alpha- and beta-layers is mainly vertical, but it is alternated in other sauropsids (Baden & Maderson, 1970, Figure 4).

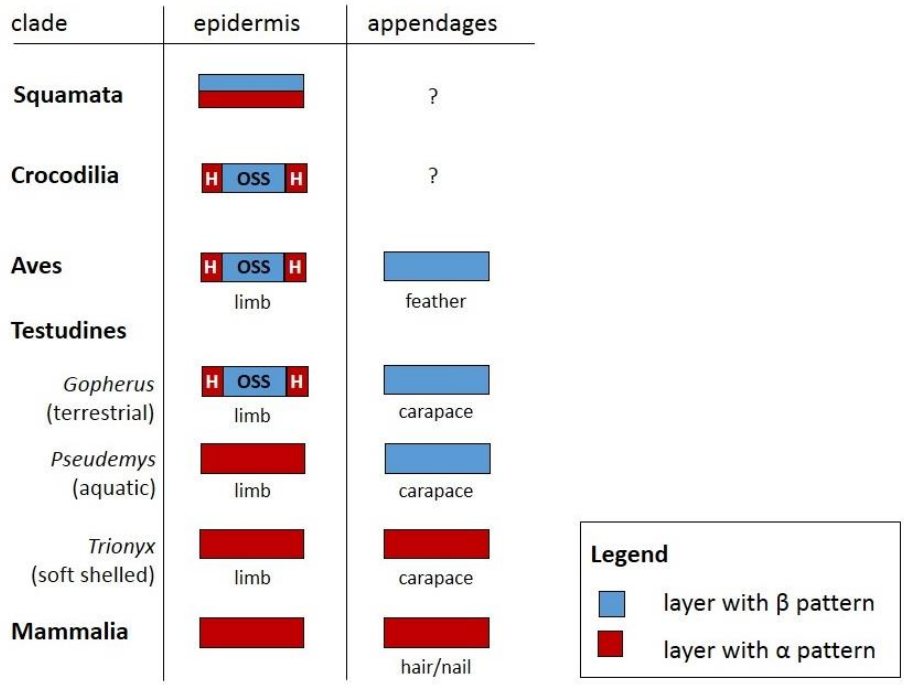


Figure 4. Schematic representation of the pattern given by X-ray diffraction in the epidermis and appendages of amniotes. The pattern although in agreement with a majority of proteins with either α or β conformation does not mean a total absence of proteins of the other type. For example in the hinge region although the alpha layer is predominant a thin a beta layer is present. H; hinge region, OSS; outer scale surface.

Scales are hard and rigid plate-like structures where the keratinocytes of the beta layer are packed with bundled CBPs, while scutes are larger, have a thicker beta-layer than scales and often contain underlying dermal plates. Scales and scutes can come in various measures, shapes and degree of overlap according to the body region (Maderson et al. 1998; Alibardi &Thompson, 2000; Maderson & Alibardi, 2000; Price, 1982), but also the environment might have an impact on their shapes and dimensions (Allam & Abo-Eleneen, 2012; Spearman, 1973; Regal, 1975; Lillywhite & Maderson, 1982). Often scales and scutes have been used to classify reptiles (Maderson, 1964; Soulé & Kerfoot, 1972; Spearman, 1973; Landmann, 1986; Jayne, 1988; Arnold et al., 2002; Brazaitis, 1987; Richardson et al., 2002). The majority of sauropsid scales presents a larger outer surface (dorsal) and shorter inner surface (ventral) (Landmann, 1986). In general archosaurs and

turtles have little overlap in their scales or scutes, while most squamates have smaller scales of the overlapping type.

Although in Sauropsida the exact epidermal composition is not known as for mammals, many keratin intermediate filaments of Sauropsida are annotated in the ncbi database. They also have been studied in sauropsids with available genome sequences (Greenwold et al., 2014; Eckhart et al., 2008; Hallahan et al., 2009; Alibardi et al., 2011; Dalla Valle et al., 2011; Vandebergh & Bossuyt, 2012). For filaggrin and involucrin which are the most abundant mammalian structural keratinocyte proteins after KIFs no homologs have been identified in sauropsids. In fact, keratohyalin granules which contain profilaggrin have not been observed in sauropsids. Many other proteins of the cornified cell envelope have been found in sauropsids and clade specific candidates were identified in sauropsids (Vanhoutteghem et al., 2008; Mlitz et al., 2014; Strasser et al., 2014; **Paper I & IV**). According to several bibliographical sources, CBPs (beta-keratins) are the most abundant proteins in the beta layers (Sawyer et al., 2000; Knapp et al., 1993; Gregg & Rogers, 1986; Klein et al., 2010; Klein & Gorb, 2012; Alibardi et al., 2012; Alibardi, 2013) while the alpha layer consists mainly in KIFs and cornifed cell envelope proteins like loricrin, although CBPs are found as well (Spearman, 1969; Alibardi, 2002, 2003b, 2012, 2015; Alibardi & Toni, 2004; Mlitz et al., 2014; Strasser et al., 2015; Alibardi et al., 2015; Klein et al., 2010; Klein & Gorb, 2012). CBPs are present in the epidermal appendages of sauropsids, instead no KRAPTs are found, proteins typical of mammalian appendages (Gregg & Rogers, 1986; Brush, 1993; Fraser & Parry, 1996). Furthermore in epidermal appendages of sauropsids CBPs also seem to form the dominant component (Presland et al., 1989; Knapp et al., 1993; Sawyer et al., 2000; Wu et al., 2004; Hallahan et al., 2009). The nature of CBPs will be further discussed in section 1.6.

1.4.2 Bird epidermis

Apart from the regions with scales and scutes such as legs, feet and in some birds part of the head (comb and wattles), the avian integument is quite different from that of reptiles. The skin

underneath the feathers and in unscaled regions is fatty, thin, pliable and soft, appearing more similar to the mammalian integument than the reptilian one (Lucas & Stettenheim, 1972; Spearman & Hardy, 1985; Menon et al., 1996; Sawyer & Knapp, 2003). Most of the body is covered by feathers, bird specific epidermal appendages, that exhibit the most complex topology of all epidermal appendages and differ for their more complex morphogenesis. Like the appendages of other sauropsids they contain KIFs, CBPs and other keratin associated proteins, but several proteins have been found specifically in feathers like EDCRP (Strasser et al., 2015) and a histidine-rich protein EDMTFH (**Paper III**) which is the updated version of the earlier identified HRP or FP (Walker & Rogers, 1976; Rogers, 1985). Another feature linked to the feather evolution is the massive expansion of CBP (feather keratin) genes whose products account for most of the feather composition (Glenn et al., 2008; Greenwold & Sawyer, 2010, 2011, 2013; Greenwold et al., 2014; Ng et al., 2014).

Five main types of feathers are distinguished namely filoplumes, bristles, downy-, contour- and flight feathers, which manifest different morphological and functional features and can exhibit many variations (Lucas & Stettenheim, 1972). Different feather phenotypes are obtained by the absence or not of some typical feather building components (Lucas & Stettenheim, 1972) and by the differential combinations of KIF and CBP gene expression that confer diverse structural properties (Ng et al., 2014).

The basic building plan for a typical pennaceous feather consists in a central shaft divided into the proximal calamus partially embedded in the skin and the distal rachis that initiates from the starting point of the ramifications. The latter are called barbs and can develop secondarily ramifications known as barbules which can be adorned by hooklets that are connected to the barbules of the next barb (Prum, 1999). In general KIFs are more represented in the rachis and calamus (Gregg & Rogers, 1986; Ng et al., 2012, 2015; Rice et al., 2013; Wu et al., 2015), while feather corneous beta proteins are more abundant in the barbs and barbules (Ng et al., 2012, 2014; Greenwold et al., 2014; Wu et al., 2015).

The morphogenesis of feathers initiates in barb ridges generated in the feather follicle, possibly derived from the embryonic subperiderm (Sawyer et al., 2005; Sawyer & Knapp, 2003; Alibardi et al., 2016b). Even though feathers are unique to birds this embryonic layer is present as well in crocodilians, the other extant archosaur clade. Furthermore feather keratin specific antibodies have shown to react also in embryonic crocodilian scales (Alibardi & Thompson, 2002; Sawyer et al., 2003a; Alibardi & Sawyer, 2006) making a common origin of scales and feathers plausible. This is further supported by the use in all amniotes of both common developmental pathways like Hedgehog, Wnt/beta-catenin and Bmp (Dhouailly, 1975, 2009; Chuong, 1998;Wu et al., 2004) and anatomical placodes (DiPoi & Milinkowitch, 2016). However it is retained from other researchers that feathers did not evolve from archosaurian scales (Dhouailly, 2009). Several phylogenetic analyses of archosaur CBPs have dated the origin of feather CBPs before the one of bird and crocodilian scale CBPs (Greenwold et al., 2013; Ye et al., 2010; Dalla Valle et al., 2008).

Feather cells (barb and barbules) grow out from the barb ridges, elongating into chains which form filaments that later become cornified and accumulate CBPs, traditionally named feather (beta-) keratins (Alibardi, 2002; Chuong & Widelitz, 1999; Sawyer et al., 1986; Lucas & Stettenheim, 1972; Gregg & Rogers, 1986; Sawyer et al, 2005; Sawyer & Knapp, 2003; Alibardi et al., 2006).

Within the forming barb ridge, the chains of united cells forming barbule cells or the rod forming the barbs are subsequently divided by the degeneration of supportive cells that form spaces and give rise to the rachis, barbs and barbules (Chuong, 1998; Alibardi & Toni, 2008; Alibardi, 2016b). Around the forming feather, a sheath derived from the stratification of the embryonic periderm, is formed but is later lost with the emergence of the feather from its follicle. Due to the similarities observed between embryonic skin development and feather morphogenesis a model has been proposed in which the layered organization of feather follicle reflects that of the embryonic epidermis (Sawyer et al, 2005.; Sawyer & Knapp, 2003). As already mentioned, the feather sheath corresponds to the embryonic secondary periderm, the barbs and barbules to the embryonic

subperiderm, and the marginal plate of barb ridges to the proliferative layer of the embryonic epidermis proper (Figure 5).

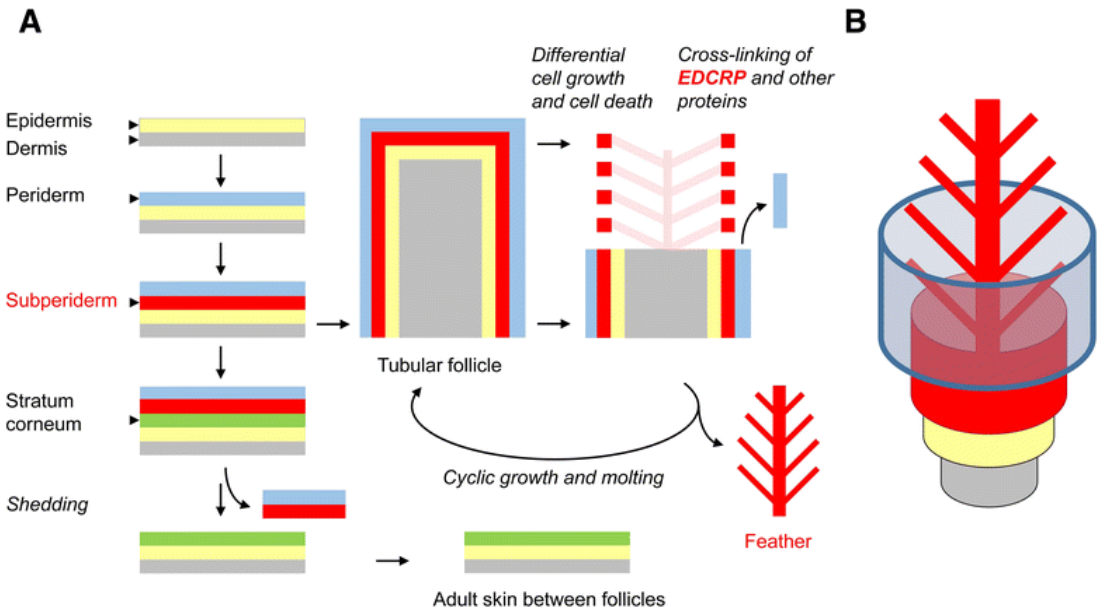


Figure 5. Schematic representation of the embryonic epidermal and feather morphogenesis. The diagram shows the equivalent stratification **(A)** in the embryonic skin and **(B)** in a feather follicle. Source: Strasser et al., 2015. BMC Evol Biol. 15: 82.

Several hypothesis are formulated about the origin of bird feathers and scales. One proposes that feathers and hairs derive from the overlapping epidermal scales of a common tetrapod ancestor of amniotes (Maderson, 1972; Chuong, 1998; Maderson & Alibardi, 2000; Sharpe, 2001; Alibardi, 2004). Another related view is that avian and reptile scales do not have a separate origin, but the same (Wagner, 2014; Alibardi & Sawyer, 2002; Sawyer & Knapp, 2003; Sawyer et al., 2005). A third one believes that avian scales and scutes actually derive secondarily from feathers and that neither feathers nor hairs derive from overlapping scales (Dhouailly, 2009).

Only specialized feathers (asymmetric pennaceous) serve to fly, but the original role was likely essential for insulation of the homoeothermic birds, and also assumed functions in behavioral communication, camouflage and sound production as well (Bostwick & Prum, 2003; Clark & Feo, 2008). Other hard epidermal appendages encountered in birds are the claws and the highly

diversified beak, an analogous structure to the turtle rhamphotheca, where the corneous layer functionally substitutes the loss of teeth.

1.4.3 Crocodilian epidermis

What is most striking about the crocodilian integument is its armored aspect, which is the result of a combination of thick epidermal scales and underlying same sized and shaped dermal plates, called osteoderms. This thick armor with protruding scutes is generally confined to the dorsal part of the crocodilian body and keeled tail scutes, while on the ventral side and the head scutes are mostly flat. Only few crocodilians have a full armor covering also the belly and in the case of the Chinese alligator even the eyelids. Scutes are of the non-overlapping type and arranged in regular rows and patterns. In crocodiles as in many reptiles the different morphology of scales in various body parts has been used for identification (Brazaitis, 1987; Richardson et al., 2002), but these diverse looking scales are histologically and biochemically uniform (Spearman, 1966; Spearman & Riley, 1969; Baden & Maderson, 1970; Parakkal & Alexander, 1972; Sawyer et al., 2000; Alibardi, 2005a). In crocodilians, it has been shown that the pattern of scales observed on their heads is not caused by differential developmental regulation, but simply by physical cracking of the thick hard beta layer in response to increasing pressure from underneath by the fast expanding skull during growth (Milinkovitch et al., 2013).

Not many studies have focused on the crocodilian epidermis which is mostly used in comparison to the well-studied avian integument. As already discussed in the previous section the embryonic epidermis of crocodilians reflects the avian embryonic epidermis and contains the subperiderm, an embryonic layer with similar periderm granules (Alibardi, 2003b) unique to archosaurs. In adults, the epidermis comprises a basal layer with cells that become polygonal during active scale growth, 3-6 suprabasal layers with flat cells, 1-2 pseudostratified transitional or pre-corneous layers and a variably thick cornified layer with relatively thin cells at maturity (Alexander, 1970; Alibardi & Thompson, 2002; Alibardi, 2003b, 2005a, Figure 6). With respect to the outer

scale region, hinge regions not only have a much thinner epidermis, but also thinner mature corneocytes that resemble avian (sebo) keratinocytes (Alibardi & Toni, 2006; Alibardi, 2005a). The protein distribution of crocodilian scutes has been found similar to the one present in avian scutes (Baden & Maderson, 1970) which is in line with their common ancestral origin.

The mechanism of scale growth and shedding is not well know in crocodylians, but it is believed that through environmental wearing corneocytes are lost more or less continuously (Alibardi & Toni, 2006).

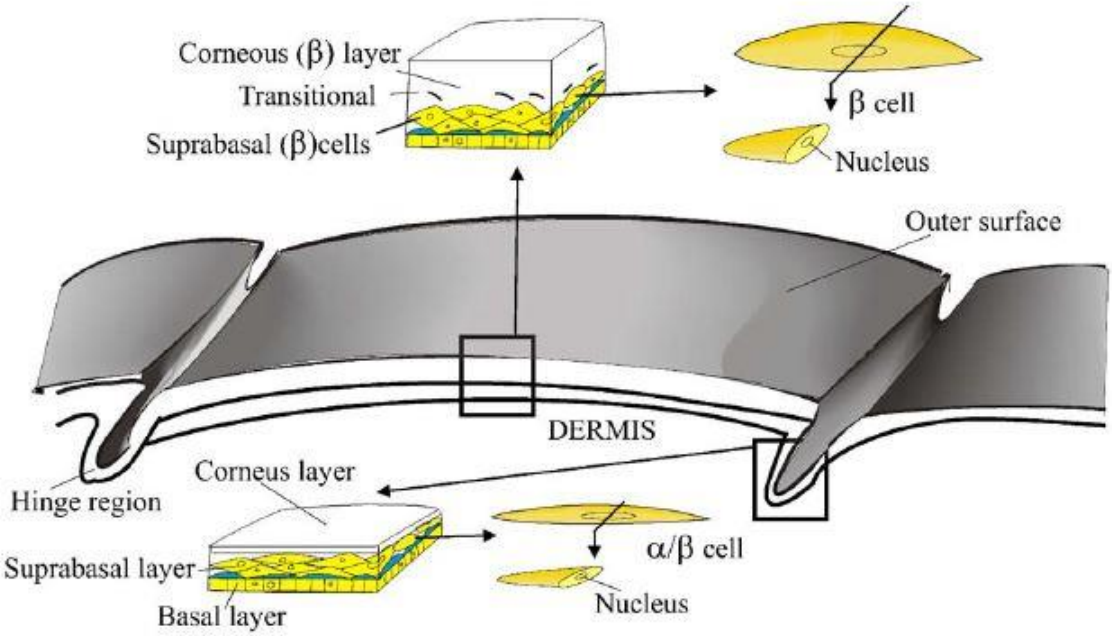


Figure 6. Schematic representation of the crocodilian epidermis. At the top part the outer scale surface with a thick corneous beta layer, while at the bottom part the epidermis in the hinge region with a thin corneous layer. Reprinted from: Prog. Histochem. Cytochem. 40, Alibardi & Toni, Cytochemical, biochemical and molecular aspects of the process of keratinization in the epidermis of reptilian scales. Pp 173-134 © 2006 with permission from Elsevier.

1.4.4 Turtle epidermis

Most studies of the turtle shell have concentrated on the anatomy, developmental regulation and origin of the bony component of this unique structure (Ruckes, 1929; Burke, 1989; Gilbert et al., 2001; Reisz & Head, 2008; Nagashima et al., 2007, 2009; Hirasawa et al., 2013, 2015; Rice et

al., 2015) and not much attention has been given to the epidermal component, but it has been well described in various studies (Alibardi, 2005b; Alibardi &Thompson, 1999; Zangerl, 1969; Alibardi & Dipietrangelo, 2005; Alibardi & Minelli, 2016; Alibardi & Toni, 2006). Various studies investigating the molecular basis of skin morphogenesis and epidermal differentiation in turtles have been published (Dalla Valle et al., 2009, 2013; Li et al., 2013; Moustakas-Verho et al., 2014, 2015; **Paper I**).

The hard skin of the turtle shell (dorsal carapace and ventral plastron) has quite different characteristics than the soft skin found on neck, head, tail and paws of turtles as well as in the region between scutes called hinge region (Parakkal & Alexander, 1972; Maderson, 1985; Landmann, 1986; Alibardi, 2005b, 2013b-c). Within testudines variation can be found in the epidermal component of the "soft" skin, namely aquatic turtles usually do not contain any scales thus having truly soft skin, while tortoises (terrestrial testudines) present overlapped and hard scales similar to those of the alligator (Baden & Maderson, 1970). The hard epidermis of the shell forms large flat plate-like modified scales called scutes that overly the dermal bones. According to their position in the carapace nuchal, vertebral, caudal, costal, and marginal scutes are distinguished, while in the plastron gular, humeral, pectoral, abdominal, femoral, anal ones (Zangerl, 1969). A narrow hinge region is present among the scutes whose sutures do not match those of the underlying dermal bones. The double horny and bone armor strengthens the mechanical resistance of the shell. Aside from the dermal and epidermal skin components various parts of the skeleton like ribs, vertebrae and pectoral and pelvic girdle have fused with the turtle shell. The result is a deep bony dermal-endoskeleton covered by a superficial corneous epidermis, which has given rise to a very resistant, hard, but also inflexible and rigid cage-like structure.

The soft shelled turtles and the leather back sea turtle do not contain scutes in the epidermis and lack or have the bony shell elements reduced. Both have the shell covered by a thin leather-like skin. The shell of these turtles has become flattened, streamlined and pliable with respect to hard shelled turtles as an secondarily derived adaptation to the ecology and feeding habits of these

turtles. Turtles have various hard epidermal appendages which in the case of claws are shared with other sauropsids (Alibardi, 2003a, 2014b,g), but specific in the case of the shell and rhamphotheca, a beak-like structure replacing the function of the lacking teeth (Alibardi, 2016c).

From a histological point of view the soft epidermis of turtles consists in a basal layer, 2-4 stratified layers of differentiating keratinocytes (suprabasal layers), a transitional and multi-stratified corneous layer (Spearman, 1969; Matoltsy & Huszar, 1972; Matoltsy & Bednarz, 1975; Wyld & Brush, 1979; Alibardi, 2002). On the other hand the hard epidermis of the scutes presents only a thin living epidermis and a thick, horny layer of CBPs also called beta-keratins (Baden et al., 1974; Wyld & Brush, 1979, 1983; Homer et al., 2001; Alibardi, 2002; Alibardi et al., 2004, Figure 7). The soft turtle skin corresponds to predominant alpha layers with high level of KIFs whereas the hard one with scutes and scales to predominant beta layers (Alibardi et al., 2004, 2009; Dalla Valle et al., 2013a; Alibardi, 2005b, 2013c, 2014b).

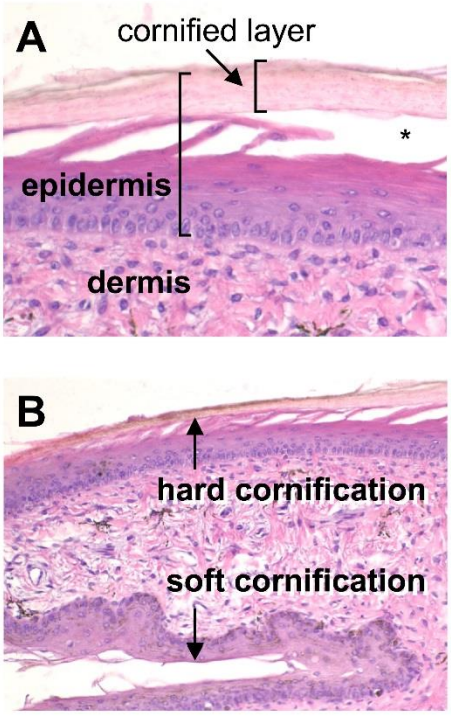


Figure 7. Histological sections of embryonic turtle skin. **(A)** carapace showing a scute with the thick beta layer of the cornified layer in pale pink colouring and **(B)** with soft cornification in the hinge region lacking the thick cornified layer as seen in the scutes that present hard cornification. **(A)** modified from Holthaus et al., 2016. Mol Biol Evol. 33(3):726-37.

The expansion of the CBP cluster has been hypothesized to be related to the novelty of the turtle shell (**Paper I**; Greenwold et al., 2014; Li et al., 2013; Alibardi et al., 2009).

A special note has to be made for the shedding mechanism of scutes which in some testudines is variable and different for land dwelling turtles (tortoises) and water dwelling ones. In most tortoises no shedding of the scutes takes place making that the thickness and weight of the shell increases with age. In this case the external corneous layer is reduced through slow wearing. In other turtles shedding can be either absent, irregular or periodic. In the latter case the whole outer and smaller scutes flake off following the formation of a peculiar shedding layer (Alibardi, 2005b).

1.4.5 Squamate epidermis

Squamate epidermis is the most complex structured epidermis found in all reptiles, and in vertebrates in general, due to its particular shedding mechanism which involves a whole epidermal generation and not just superficial parts of the cornified layer as in other sauropsids. The first detailed histological description of the shedding cycle of squamates was made on the snake and gecko epidermises (Maderson, 1965, 1966). Many studies highlighting the ultrastructural, histological and biochemical aspects of the squamate epidermis have followed (Banjeree & Mittal, 1972; Maderson et al., 1972, 1998; Baden et al., 1974; Landmann, 1979, 1986; Mittal & Singh, 1987a-b; Alibardi et al., 2012; Alibardi, 2012, 2014a, c-f). The process of shedding has been poorly clarified, although it is known that hydrolases and lysosomal enzymes are involved in the degradation of desmosomes between outer and inner epidermal generations (Goslar, 1958; Alibardi, 1997). Recent molecular studies concerning some of the epidermal process have been published clarifying the basic information on the gene structure and, for the first time, also report many CBP sequences from all reptilian groups (Dalla valle et al., 2005, 2007a-b, 2008, 2009a-c2010, 2013a; Strasser et al., 2014, 2015; **Paper IV**), but our basic knowledge of the epidermal histology and shedding of squamates has not changed much over the last forty years.

The epidermal generation that will be shed is called outer generation while the new one forming underneath it is called inner generation. A mature outer generation consist in an oberhautchen, beta, mesos, alpha, lacunar tissue and clear layer, while the inner generation during

most of the resting phase contains an oberhautchen, beta-, mesos and alpha layer. Prior to shedding the squamate epidermis contains two epidermal generations and the basal layer for a total of eleven different layers (Figure 8).

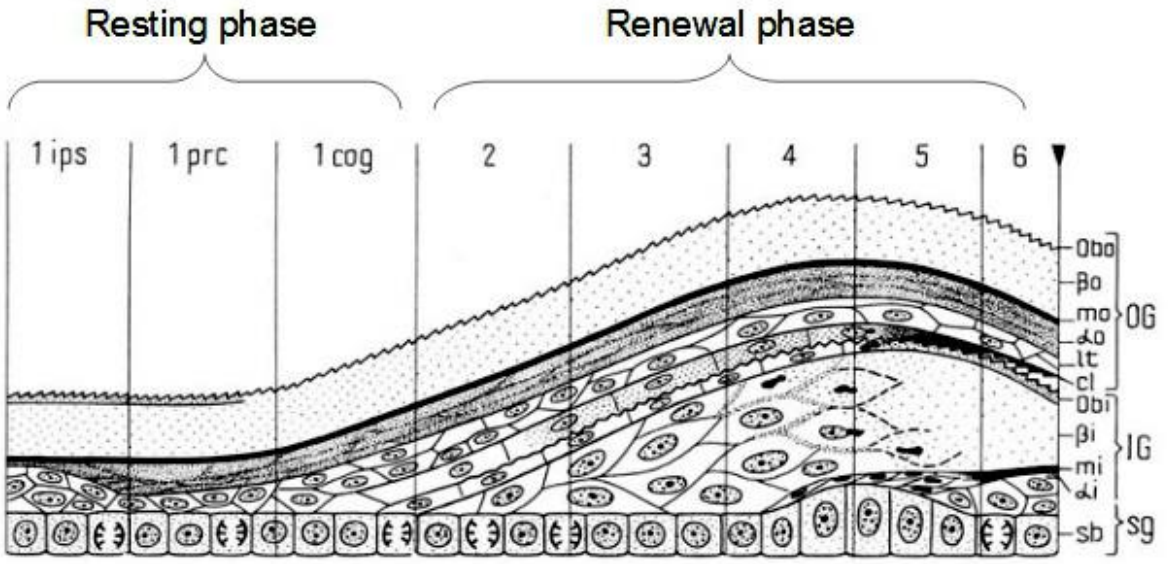


Figure 8. Schematic representation of the shedding cycle in squamate skin. The epidermis consist in multiple layers indicated on the right side of the drawing. The shedding cycle can be divided into 2 phases; the resting and renewal phase. The resting phase is represented by stages 1 with **(1ips)** immediate post-shedding phase, **(1prc)** the perfect resting phase, and **(1cog)** the completion of the outer generation. The renewal phase consists in 6 stages during which new layers are being formed and specified. The outer generation (OG) is shed at the end of stage 6 while the new inner generation (IG) takes its place. Obo (outer oberhautchen), βo (outer β-layer), mo (outer mesos layer), αo (outer α-layer), lt (lacunar tissue) and cl (clear layer), Obi (inner oberhautchen), βi (inner β-layer), mi (inner mesos layer), αi (a partially formed α-layer). sg, germinal layer; sb, stratumbasale. At the right side of stage 6, the reverse triangle marks the time of shedding. Adapted with permission of Springer Nature from: Biology of the Integument, Vol. 2 Vertebrates, Chpt 9 The Skin of Reptiles: Epidermis and Dermis, Pp. 150-187, Landmann L. © Springer-Verlag Berlin Heidelberg 1986.

The shedding cycle is divided into two phases, the resting and the renewal phase this latter is further divided into six stages. The duration of the resting phase is very variable according to the

environmental temperature, circadian clock and physical condition of the animal. Although called resting phase the formation of the epidermis is not completely at a halt during the entire phase. Just after shedding when entering the resting phase the main part of the alpha layer is formed (stage 1ips), followed by a true resting condition where structural protein synthesis is practically absent (stage 1prc or perfect resting stage, Maderson, 1965, 1966, 1985). The resting phase ends with the completion of what will become the outer generation by generating a lacunar and clear layer (stage 1cog). The renewal phase starts with the formation of a new inner generation while the "old" inner generation becomes the outer one. During the various renewal stages an oberhautchen-, beta-, mesos- layer and part of the alpha-layer are build up underneath the outer generation. At the end of the final stage (stage 6), the mature outer generation is shed and the inner generation with a mature oberhautchen and beta-layer but a still immature alpha layer comes to be exposed to the surface (Maderson, 1965; Landmann, 1979, 1986, Figure 8).

Squamate scales can manifest a broad variety in morphology like tuberculated, rectangular, cycloid or keeled the latter often displaying micro ornamentations such as spinules (Arnold, 2002; Maderson et al., 1998; Allam & Abo-Eleneen, 2016; Irish et al., 1988; Price, 1982). Most squamate scales are of the overlapping (imbricated) type. Several hard epidermal appendages like claws, adhesive setae for climbing, crests, frills and different types of spines are encountered in squamates (Chang et al., 2009).

Snakes distinguish themselves for having lost any kind of these appendages with the exception of spurs, claw-like structures in the Boa and pythons. Actually this loss of hard appendages is not unique to snakes, but shared with some other squamate clades. What makes the snake integument unique is the shedding of the outer generation in one piece and their capacity to enlarge their body diameter which requires notable extension of the integument. This extending is obtained mainly by the unfolding of normally pleated interscale (hinge) regions combined with flattening of the scale surface and a change in cell shape that altogether can produce an enlargement of over 300% (Close & Cundall, 2014). The increased friction to which the ventral part of the snake

integument is exposed with respect to squamates with limbs, has led to adaptations for high mechanical resistance to abrasion during sliding locomotion (Klein and Gorb, 2014).

1.5 The epidermal differentiation complex (EDC)

A gene cluster whose protein products are involved in the process of epidermal differentiation and therefor called Epidermal Differentiation Complex (EDC) was first identified and described in human (Mischke et al., 1996). Conservation of this gene cluster was confirmed in different mammalian clades (de Guzmang Strong et al., 2010). This complex was identified for the first time in a sauropsid, the chicken in 2008 (Vanhoutteghem et al., 2008). In 2014 it was completely characterized in chicken as well as in the green anole (Strasser et al., 2014) revealing its synteny with mammals. During this PhD project the analysis started by Strasser and colleagues was extended to turtles (**Paper I**), crocodilians (**Chapter 2.6**) and a specialized squamate clade, the snakes (**Paper IV**).

The EDC includes the S100A1-9 proteins found at one extreme of the complex, the S100A10-11 found at the other extreme and all the genes positioned in between. Many proteins of the EDC are structural components of the cornified cell envelope (CE) that replaces the plasma membrane in terminal differentiated keratinocytes, also called corneocytes (Kalinin et al., 2002). Examples of some of the best known EDC proteins of mammals are involucrin, filaggrin, small proline-rich proteins (SPRRs), trichohyalin and loricrin. The latter protein has also been identified in sauropsids where it is even present with two to three copies. Although filaggrin is a major component of the epidermis in mammals it has not been found in sauropsids. Another group of important EDC proteins present in both mammals and sauropsids is the S100 fused type protein (SFTP) family. While in mammals it consists in seven proteins of which the best known are cornulin, trichohyalin and filaggrin (Henry et al., 2012; Kypriotou et al., 2012), in sauropsids there are only two SFTPs, cornulin and a trichohyalin-like protein called scaffoldin (Mlitz et al., 2014;

Strasser et al., 2014; **Paper I; Chapter 2.6**). Snakes were the only clade where two scaffoldins (SCFNs) were identified bringing the total number of SFTPs to three (**Paper IV**). The SFTPs of sauropsida have been studied in detail and compared to mammals (Mlitz et al., 2014). Thus, in amniotes conservation of the EDC that reflects common phylogenetic origin is found at the level of the S100A proteins, the peptidoglycan recognition protein 3 (PGLYRP3), loricrin, SPRR-like proteins and SFTPs. The main differences between the mammalian and sauropsid EDC are found before and after the loricrin gene/s. In sauropsids many genes for EDC proteins are present before the loricrin(s) and after the PGLYRP3 gene, and other sauropsid specific EDC genes are found after the loricrin gene(s), and are mainly represented by the conspicuous gene cluster of the corneous beta proteins (beta-keratins) as well as *EDAAs* in turtles and archosaurs. In mammals no genes are present between the PGLYRP4 and loricrin and as already mentioned no corneous beta protein genes have been found.

Based on similar head and tail domains and repeat sequences in EDC proteins as well as the organization of EDC genes in tandem arrays, it has been hypothesized that EDC genes have originated by repeated duplications from an ancestral gene driven by the evolutionary selection for improved barrier function (Backendorf & Hohl, 1992; Markova et al., 1993; Strasser et al., 2014).

1.5.1 Epidermal differentiation and cornified cell envelope formation

Most of the knowledge about keratinocyte differentiation and the formation of the cornified cell envelope (CE) has been gained in humans and mammalian model species (Figure 9). The onset of the keratinocyte terminal differentiation pathway which includes the formation of the CE is triggered by different factors. It is believed that the switch of expression from K5 and K14 keratins, which are typical of basal cell gene expression, to K1 and K10 keratins (Watt, 1984; Koster et al., 2007), is initiated by the loss of contact to the basement membrane when keratinocytes move to the suprabasal layers. Later a rise in the intracellular Ca2+ concentration (Kalanin et al., 2002; Kypriotou et al., 2012) and protein kinase C activation (Kypriotou et al., 2012) are important

regulators of differentiation. Synthesis of the early CE precursor involucrin starts already in the spinous layers, while the effective CE assembly starts in the granular layer with the formation of an involucrin and non EDC envoplakin and periplakin scaffold (Kalanin et al., 2002). Somewhat later reinforcement of the forming CE begins by crosslinking loricrin, the main CE component, through calcium dependent enzymes called transglutaminases to other loricrins and to SPRRs (Candi et al., 1999, 2005; Kalanin et al., 2002; Steinert & Marekov, 1995; Steinert et al., 1999). Loricrin and SPRRs are also involved in crosslinking the CE to the KIF-filaggrin network of the keratinocyte. In the CE many other minor EDC proteins become cross-linked by transglutaminases mediated isopeptide bonds to further reinforce the structure. Besides proteins lipids such as ceramides are reversed from lamellar bodies produced in the Golgi apparatus into the extracellular space. These lipids covalently bind to the outer surface of the mature CE forming the lipid envelope, in addition they compose lipid laminae filling the inter corneocyte spaces further impermeabilizing this way the cornified layer (Kalanin et al., 2002; Candi et al., 2005; Kypriotou et al., 2012; Eckhart et al., 2013). Filaggrin has an important role for the structural shape of future corneocytes. This protein is synthesyzed in the granular layer under the form of keratohyalin granules containing phosphorylated profilaggrin. Profilaggrin once dephosphorylated and cleaved into filaggrin determines the aggregation of KIFs into tight bundles and, together with cell dehydration and condensation, causes the collapse of keratinocytes into the flat shaped corneocytes (Resing & Dale, 1991; Candi et al., 2005).

In summary, the terminally differentiated keratinocytes undergo cornification, programmed cell death and replacement of the plasma membrane by the CE in mature corneocytes. At the end of the process a resistant and insoluble structure, known as the cornified cell envelope (CE), composed of keratins embedded in an amorphous matrix of proteins and surrounded by a lipid envelope, is formed (Figure 9).

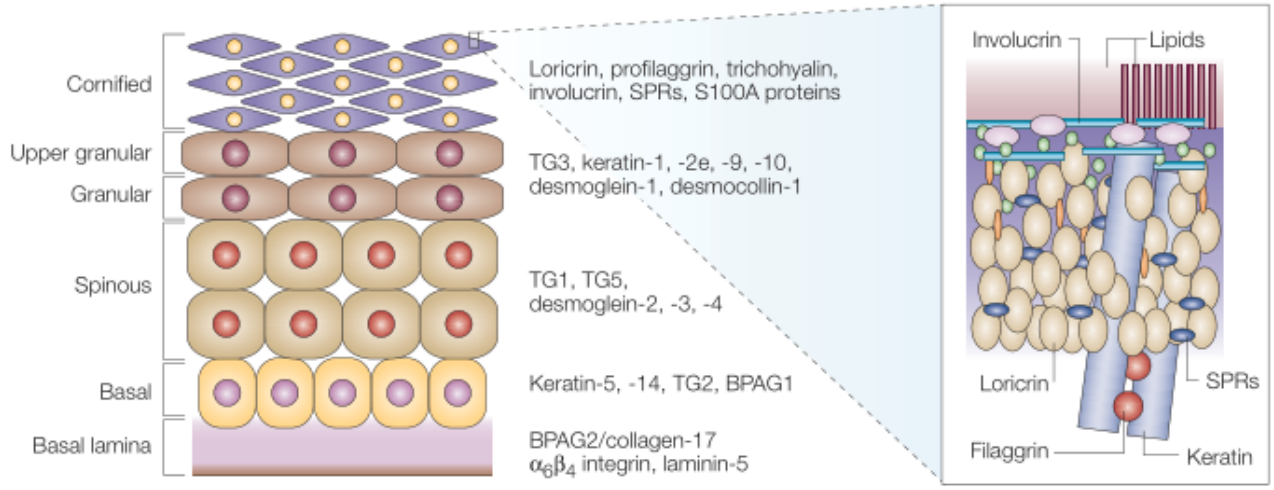


Figure 9. Terminal differentiation of keratinocytes and the cornified cell envelope. On the left the differentiation of keratinocytes in various layers of the epidermis. In the inset on the right the cornified cell envelope where proteins become highly cross-linked by transglutaminases while lipids are deposited on the external surface. TG, transglutaminase. Reprinted by permission from Macmillan Publishers Ltd: Nat Rev Mol Cell Biol. 6:328–340, Candi et al., © 2005.

Lamellar ichthyosis, atopic dermatitis (AD), psoriasis and Vohwinkel syndrome are all diseases caused by faulty cornified cell envelope formation and/or mutations in EDC genes like filaggrin, PGLYRP3 and loricrin, showing the importance of the EDC complex and of the function of the cornified layer in the epidermal barrier (Eckhart et al., 2013; Kypriotou et al., 2012; Irvine et al., 2011; Sun et al., 2006; Palmer et al., 2006; Candi et al., 2005; Maestrini et al., 1996; Huber et al., 1995).

1.6 EDC proteins and their characteristics

Most EDC proteins can be defined as simple epidermal differentiation complex proteins (SEDCs) based on their gene structure that consist in two exons, only one of which coding, separated by an intron (Strasser et al., 2014). Exceptions to this rule are the S100A proteins, PGLYRP3, EDKM and the SFTPs. Usually the 5'- terminal of the non-coding exon is preceded by a canonical TATA box (Strasser et al., 2014). Other features often present are repetitive amino acid

sequences and dominance of a few amino acids like glycine (G), proline (P), cysteine (C), serine (S), lysine (K) and glutamine (Q). The majority of EDC proteins have relatively short sequences and are approximately 60 up to 250 amino acids long. Apart from the beta sheet motif of the corneous beta proteins and the Ca^{2+}-binding EF-hand motif of the S100A and SFTP family no specific domains are present in EDC proteins. But conserved N-and C-terminals have been identified both within the EDC of sauropsids and of amniotes for proteins like loricrin, PGLYRP3, SPRRs and sauropsid SPRR-like proteins (Strasser et al., 2014; **Paper I** & **IV**). Conservation in these terminals is mostly regarding glutamine (Q) and lysine (K) residues that are targets for transglutaminases, and that cross link EDC proteins in the cornified layer. It has been shown that conserved lysine (K) residues of KIFs are implicated in isopeptide binding to glutamine (Q) in the terminal regions of proteins like loricrin, SPRR and involucrin in humans (Candi et al., 1998).

Loricrin is with its 65-70% in human and 80-85% in mouse (Steven & Steinert, 1994) the most conspicuous component of the cornified cell envelope in mammals, even though multiple copies are found in some sauropsids, little is known about its abundance and function in this clade. Its localization has been studied in crocodiles, turtles and anolis lizards (Alibardi, 2003b; Alibardi et al., 2004, 2015). This apolar insoluble protein is extremely rich in glycine residues which because of their minal side chains have a great degree of free rotation resulting in chain flexibility associated with bends (Fraser & Parry, 2014; Steinert et al., 1991). Therefore, the glycine rich sequences of loricrin are presumed to assume a formation with alternating glycine loops (Steinert et al., 1991; Hole et al., 1991). These glycine loop domains display a non-structured organization permitting great mobility and are believed to confer elasticity to the protein in a spring-like nature (Figure 10). At the same time the N- and C-terminal of loricrin which are rich in glutamine an lysine targets provide resistance and insolubility through transglutaminase crosslinking (Steinert et al., 1991; Candi et al., 1998, 1999, 2005) as well as by intramolecular disulfide binding (Hohl et al., 1991).

The presence of proline residues is believed to be correlated to changes in the direction of the polypeptide chain acting as a structural disruptor in secondary elements like alpha helixes and

beta sheets (Levitt, 1978; MacArthur & Thornton, 1991). This is due to the conformational rigidity of proline given by the cyclic structure of its side chain. Proteins such as SPRRs and sauropsid SPRR-like proteins contain a good number of proline repeats. Mammalian SPRRs have, like loricrin, a disorganized structure and glutamine- and lysine-rich terminals. For this reason, SPRRs are believed to confer elasticity by the central domain and mechanical resistance by the N- and C-termini as loricrin does (Steinert et al., 1999; Candi et al., 1999, 2005, Figure 10). Different ratios of loricrin and SPRRs alter physical properties like toughness and rigidity of the epidermis (Steinert et al., 1998a-b; Candi et al., 2005). In addition to their structural function as cross-bridging proteins, SPRRs seem to have anti-oxidative properties and detoxify reactive oxygen species (ROS) through cysteine residues (Vermeij & Backendorf, 2010; Vermeij et al., 2011).

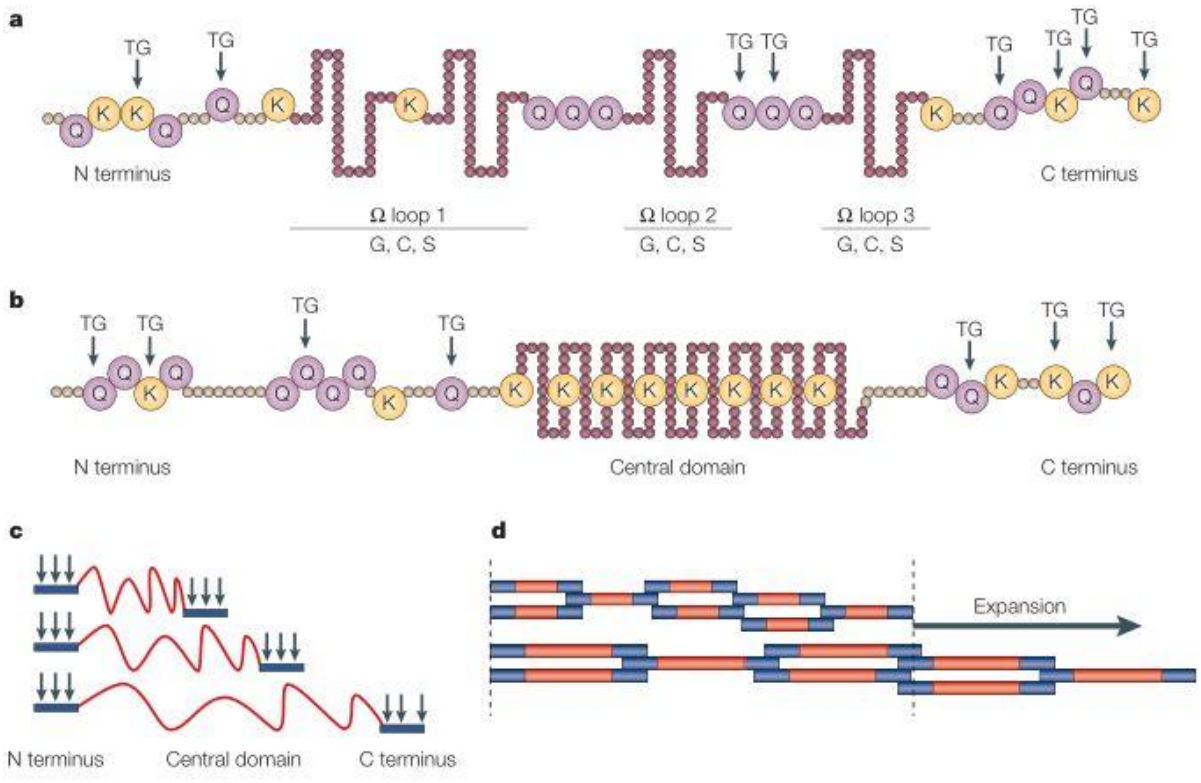


Figure 10. Protein structure and proposed mechanical properties of cornification proteins. Schematic drawings depict the protein structures of loricrin with glycine loops **(a)** and of a SPRR **(b),** while in **(c)** and **(d)** the spring-like mechanism due to the disorganized structure of the central domain of these proteins is shown. Reprinted by permission from Macmillan Publishers Ltd: Nat Rev Mol Cell Biol. 6:328–340, Candi et al., © 2005.

Cysteine residues present in EDC proteins can form disulfide bonds which play a role in rendering appendages hard, insoluble and resistant. Most disulfide bonds are supposed to be intramolecular stabilizing proteins more than intermolecular involved in assembly (Fraser & MacRae, 1963; Filshie et al., 1964; Hohl et al., 1991; Parry et al., 2006), but in mammalian appendages intermolecular disulfide bonds are found (Gillespie, 1972; Dedeurwaerder et al., 1964). Cysteine-rich proteins like hair keratins and KRTAPs are known to be involved in the hard cornification of nail and hair (Wang et al., 2000; Eckhart et al., 2008; Alibardi et al., 2011; Langbein & Schweizer, 2005; Rogers et al., 2001; Powell & Rogers, 1997). It has been proposed that cysteine-rich proteins might have a similar role in the hard appendages of sauropsids (Eckhart et al., 2008, 2013). Notably, a pattern of cysteine residues duplets (CC) found in mammalian KRTAPs and believed to facilitate crosslinking (Parry et al., 2006), has also been found in the bird EDC protein EDCRP (Strasser et al., 2015). A good number of EDC proteins with cysteine duplets and even triplets have been observed in snakes as well (**Paper IV**).

The amino acid residue tyrosine (Y) contains an aromatic ring and is higly interactive either by ring stacking of the aromatic rings (McGaughey et al., 1998) or hydrogen bonding of the hydroxyl group (Levitt & Perutz, 1988). These characteristics endow tyrosine residues with physical properties that can strengthen the material in which they are present. In fact, segments rich in tyrosine and glycine are found in hard appendages of mammals (Gillespie, 1972, 1991) and in sauropsids the terminal domains of many CBPs contain tyrosine- and glycine-rich segments (Dalla Valle et al., 2008, 2009b; Gregg & Rogers, 1986). The combination of these two amino acids is believed to confer pliability by the glycine residues and strength through the tyrosine ones obtaining a glue like effect (Parry et al., 2006). Tyrosine-rich proteins called EDAAs (EDMTFs in birds) have been identified in turtles and archosaurs (Strasser et al., 2014; **Paper I** & **III**; **Chapter 2.6**) and also in squamates four EDC proteins with a high aromatic acid content were identified (**Paper IV)**.

In addition to their conserved motif, SFTPs have a conserved C-terminal sequence motif and are rich in arginine, glutamic acid, glutamine, glycine, serine, and/or histidine residues (Mlitz et al.,

2014; Henry et al., 2012). In both tricohyalin (THCC) and scaffoldin (SCFN) a carboxy-terminal domain and a high content of glutamine in the central domain has been conserved in amniotes (Mlitz et al., 2014). Expression of cornulin and SCFN in sauropsids was similar to expression of cornulin and THCC in human and mouse, namely in structures that will support the formation of hard appendages like the filiform papillae of the tongue, nails, feathers, and hair (Mlitz et al., 2014). So it seems that in both mammals and sauropsids their function, which is to offer scaffolds for the growth of diverse skin appendages such as claws, nails, hair, and feathers, is conserved.

1.7 Corneous beta proteins, previously termed "beta-keratins"

The best studied epidermal proteins of sauropsids are the corneous beta proteins (CBPs), previously also known as beta-keratins. These proteins have been analyzed in sauropsids in general (Alibardi, 2013a; Fraser & Parry, 2014; Alibardi et al., 2007, 2009; Toni et al., 2007; Sawyer et al., 2000) and specifically in birds (Wu et al., 2015; Ng et al., 2014; Greenwold & Sawyer, 2010, 2011, 2013; Alibardi et al., 2006; Gregg et al., 1984; Presland et al., 1989; Whitbread et al., 1991), in crocodiles (Greenwold & Sawyer, 2013; Dalla Valle et al., 2009c; Ye et al., 2010; Alibardi, 2003b, 2005a; Alibardi et al., 2006), in turtles (**Paper I**; Alibardi, 2005b, 2013c, 2014b; Dalla Valle et al., 2009b, 2013a; Li et al., 2013) and in squamates (**Paper IV;** Liu et al., 2015; Dalla valle et al., 2005, 2007a-b, 2009a, 2010; Staudt et al., 2012; Alibardi, 2012, 2013d-e, 2014a, c-f, 2015a-b).

First identified by their X-ray diffraction pattern that exhibited a beta sheet secondary structure as opposed to the keratin intermediate filaments that showed an alpha pattern deriving from their secondary alpha helical structure (Marwick, 1931; Astbury & Marwick, 1932; Rudall, 1947). Later under electron microscopy (Filshie & Rogers, 1962) it was shown that while the beta pattern consisted in filaments of 3.4 nm, the alpha keratin one had larger filaments of 8–10 nm. Afterwards the name beta keratins was assigned to these proteins as it was believed they were a special kind of hard keratin only found in sauropsids (Baden & Maderson, 1970; Fraser et al., 1972; Fraser & Parry, 1996). By now it has become clear that other than both being part of the epidermis

and forming filaments, KIFs and CBPs (beta keratins) have nothing in common (Alibardi. 2016a, Pp. 288-293). First of all, they do not share the typical alpha-helical core domain of the keratin superfamily, instead CBPs have a typical conserved core domain of 34 amino acids predicted to form a pleated beta sheet which characterizes them (Calvaresi et al., 2016; Fraser & Parry, 1996, 2011, 2014; Alibardi & Toni, 2007). Secondly, the gene structure of the CBPs is the same as most of the EDC proteins with two exons separated by an intron (Strasser et al., 2014), while the one of the KIFs is completely different containing seven to eight exons separated by introns (Eckhart et al., 2008; Lehnert et al., 1984). Furthermore, the exons of KIFs are all coding, but only one is in CBPs. Thirdly the CBPs are located as already mentioned on the Epidermal Differentiation Complex (EDC), whereas KIFs are both in mammals and sauropsids located in a different locus and even on a different chromosome in animals with a completely characterized genome (Greenwold et al., 2014; Vandebergh & Bossuyt, 2012). Finally, the length and consequently the molecular weight of KIFs and CBPs is quite different, namely KIFs generally range from 40-70 kDa while CBPs are low weight proteins ranging mainly between 8-25 kDa. Thus no evidence of evolutionary or biochemical relation exists and it has been proposed to change the incorrect and confusing name of beta keratins into corneous beta proteins (Alibardi et al., 2012).

Recently genes encoding CBPs have been identified as part of the EDC in all sauropsid orders apart from the Sphenodontia for which no genome is available yet (**Paper I** & **IV**; Strasser et al., 2014; **Chapter 2.6**). This makes it likely that the proposed common origin of all CBP diversity from a single ancestral sequence (Gregg & Rogers, 1986) was in the EDC. The number of CBPs that have been found in the homonymous named cluster ranges from 25 in the American alligator (**Chapter 2.6**) to 65 in chicken (**Paper I**), but could even be 71 in the Japanese gecko (Liu et al., 2015), if these CBPs will be confirmed to be all on the EDC. The absolute highest number of CBPs identified so far is found in birds (120) where many feather CBPs are located outside the EDC and related to the innovation of feathers as well as to the enormous feather phenotype variation (Alibardi, 2016b; Alibardi & Toni, 2008; Greenwold & Sawyer, 2010, 2013; Greenwold et al.,

2014; Ng et al., 2015; Wu et al., 2015). The same tendency of gene translocation was seen for the turtles and also in this case has been put into relation to a novelty skin appendage like the shell (**Paper I**; Greenwold & Sawyer, 2010; Li et al., 2013).

1.7.1 CBPs: structural, mechanical and physical properties

Even though characterized by their beta sheet secondary structure, the CBPs are not the only epidermal proteins containing beta sheets. But they are the only ones with antiparallel beta strands in the central domain that through beta turns give rise to the formation of a pleated and twisted beta sheet. Both X-ray diffraction experiments and molecular modeling have suggested that CBPs dimerize and that dimerization occurs between the beta sheets that pair up forming a β-sandwich (Fraser & Parry, 1996, 2017; Calvaresi et al., 2016). Van der Waals and hydrogen-bond interactions between beta sheets are thought to be responsible for intermolecular interactions of CBPs (Calvaresi et al., 2016; Cheng et al., 2013; Fraser & Parry, 2009). The dimers (β-sandwiches) represent the individual structural units that form through head-tail assembly filaments with a helical configuration in which four repeat units form one turn (Figure 11). Different conformations have been proposed for the filaments, one where the dimers axis is perpendicular to the one of the filament and sheets consist in three complete and two partial beta strands (Fraser & Parry, 2008, 2014, 2017), another where the dimers axis is parallel to the one of the filament and sheets are composed of four beta strands (Calvaresi et al., 2016, Figure 11).

While the central beta sheet domain is responsible for dimerization and forming fibrils of the helical structure, it is believed that the N-and C-termini of CBPs form bonds with other proteins constituting this way the matrix component of CBPs (Calvaresi et al., 2016; Alibardi, 2016a; Fraser & Parry, 1996, 2008, 2011, 2014). The fact that CBPs contain both the filament and matrix component in a single chain is what makes them unique with respect mammalian proteins where different protein types fulfill these functions (Filshie & Roger, 1962; Alibardi et al., 2006, 2009).

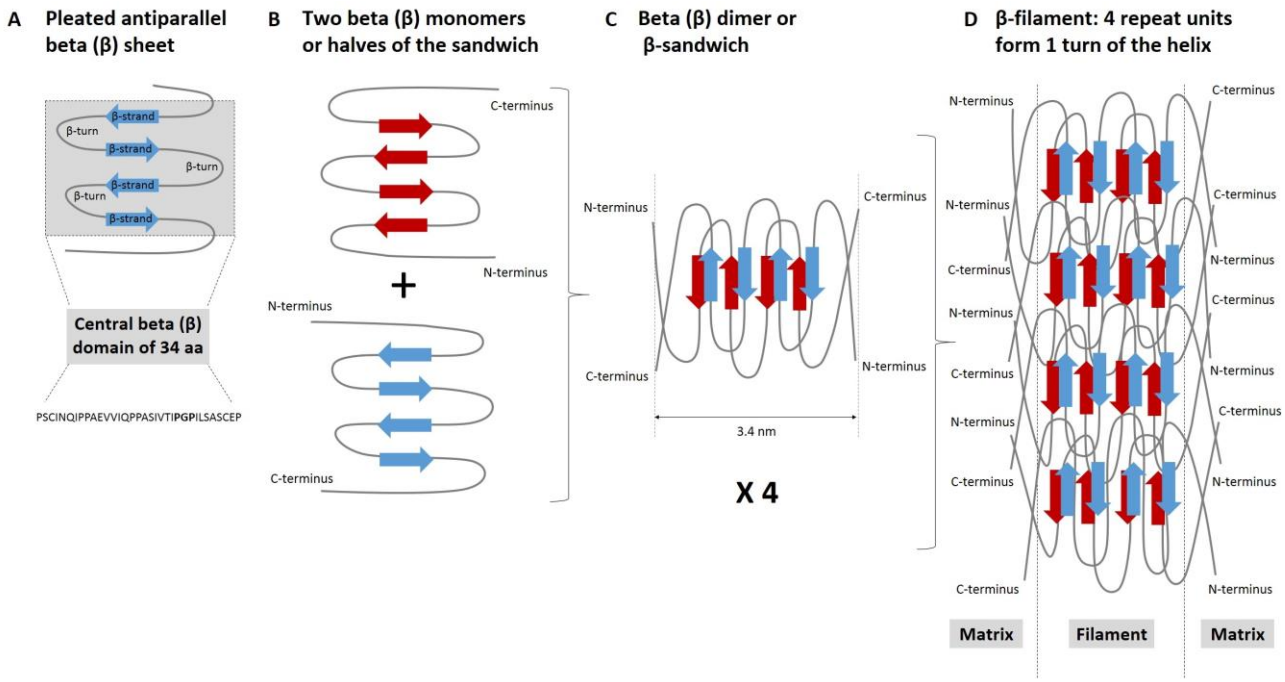


Figure 11. Simplified and schematical illustration of CBP structure, dimerization and filament formation. In **(A)** the CBP protein structure with pleated beta sheet in **(B)** two CBP monomers **(C)** a beta dimer composed of the two monomers shown in B and in **(D)** a beta filament composed of 4 structural repeat units as in C. The beta dimers form the filament and the C- and N- termini the matrix. Note: the illustration is in one dimension and does not show the twist of the beta sheet. The shown conformation and beta sheet is according to Calvaresi et al., 2016. For an alternative beta sheet and filament conformation see Fraser & Parry, 2017. For two-dimensional models of the proposed conformations see the mentioned papers.

The CBP cluster can be divided into subclusters based on specific CBP characteristics related to length, N-and C-termini, beta sheet domain and amino acid content. Terminal segments rich in specific amino acid residues can endow the tissue with particular physical properties (Fraser & Parry, 2017). For example some groups of CBPs contain terminal segments particularly rich in glycine, cysteine, tyrosine or proline residues or combinations of these which can attribute particular mechanical and physical properties to the proteins (Dalla valle et al., 2005, 2007a-b, 2009a-b, 2010; Hallahan et al., 2009; Alibardi et al., 2007, 2009; Gregg & Rogers, 1986). Repeated motifs rich in glycine and aromatic amino acids are found in the C-terminal domain of chicken scale and claw (Aral et al., 1983; Gregg et al., 1984; Whitbread et al., 1991) presumably conferring

favorable mechanical properties and hydrophobicity to these appendages. In extensions rich in charged and cysteine residues at the N-terminal domain, the charged residues are believed to make tissue softer and more flexible by increasing its water content while the ones with cysteine take care that the tissue's insolubility and resistance is not compromised (Fraser & Parry, 2014; Taylor et al., 2004; Watt & Leeder, 1968). On the other hand as already mentioned cysteine residues could play a role in disulfide binding providing insolubility and resistance to sauropsid appendages. CBPs rich in cysteine have been immunolocalized in lizard nails (Alibardi, 2015b; Alibardi & Toni, 2009). Disulfide bonds have been proved to be present in the squamate epidermis, especially in the beta and oberhautchen layer (Goslar, 1958; Matoltsy, 1962; Spearman & Riley, 1969; Baden et al., 1974; Banjeree & Mital, 1978; Mittal & Sing, 1987a-b). Interaction between KIFs and CBPs have been proposed in studies of various nature (Wyld & Brush, 1979, 1983; Alibardi &Toni, 2006; Toni et al., 2007; Ripamonti et al., 2009; Alibardi, 2013, 2014) and it has been suggested that some CBP interaction is mediated by disulfide bonds (**Paper V**; Fraser et al., 1972; Fraser & Macrae, 1978).

In brief, CBPs are believed to be responsible for the hard and resistant character of the sauropsid integument and their epidermal appendages with a different quantitative and qualitative ratio of CBP composition determine the degree of rigidness, hydrophobicity and poor staining to histological dyes (Alibardi, 2012, 2016a). Their presence is abundant in the sauropsid epidermis (O'Guin et al., 1987; Knapp et al., 1991; Sawyer et al., 2000; Alibardi & Toni, 2006, 2008; Toni et al., 2007), and in epidermal appendages like claw, beak, feathers and scales (Alibardi et al., 2006, 2007; Toni et al., 2007; Sawyer & Knapp, 2003; Gregg & Rogers, 1986; Sawyer et al., 1986; Presland et al., 1989). CBPs are likely to have an analogous function to the KRTAPs found in mammalian appendages (Alibardi, 2006, 2009, 2013) thus depicting a scenario of convergent evolution.

Although a lot of research has been performed on CBPs, still little is known about the regulation of their gene expression, the nature of their chemical or chemical-physical interactions

with other proteins and one another, their role during formation of the sauropsid epidermal barrier and epidermal appendages, and the molecular evolution of these proteins.

1.8 Skin defense mechanisms by proteins and peptides

The intact epidermis is a physical barrier to the attacks of microbes, but in case of loss of integrity and wounds an additional defense mechanism, the innate immune system, is put into action to prevent bacterial infection.

On the mammalian EDC we can find two proteins, the PGLYRP3 and 4 (Peptidoglycan Recognition Proteins 3 and 4), involved in protecting the epidermis against microbial aggression (Sun et al., 2006). In sauropsids EDC only one of these, the PGLYRP3, is present with exception of the chicken where none was found (Strasser et al., 2014; **Paper I** & **IV; Chapter 2.6**). In mammals these proteins have been found specifically expressed in organs exposed to the environment like epidermis, eyes, and mucous membranes (Lu et al., 2005; Sun et al., 2006). The protein's name explains its function in defending the epidermis, it recognizes the peptidoglycan typically present on bacterial cell walls. This receptor binds to the murein peptidoglycans (PGN) of Gram-positive bacteria and is bactericidal to these types of microbes. Even though it can also bind to other pathogens like Gram-negative bacteria and fungi, its effect exerted on these pathogens is only bacteriostatic. PGLYRP3 (and 4) kill bacteria by binding to the cell wall or outer membrane and exploiting the bacterial stress defense response. More precisely the bactericidal effect is mediated by a two component system inducing membrane depolarization and production of hydroxyl radicals (-OH) in the cytoplasm and this is accompanied by cessation of intracellular biosynthesis of DNA, RNA, peptidoglycan and proteins (Kashyap et al., 2011). This mechanism is different than the one general used by antimicrobial peptides that instead kill microbes by permeabilizing bacterial membranes (Lu et al., 2005; Kashyap et al., 2011).

Several antimicrobial peptides (AMPs) have been identified in reptiles (Stegemann et al., 2009; Chattopadhyay et al., 2006; Lakshminarayanan et al., 2005, 2008) and some specifically in

the skin (Dalla Valle et al., 2012, 2013b; Benato et al., 2013). These peptides are produced by both eukaryotes and prokaryotes as part of the innate immune system. They show a broad diversity in interactions and antimicrobial spectrum (Gram positive and negative bacteria, fungi and sometimes enveloped viruses and protozoa). Although most AMPs are amphipathic and positively charged, they show variability based on their structure and their amino acid composition. Several classes are recognized, like the defensins that contain 6 cysteines forming specific intramolecular disulfide bonds, and the cathelicidines that are linear cationic alpha-helical containing peptides (Ganz, 2003; Brogden, 2005). Characteristics like the net charge and hydrophobicity of AMPs are thought to determine their functionality and changes in these characteristics can influence both antimicrobial activity and selectivity (Zelezetsky et al., 2005).

Antimicrobial peptides have been well studied since they are believed to be the solution to the increasing resistance of bacteria to antibiotics. However, their application in the medical field has resulted more complicated than first thought. These peptides are very reactive and can interact with the testing medium and various environmental factors like pH, ionic strength and salt concentration (Walkenhorst et al., 2013; Wu et al., 2008; Lee et al., 1997; Friedrich et al., 1999; Wei et al., 2007). Therefore testing their efficiency has resulted highly dependent on the medium implied and conditions applied during trials. The exact physiological conditions under which these peptides function are often not known and cannot be reproduced during bioassays. In any case several studies have demonstrated the bactericidal effect of reptile AMPs (**Paper II**; Stegemann et al., 2009; Chattopadhyay et al., 2006; Lakshminarayanan et al., 2005, 2008).

It has been shown that these peptides function through various mechanisms; usually they attack the bacterial membrane causing the formation of pores, but they can also attack membrane bound or specific intracellular targets like DNA, RNA or protein synthesis (**Paper II**; Klüver et al., 2006; Falla et al., 1996; Boman et al., 1993; Hsu et al., 2005).

Chapter 2 Results

2.1 Results: Epidermal proteins of Testudines

2.1.1 Introduction to paper I: Comparative Genomics Identifies Epidermal Proteins Associated with the Evolution of the Turtle Shell.

The first paper reports the analysis of the epidermal differentiation complex (EDC) of the sauropsid clade of the Testudines. The EDC gene cluster encodes many structural proteins of the epidermis in mammals. Homologs of the EDC had been reported in the chicken and in the green anole lizard but in no other sauropsids prior to this PhD study.

We have investigated the hypothesis that the evolution of mechanically resilient modifications of the epidermis in turtles such as the scutes of the shell, might have been associated with specific adaptations of the EDC. Using newly available genome sequences of four turtle species, we show that the EDC of turtles comprises more than 100 genes, including at least 48 genes that encode corneous beta proteins (CBPs), also known as beta-keratins. Furthermore we identified several EDC proteins that contained cysteine/proline contents beyond 50% of total amino acid residues as well as conserved sequence motifs also present in loricrin, PGLYRP3 and small proline-rich proteins (SPRRs) of mammals. Distinct subfamilies of EDC genes have been expanded and partly translocated to loci outside of the EDC in turtles.

Experimental data obtained by RT-PCR analysis on tissues of *E. orbicularis* showed that EDC genes are differentially expressed in the skin of various body sites and that a subset of CBP genes within the EDC as well as a subset located outside of the EDC are expressed predominantly in the scutes of the turtle shell. Also other gene clusters that have undergone expansion like EDPCV and EDQM contain members predominantly expressed in the shell. Further confirmation of skin specific expression was obtained by screening the published transcriptome of *T. scripta*.

Phylogenetic analysis of the CBPs (beta keratins) indicated that those predominantly expressed in the scutes of the shell belong to subcluster A being similar to the chicken claw keratins as well as to turtle genes duplicated and translocated outside the EDC. The latter seem to have

originated from the cluster A on the EDC. This strengthens the hypothesis that duplication and translocation of members of this subcluster might be related to the origin and the specialization of the turtle shell.

The reported results suggest that the evolutionary novelty of the turtle shell involved specific molecular adaptations of epidermal differentiation as well as the origin and expansion of shell-related genes. Similarities in the overall structure of the EDC and amino acid sequence similarities of EDC-encoded proteins support the hypothesis that an EDC was already present in a common ancestor of turtles and mammals.

Comparative Genomics Identifies Epidermal Proteins Associated with the Evolution of the Turtle Shell

Karin Brigit Holthaus,[†,1,2] Bettina Strasser,[†,1] Wolfgang Sipos,[3] Heiko A. Schmidt,[4] Veronika Mlitz,[1] Supawadee Sukseree,[1] Anton Weissenbacher,[5] Erwin Tschachler,[1] Lorenzo Alibardi,[2] and Leopold Eckhart*[,1]

[1]Research Division of Biology and Pathobiology of the Skin, Department of Dermatology, Medical University of Vienna, Vienna, Austria

[2]Dipartimento di Scienze Biologiche, Geologiche ed Ambientali (BiGeA), University of Bologna, Bologna, Italy

[3]Clinical Department for Farm Animals and Herd Management, University of Veterinary Medicine Vienna, Vienna, Austria

[4]Center for Integrative Bioinformatics Vienna (CIBIV), Max F. Perutz Laboratories, Medical University of Vienna, University of Vienna, Vienna, Austria

[5]Vienna Zoo, Vienna, Austria

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: leopold.eckhart@meduniwien.ac.at.

Associate editor: Nicolas Vidal

Abstract

The evolution of reptiles, birds, and mammals was associated with the origin of unique integumentary structures. Studies on lizards, chicken, and humans have suggested that the evolution of major structural proteins of the outermost, cornified layers of the epidermis was driven by the diversification of a gene cluster called Epidermal Differentiation Complex (EDC). Turtles have evolved unique defense mechanisms that depend on mechanically resilient modifications of the epidermis. To investigate whether the evolution of the integument in these reptiles was associated with specific adaptations of the sequences and expression patterns of EDC-related genes, we utilized newly available genome sequences to determine the epidermal differentiation gene complement of turtles. The EDC of the western painted turtle (*Chrysemys picta bellii*) comprises more than 100 genes, including at least 48 genes that encode proteins referred to as beta-keratins or corneous beta-proteins. Several EDC proteins have evolved cysteine/proline contents beyond 50% of total amino acid residues. Comparative genomics suggests that distinct subfamilies of EDC genes have been expanded and partly translocated to loci outside of the EDC in turtles. Gene expression analysis in the European pond turtle (*Emys orbicularis*) showed that EDC genes are differentially expressed in the skin of the various body sites and that a subset of beta-keratin genes within the EDC as well as those located outside of the EDC are expressed predominantly in the shell. Our findings give strong support to the hypothesis that the evolutionary innovation of the turtle shell involved specific molecular adaptations of epidermal differentiation.

Key words: **turtles, skin, gene family, integument, gene duplication.**

Article

Introduction

Turtles are a clade of reptiles that have evolutionarily diverged from their next relatives, that is, the archosaurs (crocodilians and birds) approximately 240–260 Ma (fig. 1*A*; Iwabe et al. 2005; Kumar and Hedges 2011; Shaffer et al. 2013; Wang et al. 2013; Thomson et al. 2014; Bever et al. 2015; Crawford et al. 2015). The most important morphological innovation in the evolution of turtles has been the shell which is composed of skeletal, dermal, and epidermal elements that together form the ventral plastron and the dorsal carapace (Zangerl 1969). The complex evolution and development of the bony elements of the turtle shell have been extensively studied and reviewed (Ruckes 1929; Burke 1989; Reisz and Head 2008; Nagashima et al. 2009; Hirasawa et al. 2013, 2015; Rice et al. 2015). The epidermal components of the shell are the scutes in hard-shelled turtles and the largely unpatterned epidermis in soft-shelled turtles (Thomson et al. 2014). The latter have lost both scales, an ancestral trait of reptiles, and scutes, which are generally considered to be derived from scales (Alibardi and Thompson 1999; Thomson et al. 2014). Other important epidermal structures of turtles are the claws, which are shared with other amniotes (Alibardi 2003, 2014) and the rhamphotheca, a horny sheath covering the mandibles that functionally compensates the absence of teeth in turtles. The molecular basis for the evolution of epidermal structures in turtles is only beginning to emerge (Dalla Valle et al. 2009; Li et al. 2013; Moustakas-Verho et al. 2014, 2015).

The epidermis of vertebrates is a stratified epithelium in which cells of the basal layer proliferate and start to differentiate upon detachment from the basement membrane that separates the epidermis from the underlying dermis. Keratinocyte differentiation involves the transcriptional upregulation of genes that encode structural proteins and

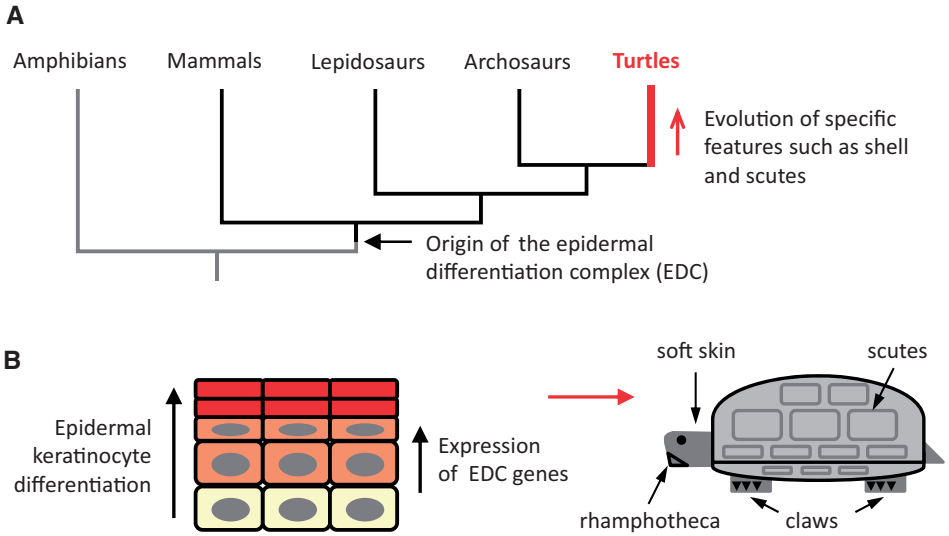


Fig. 1. Schematic overview of the phylogenetic position of turtles and keratinocyte differentiation in the epidermis of turtles. (*A*) Phylogenetic tree of turtles and other vertebrates. (*B*) Diagram of the epidermis of turtles and other amniotes. Keratinocytes proliferate in the basal layer (yellow) and, upon transition into suprabasal layers, undergo a differentiation program that ultimately converts living cells into dead components of the cornified layer (red) (left panel). Variations of the gene expression program during differentiation lead to various epidermal structures of turtles, such as the scutes of the shell (right panel).

the passive movement of cells toward the skin surface. Ultimately, keratinocytes undergo cornification, a mode of programmed cell death (Eckhart et al. 2013) that generates mechanically rigid and interconnected cell corpses (corneocytes) (fig. 1*B*). Although the molecular determinants of epidermal differentiation have been characterized only incompletely in turtles, it can be inferred from comparison with other amniotes (Strasser et al. 2014) that the epidermal features of turtles are a consequence of specific adaptations of the process of keratinocyte differentiation.

In mammals, many of the components of the cornified protein envelope of corneocytes are encoded by genes of a gene cluster known as the Epidermal Differentiation Complex (EDC) (Mischke et al. 1996). The human EDC comprises genes encoding S100A proteins, peptidoglycan recognition proteins (PGLYRP), simple EDC (SEDC) genes with one noncoding and one coding exon such as loricrin, involucrin, and small proline-rich proteins, and S100 fused-type proteins (SFTPs) such as cornulin, trichohyalin, and filaggrin (Henry et al. 2012; Kypriotou et al. 2012).

Recently, we have shown that a gene cluster with the same basic organization is also present in two sauropsidian model species, the chicken and the green anole lizard (Strasser et al. 2014). Moreover, in the above study we demonstrated that these genes are specifically expressed in epidermal keratinocytes. Loricrin contributes to the formation of the skin barrier not only in mammals but also in lizards (Strasser et al. 2014). SFTPs are expressed in human and avian epithelia that function as scaffolds for growing skin appendages such as claws, hair, and feathers (Mlitz et al. 2014). Recently, a new epidermal differentiation cysteine-rich protein (EDCRP) has been detected as a component of avian feathers (Strasser et al. 2015). Importantly, gene locus synteny (Vanhoutteghem et al. 2008; Strasser et al. 2014) and conservation of exon–intron organization (Strasser et al. 2014) have led to the hypothesis that the beta-keratins, which are widely considered the main epidermal proteins of sauropsids (Fraser and Parry 1996, 2014; Alibardi et al. 2009), have originated in the EDC and represent a sauropsid-specific subtype of SEDC gene products (Strasser et al. 2014). It is important to note that the term "beta-keratins" indicates neither common ancestry nor sequence similarity to "keratins" in the sense used by the Gene Nomenclature Committee. The latter group of proteins was originally named "alpha-keratins" and belongs to the intermediate filament protein superfamily (Schweizer et al. 2006). We advocate the renaming of beta-keratins to "corneous beta-proteins" or another term without the misleading word keratin, but we will use the traditional term here to link our report to the previous literature on skin proteins of turtles. The phylogeny of beta-keratins in turtles has been recently reported (Li et al. 2013); however, the role of the EDC in the evolution of the unique integument of turtles has remained elusive.

Here, we report the identification of the genes that constitute the EDC in turtles, the investigation of EDC gene expression in a turtle model species, and comparative analyses that suggest evolutionary trajectories for the main types of EDC genes in turtles. Our results reveal that the evolution of turtles involved expansions of gene families within the EDC, translocations of beta-keratin and other genes to novel loci outside of the EDC, and adaptations of EDC gene expression patterns to turtle-specific integumentary structures.

Results

The Basic Organization of the EDC Is Conserved in Turtles

To investigate the presence and organization of the EDC in a representative turtle species, we used the published genome sequence of the western painted turtle, *Chrysemys picta bellii*

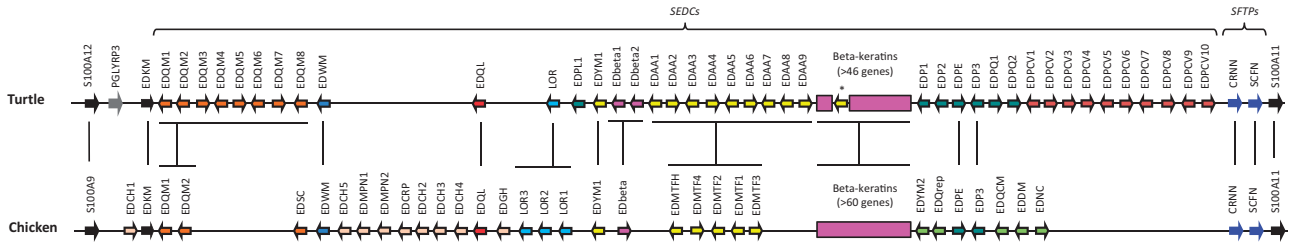


Fig. 2. Organization of the EDC in the turtle *Chrysemys picta* in comparison to that of the chicken. Genes of the EDC in chicken (chromosome 25) and the turtle *C. picta* are schematically depicted. Arrows indicate the orientation of the genes. SEDC genes with two exons are represented by colored arrows with a black frame whereas other genes are shown as filled arrows. Clusters of beta-keratin genes are shown as boxes (for more detailed information about beta-keratins, see supplementary fig. S13, Supplementary Material online). The gene *EDAA10* (*) is located within the beta-keratin gene cluster of the turtle. Colors indicate families of genes as defined in the text. Numbers indicate the position of genes within each family cluster but not 1:1 orthology to specific members of the same gene family in other species. Black vertical lines connect orthologous genes or gene families. Note that the schemes are not drawn to scale.

(Shaffer et al. 2013), and determined the set of genes located between the homologs of S100A12 and S100A11 genes. Automatic gene prediction algorithms had failed to correctly annotate many EDC genes of the chicken and lizard (Strasser et al. 2014), and were also not considered reliable for *C. picta*. Therefore, we used the existing gene annotations for S100A and PGLYRP genes only, and performed tBLASTn searches with the amino acid sequences of human, chicken, and lizard EDC-encoded proteins (Strasser et al. 2014) and predicted additional genes of the SEDC type by screening conceptual translations of the EDC nucleotide sequence. Iterative rounds of gene searches were performed in which newly predicted amino acid sequences were used as query sequences for the tBLASTn searches.

The EDC of the western painted turtle has an organization of largely shared synteny with that of the chicken (Strasser et al. 2014; fig. 2). Besides 12 S100A genes and *PGLYRP3*, we identified a homolog of *EDKM*, 90 SEDC genes (including five partial genes) and 2 SFTP genes on the EDC scaffold (GenBank accession number NW_007281429.1) of the *C. picta* genome (supplementary tables S1 and S2 and fig. S1, Supplementary Material online). Names and abbreviations were tentatively assigned to these genes according to a preliminary nomenclature system for sauropsidian EDC genes (Strasser et al. 2014; supplementary table S1, Supplementary Material online). In addition to the SEDC genes on the EDC scaffold, we identified SEDC gene homologs at two genome loci outside of the EDC as well as on several short scaffolds that did not contain any other genes than SEDCs. Because the scaffold containing the great majority of EDC genes has several sequence gaps, it is possible and even likely that some of the latter scaffolds have not yet been integrated into their correct position within the EDC and that the number of genes within the EDC is higher than that on the genomic scaffold mentioned above. Details on the SEDC genes identified at non-EDC loci are provided below.

The gene loci identified in *C. picta* were compared to those of three other turtles of which genome sequences were available in GenBank, that is, *Chelonia mydas*, *Pelodiscus sinensis*, and *Apalone spinifera*. These comparisons showed a similar organization of the EDC in *Che. mydas* and *P. sinensis* (supplementary tables S3 and S4 and figs. S2 and S3, Supplementary Material online) whereas the fragmented genome sequence assembly of *A. spinifera* did not allow alignments of sufficient length (not shown).

Proteins Encoded by Turtle EDC Genes Have Evolved Extreme Biases in Amino Acid Compositions and Highly Repetitive Sequences

The newly identified EDC gene sequences of turtles were translated in silico (supplementary figs. S1 and S2, Supplementary Material online) and the resulting amino sequences were analyzed for features that might be associated with the presumable function of the encoded proteins in the epidermis of turtles. As previous studies have suggested that the evolution of the EDC has generated SEDC proteins with highly diverse amino acid compositions (Strasser et al. 2014), we determined the amino acid contents of SEDC proteins in *C. picta*. Indeed, many SEDC proteins of *C. picta* have extremely high contents of either glycine and serine, or cysteine and proline (fig. 3*A*–*C*), and, in addition, contain lysine and glutamine residues which are supposed to be the sites of protein cross-linking via transglutamination (Strasser et al. 2014). Remarkably, the combined content of cysteine and proline exceeded 50% of the total amino acid residues in several SEDC proteins. The genes encoding glycine/serine-rich proteins were clustered in one half (fig. 2) of the EDC whereas the genes encoding cysteine/proline-rich proteins were clustered in the other half (fig. 2) of the EDC, indicating that they arose by tandem duplication events. Another group of genes encoding proteins rich in aromatic amino acids, particularly histidine and tyrosine (supplementary fig. S4, Supplementary Material online), is located in the central region of the EDC. These genes are likely homologous to chicken genes that were previously named "epidermal differentiation proteins starting with the MTF motif" (EDMTFs) (Strasser et al. 2014). For the turtle homologs of EDMTFs, we propose the name epidermal differentiation proteins rich in aromatic amino acids (EDAAs). Beta-keratins, as defined by the presence of a 34-amino acid residue segment with high propensity to form beta-sheets (Fraser and Parry 1996, 2014; Alibardi et al. 2009), are encoded by SEDC genes located on both sides of the EDAA cluster. The amino-terminal portion of most beta-keratins does not have an extreme bias in the

amino acid content whereas the carboxy-terminal portion is typically rich in glycine and tyrosine (fig. 3*D*).

Among the two SFTPs of *C. picta*, cornulin is rich in proline (18%), glutamine (10%), and glutamic acid (14%) whereas scaffoldin is rich in glutamic acid (~24%), arginine (~22%), and proline (~18%; the percentage numbers are not accurate because the gene has not been completely sequenced). In many SEDC proteins (fig. 3*B* and *C*) and in both SFTPs (supplementary fig. S5, Supplementary Material online), the amino acid sequences are dominated by repeats, possibly representing the products of inequal crossovers during the evolution of EDC genes (Strasser et al. 2014). Proteins encoded by genes at various positions distributed over the entire length of the SEDC gene cluster of *C. picta* contain conserved sequence motifs at their amino and carboxy-terminus (supplementary fig. S6, Supplementary Material online), similar to diverse proteins encoded by EDC genes of humans, chicken, and lizard (Strasser et al. 2014). The conservation of lysine and glutamine residues, that is, the target amino acids of transglutamination (Strasser et al. 2014), suggests that protein cross-linking via transglutamination is a conserved feature of EDC proteins. Common exon–intron structure, a gene arrangement compatible with an evolution by tandem duplications, and the presence of conserved sequence elements at the amino- and carboxy-termini of many (but not all, e.g., beta-keratins) SEDC proteins, support the hypothesis that SEDC genes have originated from a single or only few ancestral gene(s) (Strasser et al. 2014). The amino acid sequences of turtle SEDC proteins exemplify the remarkable sequence diversification that has accompanied the evolution of epidermal proteins in amniotes (fig. 3*E*).

Gene Duplications and Translocations Have Generated Families of SEDC Genes Both Inside and Outside the EDC of Turtles

To allow for hypotheses on the evolutionary history of individual EDC genes of turtles, we next compared the amino acid sequences of proteins encoded by genes along the EDC. Classical approaches of molecular phylogenetics were deemed not applicable for most EDC genes because of the highly repetitive nature of amino acid sequences and because of the biased amino acid compositions of the encoded proteins, which precluded unambiguous sequence alignments. However, we performed a phylogenetic analysis of beta-keratins (see below).

We found that a large portion of the EDC of *C. picta* was comprised by five distinct gene types, namely those encoding EDQMs (Epidermal Differentiation proteins containing a glutamine (Q) Motif) (supplementary fig. S7, Supplementary Material online), EDAAs (supplementary fig. S8, Supplementary Material online), EDP (Epidermal Differentiation proteins rich in Proline)-like proteins, EDPCVs (Epidermal Differentiation proteins rich in Proline, Cysteine and Valine) (supplementary fig. S9, Supplementary Material online), and beta-keratins (supplementary fig. S10, Supplementary Material online). Only the existence of the latter proteins of turtles and their homology to proteins of the chicken was reported previously (Dalla Valle et al. 2009; Li et al. 2013). Orthologs of *EDQM*, *EDAA*, and *EDP*-like genes are also present in the chicken, whereas turtle *EDPCV* genes appear to lack counterparts in the chicken (fig. 2).

The number of *EDQM* genes was higher in *C. picta* ($n = 8$) than in chicken ($n = 2$), suggesting a lineage-specific expansion of this gene family. Similarly, the number of *EDAA* genes in *C. picta* ($n = 22$) was higher than the number of the homologous *EDMTF* genes in the chicken ($n = 5$). Unexpectedly, BLAST searches identified a locus (between genes encoding SLAMF8 and NLRPs) outside of the EDC that contained *EDAA* genes (supplementary fig. S11, Supplementary Material online). This locus was conserved in *Che. mydas* and *P. sinensis*, however in the latter only *EDAA* genes carrying premature stop codons or frameshift mutations could be identified. This pattern of *EDAA* gene loci is compatible with the hypothesis that *EDAA* genes originated within the EDC, and *EDAA* copies were translocated next to the *SLAMF8* locus (supplementary fig. S11, Supplementary Material online) in the stem lineage of turtles. Fifteen *EDPCV* genes were identified in *C. picta*, whereas only four *EDPCV* genes were found in the soft-shelled turtle *P. sinensis*. In the latter we identified a scaffold (GenBank accession number NW_005854374.1) that contained *EDPCV* genes as well as the gene *Natural killer cell receptor 2B4-like*, suggesting that this scaffold is not part of the EDC. As neither *C. picta* nor *Che. mydas* had *EDPCV* genes at syntenic loci, it is likely that the *EDPCV* gene cluster has undergone a rearrangement, possibly a translocation of a subset of its genes, in *P. sinensis*.

The largest family of SEDC proteins of the turtles are the beta-keratins. In total, we identified 82 complete and more than 10 partial beta-keratin genes in the genome of *C. picta*. Sequence alignments showed that there were subfamilies with characteristic sequence motifs (supplementary fig. S10, Supplementary Material online). Comparisons of beta-keratin gene loci of *C. picta*, *Che. Mydas*, and *P. sinensis* and genomes of other vertebrates demonstrated that some of the beta-keratin genes of the turtles are located adjacent to the gene *ODF3B* outside of the EDC (supplementary fig. S12, Supplementary Material online). No other vertebrates have beta-keratin genes at this locus, suggesting that this beta-keratin gene cluster originated specifically in the evolutionary lineage leading to modern turtles. The beta-keratins encoded by genes at this locus (tentatively named Beta-O proteins, whereby O indicates the location of the genes "outside of the EDC"), are most closely related to beta-keratins encoded by a subcluster (tentatively named Beta-A) of the beta-keratin gene cluster in the EDC (supplementary fig. S13, Supplementary Material online). Within the EDC, the Beta-A gene cluster is flanked by the Beta-B cluster of beta-keratins for which we could not identify close homologs outside of the EDC. The cluster of Beta-A genes of the turtle is syntenic to "claw beta-keratins" (figure 3 in Greenwold et al. 2014) of the chicken (designated "Beta claw" in supplementary fig. S13*A*, Supplementary Material online). Phylogenetic analysis suggests that beta-keratins of the Beta-A plus Beta-O clade of turtles and claw, feather, and scale beta-keratins of the chicken form four separate strongly supported monophyletic

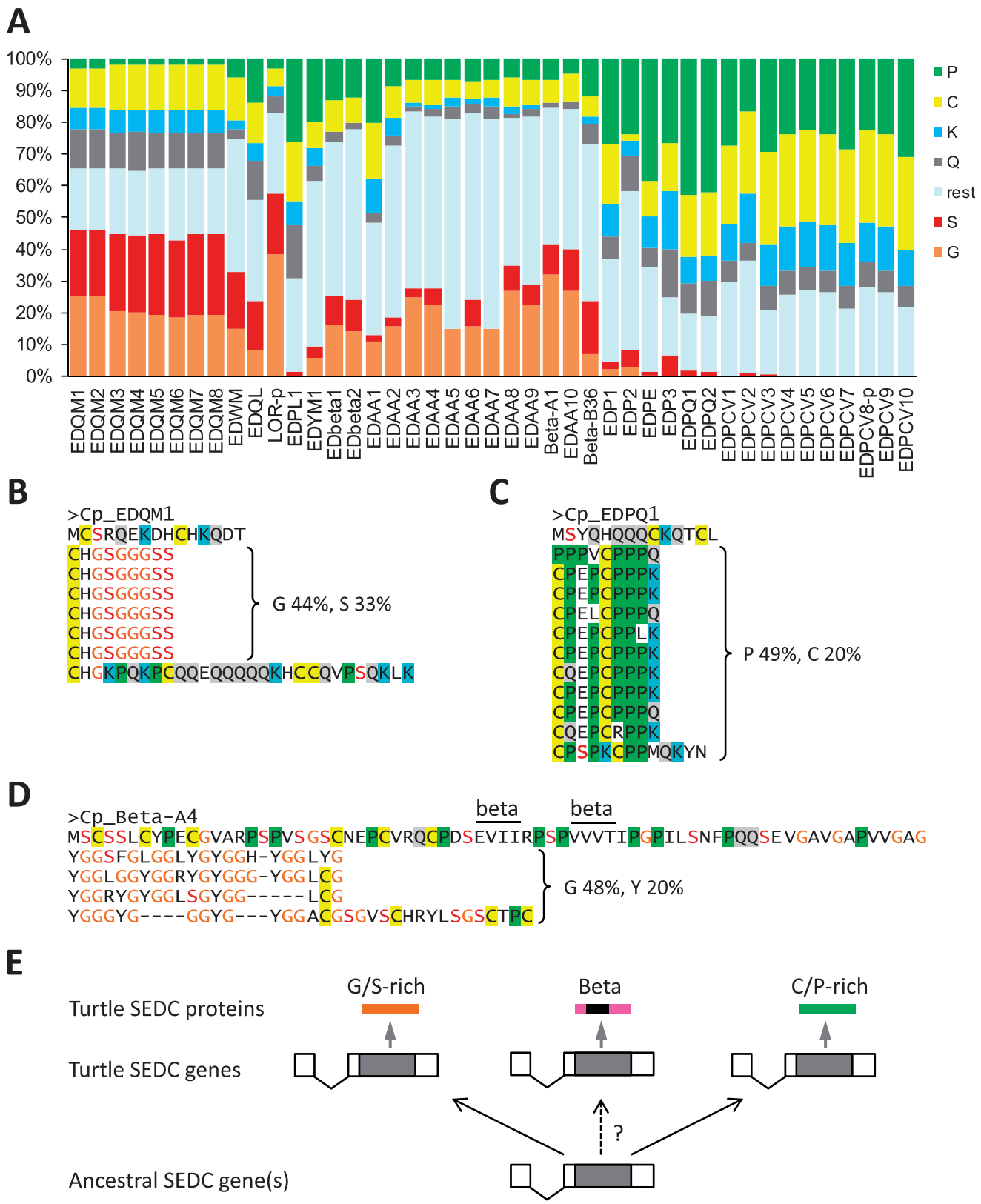


Fig. 3. SEDC genes encode proteins with extremely biased amino acid composition. (*A*) The diagram shows the amino acid compositions of SEDC proteins of *Chrysemys picta*. The protein data are shown in the order of the corresponding genes in the EDC (fig. 2). Note that out of the main beta-keratin gene cluster, only the translation products of the first and the last gene are included here. (*B–D*) Amino acid sequences of exemplary SEDC proteins. The positions of two predicted beta-sheets in Beta-A4 are indicated. (*E*) Schematic depiction of the evolutionary diversification of SEDC genes from a common ancestral gene.

groups. Furthermore, these groups cluster together to the exclusion of the other beta-keratins (supplementary fig. S13*B*, Supplementary Material online). Together with the localization of Beta-A genes within the phylogenetically ancient beta-keratin subcluster of the EDC (supplementary fig. S13, Supplementary Material online), the strong support for the joined subtree of Beta-A and Beta-O proteins suggests that the cluster of Beta-O genes arose by translocation of one or more ancestral genes from the Beta-A gene cluster, followed by gene duplications.

In addition to the above-mentioned gene families, the EDC of turtles contains several individual genes that are orthologous to EDC genes of the chicken and other amniotes (Strasser et al. 2014). Like the EDCs of the lizard and human but different from that of the chicken, the turtle EDC contains a *PGLYRP3* gene. The western painted turtle has a single *LOR* gene (fig. 2, supplementary fig. S3, Supplementary Material online) whereas the chicken has three (Strasser et al. 2014). Both in turtle and chicken, *LOR* is flanked by a gene, tentatively named *EDQL* (previously named *EDQM3* in chicken (Strasser et al. 2014)), that encodes a protein with a carboxy-terminus highly similar to that of loricrin (supplementary fig. S14*A* and S6 and table S1, Supplementary Material online). *EDWM*, an SEDC gene present in all sauropsids investigated so far (Strasser et al. 2014) is conserved in the hard-shelled turtles *C. picta* and *Che. mydas* but has acquired mutations that destroy its open reading frame in the soft-shelled turtles *P. sinensis* and *A. spinifera* (supplementary fig. S15, Supplementary Material online). *EDCRP* (Strasser et al. 2015) and other genes encoding extremely cysteine-rich proteins are absent between the *EDWM* and *LOR* genes of the turtle whereas they are present at this site of avian EDCs (fig. 2). *EDP3* genes were identified in *C. picta* and chicken (supplementary fig. S14*B*, Supplementary Material online). Most of the SEDC genes of *C. picta* had orthologs with highly conserved sequences in *Che. mydas* and *P. sinensis* (supplementary fig. S16, Supplementary Material online). However, the numbers of genes in the SEDC subfamilies of *EDQM* and *EDPCV* genes differed (supplementary fig. S3, Supplementary Material online), and SEDC genes containing multiple internal sequence repeats, such as *LOR* and *EDPE*, could not be faithfully predicted for *Che. mydas* and *P. sinensis* because of uncertainties in the genomic sequence assembly (supplementary fig. S3, Supplementary Material online, and data not shown). Thus, the evolution of individual EDC genes in the diverse subclades of turtles remains to be investigated in future studies.

Together, these data suggest that the EDC genes underwent differential evolution in the lineages leading to turtles and other sauropsids, with many genes being conserved and some genes undergoing repeated rounds of tandem duplication events to give rise to turtle-specific expansions of gene families.

EDC Genes Are Differentially Expressed in the Shell and Other Integumentary Structures of the European Pond Turtle

To test whether the predicted EDC genes are expressed, we investigated RNA-seq data of *C. picta* and *P. sinensis* (available in the National Center for Biotechnology Information (NCBI) databases, Materials and Methods) and screened the published transcriptome sequence reads of the red-eared slider turtle (*Trachemys scripta*) (Kaplinsky et al. 2013). The available RNA-seq information from *C. picta* did not include specific samples from skin, nevertheless we found sequence reads indicating expression of the predicted exons of *EDP3*, *EDPQ1/2*, and two *EDPCV* genes (Shaffer et al. 2013) (supplementary table S2*A*, Supplementary Material online). RNA-seq data from *P. sinensis* (Wang et al. 2013) demonstrated expression of most predicted EDC genes (supplementary table S4*A*, Supplementary Material online) and suggested transcriptional upregulation of these genes during the developmental maturation of the epidermis (supplementary fig. S17, Supplementary Material online). The analysis of the transcriptome data from *T. scripta* (Kaplinsky et al. 2013) confirmed expression of homologs of all genes investigated, including *cornulin*, *scaffoldin*, *EDKM*, *loricrin*, *EDQL*, and *EDPE* in the embryo of *T. scripta*. However, these data did not allow assigning the transcripts to particular tissues and body sites.

Therefore, we studied EDC gene expression in freshly prepared turtle tissues. Because *C. picta* was not available to us, 45-days old embryos of the European pond turtle (*Emys orbicularis*) from a breeding program at the Vienna Zoo were investigated. Representative histological images illustrating the epidermal layers and fully cornified skin structures present at this embryonic stage are shown in supplementary figure S18, Supplementary Material online, Supplementary Material online. Muscle, kidney, tongue (without cornifying keratinocytes), and nose/rhamphotheca, skin of neck, tail, toes including claws, carapace, and plastron (with cornifying keratinocytes) were subjected to RNA extraction and reverse transcription polymerase chain reaction (RT-PCR) analyses using primers that were designed to anneal to the predicted exons 1 and 2 of EDC genes of *C. picta*. With the exception of primers specific for *EDPE*, all the other PCRs that we performed on the cDNAs derived from different tissues of *E. orbicularis* gave single products that could be purified and sequenced (supplementary fig. S19*A* and *B*, Supplementary Material online). Alignment of cDNA sequences of *E. orbicularis* to the predicted mRNA sequences of *C. picta* confirmed the specificity for the intended targets and revealed a high degree of sequence conservation between *E. orbicularis* and *C. picta* (supplementary fig. S19*C*, Supplementary Material online). A PCR with primers specific for the housekeeping gene *GAPDH* confirmed that all preparations of tissue samples contained cDNAs accessible for PCR amplification, though differences in cDNA amounts allowed only for semiquantitative comparisons of gene expression (fig. 4, lowermost panel). A cDNA preparation from the nose and rhamphotheca (rhinotheca) of the turtle embryos contained transcripts of all the genes investigated whereas other tissues contained only transcripts of a subset of genes. The physiological significance of the broad gene expression in the skin of the nose and/or rhamphotheca is unknown.

All genes localized in the EDC were expressed in tissues that contained epidermal keratinocytes (fig. 4). Likewise, *EDAA* genes located outside the EDC (*EDAA-O*)

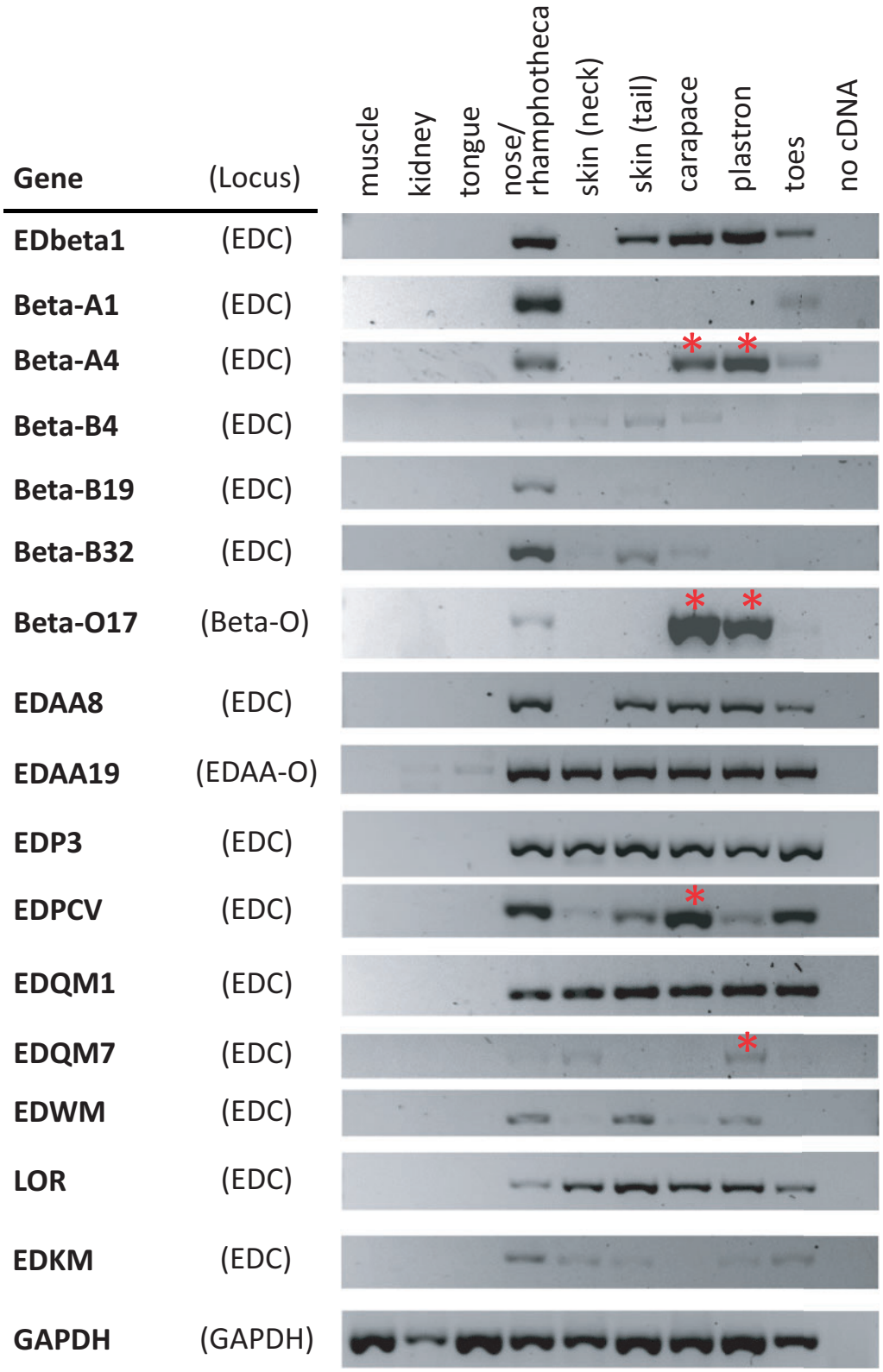


FIG. 4. EDC genes are differentially expressed in the skin of different body sites of the European pond turtle. The expression of EDC genes was determined by RT-PCR in embryonic tissues of the European pond turtle (*Emys orbicularis*). Intron-spanning primers were designed using the sequences of the EDC genes of *Chrysemys picta* and *Chelonia mydas*. The RT-PCR products were sequenced and their identity was determined by identifying the best sequence matches with EDC genes of *C. picta* (supplementary fig. S19, Supplementary Material online). Red asterisks mark transcripts that are predominantly expressed in the shell (carapace and/or plastron).

(supplementary fig. S11, Supplementary Material online) and beta-keratin genes outside the EDC (*Beta-O*) (supplementary fig. S12, Supplementary Material online) were essentially confined to tissues in which keratinocytes cornify (fig. 4). Transcripts of several EDC genes (*LOR*, *EDQM1*, *EDP3*, *EDAA19*) were detected in the skin of all body sites whereas some genes showed differential expression at the various regions of the body surface. Among beta-keratins, EDbeta1 showed a relatively wide expression pattern whereas Beta-A1 was expressed only in the nose/mouth region and the toes, perhaps indicating a role in the hard cornification of the rhamphotheca and the claws, respectively. Intriguingly, the transcripts tentatively named Beta-A4, originating from a gene within the Beta-A subcluster of the beta-keratin gene cluster of the EDC (supplementary fig. S13A, Supplementary Material online), and Beta-O17, which corresponds to a beta-keratin located outside the EDC, were present at the highest levels of expression in the carapace and the plastron. In particular, Beta-O17 was essentially specific for the shell because RT-PCR products from the nose/rhamphotheca and the toes were much weaker than those from the carapace and the plastron (fig. 4, uppermost panel). In summary, the expression analysis of EDC and EDC-related genes of *E. orbicularis* demonstrated that most genes are differentially expressed at various body sites and some of these genes, including beta-keratins of the Beta-A and Beta-O families as well as distinct SEDC genes different from beta-keratins, are expressed predominantly in the shell (fig. 4, red asterisks).

Discussion

The results of this study suggest that the evolution of the unique morphology of turtles involved specific adaptations of epidermal differentiation genes located in, or originating from the amniote-specific gene cluster known as EDC (Strasser et al. 2014). A scenario for the evolution of the EDC in turtles is schematically depicted in figure 5. According to this model, the basic organization of the EDC was inherited from a common ancestor of turtles and their next relatives, the archosaurs. In the lineage leading to turtles, *EDAA* and beta-keratin genes were independently translocated to loci outside the EDC. The *EDQM* and *EDPCV* gene families as well as *EDAA* and beta-keratin genes both within and outside the EDC expanded by repeated gene duplications. Furthermore, many EDC genes acquired differential expression patterns in various skin structures. We propose that some EDC genes, including a subset of beta-keratin genes (members of the Beta-A cluster), and beta-keratin genes at the locus outside of the EDC (Beta-O) evolved a predominant expression in scales of the dorsal and ventral aspects of the body where they contributed to the evolution of the hard scutes of the shell.

EDC genes encode structural proteins of epidermal keratinocytes (Henry et al. 2012; Kypriotou et al. 2012; Eckhart et al. 2013). In particular, proteins encoded by SEDC genes are supposed to exert their function by becoming cross-linked components of mechanically resilient structures at the skin surface (Candi et al. 2005; Eckhart et al. 2013). The relative abundance and the type of molecular interactions of individual proteins likely modulate the physicochemical parameters of cornification products such as the pliable cornified layer of the "soft" epidermis and the more rigid scutes of the shell. Our data suggest that SEDC protein families with very different amino acid contents have expanded during the evolution of turtles, namely EDQMs (containing a characteristic stretch of glutamine residues), EDPCVs (rich in proline and cysteine residues), EDAAs (rich in aromatic amino acids), and beta-keratins. The distinct sequence features of these protein families might facilitate different types of interactions with other structural proteins of cornifying keratinocytes, including keratins, cytolinkers, and cell junction proteins that are encoded by genes at loci outside of the EDC (Niessen 2007; Vandebergh and Bossuyt 2012; Wiche et al. 2015). Glutamine and cysteine residues (present in EDQMs and EDPCVs) are the main sites of intermolecular cross-linking of EDC proteins via transglutamination and disulfide bond formation, respectively (Kalinin et al. 2002; Eckhart et al. 2013; Rice et al. 2013). Stretches of glycine residues, located between transglutamination sites of EDQM proteins possibly allow for flexible changes in protein length that are supposed to contribute to the compaction of the cellular protein envelope during keratinocyte cornification (Candi et al. 2005). Aromatic amino acid residues (highly abundant in EDAAs and in the carboxy-terminal portion of beta-keratins) are potential sites of the non-covalent protein interaction mode termed pi-stacking (McGaughey et al. 1998; Waters 2002). Together with the emerging data on EDC proteins of other amniotes (Henry et al. 2012; Strasser et al. 2014; our unpublished data), the results of the present study provide the basis for theoretical and experimental studies on the molecular interactions that determine the epidermal phenotypes of amniotes.

The expression of EDC genes at the various body sites of turtles was investigated by semiquantitative RT-PCR analyses using *E. orbicularis* as a model species. This approach had several limitations such as the restricted availability of tissue samples which did not allow the analysis of biological replicates. Nevertheless, our results allow the conclusion that many turtle EDC genes are expressed in the skin of more than one body site. This is true for beta-keratins of the cluster B (within the EDC), loricrin, EDP3, EDAA, and at least one EDQM gene. However, our data also identify EDC genes expressed predominantly in the shell (*Beta-A4*) and, in some cases, predominantly in the carapace (*EDPCV*, assignment of this *E. orbicularis* RT-PCR product to an individual *EDPCV* gene family member was not possible) or the plastron (*EDQM7*) (fig. 4). The association of gene expression with the shell was most obvious for two beta-keratins investigated, one belonging to the Beta-A cluster (within the EDC) and the other belonging to the Beta-O cluster (outside the EDC). These findings suggest a specific role for these beta-keratins in the scutes of the shell but also indicate that other SEDC genes have contributed to the evolution of the shell.

The data presented here complement and extend previous studies on the roles of beta-keratins in the evolution of turtles. Beta-keratins, also referred to as corneous beta-proteins (Alibardi et al. 2009) to indicate their lack of common ancestry with keratins (Schweizer et al. 2006), are encoded by genes

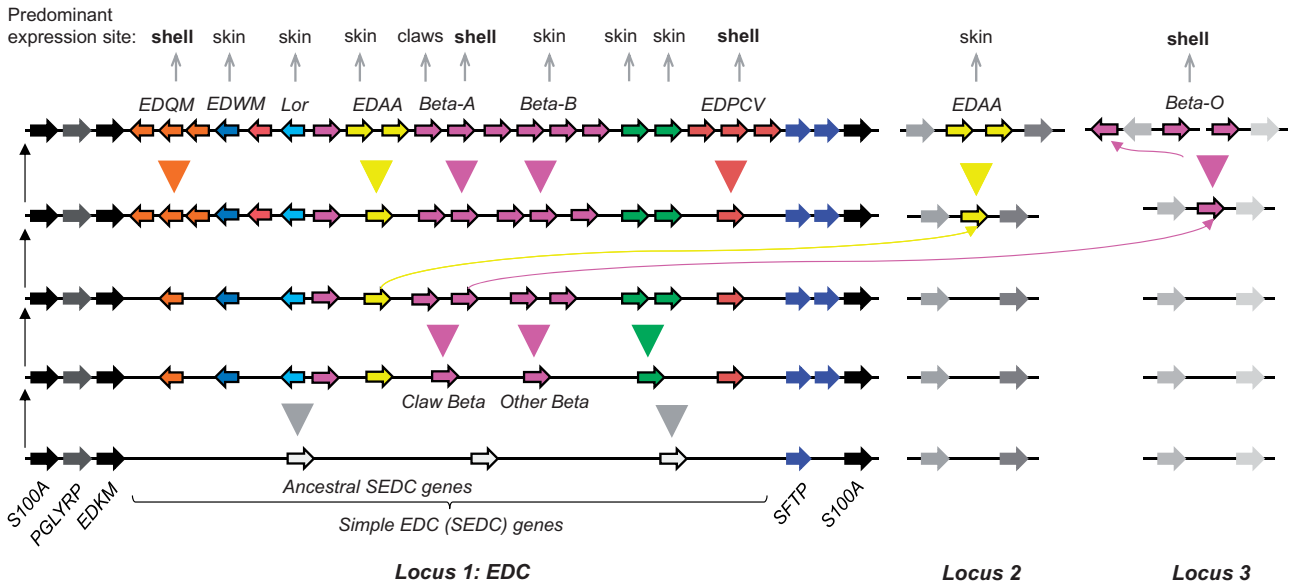


Fig. 5. A scenario for the evolution of the EDC in turtles. Based on the results of this study a scenario for the diversification of turtle EDC genes was developed. The hypothetical structures of the EDC and two other loci, that contain EDC-related genes in modern turtles, are depicted schematically. The most primitive EDC containing ancestral SEDC genes ("simple EDC genes" consisting of one noncoding and one coding exon) is shown at the bottom. The association of EDC gene expression with tissues of modern turtles, as determined by RT-PCRs, is shown on the top of the schematics. Genes are represented by arrows. Curved lines indicate gene translocations; triangles indicate gene family expansions. To provide a better overview, only a subset of EDC genes of each clade (indicated by different colors) is shown.

of the SEDC-type (one noncoding and one coding exon) (fig. 3*E*). They are defined by a central segment of amino acids that are predicted to form beta-sheets which mediate the formation of filaments (Fraser and Parry 1996, 2014). The conserved presence of beta-keratin genes within the SEDC gene clusters of lizard (Strasser et al. 2014), birds, and turtles as well as identical exon–intron structures of beta-keratin and other SEDC genes argue for an evolutionary origin of beta-keratins by derivation from a common ancestral gene. However, the lack of SEDC-typical sequence motifs (supplementary fig. S6, Supplementary Material online) at the amino- and carboxy-terminal ends and the presence of the beta-sheet-forming core sequence makes beta-keratins unique among SEDC proteins and leaves open the possibility that as-yet-unknown recombination events were involved in the origin of beta-keratins. Our semiquantitative RT-PCRs suggest that the Beta-A cluster of turtle beta-keratin genes comprises genes (e.g., *Beta-A1*) that are expressed in the toes and others (e.g., *Beta-A4*) that are also expressed in the toes but more strongly in the shell (fig. 4). Notably, the Beta-A cluster is syntenic with the claw beta-keratin gene cluster in birds (Greenwold et al. 2014; supplementary fig. S13*A*, Supplementary Material online), and phylogenetic analysis suggests that these genes belong to the same subclade of beta-keratins, which comprises Beta-A plus Beta-O proteins of turtles and claw, feather, and scale beta-keratins of the chicken (supplementary fig. S13*B*, Supplementary Material online). Based on these data, we put forward the hypothesis that turtle Beta-A proteins and chicken claw beta-keratins have probably been inherited from a common ancestor of turtles and birds in which the evolutionary precursors of Beta-A proteins might have been components of claws. It is conceivable that distinct sequence features of these ancestral proteins contributed to the hardness of the claws. Later, duplicated genes of this type might have been co-opted as components of the hard scutes of the evolving shell. A gene translocation and further duplications generating the Beta-O cluster of shell beta-keratins might have been associated with the further evolution of the shell (fig. 5). This scenario is partly analogous to the evolution of the so-called "hair keratins," that is, keratin intermediate filament proteins that likely functioned in the claws of primitive amniotes before they were co-opted as components of mammalian hair (Eckhart et al. 2008).

The above scenario of beta-keratin evolution refines the evolutionary model of a previous report (Li et al. 2013), in which "turtle-specific beta-keratins," corresponding to beta-keratins of the Beta-A and Beta-O clusters of our study, with a putative expression in the shell have been proposed. Other reports have identified mRNAs encoding 17 individual beta-keratins in the hard-shelled turtle *Pseudemys nelsoni* (Dalla Valle et al. 2009) and five beta-keratins in the soft-shelled turtle *A. spinifera* (Dalla Valle et al. 2013). The results of the present study allow assigning 14, 2 and 1 beta-keratins of *P. nelsoni* to the Beta-O, A and B clusters, respectively, whereas all previously described beta-keratins of *A. spinifera* belong to the Beta-B cluster (supplementary fig. S20, Supplementary Material online). In agreement with our RT-PCR results obtained in *E. orbicularis*, the mRNA transcripts from Beta-B genes of *P. nelsoni* and *A. spinifera* tended to be more abundant in tissues outside of the shell (Dalla Valle et al. 2009, 2013). In contrast, a Beta-O protein predominated over

a Beta-B protein in the scutes of the shell of *P. nelsoni* according to a recent immuno-labeling study (Alibardi 2014), supporting the role of Beta-O proteins in the shell, as proposed here. In future studies, it will be important to carefully consider the sequence similarities among the many beta-keratins and to further improve quantitative comparisons of individual beta-keratin expression levels at different body sites of turtles.

A hard shell was present in a common ancestor of all modern turtles and was lost during the evolution of soft-shelled turtles (Gaffney 1990; Li et al. 2008; Lyson et al. 2014). A significant role of beta-keratin pseudogenization in this degeneration process was previously suggested (Li et al. 2013). The present study confirms changes in the set of beta-keratins in *P. sinensis* and identifies further epidermal differentiation genes that have been lost in this soft-shelled turtle. Besides a rearrangement and reduction of the number of *EDPCV* genes in *P. sinensis*, we found an inactivation of *EDWM* in the two soft-shelled turtles *P. sinensis* and *A. spinifera*. Since *EDWM* is present in all other sauropsids investigated so far (Strasser et al. 2014; supplementary fig. S15, Supplementary Material online), the distribution of *EDWM* in amniote species correlates with that of scales, which are widely conserved in sauropsids with the exception of soft-shelled turtles (Crawford et al. 2015). Notably, scales and scutes share elements of their developmental program (Moustakas-Verho and Cherepanov 2015). Therefore, the loss of *EDWM* may have been associated—perhaps as a secondary event after the inactivation of a surface patterning mechanism—with the loss of scales and hard scutes in soft-shelled turtles. A scenario summarizing the changes of the EDC during the evolution of soft-shelled turtles is depicted in supplementary figure S21, Supplementary Material online. It will be interesting to explore the genomic foundations for the diversification of the integument in the various phylogenetic lineages of turtles in future studies.

Collectively, the results of the present comparative genomics study and our gene expression data indicate that the evolution of the integument of turtles was associated with numerous adaptations of genes involved in epidermal differentiation and with the origin and expansion of shell-associated proteins. As this study provides a comprehensive catalog of EDC genes expressed in the epidermis and distinct skin appendages of turtles, these data will facilitate further in-depth investigations of the evolution of claws, rhamphotheca, scutes, and scales of turtles, and reptiles in general.

Materials and Methods

Genome Sequences and Gene Identification

Genome sequences from the following turtle species were used for gene predictions: western painted turtle (*C. picta bellii*) (Shaffer et al. 2013), Chinese soft-shelled turtle (*P. sinensis*), and green sea turtle (*Che. mydas*) (Wang et al. 2013). The accession numbers of genome sequences are listed in supplementary tables S2–S4, Supplementary Material online. Coding sequences and exon–intron borders were predicted according to a published approach (Strasser et al. 2014). Briefly, the genomic regions between S100A12 and S100A11 genes were screened for EDC genes using the following three methods. First, the amino acid sequences of EDC proteins from other amniotes were used as queries in tBLASTn searches. Second, RNA-seq data available in the Sequence Read Archive and information about RNA-seq exon coverage available in the NCBI browser for "genomic regions, transcripts, and products" were used to identify transcribed regions, which were subsequently investigated for the potential to encode proteins with amino acid sequences similar to known EDC proteins. Third, for the prediction of SEDC genes, the genomic sequence was conceptually translated, and open reading frames encoding proteins of 50–500 amino acids were identified. Putative protein-coding sequences were scrutinized for the presence of a splice acceptor site at a distance of 10–30 nt upstream of the start codon and for the presence of a putative noncoding exon 1, as defined by a TATA box followed by a splice donor site at a distance of 60–90 nt. The gene predictions were validated by BLAST searches in the transcriptome of *T. scripta* (Kaplinsky et al. 2013) and by RT-PCR tests in *E. orbicularis* (see below).

Sequence Alignment and Phylogenetic Analysis

For phylogenetic analysis, the amino acid sequences of beta-keratins of *C. picta* (supplementary fig. S1*B*, Supplementary Material online) and chicken were used. Chicken beta-keratin genes within the EDC (chromosome 25) were identified at the genomic loci indicated in supplementary table S6, Supplementary Material online, and translated in silico. Amino acid sequences of feather beta-keratins encoded by genes outside of the EDC were obtained from Ng et al. (2014). The beta-keratin sequences were aligned using Multalin (Corpet 1988) with default settings. After checking for alignment errors, only the unambiguously aligned core segment (positions 67–126 of the overall alignment, supplementary Material online: FASTA file) was used for subsequent phylogenetic analysis. A phylogenetic tree was reconstructed by maximum likelihood (ML) using IQ-TREE 1.3.8 (Nguyen et al. 2015) using the JTT + G4 model (Jones et al. 1992; Yang 1994). The evolutionary model was determined by model selection according to Posada (2008) as implemented in IQ-Tree using the Bayesian information criterion. Tree searches were performed for three different perturbation strengths (-pers 0.5, 0.2, and 0.1) and two different stop conditions (-numstop 200 and 400). For each pair of search options, five replicates were performed and the reconstructed tree with the highest likelihood was taken as the ML estimate. Support values were obtained by ultrafast bootstrap approximation (UFBoot) (Minh et al. 2013) with 10,000 samples in IQ-TREE. Since UFBoot support values behave like posterior probabilities (Minh et al. 2013), branches with support values of at least 90% are regarded as supported, whereas values of at least 95% are regarded as strongly supported.

Animal Tissues

Tissues were sampled from 45 days old embryos of the European pond turtle (*E. orbicularis*) in agreement with the

national laws regulating animal welfare, the guidelines of Good Veterinary Practice, and the guidelines of the Ethics committee of the Medical University of Vienna. The embryos were derived from an *E. orbicularis* breeding program at the Vienna Zoo.

RT-PCR

RNA was prepared from tissues of *E. orbicularis* according to a published protocol (Mlitz et al. 2014; Strasser et al. 2014). The RNA was reverse-transcribed to cDNA which was subsequently amplified by PCRs with primers specific for EDC genes. The sequences of the primers were chosen to anneal to conserved regions of EDC genes predicted in the genomes of *C. picta* and *Che. mydas*. Primer sequences are listed in supplementary table S5, Supplementary Material online. PCR products were purified and sequenced. Nucleotide sequences of cDNAs were submitted to GenBank (accession numbers KR632557–KR632565).

Supplementary Material

Supplementary figures S1–S21 and tables S1–S6 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

The authors thank Michael Mildner and Karin Jaeger for helpful discussions and technical support. This work was supported by the Austrian Science Fund (FWF): P23801.

References

Alibardi L. 2003. Adaptation to the land: the skin of reptiles in comparison to that of amphibians and endotherm amniotes. *J Exp Zool B Mol Dev Evol.* 298:12–41.

Alibardi L. 2014. Immunocytochemistry suggests that the prevalence of a sub-type of beta-proteins determines the hardness in the epidermis of the hard-shelled turtle. *J Exp Zool B Mol Dev Evol.* 322:54–63.

Alibardi L, Dalla Valle L, Nardi A, Toni M. 2009. Evolution of hard proteins in the sauropsid integument in relation to the cornification of skin derivatives in amniotes. *J Anat.* 214:560–586.

Alibardi L, Thompson MB. 1999. Epidermal differentiation during carapace and plastron formation in the embryonic turtle *Emydura macquarii*. *J Anat.* 194:531–545.

Bever GS, Lyson TR, Field DJ, Bhullar BAS. 2015. Evolutionary origin of the turtle skull. *Nature* 525:239–242.

Burke AC. 1989. Development of the turtle carapace: implications for the evolution of a novel bauplan. *J Morphol.* 199:363–378.

Candi E, Schmidt R, Melino G. 2005. The cornified envelope: a model of cell death in the skin. *Nat Rev Mol Cell Biol.* 6:328–340.

Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. *Nucleic Acids Res.* 16:10881–10890.

Crawford NG, Parham JF, Sellas AB, Faircloth BC, Glenn TC, Papenfuss TJ, Henderson JB, Hansen MH, Simison WB. 2015. A phylogenomic analysis of turtles. *Mol Phylogenet Evol.* 83:250–257.

Dalla Valle L, Michieli F, Benato F, Skobo T, Alibardi L. 2013. Molecular characterization of alpha-keratins in comparison to associated beta-proteins in soft-shelled and hard-shelled turtles produced during the process of epidermal differentiation. *J Exp Zool B Mol Dev Evol.* 320:428–441.

Dalla Valle L, Nardi A, Toni M, Emera D, Alibardi L. 2009. Beta-keratins of turtle shell are glycine-proline-tyrosine rich proteins similar to those of crocodilians and birds. *J Anat.* 214:284–300.

Eckhart L, Dalla Valle L, Jaeger K, Ballaun C, Szabo S, Nardi A, Buchberger M, Hermann M, Alibardi L, Tschachler E. 2008. Identification of reptilian genes encoding hair keratin-like proteins suggests a new scenario for the evolutionary origin of hair. *Proc Natl Acad Sci U S A.* 105:18419–18423.

