Conservation genetics of local and wild pig populations: insight in genetic diversity and demographic history

Juan Manuel Herrero Medrano

Thesis committee

Thesis supervisor
Prof. Dr. M.A.M. Groenen
Personal chair at the Animal Breeding and Genomics Centre
Wageningen University

Thesis co-supervisors
Dr. R.P.M.A. Crooijmans
Assistant professor at the Animal Breeding and Genomics Centre
Wageningen University

Dr. H.-J. Megens
Research Associate at the Animal Breeding and Genomics Centre
Wageningen University

Other members
Prof. Dr. B. Zwaan, Wageningen University.
Dr. G. Larson, Durham University, United Kingdom.
Dr. P. Wiener, University of Edinburgh, United Kingdom.
Prof. Dr. Ir. F. Govers, Wageningen University.

This research was conducted under the auspices of the Graduate School of Wageningen Institute of Animal Sciences (WIAS).

Conservation genetics of local and wild pig populations: insight in genetic diversity and demographic history

Juan Manuel Herrero Medrano

Thesis

submitted in fulfillment of the requirements for the degree of doctor
at Wageningen University
by the authority of the Rector Magnificus
Prof. dr. M.J. Kropff,
in the presence of the
Thesis Committee appointed by the Academic Board
to be defended in public
on Monday 4 November 2013
at 4 p.m. in the Aula.

Herrero-Medrano, JM.
Conservation genetics of local and wild pig populations: insight in genetic diversity and demographic history.
160 pages

PhD thesis, Wageningen University, the Netherlands (2013)
With references, with summaries in Dutch and English

ISBN 978-94-6173-751-9

Abstract

Herrero-Medrano, JM. (2013). Conservation genetics of local and wild pig populations: insight in genomic diversity and demographic history. PhD thesis, Wageningen University, the Netherlands

A limited number of highly productive populations has progressively led to many local breeds becoming endangered or extinct. Genetic characterization of the genetic resources is, thereby, needed to prevent further loss of genetic and cultural heritage. The development of new genotyping technologies provides unprecedented opportunities for the implementation of effective conservation programs. The aim of the study described in this thesis was to explore the genetic diversity and demographic history of local pig populations, and the applicability of the results for the long-term future conservation of livestock genetic resources. In this thesis, mitochondrial DNA (mtDNA) and nuclear genetic marker systems were used to explore the past demographic history and genetic diversity of domestic and wild pigs. A Bayesian phylogeographic analysis using mtDNA allowed the detection of past dispersal events of *Sus scrofa* across Eurasia. Dispersal patterns consistent with fossil records described in other species as well as hitherto untested dispersal routes were detected. Insight in the demographic history of local pigs was obtained by using 60K SNP data. The study of regions of homozygosity (ROH) and past effective population size (N_e) showed genetic signs of past bottlenecks in some populations. The estimation of N_e revealed the bottleneck suffered by wild pigs during the last glacial maximum, and an increase of N_e of Iberian pigs may due to the domestication. The SNP panel proved to be highly efficient for population structure analyses, as it was able to differentiate 13 European local breeds and correctly assigning the pigs to their population of origin. Within the population structure analysis, identification of admixture is a relevant issue in conservation management of livestock species. A population structure analysis combined with an analysis of ROH and the calculated inbreeding factor at the individual level provided suitable parameters to identify pigs that have been recently crossed with other breeds. I observed a high correlation between genetic diversity computed with the 60K SNP and whole genome re-sequence data. This high correlation inferred indicates that the Porcine 60K SNP Beadchip provides reliable estimates of genomic diversity in European pig populations. The study of NGS data of local and commercial European breeds demonstrated that, despite the higher inbreeding observed in many local pigs, local pigs harbour different genomic variants that may represent a valuable genetic reservoir for the livestock breeding industry in the future. This thesis provides a benchmark to address rational management and

exploitation of local genetic resources. Moreover, the large representation of pig populations, the choice of genetic marker systems and the approaches utilized may benefit future studies that aim to genetically characterize livestock populations.

Contents

1

General introduction

1.1 Introduction

There are at least six species of the Genus *Sus*, of which *Sus scrofa* shows the largest geographic distribution [1]. It is estimated that around 3.0-3.5 million years ago *Sus scrofa* emerged from South East Asia and colonized Asia, Europe and North Africa [2, 3]. Similar to other mammals, the glaciation events that occurred during the middle to end of the Pleistocene had a great impact on the spatial distribution of *Sus scrofa* with some populations becoming extinct and others being marginalized in warmer areas in so-called refugia [4]. After the Last Glacial Maximum, mammalian populations re-colonized regions previously covered by the permafrost, while others became isolated on islands (Figure 1.1).

Two main factors shaped the current genetic variability of pig populations and other mammals. The first comprises of large climate changes in the Pleistocene [5] while the second is human intervention during the Holocene (i.e. roughly the last 10,000 years). Interactions between humans and *Sus scrofa* have contributed markedly to the current variation in the latter species due to, most notably, domestication [3, 6, 7]. The most ancient archaeological evidence of pig domestication was found in Anatolia, suggesting that *Sus scrofa* was first domesticated in the Near East ~10,000 years ago [8–10]. Archaeological findings such as the sharp decreased molar tooth size observed by Ervynck et al. [11] in pig remains, strongly supports that this geographic region represented an ancient domestication centre for the species *Sus scrofa*, as it proved to be for other species too, including for instance cattle and wheat. The specific geographic and temporal context in which domestication of the pig took place across Eurasia is still debated. The development of new molecular techniques applied to modern and ancient archaeological data have supported the theory of multiple independent domestication events throughout Eurasia [3, 6, 12]. In Asia, four regions of possible pig domestication centre were suggested: 1) China, 2) South Asia, 3) the island of Lanyu and 4) Sulawesi [3, 12]. Regarding pig domestication in Europe, dispersal from the Near Eastern pigs across Europe by Neolithic farmers has been suggested based on the presence of Near East mtDNA haplotypes in pig remains excavated from early farming sites [6]. However, the total absence of a genetic signal of Near Eastern haplotypes in modern European breeds [3, 13, 14], together with the archaeological remains of bones of small domesticated pigs found in Switzerland and Italy among others, seemed to indicate independent domestication events from European wild boars [3]. Recently, the role of Europe and other regions as domestication centres has been challenged. Larson and Burger [15] pointed out the high similarity between mtDNA from wild and domestic pigs, but also the lack of

strong archaeological evidence of long-term domestication in Europe and in South East Asia. These authors suggested that this discrepancy between archaeological and genetic data can be explained by repeated events of admixture between domestic pigs imported from Near Eastern regions and European wild pigs in Europe [16], and similarly, between domestic pigs imported from China and wild boars from South East Asia [15]. Eurasian domestic pigs were subsequently transported to Oceania, Africa and America [17–19] explaining the current worldwide distribution of *Sus scrofa*.

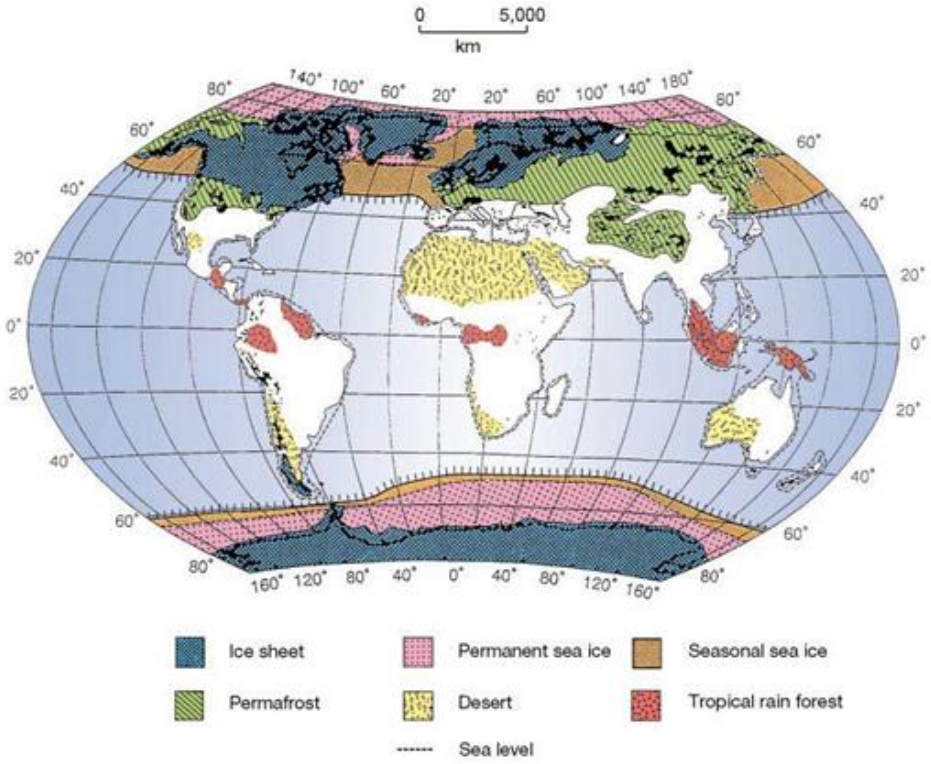

Figure 1.1 Reproduced from Hewitt, 2000 [4]. The maximum extent of ice and permafrost at the end of the last ice age 20,000 yr BP. The lowered sea level, large deserts and main blocks of tropical forest are indicated.

Pigs were independently domesticated in Asia and Europe, and subsequently selected for traits valuable to humans for thousands of years [20]. Asian domestic pigs were reared in family farms and fed by the farmers at very initial stages of domestication. However, the transition from wild to farmed was much slower in Europe, where domestic pigs were commonly released in the forest where they fed on acorns or chestnuts (Figure 1.2). Therefore, while Asian domestic pigs started to show the classic phenotype of a domestic pig, including shorter legs and a pot-belly, European pigs retained many similarities to their wild ancestor [20]. In addition, it is likely that human migrants transported domesticated pigs to different geographic locations. Wherever these pigs escaped, the result could be feral or hybrid pigs, potentially distorting the local wild boar population genetic structure [21]. A remarkable example of the human influence on the current genetic structure of *Sus scrofa* was the introgression of Asian domestic pigs into European pigs during the late 18th and 19th centuries [22, 23]. Chinese pigs were imported to England and used to improve local pigs for important production traits such as litter size and rapid weight gain. The ensuing English breeds, due to their improved production characteristics, subsequently became the founders of several of the currently recognized international, commercial pig breeds. This historical period constitutes the origin of the transformation from traditional systems into highly productive systems in Europe. The last 60-70 years have seen a revolution in livestock genetic improvement in general, and pigs in particular, due to strong selection schemes [24]. The modern animal production industry is characterized by the use of animals that are optimized for feed conversion, rapid growth and prolificacy in high-input production systems. While intensive systems are common in the developed world, the so-called low-input systems represent a fundamental source of meat in the developing world [25]. These systems are often based on the exploitation of local breeds reared under extensive or semi-extensive conditions.

Like the domesticated pig, the demography of wild populations has been strongly influenced by human activities, even a long time before domestication [26]. Hunting, habitat degradation, and re-stocking have resulted in continuous change of wild populations for centuries [7, 27]. Wild boar populations have been subjected to strong bottlenecks that sharply reduced their population size during the last centuries [28]. On the other hand, in the last 50 years, an expansion of the population size occurred, particularly in Europe, due to rapid growth in population densities due to improved nature conservation measures, even leading to active management (i.e. hunting) to prevent overpopulation in many regions [29, 30].

Nowadays, from a practical point of view, three groups of pig populations can be distinguished: (i) Commercial or international breeds, representing the pig

populations spread across different geographic regions; (ii) local breeds, restricted to one country or geographic region; (iii) wild, including feral, pigs that have not been domesticated [31].

Figure 1.2 Men knocking down acorns to feed swine, from the 14th century English Queen Mary Psalter, MS. Royal 2 B VII f.81v.

1.2 Genetic marker systems used in livestock populations

The development of genetic marker systems has had an overwhelming impact in population genetic research [32, 33]. Microsatellites and mitochondrial DNA (mtDNA) revolutionized our understanding of evolutionary process and population history in humans and animals. Various limitations, e.g. difficulty of automation of genotyping assays, highlighted the need for marker systems that would better scale to genotyping at high density [32]. High-density SNP panels are currently widely used in a large variety and number of studies, thereby greatly improving our knowledge of genome variation both in wild and domesticated species. Technological advances have culminated in complete sequencing of livestock species such as chicken [34], pig [2], and cattle [35], opening up new possibilities in population genetics studies [15, 36].

Direct sequencing of the hypervariable region, or control region, of the **mtDNA** can be rapidly obtained at relatively low cost. The study of mtDNA has proven to be informative for phylogenetic studies, and has unravelled key elements of the domestication process in many species, such as the fact that many species, including pig, were domesticated independently in different regions (e.g. see references [3, 23, 37, 38]). MtDNA evolves rapidly –faster than nuclear DNA–, is maternally inherited and does not recombine [39–41] which largely explains its usefulness for phylogenetic studies. However, the fact that mtDNA represents a single locus, together with being exclusively inherited maternally, limits detailed studies to unravel some historical processes acting on populations [32, 42]. Thus, complex demographic events such as admixture [15] and sex-specific breeding practices, which have a very important role in livestock production, cannot be accurately estimated with mtDNA. These limitations have justified the need of incorporating nuclear data in population genetic studies.

Microsatellites have been the marker system of choice in population genetic studies [32] for the better part of the last twenty-five years. The popularity of microsatellites stems from the possibility to genotype individuals with a very polymorphic and co-dominant genetic marker [43], at reasonable cost. Microsatellites have been harnessed by researchers on population genetic studies to test parentage or relatedness (e.g. [44]), genetic diversity (e.g. [45]), and population structure (e.g. [46]), due to the high polymorphism and distribution across the genome. The high level of heterozygosity and the genome-wide distribution made this type of marker suitable for developing genetic maps in humans [47] and livestock species such as cow [48], pig [49], chicken [50] and sheep [51] during the 1990s. However, several disadvantages have been associated with microsatellites. Complex patterns of mutation within and between loci add uncertainty to the use of microsatellites in phylogenetic studies [32]. Recurrent mutation may lead to homoplastic alleles that are identical by state but not by descent [52, 53]. From a practical point of view, genotyping a large number of microsatellites is labour-intensive and it is also difficult to compare studies across different laboratories unless reference control samples are genotyped to calibrate allele sizes. Finally, the density of microsatellite panels, usually not exceeding several hundred markers, is too low for fine mapping of QTLs [54].

Single nucleotide polymorphisms (**SNPs**) are (usually) bi-allelic markers and therefore, less polymorphic than microsatellites. However, the low polymorphism may be compensated by genotyping large SNP panels [46]. Modern high-throughput SNP panels allows for single-reaction assays of tens of thousands to millions of SNPs. The high density of such marker assays can provide insight in the

genome not achievable with microsatellites. For instance, the use of a high-density SNP panel demonstrated that linkage disequilibrium (LD) extends to larger distances than expected in humans [55]. The high density of SNPs, the existence of automated and standardized panels, and its cost-effectiveness largely explain that dense SNP panels have revolutionized genome research in recent years. In fact, commercially available high-density SNP panels, now available for almost all important livestock species [56–58] at relatively low cost, have led to the adaptation of highly standardized marker assays by livestock industries [59].

The high number of neutral genetic markers included in the SNP panels has increased the accuracy to assess classic parameters within population genetics, including population structure and relationship among populations [60–62], although there is not a full agreement regarding which genetic marker –SNPs or microsatellites– better reflect genome-wide genetic diversity [45]. High-density SNP panels have proven their usefulness to study patterns of LD, for QTL mapping, and for genome-wide association studies [63–65]. An issue of some concern in the use of chip-based high-throughput genotyping is the presence of so-called ascertainment bias. Commercially available SNP chips are usually designed on the basis of genetic variation of a small number of individuals from selected populations, such as commercial breeds. Therefore, the utilization of this set of markers in a wider set of populations may distort the results, especially when used in populations genetically differentiated from those that were used to design the chip [66].

A recent development to analyse entire genome-wide variation in livestock is the utilization of Next Generation Sequencing (**NGS**) technology. For instance, the study of the pig genome provided new insights of the demographic history of *Sus scrofa* [2], and also provided the opportunity to perform unbiased and accurate studies to estimate genomic diversity [67], regions of homozygosity [65] and to detect signatures of selection [68]. One of the most promising applications of NGS data is the study of functional consequences of mutations, and thereby to investigate the molecular basis of phenotypic traits of importance. Owing to its efficiency and wide applicability, NGS, is expected to become the most important tool in the coming years to study genomic variability.

The use of NGS data involves new approaches, but also leads to new challenges that need to be met. Recently, genome-wide association analyses (GWAS) in humans have been replaced by studies based on NGS data [69]. While GWAS aims to identify genomic regions or mutations underlying production traits and disease phenotypes, whole-genome sequence analyses make it possible to point out target genes through a functional genomics approach. The great accumulation of data

generated with new sequencing technologies implies new approaches in functional genomics [70]. Thus, there is a shift from forward genetics to reverse genetics. Classically, in forward genetics, the starting point to study functional genetics is a phenotypic variation for which the researcher tries to determine the genetic basis by eventually sequencing genomic regions that may be involved in the observed phenotypic trait. The so-called reverse genetics approach operates in the opposite direction. Since NGS allows the researchers access to large volumes of genomic data, generally faster than to phenotypic data, it promotes a scenario where the gene sequence is known but its biological effect is unknown. Another consequence of the transition from marker systems such as microsatellites and SNP chips to whole-genome sequences is that, instead of a moderate number of individuals genotyped for a small or medium number of markers, a few or even a single individual is genotyped for (almost) all the variation in the genome [71]. This approach has some challenges. First, the use of a few individuals may not represent the idiosyncrasy of the population, leading to insufficient power to extrapolate biological questions to the population level. Second, random shot gun sequencing implies variable sequence coverage, with nucleotide sites genotyped with low coverage that may represent only one of the two parental chromosomes [71]. Moreover, sequencing errors at regions with high coverage can be misinterpreted as polymorphic sites [71].

1.3 Conservation genetics in livestock species

Introduction

Human activities are directly and indirectly decreasing the global biodiversity as reflected by the large number of species that have become extinct, while many others have declined to dangerously low population sizes that could lead to extinction [72]. An active response is needed to improve the management of endangered species in order to ensure their long-term survival. To achieve this goal, genetic factors must be considered [73]. Conservation genetics is the application of evolutionary and molecular genetics to preserve biodiversity. Preservation of genetic variation and population distinctiveness are major concerns in conservation genetics of livestock species since loss of genetic diversity increases the risk of extinction [73], while high genetic variation is related to higher adaptive potential to new environmental conditions such as changing climate conditions or emergence of new pathogens [73, 74].

Another concern in conservation genetics is inbreeding depression, which refers to the accumulation of homozygous alleles with a negative effect on fitness. While damaging mutations are efficiently removed, or kept at low frequencies, from large

populations by natural selection, small populations have a higher probability to accumulate high frequencies of damaging mutations through random genetic drift, thereby reducing fitness and productivity [75, 76]. As in any managed population, in domestic breeds, the need to preserve genetic variation to avoid the risk of inbreeding depression must be considered. Avoiding health problems is a direct concern for animal breeders. Furthermore, the genetic variation is the raw material used by the breeders to accelerate genetic gain. The large amount of data provided by new genotyping technologies will aid to improve traditional approaches of maintaining genetic variation in populations, and, at the same time, will make it possible to address new approaches in conservation [36, 77, 78].

Conservation of local breeds

Management practices, developed mainly in Europe during the last two centuries, have tended to increase productivity and economic profitability to the detriment of the conservation of livestock genetic resources [79]. Specifically, advances in the livestock industry during the last decades have led to the existence of a limited number of highly productive breeds that have progressively marginalized, and even replaced, many local breeds. The FAO has cautioned that nearly 20% of the domestic animal breeds are threatened with extinction, while 30% lack data regarding the status [80]. Local breeds are particularly threatened due to their typically small population size and low productivity [25]. Low production of local breeds may compel farmers to choose different breeds on economic grounds, thereby increasing the risk of local breeds becoming extinct. Small-scale farmers play a key role in the management of livestock genetic recourses and therefore biodiversity, as recently recognized by FAO [81]. There are two major concerns in the conservation genetics of livestock domestic populations: inbreeding and crossbreeding. Inbreeding may erode genetic diversity and threaten long-term survival because of inbreeding depression. Crossbreeding, may pose a threat to the historical significance of the population by losing what is perceived to be its 'genetic integrity' [82], as well as losing alleles particular to the population. Thus, rational use and protection of local breeds from genetic introgression from highly productive breeds are major goals in conservation genetics of livestock populations.

Local breeds do provide a large contribution to the genetic diversity of the domestic stock [79, 83, 84]. The genetic variability of local breeds has been related to high adaptation to harsh environments and meat quality characteristics. While highly-productive breeds need proper nutritional and environmental conditions to be fully productive, local breeds may be productive under harsh local conditions. A

clear example of adaptation of local breeds in a local environment is observed in populations reared in tropical regions. Indigenous breeds from those geographic areas have to cope with heat stress, poor nutrition and disease pressure [25, 85]. Hansen [86] reviewed the differences between Zebu breeds, indigenous from tropical areas, and European breeds (*Bos taurus*). The Zebu breeds show physiological differences that confer superior ability to regulate body temperature during heat stress. Finally, it cannot be dismissed that local breeds are often regarded as part of the cultural heritage of local and national communities, having an important socio-economic value in their geographic regions.

Once the need to preserve local breeds has been accepted, the genetic characterization of livestock resources arises as a major step in the design of proper conservation and management programs. The maintenance of genetic resources will increase the capability to respond to present and future needs of livestock production.

New approaches in conservation genetics: conservation genomics

The term conservation genomics has recently been introduced and refers to the use of genome-wide data to solve problems in conservation biology [36]. Traditionally, genetic studies to preserve biodiversity focused on the analysis of neutral variation by analysing neutral markers such as microsatellites and SNP panels. Neutral variation is shaped by evolutionary forces, i.e. genetic drift, mutation, migration and recombination, and it can be used to assess demography, population distinctiveness, gene flow and conservation status [77, 87]. Low levels of genetic diversity, as inferred from neutral loci, has been indirectly related to reduced individual fitness due to inbreeding depression [88, 89]. However, correlation between neutral variation and fitness of the population may be weak [87, 90].

