Alma Mater Studiorum – Università di Bologna

DOTTORATO DI RICERCA IN

Scienze e Tecnologie Agrarie Ambientali e Alimentari

Ciclo XXIX

Settore Concorsuale di afferenza: 07/G1 SCIENZE E TECNOLOGIE ANIMALI

Settore Scientifico disciplinare: AGR/17 ZOOTECNICA GENERALE E MIGLIORAMENTO GENETICO

TITOLO TESI

Towards the development of nutrigenetic concepts in pigs by merging genomics, transcriptomics and metabolomics

Presentata da: ANISA RIBANI

Coordinatore Dottorato

Prof. Giovanni Dinelli

Relatore

Prof. Luca Fontanesi

Esame finale anno 2017

Abstract

In a context of system biology concepts applied to livestock, animal nutrition can be considered as the netwotk interconnecting metabolic, physiologic and genetic aspects. The aim of this Thesis was to detect markers associated with productive traits and metabolic pathways in pigs, merging analyses on pig genome, transcriptome and metabolome. Several approaches of target metabolomics, target re-sequencing of pig genome portions and RNA-seq have been performed, adding classical lab validations. Analysis of the variability in pig genes related to metabolism like bitter taste receptors genes (*TAS2R*), fatty acids receptors genes (*GPR120*), *KMO* and others have been carried out in different pig populations, including commercial breeds (Large White, Duroc, Landrace, Pietran, Meishan) and Italian local pig breeds (Mora Romagnola, Nero Siciliano, Apulo-Calabrese, Casertana and Cinta Senese) as well as wild boars. Moreover, genome wide association analyses based on metabolites and, for most metabotypes, significant SNPs were close or within genes directly involved in the catabolic or anabolic pathways of the targeted metabolites. A few of these markers were associated (P nominal value <0.01) with production and carcass traits. According to our results, the development of precision feeding strategies focused on specific amino acid needs of the animals according to their genotype in genes involved in the amino acid metabolism pathways would be one of the envisaged perspectives of application in pig nutrigenetics and, more generally, in livestock nutrition.

Contents

1. General Introduction

Nutrition has always been one of the most important issue related to human health and welfare. As already Hippocrates stated long time ago ("Let food be thy medicine and medicine be thy food", Hippocrates, 431 B.C.), there is a bond, a very close connection, between nutrition and physiology not only in humans, but more in general in all the living organisms. In fact, food components introduced in a organism through diet could affect not only its metabolism but also its immune system response, its gut microbiota or its cell proliferation and signalling, preventing in some cases the development of diseases like cancer (Pardee, 1974; Donaldson, 2004; Kau *et al.*, 2011).

In livestock, nutrition and diet are fundamental for animal health and welfare, and are strictly connected with the rearing methods and breedings. The choice of the right raw materials to feed animals could prevent the massive use of antibiotics and drugs for prophylaxis, increasing some disease resistances in several livestock animals (Stein, 2002; Patterson and Burkholder, 2003; Castillo *et al.*, 2004). Moreover, animal diet could also affect different phenotypes related to productive traits, such as animal growth, feed intake, reproductive performances, physiology conditions, litter size and so on (Romanov, 1999; Chagas *et al.*, 2007; Gonzàles-Alvarado *et al.*, 2007; Santos *et al.*, 2011).

Finally, nutrition is related to feed intake which is the centre of a network interconnecting multiple metabolic, physiologic and genetic aspects. This network of relationships can be included in the complex concept of system biology, called also integrated metabolism (Zeisel *et al.*, 2001; Nicholson *et al.*, 2004).

1.1. "Omics" Sciences: genomics, transcriptomics and metabolomics

The Central Dogma of Molecular Biology, enunciated by Crick in 1958, states that in living organism DNA carries the genetic information that is tranfered to RNA through a biological process called transcription and then the sequence information is translated in proteins following the genetic code thanks to the ribosomial process called translation (Crick, 1970). Proteins are the biomolecules responsible for the phenotype, including all the external observable characteristics and features of an individual. The role of proteins is also extended to organism physiology and metabolism because a substantial proportion of them are enzymes involved in all the biochemical pathways for synthesis and transformation of all biological molecules (Fairbanks and Andersen, 1999).

The understanding of the path of the information flow from genes to phenotype (proteins) is one of the main concern of modern molecular biology, considering all the steps included in this process of information transfer. In order to describe the global study and analysis of genes, transcripts (RNAs), proteins and metabolites (including all the biological molecules obtained from a whatever metabolic pathway) 4 popular terms have been widely estabilshed and accepted, ending with the Greek suffix "-ome" with the meaning of "complete": genomics, transcriptomics, proteomics and metabolomics (Mooser and Ordovas, 2003).

Genomics is considered the analysis of all the genes present in a genome of a given organism, describing their DNA sequence, structure in exons/introns and mapping on chromosomes. Genomics can also include functional and comparative approaches studying how genes interact with each other in order to define better the role of different genes (Tyers and Mann, 2003). Transcriptomics can be defined as the analysis of all the RNAs species (including mRNAs, non-coding RNAs and small RNAs) present in a tissue through cDNA sequencing or using oligonucleotide microarray technology (Ozsolak and Milos, 2011). This analysis level is very complemetary to genomics and allows to understand the role played by functional elements of a genome, their expression and their quantity in a particular tissue in a particular physiological condition of an organism (Wang *et al.*, 2009).

Proteomics and metabolomics can be considered the other facets of biological complexity in a living organism: even if they are strictly related because most of the times enzymes involved in a metabolic pathway lead to the production of several metabolites, proteomics deals with the characterization of all the proteins present in a particular tissue including the analysis of their distribution, abundance, post-translation modifications and functions (Mooser and Ordovas, 2003). Metabolomics, on the other hand, aims to analyze all the metabolites in a biological sample, which could include a large range of biomolecules such as proteins, small oligopeptides, aminoacids, glicoproteins, sugars, lipids, fatty acids, nucleotides and so on. These numerous analytes have very different physiological and chemical properties occurring at different abundance levels (Dettmer *et al.*, 2006) and can be defined as the biochemical substrates or products of enzymatic reactions (Fontanesi, 2016). However, obtaining a complete characterization of all the metabolites present in a complex biological organism in several conditions is still a tricky issue (Patti *et al.*, 2012).
Because proteome and metabolome are at the end of the "Omics" cascade, they can be considered as internal phenotypes, a sort of expressed (and detectable) profiles with morfological and functional relevance in an organism (Raamsdonk *et al.*, 2001).

Finally, an integrative analysis combining genomics, transcriptomics, proteomics and metabolomics would provide a complementary and global view of genome organization, cellular function and responses to enviroment of complex systems (Hawkins *et al.*, 2010).

1.2. Molecular approaches and biotechnologies behind "Omics" Sciences

During the last few years, the technological advances in biology has led to the comprehension of the emerging field of system biology. In particular, the possibility to investigate and analyze at the same time the whole genome, transcriptome, proteome and metabolome of an organism, in particular physiologic conditions or in response of a given stimulus, is fundamental for the better understanding of the biochemical and biological mechanisms in a complex system (Dettmer *et al.*, 2006). Due to their different biological features, DNA, RNA, proteins and metabolites can be detected using different technologies, according to the investigation. In general, DNA and RNA have been so far analyzed with DNA sequencing approaches or using array technologies (Ozsolak and Milos, 2011). On the other hand, comprehensive metabolomics analyses require several approaches based on mass spectometry, due to the large variety of physical and chemical properties of metabolites (Bino *et al.*, 2004). Another strategy used in metabolomic investigations is the Nucleic Magnetic Resonance (NMR) spectroscopy, used mainly for high-throughtput metabolic profiles (Dettmer *et al.*, 2006). More detailed metabolomic technologies are explained in Paragraph 1.2.2.

1.2.1 Genomics and Transcriptomics: Next Generation Sequencing technologies

Recently, the development of high-throughput next generation DNA sequencing (NGS) technologies revoluzioned genomics and transcriptomics, including the possibility to obtain Whole Genome Sequencing (WGS) from a pool or from an individual or to obtain Gene Targeted Resequencing with very little costs (for example, the barrier of $1000 genome has been broken; van Dijk *et al.*, 2014). NGS can also detect some genomic rearrangements such as copy-number variants or gene-fusion events, allowing the better understanding of genome dynamics (Hawkins *et al.*, 2010). In trascriptomics, RNA-Seq method (RNA sequencing), based on cDNA libraries sequencing through reverse transcription, allows to both mapping and quantifying the population of RNA (total or fractioned) present in a given tissue or even cell (Beyer *et al.*, 2007; Wang *et al.*, 2009; Ozsolak and Milos, 2011). RNA-Seq can detect alternative splice variants and includes also small RNAs sequencing which are involved in several trascriptional and post-trascriptional regulation mechanisms. NGS platforms that have been mainly used over the past 10 years are 454 (Roche),

Illumina, SOLiD (Thermo Fisher Scientific), Ion Torrent (Thermo Fisher Scientific) and Pacific Biosciences (PacBio). These platforms are based on different NGS methods: while Illumina, SOLiD and 454 use optical detection of incorporated nucleotides using fluorescence and camera scanning, Ion Torrent does not require any imaging technology because it is based on a semiconductor sequencing in which a proton is released during nucleotide incorporation and it is subsequently detected by ion sensors. Sequencing platforms described above generate short reads, in a range from 35 to 400 bp excepted the 454 instruments which can generate reads of maximum 1 kb. PacBio platforms, on the other hand, provide extremely long reads of 20 kb and more and it is considered the ideal tool to finish *de novo* genome assemblies or improve genome drafts (van Dijk *et al.*, 2014).

1.2.2. Metabolomics data analysis and application in livestock

In livestock, metabolomics has alreadly contribute to add more information on the physiologic and metabolic state of animals detecing and quantifying hundreds of metabolites in a biological sample (e.g. serum, plasma, milk and so on). Metabotypes constitute internal or molecular phenotypes which are defined by the amount of metabolites in a sample and can be influenced by genetical and environmental factors (Fontanesi, 2016). Application of metabolomics to livestock and animal breeding can refine and improve trait descriptions, characterizing more intermediate phenotypes related to productive traits such as growth rate, fat deposition, milk production and so on.
In general, as described by Junot and colleagues (2014), metabolites can be classified in endogenous metabolites and xenobiotics; the first ones are produced directly by the biochemical processes of the organism while the second ones are all the chemical compounds present in the biological sample derived from external molecules introduced and, in some cases, transformed by the organism.

Because of the heterogeneity and the instability of the molecular species constituiting metabolites, there is no a technology able to detect precisely the whole metabolome (Fontanesi, 2016). However, the analytica platforms mainly used for metabolomics are based on mass spectrometry (MS), high-performace liquid-phase chromatography (HPLC) and nuclear magnetic resonance (NMR) spectrometry (Dettmer *et al.*, 2006). Approaches such as liquid-cromatorgaphy/mass-spectrometry (LC/MS) an NMR are considered untargeted because they detect as many metabolites as possible through different peaks (MS) or spectra (NMR) but not always they are able to characterize all of them in terms of metabolites structure and require appropriate and very complex statistical methods with multivariate approaches (Alonso *et al.*, 2015; Yi *et al.*, 2016). On the other hand, the targeted metabolomics approaches like gas-cromatography/mass-spectrometry (GC/MS) and flow injection

analysis MS/MS are based on preselected metabolites which are considered enough informative to describe biological and physiological conditions of an organism. Therefore, in targeted approaches, specific known metabolites are profiled and quantified with precision (Yi *et al.*, 2016).

Metabolomics data analyses require an *a priori* biological pathway and gene network analyses, for which there are some dedicated databases, for example the Kyoto Encyclopedia of Genes and Genomes (KEGG; Kanehisha *et al.*, 2014). Generally speaking, raw metabolomics data must be processed to reduce complexity and enhance metabolically significant pathways in order to obtain practicable data matrices. According to Yi and colleagues (2016), extracting information from metabolomics data requires 4 steps: *a)* Pre-processing which includes noise filtering, peak detection and normalization steps; *b)* Identification of metabolites analysing accurately peaks and spectra, for which there are many tools and softwares according to the technology used for metabolomics analysis; *c)* Variable selection which aims to extract relevant metabolites among all the detected ones with essential roles in metabolism; *d)* Modeling of the data, the final step able to extract information from biological and physiological events occurring in the organism.

Biological pathways are fundamental in metabolomics analysis to group metabolites that are linked to each other because of subsequent enzymatic and biochemical reactions: for this reason, several methodologies have been developed to perform metabolite enrichment analysis (for example the web-based tool Metabolite Set Enrichment Analysis, MSEA; Xia and Wishart, 2010). Finally, network reconstruction methods based on system biology can help to better understand the complexity of metabolomics in livestock and to link metabolomics with genomics and transcriptomics, reaching a more complex and complete overview of the state or the condition of an animal (Fontanesi, 2016).

1.2.3. Merging Genomics and Metabolomics in livestock: mGWAS approach

In literature, only few studies have been published on the integration of metabolomics with livestock genomics (Fontanesi, 2016). In particular, genetic variants occurring in some key genes of metabolic pathways might affect the level of many metabolites in animal biofluids and tissues, leading to changes of their metabolomic profiles. These metabolites controlled and influenced by genetics are defined as metabotypes and several studies in humans showed that mutations in genes related to metabolic pathways are associated with the level of one or more linked metabolites (Suhre and Gieger, 2012). These metabolite-based genome-wide association studies (mGWAS) provide the possibility to analyze and detect the link between a genetic variant and a presence/absence or the level of a metabolite.

In livestock, mGWAS have been performed in some pig populations on plasma using a target metabolomics approach (Fontanesi *et al.*, 2014; 2015) and in Danish Holstein dairy cattles on milk using an untargeted metabolomics apporach (Buitenhuis *et al.*, 2013). Fontanesi and collaborators (2014 and 2015) demonstrated that the level of metabotypes derived from diet in circulating plasma of pigs were associated with the variability in some specific and linked genes, suggesting to open new opportunities for novel applications of metabotypes analysis in animal breeding, such as nutrigenomic approaches.

1.3 Nutrigenetics and Nutrigenomics

In the era of "Omic" sciences, in which it is possible to better investigate and link almost all the biological processes occurring in an organism, nutrition can be considered a true integrative science in this context (Mutch *et al.*, 2005). It is largely known that nutrients can interact with molecular mechanisms and modulate biochemical processes affecting organism's physiological and metabolic functions; in particular, as Garg and colleagues reported (2014), nutrients are considered the most influential environmental factors to which genomes are exposed. In fact, at the cellular level nutrients can act for example *a)* as direct ligands for transcription factor receptors; *b)* in signal transduction pathways and *c)* modifying both primary and secondary pathways of metabolism because concentrations of substrates can alter gene regulation or cell signalling (Garg *et al.*, 2014).

The scientific discipline studying relationships between the individual genetic variability and the diet is the nutrigenetics, a concept introduced by Brennan and Mulligan in 1975 studying the hypoglycemia in humans (Brennan and Mulligan, 1975). In general, nutrigenetics aims to understand how the genetic composition, in terms of SNPs or other variants, of an organism responds to diet or is associated with different responses to nutrients (Ordovas and Corella, 2004; Mutch *et al.*, 2005). On the other hand, nutrigenomics applies genomic concepts in nutritional research describing the influence of common dietary ingredients on the genome (in terms of gene expression, genome stability and epigenome alterations; Garg *et al.*, 2014) and allowing a better understanding of how nutrients affect metabolism and homeostasis (Ordovas and Corella 2004; Muller and Kersten, 2003). Thus, with nutrigenomics it will be possible to define personalized diet according to genotypes in humans but also in livestock, developing personalized nutrition or precision feeding strategies in order to prevent some diseases and improve feed intake (Fenech *et al.*, 2011; Simopoulos, 2010).

Nutrigenetics and nutrigenomics are different approaches aiming to investigate the interaction bewteen genetics and diet from two opposite points of view and can be applied successfully in livestock (Garg *et al.*, 2014).

1.4 Nutrigenetics, taste perception and feed intake in pigs

In general, nutrition and feed intake are strictly linked with feeding behaviour in both wild and domesticated animals. The chemosensory system is one of the most relevant factor determining food preferences and has evolved to allow animals to discriminate between foods in their environment (Kats and Dill, 1998; Li and Zhang, 2013). Nutritional chemosensing allows the understanding of the perception of nutrients related to genomic, metabolic, physiological mechanisms and body homeostasis; in livestock, nutritional chemosensing science has been unexplored so far comparing to what has been well documented in human and animal models (Clop *et al.*, 2016). In fact, the diet offered to farm animals is usually a single diet nutritionally balanced but with no possibilty for the livesock to choose food, assuming that animals can adapt their taste perception according to whatever diet.

Domesticated pigs belong to *Sus scrofa* species (Linneaus 1758) and originated from wild boars with two independent domestication events starting about 9000 years ago in the Middle-East Asia and in Central Asia (Groenen *et al.*, 2012; Rubin *et al.*, 2012); since then, pigs have spread worldwide becoming one of the most important livestock for meat production.

Even if pigs are mainly reared for the production of several high-quality meat and food products, their suitability as models for human biomedical research is becoming accepted in the last decade (Gandarillas and Bas, 2009; Bendixen *et al.*, 2010; Verma *et al.*, 2011). In fact, both human and pigs are omnivorous mammals and share physiological similarities related to anatomy of the gastrointestinal tract (GIT), for example the size of the organs or the use of colon as the main fermentation site of dietary components of plant origins (Roura *et al.*, 2016).

As occurred in all mammals, also in pigs nutrients introduced with diet are primarly perceived through the taste system in the oral cavity. There, food compounds are solubilised in saliva and detected by the taste system thanks to sensory cells of the taste buds in the tongue *papillae* (Matsuo, 2000); then, a depolarisation occurrs in the sensory cells after biochemical signalling and reach the brain through a dedicated neuronal network (Barretto *et al.*, 2015; Roura *et al.*, 2016). In general there are five basic tastes detected by mammals: sweet, salty, sour, umami and bitter (Bachmanov and Beauchamp, 2007; Chaudhari and Roper, 2010). While salty and sour tastes are detected by trans-membrane ion

channels, sweet, umami and bitter are sensed by G-protein coupled receptors (*GPCRs*) belonging to the taste receptors class genes (*TASRs*). Tastes such as sweet and umami are appetizing and characterize high energy foods containing sugars and amino acids, while bitter taste is unpleasant and evolved as a warning system to protect against the presence of toxic and dangerous compounds in food (Li and Zhang, 2013).

In livestock, including pigs, taste perception can influence production and performance traits promoting different feeding behaviours (for example the appetite and the reward circuits) affecting the energy balance and body homeostasis (Depoortere, 2014; Loper *et al.*, 2015; Ribani *et al.*, 2016; Clop *et al.*, 2016). Genes encoding for taste receptors (*TASRs)*, taste sensitivity and the related molecular mechanisms are in general similar between pigs and humans, especially for sweet, sour, umani, and fatty acid tastes (Roura *et al.*, 2016). On the other hand bitter perception is the most different between the two species, maybe because of the different environment in which they evolved and adapted (Li and Zhang, 2013). However, some components known to be bitter to humans like alkaloids such as caffeine or pharmaceuticals including antibiotics are avoided by pigs (Nelson and Sangeret, 1997; Danilova *et al.*, 1999).

In pigs, recent studies have repoterd associations between taste receptors genes (*TASR*) and some important productive traits such as growth and fat deposition (Fontanesi *et al.*, 2015; Ribani *et al.*, 2016; Clop *et al.*, 2016). In particular, variants in bitter taste receptors genes (TAS2R gene family) in particular in *TAS2R38* and *TAS2R39* genes show significant association with Back Fat Thickness (BFT), suggesting their involvment in fat deposition (Ribani *et al.*, 2016). Compared to humans, pigs have a lower number of *TAS2R* genes, indicating that probably pigs evolved a higher resilience to bitter dietary components (Groenen *et al.*, 2012).

Generally, pigs show preferences for sweeteners (Roura and Tedo, 2009; Ripken *et al.*, 2014) probably due to the endocrinological responses determining short-term feed intake (Houpt *et al.*, 1979). In addition, several studies demonstrate that pigs show preferences for umami tastants especially amino acids including also the limiting essential amino acids used as supplements in swine diets: Lysine, Methionine and Threonine (Ettle *et al.*, 2010; Roura *et al.*, 2011). Moreover, fats introduced by diet are appetizing for pigs and are the main sources of energy supply; they can have an influence in the carcass meat quality in pig production. Several long chain fatty acids receptors have been detected in pigs encoded by *GPR120*, *GPR40*, *GPR41*, *GPR43* and *GPR84* genes (De Jager *et al.*, 2013). In particular, variants in *GPR120* have been significantly associated with fat deposition and growth (Fontanesi *et al.*, 2015; Song *et al.*, 2015).

Finally, the understanding of how genes involved in the taste perception system interact with each other building a nutrient sensing regolatory network is crucial to develop nutrigenetic approaches in pig breeding and nutrition.

1.5 Aim

The aim of this Thesis was to detect markers associated with productive traits and feed intake in pigs, merging analyses on pig genome, transcriptome and metabolome. Several approaches of re-sequencing of pig genome portions and RNA-seq using different Next Generation Sequencing platforms have been performed, adding classical lab validations such as Sanger sequencing, PCR-RFLP, fragment analyses or Real-Time PCRs for gene expression analyses. Analysis of the variability in pig genes related to metabolism like bitter taste receptors genes (*TAS2R*), fatty acids receptors genes (*GPR120*), *KMO* and others have been carried out in different pig populations, including commercial breeds (Large White, Duroc, Landrace, Pietran, Meishan) and Italian local pig breeds (Mora Romagnola, Nero Siciliano, Apulo-Calabrese, Casertana and Cinta Senese) as well as wild boars.

In the first chapter we performed an association study beetween a nutrigenetics target gene, the fatty acid receptor *GPR120*, and the average growth rate in Italian Large White pigs, re-sequencing *GPR120* gene in different pig populations using the next generation sequencing Ion Torrent PGM platform. We identified 3 SNPs in the gene among the populations and significant differences of allele and genotype frequencies distribution associated with Average Daily Gain (ADG) estimated breeding value (EBV) trait. The second chapter show the study of the re-sequencing of the bitter taste receptors genes in different pig populations detecting SNPs and some mutations affecting protein conformation of some of the receptors. Morevore, we identified significant association with some variants and the Back Fat Thickness (BFT) estimated breeding value (EBV) trait in Italian Large White pigs. Finally, in the last chapter we investigated the functional interactions between pig genome and metabolome performing genome wide association studies based on 200 plasma metabolites and on the Illumina PorcineSNP60 BeadChip genotyping in two pig comercial breeds, Italian Large White and Italian Duroc. We added moreover a gene expression analysis as a nutrigenetic pilot study. Our result will be useful to integrate molecular phenotypes and genotyping for breeding purposes.

2. Chapters

2.1. Next generation semiconductor based-sequencing of a nutrigenetics target gene (*GPR120*) and association with average growth rate in Italian Large White pigs

Luca Fontanesi[1,2], Francesca Bertolini[1*], Emilio Scotti[1*], Giuseppina Schiavo[1], Michela Colombo[1], Paolo Trevisi[1], Anisa Ribani[1], Luca Buttazzoni[3], Vincenzo Russo[1], Stefania Dall'Olio[1]

[1] Department of Agricultural and Food Sciences (DISTAL), Division of Animal Sciences, University of Bologna, Viale Fanin 46, 40127 Bologna, Italy

[2] Centre for Genome Biology, University of Bologna, 40126 Bologna, Italy

[3] Consiglio per la Ricerca e la Sperimentazione in Agricoltura, Centro di Ricerca per la Produzione delle Carni e il Miglioramento Genetico, Via Salaria 31, 00015 Monterotondo, Roma, Italy

* These authors contributed equally to this work.

Running head: *GPR120* SNP in pigs and growth rate

Published in Animal Biotechnology 2015 *26*(2), 92-97

http://dx.doi.org/10.1080/10495398.2014.881369

ABSTRACT

The *GPR120* gene (also known as *FFAR4* or *O3FAR1*) encodes for a functional omega-3 fatty acid receptor/sensor that mediates potent insulin sensitizing effects by repressing macrophage-induced tissue inflammation. For its functional role, *GPR120* could be considered a potential target gene in animal nutrigenetics. In this work we resequenced the porcine *GPR120* gene by high throughput Ion Torrent semiconductor sequencing of amplified fragments obtained from 8 DNA pools derived, on the whole, from 153 pigs of different breeds/populations (two Italian Large White pools, Italian Duroc, Italian Landrace, Casertana, Pietrain, Meishan and wild boars). Three single nucleotide polymorphisms (SNPs), two synonymous substitutions and one in the putative 3'-untranslated region (g.114765469C>T), were identified and their allele frequencies were estimated by sequencing reads count. The g.114765469C>T SNP was also genotyped by PCR-RFLP confirming estimated frequency in Italian Large White pools. Then, this SNP was analyzed in two Italian Large White cohorts using a selective genotyping approach based on extreme and divergent pigs for back fat thickness (BFT) estimated breeding value (EBV) and average daily gain (ADG) EBV. Significant differences of allele and genotype frequencies distribution was observed between the extreme ADG-EBV groups ($P<0.001$) whereas this marker was not associated with BFT-EBV.