Eckhart L, Lippens S, Tschachler E, Declercq W. 2013. Cell death by cornification. *Biochim Biophys Acta.* 1833:3471–3480.

Fraser RD, Parry DA. 1996. The molecular structure of reptilian keratin. *Int J Biol Macromol.* 19:207–211.

Fraser RD, Parry DA. 2014. Amino acid sequence homologies in the hard keratins of birds and reptiles, and their implications for molecular structure and physical properties. *J Struct Biol.* 188:213–224.

Gaffney ES. 1990. The comparative osteology of the Triassic turtle Proganochelys. *Bull Am Mus Nat Hist.* 194:1–263.

Greenwold MJ, Bao W, Jarvis ED, Hu H, Li C, Gilbert MT, Zhang G, Sawyer RH. 2014. Dynamic evolution of the alpha (α) and beta (β) keratins has accompanied integument diversification and the adaptation of birds into novel lifestyles. *BMC Evol Biol.* 14:249.

Henry J, Toulza E, Hsu C, Pellerin L, Balica S, Mazereeuw-Hautier J, Paul C, Serre G, Jonca N, Simon M. 2012. Update on the epidermal differentiation complex. *Front Biosci.* 17:1517–1532.

Hirasawa T, Nagashima H, Kuratani S. 2013. The endoskeletal origin of the turtle carapace. *Nat Commun.* 4:2107.

Hirasawa T, Pascual-Anaya J, Kamezaki N, Taniguchi M, Mine K, Kuratani S. 2015. The evolutionary origin of the turtle shell and its dependence on the axial arrest of the embryonic rib cage. *J Exp Zool B Mol Dev Evol.* 324:194–207.

Iwabe N, Hara Y, Kumazawa Y, Shibamoto K, Saito Y, Miyata T, Katoh K. 2005. Sister group relationship of turtles to the bird-crocodilian clade revealed by nuclear DNA-coded proteins. *Mol Biol Evol.* 22:810–813.

Jones DT, Taylor WR, Thornton JM. 1992. The rapid generation of mutation data matrices from protein sequences. *Comput Appl Biosci.* 8:275–282.

Kalinin AE, Kajava AV, Steinert PM. 2002. Epithelial barrier function: assembly and structural features of the cornified cell envelope. *Bioessays* 24:789–800.

Kaplinsky NJ, Gilbert SF, Cebra-Thomas J, Lilleväli K, Saare M, Chang EY, Edelman HE, Frick MA, Guan Y, Hammond RM, et al. 2013. The embryonic transcriptome of the red-eared slider turtle (*Trachemys scripta*). *PLoS One* 8:e66357.

Kumar S, Hedges SB. 2011. TimeTree2: species divergence times on the iPhone. *Bioinformatics* 27:2023–2034.

Kypriotou M, Huber M, Hohl D. 2012. The human epidermal differentiation complex: cornified envelope precursors, S100 proteins and the 'fused genes' family. *Exp Dermatol.* 21:643–649.

Li C, Wu XC, Rieppel O, Wang LT, Zhao LJ. 2008. An ancestral turtle from the Late Triassic of southwestern China. *Nature* 456:497–501.

Li YI, Kong L, Ponting CP, Haerty W. 2013. Rapid evolution of Beta-keratin genes contribute to phenotypic differences that distinguish turtles and birds from other reptiles. *Genome Biol Evol.* 5:923–933.

Lyson TR, Schachner ER, Botha-Brink J, Scheyer TM, Lambertz M, Bever GS, Rubidge BS, de Queiroz K. 2014. Origin of the unique ventilatory apparatus of turtles. *Nat Commun.* 5:5211.

McGaughey GB, Gagné M, Rappé AK. 1998. pi-Stacking interactions. Alive and well in proteins. *J Biol Chem.* 273:15458–15463.

Minh BQ, Nguyen MA, von Haeseler A. 2013. Ultrafast approximation for phylogenetic bootstrap. *Mol Biol Evol.* 30:1188–95.

Mischke D, Korge BP, Marenholz I, Volz A, Ziegler A. 1996. Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex ("epidermal differentiation complex") on human chromosome 1q21. *J Invest Dermatol.* 106:989–992.

Mlitz V, Strasser B, Jaeger K, Hermann M, Ghannadan M, Buchberger M, Alibardi L, Tschachler E, Eckhart L. 2014. Trichohyalin-like proteins have evolutionarily conserved roles in the morphogenesis of skin appendages. *J Invest Dermatol.* 134:2685–2692.

Moustakas-Verho JE, Cherepanov GO. 2015. The integumental appendages of the turtle shell: an evo-devo perspective. *J Exp Zool B Mol Dev Evol.* 324:221–229.

Moustakas-Verho JE, Zimm R, Cebra-Thomas J, Lempiäinen NK, Kallonen A, Mitchell KL, Hämäläinen K, Salazar-Ciudad I, Jernvall J, Gilbert SF. 2014. The origin and loss of periodic patterning in the turtle shell. *Development* 141:3033–3039.

Nagashima H, Sugahara F, Takechi M, Ericsson R, Kawashima-Ohya Y, Narita Y, Kuratani S. 2009. Evolution of the turtle body plan by the folding and creation of new muscle connections. *Science* 325:193–196.

Ng CS, Wu P, Fan WL, Yan J, Chen CK, Lai YT, Wu SM, Mao CT, Chen JJ, Lu MY, et al. 2014. Genomic organization, transcriptomic analysis, and functional characterization of avian α- and β-keratins in diverse feather forms. *Genome Biol Evol.* 6:2258–2273.

Niessen CM. 2007. Tight junctions/adherens junctions: basic structure and function. *J Invest Dermatol.* 27:2525–2532.

Nguyen LT, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol.* 32:268–274.

Posada D. 2008. jModelTest: phylogenetic model averaging. *Mol Biol Evol.* 25:1253–1256.

Reisz RR, Head JJ. 2008. Palaeontology: turtle origins out to sea. *Nature* 456:450–451.

Rice R, Riccio P, Gilbert SF, Cebra-Thomas J. 2015. Emerging from the rib: resolving the turtle controversies. *J Exp Zool B Mol Dev Evol.* 324:208–220.

Rice RH, Winters BR, Durbin-Johnson BP, Rocke DM. 2013. Chicken corneocyte cross-linked proteome. *J Proteome Res.* 12:771–776.

Ruckes H. 1929. The morphological relationships between the girdles, ribs, and carapace. *Ann N Y Acad Sci.* 13:81–120.

Schweizer J, Bowden PE, Coulombe PA, Langbein L, Lane EB, Magin TM, Maltheis L, Omary MB, Parry DA, Rogers MA et al. 2006. New consensus nomenclature for mammalian keratins. *J Cell Biol.* 174:169–174.

Shaffer HB, Minx P, Warren DE, Shedlock AM, Thomson RC, Valenzuela N, Abramyan J, Amemiya CT, Badenhorst D, Biggar KK, et al. 2013. The western painted turtle genome, a model for the evolution of extreme physiological adaptations in a slowly evolving lineage. *Genome Biol.* 14:R28.

Strasser B, Mlitz V, Hermann M, Rice RH, Eigenheer RA, Alibardi L, Tschachler E, Eckhart L. 2014. Evolutionary origin and diversification of epidermal barrier proteins in amniotes. *Mol Biol Evol.* 31:3194–3205.

Strasser B, Mlitz V, Hermann H, Tschachler E, Eckhart L. 2015. Convergent evolution of cysteine-rich proteins in feathers and hair. *BMC Evol Biol.* 15:82.

Thomson RC, Plachetzki DC, Mahler DL, Moore BR. 2014. A critical appraisal of the use of microRNA data in phylogenetics. *Proc Natl Acad Sci U S A.* 111:E3659–E3668.

Vandebergh W, Bossuyt F. 2012. Radiation and functional diversification of alpha keratins during early vertebrate evolution. *Mol Biol Evol.* 29:995–1004.

Vanhoutteghem A, Djian P, Green H. 2008. Ancient origin of the gene encoding involucrin, a precursor of the cross-linked envelope of epidermis and related epithelia. *Proc Natl Acad Sci U S A.* 105:15481–15486.

Wang Z, Pascual-Anaya J, Zadissa A, Li W, Niimura Y, Huang Z, Li C, White S, Xiong Z, Fang D, et al. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet.* 45:701–706.

Waters ML. 2002. Aromatic interactions in model systems. *Curr Opin Chem Biol.* 6:736–741.

Wiche G, Osmanagic-Myers S, Castañón MJ. 2015. Networking and anchoring through plectin: a key to IF functionality and mechanotransduction. *Curr Opin Cell Biol.* 32:21–29.

Yang Z. 1994. Maximum likelihood phylogenetic estimation from DNA sequences with variable rates over sites: approximative methods. *J Mol Evol.* 39:306–314.

Zangerl R. 1969. The turtle shell. In: Gans C, d'A Bellairs, Parsons T, editors. Biology of reptilia, I. New York (NY): Academic Press. p. 311–339.

2.2 Results for skin defenses: Microbiological assays with skin specific reptile antimicrobial peptides

2.2.1 Introduction to paper II: Microbicide Activity of Two Reptilian Antimicrobial Peptides on Gram Positive and Gram Negative Bacteria.

In the following paper, we have studied the defense mechanism exerted by reptile antimicrobial peptides in the epidermis. Two previously identified skin-specific reptile antimicrobial peptides, namely a turtle beta defensin and a lizard cathelicidin, were tested using the broth microdilution assay on Gram-negative (*E. coli*) and Gram-positive bacteria (*S. aureus*).

The bactericidal activities of these antimicrobial peptides (AMPs) are demonstrated by both growth inhibition during microbial assays and by an electron microscopic study on the ultrastructural damage produced by these the bacteria.

The bactericidal effect was assessed by the MIC and IC_{50} values after 3 hours of incubation with the peptides. Values differed between the beta defensin and cathelicidin and were in the range from 0.69-4.14 mg/ml and 0.05-1.9 mg/ml, respectively.

On the ultrastructural level the effect of the peptides was visible by alteration and rupture in the plasma membrane, lowering of the ribosomes, swelling and clumping in the nucleoid region of bacteria. Immunogold labeling against the two peptides indicated that their localization was not limited to the plasma membrane and the cytoplasm of the treated bacteria, but it was also observed in the nucleoid region and its protein scaffold.

The bactericidal activity was observed at peptide concentrations that were relatively high as compared to other studies on AMPs. Differences in the activities of the AMPs could be due to differences in technical protocols and peptide solubility of the *in vitro* assays.

This study demonstrates that two reptile skin-specific antimicrobial peptides here tested inhibit bacterial growth. Additionally, the ultrastructural data suggests that these peptides initially

operate at the plasma membrane but later they are also found associated with ribosomes and can even enter the nucleus. It has been hypothesized that reptiles have an efficient innate immunity, in part based on anti-microbial peptides, because their acquired immune system is relatively slow and not as efficient as that of mammals. Our study offers experimental evidence that supports this hypothesis.

Journal of Immunobiology

Holthaus et al, J Immuno Biol 2016, 1:1
http://dx.doi.org/10.4172/jib.1000104

Research Article Open Access

Microbicide Activity of Two Reptilian Antimicrobial Peptides on Gram Positive and Gram Negative Bacteria

Holthaus KB, Spisni E and Alibardi L*

Department of Bigea, University of Bologna, Via Selmi 3, Italy

***Corresponding author:** Alibardi, L, Department of Bigea, University of Bologna, Via Selmi 3, Italy, Tel: +39 051 209 4257; E-mail: lorenzo.alibardi@unibo.it

Received date: December 10, 2015; **Accepted date:** January 19, 2016; **Published date:** February 4, 2016

Abstract

Previous *in-vivo* studies have isolated and identified peptides with typical molecular anti-microbial characteristics in reptiles. In the present study we have tested the putative antimicrobial action of a lizard cathelicidin and of a turtle beta-defensin using the broth microdilution assay on Gram positive and Gram negative bacteria. The addition of the peptides at concentrations indicatively ranging between 0.05-1.9 mg/ml (cathelicidin) and 0.69-4.14 mg/ml (beta-defensin) inhibited bacterial growth after 3 hours of incubation as determined by their MIC and IC_{50} values. Due to the poor solubility and the medium interference the real concentration of the delivered peptides to the bacterial cultures was uncertain. The qualitative evaluation of the anti-microbial damage after treatment with the peptides was done under the electron microscope that showed some alteration and rupture in the plasma membrane, lowering of the ribosomes, swelling and clumping in nucleoid region of Gram negative (*E. coli*) and Gram positive (*S. aureus*) bacteria. Immunogold labeling against the two peptides indicated that the peptides were localized not only on the plasma membrane and in cytoplasm of the treated bacteria, but also in the nucleoid region and its protein scaffold. The present ultrastructural study suggests that these peptides operate a cellular damage initially on the plasma membrane but further also in the ribosomes and on the DNA or its associated proteins.

Keywords: Reptiles; Antimicrobial peptides; Bacteria; Antimicrobial tests; Ultrastructure.

Introduction

Protection from potentially pathogenic infections from microbes occurs through different mechanisms, including the production of antimicrobial peptides [1]. Numerous antimicrobial peptides responsible for a strong innate immunity have been discovered in prokaryotes and eukaryotes [2-5]. Antimicrobial peptides are composed of 8-60 or more amino acids and include several categories among which the best known include the beta-defensins and cathelicidins [6,7]. The potential utilization of these molecules as effective new antibiotics is of paramount importance in recent times due the mounting resistance of numerous pathogenic microbes to old and new classes of antibiotics, and therefore efforts in discovering effective new drugs is a very active field of modern infective research [8]. Antimicrobial peptides are not a homogeneous class of compounds, but show a broad diversity in structure and antimicrobial spectrum and interactions [9].

Previous studies, based on the observation of the high resistance of lizards and turtle to wounds which showed the presence of numerous intercellular and intracellular bacteria in the epidermis [10-13], suggested that potent antimicrobial peptides were possibly involved in the outstanding immunity present in these reptiles. This hypothesis was later confirmed by the isolation of numerous beta-defensins and some cathelicidins from lizard and turtle [14-16], and from their prevalent localization in granulocytes and activated keratinocytes [13,17]. The association of immunoreactivity for both beta-defensins and cathelicidins with bacteria localized in the stratum corneum further suggested the presence of an antimicrobial barrier in the epidermis, possibly derived from the release of antibacterial molecules that can reach the superficial part of the stratum corneum. Therefore reptiles among amniotes may represent an interesting source of potentially useful peptide antibiotics for medical utilization [8,18,19].

A direct proof of a true antimicrobial affect for the peptides characterized in both turtles and lizards awaits further studies testing the identified molecules on microbial cultures. The present study address the above goal, documenting a cytotoxic effect of two among the most abundant antimicrobial peptides previously characterized in a turtle and a lizard, on cultured bacteria. The antimicrobial effect has been detected using microbial cultures of Gram negative and Gram positive bacteria, and the microbicide action was documented by determining the degree of growth inhibition and evaluating the ultrastructural damage on bacterial cells.

Materials and Methods

Bacterial strains

We used as test organisms Gram negative and Gram positive bacteria represented respectively by Escherichia coli (strain DH5a) and *Staphylococcus aureus* (strain ATCC 2913). The *E. coli* strain was stored in our lab at -80°C and the *S. aureus* strain came from an LB/agar plate. Both strains were cultured freshly for 24 hrs, and shaken at 220 rpm at 37°C in LB Lennox broth for the experiments.

Peptides

Two reptile antimicrobial peptides of 40 amino acids selected by us were synthesized by ProteoGenix Biotec Company, France, as a peptide synthesis service. These cationic antimicrobial peptides were selected on the sequences of a cathelicidine detected in the lizard *Anolis carolinensis* (AcCATH-1, [15]) and of a beta defensin detected in the

turtle *A. spinifera* (TuBD-1, [14]). The amino acid sequences of both antimicrobial peptides are shown in Table 1. Dissolved stock solutions were prepared by the producing Company. In order to avoid that the peptide solvents used in the experiments could also produce damaging effects on our tested microorganisms, the peptide solutions were tested in two different trials for their inhibitory effects. In one experiment the peptides were removed through filtering and in another experiment by reproducing the composition of the solvent and utilizing this solution on the bacterial strains omitting the peptides. The vehicle solution for the turtle beta defensin did not influence bacterial growth at any tested concentration and also the solvent of the lizard cathelicidin did not affect bacterial growth at the employed concentration.

Peptide	Concentration in mg/ml							
	Gram negative *E. coli*				Gram positive *S. aureus*			
	No effect	IC50	MIC least effect	MIC 100%	No effect	IC50	MIC least effect	MIC 100%
TuBD-1 IIGTAICIRRRGACFPIRCPL-YTVRIGRCGLALPCCRWYR	0.5	0.81	0.69	4.14	0.5	1.14	0.69	4.14
AcCATH-1 SLIVVTCDAAVQDDPQMTR-FRGLGHFFKGFGRGFIWGLNH	0.037	0.04	0.062†	*1.90	0.05	0.15†	0.095	*1.90

Table 1: Antimicrobial activity of TuBD-1 and AcCATH-1 against *E. coli and S. aureus* *, at this concentration the solvent likely has also an antimicrobial effect. † value is indicative, because in the upper range of doses there was interference with the solvent. Note: MIC of 100% was defined as inhibiting ≥ 99.9% of bacterial growth.

Media used

Lennox Broth (LB) containing 10 g/l tryptone, 5 g/l yeast extracts and 5 g/l NaCl, was utilized as the standard medium. This medium had a physiological pH and salt concentration of 86 mM. Due to the low solubility of the peptides in the medium and to the possible interactions with salts and organic components we attempted to introduce some variants in order to increase the peptide solubility, like a low salt Medium (10 mM instead of 86 mM as in the original medium), and the addition of 0.01% or 0.025% acetic acid. Another medium utilized was a 0.1 M Tris-HCl buffer solution at pH 6.8, modified with the addition of NaCl to obtain a final concentration of 36.7 mM NaCl.

Antimicrobial assays

Initially the peptides were tested on LB/agar in Petri plates using the colony counting assay but, probably due to the binding of the charged peptides to complex carbohydrates present in the agar [20], no antimicrobial activity was detected. Therefore, the broth micro dilution assay was applied to samples, and the incubation with the peptides was done using LB as described before as a medium. This procedure was followed by plating the surviving bacteria from the test solution on agar in order to determine the antimicrobial activity of the peptides using the colony counting assay.

Prior to testing a subculture of the bacterial strain, the culture was grown at 37°C until the concentration of bacteria reached a mid-logarithmic phase (about 3 hours). After measuring the Optical Density at 600 nm (OD600), the bacterial culture was diluted in the standard medium (LB) to obtain 106 colony-forming units per ml (CFU/ml).

We tested the peptides at concentrations ranging from a minimum of 0.05 µg/ml up to 4.14 mg/ml. Peptides were diluted to the different testing concentrations in 50 µl LB and added to an equal volume of bacterial solution in a 1:1 dilution, and therefore the final bacterial solutions contained 5×10^5 CFU/ml. The final inoculated volumes of 100 µl were then incubated for 3 hrs at 37°C, and shook at 220 rpm. After this period the bacterial solutions were diluted on a 10 fold base, and they were plated in duplicate on Petri dishes (60 mm Ø). After incubation for 18-20 hrs at 37°C in the Petri dishes, the CFU were counted and compared to control cultures grown with no addition of the peptide. The antimicrobial activity was expressed as % of bacterial growth inhibition with respect to the controls, and it was plotted against the tested concentrations of peptides. Using linear regression, the half maximum inhibitory concentration (IC_{50}) was calculated with Excell's ED50V10 add-in method. We also determined the minimal inhibitory concentrations (MIC) at 100% growth inhibition and at the minimum effect on the bacterial cultures. All the results were based on the mean value obtained by at least three independent trials performed in duplicate. The peptides did not show any activity when plated on agar; therefore the colony counting assay was not used for testing the peptides but merely to quantify the growth inhibition obtained by the broth micro dilution assay.

Ultrastructural evaluation of the damage

We sampled controls and tested colonies (2 × 4 mm large) that were growing on the LB-agar substrate (arrowheads and arrows in Figure 1), and that showed different degrees of inhibition related to the peptide used (CATH-1 at 95 µg/ml and TBD-1 at 1.0 mg/ml). Using a sharp razor blade and a tweezer, the colonies of interest grown on the Agar substrate were collected from their Petri dish and immediately fixed. The fixative contained 4% Paraformaldehyde in 0.12 M Phosphate buffer at pH 7.2, and fixation lasted 3 hours at room temperature. After rinsing in the Buffer, the samples were dehydrated in ethanol up to 90% and embedded in Bioacryl Resin under UV at 0-4°C (Scala et al. 1992). Using an ultramicrotome, 1-2 µm thick sections of the samples with their agar support were collected, and the presence of bacteria was systematically checked after staining the sections with 0.5% Toluidine blue. After identifying useful area containing groups of bacteria, thin sections of 40-90 nm were collected on 200-300 mesh Copper or nickel grids for the following study under the transmission electron microscope.

For the routine morphological study, the samples were stained for 30 minutes in 1% uranyl acetate and 5 min in 0.01 M lead citrate, rinsed in water and dried. For ultrastructural immunocytochemistry, two polyclonal rabbit antibodies against AcCATH-1 and TuBD-1 were utilized, as previously specified [12,13]. Briefly, sections on nickel grids were incubated for 3-4 hours at room temperature with the primary antibody at a dilution 1:100 in 0.12 M Tris buffer pH 7.2 containing 1%

Bovine Serum Albumine and 0.01% Triton-X. In control sections, the antibody was omitted in the incubation step. After rinsing in the buffer, the sections were incubated for 1 hour at room temperature with an anti-Rb secondary antibody conjugated with 5 or 10 nm gold particles, rinsed in buffer, in distilled water, and dried. The sections were observed under a Zeiss C10 Transmission Electron Microscope operating at 60 kV, and the images were recorded by a digital camera or photographed with Kodak films (EM Film 4489).

Results

Antimicrobial assays

Both antimicrobial peptides tested showed to negatively influence bacterial growth in *E. coli* and *S. aureus* (Figure 1 and Table 1). One of the problems we encountered in trying to establish a testing protocol was that both peptides did not dissolve well in the medium and this probably diminished their potential activity and availability to the bacterial targets. Another problem was that the solvent of the cathelicin utilized by the producer for the production of the peptide showed inhibitory side effects. Since we could not identify a suitable testing medium where peptides were solubilized efficiently, the calculated concentration in our tests should be considered only indicative.

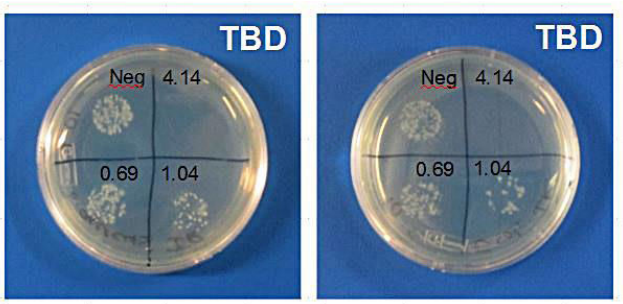


Figure 1: Examples of visible antimicrobial effect on colonies of *E. coli* (left) and *S. aureus* (right) grown in Petri dishes after treatment with TuBD-1 (TBD). The concentrations of the peptides are indicated in mg/ml. Neg is the negative control (untreated, arrowheads) while 4.14 is the MIC (complete inhibition). The other concentrations indicate the least effect (0.69) and an intermediate concentration (1.04). The latter was utilized for the study under the electron microscope (arrows).

As a general result in our tests, although the turtle beta defensin (TuBD-1) did inhibit the growth in both tested bacterial species, its effect was less pronounced compared to the lizard cathelicidin. TuBD-1 in particular did not solubilize well in the employed medium, and form irregular precipitating aggregates. Despite of this drawback it was determined that the IC_{50} for the turtle peptide was indicatively at 0.81 mg/ml for *E. coli* and 1.14 mg/ml for *S. aureus*. There was no sign of inhibition under 0.5 mg/ml of peptide concentration, while the MIC was at 0.69 mg/ml. No growth at all was seen at 4.14 mg/ml of turtle defensin for both bacterial species (Figure 1 and Table 1). Both Gram positive and Gram negative species showed the same MIC for TuBD-1, but *E. coli* was more sensitive, and showed an average inhibition of 78% against 51% inhibition for *S. aureus* with a concentration of 1.04 mg/ml. Also, the IC_{50} of *E. coli* was lower than the IC_{50} for *S. aureus* (Table 1)

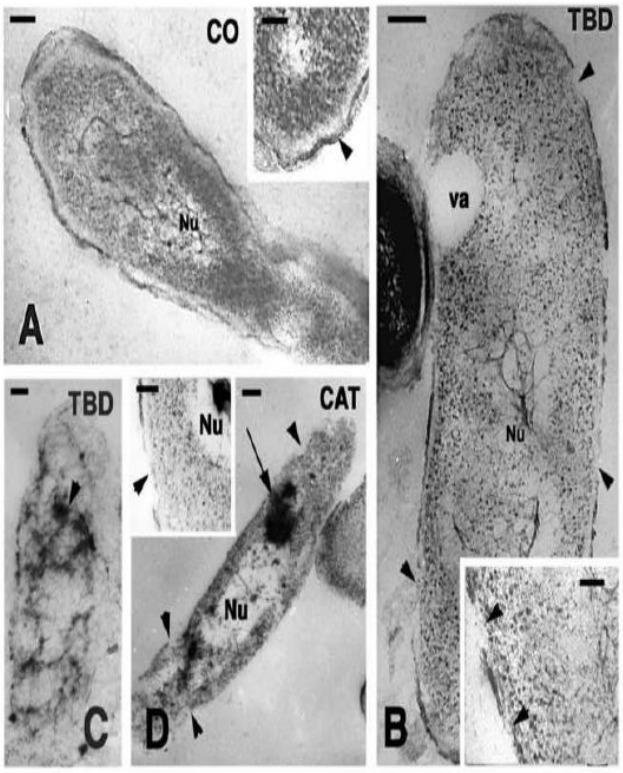


Figure 2: Ultrastructure of normal (A) and damaged (B-D) *E. coli* after TuBD-1 (TBD) and AcCATH-1 (CAT) treatment. A, untreated control cell (CO) showing the central nucleoid region (Nu). Bar: 300 nm. The inset shows the continuity of the cell membrane (arrowhead). Bar: 200 nm. B, damaged bacterial cell after treatment with TuBD-1 (TBD). The cell membrane is discontinuous (arrowheads), the electron-pale cytoplasm is vacuolated (va) and ribosomes are diluted, and the nucleoid (Nu) is not well distinguished from the cytoplasms. Bar: 250 nm. In the inset, the arrowheads point to a discontinuous cell wall and plasma membrane, Bar: 100 nm. C, advance degenerated bacterium after TuBD-1 administration. The arrowhead indicates clumped electron-dense globules while no ribosomes and plasma membrane are present and the cell content directly contacts the extracellular medium. Bar: 100 nm. D, damaged bacterial cell after AcCATH-1 application featuring the enlarged empty nucleoid region (Nu), large electron-dense globules (arrow) and loss of the cell wall and plasma membrane (arrowheads) so that the cytoplasm is exposed. Bar: 100 nm. The inset details on the discontinuity of the cell wall and plasma membrane (arrowhead). Nu, pale nucleoid. Bar: 100 nm.

Also the lizard peptide (AcCATH-1) did not completely dissolve and tended to precipitate, so that the effective concentration available for the anti-microbial effect was lower than the initial concentration. Despite of this drawback, the lizard cathelicidin (AcCATH-1) showed an IC_{50} of 62 µg/ml on *E. coli* (Table 1), but started to inhibit growth at 50 µg/ml and showed no effect at 37 µg/ml. The test using AcCATH-1 on *S. aureus* showed an IC_{50} of 150 µg/ml and the concentration with no inhibitory effect was at 50 µg/ml, therefore higher when compared to that for *E. coli*.

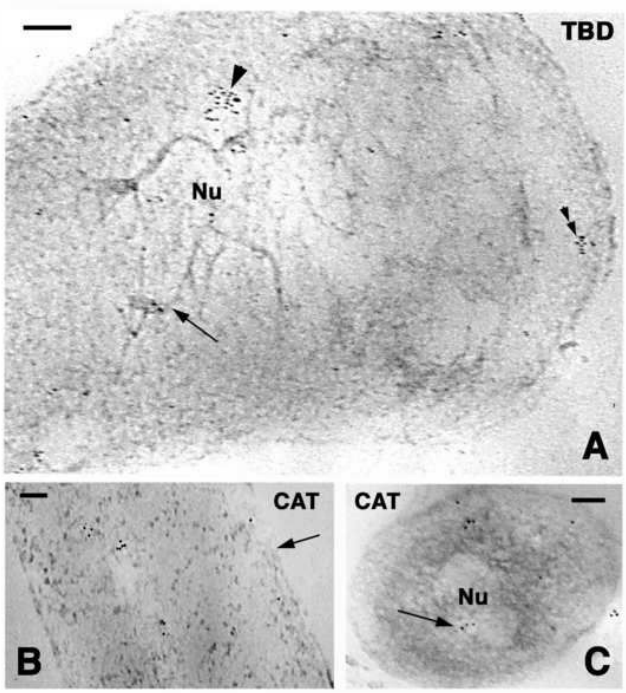


Figure 3: Immunolabeling for TuBD-1 (TBD) and for AcCATH-1 (CAT) in *E. coli* 3 hours after the treatment. A, detail of a cell showing gold particles localized in the peripheral cytoplasm (double arrowhead), nucleoid (arrowhead), and on the protein scaffold (arrow). Bar: 100 nm. B, intracellular labelling in a bacterium with rupture of the plasma membrane (arrow). Bar: 100 nm. C, cross-sectioned bacterium showing labelling in the cytoplasm and Nucleoid (Nu) region (arrow). Bar: 100 nm.

The total inhibition was the same in both species at the concentration of 1.9 mg/ml of cathelicidin but, due to the possible damaging effects of the solvent at this relatively high concentration, these results were discharged. In conclusion, as the results obtained with the turtle beta-defensin, also the lizard cathelicidin showed a stronger inhibitory effect on *E. coli* than on *S. aureus*.

We tried alternative protocols in the attempt to improve the peptides solubility but with no success. When glacial acetic acid at 0.01 and 0.025% was added to the peptide solutions, no effect was elicited aside a negative impact on the growth of bacteria . In another attempt to increase the antimicrobial peptide activity on the bacteria we tested another medium (0.1M Tris HCL) as well as a low salt variant of the LB-medium (10 mM NaCl instead of 86 mM). In the modified 0.1 M Tris HCL buffered medium the turtle beta defensin eventually showed a good solubility, but the buffer alone had a strong inhibitory effect (over 90%) on bacterial growth, which made it unsuitable . The low salt LB variant did not increase the peptide activity but likely influenced bacterial growth, and no further work was carried out following these alternative protocols.

Ultrastructural analysis on *E. coli*

The number of bacteria observed in each thin section analyzed under the electron microscope (12 thin sections in total) ranged between 30 and 60 (*E. coli*). The qualitative observations on untreated *E. coli* showed the typical ultrastructure with numerous free ribosomes surrounding the nucleoid region, and a complete cell wall and plasma membrane surrounding the perimeter of the cell (Figure 2A). Damaged bacteria, with membrane or cytoplasmic alterations, were occasionally seen in untreated cultures.

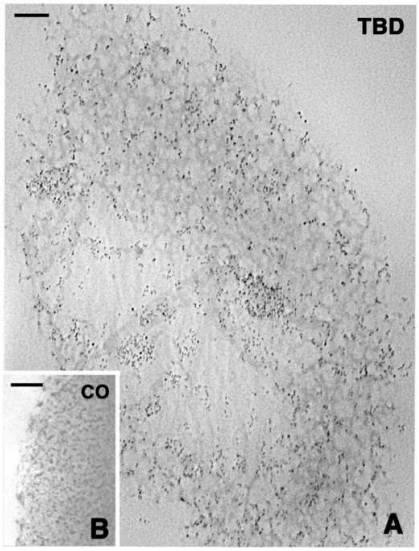


Figure 4: Degenerating immunolabeled *E. coli.* A, Largely degenerated bacterium missing of cell membranes, ribosomes and nucleoid region, and intensely immunolabeled for TuBD-1 (TBD). Bar: 100 nm. B, immunonegative control section. Bar: 200 nm.

The observations on samples after 3 hours of incubation with 1.0 mg/ml of Turtle BD-1, showed that most bacteria (roughly over 80% of recognizable bacteria) appeared damaged in both the cell wall and plasma membrane as well as in the ribosome number (decreased) and in the nucleoid region (Figure 2B). The degree of damage varied from swollen bacterial cells to completely degenerated cells without recognizable cell organelles. In the slightly altered bacteria, the number of ribosomes appeared reduced and the protein scaffold in the nucleoid region appeared irregularly dilated while numerous discontinuities were present along the cell wall (Figure 2B). In other bacteria, cell degeneration was more advanced to the point that not only the cell wall was largely absent but also the cytoplasm appeared devoid of ribosomes while numerous irregular clumps of electron-dense material were present (Figure 2C).

A similar damage over many bacterial cells (roughly over 80% of recognizable bacteria, but likely clumped material derived from completely destroyed bacteria was also present in the sections) was also detected after treatment with 95 μg/ml of the lizard cathelicidin (Ac-CATH-1). The damaged bacteria after 3 hours of peptide incubation appeared generally in a very advanced stage of degeneration, featuring numerous discontinuities along the cell wall and plasma membrane, strong reduction of ribosomes, appearance of flocculent material in the cytoplasm and of dense roundish clumps of material often associated to the nucleoid (Figure 2D). The nucleoid region in particular was swollen and scarce protein scaffolds were seen.

The immunogold observations on damaged but still recognizable bacterial cells of *E. coli* showed the presence of gold particles over the cytoplasm and the nucleoid area using both the turtle beta-defensin and lizard cathelicidin, including the protein scaffold of the nucleoid (Figure 3). This observation indicated a complete penetration of the

peptides in all regions of the bacterial cells. In some residual bodies resulting from advanced stages of bacterial cell degeneration, the immunolabeling was seen over most of the bacterial remnants where a nucleoid and cytoplasmic regions were no longer detectable (Figure 4A). No labeling was seen in control sections (Figure 4B).

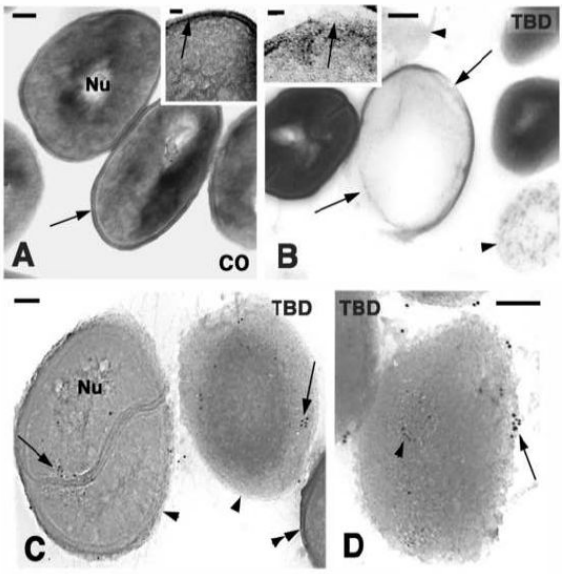


Figure 5: Structure (A,B) and immunogold labeling (B,C) of *S. aureus* cells treated with TuBD-1 (TBD). A, untreated samples (CO, control) showing a dense cytoplasm, a complete cell wall (arrow) and centered nucleoid region (Nu). Bar: 100 nm. The inset shows the continuity of the cell wall (arrow). Bar: 50 nm. B, after 3 hour of treatment with among normal cells (darker) degenerated and electron-pale cells with coagulated cytoplasm devoid of ribosomes are visible (arrowheads). The arrows point to discontinuities on the cell wall of a ghost cell. Bar: 100 nm. The inset details the discontinuity of the cell wall (arrow). Bar: 50 nm. C, Two treated cells (Nu, nucleoid), one in division (left), show labeling in the peripheral cytoplasm and along the cell wall (arrows), the latter largely missing (arrowheads, compare with the cell wall in the nearby cell, double arrowhead). Bar: 100 nm. D, other treated and degenerating protoplast showing cluster labeling (arrow) along the irregular cell periphery while gold particles are also present in the cytoplasm (arrowhead). Bar: 100 nm.

Ultrastructural analysis on *S. aureus*

The number of bacteria observed in each thin section analysed under the electron microscope (12 thin sections in total) ranged between 150-200 in *S. aureus*. In the untreated cultures most of the cells were intact and typically surrounded by a thick cell wall (Figure 5A), and few protoplasts (cells without the cell wall) but rare degenerated cells were present. In the treated culture at 95 μg/ml of Ac-CATH1, a clearly visible damage on the cell morphology interested a higher number of bacteria (roughly 30-40%) that in normal controls.

The cellular alteration varied from the disappearance of the cell wall in numerous bacterial cells that gave rise to more frequently detected protoplasts, to a cytoplasmic coagulation within the damaged protoplasts or, in other cases, to the formation of ghost cells devoid of cytoplasm content and the rupture of the cell wall and plasma membrane. The observation of the immunolabeling detected under the electron microscope, aiming to evaluate the penetration and localization of the turtle beta-defensin (TuBD-1) in the treated cells of *S. aureus*, showed that the gold particles were mainly distributed on the peripheral areas of the bacterial cells and along the cell wall (Figure 5B and 5C). Also the central cytoplasm of damaged cells and the plasma membrane of protoplasts were immunolabeled. Often the gold particles formed clusters, especially along the damaged cell wall and the plasma membrane that appeared frequently discontinuous (Figures 5C and 5D). Although observed less frequently, also the nucleoid region was immunolabeled for the turtle beta-defensin. No labelling was seen in control sections.

Discussion

Antimicrobial assays

The protocols established for testing anti-microbial peptides may give un-accurate results due to a variety of conditions such as poor solubility of the peptides, medium interactions, pH, ionic strength and salt concentration, all factors that can influence the effectiveness on the tested bacterial strains. Furthermore a medium should mimic the in vivo environmental conditions of the organism from which the peptide was originated to assure a realistic functional test, but this was not possible in our case. Various studies have analysed these interactions [21-25], but the mechanism of peptide availability to bacteria in culture has not been fully elucidated.

In our attempts to test some antimicrobial activity of our peptides, different problems arose in order to obtain a realistic MIC value that could actually correspond to the effective MIC of the condition in vivo. One problem is related to the right folding of the peptides utilized in our test since it is known that antimicrobial peptides must have a specific three-dimensional form (the effective folded peptide) in order to exert their anti-microbial effect [2,26]. Peptides without the right folding can have very little to no antimicrobial effect at all. In the present study we could not determine the concentration of the effective folded peptides within the available mix of peptides provided by the Peptide Synthetic Company, therefore the reported concentrations are only indicative and the real MIC is likely much lower. Another problem, which was mentioned above, is the poor solubility of the peptides that probably diminished the effective peptide availability in solution compared to the calculated inhibithory concentrations (Table 1).

In order to improve the antimicrobial activity of our peptides we tested some LB variants, but without success since the changes introduced influenced themselves the growth of bacterial and made it impossible to compare the results. Although the Tris/HCl medium gave similar results as in a previous study [20], negative controls showed that the medium alone caused over 90% inhibition, and therefore we could not consider this medium. Despite these problems, the qualitative results clearly showed that a sensible number of bacteria (30-80% or higher) were affected by the peptide solutions, the basis for further more quantitative pharmacological studies. In future studies, the solubility problems should be overcome if these antimicrobial peptides of reptilian origin will be tested in vivo for possible medical applications.

It is believed that antimicrobial peptide characteristics like their net charge and hydrophobicity determine their functionality. Changes in their net charge and hydrophobic ratio can influence both their antimicrobial activity and selectivity [27]. Both our peptides are cationic although their net charge is different, +8 for TuBD-1 and +1 for AcCATH-1, and the index of hydropaty is fairly low for AcCATH-1

(+0.077) when compared to TuBD-1 (+0.463). The lizard cathelicidin showed a MIC over 10 folds lowers than that the MIC of the turtle beta defensin (0.05 against 0.69 mg/ml). Also the MIC value for the total inhibition was 2 fold smaller for AcCATH-1 with respect to TuBD-1. In our testing conditions the lizard cathelicidin (AcCATH-1) that presents a moderate net charge and hydrophobicity appears to function more efficiently than the turtle beta defensin (TuBD-1). Furthermore, *E. coli* strains seem to be more sensitive to both peptides with respect to those of *S. aureus*, perhaps due to the presence of the thicker cell wall in the latter, G+ bacteria. The charge of the tested peptides resembles that of other antimicrobial peptides that in physiological conditions are generally cationic and that assume a secondary amphipathic structure in a hydrophobic environment or when encountering a cell membrane. The conformation of an amphipathic structure seems to be essential, since it forms an alpha helix which lipophilic face allows the solubilization of the peptide when it contacts the phosholipids of the bacterial membrane [28-30]. The initial target of cationic peptides is the anionic bacterial cell membrane where the positively charged peptide binds to the negatively charged phospholipids [31]. No specific receptors are involved in the binding, and this makes difficult for bacteria to develop resistance to these molecules.

In comparison to previous microbiological tests using reptilian antimicrobial peptides, the microbicide concentrations of AcCATH-1 and TuBD-1 peptides appear much higher in the conditions of our experiments mentioned above. In fact, TuBD-1 featured an inhibitory activity from 690 μg/ml up while the lower values for AcCATH-1 was at 37-69 μg/ml. Similar inhibitory concentration to those for AcCATH-1 were obtained using an antimicrobial peptide derived from snake venom (120-130 g/ml for *E. coli*, over 200-250 μg/ml for *S. aureus*; [32]). However a cathelicidin isolated from the snake Bungarus fasciatus was reported to express very low MIC values (0.6-2.3 μg/ml for *E. coli*, and 4.7 μg/ml for *S. aureus*), but for some *S. aureus* strains >100 μg/ml of peptide were needed [33]. A turtle beta-defensin (from Emys orbicularis) also showed very low MIC values, 0.65 and 5.6 μmol/L for respectively *E. coli* and *S. aureus* [34]. Another beta-defensin from the turtle Caretta caretta showed IC_{50} values of 3.3 μM for *E. coli* and 5.1 μM for *S. aureus* [35]. Finally crocodilian antimicrobial peptides (leucrocins) showed very different MIC values from as low as 0.66 up to >156 μg/ml for Staphylococcus sp. Not only the various leucrocins had a variable impact on different bacterial strains but also the bacterial strains showed varying sensitivity to the peptides [36]. These results suggest obvious differences in sensitivity among bacterial species.

Despite the IC_{50} and MIC values for AcCATH-1 and TuBD-1 peptides are apparently higher than other reptilian peptides, their morphological effect on the bacteria seen under the electronic microscope was however impressive.

Morphological alterations

The present ultrastructural study shows that the lizard Ac-CATH-1 and the turtle Tu-BD-1 peptides determine some inhibition of microbial growth that derives from the cell damage to both *E. coli* and *S. aureus* strains. After only 3 hours from the treatment, the ultrastructural analysis has clearly shown signs indicating that both the cells of *E. coli* and *S. aureus* strains are damaged at various degrees. The damage on bacterial cells was variable, often advanced, and numerous aggregates of clumped amorphous masses, often labeled with immunogold likely representing residual bacterial cells, were observed. These uncertain remnants of degenerated bacteria are also a problem in our attempt to give a quantitative esteem of the damage, another reason that makes quantitative determination of the damage very difficult in this study. Therefore in the present qualitative study the main goal was to document the degree of damage in bacterial cells (plain ultrastructure) and the penetration of the peptide inside bacterial cells (immunogold labeling).

We have not determined the effects of the peptides in strains of bacteria treated for longer periods (24 hours is a standard period for traditional antibiotics) but they would have likely been much more dramatic than those here observed after few hours from the treatment. In previous morphological studies on the damage elicited by the treatment with antimicrobial peptides on different strains of bacteria, clear signs of cytological alterations in bacterial cells were seen at 30 minutes, 1, 2, 6, 12, and 21 hours after the incubation with the different AMPs tested [19,37,38].

From the present observations it appears that the cytolytic effects are directly or un-directly elicited not only on the cell wall and the plasma membrane, possibly the primary or however the initial targets of the peptides, but also the ribosomes and the nucleoid region appear subjected to some effect of the peptides. Furthermore, the immunolocalization of the peptides within bacterial cells indicates that after the peptides have penetrated and crossed the cell wall and the plasma membrane, they localize in the nucleoid, suggesting a possible interaction with the bacterial DNA. Previous studies in the vast literature on the sites of action of AMPs have indicated that not only the plasma membrane but also the DNA, RNA and indirectly also protein synthesis can be the targets of some peptides [4,39].

Different models on possible antimicrobial mechanisms for entering the bacterial cell and act on intracellular targets have been proposed [26,40-42]. Once inside the cell the peptides may interact with RNA, DNA and protein synthesis causing their inhibition, and from our observations with immunogold labelling a possible interaction with the DNA of the bacteria is suggested.

Previous studies on the antimicrobial effect from snake peptides that were examined under the electron microscope [38] using peptide concentrations varying from 4-10 μg/ml have shown similar cytolitic effects on bacterial cells as the damages shown in the present study. The damage initially included blebbing of the plasma membranes, the rupture of the membranes with loss of cytoplasmic content, and later the clarification of the nucleoid region while ribosomes disappeared [37]. The deterioration of the bacterial cell structure later leads to the formation of ghost cells that feature a discontinuous cell wall and cell membrane, an extracted content in the cytoplasm or the presence of sparse clumped material without ribosomes, damages frequently observed in our material (Figsure 2 and 4). Similar ultrastructural degenerative aspects were also observed using another beta-defensin peptide, pelovaterin, derived from a soft shelled turtle on Gram+ *Pseudomonas sp* [18] or a snake cathelicidin on Gram- *E. coli* [19], but at a much lower dosage than in our study (12 μg/ml).

In conclusion, these data further indicate that antimicrobial peptides produced in reptiles may represent potential pharmacological drugs after a further trial of pre-clinical tests once their solubilization will be improved [8]. It has been indicated that reptiles have a very efficient innate immunity in part based on anti-microbial peptides since their acquire immunitary system is relatively slow and not as efficient as that of mammals [43]. Particularly in lizards the presence of effective peptides may be linked to the relatively low inflammatory response after wounding, a process that favors the following re-epitelialization and tissue regeneration [10], while in turtle the anti-microbial barrier impedes microbe invasion in the skin [13]. Based on this hypothesis the present explorative study has shown that a lizard

cathelicidin and a turtle beta-defensin are bacterial killers, but their potential as anti-infective agents has to be fully evaluated in further and more specifically designed microbiological and pharmacological studies.

Acknowledgments

The microscopic study was financed through the Comparative Histolab and a 2014 RFO from the University of Bologna. Dr. Sandra Turroni (Department of Pharmacy and Biotechnology, University of Bologna) supplied the *S. aureus* strains for testing. Drs. Alessio Papi and Francesca Borsetti of the Department of Bigea kindly advised in some laboratory operations.

References

1. Danilova N (2006) The evolution of immune mechanisms. J Exp Zool B Mol Dev Evol 306: 496-520.
2. Tossi A, Sandri L (2002) Molecular diversity in gene-encoded, cationic antimicrobial polypeptides. Curr Pharm Des 8: 743-761.
3. Ganz T (2003) Defensins: antimicrobial peptides of innate immunity. Nat Rev Immunol 3: 710-720.
4. Brogden KA (2005) Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol 3: 238-250.
5. Sahl HG, Pag U, Bonness S, Wagner S, Antcheva N, et al. (2005) Mammalian defensins: structures and mechanism of antibiotic activity. J Leukoc Biol 77: 466-475.
6. Zanetti M (2004) Cathelicidins, multifunctional peptides of the innate immunity. J Leukoc Biol 75: 39-48.
7. Selsted ME, Ouellette AJ (2005) Mammalian defensins in the antimicrobial immune response. Nat Immunol 6: 551-557.
8. van Hoek ML (2014) Antimicrobial peptides in reptiles. Pharmaceuticals (Basel) 7: 723-753.
9. Thomas S, Karnik S, Barai RS, Jayaraman VK, Idicula-Thomas S (2010) CAMP: a useful resource for research on antimicrobial peptides. Nucleic Acids Res 38: D774-780.
10. Alibardi L (2006) Ultrastructural and immunohistochemical observations on the process of horny growth in chelonian shells. Acta Histochem 108: 149-162.
11. Alibardi L (2010) Ultrastructural features of the process of wound healing after tail and limb amputation in lizard. Acta Zool 91: 306-318.
12. Alibardi L (2013) Granulocytes of reptilian sauropsids contain beta-defensin-like peptides: a comparative ultrastructural survey. J Morphol 274: 877-886.
13. Lorenzo A (2013) Immunolocalization of a beta-defensin (Tu-BD-1) in the skin and subdermal granulocytes of turtles indicate the presence of an antimicrobial skin barrier. Ann Anat 195: 554-561.
14. Dalla Valle L, Benato F, Maistro S, Quinzani S, Alibardi L (2012) Bioinformatic and molecular characterization of beta-defensins-like peptides isolated from the green lizard Anolis carolinensis. Dev Comp Immunol 36: 222-229.
15. Dalla Valle L, Benato F, Paccanaro MC, Alibardi L (2013) Bioinformatic and molecular characterization of cathelicidin-like peptides isolated from the green lizard Anolis carolinensis. Ital. J. Zool 80: 177-186.
16. Benato F, Dalla Valle L, Skobo T, Alibardi L (2013) Biomolecular identification of beta-defensin-like peptides from the skin of the soft-shelled turtle Apalone spinifera. J Exp Zool B Mol Dev Evol 320: 210-217.
17. Alibardi L, Celeghin A, Dalla Valle L (2012) Wounding in lizards results in the release of beta-defensins at the wound site and formation of an antimicrobial barrier. Dev Comp Immunol 36: 557-565.
18. Lakshminarayanan R, Vivekanandan S, Samy RP, Banerjee Y, Chi-Jin EO, et al. (2008) Structure, self-assembly, and dual role of a beta-defensin-like peptide from the Chinese soft-shelled turtle eggshell matrix. J Am Chem Soc 130: 4660-4668.
19. Chen XX , Yu GY, Zhan Y, Zhang Y, Shen JH, et al. (2009) Effects of the Antimicrobial Peptide OH-CATH on Escherichia coli. Zoological Research 30: 171-177.
20. Schwab U, Gilligan P, Jaynes J, Henke D (1999) In Vitro Activities of Designed Antimicrobial Peptides against Multidrug-Resistant Cystic Fibrosis Pathogens. Antimicrob. Agents Chemother 43:1435-1440.
21. Walkenhorst WF, Wolfgang Klein J, Vo P, Wimley WC (2013) pH Dependence of Microbe Sterilization by Cationic Antimicrobial Peptides. Antimicrob. Agents Chemother 57: 3312–3320.
22. Wu G, Ding J, Li H, Li L, Zhao R, et al. (2008) Effects of cations and pH on antimicrobial activity of thanatin and s-thanatin against Escherichia coli ATCC25922 and B. subtilis ATCC 21332. Curr Microbiol 57: 552-557.
23. Lee IH, Cho Y, Lehrer RI (1997) Effects of pH and salinity on the antimicrobial properties of clavanins. Infect Immun 65: 2898-2903.
24. Friedrich C, Scott MG, Karunaratne N, Yan H, Hancock RE (1999) Salt-resistant alpha-helical cationic antimicrobial peptides. Antimicrob Agents Chemother 43: 1542-1548.
25. Wei G, Campagna AN, Bobek LA (2007) Factors affecting antimicrobial activity of MUC7 12-mer, a human salivary mucin-derived peptide. Annals of Clinical Microbiology and Antimicrobials 6: 1-10.
26. Klüver E, Adermann K, Schulz A (2006) Synthesis and structure-activity relationship of beta-defensins, multi-functional peptides of the immune system. J Pept Sci 12: 243-257.
27. Zelezetsky I, Pag U, Sahl HG, Tossi A (2005) Tuning the biological properties of amphipathic alpha-helical antimicrobial peptides: Rational use of minimal amino acid substitutions. Peptides. 26: 2368-2376.
28. Maloy WL, Kari UP (1995) Structure-activity studies on magainins and other host defense peptides. Biopolymers 37: 105-122.
29. Saberwal G and Nagaraj R (1994) Cell-lytic and antibacterial peptides that act by perturbing the barrier function of membranes: Facets of their conformational features, structure-function correlations and membrane-perturbing abilities. Biochim. Biophys. Acta. 1197: 109–131.
30. Kaiser ET, Kézdy FJ (1987) Peptides with affinity for membranes. Annu Rev Biophys Biophys Chem 16: 561-581.
31. Powers JP, Hancock RE (2003) The relationship between peptide structure and antibacterial activity. Peptides 24: 1681-1691.
32. Sachidananda MK, Murari SK, Channe Gowda D (2006) Characterization of an antibacterial peptide from Indian cobra (Naja naja) venom. J. Venom. Anim. Toxins incl. Trop. Dis. 13: 446-461.
33. Wang Y, Hong J, Liu X, Yang H, Liu R, et al. (2008) Snake cathelicidin from Bungarus fasciatus is a potent peptide antibiotics. PLoS One 3: e3217.
34. Stegemann C, Kolobov A Jr, Leonova YF, Knappe D, Shamova O, et al. (2009) Isolation, purification and de novo sequencing of TBD-1, the first beta-defensin from leukocytes of reptiles. Proteomics 9: 1364-1373.
35. Chattopadhyay S, Sinha NK, Banerjee S, Roy D, Chattopadhyay D, et al. (2006) Small cationic protein from a marine turtle has beta-defensin-like fold and antibacterial and antiviral activity. Proteins 64: 524-531.
36. Pata S, Yaraksa N, Daduang S, Temsiripong Y, Svasti J, et al. (2011) Characterization of the novel antibacterial peptide Leucrocin from crocodile (Crocodylus siamensis) white blood cell extracts. Dev Comp Immunol 35: 545-553.
37. Friedrich CL, Moyles D, Beveridge TJ, Hancock RE (2000) Antibacterial action of structurally diverse cationic peptides on gram-positive bacteria. Antimicrob Agents Chemother 44: 2086-2092.
38. Lee HY, Andalibi A, Webster P, Moon SK, Teufert K, et al. (2004) Antimicrobial activity of innate immune molecules against Streptococcus pneumoniae, Moraxella catarrhalis and nontypeable Haemophilus influenzae. BMC Infectious Diseases. 4:12.
39. Lehrer RI, Barton A, Daher KA, Harwig SS, Ganz T, et al. (1989) Interaction of human defensins with Escherichia coli. Mechanism of bactericidal activity. J Clin Invest 84: 553-561.
40. Falla TJ, Karunaratne DN, Hancock RE (1996) Mode of action of the antimicrobial peptide indolicidin. J Biol Chem 271: 19298-19303.

41. Boman HG, Agerberth B, Boman A (1993) Mechanisms of action on Escherichia coli of cecropin P1 and PR-39, two antibacterial peptides from pig intestine. Infect Immun 61: 2978-2984.
42. Hsu CH, Chen C, Jou ML, Lee AY, Lin YC, et al. (2005) Structural and DNA-binding studies on the bovine antimicrobial peptide, indolicidin: Evidence for multiple conformations involved in binding to membranes and DNA. Nucleic Acids Res. 33: 4053–4064.
43. Zimmerman LM, Vogel LA, Bowden RM (2010) Understanding the vertebrate immune system: insights from the reptilian perspective. J Exp Biol 213: 661-671.

2.3 Results for EDMTFH: a bird specific EDC protein found in feathers and embryonic subperiderm

2.3.1 Introduction to paper III: Immunolocalization of a histidine-rich Epidermal Differentiation Protein in the chicken supports the hypothesis of an evolutionary developmental link between the embryonic subperiderm and feather barbs and barbules.

In this paper, my colleagues and I report the immunolocalization of the chicken protein EDMTFH under light microscopy and electron microscopy and the detection of isolated EDMTFH by immunoblotting. EDMTFH (Epidermal Differentiation Protein starting with a Met-Thr-Phe motif and rich in Histidine) is shown to be identical to histidine-rich protein (HRP), which had previously been identified as a component of feathers. The gene encoding this protein is positioned on the epidermal differentiation complex (EDC) which is involved in providing numerous structural proteins for cornifying skin cells in amniotes including birds.

During this study EDMTFH was localized in both the subperiderm, a layer of the embryonic epidermis and in feather barbs and barbules of the feather follicle. This co-localization of EDMTFH supports an important hypothesis that links the evolution and morphogenesis of feathers. The subperiderm is a transient embryonic layer present in birds, but also in crocodilians; therefore its origin predates the one of feathers. A model of feather morphogenesis (Sawyer & Knapp, 2003; Sawyer et al., 2005) proposes that the feather sheath corresponds to the embryonic secondary periderm, barbs and barbules to the embryonic subperiderm and the marginal plate of barb ridges to the proliferative layer of the embryonic epidermis proper.

In the present study we found EDMTFH (or HRP) is expressed in feather cells that undergo hard cornification, which is the process that converts cells into components of a hard cornified skin appendage. This is similar to the expression pattern of feather CBPs, as also shown in this study by

the co-localization with immunogold labelling of EDMTFH and feather CBP (beta-keratin). Besides EDMTFH, the epidermal differentiation complex genes encoding feather CBP and EDCRP are expressed in the embryonic subperiderm and in feathers.

The specific position in the scale subperiderm and in the feather follicle of EDMTFH, combined to previous obtained results for other EDC proteins, provides evidence that supports the hypothesized model in which the cyclically regenerating feather follicle is topologically, developmentally and evolutionarily related to the embryonic epidermis of archosaurs.

PLOS ONE

RESEARCH ARTICLE

Immunolocalization of a Histidine-Rich Epidermal Differentiation Protein in the Chicken Supports the Hypothesis of an Evolutionary Developmental Link between the Embryonic Subperiderm and Feather Barbs and Barbules

Lorenzo Alibardi[1], Karin Brigit Holthaus[1,2], Supawadee Sukseree[2], Marcela Hermann[3], Erwin Tschachler[2], Leopold Eckhart[2]*

1 Comparative Histolab and Dipartimento di Scienze Biologiche, Geologiche ed Ambientali (BiGeA), University of Bologna, Bologna, Italy, **2** Research Division of Biology and Pathobiology of the Skin, Department of Dermatology, Medical University of Vienna, Vienna, Austria, **3** Department of Medical Biochemistry, Medical University of Vienna, Vienna, Austria

* leopold.eckhart@meduniwien.ac.at

Citation: Alibardi L, Holthaus KB, Sukseree S, Hermann M, Tschachler E, Eckhart L (2016) Immunolocalization of a Histidine-Rich Epidermal Differentiation Protein in the Chicken Supports the Hypothesis of an Evolutionary Developmental Link between the Embryonic Subperiderm and Feather Barbs and Barbules. PLoS ONE 11(12): e0167789. doi:10.1371/journal.pone.0167789

Editor: Michel Simon, INSERM, FRANCE

Received: September 17, 2016

Accepted: November 21, 2016

Published: December 9, 2016

Data Availability Statement: All relevant data are within the paper and its Supporting Information files.

Funding: The study was supported by the Austrian Science Fund (FWF): P23801 (LE) and Austrian Science Fund (FWF): P28004 (LE) (https://www.fwf.ac.at/en/). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Abstract

The morphogenesis of feathers is a complex process that depends on a tight spatiotemporal regulation of gene expression and assembly of the protein components of mature feathers. Recent comparative genomics and gene transcription studies have indicated that genes within the epidermal differentiation complex (EDC) encode numerous structural proteins of cornifying skin cells in amniotes including birds. Here, we determined the localization of one of these proteins, termed EDMTFH (Epidermal Differentiation Protein starting with a MTF motif and rich in Histidine), which belongs to a group of EDC-encoded proteins rich in aromatic amino acid residues. We raised an antibody against an EDMTFH-specific epitope and performed immunohistochemical investigations by light microscopy and immunogold labeling by electron microscopy of chicken embryos at days 14–18 of development. EDMTFH was specifically present in the subperiderm, a transient layer of the embryonic epidermis, and in barbs and barbules of feathers. In the latter, it partially localized to bundles of so-called feather beta-keratins (corneous beta-proteins, CBPs). Cells of the embryonic periderm, the epidermis proper, and the feather sheath were immunonegative for EDMTFH. The results of this study indicate that EDMTFH may contribute to the unique mechanical properties of feathers and define EDMTFH as a common marker of the subperiderm and the feather barbules. This expression pattern of EDMTFH resembles that of epidermal differentiation cysteine-rich protein (EDCRP) and feather CBPs and is in accordance with the hypothesis that a major part of the cyclically regenerating feather follicle is topologically, developmentally and evolutionarily related to the embryonic subperiderm.

Competing Interests: The authors have declared that no competing interests exist.

Introduction

The cornified skin barrier of amniotes and cornified skin appendages such as claws, hair, and feathers are formed by epidermal keratinocytes that differentiate by inducing the expression of specific sets of genes [1–4]. The stratification of the epidermis begins in the embryo and involves the establishment of the periderm as the superficial layer in all amniotes and the formation of a subperiderm in archosaurs [5–10]. Periderm and subperiderm are shed during late development when a mature cornified layer (stratum corneum) has been established by the definitive epidermis. During adult life the cornified epidermis provides the essential protection against water loss and mechanical stress whereas cornified skin appendages serve various functions including, but not limited to, grasping (claws), thermoinsulation (hair, feathers), and facilitating flight (feathers).

Many of the structural components of cornifying keratinocytes are encoded in a gene cluster termed the Epidermal Differentiation Complex (EDC). In humans and other mammals, the EDC comprises genes for proteins that interact with each other to form, via transglutamination, a cornified cell envelope or with the keratin intermediate filaments during compaction of the cytoskeleton [11, 12]. Recent studies have shown that non-mammalian amniotes also have an EDC in which both orthologs of human EDC genes, such as *loricrin* and *cornulin*, and clade-specific genes which, for example, code for proteins of the scutes of turtle or the feathers of birds are located [13–15]. Proteins traditionally termed beta-keratins [16–18] but now identified as Corneous Beta Proteins (CBPs) represent a major sub-cluster of EDC genes in sauropsids while they are absent in mammals [8, 19]. These proteins of a molecular mass typically in the range of 10–18 kDa possess a characteristic central region (with a most highly conserved stretch of 34 amino acid residues) that folds into an anti-parallel beta-sheet and facilitates the formation of CBP filaments of 3–4 nm thickness [20, 21]. The intra- and intermolecular interactions of sub-domains and sequence motifs in other avian EDC proteins have remained unknown so far. Conserved sequence motifs at the amino- and carboxy-terminus of EDC proteins are likely sites of transglutamination whereas an extremely high cysteine content of epidermal differentation cysteine-rich protein (EDCRP) has been proposed to form multiple disulfide bonds that may contribute to the mechanical strengthening of feathers [22]. Labeling with tritiated histidine and autoradiography suggested that histidine-containing proteins are present in the cytoplasm and in corneous bundles of barbules, however, the identity of these protein(s) was not determined in that study [23].

The morphogenesis and maturation of feathers depends on a complex spatio-temporal cell differentiation program in which EDC-encoded and non-EDC-encoded proteins form the body of the feather whereas other proteins regulate the scaffolding function and programmed cell death of intermediate cells [14, 24–27]. Similarities in the topology and gene expression profiles have suggested that the layered organization of feather follicles is equivalent to that of the embryonic epidermis, with the feather sheath corresponding to the embryonic periderm, the barbules corresponding to the embryonic subperiderm and the marginal plate of barb ridges corresponding to the proliferative layer of the embryonic epidermis proper [7, 9, 28]. As the evolutionary origin of the subperiderm in a common ancestor of birds and alligators [28] predated that of feathers, the evolutionary origin of feathers has probably involved the coption of embryonic cell type-specification and differentiation mechanisms for a morphogenesis program of a skin appendage that is cyclically renewed in adult birds [22, 28].

One of the EDC proteins of the chicken is EDMTFH (Epidermal Differentiation protein starting with MTF motif, Met-Thr-Phe, and rich in Histidine) [13]. Its expression was detected by RT-PCR in embryonic skin and feathers and, by proteomics, in feathers [13]. Here, we compared the sequence of EDMTFH to that of chicken histidine-rich protein (HRP), also

known as fast protein (Fp) [29], a major feather protein [30]. We show that, due to local sequence mismatches, previously reported HRP/Fp sequences are not compatible with the reference genome sequence of the chicken, and we suggest that the amino acid sequence of EDMTFH represents the translation product of the gene that has previously been referred to as HRP or Fp. Furthermore, we demonstrate that EDMTFH is expressed in feather barbules and in the subperiderm, thereby adding support to the hypothesis of a close relationship between these two epithelial derivatives.

Materials and Methods

Ethics statement

All animal procedures were approved by the Animal Care and Use Committee of the Medical University of Vienna (Permit Number: 66.016/0014-II/3b/2011), all efforts were made to minimize suffering of animals, and all procedures were conducted according to the guidelines established by the Committee.

Animals, tissue preparation and fixation

Sexually mature Derco brown (TETRA-SL) laying hens and roosters were purchased from Diglas Co. (Feuersbrunn, Austria), maintained on open floor space with free access to water and feed (standard diet, ssniff, Germany) with a daily light period of 16 hours. For fertilized eggs, hens and roosters were housed together in flocks in the animal facility. Freshly laid and fertilized eggs were incubated at 37.5˚C and 60–70% humidity to maintain normal embryonic development. For tissue and organ retrieval, chicken embryos on embryonic days E12 through E19 were euthanized by decapitation.

Tissue samples from chicken embryos were prepared, fixed with 7.5% formaldehyde and embedded in paraffin as described previously [14]. For ultrastructural investigations, skin samples were collected from chick embryos at stages 38–40 as previously reported [31]. The collected tissues were immediately fixed for 5 hours in cold (0–4˚C) 4% paraformaldehyde in 0.1 M phosphate buffer at pH 7.4, rinsed in buffer for about 30 minutes, dehydrated in ethanol (70%, 80%, 95%, 100%), and immersed in Bioacryl resin for 3–5 hours (pieces sunk to the bottom of the container) before curing them under ultraviolet light at 0–4˚C for 3 days [32].

Generation of an antibody against EDMTFH

The peptide DHRFKHLYGLHRDHHHD, corresponding to amino acid residues 29–45 of chicken EDMTFH, was synthesized and coupled to keyhole limpet haemocyanin (KLH) by Davids Biotechnologie GmbH, Regensburg, Germany. The KLH-coupled peptide was used as immunogen for the generation of mouse antiserum (Davids Biotechnologie GmbH), essentially according to a published protocol [33]. Immunohistochemistry with antiserum dilutions of 1:250–1:1000 gave specific signals which were not obtained when the primary antibody was omitted or when pre-immune serum or mouse antibodies of unrelated specificities were used instead of anti-EDMTFH.

Western blot analysis

Chicken embryonic feather samples (stage 38) were homogenized in a solubilization buffer containing 8 M urea, 50 mM Tris-HCl (pH 7.6), 0.1 M 2-mercaptoethanol, 1 mM dithiothreitol and protease inhibitor (Sigma). The particulate material was removed by centrifugation at 10,000 g for 10 minutes. Laemmli buffer was added and samples were denatured at 100˚C for 5 minutes. Proteins (40 μg per lane) were separated by sodium dodecyl sulfate-

polyacrylamide gel electrophoresis (SDS-PAGE) at a polyacrylamide concentration of 15% using a Biorad apparatus. The Sigma Wide Range molecular weight marker (10–250 kDa) was used for estimating protein masses. After electrophoresis, proteins were transferred onto a nitrocellulose membrane. The membrane was stained with Ponceau Red to visualize the protein transfer. Mouse anti-EDMTFH at a dilution of 1:1000 was used as the primary antibody, and a fluorescence labeled goat anti-mouse immunoglobulin G (IgG-h+I Cy5 conjugated, Bethyl) was used as secondary antibody. Bands were detected using the Biorad external laser Molecular Imager FX combined with the program PharosFX. In negative control experiments, the samples were subjected to the same procedure but the primary antibody was omitted.

Light and electron microscopy immunolabeling analyses

Immunohistochemical stainings for light microscopy were performed according to a published protocol [14]. Mouse anti-EDMTFH was used at dilutions of 1:250 and 1:500, and biotinylated sheep anti-mouse IgG (1:200; GE, Chalfont, UK) was used as secondary antibody. Sheep serum (10%) was added to the secondary antibody to prevent unspecific binding. Finally, the sections were incubated with streptavidin-biotin-horseradish peroxidase (HRP) complex and 3-amino-9-ethylcarbazole (DakoCytomation, Glostrup, Denmark), and counterstained with hematoxylin. In control experiments, the primary antibody was either replaced by pre-immune serum or preabsorbed with the antigenic peptide (Davids Biotechnologie GmbH, Regensburg, Germany). The preabsorption procedure was modified from a published protocol [34]. Two μl anti-EDMTFH antibody, 0.5 μl antigenic peptide (10 mg/ml) and 37.5 μl phosphate-buffered saline containing 2% BSA were mixed and incubated at room temperature for 30 min. Subsequently, the antibody was further diluted to the final concentration and used for the immunostaining protocol as described above.

For the electron microscopy study, samples of wing downfeathers (stages 38 Hamburger–Hamilton (HH), n = 3, and 39 HH, n = 3) embedded in Bioacryl resin were sectioned using an ultramicrotome, and 2–4 μm thick sections were collected on glass slides, stained with 1% toluidine blue and observed under a light microscope for general histology. From areas of interest, thin sections of 40–90 nm in thickness were collected on Nickel grids for the immunodetection by immunogold under a transmission electron microscope. In order to improve antibodies penetration, a hatching treatment for 10 minutes with 2% HIO_4 was done on grid sections, and the grids were rinsed in distilled water for 23 minutes with two changes. Thin sections were pre-incubated for 10 minutes in 0.05 M TRIS-HCl buffer at pH 7.4, containing 1% Cold Water Fish Gelatin. The sections were then incubated for 5 hours at room temperature in primary antibodies diluted 1:100 in buffer. Mouse anti-EDMTFH and, in some experiments, a rabbit "feather keratin" antibody (generously donated from Dr. R. H. Sawyer, University of South Carolina, USA, see [9, 35]) were used as primary antibodies. In controls, the primary antibody was omitted in the first incubation step. The sections were rinsed in buffer and incubated for 1 hour at room temperature with secondary anti-mouse immunoglobulin (IgG) (for detection of EDMTFH) or anti-rabbit IgG (for detection of feather beta keratin) gold-conjugated antibodies (Sigma, USA, 5 or 20 nm gold particles). In double-labeling experiments anti-rabbit IgG 20 nm diameter gold-conjugates and anti-mouse IgG 5 nm diameter gold conjugates were used for the detection of corneous feather beta-proteins (feather keratins) and for the detection of EDMTFH, respectively. After incubation, the grids were rinsed in buffer, dried, stained for 5 minutes with 2% uranyl acetate, and observed under the electron microscope Zeiss 10C/CR operating at 60 kV.

Results

EDMTFH corresponds to the previously reported histidine-rich protein (HRP) of chick feather

Amino acid sequence alignments showed that EDMTFH is identical to the previously reported chicken histidine-rich protein (HRP) [29, 36] with the exception of the carboxy-terminal segment (Fig 1). Re-investigation of the previously published cDNA sequence from which the carboxy-terminus of HRP had been derived by translation *in silico* [29] suggested that two indel changes in the nucleotide sequence, inducing a frameshift relative to the chicken genome sequence and the sequence of EDMTFH cDNA [13], had caused an incorrect prediction of the carboxy-terminal amino acid sequence of HRP (Fig 1A). Our previous search for EDMTFH peptides in the chicken feather proteome [13, 37] revealed two EDMTFH-derived peptides (Fig 1A, green underlines) of which one comprised a part of the carboxy-terminal amino acid sequence present in EDMTFH but not in the predicted HRP.

The amino-terminal sequence of EDMTFH is identical to a 20-amino acid peptide previously identified by direct peptide sequencing of HRP [29] (Fig 1A, blue underline) and highly similar to the sequences of peptides reported for so-called HRP-B proteins [38] (Fig 1B). The 5′-untranslated region of HRP/Fp [39] matches perfectly to the non-coding sequences in exon 1 and at the 5′-end of exon 2 of *EDMTFH* (S1 Fig), while the coding sequence of the HRP cDNA [29] (with the sequence differences shown in Fig 1A) is entirely derived from exon 2 of the *EDMTFH* gene (S1 Fig). As the EDMTFH sequence, determined from a chicken cDNA [13], matches perfectly with the chicken reference genome sequence whereas the previously reported HRP and HRP-B sequences show only partial identities, we keep using the name EDMTFH instead of HRP.

EDMTFH belongs to a group of epidermal differentiation proteins rich in aromatic amino acid residues

The EDMTFH gene is located in the EDC and is flanked by the CBP gene EDbeta and EDMTF4 [13] (Fig 2A). EDMTF4 is most similar to EDMTFH among chicken proteins, followed by EDMTF1 through 3, which are located next to EDMTF4 (Fig 2B). An internal peptide of EDMTFH that differs in sequence from all its EDMTF paralogs (Fig 2B, underlined) was selected as an immunogen for raising an EDMTFH-specific antibody for *in situ* immunolocalization studies (see below). The cysteine contents of EDMTFH and EDMTF4 are much lower than that of other EDMTF proteins (1–2% versus 11–13%). A high histidine content is present only in EDMTFH, however, aromatic amino acids (F, W, Y, and H) are enriched in all EDMTF proteins.

To determine the evolutionary conservation of EDMTFH among birds, we screened avian genome and protein data using the amino acid sequence of chicken EDMTFH as a query. Homologs of EDMTFH were identified in all birds investigated and the carboxy-terminal sequence was highly conserved (Fig 2C). The genes encoding EDMTFH homologs in other species were located at genome positions of shared synteny as compared to chicken EDMTFH and the encoded proteins showed higher sequence similarity to EDMTFH than to other chicken proteins.

Histidine was present at high amounts in EDMTFH of chicken, turkey and quail, which are representatives of the family Phasianidae (Fig 2C). In other birds, the proteins most similar to EDMTFH had lower contents of histidine (15.2% in chicken EDMTFH) but shared with EDMTFH the high content of aromatic residues (40.4% in chicken EDMTFH). Notably, chicken EDMTFH (total number of amino acid residues: 99) contains 12 sites in which an

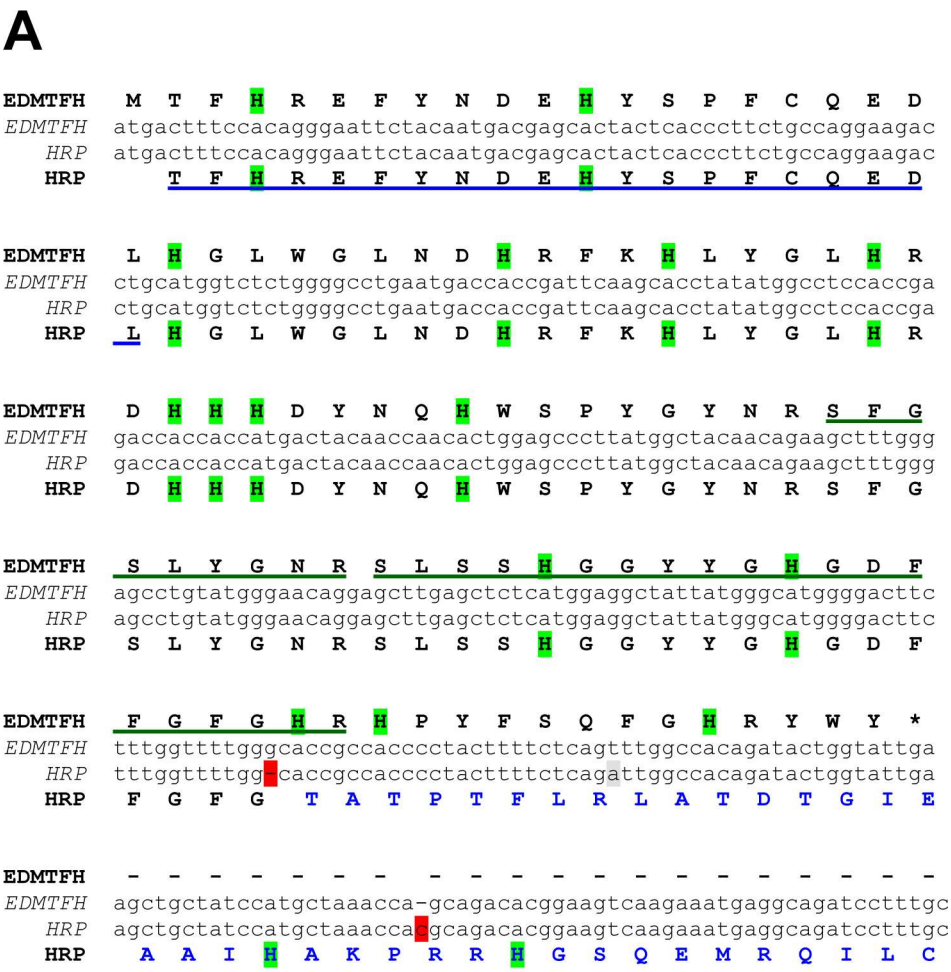


Fig 1. Nucleotide and amino acid sequence alignments of EDMTFH versus histidine-rich protein (HRP). (**A**) The nucleotide sequences of the coding region of chicken EDMTFH [13] and of the chicken HRP cDNA reported previously [29] were aligned. Translations into amino acid sequences are shown above and below the sequences, respectively. Note that insertions and deletions (red shading) in the cDNA sequence relative to the chicken EDMTFH gene in the current genome assembly cause reading frameshifts leading to the prediction of a different carboxy-terminus of HRP (blue fonts) relative to EDMTFH. Sequences corresponding to peptides that were previously identified in feather extracts are marked underlined feather proteins peptides (underlined) corresponding to EDMTFH were identified by (blue underline [29], green underlines [13]). Histidine (H) residues are highlighted by green shading. The stop codon of EDMTFH is marked with an asterisk. (**B**) Alignment of amino-terminal amino acid sequences of EDMTFH [13] and HRP, as determined by direct sequencing of proteins isolated from feathers [29, 38]. Predicted HRP-B residues that deviate from the EDMTFH sequence at positions of histidines (H) are shaded grey.

doi:10.1371/journal.pone.0167789.g001

aromatic residue is followed by glycine, and a similar enrichment for such dipeptides is present in the most similar proteins of other avian species (Fig 2C). Together, the sequence features of

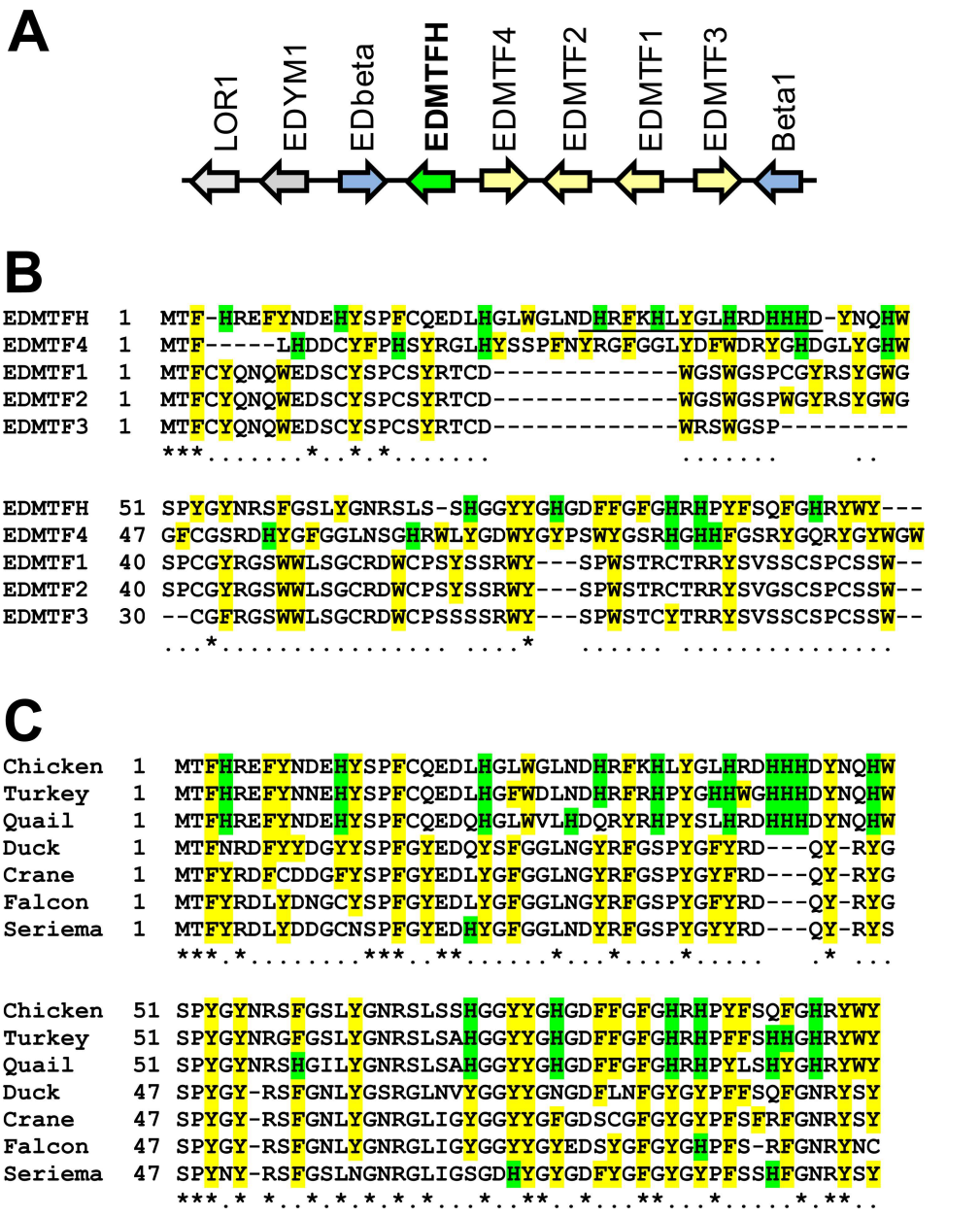


Fig 2. Comparison of chicken EDMTFH versus other EDMTF proteins of the chicken and homologs in other species. (**A**) EDMTFH gene locus in the chicken. Arrows indicate the origientation of gene transcription. The genes *EDbeta* and *Beta1* (preliminary names [13]) encode corneous beta-proteins (CBPs), also known as beta-keratins. (**B**) Amino acid sequence alignment of EDMTFH versus other chicken EDMTF proteins. The peptide used as an epitope for raising the anti-EDMTFH antibody is underlined. (**C**) Amino acid sequence alignment of chicken EDMTFH versus the most similar proteins of other birds. Formats in (**B**) and (**C**) indicate the following: Histidine (H) residues are highlighted by green shading, other aromatic residues are highlighted by yellow shading. Numbers indicate amino acid sequence positions. Identity of residues in all sequences is indicated by an asterisk and conservation in at least 50% of the sequences is indicated by "." below the alignments. Hyphens were introduced to maximize the alignment of the sequences. Sequences of EDMTFH orthologs of turkey (*Meleagris gallopavo*) and quail (*Coturnix japonica*) were predicted from genomic DNA. Accession numbers of other EDMTFH sequences: AHA62422.1 (chicken, *Gallus gallus*), XP_012964640.1 (duck, *Anas platyrhynchos*), XP_010299618.1 (crane, *Balearica regulorum gibbericeps*), XP_013153676.1 (falcon, *Falco peregrinus*), XP_009701209.1 (seriema, *Cariama cristata*).

doi:10.1371/journal.pone.0167789.g002

EDMTFH and their differential conservation during evolution suggest that the high histidine content arose specifically in the avian clade Phasianidae while the high content of aromatic amino acid residues is conserved and therefore likely important for the function of EDMTFH.

EDMTFH is present in the subperiderm and in feather barbules

For *in situ* immunolocalization studies, antibodies were raised against the EDMTFH-specific peptide DHRFKHLYGLHRDHHHD (Fig 2B). Western blot analysis confirmed that the antibody bound to a feather protein of the expected size of approximately 12 kDa whereas larger proteins such as CBPs and keratins were not labeled (S2 Fig). Embryonic skin and feather follicles as well as adult skin of chickens were immunohistochemically stained with this antiserum and, as negative controls, with preimmune serum and antibodies of unrelated specificities. Furthermore, in some control experiments the primary antibody was preadsorbed with the antigenic peptides. EDMTFH was most strongly expressed in growing feathers on embryonic days E14 and E18 whereas the skin between feather follicles was EDMTFH-negative. EDMTFH was concentrated in barbule cells (Fig 3A–3C, 3E and 3F) whith external cells of barbule plates being most strongly labeled (Fig 3C). Barb cortical and medullary cells were also labeled for EDMTFH (Fig 3F), however, barbs (rami) were immunonegative in feathers that appeared to have progressed further in differentiation, perhaps indicating that the EDMTFH epitope was masked during cornification. Negative control experiments did not yield staining of barbules and, thereby, confirmed the specificity of the EDMTFH immunostaining (Fig 3D and 3F). The feather sheath lacked EDMTFH (Fig 3A–3C and 3F).

EDMTFH was also expressed in the subperiderm of scutate scales on the legs of chicks on embryonic days E18 and E19 (Fig 3H). The immunolabeling pattern was irregular, and the labeling intensity appeared to decrease with cornification of the subperidermal cells. Negative control stainings in which the primary antibody was either replaced by the preimmune serum or preabsorbed with the antigen were negative, confirming the specificity of the staining (Fig 3I). Other parts of the embryonic skin, including the dermis, the basal and lower suprabasal layers of the epidermis, the cornified cell layers of scales and the periderm consistently lacked expression of EDMTFH (Fig 3A).

Ultrastructural localization of EDMTFH

To determine the subcellular localization of EDMTFH, immunogold labeling and transmission electron microsopy were performed using the anti-EDMTFH antibody. EDMTFH immunogold labeling was consistently detected in the external barbules cells while it became uneven in barb cortical cells and disappeared in barb medullary cells of the feather samples investigated here. The labeling was mainly, but not exclusively, observed over the dense CBP packets accumulated among the paler cytoplasm (Fig 4A, S3 Fig). A similar diffuse labeling was detected in the subperiderm of scales (Fig 4B). No antibody-conjugated gold particles were present in control sections (Fig 4C). The gold labeling with the anti-EDMTFH antibody over the linear filaments of most barbule cells and some barb cortical cells was further observed after silver enhancement, a technique that increases the size of the ultrastructural label to allow a more panoramic view of the labeling over broader areas of barbule cells (Fig 4D–4F). Double labeling with anti-EDMTFH, conjugated to 5 nm gold particles, and anti-feather CBP, conjugated to 20 nm gold particles, suggested that EDMTFH at least partly co-localized with feather CBP (Fig 4G). The labeling for feather CBP was observed over the corneous bundles in most barbule and barb cells and, consequently, the entire cell appeared labeled, except for the central cytoplasm where scattered gold particles were present. The total amount of label for EDMTFH

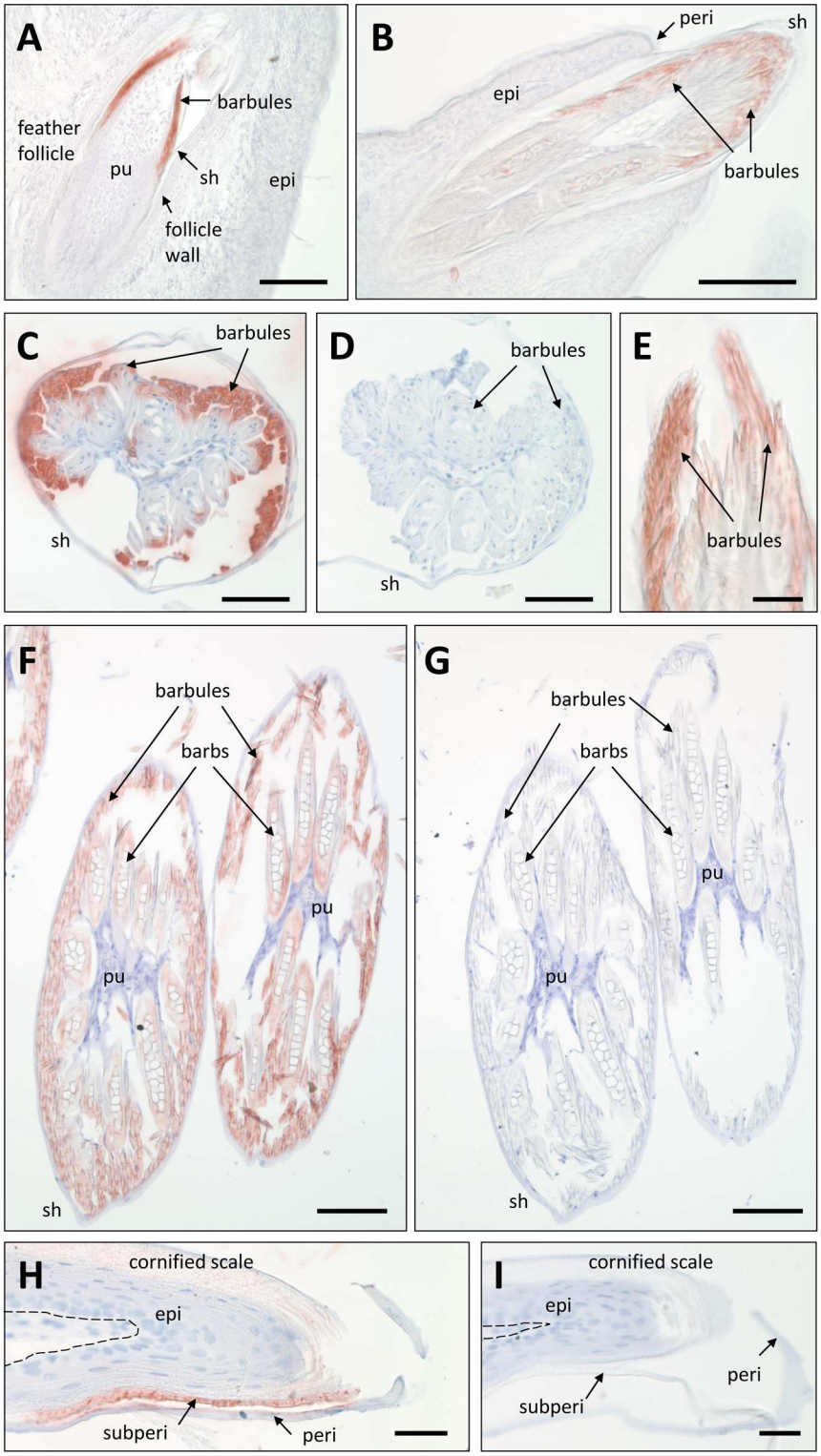


Fig 3. Light microscopic immunohistochemistry of EDMTFH. (**A-C**, **E, F**) Feather follicles and feathers on the wings of chick embryos on day E18 of development were immunostained with anti-EDMTFH (red). In negative control experiments the primary antibody was replaced with an antiserum raised against an unrelated peptide (**D**) or preabsorbed with the antigen (**G**). Scutate scales on the legs of chick embryos (day E19) were immunolabeled for EDMTFH (**H**) or subjected to the negative control experiment (**I**). The dermo-epidermal junction is indicated by a dashed line (**H**, **I**). epi, epidermis; peri, periderm; pu, pulp; sh, feather sheath; subperi, subperiderm. Bars: 100 μm (**A**, **B**, **F**, **G**), 50 μm (**C-E**), 25 μm (**H**, **I**).

doi:10.1371/journal.pone.0167789.g003

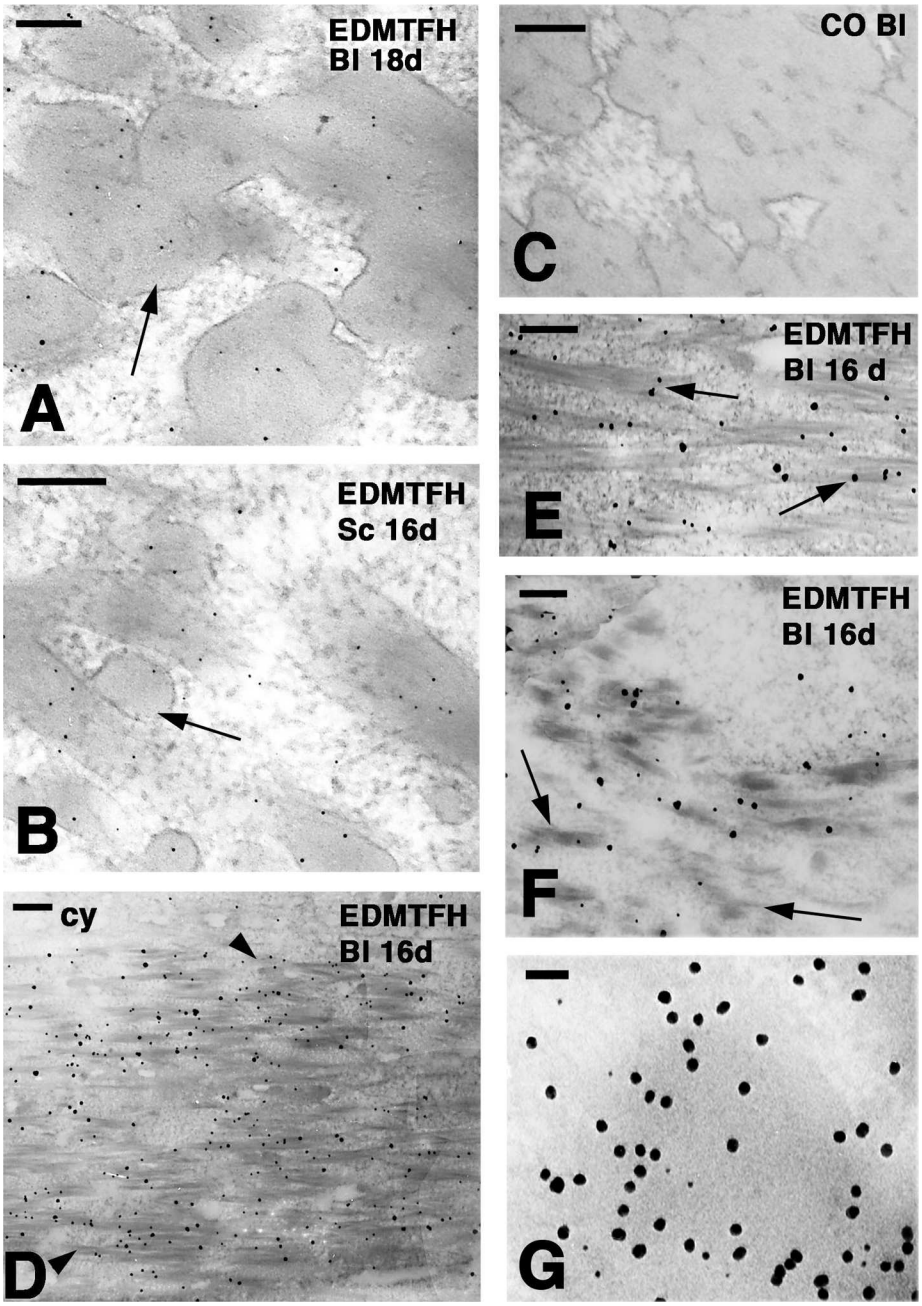


Fig 4. Ultrastructural localization of EDMTFH by immunogold labeling. Downfeathers and scales of chicken embroys at days 16 (16d) and 18 (18d) of development were labeled for EDMTFH either without (**A-C, G**) and with (**D-F**) silver enhancement. (**A**) Diffuse labeling over corneous bundles (arrow) of barbule cells (bl). (**B**) Diffuse labeling in the corneous bundles (arrow) of a subperiderm cell in a scale. (**C**) Immuno-negative control section of a barbule. (**D**) Labeling cytoplasmic corneous bundles (arrowheads) but not the cytoplasm (cy) in a barbule cell (bl). (**E**) Close-up to show the association of the labeling with corneous bundles (arrows). (**F**) Early differentiating barbule cell with short corneous bundles (arrows). (**G**) Double-labeling for EDMTFH (5 nm gold particles) and feather beta-keratin (20 nm gold particles) in a barbule cell. Note that the large particles appear to be more abundant than the small particles. A lower magnification image of the double-labeling is shown in S3 Fig. Bars: 100 nm (**A**, **B**); 200 nm (**C-F**); 50 nm (**G**).

doi:10.1371/journal.pone.0167789.g004

appeared to be lower than that for feather CBP, and in contrast to the even distribution of feather CBP, EDMTFH was variably concentrated in different areas of barbule cells (S3 Fig).

Discussion

The results of this study show that EDMTFH is expressed in the subperiderm of the embryonic epidermis and in barbs and barbules of feathers (Fig 5). Our data improve and extend previous studies in which HRP/Fp was suggested to have a main role in feathers [29, 30, 36]. Together with comparative sequence analysis, the immunolabeling results point to a role of EDMTFH in the maturation of the cornified components of mature feathers.

Research on EDMTFH, then termed HRP or Fp, was initiated in the 1970ies and yielded the important insight that a histidine-rich protein is present at relatively high amounts in chicken feathers as well as in embryonic epidermis [29, 30, 36, 38, 39]. Revisiting this topic, we found that a published cDNA sequence of chicken HRP contained nucleotide changes relative to the chicken genome sequence, leading to an aberrant prediction of the carboxy-terminus of the protein [29]. In our studies, an EDMTFH cDNA (GenBank accession number KC963987) [13] showed sequence identity to the chicken genome sequence, and a peptide identified by proteomic analysis of feathers supported the carboxy-terminal sequence of EDMTFH but not that of HRP (Fig 1A). These data validate the EDMTFH sequence presented here and indicate that the prediction of the carboxy-terminus of the HRP sequence was caused by the presence of a polymorphism in the animal from which the HRP mRNA was isolated [29] or, more likely, by a cDNA cloning artifact.

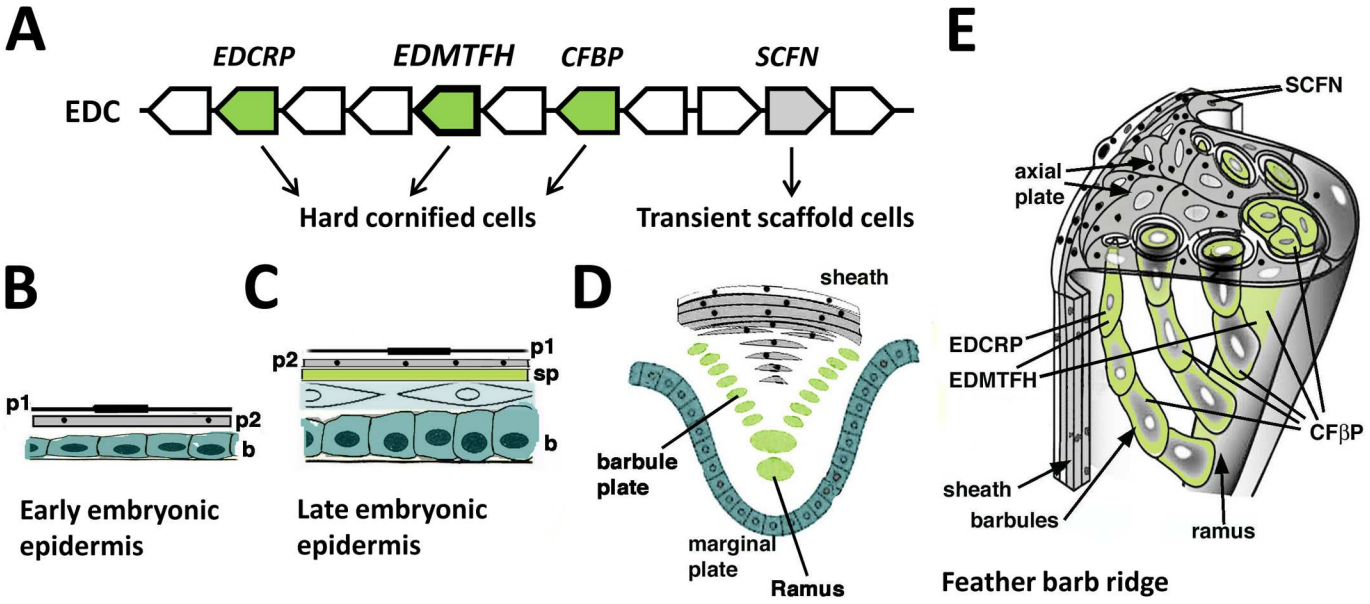


Fig 5. Schematic representation of the contribution of EDMTFH to feather cornification. (**A**) EDMTFH and other genes, that encode proteins of the feather follicle, are located in the same gene cluster, the avian epidermal differentiation complex (EDC). EDMTFH, EDCRP and corneous feather beta proteins (CFBPs) are components of hard cornified cells whereas scaffoldin (SCFN) is a component of cells that form a transient scaffold of growing feathers. (**B-D**) The embryonic epidermis of birds increases the number of layers during development. During late development, 2 layers of periderm (p1, p2) and a subperiderm (sp) are present above the definitive epidermis (**C**). The feather barb ridge has a topologically similar organization as late embryonic epidermis on scutate scales whereby the equivalents of the periderm form the feather sheath and the axial plate, and barbules cells and rami are equivalent to the subperiderm (**D**), as suggested by common expression of protein markers. A 3-dimensional depiction of a growing barb ridge of a down feather indicates the different roles of cornifying cells (green) and transient scaffolding cells (grey) in the morphogenesis of feathers (**E**). Periderm granules containing scaffoldin are indicated as black dots.

doi:10.1371/journal.pone.0167789.g005

The antibody that was generated in our study was directed to an internal EDMTFH-specific peptide. Detection by Western blot and immunochemical staining of EDMTFH in chicken feathers confirmed peptide isolation and sequencing [29] and proteomics results of feathers (see [13]). The generation of the EDMTFH antibody facilitated the localization of this protein *in situ*. Immunohistochemical staining could be done by standard antigen retrieval with citrate buffer (pH 6) whereas ultrastructural immunogold labeling required an etching protocol [31]), indicating that the EDMTFH epitope was at least partly masked in the embedded tissues. Labeling was observed in many cells of feather barbules and in the subperiderm of scutate scales, however, it is possible that the epitope of EDMTFH is masked by tight protein-protein interactions including transglutamination so that a fraction of the EDMTFH proteins is not accessible for immunolabeling. In agreement with this notion, EDMTFH immunolabeling was detected in the rami (barbs) of many but not all developing feathers. Thus, the immunolabeling pattern likely represents the distribution of EDMTFH at cellular terminal differentiation stages prior to full cornification.

The present study shows that EDMTFH is expressed in feather cells that undergo hard cornification, i.e. the conversion into components of a hard cornified skin appendage [40, 3] (Fig 5). Double-labeling of EDMTFH protein and feather CBPs at the ultrastructural level (Fig 4G) supported the expectation that in these cells EDMTFH is less abundant than CBPs. The latter are encoded by a family of more than a hundred genes, many of which are co-expressed in all types of feathers [25]. The EDMTFH protein appeared to be added to the corneus bundles of barbule cells while it was present at smaller (immuno-detectable) amounts in barb cortical and absent in medullary cells. These results indicate that EDMTFH is one of the structural proteins, besides CBPs, that form the cytoskeleton of maturing feathers.

The expression pattern of EDMTFH is similar to that of EDCRP and clearly different from that of scaffoldin, another EDC-encoded protein that we have previously detected in the embryonic periderm and feather sheath [14, 41]. EDCRP has an extraordinarily high content of cysteine residues and, therefore, it likely forms multiple disulfide bonds which might contribute to the cross-linking of cytoskeletal proteins and the hardening of feathers. This process is supposed to resemble the maturation of mammalian hair fibers in which cysteine-rich keratin-associated proteins (KAPs, also known as Krtaps) are expressed [22, 42]. Interestingly, another class of KAPs is rich in glycine and tyrosine [43], and these proteins have been suggested to contribute to the mechanical properties of hair by establishing bidirectional protein interactions via cation-π interactions or π stacking [44, 45]. The latter type of protein-protein interactions depends on regularly arranged aromatic residues (in this case, tyrosine). It is interesting to note that glycine and tyrosine-rich KAPs are similar to EDMTF proteins, including EDMTFH, with regard to size and amino acid sequence (S4 Fig). As both types of proteins are expressed at sites of hard cornification (KAPs in hair, EDMTFH in feathers), functional analogy may be presumed. This hypothesis remains to be tested in future studies.

EDMTFH and related EDMTF proteins are conserved among diverse species of birds whereby only the high content of aromatic residues but not the high content of histidine is conserved outside the clade comprising chicken, turkey and quail (Phasianidae) (Fig 2C). EDMTFH is not homologous to mammalian filaggrin, which has also been referred to as histidine-rich protein [46]. Filaggrin is an S100 fused-type protein encoded by a gene in the mammalian EDC [47, 12]. It has a histidine content of 10%, undergoes proteolytic degradation and gives rise to free histidine as a precursor of the UV-absorbing substance urocanic acid in the cornified layer of the epidermis of mammals, whereas it is absent in birds [48–50, 14]. A recent paper found that a chromosomal locus containing *EDMTFH* appeared to be associated with red feather coloration in a crossing experiment of common canaries and red siskins [51]. However, no mechanistic link between *EDMTFH* and the color of feathers was identified. In further

studies, it will be interesting to investigate whether adaptation of EDMTF genes, such as the rise of the histidine content of the EDMTFH protein in Phasianidae, was associated with specific changes in feather properties.

Our results demonstrate that, besides the feather follicle, EDMTFH is expressed in the subperiderm of the embryonic epidermis (Figs 3–5). The immunolabeling obtained with the anti-EDMTFH antibody is similar to the previously reported distribution pattern of the so-called HRP-B protein, as determined using an antibody against HRP isolated from feathers [38]. This congruence of immunolabelings further supports the identity of EDMTFH and HRP, provided that the previously reported HRP sequence is corrected at the carboxy-terminus as outlined in Fig 1). A similar expression pattern as that of EDMTFH/HRP has been detected, by immunohistochemistry, for feather-type CBP (beta-keratin) [9, 35], and, by mRNA *in situ* hybridization, for EDCRP [22]. Together, these studies provide substantial amount of evidence in support of the hypothesis that feather barbs and barbules are related, in terms of evolution and development, to the embryonic subperiderm [7, 9, 22, 29, 52]. Further studies of EDMTFH may help to shed more light into the molecular basis of the evolutionary origin, the growth and the material properties of feathers.

Supporting Information

S1 Fig. The nucleotide sequence of a partial HRP/Fp cDNA matches the 5′-untranslated region of EDMTFH. (**A**) Nucleotide sequence published by Presland and colleagues [39]. (**B**) Alignment of the HRP/Fp cDNA sequence and the *EDMTFH* gene sequence. The *EDMTFH* sequence was derived from the current chicken reference genome sequence (GenBank Accession number NC_006112.3). Nucleotide number 1 of *EDMTFH* in this alignment corresponds to NC_006112.3 nucleotide number 1977520, and *EDMTFH* is transcribed from the minus strand. The proximal promoter region including a TATA box-like element (underlined) is shown with blue fonts. Intronic sequences are marked by red fonts. The start codon is highlighted by green shading. (**C**) Schematic depiction of the exon-intron structure of the chicken *EDMTFH* gene. Color code: blue, non-transcribed regions flanking the gene; black, exons; red, intron; green, start codon; yellow, stop codon.
(PDF)

S2 Fig. Western blot analysis using the anti-EDMTFH antibody. Protein was extracted from embryonic feathers of chicken, electrophoresed through a 15% polyacrylamide gel and blotted onto a nitrocellulose membrane. After Ponceau staining of total protein (right panel), the membrane was probed with anti-EDMTFH (primary antibody) and fluorescence-labeled goat anti-mouse immunoglobulin G (secondary antibody). In the negative (neg.) control experiment, the primary antibody was omitted. Positions of molecular mass markers are indicated on the left. kDa, kilo-Dalton.
(PDF)

S3 Fig. Double immunolabeling of EDMTFH and feather-type corneous beta protein in barbule cells. Low magnification view of double (DOUB) immunolabeling for EDMTFH (small gold particles, highlighted by red circles) and feather corneous beta protein (large gold particles) in barbule cells at stage 37–38 of development. Feather beta keratin labeling was concentrated over beta packets (dark) that are surrounded by the less electron-dense cytoplasm (cy). EDMTFH labeling is sparse in both cytoplasm and beta packets. n, nucleus. Bar, 200 nm.
(PDF)

S4 Fig. Amino acid sequence alignment of chicken (*Gallus gallus*, Gg) EDMTFH and human (*Homo sapiens*, Hs) keratin-associated protein (KAP/Krtap)7-1. Positions of

identical residues in both proteins are indicated by * below the alignment. Conservation of aliphatic (I, L, M, V) and hydrophilic (S, T, N, Q, E, D, K, R) residues in indicated by ":" and ".". Aromatic residues (F, H, W, Y) are highlighted by yellow shading and glycine (G) residues are highlighted by grey shading.
(PDF)

Acknowledgments

The authors thank Bahar Golabi, Bettina Strasser, Veronika Mlitz, Julia Lachner, and Michael Mildner for excellent technical support and helpful discussions.

Author Contributions

Conceptualization: LA KBH SS MH ET LE.

Funding acquisition: LE.

Investigation: LA KBH SS LE.

Methodology: LA KBH SS MH LE.

Resources: LA MH LE.

Supervision: LA LE.

Visualization: LA LE.

Writing – original draft: LA KBH LE.

Writing – review & editing: LA KBH SS MH ET LE.

References

1. Widelitz RB, Jiang TX, Yu M, Shen T, Shen JY, Wu P, et al. Molecular biology of feather morphogenesis: a testable model for evo-devo research. J Exp Zool B Mol Dev Evol. 2003; 298:109–122. doi: 10.1002/jez.b.29 PMID: 12949772
2. Alibardi L. Adaptation to the land: The skin of reptiles in comparison to that of amphibians and endotherm amniotes. J Exp Zool B Mol Dev Evol. 2003; 298:12–41. doi: 10.1002/jez.b.24 PMID: 12949767
3. Eckhart L, Lippens S, Tschachler E, Declercq W. Cell death by cornification. Biochim Biophys Acta. 2013; 1833:3471–3480. doi: 10.1016/j.bbamcr.2013.06.010 PMID: 23792051
4. Alibardi L. Sauropsids cornification is based on corneous beta-proteins, a special type of keratin-associated corneous proteins of the epidermis. J Exp Zool B Mol Dev Evol. 2016; 326:338–351. doi: 10.1002/jez.b.22689 PMID: 27506161
5. Sawyer RH, Abbott UK, Fry GN. Avian scale development. III. Ultrastructure of the keratinizing cells of the outer and inner epidermal surfaces of the scale ridge. J Exp Zool. 1974; 190:57–70. doi: 10.1002/jez.1401900105 PMID: 4436621
6. Sawyer RH, Knapp LW, O'Guin WM. Epidermis, dermis and appendages. In: Bereiter-Hahn J, Matoltsy G, Richards KS, editors. Biology of the Integument. 2 Vertebrates. Berlin, Heidelberg: Springer; 1986. pp. 194–238.
7. Sawyer RH, Knapp LW. Avian skin development and the evolutionary origin of feathers. J Exp Zool B Mol Dev Evol. 2003; 298:57–72. doi: 10.1002/jez.b.26 PMID: 12949769
8. Alibardi L, Knapp LW, Sawyer RH. Beta-keratin localization in developing alligator scales and feathers in relation to the development and evolution of feathers. J Subm Cytol Pathol. 2006; 38:175–192.
9. Sawyer RH, Salvatore BA, Potylicki TT, French JO, Knapp LW. Origin of feathers: feather beta-keratins are expressed in discrete epidermal cell populations of embryonic scutate scales. J Exp Zool B Mol Dev Evol. 2003; 295:12–24. doi: 10.1002/jez.b.5 PMID: 12548540
10. Greenwold MJ, Sawyer RH. Linking the molecular evolution of avian beta keratins to the evolution of feathers. J Exp Zool B Mol Dev Evol. 2011; 316:609–616. doi: 10.1002/jez.b.21436 PMID: 21898788

11. Candi E, Schmidt R, Melino G. The cornified envelope: a model of cell death in the skin. Nat Rev Mol Cell Biol. 2005; 6:328–340. doi: 10.1038/nrm1619 PMID: 15803139
12. Henry J, Toulza E, Hsu CY, Pellerin L, Balica S, Mazereeuw-Hautier J, et al. Update on the epidermal differentiation complex. Front Biosci (Landmark Ed). 2012; 17:1517–1532.
13. Strasser B, Mlitz V, Hermann M, Rice RH, Alibardi L, Tschachler E, et al. Evolutionary origin and diversification of epidermal barrier proteins in amniotes. Mol Biol Evol. 2014; 31:3194–3205. doi: 10.1093/molbev/msu251 PMID: 25169930
14. Mlitz V, Strasser B, Jaeger K, Hermann M, Ghannadan M, Buchberger M, et al. Trichohyalin-like proteins have evolutionarily conserved roles in the morphogenesis of skin appendages. J Invest Dermatol. 2014; 134:2682–2692.
15. Holthaus KB, Strasser B, Sipos W, Schmidt HA, Mlitz V, Sukseree S, et al. Comparative genomics identifies epidermal proteins associated with the evolution of the turtle shell. Mol Biol Evol. 2016; 33:726–737. doi: 10.1093/molbev/msv265 PMID: 26601937
16. Greenwold MJ, Sawyer RH. Genomic organization and molecular phylogenies of the beta (beta) keratin multigene family in the chicken (Gallus gallus) and zebra finch (Taeniopygia guttata): implications for feather evolution. BMC Evol Biol. 2010; 10:148. doi: 10.1186/1471-2148-10-148 PMID: 20482795
17. Greenwold MJ, Sawyer RH. Expression of archosaurian beta-keratins: diversification and expansion of archosaurian beta-keratins and the origin of feather beta-keratins. J Exp Zool B Mol Dev Evol. 2013; 320:393–405. doi: 10.1002/jez.b.22514 PMID: 23744807
18. Greenwold MJ, Bao W, Jarvis ED, Hu H, Li C, Gilbert MTP, et al. Dynamic evolution of the alpha (α) and beta (β) keratins has accompanied integument diversification and the adaptation of birds into novel lifestyles. BMC Evol Biol. 2014; 14:249–265. doi: 10.1186/s12862-014-0249-1 PMID: 25496280
19. Alibardi L. Keratinization and lipogenesis in epidermal derivatives of the zebrafinch Taeniopygia guttata castanotis (Aves, Passeriformes, Ploecidae) during embryonic development. J Morphol. 2002; 251:294–308. doi: 10.1002/jmor.1090 PMID: 11835366
20. Fraser RD, Parry DA. Molecular packing in the feather keratin filament. J Struct Biol. 2008; 162:1–13. doi: 10.1016/j.jsb.2008.01.011 PMID: 18334302
21. Calvaresi M, Eckhart L, Alibardi L. The molecular organization of the beta-sheet region in corneous beta-proteins (beta-keratins) of sauropsids explains its stability and polymerization into filaments. J Struct Biol. 2016; 194:282–91. doi: 10.1016/j.jsb.2016.03.004 PMID: 26965557
22. Strasser B, Mlitz V, Hermann M, Tschachler E, Eckhart L. Convergent evolution of cysteine-rich proteins in feathers and hair. BMC Evol Biol. 2015; 15:1–11.
23. Alibardi L. Ultrastructural localization of tritiated histidine in downfeathers of the chick in relation to the synthesis of keratin-associated proteins. Cell Tiss Org. 2005; 182:35–47.
24. Alibardi L. Review: cornification, morphogenesis and evolution of feathers. Protoplasma 2016.
25. Ng CS, Chen CK, Fan WL, Wu P, Wu SM, Chen JJ, et al. Transcriptome analyses of regenerating adult feathers in chicken. BMC Genomics 2015; 16:1–16.
26. Wu P, Ng CS, Yan J, Lai YC, Chen CK, Lai YT, et al. Topographical mapping of -and -keratins on developing chcicken skin integument: functional interaction and evolutionary perspectives. Proc Natl Acad Sci U S A. 2015; 112:E6770– E6779. doi: 10.1073/pnas.1520566112 PMID: 26598683
27. Yu M, Wu P, Widelitz RB, Chuong CM. The morphogenesis of feathers. Nature. 2002; 420:308–312. doi: 10.1038/nature01196 PMID: 12442169
28. Sawyer RH, Rogers L, Washington L, Glenn TC, Knapp LW. Evolutionary origin of the feather epidermis. Dev Dyn. 2005; 232:256–267. doi: 10.1002/dvdy.20291 PMID: 15637693
29. Rogers GE. Genes for hair and avian keratins. Ann N Y Acad Sci. 1985; 455:403–25. PMID: 2417523
30. Walker ID, Rogers GE. Differentiation in avian keratinocytes. The properties of the proteins of the chick down feather. Eur J Biochem. 1976; 69:329–339.
31. Alibardi L. Immunolocalization of alpha-keratins and feather beta-proteins in feather cells and comparison with the general process of cornification in the skin of mammals. Ann Anat. 2013; 195:189–198. doi: 10.1016/j.aanat.2012.08.005 PMID: 23228339
32. Scala C, Cenacchi G, Ferrari C, Pasquinelli G, Preda P, Manara G. A new acrylic resin formulation: a useful tool for histological, ultrastructural, and immunocytochemical investigations. J Histochem Cytochem. 1992; 40:1799–1804. PMID: 1431065
33. Eckhart L, Dalla Valle L, Jaeger K, Ballaun C, Szabo S, Nardi A, et al. Identification of reptilian genes encoding hair keratin-like proteins suggests a new scenario for the evolutionary origin of hair. Proc Natl Acad Sci U S A 2008; 105:18419–18423. doi: 10.1073/pnas.0805154105 PMID: 19001262

34. Fischer H, Rossiter H, Ghannadan M, Jaeger K, Barresi C, Declercq W, et al. Caspase-14 but not caspase-3 is processed during the development of fetal mouse epidermis. Differentiation. 2005; 73:406–413. doi: 10.1111/j.1432-0436.2005.00046.x PMID: 16316411

35. Sawyer RH, Glenn T, French B, Mays B, Shames RB, Barnes GL, et al. The expression of beta (β) keratins in the epidermal appendages of reptiles and birds. Am Zool. 2000; 40:530–539.

36. Powell BC, Rogers GE. Isolation of messenger RNA coding for the fast protein of embryonic chick feather. Nucl Acid Res. 1979; 7:2165–2176.

37. Rice RH, Winters BR, Durbin-Johnson BP, Rocke DM. Chicken corneocyte cross-linked proteome. J Proteome Res. 2013; 12:771–776. doi: 10.1021/pr301036k PMID: 23256538

38. Barnes GL, Sawyer RH. Histidine-rich protein B of embryonic feathers is present in the transient embryonic layers of scutate scales. J Exp Zool. 1995; 271:307–314. doi: 10.1002/jez.1402710408 PMID: 7722473

39. Presland RB, Gregg K, Molloy PL, Morris CP, Crocker LA, Rogers GE. Avian keratin genes. I. A molecular analysis of the structure and expression of a group of feather keratin genes. J Mol Biol. 1989; 209:549–559. PMID: 2479754

40. Alibardi L, Toni M, Dalla Valle L. Hard cornification in reptilian epidermis in comparison to cornification in mammalian epidermis. Exp Dermatol. 2007; 16:961–976. doi: 10.1111/j.1600-0625.2007.00609.x PMID: 18031455

41. Alibardi L, Mlitz V, Eckhart L. Immunolocalization of scaffoldin, a trichohyalin-like protein in the epidermis of the chicken embryo. Anat Rec. 2015; 298:479–487.

42. Rogers MA, Langbein L, Winter H, Ehmann C, Praetzel S, Korn B, et al. Characterization of a cluster of human high/ultrahigh sulfur keratin-associated protein genes embedded in the type I keratin gene domain on chromosome 17q12-21. J Biol Chem. 2001; 276:19440–19451. doi: 10.1074/jbc.M100657200 PMID: 11279113

43. Rogers MA, Langbein L, Winter H, Ehmann C, Praetzel S, Schweizer J. Characterization of a first domain of human high glycine-tyrosine and high sulfur keratin-associated protein (KAP) genes on chromosome 21q22.1. J Biol Chem. 2002; 277:48993–49002. doi: 10.1074/jbc.M206422200 PMID: 12359730

44. Waters ML. Aromatic interactions in model systems. Curr Opin Chem Biol. 2002; 6:736–741. PMID: 12470725

45. McGaughey GB, Gagné M, Rappé AK. pi-Stacking interactions. Alive and well in proteins. J Biol Chem. 1998; 273:15458–15463. PMID: 9624131

46. Ramsden M, Loehren D, Balmain A. Identification of a rapidly labelled 350K histidine-rich protein in neonatal mouse epidermis. Differentiation. 1983; 23:243–249. PMID: 6406285

47. Irvine AD, McLean WH, Leung DY. Filaggrin mutations associated with skin and allergic diseases. N Engl J Med. 2011; 365:1315–1327. doi: 10.1056/NEJMra1011040 PMID: 21991953

48. Mildner M, Jin J, Eckhart L, Kezic S, Gruber F, Barresi C, et al. Knockdown of filaggrin impairs diffusion barrier function and increases UV sensitivity in a human skin model. J Invest Dermatol. 2010; 130:2286–2294. doi: 10.1038/jid.2010.115 PMID: 20445547

49. Pendaries V, Malaisse J, Pellerin L, Le Lamer M, Nachat R, Kezic S, et al. Knockdown of filaggrin in a three-dimensional reconstructed human epidermis impairs keratinocyte differentiation. J Invest Dermatol. 2014; 134:2938–2946. doi: 10.1038/jid.2014.259 PMID: 24940654

50. Barresi C, Stremnitzer C, Mlitz V, Kezic S, Kammeyer A, Ghannadan M, et al. Increased sensitivity of histidinemic mice to UVB radiation suggests a crucial role of endogenous urocanic acid in photoprotection. J Invest Dermatol. 2011; 131:188–194. doi: 10.1038/jid.2010.231 PMID: 20686493

51. Lopes RJ, Johnson JD, Toomey MB, Ferreira MS, Araujo PM, Melo-Ferreira J, et al. Genetic basis for red coloration in birds. Curr Biol. 2016; 26:1427–1434. doi: 10.1016/j.cub.2016.03.076 PMID: 27212400

52. Alibardi L. Cells of embryonic and regenerating germinal layers within barb ridges: implication for the development, evolution and diversification of feathers. J Submicr Cytol Path. 2006; 38:51–76.

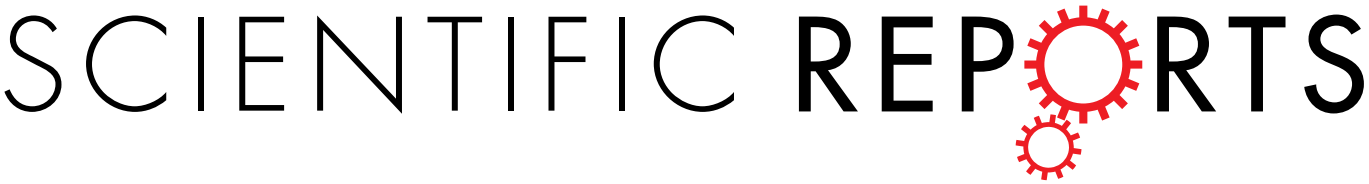
2.4 Results for the epidermal differentiation proteins of Serpentes

2.4.1 Introduction to paper IV: Identification and comparative analysis of the epidermal differentiation complex in snakes.

This paper reports the results of a study on the clade of serpentes (snakes), specialized squamates that have lost their limbs and limb-associated appendages like claws. The genes of the EDC of snakes were identified and a scenario for the evolution of the cornification proteins in squamates developed. Comparative genomics and gene expression screening were performed for two snake species, the king cobra and the Burmese python, and the green anole lizard. We identified snake EDC genes that encode homologs of human skin barrier proteins, such as loricrin, cornulin, SPRR-like proteins and antimicrobial peptidoglycan recognition protein 3. The EDC of snakes also contains genes encoding corneous beta proteins (beta-keratins), which are similar to those of the green anole lizard. A corneous beta protein containing not one but four pleated beta sheets, unique to squamates, was investigated during this analysis.