New genomic tools will change conservation genetics approaches [36, 77]. Various applications will focus on estimating parameters that require the use of neutral markers at a higher accuracy, such as levels of inbreeding, and population differentiation [36]. Perhaps the most exiting application of whole-genome sequencing is the study of functional genetic variation. For instance, NGS enables direct assessment of the consequences of loss of genetic variation on the adaptability of populations [36, 77] (Figure 1.3). It furthermore enables identification of the effect and distribution of those loci involved in fitness [91]. The knowledge provided by NGS data combined with demographic history, may allow the identification of local populations with higher adaptive potential, as well as

detection of endangered populations with high levels of detrimental variation. Both situations may require specific actions from a conservation point of view [77, 92].
The application of genomics tools in conservation biology will require overcoming several challenges and limitations. Understanding the relation between genomic variation within genes and its consequence in adaptation is still very much under development [78]. From a practical point of view, NGS data provides large amounts of data that may be challenging in terms of analysis and storage [36]. Moreover, the cost of re-sequencing representative samples of a population is still high.

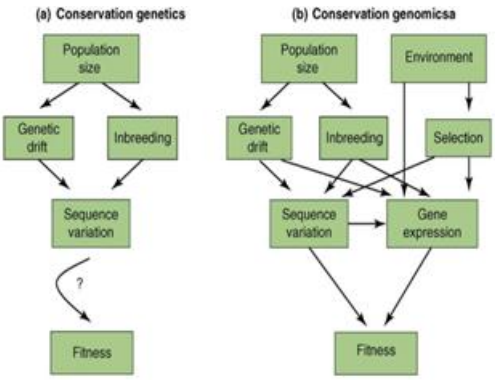

Figure 1.3 Reproduced from Ouborg et al. [78]. Schematic representation of conservation genetics (a) and conservation genomics (b) approaches.

1.4 Aim and outline of the thesis

The first aim of this thesis is to provide further understanding of the past events that shaped the current genetic structure of *Sus scrofa*. To achieve this goal, historical migration patterns of wild boar populations across Eurasia are estimated as well as the genetic diversity and differentiation based on mtDNA data (Chapter 2). Chapter 3 comprehensively characterizes a Spanish local breed (the Chato Murciano pig) by investigating the majority of all remaining breeding stock, and by integrating microsatellites, mtDNA and high-density SNP panel. This breed exemplifies the risks to which many other livestock populations are exposed as a result of inbreeding and crossbreeding with highly productive breeds. A broader study of local pig populations, including domestic and wild populations from the Iberian Peninsula is performed in the Chapter 4. Population structure, inbreeding and demographic history in each population was assessed using high-density SNP data to provide valuable insights in the conservation genetics of pigs from the Iberian Peninsula. Integration of high-density SNP and NGS data to assess genomic diversity in local pig breeds with a wide European distribution is presented in Chapter 5. Moreover, the analysis of NGS data from local and commercial breeds is compared to identify candidate mutations underlying phenotypic differences between highly productive breeds and low-input-breeds. Finally, in the general discussion the relevant findings of this thesis and the practical implication of the results in conservation genetics of livestock species are discussed.

References

1. Ruvinsky A, Rothschild MF, Larson G, Gongora J: **Systematics and Evolution of the Pig**. In *The genetics of the pig, 2nd edition*. edited by Rothschild MF, Ruvinsky A CABI, Wallingford, UK: 2011:1–13.
2. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-J, Li S, Larkin DM, Kim H, Frantz LAF, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S, Bujie Z, et al.: **Analyses of pig genomes provide insight into porcine demography and evolution.***Nature* 2012, **491**:393–8.
3. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A: **Worldwide phylogeography of wild boar reveals multiple centers of pig domestication.***Science (New York, N.Y.)* 2005, **307**:1618–21.
4. Hewitt G: **The genetic legacy of the Quaternary ice ages.***Nature* 2000, **405**:907–13.
5. Hofreiter M, Serre D, Rohland N, Rabeder G, Nagel D, Conard N, Münzel S, Pääbo S: **Lack of phylogeography in European mammals before the last glaciation.***Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**:12963–8.
6. Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, Vigne J-D, Edwards CJ, Schlumbaum A, Dinu A, Balaçsescu A, Dolman G, Tagliacozzo A, Manaseryan N, Miracle P, Van Wijngaarden-Bakker L, Masseti M, Bradley DG, Cooper A: **Ancient DNA, pig domestication, and the spread of the Neolithic into Europe.***Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**:15276–81.
7. Scandura M, Iacolina L, Crestanello B, Pecchioli E, Di Benedetto MF, Russo V, Davoli R, Apollonio M, Bertorelle G: **Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable?***Molecular ecology* 2008, **17**:1745–62.
8. Redding R, Rosenberg M: *Ancestors for the Pigs: Pigs in Prehistory.* Philadelphia: Univ of Pennsylvania MASCA Researcher Papers in Science and Archaeology; 1998, **15**:65–76.
9. Zeder MA: **Domestication and early agriculture in the Mediterranean Basin: Origins, diffusion, and impact.***Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:11597–604.

10. Conolly J, Colledge S, Dobney K, Vigne J-D, Peters J, Stopp B, Manning K, Shennan S: **Meta-analysis of zooarchaeological data from SW Asia and SE Europe provides insight into the origins and spread of animal husbandry**. *Journal of Archaeological Science* 2011, **38**:538–545.
11. Meadow R, Hongo H, Dobney K, Ervynck A: **Born Free ? New Evidence for the Status of *Sus scrofa* at Neolithic Çayönü Tepesi (Southeastern Anatolia, Turkey)**. *Paléorient* 2001, **27**:47–73.
12. Larson G, Liu R, Zhao X, Yuan J, Fuller D, Barton L, Dobney K, Fan Q, Gu Z, Liu X-H, Luo Y, Lv P, Andersson L, Li N: **Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA.***Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:7686–91.
13. Davies TJ, Buckley LB: **Phylogenetic diversity as a window into the evolutionary and biogeographic histories of present-day richness gradients for mammals.***Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2011, **366**:2414–25.
14. Manunza A, Zidi A, Yeghoyan S, Balteanu VA, Carsai TC, Scherbakov O, Ramírez O, Eghbalsaied S, Castelló A, Mercadé A, Amills M: **A High Throughput Genotyping Approach Reveals Distinctive Autosomal Genetic Signatures for European and Near Eastern Wild Boar**. *PLoS ONE* 2013, **8**:e55891.
15. Larson G, Burger J: **A population genetics view of animal domestication.***Trends in genetics : TIG* 2013, **29**:197–205.
16. Ottoni C, Flink LG, Evin A, Geörg C, De Cupere B, Van Neer W, Bartosiewicz L, Linderholm A, Barnett R, Peters J, Decorte R, Waelkens M, Vanderheyden N, Ricaut F-X, Cakirlar C, Cevik O, Hoelzel a R, Mashkour M, Karimlu AFM, Seno SS, Daujat J, Brock F, Pinhasi R, Hongo H, Perez-Enciso M, Rasmussen M, Frantz L, Megens H-J, Crooijmans R, Groenen M, et al.: **Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics.***Molecular biology and evolution* 2013, **30**:824–32.
17. Larson G, Cucchi T, Fujita M, Matisoo-Smith E, Robins J, Anderson A, Rolett B, Spriggs M, Dolman G, Kim T-H, Thuy NTD, Randi E, Doherty M, Due RA, Bollt R, Djubiantono T, Griffin B, Intoh M, Keane E, Kirch P, Li K-T, Morwood M, Pedriña LM, Piper PJ, Rabett RJ, Shooter P, Van den Bergh G, West E, Wickler S, Yuan J, et al.: **Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in Island Southeast Asia and Oceania.***Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**:4834–9.
18. Ramírez O, Ojeda a, Tomàs a, Gallardo D, Huang LS, Folch JM, Clop a, Sánchez a, Badaoui B, Hanotte O, Galman-Omitogun O, Makuza SM, Soto H, Cadillo J, Kelly L,

Cho IC, Yeghoyan S, Pérez-Enciso M, Amills M: **Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds.***Molecular biology and evolution* 2009, **26**:2061–72.

19. Burgos-Paz W, Souza CA, Megens HJ, Ramayo-Caldas Y, Melo M, Lemús-Flores C, Caal E, Soto HW, Martínez R, Alvarez LA, Aguirre L, Iñiguez V, Revidatti MA, Martínez-López OR, Llambi S, Esteve-Codina A, Rodríguez MC, Crooijmans RPMA, Paiva SR, Schook LB, Groenen MAM, Pérez-Enciso M: **Porcine colonization of the Americas: a 60k SNP story.***Heredity* 2013, **110**:321–30.
20. White S: **From Globalized Pig Breeds to Capitalist Pigs: A Study in Animal Cultures and Evolutionary History**. *Environmental History* 2011, **16**:94–120.
21. Goedbloed DJ, Megens HJ, Van Hooft P, Herrero-Medrano JM, Lutz W, Alexandri P, Crooijmans RPM a, Groenen M, Van Wieren SE, Ydenberg RC, Prins HHT: **Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations.***Molecular ecology* 2012, **22**:856–866.
22. Darwin C: *The Variation of Animals and Plants under Domestication*. John Murry. London: 1868.
23. Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L: **The origin of the domestic pig: independent domestication and subsequent introgression.***Genetics* 2000, **154**:1785–91.
24. Porter V: *Pigs. A Handbook to the Breeds of the World*. Mountfield, East Sussex: Helm Information, Ltd.; 1993.
25. FAO: **Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration.***(available at http://www.fao.org/docrep/010/a1404e/a1404e00.htm).* 2007.
26. Vigne J-D, Zazzo A, Saliège J-F, Poplin F, Guilaine J, Simmons A: **Pre-Neolithic wild boar management and introduction to Cyprus more than 11,400 years ago.***Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**:16135–8.
27. Scandura M, Iacolina L, Cossu A, Apollonio M: **Effects of human perturbation on the genetic make-up of an island population: the case of the Sardinian wild boar.***Heredity* 2011, **106**:1012–20.
28. Apollonio M, Randi E, Toso S: **The systematics of the wildboar (Sus scrofa L.) in Italy.***Bollettino Di Zoologia* 1988, **3**:213–221.
29. Lang S, Pesson B, Klein F, Schreiber A: **Wildlife genetics and disease: allozyme evolution in the wild boar (*Sus scrofa*) caused by a swine fever epidemy.***Genetics, selection, evolution: GSE* 2000, **32**:303–10.

30. Beltrán-Beck B, Ballesteros C, Vicente J, De la Fuente J, Gortázar C: **Progress in Oral Vaccination against Tuberculosis in Its Main Wildlife Reservoir in Iberia, the Eurasian Wild Boar.***Veterinary medicine international* 2012, **2012**:978501.
31. Lauvergne JJ: **Genetics in animal populations after domestication: the consequences for breeds conservation.** In *Proceedings of the Second World Congress on Genetic Applied to Livestock Production*. Madrid: 1982:77–85.
32. Zhang D-X, Hewitt GM: **Nuclear DNA analyses in genetic studies of populations: practice, problems and prospects.***Molecular ecology* 2003, **12**:563–84.
33. Fu YX, Li WH: **Coalescing into the 21st century: An overview and prospects of coalescent theory.***Theoretical population biology* 1999, **56**:1–10.
34. International Chicken Genome Sequencing Consortium.: **Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution.***Nature* 2004, **432**:695–716.
35. Zimin A V, Delcher AL, Florea L, Kelley DR, Schatz MC, Puiu D, Hanrahan F, Pertea G, Van Tassell CP, Sonstegard TS, Marçais G, Roberts M, Subramanian P, Yorke JA, Salzberg SL: **A whole-genome assembly of the domestic cow, Bos taurus.***Genome biology* 2009, **10**:R42.
36. Allendorf FW, Hohenlohe P a, Luikart G: **Genomics and the future of conservation genetics.***Nature reviews. Genetics* 2010, **11**:697–709.
37. Tishkoff SA, Gonder MK, Henn BM, Mortensen H, Knight A, Gignoux C, Fernandopulle N, Lema G, Nyambo TB, Ramakrishnan U, Reed FA, Mountain JL: **History of click-speaking populations of Africa inferred from mtDNA and Y chromosome genetic variation.***Molecular biology and evolution* 2007, **24**:2180–95.
38. Liu Y-P, Wu G-S, Yao Y-G, Miao Y-W, Luikart G, Baig M, Beja-Pereira A, Ding Z-L, Palanichamy MG, Zhang Y-P: **Multiple maternal origins of chickens: out of the Asian jungles.***Molecular phylogenetics and evolution* 2006, **38**:12–9.
39. Wolstenholme DR: **Animal mitochondrial DNA: structure and evolution.***International review of cytology* 1992, **141**:173–216.
40. Avise JC: *Molecular markers, natural history and evolution*. 1st edition. London: 1994.
41. Brown WM: **Rapid Evolution of Animal Mitochondrial DNA**. *Proceedings of the National Academy of Sciences* 1979, **76**:1967–1971.
42. Hoelzer G: **Inferring phylogenies from mtDNA variation: mitochondrial-gene trees versus nuclear-gene trees revisited.***Evolution* 1997, **51**:622–626.
43. Weber JL, Wong C: **Mutation of human short tandem repeats.***Human molecular genetics* 1993, **2**:1123–8.

44. Usha AP, Simpson SP, Williams JL: **Probability of random sire exclusion using microsatellite markers for parentage verification.***Animal genetics* 1995, **26**:155–61.

45. Ljungqvist M, Mikael A: **Do microsatellites reflect genome-wide genetic diversity in natural populations ?** 2010:851–855.

46. Haasl RJ, Payseur B a: **Multi-locus inference of population structure: a comparison between single nucleotide polymorphisms and microsatellites.***Heredity* 2011, **106**:158–71.

47. Weissenbach J: **A second generation linkage map of the human genome based on highly informative microsatellite loci.***Gene* 1993, **135**:275–8.

48. Kappes SM, Keele JW, Stone RT, McGraw RA, Sonstegard TS, Smith TP, Lopez-Corrales NL, Beattie CW: **A second-generation linkage map of the bovine genome.***Genome research* 1997, **7**:235–49.

49. Rohrer GA, Alexander LJ, Hu Z, Smith TP, Keele JW, Beattie CW: **A comprehensive map of the porcine genome.***Genome research* 1996, **6**:371–91.

50. Groenen MA, Cheng HH, Bumstead N, Benkel BF, Briles WE, Burke T, Burt DW, Crittenden LB, Dodgson J, Hillel J, Lamont S, De Leon AP, Soller M, Takahashi H, Vignal A: **A consensus linkage map of the chicken genome.***Genome research* 2000, **10**:137–47.

51. Maddox JF, Davies KP, Crawford AM, Hulme DJ, Vaiman D, Cribiu EP, Freking BA, Beh KJ, Cockett NE, Kang N, Riffkin CD, Drinkwater R, Moore SS, Dodds KG, Lumsden JM, Van Stijn TC, Phua SH, Adelson DL, Burkin HR, Broom JE, Buitkamp J, Cambridge L, Cushwa WT, Gerard E, Galloway SM, Harrison B, Hawken RJ, Hiendleder S, Henry HM, Medrano JF, et al.: **An enhanced linkage map of the sheep genome comprising more than 1000 loci.***Genome research* 2001, **11**:1275–89.

52. Ohta T, Kimura M: **A model of mutation appropriate to estimate the number of electrophoretically detectable alleles in a finite population*.***Genetical research* 2007, **89**:367–70.

53. Estoup A, Jarne P, Cornuet J-M: **Homoplasy and mutation model at microsatellite loci and their consequences for population genetics analysis.***Molecular ecology* 2002, **11**:1591–604.

54. Ignal A V, Ilan DM, Vignal A, Milan D, SanCristobal M, Eggen A: **A review on SNP and other types of molecular markers and their use in animal genetics.***Genetics, selection, evolution : GSE* 2002, **34**:275–305.

55. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES: **Linkage disequilibrium in the human genome.***Nature* 2001, **411**:199–204.

56. Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TPL, Sonstegard TS, Van Tassell CP: **Development and characterization of a high density SNP genotyping assay for cattle.***PloS one* 2009, **4**:e5350.

57. Ramos AM, Crooijmans RPM a, Affara N a, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu Z-L, Kerstens HH, Law AS, Megens H-J, Milan D, Nonneman DJ, Rohrer G a, Rothschild MF, Smith TPL, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen M: **Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology.***PloS one* 2009, **4**:e6524.

58. Groenen MAM, Megens H-J, Zare Y, Warren WC, Hillier LW, Crooijmans RPMA, Vereijken A, Okimoto R, Muir WM, Cheng HH: **The development and characterization of a 60K SNP chip for chicken.***BMC genomics* 2011, **12**:274.

59. Ernst CW, Steibel JP: **Molecular advances in QTL discovery and application in pig breeding.***Trends in genetics : TIG* 2013, **29**:215–224.

60. Downing T, Imamura H, Decuypere S, Clark TG, Coombs GH, Cotton JA, Hilley JD, Doncker S De, Maes I, Mottram JC, Quail MA, Rijal S, Sanders M, Stark O, Sundar S, Vanaerschot M, Hertz-fowler C, Dujardin J, Berriman M: **Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance**. 2011:2143–2156.

61. Gautier M, Laloë D, Moazami-Goudarzi K: **Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds.***PloS one* 2010, **5**.

62. Decker JE, Pires JC, Conant GC, McKay SD, Heaton MP, Chen K, Cooper A, Vilkki J, Seabury CM, Caetano AR, Johnson GS, Brenneman RA, Hanotte O, Eggert LS, Wiener P, Kim J-J, Kim KS, Sonstegard TS, Van Tassell CP, Neibergs HL, McEwan JC, Brauning R, Coutinho LL, Babar ME, Wilson GA, McClure MC, Rolf MM, Kim J, Schnabel RD, Taylor JF: **Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics.***Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**:18644–9.

63. Ponsuksili S, Murani E, Brand B, Schwerin M, Wimmers K: **Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model.***Journal of lipid research* 2011, **52**:668–78.

64. Wang JY, Luo YR, Fu WX, Lu X, Zhou JP, Ding XD, Liu JF, Zhang Q: **Genome-wide association studies for hematological traits in swine.***Animal genetics* 2012.

65. Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz L a. F, Schook LB, Crooijmans RPM a., Groenen M a. M: **Regions of Homozygosity in the Porcine Genome:**

Consequence of Demography and the Recombination Landscape. *PLoS Genetics* 2012, **8**:e1003100.

66. Albrechtsen A, Nielsen FC, Nielsen R: **Ascertainment Biases in SNP Chips Affect Measures of Population Divergence Research article**. 2010, **27**:2534–2547.
67. Esteve-Codina A, Kofler R, Himmelbauer H, Ferretti L, Vivancos AP, Groenen MAM, Folch JM, Rodríguez MC, Pérez-Enciso M: **Partial short-read sequencing of a highly inbred Iberian pig and genomics inference thereof.***Heredity* 2011, **107**:256–64.
68. Rubin C-JC-J, Megens H-JH-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg O, Jern P, Jorgensen CB, Archibald AL, Fredholm M, Groenen MAM, Andersson L, Martinez Barrio A, Carlborg Ö, Jørgensen CB: **Strong signatures of selection in the domestic pig genome.***Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**:19529–36.
69. Singh AP, Zafer S, Pe'er I: **Metaseq: privacy preserving meta-analysis of sequencing-based association studies.***Pacific Symposium on Biocomputing. Pacific Symposium on Biocomputing* 2013:356–67.
70. Mir KU: **Sequencing genomes: from individuals to populations.***Briefings in functional genomics & proteomics* 2009, **8**:367–78.
71. Lynch M: **Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects.***Molecular biology and evolution* 2008, **25**:2409–19.
72. Frankham R, Ballou JD, Briscoe DA: *Introduction to Conservation Genetics*. Cambridge . Cambridge: 2002.
73. Frankham R: **Genetics and conservation biology**. *Comptes Rendus Biologies* 2003, **326**:22–29.
74. Barrett RDH, Schluter D: **Adaptation from standing genetic variation.***Trends in ecology & evolution* 2008, **23**:38–44.
75. Frankham R, Ralls K: **Conservation biology: Inbreeding leads to extinction**. *Nature* 1998, **392**:441–442.
76. Charlesworth B, Charlesworth D: **Some evolutionary consequences of deleterious mutations.***Genetica* 1998, **102-103**:3–19.
77. Kohn MH, Murphy WJ, Ostrander E a, Wayne RK: **Genomics and conservation genetics.***Trends in ecology & evolution* 2006, **21**:629–37.
78. Ouborg NJ: **Integrating population genetics and conservation biology in the era of genomics.***Biology letters* 2010, **6**:3–6.
79. Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen M a M, Crooijmans RPM a, Megens H-J, Zhang H, Okimoto R, Vereijken A, Jungerius A, Albers GAA, Lawley CT, Delany ME, MacEachern S, Cheng HH: **Genome-wide assessment of**

worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds.*Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:17312–7.

80. FAO: **Building on gender agrobiodiversity and local knowledge**. *Food and Agriculture Organization of the United Nations* 2006, **1**:1–5.
81. FAO: *Intergovernmental Technical Working Group on Animal Genetic Resources for Food and Agriculture. Seventh Session*. Rome: 2012.
82. Berthouly-Salazar C, Thévenon S, Van TN, Nguyen BT, Pham LD, Chi CV, Maillard J-C: **Uncontrolled admixture and loss of genetic diversity in a local Vietnamese pig breed.***Ecology and evolution* 2012, **2**:962–75.
83. Ollivier L: **European pig genetic diversity: a minireview.***Animal : an international journal of animal bioscience* 2009, **3**:915–24.
84. Hall SJ, Bradley DG: **Conserving livestock breed biodiversity.***Trends in ecology & evolution* 1995, **10**:267–70.
85. Renaudeau D, Collin a, Yahav S, De Basilio V, Gourdine JL, Collier RJ: **Adaptation to hot climate and strategies to alleviate heat stress in livestock production.***Animal : an international journal of animal bioscience* 2012, **6**:707–28.
86. Hansen PJ: **Physiological and cellular adaptations of zebu cattle to thermal stress.***Animal reproduction science* 2004, **82-83**:349–60.
87. Holderegger R, Kamm U, Gugerli F: **Adaptive vs. neutral genetic diversity: implications for landscape genetics**. *Landscape Ecology* 2006, **21**:797–807.
88. Markert JA, Champlin DM, Gutjahr-Gobell R, Grear JS, Kuhn A, McGreevy TJ, Roth A, Bagley MJ, Nacci DE: **Population genetic diversity and fitness in multiple environments.***BMC evolutionary biology* 2010, **10**:205.
89. Hedrick PW: **Conservation genetics: where are we now?***Trends in Ecology & Evolution* 2001, **16**:629–636.
90. Reed DH, Frankham R: **Correlation between Fitness and Genetic Diversity**. *Conservation Biology* 2003, **17**:230–237.
91. Ellegren H, Sheldon BC: **Genetic basis of fitness differences in natural populations.***Nature* 2008, **452**:169–75.
92. Crandall K, Bininda-Emonds O, Mace G, Wayne R: **Considering evolutionary processes in conservation biology.***Trends in ecology & evolution* 2000, **15**:290–295.

