

Alma Mater Studiorum - Università degli Studi di Bologna

Dipartimento di Biologia Evoluzionistica Sperimentale

Dottorato di Ricerca in Biodiversità ed Evoluzione – XIX° ciclo

Settore scientifico-disciplinare: BIO/05

Evolution of repetitive DNA in model arthropods

Tesi di Dottorato

Dr. Andrea Luchetti

Coordinatore:

Chiar.mo Prof.

Giovanni Cristofolini

Relatore:

Prof.ssa Barbara Mantovani

...And after several round
of gene conversions, that
small repeat said: “After all,
I can transpose, now....”

To my family

Acknowledgement

I am very grateful to Prof. Barbara Mantovani because of her complete trust in me and for her continuous encouragements and supports (...and control). I also wish to thank the BECoN group, the Termites group, the *Tunga* group, and all the staff of the Molecular Zoology Lab. And many thanks also to Ax and its clushti!

A lot of other people I wish to thank, even simply because they exist. Therefore, I'll not mention anyone here for the risk to forget someone. It is sufficient to say that I'll be always grateful to all you.

Index

Chapter 1. Introduction pag. 1

Chapter 2. RET76 satellite DNA dynamics in termites pag. 29

Chapter 3. Ribosomal intergenic spacer variability in tadpole shrimps
pag. 50

Chapter 4. LEP150 repetitive DNA evolution in clam shrimps pag. 112

Chater 5. Conclusions pag. 173

Chapter 6. Other research activities carried out during the PhD course.
pag. 179

Chapter 1. Introduction

*R*epetitive DNA sequences and their evolution are topics of particular interest for two fundamental and obviously linked reasons. First of all, this kind of DNA constitutes the majority of nuclear DNA and it is highly differentiated in its genomic role, involving protein and non protein coding genes and non transcribed sequences. Secondly, owing to its involvement in so different nuclear dynamics, its mechanisms of evolution are far from being defined: concerted evolution, the only precisely described pattern, requires particular prerequisites and these are not always met in nature. In this work, three cases of study will be presented in order to test the impact of peculiar organismal (i.e. eusociality and unisexuality) and molecular (i.e. proximity of functional gene/domain located outside the repeats array) conditions on the evolutionary pattern of repetitive sequences, trying to highlight described mechanisms and suggesting a new one. Three animal models from the phylum Arthropoda has been chosen: two crustaceans Branchiopoda taxa and a termite species complex. For each of the three models, population genetic and phylogenetic studies have been preliminarily carried out, in order to better define the taxonomy and systematics of these animal systems.

Branchiopoda: general features

*T*he class Branchiopoda pertains to the sub-phylum Crustacea (phylum Arthropoda), and it is considered to be one of the most diversified group (Martin, 1992). As a consequence, Branchiopoda monophyly is still debated even if some evidences, such as larval characters (Sanders, 1963) and sperm cell morphology (Wingstrand, 1978), support this hypothesis.

Branchiopoda are characterized by flat, leaf-shaped trunk appendages (phyllopora), often only poorly differentiated. They are used to swim and breath as well as for feeding; branchiopods feed removing organic particles from the water as they swim or stirring up sediments to resuspend organic particles.

Others can scrape organic matter from sand grains and rock, or may actively prey on other small animals.

The number of trunk segments can vary, depending on the taxon, with several cases of fusions. The posterior region usually lacks of appendages and in most cases end with a caudal furca. Several orders possess a carapace derived from an expansion of the dorsal cuticula, probably originating from the maxillary somite. It can extend dorsally and can fuse, assuming the shape of a univalve or bivalve shield.

As other Crustacea, Branchiopoda larval development occurs through a nauplius stage: this larva is not segmented and is characterized by three pairs of swimming appendages and a single eye.

Branchiopoda are fresh-water crustaceans, usually living in ephemeral ponds; few species are known to be marine and some taxa live in salt water (for instance *Artemia salina*). Being well adapted to unstable environment, Branchiopoda are able to survive for long time without water owing to the production of resistant eggs. These eggs can remain quiescent until environmental conditions allow the hatching with the subsequent development of larvae and adults.

These crustaceans are known from the Devonian era (400 Mya) and they are, therefore, among the oldest living crustacean (Fryer, 1987). Branchiopoda systematics has been revised several times (Fryer, 1987; Hanner & Fugate, 1997; Braband et al., 2002; Stenderup et al., 2006), the definition of both higher and lower taxonomic entities being rather complex due to the wide morphological variation and differences in reproductive strategies. The four orders classically recognised are: Anostraca, Conchostraca, Notostraca and Cladocera.

Anostraca, such as *Artemia* spp, completely lacks a carapace and have elongate bodies compared to most other living branchiopods, with leaf-like appendages used to swim belly up. One centimeter is the average size for adult anostracans, but some species can reach 10 cm.

The Notostraca are called "tadpole shrimp" because of their broad, plate-like carapace and narrow, elongate thorax/abdomen. Some species can be found in rice fields and can cause considerable damages by burrowing into the sediment

and inadvertently dislodging the young rice plants. Adults are predators, also preying on relatively large animals, such as other Branchiopoda or even tadpoles. The order Notostraca comprises only two worldwide distributed genera, *Triops* and *Lepidurus*, whose gross morphology is very similar. Present *Triops cancriformis* individuals are indistinguishable from the Triassic form, despite 200 Myr have passed (Trusheim, 1938; Longhurst, 1955): therefore, it is considered the oldest living animal species (Fryer, 1985).

Cladocera are called "water fleas" and are typically quite small ranging from 0.5 millimetre up to three millimeters. The best known member of the Cladocera is *Daphnia pulex*. The bivalved carapace of cladocerans covers the thorax/abdomen, but not the cephalon. Cladocerans dominate the plankton found in freshwater habitats. They are mostly benthic and swim with their antennae instead of their thoracic appendages. The few cladoceran groups that occur in marine habitats are morphologically very differentiated, having huge eyes and, instead of being filter feeders like most of their freshwater relatives, they are active predators.

The Conchostraca, or "clam shrimps", have appendages along the entire thorax/abdomen and their carapace is bivalved with a hinge that allows the opening and closing of the two halves. The conchostracan body is completely enclosed in the carapace. Conchostracans are found exclusively in freshwater.

Branchiopoda show a wide range of reproductive strategies: from the bisexuality, both gonochoric and hermaphroditic, to parthenogenesis.

The species analysed in this thesis are the notostracan *T. cancriformis* and the conchostracan *Leptestheria dahalacensis*.

***Triops cancriformis* (Bosc, 1801)**



In the head, the dorsal organ is rounded off or oval, the second pair of jaws is of greater dimensions with respect to the congeneric species. The carapace is oval and its carena finishes with a spike that is equivalent in length or larger than those present in the groove. Some specimens show spikes also on carena. The trunk is composed from 32-35 segments, 4-9 of which apods, with no thorns on the ventral side. The number of appendages is highly

variable (48-57). In the telson there are 1-4 great thorns and one long furca. In some individuals, the telson has a structure similar to the supra-anal plate: such structure is completely absent in congeneric taxa, while it is clearly smaller with respect to the one found in individuals of the other notostracan genus *Lepidurus*. The presence of a kind of a supra-anal plate induced Linder (1952) to hypothesize that *T. cancriformis* pertains to a third, monospecific genus of the family Triopsidae. Recent molecular analyses support this hypothesis, showing a clear-cut differentiation of this taxon from other *Triops* species (Mantovani et al., 2004). There is not marked sexual dimorphism and the only useful sexual character is the presence/absence of eggs.



Figure 1. *T. cancriformis* samples.

T. cancriformis is distributed in Europe (from Spain to Russia), North Africa, the Balkans, Anatolia and from the Middle East to India. It has been also imported in Japan in the last century. It lives in temporary pools, and its presence is related to the absence of predators and to high temperatures, necessary for the hatching of resting eggs. In Japan, Spain, France and Italy, they are considered

harmful for rice cultivations. The larva is a metanauplius, with the dorsal organ of the first stage rounded off. Males are present sporadically across Europe and there are all-female populations. In some cases histological analyses have shown that individuals considered females actually are hermaphrodites, having a male portions in their gonad (Zaffagnini and Trentini 1980). Italian populations are composed only by females and probably they reproduce themselves through parthenogenesis (for an overview see Scanabissi et al., 2005). *T. cancriformis* therefore show both anfigonic and parthenogenetic populations.

On morphological grounds, three subspecies have been recognized: *T. cancriformis cancriformis* (Bosc), *T. cancriformis simplex* Ghigi and *T. cancriformis mauretanicus* Ghigi. It should be noted that the high morphological variability makes very difficult to unequivocally diagnose a single individual, especially as far as *T. c. simplex* is concerned (Longhurst 1955). Recent phylogenetic studies, based on mitochondrial data, individuate two main lineages: one is formed of *T. c. cancriformis* populations and samples from Northern Spain that were classified as *T. c. simplex* in the most recent literature; the second lineage comprises all populations of *T. c. mauritanicus* and northern African populations of *T. c. simplex* (Korn et al., 2006).

***Leptestheria dahalacensis* Rüppel, 1937**

L is a canonical gonochoric taxon, with sex ratio ranging from 49% to 68.4% biased toward males (Tinti and Scanabissi, 1996). It presents a prominent occipital angle in both sexes, always with rostral thorn. The cephalic rostrum is rounded off in males and acuminate in females, and the inferior part of the careen shows small bristles in males that are absent in females. In males the first two pairs of toracopods are modified in claspers, with the inner tubercle rather prominent. Six epipodites (X-XV appendage pairs) are modified in cylindrical ovopositors. Last abdominal tergites show short conical spikes.



Figure 2. *L. dahalacensis* samples.

L. dahalacensis is distributed across Eurasia: in Europe it is present in the Baleari, Belgium, Italy and Danube area. In Asia it is found in Anatolia, Middle-East, Daghestan, Uzbekistan, Mongolia and China. It takes the name from the locality where it was first described, the Dahalac islands in the Red Sea.

It is found in temporary pools and in rice fields. All the species pertaining to the *Leptestheria* genus seem able to support consistent variations in environmental conditions. The adults are mainly benthonic and continuously move the surface of the substrate, from which they gain the food by filtration. The development take place through 5 larval stages: nauplius I and II, metanauplius, peltatulus and heilophore (Eder 2002). No molecular data are to the present available on this taxon.

Isoptera: general features

The order Isoptera is the only other insect order, beside Hymenoptera, in which can be found populations organised in colonies with rigid social structure (i.e. eusocial colonies).

Termite colonies are comprised of three basic castes: workers, soldiers, and reproductives. They are small insects with body coloration varying from white, typical of the great majority of the individuals, to tawny, exclusive of the sexually mature individuals (king and queen). The head shows a pair of composed eyes, well developed in fertile termites, while undifferentiated and little functional in the sterile individuals (juvenile forms, soldiers and workers). Simple eyes are

generally absent, but they appear in the more ancient groups. The antennae are long approximately two times the head, which presents a chewing mouth-parts. Fertile individuals are equipped with two equal wings (= isoptera), while juvenile forms, soldiers and workers are always wingless; in secondary reproductives wings never reach the complete development. In the gut symbiotic bacteria and protozoans able to digest the cellulose and allowing its assimilation are present. Gonads are developed and functional in the royal couple and, more generally, in swarming alates; on the contrary, wings are atrophic in individuals of all the other castes, with the exception of the secondary reproductives. In these individuals, gonads reach the complete development and functionality in particular circumstances, such as the death of the royal couple or the foundation of a new colony by budding. With age, the ovaries of the adult termites are able to increase the number of their ovarioles, with ageing of the individual. The enormous development of the ovaries, leading to a remarkable increment of the fecundity, and the considerable increase of the emolymph mass, produce an abdominal hypertrophy (physogastry).

Eggs hatch into small larvae that are genetically able to develop into any caste. Epigenetic factors that may direct the developmental pathway followed by a single termite are represented by time of the year, diet and pheromones.

Workers constitute the bulk of the population. In lower termites there is a false worker caste called pseudergates; these retains the potential to become alates (for instance, in *Kaloterme flavicollis*). Workers feed all the other individuals: larvae, nymphs, soldiers and reproductives. They also organise the colony life. Soldiers may develop from nymphs, pseudergates, or workers through two moults. Owing to their specialized defensive features, soldiers provide colony defence against predators.

Reproductives develop either from alates or neotenic. Alates are winged termites, and each species produces alates at a particular season. The alates of each species fly at a unique time of the day and under specific conditions. They develop from nymphs by growing wings and compound eyes. After flying, the alates break off their wings along a basal suture and are then called dealates. Dealates form tandem courtship pairs and dig into the soil adjacent to wood, mate, and start a colony (in this case called also family). Primary

reproductive females, or queens, vary in size depending on the species. The enlarged abdomen makes them relatively immobile and dependent on the workers. There is usually just one pair of primary reproductives per colony, but some species have a low incidence of colonies with multiple reproductives (polygamy). Secondary reproductives may develop from unswarmed alates (adultoids), nymphs (nymphoids) or workers (ergatoids). If a primary reproductives dies, it is usually quickly replaced by a secondary reproductive of the same sex. In the more primitive, wood-inhabiting termites, large numbers of pseudergates quickly moult to neotenics when removed from the pheromonal inhibition of a primary reproductive. While primary reproductives usually outbreed, secondary reproductives always mate with their kin, obviously enhancing the inbreeding.

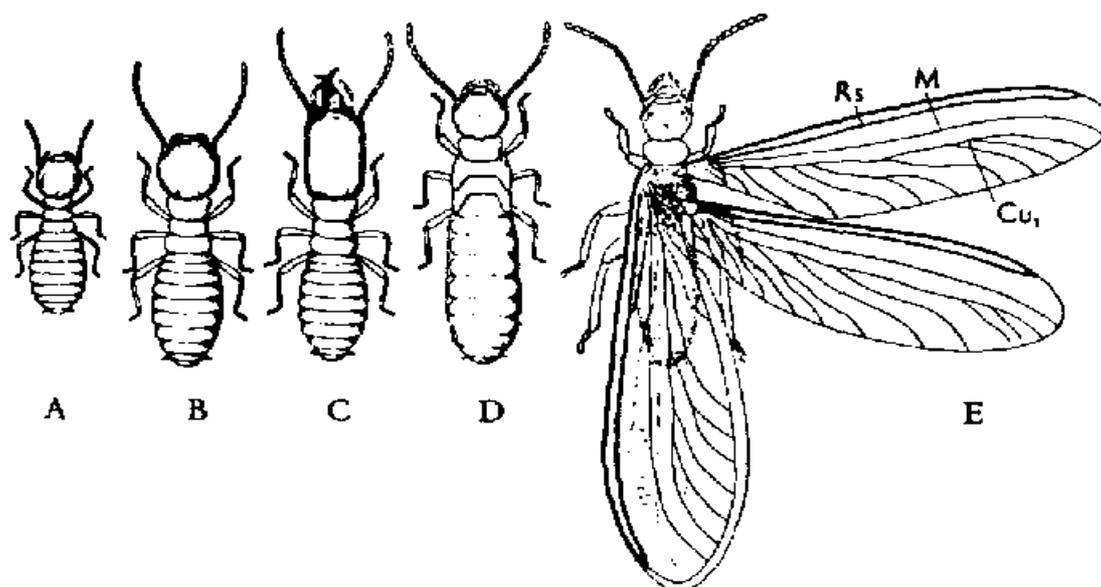


Figure 3. *Reticulitermes* castes. A: nymph; B: worker; C: soldier; D: secondary reproductive; E: alate, primary reproductive.

The genus Reticulitermes Rossi

The genus *Reticulitermes* pertains to the family Rhinotermitidae (sub-family = Heterotermitinae), that is the apical lineage of the lower termites (Fig. 4).

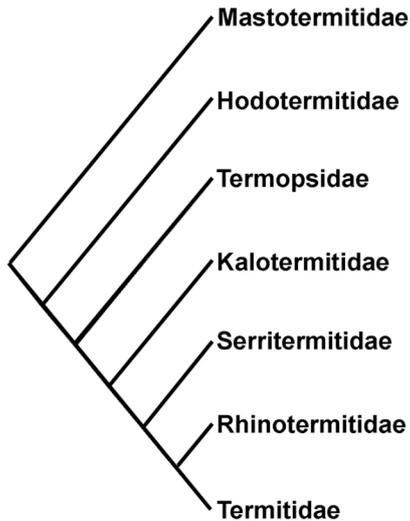


Figure 4. Higher phylogeny of Isoptera. Modified from Eggleton, 2001.

Reticulitermes colonies are huge families composed by thousand of individuals; the colony ontogeny has been long debated due to the cryptic life style of these termites (Thorne et al., 1999; Lainé and Wright, 2003). However, gathering ecological, behavioural and genetic data (reviewed in Thorne et al., 1999) it has been hypothesized that a colony is founded by dealates and slowly growth until the number of new workers supports an increase in egg production. Once this simple family is established, multiple secondary reproductives develop at the edges of the colony, thus promoting further increase in growth rate. New colonies are then founded by alates swarming or colony budding (splitting).



Figure 5. Left: *Reticulitermes flavipes* alate before the swarming; right: worker and soldier within a nest.

This happens when reproduction occurs between nestmates spending much or all their time separated from the primary reproductives, thus creating a “satellite” colony unit (Thorne et al., 1999). It is to be noted that budding does not always produce colony division, since contacts with parent colony remains. This has been demonstrated by genetic studies, evidencing different kind of colony structures in *Reticulitermes*: these ranges from simple families, headed by single dealate royal couple, to complex and interconnected nests with several neotenic secondary reproductives (reviewed in Thorne et al., 1999).

The genus has a wide distribution in temperate areas of Europe, America, North Africa and Asia. The taxonomic rank of some *Reticulitermes* taxa, especially from the Mediterranean area, as well as their phylogenetic relationships are still debated, particularly for the lack of reliable morphological markers (Clément et al., 2001). The genus *Reticulitermes* embodies six fossil species (some probably dating back to the Eocene) and fourteen living ones.

The taxonomic situation derived from literature data at the beginning of this work is reported in Figure 6, with the seven species and subspecies then recognised in Europe.

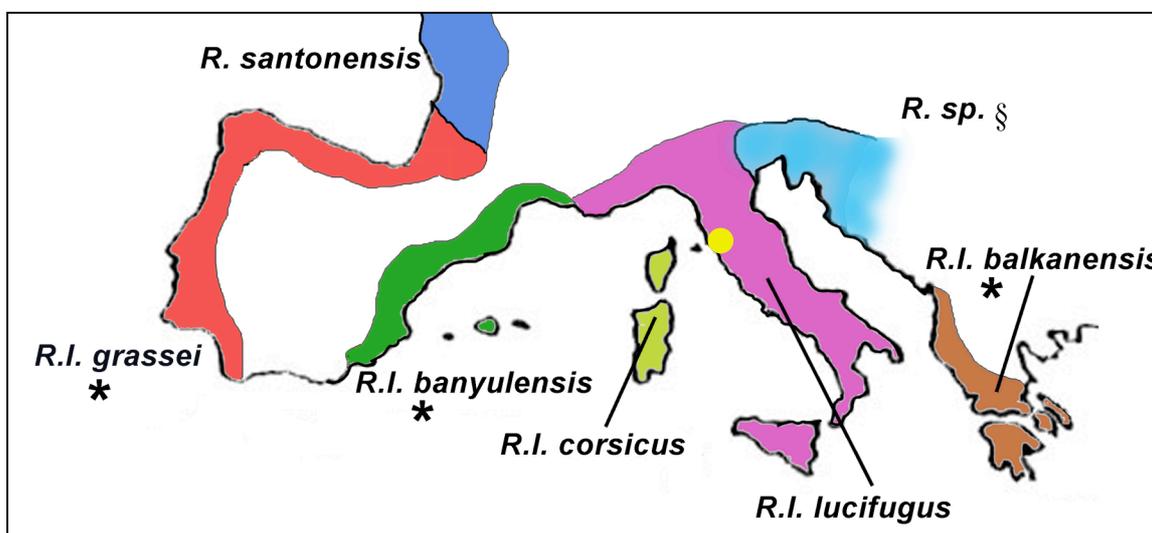


Figure 6. *Reticulitermes* European distribution as resulted from literature data (Plateaux and Clément, 1984; Campadelli, 1987; Bagnères et al., 1990; Clément et al., 2001; Jenkins et al., 2001; Marini and Mantovani, 2002).

*: These taxa have been raised to a specific rank (Clément et al., 2001).

§: The taxon *R. sp.* has been described as a new species (*R. urbis*; Bagnères et al. 2003).

Other two taxa have been described in the Mediterranean area. The first, *R. clypeatus*, was described on morphological ground by Lash (1952) in the Jerusalem area. The second taxon was retrieved in Turkey and it was first described as *R. lucifugus*; however molecular analyses, based on mitochondrial DNA, evidenced a clear differentiation with respect to the *R. lucifugus* species complex (Austin et al., 2002).

The repetitive DNA

Eukaryotic genome shows a substantial fraction constituted by repeated DNA sequences, either tandemly or non-tandemly arranged. They can be localized in specific chromosomic regions or dispersed in the genome. Repetitive sequences can be further distinguished, on the basis of their copy number, in middle repetitive and highly repetitive sequences. Typical examples of the former are gene families such as ribosomal DNA (rDNA) and histone genes, while the latter are represented by the so-called satellite DNA (satDNA), which is usually not transcribed (Elder and Turner, 1995).

Ribosomal DNA (rDNA)

The eukaryotic ribosomal DNA gene family is arranged in a cluster known as nucleolar organizer region (NOR); it can be localized on one or more chromosomes. This region is composed of hundred to thousand of tandemly arranged rDNA members (repeats). Each repetitive unit consists of 18S, 5.8S and 28S coding regions, separated by internal transcribed spacers (ITS1 and ITS2, respectively), while 18S and 28S of adjacent repeats are separated by a non transcribed spacer (NTS) and an external transcribed spacer (ETS) (generally referred to as intergenic spacer, IGS; Fig. 7).

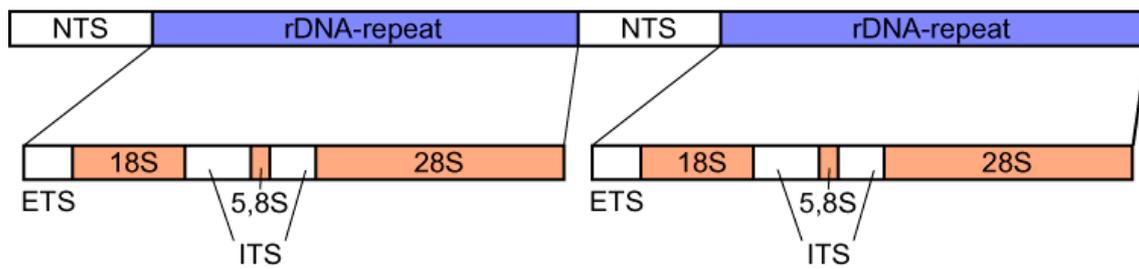


Figure 7. Schematic drawing of the tandem organization of eukaryotic rDNA.

The three ribosomal genes are co-transcribed as a part of a single poly-cistronic precursor called 35/37S in yeasts, 40S in insects and amphibians and 45S in mammals. The transcription starting point (TSP) can be found in the IGS region, upstream the 18S, while the synthesis finishes in the following IGS in a point located downstream the 28S. The transcript therefore is a pre-rRNA constituted by the 18S / 5.8S / 28S coding regions, preceded from an external transcribed spacer (5' - ETS), separated by the two internal transcribed spacers (ITS1 and ITS2) and followed by a short external transcribed spacer (3' - ETS). The IGS sequence structurally separates one transcriptional unit from the following one. Beside the presence of initiation and termination transcription signals, this region is of particular interest for the occurrence of clusters of repetitive sequences (sub-repeats): these sub-regions seem to play an adaptive role to local environment (Gorokhova et al., 2002). It is known that sub-repeats could act as enhancer of rDNA transcription by carrying a gene promoter duplication in their sequence. However, in the dipteran *Simulium sanctipauli*, IGS sub-repeats lack of promoter sequences and it has been suggested that the simple repetitive nature of this region could act as enhancer of transcription, as observed in some vertebrates (Morales-Hojas et al., 2002 and reference therein). Sub-repeats organisation may vary from a cluster of tandem repeats, such as in *Aedes albopictus* (Baldrige and Fallon, 1992) to two different clusters, like in *Artemia* and *Daphnia pulex* (Gil et al., 1987; Crease, 1993) to four, as observed in the swimming crab *Charybdis japonica* (Ryu et al., 1999). In contrast, in *Aedes aegypti* sub-repeats are not organised in tandem but they are interspersed with unrelated sequences (Wu and Fallon, 1998), while in the

copepod *Tigriopus* the IGS region completely lacks of subrepeats (Burton et al., 2005).

A fourth ribosomal gene, the 5S, is not usually included in the tandem organization of the NOR, because it is transcribed by a different RNA polymerase. The 5S genes are usually found either as tandem arrays residing outside the ribosomal DNA repeat units or dispersed in numerous genomic sites. However, several studies have shown that 5S genes can be linked to diverse tandemly repeated gene families (Drouin and Moniz de Sà, 1995 and reference therein). It has been shown that 5S genes are linked to the rDNA in the copepod crustacean species *Calanus finmarchicus* (Drouin et al. 1987), as well as in the nematode *Meloidogyne arenaria* (Vahidi et al. 1988). On the other hand, the 5S gene of the nematode *Caenorhabditis elegans* is linked to the tandemly repeated trans-spliced leader sequences (Krause and Hirsh 1985). In other crustacean species, such as the branchiopod *Artemia salina* and the isopod *Asellus aquaticus*, 5S genes are found linked to the histone gene repeats (Andrews et al. 1987; Barzotti et al., 2000). In crustaceans, two new organizations have been recorded recently: a cluster of tandemly arranged 5S genes in the isopod *Proasellus coaxalis* (Pelliccia et al., 1998) and a further linkage type in *A. aquaticus* with the U1 small nuclear RNA gene (Pelliccia et al., 2001).

Satellite DNA (satDNA)

Highly repeated sequences are often indicated in the literature as  satellite DNA (Elder and Turner, 1995). This structure takes the name from the fact that ultracentrifuging the genomic DNA at 60,000 RPM, for 72 hours in cesium chloride gradient (CsCl), the genome is separated in two fractions: one constitutes the so-called main band, while the other constitutes the satellite band. This band can be heavier or lighter than the main one depending on G+C or A+T richness. This band always comprises highly repeated sequences. Satellite DNA is constituted by sequences from ten to thousand of bp long, typically organized in arrays up to 100 megabases. Usually, they are localized in the heterochromatin of pericentromeric and/or

telomeric regions, but sometimes they are dispersed in the genome. The copy number is substantially conserved within the populations, but the monomeric unit may show several variants regarding the nucleotide sequence (Charlesworth et al., 1994; Ugarkovic and Plohl, 2002). The possible role of this fraction of the genome has been long discussed; recently some evidences have concurred to formulate some hypotheses. Some surveys conducted on chromosome evolution indicate that satellite DNA is directly correlated with the dynamics of the chromosome restructuring, with particular impact on the chromosome changes associated to the speciation: genomes showing frequent chromosomal rearrangements, as in the case of *Equus zebra*, embody a greater number of satellite DNA families, with respect to those with a highly conserved karyotype, like *Macrotus waterhousii* (Phyllostomidae) (Wichman et al., 1991; Bradley and Wichman, 1994; Garagna et al., 1997, 2001; Li et al., 2000). Slamovits et al. (2001) have proposed that the active involvement of satellite DNA in the processes of expansion, contraction and mobilization promotes the chromosome restructuring, rather than its amount.

The most significant studies focusing on the function of satellite DNA are those regarding centromere functionality, in particular the association with CEN proteins, that are specific centromeric proteins (Earnshaw and Rattner, 1989). It has been evidenced that CENP-B interacts with one portion of 17 pb of the human α satellite DNA called CENP-B box (Masumoto et al., 1989); subsequently a CENP-B box has been observed also in *Mus musculus* as well as in other mammals and plants (Stitou et al., 1999 and reference therein). It has been therefore proposed a structural role for satellite DNA through interaction with specific proteins needed for the normal centromeric activity (Schueler et al., 2001; Henikoff et al., 2001; Cooper and Henikoff, 2004; Henikoff and Dalal, 2005). Even if satDNA is usually described as non-coding element, it has been observed that some satellite DNAs are transcribed, as in hymenopterans *Diadromus pulchellus* and *Aphaenogaster subterranea* (Renault et al., 1999; Lorite et al., 2002). The transcription can be specific: in *Diprion pini* (Hymenoptera) is sex-specific (Rouleux-Bonnin et al., 1996), in *Gecarcinus lateralis* (Crustacea) is tissue- and stage-specific (Varadaraj and Skinner, 1994); in *Mus musculus* it is regulated with a cell-specific pattern

(Rudert et al., 1995; Sam et al., 1996). These transcripts have all ribozymatic activities; however, it is not still clear their cellular function (for a review see Ugarkovic, 2005).

Besides satellite DNAs, on the basis of the repeat unit size, other two kind of tandem repeats array can be found. The microsatellite sequences are array of short units (2-5 bp) moderately repeated up to hundred of times. The copy number varies both within populations and between populations. The minisatellite sequences are unit of approximately 15 bp that form repetitions of 0.5-30 kbp. Also minisatellites vary remarkably in the number of copies.

Evolution of repetitive DNA

A non-independent variation of sequence divergence is commonly observed for repetitive DNA sequences: the divergence between sequences within an evolutionary unit (either strain, population, sub-species, species, etc) decidedly turns out smaller of that one observed between different ones (reviewed by Nei and Rooney, 2005; Fig. 8). This pattern of variability cannot be described by the classical model of divergent evolution and led to the concept of concerted evolution.

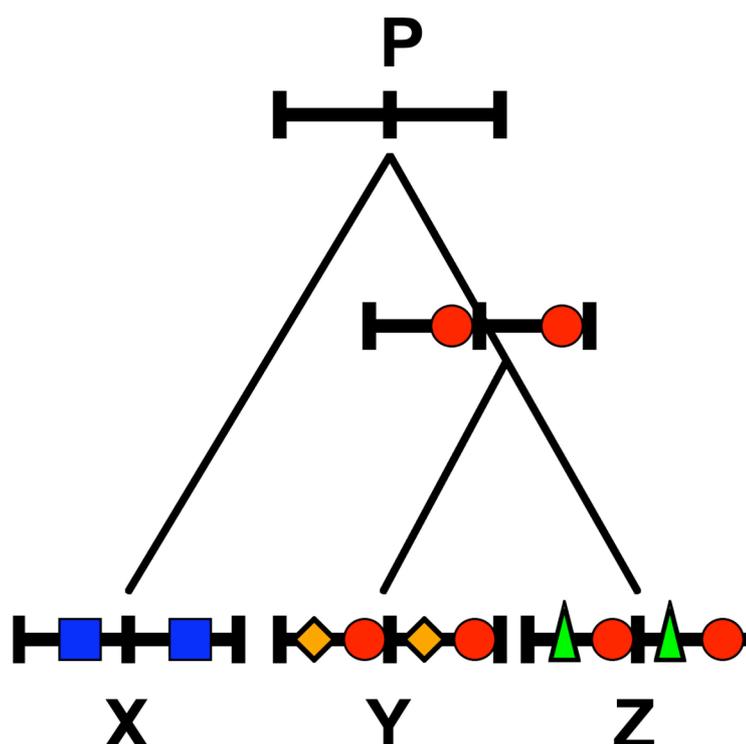


Figure 8. Pattern of concerted evolution. P: parental species; X, Y, Z: derived species.

Several studies on the genetic variation of repetitive DNA have suggested that selection and random genetic drift cannot describe the pattern of concerted evolution. Natural selection, in fact, would fix one determined mutation at each locus, but it is unlikely that the same mutation can be fixed at all loci. On the other hand, random genetic drift would fix different mutations at each locus, independently for individuals, populations and taxa. In order to explain the observed pattern of concerted evolution, therefore, it has been proposed the mechanism of molecular drive. This process is realized by means of a variety of mechanisms of genomic turnover that favour the spread of new mutations among members of a multigenic family (homogenization), and, thanks to bisexual reproduction, among individuals of the same population (fixation) (Dover, 1982, 1986, 2002). The mechanisms of genomic turnover (slippage replication, unequal crossing-over, rolling circle replication, gene conversion, duplicative transposition) can cause, in different ways, increases or losses of mutations during the life of the individual, leading to non-mendelian segregations. This can determine the random spreading of a mutation within its multigene family on the same chromosome and among homologs and non homologs chromosomes. If an unequal exchange between chromosomes happens, and such fluctuations increase the frequency of that mutation, then, due to panmixis, in the next generation more individuals bearing such mutation will occur. Even if it is probable that the greater part of the new mutations will be eliminated, some possibility remains to have some of these mutations fixed and replacing the former variant. This is obviously correlated to the effective population size (N_e) and to the rate of unequal exchanges (if not by eventual biases). Therefore the fixation and the homogenization are closely correlated by means of the continuous chromosomes reshuffling in every generation, due to bisexuality. Moreover, the great difference between the rates of genomic turnover and those of chromosome segregation guarantees the genetic cohesion within the population. Therefore, contrary to selection and genetic drift, molecular drive in bisexual taxa allows the spreading of new mutations among repeats that otherwise would evolve independently.

Other evolutionary dynamics observed studying satellite DNAs, related to the co-existence within the same genome of more than one family. It is not unusual that in a taxon more satellite DNA families are present. In particular, it has been observed that some families are preferentially amplified with respect to others in closely related species. In order to explain this phenomenon Salser et al. (1976) have proposed the "library" hypothesis: some sequences, among the several present in the genome, can be species-specifically amplified, while the others would be maintained with a low copy number (reviewed by Ugarkovic and Plohl, 2002); this theory has been confirmed from different studies on congeneric species (Mestrovic et al., 1998; Miller et al., 2000; Cesari et al., 2003). Another hypothesis explaining the considerable fluctuations in copy number and variability of satellite DNAs coexisting in the same genome, has been put forward by Nijman and Lenstra (2001) in the "Feedback Model". Briefly, a satellite DNA family encounters three phases during its life: in the first phase, interactions of homogeneous repeats cause rapid expansions as well as contractions with saltatory fluctuations in the copy number; then, in the second phase, mutations and recombination events lead to new variants, evolving independently. The third, terminal phase is reached when degeneration by mutations stops interactions between old monomers and a new satellite DNA family takes their place.

Particular cases of repetitive DNA evolution have been studied, in which concerted evolution is not achieved.

A first example comes from studies on the Bag320 satellite DNA family that is shared by bisexual and unisexual taxa of *Bacillus* stick insects. The parthenogenetic taxon *B. atticus* does not show fixation in different sub-species, as instead observed in the bisexual *B. grandii*. This was explained as due to the lack of chromosome reshuffling and panmixis in *B. atticus* (Luchetti et al., 2003).

Another particular situation is related to repeats residing at edge of the array: in these regions, genomic turnover mechanisms should be less efficient, as demonstrated by theoretical simulations (Smith, 1976). Experimental studies on human α -satellite DNA have shown that monomers at the cluster ends are less homogeneous than other repeats (Mashkova et al. 1998, 2001; Bassi et al.

2000). This particular dynamics is also evident in the analysis of the ribosomal IGS subrepeat arrays: in *D. pulex* (Crease 1995) and in the swimming crab *Charybdis japonica* (Ryu et al. 1999) external repeats are less homogenized with respect to the inner ones.

References

Andrews MT, Vaughn JC, Perry BA, Bagshaw JC (1987). Interspersion of histone and 5S RNA genes in *Artemia*. *Gene* 51: 61-67.

Austin JW, Szalanski AL, Uva P, Bagnères A-G, Kence A. (2002). A comparative genetic analysis of the subterranean termite genus *Reticulitermes* (Isoptera: Rhinotermitidae). *Annals of the Entomological Society of America* 95: 753-760.

Bagnères A-G, Clément J-L, Blum MS, Severson RF, Joulie C, Lange J (1990). Cuticular hydrocarbons and defensive compounds of *Reticulitermes flavipes* (Kollar) and *R. santonensis* (Feytaud) polymorphism and chemiotaxonomy. *Journal of Chemical Ecology* 16: 3213-3244

Bagnères A-G, Uva P, Clément J-L (2003) Description d'une nouvelle espèce de Termite: *Reticulitermes urbis* n. sp. (Isopt., Rhinotermitidae). *Bulletin de la Société Entomologique de France* 108: 434-436.

Baldrige GD, Fallon AM (1992). Primary structure of the ribosomal DNA intergenic spacer from the mosquito, *Aedes albopictus*. *DNA and Cell Biology* 11: 51-59.

Barzotti R, Pelliccia F, Bucciarelli E, Rocchi A (2000). Organization, nucleotide sequence, and chromosomal mapping of a tandemly repeated unit containing the four core histone genes and a 5S rRNA gene in an isopod crustacean species. *Genome* 43: 341-435.

Bassi C, Magnani I, Sacchi N, Saccone S, Ventura A, Rocchi M, Marozzi A, Ginelli E, Meneveri R (2000). Molecular structure and evolution of DNA sequences located at the alpha satellite boundary of chromosome 20. *Gene* 256: 43-50.

Braband A, Richter S, Hiesel R, Scholtz S (2002). Phylogenetic relationships within the Phyllopoda (Crustacea, Branchiopoda) based on mitochondrial and nuclear markers. *Molecular Phylogenetics and Evolution* 25: 229-244.

Bradley RD, Wichman HA (1994). Rapidly evolving repetitive DNAs in a conservative genome: a test of factors that affect chromosomal evolution. *Chromosome research* 2: 354-360.

Burton RS, Metz EC, Flowers JM, Willet CS (2005). Unusual structure of ribosomal DNA in the copepod *Tigriopus californicus*: intergenic spacer sequences lack internal subrepeats. *Gene* 344: 105-113.

Campadelli G (1987). Prima segnalazione di *Reticulitermes lucifugus* Rossi per la Romagna. *Bollettino dell'Istituto di Entomologia 'G. Grandi'* 42: 175-178.

Cesari M, Luchetti A, Passamonti M, Scali V, Mantovani B (2003). Polymerase chain reaction amplification of the Bag320 satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea). *Gene* 312: 289-295.

Charlesworth B, Sniegowski P, Stephan W (1994). The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371: 215-220.

Clément J-L, Bagnères A-G, Uva P, Wilfert L, Quintana A, Reinhard J, Dronnet S (2001). Biosystematics of *Reticulitermes* termites in Europe: morphological, chemical and molecular data. *Insectes Sociaux* 48: 202-215.

Cooper JL, Henikoff S (2004). Adaptive evolution of the histone fold domain in centromeric histones. *Molecular Biology and Evolution* 21: 1712–1718.

Crease TJ (1993). Sequence of the intergenic spacer between the 28S and 18S rRNA-encoding gene of the crustacean, *Daphnia pulex*. *Gene* 134: 245-249.

Crease TJ (1995). Ribosomal DNA evolution at the population level: nucleotide variation in intergenic spacer arrays of *Daphnia pulex*. *Genetics* 141: 1327-1337.

Dover GA (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111-117.

Dover GA (1986). Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends in Genetics* 2: 159-165.

Dover GA (2002). Molecular drive. *Trends in Genetics* 18: 587-589.

Drouin G, Hofman JD, Doolittle WF (1987). Unusual ribosomal gene organization in coepods of the genus *Calanus*. *Journal of Molecular Biology* 196: 943-946

Drouin G, Moniz de Sa M (1995). The concerted evolution of 5S ribosomal genes linked to the repeats units of other multigenes families. *Molecular Biology and Evolution* 12: 481-493

Eder E (2002). SEM investigations of the larval development of *Imnadia yeyetta* and *Leptestheria dahalacensis* (Crustacea: Branchiopoda: Spinicaudata). *Hydrobiologia* 486: 39-47.

Eggleton P (2001). Termites and trees: a review of recent advances in termite phylogenetics. *Insectes Sociaux* 48: 187-193.

Elder JF, Turner BJ (1995). Concerted evolution of repetitive DNA sequences in eukaryotes. *Quarterly Review in Biology* 70: 297-320.

Earnshaw WC, Rattner JB (1989). A map of the centromere (primary constriction) in vertebrate chromosomes at metaphase. *Progress in Clinical Biology Research* 318: 33-42.

Fryer G (1985). Structure and habits of living branchiopod crustaceans and their bearing on the interpretation of fossil forms. *Transactions of the Royal Society Edinburgh* 76: 103-113.

Fryer G (1987). A new classification of the branchiopod Crustacea. *Zoological Journal of Linnean Society* 91: 357-383.

Garagna S, Pérez-Zapata A, Zuccotti M, Mascheretti S, Marziliano N, Redi CA, Aguilera M, Capanna E (1997). Genome composition in Venezuelan spiny-rats of the genus *Proechimys* (Rodentia, Echimyidae). I. Genome size, C-

heterocromatin and repetitive DNAs *in situ* hybridization patterns. *Cytogenetics and Cell Genetics* 78: 36-43.

Garagna S, Marziliano N, Zuccotti M, Searle JB, Capanna E, Redi CA (2001). Pericentromeric organization at the fusion point of mouse Robertsonian translocation chromosomes. *Proceeding of the National Academy of Sciences USA* 98: 171-175.

Gil I, Gallego ME, Renart J, Cruces J (1987). Identification of the transcriptional initiation site of ribosomal RNA genes in the crustacean *Artemia*. *Nucleic Acids Research* 15: 6007-6016.

Gorokhova E, Doeling TE, Weider LJ, Crease TJ, Elser JJ (2002). Functional and ecological significance of rDNA intergenic spacer variation in a clonal organism under divergent selection for production rate. *Proceedings of the Royal Society of London Series B* 269: 2373-2379.

Hanner R, Fugate M (1997). Branchiopod phylogenetic reconstruction from 12S rDNA sequence data. *Journal of Crustacean Biology* 17: 174-183.

Henikoff S, Ahmad K, Malik HS (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293: 1098-1102.

Henikoff S, Dalal Y (2005). Centromeric heterochromatin: what makes it unique? *Current Opinions in Genetic and Development* 15:177-184

Jenkins TM, Dean RE, Verkerk R, Forschler T (2001). Phylogenetic analyses of two mitochondrial genes and one nuclear intron region illuminate European subterranean termite (Isoptera: Rhinotermitidae) gene flow, taxonomy and introduction dynamics. *Molecular Phylogenetics and Evolution* 20: 286-293.

Korn M, Marrone F, Pérez-Bote JL, Machado M, Cristo M, Cancela da Fonseca L, Hundsdoerfer AK (2006). Sister species within the *Triops cancriformis* lineage (Crustacea, Notostraca). *Zoologica Scripta* 35: 301-322

Krause M, Hirsh D (1987). A trans-spliced leader sequence on actin mRNA in *C. elegans*. *Cell* 49: 753-76 1.

Lainé LV, Wright DJ (2003). The life cycle of *Reticulitermes* spp. (Isoptera: Rhinotermitidae): what do we know? *Bulletin of Entomological Research* 93: 267-378.

Li Y-C, Lee C, Sanoudou D, Hsu T-H, Li S-Y, Lin C-C (2000). Interstitial colocalization of two cervid satellite DNAs involved in the genesis of the Indian muntjac karyotype. *Chromosome Research* 8: 363-373.

Linder F (1952). Contributions to the morphology and taxonomy of the Branchiopoda Notostraca, with special reference to the North American species. *Proceedings of the United States National Museum* 102: 1-69.

Longhurst AR (1955). A review of the Notostraca. *Bulletin of the British Museum (Natural History), Zoology* 3: 1-57.

Lorite P, Renault S, Rouleux-Bonnin F, Bigot S, Periquet G, Palomeque T (2002). Genomic organization and transcription of satellite DNA in the ant *Aphaenogaster subterranea* (Hymenoptera, Formicidae). *Genome* 45: 609-616.

Luchetti A, Cesari M, Carrara G, Cavicchi S, Passamonti M, Scali V, Mantovani B (2003). Unisexuality and molecular drive: Bag320 sequence diversity in *Bacillus* taxa (Insecta Phasmatodea). *Journal of Molecular Evolution* 56: 587-596.

Mantovani B, Cesari M, Scanabissi F (2004). Molecular taxonomy and phylogeny of the 'living fossil' lineages *Triops* and *Lepidurus* (Branchiopoda: Notostraca). *Zoologica Scripta* 33: 367-374.

Marini M, Mantovani B (2002). Molecular relationships among European samples of *Reticulitermes* (Isoptera: Rhinotermitidae). *Molecular Phylogenetics and Evolution* 22: 454-459.

Martin JW (1992). Branchiopoda. In: *Microscopic Anatomy of Invertebrates*. Vol 9: Crustacea, pp. 25-224.

Mashkova T, Oparina N, Alexandrov I, Zinovieva O, Marusina A, Yurov Y, Lacroix MH, Kisselev L (1998). Unequal crossing-over is involved in human alpha satellite DNA rearrangements on border of the satellite domain. *FEBS Letters* 441: 451-457.

Mashkova TD, Oparina NYu, Lacroix MH, Fedorova LI, Tumeneva IG, Zinovieva IG, Kisselev LL (2001). Structural rearrangements and insertions of dispersed elements in pericentromeric alpha satellites occur preferably at kinkable DNA sites. *Journal of Molecular Biology* 305: 33-48.

Masumoto H, Sugimoto K, Okazaki T (1989). Alphoid satellite DNA is tightly associated with centromere antigens in human chromosomes throughout the cell cycle. *Experimental Cell Research* 181: 181-196.

Mestrovic N, Plohl M, Mravinac B, Ugarkovic D (1998). Evolution of satellite DNAs from the genus *Palorus* – Experimental evidence for the "library" hypothesis. *Molecular Biology and Evolution* 15: 1062-1068.

Miller WJ, Nagel A, Bachmann J, Bachmann L (2000). Evolutionary dynamics of the *SGM* transposon family in the *Drosophila obscura* species group. *Molecular Biology and Evolution* 17: 1597-1609.

Morales-Hojas R, Post RJ, Wilson MD, Cheke RA (2002). Completion of the sequence of the nuclear ribosomal DNA subunit of *Simulium sanctipauli*, with description of the 18S, 28S and the IGS. *Medical and Veterinary Entomology* 16: 386-394.

Nei M, Rooney AP (2005). Concerted and Birth-and-Death evolution of multigenes families. *Annual Review in Genetics* 39:121-52.

Nijman IJ, Lenstra JA (2001). Mutation and recombination in cattle satellite DNA: a feedback model for the evolution of satellite repeats. *Journal of Molecular Evolution* 52: 361-371.

Pelliccia F, Barzotti R, Volpi EV, Bucciarelli E, Rocchi A (1998). Nucleotide sequence and chromosomal mapping of the 5S rDNA repeat of the crustacean *Proaxellus coxalis*. *Genome* 41: 129-133.

Pelliccia F, Barzotti R, Bucciarelli E, Rocchi A (2001). 5S ribosomal and U1 small nuclear RNA genes: a new linkage type in the genome of a crustacean that has three different tandemly repeated units containing 5S ribosomal DNA sequences. *Genome* 44: 331-335.

Plateaux L, Clément JL (1984). La spéciation récente des termites *Reticulitermes* du complexe *lucifugus*. *Revue de la Faculté de Science de Tunis* 3: 179-206.

Renault S, Rouleux-Bonnin F, Periquet G, Bigot Y (1999). Satellite DNA transcription in *Diadromus pulchellus* (Hymenoptera). *Insect Biochemistry and Molecular Biology* 29: 103-111.

Rouleux-Bonnin F, Renault S, Bigot Y, Periquet G (1996). Transcription of four satellite DNA subfamilies in *Diprion pini* (Hymenoptera Symphita, Diprionidae). *European Journal of Biochemistry* 238: 752-759.

Rudert F, Bronner S, Garnier J-M, Dollé P (1995). Transcripts of opposite strands of γ -satellite DNA are differentially expressed during mouse development. *Mammalian Genome* 6: 76-83.

Ryu SH, Do YK, Hwang UW, Choe CP, Kim W (1999). Ribosomal DNA intergenic spacer of the swimming crab, *Charybdis japonica*. *Journal of Molecular Evolution* 49: 806-809.

Salser W, Bowen S, Browne D, et al. (11 co-authors) (1976). Investigation of the organization of mammalian chromosome at the DNA sequence level. *Fed. Proc.* 35: 23-35.

Sam M, Wurst W, Forrester L, Vauti F, Heng H, Bernstein A (1996). A novel family of repeat sequences in the mouse genome responsive to retinoic acid. *Mammalian Genome* 7: 741-748.

Sanders HL (1963). The Cephalocarida. Functional morphology, larval development, comparative external anatomy. *Memoirs of the Connecticut Academy of Arts and Sciences* 15: 1-80.

Scanabissi F, Eder E, Cesari M (2005). Male occurrence in Austrian populations of *Triops cancriformis* (Branchiopoda, Notostraca) and ultrastructural observation of the male gonad. *Invertebrate Biology* 124: 57-65.

Schueler MG, Higgins AW, Rudd MK, Gustashaw K, Willard HF (2001). Genomic and genetic definition of a functional human centromere. *Science* 294: 109-115.

Slamovits CH, Cook JA, Lessa EP, Rossi MS (2001). Recurrent amplifications and deletions of satellite DNA accompanied chromosomal diversification in south american Tuco-tucos (genus *Ctenomys*, Rodentia: Octodontidae): a phylogenetic approach. *Molecular Biology and Evolution* 18: 1708-1719.

Smith GP (1976). Evolution of repeated DNA sequences by unequal crossover. *Science* 191: 528-535.

Stenderup JT, Olesen J, Glenner H (2006). Molecular phylogeny of the Branchiopoda (Crustacea) - Multiple approaches suggest a 'diplostracan' ancestry of the Notostraca. *Molecular Phylogenetics and Evolution* 41: 182-194.

Stitou S, Diaz de la Guardia R, Jimenez R, Burgos M (1999). Isolation of a species-specific satellite DNA with a novel CENP-B-like box from the North African rodent *Lemniscomys barbarus*. *Experimental Cell Research* 250: 381-386.

Thorne BL, Traniello JFA, Adams ES, Bulmer M (1999). Reproductive dynamics and colony structure of subterranean ermites of the genus *Reticulitermes* (Isoptera Rhinotermitidae): A review of the evidence from behavioural, ecological, and genetic studies. *Ethology Ecology and Evolution* 11: 149-169.

Tinti F, Scanabissi F (1996). Reproduction and genetic variation in clam shrimps (Crustacea, Branchiopoda, Conchostraca). *Canadian Journal of Zoology* 74: 824-832.

Trusheim F (1938). Triopsiden aus dem Keuper-Frankens. *Palaeontologische Zeitschrift* 19: 198-216.

Ugarkovic D (2005). Functional elements residing within satellite DNAs. *EMBO Reports* 6: 1035-1039.

Ugarkovic D, Plohl M (2002). Variation in satellite DNA profiles: causes and effects. *EMBO Journal* 21: 5955-5959.

Vahidi H, Curran J, Nelson DW, Webster JM, McClure MA, Honda BM (1988). Unusual sequences, homologous to 5s RNA, in ribosomal DNA repeats of the nematode *Meloidogyne arenaria*. *Journal of Molecular Evolution* 27: 222-227.

Varadaraj K, Skinner DM (1994). Cytoplasmic localization of transcripts of a complex G+C-rich crab satellite DNA. *Chromosoma* 103: 423-431.

Whicman HA, Payne SS, Reeder TW (1991). Intrageneric variation in repetitive sequences isolated by phylogenetic screening of mammalian genomes. In (M. Clegg, S. J. O'Brien eds.) *Molecular Evolution*. New York: Alan R. Liss Inc., 153-160.