Several EDC genes present in the green anole lizard and the Japanese gecko and, by phylogenetic inference, also in an ancestor of snakes, have been lost in snakes. Screening the green anole trascriptome database did not suggest limb-specific roles of these proteins. Unique to this clade was the discovery of two scaffoldin (SCFN) proteins that are trichohyalin-like proteins. Therefore, the total number of members belonging to the S100 fused-type family in snakes is three instead of two like in most sauropsids and only one in the anole lizard. Divergent evolution of the two snake SCFNs was suggested by differences in the promoter sequences and by the loss of the conserved C-terminal motif in snake SCFN2. Unexpectedly, we identified multiple genes encoding short proteins with cysteine contents between 20 and 45% in the EDC of both snakes and the lizard. Cysteine-rich amino acid sequences were previously considered to be specific for hard skin appendages such as claws that require disulfide bonds between cysteine residues for stabilization.

Our results provide a characterization of the snake EDC and a catalogue of its protein products, an information which will be helpful in future studies. Furthermore these results suggest that adaptations of the EDC, including an increase in cysteine-dependent protein cross-linking, have facilitated the evolution of a mechanically highly resilient cornified skin surface in snakes and other squamates.



OPEN

Identification and comparative analysis of the epidermal differentiation complex in snakes

Received: 14 November 2016
Accepted: 22 February 2017
Published: 27 March 2017

Karin Brigit Holthaus[1,2], Veronika Mlitz[1], Bettina Strasser[1], Erwin Tschachler[1], Lorenzo Alibardi[2] & Leopold Eckhart[1]

The epidermis of snakes efficiently protects against dehydration and mechanical stress. However, only few proteins of the epidermal barrier to the environment have so far been identified in snakes. Here, we determined the organization of the Epidermal Differentiation Complex (EDC), a cluster of genes encoding protein constituents of cornified epidermal structures, in snakes and compared it to the EDCs of other squamates and non-squamate reptiles. The EDC of snakes displays shared synteny with that of the green anole lizard, including the presence of a cluster of corneous beta-protein (CBP)/beta-keratin genes. We found that a unique CBP comprising 4 putative beta-sheets and multiple cysteine-rich EDC proteins are conserved in all snakes and other squamates investigated. Comparative genomics of squamates suggests that the evolution of snakes was associated with a gene duplication generating two isoforms of the S100 fused-type protein, scaffoldin, the origin of distinct snake-specific EDC genes, and the loss of other genes that were present in the EDC of the last common ancestor of snakes and lizards. Taken together, our results provide new insights into the evolution of the skin in squamates and a basis for the characterization of the molecular composition of the epidermis in snakes.

Snakes are reptiles that have lost their limbs during evolution and developed a unique predatory lifestyle that involves the ability to swallow prey of a diameter larger than that of their own body[1–3]. The skin of snakes, and more specifically the epidermis, consists of rigid scales and soft inter-scale regions, which together provide both mechanical resistance and flexibility[4,5]. Snakes belong to the squamate reptiles which are characterized by the regular shedding of the outer layers of the epidermis, also known as ecdysis[6,7]. In snakes, the superficial layers of the epidermis are detached as a single, coherent sheet whereas other squamates (lizards and geckos) shed multiple smaller flakes. While the dynamic regulation and composition of snake skin has been partially revealed over the past fifty years[7–9], the recent availability of whole genome sequences of snakes and other reptiles allows, for the first time, to identify genes that encode epidermal proteins and, by comparative genomics, to establish a basis for building hypotheses on the molecular evolution of the epidermis in snakes[10–12].

In all amniotes, keratinocytes proliferate in the basal layer and differentiate in the suprabasal layers of the epidermis[13,14]. Signaling between the epidermis and the underlying dermis controls the patterning of the epidermis and the formation of skin appendages[15–18]. While differentiation of keratinocytes in the mammalian epidermis involves continuous alterations of cell structures and movement of cells towards the body surface, keratinocyte differentiation in squamates results in the formation of distinct non-interconvertible layers that remain stable for several weeks before they are shed together[8,9]. In its final differentiation stage, the outer generation of the epidermis comprises a clear, lacunar, alpha, mesos, beta, and oberhautchen layer[7–9]. Prior to shedding, the outer generation of the epidermis protects the resting and newly forming inner generation of the epidermis. At the beginning of skin development, the cell layers of the embryo-specific periderm cover the epidermis[19]. The histological and ultrastructural features of squamate skin have been reported[7–9,20]. However, only few aspects of the molecular architecture of alpha and beta-layers of squamate epidermis have been determined so far[21–24].

Corneous beta-proteins, traditionally called beta-keratins, have been identified as components of the epidermis in snakes, like in other reptiles[22,25–27]. Keratin intermediate filament proteins, previously referred to as alpha-keratins, are the main cytoskeletal proteins in the epidermis and skin appendages of vertebrates. A

[1]Research Division of Biology and Pathobiology of the Skin, Department of Dermatology, Medical University of Vienna, Vienna, Austria. [2]Dipartimento di Scienze Biologiche, Geologiche ed Ambientali (BiGeA), University of Bologna, Bologna, Italy. Correspondence and requests for materials should be addressed to L.E. (email: leopold.eckhart@meduniwien.ac.at)

cysteine-rich keratin component of reptilian claws has been lost due to gene inactivation during the evolution of snakes[28]. The presence of various ultrastructurally, but not biochemically, defined epidermal components such as fibers in the beta-layer[29] and different granules in the oberhautchen and the clear layer of reptiles[21] indicate that many structural proteins of snake epidermis remain to be identified.

The recent availability of genome and transcriptome sequences from multiple vertebrates has allowed the determination of genes implicated in epidermal structure and function. Based on the dermatologically relevant characterization of human epidermal barrier genes, we have screened non-mammalian tetrapods for homologs of a gene cluster known as the Epidermal Differentiation Complex (EDC)[30–34]. Genes encoding S100 fused-type proteins (SFTPs), which are homologous to a subgroup of mammalian EDC genes[35], were found in amphibians[34], and more complex gene clusters homologous to the mammalian EDC were identified in the chicken[30,32], in the green anole lizard (*Anolis carolinensis*)[30] and in turtles[33].

Here we extend the comparative analysis of the EDC in sauropsids and determine the gene complement of the EDC in snakes. We characterize the amino acid sequences of EDC-encoded proteins, suggest hypotheses about their contributions to the molecular architecture of the epidermis, and identify cases of gain and loss of specific EDC genes during the evolution of stem lepidosaurs and snakes.

Results

Identification of epidermal differentiation complex (EDC) genes in snake genomes. The EDCs of the Burmese python (*Python bivittatus*) and the king cobra (*Ophiophagus hannah*) were defined as the genomic regions flanked by S100A genes, like in other amniotes[30,33]. The gene complement of the EDC of these snakes was identified by tBLASTn searches using EDC-encoded proteins of *A. carolinensis*, chicken and humans as queries and by *de novo* prediction of genes in an iterative process, as described previously[33]. The predicted amino acid sequences of snake EDC proteins were used as queries in tBLASTn searches in the published transcriptomes of snakes to test for the expression of the predicted genes.

The nomenclature for EDC genes follows the preliminary system defined in previous studies[30]. In brief, gene names consist of the term Epidermal Differentiation (ED) followed by a term that describes the amino acid composition or the presence of particular amino acid sequence motifs in the encoded protein. For easier readability, only the abbreviations are used in the text whereas the full names of genes are summarized in Supplementary Table S1. Exceptions to this naming convention were made to indicate orthologs of human loricrin and cornulin and chicken scaffoldin.

The EDC of snakes is largely syntenic with that of the green anole lizard. The structure of the EDC is very similar in the Burmese python, the king cobra, and the green anole lizard (Fig. 1). Like in other amniotes[30,33], the EDC of snakes is bordered by S100A genes and comprises a peptidoglycan recognition protein 3 (*PGLYRP3*) gene, simple (single coding exon) EDC (SEDC) genes, and SFTP genes (Suppl. Tables S2-6; Suppl. Figures S1 and S2). A gene homologous to *EDKM* of lizards, turtles and birds is located between the *PGLYRP3* and the SEDC genes (Fig. 1). In the draft genome of the king cobra, the genes *EDSQ* and *EDEPT*, the orthologs of which are neighbors in the python genome, were separated by a series of genes not related to the classical EDC genes and a sequence gap. This pattern indicates that a gene rearrangement event might have disrupted the canonical organization of the EDC in this species (Fig. 1, §).

To test whether the predicted EDC genes of snakes are expressed, tissue transcriptomes of snakes were screened by tBLASTn searches. Indeed, transcript reads corresponding to most EDC genes (Suppl. Tables S2 and S3), were detected in skin transcriptomes of the ball python (*Python regius*)[36] (Suppl. Fig. S3) and of the painted saw-scaled viper (*Echis coloratus*)[37], whereas the transcriptomes of internal organs of snakes included no or only very small numbers of EDC gene transcripts (Suppl. Fig. S4), suggesting a skin-specific expression of most EDC genes.

A unique corneous beta-protein (CBP) comprising 4 beta-sheets is conserved in squamates. SEDC genes form a continuous cluster in snakes and include a sub-cluster of genes that encode corneous beta-proteins (CBPs), also known as beta-keratins[14]. These proteins are characterized by a conserved core domain that is predicted to form a beta-sheet[38]. The CBP cluster gene is located between the *loricrin* and *EDYM2* genes of snakes (Fig. 1, Fig. 2). It is syntenic with the CBP locus of the green anole lizard and with the main CBP loci of birds and turtles[30,33]. Within the CBP gene cluster, 35 and 36 CBP genes, here termed *Beta1* through *Beta36* in order of the arrangement of the genes, were identified in the python and cobra, respectively, which is comparable to the 40 CBP genes present in the green anole lizard[26], but lower than the 71 CBP genes reported for the Japanese gecko[39].

Both the python and cobra have an ortholog of the lizard gene *Beta1*, previously termed *Li-Ac40*[26] (Fig. 2). This gene encodes a protein that contains 4 CBP core sequence motifs and therefore is predicted to form 4 beta-sheets (Fig. 3; Suppl. Fig. S5) whereas all other CBPs identified so far comprise only a single beta-sheet domain. Remarkably, classical CBPs undergo dimerization via face-to-face interactions between their beta-sheets and subsequently they assemble into a beta-fibril (beta-filament) in which, according to the classical model[38,40,41], 4 dimers form one turn of the helical structure. We put forward a hypothetical model in which 2 Beta1 proteins dimerize via their 4 beta-sheets and thereby form one complete turn of the helical structure of a beta-fibril (Fig. 3). The integration of Beta1 dimers into fibrils likely occurs via edge-to-edge interactions in a manner equivalent to that proposed for dimers of CBPs with 1 beta-sheet domain[41,42]. As there are currently no experimental data on sauropsidian CBPs, that could resolve the structure of beta-fibrils at atomic resolution[41], the integration of the complete Beta1 sequences of squamates into ongoing computer modelling attempts[42] will extend the scope of these studies beyond the investigation of interactions between isolated beta-sheets. As Beta1 homologs are

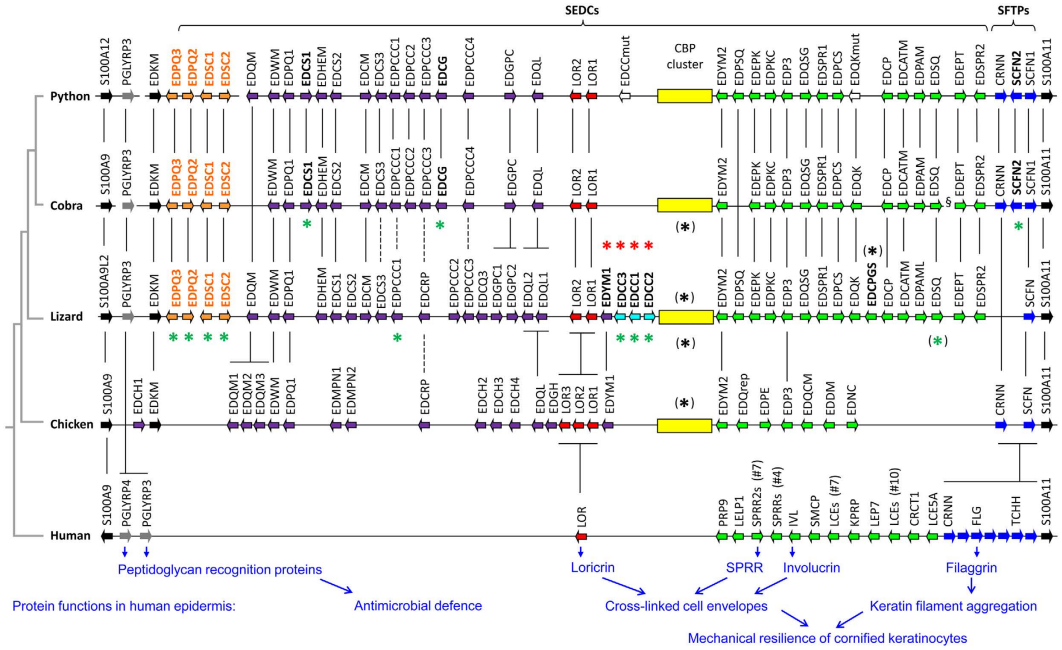


Figure 1. Organization of the epidermal differentiation complex (EDC) in the Burmese python and the king cobra. Genes of the EDC in snakes *Python bivittatus* and *Ophiophagus hannah*, in comparison to those of the lizard (*Anolis carolinensis*), the chicken, and human, are schematically depicted. Arrows indicate the orientation of the genes. Simple EDC (SEDC) genes with 2 exons are represented by colored arrows with a black frame whereas other genes are shown as filled arrows. Corneous beta-protein (CBP) gene clusters are shown as boxes in this diagram while detailed information about the genes in these clusters are depicted in Fig. 2. Members of gene families are numbered according to the positions of genes without indicating 1:1 orthology to specific members of the same gene family in other species. The depiction of the human EDC is simplified by representing gene family clusters with arrows and indicating the total number (#) of genes within each cluster. Black vertical lines connect orthologous genes or gene families. Green and red asterisks indicate putative gene gain and loss events whereas black asterisks indicate gene differences that could not be unambiguously assigned to an evolutionary event in a particular lineage. Note that the diagram is not drawn to scale. The symbol § marks a locus in which genes unrelated to classical EDC genes are present in the current genome sequence assembly of the cobra. Because of improved delineation of orthology relationships, the following gene names have been newly assigned to replace previous names[30]: lizard EDQL instead of EDCQ, chicken EDQM3 instead of EDSC, and EDPQ1 instead of EDCH5.

present in snakes, lizards, and geckos (Fig. 3), but not in other sauropsids, these proteins represent an evolutionary innovation of squamates.

Interestingly, several genes within the CBP locus encode proteins that lack a beta-sheet-forming domain, i.e. the defining feature of CBPs, but share the exon-intron organization with CBP and other SEDC genes (Fig. 2, Suppl. Fig. S1 and S2). The positions of these non-CBP genes relative to specific *CBP*s are largely conserved between snakes and the green anole lizard (Fig. 2). BLAST searches with the sequences of newly identified EDC genes allowed us to identify previously uncharacterized homologs of these genes in the green anole lizard (Suppl. Table S6; Suppl. Fig. S6).

The EDC of squamates contains multiple genes that encode proteins with extremely high cysteine contents. The amino acid sequences of snake EDC proteins were analyzed for the presence of conserved sequence motifs. Snake S100A genes and SFTPs contain an amino-terminal S100 domain[43] while snake PGLYRP3 is predicted to acquire the characteristic structural fold also found in other PGLYRPs[43]. Among SEDC proteins, only CBPs contain a structural motif, *i.e.* the beta-sheet forming region, whereas other SEDCs of snakes do not contain sequences associated with the propensity to fold into a known protein domain. However, many snake SEDC proteins have amino-terminal and carboxy-terminal sequence motifs that are conserved in SEDCs of a diverse range of amniotes, including humans (Suppl. Fig. S7).

SEDC proteins of amniotes are generally characterized by amino acid sequence repeats and high abundance of one or several of the following amino acid residues: glycine (G), serine (S), proline (P), glutamine (Q), and cysteine (C)[30,33]. Likewise, the EDC of the python comprises genes that encode proteins with high contents of G (e.g., loricrins 1 and 2, > 34% G), S (EDCS3, 36% S; EDPS1, 33% S), P (EDPQ2, 34% P; EDPCK, 38% P) and Q (EDPQ2, 24% Q) (Fig. 4 and Suppl. Fig. S1). Cysteine-rich amino acid sequences are encoded by many SEDC genes of the snakes, which is surprising in the context of the hypothesis that cysteine-rich proteins function mainly in hard and resistant skin appendages such as claws, hair and feathers, and tend to be lost in species, such as snakes, that lack these skin appendages[28].

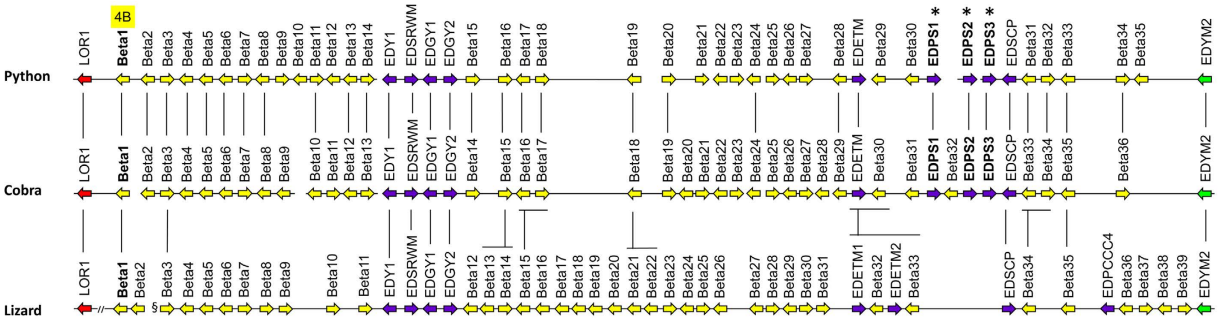


Figure 2. Organization of the corneous beta-protein (CBP), also known as beta-keratin, gene cluster in snakes. The CBP gene clusters in python (*P. bivittatus*), cobra (*O. hannah*) and lizard (*A. carolinensis*) are schematically depicted. The CBP genes were tentatively named "Beta" followed by a number that indicates the position in the cluster. The CBP genes of the lizard correspond to the "Li-Ac" genes reported previously[26], whereby Beta1 and 2 are identical to Li-Ac40 and 39, respectively, and Beta3 through 39 are identical to Li-Ac37 through 1. A gap in the lizard genome sequence assembly (§) is the likely locus of Beta40 (Li-Ac38). Arrows indicate the orientation of the genes. CBP genes are represented by yellow arrows whereas non-CBP genes inside the cluster are shown by violet arrows. The label "4B" marks the presence of 4 beta-sheets in the encoded protein (Fig. 3) and asterisks indicate snake-specific genes (Suppl. Fig. S8). *LOR1* and *EDYM2* are conserved genes flanking the CBP gene cluster. Vertical lines indicate orthologs. Note that the schemes are not drawn to scale.

Fifteen EDC genes of the python encode proteins containing at least 20% cysteine, and among them 5 proteins have a cysteine content above 35% (Fig. 4B). Five python EDC genes encoding proteins with high cysteine content were clustered between *EDPQ1* and *EDQL*, i.e. two genes evolutionarily conserved in lizard and chicken (Fig. 1). This region of the python EDC is syntenic with the locus of a gene encoding the cysteine-rich feather protein (EDCRP) in the chicken[32], while the homologous locus of the green anole lizard contains a gene encoding an EDCRP-like protein[32] and 9 other cysteine-rich proteins (EDCS1, EDCS3, EDPCCC1-4, EDGPC1-2 and EDCQ3). Among the latter, 8 were identified in the present study (Fig. 1; Suppl. Fig. S6). The highest cysteine contents so far detected among squamate EDC proteins are present in *A. carolinensis* EDPCCC4 (45% cysteine) and EDPCCC1 of the king cobra (46.6% cysteine). EDPCCC1 is conserved among snakes, and by the analysis of tissue transcriptomes of *E. coloratus* we could confirm expression in the skin whereas internal organs lacked EDPCCC1 (Suppl. Fig. S4C). The cysteine-rich proteins of snakes and other squamates are characterized by repetitive amino acid sequences with clusters of two or more cysteine residues (Fig. 4B). These proteins are candidates to become cross-linked components of the hard scales of squamates but this hypothesis remains to be tested in future studies.

Identification of snake-specific EDC gene innovations and losses. Differences in the EDCs of snakes and the anole lizard (Fig. 1) suggest that gene innovations or gene losses have occurred in either one of the evolutionary lineages leading to snakes and iguanians (represented here by *A. carolinensis*). To determine the ancestral condition for each of the clade-specific genes, we searched for orthologous genes in the Japanese gecko (*G. japonicus*), representing Gekkota, a basal clade of squamates and the closest outgroup to Toxicofera[44–46] with a sequenced genome[39]. These comparisons suggested that some EDC genes have been lost in snakes whereas others have originated in the snake lineage.

Genes of the EDCC family (Fig. 5A), which encode proteins rich in cysteine-cysteine motifs, are present in the Japanese gecko (n = 6) and in the green anole lizard (n = 3) but not in snakes (a sister group of the green anole lizard). This species distribution of EDCCs suggests that EDCC gene(s) were present in the last common ancestor of snakes and lizards and later underwent inactivation in snakes. In agreement with this hypothesis, we identified a mutated remnant of an EDCC gene located between *loricrin 1* and the CBP gene cluster of the python (Fig. 5B). Orthologs of EDCC are absent from the EDCs of birds and turtles[30,33], indicating that EDCCs represent a squamate-specific gene innovation. The gene *EDYM1*, which is conserved in turtles and birds[33], is located between *EDCC* and *loricrin* genes in the anole lizard and the gecko whereas it is absent in snakes (Fig. 1). These data suggest that *EDYM1* and the *EDCC* genes of the last common ancestor of Toxicofera (snakes and iguanid lizards) were lost in snakes.

A group of apparently snake-specific proteins are encoded by the genes *EDPS1* through *EDPS3* (*epidermal differentiation proteins rich in proline and serine 1-3*) (Suppl. Fig. S8), which are located within the CBP gene cluster of snakes but not in the CBP cluster of other squamates (Fig. 2). The investigation of tissue transcriptomes of *E. coloratus* suggested that EDPS homologs are expressed in the skin but not in internal organs (Suppl. Fig. S4E).

Duplication of scaffoldin in snakes. Snakes have 3 SFTP genes (*Crnn*, *Scfn1*, *Scfn2*) (Fig. 6) whereas the green anole lizard has only one (*Scfn*) while the bearded dragon, the Japanese gecko, the American alligator and the chicken have two (*Crnn*, *Scfn*)[30,31]. Both isoforms of snake scaffoldin proteins are rich in glutamic acid (E) and arginine (R) residues (Fig. 6A), which are also highly abundant in human trichohyalin but not in cornulin[31]. Expression of both *Scfn* genes was confirmed by intron-spanning RNA-seq reads in the painted saw-scaled viper (*Echis coloratus*) (Suppl. Fig. S9).

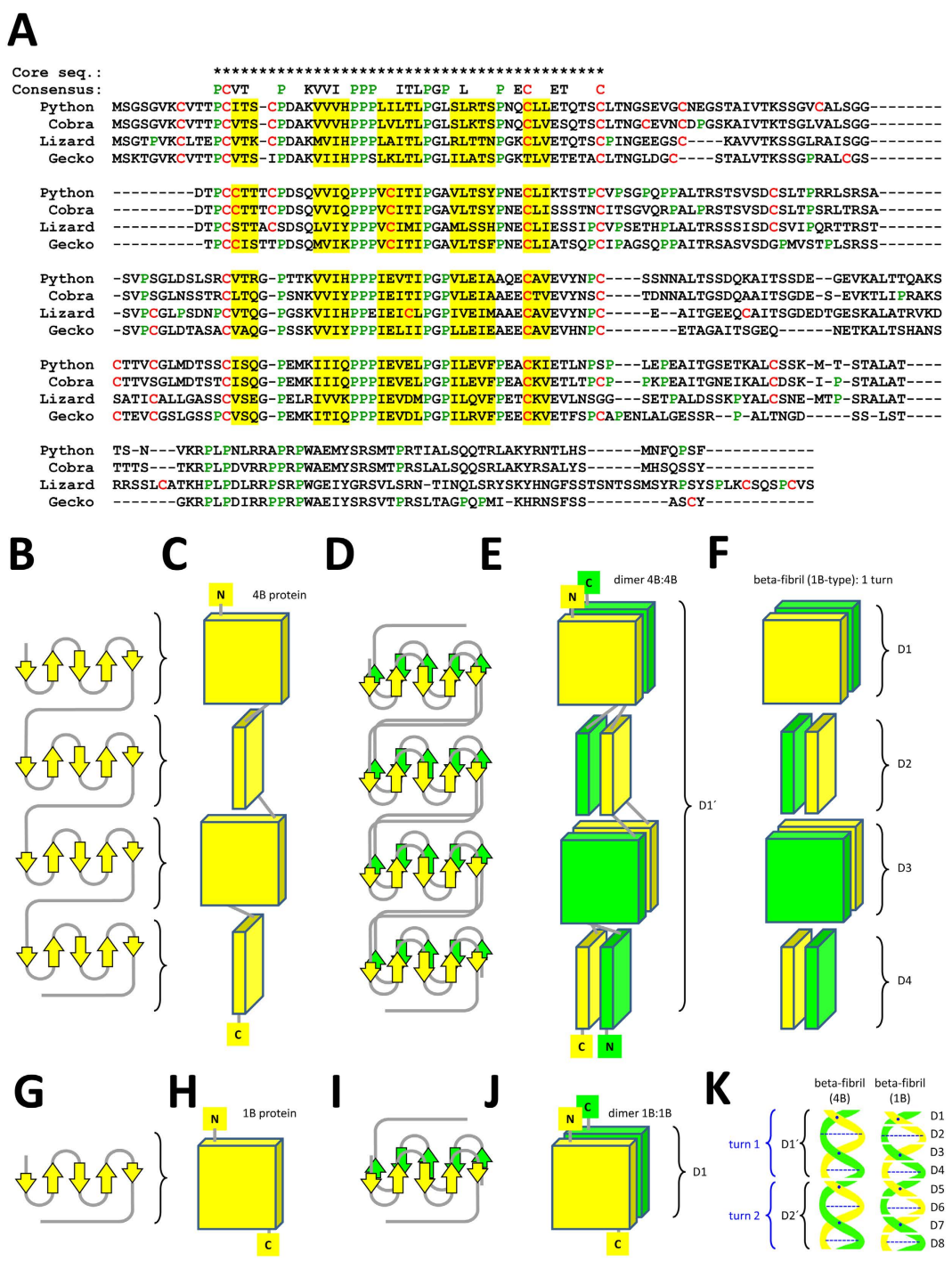


Figure 3. A corneous beta-protein with 4 beta-core sequence motifs is predicted to facilitate the formation of unique beta-fibrils in squamates. (**A**) Amino acid sequence alignment of Beta1 proteins of python (*P. bivittatus*), cobra (*O. hannah*), lizard (*A. carolinensis*) and gecko (*G. japonicus*). Internal sequence repeats were aligned to illustrate the conservation of segments corresponding to the CBP core sequences (indicated by * above the alignment) (see also Suppl. Fig. S5). Key residues of the repeat consensus sequence are indicated. Putative beta-strand-forming residues are indicated by yellow shading. Proline and cysteine residues are shown in green and red fonts, respectively. (**B–K**) Modeling of protein structures. The model for the folding (**B**,**C**) and dimerization (**D**,**E**) of CBPs with 4 beta-sheets (B4 proteins) was adapted from the model for CBPs with 1 beta-sheet (B1 proteins) (**G–J**) which also proposes a helical arrangement of 4 CBP dimers in 1 turn of the helical structure of a beta-fibril[41] (**F**). The model for B4 proteins suggests that a dimer (D1′) (**E**) is equivalent to 1 turn of a beta-fibril (**F**) and 2 B4-protein dimers (D1′ and D2′) substitute for 8 B1-protein dimers to form 2 turns of a beta-fibril helix (**K**). The orientation of the dimerization interfaces is indicated by dashed lines and dots in (**K**). Note that the schematic depiction of beta-sheets is simplified and does not show a twist typically present in beta-sheets. C, carboxy-terminus; N, amino-terminus.

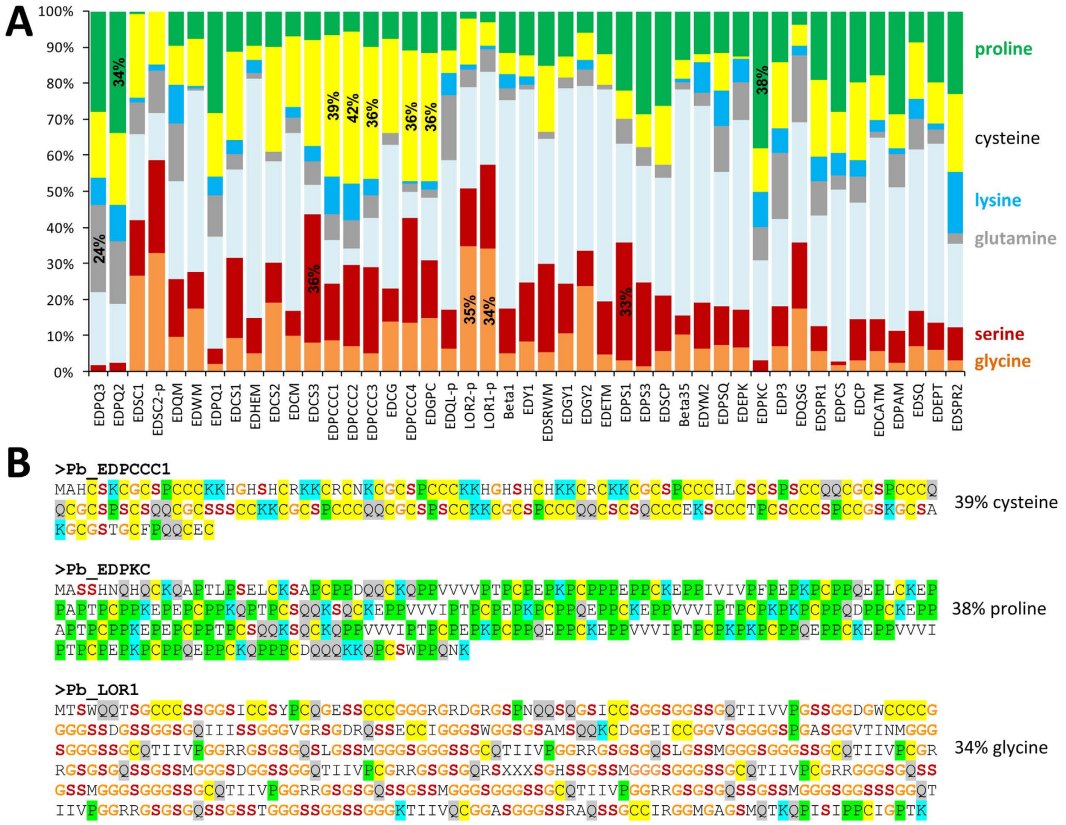


Figure 4. SEDC genes encode proteins with extremely biased amino acid composition. (**A**) The diagram shows the amino acid compositions of SEDC proteins of *Python bivittatus*. The protein data are shown in the order of the corresponding genes in the EDC (Fig. 1). Translation products of genes within the corneous beta-protein (CBP) gene cluster (Fig. 2) are not included here, with the exception of proteins encoded by the first and the last CBP. (**B**) Amino acid sequences of exemplary SEDC proteins of the *P. bivittatus*.

Phylogenetic profiling and gene locus comparison suggested that the *Scfn2* gene originated by duplication of the primordial *Scfn* gene specifically in snakes (Fig. 6B). The sequences of the proximal promoters of *Scfn1* and *Scfn2* genes were partly conserved, and they contained homologous TATA boxes. Remarkably, the putative binding sites for 2 transcription factors (KLF4 and AP-1) in promoters of SFTP genes[31] were differentially conserved in snake *Scfn1* and *Scfn2* (Fig. 6B, Suppl. Fig. S10). The predicted binding site for AP-1 was present in the promoters of *Scfn1* but not of *Scfn2* genes whereas the KLF4 binding site was present in the promoters of *Scfn2* but not of *Scfn1* genes of both python and cobra. Taken together, these data suggest a scenario in which a single ancestral *Scfn* gene was present in the last common ancestor of snakes and the green anole lizard, this gene was duplicated in primitive snakes, and the derived genes underwent divergent evolution of their promoter sequences.

Moreover, the SCFN2 proteins lack a carboxy-terminal sequence motif (CTM) that is present in SCFN1 of snakes and in most other SFTPs of amniotes[31,47] (Fig. 6). As this motif has been implicated in keratin filament binding of SFTPs[47], divergent amino acid sequence evolution appears to have caused also differences in the functions of scaffoldins 1 and 2 of snakes.

Discussion

The results of the present study shed new light on the molecular composition and evolution of the epidermis in snakes and other squamates. The comparative analysis of snake EDCs suggests that the epidermis of snakes contains many more proteins than the small set of CBPs (beta-keratins) identified in previous studies[25,26]. Both the number of CBPs and the number of other EDC genes of snakes is similar to those of the green anole lizard, and the total number and sequence diversification of epidermal differentiation genes (including CBPs) in snakes exceeds that present in mammals (Fig. 1). The new data therefore indicate that the process of keratinocyte cornification in the epidermis of snakes requires the participation of numerous proteins aside CBPs, like in other sauropsids. While crucial roles of mammalian EDC genes such as loricrin, LCEs, trichohyalin, and filaggrin in the skin barrier of mammals have been defined by a long series of studies[48–52], the identification of the EDC gene complement of snakes is the pivotal starting point for a comprehensive investigation of epidermal differentiation in this important subgroup of reptiles.

Important limitations of this study were the quality of the genome sequences that were available for analysis and the focus of our study on genes of the EDC. The current genome sequences of squamates are not of the same quality as those of mammalian model species and, therefore, some gaps are present in our model of the EDC in snakes (Fig. 1). The aim of this study was the characterization of the EDC in snakes; and other genome loci that control

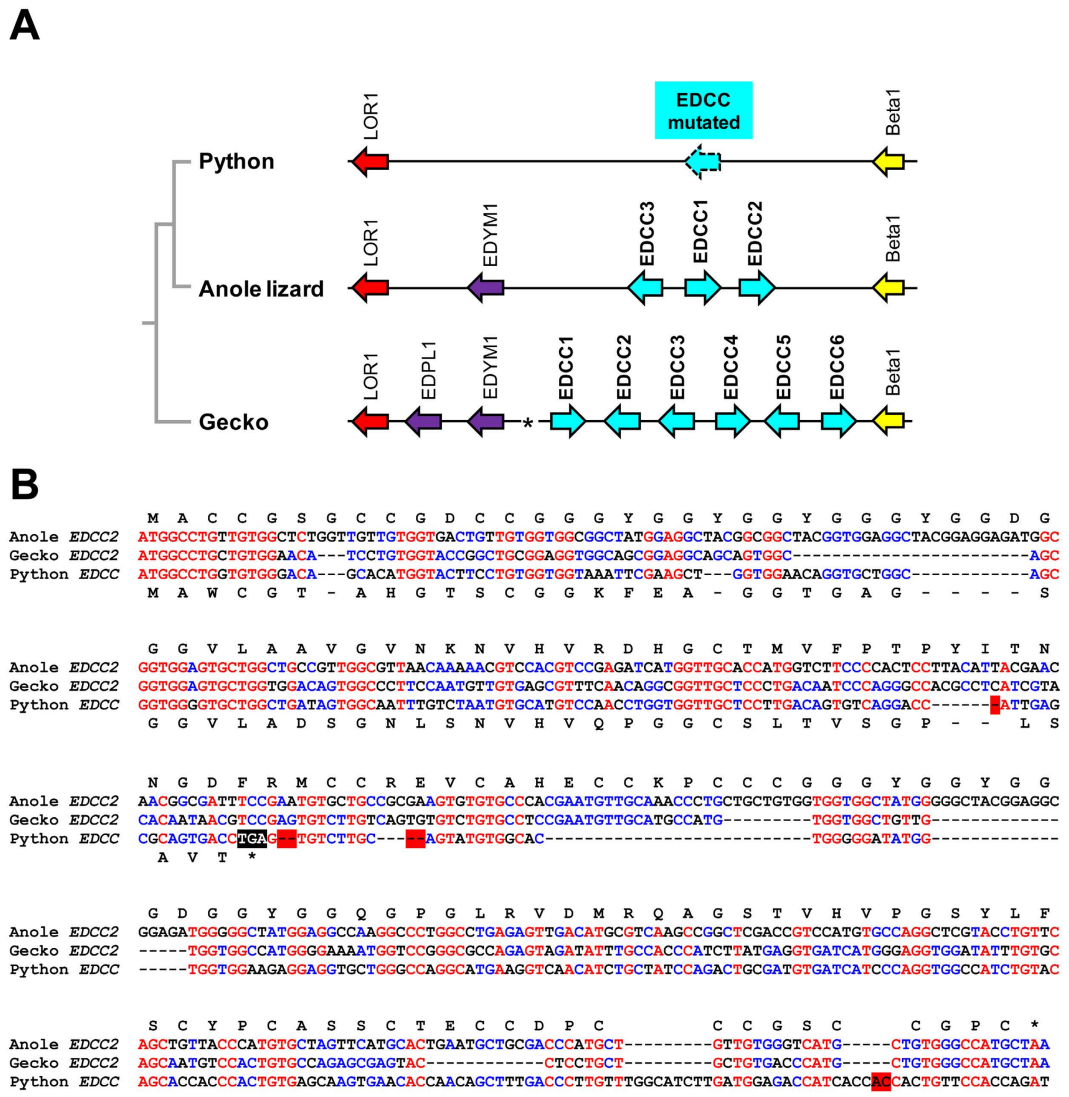


Figure 5. EDCC genes have accumulated inactivating mutations in snakes. (**A**) The gene loci of EDCC and EDYM1 genes in the green anole lizard (Ac, *Anolis carolinensis*) and the gecko (Gj, *Gekko japonicus*) were compared to that of the python (Pb, *Python bivittatus*). (**B**) Nucleotide sequence alignment of the mutated EDCC sequence of the python and EDCC genes of the anole lizard and the gecko. Red fonts indicate conservation in all 3 sequences and blue fonts indicate conservation in 2 of 3 sequences. Red shading indicates frame-shift mutations and a premature stop codon in the python genes is shown with white fonts on black background. Amino acid sequences obtained by in silico translation of anole and python genes are shown above and below the nucleotide sequence alignment, respectively.

distinct aspects of epidermal differentiation, such as the enzymatic control of protein cross-linking during keratinocyte cornification and the disruption of intercellular junctions during ecdysis have not been investigated here.

The EDC of snakes is largely syntenic with that of the green anole lizard, and only few genes are not orthologous between the two taxa. As snakes lack the specialized epidermal differentiation pathways that lead to the formation of claws and toe pad lamellae, the large degree of EDC gene conservation between snakes and the anole lizard suggests that the great majority of EDC genes play essential roles in skin structures unrelated to limb-specific appendages. Nevertheless, the present identification of several snake-specific changes in the EDC gene complement (loss of EDCCs and EDYM1, origin of EDPS, duplication of scaffoldin) points to unique characteristics of the epidermal differentiation in snakes, likely evolved during their specific adaptation to their environment. EDCCs contain multiple CC dipeptides and stretches of proline residues, respectively, and these proteins show limited similarity with other SEDC proteins, indicating that they have non-redundant roles. Likewise, the amino acid sequences of scaffoldins 1 and 2, as well as their promoters, differ substantially. Recently, we have shown that scaffoldin of the chicken is expressed in the embryonic periderm and in epithelial cells that support the morphogenesis of claws and feathers by providing a transient epithelial scaffold which degenerates after maturation of these skin appendages[31]. These data contributed to the evolutionary-developmental model that connects the embryonic archosaur scale and feathers[16–18,53]. The identification of the genes encoding

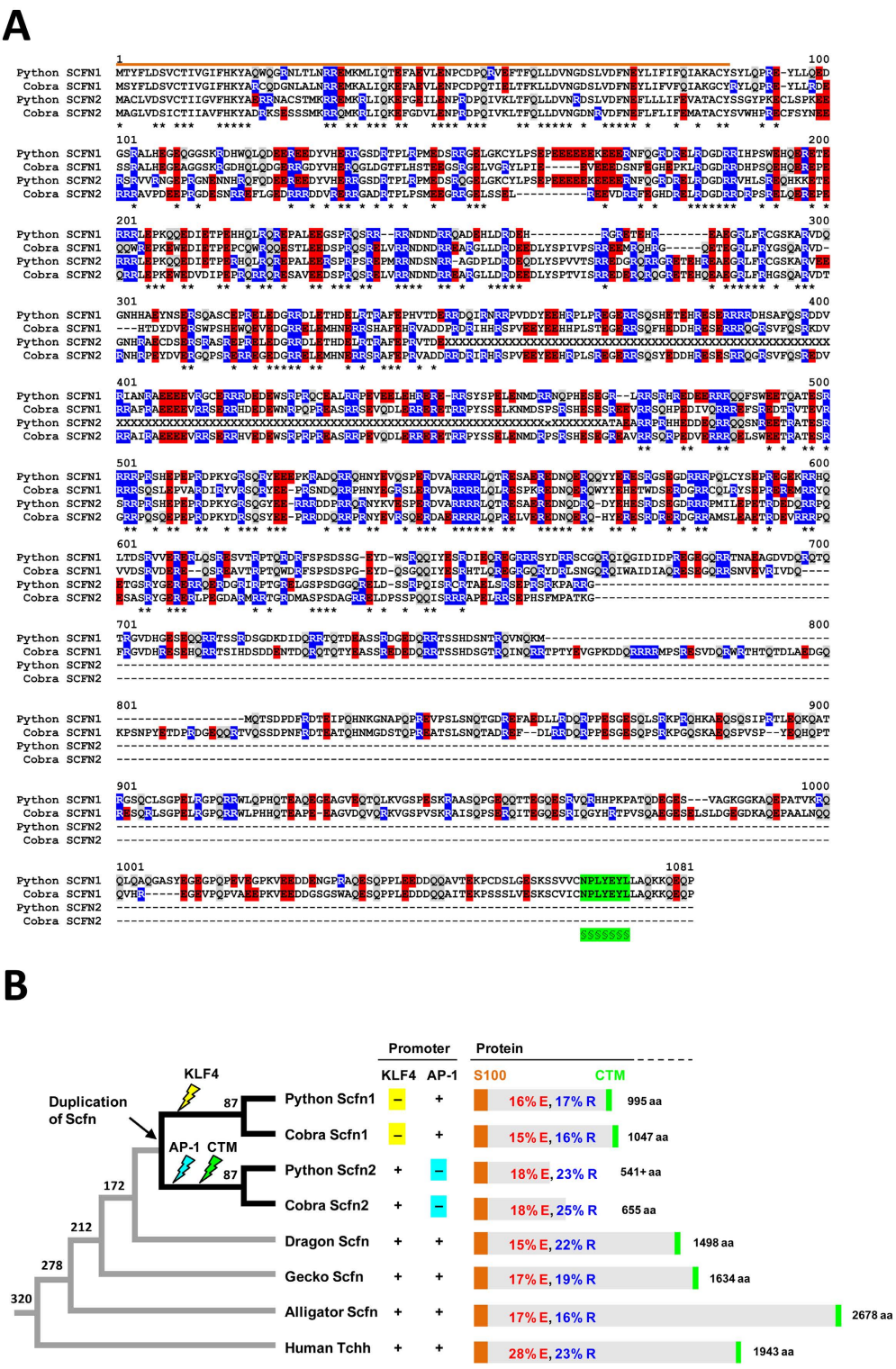


Figure 6. Duplication and sequence diversification of scaffoldin in snakes. (**A**) Amino acid sequence alignment of scaffoldin (SCFN) 1 and 2 proteins of python and cobra. The S100 domain is indicated by an orange-colored line above the alignment. A carboxy-terminal motif (CTM) (§) implicated in keratin binding is highlighted by green shading. The predominant amino acid residues, i.e. glutamic acid (E) and arginine (R), are highlighted by red and blue shading. Positions with identical amino acid residues in all 4 sequences are indicated by asterisks below the alignment. (**B**) Schematic phylogenetic tree for SFTPs in amniotes. Numbers at the branching points of the species tree indicate the divergence times (million years ago) of phylogenetic lineages. The presence of putative binding sites for the transcription factors KLF4 and AP-1 in the promoters (Suppl. Fig. S10) are indicated. The organization of the proteins is depicted schematically with orange boxes indicating the S100 domain and green boxes indicating the CTM. The contents of glutamate (E) and arginine (R) of each protein are shown. aa, amino acids; Tchh, trichohyalin.

scaffoldins 1 and 2 will facilitate the determination of their expression during embryonic development and during the shedding cycle of snakes in future studies. Likewise, the future investigation of other EDC genes, that have either been lost or acquired specifically in the snake lineages, will help to elucidate differences between the epidermis of snakes and other squamates.

The presence of multiple genes encoding cysteine-rich EDC proteins in snakes suggests that a high cysteine content of proteins is not only required for hard skin appendages such as claws, hair and feathers. These skin appendages consist of entirely cornified proteinaceous structures and are absent in snakes. Hard skin appendages utilize cysteine-dependent disulfide protein cross-linking to acquire high mechanical resilience. In mammals, cysteine-rich keratins are components of hair and nails whereas keratins with low cysteine content are components of the soft epidermis[54]. Cysteine-rich keratin-associated proteins (Krtaps) are further components of mammalian hair and nails, and a cysteine-rich EDC protein (EDCRP) is a component of avian feathers[30,32]. Our detection of multiple EDC genes for high-cysteine proteins in snakes suggests that these epidermal proteins can have functions unrelated to claws, hair and feathers. Pythons and boas have spurs (rudimentary claws) that are located next to their cloaca, but other snakes do not have homologs of claws. Thus, disulfide bond-mediated cross-linking of cysteine-rich proteins may contribute to the maturation of hard scales in snakes and probably also in other squamates. In this regard, it will be interesting to compare the expression pattern of cysteine-rich EDC proteins in different types of snake scales, i.e. flat and tough scales on the head, keeled and perhaps softer scales on the dorsum and the sides of the body, and large, mechanically resistant ventral scales (gastrosteges) that are utilized for movement. However, the presence of high cysteine contents in the absence of hard skin appendages may also point to a role of cysteine residues that is unrelated to disulfide bond formation. Cysteine residues have been identified as attachment sites for palmitic acid which allows the anchoring of proteins into membranes[55]. The process of cysteine palmitoylation has been demonstrated in mammalian skin proteins, but whether a similar process also occurs in snakes requires further investigations.

Our finding that a unique CBP (beta-keratin) comprising 4 beta-sheets is conserved among and specific for squamates indicates that this protein contributes to unique properties of epidermal keratinocytes in squamates. Previous immunolabeling studies in the green anole lizard have suggested expression of Beta1, also referred to as Li-Ac40, in the beta-layer of scales on different body sites investigated[56]. The immunolabeling for this large beta-protein was associated with filaments of 3 nm thickness, i.e. the characteristic diameter of beta-fibrils. The classical model of the beta-fibril structure was developed more than 30 years ago on the basis of X-ray diffraction studies[38,40]. According to this model, 4 CBP (then called beta-keratin) dimers form a turn of a left-handed helix with four repeating units per turn. It is now striking that the only CBP comprising more than 1 beta-sheet-forming segment contains 4 such elements, indicating that it may span exactly 1 turn of the fibril. Of note, an alternative model in which 4 CBP dimers would be arranged to form half of a helix turn has also been reported[42]. Many aspects of beta-fibril formation in squamates and sauropsids in general are still open and comparative studies on CBPs with 1 and 4 beta-sheet regions may yield valuable insights in future studies.

In conclusion, the results of this study establish a comprehensive catalog of EDC genes of snakes and, thereby, provide the basis for further studies on the molecular organization and evolution of the epidermis in snakes and other squamates.

Methods

Genome sequences and gene identification. Genome sequences from the following squamate species were used for gene predictions: Burmese python (*Python bivittatus*)[10], king cobra (*O. hannah*)[11], painted saw-scaled viper (*Echis coloratus*)[36], bearded dragon (*Pogona vitticeps*)[57] and the Japanese gecko (*Gekko japonicus*)[39]. The accession numbers of genome sequence scaffolds corresponding to the EDC are listed in Supplementary Tables S2 through S6. Coding sequences of EDC genes were predicted using a combination of the following approaches. Amino acid sequences of EDC proteins of *A. carolinensis*[30] and humans[35] were used as queries in tBLASTn searches against the nucleotide sequence between *S100A12* and *S100A11* genes of the target genome. Information about exon coverage by RNA-seq reads, available in the NCBI browser for "genomic regions, transcripts, and products", was used to identify transcribed regions in the EDC of *P. bivittatus*. The transcribed regions were translated and the resulting amino acid sequences were compared to those of known EDC proteins. The nucleotide sequence of EDC regions of apparently low gene density was translated *in silico*, and additional open reading frames of candidate EDC genes were identified according to published criteria[33]. Predictions of snake EDC genes were validated by BLAST searches in the transcriptomes of *P. regius* and *E. coloratus*[36,37].

Bioinformatic analysis of gene promoters and amino acid sequences encoded by EDC genes. For the assessment of transcription factor binding scores in the promoter sequences of SFTP genes, the JASPAR 2016 server (http://jaspar.genereg.net) was used[58]. Primary and secondary structure analyses of the proteins were performed on the PSIPRED protein structure prediction server[59] and using the software tools at the ExPASy SIB Bioinformatics Resource Portal[60].

References

1. Shine, R. Reptiles. *Curr. Biol.* **23,** R227–31 (2013).
2. Pincheira-Donoso, D., Bauer, A. M., Meiri, S. & Uetz, P. Global taxonomic diversity of living reptiles. *PLoS One.* **8:e59741** (2013).
3. Close, M. & Cundall, D. Snake lower jaw skin: extension and recovery of a hyperextensible keratinized integument. *J. Exp. Zool. A. Ecol. Genet. Physiol.* **321,** 78–97 (2014).
4. Klein, M. C. & Gorb, S. N. Epidermis architecture and material properties of the skin of four snake species. *J. R. Soc. Interface*. **9,** 3140–55 (2012).
5. Filippov, A. E. & Gorb, S. N. Modelling of the frictional behaviour of the snake skin covered by anisotropic surface nanostructures. *Sci. Rep*. **6,** 23539 (2016).

6. Landmann, L. The skin of reptiles. Epidermis and dermis. In: *Biology of the Integument. 2 Vertebrates* (ed. Bereiter-Hahn, J., Matoltsy, G., Richards, K. S.) 150–187 (Springer, 1986).
7. Maderson, P. F. A., Rabinowitz, T., Tandler, B. & Alibardi, L. Ultrastructural contributions to an understanding of the cellular mechanisms involved in lizard skin shedding with comments on the function and evolution of a unique Lepidosaurian phenomenon. *J. Morphol.* **236,** 1–24 (1998).
8. Maderson, P. F. A. Histological changes in the epidermis of snakes during the sloughing cycle. *J. Zool.* **146,** 98–113 (1965).
9. Landmann, L. Keratin formation and barrier mechanisms in the epidermis of Natrix natrix (Reptilia, Serpentes): an ultrastructural study. *J. Morphol.* **162,** 93–126 (1979).
10. Castoe, T. A. *et al.* The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. *Proc. Natl. Acad. Sci. USA* **110,** 20645–50 (2013).
11. Vonk, F. J. *et al.* The king cobra genome reveals dynamic gene evolution and adaptation in the snake venom system. *Proc. Natl. Acad. Sci. USA* **110,** 20651–6 (2013).
12. Tollis, M., Hutchins, E. D. & Kusumi, K. Reptile genomes open the frontier for comparative analysis of amniote development and regeneration. *Int. J. Dev. Biol.* **58,** 863–871 (2014).
13. Eckhart, L., Lippens, S., Tschachler, E. & Declercq, W. Cell death by cornification. *Biochim. Biophys. Acta* **1833,** 3471–3480 (2013).
14. Alibardi, L. Sauropsids cornification is based on corneous beta-proteins, a special type of keratin-associated corneous proteins of the epidermis. *J. Exp. Zool. B. Mol. Dev. Evol.* **326,** 338–351 (2016).
15. Chang, C. *et al.* Reptile scale paradigm: Evo-Devo, pattern formation and regeneration. *Int. J. Dev. Biol.* **53,** 813–826 (2009).
16. Dhouailly, D. A new scenario for the evolutionary origin of hair, feather, and avian scales. *J. Anat.* **214,** 587–606 (2009).
17. Musser, J. M., Wagner, G. P. & Prum, R. O. Nuclear β-catenin localization supports homology of feathers, avian scutate scales, and alligator scales in early development. *Evol. Dev.* **17,** 185–194 (2015).
18. Di-Poï, N. & Milinkovitch, M. C. The anatomical placode in reptile scale morphogenesis indicates shared ancestry among skin appendages in amniotes. *Sci. Adv.* **2,** e1600708 (2016).
19. Alibardi, L. Transition from embryonic to adult epidermis in reptiles occurs by the production of corneous beta-proteins. *Int. J. Dev. Biol.* **58,** 829–839 (2014).
20. Alibardi, L. Ultrastructure of the embryonic snake skin and putative role of histidine in the differentiation of the shedding complex. *J. Morphol* **251,** 149–68 (2002).
21. Alibardi, L. & Toni, M. Cytochemical, biochemical and molecular aspects of the process of keratinization in the epidermis of reptilian scales. *Prog. Histochem. Cytochem.* **40,** 73–134 (2006).
22. Alibardi, L. Review: mapping epidermal beta-protein distribution in the lizard Anolis carolinensis shows a specific localization for the formation of scales, pads, and claws. *Protoplasma.* **253,** 1405–1420 (2016).
23. Roberts, J. B. & Lillywhite, H. B. Lipid barrier to water exchange in reptile epidermis. *Science* **207,** 1077–9 (1980).
24. Tu, M. C., Lillywhite, H. B., Menon, J. G. & Menon, G. K. Postnatal ecdysis establishes the permeability barrier in snake skin: new insights into barrier lipid structures. *J. Exp. Biol.* **205,** 3019–30 (2002).
25. Dalla Valle, L., Nardi, A., Belvedere, P., Toni, M. & Alibardi, L. Beta-keratins of differentiating epidermis of snake comprise glycine-proline-serine-rich proteins with an avian-like gene organization. *Dev. Dyn.* **236,** 1939–53 (2007).
26. Dalla Valle, L., Nardi, A., Bonazza, G., Zucal, C., Emera, D. & Alibardi, L. Forty keratin-associated beta-proteins (beta-keratins) form the hard layers of scales, claws, and adhesive pads in the green anole lizard, Anolis carolinensis. *J. Exp. Zool. B. Mol. Dev. Evol.* **314,** 11–32 (2010).
27. Dalla Valle, L., Nardi, A. & Alibardi, L. Isolation of a new class of cysteine-glycine-proline-rich beta-proteins (beta-keratins) and their expression in snake epidermis. *J. Anat.* **216,** 356–67 (2010).
28. Dalla Valle, L., Benato, F., Rossi, C., Alibardi, L., Tschachler, E. & Eckhart, L. Deleterious mutations of a claw keratin in multiple taxa of reptiles. *J. Mol. Evol.* **72,** 265–73 (2011).
29. Klein, M. C. & Gorb, S. N. Ultrastructure and wear patterns of the ventral epidermis of four snake species (Squamata, Serpentes). *Zoology* (*Jena*) **117,** 295–314 (2014).
30. Strasser, B. *et al.* Evolutionary origin and diversification of epidermal barrier proteins in amniotes. *Mol. Biol. Evol.* **31,** 3194–3205 (2014).
31. Mlitz, V. *et al.* Trichohyalin-like proteins have evolutionarily conserved roles in the morphogenesis of skin appendages. *J. Invest. Dermatol.* **134,** 2685–2692 (2014).
32. Strasser, B., Mlitz, V., Hermann, H., Tschachler, E. & Eckhart, L. Convergent evolution of cysteine-rich proteins in feathers and hair. *BMC Evol. Biol.* **15,** 82 (2015).
33. Holthaus, K. B. *et al.* Comparative genomics identifies epidermal proteins associated with the evolution of the turtle shell. *Mol. Biol. Evol.* **33,** 726–37 (2016).
34. Mlitz, V., Tajamul, H., Tschachler, E. & Eckhart, L. Filaggrin has evolved from an "S100 fused-type protein" (SFTP) gene present in a common ancestor of amphibians and mammals. *Exp Dermatol.* Electronic publication ahead of print. doi: 10.1111/exd.13317.
35. Henry, J. *et al.* Update on the epidermal differentiation complex. *Front. Biosci.* **17,** 1517–32 (2012).
36. Hargreaves, A. D., Swain, M. T., Logan, D. W. & Mulley, J. F. Testing the Toxicofera: comparative transcriptomics casts doubt on the single, early evolution of the reptile venom system. *Toxicon* **92,** 140–56 (2014).
37. Hargreaves, A. D., Swain, M. T., Hegarty, M. J., Logan, D. W. & Mulley, J. F. Restriction and recruitment-gene duplication and the origin and evolution of snake venom toxins. *Genome Biol. Evol.* **6,** 2088–2095 (2014).
38. Fraser, R. D. & Parry, D. A. The molecular structure of reptilian keratin. *Int. J. Biol. Macromol.* **19,** 207–11 (1996).
39. Liu, Y. *et al.* Gekko japonicus genome reveals evolution of adhesive toe pads and tail regeneration. *Nat. Commun.* **6,** 10033 (2015).
40. Gregg, K. & Rogers, G. Feather keratins: composition, structure and biogenesis. In: *Biology of the Integument* (ed. Bereither-Hahn, J., Matoltsy, G., Sylvia-Richards, K.) 666–694 (Springer, 1986).
41. Fraser, R. D. B. & Parry, D. A. D. Filamentous structure of hard β-keratins in the epidermal appendages of birds and reptiles. *Subcell. Biochem.* **82,** 231–252 (2017).
42. Calvaresi, M., Eckhart, L. & Alibardi, L. The molecular organization of the beta-sheet region in Corneous beta-proteins (beta-keratins) of sauropsids explains its stability and polymerization into filaments. *J. Struct. Biol.* **194,** 282–91 (2016).
43. Marchler-Bauer, A. *et al.* CDD: NCBI's conserved domain database. *Nucleic Acids Res.* **43,** D222–6 (2015).
44. Wiens, J. J. *et al.* Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. *Biol. Lett.* **8,** 1043–1046 (2012).
45. Vidal, N. & Hedges, S. B. The phylogeny of squamate reptiles (lizards, snakes, and amphisbaenians) inferred from nine nuclear protein-coding genes. *C. R. Biol.* **328,** 1000–8 (2005).
46. Reeder, T. W. *et al.* Integrated analyses resolve conflicts over squamate reptile phylogeny and reveal unexpected placements for fossil taxa. *PLoS One* **10,** e0118199 (2015).
47. Strasser, B., Mlitz, V., Fischer, H., Tschachler, E. & Eckhart, L. Comparative genomics reveals conservation of filaggrin and loss of caspase-14 in dolphins. *Exp. Dermatol.* **24,** 365–369 (2015).
48. Smith, F. J. *et al.* Loss-of-function mutations in the gene encoding filaggrin cause ichthyosis vulgaris. *Nat. Genet.* **38,** 337–42 (2006).
49. Palmer, C. N. *et al.* Common loss-of-function variants of the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. *Nat. Genet.* **38,** 441–6 (2006).

50. de Cid, R. *et al.* Deletion of the late cornified envelope LCE3B and LCE3C genes as a susceptibility factor for psoriasis. *Nat. Genet.* **41,** 211–5 (2009).
51. Medland, S. E. *et al.* Common variants in the trichohyalin gene are associated with straight hair in Europeans. *Am. J. Hum. Genet.* **85,** 750–5 (2009).
52. Ishitsuka, Y. *et al.* Lce1 family members are Nrf2-target genes that are induced to compensate for the loss of loricrin. *J. Invest. Dermatol.* **136,** 1656–63 (2016).
53. Wagner, G. P. *Homology, Genes and Evolutionary Innovation* (Princeton University Press, 2014).
54. Moll, R., Divo, M. & Langbein, L. The human keratins: biology and pathology. *Histochem. Cell. Biol.* **129,** 705–733 (2008).
55. Liu, K. M. *et al.* Cyclic alopecia and abnormal epidermal cornification in Zdhhc13-deficient mice reveal the importance of palmitoylation in hair and skin differentiation. J. Invest. *Dermatol.* **135,** 2603–10 (2015).
56. Alibardi, L. Immunolocalization of large corneous beta-proteins in the green anole lizard (Anolis carolinensis) suggests that they form filaments that associate to the smaller beta-proteins in the beta-layer of the epidermis. *J. Morphol.* **276,** 1244–1257 (2015).
57. Georges, A. *et al.* High-coverage sequencing and annotated assembly of the genome of the Australian dragon lizard Pogona vitticeps. *Gigascience* **4,** 45 (2015).
58. Mathelier, A. *et al.* JASPAR 2016: a major expansion and update of the open-access database of transcription factor binding profiles. *Nucleic Acids Res.* **44,** D110–5 (2016).
59. McGuffin, L. J., Bryson, K. & Jones, D. T. The PSIPRED protein structure prediction server. *Bioinformatics*. **16,** 404–405 (2000).
60. Artimo, P. *et al.* ExPASy: SIB bioinformatics resource portal. *Nucleic Acids Res.* **40,** W597–603 (2012).

Acknowledgements

We thank Wolfgang Sipos for helpful discussions. This work was supported by the Austrian Science Fund (FWF): P23801 and P28004.

Author Contributions

L.E. conceived the study, K.B.H., B.S., and V.M. performed genome and transcriptome sequence studies, K.B.H., B.S., V.M., and L.E. analyzed the results, K.B.H., L.A., E.T., and L.E. wrote the manuscript. All authors reviewed the manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing Interests: The authors declare no competing financial interests.

How to cite this article: Holthaus, K. B. *et al.* Identification and comparative analysis of the epidermal differentiation complex in snakes. *Sci. Rep.* **7**, 45338; doi: 10.1038/srep45338 (2017).

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

2.5 Disulfide binding of corneous beta proteins in the squamate skin

2.5.1 Introduction to paper V: Disulfide-bond-mediated cross-linking of corneous beta-proteins in lepidosaurian epidermis.

In paper number V, we have investigated possible covalent disulfide binding of corneous beta proteins (CBPs), also known as beta-keratins, in the cornified layer of squamates. The genes encoding corneous beta proteins form a cluster of variable length on the epidermal differentiation complex of sauropsids. These proteins have been identified as an important component of the sauropsid epidermis and appendages. In the sauropsidian epidermis an alpha layer and beta layer are distinguished, which correlates respectively to an alpha helix X-ray pattern caused by keratin intermediate filaments (KIFs) and a beta sheet pattern given by the CBPs. Especially in the part exposed to environmental insults such as the scutes and scales the beta layer is dominant and both CBPs and disulfide bondings were found to be prominent. Disulfide bonding is a particular resistant protein-protein linkage that has been found in mammalian hair keratins and KRTAPs. The hard and resistant nature of sauropsid tegument has led to hypothesis that this kind of crosslinking could be present. Furthermore co-localization of keratin intermediate filaments and corneous beta proteins in the cornified layer made binding between these proteins a likely option.

In this paper we report evidence by immunogold labeling that CBPs partly co-localize with KIFs in differentiating and mature corneous layers (beta- and alpha-) of the squamate epidermis. Furthermore we applied reduction/alkylation and oxidation protocols to squamate skin samples prior to electrophoresis to either disrupt or enhance possible disulfide bindings present in CBPs. The results were visualized by western blotting with a CBP specific antibody. Positive results showed the appearance or disappearance of bands in both the size range of CBP monomers as well as higher MW ranges that could be in some cases compatible with KIF and CBP bonding.

Although the cross-reactivity of the CBP antibodies with more than one CBP species precludes the identification of a specific CBP responsible for the marking, we were able, by

combining the experimental approach with bioinformatics analysis of *A. carolinensis* and *P. bivittatus* CBPs, to identify likely candidate proteins for some CBP bands.

In conclusion, the following paper shows by using immunolabeling and western blotting after a reduction/alkylation or oxidation treatment that KIFs and CBPs likely form very stable chemical bonds in the cornified layers. Our pre-western blot treatments modified the association of CBPs to other proteins. This was seen by an alteration in the molecular weight band pattern for the CBP antibody. All together this supports the hypothesis of possible disulfide binding between CBPs and other proteins that in some cases could be KIFs.

2.5.2 Paper V: Disulfide-bond-mediated cross-linking of corneous beta-proteins in lepidosaurian epidermis.

Disulfide-bond-mediated cross-linking of corneous beta-proteins in lepidosaurian epidermis

K.B. Holthaus^, L. Alibardi*^

^*Department of Bigea, University of Bologna, Bologna, Italy*

**Comparative Histolab, Padua, Italy*

Running head: epidermis, keratins, beta-proteins, electrophoresis, immunological methods

Abstract

Corneous beta-proteins (CBPs), formerly indicated as beta-keratins, are major protein components of the epidermis in lepidosaurian reptiles and are largely responsible for their material properties. These proteins have been suggested to form filaments of 3.4 nm in thickness and to interact with themselves or with other proteins, including Intermediate Filament keratins (IF-keratins). Here, we performed immunocytochemical labelings of CBPs in the epidermis of different lizards and snakes, and investigated by immunoblotting analysis whether the reduction of disulfide bonds or protein oxidation affects the solubility and mobility of CBPs. Immunogold labeling suggested that CBPs partly co-localize with IF-keratins in differentiating and mature beta-cells. The chemical reduction of epidermal proteins from lizard and snake epidermis increased the abundance of CBP-immunoreactive bands in the size range of CBP monomers on western blots. Conversely, *in vitro* oxidation of epidermal proteins reduced the abundance of putative CBP monomers. Some modifications in the IF-keratin range were also noted. These results strongly indicate that CBPs associate with IF-keratins and other proteins via disulfide bonds in the epidermis of lizards and snakes, which likely contributes to the resilience of the cornified beta- and alpha-layers of lepidosaurian epidermis during life and after shedding.

Key words: lepidosaurian epidermis, alpha-keratins (IF-keratins), corneous beta-proteins, immunoblotting immunocytochemistry

Introduction

The epidermis of reptiles contains intermediate filaments (alpha)-keratins (IF-keratins) that associated to specialized corneous proteins (CPs) and give rise to corneous layers of different texture indicated as beta- and alpha-layers (Baden and Maderson, 1970; Maderson et al., 1998; Alibardi and Toni, 2006). The beta-layer forms the thick corneous layer of the shell (carapace, bridge and plastron) in turtles and in crocodilian scutes. In lizards, snakes, amphisbenids, and sfenodontids (lepidosaurians), the beta-layer is variably thin and pliable but still constitutes the external and hardest layer of scales. The thickness of the beta-layer determines the stiffness and resistance to mechanical insults and wearing of scales in these reptiles.

Most of the proteins associated to IF-keratins, formerly indicated as beta-keratins, constitute a group of small proteins of 10-24 kDa produced and packed in the hard corneous layers of scales indicated as oberhautchen and beta-layers (Baden and Maderson, 1970; Wyld and Brush, 1979, 1983; Alibardi and Toni, 2006; Alibardi et al., 2009). Beta-keratins are very different from IF-keratins and are considered Corneous Beta-Proteins (CBPs, Alibardi et al. 2009; Calvaresi et al., 2016) containing a 34 amino acid long central region conformed in 4-5 anti-parallel beta-pleated sheets that produce filaments of 3.4 nm in thickness (Fraser and Parry, 1996, 2011). These small CBPs likely associate to IF-keratins during the differentiation of the oberhautchen and beta-cells in the epidermis but other CBPs have been also found in the alpha-layer indicating that beta-layer contains a higher proportions of CBPs and lower quantity of IF-keratins while the alpha-layer contains a lower proportion of CBPs and a higher proportion of IF-keratins (Alibardi et al., 2012; Alibardi, 2013, 2015).

Differently from IF-keratins, CBPs genes are encoded in the Epidermal Differentiation Complex (EDC), a locus where other genes coding for numerous types of corneous proteins, such as loricrin, cornulin, and trichohyalin-like etc., are present (Vanhouttegem et al. 2008; Strasser et al., 2014; Holthaus et al., 2016, 2017). Differently from CBPs, the other corneous proteins of the EDC, mainly present in the alpha-layers of lepidoaurian epidermis (Alibardi and Toni, 2004; Mlitz et al., 2014; Strasser et al., 2015; Alibardi 2016), do not possess a central beta-sheet region. It is believed that during the formation of the beta-layer, CBPs are deposited over an IF-keratin meshwork. Proofs for this association derives from the isolation of alpha-keratins and CBPs from the beta-layer of various reptiles (Wyld and Brush, 1979, 1983; Alibardi and Toni, 2006; Toni et al., 2007), from the co-localization of alpha-keratins (IF-keratins) and CBPs in the beta-layer of numerous reptiles by immunolabeling (Alibardi, 2013, 2015, 2016), and from detailed X-ray diffraction studies that showed some IF-keratin components mixed with the prevalent CBPs (beta-keratins) components in

the beta-layers of snakes (Ripamonti et al., 2009). Also the gradual degree of mechanical properties, from hard and inflexible to soft and stretchable, detected by micro-nanoindentation and chemical analysis in snake epidermis suggests that the tougher layers (oberhautchen and beta-layer) contain prevalent CBPs (beta-keratins) that decrease and are replaced by IF-keratins and lipids in the inner (alpha) layers (Klein et al., 2010; Klein and Gorb, 2012; Torri et al., 2014).

Standard biochemical methods for separating epidermal proteins and showing interaction are generally not applicable to epidermal proteins present in the cornified layer due to the highly cross-linked resilient nature of this structure. Despite treatments proteins often remain bound forming insoluble complexes and require harsh conditions for separation that can negatively influence further sample processing. In this study we applied additional treatments and attempted an alternative approach to this problem.

Anomalies of western blot-labeling for CBPs (beta-keratins), producing some labeled bands above the beta-keratin range were previously noted, and were interpreted as due to the formation of polymers or to the presence of strong and stable associations between IF-keratins and CBPs, that were not cleaved in conventional preparations for electrophoresis, giving rise to immunolabeled bands outside the expected molecular weight (MW, Alibardi and Toni, 2006; Toni et al., 2007). In order to definitely provide a further proof on the existence of direct interactions between IF-keratins and CBPs responsible for the different mechanical properties of the beta- and alpha-layers in lepidosaurian epidermis, we have utilized electrophoretic separation of epidermal proteins extracted under different chemical conditions to determine some cleavage and separation as well as bonding between the two types of proteins.

Materials and methods

Tissues collection and embedding

The samples were collected, fixed and embedded as indicated in previous studies (Toni et al., 2007; Alibardi, 2013, 2014, 2015), following the Italian Guidelines for animal care and handling (art. 5, DL 116/92). In the present study we have utilized fresh epidermis from the lizards *Podarcis sicula* and *Tarentola mauritanica*, whole fresh skin from *Anolis carolinensis* and molts from *Pogona vitticeps*, while for the snakes other molts derived from *Morelia bredli* and *Agkistrodon contortrix* and fresh skin from *Python bivittatus, Liasis fuscus,* and *Natrix natrix*. For immunocytochemical detection, 2 x 3 mm large samples were immediately fixed at 0-4°C for 5-8 hours in 4% paraformaldehyde in 0.1 M phosphate buffer at pH 7.4 (Sigma, St Louis, MI, USA), dehydrated in ethanol and infiltrated in

Bioacryl resin for 3-4 hours at 0-4°C. This resin was made following the indications reported in Scala et al. (1992), and the infiltrated tissues were finally embedded in pure Bioacryl resin in gelatin capsules for polymerization under UV light at 0-4°C for 3 days. Using an ultramicrotome (Ultrotome III, LKB, Bromma, Sweden), sections of 2-4 μm were collected on glass slides, and dried for the histological examination. Sections were stained on a hot plate using a 1% Toluidine blue solution. Other sections were instead collected over chromoallume-gelatin-coated slides over a hot plate at 40-45 °C, dried and later utilized for immunocytochemical detection.

Immunocytochemistry

The semi thin sections were pre-incubated for 30 minutes at room temperature with 2% BSA in 0.05 M Tris/HCl buffer at pH 7.6 containing 3% normal goat serum. The plastic sections were incubated for 8 hours at room temperature in the primary antibody (rabbit Pre-core box antibody (Alibardi, 2015), rabbit IF-keratin AK2 antibody (Alibardi, 2013)) diluted in the Tris buffer (1:100). In controls, the primary antibody was omitted. Sections were rinsed in buffer and incubated with secondary anti-rabbit FITC-conjugated antibody (Fluorescein Isothiocyanate, Sigma, dilution 1: 100), and were observed using a fluorescence microscope (Euromex, The Netherlands) equipped with a fluorescein filter.

Immunogold labeling was detected on thin sections of the epidermis collected with an ultramicrotome on Nickel grids of 200 mesh. The sections were incubated for 5 hours at room temperature with the above primary antibodies (dilution 1: 50-100 in buffer), and rinsed in the buffer. Some sections were double-immunolabeled with a beta-protein antibody raised in Goat (Beta-keratin G30 antibody, a gift form Dr. RH Sawyer, University of South Carolina, USA) and with the AK2 antibody raised in rabbit against IF-keratins. In control sections, the primary antibodies were omitted in the incubation step. Secondary antibodies (anti-rabbit 5-10 nm Gold Conjugated IgG, Sigma, USA, anti-Goat 20 nm Gold conjugated) were diluted in buffer (1: 80), and the sections were incubated for 1 hour at room temperature. The sections were finally stained for 4 minutes at room temperature with 1% uranyl acetate, rinsed in distilled water, and dried. The grids were studied using a 10C/CR Zeiss transmission electron microscope operating at 60 kV.

Protein extraction and reductive or oxidative treatments

Fresh skin, epidermis and molt samples from various Lepidosaur species were collected and homogenized in a solubilization buffer (modified from Sybert et al., 1985) with 8 M urea (Sigma, Steinheim, Germany), 50 mM Tris-HCl (pH 7.6, Sigma, Steinheim, Germany), 0.1 M 2-

mercaptoethanol (Sigma, USA), 1 mM DTT (Biorad, Hercules, USA) and protease inhibitor (Sigma, Steinheim, Germany). The particulate material was removed by centrifugation at 10,000 × g for 10 minutes. Using the Bradford protein assay (Biorad, München, Germany) with bovine serum albumin as the standard the sample protein concentration was determined. The obtained solutions underwent different treatments before performing electrophoresis aiming to further reduce the disulfide bonds or to oxidize them in order to detect variation in the electrophoretic pattern with the control solution without any treatment.

In the reduction treatment, samples were treated with a reducing and alkylation protocol normally used for two-dimensional blotting to enhance the breaking of disulfide bonds. The protocol consisted in adding first a reduction and alkylation buffer (pH 9) then a reducing solution with DTT (final concentration 50 Mm) and leaving the samples in agitation for 1 hour at room temperature. These first steps were followed by an incubation in the dark in agitation for 1 hour at room temperature with an alkylation solution containing Iodioacetamide (final concentration 50 mM, GE Healthcare, Buckinghamshire, UK). After this pre-treatment Laemmli buffer was added as usual and samples were denatured for 5 minutes at 100 °C. Samples were generally loaded with 40 µg of protein and separated on a SDS-PAGE (Sodium Dodedecyl Sulfate-polyacrylamide gel) at 12 or 15 % using a Biorad apparatus. Gels at 15% have only been used to check for the presence of bands below 10 kDa. As markers the Biorad (Hercules,USA) wide range MW (10–250 kDa) markers were used. A sample, which had not undergone the reduction and alkylation treatment, was used as a control. For its well-known reducing capacity tributylphosphine (TBP) was tested as an alternative to DTT, but results did not yield more signs of reduction (not shown). We experimented the reduction and alkylation agents both with longer exposure times and higher concentrations, but this has not lead to better results.

In the oxidative treatment, we instead attempted to enhance the disulfide binding of corneous beta proteins by adding an oxidation agent (H_2O_2, Sella, Schio, Italy) to the solubilization buffer. Samples were incubated for 5 minutes with two concentrations, one more physiological (10 µM) and another using a high concentration (1 mM, data not shown) according to Cumming et al. (2004). In the case of the oxidation protocol a non-reducing Laemmli buffer was added before loading and samples were not denatured by boiling the solution. Electrophoresis was performed as usual and a sample, which had not undergone oxidation and was loaded with the standard Laemmli buffer was used as a control.

Electrophoresis and western blotting

After electrophoresis in acrylamide gels at 12 or 15%, proteins were transferred on a nitrocellulose membrane. For the immunoblotting the following previous designed corneous beta protein specific

antibodies were used: HgGC10 (Dalla Valle et al., 2012) and Pre-core box (PCB, Alibardi, 2015). Also two IF-(alpha-) keratin specific antibodies were used: AK2 and α1 (Alibardi, 2013 and 2014 respectively). The epitope of these antibodies and their characteristics are indicated in Table 1. Membranes were stained with Ponceau red to check for protein transfer before incubation with the primary antibody (Ab) at 1:500 dilution. For incubation with the secondary Ab usually a fluorescent anti-rabbit IgG antibody (ECL plex goat-o-rabbit IgG CYTM5, GE Healthcare, Buckinghamshire, UK) was utilized, but in case of a second blotting on the same membrane also a CY3 conjugated antibody of the same supplier was used. When using CY3 as a secondary Ab staining with Ponceau red was done after acquisition to avoid a false positive signal in the red range. Bands were detected using the Biorad external laser Molecular Imager FX combined with the program Pharos FX. After performing CBP protein and IF-keratin blotting on the same membrane, the acquisition was done in different fluorescent canals (CY3 and CY5). Controls omitting the primary Ab to check for auto immunofluorescence of the samples were done for all samples.

Bioinformatics analysis

The lizard *A. carolinensis* corneous beta protein (CBP) sequences (Dalla Valle et al. 2010), have been updated according to the latest database version and orientation and numbering of this cluster was inverted to comply with other squamate studies (Holthaus et al., 2017, Suppl. Table S1).

CBP sequences of the python snake *P. bivittatus* were recently identified (Holthaus et al., 2017), based on the published genome of this specie (Castoe et al., 2013). In order to identify the keratin intermediate filaments (IF-keratin), sequences of *A. carolinensis* the ncbi protein database was consulted. Alignments of antibody epitopes with the database scaffolds of *A. carolinensis* and *P. bivittatus* were performed with the tool Tblastn at https://blast.ncbi.nlm.nih.gov. For aligning protein sequences the server MultAlin (Corpet, 1988) was used. ExPASy (Artimo et al., 2012) is another portal that was used for both calculating molecular weights (ProtParam tool) and verifying the epitope compatibility (LALIGN).

Results

Histology, light and ultrastructural immunocytochemistry

The epidermis of lizards and snakes during most of the shedding cycle is formed by a basal layer with one during resting phase or few layers during different periods of renewal phase, when a new epidermal generation if produced before molting (Fig. 1 A, B). Detailed explanation of the structure

and immunoreactivity of the different layers of snakes and lizards have been extensively reported (Alibardi and Toni, 2006; Alibardi, 2013, 2014, 2015), and only essential features of the alpha- and beta-layer are here reported. Supra-basal layers consist in new alpha-cells (Fig. 1 A), or in new alpha- and fusiform and dark-stained beta-cells in the mid stage of the renewal phase (Fig. 1 B). Beta-cells eventually condense and form a stiff and poorly stained or chromophobic beta-layer. The molts derived from shedding consist in a chromophobic and stiff beta-layer with the chromophilic and softer alpha-layer, forming the shed outer epidermal generation (Fig. 1 C).

Immunolabeling for IF-keratin shows that the basal and suprabasal layer are labeled but not the outer beta-layer CBPs (Fig. 1 D) while labeling for CBP shows that both forming beta-layer (Fig. 1 E) or the mature and compact beta-layer (Fig. 1 F) are immunolabeled. The immunogold localization of CBPs shows labeling in the corneous (β) packets of fusiform beta-cells (Fig. 1 G) and in the compact beta-layer (Fig. 1 H). Finally, double immunolabeling using small and larger gold particles shows co-localization of IF-keratin and CBP in the beta-packets of differentiating beta-cells (Fig. 1 I).

General western blot observations after reduction-alkylation or oxidation treatments

Both lizards and snakes showed a protein band around 16-19 kDa that responded to the treatment also when using two different antibodies. After reduction-alkylation in the snake *M. bredlii* a band appeared around 17-18 kDa and a weakened one around 37 and 50 kDa (Fig. 2 A). Also in the lizard *P. sicula,* bands around 17-19 as well as around 21-22 kDa appeared but not in the lizards *A. carolinensis* and *P. vitticeps* (data not shown). In the snake *A. contortrix* a band around 15 kDa was intensified after the treatment (Fig. 2 B). and the band visible at the bottom of the control lane which disappears in the reduced lane (Fig. 2 B) actually gives a band around 10 kDa (15% gel data not shown). Bands around 17-18 and 22-23 kDa appeared while one approximately around 65 kDa disappeared in the lizard *T. mauritanica* (Fig. 2B).

We also used the PCB antibody, which gave weaker bands than HgGC10 as it is shown in Suppl. Fig. S1.

Specific variations of CBP bands after reduction-alkylation

Both lizards and snakes showed a protein band around 16-19 kDa that responded to the treatment also when using two different antibodies. After reduction-alkylation in the snake *M. bredlii* a band appeared around 17-18 kDa and a weakened one around 37 and 50 kDa (Fig. 2 A). Also in the lizard *P. sicula,* bands around 17-19 as well as around 21-22 kDa appeared but not in the lizards *A.*

carolinensis and *P. vitticeps* (data not shown). In the snake *A. contortrix* a band around 15 kDa was intensified after the treatment (Fig. 2 B). Bands around 17-18 and 22-23 kDa appeared while one approximately around 65 kDa disappeared in the lizard *T. mauritanica* (Fig. 2B).
We also used the PCB antibody, which gave weaker bands than HgGC10 as it is shown in Suppl. Fig. S1.

Specific variations of CBP bands after oxidation
In the lizard *P. sicula*, bands in the 50-70 range and approximately around 45 kDa increased in intensity while in another lizard *A. carolinensis* mainly one band around 16-17 kDa disappeared without an increase in other regions (Fig. 2 C). In the snake *P. bivittatus* a band around 15-16 kDa disappeared while bands increased slightly around 33-36 kDa (Fig. 2 D). Finally, in the lizard *P. vitticeps*, mainly a band around 23-24 kDa tended to disappear but no intensification of protein bands was observed in other regions (Fig. 2 D).

Peak intensity analysis after reduction-alkylation or oxidative treatment
The analysis of the band variations using the peak intensity analysis of the program Pharos Fx more precisely indicated the variation of pattern in different samples (Fig. 3). In the epidermis from the lizard *P. sicula* after reduction-alkylation four bands appeared different between the untreated control and the reduced sample (Fig. 3 A-A1). While two bands around 70 and 42-44 kDa decreased two main bands around 21-23 kDa and 17-19 kDa appeared after reduction and alkylation. In the skin samples of the snake *P. bivittatus* also bands at higher MW one around 37 kDa and two around 27-29 kDa decreased, while one to two peaks increased around 15-18 after reduction-alkylation and marking with PCB (Fig. 3 B-B1). In the epidermis of the lizard *A. carolinensis*, the oxidative treatment determined the disappearance of mainly a 16-18 kDa peak with respect to the control but no increase was seen in other protein bands (Fig. 3 C-C1).

Variations of protein bands in IF-keratin range (32-75 kDa) using IF-keratin and CBP Abs
Since it is not unusual to see marking for CBP Abs in a MW range which actually belongs to the IF-(alpha-) keratins (37-75 kDa), we also performed immunoblotting with two IF-keratin antibodies (AK2 and α1, see Table 1) followed by direct labeling for CBPs on the same membrane. In this way we tried to identify any sign of possible interaction between IF-keratins and CBPs. When we used IF-keratin specific abs, a lowered weight band could be seen in some samples after applying the reduction-alkylation protocol. The reduction in weight is estimated around 10-20 kDa, which could

fit the MW of a CBP. The second blotting with a CBP specific Ab was performed using a different fluorescent secondary Ab and acquisition canal. The results only show the main beta overlap on the alpha network, since the second blotting with a beta Ab is not the same as when performed on a "clean" membrane as shown by the beta controls (Suppl. Fig. S 2).

Bioinformatics were used to verify specificity and identify other keratins possibly reacting with the IF-keratin α1 antibody (ncbi database cytoskeletal 14-like XP_003222513.1) as was done for CBP antibodies of the lizard *A. carolinensis*. The α1 keratin should give bands at 50,7 kDa, but besides bands in this range it also gives two lower bands around 37 and 42-43 kDa (Suppl.Fig. S2 A). The latter probably belong to K17-like X1 (37.4 kDa) and X2 (46.2 kDa) (ncbi: cytoskeletal 17 isoforms XP_008111601.1 and XP_008111602.1) which have a MW compatible with the results and reasonable identity with the α1 antibody (Suppl. Table S2).

General electrophoretic and western blot observations

The employed antibodies for CBPs immunolabeled proteins after western blotting, producing bands in the beta-range (HgGC10 and Pre-Core Box, PCB) but also in the IF-keratin range (HgCG10) (Suppl. Fig. S3 A). The HgGC10 antibody showed bands at 16-18 kDa and stronger bands around 37-38, 43-45 and 50-60 kDa (Suppl. Fig. 3 A) while the pre-core box antibody showed stronger bands at 16-18 and 21-23 kDa, and weaker at around 37, 43-45 and 50-60 kDa. Control experiments in which the primary antibodies were omitted showed no labeled bands (Suppl. Fig. S1). Based on the degrees of conservation of the known epitope sequences of these antibodies in the sequences of CBPs, we identified candidate CBPs matching in size to the immunodetected bands (Suppl. Fig. S3 A, but see later Discussion).

Identification of candidate CBPs immunoreactive with the tested antibodies

Due to the high similarity among different CBPs, even a specific antibody can cross-react with other CBPs (Alibardi, 2016). In the following discussion we speculate on the possible CBPs responsible for bands that were detected in the western blots (WBs) from the green anole lizard *A. carolinensis* and the Burmese python *P. bivittatus* (see Suppl. Fig. S4). Initially the epitopes of Abs used were aligned (Tblastn) against CBP scaffolds previously identified in *A. carolinensis* and *P. bivittatus* (Dalla Valle et al, 2010; Holthaus et al, 2017). The resulting hits were screened against the list of known CBP proteins of these species. The MW of CBP candidates was predicted and in the last step confronted with the MW bands in the WBs. From this combined approach the most likely candidates for bands around 16-19 kDa were identified in *A. carolinensis* as Beta-21, -22 and -24 (Suppl. Table

2 and Suppl.Fig. S4 A). These most likely candidates are part of sub cluster IIIB (Beta21-28) of the CBP cluster present in both the green anole and the python (red box in Suppl. Fig. S4 B). The CBP cluster of python and the green anole lizard is highly syntenic and many CBPs are orthologous between the two species (Holthaus et al., 2017). In alternative, also Beta-27 and -28 or Beta-35 could be responsible for the bands at 17-19 kDa, although presumed less likely due to lower compatibility for size (Beta-27/28) and lower compatibility with the Burmese python (Beta-35) (Suppl. Table 2).

Most likely candidates in the Burmese python for the 15-17 kDa bands observed were Beta-19 and -21 (Suppl. Fig. S2 and Suppl. Table S2) belonging to the python sub cluster IIIB (Beta19-27). No good match was found for the 17-19 kDa CBP band (Suppl. Fig. S2) probably due to various gaps present in the python's database exactly where most likely candidates are positioned.

Discussion

Pitfalls in studying proteins extracted from the stratum corneum

While soluble, cytoplasmic or even structural proteins linked to cell membrane can be dissociated and solubilized to provide a reliable electrophoretic migration and immunoblotting, this is often not the case for proteins that form the corneous layers since they chemically react to each other (Gillespie, 1991; Powell and Rogers, 1994). Therefore a study on proteins that are likely altered and cross-linked as the cornification process progress, presents several interpretative problems linked to: 1) possible modification of the original epitopes and immunoreactivity after extraction, reduction or oxidation treatments, 2) formation of un-cleaving stabilized bonds that link IF and CBPs altering their electrophoretic migration and apparent MW, 3) alteration of the expected molecular weight due to the alkylation or degradation process, 4) unfolding of the protein after the reduction or the oxidation treatments so that the epitopes can become available for immune-detection. All the above possibilities can introduce bias or uncertainties for the interpretation of the results. Therefore the main take-away message of the present study on mobile protein bands is the realization that strong covalent bonds indeed occurs between IF-keratins and CBPs in corneous layers of lizards and snakes, forming the beta- and alpha-layers.

Although the appearance of bands at 15-16 kDa may derived from the increase of the MW of the HgGC10 CBPs following alkylation, it appears more likely that other CBPs with some epitope identity with that of the HgGC10 are identified using this antibody (Suppl. Fig. S4 A). Also the labeled bands at 16-18 kDa using the PCB antibody may derive from the recognition of the epitope in the electrophoretically isolated CBPs but labelling did not occur to those linked to IF-keratins (or

alpha keratins), perhaps for epitope inaccessibility in this higher MW fraction. Whether this result is due to the external epitope recognized by the HgGC10 antibody while the PCB epitope is internal to the protein (Alibardi, 2016), remains unknown.

The variations observed in our study most likely derive from the different origin of the samples, with the epidermis at different stages of the shedding cycle, although the other variables indicated above can be accounted for. It could be argued that the treatments (reductive or oxidative) causes a change in the epitope that is no longer recognized by the antibodies after the treatment and therefore some bands disappear. Although we cannot completely exclude this epitope alteration, the treatments used were not particularly aggressive and more importantly it would be hard to explain both the disappearance and the appearance of bands in the same sample (Fig.2A) if the epitope was changed by the treatment. It seems more likely that these simultaneous changes are due to the reduction of disulfide bonds in the reductive treatment that determine the release of some proteins from the initial combination. Since the antibody can recognize more than one protein (Alibardi, 2016), the effect of the treatment could have a different impact on different epitopes. However, as seen from the protein alignments (Suppl. Fig. S4A), the differences among these proteins are relatively small and do not involve redox sensitive residues like cysteine, methionine and tyrosine.

Molecular structure and interactions in the mature vertebrate corneous material

The complexity of the chemical interactions that form the mature material in the corneous layer of vertebrate epidermis and of their appendages (claws, nails, hairs, scales, feathers etc.) is poorly known (Gillespie, 1991; Powell and Rogers, 1994). It is known that two main types of covalent bonds are involved, disulfide bonds joining cysteines among keratins and other corneous proteins and isopeptide bonds linking lysine residues mainly with aspartic or arginine lateral residues present in EDC proteins (Polakowska and Goldsmith, 1991; Hashimoto et al., 2001; Kalinin et al., 2002; Eckhart et al., 2003). Disulfide bonds derive from the action of sulfhydryl oxidase on proteins containing sulfhydryl groups while isopeptide bonds derive from the catalysis of transglutaminase on proteins containing the ε amino-groups of lysine that form carbo-amide bonds with the carbossile group of aspartate or the amine-group of arginine (Fig. 4). In the hardest corneous material of hairs, claws or nails, the nature of the chemical association between IF-keratins and keratin associated proteins (KAPs or KRTAPs) is poorly known, and only general proofs on the presence of covalent bonds have been identified so far between these molecules (Fujikawa et al., 2012. Matsunaga et al., 2013). Even less known is, in feathers, scales and claws, the specific chemical bonds formed between IF-keratins and the prevalent

Feather CBPs, or other proteins of the EDC that are present (Ng et al., 2015; Wu et al., 2016; Mlitz et al., 2014; Strasser et al., 2015).

Previous immunohistochemical studies (Banjerjee and Mittal, 1978; Mittal and Sing, 1987a,b; Alibardi, 2001) indicated the almost complete disappearing of sulfhydryl groups in the mature beta-layer of snakes and lizards where instead only or mainly disulfide groups are present, while sulfhydryl groups remain in the mature alpha-layers (Fig. 4), explaining the stiffness and un-elasticity of the harder beta-layer and the pliability and stretching of the alpha-layer. Also, Sulfhydryl oxidase, the enzyme determining the formation of disulfide bonds, is mainly present in the forming beta-cells and is lower to absent in alpha-cells (Alibardi, 2015). Transglutaminase is instead present in alpha-layers but appears absent in beta-layers of sauropsid scales (Alibardi and Toni, 2002). During alpha- or beta-cornification, the chemical combination of the initially synthesized IF-keratin and CBPs gives rise to the resistant corneous material and the specific material properties that ensure enzymatic endurance and mechanical protection (beta-layer) or pliability and water impermeability (alpha-layer) of the scales in these reptiles (Maderson et al., 1998; Klein et al., 2010; Klein and Gorb, 2012). These different material properties are needed for scale growth, water conservation, mechanical protection, and shedding in lizards and snakes (Maderson et al., 1998).

In conclusion the present study, using immunolabeling and western blotting after reduction or oxidation, indicate that IF-keratins and other proteins of the EDC, in particular the CBPs of lower MW, can form very stable chemical bonds in the corneous layers, in the beta-layer but likely also in the alpha-layer at complete cornification (Fig. 4). This can explain the appearance of immunolabeled bands for CBPs at a higher MW than that in the expected range (9-18 kDa), since the epitope is present in this resistant and stable combination. As suggested from Fig. 4, during cornification in both alpha- and beta-layers, the IF-keratins bind to CBPs. However while in beta-cells CBPs are produced in high amount and form filaments and then beta-corneous packets and eventually a dense beta-layer, CBPs are less abundant in alpha-cells forming a softer alpha-layer at maturity. While disulfide bonds are likely prevalent in mature beta-cells, sulfhydril groups, disulfide bonds and iso-peptide bonds are present in alpha-cells.

Acknowledgments: The study was in part supported by a University of Bologna RFO 2014 grant and by self-support (LA). We thank Dr. E. Spisni for lab facility use, Dr. F. Borsetti (Department of Bigea) for the useful technical suggestions during the experiments, and Dr. L. Eckhart (Department of Dermatology, Medical University of Vienna, Austria) for helpful comments on the manuscript.

References

Alibardi, L., 2001. Keratohyalin-like granules in lizard epidermis: evidence from cytochemical, autoradiographic and microanalytic studies. J. Morphol. 248, 64–79.

Alibardi, L., 2013. Cornification in reptilian epidermis occurs through the deposition of keratin associated beta proteins (beta-keratins) onto a scaffold of intermediate filament keratins. J. Morphol. 274, 175–193.

Alibardi, L., 2014. Immunolocalization of alpha-keratins and associated beta-proteins in lizard epidermis shows that acidic keratins mix with basic keratin-associated beta-proteins Protoplasma 251: 827-837.

Alibardi, L., 2015. Immunolocalization of sulfhydryl oxidase in reptilian epidermis indicates that the enzyme participates mainly to the hardening process of the beta-corneous layer. Protoplasma DOI 10.1007/s00709-015-0782-9.

Alibardi, L., 2016. Review: mapping epidermal beta-proteins distribution in the lizard *Anolis carolinensis* shows a specific localization for the formation of scales, pads and claws. Protoplasma DOI 10.1007/s00709-015-0909-z.

Alibardi, L., Toni, M., 2004. Immunolocalization of transglutaminase and cornification markers proteins in the epidermis of vertebrates suggests common processess of cornifi-cation across species. J. Exp. Zool. 302, 526–549.

Alibardi, L., Toni, M., 2006. Cytochemical, biochemical and molecular aspects of the process of keratinization in the epidermis of reptilian scales. Prog. Histochem. Cytochem. 40, 73−134.

Alibardi, L., Segalla, A., Dalla Valle, L., 2012. Distribution of specific keratin associated beta-proteins (beta-keratins) in the epidermis of the lizard *Anolis carolinensis* helps to clarify the process of cornification in lepidosaurians. J. Exp. Zool 318B, 388–403.

Alibardi, L., Dalla Valle, L., Nardi, A., Toni, M., 2009. Evolution of hard proteins in the sauropsid integument in relation to the cornification of skin derivatives in amniotes. J. Anat. 214, 560–586.

Artimo, P., et al. 2012. ExPASy: SIB bioinformatics resource portal, Nucleic Acids Res, 40(W1):W597-W603.

Baden, H.P., Maderson, P.F., 1970. Morphological and biophysical identification of fibrous proteins in the amniote epidermis. J. Exp. Zool. 174, 225–232.

Banerjee, T.K., Mittal, A.K., 1978. Histochemistry of the epidermis of the Chequered water snake *Natrix piscator* (Colubridae, Squamata). J. Zool. 185, 415–435.

Calvaresi, M., Eckhart, L., Alibardi, L., 2016. The molecular organization of the beta-sheet region in corneous beta-proteins (beta-keratins) of sauropsids explains its stability and polymerization into filaments. J. Struct. Biol. 194, 282–291.

Castoe T.A. et al. 2013. The Burmese python genome reveals the molecular basis for extreme adaptation in snakes. Proc. Natl. Acad. Sci. U.S.A. 110, 20645-50.

Corpet, F., 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 16, 10881-10890.

Cumming R.C., Andon N.L., Haynes P.A., Park M., Fischer W.H., Schubert D., 2004. Protein disulfide bond formation in the cytoplasm during oxidative stress. J. Biol. Chem. 279, 21749–21758.

Dalla Valle, L., Nardi, A., Bonazza, G., Zuccal, C., Emeera, D., Alibardi, L., 2010. Forty keratin-associated β-proteins (β-keratins) form the hard layers of scales, claws and adhesive pads in the green anole lizard, *Anolis carolinensis*. J. Exp. Zool. 314B, 11-32.

Eckhart, L., Lippens, S., Tschachler, E., Declercq, W. 2013. Cell death by cornification. Biochem. Biophys. Acta 1833, 3471-3480.

Fujikawa, H., Fujimoto, A., Farooq, M., Ito, M., Shimomura, Y. 2012. Characterization of the human hair keratin-associated protein 2 (KRTAP2) gene family. J. Invest. Dermatol. 132, 1806-1813.

Fraser, R.D.B., Parry, D.A.D., 1996. The molecular structure of reptilian keratin. Int. J. Biol. Macromol. 19, 207–211.

Fraser, R.D.B., Parry, D.A.D., 2011. The structural basis of the filament-matrix texture in the avian/reptilian group of hard beta-keratins. J. Struct. Biol. 173, 391–405.

Gillespie, J.M., 1991. The structural proteins of hair: isolation, characterization and regulation of biosynthesis In: Physiology, biochemistry and molecular biology of the skin, (ed Goldsmith LA), pp. 625-659. Oxford, Oxford University Press.

Hashimoto, Y., Suga, Y., Matsuba, S., Mizoguchi, M., Takamori, K., Seitz, J., Ogawa, H., 2001. Inquiry into the role of skin sulfhydryl oxidase in epidermal disulfide bond formation: implications of the localization and regulation of skin as revealed by TPA, retinoic acid, and UVB radiation. J Investig Dermatol 117:752–754.

Holthaus, K.B., Strasser, B., Sipos, W., Schmidt, H.A., Mlitz, V., Sukseree, S., Weissenbacher, A., Tschachler, E., Alibardi, L., Eckhart, L., 2016. Comparative genomics identifies epidermal proteins associated with the evolution of the turtle shell. Mol. Biol. Evol 33, 726-737.

Holthaus, K.B., Mlitz, V., Strasser, B., Tschachler, E., Alibardi, L, Eckhart, L., 2017. Comparative genomics of the epidermal differentiation complex suggests evolutionary adaptations of snake skin. (in revision).

Klein, M.C., Deuschle, J., Gorb, S. 2010. Material properties of the skin of the Kenyan sand boa *Gongylophis colubrinus* (Squamata, Boidae). J. Comp. Physiol. A 196, 659–668.

Klein, M.C., Gorb, S.N. 2012. Epidermis architecture and material properties of the skin of four snake species. J. R. Soc. Interface. 9, 3140-55.

Maderson, P.F.A., Rabinowitz, T., Tandler, B., Alibardi, L., 1998. Ultrastructural contributions to an understanding of the cellular mechanisms involved in lizard skin shedding with comments on the function and evolution of a unique lepidosaurian phenomenon. J. Morphol. 236, 1-24.

Matsunaga, R., Abe, R., Ishii, D., Watanabe, S-I., Kiyoshi, M., Nöcker, B., Tsuchiya, M., Tsumoto, K. 2013. Bidirectional binding property of high glycine-tyrosine keratin-associated protein contributes to the mechanical strength and shape of hair. J. Struct. Biol. 183, 484-494.

Mittal, A.K., Singh, P.N., 1987a. Scale epidermis of *Natrix piscator* during its sloughing cycle: structural organization and protein histochemistry (Reptilia: Colubridae). J.Zool. 213, 545-568.

Mittal, A.K., Singh, P.N., 1987b. Hinge epidermis of *Natrix piscator* during its sloughing cycle: structural organization and protein histochemis- try. J. Zool. 213, 685–695.

Mlitz, V., Strasser, B., Jaeger, K., Hermann, M., Ghannadan, M., Buchberger, M., Alibardi, L., Tschachler, E., Eckhart, L., 2014. Trichohyalin-like proteins have evolutionarily conserved roles in the morphogenesis of skin appendages. J. Invest. Dermatol. 134, 2682–2692.

Ng C.S.,Chen C.K., Fan W.L., Wu, P.,Wu, S.M., Chen, J.J., Lai, Y.T., Mao, C.T., Lu, M.Y.J., Chen, D.R., Lin, Z.S., Yan, K.J., Sha, Y.A., Tu, T.C., Chen, C.F., Chuong, C.M., Li, W.H., 2015. Transcriptome analyses of regenerating adult feathers in chicken. BMC Genomics 16, 1-16.

Polakowska, R.R., Goldsmith, L.A., 1991. The cell envelope and transglutaminases. In: Physiology, Biochemistry and Molecular Biology of the Skin, vol 1. Goldsmith L.A. (Editor). Oxford University Press, New York-Oxford. pp. 168–201.

Powell, B., Rogers, G., 1994. Differentiation in hard keratin tissues: hair and related structure In: The keratinocyte handbook, (eds Leigh I, Lane B, Watt F), pp. 401-436. Cambridge University Press.

Scala, C., Cenacchi, G., Ferrari, C., Pasquinelli, G., Preda, P., Manara, G., 1992. A new acrylic resin formulation: a useful tool for histological, ultrastructural, and immunocytochemical investigations. J. Histochem. Cytochem. 40, 1799-1804.

Strasser, B., Mlitz, V., Hermann, M., Rice, R.H., Eigenheer, R.A., Alibardi, L., Tschachler, E., Eckhart, L., 2014. Evolutionary origin and diversification of epidermal barrier proteins in amniotes. Mol. Biol. Evol. 31, 3194–3205.

Strasser, B., Mlitz, V., Hermann, M., Tschachler, E., Eckhart, L., 2015. Convergent evolution of cysteine-rich proteins in feathers and hair. BMC Evol. Biol. 15, 82.

Sybert, V.P., Dale, B.A., Holbrook, K.A. 1985. *Ichthyosis vulgaris*: Identification of a defect in synthesis of filaggrin correlated with an absence of keratohyaline granules. J. Invest. Dermatol. 84, 191–194.

Toni, M., Dalla Valle, L., Alibardi, L., 2007. Hard (Beta-) keratins in the epidermis of reptiles: composition, sequence, and molecular organization. J. Prot. Res. 6, 3377–3392.

Torri, C., Mangoini, A., Teta, R., Fattorusso, E., Alibardi, L., Fermani, S., Bonacini. I., Gazzano, M., Burghammer, M., Fabbri, D., Falini, G., 2014. Skin lipid structure controls water permeability in snake molts. J. Struct. Biol. 185, 99-106.

Vanhoutteghem, A., Djian, P., Green, H., 2008. Ancient origin of the gene encoding involucrin, a precursor of the cross-linked envelope of epidermis and related epithelia. PNAS 105, 15481–15486.

Wu, P., Ng, C.S., Yan, J., Lai, Y.C., Chen, C.K., Lai, Y.T., Wu, S.M., Chen, J.J., Luo, W., Widelitz, R.B., Li, W.H., Chuong, C.M., 2015. Topographical mapping of α-and β-keratins on developing chicken skin integument: functional interaction and evolutionary perspectives. PNAS, doi: 10.1073 /pnas.1520566112.

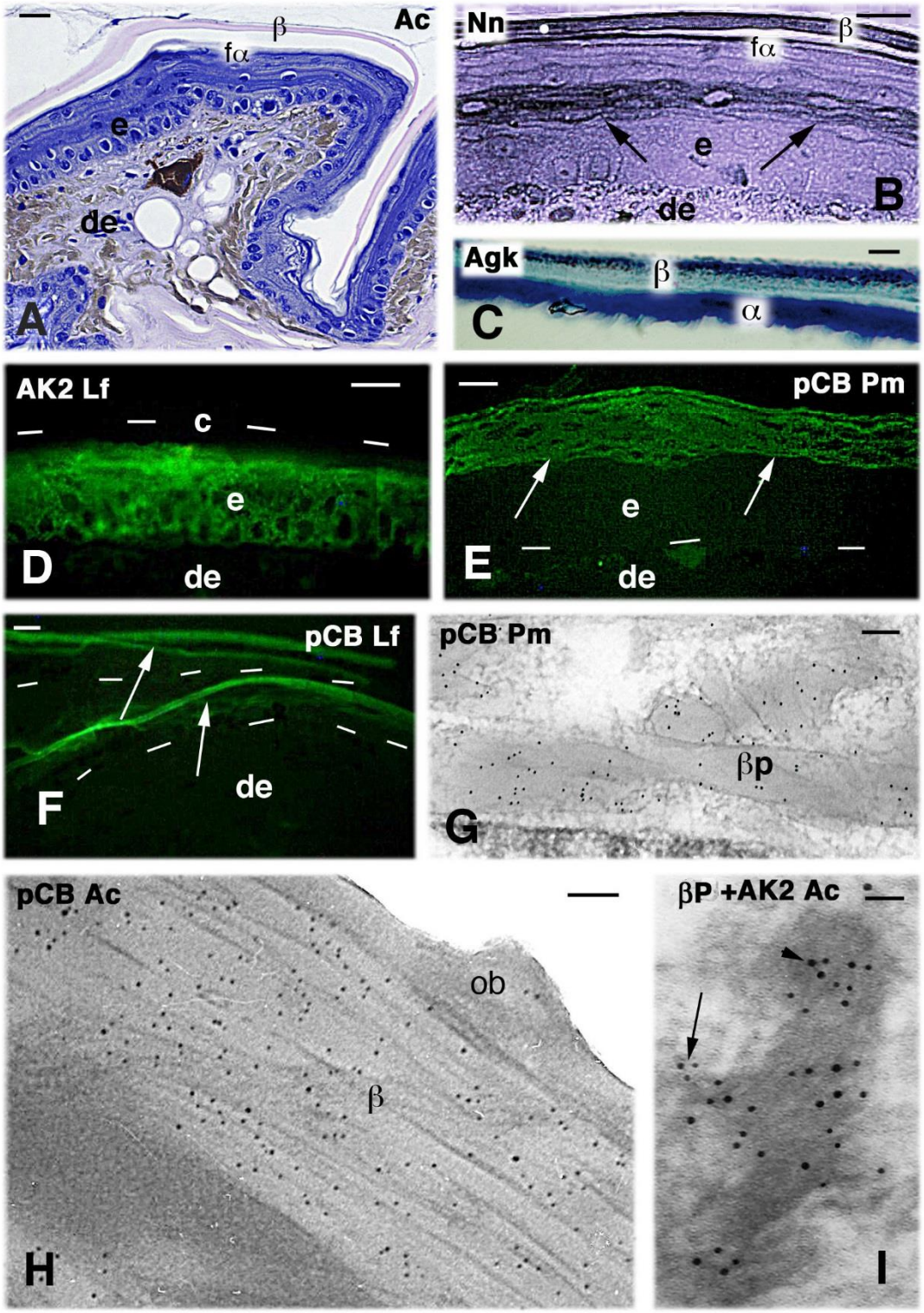
Wyld, J.A., Brush, A.H., 1979. The molecular heterogeneity and diversity of reptilian keratins. J. Mol. Evol., 12, 331-347.

Wyld, J.A., Brush, A.H., 1983. Keratin diversity in the reptilian epidermis. J. Exp. Zool. 225, 387-396.

Figures paper V

Content:

Fig. 1. Histological (**A-F**) and electron microscopic (**G-I**) views of lepidosaurian epidermis. Scale bars in all histological figures correspond to 10 μm while they represent 100 nm in all electron microscopic images. **A**, scale of the lizard *Anolis carolinensis* (Ac) showing the epidermis in post-shedding stage. **B**, detail of epidermis of the snake *N. natrix* (Nn) in renewal stage with still forming outer alpha-layer and with differentiatiating cells of the inner beta-layer (arrows). **C**, molt of the snake *Agkistrodon contortrix* (Agk) showing the paler beta-layer and the blue-stained alpha-layer. **D**, immunolabeling using the AK2 antibody for IF-keratins of the viable part of the epidermis in the snake *Liasis fuscus* (Lf, the upper dashes indicate the unstained corneous layer). **E**, scale of the lizard *Podarcis muralis* (Pm) where the forming beta-layer (arrows) is immunofluorescent using a pre-core box antibody. Dashes indicate the base of the epidermis. **F**, overlapped scales of the snake *L. fuscus* (Lf) in resting stage with the intensely labeled compact beta-layer stained by the pre-core box antibody. Dashes indicate the base of the epidermis. **G**, detail of immunolabeled beta-packets in a differentiating beta-cells of the lizard *P. muralis* (Pm). **H**, distributed immunolabeling of the mature beta-layer merged with the oberhautchen of the lizard *A. carolinensis* (Ac) using the pre-core box antibody. **I**, double labeling of a corneous packets in a differentiating beta-cell of *A. carolinensis* (Ac), using a beta-protein antibody (the arrowhead indicates the large gold particles) and the AK2 antibody (the arrow indicates the smaller gold particles) for IF-keratins.


Ac
β
fα
e
de
A
Nn
β
fα
e
de
B
Agk
β
α
C
AK2 Lf
c
e
de
D
pCB Pm
e
de
E
pCB Lf
de
F
pCB Pm
βp
G
pCB Ac
ob
β
H
βP +AK2 Ac
I

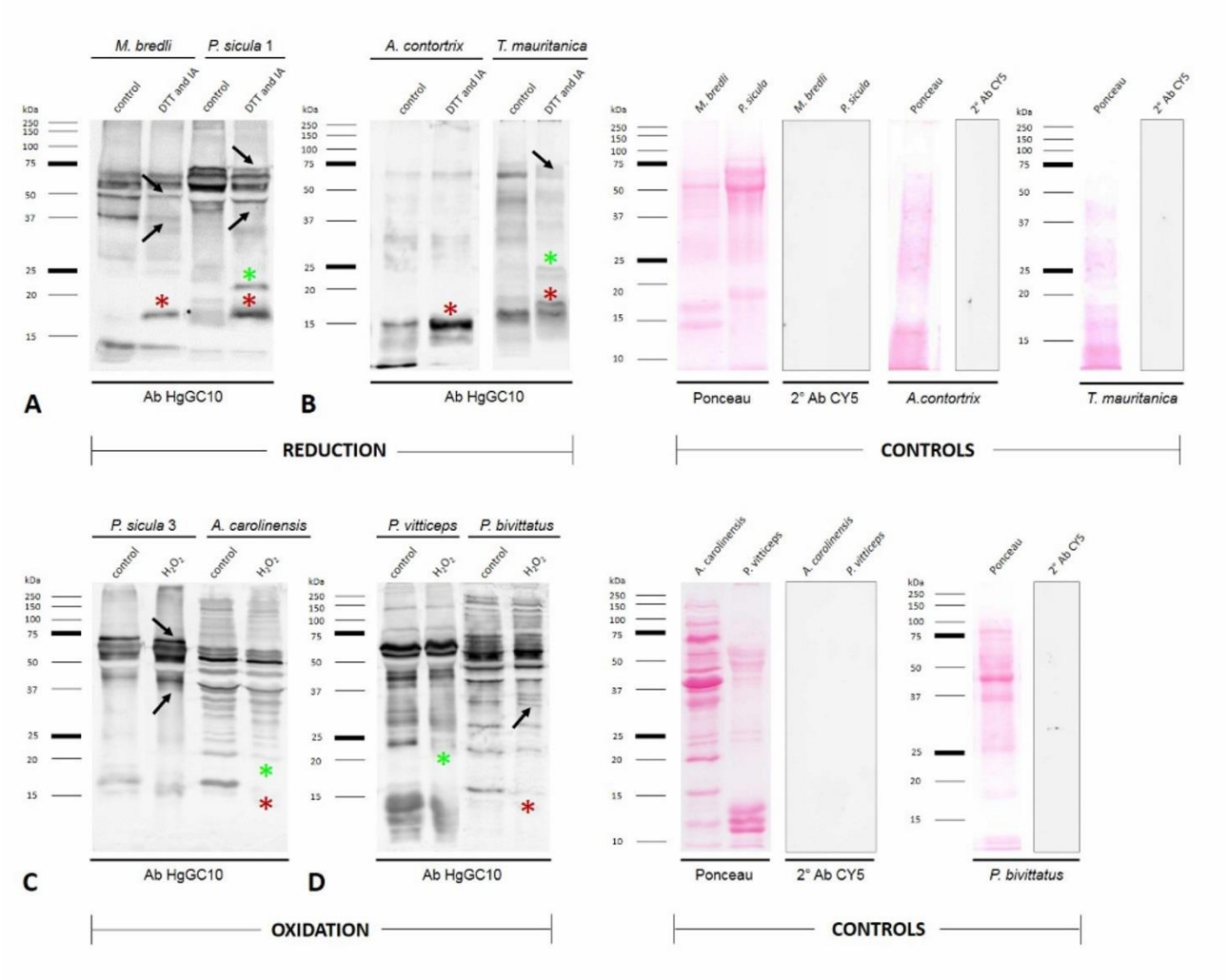


Fig. 2. The effect of a reducing treatment with DTT followed by an alkylation with IA on a snake *M. bredli* and lizard *P. sicula* with Ab HgGC10 in **3A** and on another snake *A. contortrix* and a gecko *T. mauritanica* with Ab HgGC10 in **3B**. In **C-D** the effect of oxidation with H_2O_2 on lizard skin samples of *P. sicula*, *A. carolinensis* and *P. vitticeps* and a snake skin sample of *P. bivittatus* (**3D**). **D-E)** Controls for protein presence (Ponceau) and absence of auto-immunofluorence (CY5) are shown at the right of results. The modifications of the 15-19 kDa CBP band are shown by red asterisks, while the ones of the 21-24 kDa bands by green asterisks. Black arrows indicate the intensification or weakening of bands in the MW range above the CBP one. Molecular weight is indicated by markers on the side (in kDa). A 12% polyacrylamide gel and nitrocellulose membrane were implied.

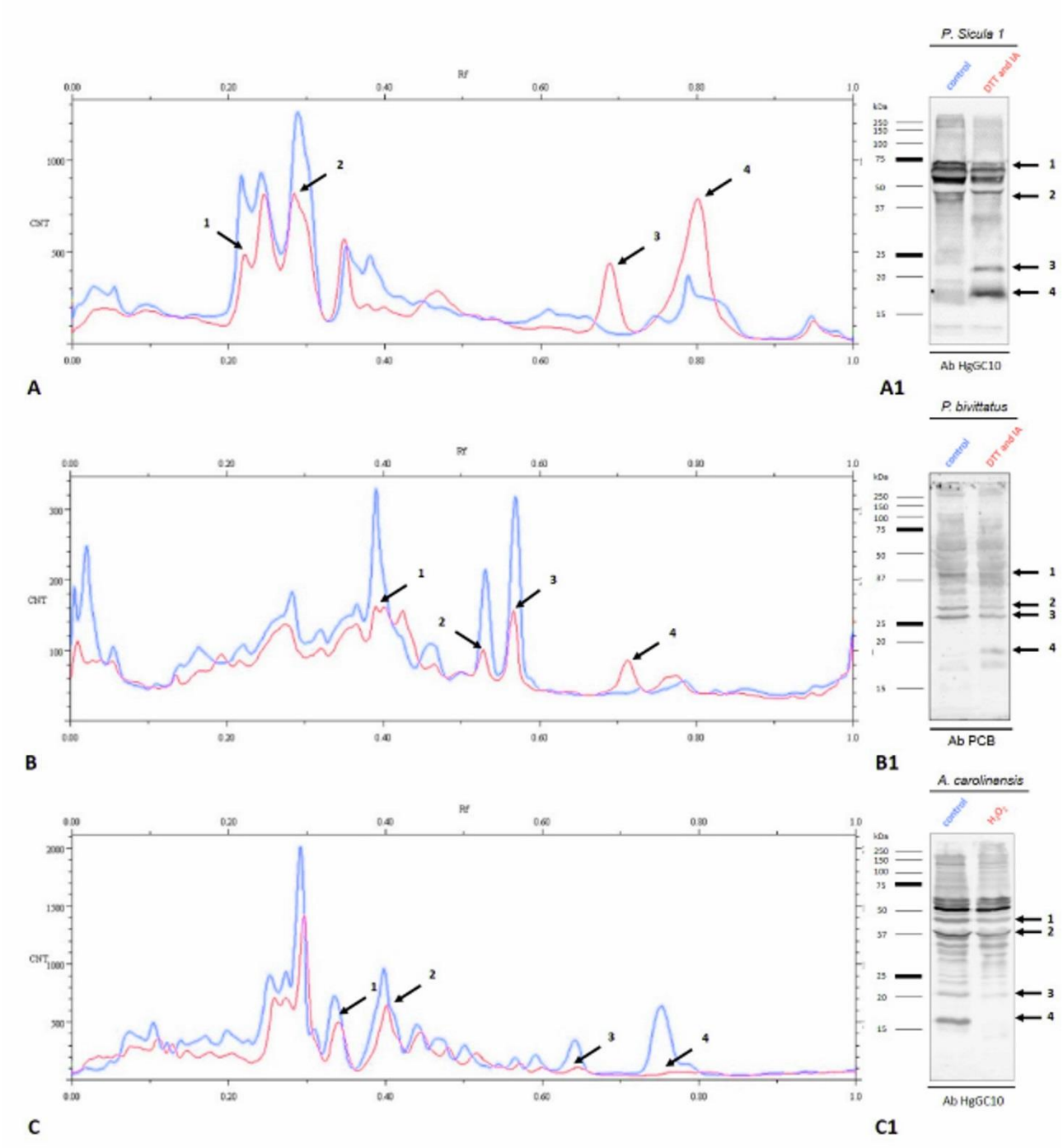


Fig. 3. Peak intensity comparison of WB bands before and after treatment. **A**) results for the lizard *P. sicula* 1 and in **A1** the WB bands corresponding to the observed peaks. **B**) results for the snake *P. bivittatus* and in **B1** the WB bands corresponding to the observed peaks. **C**) results for the lizard *A. carolinensis* and in **C1** the WB bands corresponding to the peaks. The blue line represents the controls while the red one samples threatened with reduction and alkylation (**A** and **B**) or oxidation (**C**). Black numbered arrows indicated lowered or heightened peaks corresponding to WB bands that either were reduced or oxidized.

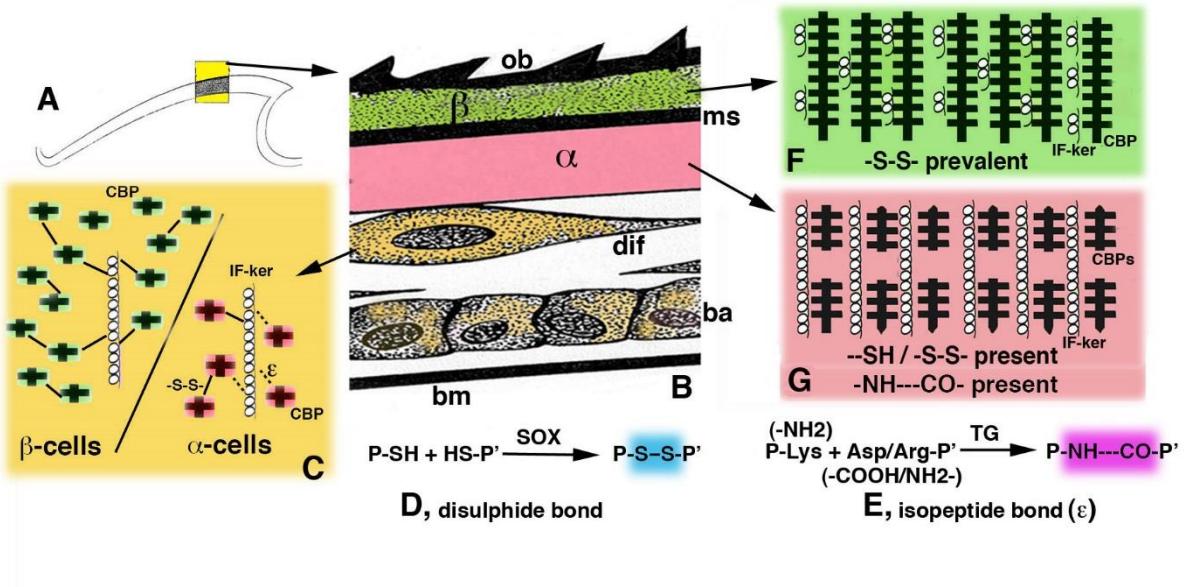


Fig. 4. Schematic drawing showing the main events occurring during cornification in a scale (**A**) with the formation of the spiny oberhautchen, followed by the beta-layer and alpha-layer (**B**). Both β- and α-cells have origin from suprabasal differentiating cells, initially the β-cells and later the α-cells (**C**). **D** and **E** show the two main types of definitive chemical bonds while **F** and **G** show the final compaction of IF-keratins and polymeric CBPs (ladder) in the mature β-layer and α-layer. **Legend**: ba, basal cells; bm, basement membrane; CBPs, corneous beta-proteins; dif, differentiating suprabasal cell; IF-ker, Intermediate filament keratins (α-keratins); ob, oberhautchen; P,P', proteins; SOX, sulfhydryl oxidase; TG, transglutaminase. Disulfide bonds are indicated with a line while isopeptide bonds (ε) are indicated with dashes.

2.6 The epidermal differentiation proteins of crocodilians

2.6.1 Introduction

Crocodilians are the closest living relatives of birds with whom they form the monophyletic clade of the Archosauria. The time of divergence between crocodilians and birds is estimated around 219-255 Mya (Chiari et al. 2012; Shen et al., 2011; Hugall et al., 2007; Janke & Arnason, 1997). Within the sauropsid (reptiles and birds) clade, archosaurs are believed to be the closest relatives to turtles from which they have separated approximately 250-257 Mya during the Late Permian (Wang et al., 2013; Chiari et al., 2012). Today the once broadly represented clade of the crocodilians has only 24 species left (Uetz & Hošek, 2017), divided into 3 families (Crocodylidae, Gavialidae and Alligatoridae). Most crocodilians live in tropical and subtropical fresh waters, but some of these predators are also found in brackish and marine waters and at higher, temperate latitudes.

At first sight crocodilians and birds do not have much in common and in systematics they have been considered separated classes, but their common ancestral origin has been confirmed by phylogenetic studies (Hugall et al., 2007; Field et al., 2014; Janke & Arnason, 1997). These diverse looking clades share oviparity, parental care for their offspring, well developed social interactions and vocal communications. At the anatomical level they exhibit similar characteristics in the skeleton (Pough et al., 1999) and in the embryonic epidermis where is present during morphogenesis the subperiderm, a layer unique to archosaurs (Sawyer et al., 2005; Sawyer & Knapp, 2003).

Crocodilian epidermal processes have been extensively studied especially for comparison of crocodilian scale morphogenesis versus the morphogenesis of avian scales and feathers (Alibardi & Thompson, 1999-2002; Alibardi et al., 2006; Alibardi, 2005b; Sawyer et al,.2000). What distinguishes the crocodilian skin is its thick armor which is formed by epidermal scutes that in

some species or regions of the body also comprise underlying same-sized and shaped bony dermal plates (osteoderms). This armor is mainly present on the dorsal part of the body and on the keeled tail, while the ventral body part and head usually exhibit flatter scales. However in some fully armored species, the armor covers the belly as well and in the Chinese alligator even the eyelids. Like in all reptiles, the epidermis is traditionally divided in layers with either an alpha or beta pattern (Maderson, 1965, 1985; Landmann, 1986; Alibardi, 2013a). In regions where protection from the environment is needed, scutes are present and the beta layer is very thick while the layer of living cells is thin (Alibardi, 2013a; Spearman, 1969). On the other hand regions requiring more flexibility like the regions between scutes (hinge regions) have a thinner epidermis and a more predominant alpha layer (Spearman & Riley, 1969; Alibardi 2003b, 2005b). The crocodilian epidermis presents thus varying thickness.