2

Phylogeography of worldwide pig populations derived from mitochondrial DNA analysis

Herrero-Medrano JM[1], Larson G[2], Shapiro B[3] Groenen MAM[1], Megens HJ[1], Crooijmans RP[1]

[1]Animal Breeding and Genomics Centre, Wageningen University, The Netherlands;
[2]Department of Archaeology, Durham University, Durham DH1 3LE, UK;
[3]Department of Ecology and Evolutionary Biology, University of California Santa Cruz, Santa Cruz, California, United States of America

Paper in preparation (2013)

Abstract

Humans have influenced genetic variation of *Sus scrofa* by domestication, but also by marginalizing the natural environment, hunting and re-stocking of the wild species over thousands of years. As a result, the phylogeography of *Sus scrofa* is complex and not fully understood. In order to unravel historical dispersal patterns of wild boar, we applied a Bayesian phylogeographic inference to identify historical dispersal patterns. We analysed mtDNA sequences from 850 wild pigs selected from a wide distribution across Eurasia and North Africa. In addition, to explore the potential effect of domestic introgression in the analysis, we compared these to 2423 mtDNA sequences from domestic and feral pigs from across the same wide region. We observed significant dispersal events consistent with the expansion of *Sus scrofa* from South East Asia to the rest of the continent and with the arrival of wild pigs to Japan and Ryukyu islands. In addition, we propose a novel dispersal route linking Siberia with South Asia. Regarding Europe, we observed signs of postglacial re-colonization from southern refugia to centre Europe. Finally, we discuss the limitations of phylogeographic analysis to infer the dispersal history of wild pigs. In particular, admixture between domestic and wild pigs strongly influences the analysis. Moreover, the prior determination of the geographic regions as well as the inclusion on the analysis of specific samples often conditioned the significance of the dispersal event. We conclude that the asymmetric discrete diffusion model as implemented in BEAST provides insights of past dispersal events of wild pigs, although the large human influence of pigs history reinforce the need of being cautions in the design of the study.

Key words: demography, mtDNA, pig, phylogeography, Bayesian inference

2.1 Introduction

Sus scrofa originated in South East Asia 5.3-3.5 million years ago [1]. Larson et al. [2] studied mtDNA from wild and domestic pigs and concluded a scenario where wild pigs expanded from Island Southeast Asia (ISEA) to East Asia before migrating to the West and colonizing Europe through Middle East. By adding whole-genome sequence data, Groenen et al. [1] supported the East-West expansion, specifying that migration may have taken place from North-eastern Eurasian populations to Europe along the Pleistocene. During this period, repeated glaciations, especially the Last Glacial Maximum (~ 20,000 years ago), likely shaped current genetic variability of *Sus scrofa*, as it did for other species [3–5]. Pleistocene glaciations caused an important decrease of the population size of wild pigs throughout Eurasia [1], and forced migrations of surviving populations that became marginalized in warmer areas, so-called refugia [4]. The repeated glaciations also brought about sea-level oscillations [6, 7], facilitating migrations or causing isolation of island populations. While the phylogeography of many species has been mostly influenced by geo-climatic events, the demographic history of wild pig populations have also been influenced by humans activities, especially domestication, deforestation, hunting and population re-stocking [2, 5, 8]. These extrinsic factors, together with intrinsic factors such as its high reproduction rate and adaptability to different environments [9], have resulted in a complex phylogeography.

Mitochondrial DNA (mtDNA) evolves faster than nuclear DNA [10], it is maternally inherited and does not recombine [11, 12], which, combined, explains its use in phylogenetic studies. Dispersal or migration patterns of wild boar have been hypothesized by analysing the topology of phylogenetic trees based on molecular data (e.g. [2, 13]). These studies used parsimony inferences or Bayesian approach that did not include prior specifications of the geographical distribution of the sampling locations. Recent advances have been made in Bayesian phylogeographic modeling. The method conducted by Lemey et al. [14] to test the phylogeographic history of virus populations did overcome limitations of previous approaches. This improved methodology has been recently applied in several studies [15–18]. For instance, Marske et al. [16] studied the role of environmental changes in the dispersal histories of four New Zealand forest beetles. Applying the method to large mammals, Edwards et al. [15] explored the present and past dynamics of brown bear and polar bear, pointing out the large influence of climate events on disperal of bears and the probably admixture origin of the modern polar bear.

Here we implemented the same spatially explicit Bayesian inference approach in order to infer the historical dispersal patterns of wild boar populations across Eurasia using mtDNA. This Bayesian framework is applied for the first time in pig. Differently to other species such as beetle or bears, the aspect of domesticated pigs overlapping with much of the vast natural range of *Sus scrofa* has the potential to complicate the analysis in this species. Therefore, the effects of recent introgression from domestic pigs into wild boar were assessed by analysing more than 200 populations of wild, feral and domestic pigs with wide distribution throughout Eurasia. This study explores the historical dispersal patterns of wild boar and, at the same time, gives insight of the sensitiveness to sampling design and limitations in applying spatially explicit Bayesian inference in *Sus Scrofa*.

2.2 Results

A total of 3249 sequences were grouped into 467 haplotypes (Supplementary material_1). Of those, 231 haplotypes were exclusively present in wild pigs and 137 were carried by both wild boars and domestic/feral pigs. The initial dataset of wild boar sequences was split into two groups based on their haplotype (Table 2.1). The first group, hereafter called SET_1, included exclusively those wild pigs carrying haplotypes not found in any domestic or feral pig. The second group, SET_2, included all the sequences of wild pigs regardless its haplotype. The relationships between the inferred haplotypes were visualized using median-joining networks. The discrete traits model as implemented in BEAST [14] was implemented to obtain a realistic estimation of historical dispersal patterns of wild boar across Eurasia. To implement this method, 14 geographic regions were determined in accordance with biogeographic features and incorporated to the model (Table 2.1). The significance of the migration pathways between those regions was computed by Bayes factor test (BF). These Phylogeographic analyses were performed separately for SET_1 and SET_2 to evaluate the sensitivity of the method to recent domestic introgression. Since the number of samples per geographic regions varied widely, we developed a random selection process to sample 15 sequences per geographic region for each group to avoid bias due to differences in sample size (Table S3, Supplementary material_2). Additionally, genetic diversity estimates and population expansion analysis were computed (Supplementary material_2).

Table 2.1 Number of sequences in each set of data.

Regions	Geographic Region	Code	SET_1	SET_2
R01	Japan	R01_JAP	18	21
R02	Ryukyu Islands	R02_RYU	13	13
R03	Peninsula Malaysia/ISEA	R03_MAL	14	15
R04	North East Asia	R04_NEA	117	123
R05	Central Asia/China	R05_CAS	45	74
R06	South Asia/India	R06_SAS	42	47
R07	South East Asia/Thailand	R07_SEA	59	89
R08	Taiwan island	R08_TAW	12	13
R09	Middle East	R09_NES	39	45
R10	North Africa	R10_NAF	1	9
R11	Iberian Peninsula	R11_IBP	40	43
R12	Italic Peninsula	R12_ITP	20	60
R13	Balkan peninsula	R13_BKP	58	74
R14	Central Europe	R14_CEU	117	295
		Total	595	921

Diffusion in Asia

South East Asia (R07:SEA) presented the most significant migration pathways in Asia, both towards the south with the Islands of South East Asia (R03:MAL) and towards the north with Central Asia (R05:CAS) (Figure 2.1). The South East Asia region also presented the highest number of haplotypes (Number_Hap = 52) and mtDNA diversity (h = 0.98) (Table S1, Supplementary material_2).

North East Asia (R04:NEA) had a significant link with Japan (R01:JAP) (SET_1, BF = 32) representing the best supported link between the main land and the Asian eastern islands –Japan, Ryukyu and Taiwan–. This pathway was in agreement with the similarity observed between haplotypes carried by pigs from Japan and pigs from south east Siberia (Figure 2.2). A detailed study of Asian eastern islands revealed significant dispersal patterns (BF> 39) between Japan (R01:JAP) and Ryukyu Islands (R02:RYU). No related haplotypes and no significant link were found between Taiwan and Ryukyu Islands. Taiwan, however, was linked with Central Asia in the analysis performed with the SET_2, in agreement with the occurrence of related and common haplotypes carried by wild boars from Taiwan and Chinese wild and domestic pigs (SS477 and SS320 respectively). Significant links were observed between Japan and both Central Asia and South East Asia, but only for SET_2. The study of the Network graphic showed that a haplotype common in

Japanese and Vietnamese wild boar was also highly frequent in domestic pigs (SS207).

South Asia/India region (R06:SAS) was the region with fewest number of significant links in Asia, presenting only significant BF with the adjacent region of South East Asia/Thailand (R07:SEA) and with North East Asia (R04:NEA). Of these connections only four out of the 20 analysis presented a BF > 8 in both pathways. Two highly differentiated clusters of wild boars from South Asia/India were observed in the network analysis. The first cluster included pigs from this region and Russian wild boars, and the second cluster encompassed India and South East Asia.

Diffusion in Middle East

Two significant migration pathways were found between Middle East (R09:NES) and European regions. (Figure 2.3). The pathway between Middle East and central Europe (R14:CEU) was only supported by the SET_1 with a BF = 10 when samples from Malcantone region (Swiss Alps) were included. The pathway between Middle East and the Italian Peninsula (R12:ITP) was identified only in SET_2 (BF= 16). Network analysis revealed that Middle East haplotypes occupied an intermediate position between Asia and Europe (Figure 2.4), including five Iranian wild boar clustering in the Asian clade. Wild pigs from Armenia, Iraq and Israel had haplotypes in common with domestic and wild European pigs (Figure 2.4). No dispersal patterns were detected between Middle East and Asia. The most likely migration pathway that linked the regions Middle East and North East Asia (R04:NEA) gave a BF lower than 8 (BF= 5.3). This North East Asia to Middle East pathway depended entirely on the inclusion of a single Russian wild boar collected around the Volga river, the only one we could sample for the analysis, that clustered together with Middle East pigs in the network analysis.

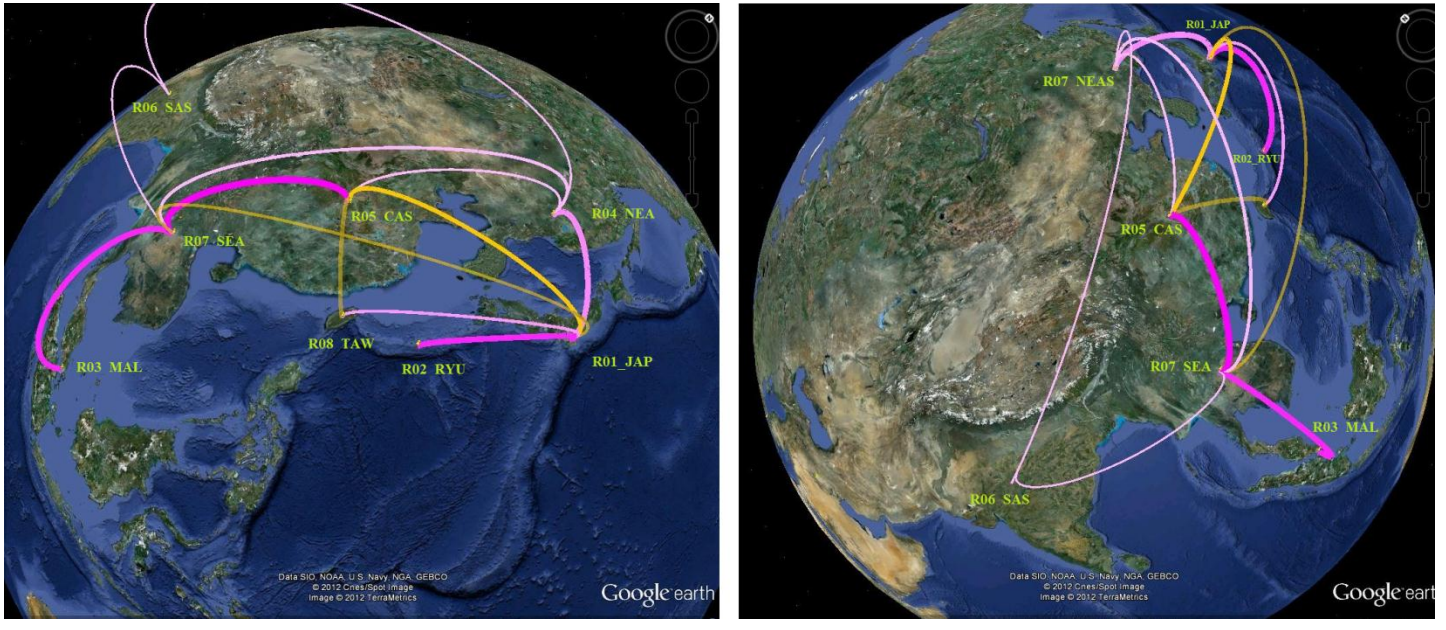

Figure 2.1 Migration pathways in Asia with BF > 8. Colour code: Pink, consistent pathway across sets 1 and 2; Red: path ways consistent only in SET_1; Yellow: pathway consistent only in SET_2 . The intensity of the colour is proportional to the BF value. The thickness of the pathway is proportional to the consistence of the pathway.

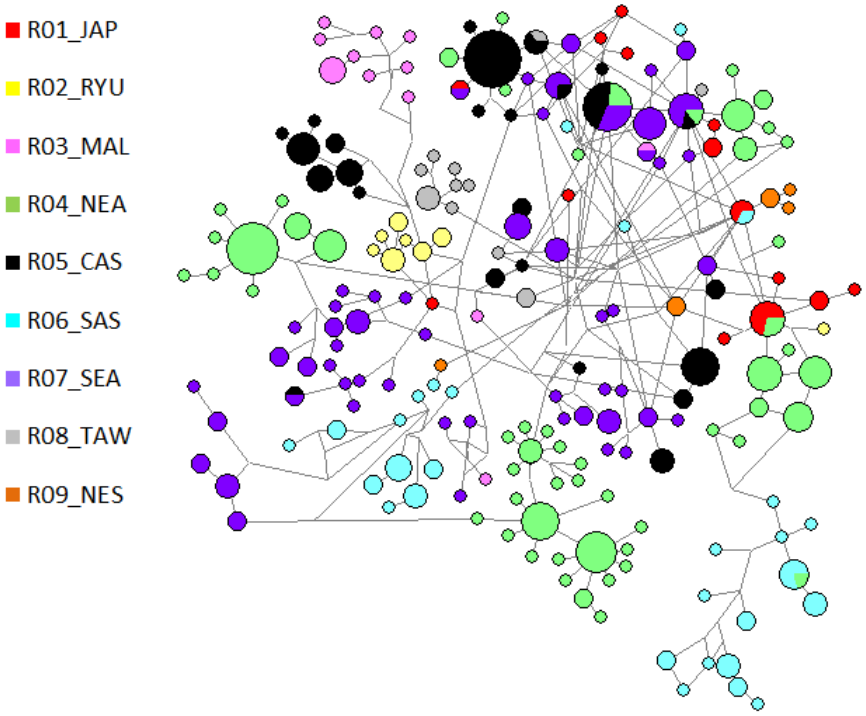

Figure 2.2 Medium-joining networks of Asian mtDNA haplotypes. The size of the circle is proportional to the frequency of the haplotype.

Diffusion in Europe and North Africa

The number of well-supported dispersal pathways was much lower in Europe than in Asia. Three highly significant and consistent pathways linked Central Europe (R14:CEU) with the Mediterranean Iberian (R11:IBP) (SET_1; BF = 70), Italian (R12:ITP) (SET_1; BF > 2000) and Balkan (R13:BKP) (SET_1; BF = 92) peninsulas. The Central Europe region (R14:CEU) contained 53 haplotypes shared between wild and domestic pigs representing the highest number among all regions. The three Mediterranean Peninsulas had haplotypes clustering together with Central European pigs. The exceptions were pigs from Italy and from Malcantone region (South of Swiss Alps), which carried haplotypes that clustered with Middle East haplotypes (Figure 2.4).

North Africa (R10:NAF) was only linked with Iberian Peninsula by a migration pathway that was significant in SET_1 (BF = 17) (Figure 2.3). A wild boar from Morocco carried one of the most common haplotypes in domestic pigs from Europe (SS029) while wild boars from Tunisia had haplotypes that cluster with European pigs (SS308 and SS611).

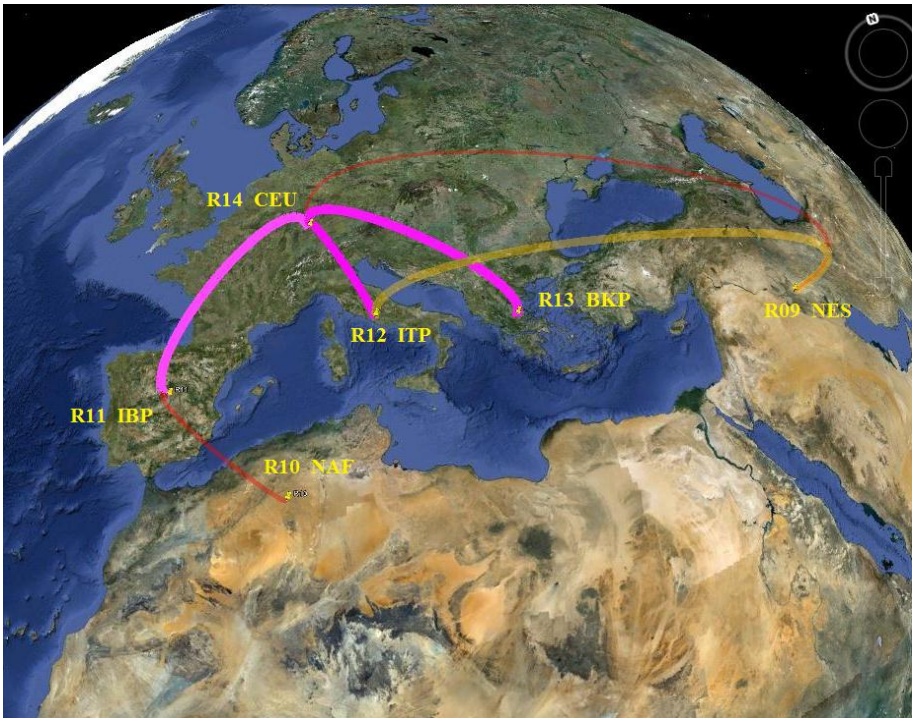

Figure 2.3 Migration pathways in Middle East, Europe and North Africa with BF > 8. Colour code: Pink, consistent pathway across sets 1 and 2; Red: pathways consistent only in SET_1; Yellow: pathway consistent only in SET_2 . The intensity of the colour is proportional to the BF value. The thickness of the pathway is proportional to the consistence of the pathway in each set.

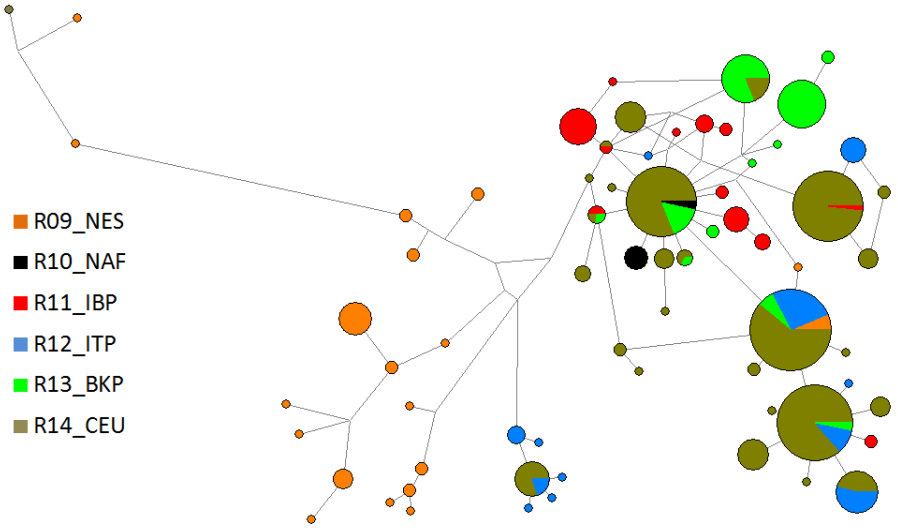

Figure 2.4 Medium-joining network of Middle East, Europe and North Africa mtDNA haplotypes. The size of the circle is proportional to the frequency of the haplotype.

2.3 Discussion

Biogeographic research in *Sus scrofa* had led to general assumptions of dispersal patterns based on the topology of phylogenetic trees and the geographic origin of the samples, e.g. [2]. This approach implies that multiple, different migratory scenarios are consistent with the genetic data, but they do not explicitly test dispersal patterns. We performed a Bayesian phylogeographic analysis on mitochondrial DNA to reconstruct dispersal patterns of wild boar between 14 geographic regions throughout Eurasia. Admixture between domestic and wild populations hamper the study of demographic history, particularly in studies based on mtDNA [19]. As a result of the overlapping ranges of wild and domestic pigs during thousands of years, repeated events of hybridization have occurred [20] as demonstrated by molecular data [21]. The effect of hybridization was taken into account, and thus mtDNA sequences of wild boars were compared with a comprehensive dataset of domestic and feral pigs, the largest compiled to date, resulting in the differentiation of SET_1 (wild pigs carrying haplotypes not detected in domestic pigs) and SET_2 (those wild pigs with haplotypes observed in domestic pigs).