Wingstrand KG (1978). Comparative spermatology of the Crustacea Entomostraca. 1. Subclass Branchiopoda. Det Kongelige Danske Videnskabernes Selskab Biologiske Skrifter 22: 1-66.

Wu CCN, Fallon AM (1998). Analysis of a ribosomal DNA intergenic spacer region from the yellow fever mosquito, *Aedes aegypti*. *Insect Molecular Biology* 7: 19-29.

Zaffagnini F, Trentini M (1980). The distribution and reproduction of *Triops cancriformis* (Bosc) in Europe (Crustacea, Notostraca). *Monitore Zoologico Italiano* 14: 1-8.

Chapter 2. RET76 satellite DNA dynamics in termites

The study of the RET76 satellite DNA family in the European *Reticulitermes* taxa is here given together with the preliminary investigations on *Reticulitermes* taxonomy, phylogenetic relationships and timing of cladogenetic events. The evolution of satI and satII RET76 sub-families is discussed in the light of the resulting systematics, taking into account the eusocial way of life of these insects.

Results are presented in the following papers:

Luchetti A, Trenta M, Mantovani B, Marini M (2004). Taxonomy and phylogeny of North Mediterranean *Reticulitermes* termites (Isoptera, Rhinotermitidae): a new insight. *Insectes Sociaux* 51: 117-122.

Luchetti A, Marini M, Mantovani M (2005). Mitochondrial evolutionary rate and speciation in termites: data on European *Reticulitermes* taxa (Isoptera, Rhinotermitidae). *Insectes Sociaux* 52: 218-221.

Luchetti A, Marini M, Mantovani B (2006). Non-concerted evolution of the RET76 satellite DNA family in *Reticulitermes* taxa (Insecta, Isoptera). *Genetica* 128:123–132.

The results of this part of the thesis were also presented at the following symposia/Congresses:

Mantovani B, Luchetti A, Marini M (2004). Le termiti eusociali del genere *Reticulitermes* (Isoptera, Rhinotermitidae): rilevanza dei marcatori mitocondriali e del DNA satellite per la tassonomia, filogenesi e speciazione dei taxa europei. *Atti dell'Accademia Nazionale Italiana di Entomologia*, LII: 153-170.

Luchetti A, Bergamaschi S, Marini M, Mantovani B (2005). Analisi molecolare di campioni di *Reticulitermes* (Isoptera: Rhinotermitidae) dell'areale balcanico. XI° Convegno Nazionale A.I.S.A.S.P. Sezione Italiana I.U.S.S.I. - International Union for the Study of Social Insects, Firenze, 1-2-3- Febbraio.

Research article

Taxonomy and phylogeny of north mediterranean *Reticulitermes* termites (Isoptera, Rhinotermitidae): a new insight

A. Luchetti, M. Trenta, B. Mantovani* and M. Marini

Dipartimento Biologia Evoluzionistica Sperimentale, Via Selmi 3, 40126 Bologna, Italy, e-mail: barman@alma.unibo.it

Received 8 May 2003; revised 1 July and 28 August 2003; accepted 24 September 2003.

Summary. The molecular characterisation of 18 new populations of *Reticulitermes* is here presented for COII and 16S genes; results are elaborated and compared to all available ones, at distances and gene trees (Maximum Parsimony, Maximum likelihood, Bayesian analysis) levels. Within the *R. lucifugus* complex, a subspecific rank of differentiation appears tenable for Italian (*R. lucifugus lucifugus*, *R. lucifugus corsicus*) and European (*R. lucifugus banyulensis*, *R. lucifugus grassei*) taxa; a subspecific differentiation emerges also for the Sicilian samples. The existence of two different entities in the area formerly defined as inhabited by *R. balkanensis* is demonstrated; The north-eastern Italian *Reticulitermes* sp. is found to be more widely distributed in northern and south-eastern Italy and shows a close relationship to the sample from Peloponnese; the GenBank sample from continental Greece, on the contrary, appears more related to other eastern taxa such as *R. lucifugus* from Turkey and *R. clypeatus* from Israel; the distribution and differentiation of eastern *Reticulitermes* taxa are explained through the role that southern Balkans should have played as glacial refuge. Present data also evidences instances of anthropic involvement in taxa distribution; for one of these, the importation of at least a family group is taken into account.

Key words: *Reticulitermes*, termites, cytochrome oxidase II, 16S, parsimony, likelihood, Bayesian analysis.

Introduction

The genus *Reticulitermes* comprises worldwide distributed subterranean termites of both ecological and economical importance: besides their natural habitat, they also adapt well to urban areas, thereby constituting one of the major insect pests of wooden structures (Austin et al., 2002). Their distribution is firstly due to paleogeographic events, but it is also

deeply influenced by human activities (Verkerk and Lainé, 2000; Jenkins et al., 2001).

Both taxonomic ranks and phylogenetic relationships within this genus are highly debated especially in the Mediterranean area where seven species are at present reported (Clément et al., 2001).

R. lucifugus is widely distributed with a series of entities mainly described in the past as subspecies: *R. lucifugus balkanensis* in the Balkans, *R. lucifugus lucifugus* in continental Italy, *R. lucifugus corsicus* in Corsica, Sardinia and Tuscany, *R. lucifugus grassei* in south-western France, Spain, Portugal and Devon (U.K.), *R. lucifugus banyulensis* in southern France and north-eastern Spain. *R. lucifugus grassei* and *R. lucifugus banyulensis* are connected through the Iberian peninsula by a series of interfertile populations (Plateaux and Clément, 1984; Bagnères et al., 1990). *R. lucifugus balkanensis*, *R. lucifugus grassei* and *R. lucifugus banyulensis* have recently been raised to species level by Clément et al. (2001).

R. santonensis is limited to a small area on the French Atlantic coast. Since its first description by Feytaud (1924) the taxonomic status of this taxon has repeatedly changed, being alternatively considered as a distinct species, a form of *R. flavipes* imported in Europe, or a subspecies of *R. lucifugus* (review in Lozzia, 1990). Recent molecular analyses agree in its derivation from north American *R. flavipes* samples introduced in Europe through commerce (Jenkins et al., 2001; Marini and Mantovani, 2002). Records indicate that *R. santonensis* expanded its range during the 20th century along trade corridors such as canals and railroads (Verkerk and Lainé, 2000).

Reticulitermes sp. (sensu Clément et al., 2001) has been indicated as a new entity found in urban habitats in north-eastern Italy and south-eastern France. Its genetic differentiation had already emerged on the basis of morphological (Campadelli, 1987) and mitochondrial DNA analyses (Marini and Mantovani, 2002). Uva (2002) proposed to describe this taxon as a new species.

Finally, *R. clypeatus* is the specific name given by Lash (1952) on a morphological basis to termites collected in Jerusalem.

* Author for correspondence.

Recently, the 677 bp of the COII gene were utilised for a wide analysis on 21 north American, European and Asiatic *Reticulitermes* taxa (Austin et al., 2002). As far as the Mediterranean entities are concerned and following the terminology proposed by Clément et al. (2001), this study evidenced three main clusters: one comprising *R. grassei*, one with Italian *R. lucifugus*, *R. lucifugus corsicus* and *R. banyulensis*, and a third one with *R. balkanensis*, *R. clypeatus* and *R. lucifugus* from Turkey. It is to be noted that the Turkish samples were validated as *R. lucifugus* using a cuticular hydrocarbon analysis by gas chromatography-mass spectrometry. *R. santonensis* is here suggested as the result of some limited hybridisation event in France from introduced *R. flavipes*. *R. sp.* from Domène (France), corresponding to the newly described *R. urbis* (Uva, 2002), is explained either as the result of hybridisation events or as an introduced *Reticulitermes* form (Austin et al., 2002).

In the present paper we present the mitochondrial DNA characterisation of newly sampled, mainly Italian, *Reticulitermes* populations using the cytochrome oxidase subunit II (COII) and the large ribosomal subunit (16S) genes. As evidenced in our previous paper (Marini and Mantovani, 2002), the two mitochondrial markers show a different resolution power, COII sequences being more variable than 16S ones; more consistent results are therefore obtained when the two data sets are compared. The main aims of the present work are to verify the taxonomic status of the entities under study and to discover the dynamics of the evolutionary processes that led to present taxa distribution and differentiation. A particular focus will be on *Reticulitermes sp.* (sensu Clément et al., 2001), which in previous analyses (Marini and Mantovani, 2002), clearly appeared a possible representative of an eastern taxon, whose transadriatic distribution was to be verified.

Materials and methods

Analysed specimens, either field caught or taken from laboratory stocks, were preserved in absolute alcohol or frozen at -80°C until they were used for molecular investigation. All pertinent informations on samples are given in Table 1 and Fig. 1. Two individuals for each infestation site were analysed. The Peloponnese sample was determined as *R. balkanensis* on morphological bases (Clément et al., 2001). Samples of *R. lucifugus corsicus* (COR) and *R. banyulensis* (BAN) were a kind gift of J.-L. Clément.

For total DNA extraction, a single termite head was ground in a quick extraction buffer (PCR buffer 0,1 \times , SDS 0,1 \times), added with proteinase K, then frozen at -80°C , warmed at 65°C for 1 h and at 95°C for 15 min.

PCR amplification was performed in 50- μl mixture using the Taq Polymerase Recombinant kit (Invitrogen), following the kit protocol. Thermal cycling was done in a Gene Amp PCR System 2400 (Applied Biosystems) programmable cyclic reactor, using 30 of the following cycles: denaturation at 94°C for 30 s, annealing at 48°C for 30 s, extension at 72°C for 30 s. The amplified products were purified with the Nucleo Spin kit (Macherey-Nagel). Both strands were directly sequenced with the DNA sequencing kit (BigDye terminator cycle sequencing, Applied biosystems) in a 310 Genetic Analyzer (ABI) automatic sequencer. The primers for PCR amplification and sequencing were mtD-13 = TL2-J-3034 (5'-AAT ATG GCA GAT TAG TGC A-3')/mtD-20 = TK-N-3785 (5'-GTT TAA GAG ACC AGT ACT TG-3') for the COII gene, and mtD-32 = LR-J-12887 (5'-CCG GTC TGA ACT

Table 1. Analyzed samples, collecting sites and acronyms are given together with scored haplotypes. Numbers in the first column refer to map of Fig. 1

	Collecting sites	Acronyms	Haplotypes	
			COII	16S
Italy:				
1	Verona	VER	H1	H1
2	Ravenna	RAV	H2	H2
				H1
3	Lecce	LEC	H3	H2
4	Galatina	GAL	H3	H3
5	Trieste site a	TR1a	H4	H4
6	Trieste site b	TR1b	H4	H5
7	Desenzano sul Garda	DGA	H5	H6
8	Firenze Peretola	FPE	H6	H6
9	Castel Porziano	CPO	H7	H6
10	Canosa Sannita	CSA	H8	H6
11	Spezzano Albanese	SAL	H5	H7
12	Rosarno	ROS	H9	H8
13	Palermo	PAL	H10	H9
14	Agrigento	AGR	H11	H10
15	Santo Stefano Quisquina	SSQ	H11	H11
16	Capitana di Quartu	CQU	H12	H12
17	Pula Is Molas	PMO	H13	H13
Greece:				
18	Areopolis	ARE	H14	H14

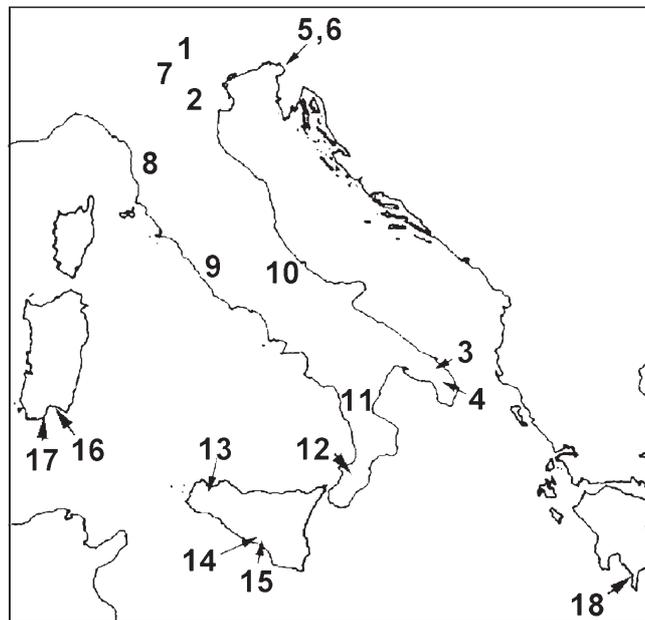


Figure 1. Map showing collecting sites of *Reticulitermes* samples (see Table 1 for reference numbers)

CAG ATC ACG T-3')/mtD-34 = LR-N-13398 (CGC CTG TTT AAC AAA AAC AT) for the 16S gene, obtained by the Biotechnology Laboratory (N.A.P.S.), Vancouver, University of British Columbia.

Sequences were aligned with CLUSTAL algorithm of the Sequence Navigator program (Version 1.0.1, Applied Biosystems), and alignments were edited by eye.

Haplotype sequences have been entered into GenBank under the accession numbers AY267857 – AY267868 (COII) and AY268356 – AY268369 (16S).

Kimura 2-parameter distances, nucleotide and amino acid compositions were obtained using Mega 2.1 package (Kumar et al., 2001). Neighbor-Joining (NJ) on Kimura 2-parameter distances, Maximum Parsimony (MP) and Maximum Likelihood (ML) analyses were performed using PAUP* program (version 4.0b, Swofford, 2001) with bootstrap evaluation corresponding to 1000, 500 and 100 replicates, respectively. For Maximum Likelihood analysis a Modeltest (version 3.06; Posada and Crandall, 1998) was run to determine the best substitution models (Hasegawa-Kishino-Yano + Gamma for 16S; Tamura-Nei + Gamma for COII and combined data set; Posada and Crandall, 1998) with the evaluation of base frequencies, R-matrix, proportion of invariable sites and value of gamma shape parameter.

Bayesian analysis was performed with MrBayes 2.01 (Huelsenbeck, 2000), setting ML parameters (lset) as follow: nst = 6 and rates = gamma (corresponding to General-Time-Reversible + Gamma model), basefreq = estimate (estimated proportion of base types from the data). The Markov chain Monte Carlo process was set so that four chains ran simultaneously for 10^6 generations with trees being sampled every 100 generations, for a total of 10^4 trees. The improvement of $-lnL$ was analysed graphically, and the 'steady state' of ML scores was determined to have occurred by the 2000th tree. Thus, the first 2000 trees were discarded (burnin = 2000, in MrBayes jargon) and a strict consensus tree was computed on the remaining 8000 trees.

For appropriate comparisons we utilised *Reticulitermes* sequences obtained in a previous paper (Marini and Mantovani, 2002) and the sequences drawn from GenBank of the following taxa: *R. flavipes* (16S: A.C. U17824; COII: A.C. AF107484); *R. speratus* (16S: A.C. D89827 – *R. speratus* 1; A.C. D89829 – *R. speratus* 2; COII: A.C. AB005463 – *R. speratus* 1, A.C. AB005584 – *R. speratus* 2); *R. clypeatus* (COII: A.C. AF525320); *R. lucifugus* (COII: A.C. AF525330 – *R. lucifugus* Turkey); *R. grassei* (COII: A.C. AF525327); *R. banyulensis* (COII: A.C. AF525319); *R. balkanensis* (COII: A.C. AF525318). *Coptotermes formosanus* was utilised as outgroup (16S: A.C. D89831 – *C. formosanus* 1, A.C. U17778 – *C. formosanus* 2; COII: A.C. AF107488, *C. formosanus*).

Results

The sequencing analysis covered a 567–685 bp fragment of the COII gene, encoding for 188–228 amino acids. The mean AT content is 62.2%. Of the 14 haplotypes scored (Table 1), haplotypes 5, 6, 7 and 8 equal the haplotypes previously found in the *R. lucifugus lucifugus* samples from Napoli, Firenze, Roma and Chieti, respectively (Marini and Mantovani, 2002). The ten new haplotypes observed differ by 3 (H10 vs H11) – 57 (H2 vs H9) substitutions; of 104 variable sites, 78,8% occurred at the third codon position, 16,4% at the first and 4,8% at the second.

Private (or diagnostic) substitutions characterise the haplotypes scored for the north-eastern samples (8 transitions, 3 transversions), the Sicilian populations (3 transitions, 1 transversion) and the Corsican/Sardinian samples (2 transitions).

Inferred amino acid sequences show 16 variable sites. In the north-eastern haplotypes, one of the two private replacements is non conservative, while in the Sicilian samples only one conservative replacement was found.

The sequencing analysis covered a 392–503 bp fragment of the 16S gene, with a mean AT content of 62%. Fourteen haplotypes were scored (Table 1). Presently determined H6 corresponds to the haplotype found in the majority of *R. lucifugus lucifugus* populations present in central Italy, while H12 equals the haplotype scored in the *R. lucifugus corsicus*

sample from Capitana di Quartu (Sardinia) (Marini and Mantovani, 2002). The twelve new haplotypes scored differ by 2 (H7 vs H8; H1 vs H2, H3, H14) – 21 (H7 vs H1, H2, H14) substitutions. Of the 35 variable sites, only one transition and one transversion are diagnostic of the Sicilian samples.

Dendrograms were built on COII haplotypes, on 16S sequences or on combined data sets, taking into account also previously determined haplotypes (Marini and Mantovani, 2002). The same terminal branching pattern can be observed in all elaborations (Fig. 2a–b). In general terms, the Verona (VER) and Ravenna (RAV) samples cluster with other north-eastern populations of peninsular Italy, as do new Sardinian (CAP, PUL) and Corsican (*R. lucifugus corsicus*) samples with previously analysed Sardinian and coastal Tuscan populations. The Rosarno (ROS) sample shows a higher affinity with *R. lucifugus lucifugus* populations of central Italy. On the other hand, the south-eastern Italian population of Galatina and Lecce (GAL, LEC) together with the Greek sample of *R. balkanensis* (ARE) appear highly related to the north-eastern Italian taxon (cluster 1). The Sicilian populations (PAL, AGR, SSQ) and the north-easternmost Italian samples of Trieste (TR1a,b), quite unexpectedly, cluster together, being less differentiated from the *R. lucifugus lucifugus* group.

Kimura 2-parameter distances were computed for all data sets (available from the Authors); Neighbor-Joining analyses on these distance values gave dendrograms (not shown) fully comparable to the MP ones.

The ML tree on all available COII sequences (Fig. 2a) show *R. speratus* as the most basal taxon followed by a dichotomy between north-eastern Italian samples/GAL/ARE (cluster 1) and all other samples; among these, the first group to emerge is the well differentiated *R. santonensis/R. flavipes* cluster followed by a geographically partitioned *R. lucifugus* cluster. To be noted is the subclustering (even though with high genetic divergence) of *R. grassei* and *R. banyulensis*. The MP dendrogram (not shown) agrees in the terminal branching pattern of the main groups, with the only exception of the splitting of *R. banyulensis* and *R. grassei* haplotypes. Further, the *R. santonensis/R. flavipes* cluster becomes basal to a tritomy among *R. speratus* haplotypes, cluster 1 and the geographically subdivided *R. lucifugus* cluster.

The 16S ML (Fig. 2b) and MP (not shown) trees agree in a polytomic topology among i) the *R. lucifugus lucifugus/R. lucifugus corsicus* cluster, ii) the Sicilian and Trieste haplotypes and iii) all other sequences; this main cluster splits in a tritomy showing a higher affinity of i) *R. banyulensis* with *R. grassei* haplotypes, ii) *R. santonensis* and *R. flavipes* sequences; iii) *R. speratus* haplotypes with cluster 1 sequences.

The ML tree on combined data sets (Fig. 2c) agrees with that obtained on COII sequences, with the only collapse of the *R. speratus* node. On the other hand, the MP tree (not shown) combines the basal position of *R. santonensis* and *R. flavipes* sequences (already shown in the MP tree on COII data set) and the clustering of *R. speratus* sequences with those of cluster 1 (already observed in 16S analyses).

A final elaboration took into account one sequence for each of the presently observed clusters plus other European haplotypes available from GenBank.

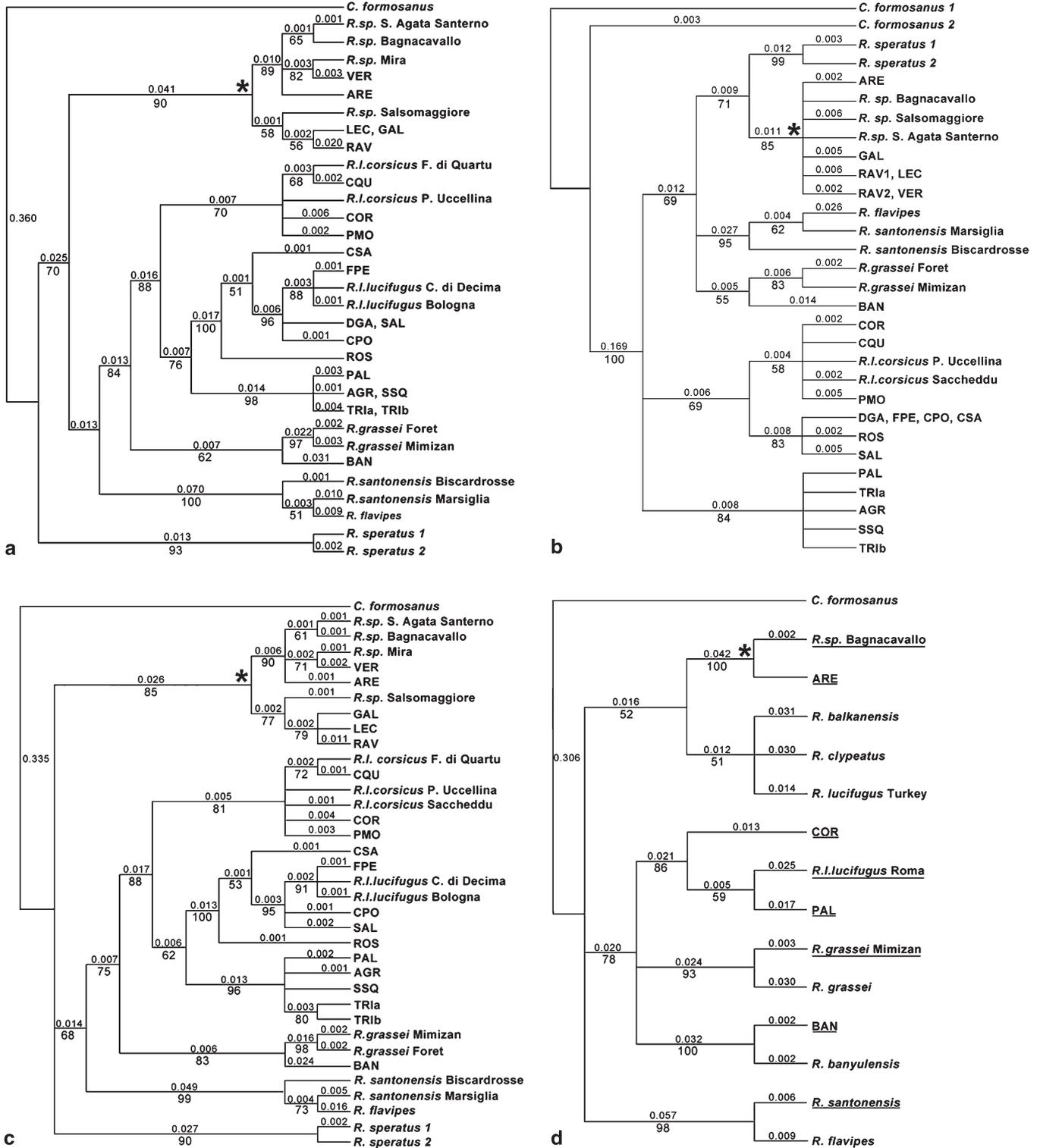


Figure 2. Maximum Likelihood trees (-Ln likelihood: a = 2424.387; b = 1355.355; c = 3837.325; d = 2411.641) obtained on COII (a, d), 16S (b) and combined sequences (c). Acronyms as in Table 1 and in Materials and methods; asterisk marks cluster 1 (see Results). Values above branches indicate the number of substitution/site, while numbers below branches represent bootstrap percentages.

The ML tree (Fig. 2d) shows three well defined groups. The first embodies north- and south-eastern Italian populations and the presently analysed *R. balkanensis* sample; these are related to – but well differentiated from – the eastern mediterranean taxa available from GenBank (*R. balkanensis* from eastern Greece, *R. lucifugus* from Turkey and *R. clypeatus*). Another cluster is built up by Italian and French *R. lucifugus* samples with a subclustering of *R. lucifugus lucifugus*, *R. lucifugus corsicus* and Sicilian *R. lucifugus*, differentiated from both *R. grassei* and *R. banyulensis* haplotypes. The third cluster comprises *R. santonensis* and *R. flavipes* sequences. In the MP dendrogram (not shown) *R. santonensis* and *R. flavipes* sequences are basal to a polytomy given by the above reported clusters.

A Bayesian strict consensus tree obtained from the same data set as figure 2d (Fig. 3) confirms the higher relationships of *R. lucifugus lucifugus*, *R. lucifugus corsicus*, Sicilian *R. lucifugus*, *R. banyulensis* and *R. grassei* (97% of posterior probability); further, it resolves the tritomy observed in the ML elaboration (Fig. 2d): eastern taxa are more differentiated from *R. lucifugus* than *R. santonensis*/*R. flavipes* and their divergence follows the geographic distribution.

Discussion

The molecular characterisation of 18 newly studied populations of *Reticulitermes* is here presented for COII and 16S genes.

For the *R. lucifugus* complex, our analyses widen the known range of *R. lucifugus lucifugus* and *R. lucifugus corsicus*, but more clearly demonstrate that a subspecific rank of differentiation should be maintained for *R. lucifugus lucifugus*, *R. lucifugus corsicus*, *R. lucifugus banyulensis* and *R. lucifugus grassei* owing to the lower differentiation scored in both ML and Bayesian analyses. It should be further noted that the reproductive isolation between the last two taxa – often recalled as demonstrative of a specific differentiation following the biological species concept (Clément et al., 2001) – is also contradicted by the repeatedly reported area of interfertile Iberian forms (Austin et al., 2002).

The present study confirms a higher affinity among Italian *Reticulitermes* with respect to other European *R. lucifugus* subspecies. It also evidences for the first time a certain degree of genetic divergence of the Sicilian samples from both *R. lucifugus lucifugus* and *R. lucifugus corsicus*. The divergence of Sicilian *Reticulitermes* had already been suggested on morphometric grounds (Lozzia, 1990). The congruence between morphological and genetic data suggests a rank of subspecific differentiation of the Sicilian populations.

Comparisons with GenBank sequences are available only for the protein codifying gene. In this regard, it should be noted that while for *R. banyulensis* and *R. grassei* both presently determined and GenBank derived haplotypes cluster together, the sample of *R. balkanensis* analysed here and the haplotype reported for apparently the same taxon in GenBank do not; this demonstrates a divergence between our sample from Peloponnese and the GenBank derived one from continental Greece.

The new entity formerly recorded in north-eastern Italy and indicated as *Reticulitermes sp.* by Clément et al. (2001) is found to be more widely distributed in northern Italy (Verona and Ravenna) and also in south-eastern Italy (Lecce and Galatina). In all elaborations, this entity shows a low genetic divergence with the presently analysed Peloponnese sample of *R. balkanensis*. The poor differentiation between *Reticulitermes sp.* and *R. balkanensis* is also evident on the basis of other approaches: the morphology and cuticular hydrocarbons are very similar and only slight differences in soldier defensive compounds were identified between the two phenotypes (Clément et al., 2001). It appears therefore clear that we are dealing with samples of the same taxon, genetically and morphologically homogeneous and well differentiated from congeneric taxa; this specific entity spreads from Peloponnese to south- and north-eastern Italy and southern France. It shows a parapatric distribution with respect to *R. lucifugus*. Whether its transadriatic distribution in southern Italy is due to natural paleogeographic events or to human activities remains to be defined. On the whole, we suggest contrary to Uva (2002), that *Reticulitermes sp.* is conspecific with *R. balkanensis* and therefore does not warrant a separate description.

As far as eastern Mediterranean entities are concerned, genetic distances and dendrogram topologies demonstrate

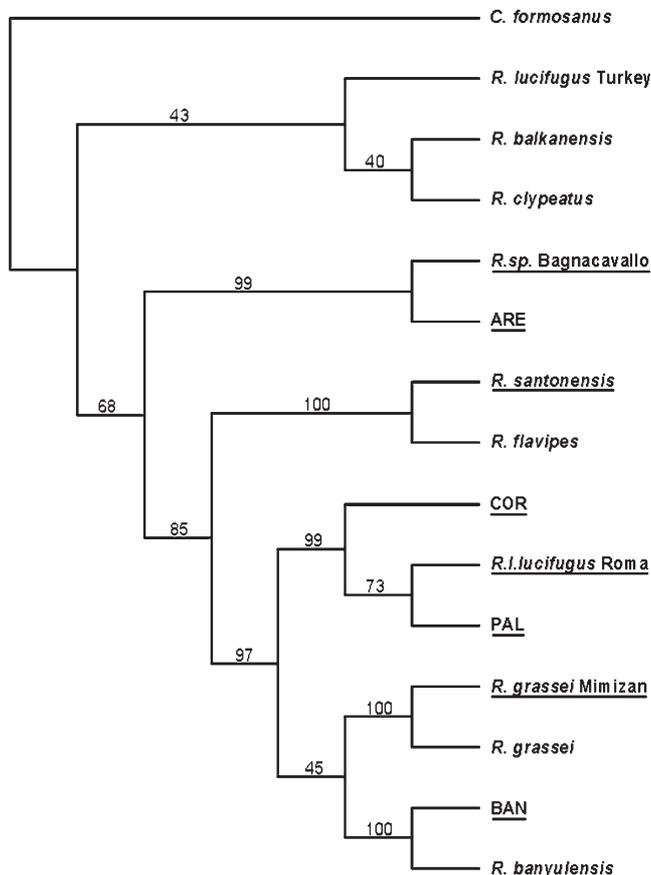


Figure 3. Strict consensus tree (contype = allcompat, in bayesian parlance) obtained with Bayesian analysis. Acronyms are as in Table 1 and in Material and methods. Numbers above the branches indicate posterior probability values of each cluster.

that the so-called '*R. lucifugus*' from Turkey does not pertain to the *R. lucifugus* complex; this was also observed by Austin et al. (2002). This taxon appears strictly related to both *R. clypeatus* from Israel and *R. balkanensis* from continental Greece; all these entities share a genetic differentiation level which is not of clear specific rank: in dendrograms, their haplotypes always cluster together, but with low bootstrap/posterior probability values (Fig. 2d, 3 respectively), suggestive of a polytomic topology. A possible explanation of present eastern taxa distribution and differentiation may be linked to the role that southern Balkans should have played as a glacial refuge (Clément et al., 2001), about 13,000 years ago; from that time onwards, the surviving termites could have spread westerly, to give the present *Reticulitermes sp.* (sensu Clément et al., 2001), and easterly producing an eastern Mediterranean complex of entities (here represented by eastern Greek *R. balkanensis*, the Turkish '*R. lucifugus*' and the Palestinian *R. clypeatus*). This hypothesis which also accounts for the above mentioned incongruence between the presently analysed *R. balkanensis* (from Peloponnese) and the *R. balkanensis* sample from GenBank (from continental Greece) could be tested with further collections in the eastern Mediterranean area.

The present data also pinpoint at least two instances of peculiar distributions: *R. lucifugus corsicus* in Tuscany and Sicilian colonies in northern Italy. As far as the former is concerned the present collections indicate its presence in the only location of Parco dell'Uccellina known from previous work (Marini and Mantovani, 2002). Even if the transthyrreanean range of *R. lucifugus corsicus* is to be widened by other samplings (see e.g. Uva et al. 2003), the spotty distribution in Tuscany together with the presence of the very same haplotypes for both COII and 16S genes in the Tuscan sample and in one Sardinian location, suggest that the presence of *R. lucifugus corsicus* in Tuscany is the outcome of human trade. This situation well matches Jenkins et al. (2001) observations on *R. lucifugus grassei* imported into the United Kingdom.

Human activities can also be considered to explain the introduction of Sicilian colonies in the Trieste area, but a quite different picture emerges here: the two Trieste samples collected in the same location at various times show two haplotypes for the 16S gene and both COII and 16S haplotypes of the Trieste colony differ from the Sicilian ones. Given that *R. lucifugus* colonies may be structured as family, tribe or population (Clément et al., 2001), it is possible to suggest the importation of at least a family group. It will be of interest to widen the Sicilian collections to verify the area of origin of the importation.

The present analyses again support the identity between *R. santonensis* and *R. flavipes* (Jenkins et al., 2001; Marini and Mantovani, 2002); no hybridisation events (Austin et al., 2002) can be taken into account to explain the relationships between the two taxa, since mitochondrial DNA in termites is, as far as we know, a pure maternally inherited genome.

With respect to the hymenopteran counterpart, it is finally to be noted that an array of genetic approaches (such as satellite DNA, microsatellite markers, and fine chromosomal charac-

terisation) are still lacking to widen our knowledge on population dynamics and reproductive biology of these eusocial insects.

Acknowledgments

This work was funded by Canziani and 'Biodiversità: livelli di scala e interazioni' (Università of Bologna) grants. We wish to thank dr Valeria Zaffagnini for sample collections

References

- Austin, J.W., A.L. Szalanski, P. Uva, A.G. Bagnères and A. Kence, 2002. A comparative genetic analysis of the subterranean termite genus *Reticulitermes* (Isoptera: Rhinotermitidae). *Ann. Entomol. Soc. Am.* 95: 753–760.
- Bagnères, A.-G., J.-L. Clément, M.S. Blum, R.F. Severson, C. Joulie and J. Lange, 1990. Cuticular hydrocarbons and defensive compounds of *Reticulitermes flavipes* (Kollar) and *R. santonensis* (Feytaud) polymorphism and chemiotaxonomy. *J. Chem. Ecol.* 16: 3213–3244.
- Campadelli, G., 1987. Prima segnalazione di *Reticulitermes lucifugus* Rossi per la Romagna. *Boll. Ist. Entomol. 'G. Grandi'* 42: 175–178.
- Clément, J. L., A.G. Bagnères, P. Uva, L. Wilfert, A. Quintana, J. Reinhard and S. Dronnet, 2001. Biosystematics of *Reticulitermes* termites in Europe: morphological, chemical and molecular data. *Insect. Soc.* 48: 202–215.
- Feytaud, J., 1924. Le termite de Saintonge. *C. R. Acad. Sci., Paris* 178: 241–244.
- Huelsenbeck, J. P., 2000. MrBayes: Bayesian inference of phylogeny, version 1.1. Available via <http://brahms.biology.rochester.edu/software.html>
- Jenkins, T.M., R.E. Dean, R. Verkerk and B. T. Forschler, 2001. Phylogenetic analyses of two mitochondrial genes and one nuclear intron region illuminate European subterranean termite (Isoptera: Rhinotermitidae) gene flow, taxonomy and introduction dynamics. *Mol. Phylogenet. Evol.* 20: 286–293.
- Kumar, S., K. Tamura, I.B. Jakobsen and M. Nei, 2001. *MEGA2: Molecular Evolutionary Genetics Analysis Software*. Arizona State University, Tempe, Arizona, USA.
- Lash, J., 1952. A new species of *Reticulitermes* from Jerusalem, Palestine. *Am. Mus. Novitates* 1575: 1–7.
- Lozzia, G.C., 1990. Indagine biometrica sulle popolazioni italiane di *Reticulitermes lucifugus* Rossi (Isoptera Rhinotermitidae). *Boll. Zool. Agr. bachic.* 22: 173–193.
- Marini, M., and B. Mantovani, 2002. Molecular relationships among European samples of *Reticulitermes* (Isoptera: Rhinotermitidae). *Mol. Phylogenet. Evol.* 22: 454–459.
- Plateaux, L. and J.L. Clément, 1984. La spéciation récente des termites *Reticulitermes* du complexe *lucifugus*. *Rev. Fac. Sc. Tunis* 3: 179–206.
- Posada, D. and K.A. Crandall, 1998. Modeltest: testing the model of DNA substitution. *Bioinformatics* 14: 817–818.
- Swofford, D.L., 2001. *PAUP* – Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Ver. 4b. Sinauer Associates, Sunderland, Massachusetts.
- Uva, P., 2002. *Relations phylogénétiques chez les termites du genre Reticulitermes en Europe. Description d'une nouvelle espèce*. Thèse de doctorat, Université F. Rabelais, Tours, pp 1–44.
- Uva, P., J.L. Clément and A.G. Bagnères, 2003. Colonial and geographic variations in behaviour, cuticular hydrocarbons and mtDNA of Italian populations of *Reticulitermes lucifugus* (Isoptera, Rhinotermitidae). *Insect. Soc.*, in press.
- Verkerk, R.H.J. and L.V. Lainé, 2000. Termites in Europe: perspective of research and management from the last century. In: XXI Int. Congr. Entomol., Brazil (D.L. Gazzoni, Ed.), p. 859.

Research article

Mitochondrial evolutionary rate and speciation in termites: data on European *Reticulitermes* taxa (Isoptera, Rhinotermitidae)

A. Luchetti, M. Marini and B. Mantovani

Dipartimento Biologia Evoluzionistica Sperimentale, Via Selmi 3, 40126, Bologna, Italy, e-mail: andluche@alma.unibo.it

Received 30 March 2004; revised 29 July and 15 November 2004; accepted 1 December 2004.

Summary. The rate of mitochondrial DNA evolution and the speciation pattern in relation to glacial periods are tested in the European taxa of the eusocial genus *Reticulitermes*. The linearized tree obtained from cytochrome oxidase II sequences and a geological event calibration shows a substitution rate 100-fold higher than that usually applied for insect mitochondrial DNA. An accelerated rate of evolution has also been observed in social Vespidae (Hymenoptera); we therefore suggest the involvement of eusociality in mediating gene pool drift. The role of the last ice age in speciation pattern of *Reticulitermes* taxa is supported by molecular data, but a four refugia model better explains genetic diversity, phyletic relationships and present-day distribution of these termites.

Key words: Glacial refugia, mitochondrial DNA, molecular clock, rate of evolution, two-cluster test.

Introduction

Mitochondrial DNA analyses are widely exploited in phylogenetic reconstruction mainly because of the practical advantages that this approach offers. This genetic compartment is also often utilised as a source of historical information. Yet, little is known about the evolutionary dynamics of mitochondrial genome in organisms deviating from canonical bisexuality, i.e. females and males freely interbreeding. In eusocial animals, clusters of individuals constitute colonies in which only a single or a few females produce offspring. This leads to a consistent bias in haplotype sampling from one generation to the next. A quicker divergence of mitochondrial genomes has been proved in Hymenoptera where the mitochondrial genomic compartment shows a significantly higher evolutionary rate in social Vespidae with respect to solitary ones (Schmitz and Moritz, 1998). This has been explained as due to the presence of a unique or few reproductive females

(queens) laying a large number of eggs, while thousands of infertile workers do not produce offspring. In such conditions, mutations accumulate and fix, while ancestral haplotypes are lost more rapidly than in solitary insects.

The heterometabolous order Isoptera comprises the largest and best known other group of eusocial insects. Among these, the subterranean termites of the genus *Reticulitermes* are the most abundant, naturally residing termites in Europe. Many taxa have been analysed for morphology, cuticular hydrocarbons, defensive compounds, behaviour, mitochondrial and nuclear DNA (Clément et al., 2001; Marini and Mantovani, 2002; Austin et al., 2002; Uva et al., 2004a,b; Luchetti et al., 2004).

The distribution of *R. lucifugus* subspecific entities follows a geographic pattern. The Iberian Peninsula and southern France host *R. lucifugus grassei* on the Atlantic coasts and *R. lucifugus banyulensis* on the Mediterranean ones. *R. lucifugus lucifugus* is present on Italian mainland, *R. lucifugus corsicus* in Corsica, Sardinia and Tuscany and *R. lucifugus* subsp. nov. in Sicily (description in progress; Lozzia, 1990; Luchetti et al., 2004) (Fig. 1). *R. lucifugus balkanensis* inhabits the Balkanic area. Recent molecular analyses suggest the existence of two genetically distinct taxa: a trans-Adriatic form, ranging from Peloponnese to north- and south-eastern Italy, and an eastern continental Greek entity (Fig. 1) (Luchetti et al., 2004). While the latter may correspond to the balkanic taxon, the taxonomic position of the former is still debated (Clément et al., 2001; Austin et al., 2002; Uva et al., 2004a; Luchetti et al., 2004). Recently, a specific rank of differentiation has been proposed for the Iberian taxa and for *R. lucifugus balkanensis* (Clément et al., 2001).

In the eastern Mediterranean basin, populations from Turkey were discovered as belonging to a new taxon with a high genetic divergence from *R. lucifugus* (Austin et al., 2002). The Israelian samples are classically known as *R. clypeatus* (Lash, 1952).

Finally, *R. santonensis* is present in a small area on the French Atlantic coast. This entity is genetically indistinguish-

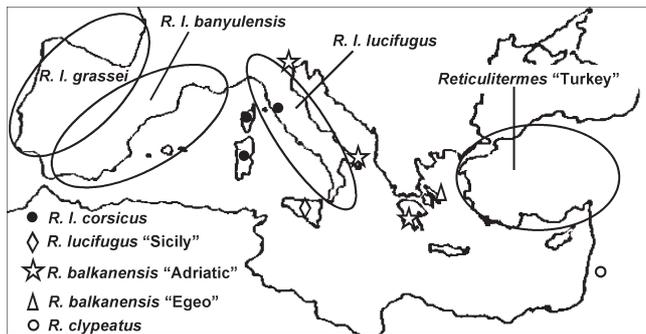


Figure 1. Map showing the present-day distribution of *Reticulitermes* taxa

able from the north-American *R. flavipes*, and its presence in Europe has been explained as the product of an anthropic introduction (Jenkins et al., 2001; Vieau, 2001; Marini and Mantovani, 2002).

Quaternary cold periods have significantly affected genome diversity in many organisms through refugia isolations and post-glacial spreading. The three Mediterranean peninsulas and the Near East are known as centres of speciation and starting points for recolonisation in different organisms (Hewitt, 2001). Biogeographic data and genetic relationships among European *Reticulitermes* termites suggest the involvement of recent climatic oscillations in their distribution and differentiation. Clément et al. (2001) have proposed two alternative hypotheses to explain the evolution of *Reticulitermes* taxa since the last ice age (Würmian era). The first model involves the three Mediterranean peninsulas as glacial refugia with the contemporary speciation of *R. balkanensis*, *R. lucifugus* and *R. l. banyulensis* – *R. l. grassei* group; the latter should have diverged approximately 8000 year ago by northward recolonisation. This could explain the genetic affinities between the Iberian taxa. The second model involves only the Balkanic and Iberian peninsulas as possible refugia: the Italian peninsula should have been recolonised after the end of the ice age from Spanish populations. This scenario may explain the similarities between *R. lucifugus*, *R. l. grassei* and *R. l. banyulensis*.

In the present paper, we analyse the data presented in Luchetti et al. (2004) through the “linearized tree” method (Takezaki et al., 1995) in order to verify the substitution rate of the mitochondrial compartment in these eusocial insects and to clarify their pattern of speciation. The analysis was extended to the Turkish and Israelian samples (Austin et al., 2002) to test the possibility of a fourth glacial refugium in the Near East. *R. santonensis* was excluded from this study given its allopatric origin.

Materials and methods

Cytochrome oxidase II (COII) sequences have been downloaded from Genbank for each European taxon. Accession numbers and collecting sites are given in Table 1.

For reasons of clarity, taxa of uncertain or still debated taxonomic status were hence indicated as follow:

Table 1. Analysed taxa and collection data (asterisks mark termites that were a kind gift of Prof. J. L. Clément)

Taxon	Sampling localities	GenBank A. N.
<i>R. l. lucifugus</i>	Chieti (Italy) Rosarno (Italy)	AF291738 AY267863
<i>R. l. lucifugus</i> “Sicily”	Palermo (Sicily) Agrigento (Sicily)	AY267857 AY267864
<i>R. l. corsicus</i>	* Parco dell’ Uccellina (Italy) Flumini di Quartu (Sardinia)	AY267858 AF291727 AF291728
<i>R. l. banyulensis</i>	* Béziers (France)	AY267859 AF525319
<i>R. l. grassei</i>	Foret (France) Mimizan (France)	AF291731 AF291733
<i>R. balkanensis</i> “Adriatic”	Areopolis (Peloponnese) Bagnacavallo (Italy) Galatina (Italy)	AY267867 AF291736 AY267862
<i>R. balkanensis</i> “Egeo”	Schinias (Eastern Greece)	AF525318
<i>Reticulitermes</i> “Turkey”	Antalya (Turkey)	AF525330
<i>R. clypeatus</i>	Ben Shemen (Israel)	AF525320
<i>C. formosanus</i>	New Orleans (U.S.A.)	AF107488

- populations from north- and south-eastern Italy and Peloponnese = *R. balkanensis* “Adriatic”
- sample from eastern Greece = *R. balkanensis* “Egeo”
- the Turkish entity = *Reticulitermes* “Turkey”.

Termites from Sicily, whose taxonomic description is still in progress, will be reported as *R. lucifugus* “Sicily”.

Neighbor-Joining (NJ) on Kimura 2-parameter distances, Maximum Parsimony (MP) and Maximum Likelihood (ML) were computed with PAUP* v. 4.0b (Swofford, 2001), with the parameters described in Luchetti et al. (2004). Linearized tree and two-cluster test (Takezaki et al., 1995) were calculated through LINTREE server program, located at web site <http://shangai.bio.psu.edu/lintree.html>.

Two methods were applied to estimate the divergence time between analysed taxa. First, an independent molecular clock was calibrated on a geological event, i.e. the last separation between Sicily and peninsular Italy at the end of the Würmian era (12 kyr). This geological event should have isolated the Sicilian entity from *R. l. lucifugus*. Furthermore, the substitution rate commonly estimated for insects mtDNA, i.e. 2.3%/Myr (Brower, 1994), was considered.

Substitution rates for 16S rDNA (Luchetti et al., 2004 and Marini and Mantovani, 2002 datasets) and for 16S rDNA - ND1 genes (Uva et al., 2004a,b) were also calculated applying the geological calibration.

Results

The MP analysis performed on COII sequences gave two equally parsimonious trees, with length equal to 266 steps (C.I. = 0.744; not shown). The bootstrap consensus tree (268 steps; C.I. = 0.739; not shown) completely agree with the NJ one (Fig. 2): the easternmost taxa (*R. balkanensis* “Egeo”, *Reticulitermes* “Turkey” and *R. clypeatus*) and the trans-Adriatic entity constitute two sister clades, well differentiated from *R. lucifugus* entities. *R. l. grassei* and *R. l. banyulensis* are basal to the cluster of Italian *R. lucifugus* sequences.

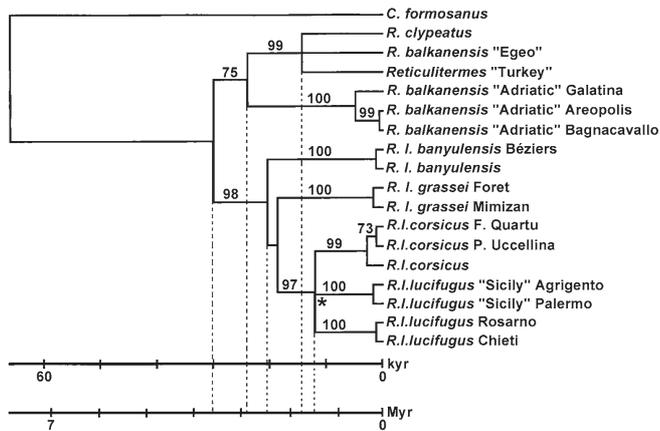


Figure 2. Schematic tree based on Neighbor-Joining and linearized tree following Takezaki et al. (1995). Bootstrap values >70% are reported above branches. The upper scale bar represents the time scale according to the geological calibration (end of the Würmian era, 12 kyr ago; asterisk marks the node to which the calibration was applied). The lower scale bar indicates the time scale obtained by applying the general molecular clock estimate (2.3%/Myr; Brower, 1994)

In this cluster, haplotypes segregate in three well differentiated groups, i.e. *R. l. lucifugus*, *R. l. corsicus* and *R. lucifugus* "Sicily". The ML tree ($-\ln L = 2199.13418$; not shown) shows the same terminal branching pattern of NJ and MP trees. The only remarkable differences are the clustering of easternmost taxa in the most basal clade and the grouping of *R. l. grassei* – *R. l. banyulensis* in a sister clade of Italian *R. l. lucifugus*. On the whole, these results are in line with those obtained in previous studies (Marini and Mantovani, 2002; Luchetti et al., 2004).

The two-cluster test evidences a substitution rate significantly constant among clusters (available from the Authors). Taking as a reference the separation between *R. lucifugus* "Sicily" and the peninsular *R. l. lucifugus* at 12 kyr ago, the COII substitution rate is equal to 0.25–0.28%/kyr. Substitution rates of the same magnitude are also scored in the other data sets considered: $\approx 0.17\%/kyr$ for the 16S rDNA gene, and $\approx 0.25\%/kyr$ for the 16S rDNA-ND1 region.

Following the COII substitution rate, the main cladogenetic event between the western- and the eastern-Mediterranean lineages should date back to 31.34 ± 3.43 kyr ago (Fig. 2).

The Iberian clades *R. l. grassei* and *R. l. banyulensis* and the Italian *R. lucifugus* have originated, almost contemporary, about 20.90 ± 3.07 kyr ago, near the Last Glacial Maximum (LGM, 18 kyr). The trans-tyrrhenian *R. l. corsicus* appears to have differentiated at the same time of *R. l. lucifugus* – *R. lucifugus* "Sicily" splitting.

About 24.97 ± 3.49 kyr ago, the ancestors of *R. balkanensis* "Adriatic" should have diverged from those of the easternmost complex *R. clypeatus* – *Reticulitermes* "Turkey" – *R. balkanensis* "Egeo". These latter entities should have differentiated each other 14.43 ± 2.52 kyr ago.

When the common estimate of insects mtDNA substitution rate (Brower, 1994) is applied, all cladogenetic events are

significantly predated. Western- and eastern-Mediterranean clades diverged 3.74 ± 0.41 Myr ago. The differentiation of the Iberian taxa from the ancestors of the Italian ones should date back to 2.49 ± 0.36 Myr, while the splitting of the Sicilian entity from the peninsular one is predated to 1.43 ± 0.26 Myr ago. The easternmost taxa, whose splitting took place 1.72 ± 0.30 Myr ago, diverged from the other eastern-Mediterranean clade, *R. balkanensis* "Adriatic", about 2.98 ± 0.41 Myr ago.

Discussion

The most interesting datum is the extraordinary discrepancy between the substitution rate usually estimated and applied for insect mtDNA (Brower, 1994) and the one scored for the COII gene with the geological calibration: the latter is nearly 100-fold higher than the former.

It is important to understand which of these calibrations is the most tenable for both the right estimation of divergence time within Mediterranean *Reticulitermes* and the long running debate on the role of glaciations in speciation.

First of all, we should consider that the high substitution rate computed for the COII gene does not appear specific of this protein coding tract: the evolutionary rates computed for other mtDNA regions (16S and 16S-ND1) are of comparable magnitude.