In crocodilians as in other sauropsids corneous beta proteins (beta-keratins) have been identified in the epidermis and epidermal appendages (Alibardi & Toni, 2006; Toni & Alibardi, 2007; Wyld & Brush, 1979; Maderson, 1985; Sawyer et al., 2000, 2003a; Alibardi & Thompson, 2002; Alibardi, 2003b).

Recently the Epidermal Differentiation Complex (EDC), which is also present in mammals, has been analyzed in several sauropsids like chicken, the green anole, turtles and snakes (Strasser et al., 2014; **Paper I** & **IV**). This complex comprises genes which are involved in the epidermal differentiation of keratinocytes, mainly by encoding structural protein components for the cornified cell envelope. In the stratified epithelium of terrestrial vertebrates the epidermal differentiation process begins in the basal layer where cells (keratinocytes) proliferate and start to differentiate when moving upwards through the suprabasal layers toward the skin surface. When reaching the most external stratified layer of the skin, called the cornified cell layer, cells have become terminally differentiated meaning they are cornified, dead, flat and have their plasma membrane replaced by the cornified cell envelope. Whilst the nature of this corneous envelope in reptilian

beta-cells remains unknown, the mammalian cornified layer is composed of a cytoskeleton of keratin intermediated filaments immersed in an amorphous matrix of proteins, which confers a highly resistant and insoluble structure. This layer, only present in terrestrial vertebrates, is essential for life outside an aquatic environment since it provides protection against cutaneous water loss and shields more efficiently against biological, physical, mechanical, and micobiological assaults (Eckhart et al., 2013; Kypriotou et al., 2012; Candi et al., 2005; Kalinin et al., 2002).

In this chapter the preliminary results of the study on the crocodilian EDC are reported. Although crocodilians have not evolved specialized appendages like the bird feathers or the turtle shell, analysis of their EDC is essential for revealing adaptations specific to this aquatic-adapted branch of archosaurs. Moreover as closest relatives to birds, that have evolved a highly adapted feather covered integument, the comparison with the crocodilian EDC can clarify bird specific gene losses and gains. The present results and ongoing studies also indicate that CBPs are more numerous than the 20-21 types reported for crocodilians (Greenwold & Sawyer, 2013). Comparison to other clades has confirmed synteny among the amniote EDC also for crocodilians. This study, although non conclusive, provides an inventory of EDC proteins which are involved in epidermal differentiation in crocodilians. Altogether, the present report is the basis for further in depth analysis of the proteins encoded by the crocodilian EDC.

2.6.2 Material and methods

To perform the comparative analysis, the EDC genes of crocodilians were first of all predicted using the genome sequences derived from two species: the Chinese alligator (*Alligator sinensis*) (Wan et al., 2013), and the Australian saltwater crocodile (*C. porosus*) (Green et al., 2014). To determine the ancestral condition for observed genes losses and duplications in the two species principally investigated, the available genomic sequences of the American alligator (*A. mississippiensis*) and the Indian gharial (*G. gangeticus*) were scrutinized (Green et al., 2014). In

Supplementary Tables S2–S5 the accession numbers of genome sequence scaffolds corresponding to the crocodilian EDCs are listed.

Coding sequences and exon–intron borders of EDC genes were predicted screening the nucleotide sequence between *S100A12* and *S100A11* genes of the target genome, as previously done for lizard, turtle and snakes (Strasser et al., 2014; **Paper I & IV**). This screening was primarily performed with tBLASTn using as queries the amino acid sequences of chicken and turtle (*C. picta*) EDC proteins. In case of EDC regions with apparently low gene density the nucleotide sequence was translated *in silico*, which permitted the identification of additional open reading frames of candidate EDC genes. These likely EDC genes were verified for the presence of both a splice acceptor site localized 10–30 nt upstream of the start codon and a putative noncoding exon 1, which was defined by a TATA box and a splice donor site 60–90 nt away. Furthermore RNA-seq reads, available in the NCBI browser for "genomic regions, transcripts, and products", were used to identify transcribed regions of the *A. sinensis* EDC. Subsequently the transcribed regions were translated using ExPASy, and the resulting amino acid sequences were investigated for the potential to encode proteins with amino acid sequences similar to known EDC proteins. The methods here applied followed the criteria published in Strasser et al. 2014.

Several bioinformatics portals were used such as Multalin (Corpet 1988) for aligning amino acid sequences and ExPASy (Artimo et al. 2012) for calculating amino acid composition (ProtParam tool).

2.6.3 Results

Identification of the epidermal differentiation complex (EDC) in crocodilian genomes

As in all amniotes the EDC of crocodilians was defined as the genomic region bordered by S100A genes (Strasser et al., 2014; **Paper I & IV**). The EDC gene complement of the two principal

investigated species, the Chinese alligator (*Alligator sinensis*) and the Australian saltwater crocodile (*Crocodylus porosus*) was identified primarily by tBLASTn searches using EDC-encoded proteins of chicken and turtle (*Chrysemys picta*) as queries, but also by de novo prediction of genes in an iterative process, that was reported in Strasser et al., 2014.

The nomenclature for EDC genes follows the preliminary system that was defined in Strasser et al., 2014 and consists in the term Epidermal Differentiation (ED) followed by a term describing the amino acid composition or the presence of particular amino acid sequence motifs in the encoded gene. In the text only name abbreviations are used to simplify reading, while full gene names can be found in Supplementary Table S1. In the case of orthology with human loricrin, peptidoglycan recognition protein 3 (*PGLYRP3*) and cornulin and sauropsid scaffoldin (*SCFN*) this naming convention was not applied.

To verify expression of the predicted EDC genes of crocodilians, the RNA sequencing data of the Chinese alligator was screened for transcribed regions and most genes were confirmed to be expressed (Suppl. Table S2).

The EDC of crocodilians confirms conservation within amniotes and sauropsids

The EDC of crocodilians is the last of the sauropsid clade with available genome sequence to be analyzed and confirms the conservation of various genes and their proteins within amniotes, as was already observed in other investigated sauropsids (Strasser et al., 2014; Vanhoutteghem et al., 2008; **Paper I & IV**). This conservation is encountered in the bordering S100A genes, S100 fused-type genes or SFTPs located before S100A11, the peptidoglycan recognition protein 3 (*PGLYRP3*), loricrin and SPRR-like genes (Fig.1). Both the S100A and SFTPs genes contain a specific amino-terminal S100A domain (Ca^{2+}-binding EF-hand motifs), while *PGLYRP3* contains a characteristic structural fold also found in other PGLYRP proteins (Marchler-Bauer et al., 2016). Most of the EDC of crocodilians comprises, like in other amniotes, simple (single coding exon) EDC (SEDC) genes,

which apart from CBPs do not contain sequences associated with a known protein domain (Fig. 2-3). In previous studies several conserved sauropsid specific genes were identified on the EDC of which *EDKM*, *EDWM*, *EDQL*, *EDYM1*, *EDP3* and *SCFN* have also been found in crocodilians (Fig. 1-3). As in the EDC of all sauropsids the crocodilian EDC contains a CBP gene cluster which in this cluster has the lowest number of genes (n=25-32) identified up to now.

Apart from specific proteins, conservation is found as well at the level of sequence motifs at the amino and carboxy-terminus (Fig. 4A-D) and at the level of amino acid repeats with high abundance of certain amino acids, as previously reported (Strasser et al. 2014; **Paper I & IV**). The terminal sequence motifs concern mainly lysine and glutamine residues, which are the target amino acids of transglutamination suggesting that this type of protein cross-linking is a conserved feature of amniote EDC proteins. The amino acids that characterize amniote SEDC proteins by their abundance or/and tandem repeats are glycine (G), serine (S), proline (P), glutamine (Q), and cysteine (C) (Strasser et al., 2014; **Paper I & IV**; Fig.5A-D). Likewise, the EDC of the Chinese alligator comprises genes that encode proteins with high contents of G (loricrin 40,7 %), C (EDC3 23,6 %), P (EDPQ1 38,2 %, EDPE 37,1 %) and Q (EDPQ1 19,1 %, EDP3 18,1 %) (Fig. 2).

Genes unique to crocodilians were identified in the small cluster of C-rich genes *(EDCs)*, the EDPCQ and EDRYA gene and a second EDDM-like gene (Fig.1).

Three large gene clusters compose most of the EDC in crocodilians

What is most striking about the EDC of crocodilians is that most of the genes present belong mainly to three large gene clusters containing each dozens of genes. These clusters consist in EDCH, EDAA and CBP genes that together comprise about 75% of all the EDC genes. The CBPs are known for their specific central domain which is predicted to form a pleated beta sheet involved in dimerization of these proteins that can compose both the filament and matrix component (Fraser & Parry, 1996, 2014, 2017; Alibardi et al. 2009; Calvaresi et al., 2016). *EDAAs* (Epidermal

Differentiation proteins rich in Aromatic Amino acids) form a cluster of genes which is located in the central region of the EDC before the CBPs (Fig.1) and encodes proteins rich in aromatic amino acids, particularly tryptophan and tyrosine (Fig. 5D). These genes are believed to be homologous to the turtle EDAA genes (**Paper I**) and to chicken genes that were previously named EDMTFs (Strasser et al. 2014). The last of the large gene clusters is composed of *EDCHs* (Epidermal Differentiation proteins containing Cysteine Histidine motifs Fig. 5C), which are characterized by the particularity of having the first half of the protein rich in serine and these CH motifs, while the second half is rich in proline, glutamine and cysteine duplets and lysine residues.

Other genes that form clusters in crocodilians although to minor extent are the *EDPQs* with 2-6 members (Fig. 5B). These proline-rich proteins are not found in chicken, but show likely orthology with the 2 EDPQ genes present in the turtle *C. picta*. Finally, cysteine-rich amino acid sequences are encoded by a small cluster of 2-4 SEDC genes (Fig. 5A).

The fact that the crocodilian EDC comprises five clusters (3 large and 2 small ones) of distinct gene types (*EDCHs*, *EDAAs CBPs*, *EDPQs* and *EDCs*) is similar to what was observed for the structure of the turtle EDC (**Paper I**). In turtles though, the gene type of some of the expanded clusters was different than the one in crocodilians and the CBP cluster composed by far the most dominant cluster. Moreover part of the genes belonging to two of these expanded clusters were translocated to loci outside the EDC in turtles.

The crocodilian EDC reveals gene losses and gains within the archosaurs-turtle branch

Analysis of the crocodilian EDC has made it possible to reveal genes specific to the archosaur clade as well as bird specific adaptations within the archosaur-turtle lineage. Additionally the crocodilian gene complement has confirmed that *EDAAs* and *EDPE* are found in all clades of the archosaur-turtle branch, but not in lepidosaurs. These genes are therefore specific to the archosaur-turtle lineage, while EDCH and EDDM-like genes have turned out to be unique to just the archosaur

lineage. The crocodilian *EDDM1* and *2*, are similar to chicken *EDDM* (Fig. 1), but also EDQCM, thus orthology is not absolutely certain.

Even though birds are the crocodilians closest relatives, avians have evolved a highly differentiated tegument. In fact, bird specific genes and gene duplications as well as bird specific gene loss and reduction that could be involved in this adaptational process have been identified by comparison to crocodilians. One example of expansion in birds is loricrin which has expanded to three copies with respect to the one present in crocodilians, like in most amniotes (Fig.1). Whereas the lower number of EDCH genes in chicken (n=4) with respect to *A. sinensis* (n=22, Fig. 2) can be due to both a crocodilian specific expansion or bird specific loss of this gene family. Notably a high presence of proline-rich genes, namely an EDPL, EDP1, EDP2 and EDPCV gene as well as an EDPQ cluster was found in crocodilians, feature shared with the turtle clade, but not prominent in the chicken which is the closest relative to crocodilians (Fig.1). An *EDPCV* gene orthologous to the turtle one was identified in *A. sinensis*, whereas in the chicken it appears to lack counterparts. These *EDPCVs* formed a cluster of 15 genes in turtles while in crocodilians only one gene was found that was even lost in *C. porosus* (Fig.1, 3). The number of EDAA genes (n=18-19) of most crocodilian species (Fig. 2-3, Suppl. Fig. 1-2) was higher than the number of homologous EDMTF genes in the chicken (n=5). Eventhough the total number of EDAA genes of the turtle *C. picta* (n=22) was similar to crocodilians, part of these genes were translocated to another locus in turtles. For the EDC locus, the number of EDAAs in all turtle species investigated was around 5-10 (**Paper I**), so lower than in crocodilians.

Thus the structure of the crocodilian EDC shows similarity to both the chicken and the turtle (*C. picta*) (Fig. 1). On one hand the presence of large gene clusters and proline-rich genes is shared with the turtle clade. On the other hand EDCH and EDDM-like genes are in common with the chicken, even though the *EDCHs* were either reduced in chicken or expanded in crocodilians (Fig. 1).

Differences in gene duplications and reduction between crocodilian clades

Although most of the SEDC gene sequences of the different crocodilian species analyzed, namely *A. sinensis*, *C. porosus*, *G. gangeticus* and *A. mississippiensis*, were highly conserved and displayed orthology (Fig. 5A-D), several genes underwent differential expansion or loss in diverse species and subclades (Fig. 1-3). To determine the ancestral condition for the observed differences between the Chinese alligator (*A. sinensis*) and the Australian saltwater crocodile (*C. porosus*), results were confronted with other available genomic sequences which were of the Indian gavial (*G. gangeticus*), the closest outgroup to Crocodylidea, and of the American alligator (*A. mississippiensis*) the only other extant alligator (Suppl. Fig. 1-2).

The most remarkable difference is found for *A. sinensis* which has a higher number of both CBPs (n=43) and EDAAs (n=41). This additional expansion, which brings EDAAs and CBPs at practically the same number, seems specific to the *A. sinensis*, since it is not found in the *A. mississippiensis* (Suppl. Fig. 1). Further comparison suggested that the *C. porosus* has lost some proline-rich genes like *EDPCV* and reduced others like *EDPQs* from 6 to only 2 copies (Fig. 1-3). Reduction is also seen for genes such as *EDCs* that are present in a lower number in the Crocodylidae and Gavialidea lineage with respect to the Alligatoridae family (Fig. 1-3, Suppl. Fig. 1-2). This could be confirmed by the presence of mutated *EDCs* in both *G. gangeticus* and *C. porosus*. Furthermore also the EDCH cluster contains a lower number of genes in Gavialidea and Crocodylidae with respect to the Alligatoridae family (Fig. 1-3, Suppl. Fig. 1-2).

Thus, the differential evolution with expansions and losses of specific EDC genes in the diverse subclades and/or species of crocodilians remains to be investigated in future studies.

2.6.4 Discussion

The preliminary results of this comparative study suggest that crocodilians like other sauropsids show conservation with the amniote EDC (Strasser et al., 2014; de Guzmang Strong et

al., 2010; Vanhoutteghem et al., 2008; **Paper I & IV**). Apart from the identification of S100A, CRNN, PGLYRP3, loricrin and SPRR-like genes orthologous to the human ones, also specific amino-and carboxy-terminal sequence motifs were found to be conserved with mammalian proteins. In addition, the amino acid composition of EDC proteins biased towards several amino acids such as glycine, cysteine, proline, glutamine, serine and lysine, often in the form of tandem repeats, was confirmed for crocodilians like for other amniotes (Strasser et al., 2014; **Paper I &IV**; Candi et al., 2005; Kalanin et al., 2002). The observed conservation suggests that probably in all amniotes the same basic mechanisms are involved in forming the epidermal barrier. In mammals cross-linking of several EDC proteins is performed by transglutaminase enzymes that use lysine and glutamine residues as targets (Candi et al., 1998, 1999; Kalanin et al., 2002; Steinert et al., 1999) and based on observed terminal sequence homology for various proteins this likely occurs in crocodilians and other sauropsids as well (Strasser et al., 2014; Vanhoutteghem et al., 2008; **Paper I & IV**).

The epidermal barrier is a highly resistant and insoluble structure forming a first-line defense against dehydration and environmental assaults of biological, mechanical, chemical and physical nature (Kypriotou et al., 2012; Henry et al., 2012; Kalinin et al., 2002; Candi et al, 2005). In amniotes it consists in the cornified layer that is buildup by layers of keratinocytes that have undergone terminal differentiation, cornification and replacement of their plasma membrane with the cornified cell envelope (CE) (Eckhart et al., 2013; Henry et al., 2012; Kalinin et al., 2002; Candi et al, 2005). Many proteins encoded on the EDC are involved in the epidermal differentiation process as structural components of the CE. Other EDC proteins associate to keratin intermediate filaments (KIFs) during the cornification process in the epidermis and appendages. Whereas the KIFs form the filament of the cytoskeleton in keratinocytes, the proteinaceous matrix component is constituted of keratin-associated proteins to which also some EDC proteins such as mammalian filaggrin and sauropsid CBPs belong. KIFs confer both strength and flexibility to the epidermis, while on the other hand proteins associated to keratin endow the cornified layer with inflexibility,

insolubility and extreme mechanical resilience (Alibardi, 2006, 2013a; Resing & Dale, 1991; Kalinin et al, 2002). So the mechanical and physical properties of the epidermis and appendages depend on the type and proportion of proteins associated to keratins or matrix proteins, but also on the kind of protein bonds formed. For example in mammalian appendages both the type of proteins (KRTAP) and bonds (disulfide bonds) are different than in the epidermis (Eckhart et al., 2013; Rogers et al., 2006; Gillespie, 1991; Powell & Rogers, 1994; Matoltsy, 1987). At the same time both the amino acid composition and the structural organization of EDC proteins can influence the characteristics of the integument in which they are present. An example are the CBPs that with their characteristic secondary pleated beta sheet structure are believed to be responsible for the toughness of sauropsid epidermis and appendages (Alibardi, 2003; 2016a; Spearman, 1969; Landmann, 1986). EDAA genes, that form one of the big clusters in crocodilians, are rich in tyrosine which has the property to strengthen the corneous material. When glycine residues are combined to tyrosine, this adds pliability to the material characteristics (Fraser & Parry, 2014; Parry et al., 2006). This tyrosine-glycine combination is found in some mammalian KRTAPs (Gillespie, 1991) and terminal sequences of some sauropsid CBPs including those of crocodilians (Dalla Valle et al., 2008, 2009b; Gregg & Rogers, 1986; Greenwold & Sawyer, 2011, 2013).

The basic organization of the EDC was probably inherited from a common ancestor of archosaurs (crocodilians and birds) and their next relatives, the testudines. In fact some orthologous proteins such as EDAAs and EDPE were found in both archosaurs and turtles. Furthermore the crocodilian EDC complement reveals that the EDCH and EDDM proteins that they have in common with the chicken are specific to the archosaur clade. In comparison to crocodilians and turtles, birds seem not only to have reduced proline-rich genes through loosing EDPCV, EDP1-2 and EDPQ genes, but also aromatic acid rich genes such as EDAAs. Possible further reduction in birds could include EDCH genes, although it cannot be excluded that instead these genes were expanded in crocodilians.

Expansion of the CBP cluster in birds and turtles has been put in relation to the evolution of specific appendages such as feathers and the shell respectively (Greenwold & Sawyer, 2010; Li et al., 2013). In both birds and turtles genes involved in forming specific appendages have been translocated outside the EDC (**Paper I**; Greenwold & Sawyer, 2010; Ng et al., 2014). Even though the crocodilians exhibit, like the turtle the presence of large clusters on the EDC, no evidence of translocation was found in the species investigated. Only in *A. sinensis* a scaffold with non EDC proteins on the same scaffold as some EDCHs was identified, but this difference may be due to the fact that the genome assemblies of crocodilian species are not of the same quality as those of mammals neither the position or assembly of many scaffolds is certain. Since crocodilians have not evolved any specific appendage for which gene duplication was needed, the presence of large gene clusters and gene redundancy could be a primitive characteristic of the archosaur-turtle lineage. Subsequently this characteristic could have been eliminated in birds, which are known to have reduced their genome size. In none of the squamates investigated so far large gene clusters were found with exception of the CBPs. Instead the squamate EDC consisted in many individual genes and only some small gene clusters (Strasser et al., 2014; **Paper IV**). Apart from the CBPs, abundant in the beta layer located in the outermost part of the epidermis involved in protecting vertebrates from the environment (Baden & Maderson, 1970; Landmann, 1986; Sawyer et al., 2000; Alibardi & Toni, 2006; Alibardi, 2003b, 2005a, 2013a), nothing is known about the function or localization of proteins such as EDCHs and EDAAs that are encoded by the other two major gene clusters.

Notably the expansion of several gene clusters, namely EDAAs, EDCs, EDCHs, EDPQs and CBPs, diverged during evolution quite markedly between crocodilian families and even species of the same family. Crocodilians are considered a primitive slow evolving clade, but the differences encountered in the EDC genes reveal the existence of specific adaptations of the epidermal barrier in these reptiles,that could be related to adaptation to the aquatic environment. The degree of armored scutes covering the body varies between different species of crocodilians (Brazaitis, 1987;

Richardson et al., 2002) and this might be involved in the observed differences, but at the moment no experimental evidence can confirm this hypothesis. In particular, the Chinese alligator has undergone further expansion of the CBP, EDAA and EDCH gene clusters compared to other crocodilians. The reason for this species specific additional expansion of EDC genes in comparison to its closest relative, the American alligator, remains a mystery and needs further investigation.

2.6.5 Conclusion

In summary, the preliminary results of the present comparative genomics analysis confirm synteny to the amniote EDC for crocodilians, as was the case for all sauropsids investigated so far (Strasser et al., 2014; Vanhoutteghem et al., 2008; **Paper I** & **IV**). In addition, during evolution in crocodilians several genes involved in the epidermal barrier were expanded leading to dominance of only a few gene types in the crocodilian EDC. While the expansion of EDAA and EDCH genes seems to be crocodile specific, the CBP expansion is common to all sauropsids, but is minor in crocodilians compared to other sauropsids.

The comparative analysis with birds, the closest relatives of crocodilians, has revealed genes that have evolved specifically in the archosaurian lineage such as EDCHs and EDDM-like. In crocodilians though, the gene clusters of *EDCH* and *EDAA* have much higher numbers compared to birds (Strasser et al., 2014). Furthermore some genes shared between archosaurs and turtles such as *EDAAs* and *EDPE* are not found in squamates (Strasser et al., 2014; **Paper IV**) thus indicating a gene adaptation specific to the archosaur and turtle lineage (**Paper I**). Notably, crocodilians share with the turtle a good number of proline-rich genes that are not present in chicken (Strasser et al., 2014; **Paper I**). Several genes such as a small cluster of cysteine-rich genes *(EDCs)*, the *EDPCQ, EDRYA* and the second *EDDM* gene were found to be unique to crocodilians.

In conclusion this study provides for the first time a comprehensive catalog of EDC genes identified in the crocodilian clade. These molecular data will serve future investigations into the

expression of crocodilian EDC genes and the evolution of claws, scutes, and scales in crocodilians in comparison to birds, their closest extant relatives.

2.6.6 Figures: The epidermal differentiation proteins of crocodilians

Content:

Figures 1-5

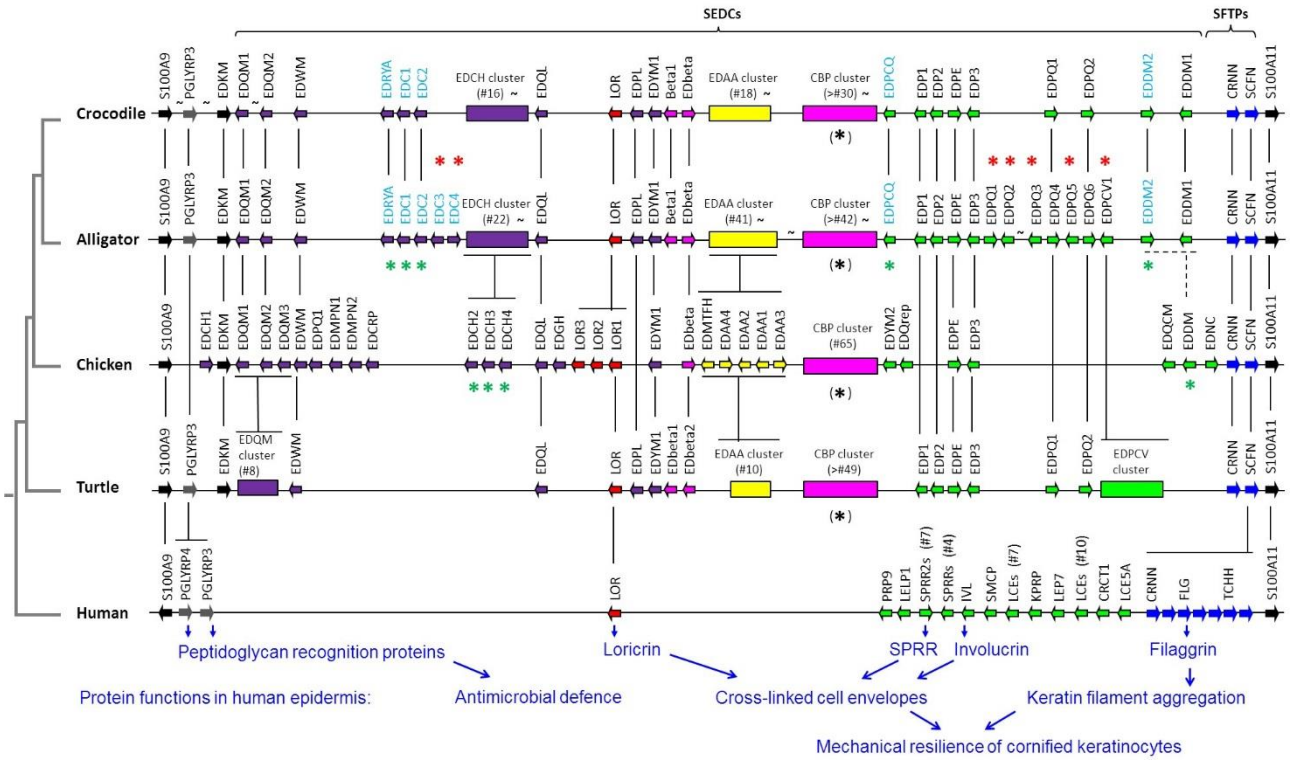


Figure 1. Organization of the epidermal differentiation complex (EDC) in the Chinese alligator and the Australian saltwater crocodile. Genes of the EDC in crocodilians *Crocodylus porosus* and *Alligator sinensis*, in comparison to those of the chicken (*G. gallus*), the turtle (*C. picta*), and human, are schematically depicted. Arrows indicate the orientation of the genes. Simple EDC (SEDC) genes with 2 exons are represented by colored arrows with a black frame whereas other genes are shown as filled arrows. Corneous beta-protein (CBP) gene clusters are shown as boxes in this diagram while detailed information about the genes in these clusters are depicted in Fig. 2. Members of gene families are numbered according to the positions of genes without indicating 1:1 orthology to specific members of the same gene family in other species. The depiction of the human EDC is simplified by representing gene family clusters with arrows and indicating the total number (#) of genes within each cluster. Black vertical lines connect orthologous genes or gene families. Green and red asterisks indicate putative gene gain and loss events whereas black asterisks indicated gene differences that could not be unambiguously assigned to an evolutionary event in particular lineage. Note that the diagram is not drawn to scale. The symbol ~ marks interruptions in the crocrodilian EDC scaffold. Because of improved delineation of orthology relationships, the following gene names have been newly assigned to replace previous names (Strasser et al., 2014): chicken EDQM3 instead of EDSC, and EDPQ1 instead of EDCH5.

A

>Asi_CRNN_partial
MTQLQGNIEGIISAFNAYAKKDGGCITLSKGELKQLIQQEFADVLVKPHDLQTIDQVLQRLDAESKDRIDFDEFLVLVFQVAKACHKKL
NPCQPSGDGQGSASQGDASRGQAQKADQEQAPEQGQKQPEAPEQDPTRPQAPETRTAEGNLSRRHTQDPEVSGTQDPEVSQGDGNHEAQ
AAETPEHDSIHRQGQEPEQDPSHHRAQEQDPNRGGQDPQVPQQDVKHETLESGAPEQAFNHHPVLQPSVSERDLDHCQSSASERDLDHH
PTLQPSTLETDLDRHPTLQPNTLETDLDRHPTLQPNTLETDLDRHPTLQPNTLETDLDRHPTLQPNTLETDLDRHPTLQPNTLETDLDR
HPTLQPNTLETDLDRHPTLQPNTLETDLDRHPTLQPNTLETDLDRHPTLQPNTLETDLDRHPTLQPNTLETDLDRHPTLQPNTLETDLD
RHPTLQPNTLETDLDRLPTPHPSTLETDLDRHPTLQPNTLETDLDRLPTPHPSTLETDLDRHPTLQPNTLETDLDRLPTPHPSTLETDL
DHHQGLETEAPEQDLSSTTGRESTHNEAHESQALEQESQESQPDEQDQNRHQSKEPEASDHTLRHQPQEAQPTEQDLTWETQTPRVPEG
DVSRGGTPFSPALQQXXXARQDPREEALTAYRPYIYQCQKPPTFPYQWLPKQ

>Asi_EDAA1
MFDSLDTIEDLCYQGCWDLCYRRPYWNCCWDPCTYRRPSWSGCWDPCTYRRPYIYDNCYGYGGLYGLGGCYPYSSRWGHKYSYGNCWPC

>Asi_EDAA2
MFDNLDTTEDLWYKGQSGCWDPCYRRPSWSGCWDPYYRRSSWSGCWDPCYRRPSWSSCWDPCYSRSSWNSCWDPCNYRRPYSYGLGSSY
GSGFCYPYASRWGRRGSYGKCWPC

>Asi_EDAA3
MFDSLDTIEDLCYQGCWDPCYRRPYWNCCWDPCTYRRPSWSGCWDPCTYRRPYIYDNCYGYGGLYGLGGCYPYSSRWGHKYSYGNCWPC

>Asi_EDAA4
MFDNLDTTEDLWYKGQSGCWDPCYRRPSWSGCWDPYYRRSSWSGCWDPCYRRPSWSSCWDPCYSRSSWNSCWDPCNYRRPYSYGLGSSY
GSGFCYPYASRWGRRGSYGKCWPC

>Asi_EDAA5
MSDSLGMLEDLCFQGPNCCFPPCPKRLCYCCCYDPCTGKLLWKGLCCCPCWGGLGGHGKFRPQSTWSKTDVKSKIQIPG

>Asi_EDAA6
MFDSLDMIEDLSYHGQSDCFPHYRRRPPYTCSCYDQCGRLVCRGCCWSIPPWWCHQCSSGGCWPC

>Asi_EDAA7_partial
MSDSLDMFEDLYYQSPDCCWXXX

>Asi_EDAA8
MFDSLDTVEDLFYPGQSDFCWPQYPQRRYWCCCYDRYGCLVWQGPCCYPGPWCRR

>Asi_EDAA9
MSDSLDMLEDLHYQDSSCCWRPPCRRRCWCCCYDPCTGQLIWQGMCWCPGWRSRGRYGRCWPC

>Asi_EDAA10
MFDSLDTIEDLSYQGQSDCFPYQRRRPRYTCCCYDQCGRLVWRGCCWSIPPWWCHQGSPGGCWPC

>Asi_EDAA11
MSDSLGMLEDLCFQGPNCCFPPCPKRLCYCCCYDPCTGKLLWKGLCCCPCWGGLGGHGKFRPC

>Asi_EDAA12
MSDSQDMFEDLYYQSPDCCWRPRHKKPCYCCCYDPCTGELVWEGWCCCPCWRGRGRYGRCWPC

>Asi_EDAA13
MSDSLDMLEELSFQDPYCCWRPRRTKLCWCCCCDPCTGEVVWEGWCCCPCWHGRGRYGRHWPC

>Asi_EDAA14
MSDSLGMLEDLCYQDPSCCWRPRRTKLCWCCCCDPCTGEVVWEGWCCCPCWRGRGSYGRRWPC

>Asi_EDAA15
MIDSLDTIEDLSYPGQSDCFPHYCRRPPYTCCCCDQCGRLVWHGCCWSIPRWWCSKGSSGSSCPC

>Asi_EDAA16
MFDSLDMIEDLHYPGQSDCLPYQRRRPPYTCCCYDQWGRLVWRGCCWSAPPWGCR

>Asi_EDAA17
MFDSLDMIKGLSYQGQSGCFLYQRRRPPYTCCCYDQCRRLVWRGCCWSISPWWCRQGSSGGCWPC

>Asi_EDAA18

MFDSLDMIEDLHYPGQSDCFPHQRRRPPYTCCCYDQWGRLVWRGCCWSAPPWWCR

>Asi_EDAA19
MSDSLGMLGDLCYQDSSCCWRPRRRKPCYYCCYDPCTGQLIWQGMCCCPLWRGRGRYGRFWPC

>Asi_EDAA20
MFDSLDTVEDLFYPGQSDFHWLPYPQRRYWCCCYDRYGRLVWQGPCCYPGPWYRR

>Asi_EDAA21
MFDSLDMIENLSYQGQSDCFPYYHRRPSYTCCCFDQCGRLVCRGCCWSIPS

>Asi_EDAA22
MIEDLFYQGQSDCFPHYRRRPLYMCCYYDQCGRLVWQGCCWSIPRWWCCQGSSGGCWPC

>Asi_EDAA23
MSDSLDMLEDLYYQSPDCCWKPRRTKPCYCCFYDPCTGELVWEGRCCCPCWAGRGRYGGCCPC

>Asi_EDAA24
MFDSLDTIEDLSYHGQAGCFPHYRRRPRYTCCCYDQCGRLVWQGCCWSITHWWCCQGSSGGCWPC

>Asi_EDAA25
MFDSLDTIEDLSYHGQAGCFPHYRRRPRYTCCCYDQCGRLVWQGCCWSITHWWCCQGSSGGCWPC

>Asi_EDAA26
MSDSLGMFEDLYYQYPTCCWRPRRRKPCYCCCYDPCTGELVWQGWCCCPCWGGRGSYGRRWPC

>Asi_EDAA27
MFDSLDMIEDLSYHGQSDCFPHYRRRPLYTCCCYDQCGRLVWRGCCWSIPRWWCYRRSSGSSWPC

>Asi_EDAA28
MSDSLDMFEDLYYQSPICCWRPRRRKPCYCCCYDPCTGELVWQGWCCCPCWGGRGSYGRRWPC

>Asi_EDAA29
MSDSLGMLEDLCFQGPDCCFTPRPIKLCYCCCYDPCTGKLLWKGLCCCPCWGGLGGHCKFRPC

>Asi_EDAA30
MFDSLDMIEDLSYPGQSDCFPHYHRRPPYTCCCFDQCGRLVCRGCCWSIPS

>Asi_EDAA31
MFDSLDTIEDLIYPGQLDCRSPYPGRRYWCCCCDRCGRLVWQGPCCYPRLLCPHKDSSGSSWPC

>Asi_EDAA32
MFDSLDMIEVLIYTGQLDCRPPYPGRRYWCCCCDRCGRLVWQGPCCYPGLWCPHKDSSGSSWPC

>Asi_EDAA33
MFDSLDMIEDLHYPGQLDCRPPYPGRRYWCCCYDRYGCLVWQGPCCYPGPWCRR

>Asi_EDAA34
MSDSLNMLENFRYPGQSYCWDPCYRRPYWNNWWDPCTYRRPYIYDNCYGYGGLYGLGGCYPYSSRWGRRGSWGNCWPC

>Asi_EDAA35
MSDSLNMLENFCYPGQSNCWDPCYRRPYWNNWWDPCTYRRPFWSGCWDPYTYRRPYIYDNCYGYGGLYGLGGCFPSSSRWGRRGSWGSC
WPC

>Asi_EDAA36
MTYHQSGCDDVCYTPCNYGGLYGYQGLTGCWEPWTYGRPYSYGCWNPRTYRWPDSYWEPCGYGSG

>Asi_EDAA37
MSDSLNMLENFRYPGQSYCWDPCYRRPYWNNWWDPCTYRRPYIYDNCYGYGGLYGLGGCYPYSSRWGRRGSWGNCWPC

>Asi_EDAA38
MTYHQSGCDNVCCTPCSYGGLYGYQGLTGCWEPWTYGRPYSYGCWNPRTYRWPDSYWDPCGYGSG

>Asi_EDAA39
MSDSLDMLEDLWYPGQSNCWDPCYRRPYWNNCWDPCTYRRPFRSGCWDPCTYRKPYIYNRCYGYGGLYGAGGCFPYSTRWGRRYSAGNC
WPC

>Asi_EDAA40
MSDSLDMTEDLYYQYPTCCWRPPCRRRCWCCCYNPCTGQLIWQGWCWCPGWRSRGRYGRCWPC

>Asi_EDAA41
MLDSLDTLYDLFYQGQSDCWPPTPRRPPYTCCCYDRCRRLVLHGCCWSIPPW

>Asi_EDC1
MCSCCSGCHGTRSVQPICYVQPVCCEPVYIHRSSGSCCQPCGSCCGSCCGFCCRGSRSCPRVVIQRRPMPVCCPPLQYSAPMQQHCSPL
KKC

>Asi_EDC2
MCSCCSGCHGTETICYVQPVCCKPVYIQRSSGSCCQPCGSCCGSCCWGSRSCPRVLIQRRPMPVCCPPLQYSSPMQQCCSPLKKC

>Asi_EDC3
MCSGCSGCHGTETICYVQPICCEPVYIQRSSESCCQPCGSCCGSCCGSCCRGSRSCPRVVIQRWPMPVCCPPLQYSAPMQQCCSPLKKC

>Asi_EDC4
MCSCCSGCHGTESVQPICCEPVYIQRSLGSCCQPCGSCCGSCCGGSRPFPRVVIQRRPMPVCCPPLQYSAPMRKYSAPMQQCCPPLKKC

>Asi_EDCH1
MCSRRSCHDHGSSSHGCHRHESSCHGSSSSINCVIEKPVPICPMPQCCPPIQQCCPPMQQCCPPIQQCCPPVKCCQQNQQCCKFPPQYP
K

>Asi_EDCH2
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSINCVIEKPVPICPVPQCCPQLPQCCVPVQPSPMQYCQQSKQCC

>Asi_EDCH3
MCSRRSCHDHGSSSHGCHSHESPCHSSSSSINCVIEKPVPVCPVPQCCPQLPQCCVPVQQCCPPVQCCQQSKQCCKIPPPCPK

>Asi_EDCH4
MCSRRSCHDHGSSSHGCHSHESSCHDSSSSINCVIEEPVPVCPVPQCCPQLPQCCVPVQQCCPPVQCCQQSKQCCKIPPQCPK

>Asi_EDCH5
MGSHGSHHDHGSSSQGCHGHESSCHGSSSSTTCVIEKPVPVCPPQPCCPPVNGGQQSNPCCKFPTQYPK

>Asi_EDCH6
MCSRRSCHDHGSSSHGCHGHESSCHGASSSIPCIIEKPVPVCPVQPCCPPVQQWCPPMQKCCPPVQQCCPPMQKCCPPVQQCCPPMQKC
CPPVKCCQQSKQCCKFPPQCPK

>Asi_EDCH7
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSIPCVIEKPVPVCPVQPCCPPVQQCCPPMQKCCPPVQQCCPPMQKCCPPVQQCCPPRQKC
CPPVKCCQQSKQCCKFPPQCPK

>Asi_EDCH8_partial
MCSRRSCHDHGSSSHGCHXXXSSIPCIIEKPVPVCPVQPCCPPVQQCCPPMQKCCPPVQQCCPPMQKCCPPVQQCCPPMQKCCPPVKCC
EQSKQCCKFPPQCPK

>Asi_EDCH9_partial
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSIHCVIEKTVPVCPVQPCCPPVQQCCPPMQKCCXXXPVQQCCPPMQKCCPPVKCCQQSKQ
CCKFPPQCPK

>Asi_EDCH10_partial
MCSRRSCHGHGSSSHGCHRHESSCHGSSSSVHCVIENXXXVQQCCPPMQKCCPPVKCCQQSKQCCKFPPQCPK

>Asi_EDCH11_partial
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSIPCIIEKPVPVCPVQPCCPPVQQCCPPMQKCCXXX

>Asi_EDCH12
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSINCVIEKPVPICPMPQCCPPVQQCCPPVQKCCPPVQCCQQSQQCCKIPPQFPK

>Asi_EDCH13
MCSHGSCHNRHRSCHGSSSHCHESRPSCNIVVVEKPYCCPVQRYCPPVSCCYPRYQYSQCCKFPQYPKCPPQYPK

>Asi_EDCH14
MCSHGSCHNRHRSCHGSSSHCHESRPSCNIVVVEKPYCCPVQRYCPPVSCCYPGYQYSQCCKFPQYPKCPPQYPK

>Asi_EDCH15
MCSHGSWHNRPESCHNRHRSCHGSSSDCHESRPSCNIIVVEKPYCCPVQRYCPPVSCCYPGYQYSQCCKFPQYPKYPQYPK

>Asi_EDCH16
MCSHGSCHNRHRSCHGSSSHCHESRPSCNIVVVEKPYCCPVQRYCPPVSCCYPGYQYSQCCKFPQYPKYPQYPK

>Asi_EDCH17
MCSRGSCHGSSSHCHEPRPSCNIVVVEKPYVQACCPVPCYCPPMSCCYPRYQYSQCCKFPQYPK

>Asi_EDCH18
MCSRGSCHDHHNTCHRSRRGSHCHESRPSCNITVVERPSMQSWCPVQQYCPVQRYCPPICCYPRYQYSQCCKFPPQYPKCPPQYPK

>Asi_EDCH19
MCSRGSCHDHHSSCHRSRRGSHCHESRPSCNIMVVDRPSMQSWCPVQQYCPVQRYCPPTCCYPHYQYSQCYKFPPQYPKCPPQYPK

>Asi_EDCH20
MRSRGSCHDHHRSCHGSRCGSHCHEARPSCNVVVVERPSVQQYCPVQCYCPPTCCYPRYQYSQYPKCPPQYPK

>Asi_EDCH21_partial
XXXSRRRCHDRETSCHESRPSCHSSGSRCHDSRPSCHGSVIERPPYCQWQQYCPVQPYYPGYQYSQCCKFPQYPKYPQYPK

>Asi_EDCH22
MCSRRRCHERETSCHSSGSRCHESRPSCNFTVVERPSCWQGQQYCPPYYPVQPIYPDYQQCCKYPKIPQGPPCPHSTHSTPSPQSTPSN
SQKNSKERSQDLHVPFLLL

>Asi_EDDM1
MSYPQQHQCKQPCLPPPIVNNCPLQTVEPCGTVHVSQCATRCVDTCGTASVAQAASGSLSPCGSISEAQAMGTSLSPCGGVSVIQAASK
SVDPRGGVSMAQAAGKSVDPCGGVSMAQAASKSVDPCGGVSVAQAASKSVDPHSGVSMAQAAGTSLCPCGGVCVACAASRSVGPYGGVS
VAQAASKSVDPHGGVSMAQAAGKSVDPCGGVSMAQAASKSVDPCGGVSMAQAASISVDPCGGASVAQAASKSVDPCGGVCVACAASRSV
DPSGGVRVSQAASKSVDPFDGVCVAQAASKSVDPCGGVSMAQATGTSLNPCAGVSVAQAATKCIEPGAKGCVRPYAIQCADVCRSKCVA
SYGQVKVDPCAPGCAKTYPLQNVDPCLPKGSPMQQCCRKSKQC

>Asi_EDDM2
MAFPNQQQYKQPCLPPLVCIQKCPPRCMDQCDAACVKKHTDPCGNICAKSCTTKCVDSCDGISTMLCVTKCMDPCGAACMKKCTTKCMC
PSNTVCTKPCATKYVDPCSTSCVTSCVTPCPEPCNTVCVKECVTKCMDPCDTFCAEPYVTKCGDPGCSSSAKPCITKCVDSCNAVCVKE
CTTKCVDLCSTRCAKPCVTICVDLCGTVCAKPCITKCVDSCGTGCGKLCVTKCMDPCGTICAKPCLTMCMNPCSTRCAKPSVTKCVDLC
GTVCGKPCITKHGDPCGTICVKECTTKCMDPCGIICTKPCVTKCVDPCTTSCVTSCVTKSTESCRTVCVKKCTVKCVDTCSTVCAKPFV
PKCTDPCCPRCTASSGTMCVDPCAPVCKKTYPLQSVDLRLSKCPPVQQCCQKPKRCQQPLDHQQLWHEDK

>Asi_EDKM_partial
MSKLLKVITDLIDSNQGNSRKEGKAEMFCRSEFKKLIQXXXSPSYRYRHIKSLPESETEPVNHKKISTVKHCVY

>Asi_EDP1
MSYYGQQCKQRCLPICQDQLPVKCPPMCPQQWTPQYSTKSYSGYDGYCESSSLQCMEPCGPKGWSKCRPQCEQVCVPPCPPSCVTQCIT
KCPPPCITKCPPPCVKCPQQCVPQCVTKCPPPCVKCPQQCVPQCVTKCPPPCVTKCPPPCVKCPPPCVKCPQPCVTQCITKCPPPCVTK
CPQQCVTKCPPPCVTKCPQQCVTKCSPPCVTKCPQQCVTKCPPPCVTKCPQQCVTKCPQQCVTKCPQQCVTKGQSSKVKISSNNKKYCS
ASKWF

>Asi_EDP2
MSSRQNQQQCKQVLTLPPALSKAIPDPAPVLDPEVPEPVPAGPEPCPAPVKEPENAPRRCQEEEHCKQPLGQPPTLAPKLEPEPESKLG
GFLVPEEPEDSVPVQPPPLVEKQQQPSL

>Asi_EDP3
MNLQKQEKQVPVSFKASEKPSPFLFSDLLSSQQQQQQWVPPKCQDKCPPKCVELCKPPKCQDKCPPKCPPQQ

>Asi_EDPCQ
MPWYRQQCKHQPLLGSSHVNKSLWQQDCEYCPLQCPQPRGKQCMQQCLALYPLHQSVCKGSLPGMTVSSQHTCVAECLWPCESERPQPG
CMAEQPLQQHVINSQLCCTRTAQNWHQTANSQPCCTTMTTEPQRATKGKAPCAPSVPNQECVNNRPPCAPSAPNQKCVNNRPPCAPSTP
NQKCVNNCPPCGPPAANHSCCTKSLPCKPESPAEERQEESLLWESVPECLQLQDMTDSQGMTKCPQLLCMMDDLPPCMHTYPALQCKTT
CTPPSSAQCPALQQCLASCPPLQAIQHPPLPRATNCSLPYDATHLPPPQQCWTTCPLTHPSNSTCPLLEQSIATCHLPPCVMEGLLPAT

TKYPRGRQHLHKHLPPSFMTKCFPSGMTAAAQQSCLTKRRRQRILRRPSLRPRMKCPLPQQCFAMRPLPPNATECLLPGTTECPPEQRA
PAHLPQQHIAKARPWGVTGFPKYHRK

>Asi_EDPCV
MSFQHQCKQPCLPPPICGQTVPSQPCVAPCSTVRVDPCPPVCVNPCPPQCVDPCPPGCVKPCPPQCVDPCPPVCVNPCPPQCVEPCPPG
CVNPCPPQCVDPCPPGCVNPCPPQCVDPCPPKCPPSQQCCAQTKLC

>Asi_EDPE
MSSHQMQCKQKTTLPPPLCKGPPNQGQDPVPLPEPVPLPAPIPEPGQGKTPDIKIPECPPQQQQQCKLPPIIPPCPPPCKEPPMPEPMP
FPEPVPCPEKPVPLPAPLPEPGKGKTPDIKIPECPPQQQQQCNEPPIIIPPCPPPCKEPPMPEPMPFPEPVPCPEKPVPLPAPLPDPGQ
GKTPDIKIPECPPQQQQQCKLPPIIPPCPPPCKEPPVPEPMPFPEPVPCPEKPVPLPAPLPAPMPDPGQGKTPDIKIPVCPPQQQQQCK
EPPVVIPPCPPPCKEPPVKCPPPCPPIQQQQQKQPCQWPPQQK

>Asi_EDPL
MSHDQQQIKQPFQPPPESSQLCPPLKFLESCPSAPPKCPEPAPPPPKGPEPSPQPPTTPTC

>Asi_EDPQ1
MSYPNQQQWQQVVRPPPVIPPQKCPPLVFPPQKCPPPQIPPPKCPVPDIPPQKCPPPQWPQQKCPPQK

>Asi_EDPQ2
MSYPNQQQCRQVVYPPPVIPPQKCPPPVCPPQKCPPLVIPPQKCPALQIPPPKCPVPDIPLQKCPPPQWPQQQCPLQK

>Asi_EDPQ3
MSYPNQQQCKQVVCPPPLIPPQKCLPPQCPLPKCPPTKCPPPEWPDQKCPPPKCPDQKCPPPQCPQLQ

>Asi_EDPQ4
MSYPNQQQCRQGVYPPPVIPPQKCPPPVCPPQKCPPLVIPPQKCPALQIPPPKCPVPDIPLQKCPPPQWPQQQCPPQK

>Asi_EDPQ5
MSYPNQQQCKQVVFPPPVIPPQKCPPPQCPLPKCPPSKCPPPEWPDQKCPPPQCPDQKCPPPQCPQLQ

>Asi_EDRYA
MYSPGHYTAFCHGPSGSCLGKGGPYSCGQTFCSGIQHSQGDSRSSCHSSGPLCHGSGPFDQKTWRRQHVRKRRPVWPCQAPVQKCCAPA
QLRYA

>Asi_EDQL
MCSRENRGCHPRESSSCHNSGRSSCHGSGDVICHEVTPLPDIQTVPPMPMPTPLPMPVPAPIPCQQQQTKQPIHWPPQQQHQK

>Asi_EDQM1_partial
MSWQNQQGSGCYRSGGCHSRXXXGGSSGGGCHSSGGCHSSGGYYGGRSFHVISGGHSQGSQQHKQISKVPSQKLK

>Asi_EDQM2_partial
MSQQNQQGSGCHGSGGCHGTRSSSGGCHGGGGCHGGGSTGGGCHGGGXXXGGSSSGGCHSSAGCHGGRFSQVIGGGHSQGSQQCKQISQ
VPSQKLK

>Asi_EDWM
MTCSSGRESYFNLNSTWYDPSGSWLENHRIPLCYADDSCCGGCNPDVRGVGGHNYRPCWYRRSVCSEAERGSSSGYCGSEDSGCARRPT
LGYSDGCGGYRRGPDRCNGECSSHEFGRRPTYHYAADVYLANERLACSEGCHGSSGGFYGSSGGCHRRRRCGEPCHGSGSYGSSRGCHG
RRRSVCGEPCHDSGSSGYLQRVCVKPGPCIPRCPPRQKYVRSTQSCCIPVQTYCAPVQAYCPPVGKYSSGGQQCKQTSKLPILKAK

>Asi_EDYM1
MSYYGYQLKQQCYVPPGVKYSSCVTRCPKPPATKWTTPCATKCTEPCITKKLPEPCATTCVKTRVVRCPLPCTPTRPEPCAAKCLTPCA
TGYLEPCGLQQPQPFSKGWEHQWAPQYIQPCPMRCPPACGPAYMQPPAQKCTAPYYFQWSNRCGYGNCGPC

>Asi_LOR
MSSTQQKTACQEIPHQSGGLQGSSCHGGGSSVSSGNGIGGASSCCGGSSGQNIHVSSSGGSSCCTGGSGYGIGGGSNYGGGSSGQKIGV
TGGSSSGSSYGGGLSSGISRGSAQKVVIADGGSSGSSSCSGGGSSYGIGGGSSSGGVSSGQTIVVGSGSGGSGCGVSGGSSYGTGGGSS
CGGVSSGQATVVGSGSGGSGYGIGGGSGGSGYGIGGGSSYGTGGGSSSGGVSSGHKTIVGSGSGGSGYGIGGGSSYGTGGGSSSGGVSS
GHKTIVGSGSGGSGYGIGGGSSYGTGGGSSSGGVSSGHKTIVGSGSGGSGYGIGGGSSYGTGGGSSSGGVSSGHKTIVGSGSGGSGYGI
GGGSSYGTGGGSSSGGVSSGHKTIVGSGSGGSGYGIGGGSSYGTGGGSSSGGVSSGHKTIVGSGSGGSGYGIGGGSSYGTGGGSSSGGV
SSGHKTIVGSGSGGSGYGIGGGSSYGTGGGSSSGGVSSGHKTIVGSGSGGSGYGIGGGSSYGTGGGSSSGGVSSGHKTIVGSGSGGSGY
GIGGGSSYGTGGGSSSGGVSSGHKTIVGSGSGGSGYGIGGGSSYGTGGGSSSGGVSSGHKTIVGSGSGGSGYGIGGGSSYGTGGGSSSG
GVSSGHKTIVSSSSGGSGYGIGGGSSSGGVSSGHKTIVGSGSGGSGYGIGGGSSYGTGGGSSSGGVSSGHKTIVGSGSGGSGYGIGGGS
GYGIGGGSSCGGGSSGQTISIGSGSGSGSGGSSGISIGGPGIGGGLSSGGSGGSSSKVIINSGGSGGSCSFSGSSSSGSGGSLQSVPQH
QTKQPCQWPPQQK

>Asi_PGLYRP3

MMLRLAVLFSALCASSCQLACPPIISPAKWGSRPAKCAAPLSKVPPGNVIIIHTAGSACHTQPECSELLRNIQVFHRDMKEWCDISYNF
LIGEDGNVYEGRGWLLEGAHTYGYNDLSLGIAFIGNFTERSPNEAAWKTLKNLLAYAVQSGYLASDYLLMAHSDVSNTVSPGKLIRETI
KMWPHYKH

>Asi_S100-A9
MKCPDAQTELEKAIEKIIDIFHQYSVRVGHFDTLTKMEMKLLIDKQLPNYMKNKIKPGEIDALFKDLDKNKDQQLSFGEFMGLISKVTI
ATHEHLHHCGEEEGHHHHQDEHHHH

>Asi_S100-A11
MSKVPVAPTETERCIESLLAVFQRYAGRSDRDETKLSKTEFLAFMNSELASFTKNQKDPAILDRMMKKLDLNCDGQLDFQEFLNLIGGI
AVACHDALCTGGPGGPKGPSCPKGPSGPKKL

>Asi_SCFN
MPHLLDSIGTIINVFYQYVTEDREGSSLSRRQMRLFIQKEFADVLVHLNVNLNFMRKAADNHVNQNSDSQPEEPEPRGDEGRHHQLHEL
EQGEEVRSHSQPRGNDKQTYERSQRPVLEREPQLYEESRHQPRVTEQREEARRRSPTQEPEQPVYEGGRRQPREPEQQEEVRSRFQPRD
TDTQSFEKSRRPALEREPELYEESRRQPCEPEQREKVRSRSQPQEPEPRGDEGRHRQLHELEQQEEVRGRSQSRGNDKQTHERSQHPVL
EREPQLYEESHRQPHVTEQHEEAGSCSQSREAEQQVYDGRGHQLCEPEQQKEVRSRSQPREPKPRGDERRRHLPHDPKQREEVKSHSQP
SDTDTQTYERSQHPGLECEPQLYEESHHQPREPEQRGEVRSRFQPRESEQQVYEGRRHQPRGPEQHEEVRSRSQPHDADTQNFERKQRP
GLECEAQLHEERSRQPREPEQREEVRSRFQPEQPEQQVYGRRGHQPRDPKREEVRSCTQWHETDTQNYERSQHPVLECESQLYERSCRQ
PREPEQQEEVRSRFQPQEPGPRGDEGRQRQALEPEQQVYEGRRRQPLEPEQQKDIRSRFHPSDTNTQSYVRSQRPALERESQFYEESPR
QPHEPEQREKVRSRSQPQEPEPQGDEGRHRQPHQPEQRAEARRQSQPHDNDTQTYERNQHPVLEREPQLYEESHRQPHEPEQREVRSHS
QTQKPEQQAYEGRSHQPREPEQRENVRSRSQPHDTYTQTYERNRRPILERESQLFERSHSQPLETDHREELRSRFQPQEPEPRGLNQNS
ERM

B

>Asi_EDbeta
MACSTNVCNNSSVSCGVAAPQPIADSCNEPCVRQCPDSTVVIYPPPVVLTFPGPILSCFPQESVVGSSASPVLGSSLGGSYGAGYPNGG
SRCGSRYSNGSCGPC

>Asi_Beta1
MTCYPALSSGICASPCGVAVPQPITDSSNEPCVRQCPDSTVVIQPPPVVVTIPGPMLSSFPQEGIVGSTGAPHIAAGFGGGFGSQGFYG
SRAYMGAGGPYGYGGLWGYGGLCGSGGWRSGHRYLNGNCAPC

>Asi_Beta2
MACYPALSSGICASPCGVAVPQPIADSFNEPCVRQCPDSTVVIQPPPSVVTIPGPILSSFPQDSIVGSTGAPHIGAGFGGGFGSRGYAG
SRAYMGAGGPYGYGGLWGYGGLCGSGGWRGGHRYLNGNCSPC

>Asi_Beta3
MACYPALSSGICASPCGVAVPQPIADSFNEPCVRQCPDSTVVIQPPPSVVTIPGPILSSFPQDSIVGSTGAPHIGAGFGGGFGSRGYAG
SRAYMGAGGPYGYGGLRGSGGLCGSGGWRWGHRYLNGNCAPC

>Asi_Beta4
MACYPALSSGICASPCGVAVPQPIADSFNEPCVRQCPDSTVVIQPPPSVVTIPGPILSSFPQDSIVGSTGAPHIGAGFGGGFGSRGYAG
SRAYMGAGGPYGYGGLRGYGGLCGSGGWRGGHRYLNGNCAPC

>Asi_Beta5
MACYPALSSGICASPCGVAVPQPIADSCNEPCVRQCPDSTVVIQPPPCVVTIPGPMLSSFPQDSLVGSTRAPHIGAGFGGGFGSRGYAG
SWGCMGAGGPYGYGGLWVMGASVFLGAGDGATGTSMATVHHAKPWRNGHGTGNNQAAMWIDG

>Asi_Beta6_partial
MACYPALSSGICASPCGVAVPQPIADSFNEPCVRQCPDSTVVIQPPPSVVTIPGPILSSFPQDSIVGSTGAPHIGAGFGGGFGSRGXXX
MGSGGPYGYGGLWGYGGLCGSGGWRGGHRYLNGNCAPC

>Asi_Beta7
MACYPALSSGICASPCGVAVPQPIADSFNEPCVRQCPDSTVVIQPPPSVVTIPGPILSSFPQDSIVGSTGAPHIGAGFGGGFGSRGYAG
SRAYMGAGGPYGYGGLWGYGGLCGSGGWRGGHRYLNGNCSPC

>Asi_Beta8
MSSGIGCSNPCEVNCPQPQAVTDNEPCVITCPDSRVIIYPPPVVVTFPGPILTTCPQESVVASTASADTVPAEMPASVPLTTAVSGSLE
PCAETIAPPIIPRPLSRYVPKYSHTYSSHWMHPCNTNRFGKRWAY

>Asi_Beta9

MSCYDISYPPCGVTLPPCPEPFAVTSNEIYSAQYPDRIVETELEDGQPCTVIYPGPILTTFPQQTLVGSSALFDMERLLGSRRSLEFEG
LLGLGGICGPGSLCNSEPFDDFPYGNCGPV

>Asi_Beta10
MACTDLCYPSSGIACPTPIANSYNDLCVRQCPDSRAVIQPPPVSVIFPGPILSSFPQDSIVGSSGAPVVGGYGSSFGTRFGYSGLEGSL
GYGSSGGYGGPGYEGLGGGYVGGSLGYGGGLCSSGSLYNYGRLYGSGFGYGYCSPFSYRRYNRYRRGSCGPC

>Asi_Beta11
MPCYDECYVPCRETCPSPVADSCNELPCVRQCPDSTTVIQPPPVVVTFPGPILSSCPQDSVVGLLGAPSTGSSAGSLSYGGSFGSGGLY
NYGGLYSSGLSGLGMGHCCPYSRPLNTYRYGRCFPC

>Asi_Beta12
MSCYDECYTPGAVARPRPIADSWNGLCVRQCPASRVVIQPPPAVVTIPGPILSNYPQDSVVGSAGVPAVGYSPRGYLGYGGSEGALVSG
GSGGALVSGGSLGYGSDLGYGGSLGYGVGLGYGGNLGYSGGLCYGDGLGYGSYGGSYGSPYGGLCGSGYSGFGSGYCRPFSYRRYNRSL
SGSCGPC

>Asi_Beta13
MSCYDECYTPGAVARPRPIADSWNGLCVRQGPASRVVIQPPPAVVTIPGPILSNYPQDSVVGSAGVPAVGYSPRGYLGYGGSEGALVSG
GSGGALVSGGSLGYGSDLGYGGSLGYGVGLGYGGNLGYSGGLCYGDGLGYGSYGGSYGSRSLSSYGGLCGSGYSGFGSGYCRPFSYRRY
NRFLSGSCGPC

>Asi_Beta14
MSCTDLCYPSSGIACPRPFADSCNEACIRQCPDSRVVIRPPPVVVTLPGPILSNYPQDSIVGSSGTPLVGHGAAGGTALSGGSTGPGGA
LGYGGIYGGFSGYGGLGGYGGLGGYGGLGGYGGLGGYGGLGGYGGLCGYGGWGGYGGWGGSCYGLGTGFGYGRCYSPAYCSPYYSRRSR
YGVCTPC

>Asi_Beta15
MSCTDLCYPSGGIACPKPYADSCNEACVRQCLDSRVVIRRPPAVVTFRGPILSNFPQDSIVGSAGVPAAGHGAAGGTALSNSIGGAGGF
YGYGASLDSGGLYGYGGSLGYGGLYGYGGSLGYGGLCGYGGLSSDSGSCYSSGYCSPYSYRRYGRYRYGSCGPC

>Asi_Beta16
MSCTDLCYPSSGIACPRPFADSCNEACLRQCPDSRAVIQPPPVVVTFPGPILSNFPQDSIVGSVGVPEVGHGAAGGTALSNVTGSAGGI
YGGLGYGDALGYGGGYGGSFGSRGLYGYGGSLGYGGLYGYGRSLGSGGLCGYGGSLGYGGLCGYGGLSSDSGSCYSSGYCSPYSYRRYG
RYRYGSCGPC

>Asi_Beta17_partial
XXXGDSRAVVYPPPVAITFPGPILASCPQESYVGTSEPQCIGGPYTAGGYLGYRGSVGTGYSYPSYSRQLNRYRYGGCGPC

>Asi_Beta18
MSSYGQLISSRCYNPCEVTCPRPYANAWNEPCVTSCGDSRAVVYPPPVAITFPGPILASCPQESYVGTSEPQCIGGPYTAGGYLGYRGS
VGTGYSYPSYSSQLNRYRYGGCGPC

>Asi_Beta19
MSCYDECYTPGAVARPRPIADSWNGLCVRQCPASRVVIQPPPAVVTIPGPILSNYPQDSVVGSAGVPAVGYSPRGYLGYGGSEGALVSG
GSGGALVSGGSLGYGSDLGYGGSLGYGVGLGYGGNLGYSGGLCYGDGLGYGSYGGSYGSRSLSSYGGLCGSGYSGFGSGYCRPFSYRRY
NRSLSGSCGPC

>Asi_Beta20_partial
MSSYGQLISSRCYNPCEVTCPRPYANAWNEPCVTSCGDSRAVVYPPPVAITFPGPILXXX

>Asi_Beta21_partial
MSSYGQLISSRCYNPCEVTCPRPYADAWNEPCVTSCGDSRAVVYPPPVAITFPGPILASCPQESYVGTSEPQCIGGPYTAGGYLGYRGS
XXX

>Asi_Beta22
MSSYGQLISSRCYNPCEVTCPRPYADAWNEPCVTSCGDSRAVVYPPPVAITFPGPILASCPQESYVGTSEPQCIGGPYTAGGYLGYRGA
GIFLGKKKCQEMGNKIQIRE

>Asi_Beta23
MSLYRQLISSRCSNPCEVTCPQPYADAWNQPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGSSAPISIGSSFGYGGSFTYGGS
LGYGGSYSTGSTYPCYSQRVNRYRYRSCGPCQTQKEFTCTRNTQETKCKIQTQGLADDCEKC

>Asi_Beta24

MSFNRQLLSSRCYNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGTSEPLQIGSSFLSRGSVGSGSS
LGCLSPYYSQRYNKYRYDNCGSC

>Asi_Beta25
MSFNRQLLSSRCYNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGTSEPLQIGSSFVSRGSVGSGSS
LGCLSPYYSQRYSKYCYGNYGSC

>Asi_Beta26
MSFNRQLLSSRCYNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGTSEPLQIGSSFVSRGSVGSGSS
LGCLSPYYSQRYNKYRYGNCGSC

>Asi_Beta27
MSFNRQLLSSRCYNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVVTFPGPILSSCPQESYVGTSEPLQIGSSFVSRGSVGSGSS
LGCLSPYYSQQYNKYRYGNCGSC

>Asi_Beta28
MSFTRQLLSSRCFNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGTSEPLQIGSSFVSRGSVGSGSS
LGCLSPYYSQRYNKYRYDNCGSC

>Asi_Beta29
MSFNRQLLSSRCYNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGTSEPLQIGSSFVSRGSVGSGSS
LGCLSPYYSQRYNKYRYDNCGSC

>Asi_Beta30
MSFNRQLLSSRCYNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGTSEPLQIGSSFVSRGSVGSGSS
LGCLSPYYSQRYNKYRYDNCGSC

>Asi_Beta31
MSFNRQLLSSRCYNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGTSEPLQIGSFFLSRGSVGSGSS
LGCLSPYYSQRYNCGSC

>Asi_Beta32
MSFNHQLLSSRCYNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVVTFPGPILSSCPQESYVGTSEPLQIGSSFVSRGSVGSGSS
LGCLSPYYSQRYNKYRYDNCGSC

>Asi_Beta33
MSFNRQLLSSRCYNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVVTFPGPILSSCPQESYVGTAEPLQIGSSFVSRGSVGSGSS
LGCLSPYYSQRYNKYRYGNCGSC

>Asi_Beta34
MTSTGALCCYPPRPSCDVVCPRPYADAWNEPCVKSCGDSRAVVHPPPVVVTFPGPILASCPQESYVGTSLPEVNGGSFGSGSSFGAGFG
YRGNLGSRGYLGYGGGSLGYGGGSLGYGGGSLGYGGGSLGYGSYLGSGGSLGYEGPFCGLGGPICGYGSYGGSYGSGLSSYGGGYSSPY
YSRRYSKYRYGSCGTF

>Asi_Beta35
MSQSLSSRCLPPCDVTWPRPCADAWNWPCVTSCGDSRAVVYPPPVVVHFPGPILASCPQDSIVGTVLPRQSGDIGPFPYGSGSAYGSGG
SAYGFGVGHGSGSGYGSGSSYGFGGGYGSGSRYGFGSGYGFGSGYGSGSCYGSVSSRRYRKFSSGNCGPC

>Asi_Beta36
MSQSLSSRCLPPCDVTCPRPCADAWNWPCVTSCGDSRAVVYPPPVVVNFPGPILASCPQESIVGTVLPQQSGDIGPFPYGSSSAYGSGS
SYGFGGGYGSGSGYGSGSRYGFGSGYGFGSGYGSGSCYGSVSSRRYHKFSSGSCGPC

>Asi_Beta37
MSQSLSSRCLPPCDVTCPRPCVNAWNWPCVTSCGDSRAVVYPPPVVVNFPGPILVSCPQESIVGTVLPKSTGGIGPFPYGSGGGYGSGS
SYGFGGGYGSGSSYGSGGGFEIGSGYGSGGSYGVGSGYGSGSGYGSGFDCGLGSGYGFGGGYRSGSCYGSVSSRRRRRYSSVSCGPC

>Asi_Beta38
MALSSRCCPSVMCPKPCVDACNWPCVTSCGDSKAVVYAPPVIVHFPGPILTSCPQESIVGTSMPTSIRGGGGPYSSGTFGSGFSSGFGS
GFSSGFGSGFGGGYGLGGGYGLGGGYGMSGGYGLGGGYGSGGCYGFGGGYKSGSYYGSVSSRRRRFSSVGCGPC

>Asi_Beta39
MSFNRQLISSCCWPLCDVMCPQPYADVWNEPCVTSCGDSRAMVYPAPVAITFLGPILSSCPQESYVGTSFPGEICPRYASGGFFGLGPH
MVLGSPIGAVNRALDIACHAKPPKNSHRTKENL

>Asi_Beta40
MGSYGALVSSGCYNPCEVTCPEPCVQACNQPCVTSCGDSRAVVYAPPVIVTFPGPILSTCPQESLVGTVLPYESGRLIPMGGSSYGGGS
SFGSGGFNGGSYGGGFSGGSGGGSYRGGYGGFSGGGSYGGGSYGGSSGGGGFSSGGFIGGGSGYGGLGCGGGSFGGGNFSGGSSGGRFG
SSCGYRRSYTSARSSFGNAGGSRGGSCGPC

>Asi_Beta41
MGSYGPLVSSRCSNPCEVMCPEPCVQACNQPCVTSCGDSRAVVYAPPVIVTFPGPILSTCPQESLVGTVLPYESGRPIPMGGSSYGAGS
SSGSGGFTGGSYGGGVSGGSGSGPYGGGYGGYGGYSGGGGSYGGGSYGGSSGGGGFFSGGFIGGGSGYGGLGCGGGSFGGGSFTGGSSS
GGGYRRSCGYRRSYTSGRSSFGSSGGSRRGSCGPF

>Asi_Beta42
MSCGQMISSRCLPPCEMMCPEPYAAACNYPNTTSFGDSRAVVFAPPVIMTFPGPILASCPQESVVGAAEPYPIGGYPGGPYEGSGGSYG
GSSGIGGSYGASYGHSGGSYGYSGGSYGGSSGIGGSYGTSGRSYGGSSRTGGSSGATYGSSGRSSGISGGARSSGTSYGVSGGSREGSR
GSRGVSGRTGVSSASSGDSYGISGGSHRSGTSSGGSGESPKVYGESHEGPGESEHSTGISGGSYGISGGSRRSGTSSGGSGESPKVYGE
SHEGPGESEHSTGISGGSYGISGGSRRCGTSSGGSGESCKVYGESHEGPGESEHSSGRSGGSYGISGGSRRSGTSPGGFGESHEGSRRN
EDSSESSKGSYVISGESYGSGDPYEGYGGSYGVGGSGIFRSSFFSRCPPGSFSSSFPFTRFAYQRQFGNNEFF

>Asi_Beta43
MSCNTDPCPEGHPSPCEVKCPQPIVTSTNEACVVSCGDSRVIIYPPPVIVTFPGPILSTCPQESLVGAAVPCESGVSQSATTVPLTSEI
GGNSGFSVPLRSEIGGNSGFSVPLRSEIGGNSGFSVPLRSEIMGNSGCSAERLYLNREPEQPSTYTYSFTSQWRHPCNRPGWNRYRSSY
MKKEEPEEEEKPKEWHVGTEESS

Figure 2. Amino acid sequences of proteins encoded by EDC genes of *Alligator sinensis* (Asi). **(A)** Amino acid sequences of EDC proteins other than corneous beta proteins (CBP). **(B)** Amino acid sequences of CBPs, also known as beta-keratins. Cysteine (C) is highlighted in yellow, proline (P) in green, lysine (K) in cyan, glutamine (Q) in grey. Serine (S) and glycine (G) are bolded and in red and orange respectively. Stretches of X´s indicate unknown numbers of amino acid residues, that could not be predicted because of gaps in the corresponding gene sequences.

A

>Cpo_CRNN
MTQLQGNIEGIISAFNAYAKEDGGCITLSKGELRQLIQQEFPDVLVKPHDLQTIDQVLQHLDAESEDRIDFDEFLVLVFQVAKACHKEL
NPCQPSGDGQSSASQGDASRGQAQRAEEDPGRRHIQEPQALEQGQKQPEAPEQDASRPQAPETPTAEGDLSHRHTQDPEVSQGDGNREA
QAAETPEHDSVRRQGQELEQDFSHHRAQEQDPNREGQDPQVPQQDVTHEALESGALEQAPNRHPALQSSTSETDLDRHQGLQSEAPEQD
LSPTAERESTHYEAHEPQAPEQGSQESQPAEQDQNRHQSEEPETLEHTLRHQPQEAQPTKQDLTQETQTPRVPEGDVSRGGTPFSPALQ
QDRGAQQDPREEALTAYRFYIYQCQTPPTFPYQWLPKQ

>Cpo_EDAA1
MSDSLGMLEDLCYQNSSCCWRPHCRRPCYCCCYDPCTGELIWEGWCCCPGWRGSGRYGSCWPC

>Cpo_EDAA2
MSDSLGMLEDLYYQDPSCCWRPRRRRPCYCCCYDPCTGELLWEGWCCCPCWPWRGKYGRRWPC

>Cpo_EDAA3
MSDSLGMLEDLCYQDSSCCWRPRRRRPCYCCCYDPCTGELIWEGWCCCLGWRGSGSYGSCWPC

>Cpo_EDAA4
MLEDLCYQGLDCGWRSRYGRPCYCCCYDPCTEELIWEGWCSCPWWRGRGRYGRCWPC

>Cpo_EDAA5
MFDSLDTLDDLFYQGQSDCWPPTPKRPPYTCCCYDRWGRLVWRGCCWSIPPWWCRTKIPGKGWPC

>Cpo_EDAA6
MFDSLDMIEDLSYQGQSDCFLHQHRRPRYMCCCYDQCGRLVWSGCCWSVPPWWCCQGSPGGCWPCQIQTNYPRHLEQHSYEKQDMDSQL
TGRTRQC

>Cpo_EDAA7
MFPLLPAVFQVCLQHRKVFDSLDTVEDLYYPAQLDCTPPYPGRRYWCCCYDRYGRLVWQGPCCYPGPWCPRRGSSGRGWLC

>Cpo_EDAA8
MFDSLDTIEDLYYPGQLDCTPRYPRRRYWCCSYDRCGRLVWQGPCCYPGPWCPHKGSSGSSWPC

>Cpo_EDAA9
MFDSLDMIKDLSYQGQSDCFSHQHRRPPYTCCCYDQCGRLVWRGCCWSIPPWWCCQGSSGSSWPY

>Cpo_EDAA10
MFDSLDAIEDLCYQGQYDCWDPCYRRPYWYGCWDPCTYRRPYNYGNCYGYGGLYRLGGCYPYSSRWGRKYSYGNCWPC

>Cpo_EDAA11
MSESLDMLENLWYPGQSNCWDPWYRRPYWNSCWDPCTYRRPYIYDNCYGYGGLYGFGRCYPYSTRWGRRGSWGSCWPC

>Cpo_EDAA12
MFDSLDMIEDLPYQGQSDYFPHYCRRPTYMCCCYDWSGVVAAAPYNPGWCCRSSSGNCWLC

>Cpo_EDAA13
MSDSLDMLENLWYPGQSNCWDPWYRRPYWNSCWDPCTYRRPSWSGCWDPCTYRRPYIYNSCYGYGGLYGAGGCYPYTTRWGRRYSAGSC
WPC

>Cpo_EDAA14
MFDSLDMIEDLHFPGQSDCFAPQRRRPPYTCCCYDQWGRLVWRGCCSSIPPWWCYQGSS

>Cpo_EDAA15
MTYHQSGSDDMCYTPCSYGGLYGCQSLTGCWKPWTYRRPHSYGCWNPCTYRWPDSYWDPRGYGSG

>Cpo_EDAA16_partial
MSESLDMLENLWYPGQSNCWDPWYRRPYWNSCWDPCTYRRPYWSGCWDPCTYRRPYIYXXX

>Cpo_EDAA17_partial
XXXGCWDPWYRRSSWSGCWDPWYRRSSWRGCWDPCGYGRPYIYGLGGLSGSGSCYPYYSRWGRRGSYGKCWPC

>Cpo_EDAA18_partial
MFDSLDTTEDLYYKGQSGCWDPWYRRSSWSGCWDPWYRRSSWXXX

>Cpo_EDDM1_partial

MSYPHQHQCKQPCLPPPIVKNCPLQTVEPCGTVHVSQCTTRCVDTCSTASIAQAASRSLSPCGGISEAQAMGTSLSPCGGVSVAQAASR
SVDPCGGVCVAQAASKSVDPCGGVSVAQAVGTSLCPCGGVCVACAASRNVRPCGGVSVAQAASKSVDPCGGVSVAQAAGTXXXCGGVSV
AQAASRSVDPCGGVCVAQAASKSVDLCGGASMAQAKGTSLSPCAGASVAQATTKCIEPGAKGCVRPYAIQCADVCRPKCVASYGQVKVD
PCASGCAKTYPLQNVDPCLPKGSPVLQSKKC

>Cpo_EDDM2
MAFPNQQQCKQPCLPSLVCIQKCPPGCVDQRDAACVKKHTDPCGNICTKSCTTKCVDSCNDISTMLCVTKCVDPCGAACVKECTTKCMC
PSNTVCEKPCVTKYVDPCGTSCVTSCVTPCPEPGNTVCVKECITKCMDPCGTFCAEPYVTKCVDPSCSSSAKLCITKCVDLCNTVCVKE
CTTKCVDPCSTRCAKPCVTNCVDLCGTVCAKPCITKCVDSCCTGCGKLCVTKYMDPCGTICAKPCLTMCMNPCSTRCAKPTVTKCVELC
STVCPKPCIAKHGDPCGTICVKECTTKCMDPHDVICTKPCVTKCVDPCTTSCVTSCVTKCTESCNTVCIKKCTVKCMDTCSTVCAKPFV
PKCMDPCCPRCTASSGTMCMDPCAPVCKKTYPLQSVDPHLPKRPPVQQCCQKPKQC

>Cpo_EDC1
MCSCCSGCHGTGSVQPICYVQPVCCELVYIQRSSGSCCQPCGSCCGSCCGGSRSCPRVVMQRCPMLVCCPPLQYLAPMQQCWLPLKKC

>Cpo_EDCH1
MCSRGSCHDHGSSSHGCHGRESSCHGSSSSINCVIEKPVPVCPVPQCCPQLPQCCVPVQPCCPPVQCCQQSKQCFKCPPQCPK

>Cpo_EDCH2
MCSRGSCHDHRSSSHGCHSHESSCHGSTSSINCVIEKPVPVCPVPQCCPQLPQCCVPTQQCCPPVQQCCPPVQCCQQSKQCCKIPPPCP
K

>Cpo_EDCH3
MCSRRSCHDHGSSSHGCHGHESSCHSSSTSVNCVIEKPVPICPVPQCCPQLPQCCVPTQQCCPPVQCCQQSKQCCKIPPPCPK

>Cpo_EDCH4
MCSRRSCHDHGSSSHGCHGHESSCHSSSTSINCVIEKPVPVCPVPQCCPQLPQCCVPTQQCCPPVQCCQQSKQCCKIPPPCPK

>Cpo_EDCH5
MCSRRSCHDHGSSSHGCHRHESSCHGSSSSLHCVIEKPVPICPVQPCCPPVQQCCPPVQKCCPPVQQCCPPMKCCQQSKQCCKIPPPCP
K

>Cpo_EDCH6_partial
XXXSPVKCCQQTKQCCKFPPPCPK

>Cpo_EDCH7
MCSRGSCHDHHRSCHGSQYGSHCQEFRPSCNITVVERPSVQAWCPVQQFCPVQCYRSPTCCYPRYQYSQCCKFPPQYPKCPAQYPI

>Cpo_EDCH8
MCSRRSCHDRDTSCHRSRPSCHSSGSGCHESRRSCHIPVVERPPYCQWQQYRPVQPYYPGYQYSQCCKFPQYPQYPQYPK

>Cpo_EDCH9_partial
XXXHEPTCSCNVVVVEKPYVQACCPVPCYCPPVSCCYPRYQSSQCCKFPQYPKCSPQYPK

>Cpo_ECH10
MCSHGSCHSHHNSCHGSSSRCHEPRRSCNIVVVEKPYVQACCPVPRYCPVPSYCPPVSCCYPRYQSSQCCKFPQYPK

>Cpo_EDCH11
MCSHGSCHSHRNSCHGSSSRCHEPRCSCNIVVVEKPYVQACCPVPRYCPVPSYCPPVSCCYPRYQSSQCCKFPQYPK

>Cpo_EDCH12
MCSRGSCHSHHNSCHGSSSHCHEPRRCCNVVVVEKPYVQACCPVPCYCPPVSCCYPRYQSSQCCKFPQYPKCPPQYPK

>Cpo_EDCH13
MCSRGSCHDHGSSSHGCHGHESSCHSSSSSVNCIIEKPVPICPMPQCCPPVPQCCPPVQQCCPPVQQCCPPVQCCQQSQQCCKIPPQCP
K

>Cpo_EDCH14_partial
MCSRGSCHDHGSSSHGCHGHESSCHGSSSSLHCVIEKPVPICPVQPCCPPVQQCCPPVPKCCLWQASRGRXXX

>Cpo_EDCH15
MCSRRPCHERETSCHESRPSCHSSGSRCHESRPSCHISVVERPSYCQWQQYRPPYYPDYQQCYKYPVQYPKFPQCPQYPQYPKYPQFPQ
YPK

>Cpo_EDKM
MSKLLKAITDLIDSNQGNSRKECKAEMFCRSEFKKLIQQDLAPVTLSSSYRYRHIKSLAESETEPVNHKEISTAKHCAY

>Cpo_EDP1_partial
MSYYGQQCKQRCLPTPICQDQLPVKCPPMCPQQWTPQYSTKSYSGYDGYCESSSLQCIEPCGPKGWQKCRPQCEQVCLPPCPLPCITKC
PPPCVKCPQQCVTKCPPPXXXQCVTECVTKCPQQCVTKCPQQCVTECVTKCPPPCVTKCPPPCVTKCPQQCVTECVTKCPPPCVTKCPQ
QCVTECVTKCPQQCVTKCPQQCVTKGQSSKVKISSNNKKYCSASQWFW

>Cpo_EDP2
MSSRQNQQQCKQVPTLPPALSKVIPDPAPVLDPAVPEPVPAELEPCAAPVEDLENAPGRHQEEEYCKQPLGQPLILAPKLEPEPESKLG
GFLIPEEPEASAPVQPSPLVEQQQQQQQPPPSL

>Cpo_EDP3
MSSQQQQQQQERVPPKCQDKCPPKCVELCKPPKCQDKCPPKCPSQQQQQQQQQRPKQK

>Cpo_EDPCQ
MPEDALYRQQCKHQPLTGSSHVNKSLRQQVSEYCPLQCPRLRVKKRVALYPLHQSVCGALLPSMTVSSQHMCVAECLWPCETECLQPGC
MAEQPLQQHVTNSQLRCMHTAQNWHQTVNSQPCCTSVTTEPRRATKCKAPCAPSSPNQECENNRPPCAPSAPNQKCVNNRPPCAPSTPN
QKCINNCPPCGPPAANHSCCAKRPPGMPESPAEKGQEESPLWESGPERLQLQDMTENQGMTKCPQLLCMTHDLPPCMHTCPALQCRTTC
PPPSSAKCPALQQCLASCPLLQAIQHPLLPKATSCPLPYDATHLPPPQQCWTTCPLLNSSNSTCPLLEQRITTCHWPPCVTKGMLPATT
KYPRQRQHLRKHLPPSFVTKRLPSGMTAAALQRGVTKWHRQCILKRPSLQPRMKRPLPQQCFAMHPVLPCVSEYLLPGTTECPPEQRAP
AHLPQQRITKAHPPGVTGFPQYHRKRGFEGILGGHI

>Cpo_EDPE
MSSHQMQCKQKTTLPPPLCKEPPNPGQDVVPLPDPVPLPAPMPEPGPGKTPDIKIPECPLPQQQQCKQAPIIPPCPPPCKEPQVPEPVP
FPEPGPCPEKPVPLPAPVPEPGKGKTPDIKIPECPPQQQQQCKQPPVIIPPCPPPCKEPPVQEPMPFPEPGPCPENPVPLPAPVPDPGQ
GKTPDIKIPECPPQQQEQCKQPPVIIPPCPPPCKETPVPVPCPEKPVPLPVPVPDLGQGKTPDIKIPECPPQQQQQCKQPPVIIPPCPP
PCKEPQVPEPMPFPEPGPCPENPVPLPAPVPDPGQGKTPDIKIPECPPQQQEQCKQLPVIPPCPPPCKEPPVKCPPPQQTCPPIQQQQQ
KQPCQWPPQQK

>Cpo_EDPL
MSHDQQQIKQPLQPPPESSKLCPHLKFLESCPSAPPKCPEPAPPPPKGLEPSPQPLLCPPAEKGHMAEPF

>Cpo_EDPQ1
MSYPNQQQCKQVVCPPPLCPPPKCSPPVCPPPKCTPPDWPDQKCPPPKCPPPKCPPPYWPDQKCPPQK

>Cpo_EDPQ2
MSYPNQQQCKQVVCPPPVCPPTKCPPQKCPPPDCPPLKCPPQKCPPPVCPPQKCPPPECPQKCPPQK

>Cpo_EDQL
MCSRENRGCHDTESSSCHDSGRSSCHSSGDVICHEVTPLPDIQTVPPMPLPLPAPAPIPCQQQQTKQPIHWPPQQQHQK

>Cpo_EDQM1
MSQQNQQGSGCHRSGGCHGRGSSGGSYYSGGGCHGGGSSGGSCYGGGGCRGGGSSGGRCYGGGSSGGSCHGGGSYHGGGGSGGRCYGGG
GCYSGGSIGRGCHGGGSSGGGCHSGGGCHSGGGCHSGRSFHVISGGHFQGSRQHKQISHVPSQKLK

>Cpo_EDQM2
MSQQNQQGSGCHGGGGCHSRRSSSGGCHGGGDCHSGGSSGGSCHGGGSSGGGCHGGGGCHSGGGCHGGGSSGGGCHGGGGCHGGGSSGG
GCHGGGGCHGGRSSQVIIRRHSQGSQQCKQISQVLSQKLK

>Cpo_EDRYA
MYSPGHYAAFCHGSSGSCLGKGGPYGCGQAFCSGIQHSQGDASSSCHSSGPLCHGSGPFDHKAWRRQHVRKRRPVWLCQAPVQKCCTPA
QLRYA

>Cpo_EDWM
MICSSGRESYFNLNSTWYDPSGSWLENHRIPLRYADDSCCGGCNPDVRGVGGHNYRPYWYRRSVCSEAERGSSSGYCGSEDSGCARRPT
LGYSEGCDCYRRRPDCCNGECSSHEFGRRPTYHYAADVYLANERLACSDGCHGSSGGFYGSSGGCHRRRRCGEPCHGSGSYGFSRGCHG
RCRSVCGEPCHDSGSSSHLLPVCVKPEPCIPRCPPKQKYVRSTQSCCIPVQSYCTPVKACCPPIQAYCPPVGKYSSGGQQCKQTSKLPT
LKAK

>Cpo_EDYM1
MSYYGYQLKQQCYVPPGVKYSSCVTRCPKPPAMKWTTPCTTKCTEPCIAKKLPEPCATTCVKTRVVRCPLPCTPTCPEPCAAKCVTPCA
TGYLEPCGLQQPQPFSEGWEHQWAPQYIWPCPTRCPSACGPAYMQPPAQKCTAPYYFQWSNRHGYGNCGPC

>Cpo_LOR_partial
MSSSQQKTACQEIPHQSGGLQGSTCHGGGSSVSSGCGIGGGSSYGGGSSGQKIGVTGGSSSGSSYGGGLSSGISGGSAQKVVIAGGSSG
GCSCCSSGGSSYGIGGGSSCGSVTSGQTTVVGSGSGGSSYGSGGGSSCSGVSSGQKTIVGSGSGGSSYGSGGGSSCGGVSSGQTTVVGS
GSGGSSYGIGGGSSCGGVSXXXSYGISGGSSYGSVSSGQTTVVGSGSGGSSYGSGGGSSCGGVSSGQTIVVGSGSGGSSCGVSGGSSYG
TGGGSNCGGVTSGQKTIVDSGSGGSSYGIGGGSSCGGVSSGQTTVVGSGSGGSGYGIGGGSSHGSGSSGQKTIVGSGSRGSSYGISGGS
SYGRGGGSNCGGVSSGHKTIVGSGSAGSGYGIGGGSSYGIGGGSSCGGGSSGQTVSIGSGGSSCGSGGSSGISIGGPGIGGGLSSGGSG

GSLSKVITTSGGSGGSSGGSLQSVPQHQTKQPCQWPPQQK

>Cpo_PGLYRP3
MMLRLAVLFSALCAASCQLACPFIVSPAKWGSRPAKCASPLSKVPPGNVVIIHTAGSACHTQPECSELLRNIQVFHRDMKEWCDISYNF
LIGEDGNVYEGRGWLLEGAHTYGYNDLSLGIAFIGNFTERSPNEAAWKALKNLLAYAVQSGYLASDYLLMAHSDVSNTISPGKLIRETI
KLWPHYKH

>Cpo_S100A9_partial
XXXKCPEAQTELEQSIEKIIDVFHVYSVRVGHFDTLTKMELKLLIQKQLPNYLKNQTSPGQIDALFKDLDKNKDQQLSFGEFMVLITRV
TIATHEHLHHCGEGEGQHQHQDEHHHH

>Cpo_S100A11_partial
XXXSKIPVAPTETERCIESLLAVFQRYAGRSDRDETMLSKTEFLAFMNSELASFTKNQKDPAVLDRMMKKLDLNCDGQLDFQEFLNLIG
GIAVACHEALCKGGPGGPKGPGGPKKL

>Cpo_SCFN
MPRVLDSIGTIIKVFYQYATEDREGSRLSRRQMRLFIQKEFADVLVKPYDPLTIDMVLRLLDQDGDGSIDFSEFLILAFRVAQACYSYL
APKPELQERQQQGRRGKELNEPEAKADRGRGHQLREPEPRVGRRSHSEDPEQEPEPRRDEGRQRQSLEPEQQVYEGRRHQSRDPEWREE
VRSRSQPRDIMQSHERSQRQLLEHEPQLYEESHHRPHERECREEVRSHSQPRDTSTQAYERSRRPLLERESQLDKESHHQPREPEKQEE
AWSRFQSRKPEPQWDEGRQRQPLEPKERQEARGRSQLHDTDTQSYERSQRPVLEPEPQLYEENRHQPHEREQRQEAKSRSQLQEPEQPV
YEGRGQQPHEPEEGKKVKSWSQPCDNDAQNYERGQYSICEREPQLYEESRRQPCNQEQQEAVRSHSQTQESEQVYAGRRHQPLEPEQRE
QVRGRSQPRDTDTKSYERSRCPALKPEPQLYEESRHQPREQEQREEVRSHFQPRQPEPQRDESRQRQPLEPEQQVYEGRRHQPQEPKQR
EEVRGRSQPRNAETQSHERSQCPVFEREPRLYEESRRQPREPEQQEEVRNCSQPCDNDTQTYERSQRPVLECEPRLYKESRRQPREPEQ
REEARTRSQPQEPEQPVYKGGRCQPREPEQREKVRSCSQPQDPEPRGNERRHRQLHEPEQRAEMRSHSQPCGSDKQTYERSQHPVLERE
PQPYKESRRQPRETEQQKEAGSHSQPQEAEQQVYDGRGHQPREPQQREEVRSHSQPRGTNTPTYERSQRPVLEHESQLYKGSHCQPLET
EQREEVRSRFQPREPKLRGEEGRRHLPHDPEQREEVKSHSQPSDIDTQTYERSQHPVLECEPQLYEESHHQPREPEQRIEVRSRFQPQE
SKRQVYEGRYHQPHGPEQQEEVRSRSQPHDADTQNIERNQRPALELEPQLHEKRSHQPREPEQRVEVRSRFQPQAPEQQVYGRRGHQPR
DPKRDEVRSCTQWHETDTQNYERSRHPVLEHESQLYERSHQQPREPEQQEEVRSHFQPQESEPRGNEGRRLALEPEQQVYEGRRHQPLE
SELQKDIRSRFHPSNTYTQSYVRSQHPALERESQFHDESPLQPREPEQREKVRSRSQPQEPEPQGDEGRHRQPHRLEQRAQVRSHSQPH
DTYRQTYEKNRHPILERESQLYERSRHQPLETHHREELRSRFQPQEPEPRGDEGRQRQSLEPEQQVYEERHHQPREPEQREEVRSRSQP
RNAETQSHERSQCPVFEREPRLYEESHRQPREPEQREEVRSRSQPRDNDTQTYERSQRPVLERESRLYKESHHQRRDPEQREEARTRSQ
PQEPEQPVYDGGRHQPREPEQQEEVRSHSQPHKPEHVSGESYCQPRDPEQREVKDCYQSREPEQVYEGSRHQVFEPDQQREVRSCYQPH
ISKPRVNQGGQHQLRVSEERYEGRRHQPHEAETQAYERSRHPFLECEPEEYEGSHIQLRELAQEGVRNRYWPRERGAQTDERRHHELRE
REPQVDEERCRQLCEPEHLGDVRRPYQPYERERQVDGRRLRQSREPEQEHQPSEAEWPEEVRTRYLPREPEYVHEGRRHQVQDTDRQGEV
RSRYQPREPDTRAFERAPRPLFEQKPQLYERSQRQPPEPDQLEKERCHYKPHVPERQVNAESRHQAHRTELQAYERSRDLLHEPESRVD
EQRLRRHQDPEQQVYERSCHQLSEVEQGEVRSHSQPHEHQQVPKKSLHQPWDPEQRGAKDRYQSREPEQVYEGSHHQVCQPNQQGEVRS
CYQPHETEPQAFKGRHYQLLECEPKLHEGRHEVKQVDEGNHLPLEPQQRETRSLYCLYNGQRQPHQLEQRVDDGSRYQHHEPEQLRDGR
SHYQPREPGPRGDARSRRPFHECEPEVHEGSHHQPREQRDVWSHSQPRREPDPRVDERSQPRPRQPELQVYEGSRHQLRDREQGELRDRS
LPLDPETQTYERSRAQAHDPEPQRDGQTREPCDTELQLDEKSLGQPRASIQLGNGRRRYQPRELEPEANEGSCYQPRGPEQQRNSQSHG
QPHEPEQRVKEGSRSHLPEQVQQGDAQPHYVPLETRPQAHDGSRRTLQEPKIQHYEGTRHQSQNPEQKGDVRSHYQPREPTQQVHKGNR
RELRESETQKGEGSQRQPREPKEQVYEGSHRPTPLCDPEQQRAVQSSGQSRDPTPQQDEVNCHQPHEPEPRVVAGNVRQRYEVLPPRDD
QSLLQPDQLVPQRGDRSDDPQPEVKPHDRSLPQSLEPERQSNERTQHQTLEPEHREAEKNGQRPYTPDPKRDGEICHLVSAKQRDDGGR
QPQSHETEPRGGTGTCLQQREAEAQGSTTTSRNPQEPEAQGPERARQPRYPESQGVKSSPHQPQKAAPQEAERDNQPPQNTEPQDGERS
RPQAGEAKPSKEEASQSEPHNPDSTDDNRSRAAPAPPPTGDPGSQTQPREGELRDGSRHQSSRPEGRGDEGSQPQAYEPGCRAVPSQTP
ERGLREGEGSRQQPQALEAPAEGGSRQQPPQGDAASRRQQEVAPEEAAGSHLPREALAQLQEESPTRAAESQEGEQSHHPPEPVVSQEG
LGDPRLDEAKASLPCSPLYVYLLAQKAEQQLCSAPAPQEQP

B

>Cpo_EDbeta
MACSTNVCNNSAVSCGVAAPQPIADSCNEPCVRQCPDSKVLIYPPPVVVTFPGPIITSFPQESVVGSSAAPVCGSSLGGSYGPGHPYGG
SQWGSRYINGSCGPC

>Cpo_Beta1
MACSPALSSGICASPCGVAVPQPIADSYNEPCVRQCPDSRVVIQPPASVVTFPGATLSSFPQDSFVGSAGLPHLGAGFGGSFGSRGFYG
SPGYTGAGGPYGYGGLWGYGGLCGSGGLRWGHRYLNGNCGPC

>Cpo_Beta2
MACPPALSSGICASPCGVAVPQPIADSYNEPCVRQCPDSRVVIQPPASVVTFPGATLSSFPQDSFVGSAGLPHLGAGYGGSFGSRGFYG
SPGYFGAGGPYGYGGLWGYGGLCGSGGWRWGHRYLNGNCAPC

>Cpo_Beta3
MSSVTGCSNPCEVSCPQPQAVTANEPCVITCPDSRVIIYPPPVVVTFPGPILTTCPQESVVASADTVPAELPAAAPLTTEVSGSLEPCA
ETVAPPVIPRRLPRYVPKYSYTYSSQWMHPCNTNRFGKRWAY

>Cpo_Beta4
MSCYDIPYPPCGVTLPPCPEPFAVTSSEIHAVQYPDRIVETELEDGQPCTVIYPGPILTTFPQQTLVGSSALFDMERLLGSRRSFEFEG
PLGLGGICGPGSLCNSEPFGDFPYGNCGPV

>Cpo_Beta5
MACTDLCYPSSDIVCPTPIANSYNDLCVRQCPDSRAVIQPPPVVVTFPGPILNSFPQESIVGSSGAPVVGDYGSSFGARFGYSGLGGSL
GYGSYGGYGGDGYEGLGGGYVGGSLGYGSGSLGYGSGLCGSGSLYNYGRLYGSGFGYGYCSPYSYRRYNRYRRGSCGPC

>Cpo_Beta6
MFCSNECYTPCTVPCPQPTANSCNEPCVSQCPDSTVVIYPPPIVVSFPGPILSSCPQGSIVGFSGPARIGGSGSSGSSLAIRVGYENSG
LSSSRWINRYHLGSCGPC

>Cpo_Beta7
MSCYNECYTPGAVACPRPIADSWNGLCVRQCPASRVVIQPPPAVVTVPGPILSNYPQDSVVGSAGVPAVGYSPRGYLGYGGSEGALVSG
GSGGALVSGGALGYGSDVGYGVGLGYRGGLGYGLGLGYGGSLGYSGGLCYGDGLGYGSYGGCYGSRSLSSYGGLCGSGYSGFGSGYCRP
FSYRRYNRSLSGSCGPC

>Cpo_Beta8
MSCTDLCYPSSGIACPRPFADSCNEACIRQCPDSRAVIQPPPVVVTLPGPILSSFPQDSIVGSAGVPAVGHGAAGGTALSGGPSGPGGH
LGYGGLYGSLGGYGGLGGYGGLGGYGGLGGYGGLGGYGGLGGYGGWGGSCYGLGSGYGYGRCYVPTYCGPYYSRRSYYGICRPC

>Cpo_Beta9
MSCTDLCYPSSGIACPRPFADSCNEACIRQCPDSRAVIQPPPVVVTIPGPILSNFPQDSVVGSAGVPAVGHGAAGGTALSGGPGGAGGI
YGSGGSLGYGGLYGYGRSLGYGGLCGYGGSLGYGGLYGYGGSLGYGGLCGYG

>Cpo_Beta10_partial
MSCTDLCYPSSGIACPRPFADSCNEACIRQCPDSRAVIQPPPVVVTIPGPILSNFPQXXX

>Cpo_Beta11_partial
XXXDAWNEPCVTSCGDSRAVVYPPPVAITFPGPILSSCPQESYVGTSEPLCIGGPYPAGGYLGYRGSVGTGCSYSSYSRQLNTYRYGSC
GPC

>Cpo_Beta12_partial
XXXAVVYPPPVAITFPGPILSSCPQESYVGTSEPLCIGGPYTAGGYLGYRGSVGTGYSYPSYSRQLNRYRYGGCGPC

>Cpo_Beta13
MSSYGQLISSRCYNPCEVTCPRPYADAWNEPCVTSCGDSRAVVYPPPVAITFPGPILSSCPQESYVGTSEPLCIGGLYPAGGYLGYRGS
VGTGCSYPSYSRQLNRYRYGGCGPC

>Cpo_Beta14
MSSYGQLISSRCYNPCEVTCPRPYADAWNEPCVTSCGDSRAVVYPPPVAITFPGPILSSCPQESYVGTSEPLCIGGPYPAGGYLGYRGS
VGTGYSYPSYSRQLNRYRYGGCGPC

>Cpo_Beta15
MSSYGQLISSRCYNPCEVTCPRPYADAWNEPCVTSCGDSRAVVYPPPVAITFPGPILSSCPQESYVGTSEPLCIGGPYPAGGYLGYRGS
VGTGYSYPSYSRQLNRYRYGGCRPC

>Cpo_Beta16
MSLYRQLLSSRCSNPCEVTCPQPYADAWNQPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGSSAPISIGSSFGYGGSFTYGGS
LSYGGSFSTGSTYPCYSQRVKRYRYRSCGPCQTQKEFTCTRNTQETECKIQAQGLADDCEKC

>Cpo_Beta17_partial
XXXSFLSGGSAGSGSSLGCLSPYYSQQYNKYRYGNCGSC

>Cpo_Beta18
MSQSLSSRCLPPCSDICPKPCADAWNWPCVTSCGDSRAVVYPPPVVVHFPGPILASCPQESIVGTVEPRFPSNTGPYYPVGSGSGYGSG
GGFGFGSGYGSGSSYGFGSGYGFGSGYGSGSCYRSSRRYRKFSSGSCGPC

>Cpo_Beta19
MSFNRQLLSSRCFNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVVTFPGPILASCPQESYVGTSEPLQIGGSFLSGGSAGSGSS
LGCLSPYYSQRYNKYRYGNCGSC

>Cpo_Beta20
MSFNRQLLSSRCFNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVVTFPGPILASCPQESYVGTSEPLQIGGSFLSGGSAGSGSS
LGCLSPYYSQRYNKYRYGNCGSC

>Cpo_Beta21

MSTSGALCCYPPQPPCEVTCPRPYADAWNEPCVKSCGDSRAVVHPPPVVVTFPGPILASCPQESYVGTSLPQLSGSLSGSGGFIGSGGG
YGGSLGYRGSLGYGSSLGYGGSLGYGGSLGYGGSLGYGGSLGYGGPFCGLGGPICGYGSYGGSYSSGLSSYGGGYSSPFCSRRYSKYRY
GSCGPC

>Cpo_Beta22
MSCSENPCNDPCSTPCEAKCPKPQGITSNEPCVIACEDTRVIIYPPPVVVTFPGPILTTCPQETLVASTVTLAESSDDVTLAESPAMLP
SVPEVTRRSVPCDEICPPCIIPRPMPCYLPNYSYTFSTQWTHPCNRSGFKKYKSS

>Cpo_Beta23
MSQSLSSRCLPPCSDICFRPCVDAWNWPCVTSCGDSRAVVHPPPVVVHFPGPILASCPQESIVGTVEPRLLSNTGPYYSVGSGSGYGSG
GSYGIGSSYGSGGGYRIGSGYGSVGGYGMGSGYGSGGGYGMGSGYGYGSGYGPGFGCGFGSGYGSGGGYGFGGGYRSGSCYGSVSSRRY
RRYSSGSCGPC

>Cpo_Beta24
MALSSRCCPSVICPKPCVDACNWPCVTSCGDSKAVVYAPPVIVHFPGPILASCPQESIVGTVLPNPMRGGVGPYTSGSFGSGSSYGSGA
SYGSSSGFRSGSGFGFSDGYGLGGGYGSGGGYGSGGGYGFGCGYGSDGGYGFGGGYRSSSCYGSVSSRRRRRYSSAGCGPC

>Cpo_Beta25
MSLQGGQQWELSKVVCLEPWAEAWNKPCITSCGDSGAVVYLTPVAITFLGPILSSCLQDSYVGTSLPEEIYPCYVSGGLLWFSERA

>Cpo_Beta26
MSFNRQLISFRCWPPCNVMCPQPYADAWNEPCITSCSDSRAVVYPPPVAITFPGPILSSCLQESYVGTSLTEEIRPRYASGGCFGSRAS
YGFGSSHWSRQYSSGYC

>Cpo_Beta27_partial
MGSYGPLVSSGCYNPCEVACPEPCVQACNQPCVTSCGDSRAVVYAPPVIVTFPGPILSTCPQESLVGTVLPYESGRPIPMRGSSYGGGS
SFGSGGFTGGXXX

>Cpo_Beta28_partial
MGSYGPLVSSGCYNPCEVACPEPCVQACNQPCVTSCGDSRAVVYAPPVIVTFPGPILSTCPQESLVGTVLPYESGRPIPMRGSSYGGGS
SFGSGGFTGGXXX

>Cpo_Beta29
MGSYGPLVSSGCYNPCEVACPEPCVRACNQPCVTSCGDSRAVVYAPPVIVTFPGPILSTCPQESLVGTVLPYESGRPIPMRGSSYGGGS
SFGSGGFTGGSYGGGFSGGLSSGGGSYGGGYGGLSGGGGSYGGGSYGGSSGGGGFSSGGFGYGGLCGGGGSFGGGSFSGGSSGGGYRRS
CGYRRSYTSGRATFGSSGGSRRGSIGPC

>Cpo_Beta30_partial
MSCGQMLSSRCLPPCEMMCPEPYATACNYPCTTSFGDSKAVVFAPPVIMTFPGPILATCPQESVVGAAEPYPVGGFPGGPYEGSGGSYG
GSYGNYGGSYGGSSGFGGSYGTSRRSYGGSSGTGGSSGATYGSSGHSYGISGGSHSSGTSYGGSGGSRGGSGGIGGSSGSSGGSYGMSG
GSRGSSTSYGGTGGSRGGSGRIGVSSGSSGGSYGMSGGSHGSTTSYGGTGGSRGGSGGIGVSSGSSGGSYGTSGGSHESAXXXGGSGGI
GVSSGSSGGSYGNSGGSHESETSYGGSGGSHEGSGGSHGVSGAIKGSSGSSGCFYGMSGDSYGFGDSCEGSGGSCDVEGSGIFRSSFFS
RYSPGNFRNTFPFTRFSYQRQFGNNEFF

>Cpo_Beta31
MSCNTDHCTEGRPSPCEVKCPQPIVTSTNEACVVSCGDSRVIIYPPPVIVTFPGPILSTCPQESLVGAAVPCESGVPQSATTVPLTSEI
GGSSGFSVPLRSEIGGNSGFSVPLRSEIMGNSGFSAERLYLNREPQQPSTYTYSFTSQWRHPCNRPGWNRYQSSYMKKEEPEEEEKPKE

Figure 3. Amino acid sequences of proteins encoded by EDC genes of *Crocodylus porosus* (Cpo). **(A)** Amino acid sequences of EDC proteins other than corneous beta proteins (CBP). **(B)** Amino acid sequences of CBPs, also known as beta-keratins. Cysteine (C) is highlighted in yellow, proline (P) in green, lysine (K) in cyan, glutamine (Q) in grey. Serine (S) and glycine (G) are bolded and in red and orange respectively. Stretches of X´s indicate unknown numbers of amino acid residues, that could not be predicted because of gaps in the corresponding gene sequences.

Fig. 4. Conserved amino acid sequence motifs of crocodile EDC proteins. Amino acid sequence alignments of motifs present at the amino-terminus **(A)** and carboxy-terminus **(B, C)** of some but not all EDC proteins of the chicken (*Gallus gallus*, Gg), turtle (Chrysemys picta, Cp), snake (*Pyton bivittatus*, Pb) and human (*Homo sapiens*, Hs). The amino acid sequence motifs shown in A and B were discussed in detail in Strasser et al. (2014). The present study shows that these motifs are also conserved in several EDC proteins of crocodilians, represented here by the Chinese alligator (*Alligator sinensis*, Asi). Panel **C** shows a sequence motif at the carboxy-terminus of proteins that are encoded by gene neighbors of the beta-protein gene cluster. In panel **D** a carboxy-terminus which is only found conserved in the archosaurs and turtle branch. Amino acid residues tryptophan (W) in purple and tyrosine (Y) in dark green and phenylalanine (F) in magenta, K and Q (potential transglutamination sites), C (potential disulfide bonding sites), P, G and S are highlighted by specific colors corresponding to those in Figure 2. *, end of the protein.

A

```
Asi EDPE    MSSH--QMQCKQKTTLPP
Asi EDP3    MNL---QKQEKQ---VPV
Asi EDPCV   MSF---QHQCKQPC-LPP
Asi EDYM1   MSYYGYQL--KQQCYVPP
Asi EDP2    MSSRQNQQQCKQVLTLPP
Asi EDDM1   MSY-PQQHQCKQPCLPPP
Asi EDDM2   MAF-PNQQQYKQPCLPPL
Asi EDPQ3   MSY-PNQQQCKQVVCPPP
Gg  EDQCM   MSYY---EQCKQPCL-PP
Gg  EDPE    M-------QCKQEVTLPP
Gg  EDYM1   MSYWY---QYKQQCFIPS
Gg  EDP3    MSSH---QQ-KQQQQIPA
Cp  EDP3    MSSD--QQQCKQTCPPPP
Cp  EDYM1   MSYFAY--QYKQRNYTPY
Cp  EDPCV1  MAY---QQQCKQPCLPPP
Cp  EDPE    MSLHQDQQQCKQGITLPP
Hs  Lor     MSY-----QKKQPTPQPP
Hs  PRR9    MSFS--EQQCKQPCVPPP
Hs  SPRR1A  MNS---QQQ-KQPCTPPP
Hs  SPRR2A  MSY--QQQQCKQPCQPPP
Hs  SPRR2G  MSY--QQQQCKQPCQPPP
Hs  SPRR4   MSS--QQQQRQQQQCPPQ
Pb  EDSPR1  MACPY--QQCKQPCLPPP
Pb  EDSPR2  MS------QCKQACKAPP
Pb  EDCP    MSF-----QCKQACPCPS
Pb  EDP3    MS---QQQQCKQIPCTPP
```

B

```
PGLYRPs  Asi PGLYRP3 PG-KLIRETIKMWPHY--KH-*
         Cp  PGLYRP3 PI----RKVLKTWPHY--KH-*
         Pb  PGLYRP3 PGEFVRAEISK-WPNY--KH-*
         Hs  PGLYRP3 ----ALYNIISTWPHF--KH-*
SFTPs    Asi Crnn    YQCQK-----PPTFPYQWLPKQ*
         Pb  Crnn    QLRQK-----FLHFPPPWSTKQ*
         Gg  Crnn    EQEHL----QPQ-WPP----RK*
         Cp  Crnn    WHSQKPRP-FPHWWPP----KK*
SEDCs    Asi Lor     HQT-KQPCQ----WPP---QQK*
         Asi EDQL    QQT-KQPIH----WPPQQQHQK*
         Asi EDPE    QQQQKQFCQ----WPP---QQK*
         Asi EDPCQ   QHIAKARPWGVTGFP--KYHRK*
         Gg  Lor1    QQT--QPIS----WPPQT-KHK*
         Gg  EDGH    QQI-KQ----SSQWPPS--QKK*
         Gg  EDPE    QQV-KQP----SPWPLT---QK*
         Gg  EDQL    QQI-KQP----VQWPT--QQQK*
         Cp  EDP2    EQQQKQP----HHWPP-K--RK*
         Cp  Lor     QQT-KQPC----QWPPNP--RK*
         Cp  EDQL    QQT-KQPC----QWPPQKHQ-K*
         Pb  EDPKC   DQQKKQPCS----WPP---QNK*
         Pb  EDQL    QQQKKQGCQ----LPP----QK*
         Pb  Lor1    Q-TKQPISIPPCIGPT-----K*
         Hs  Ivl     QQ--KQ----EVQWPP-K--HK*
         Hs  Lor     QQ--KQ----APTWPS-----K*
```

C

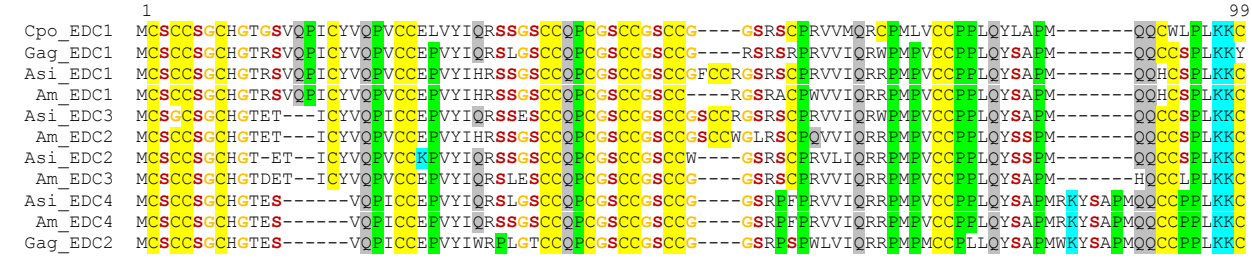
```
Asi EDP1    NNKK-YCSAS-KW-F*
Gg  EDQrep  HAKK-YCSAS-KWPW*
Gg  EDYM2   HSKKSRC-AS-KWLW*
Cp  EDP1    HCKK-YCSAP-KWPW*
Pb  EDPSQ   GQK--YCSASNNWPW*
Pb  EDYM2   TGKK-YCSTT-KWPF*
```

D

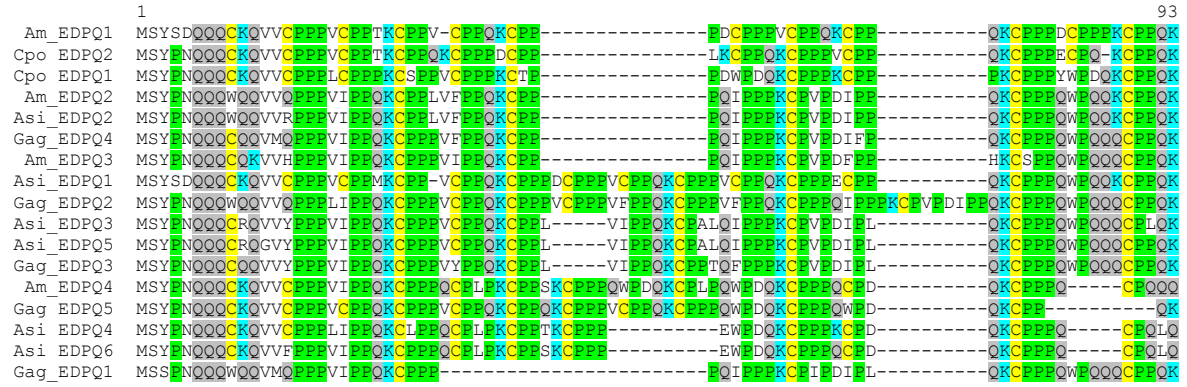
```
Asi EDQM1   QQ-CKQISQVPSQKLK*
Asi EDWM    QQ-CKQTSKLPILKAK*
Gg  EDWM    QQVCK----VPARKIK*
Gg  EDQM1   QQQ--QVHQLPSQKMK*
Cp  EDQM2   QKICK----VPCQKLK*
Cp  EDQM1   QKHCCQ---VPSQKLK*
```

Fig. 5. Conservation and amino acid sequence characteristics of crocodilian EDC clusters. Amino acid sequence alignments of proteins encoded by the EDC cluster **(A)** by the EDPQ cluster **(B)** some but not all proteins encoded by the EDCH cluster **(C)** some but not all proteins encoded by the EDAA cluster **(D)**. Asi, *Alligator sinensis* (Chinese alligator), Am, *Alligator mississippiensis* (American alligator), Cpo, *Crocodylus porosus* (Australian saltwater crocodile) and Gag, *Gavialis gangeticus* (Indian gharial). Coloring as defined in Fig. 2.

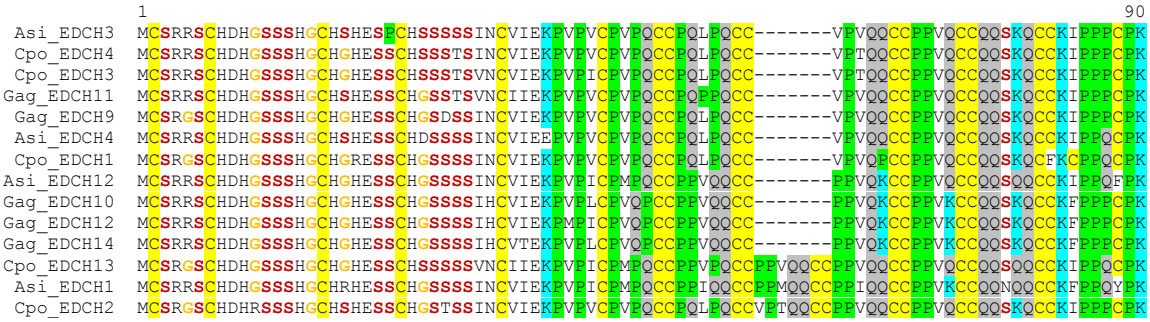
A


1
99
Cpo_EDC1
Gag_EDC1
Asi_EDC1
Am_EDC1
Asi_EDC3
Am_EDC2
Asi_EDC2
Am_EDC3
Asi_EDC4
Am_EDC4
Gag_EDC2

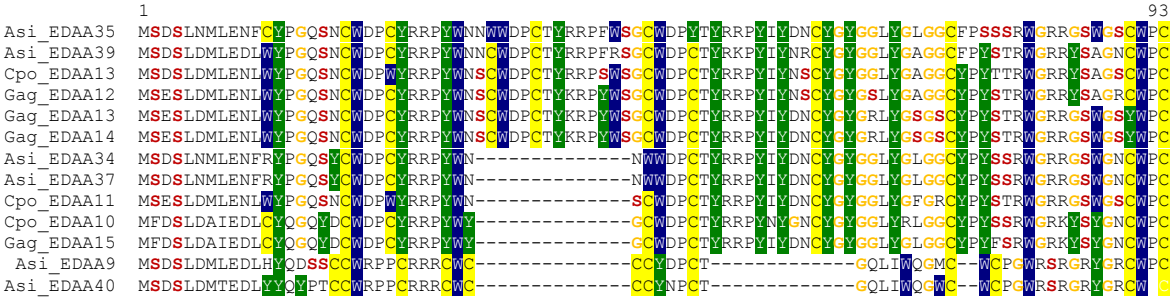
B


1
93
Am_EDPQ1
Cpo EDPQ2
Cpo EDPQ1
Am_EDPQ2
Asi_EDPQ2
Gag_EDPQ4
Am_EDPQ3
Asi_EDPQ1
Gag_EDPQ2
Asi_EDPQ3
Asi_EDPQ5
Gag_EDPQ3
Am_EDPQ4
Gag EDPQ5
Asi EDPQ4
Asi EDPQ6
Gag_EDPQ1

C


1
90
Asi_EDCH3
Cpo_EDCH4
Cpo_EDCH3
Gag_EDCH11
Gag_EDCH9
Asi_EDCH4
Cpo_EDCH1
Asi_EDCH12
Gag_EDCH10
Gag_EDCH12
Gag_EDCH14
Cpo_EDCH13
Asi_EDCH1
Cpo_EDCH2

D


1
93
Asi_EDAA35
Asi_EDAA39
Cpo_EDAA13
Gag_EDAA12
Gag_EDAA13
Gag_EDAA14
Asi_EDAA34
Asi_EDAA37
Cpo_EDAA11
Cpo_EDAA10
Gag_EDAA15
Asi_EDAA9
Asi_EDAA40

Chapter 3 Conclusive remarks

3. Conclusive remarks

This PhD thesis reports the characterization of the epidermal differentiation complex (EDC) of the sauropsid clades Testudines, Serpentes, and Crocodylia and provides a catalogue of the protein products of these gene clusters. The comparative analysis performed on the here newly identified EDCs and previously identified ones (Strasser et al., 2014; Vanhoutteghem et al., 2008; Mischke et al., 1996) has revealed a common organization of the EDC in all amniotes, suggesting a shared ancestry and a similar genetic control of the epidermal barrier in fully terrestrial vertebrates. As an adaptation to terrestrial life, the epidermal barrier was strengthened by a cornified layer in which many of the proteins coded by the EDC have a role, mainly by providing structural components for the cornified cell envelope of keratinocytes (Eckhart et al., 2013; Henry et al., 2012; Kalanin et al., 2002; Kypriotou et al., 2012; Candi et al., 2005). Common functional requirements of the amniote cornified layer can be deduced from the conservation in nearly all amniotes of proteins such as loricrin, PGLYRP3, cornulin and SPRR-like proteins (Strasser et al., 2014; Vanhoutteghem et al., 2008) and from the presence of conserved sequence motifs (Strasser et al., 2014, **Paper I & IV**) involved in crosslinking of protein components by transglutaminases (Steinert & Marekov, 1995; Steinert et al., 1999; Candi et al., 1998, 1999; Rice et al., 1977). Some cysteine-rich EDC proteins found in birds (Strasser et al., 2015) and squamates (Strasser et al., 2015; **Paper IV**) share cysteine repeat sequences with some ultrahigh/high sulfur keratin associated proteins (KRTAPs) of mammalian appendages (Powell & Rogers, 1986, 1994; Gillespie,1991). Since the KRTAPs are not found on the EDC and exhibit a different gene structure, this homology is a likely case of convergent evolution (Strasser et al., 2015).

Due to the fact that sauropsids possess corneous beta proteins (CBPs or beta-keratins) which have been considered a special type of hard keratin, their cornification process was presumed different from that in other amniotes. By now it has become clear that CBPs are biochemically very different from true keratins, which belong to the intermediate filament proteins (Alibardi et al.,

2009; Alibardi, 2016a; Calvaresi et al., 2016). CBPs are structural proteins that likely bind to keratin intermediate filaments (KIFs) forming the corneous material component deposited on and maybe even replacing the KIF network. CBPs distinguish themselves from mammalian keratin-associated proteins though, for being capable of forming not only the interfilamentous (matrix) component but also the filamentous component in the epidermis (Fraser & Parry, 1996, 2017). In all amniotes the cornification process seems to imply the same basic mechanism, in which KIFs compose the cytoskeleton to which structural (matrix) proteins start to bind, accumulating and forming in the end a resistant, insoluble and amorphous (or cornified) structure. In fact, in sauropsids like other amniotes KIFs have been identified (Greenwold et al., 2014; Eckhart et al., 2008; Hallahan et al., 2009;Vandebergh & Bossuyt, 2012) and localized in the epidermis in co-localization with proteins involved in cornification like CBPs (Alibardi, 2013, 2015; **Paper V**). Many of the structural proteins involved in cornification of the epidermis and epidermal appendages are encoded on the EDC in sauropsids. A huge difference with mammals exists for the cornification of the mammalian skin appendages, since most of the interfilamentous (matrix) forming proteins involved (KRTAPs) are not encoded on the EDC.

Furthermore several in detail studies were performed looking into specific aspects of the epidermal barrier. One of these has investigated the defense mechanism exerted by skin specific reptile antimicrobial peptides. Their bactericidal effect was demonstrated by growth inhibition in bacterial cultures and evaluation of ultrastructural damage under electron microscopy.

In the paper on the chicken protein EDMTFH its localization during embryogenesis and feather morphogenesis was analyzed to elucidate on the role of this protein. Its co-localization in embryonic subperiderm and feather follicle correlates the layered organization of the epidermis during embryogenesis to the one of feather follicles during feather morphogenesis. Therefore obtained results supported the model (Sawyer et al., 2005; Sawyer & Knapp, 2003) in which the cyclical growth and shedding of feathers is a modified replication of a series of steps in embryonic

skin development.

Another aspect analyzed by electrophoretic and immunogold labelling methods was the covalent disulfide binding of corneous beta proteins (CBPs) in the epidermis of squamates. The experimental data showed ultrastructurally co-localization of keratin intermediate filaments (KIFs) and CBPs in the epidermis. Furthermore reduction/alkylation and oxidation of CBPs altered the pattern of marked bands in western blots, in particular CBP monomer bands appeared or disappeared after treatments and modifications in the KIF size range were observed as well. Observations fit to the resilient character of the cornified layers of squamate epidermis (Maderson et al., 1998; Klein et al., 2010; Klein & Gorb, 2012) which presumes the formation of very stable chemical bonds (Banjerjee & Mittal, 1978; Mittal & Sing, 1987a-b; Alibardi, 2001). Results allude to the likely presence of disulfide bonds in at least some CBPs and to possible interaction with KIFs.