Diffusion in Asia

Fossil records of Pleistocene mammalian fauna and molecular data indicate an expansion of *Sus scrofa* species from southeast to northeast on the Asian Continent [2, 22]. Our results support the South East Asia region as the origin of dispersal events of *Sus scrofa*. The low sea-level during glacial periods of the Pleistocene formed a land bridge between South East Asian mainland and Island South East Asia (ISEA) enabled dispersal events of large mammals and other organisms[23], which explains the significant dispersal pathway linking these regions. Moreover, signs of recent demographic expansion, network analysis, and the significant dispersal pathway all indicate repeated dispersal events between South East Asia and Central Asia. We observed a significant pathway between South East Asia and North East Asia regions, mainly in SET_1. This is consistent with Cho et al. [13] who suggested a land bridge between Korea and Southern China based on mtDNA data. Given the weak link between Central Asia and North East Asia regions, we suggest a stepping stone model, where only a subset of migrants that settled in central China moved further toward northern East Asian regions.

Land bridges enabled the gradual dispersion of plants, birds and mammals from continental to islands. Vertebrate paleontological studies revealed the existence of two main land bridges connecting Japan with the main land. The first land bridge connected the Korean peninsula with the island of Kyushu during the Middle Pleistocene [24, 25]. The second land bridge connected Siberia and what is now known as Hokkaido via the island of Sakhalin, serving as a migration route for mammals in the middle and late Pleistocene [22, 26]. The well-supported migration pathway between North East Asia and Japan, together with the existence of highly related haplotypes carried by wild boar from Japan and from southeast Siberia strongly indicates the second land bridge as the major migration pathway from the continent to Japan. The link between Japan and Ryukyu Islands forms the most significant connection among Eastern Asian islands. This is in agreement with Kawamura [22], who observed that migrant species of animal and plants arrived via land bridges from Southern Japan to southern Ryukyu Islands in the Middle-Late Pleistocene.

Larson et al. [27] observed two well-differentiated wild boar lineages in the South Asia –India region– noting the peculiarity of such high genetic differentiation in a relatively small geographic region. Applying a far larger sample of wild boar mitochondrial haplotypes from India, we observed a similar signature in the network analysis. The two migration pathways linking South Asia with South East Asia and with North East Asia region may explain the existence of those highly differentiated mtDNA lineages in India. The significant migration pathway that links

South Asia with North East Asia regions is particularly interesting since it reveals a migration route not previously described in wild boar between Siberia and India. The inclusion in the analysis of a Russian wild boar from Lake Baikal that shared its haplotype with a Northwest Indian pigs was important for the statistical consistency of the migration pathway. Since only samples from the Northwest of India had haplotypes related with Russian haplotypes, and considering Himalaya as a geographic barrier, this novel north-south migration pathway would have extended across the western boundary of the Himalayas. A higher number of samples from Siberia are needed to confirm this migration route.

The analysis of the SET_2 revealed three dispersal pathways that were not observed in the analysis performed in the SET_1 (see yellow links in Figure 2.2). While the link between Central Asia/China region with Taiwan could be due to the land bridge that appeared in the Holocene period, it is very unlikely that the significant dispersal events linking Central China with Japan, and the latter with South East Asia, correspond to a natural dispersal events of wild boar. Indeed, the links between these regions were only observed when haplotypes shared with domestic pigs were included in the analysis. This highlights the limitation of this Bayesian inference approach when admixed animals are included.

Diffusion in Middle East

In order to unravel the dispersal pathways used by the Asian wild boars to colonize Europe, we tested the significance of all the Asian regions with Middle East and Europe. The only trace of connection between Asia and western regions was found between North East Asia and Middle East. This link was found in the 50% of the analyses using SET_1 samples, though with a BF= 5.3 that was not significant considering our threshold of BF = 8 [15]. However, it is interesting that the use of whole-genome data supports the same dispersal pathway [1], and thus it can be speculated that North Asian pigs moved from North East Asia north of the Himalayas. From this location, *Sus scrofa* could have migrated either westbound to Caspian Sea, directly to Europe, or via a route south of the Caucasus to the Middle East. The antiquity of this East-West dispersal process –Early Pleistocene– could have obscured mtDNA signs of migration.

We did observe significant pathways linking Middle East with Europe. Larson et al. [28] and more recently Ottoni et al. [20] demonstrated that European mtDNA haplotypes replaced Near Eastern domestic pigs as a result of human migrations or trade of animals from Europe to Middle East. Near Eastern wild pigs carrying European type haplotypes, as a result of admixture between wild and domestic pigs in Middle East, explains the link between Middle East and Italy in the Set_2.

Regarding the dispersal event between Middle East and Central Europe, the inclusion of Swiss pigs from Malcantone region carrying the mtDNA haplotypes indigenous from Italy (haplotype code E2 in [5, 29] and D4 in [2]) was decisive for the significance of this dispersal event.

Diffusion in Europe and North Africa

During the Last Glacial Maximum (LGM), temperate species migrated to the southern peninsulas of Europe. This was followed by a postglacial re-colonization of north and central Europe [4, 5, 30–33]. Our analysis revealed highly significant and consistent pathways from Iberian, Italic and Balkan Peninsulas to Central Europe region reflecting genetic signs of this re-colonization event. Participation of the Balkan region in the postglacial colonization of central and north Europe in *Sus scrofa* has been observed using mtDNA [34]. The consistent dispersal event detected and the large proportion of haplotypes shared by wild boar from the Balkan and central European regions, even larger than haplotypes shared by the Iberian and Italic Peninsulas with the central European regions, support the Balkans as a major re-colonization area [4, 30].

Two different clades of haplotypes has been described in Italy: clade E2, endemic to Italy, and clade E1 common to central European wild boars [5, 28]. The dispersal pathway between the Italian Peninsula and the central European region was solely due to the haplotypes of clade E1 (results not shown). Endemic wild boar from Italy would have been highly adapted to the Mediterranean environment and with the Alps representing an important geographic barrier hampering natural migrations. The existence of a pathway with an extremely high BF (BF > 2000) could be due to (i) recent and bidirectional migrations of wild pigs between central Europe and Northern regions of Italy [35]; (ii) the reintroduction of central European wild boars into Italy after World War II [36]; (iii) The inclusion of Malcantone region within the region central Europe. This region has the peculiarity of being located in the Swiss Alps, exactly in the boundary separating Italian and European regions, and also of having high variability of mtDNA haplotypes. Thus, the analysis performed can not unravel the role of Italian pigs in the postglacial re-colonization and exemplify the difficulty of determining the boundaries between geographic regions.

A well-supported pathway connected North Africa with the Iberian Peninsula only in the SET_1 of samples. Ramirez et al. [37] observed a shared mtDNA haplotype between Europe, North Africa and Middle East suggesting a common ancient haplotype to those regions. Differently, our results are consistent with an Iberian origin of the pig populations from North Africa. The similarities between haplotypes detected in North Africa and central Europe might be due to a common haplotype

originated in Spain. In this scenario, Iberian wild pigs migrated to both, north Africa swimming through the strait of Gibraltar or carried by humans, possibly by the Romans who transported many large mammalian species across the Mediterranean to feature in arenas.

2.4 Material and methods

Sample sequencing

Blood, tissue or hair samples from 390 wild boars, 43 feral and 884 domestic pigs were collected and the DNA was extracted using standard protocols. A 722 bp fragment of the mtDNA D-loop region, corresponding to positions 15451–16088 of the reference mtDNA genome, was amplified by polymerase chain reaction (PCR). The PCR amplicons were purified and sequenced for both strands on an ABI 3130® DNA sequencer (Applied Biosystems, USA). Since not all of the samples yielded the entire sequenced fragment, a 642 bp fragment was finally used for the analysis. The *novo* sequences were combined with sequences from 460 wild boars 57 feral and 1439 domestic pigs from Genebank entries with worldwide distribution. In total, the dataset encompassed 230 populations of wild pigs (850 individuals) from Eurasia and North Africa as well as domestic (2323 individuals) and feral pigs (100 individuals) from 70 countries throughout Europe, Asia, Africa, America and Oceania (Table S2, Supplementary material_2).

Alignment and haplotype determination

D-loop region was visualised and exported with Genome Assembly Program (GAP4, [39] using the mtDNA pig sequence GenBank AJ002189 [40] as reference sequence. ClustalX V.2 [41] and ALTER [42] were respectively used to the alignment and subsequent grouping of the entire dataset of sequences -wild, domestic and feral- in haplotypes.

Groups design

Eurasia was split into 14 geographic regions according to biogeographic features such as zoogeography, vegetation and geographic barriers [23, 43–45]. Wild boars were assigned to their geographic region based on the origin of the sample. Wild boars whose haplotypes were exclusively carried by wild pigs but not by any domestic pig were assigned to the group SET_1. The entire dataset from wild boar, regardless of the haplotype, were assigned to the group SET_2 (Table 2.1). Additionally, European wild boars that had Asian-like haplotypes or vice versa where excluded to be considered a compelling sign of domestic introgression.

Genetic diversity and mismatch distribution

We used ARLEQUIN 3.5 [46] to estimate haplotype diversity (h), nucleotide diversity (π), number of haplotypes and number of polymorphic sites (κ) within all the geographic regions of SET_1 and SET_2 separately. This software was also used to test for historical demographic patterns by calculating mismatch distributions of pairwise nucleotide differences between haplotypes. The sum of square deviations (SSD) between the observed and the expected mismatch was used as a statistic test for the departure from the model of population expansion [47]. Additionally, Fu's Fs and Tajima's D neutrality tests were calculated to obtain additional insights of population expansion history.

Discrete phylogeographic diffusion model

The spatial dynamics of wild boars were inferred using an asymmetric discrete diffusion model implemented within the Bayesian inference framework of BEAST v1.7.2 [47].

The mtDNA sequences were assigned to one of the 14 discrete geographic regions previously determined. Both, mtDNA sequences and discrete geographic regions were used as BEAST input data. Phylogeographic relationships between geographic regions were estimated using a continuous-time Markov chain (CTMC) model [14] extended to allow for different rates of diffusion between locations depending on the direction traveled [15]. For a number of k discrete geographic regions, k(k-1) different dispersal patterns are possible, but not all of them are realistic in nature. Therefore, the discrete phylogeographic analysis was extended with the Bayesian Stochastic Search Variable Selection (BSSVS) to select only the most informative dispersal scenarios. Under BSSVS, we assume a truncated Poisson prior on the number of nonzero rates and that rates are, a priori, independent and gamma-distributed [14]. This analysis enables computation of a Bayes Factor (BF) test [48] that establishes the parsimonious descriptions of the phylogeographic diffusion in order to identify significant migration pathways between geographic regions [14, 15]. Those migration pathways with a BF > 8 were considered as well-supported diffusion rates between geographic regions [15].

We assumed the HKY substitution model [49] and Bayesian skyline [50] coalescent priors [51]. To estimate probable migration pathways, diffusion rate parameters and gene genealogies we ran Markov chain Monte Carlo (MCMC) simulations of 100,000,000 iterations. We discarded the initial 10% of realizations as chain burn-in and sub-sampling every 10,000 iterations to decrease autocorrelation. We used Tree Annotator [47] to yield a Maximum Clade Credibility (MCC) consensus tree. Finally, to provide a spatial projection of the diffusion patterns, the MCC tree was

converted into a keyhole markup language (KML) file suitable for viewing with Google Earth (http://earth.google.com).

The analyses described above were applied sequentially to the two sets of data previously described –SET_1 and SET_2–. To overcome the potential bias due to the differences in sample size between geographic regions, the analysis was computed multiple times selecting randomly 15 sequences per geographic region and group of samples (Table S3, supplementary material_2). As a result of this resampling procedure, 20 subsets were generated: 10 subsets corresponding to SET_1 and 10 to the SET_2.

Network analysis

To obtain additional information of the relations among haplotypes we used the Median-Joining method implemented by Network 4.6.0.0 (Fluxus Technology, http://www.fluxus-engineering.com).

Supporting Information

Chapter_2.zip

https://mega.co.nz/#!DBkxHKyL!WN-_uzv5Y_Em7Bu-DMZjpHO2Irh7eGOEXWQ3SlmLPDU

References

1. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-J, Li S, Larkin DM, Kim H, Frantz LAF, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S, Bujie Z, et al.: **Analyses of pig genomes provide insight into porcine demography and evolution.***Nature* 2012, **491**:393–8.
2. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, Rowley-Conwy P, Andersson L, Cooper A: **Worldwide phylogeography of wild boar reveals multiple centers of pig domestication.***Science (New York, N.Y.)* 2005, **307**:1618–21.
3. Hofreiter M, Serre D, Rohland N, Rabeder G, Nagel D, Conard N, Münzel S, Pääbo S: **Lack of phylogeography in European mammals before the last glaciation.***Proceedings of the National Academy of Sciences of the United States of America* 2004, **101**:12963–8.
4. Hewitt G: **The genetic legacy of the Quaternary ice ages.***Nature* 2000, **405**:907–13.
5. Scandura M, Iacolina L, Crestanello B, Pecchioli E, Di Benedetto MF, Russo V, Davoli R, Apollonio M, Bertorelle G: **Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable?***Molecular ecology* 2008, **17**:1745–62.
6. Siddall M, Rohling EJ, Almogi-Labin A, Hemleben C, Meischner D, Schmelzer I, Smeed DA: **Sea-level fluctuations during the last glacial cycle.***Nature* 2003, **423**:853–8.
7. Miller KG, Kominz MA, Browning J V, Wright JD, Mountain GS, Katz ME, Sugarman PJ, Cramer BS, Christie-Blick N, Pekar SF: **The Phanerozoic record of global sea-level change.***Science (New York, N.Y.)* 2005, **310**:1293–8.
8. Larson G, Albarella U, Dobney K, Rowley-conwy P, Tagliacozzo A, Manaseryan N, Miracle P, Wijngaarden-bakker L Van, Masseti M, Bradley DG, Cooper A: **Ancient DNA , pig domestication , and the spread**. 2007.
9. Bieber C, Ruf T: **Population dynamics in wild boar *Sus scrofa*: ecology, elasticity of growth rate and implications for the management of pulsed resource consumers**. *Journal of Applied Ecology* 2005, **42**:1203–1213.
10. Brown WM: **Rapid Evolution of Animal Mitochondrial DNA**. *Proceedings of the National Academy of Sciences* 1979, **76**:1967–1971.
11. Wolstenholme DR: **Animal mitochondrial DNA: structure and evolution.***International review of cytology* 1992, **141**:173–216.

12. Avise JC: *Molecular markers, natural history and evolution*. 1st edition. London: 1994.
13. Cho I-C, Han S-H, Fang M, Lee S-S, Ko M-S, Lee H, Lim H-T, Yoo C-K, Lee J-H, Jeon J-T: **The robust phylogeny of Korean wild boar (Sus scrofa coreanus) using partial D-loop sequence of mtDNA.***Molecules and cells* 2009, **28**:423–30.
14. Lemey P, Rambaut A, Drummond AJ, Suchard M a: **Bayesian phylogeography finds its roots.***PLoS computational biology* 2009, **5**:e1000520.
15. Edwards CJ, Suchard M a, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, Connell TCO, Coxon P, Monaghan N, Valdiosera CE, Lorenzen ED, Willerslev E, Baryshnikov GF, Rambaut A, Thomas MG, Bradley DG, Shapiro B, Tables S, O'Connell TC: **Ancient hybridization and an Irish origin for the modern polar bear matriline.***Current biology : CB* 2011, **21**:1251–8.
16. Marske KA, Leschen RAB, Buckley TR: **Concerted versus independent evolution and the search for multiple refugia: comparative phylogeography of four forest beetles.***Evolution; international journal of organic evolution* 2012, **66**:1862–77.
17. Marske KA, Leschen R a B, Buckley TR: **Reconciling phylogeography and ecological niche models for New Zealand beetles: Looking beyond glacial refugia.***Molecular phylogenetics and evolution* 2011, **59**:89–102.
18. De Bruyn A, Villemot J, Lefeuvre P, Villar E, Hoareau M, Harimalala M, Abdoul-Karime AL, Abdou-Chakour C, Reynaud B, Harkins GW, Varsani A, Martin DP, Lett J-M: **East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus.***BMC evolutionary biology* 2012, **12**:228.
19. Larson G, Burger J: **A population genetics view of animal domestication.***Trends in genetics : TIG* 2013, **29**:197–205.
20. Ottoni C, Flink LG, Evin A, Geörg C, De Cupere B, Van Neer W, Bartosiewicz L, Linderholm A, Barnett R, Peters J, Decorte R, Waelkens M, Vanderheyden N, Ricaut F-X, Cakirlar C, Cevik O, Hoelzel a R, Mashkour M, Karimlu AFM, Seno SS, Daujat J, Brock F, Pinhasi R, Hongo H, Perez-Enciso M, Rasmussen M, Frantz L, Megens H-J, Crooijmans R, Groenen M, et al.: **Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics.***Molecular biology and evolution* 2013, **30**:824–32.
21. Goedbloed DJ, Megens HJ, Van Hooft P, Herrero-Medrano JM, Lutz W, Alexandri P, Crooijmans RPM a, Groenen M, Van Wieren SE, Ydenberg RC, Prins HHT: **Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations.***Molecular ecology* 2012, **22**:856–866.

22. Kawamura Y: **Immigration of mammals into the Japanese islands during the Quaternary**. *The Quaternary Research* 1998, **37**:251–257. [in Japanese with English abstract].

23. Tougard C: **Biogeography and migration routes of large mammal faunas in South–East Asia during the Late Middle Pleistocene: focus on the fossil and extant faunas from Thailand**. *Palaeogeography, Palaeoclimatology, Palaeoecology* 2001, **168**:337–358.

24. Kawamura Y: **Honyurui (Mammals)**. *Gekkan Kokogaku (Archaeology Quarterly)* 2001, **74**:36–40. [in Japanese].

25. Kawamura Y: **Last glacial and Holocene land mammals of the Japanese islands: their fauna, extinction and immigration.***The Quaternary Research* 2007, **46**:171–177.

26. Takahashi K: **Nihon Retto no Senshin-Koshinsei ni okeru rikusei honyu dobutsu-so no keisei katei (The formative history of the terrestrial mamalian fauna of the Japanese Islands during the Plio-Pleistocene).***Kyusekki Kenkyu (Palaeolithic Research)* 2007, **3**:5–14.

27. Larson G, Liu R, Zhao X, Yuan J, Fuller D, Barton L, Dobney K, Fan Q, Gu Z, Liu X-H, Luo Y, Lv P, Andersson L, Li N: **Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA.***Proceedings of the National Academy of Sciences of the United States of America* 2010, **107**:7686–91.

28. Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, Vigne J-D, Edwards CJ, Schlumbaum A, Dinu A, Balaçsescu A, Dolman G, Tagliacozzo A, Manaseryan N, Miracle P, Van Wijngaarden-Bakker L, Masseti M, Bradley DG, Cooper A: **Ancient DNA, pig domestication, and the spread of the Neolithic into Europe.***Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**:15276–81.

29. Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L: **The origin of the domestic pig: independent domestication and subsequent introgression.***Genetics* 2000, **154**:1785–91.

30. Hewitt G: **Mediterranean Peninsulas: The Evolution of Hotspots**. In *Biodiversity Hotspots*. edited by Zachos FE, Habel JC Berlin, Heidelberg: Springer Berlin Heidelberg; 2011:123–147.

31. Taberlet P, Fumagalli L, Wust-Saucy AG, Cosson JF: **Comparative phylogeography and postglacial colonization routes in Europe.***Molecular ecology* 1998, **7**:453–64.

32. Stewart JR, Lister AM, Barnes I, Dalén L: **Refugia revisited: individualistic responses of species in space and time.***Proceedings. Biological sciences / The Royal Society* 2010, **277**:661–71.
33. Avise JC, Walker D, Johns GC: **Speciation durations and Pleistocene effects on vertebrate phylogeography.***Proceedings. Biological sciences / The Royal Society* 1998, **265**:1707–12.
34. Alexandri P, Triantafyllidis A, Papakostas S, Chatzinikos E, Platis P, Papageorgiou N, Larson G, Abatzopoulos TJ, Triantaphyllidis C: **The Balkans and the colonization of Europe: the post-glacial range expansion of the wild boar, *Sus scrofa***. *Journal of Biogeography* 2012, **39**:713–723.
35. De Beaux E, Festa E: **La ricomparsa del cinghiale nell'Italia Settentrionale-Occidentale.***Memorie Soc. It. Sc. Nat.* 1927, **9**:263–322.
36. Apollonio M, Randi E, Toso S: **The systematics of the wildboar (*Sus scrofa* L.) in Italy.***Bollettino Di Zoologia* 1988, **3**:213–221.
37. Ramírez O, Ojeda a, Tomàs a, Gallardo D, Huang LS, Folch JM, Clop a, Sánchez a, Badaoui B, Hanotte O, Galman-Omitogun O, Makuza SM, Soto H, Cadillo J, Kelly L, Cho IC, Yeghoyan S, Pérez-Enciso M, Amills M: **Integrating Y-chromosome, mitochondrial, and autosomal data to analyze the origin of pig breeds.***Molecular biology and evolution* 2009, **26**:2061–72.
38. Bonfield JK, Smith K f, Staden R: **A new DNA sequence assembly program.***Nucleic acids research* 1995, **23**:4992–9.
39. Ursing BM, Arnason U: **The complete mitochondrial DNA sequence of the pig (*Sus scrofa*).***Journal of molecular evolution* 1998, **47**:302–6.
40. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.***Bioinformatics (Oxford, England)* 2007, **23**:2947–8.
41. Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D: **ALTER: program-oriented conversion of DNA and protein alignments.***Nucleic acids research* 2010, **38**:W14–8.
42. Wallace AR: *The Geographical Distribution of Animals*. Macmillan . Londres: 1876.
43. Branca A, Paape TD, Zhou P, Briskine R, Farmer AD, Mudge J, Bharti AK, Woodward JE, May GD, Gentzbittel L, Ben C, Denny R, Sadowsky MJ, Ronfort J, Bataillon T, Young ND, Tiffin P: **Whole-genome nucleotide diversity, recombination, and linkage disequilibrium in the model legume Medicago truncatula.***Proceedings of the National Academy of Sciences of the United States of America* 2011, **108**:E864–70.

44. Encyclopædia Britannica.: **Encyclopædia Britannica Online Academic Edition. Encyclopædia Britannica Inc.** 2012.
45. Excoffier L, Lischer HEL: **Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.***Molecular ecology resources* 2010, **10**:564–7.
46. Schneider S, Excoffier L: **Estimation of past demographic parameters from the distribution of pairwise differences when the mutation rates vary among sites: application to human mitochondrial DNA.***Genetics* 1999, **152**:1079–89.
47. Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees.***BMC evolutionary biology* 2007, **7**:214.
48. Suchard MA, Weiss RE, Sinsheimer JS: **Bayesian selection of continuous-time Markov chain evolutionary models.***Molecular biology and evolution* 2001, **18**:1001–13.
49. Hasegawa M, Kishino H, Yano T: **Dating of the human-ape splitting by a molecular clock of mitochondrial DNA.***Journal of molecular evolution* 1985, **22**:160–74.
50. Drummond AJ, Rambaut A, Shapiro B, Pybus OG: **Bayesian coalescent inference of past population dynamics from molecular sequences.***Molecular biology and evolution* 2005, **22**:1185–92.
51. Minin VN, Bloomquist EW, Suchard MA: **Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics.***Molecular biology and evolution* 2008, **25**:1459–71.