Key words: GPR120; Ion Torrent semiconductor sequencing; SNP; Association study; Heavy pigs

INTRODUCTION

Nutrigenetics applied to the livestock industry aims to improve feed efficiency, growth rate, and health status of animals and, in turn, several related economic traits. In order to maximize these objectives it is important to investigate the genetic factors involved in the biological mechanisms affecting nutrient-gene interactions. For example, several studies have already identified mechanisms by which the gut senses luminal nutrients and regulates homeostatic mechanisms of energy metabolism in response to feeding or fasting through activation of gut-brain networks [reviewed in (1)]. Specific taste receptors for different nutrients mediate sensing responses of specialized enteroendocrine cells of the intestinal epithelium that transmit signals by releasing gastro-intestinal regulatory peptides.

Free fatty acids (FFAs) are essential dietary nutrients that activate different G-protein-coupled receptors (GPCR) that function on the cell surface and play essential roles also in nutritional regulation. Among these receptors, FFAR2 (GPR43) and FFAR3 (GPR41) are activated by short-chain FFAs, whereas medium and long-chain FFAs activate FFAR1 and GPR120 [also known as FFAR4 or O3FAR1; (2)]. In particular, GPR120 is activated by saturated (C14-C18) and unsaturated FFAs (C16-C22) that induce a rise in cytosolic free Ca^{2+}, but do not promote cAMP production. Recently Oh et al. (3) demonstrated that GPR120 is a functional omega-3 fatty acid receptor/sensor and mediates potent insulin sensitizing and antidiabetic effects by repressing macrophage-induced tissue inflammation. This receptor is highly expressed in human and mouse intestinal tracts, adipose tissues, and macrophages (3, 4). Colombo et al. (5) confirmed that the *GPR120* gene is expressed also in the gastrointestinal tract of young pigs, with higher expression in the colon as compared to three different gastric sites and jejunum. This finding would support a major role of this receptor in the colon, where it might mediate incretin rise after stimulation with FFAs, as reported in mice (6). Moreover, *GPR120* transcription level in intestinal epithelial cells is significantly increased by bacteria categorized as either probiotics or bacteria capable of inducing anti-inflammatory effects.

These microorganisms can produce FFAs in the intestine by microbial fermentation, so providing a signaling pathway finally turning out in an anti-inflammatory effect (7).

The role of GPR120 was further investigated by Ichimura et al. (8) who showed that GPR120 knock-out mice fed a high-fat diet developed obesity, glucose intolerance, reduced insulin signalling, enhanced inflammation in adipose tissue, fatty liver with decreased adipocyte differentiation and lipogenesis and enhanced hepatic lipogenesis through a reduced production of lipid hormone C16:1n7 palmitoleate, while no effect was observed in control mice. In humans, exon sequencing of the *GPR120* gene revealed that a deleterious non synonymous mutation (p.R270H) inhibiting signalling activity increased the risk of obesity in European and Japanese populations (8,9).

Using a genome wide candidate gene approach coupled with genome wide association studies with anonymous single nucleotide polymorphisms (SNPs), we recently started a systematic investigation to identify genetic factors affecting fat deposition and growth efficiency in Italian heavy pigs (10-14). These two traits are final complex phenotypes related to feed consumption and energy metabolism that might be explained, at least in part, by genetic factors that govern gene-nutrient interactions.

Next generation sequencing platforms are changing the way in which sequencing experiments for SNP discovery are designed. In particular, the Ion Torrent semiconductor-based sequencing can be applied to different experimental designs and targets due to its scalability (15).

In this work we resequenced the porcine *GPR120* gene by high throughput semiconductor sequencing (15) to identify polymorphisms in the pig. Then, we genotyped an SNP in two Italian Large White cohorts to evaluate if this marker is associated with back fat thickness (BFT) and average daily gain (ADG).

MATERIALS AND METHODS

Animals and production traits

Two groups of Italian Large White pigs were used in the association studies. The first group was constituted by 560 two-generations unrelated gilts with extreme and divergent estimated breeding values (EBVs) for BFT (280 with the most negative BFT EBV and 280 with the most positive BFT EBV), selected among about 12,000 pigs individually performance tested at the Central Test Station of the National Pig Breeder Association (ANAS) for the sib-testing evaluation of candidate boars within the national selection program of the Italian Large White breed (12). Average and standard deviation of BFT EBV of the pigs in the negative and positive tails was: –9.40 ± 1.60 mm and +8.00 ± 5.95 mm, respectively.

The second group of Italian Large White pigs was constituted by 380 two-generation unrelated gilts with extreme and divergent ADG EBV, selected among the same population described above (190 with the most negative ADG EBV and 190 with the most positive ADG-EBV). Average and standard deviation values of ADG EBV of the pigs in the negative and positive tails were –30 ± 14 g and +81 ± 12 g, respectively. More details about these animals are reported in Fontanesi et al. (10, 14).

Estimated breeding values for these traits were predicted by a BLUP-multiple trait animal model including the fixed factors of batch, age at the beginning of test, date of slaughtering and inbreeding coefficient, besides the random factors of animal and litter. For BFT, also body weight at slaughter and age at slaughter were considered. Fifty Italian Large White pigs with the lowest BFT EBV and 50 with the highest BFT EBV from the first group were used for resequencing the *GPR120* gene. Other pigs randomly selected from different breeds (10 Italian Duroc, 10 Italian Landrace, 8 Pietrain, 10 Casertana, 5 Meishan and 10 Italian wild boars), for which no phenotypic data were available, were used for resequencing this gene.

Resequencing and identification of polymorphisms

Genomic DNA was extracted from blood, muscle or ear tissue using the Wizard® Genomic DNA Purification kit (Promega Corporation, Madison, WI, USA). Extracted DNA was quantified using a NanoPhotometer P-330 instrument (Implen GmbH, München, Germany) and pooled at equimolar concentration as it follows: one pool was constituted from the sub-sample of 50 Italian Large White pigs with the lowest BFT EBV, one pool was constituted from the 50 Italian Large White pigs with the highest BFT EBV, and one pool was constituted for each of the other breeds with no phenotypic information (Italian Duroc, Italian Landrace, Pietrain, Casertana, Meishan and Italian wild boars).

PCR primers were designed on the sequence of the porcine *GPR120* gene annotated in the Sscrofa10.2 gene (Ensembl accession no. ENSSSCG00000010478) to amplify all three coding exons including non coding transcribed and intronic regions (Table 1). PCR was carried out using the DNA pools prepared as described above and the Phusion® Hot Start Flex 2X Master Mix (Euroclone) using PCR conditions reported in Table 1. Amplified fragments were purified using ExoSAP-IT® (USB Corporation, Cleveland, Ohio, USA), pooled according to the breed/DNA pool of origin using equimolar DNA from each amplified product. Then, obtained pooled fragments were used for library preparation and sequencing with the Ion Torrent PGM (Life Technologies). Briefly, 200 ng of every pool of amplified products was enzymatically sheared, end repaired and ligated with different barcodes using the Ion Xpress™ Plus Fragment Library kit (Life Technologies) and the Ion Xpress™ Barcode Adapters 1-16 Kit (Life Technologies). Then, resulting DNA material was size selected using the e-gel system (Invitrogen, Carlsbad, CA, USA) and bands corresponding to 100 bp of inserts were collected and quantified by qPCR using a StepOnePlus™ Real-Time PCR System (Life Technologies). Barcoded fragments were pooled again with the same concentration, clonally amplified, purified and sequenced using the Ion One Touch™ 100 Template kit and the Ion PGM™ Sequencing kit with a Ion 316 chip (Life Technologies). Obtained sequencing reads for the *GPR120*

gene were filtered, trimmed, automatically assigned to the different barcodes, and analysed for the presence of differences between the reference sequence (ENSSSCG00000010478) and obtained sequences with the Ion Torrent suite v.2.2 (Life Technologies). Several other amplified fragments were also sequenced in the same chip and results will be reported elsewhere. Average coverage of the sequenced *GPR120* gene was 20,535 X.

SNP genotyping

An SNP identified by sequencing (g.114765469C>T) was genotyped by PCR-RFLP. Briefly, genomic DNA from individual samples was amplified by using primer pair ExSNP (Table 1) and the obtained DNA fragment (5 μL of PCR product) was digested overnight at 37 °C with 2 U of *Fsp*BI restriction enzyme (Thermo Scientific - Fermentas, Vilnius, Lithuania) in a total of 25 μL of reaction volume with 1X reaction buffer. Resulting DNA fragments were electrophoresed in TBE 1X 2.5% agarose gels. DNA bands were visualized with 1X GelRed Nucleic Acid Gel Stain (Biotium Inc., Hayward, CA, USA). Allele C resulted in an undigested fragment of 170 bp, whereas allele T was detected by the occurrence of two fragments of 126 and 44 bp (the latter not completely visible in the gel).

Data analysis

Estimated SNP allele frequencies in the different breeds and groups of pigs was obtained by counting alternative Ion Torrent filtered reads produced within different barcoded libraries and normalizing according to the number of sequenced chromosomes.

The miRNA_Targets server [http://mamsap.it.deakin.edu.au/~amitkuma/mirna_targetsnew/sequence.html (16)] that includes two target prediction algorithms (miRanda and RNAhybrid) was used to screen the putative 3'-UTR of

the porcine *GPR120* gene for possible microRNA target sites in the region of the g.114765469C>T SNP, using human and mouse microRNA databases. In addition, RNAhybrid web server (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/submission.html) was used to predict potential porcine microRNA target sites in the same *GPR120* gene region. Porcine microRNAs were obtained from Li et al. (17).

Chi square and Cochran-Armitage trend tests of significance of allele and genotype frequency differences for the g.114765469C>T *GPR120* SNP between the two extreme tails of genotyped pigs was calculated.

RESULTS AND DISCUSSION

Ion Torrent resequencing of 2570 bp of the porcine *GPR120* gene (including the whole coding region, 5'- and 3'-untranslated regions (UTR) and portions of adjacent 5'- or 3'-flanking and intronic regions; Table 1) in a total of 153 animals identified just 3 SNPs (g.114743623G>C; g.114764865G>A; and g.114765469C>T) suggesting that this gene is quite conserved in *Sus scrofa*. Two SNPs were synonymous substitutions in exon 1 (g.114743623G>C or c.42G>C) and exon 3 (g.114764865G>A or c.837G>A) already annotated in Ensembl Sscrofa10.2 version (rs335415655 and rs343994284, respectively). The third SNP (g.114765469C>T indicated also as rs327485208 in Ensembl Sscrofa10.2), was located in the putative 3'-UTR of the gene. Analysis of this region with a few microRNA target prediction tools suggested that this polymorphism could potentially have a functional role. Several potential binding sites for different microRNA could be altered by the g.114765469C>T SNP. For example, considering a threshold of minimum free energy of -20 for the microRNA-target duplex, a few target sites for microRNAs (miR-30b-3p, PC220-3p, miR-216-3p, and PC167-3p among a few others) were modified by the polymorphism in this putative 3'-UTR. However, these modifications did not substantially distinguish the putative effect of the alternative

alleles (data not shown). This preliminary *in silico* analysis should be further evaluated and supported demonstrating that i) this SNP is included in a transcribed 3'-UTR (it is located 355 downstream the stop codon), as the full 3'-UTR of this gene is not characterized yet; ii); microRNAs actually bind the predicted sites and alter gene expression.

Frequencies of the two alleles of the 3 *GPR120* SNPs, as estimated on the basis of Ion Torrent reads, are reported in Table 2. To our knowledge, this study has been the first to apply the Ion Torrent semiconductor sequencing platform to estimate allele frequencies in DNA pools. The g.114764865G>A SNP was polymorphic only in the Pietrain breed in which the reference allele (G) showed a frequency of 0.590. In Italian Duroc, Meishan and Italian wild boars all three SNPs were monomorphic for the reference allele. Estimates of allele frequencies of the g.114765469C>T SNP in the two Italian Large White sequenced pools were also evaluated by genotyping this mutation by PCR-RFLP in the same animals that constituted the pools. Allele frequency in the negative and positive BFT EBV pools was 0.370 and 0.480 respectively, very close to what was estimated on pooled DNA by high throughput sequencing (negative pool 0.324; positive pool 0.444). Small discrepancies between the two genotyping protocols might be due to technical errors in constructing the DNA pools to be amplified, related to DNA quantification or by pipetting errors, or derived by difference of DNA quality among the samples. Other errors might come from PCR amplification, allele preferential bias or from barcoding and emulsion PCR steps (18). Comparison among estimated allele frequencies of the different SNPs could provide an indication about linkage disequilibrium among markers. Based on a first raw evaluating of the sequencing results, g.114743623G>C and g.114765469C>T SNPs may not be in complete linkage disequilibrium in Italian Large White, whereas these two polymorphic sites might be in complete or almost complete linkage disequilibrium in Italian Landrace, Pietrain and Casertana, where the reference allele of both SNPs had very close frequency. The g.114764865G>A polymorphism creates additional haplotypes in the Pietrain breed.

The g.114765469C>T SNP, because of its potential functional effect, has been chosen for genotyping by PCR-RFLP the two Italian Large White cohorts with extreme and divergent EBVs for BFT or ADG. These traits are end phenotypes that are related to functional relevance of the *GPR120* gene already demonstrated in humans and mice (8). The genotyping of the Italian Large White cohort selected according to extreme BFT EBV completed the genotyping carried out to verify allele frequencies estimated by high throughput sequencing. Allele and genotype differences between extreme tails in the BFT and ADG populations are reported in Table 3. In the whole BFT EBV cohort, allele and genotype frequency differences were not statistically significant ($P > 0.10$) whereas allele and genotype distributions in the extreme tails for ADG EBV were highly significant ($P = 9.50E-05$ and $P<0.001$, respectively). The reference allele (C) was less frequent in the negative tail than in the positive one. In this tail there was a marked increase of genotype CC and a decrease of genotype TT. Heterozygous animals remained almost of the same frequency in the two extreme groups (Table 3). These genotype distributions in the two compared ADG tails and the highly significant Cochran-Armitage trend test ($P= 3.20E-04$) may indicate an additive effect of this marker on ADG.

A few QTL for ADG or related traits [http://www.animalgenome.org/cgi-bin/QTLdb/SS/index (19)] have been already reported on porcine chromosome 14 (SSC14) where this gene is located (between positions 114742975 and 114765758). In a genome wide association study for ADG that we recently carried out in Italian Large White pigs (14) an SNP on SSC14 at position 107939105 (ALGA0080306) was one of the most significant marker associated with this trait ($P=1.08E-07$). Additional SNPs closer to the *GPR120* had low P values (e.g. ASGA0065801, $P=4.47E-04$) suggesting that the region around this gene may contain QTL for growth in this heavy pig breed.

GPR120 is a 2 G protein coupled taste-active fatty acid receptor involved in various physiological homeostasis processes such as fat deposition, regulation of appetite and food preference (4,20). In humans, a deleterious missense mutation in the *GPR120* gene is associated with increased risk of obesity and *GPR120* knocked-out mice show reduced insulin signalling and enhanced

inflammation in adipose tissue and develop obesity when fed a high-fat diet (8). No missense mutation was identified in the porcine gene even if a large number of animals from different breeds was sequenced. In addition, the genotyped polymorphism was not associated with BFT in Italian Large White pigs indicating that *GPR120* might not play an important role in defining this obesity related phenotype in the analyzed cohort, even if results reported in Table 3 indicated a tendency towards suggestive differences of allele frequencies in the two extreme BFT EBV tails. GPR120 is expressed in the gastrointestinal tract of pigs (5) where it could be directly involved in fatty acid perception and feed uptake, and, indirectly, in growth rate and efficiency, potentially linking molecular functions with a final phenotype. Association between the g.114765469C>T SNP and ADG may add evidences towards a potential functional role of this SNP and, in general, of the GPR120 receptor in the biological mechanisms affecting this important production trait. Nutrigenetics approaches to define appropriate feeding practices and more efficient feedstuff in heavy pigs could consider variability in the *GPR120* gene worth of further investigation to evaluate host gene-feeding interactions. Additional studies will be needed to experimentally validate a potential functional role of g.114765469C>T SNP, in altering putative 3'-UTR regulatory regions (e.g. microRNA target sites) and confirm the results of the association study before considering this polymorphism in selection plans for the Italian Large White breed to improve ADG.

Acknowledgements

We thank ANAS for providing data and samples, Sara De Fanti (University of Bologna) for technical assistance, members of the Centre for Genome Biology for their support and Rita Casadio and Pier Luigi Martelli (Biocomputing Group) for their advises on data analysis. This work was funded by Italian MiPAAF (INNOVAGEN Project).

REFERENCES

1. Rosoamanana R, Darcel N, Fromentin G, Tomé D. Nutrient sensing and signalling by the gut. *Proc Nutr Society*. 2012;71:446–455.

2. Ichimura A, Hirasawa A, Hara T, Tsujimoto G. Free fatty acid receptors act as nutrient sensors to regulate energy homeostasis. *Prostaglandins Other Lipid Mediat.* 2009;89:82–88.

3. Oh DY, Talukdar S, Bae EJ, Imamura T, Morinaga H, Fan W, Li P, Lu WJ, Watkins SM, Olefsky JM. GPR120 is an omega-3 fatty acid receptor mediating potent anti-inflammatory and insulin-sensitizing effects. *Cell.* 2010;142:687-698.

4. Miyauchi S, Hirasawa A, Iga T, Liu N, Itsubo C, Sadakane K, Hara T, Tsujimoto G. Distribution and regulation of protein expression of the free fatty acid receptor GPR120. *Naunyn Schmiedebergs Arch Pharmacol.* 2009;379:427-434.

5. Colombo M, Trevisi P, Gandolfi G, Bosi P. Assessment of the presence of chemosensing receptors based on bitter and fat taste in the gastrointestinal tract of young pig. *J Anim Sci.* 2012;90: Suppl 4:128-130.

6. Hirasawa A, Tsumaya K, Awaji T, Katsuma S, Adachi T, Yamada M, Sugimoto Y, Miyazaki S, Tsujimoto G. Free fatty acids regulate gut incretin glucagon-like peptide-1 secretion through GPR120. Nat Med. 2005;11:90-94.

7. Fredborg M, Theil PK, Jensen BB, Purup S. G protein-coupled receptor120 (GPR120) transcription in intestinal epithelial cells is significantly affected by bacteria belonging to the Bacteroides, Proteobacteria, and Firmicutes phyla. *J Anim Sci.* 2012;90 Suppl 4:10-12.

8. Ichimura A, Hirasawa A, Poulain-Godefroy O, Bonnefond A, Hara T, Yengo L, Kimura I, Leloire A, Liu N, Iida K, Choquet H, Besnard P, Lecoeur C, Vivequin S, Ayukawa K, Takeuchi M, Ozawa K, Tauber M, Maffeis C, Morandi A, Buzzetti R, Elliott P, Pouta A, Jarvelin MR, Körner A, Kiess W, Pigeyre M, Caiazzo R, Van Hul W, Van Gaal L, Horber F, Balkau B, Lévy-Marchal

C, Rouskas K, Kouvatsi A, Hebebrand J, Hinney A, Scherag A, Pattou F, Meyre D, Koshimizu TA, Wolowczuk I, Tsujimoto G, Froguel P. Dysfunction of lipid sensor GPR120 leads to obesity in both mouse and human. Nature. 2012;483:350-354.

9. Waguri T, Goda T, Kasezawa N, Yamakawa-Kobayashi K. The combined effects of genetic variations in the GPR120 gene and dietary fat intake on obesity risk. *Biomed Res.* 2013;34:69-74.

10. Fontanesi L, Bertolini F, Dall'Olio S, Buttazzoni L, Gallo M, Russo V. Analysis of association between the *MUC4* g.8227C>G polymorphism and production traits in Italian heavy pigs using a selective genotyping approach. *Anim Biotechnol.* 2012;23:147-155.

11. Fontanesi L, Bertolini F, Scotti E, Trevisi P, Buttazzoni L, Dall'Olio S, Davoli R, Bosi P, Russo V. Polymorphisms in an obesity-related gene (*PCSK1*) are associated with fat deposition and production traits in Italian heavy pigs. *Animal*. 2012;**6:1913-1924**.

12. Fontanesi L, Galimberti G, Calò DG, Fronza R, Martelli PL, Scotti E, Colombo M, Schiavo G, Casadio R, Buttazzoni L, Russo V. Identification and association analysis of several hundred single nucleotide polymorphisms within candidate genes for back fat thickness in Italian Large White pigs using a selective genotyping approach. *J Anim Sci.* 2012;90:2450-2464.

13. Fontanesi L, Schiavo G, Galimberti G, Calò DG, Scotti E, Martelli PL, Buttazzoni L, Casadio R, Russo V. A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes. *BMC Genomics*. 2012;13:583.

14. Fontanesi L, Schiavo G, Galimberti G, Calò DG, Russo V. A genome wide association study for average daily gain in Italian Large White pigs. Submitted.

15. Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A,

Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J. An integrated semiconductor device enabling non-optical genome sequencing. Nature. 2011;475:348-352.

16. Kumar A, Wong AK, Tizard ML, Moore RJ, Lefèvre C. miRNA_Targets: a database for miRNA target predictions in coding and non-coding regions of mRNAs. *Genomics*. 2012;100:352-356.

17. Li M, Xia Y, Gu Y, Zhang K, Lang Q, Chen L, Guan J, Luo Z, Chen H, Li Y, Li Q, Li X, Jiang AA, Shuai S, Wang J, Zhu Q, Zhou X, Gao X, Li X. MicroRNAome of porcine pre- and postnatal development. *PLoS One*. 2010;5:e11541.

18. Chen X, Listman JB, Slack FJ, Gelernter J, Zhao H. Biases and errors on allele frequency estimation and disease association tests of next-generation sequencing of pooled samples. *Genet Epidemiol*. 2012;36:549-560.

19. Hu ZL, Park CA, Wu XL, Reecy JM. Animal QTLdb: an improved database tool for livestock animal QTL/association data dissemination in the post-genome era. *Nucleic Acids Res.* 2013;41(Database issue):D871-D879.

20. Miyauchi S, Hirasawa A, Ichimura A, Hara T, Tsujimoto G. New frontiers in gut nutrient sensor research: free fatty acid sensing in the gastrointestinal tract. *J Pharmacol Sci*. 2010;112:19-24.

Table 1. PCR primer used in this study

Primer pair name	**Primer forward and reverse (5'-3')**	**Amplified region (bp)**	**Annealing Temperature (°C)[a]**	**Use**
Ex1	TGTTTCCAGCTGGCACTTC TGGGTACGGTGACACGATTA	5'-flanking region, 5'-UTR, exon 1, part of intron 1 (912)	59	Resequencing (Ion Torrent)
Ex2	GAGGACAGAGTTGCCGAGTC TGTCCCTCAGTGTTGCAGAG	Part of intron 1, exon 2, part of intron 2 (925)	65	Resequencing (Ion Torrent)
Ex3	GAACCCCACCCTGATGTG GGGTGGGGAACTAGAGTGG	Part of intron 2, exon 3, part of 3'-UTR and part of the putative 3'-flanking region (850)	59	Resequencing (Ion Torrent)
Ex3SNP	GGTGTGTTAGTAATATGATTG GGGTGGGGAACTAGAGTGG	Putative 3'-UTR (170)	56	PCR-RFLP (*Fsp*BI)

[a] PCR profile for primer pairs Ex1, Ex2, and Ex3 was the following : 98 °C for 1 min, followed by 35 cycles with 98 °C for 10 sec, the appropriate annealing temperature for 30 sec, 72 °C for 40 sec, the final elongation step was at 72 °C for 8 min. PCR Profile for primer pair Ex3SNP was the following: 95 °C for 5 min, then 35 cycles at 95 °C for 30 sec, the reported annealing temperature for 30 sec, 72 °C for 30 sec, the final elongation step was for 5 min at 72 °C.

Table 2. Estimated allele frequencies of the three identified *GPR120* polymorphisms based on Ion Torrent reads count. Only the frequency of the first allele is reported.

Breeds	**No. of animals**	**g.114743623G>C**	**g.114764865G>A**	**g.114765469C>T**
Italian Large White (negative BFT EBV)	50	0.437	1.000	0.324
Italian Large White (positive BFT EBV)	50	0.590	1.000	0.444
Italian Landrace	10	0.585	1.000	0.588
Italian Duroc	10	1.000	1.000	1.000
Pietrain	8	0.482	0.590	0.486
Casertana	10	0.564	1.000	0.500
Meishan	5	1.000	1.000	1.000
Italian wild boars	10	1.000	1.000	1.000

Table 3. Differences of allele and genotype frequencies of the g.114765469C>T polymorphism between the two extreme and divergent tails chosen using a selective genotyping approach based on estimated breeding values for two production traits in Italian Large White pigs.