During this study evidence for clade-specific adaptations in the epidermal barrier of sauropsids has been brought forward as well as conservation of proteins coded on the EDC and involved in the cornification process of keratinocytes in sauropsids and in some cases in all amniotes. A series of different experimental methods were used to shed light on specific properties of reptilian skin defense mechanisms, the role of the avian structural protein EDMTFH and the presumed disulfide binding of CBPs in sauropsids. Beyond the characterization of several specific aspects of the epidermal proteins, this study has led to the identification of a great number of new genes that merit further investigation as likely determinants of the vital skin barrier of sauropsids.

References cited

Alexander N.J. 1970. Comparison of alpha and beta keratin in reptiles. Zellforsch Mikrosk Anat. 110:153–65.

Alibardi L. & Maderson P.F. 2003. Observations on the histochemistry and ultrastructure of the epidermis of the tuatara, *Sphenodon punctatus* (Sphenodontida, Lepidosauria, Reptilia): a contribution to an understanding of the lepidosaurian epidermal generation and the evolutionary origin of the squamate shedding complex. J Morphol. 256(2):111–33.

Alibardi L. & Minelli D. 2016. Sites of cell proliferation during scute morphogenesis in turtle and alligator are different from those of lepidosaurian scales. Acta Zoologica (Stockholm) 97:127–141.

Alibardi L. & Sawyer R.H. 2002. Immunocytochemical analysis of beta keratins in the epidermis of chelonians, lepidosaurians, and archosaurians. J Exp Zool. 15;293(1):27–38.

Alibardi L. & Sawyer R.H. 2006. Cell structure of developing downfeathers in the zebrafinch with emphasis on barb ridge morphogenesis. J Anat. 208:621–42.

Alibardi L. & Thompson M.B. 1999. Epidermal differentiation during carapace and plastron formation in the embryonic turtle *Emydura macquarii*. J Anat. 194:531–545.

Alibardi L. & Thompson M.B. 2000. Scale morphogenesis and ulstrastructure of the dermis during embryonic development in the alligator (*Alligator mississippiensis*, Crocodilia, Reptilia). Acta Zoologica (Stockholm). 81:325–338.

Alibardi L. & Thompson M.B. 2001. Fine structure of the developing epidermis in the embryo of the American alligator (*Alligator mississippiensis*, Crocodilia, Reptilia). J. Anat. 198:265–282.

Alibardi L. & Thompson M.B. 2002. Keratinization and ultrastructure of the epidermis of late embryonic stages in the alligator (*Alligator mississippiensis*). J. Anat. 201:71–84.

Alibardi L. & Toni M. 2004. Immunolocalization of transglutaminase and cornification markers proteins in the epidermis of vertebrates suggests common processes of cornification across species. J. Exp. Zool. 302:526–549.

Alibardi L. & Toni M. 2006. Cytochemical, biochemical and molecular aspects of the process of keratinization in the epidermis of reptilian scales. Prog. Histochem. Cytochem. 40:173–134.

Alibardi L. & Toni M. 2007. Beta keratins of reptilian epidermis share a conserved common epitope termed the core-box. Res. J. Biol. Sci. 2:329–339.

Alibardi L. & Toni M. 2008. Cytochemical and molecular characteristics of the process of cornification during feathermorphogenesis. Progress Histoch Cytoch. 43:1–72.

Alibardi L. & Toni M. 2009. Immunocytochemistry and protein analysis suggest that reptilian claws contain small high cysteine-glycine proteins. Tissue Cell. 41:180–192.

Alibardi L. 1997. Morphogenesis of the digital pad lamellae in the embryo of the lizard Anolis lineatopus. J Zool. Lond. 243:47–55.

Alibardi L. 2001. Keratohyalin-like granules in lizard epidermis: evidence from cytochemical, autoradiographic and microanalytic studies. J. Morphol. 248:64–79.

Alibardi L. 2002. Keratinization and lipogenesis in epidermal derivatives of the zebrafinch *Taeniopygia guttata castanotis* (Aves, Passeriformes, Ploceidae) during embryonic development. J Morph 251:294–308.

Alibardi L. 2003a. Adaptation to the land: the skin of reptiles in comparison to that of amphibians and endotherm amniotes. J Exp Zool B Mol Dev Evol. 298:12–41.

Alibardi L. 2003b. Immunocytochemistry and keratinization in the epidermis of crocodilians. Zool Stud. 42:346–56.

Alibardi L. 2004. Dermo-epidermal interactions in reptilian scales: speculations on the evolution of scales, feathers and hairs. J Exp Zool. 302B:365–383.

Alibardi L. 2005a. Keratinization in crocodilian scales and avian epidermis: evolutionary implications for the origin of avian apteric epidermis. Belg J Zool. 135:9–20.

Alibardi L. 2005b. Proliferation in the epidermis of chelonians and growth of the horny scutes. J. Morphol. 265:52–69.

Alibardi L. 2006. Structural and immunocytochemical characterization of keratinization in vertebrate epidermis and epidermal derivatives. Int Rev Cytol. 253:177–259.

Alibardi L. 2012. Immunolocalization of keratin associated beta-proteins (beta-keratins) in the regenerating lizard epidermis indicates a new process for the differentiation of the epidermis in lepidosaurians. J. Morphol. 273:1272–1279.

Alibardi L. 2013a. Cornification in reptilian epidermis occurs through the deposition of keratin associated beta proteins (beta-keratins) onto a scaffold of intermediate filament keratins. J. Morphol. 274:175–193.

Alibardi L. 2013b. Immunocytochemistry suggests that a prevalence of a sub-type of betaproteins determines the hardness in the epidermis of the hard-shelled turtle. J. Exp. Zool. 322:54–63.

Alibardi L. 2013c. Ultrastructural immunolocalization of alpha-keratins and associated beta-proteins (beta-keratins) suggests a new interpretation on the process of hard and soft cornification in turtle epidermis. Micron. 52–53:8–15.

Alibardi L. 2013d. Immunolocalization of Keratin-associated beta-proteins (beta-keratins) in pad lamellae of geckos suggest that glycine-cysteine-rich proteins contribute to their flexibility and adhesiveness. J. Exp. Zool. 319:166–178.

Alibardi L. 2013e. Immunolocalization of specific keratin associated beta-proteins (beta-keratins) in the adhesive setae of Gekko gecko. Tissue Cell 45:231–240.

Alibardi L. 2014a. Formation of adherens and communicating junctions coordinate the differentiation of the shedding-layer and beta-epidermal generation in regenerating lizard epidermis. J. Morphol. 275:693–702.

Alibardi L. 2014b. Immunolocalization of beta-proteins and alpha-keratins in the epidermis of the soft-shelled turtle explains the lack of formation of hard corneous material. Acta Zool. 96:218–224.

Alibardi L. 2014c. Comparative immunolocalization of keratin-associated beta-proteins (beta-keratins) support a new interpretation for the cyclical process of keratinocytes differentiation in lizard epidermis. Acta Zool. 95:32–43.

Alibardi L. 2014d. Immunocytochemistry indicates that glycine-rich beta-proteins are present in the beta-layer while cysteine-rich beta-proteins are present in beta- and alpha-layers of snake epidermis. Acta Zool. 95:330–340.

Alibardi L. 2014e. Presence of a glycine-cysteine-rich beta-protein in the oberhautchen layer of snake epidermis marks the formation of the shedding layer. Protoplasma. 251:1511–1520.

Alibardi L. 2014f. Immunogold labeling shows that glycine-cysteine-rich beta-proteins are deposited in the oberhautchen layer of snake epidermis in preparation to shedding. J. Morph. 276:144–151.

Alibardi L. 2014g. Immunocytochemistry suggests that the prevalence of a sub-type of beta-proteins determines the hardness in the epidermis of the hard-shelled turtle. J Exp Zool B Mol Dev Evol. 322:54–63.

Alibardi L. 2015a. Review: mapping epidermal beta-proteins distribution in the lizard Anolis carolinensis shows a specific localization for the formation of scales, pads and claws. Protoplasma. doi:10.1007/s00709-015-0909-z.

Alibardi L. 2015b. The corneous layer of the claw in the lizard *Anolis carolinensis* mainly contains the glycine–cysteine-rich beta-protein HgGC3 in addition to hard keratins. Tissue Cell. 46:326–333.

Alibardi L. 2016a. The process of cornification evolved from the initial keratinization in the epidermis and epidermal derivatives of vertebrates: A new synthesis and the case of sauropsids. In: Kwang Jeon, editors: International Review of Cell and Molecular Biology. Vol 327, Chennai: Academic Press:263–320.

Alibardi L. 2016b. Review: cornification, morphogenesis and evolution of feathers. Protoplasma. Sep 10. [Epub ahead of print]

Alibardi L. 2016c. Microscopic and immunohistochemical study on the cornification of the developing beak in the turtle *Emydura macquarii*. J Morphol. 277(10):1309–19.

Alibardi L. Dalla Valle L. Nardi A. Toni M., 2009. Evolution of hard proteins in the sauropsid integument in relation to the cornification of skin derivatives in amniotes. J. Anat. 214:560–586.

Alibardi L.& Dipietrangelo L. 2005. Differentiation of the epidermis of scutes in embryos and juveniles of the tortoise *Testudo hermanni* with emphasis on beta-keratinization. Acta Zoologica. 86:206–216.

Alibardi L., Jaeger K., Dalla Valle L., Echkart L. 2011. Ultrastructural localization of hair keratin homologs in the claw of the lizard *Anolis carolinensis*. J. Morphol. 272:363–370.

Alibardi L., Knapp L.W., Sawyer R.H. 2006. Beta-keratin localization in developing alligator scales and feathers in relation to the development and evolution of feathers. J. Subm. Cytol. Pathol. 38:175–192.

Alibardi L., Segalla A., Dalla Valle L. 2012. Distribution of specific keratin associated beta-proteins (beta-keratins) in the epidermis of the lizard *Anolis carolinensis* helps to clarify the process of cornification in lepidosaurians. J. Exp. Zool 318B:388–403.

Alibardi L., Spisni E., Toni M. 2004. Differentiation of the epidermis in turtle: an immunocytochemical, autoradiographic and electrophoretic analysis. Acta Histochem. 106:379–395.

Alibardi L., Strasser B., Eckhart L. 2015. Immunolocalization of loricrin in the maturing α-layer of normal and regenerating epidermis of the lizard *Anolis carolinensis*. J Exp Zool B Mol Dev Evol. 324(2):159–67.

Alibardi L., Toni M., Dalla Valle L. 2007. Hard cornification in reptilian epidermis in comparison to cornification in mammalian epidermis. Exp Dermatol. 16:961–976.

Allam A.A. & Abo-Eleneen R.E. 2012. Scales microstructure of snakes from the Egyptian area. Zool Sci. 29:770–775.

Arnold E.N. 2002. History and function of scale microornamentation in lacertid lizards. J Morphol. 252:145–69.

Artimo P., Jonnalagedda M., Arnold K., Baratin D., Csardi G., de Castro E., Duvaud S., Flegel V., Fortier A., Gasteiger E. et al. 2012. ExPASy: SIB bioinformatics resource portal, Nucleic Acids Res, 40(W1):W597–W603.

Astbury W.T. & Marwick T.C. 1932. X-ray interpretation of the molecular structure of feather keratin. Nature (London) 130:309–310.

Backendorf C. & Hohl D. 1992. A common origin for cornified envelope proteins? Nat. Genet 2:91.

Baden H., Sviokla S., Roth I. 1974. The structural protein of reptilian scales. J. Exp. Zool. 187:287–294.

Baden H.P. & Maderson P.F. 1970. Morphological and biophysical identification of fibrous proteins in the amniote epidermis. J. Exp. Zool. 174:225–232.

Banjeree T.K. & Mittal A.K. 1978. Histochemistry of the epidermis of the Chequered water snake *Natrix piscator* (Colubridae, Squamata). J. Zool. (Lond.) 185:415–435.

Barrowclough G.F., Cracraft J., Klicka J., Zink R.M. 2016. How Many Kinds of Birds Are There and Why Does It Matter? PLOS ONE. 11 (11).

Benato F., Dalla Valle L., Skobo T., Alibardi L. 2013. Biomolecular identification of beta-defensin-like peptides from the skin of the soft-shelled turtle *Apalone spinifera*. J Exp Zool B Mol Dev Evol. 320:210–217.

Boman H.G., Agerberth B., Boman A. 1993. Mechanisms of action on *Escherichia coli* of cecropin P1 and PR-39, two antibacterial peptides from pig intestine. Infect Immun. 61:2978–2984.

Bostwick K.S. & Prum R.O. 2003. High-speed video analysis of wing-snapping in two manakin clades (Pipridae: Aves). J Exp Biol. 206(Pt 20):3693–706.

Brazaitis P. 1987. The identification of crocodilian skins and products. In Wildlife Management: Crocodiles and Alligators (Web G.J.W., Manolis S.C.,Whitehead P.J. eds). Surrey Beatty & Sons Pty, Chipping Norton, New South Wales. Pp. 373–386.

Brochu C.A. 2003. Phylogenetic approaches toward crocodylian history. Annual Review of Earth and Planetary Sciences. 31:357.

Brogden K.A. 2005. Antimicrobial peptides: Pore formers or metabolic inhibitors in bacteria? Nat. Rev. 3:238–250.

Brush A.H. 1993. The origin of feather: a novel approach. In: Farner, D., Kling, J., Parker, K. (Eds.), Avian Biology. Academic Press, New York. Pp. 121–162.

Burke A.C. 1989. Development of the turtle carapace: implications for the evolution of a novel bauplan. J Morphol. 199:363–378.

Calvaresi M., Eckhart L., Alibardi L. 2016. The molecular organization of the beta-sheet region in corneous beta-proteins (beta-keratins) of sauropsids explains its stability and polymerization into filaments. J. Struct. Biol. 194:282–91.

Candi E., Schmidt R., Melino G. 2005. The cornified envelope: a model of cell death in the skin. Nat Rev Mol Cell Biol. 6:328–340.

Candi E., Tarcsa E., Digiovanna J.J., Compton J.G., Elias P.M., Marekov L.N., Steinert P.M. 1998. A highly conserved lysine residue on the head domain of type II keratins is essential for the attachment of keratin intermediate filaments to the cornified cell envelope through isopeptide crosslinking by transglutaminases. Proc. Natl Acad. Sci. USA. 95:2067–2072.

Candi E., Tarcsa E., Idler W.W., Kartasova T., Marekov L.N., Steinert P.M. 1999. Transglutaminase cross-linking properties of the small proline-rich 1 family of cornified cell envelope proteins. Integration with loricrin. J Biol Chem. 274:7226–7237.

Castoe T.A., & Pollock D.D. 2013. Chinese alligator genome illustrates molecular adaptations. Cell Research. 23:1254–1255.

Chang C., Wu P., Baker R.E., Maini P.K., Alibardi L., Chuong C-M. 2009. Reptile scale paradigm. Evo-devo, pattern formation and regeneration. Int. J. Dev. Biol. 53:813–826.

Chattopadhyay S., Sinha N. K., Banerjee S., Roy D., Chattopadhyay D., Roy S. 2006. Small cationic protein from a marine turtle has beta-defensin-like fold and antibacterial and antiviral activity. Proteins 64:524–31.

Cheng P.N., Pham J.D., Nowick J.S. 2013. The supramolecular chemistry of β-sheets. J. Am. Chem. Soc. 135:5477–5492.

Chiari Y, Cahais V, Galtier N, Delsuc F. 2012. Phylogenomic analyses support the position of turtles as sister group of birds and crocodiles. BMC Biol. 10:65.

Chuong C-M. & Widelitz R.B. 1999. Feather morphogenesis: a model of the formation of epithelial appendages. In: Chuong CM (ed) Molecular basis of epithelial appendage morphogenesis. Landes Bioscience, Georgetown, Texas, USA. pp 57–73.

Chuong C-M. 1998. Molecular Basis of Epithelial Appendage Morphogenesis. Austin, TX: Landes Bioscience.

Chuong C-M., Wu P., Zhang F.C., Xu X.,Yu M., Widelitz R.B., Jiang T.X., Hou L. 2003. Adaptation to the sky: defining the feather with integument fossils from Mesozoic china and experimental evidence from molecular laboratories. J Exp Zool. 298B:42–56.

Clark C.J. & Feo T.J. 2008. The Anna's hummingbird chirps with its tail: a new mechanism of sonation in birds. Proc Biol Sci. 22;275(1637):955–62.

Close M. & Cundall D. 2014. Snake lower jaw skin: extension and recovery of a hyperextensible keratinized integument. J. Exp. Zool. A. Ecol. Genet. Physiol. 321:78–97.

Corpet F. 1988. Multiple sequence alignment with hierarchical clustering. Nucleic Acids Res. 16:10881–10890.

Dalla Valle L., Michieli F., Benato F., Skobo T., Alibardi L. 2013a. Molecular characterization of alpha-keratins in comparison to associated beta proteins in soft-shelled and hard-shelled turtles produced during the process of epidermal differentiation. J Exp Zool B Mol Dev Evol. 320:428–441.

Dalla Valle L., Benato F., Paccanaro M.C., Alibardi L. 2013b. Bioinformatic and molecular characterization of cathelicidin-like peptides isolated from the green lizard Anolis carolinensis. Ital. J. Zool. 80:177–186.

Dalla Valle L., Benato F., Maistro S., Quinzani S., Alibardi L. 2012. Bioinformatic and molecular characterization of beta-defensins-like peptides isolated from the green lizard Anolis carolinensis. Dev Comp Immunol. 36:222–229.

Dalla Valle L., Nardi A., Alibardi L. 2009a. Isolation of a new class of cysteine-glycine-proline rich beta-proteins (beta-keratins) and their expression in snake epidermis. J. Anat. 216:356–367.

Dalla Valle L., Nardi A., Toni M., Emeera D., Alibardi L. 2009b. Beta-keratins of turtle shell are glycine-proline-tyrosine-rich proteins similar to those of crocodilians and birds. J. Anat. 214:284–300.

Dalla Valle L., Nardi A., Gelmi C., Toni M., Emeera D., Alibardi L. 2009c. Beta-keratins of the crocodilian epidermis: composition, structure, and phylogenetic relationships. J Exp Zool B Mol Dev Evol. 312(1):42–57.

Dalla Valle L., Nardi A., Toffolo V., Niero C., Toni M., Alibardi L. 2007a. Cloning and characterization of scale beta-keratins in the differentiating epidermis of geckos show they are glycine-proline-serine-rich proteins with a central motif homologous to avian beta-keratins. Dev. Dyn. 236:374–388.

Dalla Valle L., Nardi A., Belvedere P., Toni M., Alibardi L. 2007b. Beta-keratins of differentiating epidermis of snake comprise glycine-proline-serine-rich proteins with an avian-like gene organization. Dev. Dyn. 236:1939–1953.

Dalla Valle L., Toffolo V., Belvedere P., Alibardi L. 2005. Isolation of a mRNA encoding a glycine-proline-rich beta-keratin expressed in the regenerating epidermis of lizard. Dev. Dyn. 234:934–947.

De Guzman Strong C., Conlan S., Deming C.B., Cheng J., Sears K.E., Segre J.A. 2010. A milieu of regulatory elements in the epidermal differentiation complex syntenic block: implications for atopic dermatitis and psoriasis. Human Molecular Genetics. 19(8):1453–1460.

Dedeurwaerder R.A., Dobb M.G., Sweetman B.J. 1964. Selective extraction of a protein fraction from wool keratin. Nature 203:48–49.

Dhouailly D 1977. Dermo-epidermal interactions during morphogenesis of cutaneous appendages in amniotes. Front Matrix Biol. 4:86–121.

Dhouailly D. 1975. Formation of cutaceous appendages in dermo-epidermal recombinaitons between reptiles, birds and mammals. Wilhelm Roux' Arch. Entwicklungsmech Org. 177:323–340.

Dhouailly D. 2009. A new scenario for the evolutionary origin of hair, feather, and avian scales. J Anat. 214:587–606.

Di-Poï N. & Milinkovitch M.C. 2016. The anatomical placode in reptile scale morphogenesis indicates shared ancestry among skin appendages in amniotes. Science Advances. Vol. 2, no. 6, e1600708.

Eckhart L., Dalla Valle L., Jaeger K., Ballaun C., Szabo S.,Nardi A., Buchberger M., Hermann M., Alibardi L., Tschachler E. 2008. Identification of reptilian genes encoding hair keratin-like proteins suggests a new scenario for the evolutionary origin of hair. Proc Natl Acad Sci U S A. 105:18419–18423.

Eckhart L., Lippens S., Tschachler E., Declercq W. 2013. Cell death by cornification. Biochim Biophys Acta. 1833:3471–3480.

Evans S.E. 1995. Lizards: evolution, early radiation and biogeography. Sun, A. and Wang Y. (eds.). Sixth symposium on Mesozoic Terrestrial Ecosystems and Biota Published by China Ocean press:51–55.

Falla T.J., Karunaratne D.N., Hancock R.E. 1996. Mode of action of the antimicrobial peptide indolicidin. J Biol Chem. 271:19298–19303.

Field D.J., Gauthier J.A., King B.L., Pisani D., Lyson T.R., Peterson K.J. 2014. Toward consilience in reptile phylogeny: microRNAs support an archosaur, not a lepidosaur affinity for turtles. Evol Dev. 16(4):189–196.

Filshie B.K. & Rogers G.E. 1962. An electron microscope study of the fine structure of feather keratin. J. Cell Biol. 13:1–12.

Filshie B.K., Rogers G.E., Fraser R.D.B., MacRae T.P. 1964. X-ray diffraction and electron-microscope observations on soluble derivatives of feather keratin. Biochem. J. 92, 19.

Fong J.J., Brown J.M., Fujita M.K., Boussau B. 2012. Phylogenomic approach to vertebrate phylogeny supports a turtle-archosaur affinity and a possible paraphyletic Lissamphibia. PLoS ONE. 7:e48990.

Fraser N.C. & Benton M.J. 1989. The Triassic reptiles Brachyrhinodon and Polysphenodon and the relationships of the sphenodontids. Zool J Linn Soc. 96:413–445.

Fraser R.D.B. & MacRae T.P. 1963. Structural organization in feather keratin. J. Mol. Biol. 7:272–280.

Fraser R.D.B. & Macrae T.P. 1978.Current views on the keratin complex. In: Spearman RIC, Riley PA (eds.) Skin in vertebrates. London: University College Pp. 67–87.

Fraser R.D.B. & Parry D.A.D. 1996. The molecular structure of reptilian keratin. Int. J. Biol. Macromol. 19:207–11.

Fraser R.D.B. & Parry D.A.D. 2008. Molecular packing in the feather keratin filament. J. Struct. Biol. 162:1–13.

Fraser R.D.B. & Parry D.A.D. 2009. The role of b-sheets in the structure and assembly of keratins. Biophys. Rev. 1, 27–35. Biochemistry. 17:4277–4285.

Fraser R.D.B. & Parry D.A.D. 2011. The structural basis of the filament-matrix texture in the avian/reptilian group of hard beta-keratins. J. Struct. Biol. 173:391–405.

Fraser R.D.B. & Parry D.A.D. 2014. Amino acid sequence homologies in the hard keratins of birds and reptiles, and their implications for molecular structure and physical properties. J. Struct. Biol. .J Struct Biol. 188(3):213-24.

Fraser R.D.B. & Parry D.A.D. 2017. Filamentous structure of hard β-keratins in the epidermal appendages of birds and reptiles. In: D.A.D. Parry, J.M. Squire (eds.), Fibrous Proteins: Structures and Mechanisms, Springer International Publishing. Pp. 231–252.

Fraser R.D.B., MacRae T.P., Rogers G.E. 1972. Keratins: Their Composition, Structure and Biosynthesis. Charles C., Thomas, Springfield, Il, USA.

Friedrich C., Scott M.G., Karunaratne N., Yan H., Hancock R.E. 1999. Salt resistant alpha-helical cationic antimicrobial peptides. Antimicrob Agents Chemother. 43:1542–1548.

Fujikawa H., Fujimoto A., Farooq M., Ito, M., Shimomura Y. 2012. Characterization of human hair keratin-associated protein 2 (KRTAP2) gene family. J. Invest. Dermatol. 132:1806–1813.

Gaffney E.S. 1986. Triassic and Early Jurassic turtles. In PADIAN, K. (ed.). The beginning of the age of dinosaurs.Cambridge University Press, Cambridge. Pp. 183–188.

Gaffney E.S. 1990. The comparative osteology of the Triassic turtle Proganochelys. Bull Am Mus Nat Hist. 194:1–263.

Ganz T. 2003. The role of antimicrobial peptides in innate immunity. Integr. Comp. Biol. 43:300–304.

Gilbert S.F., Loredo G.A., Brukman A., Burke A.C. 2001. Morphogenesis of the turtle shell: the development of a novel structure in tetrapod evolution. Evol Dev. 3(2):47–58.

Gillespie J.M. 1972. Proteins rich in glycine and tyrosine from keratins. Comp. Biochem. Physiol. B 41:723–734.

Gillespie J.M. 1991. The structural proteins of hair: isolation, characterization and regulation of biosynthesis In: Goldsmith L.A. (eds.), Physiology, Biochemistry and Molecular Biology of the Skin, Oxford: Oxfords University Press. Pp 625–659.

Glenn T.C., French J.O., Heincelman T.J., Jones K.L., Sawyer R.H. 2008. Evolutionary relationships among species of feather beta (β) keratin genes from several avian orders. Integ Comp Biol. 48:463–475.

Godefroit P., Cau A., Dong-Yu H., Escuillié F., Wenhao W., Dyke G. 2013. A Jurassic avialan dinosaur from China resolves the early phylogenetic history of birds. Nature. 359–62.

Gordon M.S., Graham J.B., Wang T. 2004. Introduction to the special collection: revisiting the vertebrate invasion of the land. Physiological and biochemical zoology. Chicago, IL: University of Chicago Press on behalf of the Society for Integrative and Comparative Biology. 77 (5):697–699.

Goslar H.G. 1958. Beiträge zum Häutungsvorgang der Schlangen 1. Mitteilung: Histologische und topochemische Untersuchungen an der Haut von Natrix natrix L. während der Phasen des normalen Hautungszyklus. Acta Histochem. 5:182–212.

Green R.E., Braun E.L., Armstrong J., Earl D., Nguyen N., Hickey G., Vandewege M.W., St John J.A., Capella-Gutiérrez S., Castoe T.A. et al. 2014. Three crocodilian genomes reveal ancestral patterns of evolution among archosaurs. Science. 346,1254449:1–9.

Greenwold M.J. & Sawyer R.H. 2010. Genomic organization and molecular phylogenies of the beta (beta) keratin multigene family in the chicken (*Gallus gallus*) and zebra finch (*Taeniopygia guttata*): implications for feather evolution. BMC Evol Biol. 10:148.

Greenwold M.J. & Sawyer R.H. 2011. Linking the molecular evolution of avian beta keratins to the evolution of feathers. J Exp Zool B Mol Dev Evol. 316:609–616.

Greenwold M.J. & Sawyer R.H. 2013. Expression of archosaurian beta-keratins: diversification and expansion of archosaurian beta-keratins and the origin of feather beta-keratins. J Exp Zool B Mol Dev Evol. 320:393–405.

Greenwold M.J., Bao W., Jarvis E.D., Hu H., Li C., Gilbert M.T.P. , Zhang, G., Sawyer, R.H. 2014. Dynamic evolution of the alpha (α) and beta (β) keratins has accompanied integument diversification and the adaptation of birds into novel life-styles. BMC Evol Biol. 14:249–265.

Gregg K.& Rogers G. 1986. Feather keratins: composition, structure and biogenesis. In: Bereither-Hahn J., Matoltsy G., Sylvia-Richards K. editors. Biology of the integument: vertebrates. New York: Springer-Verlag Pp. 666–694.

Gregg K., Wilton S.D., Parry D.A.D., Rogers G.E. 1984. A comparison of genomic coding sequences for feather and scale keratins: structural and evolutionary implications. EMBO J. 3:175–178.

Hallahan D.L., Keiper-Hrynko N.M., Shang T.Q., Ganzke T.S., Toni M., Dalla Valle L., Alibardi L. 2009. Analysis of gene expression in gecko digital adhesive pads indicates significant production of cysteine- and glycine-rich beta-keratins. J Exp Zool B Mol Dev Evol. 312:58–73.

Hedges B. 2012. Amniote phylogeny and the position of turtles. BMC Biology. 10:64.

Henry J.,Toulza E., Hsu C., Pellerin L., Balica S., Mazereeuw-Hautier J., Paul C., Serre G., Jonca N., Simon M. 2012. Update on the epidermal differentiation complex. Front Biosci. 17:1517–1532.

Hirasawa T, Nagashima H, Kuratani S. 2013. The endoskeletal origin of the turtle carapace. Nat Commun. 4:2107.

Hirasawa T., Pascual-Anaya J., Kamezaki N., Taniguchi M., Mine K., Kuratani S. 2015. The evolutionary origin of the turtle shell and its dependence on the axial arrest of the embryonic rib cage. J Exp Zool B Mol Dev Evol. 324:194–207.

Hohl D., Mehrel T., Lichti U., Turner M.L., Roop D.R., Steinert P.M. 1991. Characterization of Human Loricrin. J Biol Chem. 266:6626–6636.

Homer B.L., Li C., Berry K.H., Denslow N.D., Jacobson E.R., Sawyer R.H., Williams J.E. 2001. Soluble scute proteins of healthy and ill desert tortoises (*Gopherus agassizii*). Am J Vet Res. 62:104–10.

Hsu C.H., Chen C., Jou M.L., Lee A.Y., Lin Y.C., Yu Y.P., Huang W.T., Wu S.H. 2005. Structural and DNA-binding studies on the bovine antimicrobial peptide, indolicidin: Evidence for multiple conformations involved in binding to membranes and DNA. Nucleic Acids Res. 33:4053–4064.

Huber M., Rettler I., Bernasconi K., Frenk E., Lavrijsen S.P., Ponec M., Bon A., Lautenschlager S., Schorderet D.F., Hohl D. 1995. Mutations of keratinocyte transglutaminase in lamellar ichthyosis. Science. 267:525–528.

Hugall A.F., Foster R., Lee M.S.Y. 2007. Calibration choice, rate smoothing, and the pattern of tetrapod diversification according to the long nuclear gene RAG-1. Syst Biol. 56:543-563.

Irish F., Williams E., Seling E. 1988. Scanning electron microscopy of changes in epidermal structure occurring during the shedding cycle in squamate reptiles. J Morphol. 197:105–26.

Irvine A.D., McLean W.H., Leung D.Y. 2011. Filaggrin mutations associated with skin and allergic diseases. N. Engl. J. Med. 365:1315.

Iwabe N., Hara Y., Kumazawa Y., Shibamoto K., Saito Y., Miyata T., Katoh K. 2005. Sister group relationship of turtles to the bird-crocodilian clade revealed by nuclear DNA–coded proteins. Mol. Biol. Evol. 22(4):810–813.

Janke A. & Arnason U. 1997. The complete mitochondrial genome of *Alligator mississippiensis* and the separation between recent archosauria (birds and crocodiles). Mol Biol Evol. 14(12):1266-72.

Jarvis E.D. Mirarab S., Aberer A.J., Li B., Houde P., Li C., Ho S.Y., Faircloth B.C., Nabholz B., Howard J.T. et al. 2014. Whole-genome analyses resolve early branches in the tree of life of modern birds. Science. 346 (6215):1320–1331.

Jayne B.C. 1988. Mechanical behavior of snake skin. J Zool Lond. 214:125–140.

Jones M.E., Anderson C.L., Hipsley C.A., Müller J., Evans S.E., Schoch R.R. 2013. Integration of molecules and new fossils supports a Triassic origin for Lepidosauria (lizards, snakes, and tuatara). BMC Evol. Biol. 13:208.

Kalinin A.E., Kajava A.V., Steinert P.M. 2002. Epithelial barrier function: Assembly and structural features of the cornified cell envelope. BioEssays. 24:789–800.

Kashyap D.R., Wang M., Liu L.H., Boons G.J., Gupta D., Dziarski R. 2011. Peptidoglycan recognition proteins kill bacteria by activating protein-sensing two-component systems. Nature Med. 17:676–683.

Kielan-Jaworowska Z., Cifelli R. L., Luo, Z. 2004. Mammals from the age of dinosaurs. Columbia University Press, New York, NY, USA.

Klein M.C. & Gorb S.N. 2012. Epidermis architecture and material properties of the skin of four snake species. J. R. Soc. Interface. 9:3140–55.

Klein M.C. & Gorb S.N. 2014. Ultrastructure and wear patterns of the ventral epidermis of four snake species (Squamata, Serpentes). Zoology (Jena) 117:295-314.

Klein M.C., Deuschle J., Gorb S. 2010. Material properties of the skin of the Kenyan sand boa *Gongylophis colubrinus* (Squamata, Boidae). J. Comp. Physiol. A 196:659–668.

Klüver E., Adermann K., Schulz A. 2006. Synthesis and structure-activity relationship of beta-defensins, multi-functional peptides of the immune system. J Pept Sci. 12:243–257.

Knapp L.W., Linser P.J., Carver W.E., Sawyer R.H. 1991. Biochemical identification and immunological localization of two non-keratin polypeptides associated with the terminal differentiation of avian scale epidermis. Cell Tissue Res. 265:535–545.

Knapp L.W., Shames R.B., Barnes G.L., Sawyer R.H. 1993. Region-specific patterns of beta keratin expression during avian skin development. Dev Dyn. 196:283–90.

Koster M.I. & Roop D.R. 2007. Mechanisms regulating epithelial stratification. Annu. Rev. Cell Dev. Biol. 23:93–113.ka

Kypriotou M., Huber M., Hohl D. 2012. The human epidermal differentiation complex: cornified envelope precursors, S100 proteins and the ‘fused genes’ family. Exp Dermatol. 21:643–649.

Lakshminarayanan R., Chi-Jin E. O., Loh X.J., Kini R.M.,Vivekanandan S. 2005. Purification and characterization of a vaterite-induced peptide, pelovaterin, from the eggshells of *Pelodiscus sinensis* (Chinese soft-shelled turtle). Biomacromolecules. 6:1429–1437.

Lakshminarayanan R., Vivekanandan S., Samy R.P., Banerjee Y., Chi-Jin E.O., Teo K.W., Jois S.D., Kini R.M., Valiyaveettil S. 2008. Structure, self-assembly, and dual role of a bdefensin-like peptide from the Chinese soft-shelled turtle eggshell matrix. J. Am. Chem. Soc. 130:4660–4668.

Landmann L. 1979. Keratin formation and barrier mechanisms in the epidermis of *Natrix natrix* (Reptilia: Serpentes): an ultrastructural study. J. Morphol. 162:93–126.

Landmann L. 1986. The skin of Reptiles: epidermis and dermis. In: Bereiter-Hahn J, Matoltsy AG, Sylvia-Richards K. (eds), Biology of the integument, Vol. 2 Vertebrates. Berlin-Heidelberg: Springer. Pp 150–187.

Langbein L. & Schweizer J. 2005. Keratins of the human hair follicle. Int Rev Cytol. 243:1–78.

Lee I.H., Cho Y., Lehrer R.I. 1997. Effects of pH and salinity on the antimicrobial properties of clavanins. Infect Immun. 65: 2898–2903.

Levitt M. & Perutz M.F. 1988. Aromatic rings act as hydrogen-bond acceptors. J. Mol. Biol. 201:751–754.

Levitt M. 1978. conformational preferences of amino acids in globular proteins. Biochemistry. 17 (20):4277–4285.

Li Y.I, Kong L., Ponting C.P., Haerty W. 2013. Rapid evolution of beta keratin genes contribute to phenotypic differences that distinguish turtles and birds from other reptiles. Genome Biol Evol. 5:923–933.

Lillywhite H.B. & Maderson P.F.A. 1982. Skin structure and permeability. Gans C, Pough F, editors. In Biology of the reptilia, physiology. London: Academic Press. C12:397–442.

Lillywhite H.B. 2006. Review: water relations of tetrapod integument. J. Exp. Biol. 209:202–226.

Liu Y., Zhou Q., Wang Y., Luo L., Yang J., Yang L., Liu M., Li Y., Qian T., Zheng Y. et al. 2015. *Gekko japonicus* genome reveals evolution of adhesive toe pads and tail regeneration. Nat. Commun. 6:10033.

Lu X., Wang M., Qi J., Wang H., Li X., Gupta D., Dziarski R. 2006. Peptidoglycan recognition proteins are a new class of human bactericidal proteins. J Biol Chem. 3;281(9):5895-907.

Lucas A.M. & Stettenheim P.R. 1972. Avian anatomy integument. In Agriculture Handbook 362. Agricultural Research Services. Washington DC: US Department of Agriculture.

MacArthur M.W. & Thornton J.M. 1991. Influence of proline residues on protein conformation. J. Mol. Biol. 218:397–412.

Maderson P. F. A., Baranowitz S., Roth S. I. 1978. A histological study of the long-term response to trauma of squamate integument. J. Morphol. 157:121−136.

Maderson P.F.A. & Alibardi L. 2000. The development of the sauropsid integument: a contribution to the problem of the origin and evolution of feathers. Am Zool. 40:513–529.

Maderson P.F.A. 1964. The skin of lizards and snakes. British Journal of Herpetology 3(6):151–154.

Maderson P.F.A. 1965. Histological changes in the epidermis of snakes during the sloughing cycle. J. Zool. (London) 146:98–113.

Maderson P.F.A. 1966. Histological changes in the epidermis of the Tokay (Gekko gecko) during the sloughing cycle. J. Morphol. 119:39–50.

Maderson P.F.A. 1985. Some developmental problems of the reptilian integument. In: Gans C, Billett F., Maderson P.F., editors. Biology of Reptilia. New York: John Wiley & Sons. Pp. 525–598.

Maderson P.F.A., Flaxman B.A., Roth S.I., Szabo G. 1972. Ultrastructural contribution to the identification of cell types in the lizard epidermal generation. J. Morphol. 136:191–210.

Maderson P.F.A., Rabinowitz T., Tandler B., Alibardi L. 1998. Ultrastructural contributions to an understanding of the cellular mechanisms involved in lizard skin shedding with comments on the function and evolution of a unique lepidosaurian phenomenon. J. Morphol. 236:1–24.

Maestrini E., Monaco A.P., McGrath J.A., Ishida-Yamamoto A., Camisa C., Hovnanian A., Weeks D.E., Lathrop M., Uitto J., Christiano A.M. 1996. A molecular defect in loricrin, the major component of the cornified cell envelope, underlies Vohwinkel's syndrome. Nature Genet. 13:70–77.

Marchler-Bauer A., Derbyshire M.K., Gonzales N.R., Lu S., Chitsaz F., Geer L.Y., Geer R.C., He J., Gwadz M., Hurwitz D.I. et al. 2016. CDD: NCBI's conserved domain database. Nucleic Acids Res. 43:D222–6.

Markova N.G., Marekov L.N., Chipev C.C., Gan S.Q., Idler W.W., Steinert P.M. 1993. Profilaggrin is a major epidermal calcium-binding protein. Mol Cell Biol. 13:613-625.

Marwick T.C. 1931. The X-ray classification of epidermal proteins. J. Text Sci. 4:31-33.

Matoltsy A.G. & Bednarz J.A. 1975. Lamellar bodies of the turtle epidermis. J Ultrastruct Res. 53:128–132.

Matoltsy A.G. & Huszar T. 1972. Keratinization of the reptilian epidermis: an ultrastructural study of the turtle skin. J Ultrastruct Res. 38:87–101.

Matoltsy A.G. 1962. Mechanism of keratinization, Fundamentals of Keratinization. Edited by E. Butcher, R.F. Sognnaes. Washington, DC, American Assoc for the Advancement of Science. Pp. 1–25.

Matoltsy A.G. 1987. Cell and molecular biology of keratins. Concluding remarks and future directions. In: Sawyer RH, editor. Topics in Developmental Biology, Vol. 22. Orlando, FL: Acc. Press Inc. Pp 255–264.

Matsunaga R., Abe R., Ishii D., Watanabe S-I., Kiyoshi M., Nöcker B., Tsuchiya M., Tsumoto K. 2013. Bidirectional binding property of high glycine-tyrosine keratin-associated protein contributes to the mechanical strength and shape of hair. J. Struct. Biol. 183:484–494.

McGaughey G.B., Gagné M, Rappé A.K. 1998. pi-Stacking interactions. Alive and well in proteins. J Biol Chem. 273:15458–15463.

Menon G.K., Maderson P.F.A., Drewes R.C., Baptista L.F., Price L.F., Elias P.M. 1996. Ultrastructural organization of avian stratum corneum lipids as the basis for facultative cutaneous waterproofing. J. Morphol. 227:1–13.

Milinkovitch M.C., Manukyan L., DebryA., Di-Poï N., Martin S., Singh D., Lambert D., Zwicker M. 2013. Crocodile head scales are not developmental units but emerge from physical cracking. Science. 339, 6115:78–81.

Mischke D., Korge B.P., Marenholz I., Volz A., Ziegler A. 1996. Genes encoding structural proteins of epidermal cornification and S100 calcium-binding proteins form a gene complex (“epidermal differentiation complex”) on human chromosome 1q21. J Invest Dermatol. 106:989–992.

Mittal A.K. & Singh J.P.N. 1987a. Scale epidermis of Natrix piscator during its sloughing cycle structural organization and protein histochemistry. J. Zool. (Lond.) 213:545–568.

Mittal A.K. & Singh J.P.N. 1987b. Hinge epidermis of Natrix piscator during its sloughing cycle structural organization and protein histochemistry. J. Zool. (Lond.) 213:685–695.

Mlitz V., Strasser B., Jaeger K., Hermann M., Ghannadan M., Buchberger M., Alibardi L., Tschachler E., Eckhart L. 2014. Trichohyalin-like proteins have evolutionarily conserved roles in the morphogenesis of skin appendages. J. Invest. Dermatol. 134:2682–2692.

Moustakas-Verho J.E. & Cherepanov G.O. 2015. The integumental appendages of the turtle shell: An evo-devo perspective. J. Exp. Zool. (Mol. Dev. Evol.) 324B:221–229.

Moustakas-Verho J.E., Zimm R., Cebra-Thomas J., Lempiäainen N.K., Kallonen A., Mitchell K.L., Häamäaläainen K., Salazar-Ciudad I., Jernvall J., Gilbert S.F. 2014. The origin and loss of periodic patterning in the turtle shell. Development. 141:3033–3039.

Nagashima H., Kuraku S., Uchida K., Kawashima-Ohya Y., Narita Y., Kuratani S. 2007. On the carapacial ridge in turtle embryos: its developmental origin, function and the chelonian body plan. Development. 134(12):2219–2226.

Nagashima H., Sugahara F, Takechi M, Ericsson R, Kawashima-Ohya Y, Narita Y, Kuratani S. 2009. Evolution of the turtle body plan by the folding and creation of new muscle connections. Science. 325(5937):193–196.

Ng C.S., Chen C.K., Fan W.L., Wu P., Wu S.M., Chen J.J., Lai Y.T., Mao C.T., Lu M.Y., Chen D.R. et al. 2015. Transcriptome analyses of regenerating adult feathers in chicken. BMC Genomics. 16:1–16.

Ng C.S., Wu P., Fan W.L., Yan J., Chen C.K., Lai Y.T., Wu S.M., Mao C.T., Chen J.J., Lu M.Y. et al. 2014. Genomic organization, transcriptomic analysis, and functional characterization of avian alpha- and beta-keratins in diverse feather forms. Genome Biol Evol. 6(9):2258–73.

Ng C.S., Wu P., Foley J., Foley A., McDonald M.L., Juan W.T., Huang C.T., Lai Y.T., Chen C.F., Leal S.M. et al. 2012. The chicken frizzle feather is due to an alpha-keratin (KRT75) mutation that causes a defective rachis. PLoS Genet. 8 (7), e1002748.

Niessen C.M. 2007. Tight junctions/adherens junctions: basic structure and function. J Invest Dermatol. 27:2525–2532.

O'Guin M.W., Galvin S., Schermer A., Sun T.T. 1987. Pattern of keratin expression define distinct pathways of epithelial development and differentiation. In: Sawyer, R.H. (Ed.), Topics in Developmental Biology. 22. Acc. Press Inc, Orlando, FL. Pp. 97–125.

Palmer C.N., Irvine A.D., Terron-Kwiatkowski A., Zhao Y., Liao H., Lee S.P., Goudie D.R., Sandilands A., Campbell L.E., Smith F.J. et al. 2006. Common loss-of-function variants of

the epidermal barrier protein filaggrin are a major predisposing factor for atopic dermatitis. Nat. Genet. 38:441.

Parakkal P.F. & Alexander N.J. 1972. Keratinization: A Survey of Vertebrate Epithelia. Academic Press, New York, Ch.4.

Parry D.A., Smith T.A., Rogers M.A., Schweizer J. 2006. Human hair keratin-associated proteins: sequence regularities and structural implications. J Struct Biol. 155:361–9.

Pough F.H., Andrews R.M., Cadle J.E., Crump M.L., Savitzky A.H., Wells K.D. 2015. Herpetology, 4th edition, Sinauer Associates.

Pough F.H., Andrews R.M., Cadle J.E., Crump M.L., Savitzky A.H., Wells K.D. 2001. Herpetology, 2nd edition. Upper Saddle River, NJ: Prentice Hall.

Pough F.H., Janis C.M., Heiser J.B. 1999. Vertebrate life, 5th edition. Upper Saddle River, NJ: Prentice Hall.

Powell B.C. & Rogers G.E. 1994. Differentiation in hard tissues: hair and related structures. In: Leigh I, Lane B, Watt F (Eds): The keratinocyte handbook. pp 401–436. Cambridge University Press, Cambridge, UK.

Powell B.C. & Rogers G.E. 1997. The role of keratin proteins and their genes in the growth, structure and properties of hair. EXS. 78:59–148.

Presland R.B., Gregg K., Molloy P.L., Morris C.P., Crocker L.A., Rogers G.E. 1989. Avian keratin genes I. A molecular analysis of the structure and expression of a group of feather keratin genes. J. Mol. Biol. 209:549–560.

Price R.M. 1982. Dorsal scale microdermatoglyphics: ecological indicator or taxonomic tool. J Herpetol 16:294–306.

Prum O.R. 1999. Development and evolutionary origin of feathers. J. Exp. Zool. 285:291–306.

Pyron R.A., Burbrink F.T., Wiens J.J. 2013. A phylogeny and revised classification of Squamata, including 4161 species of lizards and snakes. BMC Evol. Biol. 13:93.

Regal P.J. 1975. The evolutionary origin of feathers. The Quarterly Review of Biology 50(1):35-66.

Reisz R.R. & Head J.J. 2008. Palaeontology: turtle origins out to sea. Nature. 456:450–451.

Resing K.A. & Dale B.A. 1991. Proteins of keratohyalin. In: Goldsmith LA, editor. Physiology, Biochemistry and Molecular Biology of the Skin, Vol. 1, 2nd ed. New York: Oxford University Press. Pp 148–167.

Rice R., Riccio P., Gilbert S.F., Cebra-Thomas J. 2015. Emerging from the rib: resolving the turtle controversies. J Exp Zool B Mol Dev Evol. 324:208–220.

Rice R.H. & Green H. 1977. The cornified envelope of terminally differentiated human epidermal keratinocytes consists of cross-linked protein. Cell.11:417–422.

Rice R.H., Winters B.R., Durbin-Johnson P.D., Rocke D.M. 2013. Chicken corneocyte cross-linked proteome. J Prot Res. 12:771–776.

Richardson K.C., Webb G.J.W., Manolis S.C. 2002. Crocodiles: Inside Out. A Guide to the Crocodilians and their Functional Morphology. Surrey Beatty & Sons, Chipping Norton.

Ripamonti A., Alibardi L., Falini G., Fermani S., Gazzano M. 2009. Keratin-lipid structural organization in the corneous layer of snake. Biopolymers. 91:1172–1181.

Rogers G.E. 1985. Genes for hair and avian keratins. Ann N Y Acad Sci. 455:403–25.

Rogers M.A., Langbein L., Praetzel-Wunder S., Winter H., Schweizer J. 2006. Human hair keratin-associated proteins (KAPs). Int. Rev. Cytol. 251:209–263.

Rogers M.A., Langbein L., Winter H., Ehmann C., Praetzel S., Korn B., Schweizer J. 2001. Characterization of a cluster of human high/ultrahigh sulfur keratin-associated protein genes embedded in the type I keratin gene domain on chromosome 17q12-21. J Biol Chem. 276:19440–51.

Rougier G.W., de la Fuente M.S., Arcucci A.B. 1995. Late Triassic turtles from South America. Science. 268:855–858.

Ruckes H. 1929. The morphological relationships between the girdles, ribs, and carapace. Ann N Y Acad Sci. 13:81–120.

Rudall K.M. 1947. X-ray studies of the distribution of protein chain types in the vertebrate epidermis. Biochim. Biophys. Acta 1:549–562.

Sawyer R.H. & Knapp L.W. 2003. Avian skin development and the evolutionary origin of feathers. J Exp Zool B Mol Dev Evol. 298:57–72.

Sawyer R.H., Glenn T.C., French J.O., Mays B., Shames R.B., Barnes G.L., Rhodes W., Ishikawa Y. 2000. The expression of beta (β) keratins in the epidermal appendages of reptiles and birds. Am. Zool. 40:530–539.

Sawyer R.H., Knapp L.W., O'Guin W.M. 1986. The skin of birds: epidermis, dermis and appendages. In: Bereiter-Hahn J, editor. Biology of the integument. Vol. 2. Vertebrates. Springer Verlag; New York: Pp. 194.

Sawyer R.H., Rogers L., Washington L., Glenn T.C., Knapp L.W. 2005. Evolutionary origin of the feather epidermis. Dev Dyn. 232:256–267.

Sawyer R.H., Salvatore B.A., Potylicki T.T., French J.O., Glenn T.C., Knapp L.W. 2003a. Origin of feathers: Feather beta (beta) keratins are expressed in discrete epidermal cell populations of embryonic scutate scales. J Exp Zool B Mol Dev Evol. 295:12–24.

Sawyer R.H., Washington L.D., Salvatore B.A., Glenn T.C., Knapp L.W. 2003b. Origin of archosaurian integumentary appendages: The bristles of the wild turkey beard express feather-type beta keratins. J Exp Zool. 297B Mol. & Dev. Evol. 27–34.

Sharpe P.T. 2001. Fish scale development: hair today, teeth and scales yesterday? Curr Biol. 11:R751–R752.

Shen X.X., Liang D., Wen J.Z., Zhang P. 2011. Multiple genome alignments facilitate development of NPCL markers: a case study of tetrapod phylogeny focusing on the position of turtles. Mol Biol Evol. 28:3237–3252.

Shu D.G., Luo H.L., Conway Morris S., Zhang X.L., Hu S.X., Chen L., Han J., Zhu M., Li Y., Chen L.Z. 1999. Lower Cambrian vertebrates from south China. Nature. 402 (6757):42–46.

Soulé M. & Kerfoot W.C. 1972. On the climatic determination of scale size in a lizard. Systematic Zoology. 21(1):97–105.

Spearman R.I.C. 1969. The epidermis of gopher tortoise Testudo polyphemus (Daudin). Acta Zool. 50:1–9.

Spearman R.I.C. & Riley P.A. 1969. A comparison of the epidermis and pigment cells of the crocodile with those in two lizard species. Zool J Linn Soc. 48:453–66.

Spearman R.I.C. 1966. The keratinization of epidermal scales, feathers and hairs. Biol Rev. 41:59–96.

Spearman R.I.C. 1973. In The Integument: A Textbook of Skin Biology. Cambridge University Press, London. The skin of reptiles. Pp. 83–90.

Staudt K., Saxe F.P.M., Schmied H., Soeur R., Bohme W., Baumgartner W. 2012. Comparative investigations of the sandfish's beta-keratin (Reptilia: Scincidae: Scincus scincus). Part 1: surface and molecular examinations. J. Biomim. Biomater. Tissue Eng. 15:1–16.

Steinert P.M & Marekov L.N. 1995. The proteins elafin, filaggrin, keratin intermediate filaments, loricrin, and small proline-rich proteins 1 and 2 are isodipeptide cross-linked componentsof the human epidermal cornified cell envelope. J. Biol. Chem. 270:17702.

Steinert P.M., Mack J.W., Korge B.P., Gan S.Q., Haynes S.R., Steven A.C. 1991. Glycine loops in proteins: their occurrence in certain intermediate filament chains, loricrins and single-stranded RNA binding proteins. Int J Biol Macromol. 13(3):130–9.

Steinert, P. M., Kartasova, T. & Marekov, L. N. 1998a. Biochemical evidence that small proline-rich proteins and trichohyalin function in epithelia by modulation of the biomechanical properties of their cornified cell envelopes. J. Biol. Chem. 273:11758–11769.

Steinert, P. M., Candi, E., Kartasova, T. & Marekov, L. 1998b. Small proline-rich proteins are cross-bridging proteins in the cornified cell envelopes of stratified squamous epithelia. J. Struct. Biol. 122:76–85.

Steinert P.M., Candi E., Tarcsa E., Marekov L.N., Sette M., Paci M., Ciani B., Guerrieri P., Melino G. 1999. Transglutaminase crosslinking and structural studies of the human small proline rich 3 protein. Cell Death Differ. 6:916–930.

Steven A.C. & Steinert P.M. 1994. Protein composition of cornified cell envelopes of epidermal keratinocytes. J. Cell Sci. 107:693–700.

Strasser B., Mlitz V., Hermann M., Rice R.H., Eigenheer R.A., Alibardi L., Tschachler E., Eckhart L. 2014. Evolutionary origin and diversification of epidermal barrier proteins in amniotes. Mol. Biol. Evol. 31:3194–3205.

Strasser B., Mlitz V., Hermann M., Tschachler E., Eckhart, L. 2015. Convergent evolution of cysteine-rich proteins in feathers and hair. BMC Evol. Biol. 15:82.

Sun C., Mathur P., Dupuis J., Tizard R., Ticho B., Crowell T., Gardner H., Bowcock A. M., Carulli J. 2006. Peptidoglycan recognition proteins Pglyrp3 and Pglyrp4 are encoded from the

epidermal differentiation complex and are candidate genes for the Psors4 locus on chromosome 1q21. Hum. Genet. 119:113–125.

Taylor A.M., Bonser R.H.C., Farrent J.W. 2004. The influence of hydration on the tensile and compressive properties of avian keratinous tissues. J. Mater. Sci. 39:939–942.

Toni M., Dalla Valle L., Alibardi L. 2007. Hard (beta-) keratins in the epidermis of reptiles: Composition, sequence, and molecular organization. J Proteome Res. 6:3377–3392.

Tzika A.C., Helaers R., Schramm G., Milinkovitch M.C. 2011. Reptilian-transcriptome v1.0, a glimpse in the brain transcriptome of five divergent Sauropsida lineages and the phylogenetic position of turtles. EvoDevo. 2:1–18.

Uetz P. & Hošek J. (Eds.). The Reptile Database: http://www.reptile-database.org, based on access Jan 2017.

van Tuinen M. & Hadly E.A. 2004. Error in estimation of rate and time inferred from the early amniote fossil record and Avian molecular clocks. J Mol Evol. 59:267.

Vandebergh W. & Bossuyt F. 2012. Radiation and functional diversification of alpha keratins during early vertebrate evolution. Mol Biol Evol. 29:995–1004.

Vermeij W.P. & Backendorf C. 2010. Skin cornification proteins provide global link between ROS detoxification and cell migration during wound healing. PLoS ONE. 5(8):e11957.

Vermeij W.P., Alia A., Backendorf C. 2011. ROS quenching potential of the epidermal cornified cell envelope. J Invest Dermatol. 131:1435–1441.

Wagner G. P. 2014. Homology, genes and evolutionary innovation. Princeton University Press.

Walkenhorst W.F., Wolfgang Klein J., Vo P., Wimley W.C. 2013. pH Dependence of microbe sterilization by cationic antimicrobial peptides. Antimicrob. Agents Chemother. 57:3312–3320.

Walker I.D. & Rogers G.E. 1976. Differentiation in avian keratinocytes. The properties of the proteins of the chick down feather. Eur J Biochem. 69:329–339.

Wan Q.H., Pan S.K., Hu L., Zhu Y., Xu P.W., Xia J.Q., Chen H., He G.Y., He J., Ni X.W. et al. 2013. Genome analysis and signature discovery for diving and sensory properties of the endangered Chinese alligator. 23(9):1091–105.

Wang H., Parry D.A.D., Jones L.N., Idler W.W., Marekov L.N., Steinert P.M. 2000. In vitro assembly and structure of trichocyte keratin intermediate filaments:A novel role for stabilization by disulfide bonding. J Cell Biol. 151:1459–1468.

Wang Z., Pascual-Anaya J., Zadissa A., Li W., Niimura Y., Huang Z., Li C., White S., Xiong Z., Fang D. et al. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. Nat Genet. 45:701–706.

Waters M.L. 2002. Aromatic interactions in model systems. Curr Opin Chem Biol. 6:736–741.

Watt F. M. 1984. Selective migration of terminally differentiating cells from the basal layer of cultured human epidermis. J. Cell Biol. 98:16–21.

Watt I.C. & Leeder J.D. 1968. Role of carboxyl groups in water absorption by keratin. J. Chem. Tech. Biol. 18:1–4.

Wei G., Campagna A.N., Bobek L.A. 2007. Factors affecting antimicrobial activity of MUC7 12-mer, a human salivary mucin-derived peptide. Annals of Clinical Microbiology and Antimicrobials. 6:1–10.

Whitbread L.A., Gregg K., Rogers G.E. 1991. The structure and expression of a gene encoding chick claw keratin. Gene. 101:223–229.

Wiche G., Osmanagic-Myers S., Castañón M.J. 2015. Networking and anchoring through plectin: a key to IF functionality and mechanotransduction. Curr Opin Cell Biol. 32:21–29.

Wiens J.J., Hutter C.R., Mulcahy D.G., Noonan B.P., Townsend T.M., Sites Jr. J.W., Reeder T.W. 2012. Resolving the phylogeny of lizards and snakes (Squamata) with extensive sampling of genes and species. Biol. Lett. 8:1043–1046.

Wu G., Ding J., Li H., Li L., Zhao R., Shen Z., Fan X., Xi T. 2008. Effects of cations and pH on antimicrobial activity of thanatin and s-thanatin against Escherichia coli ATCC25922 and B. subtilis ATCC 21332. Curr Microbiol. 57:552-557.

Wu P., Hou L., Plikus M., Hughes M., Scehnet J., Suksaweang S., Widelitz R., Jiang T.X., Chuong C-M. 2004. Evo-Devo of amniote integuments and appendages. Int J Dev Biol. 48:248–267.

Wu P., Ng C.S., Yan J., Lai Y-C., Chen C-K., Lai Y-T., Wu S-M, Chen J-J, Luo W., Widelitz R.B. et al. 2015. Topographical mapping of α-and β-keratins on developing chicken skin integument: functional interaction and evolutionary perspectives. Proc Natl Acad Sci USA. 112:E6770– E6779.

Wyld J.A. & Brush A.H. 1979. The molecular heterogeneity and diversity of reptilian keratins. J Mol Evol. 12:331–347.

Wyld J.A. & Brush A.H. 1983. Keratin diversity in the reptilian epidermis. J. Exp. Zool. 225:387–396.

Xu X., You H., Du K., Han F. 2011. An Archaeopteryx-like theropod from China and the origin of Avialae. Nature.475:465–470.

Ye C., Wu X., Yan P., Amato G. 2010. beta-Keratins in crocodiles reveal amino acid homology with avian keratins Mol Biol Rep. 37(3):1169–74.

Zangerl R. 1969. The turtle shell. In: Gans C. (ed.), Biology of the Reptilia, Vol 1, Ch 6, Accad Press, London-New York, Pp. 311-339.

Zelezetsky I., Pag U., Sahl H.G., Tossi A. 2005. Tuning the biological properties of amphipathic alpha-helical antimicrobial peptides: Rational use of minimal amino acid substitutions. Peptides. 26:2368-2376.

Zheng Y. & Wiens J.J. 2016. Combining phylogenomic and supermatrix approaches, and a time-calibrated phylogeny for squamate reptiles (lizards and snakes) based on 52 genes and 4162. Molecular Phylogenetics and Evolution 94:537–547.

Supplementary Data: Supplementary Figures and Tables

Content

Supplementary Data: Supplementary Figures

Comparative genomics identifies epidermal differentiation proteins associated with the evolution of the turtle shell

Karin Brigit Holthaus, Bettina Strasser, Wolfgang Sipos, Heiko A. Schmidt, Veronika Mlitz, Supawadee Sukseree, Anton Weissenbacher, Erwin Tschachler, Lorenzo Alibardi, Leopold Eckhart

Content

A

>Cp_CRNN
MTQLLSNIKGIINAFYVFAKKDGACPTLSKGELRQLIHQEFADVTVVPQGLQTIDKLLQLLDTDSDGRLDFNGFLVLIFQVAK
ACYGEVSQGQRPGHGGSSASQGEANCGERTKEPPTPERDPSPRQAPEPQTPERDSIPCQAPEPQIPEQDSSPCQAPELQIPEQ
DSNPCQAPEAQIPERDPSPCQAPELQTPEQDSSPRQAPEPQIPEQDSSPRQALEPQIPEQDSIPCQAPEAQIPERDPSPHQAP
EPQTPERDPSPHQGLEPQTPERDPSPCHAPEPQIPEQDPSLHQAPEPQTPEQDSSSRHTPEPQIPEQEPSPCRGPEPQTPEQD
LSPSQAPEPQTPEQDLSHSETQLPPTQQRNQGAQDPTEPAAGQASKSSQCLYSWHSQKPRPFPHWWPPKK

>Cp_EDAA1
MFHHQKICKPWCCKPHQKICKPWCCKPWGYGGSSGYGGDYGYCPPFWCKKPFKCCYPYPYPGCYPYPKPCCYPCPYPYPCGPG
YQYPCLAEEE

>Cp_EDAA2
MSFNKSIIGELYYNPCCYGGYRGYRGYGYCRPWCYQRPYKYGWGHHYKCCYPYPYQWGYGKGWPCFAEEE

>Cp_EDAA3
MNYHHQKLSHHWGCDPCWNGGWGGYGGHYGCYRPWGYYRPYGWGWGHSYGYPYRWGGGYGYGRCWPCFAEEE

>Cp_EDAA4
MTYHHQKLSHHWGCDPCWNGGWGGYGGHYGCYRPWGYYRPYSYGWGHNYGSCYSYPYRWGGGYGYGRCWPCFAEEQ

>Cp_EDAA5
MTFDELMNEELYYNPYCYKGWRGYRGHYGCYRPWGYQRPYRYGWGHQYDCHYPYRWGHGYGYGKFWPCFAEEQ

>Cp_EDAA6
MTFDESINDELYYNPWSHGCWHGSRGHYGCGRPWGYGRQSRWGWGHGYDCYYPYSSRWGHWYPYVKQWPC

>Cp_EDAA7
MTFDELMNEELYYNPYCYKGWRGYRGHYGCYRPWGYQRPYRYGWGHQYDCHYPYRWGHGYGYGKFWPCFAEEQ

>Cp_EDAA8
MTFHHQKLSHHWGCDPCSSGSWGGYRGHYDCYRPWGYSRPYGCGWGYNDGCYYPYSSRWGHGYGGYGYGGCGYGYGGHGYGKC
WPC

>Cp_EDAA9
MNYHHQKLSHHWGCDPCWNGGWGGYGGHYGCYRPWGYYRPYSYGWGHNSGSCYSYPYRWGGGYGYGRCWPCFAEEQ

>Cp_EDAA10
MTWSGYGYNDGCYSPCGYGGRWAYGSPCGYRGLCGYGGHSSHGGSWGYRGSYGYRGAYHSGYCYPFSSQQGHRYSYGNCGPC

>Cp_EDAA11
MTYHHQKLSHHWGCDPCWNGGWGGYGGHYGCYRPWGYYRPYSYGWGHNSGSCYSYPYRWGGGYGYGRCWPCFAEEQ

>Cp_EDAA12
MNYHHQKLSHHWGCDPCWNGGWGGYGGYYGCYRPWGYYRPYSYGWGHNSGSCYSYPYRWGGGYGYGRCWPCFAEEQ

>Cp_EDAA13
MTYHHQKLSHHWGCDPCWNGGWGGYGGHYGCYRPWGYYRPYSYGWGHNYGSCYSYPYRWGGGYGYGRCWPCFAEEQ

>Cp_EDAA14
MTYHHQKLSHHWGCDPCWNGGWGGYGGHYGCYRPWGYYRPYSYGWGHNYGSCYSYPYRWGGGYGYGRCWPCFAEEQ

>Cp_EDAA15
MNYHHQKLSHHWGCDPCWNGGWGGYGGHYGCYRPWGYYRPYSYGWGHNSGSCYSYPYRWGGGYGYGRCWPCFAEEE

>Cp_EDAA16
MNYHHQKLSHHWGCDPCWNGGWGGYGGHYGCYRPWGYYRPYSYGWGHNYGSCYSYPYRWGGGYGYGRCWPCFAEEE

>Cp_EDAA17
MTFDELMNEELYYNPYCYKGWRGYRGHYGCYRPWGYQRPYRYGWGHQYDCHYPYRWGHGYGYGKFWPCFAEEQ

>Cp_EDAA18
MTFDELMNEELYYNPYCYKGWRGYRGHYGCYRPWGYQRPYRYGWGHQYDCHYPYRWGHGYGYGKFWPCFAEEQ

>Cp_EDAA19
MTFDENFSDELYYKPYHYGGWGGRGYGYCRPWCYQRPYKCCWGYPKGCWYPDPCHWGWGYGYGKGWPCFAQEE

>Cp_EDAA20
MTFDENFSEKLDYKPCHYGGWRGRGYGWGRPWCYQRPYRCCWGYPKGCWYPYPCHWGWGYGYGKGWPCFAQEE

>Cp_EDAA21
MTFDENFSDELYYKPYHYGGWGGRGYGYCKPWCHQRPYKCCWGYPKGCWYPYPCHWGWGYGYGKDWPCFAQEE

>Cp_EDAA22
MTFDENFSDELYYKPYHYGGWGGRGYGWCRPWCYQRPYKCCWGYPKGCWYPYPCHWGWGYGYGKGWPCFAQEE

>Cp_EDKM
MSNLIKAIADMIDSYQSNSRKGRESERFRRCEFKKLVQQDPTPAKRSSSNKHKHTTSLPDSDAELMNKKELITANPCVQ

>Cp_EDP1
MPYYGQQHKHLPAPVCVTKCSQPCPPQYEQHCVPKCRPVYVTKCPPLYGPQYAYPCAPQCPPRCVTKCPPRCVTKCPPPCVTK
CPPPCVTKCPPPCVTKCPPPCVTKCPPPCVTKCPPPCVTKCPPPCVTKCPPPCVTKCPPPCVTKCPPPCMTKCPQQCVTQCPG
QYQSGKVQISSHCKKYCSAPKWPW

>Cp_EDP2
MASRQNQQQRKQTLTLPPALSNATSEPAPPPEAVPEPCPATVEEPENSPQEEEGPQEEYKQPLNQPLGPAPELEPEPVLCPEP
ESNPPEVKEIEYLQPDHQQYKHPPTLPPAPGMETSKEYQQAESEPELGRCPPPIREPEGPPFVQPSSPVEEQQQKQPHHWPPK
RK

>Cp_EDP3
MSSDQQQCKQTCPPPPKCQEKCPPPCKEPVKTPKCQEKCPPPSKEPKCPPPKQSQDWKQC

>Cp_EDPCV1
MAYQQQCKQPCLPPPCCVKQCKTKCVDPCPCPPQCVDPCPPCPPKCVDPCPPKCVDPCPPKCVDLCPPKCVDQFPCPPKCVDV
CPPKCVDLCPPKCVDQCPCPPKCVDVCPPKCVDVCPPKCVDVCPPCPPLQHCCQEKKHY

>Cp_EDPCV2
MAYQQQCKQTCLPPPCCVTKCVTKCLDPCYKVCVTKCVTKCLDPCCKVCVKKCTRCVHPCSCPPKHVDPCPPCLPKCPPVQHC
CKEKKPC

>Cp_EDPCV3
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPQKCLPCPPCPQKCPPCPP
CPPCPQKCPPCPQKCPPCPPCPPCPQKCPPCPPCPQKCPPCPPCPPKCSPVQHCCKEKKLC

>Cp_EDPCV4
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPQKCIPCPQKCIPCPPCPQ
KCPPCPPCPQKCPPCPPCPPCLPKCPPVQHCCKEKKLC

>Cp_EDPCV5
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPQKCLPCPPCPQKCPPCPP
CPQKCPPCPPCLPKCPPVQHCCKEKKLC

>Cp_EDPCV6
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPQKCLPCPPCPPCHQKCPP
CPQKCPPCPPCPPCPPKCPPVQHCCKEKKLC

>Cp_EDPCV7
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTKCVHPCPCPCPQKCIPCPPCPQKCPPCPP
CPPCPQKCPPCPPCPPCPQKCPPCLPCPQKCPPCPPCPPCLPKCPPVQHCCKEKKLC

>Cp_EDPCV8_partial
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPQKCLPCPPCPQKCPPCPP
CPPCPQXXXXXXXXXX

>Cp_EDPCV9
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPQKCPPCAPCPPCPQKCPP
CPPCPPCHQKCPPCPPCLPKCPPVQHCCKEKKLC

>Cp_EDPCV10
MAYQQQCKQPCLPPPCCVTKCTTKCLDPCCKVCVKKCTTCVHPCPCPCPCPQKCLPCPPCPPCPQKCPPCPPCPPCPEKCPPC
PPCPPCPPECPPVQHCCKEKKLC

>Cp_EDPCV11
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPQKCIPCPPCPQKCPPCPPCP
QKCPPCPPCPPCLPKCPPVQHCCKEKKLC

>Cp_EDPCV12
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCNVCVKKCTTCVHPCPCPCPQKCIPCPPCPQKCPPCPP
CPQKCPPCPPCPPCLPKCPPVQHCCKEKKLC

>Cp_EDPCV13
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPQKCIPCPPCPQKCPPCPQ
KCLPCPQKCIPCPPCPQKCPPCPQKCLPCPQKCLPCPPCPQKCPPCPPCPPCPQKCPPCPPCPQKCPPCPPCPPCLPCPQ
KCPPCPPKCPPVQHCCKEKKLC

>Cp_EDPCV14
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPQKCLPCPQKCLPCPPCPQ
KCPPCPPCPPCPQKCPPCLPCPPCPQKCPPCPPCPPCLPCPQKCPPCPPCPPKCPPVQHCCKEKKLC

>Cp_EDPCV15
MAYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPQKCLPCPPCPQKCPPCPP
CPPCPQKCPPCPPCPPCLPCPQKCPPCPPCPPKCPPVQHCCKEKKLC

>Cp_EDPE
MSLHQDQQQCKQGITLPPALCKEKCPEMVPCPEPVKCPEPIPCCPVKPPCKEPPVPIPTPCPEPIPCPQQKQQCKLPPVPVPL
PHPEPIPCSPEKPPCKEPPFPHPLPHPEPIPHCPEKPPCKEPKVPVPLPHPEPIPYCPDKSPCKQPPAPVTPPCPKPIPCSPE
KPPCKEPPFPHPLPHPEPIPCPQHKQQCKLPPVPVPLPHPEPIPHCPEKPPCKEPKVPVPLPHPELIPHCPEKPPCKEPPFPH
PLPHPEPIPHCPEKPPCKEPKVPVPLPHPEPIPCPQQKQQCKLPPVPVPTPCPEPIPCSPEKPPCKEPPAPVTPPCPKPRPCP
ENPPCKEPPVPVTPPCPEPVKCPPCVKTCPPIEQQQCKQQCQLPPNWK

>Cp_EDPL1
MSCHQHQQQCKQPCMPPPCKEPDRPKTTEPCPQQCTEPCPQQCTEPRPPKCVETCPPKCPPVQQC

>Cp_EDPQ1
MSYQHQQQCKQTCLPPPVCPPPQCPEPCPPPKCPEPCPPPKCPELCPPPQCPEPCPPLKCPEPCPPPKCQEPCPPPKCPEPCP
PPKCPEPCPPPQCQEPCRPPKCPSPKCPPMQKYN

>Cp_EDPQ2
MSYQHQQQCKQTCLPPPVCPPPQCPEPCPPPKCPEPCPPPKCPEPCPAPQCPEPCPPPKCPEPCPAPQCPEPCPPLKCPEPCP
PPKCPEPCPQPQCQEPCPPPQCQEPCKPPKCPSPKCPPMQKYN

>Cp_EDQM1
MCSRQEKDHCHKQDTCHGSGGGSSCHGSGGGSSCHGSGGGSSCHGSGGGSSCHGSGGGSSCHGSGGGSSCHGKPQKPCQQEQQ
QQQKHCCQVPSQKLK

>Cp_EDQM2
MCSRQEKDHCHKQDVSGGCHSSGSSCHSSGGSSCHSGGSSCHGSGGSSCHSSGGSSCHGGGSSCHSGGSCHGKPQQPCQQQQQ
QKICKVPCQKLK

>Cp_EDQM3
MCSRQEKDHCHKQDGCHSSGGCHSSGSSCHSGGGSSCHSGGSSCHGSGGSSCHSSGGSSCHGGGSSCHSGGSCHGKPQQHCQQ
QQQQKICKVPCQKLK

>Cp_EDQM4
MCSRQEKDHCHKQDGCHSSGGCHSSGSSCHSGGGSSCHSGGSSCHGSGGSSCHSSGGSSCHGGGSSCHSGGSCHGKPQQHCQQ
QQQQQKICKVPCQKLK

>Cp_EDQM5
MCSRQEKDHCHKQDGCHSSGGCHSSGSSCHSSGGSSCHSGGSSCHGSGGSSCHSSGGSSCHGGGSSCHSGGSCHGKPQQHCQQ
QQQQKICKVPCQKRK

>Cp_EDQM6
MCSRQEKDHCHKQDGCHSSGGCHSSRSSCHESGGSSCHSGGSSCHGSGGSSCHSSGGSSCHGGGSSCHSGGSCHGKPQQHCQQ
QQQQKICKVPCQKLK

>Cp_EDQM7
MCSRQEKDHCHKQDGCHSSGGCHSSGSSCHSSGGSSCHSGGSSCHGSGGSSCHSSGGSSCHGGGSSCHSGGSCHGKPQQHCQQ
QQQQKICKVPCQKLK

>Cp_EDQM8
MCSRQEKDHCHKQDGCHSSGGCHSSGSSCHSSGGSSCHSGGSSCHGSGGSSCHSSGGSSCHGGGSSCHSGGSCHGKPQQHCQQ
QQQQKICKVPCQKLK

>Cp_EDQL
MCSREPRGCHDSGSSSCHDSGSSTCHSSGGGSCHDVKPLPQCPTPVPCQTTTLPCQQQTKQPCQWPPQKHQK

>Cp_EDWM
MIYSSGRESYFNLNSTWYDPAGSWLDTRRTPFTYAYSTCCSSGGCPRGGHDNRCYEYRRSGCGENCHGSSGSCHGSGGHCCVR
RPSYFHGYSGGCHGHGRSVCSERSCHGSGSSCHGSGSSCHGSGSSCHGSGSSCHNTSGACHSTPIYVKPKQYVQQCCPPVQQC
CLPVKKCCPPVQKC

>Cp_EDYM1
MSYFAYQYKQRNYTPYSTTRLIPHAEPCVVKGPAPRVTKCADPCAVKHPAPCTTKCRDPCAGKPSVPCATKCFEPHAQRHPAK
HYPKFSEPAGVKCSTPCDTRYHEPYGLIHPQPFPERWNPCAPPYVHPYVTGYPQACGPTYVPSFPKYPYPYAPQWPNTWGYGN
CGPC

>Cp_LOR_partial
MCSHQEKKACHEIPTQAGGCHAGGGGSSGSGSGALLGPHILGSSSYGVGGGSSYCGSGESCQKIIIAGGSSGSSSGSSGGSYG
CGVGGGSGGGSGGSYGCGVGGGSGGGSGGGSGQKIIIAGGESGGSYGCGVGGGSGGGSGGSGQKIIIAGGESGGSYXXXXXXXX
XXYGCGVGGGSGGGSGGSGQKIIIAGGGSGGSYGCGVGGGSGGGSGGSSGGCTGGGSGYSSGGSAGYCSGGSDDFGSGGSLQS
MQQKCPIVIPCMEQQQTKQPCQWPPNPRK

>Cp_PGLYRP3
MFRLAVFLCALCAVSWGFLCLRIVSPGKWGGRPANCSSPLKAVQTGYVIVLHTAGGSCKTGAECNQQMTNIQHYHMNNKGWCN
IAYNFLIGEDGKVYEGRGWNTEGAHTYGYNDISLGIAFMGDFTGRSPNAAAWIALKHLLHFAVENGYLSSDYLLMAHGDVSHT
ISPGQPIRKVLKTWPHYKH

>Cp_SCFN_partial
MPQLLDSISTLISVFYKPGKKDEDCSTISKREVKRFIQRGFADITVNPYDAHAIEAVLQLLDHDGDGVEDFNEFLLLVFRVAK
VCYWYLQPKQRLPQRTETGLSGERRQEPEAGRAEGSRDQPKPLGTEKGYYETWEPETRETERSRRQPREPEPRGDERANYETY
EPEKREEERRRRQPHGPEPRRDERSWYGTQRSEHRENERSHHRSCEPEPRGDERSRYERRDLETREVDRSRRHPHERERREDE
RSSHQPYEPEPQSRERRCLEPESRQDEKSGYWPRESERSRYETCEPEAREEERSHRQPREPEPRRDERSRSRPHEPQPQGDER
SHYRAHAPETREEERSRCPQQQPEPRRDERSRYEIYDHDTREEERSRCHPHEPEHEEEERSRHQPREPEHEEEERSADSHVSL
NXXXXXXXXXXSRRQPRKPESREDEGSRRQPREPEPREEERSRRQPREPEPREDEGSRRQPRKPEPREDEGSRRQPREPEPRE
DERRRRQPHEPEPREDNRSRRQPHGPELRGDGWSRHQSCEPEAREGERSRCEPCEPEQREEKRSHQQPREPQWREDERSRRQP
REPEPREDEGAADSHVSLNXXXXXXXXXXDEGSRRQPREPEPREDEGSRRQPREPEPREGERCSIQSQTPEPPHEEGNLPPSS
EPVENEGTLREPHDPESTDDSRRCPATHEPPPTGDEGSQLQPRESAAPGDEGNPQKTSESEPREDDGSQPQLCEPEQKEGDGV
LRQTPESQPLEEVESQEQPRGPEPRTEGSRHQPPQGGEASHLQPDIEPQQGDGSRHRPRDPQEERERSSHQARELPLPRLGEEP
EEGEWSQSHPEPANAEEEPGETDPDDVKASLPCNPLYVYLLEQTVEKQLYLAPPHQDHP

B

>Cp_EDbeta1
MSCGANLCIDGGSACGVARPRPCADSCNQPCVTQCPDSRVIIYPPPVVVTFPGPILTTFPQESVVESVGAPVVASGYGGTSGS
GAFGVGHGNCGPCGPC

>Cp_EDbeta2
MSCSRNVCTAGGSACGVARPRPFTDSCNQPCVTRCPDSRVIIYPPPVVVTFPGPILTTFPQESVVESVGAPVVASGYGGTSGS
GAFGVGHGNRDLCGPC

>Cp_Beta-A1
MSCSSLCYPECGVARPSPVSGSCNEPCVRQCPDSEVIIRPSPVVVTIPGPILSNFPQQSEVGAVGAPVVGAGYGGSFGLGGLY
GYGGHYGGLYGYGGLGGYGGRYGYGGLSGYGGLCGYGGRYGYGGLCGYGGRYGYGGLSGYGGLCGYGGGYGGGYGYGGACGSG
VSCHRYLSGSCTPC

>Cp_Beta-A2
MSCSSLCYPECGVARPSPVSGSCNEPCVRQCPDSEVIIRPSPVVVTIPGPILSNFPQQSEVGAVGAPVVGAGYGGSFGLGGLY
GYGGHYGGLYGYGGLGGYGGRYGYGGGYGGLCGYGGRYGYGGLSGYGGRYGGLCGYGGGYGGGYGYGGACGSGVSCHRYLSGS
CTPC

>Cp_Beta-A3
MSCSSLCYPECGVARTSPDSGSCNELCVRQCPDSEVIIRPSPVVVTIPGPILSNFPYRGHYGRLYCYGGLGGYGGHYGYGGLC
GYRGRYGYGGLCGYRGRYGYGGLSGYGGHYGGLCD

>Cp_Beta-A4
MSCSSLCYPECGVARPSPVSGSCNEPCVRQCPDSEVIIRPSPVVVTIPGPILSNFPQQSEVGAVGAPVVGAGYGGSFGLGGLY
GYGGHYGGLYGYGGLGGYGGRYGYGGGYGGLCGYGGRYGYGGLSGYGGLCGYGGGYGGGYGYGGACGSGVSCHRYLSGSCTPC

>Cp_Beta-A5
MSCSSLCYPECGVARPSPVSGSCNEPCVRQCPDSEVIIRPSPVVVTIPGPILSNFPQQSEVGAVGAPVVGAGYGGSFGLGGLY
GYGGHYGGLYGYGGLGGYGGHYGYGGLSGYGGLCGYGGRYGYGGLSGYGGRYGGLCGYGGGYGYGGACGSGVSCHRYLSGSCT
PC

>Cp_Beta-A6
MSCSSLCYPECGVARPSPVSGSCNEPCVRQCPDSEVIIRPSPVVVTIPGPILSNFPQQSEVGAVGAPVVGAGYGGSFGLGGLY
GYGGHYGGLYGYGGLGGYGGRYGYGGGYGGGYGGLCGYGGRYGYGGLSGYGGRYGGLCGYGGGYGGGYGYGGACGSGVSCHRY
LSGSCTPC

>Cp_Beta-A7
MSCSSLCYPECGVARPSPVSGSCNEPCVRQCPDSEVIIRPSPVVVTIPGPILSNFPQQSEVGAVGAPVVGAGYGGSFGLGGLY
GYGGHYGGLYGYGGLGGYGGHYGYGGLSGYGGGYGGLCGYGGRYGYGGLSGYGGRYGGLCGYGGGYGGGYGYGGACGSGVSCH
RYLSGSCTPC

>Cp_Beta-A8
MSCSSRCYPECGVARPSPVSGSCNELCVRQCPDSEVIIRPSPVVVTIPGQILSNFPYGGHYGRLYGYGGLGGYGGHYGYGGLC
GYGGRYGYGGLSGYGGHYGGLCD

>Cp_Beta-A9
MSCSSLCYPECGVARPSPVSGSCNEPCVRQCPDSEVIIRPSPVVVTIPGPILSNFPQQSEVGAVGAPVVGAGYGGSFGFGGLN
GYGGHYGGLYGYGRGYGGLSGYGGGYGGLCGYGGHYGYGGLSGYGGRYGGLCGYGGGYGYGGACGSGVSCHRYLSGSCTPC

>Cp_Beta-A10
MSCSSMCYPECGVTRPSPVSGSCNEPCVRQCPDSEVIIRPSPVVVTIPGPILSNFPQQSRVGAIGAPVVGPGYGASFSLGGLY
GSGSCYGGLYSYGILYGYGGLGGYRGWLWGIRSLWGIM

>Cp_Beta-A1L1
MSCSSLSYPECGVARPSPVSGTCNEPCIRQCPDSEVVIRPSPVVVTLPGPILSTFPQQSEVAAVGAPVVGAGYGGSFGLGGLY
GSGGRYGGLYGLGGFGGYGGLYGYGGLGGYGGLCGYGGGYGYGGLGGYGGLGRYGGLCGYGGGYGGLCGYGGGYGYGGLGGYG
GLCGYGGYGRRYRGGYCGPC

>Cp_Beta-A1L2
MSCSSLSYPECGVARPSPVSGSCNEPCVRQCPDSEVVIRPSPVVVTLPGPILSTFPQQSGVGAVGAPVVGAGYGGSFGLGGLY
GSGGHYGGLYGLGGLGGYGSHYGYGGLGGYGGLCGYGGGYGYGGLGGYGGLGRYGGLCGYGGGYGGLCGYGGGYGYGGLGGYG
GLCGYGGYGRRYRGGYCGPC

>Cp_Beta-B1
MSFNGPQTGAQGSLPCGVKCSEPYIATASEPCVVKCKDSRVIIYPPPVVVTFPGPILTTCPQESIVASSGPPDTGVAESAARI
SAAPRVTGSLGPHLDRCPASINIRHEAQYTPKYSYTYSSRWSHPGKSLETTGYSQTTRNIDRTK

>Cp_Beta-B2
MSCYGLRNIPCEVPRPTPAAVTYNEPCVIQCPDSIFESDSPPGIAIIPGPILTTFPHYSVVETSPLFDTERSFCSERSLGSQG
FMNLYN

>Cp_Beta-B3
MFSDEEFFYKNKQPQKPQKGQNPCLPQKKPKPCPPQKVPCTPKPPRCPTYPPYIPPPRPCPPQYPSICPQPYIGVWNEPCVTE
CGDSTAVVFAPPVVVNFPGPTLATCPQDSVVGSSLPRGIIGPYGPGGSLSSGGAFGTGSSFGSSVSSFSSAGSFGSGGYGGSG
GYLGSGGFLGSGGYCGSGGWNPCHYGRCGPC

>Cp_Beta-B4
MFSDEEFWYKSQQPQKPQKGQNPCLPQKKPKPCPPQKVPCTPKPPRCPTYPPYIPPPRPCPPQYPSIWPQPYIGVWNEPCVTE
CGDSTAVVFAPPVVVNFPGPTLATCPQDSVVGSSLPRGIIGPYGSGGSLGSGGSFGAGSSFGSSVSSFSSGGSFGSGGYGGSR
GFLGSGGYGGSGGYCGFGGWNPCHYGRCGPC

>Cp_Beta-B5
MSSYEQLCNTQCYAPCNVTCPQPIVDTCNEPCITSCSDSRAVVYPPLIVVTFPGTLLSFCPQESVEESSAHVGIRSS

>Cp_Beta-B6
MSSYRQLCNTQCYAPCNVTCPRPFVDACNEPCFTSCGDSSAVLYPPPVIVRFPGPILATCPQESVVGSSAPFGIGSSLGIGGP
YVSGSLGNYGGSYTSGLSARGNGCSYPSSSSQRFTTYRSGSCQPYQTQK

>Cp_Beta-B7
MSSYRQLCNTQCYAPCNVTCPRPFVDACNEPCFTSCGDSSAVLYPPPVIVRFPGPILATCPQESVVGSSAPFGIGSSLGIGGP
YVSGSLGNYGGSYTSGLSARGNGCSYPSSSSQRFTTYRSGSCQPYETQK

>Cp_Beta-B8
MSFCRDLCKYPSYPSCDVTCPQPFVDACNQPCVTSCGDSSVVVYPPPVVVRFPGPILATCPQESVVGSSEPLGIGSSFGYRGS
YLSGSSYGYKSLYNDRRSYTPGLSSLGRGSSDPCSSRWLNMYGCGPRQTQQE
>Cp_Beta-B9
MSFYGDPARSQCYLPCEGTCQQPVANVCNEPWVRSGGDSRGVGYAPLVVVTFPGPSSQYLLSGKHDWNGTAK

>Cp_Beta-B10
MSSHRQLVSPRCATPWEVTCPQPGANICSQPCVTSCEDSRVMVYAPPVVVAFPGPILSTCPQKSITGSEVPGEMGAYLDLEGH
MVLGAHMVSGLPMALGKHMVMKSHLVLGDHMVQTTHGSGGSYGHGGSYSASNSYGARGPYGAGRFLVFRGSYGSGGSYSHSRA
YTSRLSPLGTGNSCPYSSQRTSMSHYKNCGPY

>Cp_Beta-B11
MSSCKDLSCRPSPCYPDICPDPCVVARNEPCITSCADSTAVVYPPPVSVLFPGPILSSSPQHSLVGSTLPALPYGARGSFGGG
ALGGPIGYGSGYGGALEGGYGYGGLSGYGGSYGYGGLSGYGGSYGYGGLCGYGGGYGGGYGGLCGYGRRYGGRCYSSRRGSCG
PC

>Cp_Beta-B12
MSSCKDLSCRPSPCYPDICPDPCVVARNEPCITSCADSTAVVYPPPVSVLFPGPILSSSPQHSVVGSTLPALPYRAGGSFGGG
ALGGPIGYGSGYGGALEGGYGYGGLSSYGGSYGYGGLSGYGGSYGYGGLCGYGAGYGGGYGGLCGYGAGYGGGYGGLCGYGRR
YGGRCYSSRRGSCGPC

>Cp_Beta-B13
MSSCKDLSCRPSPCYPDICPDPCVVARNEPCITSCADSTAVVYPPPVSVLFPGPILSSCPQHSLVGSTLPALPYRAGGSFGGG
ALGGPIGYGSGYGGALEGGYGYGGLSSYGGSYGYGGLCGYGGGYGGGYGGLCGYGRRYGGRCYSSRRGSCGPC

>Cp_Beta-B14
MSSSKDLCYPRPPCYPDICPNPYVDAWNEPCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSVVGSTLPNLPYGYGSPYGGG
SFGGSVSSGGAYGGGYGAGYGGGYGGLYGYGKGYGRKCYSSRFGSCGPC

>Cp_Beta-B15
MSSSKALCYPRPPCYPDICPDPYVDACNEPCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSVVGSTLPNLPYGYGGSYGGG
SFSVSVGSGGAYGGGYGARYGGGYGGLYGYGKGYGRKCYSSRFGSCGPC

>Cp_Beta-B16
MSSSKALCYPRPPCYPDICPDPYVDACNEPCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSVVGSTLPNLPYGYGGSYGGG
SFSGSVGSGGAYGGGYGARYGGGYGGLYGYGKGYGRKCYSSRFGSCGPC

>Cp_Beta-B17
MSSSKALCYPRPPCYPDICPDPYVDACNEPCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSVVGSTLPNLPYGYGGSYGGG
SFSGSVGSGGAYGGGYGAGYGGGYGGLYGYGKGYGRKCYSSRFGSCGPC

>Cp_Beta-B18
MSSSKDLCYPRPSCYPDICPNPYVDAWNEPCVTSCGDSSAVVYAPPVVVRFPGPILSTCPQDSVVGSTLPNLPYGYGGSYGAG
SFGGSVISGGAYGGRYGAGYGGGYGGLYGYGKGYGRKCYSSRFGSCGPC

>Cp_Beta-B19
MSSSKALCYPRPPCYPDICPDPYVDAWNEPCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSIVGSTLPNLPYGYGGSYGGG
SFGGSVGSRGAYGGGYGAGYGGGYGGLYGYGKGYGRKCYSSRFGSCGPC

>Cp_Beta-B20
MSSSKALCYPRPPCYPDICPNPYVDACNEPCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSVVGSTLPNLPYGYRGSYESG
SFGGSVGSRGAYGGGYGAGYGGGYGGLYGYGKGYGRKCYSSRFGSCGPC

>Cp_Beta-B21
MSSSKALCYPRPPCYPDICPDPYVDAWNEPCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSVVGSTLPNLPYGYGGPYGGG
SFSGSVGSGVAYGGGYGAGYGGGYGGLYGYGKGYGRKCYSSRFGSCGPC

>Cp_Beta-B22
MSSSKALCYPRPPCYPDICPDPYVDAWNEPCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSVVGSTLPNLPYGYRGSYGGG
SFGGSVGSSGVYGGVYGGVYGGGYDGGYGGLCGYGRRYGRKSYSSRFGSCGPC

>Cp_Beta-B23
MSSSKALCYPRLPCHPDICPNPYVDAWNESCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSVVGSTLPNLPYRYEGPYGGG
SFGGSGGSGRAYGGRYNVGYGSRYGDLCGHGRRFGRKCYSSRFESCRPC

>Cp_Beta-B24
MFAEFLCWQPRPYCPDICPDPCAYVCNKPFVTSCGDSNGVVYAPPDVMRFPGQALTTCPQDSFVGTVGPRLFYSSWGGLGGHA
GGGFGGGSWGWYGGGSGGGFRGGYGGSFRGGYGGRYGSGYGGGYGGGYGGGFGSGSGGRFGGGLGGSYGYGGSYGGNSYANRW

>Cp_Beta-B25
MSPVKDLCCQPGPYYPDICPDPCAYVCNEPCVTTCGDSNAVVFAPPVVVRFPGPTLATCPQDSFVGTSLPNFPYRLGGGLGGR
IGGGLGGGYGGGYGGGYGGGSFGGFGGGFGGGFGGGIGGGYRGSYGYGGRYGRNCYANRWECCPW

>Cp_Beta-B26
MSPVKDLCYQPRPYCPDICPEPCAYVCNEPCVTTCGDSNAVVFAPPVVVRFPGPTLATCPQDSFVGTSLPNFPYRLGGGLGGR
IGGGLGGGYGGGYGGGYGVGSFGGFGGGSGGGFGGGFGGGIGGGYRGSYGYGGRYGRNCYANRWECCPW

>Cp_Beta-B27
MSPVKDLCCQPRPYCPDICPDPCAYVCNEPCVTSCGDSSAVVFAPPVVVRFPGPTLATCPQDSFVGTSLPNFPYTPWGGLGGR
AGAGLGGGSWSGYGGGFGVGAGGGFGGGFGGGFGAGYGGSYGNWGRYGRRCSYTSYRCPC

>Cp_Beta-B28
MSCCPPQDCIPDICPCPYIDVRNEPCISSCGDSTAVVYAPPVVVNFPGPTMATCPQDSFIGTSLPNMPVRAGASYGSGGGFSG
SIGSRGSYGAGFGGGYGGGFRGGYGGGHGGGYGGGFGGGLRCGYGGSYGYGGPCGYGRRSQRGITVLGGGYSGSSYGNCGPC

>Cp_Beta-B29
MSYCPPQDCIPDICPRPYIDVCNEPCISSCGDSNAVVFAPPVVVRFPGPTMATCPQDSIVGSSLPNMPIRAGGLYGSGGGFSG
SITSGGSYGGGFGGGYSGGYGGGSSIVYGGGAGGGYGGGAGGGYGAGYGGSYGCGGSRGYIRKSYRSISGGGYSGFNRGNCGP
C

>Cp_Beta-B30
MSYCPPQDCIPDICPRPYIDVCNEPCISSCGDSNAVVFAPPVVVRFPGPTMATCPQDSIVGSSLPNMPIRAGGSYGSGGGFSG
SITSGGSYGGGFGGGYSGGYGGGSSIVYGGGAGGGYGGGAGGGYGAGYGGSYGCGGSRGYIRKSYRSISGGGYSGFNRGNCGP
C

>Cp_Beta-B31
MSYCPPQDCIPDICPRPYIDVCNEPCISSCGDSTAVVFAPPVVVRFPGPTMATCPQDSFVGSSLPNMPIRAGASYGSGGGFSG
SITSGGSYGGGLGGGYSGGYGGGSSIVYGGGAGGGYGGGAGGGYGAGYGGSYGCGGSRGYIRKSYRSISGGGYSGFNRGNCGP
C

>Cp_Beta-B32
MSYCPPQDCYPDICPRPCIDVRNEPCISSCGDSTAVVYAPPVVVRFPGPTMATCPQDSFVGSSLPNLPIRPGGSYGGSISYGG
GYGGGYGGGNSVVGSGGGFGGSTGYGGVYGGGAGSGYGGGYGGGAGGGYGGCYGGSYGSGGSRGYSKKSYRSISGGGYSGVNR
GNCVQS

>Cp_Beta-B33
MSYCPPQDCYPDICPRPCIDVRNEPCISSCGDSTAVVYAPPVVVRFPGPTMATCPQDSFVGSSLPNLPIRPGGSYGGSLTYAG
GYGGGYGGGNSVVGSGGGFGGSTGYGGVYGGGVGGGYGGGYGGGAGGGYGGGAGGGYGGCYGGSYGSGGSRGYSKKSYRSICG
GGYSGVNRGNCGPC

>Cp_Beta-B34
MTYCPPQDCYPDICPRPCIDVRNEPCISSCGDSTAVVFAPPVVVRFPGPTMATCPQDSFVGSSLPNLPIRPGGSYGGSITYGG
GYGGGYGGGNSVVGSGGGFGGSTGYGGVYGGGVGGGYGGGYGGGAGGGYGGGAGGGYGGCYGGSYGSGGSYGYNKKSYRSISG
GGYCGVKSGNYGPC

>Cp_Beta-B35
MSYCPPQDCYPDICPRPCIDVRNEPCISSCGDSTAVVFAPPVVVRFPGPTMATCPQDSFVGTSLPNLPIRPVGSNGGSISYGG
GYGGGYGGGNSVVGSGGGFGGSTGYGGVYGGGVGGGYGGVYGGGVGGGYGGGYGAGVGCGYGGGAAGGYGGCYGGSYGSGGSR
GYSKKSYRSISGGRSSGVKSGNPEKPSME

>Cp_Beta-B36
MSFNGVPCNDQCHNPCEVTCPQPIVNSSNQPCVVSCGDSRVVIYPPPVVVTFPGPILSTCPQDSIVGSSAASGSRISGSASVI
SSTPGVTGCSKPYAESVFVRSEPQYTPKYSYTYSSPWIHPGNTSGSGHYRSSYVQKACPRNEEPQQNEKQDAEQCS

>Cp_Beta-B17L
MSSSKALCYPRPPCYPDICPDPYVDACNEPCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSVVGSTLPNLPYGYGGSYGGG
SFSGSVGSGGAYGGGYGARYGGGYGGLYGYGKGYGRKCYSSRFGSCGPC

>Cp_Beta-B18L
MSSSKALCYPRPPCYPDICPNPYVDAWNEPCVTSCGDSSAVVYAPPVVVRFPGPILATCPQDSVVGSTLPNLPYGYRGSYGSG
SFGGSVSSGGAYAGGYGAGYGGGYGGLYGYGKGYGRKCYSSRFGSCGPC

>Cp_Beta-O1
MTFSSLCYPECGVARPSPVTGSSNEPCVRQCPDSEVVIRPSPVVVTLPGPILSNFPQQSEVAAVGAPVVGAGFGGSFGLGGLY
GYGGHYGGLYGLGRLGGYGGRYGYGGLLGYGGHCGYPGLYGYGGLWGYGGYGRRYLSGYCGPC

>Cp_Beta-O2
MTFSSLCYPECGVARPSPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPILSNFPQQSEVAAVGAPVVGAGFGGSFGLGGLY
GYGGHYGGLYGLGRLGGYGGHYGYGGLLGHGGYCGYPGLYGYGGLWGYGGYGRRYLGGYCGPC

>Cp_Beta-O3
MTFSSLCYPECGVARPSPVTGSANEPCVRQCPDSEVVIRPSPVVVTLPGPILSNFPQQSEVAAVGAPVVGAGFGGSYGLGGLY
GYGGHYGGLYGLGRLGGYGGRYGYGGLLGNGGYCGYPGLYGYGRLWGNGGHCGYPGLYGYGGLWGYGGHCGYPGLYGYGGLSG
SGVSNHRYLSGSSGPC

>Cp_Beta-O4
MTFSSLCYPECGVARPSPITGSSNEPCVRQCPDSEVVIRPSPVVVTLPGPILSNFPQESEVAAVGAPVVGAGFGGSYGLGGLY

GYGGHYGGLYGLGRLGGYGGHYGYGGLLGYGGHCGYPGLYGYGGLWGYGGYGRRYLGGYCGPC

>Cp_Beta-O5
MTFSSLCYPECGVARPSPVTGSSNEPCVRQCPDSEVVIRPSPVVVTLPGPILSNFPQESEVAAVGAPVVGAGFGGSFGLGGLY
GYGGHYGGLYGLGRLGGYRGLYGYGRLLGHGGYCGYPGLYGYGGLWGYGGYGRRYLGGYCGPC

>Cp_Beta-O6
MTFSSLCYPECGVARPSPVTGSCNEPCVRQCQDSQVVIRPSPVVVTLPGPIMSNFPQHSAVGAVGAPVVGAGFGGSYGLGGLN
GSGGQYGGLSGLGGYGGYGGLCGSGVSCHRYLSGSGGLC

>Cp_Beta-O7
MTFSSLCYPECGVARPSPVTGSCNEPCVRQCQDSQVVIRPSPVVVTLPGPIMSNFPQHSAVGTVGAPVVGAGFGGSYGLGGLN
GSGGQYGGLSGLGGYGGYGGLCGSGVSCHRYLSGSCGPC

>Cp_Beta-O8
MTFSSLCYPECGVAQPSPVTGSCNEPCVRQCPDSQVVIRPSPVVMTLPGPILSNFPQHSVVGAVGAPVVGAGFGGSYGLGGLN
GSGGHYGGWSGLGGYGGYGGLSGSGVSYHRYLSGSCGPC

>Cp_Beta-O9
MTFSSLCYPECGVARPSPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAIGAPVVGPGFGGSFGHGGFG
YGGLYGGLHGLGGYGGYGGHYGYGGLGGYLGGYGYGGLCGSGVSCHRYLSGNCGPC

>Cp_Beta-O10
MTFSSLCYPECGVARPSPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGFG
YGGLYGGLHGLGGYGGYGGHYGYAGLGGYLGGYGYGGLCGSGVSCHRYLSGNCGPC

>Cp_Beta-O11
MTFSSLCYPECGVARPSPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGFG
YGGLYGGLYGLGGYGGYGGHYGYGGLGGYLGGYGYGGLCGSGVSCHRYLSGNCGPC

>Cp_Beta-O12
MTFSSLCYPECGVARPSPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGLG
YGGLYGGLYGLGGYGGYGGHYGYGGLGGYLGGYGYGGLCGSGVSCHRYLSGNCGPC

>Cp_Beta-O13
MTFSSLCYPECGVARPSPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGFG
YGGLYGGLYGLGGYGGYGGHYGYAGLGGYLGGYGYGGLCGSGLSCHRYLSGNCGPC

>Cp_Beta-O14
MTFSSLCYPECGVARPSPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGRGGFG
YGGLYGGLYGLGGYGGYGGHYGYGGLGGYLGGYGYGGLCGSGLSCHRYLSGNCGPC

>Cp_Beta-O15
MTFSSLCYPECGVARPSPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGFG
YGGLYGGLYGLGGYGGYGGHYGYGGLGGYGGLGGYLGGYGYGGLCGSGVSCHRYLSGNCGPC

>Cp_Beta-O16
MTFSSLCYPECGVARPSPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGFG
YGGLYGGLYGLGGYGGYGGHYGYGGLGGYGGLGGYLGGYGYGGLCGSGVSCHRYLSGNCGPC

>Cp_Beta-O17
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPILSNFPQHSGVGALGAPVVGPGFGGSFGHG
GFGYGGLYGGLYGLGGYGGYGGRYGYGGLWGHGGYCGYPGLYGYGGLLGYGGYGRRYLGGRCGPC

>Cp_Beta-O18
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPILSNFPQHSAVGALGAPVVGPGFGGSFGHGALY
GYGGLYGGWCGLGGYGGYCGPYGYGGLGGYLGCYGYGGICGSGVSCHRYLSGSCGPC

>Cp_Beta-O19
MTFSSLCYPECGVAQPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSHYTQESVVGALGAPVVGPGFGGSFGHGGLY
GYGGRYSGWYGLGGYGGYCGPYGYGSLGGYGGLGGYVGGYGYGGICGSGVSCHRYLSGSCGTC

>Cp_Beta-O20
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPILSNFPQHSAVGAVGAPVVGPGFGGSFGHGGYG
YGGLYGGLYGLGGYGGYGGHYGYGGLWGHGGYCGYPGLYGYGGLWGYGGYGRRYLGGHCGAW

>Cp_Beta-O21
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGYG
YGGLYGGLYGLGGYGGYGGHYGYGGLLGHGGYCGYPGLYGYGGLWGYGGYGRRYLGGRCGTW

>Cp_Beta-O22
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPILSNFPQHSAVGAVGAPVVGPGFGGSFGHGGYG
YGGLYGGLYGLGGYGGYGGHYGYGGLWGHGGYCGYPGLYGYGGLWGYGGYGRRYLGERCGTC

>Cp_Beta-O23
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGFG
YGGLYGGLYGLGGYGGYGGHYGYGGLWGHGGYCGYPGLYGYGGLWGYGGYGRRYLGGRCGTC

>Cp_Beta-O24
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGFG
YGGLYGGLYGLGGYGGYGGHYGYGGLWGHGGYCGYPGLYGYGGLWGYGGYGRRYLGGRCGTC

>Cp_Beta-O25
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGFG
YGGLYGGLYGLGGYGGYGGHCGYGGLWGHGGYCGYPGLYGYGGLWGYGGYGRRYLGGRCGTC

>Cp_Beta-O26
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGFG
YGGLYGGLYGLGGYGGYGGHYGYGGIIGPWGILRLPGSLWLRGIMGIWGIWP

>Cp_Beta-O27
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGYG
YGGLYGGLYGLGGYGGYGGHYGYGGLWGHGGYRGYPGLYGYGGVMGIWGIWP

>Cp_Beta-O28
MISSSLCYPECGVARPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPILSNFPQHSAVGALGAPVVGPGFGGSFGHGGYG
YGGLYGGLYGLGGYGGYGGRYGYGGLYGFGGLGGYGYGGLCGSRLSCHRYLSGNCGPC

>Cp_Beta-O29
MISSSLCYPECGVARPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPILSNFPQHSAVGAVGAPVVGPGFGGSFGHGGFG
YGGLYGGLYGLGGYGGYGGRYGYGGLYGFGGLGGYGYGGLCGSRLSCHRYLSGNCGPC

C

>Cp_Beta-p1
MSCSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVVVTIPGPILSNFPQQSEVGAVGAPVVGAGYGGSFGLGGLN
GYGGHYGGLYGXXXXXXXXXX

>Cp_Beta-p2
MSCSSLCYPECGVARPSPVSGSCNEPCVRQCPDSEVVIRPSPVVVTLPGPILSTFPQQSEVAAVGAPVVGAGYGGSFGLGGLY
GSGGYYGGLYGLGGFGGYGSHYGYGGLGGYGGLGGYGGGYGGXXXXXXXXXX

>Cp_Beta-p3
MSPVKDLCCQPGPYCPDICPDPCAYVCNEPCVRTCGDSSAVVFAPPVVVRFPGPTLATCPQDSFVGTSLPNFPYTPWGGLGGR
AGAGLGGGSWSGYGGGFGVGARGGLGGGFGAGFGGGFGGGFXXXXXXXXXX

>Cp_Beta-p4
MSCSSLSYPECGVARPSPVSGSCNEPCVRQCPDSEVVIRPSPVVVTIPGPILSNFPQQSEVGAVGAPLVGAGYGGSFGLGGLN
GYGGHYGGLYGXXXXXXXXX

>Cp_Beta-p5
MTFSSLCYPECGVARPSPVTGSFNEPCVRQCPDSEVVIRPSPVVVTLPGPILSNFPQESEVAAVGAPVVGAGFGGSFGLGGLY
GYGGHYGGVDGLGGFSRXXXXXXXXXX

>Cp_Beta-p6
MTFSSLCYPECGVARPSPVTGSANEPCVRQCQDSEVVIRPSPVVVTLPGPILSNFPQQSEVAAVGAPVVGAGFGGSFGLGDCT
AMXXXXXXXXXXCWAFFI

>Cp_Beta-p7
MSCSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVVVTIPGPILSNFPQQSEVGAVGAPVVGAGYGGSFGLGGLN
GYGGHYGGLYGLGGFGGYGGLYGYGGLGGYGGLCGYGGGYGYGGLGRYGGLCGYGGGYGGLCGYGGGYGYGGLGGYGGLCGYG
GGYGYGGLGRXXXXXXXXXX

>Cp_Beta-p8
MSCSSLSYPECGVARPSPVSGSCNEPCIRQCPDSEVVIRPSPVVVTLPGPILSTFPQQSEVAAVGAPVVGAGYGGSFGLGGLY
GSGGHYGGLYGLGGFGGYGGLYGYGGXXXXXXXXX

>Cp_Beta-p9
MSCSSLSYPECGVARPSPVSGSCNEPCVRQCPDSEVVIRPSPVAVTLPGPILSTFPQQSEVAAVGAPVVGAGYGGSFGLGGLY
GSGGHYGGLYGLGGFGGYGGLYGYGGLGGYGGLCGYGGGYGYGGLGGYGGLGRYGGLCGYGGGYGGLCGYGGXXXXXXXXXX

>Cp_Beta-p10
MSCSSLSYPECGVARPSPVSGSCNEPCVRQCPDSEVVIRPSPVAVTLPGPILSTFPQQSEVAAVGAPVVGAGYGGSFGLGGLY
GSGGQYGGLYGLGGFGGYGGLYGYGGLGRYGGLCGYGGGYGYGGLGGYGGLGRYGGLCGYRGGYGYGGLGGYGGLGRYGGLCG
YGGGYGYGGLGGYXXXXXXXXXX

>Cp_Beta-p11
XXXXXXXXXXCGVARPSPVSGSCNEPCVRQCPDSEVVIRPSPVVVTLPGPILSNFPQQSEVAAVGAPVVGAGFGGSYGLGGLY
GYGGHYGGLYGLGGYGGYGGRYGYGGLLGNGGYYGYPGLYGYGGLWGNGGHCGYPGLYGYGRLWGYGGHCGYPGLYGYGGLSG
SGVSCHRYLSGSRGPC

>Cp_Beta-p12
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGAPVVGPGFGGSFGHGGXX
XXXXXXXX

>Cp_Beta-p13
MTFSSLCYPECGVARPCPVTGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPILSNFPQHSGVGALGAPVVGPGFGGSFGHGGXX
XXXXXXXX

>Cp_Beta-p14
MTFSSLCYPECGVARPSPVTGTCNEPCVRQCQDSEVVIRPSPVVVTLPGPIMSNFPQHSGVGAVGPPVVGPGFGGSFGHGGFG
YGGLYGGLHGLGGYGGYGGHYGYGGLGGYLXXXXXXXXXX

>Cp_Beta-p15
XXXXXXXXXXVARPSPVTGSSNEPCVGQCPDSEVVIRPSPVVVTLPGPILSNFPQQSEVAAVGAPVVGAGFGGSFGLGGLYGY
GGHYGGLYGLGRLGGYRGLYGYGGLLGHGGYCGYPGLYGYGGLWGYGGYGRRYLGGYCGPR

>Cp_Beta-p16
XXXXXXXXXXACNEPCVRQCQDSEVVIRPSPVVVTLPGPILSNFPQHSAVGALGAPVVGPGFGGSFGHGGFGYGGLYGGLYGL
GGYGGYGGHYGYGGLWGHGGYCGYPGLYGYGGLWGYGGYGRRYLGGRCGTC

>Cp_Beta-p17
XXXXXXXXXXNPSPVVMTLPGPILSNFPQHSVVGAVGAPVVGAGFGGSYGLGGLNGSGGHYGGLSGLGGYGGYGGLCGSGVTC
HRYLSGSCGPC

Supplementary Figure S1. Amino acid sequences of proteins encoded by EDC genes of *C. picta bellii* (Cp). (A) Amino acid sequences of EDC proteins other than beta-proteins. (B) Amino acid sequences of beta-keratins of completely sequenced genes. (C) Amino acid sequences of beta-keratins of partially sequenced genes. Amino acid residues K and Q (potential transglutamination sites), C (potential disulfide bonding sites), P, G and S are highlighted by specific colors corresponding to those in Figure 3. XXXXXXXXXX, missing amino acid residues (number unknown).