A second hint derives from the comparison with nuclear data on *Reticulitermes* taxa (Uva et al., 2004a): due to its repetitive nature, ITS2 is a fast evolving region undergoing concerted evolution, which should rapidly amplify the inter-taxa divergences (Dover, 2002). Yet, the sequencing of ITS2 nuclear region does not resolve the entities identified with mtDNA markers: for example, *R. clypeatus* shows the same genotype of *R. balkanensis* "Egeo". The poor variation scored (only 14 variable sites out of the 382 sites sequenced) could be the result of gene flow between *Reticulitermes* populations, either due to recent migrations or introductions. However, this hypothesis appears consistent only for sub-specific entities, but it is unlikely for those taxa showing a specific rank of differentiation. Thus, the low variability scored could be better explained as the result of recent cladogenetic events, and this hardly fit with the divergence time computed by applying the generalised mtDNA evolutionary rate (3.7 Myr–1.5 Myr). Present analyses therefore support the conclusions drawn from Hymenoptera (Schmitz and Moritz, 1998). Obviously, it is not possible to exclude that *Reticulitermes* termites could have a dramatically accelerated molecular clock because of some undefined molecular mechanism. Further, contraction and isolation of termite populations during the glacial period could have increased the haplotype drift.

On the whole, our results support a high substitution rate, mainly due to the bias in the number of reproductive individuals per generation/population. The reproductive biology of a taxon appears, therefore, to have a leading role in determining the mtDNA evolutionary rate. This is also supported by data on organisms unrelated to insects: in the

matriarchal society of macaque monkeys, characterised by few dominant females, a computer simulation showed that social and geographical population structures could significantly increase the mtDNA substitution rate (Hoelzer et al., 1998).

Concerning the pattern of *Reticulitermes* cladogenesis, our analysis confirms the role that the last cold period and the glacial refugia have played in speciation and present-day distribution of animals. Our results well fit with the climate oscillations during the last ice age and, at variance of previous hypothesis (Clément et al., 2001), point to a different timing and pattern of cladogenesis.

In particular, it appears that in the Iberian area *R. l. banyu-lensis* and *R. l. grassei* diverged during the LGM and not after the glacial period by the northern spreading of post-glacial recolonisation (Clément et al., 2001). It is assumable that Iberian taxa started to differentiate since their isolation at the opposite side of the Iberian Peninsula. As climate warmed, they newly become into contact producing the current distribution with sympatric zones.

For the Italian peninsula, *R. l. corsicus* and *R. lucifugus* “Sicily” have started to diverge contemporarily. During the LGM, only a strait separated the north-east of the Corsican-Sardinian plaque from the mainland (Thiede, 1978). It is impossible to suggest whether island colonisation has been anthropically mediated and how many of these events took place.

Finally, our data suggest the existence of two refugia in the eastern Mediterranean area, i.e. the southern Balkans – Greece and the Near East. When the climate warmed, the trans-Adriatic lineage recolonised Balkans and eastern Italy, while the easternmost one (comprising Israelian, Turkish and eastern Greek populations) could have further differentiated during a huge westward colonisation. Thus, the low genetic divergence observed between *R. balkanensis* “Egeo”, *Reticulitermes* “Turkey” and *R. clypeatus* (Luchetti et al., 2004) could be well explained by their recent separation (14.43 +/- 2.52 kyr).

In conclusion, the present paper indicates that when tackling mitochondrial markers and divergence times in the absence of a geological calibration, the reproductive biology of the studied organism must be well defined. Otherwise, the time since isolation based on the assumption of a generalised molecular clock may be consistently biased. Further, a more clear cladogenetic picture of European *Reticulitermes* taxa is put forward.

Acknowledgments

This work was funded by Canziani and “Biodiversità: livelli di scala e interazioni” (University of Bologna) grants.

References

- Austin, J.W., A.L. Szalanski, P. Uva, A.-G. Bagnères and A. Kence, 2002. A comparative genetic analysis of the subterranean termite genus *Reticulitermes* (Isoptera: Rhinotermitidae). *Ann. Entomol. Soc. Am.* 95: 753–760.
- Brower, A.V.Z., 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proc. Natl. Acad. Sci. U.S.A.* 91: 6491–6495.
- Clément, J.-L., A.-G. Bagnères, P. Uva, L. Wilfert, A. Quintana, J. Reinhard and S. Dronnet, 2001. Biosystematics of *Reticulitermes* termites in Europe: morphological, chemical and molecular data. *Insect. Soc.* 48: 202–215.
- Dover, G.A., 2002. Molecular drive. *Trends Genet.* 18: 587–589.
- Feytaud, J., 1924. Le termite de Saintonge. *Comptes Rendus Acad. Sci. Paris* 178: 241–244.
- Hewitt, G.M., 2001. Speciation, hybrid zones and phylogeography – or seeing genes in space and time. *Mol. Ecol.* 10: 537–549.
- Hoelzer, G.A., J. Wallman and D.J. Melnick, 1998. The effects of social structure, geographical structure, and population size on the evolution of mitochondrial DNA: II. Molecular clocks and the lineage sorting period. *J. Mol. Evol.* 4: 21–31.
- Jenkins, T.M., R.E. Dean, R. Verkerk and B.T. Forschler, 2001. Phylogenetic analyses of two mitochondrial genes and one nuclear intron region illuminate European subterranean termite (Isoptera: Rhinotermitidae) gene flow, taxonomy and introduction dynamics. *Mol. Phylogenet. Evol.* 20: 286–293.
- Lash, J., 1952. A new species of *Reticulitermes* from Jerusalem, Palestine. *Am. Mus. Novitates* 1575: 1–7.
- Lozzia, G.C., 1990. Indagine biometrica sulle popolazioni italiane di *Reticulitermes lucifugus* Rossi (Isoptera Rhinotermitidae). *Boll. Zool. Agr. Bachic.* 22: 173–193.
- Luchetti, A., M. Trenta, B. Mantovani and M. Marini, 2004. Taxonomy and phylogeny of north mediterranean *Reticulitermes* termites (Isoptera, Rhinotermitidae): a new insight. *Insect. Soc.* 51: 117–122.
- Marini, M. and B. Mantovani, 2002. Molecular relationships among European samples of *Reticulitermes* (Isoptera: Rhinotermitidae). *Mol. Phylogenet. Evol.* 22: 454–459.
- Plateaux, L. and J.-L. Clément, 1984. La spéciation récente des termites *Reticulitermes* du complexe *lucifugus*. *Rev. Fac. Sci. Tunis* 3: 179–206.
- Schmitz, J. and R.F.A. Moritz, 1998. Sociality and the rate of rDNA sequence evolution in wasps (Vespidae) and honeybee (*Apis*). *J. Mol. Evol.* 47: 606–612.
- Swofford, D.L., 2001. *PAUP* – Phylogenetic Analysis Using Parsimony (*and Other Methods)*. Ver. 4b. Sinauer Associates, Sunderland, Massachusetts.
- Takezaki, N., A. Rzhetsky and M. Nei, 1995. Phylogenetic test of the molecular clock and linearized trees. *Mol. Biol. Evol.* 12: 823–833.
- Thiede, J., 1978. A glacial Mediterranean. *Nature* 276: 680–683.
- Uva, P., J.-L. Clément, J.W. Austin, J. Aubert, V. Zaffagnini, A. Quintana and A.-G. Bagnères, 2004a. Origin of a new *Reticulitermes* termite (Isoptera, Rhinotermitidae) inferred from mitochondrial and nuclear DNA data. *Mol. Phylogenet. Evol.* 30: 344–353.
- Uva, P., J.-L. Clément and A.-G. Bagnères, 2004b. Colonial and geographic variations in behaviour, cuticular hydrocarbons and mtDNA of Italian populations of *Reticulitermes lucifugus* (Isoptera, Rhinotermitidae). *Insect. Soc.* 51: 163–170.
- Vieau, F., 2001. Comparison of the spatial distribution and reproductive cycle of *Reticulitermes santonensis* Feytaud and *Reticulitermes lucifugus grassei* Clément (Isoptera, Rhinotermitidae) suggests that they represent introduced and native species, respectively. *Insect. Soc.* 48: 57–62.

Non-concerted evolution of the RET76 satellite DNA family in *Reticulitermes* taxa (Insecta, Isoptera)

Luchetti Andrea, Mario Marini & Barbara Mantovani*

*Dipartimento di Biologia Evoluzionistica Sperimentale, Università di Bologna, Via Selmi 3, 40126, Bologna, Italia; *Author for correspondence (Phone: +39-51-2094169; Fax: +39-51-2094286; E-mail: barman@alma.unibo.it)*

Received 26 September 2005 Accepted 25 November 2005

Key words: concerted evolution, eusociality, Isoptera, repetitive sequence variability, *Reticulitermes*, satellite DNA

Abstract

The evolutionary dynamics of satellite DNA is most often studied in canonical mating systems, where bisexuality and panmixis are the rule. In eusocial termites, the limited number of reproducers starting a new colony and the maintenance of the colony through few neotenic act as bottle-necks both in space and time. No data on repetitive DNA are available for Isoptera and for their peculiar reproductive strategy. Here we present the first satellite DNA family isolated in European *Reticulitermes*. RET76 is a G + C rich satellite embodying two sub-families with a 76 bp monomer. RET76 sequences are highly variable (sequence homology is lower than 80% within sub-families and lower than 68% in the entire family) and this variability is equally distributed among the eight analysed taxa, thus depicting a pattern of non-concerted evolution. The absence of variant fixation – together with the strict monomer length conservation – may be explained at the molecular level as due to functional constraints acting on these sequences, and/or at the organismic level by considering the involvement of eusociality in preventing or greatly reducing variant fixation, somehow mimicking an unisexual strategy.

Introduction

Satellite DNA (satDNA) are highly, tandemly repeated sequences representing from 1 to 70% of an eukaryotic genome; they are usually organised in large heterochromatic clusters, mainly located in pericentromeric and/or telomeric regions (Charlesworth, Sniegowski & Stephan, 1994). The role of this genomic compartment is still debated. However, a variety of evidences had already demonstrated its involvement in centromere structure and dynamics, karyotypic evolution, sex/tissue specific transcription and the maintenance of heterochromatin by RNA interference (Renault et al., 1999; Henikoff, Ahmad & Malik, 2001; Schueler et al., 2001; Slamovits et al., 2001; Lorite et al., 2002b; Slamovits & Rossi, 2002;

Martienssen, 2003; Martienssen, Zaratiegui & Goto, 2005). The repeat length conservation, observed in different satellite DNAs, further suggests that these sequences may be relevant in nucleosome phasing and/or in modulating higher-order structures along the entire array (Hall, Kettler & Preuss, 2003).

The evolutionary pattern observed for repeated sequences is known as concerted evolution and leads to higher sequence homogeneity within than among evolutionary units (either strains, populations, subspecies or species). Concerted evolution is realised by molecular drive, which acts through genomic turnover mechanisms and population dynamic processes (Dover, 2002). The former involves non-reciprocal DNA exchanges within and between chromosomes, leading to variant

homogenisation within genomes; the latter determines variant fixation in different lineages through bisexuality acting as a driving force in variant fixation (Dover, 1986, 2002) followed by geographic isolation, bottle-necks, and other population processes. A well-known example of concerted evolution and of the role played by unisexuality is given by *Bacillus* stick insects (Luchetti et al., 2003).

Within insect species, sequence variability among satellite DNA monomers mainly ranges from 1 to 15%, lower or higher values being rarely observed (King & Cummings, 1997).

Beside variation in nucleotide sequence, satDNAs could also vary in copy number between related taxa (Ugarkovic & Plohl, 2002). Thus, many satellite DNA families are species-specific or can be species-specifically amplified to high copy numbers (Miller et al., 2000 and references therein; Lorite et al., 2001). Their origin is explained through the ‘expansions–contractions’ model which gave rise to the so-called ‘library hypothesis’: related taxa share a library of different satellite modules which may be differently amplified to high copy numbers during cladogenesis (Southern, 1975; Salser et al., 1976). This has been firstly demonstrated through the analyses of four satellites in congeneric *Palorus* species (Mestrovic et al., 1998; Mravinac et al., 2002) and for the Bag320 satellite in *Bacillus* stick insects (Cesari et al., 2003).

The great bulk of information on satellite DNA is found for canonical mating systems, i.e. for gonochoric and panmictic taxa. Eusociality is a way of life found in colonial insects where one or a few reproducing individuals head hundreds/thousands of sterile individuals, either queens in haplodiploid Hymenoptera or kings and queens in diplodiploid Isoptera. In many termites, in particular, a new colony starts with two alates, which will be substituted at their death by their neotenic offspring. Obviously, mechanisms reducing inbreeding depression occur and are suggested to range from the molecular level (f.e. karyotype repatterning; Fontana, 1990, 1991) to the population level (sex-biased dispersal or sex-biased alate production at the colony level; see f.i. Crosland et al., 1994; Shellman-Reeve, 1996). Yet, with respect to gonochoric non-social animals, panmictic reproduction is prevented: the limited number of reproducers starting a new colony and the

maintenance of the colony through related individuals act as bottle-necks both in space and time. An evident effect of eusociality is found for the mitochondrial genome for which an accelerated molecular clock has been found (Schmitz & Moritz, 1998; Luchetti, Marini & Mantovani, 2005).

Among Isoptera, the olarctic genus *Reticulitermes* embodies the most abundant, naturally residing subterranean termites in Europe.

From a taxonomic point of view, four species (one of which differentiated in five subspecific taxa) are recognised across Europe, identified through morphological, biochemical, behavioural and molecular analyses (Clément et al., 2001; Austin et al., 2002; Marini & Mantovani, 2002; Luchetti et al., 2004; Uva et al., 2004). In particular, the Iberian Peninsula and southern France host *R. lucifugus grassei* on the Atlantic coasts and *R. lucifugus banyulensis* on the Mediterranean ones. *R. lucifugus lucifugus* is present on Italian mainland, *R. lucifugus corsicus* in Corsica, Sardinia and Tuscany and *R. lucifugus* subsp. nov. in Sicily (description in progress; Lozzia, 1990; Luchetti et al., 2004). In the Balkanic area two entities can be recognised: *R. balkanensis*, previously identified as a *R. lucifugus* subspecies and now elevated to a specific rank (Clément et al., 2001), and a trans-Adriatic form, *Reticulitermes* sp., ranging from Peloponnese to north- and south-eastern Italy (Luchetti et al., 2004). The taxonomic position of the latter taxon remains still debated (Clément et al., 2001; Marini & Mantovani, 2002; Luchetti et al., 2004; Uva et al., 2004).

Finally, *R. santonensis* is present in a small enclave on the French Atlantic coast: this entity is genetically homogeneous with the North-American *R. flavipes*, and its presence in Europe has been explained as the consequence of anthropic introductions (Marini, Zaffagnini & Mantovani, 2000; Jenkins et al., 2001; Vieau, 2001; Marini & Mantovani, 2002).

No data on repetitive DNA are available for Isoptera and for their peculiar reproductive strategy. Here we present the characterisation of the first satellite DNA family isolated in the European termites of the genus *Reticulitermes*, with the aim to evaluate the role of eusociality on the evolutionary dynamics of repetitive DNA, in particular the effects that the absence of panmixia may have on variant fixation.

Materials and methods

Samples collection and DNA extraction

Alcohol preserved *Reticulitermes* specimens were field collected or taken from laboratory colonies; sample informations are given in Table 1. Total DNA was extracted from 10 termite heads (taken from a single colony per taxon) as described by Preiss, Hartley & Artavanis Tsakonas (1988).

Isolation of repetitive sequences

Repetitive sequences were searched in the *R. lucifugus lucifugus* genome by DNA digestions. Among the several restriction enzymes utilised, *AluI* produced a prominent, smeared band comprised between 350 and 400 bp. This band was excised from the gel, ligated to a pUC18 plasmid and used to transform *E. coli* competent cells (Invitrogen). Recombinant colonies were screened for blue–white colour (Sambrook, Fritsch & Maniatis, 1989). Fifty recombinant colonies were amplified and both strands sequenced using M13 primers and Big Dye Terminator kit (Applied). Sequences were screened by eye and BLAST search, and were found to contain some micro-satellite loci, ribosomal DNA, short interspersed elements (Luchetti, 2005) and several regions of unidentified origin: among them, a clone (Rllalu5) was found to contain four repetitions of a 76 bp sequence (A.N.: DQ205574). These repeated sequences were called RET76.

Restriction enzymes and Southern blot analyses

A majority rule consensus sequence was derived from the four monomers in order to find conserved restriction enzymes sites: only partially conserved sites for *HhaI* and *HaeIII* were found. Total DNA digestions with these restriction enzymes were performed on *R. lucifugus lucifugus*, *R. lucifugus grassei*, *Reticulitermes* sp., *R. balkanensis* and *R. flavipes*. Portions of digested DNA at 225 and 300 bp were eluted from the gel; the DNA was then cloned in a pGEM T-Easy vector (Promega) as described by Sanchez et al. (1996) and recombinant clones were identified and sequenced as described above.

Southern blot was carried out according to standard procedures (Sambrook, Fritsch & Maniatis, 1989), at medium stringency conditions ($1 \times$ SSC at 65°C). Hybridisations were carried out with appropriate probes excised from plasmid (clone Rllalu5 for RET76 satI and clone RlgI for RET76 satII; see below). Labelling and detection followed the instructions provided with the CDP Star System (Roche).

PCR analysis

In order to obtain further sequences, two primers were designed: mon1F (5'-AGW GCA GCG CCC TCA CAT-3') and mon1R (5'-MCT CTG TTC GCT YTG TCR GC-3'). In some cases, these primers amplified also a different repeat variant; for this monomer, new specific primers were

Table 1. Sample informations and mean p -Distance \pm Standard Error (p -D \pm SE) of RET76 sub-families (satI and satII)

Taxon	Sampling localities	Acronym	Mean p -D \pm SE	
			satI	satII
<i>R. l. lucifugus</i>	Roma (Italy)	Rll	0.244 \pm 0.023	0.226 \pm 0.020
<i>R. lucifugus</i> subsp. nov.	Palermo (Sicily)	Rls	0.255 \pm 0.022	0.219 \pm 0.024
<i>R. l. corsicus</i>	Lab reared ^a	Rlc	0.250 \pm 0.021	0.222 \pm 0.024
<i>R. l. banyulensis</i>	Lab reared ^a	Rlb	0.241 \pm 0.020	0.207 \pm 0.022
<i>R. l. grassei</i>	Mimizan (France)	Rlg	0.260 \pm 0.024	0.208 \pm 0.020
<i>Reticulitermes</i> sp.	Bagnacavallo (Italy)	Rsp	0.240 \pm 0.022	0.197 \pm 0.022
<i>R. balkanensis</i>	Shinias (Greece)	Rbk	0.248 \pm 0.023	0.238 \pm 0.022
<i>R. flavipes</i>	New Orleans (USA)	Rfl	0.281 \pm 0.022	0.203 \pm 0.024
Overall mean p -D \pm SE			0.254 \pm 0.013	0.216 \pm 0.012

^aTermites kindly sent by Prof JL Clément.

therefore designed: mon2F (5'-CAG TGA CTG AGV CCA CGN CGA C-3') and mon2R (5'-GTC TCG CCT CTB TTC CTT TK-3'). On the whole, we therefore found sequences pertaining to two different classes of variants (sub-families); these were then indicated as RET76 satI and RET76 satII, respectively.

Thermal cycling was carried out in a Gene Amp PCR System 2400 (Applied Biosystems), using the following program: initial denaturation at 94°C for 5 min, 25 cycles of 30 s at 94°C for, 30 s at 48°C, 1 min at 72°C, and a final extension for 7 min at 72°C. PCR amplifications were performed on all available taxa of the *Reticulitermes* genus and from the evident ladder obtained, bands corresponding to trimers, tetramers and pentamers were eluted from the gel, cloned and sequenced as described above. Analysed sequences were submitted to Genbank under the Accession Numbers: DQ205575–DQ205610 (satI) and DQ205611–DQ205640 (satII).

Sequence analyses

Concatenated monomers were first separated and primers, together with partial monomers, cut off. Sequences were aligned with the CLUSTAL algorithm (Sequence Navigator program v.1.1; Applied Biosystem). The level of differentiation between restriction- and PCR-obtained monomers was calculated with Arlequin v. 2000 software (Schneider, Roessli & Excoffier, 2000).

Average monomer length, nucleotide composition and *p*-distances were calculated with MEGA 2.1 software (Kumar et al., 2001); putative gene conversion events were scored with DnaSP 3.1 (Rozas & Rozas, 1999).

Neighbour Joining and Maximum Parsimony dendrograms were computed using PAUP* v. 4.0b8a (Swofford, 2001), with 2000 bootstrap replicates.

Nucleotide variation at each position was evaluated with the method of Strachan, Webb and Dover (1985), reintroduced by Pons, Petipierre and Juan (2002) for the study of homogenisation and fixation of repetitive sequences. This method allows to describe the transition stages at each position by distinguishing them into six classes. Class 1 includes sites identical among monomers of the two examined taxa, i.e. sites unchanged from the ancestor. Class 2 represents sites at which

mutations with a frequency below 50% occurred in a taxon, while the other sequences remained homogeneous for the ancestral nucleotide. Classes 3–5 comprise sites which retain the ancestral stage in a taxon, while in the other taxon a mutation is present and shows a frequency equal to 50% (class 3), between 50 and 99% (class 4) and 100% (class 5). Class 6 includes sites with new mutations occurring at an already fixed nucleotide position in the given taxon. Sites that cannot be assigned to any class 1–6 fall in class *N*. Lorite et al. (2004) have further distinguished this class in *N*1 and *N*2, the former comprising shared mutations between species (already present in the ancestor) and the latter comprising all other mutations (occurred after the species splitting).

Results

Isolation of RET76 repeats

Reticulitermes genomic restrictions with *Hae*III (Figure 1(a)) and *Hha*I (not shown) gave smeared lanes, with no evidence of prominent nor faint bands. DNA was cut from the gel at the presumptive positions of trimers and tetramers (225 and 300 bp, respectively), but cloning and subsequent sequencing revealed that only one out of 70 positive clones contained 3 RET76 satI sequences from the *R. flavipes* sample.

The cloning of the amplicons obtained with mon1F/mon1R and mon2F/mon2R primers produced further 65 positive clones containing 147 complete monomers. The alignment of these sequences reveals we are dealing with two RET76 sub-families: consensus sequence comparison shows that the first variant (satI) presents a one-base pair insertion at position 23, and a 4 bp insertion located between positions 44 and 47 of the alignment; the second variant (satII) has the same satI monomer length owing to a tandem duplication of the first five base pairs (Figure 2).

On the whole we analysed 73 satI and 74 satII repeats obtained by PCR, plus 4 monomers from *R. lucifugus lucifugus* and 3 monomers from *R. flavipes* (satI variant in both cases) obtained through genomic restriction. Therefore, 10 satI sequences have been analysed for each taxon, while the number of satII monomers sequenced ranges from 7 to 10.

tract moved from sub-family satI to satII, while the other moved from satII to satI.

RET76 sequence diversity among species

Maximum Parsimony analysis fails to identify specific clusters in both sub-families: repeat units intermingle in an almost complete polytomy, disregarding species, subspecies and isolation procedure (DNA restriction/PCR amplification; Figure 3). Dendrogram built with distance method (Neighbour Joining; not shown) completely agree with the parsimony tree.

The results obtained with the Strachan, Webb and Dover method (1985; Table 2) well reflect those observed in the phylogenetic elaborations. Only 1.3% of satI mutations, observed in the *R. l.*

lucifugus-R. lucifugus subsp. comparison, fall in class 3, while all the other mutations of both satI and satII fall in class 1, 2 and N, with a higher percentage in classes 2 and N2. These represent nucleotide substitutions occurred after cladogenesis.

Discussion

In this study we describe the G+C rich RET76 satellite DNA family, isolated from the *Reticulitermes* genome, and represented by two sub-families.

The most interesting datum is the distribution of variability at both satellite sub-families and taxa levels. This condition reflects the expected pattern

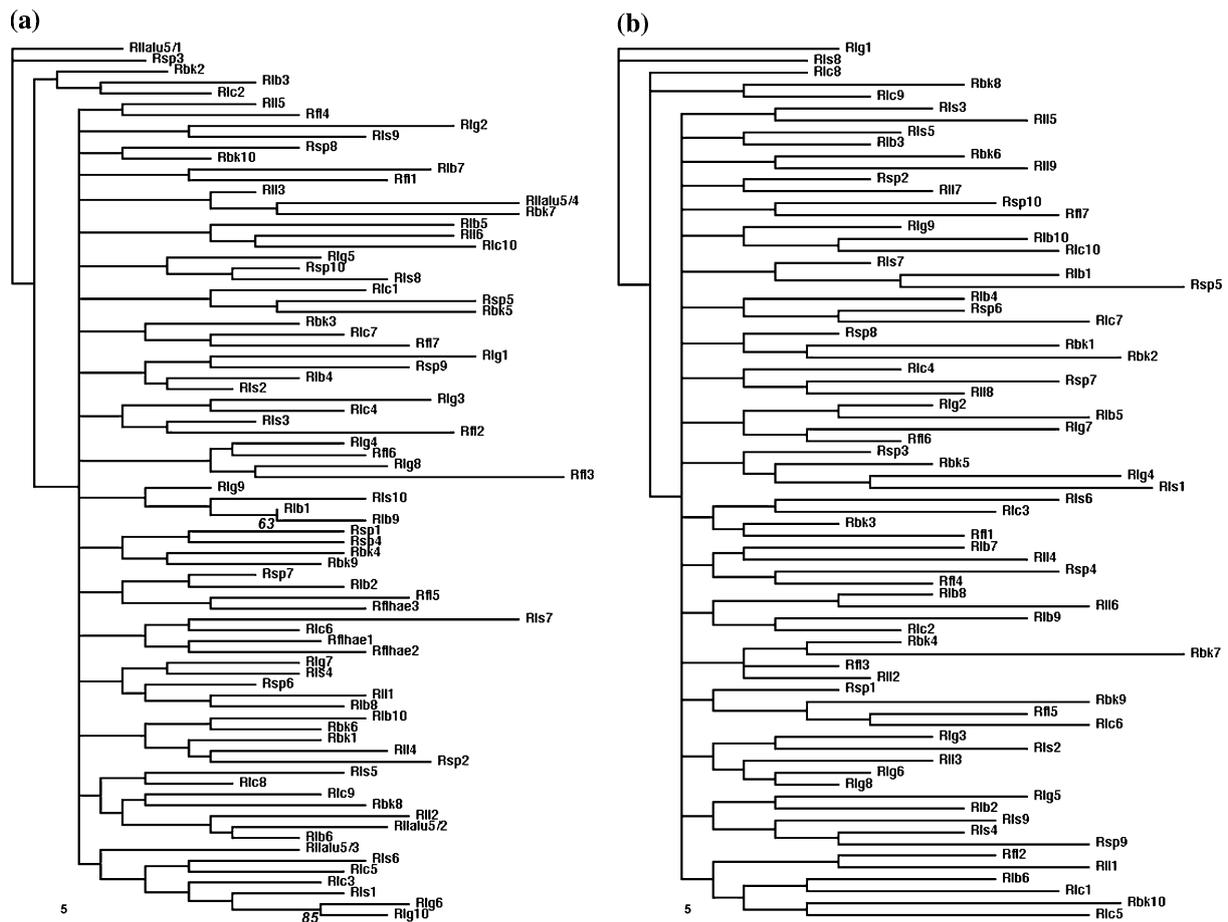


Figure 3. Maximum Parsimony bootstrap consensus trees for: (a) RET76 satI (T.L.: 722; C.I.: 0.273), and (b) RET76 satII (T.L.: 564; C.I.: 0.346). Rllalu5/1-4 and Rflhae1-3 are monomers from restriction-obtained clones. Italic numbers at nodes represent bootstrap values.

Table 2. Nucleotide variation of satI and satII sub-families evaluated following Strachan, Webb and Dover (1985)

		satI								satII							
		1	2	3	4	5	6	N1	N2	1	2	3	4	5	6	N1	N2
Rll	Rlc	3.9	27.6	0.0	0.0	0.0	0.0	5.3	63.2	6.6	27.6	0.0	0.0	0.0	0.0	19.7	46.1
	Rls	5.3	27.6	1.3	0.0	0.0	0.0	14.5	51.3	10.5	19.7	0.0	0.0	0.0	0.0	5.3	64.5
	Rlb	6.6	32.9	0.0	0.0	0.0	0.0	17.1	43.4	11.8	28.9	0.0	0.0	0.0	0.0	7.9	51.3
	Rlg	7.9	22.4	0.0	0.0	0.0	0.0	13.2	56.6	12.3	33.3	0.0	0.0	0.0	0.0	8.6	45.7
	Rsp	13.2	23.7	0.0	0.0	0.0	0.0	17.1	46.1	10.5	31.6	0.0	0.0	0.0	0.0	9.2	43.4
	Rbk	9.6	35.1	0.0	0.0	0.0	0.0	9.6	45.7	7.4	21.3	0.0	0.0	0.0	0.0	9.6	42.6
	Rfl	3.9	23.7	0.0	0.0	0.0	0.0	13.2	59.2	12.2	40.2	0.0	0.0	0.0	0.0	4.9	42.7
Rlc	Rls	6.6	19.7	0.0	0.0	0.0	0.0	21.1	52.6	7.9	30.3	0.0	0.0	0.0	0.0	9.2	52.6
	Rlb	9.2	9.2	0.0	0.0	0.0	0.0	19.7	61.8	9.2	23.7	0.0	0.0	0.0	0.0	6.6	60.5
	Rlg	9.2	15.8	0.0	0.0	0.0	0.0	17.1	57.9	4.9	42.0	0.0	0.0	0.0	0.0	9.9	43.2
	Rsp	6.6	22.4	0.0	0.0	0.0	0.0	13.2	57.9	7.9	38.2	0.0	0.0	0.0	0.0	7.9	46.1
	Rbk	4.3	41.5	0.0	0.0	0.0	0.0	16.0	38.3	5.3	20.2	0.0	0.0	0.0	0.0	14.9	40.4
	Rfl	7.9	14.5	0.0	0.0	0.0	0.0	7.9	69.7	15.9	29.3	0.0	0.0	0.0	0.0	13.4	41.5
Rls	Rlb	11.8	13.2	0.0	0.0	0.0	0.0	21.1	53.9	14.5	26.3	0.0	0.0	0.0	0.0	2.6	56.6
	Rlg	13.2	14.5	0.0	0.0	0.0	0.0	18.4	53.9	17.3	21.0	0.0	0.0	0.0	0.0	12.3	49.4
	Rsp	5.3	31.6	0.0	0.0	0.0	0.0	13.2	50.0	21.1	17.1	0.0	0.0	0.0	0.0	9.2	52.6
	Rbk	5.3	42.6	0.0	0.0	0.0	0.0	11.7	40.4	7.4	22.3	0.0	0.0	0.0	0.0	8.5	42.6
	Rfl	5.3	23.7	0.0	0.0	0.0	0.0	9.2	61.8	17.1	24.4	0.0	0.0	0.0	0.0	6.1	52.4
Rlb	Rlg	13.2	21.1	0.0	0.0	0.0	0.0	17.1	48.7	12.3	35.8	0.0	0.0	0.0	0.0	8.6	43.2
	Rsp	9.2	27.6	0.0	0.0	0.0	0.0	11.8	51.3	13.2	30.3	0.0	0.0	0.0	0.0	11.8	44.7
	Rbk	5.3	41.5	0.0	0.0	0.0	0.0	17.0	36.2	8.5	18.1	0.0	0.0	0.0	0.0	8.5	45.7
	Rfl	6.6	21.1	0.0	0.0	0.0	0.0	14.5	57.9	15.9	30.5	0.0	0.0	0.0	0.0	8.5	45.1
Rlg	Rsp	5.3	31.6	0.0	0.0	0.0	0.0	14.5	48.7	18.5	27.2	0.0	0.0	0.0	0.0	11.1	43.2
	Rbk	6.4	37.2	0.0	0.0	0.0	0.0	10.6	45.7	11.1	24.7	0.0	0.0	0.0	0.0	8.6	55.6
	Rfl	6.6	23.7	0.0	0.0	0.0	0.0	15.8	53.9	11.5	46.0	0.0	0.0	0.0	0.0	8.0	34.5
Rsp	Rbk	9.6	37.2	0.0	0.0	0.0	0.0	18.1	39.4	10.6	14.9	0.0	0.0	0.0	0.0	9.6	43.6
	Rfl	7.9	23.7	0.0	0.0	0.0	0.0	15.8	52.6	22.0	36.6	0.0	0.0	0.0	0.0	8.5	32.9
Rbk	Rfl	10.6	30.9	0.0	0.0	0.0	0.0	13.8	44.7	13.4	26.8	0.0	0.0	0.0	0.0	3.7	48.8

of variability in the absence of concerted evolution: if each monomer accumulates mutations independently, differences between repeats randomly taken from a given taxon should be the same as those observed between repeats chosen from another related taxon (Dover, 1982).

The absence of homogenisation of RET76 sequences is expressed by their high variability: their homology is lower than 80% within each sub-family and falls below 68% when we consider the entire family. As a comparison, in *Pimelia radula ascendens*, four satellite DNA sub-families show homology values ranging from 73 to 85%, and exhibit within sub-families homology up to 91% (Pons, Juan & Petitpierre, 2002). Further, RET76 variability is equally distributed among analysed

taxa: its high sequence diversity therefore cannot be ascribed to the fixation of different variants. Even if this latter feature is quite unexpected in the light of concerted evolution, it is true that in many organisms satellites are conserved across species. In particular, in eusocial Hymenoptera of the genera *Messor* and *Formica* the analyzed satellite DNAs appear unfixed, but they share values of variability lower than RET76 (Lorite et al., 2002a, 2004).

A technical bias may be invoked to explain the situation observed: the use of PCR amplification to gain RET76 monomers could have hidden the inter-taxa differences through the amplification of specific subsets of repeated sequences. However, both the high variability observed and the absence

of significance scored between monomers obtained through genomic restriction and PCR amplification suggest that monomers sampling may be unbiased by the technique utilised. Further, in many satellite studies PCR amplification is routinely applied and often it allowed to isolate different satDNAs sub-families (Bruvo et al., 2003; Cesari et al., 2003). The possibility of a bias cannot completely be ruled out, but the above reported considerations allow to suggest that such a bias has been at least minimised.

It should be considered that one of the leading forces in satellite DNA dynamics is represented by bisexual reproduction usually assumed to take place in a panmictic scenario. Chromosome re-shuffling within populations has a consistent impact on variant fixation, but eusociality hinders random matings reducing the number of reproducers to few units. In this case, the great majority of the new chromosome combinations produced in each offspring (thousand of individuals) falls in an evolutionary 'blind alley' since they cannot mix again. The new mutations occurring within non-reproductive castes do not have the possibility to spread (or to be eliminated) among individuals. In such a context, colony budding and the onset of secondary reproducers (neotomics) can introduce further satDNA variants within the population. Thus, the process of fixation could be never achieved, neither at the population level nor at the taxon one.

This would be also the case of satellite DNA conservation across the genus *Formica* (Lorite et al., 2004), since they are eusocial insects with only few reproducers among thousand of colony members. However, Authors point their attention to the haplodiploidy of these ants in order to explain the lack of both homogenisation and fixation. Actually, haplodiploidy well explains the relative low homogenisation found in satDNA, because in haploid males mutation rate can overcome the efficiency of genomic turnover mechanisms (Lorite et al., 2004).

Therefore, eusocial condition could only explain why RET76 sequences appear unfixed, but the lack of homogenisation should be the outcome of some other processes, also considering that termites are diploid.

The RET76 repeat unit length is strictly conserved between the two sub-families, as it is particularly evident when comparing the two

consensus sequences: in satII the five nucleotides lost by deletion are counterbalanced by the five-base pair duplication. This strict conservation could indicate that monomer length, rather than the nucleotide sequence itself, is involved in some function. If this is the case, the observed lack of homogenisation and fixation suggests that this repeated sequence may retain a general and conserved function across taxa. As argued for a number of satDNAs (Hall, Kettler & Preuss, 2003), this function could involve the correct positioning of nucleosomes and/or modulation of higher-order structures. On the other hand, given the well-known role of satellite DNA in chromosome repatterning (Niedermaier & Moritz, 2000; Slamovits et al., 2001), the involvement of RET76 in *Reticulitermes* translocations (Fontana, 1990, 1991) may be suggested. The gene conversion events scored also demonstrate the possibility of pairing even between satI and satII RET76 clusters. Obviously, a detailed analysis on chromosome location of RET76 loci is required to confirm this hypothesis. In particular it must be verified the possibility that the two sub-families could be located and compartmentalised in different chromosome sets (f.i. Pons, Juan & Petitpierre, 2002).

Alternatively, a different scenario may be depicted by taking into account the 'Feedback Model' (Nijman & Lenstra, 2001), describing the life history of satDNAs. Satellite DNAs should encounter three phases during their lives: in phase I interactions of homogeneous repeats cause rapid expansions as well as contractions with saltatory fluctuations in the copy number. Phase II starts with mutation and recombination events leading to independent contractions and expansions of new sequence variants: the pattern known as concerted evolution becomes evident. A satellite family may enter in the terminal phase III when degeneration by mutations stops interactions between old monomers while a younger satDNA takes their place. Applying this theory, it is possible to assume that RET76 had gone through phase II, giving origin to satI and satII; now both sub-families are possibly entering in phase III, thus are inactive and degenerating satellite DNAs. The antiquity of RET76 could be supported by the distribution of analysed species: these repeats have been isolated from European native taxa and from *R. flavipes*, which is a native North-American species. While the European taxa have originated

in the glacial refugia during the last ice age (cladogenetic events took place between 31.34 ± 3.43 kyr and 12 kyr ago, Luchetti, Marini and Mantovani, 2005), the origin of *R. flavipes* should date back up to the separation of the North-American plaque from the European one.

In conclusion, data presented here would confirm that, beside general molecular processes, the evolution of repetitive sequences is linked to the specific biology of the organism examined. In particular reproductive strategies appear to interfere with the process of the fixation: in the case of termites, the lack of panmixia could prevent mutation spreading within the population and then within the taxon. It should be recalled that in unisexual stick-insects of the genus *Bacillus* the same range of diversity observed at individual and supra-individual levels has been hypothesised as due to the lack of variant fixation in the absence of mixis (Luchetti et al., 2003)

Given that the data presented here are the first available for highly repeated sequences in eusocial diplo-diploid organisms, it is not possible to perform any kind of comparison, but the topic requires further analyses in other Isoptera taxa.

Acknowledgements

This work was supported by M.U.R.S.T. 60% and Canziani funds.

References

- Austin, J.W., A.L. Szalanski, P. Uva, A.G. Bagnères & A. Kence, 2002. A comparative genetic analysis of the subterranean termite genus *Reticulitermes* (Isoptera: Rhinotermitidae). *Ann. Entomol. Soc. Am.* 95: 753–760.
- Bruvo, B., J. Pons, J.C. Ugarkovic, E. Petitpierre & M. Plohl, 2003. Evolution of low-copy number and major satellite DNA sequences coexisting in two *Pimelia* species-groups (Coleoptera). *Gene* 312: 85–94.
- Cesari, M., A. Luchetti, M. Passamonti, V. Scali & B. Mantovani, 2003. PCR amplification of the *Bag320* satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea). *Gene* 312: 289–295.
- Charlesworth, B., P. Sniegowski & W. Stephan, 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371: 215–220.
- Clément, J.L., A.G. Bagnères, P. Uva, L. Wilfert, A. Quintana, J. Reinhard & S. Dronnet, 2001. Biosystematics of *Reticulitermes* termites in Europe: morphological, chemical and molecular data. *Insect. Soc.* 48: 202–215.
- Crosland, M.W.G., G.X. Li, L.W. Huang & Z.R. Dai, 1994. Switch to single sex alate production in a colony of the termite *Coptotermes formosanus*. *J. Entomol. Sci.* 29: 523–525.
- Dover, G.A., 1982. Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111–117.
- Dover, G.A., 1986. Molecular drive in multigene families: how biological novelties arise, spread and are assimilated. *Trends Genet.* 2(6): 159–165.
- Dover, G.A., 2002. Molecular drive. *Trends Genet.* 18(11): 587–589.
- Fontana, F., 1990. Restriction of chromosome interchanges to males of *Reticulitermes lucifugus* (Isoptera: Rhinotermitidae). *Cytobios* 63: 91–94.
- Fontana, F., 1991. Multiple reciprocal chromosomal translocations and their role in the evolution of sociality in termites. *Ethol. Ecol. Evol.* 1: 15–19.
- Hall, S.E., G. Kettler & D. Preuss, 2003. Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. *Genome Res.* 13: 119–205.
- Henikoff, S., K. Ahmad & H.S. Malik, 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293: 1098–1102.
- Jenkins, T.M., R.E. Dean, R. Verkerk & T. Forschler, 2001. Phylogenetic analyses of two mitochondrial genes and one nuclear intron region illuminate European subterranean termite (Isoptera: Rhinotermitidae) gene flow, taxonomy and introduction dynamics. *Mol. Phylogenet. Evol.* 20: 286–293.
- King, L.M. & M.P. Cummings, 1997. Satellite DNA repeat sequence variation is low in three species of burying beetles in the genus *Nicrophorus* Coleoptera: Silphidae. *Mol. Biol. Evol.* 14: 1088–1095.
- Kumar, S., K. Tamura, I.B. Jakobsen & M. Nei, 2001. MEGA2: Molecular Evolutionary Genetics Analysis software. Arizona State University, Tempe, Arizona, USA.
- Lorite, P., T. Palomeque, I. Garneria & E. Petitpierre, 2001. Characterization and chromosome location of satellite DNA in the leaf beetle *Chrysolina americana* Coleoptera, Chrysomelidae. *Genetica* 110: 143–150.
- Lorite, P., J.A. Carrillo, A. Tinaut & T. Palomeque, 2002a. Comparative study of satellite DNA in ants of the *Messor* genus. *Gene* 297(1–2): 113–122.
- Lorite, P., S. Renault, F. Rouleux-Bonnin, S. Bigot, G. Periquet & T. Palomeque, 2002b. Genomic organization and transcription of satellite DNA in the ant *Aphaenogaster subterranea* (Hymenoptera, Formicidae). *Genome* 45(4): 609–616.
- Lorite, P., J.A. Carrillo, A. Tinaut & T. Palomeque, 2004. Evolutionary dynamics of satellite DNA in species of the genus *Formica* (Hymenoptera, Formicidae). *Gene* 332: 159–68.
- Lozzia, G.C., 1990. Indagine biometrica sulle popolazioni italiane di *Reticulitermes lucifugus* Rossi (Isoptera Rhinotermitidae). *Boll. Zool. Agr. bachic.* 22: 173–193.

- Luchetti, A., 2005. Identification of a short interspersed repeat in *Reticulitermes lucifugus* (Isoptera Rhinotermitidae) genome. *DNA Seq.* 16(4): 304–307.
- Luchetti, A., M. Cesari, G. Carrara, S. Cavicchi, M. Passamonti, V. Scali & B. Mantovani, 2003. Unisexuality and molecular drive: Bag320 sequence diversity in *Bacillus* taxa (Insecta Phasmatodea). *J. Mol. Evol.* 56(5): 587–596.
- Luchetti, A., M. Trenta, B. Mantovani & M. Marini, 2004. Taxonomy and phylogeny of north mediterranean *Reticulitermes* termites (Isoptera, Rhinotermitidae): a new insight. *Insect. Soc.* 51: 117–122.
- Luchetti, A., M. Marini & B. Mantovani, 2005. Mitochondrial evolutionary rate and speciation in termites: data on European *Reticulitermes* taxa (Isoptera, Rhinotermitidae). *Insect. Soc.* 52: 218–221.
- Marini, M. & B. Mantovani, 2002. Molecular relationships among European samples of *Reticulitermes* (Isoptera: Rhinotermitidae). *Mol. Phylogenet. Evol.* 22: 454–459.
- Marini, M., V. Zaffagnini & B. Mantovani, 2000. II genere *Reticulitermes* (Isoptera: Rhinotermitidae) in Italia: un approccio molecolare. *Atti 61° Congr. Naz. UZI, S. Benedetto del Tronto*, p. 84.
- Martienssen, R.A., 2003. Maintenance of heterochromatin by RNA interference of tandem repeats. *Nat. Genet.* 35(3): 213–214.
- Martienssen, R.A., M. Zaratiegui & D.B. Goto, 2005. RNA interference and heterochromatin in the fission yeast *Schizosaccharomyces pombe*. *Trends Genet.* 21(8): 450–456.
- Mestrovic, N., M. Plohl, B. Mravinac & D. Ugarkovic, 1998. Evolution of satellite DNAs from the genus *Palorus* – Experimental evidence for the library hypothesis. *Mol. Biol. Evol.* 15: 1062–1068.
- Miller, W.J., A. Nagel, J. Bachmann & L. Bachmann, 2000. Evolutionary dynamics of the *SGM* transposon family in the *Drosophila obscura* species group. *Mol. Biol. Evol.* 17(11): 1597–1609.
- Mravinac, B., M. Plohl, N. Mestrovic & D. Ugarkovic, 2002. Sequence of PRAT satellite DNA “frozen” in some Coleopteran species. *J. Mol. Evol.* 54: 774–783.
- Niedermaier, J. & K.B. Moritz, 2000. Organization and dynamics of satellite and telomere DNAs in *Ascaris*: implications for formation and programmed breakdown of compound chromosomes. *Chromosoma* 109: 439–452.
- Nijman, I.J. & J.A. Lenstra, 2001. Mutation and recombination in cattle satellite DNA: a feedback model for the evolution of satellite repeats. *J. Mol. Evol.* 52: 361–371.
- Pons, J., C. Juan & E. Petitpierre, 2002. Higher-order organization and compartmentalization of satellite DNA PIM357 in species of the coleopteran genus *Pimelia*. *Chromosome Res.* 10: 597–606.
- Pons, J., E. Petitpierre & C. Juan, 2002. Evolutionary dynamics of satellite DNA family PIM357 in species of the genus *Pimelia* (Tenebrionidae, Coleoptera). *Mol. Biol. Evol.* 19: 1329–1340.
- Preiss, A., D.A. Hartley & S. Artavanis Tsakonas, 1988. Molecular genetics of enhancer of split, a gene required for embryonic neural development in *Drosophila*. *EMBO J.* 12: 3917–3927.
- Renault, S., F. Roulex-Bonnin, G. Periquet & Y. Bigot, 1999. Satellite DNA transcription in *Diadromus pulchellus* (Hymenoptera). *Ins. Biochem. Mol. Biol.* 29: 103–111.
- Rozas, J. & R. Rozas, 1999. DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* 15: 174–175.
- Salser, W., S. Bowen & D. Browne, et al. (11 co-authors), 1976. Investigation of the organization of mammalian chromosomes at the DNA sequence level. *Fed. Proc.* 35: 23–35.
- Sambrook, J., E.T. Fritsch & T. Maniatis, 1989. *Molecular Cloning. A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, New York.
- Sanchez, A., M. Bullejos, M. Burgos, R. Jimenez & R. Diaz, 1996. An alternative to blunt-end ligation for cloning DNA fragments with incompatible ends. *Trends Genet.* 12(2): 44.
- Schmitz, J. & R.F.A. Moritz, 1998. Sociality and the rate of rDNA sequence evolution in wasps (Vespidae) and honeybee (*Apis*). *J. Mol. Evol.* 47: 606–612.
- Schneider, S., D. Roessli & L. Excoffier, 2000. Arlequin: A software for population genetics data analysis. Ver 2.000. Genetics and Biometry Lab, Dept. of Anthropology, University of Geneva.
- Schueler, M.G., A.W. Higgins, M.K. Rudd, K. Gustashaw & H.F. Willard, 2001. Genomic and genetic definition of a functional human centromere. *Science* 294: 109–115.
- Shellman-Reeve, J.S., 1996. Operational sex ratios and lipid reserves in the dampwood termite *Zootermopsis nevadensis* (Hagen) (Isoptera: Termitidae). *J. Kansas Ent. Soc.* 69: 139–146.
- Slamovits, C.H., J.A. Cook, E.P. Lessa & M.S. Rossi, 2001. Recurrent amplifications and deletions of satellite DNA accompanied chromosomal diversification in south american Tuco-tucos genus *Ctenomys*, Rodentia: Octodontidae: a phylogenetic approach. *Mol. Biol. Evol.* 18: 1708–1719.
- Slamovits, C.H. & M.S. Rossi, 2002. Satellite DNA: agent of chromosomal evolution in mammals. A review. *J. Neotrop. Mammal.* 9(2): 297–308.
- Southern, E.M., 1975. Long range periodicities in mouse satellite DNA. *J. Mol. Biol.* 94: 51–69.
- Strachan, T., D. Webb & G.A. Dover, 1985. Transition stages of molecular drive in multiple-copy DNA families in *Drosophila*. *EMBO J.* 4: 1701–1708.
- Swofford, D.L., 2001. PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4b. Sinauer Associates, Sunderland, Massachusetts.
- Ugarkovic, D. & M. Plohl, 2002. Variation in satellite DNA profiles causes and effects. *EMBO J.* 21: 5955–5959.
- Uva, P., J.L. Clément, J.W. Austin, J. Aubert, V. Zaffagnini, A. Quintana & A.G. Bagnères, 2004. Origin of a new *Reticulitermes* termite (Isoptera, Rhinotermitidae) inferred from mitochondrial and nuclear DNA data. *Mol. Phylogenet. Evol.* 30: 344–353.
- Vieau, F., 2001. Comparison of the spatial distribution and reproductive cycle of *Reticulitermes santonensis* Feytaud and *Reticulitermes lucifugus grassei* Clément (Isoptera, Rhinotermitidae) suggests that they represent introduced and native species, respectively. *Insectes Soc.* 48: 57–62.

Chapter 3. Ribosomal intergenic spacer variability in tadpole shrimps

In this chapter, the molecular characterization of the ribosomal intergenic spacer (IGS) in the tadpole shrimp *Triops cancriformis* will be presented. Firstly, European populations with different reproductive strategies were genotyped at nuclear (microsatellite) and mitochondrial loci in order to define genetic variability and gene flow. Secondly, the IGS region was characterised and analysed both for nucleotide diversity and length heterogeneity. The possible effects of reproductive modes on the scored variability are then evaluated.

The results will be presented through the following papers:

Mantovani B, Cesari M, Luchetti A, Scanabissi F - Mitochondrial and nuclear DNA variability in the living fossil *Triops cancriformis* (Bosc, 1801) (Crustacea, Branchiopoda, Notostraca). Submitted

Luchetti A, Scanabissi F, Mantovani B (2006). Molecular characterization of ribosomal intergenic spacer in the tadpole shrimp *Triops cancriformis* (Crustacea, Branchiopoda, Notostraca). *Genome* 49: 888-893.

Luchetti A, Scanabissi F, Mantovani B - Ribosomal intergenic spacer in gonochoric population of *Triops cancriformis*: nucleotide diversity and length variation among European samples. Submitted

This work has also been presented at the following symposium:

Cesari M, Luchetti A, Scanabissi F, Mantovani B (2004). Genetic variability in *Triops cancriformis* (Bosc 1801) populations revealed by molecular markers. V° International Large Branchiopod Symposium, Toodyay, Western Australia 16-20 August.