3

Farm-by-farm analysis of microsatellite, mtDNA, and SNP genotype data reveals inbreeding and crossbreeding as threats to the survival of a native Spanish pig breed

Herrero-Medrano JM[1], Megens HJ[2], Crooijmans RP[2], Abellaneda JM[1], Ramis G[1]

[1]Departamento de Producción Animal, Facultad de Veterinaria, Universidad de Murcia, 30100, Murcia, Spain; [2]Animal Breeding and Genomics Centre, Wageningen University, 6708WD, Wageningen, The Netherlands.

Animal Genetics (2013) 44, 259-66

Abstract

The Chato Murciano (CM), a pig breed from the Murcia region in the South-East of Spain, is a good model for endangered livestock populations. The remaining populations are bred on around 15 small farms, and no herd book exists. To assess the genetic threats to the integrity and survival of the CM breed, and to aid in designing a conservation program, three genetic marker systems –microsatellites, SNPs and mtDNA– were applied across the majority of the total breeding stock. In addition, mtDNA and SNPs were genotyped in breeds that likely contributed genetically to the current CM gene pool. The analyses revealed levels of genetic diversity within the range of other European local breeds (He = 0.53). However, when the eight farms that rear at least 10 CM pigs were independently analysed, high levels of inbreeding were found in some. Despite the evidence for recent crossbreeding with commercial breeds on a few farms, the entire breeding stock remains readily identifiable as CM, facilitating design of traceability assays. The genetic management of the breed is consistent with farm size, farm owner, and presence of other pig breeds on the farm, demonstrating the highly ad-hoc nature of current CM breeding. The results of genetic diversity and substructure of the entire breed, as well as admixture and crossbreeding obtained in the present study, provide a benchmark to develop future conservation strategies. Furthermore, this study demonstrates that identifying farm-based practices and farm-based breeding stocks can aid in design of a sustainable breeding program for minority breeds.

Key words: Genetic diversity, pig, endangered breed, mitochondrial DNA, microsatellites, SNP, Chato Murciano.

3.1 Introduction

Local breeds are important for the maintenance of the genetic diversity and future food security as stated by the United Nations Food and Agricultural Organization (FAO). Moreover, domesticated animal breeds often are regarded as part of the cultural heritage of local and national communities. The FAO has warned that the decline of animal genetic resources is proceeding at an alarming rate [1], highlighted by the fact that nearly 20% of domestic animal breeds are threatened with extinction [2].

One of the local, heritage pig breeds that is categorized as endangered on the List for Domestic Animal Diversity Information System (DAD-IS, http://dad.fao.org/) is the Chato Murciano (CM) pig breed. The CM is autochthonous to the Region of Murcia (Spain) (Google Earth projection, supplementary material), and 'Chato' refers to the snub-nose characteristic of the breed. The ancestor of the breed is a rustic black Mediterranean pig, also referred to as primitive Murcia pig, that lived in the South-East of Spain at least 150 years ago [3]. Due to the high rusticity of the black Mediterranean pig, farmers designed crosses with breeds such as Berkshire, Large White, Retinto Iberian pig and Tamworth in the late 19th century and early 20th century in order to improve production parameters [3]. As a result, a phenotypically differentiated breed was described as "Chato Murciano". The depreciation of animal fat by the consumer in the second part of the 20th century, and the low growth rate and feed conversion efficiency, lead the CM breed to the edge of extinction with 20-30 breeders in 1997. However, conservation programs have raised the population number to 287 individuals in 2009.

Despite favourable appreciation of CM meat products by local customers [4], several concerns remain regarding the long-term survival of the breed. In the early years of this century, a conservation program based on the supply of semen and replacement gilts to the farms was very successful. However, after the year 2008 the program lost popularity, perhaps because of the increase of the population number of the CM pig. Nowadays, breeding stocks tend to be isolated by farm, and the farmers each tend to apply their own breeding strategies, which in some cases may even include crossing with commercial breeds. While crossbreeding may potentially enhance production traits, it simultaneously threatens the heritage status of the breed and with that the higher value of its meat. For the smallest farms in particular, there is a concern of high inbreeding levels and related risks of inbreeding depression and subsequent decrease in productivity of the pigs. Further loss of breeding stock may result from farmers being unable to cope with the economic consequences of the low productivity of the CM breed. All these factors

directly threaten the loss of this breed that represents over 150 years of cultural and agricultural history and that remains to have an important socio-economic influence in the Region of Murcia. However, in order to design an effective conservation program, a detailed genetic study of the CM breed diversity is a necessary first step.

Endangered domesticated breeds and populations generally face two major threats: inbreeding and crossbreeding. Because the CM breed is currently almost exclusively confined to a single region in Spain (i.e. Murcia), where it is being bred in small numbers on a small number of farms, together with the high propensity of being crossbred with other pig breeds, the CM is a very suitable model to study the effect of both threats on endangered domesticated populations. This study surveys eight farms (50% of all farms that currently are listed as official CM breeding farms) that breed at least 10 animals, representing 70% of the current breeding population, and include all major pig breeds that are thought to have contributed - historically and recently – to the genetic make-up of the CM breed. We apply three different marker systems comprehensively, to investigate 1) inbreeding and farm-based population stratification, and 2) genetic contributions from other pig populations, especially recent crossbreeding with commercial breeds. The results of this study are important to address a basis for rational exploitation of the CM breed in the future.

3.2 Materials and Methods

Animals and sampling

Genomic DNA was extracted from blood and hair using the Gentra Pure Gene Blood kit (Qiagen) and the Danapure Spin kit (Genedan SL, Spain) respectively according to the manufacturer‘s protocol. The study included 194 Chato Murciano (CM) pigs representing 70% of the current breeding population. All the breeders reared in each of the eight most important registered farms were sampled. In addition, 194 domestic pigs from the breeds Large White (LW, n = 53), Landrace (LR, n = 29), Duroc (DU, n = 42), Tamworth (TA, n = 15), Berkshire (BK, n = 19), Iberian pig (IP, n = 26) and Meishan (MS, n = 10) were included. The samples used for each molecular marker are detailed in the Table 3.1.

Table 3.1 Sampling information and analysis performed in each pig population.

Breed	Country	N	D-Loop*	60k*	Microsatellites*
Chato Murciano					
Farm1	Spain	44	32	8	36
Farm2	Spain	18	14	5	16
Farm3	Spain	22	18	5	21
Farm4	Spain	32	26	6	32
Farm5	Spain	9	7	3	7
Farm6	Spain	42	35	8	42
Farm7	Spain	16	10	2	11
Farm8	Spain	11	10	2	10
Total	Spain	194	152	39	175
Large White	Commercial	53	48	53	0
Landrace	Commercial	29	34	29	0
Duroc	Commercial	42	42	42	0
Tamworth	United Kingdom	15	14	15	0
Berkshire	United Kingdom	19	19	19	0
Iberian Retinto	Spain	11	11	11	0
Iberian Negro	Spain	15	15	15	0
Meishan	China	10	10	10	0

* Number of animals genotyped

Microsatellite genotyping

A total of 34 autosomal markers were analyzed, of which 24 belong to the panel recommended by the ISAG-FAO Advisory Group on Animal Genetic Diversity [5] and ten to the panel designed by the Roslin Institute, UK [6]. Microsatellites were chosen based on their absence of null alleles, sharpness of peaks and possibility of being grouped into six multiplex PCR. Each multiplex set PCR contained markers without overlapping of alleles of the same dye (Table 7, supplementary material). Amplified products were electrophoresed in an ABI 3130® (Applied Biosystems, USA) and allele calling was performed in Genemapper v.3.7 (Applied Biosystems, USA). The French Pig Map reference samples DNAs F9110010 and F9119912 (courtesy of INRA, http://www.toulouse.inra.fr) were used to normalize the allele sizes.

Mitochondrial sequencing

Part of the D-loop region of the mitochondrial DNA was amplified using the primers described by Luetkemeier et al. (2010) yielding a 772 bp fragment. The PCR amplicons were purified and sequenced for both strands on an ABI 3130® DNA sequencer (Applied Biosystems, USA). The Genbank accession numbers of the sequences are described in Table 8 (supplementary material).

SNP genotyping

High density SNP genotyping was performed using the PorcineSNP60 BeadChip (IlluminaInc, USA; Ramos et al. 2009) according to manufacturer's protocol. For this study, only SNPs mapped to one of the 18 autosomes on *Sus scrofa* build 10.2 were included in the analysisand SNP markers with more than 5% missing genotypes were excluded by using Plink software [9]. Finally, 46,887 SNPs of the 62,163 potential SNPs were used for the analysis.

Data analysis

The numbers of alleles per marker microsatellite and allele frequencies, as well as observed and expected heterozygosity were calculated with Genalex V.6.3 [10]. To analyze the genetic differentiation between farms, Wright's F-statistics [11] was used as defined by [12] and as implemented in the GENEPOP 4.0.10 software [13]. The matrix of genetic distances [14] between the farms was calculated with the software Genalex V.6.3 and used to construct a Neighbor-Joining (NJ) tree with Mega 5.03 [15].

The Structure software version 2.0 [16] was used to infer population stratification in CM breed based on microsatellite data. Structure uses a Bayesian clustering algorithm to identify clusters with distinctive allele frequencies. To detect the best value of K (number of assume clusters), the Bayesian Information Criterion (BIC) implemented by the package of R Adegenet[17] was used. All Structure runs used 10,000 iterations after a burn-in of length 10,000 MCMC replications. The Structure results were graphically displayed by using Distruct 1.1 [18].

Genome Assembly Program (GAP4) [19] was used to view and obtain the consensus sequence of D-loop region for each individual relative to pig mtDNA sequence GenBank ID AJ00218 as a reference. Sequences were subsequently aligned by Clustal X V.2 [20] and grouped into haplotypes using the program ALTER [21]. Phylogenetic relationships among the haplotypes were determined with Mega 5.03 using the NJ method based on the maximum composite likelihood.

The Plink software was used to compute the Identity by State (IBS) matrix for all pairs of individuals based on genome similarity derived from the 60K SNP data. The IBS matrix was used to perform multidimensional scaling analysis (MDS). The Structure software was used to examine relatedness among breeds and also

population stratification between farms in order to compare the results of Structure with the microsatellite markers. Pairwise F_{st} values and genetic distances between breeds were calculated using Powermarker [22].

3.3 Results

Genetic diversity and population substructure of Chato Murciano pig

All tested microsatellite markers were polymorphic for all the Chato Murciano (CM) pigs with an average number of detected alleles of 6.4 ranging from 2 (SW796) to 12 (S0005). The expected heterozygosity was 0.539 and the observed heterozygosity was 0.516 (Table 3.2). The heterozygosity analysis performed by farms showed that Farm 6 had the highest genetic diversity (He = 0.563) while Farm 8 had the lowest (He = 0.262). The occurrence of allele frequencies is summarized in Table 4 (supplementary material). There were at least two monomorphic loci in seven out the eight farms under study.

The average value of F_{st} across all loci was 0.114 indicating that 11% of genetic variation was explained by differences between farms. The pairwise genetic distances (DA) between farms are given in Table 5 (supplementary material) and graphically represented in the Figure 3.1. The phylogenetic tree had two differentiated clusters (Farms 2, 3, 4, 5 and 7, in green, and Farms 6 and 8, in red), while one farm (Farm 1) occupied an intermediate position.

To estimate the existence of different genetic clusters in the CM population, the admixture model implemented by Structure software was used for the microsatellites and SNP genotyping data separately (Figure 3.2). For K = 2 consistent results were obtained across microsatellites and SNPs data and also with the tree of genetic distances. Farms 2-5 and 7 (green), and Farms 6 and 8 (red) each appeared to represent different gene pools, while pigs from Farm 1 were assigned to each of these gene pools. However, for K = 3 a third cluster represented by Farm 1 (yellow) appeared with microsatellite data but not with SNPs data.

Table 3.2 Number of alleles (Na), Na mean (NMa), and observed (Ho) and expected (He) heterozygosity of microsatellite analysis in Chato Murciano breed as a whole population and in each farm separately.

	N*	N**	N_a	NM_a	H_o	H_e
Farm1	44	36	122	3.588	0.495	0.482
Farm2	18	16	114	3.353	0.576	0.465
Farm3	22	21	116	3.412	0.460	0.429
Farm4	32	32	138	4.059	0.486	0.460
Farm5	9	7	95	2.794	0.608	0.487
Farm6	42	42	161	4.735	0.585	0.563
Farm7	16	11	77	2.265	0.338	0.324
Farm8	11	10	40	1.176	0.318	0.262
CM	194	175	217	6.386	0.516	0.539

* Number of animals reared in each farm. ** Number of animals analysed.

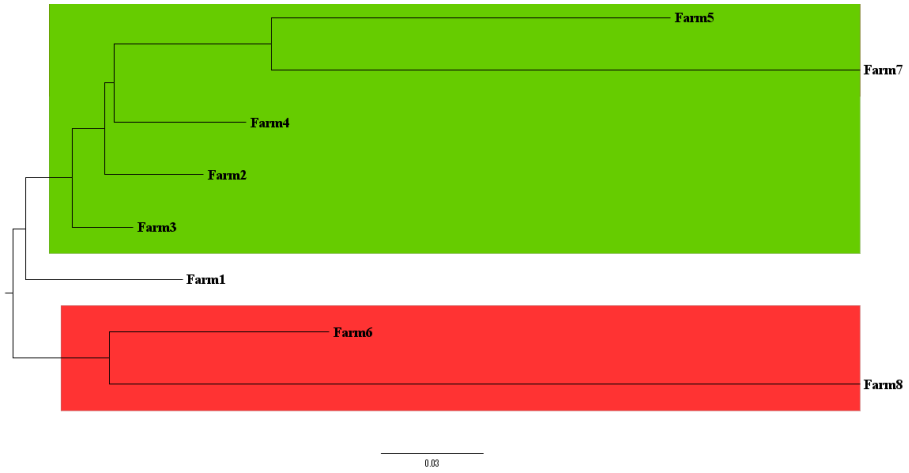

Figure 3.1 NJ tree of Chato Murciano farms constructed from Nei genetic distances (D_A).

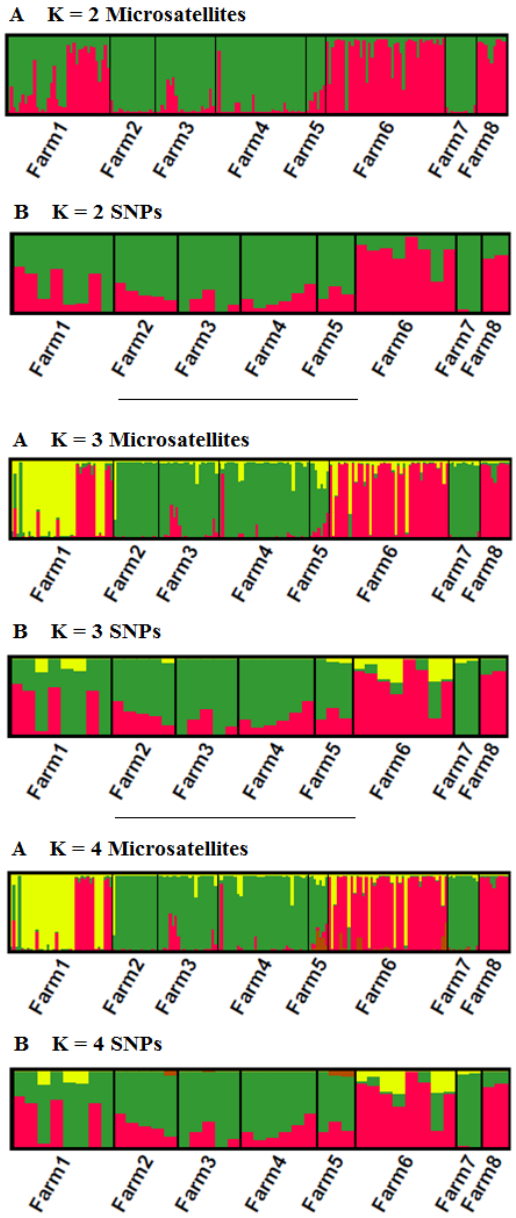

Figure 3.2 Estimated membership coefficients from K = 2 - 4. (A) microsatellites data; (B) SNPs data. Each color represents the proportion of the genome assigned to each assumed cluster.

Relationships of Chato Murciano breed with other domestic breeds

Based on the 60K SNP data three main clusters were identified by the MDS analysis (Figure 3.3). The first one was the Asian cluster represented by the Meishan breed, which was the most divergent population. The second cluster was in the opposite end of the first component relative to the Asian cluster, and was represented by the Duroc breed. Finally, a third cluster that included two commercial breeds, Large White and Landrace, the English breeds Tamworth and Berkshire, and the Spanish breeds Iberian and Chato Murciano (CM).

Since the Meishan breed and the Duroc breed accounted for a large part of the variation observed, these two breeds were removed subsequently to analyze the European cluster (Figure 3.3B). All the breeds formed discrete clusters with the exception of Retinto and Negro Iberico that are actually two varieties of Iberian Pig. CM occupied an intermediate position, in between the commercial breeds and Tamworth, Berkshire, and Iberian Pig, in the first component, but the second component explained the difference between CM and the other breeds. All the CM pigs were grouped in the same area of the plot, although seven individuals separated slightly from the main CM cluster. Four of those animals belonged to Farm 6 (Figure 8, supplementary material).

To infer whether CM is a distinct genetic population and to detect admixed individuals, an admixture model was tested without defining the population from which the individuals were obtained. K values from six to nine were tested using the 60K SNP data (Figure 5, supplementary material), with K = 8 determined to be the optimum value (Figure 7, supplementary material). Membership Coefficients in the eight clusters inferred by Structure are detailed in Table 3.3. Large White and Landrace were the breeds with lower membership coefficient (0.77 and 0.71 respectively) while Meishan showed the highest (1.00). For the rest of the breeds the average membership coefficient was over 0.93 with the exception of CM, for which it was 0.85. A detailed study of the individual membership coefficient in CM breed was made, which revealed that every CM pig clustered together with a membership coefficient over 0.75 except for four pigs that had a membership coefficient of around 0.5.

The results from Nei´s genetic distances and the pairwise F_{st} between breeds are detailed in Table 6 (supplementary material) and Nei´s genetic distances graphically represented in Figure 3.4. Meishan was the most divergent population while Negro Iberico and Retinto were the least divergent groups. CM occupied an intermediate position between Berkshire and Large White - Landrace populations.

Figure 6 (supplementary material) shows a NJ phylogenetic tree based on genetic distances between mtDNA sequences. The tree showed two main clades: the first

included exclusively European and commercial breeds and the second represented an Asian mitochondrial clade, and included all pigs of the Chinese Meishan breed, and some commercial and English pigs. Asian mtDNA haplotypes were highly frequent in British pigs, with 100% of Berkshire pigs and in 93% of Tamworth pigs having Asian haplotypes. The only two European breeds that exclusively carried European haplotypes were Iberian Pig and CM. Three haplotypes common in CM (HP2-4, together 77% occurrence) were shared with commercial breeds. One haplotype was shared with Iberian Pig (HP1). All CM pigs carrying HP1 belonged to Farm 7 which rears both Iberian Pig and CM pigs. One haplotype was unique to CM (HP5) and occurred in 14% of CM pigs.

Table 3.3 Membership Coefficient of the breeds tested in the eight clusters inferred by Structure software.

		Inferred clusters								
		1	2	3	4	5	6	7	8	N
Predefined population	LW	0.011	0.025	0.017	0.036	0.768	0.030	0.045	0.069	53
	LR	0.005	0.036	0.002	0.710	0.087	0.025	0.048	0.087	29
	DU	0.957	0.005	0.001	0.001	0.000	0.003	0.001	0.033	42
	TA	0.001	0.000	0.000	0.000	0.000	0.998	0.000	0.001	15
	BK	0.003	0.950	0.001	0.007	0.005	0.006	0.008	0.020	19
	IP_RE	0.004	0.008	0.000	0.002	0.004	0.010	0.002	0.969	11
	IP_NI	0.063	0.002	0.000	0.000	0.004	0.001	0.001	0.929	15
	MS	0.000	0.000	1.000	0.000	0.000	0.000	0.000	0.000	10
	CM	0.064	0.011	0.001	0.003	0.019	0.007	0.854	0.041	39

* Large White (LW); Landrace (LR); Duroc (DU); Berkshire (BK); Tamworth (TA); Iberian Pig Retinto (IP_RE) and Negro Ibérico IP_NI; Meishan, (MS); Chato Murciano (CM).