Trait[a]	**Tail**[b]	**No. of pigs**[b]	**Allele frequencies**			**Genotype frequencies**				
			C	**T**	***P***[c]	**CC**	**CT**	**TT**	**P**[d]	**P**[e]
BFT	negative	275	0.418	0.582	0.139	0.186	0.465	0.349	0.346	0.155
	positive	279	0.462	0.538		0.222	0.480	0.298		
ADG	negative	189	0.352	0.648	9.50E-05	0.127	0.450	0.423	6.59E-04	3.20E-04
	positive	189	0.492	0.508		0.249	0.487	0.264		

[a]BFT = back fat thickness estimated breeding value (EBV); ADG = average daily gain EBV.

[b]The two extreme and divergent tails for BFT EBV and ADG EBV.

[c]P value of the chi square test for allele frequency difference between the negative and positive tails.

[d]P value of the chi square test for genotype frequency difference between the negative and positive tails.

[e]P value of the Cochran-Armitage trend test for genotype frequency differences between the negative and positive tails.

2.2. Next generation semiconductor based sequencing of bitter taste receptor genes in different pig populations and association analysis using a selective DNA pool-seq approach

A. Ribani[1*], F. Bertolini[1*], G. Schiavo[1*], E. Scotti[1], V. J. Utzeri[1], S. Dall'Olio[1], P. Trevisi[1], P. Bosi[1], and L. Fontanesi[1]

[1]Department of Agricultural and Food Sciences (DISTAL), Division of Animal Sciences, University of Bologna, Viale Fanin 46, 40127 Bologna, Italy

* These authors contributed equally to this work

Running title: **Pool-seq for porcine bitter taste receptor genes**

Published in *Animal Genetics* (2016), doi: 10.1111/age.12472

Summary

Taste perception in animals affects feed intake and may influence production traits. In particular, bitter is sensed by receptors encoded by a family of *TAS2R* genes. In this research, using a DNA pool-seq approach, coupled with next generation semiconductor based target re-sequencing, we analysed 9 porcine *TAS2R* genes (*TAS2R1, TAS2R3, TAS2R4, TAS2R7, TAS2R9, TAS2R10, TAS2R16, TAS2R38* and *TAS2R39*) to identify variability and, at the same time, estimate single nucleotide polymorphism (SNP) allele frequencies in several populations and testing differences in an association analysis. Equimolar DNA pools were prepared for five pig breeds (Italian Duroc, Italian Landrace, Pietrain, Meishan and Casertana) and wild boars (5-10 individuals each) and for two groups of Italian Large White pigs with extreme and divergent back fat thickness (BFT; 50 + 50 pigs). About 1.8 million of reads were obtained by sequencing amplicons generated from these pools. A total of 125 SNPs were identified, of which 37 were missense mutations. Three of them (p.I53F and p.W85L in T*AS2R4*; p.L37S in *TAS2R39*) could have important effects on the function of these bitter taste receptors, based on in silico predictions. Variability in wild boars seems lower than that in domestic breeds potentially as a result of selective pressure in the wild towards defensive bitter taste perception. Three SNPs in *TAS2R38* and *TAS2R39* were significantly associated with BFT. These results might be important to understand the complexity of taste perception and their associated effects that could be useful to develop nutrigenetic approaches in pig breeding and nutrition.

Keywords: Nutrigenetics**,** SNP, Ion Torrent, Back fat thickness, Taste perception, TAS2R38, TAS2R39

Text

Taste perception mainly affects preferences and consequently food/feed intake (Chaudhari & Roper 2010) that, in turn, can influence production and performance traits in livestock (Patience *et al.* 2015). The mammalian chemosensory system usually discriminates five major basic taste classes: salty, sour, sweet, umami and bitter (Kinnamon & Cummings 1992; Lindemann 1996; Drayna 2005; Behrens & Meyerhof 2009; Chaudhari & Roper 2010). Among these taste classes, bitter perception, that evolved as a central warning system to protect against ingesting potentially toxic bitter-tasting substances (Li & Zhang 2014), could be particularly relevant to design appropriate feeding strategies in livestock dealing with bitter compounds present in feedstuff (e.g. Roura & Tedò 2009; Solà-Oriol *et al.* 2014; Lombardi *et al.* 2015). Bitter is sensed by a family of bitter taste receptors (referred as TAS2Rs) that are seven-transmembrane receptors encoded by a family of approximately 10-40 functional *TAS2R* genes in mammalian genomes (Chandrashekar *et al.* 2000; Wu *et al.* 2005; Bachmanov & Beauchamp 2007). Polymorphisms in *TAS2R* genes are associated with different taste responses to several compounds, both natural and synthetic (Bufe *et al.* 2002; Kim *et al.* 2003; Kim & Drayna 2005; Hayes *et al.* 2011). Bitter taste receptors also influence glucose homeostasis and regulate thyroid function (Dotson *et al.* 2008; Clark *et al.* 2015) and a few variants in these genes are associated with body weight and obesity in humans (Goldstein *et al.* 2005; Keller *et al.* 2010, 2014).

Few studies have investigated the structure, expression and variability of *TAS2R* genes in pigs (Colombo *et al.* 2012; Groenen *et al.* 2012; da Silva *et al.* 2014). Da Silva *et al.* (2014), by querying the pig genome, identified 15 *TAS2Rs* genes (11 annotated as genes and 4 described as pseudogenes), in addition to other taste receptor genes. Subsequently, they analysed the variation in these genes by comparing whole re-sequencing data from 79 pigs of different breeds and provided evolutionary and phylogeographical relationships in comparison with non-bitter taste genes (da Silva *et al.* 2014). As far as we know, no association studies have been carried out thus far using variants in these porcine genes.

We recently evaluated the use of next generation semiconductor based sequencing (i.e. Ion

Torrent technology; Rothberg *et al.* 2011) for partial and whole genome re-sequencing and single nucleotide polymorphism (SNP) discovery in a few livestock and aquaculture species (Bertolini *et al.* 2014, 2015, 2016; Bovo *et al.* 2015). In addition, we used this sequencing platform for target resequencing combining, at the same time, SNP discovery and association analyses using a selective DNA pool-seq approach (Fontanesi *et al.* 2015). Allele frequency estimation in DNA pools constructed from divergent groups of animals has been commonly used in different livestock species to evaluate association between DNA markers and productive traits (e.g. Darvasi & Soller 1994). In this work we applied the DNA pool-seq approach previously described (Fontanesi *et al.* 2015) to re-sequence nine *TAS2R* genes (*TAS2R1, TAS2R3, TAS2R4, TAS2R7, TAS2R9, TAS2R10, TAS2R16, TAS2R38* and *TAS2R39*), three of them (*TAS2R1, TAS2R3* and *TAS2R4*) not previously investigated by others (da Silva *et al.* 2014; Table 1), as they were not annotated or assembled in any Sscrofa10.2 chromosomes or they were considered pseudogenes. Resequencing was carried out in six pig breeds and wild boars for polymorphism identification and association analysis with back fat thickness (BFT). A fat deposition trait was included in the association analysis, as *TAS2R* gene variants are associated with human obesity and related traits, even if the molecular mechanisms involving TAS2R in fat deposition are not understood yet (Tepper & Ullrich 2002; Goldstein *et al.* 2005; Keller *et al.* 2010, 2014).

The porcine *TAS2R* genes indicated above (and identified by BLAST analysis with the corresponding genes in humans and mice; Table 1) were amplified from DNA pools using primer pairs reported in Table S1. DNA pools were prepared with equimolar quantity of DNA from each of five pig breeds (Italian Duroc, n. of pigs in the pool = 10; Italian Landrace, n. = 10; Pietrain, n. = 8; Casertana, n.= 10 and Meishan, n. = 5) and wild boars (n. = 10) to obtain allele frequencies in different populations. Moreover, two DNA pools (50 + 50 animals) were prepared from Italian Large White pigs (all gilts) with extreme and divergent estimated breeding value (EBV) for BFT. EBV were calculated as previously described (Fontanesi *et al.* 2012a, 2012b). Average and standard deviation of BFT EBV of the pigs in the negative and positive tails were: –9.40 ± 1.60 mm and +8.00 ± 5.95

mm, respectively. These pigs were selected from ~12,000 pigs individually performance tested at the Central Test Station of the National Pig Breeder Association (ANAS) for the sib-testing evaluation of candidate boars within the national selection program of the Italian Large White breed (Fontanesi *et al.* 2012a, 2012b, 2015). PCR was carried out from DNA pools (Table S1) and all amplicons obtained from the same breed/DNA pool were in turn pooled for library preparation and sequencing with the Ion Torrent PGM (Life Technologies) using different barcodes (Table S1). Barcoded fragments were pooled again at the same concentration, clonally amplified, purified and sequenced using the Ion One Touch™ 100 Template kit and the Ion PGM™ Sequencing kit with an Ion 316 chip (Life Technologies). The sequencing reads generated were filtered, trimmed, automatically assigned to the different barcodes and aligned to the reference sequences using the Ion Torrent Suite 2.2 (Life Technologies), that includes TMAP aligner (https://github.com/iontorrent/TMAP). Average depth and coverage for the sequenced regions is reported in Table S2. Differences in coverage might be due to different efficiencies in the construction of the barcoded sub-libraries or to other technical issues derived by the starting DNA quantity used for each amplicons (see notes to Table S1 about the Ion Torrent sequencing protocols). A total of 1,821,843 reads successfully aligned to the targeted regions with a mean depth of 2,711 X. The filtered *bam* files were deposited in the EMBL-EBI European Nucleotide Archive (ENA) with the project accession number PRJEB11635. Polymorphism detection was carried out using *mpileup* function of SAMTOOLS (Li *et al.* 2009). Then, only SNPs with quality scores ≥20 were retained for further analyses. Sanger sequencing was carried out to confirm Ion Torrent sequencing data on a few fragments (Table S1). A combination of *mpileup* counts with SAMTOOLS and Python scripts was then used to determine the number reads with the reference and with the alternative nucleotide at each SNP site, that was weighted according to the number of sequenced chromosomes for allele frequency estimation (Table S3). Five SNPs were genotyped by PCR-RFLP on individual pigs included in the DNA pools to validate allele frequency estimates as determined by read count (Tables S1 and S4), confirming these values as previously demonstrated (Fontanesi *et al.* 2015). Amplicons obtained for the same gene were partially

overlapped. Overlapping amplified regions reported the presence of 6 SNPs in both amplicons (Table S3) further confirming that the SNP calling pipeline was set up correctly and was able to eliminate false positives that might be produced by the Ion Torrent sequencing technology. The effect of each SNP was evaluated using VEP (Variant Effect Predictor) and the effect of all missense mutations was analysed using SIFT (Kumar et *al.* 2009).

Sequence data we obtained with Sanger and Ion Torrent indicated that the porcine *TAS2R1* is not a pseudogene (EMBL accession no. LT221026), in contrast to the previous analyses (da Silva *et al.* 2014) and the current annotation in Sscrofa10.2 (Ensembl release n. 84, March 2016). Figures 1a and 1b give an overview of the distribution of the SNPs in the investigated pig populations and in the analysed *TAS2R* genes. Detailed information is reported in Table S3. A total of 125 SNPs were identified in at least one population (Table 1, Fig. 1 and Fig. 2). Among these SNPs, 70 were already present in dbSNP (http://www.ncbi.nlm.nih.gov/SNP/) build 146 (November 24th, 2015), 53 of which were also reported by da Silva *et al.* (2014), whereas 55 were novel (Table S3; dbSNP ID: from ss1971458462 to ss1971458516). *TAS2R3* (n. of SNPs = 26), *TAS2R38* (n. = 23) and *TAS2R1* and *TASR29* (n. = 22) genes had the largest number of detected SNPs among the investigated genes (Table 1). Among the 125 identified polymorphisms, 66 occurred in flanking or untranslated regions and 59 were located in coding regions, 22 of which were synonymous and 37 were missense mutations. These latter variants might provide the most interesting polymorphisms, considering the functional roles of different protein domains in these receptors related to bitter molecule sensing and chemo-signal transmission (Fig. 3). Ten missense mutations (the largest number) were identified in *TAS2R39*, one of them (p.L37S, located in the first intracellular domain; polymorphic in wild boars, Italian Large White, Casertana and Pietrain breeds) was predicted to significantly modify the protein function (SIFT $P < 0.05$). Two other missense mutations in *TAS2R4* (p.I53F and p.W85L, located in the second and third transmembrane domains, respectively, and polymorphic in Italian Landrace, Italian Duroc and Meishan pigs) were predicted to be deleterious (SIFT $P< 0.05$).

Cluster analyses of the heat map based on SNP allele frequency distribution in the analysed

breeds confirmed the different origin of the Chinese Meishan pigs compared to all other European breeds/populations, as already well established (Fig. 1b). The largest number of SNPs was identified in Italian Duroc pigs (n. = 70) and the lowest in wild boars (n. = 43). According to these results, it seems that the domestication process might have relaxed the variability in *TAS2R* genes compared to wild boars (the number of SNPs identified in domestic breeds and in wild boars differed at $P<0.10$) in which the exposure to natural potential toxic feeds could have contributed to the selection and fixation of some of the bitter taste receptor variants. However, more detailed population genetic analyses should be carried out to confirm this hypothesis as we investigated a small number of wild boars.

A total of 13 SNPs (7 in *TAS2R39*, 3 in *TAS2R38* and one in *TAS2R4*, *TAS2R10* and *TAS2R16*) were significantly associated ($P_{nominal\ value}<0.05$; Chi square tests) with BFT, by comparing estimated allele frequencies in the two extreme and divergent groups of Italian Large White pigs (Table S5). Considering that SNPs in the same genes might be in linkage disequilibrium and the tests are not independent (a few genes are on the same chromosome; Table 1), correction for multiple testing was obtained using the Proportion of False Positives (PFP; Fernando *et al.* 2004), as described in Fontanesi *et al.* (2012). One of the analysed SNPs in *TAS2R39* (a missense mutation: p.N71T; rs342835508) had a $P_{PFP}<0.05$ and two additional SNPs, one in the same *TAS2R39* (rs326928677, located in the 5'-flanking or untranslated region) and one in *TAS2R38* (a missense mutation not reported before: p.I277M) had a $P_{PFP}<0.10$. As *TAS2R38* and *TAS2R39* are both on SSC18, it might be possible that their potential effect on BFT could be due to linkage disequilibrium with other QTL segregating on this chromosome region. However, our previous genome wide association studies on BFT in the Italian Large White pig breed did not identify any significant signals in the regions in which these genes are located (Fontanesi *et al.* 2012b; Bovo *et al.* 2015) even if several QTLs for fat deposition traits have been reported in this region by other studies (e.g. Hu *et al.* 2016). These two associated *TAS2R39* SNPs might capture two major haplotypes segregating in Italian Large White pigs for this gene, as it could be also deduced in part by the signals of association (even if at a $P_{nominal}$

$_{value}$<0.05) that come from other 5 polymorphisms of the same gene (Table S5). To further validate the results obtained for *TAS2R39*, the most significant SNP (p.N71T) was also genotyped by PCR-RFLP in a larger number of pigs (75 + 75 animals with extreme and divergent EBV for BFT; all gilts with BFT EBV < - 8.5 and all gilts with BFT EBV > +5.5 mm recruited from a population of about 12,000 performance tested pigs; Fontanesi *et al.* 2012a, 2012b, 2015). Results of individual genotyping confirmed the effect already observed ($P_{nominal\ value}$ <0.01). The human bitter taste receptor *hTAS2R39* has been recently characterized as a receptor for several bitter compounds, e.g. flavonoids (Roland *et al.* 2014; Yamazaki *et al.* 2014) but no study investigated its association with obesity or fat deposition traits in human or other species yet. TAS2R38 is a major bitter compound receptor in humans whose variability is strongly associated with the perception of some bitter synthetic compounds such as phenylthiocarbamide (PTC) and the related 6-n-propylthiouracil (PROP; Suo & Reed 2001). Variability in PROP sensitivity and variants in the human *TAS2R38* were the subject of controversial association studies with body mass index (Tepper & Ullrich 2002; Goldstein *et al.* 2005; Tepper *et al.* 2008; Sausenthaler *et al.* 2009). From these studies, it is not clear if there is a direct or indirect involvement of variability in this gene in the final effects of food ingestion, i.e. body weight. In our study in pigs, as the animals were from a performance test station, there should not be any difference in terms of diet components given to the animals. It will be interesting to further evaluate if the significant effect on BFT might be indirectly due to a different feed ingestion rate (due to different feed taste perception) or due to other factors that might contribute to understand the physiological effects of *TAS2R38* variants in pigs. Association studies with feed conversion rate and daily gain will be carried out to dissect, at least in part, the effects on related traits routinely measured, as association studies on feed preference in livestock might be difficult to carry out for a large number of animals needed to obtain meaningful results.

In conclusion, this study, using a methodological approach based on next generation sequencing, was able to obtain in a single experiment multiple levels of results. The identified variability in the investigated family of genes (i.e. bitter taste receptors) might be important to

understand the complexity of taste perception (Roundnitzky *et al.* 2015) and their associated effects on production and performance traits. Further studies are needed to develop nutrigenetic approaches starting from the obtained results and integrate them in pig breeding and nutrition.

Acknowledgements

We would like to thank ANAS to provide samples and data and Dr. Sara De Fanti (University of Bologna) for technical support. We thank Prof. Vincenzo Russo (University of Bologna) for his support at the beginning of this study. This study has been funded by MiPAAF Innovagen Project.

References

Bachmanov A.A. & Beauchamp G.K. (2007) Taste receptor genes. *Annual Review of Nutrition* **27**, 389-414.

Behrens M. & Meyerhof W. (2009) Mammalian bitter taste perception. *Results and Problems in Cell Differentiation* **47**, 1-18.

Bertolini F., Schiavo G., Scotti E., Ribani A., Martelli P.L., Casadio R. & Fontanesi L. (2014) High-throughput SNP discovery in the rabbit (*Oryctolagus cuniculus*) genome by next-generation semiconductor-based sequencing. *Animal Genetics* **45**, 304-7.

Bertolini F., Scimone C., Geraci C., Schiavo G., Utzeri V.J., Chiofalo V. & Fontanesi L. (2015) Next generation semiconductor based sequencing of the donkey (*Equus asinus*) genome provided comparative sequence data against the horse genome and a few millions of single nucleotide polymorphisms. *PLoS One* **10**, e0131925.

Bertolini F., Geraci C., Schiavo G., Sardina M.T., Chiofalo V. & Fontanesi L. (2016) Whole genome semiconductor based sequencing of farmed European sea bass (*Dicentrarchus labrax*) Mediterranean genetic stocks using a DNA pooling approach. *Marine Genomics*, doi: 10.1016/j.margen.2016.03.007.

Bovo S., Bertolini F., Schiavo G., Mazzoni G., Dall'Olio S. & Fontanesi L. (2015) Reduced representation libraries from DNA pools analysed with next generation semiconductor based-sequencing to identify SNPs in extreme and divergent pigs for back fat thickness. *International Journal of Genomics* **2015**, 950737.

Bufe B., Hofmann T., Krautwurst D., Raguse J.D. & Meyerhof W. *(2002)* The human TAS2R16 receptor mediates bitter taste in response to beta-glucopyranosides. *Nature Genetics* **32**, 397-401.

Chandrashekar J., Mueller K.L., Hoon M. A., *et al.* (2000) T2Rs function as bitter taste receptors. *Cell* **100**, 703-11.

Chaudhari N. & Roper S.D. (2010) The cell biology of taste. *Journal of Cell Biology* **190**, 285-96.

Clark A.A., Dotson C.D., Elson A.E., *et al.* (2015) TAS2R bitter taste receptors regulate thyroid function. *FASEB Journal* **29**, 164-72.

Colombo M., Trevisi P., Gandolfi G. & Bosi P. (2012) Assessment of the presence of chemosensing receptors based on bitter and fat taste in the gastrointestinal tract of young pig. *Journal of Animal Science* 90, 128-30.

da Silva E. C., de Jager N., Burgos-Paz W., Reverter A., Perez-Enciso M. & Roura E. (2014) Characterization of the porcine nutrient and taste receptor gene repertoire in domestic and wild populations across the globe. *BMC Genomics* **15**, 1057.

Darvasi A. & Soller M. (1994) Selective DNA pooling for determination of linkage between a molecular marker and a quantitative trait locus. *Genetics* **138**, 1365-73.

Dotson C.D., Zhang L., Xu H., *et al.* (2008) Bitter taste receptors influence glucose homeostasis. *PLoS One* **3**, e3974.

Drayna D. (2005) Human taste genetics. *Annual Review of Genomics and Human Genetics* **6**, 217–35.

Fernando R.L., Nettleton D., Southey B.R., Dekkers J.C., Rothschild M.F. & Soller M. (2004) Controlling the proportion of false positives in multiple dependent tests. *Genetics* **166**, 611-9.

Fontanesi L., Galimberti G., Calò D.G., *et al.* (2012a) Identification and association analysis of several hundred single nucleotide polymorphisms within candidate genes for back fat thickness in Italian Large White pigs using a selective genotyping approach. *Journal of Animal Science* **90**, 2450-64.

Fontanesi L., Schiavo G., Galimberti G., *et al.* (2012b) A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes. *BMC Genomics* **13**, 583.

Fontanesi L., Bertolini F., Scotti E., *et al.* (2015) Next generation semiconductor based-sequencing of a nutrigenetics target gene (*GPR120*) and association with growth rate in Italian Large White pigs. *Animal Biotechnology* **26**, 92-7.

Groenen M.A., Archibald A.L., Uenishi H., *et al.* (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**, 393-8.

Goldstein G.L.,Daun H. & Tepper B.J.(2005) Adiposity in middle-aged women is associated with genetic taste blindness to 6-n-propylthiouracil. *Obesity Research* **13**, 1017–23.

Hayes J.E., Wallace M.R., Knopik V.S., Herbstman D.M., Bartoshuk L.M. & Duffy V.B. (2011) Allelic variation in TAS2R bitter receptor genes associates with variation in sensations from and ingestive behaviors toward common bitter beverages in adults. *Chemical Senses* **36**, 311-9.

Hu Z.L., Park C.A. & Reecy J.M. (2016) Developmental progress and current status of the Animal QTLdb. *Nucleic Acids Research* **44**, D827-33.

Keller K.L., Olsen A., Cravener T.L., *et al.* (2014) Bitter taste phenotype and body weight predict children's selection of sweet and savory foods at a palatable test-meal. *Appetite* **77**, 113-21.

Keller K.L., Reid A., MacDougall M.C., *et al.* (2010) Sex differences in the effects of inherited bitter thiourea sensitivity on body weight in 4-6-year-old children. *Obesity* **18**, 1194-200.

Kim U.K. & Drayna D. (2005) Genetics of individual differences in bitter taste perception: lessons from the PTC gene. *Clinical Genetics* **67**, 275-80.

Kim U.K., Jorgenson E., Coon H., Leppert M., Risch N. & Drayna D. (2003) Positional cloning of the human quantitative trait locus underlying taste sensitivity to phenylthiocarbamide. *Science* **299**, 1221-5.

Kinnamon S.C. & Cummings T.A. (1992) Chemosensory transduction mechanisms in taste. *Annual Review of Physiology* **54**, 715–31.

Kumar P., Henikoff S. & Ng P.C. (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature Protocols* 4, 1073-81.

Li D. & Zhang J. (2014) Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. *Molecular Biology and Evolution* **31**, 303-9.

Li H., Handsaker B., Wysoker A., *et al.* (2009) 1000 Genome Project Data Processing Subgroup. The Sequence alignment/map (SAM) format and SAMtools. *Bioinformatics* **25**, 2078–9.

Lindemann B. (1996) Taste reception. *Physiological Reviews* **76**, 718-66.

Lombardi D., Vasseur E., Berthiaume R., DeVries T.J. & Bergeron R. (2015) Feeding preferences and voluntary feed intake of dairy cows: Effect of conservation and harvest time of birdsfoot trefoil and chicory. *Journal of Dairy Science* **98**, 7238-47.

Patience J.F., Rossoni-Serão M.C. & Gutiérrez N.A. (2015) A review of feed efficiency in swine: biology and application. *Journal of Animal Science and Biotechnology* **6**, 1-9.

Rothberg J.M., Hinz W., Rearick T.M., et al. (2011) An integrated semiconductor device enabling non-optical genome sequencing. *Nature* **475**, 348-52.

Roura E. & Tedò G. (2009) Feed acceptance in pigs: oronasal sensing perspective. *Voluntary Feed Intake in Pigs. Wageningen Academic Publishers, Wageningen*: 105-28.