1 **Mitochondrial and nuclear DNA variability in**
2 **the living fossil *Triops cancriformis* (Bosc, 1801)**
3 **(Crustacea, Branchiopoda, Notostraca)**

4
5 Barbara Mantovani, Michele Cesari, Andrea Luchetti, Franca
6 Scanabissi

7 Dipartimento di Biologia Evoluzionistica Sperimentale, Università di
8 Bologna, via Selmi 3, 40126 Bologna, Italy

9

10 Corresponding Author: Barbara Mantovani, Dipartimento di Biologia
11 Evoluzionistica Sperimentale, Università di Bologna, via Selmi 3,
12 40126 Bologna, Italy; Phone: +390512094169; Fax: +390512094286;
13 E-mail address: barman@alma.unibo.it

14

15 Keywords: cytochrome oxidase I, mitochondrial control region,
16 microsatellites, sexuality, tadpole shrimp, monopolization hypothesis

17

18 Running title: Genetic variability in *Triops cancriformis*

19

20 Number of words: 6868

21

1 **Abstract**

2 The living fossil *Triops cancriformis* exhibits bisexual and unisexual
3 populations, the former being either gonochoric or hermaphroditic.
4 Genetic surveys have recently revealed a general trend of low
5 differentiation of 12S and 16S mitochondrial genes. We utilise new
6 mitochondrial (COI gene and control region) and nuclear
7 microsatellite dinucleotide markers to verify genetic variability levels
8 and to correlate them with reproductive modes of different European
9 populations. The mitochondrial analyses unexpectedly confirmed the
10 pattern of low variability among *T. cancriformis* specimens, and even
11 if a high number of specimens was analysed for the 16S gene. The
12 lack of genetic variability could be explained by rejecting the
13 assumption that mitochondrial DNA evolves as a strictly neutral
14 marker, with strong effects of selection causing different fitness of
15 mitochondrial haplotypes. In microsatellite loci analyses Italian
16 populations were monomorphic or exhibited little polymorphism,
17 while other European samples displayed a higher degree of
18 polymorphism and private alleles. The clear-cut difference emerging
19 from the comparison between Spanish, Austrian and Italian samples
20 may be either linked to null allele presence, or to the peculiar gamete
21 distribution and reproductive behaviours exhibited by the different
22 populations. Sample differentiation levels and the loss of variability
23 scored for an Italian population add data to the Monopolization
24 hypothesis. Moreover, the probable presence of a sex-linked
25 microsatellite locus is discussed.

26

1 **Introduction**

2 The living fossil Notostracan *Triops cancriformis* (Bosc, 1801) lives in
3 Eurasian and North African astatic waters with bisexual and
4 unisexual populations. While the latter comprise thelytokous
5 parthenogenetic females, the former can be either gonochoric with
6 distinct female and male individuals or hermaphroditic with
7 individuals producing both male and female gametes (Trusheim,
8 1938; Longhurst, 1955; Wingstrand, 1978; Zaffagnini & Trentini,
9 1980; Fryer, 1985; Engelmann et al., 1997; Scanabissi & Mondini,
10 2002; Scanabissi et al., 2005). The sexuality of a population is
11 difficult to define on morphological grounds, the only diagnostic
12 characters being the modification of the eleventh pair of trunk
13 appendages in both sexes and the presence/absence of eggs
14 (Mathias, 1937; Fryer, 1988; Engelmann et al., 1996). These
15 characters are often misleading (hermaphrodites with eggs can be
16 mistakenly recognized as females) or inapplicable (young
17 individuals). Moreover, the diffusion of resistant eggs by means of
18 wind or birds (Figuerola et al., 2005) promotes deme intermingling,
19 so that individuals deriving from populations with different sexuality
20 may occur in the same population. A remarkable example is the
21 finding of functional males (Scanabissi et al., 2005) in a
22 hermaphroditic Austrian population (sensu Wingstrand, 1978). This
23 could be due to the introduction of new resistant eggs, but the co-
24 occurrence of males and hermaphrodites may also lead to consider
25 this *T. cancriformis* population as a possible candidate for
26 androdioecy (Pannell, 2002, Weeks et al., 2006).

1 Besides sexuality, also the taxonomy of *T. cancriformis* populations
2 has always been controversial (Ghigi, 1921, 1924; Colosi, 1923;
3 Gurney, 1923; Gauthier, 1934) owing to the high variability of
4 individual morphological characters (Longhurst, 1955; Alonso, 1985).
5 In his monograph on Notostraca, Longhurst (1955) finally
6 acknowledged the presence of three subspecies, *T. cancriformis*
7 *cancriformis*, *T. cancriformis mauritanicus* Ghigi, 1921 and *T.*
8 *cancriformis simplex* Ghigi, 1921. Among the morphological
9 characters used by Longhurst, Alonso (1985) recognized the
10 presence/absence of spines in the carapace carina as the only useful
11 taxonomic criterion for taxon identification, the other characters being
12 exceedingly variable with differences observed even between the left
13 and right sides of the same individual.

14 On a genetic ground, a recent molecular analysis on 12S and 16S
15 mitochondrial markers based on a wide taxon sampling (Korn et al.,
16 2006) indicates that *T. cancriformis* is divided into two distinct
17 lineages. One lineage comprises European *T. cancriformis*
18 *cancriformis* populations and samples from northern Spain that had
19 been classified as *T. cancriformis simplex* in the most recent
20 literature (Alonso, 1985; Alonso, 1996; Boix et al., 2002). The second
21 lineage comprises Iberian and northern African *T. cancriformis*
22 *mauritanicus* demes and northern African populations of *T.*
23 *cancriformis simplex*. The Authors therefore propose to recognize
24 them as two species, *Triops cancriformis* and *Triops mauritanicus*.
25 Even if the level of divergence emerging from the analyses appears

1 low, for our present purposes we accept the new terminology that will
2 help us in clarity.

3 While the strictly gonochoric *T. mauritanicus* shows high
4 substructuring and may include at least five subspecies, the
5 amphigonic and parthenogenetic populations of *T. cancriformis* lacks
6 diversification (Korn et al., 2006). This was already observed in a
7 previous survey within *T. cancriformis* populations (sensu Korn et al.,
8 2006), where a very low level of variability was found for the 12S and
9 16S genes (Mantovani et al., 2004). The same paper demonstrated
10 also that *T. cancriformis* was significantly differentiated from available
11 congeneric taxa, thus supporting the hypothesis that this species
12 should be ascribed to a separated genus (Linder, 1952).

13 Present study has been undertaken to verify if the low level of
14 genetic variability in *T. cancriformis* is due to previously utilized
15 molecular markers (12S and 16S) or to some specific properties of
16 the mitochondrial/nuclear genomes. Therefore, different
17 mitochondrial genes are here taken into account (the cytochrome
18 oxidase I gene and the mitochondrial control region), and the nuclear
19 genome is analysed using previously identified dinucleotide
20 microsatellite markers (Cesari et al., 2004). The former have been
21 very useful in previous Crustacean genetic studies (Remigio &
22 Hebert, 2000; Chu et al., 2003), while the latter are known for their
23 high polymorphism levels and usually represent a powerful tool for
24 genetic variability studies.

25 A particular interest in the study of Notostracan taxa derives from
26 their living in astatic waters, such as temporary pools. These habitats

1 are now seriously endangered due to urbanization, drainage and
2 global warming changes. Therefore, the study the biodiversity of
3 these organisms could help in protecting such threatened
4 environments: the presence of several endangered branchiopod
5 species (including *T. cancriformis*) was fundamental in creating
6 Austria's first National Park meeting the IUCN criteria (Neusiedler
7 See-Seewinkel Park, Eder et al., 1996). Moreover, given the
8 reproductive variability harboured by *T. cancriformis* and its ancient
9 age, this species constitutes a model for the study of sexuality and
10 reproductive modes. In this paper we analyse gonochoric,
11 hermaphroditic and parthenogenetic *T. cancriformis* samples in order
12 to gain an insight on the boundaries and relationships between
13 genetic variability and sexuality.

14

15 **Materials and Methods**

16 ***Animals***

17 Present work was carried out on seven populations. Most localities
18 were sampled once, with the exception of Grosseto and Espolla,
19 which were examined in two different years (Table 1); a total number
20 of nine samples were therefore considered.

21 In each sample, two individuals were analyzed for all four
22 mitochondrial genes (Table 1). Ten additional specimens were
23 analyzed for the 16S gene only in all samples but the Palermo one.
24 From 17 to 22 individuals were genotyped at the five microsatellite
25 loci (MSL) in six samples: for Marchegg and Oristano populations we

1 considered previous data (Cesari et al., 2004), while Palermo was
2 not taken into account because too few specimens were available.

3 Total DNA was extracted from single individuals, following CTAB
4 (Doyle & Doyle, 1987) or phenol/chloroform (Sambrook et al., 1989)
5 protocols.

6 ***Mitochondrial analyses***

7 PCR amplification was performed in 50 µl reactions using the
8 Invitrogen PCR kit with recombinant *Taq* DNA polymerase. 35 cycles
9 were scheduled as follows: denaturation at 94°C for 30 sec,
10 annealing at 48°C for 30 sec, extension at 72°C for 30 sec. The
11 amplified products were purified with the Wizard PCR cleaning
12 (Promega) kit and both strands were sequenced in an ABI PRISM
13 310 Genetic Analyzer (Applied). The primers for PCR amplification
14 and sequencing (Invitrogen) were mt-35 (5'-AAG AGC GAC GGG
15 CGA TGT GT-3') and mt-36 (5'-AAA CTA GGA TTA GAT ACC CTA
16 TTA T-3') for the 12S gene; mt-32 (5'-CCG GTC TGA ACT CAG ATC
17 ACG T-3') and mt-34 (5'-CGC CTG TTT AAC AAA AAC AT-3') for
18 the 16S gene; TCMCR-F (CCC GTC GCT CTC TCC TCT A) and
19 TCMCR-R (GCC ACA TGA TTT ACC CTA TCA AA) for the
20 Mitochondrial Control Region (MCR); COI-F (5'-GGT CAA CAA ATC
21 ATA AAG ATA TTG G-3') and COI-R (5'-TAA ACT TCA GGG TGA
22 CCA AAA AAT CA-3') for the cytochrome oxidase I (COI) gene.
23 Primers were derived from Simon et al. (1994; 12S and 16S genes),
24 from Folmer et al. (1994; COI gene) or specifically designed on the *T.*
25 *cancriformis* complete mitochondrial sequence (MCR; Genbank
26 Accession Number NC_004465). Alignments were performed with

1 the Clustal algorithm of the Sequence Navigator program (ver 1.0.1,
2 Applera) and were also checked by visual inspection. The nucleotide
3 sequences of the newly analyzed specimens have been submitted to
4 the GenBank (A.N.: DQ369307-8, 12S; DQ369309, DQ664195 and
5 EF190477-81, 16S; DQ369312-7, DQ664196, COI; AY764144-6,
6 DQ369310-1 and EF190476, MCR). For opportune comparisons,
7 homologous sequences were drawn from the complete mitochondrial
8 sequences of a Japanese *T. cancriformis* and of *T. longicaudatus*
9 LeConte, 1846 (A.N. NC_006079), the latter used as outgroup.
10 Substitutions were determined using MEGA version 3.1 (Kumar et
11 al., 2004), while Maximum Parsimony (MP) and Maximum Likelihood
12 (ML) dendrograms were computed using PAUP* 4.0b10 (Swofford,
13 2001); bootstrap values were obtained after 2000 and 200 replicates,
14 respectively. The possibility to analyse mtDNA genes as a combined
15 dataset was tested with an Incongruence Length Difference test (ILD
16 test; Farris et al., 1994, 1995), implemented on PAUP*, after 500
17 replicates. In the MP analysis, gaps were considered as fifth state.
18 For ML analysis, a Modeltest (version 3.06; Posada & Crandall,
19 1998) was run to determine the best substitution model (TVM+G),
20 with the evaluation of base frequencies, rate matrix, proportion of
21 invariable sites and value of gamma shape parameter (0.2113). A
22 parsimony network was determined for the 16S haplotypes coming
23 from 12 individuals for each population, but Palermo, by applying the
24 method of Templeton et al. (1992) as implemented in TCS 1.21
25 (Clement et al., 2000), with the gaps considered as fifth state. The
26 sequences of the COI gene were analyzed for deviations from

1 neutrality (McDonald-Kreitman test; McDonald & Kreitman, 1991)
2 using DnaSP (Rozas et al., 2003).

3 ***Microsatellites analyses***

4 Populations were genotyped at five dinucleotide microsatellite loci
5 (tcAC-8p1, tcAC-9p1, tcAC-10p1, tcAC-10p2 and tcAC-14p1)
6 following the protocols described in Cesari et al. (2004; Table 2).
7 Observed and expected heterozygosities, allelic frequencies and Nm
8 (number of migrants, following Wright, 1969 algorithm) were
9 computed using Genetix 4.05 (Belkhir et al., 2004); probability to fit to
10 Hardy-Weinberg equilibrium (HWE), linkage disequilibrium test,
11 relationship between population differentiation and geographical
12 distance, genic and genotypic differentiation were calculated using
13 Genepop 1.2 (Raymond & Rousset, 1995). Genic diversity
14 (according to Nei 1987 algorithm), allelic richness and F-statistics
15 were computed using FSTAT 2.9.3 (Goudet, 2001). F-statistics was
16 performed taking also into account previously analysed Italian and
17 Austrian samples (Cesari et al., 2004). Sample differentiation based
18 on haplotype frequencies and M values (number of migrants) were
19 calculated with Arlequin 3.0b (Excoffier et al., 2005). Given that
20 polymorphism at annealing sites of the MSL primers can prevent the
21 amplification of a particular allele, therefore resulting in heterozygote
22 deficiencies, null allele frequencies were estimated according to
23 Chakraborty et al. (1992), Brookfield (1996) and Van Oosterhout et
24 al. (2004) algorithms with Bonferroni corrections, using Microchecker
25 2.2.1 (Van Oosterhout et al. 2004).

26

1 **Results**

2 ***Mitochondrial analysis***

3 *mtDNA diversity*

4 Overall, 1816-1822 base pairs were sequenced in each individual
5 (347 bp for the 12S gene, 503-509 bp for the 16S gene, 595 bp for
6 the COI gene and 371-372 bp for the MCR), and fourteen mitotypes
7 (i.e. combined mitochondrial haplotypes) were found (Table 1). All
8 populations showed private mitotypes. A single mitotype
9 characterized Lecce and Espolla 2006. In the Grosseto pond,
10 differences were scored between samples obtained in consecutive
11 years (2002-2003), but not between specimens sampled the same
12 year. In each of the other populations, the two individuals showed
13 mitotypes diverging for one/two point mutations or for one indel, with
14 the exception of the two Marchegg individuals, which differed for 29
15 substitutions.

16 The Marchegg mitotype A was the more diverse from the other
17 mitotypes (29-33 substitutions), while mitotype B differed for only 1-
18 15 substitution(s). The comparison between Italian mitotypes showed
19 at most 4 substitutions, with the exception of the Sicilian and Apulian
20 populations, which were differentiated for 13-17 substitutions. On the
21 whole, the four Spanish specimens were more similar to the
22 Sardinian sample (two indels or 2-3 substitutions).

23 The ILD test was not significant ($P = 0.29$); therefore all mitochondrial
24 loci were analysed as a combined dataset, on a total of 1822
25 characters. MP and ML dendrograms (Fig. 1) differ in deep branching
26 topology. In the MP tree (Fig. 1A) the Austrian mitotype A is basally

1 located and two further highly supported clusters can be recognised.
2 On the other hand, the ML dendrogram (Fig. 1B) is mainly polytomic,
3 with the Austrian mitotype A occurring in the only supported cluster
4 with Apulian and Sicilian sequences. The Spanish and Sardinian
5 mitotypes appear related in the MP analysis, while their cluster
6 collapses in the ML analysis.

7 *16S gene haplotype differentiation*

8 The mitochondrial analysis was performed on ten more individuals in
9 all populations but Palermo (total number of analyzed
10 specimens=96) for the 16S gene, which was chosen as it proved to
11 be the most variable in the first part of the analysis (Table 1). The
12 newly sequenced specimens revealed five new haplotypes, four in
13 Espolla (i, j, k, l) and one in Marchegg (m), differing respectively for
14 one or two substitutions with respect to the most common haplotype
15 b.

16 The network analysis produced two different lineages, one
17 embodying only haplotype a, which was found in most Marchegg
18 individuals and in one Oristano specimen, and the other comprising
19 the remaining sequences (Fig. 2). In the latter lineage, haplotype b is
20 the most frequent, being found in the great majority of Ferrara,
21 Grosseto, Oristano and Espolla individuals, and in one Austrian
22 specimen. Lecce is the only sample presenting a single haplotype
23 (e).

24 ***Microsatellites analysis***

25 Locus tcAC10-p2 is monomorphic in all six newly genotyped
26 samples. The other loci are all polymorphic in the bisexual *Espolla*

1 samplings, with the presence of private alleles (Table 2, Fig. 3).
2 Italian samples show a decidedly low variability: in particular,
3 *Grosseto 2003* and *Lecce* are homozygous at all loci, whereas only
4 one (tcAC9-p1) or two (tcAC9-p1, tcAC10-p1) MSL are polymorphic
5 in the *Ferrara* and *Grosseto 2002* populations, respectively. In the
6 Italian populations genic diversity ranges from 0.091 to 0.279, while
7 Spanish samples retain overall higher values, varying from 0.053 to
8 0.474. Null alleles presence could not be discarded in *Ferrara* (locus
9 tcAC10-p1), *Grosseto 2002* (locus tcAC9-p1) and *Espolla 2004* (loci
10 tcAC8-p1 and tcAC9-p1) samples (Table 2). Moreover, the re-
11 analysis of the data presented in Cesari et al. (2004) with presently
12 used algorithms (Chakraborty et al. 1992 and Van Oosterhout et al.
13 2004) revealed a probable presence of null alleles in all loci (but
14 tcAC8-p1) in the *Marchegg* population.

15 No evidence of significant linkage was found among the five loci
16 ($P > 0.05$). *Ferrara* and *Grosseto 2002* samples deviate significantly
17 from HWE ($P < 0.05$ and $P < 0.001$, respectively). HW disequilibrium
18 over all loci is well reflected by highly significant F_{IS} and F_{IT} values
19 (0.451 and 0.836, respectively; $P < 0.001$). In the F_{ST} test, a significant
20 value (0.702; $P < 0.001$) points to a substantial genetic differentiation
21 over all loci.

22 The pairwise F_{ST} values (Table 3) highlight a genetic structuring both
23 between the hermaphroditic Austrian and gonochoric Spanish
24 samples and between them and the Italian populations. Among the
25 latter, the F_{ST} values point to a high differentiation of *Lecce* from the
26 other Italian populations. A significant value is also obtained in the

1 comparison between *Grosseto 2003* and *Oristano*. Genic, genotypic
2 and haplotypic frequency differentiation completely confirm this
3 pattern even though in the latter analysis the comparison between
4 *Grosseto 2003* and *Oristano* samples is not significant (data
5 available from the authors).

6 The number of migrants has been estimated following two different
7 algorithms (Table 3). Both analyses agree in showing a very low
8 number of migrants between Spanish and Austrian samples and
9 between these populations and the Italian ones. Values indicative of
10 dispersal have been scored among Italian samples with the
11 exception of *Lecce*, with the highest score found between *Ferrara*
12 and *Grosseto 2003* in the Wright's model. Population differentiation
13 confirms this aspect and appears correlated to the increasing
14 geographical distance between samples ($R^2=0.699$, $P<0.05$).

15 In the *Espolla* samples, the MSL polymorphic genotypes were also
16 considered in the light of the specimens' sex. It was noted in the
17 2004 sample a peculiar sex-linked pattern of variability at the tcAC8-
18 p1 locus: the 10 analyzed females (Table 2) presented the same
19 homozygous genotype (150/150), while the ten male specimens
20 showed either homozygous or heterozygous genotype (144/144, 5
21 males; 144/150, 4 males; 150/150, 1 male). Seventeen additional
22 females of the *Espolla 2004* sample were analyzed at this locus, and
23 all of them exhibited the 150/150 genotype. A comparable situation
24 (homozygous females and homozygous or heterozygous males) was
25 scored for the 2006 *Espolla* sampling. Further, all 92 females of the
26 Italian populations so far analyzed shared the 150/150 genotype

1 (present data; Cesari et al., 2004), while the hermaphroditic Austrian
2 sample presents homozygous or heterozygous genotypes with the
3 occurrence of a different allele (148) together with allele 150
4 (148/148, 4 individuals; 148/150, 6 individuals; 150/150, 2
5 individuals; Cesari et al., 2004). On the whole, alleles 144 or 148
6 occur only in males or hermaphrodites either in homozygous
7 (144/144; 148/148) or heterozygous (144/150; 148/150) condition,
8 while the female sex presents only the 150/150 genotype.

9

10 **Discussion**

11 The mitochondrial analyses confirmed the low variability among *T.*
12 *cancriformis* specimens (Mantovani et al., 2004). If this is somewhat
13 acceptable for a protein-coding gene such as COI, it is absolutely
14 unexpected for the MCR, a mitochondrial marker widely used for
15 population analysis (Chu et al., 2003; Kang et al., 2005; Vianna et al.,
16 2006). Also the widening of the analysis to a higher number of
17 individuals per population for the 16S gene confirms the low
18 differentiation rate. The most differentiated mitotype (A) shows a
19 pairwise sequence difference percentage ranging from 1.64% to
20 1.86%: this datum is in line with that found by Murugan and
21 coworkers (2002) in American *Triops longicaudatus* samples (2%),
22 but it is lower with respect to sequence divergence found within taxa
23 of the other Notostraca genus, *Lepidurus* (up to 3.4%; King &
24 Hanner, 1998).

25 Mutation rate of mitochondrial DNA is usually very high, and with
26 enough sequence length the error in reconstructing the true species'

1 genealogy should be small (even if the definition of “enough
2 sequence” can be problematic; Ballard & Whitlock, 2004). If the
3 divergence time-scale is too small, the number of mutational
4 differences in the populations would be too little, thus preventing to
5 accurately reconstruct the gene trees, even with complete
6 mitochondrial DNA sequence data (Ballard & Whitlock, 2004). Even if
7 *T. cancriformis* lineage is 200 Myr old (Fryer, 1985), the origin of
8 European populations here analysed should date back to 1.08 - 0.26
9 Myr ago (Korn et al., 2006). The very recent divergence from the
10 common ancestor could be the cause of the observed low variability.
11 As a matter of fact, the wide array of analysed specimens for the 16S
12 gene reveals the presence of an overrepresented haplotype (b),
13 which is found in 64 out of 96 individuals and in five out of seven
14 populations.

15 Another explanation for such low variability could also be done
16 rejecting the assumption that mitochondrial DNA evolves as a strictly
17 neutral marker. The McDonald-Kreitman test computed on COI
18 sequences gave a significant P-value (<0.05), thus suggesting that
19 selection may have played a role in the evolution of *T. cancriformis*
20 mitochondrial DNA. Strong direct and indirect effects of selection on
21 other parts of the genome may also influence mitochondria:
22 mitochondrial fitness effects can be context-dependent, as they can
23 be conditioned by the nuclear genotype or by the environment that
24 the organism inhabits (Ballard & Whitlock, 2004). For example,
25 different fitness of mitochondrial haplotypes has been found in
26 copepods (Schizas et al., 2001), mice (Takeda et al., 2000) and

1 *Drosophila* (Rand et al., 2001; James & Ballard, 2003). Presently
2 scored low variability could therefore be linked to the habitat of *T.*
3 *cancriformis*. Temporary pools are often muddy and may become
4 anoxic: tadpole shrimps are thought to withstand this challenging
5 environment by varying the haemoglobin concentration in their
6 haemolymph (Fox, 1949), and by breathing atmospheric oxygen in
7 cases of strong anoxia (Fryer, 1988). Such peculiar conditions could
8 exert a strong selective pressure on the encoding apparatus of
9 phosphorylative oxidation. Obviously, if this will be demonstrated,
10 mitochondrial markers will reduce their informative role. It should be
11 noted though, that this interpretation does not fit with genetic
12 variability data on *Lepidurus*, which, especially for the COI gene,
13 appears highly variable at the intraspecific level (Mantovani et al.,
14 2004; in preparation). Considering the scored low variability levels,
15 our investigation suggests at most an affinity between Southern Italy
16 samples (Lecce and Palermo) and between Oristano and Espolla
17 mitotypes.

18 While mitochondrial analysis fails to reveal patterns of population
19 structure, MSL analyses point out to a clear differentiation between
20 Italian, Austrian and Spanish populations, possibly consistent with a
21 model of isolation by geographic distance.

22 However, variability levels are quite different in analyzed samples:
23 Italian populations are monomorphic (*Grosseto 2003* and *Lecce*) or
24 show little polymorphism and deviate from HWE owing to a
25 significant heterozygote deficiency (*Grosseto 2002* and *Ferrara*). The
26 overall low polymorphism level found in Italian samples is

1 comparable with that previously scored in the Sardinian population
2 (Cesari et al., 2004). On the other hand, the Spanish sample and the
3 previously analyzed Austrian one display a higher degree of
4 polymorphism and private alleles.

5 The clear-cut difference emerging in variability behaviour within
6 Spanish, Austrian and Italian samples may be explained as due to
7 different reasons. First, HW disequilibrium could be linked to null
8 allele presence; however, it should be reminded that the methods to
9 estimate the presence of null alleles assume that we are dealing with
10 gonochoric populations: population subdivision and/or local breeding
11 structure (Brookfield, 1996) are not taken into account. HWE
12 deviations may therefore be related to the presence of null alleles in
13 the Spanish gonochoric population, while in Italian and Austrian
14 samples, HWE deviations are better linked to their parthenogenetic
15 (Scanabissi & Mondini, 2002) and hermaphroditic (Wingstrand, 1978)
16 condition, respectively. The presence of a male in the *Lecce* sample
17 represent the first and so far unique finding of a *T. cancriformis* male
18 in Italy and may explain the differentiation of this population from the
19 other Italian demes. Obviously, the reproductive role of this male
20 needs to be clarified, especially in the light of the variability lack
21 found in this sample.

22 Our analyses reveal a loss of variability experienced by the *Grosseto*
23 sample in two consecutive years (2002-2003) and a yearly difference
24 was scored also with the mitochondrial markers (Table 1). This could
25 be due to the fact that the 2002 genetic variability was embodied in
26 new resistant eggs that came through migration by birds, even if our

1 data on migration do not confirm this pattern. The new eggs could
2 have hatched in 2002, but then individuals may have been selected
3 against. In fact, it has been hypothesized by De Meester et al. (2002)
4 that strong founder events shape population structure in many
5 aquatic organisms (Monopolization Hypothesis), with the presence of
6 egg banks creating a powerful buffer against the impact of new
7 migrants. However, the levels of sample differentiation scored among
8 Italian populations do not support the theory that the capacity of
9 resources monopolization by obligate parthenogens is low (De
10 Meester et al., 2002). Further, in rotifers associated to resting egg
11 banks (Gómez & Carvalho, 2000) and in *T. longicaudatus* (Scott &
12 Grigarick, 1979) single resistant eggs need different conditions of
13 flooding, soil depth and temperature variations before hatching. It is
14 therefore possible that the genetic variability embodied by the
15 *Grosseto 2002* sample was absent in the *Grosseto 2003* sample
16 because most eggs may not have hatched.

17 Interestingly, the *Espolla* sample exhibits all polymorphic loci in
18 HWE, with the notable exception of tcAC8-p1 in the 2004 sample
19 (Table 2). Again, this peculiar situation could be caused by null
20 alleles, but an alternative explanation may be that tcAC8-p1 is a sex-
21 linked locus. In fact, despite the high number of analysed individuals,
22 all Italian and Spanish females exhibit the same homozygous
23 genotype (150/150), while the Spanish males and the hermaphroditic
24 Austrian specimens display also heterozygous genotypes. It could
25 therefore be argued that this locus is sex-linked and that female may
26 represent the heterogametic sex. The variability found in males and

1 hermaphrodites may be further explained taking into account that in
2 some species a larger mutation rate has been detected in males at
3 particular loci possibly owing to a larger number of germ cell divisions
4 (Ellegren, 2000). Obviously, the absence of a linkage map and the
5 inability to define sex chromosomes due to their very small size
6 (Marescalchi et al., 2005) constitute a consistent drawback. The
7 presence of diagnostic genotypes could nevertheless prove very
8 useful in future studies on reproductive biology also for conservation
9 purposes.

10 On the whole, even if reproductive mode is taken into account,
11 scored variability levels are actually low, at variance with other
12 studies also in Crustacean parthenogenetic taxa (Pálsson, 2000;
13 Pfrender et al., 2000). This limited genetic variability embodied by *T.*
14 *cancriformis* both at the mitochondrial and nuclear levels agrees with
15 the well-known morphological stasis experienced by this taxon, but
16 the limited intraspecific differentiation requires to be explained, also
17 in the light of different reproductive behaviour.

18

1 **References**

2

3 Alonso M (1985). A survey of the Spanish Euphyllopoda. *Misc Zool*
4 **9**: 179-208.

5

6 Alonso M (1996). Crustacea, Branchiopoda. In: Fauna Ibérica, vol. 7,
7 Ramos MA et al. (eds). Museo Nacional de Ciencias Naturales.
8 CSIC: Madrid.

9

10 Ballard JWO, Whitlock MC (2004). The incomplete natural history of
11 mitochondria. *Mol Ecol* **13**: 729-744.

12

13 Belkhir K, Borsa P, Chikhi L, Raufaste N, Bonhomme F (2004).
14 GENETIX 4.05, logiciel sous Windows™ pour la génétique des
15 populations. Laboratoire Génome, Populations, Interactions, CNRS
16 UMR 5000, Université de Montpellier II, Montpellier (France). (code
17 available at <http://www.univ-montp2.fr/~genetix/genetix/genetix.htm>)

18

19 Boix D, Sala J, Moreno-Amich R (2002). Population dynamics of
20 *Triops cancriformis* (Crustacea: Branchiopoda: Notostraca) of the
21 Espolla temporary pond in the northeastern Iberian peninsula.
22 *Hydrobiologia* **486**: 175-183.

23

24 Brookfield JFY (1996). A simple new method for estimating null allele
25 frequency from heterozygote deficiency. *Mol Ecol* **5**: 453-455.

1

2 Cesari M, Mularoni L, Scanabissi F, Mantovani B (2004).
3 Characterization of dinucleotide microsatellite loci in the living fossil
4 tadpole shrimp *Triops cancriformis* (Crustacea Branchiopoda
5 Notostraca). *Mol Ecol Notes* **4**: 733-735.

6

7 Clement M, Posada D, Crandall K (2000). TCS: a computer program
8 to estimate gene genealogies. *Mol Ecol* **9**: 1657-1660.

9

10 Chakraborty R, De Andrade M, Daiger SP, Budowle B (1992).
11 Apparent heterozygote deficiencies observed in DNA typing data and
12 their implications in forensic applications. *Ann Hum Genet* **56**: 45-47.

13

14 Chu K, Li CP, Tam YK, Lavery S (2003). Application of the
15 mitochondrial control region in population genetic studies of the
16 shrimp *Penaeus*. *Mol Ecol Notes* **3**: 120-122.

17

18 Colosi G (1923). Note sopra alcuni Eufillopodi. *Atti Soc Ital Sci Nat*
19 **62**: 75-80.

20

21 Doyle JJ, Doyle JL (1987). A rapid DNA isolation method for small
22 quantities of fresh tissues. *Phytochemical Bulletin* **19**: 11-15.

23

24 De Meester L, Gómez A, Okamura B, Schwenk K (2002). The
25 Monopolization Hypothesis and the dispersal-gene flow paradox in
26 aquatic organism. *Acta Oecol* **23**: 121-135.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

Ellegren H (2000). Evolution of the avian sex chromosomes and their role in sex determination. *Trends Ecol Evol* **15**: 188-192.

Engelmann M, Hoheisel G, Hahn T, Joost W, Vieweg J, Naumann W (1996). Populationen von *Triops cancriformis* (Bosc) (Notostraca) in Deutschland Nördlich 50°N sind nicht klonal und höchstens fakultativ hermaphroditisch. *Crustaceana (Leiden)* **69**: 755-768.

Engelmann M, Hahn T, Hoheisel G (1997). Ultrastructural characterization of the gonads of *Triops cancriformis* (Crustacea, Notostraca) from populations containing both females and males: no evidence for hermaphroditic reproduction. *Zoomorphology* **117**: 175-180.

Eder E., Hödl W, Milasowszky N (1996). Die Gross-Branchiopoden des Seewinkels. *Stapfia* **42**, zugleich *Kataloge des O. Ö. Landesmuseums N.F.* **100**: 93-101.

Excoffier L, Laval G, Schneider S (2005). Arlequin ver. 3.0: An integrated software package for population genetics data analysis. *Evolutionary Bioinformatics Online* **1**: 47-50.

Farris JS, Källersjö M, Kluge AG, Bult C (1994). Testing significance of incongruence. *Cladistics* **10**: 315-319.

1 Farris JS, Källersjö M, Kluge AG, Bult C (1995). Constructing a
2 significance test for incongruence. *Syst Biol* **44**: 570-572.
3
4 Figuerola J, Green AJ, Michot TC (2005). Invertebrate Eggs Can Fly:
5 Evidence of Waterfowl-Mediated Gene Flow in Aquatic Invertebrates.
6 *Am Nat*, **165**: 274-280.
7
8 Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994). DNA
9 primers for amplification of mitochondrial cytochrome oxidase subunit
10 I from diverse metazoan invertebrates. *Mol Mar Biol Biotech* **3**: 294-
11 299.
12
13 Fox HM (1949). On *Apus*: its rediscovery in Britain, nomenclature
14 and habits. *Proc Zool Soc Lond* **119**: 693-702.
15
16 Fryer G (1985). Structure and habits of living branchiopod
17 crustaceans and their bearing on the interpretation of fossil forms. T
18 *Roy Soc Edin* **76**: 103-113.
19
20 Fryer G (1988). Studies on the functional morphology and biology of
21 the Notostraca (Crustacea: Branchiopoda). *Philos Trans R Soc Lond*
22 *B Biol Sci* **321**: 27-124.
23
24 Gauthier H (1934). Contribution à l'étude de l'*Apus cancriformis* et de
25 ses variations dans l'Afrique du Nord. *Bull Soc Sci Nat Maroc*
26 **14**:125-139.

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26

Ghigi A (1921). Ricerche sui Notostraci di Cirenaica a di altri paesi del Mediterraneo. *Atti Soc Ital Sci Nat* **60**: 161-188.

Ghigi A (1924). Ancora sulla sistematica delle specie mediterranee del genere *Triops*. *Atti Soc Ital Sci Nat* **63**: 193-202.

Gómez A, Carvalho GR (2000). Sex, parthenogenesis and genetic structure of rotifers: microsatellite analysis of contemporary and resting egg bank populations. *Mol Ecol* **9**: 203-214.

Goudet J (2001). FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). (code available from <http://www.unil.ch/izea/software/fstat.html>)

Gurney R (1923). Notes on some British and North African specimens of *Apus cancriformis*, Schaeffer. *Ann Mag Nat Hist* **11**: 496-502.

James AC, Ballard JWO (2003). Mitochondrial genotype affects fitness in *Drosophila simulans*. *Genetics* **164**: 187-194.

Kang TW, Lee EH, Kim MS, Paik SG, Kim S, Kim CB (2005). Molecular phylogeny and geography of Korean medaka fish (*Oryzias latipes*). *Mol Cells* **20**:151-156.

1 King JL, Hanner R (1998). Cryptic species in a “living fossil” lineage:
2 taxonomic and phylogenetic relationships within the genus *Lepidurus*
3 (Crustacea: Notostraca) in North America. *Mol Phylogenet Evol* **10**:
4 23-36.

5

6 Korn M, Marrone F, Pérez-Bote JL, Machado M, Cristo M, Cancela
7 da Fonseca L, Undsdoerfer AK (2006). Sister species within the
8 *Triops cancriformis* lineage (Crustacea, Notostraca). *Zool Scr* **35**:
9 301-322

10

11 Kumar S, Tamura K, Nei M (2004). MEGA3: Integrated software for
12 Molecular Evolutionary Genetics Analysis and sequence alignment.
13 *Brief Bioinform* **5**:150-163.

14

15 Linder F (1952). Contributions to the morphology and taxonomy of
16 the Branchiopoda Notostraca, with special reference to the North
17 American species. *Proc U S Nat Mus* **102**: 1-69.

18

19 Longhurst AR (1955). A review of the Notostraca. Bulletin of the
20 British Museum (Natural History). *Zoology* **3**: 1-57.

21

22 Mantovani B, Cesari M, Scanabissi F (2004). Molecular taxonomy
23 and phylogeny of the ‘living fossil’ lineages *Triops* and *Lepidurus*
24 (Branchiopoda: Notostraca). *Zool Scr* **33**: 367-374.

25

1 Marescalchi O, Cesari M, Eder E, Scanabissi F, Mantovani B (2005).
2 Chromosomes in sexual populations of Notostracan and
3 Conchostracan taxa (Crustacea, Branchiopoda). *Caryologia* **58**: 164-
4 170.

5

6 Mathias P (1937). *Biologie des Crustacés Phyllopoetes*. Hermann
7 Editeurs: Paris.

8

9 McDonald JH, Kreitman M (1991). Adaptive protein evolution at the
10 *Adh* locus in *Drosophila*. *Nature* **351**: 652-654.

11

12 Murugan G, Maeda-Martinez AM, Obrégon-Barboza H, Hernández-
13 Saavedra (2002). Molecular characterization of the tadpole shrimp
14 *Triops* (Branchiopoda: Notostraca) from the Baja California
15 Peninsula, México: new insights on species diversity and phylogeny
16 of the genus. *Hydrobiologia* **486**: 101-113.

17

18 Nei M (1987). *Molecular evolution genetics*. Columbia University
19 Press: New York.

20

21 Pálsson S (2000). Microsatellite variation in *Daphnia pulex* from both
22 sides of the Baltic Sea. *Mol Ecol* **9**: 1075-1088.

23

24 Pannell JR (2002). The evolution and maintenance of androdioecy.
25 *Annu Rev Ecol Syst* **33**: 397-425.

26

1 Piry S, Alapetite A, Cornuet JM, Paetkau D, Baudouin L, Estoup A
2 (2004). GeneClass2: A Software for Genetic Assignment and First-
3 Generation Migrant Detection. *J Hered* **95**: 536-539.
4
5 Posada D, Crandall KA (1998). Modeltest: testing the model of DNA
6 substitution. *Bioinformatics* **14**: 817-818.
7
8 Pfrender ME, Spitze K, Lehman N (2000). Multi-locus genetic
9 evidence for rapid ecologically based speciation in *Daphnia*. *Mol Ecol*
10 **9**: 1717-1735.
11
12 Rand DM, Clark AG, Kann LM (2001). Sexually antagonistic
13 cytonuclear fitness interactions in *Drosophila melanogaster*. *Genetics*
14 **159**: 173-187.
15
16 Raymond M, Rousset F (1995). GENEPOP (version 1.2): population
17 genetics software for exact tests and ecumenicism. *J Hered* **86**: 248-
18 249.
19
20 Remigio EA, Hebert PDN (2000). Affinities among anostracans
21 (Crustacea: Branchiopoda) families inferred from phylogenetic
22 analyses of multiple gene sequences. *Mol Phylogenet Evol* **18**: 117-
23 128.
24

1 Rozas J, Sánchez-DelBarrio JC, Messeguer X, Rozas R (2003).
2 DnaSP, DNA polymorphism analyses by the coalescent and other
3 methods. *Bioinformatics* **19**: 2496-2497.
4
5 Sambrook E, Fritsch F, Maniatis T (1989). Molecular cloning: a
6 laboratory manual. Cold Spring Harbor Press: Cold Spring Harbor,
7 NY.
8
9 Scanabissi F, Mondini C (2002). A survey of the reproductive biology
10 in Italian branchiopods. Part B. The male gonad of *Lepidurus apus*
11 *lubbocki* Brauer, 1873 (Notostraca). *Hydrobiologia* **486**: 273-278.
12
13 Scanabissi F, Eder E, Cesari M (2005). Male occurrence in Austrian
14 populations of *Triops cancriformis* (Branchiopoda, Notostraca) and
15 ultrastructural observations of the male gonad. *Invertebr Biol* **124**: 57-
16 65.
17
18 Scott SR, Grigarick AA (1979). Laboratory studies of factors affecting
19 egg hatch of *Triops longicaudatus* (LeConte) (Notostraca:
20 Triopsidae). *Hydrobiologia* **63**: 145-152.
21
22 Schizas NV, Chandler GT, Coull BC, Klosterhaus SL, Quattro JM
23 (2001). Different survival of three mitochondrial lineages of a marine
24 benthic copepod exposed to a pesticide mixture. *Environ Sci Technol*
25 **35**: 535-538.
26

1 Simon C, Frati F, Beckenbach A, Crespi B, Liu H, Flook P (1994).
2 Evolution weighting and phylogenetic utility of mitochondrial gene
3 sequences and a compilation of conserved polymerase chain
4 reaction primers. *Ann Entomol Soc Am* **87**: 651-701.
5
6 Swofford DL (2001). PAUP* - Phylogenetic Analysis Using
7 Parsimony (*and other methods), Version 4.0. Sinauer Associates,
8 Sunderland, Massachusetts.
9
10 Takeda K, Takahashi S, Onishi A, Hanada H, Imai H (2000).
11 Replicative advantage and tissue-specific segregation of RR
12 mitochondrial DNA between C57BL/6 and RR heteroplasmic mice.
13 *Genetics* **155**: 777-783.
14
15 Templeton AR, Crandall KA, Sing CF (1992). A cladistic analysis of
16 phenotypic association with haplotypes inferred from restriction
17 endonuclease mapping and DNA sequence data. III. Cladogram
18 estimation. *Genetics* **132**: 619-633.
19
20 Trusheim F (1938). Triopsiden aus dem Keuper-Frankens.
21 *Palaeontol Z* **19**: 198-216
22
23 Van Oosterhout C, Hutchinson WF, Wills DPM, Shipley P (2004).
24 Micro-Checker: software for identifying and correcting genotyping
25 errors in microsatellite data. *Mol Ecol Notes* **4**: 535-538.
26

1 Vianna JA, Bonde RK, Caballero S, Giraldo JP, Lima RP, Clark A,
2 Marmontel M, Morales-Vela B, De Souza MJ, Parr L, Rodriguez-
3 Lopez MA, Mignucci-Giannoni AA, Powell JA, Santos FR (2006).
4 Phylogeography, phylogeny and hybridization in trichechid sirenians:
5 implications for manatee conservation. *Mol Ecol* **15**: 433-447.
6
7 Weeks SC, Benvenuto C, Reed SK (2006). When males and
8 hermaphrodites coexist: a review of androdioecy in animals. *Integr*
9 *Comp Biol* **46**: 449-464
10
11 Wingstrand KG (1978). Comparative spermatology of the Crustacea
12 Entomostraca. 1. Subclass Branchiopoda. *Biol Skr Dan Vid Sel* **22**:
13 1-66.
14
15 Wright S (1969). Evolution and the genetics of populations. Vol. 2:
16 The theory of gene frequencies. University of Chicago Press:
17 Chicago.
18
19 Zaffagnini F, Trentini M (1980). The distribution and reproduction of
20 *Triops cancriformis* (Bosc) in Europe (Crustacea, Notostraca). *Monit*
21 *Zool Ital* **14**: 1-8.
22

1 **Acknowledgements**

2 The authors are indebted to Dr. Dani Boix, Dr. Jordi Sala (Girona
3 University) and Dr. Loris Mularoni (IMIM, Barcelona) for collecting the
4 Espolla specimens and to Dr. Giuseppe Alfonso (Lecce University)
5 for sampling the Lecce individuals. We also thank Dr. Erich Eder
6 (Wien University) for supplying the Austrian sample. This work was
7 funded by M.I.U.R. 40% and Donazione Canziani (Università di
8 Bologna) grants.

9

1 **Legends to figures**

2

3 **Figure 1 A.** Maximum Parsimony dendrogram (consistency index:
4 0.974; retention index: 0.839; tree length: 349) computed on the
5 combined mitochondrial dataset (12S, 16S, MCR and COI). Values
6 above the branches indicate mutational steps, while those under the
7 branches show bootstrap values.

8 **B.** Maximum Likelihood (-lnL: 3824.69) phylogram obtained from
9 combined analyses of the four mitochondrial genes. Values under the
10 branches point out bootstrap percentages.

11

12 **Figure 2.** 16S haplotype network. Lines represent a single mutational
13 event or an indel regardless of their length, while circles represent
14 haplotypes, with size proportional to the frequency of occurrence.
15 Open dots indicate hypothetical mitotypes. Letters denoting different
16 haplotypes as in table 1, with the addition of haplotypes i-m (see text
17 for details).

18

19 **Figure 3.** Allelic frequencies at the five MSL in the presently
20 analysed populations.

Table 1. Sampling information, scored haplotypes in taxa analyzed for mitochondrial genes and mean number of analyzed individuals for MSL. Asterisks denote haplotypes and samples scored in previous papers (*Mantovani et al. 2004, GenBank Accession Number 12S; AY1595634; 16S, AY159571-7; **Cesari et al. 2004). (H = hermaphrodite; F = female; M = male; n.a.: not available).

Collecting Site	Year	Sex	Mitochondrial Analysis					Nuclear Analysis	
			12S	16S	Haplotype		mt-type	Mean sample size	
<i>Austria</i>									
Marchegg	2002	H	Marchegg-1	a	a	a	a	A	16.8**
	2002	H	Marchegg-2	b	b	b	b	B	
<i>Italy</i>									
Ferrara - Emilia Romagna	2001	F	Ferrara-1	b*	c*	b	c	C	21.0
	2001	F	Ferrara-2	b*	b*	c	c	D	
Grosseto - Tuscany	2002	F	Grosseto-1	b*	d*	b	c	E	19.8
	2002	F	Grosseto-2	b*	d*	b	c	E	
	2003	F	Grosseto-3	b	b	b	c	F	
	2003	F	Grosseto-4	b	b	b	c	F	
Lecce - Apulia	2005	F	Lecce-1	b	e	d	d	G	19.2
	2005	M	Lecce-2	b	e	d	d	G	
Oristano - Sardinia	1995	F	Oristano-1	b*	f*	e	c	H	16.0**
	1995	F	Oristano-2	b*	a*	e	c	I	
Palermo - Sicily	2001	F	Palermo-1	b*	g*	f	d	J	n.a.
	2001	F	Palermo-2	b*	h*	f	d	K	
<i>Spain</i>									
Espolla	2004	F	Espolla-1	b	b	e	c	L	18.6
	2004	M	Espolla-2	c	b	e	e	M	
	2006	F	Espolla-3	c	b	e	c	N	
	2006	M	Espolla-4	c	b	e	c	N	

Table 2. Number of alleles (A), allelic richness (A_C), possible null alleles presence (NA), genic diversity (GD; Nei 1987), observed (H_O) and expected (H_E) heterozygosity for each MSL for the six *T. cancriformis* samples (N=number of analysed specimens, distinguished in females/males; asterisks denote P values of the HW exact test: * $P < 0.05$; *** $P < 0.001$).

	Ferrara	Grosseto (2002)	Grosseto (2003)	Lecce	Espolla (2004)	Espolla (2006)	Locus
A / A_C	1 / 1.00	1 / 1.00	1 / 1.00	1 / 1.00	2 / 2.00	2 / 2.00	
NA	-	-	-	-	yes	-	
GD	0.000	0.000	0.000	0.000	0.474	0.288	
H_O	0.000	0.000	0.000	0.000	0.200	0.222	tcAC-8p1
H_E	0.000	0.000	0.000	0.000	0.455	0.278	
					*		
N	19/0	20/0	20/0	17/1	10/10	10/8	
A / A_C	1 / 1.00	3 / 2.91	1 / 1.00	1 / 1.00	3 / 2.95	3 / 2.94	
NA	-	yes	-	-	yes	-	
GD	0.000	0.279	0.000	0.000	0.224	0.325	
H_O	0.000	0.000	0.000	0.000	0.118	0.263	tcAC-9p1
H_E	0.000	0.265	0.000	0.000	0.215	0.314	

N	21/0	20/0	20/0	19/1	10/7	10/9	
A / A_C	2 / 1.87	3 / 2.68	1 / 1.00	1 / 1.00	5 / 4.33	4 / 3.71	
NA	yes	-	-	-	-	-	
GD	0.091	0.191	0.000	0.000	0.386	0.332	
H_O	0.000	0.100	0.000	0.000	0.333	0.263	tcAC-10p1
H_E	0.087	0.184	0.000	0.000	0.374	0.321	
	*						
N	22/0	20/0	20/0	19/1	10/8	10/9	
A / A_C	1 / 1.00	1 / 1.00	1 / 1.00	1 / 1.00	1 / 1.00	1 / 1.00	
NA	-	-	-	-	-	-	
GD	0.000	0.000	0.000	0.000	0.000	0.000	
H_O	0.000	0.000	0.000	0.000	0.000	0.000	tcAc-10p2
H_E	0.000	0.000	0.000	0.000	0.000	0.000	
N	22/0	19/0	20/0	18/1	9/10	10/9	
A / A_C	1 / 1.00	1 / 1.00	1 / 1.00	1 / 1.00	2 / 1.99	2 / 1.74	
NA	-	-	-	-	-	-	
GD	0.000	0.000	0.000	0.000	0.234	0.053	
H_O	0.000	0.000	0.000	0.000	0.263	0.053	tcAC-14p1
H_E	0.000	0.000	0.000	0.000	0.229	0.051	
N	21/0	20/0	20/0	18/1	10/9	10/9	
H_O	0.000	0.020	0.000	0.000	0.183	0.160	
H_E	0.017	0.090	0.000	0.000	0.254	0.193	Over all loci
	*	***					

Table 3. Below the diagonal, pairwise F_{ST} values and between parentheses its significance (***: $P < 0.001$). Above the diagonal, pairwise estimated number of migrants (N_m), following Wright (1969; $N_m = (1 - F_{ST}) / (4 * F_{ST})$) and M values ($M = 2N_m$, where $N_m = (1 - F_{ST}) / (2 * F_{ST})$; in parentheses) between samples over all loci.