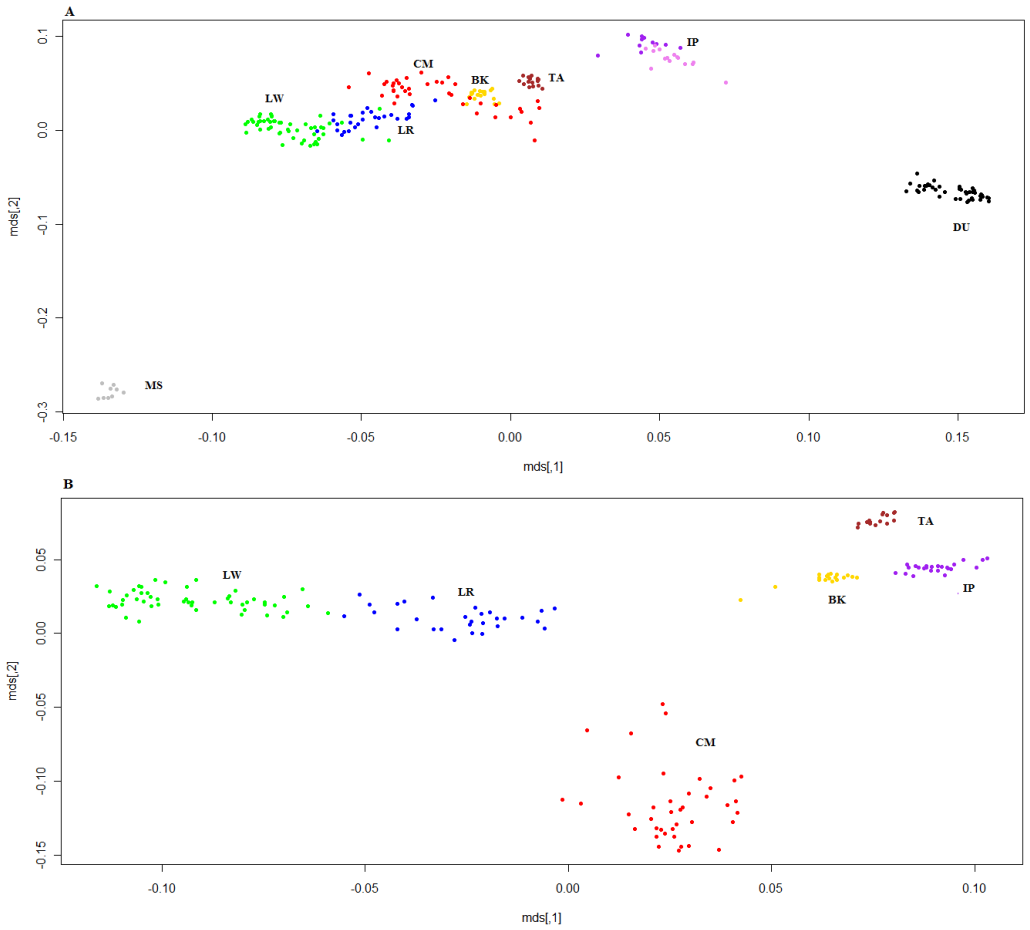

Figure 3.1 MDS (A) All breeds; (B) European breeds.

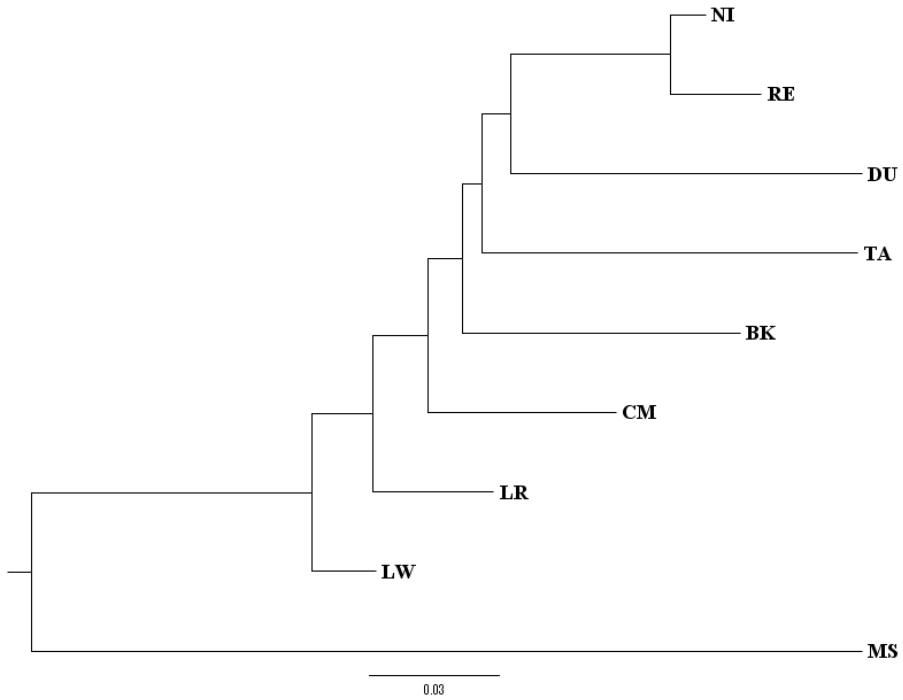

Figure 3.4 NJ tree of the breeds tested constructed from Nei genetic distances based on SNPs data.

3.4 Discussion

In the conservation genetics of specific populations or breeds of domestic animals, there are always two major threats to consider: inbreeding and crossbreeding. The former may erode genetic diversity and threaten long-term survival because of reduced productivity or inbreeding depression [23]. The latter may threaten the historical significance of the population or breed by loosing what is perceived to be the 'genetic integrity'. Although difficult to quantify, this may result in decreased value of the breed for being a unique reservoir for genetic or phenotypic variation. In addition, crossbreeding may reduce the perceived value of a breed for being a representative for cultural or culinary heritage.

The Chato Murciano (CM) breed is vulnerable to both threats. CM pigs are currently being bred and reared on approximately 15 farms, and only eight of them are breeding at least 10 breeders. The limited number of farms in the Murcia

region that breed this pig together with their geographical proximity, also provides the unique opportunity for highly, and so unprecedented, detailed dissection of how both threats influence the genetic make-up of an endangered domesticated population.

The current study found levels of genetic diversity in the CM breed to be within the range of other European local breeds such as Iberian Pig [24], Magalitsa [25] and Portuguese breeds [26] in terms of expected heterozygosity and mean number of alleles per microsatellite locus. A farm-by-farm analysis, however, revealed marked differences between farms in the genetic diversity, with particularly Farms 7 and 8 displaying the characteristics of isolated small farms with high inbreeding and genetic drift. The CM breed shows a high degree of substructure that is correlated with the farms they are reared on. Specifically, there appear to be two distinct gene pools. The first one includes Farms 2, 3, 4, 5 and 7, and within these farms, Farms 2, 3 and 4 displaying an even closer relationship(DA < 0.07). These three farms belong to the same farmer and occupy the same geographic area (Google Earth projection, supplementary material). Therefore, it is likely that an increased rate of swapping of boars between farms enhances gene flow between them, explaining the high similarity. Farms 6 and 8 represented the second gene pool, while Farm 1 seemed to represent a mix of animals from both clusters instead of forming a distinct third group, since Farm 1 had animals assigned to both clusters inferred by Structure using SNPs data and occupied an intermediate position in the DA tree (Figure 3.1)

These results demonstrate that all farmers appear to have their own breeding strategies. It is clear that breeding strategy is not limited to effective population size on farms and exchange of breeding stock of the CM pigs between farms, however Farms 6 and 7 showed signs of genetic introgression with other breeds. In Farm 7, 60% of the pigs presented a mtDNA haplotype found only in Iberian Pigs (IP, haplotype HP1). Considering that Farm 7 reared both CM and IP, and high level of inbreeding (Table 3.2), it is likely that the farmer preferred to cross animals of those breeds to improve production parameters sometime in the past, rather than exchange CM breeding stocks with other farms. Similarly, for Farm 6, several indications of genetic introgression were found. Firstly, Farm 6 showed the lowest allele frequency in two loci microsatellite previously described as exclusive of CM pure animals [27]. Specifically, allele 123 of loci SW951had a frequency of 0.54 in Farm 6 but it was close to 1.00 in the other farms. Secondly, animals belonging to Farm 6 were clearly separated from the CM cluster when the Identity by State (IBS) matrix was computed. Finally, four pigs from Farm 6 were assigned to CM breed with membership coefficients lower of ~0.5 indicating a high level of genetic

admixture. It seems, therefore, that breeding practices of Farms 6 and 7 include, or included, crossbreeding strategies. This is noteworthy since breeders with high level of admixture maybe excluded from a future conservation program in order to preserve the genetic identity of the CM breed.

All the pig breeds, including CM, were unambiguously identified by the stratification models used, showing that high density SNP genotyping data provided a powerful tool for assessing genetic differentiation between populations and breed assignment [28]. Meishan was the most divergent population whilst Large White and Landrace were less differentiated as was previously observed e.g. [29]. Structure analysis showed a lower membership coefficient in the CM breed (0.85) than in Berkshire, Tamworth, Duroc and Iberian Pig breeds (>0.90), but this was mainly due to the existence of a limited number of admixed animals within the CM pigs tested.

We found evidence of genetic contributions from other pig populations into CM breed. Specifically, the breeds Large White, Berkshire and Iberian Pig appear to have had large influence on the CM gene pool (Table 3.3). All three breeds showed lower genetic distances to CM than Tamworth for instance, which is in agreement with the recorded history of the CM breed. Interestingly, the Duroc breed seemingly had a large contribution to the CM gene pool (6%), although no historical records exist that support such crossbreeding. Crossbreeding with Duroc appears to have been recent and localized, as it is mostly evident in CM pigs from Farm 6.

The analysis of mtDNA has been an important tool to identify genetic introgression of Asian pigs into European pigs . Occurrence of Asian haplotypes can be particularly high in commercial and English breeds [30, 31], as confirmed by this study. However, despite the very high occurrence of Asian haplotypes in English breeds such as Berkshire, which are thought to have been used in the past for crossbreeding with the CM pigs, no Asian haplotypes were found in this Spanish breed. Because of recorded breed history, the presence of Asian haplotypes in CM might be expected since English pig breeds could have carried Asian haplotypes into the CM breed as happened with other local breeds such as Jabugo Spotted [32]. The absence of Asian haplotypes is however consistent with historical records that report that only males were imported to Murcia Region [3]. Therefore, Asian haplotypes may have never been introduced in the CM population, or, alternatively, may have disappeared during the bottleneck suffered by the breed in the second part of the 20th century. Most of the CM (55%) carried haplotypes that are common in other European pigs. Although this may reflect past and present day crossbreeding with commercial pig breeds, these common haplotypes may have had a wide distribution area even at the time the first CM pigs were

historically described. It is noticeable that one haplotype (HP5), which was present in 14% of the CM pigs analyzed, was not detected in any other breed. It is possible that HP5 had been inherited from the primitive Murcia pig, or, alternatively, a more recent mutation may have gained prominence during recent bottlenecks because of drift. Unfortunately, no biological material (e.g. mounted skeletons or stuffed animals) from the primitive Murcia pig remain, which makes a direct comparison to the present work impossible.

Despite historical records and genetic observations that support past and present introgression, the CM breed remains genetically distinct from other breeds. This distinctness offers the opportunity for designing genetic marker sets that could unambiguously trace products sold to be from CM origin back to the origin. This distinctness therefore will greatly facilitate creating tools to enforce product identity certification, in this way stimulating local farmers, but also local butchers and the local government, to maintain their breeding stock sustainably.

In conclusion, the joint analysis of microsatellite, SNPs and mtDNA has showed a farm-based population stratification due to both, the existence of small genetically isolated farms with concerning levels of inbreeding and independent breeding strategies developed by the farmers. Moreover, we were able to detect signs of past and recent introgression from other breeds, and even locate in which farms- it had occurred. Therefore, identifying farm-based practices and farm-based breeding stocks seems to be an accurate approach to set up sustainable breeding programs for minority livestock breeds when pedigree information is absent.

Acknowledgements

The research related to the Genetic Characterization of Chato Murciano pig has been granted by the Servicio de Formación y Transferencia Tecnológica de la Dirección General de Modernización de Explotaciones y Transferencia Tecnológica de la Consejería de Agricultura y Agua de la Comunidad Autónoma de la Región de Murcia through the Research Contract 12603.

Supporting Information

Chapter_3.zip

https://mega.co.nz/#!WFFA1LqS!Ykc-SitdeSpk19P82dUqfSiP7HxBxgpYMGSOsres1CY

References

1. FAO: **Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration.** *(available at http://www.fao.org/docrep/010/a1404e/a1404e00.htm).* 2007.
2. FAO: **Building on gender agrobiodiversity and local knowledge**. *Food and Agriculture Organization of the United Nations* 2006, **1**:1–5.
3. Lobera J: *El cerdo Chato Murciano: Orígenes e Historia*. Murcia (Spain).: Consejería de Medio Ambiente, Agricultura y Agua.; 1998.
4. Peinado B, Poto A, Gilb F, López G: **Characteristics of the carcass and meat of the ChatoMurciano pig.***Livestock Production Science* 2004, **90**:285–92.
5. Hoffmann I, AjmoneMarsan P, Barker JSF, Cothran, E.G. Hanotte O, Lenstra JA, Milan D, Weigend S, Simianer H: **New MoDAD marker sets to be used in diversity studies for the major farm animal species: recommendations of a joint ISAG/FAO working group.** In *Proceedings of 29th International Conference on Animal Genetics*. Tokyo, Japan: 2004:123.
6. Russell GA, Archibald AL, Haley CS, Law AS: **The pig genetic diversity database and the WWW.***Archivos de Zootecnia* 2003, **52**:165–72.
7. Luetkemeier ES, Sodhi M, Schook LB, Malhi RS: **Multiple Asian pig origins revealed through genomic analyses.***Molecular phylogenetics and evolution* 2010, **54**:680–6.
8. Ramos AM, Crooijmans RPM a, Affara N a, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu Z-L, Kerstens HH, Law AS, Megens H-J, Milan D, Nonneman DJ, Rohrer G a, Rothschild MF, Smith TPL, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen M: **Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology.***PloS one* 2009, **4**:e6524.
9. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M a R, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.***American journal of human genetics* 2007, **81**:559–75.
10. Peakall R, Smouse PE: **GenAlEx 6.5: genetic analysis in Excel. Population genetic software for teaching and research-an update.***Bioinformatics (Oxford, England)* 2012, **28**:2537–9.
11. Wright: **Evolution and the Genetics of Populations**. In *The Theory of Gene Frequencies*. University. Chicago: 1969.

12. Weir BS, Cockerham C. C: **Estimating F-Statistics for the Analysis of Population Structure**. *Evolution* 2008, **38**:1358–1370.
13. Rousset F: **genepop'007: a complete re-implementation of the genepop software for Windows and Linux.***Molecular ecology resources* 2008, **8**:103–6.
14. Nei M: **Genetic Distance between Populations**. *The American Naturalist* 1972, **106**:283–292.
15. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.***Molecular biology and evolution* 2007, **24**:1596–9.
16. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.***Genetics* 2000, **155**:945–59.
17. Jombart T: **adegenet: a R package for the multivariate analysis of genetic markers.***Bioinformatics (Oxford, England)* 2008, **24**:1403–5.
18. Rosenberg NA, Ave W, Mi AA: **Distruct : a program for the graphical display of population structure**. *Molecular Ecology Notes* 2004, **4**:137–8.
19. Bonfield JK, Smith K f, Staden R: **A new DNA sequence assembly program.***Nucleic acids research* 1995, **23**:4992–9.
20. Larkin MA, Blackshields G, Brown NP, Chenna R, McGettigan PA, McWilliam H, Valentin F, Wallace IM, Wilm A, Lopez R, Thompson JD, Gibson TJ, Higgins DG: **Clustal W and Clustal X version 2.0.***Bioinformatics (Oxford, England)* 2007, **23**:2947–8.
21. Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D: **ALTER: program-oriented conversion of DNA and protein alignments.***Nucleic acids research* 2010, **38**:W14–8.
22. Liu K, Muse S V: **PowerMarker: an integrated analysis environment for genetic marker analysis.***Bioinformatics* 2005, **21**:2128–9.
23. Santana ML, Oliveira PS, Eler JP, Gutiérrez JP, Ferraz JBS: **Pedigree analysis and inbreeding depression on growth traits in Brazilian Marchigiana and Bonsmara breeds.***Journal of animal science* 2012, **90**:99–108.
24. Alves E, Fernández AI, Barragán C, Ovilo C, Rodríguez C, Silió L: **Inference of hidden population substructure of the Iberian pig breed using multilocus microsatellite data**. *Spanish Journal of Agricultural Research* 2006, **4**:37–46.
25. Manea MA, Georgescu SE, Kevorkian S, Costache M: **Genetic diversity analyses of seven Romanian pig populations based on 10 microsatellites**. *Romanian Biotechnological Letters* 2009, **14**:4827–4834.
26. Vicente a a, Carolino MI, Sousa MCO, Ginja C, Silva FS, Martinez a M, Vega-Pla JL, Carolino N, Gama LT: **Genetic diversity in native and commercial breeds of**

pigs in Portugal assessed by microsatellites.*Journal of animal science* 2008, **86**:2496–507.

27. Vega-Pla JL, Martínez AM, Peinado B, Poto A, Delgado JV: **The use of molecular techniques in the support to the conservation of the Chato Murciano breed of pig**. *Archivos Latinoamericanos de Produccion Animal* 2004, **12**:45–48.
28. Ramos AM, Megens HJ, Crooijmans RPMA, Schook LB, Groenen MAM: **Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing.***Animal genetics* 2011, **42**:613–20.
29. Megens HH-J, Crooijmans RPM a., San Cristobal M, Hui X, Li N, Groenen MAM, Magali SC, Xiao H, Ning L, Martien AMG: **Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication.***Genetics, selection, evolution : GSE* 2008, **40**:103–28.
30. Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L: **The origin of the domestic pig: independent domestication and subsequent introgression.***Genetics* 2000, **154**:1785–91.
31. Kim KI, Lee JH, Li K, Zhang YP, Lee SS, Gongora J, Moran C: **Phylogenetic relationships of Asian and European pig breeds determined by mitochondrial DNA D-loop sequence polymorphism.***Animal genetics* 2002, **33**:19–25.
32. Alves E, Ovilo C, Rodríguez MC, Silió L: **Mitochondrial DNA sequence variation and phylogenetic relationships among Iberian pigs and other domestic and wild pig populations.***Animal genetics* 2003, **34**:319–24.

4

Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula

Herrero-Medrano JM[1,2] , Megens HJ[2], Groenen MAM[2], Ramis G[1], Bosse M[2], Pérez-Enciso M[3,4], Crooijmans RPMA[2].

[1] Departamento de Producción Animal. Fac. de Veterinaria, Área de Genética. C.P. 30100, Murcia. Spain; [2] Animal Breeding and Genomics Centre. Wageningen University. De Elst 1,6708 WD. Wageningen, The Netherlands; [3] Centre for Research in Agricultural Genomics, Consortium CSIC-IRTA-UAB-UB. Edifici CRAG, Campus Universitat Autonoma Barcelona, 08193 Bellaterra, Spain; [4] Institut Català de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain.

BMC Genetics (2013) Submitted

Abstract

Background

Inbreeding is among the major concerns in management of local livestock populations. The effective population size of these populations tends to be small, which enhances the risk of fitness reduction and extinction. High-density SNP data make it possible to undertake novel approaches in conservation genetics of endangered breeds and wild populations. A total of 97 representative samples of domestic and wild pig populations from the Iberian Peninsula, subjected to different levels of threat with extinction, were genotyped with a 60K SNP panel. Data analyses based on: (i) allele frequency differences; (ii) linkage disequilibrium and (iii) regions of homozygosity were integrated to study population relationships, inbreeding and demographic history.

Results

The domestic pigs analyzed belonged to local Spanish and Portuguese breeds: Iberian – including the variants Retinto Iberian, Negro Iberian and Manchado de Jabugo –, Bisaro and Chato Murciano. The population structure and persistence of phase analysis demonstrated high genetic relations between Iberian variants, with recent crossbreeding of Manchado de Jabugo with other pig populations. Chato Murciano showed a high frequency of long runs of homozygosity indicating recent inbreeding and reflecting the recent bottleneck reported by historical records. The Chato Murciano and the Manchado de Jabugo breeds presented the lowest effective population sizes in accordance with their status of highly inbred breeds. The Iberian wild boar presented a high frequency of short runs of homozygosity indicating past small population size but no signs of recent inbreeding. The Iberian breed showed higher genetic similarities with Iberian wild boar than the other domestic breeds.

Conclusions

High-density SNP data provided a consistent overview of population structure, demographic history and inbreeding of minority breeds and wild pig populations from the Iberian Peninsula. Despite the very different background of the populations used, we found a good agreement between the different analyses. Our results are also in agreement with historical reports and provide insight in the events that shaped the current genetic variation of pig populations from the Iberian Peninsula. The results exposed will aid to design and implement strategies for the future management of endangered minority pig breeds and wild populations.

Key words: Local breeds, population genetics, SNP, genetic diversity, effective population size, pig, Iberian Peninsula

4.1 Background

Progressive population decline has called the attention of the conservation management and scientific communities. Both for wild and domesticated populations alike there is a fear that inbreeding may lead to loss of allelic variation and adverse phenotypic consequences [1]. In addition, loss of variation may lead to reduced response to changing environments, of which genetic susceptibility to novel infectious diseases is a specific concern. Agricultural diversity in particular is of concern for future food safety [2]. Variation conserved in local breeds is often related to important traits that classically are attributed to traditional populations, such as adaptation to the environment and greater resistance to local pathogens.

In addition to these concerns, local populations are often considered to be part of the local culture and history. For instance, local pigs are often linked to local cuisine and the local landscape. The Spanish Iberico or Iberian, and Portuguese Alentejano pigs, for instance, are used to produce highly priced products due to their quality that in part results from feeding with acorns from sparse Mediterranean oak forests, the so called 'Dehesas'. The wild relative of the pig, the wild boar, on the other hand, plays a significant role in the wildlife of the Iberian Peninsula. It is among the main prey species of an iconic predator from the Iberian Peninsula such as Iberian wolf [3]. Moreover, wild boar is an important reservoir of infectious diseases as relevant as tuberculosis in the Iberian Peninsula [4], and therefore also of concern for public and animal health.

Local pig populations, both wild and domestic, have been highly affected by human-induced changes. Local, usually fat, breeds for instance, were affected by changes in consumer preference in the middle of the 20th century when consumers started to avoid high-fat meat. As a result, a relatively small number of highly productive pig breeds progressively replaced and marginalized the traditional breeds. Many breeds became extinct in the past decades, while many other traditional breeds today face near-extinction either through dwindling population numbers or hybridization with highly productive breeds [5]. At the same time, the increase in woodland across Europe has allowed wild boar populations to increase in many countries, after having been marginalized for centuries [6]. The Iberian Peninsula provides a good representation of local pig populations, both wild and domestic. While sharing the same geography, these populations have undergone different historical events, have different phenotypic attributes, and have a different conservation status. The Iberian pigs have been reared in an extensive traditionally system in the South and West of the Iberian Peninsula for centuries, remaining isolated from the modern breeding practices developed in the late 18th

and 19th century in NW Europe [7]. Iberian pigs are related to other Mediterranean pigs of Italy and The Balkans [5], which are thought to have a smaller influence from Asian pigs than the NW European pigs. Conversely, Chato Murciano and Manchado de Jabugo, now both highly endangered populations, and Bisaro resulted from crosses between native pigs from the Iberian Peninsula and foreign pigs at the end of the 19th century [8]. Beside these domestic populations, the Iberian Peninsula is also inhabited by wild boars that may represent the ancestor of these local breeds, and also constitute an important wildlife species of the Iberian Peninsula.