Roudnitzky N., Behrens M., Engel A., *et al.* (2015) Receptor polymorphism and genomic structure interact to shape bitter taste perception. *PLoS Genetics* **11**, e1005530.

Sausenthaler S., Rzehak P., Wichmann H.E. & Heinrich J. (2009) Lack of relation between bitter taste receptor TAS2R38 and BMI in adults. *Obesity* **17**, 937-8.

Solà-Oriol D., Roura E. & Torrallardona D. (2014) Feed preference in pigs: relationship between cereal preference and nutrient composition and digestibility. *Journal of Animal Science* **92**,220-8.

Tepper B.J., Koelliker Y., Zhao L., *et al.* (2008) Variation in the bitter-taste receptor gene *TAS2R38* and adiposity in a genetically isolated population in southern Italy. *Obesity* **16**, 2289-95.

Tepper B.J. & Ullrich N.V. (2002) Influence of genetic taste sensitivity to 6-n-propylthiouracil (PROP), dietary restraint and disinhibition on body mass index in middle-aged women. *Physiology & Behavior* **75**, 305-12.

Wu S.V., Chen M.C. & Rozengurt E. (2005) Genomic organization, expression, and function of bitter taste receptors (T2R) in mouse and rat. *Physiological Genomics* **22**, 139-49.

Yamazaki T., Sagisaka M., Ikeda R., *et al.* (2014) The human bitter taste receptor hTAS2R39 is the primary receptor for the bitterness of theaflavins. *Bioscience, Biotechnology and Biochemistry* **78**, 1753-6.

Figure 1. Number of segregating single nucleotide polymorphisms (SNPs - missense; synonymous; 5'-flanking/5'-untranslated regions, 5'-UTR; 3'-flanking/3'-untranslated regions, 3'-UTR) identified in the different breeds or populations (P, Pietrain; M, Meishan; CA, Casertana; WB, wild boars; L, Italian Landrace; D, Italian Duroc; LW (-), Italian Large White with negative back fat thickness estimated breeding value; LW (+), Italian Large White with positive back fat thickness estimated breeding value.

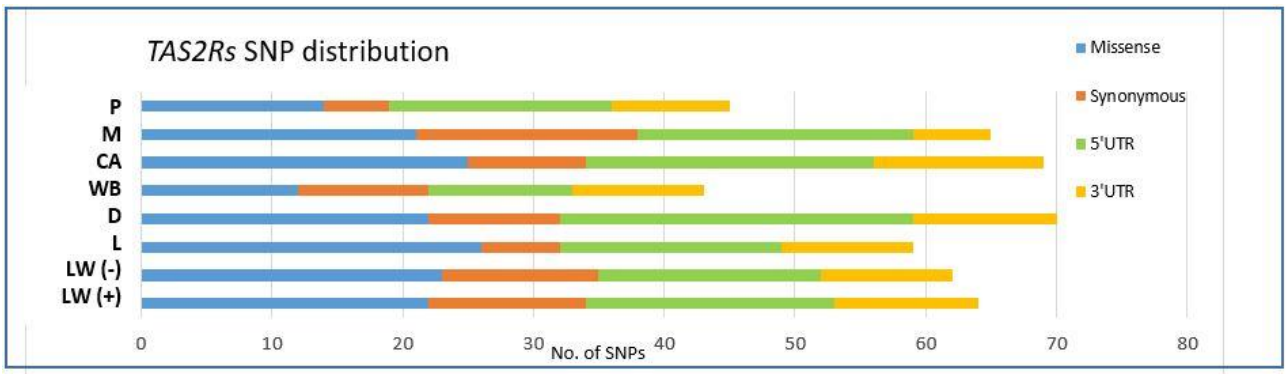


Figure 2. Heatmap representing allele frequencies distribution of the SNPs identified in the corresponding bitter taste receptor genes among the analysed pig breeds or populations (indicated as reported above).

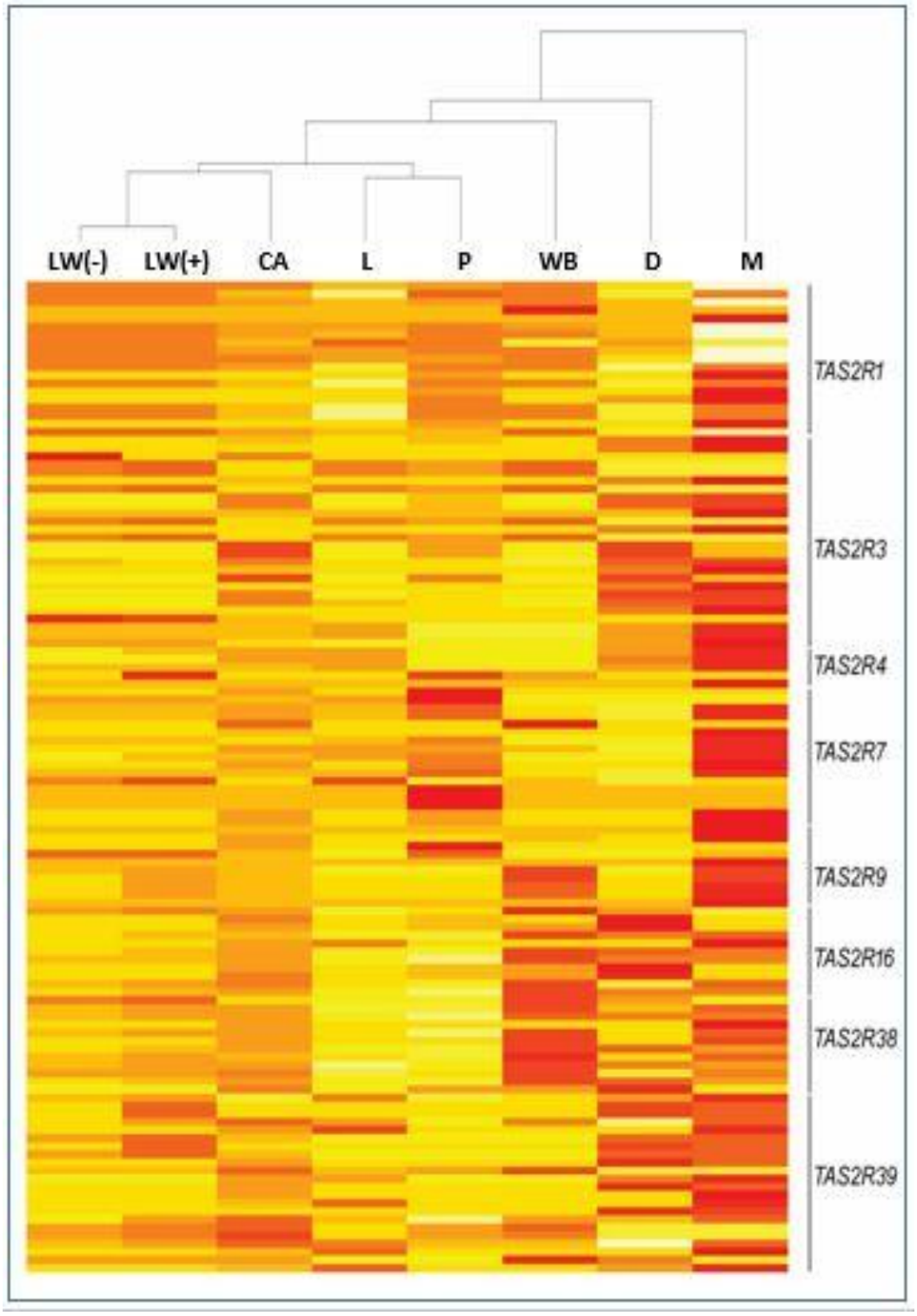


Figure 3. Snake plot of general bitter taste receptor protein (modified from Roudnitzky *et al.* 2015) with information on the position of the missense mutations identified in the deduced protein of the analysed porcine taste receptor genes. Colours correspond to different genes. Bolded circled amino acids indicate deleterious substitutions. Positions of the substituted amino acids have been deduced from multiple alignments of the protein sequences and Prediction of Transmembrane Regions and Orientation using TMpred server at the ExPASy Bioinformatic resource portal (http://embnet.vital-it.ch/software/TMPRED_form.html). Numbers in the amino acid positions in the snake plots indicate a progressive order reported below the figure with the corresponding amino acid substitution details. TMI-VII indicates the seven transmembrane domains.

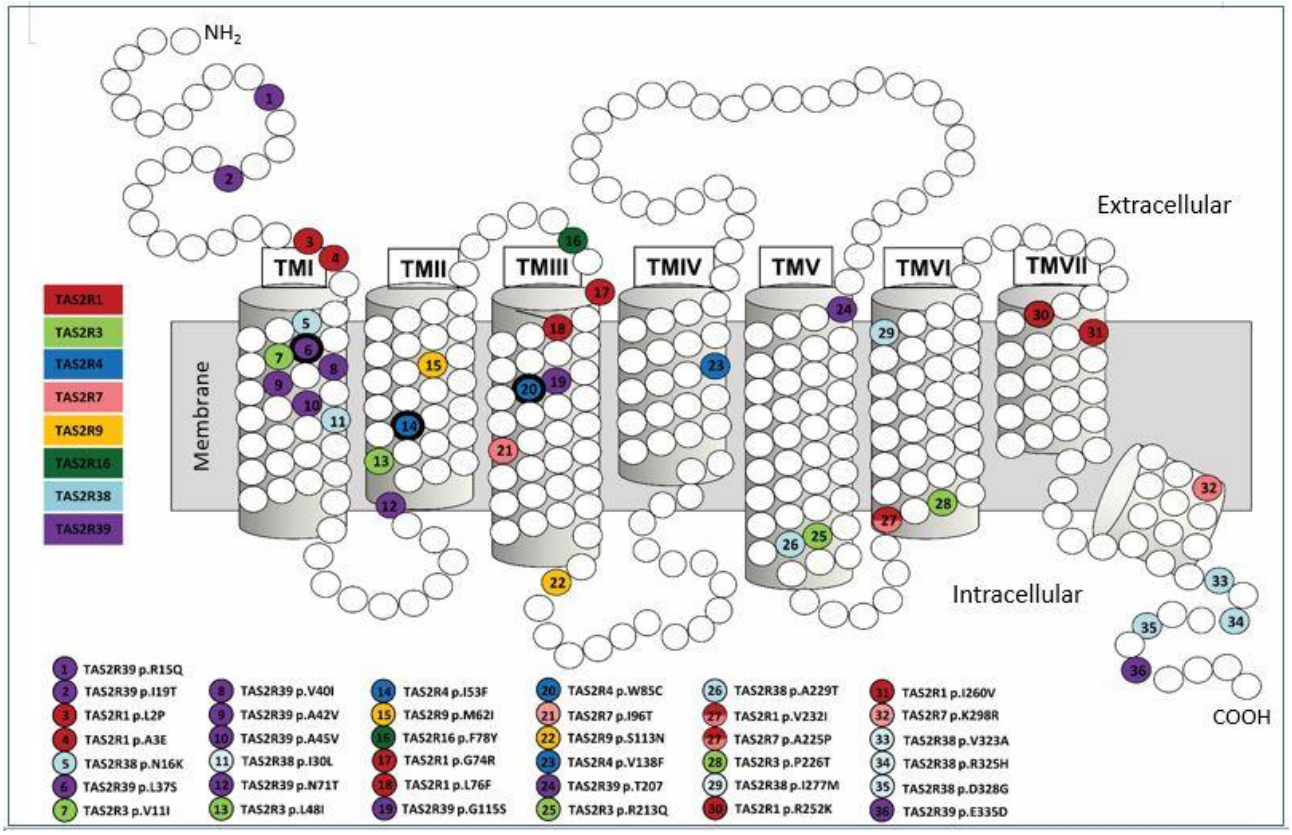


Table 1. Summary of information from the analysed porcine bitter taste receptor genes, comparison with putative homologous genes in humans and mice and number of single nucleotide polymorphisms (SNPs) identified in this study.

***Sus scrofa* gene symbol (NCBI)[1]**	**SSC (coordinates and gene size)[2]**	***Homo sapiens* (identity %)[4]**	***Mus musculus* (identity %)[5]**	**No. of SNPs[6]**
TAS2R1 (LOC106508396)	GL893464.1 (28,052-29,033; 981 bp)[3]	*TAS2R1* (74)	*Tas2r119* (71)	22 (2, 7, 2, 11)
TAS2R3 (LOC100621051)	GL892960.2 (34,965-35,915; 950 bp)[3]	*TAS2R3* (81)	*Tas2r137* (76)	26 (20, 4, 2, 0)
TAS2R4 (LOC100620853)	GL892960.2 (41,686-42,576; 890 bp)[3]	*TAS2R4* (80)	*Tas2r108* (74)	5 (1, 3, 1, 0)
TAS2R7 (LOC100523246)	5 (63,985,142- 63,986,080; 938 bp)	*TAS2R7* (81)	*Tas2r130* (78)	6 (1, 3, 0, 2)
TAS2R9 (LOC100522867)	5 (63,976,739-63,977,674; 935 bp)	*TAS2R9* (83)	*Tas2r130* (69)	3 (1, 2, 0, 0)
TAS2R10 (LOC100522675)	5 (63,965,446-63,966,375; 929 bp)	*TAS2R10* (84)	*Tas2r104* (72)	8 (2, 0, 2, 4)
TAS2R16 (LOC100513769)	18 (25,883,452-25,884,354; 902 bp)	*TAS2R16* (77)	*Tas2r118* (71)	10 (2, 1, 3, 4)
TAS2R38 (LOC100624167)	18 (8,357,518-8,358,525; 1007 bp)	*TAS2R38* (77)	*Tas2r138* (75)	23 (8, 7, 8, 0)
TAS2R39 (LOC100621890)	18 (7,358,848-7,359,855; 1007 bp)	*TAS2R39* (81)	*Tas2r139* (71)	22 (6, 10, 4, 3)

[1] Within brackets: Gene symbol in NCBI Gene database (http://www.ncbi.nlm.nih.gov/gene/).

[2] SSC: porcine chromosome; in parenthesis: gene coordinates in the corresponding chromosome or scaffold as reported in the Sscrofa10.2 genome version Ensembl Release 84, March 2016, and size on the gene in bp).

[3] Scaffold ID and related gene coordinates within it are reported.

[4] Human gene with the highest identity.

[5] Mouse gene with the highest identity.

[6] The numbers within brackets correspond to SNPs in the 5'-flanking region/5'-untranslated region, missense mutations, synonymous mutations, SNPs in the 3'-flanking region/3'-untranslated region, respectively.

Supplementary material

Table S1. Primers, PCR conditions and methods used for Ion Torrent sequencing, Sanger sequencing and PCR-RFLP analyses.

Gene symbol	Primer name	Primer sequence[1,2]	Primer start position (SSC:coordinate)[3]	Amplified fragments size (in bp, excluding primers)	Use[4]
TAS2R1	TAS2R1-like_pig_p1_F	GTCTTCGGCTTCCTCTCTGA	GL893464.1:27,505	381	Ion Torrent/Sanger
	TAS2R1-like_pig_p1_R	TGGATTTGCTCCTTTTTCAA	GL893464.1:27,926		
	TAS2R1-like_pig_p2_F	TGCAAGATGAGACTGCAAGG	GL893464.1:27,816	409	Ion Torrent/Sanger
	TAS2R1-like_pig_p2_R	ATCCTGGTCTTCCTGGGTCT	GL893464.1:28,264		
	TAS2R1-like_pig_p4_F	GCAAACCAAAGTCCCGATT	GL893464.1:28,747	450	Ion Torrent/Sanger
	TAS2R1-like_pig_p4_R	AGCAAATCCTCGCCAGAGTA	GL893464.1:29,235		
	TAS2R1_773_SS_F	GGCAGCAGTTTTGTTCTTTT	GL893464.1:28,079	108	PCR-RFLP (*Ssp*I)
	TAS2R1_773_SS_R	CAACTTCTTCCCAATCTGCT	GL893464.1:28,226		
TAS2R3	TAS2R3_IT_1_F	TGAGAAACCTTTGTATTCCCAGT	GL892960.2:33,979	940	Ion Torrent
	TAS2R3_IT_1_R	AGCTTATTCGGCAGACCTGA	GL892960.2:34,961		
	TAS2R3_IT_2_F	GGGCAGAGACAAGAGACAGG	GL892960.2:34,838	952	Ion Torrent

	TAS2R3_IT_2_R	TGTTTCCCAGAATGAGGACA	GL892960.2:35,829		
TAS2R4	TAS2R4_IT_1_F	AGATGGGGAAGATGGTTGC	GL892960.2:41,161	809	Ion Torrent
	TAS2R4_IT_1_R	TGGCAATCTTCACACAGTACAA	GL892960.2:42,010		
	TAS2R4_IT_2_F	TTTGGTTTGTAACCTTGCTGAA	GL892960.2:41,963	801	Ion Torrent
	TAS2R4_IT_2_R	TGCCCATGTAAAACACATGC	GL892960.2:42,805		
TAS2R7	TAS2R7_IT_1_F	TGGGTTTCTGGTAAGTTTATTCG	5:63,984,702	826	Ion Torrent
	TAS2R7_IT_1_R	AAACACAGAGAGGGCCAAAC	5:63,985,570		
	TAS2R7_IT_2_F	TTCCTCTGGATGAGGTGGAG	5:63,985,496	925	Ion Torrent
	TAS2R7_IT_2_R	GTTGAGCCAACAGCTTTTCA	5:63,986,461		
TAS2R9	TAS2R9_IT_1_F	GGCTACCCATCTCTTTCATTCC	5:63,976,173	892	Ion Torrent
	TAS2R9_IT_1_R	TTCAGCCAGAGGAAAAATGG	5:63,977,106		
	TAS2R9_186_deg_SS_F	CCAGAATCTGCTTGTTGTCTGTaAT[2]	5:63,976,899	147	PCR-RFLP (*Ssp*I)
	TAS2R9_186_SS_R	AAAATGGGTGGGATATACTGGCT	5:63,977,093		
TAS2R10	TAS2R10_IT_1_F	TTCAAAGAAACTTGTGTCTTCAGTG	5:63,964,924	880	Ion Torrent
	TAS2R10_IT_1_R	CCCATCAGAATGAGAAGAACC	5:63,965,849		
	TAS2R10_IT_2_F	GCAAATTTTTCCCACCACAT	5:63,965,779	939	Ion Torrent
	TAS2R10_IT_2_R	TTTGGGTTTTCCTGGTGAAG	5:63,966,757		
TAS2R16	TAS2R16_IT_1_F	TCTCAGGCTGATAATGAAGGAA	18:25,883,129	803	Ion Torrent
	TAS2R16_IT_1_R	GCAGAAATGTCTCAAGCCTCT	18:25,883,974		
	TAS2R16_IT_2_F	TTCTCTAGAAACAGCACCGTGA	18:25,883,929	806	Ion Torrent

	TAS2R16_IT_2_R	TGGGCAATAAATTCTTGTGG	18:25,884,776		
	TAS2R16_233_deg_SS_F	CTCCCACTTCCATCCTCACTGTGaAT[2]	18:25,883,658	88	PCR-RFLP (*Ssp*I)
	TAS2R16_233_SS_R	ATGGGGCAGCTGAAGGAGGAG	18:25,883,792		
TAS2R20	TAS2R20_IT_1_F	CAATCCCTGGCCTCTCAGT	5:63,903,604	938	Ion Torrent
	TAS2R20_IT_1_R	TGAAGAAAAACTAGGCTGGAGAA	5:63,904,581		
	TAS2R20_IT_2_F	CTTCTGGGAGCTTCGTTCTT	5:63,904,533	804	Ion Torrent
	TAS2R20_IT_2_R	TTCATTTGGACGATGATAACG	5:63,905,377		
TAS2R38	TAS2R38_IT_1_F	AAAACAGTTCATGGTAAGAGTCTCC	18:8,356,992	802	Ion Torrent
	TAS2R38_IT_1_R	CAGGCCAGCTTGATTTATGA	18:8,357,838		
	TAS2R38_IT_2_F	CAGCTACCAAACCACCATCA	18:8,357,787	805	Ion Torrent
	TAS2R38_IT_2_R	TGGATCTTTAACCCAGTGTGC	18:8,358,632		
TAS2R39	TAS2R39_IT_1_F	CCACATCTGCAAAATGGAG	18:7,358,370	808	Ion Torrent
	TAS2R39_IT_1_R	AGGCCACAGTAATTGAAGAACA	18:7,359,218		
	TAS2R39_IT_2_F	TTCAACATCCCCAAGCTTTT	18:7,359,135	876	Ion Torrent
	TAS2R39_IT_2_R	AATGGAAGACGCAGGTGAAG	18:7,360,050		
	TAS2R39_118_SS_F	ATGATCAAAACCAGCAGTCC	18:7,358,899	138	PCR-RFLP (*Psi*I)
	TAS2R39_118_SS_R	AAAGCAGGATCTTGCCATTT	18:7,359,077		
	TAS2R39_N71T_SS_F	TGGGTTCATTGCAGCTATAA	18:7,359,000	328	PCR-RFLP (*Bsr*I)
	TAS2R39_N71T_SS_R	AAACCAGGGCATCAATCCAG	18:7,359,327		

[1] PCR primers were designed on the corresponding regions retrieved from Sscrofa10.2. Primers were designed to amplify all coding exons including non-coding transcribed and flanking regions. Genomic DNA was extracted from blood, muscle or ear tissue using the Wizard® Genomic DNA Purification kit (Promega Corporation, Madison, WI, USA). Extracted DNA was quantified in triplicate using a NanoPhotometer P-330 instrument (Implen GmbH, München, Germany) and pooled at equimolar concentration (as reported in the text for the different breeds/populations) for the Ion Torrent sequencing. For all other applications, each DNA sample was considered separately.

[2] Nucleotides in lower case letter have been included to create artificial restriction sites for PCR-RFLP analysis.

[3] The nucleotide coordinate in the corresponding *Sus scrofa* chromosome (SSC) in the Sscrofa10.2 genome version is reported. The scaffold ID is reported for the genes not assembled in any chromosome yet.

[4] Use of the amplified fragments. Ion Torrent: sequencing with Ion Torrent; Sanger: sequencing with Sanger technology; PCR-RFLP: analisys of polymorphisms with this genotyping approach.

Ion Torrent sequencing. PCRs on equimolar pools were performed in a total volume of 25 µl of final volume in a 2720 thermal cycler (Life Technologies, Carlsbad, CA, USA): the reaction included 50 ng of genomic DNA, 1X of Phusion® Hot Start Flex 2X Master Mix, 10 µM of Forward Primer, 10 µM of Reverse Primer with the following amplification condition: 98 °C for 1 min, followed by 35 cycles with 98 °C for 10 sec, 57-59 °C for 30 sec, 72 °C for 40 sec and final elongation step at 72 °C for 8 min. Amplified fragments were purified using ExoSAP-IT® (USB Corporation, Cleveland, Ohio, USA), pooled according to the breed/DNA pool of origin using putative equimolar DNA from each amplified product estimated from gel electrophoresis. The resulting pooled fragments were then used for library preparation and sequencing with the Ion Torrent PGM (Life Technologies). Briefly, 200 ng of every pool of amplified products was enzymatically sheared, end repaired and ligated with different barcodes using the Ion Xpress™ Plus Fragment Library kit (Life Technologies) and the Ion Xpress™ Barcode Adapters 1-16 Kit (Life Technologies). The resulting DNA was then size selected using the e-gel system (Invitrogen, Carlsbad, CA, USA) and bands corresponding to 100 bp of inserts were collected and quantified by qPCR using a StepOnePlus™ Real-Time PCR System (Life Technologies). Barcoded fragments were pooled again with the same concentration, clonally amplified, purified and sequenced using the Ion One Touch™ 100 Template kit and the Ion PGM™ Sequencing kit with an Ion 316 chip (Life

Technologies). Obtained sequencing reads were filtered, trimmed, automatically assigned to the different barcodes, and analysed for the presence of differences between the reference sequences (from Sscrofa10.2) and obtained sequences with the Ion Torrent suite v.2.2 (Life Technologies).

PCR-RFLP. PCRs were obtained in a 2720 Life Technologies thermal cycler (Life Technologies) with the following profile: 5 min at 95 °C; 35 amplification cycles of 30 sec at 95 °C, 30 sec at the appropriate annealing temperature (53-62 °C), 30 sec at 72 °C; 10 min at 72 °C. The final reaction volume was of 20 µL and included about 50 ng of template DNA; the Phusion® High-Fidelity PCR Master Mix (New England Biolabs, Ipswich, UK); 10 pmol of each primer. A total of 5 µl of PCR product was digested overnight with 2 U of the enzyme indicated in brackets and 1X reaction buffer in a total of 20 µL of reaction volume, following the temperature indicated in each manufacturer protocol. Resulting DNA fragments were electrophoresed in 2.5% agarose gels in TBE 1X and stained with 1X GelRed Nucleic Acid Gel Stain (Biotium Inc., Hayward, CA, USA).