	Marchegg	Ferrara	Grosseto (2002)	Grosseto (2003)	Oristano	Lecce	Espolla (2004)	Espolla (2006)
Marchegg		0.22 (0.8861)	0.34 (0.8989)	0.20 (0.5746)	0.27 (0.7223)	0.09 (0.3692)	0.16 (0.5517)	0.14 (0.4611)
Ferrara	0.5293 (***)		6.21 (9.2002)	infinity (7.2131)	4.10 (8.0375)	0.01 (0.2346)	0.07 (0.2730)	0.05 (0.2166)
Grosseto (2002)	0.4215 (***)	0.0387		3.88 (6.5909)	4.10 (6.3778)	0.07 (0.2673)	0.11 (0.3002)	0.08 (0.2372)
Grosseto (2003)	0.5566 (***)	-0.0045	0.0605		2.62 (3.9024)	0.00 (0.0939)	0.07 (0.2027)	0.05 (0.1530)
Oristano	0.4836 (***)	0.0574	0.0575	0.0872 (***)		0.03 (0.1533)	0.10 (0.2740)	0.07 (0.2124)
Lecce	0.7290 (***)	0.9544 (***)	0.7865 (***)	1.0000 (***)	0.9049 (***)		0.07 (0.2545)	0.05 (0.1938)
Espolla (2004)	0.6035 (***)	0.7743 (***)	0.7033 (***)	0.7880 (***)	0.7244 (***)	0.7837 (***)		19.45 (52.9202)
Espolla (2006)	0.6475 (***)	0.8211 (***)	0.7513 (***)	0.8349 (***)	0.7730 (***)	0.8368 (***)	0.0127	

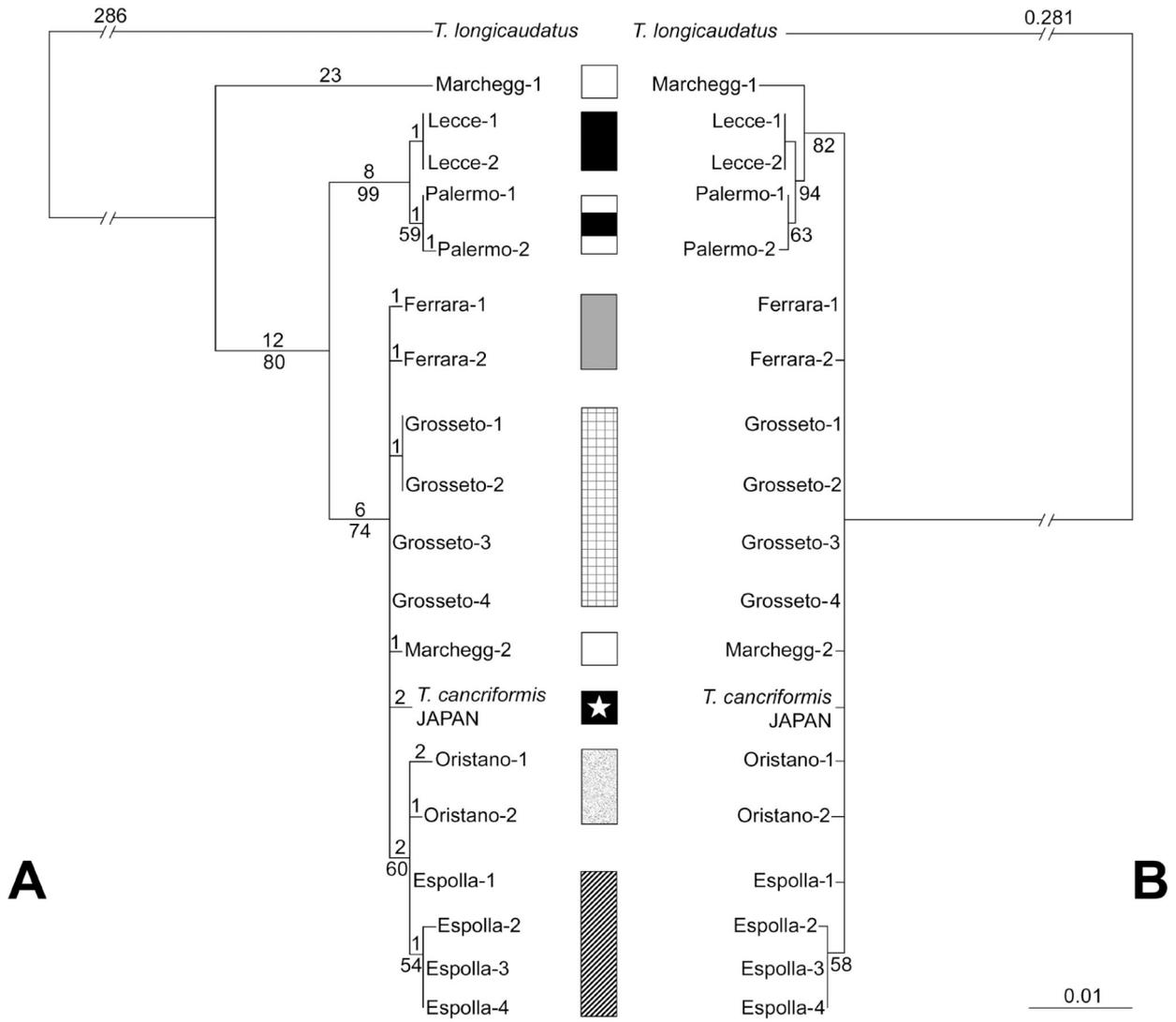


Figure 1

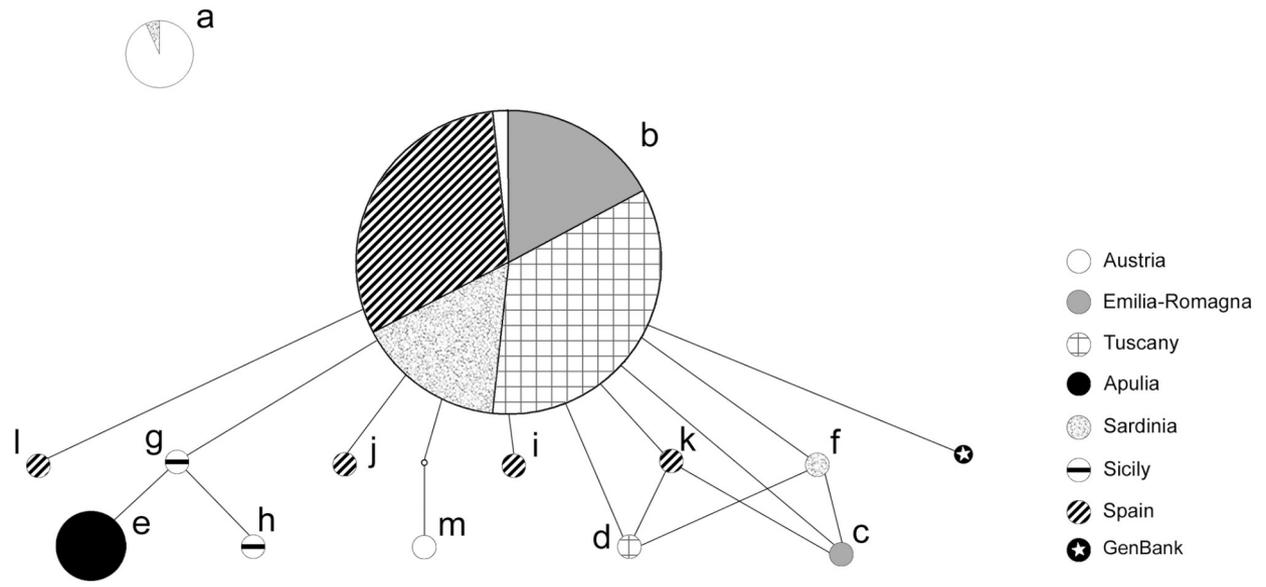


Figure 2

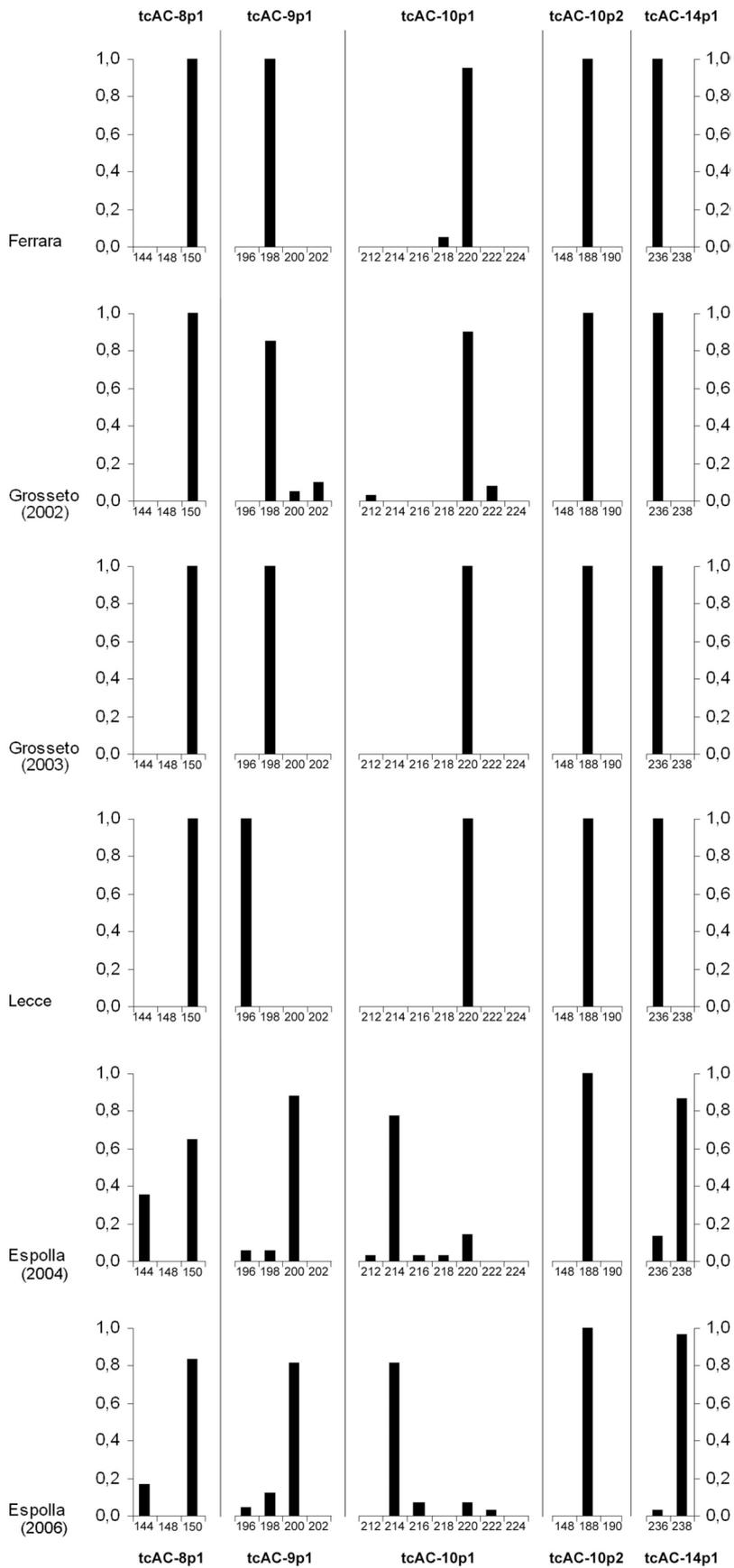


Figure 3

Molecular characterization of ribosomal intergenic spacer in the tadpole shrimp *Triops cancriformis* (Crustacea, Branchiopoda, Notostraca)

Andrea Luchetti, Franca Scanabissi, and Barbara Mantovani

Abstract: Nuclear ribosomal DNA constitutes a multigene family, with tandemly arranged units linked by an intergenic spacer (IGS), which contains initiation/termination transcription signals and usually tandemly arranged subrepeats. The structure and variability of the IGS region are analyzed here in hermaphroditic and parthenogenetic populations of the "living fossil" *Triops cancriformis* (Branchiopoda, Notostraca). The results indicate the presence of concerted evolution at the population level for this G+C-rich IGS region as a whole, with the major amount of genetic variability found outside the subrepeat region. The subrepeats region is composed of 3 complete repeats (a, c, d) intermingled with 3 repeat fragments (b, e, f) and unrelated sequences. The most striking datum is the absolute identity of subrepeats (except type d) occupying the same position in different individuals/populations. A putative promoter sequence is present upstream of the 18S rRNA gene, but not in subrepeats, which is at variance with other arthropod IGSs. The absence of a promoter sequence in the subrepeats and subrepeat sequence conservation suggests that this region acts as an enhancer simply by its repetitive nature, as observed in some vertebrates. The putative external transcribed spacer (840 bp) shows hairpin structures, as in yeasts, protozoans, *Drosophila*, and vertebrates.

Key words: concerted evolution, Crustacea, external transcribed spacer, intergenic spacer, ribosomal DNA, subrepeats, *Triops cancriformis*.

Résumé : L'ADN ribosomal nucléaire constitue une famille de multigène, avec les unités tandemly disposées liées par une entretoise intergénique (IGS), qui contient des signaux d'initiation/terminateur de la transcription et habituellement tandemly disposé secondaire-répète. Dans ce travail, la structure et la variabilité de la région d'IGS est analysée dans les populations hermaphrodites et parthenogenetic du fossile vivant *T. cancriformis* (Branchiopoda, Notostraca). Les résultats démontrent la présence de l'évolution concertée au niveau de population pour cette IGS région riche de G+C dans l'ensemble, avec la quantité principale de variabilité génétique ont trouvé en dehors du secondaire-répète la région. Le dernier est accumulation par trois répétitions complètes (a, c, d) mélangeant avec trois fragments de répétition (b, e, f) et ordres indépendants. Les informations les plus saisissantes sont l'identité absolue des subrepeats (moins que le d) occupant la même position entre individuals/populations différent. Un putatif promoteur est ascendant le gène 18S, mais pas dans secondaire-répète, au désaccord d'autres arthropodes IGS. L'absence de promoteurs dans secondaire-répète et la conservation des secondaire-répète suggèrent que cette région pourrait agir en tant que renforceur simplement par leur nature réitérée, comme observé dans quelques vertébrés. La putative entretoise transcrite externe (ETS; 840 bp) montre des structures d'épingle comme déjà observé dans les levures, les protozoans, la drosophile et les vertébrés.

Mots clés : évolution concertée, Crustacea, entretoise transcrite externe, entretoise intergénique, ADN ribosomal, secondaire-répète, *Triops cancriformis*.

Introduction

In eukaryotes, nuclear ribosomal RNA genes constitute a multigene family, composed of hundreds or thousands of tandemly arranged members (repeats). Each rDNA repetitive

unit contains 18S, 5.8S, and 28S rRNA gene coding regions, separated by internal transcribed spacers (ITS1 and ITS2, respectively). Each unit is linked to the following one by a long intergenic spacer (IGS). This IGS region is of particular interest, given the presence of the transcription initiation and

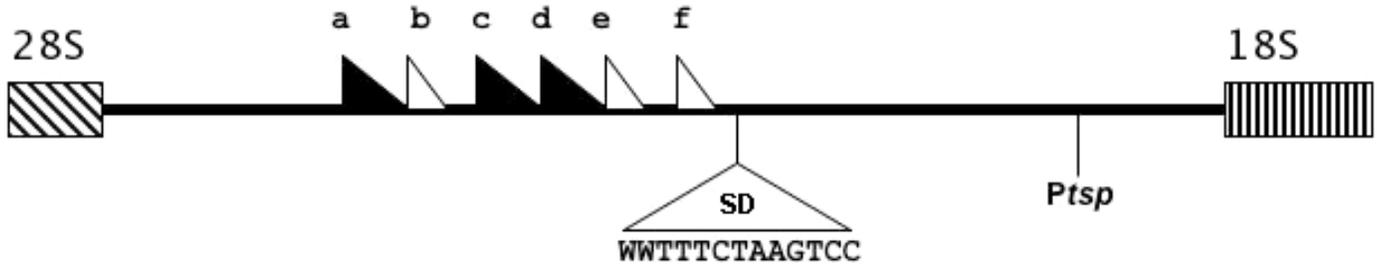
Received 17 October 2005. Accepted 28 March 2006. Published on the NRC Research Press Web site at <http://genome.nrc.ca> on 13 September 2006.

Corresponding Editor: B. Golding.

A. Luchetti,¹ F. Scanabissi, and B. Mantovani. Dipartimento di Biologia Evoluzionistica Sperimentale, via Selmi 3, 40126 Bologna, Italy.

¹Corresponding author (e-mail: andluche@alma.unibo.it).

Fig. 1. Schematic drawing of intergenic spacer (IGS) sequence structure. Lowercase letters indicate subrepeats a–f; filled triangles represent complete subrepeats; empty triangles indicate 5' end subrepeat fragments. *Ptsp* indicates the putative transcription starting point. The small domain (SD) shared with *Daphnia pulex* is also indicated.



termination signals, and because of the occurrence of repetitive sequences (subrepeats), which seem to play an adaptive role in local environments (Gorokhova et al. 2002). In Arthropoda, subrepeat organisation varies from a cluster of tandem repeats, as in *Aedes albopictus* (Baldrige and Fallon 1992), to 2 different clusters, as in *Artemia* and *Daphnia pulex* (Gil et al. 1987; Crease 1993), to 4 different clusters, as observed in the swimming crab *Charybdis japonica* (Ryu et al. 1999). In contrast, subrepeats in *Aedes aegypti* are interspersed with unrelated sequences (Wu and Fallon 1998), and the IGS regions in the copepod *Tigriopus* completely lack repetitive DNA (Burton et al. 2005).

As do other repeated sequences, rDNA units show more sequence similarity within than between evolutionary units (i.e., strains, population, subspecies, species, and so on). This pattern is known as concerted evolution, and it is achieved through molecular drive, a process involving the effects of both genomic turnover mechanisms (unequal crossing-over, gene conversion, etc.) and population dynamics (Elder and Turner 1995; Dover 2002). Concerted evolution is particularly evident in IGSs because they experience a relaxed selective force, whereas rRNA coding regions are under purifying selective pressure (Nei and Rooney 2005).

The tadpole shrimp *Triops cancriformis* inhabits ephemeral ponds from Eurasia to Africa, and represents one of the most intriguing taxa to be studied. Its morphological stasis as a living fossil since the Triassic age (Longhurst 1955) contrasts with the consistent variability in sexual reproductive strategies, which range from bisexuality (either gonochoric or hermaphroditic) to unisexuality (parthenogenesis). Recent molecular analyses performed on the 12S and 16S rRNA genes of *T. cancriformis* samples from Europe revealed a substantial differentiation of this taxon from the other species ascribed to the genus *Triops*, and a consistent genetic homogeneity over the analysed range, even for hyper-variable mitochondrial markers (AT-rich region) (Cesari et al. 2004a; Mantovani et al. 2004). Only microsatellite analyses detected some degree of variability in *T. cancriformis*. In particular, the population from Marchegg, Austria, was polymorphic at the 5 analyzed loci for 2 or 3 alleles, whereas the Italian samples were either monomorphic or polymorphic at only 1 or 2 loci. It is possible that the different levels of genetic variability scored for these nuclear markers are linked to the different reproductive strategies of Austrian (hermaphroditism) and Italian (parthenogenesis) populations (Cesari et al. 2004a, 2004b).

The work presented here aims to describe the IGS struc-

ture and variability in Austrian and Italian populations of *T. cancriformis*, to add new data about the IGS region in arthropods, and to obtain a new genetic marker to evaluate *T. cancriformis* population differentiation.

Material and methods

Tadpole shrimps were collected in a rice field (individuals ITA1 and ITA2 from Ferrara, Italy) and from natural ponds (individuals AUS1 and AUS2 from Marchegg, Austria), and preserved in absolute alcohol. Total DNA was obtained with standard phenol-chloroform extraction from the pleon tissues.

Long PCR was performed on 150 ng of DNA template with Takara La Taq kit (Takara Bio Inc., Shiga, Japan), in accordance with the manufacturer's instructions. Amplification was done in a PCT-100 thermocycler (MJ Research), as follows: denaturation for 5 min at 94 °C, 30 cycles at 94 °C for 30 s, 50 °C for 30 s, 70 °C for 10 min, and a final extension at 72 °C for 12 min. The 2 primers were 28ii, modified for branchiopods (5'-GGCTCTTCCTATCATTGCGAAGCA-GTATTCGC-3'), and 18i (5'-TTTCTCAGGCTCCCTCTC-CGGAATCGAACCCCT-3') (Hillis and Dixon 1991).

Amplicons were gel-extracted with the Wizard SV Gel and PCR Purification kit (Promega, Madison, Wis.). They were sequenced using the primer walking method (internal primer sequences are available from the authors) and the Big Dye Terminator Sequencing kit (Applied Biosystems, Norwalk, Conn.) in an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). Sequences were aligned with the CLUSTAL algorithm of the Sequence Navigator software package (Applied Biosystems).

The 18S and 28S rRNA gene boundaries were defined after comparison with other Notostracan sequences found in GenBank.

Internal subrepeats were defined by dot-plot analysis, using the ResDotPlot server program (available at <http://www.changbioscience.com/res/resd.html>) with an 8-bp window size and 100% identity. P-distances and nucleotide compositions were calculated using MEGA v. 2.1 (Kumar et al. 2004); indels were not considered in each pairwise comparison. The external transcribed spacer (ETS) secondary structure was calculated with the MFOLD server program (available at <http://www.bioinfo.rpi.edu/~zukerm/rna/>) (Zucker 2003).

Sequences have been submitted to GenBank (accession Nos. DQ205641–DQ205644).

Fig. 2. Alignment of IGS subrepeats. Lowercase letters (a-f) in the acronyms refer to specific cluster position). Nucleotide differences between the Austrian and Italian subrepeat d are shaded in grey.

80

```

ITA1a GAGGCCGACG TCG-GCTTTC AGCCTCGCCG GGGCAAGGCA GCCTCGGCCG GCCAGGGACT GGGTGCCGCC CGA-GTCGG
ITA2a .....
AUS1a .....
AUS2a .....
ITA1c ..... G . C . GG . T . A . . . . . A . AA . . . . .
ITA2c ..... G . C . GG . T . A . . . . . A . AA . . . . .
AUS1c ..... G . C . GG . T . A . . . . . A . AA . . . . .
AUS2c ..... G . C . GG . T . A . . . . . A . AA . . . . .
ITA1d ..... A . . C . - . . T . . . . . G . . . GG . T . A . . . . . - . A . . . . .
ITA2d ..... A . . C . - . . T . . . . . G . . . GG . T . A . . . . .
AUS1d ..... A . . C . - . . T . . . . . G . . . GG . T . A . . . . .
AUS2d ..... A . . C . - . . T . . . . . G . . . GG . T . A . . . . .
ITA1b ..... A . . C . - . . T . . . . . G . . . GG . T . A . . . . . G . . . . . A . A . . . . . -
ITA2b ..... A . . C . - . . T . . . . . G . . . GG . T . A . . . . . G . . . . . A . A . . . . . -
AUS1b ..... A . . C . - . . T . . . . . G . . . GG . T . A . . . . . G . . . . . A . A . . . . . -
AUS2b ..... A . . C . - . . T . . . . . G . . . GG . T . A . . . . . G . . . . . A . A . . . . . -
ITA1e ..... A . . C . - . . G . . . . . G . . . GG . T . . . . . A . - - - - - - - - - - - - - - -
ITA2e ..... A . . C . - . . G . . . . . G . . . GG . T . . . . . A . - - - - - - - - - - - - - - -
AUS1e ..... A . . C . - . . G . . . . . G . . . GG . T . . . . . A . - - - - - - - - - - - - - - -
AUS2e ..... A . . C . - . . G . . . . . G . . . GG . T . . . . . A . - - - - - - - - - - - - - - -
ITA1f C . . . . . . . GTGAC . - . . . A . . . . . . . GG AGT . . A . . . . . C . - - - - - - - - - - - - - - -
ITA2f C . . . . . . . GTGAC . - . . . A . . . . . . . GG AGT . . A . . . . . C . - - - - - - - - - - - - - - -
AUS1f C . . . . . . . GTGAC . - . . . A . . . . . . . GG AGT . . A . . . . . C . - - - - - - - - - - - - - - -
AUS2f C . . . . . . . GTGAC . - . . . A . . . . . . . GG AGT . . A . . . . . C . - - - - - - - - - - - - - - -

```

160

```

ITA1a CGAAAATGTT GGGTCATATG GCCCAAAACT C-ACAATGTA GGCTAGGGGC ATCCA--GCA GACTCCATAG GAGTGCTTTT
ITA2a .....
AUS1a .....
AUS2a .....
ITA1c . C . G . AT . . TT . C . . C . . . . . . . T . . . C . A . . TCC . . . . . A . . . . . TTCC-- . T . . TTCGA . . . . C . . . .
ITA2c . C . G . AT . . TT . C . . C . . . . . . . T . . . C . A . . TCC . . . . . A . . . . . TTCC-- . T . . TTCGA . . . . C . . . .
AUS1c . C . G . AT . . TT . C . . C . . . . . . . T . . . C . A . . TCC . . . . . A . . . . . TTCC-- . T . . TTCGA . . . . C . . . .
AUS2c . C . G . AT . . TT . C . . C . . . . . . . T . . . C . A . . TCC . . . . . A . . . . . TTCC-- . T . . TTCGA . . . . C . . . .
ITA1d . . . . . T . . . . . . . . . . . . . . . . C . . . . . T . . . . . C . . . . . CC . . . . . C . G . . . . A . . . .
ITA2d . . . . . T . . . . . . . . . . . . . . . . C . . . . . T . . . . . C . . . . . CC . . . . . C . G . . . . A . . . .
AUS1d . . . . . T . . . . . . . . . . . . . . . . C . . . . . T . . . . . C . . . . . CC . . . . . C . G . . . . A . . . .
AUS2d . . . . . T . . . . . . . . . . . . . . . . C . . . . . T . . . . . C . . . . . CC . . . . . C . G . . . . A . . . .
ITA1b -----
ITA2b -----
AUS1b -----
AUS2b -----
ITA1e -----
ITA2e -----
AUS1e -----
AUS2e -----
ITA1f -----
ITA2f -----
AUS1f -----
AUS2f -----

```

206

```

ITA1a CTGTCATTGT TGTCGGGGTC CGGGCTTTTC AAGACGCGTT CCACCGA
ITA2a .....
AUS1a .....
AUS2a .....
ITA1c . C . . . . . . . A . . . . . A . . - . . . . . G . . G . . TG . .
ITA2c . C . . . . . . . A . . . . . A . . C . . . . . G . . G . . TG . .
AUS1c . C . . . . . . . A . . . . . A . . C . . . . . G . . G . . TG . .
AUS2c . C . . . . . . . A . . . . . A . . C . . . . . G . . G . . TG . .
ITA1d . . . . . . . . . . . . . . . . C . . . . . G . . G . . TG . C
ITA2d . . . . . . . . . . . . . . . . C . . . . . G . . G . . TG . C
AUS1d . . . . . . . . . . . . . . . . C . . . . . G . . G . . TG . .
AUS2d . . . . . . . . . . . . . . . . C . . . . . G . . G . . TG . .
ITA1b -----
ITA2b -----
AUS1b -----
AUS2b -----
ITA1e -----
ITA2e -----
AUS1e -----
AUS2e -----
ITA1f -----
ITA2f -----
AUS1f -----
AUS2f -----

```

Fig. 3. Putative promoter sequence of *Triops cancriformis* (Tca), *D. pulex* (Dpu), and *Artemia franciscana* (Afr) gene (g) and spacer (s). The beginning of the promoter sequence is indicated by >; 7 bp of the upstream sequence are reported. The transcription start point identified in *A. franciscana* (Koller et al. 1987) is underlined.

Tca TGGTTTT>TATACGTGAAAGC
Dpu (g) ATTGATA>TATACGAGAGAAT
Dpu (s) TTTCCTA>TATAGGGAAGAGG
Afr (g) GACCTTA>TATAAGGGGGTCA
Afr (s) TCTTGTA>TATAGAGAAGTCA

Results and discussion

Long PCR amplification products obtained for each individual tadpole shrimp ranged from 4938 bp in the Austrian samples to 4939 and 4940 bp in the ITA1 and ITA2 individuals, respectively. The consensus sequence was used in a BLAST search, and no significant sequence similarity was observed with any other sequence/domain in public databases, with the exception of the 5' (492 bp) and 3' (367 bp) termini, which align with other arthropod 28S and 18S rRNA genes, respectively. The sequenced rRNA genes showed a G+C content equal to 55.1% (28S) and 47.7% (18S) and no nucleotide substitution among the analysed samples.

The IGS region is 4079–4081 bp long. It has a G+C content of 52.9%. This G+C richness contrasts with data on other crustaceans, where an A+T-rich IGS is usually found (Crease 1993 and references therein).

The sequences retrieved exhibit complete identity between the Austrian samples, whereas the Italian ones differ for 3 indels. At the interpopulation level, 0.17% nucleotide diversity was observed. Even if based on a small sampling, these estimates indicate that the evolution of the IGS region appears concerted at the population level in tadpole shrimps. Genetic isolation is the basis for the observed pattern of concerted evolution, because mutations can spread between populations only by gene flow. Data on mitochondrial and microsatellite markers suggest restricted gene flow between the Italian and Austrian *T. cancriformis* populations (Cesari et al. 2004a, 2004b; Mantovani et al. 2004); this may explain the results obtained here.

Dot-plot analysis showed a repetitive block located between nucleotide positions 1300 and 2500 of the complete amplicon; however, repeats appeared to be poorly conserved and discontinuous (not shown). Visual inspection of the sequences revealed an unusual organisation of this region (Fig. 1), with complete repeats (a, c, d) intermingling with repeat fragments containing only the 5' end (b, e, f) and nonrepetitive sequence. This is a peculiar configuration among the IGS regions analysed to date (Gil et al. 1987; Crease 1993; Baldrige and Fallon 1992; Ryu et al. 1999).

Sequence diversity between subrepeats within the same cluster varies from 3.4% to 24.9%. Repeat d is less divergent from the other complete repeats (a, c), and fragment f is the most divergent one within each cluster. When subrepeats occupying the same position are compared both within and between the Italian and Austrian samples, a complete identity is observed. This means that the same sequence is retrieved,

for example, for subrepeat a, disregarding individual or population. Only d subrepeats behave differently: they show 1 nucleotide substitution between Italian clusters, and 2 fixed nucleotide substitutions between Italian and Austrian samples (Fig. 2).

In the subrepeat cluster of *D. pulex*, there are position-specific subrepeat variants; significant differences between isolated populations can be seen when analysing the cluster as a whole (Crease 1995). In the case of *T. cancriformis*, p-distance analysis indicates the presence of position-specific subrepeats, which are absolutely identical between individuals and populations, with the exception of subrepeat d (Fig. 2). Considering the molecular data on genetic isolation and the differences in reproductive strategies between the Italian (parthenogenetic) and Austrian (hermaphroditic) samples (Cesari et al. 2004a, 2004b; Mantovani et al. 2004), this position-specific conservation is difficult to explain. Given the concerted evolution observed for the IGS region as a whole, the absence of genomic turnover mechanisms able to homogenize subrepeats within populations might be related to some functional constraints in this part of the spacer (see below).

The transcription start point and the promoter sequence have not been experimentally determined in this work, but a pyrimidine–purine-rich sequence, similar to *D. pulex* and *Artemia franciscana* promoters (Fig. 3) (Koller et al. 1987; Crease 1993), has been found with a putative transcription start point at nucleotide 3735 of the alignment (Fig. 1). In several arthropods, subrepeat elements can act as transcription enhancers by carrying duplications of gene promoters in their sequence (see Crease 1993 and references therein; Wu and Fallon 1998 and references therein). However, subrepeats analysed in this work do not show any sequence similarity with gene promoter sequences. In the dipteran *Simulium sanctipauli*, IGS subrepeats lack a promoter sequence, but it has been suggested that the simple repetitive nature of this region acts as an enhancer of transcription, as observed in some vertebrates (Morales-Hojas et al. 2002 and references therein). This hypothesis might also be proposed for tadpole shrimps, explaining both the absence of a promoter sequence in their subrepeats and the sequence conservation of specifically located subrepeats.

Visual inspection of the sequences also reveals a small domain already observed in other crustacean IGS regions. In *D. pulex*, the sequence 5'-WWTTTCTAAGTCC-3' is present both within subrepeats and immediately downstream of the repetitive region (Crease 1993). In *T. cancriformis* IGS, this sequence was found in the latter position, starting from nucleotide 2794 of the alignment, but not within subrepeats (Fig. 1). The significance of this small sequence is unknown; however, its sequence conservation (100% of identity) and the fact that the same position is observed indicates involvement in some function.

The length of the ETS from the putative transcription start point is 840 bp long, roughly the same length observed for other arthropods (*A. franciscana*, 790 bp (Koller et al. 1987); *Drosophila melanogaster*, 861 bp (Simeone et al. 1985); *A. aegypti*, 795 bp (Wu and Fallon 1998)). The ETS sequence does not show repetition or similarity with other published sequences, but it has the potential to form stable secondary structures with long hairpins (Fig. 4A). Such structures have already been observed in vertebrates, yeast,

Fig. 4. Secondary structure of the putative ETS. A. Numbers refer to nucleotide positions. B. ETS/18S junction (arrow indicates the beginning of the 18S rRNA gene. Calculated free energy is reported below each structure.



and trypanosomatid protozoa (Michot and Bachellerie 1991; Yeh and Lee 1992; Schnare et al. 2000). Furthermore, a stem-loop structure can be formed at the ETS/18S rRNA gene junction (Fig. 4B), as in *D. melanogaster* and trypanosomatid protozoa (Simeone et al. 1985; Schnare et al. 2000). The ability to form secondary structures seems to be a shared feature of ETS regions, suggesting a potential role in pre-rRNA processing mechanisms.

Acknowledgements

We wish to thank Eric Eder for providing the Austrian sample. This work was carried out with the financial support of RFO and Canziani funds.

References

- Baldrige, G.D., and Fallon, A.M. 1992. Primary structure of the ribosomal DNA intergenic spacer from the mosquito, *Aedes albopictus*. *DNA Cell Biol.* **11**: 51–59.
- Burton, R.S., Metz, E.C., Flowers, J.M., and Willet, C.S. 2005. Unusual structure of ribosomal DNA in the copepod *Tigriopus californicus*: intergenic spacer sequences lack internal sub-repeats. *Gene*, **344**: 105–113.
- Cesari, M., Luchetti, A., Scanabissi, F., and Mantovani, B. 2004a. Genetic variability in *Triops cancriformis* (Bosc 1801) populations revealed by molecular markers. V^o International Large Branchiopod Symposium, Toodyay, Western Australia, p. 16.
- Cesari, M., Mularoni, L., Scanabissi, F., and Mantovani, B. 2004b. Characterization of dinucleotide microsatellite loci in the living fossil tadpole shrimp *Triops cancriformis* (Crustacea Branchiopoda). *Mol. Ecol. Notes*, **4**: 733–735.
- Crease, T.J. 1993. Sequence of the intergenic spacer between the 28S and 18S rRNA-encoding gene of the crustacean, *Daphnia pulex*. *Gene*, **134**: 245–249.
- Crease, T.J. 1995. Ribosomal DNA evolution at the population level: nucleotide variation in intergenic spacer arrays of *Daphnia pulex*. *Genetics*, **141**: 1327–1337.
- Dover, G.A. 2002. Molecular drive. *Trends Genet.* **18**: 587–589.
- Elder, J.F., and Turner, B.J. 1995. Concerted evolution of repetitive DNA sequences in eukaryotes. *Quart. Rev. Biol.* **70**: 297–320.
- Gil, I., Gallego, M.E., Renart, J., and Cruces, J. 1987. Identification of the transcriptional initiation site of ribosomal RNA genes in the crustacean *Artemia*. *Nucleic Acids Res.* **15**: 6007–6016.
- Gorokhova, E., Doeling, T.E., Weider, L.J., Crease, T.J., and Elser, J.J. 2002. Functional and ecological significance of rDNA intergenic spacer variation in a clonal organism under divergent selection for production rate. *Proc. R. Soc. Lond. B*, **269**: 2373–2379.
- Hillis, D.M., and Dixon, M.T. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. *Quart. Rev. Biol.* **66**: 411–453.
- Koller, H.T., Frondorf, K.A., Maschner, P.D., and Vaughn, J.C. 1987. In vivo transcription from multiple spacer rRNA gene promoters during early development and evolution of the intergenic spacer in the brine shrimp *Artemia*. *Nucleic Acids Res.* **15**: 5391–5411.
- Kumar, S., Tamura, K., Jakobsen, I.B., and Nei, M. 2004. MEGA2: Molecular Evolutionary Genetics Analysis software. Arizona State University, Tempe, Ariz.
- Longhurst, A.R. 1955. A review of the Notostraca. *Bull. British Mus. Nat. Hist. D*, **3**: 1–57.
- Mantovani, B., Cesari, M., and Scanabissi, F. 2004. Molecular taxonomy and phylogeny of the 'living fossil' lineages *Triops* and *Lepidurus* (Branchiopoda: Notostraca). *Zool. Scr.* **33**: 367–374.
- Michot, B., and Bachellerie, J.P. 1991. Secondary structure of the 5' external transcribed spacer of vertebrate pre-rRNA. Presence of phylogenetically conserved features. *Eur. J. Biochem.* **195**: 601–609.
- Morales-Hojas, R., Post, R.J., Wilson, M.D., and Cheke, R.A. 2002. Completion of the sequence of the nuclear ribosomal DNA subunit of *Simulium sanctipauli*, with description of the 18S, 28S and the IGS. *Med. Vet. Entomol.* **16**: 386–394.
- Nei, M., and Rooney, A.P. 2005. Concerted and birth-and-death evolution of multigenes families. *Annu. Rev. Genet.* **39**: 121–152.
- Ryu, S.H., Do, Y.K., Hwang, U.W., Choe, C.P., and Kim, W. 1999. Ribosomal DNA intergenic spacer of the swimming crab, *Charybdis japonica*. *J. Mol. Evol.* **49**: 806–809.
- Schnare, M.N., Collings, J.C., Spencer, D.F., and Gray, M.W. 2000. The 28S-18S rDNA intergenic spacer from *Crithidia fasciculata*: repeated sequences, length heterogeneity, putative processing sites and potential interaction between U3 small nu-

- cleolar RNA and the ribosomal RNA precursor. *Nucleic Acids Res.* **28**: 3452–3461.
- Simeone, A., La Volpe, A., and Boncinelli, E. 1985. Nucleotide sequence of a complete ribosomal spacer of *D. melanogaster*. *Nucleic Acids Res.* **13**: 1089–1101.
- Wu, C.C.N., and Fallon, A.M. 1998. Analysis of a ribosomal DNA intergenic spacer region from the yellow fever mosquito, *Aedes aegypti*. *Insect Mol. Biol.* **7**: 19–29.
- Yeh, L.C., and Lee, J.C. 1992. Structure analysis of the 5' external transcribed spacer of the precursor ribosomal RNA from *Saccharomyces cerevisiae*. *J. Mol. Biol.* **228**: 827–839.
- Zucker, M. 2003. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**: 3406–3415.

1 **Ribosomal intergenic spacer in gonochoric populations of *Triops***
2 ***cancriformis*: nucleotide diversity and length variation among European**
3 **samples.**

4

5 Andrea Luchetti*, Franca Scanabissi and Barbara Mantovani

6

7 Dipartimento di Biologia Evoluzionistica Sperimentale, via Selmi 3, 40126,
8 Bologna, Italy

9

10 *Corresponding Author: Dr. Andrea Luchetti, Dipartimento di Biologia
11 Evoluzionistica Sperimentale, via Selmi 3, 40126 Bologna, Italy. Tel.: +39-051-
12 2094169; Fax: +39-051-2094286; e-mail: andluce@alma.unibo.it

13

14

1 **Abstract**

2 The eukaryotic ribosomal DNA gene family is arranged in a cluster of tandemly
3 arranged rDNA repeats, with each unit linked to the following one by a long
4 intergenic spacer (IGS). The IGS sequence of the crustacean *Triops*
5 *cancriformis* is 4.08 kbp long with a subrepeat cluster intermingled with non-
6 repetitive sequences. In this paper, the ribosomal IGS of a bisexual gonochoric
7 population of *Triops cancriformis* is first characterised and then IGS length
8 heterogeneity is compared among gonochoric, hermaphroditic and
9 parthenogenetic populations. Nucleotide variability is in line with population
10 genetics data, and hermaphrodites and parthenogenetics seems to share an
11 ancestral sequence. Mutations in functional sites suggest that gonochoric
12 population are progressively diverging from other populations. A fixed genotype
13 is scored in subsequent year samplings both in unisexual and bisexual
14 populations. In the latter, the lack of subrepeats copy number variation
15 suggests a strong impact of selection, possibly due to local environment
16 adaptation.

17

18

19 Key words: ribosomal intergenic spacer (IGS); *Triops cancriformis*; length
20 heterogeneity; molecular drive; reproductive strategies.

21

22

1 **Introduction**

2 The eukaryotic ribosomal DNA gene family is arranged in a cluster known as
3 nucleolar organizer region (NOR); it can be localized on one or more
4 chromosomes. This region is composed of hundreds of tandemly arranged
5 rDNA repeats. Each repetitive unit comprises 18S, 5.8S and 28S coding
6 regions, separated by internal transcribed spacers (ITS1 and ITS2,
7 respectively); each unit is linked to the following one by a long intergenic spacer
8 (IGS). This latter region is of particular interest for the presence of the initiation
9 and termination transcription signals, and for the occurrence of a cluster of
10 repetitive sequences (subrepeats), which seem to play an adaptive role to local
11 environment (Gorokhova et al. 2002).

12 As repetitive DNA sequences, rDNA units evolve following the pattern of
13 concerted evolution with a higher sequence similarity within than between
14 evolutionary units (i.e. strains, population, sub-species, species and so on; Nei
15 and Rooney 2005). This pattern is achieved through molecular drive, a dual
16 process summing up the effects of both genomic turnover mechanisms
17 (unequal crossing-over, gene conversion, etc.) and chromosome reshuffling by
18 bisexual reproduction (Dover 2002). Beside nucleotide sequence profiles,
19 molecular drive also provokes fluctuations in repeat length and copy number
20 (Ugarkovic and Plohl, 2002).

21 In the ribosomal IGS, length heterogeneity due to subrepeat copy number
22 variation is usually observed and this has been often correlated with differential
23 transcription and growth rate (reviewed in Weider et al., 2005). Interestingly,
24 length heterogeneity comparisons between unisexual and cyclical
25 parthenogenetic populations of the greenbug aphids *Schizaphis graminum* and
26 of the cladoceran *Daphnia pulex* showed that gonochoric reproduction

1 produces new IGS size variants through unequal crossing over (Crease and
2 Lynch, 1991; Schufran et al., 1997, 2003). This probably happens because in
3 unisexual taxa the rate of recombination is decidedly low, as observed in
4 thelytokous-laying workers of the Cape honeybees *Apis mellifera capensis*
5 (Baudry et al., 2004).

6 The tadpole shrimp *Triops cancriformis* can be found in temporary ponds and
7 rice fields of Eurasia and North Africa. It exhibits an exceptional morphological
8 stasis since the Triassic age (Longhurst 1955), thus being considered a living
9 fossil. On the other hand, it shows a consistent variation in sexual reproductive
10 strategies that range from bisexuality (either gonochoric or hermaphroditic) to
11 unisexuality (parthenogenesis; for an overview see Scanabissi et al., 2005). In a
12 previous study, the ribosomal IGS of Italian (parthenogenetic) and Austrian
13 (hermaphroditic) samples has been characterised (Luchetti et al., 2006a): the
14 sequence is 4.08 kbp long with a subrepeat cluster intermingled with non-
15 repetitive sequences (Fig. 1). The subrepeat array comprises six units (a-f)
16 either complete or incomplete. Downstream the cluster a small domain shared
17 with *D. pulex* can be found, but with still unknown function; moreover, a putative
18 promoter sequence has been identified 840 bp upstream the 18S rRNA gene
19 (Fig. 1).

20 In the present paper, the molecular characterization of the ribosomal IGS of a
21 bisexual population of *Triops cancriformis* is reported. Furthermore, IGS length
22 heterogeneity is studied in several tadpole shrimp populations characterized by
23 different reproductive strategies (gonochoric and hermaphroditic bisexuality and
24 parthenogenetic unisexuality).

25

26

1 **Material and methods**

2 Tadpole shrimps were collected in Espolla (Gerona, Spain) in a natural pond,
3 and preserved in absolute alcohol; total DNA was obtained with standard
4 phenol-chloroform extraction from the pleon tissues.

5 Amplification, cloning and sequencing of ribosomal IGS of the Spanish sample
6 were performed as described in Luchetti et al. (2006a). The obtained sequence
7 was submitted to public database Genbank, under the accession number:
8 XXYYYY. Sequences of both Italian and Austrian populations were considered
9 for appropriate comparison (Genbank acc. nos.: DQ205641 – DQ205644).

10 PCR for IGS length heterogeneity evaluation was performed on 5 – 12 samples
11 of 5 populations collected in Italy, Austria and Spain (Table 1), with Genespin S
12 to S Taq Polymerase. The subrepeat region (cluster a-f) was splitted in two
13 amplification boxes. The first box (Rep1 subcluster) comprises subrepeats a
14 and b; this was amplified with primers Rep1D (5' – TAC CCG CGT TTG ATG
15 ACT CT -3')/Rep1R (5' – CTA CCG GGC ATG GAT TTT ACG - 3'). The second
16 box (Rep2 subcluster) spans subrepeats c-f; it was amplified with primers
17 Rep2D (5' – GCC CAA TAC CCC AAC CAT AC - 3')/Rep2R (5' – ACG CTC
18 TTT GCA TCC ACT TT - 3') (Fig. 1). Southern blot analysis was done with Dig-
19 DNA Labeling and Detection kit (Roche), following the manufacturer
20 instructions. Hybridization was performed at 65°C and high stringency washes
21 were done with a solution 0.1x SSC/1% SDS.

22 Uncorrected p distances were calculated with Phylo_Win software (Galtier et
23 al., 1996; available at http://pbil.univ-lyon.fr/software/phylo_win.html).

24 Distribution of nucleotide diversity (π) across IGS sequences was calculated
25 with a sliding window analysis (10 bp window size, jumping each 5 bp), using

1 the program VariScan v. 2.0 (Vilella et al., 2005; available at
2 <http://ub.es/softevol/variscan>).

3

4 **Results and Discussion**

5 Long-PCR amplification on Espolla sample genomic DNA gave a 4752 bp DNA
6 fragment corresponding to the complete IGS sequence (3892 bp) linked to
7 neighbouring sequences (28S 3' end - 492 bp and 18S 5' end - 367 bp). The
8 sequenced 28S and 18S gene regions are completely identical to those found
9 in Italian and Austrian samples (Luchetti et al., 2006a), with the only exception
10 of a single point mutation in the 28S.

11 The IGS region shows G+C richness equal to the 52.3%. This value is in line
12 with those found in previously analysed populations, but it sharply differs from
13 other Crustacean IGS that were found A+T rich (Crease et al., 1993 and
14 reference therein). At the sequence level, the Spanish sample differs
15 significantly from Italian and Austrian ones, with nucleotide divergence of 2.6%-
16 2.7%. Furthermore, several indels can be observed in this comparison, and a
17 large deletion occurs in the Spanish sample within the second box of the sub-
18 repeat array: indeed, the sub-repeat "d" is lacking. This observation is in line
19 with microsatellite analysis that points to a substantial differentiation of the
20 Espolla sample from other European populations (Mantovani et al., submitted).

21 A sliding window analysis was performed comparing Italian, Austrian and
22 Spanish sequences to verify if the nucleotide diversity is equally distributed
23 across the sequence or if there are conserved regions. Any specific domain
24 appears well conserved among the three samples (Fig. 2) and nucleotide
25 variation is scored also at the sub-repeats loci, at the small domain and at the
26 gene transcription promoter. As far as the small domain is concerned, its

1 sequence (WWTTCTAAGTCC) is disrupted in the Spanish sample by the
2 tandem duplication of a short motif (TTTTGGGAA), producing a direct and an
3 inverse repeat (Fig. 2). This condition sharply differs from the domain
4 conservation observed between parthenogenetic (Italian) or hermaphroditic
5 (Austrian) *T. cancriformis* populations and the cladoceran *D. pulex* (Crease,
6 1993). The domain sharing suggests their sequence as the ancestral one. The
7 different behaviour of *T. cancriformis* populations may be linked to their different
8 reproductive strategies. Unisexuality and hermaphroditism, due to the
9 inbreeding, conserved and fixed the ancestral sequence, while gonochoric
10 bisexuality in the Spanish sample allowed the evolution toward a new domain.
11 The gene promoter results quite conserved, as it is somehow expected, being
12 identical in previously analysed taxa and with only one point mutation scored in
13 the Spanish sample (Fig. 2). Transcription of pre-rRNA requires a dedicated set
14 of proteins able to bind IGS in the promoter region and to recruit the RNA pol-I
15 (Moss et al., 2006). Each mutation arising in the promoter would spread among
16 rDNA repeats accordingly to the concerted evolution principle in a gradual and
17 cohesive manner. This would allow the elimination through selection or drift of
18 those protein variants (and then alleles) that do not fit the new promoter variant,
19 leading to the (co)evolution of a species-specific transcriptional machinery
20 (Dover and Flavell, 1984). Therefore, the presence of a point mutation in the
21 Spanish promoter may indicates that, given enough time, this population would
22 become genetically isolated from the other *T. cancriformis* populations
23 analysed.
24 Length heterogeneity analysis did not evidence multiple bands when amplifying
25 the Rep1 subcluster. On the ther hand, different banding patterns were
26 observed among populations in the Rep2 box analysis (Table 1; Fig. 3).

1 Austrian and Sardinian samples show a single band of equal length,
2 corresponding to the complete Rep2 region (c+d+e+f). The other Italian
3 populations (Ferrara and Grosseto) show a four-banded pattern, with the band
4 corresponding to the complete Rep2 region as the most represented and the
5 one completely lacking subrepeats only lightly detectable. The Spanish samples
6 show only two bands: accordingly to sequence data, the most represented band
7 lacks subrepeat d, and a second band, shared with Italian specimens, lacks of
8 both c and d subrepeats. The former band it is completely absent in all other
9 analysed *T. cancriformis* populations, indicating that this genotype has been
10 fixed only in the Iberian specimens. On the whole, within-sample IGS length
11 heterogeneity can be observed both for gonochoric and parthenogenetic
12 samples, while a fixed genotype occurs in Austrian hermaphrodites and in the
13 parthenogenetic population of Oristano (Table 1; Fig. 3).

14 In greenbug aphid, parthenogenetic lineages show stable IGS length profiles
15 over the time. Variation in length heterogeneity was observed only after inter-
16 and intra-clone mating (Shufran et al., 1991, 1997). This led to the hypothesis
17 that mechanisms of genomic turnover, such as unequal crossing over, while
18 maintaining sequence homogeneity, may produce IGS length variation. The fact
19 that variation in IGS genotype is observed only after mating, i.e. gonochoric
20 reproduction, point to the absence, or very reduced, recombination rate in
21 parthenogenetic taxa (Baudry et al., 2004).

22 No IGS genotype changes occurred in the two years sampling of both Espolla
23 and Grosseto (Table 1). While length heterogeneity conservation is somehow
24 expected in the “frozen” parthenogenetic genotype of Grosseto, it is surprising
25 for the gonochoric sample of Espolla, where genomic turnover mechanisms and
26 amphimixis are the rule. Further, the gonochoric sample shows a lower number

1 of IGS variants with respect the unisexual ones, at variance of *D. pulex* (Crease
2 and Lynch, 1991). This is very difficult to explain only in the light of molecular
3 drive. Possibly, some kind of selective forces are maintaining those IGS length
4 variants in the Spanish sample, being well adapted to local environment
5 (Gorokova et al., 2002).

6 Austrian hermaphrodite and Oristano parthenogenetic populations show the
7 same, single banded IGS genotype: this shifting toward a specific IGS variant
8 could be again an effect of selection. Clonal lineages of greenbug aphids
9 showed reduced number of IGS length variants and the preferential
10 amplification of a particular size variant as an effect of insecticide selection
11 (Shufran et al., 2003). Thus, the fixed genotype observed in the hermaphrodites
12 and in the Oristano parthenogenetic sample could be the results of a similar
13 process, where environmental condition may have selected the (c+d+e+f) IGS
14 variant. For example, it has been demonstrated that longer IGS size would
15 correspond to increased growth rates (Weider et al., 2005). However, while both
16 populations live in natural ponds, environmental conditions of Austrian and
17 Sardinian (Oristano) ponds are rather different. On the other hand,
18 mitochondrial DNA analyses (Mantovani et al., submitted) indicate gene flow
19 (migration) between these samples. Therefore, it is possible to hypothesize that
20 this particular IGS genotype originated in one population and then spread in the
21 other one simply by gene flow.