The recent availability of a high-density porcine SNP panel [9] provides an essential tool for genome wide association studies and genomic selection for economically important traits [10, 11]. Besides the use of high-density SNP arrays for economic purposes, these panels have demonstrated their power to assess major questions in conservation genetics [12, 13]. The study of linkage disequilibrium (LD) and genetic distances enable the estimation of effective population size from genetic data [14], which is of major interest in conservation genetics, especially when pedigree information is unavailable as is frequently the case for minority breeds and wild populations. In addition, high-density SNP arrays allow assessing similarities in the patterns of LD across populations (i.e. persistence of phase), providing information about the relatedness of populations [15]. The occurrence of runs of homozygosity (ROH) is indicative of demographic history and recent inbreeding [13, 16]. While the same parameters can be interpreted as signatures of selection on genomic regions [17, 18], when taken as global genomic parameters, they are highly indicative of demographic history [19], if properly corrected for local recombination rate [13]. A genome-wide SNP assay, combined with a detailed recombination map for the species [20] can therefore aid in giving insight into the conservation management of pig populations. Despite the fact that SNP assays are gaining interest for traceability purposes [21, 22], only few studies have used a high-density SNP assays for conservation purposes [1, 12, 23–25].

Here we present a comprehensive study in which high-density SNP data from domestic and wild pigs were used to address questions important to conservation genetics. First, we assessed the relationships between pigs by population structure analysis and by investigating the persistence of LD phase. Secondly, patterns of LD in each population were used together with a high-density recombination map to estimate past and present effective population size. Finally, the number and size of ROH were investigated in each individual. The joint analysis of all those parameters allowed us to obtain reliable and consistent data of population structure,

inbreeding and demographic history in each population providing valuable insights for future management strategies in pig from the Iberian Peninsula.

4.2 Results

A total of 97 pigs from domestic and wild autochthonous populations from the Iberian Peninsula were genotyped with the Porcine SNP60 Beadchip [9]. The SNPs located on the sex chromosomes and those with more than 5% missing genotypes were excluded from the analysis, resulting in a total of 47,594 SNPs used for the analysis.

Population structure

The Principal Component Analysis (PCA) revealed four main clusters represented by wild boar, Iberian, Bisaro and Chato Murciano (Figure 4.1). Among populations, Chato Murciano was the most divergent breed, showing a pairwise $F_{st} \geq 0.22$ with all populations except for Bisaro where it was 0.18. The variants of Iberian –Retinto, Negro Iberian, Manchado de Jabugo and five unclassified Iberian pigs– showed low Nei's genetic distances between them (≤ 0.06) and low F_{st} (≤ 0.055), and likewise the two populations of wild boar from Spain and Portugal (0.04 and 0.056 respectively). Among domestic pigs, Iberian variants showed the lowest Nei´s genetic distance to wild boar (0.10 - 0.12). The results from Nei's genetic distances and the pairwise Fst between populations are detailed in Additional file 1.

The Bayesian clustering algorithms implemented in the Structure software assigned all individuals to their expected clusters (K), with the exception of one Manchado de Jabugo pig (MJ_02) that was placed in the cluster of the other Iberian variants. Pairwise genetic distances between individuals (Additional file 5) and PCA analysis done using the Iberian pig data only (Additional file 2) confirmed this finding. K values from 2 to 7 were tested (Figure 4.2). The optimal K-value was estimated using the method described by Evanno et al. [26], indicating that K = 4 was the most parsimonious number of clusters (Additional file 3) in full agreement with the PCA analysis. Chato Murciano and Bisaro appeared as differentiated clusters when K=2-3, while Iberian and wild pigs shared the same cluster for those K values. The Iberian cluster (yellow) contributed to all the other populations, in particular to Bisaro (25%) and wild pigs from Spain (12%) (Additional file 3). No differentiation between wild boar from Portugal and Spain was apparent, nor between the Iberian, Retinto and Negro Iberian variants, for any of the K values tested. However, signs of admixture from an unspecified origin were observed in Manchado de Jabugo for K ≥ 5. Thus, for subsequent analysis, samples were grouped as follows: Bisaro, Chato Murciano, Manchado de Jabugo, Iberian pig and wild boar. In addition, the Iberian variants Retinto and Negro Iberico were also

analyzed separately. Finally, for population-based analyses the pig MJ_02 was removed since it fell outside of any of the groups considered in the analysis.

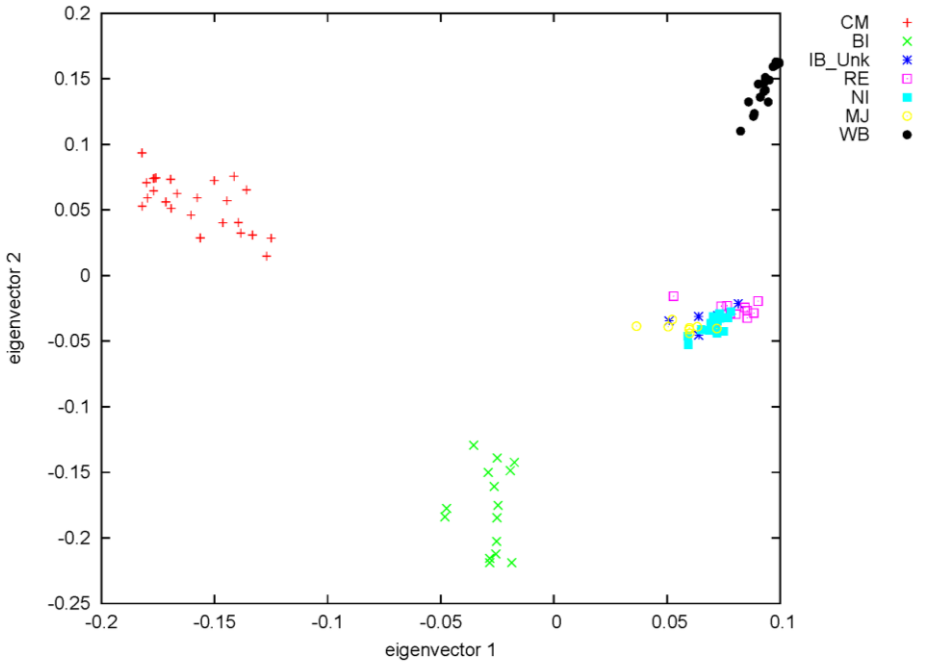

Figure 4.1 Different population groups defined with PCA analysis. WB, wild boar; IB_Unk, Iberian unidentified variant; NI, Negro Iberian; RE, Retinto; MJ, Manchado de Jabugo; BI, Bisaro; CM, Chato Murciano.

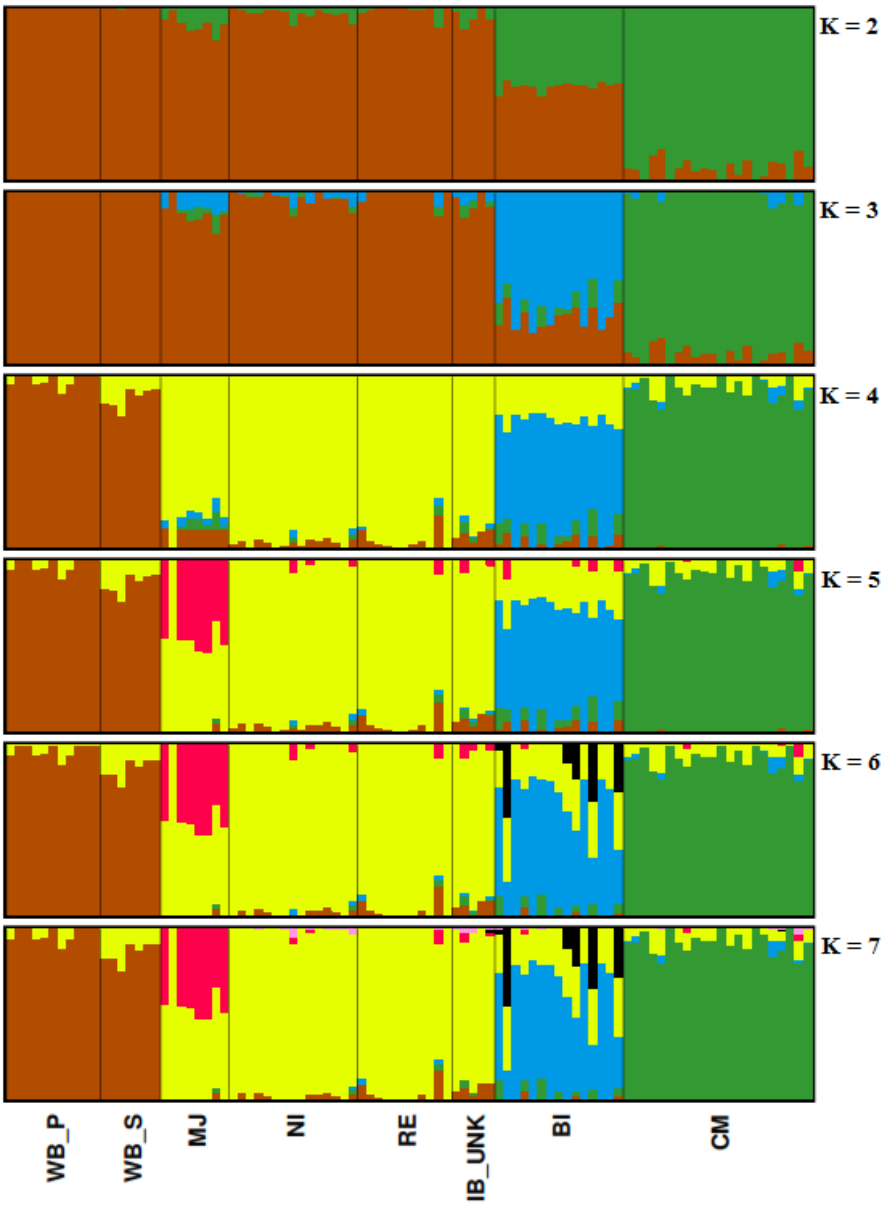

Figure 4.2 Graphic representation of estimated membership coefficients for each individual for K = 5. Each color represents the proportion of the genome assigned to each assumed cluster.WB, wild boar; IB_Unk, Iberian unidentified variant; NI, Negro Iberian; RE, Retinto; MJ, Manchado de Jabugo; BI, Bisaro; CM, Chato Murciano.

Linkage disequilibrium among populations

Markers that deviated from Hardy-Weinberg equilibrium ($P < 0.001$) or had a minor allele frequency (MAF) below 0.05 were excluded prior to LD analysis. Subsequently, 24,703 SNPs in wild boar, 29,856 SNPs in Bisaro, 33,454 in Chato Murciano, 27,858 in Iberian and 26,246 in Manchado de Jabugo were used to estimate LD for all SNP pairs less than 3 Mbp apart. Pairwise r^2 values were averaged over all 18 autosomes and plotted as a function of increasing genetic distance in all populations studied (Figure 4.3). The persistence of LD as the distance between loci increased and the strength of LD, varied widely between populations and between chromosomes (Table 4.1). The decay of LD as a function of marker distance was greater in wild boar ($r^2 < 0.2$ within 0.1 Mbp) than in the domestic breeds and showed the lowest average r^2 across all genetic distances. Among the domestic breeds, LD was the lowest in Iberian ($r^2 < 0.2$ within 0.2 Mbp). By contrast, Manchado de Jabugo and Chato Murciano had the most pronounced extent of LD at short genetic distances, although LD decreased faster in Chato Murciano than in Manchado de Jabugo for genetic distances higher than 1 Mbp.

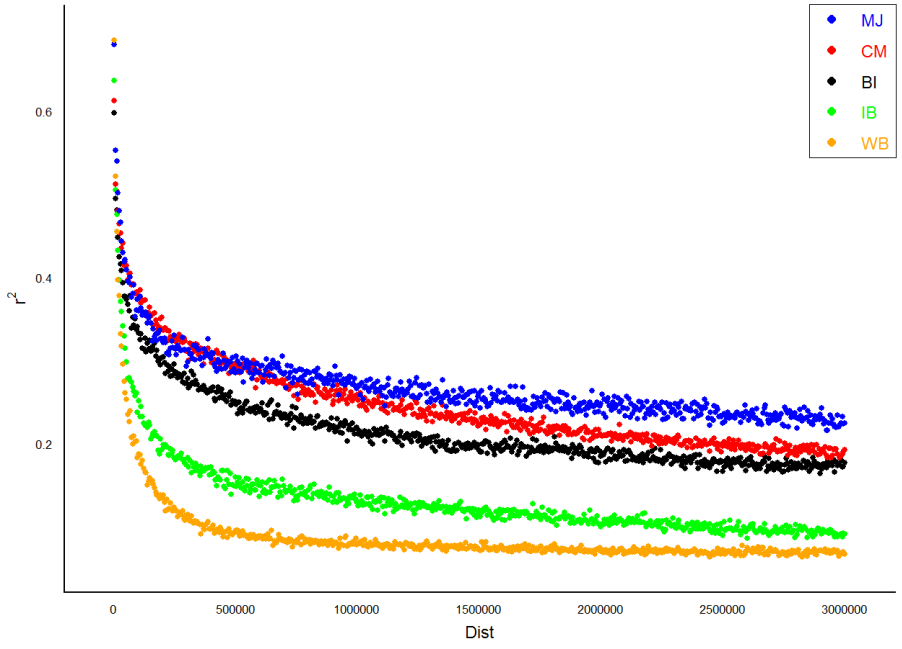

Figure 4.3 Average LD measure as r^2 across all chromosomes. WB, wild boar; IB, Iberian; MJ, Manchado de Jabugo; BI, Bisaro; CM, Chato Murciano.

Table 4.1 Linkage disequilibrium (r^2) and recombination rate averaged per chromosome and per population.

Chrom	Rec. Rate (cM/Mb)*	r^2 ± SD				
		CM	BI	IB	MJ	WB
1	0.36	0.29±0.27	0.25±0.24	0.13±0.15	0.26±0.29	0.11±0.15
2	0.64	0.22±0.24	0.24±0.22	0.16±0.17	0.3±0.31	0.1±0.13
3	0.71	0.25±0.25	0.17±0.19	0.12±0.13	0.26±0.29	0.09±0.13
4	0.67	0.27±0.27	0.18±0.2	0.12±0.12	0.24±0.26	0.11±0.13
5	0.86	0.25±0.25	0.2±0.21	0.11±0.13	0.25±0.28	0.09±0.12
6	0.80	0.26±0.23	0.22±0.23	0.15±0.17	0.27±0.29	0.1±0.14
7	0.85	0.24±0.24	0.21±0.22	0.12±0.12	0.3±0.31	0.09±0.12
8	0.65	0.28±0.26	0.24±0.23	0.15±0.17	0.24±0.27	0.11±0.14
9	0.73	0.24±0.25	0.24±0.23	0.12±0.13	0.3±0.3	0.09±0.12
10	1.14	0.24±0.25	0.18±0.2	0.09±0.11	0.23±0.27	0.08±0.1
11	0.75	0.23±0.23	0.19±0.2	0.16±0.18	0.26±0.26	0.1±0.12
12	1.24	0.26±0.24	0.2±0.2	0.13±0.13	0.24±0.27	0.09±0.12
13	0.46	0.3±0.26	0.27±0.26	0.17±0.19	0.36±0.29	0.11±0.15
14	0.73	0.3±0.26	0.24±0.24	0.17±0.18	0.3±0.28	0.09±0.12
15	0.61	0.3±0.26	0.21±0.22	0.14±0.15	0.26±0.27	0.11±0.14
16	0.78	0.31±0.26	0.21±0.21	0.13±0.15	0.27±0.29	0.1±0.13
17	0.95	0.27±0.26	0.22±0.23	0.11±0.13	0.24±0.29	0.09±0.11
18	0.81	0.21±0.22	0.2±0.2	0.09±0.12	0.23±0.28	0.09±0.12
Total	**0.76**	**0.26±0.25**	**0.22±0.22**	**0.13±0.14**	**0.27±0.28**	**0.10±0.13**

*Averaged recombination rate among 4 pig populations[20].

Persistence of phase

The persistence of LD phase was calculated as the Pearson correlation (r) between SNP pairs in all possible population pairs. Similar to LD, r decreased as the distance between markers increased (Figure 4.4). This was observed for all pairs of populations, although at different degrees (Figure 4, Additional file 4). Bisaro and Chato Murciano showed the greatest correlation of phase at short genetic distance. However, for SNP pairs spaced more than 1.5 Mb apart, Iberian and Manchado de Jabugo showed the highest correlation of phase. Correlations between the other pairs of domestic pig populations (CM-MJ, CM-IB, BI-IB, BI-MJ) tended to be similar.

The persistence of phase found between wild boar and all domestic pigs was lower than between domestic populations.

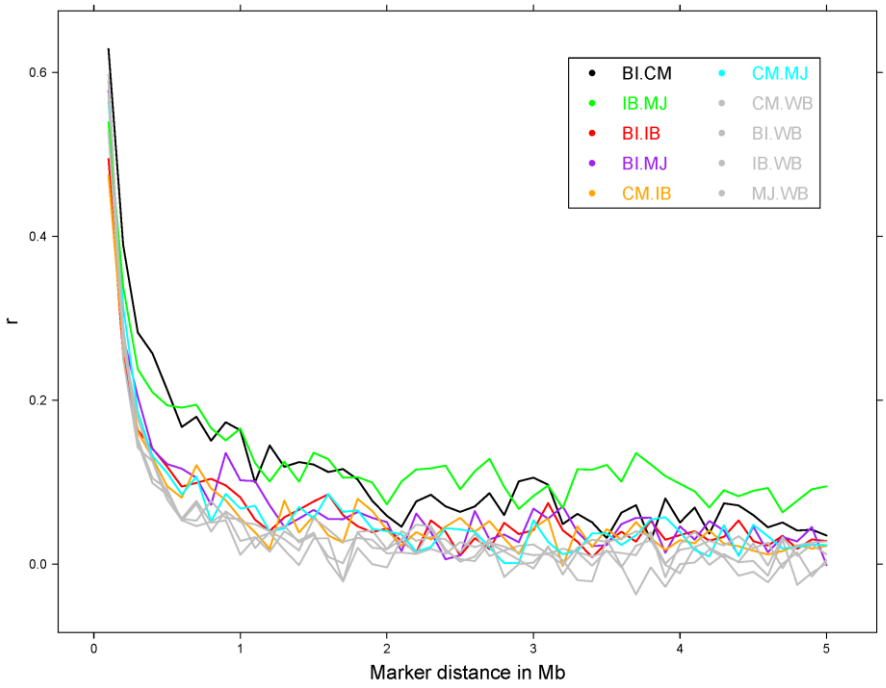

Figure 4.4 Correlation of Phase between populations for SNP pairs grouped by distance across the whole genome. The pairs between wild boar and domestic pigs (WB – IB/MJ/CM/BI) were uniformly plotted in gray for ease of reading. WB, wild boar; IB, Iberian; MJ, Manchado de Jabugo; BI, Bisaro; CM, Chato Murciano.

Current effective population size

The mean values of LD for all 1 Mb bins across the entire genome were used to estimate the current effective population size (N_e) implementing the equation $r^2 = 1/(4N_ec + 1)$ [27]. This estimation was performed taking into account the recombination for each of these bins [20]. Large N_e was observed for the wild boar population (Table 4.2). Among the domestic breeds, Iberian also had a high N_e (N_e = 151±84) while Manchado de Jabugo and Chato Murciano had smaller effective

population sizes (N_e = 46±50 and 59±31 respectively). Nevertheless, notice the large SD of these estimates.

Table 4.2 Current Effective population size (N_e) in each population. Sample size (N); Current effective population size (Ne); Standard deviation (SD).

POP	N	Ne ± SD
BI	15	74±37
CM	25	59±31
MJ	7	46±50
NI	15	95±49
RE	10	88±126
IB*	31	151±84
WB	18	180±61

*Ne in Iberian Breed, considering RE, NI and IB_Unk as a single population.

Past effective population size

The past N_e at generation T, where T = 1/2c [14], was similarly estimated for each bin of 1 Mb and sorted based on decreasing recombination rate values. This approach allowed studying N_e from as few as 5 to 20,000 generations ago (Figure 4.5). Similar to the estimation of the current N_e, wild boar tended to have the highest past N_e, followed by Iberian pigs. A noteworthy drop in N_e was observed in wild boar 10,000 – 20,000 generations ago, with a decrease of N_e from over 70,000 to below 30,000. The N_e increased rapidly in Iberian pigs at ~3,500 - 5,000 showing a maximum N_e at ~3,500 generations ago (N_e ~ 12,000). This increase in N_e was not observed in any other population (Additional file 7).

Runs of homozygosity

ROH of a minimum of 10 kbp containing at least 20 homozygous SNPs were studied in each individual separately. All individuals included in this study showed ROH. However, there were marked differences between populations in terms of number and length of ROH (Figure 4.6). The sums of all ROH per animal allowed the estimation of the percentage of the genome covered by ROH in each population (Additional file 6). The Chato Murciano had the largest mean proportion of its genome covered by ROH (29%). Other populations had mean values lower than 20%, with Bisaro displaying the lowest mean proportion (10%). The mean of the

total number of ROH per population was higher in Chato Murciano and wild boar (34 and 30 respectively) than in Iberian and Manchado de Jabugo (26 and 24 respectively). The Bisaro breed showed the lowest mean number of ROH (13). Regarding the length of ROH, approximately 36% of the Chato Murciano pigs analyzed had long ROH (> 100 Mbp) and 92% of the pigs of this breed had ROH in the range of 50-100 Mbp, making Chato Murciano the population with the highest proportion of long ROH. By contrast, none of the wild boars analyzed had long ROH, and only 20% of wild pigs analyzed contained ROH in the range of 50-100 Mbp, indicating that wild boar had shorter ROH than the other populations.