Sanger sequencing. Amplicons obtained by PCR from individual DNA samples were purified with ExoSAP-IT (USB Corporation, Cleveland, Ohio, USA) and then sequenced using the Sanger method with the BrightDye® Terminator Cycle Sequencing Kit (NIMAGEN, Nijmegen, the Netherlands). Sequencing reactions were loaded on an ABI3100 Avant sequencer (Applied Biosystems). All sequences were visually inspected and aligned with the help of CodonCode Aligner (version 5.1.5) software.

Table S2. Average depth and coverage for the sequenced gene regions in the different breeds/populations and average across all genes. Breeds/populations are defined below.

Gene/items	Mean depth (X)								Coverage (%)							
	LW (+)	LW (-)	L	D	WB	CA	M	P	LW (+)	LW (-)	L	D	WB	CA	M	P
TAS2R1	417	3748	5851	2054	268	143	107	59	100	100	100	100	100	100	100	76
TAS2R3	1956	10978	1936	6432	1005	858	590	777	100	100	100	100	100	100	100	100
TAS2R4	2119	10270	6894	7845	1186	905	425	457	100	100	100	100	100	100	100	100
TAS2R7	1801	11346	5717	6100	978	995	484	835	100	100	100	100	100	100	100	100
TAS2R9	474	2854	2023	2645	342	388	252	253	100	100	100	100	100	100	100	100
TAS2R10	498	3133	2252	2718	157	246	168	153	100	100	100	100	100	100	100	100
TAS2R16	2212	12758	8107	9116	1127	1068	568	825	100	100	100	100	100	100	100	100
TAS2R38	2107	10489	5788	6499	1069	898	547	575	99	100	100	100	100	100	100	100
TAS2R39	1769	4083	4230	5868	392	542	328	149	100	100	100	100	100	100	100	100
Averaged	1484	7740	4755	5475	725	672	385	454	100	100	100	100	100	100	100	97

Mean depth indicates the mean number of times that each nucleotide in the analysed genes is covered by the reads (calculated per pool and then averaged per breed/group). Coverage indicates the percentage of each gene with depth of at least 1/10 the mean depth. LW (+), Italian Large White with positive back fat thickness estimated breeding value, LW (-), Italian Large White with negative back fat thickness estimated breeding value; L, Italian Landrace; D, Italian Duroc; WB, wild boars; CA, Casertana; M, Meishan; P, Pietrain.

Table S3. Detailed information on the identified single nucleotide polymorphisms (SNPs).

Gene symbol	SNP position (SSC:chromosome coordinate or scaffold) in Sscrofa10.2[1,2,3]	Position in the amplicon	Gene region/type of mutation[4]	Effects on protein	SIFT	Novel or reported also by previous studies	dbSNP ID	LW (+)[5]	LW (+)	LW (-)	LW (-)	L	L	D	D	WB	WB	CA	CA	M	M	P	P
								freq_ref[6]	freq_alt[7]	freq_ref	freq_alt	freq_ref	freq_alt	freq_ref	freq_alt	freq_ref	freq_alt	freq_ref	freq_alt	freq_ref	freq_alt	freq_ref	freq_alt
TAS2R1	GL893464.1:g.27529G>C	24	3'UTR			This study	ss1971458462	-	1,00	0,02	0,98	0,28	0,72	0,49	0,51	-	1,00	0,10	0,90	1,00	-	0,22	0,78
	GL893464.1:g.27703C>T	198	3'UTR			This study	ss1971458463	0,10	0,90	0,11	0,89	0,43	0,57	0,39	0,61	0,10	0,90	0,22	0,78	0,08	0,92	0,06	0,94
	GL893464.1:g.27727C>G	222	3'UTR			This study	ss1971458464	0,01	0,99	0,00	1,00	0,22	0,78	0,21	0,79	-	1,00	0,10	0,90	1,00	-	0,18	0,82
	GL893464.1:g.27781C>G	276	3'UTR			This study	ss1971458465	1,00	-	1,00	-	1,00	-	1,00	-	0,63	0,37	1,00	-	1,00	-	1,00	-
	GL893464.1:g.27819A>C	314	3'UTR			This study	ss1971458466	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	0,26	0,74	1,00	-
	GL893464.1:g.27822A>G	317	3'UTR			This study	ss1971458467	0,01	0,99	0,04	0,96	0,18	0,82	0,24	0,76	0,07	0,93	0,13	0,87	1,00	-	-	1,00
	GL893464.1:g.27843G>A	28	3'UTR			This study	ss1971458468	0,01	0,99	0,01	0,99	0,19	0,81	0,27	0,73	-	1,00	0,05	0,95	1,00	-	-	1,00
	GL893464.1:g.27847A>C	32	3'UTR			This study	ss1971458469	0,37	0,63	0,29	0,71	0,18	0,82	0,39	0,61	1,00	-	0,51	0,49	1,00	-	0,38	0,63
	GL893464.1:g.27851G>T	36	3'UTR			This study	ss1971458470	0,01	0,99	0,01	0,99	0,16	0,84	0,23	0,77	0,01	0,99	0,03	0,97	1,00	-	-	1,00
	GL893464.1:g.27863A>G	48	3'UTR			This study	ss1971458471	0,00	1,00	0,01	0,99	0,16	0,84	0,38	0,62	0,01	0,99	0,03	0,97	1,00	-	0,15	0,85
	GL893464.1:g.27869G>C	54	3'UTR			This study	ss1971458472	-	1,00	0,00	1,00	0,17	0,83	0,24	0,76	-	1,00	0,03	0,97	-	1,00	-	1,00
	GL893464.1:g.28156T>C	341	Missense	I260V	0,88	This study	ss1971458473	-	1,00	0,02	0,98	0,46	0,54	0,39	0,61	-	1,00	0,27	0,73	-	1,00	NA[8]	NA
	GL893464.1:g.28179C>T	364	Missense	R252K	0,2	This study	ss1971458474	-	1,00	0,02	0,98	0,46	0,54	0,44	0,56	-	1,00	0,26	0,74	1,00	-	NA	NA
	GL893464.1:g.28240C>T	425	Missense	V232I	0,67	This study	ss1971458475	-	1,00	0,02	0,98	0,42	0,58	0,40	0,60	-	1,00	0,25	0,75	1,00	-	NA	NA
	GL893464.1:g.28808G>A	60	Missense	L76F	0,71	This study	ss1971458476	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	0,01	0,99	0,71	0,29
	GL893464.1:g.28814C>T	68	Missense	G74R	0,57	This study	ss1971458477	-	1,00	0,01	0,99	0,52	0,48	0,40	0,60	-	1,00	0,21	0,79	0,01	0,99	-	1,00
	GL893464.1:g.28842G>A	96	Synonymous	C64		This study	ss1971458478	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	0,03	0,97	0,77	0,23
	GL893464.1:g.28953A>G	207	Synonymous	I27		This study	ss1971458479	1,00	-	1,00	-	1,00	-	0,66	0,34	1,00	-	1,00	-	0,00	1,00	0,61	0,39
	GL893464.1:g.29026G>T	280	Missense	A3E	1	This study	ss1971458480	0,00	1,00	0,01	0,99	0,47	0,53	0,37	0,63	0,01	0,99	0,20	0,80	0,00	1,00	-	1,00
	GL893464.1:g.29029A>G	283	Missense	L2P	0,1	This study	ss1971458481	0,00	1,00	0,01	0,99	0,47	0,53	0,37	0,63	0,00	1,00	0,20	0,80	-	1,00	-	1,00

	GL893464.1:g.29037G>A	291	5'UTR			This study	ss1971458482	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	0,02	0,98	0,73	0,27
	GL893464.1:g.29058T>C	312	5'UTR			This study	ss1971458483	0,00	1,00	0,02	0,98	0,47	0,53	0,70	0,30	0,00	1,00	0,21	0,79	1,00	-	0,46	0,54
TAS2R3	GL892960.2:g.33976G>A	46	5'UTR			This study	ss1971458484	1,00	-	1,00	-	0,78	0,22	0,61	0,39	1,00	-	0,85	0,15	0,19	0,81	0,87	0,13
	GL892960.2:g.33988T>C	58	5'UTR			This study	ss1971458485	1,00	-	1,00	-	1,00	-	0,62	0,38	1,00	-	1,00	-	0,19	0,81	1,00	-
	GL892960.2:g.34100G>A	170	5'UTR			This study	ss1971458486	0,80	0,20	0,71	0,29	0,94	0,06	1,00	-	1,00	-	0,90	0,10	1,00	-	1,00	-
	GL892960.2:g.34112T>A	182	5'UTR			This study	ss1971458487	0,10	0,90	0,18	0,82	0,27	0,73	1,00	-	0,02	0,98	0,71	0,29	1,00	-	0,37	0,63
	GL892960.2:g.34131A>G	201	5'UTR			This study	ss1971458488	0,11	0,89	0,18	0,82	0,27	0,73	1,00	-	0,03	0,97	0,73	0,27	1,00	-	0,37	0,63
	GL892960.2:g.34132T>C	202	5'UTR			This study	ss1971458489	0,95	0,05	0,93	0,07	0,91	0,09	0,64	0,36	0,95	0,05	0,90	0,10	0,21	0,79	0,95	0,05
	GL892960.2:g.34133A>C	203	5'UTR			This study	ss1971458490	0,11	0,89	0,18	0,82	0,26	0,74	1,00	-	0,02	0,98	0,72	0,28	1,00	-	0,37	0,63
	GL892960.2:g.34179G>A	249	5'UTR			This study	ss1971458491	0,89	0,11	0,84	0,16	0,79	0,21	0,15	0,85	0,98	0,02	0,36	0,64	-	1,00	0,71	0,29
	GL892960.2:g.34181C>A	251	5'UTR			This study	ss1971458492	0,90	0,10	0,85	0,15	0,81	0,19	0,16	0,84	1,00	-	0,37	0,63	-	1,00	0,72	0,28
	GL892960.2:g.34205G>A	275	5'UTR			This study	ss1971458493	0,99	0,01	0,98	0,02	0,95	0,05	0,66	0,34	1,00	-	1,00	-	0,27	0,73	1,00	-
	GL892960.2:g.34218T>A	288	5'UTR			This study	ss1971458494	0,10	0,90	0,17	0,83	0,25	0,75	0,76	0,24	0,02	0,98	0,71	0,29	1,00	-	0,37	0,63
	GL892960.2:g.34285C>T	355	5'UTR			This study	ss1971458495	1,00	-	1,00	-	1,00	-	0,65	0,35	1,00	-	1,00	-	0,20	0,80	1,00	-
	GL892960.2:g.34410A>G	480	5'UTR			This study	ss1971458496	0,10	0,90	0,17	0,83	0,25	0,75	0,92	0,08	0,02	0,98	0,73	0,27	1,00	-	0,35	0,65
	GL892960.2:g.34488A>G	558	5'UTR			This study	ss1971458497	0,93	0,07	0,86	0,14	0,83	0,17	0,42	0,58	1,00	-	0,31	0,69	0,81	0,19	0,66	0,34
	GL892960.2:g.34516C>G	586	5'UTR			This study	ss1971458498	0,92	0,08	0,86	0,14	0,84	0,16	0,44	0,56	1,00	-	0,33	0,67	0,82	0,18	0,69	0,31
	GL892960.2:g.34522C>G	592	5'UTR			This study	ss1971458499	0,73	0,27	0,59	0,41	0,75	0,25	0,13	0,87	1,00	-	0,26	0,74	0,06	0,94	0,69	0,31
	GL892960.2:g.34537A>G	607	5'UTR			This study	ss1971458500	1,00	-	1,00	-	0,82	0,18	0,70	0,30	1,00	-	0,83	0,17	0,20	0,80	1,00	-
	GL892960.2:g.34538C>T	608	5'UTR			This study	ss1971458501	0,93	0,07	0,87	0,13	1,00	-	0,42	0,58	1,00	-	0,50	0,50	0,82	0,18	0,69	0,31
	GL892960.2:g.34555G>A	625	5'UTR			This study	ss1971458502	1,00	-	1,00	-	1,00	-	0,71	0,29	1,00	-	1,00	-	0,18	0,82	1,00	-
	GL892960.2:g.34759T>C	829	5'UTR			This study	ss1971458503	0,89	0,11	0,87	0,13	0,76	0,24	0,11	0,89	1,00	-	0,33	0,67	0,01	0,99	0,85	0,15
	GL892960.2:g.34995G>A	158	Missense	V11I	0,47	This study	ss1971458504	0,88	0,12	0,83	0,17	0,68	0,32	0,16	0,84	1,00	-	0,35	0,65	0,01	0,99	0,82	0,18
	GL892960.2:g.35106C>A	269	Missense	L48I	1	This study	ss1971458505	1,00	-	1,00	-	0,65	0,35	0,63	0,37	1,00	-	1,00	-	0,22	0,78	1,00	-
	GL892960.2:g.35387A>T	550	Synonymous	S141		This study	ss1971458506	0,78	0,22	0,76	0,24	0,93	0,07	1,00	-	1,00	-	0,94	0,06	1,00	-	1,00	-
	GL892960.2:g.35602G>A	765	Missense	R213Q	1	This study	ss1971458507	0,72	0,28	0,76	0,24	0,68	0,32	0,64	0,36	1,00	-	0,75	0,25	0,28	0,72	1,00	-
	GL892960.2:g.35640C>A	803	Missense	P226T	1	This study	ss1971458508	0,72	0,28	0,75	0,25	0,64	0,36	0,61	0,39	1,00	-	0,74	0,26	0,24	0,76	1,00	-

	GL892960.2:g.35729C>T	892	Synonymous	S255		This study	ss1971458509	0,73	0,27	0,74	0,26	0,83	0,17	0,68	0,32	1,00	-	0,96	0,04	0,22	0,78	1,00	-
TAS2R4	GL892960.2:g.41356T>C	238	5'UTR			This study	ss1971458510	0,77	0,23	0,75	0,25	0,68	0,32	0,66	0,34	1,00	-	0,73	0,27	0,26	0,74	1,00	-
	GL892960.2:g.41842A>T	724	Missense	I53F	0,01	This study	ss1971458511	0,74	0,26	0,76	0,24	0,68	0,32	0,67	0,33	1,00	-	0,75	0,25	0,24	0,76	1,00	-
	GL892960.2:g.41940G>C	822	Missense	W85C	0,01	This study	ss1971458512	0,74	0,26	0,78	0,22	0,69	0,31	0,67	0,33	1,00	-	0,79	0,21	0,28	0,72	1,00	-
	GL892960.2:g.42100G>T	138	Missense	V139F	0,89	This study	ss1971458513	0,64	0,36	0,79	0,21	0,94	0,06	1,00	-	0,86	0,14	1,00	-	1,00	-	0,69	0,31
	GL892960.2:g.42117C>T	155	Synonymous	Y144		This study	ss1971458514	1,00	-	1,00	-	1,00	-	0,87	0,13	1,00	-	1,00	-	0,57	0,43	1,00	-
TAS2R7	5:g.63984901T>G	200	5'UTR			da Silva et al. (2014)	rs332658718	0,85	0,15	0,84	0,16	0,81	0,19	1,00	-	1,00	-	0,82	0,18	1,00	-	0,40	0,60
	5:g.63985428T>C	727	Missense	I96T	0,11	da Silva et al. (2014)	rs325458119	0,70	0,30	0,70	0,30	0,68	0,32	1,00	-	1,00	-	0,83	0,17	1,00	-	0,18	0,82
	5:g.63985814G>C	319	Missense	A225P	1,00	da Silva et al. (2014)	rs344408296	0,63	0,37	0,67	0,33	0,61	0,39	1,00	-	0,78	0,22	0,46	0,54	-	1,00	0,29	0,71
	5:g.63986034G>A	539	Missense	K298R	1	da Silva et al. (2014)	rs335556860	0,65	0,35	0,66	0,34	0,61	0,39	1,00	-	0,77	0,23	0,46	0,54	0,00	1,00	0,29	0,71
	5:g.63986142A>G	646	3'UTR			dbSNP database	rs331370913	0,96	0,04	1,00	-	0,86	0,14	1,00	-	0,71	0,29	0,84	0,16	1,00	-	1,00	-
	5:g.63986415A>C	917	3'UTR			This study	ss1971458515	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	0,94	0,06	0,07	0,93	1,00	-
TAS2R9	5:g.63976430C>T	258	5'UTR			da Silva et al. (2014)	rs318422340	0,75	0,25	0,69	0,31	0,68	0,32	1,00	-	1,00	-	0,83	0,17	0,00	1,00	0,45	0,55
	5:g.63976924G>A	752	Missense	M62I	1	da Silva et al. (2014)	rs341774888	0,67	0,33	0,68	0,32	0,52	0,48	1,00	-	0,60	0,40	0,54	0,46	-	1,00	0,48	0,52
	5:g.63977076G>A	904	Missense	S113N	1	da Silva et al. (2014)	rs81384489	0,75	0,25	0,72	0,28	0,67	0,33	1,00	-	1,00	-	0,74	0,26	-	1,00	0,45	0,55
TAS2R10	5:g.63965383G>A	460	5'UTR			da Silva et al. (2014)	rs337239138	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	0,73	0,27	-	1,00	0,53	0,47
	5:g.63965434G>A	511	5'UTR			dbSNP database	rs344116619	0,68	0,32	0,67	0,33	0,57	0,43	1,00	-	0,81	0,19	0,60	0,40	-	1,00	0,29	0,71
	5:g.63965817T>A	39	Synonymous	G124		This study	ss1971458516	0,30	0,70	0,45	0,55	0,28	0,72	0,69	0,31	0,69	0,31	0,85	0,15	0,79	0,21	0,90	0,10
	5:g.63966066G>A	288	Synonymous	R207		da Silva et al. (2014)	rs318837222	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	0,83	0,17

	5:g.63966540G>A	762	3'UTR			da Silva et al. (2014)	rs330407336	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	0,71	0,29
	5:g.63966654A>G	876	3'UTR			da Silva et al. (2014)	rs333662429	0,90	0,10	0,83	0,17	0,89	0,11	1,00	-	1,00	-	1,00	-	1,00	-	0,68	0,32
	5:g.63966666A>G	888	3'UTR			da Silva et al. (2014)	rs345144976	0,89	0,11	0,89	0,11	0,84	0,16	1,00	-	1,00	-	0,69	0,31	0,03	0,97	0,69	0,31
	5:g.63966675A>G	897	3'UTR			da Silva et al. (2014)	rs325493324	0,89	0,11	0,89	0,11	0,82	0,18	1,00	-	1,00	-	0,74	0,26	-	1,00	0,67	0,33
TAS2R16	18:g.25883270A>G	142	5'UTR			da Silva et al. (2014)	rs319457013	0,98	0,02	1,00	-	0,97	0,03	1,00	-	1,00	-	1,00	-	0,56	0,44	1,00	-
	18:g.25883439T>C	311	5'UTR			da Silva et al. (2014)	rs321449291	0,99	0,01	0,98	0,02	1,00	-	1,00	-	0,95	0,05	0,88	0,12	0,58	0,42	1,00	-
	18:g.25883604C>T	476	Synonymous	S51		da Silva et al. (2014)	rs322392839	0,87	0,13	0,80	0,20	1,00	0,00	0,98	0,02	1,00	-	0,93	0,07	1,00	-	0,58	0,42
	18:g.25883684T>A	556	Missense	F78Y	0,84	da Silva et al. (2014)	rs331863796	0,59	0,41	0,59	0,41	1,00	-	1,00	-	1,00	-	0,82	0,18	0,88	0,12	0,61	0,39
	18:g.25883991T>C	63	Synonymous	Y180		da Silva et al. (2014)	rs334023278	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	0,63	0,37	1,00	-
	18:g.25884087A>G	159	Synonymous	Q212		da Silva et al. (2014)	rs341414049	0,59	0,41	0,74	0,26	0,90	0,10	0,79	0,21	0,01	0,99	0,67	0,33	0,00	1,00	1,00	-
	18:g.25884466G>A	538	3'UTR			dbSNP database	rs321411308	0,73	0,27	0,81	0,19	0,90	0,10	0,97	0,03	0,29	0,71	0,80	0,20	0,21	0,79	1,00	-
	18:g.25884532T>C	604	3'UTR			da Silva et al. (2014)	rs336837976	0,67	0,33	0,75	0,25	0,91	0,09	0,80	0,20	0,31	0,69	0,80	0,20	-	1,00	1,00	0,00
	18:g.25884711C>T	783	3'UTR			da Silva et al. (2014)	rs340170229	0,62	0,38	0,74	0,26	0,91	0,09	0,79	0,21	0,31	0,69	0,74	0,26	-	1,00	1,00	-
	18:g.25884734A>G	806	3'UTR			da Silva et al. (2014)	rs319877687	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	1,00	-	0,65	0,35	1,00	-
TAS2R38	18:g.8357099C>T	108	5'UTR			da Silva et al. (2014)	rs332943020	0,40	0,60	0,50	0,50	0,93	0,07	0,44	0,56	0,01	0,99	0,51	0,49	1,00	-	0,76	0,24
	18:g.8357181G>A	190	5'UTR			da Silva et al. (2014)	rs327942989	0,96	0,04	0,90	0,10	0,99	0,01	0,49	0,51	1,00	-	0,75	0,25	1,00	-	0,90	0,10
	18:g.8357288C>T	297	5'UTR			da Silva et al. (2014)	rs339063186	0,95	0,05	0,89	0,11	1,00	-	0,24	0,76	0,80	0,20	0,74	0,26	1,00	-	0,90	0,10
	18:g.8357304T>C	313	5'UTR			da Silva et al. (2014)	rs318914269	0,46	0,54	0,58	0,42	0,69	0,31	0,30	0,70	0,02	0,98	0,44	0,56	0,21	0,79	0,82	0,18

	18:g.8357408G>C	417	5'UTR			da Silva et al. (2014)	rs336180080	0,78	0,22	0,78	0,22	0,68	0,32	1,00	-	1,00	-	0,74	0,26	0,22	0,78	1,00	-
	18:g.8357418A>G	427	5'UTR			da Silva et al. (2014)	rs320217700	0,44	0,56	0,51	0,49	0,68	0,32	0,16	0,84	0,02	0,98	0,32	0,68	0,22	0,78	0,80	0,20
	18:g.8357446A>C	455	5'UTR			da Silva et al. (2014)	rs339114943	0,44	0,56	0,52	0,48	0,69	0,31	0,24	0,76	0,01	0,99	0,33	0,67	0,23	0,77	1,00	-
	18:g.8357461G>C	470	5'UTR			da Silva et al. (2014)	rs319199758	0,90	0,10	0,84	0,16	1,00	-	0,25	0,75	0,80	0,20	0,74	0,26	1,00	-	0,90	0,10
	18:g.8357565C>A	574	Missense	N16K	0,45	da Silva et al. (2014)	rs344373699	0,90	0,10	0,84	0,16	1,00	-	0,56	0,44	0,81	0,19	0,76	0,24	1,00	-	0,91	0,09
	18:g.8357604A>G	613	Synonymous	G29		da Silva et al. (2014)	rs329663593	0,42	0,58	0,52	0,48	0,68	0,32	1,00	-	0,02	0,98	0,35	0,65	0,21	0,79	0,80	0,20
	18:g.8357605A>C	614	Missense	I30L	0,21	dbSNP database	rs340228133	0,39	0,61	0,49	0,51	0,65	0,35	0,17	0,83	0,01	0,99	0,32	0,68	0,18	0,82	0,82	0,18
	18:g.8358006T>C	220	Synonymous	F163		da Silva et al. (2014)	rs324968198	0,44	0,56	0,55	0,45	1,00	-	0,59	0,41	0,23	0,77	0,83	0,17	1,00	-	1,00	-
	18:g.8358067T>C	281	Synonymous	L184		da Silva et al. (2014)	rs344498111	0,49	0,51	0,66	0,34	1,00	-	0,66	0,34	0,23	0,77	0,60	0,40	0,28	0,72	1,00	-
	18:g.8358202G>A	416	Missense	A229T	0,68	dbSNP database	rs336060799	0,25	0,75	0,31	0,69	0,51	0,49	0,09	0,91	0,01	0,99	0,20	0,80	0,15	0,85	0,70	0,30
	18:g.8358264C>T	478	Synonymous	L249		da Silva et al. (2014)	rs343260652	0,99	0,01	0,99	0,01	0,83	0,17	1,00	-	1,00	-	0,78	0,22	0,22	0,78	1,00	-
	18:g.8358303C>T	517	Synonymous	A262		da Silva et al. (2014)	rs327214552	0,46	0,54	0,62	0,38	0,67	0,33	0,66	0,34	0,19	0,81	0,52	0,48	0,23	0,77	1,00	-
	18:g.8358348C>G	562	Missense	I277M	0,27	dbSNP database	rs337735554	0,75	0,25	0,92	0,08	0,87	0,13	0,72	0,28	0,31	0,69	0,67	0,33	0,37	0,63	0,96	0,04
	18:g.8358474A>G	688	Synonymous	L319		da Silva et al. (2014)	rs328044334	0,42	0,58	0,51	0,49	0,70	0,30	0,20	0,80	0,01	0,99	0,34	0,66	0,36	0,64	0,81	0,19
	18:g.8358480A>G	694	Synonymous	L321		da Silva et al. (2014)	rs335407353	0,53	0,47	0,67	0,33	0,74	0,26	0,72	0,28	0,18	0,82	0,63	0,37	0,38	0,62	0,96	0,04
	18:g.8358485T>C	699	Missense	V323A	1	dbSNP database	rs319573015	0,44	0,56	0,55	0,45	1,00	-	0,22	0,78	0,02	0,98	0,39	0,61	0,39	0,61	0,81	0,19
	18:g.8358486A>G	700	Synonymous	V323		da Silva et al. (2014)	rs330707292	0,45	0,55	0,55	0,45	1,00	-	1,00	-	0,01	0,99	0,40	0,60	0,39	0,61	0,81	0,19