22 As far as the hermaphrodites are concerned, data presented here are the first
23 analyzing IGS length heterogeneity in this peculiar reproductive dynamics.
24 Austrian populations were described as hermaphrodites, but it is not known to
25 date if they reproduce by selfing (Longhurst, 1955). However, as the gamete

1 production proceeds as in gonochoric populations, and then unequal crossing
2 over should take place, IGS length heterogeneity is expected to vary.
3 It has been demonstrated that molecular drive could be affected by specific
4 organismal traits, such as reproductive modes (Luchetti et al., 2003, 2006b;
5 Lorite et al., 2004); furthermore, unisexuality seems to influence the evolution of
6 IGS size variants (Crease and Lynch, 1991; Schufran et al., 1997, 2003). Data
7 presented here, on the other hand, point to an extreme conservation of the IGS
8 genotype within populations and on the lack of significant variation in
9 gonochoric individuals; therefore, the effect of natural selection on the IGS size
10 in *T. cancriformis* seems to play a role greater than molecular drive.

11

12 **Acknowledgements**

13 We wish to thank Dani Boix, Michele Cesari and Loris Mularoni for collecting
14 tadpole shrimps in Espolla. This work has been supported by RFO and
15 Canziani funds.

16

1 **References**

2

3 Baudry E, Kryger P, Allsop M, Koeniger N, Vautrin D, Mougél F, Cornuet
4 J-M, Solignac M (2004). Whole-genome scan in thelytokous-laying workers of
5 the Cape Honeybee (*Apis mellifera capensis*): central fusion, reduced
6 recombination rate and centromere mapping using half-tetrad analysis.
7 *Genetics* 167: 243-252.

8 Crease TJ (1993). Sequence of the intergenic spacer between the 28S
9 and 18S rRNA-encoding gene of the crustacean, *Daphnia pulex*. *Gene* 134:
10 245-249.

11 Crease TJ, Lynch M (1991). Ribosomal DNA variation in *Daphnia pulex*.
12 *Molecular Biology and Evolution* 8: 620-640.

13 Dover GA (2002). Molecular drive. *Trends in Genetics* 18: 587-589.

14 Dover GA, Flavell RB (1984). Molecular coevolution: DNA divergence and
15 the maintenance of function. *Cell* 38: 622-623.

16 Galtier N, Gouy M, Gaultier C (1996). SEAVIEW and PHYLO_WIN: two
17 graphic tools for sequence alignment and molecular phylogeny. *CABIOS* 12:
18 543-548.

19 Gorokhova E, Doeling TE, Weider LJ, Crease TJ, Elser JJ (2002).
20 Functional and ecological significance of rDNA intergenic spacer variation in a
21 clonal organism under divergent selection for production rate. *Proceedings of*
22 *the Royal Society of London Series B* 269: 2373-2379.

23 Longhurst AR (1955). A review of the Notostraca. *Bulletin of the British*
24 *Museum (Natural History), Zoology* 3: 1-57.

1 Lorite P, Carrillo JA, Tinaut A, Palomeque T (2004). Evolutionary
2 dynamics of satellite DNA in species of the genus *Formica* (Hymenoptera,
3 Formicidae). *Gene* 332: 159-168.

4 Luchetti A, Cesari M, Carrara G, Cavicchi S, Passamonti M, Scali V,
5 Mantovani B (2003). Unisexuality and molecular drive: Bag320 sequence
6 diversity in *Bacillus* taxa (Insecta Phasmatodea). *Journal of Molecular Evolution*
7 56: 587-596.

8 Luchetti A, Scanabissi F, Mantovani B (2006a). Molecular characterization
9 of ribosomal intergenic spacer in the tadpole shrimp *Triops cancriformis*
10 (Crustacea, Branchiopoda, Notostraca). *Genome* 49: 888-893.

11 Luchetti A, Marini M, Mantovani B (2006b). Non-concerted evolution of the
12 RET76 satellite DNA family in *Reticulitermes* taxa (Insecta, Isoptera). *Genetica*
13 128:123–132.

14 Moss T, Stefanovsky, Langlois F, Gagnon-Kugler T (2006). A new
15 paradigm for the regulation of the mammalian ribosomal RNA genes.
16 *Biochemical Society Transaction* 34: 1079-1081.

17 Nei M, Rooney AP (2005). Concerted and Birth-and-Death evolution of
18 multigenes families. *Annual Review in Genetics* 39:121-52.

19 Scanabissi F, Eder E, Cesari M (2005). Male occurrence in Austrian
20 populations of *Triops cancriformis* (Branchiopoda, Notostraca) and
21 ultrastructural observation of the male gonad. *Invertebrate Biology* 124: 57-65.

22 Schufran KA, Black WC IV, Margolies DC (1991). DNA fingerprinting to
23 study spatial and temporal distributions of an aphid, *Schizaphis graminum*
24 (Homoptera: Aphididae). *Bulletin of Entomological Research* 81: 303-313.

1 Schufran KA, Peters DC, Webster JA (1997). Generation of clonal
2 diversity by sexual reproduction in the greenbug, *Schizaphis graminum*. Insect
3 Molecular Biology 6: 203-209.

4 Schufran KA, Mayo ZB, Crease TJ (2003). Genetic changes within an
5 aphid clone: homogeneization of rDNA intergenic spacer after insecticide
6 selection. Biological Journal of the Linnean Society 79: 101-105.

7 Ugarkovic D, Plohl M (2002). Variation in satellite DNA profiles: causes
8 and effects. EMBO Journal 21: 5955-5959.

9 Vilella AJ, Blanco-Garcia A, Hutter S, Rozas J (2005). VariScan: analysis
10 of evolutionary patterns from large-scale DNA sequence polymorphism data.
11 Bioinformatics 21: 2791-2793.

12 Weider LJ, Elser JJ, Crease TJ, Mateos M, Cotner JB, Markow TA (2005).
13 The functional significance of ribosomal (r)DNA variation: impacts on the
14 evolutionary ecology of organisms. Annual review in Ecology Evolution and
15 Systematics 36: 219-242.

16

17

1 **Figure legends**

2

3 **Fig. 1** Schematic drawing of *T. cancriformis* ribosomal IGS (modified from
4 Luchetti et al., 2006a). Ovals and emi-ovals represent complete and incomplete
5 subrepeats, respectively. Arrows represent primers for length heterogeneity
6 analysis (1/1' for Rep1D/R and 2/2' for Rep2D/R; see text). SD: small domain
7 found in *D. pulex*; Ptsp: putative promoter sequence and transcription starting
8 point.

9

10 **Fig. 2** Sliding window analysis of nucleotide diversity. The absence of the “d”
11 subrepeat, the disruption of the small domain shared with *D. pulex* (Dpu) and
12 the mutation within the putative promoter sequence occurring in the Spanish
13 sample (SPA) with respect to Italian (ITA) and Austrian (AUS) ones are
14 reported.

15

16 **Fig. 3** Comparative Southern blot analysis of Rep2 subcluster among
17 populations. Arrowheads indicate bands and on the left side the IGS genotypes
18 are reported. Sample acronyms are as in Table 1.

19

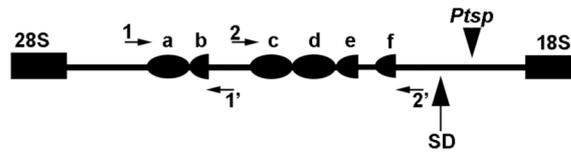
20

1 **Table 1.** Sampling localities, reproductive mode (RM), number of analysed
 2 specimens (N) and resulting IGS genotype of *T. cancriformis* populations (§: H
 3 = hermaphrodite, P = parthenogenetic, G = gonochoric; °: letters refer to
 4 subrepeat as in Fig 1; *: number refers to collecting years)

Sample site	Acronym	RM§	N	IGS genotype°				
				(c+d+e+f)	-e or -f	-d	-(c+d)	-(c+d+e+f)
<i>Austria</i>								
Marchegg	MAR	H	5	+	-	-	-	-
<i>Italy</i>								
Ferrara	FER	P	10	+	+	-	+	+
Grosseto 02*	GRO02	P	10	+	+	-	+	+
Grosseto 03*	GRO03	P	10	+	+	-	+	+
Oristano	ORI	P	5	+	-	-	-	-
<i>Spain</i>								
Espolla 04*	SPA04	G	12	-	-	+	+	-
Espolla 06*	SPA06	G	8	-	-	+	+	-

5

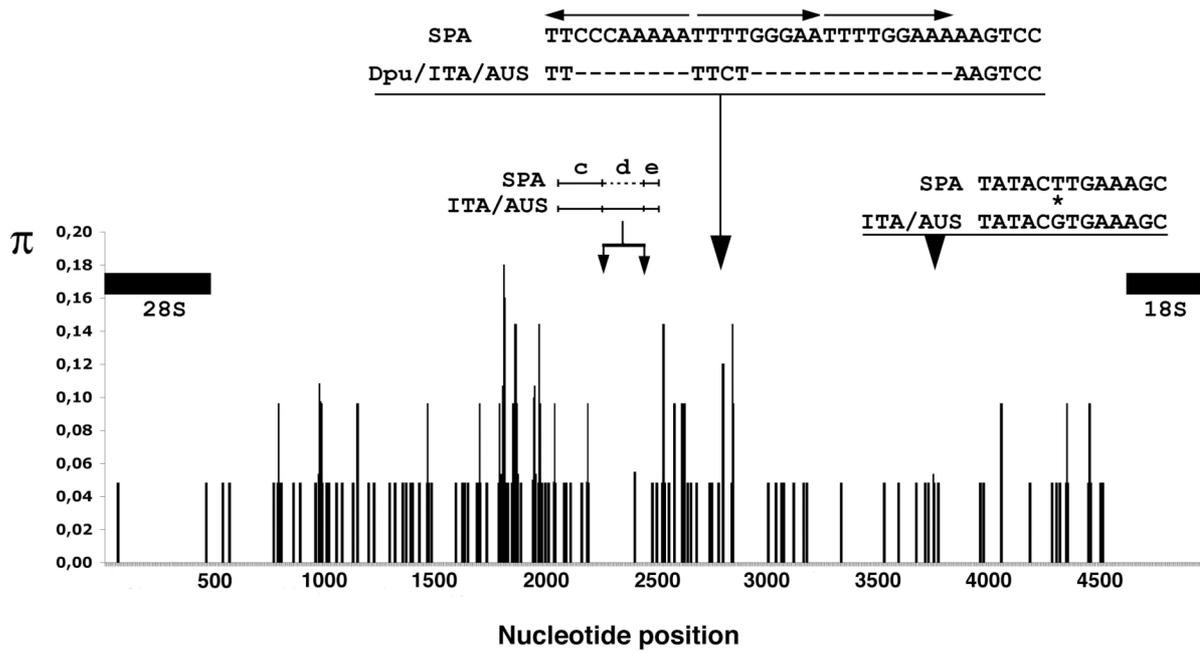
6



1

2 Figure 1

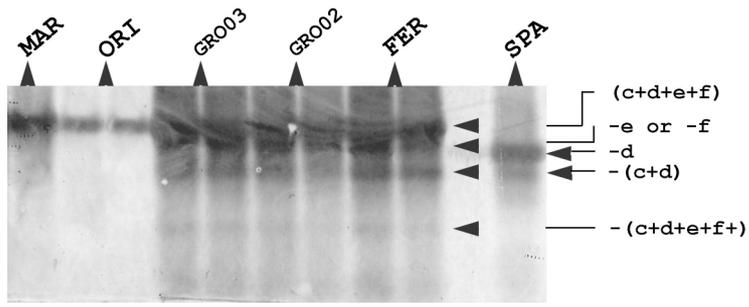
3



1

2 Figure 2

3



1

2 Figure 3

3

Chapter 4. LEP150 repetitive DNA evolution in clam shrimps

The studies on the evolution of LEP150 repetitive DNA in the clam shrimp *Leptestheria dahalacensis* are here reported. As a preliminary analysis, clam shrimp European populations were genotyped at microsatellite and mitochondrial loci. Then, LEP150 sequence variability was analysed among Italian, Austrian and German populations. Repeats neighbouring the 5S rRNA gene were of particular interest and evidenced the co-evolution of the two repetitive DNA families. In this context, a particular condition of molecular drive, called “molecular sweep”, was suggested for the first time.

The results will be presented in the following papers:

Cesari M, Luchetti A, Scanabissi F, Mantovani B (2007). Genetic variability in European *Leptestheria dahalacensis* (Rüppel, 1837) (Crustacea, Branchiopoda, Spinicaudata). *Hydrobiologia*, in press

Luchetti A, Marino A, Scanabissi F, Mantovani B (2004). Genomic dynamics of a low copy number satellite DNA family in *Leptestheria dahalacensis* (Crustacea, Branchiopoda, Conchostraca). *Gene* 342: 313-320

Luchetti A, Scanabissi F, Mantovani B - Evolution of LEP150 sub-repeat array in the ribosomal IGS of the clam shrimp *Leptestheria dahalacensis* (Crustacea Branchiopoda Conchostraca): the molecular sweep hypothesis. Submitted

These results were also presented at the following symposia/Congresses:

Luchetti A, Cesari M, Scanabissi F, Mantovani B (2004). Genetic variability of repetitive sequences in Italian and Austrian populations of *Leptestheria dahalacensis* (Rüppel 1837) (Conchostraca). V° International Large Branchiopod Symposium, Toodyay, Western Australia 16-20 August

Luchetti B, Scanabissi F, Mantovani B (2006). Co-evolution of 5S rDNA and LEP150 repetitive sequences within ribosomal IGS in the clam shrimp

Leptestheria dahalacensis (Crustacea, Branchiopoda). Second Meeting of Italian Evolutionary Biologists, 4-7 September 2006, Florence, Italy

Hydrobiologia: in press

DOI 10.1007/s10750-007-0645-2

**Genetic variability in European *Leptestheria dahalacensis* (Rüppel, 1837)
(Crustacea, Branchiopoda, Spinicaudata)**

Michele Cesari, Andrea Luchetti, Franca Scanabissi, Barbara Mantovani

*Dipartimento di Biologia Evoluzionistica Sperimentale, Università di Bologna,
via Selmi 3, 40126 Bologna, Italy*

Correspondence: Michele Cesari; *Fax* +39 051 2094286; *E-mail:*
cesari_michele@yahoo.it

Keywords: 12S, 16S, cytochrome oxidase I, clam shrimp, geographic dispersal, microsatellites.

Abstract

The genetic variability of the gonochoric *Leptestheria dahalacensis* (Rüppel, 1837) was studied through the analysis of mitochondrial and nuclear (microsatellite loci) markers in eight Italian and two Central European populations. Mitochondrial data exhibited a low variability, as only six mitotypes were scored: five in Italy and one for both Central European samples, with a very low number of substitutions. All analysed microsatellite loci were variable, with 3-5 alleles per locus and 1-4 alleles per population. All populations were at the Hardy-Weinberg equilibrium, with the exceptions of two samples for locus IdAC-16, due to heterozygote excess, and of four populations for locus IdAC-11, probably linked to the presence of null alleles. A substantial population structuring was found between Central European and Italian samples for both utilized markers. This observation may be explained by isolation by distance and/or recent isolation events. On the other hand, the absence of a clear inter-pond variability in Italian sample comparisons may be ascribed to high dispersal ability in the short range.

Introduction

Branchiopoda are primitive Crustacea mainly inhabiting astatic waters (i.e. water bodies with fluctuating surface levels). These may be interpreted as island-like habitats in a terrestrial landscape (De Meester et al., 2002). Branchiopoda are unable of active dispersal but they allow passive transport, producing resting eggs (Thiery and Pont, 1987; Brendonck and Riddoch, 1999; Figuerola and Green, 2002; Figuerola et al., 2005), ensuring survival in unsuitable conditions (Martin, 1992).

DNA molecular studies are generally limited in Branchiopoda, a group of interest for its ancient origin, reproductive biology and life cycles (Dumont and Negrea, 2002).

As part of a continuing project aiming at the genetic characterization of European Spinicaudata and Notostraca, also in the light of different modes of reproduction, ranging from unisexual to bisexual gonochoric and hermaphroditic populations, we approached the analysis of genetic variability levels in the gonochoric *Leptestheria dahalacensis* (Rüppel, 1837), a species widely distributed in central-southern Europe, by means of biparental (nuclear) and matrilinear (mitochondrial) DNA markers. So far, the only available genomic analysis in *L. dahalacensis* relates to the finding of a highly repeated satellite DNA, the LEP150 family. This is a low copy number satellite DNA (0.5% of the genome), with 150 bp monomeric units and mean sequence diversity among them of up to 10.3%. (Luchetti et al., 2004).

As nuclear markers we chose microsatellites: these are short units of 2-6 bp, repeated up to 100 times in the tandem array (Chambers & MacAvoy, 2000), and are found at very high frequencies in the genome of every organism analyzed so far (Li et al., 2002). They are usually characterised by an elevated

polymorphism, possibly due to slippage events during replication (Ellegren, 2000). Such peculiarity makes them very powerful genetic markers: microsatellites proved to be valuable tools for phylogeography and reproductive biology studies at the population level in branchiopods and other animals exhibiting resting stages (Pálsson, 2000; Pfrender et al., 2000; Limburg & Weider, 2002, Gómez et al., 2002; Figuerola et al., 2005; Colautti et al., 2005). On the other hand, mitochondrial analyses of the Spinicaudata, in particular the Leptestheriidae family, are scarce and data are available only at the family or genus level. A survey on phylogenetic relationships based on mitochondrial 12S and nuclear EF1- α genes by Braband et al. (2002) confirmed the monophyly of Spinicaudata and the relationships among the taxa within this order as already inferred by Fryer (1987) on morphological data. A recent analysis of the family Limnadiidae on nuclear (28S) and mitochondrial (12S and cytochrome b) genes found inconsistencies with prior hypotheses of inter-generic relationships (Hoeh et al., 2006).

In this study we report the isolation and analysis of three microsatellite loci and the analyses of three mitochondrial genes (12S, 16S, cytochrome oxidase I) in *L. dahalacensis* European populations to verify population variability and to contribute to the knowledge of both nuclear and mitochondrial genomes in the order Spinicaudata.

Materials and Methods

Animals

Specimens were sampled in different years from five Italian rice ponds located in an area of 131 Km² near Ferrara, from two different rice paddies in Isola della Scala, near Verona (about 115 km NW from Ferrara), and from natural pools in

Marchegg (Austria) and in Hadamar (Germany) (about 530 km and 990 km from Ferrara, respectively; Fig. 1; Table 1). Total DNA was extracted from single individuals, following CTAB (Doyle & Doyle, 1987) or phenol/chloroform (Sambrook et al., 1989) protocols. Two specimens for each population were analysed for mitochondrial genes (Table 1), while 6 to 24 individuals for each sample were genotyped at the three microsatellite loci (MSL; Table 3). Specimens analysed for mitochondrial data were also genotyped at the three MSL.

Mitochondrial analyses

PCR amplification was performed in 50 µl reactions using the Invitrogen PCR kit with recombinant *Taq* DNA polymerase and the following protocol: initial denaturation at 94°C for 5 min, 35 cycles with denaturation at 94°C for 30 sec, annealing at 48°C for 30 sec, extension at 72°C for 30 sec, followed by a final extension at 72°C for 7 min. The amplified products were purified with the Wizard PCR cleaning (Promega) kit and both strands were sequenced in an ABI PRISM 310 Genetic Analyzer (Applied Biosystems). The primers for PCR amplification and sequencing (Invitrogen) were: mt-35 (5'-AAG AGC GAC GGG CGA TGT GT-3') and mt-36 (5'-AAA CTA GGA TTA GAT ACC CTA TTA T-3') for the 12S gene; mt-32 (5'-CCG GTC TGA ACT CAG ATC ACG T-3') and mt-34 (5'-CGC CTG TTT AAC AAA AAC AT-3') for the 16S gene; COI-F (5'-GGT CAA CAA ATC ATA AAG ATA TTG G-3') and COI-R (5'-TAA ACT TCA GGG TGA CCA AAA AAT CA-3') for the cytochrome oxidase I (COI) gene. Primers were derived from Simon et al. (1994; 12S and 16S genes), and from Folmer et al. (1994; COI gene). Alignments performed with the Clustal algorithm of the Sequence Navigator program (ver 1.0.1, Applied Biosystems) were also checked by visual inspection. The nucleotide sequences of the newly analysed specimens have been

submitted to GenBank (A.N.: DQ872781-2, 12S; AY159586, DQ872783-5, 16S; DQ872786, COI). Absolute numbers of nucleotide substitutions between mitotypes (i.e. combined mitochondrial haplotypes) were determined using MEGA version 3.1 (Kumar et al., 2004).

Microsatellite analysis

A dinucleotide microsatellite enriched library was obtained using the F.I.A.S.CO. protocol (Zane et al., 2002). Thirty-two recombinant colonies were screened through amplification, using M13 primers. The amplicon was purified with the Wizard PCR cleaning kit (Promega) and sequenced in an ABI 310 Genetic Analyzer (Applied Biosystems). Oligonucleotide primers were then designed using Primer3 software (Rozen & Skaletsky, 2000; Table 2) and were optimized for PCR amplification by testing over a range of MgCl₂ concentrations and annealing temperatures. The PCR reactions (10 µl total volume) included 4 ng of genomic DNA, a variable MgCl₂ concentration (see Table 2), 10 mM of each primer, 200 mM dNTPs, 1 µl of 10x Buffer (Invitrogen kit) and 1 U of *Taq* polymerase (Invitrogen). Amplifications were performed in a GeneAMP PCR System 2400 (Applied Biosystems), as follows: initial denaturation at 94° C for 5 min, 35 cycles at 94° C for 30 s, 58° C for 30 s, 72° C for 30 s, followed by a final holding at 72° C for 7 min. Genotyping of individuals was performed in a Beckman CEQ8000, using 5' labelled (Invitrogen) forward primers. Observed and expected heterozygosities, allelic frequencies and number of migrants per generation ($Nm=(1-F_{ST})/(4*F_{ST})$; Wright, 1969) were computed using Genetix 4.05 (Belkhir et al., 2004); probability to fit to Hardy-Weinberg equilibrium (HWE), linkage disequilibrium and correlation tests between population differentiation and geographical distance were carried out using Genepop 1.2 (Raymond & Rousset, 1995).

Allelic richness and F-statistics were computed using FSTAT 2.9.3 (Goudet, 2001). Possible link between number of analysed individuals and observed heterozygosity was tested by means of correlation coefficient (R^2). An assignment test to evaluate the probability for each genotype/individual in each population to occur in a different one was computed with GeneClass2 (Piry et al., 2004) based on Paetkau et al. (1995) algorithm. Given that polymorphism at annealing sites of the microsatellite primers can prevent the amplification of a particular allele, therefore resulting in heterozygote deficiencies, the presence of null alleles was tested with Microchecker 2.2.1 (Van Oosterhout et al., 2004). The Neighbor Joining tree was calculated with Nei et al. (1983) distances using Populations 1.2.28 (written by Olivier Langella), with 10000 bootstrap replicates.

Results

Mitochondrial analysis

Overall, 1438 base pairs (bp) were sequenced: 341 bp for the 12S gene, 502 bp for the 16S gene and 595 bp for the COI gene. The scored variability is low: the 12S gene exhibits 2 haplotypes, differing for 2 substitutions; the 16S gene is more variable, with 4 haplotypes discriminated by 1-3 mutations and the COI gene exhibits the same haplotype for all analyzed individuals. When combining the three data sets, six mitotypes are scored out of 20 total sequences (Table 1). Mitotype F characterises the Austrian and German specimens, and it is differentiated from the others by 1-4 substitution(s). In the Italian populations, the Verona ponds share mitotype D, which was also scored in the Leona and Gran Linea samples, while mitotype B has been found in all Ferrara pools except Leona. Among the Ferrara samples private mitotypes occur: A and C are found only in Contane and Amiani ponds, respectively, while mitotype E is

observed only in Leona and, together with mitotype D, it is conserved in subsequent years (2002-2003). It is to be noted that in all Ferrara populations but Mezzogoro, the two analysed individuals show different mitotypes. The Italian mitotypes differ from each other by 1–5 nucleotide substitution(s).

Microsatellite analysis

The three microsatellite loci (MSL) are polymorphic in all samples, with the only exception of IdAC-11 for the sample Isola della Scala A (Fig. 2), with allelic richness ranging between 1.953 and 3.208 (Table 3). At the IdAC-10 locus, all populations share a most common allele (153); a private allele occurs in the German sample (Fig. 2). For locus IdAC-11, allele 207 is the most represented, with the only exception of Leona A sample. At locus IdAC-16, allele 174 is present at higher frequencies only in Italian populations. At loci IdAC-11 and IdAC-16 the Austrian and German samples share two alleles absent in the Italian specimens.

Observed heterozygosity at polymorphic loci ranges from 0.063 (IdAC11; Contane) to 0.818 (IdAC11; Hadamar). In all instances but six, observed heterozygosity matches the expected one. In particular, at the IdAC-11 and IdAC-16 loci a significantly lower or higher number of heterozygous genotypes is scored in four (Contane, Mezzogoro, Isola della Scala B and Marchegg) and two (Mezzogoro, Leona B) populations, respectively (Table 3). The absence of correlation between number of analysed individuals and observed heterozygosity ($R^2 = 0.04$; $P = 0.57$) suggests that the scored variability is not linked to sample size.

F_{IT} and F_{IS} values are significant only for the IdAC11 locus. This reflects the observed deviations from HWE, which may be explained by the presence of null alleles in 5 out of 10 samples (Table 3). Linkage disequilibrium analyses are

non significant for all pair of loci, indicating no association between genotypes (IdAC-10 & IdAC-11 $P=0.668$; IdAC-10 & IdAC-16 $P=0.997$; IdAC-11 & IdAC-16 $P=0.426$).

The overall F_{ST} significant value indicates a substantial genetic structuring of *L. dahalacensis* samples (Table 3). Among Italian populations, the pairwise F_{ST} values point out the Leona A and Isola della Scala A demes as the most differentiated from the other samples. Furthermore, a significant differentiation of the Central European populations is found both between them and in comparison with the Italian samples (Table 4).

Migration rate estimates (Nm) are very low between the Central European samples and the Italian ones. On the other hand, a general high rate was found among the Italian populations, with the exceptions of Leona A and Isola della Scala A samples (Table 4).

In the assignment test only three individuals of different Italian populations (Mezzogoro, Leona B and Isola della Scala B) could be cross-assigned to other Italian samples (Leona B, Mezzogoro and Leona A, respectively). No relationship emerges between newly assigned individuals and sampling site or year (not shown).

The unrooted Neighbor-Joining dendrogram (Fig. 3) shows the two related populations from Verona clustering together with the Ferrara sample Contane. Leona A and Mezzogoro samples appear related, but no other clear relationship emerges between Ferrara samples. On the other hand, a higher affinity of Austrian and German samples is scored, together with their high divergence from Italian demes.

The significant positive correlation between genetic and geographic distances ($R^2 = 0.541$; $P < 0.05$) indicates that the divergence observed between analyzed populations could be due to isolation by distance (Fig. 4).

As far as collecting years are concerned, the only significant F_{ST} value is that scored between Leona A and Leona B, i.e. samples derived from the same pond but sampled in two subsequent years.

Discussion

The evaluated genetic parameters point to a general low variability of presently analysed *L. dahalacensis* samples with respect to other crustaceans both for mitochondrial and microsatellite markers.

Despite having analysed more than 1400 mitochondrial bp, little variability is observed, with an overall differentiation limited to a maximum of five polymorphic sites in the combined dataset. Two out of three analysed genes (12S and 16S) were successfully used in previous surveys on phylogenetic relationships in Branchiopoda (Hanner & Fugate, 1997; Murugan et al., 2002; Mantovani et al., 2004; Korn et al., 2006; Hoeh et al., 2006; Stenderup et al., 2006) and proved to be informative. They also revealed the presence of cryptic species within some Notostracan taxa (King & Hanner, 1998; Korn & Hundsdoerfer, 2006). Moreover, the scoring of only one haplotype for the COI gene is of particular interest, because this gene, albeit its protein-coding role, often shows a consistent variability (Folmer et al., 1994; deWaard et al., 2006). Therefore, our results seem to suggest mtDNA mutation rate heterogeneity in different Branchiopod taxa. It should be noted, on the other hand, that even if mitochondrial DNA analysis does not provide evidence for any detailed phylogeographic pattern, the Central European samples exhibit the same

private mitotype, which appears therefore distributed over a wide geographic area. On the contrary, variability is observed in the Italian populations of Ferrara, where different mitotypes occur in the same pond in four out of five localities.

Nuclear data point out that the level of among population differentiation across all loci ($F_{ST} = 0.082$) is low when compared to previous analyses on Spinicaudata (*Eulimnadia texana*: $F_{ST} = 0.284$, Weeks & Duff, 2002), Anostraca (*Branchipodopsis paludosa*: $F_{ST} = 0.360$, Boileau et al., 1992; *Branchipodopsis wolffi*: $F_{ST} = 0.291$, Brendonck et al., 2000) and Cladocera (*Daphnia pulex*: $F_{ST} = 0.414-0.445$, Pålsson, 2000), while being comparable to data obtained on other Anostraca (*Artemiopsis stefanssoni*: $F_{ST} = 0.075$, Boileau et al., 1992; *Artemia franciscana*: $F_{ST} = 0.12$; Abreu-Grobois, 1987). Nevertheless, these comparisons should be considered with care because different factors have to be taken into account (f.i. number and types of loci, either allozymes or microsatellites, and geographic scale).

Actually, polymorphism values at the presently analyzed MSL are decidedly higher than those scored in five *L. dahalacensis* Italian populations through the analysis of 22 allozyme loci (Tinti & Scanabissi, 1996). In that paper, mean observed heterozygosities ranged from 0.10 to 0.14 with some instances of HW departures due to heterozygote deficiency. This observation confirms the more powerful resolution of microsatellite markers, a well established pattern found both in animals (see f.i. Wirth and Bernatchez, 2001; Corujo et al., 2004) and in plants (Nybom, 2004). Present analysis highlights significant deviations from HW equilibrium owing to either heterozygote deficiency (locus IdAC-11) or heterozygote excess (locus IdAC-16). Departures from Mendelian ratios in locus IdAC-11 may be explained by the presence of null alleles, while the homozygote

deficiency at the IdAC-16 locus in two populations is more difficult to explain, even if it is not uncommon in microsatellite analyses (see f.e. Reece et al., 2004).

With regard to the observed pattern of genetic differentiation in Ferrara samples, pairwise F_{ST} values show a high relatedness between them, with the exception of the Leona A sample, which is very differentiated also from the specimens collected in the same pond one year later (sample Leona B).

The significant differentiation of the Leona A sample seems mainly due to a switch in the frequency between alleles 207 and 209 at the IdAC-11 locus. However, it is to be noted that both Leona A and Leona B share the same mitotypes (D and E).

Given this pattern of differentiation and the absence of private alleles, it is possible that analyzing more samples or a higher number of MSL may lead to an absence of differentiation between these samples. At the mitochondrial level only Contane and Amiani samples show a private mitotype (A or C), but it always co-occurs with a more widespread one. On the whole, therefore, the absence of clear-cut inter-pond variability in Ferrara samples at both nuclear and mitochondrial level seems to reflect high short-range dispersal ability (and then gene flow) rather than historical colonization of new habitats (De Meester et al., 2002). Such mobility could be enhanced by other media than birds (Thiery and Pont, 1987), such as wind (a short-distance factor in Anostraca; Brendonck and Riddoch, 1999), and anthropogenic causes, for example in rice fields.

On the other hand, when analyzing all samples together, a significant isolation by distance can be observed. A reduced long-range dispersal ability can be invoked to explain the differentiation of the Verona populations: branchiopod

migrations are thought to cover long distances only thanks to birds acting as a dispersal agent for the resting eggs (Dumont & Negrea, 2002; Figuerola & Green, 2002; Figuerola et al., 2005).

Only MSL data show divergence between Austrian and German *L. dahalacensis* populations, while, even if with different resolution power, both analyzed molecular markers point to the differentiation of the two Central European samples with respect to the Italian ones. Again, low long-range dispersal ability can be taken into account to explain this pattern of divergence. It should be further considered that the Alps constituted a geographical barrier for many organisms during postglacial recolonization (Hewitt, 1996), so they might be responsible for the reduction of the amount of genetic exchange and thus strengthening the isolation. The low values of migration rate observed between Italian and Central European samples reinforce this hypothesis. Even if no data are available for Spinicaudata, and considering that only two Central European populations were taken into account, it could be speculated that some *L. dahalacensis* individuals may have crossed the Alps, possibly in a single event, giving rise to a new deme from which both Marchegg and Hadamar populations originated. The Central European F mitotype, highly related with those scored in Italian demes, suggests that the colonization happened fairly recently.

Conclusion

This is the first analysis assessing genetic variability in *L. dahalacensis*, using both mitochondrial and nuclear markers. Even if scored differentiation is generally low, interesting data on population structuring and intra-pond diversity emerges. Future comparative studies on other European spinicaudatan taxa,

which also exhibit different modes of reproduction, will contribute to the knowledge of genetic variability levels and trends in relationships to the peculiar habitat represented by astatic waters and to the different reproductive strategies.

Acknowledgements

We wish to thank E. Eder and I. Frolova for supplying the Austrian and German specimens, respectively. This work was supported by M.U.R.S.T. 40% funds.

References

Abreu-Grobois, F.A., 1987. A review of the genetics of *Artemia*. In: Sorgeloos, P., D.A. Bengtson, W. Decler & E. Jaspers (eds) *Artemia* research and its applications, vol. 1 Universa, Wetteren, pp. 61-99.

Belkhir, K., P. Borsa, L. Chikhi, N. Raufaste & F. Bonhomme, 1996-2004. GENETIX 4.05, logiciel sous Windows™ pour la génétique des populations. Laboratoire Génome, Populations, Interactions, CNRS UMR 5000, Université de Montpellier II, Montpellier (France). (code available at <http://www.univ-montp2.fr/~genetix/genetix/genetix.htm>)

Boileau, M.G., P.D.N. Hebert & S.S. Schwartz, 1992. Nonequilibrium gene frequency divergence: persistent founder effects in natural populations. *Journal of Evolutionary Biology* 5: 25-39.

Braband, A., S. Richter, R. Hiesel & G. Scholtz, 2002. Phylogenetic relationships within the Phyllopoda (Crustacea, Branchiopoda) based on mitochondrial and nuclear markers. *Molecular Phylogenetics and Evolution* 25: 229-244.

Brendonck, L. & B.J. Riddoch, 1999. Wind-borne short-range egg dispersal in anostracans (Crustacea: Branchiopoda). *Biological Journal of the Linnean Society* 67: 87-95.

Brendonck L., L. De Meester & B.J. Riddoch, 2000. Regional structuring of genetic variation in short-lived rock pool populations of *Branchipodopsis wolffi* (Crustacea: Anostraca). *Oecologia* 123: 506–515.

Chambers, G.K. & E.S. MacAvoy, 2000. Microsatellites: consensus and controversy. *Comparative Biochemistry and Physiology Part B* 126: 455-476.

Colautti, R.I., M. Manca, M. Viljanen, H.A. Ketelaars, H. Bürgi, H.J. Macisaac & D.D. Heath, 2005. Invasion genetics of the Eurasian spiny waterflea: evidence for bottlenecks and gene flow using microsatellites. *Molecular Ecology* 14: 1869-1879.

Corujo, M., G. Blanco, E. Vazquez, J.A. Sanchez, 2004. Genetic structure of northwestern Spanish brown trout (*Salmo trutta* L.) populations, differences between microsatellite and allozyme loci. *Hereditas* 141: 258-71.

De Meester, L., A. Gómez, B. Okamura & K. Schweink, 2002. The Monopolization Hypothesis and the dispersal-gene flow paradox in aquatic organisms. *Acta Oecologica* 23: 121-135.

deWaard, J.R., V. Sacherova, M.E.A. Cristescu, E.A. Remigio, T.J. Crease & P.D.N. Hebert, 2006. Probing the relationships of the branchiopod crustaceans. *Molecular Phylogenetics and Evolution* 39: 491-502.

Doyle, J.J. & J.L. Doyle, 1987. A rapid DNA isolation method for small quantities of fresh tissues. *Phytochemical Bulletin* 19: 11-15.

Dumont, H.J. & S.V. Negrea, 2002. Introduction to the Class Branchiopoda. Backhuys Publishers, Leiden.

Ellegren, H., 2000. Microsatellites mutations in the germline: implications for evolutionary inference. *Trends in Genetics* 16: 551-558.

Folmer, O., M. Black, R. Hoeh, R. Lutz, & R. Vrijenhoek, 1994. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology* 3: 294-299.

Figuerola, J. & A.J. Green, 2002. Dispersal of aquatic organisms by waterbirds: a review of past research and priorities for future studies. *Freshwater Biology* 47: 483-494.

Figuerola, J., A.J. Green & T.J. Michot, 2005. Invertebrate eggs can fly: evidence of waterfowl-mediated gene flow in aquatic invertebrates. *American Naturalist*, 165:274-280.

Fryer, G., 1987. A new classification of the branchiopod Crustacea. *Zoological Journal of the Linnean Society* 91: 357-383.

Gómez, A., G.J. Adcock, D.H. Lunt & G. R. Carvalho, 2002. The interplay between colonization history and gene flow in passively dispersing zooplankton: microsatellite analysis of rotifer egg banks. *Journal of Evolutionary Biology* 15: 158-171.

Goudet, J., 2001. FSTAT, a program to estimate and test gene diversities and fixation indices (version 2.9.3). (code available from <http://www.unil.ch/izea/software/fstat.html>)

Hanner, R. & M. Fugate, 1997. Branchiopod phylogenetic reconstruction from 12S rDNA sequence data. *Journal of Crustacean Biology*, 17: 174-183.

Hewitt, G.M., 1996. Some genetic consequences of ice ages, and their role in divergence and speciation. *Biological Journal of the Linnean Society* 58: 247-276.

Hoeh, W.R., N.D. Smallwood, D.M. Senyo, E.G. Chapman & S.C. Weeks, 2006. Evaluating the monophyly of *Eulimnadia* and the Limnadiinae (Branchiopoda: Spinicaudata) using DNA sequences. *Journal of Crustacean Biology* 26:182-192.

King, J.L. & R. Hanner, 1998. Cryptic species in a "living fossil" lineage: taxonomic and phylogenetic relationships within the genus *Lepidurus* (Crustacea: Notostraca) in North America. *Molecular Phylogenetics and Evolution* 10:23-36.

Korn, M., F. Marrone, J.L. Pérez-Bote, M. Machado, M. Cristo, L. Cancela da Fonseca & A.K. Hundsdoerfer, 2006. Sister species within the *Triops cancriformis* lineage (Crustacea, Notostraca). *Zoologica Scripta* 35: 301-322.

Korn, M. & A.K. Hundsdoerfer, 2006. Evidence for cryptic species in the tadpole shrimp *Triops granarius* (Lucas, 1864) (Crustacea: Notostraca). *Zootaxa* 1257: 57-68.

Kumar, S., K. Tamura & M. Nei, 2004. MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Briefings in Bioinformatics* 5: 150-163.

Li, Y., A.B. Korol, T. Fahima, A. Beiles & E. Nevo, 2002. Microsatellites: genomic distribution, putative functions and mutational mechanisms: a review. *Molecular Ecology* 11: 2453-2465.

Limburg, P.A. & L.J. Weider, 2002. 'Ancient' DNA in the resting egg bank of a microcrustacean can serve as a palaeolimnological database. *Proceedings of the Royal Society of London. Series B: Biological Sciences* 269: 281-7.

Luchetti, A., A. Marino, F. Scanabissi & B. Mantovani, 2004. Genomic dynamics of a low copy number satellite DNA family in *Leptestheria dahalacensis* (Crustacea, Branchiopoda, Conchostraca). *Gene* 342: 313-320.

Mantovani, B., M. Cesari & F. Scanabissi, 2004. Molecular taxonomy and phylogeny of the 'living fossil' lineages *Triops* and *Lepidurus* (Branchiopoda: Notostraca). *Zoologica Scripta* 33: 367-374.

Martin, J.W., 1992. Branchiopoda. In: *Microscopy Anatomy of Invertebrates*, Vol. 9: Crustacea, pp. 25-224. Wiley-Liss, Inc.

Murugan, G., A.M. Maeda-Martinez, H. Obregòn-Barboza & N.Y. Hernandez-Saavedra, 2002. Molecular characterization of the tadpole shrimp *Triops* (Branchiopoda: Notostraca) from the Baja California peninsula, México: new insights on species diversity and phylogeny of the genus. *Hydrobiologia* 486: 101-113.

Nei, M., F. Tajima & Y. Tateno, 1983. Accuracy of estimated phylogenetic trees from molecular data. II. Gene frequency data. *Journal of Molecular Evolution* 19:153-170.

Nybom, H., 2004. Comparison of different nuclear DNA markers for estimating intraspecific genetic diversity in plants. *Molecular Ecology* 13: 1143-55.

Paetkau, D., W. Calvert, I. Stirling & C. Strobeck, 1995. Microsatellite analysis of population structure in Canadian polar bears. *Molecular Ecology* 4: 347-354.

Pálsson, S., 2000. Microsatellite variation in *Daphnia pulex* from both sides of the Baltic Sea. *Molecular Ecology* 9:1075-1088.

Piry, S., A. Alapetite, J.M. Cornuet, D. Paetkau, L. Baudouin & A. Estoup, 2004. GeneClass2: A Software for Genetic Assignment and First-Generation Migrant Detection. *Journal of Heredity* 95: 536-539

Pfrender, M.E., K. Spitze & N. Lehman, 2000. Multi-locus genetic evidence for rapid ecologically based speciation in *Daphnia*. *Molecular Ecology* 9:1717-1735.

Raymond, M. & F. Rousset, 1995. GENEPOP (version 1.2): population genetics software for exact tests and ecumenicism. *Journal of Heredity* 86:248-249.

Reece, K.S., W.L. Ribeiro, P.M., Gaffney, R.B., Carnegie & S.K. Allen, Jr., 2004. Microsatellite marker development and analysis in the Eastern Oyster (*Crassostrea virginica*): confirmation of null alleles and non-Mendelian segregation ratios. *Journal of Heredity* 95 (4):346–352.

Rozen, S. & H.J. Skaletsky, 2000. Primer3 on the WWW for general users and for biologist programmers. In Krawetz, S. & S. Misener (eds), *Bioinformatics Methods and Protocols: Methods in Molecular Biology*. Humana Press, Totowa, NJ: 365-386. (Code available at http://www-genome.wi.mit.edu/genome_software/other/primer3.html)

Sambrook, J., E.T. Fritsch & T. Maniatis, 1989. *Molecular cloning. A laboratory manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY .

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu & P. Flook, 1994. Evolution weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annals of the Entomological Society of America* 87: 651-701.

Stenderup, J.T., J. Olesen & H. Glenner, 2006. Molecular phylogeny of the Branchiopoda (Crustacea)—Multiple approaches suggest a ‘diplostracan’ ancestry of the Notostraca. *Molecular Phylogenetics and Evolution* 41: 182-194.

Thiery, A. & D. Pont, 1987. *Eoleptestheria ticinensis* (Balsamo-Crivelli, 1859) Conchostracé nouveau pour la France (Crustacea, Branchiopoda, Conchostraca). Vie et Milieu 37: 115-121.

Tinti, F. & F. Scanabissi, 1996. Reproduction and genetic variation in clam shrimps (Crustacea, Branchiopoda, Conchostraca). Canadian Journal of Zoology 74: 824-832.

Van Oosterhout, C., W.F. Hutchinson, D.P.M. Wills & P. Shipley, 2004. Micro-Checker: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes 4: 535-538.

Weeks, S.C. & RJ Duff, 2002. A genetic comparison of different populations of clam shrimp in the genus *Eulimnadia*. Hydrobiologia 486: 295-302.

Wirth, T., L. Bernatchez, 2001. Genetic evidence against panmixia in the European eel. Nature 409: 1037-1040.

Wright, S., 1969. Evolution and the genetics of populations. Vol. 2: The theory of gene frequencies. University of Chicago Press, Chicago. 511 pp.

Zane, L., L. Bargelloni & T. Patarnello, 2002. Strategies for microsatellite isolation: a review. Molecular Ecology 11: 1-16.

Table 1. Sampling information, scored haplotypes for mitochondrial genes and mean number of analyzed individuals for MSL. Within specific localities, capital letters indicate different years (Leona) or different ponds (Isola della Scala).

Locality	Collecting site	Year	Individual	Mitochondrial Analysis				Nuclear Analysis
				Haplotype				Mean sample size
				12S	16S	COI	mt-type	
<i>Italy</i>								
Ferrara	Contane	2000	Contane 2000-1	a	a	a	A	16.3
		2000	Contane 2000-2	b	b	a	B	
	Mezzogoro	2002	Mezzogoro 2002-1	b	b	a	B	20.7
		2002	Mezzogoro 2002-2	b	b	a	B	
	Amiani	2002	Amiani 2002-1	a	b	a	C	11.7
		2002	Amiani 2002-2	b	b	a	B	
	Leona A	2002	Leona A 2002-1	b	c	a	D	20.3
		2002	Leona A 2002-2	a	c	a	E	
	Leona B	2003	Leona B 2003-1	b	c	a	D	24.0
		2003	Leona B 2003-2	a	c	a	E	
	Gran Linea	2003	Gran Linea 2003-1	b	b	a	B	12.0
		2003	Gran Linea 2003-2	b	c	a	D	
Verona	Isola della Scala A	2006	I.Scala A 2006-1	b	c	a	D	11.3
		2006	I.Scala A 2006-2	b	c	a	D	
	Isola della Scala B	2006	I.Scala B 2006-1	b	c	a	D	10.7
		2006	I.Scala B 2006-2	b	c	a	D	
<i>Austria</i>								
Vienna	Marchegg	2001	Marchegg 2001-1	b	d	a	F	23.7
		2001	Marchegg 2001-2	b	d	a	F	
<i>Germany</i>								
Hessen	Hadamar	2001	Hadamar 2001-1	b	d	a	F	9.0
		2001	Hadamar 2001-2	b	d	a	F	

Table 2. General features of the three microsatellite loci (MSL) identified in *L. dahalacensis*. (*: labelled primers; A: number of alleles).

Locus	Primer Sequences (5'-3')	[MgCl₂]	Motif	A	Product Size Range	Genbank A.N.
ldAC-10	F*: ACGCGCTATCTGTTAGGAAT R: TCGTTTGTGTCTTTTGTATTTC	1 mM	(AC) ₁₀	3	149-157 bp	AY765352
ldAC-11	F*: AAGATGTCCGCCTTTTCCT R: AAGGACAGGGGTGATGACTG	1.25 mM	(AC) ₁₁	5	199-211 bp	AY765350
ldAC-16	F*: TCCGACATCGTTTTCTTCC R: CAAGTGCAAGGTTTGGGAGT	1.25 mM	(AC) ₁₆	4	160-174 bp	AY765351

Table 3. Allelic richness (A_C), observed (H_O) / expected (H_E) heterozygosity per locus and population computed for the three MSL, and F-statistics among populations (F_{IT}), between populations (F_{ST}) and within population (F_{IS}) (N: number of analysed individuals; §: possible presence of null alleles; * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$).

Locus		Contane 2000	Mezzogoro 2002	Amiani 2002	Leona A 2002	Leona B 2003	Gran Linea 2003	I.Scala A 2006	I.Scala B 2006	Marchegg 2001	Hadamar 2001	F_{IT}	F_{ST}	F_{IS}
IdAC-10	A_C	1.994	1.996	1.981	1.999	1.988	2.000	1.996	2.000	2.000	2.000			
	H_O / H_E	0.563 / 0.404	0.421 / 0.432	0.417 / 0.330	0.556 / 0.475	0.458 / 0.395	0.417 / 0.497	0.500 / 0.375	0.700 / 0.455	0.583 / 0.486	0.167 / 0.153	-0.128	0.017 *	-0.148
	N	16	19	12	18	24	12	10	10	24	6			
IdAC-11	A_C	2.998 §	3.208 §	1.990	2.494 §	2.223	1.993	1.000	2.000 §	2.718 §	2.921			
	H_O / H_E	0.063 / 0.315 ***	0.409 / 0.600 **	0.273 / 0.351	0.333 / 0.534	0.333 / 0.385	0.500 / 0.375	0.000 / 0.000	0.083 / 0.156 *	0.217 / 0.542 ***	0.818 / 0.583	0.373 ***	0.097 ***	0.306 ***
	N	16	22	11	21	24	12	12	12	23	11			
IdAC-16	A_C	2.000	2.490	1.981	1.872	2.436	1.998	1.953	1.996	2.980	2.985			
	H_O / H_E	0.765 / 0.493	0.524 / 0.489 *	0.417 / 0.330	0.273 / 0.236	0.542 / 0.484 *	0.417 / 0.413	0.333 / 0.278	0.300 / 0.375	0.792 / 0.656	0.800 / 0.635	-0.009	0.125 ***	-0.153
	N	17	21	12	22	24	12	12	10	24	10			
Over all loci	H_O / H_E	0.463 / 0.404	0.451 / 0.507	0.369 / 0.337	0.387 / 0.415	0.444 / 0.421	0.444 / 0.428	0.278 / 0.218	0.361 / 0.329	0.531 / 0.561	0.595 / 0.457	0.081 *	0.082 ***	-0.002

Table 4. Pairwise F_{ST} values and their significance (below the diagonal) and pairwise estimated number of migrants ($N_m=(1-F_{ST})/(4*F_{ST})$; Wright, 1969; above the diagonal) between populations (* $P<0.05$; ** $P<0.01$; *** $P<0.001$).

	Contane 2000	Mezzogoro 2002	Amiani 2002	Leona A 2002	Leona B 2003	Gran Linea 2003	I.Scala A 2006	I.Scala B 2006	Marchegg 2001	Hadamar 2001
Contane 2000		9.95	12.26	1.38	∞	18.94	3.38	21.59	2.77	1.38
Mezzogoro 2002	0.025		16.53	9.43	69.56	72.74	1.90	4.25	4.91	1.66
Amiani 2002	0.020	0.015		2.60	∞	14.63	9.57	41.19	2.06	1.02
Leona A 2002	0.154**	0.026	0.088**		2.35	3.71	0.86	1.29	2.00	0.73
Leona B 2003	-0.010	0.004	-0.014	0.096**		∞	5.02	40.13	2.96	1.37
Gran Linea 2003	0.013	0.003	0.017	0.063*	-0.001		3.18	∞	3.96	0.95
I.Scala A 2006	0.069*	0.116**	0.026	0.226***	0.047*	0.073		∞	1.06	0.54
I.Scala B 2006	0.012	0.056*	0.006	0.163**	0.006	-0.003	-0.015		1.81	0.75
Marchegg 2001	0.083*	0.048*	0.108**	0.111**	0.078**	0.060*	0.190***	0.122**		3.73
Hadamar 2001	0.153***	0.131**	0.196***	0.256***	0.154***	0.208**	0.318***	0.251*	0.063*	

Figure captions

Figure 1. Geographic location of the studied populations. In the box, the sampled rice pools of the Ferrara region are displayed as black squares.

Figure 2. Allelic frequencies at the three MSL in the presently analysed populations.

Figure 3. Unrooted Neighbor-Joining dendrogram computed on Nei et al. (1983) distances derived from the three MSL; numbers above branches indicate bootstrap values.

Figure 4. Isolation by distance analysis: correlation between pairwise population differentiations and geographic distances ($R^2 = 0.541$; $P < 0.05$).