A

>Cm_CRNN
MTQLLSNIKGIINAFYVSAKKDGACPTLSKGELRQLICQEFADVTVVPPGLQTIDKMLQLLDTDSDGRLDFNGFLVLVFQVAK
ACYREVSQGQQPGHGERSASPGEAKCGEHTQEPQTPERDPTPRQAPEFQISERDPSPHQAPEPQTPEHDPSPHQALKPQTPER
DPSPHQVLEPQKPEQDLSPRQALEPPTPERDLSPRQALEPQMSEWDPSPRQAPEPQKPEQNPSSHQAPEPQTPKQDSSPRQTP
EPQTPERDLSPRQAPEFQISEQDPSPHQAPKPQTPEQDPSPRQALEPQMSEWDPSPRQAPEPQKPEQDPSSHQAPEPQTPKQD
SSPCQAPEPQTPEQDPSPQQVREPQTQEQDPSLRQAPEPQTPEQDSSPRQASEPQIPEQNPSPHQDLEPQTPEQDLNHSETEL
PPTQQRNQGVQDPREPAAGQASKSSQCLYSWHYQKPLPFPHWWPPKK

>Cm_EDAA1
MTYHHQKISHHWGCDPRWNGGWGGYRGHYDCYRPWGYYRPYSYGWCPNYDSCYSYPYRWGSGYGYGRCWPCFAEEQ

>Cm_EDAA2_partial
MTFDESINEELFYNPWSHGCWHGSRGHYGCGRPWGYGRQSRLGWGHGYDCYYPYSSRWGHWYPYGXXXXXXXXXX

>Cm_EDAA3
MTFDELMNDELYYNPYCYQGWRGYRGHYGCYRPWGSWKPYRYGWGHQYGGHYPYRWGHGYGYGKFWPCLAEEQ

>Cm_EDAA4
MTYHHQKISHHWGCDPCSYGSWGGYRGHYDCYRPWSYSRPYGYGCSYNDGCYYPYSSRWGHGYGGYGYGSHGYGKCWQE

>Cm_EDAA10L
MTWSGYGCNDGCYSPCGYGGQWACGSPCGYRGLCGYGGRSGHGGSCGYRGSYGYRGSYDSGHCYPFAFQRGHRYSYGNCGPC

>Cm_EDAAO1
MTFYENFSDELYYKPYYYGGWGGSRGYGYCRPWCYRRPYKCCWGYPKGCWHPYPCHWGWGHGYGKGWPCFAEEE

>Cm_EDAAO2_partial
MTFDENFSDELYYKPYYSGARGGSRCNGYHKPCCYQRPXXXXXXXXXX

>Cm_EDbeta1
MSCGGNLCIDGGSACGVARPKPYTDSCNQPCVTQCPDSRVVIYPPPVVVTFPGPILTTFPQESVVESVGAPVVASGYGGTSGS
GAFGVGHGNCGPCGPC

>Cm_EDbeta2
MSCSRNVCTAGGSACGVARPRPFADSCNQPCVTQCPDSRVVIYPPPVVVTFPGPILTTFPQESVVESVGAPVVASGYGGTSDS
GAFGVGRGNRDLCGPC

>Cm_EDKM
MSALIKAIADMIDSYQRNAKKGCESERIRRCEFKKLLQQEPSPAKISSSNKYEHTTSLPDSDAELMIKKELITANPCVY

>Cm_EDP1
MPYYGQQHKQLCLPPPACVTKCSQPYPPQYEQQCVPKCRPVYVTKCPPWYGPQYAYPCAPQCPPPCVTKCPPPCPPPCVTKCP
PPCPPPCVTKCPPPCVTKCPQHCVTQYPDQYQSGKVQISSHGKKYCSGPKWPW

>Cm_EDP2
MASQQNQQQRKQTLTLPLALSNATSEPAPTPEAGPEPCPATVEERENSPQEEEESQEEYKRPLNQPLGPAPELEPEPVLGPEP
ESNPSEVKEIEYLQLDQQQYKHPPTLPPAPGIETSKEYQQAEPELEPEPGRCPPPISEAEGPLFVQPSSPVEEQQQKQPHHWP
PKRK

>Cm_EDP3
MSADQQQCKQTCPPPPKCQEKCPPPCKEPCKPSKCQEQCPPPCKDPCPPKCPPPQQSQDWKHC

>Cm_EDP3L_partial
XXXXXXXXXXDQQQCKQTCPPPPKCQEKCPPPCKEPCKPSKCQEQCPPPCKDPCPPKCPPPQQSQDWKHC

>Cm_EDPCV1
MVYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVKKCVDPCCNVCVKKCTTKCLDPCCKVCVTKCVKKCVDPCCKVCVK
KCTTCVHPCPCPCPQKCPPCPPCFPKCPPVEHCCKEKKFW

>Cm_EDPCV2
MVYQQQCKQPCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPEKCIPCPEKCIPCPQKCP
PCPPCPEKCPPVQHCCKEKKLC

>Cm_EDPCV3
MVYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTSCVHPCPCPCPQKCIPCPQKCIPCPQKCI
PCPPCPPCLPKCPPVQHCCKEKKFW

>Cm_EDPCV4
MAYQQQCKQPCLPPPCCVTQCTTKCLDPCCKVCVTKCVTKCMDPCCKVCVKKCTTCVHPCPCPPKHVDPCPLCFPKCPPVQHC
CKEKKLC

>Cm_EDPCV5
MSYQHQCKQPCLPPPCCVKQCKTKCVDPCPCPPKCVDPCPPKCMDPCPPKCVDPCPLKCVDQWPCPPKCVDLCPPKCVDPCPP
KCVDPCPPCPPLQHCCKEKKFC

>Cm_EDPCV6_partial
MVYQLQCKQTCLPPPCCVTKCTTKCLDPCCKVCVTKCVKKCVDPCCKVCVXXXXXXXXXX

>Cm_EDPCV7
MVYQQQCKQPCLPPPCCVTKCTTKCLDPCCKVCVTKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPEKCIPFPEKCIPCPEKCI
PCPQKCIPCPEKCIPCPPCPEKCPPVQHYCKEKKIC

>Cm_EDPCV8
MVYQQQCKQTCLPPPCCVTKCTTKCLDPCCKVCETKCVTKCVDPCCKVCVKKCTTCVHPCPCPCPQKCPPKCPPVQHCCKEKK
FW

>Cm_EDPL1-like
MSCQQPQQQCKQPCMPPPCKEPGHPHTTDPGPQQCTEPRAQLCPEPCPPKCVETCSPKCLSSAQSHVPSCAQSHVLQNAWRHV
PQNAHLYSSARNSSVTNKIHSSLARKKPIPNAGSQVM

>Cm_EDQL_partial
MCSREPHGCHDTGSSSCHDTGSSSSPDTGSSFCXXXXXXXXXXSSSCHDTGSSSCHGSGGGTCHDVKPLPQCPIPVPCQTTTI
PCQQQTKQPCQWPPQKHQK

>Cm_EDQM1
MCSRQEKDQCHKQDHCHKQDHCHKEDHCHGSGGGSSCHGSGGSSCHGKPQKPCQQEQQQQQKHCCPVPSQKLK

>Cm_EDQM2
MCSRQEKDHCHKQDHCHSSGSSCHGSGGSSCHSGGSSCHSSGGSSCHGGGSSCHSGGSSCHGKPQQHCQQQQQQKICKVPCQK
LK

>Cm_EDQM3
MCSRQEKDHCHKQDHCHSSGSSCHGSGGSSCHSGGSSCHSSGGSSCHGGGSSCHSGGSSCHGKPQQHCQQQQQQKICKVPCQK
LK

>Cm_EDQM4_partial
MCSRQEKDHCHSSXXXXXXXXXX

>Cm_EDWM1
MIYSSGRESYFNRNSTWYDPAGSWLDTRRTPFTYAYSTCCSSGCGPRGGHDNRCYEYRRSGCAENCHGSSGSCHGSGGHCCVR
RPSYFHGSSGGCHGHGWSVCSERSCHSSGSSCHGSGSSCHGSGSSCHGTSGACHSAPIYVKPKQHVQQCCPPVQKCCPPMQQC
CLPVKKC

>Cm_EDWM2
MIYSSGRESYFNRNSTWYDPAGSWLDTRRTPFTYAYSTCCSSGCGPRGGHDNRCYEYRRSGCAENCHGSSGSCHGSGGHCCVR
RPSYFHGSSGGCHGHGWSVCSERSCHSSTVLEHHATVLDRHATVPLEHATVHQFM

>Cm_EDYM1
MSYFAYQYKQRNYTPYSATRLVPPAEPCVVKGPAPPGTKCAETCAVKHPAPCTTQCRDPCAAKPSVPCATKCFEPHAQRHPAQ
YIPKFSEPVGVKCSTPCVTRYHEPYGLIHPQPFPERWNPCAPPYVHGGYPQACGPTYVPSFPKYPYPYAPQWPDTWGYGNCGP
C

>Cm_LOR_partial
MCSHQEKQDCREIPAQAGGHHASGAGSSGSCPGALRGQHILGSSGCGGGGGSSCCSSGESCQKIIITGGGSGGSSGGSGGSYG
CGVGGGSGIGSGGSGQKIIIADGGSGGLSGCGVGGGSGSGSGGSYGWGVGGGCGGEGXXXXXXXXX

>Cm_PGLYRP3
MSLCILDFIVWFFCLHIVTPGKWGGRPANCSSPLKSVQPGYVIVLHTAGGSCKTRAECNQQMTNIQDYHMNNKGWCNIAYNFL
IGEDGKVYEGRGWHTEGAHTYGFNDISLGIAFIGDFTGRSPNAAAWKALKHLLHFAVENGYLSSNYLLMAHGDVSNTLSPGQP
IRKTLKKWPHYKH

>Cm_SCFN_partial
MPQLLDSISTIISVFYKHGKKDEDCSTISKREVKRFIQREFDNITVNPYDAHTIEAVLQLLDRDGDGAVDFNEFLLLVFRVAK
VCYWYLQPKQHLPQRTETGLSGERRQEPEAGRAEGSRDQPEPPGTEKGYYETRESETREIERSCRQPREPEPRGDQGGHYEIR
EPKPRENERSRHQLGEPEPRGDERRRYETREPEPREEERRHPQPREPEPRGNERSHREPFEQPREPEPRGESPXXXXXXXXXX
PHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPRE
EERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQP

REPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPRED
ERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPR
EPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDE
SRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEER
SRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREP
QPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERS
HRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPE
PREDERSHRQPREPQPREEERSRRQPHEPEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHEPEPREDERSR
RQPREPEPREEERSRHQPQEPESREDEGSCCHPRKPEPREDGGSHLQPPEPESREDERSHHQPREPESQEDESSLHQPREPEP
REDERSRRQPNEPERREGERCSIQPQTPEPPHEEGYLPPSSEPEAREDLGTLHEPNDPESTDDYRKFPATHEPSPTGDEGSQL
QPRESALPGDEGSPQKRRESEPREDDGSQPQLCEPEHREGDGVLRHTPESQPLEEAGSQEQPHDLEPRTEGNRHQLPQGGEAS
HVQRDAEPQQGNGSRHQPREPQEESERSSHQAREPLPQLGEEPQGEWSQSHPEPANAEEEPGETDPDEAKASLPCNPLYVYLL
EQTVEKQLYLAPPCQKRP

B

>Ps_EDAA1_partial
XXXXXXXXXXGCYPYPCRWGCGYGSGNCWPCFAAEE

>Ps_EDAA2_partial
XXXXXXXXXXGCYPYPCRWGCGYGSGNCWPCFAAEE

>Ps_EDAA3_partial
MTFHHQKISRPWGCDPCCYGGWGGCRGRYDCYRAGXXXXXXXXXXAACSSSPSRWGCGYGYGNCWPCFAEQE

>Ps_EDAA4_partial
MTFHHQKISRPWGCDPCCYGGWGGCRGRYDCYRAGXXXXXXXXXXSHPCSSSPSRWGCGYGYGNCWPCFAEQE

>Ps_EDAA5
MTFDGSIIELCYNPWSHGCWHGSRSHFSYTRPCGNWGASRGGWNHVYDCSRPCSSQWGHWYPCGKHWPC

>Ps_EDAA6
MTYQHQKISHCWGYSPCSYGSGGGYRGHYDCYRPGCYSRPYRCGWGYDNGCYYPSSSRWGHGCGYQYGRCWPC

>Ps_EDAA7
MTFDDSPISELCYCPGSYGGCCGYGGSGYCRPWCYPRPYKCCWRRQYKCCYPYPYRWGCGYGSGNCWPCFAAEE

>Ps_EDAA8
MIFDENFRDELYYWSYHYGGGSRGYGCCRPCFAQEE

>Ps_EDbeta1
MSCSRSVCAEGAAACVAPPQPCGDTWTQPCVTQCPDSRVVIFPPPVVVTFPGPILTTFPQESVVDSAGTPAVASGYGGSHGAG
ALGLGRGACGLCGPC

>Ps_EDKM_partial
MSKFIKAITDLISSYQDNSRKGRESERFQRCXXXXXXXXXX

>Ps_EDP1L1_partial
MTYYGRKHQQHCLPSPACVAKCPQPCRPQYEQHCAPKCQPVYVTKCTPLYXXXXXXXXXXCVTKCPPQCVTKCPPPCVTKCPP
QCVTKCPPPCVTKCPPQCVTKCPPPCVTKCPPRCVTKCPQQCVTQYPGQCQSGNIKMSSQCKKYCSAPNWPW

>Ps_EDP1L2
MTYYGRKHQQHCLPSPACVAKCPQPCRPQYEQHCAPKCQPVYVTKCPPLYGPQYAFPCAAQCPPRCVTKCPPPCVTKCPPQCV
TKCPPPCVTKCPPQCVTKCPPPCVTKCPPQCVTKCPPPCVTKCPPQCVTKCPPPCVTKCPPQCVTKCPPPCVTKCPPQCVTKC
PPPCVTKCPPQCVTKCPPQCVTKCPPRCVTKCPQQCVTQYPGQCQSGNIKMSSQCKKYCSAPNWPW

>Ps_EDP2
MASPQNQQQRRQSLPLPPALSNAAPEPEPSPGPRTVKEPENAPREEEKPQKEPLDQPPGPVPELEPEPEPEPEPAPEPNPPEA
EEAGYLQPEQQQYKQPPALPPAPGAETSTECEEAKPEPEPEPGRCPPPISEPEGPGPVQPSPPGEEKQQKQPCRWPPARK

>Ps_EDP4
MSSQQHQCKQTCPPPKCQEKCPPPCKEPCPPPKCPSPCKEPCPPPKCQDPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPP
PKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPS
PCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPTPPPLSPSGFLFVSLFLPPPFTLLPRDPVSHKGWGGGAGRKLCVSAEA
RGSAGAP

>Ps_EDP5
MPHHQQQCKQPCQPPPVCPPLCQETCPTPKCPEPCSPKCPEPCSPKCPEPCSPKCPEPCSPKCPEPCSPKCPEPC
SPKCPEPCPKPCPEPCPEPKCPEPKCPEPKCPEPCPPPHCQEKCPPPHCQEKCPPPHCQEKCPPPHCQEKCPPPHCQEKCPPP
HCQEKCPPPHCQEKCPPPKCPPTQK

>Ps_EDP6
MSSQQHQCKQTCPPPKCQEKCPPPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPP
PKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPS
PCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEP
CPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCPSPCKEPCPPPKCQDKWEQTLPGTMVPATLTQLPLSAA
SHTAPCLAFNYGKRRPPAECGVPHSYPLGGGLTQTDRHTQLPAETWPCVGLLSASPRPVSQASKNQGRSLQAST

>Ps_EDPCV1
MSHQQQCKQTCPPPPCCVKHCETKCVDPCPCPPKCVDPCPCPPKCPPCCPCPPCPPKCIDPCPCPPKCVDPCPCPPKCPPCPP
CPPVQQCCQEKKSC

>Ps_EDPCV2
MAHQQQCKQTCPPPQCCVTKCTTKCLDPCCNVCETKCVTKCVDPCCNVCVKKCTTCVHPCPCPPKCVDPCPCPCPPKCVDPCP
CPCPPKCVDPCPCPCPPKCVDPCPCPCPQKCVPCPPKCPPVQQCCKEKRLC

>Ps_EDPCV3
MAHQQQCKQTCPPLQCCATKCTTKCLDPWCNICGTKCVAKCTDPCCNVCVKKCTKCVHPCPCPPKCVDSCPCPCSCPQKCPPC
PLICPPVQQCCKEKKLC

>Ps_EDPCV4
MAHQQQCKQTCPPPQCCVTKCTTKCLDPCCNVCETKCVTKCVDPCCNVCVKKCTTCVHPCPCPPKCVDPCPCPCPCPCPQKCV
PCPPCPPKCPPVQQCCKEKRLC

>Ps_EDPL1
MSCQQPQQQCKQACMPPPCKEPCLPKTTEPCPQQCTEPCPQQGTEPCPPKCAEPRPGPRRDPRLSSSARTAAPRINCSRRARR
RQIPTREPSPLAPGCFTVALLVFQLRTGLAAGTGRWEGFGTDTGSQAAQF

>Ps_EDQL
MCSREPRGCPDSERSSCPSSERSSCHGSEATTCHDVKPHPQYPTTVPCQTPTSPCQQQTKQTCPWPPQKHQK

>Ps_EDQM1_partial
XXXXXXXXXXGSGGGSCHGKPQKPCVPDQQQHKDCCQVPSQKLK

>Ps_EDQM2
MCSRQGGDHCHSSDSCLSSGSSCHGSGGSSCHSSGGSSCHGGGSSCHSGGSSCQGGGSSCHSGGSSCHGGGSSCHGGGSCHGK
PQQPCQQQQPQQQQKVCNVPCKKLK

>Ps_EDYM1
MSYFAYQYKQRNYTPYSTTRLLACAEPCVVKGPAPCGTKCVEPCATKRPAPCVPKCRDPCAGKAPVHCEPKCLEPHAQRGPAH
CAPKFSEPAGVKCSVPWVPRCHEPYGPIPARPFPERWNPCAPPYGQPFVTGYPQACGPSYGPSFPKYSYPCAPQWPGGWGYGG
CGPC

>Ps_LOR_partial
MCSHQQKQDCHEIPAGPGGDCSCTGGGSVGGGSGAVLHQSTLGNSSGGSACCGVGGGSGGVGGIAAGGCLGGGSGGVGGGPGQ
KGVVCGGSGGXXXXXXXXXX

>Ps_PGLYRP3
MSQLAVLLWAVSAVSWGLSCLHIITPGKWGGRPANCSTPLKDVQTGYVVTLHTAGRSCSTQAECNQELLNIQQHHMNSKGWCN
LAYNFLIGEDGNVYEGRGWNTEGAHTYGSPNAAAWVALKRLLHFAVEAGYLSSNYLIMAHGDVSNTISPGQPIPNVLKTWPHY
KH

>Ps_SCFN
MPQLLDSISTIISIFCTQGEKDEGCSTVSRREMKRFLQREFADVTLTPYDPPTIEAVLQLLDHDEDGAVDFNDFLLLVFRVAK
VCYWYLQPKQRLPPRAEKEMSGKQGPAPKAGRAKGSCEQPCEKGGYEMPEPEARETEKSHWQPHETEPRRDQGGRHKIHEPKP
RENSRSHHQRGEPELQGDESWSYEPCEPEPREAETRRHQPRESEPRGDERSCCEPFEYEPKPSEHPHEPEPRGAESWCYEPHE
PEPMEEERYRHQPREPKPQGDERRCHEIRVSEERENDRRHHQPRESEPQGAERSWAGACEPEAREEERHHPQPRQPEIQKDEW
RHYGIRDHERREVDRRQQHLHDPEWRGDDRSSHQPREPEPQGYERRRSEQEPIEGGKSHYWPREQKPRGGAKTHYQPSRPEQR
RDERHGYERCDREARREESRHQPREPELRGAERHHYETDAAELRENDGTSHQSHETQPREDETRLYDRHILQRRNRDEHCPQR
RESGTREDKMSHSERHAPKSQGAKRRHHQPNSPEHQGDDRYSHERCETEYRRNDKRHNQPGQPKLRRDERSHYQSWEPELRVD
KWSCRQPREPEPREDDKSHRQPCDPEPREGERSCHQPRPREDKRSLHLPHEPESREDGSSPRQPWQPEPREDDKSHHQPCDPE
PREGERSCHQPRPREDKRSLHLPREPESREDGSSPRQPWQPEPREDDKSHHQPCDPEPREGERSCHQPRLREDKRSLHLPREP
ESREDGSSPRQPWQPEPREDDKSHRQPCDPEPREGERSCRQTHPRKDKRSLHLPREPESREEERRHQPSEFEPREAEQSLHQP
HELEPKGEERSHRQPHEPEPREEEQSLHQPREPEPRGEERSHRQPHEPETTEKERSLHKPHEPELREERSYRQPLDPEPREEE
QSLHRPHEPELRGEERRHRPRDFEHREDEQSLHQPREPEPRGEEKSLHQPDPEPRKDEQSHHQPLDPEPRKDEQSHHQPLDPE
PRKDEQSHHQPLDPEPRKDEQSHHQPLDPEPRKDEQSFHQPHELEPKEGERSHQQPNKPEPREEESSH

RQPREPEPREEQRRCHQPPEPEAGVEEVSQQQPGSHEPEQEKGHLPLWREPEAREDVGTQRDADAPGDYTRCPAMPKPSPTGE
VGIPLQPWDGSLAGGEGSRQAATDSERREDSQPQLREPEHREEDGILRETPESEPRAATGSQHQPCDLEGQREGSRPQRPPGG
DRSCLPLDTEPPLANRSLPCPKEPQEHASKRSSPQASELQPWLGEAPQEGEWSQSHPESASSQEEPDETGAEDAKACLPCSPL
YVYLLEQTVEKQLYLAPPHQEWP

C

>Ps_Beta-1
MSCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVGAVGAPVVGSGYGGSFGAGALSGYGA
PYGGLYGLGGFGGYGGHFGGLCGYGGRYGYGGLNGYGRGFGGLCGYGGGYGGLCGYGGGYGGLCGYGGGYGGLCGYRGGYGYG
GACGTGVSCHRYLSACGSPC

>Ps_Beta-2
MSCYSECGVARPSPVSGTFNEPCVRQCPDSQVLIMPSPIAVTIPGPIMSTFPQHSEVGAVGAPVVGSGYGGSFGAGGLFGSGA
GYGGLFGLGGSGGYGGLCGYGGRYGYGGGYGTGVSCHRYLSACGSPC

>Ps_Beta-3
MACSSLSYPECGVARPSPVSGTFNEPCVRQCPDSHVLIRPSPIAMTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLF
GSGAGYGGLSVLGGSGGYGGLCGYGGRYGYGGGYGSGVSCHRYLSACGSPC

>Ps_Beta-4
MSCYPECGVARPSPVSGTFNEPCVRQCPDSQVLIRPSPIAMTIPGPIMSTFPQHSEVGAVGAPFVGSGYGGSFGAGGLFGSGA
GYGGLFGLGGSGGYGGRYGYGGGYGYGGSCGTGVSCHRYLSACGSPC

>Ps_Beta-5
MSCYPECGVARPSPVSGTFNEPCVRQCPDSQVLIMPSPVAMTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLFGSGA
GYGGLSGLGGSGGYGGLCGYGGRYGYGGGYGIGLSCHRYLSACGSPC

>Ps_Beta-6
MSCYSECGVAQPRPVSGTFNEPCVRQCPDSHVLIMPSPIAVTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLFGSGA
GYGGLLGLGGSGGYGGLCGYGGRYGYGGGYGSGVSCHRYLSACGSPC

>Ps_Beta-7
MIKSLCAPRCYPCPDICPDPCAYVCNEPCVTSCGDSTAVVYAPPVAIRFPGPILATCPQDSVMGTSLPQIPYGPYGGGAVFAL
GVL

>Ps_Beta-8
MSSHKDLCCPRPSCCPDVCPQPYVDAWNQPCVTSCGDSSAVVYPPPVIVRFPGPILATCPQESFVGTSLPNVPYGYGGSYAGG
NFGGSVSAGSVSGGTYGGLYSYGKRYDRKCYPSRFGGCGPC

>Ps_Beta-9
MSSCKDLCLPRPACCPNVCPQPYVDAWNQPCVTSCGDSSAVVYPPPVIVRFPGPILATCPQESFVGTALPNVPYGYGGSYGGG
SISGSVGSGHVSGGAYGGGYAGFYGYGSKYDRKCSSSLFGRCGPC

>Ps_Beta-10
MASRKDLCCPRPPCCPDICPQPYVDAWNGPCVTSCGDSSAVVYPPPVVVRFPGPILATCPQDSVVGTALPNVPYGYGGSYGGS
NFGGSLSSGSVSGGAYGGLCSYGKRYERKCYSSRFGGCGSC

>Ps_Beta-11
MSSRKELCCPRPQCYPDVCPQPYVDAWNGPCVTSCGDSSAVVYPPPVVVRFPGPILATCPQESVVGTALPNVPYGSGGAYAGG
KFGGSVSSGGVYGRGYTGGYGAGYGGLFGDGSKYGRNCYSSRFGGCGPC

>Ps_Beta-12
MASRKDLCCPLPPCCPDICPQPYVDAWNGPCVTSCGDSSAVVYPPPVVVRFPGPILSTCPQDSVVGTALPNVPYGYGGSYRGS
NFGGSLGSGSVSGGAYGGLCSYGKRYERKCYSSRFGGCGPC

>Ps_Beta-13
MSSRKELCCPRPQCYPDVCPQPYVDAWNGPCVTSCGDSSAVVYPPPVVVRFPGPILATCPQESVVGTALPNVPYGSGGAYAGG
KFGGSVSSGGVYGRGYTGGYGAGYGGSFGDGSKYGRNCYSSRFGGCGPC

>Ps_Beta-14
MSSRKELCCPRPQCYPDICPQPYVDAWNGPCVTSCGDSSVVVYPPPVVVRFPGPILATCPQESVVGTALPNVPYGSGGAYAGG
KFGGLAGSGGAYGSGYTGGYGAGYGGSFGDGSKYGRNCYSSRFGGCGPC

>Ps_Beta-15
MSSCRELCYQPSPCWPDICPDPCAVARNEPCITSCGDSTAVVYPPPVSVLFPGPILSTLPQHSVVGSTLPAIPYGARSSSGGG
ILGGSLGYGGGYDGAYGGGYGGGIGAFGYGGLCGYGKRYGRRCYPYRCGPCWPC

>Ps_Beta-16
MSFCRDLCPSPSYPACQVTCPQPFVDACNGPCVTSCGDSTAVVYPPPVIVNFPGPILATCPQESIVGSSEPLGIGSAIGYGGS
NLSVSSYGYRPSLGYGGSSGSQSLNSLRRSYTSGVSSVGRGGSDPCSSRWLMMYGCGPRPTQQH

>Ps_Beta-17
MSTARLWLNNLNSEDSEQDQDSKTKDPPKDCPKGCKCPKCEPCKPCPRPCPRPCPRPCPPCWPCPRPCPPRCPPICPQPCVNV
CNEPCVTSCGDSTAVVYAPPVQVIFPGPILATCPQDSVVGSSLPAGMFGPRPSGGFLGAGSSVGSGGYLGAGGYLGAVGSFGP
GGCVGGGGSSFGTVCCNPCRYGSCWPWSSQEGSPRKTTLPGNDKRAADSRLS

>Ps_Beta-18
MSTARLWLNNLNSEDSEQDQDSKTKDPPKDCPKGCKCPKCCKCPPIKPCKPYPCFPRPCPRPCPPCWPCPRPCPPRCPPICPQ
PCVNVCNEPCVTSCGDSTAVVYAPPVQVIFPGPILATCPQDSVVGSSLPAGMFGPRPSGGFLGAGSSVGSGGYLGAGGYLGAV
GSFGPGGCVGGGGSSFGTVCCNPCRYGSCWPC

>Ps_Beta-19
MSFQGPATCTQSSQPCEVKCGEPCVLTHNEPCVVQCPDTRVIIYPPPVVVTFPGPILTTCPQESIVTAAGPPDISLAESSGMI
SSATGGSGSVGPQIHGCAPKATAGLESRYTSKYSSTSRLSGQGLGESYCYGNRQL

>Ps_Beta-20
MSFYGPCNVPCEGPHPMPSAVTYHEPCVIQCPDSMHETDSPPGVAIIPGPTLTTLSHYSLVGSSALLDMERPFGPHMCLGSEG
CMNSFD

>Ps_Beta-21
MSFYGPCNVPCEGPHPMPSAVTYHEPCVIQCPDSMHETDSPPGVAIIPGPTLTTLSHYSLVGSSALLDMERPFGPHMCLGSEG
CMNSFD

>Ps_Beta-22
MYSSMAFIGKDQKPQEPQQDEKTCPPPHKRCPPPEKTCPPPQKTCPPPCPPCPPCPKCPPCPPCRPCPPLCPPRCPLICPEPC
INVCNEPCVTGCGDSTALIFAPPVRVIFPGPILATCPQDSVVGSSLPFGVLGPRRAGGSLGSGGAFGAGSSFNAGGYLGSGRY
LGSGGSSFGTVCCNPCSCVRPRPC

>Ps_Beta-23
MYSTMALIGKDQKPREPQQDQKIPPPHKGCPPCPECPPQCPPCPKCPPCPKCPPCPKCPPCPKCPPCPKCPPCPKCPPCPKCP
PCPKCPPCPKCPPCPKCPPCPKCPPCPPCRPCPRLCPPRCPPICPEPCINVCNEPCVTSCGDSTAVVFAPPVRVIFPGPILAT
CPQDSVVGSSLPFGILGPRRAGGSLGSGGAFGAGSSFNAGGYLSSGGYLGSGGSSFGTVSCNPCSCVRPWPC

>Ps_Beta-24
MYSTMALIGKDQKPREPQQDQKIPPPQKCCPPCPECPPPCKPCPPCPPCPPCPPCPKCPPCPPCRPFCPPRCPAICPEPCINV
CNEPCVTSCGDSTAVVFAPPVRVIFPGPILATCPQDSVVGSSLPFGILGPRRAGGSLGSGGAFGAGSSFNAGGYLSSGGYLGS
GGSSFGTVSCNPCSCVRSWPC

>Ps_Beta-25
MSTAKLWLNNLNSEDSHQDQDTKTKHQKKCSPDCLPCPQDPQCPECPKCPDPSKCPDCLPCPQDPQCPECRKCPDPSKCPDCL
PCPQDAQCPECPKCPVPPKCPDCLPCPQDPQCPECPKCPDPPKCPDCLPCPQDPQCPECPKCPDPPKCPDCLPCPQNHQCPEC
PKCPDPCKCPECRPCPQDPQCPECPKCPDPPKCPDCLPCPQNHQCPECPKCPDPCKCPECRPCPQDPQCPECPKCPYPPKCPN
CLPCPQDPQCPECPQCPDPSKCPECPKCPNPPKCPDCPKCPEPTKCPDPSKCPDCPKCPPCLPCPLCRPCPRLCPPRCPPICP
PPCVNVCNEPCVTHCGDSTAVVYAPPVQVIFPGPILSTCPQDSVVGSSLPAGMFGPRPSGGLLGAGGLLGLGGSFGSGGCFGG
GGSSFGTVRCNPCRCGRCWPC

>Ps_Beta-26
MSRPICPECPKCPDPSKCPDCLSCPQDPQCPKCPKCPNPSKCPDCFPCPQDPQCPECPKCPYPPKCPDCLPCPQDPQCPECPK
CPEPPKCPDCPKCPEPTKCPDPSKCPDCPKCPPCLPCPPCRPCPRPCPPRCPPICPPPCVNVCNDTCVTHCGDSTAVVYAPPV
QVCGDSTAVVYAPPVQVIFPGPILSTCPQDSVVGSSLPAGMFGPRPSGGLLGAGGLLGLGGSFGSGGCFGGGGSSFGTVRCNP
CRCGRCWPC

>Ps_Beta-27
MSTAKLGLNDLKSQDSQQDQDTKTKDPTKGCPECPQRTECPKGCNCPKCCECPKCEPCEPCPPCPRPCPRPCPPCWPCPRPCP
PRCPPICPSPCVNVCNEPCVTRCGDSTAVVYAPPVQVIFPGPILSTCPQDSMVGSSLPAGMFGPRPSGGYLGAGGSVGPRGSF
GPGGCVGGGGSSFGTVCCNPCHYGRRWPC

>Ps_Beta-28
MACSSLSYPECGVARPSPVSGTFNEPCVRQCPDSHVLIMPSPVAMTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLF
GSGAGYGGLSGLGGSGGYGELCGYGGRYSYGGGYGTGLSCHRYLSACGSPC

>Ps_Beta-29
MACSSLSYPECGVARPRPVSGTFNEPCVRQCPDSHVLIMPSPIAVTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLF
GSGAGYGGFSGLGGAGGYGGLCGYGGRSGYGGGYGTGVSCHRYLSARGSPC

>Ps_Beta-30
MSCYSECGVARPRPVSGTFNEPCVRQCPDSQVLIMPSPVAVTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLFGSGA

GYGGLFGLGGSGGYGGLCGYGGRSGYGGGYGTGVSCHRYLSACGSPC

>Ps_Beta-31
MACSSLSYPECGVARPRPVSGTFNEPCVRQCPDSHVLIMPSPIAVTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLF
GSGAGYGGLSGLGGSGGYGGLCGYGGRYGSGRGYGTGVSCHIKAKGKTKD

>Ps_Beta-32
MSCYTECGVARPSPVSGTFNEPCVRQCPDSQVLIMPSPVAMTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLFGSGA
GYGGLFGLGGSGGYRELCGYGGRYGYGGGYGTGVSCHRYLSACGSPC

>Ps_Beta-33
MSCYTECGVARPSPVSGTFNEPCVRQCPDSQVLIMPSPVAMTIPGPIMSTFPQHSEVGAVGAPVVGSGYGGSFGAGGLFGSGA
GYGGFSGLGASGGYGGLCGSPVVGSGYGGSFGAGGLFGSGAGYGGLFGLGGSGGYGGLCGYGDRYGYGGGYGSGVSCHRYLSA
CGSPC

>Ps_Beta-34
MSCYSECGVARPRPVSGTFNEPCVRQCPDSQVLIMPSPIAMTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLFGSGA
GYGGLSGLGGSGGYGELCGYGGRYSYGGGYGTGLSCHRYLSACGSPC

>Ps_Beta-35
MACSSLSYPECGVARPRPVSGTFNEPCVRQCPDSHVLIMPSPIAVTIPGPIMSTFPQHSEVGAVGAPVVGSGYGGSFGAGGLF
GSGAGYGGLFGLGGSGGYGGLCGYGGRYGYGGGYGSGVSCHRYLSACGSPC

>Ps_Beta-36
MACSSLSYPECGVARPRPVSGTFNEPCVRQCPDSQVLIMPSPVAMTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLF
GSGAGYGGFSGLGGAGGYGGLCGYGGRYGYGGGYGTGVSCHRYLSACGSPC

>Ps_Beta-37
MSCYTECGVARPRPVSGTFNEPCVRQCPDSQVLIMPSPVAMTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLFGSGA
GYGGLSGLGGSGGYGGLCGYGGRYGSGRGYGTGVSCHRYLSACGSPC

>Ps_Beta-38
MSCYPECGVARPSPVSSTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVAALGAPVVGSGYGGSFGAGALSGYGA
PYGGYYGLGGLGGYGGRFGYGGLCGYGGGYGYGRGFGGLCGYGGGYGGLCGYRGGYGYGGSCGTGVSCHRYLSGSCTPC

>Ps_Beta-39
MSCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVGAVGIGAPVVGSGYGGSFGAGALSGY
GAPYGGYYGLGGFGGYGGRFGYGGLCGYGGGYGYGRGFGGLCGYGGGYGGLCGYGGGYGGLCGYRGGYGYGGSCGTGVSCHRY
LSACGSPC

>Ps_Beta-40
MSCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVAALGAPVVGSGYGGSFGAGALSGYGA
PYGGYYGLGGFGGYGGRFGYGGLCGYGGGYGYGRGFGGLCGYGGGYGGVWGEVTVMEDLVAPGYLVIGT

>Ps_Beta-41
MSCYTECGVARPRPVSGTFNEPCVRQCPDSQVLIMPSPVAMTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLFGSGA
GYGGLFGLGGSGGYGGLCGYGGRYGYGGGYGTGVSCHRYLSACGSPC

>Ps_Beta-42
MSCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVAALGAPVVGSGYGGSFGAGGLFGSGA
GYGGFSGLGASGGYGGLCGYGGRYGYGGGYGTGVSCHRYLSACGSPC

>Ps_Beta-43
MSCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVAALGAPVVGSGYGGSFGAGALSGYGA
GGYGGLCGYGGGYGGLCGYGGGYGGLCGYGGGYGGLCGYGGGYGGLCGYGGGYGGLCGYGGGYGGLCGYGGGY
GGLCGYRGGYGYGGACGTGVSCHRYLSACGSPC

>Ps_Beta-44
MACSSLSYPECGVARPSPVSGTFNEPCVRQCPDSHVLIMPSPVAMTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLF
GSGAGYGGLSGLGGSGGYGELCGYGGRYSYGGGYGTGLSCHRYLSACGSPC

>Ps_Beta-45
MSCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVGALGAPVVGSGYGGSFGAGALSGYGA
PYGGMYGLGGLGGYGGHFGYGGLCGYGGGYGYGRGFGGLCGYGGGYGGLCGYGGGYGGLCGYRGGYGYGGSCGTGVSCHRYLS
GSCTPC

>Ps_Beta-46
MSCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVAALGAPVVGSGYGGSFGAGALSGYGA
PYGGYYGLGGFGGYGGRFGYGGLCGYGGGYGYGRGFGGLCGYGGGYGGLCGYGGGYGGLCGYRGGYGYGGSCGTGVSCHRYLS
GSCTPC

>Ps_Beta-47
MACSSLYYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVAALGAPVVGSGYGGSFGAGALS
GYGAPYGGYYGLGGFGGYGGRFGYGGLCGYGGGYGYGRGFGGLCGYGGGYGGLCGYGGGYGGLCGYGGGYGGLCGYRGGYGYG
GACGTGVSCHRYLSGSCTPC

>Ps_Beta-48
MSCSSLYYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVAALGAPVVGSGYGYGAPYGGMY
GLGGLGGYGGHFGYGGLCGYGGGYGYGRGFGGLCGYGGGYGGLCGYGGGYGGLCGYRGGYGYGGSCGTGVSCHRYLSGSCTWP
KEP

>Ps_Beta-49
MSCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVGAVGAPVVGSGYGGSFGAGALSGYGA
PYGGYYGLGGFGGYGGRFGYGGLCGYGGGYGYGRGFGGLCGYGGGYGGLCGYGGGYGGLCGYRGGYGYGGACG
TGVSCHRYLSACGSPC

>Ps_Beta-50
MSCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQQSEVGAVGAPVVGSGYGGSFGAGALSGYGA
PYGGLYGLGGFGGYGGHFGGLCGYGGRYGYGGLNGYGRGFGGLCGYGGGYGGLCGYGGGYGGLCGYRGGYGYGGACGTGVSCH
RYLSACGSPC

>Ps_Beta-51
MACVPQDCGSDICPRPYIDVCNSPCVSSCGDSTAVVFAPPVVVRFPGPTLATCPQDSIVGSALPQLPYGPGGFPGVGGGVGGP
FAGGYGGVSGGRFGGNYGGYSGGSGGGYAAGGCGGGYSGGYGGSCGSRRSYRSISSCGGGYSSKGGCGPC

>Ps_Beta-52
MACIPQNCCLDICPCPYIDVCNSPCSSSCSDSTAVVYAPPAVVRFPGPTLATCPQDSFVGSVLPQLPSGPGGFPGDGGGVGGS
LGSGG

>Ps_Beta-53
MACVPQDCGSDICPRPYIDVCNSPCISSCGDSTAVVFAPPVVVRFPGPTLATCPQDSFVGTSLPQLPSGPGGFPGVGGGVGGA
YGGRFGGNYGGYSGGYRGGYRGGYGGGCGGGYCGGNGGSCGFRKSYSSISSCGGGYSSKGGCGPC

>Ps_Beta-54
MACVPQDCGSDICPRPYIDVCNSPCVTSCGDSTAVVYAPPVVVRFPGPTLATCPQDSFVGTSLPQLPSGPGGFPGVGGGAGGA
YGGRFGGNYGGYSGGYRGGYRGGYGGGCGGGYCGGNGGSCGFRKSYSSISSCGGGYSSKGGCGPC

>Ps_Beta-55
MSRAPPVCVKFPGPTLATCPQESFVGTSIPYLAGGPGGFPGSGVGSRGPFTAKSSSSGVYAGGYNGGSGGGYVYGSGFQGGYG
GSCGYGRKLYGSRSVCGGQYSGYQGRNCGPC

>Ps_Beta-56
MKFPCAPRCYPCPDICPEPCAYVCNEPCVTSCGDSTAVVYAPPVAVRFPGPILATCPQDSVVGTTLPLPPYGPYRGGAGGGAG
SFLGGGGSGVFGGGSGGGAGGLGGLGGGYCYGSSSGGYGRHCSYTCVPCPRYRPC

>Ps_Beta-57
MTFFPCAPRCHPCPDLCPEPCAVVCNDPCVTSCGDSTAVVFAPPVAIRFPGPTLATCAQDSVVGSSAPWLPYGPYGPSRGGAGG
GAGSGAGGGVYGGVFGGGVYGGALGGGSGSGFGGCSGGGYGGSFRQWSRYDRKCYANRYDCCPPC

>Ps_Beta-58
MKFPCAPRCHPCPDICPEPCAYVCNEPCVTSCGDSTAVVYAPPVAIRFPGPILATCPQDSVVGSSLPQIPYGPYGGGAGGGGA
LGGAGGLGGGALVGAGSGFGGGSGGGFGSGFGGCHGGYWGRRCYANRYDCCPW

>Ps_Beta-59
MKSLCAPRCYPCPDICPDPCAYVCNEPCVTSCGDSTAVVYAPPVAIRFPGPILATCPQDSVVGTSLPQIPYGPYGGGAVFALG
VL

>Ps_Beta-60
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVAVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGYPGFYGYGGYCGYPGFFGPC

>Ps_Beta-61
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVAVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGYPGFYGYGGYCGYPGFYGNGGYCGYPGYYGNGGYCGYPGYYGNGGY
CGYPGYYGNGGYCGYPGYYGNGGYCGYPGYYGNGGYCGYPGYYGNGGYCGYPGYYGNGGYCGYPGYYGNGG
YCGYPGYYGNGGYCGYPGYYGNGGYCGYPGYYGNGGYCGYPGYYGNGGYCGYPGYYGYGGVVGSGVSCHRYRSGSCAPC

>Ps_Beta-62
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGYWGRPLRLWGLMGLPRLLWEWGILRLPRLLW

>Ps_Beta-63_partial
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVAVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGYPGFYGYGGYCGYPGFYGYGGYCGYPGFYGNGGYCGYPGXXXXXXXXXX

>Ps_Beta-64
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVAVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGYPGFYGYGGYCGYPGFFGGGDCGYPGYYGNWGYPGYYGGLCGGYGGYGRRYLGGYCGP
C

>Ps_Beta-65
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGFGGYPGFYGYGGYCGYPGFYGNGGFWGYPGRYAHGGYCGYPGYYGGLCGGYGGYGRRYLGG
YCGPC

>Ps_Beta-66
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVAVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGYPGFYGYGGYCGYPGFFGGGGFWGYPGRYAHGGYCGYPGYYGNGGYCGYPGFYGYGGV
VGSGVSCHRYRSGSWAPC

>Ps_Beta-67
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGLWGYPGFYGNGGYWGYPGRYAHGGYCGYPGFYGNWGYCGYPGYYGGLCGGYGGYGRRY
LGGYCGPC

>Ps_Beta-68
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGLWGYPGFYGNGGYWGYPGRYAHGGYCGYPGYYGNWGYPGYYGGLCGGYGGYGRRYLGG
YCGPC

>Ps_Beta-69
MTFSSMCYPECGVARPSPISVSCNEPCVRQCPDSHVTIIPTPVDVTLPGPILETHPQQSTVGAVEDPALGVGYGSSLGLGGLG
DNGGFYGLRKFGGFGGYLGGCGYGGYCGYPGYYGYGGYCGYPGYYGYGRYCGYPGFYGSRRLYGSGVSCHSYLGGYCGPC

>Ps_Beta-70
MTFSSMCYPECGVARPSPISVSCNEPCVRQCPDSHVTIIPTPVDVTLPGPILETHPQQSTVGAVEDPALGVGYGSSLGLGGLG
DNGGFYGLRKFGGFRGYLGGYRYGRYCGYPGYYGYGGYCGYPGYYGYGRYCGYPGFYGSRRLYGSGVSCHSYLGGYCGPC

>Ps_Beta-71
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGLWGYPGFYGNGGYWGYPGRYAHGGYCGYPGFYGNGGFWGYPGRYAHGGYCGYPGYYGN
WGYPGYYGGLCGGYGGYGRRYLGGYCGPC

>Ps_Beta-72_partial
MALSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGLWGYPGFYGNGGYWGYPGFFGNRGDWGYPRXXXXXXXXXX

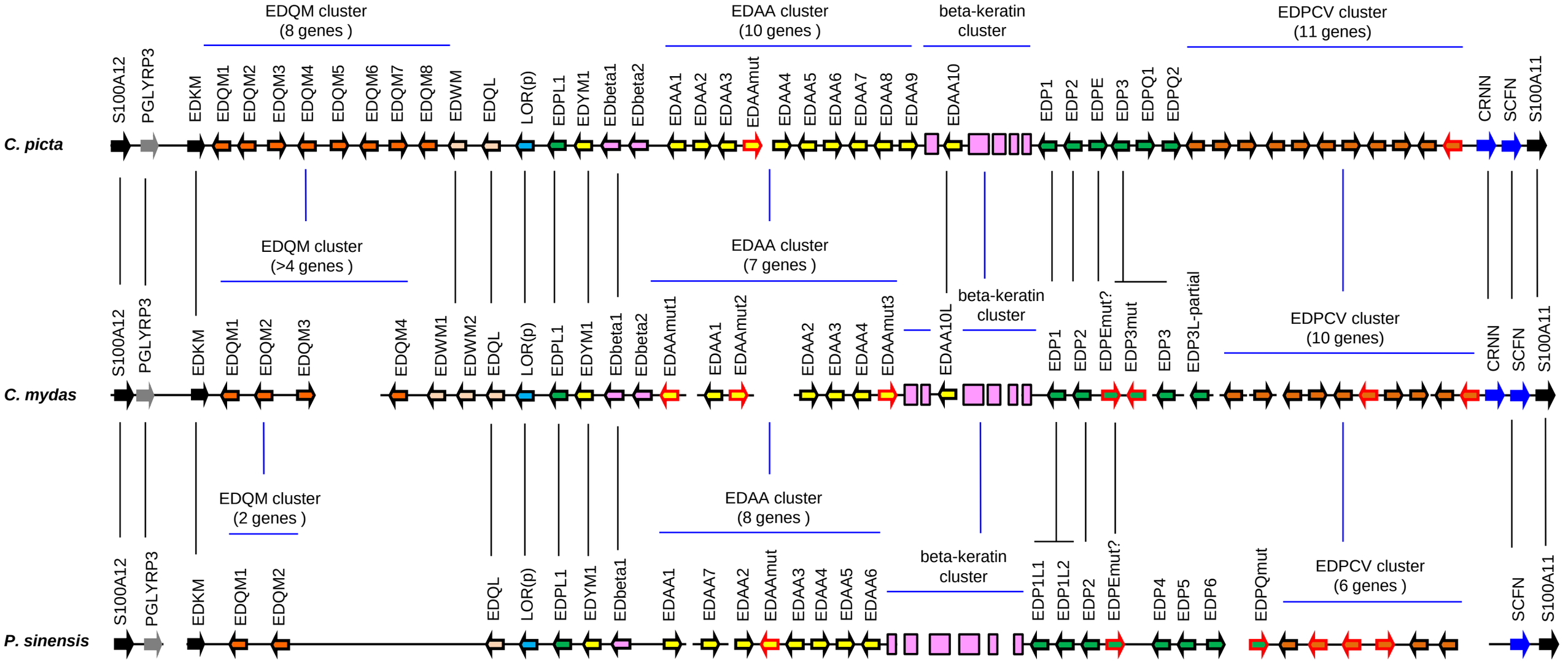
>Ps_Beta-73
MAFSSLCYPECGVARPSPVSGTCNEPCVRQCPDSEVVIRPSPVVVTLPGPVLSTFPQQSEVAAVGEPVVGAGYGGSFANGGLY
GYGGRYGGLFGYGGYGYGGRYGYGGYPGFYGYGGYCGYPGYYGNGGFWGYPGRYAHGGYCGYPGYYGNGGYCGYPGYYGNGGY
CGYPGYYGYGGVVGSGVSCHRYRSGNCAPC

>Ps_Beta-74
MSCYPECGVARPSPVSGTFNEPCVRQCPDSEVVIRPSPIAVTIPGPIMSTFPQHSEVGAVGAPLVGSGYGGSFGAGGLFGSGA
GYGGLFGLGGLGGYGGHFGYGGLCGYGGRYGYGGGYGYGGSCGTGVSCHRYLSACGSPC

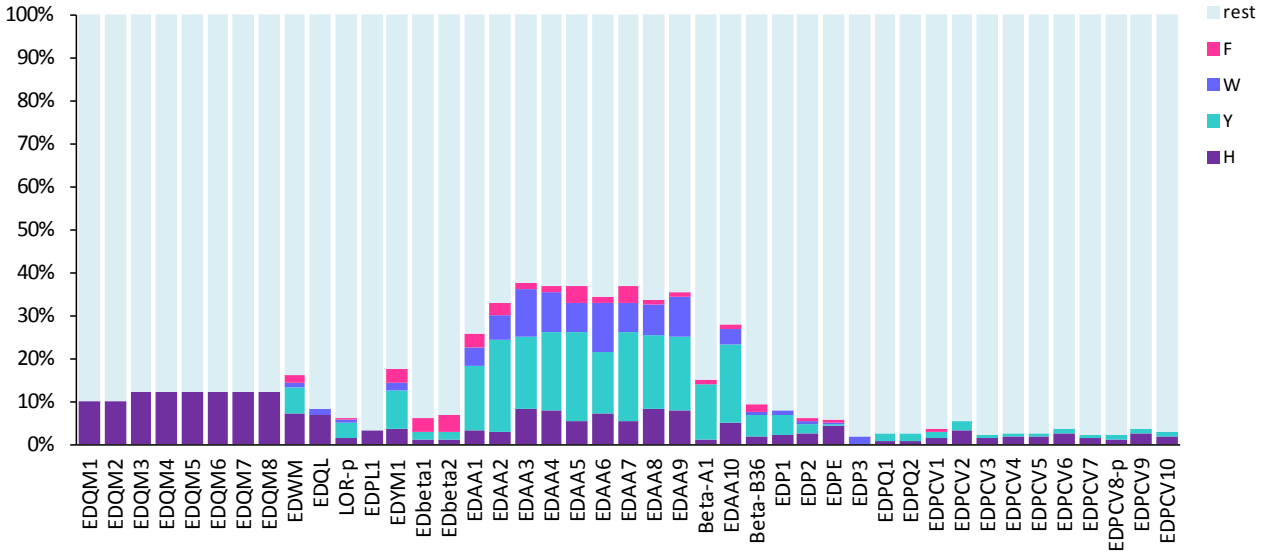
>Ps_Beta-75_partial
XXXXXXXXXXIRFPGPTLATFAQDSVVGSSAPWLPYGPYGPSRGGAGGGAGSGAGGGVYGGVFGGGVYGGALGGGSGSGFGGC
SGGGYGGSFRQWSRYDRKCYANRYDCCPPC

>Ps_Beta-76
MSFAPIIPLTLVLYVESTPLLSSSGLPPSLKTSFNGEPCTDRCHNPYEVTCPQPIGNSSNQPCVVACGDSRVVVCPPLVVVTC
PGPILSTCPQDSIVGSSIPSGSRIFGSTSMISWHQESLGTQSHMQGESS

Supplementary Figure S2. Amino acid sequences of EDC genes of *Chelonia mydas* (Cm) and *Pelodiscus sinensis* (Ps). (A) Amino acid sequences of EDC proteins of Cm (beta-keratins are not included). (B) Amino acid sequences of EDC proteins of Ps (beta-keratins are not included). (C) Amino acid sequences of beta-keratins of Ps. Amino acid residues C, G, K, P, Q, and S are highlighted by specific colors (see main text and legend of Suppl. Fig. S1).



Supplementary Figure S3. Comparison of the EDCs in 3 species of turtles. The arrangement of genes on the EDC of *C. picta*, *C. mydas*, and *P. sinensis* is schematically depicted. Simple EDC (SEDC) genes with 2 exons are represented by colored arrows with a black frame whereas other genes are shown as filled arrows without frame; red frames indicate SEDC genes that are inactivated by mutations. Clusters of more than 2 beta-keratin genes are shown as boxes. Colors indicate groups of genes as defined in the text. Black vertical lines connect orthologs. Note that the schemes are not drawn to scale. The complete sequences of SEDC genes containing multiple internal repeats, such as LOR and EDPE, could not be faithfully predicted because of uncertainties in the genomic sequence assembly (p, partial sequence available; mut, mutation inactivating the gene). Note that the EDC of the soft-shelled turtle (*P. sinensis*) lacks the genes EDWM and CRNN and contains fewer functional genes of the families EDQM and EDPCV.



Supplementary Figure S4. Aromatic amino acid contents of turtle SEDC proteins. The contents of aromatic amino acid residues (% of total residues) in *Chrysemys picta* SEDC proteins are depicted for comparison with Figure 3A. Note that only the beta-keratins encoded by the first (Beta-A1) and last (Beta-B36) gene of the beta-keratin cluster are included here.

A

```
>Cp_CRNN
MTQLLSNIKGIINAFYVFAKKDGACPTLSKGELRQLIHQEFADVTVVPQGLQTIDKLLQLLDTDSDGRLDFNGFLVLIFQVAK
ACYGEVSQGQRPGHGGSSASQGEANCGERTKEPP
TPERDPSPRQAPEPQ
TPERDSIPCQAPEPQ
IPEQDSSPCQAPELQ
IPEQDSNPCQAPEAQ
IPERDPSPCQAPELQ
TPEQDSSPRQAPEPQ
IPEQDSSPRQALEPQ
IPEQDSIPCQAPEAQ
IPERDPSPHQAPEPQ
TPERDPSPHQGLEPQ
TPERDPSPCHAPEPQ
IPEQDPSLHQAPEPQ
TPEQDSSSRHTPEPQ
IPEQEPSPCRGPEPQ
TPEQDLSPSQAPEPQ
TPEQDLSHSETQLPP
TQQRNQGAQDPTEPAAGQASKSSQCLYSWHSQKPRPFPHWWPPKK
```

B

```
>Cm_SCFN_partial
MPQLLDSISTIISVFYKHGKKDEDCSTISKREVKRFIQREFDNITVNPYDAHTIEAVLQLLDRDGDGAVDFNEFLLLVFRVAK
VCYWYLQPKQHLPQRTETGLSGERRQEPEAGRAEGSRDQ
PEPPGTEKGYYETRESETREIERSCRQPREPEPRGDQGGHYEIRE
PKPRENERSRHQLGEPEPRGDERRRYETREPEPREEERRHPQPRE
PEPRGNERSHREPFE-----------QPREPEPRGESPXXXXPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDESRRRQPREPEPREDERSHRQPREPQPREEERSRRQPHE
PEPREDERSRRQPREPEPREEERSRHQPQEPESREDEGSCCHPRK
PEPREDGGSHLQPPEPESREDERSHHQPREPESQEDESSLHQPRE
PEPREDERSRRQPNEPERREGERCSIQPQTPEPPHEEGYLPPSSE
PEAREDLGTLHEPNDPESTDDYRKFPATHEPSPTGDEGSQLQPRE
SALPGDEGSPQKRRESEPREDDGSQPQLCEPEHREGDGVLRHTPE
SQPLEEAGSQEQPHDLEPRTEGNRHQLPQGGEASHVQRDAEPQQGNGSRHQPREPQEESERSSHQAREPLPQLGEEPQGEWSQ
SHPEPANAEEEPGETDPDEAKASLPCNPLYVYLLEQTVEKQLYLAPPCQKRP
```

Supplementary Figure S5. S100 fused-type proteins (SFTPs) of turtles contain sequence repeats. The amino acid sequences of 2 exemplary SFTPs (CRNN, cornulin (A); SCFN, scaffoldin (B)) of turtles (Cm, *C. mydas*; Cp, *C. picta bellii*) are shown. Sequence repeat elements are aligned. For comparison with other EDC proteins, amino acid residues C, G, K, P, Q, and S are highlighted with the same colors as in the amino acid sequences shown in Supplementary Figures S1 and S2. The sequence repeats of CRNN are not perfect. Sequence repeat elements of SCFN are composed of 3 internal imperfect repeats with the consensus sequence P(E/Q)PRE(D/E)E(R/S)(S/R)(R/H)RQP(R/H)E. Sequence repeats differ among turtle species (not shown). X, unknown residues because of incomplete gene sequence.

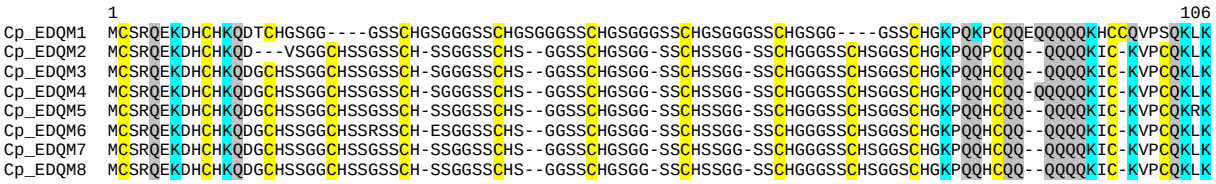
A

```
Hs Lor      MSY-----QKKQPTPQPP
Hs PRR9     MSFS--EQQCKQPCVPPP
Hs SPRR1A   MNS---QQQ-KQPCTPPP
Hs SPRR1B   MSS---QQQ-KQPCTPPP
Hs SPRR2A   MSY--QQQQCKQPCQPPP
Hs SPRR2B   MSY--QQQQCKQPCQPPP
Hs SPRR2D   MSY--QQQQCKQPCQPPP
Hs SPRR2E   MSY--QQQQCKQPCQPPP
Hs SPRR2F   MSY--QQQQCKQPCQPPP
Hs SPRR2G   MSY--QQQQCKQPCQPPP
Hs SPRR4    MSS--QQQQRQQQQCPPQ
Ac EDCP     MSY-----QCKQRCLPPP
Ac EDPQ2    MSY--QQQQCKQPCQPPP
Ac EDPQ3    MSSDSF--QCTQPCKAPP
Ac EDSPR2   MS-----QQCKQGCKAPP
Ac EDSQ     MSY-----QVKQASLPPP
Ac EDEPT    MSY-----QARQPCTAPP
Ac EDSPR1   MACPH--QQCKQPCLPPP
Ac EDPSQ    MYC--TDQQCKQACLPPP
Gg EDCQCM   MSYY---EQCKQPCLPPP
Gg EDPE     M-------QCKQEVTLPP
Gg EDYM1    MSYWY---QYKQQCFIPS
Cp EDYM1    MSYFAY--QYKQRNYTPY
Cp EDAA1    MFH--HQKICKPWCCKPH
Cp EDP3     MSS--DQQQCKQTCPPPP
Cp EDPCV1   MAY---QQQCKQPCLPPP
Cp EDPCV2   MAY---QQQCKQTCLPPP
Cp EDPCV3   MAY---QQQCKQTCLPPP
Cp EDPCV4   MAY---QQQCKQTCLPPP
Cp EDPCV5   MAY---QQQCKQTCLPPP
Cp EDPCV6   MAY---QQQCKQTCLPPP
Cp EDPCV7   MAY---QQQCKQTCLPPP
Cp EDPCV8   MAY---QQQCKQTCLPPP
Cp EDPCV9   MAY---QQQCKQTCLPPP
Cp EDPCV10  MAY---QQQCKQPCLPPP
Cp EDPE     MSLHQDQQQCKQGITLPP
Cp EDPQ1    MSY-QHQQQCKQTCLPPP
Cp EDPQ2    MSY-QHQQQCKQTCLPPP
Cp EDPL1    MSCHQHQQQCKQPCMPPP
```

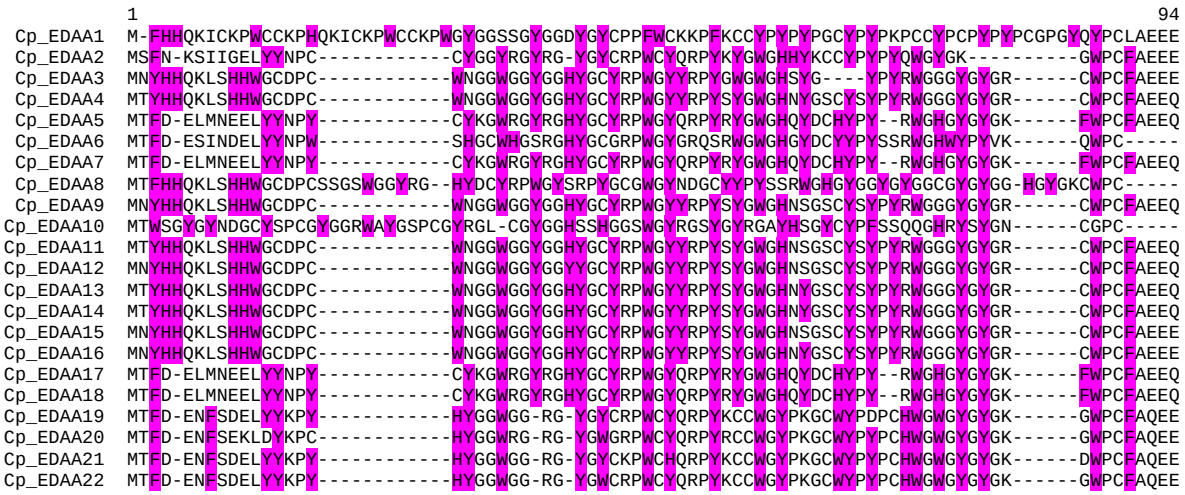
B

```
PGLYRPs     Hs PGLYRP3   ALYNIISTWPHF------KH-*
            Ac PGLYRP3   PIRAEISKWPNY------KHN*
            Cp PGLYRP3   PIRKVLKTWPHY------KH-*
SFTPs       Gg Crnn      QEHLQP-QWPP--------RK*
            Cp Crnn      KPRPFPHWWPP--------KK*
SEDCs       Hs Ivl       QQ-KQEVQWPP-------KHK*
            Hs Lor       QQ-KQAPTWPS---------K*
            Ac EDCQ1     QQVKQPTQWPS------QNQK*
            Ac EDCQ2     QQVKQPTQWPP------QNAK*
            Ac EDEPK     QQRKQPSTWPL---------K*
            Ac EDPKC     HQKKQPCYWPH--------HK*
            Ac Lor1      Q-TKQMNTWPSG-------QK*
            Gg EDGH      QQIKQSSQWPPS------QKK*
            Gg EDPE      QQVKQPSPWPLT-------QK*
            Gg EDQrep    KKYCSASKWPW---------*
            Gg EDQL      QQIKQPVQWPT------QQQK*
            Gg Lor1      QQT-QPISWPPQT-----KHK*
            Cp Lor       QQTKQPCQWPPNP------RK*
            Cp EDQL      QQTKQPCQWPPQ-----KHQK*
            Cp EDAA5     HGYGYGKFWPCFA----EEQ-*
            Cp EDP2      QQQKQPHHWPP-------KRK*
            Cp EDYM1     YPYPYAPQWPNTWGYGNCGPC*
```

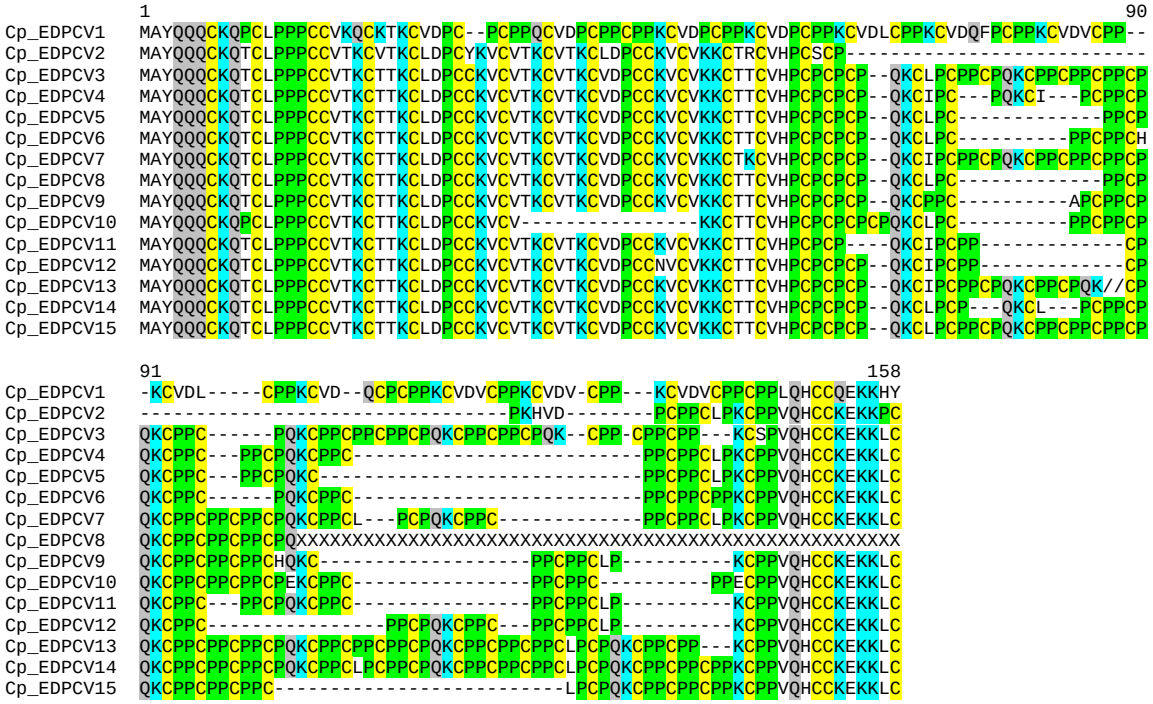
Supplementary Figure S6. Conserved amino acid sequence motifs at the amino-terminus (A) and carboxy-terminus (B) of EDC proteins. The amino acid sequence motifs of the lizard (*Anolis carolinensis*, Ac), chicken (*Gallus gallus*, Gg) and human (*Homo sapiens*, Hs) were discussed in detail in Strasser et al. (2014). The present study shows that these motifs are also conserved in several but not all EDC proteins of the turtle (*Chrysemys picta*, Cp). *, end of the protein.



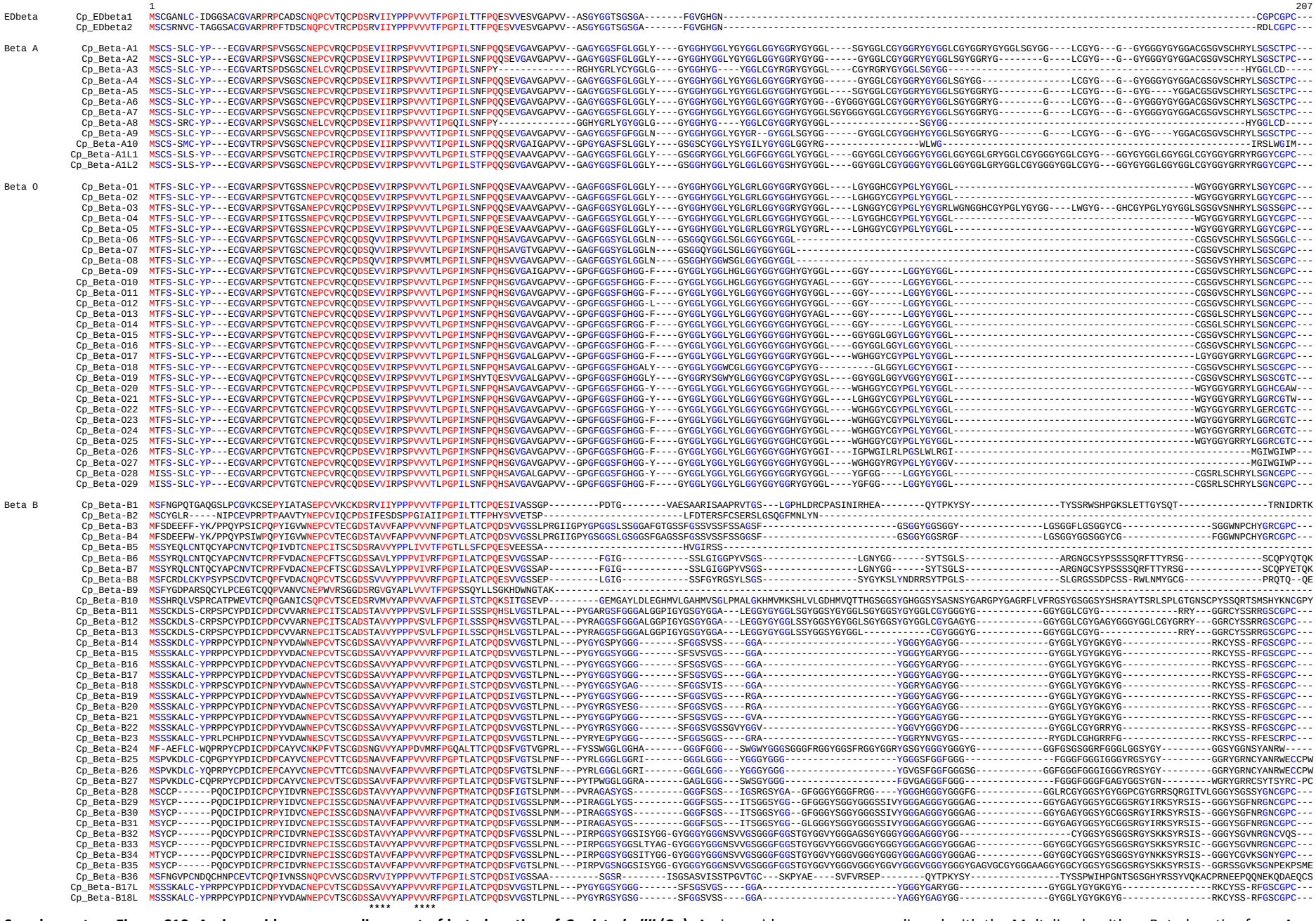
Supplementary Figure S7. Amino acid sequence alignment of epidermal differentiation proteins containing a glutamine (Q)-rich motif (EDQMs) of *C. picta bellii* (Cp). Amino acid residues implicated in covalent protein cross-linking (C-C, Q-K) are highlighted.



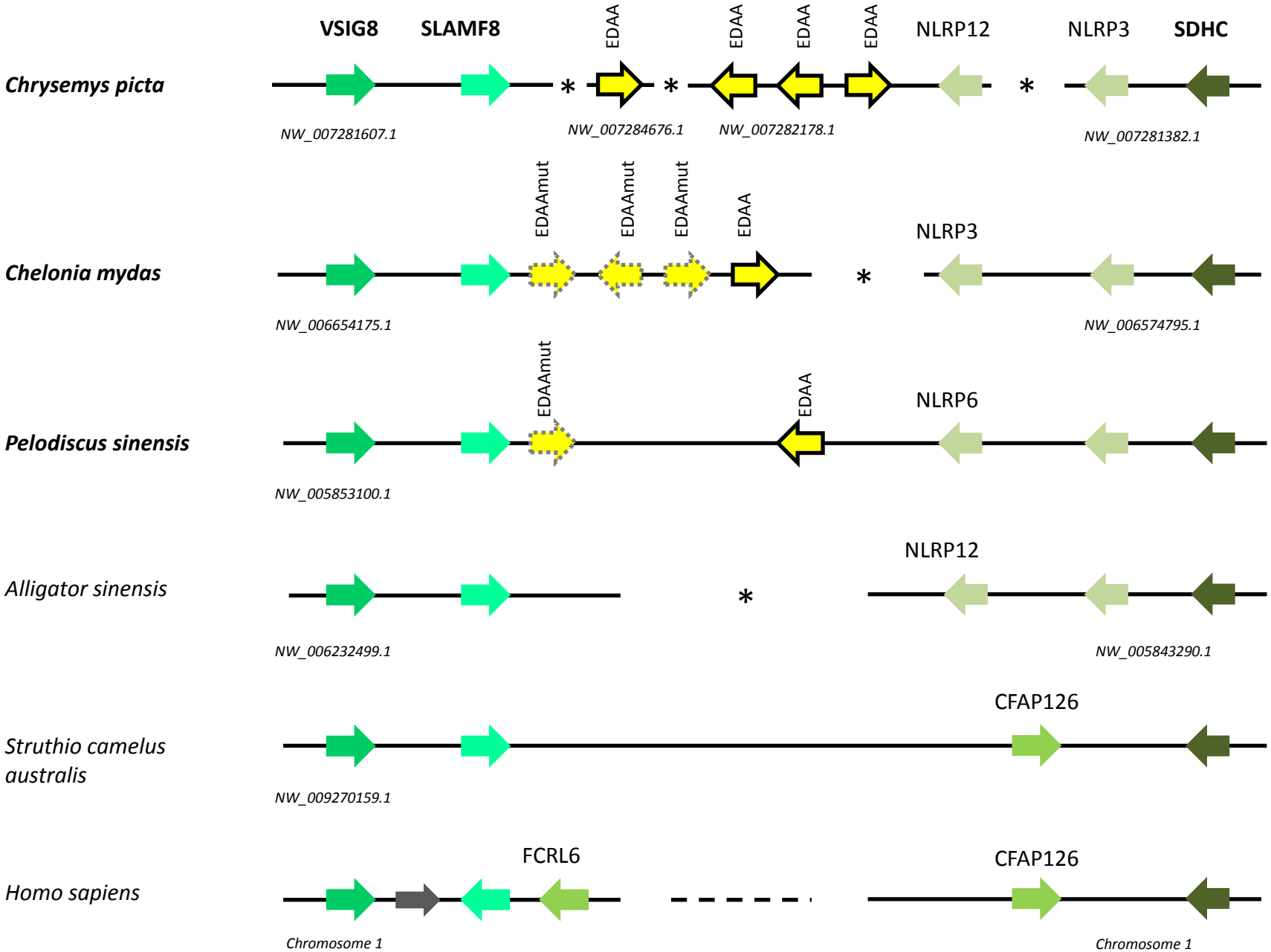
Supplementary Figure S8. Amino acid sequence alignment of epidermal differentiation proteins rich in aromatic amino acids (EDAAs) of *C. picta bellii* (Cp). Aromatic amino acids are highlighted.



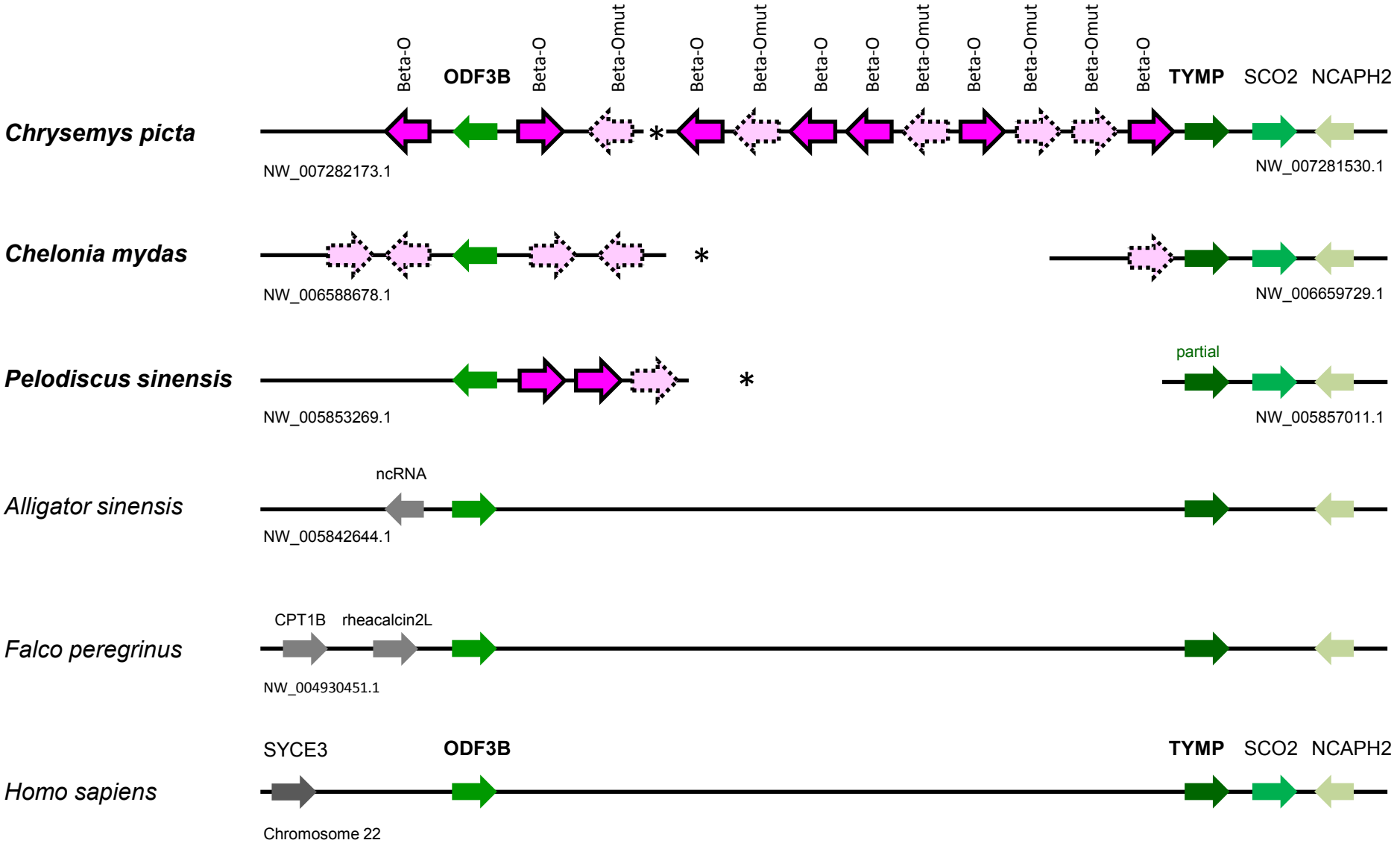
Supplementary Figure S9. Amino acid sequence alignment of epidermal differentiation proteins rich in proline, cysteine and valine (EDPCV) of *C. picta bellii* (Cp). Proline and amino acids implicated in covalent protein cross-linking (C-C, Q-K) are highlighted. At the position indicated by "//" a part of the amino acid sequence of EDPCV13 has been removed to facilitate alignment of the carboxy-terminal sequences. X, unknown residues because of incomplete gene sequence.



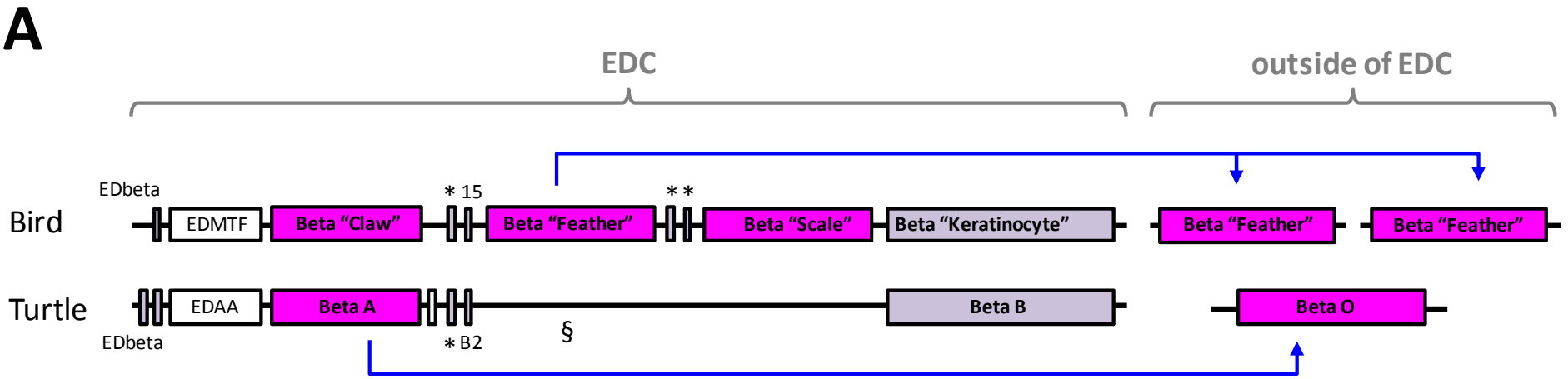
Supplementary Figure S10. Amino acid sequence alignment of beta-keratins of *C. picta bellii* (Cp). Amino acid sequences were aligned with the Multalin algorithm. Beta-keratins form 4 groups encoded by genes at different loci. EDbeta, Beta A, and Beta B proteins are encoded by genes arranged in this order within the EDC whereas Beta O proteins are encoded by genes outside the EDC. The group of Beta A proteins shows high sequence similarity to Beta O proteins. Beta-A1L1, A1L2, B17L, and B18L are located on unplaced single-contig-scaffolds. Beta A and B proteins are numbered according to the order of their genes in the EDC. Red, conserved in 90% or more sequences; blue, conserved in 50% or more sequences; /, additional sequence was removed to allow alignment of the N-terminus. Asterisks indicate positions of 2 beta-sheets.



Supplementary Figure S11. Turtle EDAA gene locus outside of the EDC and homologous loci without EDAA genes in other amniotes. The conserved genes VSIG8, SLAMF8 and SDHC flank a chromosomal locus that contains EDAA genes in turtles but not in other amniotes. The relative arrangement and the orientation of genes is schematically depicted. The schematics are not drawn to scale. For detailed information about EDAA gene names and positions, see Suppl. Tables S1-S4. Colors indicate gene homology. EDAA genes carrying deleterious mutations have frames with a broken line. The numbering of NLRP (NLR family, pyrin domain containing) proteins in reptiles is uncertain. GenBank accession numbers are shown below the various scaffolds. A broken lines indicates a long region of human chromosome that contains too many genes to be shown here. The genome of the ostrich (*S. camelus australis*) is shown as representative of phylogenetically basal birds. Asterisks indicate genome sequence gaps.



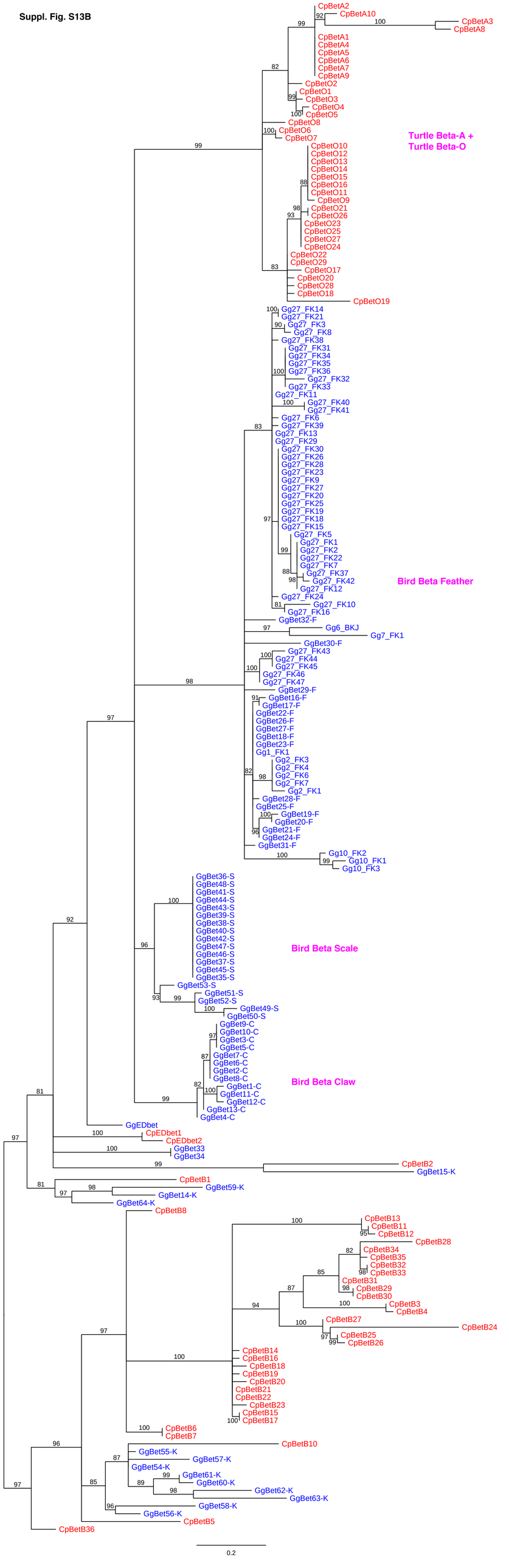
Supplementary Figure S12. Beta-keratin (Beta-O) gene locus outside of the EDC in turtles and homologous loci without beta-keratin genes in other amniotes. The conserved genes *ODF3B* and *TYMP* flank a chromosomal locus that contains beta-keratin genes in turtles but not in other amniotes. The relative arrangement and the orientation of genes is schematically depicted. The schematics are not drawn to scale. Colors indicate gene homology. Pink arrows indicate beta-keratin genes comprising apparently functional open reading frames. Light pink and frames with broken lines indicate the presence of deleterious mutations. For detailed information about gene names and loci, see Suppl. Tables S1-S4. GenBank accession numbers are shown below the various scaffolds. In birds, represented here by peregrine falcon (*F. peregrinus*), this locus is not well conserved but consistently devoid of beta-keratin genes. Note that high sequence similarities to the genes on the above scaffolds indicate that additional genes (also designated Beta-O) on other short scaffolds (Suppl. Tables S2-S4) belong to the same locus. Asterisks indicate genome sequence gaps.



(continued on the next page)

Supplementary Figure S13. Gene locus comparison and phylogenetic analysis of beta-keratins (also known as corneous beta-proteins) of a turtle (*C. picta bellii*) and a bird (*G. gallus*). (A) Schematic diagram of beta-keratin gene loci of a bird (chicken, *G. gallus*) and a turtle (western painted turtle, *C. picta*). For an overview of the genes of the surrounding locus, see Figure 2. Clusters of more than 2 similar beta-protein genes are shown as long boxes (length not proportional to the number of genes) whereas single genes are depicted as short boxes. The names of the gene clusters of the turtle are defined in the main text. EDMTF (Strasser et al. 2014) and EDAA genes are indicated as white boxes whereas beta-keratins are indicated by pink and violet shading with pink clusters being closely related according to molecular phylogenetics (panel B). The names of chicken beta-keratin clusters clade are adapted from Ng et al. (2014) and Greenwold et al. (2014). Putative gene translocations linking clusters of related beta-keratin genes (see panel B) within the EDC (left) and outside of the EDC (right) are indicated by blue arrows. Genes marked by asterisks belong to the groups Beta-B or bird Beta "Keratinocyte". The turtle gene EDAA10 is located between the clusters Beta A and Beta B of the turtle. §, the length of the line between Beta B2 (orthologous to chicken beta 15) and the rest of the Beta B cluster of the turtle is not proportional to the physical distance but indicates that this locus does not contain orthologs of avian feather and scale beta-keratins. **(B, next page)** Maximum likelihood phylogeny of beta-keratins of the western painted turtle *C. picta* (Cp, in red) and chicken *G. gallus* (Gg, in blue). Support of phylogenetic groups was computed by the ultrafast bootstrap approximation approach (UFBoot) (see Materials and Methods for more information). Since UFBoot support values behave like posterior probabilities, branches with support values of at least 90% and 95% are regarded as supported and strongly supported, respectively. Branches with support lower than 80% are not shown (see Materials and Methods for more information). For clarity, mid-point rooting was used to draw the tree, however, the true position of the root is unknown. The strongly supported monophyletic groups of avian feather, scale and claw Beta proteins as well as the monophyletic group of turtle Beta-A with Beta-O proteins are indicated (pink fonts) on the right. The relationships within these clades could not be resolved because of the limited phylogenetic information in the underlying sequence alignment. Partial Cp beta-keratins (Suppl. Fig. S2C), the highly derived and possbily pseudogenic sequence Cp_Beta-B9, Cp beta-keratins Beta-A1L1, A1L2, B17L, and B18L encoded by genes on unplaced single-contig-scaffolds, and incomplete feather beta-keratin sequences (Ng et al. 2014) were not included in the phylogenetic analysis. In the names of feather beta-keratins encoded by genes outside the EDC (labeled FK), the chromosome number is indicated after the species code. For example, Gg27 indicates *Gallus gallus* chromosome 27). Other labels of Gg sequences: F, feather beta-keratin encoded by gene within the EDC on chromosome 25; S, scale beta-keratin (EDC); C, claw beta-keratin (EDC); K, keratinocyte beta-keratin (EDC).

Suppl. Fig. S13B



A

```
          1                                                          60
Cp_EDQL   MCSREPRGCHDSGSSSCHDSGSSTCHSSGGGSCHDVKPL----------PQCPT-PVPCQ
Gg_EDQL   MCSRADRGCHSSESSSCH-SGGSSCHGSEEVTCHEVSAVQDGTPVVVLQPQCPVVTVPTQ
          ****  **** * ***** ** * ** *    ** *             ****   ** *

          61                          88
Cp_EDQL   --TTTLPC---QQQTKQPCQWPPQKHQK
Gg_EDQL   GPVAPVPCQQQQQQIKQPVQWPTQQ-QK
             **   *** *** *** *  **
```

B

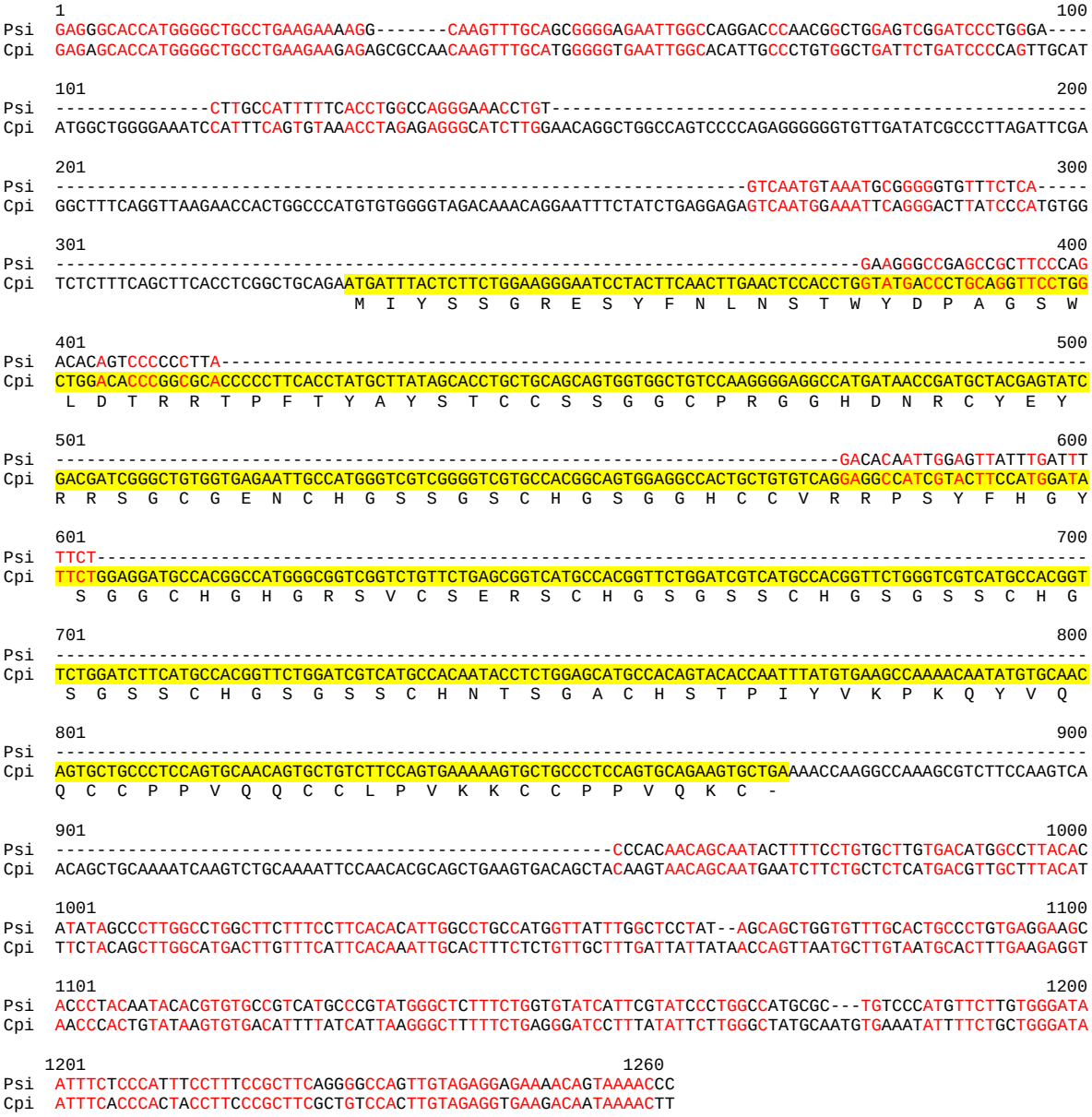
```
          1                                                          60
Cp_EDP3   MSSDQQQCKQTCPPPPKCQEKCPPPCKEPVKTPKCQEKCPPPSKEPKCPPPKQSQDWKQC
Gg_EDP3   MSSHQQ--KQQQQIPAQCQQKCPPKGVEQCQAPKGQTKCPVKSIPQQ---QQQQQCPKQK
          *** **  **    *  ** **** *    ** * ***  *        * *  **
```

Supplementary Figure S14. Examples of proteins encoded by orthologous SEDC genes of the western painted turtle and the chicken. Orthology of genes was inferred from reciprocal highest sequence similarity and gene locus synteny. Amino acid sequences of EDQL **(A)** and EDP3 **(B)** proteins were aligned. Note that EDQL of the chicken was previously named EDQM3 (Strasser et al., 2014). The EDP3 gene of the chicken had been missed in our previous study of the chicken EDC (Strasser et al., 2014). Here, we identified the coding sequence of chicken EDP3 in the EDC on chromosome 25 (Accession number: NC_006112.2, nucleotide positions 1108303-1108470, reverse orientation). Asterisks below the alignments indicate identical amino acid residues. In the alignment of EDQL sequences, the carboxy-terminal sequence similar to the carboxy-terminus of loricrin (Suppl. Fig. S6B) is underlined. Amino acids involved in cross-linking via cysteine bridges (C) and transglutamination (Q, K) as well as proline residues (P) are highlighted by color-shading as in Suppl. Fig. S1. Cp, *Chrysemys picta bellii*; Gg, *Gallus gallus*.

A

Suppl. Fig. S15. The *EDWM* gene is deleted in the soft-shell turtles. Continued on the next page.

B



Supplementary Figure S15. The *EDWM* gene is deleteriously mutated in soft-shell turtles. (A) Amino acid sequence alignment of EDWM proteins. *EDWM* is conserved in representatives of all main clades reptiles whereas the conceptual translation of an *EDWM* gene fragment of the softshell turtle *Apalone spinifera* (Asp) shows inactivation. An "X" on black background indicates the premature end of EDWM because of an in-frame stop codon in the *Apalone spinifera* sequence. Because of the end of the genomic sequence contig, the conceptual translation of Asp EDWM is incomplete (indicated by question marks). Red letters indicate residues present in all species and blue letters indicate residues present in more than 50% of the species (except *Apalone spinifera*). Aca, *Anolis carolinensis*; Asi, *Alligator sinensis*; Cmy, *Chelonia mydas*; Cpi, *Chrysemys picta bellii*; Gga, *Gallus gallus*; Oha, *Ophiophagus hannah*. **(B)** The coding sequence of *EDWM* has been lost in the softshell turtle *Pelodiscus sinensis*. DNA sequences from the predicted *EDWM* locus within the EDC region of *Pelodiscus sinensis* (Ps) and *Chrysemys picta bellii* (Cp) were aligned. Identical nucleotides are shown in red. The coding sequence of the *EDWM* gene of *C. picta* is highlighted by yellow shading. The amino acid sequence is shown below the coding sequence. Cp, Chrysemys_picta_bellii-3.0.3 Scaffold107, whole genome shotgun sequence, gi|636526453:c1011496-1010237 (reverse complement); Ps, PelSin_1.0 scaffold1810, whole genome shotgun sequence, gi|557455322:c36333-35817 (reverse complement).

A

```
            1                                                          60
  Cp_EDP1   MPYYGQQHKH--LPAPVCVTKCSQPCPPQYEQHCVPKCRPVYVTKCPPLYGPQYAYPCAP
  Cm_EDP1   MPYYGQQHKQLCLPPPACVTKCSQPYPPQYEQQCVPKCRPVYVTKCPPWYGPQYAYPCAP
Ps_EDP1L2   MTYYGRKHQQHCLPSPACVAKCPQPCRPQYEQHCAPKCQPVYVTKCPPLYGPQYAFPCAA
 As_EDP1p   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

            61                                                        120
  Cp_EDP1   QCPPRCVTKCPPRCVTKCPPPCVTKCPPPCVTKCPPPCVTKCPPPCVTKCPPPCVTKCPP
  Cm_EDP1   QCPPPCVTKCPPPC----PPPCVTKCPPPC----------------------------
Ps_EDP1L2   QCPPRCVTKCPPPCVTKCPPQCVTKCPPPCVTKCPPQCVTKCPPPCVTKCPPQCVTKCPP
 As_EDP1p   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

            121                                                       180
  Cp_EDP1   PCVTKCPPPCVTKCPPPCVTKCPPPCVTKCPPPCMTKCPQQCVTQCP------------
  Cm_EDP1   ----------------------PPPCVTKCPPPCVTKCPQHCVTQYP------------
Ps_EDP1L2   PCVTKCPPQCVTKCPPPCVTKCPPQCVTKCPPPCVTKCPPQCVTKCPPPCVTKCPPQCVT
 As_EDP1p   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXPPCVTKCPPPCVTKCPPPCVT

            181                                                   232
  Cp_EDP1   ---------------------------GQYQSGKVQISSHCKKYCSAPKWPW
  Cm_EDP1   ---------------------------DQYQSGKVQISSHGKKYCSGPKWPW
Ps_EDP1L2   KCPPQCVTKCPPRCVTKCPQQCVTQYPGQCQSGNIKMSSQCKKYCSAPNWPW
 As_EDP1p   KCPPPRVTKCPPRCVTKCPQQCVTQYPGQCQSGNIKMSSHCKSYCSTPKWPW
```

B

```
            1                                                          60
  Cp_EDP2   MASRQNQQQRKQTLTLPPALSNATSEPAPPPEAVPEPCPATVEEPENSPQEEEGPQEEYK
  Cm_EDP2   MASQQNQQQRKQTLTLPLALSNATSEPAPTPEAGPEPCPATVEERENSPQEEEESQEEYK
  Ps_EDP2   MASPQNQQQRRQSLPLPPALSNAAPEPEPSPG------PRTVKEPENAPREEEKPQKE--
 As_EDP2p   MASPQNQQQRRQILTLPPALSNATPEPEPSPEVARDPGPTTVXXXXXXXXXXXXXXXXXX

            61                                                        120
  Cp_EDP2   QPLNQPLGPAPELEPEPVLCPEP--ESNPPEVKEIEYLQPDHQQYKHPPTLPPAPGMETS
  Cm_EDP2   RPLNQPLGPAPELEPEPVLGPEP--ESNPSEVKEIEYLQLDQQQYKHPPTLPPAPGIETS
  Ps_EDP2   -PLDQPPGPVPELEPEPEPEPEPAPEPNPPEAEEAGYLQPEQQQYKQPPALPPAPGAETS
 As_EDP2p   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

            121                                                   172
  Cp_EDP2   KEYQQAESE--PELGRCPPPIREPEGPPFVQPSSPVEEQQQKQPHHWPPKRK
  Cm_EDP2   KEYQQAEPELEPEPGRCPPPISEAEGPLFVQPSSPVEEQQQKQPHHWPPKRK
  Ps_EDP2   TECEEAKPEPEPEPGRCPPPISEPEGPGPVQPSPPGEEKQQKQPCRWPPARK
 As_EDP2p   XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

Supplementary Figure S16. Amino acid sequence alignments of orthologous EDC proteins of 4 species of turtles. Continued on the next page.

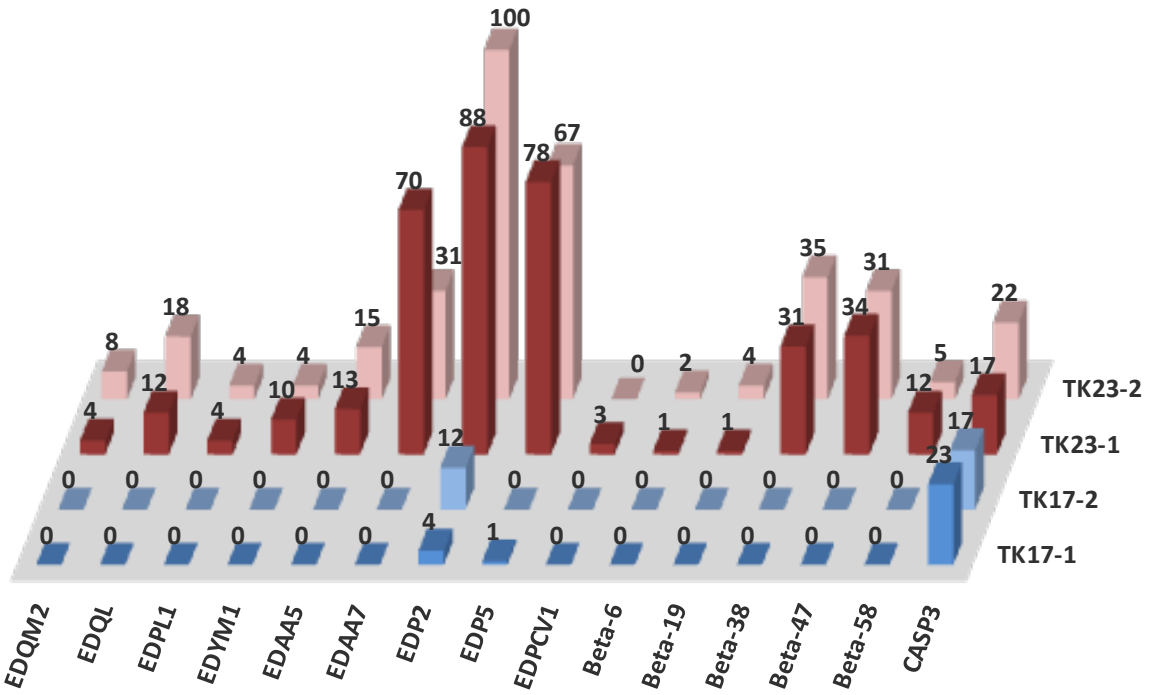
C

```
          1                                                                 60
 Cp_EDQL  MCSREPRGCHDSGSSSCHDSGSST------------------------------CHSSGG
Cm_EDQLp  MCSREPHGCHDTGSSSCHDTGSSSSPDTGSSFCXXXXXXXXXXSSSCHDTGSSSCHGSGG
 Ps_EDQL  MCSREPRGCPDSERSSCPSSERSS------------------------------CHGSEA
As_EDQLp  MCSREPRGCPDSERSSCPDSERSXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX

          61                                        102
 Cp_EDQL  GSCHDVKPLPQCPTPVPCQTTTLPCQQQTKQPCQWPPQKHQK
Cm_EDQLp  GTCHDVKPLPQCPIPVPCQTTTIPCQQQTKQPCQWPPQKHQK
 Ps_EDQL  TTCHDVKPHPQYPTTVPCQTPTSPCQQQTKQTCPWPPQKHQK
As_EDQLp  XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

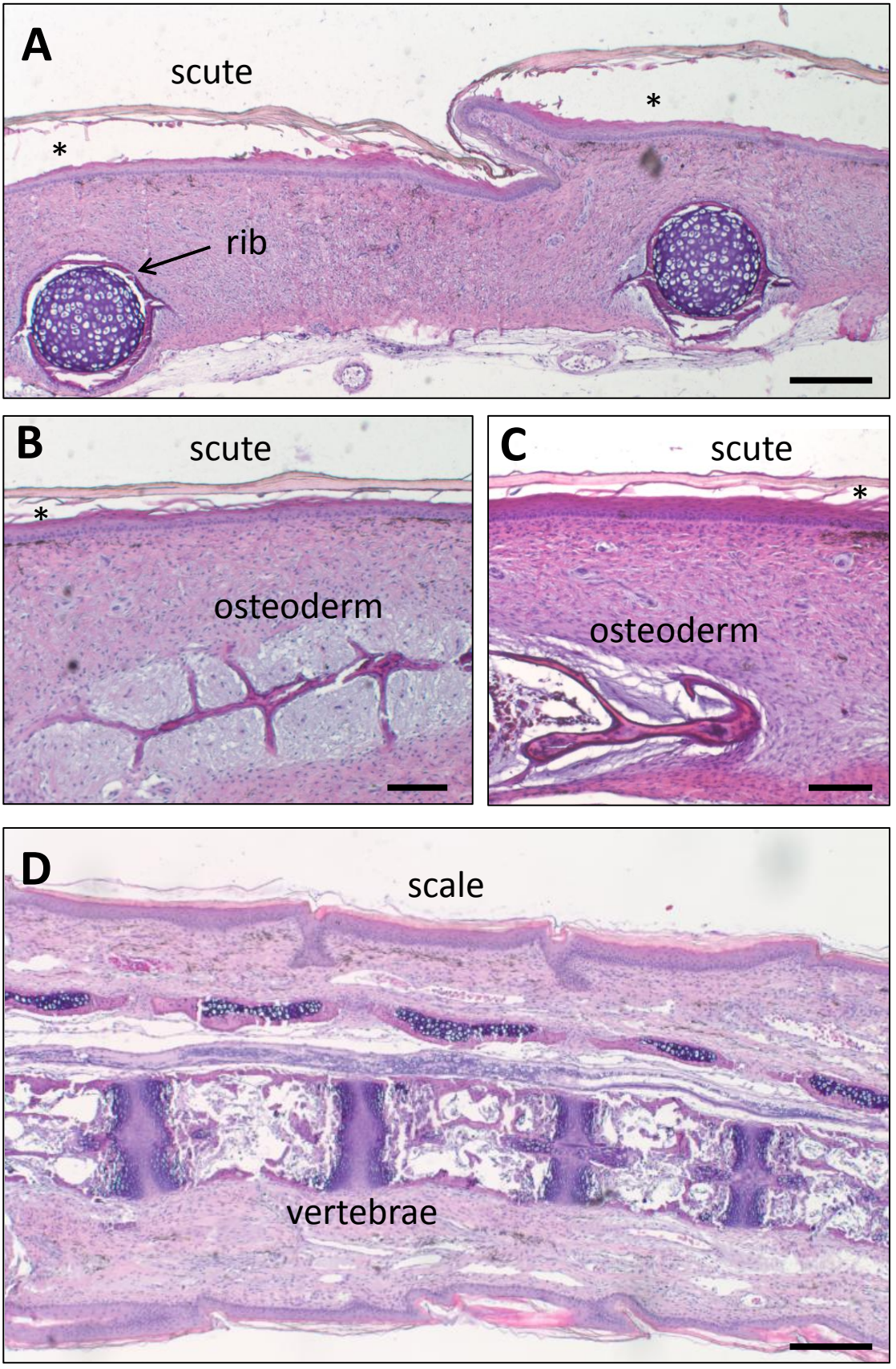
D

```
          1                                                                 60
Cp_EDYM1  MSYFAYQYKQRNYTPYSTTRLIPHAEPCVVKGPAPRVTKCADPCAVKHPAPCTTKCRDPC
Cm_EDYM1  MSYFAYQYKQRNYTPYSATRLVPPAEPCVVKGPAPPGTKCAETCAVKHPAPCTTQCRDPC
Ps_EDYM1  MSYFAYQYKQRNYTPYSTTRLLACAEPCVVKGPAPCGTKCVEPCATKRPAPCVPKCRDPC
As_EDYM1p XXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXKCRDPC

          61                                                        120
Cp_EDYM1  AGKPSVPCATKCFEPHAQRHPAKHYPKFSEPAGVKCSTPCDTRYHEPYGLIHPQPFPERW
Cm_EDYM1  AAKPSVPCATKCFEPHAQRHPAQYIPKFSEPVGVKCSTPCVTRYHEPYGLIHPQPFPERW
Ps_EDYM1  AGKAPVHCEPKCLEPHAQRGPAHCAPKFSEPAGVKCSVPWVPRCHEPYGPIPARPFPERW
As_EDYM1p AGKTSIHCEPKCLEPHAQRGPAHCAPKFSEPVGVKCSVPWVPRCHEPYG-----PVPERW

          121                                                 170
Cp_EDYM1  NPCAPPYVHPYVTGYPQACGPTYVPSFPKYPYPYAPQWPNTWGYGNCGPC
Cm_EDYM1  NPCAPPYVHG---GYPQACGPTYVPSFPKYPYPYAPQWPDTWGYGNCGPC
Ps_EDYM1  NPCAPPYGQPFVTGYPQACGPSYGPSFPKYSYPCAPQWPGGWGYGGCGPC
As_EDYM1p NPCAPPYGQPFVTGXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXXX
```

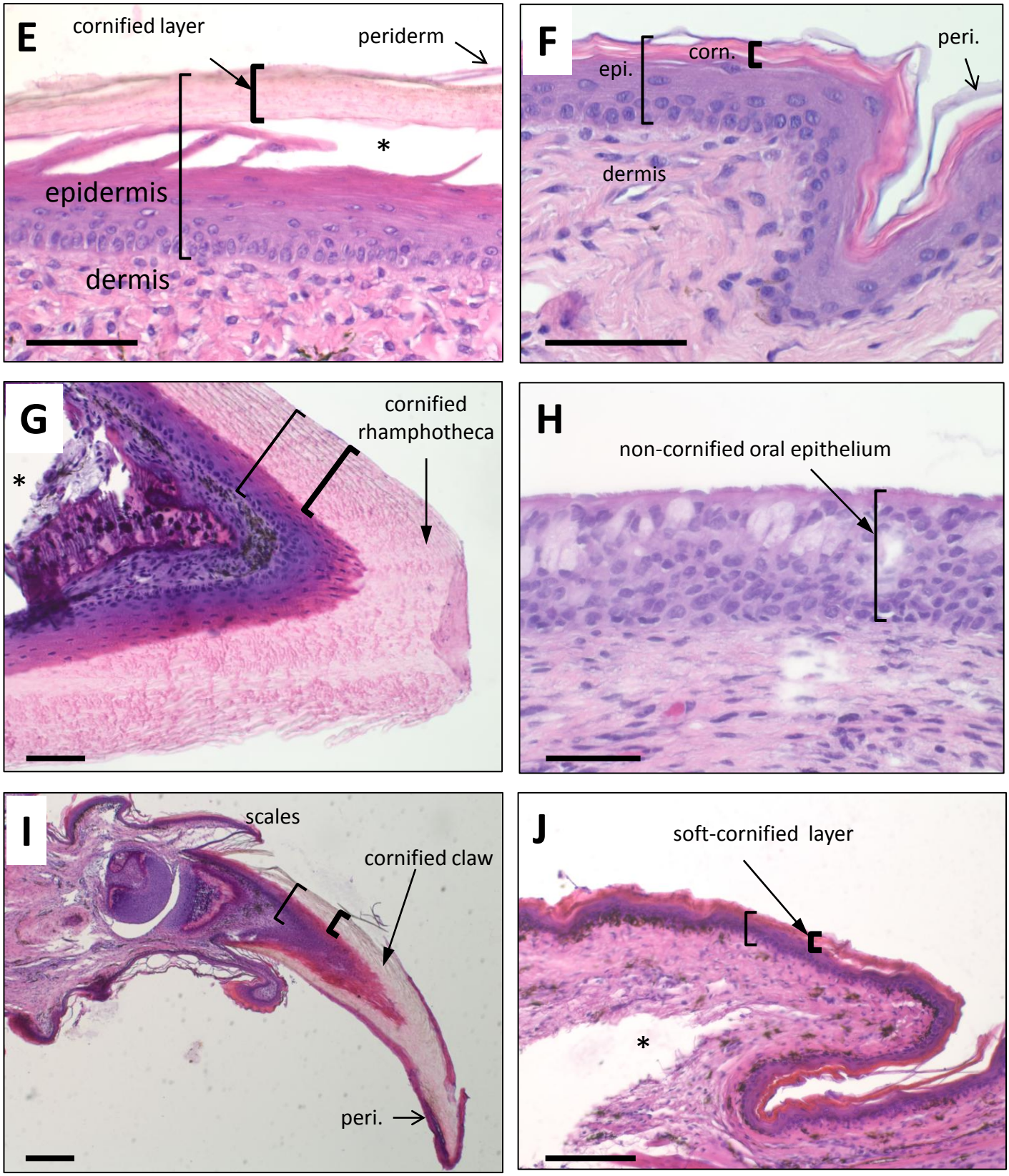
Supplementary Figure S16. Amino acid sequence alignments of orthologous EDC proteins of 4 species of turtles. The amino acid sequences of EDP1 (A), EDP2 (B), EDQL (C) and EDYM1 (D) of *Chrysemys picta* (Cp), *Chelonia mydas* (Cm), *Pelodiscus sinensis* (Ps) and *Apalone spinifera* (As) were aligned using the Multalin algorithm. Red fonts, residues conserved in all sequences; blue fonts, conserved in 50-75% of the sequences; p, partial amino sequence; X, unknown amino acid residue. Note that the highly fragmented genome sequence of *A. spinifera* did not allow the prediction of complete sequences.



Supplementary Figure S17. RNA sequencing (RNA-seq) data suggest upregulation of EDC gene expression during epidermal maturation in embryos of *Pelodiscus sinensis*. To estimate the expression levels of EDC genes during embryonic development of the soft-shelled turtle *P. sinensis*. RNA-seq data deposited in the GenBank sequence read archive (SRA) (Wang et al., 2013) were screened by tBLASTn using the sequence of the first 17 amino acid residues of each protein as query. The RNA-seq reads yielding a 100% match to the query sequence were counted and plotted over each sample (Tokita and Kuratani (TK) development stage - number of replicate). Data from 2 biological replicates of various embryonic stages were analyzed. The results obtained for the two final developmental stages, TK17 and TK23, of Wang et al. (2013) are shown. A subset of the EDC genes of *P. sinensis* (Supplementary Figures S2B, C) were investigated. The ubiquitous pro-apoptotic protease, caspase-3 (CASP3), was used for comparison. As the amino-terminus of the predicted caspase-3 protein of *P. sinensis* (XP_006128558.1) appears to be incorrect, we used the sequence of residues 28-44 as query for CASP3. Note that the expression levels of all EDC genes increased from development stage TK17 to stage TK23 whereas the expression of CASP3 remained unchanged. Accession numbers of transcriptome data: DRX001551 (TK17, sample 1), DRX001552 (TK17, sample 2), DRX001553 (TK23, sample 1), DRX001554 (TK23, sample 2). References: Tokita M, Kuratani S. 2001. Normal embryonic stages of the chinese softshelled turtle Pelodiscus sinensis (Trionychidae). *Zool Sci*. 18:705-715. Wang Z, Pascual-Anaya J, Zadissa A, Li W, Niimura Y, Huang Z, Li C, White S, Xiong Z, Fang D, et al. 2013. The draft genomes of soft-shell turtle and green sea turtle yield insights into the development and evolution of the turtle-specific body plan. *Nat Genet*. 45:701-706.



Supplementary Figure S18. Hematoxylin & eosin (H&E) staining of tissues of the European pond turtle (*E. orbicularis*) on embryonic day 45. Continued on next page.



Supplementary Figure S18. Hematoxylin & eosin (H&E) staining of tissues of the European pond turtle (*E. orbicularis*) on embryonic day 45. Tissue samples were taken from embryos of the same developmental stage that was used for RNA preparations (Fig. 4). (A, B, E) Carapace, (C) plastron, (D) tail, (F) leg, (G) rhamphotheca, (H) oral epithelium, (I) toe including claw, (J) neck skin. Asterisks indicate artifacts of tissue preparation. In panels E-J the epidermal compartment is indicated by a thin-lined bracket, and the cornified epidermal compartment is indicated by a thick-lined bracket. Note the presence of the periderm (peri), an embryo-specific layer. Scale bars: (A, D, I) 200 µm; (B,C, J) 100 µm; (E, F, G) 50 µm; (H) 40 µm.

A

>Eo_Beta-A1_partial
GGTTTACCTCCATCACAGAAAGATGTCTTGCTCCAGCCTGTGCTATCCAGAATGCGGGGTGGCCCGNCCCAGTCCAGTT

>Eo_Beta-A4_partial
GGTTTACCTCCATCACAGAAAGATGTCTTGCTCCAGCCTGTGCTATCCAGAATGCGGGGTGGCCCGTCCCAGTCCAGTT

>Eo_Beta-B4_partial
GGCTCANCCCCATTGCAGAAAGATGTTTTCCGATGAGGAGTTCTCTTACAAAAGTCAGCAGCGACATATACCGCAGAAAGGCCAAAACCCATGTCTGCCGCAGAAAG
AGACAAAACCGTGTCCGCCGCAGAAAGTGCCATGTCCGCCAAAGCGTCCTCCGTGTCCAAAGTATCCACCATGCATTCCGTATCCACGGCCATGCCCACCTCAGTAT
CCATCAATATGCCCCAACCATATATTGGGGTCTGGAACGAACCGTGTGTCACAGAATGCGGTGACTCGACTGCAGTCGTCTTTGCACCACTGGTTGTCGTGAATTTC
CCAGGCCCAACTCTTGCCACTTGCCCACAGGATAGCGTAGTGGGATCCTCTCTACCAAGAGGAATTATAGGCCCATATGGTTCCGGGGGCTCTTTCGGTAACGGGGG
CTTATCCTTCAGCAACGGGAGCTCATCCTTCGGTACCGGGAGCTCATTCTTCGGCTCCTGGGGT

>Eo_Beta-B19_partial
GGNTATACCACAATCACAAAAGATGTCTTCCAGCAAAGCTTTGTGTTACCCACGCCCGCCATGTTATCCCGACATCTGCCCCGACCCATATGTTGATGCCTGCAACG
AGCCTTGTGTCACATCATGTGGTGACTCGAGCGCAGTCGTCTATGCCCCACCGGTTCTCGTGAGATTCCCGGGACCAATTCTCGCTACCTGCCCTCAAGACAGC

>Eo_Beta-B32_partial
GGATTCCCTCCATCACAGAAGATGWCTTACTGCCCACCCCAGGATTGTTATCCCGATATATGCCCACGTCCATGTATTGACGTCCGCAACGAGCCGTGTATCTCATC
ATGYGGAGATTCGACTGCAGTG

>Eo_Beta-O17_partial
GGTTTACCTCCATCCCAGAAAGATGACTTTCTCCAGCCTGTGCTATCCAGAATGCGGGGTGGCCCGGCCCTGTCCAGTCACTGGNACCTGCAACGAGCCGTGCGTTA
GGCAGTGCCAGGACTCCGAAGTGGTGATCAGACCCTCACCGGTTGTC

>Eo_EDAA8_partial
TTGGTGAACGGGGTTTACCTCCAAGGCAGAAAGATGACTTTCCACCACCAAAAGCTTTCTCACCATTGGGGCTGTGACCCCTGTTGGAAAGGGGGCTGGGGCGGTTA
TGGGGGCCATTACGGCTGCTATCAGCCATGGGGTTATTCTAGACCATACGGATGGGGGTGGGGCCACAATTCCGGTTCCTGCTACTCTTACCCTTACCGATGGGGGG
GTGGTTACGGCTATGGA

>Eo_EDAA19_partial
CTGGGTTTACCGCCYACACAGAAAGATGACTTTCGATGAAAACTTTAGTGATGAATTGTATTATAAGCCCTACCATTATGGAGGCTGGGGTGGTAGAGGATACGGCT
ACTGTAAGCCATGGTGCTACCAAAGGCCATACAAGTGTTGTTGGGGCT

>Eo_EDbeta1_partial
TGTTACCTCCATCTCAGAAAGATGTCGTGCGGGGCAAACTTGTGCATTGATGGGGGCTCAGCGTGCGGTGTGGCACGCCCCAGGCCATGCGCTGACAGCTGCAACCA
ACCATGCGTCACGCAGTGCCCCGACTCGAGAGTGGTGAT

>Eo_EDKM_partial
TCTAACCTCATAAAGGCCATCGCAGATATGATTAACAGCTACCAGAGCAACTCCAGAAAGGGCAGGGAATCAGAGAGGTTCCGGAGGTGTGAATTTAAGAAGTTGGT
CCAGCAGGAGCCGTCTCCTGCCAAGAGATCATCCAGCAATAAATATAAACACACTACGAGCTTGCCAGATTCTGACGCGGAGCTGATGAACAAGAAAGAACTCATCA
CTGCCAATCCCTGTGTGTACTGAAG

>Eo_EDP3_partial
AGGTTTACCTAGATCATTGAAAGATGTCTTCTGATCAGCAACAGTGTAAGCAGACCTGCCCNCCTCCTCCANAATGCCAGGAGAAATGCCCTCCACCATGCAAGGAG
CCCTGCAAGACTCCNAAATGCCAGGAGAAATGCCCTCCGCCATGCAAGGAGCCCTGCCCTCCCAAATGCCCCCCTCCACAG

>Eo_EDPCV_partial
TTTTGCCCNCNTCCCAGGAAGATGGCTTACCAGCAGCAATGCAAACAGACCTGCCTKCCCCCTCCTTGCTGTGTGACCAAGT

>Eo_EDPE_gDNA_partial
CCTGAGCCAATACCATGTTGTCCAGNGAAGCCACCATGCAAGGAACCACCAGTCCCAATTCNTACTCCATGCCCTGAGCCAATACCATGTCCTCANGAGAAGCAACA
GTGCAAGNTGCCACCAGTCCCAGTTCCTCTTCCACACCCTGAGCCAATACCNTGTTCTCCAGAGAAGCCACCATGCAAGGAACCCCCATTCCCANTTCCTACTCCAT
GCCCTGAGCCAATACCTTATTGTCCAGAAAAGCCACCATGCAAGGAACCAAAAGTCCCACNTCCTCTTCCACACCCTGAGCCAATACCATGTNNTCMAGAGAAGCNR
CNNTGCAAGNAGCCACCAGTCCCAGTTCCTCTTCCACGCCCTGAGCCAATACCNTGTTCTCCAGANAAGCCACCATGCAAGGAACCACCAGTCCCAGTTCCTNTTCC
ACACCCTGAGCCAATACCTTATTGTCCAGANAAGCCGCCATGCAAGNAACCACCAGNCCCNGTTNCTCNTCCANGCCCTGAGCCAATACCATGTTNTCCAGAGAAGC
CNCCATGCAAGGAACCACCAKTCCCAGTTCCTCYTCCACACCCTGAGCCAATACCATGTCCTCAACAGAAGCAGCAGAGCAAGTTGCCACCAGTCCCAGTTCCTACT
CCATGCCCTGAGCCAATACCTTGTTCTCCAGAGAAGCCGCCATGCAAGGAACCACCAGCCCCCGTTACTCCTCCATGCCCNGAGCCAGT

>Eo_EDQM1_partial
AGGGGTTACCCAAGCTACAGAAAGATGTGCTCCCGCCAGGAGAAAGACCACTGCCACAAACAAGATACCTGCCACGGGAGCGGAGGAGGGTCATCTTGCCACGGGAG
CGGAGGAGGATCATCTTGCCACGGGAGCGGAGGAGGATCATCTTGCCACGGGAGCGGAGGAGGGTCATCTTGCCACGGGAGCGGAGGAGGGTCATCTTGCCATGGGA
GCGGAGGAGGGTCATCTTGCCACGGGAAACCACAGAAGCCTTGCCAACAGGAGCAGCAACAGCAG

>Eo_EDQM7_partial
GGGTTACCCAAGCTGCAGAAAGATGTGCTCCCGCCAGGAGAAAGACCAGTGCCACAAACAAGACGGCAGCGGAGGATGCCACAGCAGCGGATCCTCTTGCCATAGTA
GTGGAGGATCTTCTTGCCACAGTGGAGGTTCATCTTGCCATGGGAGCGGAGGGTCTTCTTGCCATAGTAGTGGAGGTTCATCTTGTCATGGTGGAGGMTCATCTTGC
CACAGCGGAGGGTCTTGCCATGGGAAACCACAGCAGCACTGCCAGCAGCAGCAGCAGAA

>Eo_EDWM_partial
GCTTCACCTCGGCTGCAAAATGATTTACTCTTCTGGAAGGGAATCCTACTTCAACTTGAACTCCACCTGGTATGACCCTGCAGGTTCCTGGCTGGACACCCGGCGCA
CCCCCTTCACCTATGCTTATAGCACCTGCTGCAGCAGTGGTGGCTGTCCAAGGGGAGGCCATGATAACCGATGCTACGAGTATCGACGATCGGGCTGTGCTGAGAAT
TGCCATGGGTCGTCGGGGTCGTGCCACGGCAGTGGAGGCCACTGCTGTGTCAGGAGGCCATCGTACTTCCATGGATCCTCTGGAGGATGCCACGGCCATGGGCGGTC
GGTCTGTTCTGAGCGGTCATGCCACGGTTCTGGAGCTTCATGCCACGGTTCTGGATCTTCATGCCACGGTTCTGGATCTTCATGCCACAATACCTCTGGAGCATGCC
ACAGTACACCAATTTATGTGAAGC

>Eo_GAPDH_partial
GACAACTATGGCATAGTGGAAGGTCTCATGACCACTGTCCATGCCATCACAGCCACACAGAAGACTGTGGATGGCCCCTCTGGAAAGCTGTGGCGTGATGGCAGAGG
TGCTGCC

>Eo_LOR_partial
AAGGGTTCCCTAGCCTGCCGGAAGATGTGTTCGCATCAAGAGAAACAGGACTGCTATGAGATCCCAGCTCAGGCTGGAGGATGCCACGCCAGTGGTGGAGGATCCTC
GGGCAGTGGCAGCGGAGCTTTGCTGGGCCAGCCCATTCTGGGCTCGTCGAGCTACGGTGTTGGAGGGGGCTCGTCCTACTGCGGCTCCGGGGAGTC

Supplementary Figure S19. Continued on next pages.

B

>Eo_Beta-A1_partial
MSCSSLCYPECGVAXPSPV

>Eo_Beta-A4_partial
MSCSSLCYPECGVARPSPV

>Eo_Beta-B4_partial
MFSDEEFSYKSQQRHIPQKGQNPCLPQKETKPCPPQKVPCPPKRPPCPKYPPCIPYPRPCPPQYPSICPNHILGSGTNRVSQNAVTRLQSSLHHWLS

>Eo_Beta-B19_partial
MSSSKALCYPRPPCYPDICPDPYVDACNEPCVTSCGDSSAVVYAPPVLVRFPGPILATCPQDS

>Eo_Beta-B32_partial
MXYCPPQDCYPDICPRPCIDVRNEPCISSCGDSTAV

>Eo_Beta-O17_partial
MTFSSLCYPECGVARPCPVTXTCNEPCVRQCQDSEVVIRPSPVV

>Eo_EDAA8_partial
MTFHHQKLSHHWGCDPCWKGGWGGYGGHYGCYQPWGYSRPYGWGWGHNSGSCYSYPYRWGGGYGYG

>Eo_EDAA19_partial
MTFDENFSDELYYKPYHYGGWGGRGYGYCKPWCYQRPYKCCWG

>Eo_EDbeta1_partial
MSCGANLCIDGGSACGVARPRPCADSCNQPCVTQCPDSRVV

>Eo_EDKM_partial
SNLIKAIADMINSYQSNSRKGRESERFRRCEFKKLVQQEPSPAKRSSSNKYKHTTSLPDSDAELMNKKELITANPCVY

>Eo_EDP3_partial
MSSDQQQCKQTCXPPPXCQEKCPPPCKEPCKTXKCQEKCPPPCKEPCPPKCPPPQ

>Eo_EDPCV_partial
MAYQQQCKQTCXPPPCCVTK

>Eo_EDPE_gDNA_partial
PEPIPCCPXKPPCKEPPVPIXTPCPEPIPCPXEKQQCKXPPVPVPLPHPEPIXCSPEKPPCKEPPFPXPTPCPEPIPYCPEKPPCKEPKVPXPLPHPEPIPCXXEKX
XCKXPPVPVPLPRPEPIXCSPXKPPCKEPPVPVPXPHPEPIPYCPXKPPCKXPPXXVXXPXPEPIPCXPEKXPCKEPPXPVPXPHPEPIPCPQQKQQSKLPPVPVPT
PCPEPIPCSPEKPPCKEPPAPVTPPCXEP

>Eo_EDQM1_partial
MCSRQEKDHCHKQDTCHGSGGGSSCHGSGGGSSCHGSGGGSSCHGSGGGSSCHGSGGGSSCHGSGGGSSCHGKPQKPCQQEQQQQ

>Eo_EDQM7_partial
MCSRQEKDQCHKQDGSGGCHSSGSSCHSSGGSSCHSGGSSCHGSGGSSCHSSGGSSCHGGGSSCHSGGSCHGKPQQHCQQQQQQ

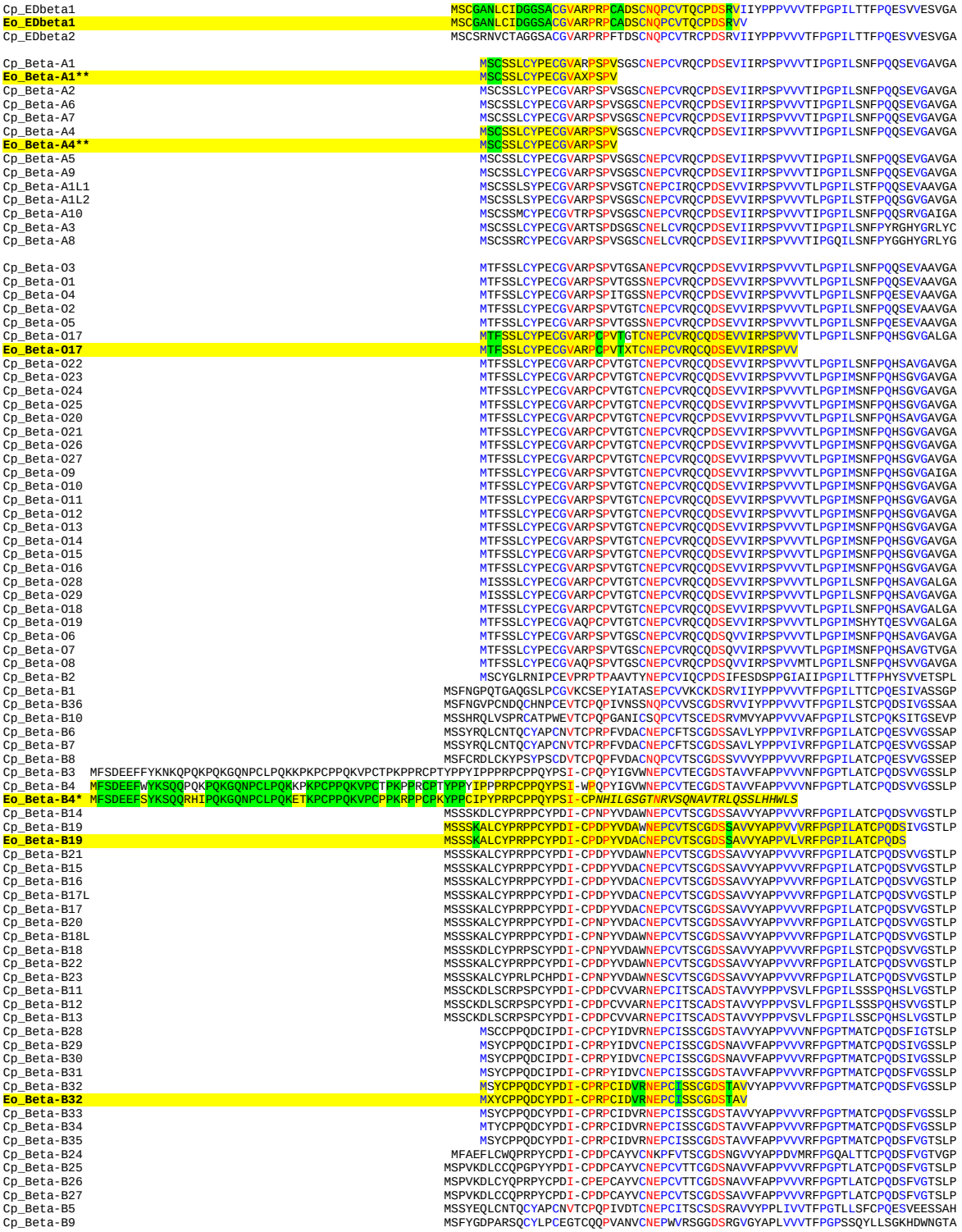
>Eo_EDWM_partial
MIYSSGRESYFNLNSTWYDPAGSWLDTRRTPFTYAYSTCCSSGGCPRGGHDNRCYEYRRSGCAENCHGSSGSCHGSGGHCCVRRPSYFHGSSGGCHGHGRSVCSERS
CHGSGASCHGSGSSCHGSGSSCHNTSGACHSTPIYVK

>Eo_GAPDH_partial
LAKVINDNYGIVEGLMTTVHAITATQKTVDGPSGKLWRDGRGAAQNIIPA

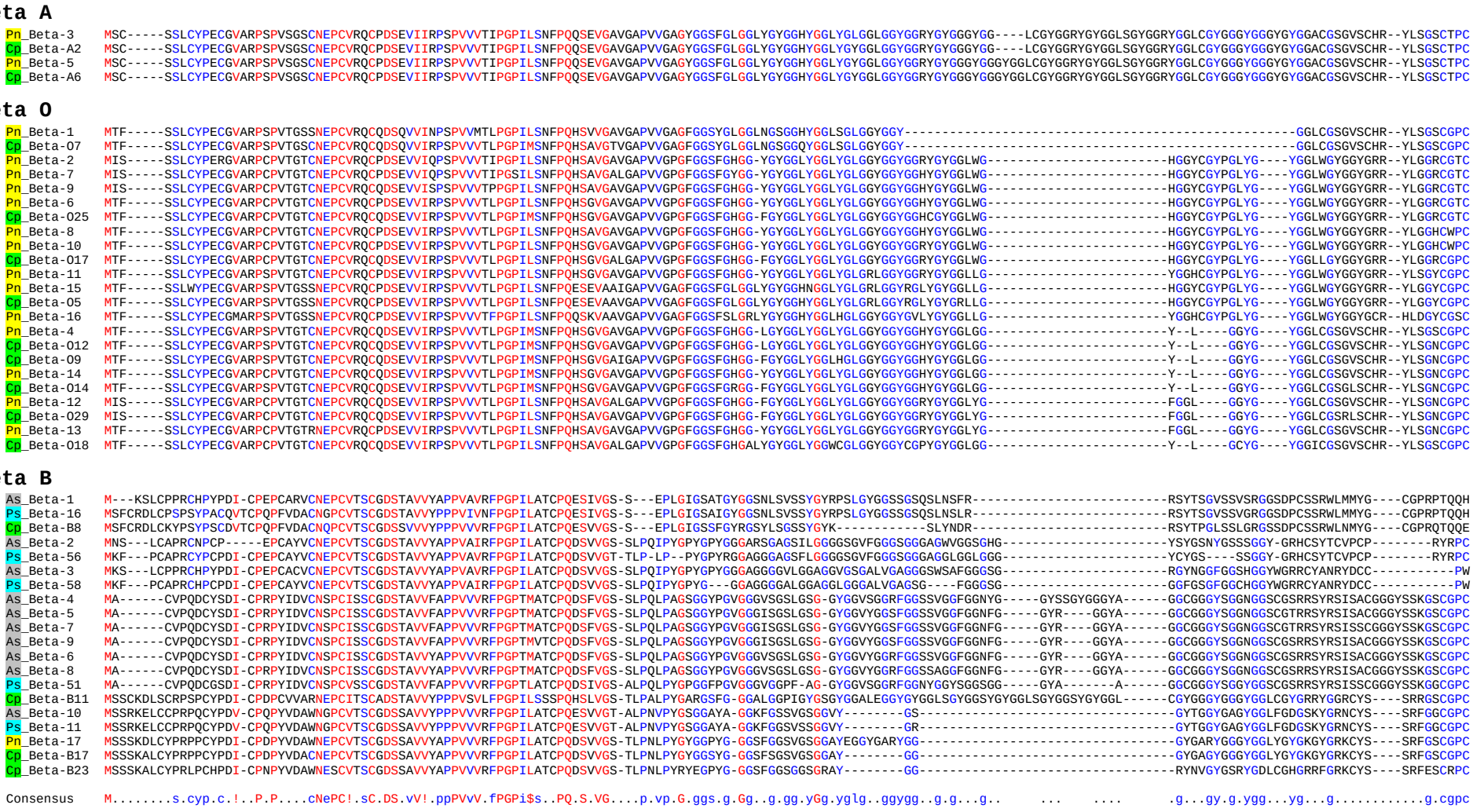
>Eo_LOR_partial
MCSHQEKQDCYEIPAQAGGCHASGGGSSGSGSGALLGQPILGSSSYGVGGGSSYCGSGE

Supplementary Figure S19. Continued on next page.