	18:g.8358491G>A	705	Missense	R325H	1	dbSNP database	rs340948118	0,44	0,56	0,53	0,47	0,71	0,29	0,22	0,78	0,01	0,99	0,37	0,63	0,40	0,60	0,81	0,19
	18:g.8358500A>G	714	Missense	D328G	0,53	dbSNP database	rs318262391	0,90	0,10	0,86	0,14	0,98	0,02	0,57	0,43	0,85	0,15	0,77	0,23	1,00	-	0,84	0,16
TAS2R39	18:g.7358453A>T	84	5'UTR			da Silva et al. (2014)	rs342293575	0,47	0,53	0,67	0,33	0,40	0,60	0,46	0,54	0,75	0,25	1,00	-	-	1,00	1,00	-
	18:g.7358501G>C	132	5'UTR			dbSNP database	rs334394369	0,54	0,46	0,73	0,27	0,87	0,13	0,48	0,52	1,00	-	1,00	-	0,54	0,46	1,00	-
	18:g.7358533G>A	164	5'UTR			dbSNP database	rs343194200	0,62	0,38	0,74	0,26	0,86	0,14	0,56	0,44	1,00	-	1,00	-	0,58	0,42	1,00	-
	18:g.7358549A>G	180	5'UTR			da Silva et al. (2014)	rs326928677	0,12	0,88	0,30	0,70	0,14	0,86	0,47	0,53	0,10	0,90	-	1,00	-	1,00	0,38	0,62
	18:g.7358626A>C	257	5'UTR			da Silva et al. (2014)	rs345479433	0,96	0,04	0,94	0,06	0,55	0,45	0,97	0,03	1,00	-	0,83	0,17	0,45	0,55	1,00	-
	18:g.7358730G>C	361	5'UTR			da Silva et al. (2014)	rs325807180	0,57	0,43	0,71	0,29	0,85	0,15	0,51	0,49	1,00	-	0,84	0,16	0,54	0,46	1,00	-
	18:g.7358891G>A	522	Missense	R15Q	0,58	da Silva et al. (2014)	rs331832991	0,56	0,44	0,74	0,26	0,88	0,12	0,51	0,49	1,00	-	1,00	-	0,58	0,42	1,00	-
	18:g.7358903T>C	534	Missense	I19T	0,5	da Silva et al. (2014)	rs335568369	0,55	0,45	0,68	0,32	0,87	0,13	0,51	0,49	1,00	-	0,82	0,18	0,58	0,42	1,00	-
	18:g.7358928T>C	559	Synonymous	N27		da Silva et al. (2014)	rs319985660	0,96	0,04	0,90	0,10	0,99	0,01	0,48	0,52	0,99	0,01	0,79	0,21	0,65	0,35	0,94	0,06
	18:g.7358957T>C	588	Missense	L37S	0,02	da Silva et al. (2014)	rs331114472	0,62	0,38	0,65	0,35	0,69	0,31	1,00	-	0,12	0,88	0,32	0,68	1,00	-	0,53	0,47
	18:g.7358965G>A	596	Missense	V40I	0,47	da Silva et al. (2014)	rs345006834	0,91	0,09	0,86	0,14	0,56	0,44	0,48	0,52	1,00	-	0,67	0,33	-	1,00	1,00	-
	18:g.7358970C>T	601	Synonymous	S41		da Silva et al. (2014)	rs322441728	0,96	0,04	0,91	0,09	1,00	-	0,53	0,47	1,00	-	0,80	0,20	0,64	0,36	1,00	-
	18:g.7358972C>T	603	Missense	A42V	0,98	dbSNP database	rs339810026	0,96	0,04	0,86	0,14	0,86	0,14	1,00	-	1,00	-	0,88	0,12	0,37	0,63	1,00	-
	18:g.7358980A>G	611	Missense	I45V	0,71	dbSNP database	rs323606521	0,95	0,05	0,96	0,04	0,61	0,39	1,00	-	1,00	-	0,90	0,10	0,43	0,57	1,00	-
	18:g.7359027G>A	658	Synonymous	A60		da Silva et al. (2014)	rs334755846	1,00	-	1,00	-	1,00	-	0,52	0,48	1,00	-	1,00	-	0,62	0,38	1,00	-
	18:g.7359059A>C	690	Missense	N71T	1	dbSNP database	rs342835508	0,09	0,91	0,29	0,71	0,14	0,86	0,16	0,84	0,11	0,89	0,00	1,00	0,01	0,99	0,36	0,64

	18:g.7359190G>A	56	Missense	G115S	0,5	dbSNP database	rs323238022	0,30	0,70	0,41	0,59	0,72	0,28	0,71	0,29	0,24	0,76	0,18	0,82	0,89	0,11	0,40	0,60
	18:g.7359467C>T	333	Missense	T207I	0,72	dbSNP database	rs337700679	0,33	0,67	0,47	0,53	0,70	0,30	1,00	-	0,34	0,66	0,15	0,85	0,97	0,03	0,44	0,56
	18:g.7359534T>C	400	Synonymous	T229		da Silva et al. (2014)	rs81209906	0,23	0,77	0,40	0,60	0,15	0,85	0,90	0,10	0,34	0,66	0,01	0,99	0,05	0,95	0,43	0,57
	18:g.7359852G>C	718	Missense	E335D	0,46	dbSNP database	rs319025082	0,86	0,14	0,89	0,11	0,42	0,58	0,91	0,09	1,00	-	0,83	0,17	0,02	0,98	1,00	-
	18:g.7359892G>C	758	3'UTR			da Silva et al. (2014)	rs330304212	0,50	0,50	0,56	0,44	0,77	0,23	0,45	0,55	0,01	0,99	0,56	0,44	1,00	-	0,94	0,06
	18:g.7359969T>G	835	3'UTR			da Silva et al. (2014)	rs340176036	0,87	0,13	0,88	0,12	0,39	0,61	0,70	0,30	1,00	-	0,84	0,16	0,01	0,99	1,00	-

In bold, SNPs identified by sequencing overlapping amplicons. 2: underlined SNPs were genotyped also by PCR-RFLP. 3: the first allele is the reference allele, the second allele is the alternative allele. 4: 5'UTR and 3'UTR indicate SNPs in the flaking or untranslated regions. 5: Breeds or Populations acronyms are as reported as note to Table S2. 6,7: freq_ref and freq_alt indicate the estimated frequency based on Ion Torrent read counts of the reference and alternative alleles, respectively. 8: NA: not available

Table S4. Comparison of allele frequency estimation methods: Ion Torrent by sequencing DNA pools vs PCR-RFLP by genotyping individual samples of the DNA pools (see text).

Gene	**SNP position**[1]	**Method**	**Allele frequencies**[2,3]							
			LW (+)		**LW (-)**		**WB**		**CA**	
			Ref	**Alt**	**Ref**	**Alt**	**Ref**	**Alt**	**Ref**	**Alt**
TAS2R1	GL893464.1:g.28179C>T	Ion Torrent	-	1.00	0.02	0.98	-	1.00	0.26	0.74
		PCR-RFLP	-	1.00	0.03	0.97	-	1.00	0.20	0.80
TAS2R9	5:g.63976924G>A	Ion Torrent	0.67	0.33	0.68	0.32	0.60	0.40	0.54	0.46
		PCR-RFLP	0.66	0.34	0.67	0.33	0.67	0.33	0.50	0.50
TAS2R16	18:g.25883684A>T	Ion Torrent	0.59	0.41	0.59	0.41	1.00	-	0.82	0.18
		PCR-RFLP	0.57	0.43	0.56	0.44	1.00	-	0.65	0.35
TAS2R39	18:g.7358957T>C	Ion Torrent	0.62	0.38	0.65	0.35	0.12	0.88	0.32	0.68
		PCR-RFLP	NA	NA	NA	NA	NA	NA	0.28	0.72
TAS2R39	18:g.7359059A>C	Ion Torrent	0.09	0.91	0.29	0.71	0.11	0.89	-	1.00
		PCR-RFLP	0.13	0.87	0.27	0.73	NA	NA	NA	NA

[1]The first allele is the reference (Ref) allele, the second allele is the alternative (Alt) allele.

[2] LW (+), 50 Italian Large White pigs with positive back fat thickness estimated breeding value; LW (-), 50 Italian Large White pigs with negative back fat thickness estimated breeding value; WB, 10 wild boars; CA, 10 Casertana; NA: not amplified.

[3] A few differences between the two methods were more evident when estimated allele frequencies were obtained in DNA-pools with a low number of pigs (e.g. 10). This is probably due to technical errors that might be amplified when a small number of individuals are used in the construction of the pools. Differences were $\leq \pm 0.04$ in the DNA pools with 50 pigs.

Table S5. Single nucleotide polymorphisms with $P_{\text{nominal value}}<0.05$ (chi square test) with corresponding Proportion of False Positives (P_{PFP}) in the comparison of allele frequencies between the two groups of Italian Large White with extreme divergent estimated breeding values for back fat thickness.

Gene symbol	**dbSNP ID and/or SNP effect/position**	**SNP position**	$P_{nominal\ value}$	P_{PFP}
TAS2R39	rs342835508 (p.N71T)	18:7359059	0.0006	0.0451
TAS2R38	rs337735554 (p.I277M)	18:8358348	0.0023	0.0774
TAS2R39	rs326928677 (5'-UTR)	18:7358549	0.0032	0.0845
TAS2R39	rs342293575 (5'-UTR)	18:7358453	0.0067	0.120
TAS2R39	rs334394369 (5'-UTR)	18:7358501	0.0082	0.122
TAS2R39	rs331832991 (p.R15Q)	18:7358891	0.0117	0.143
TAS2R39	rs81209906 (synonymous)	18:7359534	0.0149	0.156
TAS2R38	rs344498111 (synonymous)	18:8358067	0.0221	0.203
TAS2R39	rs339810026 (p.A42V)	18:7358972	0.0262	0.208
TAS2R4	Novel, ss1971458513 (p.V139F)	GL892960.2:42100	0.0283	0.213
TAS2R38	rs327214552 (synonymous)	18:8358303	0.0333	0.220
TAS2R16	rs341414049 (synonymous)	18:25884087	0.0360	0.222
TAS2R10	Novel, ss1971458516 (synonymous)	5:63965817	0.0409	0.231

2.3. Deconstructing the pig genome-metabolome functional interactions: association studies and a nutrigenetic experiment in pigs

L. Fontanesi[1], S. Bovo[1], G. L. Mazzoni[1], G. Schiavo[1], A. Ribani[1], V. J. Utzeri[1], F. Fanelli[2], F. Bertolini[1], S. Dall'Olio[1], M. Gallo[3], L. Buttazzoni[4], G. Galimberti[5], D. G. Calò[5], M. Mezzullo[2], P. L. Martelli[6], R. Casadio[6], U. Pagotto[2] and P. Bosi[1]

[1]Department of Agricultural and Food Sciences, University of Bologna, Italy,
[2] Department of Surgical and Medical Sciences, Endocrinology Unit, University of Bologna, Italy,
[3]Associazione Nazionale Allevatori Suini, Roma, Italy,
[4]Consiglio per la Ricerca e la Sperimentazione in Agricoltura - Centro di Ricerca per la Produzione delle Carni e il Miglioramento Genetico, Monterotondo Scalo (RM), Italy,
[5]Department of Statistical Sciences “Paolo Fortunati”, University of Bologna, Italy,
[6]Biocomputing Group, University of Bologna, Italy

Paper in preparation

ABSTRACT

In this study we present an innovative genomics-metabolomics approach in pigs. A total of about 1500 performance tested pigs (900 Italian Large White and 400 Italian Duroc pigs) have been genotyped with the Illumina PorcineSNP60 BeadChip. Blood was collected at slaughtering for haematological analyses. Plasma was used for the determination of about 200 metabolites using mass spectrometry. Performance and carcass traits were measured on the same animals. Heritability of metabotypes and haematological parameters ranged from 0.04 to 0.93. A Graphical Gaussian Model (GGM) was generated including partial correlation coefficients. The structure of the obtained network represented clearly the general biological relationships among metabolites. GGM analyses suggested that pig metabolism is very close to the human metabolism. Genome wide association studies identified a large number of significant regions affecting metabotypes in both breeds, partially overlapping comparing the two groups of pigs, suggesting that common genetic determinants and, on the other hand, different genetic factors are acting in Italian Large White and Italian Duroc populations to shape their breed-defined metabolomic profiles. One of the most significant results was observed for the level of kynurenine on porcine chromosome 10 in the correspondence of the kynurenine 3-mono oxygenase gene (*KMO*). The expression of this gene was investigated in the liver of piglets having different genotypes at the *KMO* gene suggesting that different activities in the KMO enzyme might be due to different affinities or functionality of the enzyme and not due different expression levels of this gene. Metabotypes from metabolite families were influenced by the same genomic regions providing useful information to indirectly disentangle unknown metabolomic pathways. Finally, genome wide association studies basing on metabolites and the integration of omics approaches could open interesting applications in pig nutrigenetics, for which we here present the first pilot experiment.

Keywords: metabolomics, phenomics, systems biology, genome wide association study

Introduction

Holistic approaches based on the integration of different omics technologies are clarifying the biological processes underlying many phenotypic and pathological traits in humans and in many other species (Mooser and Ordovas, 2003). Metabolomics has been changing the way in which differences among animals can be investigated. Metabolomics can detect and quantify hundreds of metabolites that constitute internal (or molecular) phenotypes, providing information on the metabolic state of the animals that is influenced by genetic and environmental factors (Dettmer et al., 2006). Metabotypes are referred as phenotypes defined by the level of metabolites in a biological fluid.

In livestock, high throughput genotyping platforms, such as Illumina and Affymetrix, have changed the possibility to dissect genetic variability of performance traits applying genome wide association studies (GWAS). In pigs, GWASs were mainly conducted for a few performance traits (e.g. Fontanesi et al., 2012; Fowler et al., 2013; Diniz et al., 2014; Qiao et al., 2015).

However, despite this approach has produced a quite large number of single nucleotide polymorphisms (SNPs) associated to the target traits, only in few cases significant SNPs could be directly linked to a biological process explaining variability of the investigated parameter. Therefore it seems that a quite large distance exists between the genotype space (a portion of the whole genome information of the animals) and the final production traits that dilutes the effects of the markers reducing the proportion of variability explained in GWAS, usually conducted with a relatively low number of animals.

Metabolomics is the study of a large range of metabolites generated from the characterization of biological samples (Suhre and Gieger, 2012). Metabolites represent phenotypes (metabotypes) that are the direct products of the activities of enzymes included in all metabolic pathways. Therefore, they represent internal (or molecular) phenotypes that can be used to dissect more complex phenotypes like performance traits in livestock (Houle et al., 2010).

In this study we present an innovative genomics-metabolomics approach in two populations of highly phenotyped pigs, Italian Large White and Italian Duroc pigs. This dataset provided different levels of information for GWAS with about 230 different traits (performance and carcass traits, haematological parameters and metabotypes). Metabotype data were also used to model a systems biology approach that identified close metabolic similarities between humans and pigs.

Moreover, we present the first nutrigenetic pilot study of integration of omics approaches in a small population of 24 post-weaning Italian Large White piglets from the same litters balanced inside for the genotypes of a significant SNP present in the Illumina PorcineSNP60 BeadChip (Ramos et al., 2009) near

to kynurenine 3-monooxygenase (KMO) gene, involved in tryptophan (Trp) catabolism in mammals and associated with kynurenine levels in pig plasma (Bovo et al., 2016). Tryptophan (Trp) catabolism in mammals is fundamental for the immune system response to inflammation and other important physiological processes, including the conversion of Trp in essential biochemical active compounds (Moroni, 1999). This pilot study is the first application of nutrigenetic concepts in livestock.

Materials and Methods

Animals. About 900 performance tested Italian Large White pigs and about 400 Italian Duroc pigs were included in this study, for a total of 1500 animals. These pigs were from the national selection program and are individually performance tested at the Central Test Station of the National Pig Breeder Association (ANAS) for the genetic evaluation of a boar from the same litter (sib-testing). Pigs were slaughtered in a commercial abattoir in different groups of 30-60 pigs. For nutrigenetic experiment we reared 24 Italian Large White post-weaning piglets from the same litters balanced inside for the KMO genotype (11, 12, 22). After a treatment with a normal meal and dietary Trp load for all the three groups of piglets according to their genotype, we collected whole blood and, after slaughtering, liver tissues to perform DNA and RNA analyses.

Metabolomics data. Several performance traits (average daily gain and feed gain ratio), carcass traits (ham weight, backfat thickness, weight of several cuts) were determined on all animals in vivo or after slaughtering.
In addition, blood and livers were collected at slaughtering for biochemical, haematological and molecular analyses. Plasma was used for the determination of 186 metabolites using a combined Direct Flow Injection (DFI-) and liquid chromatography (LC-) coupled with tandem mass spectrometer (LC-MS/MS) Triple Quadrupole. Metabolite data were filtered using inter- and intra-plates coefficient of variation <0.20. The normalization of metabolite and haematological data was performed with a BoxCox transformation using a lambda which presented the best log likelihood.

DNA and RNA analyses. DNA was extracted from pig whole blood and liver with standard protocols or using the Wizard Genomic Purification Kit (Promega, Madison, Wisconsin, USA). Pigs have been genotyped with the Illumina PorcineSNP60 BeadChip (Ramos et al., 2009). After that, we sequenced the whole KMO pig gene designing 16 primer pairs covering the UTRs and all the 17 exons of the gene in a subset of Italian Large White and Italian Duroc genotyped pigs, using the next generation sequencing

(NGS) platform Ion Torrent Personal Genome Machine (PGM, Thermo Fisher Scientific, Waltham, Massachussets, USA). Following manufacturer instructions, we constructed 3 libraries, each representing a genotype, using different barcodes. Barcoded fragments have been pooled together, clonally amplified with emulsion PCR, purified and sequenced using the Ion PGM Sequencing kit with an Ion 316 chip (Thermo Fisher Scientific).

Total RNA was extracted from piglets' livers with the RNeasy Mini Kit and the RNeasy MinElute CleanUp Kit (QIAGEN, Venlo, The Netherlands) including smallRNAs. A RNA-seq experiment has been performed using the NGS platform Illumina Hi-Seq 2500 (Illumina, San Diego, California, USA). Following the manufacturer protocols, 3 barcoded libraries have been sequenced, each representing one KMO genotype, in order to detect differential gene expression patterns according to the different genotypes of KMO gene. Validations of KMO expression levels have been performed with Real Time PCR using the ABI PRISM 7000 instrument (Thermo Fisher Scientific, Waltham, Massachussets, USA) using the Kapa SYBR Fast qPCR MasterMix Kit (Roche, Basel, Switzerland). We analyzed standard curves using the $-2\Delta\Delta C_t$ method (Schmittgen and Livak, 2008), using reference genes according to Park et al (2015).

Metabolomics statistical analyses.

The Pearson phenotypic correlation coefficients among metabotypes and performance and carcass traits were calculated with the CORR procedure of SAS (SAS Inst. Inc., Cary, NC). Variance components, genetic parameters and their standard errors were estimated by using VCE software (Neumaier and Groeneveld, 1998; Groeneveld et al., 2010). Bivariate mixed linear animal models, for each parameter with backfat thickness, considered the significant effects of date of slaughtering (26 levels), sex (2 levels), and carcass weight (covariates). Animal (22785 pigs) and residual random effects were assumed to follow normal distributions with zero mean.

Gaussian Graphical Model (GGM) was obtained with metabolomic data. GGM is based on pairwise Pearson correlation coefficients corrected for the correlations with the other metabolites. The partial correlation coefficients were computed with R package "*corpcor"*. Then the list of all the correlations coefficient was extracted from the matrix and used as input in Cytoscape (Version. 3.0.1).

GWAS was carried out with GEMMA (Zhou and Stephens, 2012) after filtering for minor allele frequency <0.05 and Hardy Weinberg equilibrium <0.001. Different covariates were included in the model according to the traits (sex, weight, slaughtering date (range: 1-25), kit plate (range:1-14) and the centered genomic matrix, to exclude the stratification of the population due to the relatedness.

NGS analyses. Reads obtained by re-sequencing of KMO pig gene were filtered, trimmed and aligned to the reference sequences using TMAP aligner, included in the Ion Torrent Suite 2.2 software (Thermo Fisher Scientific). The filtered bam files were then processed with SAMTOOLS (Li et al., 2009), using mpileup function for SNPs detection. The effect of each SNP has been evaluated using the Variant Effect Predictor (VEP) and all missense mutations have been analyzed using the online SIFT tool (Kumar et al., 2009). For total RNA-seq reads, we performed preliminary bioinformatics analyses of transcripts filtering and trimming low-quality sequences: then we aligned on pig genome reference the remaining reads using TopHat algorithm and we performed the differentially expression analysis using Cufflinks implemented in Galaxy platform (Trapnell et al., 2009; 2012).

Results and Discussion

Correlations and heritability. After quality control and filtering 113 metabolites were selected for further analyses. Figure 1 shows the Pearson phenotypic correlation coefficients among metabotypes and production and carcass traits. Metabolites within classes were usually more correlated to each other than with metabolites of other classes. Heritability of metabolites and haematological parameters within the same classes varied substantially (Table 1). Heritability of backfat thickness was similar to previous estimates for this trait (0.60).

Gaussian Graphical Model. A GGM was generated including all partial correlation coefficients (PCC) that were above 0.21. This threshold corresponded to a Bonferroni corrected threshold of 0.05. The structure of the obtained network represented clearly the general biological relationships among metabolites (within classes) and among different classes of metabolites and included several substructures (Figure 2). The highest PCCs were obtained between metabolites that are very close in terms of positions in the metabolic pathway. In particular, SM C18:0 and SM C18:1 that showed a PCC of 0.77 are separated to each other by just one enzymatic reaction. These results, together with many other PCC between different pairs of metabolites (data not shown) confirmed in pigs the metabolic relationships and pathways already described in humans (Krumsiek et al., 2011). GGM analyses suggested that pig metabolism is very close to the human metabolism.

Genome Wide Association Studies. GWAS was carried out for a total of 153 traits (113 metabotypes, 33 haematological parameters and 7 production and carcass traits. Significant SNPs ($P<0.10$, Bonferroni corrected) were obtained for 20 metabotypes, 6 haematological parameters and 4 production/carcass traits.

In particular, for most metabotypes, significant SNPs were close or within genes directly involved in the catabolic or anabolic pathways of the targeted metabolites. A few of these markers were associated (P nominal value <0.01) with production and carcass traits.

One of the most important peaks in the GWAS in both breeds was observed on porcine chromosome 10 for the level of kynurenine in the correspondence of the KMO gene (Figure 3). KMO encodes for an enzyme that is involved in the tryptophan catabolism and transforms L-kynurenine in 3-hydroxykynurenine. Two alleles were identified in the tag SNPs. One that increases the level of kynurenine and another one that decreases the level of this metabolite. Considering the role that KMO play in the metabolism of an essential amino acid, the porcine gene encoding for this enzyme might be considered an interesting candidate for nutrigenetic studies.

Nutrigenetic experiment and KMO gene expression analysis. In order to investigate if KMO gene was differentially expressed in piglets' liver with different genotypes, we analyzed their trascriptome using RNA-seq approach: from each barcoded library (representing KMO genotypes 11, 12, 22), we obtained more than 70 million of reads and we detected the same pattern of expression of KMO gene. We confirmed these results using RT-qPCR for which we did not find any difference in KMO expression in piglets with different genotypes ($-2\Delta\Delta C_t$ value: 1.3124).