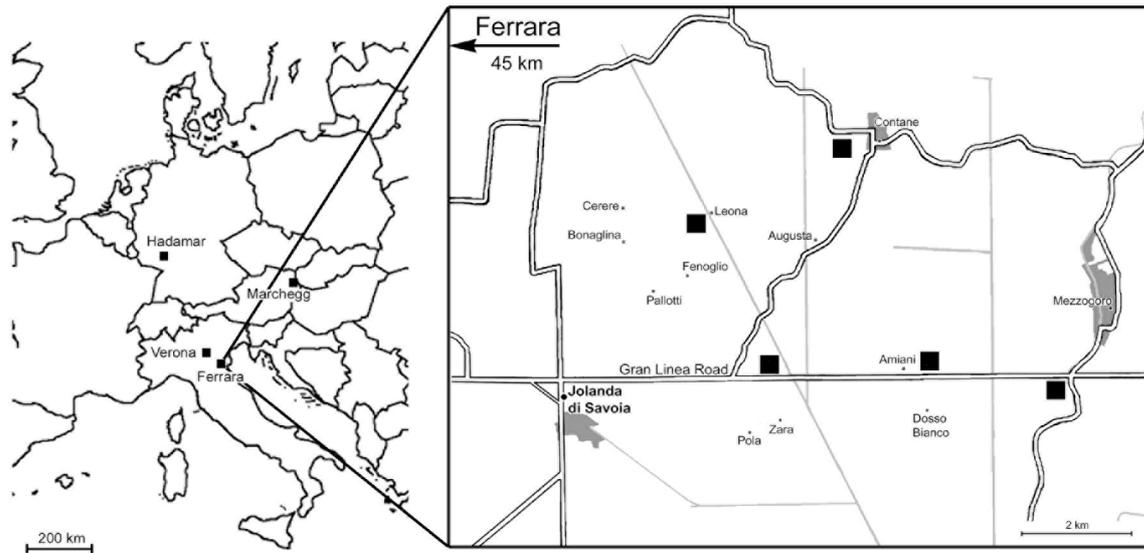


Figure 1

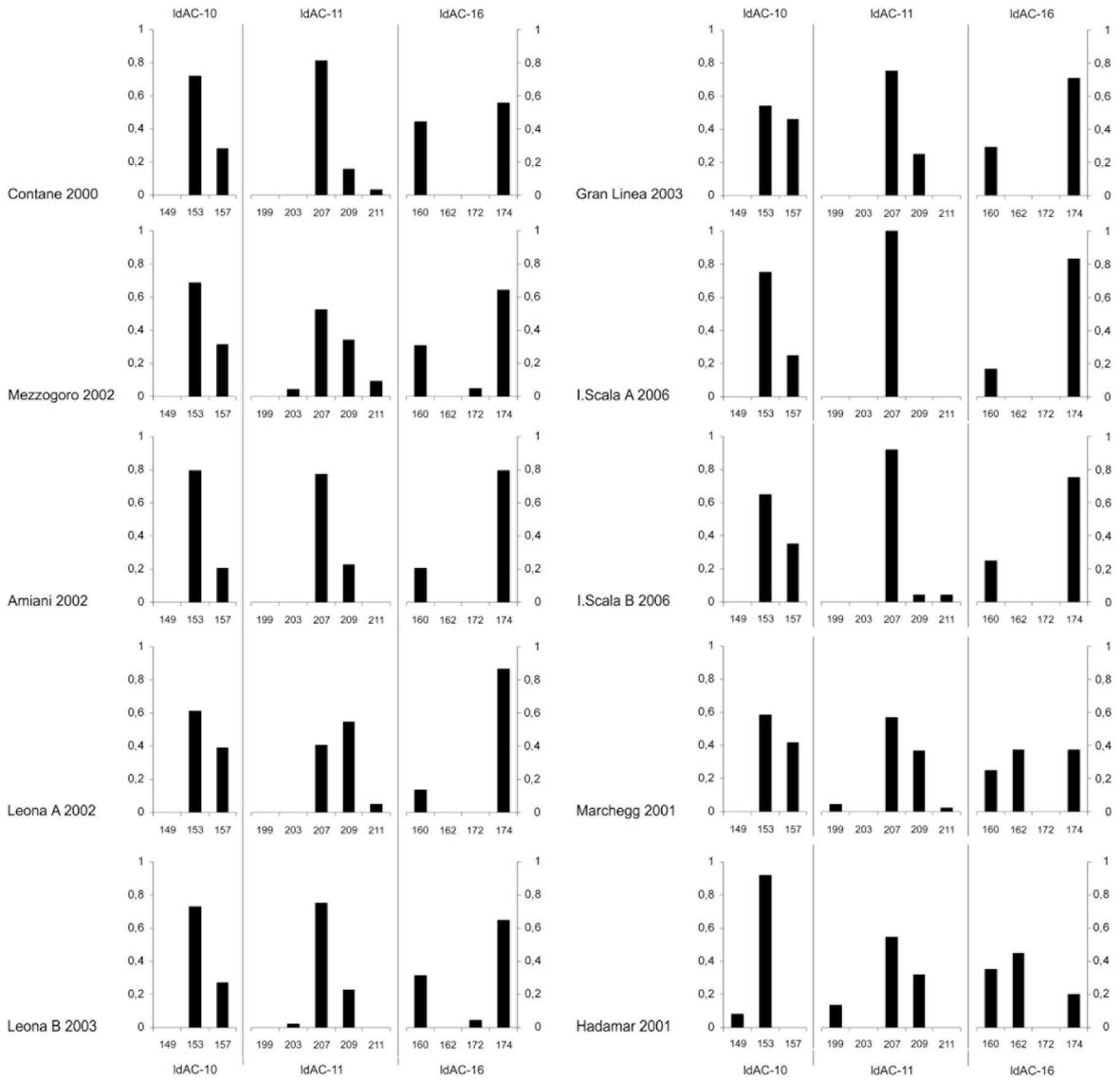


Figure 2

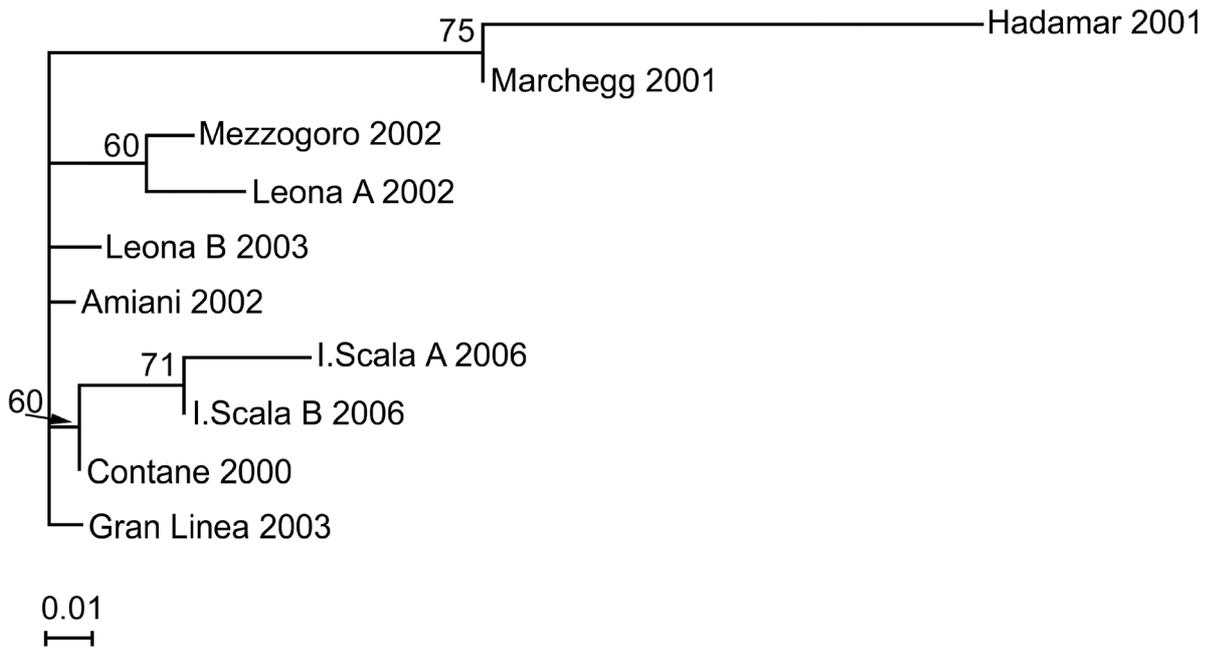


Figure 3

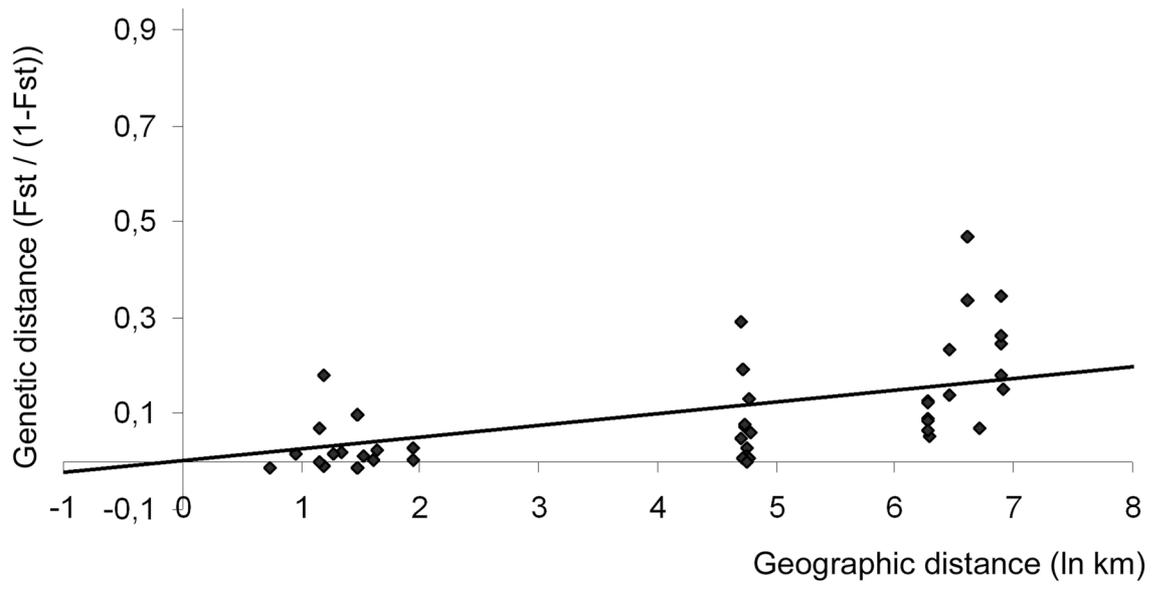


Figure 4

Genomic dynamics of a low-copy-number satellite DNA family in *Leptestheria dahalacensis* (Crustacea, Branchiopoda, Conchostraca)

Andrea Luchetti, Alberto Marino, Franca Scanabissi, Barbara Mantovani*

Dipartimento di Biologia Evoluzionistica Sperimentale, Università di Bologna, Via Selmi 3, Bologna 40126, Italy

Received 21 June 2004; received in revised form 6 August 2004; accepted 19 August 2004

Available online 1 October 2004

Received by Takashi Gojobori

Abstract

The LEP150 satellite DNA (satDNA) family found in *Leptestheria dahalacensis* (Rüppel, 1837) (Conchostraca) is a low-copy-number satellite with a canonical monomer of 150 bp. Nucleotide variation analyses suggest a 14-bp palindromic region as a possible protein binding site with constraints acting on the whole sequence but a 25-bp variable box. Besides the head-to-tail arrangement of 150 bp monomers, multimers analyses evidenced incomplete monomers, one duplication event, and three inversions. Both observed rearrangements and the higher values of sequence variability scored suggest that rearranged monomers reside in regions with a lower degree of homogenisation efficiency. Sixty-seven percent of the breakpoints occurs at kinkable dinucleotides, thus supporting their role in rearrangements as documented in alphoid satDNA recombination events. Monomers of different lengths may result from crossing over between repeats misaligned through the direct and inverted subrepeats of LEP150 monomers. ANOVA results indicate that the same range of sequence diversity is experienced at the individual and population ranks; therefore, the evolution of the *L. dahalacensis* satDNA is concerted.

© 2004 Elsevier B.V. All rights reserved.

Keywords: DNA dynamics; Concerted evolution; Highly repeated sequences; Genomic organisation; Selective pressure; Variable domains

1. Introduction

Satellite DNA (satDNA) represents a substantial fraction of eukaryotic genomes and is composed of highly, tandemly repeated sequences organised in large heterochromatic clusters, usually located in pericentromeric and/or telomeric regions (Charlesworth et al., 1994). Repeated sequences are known to evolve through a pattern of concerted evolution, resulting from variant homogenisation within genomes and variant fixation in different lineages (either strains, populations, subspecies, or species) through a process known as molecular drive (Dover, 2002). Molecular drive results from genomic turnover mechanisms, involving nonreciprocal DNA exchanges within and between chromosomes and

population dynamic processes. Beside nucleotidic sequence homogenisation, these processes can also cause fluctuations in repeat length and copy number (Ugarkovic and Plohl, 2002). The extreme consequences of concerted evolution may lead to chromosome-specific repeat variants, as observed in the *Pimelia* genome (Pons et al., 2002) or in the human α -satDNA (Willard and Wayne, 1987).

Genomic organisation of monomers has a strong impact on their dynamics. For example, contiguous monomers show a higher degree of similarity than monomers randomly present in different clusters (Durfy and Willard, 1989; Schindelbauer and Schwarz, 2002). On the other hand, a lower degree of similarity between repeats found at the border of satellite clusters has been reported in both theoretical and empirical studies (Smith, 1976; Mashkova et al., 1998, 2001; Bassi et al., 2000). The occurrence of large deletions, duplications, expansions, and inversions, as well as non-satellite sequence insertions, has been demonstrated in bordering regions. This is possibly due to the accumulation of several products of unequal crossing over and trans-

Abbreviations: bp, base pairs; satDNA, satellite DNA; ANOVA, analysis of variance; MUM, multimer; LSD, least significant differences; MP, maximum parsimony; RF, repeat fragment; S.D., standard deviation; S.E., standard error.

* Corresponding author. Tel.: +39 51 209 4169; fax: +39 51 251208.

E-mail address: barman@alma.unibo.it (B. Mantovani).

position events. Less efficient homogenisation mechanisms may explain the inability to eliminate such noncanonical variants (Mashkova et al., 1998, 2001).

Although the evolutionary pattern of satDNA has been extensively analysed, its biological significance remains unclear. Many studies indicate that it may serve to form higher-order structures as suggested by its intrinsic ability to bend (Bigot et al., 1990; Ugarkovic et al., 1992; Fitzgerald et al., 1994; Ugarkovic and Plohl, 2002). Such structural features may be relevant for the correct and tight packaging of DNA proteins complex in heterochromatic domains, and could be subject to selective pressure (Plohl et al., 1998). Constraints acting on satDNA, however, have been demonstrated only in human α satellite and *Arabidopsis* centromeric sequences (Romanova et al., 1996; Hall et al., 2003). In those cases, protein recognition sites and/or structural sequence features have been considered as constrained elements.

Most data on satDNA found in the literature mainly refer to mammals and insects. Instead, satDNA studies in Crustacea are a limited number (Varadaraj and Skinner, 1994; Bagshaw and Buckholt, 1997) and especially in Branchiopods. Indeed, analyses are available only for *Artemia* spp. (Anostraca; Maiorano et al., 1997; Motta et al., 1998). From an evolutionary point of view, Notostracan and Conchostracan branchiopoda are of particular interest because of their ancient origin and reproductive biology variations. Here, we present data relative to a satDNA family found in the gonochoric *Leptestheria dahalacensis* as a starting point for a deeper knowledge of this peculiar genomic compartment in Conchostraca.

2. Materials and methods

2.1. DNA analysis, cloning, and sequencing

Total DNA was extracted from single alcohol preserved or frozen individuals (field-caught or taken from laboratory cultures) of *L. dahalacensis*, following the CTAB method (Del Sal et al., 1989). Sample information is given in Table 1.

DNA samples of the Jolanda di Savoia population were digested with different restriction enzymes (*AccI*, *AluI*, *AvaI*, *BamHI*, *BglII*, *CfoI*, *ClaI*, *DraI*, *EcoRI*, *EcoRII*, *HaeIII*, *HincII*, *HindIII*, *HpaI*, *HpaII*, *MaeI*, *MaeII*, *MseI*, *MboI*,

MspI, *NdeI*, *NsiI*, *PstI*, *PvuI*, *RsaI*, *Sall*, *SauIII*, *ScrFI*, *SecI*, *SmaI*, *SmlI*, *SstI*, *TaqI*, and *XhoI*). Agarose gel electrophoresis was used to check for the presence of multimeric bands. Only *AluI*-restricted genomic DNA of one female (fJS) showed a ladder of faint bands. The monomer band of approximately 150 bp length was cloned into pGEM7zf(+) vector (Amersham Pharmacia Biotechnology) and transformed in *Escherichia coli* DH5 α -competent cells. Recombinant colonies were screened for blue-white color (Sambrook et al., 1989) and plasmids were sequenced with the Dye terminator cycle sequencing kit (Applied Biosystems) in a 310 Genetic Analyzer (ABI) automatic sequencer.

To avoid *AluI* restriction often producing poorly resolved bands or no bands at all, the following primers were designed on the consensus sequence of restriction-isolated fJS repeat units: dimF (5' -CGCCAGAATCCCARATARTC-3') and dimR (5' -TYGAGATTCCTGGGTTTRTTD-3').

All products of the polymerase chain reaction (PCR) were ligated in a pGEM T-Easy vector (Promega), and recombinant clones were identified and sequenced as described above.

This cloning procedure may produce ligation artifacts. However, these can be discriminated on the basis of designed primers that anneal within the monomeric unit: multimers produced by amplicon fusions thereby present duplications and deletions at the 5' and 3' ends of the primer annealing sites. Furthermore, in case of partial PCR amplification, incomplete repeats embodied in multimers should show tracts congruent with primer sequences.

Nucleotide sequences were aligned with the CLUSTAL algorithm, included in Sequence Navigator program v. 1.1 (Applied Biosystems).

Presently analysed sequences were entered, as single monomer or as multimers (MUMs), in GenBank with accession nos. AY437561–AY437608.

2.2. Southern blot and dot blot analysis

Southern blot was carried according to standard procedures (Sambrook et al., 1989). Cloned satellite monomer (fJS16) excised from plasmid was used as a hybridisation probe. Labelling and detection followed the instructions provided with the DIG DNA Labelling and Detection Kit (Roche).

The contribution of LEP150 satDNA to the *L. dahalacensis* genome was defined by dot blot. The genomic DNAs of four specimens (two females and two males) were transferred onto a Hybond-N⁺ filter (Amersham) with a Bio-Dot Apparatus (Bio-Rad), in a series of dilutions ranging from 15 ng to 2 μ g.

Cloned satellite dimers (mGL5, mGL6, fGL3, mAUI, fAU1, mAM4, and fAM1) excised from plasmids and labelled with α -³²P with Megaprime kit (Amersham) were used as hybridisation probes. The same satellite dimers, dot-blotted in a range between 0.02 and 2 ng, were used as a calibration curve. Prehybridisation and hybridisation were

Table 1

Specimen list, acronyms, collection sites, number of complete monomers obtained, mean *p* distance within each individual \pm S.E.

Specimen	Collection site	Acronym	<i>N</i>	Mean <i>p</i> distance	S.E.
Male	Gran Linea (Italy)	mGL	12	0.093	0.014
Female		fGL	11	0.097	0.014
Male	Amiani (Italy)	mAM	9	0.081	0.014
Female		fAM	8	0.080	0.013
Male	Vienna (Austria)	mAU	10	0.086	0.013
Female		fAU	6	0.121	0.016
Female	Jolanda di Savoia (Italy)	fJS	16	0.121	0.015

performed at 65° with a 7% sodium dodecyl sulfate (SDS)/0.5 M Na₃PO₄ (pH 7.0) solution; after hybridisation, the filter was washed in a 40 mM Na₃PO₄ (pH 7.0)/1 mM EDTA (pH 8.0)/1% SDS solution at the same temperature; these conditions allow 80% homology. Densitometric quantification from autoradiographic film was performed with ImageJ 1.25t software.

2.3. Structural analyses

Detection of internal repeats was performed on the consensus sequence through the server program OligoRep (available at <http://www.mgs.bionet.nsc.ru/mgs/>) with the following parameters: minimum repeat length=7; maximum repeat length=20; and maximum mismatch number=0. Curvature evaluations were performed through the DNA Curvature Analysis server (available at <http://www.lfd.uiuc.edu/staff/gohlke/curve/>) and according to the dinucleotide wedge model by Bolshoy et al. (1991).

2.4. Statistical analyses

Statistical analyses were performed only on entire monomers. Neighbor-joining and maximum parsimony (MP) dendrograms were computed using PAUP* v. 4.0b8a (Swofford, 2001), with gaps treated as missing data and 2000 bootstrap replicates.

To analyse LEP150 variability, *p* distances were calculated with MEGA 2.1 package (Kumar et al., 2001) and analysed by means of one-way ANOVA, plus comparisons between means based on least significant differences (LSD; Luchetti et al., 2003; Cesari et al., 2003). *Z* test was used to verify if monomers belonging the same MUM were significantly less variable than the average.

Sequence variation across satellite repeats was investigated as described in Hall et al. (2003): the occurrence of the most frequent base in each nucleotide position was calculated, and either conserved or variable region was defined by overlapping sliding windows analysis of the percent occurrence data. Three different windows were analysed (5, 10, and 15 bp), and since no substantial differences were observed, the 15-bp window is presented. A significant deviation of the window from the average was defined by a *Z* score with two levels of significance (± 1.2 S.D. from the mean as in Hall et al., 2003 and ± 1.96 S.D. from the mean; error probability <5%).

3. Results

3.1. LEP150 sequences

Southern blot analyses of *AluI*-restricted genomic DNA of the fJS sample showed an evident ladder of bands typical of a highly repeated DNA, with a monomeric unit of about 150 bp (Fig. 1). This family was therefore indicated as



Fig. 1. Southern blot of *AluI*-digested genomic DNA from fJS sample.

LEP150. Besides bands corresponding to monomers, dimers, and trimers, intermediate bands are also visible; these might correspond to multimers with incomplete repeats (see Section 3.3).

Densitometric measurements in dot blot analyses (not shown) indicate that LEP150 sequences represent a decidedly low-copy-number satellite constituting only 0.4–0.5% of the *L. dahalacensis* genome. This well explains the unreliability of *AluI* restrictions, usually producing only a faint ladder, but often no ladder at all. Moreover, it explains the absence of bands with other enzymes that, as shown subsequently on the consensus sequence, have a restriction site in the monomer (i.e., *AccI*, *AvaI*, *EcoRII*, *MaeI*, *MaeII*, *MseI*, *ScrFI*, *SecI*, *SmlI*, and *XhoI*; not shown).

Genomic DNA of other *L. dahalacensis* samples (Table 1) was therefore PCR-amplified. Amplicons revealed a well-defined ladder of bands in all analysed specimens, with a monomeric unit of 150 bp, consisting of two half monomers owing to the primer internal annealing sites. To avoid the generation of artificial monomers, we choose to sequence only those clones with a minimum length equal to 300 bp (one monomer flanked by two half monomers). Dimers, trimers, and one pentamer, together with intermediate products, were obtained. We carefully checked all sequenced clones: only one multimer seemed to be the result of a ligation artifact, and then was excluded from the analysis. On the whole, 72 complete monomers were obtained from clones containing either a single monomer, or more than one repeat as multimers (MUMs). The number of sequenced MUMs ranges from two in fAU to four in mGL, fAM, and mAU. Further, three multimers obtained from mGL, mAM, and fAU specimens were found to contain incomplete repeats.

Analysed sequences were labelled with the specimen acronym plus an Arabic number to distinguish different single monomers obtained from the same individual. MUM-derived monomers are further identified by a lower-case

letter (e.g., mAU1a and mAU1b are two monomers obtained from the same MUM found in the mAU specimen).

Monomer mean sequence length is $149.73 \text{ bp} \pm 1.18 \text{ S.D.}$, with a mean A+T content equal to $60.78\% \pm 1.18 \text{ S.D.}$ BLAST search does not give significant similarity with other sequences in the GenBank database; neither were particular subdomains shared with other satDNAs recognised.

In the MP analysis, the MAXTREE limit was reached at 10,600 equally parsimonious trees, with length equal to 228 steps. The MP bootstrap consensus tree (Fig. 2) results in an overall polytomy where well-differentiated sequences intermingle despite geographic, individual, sex, and MUMs origins. The majority of clusters appears weakly supported by bootstrap values. The only supported clustering is given by six fJS units and fAU2a monomer: in this regard, the different techniques used for sequences isolation (i.e., restriction procedure vs. PCR amplification) could explain the higher variability of fJS monomers with respect to the others. However, doubts are cast on this possibility given the presence of a PCR-isolated sequence (fAU2a) in this cluster. Furthermore, data in progress suggest that these six fJS monomers may constitute the flanking regions of the 5S DNA genes (Luchetti et al., in preparation). Neighbor-joining dendrogram (not shown) completely agrees with the MP one.

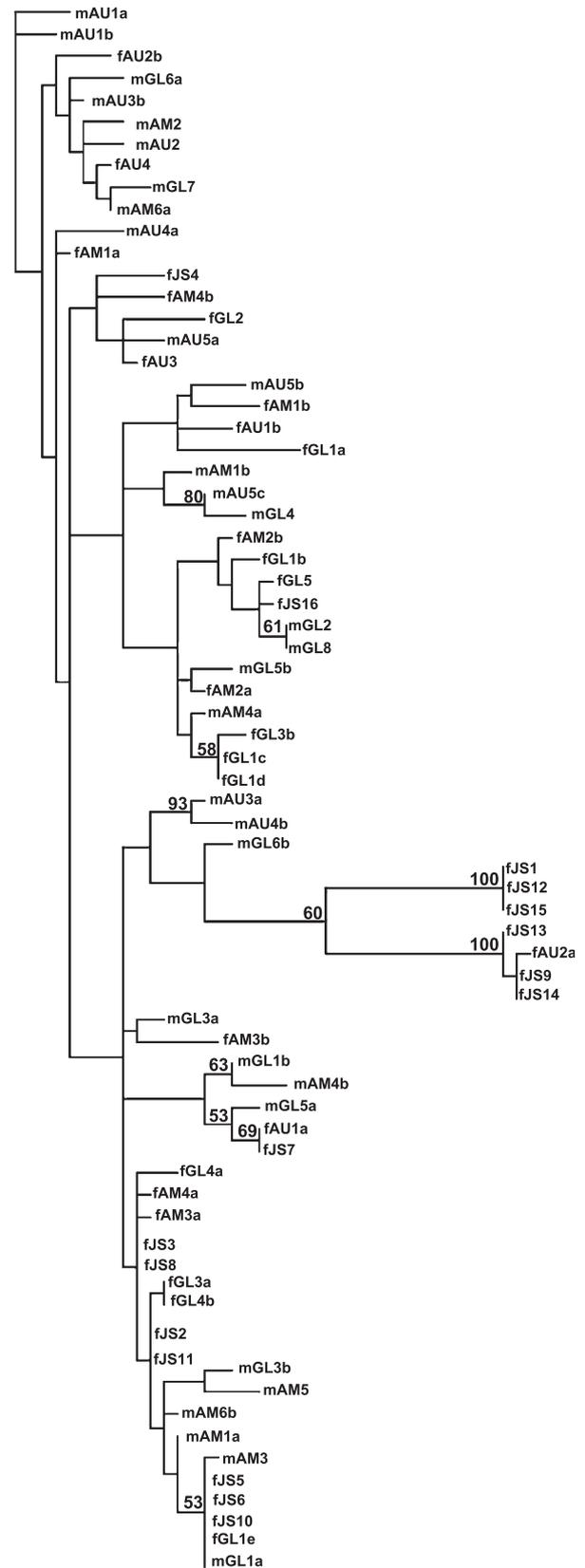
3.2. Internal sequence organisation and structural analysis

OligoRep program run on the consensus sequence shows two short direct repeats (7 and 8 bp) and one inverted repeat (8 bp). However, such subrepeats are not conserved in all analysed monomers and the 8-bp inverted repeat partially overlaps with the variable domain (see Section 3.5). A well-conserved 14-bp palindrome tract was found starting from base 109, partially overlapping with the 7-bp direct repeat. On the whole, the absence of conserved subelements does not allow to infer any within-monomer evolutionary pattern.

Curvature evaluation (available from the authors) shows that LEP150 sequences are bent, with maximum peaks corresponding to areas composed mainly of A/T stretches 3–4 bp long; these can be mainly observed in the first 100 bp of the repeats, but they never occur as consecutive. Bending periodicity is maintained in all analysed MUMs.

3.3. Genomic organisation

Repeat length is 150 bp in 90.2% of the sequenced monomers. In the remaining 9.8% of monomers, length ranges from 149 to 141 bp. Besides these complete monomers, in three MUMs, we found some repeat fragments (RFs) ranging from 7 to 115 bp (Fig. 3). In one instance, we have also found two consecutive RFs partially overlapping in a 10-bp region located at the 3' end of the first RF and at the 5' end of the second (Fig. 3c).



— 1

Fig. 2. MP bootstrap consensus tree based on all available complete repeats (TL: 240; CI=0.375; RI=0.771). Numbers above the branches are bootstrap values (>50%) obtained from 2000 replicates.

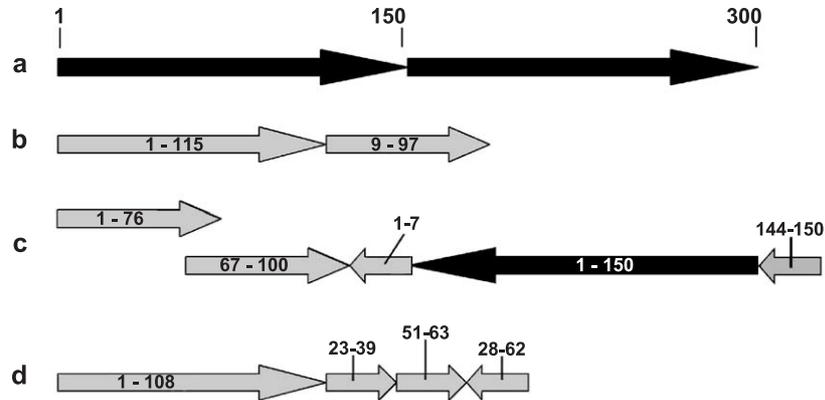


Fig. 3. Schematic representation of clones showing repeat fragments and rearrangements. (a) Consensus LEP150 dimer; (b) clone mAM2; (c) clone mGL4; and (d) clone fAU3. Arrows indicates the 5' → 3' direction. Complete monomers are in black; numbers indicate nucleotide positions of fragment ends with respect to consensus sequence.

LEP150 monomers are mainly arranged in a head-to-tail fashion, but in two MUMs we observed a tail-to-tail orientation (Fig. 3c and d); these rearrangements are always associated with RFs.

A similar scenario has been described for alphoid sequences located at the edge of satellite domain on human chromosome 21 (Mashkova et al. 1998, 2001). These repeats result less homogenised than the centromeric ones and show various structural rearrangements, with the great majority of the rearrangement breakpoints nonrandomly corresponding to kinkable dinucleotides (i.e., TG, TA, and CA). By analysing dinucleotide distribution at the edge of each RF and their adjacent positions reconstructed with the consensus sequence, we found that nearly 67% of rearrangement breakpoints scored in LEP150 satDNA occurs at kinkable dinucleotides (Table 2). The rearrangement breakpoints never occur in the variable domain found in this satDNA (see below).

Table 2
Analysis of breakpoints in rearranged clones

Clone	Breakpoints ^a	Sequence
mGL4	7	TGGT-tt tc
	67	at ta -GATT
	76	CG TT - aaaa
	100	TTAA- caac
mAM2	9	g gtt - AACT
	115	TCTC-gaga
fAU3	23	taga-C GTT
	28	g tt -CTCT
	39	A ACA - ttga
	51	tt t - AGTC
	63	T CTA - ttag
	108	CCAG-ga t

^a Numbers indicate nucleotide positions of each breakpoint; lower-case sequences are reconstructed from consensus monomer. Kinkable dinucleotides are in bold italics.

3.4. Sequence diversity among genomes, individuals, and populations

A variability analysis was carried out on all entire monomers to assess the degree and trends of diversity in *L. dahalacensis*. Mean *p* distance value within individuals ranges from 0.080±0.013 (fAM) to 0.121±0.016 (fAU) (Table 1), for an overall value of 0.103±0.013.

Mean sequence divergence of LEP150 RFs is equal to 0.199±0.027; RF variability is significantly higher with respect to the one scored for entire monomers (ANOVA, *P*<0.001).

Table 3
Z test probability (*P*) indicating if monomers belonging to the same MUM diverge significantly with respect to overall mean *p* distance calculated for the pertaining individuals

MUM acronyms	MUM mean <i>p</i> distance	Individual mean <i>p</i> distance	<i>P</i>
mGL1	0.093	0.093	N.S.
mGL3	0.060	0.093	***
mGL5	0.133	0.093	N.S.
mGL6	0.085	0.093	*
fGL1	0.097	0.097	N.S.
fGL3	0.113	0.097	N.S.
fGL4	0.034	0.097	***
mAM1	0.113	0.081	N.S.
mAM4	0.153	0.081	N.S.
mAM6	0.067	0.081	**
fAM1	0.080	0.080	N.S.
fAM2	0.040	0.080	***
fAM3	0.060	0.080	***
fAM4	0.067	0.080	*
mAU1	0.047	0.086	***
mAU3	0.073	0.086	***
mAU4	0.100	0.086	N.S.
mAU5	0.090	0.086	N.S.
fAU1	0.133	0.121	N.S.
fAU2	0.193	0.121	N.S.

* 0.05>*p*<0.01.
** 0.01>*p*<0.001.
*** *p*<0.001.

Table 4
Mean p distances (below the diagonal) and S.E. (above diagonal) calculated between individuals

	mGL	fGL	mAM	fAM	mAU	fAU	fJS
mGL		0.015	0.013	0.013	0.013	0.013	0.013
fGL	0.096		0.014	0.014	0.015	0.015	0.014
mAM	0.085	0.091		0.013	0.013	0.013	0.013
fAM	0.085	0.086	0.077		0.013	0.013	0.013
mAU	0.093	0.103	0.086	0.084		0.013	0.014
fAU	0.107	0.115	0.102	0.100	0.099		0.014
fJS	0.117	0.123	0.108	0.109	0.121	0.126	

Owing to the peculiar nature of the mechanisms underlying molecular drive, it is known that repeat units belonging to the same array are homogenised more efficiently than those that are part of different arrays (Durfy and Willard, 1989; Schindelbauer and Schwarz, 2002). To verify this condition, we first computed the mean p distance between sequences of the same MUM; then we checked, through a Z test, if this variability significantly deviates from the overall mean p distance calculated for the individual. In many instances, MUMs present a significantly lower degree of diversity than the mean (Table 3).

When all available sequences for each specimen are considered, ANOVA analysis evidences that individuals embody significantly different variability values ($P < 0.001$). On the contrary, mean sequence divergences between individuals (Table 4) do not appear significantly different, with the only exception of the comparison involving fJS and mAU (LSD, $P < 0.05$).

A comparable picture is evident when the population level is considered: taking into account all samples but fJS, intrapopulation diversity values are significantly different (ANOVA, $0.05 > P > 0.01$), while interpopulation mean sequence divergences are of a comparable magnitude.

3.5. Nucleotide variation across LEP150 satellite repeats

LEP150 sequence variation has been also verified, disregarding their origin (i.e., MUMs, individuals, and

populations) and from a functional point of view with an analysis similar to the one conducted for *Arabidopsis* and *Homo* centromeric repeats (Hall et al., 2003). The percentage of occurrence of the most frequent base for each nucleotide position was taken as a variability measure and plotted against nucleotide position (Fig. 4a). This satellite family appears quite conserved: about the 44.6% of all nucleotides occurs with a frequency of 100%, and the 43.7% of the remaining nucleotides resides within 1 S.D. from the average of $92.9 \pm 12.2\%$. On the other side, only 6.6% of them are highly polymorphic, with frequency values below -2 S.D. from the mean.

A variable region was then identified through the sliding window method (Fig. 4b). When 1.2 S.D. significance level was considered, a large area spanning from base 73 to 98 shows significantly higher values of variation (shaded in grey in Fig. 4a); 4 of 10 of the highly polymorphic sites are found in this region. When the 1.96 S.D. level of significance was taken into account, the variable domain became restricted to bases 84–87 (Fig. 4b); however, some flanking windows show Z scores being distributed at the limit of significance ($-1.89 > Z$ score > -2.04 ; marked by asterisks in Fig. 4b). On the whole, both levels of significance indicate a substantial accumulation of nucleotide variation in a ~ 25 -bp area centred on site 85.

4. Discussion

Several studies on satDNAs have demonstrated the need for a wide approach to characterise families of repeated sequences. Here the low-copy-number satDNA LEP150 from *L. dahalacensis* is analysed from both structural and variability points of view.

Nucleotide variation analysis shows that the LEP150 family, which is pericentromerically located (O. Marescalchi, personal communication), is highly conserved. In all analyzed monomers, only a 25-bp box was found to be significantly more variable with respect to the whole

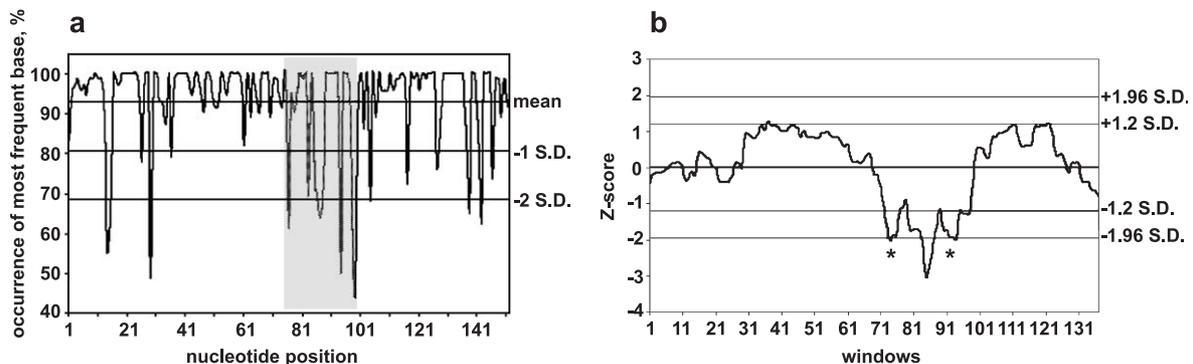


Fig. 4. Nucleotide variation across LEP150 satellite repeats. (a) Percentage of occurrence of the most frequent base for each nucleotide position is plotted against nucleotide positions. The grey shaded area indicates variable domain. (b) Plotting of Z scores measured over a 15-bp sliding window. Values over/under S.D. are significantly conserved/variable at the considered significance level (± 1.2 S.D.; ± 1.96 S.D.). Asterisks mark window with Z scores distributed at the limit of significance (see Section 3.5).

sequence. In *Arabidopsis* centromeric repeats and *Homo* α -satDNA, the same kind of analysis revealed the presence of both conserved and variable domains (Hall et al., 2003), raising the possibility of differential selective pressure to maintain a particular DNA sequence. As far as LEP150 satDNA is concerned, no significantly conserved domain was found; however, considering the low level of sequence variation, some constraints acting on the whole sequence but the 25-bp variable box can be hypothesised. This could be firstly explained through the possible interaction with specific proteins. In human α satellite, CENP-B box is a 17-bp binding site for the CENP-B protein; it contains nine fundamental nucleotides and a short palindrome whose disruption leads to nonfunctional variants (Masumoto et al., 1989; Muro et al., 1992). Further, sequences containing CENP-B boxes have been demonstrated as susceptible to selective forces (Romanova et al., 1996). In this view, constraints on *Leptestheria* satDNA could act on the 14-bp palindromic region as a possible protein binding site.

Curvature analysis reveals that LEP150 monomers are intrinsically bent, with a well-conserved periodicity: this means that also in the variable region, curvature profile does not significantly change. From a general point of view, it can be therefore suggested that a particular superstructure rather than a particular sequence variant could be under selection (Ugarkovic and Plohl, 2002). With respect to LEP150 satDNA, it also appears that although the variable region seems free to change, such changes are actually constrained to those nucleotides able to induce sequence bending.

In this regard, it is also interesting to note the maintenance of the repeat length equal to 150 bp in the 90.2% of monomers. The length of most satDNA monomers corresponds to, or is in multiples of, 150–170 bp, which is approximately the DNA length involved in a nucleosome (Hall et al., 2003).

On the whole, analyses on sequence variation and higher-order structures suggest that LEP150 satDNA could be under selection. This should act at the sequence level with strong constraints in specific regions whose nucleotide sequence must be preserved. Yet, a slightly relaxed selective force could be required also on the whole sequence to maintain the tertiary structure.

Besides MUMs composed by entire monomers arranged head-to-tail, we can also observe incomplete monomers, one duplication event, and three inversions. These rearrangements are likely to be the product of recombination events; the observed lower degree of homogenisation with respect to other monomers raises the possibility that these amplicons derive from regions with lower efficiency in genomic turnover mechanism. One hypothesis is that such monomers derive from the bordering regions of the satDNA array, as observed for α -satDNA (Mashkova et al., 1998, 2001; Bassi et al., 2000).

Monomers of different length can be the result of a crossing over occurring between misaligned repeats; such misalignments could happen by base pairing between short

homologous regions within the repeat itself. It has been observed that pairing of short, even imperfect, homologous sequences is sufficient to ensure recombination (Rubnitz and Subramani, 1984). Therefore, direct and inverted subrepeats found in LEP150 monomers can be taken into account as elements that can produce recombination.

Another interesting link between DNA structure and recombination is the nonrandom coincidence of kinkable dinucleotides and rearrangement breakpoints (Mashkova et al., 2001 and references therein). Our analysis confirms this finding, since we found 67% of scored breakpoints occurring at these peculiar sites.

The lower variability between MUM monomers confirms the existence of a short-range homogenisation process, as also observed in human α -satDNA (Durfy and Willard, 1989; Schindelbauer and Schwarz, 2002). On the other hand, ANOVA results at the individual and population ranks indicate that variability values may differ significantly, but the same range of sequence diversity is experienced at the two considered levels. Although a consistent degree of variation is allowed, our findings indicate that the evolution of the *L. dahalacensis* satDNA is concerted at the species level, since fixation of particular sequence variants does not seem to occur at any considered level.

Acknowledgment

We wish to thank Dr. Giovanni Perini for dot blot facilities and helpful discussion of the manuscript. This work was supported by Ministero Università, Ricerca Scientifica e Tecnologica 40% funds.

References

- Bagshaw, J.C., Buckholt, M.A., 1997. A novel satellite/microsatellite combination in the genome of the marine shrimp, *Penaeus vannamei*. *Gene* 184, 211–214.
- Bassi, C., Magnani, I., Sacchi, N., Saccone, S., Ventura, A., Rocchi, M., Marozzi, A., Ginelli, E., Meneveri, R., 2000. Molecular structure and evolution of DNA sequences located at the alpha satellite boundary of chromosome 20. *Gene* 256, 43–50.
- Bigot, Y., Hamelin, M.H., Periquet, G., 1990. Heterochromatin condensation and evolution of unique satellite-DNA families in two parasitic wasp species: *Diadromus pulchellus* and *Eupelmus vuilleti* (Hymenoptera). *Mol. Biol. Evol.* 7, 351–364.
- Bolshoy, A., McNamara, P., Harrington, R.E., Trifonov, E.N., 1991. Curved DNA without A–A: experimental estimation of all 16 DNA wedge angles. *Proc. Natl. Acad. Sci. U. S. A.* 88, 2312–2316.
- Cesari, M., Luchetti, A., Passamonti, M., Scali, V., Mantovani, B., 2003. Polymerase chain reaction amplification of the Bag320 satellite family reveals the ancestral library and past gene conversion events in *Bacillus rossius* (Insecta Phasmatodea). *Gene* 312, 289–295.
- Charlesworth, B., Sniegowski, P., Stephan, W., 1994. The evolutionary dynamics of repetitive DNA in eukaryotes. *Nature* 371, 215–220.
- Del Sal, G., Manfioletti, G., Schneider, C., 1989. The CTAB-DNA precipitation method: a common mini-scale preparation of template DNA from phagemids, phages or plasmids suitable for sequencing. *BioTechniques* 7, 514–520.

- Dover, G.A., 2002. Molecular drive. *Trends Genet.* 18, 587–589.
- Durfy, S.J., Willard, H.F., 1989. Patterns of intra and interarray sequence variation in alpha satellite from the human X chromosome: evidence for short range homogenization of tandemly repeated DNA sequences. *Genomics* 5, 810–821.
- Fitzgerald, D.J., Dryden, G.L., Bronson, E.C., Williams, J.S., Anderson, J.N., 1994. Conserved pattern of bending in satellite and nucleosome positioning DNA. *J. Biol. Chem.* 269, 21303–21314.
- Hall, S.E., Kettler, G., Preuss, D., 2003. Centromere satellites from *Arabidopsis* populations: maintenance of conserved and variable domains. *Genome Res.* 13, 119–205.
- Kumar, S., Tamura, K., Jakobsen, I.B., Nei, M., 2001. MEGA2: Molecular Evolutionary Genetics Analysis Software. Arizona State University, Tempe, AZ, USA.
- Luchetti, A., Cesari, M., Carrara, G., Cavicchi, S., Passamonti, M., Scali, V., Mantovani, B., 2003. Unisexuality and molecular drive: Bag320 sequence diversity in *Bacillus* Taxa (Insecta Phasmatodea). *J. Mol. Evol.* 56, 587–596.
- Luchetti, A., Cesari, M., Scanabissi, F., Mantovani, B., 2004. Genetic variability of repetitive sequences in Italian and Austrian populations of *Leptestheria dahalacensis* (Rüppel 1837) (Conchostraca). Vth International Large Branchiopod Symposium, Toodyay, Western Australia, August 16–20.
- Maiorano, D., Cece, R., Badaracco, G., 1997. Satellite DNA from the brine shrimp *Artemia* affects the expression of a flanking gene in yeast. *Gene* 189, 13–18.
- Mashkova, T., Oparina, N., Alexandrov, I., Zinovieva, O., Marusina, A., Yurov, Y., Lacroix, M.H., Kisselev, L., 1998. Unequal crossing-over is involved in human alpha satellite DNA rearrangements on border of the satellite domain. *FEBS Lett.* 441, 451–457.
- Mashkova, T.D., Oparina, N.Yu., Lacroix, M.H., Fedorova, L.I., Tumeneva, I.G., Zinovieva, O.L., Kisselev, L.L., 2001. Structural rearrangements and insertions of dispersed elements in pericentromeric alpha satellites occur preferably at kinkable DNA sites. *J. Mol. Biol.* 305, 33–48.
- Masumoto, H., Masukata, H., Muro, Y., Nozaki, N., Okazaki, T., 1989. A human centromere antigen (CENP-B) interacts with a short specific sequence in alphoid DNA, a human centromeric satellite. *J. Cell. Biol.* 109, 1963–1973.
- Motta, M.C., Landsberger, N., Merli, C., Badaracco, G., 1998. In vitro reconstitution of *Artemia* satellite chromatin. *J. Biol. Chem.* 273, 18028–18039.
- Muro, Y., Masumoto, H., Yoda, K., Nozaki, N., Ohashi, M., Okazaki, T., 1992. Centromere protein B assembles human centromeric alpha-satellite DNA at the 17-bp sequence, CENP-B box. *J. Cell. Biol.* 116, 585–596.
- Plohl, M., Mestrovic, N., Bruvo, B., Ugarkovic, D., 1998. Similarity of structural features and evolution of satellite DNAs from *Palorus subdepressus* (Coleoptera) and related species. *J. Mol. Evol.* 46, 234–239.
- Pons, J., Juan, C., Petitpierre, E., 2002. Higher-order organization and compartmentalization of satellite DNA PIM357 in species of the coleopteran genus *Pimelia*. *Chromosome Res.* 10, 597–606.
- Romanova, L.Y., Deriagin, G.V., Mashkova, T.D., Tumeneva, I.G., Mushegian, A.R., Kisselev, L.L., Alexandrov, I.A., 1996. Evidence for selection in evolution of alpha satellite DNA: the central role of CENP-B/pJ α binding region. *J. Mol. Biol.* 261, 334–340.
- Rubnitz, J., Subramani, S., 1984. The minimum amount of homology required for homologous recombination in mammalian cells. *Mol. Cell. Biol.* 4, 2253–2258.
- Sambrook, J., Fritsch, E.T., Maniatis, T., 1989. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
- Schindelhauer, D., Schwarz, T., 2002. Evidence for a fast, intrachromosomal conversion mechanism from mapping of nucleotide variants within a homogeneous α -satellite DNA array. *Genome Res.* 12, 1815–1826.
- Smith, G.P., 1976. Evolution of repeated DNA sequences by unequal crossover. *Science* 191, 528–535.
- Swofford, D.L., 2001. PAUP* Phylogenetic Analysis Using Parsimony (*and Other Methods), Version 4b. Sinauer Associates, Sunderland, MA.
- Ugarkovic, D., Plohl, M., 2002. Variation in satellite DNA profiles causes and effects. *EMBO J.* 21, 5955–5959.
- Ugarkovic, D.L., Plohl, M., Lucijanec-Justic, V., Borstnik, B., 1992. Detection of satellite DNA in *Palorus ratzeburgii*: analysis of curvature profiles and comparison with *Tenebrio molitor* satellite DNA. *Biochimie* 74, 1075–1082.
- Varadaraj, K., Skinner, D.M., 1994. Cytoplasmic localisation of transcripts of a complex G+C-rich crab satellite DNA. *Chromosoma* 103, 423–431.
- Willard, H.F., Wayne, J.S., 1987. Hierarchical order in chromosome-specific human alpha satellite DNA. *Trends Genet.* 3, 192–198.

1 **Evolution of LEP150 sub-repeat array within the ribosomal IGS of the clam**
2 **shrimp *Leptestheria dahalacensis* (Crustacea Branchiopoda**
3 **Conchostraca): the molecular sweep hypothesis.**

4

5 Andrea Luchetti*, Franca Scanabissi and Barbara Mantovani

6

7 *Dipartimento di Biologia E. S., Università degli Studi di Bologna, Bologna, Italia*

8

9

10

11

12

13 *Corresponding Author: dr Andrea Luchetti, Dipartimento di Biologia
14 Evoluzionistica Sperimentale, Università di Bologna, Via Selmi 3, 40126,
15 Bologna, Italia; Tel: +39-51-209-4169; Fax: +39-51-2094286; E-mail:
16 andluche@alma.unibo.it

17

1 **Abstract**

2 *Leptestheria dahalacensis* genome harbors repeats of the low copy number
3 LEP150 satellite DNA family linked to 5S genes, within the ribosomal intergenic
4 spacer. The sequence analysis of the region (5S, flanking region, first satellite
5 monomer: locus A, second satellite monomer: locus B) in genetically isolated
6 samples evidenced three 5S variants, α , β and γ . The α and γ variants show a
7 greater affinity and co-occur in the Central European samples, while in the
8 Italian one, the highly divergent α and β variants are present. A peculiar
9 clustering of LEP150 A and B monomers was further confirmed through the
10 sequencing for the α variant of four monomers at the 5'/ 3' tails (loci A, B, C, D
11 and D', C', B', A', respectively): mutations do not spread among bordering
12 repeats, nor they do among A and B (or A' and B'). Significantly, loci C, D, C'
13 and D' form a unique cluster. The observed pattern of variation is explained
14 taking into account the presence at the LEP150 array borders of two loci under
15 natural selection: the 5S rRNA gene, upstream, and the rDNA transcription
16 promoter, downstream. These elements may drive the dynamics of flanking
17 regions and linked repeats in a process similar to selective sweep. At variance
18 with classical genetic hitchhiking, the selective sweep here scored is realized
19 and maintained through molecular drive; therefore, one should refer to this
20 process as "molecular sweep".