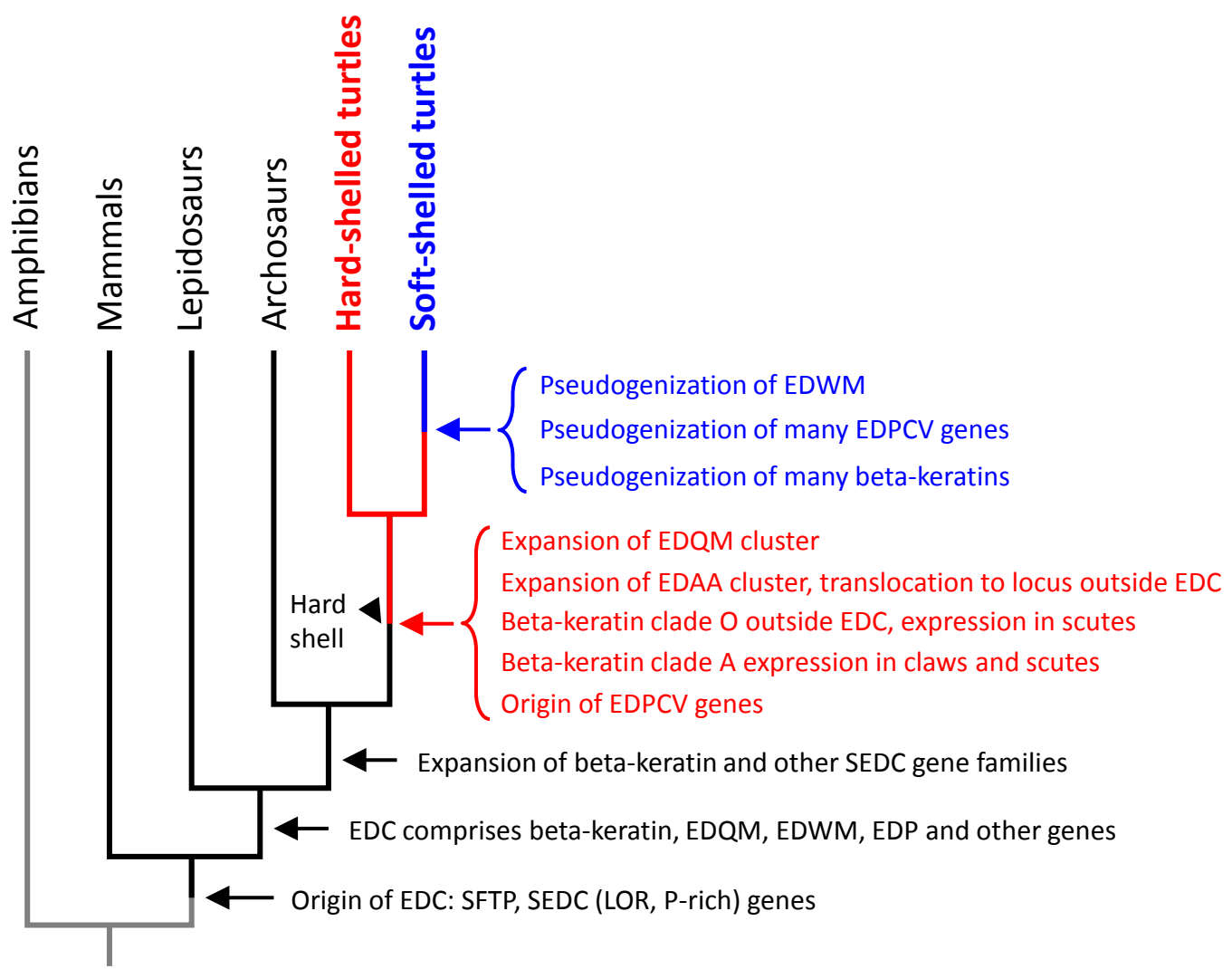
C



Supplementary Figure S19. EDC genes identified by sequencing cDNAs of *E. orbicularis*. (A) Nucleotide sequences of cDNAs from EDC-related genes of *E. orbicularis* (Eo). The cDNAs were amplified from various embryonic tissues (day 45) of *E. orbicularis* (see main text) and sequenced. In addition, the partial sequence of EDPE, amplified from genomic DNA (gDNA) is included. (B) Amino acid sequences of EDC-proteins of *E. orbicularis* (Eo). Amino acid residues are highlighted as in Suppl. Fig. S1. (C) Alignment of partial (N-terminal) amino acid sequences of beta-keratins from *E. orbicularis* (Eo) and *C. picta* (Cp). Eo sequences (highlighted by yellow shading) are inserted next to the most similar Cp sequence in which identical amino acid residues are also highlighted. Conserved residues characteristic for individual beta-keratins or clusters of beta-keratins in *C. picta* are highlighted by green shading. *The cDNA of Eo_Beta-B4 contained a frame shift (italics indicate residues downstream of the frameshift). **, annotation based on nucleotide sequence alignment (not shown).



Supplementary Figure S20. Alignment of amino acid sequences of turtle beta-keratins described in the present study and those investigated in previous gene expression studies. Amino acid sequences of beta-keratins from *Pseudemys nelsoni* (Ps, yellow) (Dalla Valle et al., 2009) and *Apalone spinifera* (As, grey) (Dalla Valle et al., 2013) were aligned with sequences of beta-keratins from *C. picta* (Cp, green) and *P. sinensis* (Ps, blue) (Supplementary Figures S1 and S2). The alignment was made with the Multalin algorithm. Red fonts, >90% conserved; blue fonts, >50% conserved.



Supplementary Figure S21. Scenario for changes of the EDC during the evolution of soft-shelled turtles. The presence and absence of EDC genes in the various clades of terrestrial vertebrates (Strasser et al. 2014; this study) and application of the principle of parsimony was used to infer features of the EDC as well as gene origin and loss events during the evolution of turtles. Fossil evidence suggests that a hard shell was a basal trait in the evolution of all extant turtles (Gaffney, 1990; Li et al. 2008; Lyson et al. 2014). Abbreviations are explained in the main text. Note that the current model is built on data from a limited set of species. In-depth analyses of the EDC in further genomes, especially among turtles as well as birds and crocodilians (together indicated as Archosaurs), will allow refinements of this model.

Supplementary Data: Supplementary Tables

Comparative genomics identifies epidermal differentiation proteins associated with the evolution of the turtle shell

Karin Brigit Holthaus, Bettina Strasser, Wolfgang Sipos, Heiko A. Schmidt, Veronika Mlitz, Supawadee Sukseree, Anton Weissenbacher, Erwin Tschachler, Lorenzo Alibardi, Leopold Eckhart

Content

Supplementary Table S1
Tentative abbreviations and full names of EDC genes in *Chrysemys picta*

Species	Gene name abbreviation	Full gene name
Chrysemys picta	Crnn	Cornulin
Chrysemys picta	EDAA1	Epidermal Differentiation protein rich in Aromatic Amino acids 1
Chrysemys picta	EDAA2	Epidermal Differentiation protein rich in Aromatic Amino acids 2
Chrysemys picta	EDAA3	Epidermal Differentiation protein rich in Aromatic Amino acids 3
Chrysemys picta	EDAA4	Epidermal Differentiation protein rich in Aromatic Amino acids 4
Chrysemys picta	EDAA5	Epidermal Differentiation protein rich in Aromatic Amino acids 5
Chrysemys picta	EDAA6	Epidermal Differentiation protein rich in Aromatic Amino acids 6
Chrysemys picta	EDAA7	Epidermal Differentiation protein rich in Aromatic Amino acids 7
Chrysemys picta	EDAA8	Epidermal Differentiation protein rich in Aromatic Amino acids 8
Chrysemys picta	EDAA9	Epidermal Differentiation protein rich in Aromatic Amino acids 9
Chrysemys picta	EDAA10	Epidermal Differentiation protein rich in Aromatic Amino acids 10
Chrysemys picta	EDAA11	Epidermal Differentiation protein rich in Aromatic Amino acids 11
Chrysemys picta	EDAA12	Epidermal Differentiation protein rich in Aromatic Amino acids 12
Chrysemys picta	EDAA13	Epidermal Differentiation protein rich in Aromatic Amino acids 13
Chrysemys picta	EDAA14	Epidermal Differentiation protein rich in Aromatic Amino acids 14
Chrysemys picta	EDAA15	Epidermal Differentiation protein rich in Aromatic Amino acids 15
Chrysemys picta	EDAA16	Epidermal Differentiation protein rich in Aromatic Amino acids 16
Chrysemys picta	EDAA17	Epidermal Differentiation protein rich in Aromatic Amino acids 17
Chrysemys picta	EDAA18	Epidermal Differentiation protein rich in Aromatic Amino acids 18
Chrysemys picta	EDAA21	Epidermal Differentiation protein rich in Aromatic Amino acids 21
Chrysemys picta	EDAA22	Epidermal Differentiation protein rich in Aromatic Amino acids 22
Chrysemys picta	EDbeta1	Epidermal Differentiation protein beta (beta-keratin) 1
Chrysemys picta	EDbeta2	Epidermal Differentiation protein beta (beta-keratin) 2
Chrysemys picta	EDKM	Epidermal Differentiation protein containing a KKLIQQ Motif
Chrysemys picta	EDP1	Epidermal Differentiation protein rich in Proline 1
Chrysemys picta	EDP2	Epidermal Differentiation protein rich in Proline 2
Chrysemys picta	EDP3	Epidermal Differentiation protein rich in Proline 3
Chrysemys picta	EDPCV1	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 1
Chrysemys picta	EDPCV2	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 2
Chrysemys picta	EDPCV3	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 3
Chrysemys picta	EDPCV4	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 4
Chrysemys picta	EDPCV5	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 5
Chrysemys picta	EDPCV6	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 6
Chrysemys picta	EDPCV7	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 7
Chrysemys picta	EDPCV8	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 8
Chrysemys picta	EDPCV9	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 9
Chrysemys picta	EDPCV10	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 10
Chrysemys picta	EDPCV11	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 11
Chrysemys picta	EDPCV12	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 12
Chrysemys picta	EDPCV13	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 13
Chrysemys picta	EDPCV14	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 14
Chrysemys picta	EDPCV15	Epidermal Differentiation protein rich in Proline, Cysteine and Valine 15
Chrysemys picta	EDPE	Epidermal Differentiation protein rich in Proline and glutamic acid (E)
Chrysemys picta	EDPL1	Epidermal Differentiation Proline-rich protein, close to Loricrin, 1
Chrysemys picta	EDPQ1	Epidermal Differentiation protein rich in Proline and glutamine (Q) 1
Chrysemys picta	EDPQ2	Epidermal Differentiation protein rich in Proline and glutamine (Q) 2
Chrysemys picta	EDQL	Epidermal Differentiation protein rich in glutamine (Q), close to Loricrin
Chrysemys picta	EDQM1	Epidermal Differentiation protein containing a glutamine (Q) Motif 1
Chrysemys picta	EDQM2	Epidermal Differentiation protein containing a glutamine (Q) Motif 2
Chrysemys picta	EDQM3	Epidermal Differentiation protein containing a glutamine (Q) Motif 3
Chrysemys picta	EDQM4	Epidermal Differentiation protein containing a glutamine (Q) Motif 4
Chrysemys picta	EDQM5	Epidermal Differentiation protein containing a glutamine (Q) Motif 5
Chrysemys picta	EDQM6	Epidermal Differentiation protein containing a glutamine (Q) Motif 6
Chrysemys picta	EDQM7	Epidermal Differentiation protein containing a glutamine (Q) Motif 7
Chrysemys picta	EDQM8	Epidermal Differentiation protein containing a glutamine (Q) Motif 8
Chrysemys picta	EDWM	Epidermal Differentiation protein containing WYDP Motif
Chrysemys picta	EDYM1	Epidermal Differentiation protein containing Y Motif 1
Chrysemys picta	Lor	Loricrin
Chrysemys picta	Pglyrp3	Peptidoglycan recognition protein 3
Chrysemys picta	Scfn	Scaffoldin

NOTE - Genes of the main beta-keratin gene cluster and S100A genes are not included here.

Suppl. Table S2A

Western painted turtle *Chrysemys picta bellii* EDC and related genes

Gene	Accession nr.	CDS start	CDS end	Expression confirmed by RNA-seq data
S100A1	NW_007281429.1	14272	17371	yes
S100A13	NW_007281429.8	34709	36274	yes
S100A14	NW_007281429.7	38691	41083	yes
S100A4-like	NW_007281429.6	76480	78681	no
S100A2-like 1	NW_007281429.5	83904	84560	no
S100A2-like 2	NW_007281429.5	92858	95973	no
S100A4	NW_007281429.4	99204	100383	yes
S100A5	NW_007281429.3	105793	106688	yes
S100A6	NW_007281429.2	108926	110289	yes
S100A7	NW_007281429.1	130785	131826	yes
S100A12	NW_007281429.0	919813	921972	yes
PGLYRP3	NW_007281429.1	945790	948573	yes
EDKM	NW_007281429.1	956038	956573	yes
EDQM1	NW_007281429.1	966063	965767	no
EDQM2	NW_007281429.1	977351	977064	no
EDQM3	NW_007281429.1	980759	981055	no
EDQM4	NW_007281429.1	985624	985325	no
EDQM5	NW_007281429.1	989554	989850	no
EDQM6	NW_007281429.1	994595	994299	no
EDQM7	NW_007281429.1	997887	998183	no
EDQM8	NW_007281429.1	1002794	1002498	no
EDWM	NW_007281429.1	1011168	1010626	no
EDQL	NW_007281429.1	1018639	1018421	no
LOR-partial	NW_007281429.1	1025940	1024173	no
EDPL1	NW_007281429.1	1049594	1049397	no
EDYM1	NW_007281429.1	1058038	1057526	no
EDAA1	NW_007281429.1	1089591	1089310	no
EDAA2	NW_007281429.1	1096092	1096304	no
EDAA3	NW_007281429.1	1101495	1101277	no
EDAA4	NW_007281429.1	1175917	1176147	no
EDAA5	NW_007281429.1	1183381	1183160	no
EDAA6	NW_007281429.1	1188493	1188705	no
EDAA7	NW_007281429.1	1194163	1193942	no
EDAA8	NW_007281429.1	1203066	1202806	no
EDAA9	NW_007281429.1	1209101	1209331	no
EDAA10	NW_007281429.1	1302661	1302413	no
EDAA11	NW_007307341.1	608	378	n.a.
EDAA12	NW_007306063.1	608	378	n.a.
EDAA13	NW_007295621.1	671	901	n.a.
EDAA14	NW_007286563.1	4325	4095	n.a.
EDAA15	NW_007284381.1	11696	11926	n.a.
EDAA16	NW_007284381.1	5762	5532	n.a.
EDAA17	NW_007284381.1	1172	1393	n.a.
EDAA18	NW_007286375.1	5287	5508	n.a.
EDAA19	NW_007282178.1	8553	8332	no
EDAA20	NW_007282178.1	17818	17597	no
EDAA21	NW_007282178.1	24458	24679	no
EDAA22	NW_007284676.1	2449	2670	n.a.
EDP1	NW_007281429.1	1745592	1745020	no
EDP2	NW_007281429.1	1758947	1758441	no
EDPE	NW_007281429.1	1772376	1773518	no
EDP3	NW_007281429.1	1778600	1778418	yes
EDPQ1	NW_007281429.1	1784025	1784378	yes
EDPQ2	NW_007281429.1	1791904	1792283	yes
EDPCV1	NW_007281429.1	1796867	1796439	no
EDPCV2	NW_007281429.1	1803508	1803780	no
EDPCV3	NW_007281429.1	1813639	1814222	no
EDPCV4	NW_007281429.1	1820502	1820137	no
EDPCV5	NW_007281429.1	1825113	1825448	no
EDPCV6	NW_007281429.1	1832160	1831816	no
EDPCV7	NW_007281429.1	1836997	1837419	yes
EDPCV8-partial	NW_007281429.1	1854445	<1854179	yes
EDPCV9	NW_007281429.1	1858298	1858651	no
EDPCV10	NW_007281429.1	1865591	1865271	yes
EDPCV11	NW_007328050.1	366	704	n.a.
EDPCV12	NW_007301624.1	43	387	n.a.
EDPCV13	NW_007284487.1	1526	1984	n.a.
EDPCV14	NW_007283637.1	1802	2254	n.a.
EDPCV15	NW_007283637.1	11177	11569	n.a.
CRNN	NW_007281429.1	1889578	1892139	no
SCFN-partial	NW_007281429.1	1905413	1919585	yes
S100A11	NW_007281429.1	1926091	1930102	yes

Notes - CDS, coding sequence; n.d., not determined; n.a. not applicable.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Contigs of EDAA11 and EDAA12 have 98% identical nucleotide sequences and identical beginning.

Beta-keratin genes are not included here.

Suppl. Table S2B

Western painted turtle *Chrysemys picta bellii* beta-keratin genes

Gene	Accession nr.	CDS start	CDS end
EDbeta1	NW_007281429.1	1068086	1067787
EDbeta2	NW_007281429.1	1073627	1073328
Beta-A1	NW_007281429.1	1221404	1221946
Beta-A2	NW_007281429.1	1227812	1227300
Beta-A3	NW_007281429.1	1232913	1232557
Beta-A4	NW_007281429.1	1238857	1239357
Beta-A5	NW_007281429.1	1244817	1244311
Beta-A6	NW_007281429.1	1255598	1255074
Beta-A7	NW_007281429.1	1262738	1263268
Beta-A8	NW_007281429.1	1266045	1265725
Beta-A9	NW_007281429.1	1272115	1272609
Beta-A10	NW_007281429.1	1277733	1277368
Beta-B1	NW_007281429.1	1311841	1311398
Beta-B2	NW_007281429.1	1323678	1323409
Beta-B3	NW_007281429.1	1358869	1359462
Beta-B4	NW_007281429.1	1363692	1363099
Beta-B5	NW_007281429.1	1392765	1392532
Beta-B6	NW_007281429.1	1400809	1400411
Beta-B7	NW_007281429.1	1414967	1414569
Beta-B8	NW_007281429.1	1426768	1426361
Beta-B9	NW_007281429.1	1431942	1432160
Beta-B10	NW_007281429.1	1452827	1452231
Beta-B11	NW_007281429.1	1463460	1462954
Beta-B12	NW_007281429.1	1473077	1473625
Beta-B13	NW_007281429.1	1479472	1479942
Beta-B14	NW_007281429.1	1493090	1493488
Beta-B15	NW_007281429.1	1504652	1505050
Beta-B16	NW_007281429.1	1516699	1517097
Beta-B17	NW_007281429.1	1532013	1532411
Beta-B18	NW_007281429.1	1536315	1535917
Beta-B19	NW_007281429.1	1540929	1541327
Beta-B20	NW_007281429.1	1547570	1547172
Beta-B21	NW_007281429.1	1551278	1551676
Beta-B22	NW_007281429.1	1555985	1555575
Beta-B23	NW_007281429.1	1563465	1563863
Beta-B24	NW_007281429.1	1568853	1568296
Beta-B25	NW_007281429.1	1584967	1584521
Beta-B26	NW_007281429.1	1592788	1593246
Beta-B27	NW_007281429.1	1630944	1630501
Beta-B28	NW_007281429.1	1641756	1641259
Beta-B29	NW_007281429.1	1646851	1646348
Beta-B30	NW_007281429.1	1655316	1655819
Beta-B31	NW_007281429.1	1661046	1660543
Beta-B32	NW_007281429.1	1673002	1672484
Beta-B33	NW_007281429.1	1685593	1685051
Beta-B34	NW_007281429.1	1698616	1698074
Beta-B35	NW_007281429.1	1708560	1707973
Beta-B36	NW_007281429.1	1730188	1729709
Beta-A1L1	NW_007295316.1	298	858
Beta-A1L2	NW_007283747.1	10082	10642
Beta-B17L	NW_007357687.1	90	488
Beta-B18L	NW_007351657.1	996	1394
Beta-O1	NW_007281530.1	1866927	1866487
Beta-O2	NW_007281530.1	1878475	1878915
Beta-O3	NW_007284315.1	7904	8452
Beta-O4	NW_007306927.1	879	439
Beta-O5	NW_007307261.1	919	479
Beta-O6	NW_007284788.1	1495	1127
Beta-O7	NW_007281980.1	49619	49987
Beta-O8	NW_007284421.1	6417	6785
Beta-O9	NW_007299123.1	79	498
Beta-O10	NW_007337571.1	2687	2268
Beta-O11	NW_007282030.1	33208	33627
Beta-O12	NW_007282723.1	2923	3342
Beta-O13	NW_007285688.1	649	1068
Beta-O14	NW_007282723.1	44024	44443
Beta-O15	NW_007284434.1	3624	4061
Beta-O16	NW_007283772.1	13629	13192
Beta-O17	NW_007282173.1	68401	68838
Beta-O18	NW_007282173.1	8154	7732
Beta-O19	NW_007282030.1	80425	80865
Beta-O20	NW_007286070.1	2133	2570
Beta-O21	NW_007285023.1	9662	9225
Beta-O22	NW_007284070.1	9973	9536
Beta-O23	NW_007282723.1	20021	20458
Beta-O24	NW_007283734.1	7236	7673
Beta-O25	NW_007285589.1	4598	4161
Beta-O26	NW_007284118.1	2918	3325
Beta-O27	NW_007282340.1	53526	53119
Beta-O28	NW_007305077.1	418	843
Beta-O29	NW_007286872.1	295	720

Note - CDS, coding sequence.

Suppl. Table S2C

Western painted turtle *Chrysemys picta bellii* partial beta-keratin genes

Gene	Accession nr.	CDS start	CDS end
Beta-p1	NW_007281429.1	1288079	>1288360
Beta-p2	NW_007281429.1	1297097	>1297471
Beta-p3	NW_007281429.1	1610400	>1610771
Beta-p4	NW_007309963.1	717	>998
Beta-p5	NW_007281530.1	1835247	>1834948
Beta-p6	NW_007281530.1	1877180	1877467
Beta-p7	NW_007284266.1	6219	>6746
Beta-p8	NW_007312745.1	602	>928
Beta-p9	NW_007303124.1	779	>1243
Beta-p10	NW_007284082.1	5517	>6053
Beta-p11	NW_007305201.1	<3	521
Beta-p12	NW_007282030.1	14653	>14895
Beta-p13	NW_007282340.1	19970	<19728
Beta-p14	NW_007282340.1	63694	<63356
Beta-p15	NW_007301237.1	>1339	935
Beta-p16	NW_007297142.1	<1	375
Beta-p17	NW_007299554.1	<1	255

Notes - CDS, coding sequence.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S3

Green sea turtle *Chelonia mydas* EDC related genes

Gene	Accession nr.	CDS start	CDS end
S100A12	NW_006666501.1	68848	66632
PGLYRP3	NW_006666501.1	49058	46106
EDKM	NW_006666501.1	39190	39044
EDQM1	NW_006666501.1	26343	26564
EDQM2	NW_006666501.1	12945	13202
EDQM3	NW_006666501.1	7329	7072
EDQM4-partial	NW_006581571.1	91560	91589
EDWM	NW_006581571.1	87240	87761
EDQL-partial	NW_006581571.1	72904	73211
LOR-partial	NW_006581571.1	65336	66987
EDPL1mut	NW_006581571.1	40144	40500
EDYM1	NW_006581571.1	30097	30603
EDbeta1	NW_006581571.1	20666	20965
EDbeta2	NW_006581571.1	15641	15940
EDAA1	NW_006618844.1	12655	12425
EDAA2-partial	NW_006605119.1	6741	>6935
EDAA3	NW_006605119.1	11428	11207
EDAA4	NW_006605119.1	20822	20583
EDAA10-like	NW_006579794.1	125068	125316
EDAAO1	NW_006654175.1	672005	672229
EDAAO2-partial	NW_006706580.1	207	>320
Beta-A1-like	NW_006605119.1	40249	40734
Beta-B36-like	NW_006589899.1	22435	21983
EDP1	NW_006589899.1	36953	36543
EDP2	NW_006589899.1	50928	50416
EDP3	NW_006620999.1	6565	6756
EDP3L-partial	NW_006593002.1	<2	184
EDPCV1	NW_006612892.1	2974	2603
EDPCV2	NW_006612892.1	8093	8410
EDPCV3	NW_006612892.1	16189	15863
EDPCV4-fused	NW_006612892.1	24874	24455
EDPCV5	NW_006612892.1	28308	28580
EDPCV6	NW_006612892.1	36193	36510
EDPCV7-partial	NW_006640425.1	475	>624
EDPCV8	NW_006579140.1	724	365
EDPCV9	NW_006647290.1	3758490	3758747
CRNN	NW_006647290.1	3732282	3729834
SCFN	NW_006647290.1	3714784	3705629
S100A11-partial	NW_006647290.1	>3685053	3684187

Notes - CDS, coding sequence.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Only the first and the last gene of the main beta-keratin cluster are indicated.

Suppl. Table S4A
EDC genes and related genes of *Pelodiscus sinensis*

Gene	Accession nr.	CDS start	CDS end	Expression confirmed by RNA-seq data
S100A12-partial	NW_005853395.1	40971	<40149	yes
PGLYRP3	NW_005853395.1	8585	4835	no
EDKM-partial	NW_005856649.1	1542	>2160	yes
EDQM1	NW_005856649.1	>14845	14759	yes
EDQM2	NW_005856649.1	29466	29140	yes
EDQL	NW_005856649.1	38597	38379	yes
Lor-partial	NW_005856649.1	49344	<49066	yes
EDPL1	NW_005856649.1	75700	75299	yes
EDYM1	NW_005856649.1	87919	87407	yes
EDAA1-partial	NW_005856649.1	<124860	124940	yes
EDAA2-partial	NW_005855424.1	<4987	5067	yes
EDAA3-partial	NW_005855424.1	25646	25443	yes
EDAA4-partial	NW_005855424.1	32648	32852	yes
EDAA5	NW_005855424.1	43933	44142	yes
EDAA6	NW_005855424.1	50778	50557	yes
EDAA7	NW_005857151.1	897	1121	yes
EDAA8	NW_005853100.1	1125816	1125706	no
EDP1A-partial	NW_005854020.1	39149	38495	yes
EDP1B	NW_005854020.1	43373	42675	yes
EDP2	NW_005854020.1	60896	60405	yes
EDP4	NW_005854020.1	87726	86956	yes
EDP5	NW_005854020.1	99701	99126	yes
EDP6	NW_005854020.1	104798	106018	yes
EDPCV1	NW_005856448.1	38543	38836	yes
EDPCV2	NW_005856448.1	10144	10548	yes
EDPCV3	NW_005856448.1	2730	3032	no
EDPCV4	NW_005854374.1	35445	35128	yes
SCFN exon 2	NW_005854801.1	1094	957	yes
SCFN exon 3	NW_005852012.1	115937	112076	yes
S100A11-partial	NW_005852012.1	>91333	91160	yes

Notes - n.a., not applicable; n.d., not determined. For beta-keratin genes, see Suppl. Table S4B.
The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.
SCFN exon 3 was partially re-sequenced and found to lack premature stop codons.
The amino acid sequence of Ps_SCFN (Fig. S2B) differs from the translation of NW_005852012.1.

Suppl. Table S4B
Beta-keratin genes of *Pelodiscus sinensis*

Gene	Accession nr.	CDS start	CDS end
EDbeta1	NW_005856649.1	98695	98399
beta1	NW_005855424.1	63829	64257
beta2	NW_005855424.1	69435	69043
beta3	NW_005855424.1	77992	78390
beta4	NW_005855424.1	85432	85040
beta5	NW_005855424.1	96312	96704
beta6	NW_005855424.1	103633	103241
beta7	NW_005859062.1	6208	5948
beta8	NW_005859062.1	12458	12084
beta9	NW_005859062.1	21403	21789
beta10	NW_005859062.1	28806	28432
beta11	NW_005859062.1	32967	33365
beta12	NW_005859062.1	38336	37962
beta13	NW_005859062.1	44144	44542
beta14	NW_005859062.1	50063	50461
beta15	NW_005859062.1	55226	54819
beta16	NW_005859062.1	98548	98985
beta17	NW_005859062.1	129236	128580
beta18	NW_005859062.1	134400	134996
beta19	NW_005856726.1	57244	56828
beta20	NW_005856726.1	70492	70223
beta21	NW_005856726.1	75358	75089
beta22	NW_005856726.1	109400	108828
beta23	NW_005856726.1	119265	118549
beta24	NW_005856726.1	131020	130457
beta25	NW_005856726.1	153300	151990
beta26	NW_005856726.1	164107	163331
beta27	NW_005856726.1	171808	172395
beta28	NW_005857404.1	2092	2496
beta29	NW_005857404.1	5648	6052
beta30	NW_005857404.1	12458	12066
beta31	NW_005857404.1	18761	19162
beta32	NW_005857404.1	29821	29429
beta33	NW_005857404.1	40813	40298
beta34	NW_005857404.1	48692	49084
beta35	NW_005857404.1	59580	59176
beta36	NW_005857404.1	66269	66673
beta37	NW_005857404.1	76424	76816
beta38	NW_005857404.1	82108	81620
beta39	NW_005857404.1	94016	93492
beta40	NW_005857404.1	99897	100355
beta41	NW_005857404.1	103517	103909
beta42	NW_005857404.1	111063	111455
beta43	NW_005857404.1	122467	123066
beta44	NW_005857404.1	127548	127952
beta45	NW_005857404.1	135528	135010
beta46	NW_005857404.1	142059	142577
beta47	NW_005857404.1	154083	153523
beta48	NW_005857404.1	160552	161061
beta49	NW_005857404.1	171148	171696
beta50	NW_005857404.1	178202	178732
beta51	NW_005851315.1	6976	7437
beta52	NW_005851315.1	14859	15125
beta53	NW_005851315.1	19991	20437
beta54	NW_005851315.1	25942	26388
beta55	NW_005851315.1	29687	30031
beta56	NW_005851315.1	48842	49258
beta57	NW_005851315.1	66396	66842
beta58	NW_005851315.1	88113	88523
beta59	NW_005851315.1	97314	97571
beta60	NW_005851515.1	864	484
beta61	NW_005851515.1	10845	11831
beta62	NW_005851515.1	22566	22147
beta63-partial	NW_005851515.1	29766	>30206
beta64	NW_005851515.1	52402	52905
beta65	NW_005851515.1	72310	72825
beta66	NW_005851515.1	83793	83239
beta67	NW_005851515.1	93560	94084
beta68	NW_005851515.1	108705	109220
beta69	NW_005853269.1	45707	45216
beta70	NW_005853269.1	60041	59550
beta71	NW_005858327.1	3703	3116
beta72-partial	NW_005858327.1	16881	>17297
beta73	NW_005856853.1	11358	10768
beta74	NW_005852318.1	1255	695
beta75-partial	NW_005870254.1	<793	488
beta76	NW_005854020.1	23969	23571

Notes - n.a., not applicable; n.d. not determined.
The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Supplementary Table S5. Primers for RT-PCR analysis of *Emys orbicularis*

Target gene(s)	Exon in which primer anneals	Orientation of primer relative to gene	Primer sequence	Scaffold containing the annealing site used for primer design (Accession number)	Annealing site start (Nucleotide number)	Annealing site end (Nucleotide number)	Notes
Beta-A1	exon 1	sense	CTTCATCCCCTCGGTGAACTG	NW_007281429.1	1220330	1220350	
Beta-A1	exon 2	anti-sense	CGGCTCGTTGCAGCTGCCAGA	NW_007281429.1	1221461	1221481	same anti-sense primer as for Beta-A4
Beta-A4	exon 1	sense	CTTCAGCTCCTCAGTCAACTG	NW_007281429.1	1237822	1237842	
Beta-A4	exon 2	anti-sense	CGGCTCGTTGCAGCTGCCAGA	NW_007281429.1	1221461	1221481	same anti-sense primer as for Beta-A1
Beta-p1	exon 1	sense	CTTCATCTCCTCAGTGAACAC	NW_007281429.1	1286991	1287011	gene not amplified in *E. orbicularis*
Beta-O17-like	exon 1	sense	CTTCATCTCCTCAGTGAACAC	NW_007282173.1	67319	67339	annealing with mismatches
Beta-p1	exon 2	anti-sense	AATTGGTCCTGGGATGGTTAC	NW_007281429.1	1288211	1288231	gene not amplified in *E. orbicularis*
Beta-O17-like	exon 2	anti-sense	AATTGGTCCTGGGATGGTTAC	NW_007282173.1	68533	68553	annealing with mismatches
Beta-B8	exon 1	sense	CTTCTTCTCCTCGGTGAACTG	NW_007281429.1	1427463	1427443	gene not amplified in *E. orbicularis*
Beta-B32	exon 1	sense	CTTCTTCTCCTCGGTGAACTG	NW_007281429.1	1673882	1673862	annealing with mismatches
Beta-B8	exon 2	anti-sense	GACAACTGGTGGTGGATAGAC	NW_007281429.1	1426642	1426622	gene not amplified in *E. orbicularis*
Beta-B32	exon 2	anti-sense	GACAACTGGTGGTGGATAGAC	NW_007281429.1	1672894	1672874	annealing with mismatches
Beta-B19	exon 1	sense	GACTTCATCCCCTTGGTGCACT	NW_007281429.1	1540259	1540280	
Beta-B19	exon 2	anti-sense	TGGTAAGGTGCTTCCCACAAT	NW_007281429.1	1541118	1541138	
Beta-B3	exon 1	sense	ACTTAATCTCCTTGGTGAACA	NW_007281429.1	1347417	1347437	
Beta-B3	exon 2	anti-sense	TCCATAGTGACATGGGTTCCA	NW_007281429.1	1359424	1359444	
EDbeta1	exon 1	sense	CGAGTTATTCTCAGTGAACTGG	NW_007281429.1	1068750	1068729	
EDbeta1	exon 2	anti-sense	ACGACAACCGGTGGTGGGTAG	NW_007281429.1	1067961	1067941	
EDAA8	exon 1	sense	ACTTCTTCTGTCTTACTCTCC	NW_007281429.1	1208021	1208041	
EDAA8	exon 2	anti-sense	GGCAAAACATGGCCAGCATCT	NW_007281429.1	1176115	1176135	
EDAA19	exon 1	sense	TCACTTACTCTCCTCGGTGAC	NW_007282178.1	9818	9798	
EDAA19	exon 2	anti-sense	GGTACCAACAACCTTTGGGAT	NW_007282178.1	8423	8403	
EDKM	exon 2	sense	CTGCACCATCACCCCGGAATG	NW_006666501.1	39208	38188	
EDKM	exon 3	anti-sense	CGACCGTGACCACTATCCAAG	NW_006666501.1	37570	37550	
EDP3	exon 1	sense	CCTTGTACCTACAGCTGAAAC	NW_007281429.1	1779587	1779567	
EDP3	exon 2	anti-sense	CTGCTTCCAGTCTTGGGACTG	NW_007281429.1	1778444	1778424	
EDPCV	exon 1	sense	TTTGTTCCTGTTGGTGACTTG	NW_007281429.1	1812055	1812075	
EDPCV	exon 2	anti-sense	CAGTGCTGCACAGGTGGGCAT	NW_007281429.1	1837374	1837394	
EDQM1	exon 1	sense	CACGAGTTCTTCTCTGCATTC	NW_007281429.1	966842	966822	
EDQM1	exon 2	anti-sense	CACCTGGCAGCAGTGCTTCTG	NW_007281429.1	965807	965787	
EDQM7	exon 1	sense	CACGTTCCTGGAGGTGAATAG	NW_007281429.1	997067	997087	
EDQM7	exon 2	anti-sense	TGGCAGGGCACCTTGCAGATC	NW_007281429.1	998150	998170	
EDWM	exon 1	sense	ACGGTCCTTGTTGGTCAATAG	NW_006581571.1	86642	86662	
EDWM	exon 2	anti-sense	ACTGTTGCACATGTTGCTTTG	NW_006581571.1	87673	87693	
LOR	exon 1	sense	ATTTGTTTCCAGTTGCTGAAC	NW_007281429.1	1028005	1027985	
LOR	exon 2	anti-sense	GCAATTATAATCTTCTGGCAG	NW_007281429.1	1025761	1025741	
GAPDH	exon 5	sense	CTTTGGCCAAGGTCATCAAT	NW_006634294.1	527040	527059	
GAPDH	exon 6	anti-sense	CAGAACATCATCCCAGCATC	NW_006634294.1	527493	527474	

Notes - The primers were designed using genome sequences of *C. picta* or *C. mydas* . The primers were used to amplify cDNAs of *E. orbicularis* .
Primer sequences are shown in 5'-3' direction. Annealing site starts and end refer to numbers on the scaffold but not to the 5´ and 3´ ends of the primers.

Suppl. Table S6
Chicken (*Gallus gallus*) beta-keratin genes within the EDC

Gene	Orientation within EDC	Accession number	CDS start	CDS end
GgEDbeta	+	NC_006112.1	825202	825525
GgBet1	-	NC_006112.2	854410	854012
GgBet2	+	NC_006112.2	855666	856082
GgBet3	-	NC_006112.2	858814	858425
GgBet4	+	NC_006112.2	860071	860460
GgBet5	+	NC_006112.2	864665	865081
GgBet6	-	NC_006112.2	867842	867435
GgBet7	+	NC_006112.2	869099	869506
GgBet8	-	NC_006112.2	872116	871700
GgBet9	+	NC_006112.2	873374	873781
GgBet10	-	NC_006112.2	876526	876110
GgBet11	+	NC_006112.2	877786	878175
GgBet12	-	NC_006112.2	880823	880434
GgBet13	+	NC_006112.2	882416	882823
GgBet14	-	NC_006112.2	885937	885545
GgBet15	-	NC_006112.2	895274	894897
GgBet16	+	NC_006112.2	906469	906762
GgBet17	+	NC_006112.2	909282	909578
GgBet18	+	NC_006112.2	912696	912992
GgBet19	+	NC_006112.2	920610	920906
GgBet20	+	NC_006112.2	921450	921746
GgBet21	+	NC_006112.2	927670	927966
GgBet22	+	NC_006112.2	929118	929414
GgBet23	+	NC_006112.2	932678	932974
GgBet24	+	NC_006112.2	936060	936356
GgBet25	+	NC_006112.2	939315	939611
GgBet26	+	NC_006112.2	944034	944330
GgBet27	+	NC_006112.2	947082	947378
GgBet28	+	NC_006112.2	950614	950910
GgBet29	+	NC_006112.2	956699	957049
GgBet30	+	NC_006112.2	960981	961331
GgBet31	+	NC_006112.2	967208	967558
GgBet32	+	NC_006112.2	973298	973642
GgBet33	-	NC_006112.2	975768	975106
GgBet34	+	NC_006112.2	979013	979657
GgBet35	-	NC_006112.2	981718	981359
GgBet36	+	NC_006112.2	982872	983237
GgBet37	-	NC_006112.2	985238	984873
GgBet38	-	NC_006112.2	989144	988779
GgBet39	+	NC_006112.2	990297	990662
GgBet40	-	NC_006112.2	993140	992775
GgBet41	+	NC_006112.2	994299	994664
GgBet42	-	NC_006112.2	997105	996740
GgBet43	+	NC_006112.2	998247	998612
GgBet44	-	NC_006112.2	1001042	1000677
GgBet45	+	NC_006112.2	1002152	1002517
GgBet46	+	NC_006112.2	1005687	1006052
GgBet47	-	NC_006112.2	1008543	1008178
GgBet48	+	NC_006112.2	1009693	1010058
GgBet49	-	NC_006112.2	1013391	1012921
GgBet50	+	NC_006112.2	1017136	1017624
GgBet51	-	NC_006112.2	1020045	1019563
GgBet52	+	NC_006112.2	1023690	1024181
GgBet53	+	NC_006112.2	1029671	1030066
GgBet54	-	NC_006112.2	1034534	1034160
GgBet55	+	NC_006112.2	1038872	1039195
GgBet56	-	NC_006112.2	1042806	1042375
GgBet57	-	NC_006112.2	1045505	1045008
GgBet58	-	NC_006112.2	1049893	1049459
GgBet59	-	NC_006112.2	1057634	1057089
GgBet60	-	NC_006112.2	1063389	1062721
GgBet61	+	NC_006112.2	1066589	1067158
GgBet62	+	NC_006112.2	1070763	1071293
GgBet63	-	NC_006112.2	1074197	1073667
GgBet64	-	NC_006112.2	1081224	1080733

Note - CDS, coding sequence.
For phylogenetic analyses, sequences of "feather" beta-keratins
encoded by genes outside of the EDC were taken from Ng et al. 2014.

Supplementary Data:

Comparative genomics of the epidermal differentiation complex suggests evolutionary adaptions of snake skin

Karin Brigit Holthaus, Veronika Mlitz, Bettina Strasser, Erwin Tschachler, Lorenzo Alibardi, Leopold Eckhart

Content

A

>**Pb_CRNN**
MSHLLRNIHSIICVFEKYAKNDGDCSTLSKGELKQLIQKEFAEVIVDPYDPKTVETVLHLLDTDCDGKVGFEEFTVLLFKVVKACYKKV
QECQVPADGPSQKHNISTQQQEIQLPQVPPCPCQKLLPPKEVLDAEELPQGQEKEKPTLEISCDQPGKEQLQQEEVSAHPRTQTSRVQS
IQKTIEEIIRGNPTCQTSQVEMRTEQGTEQAQEHQRAVPPKNRCLQEPLSPDQEPSQDVLQNVTQQSFQQHQVEERCNTRQKEETSQTL
VQERQTQQSALTMEQIPTGKVIHEPLQETHQGPFGGMNQQVQAEEKRSASGHIRNPQVVQICDQEFNSVTQGQHQIGQQRRQAQGIEQS
SVVQQDQEVQVLGQVISGPQEHQTLTPQVYQSHQEEERKPPTLGQSQYGLSEQRALPQEVIKPQVAERCPHCESPETTVAQGSQESSAR
WPFQTSRSQLRQKPLHFPPPWSTKQ

>Pb_EDCATM
MFFHKKLGWLTPTVVTSPSDDCPALSLPTCSAPCTGACVLPGLGPGSPAACTLLCEAGPNPPVSSPATGGCPACEGLSAPVLKASESPT
MLESCSPPGASTDPWREENMWEGDLPDLSVFLTVPPVPGGDSPCALGRQPDCLPPFCLSHLPLCQDPCSSLCYIPCATMTTNLPFCGVQ
CRNALSRAMGLPLARKCASAIPCTSTCPCATKCANALPCAVEIPPAPNCASTASCATKGATTIPYATKGVTTTCPCATKCVSAMPCAMA
IPAPPQCTCPAPCTTKSAISAPCTTPCPSIIPCTPLCCCKRPPSCVPQPI

>Pb_EDCG
MGCCDCCNSGRSTIIMCMPSHGYQSCYPPGYMQSYGCAPCCCMPIYCTTVRRNRSCCGCGCGCGC

>Pb_EDCM
MSCCCCCGGRGCGCCHRHRPRVVYYVGRPISVRPCCVRGYSYTSCCGHARSCCTPRYRAPYQQYYNQVKKY

>Pb_EDCP
MSFQCKQACPCPSICIKAVPVQAPCSPPSLCPCPCAPVSVKTVCPDPCPAPKGPPCPQKSCPDAKAPAAKCPPPCGTADAKSASPCQSS
DCATPCQATVCVPPCQCQCSACAPCQAQTCIPICGNPCQEQCYAYIPVSISQCSCGCSRILLVPQSSCAKSCAGSCTPVCVAPCPAPSP
PPCTCACSCGPCCS

>Pb_EDCS1
MTCSCCSSCCGSTNRIICVPGHSYYVPPSACCGSYSGSCCGSSCCGSSCCGSCCGSCCSSPPSNIVYTSPMQYCCPPMKQCPPSLQCKQ
SGYPSKKCC

>Pb_EDCS2
MTCGCGSSGSDSPTIICIPCGGQPCCSPCCCCPCYGGMQSCCCMPCCAPCGYGGCGGCCGITAYRLPSRRRSSCCGCGY

>Pb_EDCS3
MTCGGGCGSCGDSCGTCGAMLLCVPCCGCQACCSPCCSQSSCSPCSCSKCSCGQSSCSPCSCSKCSCGQSSCSPCSSSKSSCGQSSCSP
CSSSKCSCGQSSCSPCSCSKSSCGQSSCSPCSSSKSSCGQSCCSPCSCSQSSCAQSCCSPCSSSQSSCSLPCSQSCCPITCPPCCSK

>Pb_EDEPK
MSTEQQQRKQTSVLPPGLSELTTESTPEEAPQKEHDRESLSSSEQGNALPEDPKESHERKELSSEFSAIEMEPGASSEGTSGAKHGPLA
IKEEEHLPLVDQHCREPPTLPLAPSKETLKPGQPLLPVREAEGHPATQPPALGGQKQQQQQQK

>Pb_EDEPT
MAYQYKQPCPPPPACGVQCSAEADDACTLTGVAPCDPTCDASPSGACNPPDTLLYQTRQPCEATPCGPPPCPTAAPSPCALPCTSLYTV
TCAEAPSTPCGSATWQPCTPLPADAPPCPLEEACTRGPGCPETCSDPQGDMGMPYVKPGSTVCLNPCGAVSVKPSSPEEEENSIFLASY
TPHFT

>Pb_EDETM
MACPGELCSYNTSYSTSLPSSSFLLQPPSYMAETRLTGLEVLVPESTSCEIPDGCLHPCGSPAFCCHEPPVVLAFFETSLYEV

>Pb_EDGPC
MGNMESDSGGGCGCNCNCGGGSSDSHVICVPSGGGSGCCCPPCCSPCCCSPCCCSPCCCSPCCCRPCCSPCCGQSSCCAPCCKQCLK

>Pb_EDGY1
MIFYTQADSYEPYLPYYLSYCPPCYPPCYPQYGYSSGFGSYGQGDSYNYGYDSGSSSIPSIPPCSYGYRYTRRGSCYEPCAYRLGYGAR
RNFPSRRYYNPNLW

>Pb_EDGY2
MSYYDPCRSKCYSSGSSLGGYKRRIRYSTGSCYGGYGQGTRYIPSCGYGYGGYRGQTYGRICYQPLDYGYGYGYGNKYCPPQWYGGYGS
GGYDYGCRYGGLYGRRQSICYDPCYGPRYQYGSRRGSFCGPW

>Pb_EDHEM
MGCTSPNTGNEEITMMRYIRIEPLEELGLRSCPHPKHHEMMDLFCAPDLEILLEEPDLLELISIEDTSPHPEYYRRTRIYDMIELPLSA
GAGKDSGCSSHKNADPGAEHSQEDSEKATSPSSCEPKNGDTNNLPASRNPSDPATVIYSFTVEYQCSPLLKYCTLEEMGYRPIQLM

>Pb_EDKM
MSNLFRAFTEMIEGNSKVVWKKAHDSEMLKKSEFKKLIQQELSPSKTHKYKQVKHLHDSDTELMNEKELTSQLH

>Pb_EDP3
MSQQQQCKQIPCTPFTSCKCTSSCQKPCEPLACQGSAGPQCPGSTGNSSQKPTCPPGKNECFCHCHQQQQC

>Pb_EDPAM
MTHQCKLPPELPPACLKLATKTSLQTKSPPPQTVTVCAVPTPAPSPPPPQSVTCCGVPLQVPQSPPPQRVTVCAVPAPAPSPPPPQSVT
CCGVPLQVPQSPPPQRVTVCAVPAPAPSPPPPQSVTCCGVPLQVPQSPPPQRVTVCAVPAPAPSPPPPQSVTCCGVPLQVPQSPPPQRV
TVCAVPAPAPSPPPPQSVTCCGVPLQVPQSPPPQRVTVCAVPAPAPSPPPPQSVTCCGVPLQVPQSPPPQRVTVCAVPAPAPSPPPPQS
VTCCGVPLQIPKTPPPQTVTICAIPTPALAPPPAQSITCCGIPLQIPNPPQTVTICASPAPASAPAPTLVSVPCSSACTSCVPVTLSLV
QPTCVAVPQCSSCCKTCHGS

>Pb_EDPCCC1
MAHCSKCGCSPCCCKKHGHSHCRKKCRCNKCGCSPCCCKKHGHSHCHKKCRCKKCGCSPCCCHLCSCSPSCCQQCGCSPCCCQQCGCSP
SCSQQCGCSSSCCKKCGCSPCCCQQCGCSPSCCKKCGCSPCCCQQCSCSQCCCEKSCCCTPCSCCCSPCCGSKGCSAKGCGSTGCFPQQ
CEC

>Pb_EDPCCC2
MTCCSKCGCSPCCCKQSCCSPCCCNQSCCSPCCCKQSCCSPCCCKQSCCSPCCCKQSCCSPCCGQQSCCSPCCGQQSCCSQCSCGKSCC
CTPCSCCCSKGCSSKDCCSKGCSSKGCSSKGCCSKGSCSKGCCSNQCKC

>Pb_EDPCCC3_partial
MASSGTKTPKCSKCGCSPCCCPKASSSTKCSKCGSSPCCCSKASSSSKTKCSKCGCSPCCCSKSLCSKCRCSPCCCPQSRCSKCGCSPC
CCPQSQCSKCGCSPCCCPQSXXXCSPCCCAQPRCSKCGCSPCCCPQSRCSKCGCSPCRCSQSRCSKCGCSPCCCPQSRCSKCGCSPCRC
SQSRCSKCGCSPCCCPQSPCSKCGCSPCCCSQSRCSKCGCSPCRCSQSRCSKCGCSPCCCPQSPCSKCGCSPCCCSQSRCSRCGCSPCC
RSQSPCSRCRCSPCCCSQSRCSRCGRSPCCSQSCCSRCGCSPCCRSQSPCSRCRCSPCCCSQSRCSRCGRSPCCSQSCCSRCGCSPCCR
SQSPCSRCRCSPCCCSQSRCSRCGRSPCCSQSCCSRCGCSPCCRSQSPCSRCRCSPCCCSQSRCSRCGRSPCCSQSCCSRWGRSPCCCP
QSPCSKCGCSPCCCSQSQCSKCGCSPCCCPQSPCSKCGCSPCCCSQSRCSRCGCSPCRCSQSRCSKCRCSPCCSQSCCSRCGCSPCCCS
QSRCSRCRCSPCCCSQSRCSRSGSSPCCYSQSRCSRCGCSPCCCSQSRCSKCGCSPCSCCSQSRCCSKCGCFPCRCSQSRCCRCGCCPC
CCTPSCCPRCNCPYCQQNQSCCCPCCCRYSCCCPCCYCSCCNSSPLKCQPK

>Pb_EDPCCC4
MTCCPICGSSPCGCAPCYSCSPCCGCGSCGSPCCSSCGSPCCSSSCSAPCCSCSPCSGCGSCGSSCSCPQCSSSPPCCACGSCCNRQSS
GCGCGCGCSPCSSCSPCCGCGSCCSPSCSSCCGSPCCSSCGSSCCGSGCCFPKQCFS

>Pb_EDPCS
MKPLPPCGWFCFCNVFLQRLGLHDPCKKVVDPFQLHLSLFLLLPSLCTMEFEYVFPELLDETDPTFLETYRQRPWKWVDPCKPRYCPPK
YFPPCPPPCPPPCPPPKPQRPIPCCPLYPLPQCPKYFPPCCPPYPCCYPPQWCYPRPLPCVPQKPWYPCKVPCPPGWVPKYPRPYDLLE
LY

>Pb_EDPKC
MASSHNQHQCKQAPTLPSELCKSAPCPPDQQCKQPPVVVVPTPCPEPKPCPPPEPPCKEPPIVIVPFPEPKPCPPQEPLCKEPPAPTPC
PPKEPEPCPPKQPTPCSQQKSQCKEPPVVVIPTPCPEPKPCPPQEPPCKEPPVVVIPTPCPKPKPCPPQDPPCKEPPAPTPCPPKEPEP
CPPTPCSQQKSQCKQPPVVVIPTPCPEPKPCPPQEPPCKEPPVVVIPTPCPKPKPCPPQEPPCKEPPVVVIPTPCPEPKPCPPQEPPCK
QPPPCDQQQKKQPCSWPPQNK

>Pb_EDPQ1
MYTKGYGDGGGSYRNYGRSACCEPSCHSVRRRSEVKCCDTSPQSCCPPVQKCCLPRQKYCPPVQKCCPPKRKYCPPVQDYCPPVQPCCS
PVQKYCPPVEPCYPPVQKYCPPVQPCCPPVQKYCPPVQPCCPPIKKYCPPIQPCCPPAQKYCPPVQRCYPPPQRYCPPVQPCCPAPQRY
CPPVQPCCPPPQRYCPPVQQCSPPPQRYCPPVQPCCPPPQRYCPPIQPCCPPQQRYCPPVEPSCPPLEISQVQQVCRVPPHLLK

>Pb_EDPQ2
MSYHQQQCKQPCQPPPQCQKTSCPPACPPQKCVPKPCPPQECVPKPCPPQECVPKPCPPQECPPKPCPPPQQCCQNPKRC

>Pb_EDPQ3
MSYLNQQCKQPCPIPPQCKKTIPPPEQCCPPPQQYCPPPQQCCPQQQDQDPKCC

>Pb_EDPS1
MFHCCVPSCSMGPTAPLCFQPCQQSRSMPTISIPCRPVHPSMPPTSIPINLCPVPSMPPPSSSPSICIPLCPSPSPSSQRQPSIPLCLS
PQPSGSSSGSNSRISISLSPSPSSSSQSQLSIPLCLCPQPSSSSSGSSSSPSISISFSPNQSSSSQPLYCIPLSPSQPPPSISLSPSQS
QSSSSHSICIPFCPSPLPSSSSQPSFCIPLRPSPQTSSSSSSSSPRISISLSPSPSPSGQSQSCIPLCFSPQPTGSGSGSGSSPRISIS
LSPSPFSSSQSQPSYCIPLCPSQPPPSMSPRPAPPPTCCIPIIIPSNPCFCGC

>Pb_EDPS2
XXXSSSGQSQPSIPLCFSPQPSGSGSGSGSGSSPSISISLGPSHSSSGQSQPSIPLCFSPQPSGSGSGSGSSPSISISLGPSSSS
SSQSQPFYCIPLCPTQPSPSMSPSSSSQPTCCIPISIPSNPCFCVC

>Pb_EDPS3
MSSCCVPSCRVGPIAPLRVQPRDQSSSLPSCISCRPVQPSMPPTSIPISLSPVPSMSQPGPSPLSSCIPLCPMQSPPSMSLSPSPPPPT
SSQPSYCIPLCPSQPPPSMSPTPSPPPSCCIPITIPSTPCFIVF

>Pb_EDPSQ
MDCCSQQCKQPCLPPPICQQKIGIKRTLWQQRGSSQCTKPCSPSWPQSYDPECPPAYPPQGRAVCWAQPVQQCPSWPRQPCSPPCKSVS
FNSKGLRAPHGMQQSASKGKRQCLLQCSQDGPHQCLKVCAVKGCKKCANKGSPQCVTKESRGGLPCAYQPHTVKGDWTWFIKGNQQCVS
QPCVTKVNPLYSSKGSQQYRRKGGYCKTQLKVSSNGQKYCSASNNWPW

>Pb_EDQL_partial
MCSHHDRGCHGVSHDHGCHSGSSCHEDHGSCHNRSSCHDGGNPSFRVMTIPVAQDKNPMPIVQQQSPVVYPAPKLQQTQQLKQPTPYPP
QQQLQQPQQQQKQKSKTQRKSCXXX

>Pb_EDQM
MCSREDKDQCYKQEKDESPCCYSHRSSGCGNKPSSRACSGSYGWGPSSPYCQGKSDQPQQQQQKKQGCQLPPQK

>Pb_EDQSG
MQQCKQPQGFPPVGQMSTAAQCKASAVSQSQGATVDQCHGSSGNQTQGSTGIQCGGSAVSQGKGSTGNQCQGSMPSQTQSSGVSQWRGS
TGDQCGGSAVSQSKGSTGNQCQGSMPNQSQSSGVSQWRGSTGDQCGVSVVSQGQESVAFQDQESMVFQGQGSPRNQCGGPSVSQDQGST
GCQCHESAGTKY

>Pb_EDSC1
MSQQIQQGSSCCCCCGSSGGCCSRGSGGGCCGSSGGVVYVQQSQQSSRGCCINRTGACCGRCCGRRCGGGCCRRGSGGGCCSRGTGGGC
CGSSGGAIYVQQSSRCCCINRTGPCCGRCCGRRCGGGCCSGGSSQQKLK

>Pb_EDSC2_partial
MSQQQQRGGCCCCRCCGGGGGGGCGCHSGGGSQVSRSSGGGCCGRSSGGSSQQSXXXSSGGGCCGRSSGGSSQQSHSSGGGCCGRSSGG
SSQQTHSSGRCCGGGGSSQQKMK

>Pb_EDSCP
MVFSCHRPTCTSPSGPIFPISGGSHSCVPVCQPTCSIPSGPTCTRPSGPIFPISEGSHSCVPVCPPPCCIPCGPPTVCVQPPPPPCVPV
CPSPCCIPCGPPTVCVQRPQPSCIPVCPSPCSNPCGPPTVCVQPRSSPCCIPSCTTPSRPVITVSPCCSVPSCSIAPLCPQPGSPSCVV
PSGPILPVSSGQPSCIPISVHPCSPCLTFF

>Pb_EDSPR1
MACPYQQCKQPCLPPPICGKGGVKQCAVCPPACPCQSVCAVPSVKVVGDPCTPICIPPCAPKCAVDNICVQSCPPSGQSICVQPQVVCP
CDSKCPDPCCKQGSC

>Pb_EDSPR2
MSQCKQACKAPPCPEPCCPTKCPPKTCCSSPGTKAVSKKGPCCPPSADVKCPPPKVKEPSKDCCC

>Pb_EDSQ
MSYQCKQPCLPPPFMKGTTVCAGPGTTVCVSPTQGQSIEVCASPTGAICVTQGQNQAQCANVCQDCCGSVVVSPVQNQGFSPCATLCTK
CNGAVCVTKAPEQSSAVCVPQQQACQCSVCQGNISVAPIQSQSGACATICASPCSSVCVTPIATPLGPAATICTSPGACATVCQDPCGS
VCVTSKGATKCATKCATKCVAPGATVCVDPCGGVKSVKSCTTKCTSPCFAVSNEQVCVKACPSISMNQCNVKKC

>Pb_EDSRWM
MNFYMPQEYWDLNSWETNYENNYDFPSEYYYAYNPRSYILDYQSRWGSPSYSSCSDCCDPCYSAICGSPRYRSRSSYCPSCAPSSNSSW
GSPDCSSNSSCCPPCWTCGSCNPRYSPICRPQCHRSRSVSPFRSNSRSSCCPPCCRCGRWAHPSSSSSQSPDGSSNSSYCPPCWTCGSC
NPHYSPICRPRCHRSRSVSPFRSNSGSSYCPPCCRCGHWAPPSSSTSQSPDGSSNSSYYPPCWTCGSCNPHYSPICRPRRHRSRSVSPF
RSNSRSSCCPPCCRCGRWAHPSSSSSQSPDGSSNSSCYPPCCRCGGWAPPSSSCSQSPDGSSNSSYCPPCWTCGSCNPRYSPICRPRCH
RSRSVSPFRSNSGSSCYPPCCRCGGWAPPSSSCSQSPDGSSNSSYCPPCWTCGSCNPRYSPICRPRCHRSRSVSPFRSNSGSSYYPPCC
RCGGWAPPSSSCSQSPDGSSNSSYCPPCWTCGSCNPRYSPICRPRCHRSRSVSPFRSNSGSSCCPPCCRCGRWAPPSDSCSQSPDCSCH
SSCCPPCCTCGSCNPCYCPICGPQPSNFRSSSRRCPCCCRCGRCAPPCCCSSQCCRSSCSCCRPCWTCSSCNPCYCPSCRPQPPNSSFS
SCCCPPCCPCGCYGPCYDPSC

>Pb_EDWM
MSEERIYSSGREPYFNLNSTWYDPAGSWLDNRRKPFCYVENTACVTCCNPRTNVPRRGGHNYRCYCYRRSTCRPGGNPRVRCCVHNPSG
GPRDYWGRPIGDACNGCTGGYYSHGDCGSCCGSSGECGTGGGLVACAQPCVSSGGVCAEPGCRPAGRGVCAEPCVMSSGGWGSGRGVCA
EPECSASGCGSRRRGVCSEPCARASGGC

>Pb_EDY1
MIGITILSSHPSEGTKAIYCPPTIYKPSGSDSSAYPPYTSLAGTAARCYIPRKSRYLPGPTYYLSGWDCSRPMPHTYRSTYGPCSYEYL
LDNSQGYYDPCVYKYLSKGVQRSYDPYGLEYLYSYRRRYYDPSDYRYYPSYGGRCDPCVYNNPPSYGRSSYSSRGYTYRCGFSNDPCGY
SCAPGYRQPRYADRRRYSSRPISSDPYGSCPESPYASECRRPSSDSSSYCGPC

>Pb_EDYM2
MPYYAYHCKPIGPPLPLFVKKRLPRYGSQYFPLYGSRYVSSRSIPFHRRGIWMTPYSLASVAGYQPWIIPRVSEACSSKSPKLRTTRNP
VRPYVDSYQYPLQTTVPYDGMTKGPHGPINKGSLPCVIPERVLRVTKGPVLYSSKDLHPSSAQIPLRRACKGPVFYASQDPLLCSSKEE
LPRVSKGPDQYPLKGAQRSRAADSYLPMTEDPPQSGAKVTQPRATKGPRLIATKVPRFSALTVSRSFLKRGVHSFLTKTSQTNLCKGSR
SSLSKTSWLSLGKGSQPSLVRTATTPAPKTLSKNVKISSTGKKYCSTTKWPF

>Pb_LOR1_partial
MTSWQQTSGCCCSSGGSICCSYPCQGESSCCCGGGRGRDGRGSPNQQSQGSICCSGGSGGSSGQTIIVVPGSSGGDGWCCCCGGGGSSD
GSSGGSGQIIISSGGGVGRSGDRQSSECCIGGGSWGGSGSAMSQQKCDGGEICCGGVSGGGGSPGASGGVTINMGGGSGGGSSGCQTII
VPGGRRGSGSGQSLGSSMGGGSGGGSSGCQTIIVPGGRRGSGSGQSLGSSMGGGSGGGSSGCQTIIVPCGRRGSGSGQSSGSSMGGGSD
GGSSGGQTIIVPCGRRGSGSGQRSXXXSGHSSGSSMGGGSGGGSSGCQTIIVPCGRRGGGSGQSSGSSMGGGSGGGSSGCQTIIVPGGR
RGSGSGQSSGSSMGGGSGGGSSGCQTIIVPGGRRGSGSGQSSGSSMGGGSGGSSSGGQTIIVPGGRRGSGSGQSSGSSTGGGSSGGSSG
GGKTIIVQCGGASGGGSSRAQSSGCCIRGGMGAGSMQTKQPISIPPCIGPTK

>Pb_LOR2_partial
MSKQIQSSVCCGGGGSSAGCCVRSRSGGSSCCCACCCCCRGGSGHVIAVSGSQGSSCCGGIQQISSSSCCGGSSTGVVVVPGGQSSGCC
IGGGHGGGICQQKIPLIGVGGGGICCGGGSGGGSGGQTVIVPGGGSGCCGGSGGVRVVGGGSGSGWCCGGGSGGQTVIVPGSGSGCCGG
SGGVKVIGGGSGSGWCCGGGSGGQTVIVPGGGSGWWGGSXXXGGGSGGVRDVGGGSG

>Pb_PGLYRP3
MVTLQILFLILCSLSQATGCFQLITPSKWGAKPANCSQPLRDVPPEYVVIIHTAGNPCRTHADCRKEVKNIQDFHMNLKGWCDIAYSFL
IGEDGYVYEGRGWRNEGSHTYGYNDLSLGIAFIGTFVERSPQDVAWKALRCFLNFSVKIGYLAPEYILLAHSDVSDLVSPGEFVRAEIS
KWPNYKHSFYVLDRGGQ

>Pb_SCFN1
MTYFLDSVCTIVGIFHKYAQWQGRNLTLNRREMKMLIQTEFAEVLENPCDPQRVEFTFQLLDVNGDSLVDFNEYLIFIFQIAKACYSYL
QPREYLLQEDGSRALHEGEQGGSKRDHWQLQDEEREEDYVHERRGSDRTPLRPMEDSRRGELGKCYLPSEPEEEEEEKEEERNFQGRDR
ELRDGDRRIHPSWEHQERETERRRLEPKQQEDIETPEHHQLRQREPALEEGSPRQSRRRRNDNDRRQADEHLDRDEHRGRETEHREAEG
RLFRCGSKARVDQGNHHAEYNSERSQASCEPRELEDGRRDLETHDELRTRAFEPHVTDERRDQIRNRRPVDDYEEHRPLPREGERRSQS
HETEHRESERRRRDHSAFQSRDDVRIANRAEEEEVRGCERRRDEDEWSRPRQCEALRRPEVEELEHRERERRSYSPELENMDRRNQPHE
SEGRLRRSRHREDEERRRQQFSWEETQATESRRRRPRSHEPEPRRDPKYGRSQRYEEEPKRADQRRQHNYEVQSPERDVARRRLQTRES
AEREDNQERQQYYERESRGSEGDRRRPQLCYSEPREGEKRRHQLTDSRVVERERLQSRESVTRPTQRDRFSPSDSSGEYDWSRQQIYES
RDIEQREGRRRSYDRRSCGQRQIQGIDIDPREGEGQRRTNAEAGDVDQRQTQTRGVDHGESEQQRRTSSRDSGDKDIDQRRTQTDEASS
RDGEDQRRTSSHDSNTRQVNQKMMQTSDPDFRDTEIPQHNKGNAPQPREVPSLSNQTGDREFAEDLLRDQRPPESGESQLSRKPRQHKA
EQSQSIPRTLEQKQATRGSQCLSGPELRGPQRRWLQPHQTEAQEGEAGVEQTQLKVGSPESKRAASQPGEQQTTEGQESRVQRHHPKPA
TQDEGESVAGKGGKAQEPATVKRQQLQAQGASYEGEGPQPEVEGPKVEEDDENGPRAQESQPPLEEDDQQAVTEKPCDSLGESKSSVVC
NPLYEYLLAQKKQEQP

>Pb_SCFN2_partial
MACLVDSVCTIIGVFHKYAERRNACSTMKRREMKRLIQKEFGEILENPRDPQIVKLTFQLLDVNRDSLVDFNEFLLLIFEVATACYSSG
YPPKECLSPKEERSRVVRNGEPRGNENNHRQFQDEEREEDYVHERRGSDRTPLRPMEDSRQGELGKCYLPSEPEEEEEKEEEERNFQGRD
RELRDGDRRVHLSREQHKKETERRRLEPKQQEDIETPERHQLRQREPALEERSPRPSREPMRRNDSNRRAGDPLDRDEQDLYSPVVTSR
REDGRQRRGRETEHREAEGRLFRCGSKARVEEGNHRAECDSERSRASREPRELEDGRRDLETHDELRTRAFEPRVTDEXXXATAEARRP
RHHEDDEQRRQQSNREETRATESRSRRPRSHEPEPRDPKYGRSQGYEERRRDDPRRQRNYKVESPERDVARRRLQTRESAEREDNQDR
QDYEHESRDSEGDRRRPMILEPETRDEDQRRPQETGSRYGERERRQERDGRIRPTGRELGSPSDGGQRELDSSRPQISRCRTAELSRSE
PRSRKPARRG

B

>Pb_Beta1
MSGSGVKCVTTPCITSCPDAKVVVHPPPLILTLPGLSLRTSPNQCLLETQTSCLTNGSEVGCNEGSTAIVTKSSGVCALSGGDTPCCTT
TCPDSQVVIQPPPVCITIPGAVLTSYPNECLIKTSTPCVPSGPQPPALTRSTSVSDCSLTPRRLSRSASVPSGLDSLSRCVTRGPTTKV
VIHPPPIEVTIPGPVLEIAAQECAVEVYNPCSSNNALTSSDQKAITSSDEGEVKALTTQAKSCTTVCGLMDTSSCISQGPEMKIIIQPP
PIEVELPGPILEVFPEACKIETLNPSPLEPEAITGSETKALCSSKMTSTALATTSNVKRPLPNLRRAPRPWAEMYSRSMTPRTIALSQQ
TRLAKYRNTLHSMNFQPSF

>Pb_Beta2
MERRPLCYASCPPSTVTIQPPPFTLTIPGPAIFCPNQPFHIMQYNPCARDGMGIGGRMMFSDFISEDLPDLKLESSSGHGSALVTLYRS
RMALLLAMTGSN

>Pb_Beta3
MSECYASCPASTVTVQPPPFVLTIPGPALYCPDQVFGIEQYNPCAGFGGMVTRGGGGIGVEGRGGTLGALGAGRGGWGGSWSGGSGIGS
RIGGGLGIGGVGGPRGGGLGIGGVGGRRGGGLGIGGVGGPRGGGLGIGGVGGPRGGGLGIDGVGGPTGGGLDIDGPRGTTGSWSGGSGV
GGTVGGRRGSIGAGSITGGDLGTGGVGGTRGGCLSTGGTRVTTGGWSGGSRTGGSVGGRRGSIGAGGTVAGWSGGGGRRGSVGTVSVTG
GDYGLSGAGGGWSSSHRISGSHRWSGSGYGGSGGYSGGGYGGGYGGSSGYGGGFSSGRYGTSGYGGGYGHTGGYRRIKYSVGGGSGHRSG
GFSSRSYGGSYGGTLAVLPGPSDISCY

>Pb_Beta4
MSTSYNPADVCAFSCPFTMVTIQPPPFTLTIPGPTIHCPDQPLQFQQHNPCVFNKEHSPLLASRNSSFYSRALPSGSYTSSIYRY

>Pb_Beta5
MSANRSFWPNCYAQCPASTVTIQPPPFVLTIPGPALYCPDQPLGIDQYNPCEYGAYPLPLGRGGSNLTNFYSRGLPSGSFSSTTYRSSN

>Pb_Beta6
MASGWNPCCASCPPLTVTIQPPPYTVNIPGPSLYCADQPLCIEQCNPCVPIQPISHGGYSLLSSSAVSDFSSSSQKSLRSQRF

>Pb_Beta7
MASGWDQCYSNCPPMTVTIQPPPFTLNLPGPALHCPEQPFCIEQCNPCVPTPHCDNSLFSSTASGFSSSSVASQKSVSGRQSSGCFPCR
RC

>Pb_Beta8

MSFGSNSCYASCPFVTVTVQPPPFTLNIPGPSLYCPDQPLCIDQYNPCVPPFICPFRRSYTSNSISSSSSLPQKSLPRSQSEWSSVSQR
F

>Pb_Beta9
MFCGWNPCSSSRPMTAQPPFFSLNIPGISICCPEWPCCPEQYNPCAPVQPCGRMNYGRYGSYWSGRASDCNSLKFSAPKHLHQSQTKSL
HY

>Pb_Beta10
MASRWDLCSFNCPPTSVTIQPPPFTVSIPGPSLYCPDQPLCIEQCNPCVPCYGGSSYGRSRPVILASISSQKTVSQYQKGSEPFKRF

>Pb_Beta11
MSFGWNQCYPNCTPMTVTLQPPPFTMNIPGTSHCCPDQLLCVDSCSPCSPTVCAPMVCSPTVCSPTLCAPTHCGPTHCAPSQCAPVVCT
SGLNRSRYSSCFSSCSLFSRKCLPRVQSVCDPCTKY

>Pb_Beta12
MASRWDLCSFNCPATTVTIQPPPFTINIPGPSLYCPDQPLCIEQCNPCVPCYGGSNYGRFRPSSAISAASIYSQKSVSQQQSGCDPCQR
F

>Pb_Beta13
MNAARVYSGWDPRFSNCSPMTVTVHPPPLTLSIVSPAHYSPNHSPCIEPCNPCAPFQCWDGRNRYGYHALTNGVAPCLTSTNSSSQKCL
PRSKSGCVPCRSY

>Pb_Beta14
MTSYKLCTCVSCSSKLPYSKAKPSCLCMFVSEINIPSHQELVPARSETLSEAQNIRRFSRQKKSNLTCLPTSQLLGHLGRGPTRRISYN
LQRKPPLRSYRYIKC

>Pb_Beta15
MAFWDYPSFNSGLCGVSVSCVSQIPASEVIIQPPPVALTIPGPILSASTEPVAVTQNSPCALGALGACAPPGYGYGSGRRYHYGICARP
SYAQKSS

>Pb_Beta16
MSYCGPVCHVSCASSGGSCGLVPAYGSSSLSGAVSSYYGGGPITCSSQLTGSEVILQPPASVVTIPGPILSATTEPVTVGQVTPCSYSH
PLSFVGGSSYGGYGRLGYSGYGSYGTWRRYSRRCLPNYCGPC

>Pb_Beta17
MAFCGPPCPLPPCPIPSCAIPSCASCPSVGFGPGGLGGLASRSFGLPSGQPASSLGTLEGVTPSCINQIPAAEVVIQPPPVIVTLPGPI
LSASCDPVAVGGNTPCAAGGYGQGLPTGLLGGGSRQGSRYRFVGNRGSICYIPC

>Pb_Beta18
MAFCGPPCPLPPCAIPSCASSPSVGFGPGGLGGLASRSFGLPSGQPASSLGTLESVTPSCINQIPAAAVVSXXXEVVIQPPSVIVTLPG
PILSASCDPVTVGGNTPCAAGGYGQGLPTGLLGGGSSGPRPGRRFTFVGRRGSICYSPF

>Pb_Beta19
MPHCGPSCAIPSCASSPVVGFGSAGLGYRGLGLGCGYGLGYGYGLGSGYGLGSGYGLGSGYGPGALATSSGSLGTLAGVIPSSINQIPP
AEVVIQPPASIVTIPGPILAASCEPVCVGGNTPCAAGGFGRYGGYLGGRWGRLGRRGSVCPLPCPLPC

>Pb_Beta20_partial
MPYCGPSCAIPSCASAPVVGLGSTGCGPSGLGYRGLGLGYGYGGWGESASNLGTLAGVNPSCISQIPPSEVVIQPPPVVVTVPGPILSA
SCDPVSVGGYTPXXX

>Pb_Beta21
MPYCGPSCAVPSCASSPVVGFGSAGLGLGSGYGLDSGYGLGSGYGLGYGYGSGALATSSGSLGTLAGVNPSCINQIPPAEVVIQPPASV
VTIPGAILSASCEPVCVGGNTPCAVSDSGLRGSWGYGDWGYGGLGLRNRGLLGRRFPLSRRGSICYSRRGSICF

>Pb_Beta22_partial
XXXPGPILSASCDPVSVGGYTPYATGGFGRYGGSLGGRLGRFGRRGSVCPLPCPLPCPLPCPLPCTLPC

>Pb_Beta23_partial
XXXVPGPILSASCDPVSVGGYTPCATGGFGRYGGSLGGRLGRFGRRGSVCTLPCPLPCPLPCTLPC

>Pb_Beta24
MSCCVPSCTVPTCVPSGSPPICYPVGGLGSLNPCAISTAGGVSASSLGIVPGASVGCINQTPPSELVIQPPPITVVIPGPVLSASCEPV
RVGGFTACSGGSSNGGSSRVRCYPCNPCKV

>Pb_Beta25
MPHCGPSCAIPSCASSPVVGFGSAGLGYRGLGLGSGYGLGSGYGSGALATSSGSLGTLAGVIPSCINQIPPAEVVIQPPASIVTIPGPI
LSASCEPVYVGGNTPCAAGGFGRYGGYLGGRWGRLGRRGSVCTLPCPLPC

>Pb_Beta26_partial
MAFCAPVCGLPTCAIPSCASAPQYGLGSSGGSGGLGLGLGGRGGLGGDSLGGSTTSGQLGTLAGVNPQPINQIPSAEVVIRPPPVVVTI
PGPILSXXX

>Pb_Beta27_partial
XXXGRGGLGGGSLGGSTTSGQLGTLAGVNPQPINQIPSAEVVIQPPPVVVTIPGPILSASCEPVAIGGNTPCALSGSGPLGRPLALGTG
PFGRCLLGGRGNPCLPPCGF

>Pb_Beta28
MSYCGPACAIPSLASNPIVGFGSAGLGGPGYGLLPYSSGSSALAESSGGLGTLAGITPSCVNQIPPAEVVIQPPASIMTIPGPILSASC
EPVAVGGNTPCAVSGSGSDLFGNCLPGNPGLGLRRGTLLGRRSLLGNHGDVCL

>Pb_Beta29_partial
XXXCAISTAGGVSASSLGIVPGASVGCINQTPPSEVVIQPSPFLITIPGPLLSASCEPVRVGGFTAGGGGSSNGGSSRGRCYPCIPCNP
CRS

>Pb_Beta30
MSCYVPTCTIPTCVPTCPPASSSPICYPVGGVGSLSSSSSISSGGGTAASSLGIAPGATVSCINQTPASEIVIQPASVILTVPGPILSA
SCEPVRVGGYTACAGSSGRSSRSMMSGSRPYICRS

>Pb_Beta31
MSCCIPACLPPCPPPCPPTCAVPSCAAAPTIGLGSCGSGSGILGYGGGGAASASSLGILPGASVGCLSQIPPSEVVIQPPPFVITIPGA
ILSASCDPVAVGGYSPCASGSGGYLGGGLGGICRPRRKFSICKYPC

>Pb_Beta32
MSCCRPCCPPTCAIPSCASRPLIGLGGCGGGSALGGFGSGGGGAASASSLGMLAGVNPSSISQIPPSEVVIQPPPVVVTIPGPILSASC
DPVAVGGYSACSSGSYGSYSRGLLGASSRGICAPRRRYSICSSPC

>Pb_Beta33
MYCGSPCPPTCAIPSCRPTLGLGPCGLGSGSGGIGGGGSGIIGYGGGSGLAASASSLGMIPGVPVSCVSQIPPSEVVIQPPPYVLTIPG
PILSSSCEPLAVGGYSPCATGGYGASGGYGYSRYGLGSSGGYLGSGRCYGSKRRQSVCGYPC

>Pb_Beta34
MACFIPTYTTPAAASNVGLGSCRIGPNISDCGGYGYGGSGGVITGGPSSILGLTSGANIARTSRLPPSEIVIQPPLCVLTVPEPVVSAI
TPPLILGGSSSYIYGFPGYGYGPGSVGSCLRGDIMGGFQNIGQSRGRCNIYQYPS

>Pb_Beta35
MSLCNYCCNPVYRVDTPCILQIPLSEVAVQALPFLITVPGPVMSASPELVAVRGNTPRAAADCYEPYGSGGWVLEGGHGAPLAPIHIPR
GPTKGYF

Suppl. Fig. S1. Amino acid sequences of proteins encoded by EDC genes of the python (*P. bivittatus*). **(A)** Amino acid sequences of EDC proteins other than corneous beta proteins (CBP). **(B)** Amino acid sequences of CBPs, also known as beta-keratins. Amino acid residues K and Q (potential transglutamination sites), C (potential disulfide bonding sites), P, G and S are highlighted by specific colors corresponding to those in Figure 4. Stretches of X´s indicate unknown numbers of amino acid residues, that could not be predicted because of gaps in the corresponding gene sequences. Pb, *P. bivittatus*.

A

>Oh_CRNN
MSHLLRNIRSIICIFEKYAKDDGDCDTLSKGELKQLIQTEFADVIVNPHDPKTIETVLHLLDTDCDGKVGFEEFTVLVFKVVKACYKKV
QECQGTAGGPSKRGITSTQQQEIPPQQISRDGQSQRDITSIQQQEIPPQQISRDEPSQRDISRTQQQDIPPQQLSRDGPSKRDITSTQQ
QDIPPQQISRDGPSQRDTSRTQQRNIPPQQLSRDGPSKRDITRTQQQDIPPQQVSRDEPSQRDTSSTQQQEIPPQQISXDGPSQRDTSS
TQQRDIPPQQVSRDGPSQRDTSSTQQRETQTSQVSPRPCQKLVPPKKILDAEELPQEQERERTRLETSYDEKEQSQQEEASIRPRTQTS
RVLSVQRTIEDMIRGNLTCQMSLAEMRTEQAPAQGHPSALPPRRRCLQEPLSPDQEASQDVLQSVPQQSSQQRQVEERSSTCQEEVTSQ
TSGQDRRSQQSAVEMEQISSGESLRETRQRSSGGRNQQVQVEERSASGHLRIPQVVQICDQESRSVTLRQHQNDQQRCQSQGVEQSSVV
QEDQEVQAPGQVISRRQEPQMPVYRPHQEKERYPTTLECQRGPSEQRAVSQEVTRPQVAERYPRGEPGELTIVQFSQDSSTRRPFQPSQ
FQLKQRPLQFPPPWSTKQ

>Oh_EDCATM
MFFHKKLGWLTPNTVSSPTDSCLSPSTPSCSGGSPCLGSRTPPVHKVFRMAGLKPPASVTSSPLACEGLSSSIPKASGTPIVLMSCSPR
DAGFNPWNRNESAIWKERNIFTNELPNLSVFLSPLPASGGPVPLTLQCQPDCITPLCLTSFPLGHGPCGPLCVVPCATQPVNLFPCGTQ
VQLAPNCANMVPCTVKIAALGPCAMKGASLCNTEPPLASKCASPLSCTMTCPYAAKDANSLPCTMEIILASNCASPVSCATKGAAEMQY
DAKGAALTSPCGSKGADIHPCDVEILPAPNYASPTSCATKGATIIPYATKGATLTSPCGTKGANFHPCGVEILPAPKCTSPVSCATKGM
TTIPYATKGATMSCPCATKCVSAIPTLPQCACPVPCATKGAISAPLTPSWPVLTLCTPNSCHNAPAFTVPQVVQK

>Oh_EDCG
MGMWDCCSCCGNSSGRSTIVLCMPSGGYQSCYPGCMPSCGCAPMCCMPIYCTNMRRNSCCGCGCGC

>Oh_EDCM
MSCCRCCGGRGCCCCNRYRPRTVYYAGRSATAGCCCRPRGCCCGRPGCCCTSYYPSYPYYRVPVQTYGTQVKKC

>Oh_EDCP
MSYQCKQPCSCPTVCVKSAPAQTPCSPPSSCICVSEPVKVVKCPTSCQKSCTDTKAPAASCTSPCSSPCPVVCVSSCESKPACSCQTAT
CVVTTQSCDCTKPCQDTKSCAPAPEPCQTQTCTCIPVQLSQCCCGRIVLVPVSACAAPSAPPAPCCTCTCCCKPCCN

>Oh_EDCS1
MPCNCCNYCSPCCGTSTNKIICAPGQSYHVSPSCCMCGGGGCCGSCCGSCCSSCCCPPQMMCPSPMKCCCPPAMKCPPPVQLTQSCCLP
KKCC

>Oh_EDCS2
MYGGCGCGGGGGGSDGTTIICIPCGNSCGSMGCCMPCCGGMRGCCSPCCYMPCSPGGYGCGCGGIMAYRIPSSRRGSSSSCCGGY

>Oh_EDCS3
MTCCGDCCDCCGATCGNCGAVLLCVPCCGCQQCCSPCCGQSSCSSCGGQSSCSPCCGTKSSCGQSSCGGQSSCSSCGGSKSSCXGSKSS
CGQSSCSCGGSKSGCGQSSCSCGGSKSSCGQSSCSCGGSKSSCGQSSCSCGGSKSSCGQSSCGGQSSCGQSCCGGQSSCGQSCGGQSSC
GQSCCGGQSSCGQSSCGGCGQSSCSPCSCALPCIQTCCPISCPPCCKK

>Oh_EDEPK
MSTEQQQQKQTPVFPSGPSKLMAELVPEIASQKGHDGESLPSAEQGDVVPEDLKESHQKKEWTSECSAIEIPVSEDTSGAKHGSLTRKE
KEHLTSVDQHCREPPTLSQVPNKETFKSSEPPFPVREAEGHPATQPSVLGEQKQQQKKQTNQWLPK

>Oh_EDEPT
MAYQCKEPCLPPPIDVAQCSGDAQDSCTLTDMAQCVQTCDCVPLEACSAPDSLSYQTKQPCVSPSFKVALCPKAISIPCTSPCLSLRTE
MCAGSPSVPCGSVALQPGSPCPEDSQPCSTGTADACAPPSPPTCGSRCSGHCSGPQGDGGMSYVFPGSTVCLEPSGCVSVKPGVGSPEM
EGTTMFLASYQPHFM

>Oh_EDETM
MACLGLSSYNPAYSASLPSKSFLLRPPPFVAETRLTKLEILVPEHIPWDVPDVCSHPYSPLAFCYHEPPIVVGFFETSLHEV

>Oh_EDGPC
MGNMESNSGGGGGCGCSGCGCGGDYGGSDAHVICMPTGGCGCGPSCCAPCCCAPCCSPCCCAPCCSPCCCSPCCCSSGCGGQSSQCCSQ
CCKQCK

>Oh_EDGY1
MNFYTQPESIDSSYLPFYLSCYAPCYPQYGYGTGTSYCDPGNGCNYGYQPYSYGCGGLNTCSTSPCSYGYRYTDRGSCYEPWNYRMGYG
ARRNFPSRRYNCPNPW

>Oh_EDGY2
MSCYPNPSCYDPCRAKSFSSGSSGLKRRTRWSSGSCYGSYGRGVRYIPSCGYGYGGYQSPAHGGICYQPIRYDCGYGYGYGNKYCPPVC
YGGYGSGGYDYGCKYGGLYGRRQSICYDPCYGPRYQYGSRRGSFCWPC

>Oh_EDHEM
MPCTTPNKSTEEITAMHFVSIQPLDEISLGSHFHPNHHDPGIFLEDPNVMECIHFNDAHPHGEFYRRIHIVEMIELPFSPDAKESDCPL
HKHVHPGSEHSQEDPEKATSPSSSEPKSKEKLETANDGGGENIASNGDANNLSPSKDPEDAVGLMYTCKLEYQWSPMLKYCTMEELGYK
PLHLL

>Oh_EDKM
MSRLLRAFTEMIEGNSKVAPRKSTDAELLKKSEFKTLIQKELTPSKTRKHKNAKLLHESDGELMNEKELMGQWR

>Oh_EDP3
MSQQQCKQTPCVPPVSHRCTSSCSKPCEPSTCYASSGPHCPGSTGSHCPGSTGSQCPGMTGIPCSGSTGGSCQNPTCRSGNNGCFCRHP
QQQC

>Oh_EDPAM
MTHQCKLPVELPGCLKIGTKSSSQTKSSSAPSTCACCCGIPVSAPPPSPPPQTVTVCAIPAPAPAPTQSMTCCGIPVQSTSSPPPQTVS
ICAVPAPAPAQSQSITCCGIPLQTKSATPPPQTLTICVPAQAVAQSQCQSVSCCGIPVQAAKPAPAQTLTICAVPAPAPAPAPSCVPVT
LHLVQPASAPTSSCMCCSGCKT

>Oh_EDPCCC1
MTKCSKCGCSPCCCKCGCSQCCKTCCCPKCGCCPCCCCCCPKCGCCPCCCCSPCCCQQCNCAPCCCKQCVCAPCCCVCGCSPCCCQQSS
CSQSCCGQSKGCESKGCCSKCCFPKQGKS

>Oh_EDPCCC2
MTCCSKCGCSPCCCQQCCCSPCCCQQTCCQSCCEKGCGCGCGCGCKGGCGCGCXGCGCGCKGGCGCGCGCGCGCGCKGSCGCGCGCKGG
CGCGCKGGCGCGCGCKGGCGCGCGCPKQGKC

>Oh_EDPCCC3_partial
MSSGTKSPKCAKCGSSPCCCSKAASKKKCAKCGSSPCCCSKAASSLKSKCSKCGCSPCCCSKTLCSKCGCWPCCCPQSPCAKCGCXXXS
RCSRCGCSPCCCSQSRCSRCGCSPCCCSQSRCSRCGCCPCCCPQSRCCRCGCSPCCCRSQSRRCSKCGCFPCRCSQSPCCRCGCCPCCC
PQACCPRCGFPCCQQIQPYCYPCHR

>Oh_EDPCCC4_partial
MACCPICGCYPCCCVPCCSCANCCGCGPCCSSCGAPCCCSPCSSCCPCCACGGCGSPCSCPQCSSCCVCGPCCSRQSSGGGCGCGCSPC
CGCQSCGGSCGGSCGXXXCGGSCGGSCGSCGSCFPNQCFS

>Oh_EDPCS
MNYEYLFPEFLEEAEPYFFEANEDIVMPTKGGTYCPRPCPPYKCVTPRPRYCRPKYPCPCPPPCPPPCPPPCLPPCPPIRPQCPSPCYP
LYPLPQWPKPCYPSYPCRYPPQWICPKPLPCVPQRPYYPSCQLPCPPYWGRPCYPKCLPQYPCDLLEIY

>Oh_EDPKC
MASSHNQHQCKQTPTLPPEICKTAPCPPEQQCKQPPVVEVPKPKPCPPQEPPCKEPPVVVIPTPCPEPKPCPPKEPPCKEPPVVVIPTP
CPEPKPCPPKEPPCKEPPAPIPCPPEEEQCCQQKSQCKKPPVIPTPCPEPKKPQEPPCKEPPIVVIPTPTPCPEPKPCPPKEPPCKEPP
APTPCRPKEPEPCLPKEPTPDCQQKTQCKKPPVVVIPSPEPEPCPPQERPCKEPPVVVTPTPCPEPKPCPPQEPPCKQPPPCDQQQQKK
QPCSWPPQNK

>Oh_EDPQ1
MYTTENGDGEDSYSNYMRSVSCEPPSYSVRWSAEMEYCDTPPPPVRSYCPPRPIYCPPAPKGCPTKQKYCPPVQRYCPPVQKYRPPVKY
CCPPGRKSQPPAQWCSPPVQKCHPPVQSYCPPVQKYCRPVQRYCPPQQKYHPPAQPCYPPQQKYCPPAQSYRPPAQPCYPPQQKYCPPA
QSYRPPAQLCYPPQQKYCPPAQSYCPPAQPCYPPQQKYCPPAQSYCPPAEPCYPPQQKYCPPAQSYCPPAQPCYPPQQKYCPPAQPCCS
PQPKICQIKEVCKAPPHLLKK

>Oh_EDPQ2
MSYQQQQCKQPCQPPPQCQKTCPTPVCTPQVCIPKPCPPQDCIPDPCPPKVCIPRPCSPQECLPKPCPTQECLPRPCSPQECLPKPCSP
QECLPKPCPQQVCLPEPCPQDQQSCRPSQRPKRC

>Oh_EDPQ3_partial
MSYQSQQCKQPCPIPPQCKIIPPPQQCCPPPQKCCPPPQXXX

>Oh_EDPS1
MFHCCMPSCQMQPIAPLCVQPLQQPSISIPCSPPLPPMPSVSIPVSLSPPPSCIPVCRRQPSPSSSPTISIPLSSMQPPPSISLAPSQS
PSGSSPTISIPLSALIPSPPSQPCCISLCPQPSSSGPSISIPLAPSPSPSSRPPCFLPLCPGQGSPCISLVPSQPSSGSGPTISIPLSA
LQLRSSISLAPSQPSSRSSPTISIPLSALIPSPPSRPCCIPFGPSPSSSGPTISIPMNFSPRSNPSPICCFPIPMPSSPCFIC

>Oh_EDPS2
MFGYCLPSYYMGPIAPLDVQFYQQTSSMPAIGIPCRSVQSSMPPPCIPIRFSPGPSTSQSGPGSFPSCIPICPMQSPPSFSLPPYSPSP
SRQHTYCIPFDPNPSPPSSSQTQSIPLSFCPPPSSSSQTPYCIPLCPSAQSPSSGQTTSAGQPSYCIPVSTCQPSPSMSFCPAPQSSSS
RPEMSISLSPSLSSGSQSYCIPFCSYPSPSSSNQPSYSVPCCPSQTPSDANVSIPVSFSPSPSPNPPPTCRIPCSPCFLVF

>Oh_EDPS3_partial
XXXIPLSSLIPPPSKPSPPCYIPLCPSPESSNSSSGISIPLGSSSSSSNQPPCYIPICPGQSPPQPSSGSGPTISIPLSSLIPPSSSPS
QPCYIPVSPCQPSSGPTISIPMNLSPSPSSNPPPSCMIPIPMSSSPCFIY

>Oh_EDQK
MSSKDQQQKKPCGTFSTPQEPQKQCLPKNCCVVNPPETRSPCCLSPQPCTPPQKPGCLSCHPCPQNPKSYQQ

>Oh_EDQL
MCSRNNEGCSHDNTCHSGRSCCHEDRGYCGDRSNPSFTVTAIPVVQDKNPLTQTQKQCPIVYSTPQLQHTQQLKQPIPYPPQQQLKQPQ

QLKK

>Oh_EDQSG
MQQSKQPQGFPPMGQTTNQCKGSTGGQTQGPAVQQSQCQDSAPSQCQSSGVTHVHSSGVSQMQGCGVSQMQGSGVTQIQSSGLGQSQQS
DLDQCQGSGISQCQGSGGNQGKGSPMDQCGCSPVSQGKGSSGGQCKGSPVDQCGCSPVSQGKGSSGGQCKGSPVDQCGSSDVCQGKGSS
GGQVKGCPVDQCKGSKGKY

>Oh_EDSC1
MSQQQQQQGSCCHGGSSGLGGGCCRRGSSGGCCSRGTTGGGGCCGRRGSSGGCCSRGTGPVYVQQSQQTARCCCINRSGPCCGRCCGRR
CGGGCCGGGSTQLKQK

>Oh_EDSC2
MSQQQQRGGCCCCRCCGGGGGSSQVSRSSGSGCCGRSGGGSSQQAQSSGGGCCGRSGGGSSQQTQSSGGSCCGGSSQQKMK

>Oh_EDSCP
MVLSCHRPTCYSPSGPIFPVGGGSSCCVPVCHPAPCGPPTVCVQPVQSSCVPVCASPCGMSCGPPTVCVQPLQSPCVPVCASPCGISCG
PPTVCVQPLQSPCIPVCGSPCGPPAMCIQPKSSFPSCVSPSGPIFPVSSSSSIPVAPLCVQPSKPSCVSPSGPIFPVSGGHSSCMFTCA
IPVSVHPCSPCHLWF

>Oh_EDSPR1
MSYPYQQCKQPCLPPPVCGKGCATQCKVCVPACPCDVPICATPGVKVVGEACVPVCPPPCAPKCAGPAVDICVKPQGVCPCDCKCPDPC
CK

>Oh_EDSPR2
MSQQKQCCKIPPCPEACCPPKGPQKTCCSHPNKDVPQKGPCCPPSKDVKCPPQKAKQCPKHC

>Oh_EDSQ
MSYQCKQPCLPPPCLKGATVCVGPGTTVCISPTQSQGIEVCASPTGAICVTQGQSSCQCAKISHECCGPVVVTPVQTQSFSPCATIAVS
QNQGQNSTICVPQSQGCQCVTCRGSVATIQSQCGCATVCASPVCISPVAAPLAPAAVVCQDPCGNVCVASKGTAKCATKCAAPGTVVCV
DPCECVKSVKNCDDAVCAKGCQSVNQCNVKKC

>Oh_EDSRWM
MIFYTPQQYLDLDSWDINYGNDDNSPSEYFYADDPTSFVLDHKSRWDSPFYSCSSDLCGPRVRASSGSRSSCHPSCVKCHCWSPSCPDP
SCCPPWSGCGSCNPCYSPLCKPRPGRSRSAPPSMGNWGPQSSSSSCCPPWSRCGPCNPCAPCCRWGPPSMSRRGSQDSTSSCRYFPPLS
RCGPCNPCAPCCHWGPPSMSRRGSQDSTSSCRYFPPLSRCGPCNPCAPCCRWGPPSMSRQGSQDRSSCYPPRSTCGSCKPCYSPICKPR
CHSLRCPPPSMDSREPQSSSSSCCPPWSRCVPCNPCSPCCRWGPPSMSRRGSQDSSSSCRYFPPLSRCGPCNPSCPSCCWNFPSRIRWE
SQDRSSCYPPGSTCGSCNPCYSPICKPRCHSLRCPPPSMDSSCCCPPFCWSPPFMNRWGPPVCRSCPTIYYYPCYYC

>Oh_EDWM
MPEERIYSSGREAYFNLNSTWYDPAGSWLDTRRKPFRYVDNTACVTCCNPRSNVPRRGGHNYRCYCYRQCTCTPGGNPRVTCCVHNPSG
GPRDYWGRPIGDACDGCTGGHYSHAGSDCGSCCGSLGGCGTGGRTMACAQPCATSGGVCAEPGCRPAGRGVCAEPCITSSGGCSSGRGV
CAEPGCRPAGRGVCAEPCITSSGGCRTGVCAEPTCTPSGYGRRRRGVCFEPCSGTSNGC

>Oh_EDY1
MIGITILNNDPSGDTKAIYCPPTTYGPSASNPSICPPYSSLAGTAARCYIPRRYRHLFGLSPNCYLSDLSMACGYSSSLSDWRSACQPC
NYGHLFKSNQGYYDPCTFGFLSRGVQRSSDPYGLEYLYSYRRRYCDPCGYSYSPSYSRRCYDPCVYNYPSNYGRSSYPSRGYTYRCGYS
NDPCGYSCATWSCQPRYAERRRYSCQPVSCDPYGFSFESPYSSGSWRRSSDFCNDCGPY

>Oh_EDYM2
MPYYAYHCKPIGPPLPMFVKKRLPKYGSQYIPLYGSKLISSRNIPLYRRGIWVTPYSLAHVAGYQPRTIPVEFDACLNKFSKLRTTRNL
VRSDVEDYPYQIQSTVGYDGTTKGPHGPIHKGSLPCVVPERVPRMTKGPVLYSSRDLHQGLTQVPRVSRRPTFYASEDPLLCSSKEELS
QRTKRLTTQYPSKVPQGNRVRDSDLPGIDDPLQPEAKISSQPTATKGSRVITTKVPRANFLTLSHSFIKRGARSFLTKTSQAHFYRGSR
ASLVKPSRTSLVKGSQPSLPHTAIFPAPKNLSKTVKISSTGKKYSSAVNWPF

>Oh_LOR1
MASCQQSKSCCYSSGGRCVCICSGQGGSLCCCKGRGRGSVRCVSPIQQTQTSVCCGGGRRGSSGQTIIVLPGSGGGDGCCCRGRRGSSG
GRSGQIIISPGGGSGGGQSSECCIGGGICGVPGIGMNQQKCGGGIICSGGWRQQSGSAGGSCGVKVISGGSSGGRRGSGSRSRSGSPIG
RSSGGGQTIIVPCGCKGSGSRPGSPIGVCSCGDQTIIVPCPRRGSGSGSRPGSPIKGVCSGGQTIIVPCPRRGSGSGSRPGSPIKGVSA
GSQTIIVPCPRRGSGSGSRPGSPIIGGFGGGSIDGGKTIIVPCDGGSGGGGSSSVQVSGGGITVGVGAASMQTKQSIILPPCIGQTKQA
PQCPPLQKL

>Oh_LOR2
MSQQIQSSGCCSTGGGSSGHCHRRRGSGGTGCCCACCCCRGGSRHVMVVSGSQGSSCCGGIQQIPSYGCCGGSPTGVVVVPGGQSSGCC
IGGGLGGGICQQKVPLIGGDSGGGICCGGGSGGGLGGQTTIVAGGGSGSGCCGGSGVVKVIGGGSGSGSGWCGTGGGSGGQTTVITGGS
GGCCGGSGGVKVIGGGSGSGSGWCGTGSGGSGGQTTVVTGSGGGCCGGSGVKVIGGGSGSVCVDGGSEGVKVIGGGSSGGGGSFGDVKV
IGGGSGGVCGTGGGSGAKAVVVTGGSSEASACCSTGGSSGTVAVVGTGGSGGSGGKTIVVSGGGGGSDCCSGGGSGGASVVIAGGGGSS
QTKCPIVVPPCIGQTKQVCPLPPHIK

>Oh_PGLYRP3
MVILQISFLILCSLSQDLASSISPATGCFRLITPSKWEAKAANCSQPLKDVPAEYVVIIHTAGNPCRTHRDCHNEVRMIQNYHMNLKGW
CDIAYSFLIGEDGYVYEGRGWRNEGSHTYGYNDLSLGIAFIGTFVERSPEDKAWKALRCFLDFSVKIGYLSPEYIMLAHSDVSDIVSPG

EFIRAEIAKWPNYKHSLYILNRGGQ

>Oh_SCFN1
MSYFLDSVCTIVGIFHKYARCQDGNLALNRREMKALIQKEFAEVLENPCDPQTIELTFKLLDVNGDSLVDFNEYLIFVFQIAKGCYRYL
QPREYLLRDESSRALHEGEAGGSKRGDHQLQDGERRGDYVHERQGLDGTFLHSTEEGSRGELVGRYLPIEEVEEEDSNFEGHEPKLRDG
DRRDHPSQEHQERERPEQQWREPKEWEDIETPEPCQWRQQESTLEEDSPRQSRELVRRNDNDRREARGLLDRDEEDLYSPIVPSRREEMR
QHRGQETEGRLFRYGSQARVDHTDYDVERSWPSHEWQEVEDGRRELEMHNERRSHAFEHRVADDPRDRIHHRSPVEEYEEHHPLSTEGE
RRSQFHEDDHRESERRRQGRSVFQSRKDVRRAFRAEEEEVRRSERRHDEDEWNRPQPREASRRSEVQDLERRERETRRPYSSELKNMDS
PSRSHESESREEVRRSQHPEDIVQRRREFSREDTRVTEVRRRRSQSLEPVARDIRYVRSQRYEEPRSNDQRRPHNYEGRSLERDVARRR
RLQLRESPKREDNQERQWYYEHETWDSERDGRRCQLRYSEPRREREMRRYQVVDSRVDEREQSREAVTRPTQWDRFSPSDSPGEYDQSGQ
QIYESRHTLQREGRGQRYDRLSNGQRQIWAIDIAQRESEGQRRSNVEVRIVDQFRGVDHRESEHQRRTSIHDSDDENTDQRQTQTYEAS
SREDEDQRRTSSHDSGTRQINQRRTPTYEVGPKDDQRRRRMPSRESVDQRWRTHTQTDLAEDGQKPSNPYETDPRDGEQQRTVQSSDPN
FRDTEATQHNMGDSTQFREATSLSNQTADREFDLRRDQRPPESGESQPSRKPGQSKAEQSPVSPYEQHQPTRESQRLSGPELRGPQRRW
LPHHQTEAPEEAGVDQVQRKVGSPVSKRAISQPSERQITEGQESRIQGYHRTPVSQAEGESELSLDGEGDKAQEPAALNQQQVHREGEV
PQPVAEEPKVEEDDGSGSWAQESQPPLEDDDQQAITEKPSSSLVESKSCVICNPLYEYLLAQKKQEQP

>Oh_SCFN2
MAGLVDSICTIIAVFHKYADRKSESSSMKRRQMKRLIQKEFGDVLENPRDPQIVKLTFQLLDVNGDNRVDFNEFLFLIFEMATACYSVW
HPRECFSYNEERRRAVPDEEPRGDESNRREFLGEDRRRDDVRERRGADRTPLPSMEEGRRGELSSELREEVDRRFEGHDRELRDGDRRD
RPSRELQERERPEQRRLEPKEWEDVDIPEPRQRRQRESAVEEDSPRQSRELVRRNDNDRREARGLLDRDEEDLYSPTVISRREDERQRQG
RETEHQEAEGRLFRHGSQARVDTRNHRPEYDVERGQPSRERREGEDGRRELEMHNERRSRAFEPRVADDRRDRIRHRSPVEEYEEHRPL
SREGERRSQSYEDDHRESESRRQGRSVFQSREDVRRAIRAEEEEVRRSERRHVEDEWSRPRPREASRRPEVQDLERRERETRRPYSSEL
ENMDRPSRSHESEGREAVRRSQRPEDVERRRQELSWEETRATESRGRRPQSQEPEPRDPKYDRSQSYEEPRRDDQRRPRNYEVRSQERD
AERRRRLQPRELVEREDNQERQHYERESRDRERDGRRAMSLEAETRDEVRRRPQESASRYGERERLPEGDARMRRTGRDMASPSDAGRR
ELDPSSPQQISRRRAPELRRSEPHSFMPATKG

B

>Oh_Beta1
MSGSGVKCVTTPCVTSCPDAKVVVHPPPLVLTLPGLSLKTSPNQCLVESQTSCLTNGCEVNCDPGSKAIVTKTSGLVALSGGDTPCCTT
TCPDSQVVIQPPPVCITIPGAVLTSYPNECLISSSTNCITSGVQRPALPRSTSVSDCSLTPSRLTRSASVPSGLNSSTRCLTQGPSNKV
VIYPPPIEITIPGPVLEIAAEECTVEVYNSCTDNNALTGSDQAAITSGDESEVKTLIPRAKSCTTVSGLMDTSTCISQGPEMKIIIQPP
PIEVELPGPILEVFPEACKVETLTPCPPKPEAITGNEIKALCDSKIPSTALATTTTSTKRPLPDVRRPPRPWAEMYSRSMTPRSLALSQ
QSRLAKYRSALYSMHSQSSY

>Oh_Beta2
MDRPSPCYASCPPSTVVIRPPTFTLTIPGPAIFCPNQPLQIAQHNPCAHGGMGVEERMMLSDFINEDLPNLELELSSRYGSALATLYRS
RMAFPSESWL

>Oh_Beta3
MSECYASCPASTVTIQPPPFVLNIPGPALFCPDQAFGIEQHNPCAGFGGMGARGSGVGIEGTVGGSMGALGPGRGGWGGSWSGGSGIG
SRMGGGLGVAGVGGVGGLGGGGLGIGGPTGLGIGGLGGPRSGGLGIGGLGGPRSGGLGIGGLGGPRSGGLGIGGVVGPRGSLGLGGPR
GTTAGWSGGSRLGGXSVGMGSMTGGDLGSGEGGDMSGRRGSIGAGGSTVGWSGGSRLGGTVGGRRGSVGMGSMTGGDLGSGEGGVMSG
RRGSIGAGGSTVGWSGVGDTVGVRRGSVGTGSITEGDLGTGGVDGTIHGGLSSGGTRIITGGRGGGSGTGVSVGGRRGSIGTGGTIGG
GVGSSMSGRRGSIGTVSVTGGDYGLSGTRGSTVGGWTSGRRMSGSHRYGGSIGYGGGGYGGGYSSGGYGSSGYGDGYGHTGHGSGGYG
GGYGHTGHGSGGYRRIGYSVGGGDGYRSGGFSSRSYGGSFGGNLAVMPGPSDIYSYY

>Oh_Beta4
MSDHCSFSCPPLKVTIQPPPFIITVPGPSICCPDQPLCIEQHNPCAINHNSPPQAIRASSFYSRALPSTPYSPSIYRF

>Oh_Beta5
MWNECYAQCPASTVTIQPPPFVLTIPGPALYCPDQPLGIEQHNPCALPMPRGGSNLTSFYSRALPSTNFSSSTYRSSY

>Oh_Beta6
MASGWNCYTSCPPMMVTIQPPVYTLNIPGPSLYCADQSVCVEQCNPCAAIPSSGIINHGGYGLLPANVASDFSCSSPKSLENAYYF

>Oh_Beta7
MTSGWNYCYTSCPSMIVTIQPVPFAMNLTGPALHCPQQSLCIEQHNPCVPTPHCKDSLCSSTASNFSTSSVASQKSISRSQRSGCFPCR
WR

>Oh_Beta8
MSCGRMPCYSSCPPVTVTVQPPPFTLSIPGPSLYCPDQPLCIDQCNPCVPPPPLCPPITRRSYASSVVSSPSLPQKSLPRSQSEWSSVS
QRF

>Oh_Beta9
MASRWDLCSFSCPPTSVTIQPPPFTVTVPGPSLFCPDQPLCIDQYNPCVPQCLSYNRFRPLILASTSSQKTVSKF

>Oh_Beta10
MSYGWNNCYPNCTPVTVTLQPPPFSMNIPSSSLCCPDQYACIDPCRPCSPTVCSTMVCSPTVCGPSMCAPKACGPAACGPAHCGPVSCA

PVSCGPVSCAPVSCGPVSCGPVSCAPVSCAPVVCGPAFSRSRYNSYYNSCSLPSRKCLPRVQTICDPCTKY

>Oh_Beta11
MASKWDLCSFNCPPTTVTIQPPPFCLSIPGPSLHCPDQPLCIEQCNPCVPLPYRGSLPLPSSAASIYSQKSQQQQQQQQYQQRF

>Oh_Beta12
MHSGWDHCYPNCTPMTMMVQPPPLTLSIMSPAHYSPGHPACIKPHNPCAPVHCWDGRSHYGHHGLTQGAAPCFTSTSQKCLPKSKSGYV
PCRTY

>Oh_Beta13
MTSFNFCTCASCSSTRPCFHAKPSCLCMYFSQINMPLHQELVSTPREMLSTTQRIQGFLKQKTREYGCLSTSRPLKYLRGPTSGISYSL
QRNPPLQSYRYIKC

>Oh_Beta14
MCRENKLVMAIHPYFLDGRPNPPAVAVTIPGPILSACTEPIAITQHSPCAPSGGECAFPGYGNRSGRRYHNRIGPRSGGTQKSSSGNGG
PYLKNGKLEKIIRKSCKVFRNSENS

>Oh_Beta15
MSYCGSICHVPYASNSGSCALVSSYGSGTLTGLAANCYTGGHVNCSSQLTGSELIIQPPASVVSIPGPIISSTHDPVSVGQVTPCSYTH
PLNSYGAYGRLGYWGYGNYGNYGNYGGWRRYSRKCLTYN

>Oh_Beta16
MALCGPSCAIPSCASAPSVGFGSAGLGGLAPGSLGFSPFPLSESSGSLGTLAGIVPSCINQIPPAEVVLQPPSCVVTIPGPILSASCEP
VAVGGNTPCAVAGGFGQGLPLSLGGPSRLGARYGFVGNRGSICYTPC

>Oh_Beta17
MAFCGPSCTVPSCASAPSVGFGSAGLGGLAPGSLGLSPFPLSESSGSLGTLSGIVPSCINQIPPAEVVLQPPPCVVTIPGPILSASCEP
VAVGGNTPCALAGGFGQGLPTGLLGAGLLGSRLGRRYNFVGKRGSICYSPC

>Oh_Beta18_partial
MAYCGPACAVPSCASAFVVGFGSAGGKGLGWGLGYGGLGYGLGYGRGLGXXXGPILSASCEPVAVGGNTPCAPGGIGGIGGFGHYGGFY
GGRLGRLGRRGSICNLPC

>Oh_Beta19
MAYCGPACAVPSCASAPVVGFGSGGSKGLGWGHGYGLGYGYGLGYGYGRGLGYGYGAGALAETSGSLGTLAGVIPSCINQIPASEVTIQ
PPSSVVTIPGPILSASCEPVAVGGNTPCAPGGLRGGWGYGGWGYGGWGHGGLGFRGYGGLLGKRYPWNRRGSICLSRRGSVCL

>Oh_Beta20
MPYCGPSCAVPSCASAPVIGLGSSGCGPWGYGYRGLGLGYGAGALAETSGSLGTLAGVIPSCINQIPASEVTVQPPSVVVTIPGPILSA
SCEPVAVGGHTPCAAGGYGRYGGYYGGRLGRFGRRGSVCALPCNPC

>Oh_Beta21
MACCPPSCAVPSCASTPVVGLGSTGCGPCGYGSWGYGYGGLGLGYGYGSARFGESARNLGTLAGVVPSCISQIPASEVTIQPAPVVLTI
PGPILSASCDPVAVGGYTPCAPGGFGGYGGYYGGRLGRFGRRGSICSVGRRGSICTLPC

>Oh_Beta22_partial
MAYCGPACAVPSCASAXXXRLGRLGRRGSICALGSRGSICNLPC

>Oh_Beta23
MAYCGPACAVPSCASSPVVGFGSAGARGLGWGLGYGGLGYAGLGYGAGALAETSGSLGTLAGVVPQPINQIPASEVTIQPPSFVITVPG
PILSASCEPVAVGGNTPCAPGGIGRLGASYLGGRLGRLGRRGSIICNPCNLPC

>Oh_Beta24
MSCCAPACAVPTCIPACSSPVCYPVGGLGSLTSCGVGSYGMGSSAGGSSSAASLALAPGASVSCVNQIPPSEIMVQPTPIAVIIPGAIL
AATCEPVRVGGYTACASGSSSGGSSKLRYVPCNPCGPCK

>Oh_Beta25
MAYCGPSCAVPSCASAPAVGFGSAGARGLGWGLGYGGLGYGLGFGLGFGRGLGYGYGAGALAETSGSLGTLAGVIPQPINQIPASEVTI
QPPSFVITVPGPILSASCEPVAVGGYTPCAPGGIGRLGASYLGGRLGRLGRRGSIICNPCNLPC

>Oh_Beta26
MALCPPTYVIPSCASTPQFGLGSAGASAGLGLGGRGLGGGLQGGSISGLQGGSMGGLGGGLLGGGVSSGELGTLSGIVPQPINQIPPAE
IVIQPPSFIVTIPGPILSASSDPVAIGGNTPCAAPGSRILGRPSAPSRGLLGGRNLQEGRGNITLVPCGY

>Oh_Beta27
MAFCPPSCTIPSCASAPQFGLGSAGASAGLGLGGRGLGGGLQGGSMSGLGGGLLGGGVSSGELGTLSGITPQAINQIPPAEIVIQPPSF
IVTIPGPILSASCDPVAIGGNTPCAAPGSGILGRPSALSRGLLGGRNLQGGRGNVTLIPCGY

>Oh_Beta28_partial
MAFCPPSCIIPSCCYTSQLGLGSAGVSGGASAGGLXXX

>Oh_Beta29
MAYCGPACAVPSLASSPTVGFGSAGAGGLGYGILPYNPAASALAESSGSLGTLAGINPSCINQIPPAEVVLQPPSVLVTIPGPILSASC
EPTAVGGNTPCAISGSGIVGSDLYGNLLSGNLGLGLRRGTLLGRKSLLGSRGNICL

>Oh_Beta30
MSCCAPACAVPTYIPACSSPVCYPVGGLGSLTSCGMGSYGIGSAGGSMSAATLALAPGASVSCVNQTPPSELVIQPPPVAVVIPGAILA
STSEPVRVGGYTACASGSSSGGSSKLRYYPCIPCNPCRN

>Oh_Beta31
MSCCAPACAVPTCIPACSSPVCYPVGGLGSLTSCGMGSYGIGSAGGSMSAASLALAPGASVSCVNQTPPSELVIQPPPVAVVIPGAILA
STSEPVRVGGYTACASGSSGGSSKLRYFPCNPCGPCK

>Oh_Beta32
MSCYVPTCSVPTYVPTCAPPPSSPICYPVGGLGSLSSGGGMMSSGGSMMSSGGGGATLAASLGMAPGASVSCINQIPSSEVVIQPAPLM
LTIPGAILSASCEPVRVGGYTACATGPSGSSGRASRSMMYSCRPYICRS

>Oh_Beta33
MSCCPPACLPPCPPSCAIPSCASVPRIGLGSCGIGSGLLGYGGGGAASASSLGILPGANVGCINQIPSSEVVVQPPYFTVTIPGPILSA
SCEPVAVGGYSACAAGYGGYLGGGSGLGAICRPRRRASICKYPC

>Oh_Beta34
MSCCPPCVPCCPPSCVIPSCAARPTVGLGGCGTPSALGGFGGIGGYGGFGGFGGYGGFSSGSGPASASSLGTLAGVTPSPISQIPSSEV
VVQPPPVVLTIPGPILAASCEPVAVGGYSACAAGSFGGSYSRGLLGAGSRGICGPRRKFSICGSPC

>Oh_Beta35
MQSQCAPSCAIPSCRPVQALGSCGIGGGFGSGGFGFGGGYGMMGYGGGSGLATSASSLGLIPGVPVSCISQIPPSEVVIQPPPFSLTIP
GPVLASSCEPVAVGGYSPCSSGGYGVSGVYGSSRYGLGSSCGYLGGSGGYLGGGRSYGSKRRQSICGYPC

>Oh_Beta36
MACFNPSYTMPCATSNVGLGSCRTGPTVSGWGGYGYSGFGGSIGGGPSGMLGLTSGANFSRTSQLPPSEIVIQPPTCVLTIPEPVVTPI
TPPMVLSSSSPCSYGSLGYGYGSGYGYGSGSVGGCFRGDIMRGFQSIGPPRGRCNIYRCPC

Suppl. Fig. S2. Amino acid sequences of proteins encoded by EDC genes of the cobra (*O. hannah*). (A) Amino acid sequences of EDC proteins other than corneous beta proteins (CBP). **(B)** Amino acid sequences of CBPs, also known as beta-keratins. Amino acid residues K and Q (potential transglutamination sites), C (potential disulfide bonding sites), P, G and S are highlighted by specific colors corresponding to those in Figure 4. Stretches of X´s indicate unknown numbers of amino acid residues, that could not be predicted because of gaps in the corresponding gene sequences. Oh, *O. hannah*.

A

EDPCCC4

```
P. bivittatus gen  TATAAAAGGTCCCCCCATTCCCCAAACCCTCCATTCAGCTGGATTAACTCTCCTGGCCAC
P. regius RNA-seq  ----------------------------------------------CTCTCCTGGACAC

                                  <  intron  >
P. bivittatus gen  TTCCACTGCTTACAAAGGTAAG//TTCAGCTAGACCTTCGCTACTCATTGCAAAGATGAC
P. regius RNA-seq  TTCCACTGCTTACAAAG------------CTAGACCTTCGCAACTCATTGCAAAGATGAC


P. bivittatus gen  CTGCTGCCCCATCTGTGGCTCTTCACCCTGCGGTTGTGCCCCTTGCTACTCTTGTTCCCC
P. regius RNA-seq  CTGCTGCCCCATCTGTGGCTCTTCACCCTGCGGTTGTGC----------------------
```

B

EDSRWM

```
P. bivittatus gen  TATAAAAAGTCAGTTGAGTCCTGAGAAGTTCTAGAGATCCCTCCATACTGTCTATTGTCT
P. regius RNA-seq  ----------------------------------------------------CTATTGTCT

                           <  intron  >
P. bivittatus gen  GTGGTCCAGGTAAG//TTCAGACCCACCTTCATCCTCTCAAGATGAACTTCTACATGCCT
P. regius RNA-seq  GTGGTCCAG------------ATCCACCTTCATCCTCTCAAGATGAACATCTACATACCT


P. bivittatus gen  CAAGAATACTGGGACCTAAACAGCTGGGAGACCAACTATGAAAATAACTATGATTTCCCC
P. regius RNA-seq  CAAGAATACTGGGACCTAAACGGCTGGGAGACCGACTATGAAA-----------------
```

Suppl. Fig. S3. Alignment of RNA sequence reads versus genome sequences confirms the expression and the presence of an intron in the 5´-non-coding region of *EDPCCC4* and *EDSRWM* genes of pythons. (A) The nucleotide sequence of *EDPCCC4* in the *P. bivittatus* genome (gen) was aligned to the sequence of a *P. regius* RNA-seq read (GenBank sequence read archive, SRA, accession number ERR216300.5250820.1). (**B**) The nucleotide sequence of *EDSRWM* in the *P. bivittatus* genome (gen) was aligned to the sequence of a *P. regius* RNA-seq read (ERR216300.7157883.2). The complete sequences of the RNA-seq reads (100 nucleotides) were aligned to python gene segments corresponding to exon 1 and the start of exon 2 including the flanking sequences. TATA boxes are highlighted by green shading, splicing signals (GT and AG) at the ends of the intron are highlighted by blue shading, and start codons are highlighted by yellow shading. Hyphens were introduced to maximize the alignment. Red letters indicate identical nucleotides in both sequences.