From the whole *KMO* gene sequencing we detected 215 variants, of which 1 splice region variant, 5 missense and 2 synonymous SNPs and a 3 bp exonic INDEL which is in linkage disequilibrium with the SNP present in Illumina PorcineSNP60 BeadChip: this INDEL may leads to a conformational change in the enzyme structure, probably associated with its efficiency in catalyzing kynurenine (data not shown as they are under patent evaluation).

Combining gene expression analysis and the structural deduced information on the encoded protein, it could be possible to speculate that the different level of kynurenine affected by the two alleles at the *KMO* gene might be due to different affinities or functionality of the alternative protein forms and not by different gene expression levels of the two gene alleles at this locus.

Conclusion

This study reports for the first time the analysis of heritability of a large number of metabotypes in pigs and describes a systems biology analysis based only on metabolomics data (GGM). GGM identified biological relationships between metabolites already described in humans supporting indirectly the quality of the data we obtained using a targeted metabolomic approach. GWAS identified genetic variation in

genes directly involved in the metabolism of several metabolites that could open interesting applications in pig nutrigenomics. GWAS using metabolomics data can help to dissect the biological complexity of performance traits in livestock.

In conclusion, our nutrigenetic experiment using KMO genotypes as key study is the first approach using different omics technologies in livestock: the integration of these sciences will allow to develop nutrigenetic and feeding precision strategies in animal breeding.

Acknowledgements

This study was supported by Italian MiPAAF (INNOVAGEN project) and AGER-HEPIGET (Grant no. 2011-0279) funds.

References

Bovo, S., Mazzoni, G., Galimberti, G., Calò, D. G., Fanelli, F., Mezzullo, M., ... & Dall'Olio, S. (2016). Metabolomics evidences plasma and serum biomarkers differentiating two heavy pig breeds. *animal*, 1-8.

Dettmer, K., Aronov, P. A., & Hammock, B. D. (2006). Mass spectrometry-based metabolomics. *Mass spectrometry reviews*, *26*(1), 51-78.

Diniz, D. B., Lopes, M. S., Broekhuijse, M. L. W. J., Lopes, P. S., Harlizius, B., Guimarães, S. E. F., ... & Silva, F. F. (2014). A genome-wide association study reveals a novel candidate gene for sperm motility in pigs. *Animal reproduction science*, *151*(3), 201-207.

Fontanesi, L., Schiavo, G., Galimberti, G., Calò, D. G., Scotti, E., Martelli, P. L., ... & Russo, V. (2012). A genome wide association study for backfat thickness in Italian Large White pigs highlights new regions affecting fat deposition including neuronal genes. *BMC genomics*, *13*(1), 583.

Fowler, K. E., Pong-Wong, R., Bauer, J., Clemente, E. J., Reitter, C. P., Affara, N. A., ... & Griffin, D. K. (2013). Genome wide analysis reveals single nucleotide polymorphisms associated with fatness and putative novel copy number variants in three pig breeds. *BMC genomics*, *14*(1), 784.

Groeneveld, L. F., Lenstra, J. A., Eding, H., Toro, M. A., Scherf, B., Pilling, D., ... & Weigend, S. (2010). Genetic diversity in farm animals–a review. *Animal genetics*, *41*(s1), 6-31.

Houle, D., Govindaraju, D. R., & Omholt, S. (2010). Phenomics: the next challenge. *Nature reviews genetics*, *11*(12), 855-866.

Krumsiek, J., Suhre, K., Illig, T., Adamski, J., & Theis, F. J. (2011). Gaussian graphical modeling reconstructs pathway reactions from high-throughput metabolomics data. *BMC systems biology*, *5*(1), 21.

Kumar, P., Henikoff, S., & Ng, P. C. (2009). Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. *Nature protocols*, *4*(7), 1073-1081.

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., ... & Durbin, R. (2009). The sequence alignment/map format and SAMtools. *Bioinformatics*, *25*(16), 2078-2079.

Mooser, V., & Ordovas, J. M. (2003). 'Omic'approaches and lipid metabolism: are these new technologies holding their promises?. *Current opinion in lipidology*, *14*(2), 115-119.

Moroni, F. (1999). Tryptophan metabolism and brain function: focus on kynurenine and other indole metabolites. *European journal of pharmacology*, *375*(1), 87-100.

Neumaier, A., & Groeneveld, E. (1998). Restricted maximum likelihood estimation of covariances in sparse linear models. *Genetics Selection Evolution*, *30*(1), 3-26.

Park, S. J., Kwon, S. G., Hwang, J. H., Park, D. H., Kim, T. W., & Kim, C. W. (2015). Selection of appropriate reference genes for RT-qPCR analysis in Berkshire, Duroc, Landrace, and Yorkshire pigs. *Gene*, *558*(1), 152-158.

Qiao, R., Gao, J., Zhang, Z., Li, L., Xie, X., Fan, Y., ... & Huang, L. (2015). Genome-wide association analyses reveal significant loci and strong candidate genes for growth and fatness traits in two pig populations. *Genetics Selection Evolution*, *47*(1), 17.

Ramos, A. M., Crooijmans, R. P., Affara, N. A., Amaral, A. J., Archibald, A. L., Beever, J. E., ... & Hansen, M. S. (2009). Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *PloS one*, *4*(8), e6524.

Schmittgen, T. D., & Livak, K. J. (2008). Analyzing real-time PCR data by the comparative CT method. *Nature protocols*, *3*(6), 1101-1108.

Suhre, K., & Gieger, C. (2012). Genetic variation in metabolic phenotypes: study designs and applications. *Nature reviews genetics*, *13*(11), 759-769.

Trapnell, C., Pachter, L., & Salzberg, S. L. (2009). TopHat: discovering splice junctions with RNA-Seq. *Bioinformatics*, *25*(9), 1105-1111.

Trapnell, C., Roberts, A., Goff, L., Pertea, G., Kim, D., Kelley, D. R., ... & Pachter, L. (2012). Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks. *Nature protocols*, *7*(3), 562-578.

Zhou, X., & Stephens, M. (2012). Genome-wide efficient mixed-model analysis for association studies. *Nature genetics*, *44*(7), 821-824.

Table 1. Summary of heritability (h^2, minimum and maximum) of metabolites, haematological parameters and backfat thickness

Classes	h^2 min	h^2 max	# of metabolites/ parameters
Acylcarnitines	0.07	0.34	10
Amino acids	0.07	0.45	20
Biogenic amines	0.21	0.50	9
Hexoses	0.14	0.14	1
PC ae	0.05	0.72	27
PC aa	0.16	0.73	27
Lyso PC a	0.05	0.48	9
Sphingomyelines	0.14	0.47	10
Haematological parameters	0.04	0.93	33
Backfat thickness	0.60		

Figure 1. Heat map of Pearson correlations between production traits and metabolites. Hottest correlations are indicated in red-yellow.

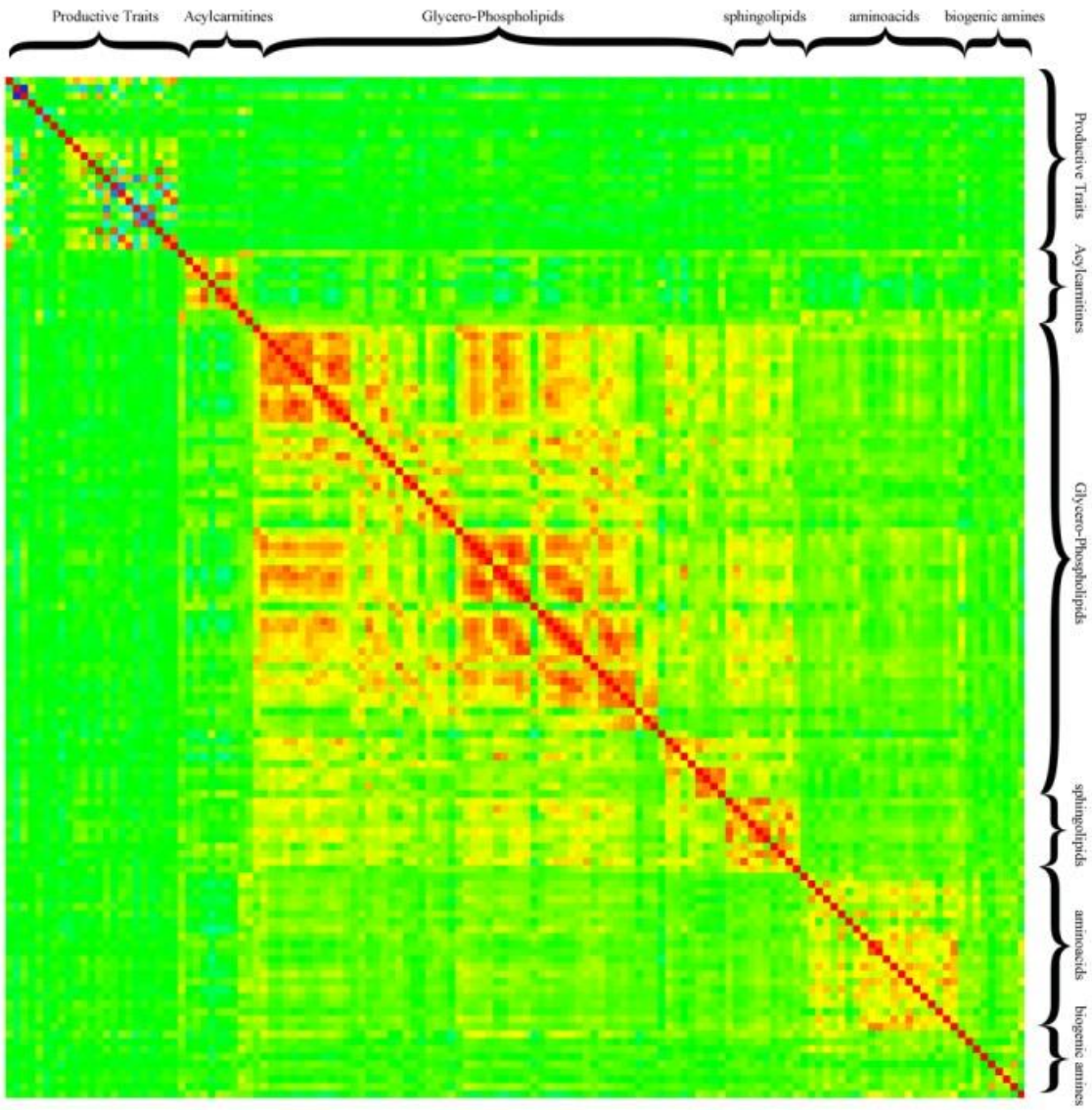


Figure 2. Gaussian Graphical Model obtained using pig metabolomic data.

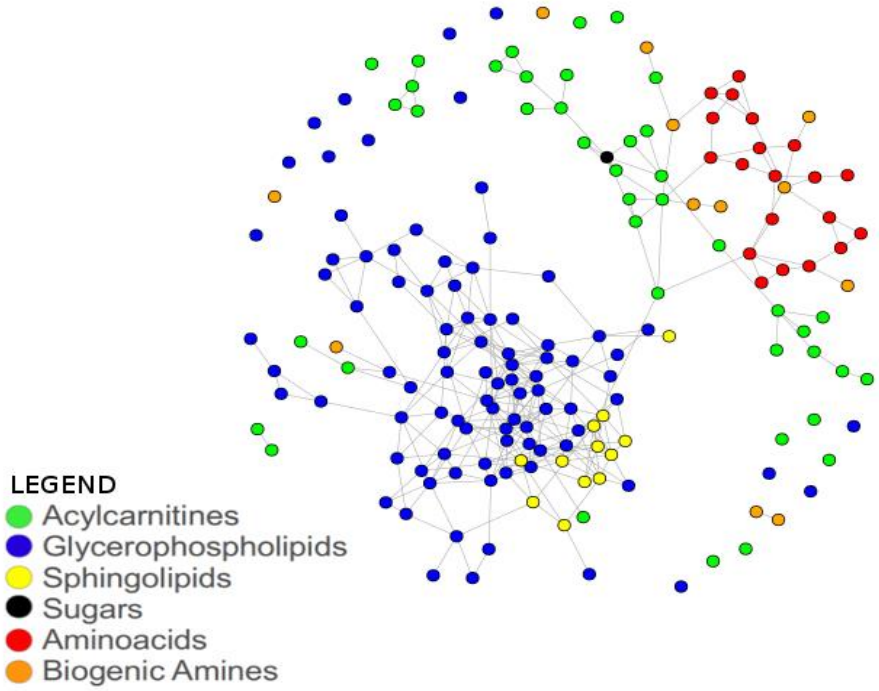
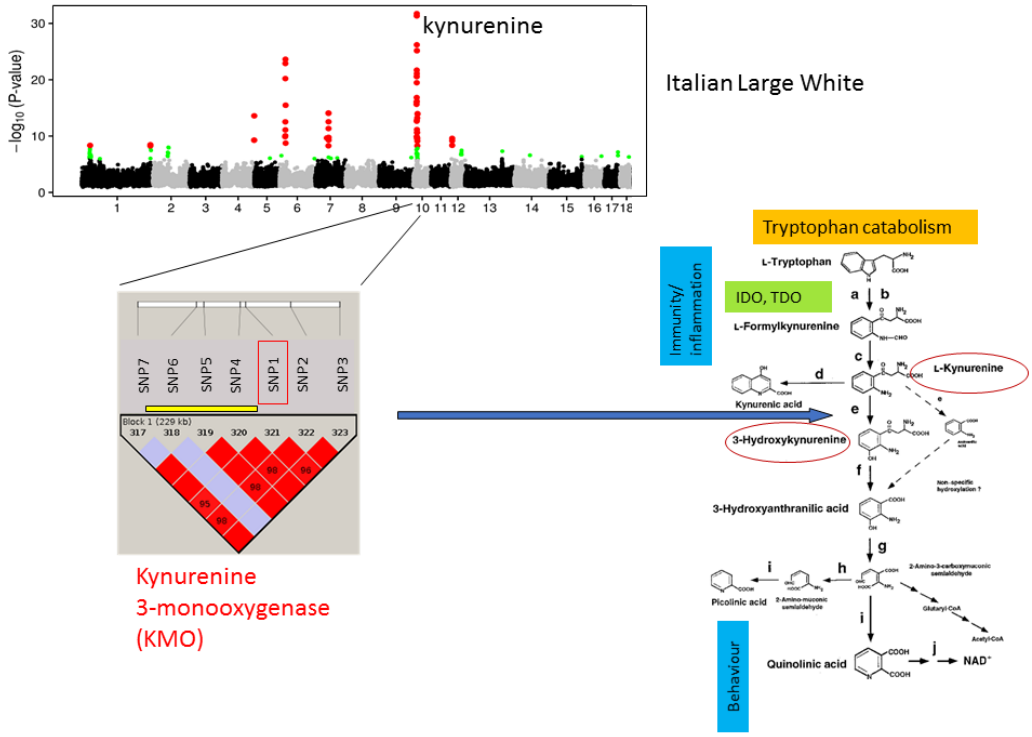


Figure 3. Manhattan plot showing genome wide association results obtained in Italian Large White pigs for a subset of metabolites (i.e. overlapping results for more than one metabolite). The peak evidenced is for the level of kynurenine in the correspondence of the KMO gene. The KMO enzyme is involved in the catabolism of the tryptophan as evidenced in the picture.



3. General Conclusions

In this Thesis we applied different omics technologies for the identification of metabolic and productive markers in several pig populations, in order to develop nutrigenetic strategies for pig breeding. In particular, we performed association studies focusing on Average Daily Gain (ADG) and Back Fat Thickness (BFT) estimated breeding value (EBV) traits mainly in Italian Heavy Pigs, like Italian Large White pigs. These traits are considered as the most important in terms of productivity in pig production chains and involve physiological processes such as growth rate and fat deposition. With a re-sequencing approach using the next generation semiconductor-based sequencing platform Ion Torrent PGM, we characterized 10 porcine taste receptors genes, 9 bitter taste receptors genes (*TAS2R* family: *TAS2R1, TAS2R3, TAS2R4, TAS2R7, TAS2R9, TAS2R10, TAS2R16, TAS2R38* and *TAS2R39*) and the long chain fatty acid receptor *GPR120* gene, all involved in the taste system affecting food preferences and feed intake in different pig populations; we included also two groups (50 + 50 animals) from Italian Large White pigs with extreme and divergent estimated breeding value (EBV) for ADG and BFT. Three single nucleotide polymorphisms (SNPs) were found in *GPR120* gene showing significant differences of allele and genotype frequencies distribution between the extreme ADG-EBV groups ($P<0.001$). In *TAS2R* genes, a total of 125 SNPs have been detected, of which 37 missense mutations and among these three of them that can have important effects on bitter taste receptors functionality (base on *in silico* predictions). Moreover, we identified a total of 13 SNPs, 7 in *TAS2R39*, 3 in *TAS2R38* and one in *TAS2R4*, *TAS2R10* and *TAS2R16* each, significantly associated with BFT trait ($P_{nominal\ value}<0.05$; Chi square tests). Our results, using a methodological approach based on next generation sequencing, provided a better understanding of the complexity of taste perception in pigs but also in humans, considering that in the recent years they have been successfully used as animal model in this field. Moreover, significant associations with production and performance traits like ADG and BFT can open the opportunities to develop nutrigenetic approaches for pig breeding and nutrition.

The last study we carried out aimed to integrate target re-sequencing NGS genes, expression analyses and metabolomics approaches for the first time in livestock. We used a target metabolomic approach on two pig populations, Italian Large White and Italian Duroc pigs, and we detected a large number of metabotypes identifying biological relationships between metabolites. We performed a GWAS detecting genetic variation in genes involved in metabolism suggesting also in this case opportunities for application in pig nutrigenetics and nutrigenomics. Finally, we performed the first nutrigenetic experiment in pigs analysing genomic sequences, transcripts and related metabolites of an enzyme involved in the Trp catabolism, integrating these different kind of data.

In conclusion, the importance of the characterization of new phenotypes including internal phenotypes (such as new metabotypes or other bioanalytes) and external ones (e.g. food preference tests) is one of the most prevalent direction undertaken by the so called phenomics, toward high-throughput phenotyping. Moreover, future studies will be needed to better understand genotypes effects of taste genes on food preferences in order to increase feed intake in pigs as well as to clarify the association between taste receptors variability and fat deposition trait.

From a methodological point of view, the integration of genomics, transcriptomics and metabolomics approaches in order to better investigate complexity traits is becoming a feasable solution thanks to the new technological advances and the reduction of costs.

Whole genome association studies might be planned following the same strategy that we have proposed in the second paper. Instead of working in a targeted approach, DNA-pool seq could be used directly in an association study by sequencing at high depth DNA pool constructed from pigs with extreme and divergent trait values.

Finally, the development of precision feeding strategies focused on specific amino acid needs of the animals according to their genotype in genes involved in the amino acid metabolism pathways would be one of the envisaged perspectives of application in pig nutrigenetics and, more generally, in livestock nutrition.

4. General References

Alonso, A., S. Marsal, and A. Julià. (2015). Analytical methods in untargeted metabolomics: State of the art in 2015. *Front. Bioeng. Biotechnol.* 3:23.

Bachmanov, A.A. & Beauchamp, G.K. (2007) Taste receptor genes. *Annu Rev Nutr* 27, 389–414.

Barretto, R. P., Gillis-Smith, S., Chandrashekar, J., Yarmolinsky, D. A., Schnitzer, M. J., Ryba, N. J., & Zuker, C. S. (2015). The neural representation of taste quality at the periphery. *Nature*, 517(7534), 373-376.

Bendixen, E., Danielsen, M., Larsen, K., & Bendixen, C. (2010). Advances in porcine genomics and proteomics—a toolbox for developing the pig as a model organism for molecular biomedical research. *Briefings in functional genomics*, 9(3), 208-219.

Beyer, A., Bandyopadhyay, S. & Ideker, T. (2007) Integrating physical and genetic maps: from genomes to interaction networks. *Nature Rev. Genet.* **8**, 699–710.

Bino, R.J., Hall, R.D., Fiehn, O., Kopka, J., Saito, K., Draper, J., Nikolau, B.J., Mendes, P., Roessner-Tunali, U., Beale, M.H., Trethewey, R.N., Lange, B.M., Wurtele, E.S., Sumner, L.W. (2004). Potential of metabolomics as a functional genomics tool. *Trends Plant Sci* 9:418–425.

Brennan, R. O., & Mulligan, W. C. (1975). Nutrigenetics: New Concepts for Relieving Hypoglycemia: Case Studies. *M Evans & Company*.

Buitenhuis, A.J., U.K. Sundekilde, N.A. Poulsen, H.C. Bertram, L.B. Larsen, and P. Sørensen. 2013. Estimation of genetic parameters and detection of quantitative trait loci for metabolites in Danish Holstein milk. J. Dairy Sci. 96:3285–3295.

Castillo, C., Benedito, J. L., Méndez, J., Pereira, V., Lopez-Alonso, M., Miranda, M., & Hernández, J. (2004). Organic acids as a substitute for monensin in diets for beef cattle. *Animal Feed Science and Technology*, *115*(1), 101-116.

Chagas, L. M., Bass, J. J., Blache, D., Burke, C. R., Kay, J. K., Lindsay, D. R., ... & Roche, J. R. (2007). Invited review: New perspectives on the roles of nutrition and metabolic priorities in the subfertility of high-producing dairy cows. *Journal of dairy science*, *90*(9), 4022-4032.

Chaudhari, N. & Roper, S.D. (2010) The cell biology of taste. *Journal of Cell Biology* 190, 285–96.

Clop, A., Sharaf, A., Castelló, A., Ramos-Onsins, S., Cirera, S., Mercadé, A., ... & van As, P. (2016). Identification of protein-damaging mutations in 10 swine taste receptors and 191 appetite-reward genes. *BMC genomics*, *17*(1), 685.

Crick, F. (1970). Central dogma of molecular biology. *Nature*, *227*(5258), 561-563.

Danilova, V., Roberts, T. & Hellekant, G. (1999) Responses of single taste fibers and whole chorda tympani and glossopharyngeal nerve in the domestic pig. *Sus scrofa*. *Chem Senses* 24, 301–316.

De Jager, N., Zhan, M., Rzepus, M., Roura, E., (2013). Towards defining the taste receptor repertoire in the pig. In: Pluske, J.R., Pluske, J.M. (Eds.), "Manipulating Pig Production XIV", *APSA Conference, Australasian Pig Science Association*, Melbourne, Australia, pp. 47.

Depoortere, I. (2014). Taste receptors of the gut: emerging roles in health and disease. *Gut*, *63*(1), 179-190.

Dettmer, K., Aronov, P. A., & Hammock, B. D. (2007). Mass spectrometry-based metabolomics. *Mass spectrometry reviews*, *26*(1), 51-78.

Donaldson, M. S. (2004). Nutrition and cancer: a review of the evidence for an anti-cancer diet. *Nutrition journal*, *3*(1), 19.

Ettle, T., Rademacher, M., Htoo, J.K., Roth, F.X., (2010). Dietary preference for methionine sources in weaned pigs. *Anim. Feed Sci.Technol.* 155, 201-205.

Fairbanks, D. J., and W. R. Andersen (1999). Genetics: The Continuity of Life. *Pacific Grove*, CA: Brooks/Cole Publishing Company.

Fenech, M., El-Sohemy, A., Cahill, L., Ferguson, L. R., French, T. A., Tai, E. S., ... & Buckley, M. (2011). Nutrigenetics and nutrigenomics: viewpoints on the current status and applications in nutrition research and practice. *Journal of nutrigenetics and nutrigenomics*, *4*(2), 69-89.

Fontanesi, L. (2016). Metabolomics and livestock genomics: Insights into a phenotyping frontier and its applications in animal breeding. *Animal Frontiers*, *6*(1), 73-79.

Fontanesi, L., Bertolini, F., Scotti, E., Schiavo, G., Colombo, M., Trevisi, P., Ribani, A., Buttazzoni, L., Russo, V., Dall'Olio, S., (2015). Next generation semiconductor based-sequencing of a nutrigenetics target gene (*GPR120*) and association with growth rate in Italian Large White pigs. *Anim. Biotechnol*. 26, 92-97.

Fontanesi, L., G. Schiavo, S. Bovo, G. Mazzoni, F. Fanelli, A. Ribani et al. (2015). Dissecting complex traits in pigs: Metabotypes illuminate genomics for practical applications. In: *Book of Abstracts of the 66th Annual Meeting of the European Federation of Animal Science*, Warsaw, Poland, 31 August-4 September. p. 152.

Fontanesi, L., S. Bovo, G. Mazzoni, A.B. Samorè, G. Schiavo, E. Scotti et al. (2014). Genome wide perspective of genetic variation in pig metabolism and production traits. Manuscript no. 359. *Proc. of the 10th World Congr. Genet. Appl. Liv. Prod.*, 17-22 Aug. 2014, Vancouver BC, Canada.