21

22

23 Key words: ribosomal DNA; concerted evolution; repetitive DNA sequences;
24 genomic organization; selective pressure; genetic hitchhiking.

25

26

1 **Introduction**

2 A large fraction of the eukaryotic genome is made by repetitive DNA
3 sequences, either interspersed or tandemly organized. Among the latter, highly
4 repeated DNAs (also called satellite DNAs or satDNAs) are usually non-
5 transcribed sequences, while middle repeated DNAs comprise ribosomal
6 (rDNAs) and histone genes (Elder and Turner 1995).

7 Repeated DNAs have been observed to evolve following a pattern known
8 as “concerted evolution”, in which a major repeat homogeneity is observed
9 within lineages (strains, populations, subspecies, species, etc.) than between
10 them (Dover 1982). Concerted evolution is achieved by the evolutionary
11 process of molecular drive, consisting of variant homogenization within
12 genomes through non-Mendelian mechanisms of genomic turnover (such as
13 gene conversion and unequal crossing-over, slippage replication, rolling circle
14 replication and reinsertion, and transposon-mediated exchange) and variant
15 fixation in the taxonomic unit by means of bisexual reproduction (Dover 2002).

16 Studies on satellite DNA also evidenced the existence of the so-called
17 “library hypothesis”: related taxa could share a number of satellite DNA families
18 that can be species-specifically amplified (Southern 1975; Salser *et al.* 1976).
19 This has been experimentally demonstrated in the coleopteran genus *Palorus*
20 (Mestrovic *et al.* 1998) and in the *Bacillus* stick insect species complex (Cesari
21 *et al.* 2003).

22 The life history of different satellite DNAs coexisting in the same genome
23 has also been described by Nijman and Lenstra (2001) through a “Feedback
24 Model”. Briefly, a satellite DNA family encounter three phases during its life: i) in
25 phase I interactions of homogeneous repeats cause rapid expansions as well
26 as contractions with saltatory fluctuations in the copy number; ii) in phase II

1 mutation and recombination events lead to new variants, evolving
2 independently; iii) the terminal phase III is reached when degeneration by
3 mutations stops interactions between old monomers and a new satDNA family
4 takes their place.

5 Albeit a great number of studies on repetitive DNA are now available,
6 most of them describe its evolutionary dynamics by analyzing monomers
7 randomly taken from the main array. On the other hand, both theoretical and
8 experimental work have shown that monomers at the cluster ends are less
9 homogeneous than other repeats, possibly because genomic turnover
10 mechanisms are less efficient in these regions (Smith 1976; Mashkova et al.
11 1998, 2001; Bassi et al. 2000). Good models for testing this particular dynamics
12 are ribosomal intergenic spacer (IGS) subrepeat arrays, as demonstrated in
13 *Daphnia pulex* (Crease 1995) and in the swimming crabs (Ryu et al. 1999), in
14 which bordering repeats are less homogenized with respect to the inner ones.

15 The genome of the clam shrimp *Leptestheria dahalacensis* embodies the
16 LEP150 family, a low copy number satellite DNA (0.5% of the genome) with
17 monomeric units of 150 bp and mean sequence diversity of 10.3% (Luchetti *et*
18 *al.* 2004). In this paper, the co-evolutionary dynamics of 5S rRNA gene and
19 LEP150 sequences located within the ribosomal IGS of three European clam
20 shrimp samples is analyzed.

21

22 **Materials and Methods**

23 Samples of *Leptestheria dahalacensis* were collected in rice field (Ferrara, Italy)
24 and natural ponds (Marchegg, Austria; Hadamar, Germany), and alcohol
25 preserved. Total DNA was extracted from a single individual per population
26 through the CTAB method (Winnepeninckx *et al.* 1993). The three samples

1 were labeled as IT, AU and GE, for Italy, Austria and Germany, respectively.

2 We first amplified the complete ribosomal IGS with primers 28ii, modified for
3 branchiopods (5'- GGC TCT TCC TAT CAT TGC GAA GCA GTA TTC GC -3')

4 and 18i (5'- TTT CTC AGG CTC CCT CTC CGG AAT CGA ACC CT -3') (Hillis
5 and Dixon, 1991), as described in Luchetti et al. (2006). Through amplicon
6 sequencing 5S rRNA gene was found upstream the sub-repeat cluster
7 composed of LEP150 sequences already identified as satellite DNA (Fig. 1;
8 Luchetti et al., 2004a, b). In order to obtain 5S-LEP150 sequences, two primers
9 allowing the amplification of fragments containing 60bp of 5S, 3' flanking region
10 and LEP150 repeats were first designed. These primers were: 5Sd (5'-GTC
11 AGA TCC CGG AAG TCA AG-3'), annealing within 5S rDNA, and monF (5'-
12 GCT GGT TTT CTH KST TGT AGA CG-3') annealing to the 3' end of LEP150
13 repeats (Fig. 1). PCR was performed in a MJ PTC-100 thermal cycler (MJ
14 Research) with the following program: initial denaturation at 95°C; 30 cycles at
15 95°C for 30 sec., 48°-54°C for 1 minute, 72°C for 1 minute; final extension at
16 72°C for 7 minutes. The amplification gave a ladder-like product with bands
17 differing by 150 bp. Bands corresponding to 550 bp and 700 bp were eluted
18 from the gel, ligated to pGEM T-easy vector (Promega), and used to transform
19 *E. coli* DH5 α -competent cells (Invitrogen). Recombinant colonies were amplified
20 with M13 primers and sequenced with the Dye terminator cycle sequencing kit
21 (Applied Biosystem) in a 310 Genetic Analyzer (ABI) automatic sequencer. Only
22 clones containing two LEP150 monomers were further considered.

23 A further primer, annealing 323 bp downstream the LEP150 array (IGLup6: 5'-
24 TGT CGT ATT CAG AGG AGT AGT AAA TCA -3'), was tested together with
25 primer 5Sd to amplify 5S rDNA, LEP150 complete array and part of ribosomal
26 IGS (Fig. 1). Amplicons were cloned and the tails, until the fourth monomer,

1 were sequenced using primers 5Sd and IGLup6 in conjunction with monF and
2 monR (5'- TCT YGA GAT TCC TGG GTT RT -3'), respectively.

3 Sequence data from this article have been deposited with the EMBL/GenBank
4 Data Libraries under accession nos.: AY772675 – AY772681, AY772684 –
5 AY772688, DQ303879 – DQ303917, XXXXXX - YYYYYY.

6 Sequences were aligned using the CLUSTAL algorithm (Sequence Navigator,
7 v.1.1, Applied Biosystem). Uncorrected p distances (p-D) and standard error
8 (S.E.) were calculated with Mega 3 package (Kumar *et al.* 2004); gene
9 conversion events and sliding window analyses of nucleotide diversity were
10 computed with DnaSP program v. 3 (Rozas and Rozas 1999). Phylogenetic
11 analyses, based on Maximum Parsimony method, were performed using
12 PAUP* v. 4.0b8a (Swofford 2001) with 1000 bootstrap replicates; gaps were
13 considered as missing data as no differences were observed when considering
14 them as 5th state.

15

16 **Results**

17 PCR amplification with 5Sd / monF primers produced 56 clones containing 60
18 bp of the 5S rDNA coding region, the 3' flanking region and two LEP150 repeat
19 units. Only four point mutations, among all clones, occur within the first 60 bp,
20 corresponding to the 3' end of 5S rDNA gene, this coding region being
21 otherwise completely homogeneous. On the other hand, three different 3'
22 flanking region - called α , β and γ - were retrieved. The α 3' flanking region is
23 present in all samples; it is 315 bp long, with sequence diversity ranging from
24 0.004 ± 0.002 to 0.012 ± 0.003 (Table 1). The β variant is found only in the IT
25 sample; it is 205 bp long and shows an internal variability of 0.004 ± 0.003 . The

1 γ variant is 355 bp with a sequence variability of $0.008 \pm 0.003 - 0.011 \pm 0.004$
2 (Table 1), and it was retrieved only in the AU and GE populations.

3 The α and γ variants are the most similar, differing for a 37 bp deletion/insertion
4 (Fig. 1) and showing a mean p-D of 0.034 ± 0.008 . The β variant differs from
5 the α and γ ones for four and five large deletions respectively (Fig. 1) and a p-D
6 of $0.152 \pm 0.026 - 0.143 \pm 0.024$.

7 The LEP150 repeat directly linked to the 3' flanking region (locus A) show a
8 sequence diversity ranging from 0.005 ± 0.004 (within IT β dataset) to $0.031 \pm$
9 0.011 (within the GE γ dataset). The following LEP150 monomer (locus B)
10 experiences a wider range of variability, with p-D values ranging from $0.006 \pm$
11 0.004 to 0.127 ± 0.023 (Table 1).

12 On the whole, LEP150 nucleotide diversity slow down when approaching the 5S
13 3' flanking region, with the exception of α and γ datasets in the AU sample,
14 where locus A is the more variable (Table 1).

15 In order to better define this variability pattern and to check if it is detectable
16 also at the other end of the LEP150 cluster, the complete array was amplified,
17 cloned, and the 5'/ 3' tails were sequenced. This analysis was performed only
18 for the α variant because it is shared by all samples and also the most
19 represented within each sample, thus allowing a comparative analysis. As
20 above, the 5' end of the array comprises 60 bp of 5S, the 3' flanking region, but
21 in this analysis four LEP150 repeats (loci A, B, C and D) were considered. The
22 3' end encompasses four monomers (loci D', C', B' and A') linked to 323 bp of
23 the IGS sequence. The sliding window analysis (Fig. 2A) at the 5' end confirms
24 the distribution of variability already observed, further indicating that in the
25 Austrian sample this general trend can be as well observed. Moreover, also at
26 the 3' end of the array the nucleotide diversity slow down when approaching the

1 IGS sequence. Visual inspection reveals that a short motif with high sequence
2 similarity to previously characterized transcription promoters is present at the
3 edge of the array (Fig. 2B).

4 Gene conversion events were searched at the intra-genomic level through
5 comparisons within and between datasets. Only nine converted tracts, 11 to 74
6 bp long, were found between 3' flanking regions always moving from γ to α
7 dataset in the AU (five conversions) and GE samples (four conversions).

8 Maximum Parsimony analysis (Fig. 3) was carried out on the 296 LEP150
9 monomers here obtained. The resulting dendrogram does not evidence any
10 sample-specific cluster. Beside some exceptions (see below), sequences from
11 loci A and B as well as from loci A' and B' form well defined, locus-specific
12 clusters and monomers linked to both α and γ 3' flanking region variants
13 intermingle. Repeats from β dataset always group in isolated clusters for both A
14 and B loci. It is to be noted that LEP150 monomers from C, D, C' and D' loci
15 cluster together, completely intermingling. Ten sequences do not fall in the
16 expected clusters, all but one pertaining to the German sample: in these
17 instances, repeats from loci D and/or D' group are found within the B and/or B'
18 cluster and vice versa.

19

20 **Discussion**

21 The sub-repeat array within the ribosomal IGS provides an interesting
22 framework to study the evolutionary behavior of tandem repetitive DNA,
23 especially for bordering repeats. In this study the analysis of the array focuses
24 on both ends of the cluster, further comprising flanking sequences: upstream,
25 part of the 5S gene and its 3' flanking region and, downstream, the putative
26 gene promoter site and 323 bp of the IGS sequence.

1 On the basis of the 5S-3' flanking region structure it is possible to hypothesize
2 that each of the three variant (α , β and γ) pertains to different rDNA loci.
3 Sequence analysis of 5S-3' flanking regions and their distribution across
4 samples confirms microsatellite data on *L. dahalacensis* population
5 differentiation (Cesari et al., in press): AU and GE populations are the closest
6 since both lack the β variant and share the γ one, which in turn does not appear
7 in the IT sample. Only the α variant is present in all samples. β and γ variants
8 may have been lost or de novo formed or simply they could pertain to a rDNA
9 "library" (Southern 1975; Salser et al. 1976; Mestrovic et al. 1998) being
10 therefore present in very low copy numbers in all samples. However, the high
11 affinity between α and γ sequences, both at the 3' flanking region and along
12 linked LEP150 monomers, strengthen the hypothesis of a recent origin of the γ
13 variant with respect the β one. This is also evident by the presence of gene
14 conversion events between α and γ sequences. The Feedback Model (Nijman
15 and Lenstra, 2001) may be applied to the rDNA evolution: at present, α and γ 3'
16 flanking region types are enough similar that DNA exchange is still allowed
17 (phase I), while α and β variants have become too much diverse so that
18 interactions between them are prevented (phase II). On the basis of this model
19 and of the molecular drive processes, given enough time, α and γ variants
20 should become completely differentiated or homogenized in a single rDNA type.
21 This clearly explains also the clustering of LEP150 sequences linked to the β
22 variant that group in isolated and well-defined clusters.
23 It has been observed that, in a tandem array, repeats located at the edges
24 significantly differentiated from the inner ones because genomic turnover
25 mechanisms have less efficiency in spreading new mutations in those loci

1 (Smith, 1976; Mashkova et al. 1998, 2001; Bassi et al. 2000). On the other
2 hand, it has been demonstrated that one of the leading forces driving
3 homogenization is distance: contiguous repeats are less divergent than
4 monomers randomly sampled from same or different array (Durfy and Willard,
5 1989; Schindelhauer and Schwarz, 2002), leading to the “short range
6 homogenization” pattern. This hold also for LEP150 sequences (Luchetti et al.
7 2004). The peculiar clustering of LEP150 sequences of A and B loci, as well as
8 A' and B', confirm that such condition cannot be applied to bordering repeats:
9 mutation do not spread among inner and external loci, nor they do among A
10 and B (or A' and B'). Indeed, it is significant that loci C, D, C' and D' form a
11 unique cluster, indicating the spreading of mutations across inner repeats of the
12 same array.

13 In the IGS sub-repeat array of *Daphnia pulex*, bordering repeats are
14 differentiated from the inner ones, and the variability among sub-repeats in that
15 position (locus) is higher than among sub-repeats at each inner position
16 (Crease, 1995). This has been observed also in the ribosomal IGS sub-repeat
17 array of the swimming crab (Ryu et al., 1999). In the case of LEP150 repeats,
18 bordering monomers (loci A-B and B'-A') are differentiated from the others, but,
19 on the contrary to the expected, they are less variable: there is a clear pattern
20 of sequence diversity decreasing when approaching the end of the array. This
21 suggests that, beside intra-locus homogenization, some other force should act
22 in such faster elimination (or fixation) rate of sequence variants in external
23 monomers with respect to the inner ones.

24 At both side of the LEP150 sub-repeat array there are two functional loci: the
25 5S rRNA gene, at the 5' end, and a gene promoter, at the 3' end. It is likely that
26 the presence of these elements drive the evolution of linked LEP150

1 monomers: they influence LEP150 variability because they are selected, in a
2 process similar to the one described as “selective sweep”. In the selective
3 sweep model a local reduction in genetic diversity is due to the rapid fixation of
4 an advantageous mutation, and selected mutations drag linked alleles in their
5 fixation (hitchhiking effect; Barton 2000). Let us consider, for instance, the 5S
6 gene: being repetitive, it evolves by molecular drive, but only the variant(s)
7 compatible with its function are homogenized. In this case, it likely that 5S drag,
8 in its way to homogenization, joined LEP150 monomers forming with them a
9 higher order repeat (HOR): in this way only those repeats linked to the
10 selectively advantageous 5S variant would be homogenized and fixed as they
11 would be also selected. In this instance, the observed variability must
12 decrease. Furthermore, locus A would be probably more often included in such
13 a HOR, thus being less variable and significantly differentiated from the locus B:
14 in this way the variability decrease when approaching to the 5S. Obviously the
15 same apply at the 3’ end of the array, where selection on the promoter
16 sequence causes the same decreasing of sequence diversity.

17 The model we propose is different from the selective sweep because molecular
18 drive plays a crucial role in maintaining the homogeneity: we shall refer to this
19 process as “molecular sweep”. Genetic hitchhiking caused by selection usually
20 lead to a valley of nucleotide diversity in the region surrounding the selected
21 site; however after the end of the selective phase the variability start to be
22 restored by accumulation of neutral mutations. Indeed, the longer is the time
23 since the selective event, the lower is the possibility to detect the selective
24 sweep. In the molecular sweep model, the loss of variability could be due to
25 both selection and homogenization by molecular drive, but the latter process

1 takes place several time among repetitive sequences: it should be therefore
2 expected that the effect of molecular sweep would be persistent.
3 On the whole, data presented here point to the non-concerted evolution among
4 repeats at the ends of an array, further providing evidences of the effects (co-
5 evolution) of flanking sequences on the tandem repeat cluster. On this base, it
6 has been hypothesized a new evolutionary model called molecular sweep,
7 which, obviously, deserve further studies in other organisms and genomic
8 contexts.
9

1 **Acknowledgements**

2 We wish to thank Gabriel A. Dover and Miroslav Plohl for stimulating
3 discussions on the evolution of repetitive DNA. We also wish to thank Erich
4 Eder and Larissa Frolova for providing Austrian and German samples,
5 respectively. This work has been supported by 60% funds - University of
6 Bologna.

7

1 **References**

- 2 Barton, N. H., 2000 Genetic hitchhiking. *Phil. Trans. R. Soc. Lond. B* **355**: 1553-
3 1562.
- 4 Bassi, C., I. Magnani, N. Sacchi, S. Saccone, A. Ventura, M. Rocchi, A.
5 Marozzi, E. Ginelli, and R. Meneveri, 2000 Molecular structure and
6 evolution of DNA sequences located at the alpha satellite boundary of
7 chromosome 20. *Gene* **256**: 43-50.
- 8 Cesari, M., A. Luchetti, M. Passamonti, V. Scali, and B. Mantovani, 2003
9 Polymerase chain reaction amplification of the Bag320 satellite family
10 reveals the ancestral library and past gene conversion events in *Bacillus*
11 *rossius* (Insecta Phasmatodea). *Gene* **312**: 289-295.
- 12 Cesari, M., A. Luchetti, F. Scanabissi and B. Mantovani 2007. Genetic
13 variability in European *Leptestheria dahalacensis* (Rüppel, 1837)
14 (Crustacea, Branchiopoda, Spinicaudata). *Hydrobiologia*, in press.
- 15 Crease, T. J., 1995. Ribosomal DNA evolution at the population level:
16 nucleotide variation in intergenic spacer arrays of *Daphnia pulex*. *Genetics*
17 **141**: 1327-1337.
- 18 Dover, G. A., 2002. Molecular drive. *Trends Genet.* **18**: 587-589.
- 19 Dover, G. A., 1982. Molecular drive: a cohesive mode of species evolution.
20 *Nature* **299**: 111-117.
- 21 Elder, J. F., and B. J. Turner, 1995 Concerted evolution of repetitive DNA
22 sequences in eukaryotes. *Quart. Rev. Biol.* **70**: 297-320.
- 23 Hillis, D.M. and M.T. Dixon, 1991. Ribosomal DNA: molecular evolution and
24 phylogenetic inference. *Quart. Rev. Biol.* **66**: 411-453
- 25 Kumar, S., K. Tamura, and M. Nei, 2004 MEGA3: Integrated software for
26 Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief.*

- 1 Bioinform. **5**: 150-163.
- 2 Luchetti, A., A. Marino, F. Scanabissi, and B. Mantovani, 2004a Genomic
3 dynamics of a low copy number satellite DNA family in *Leptestheria*
4 *dahalacensis* (Crustacea, Branchiopoda, Conchostraca). *Gene* **342 (2)**:
5 313-320.
- 6 Luchetti, A., M. Cesari, F. Scanabissi, and B. Mantovani, 2004b Genetic
7 variability of repetitive sequences in Italian and Austrian populations of
8 *Leptestheria dahalacensis* (Rüppel 1837) (Conchostraca). 5th International
9 Large Branchiopod Symposium, Toodyay, Western Australia, 16-20
10 August 2004.
- 11 Luchetti, A., F. Scanabissi, and B. Mantovani, 2006 Molecular characterization
12 of ribosomal intergenic spacer in the tadpole shrimp *Triops cancriformis*
13 (Crustacea, Branchiopoda, Notostraca). *Genome* **49**: 888-893
- 14 Mashkova, T., N. Oparina, I. Alexandrov, O. Zinovieva, A. Marusina, Y. Yurov,
15 M. H. Lacroix, and L. Kisselev, 1998 Unequal crossing-over is involved in
16 human alpha satellite DNA rearrangements on border of the satellite
17 domain. *FEBS Letters* **441**: 451-457.
- 18 Mashkova, T. D., N. Yu. Oparina, M. H. Lacroix, L. I. Fedorova, I. G. Tumeneva,
19 I. G. Zinovievax, and L. L. Kisselev, 2001 Structural rearrangements and
20 insertions of dispersed elements in pericentromeric alpha satellites occur
21 preferably at kinkable DNA sites. *J. Mol. Biol.* **305**: 33-48.
- 22 Mestrovic, N., M. Plohl, B. Mravinac and D. Ugarkovic, 1998 Evolution of
23 satellite DNAs from the genus *Palorus* – Experimental evidence for the
24 library hypothesis. *Mol. Biol. Evol.* **15**: 1062-1068.

- 1 Nijman, I. J., and J. A. Lenstra, 2001 Mutation and recombination in cattle
2 satellite DNA: a feedback model for the evolution of satellite repeats. J.
3 Mol. Evol. **52**: 361-371.
- 4 Rozas, J., and R. Rozas, 1999 DnaSP version 3: an integrated program for
5 molecular population genetics and molecular evolution analysis. Bioinf. **15**:
6 174-175.
- 7 Salser, W., S. Bowen, D. Browne, *et al.* (11 co-authors), 1976 Investigation
8 of the organization of mammalian chromosomes at the DNA sequence
9 level. Fed. Proc. **35**: 23-35.
- 10 Smith, G. P., 1976 Evolution of repeated DNA sequences by unequal
11 crossover. Science **191**: 528-535.
- 12 Southern, E. M., 1975 Long range periodicities in mouse satellite DNA. J.
13 Mol. Biol. **94**: 51-69.
- 14 Swofford, D.L., 2001 PAUP* Phylogenetic Analysis Using Parsimony (*and
15 Other Methods), Version 4b. Sinauer Associates, Sunderland,
16 Massachusetts.
- 17 Winnepeninckx B., T. Backeljau, and R. Wachter, 1993 Extraction of high
18 molecular weight DNA from molluscs. Trends Genet. **9**: 407.
- 19

1 **Table 1**

2 Mean p-Distance and S.E. for 3' flanking region, LEP150 locus A and locus B.

Dataset	Sample	3' fl. region	locus A	locus B
α				
	IT	0.004 \pm 0.002	0.017 \pm 0.006	0.077 \pm 0.014
	AU	0.011 \pm 0.003	0.010 \pm 0.004	0.006 \pm 0.004
	GE	0.012 \pm 0.003	0.026 \pm 0.075	0.124 \pm 0.017
β				
	IT	0.004 \pm 0.003	0.005 \pm 0.004	0.021 \pm 0.008
γ				
	AU	0.008 \pm 0.003	0.013 \pm 0.006	0.008 \pm 0.005
	GE	0.011 \pm 0.004	0.031 \pm 0.011	0.127 \pm 0.023

3

4

1 **Figure Legends**

2

3 Figure 1. Structure of ribosomal intergenic spacer (A) and schematic drawing of
4 obtained α , β and γ sequences (B). Dot in the 3' flanking region of amplicons
5 indicate major deletion with respect to the γ variant. Small arrows indicate the
6 positions of primers, whose name are indicated above.

7

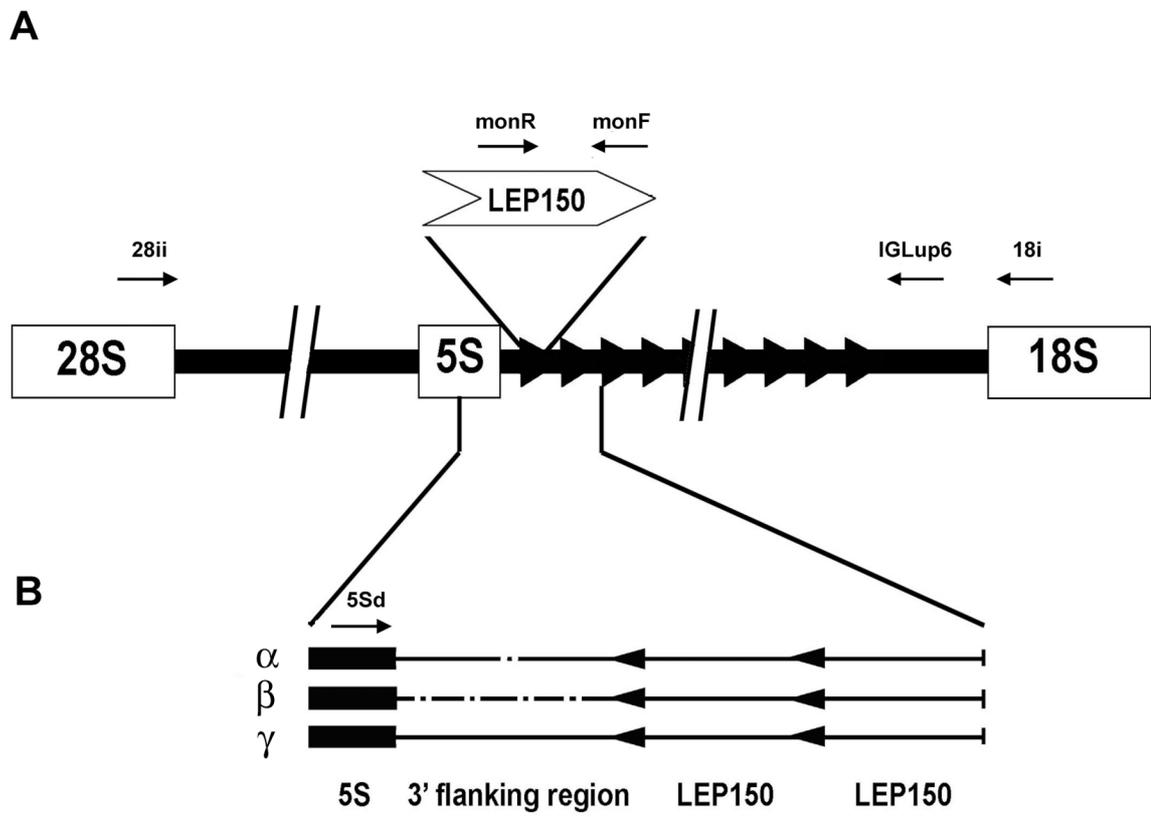
8 Figure 2. A) Distribution of nucleotide variability across the 5' and 3' end of α
9 variant LEP150 sub-repeat array in the three analysed samples. Sliding
10 windows were 150 bp widths, jumping each 75 bp. On the X and Y axes were
11 reported nucleotide positions (midpoint of sliding window) and nucleotide
12 diversity, respectively. Above, sequence structure has been represented;
13 capital letters A-D and D'-A' indicate LEP150 loci. B) Sequence of the putative
14 transcription promoter found in *L. dahalacensis* (Lda). For comparison were
15 also reported promoters of *Daphnia pulex* (Dpu), *Artemia franciscana* (Afr) and
16 *Triops cancriformis* (Tca). The beginning of promoter sequence is indicated by
17 >; g and s indicate gene and spacer promoters. Underlined A in Afr promoters
18 were identified as transcription starting point (Koller et al., 1997).

19

20 Figure 3. Maximum Parsimony dendrogram built on LEP150 monomers (T.L. =
21 492; C.I. = 0.474). LEP150 loci are indicated as in Figure 2. Numbers at nodes
22 represents bootstrap supports > 70% and values on terminal branching have
23 been omitted. The bar below indicates 1 mutational step.

24

1



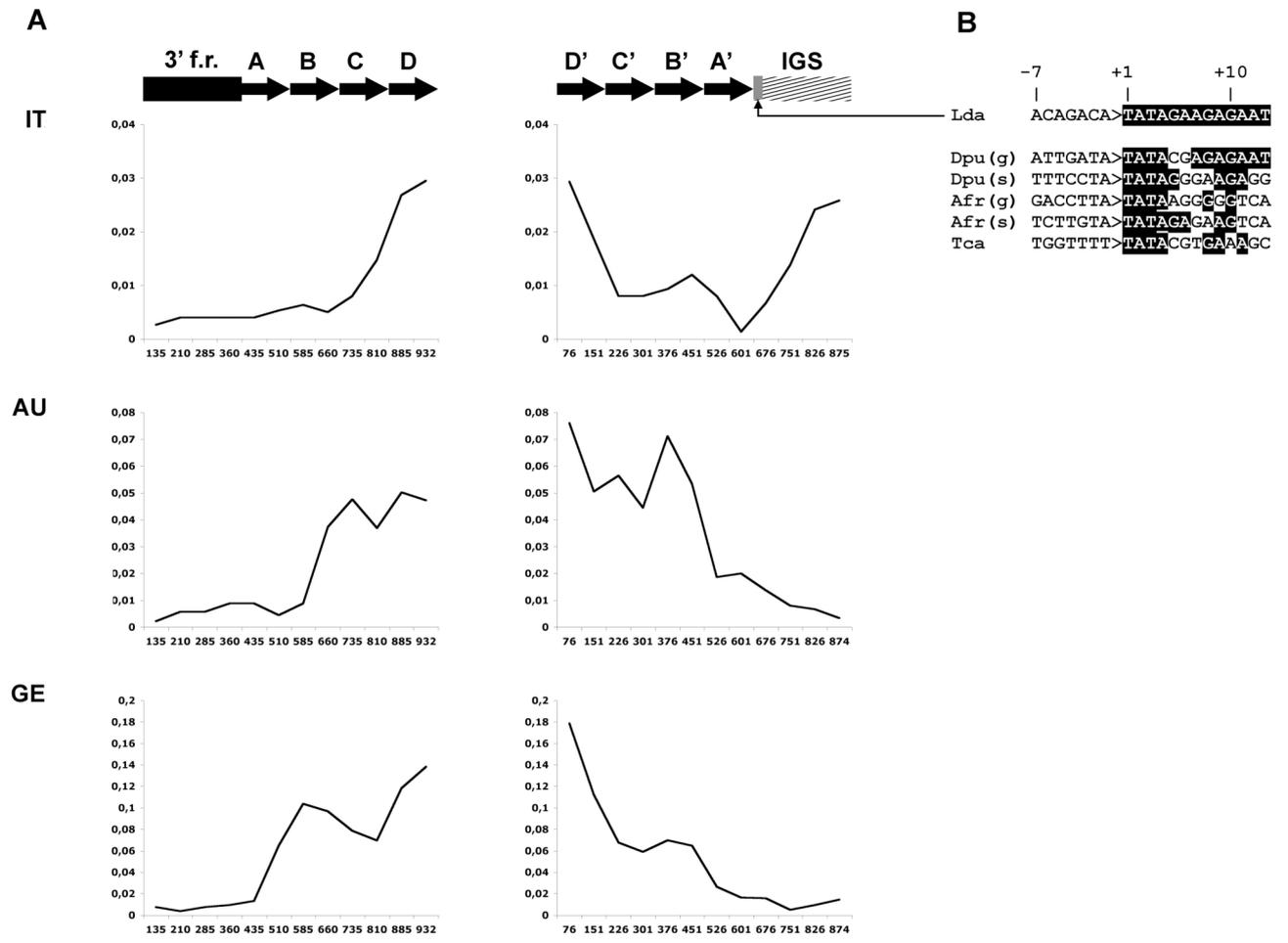
2

3

4 Figure 1

5

1



2

3 Figure 2

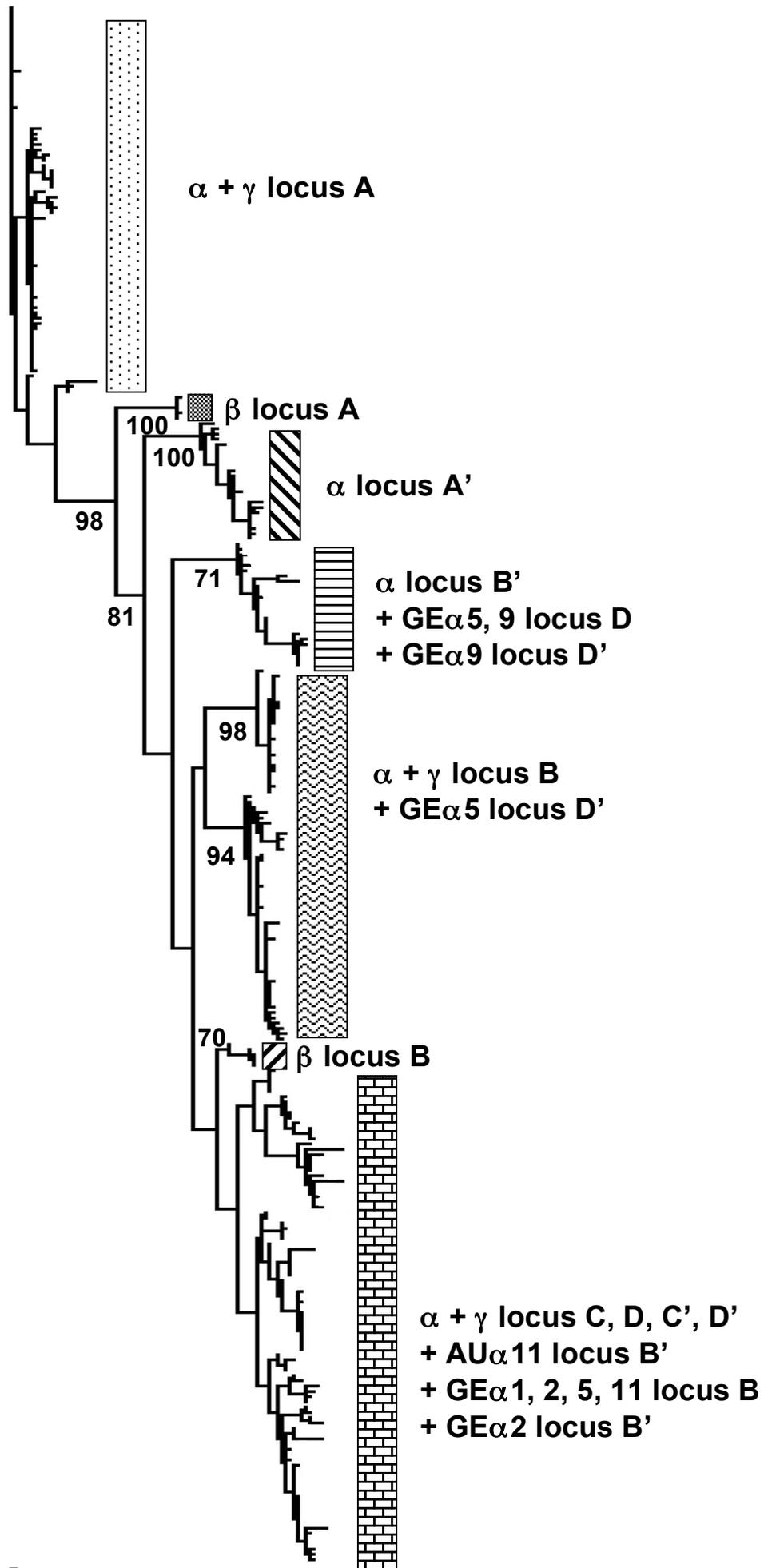


Figure 3

Chapter 5. Conclusions

The pattern of concerted evolution is achieved through the dual process known as molecular drive (Dover, 1982, 2002). It is said “dual” because it depends on both strictly molecular mechanisms (genomic turnover mechanisms: GTMs) and meiotic random chromosome segregation coupled with amphimixis and panmixis.

In many studies on repetitive DNA (either satDNA and rDNA) concerted evolution was observed and both aspects of molecular drive were substantially confirmed (for example: Bruvo et al., 2003; Pons and Gillespie, 2004; Ganley and Kobayashi, 2007). On the other hand, in the last decade a number of papers reported the absence of concerted evolution (for example: Mravinac et al., 2002; Robles et al., 2004; Mestrovic et al., 2006a).

In this thesis, three examples showing that the evolution of different repetitive DNA families does not follow the expectation are reported.

In the first instance, the RET76 satDNA in *Reticulitermes* termites does not evolve in a concerted fashion because species-specific sequence fixation is lacking. This confirmed a peculiar dynamics already observed in other animal systems in which panmixis is not the rule (Mantovani et al., 1997; Luchetti et al., 2003; Navajas and Boursot, 2003; Lorite et al., 2004), therefore establishing a sound link between specific organismal traits, such as eusociality, parthenogenesis, haplo-diploidy and the lack of fixation.

The second case concerns IGS variability and sub-repeat copy number variation in *T. cancriformis*. It highlights two different features: i) at the nucleotide level, the evolution of population-specific IGS sequence, with also significant changes in functional domains, and the conservation of an ancestral state in non-gonochoric populations is evident; ii) at the structural level, the effect of natural selection on the expected IGS length variation pattern is observed. Both aspects are somehow linked to the mechanism of the GTMs: these appear to homogenise at the sequence level but produce polymorphism in repeat copy numbers (Shufran et al., 1997). In the case here reported, while nucleotide sequence divergence follows the expectation, IGS length variation does not. This could be the result of a local adaptation, and then natural selection acts to maintain a specific IGS genotype (Gorokhova et al., 2002).

The third example comes from the evolution of a low copy number satellite DNA in the clam shrimps that was also found in the ribosomal IGS. The evolution of this repetitive DNA family, called LEP150, is quite complex being subject to both molecular drive and (relaxed) natural selection. The observed “selective sweep” effect (with selected mutations dragging linked alleles in their fixation and leading to a local reduction in genetic diversity; Barton 2000) is here due to the presence of loci under strong selective pressure (5S rDNA and IGS promoter sequence) at repeat array ends. LEP150 bordering repeat variability is therefore strongly influenced. Indeed, these monomers are more conserved than the expected on the basis of GTMs, which are freely acting on the bulk of repeats (see for instance Crease, 1995; Ryu et al., 1999; Mashkova et al. 1998, 2001). To explain this peculiar situation, in which molecular drive and natural selection co-occur, the “molecular sweep” model is here proposed.

Natural selection and random genetic drift alone cannot explain why tandemly repetitive DNA evolves in concert (Dover, 1982), but the here presented data on different repetitive DNA families and animal systems indicate that also molecular drive alone cannot explain some of the observed pattern of repetitive DNA sequence evolution. In each example reported in this thesis, the cohesive nature of repeat evolution clearly emerges; however the action of forces other than molecular drive is well evident.

Natural selection, in particular, seems to play an important role and it is probable that future studies will increase the importance of this evolutionary force in the evolution of repetitive DNA. This is quite intuitive in the evolution of rDNA, particularly regarding rRNA genes (Nei and Rooney, 2005), but during the last four years its effect also on non-coding repetitive DNA has been demonstrated (for example Hall et al., 2003; Mestrovic et al., 2006b). In the cases reported here, natural selection seems to act both directly and indirectly, through molecular sweep, and both on functional and apparently non-functional domains. Generally speaking, therefore, one can extend the conclusions of Mestrovic and co-workers (2006b) on MEL172 satDNA, that the evolution of this repetitive DNA family is an interplay between stochastic and selective events, also to other repeated DNAs.

Beside natural selection, there is also another aspect to take into account, independent from selection but equally contributing to the molecular drive failure.

Being a process strictly linked to recombination and chromosome reshuffling, all instances influencing these events directly affect also molecular drive. Unisexual reproduction, as well as all deviations from the canonical panmixis, such as eusociality and/or haplo-diploidy regarding chromosomes, has been demonstrated to upset the process of fixation (Mantovani et al., 1997; Luchetti et al., 2003; Navajas and Boursot, 2003; Lorite et al., 2004). This happens because, lacking the possibility of random mating, all new mutations cannot spread (and fix) among individuals of a population, and then across a taxon. This has particular impact on the evolution of both genomes and organisms, since repetitive DNA constitutes a substantial fraction of eukaryotic genome and it is often involved in crucial functions, such as centromere/telomere maintenance, rRNA and ribozyme syntheses. In these tasks, repetitive DNA interacts with a number of specific proteins (f.i. CENP and RNA pol-I) with which coevolves (Dover and Flavell, 1984). Henikoff and co-workers (2001), studying the DNA-protein interactions at centromeres argued that these molecules (co)evolve very rapidly, quickly diverging also between closely related species. Therefore, this can be responsible for the very different centromere organizations across organisms and for reproductive isolation of emerging species (Henikoff et al., 2001). It is clear that, in this view, conservation of mutation profile in functional repetitive DNAs (satDNA, rDNA) could cause the maintenance of cross compatibilities between diverging species. In terms of organisms life history this could mean the possibility of hybridization at least between sister species.

The mechanisms underlying the evolution of repetitive DNA are still in part not defined. While the formulation of molecular drive theory, with its dual nature, shed light on the observed pattern of concerted evolution, little is still known on the evolutionary dynamics of repetitive DNAs when requirements for the achievement of molecular drive are not met. This thesis adds some new features to take into account when tackling with repetitive DNA evolution, in line with the most recent literature on the topic.

References

Barton NH (2000). Genetic hitchhiking. *Philosophical Transactions of the Royal Society of London Series B* 355: 1553-1562.

Bruvo B, Pons J, Ugarkovic D, Juan C, Petitpierre E, Plohl M (2003). Evolution of low-copy number and major satellite DNA sequences coexisting in two *Pimelia* species-groups (Coleoptera). *Gene* 312: 85-94.

Crease TJ (1995). Ribosomal DNA evolution at the population level: nucleotide variation in intergenic spacer arrays of *Daphnia pulex*. *Genetics* 141: 1327-1337.

Dover GA (1982). Molecular drive: a cohesive mode of species evolution. *Nature* 299: 111-117.

Dover GA (2002). Molecular drive. *Trends in Genetics* 18: 587-589.

Dover GA, Flavell RB (1984). Molecular coevolution: DNA divergence and the maintenance of function. *Cell* 38: 622-623.

Ganley ARD, Kobayashi T (2007). Highly efficient concerted evolution in the ribosomal DNA repeats: total rDNA repeat variation revealed by whole-genome shotgun. *Genome Research*, in press

Gorokhova E, Doeling TE, Weider LJ, Crease TJ, Elser JJ (2002). Functional and ecological significance of rDNA intergenic spacer variation in a clonal organism under divergent selection for production rate. *Proceedings of the Royal Society of London Series B* 269: 2373-2379.

Henikoff S, Ahmad K, Malik HS (2001). The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* 293: 1098-1102.

Lorite P, Carrillo JA, Tinaut A, Palomeque T (2004). Evolutionary dynamics of satellite DNA in species of the genus *Formica* (Hymenoptera, Formicidae). *Gene* 332: 159-168.

Luchetti A, Cesari M, Carrara G, Cavicchi S, Passamonti M, Scali V, Mantovani B (2003). Unisexuality and molecular drive: Bag320 sequence diversity in *Bacillus* taxa (Insecta Phasmatodea). *Journal of Molecular Evolution* 56: 587-596.

Mantovani B, Tinti F, Bachmann L, Scali V (1997). The Bag320 satellite DNA family in *Bacillus* stick insects (Phasmatodea): different rates of molecular evolution of highly repetitive DNA in bisexual and parthenogenetic taxa. *Molecular Biology and Evolution* 14: 1197-1205.

Mashkova T, Oparina N, Alexandrov I, Zinovieva O, Marusina A, Yurov Y, Lacroix MH, Kisselev L (1998). Unequal crossing-over is involved in human alpha satellite DNA rearrangements on border of the satellite domain. *FEBS Letters* 441: 451-457.

Mashkova TD, Oparina NYu, Lacroix MH, Fedorova LI, Tumeneva IG, Zinovieva IG, Kisselev LL (2001). Structural rearrangements and insertions of dispersed elements in pericentromeric alpha satellites occur preferably at kinkable DNA sites. *Journal of Molecular Biology* 305: 33-48.

Mestrovic N, Castagnone-Sereno P, Plohl M (2006a). High conservation of the differentially amplified MPA2 satellite DNA family in parthenogenetic root-knot nematodes. *Gene* 376: 260-267.

Mestrovic N, Castagnone-Sereno P, Plohl M (2006b). Interplay of selective pressure and stochastic events directs evolution of the MEL172 satellite DNA library in root-knot nematodes. *Molecular Biology and Evolution* 23: 2316-2325.

Mravinac B, Plohl M, Mestrovic N, Ugarkovic D (2002). Sequence of PRAT satellite DNA "frozen" in some Coleopteran species. *Journal of Molecular Evolution* 54: 774-783.

Navajas M, Boursot P (2003). Nuclear ribosomal DNA monophyly versus mitochondrial DNA polyphyly in two closely related mite species: the influence of life history and molecular drive. *Proceedings of the Royal Society of London Series B* 270: S124-S127.

Nei M, Rooney AP (2005). Concerted and Birth-and-Death evolution of multigenes families. *Annual Review in Genetics* 39:121-52.

Pons J, Gillespie RG (2004). Evolution of satellite DNAs in a radiation of endemic Hawaiian spiders: does concerted evolution of highly repetitive sequences reflect evolutionary history? *Journal of Molecular Evolution* 59: 632-641.

Robles F, de la Herran R, Ludwig A, Ruiz Rejon C, Ruiz Rejon M, Garrido-Ramos MA (2004). Evolution of ancient satellite DNAs in sturgeon genomes. *Gene* 338: 133-142.

Ryu SH, Do YK, Hwang UW, Choe CP, Kim W (1999). Ribosomal DNA intergenic spacer of the swimming crab, *Charybdis japonica*. *Journal of Molecular Evolution* 49: 806-809.

Schufran KA, Peters DC, Webster JA (1997). Generation of clonal diversity by sexual reproduction in the greenbug, *Schizaphis graminum*. *Insect Molecular Biology* 6: 203-209.

Chapter 6. Other research activities carried out during the PhD course.

“More” on termites.

Papers:

Luchetti A, Bergamaschi S, Marini M, Mantovani B (2004). Mitochondrial DNA analysis of native European Isoptera: a comparison between *Reticulitermes* (Rhinotermitidae) and *Kalotermes* (Kalotermitidae) colonies from Italy and Balkans. *Redia* LXXXVII: 149-153.

Bergamaschi B, Dawes-Gromadzki T, Luchetti A, Mantovani B, Marini M (2004). Preliminary molecular analysis of Isoptera taxa from the Australian Northern Territory. *Redia* LXXXVII: 239-242.

Luchetti A (2005). Identification of a short interspersed repeat in *Reticulitermes lucifugus* (Isoptera Rhinotermitidae) genome. *DNA Sequence* 16: 304-307.

Symposia:

Luchetti A, Bergamaschi S, Marini M, Mantovani B (2005). Caratterizzazione molecolare di colonie di *Kalotermes flavicollis* (Isoptera Kalotermitidae) del Mediterraneo centro-orientale. XI° Convegno Nazionale A.I.S.A.S.P. Sezione Italiana I.U.S.S.I. - International Union for the Study of Social Insects, Firenze, 1-2-3- Febbraio.

Bergamaschi S, Dawes-Gromadzki T, Luchetti A, Mantovani B, Marini M (2005). Analisi filogenetica di isotteri australiani del Northern Territory tramite studio del gene mitocondriale 16S. XI° Convegno Nazionale A.I.S.A.S.P. Sezione Italiana I.U.S.S.I. - International Union for the Study of Social Insects, Firenze, 1-2-3- Febbraio.

Luchetti A, Mantovani B, Marini M (2006). *Reticulitermes* spp. and *Kalotermes flavicollis* (Isoptera) diversity in the Balkan Peninsula. 10° International Congress on the Zoogeography and Ecology of Greece and Adjacent Regions, 26-30 June 2006, Patras, Greece.

The Tunga sand fleas project.

Papers:

- Luchetti A, Mantovani B, Fioravanti L, Trentini M (2004). *Wolbachia* infection in the newly described Ecuadorian sand flea, *Tunga trimamillata*. *Experimental Parasitology* 108: 18-23.
- Luchetti A, Mantovani B, Pampiglione S, Trentini M (2005). Molecular characterisation of *Tunga trimamillata* and *T. penetrans* (Insecta, Siphonaptera, Tungidae): taxonomy and genetic variability. *Parasite* 12: 123-129.
- Luchetti A, Mantovani B, Trentini M (2005). Rapid identification of non-neosomic *Tunga penetrans* and *Tunga trimamillata* (Insecta Siphonaptera) specimens through PCR-RFLP method. *Bulletin of Insectology* 58: 15-18.
- Luchetti A, Mantovani B, Trentini M (2005). *Wolbachia* superinfection in an Ecuadorian sample of the sand-flea *Tunga penetrans*. *Bulletin of Insectology* 58: 93-94.
- Luchetti A, Trentini M, Pampiglione S, Fioravanti ML, Mantovani B (2007). Genetic variability of *Tunga penetrans* (Siphonaptera, Tungidae) sand fleas across South America and Africa. *Parasitology Research* 100: 593-598.

Symposia:

- Luchetti A, Trentini M, Pampiglione S, Fioravanti ML, Mantovani B (2004). New data on the molecular diversity and biology of *Tunga trimamillata* and *T. penetrans* (Siphonaptera, Tungidae). *Atti XXIII° Congresso Nazionale SolPa, Vietri sul Mare (Salerno) 9-12 Giugno, Parassitologia* 46 (Suppl. 1): 179.
- Luchetti A, Mantovani B, Trentini M (2005). Caratterizzazione e diagnostica molecolare di *Tunga penetrans* e *T. trimamillata* (Siphonaptera, Tungidae). *Atti XX° Congresso Nazionale Italiano di Entomologia, Perugia-Assisi, 13-18 Giugno.*
- Trentini M, Gustinelli A, Pampiglione S, Luchetti A, Fioravanti ML (2005). Osservazioni al SEM su *Tunga trimamillata* e *T. penetrans* (Siphonaptera,

Tungidae). Atti XX° Congresso Nazionale Italiano di Entomologia, Perugia-Assisi, 13-18 Giugno.

Luchetti A, Trentini M, Pampiglione S, Fioravanti ML, Mantovani B (2006). Genetic diversity of *Tunga penetrans* (Siphonaptera, Tungidae) across South America and Africa. Atti XXIV° Congresso Nazionale SolPa Messina, 21-24 giugno.

Trentini M, Gustinelli A, Pampiglione S, Luchetti A, Caffara M, Fioravanti ML (2006). SEM observations on *Tunga trimamillata* and *T. penetrans* (Insecta, Siphonaptera, Tungidae). 11th International Congress of Parasitology, 6th-11th August 2006 Secc, Glasgow, Scotland.

And the Diptera ...

Papers:

Masetti A, Luchetti A, Sommaggio D, Burgio G, Mantovani B (2006). Phylogeny of *Chrysotoxum* species (Diptera, Syrphidae) inferred from morphological and molecular characters. European Journal of Entomology 103: 459-467.

Masetti A, Luchetti A, Mantovani B, Burgio G (2006). PCR-RFLP assays to distinguish *Liriomyza huidobrensis* (Diptera: Agromyzidae) from associated species on lettuce cropping systems in Italy. Journal of Economic Entomology 99: 1268-1272.

Symposia:

Masetti A, Luchetti A, Mantovani B, Burgio G (2005). Analisi delle relazioni filogenetiche tra le specie polifaghe del genere *Liriomyza* (Diptera Agromyzidae) tramite marcatori molecolari mitocondriali. Atti XX° Congresso Nazionale Italiano di Entomologia, Perugia-Assisi, 13-18 Giugno.