A

>Ac_EDCATM
MFLSQLSCLLPPIFPTKPTGACSTQCMNSGTENRTSQYLDPCTAEHKSPSDSNFVSIPDPCTGVSAPKCVDLCGPEPKKSSQKCEDPSA
IAKDRLESSWGERWGIIINNNNDNGPQFPSSSAKQCTSTIRCPTKCTSMLPCTTKAYDSCATCVPRIILACARRCGIPCSNTCAA

>Ac_EDCS1
MGNCNGYGGGQSQPSMMCLPLGGCGSCCQPSCGQSCGSSGCLYLVSMGRSNSGGCCGGCCGGGGY

>Ac_EDCS2
MGQCGCCGGSGGGSMLVFTISCNPKCPPRSCPPQCCPPPYCPPPYCPPRSCPPPYYPQRYCPPSYRSSSCCSSSCCSSPCIYFIRLGGS
SNSGCCGGCGCCN

>Ac_EDCS3
MCSCGQSCKQSCCPPPPCCKSSCCAPPPCCQSSCCAPPPCCKSSCCAPPPCCKSSCCAPPPCCKSSCCAPPPCCKSSCCAPPPCCKSSC
CAPPPCCKQSCCSPPPCCAPPPCCKQSCCSPPPCCSPPPCCAPAPCCKQSCCSPPPCCSPPPCCAPAPCCKQSCCSPPPCCQSSCCAPP
PCCKSSCCAPPPCCKSSLLCTTSLLQILLLCTTSLAANPPAVHHLPAANPPAVHHPPVARARNAKGPTQVQGSLEIFHLSLNHVHHVCF
LAHLNCPNLLTWLLQPYL

>Ac_EDETM1
MVFPGEFSGCNTSYSTSLPPSTVLIQPPSYVAETSLTDLEIILPEPSPFEHPEYGFHNYHVPPLYCHLPPEQPLFLASFLTVLHQV

>Ac_EDETM2
MVFPGEFSGCNTSYSTSLPPSTVLIQPPSYVAETSLTDLEIILPEPSPFEHPEYGFHNYHVPPLYCHLPPEQPLFLASFFTVLHQV

>Ac_EDGPC1
MGNSGSYGCQCGSPYDCCSPACDQCCQSCCGQSCCAPACPSCCSPCQSTCPCCTPPCQQCCK

>Ac_EDGPC2
MGNSGSGGGCDCCGGGGGGESMMIVSCPGNNNCCPSPCCCAPRCCVPCCNQCCTQCCAPCCCQGCCK

>Ac_EDGY1
MGFYGGLDCYPPYFQPCYPFYYQPYYPQYGYGYGGYGYGYGNQGPCNYPQYGCGYGGYGYGSRYGSPGLQPCSYGYRYTNRGGCYDPRA
FRCGYGSRRSFPSRRYGYGGSYGGGCYGGYCGFF

>Ac_EDGY2
MTHIKGQSYCDPSGDKVYSGSSSGAYRQRNRRGSTGSYSGEYSQGRGCGDYGCYGGQGGQRNSDQGGYGYGGQRGRCYSGQEGYSYDGQ
RGRHYSGQESYGCDGQRGRRLSGQGGYSYDGQRARRFSGQGGYGYYGQGGNCYNNQGVYGYGGRQGGRHYSIQGGCGYGRQGGYGFGNQ
GDFGYGGLGGYGFGNQGGCGGLVGYGLGNQGGCGGLGGYGLGNQRGFGYGGLGGYGCLGGYGYGDQGRGTCLQPIGYGCRSEGYRYGNF
GCHSLGRGLLGRGCYGNTSSYGSGGYGYGSDGCGVRALLGYRHNSYQPGYSCGYPGN

>Ac_EDHEM
MLFLSLERSPPLLGTMDAQDCPSDTKPFKPTNFLSRTPRILIIRKQTSEEVTGNGDALVPDGNPNLVEVERERPPEKIQGNLPSVFCGND
SNGMGQEDPTPKEGSGDNDGGHGSVNLAATEGKPMEEMANHDHNSLPCANENCGIMTWRDCDIQQCKTDWRMGENIEAVLNWSGNRAST
GYDKAASYGYEGNNGSPESNDITITRPIVQQWSMDYRWSPLLKYCKPVQGHHECVSED

>Ac_EDP3
MSHQHQQQCVQPPSFPFVACVQDPKCVEPCPPPACQVVTDCQWPGSTGKPCQGSTGSSEQCVSGNHDGSASCVLQQEPKQC

>Ac_EDPAML
MARRKRSIVPPPCCTESEAKDIPSGDTPCESPCPSFTCNQKSPGAECTSLVEDTCCPEPTTADSLACQRYYCNAICGKKQVLALAGATV
QTISCSTLRPKPKKRGSYSTLRGSPTCGSPLQTSPCAHTCRIPLRLSPCNNRPRLPLQFTLCNTPSDPCSTLCSNTCNGQCNTPSSCPQ
SCTSCDEPLVMTAYLPITIRLCQATCSSNVPCVCASPCQSPTCSQCCQK

>Ac_EDPCCC1
MAYNYCHTSSPCCELFPCQKSGSPCYSHPCHKQCCDLCPKQCCDPYPKQCCSHPCCDPCCCPPHCKPQCCHPYYCPNPYNP

>Ac_EDPCCC2
MAKATSCSQCNCSPCCCEQCCCIPCESCACCPHCGVCPCSCRNRHASSDSYSCLCHECGYPN

>Ac_EDPCCC3_partial
XXXPCCCSKCSKCGCSPCCCPKCSKCGCSPCCCPKCSKCGCSPCCCSKSPCYQCKCSPCYSSKTPCSKCGCSPCCCPKCSKCGCSPCCC
PKSPCSTCGCSPCSCSKCSKCGCSPCCCPKSPCSTCGCSPCSCPKCSKCGCSPCCCPKSPCSNCESSPCCCYTHVKYSIYVPCCCPRCV
CARSALNRSCCLPCYRGQCYYNYSPCCCYGHSCGCGIY

>Ac_EDPCCC4
MTYGCGSCCNSCGYRNFCGPCCAPCCNSGCCHSGCCHSGCCSCCSPCCCSPCCCSPCCCSPCCCSPCCCSPCCCSPCCSPCCSPCCNPCCGGWG
CGSCCHPCCGY

>Ac_EDQK

MSSKDQQKKQQSQKPPEEQKASKNECPEKSPKTQKKQEKEEPTKKKQN

>Ac_EDQM
MCSRQDKDKAYKQEKEEKPSGCGRQNSGNESRGCGSGRQNSGSSCGKPQKPKPSQDQQQQQQQQQVCKVPDQKQK

>Ac_EDQSG
MSYQREQQYKPMYPPACCPKPYSPPEYSRPCPPSSSSGSTGIQCQGSSGTHCRWSTRIQSSSGTQCQESSGNRRQWFTGNPCQGSTWNS
CQEPSGTQYQGPTGSLWQCSSGTDGGTHSQYYHHYRQGSQDPKQC

>Ac_EDSC2
MSQQQQRGSSCCCSGGSSGGGGGGGGGGCCGRSGGSQQSQSSGGCCRSRRSSGGCCSRGGSSQQSQGSSGGCCGRSSGGSSGGCCGGSS
GGGSSQKKMPIQQMK

>Ac_EDSRWM
MTYYGCLGYCDPCNYSLPEKCSASNGYQRLYIQILPSEYYACSPCWDNYGHQISCVFPWYRMRFSYCSPCYSYNCCSPSYYCDPFCHTL
KSSDYIPRYKSTSRYCLPHGSYNYSPSYGPKTRCYSPYYSSKTKGCSHSSSNHQFTHRYGSPSSYCSPCYRTETIRGTPCYSSNECSSC
YSSPLSSCSSCYDAKTKAWSPCYSSNESSSCYGSPSSSCSSCYAFKTKGCTPCYNNHQHSPCHRCCSPYYTSRSRCCSPGYTYRRCSPC
CSSKGYSPHYNSGSKCRTPSYRYSQCLSCYRSRCHSPYYHFGSRCGAPGYRNSHYYPCRGADFGCHSPSHSSTCSSSHSTSCTYSMKSN
SYSSGCRRCSCPW

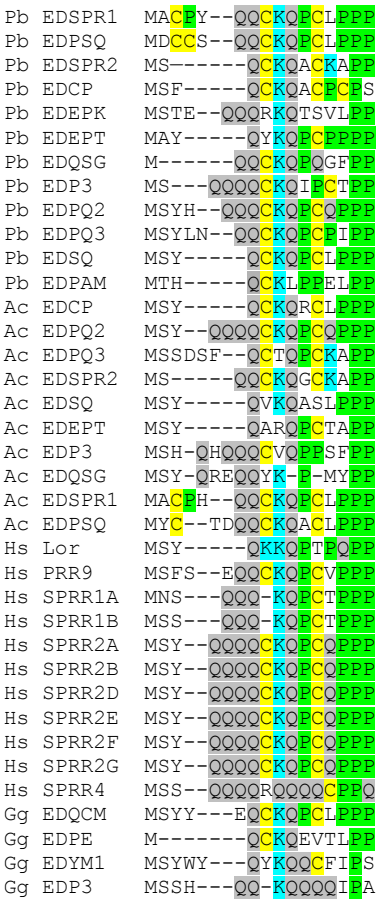
>Ac_EDY1
MAGFILFGPYPPFDLQKIYGSQGLYHPWYEPMYYEPWSYTDPSDVFANCYVPGYYNGLPNYNTSFLQPNSYGAFPGREPNGPDPLSNNG
GSGYKPRSYSHLSGLSRNGYQPGDSGNRGSNSKPHHFGSPTSDGDRYLNLYDSGCYRPRAYLSNSSRRCYDPCGYNFNSGLSNPCGYPY
LSGCGGSCYELPYSYGYPFSYGGSCYNPCGYNYPTSYGGRYCRPRGYTYISGYGGSCYDPYSAMCPYPYYTGRNRYFS

>Aca_EDYM1
MTNCNYQCWPQNFPPQGPQFGPGPQCGPGLQCGPGPQCGPGPQCGPGPQCGPGPQCGPGPQCGPGPQYGMMKPSASFMNMNQPPFMNMNQPANFPEPYKMN
FCNIQGGMPCAPCGPTPCDSKYAESSDAKSTEKHTSSCDTSGHESCTMKNSSQPTDRCSSQRPPCPSMPGISLCSPSCGPSMGRPHGFP
SSSQCFPPRNMYQYTTSKTFKSCYAK

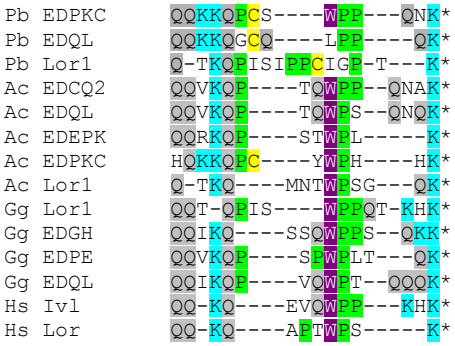
>Ac_EDYM2
MPYYRYYSKPLGPPLSPFVKKRLPKYGSQYIPFYGPRYFSSRTLPLYEKGLFSSKTLPLYQRGLYIPPYHQSGGAGYQPRGLNPRPESH
GTKLPQRNLARNPTQPHVNIYQRHLLQSPYSGMTKDRHGTITKGSVPGAPRAPLPRMTKGPVLYSPKDLEPHLPKAPLPRVTKGPVLYS
PRDLEPHLPKSPPPHIPKSPTKHSAKRIPPVVPTDSQLNVTKAQQHHTSKSMLSRATKGSRTIASKVPRFNLLNFSRSLRKKASRRNRA
KTSQPSLVKGSQPSLVRDSRGSLAKSSQPNLVHTATCPDLKKLSSNVKVSKTGKKYCSAAKWPF

Suppl. Fig. S4. Update of amino acid sequences of proteins encoded by EDC genes of *Anolis carolinensis* (Ac). (**A**) Newly identified amino acid sequences of EDC proteins in this study. Other EDC genes have already been reported by Strasser et al. (2014) Mol Biol Evol 31:3194-3205. Amino acid residues K and Q (potential transglutamination sites), C (potential disulfide bonding sites), P, G and S are highlighted by specific colors corresponding to those in Figure 4. Stretches of X´s indicate unknown numbers of amino acid residues that could not be predicted because of gaps in the corresponding gene sequences.

A

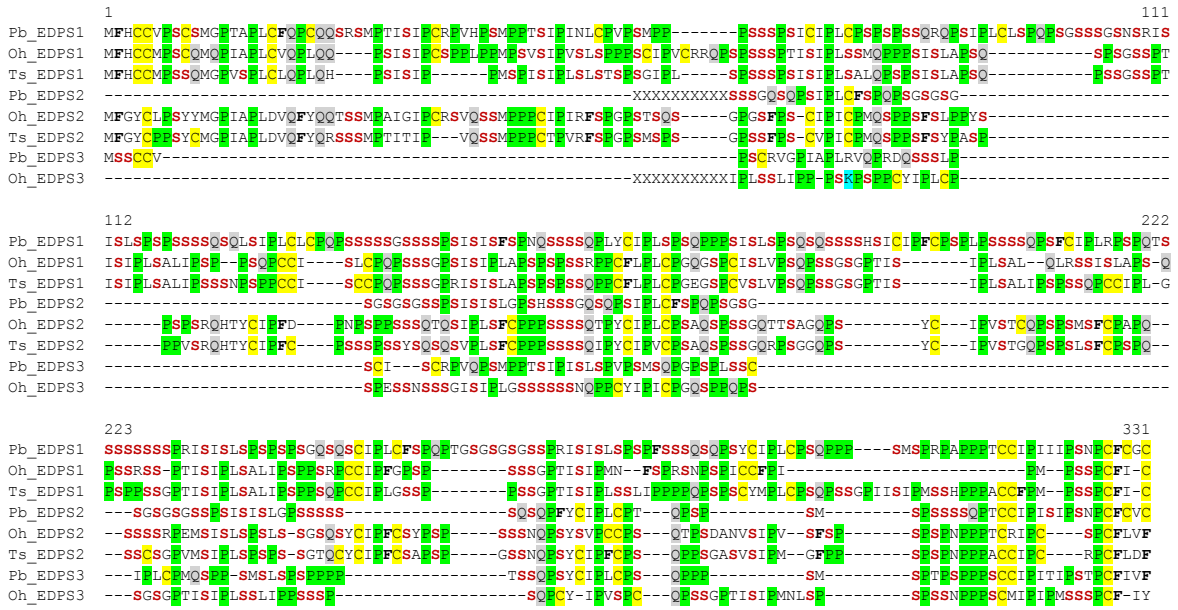


B



C

Suppl. Fig. S5. Conserved amino acid sequence motifs of snake SEDC proteins. Amino acid sequence alignments of motifs present at the amino-terminus **(A)** and carboxy-terminus **(B, C)** of some but not all SEDC proteins of the lizard (*Anolis carolinensis*, Ac), chicken (*Gallus gallus*, Gg) and human (*Homo sapiens*, Hs). The amino acid sequence motifs shown in A and B were discussed in a previous paper (27). The present study shows that these motifs are also conserved in several SEDC proteins of snakes, represented here by the python (*Python bivittatus*, Pb). Panel **C** shows a newly identified sequence motif at the carboxy-terminus of proteins that are encoded by gene neighbors of the beta-protein gene cluster. *, end of the protein.



Suppl. Fig. S6. Alignment of EDPS amino acid sequences of snakes. Amino acid residues K and Q (potential transglutamination sites), C (potential disulfide bonding sites), P, G and S are highlighted by specific colors corresponding to those in Figure 4. Stretches of X´s indicate unknown numbers of amino acid residues, that could not be predicted because of gaps in the corresponding gene sequences. Oh, *Ophiophagus hannah* (king cobra); Pb, *Python bivittatus* (Burmese python); Ts, *Thamnophis sirtalis* (common garter snake).

A

SCFN1

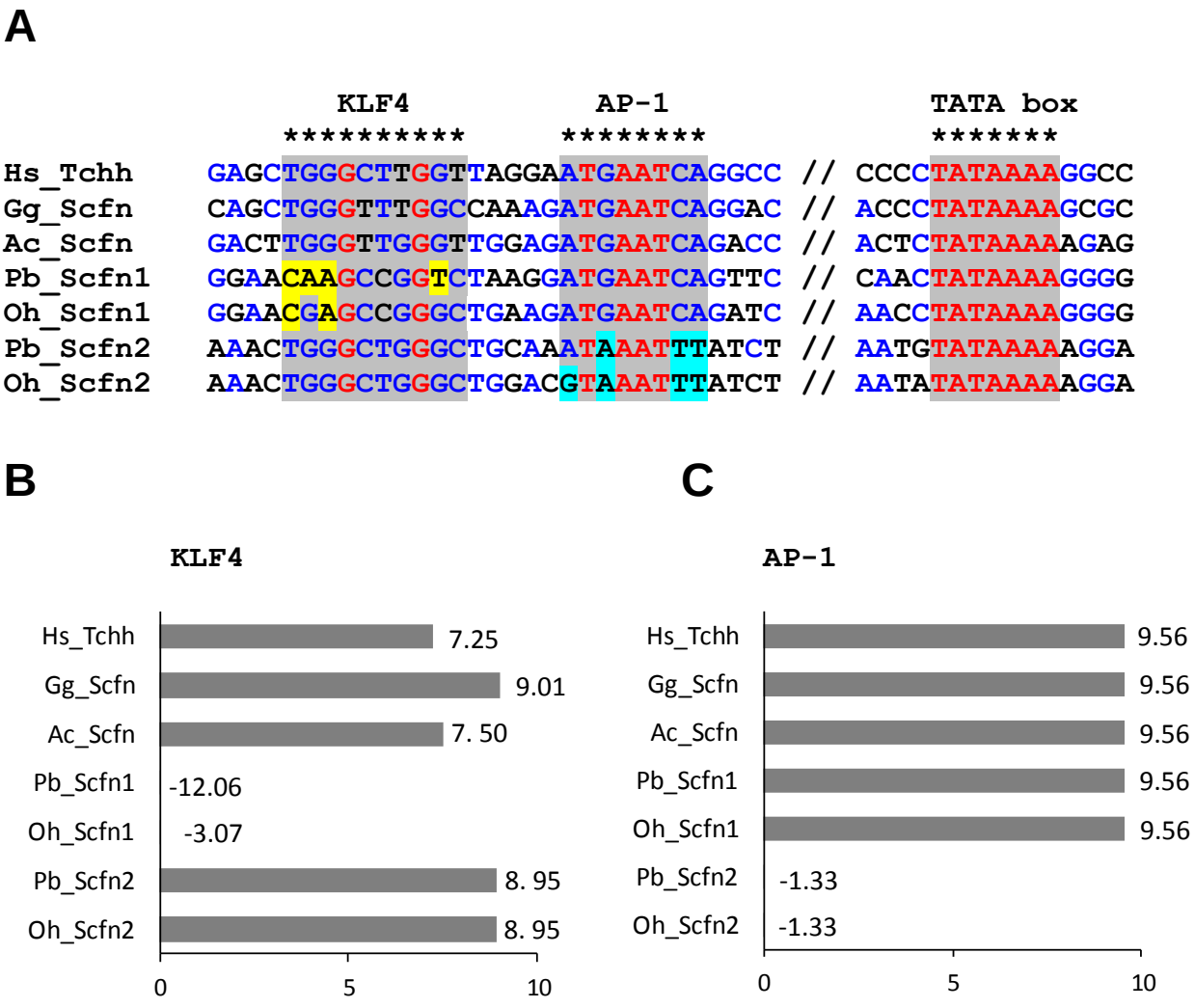
```
O. hannah genome seq. TATAAAAGGGGTTTGGCTTCCTGGCGTTTCCATACTTTTGTCTCTGAGGCTGCTCACCTG
E. coloratus RNA-seq1 ------------------------------CACACTTTTGTCTCTGAGGCTGCTCACCTA
E. coloratus RNA-seq2 ------------------------------------------------------------

                                    <  intron  >
O. hannah genome seq. ATTTTGCTGAGCTGGGTGAG//TGCAGAAAGGGTCTTCAGGATGAGCTACTTTCTGGACA
E. coloratus RNA-seq1 ATTTTGCTGAGCTGG------------AAAAAGACTTCAGGATGAGCTACTTTCTGGAAA
E. coloratus RNA seq2 ------------------------------------------------------------


O. hannah genome seq. GTGTCTGTACCATTGTTGGAATCTTTCACAAGTATGCCCGATGCCAAGATGGCAACCTCG
E. coloratus RNA-seq1 GTGTCTGCACCATTGTCGGAAT---------------------------------------
E. coloratus RNA-seq2 ------------------------TTCACAAGTATGCCCGATGTCAGGATGGCAACCTCG

                                                                              <
O. hannah genome seq. CTCTCAACCGGAGAGAAATGAAGGCGCTTATCCAGAAAGAGTTTGCTGAAGTCTTGGAGG
E. coloratus RNA-seq1 ------------------------------------------------------------
E. coloratus RNA-seq2 CTCTGAACCGGAGAGAAATGAAGACGCTTATCCAGAAAGAGTTTGCTGAAGTCTTGGAG-

                        intron  >
O. hannah genome seq. TGAG//TCCAGAATCCTTGCGACCCTCAGACAATTGAACTCACTTTCAAGCTGCTA
E. coloratus RNA-seq1 ------------------------------------------------------------
E. coloratus RNA-seq2 -----------AATCCTT-----------------------------------------
```

B

SCFN2

```
O. hannah genome seq. TATAAAAAGGAATCGGATACCTGCTATTCTCCAACAGTTCCCTGAAGGCTTCCAACC
E. coloratus RNA-seq3 ----------------------------------------CTGAAGGCTCCCTGCT
E. coloratus RNA-seq4 ------------------------------------------------------

                                       <  intron  >
O. hannah genome seq. GTTGAATGTACTGAGCTGGGTGAG//TAGAGACAAAGTCTCCAACATGGCTGGTCTC
E. coloratus RNA-seq3 GTTGAGCATACCGAGCTGG------------GCAAAGTCTCCAACATGGCTGGTCTC
E. coloratus RNA-seq4 ------------------------------------------------------


O. hannah genome seq. GTGGACAGTATCTGCACCATCATTGCTGTCTTTCACAAGTATGCTGACAGGAAGAGT
E. coloratus RNA-seq3 GTGGACAGTATCTGCAACATCATCGTGGTCTTTCAGAAG------------------
E. coloratus RNA-seq4 ------------------------------------------------------G


O. hannah genome seq. GAGAGTTCCTCCATGAAGCGAAGGCAGATGAAAAGACTCATCCAAAAAGAGTTTGGT
E. coloratus RNA-seq3 ------------------------------------------------------
E. coloratus RNA-seq4 GAGTGTTCCTCCATGAAGCGGAGGCAGATGAAAAGACTCATCCAGAAGGAATTTGGT

                                    <  intron  >
O. hannah genome seq. GACGTTCTAGAGGTAAG//CCCAGAACCCTCGTGATCCTCAGATTGTCAAGCTGACC
E. coloratus RNA-seq3 ------------------------------------------------------
E. coloratus RNA-seq4 GAAATTCTAGAG------------AACCCTCGTGACCCTCAAATTGTCAAGCTG---
```

Suppl. Fig. S7. Alignment of RNA sequence reads versus genome sequences confirms the expression and the presence of 2 introns in the *SCFN1* and *SCFN2* genes of snakes. (A) The nucleotide sequence of the *scaffoldin 1* (*SCFN1*) gene of the king cobra (*O. hannah*) was aligned to RNA sequence (RNA-seq) reads of *E. coloratus*: RNA-seq1 (GenBank sequence read archive, SRA, accession number ERR216301.7454688.2) and RNA-seq4 (ERR216301.7454688.1). **(B)** The nucleotide sequence of the *scaffoldin 2* (*SCFN2*) gene of *O. hannah* was aligned to RNA-seq reads of *E. coloratus*: RNA-seq3 (ERR216319.8005522.2) and RNA-seq4 (ERR216319.8005522.1). Green shading highlights TATA boxes, blue shading splicing signals (GT and AG) at the ends of introns, and yellow shading start codons. Hyphens were introduced to maximize the alignment. Red letters indicate identical nucleotides in 2 sequences.



Suppl. Fig. S8. Transcription factor binding sites in the promoters of SFTP genes. (A) Nucleotide sequences of the proximal promoters of the genes encoding S100 fused-type proteins (SFTPs) were aligned. As a previous investigation, which did not include sequences of snakes, suggested the presence of binding sites for the transcription factors KLF4 and AP-1 in SFTP promoters (see Figure S4 in Mlitz et al. 2014), the transcription factor binding scores were calculated using the the JASPAR 2016 server (http://jaspar.genereg.net). The positions of putative binding sites are indicated. Nucleotides compatible with the consensus binding motifs are shaded grey whereas nucleotides incompatible with the transcription factor binding are highlighted with yellow and blue background. **(B)** Binding scores for KLF4 in the promoters of SFTP genes at the site indicated in panel **A**. **(C)** Binding scores for AP-1 in the promoters of SFTP genes at the site indicated in panel **A**. Human (*Homo sapiens*, Hs), chicken (*Gallus gallus*, Gg), green anole lizard (*Anolis carolinensis*, Ac), python (*Python bivittatus*, Pb), and cobra (*Ophiophagus hannah*, Oh) sequences were analyzed. Scfn, scaffoldin; Tchh, trichohyalin.

Suppl. Table S1
Tentative abbreviations and full names of EDC genes identified in this study

Gene name abbreviation	Full gene name
Crnn	Cornulin
EDCATM	Epidermal Differentiation protein containing the CAT Motif
EDCG	Epidermal Differentiation protein rich in Cysteine and Glycine repeats
EDCM	Epidermal Differentiation protein containing a CCCC Motif
EDCP	Epidermal Differentiation protein rich in Cysteine and Proline
EDCRP	Epidermal Differentiation Cysteine-Rich Protein
EDCS1	Epidermal Differentiation protein, Cysteine-rich Short 1
EDCS2	Epidermal Differentiation protein, Cysteine-rich Short 2
EDCS3	Epidermal Differentiation protein, Cysteine-rich Short 3
EDEPK	Epidermal Differentiation protein rich in glutamic acid (E), Proline and lysine (K)
EDEPT	Epidermal Differentiation protein rich in glutamic acid (E), Proline and Threonine
EDETM	Epidermal Differentiation protein containing an ET Motif
EDETM2	Epidermal Differentiation protein containing an ET Motif 2
EDGPC	Epidermal Differentiation protein rich in Glycine, Proline and Cysteine
EDGPC2	Epidermal Differentiation protein rich in Glycine, Proline and Cysteine 2
EDGY1	Epidermal Differentiation protein rich in Glycine and tyrosine (Y) 1
EDGY2	Epidermal Differentiation protein rich in Glycine and tyrosine (Y) 2
EDHEM	Epidermal Differentiation protein containing a HEM Motif
EDKM	Epidermal Differentiation protein containing a KKLIQQ Motif
EDP3	Epidermal Differentiation protein rich in Proline 3
EDPAM	Epidermal Differentiation protein containing a PA Motif
EDPAML	Epidermal Differentiation protein containing a PA Motif Like
EDPCCC1	Epidermal Differentiation protein containing PCCC repeats 1
EDPCCC2	Epidermal Differentiation protein containing PCCC repeats 2
EDPCCC3	Epidermal Differentiation protein containing PCCC repeats 3
EDPCCC4	Epidermal Differentiation protein containing PCCC repeats 4
EDPCS	Epidermal Differentiation protein rich in Proline, Cysteine and Serine
EDPKC	Epidermal Differentiation protein rich in Proline, lysine (K) and Cysteine
EDPQ1	Epidermal Differentiation protein rich in Proline and glutamine (Q) 1
EDPQ2	Epidermal Differentiation protein rich in Proline and glutamine (Q) 2
EDPQ3	Epidermal Differentiation protein rich in Proline and glutamine (Q) 3
EDPS1	Epidermal Differentiation protein rich in Proline and Serine 1
EDPS2	Epidermal Differentiation protein rich in Proline and Serine 2
EDPS3	Epidermal Differentiation protein rich in Proline and Serine 3
EDPSQ	Epidermal Differentiation protein rich in Proline, Serine and glutamine (Q)
EDQK	Epidermal Differentiation protein containing glutamine (Q) and lysine (K) repeats
EDQL	Epidermal Differentiation protein rich in glutamine (Q), close to Loricrin
EDQL2	Epidermal Differentiation protein rich in glutamine (Q), close to Loricrin 2
EDQM	Epidermal Differentiation protein containing a glutamine (Q) Motif
EDQSG	Epidermal Differentiation protein rich in glutamine (Q), Serine and Glycine
EDSC1	Epidermal Differentiation protein rich in Serine and Cysteine 1
EDSC2	Epidermal Differentiation protein rich in Serine and Cysteine 2
EDSCP	Epidermal Differentiation protein rich in Serine, Cysteine and Proline
EDSPR1	Epidermal Differentiation protein Small Proline Rich 1
EDSPR2	Epidermal Differentiation protein Small Proline Rich 2
EDSQ	Epidermal Differentiation protein rich in Serine and glutamine (Q)
EDSRWM	Epidermal Differentiation protein containing a SRW Motif
EDWM	Epidermal Differentiation protein containing a WYDP Motif
EDY1	Epidermal Differentiation protein rich in tyrosine (Y)
EDYM1	Epidermal Differentiation protein containing Y Motif 1
EDYM2	Epidermal Differentiation protein containing Y Motif 2
Lor1	Loricrin 1
Lor2	Loricrin 2
Pglyrp3	Peptidoglycan recognition protein 3
Scfn1	Scaffoldin 1
Scfn2	Scaffoldin 2

Note - EDC genes encoding corneous beta-proteins (beta-keratins) and S100A proteins are not included here.

Suppl. Table S2

Burmese python *(Python bivittatus)* EDC genes (other than corneous beta protein genes)

Gene	Accession nr.	CDS start	CDS end	Sequence complete	Expression confirmed by *P. regius* RNA-seq data
S100A1	NW_006533184.1	1588	555	yes	yes
S100A13	NW_006533184.1	14911	15919	yes	yes
S100A14	NW_006533184.1	19278	23485	yes	yes
S100A16	NW_006533184.1	37282	38809	yes	yes
S100A2	NW_006533184.1	46616	47229	yes	yes
S100A3	NW_006533184.1	51063	52413	yes	yes
S100A4	NW_006533184.1	56042	57171	yes	yes
S100A5	NW_006533184.1	61202	61812	yes	yes
S100A6	NW_006533184.1	66291	66748	yes	yes
S100A12	NW_006539396.1	10278	9093	yes	yes
PGLYRP3	NW_006539396.1	4288	1166	yes	yes
EDKM	NW_006540970.1	21714	20641	yes	yes
EDPQ3	NW_006540970.1	15196	15360	yes	yes
EDPQ2	NW_006540970.1	9469	9227	yes	yes
EDSC1	NW_006543838.1	4076	3660	yes	no
EDSC2	NW_006543838.1	12950	13385	no	no
EDQM	NW_006533945.1	464	240	yes	yes
EDWM	NW_006533945.1	10373	9753	yes	yes
EDPQ1	NW_006533945.1	15513	14725	yes	yes
EDCS1	NW_006533945.1	21723	22019	yes	yes
EDHEM	NW_006533945.1	24848	24321	yes	yes
EDCS2	NW_006533945.1	31505	31266	yes	yes
EDCM	NW_006533945.1	35942	36157	yes	yes
EDCS3	NW_006533945.1	40476	39946	yes	yes
EDPCCC1	NW_006533945.1	46880	46335	yes	yes
EDPCCC2	NW_006533945.1	51062	50646	yes	yes
EDPCCC3	NW_006533945.1	63364	61049	no	yes
EDCG	NW_006533945.1	77683	77486	yes	yes
EDPCCC4	NW_006533945.1	81785	81345	yes	yes
EDGPC	NW_006533945.1	85309	85572	yes	yes
EDQL	NW_006533945.1	92571	<92239	no	yes
LOR2	NW_006533945.1	102481	101773	no	yes
LOR1	NW_006533945.1	109065	107293	no	no
EDY1	NW_006538280.1	8900	8205	yes	yes
EDSRWM	NW_006538280.1	15145	17079	yes	yes
EDGY1	NW_006538280.1	23044	22733	yes	yes
EDGY2	NW_006538280.1	34934	35329	yes	yes
EDETM	NW_006540169.1	33959	34210	yes	yes
EDPS1	NW_006541849.1	19165	20127	yes	yes
EDPS2	NW_006547155.1	<1262	867	no	no
EDPS3	NW_006533133.1	7040	7441	yes	yes
EDSCP	NW_006533133.1	11620	10994	yes	yes
EDYM2	NW_006533133.1	51320	50361	yes	yes
EDPSQ	NW_006533133.1	59918	58921	yes	yes
EDEPK	NW_006533133.1	66212	65754	yes	no
EDPKC	NW_006533133.1	77786	78652	yes	yes
EDP3	NW_006533133.1	83736	83521	yes	yes
EDQSG	NW_006533133.1	87350	87931	yes	yes
EDSPR1	NW_006533133.1	90689	90375	yes	yes
EDPCS	NW_006533133.1	97800	98342	yes	yes
EDCP	NW_006533133.1	107373	106795	yes	yes
EDCATM	NW_006533133.1	113018	113974	yes	yes
EDPAM	NW_006533133.1	121454	122536	yes	yes
EDSQ	NW_006533133.1	130252	131010	yes	yes
EDEPT	NW_006533133.1	135106	135657	yes	yes
EDSPR2	NW_006533133.1	139240	139043	yes	yes
CRNN	NW_006533133.1	146971	149243	yes	yes
SCFN2	NW_006533133.1	154919	152258	no	no
SCFN1	NW_006533133.1	160207	164241	yes	no
S100A11	NW_006533133.1	167015	170608	yes	yes
S100A10	NW_006533133.1	183743	184878	yes	yes

Notes - CDS, coding sequence.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S3
Burmese python *(Python bivittatus)* corneous beta protein (beta-keratin) genes

Gene	Accession nr.	CDS start	CDS end	Sequence complete	Expression confirmed by *P. regius* RNA-seq data
Beta1	NW_006533945.1	129573	128446	yes	yes
Beta2	NW_006533945.1	141471	141166	yes	yes
Beta3	NW_006533945.1	148257	149408	yes	yes
Beta4	NW_006533945.1	153207	152950	yes	yes
Beta5	NW_006533945.1	162249	161980	yes	yes
Beta6	NW_006533945.1	175889	175638	yes	no
Beta7	NW_006533945.1	178952	179227	yes	no
Beta8	NW_006533945.1	182958	182686	yes	yes
Beta9	NW_006533945.1	185558	185833	yes	no
Beta10	NW_006533945.1	187387	187124	yes	yes
Beta11	NW_006533945.1	193099	192722	yes	no
Beta12	NW_006533945.1	196728	197000	yes	no
Beta13	NW_006533945.1	204035	203727	yes	no
Beta14	NW_006533945.1	212727	213041	yes	yes
Beta15	NW_006538280.1	44064	44354	yes	no
Beta16	NW_006538280.1	49426	49821	yes	yes
Beta17	NW_006538280.1	52808	52377	yes	yes
Beta18	NW_006538280.1	57570	58020	yes	yes
Beta19	NW_006538280.1	62708	62235	yes	yes
Beta20	NW_006542926.1	15275	>15577	no	yes
Beta21	NW_006542926.1	2255	2746	yes	yes
Beta22	NW_006551345.1	>212	15	no	no
Beta23	NW_006555159.1	>522	334	no	no
Beta24	NW_006552039.1	197	556	yes	no
Beta25	NW_006540169.1	583	1002	yes	yes
Beta26	NW_006540169.1	5880	<5596	no	yes
Beta27	NW_006540169.1	<11505	11825	no	yes
Beta28	NW_006540169.1	17257	16829	yes	yes
Beta29	NW_006541849.1	>5559	5290	no	yes
Beta30	NW_006541849.1	13806	13432	yes	yes
Beta31	NW_006533133.1	21000	20548	yes	yes
Beta32	NW_006533133.1	27649	28053	yes	yes
Beta33	NW_006533133.1	31407	30952	yes	yes
Beta34	NW_006533133.1	39313	39747	yes	no
Beta35	NW_006533133.1	43812	43522	yes	no

Notes - CDS, coding sequence.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S4

King cobra (*Ophiophagus hannah)* EDC genes (other than corneous beta protein genes)

Gene	Accession nr.	CDS start	CDS end	Sequence complete
S100A9	AZIM01042421.1	45	>188	no
PGLYRP3	AZIM01003248.1	1855	5074	yes
EDKM	AZIM01003248.1	12034	12171	yes
EDPQ3	AZIM01003248.1	19465	<19349	no
EDPQ2	AZIM01003248.1	33024	33395	yes
EDSC1	AZIM01003248.1	41385	41068	yes
EDSC2	AZIM01003248.1	52467	52712	yes
EDWM	AZIM01003248.1	68925	68212	yes
EDPQ1	AZIM01003248.1	73668	72802	yes
EDCS1	AZIM01003248.1	78687	78968	yes
EDHEM	AZIM01003248.1	82623	82117	yes
EDCS2	AZIM01003248.1	87881	87576	yes
EDCM	AZIM01003248.1	92670	92894	yes
EDCS3	AZIM01003248.1	97472	96797	yes
EDPCCC1	AZIM01003248.1	104405	104049	yes
EDPCCC2	AZIM01003248.1	109843	109472	yes
EDPCCC3	AZIM01003248.1	122470	118393	no
EDCG	AZIM01003248.1	133443	133243	yes
EDPCCC4	AZIM01003248.1	137976	137235	no
EDGPC	AZIM01003248.1	141722	142009	yes
EDQL	AZIM01003248.1	145280	144999	yes
LOR2	AZIM01003248.1	151769	150621	yes
LOR1	AZIM01003248.1	157066	155969	yes
EDY1	AZIM01004187.1	73541	74254	yes
EDSRWM	AZIM01004187.1	64099	62798	yes
EDGY1	AZIM01004187.1	58166	58483	yes
EDGY2	AZIM01004187.1	47891	47478	yes
EDETM	AZIM01000954.1	23704	23952	yes
EDPS1	AZIM01000954.1	41059	41844	yes
EDPS2	AZIM01000954.1	58107	58886	yes
EDPS3	AZIM01000954.1	<63568	63978	no
EDSCP	AZIM01000954.1	69647	69066	yes
EDYM2	AZIM01000954.1	107053	106094	yes
EDEPK	AZIM01000954.1	119831	119364	yes
EDPKC	AZIM01000954.1	131139	131972	yes
EDP3	AZIM01000954.1	135690	135409	yes
EDQSG	AZIM01000954.1	140286	140879	yes
EDSPR1	AZIM01000954.1	146227	145952	yes
EDPCS	AZIM01000954.1	153869	154345	yes
EDQK	AZIM01000954.1	156359	156141	yes
EDCP	AZIM01000954.1	162072	161572	yes
EDCATM	AZIM01000954.1	167242	168537	yes
EDPAM	AZIM01000954.1	174551	175153	yes
EDSQ	AZIM01000954.1	183997	184629	yes
EDEPT	AZIM01004605.1	3309	3890	yes
EDSPR2	AZIM01004605.1	8254	8066	yes
CRNN	AZIM01004605.1	14916	17816	yes
SCFN2	AZIM01004605.1	23685	21601	yes
SCFN1	AZIM01004605.1	28678	32401	yes
S100A11	AZIM01004605.1	<38940	40166	no

Notes - CDS, coding sequence. Further S100A genes are present close to border of the EDC.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S5

King cobra (*Ophiophagus hannah)* corneous beta protein (beta-keratin) genes

Gene	Accession nr.	CDS start	CDS end	Sequence complete
Beta1	AZIM01003248.1	174867	173737	yes
Beta2	AZIM01003044.1	87502	87801	yes
Beta3	AZIM01003044.1	81345	79850	yes
Beta4	AZIM01003044.1	76531	76767	yes
Beta5	AZIM01003044.1	68820	69074	yes
Beta6	AZIM01003044.1	55588	55848	yes
Beta7	AZIM01003044.1	52619	52344	yes
Beta8	AZIM01003044.1	49125	49403	yes
Beta9	AZIM01003044.1	43757	43984	yes
Beta10	AZIM01007131.1	26847	27329	yes
Beta11	AZIM01007131.1	17992	17738	yes
Beta12	AZIM01007131.1	13225	13509	yes
Beta13	AZIM01007131.1	3381	3070	yes
Beta14	AZIM01004187.1	42024	41680	yes
Beta15	AZIM01004187.1	36898	36512	yes
Beta16	AZIM01004187.1	34624	35034	yes
Beta17	AZIM01004187.1	30955	30533	yes
Beta18	AZIM01004187.1	27140	27531	no
Beta19	AZIM01004187.1	23240	22722	yes
Beta20	AZIM01004187.1	18585	18992	yes
Beta21	AZIM01004187.1	7539	7093	yes
Beta22	AZIM01004187.1	2467	2882	no
Beta23	AZIM01041946.1	772	344	yes
Beta24	AZIM01008786.1	16443	16057	yes
Beta25	AZIM01008286.1	17037	16576	yes
Beta26	AZIM01008286.1	11937	12416	yes
Beta27	AZIM01008286.1	8808	8353	yes
Beta28	AZIM01008286.1	1958	>2062	no
Beta29	AZIM01000954.1	3170	2733	yes
Beta30	AZIM01000954.1	29658	29272	yes
Beta31	AZIM01000954.1	38016	37636	yes
Beta32	AZIM01000954.1	47356	46940	yes
Beta33	AZIM01000954.1	76707	76306	yes
Beta34	AZIM01000954.1	85348	85815	yes
Beta35	AZIM01000954.1	89721	89242	yes
Beta36	AZIM01000954.1	100589	101041	yes

Notes - CDS, coding sequence.

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S6
Green anole lizard *(Anolis carolinensis)* EDC genes newly (*) identified in the present study.

Gene	Accession nr.	CDS start	CDS end	Sequence complete	Expression confirmed by RNA-seq data (*A. carolinensis*)
EDSC2	NW_003338916.1	1040943	1040629	yes	yes
EDQM	NW_003338916.1	1036851	1037078	yes	yes
EDHEM	NW_003338916.1	995360	996070	yes	yes
EDCS1	NW_003338916.1	990209	990406	yes	yes
EDCS2	NW_003338916.1	982933	983241	yes	yes
EDCS3	NW_003338916.1	961905	962762	yes	yes
EDPCCC1	NW_003338916.1	951749	951994	yes	yes
EDPCCC2	NW_003338916.1	936661	936473	yes	yes
EDPCCC3	NW_003338916.1	<919157	919798	no	yes
EDGPC1	NW_003338916.1	893728	893540	yes	yes
EDGPC2	NW_003338916.1	876803	876600	yes	yes
EDYM1	NW_003338916.1	780003	780617	yes	yes
EDY1	NW_003338916.1	571522	572292	yes	yes
EDSRWM	NW_003338916.1	561996	560887	yes	yes
EDGY1	NW_003338916.1	556637	557008	yes	yes
EDGY2	NW_003338916.1	540103	539129	yes	yes
EDETM1	NW_003338916.1	305599	305339	yes	yes
EDETM2	NW_003338916.1	298118	297858	yes	yes
EDPCCC4	NW_003338916.1	238742	239044	yes	no
EDYM2	NW_003338916.1	184370	185365	yes	yes
EDP3	NW_003338916.1	129131	129376	yes	yes
EDQSG	NW_003338916.1	120578	120174	yes	yes
EDQK	NW_003338916.1	87721	87867	yes	yes
EDCATM	NW_003338916.1	57594	57070	yes	yes
EDPAML	NW_003338916.1	47259	46576	yes	yes

Notes - * other EDC genes have been reported in previous paper of our laboratories (27).

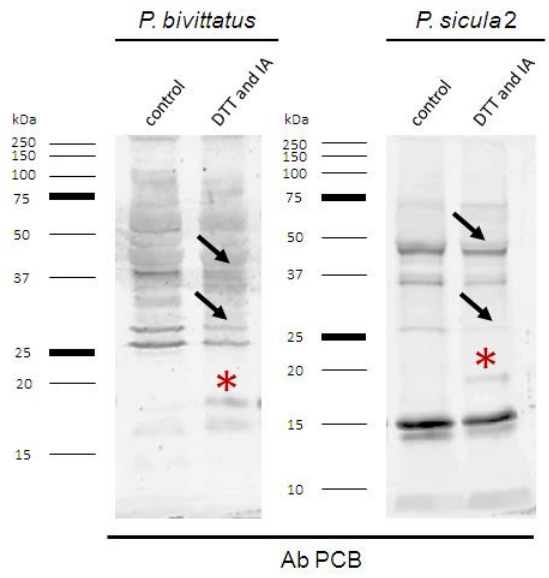
CDS, coding sequence; the symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Supplementary Data: Supplementary Figures and Tables

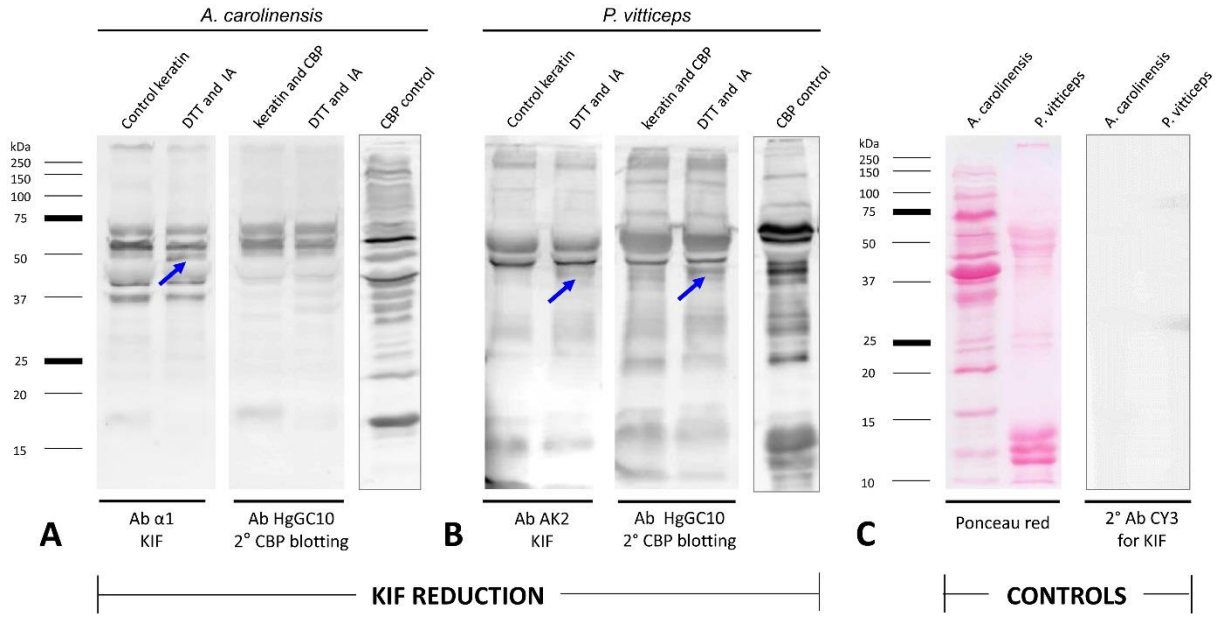
Disulfide-bond-mediated cross-linking of corneous beta-proteins in lepidosaurian epidermis

Karin Brigit Holthaus, Lorenzo Alibardi

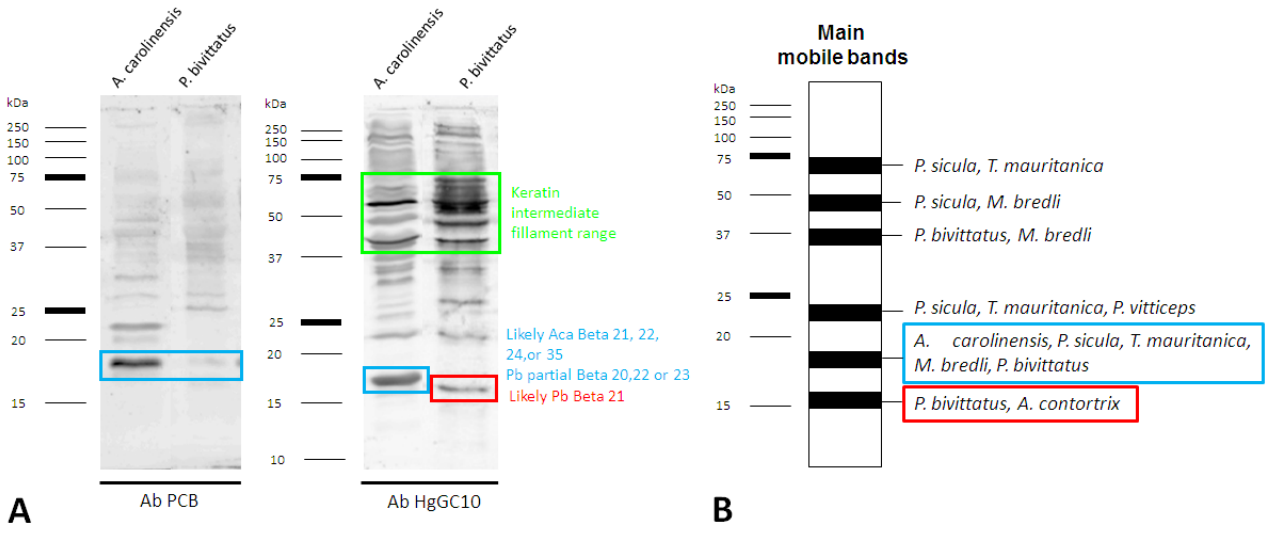
Content



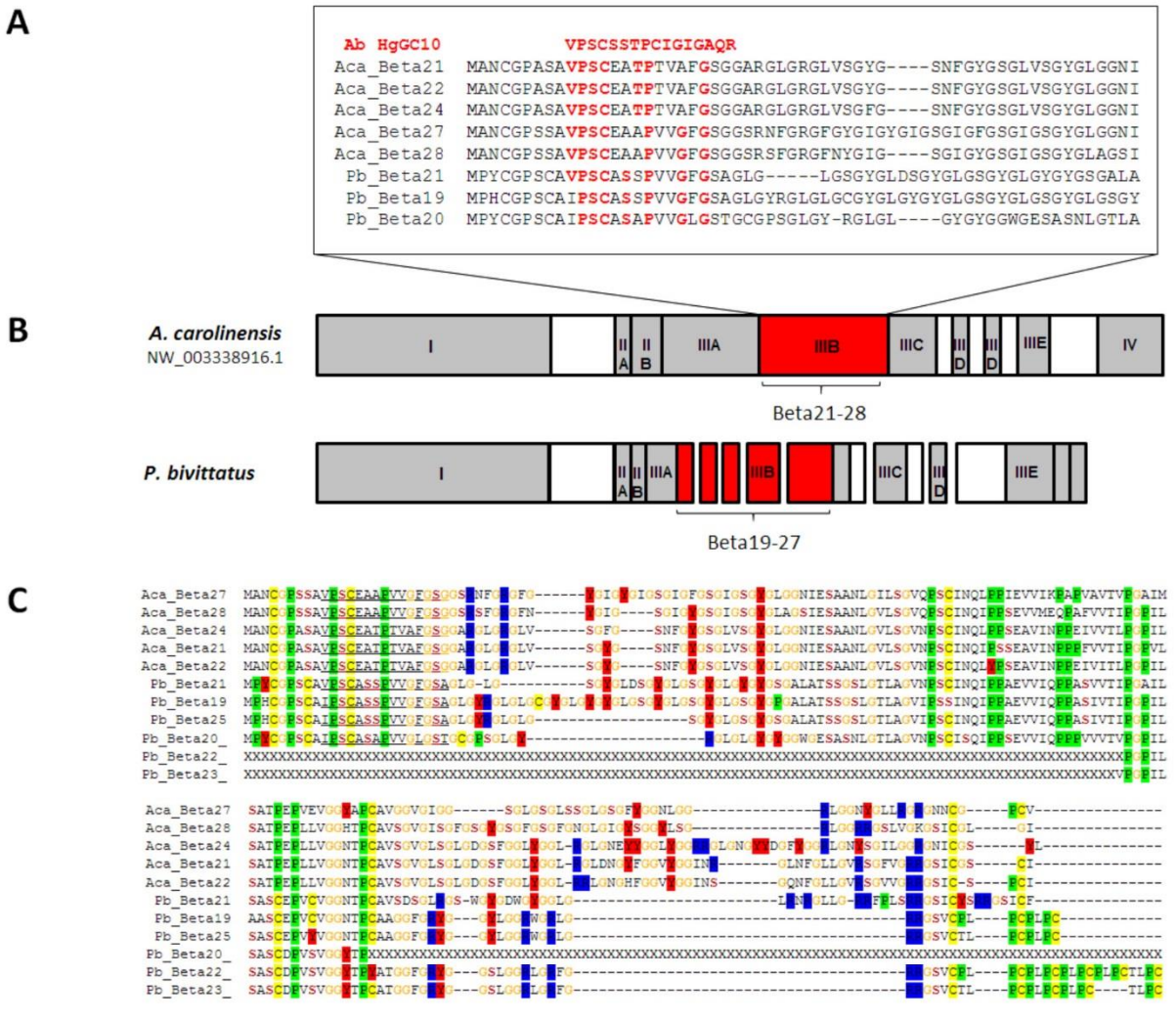
Suppl. Fig. S1. Using the PCB antibody two weak bands approximately around 18 and 19 kDa appeared while mainly a band around 37 kDa decreased in P. bivittatus. In P. sicula, only a slight decrease of bands around 43-46, 37 and 27-29 kDa was noted, a weak band appeared around 17-18 kDa. Red asterisks indicate modifications in the range of the 15-19 kDa CBP band. Black arrows indicate an intensification or weakening of bands above the CBP monomer size. Molecular weight is indicated by markers on the side (in kDa). A 12% polyacrylamide gel and nitrocellulose membrane were implied.



Suppl. Fig. S2. The effect of a reducing treatment with DTT followed by alkylation with IA on lizard skin samples using keratin intermediate filaments (KIF) Abs. ***A****) A. carolinensis with Ab α1 bands at approximately 37, 43-46, and 50, 55-70 kDa in both control and Reduced-Alkylated samples, with a slight increase in the approximately 50 kDa band and reduction in the approximately 55 kDa band in the latter. Using the HgGC10 antibody for CBPs on the same blot, the bands at approximately 37 and 42-43 kDa disappeared suggesting the latter antibody for CBPs recognizes only proteins present in the 55-70 kDa range* ***B****) P. vitticeps using Ab AK2 bands approximately at 50 and 55-70 kDa were present in both controls and reduced samples, and a low intensity band around 45-46 kDa appeared in the Reduced-Alkylated sample. Using the HgGC10 antibody for CBPs on the same blot, similar bands at approximately 50 and 55-65 kDa were seen in both control and reduced samples, but in the latter an intense band around 45-46 kDa was present.* ***C****) Control with Ponceau red and with secondary Ab CY3 used for the KIF marking. Controls for the beta Ab on "clean" membrane are shown at the right of blots. Blue arrows indicate reduced bands. A 12% polyacrylamide gel and nitrocellulose membrane were implied. The CY3 secondary Ab gives a less clean signal than CY5 and some background signal in the beta range is seen.*



Suppl. Fig. S3. ***A****) Western blot analysis of CBPs in skin protein extracts from the lizard A. carolinensis and snake P. bivittatus. The anti-CBP antibodies (Ab) PCB and HgGC10 were used.* ***B****) Schematic overview of observed mobile CBP bands with at the left the MW marker in kDa and at the right species where the band was observed. Green boxes mark size range of keratin intermediate filaments; blue and red boxes indicate main protein bands in the size range of CBPs. Candidate CBP that have predicted molecular weights of 17-18 kDa (blue) and 15-16 kDa (red) are indicated on the right.*



Suppl. Fig. S4. ***A****) Alignment of the HgGC10 epitope with the most likely protein candidates for WB results in A. carolinensis (Aca) and P. bivittatus (Pb). In red the epitope and identical amino acids in CBP sequences.* ***B****) Schematic representation of the CBP cluster in A. carolinensis and P. bivittatus. The CBP cluster is marked in grey, in red the sub cluster IIIB to which the likely candidates of WB results belong and in white non CBPs. The sub clusters are named with roman numbers and if requested a capital letter to further indicate clustering within a subcluster.* ***C****) Alignment of orthologous CBP sequences from sub cluster IIIB of A. carolinensis (Aca) and P. bivittatus (Pb).*

Suppl. Table S1

Corneous beta protein names of *Anolis carolinensis* used here and in previous pubblications.

Holthaus et al., 2017	Dalla Valle et al., 2010	Accession nr.
Beta-1	Li-Ac-40	NW_003338916.1
Beta-2	Li-Ac-39	NW_003338916.1
Beta-3	Li-Ac-37	NW_003338916.1
Beta-4	Li-Ac-36	NW_003338916.1
Beta-5	Li-Ac-35	NW_003338916.1
Beta-6	Li-Ac-34	NW_003338916.1
Beta-7	Li-Ac-33	NW_003338916.1
Beta-8	Li-Ac-32	NW_003338916.1
Beta-9	Li-Ac-31	NW_003338916.1
Beta-10	Li-Ac-30	NW_003338916.1
Beta-11	Li-Ac-29	NW_003338916.1
Beta-12	Li-Ac-28	NW_003338916.1
Beta-13	Li-Ac-27	NW_003338916.1
Beta-14	Li-Ac-26	NW_003338916.1
Beta-15	Li-Ac-25	NW_003338916.1
Beta-16	Li-Ac-24	NW_003338916.1
Beta-17	Li-Ac-23	NW_003338916.1
Beta-18	Li-Ac-22	NW_003338916.1
Beta-19	Li-Ac-21	NW_003338916.1
Beta-20	Li-Ac-20	NW_003338916.1
Beta-21	Li-Ac-19	NW_003338916.1
Beta-22	Li-Ac-18	NW_003338916.1
Beta-23	Li-Ac-17	NW_003338916.1
Beta-24	Li-Ac-16	NW_003338916.1
Beta-25	Li-Ac-15	NW_003338916.1
Beta-26	Li-Ac-14	NW_003338916.1
Beta-27	Li-Ac-13	NW_003338916.1
Beta-28	Li-Ac-12	NW_003338916.1
Beta-29	Li-Ac-11	NW_003338916.1
Beta-30	Li-Ac-10	NW_003338916.1
Beta-31	Li-Ac-9	NW_003338916.1
Beta-32	Li-Ac-8	NW_003338916.1
Beta-33	Li-Ac-7	NW_003338916.1
Beta-34	Li-Ac-6	NW_003338916.1
Beta-35	Li-Ac-5	NW_003338916.1
Beta-36	Li-Ac-4	NW_003338916.1
Beta-37	Li-Ac-3	NW_003338916.1
Beta-38	Li-Ac-2	NW_003338916.1
Beta-39	Li-Ac-1	NW_003338916.1
Beta-40	Li-Ac-38	NW_003341365.1

Note: Li, lizard and Ac, *A. carolinensis*

Suppl. Table S2

Compatibility of corneous beta protein candidates with WB results and with the beta epitope HgGC10 and keratin epitope α1.

CBPs	Compatibility WB results	Predicted MW (kDa)	Identity epitope HgGC10 (%)	Coverage epitope
Ac_Beta-21 (Li_Ac_19)	Yes	17,1	60	10|16
Ac_Beta-22 (Li_Ac_18)	Yes	17,9	60	10|16
Ac_Beta-24 (Li_Ac_16)	Yes	17,9	50	14|16
Ac_Beta-27 (Li_Ac_13)	Yes	15,8	50	14|16
Ac_Beta-28 (Li_Ac_12)	Yes	16,2	50	14|16
Ac_Beta-35 (Li_Ac_5)	Yes	16,6	58	14|16
Ac_Beta-15-Ab target	No	10,5	100	16|16
Ac_Beta-20-Ab target	No	11,5	100	16|16
Pb_Beta-19	Yes	15,1	50	14|16
Pb_Beta-21	Yes	16,1	50	14|16
Ac_K14-like-Ab target	Yes	50,7	100	23|23
Ac_K17-like X1	Yes	37,4	57	21|23
Ac_K17-like X2	Yes	46,2	57	21|23

Note: *A. carolinensis* (Ac) and *P. bivittatus* (Pb)

Supplementary Data Chpt 2.6: Supplementary Figures and Tables

Content

Suppl. Table S1

Tentative abbreviations and full names of EDC genes in crocodilians

Species	Gene name abbreviati	Full Gene Name
Asi	Crnn	Cornulin
Asi	EDAA1	Epidermal Differentiation protein rich in Aromatic Amino acids 1
Asi	EDAA2	Epidermal Differentiation protein rich in Aromatic Amino acids 2
Asi	EDAA3	Epidermal Differentiation protein rich in Aromatic Amino acids 3
Asi	EDAA4	Epidermal Differentiation protein rich in Aromatic Amino acids 4
Asi	EDAA5	Epidermal Differentiation protein rich in Aromatic Amino acids 5
Asi	EDAA6	Epidermal Differentiation protein rich in Aromatic Amino acids 6
Asi	EDAA7	Epidermal Differentiation protein rich in Aromatic Amino acids 7
Asi	EDAA8	Epidermal Differentiation protein rich in Aromatic Amino acids 8
Asi	EDAA9	Epidermal Differentiation protein rich in Aromatic Amino acids 9
Asi	EDAA10	Epidermal Differentiation protein rich in Aromatic Amino acids 10
Asi	EDAA11	Epidermal Differentiation protein rich in Aromatic Amino acids 11
Asi	EDAA12	Epidermal Differentiation protein rich in Aromatic Amino acids 12
Asi	EDAA13	Epidermal Differentiation protein rich in Aromatic Amino acids 13
Asi	EDAA14	Epidermal Differentiation protein rich in Aromatic Amino acids 14
Asi	EDAA15	Epidermal Differentiation protein rich in Aromatic Amino acids 15
Asi	EDAA16	Epidermal Differentiation protein rich in Aromatic Amino acids 16
Asi	EDAA17	Epidermal Differentiation protein rich in Aromatic Amino acids 17
Asi	EDAA18	Epidermal Differentiation protein rich in Aromatic Amino acids 18
Asi	EDAA21	Epidermal Differentiation protein rich in Aromatic Amino acids 21
Asi	EDAA22	Epidermal Differentiation protein rich in Aromatic Amino acids 22
Asi	EDAA23	Epidermal Differentiation protein rich in Aromatic Amino acids 23
Asi	EDAA24	Epidermal Differentiation protein rich in Aromatic Amino acids 24
Asi	EDAA25	Epidermal Differentiation protein rich in Aromatic Amino acids 25
Asi	EDAA26	Epidermal Differentiation protein rich in Aromatic Amino acids 26
Asi	EDAA27	Epidermal Differentiation protein rich in Aromatic Amino acids 27
Asi	EDAA28	Epidermal Differentiation protein rich in Aromatic Amino acids 28
Asi	EDAA29	Epidermal Differentiation protein rich in Aromatic Amino acids 29
Asi	EDAA30	Epidermal Differentiation protein rich in Aromatic Amino acids 30
Asi	EDAA31	Epidermal Differentiation protein rich in Aromatic Amino acids 31
Asi	EDAA32	Epidermal Differentiation protein rich in Aromatic Amino acids 32
Asi	EDAA33	Epidermal Differentiation protein rich in Aromatic Amino acids 33
Asi	EDAA34	Epidermal Differentiation protein rich in Aromatic Amino acids 34
Asi	EDAA35	Epidermal Differentiation protein rich in Aromatic Amino acids 35
Asi	EDAA36	Epidermal Differentiation protein rich in Aromatic Amino acids 36
Asi	EDAA37	Epidermal Differentiation protein rich in Aromatic Amino acids 37
Asi	EDAA38	Epidermal Differentiation protein rich in Aromatic Amino acids 38
Asi	EDAA39	Epidermal Differentiation protein rich in Aromatic Amino acids 39
Asi	EDAA40	Epidermal Differentiation protein rich in Aromatic Amino acids 40
Asi	EDAA41	Epidermal Differentiation protein rich in Aromatic Amino acids 41
Asi	EDC1	Epidermal Differentiation protein rich in Cysteine (C) 1
Asi	EDC2	Epidermal Differentiation protein rich in Cysteine (C) 2
Asi	EDC3	Epidermal Differentiation protein rich in Cysteine (C) 3
Asi	EDC4	Epidermal Differentiation protein rich in Cysteine (C) 4
Asi	EDCH1	Epidermal Differentiation protein containing Cysteine Histidine motifs 1
Asi	EDCH2	Epidermal Differentiation protein containing Cysteine Histidine motifs 2
Asi	EDCH3	Epidermal Differentiation protein containing Cysteine Histidine motifs 3
Asi	EDCH4	Epidermal Differentiation protein containing Cysteine Histidine motifs 4
Asi	EDCH5	Epidermal Differentiation protein containing Cysteine Histidine motifs 5
Asi	EDCH6	Epidermal Differentiation protein containing Cysteine Histidine motifs 6
Asi	EDCH7	Epidermal Differentiation protein containing Cysteine Histidine motifs 7
Asi	EDCH8	Epidermal Differentiation protein containing Cysteine Histidine motifs 8
Asi	EDCH9	Epidermal Differentiation protein containing Cysteine Histidine motifs 9
Asi	EDCH10	Epidermal Differentiation protein containing Cysteine Histidine motifs 10
Asi	EDCH11	Epidermal Differentiation protein containing Cysteine Histidine motifs 11
Asi	EDCH12	Epidermal Differentiation protein containing Cysteine Histidine motifs 12
Asi	EDCH13	Epidermal Differentiation protein containing Cysteine Histidine motifs 13
Asi	EDCH14	Epidermal Differentiation protein containing Cysteine Histidine motifs 14
Asi	EDCH15	Epidermal Differentiation protein containing Cysteine Histidine motifs 15
Asi	EDCH16	Epidermal Differentiation protein containing Cysteine Histidine motifs 16

Asi	EDCH17	Epidermal Differentiation protein containing Cysteine Histidine motifs 17
Asi	EDCH18	Epidermal Differentiation protein containing Cysteine Histidine motifs 18
Asi	EDCH19	Epidermal Differentiation protein containing Cysteine Histidine motifs 19
Asi	EDCH20	Epidermal Differentiation protein containing Cysteine Histidine motifs 20
Asi	EDCH21	Epidermal Differentiation protein containing Cysteine Histidine motifs 21
Asi	EDCH22	Epidermal Differentiation protein containing Cysteine Histidine motifs 22
Asi	EDDM1	Epidermal Differentiation protein containing DPCC Motifs 1
Asi	EDDM2	Epidermal Differentiation protein containing DPCC Motifs 2
Asi	EDKM	Epidermal Differentiation protein containing a KKLIQQ Motif
Asi	EDP1	Epidermal Differentiation protein rich in Proline 1
Asi	EDP2	Epidermal Differentiation protein rich in Proline 2
Asi	EDP3	Epidermal Differentiation protein rich in Proline 3
Asi	EDPCV	Epidermal Differentiation protein rich in Proline (P), Cysteine (C) and Valine (V)
Asi	EDPCQ	Epidermal Differentiation protein rich in Proline (P), Cysteine (C) and glutamine (Q)
Asi	EDPE	Epidermal Differentiation protein rich in Proline and glutamic acid (E)
Asi	EDPL1	Epidermal Differentiation Proline-rich protein, close to Loricrin, 1
Asi	EDPQ1	Epidermal Differentiation protein rich in Proline and glutamine (Q) 1
Asi	EDPQ2	Epidermal Differentiation protein rich in Proline and glutamine (Q) 2
Asi	EDPQ3	Epidermal Differentiation protein rich in Proline and glutamine (Q) 3
Asi	EDPQ4	Epidermal Differentiation protein rich in Proline and glutamine (Q) 4
Asi	EDPQ5	Epidermal Differentiation protein rich in Proline and glutamine (Q) 5
Asi	EDRYA	Epidermal Differentiation protein containing a RYA terminus
Asi	EDQL	Epidermal Differentiation protein rich in glutamine (Q), close to Loricrin
Asi	EDQM1	Epidermal Differentiation protein containing a glutamine (Q) Motif 1
Asi	EDQM2	Epidermal Differentiation protein containing a glutamine (Q) Motif 2
Asi	EDWM	Epidermal Differentiation protein containing a WYDP Motif
Asi	EDYM1	Epidermal Differentiation protein containing Y Motif 1
Asi	Lor	Loricrin
Asi	Pglyrp3	Peptidoglycan recognition protein 3
Asi	Scfn	Scaffoldin

NOTES - Asi, *Alligator sinensis* ; Genes of the main beta-keratin cluster and S100A genes are not included here.

Suppl. Table S2
Chinese alligator *(Alligator sinensis)* EDC genes (other than corneous beta protein genes)

Gene	Accession nr.	CDS start	CDS end	Sequence complete	Expression confirmed by RNA-seq data
S100-A9	NW_005842477.1	1145232	1147151	yes	yes
PGLYRP3	NW_005842477.1	1169577	1172879	yes	yes
EDKM	NW_005842477.1	1177140	1177832	yes	yes
EDQM1	NW_005842477.1	1184449	1183881	no	n.a.
EDQM2	NW_005842477.1	1193599	1193236	no	yes
EDWM	NW_005842477.1	1203609	1202815	yes	yes
EDRYA	NW_005842477.1	1208055	1207771	yes	yes
EDC1	NW_005842477.1	1211616	1211338	yes	no
EDC2	NW_005842477.1	1214928	1214671	yes	yes
EDC3	NW_005842477.1	1218546	1218277	yes	no
EDC4	NW_005842477.1	1221558	1221827	yes	yes
EDCH1	NW_005842477.1	1225947	1225675	yes	yes
EDCH2	NW_005842477.1	1229984	1229763	yes	yes
EDCH3	NW_005842477.1	1237563	1237312	yes	yes
EDCH4	NW_005842870.1	580904	580653	yes	yes
EDCH5	NW_005842870.1	585629	585420	yes	yes
EDCH6	NW_005842870.1	599749	599414	yes	yes
EDCH7	NW_005842870.1	604640	604305	yes	yes
EDCH8	NW_005842870.1	609200	609524	no	yes
EDCH9	NW_005842870.1	619754	619453	no	yes
EDCH10	NW_005844546.1	3361	3003	no	yes
EDCH11	NW_005843704.1	59713	>59898	no	yes
EDCH12	NW_005843704.1	55726	55977	no	yes
EDCH13	NW_005843704.1	52814	52587	no	yes
EDCH14	NW_005843704.1	49501	49728	no	yes
EDCH15	NW_005843704.1	47207	46962	yes	yes
EDCH16	NW_005843704.1	42440	42664	yes	yes
EDCH17	NW_005843704.1	39297	39103	yes	yes
EDCH18	NW_005843704.1	24401	24141	yes	no
EDCH19	NW_005843704.1	18101	18361	yes	no
EDCH20	NW_005843704.1	11489	11268	no	no
EDCH21	NW_005843704.1	<7754	7990	no	yes
EDCH22	NW_005843425.1	770	444	yes	yes
EDQL	NW_005843425.1	18736	18485	yes	yes
LOR	NW_005843425.1	29177	26733	yes	no
EDPL	NW_005843425.1	52261	52076	yes	yes
EDYM1	NW_005843425.1	55098	54616	yes	yes
EDAA1	NW_005843425.1	84974	84705	yes	yes
EDAA2	NW_005843425.1	90565	90906	yes	yes
EDAA3	NW_005843425.1	111579	111848	yes	yes
EDAA4	NW_005843425.1	124193	124534	yes	yes
EDAA5	NW_005843425.1	136224	136463	yes	yes
EDAA6	NW_005843425.1	140780	140583	yes	no
EDAA7	NW_005843425.1	147065	>147124	no	yes

EDAA8	NW_005843981.1	10631	10470	yes	yes
EDAA9	NW_005843981.1	16492	16301	yes	yes
EDAA10	NW_005843981.1	20879	21076	yes	yes
EDAA11	NW_005843927.1	16210	16401	yes	yes
EDAA12	NW_005843927.1	26472	26281	yes	yes
EDAA13	NW_005844332.1	5706	5897	yes	n.a.
EDAA14	NW_005843574.1	6507	6698	yes	yes
EDAA15	NW_005843574.1	17750	17553	yes	yes
EDAA16	NW_005843574.1	48584	48417	yes	yes
EDAA17	NW_005843574.1	60961	60764	yes	yes
EDAA18	NW_005843574.1	68008	67841	yes	yes
EDAA19	NW_005843574.1	73107	73298	yes	yes
EDAA20	NW_005843574.1	77958	78125	yes	yes
EDAA21	NW_005844741.1	2326	2481	yes	n.a.
EDAA22	NW_005843835.1	3825	4004	yes	n.a.
EDAA23	NW_005843835.1	12420	12229	yes	n.a.
EDAA24	NW_005845569.1	535	338	yes	n.a.
EDAA25	NW_005844011.1	7276	7079	yes	yes
EDAA26	NW_005844011.1	12653	12844	yes	yes
EDAA27	NW_005844011.1	20753	20556	yes	yes
EDAA28	NW_005844749.1	563	754	yes	yes
EDAA29	NW_005844209.1	418	609	yes	n.a.
EDAA30	NW_005844209.1	6158	6313	yes	n.a.
EDAA31	NW_005844434.1	5164	5358	yes	n.a.
EDAA32	NW_005844013.1	4582	4776	yes	no
EDAA33	NW_005844369.1	2281	2117	yes	yes
EDAA34	NW_005843959.1	5272	5508	yes	yes
EDAA35	NW_005843488.1	26363	26085	yes	yes
EDAA36	NW_005843488.1	30610	30807	yes	no
EDAA37	NW_005843488.1	37639	37875	yes	yes
EDAA38	NW_005843488.1	71905	71708	yes	no
EDAA39	NW_005843488.1	77357	77635	yes	yes
EDAA40	NW_005843488.1	95540	95731	yes	yes
EDAA41	NW_005843488.1	106922	106764	yes	no
EDPCQ	NW_005842911.1	417543	418958	yes	yes
EDP1	NW_005842911.1	409961	410779	yes	yes
EDP2	NW_005842911.1	391439	391792	yes	yes
EDPE	NW_005842911.1	381471	380539	yes	yes
EDP3	NW_005842911.1	371479	371697	yes	no
EDPQ1	NW_005842911.1	366995	366747	yes	yes
EDPQ2	NW_005842911.1	358795	359001	yes	no
EDPQ3	NW_005842911.1	341777	342013	yes	yes
EDPQ4	NW_005842911.1	339046	338840	yes	yes
EDPQ5	NW_005842911.1	334654	334890	yes	no
EDPQ6	NW_005842911.1	331432	331226	yes	yes
EDPCV	NW_005842911.1	327146	327553	yes	yes
EDDM2	NW_005842911.1	317953	316673	yes	no
EDDM1	NW_005842911.1	310213	311412	yes	yes
CRNN	NW_005842911.1	289193	285697	no	yes

SCFN	NW_005842911.1	276363	271612	yes	no
S100A11	NW_005842911.1	26699	260312	yes	yes

Notes - CDS, coding sequence; n.d., not determined; n.a. not applicable

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S3
Chinese alligator *(Alligator sinensis)* corneous beta protein (beta-keratin) genes

Gene	Accession nr.	CDS start	CDS end	Sequence complete	Expression confirmed by RNA-seq data
Beta1	NW_005843425.1	61017	60622	yes	no
EDBeta	NW_005843425.1	66835	67149	yes	yes
Beta2	NW_005843776.1	14927	15322	yes	yes
Beta3	NW_005843776.1	38892	39287	yes	yes
Beta4	NW_005843293.1	6303	6698	yes	yes
Beta5	NW_005843293.1	20473	20928	yes	yes
Beta6	NW_005843293.1	38862	39252	no	yes
Beta7	NW_005843293.1	50698	51093	yes	yes
Beta8	NW_005843293.1	57220	56816	yes	yes
Beta9	NW_005843293.1	81993	81634	yes	yes
Beta10	NW_005843293.1	98267	97782	yes	yes
Beta11	NW_005843293.1	106917	107294	yes	yes
Beta12	NW_005843293.1	128505	127948	yes	yes
Beta13	NW_005843293.1	134659	135228	yes	yes
Beta14	NW_005843293.1	141781	141224	yes	yes
Beta15	NW_005843293.1	159177	158686	yes	no
Beta16	NW_005843293.1	164860	165426	yes	yes
Beta17	NW_005843293.1	>195188	194952	no	yes
Beta18	NW_005843293.1	203180	202836	yes	yes
Beta19	NW_005844928.1	970	1539	yes	yes
Beta20	NW_005843827.1	9986	<9816	no	n.a.
Beta21	NW_005843827.1	21239	<20973	no	n.a.
Beta22	NW_005843827.1	27575	27246	yes	n.a.
Beta23	NW_005843827.1	35668	35213	yes	n.a.
Beta24	NW_005844022.1	479	141	yes	yes
Beta25	NW_005844022.1	5518	5856	yes	yes
Beta26	NW_005844022.1	9390	9052	yes	yes
Beta27	NW_005844022.1	17261	16923	yes	yes
Beta28	NW_005842911.1	582806	582468	yes	yes
Beta29	NW_005842911.1	575088	575426	yes	yes
Beta30	NW_005842911.1	571820	572158	yes	yes
Beta31	NW_005842911.1	567179	566859	yes	yes
Beta32	NW_005842911.1	554888	555226	yes	yes
Beta33	NW_005842911.1	550148	549810	yes	yes
Beta34	NW_005842911.1	543674	544258	yes	yes
Beta35	NW_005842911.1	506800	507279	yes	yes
Beta36	NW_005842911.1	502103	501663	yes	yes
Beta37	NW_005842911.1	496471	497001	yes	yes
Beta38	NW_005842911.1	492176	491685	yes	yes
Beta39	NW_005842911.1	479390	479022	yes	no
Beta40	NW_005842911.1	470088	470714	yes	yes
Beta41	NW_005842911.1	457868	458509	yes	yes
Beta42	NW_005842911.1	447145	448434	yes	yes

Beta43	NW_005842911.1	430348	430953	yes	yes

Notes - CDS, coding sequence; n.d., not determined; n.a. not applicable

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S4

Australian saltwater crocodile *(Crocodylus porosus)* EDC genes (other than corneous beta protein genes)

Gene	Accession nr.	CDS start	CDS end	Sequence complete
S100-A9	MDVP01000026.1	<243744	245030	no
PGLYRP3_part1	JRXG01058286.1	2179	<232	yes
PGLYRP3_part2	JRXG01058285.1	>626	450	yes
EDKM	JRXG01058284.1	952	252	yes
EDQM1	JRXG01006413.1	5178	5645	yes
EDQM2	MDVP01000026.1	290589	290200	no
EDWM	MDVP01000026.1	299694	298879	yes
EDRYA	MDVP01000026.1	305213	304929	yes
EDC1	MDVP01000026.1	308751	308485	yes
EDCH1	MDVP01000026.1	318243	317992	yes
EDCH2	MDVP01000026.1	322239	321967	yes
EDCH3	MDVP01000026.1	328285	328034	yes
EDCH4	MDVP01000026.1	332130	331879	yes
EDCH5	MDVP01000026.1	335663	335391	yes
EDCH6	MDVP01000026.1	>339228	339166	no
EDCH7	JRXG01099277.1	1279	1539	yes
EDCH8	JRXG01099277.1	8174	7932	yes
EDCH9	JRXG01013561.1	<1	175	no
EDCH10	JRXG01039386.1	5323	5090	yes
EDCH11	JRXG01044410.1	575	808	yes
EDCH12	JRXG01044410.1	3294	3058	yes
EDCH13	JRXG01044410.1	6192	6464	yes
EDCH14	JRXG01044410.1	8965	>9174	no
EDCH15	MDVP01000026.1	344777	344499	yes
EDQL	MDVP01000026.1	353105	352866	yes
LOR	MDVP01000026.1	363815	362264	no
EDPL	MDVP01000026.1	385663	385451	yes
EDYM1	MDVP01000026.1	388477	387995	yes
EDAA1	JRXG01066409.1	1710	1901	yes
EDAA2	JRXG01066410.1	3519	3328	yes
EDAA3	JRXG01076834.1	4212	4397	yes
EDAA4	JRXG01063821.1	2219	2046	yes
EDAA5	JRXG01104627.1	4938	4741	yes
EDAA6	JRXG01076832.1	4468	4758	yes
EDAA7	JRXG01076832.1	10472	10227	yes
EDAA8	JRXG01076832.1	17367	17561	yes
EDAA9	JRXG01076833.1	5689	5492	yes
EDAA10	JRXG01069781.1	2547	2783	yes
EDAA11	JRXG01060709.1	13537	13301	yes
EDAA12	JRXG01060709.1	4202	4387	yes
EDAA13	JRXG01060708.1	17706	17428	yes
EDAA14	JRXG01060708.1	2046	2225	yes
EDAA15	JRXG01060708.1	22931	23128	yes
EDAA16	JRXG01104928.1	2430	>2603	no

EDAA17	JRXG01076836.1	>15917	15705	no
EDAA18	JRXG01099001.1	128	<1	no
EDPCQ	MDVP01000026.1	653561	652116	yes
EDP1	MDVP01000026.1	662837	661880	no
EDP2	MDVP01000026.1	681392	681024	yes
EDPE	MDVP01000026.1	691240	692343	yes
EDP3	MDVP01000026.1	699606	699430	yes
EDPQ1	MDVP01000026.1	704196	704399	yes
EDPQ2	MDVP01000026.1	709497	709703	yes
EDDM2	MDVP01000026.1	721051	722289	yes
EDDM1	MDVP01000026.1	728952	727573	no
CRNN	MDVP01000026.1	749469	752090	yes
SCFN	MDVP01000026.1	762021	770451	yes
S100A11	MDVP01000026.1	<776095	776844	no

Notes - CDS, coding sequence; n.d., not determined; n.a. not applicable

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

Suppl. Table S5

Australian saltwater crocodile *(Crocodylus porosus)* corneous beta protein (beta-keratin) genes

Gene	Accession nr.	CDS start	CDS end	Sequence complete
Beta1	MDVP01000026.1	394371	393976	yes
EDBeta	MDVP01000026.1	400062	400376	yes
Beta2	MDVP01000026.1	425383	425778	yes
Beta3	MDVP01000026.1	430625	430230	yes
Beta4	MDVP01000026.1	454438	454079	yes
Beta5	JRXG01074025.1	3021	2515	yes
Beta6	JRXG01074025.1	16939	16616	yes
Beta7	JRXG01074025.1	28647	28060	yes
Beta8	JRXG01110871.1	14074	14595	yes
Beta9	JRXG01110870.1	10001	9576	yes
Beta10	JRXG01110870.1	15546	>15716	no
Beta11	JRXG01110870.1	<1	271	no
Beta12	JRXG01110869.1	<1	225	no
Beta13	JRXG01004993.1	2781	2437	yes
Beta14	JRXG01110864.1	1952	2296	yes
Beta15	JRXG01110865.1	2994	3338	yes
Beta16	JRXG01004994.1	5855	5400	yes
Beta17	JRXG01004994.1	>12153	12043	no
Beta18	JRXG01086413.1	2155	1736	yes
Beta19	JRXG01011795.1	3986	4324	yes
Beta20	JRXG01011795.1	10908	11246	yes
Beta21	JRXG01011795.1	17328	16774	yes
Beta22	JRXG01011795.1	38823	38389	yes
Beta23	MDVP01000026.1	583004	582435	yes
Beta24	MDVP01000026.1	587243	587755	yes
Beta25	MDVP01000026.1	590427	590167	yes
Beta26	MDVP01000026.1	595099	595419	yes
Beta27	MDVP01000026.1	600837	<600541	no
Beta28	MDVP01000026.1	605656	<605360	no
Beta29	MDVP01000026.1	611827	611207	yes
Beta30	MDVP01000026.1	625170	623850	yes
Beta31	MDVP01000026.1	641121	640585	yes

Notes - CDS, coding sequence; n.d., not determined; n.a. not applicable

The symbols < and > indicate that ends of the coding sequence were not present on the scaffold.

A

>Am_CRNN
MTQLQGNIEGIVSAFNAYAKEDGGCITLSKGELRQLIQQEFADVLVKPHDLQTIDQVLQRLDVEREDRIDFDEFLVLVFQVAK
ACHKKLSPCQPSGDGQGSAAQGDASRDQAQRADQEQGQKQPDAPEQDPTRPQAPETRTAEGDLSRRHTQDPEVSGTQDPEVSQ
GDGNHEAQAAKTPEHDSIRRQGQEPEQDPSHHRAQEQDPNREGQDPQVPQQDVKHEALESGAPEQAPNRHPVLQPSVSERDLD
HSHSSASERDLDHHPTLQPNTLETDLDRHPTLQPNTLETDLDRLPTLQPSTSETDLDHHQGLETEAPEQDLNSTTGRESTHNE
AHELQAPEQESQESQPDEQDQNRHQSEEPEASEHTLHHQPQEAQPTEQDLTWETQTPRVPEGDVSRGGTPFSPALQQDRGAQQ
DPREEALTAYRPYIYQCQKPPTFPYQWLPKQ

>Am_EDAA1_partial
MFDSLDTTEDLFYQGCWDPCYRRPYWNSCWDPCTYRRPXXXGGCYPYSSRWGRRYSYGNCWPC

>Am_EDAA2
MFDSLDVIEDLHYPGQSDCFPHQRRRPPYTCCCYDQWGRLVWRGCCWSAPPWWCRRGSSGNGWPC

>Am_EDAA3
MFDSLETLYDLFYQGQSDCWPPTPRRTPYTCCCYDRCGRLVWRGCCWSIPPW

>Am_EDAA4
MFDSLDMIEDLSYPGQSDCWPPFPRRPSYTCCCYDRCGRLVWYGCCWSIPPWGPIQTALEITGHAKSSKNIHGTGTMRTSKPI
YIFMADPKDMACTSAKHSTPLL

>Am_EDAA5
MSDSLDMIEDLYYQNPSCCWRPPCRRRCWCCCYDPCTGQLIWQGWCWCPGWRSRGRYGRCWPC

>Am_EDAA6
MFDSLHTFEDLFYLGQSDFRWPPYHQRQYLCCCYDRYECLVWQGPCCYPGPWCRR

>Am_EDAA7
MSDSLDMLENFYYPGQSNCWDPCYRRPYWNSWWDPCTYRRPSWYGCWDPCTYRRPYIYDNCYGYGGLYGLGGCYPYSSRWGRR
GSWGSCWPC

>Am_EDAA8
MSDSLDMLEELYYQSPGCCWRPRRRRPCYCCCYDPCTGELVWEGWCCCPWWGGRGRYGRRWPC

>Am_EDAA9_partial
MSDSLDMLEDLWYPGQSNCWXXXRRPYWNSCWDPYTYRRPSWSGCWNPCTYRKPYIYNRSYGYGGLYGAGGCFPYSTRWGRRY
SAGSCWPC

>Am_EDAA10_partial
MFDSLDMIEDLSYQXXX

>Am_EDAA11
MTYHQSGCDDVCYTHCSYGVLYGYHGLTACWEPWTYGRPYSYGCCNPCTYRWPNRYWEPCGYGIG

>Am_EDC1
MCSCCSGCHGTRSVQPICYVQPVCCEPVYIHRSSGSCCQPCGSCCGSCCRGSRACPWVVIQRRPMPVCCPPLQYSAPMQQHCS
PLKKC

>Am_EDC2
MCSCCSGCHGTETICYVQPVCCEPVYIHRSSGSCCQPCGSCCGSCCGSCCWGLRSCPQVVIQRRPMPVCCPPLQYSSPMQQCC
SPLKKC

>Am_EDC3
MCSCCSGCHGTDETICYVQPVCCEPVYIQRSLESCCQPCGSCCGSCCGGSRSCPRVVIQRRPMPVCCPPLQYSAPMHQCCLPL
KKC

>Am_EDC4
MCSCCSGCHGTESVQPICCEPVYIQRSSGSCCQPCGSCCGSCCGGSRPFPRVVIQRRPMPVCCPPLQYSAPMRKYSAPMQQCC
PPLKKC

>Am_EDCH1
MCSRRSCHDHGSSSHGCHRHESSCHGSSSSINCVIEKPVPICPMPQCCPPIQQCCPPVQKCCPPVKCCQQNQQCCKFPPQYPK

>Am_EDCH2
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSINCIIEKPVPICPVPQCCPQLPQCCVPVQPCCPPMDKCCPPVQCCQQSKQCCK
IPPPCPK

>Am_EDCH3
MCSRRSCHDHGSSSHGCHSHESSCHGSSSSINCVIEEPVPVCPVPQCCPQLPQCCVPVQQCCPPMQCCQQTKQCCKIPPQCPK

>Am_EDCH4
MCSRRSCHDHGSSSHGCHSHESSCHGSSSSIPCIIEKPVPICPVQPCCPPVQQCCPPVQKCCPPVKCCQQSKQCCKFPPQCPK

>Am_EDCH5
MCSRRSCHDHGSSSHGCHGHESSCHSSSSSIPCIIEKPVPVCPVQPCCPPVQQCCPPMQKCCPPVKCCQQTKQCCKFPPQCPK

>Am_EDCH6
MCSRRSCHNHGSSSHGCHGHESSCHSSSSSIPCIIEKPVPVCPVQPCCPPVQQCCPPMQKCCPPVKCCQQSKQCCKFPPQCPK

>Am_EDCH7
MCSRRSCHNHGSSSHGCHGHESSCHSSSSSIPCIIEKPVPVCPVQPCCPPVQQCCPPVQKCCPPVQQCCPPMQKCCPPVKCCQ
QTKQCCKFPPQCPK

>Am_EDCH8
MCSRRSCHDHGASSHGCHGHESSCHSSSSSIPCILEKPVPVCTVQPCCPPVQQCCPPVQKCCPPVKCCQQSKQCCKFPPQCPK

>Am_EDCH9
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSIPCIIEKPVPICPMQPCCPPVQQCCPPVQKCCPPVQQCCPPMQKCCPPVKCCQ
QSKQCCKFPPQCPKSHSDEEDTKTCPLHSSPWDGGLPA

>Am_EDCH10
MCSRRSCHDHGSSSHGCHGHESSCHSSSSSIPCIIEKPVPVCPVQPCCPPVQQCCPPVQKCCPPVQQCCPPVQKCCPPVKCCQ
QTKQCCKFPPQCPK

>Am_EDCH11
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSIPCIIEKPVPVCPMQPCCPPVQQCCPPVQKCCPPVQQWCPPMQKCCPPVKCCQ
QTKQCCKFPPQCPK

>Am_EDCH12
MCSRRSCHDHGSSSHGCHGNESSCHDSSSSINCVIEKPVPICPMPQCCPPVQQYCPPVQQCCPPVQQCCPPVQCCQQSQQCCK
IPPQFPK

>Am_EDCH13
MCSHGSCHNRHRSCHGSSSHCHESRPSCNIVVVEKPYCCPVPRYYPPVSCCYPRYQYSQCCKFPQYPKCPPQYPK

>Am_EDCH14_partial
XXXGSSSHCHESRPSCNIVVVEKPYCCPVPRYCPPVSCCYPRYQYSQCCKFPQYPKYPQYPK

>Am_EDCH15_partial
XXXGSSSHCHESRPSCNIVVVEKPYCCPVPRYCPPVSCCYPRYQYSQCCKFPQYPKYPQYPK

>Am_EDCH16_partial
XXXGSSSHCHESRPSCNIVVVEKPYVQACCPVPSYCPPVSCCYPRYQYSQCCKFPQYPKCPPQYPK

>Am_EDCH17_partial
XXXGSSSHCHESRPSCNIVVVEKPYVQACCPVPSYCPPVSCCYPRYQYSQCCKFPQYPKCPPQYPK

>Am_EDCH18
MCSRGSCHDHHNTCHRSQCGSHCHESRPSCNINEVERLSVQSWCPVQQYCPVQHYCPPTCCYPRYQYSQCCKFPPEYPKCPPQ
YPK

>Am_EDCH19
MCSRGSCHDHHSSSHGSRCRSRCHESRPSCNITVVERPSIQSWCPVQQYCPVQHYCPPTCCYPRYQYSQCCKFPPQYPKCPLQ
YPKCPLQYPK

>Am_EDCH20
MCSRGSCHDLHSSCHRSRRGSHCHESRPSCNITVVERPSVQSWCPVQQYCPVQPYCPPTCCYPRYQDSQCCKFPPQYPKCPPQ
YPKCPPQYPKCPPQYPK

>Am_EDCH21
MCSRRSCHNHHSSCHRSRYGSHCHESRPSCNITVVERPFMQSWCPVQQYCPMQRYCPPTCCYPCYQYSQCCKFPPQYPKCPPQ
YPK

>Am_EDCH22
MCSRGSCHDHHNTCHRSRCGSHCHESRPSCNITVVEWPSVKSWCPMQQYCPVQRYCPPTCCYPRYQYSQYPKCPPQYPK

>Am_EDCH23

MCSRGSCHNHHNSCHRSRRGSHCHEYRPSCNITVVDRPSMQSWCPMQQYCPTQRCCPPTCCSPRYQYSQCYKFPPQYPKCPPQ
YPK

>Am_EDCH24
MCSCRSCHDHHSSCHGSRCGSHCHESILYCNIMVVERPSLLTWCPMQQYCPVQLYCPPTCCYLRYQYSQCCQCPLQYRKCPPP
YPKYPPQYPKCPPQYPK

>Am_EDCH25
MCSRRRCHDRETSCHDSRPSCHSSGSRCHESRPSCHVSVIERPPYCQWQQYRPVQPYYPGYQYSQCCKFPQYPKYPQYPK

>Am_EDDM1_partial
MYYPQQHQCKQPCLPPPIVKNCPLQTVEPCGTVHVSQCATRCVDTCGTLSVAQAASGSLSPCGGISEAQAMGTSLSPGGGISE
AQAASKSVDPRGGVSVAQAGGTSLCPCGGVCVACAASRSVGPYGGVSVAQAAXXXSASQSVGPYGGVSVAQAASKSVDPRGGA
SMAQAAGQSVDPCGDASMAQATGQSVDPYAGVSVAQAASKSVDPRGGASMAQATGQSVDPYAGVSVAQAATKCIEPGAKGYVR
PYAIQCADVCRSKCVASYGQVKVDXCAPGCAKTYPLQNVDPCLPKGSPVQQCSRKSKQC

>Am_EDDM2
MAFPNQQQYKQPCLPPLVCIQKCPPRCVDQCDAACVKKHTDLHGNICAKSCTTKCVDSCDGISTMLCMTKCMDPCGAACVKEC
TTKCMCPSNTVCAKPCVTKYVDPCGTSCVMSCATPCLEPCNTICVKECVTKCMDPCGTFCAEPYVTKYGDPGCSNSAKPCITK
CVDSCNTVCVKECTTKCVDLCSTRYAKPCVTNCVDSCGTGCGKLCVTKCVDPCGTGCGKLCVTKCMDPCGTMCTKPCLTMCMN
PCSTRCAKPSVTKCVNLCGTMCGKPCITKHEDPCGTICVKECTTKCMDPCGIICTKPCVTKCVDPCTTSCVTSCVTKSTESCS
TVCIKKCTVKCVDTCSTICAKPFVPKCTDPCCPRCTASSGTTSMDPCAPVCKKTYPLQIVDLRLSKCPFVQQCCQKPKQC

>Am_EDKM
MSKLLKAITNLIDSNQGNSRKEGKAEMFCRSEFKKLVQQDLAPIRLSPSYRYRHIKSLPESETEPVNHKKISTVKHCVY

>Am_EDP1_partial
MSYYGQQCKQRCLPLPICQDQLPVKCPPMCPQQWTPQYSTKSYSGYDGYCESSSLQCMEPCGPKGWLKCRPQCEQVYVPPCPP
PCVTKCPQQCVTQCVTKCPPPCVKCPPPCVKCPQPCVTQCITKCPPPCVTKCPQQCVTKCPQQCVTKCPPPCVTKCPQQXXXV
TKCPPPCVTKCPQQCVTKCPPPCVTKCPQQCVTKCPQQCVTKCPQQCVTKGQSSKVKISSNNKKYCSASKWF

>Am_EDP2
MSSRQNQQQCKQVPTLPPALSKAIPDPVPVLDPEVPEPVPAGPEPCPAPVKEPENAPRRRQEEEHCKQPLGQPLTLAPKLEPE
PESKLGGFLVPEEPEDSVPVQLPPLVEQQQQPSL

>Am_EDP3
MNLQKQEKQVPVSFKASEKPSPFLFSDLFSSQQQQQQWVPPKCQDKCPPKCVELCKPPKCQDKCPPKCPPQQ

>Am_EDPQ1
MSYSDQQQCKQVVCPPPVCPPTKCPPVCPPQKCPPPDCPPPVCPPQKCPPQKCPPPDCPPPKCPPQK

>Am_EDPQ2
MSYPNQQQWQQVVQPPPVIPPQKCPPLVFPPQKCPPPQIPPPKCPVPDIPPQKCPPPQWPQQKCPPQK

>Am_EDPQ3
MSYPNQQQCQKVVHPPPVIPPQKCPPFVIPPQKCPPPQIPPPKCPVPDFPPHKCSPPQWPQQQCPPQK

>Am_EDPQ4
MSYPNQQQCKQVVCPPPVIPPQKCPPPQCPLPKCPPSKCPPPQWPDQKCPLPQWPDQKCPPPQCPDQKCPPPQCPQQQ

>Am_EDPE
MSSHQMQCKQKTTLPPPLCKGPPNQDQDPVPLPEPVPLPAPIPDPGQGKTPDIKIPECPPQQQQQCKLPPIIPPCPPPCKEPP
VPEPIPFPEPGPCPEKPVPLPAPMPEPGKGKTPDIKIPKCPPQQQQQCKEPPVIIPPCPPPCKEPPVPEPMPFPEPGPCPEKP
VPLPLPTPLPAPMPEPGQGKTPDIKIPECPQQQQQCKLPPIIPPCPPPCKEPPVPEPIPFPEPGPCPEKPLPLPTPIPDPGQG
KTPDIKIPECPPQQQQQCKLPPIIPPCPPPCKEPPVPEPMPFPEPEPCPEKPVPLPLPTPLPAPMPDPGQGKTPDIKIPECPP
QQQQQCKEPPVVIPPCPPPCKEPPVKCPPPCPPIQQQQQKQPCQWPPQQK

>Am_EDPL
MSHDQQQIKQPFQPPPESSQLCPPLKFLESCPSAPPKCPEPAPPPPKGPEPSPQPPMMPTC

>Am_EDPCQ
MPWNRQQCKHQPLLGSSHVNKSLWQQDCEYCPLQCPQPRGKQCMQQCLALYPLHQSVCKGSLPGMTVSPQHMCVAECLRPCES
ESLQPGCTAEQPLQQHVINSQLCCAHTAQNWHQTANSQPCCTTVTTEPQRVTKCKAPCAPSAPNQECVNNRPPCAPSAPNQKC
VNNRPPCAPSTPNQKCVNNRPPCGPPAANHSCCTKSPPCKPESPAEERQEESPLWESVPERLQLQDMTDSQGMTKCPQLLCMM
DDLPSCMHTCPALKCKTTCPPPSSAQCPALQQSLASCPPLQAIQHPPLPRATNCSLPYDATHLPPPQQCWTTCPLTHPSNSTC
PLLEQSIATCHLPPCVMEGLLPATTKYPRGRQHLHKHLPPSFMTKRFPSGMTAAARQSCLTKRRRQRILRRPLLRPRMKCPLP
QQCFAMRPLPPSASECLLPGTTECPPEQRAPAHLPQQHIAKAHPQGVTGFPKYHRK

>Am_EDPCV

MSFQHQCKQPCLPPLICGQTVPSQPCVAPCSTVYVDPCPPGCVNPCPPQCVDPCPPGCVKPCPPQCVDPCPPVCVNPCPPQCV
DPCLPKCPPSQQCCAQTKLC

>Am_EDQL
MCSRENRDCHDRESSSCHDSGRSSCHGSGDVICHEVTPLPDIQPMPPMPMPTPMPAPAPIPCQQQQTKQPIHWPPQQQHQK

>Am_EDQM1_partial
MSWQNQQGSGCYRSGGCHSRGSSGGGCHSGGGCHGGGSSGGGCHSGGGCHSRGSSSGSCHGGGSSSGGCHSXXXIGRGYYGGG
GCHGGGSIGGGCHSSGGCHSGGGYYGGRSFHVISGGHSQGSQQHKQISQVPSQKLK

>Am_EDQM2
MSQQSQQGSGCHGSGGCHGTRSSSGGCHGGGGCHGGGGCHGGDSSGGGCHGGGSSSGGCHSGAGCHGGRFSQVIGGGHSQGSQ
QCKQISQVPSQKLK

>Am_EDRYA
MYSPGHYTAFCHGPSGSCLGKGGPYGCGQTFCSGIQHSQGDSRSSCHSSGPLCHGSGPFDQKTWRRQHVRKRRPVWPCQAPVQ
KCCAPAQLRYA

>Am_EDWM
MTCSSGRESYFNLNSTWYDPSGSWLENHRIPLCYADDSCCGGCNPDVRGVGGHNYRPCWYRRSVCSEAERGSSSGYCGSEDSG
CARRPTLGYSDGCGGYRRGPDRCNGECSSHEFGRRPTYHYAADVYLANERLACSEGCHGSSGGFYGSSGGCHRRRRCGEPCHG
SGSYGSSRGCHGRRRSVCGEPCHGSGSSGYLQPVCVKPEPCIPRCPPRQKYVRSTQSCCIPVQTYCAPVQVYCPPVGKYSSGR
QQCKQTSKLPTLKAK

>Am_EDYM1
MSYYGYQLKQQCYVPPGVKYSSCVTRCPKPPATKWTTPCATKRTEPCITKKLPEPCATTCVKTRVVRCPLPCTPTCPEPCAAE
CITPCATGYLEPCGLQQPQPFSKGWEHQWAPQYIQPCPMICPPVCGPAYIQPPAQKCTAPYYFQWSNRCGYGNCGPC

>Am_Lor_partial
MSSTQQKTACQEIPQQSRGLQGSSCHGGGSSVSSGNGIGGASSCCGGSSGQNICVSGSSGSSCCTGGSGYGIGGG
FNYGGGSSGQKIGVTGGSSSXXXVSSGQTIVGSGSGGSDYGIGGGSSCGGVSSGHKTIVGSGSGGSGYGISGGSS
YGTGGGSSSGGVSSGHKTIVGSGSGGSSYGIGGGSSSGGVSSGHKTIVGSGSGGSGYGTGGGSSYGTGGGSSSAG
VSSGHKTIVGSGSGGSGYGIGGGSSYGIGGGSSCAGGSSGQTISIGSCSGSGGSSGISIGGPGIGGGLSSGGSGR
SSSKLIINSRGSGGSCSISGSSSSGSGGSLQSVPQHQTKQPCQWPPQQK

>Am_PGLYRP3
MMLRLAVLFSALCASSCQLACPPIISPAKWGSRPAKCAAPLSKVPPGNVIIIHTSGSACHTQPECSELLRNIQVFHRDMKEWC
DISYNFLIGEDGNVYEGRGWLLEGAHTYGYNDLSLGIAFIGNFTERSPNEAAWKTLKNLLTYAVQSGYLASDYLLMAHSDVSN
TISPGKLIRETIKMWPHYKH

>Am_S100-A9
MKCPDAQTELEKAIETIINIFHQYSVRVGHFDTLTKMEMKLLIEKQLPNYMKNQTKPXEIDALFKDLDKNKDKQLSFGEFMVL
ITRVTIATHEHLHHCGEEEGQHHHQD

>Am_S100-A11
MSKVPVAPTETERCIESLLAVFQRYAGRSDRDETKLSKTEFLAFMNSELASFTKNQKDPAVLDRMMKKLDLNCDGQLDFQEFL
NLIGGIAVACHDALCTGGPGGPKGSSGPKGSSGPKKL

>Am_SCFN_partial
MPHLLDSIGTIINVFYQYATEDREGSSLSRRQMRLFIQKEFADILVNPYDPLMIDTVLRLLDQDGDGSIDFPEFLILSFRVAQ
ACYSYLAPKPGLQERQEQGRRSKELNELEAKADRGRSHQLREPEPRVGRRRHPEDPERKPEPRRDEGRRQQSLEPEQQVYEGR
RHQSHDPEQQEEVRSRSQPRDITQSHERSQRQVLEHEPPLYEESHRQPHEQECREKVRSRSQPRDTNTQAYERNRRPVLERES
QLDKESHRQPREPKKQEEVRSRFQSREPEPRWDEGRQRQPLEPKEREEVRGRSQPRDTDTQTYERSRRPVLEPEPQLYEERHH
KPREPEQREEVRGRSQPRDTNIQIYERSRHPVLEPEPQLYEERHHQPREQEQREEARSRSQPQEPEQLVYEGKGRRPHEPEER
EKVKSRSQPRDIDAESYERGQRPILEREPQLYEESHRQPREQEQQEAVRSHSQTQEPEQVYAGRRRQPLEPEQREEVRGRSQP
RDNDTQTHERSQRPVLEHEPRLYEESRRQPHEPEQREEARRRSPPQEPEQPVYERGRRQPREPEQQEEVKSRFQPCDTDTQSF
EKSRHPALEREPELYEESHRQPREPEQREKVRSHSQPQEPEPRGDEGRHRQLHELEQQEEVRGRSQPRGNDKQTHERSQHPVL
EREPQLYEESHRQPRVTEQQEEAGSCSQSREAEQQVYDGRGHQPCEPEQQEEVRSRSQPREPKPRGGERRRHLPHDPKQREEV
KSHSQPSDTDTQTYERSQHPGLECEPQLYEESHRQPREPEQQGEVRSRFQPQESEQQVYEGRRHQPRGPEQQEEVKSRSQPHN
ADTQNSERKQHPALEREPQLHEERSRQPREPEQREEVRSRFQPQEPEQQVYGRRGHQPRDPKREEVRSCTQWHETDTQNYERS
RHPVLERESQLYERSRRQPREPEQQEEVRSHFQPQEPGPRGDEGRQRQALEPEQQVYEGKRRQPLEPEQQKDIRSRFHPSDTY
TQSYVRSQRPALERESQFYEESPRRQPREPEQREKVKSRSQPQEPEPQGNERRHGQPYGPEQRAEVRSQSQPRDSDTQTYERS
QRPVLEXXXRPVLEHEPRLYEESHRQPREPEQREEVRSRSQPRNNDIQIYERSQRPVLECEPRLYEESRRQSPEPEQREEARS
RSQPQEPEQPVYDSGRHQPREPEQQEEVRSRSQPHEPEQVSRESYCEPRVPEQREVKDCYQSRESEQVYEESRHQVYEPDQQR
EVRSCYQPHVSEPRVDQGRQHQLRVSEEQYEGRRHQPHEAKTQAYERSRHPFLEREPERYEGSRIQLRELAQEGVRNRFRPRE
RGAQTDKRSHHELSEREPQVDEERCRQLCKPEHLGDVRRPYQPYEREQQVNGRRLRQSREPEQEHEGSHHQPSEAEWPAEVRT
HYLPREPEYMHEGRRRQVQDADRQGEVRSHYQPRAPDTRAFEGAHRPLFEQEPQLYERSRRQPPEPDQLGEERRHYEPRVPER
QVNEESRHQAHRIELQVYEQSRDLLCEPEPRVDEQRQRHRQEPEQQVYKRSRHQLSEVEQGEVRSHSQTREREQVPKSHHQPR
DPEQRGAKDRYLSREPEQVYEESHRQVCEPNQQGEVRRCYQPRETKPQAFKGRHCQLLECEPQLYEGRHEVKQADEGSHLPLE

PEQRETRSLYCLYDGHQHQPHQLEQRVDDGSRYQHHEPEQLRDVRSRYQPREPGPQGDARSRRPFHEREPELHEGSRRQPREQ
RDVWSHSQPREPDPRVDERSQPRPRQPELQVYEGSGHPLRDREQGESRDRSLPLDPETQTYERSRVQARDPEPQRDGQTREPR
DSELQLDEKSLCQPRAPVQLGNGRRRYQPRELEPEANEGSCYQPRGPEQQRNLQSRGQPREPEQRVKEGSRSHLPEQVQQGDA
QSRYVPLETRPQAHDGSRRQLQEPEIQHYEGNCHQSQNPGQQGDVRSHYRPHEPAQQVHKGNRQELHESETQKGEGSQRQPRE
PKEQVYEGSHRPTPLCDPEPQRAVQNSGQSRDPTPQQEEVNCHQPREPEPRGVAGNVHQRHEVLPPRDDQSLLPPDQPVPQRG
DSSDDPQPEVRPHDRSLPQPLEPERQGNERTQRQTLEPERREAEKNGQRPYAPDPKRDDEICHLVSAKQRDDGGRQPQSRETE
PRGGTGTRLQQRESEAQGNTATSRNPREPEAQGPERARQPRYPESQGVESSRQPQKAAPQEGERDNQLPQKAEPQDGERNRPQ
AGEAKPSKEEASQREPHNPNSMDDNRSRAAPTPPPTGDPGSQTQPREGEPRDGSRHQSNRPEGRGDEGSQPRAYEPGCRAGAT
VPSQTPERGLREGEGSRQQPWALEAPAEGGSRQQPPQGDAASQRQQEIAPEQAEGSHHLPREPLAQLQEESPTRAAESQEGEQ
SHHPPEPVVSQERLGDPSLDEAKASLPCSPLYVYLLAQKAEQQLCPVPAPQEQP

B

>Am_EDbeta
MACSTNVCNNSSVSCGVAAPQPIADSCNEPCVRQCPDSTVVIYPPPVVLTFPGPILSCFPQESVVGSSASPVLGSSLGGSYGA
GYPYGGSRCGSRYSNGSCGPC

>Am_Beta1
MTCSPALSSGICASPCGVAVPQPITDSWNEPCVRQCPDSTVVIQPPPSVVTIPGAILSSFPQEGIVGSTGAPHVAAGFGGGFG
SQGFYGSRAYMGAGGPYGYGGIWGYGGLCGSGGWRSGHRYLNGNCAPC

>Am_Beta2
MSCGTGCSNPCEVNCPQPQAVTANEPCVITCPDSRVIIYPPPVVVTFPGPILTTCPQESVVASTASADTVPAEMPASVPLTAA
VSGSLEPRAETIAPPIIPRPLPRYVPKYSHTYSSHWMHPCNTNRFGKRWAY

>Am_Beta3
MSCYDISYPPCGVTLPPCPEPFAVTSNEIYSAQYPDRIVETELEDGQPCTVTYPGPILTTFPQQTLVGSSALFDMERLLGSRR
SFEFEGLLGLGGICGPGSLCNSEPFDDFPYGNCGPV

>Am_Beta4
MACTDLCYPSSGIACPTPIANSYNDLCVRQCPDSRAVIQPPPVVVTFPGPILSSFPQDSIVGSSGAPVVGGYGSSFGTRFGYS
GLEGSLGYGSSGGYGGSGYEGLGGGYVGGSLGYGSGSLGYGGGLCSSGSLYNYGRLYGSGFGYGYCSPFSYRRYNRYRRGSCG
PC

>Am_Beta5
MSCYDECYVPCRATCPSPVADSCNELPCVRQCPDSTTVIQPPPVVVTFPGPILSSCPQGSVVGLLGAPSTGSSAGSLSYVGSF
GSGGLYNYGGLYSSGLSGLGTGDCCPYSRPLNTYHYGRCFPC

>Am_Beta6
MSCTDLCYPSGGIACPKPYADSCNEARVRQCPDSRVVIWPPPAVVTFPGPILSNFPQDSIVGSTGVPAVGHGAAGGTALSNGI
GGAGGFYGYEASLDSGGLYGYGGSLGYGGLCGYGGLSSDSGSCYSSGYCSPYSYR

>Am_Beta7
MSCTDLCYPSSGIACPRPYADSCNEACVRQCPDSRAVIQPPPVVVTFPGPILSNFPQDSIVGSVGVPTVGHGAAGGTALSNGT
GGVGGIYGYGAYLGSGGLYDYGGSLVYGGGYGGSLGSGGLCGYRGSLGYGGPCGYGGLSSGSGSCYSSGYCSPYSYRRYGRYR
YGSCGPC

>Am_Beta8
MSSYGQLISSRCYNPCEVTCPRPYADAWNEPCVTSCGDSRAVVYPPPVAITFPGPILTSCPQDSYVGTSEPQCIGSPYAAGGY
LGYRGSVGTGYSYPSYSRQLNRYHYGGCGPC

>Am_Beta9
MSSYGQLISSRCYNPCEVTCPRPYADAWNEPCVTSCGDSRAVVYPPPVAITFPGPILASCPQESYVGTSEPQCIGGPYTAGGY
LGYRGSVGTGYSYPSYSRQLNRYRYGGCGPC

>Am_Beta10
MSSYGQLISSRCYNPCEVTCPRPYADAWNEPCVTSCGDSRAMVYPPPVAITFPGPILASCPQESYVGTSEPQCIGGP
YTAGGYLGYRGSVGTGYSYPSYSRQLNRYSYGGCGPC

>Am_Beta11
MSSYRQLISSRCYNPCEVTCPQPYADAWNEPCVTSCGDSRAVVYPPPVAITFPGPILASCPQESYVGTSEPQCIGGPYTFGGY
LGYRGSVGTGYSYPSYSCQINRYRYGGCRPC

>Am_Beta12
MSAYGQLISSRCYNPCEVTCPRPYADAWNEPCVTSCGDSRAVIYPPLVAITFPGPILASCPQESYVGTSEPQCVRVPYTAGSY
LGYRGSASTGYSYPSYSRQLNRYQYGGCGPC

>Am_Beta13

MFPSFHPVFQANLHCRNTSSYGQLISSRCYNPCEVTCPRPYANAWNEPCVTSCGDSRAVVYPPPVAITFPGPILASCPQESYV
GTSEPQCIGGAYTAGGYLGYRGSASTGYSYPSYYRQLNRYRYGACRPC

>Am_Beta14
MSSYGQLISSRCYNPCEVTCPRPYADAWNEPCVTSCGDSRAVVYPPPAAITFPGPILASCPQESYVGTSEPQCIGLPYTADGY
LGYRGSVGTGYSYPSYSRQLNRYRYGGCGPF

>Am_Beta15_partial
XXXLSGGAAGPGGSLGYGGLYGNNFVSGGYGGYGGYGGYGGYGGYGSLYGLGSYGGSGGLGGYGGLCGYGGYGGYGGWGSSCY
GLGSGFGYGRCYSPYYSRRYRYGICAPC

>Am_Beta16
MSLYRQLISSRCSNPCEVTCPQPYADAWNQPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGSSAPISIGSSFGYGGS
FTYGGSLGYGGSCSTGSTYPCYSQRVNRYRYRSCGPCQTQKEFTCTRNIQETKCKIQAQGLADDCEKC

>Am_Beta17
MSFNRQLLSSRCFNPCEVTCPQPYANAWNEPCVTSCGDSRAVVYPPPVVITFPGPILSSCPQESYVGTSEPLQIGSSFVSRGS
VGSGSSLGCLSPYYSQRYNKYRYDNCGSC

>Am_Beta18_partial
MTLTGALCCYQPRPPCDVVCPRPYADAWNEPCVKSCGDSRAVVHPPPVVVTFPGPILASCPQESYVGTSLPEVNGGSFGSGSS
FGAGFGYRGNLGSRGYLGYGGGSLGYGGGSLGYGGGSLGYGGXXGFLGYGGGSLGYGGGSLGYGSYLGSGGSLGYEGPFCGF
GGPICGYGSYGGSYGSGLSSYGGGYSSPYYSRRYSKYRYGSCGPC

>Am_Beta19
MSCSKNPCNDSCSTRCEAKCPEPQVITSNEPCVIACEDTRVIIYPPPVVVTFPGPILTTCPQETVVASTVTLAESSDDATLAE
SPAMIPSAPEVTRSSVPCHEICPPCIIPRPKPQYLPNYSYTFSTQWIHPCNRSGFKKYKSS

>Am_Beta20
MSQSLSSRCLPPCDMTCPRPCADAWNWPCVTSCGDSRAVVYPPPVVVNFPGPILASCPQESIVGTVLPRPSGDIGPFPYGYGG
GYGSVSSYGFGGGYRSGSDYGSGSCYGSVSSRRYRKFSSGSCGPC

>Am_Beta21
MSQSLSSRCLPPCDVTCPKPCVNAWNWPCVTSCGDSRAVVYPPPVVVNFPGPILASCPQDSIVGTVLPRSSGGIGPFPYGSGG
GYGSGSSYGSGGGYGIGSGSGYGGSYGMGSGYGSGGSYGVSSGYGSGYGSGFDCGFGSGYGYGGGYRSGSSYGSVSSRRCRRY
SSASCGPC

>Am_Beta22
MALSSRCCPSDICPKPCVDACNWPCVTSCGDSKAVVYAPPVIVHFPGPILTSCPQESIVGTSMPNIIRGGGGPYSSGTIGSGF
SSGFSSGFSSGSGSGFGGGYGLGGGYGLGGGYGIGGGYGMGGGYGIGGCYGFGGGYKSGGYYGSVSSRRRRFSSVGCGPC

>Am_Beta23_partial
MGSYGPLVSSRCSNPCEVMCPEPCVQACNQPCVTSCGDXXXRRSYTSGRSSFGSSGGSRRGSCGPF

>Am_Beta24
MSCGQISSRCLPPCEMMCPEPYVAACNYPNTTSFGDSRAVVFAPPVIMTFPGPILASCPQESVVGAAEPYPIGGYPGGPYEGS
GGSYGGSSGIGGSYGASYGHSGGSYGYSGGSYGGSSGIGGSYGTSGRSYGGSSRTGGSSGATYGSSGHSSGISGGSRSSGTSY
GGSGGSRGGSRGGSGRTGVSSVSSGDPYGISGGSHRSGSSSGGSGESPKRYGESHERPGESEHSSGSFGGSYGISGGSRHSGT
SSGGSGESPKGYGESHEGPCESEHSSGRSGGSYGICGGSRRSGTASGGFGESHEGSRRNEDSSESSKGFYGISGESYGSGDPY
EGYGGSYAVGGSGIFRSSFFSRCPPGSFSSTFPFTRFAYQRQFGNNEFF

>Am_Beta25
MSCNTDPCPEGRPSPCEVKCPQPIVTSTNEACVVSCGDSRVIIYPPPVIVTFPGPILSTCPQESLVGAAVPCESGVSQSATTV
PLTSEIGGNSGFSVPLRSEIGGNSGFSVPLRSEIMGNSGCSAERLYLNREPEQPSTYTYSFTSQWRHPCNRPGWNRYRSFYMK
KEEPEEEEKPKEWHVGTEESS

Suppl. Figure S1. Amino acid sequences of proteins encoded by EDC genes of *Alligator mississippiensis* (Am). (A) Amino acid sequences of EDC proteins other than corneous beta proteins (CBP). **(B)** Amino acid sequences of CBPs, also known as beta-keratins. Cysteine (C) is highlighted in yellow, proline (P) in green, lysine (K) in cyan, glutamine (Q) in grey. Serine (S) and glycine (G) are bolded and in red and orange respectively. Stretches of X´s indicate unknown numbers of amino acid residues, that could not be predicted because of gaps in the corresponding gene sequences. Based on database version prior to update of December 2016.

A

>Gag_EDAA1
MSDSLGMLEDLCYQGLDCGWRSHYGRPCYCCCSDPCTEELIWEEWRCCPWWHSRGRYGRCWPC

>Gag_EDAA2
MSDSLDMIEDLHFQDPSCCWRPRRTRPCYCCCYDPCTGELIWEGWCCCPCWPRRRRYGTRWPC

>Gag_EDAA3
MFGSLDMIEDLSYQGQSDCFPHQRRRPFYTCCCYDQCGRLVWSGCSWSLPPWWCRQGSPGGCWPC

>Gag_EDAA4
MSDSLGMLEDLHFQDPSCCWRPCRRRPCYCCCYDPCTGELLWEGWCCCPCWPGRGRYGSRWPC

>Gag_EDAA5
MFDSLDMIEDLSYQGQSDCFPHYHRGPPYACCCYDQCGRLVWRGCCWSIPPWWCHWDSSGSSWPC

>Gag_EDAA6
MSDSLGMLEDLHFQDPSCCWRPRRTRPCYCCCYDPCTGELIWEGWCCCPCWPRRRRYGNCWPC

>Gag_EDAA7
MSDSLGMLEGLCYKDSSCCWRPRRTRPCYCCCYDPCTGELIWEGWCCCPWWRSSSRYGSCWPC

>Gag_EDAA8
MFDSLDIIEDLSYQGQSDCFPHYCRRPFYTCCCYDQCGRLVWRGCC

>Gag_EDAA9
MFDSLDMIEDLHFPGQSDCFPPYNRRPFYTCCCYDRWGRLVWRGCCWSVPPSWCYQGSSGKSWPC

>Gag_EDAA10
MFDSLDTVEDLYYPGQLDCTPPYPGRRYWCCCYDRYGRLVWQGPCCYPGPWCPRRGSSGRGWLC

>Gag_EDAA11
MFDSLDTFEDLFYPGQSDCRWPPYPQRRYLCCCYDRYGRLVWRGWCCHPGPYYRG

>Gag_EDAA12
MSESLDMLENLWYPGQSNCWDPCYRRPYWNSCWDPCTYKRPYWSGCWDPCTYRRPYIYNSCYGYGSLYGAG_GCYPYSTRWGR
RYSAGRCWPC

>Gag_EDAA13
MSESLDMLENLWYPGQSNCWDPCYRRPYWNSCWDPCTYKRPYWSGCWDPCTYRRPYIYDNCYGYGRLYGSGSCYPYSTRWGRR
GSWGSYWPC

>Gag_EDAA14
MSESLDMLENLWYPGQSNCWDPCYRRPYWNSCWDPCTYKRPYWSGCWDPCTYRRPYIYDNCYGYGRLYGSGSCYPYSTRWGRR
GSWGSYWPC

>Gag_EDAA15
MFDSLDAIEDLCYQGQYDCWDPCYRRPYWYGCWDPCTYRRPYIYDNCYGYGGLYGLGGCYPYFSRWGRKYSYGNCWPC

>Gag_EDAA16
MLDSLDTFEDLFYQGQSDCCSPSPRRPFYMCCCYHRWGRLVWRGCCWSIPPWWCCKISPGKHWPW

>Gag_EDAA17_partial
XXXGELIWEGWCCCPCWPWKGKYGRRWPC

>Gag_EDAA18
MTYHQSACDYVCYTPCSYGGLYGYQGLTGCWEPWTYRRPYSYGC

>Gag_EDAA19
MFDSLDMIEDLHFPGQSDCFPHQRRRPFYTCCCYDQCGRLVWRGCCWSIPPLWCSKGSSGSSWPC

>Gag_EDC1
MCSCCSGCHGTRSVQPICYVQPVCCEPVYIQRSLGSCCQPCGSCCGSCCGRSRSRPRVVIQRWPMPVCCPPLQYSAPMQQCCS
PLKKY

>Gag_EDC2
MCSCCSGCHGTESVQPICCEPVYIWRPLGTCCQPCGSCCGSCCGGSRPSPWLVIQRRPMPMCCPLLQYSAPMWKYSAPMQQCC
PPLKKC

>Gag_EDCH1
MCSRRSCHDRDTSCHRSRPSCHSSGSRCHESRPSCHISVVERPPYCQWQQYRPPYYPGYQYSQCCKFPQYPQYPQYPK

>Gag_EDCH2
MCSRRPCHERETSCHESRPSCHSSGSRCHESRPSCHISVVERPAYCQWQQYRPPYYPDYQQCYKYPPQYPKIPQCPQYPQYPQ
YPQFPQYPK

>Gag_EDCH3_partial
XXXYQQCYKYPPQYPKIPQCPQYPQYPQYPQFPQYPK

>Gag_EDCH4
MCSRRSCHDHHSSCHGSRYGSHCHESRPSCNITVVERPSIQSWCPVQRYCPPTCCYPRYQYSQCCKFPPQFPKCPPQYPK

>Gag_EDCH5
MCSRGSCHDHHSSCHRSRCGSHCHESRLACNITGVERPSMQSWCPVQQYCPVQRYCPASCCYPRYQYSQCCKFPPQYPKCPPH
YPK

>Gag_EDCH6_partial
MCSHGSCHSGHNSCHGSSSHCREPRPSCNVVVVEKPYVQACCPVPRYCPVPCYCPPVSCCYPRYQSSQXXX

>Gag_EDCH7_partial
MCSRGSCHSHHNSCHGSSSHCHEPRRCCNIVVVEKPYVQACCPVPCYSPPVSCCYPRYQSSQXXX

>Gag_EDCH8
MCSRRSCHSHHRSCHGSSSHRHEPRRCCNIVVVQKPYVQACCPLPCYCPPVPCCYPGYQYSQCCKFPQYPKCPPQYPK

>Gag_EDCH9
MCSRGSCHDHGSSSHGCHGHESSCHGSDSSINCVIEKPVPVCPVPQCCPQLPQCCVPVQQCCPPVQCCQQSKQCCKIPPPCPK

>Gag_EDCH10
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSIHCVIEKPVPLCPVQPCCPPVQQCCPPVQKCCPPVKCCQQSKQCCKFPPPCPK

>Gag_EDCH11
MCSRRSCHDHGSSSHGCHSHESSCHGSSTSVNCIIEKPVPVCPVPQCCPQPPQCCVPVQQCCPPVQCCQQSKQCCKIPPPCPK

>Gag_EDCH12
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSIHCVIEKPMPICPVQPCCPPVQQCCPPVQKCCPPVKCCQQSKQCCKFPPPCPK

>Gag_EDCH13
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSIHCVIEKPMPICPVQPCCPPVQQCCPPVQKCCPPVKCCQQSKQCCKFPPPCPK

>Gag_EDCH14
MCSRRSCHDHGSSSHGCHGHESSCHGSSSSIHCVTEKPVPLCPVQPCCPPVQQCCPPVQKCCPPVKCCQQSKQCCKFPPPCPK

>Gag_EDCH15_partial
XXXCKFPQYPKCPPQYPK

>Gag_EDCH16
MCFHGSCHDHHSSCHRSRCGSHCHEPRASCNIMVVERLSVQSWCPVQQYCPVRRYCPPTCCYPRYQYSQCCKCPPQYPKYKTY
HQH

>Gag_EDPCV
MSFQHQCKQPCLPPPICGQTVPSQPCVAPCGTVSVDPCPPGCVNPCPPQCVDPCPPVCVKPCPSQCVDPCLPVCVNSCPPQCV
DPCPPACVKPCPPQCVDPCPPVCVKPCPPQCVNPCPPQCVDPCPPKCPHQCCTQTKLC

>Gag_EDPQ1
MSSPNQQQWQQVMQPPPVIPPQKCPPPPQIPPPKCPIPDIPLQKCPPPQWPQQQCPPQK

>Gag_EDPQ2
MSYPNQQQWQQVVQPPPLIPPQKCPPPVCPPQKCPPPVCPPPVFPPQKCPPPVFPPQKCPPPQIPPPKCPVPDIPPQKCPPPQ
WPQQQCPPQK

>Gag_EDPQ3
MSYPNQQQCQQVVYPPPVIPPQKCPPPVYPPQKCPPLVIPPQKCPPTQFPPPKCPVPDIPLQKCPPPQWPQQQCPPQK

>Gag_EDPQ4
MSYPNQQQCQQVMQPPPVIPPQKCPPPVFPPQKCPPPQIPPPKCPVPDIFPQKCPPPQWPQQQCPPQK

>Gag_EDPQ5
MSYPNQQQCKQVVCPPPVCPPQKCPPPVCPPQKCPPQKCPPPVCPPQKCPPPQWPDQKCPPPQWPDQKCPPQK

Suppl. Figure S2. Amino acid sequences of proteins encoded by EDC genes of *Gavialis gangeticus* (Gag) that showed differential conservation within crocodilians. **(A)** Amino acid sequences of EDC proteins other than corneous beta proteins (CBP). Note- list is limited to proteins belonging to some gene clusters and genes where differences in number was observed. Cysteine (C) is highlighted in yellow, proline (P) in green, lysine (K) in cyan, glutamine (Q) in grey. Serine (S) and glycine (G) are bolded and in red and orange respectively. Stretches of X´s indicate unknown numbers of amino acid residues, that could not be predicted because of gaps in the corresponding gene sequences.