Gandarillas, M., & Bas, F., (2009) The domestic pig (*Sus scrofa domestica*) as a model for evaluating nutritional and metabolic consequences of bariatric surgery practiced on morbid obese humans. *Cienc Investig Agrar* 36, 163–176.

Garg, R., Sharma, N., & Jain, S. K. (2014). Nutrigenomics and Nutrigenetics: Concepts and Applications in Nutrition Research and Practice.

González-Alvarado, J. M., Jiménez-Moreno, E., Lázaro, R., & Mateos, G. G. (2007). Effect of type of cereal, heat processing of the cereal, and inclusion of fiber in the diet on productive performance and digestive traits of broilers. *Poultry Science*, *86*(8), 1705-1715.

Groenen, M. A., Archibald, A. L., Uenishi, H., Tuggle, C. K., Takeuchi, Y., Rothschild, M. F., Rogel-Gaillard, C., Park, C., Milan, D., Megens, H. J., Li, S., Larkin, D. M., Kim, H., Frantz, L. A., …, and Schook, L. B., (2012). Analyses of pig genomes provide insight into porcine demography and evolution. *Nature*, vol. 491, no. 7424, pp. 393-398.

Hawkins, R. D., Hon, G. C., & Ren, B. (2010). Next-generation genomics: an integrative approach. *Nature Reviews Genetics*, *11*(7), 476-486.

Houpt, T.R., Anika, S.M., Houpt, K.A., (1979). Preabsorptive intestinal satiety controls of food intake in pigs. *Am. J. Physiol.* 236, R328-337.

Junot, C., F. Fenaille, B. Colsch, and F. Bécher. (2014). High resolution mass spectrometry based techniques at the crossroads of metabolic pathways. *Mass Spectrom. Rev*. 33:471–500.

Kanehisa, M., S. Goto, Y. Sato, M. Kawashima, M. Furumichi, and M. Tanabe. (2014). Data, information, knowledge and principle: Back to metabolism in KEGG. *Nucleic Acids Res.* 42:D199–D205.

Kats, L. B., & Dill, L. M. (1998). The scent of death: chemosensory assessment of predation risk by prey animals. *Ecoscience*, *5*(3), 361-394.

Kau, A. L., Ahern, P. P., Griffin, N. W., Goodman, A. L., & Gordon, J. I. (2011). Human nutrition, the gut microbiome and the immune system. *Nature*, *474*(7351), 327-336.

Li, D. & Zhang, J. (2013) Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. *Mol Biol Evol* 31, 303–309.

Loper, H. B., La Sala, M., Dotson, C., & Steinle, N. (2015). Taste perception, associated hormonal modulation, and nutrient intake. *Nutrition reviews*, *73*(2), 83-91

Matsuo, R., (2000) Role of saliva in the maintenance of taste sensitivity. *Crit Rev Oral Biol Med* 11, 216–229.

Mooser, V., and Ordovas, J. M. (2003) 'Omic' approaches and lipid metabolism: are these new technologies holding their promises? *Curr. Opin. Lipidol.* 14, 115–119

Muller, M., and Kersten, S. (2003) Nutrigenomics: goals and strategies. *Nat. Rev. Genet.* 4, 315–322

Mutch, D. M., Wahli, W., & Williamson, G. (2005). Nutrigenomics and nutrigenetics: the emerging faces of nutrition. *The FASEB journal*, *19*(12), 1602-1616

Nelson, S.L. & Sanregret, J.D. (1997) Response of pigs to bittertasting compounds. *Chem Senses* 22, 129–132.

Nicholson, J. K., Holmes, E., Lindon, J. C., and Wilson, I. D. (2004) The challenges of modeling mammalian biocomplexity. *Nat. Biotechnol.* 22, 1268–1274

Ordovas, J. M., and Corella, D. (2004) Nutritional genomics. *Annu. Rev. Genomics Hum. Genet.* 5, 71–118

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature reviews genetics*, *12*(2), 87-98.

Pardee, A. B. (1974). A restriction point for control of normal animal cell proliferation. *Proceedings of the National Academy of Sciences*, *71*(4), 1286-1290.

Patterson, J. A., & Burkholder, K. M. (2003). Application of prebiotics and probiotics in poultry production. *Poultry science*, *82*(4), 627-631.

Patti, G.J., O. Yanes, and G. Siuzdak. (2012). Innovation: Metabolomics: The apogee of the omics trilogy. *Nat. Rev. Mol. Cell Biol.* 13:263–269.

Raamsdonk, L. M., Teusink, B., Broadhurst, D., Zhang, N., Hayes, A., Walsh, M. C., ... & Westerhoff, H. V. (2001). A functional genomics strategy that uses metabolome data to reveal the phenotype of silent mutations. *Nature biotechnology*, *19*(1), 45-50.

Ribani, A., Bertolini, F., Schiavo, G., Scotti, E., Utzeri, V. J., Dall'Olio, S., Trevisi P., Bosi P. & Fontanesi, L. (2016). Next generation semiconductor based sequencing of bitter taste receptor genes in different pig populations and association analysis using a selective DNA pool-seq approach. *Animal Genetics*.

Ripken, D., van der Wielen, N., Wortelboer, H.M., Meijerink, J., Witkamp, R.F., Hendriks, H.F., (2014). Steviol glycoside rebaudioside A induces glucagon-like peptide-1 and peptide YY release in a porcine ex vivo intestinal model. *J. Agric. Food Chem.* 62, 8365-8370.

Romanov, M. N. (1999). Goose production efficiency as influenced by genotype, nutrition and production systems. *World's Poultry Science Journal*, *55*(03), 281-294.

Roura, E., and Tedo, G., (2009). Feed appetence in pigs: an oronasal sensing perspective. In: Torrallardona, D., Roura, E. (Eds.), Voluntary Feed Intake in Pigs, *Wageningen Academic Publishers*, Wageningen, pp. 105-140.

Roura, E., Humphrey, B., Klasing, K.C., Swart, M., (2011). Is the pig a good umami sensing model for humans? A comparative taste receptor study. *Flav. Fragr. J.* 26, 282-285.

Roura, E., Koopmans, S. J., Lallès, J. P., Le Huerou-Luron, I., de Jager, N., Schuurman, T., & Val-Laillet, D. (2016). Critical review evaluating the pig as a model for human nutritional physiology. *Nutrition research reviews*, *29*(01), 60-90.

Rubin, C.J., Megens, H.J., Martinez Barrio, A., Maqbool, K., Sayyab, S., Schwochow, D., Wang, C., Carlborg, Ö., Jern, P., Jørgensen, C.B., Archibald, A.L., Fredholm, M., Groenen, M.A., Andersson, L., (2012) Strong signatures of selection in the domestic pig genome. *Proc Natl Acad Sci U S A* 109(48):19529-36.

Santos, J. E. P., Bisinotto, R. S., Ribeiro, E. S., Lima, F. S., Greco, L. F., Staples, C. R., & Thatcher, W. W. (2011). Applying nutrition and physiology to improve reproduction in dairy cattle. *Reproduction in Domestic Ruminants VII*, *67*, 387-403.

Simopoulos, A.P. (2010) Nutrigenetics/nutrigenomics. *Annu Rev Public Health* 31: 53–68.

Song, T., Peng, J., Ren, J., Wei, H.K., Peng, J., (2015). Cloning and characterization of spliced variants of the porcine G protein coupled receptor 120. *Biomed. Res. Int.* 2015, 813816.

Stein, H. H. (2002). Experience of feeding pigs without antibiotics: a European perspective. *Animal Biotechnology*, *13*(1), 85-95.

Suhre, K., and C. Gieger. (2012). Genetic variation in metabolic phenotypes: Study designs and applications. *Nat. Rev. Genet.* 13:759–769.

Tyers, M., & Mann, M. (2003). From genomics to proteomics. *Nature*, *422*(6928), 193-197.
van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. *Trends in genetics*, *30*(9), 418-426.

Verma, N., Rettenmeier, A.W. & Schmitz-Spanke, S. (2011) Recent advances in the use of *Sus scrofa* (pig) as a model system for proteomic studies. *Proteomics* 11, 776–793

Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for transcriptomics. *Nature reviews genetics*, *10*(1), 57-63.

Xia, J., and D.S. Wishart. (2010). MSEA: A web-based tool to identify biologically meaningful patterns in quantitative metabolomic data. *Nucleic Acids Res.* 38:W71–77.

Yi, L., Dong, N., Yun, Y., Deng, B., Ren, D., Liu, S., & Liang, Y. (2016). Chemometric methods in data processing of mass spectrometry-based metabolomics: a review. *Analytica chimica acta*, *914*, 17-34.

Zeisel, S. H., Allen, L. H., Coburn, S. P., Erdman, J. W., Failla, M. L., Freake, H. C., King, J. C., and Storch, J. (2001) Nutrition: a reservoir for integrative science. *J. Nutr.* 131, 1319–1321

5. Ackwnoledgments

I would like to thank my supervisor Prof. Luca Fontanesi and my colleagues Valerio Joe Utzeri and Dr. Giuseppina Schiavo for supporting me during my PhD period. I thank also Samuele Bovo (Biocomputing group) and Dr. Sara De Fanti (University of Bologna) for technical support.

Finally, I would like to thank ANAS (National Association Pig Breeders) for providing data and samples.

6. Appendix

In Appendix it is reported the list of published papers which have not been mentioned in the Thesis and all the posters presented in Scientific Congresses attended during my PhD period.

Publications

Schiavo, G., Ivett Hoffmann, O., Ribani, A., Utzeri, V.J., Ghionda, M.C., Bertolini, F., Geraci, C., Bovo, S., and Fontanesi, L. (2017). A genomic landscape of mitochondrial DNA insertions in the pig nuclear genome provides evolutionary signatures of interspecies admixture. Accepted by *DNA Research*.

Abstract

Nuclear DNA sequences of mitochondrial origin (*numts*) are derived by insertion of mitochondrial DNA (mtDNA), into the nuclear genome. In this study, we provide, for the first time, a genome picture of *numts* inserted in the pig nuclear genome. The *Sus scrofa* reference nuclear genome (Sscrofa10.2) was aligned with circularized and consensus mtDNA sequences using LAST software. A total of 430 *numt* sequences that may represent 246 different *numt* integration events (57 *numt* regions determined by at least two *numt* sequences and 189 singletons) were identified, covering about 0.0078% of the nuclear genome. *Numt* integration events were correlated (0.99) to the chromosome length. The longest *numt* sequence (about 11 kbp) was located on SSC2. Six *numts* were sequenced and PCR amplified in pigs of European commercial and local pig breeds, of the Chinese Meishan breed and in European wild boars. Three of them were polymorphic for the presence or absence of the insertion. Surprisingly, the estimated age of insertion of two of the three polymorphic *numts* was more ancient than that of the speciation time of the *Sus scrofa*, supporting that these polymorphic sites were originated from interspecies admixture that contributed to shape the pig genome.

Motta, V., Trevisi, P., Bertolini, F., Ribani, A., Schiavo, G., Fontanesi, L., and Bosi, P. (2017). Exploring gastric bacterial community in young pigs. *PloS one*, *12*(3), e0173029.

Abstract

Microbiota plays an important role in the homeostasis of the gastrointestinal tract. Understanding the variations of the commensal microbiota composition is crucial for a more efficient control of enteric infectious diseases and for the reduction of the use of antibiotics in animal production, which are the main points of interest for improved animal healthcare and welfare and for consumer health protection. Even though the intestinal microbiota has been extensively studied, little is known about the gastric microbiota. This pilot study was aimed at a descriptive analysis of the gastric microbiota in healthy pigs and at the identification of any differences among four potentially distinct microbial niches in the stomach. Gastric mucosal samples from the oxyntic area, the pylorus and the gastric groove, and a sample of gastric contents were collected from four healthy weaned pigs. Bacterial DNA was isolated and extracted from each sample and amplicons from the V6 region of the 16S rRNA gene were sequenced using Ion Torrent PGM. The data were analysed by an "unsupervised" and a "supervised" approach in the Ribosomal Database Project (RDP) pipeline. Proteobacteria was the dominant phylum in all the samples. Differences in bacterial community composition were found between mucosal and content samples (one-way ANOSIM pairwise post hoc test, $p < 0.05$); instead, the different mucosal regions did not show differences between them. The mucosal samples were characterised by *Herbiconiux* and *Brevundimonas*, two genera which include cellulolytic and xylanolytic strains. Nevertheless, additional larger trials are needed to support the data presented in this pilot study and to increase the knowledge regarding the resident microbiota of the stomach.

Fontanesi L., Di Palma F., Flicek P., Smith A. T., Thulin C. G., Alves P. C. and the Lagomorph Genomics Consortium. LaGomiCs—Lagomorph Genomics Consortium: An International Collaborative Effort for Sequencing the Genomes of an Entire Mammalian Order. *Journal of Heredity 2016*, esw010.

Abstract

The order Lagomorpha comprises about 90 living species, divided in 2 families: the pikas (Family Ochotonidae), and the rabbits, hares, and jackrabbits (Family Leporidae). Lagomorphs are important economically and scientifically as major human food resources, valued game species, pests of agricultural significance, model laboratory animals, and key elements in food webs. A quarter of the

lagomorph species are listed as threatened. They are native to all continents except Antarctica, and occur up to 5000 m above sea level, from the equator to the Arctic, spanning a wide range of environmental conditions. The order has notable taxonomic problems presenting significant difficulties for defining a species due to broad phenotypic variation, overlap of morphological characteristics, and relatively recent speciation events. At present, only the genomes of 2 species, the European rabbit (*Oryctolagus cuniculus*) and American pika (*Ochotona princeps*) have been sequenced and assembled. Starting from a paucity of genome information, the main scientific aim of the Lagomorph Genomics Consortium (LaGomiCs), born from a cooperative initiative of the European COST Action "A Collaborative European Network on Rabbit Genome Biology - RGB-Net" and the World Lagomorph Society (WLS), is to provide an international framework for the sequencing of the genome of all extant and selected extinct lagomorphs. Sequencing the genomes of an entire order will provide a large amount of information to address biological problems not only related to lagomorphs but also to all mammals. We present current and planned sequencing programs and outline the final objective of LaGomiCs possible through broad international collaboration.

Utzeri V.J., Bertolini F., Ribani A., Schiavo G., Dall'Olio S., Fontanesi L. The albinism of the feral Asinara white donkeys (*Equus asinus*) is associated with a missense mutation in a highly conserved position of the tyrosinase (*TYR*) gene. *Animal Genetics 2016*, **47**(*1*): 120-124.

Summary

A feral donkey population (*Equus asinus*), living in the Asinara island National Park (north-west to Sardinia, Italy), includes a unique white albino donkey sub-population or colour morph, known as Asino dell'Asinara (with about 100-120 animals), that is a major attraction of this natural park. Disrupting mutations in the tyrosinase (*TYR*) gene are known to cause recessive albinisms in humans (i.e. Oculocutaneous Albinism Type 1 or OCA1 defects) and several other species. In this study, we analysed the donkey *TYR* gene as a strong candidate to identify the causative mutation of the albinism of the Asinara white donkeys. All five exons and parts of the intronic and flanking regions of the *TYR* gene were sequenced from 13 donkeys (7 Asinara white albino and 6 coloured animals). Seven single nucleotide polymorphisms were identified and distributed in five haplotypes. A missense mutation (p.H202D) in a highly conserved amino acid position (even across kingdoms), that disrupts the first copper binding site (CuA) of the TYR catalytic domain (as also confirmed by 3D protein modelling),

was identified in homozygous condition (D/D) in all Asinara white albino donkeys and in the albino son of a trio (the grey parents had genotype H/D), confirming the recessive mode of inheritance of this mutation. Genotyping 82 donkeys confirmed that Asinara white albino donkeys had genotype D/D whereas all other coloured donkeys had genotype H/H or H/D. Across populations association between the p.H202D genotypes and albino coat colour was highly significant (P=6.17E-18), further supporting a causative role of this amino acid substitution. The identification of the causative mutation of the albinism in the Asinara white donkeys might open new perspectives to study the dynamics of this putative deleterious allele in a feral population and to manage this interesting animal genetic resource.

Bertolini F., Schiavo G., Scotti E., Ribani A., Martelli P.L., Casadio R., Fontanesi L. High throughput SNP discovery in the rabbit (*Oryctolagus cuniculus*) genome by next generation semiconductor based-sequencing. *Animal Genetics 2014,* **45** *(2):*304-307

Abstract

The European rabbit (*Oryctolagus cuniculus*) is a domesticated species with one of the broadest ranges of economic and scientific applications and fields of investigation. Rabbit genome information and assembly are available (oryCun2.0), but so far few studies have investigated its variability, and massive discovery of polymorphisms has not been published yet for this species. Here, we sequenced two reduced representation libraries (RRLs) to identify single nucleotide polymorphisms (SNPs) in the rabbit genome. Genomic DNA of 10 rabbits belonging to different breeds was pooled and digested with two restriction enzymes (HaeIII and RsaI) to create two RRLs which were sequenced using the Ion Torrent Personal Genome Machine. The two RRLs produced 2 917 879 and 4 046 871 reads, for a total of 280.51 Mb (248.49 Mb with quality >20) and 417.28 Mb (360.89 Mb with quality >20) respectively of sequenced DNA. About 90% and 91% respectively of the obtained reads were mapped on the rabbit genome, covering a total of 15.82% of the oryCun2.0 genome version. The mapping and ad hoc filtering procedures allowed to reliably call 62 491 SNPs. SNPs in a few genomic regions were validated by Sanger sequencing. The Variant Effect Predictor Web tool was used to map SNPs on the current version of the rabbit genome. The obtained results will be useful for many applied and basic research programs for this species and will contribute to the development of cost-effective solutions for high-throughput SNP genotyping in the rabbit.

Fontanesi L., Ribani A., Scotti E., Utzeri V.J., Velickovic N., Dall'Olio S. Differentiation of meat from European wild boars and domesticated pigs using polymorphisms in the MC1R and NR6A1 genes. *Meat Science 2014*, **98**: 781-784

Abstract

Wild boar meat cannot be easily distinguished from domestic pig meat, especially in processed products, thus it can be fraudulently substituted with cheaper domestic pork. In this study we genotyped polymorphisms in two genes (MC1R, affecting coat color and NR6A1, associated with number of vertebrae) in 293 domestic pigs of five commercial breeds, 111 wild boars sampled in Italy, and 90 in Slovenia and other Western Balkan regions. Allele and genotype frequency datawere used to set up a DNA-based method to distinguish meat of wild boars and domestic pigs. Genotyping results indicated that domesticated genes were introgressed intowild boar populations. This complicated the determination of the origin of the meat andwould cause a high error rate if markers of only one gene were used. The combined use of polymorphisms in the two analyzed genes substantially reduced false negative results

Utzeri V.J., **Ribani A., Fontanesi L**. A premature stop codon in exon 2 of the *TYRP1* gene is associated with brown coat colour in rabbits (*Oryctolagus cuniculus*). *Animal Genetics 2014*, **45**:600 – 603

Abstract

Classical genetic studies in European rabbits (*Oryctolagus cuniculus*) suggested the presence of two alleles at the brown coat colour locus: a wild-type B allele that gives dense black pigment throughout the coat and a recessive b allele that in the homozygous condition (b/b genotype) produces brown rabbits that are unable to develop black pigmentation. In several other species, this locus is determined by mutations in the tyrosinase-related protein 1 (*TYRP1*) gene, encoding a melanocyte enzyme needed

for the production of dark eumelanin. In this study, we investigated the rabbit *TYRP1* gene as a strong candidate for the rabbit brown coat colour locus. A total of 3846 bp of the *TYRP1* gene were sequenced in eight rabbits of different breeds and identified 23 single nucleotide polymorphisms (SNPs; 12 in intronic regions, five in exons and six in the 30-untranslated region) and an insertion/ deletion of 13 bp, in the 30-untranslated region, organised in a few haplotypes. A mutation in exon 2 (g.41360196G>A) leads to a premature stop codon at position 190 of the deduced amino acid sequence (p.Trp190ter). Therefore, translation predicts a truncated *TYRP1* protein lacking almost completely the tyrosinase domain. Genotyping 203 rabbits of 32 different breeds identified this mutation only in brown Havana rabbits. Its potential functional relevance in disrupting the *TYRP1* protein and its presence only in brown animals strongly argue for this non-sense mutation being a causative mutation for the recessive b allele at the brown locus in *Oryctolagus cuniculus*.

Oral Presentations in Scientific Congresses

Next generation semiconductor based sequencing of bitter taste receptor genes in different pig breeds and populations and association study of identified polymorphisms using a DNA pooling strategy, - XXI ASPA Congress (Animal Science and Production Association) – Milan (Italy) 9-12 June 2015

Analysis of variability of the TYR gene in wild and domestic rabbits - RGB-Net Seminars and Meetings – Zagreb (Croatia) 7-8 May 2014

Posters

Utzeri V.J., Ribani A., Schiavo G., Bertolini F., Geraci C., Bovo S., Fontanesi L. *Food metagenomics against frauds: applications of next generation semiconductor based sequencing on meat and dairy products and honey* – FoodInnova 2017 – 31 January – 3 February 2017 – Cesena (Italy)

Ribani A., Utzeri V.J., Geraci C., Dall'Olio S., Nanni Costa L., Fontanesi L. *Local pig breeds are "less domesticated" than commercial populations: evidences from variability in the MC1R and NR6A1 genes in Italian autochthonous breeds* – 9th International Symposium on Mediterranean Pig – Portalegre (Portugal) – 3-5 November 2016

Ribani A., Bertolini F., Schiavo G., Scotti E., Utzeri V.J., Dall'Olio S., Trevisi P., Bosi P., Fontanesi L. *Next generation semiconductor based sequencing of bitter taste receptor genes in different pig populations and association study using a selective DNA pool-seq approach* – 67th EAAP Annual Meeting (European Federation of Animal Science) – Belfast (UK) – 29 August – 2 September 2016

Ribani A., Bertolini F., Schiavo G., Scotti E., Utzeri V.J., Dall'Olio S., Trevisi P., Bosi P., Fontanesi L. *A next generation semiconductor based target re-sequencing DNA pool-seq approach for the identification of SNPs and association studies: application to bitter taste receptor genes in different pig populations* - 35th ISAG (International Society for Animal Genetics) Conference – Salt Lake City (USA) 23-27 July 2016

Schiavo G., Hoffmann O.I., Ribani A., Utzeri V.J., Ghionda M.C., Bovo S., Fontanesi L. *A genomic landscape of mitochondrial DNA insertions in the nuclear pig genome* - 35th ISAG (International Society for Animal Genetics) Conference – Salt Lake City (USA) 23-27 July 2016

Fontanesi L., Bovo S., Schiavo G., Mazzoni G., Ribani A., Utzeri V.J., Dall'Olio S., Bertolini F., Fanelli F., Mazzullo M., Galimberti G., Calò D.G., Trevisi P., Pagotto U., Bosi P. *Deconstructing the pig genome-metabolome functional interactions* – 35th ISAG (International Society for Animal Genetics) Conference – Salt Lake City (USA) 23-27 July 2016

Ribani A., Utzeri V.J., Schiavo G., Bovo S., Geraci C., Fontanesi L. *Food genomics: application of innovative DNA analysis technologies for authentication of food products.* - TRADEIT Entrepreneurship Summer Academy – Postdam (Germany) 6-10 June 2016

Utzeri V.J., Ribani A., Dall'Olio S., Scotti E., Veličković N., Fontanesi L. *Evidence of introgression of domesticated genes in several European wild boar populations* XXI ASPA Congress (Animal Science and Production Association) – Milan (Italy) 9-12 June 2015

Utzeri V.J., Bertolini F., Ribani A., Dall'Olio S., Fontanesi L. *Identification of the causative mutation of the albinism determining the white coat colour of the Asinara donkey breed* - XXI ASPA Congress (Animal Science and Production Association) – Milan (Italy) 9-12 June 2015

Trevisi P., Ribani A., Colombo M., Utzeri V.J., Bosi P., Fontanesi L. *A first nutrigenomic trial in pigs identifies a DNA polymorphism affecting kynurenine metabolites after tryptophan addition and E. coli challenge*, 13th Digestive Physiology of Pigs – Kliczkow (Poland) 19-21 May 2015

Ribani A., Utzeri V.J., Scotti E., Bertolini F., Dall'Olio S., Fontanesi L.: "*Sequence analysis of the tyrosinase gene (*TYR, Albino *locus) in wild and domesticated rabbits (*Oryctolagus cuniculus*) and in other wild Lagomorph species*", XX ASPA Congress (Animal Science and Production Association), Bologna, 11-13 June 2013

Bertolini F., Schiavo G., Scotti E., Ribani A., Martelli P.L., Casadio R., Fontanesi L.: "*Application of the Ion Torrent technology to identify single nucleotide polymorphisms in the rabbit genome*", XX ASPA Congress (Animal Science and Production Association), Bologna, 11-13 June 2013.