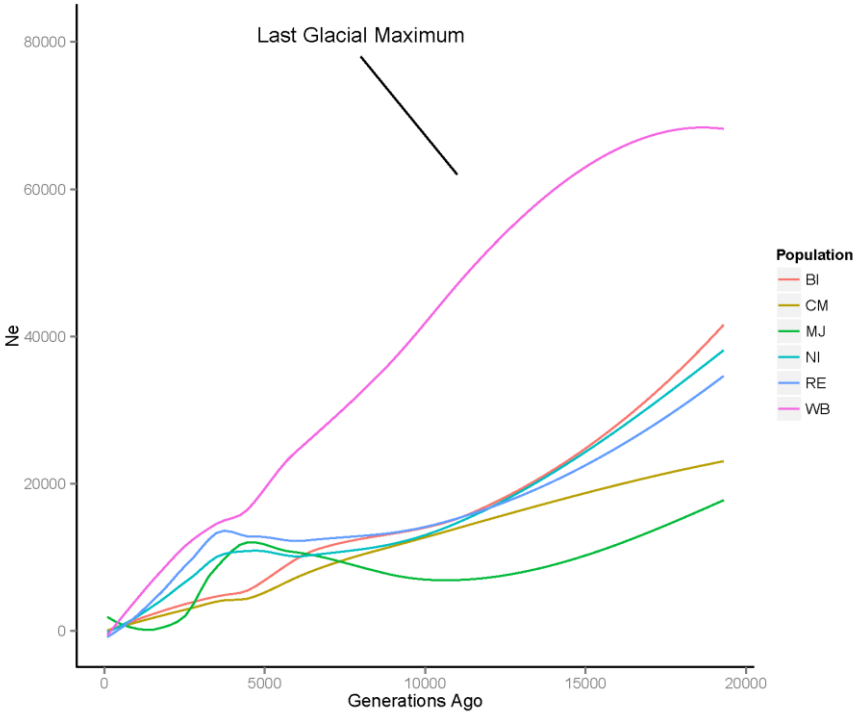

Figure 4.5 Estimated effective population size (N_e) over time. WB, wild boar; IB_Unk, Iberian unknown variant; NI, Negro Iberian; RE, Retinto; MJ, Manchado de Jabugo; BI, Bisaro; CM, Chato Murciano.

Manchado de Jabugo and Bisaro contained fewer ROH than the other populations, with Manchado de Jabugo displaying a higher proportion of long ROH than Bisaro. Twenty-five percent of Manchado de Jabugo pigs showed long ROH and 75% contained ROH in the range of 50-100 Mbp. These percentages are 6% and 33% respectively in Bisaro. Finally, Iberian had values intermediate to these breeds, since 16% of Iberian pigs displayed long ROH and 58% in the range of 50 – 100 Mb. All the pigs analyzed showed ROH shorter than 50 Mbp.

A Pearson's correlations matrix was made including LD, length of ROH and recombination rate. We found a positive correlation between mean values of LD and length of ROH per chromosome ($\rho = 0.70$, $p < 0.002$) while the correlation was negative between lengths of ROH and recombination rates ($\rho = -0.67$, $p < 0.003$).

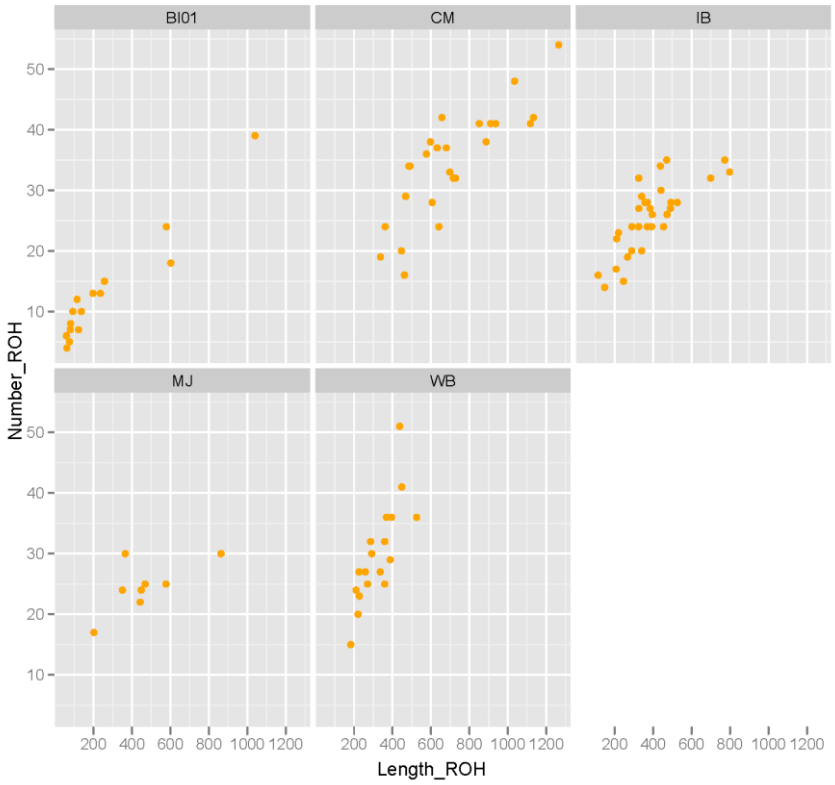

Figure 4.6 Average of Number of ROH Vs. Average of length of ROH. Each dot represents an individual. WB, wild boar; IB, Iberian; MJ, Manchado de Jabugo; BI, Bisaro; CM, Chato Murciano.

4.3 Discussion

High density SNP analysis can provide information on past and current population demography. LD is largely affected by population history and demography [28–30], constituting a potential tool to be applied to genetic population management. Specifically, LD can be used to estimate past and present effective population size [31] and to study persistence of LD phase [24]. The availability of a large number of SNPs, allows the study of parameters that can be directly relevant to assess effects of inbreeding such as the occurrence of Runs of Homozygosity (ROH). In addition, high-density SNP arrays are expected to improve the accuracy to assess population structure and relationship among populations [32–34]. Nevertheless, the applications of high-density SNP assays for investigations into genetic population management of minority breeds and wild populations in pig are scarce.

Relationships between populations

Understanding the relationships among and within populations of livestock is a necessary first step to establish conservation priorities and strategies [35, 36]. Population structure analysis, based on differences in allele frequencies, has been a proven method to assess relationships between populations [23, 37]. We combined this widely used approach with the estimates of persistence of phase as a measure of relationship between populations [24]. The different methods implemented to assess the relationships between populations showed a high degree of congruence.

The results obtained from the population structure and persistence of phase analyses indicate closer relationships between Chato Murciano and Bisaro, and between Iberian pigs and wild boar. This observed division seems to reflect the classical separation of pig populations from the Iberian Peninsula into two origins: the Celtic type and the Iberian type pigs [5]. Most of Celtic type breeds from the Iberian Peninsula are now extinct or are highly endangered. The Bisaro pig is a representative of this group [38]. All the variants of Iberian, which include Retinto, Negro Iberico and Manchado de Jabugo, belong to the Iberian type. Although the similarity between Chato Murciano and Bisaro could be due to a Celtic origin – or at least a mixed origin – of Chato Murciano, it is possible that the differentiation between the two groups of pigs actually differentiate admixed and non-admixed populations.

The high Pearson correlations for persistence of phase at long genetic distances detected between Manchado de Jabugo and other Iberian variants is typical for subpopulations of the same breed [15]. Furthermore, structure analysis confirmed the close genetic relationship between variants of Iberian pig but also signs of genetic admixture in Manchado de Jabugo. This is agreement with historical

records documenting that Manchado de Jabugo is a variant of the Iberian crossed with foreign pigs [8].

Inbreeding and effective population size

The analysis of N_e and ROH can be used to address major issues in conservation genetics such as effects of genetic drift and inbreeding [39]. The small population size inferred for the majority of local populations enhances the effect of consanguinity and genetic drift, which could compromise the long-term viability of the populations. With the absence of pedigree data in many local breeds, genetic marker data can be used as a surrogate to estimate current and past N_e, for instance through exploring the extent of LD [31]. Despite the interest in N_e for conservation of populations, the estimation of this parameter is remarkably complex [40]. The estimation of N_e assumes an ideal population that is isolated, without migration, with random mating and with a constant N_e. Although it is recognized that these assumptions are generally violated in natural populations, estimation of N_e is widely used. The estimation of recent N_e has been computed using linked [31, 41] and unlinked [42] genetic markers. We estimated N_e separately for 1 Mb bins containing information of recombination rates in order to obtain more information of demographic history [41]. While this method may provide a greater temporal dimension of N_e [43] it may make it difficult to interpret the results of current N_e [40]. Additionally, our estimate of current N_e must be treated with care due to the low sample sizes, especially in those populations with a sample size lower than 15 animals (i.e. Manchado de Jabugo and Retinto).

ROH have been used to infer the history of consanguinity in human populations [39, 44] and cattle [16]. These studies have demonstrated that long ROH are related with high consanguinity levels and also have shown the existence of a good correlation with pedigree inbreeding coefficients [16, 45]. The existence of Chato Murciano pigs with a high number of long ROH shows the importance of recent inbreeding and thus low individual genetic diversity. Indeed, we observed three Chato Murciano pigs that had more than 45% of the genome covered by ROH, but also pigs with much lower percentage. This observation is consistent with known management strategies of this breed [23] and in agreement with a strong bottleneck described for this breed about 20 years ago when the entire breed consisted of only 30-40 breeding animals. The high number of long ROH also indicates that this breed has not recently been extensively crossed with other breeds otherwise the long ROH would have broken down. The frequency of long ROH in Manchado de Jabugo was similar to the other Iberian variants. Recent admixture between Manchado de Jabugo and other pig breeds, as observed in the structure analysis, may have resulted in the break-down of long (>100 Mbp)

homozygous haplotypes. Despite the fact that Manchado de Jabugo is highly endangered with extinction as suggested from the small census population size (http://dad.fao.org/), this population did not show signs of high levels of consanguinity, likely because of its admixed origin. Thus, the conservation program currently implemented in Manchado de Jabugo is effective and necessary to assure the future viability of this population. What is also evident, however, is that this management strategy has gone at the expense of the historical genetic distinctiveness of the breed. By contrast, Bisaro showed signs of low consanguinity in agreement with its mixed origin and the strict conservation program implemented in this breed [46]. Although the Iberian pigs generally showed relatively low percentage of the genome covered by ROH, a few individuals showed a high coverage by ROH [47]. This can be expected in this heterogeneous breed which consists of local populations and different color forms.

The agreement between our observations and expectations based on historic reports highlights that analyzing the structure of ROH can aid in assessing levels of current consanguinity, and historic events such as bottlenecks in local pig populations. Furthermore, the assessment of ROH at the individual level has practical implications in conservation programs. Animals displaying high levels of ROH, for instance, could be excluded or given lower priority for breeding purposes in endangered populations. However, it must be taken into account that the 60K SNP panel applied may underestimate the number of small ROH due to ascertainment bias [13] and may inflate the length of the longest ROH [16]. Yet Bosse *et al*. [13] and Purfield *et al*. [16] concluded that high-density SNP panels allow an appropriate estimation of ROH, especially for the analysis of large ROH.

Demographic history

The study of demographic history provides a better understanding of the current risk of inbreeding, and might facilitate predicting the effects of future changes in effective population size. Despite the fact that estimation of demography implies the simplification of a complex biological reality, estimation of effective population size based on LD and recombination rate provides useful predictions and consistent comparisons between populations [31, 48]. It must be considered that the estimation of recent N_e is more inaccurate than the estimation ancient N_e owing to the increase in the variability of N_e values as the length of the segment used for the estimation increase [14]. The estimation of past N_e for wild boar tended to be higher than for domestic populations, with important drops in N_e. Moreover, wild boar had a very high number of short ROH and no long ROH. A high number of short ROH has been related with a reduced population size in the past and low inbreeding in recent times [39]. This pattern could be explained by the bottlenecks

that occurred in Europe in the last century [49], that would have reduced the N_e of wild boar drastically. Moreover, continuous events of formation of subpopulations and migration between them, favored by the lack of geographic barriers across the Iberian Peninsula, and even occasional admixture with domestic pigs, could have avoided high inbreeding in wild boar populations. Genetic signs of migration between subpopulations of Iberian wild boars have been described in Portugal [50].

The low recombination rate observed in large parts of the porcine genome essentially allows a much wider window in the past effective population size. Assuming a generation interval of approximately 2 years [24]. We observed a distinct drop of the N_e between 20,000 and 10,000 generations ago exclusively in wild boar that seems to reflect the sharp population decrease during the Last Glacial Maximum [51]. The increase in population size observed exclusively in Iberian pigs around 4000-4500 generations ago (Additional file 7) is consistent with admixture events during domestication in Europe [52, 53]. Recently the role of Europe as domestication centre has been dismissed [52, 53]. In this scenario of domestication, European domestic pigs appeared as a result of repeated events of admixture between domestic pigs imported from Near Eastern regions around 8,000 years ago [54] and wild pigs from Europe [52]. On one hand, the fact that Iberian pig allowed the study of domestication events implies that this breed represents a suitable model to study domestication in Europe, reinforcing the need to preserve the breed and avoid admixture with other pig populations. On the other hand, it confirms historic reports and previous studies with mtDNA showing that Iberian pigs did not originate from crosses with other breeds [55]. Admixture events may mask genetic signs of past demographic events [53] explaining why other domestic breeds such as Bisaro and Chato Murciano showed a different pattern of past N_e. Studies using Next Generation sequence data are needed to support and increase the accuracy of past N_e estimations in Iberian pig.

Structure analysis showed that Iberian pigs contributed to the wild boar genetic stock (Figure 4.2). This fact together with the genetic distances and F_{st} values between wild boar and Iberian pigs provide support that Iberian may have been crossbreeding with wild boar until medieval times [56]. This is in agreement with results for the European breeds studied by Groenen *et al.* [51], who describe a complex history in European breeds and incomplete lineage sorting, supporting admixture between wild and domestic pigs in Europe. It must be kept in mind that Iberian pigs were traditionally bred outdoors, which has enabled crossbreeding between wild and domestic pigs. Intriguingly, this is unlikely to have affected only the domesticated pigs, but rather also the wild boar of the Iberian Peninsula, as

suggested by recent investigations on introgression from domesticated into wild populations [12].

4.4 Conclusions

This study provides a comprehensive picture of demographic history, population structure and inbreeding of wild and domestic pig populations from the Iberian Peninsula as well as their relevance in conservation genetics. The occurrence of ROH in Chato Murciano was very high in some individuals, which may be due to a recent bottleneck and also highlights the lack of a well-designed genetic management program. Manchado de Jabugo showed a relatively high heterozygosity. This is unexpected given the extremely low census population size, and most likely reflects recent admixture with commercial pig breeds observed in this population. Conservation programs need to be maintained and carefully designed in order to avoid further loss of genetic distinctiveness. The study of N_e and ROH in Bisaro indicated high genetic diversity of this breed as a result of its mixed origin and the efforts carried out to preserve this breed. We observed that the Iberian breed may represent a good model to assess genetic signs of past demographic events as domestication, being and additional argument for the need to preserve Iberian pig breed and to avoid crossbreeding with other breeds. Previous evidence supporting the Iberian breed as being closely related to wild boar were confirmed and further evidence was provided for recurrent crossbreeding between these populations in the past. The analysis of wild populations from different regions of the Iberian Peninsula indicates that migration of wild pigs across the Iberian Peninsula may be important for the maintenance of low levels of inbreeding of Iberian wild boar.

4.5 Methods

Animals and sampling

A total of 97 unrelated pig samples were collected from populations of the Iberian Peninsula, and genomic DNA was extracted by standard protocols. The study included 18 wild boars (WB) from different regions of Portugal (n = 11) and Spain (n = 7), and 79 domestic pigs. The domestic pigs utilized for the analysis belonged to three local breeds: Iberian –including a Retinto Iberian variant (RE, n = 11), a Negro Iberian variant (NI, n = 15), an unidentified Iberian variant (IB, n = 5), and Manchado de Jabugo (MJ, n = 8) –, Bisaro (BI, n = 15) and Chato Murciano (CM, n = 25).

SNP genotyping

High-density SNP genotyping was performed using the Porcine SNP60 Beadchip (IlluminaInc, USA) designed to genotype 62,163 SNPs [9], per manufacturer's protocol. For this study, only SNPs mapped to one of the 18 autosomes on *Sus scrofa* build 10.2 and with less than 5% missing genotypes were included in the analysis.

Data analysis

Pairwise genetic distances between animals were calculated as one minus the average proportion of alleles shared using PLINK v1.07 [45]. Nei's genetic distances [57] and F_{st} values between populations were calculated using the Power marker software [58]. The pairwise distances between individuals were used to construct a Neighbor-Joining tree in Mega 5.03[59]. The admixture model implemented by the program *Structure* v2.0 [60] was used to examine relatedness among pig populations and population stratification. K values (number of assumed clusters) from two to seven were tested. Consistent results were obtained by using a burning period of 10,000 followed by 10,000 Markov chain Monte Carlo (MCMC) repetitions. This analysis was replicated after 100,000 burning steps and 100,000 MCMC repetitions. The most likely number of clusters was determined by the Evanno method [26] using the web server Structure Harvester [61]. Moreover, to obtain further detail of the population structure, the PCA was performed using the program Eigenstrat [62].

Linkage Disequilibrium analysis

Markers significantly deviating from Hardy-Weinberg equilibrium ($P < 0.001$) and with a MAF lower than 0.05 were excluded for LD analysis using PLINK v1.07 [63]. LD (r2) was estimated for all marker pairs less than 3Mbp apart across all populations and in each autosomal chromosome independently using Haploview 4.2 [64]. Graphic display of r^2 *vs.* distance per chromosome and means plot of r^2 in each breed vs. each chromosome were made in R environment (http://www.r-project.org/).

Persistence of phase

To calculate the persistence of phase and the time since two breeds diverged we followed the procedure implemented by Badke *et al.* [24]. Briefly, the SNP data was split into groups of SNPs with pairwise marker distance of 100 kbp, and the pairwise Pearson correlation between SNP was estimated across the 10 possible pairs of population.

Effective Population size

Effective population sizes were calculated in all populations implementing the equation $r^2 = 1/(4N_e c + 1)$, where r^2 is the LD, c is the marker distance in morgans

between SNP and N_e is the effective population size [27]. Additionally, past effective population size at generation T was calculated as T = 1/2*c* [14]. Logically, this equation depends on recombination rates.

Previous authors [25] tended to apply the generalization 1 Mb ~ 1 cM to calculate N_e, but this assumption may lead to incorrect estimates of N_e. Recombination rate varies considerably across and within porcine chromosomes [20], to an even larger extent than observed in other mammals [13]. Instead, we used the averaged high-density recombination map described by Tortereau *et al.* [20]. The effective population size estimates were derived by averaging multiple genomic regions in order to have a better approximation of the effective population size [48]. Towards this end, the chromosomes were divided in 1 Mb bins containing information of recombination rates and average r^2 for all possible pairs of SNPs included in each bin. These bins were subsequently used to estimate past and present effective population size. The approximation of past N_e assumes that *c* is much larger than the mutation rate (~10^{-8} per locus and generation) [15] so bins with $c < 10^{-6}$ where not considered for past N_e estimation.

Runs of homozygosity

The software PLINK v1.07 [63]was used to detect ROH for individuals separately. The ROH were defined by a minimum of 10 kbp in size and 20 homozygous SNPs. One heterozygous SNP was permitted in ROH, so that the length of the ROH was not disrupted by an occasional heterozygote. In addition, minimum SNP density of 1SNP/Mb and a largest possible gap between SNPs of 1Mb were predefined in order to assure that the ROH were not affected by the SNP density.

Number of ROH, total length of ROH and the average of ROH length in each animal were calculated for each chromosome and the mean across animals was estimated for each breed. Those ROH longer than 100 Mbp were categorized as long ROH. The percentage of the total genome length affected by ROH in each animal was also inferred.

Acknowledgements

This project was financially supported by European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant #ERC-2009-AdG: 249894 (Sel Sweep project).

Supporting Information

Chapter_4.zip

https://mega.co.nz/#!6INXXKoC!MCamaScVdm_P1FM4lRMh8DLHq3B-rM0dpUKO-_4jK1U

References

1. Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen M a M, Crooijmans RPM a, Megens H-J, Zhang H, Okimoto R, Vereijken A, Jungerius A, Albers GAA, Lawley CT, Delany ME, MacEachern S, Cheng HH: **Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds.***Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:17312–7.
2. FAO: **Building on gender agrobiodiversity and local knowledge**. *Food and Agriculture Organization of the United Nations* 2006, **1**:1–5.
3. Barja I: **Prey and Prey-Age Preference by the Iberian Wolf Canis Lupus Signatus in a Multiple-Prey Ecosystem**. *Wildlife Biology* 2009, **15**:147–154.
4. Beltrán-Beck B, Ballesteros C, Vicente J, De la Fuente J, Gortázar C: **Progress in Oral Vaccination against Tuberculosis in Its Main Wildlife Reservoir in Iberia, the Eurasian Wild Boar.***Veterinary medicine international* 2012, **2012**:978501.
5. Megens HH-J, Crooijmans RPM a., San Cristobal M, Hui X, Li N, Groenen MAM, Magali SC, Xiao H, Ning L, Martien AMG: **Biodiversity of pig breeds from China and Europe estimated from pooled DNA samples: differences in microsatellite variation between two areas of domestication.***Genetics, selection, evolution : GSE* 2008, **40**:103–28.
6. Sáez-Royuela, Telleria JL: **The increased population of the Wild Boar (*Sus scrofa* L .) in Europe**. *Mammal Review* 1986, **16**:97–101.
7. Alves E, Fernández AI, Barragán C, Ovilo C, Rodríguez C, Silió L: **Inference of hidden population substructure of the Iberian pig breed using multilocus microsatellite data**. *Spanish Journal of Agricultural Research* 2006, **4**:37–46.
8. García Dory M, Martínez S, Orozco F: *Guía de campo de las razas autóctonas españolas*. Alianza Ed. Madrid: 1990.
9. Ramos AM, Crooijmans RPM a, Affara N a, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu Z-L, Kerstens HH, Law AS, Megens H-J, Milan D, Nonneman DJ, Rohrer G a, Rothschild MF, Smith TPL, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen M: **Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology.***PloS one* 2009, **4**:e6524.
10. Duijvesteijn N, Knol EF, Merks JWM, Crooijmans RPM a, Groenen M a M, Bovenhuis H, Harlizius B: **A genome-wide association study on androstenone levels in pigs reveals a cluster of candidate genes on chromosome 6.***BMC genetics* 2010, **11**:42.

11. Ponsuksili S, Murani E, Brand B, Schwerin M, Wimmers K: **Integrating expression profiling and whole-genome association for dissection of fat traits in a porcine model.***Journal of lipid research* 2011, **52**:668–78.
12. Goedbloed DJ, Megens HJ, Van Hooft P, Herrero-Medrano JM, Lutz W, Alexandri P, Crooijmans RPM a, Groenen M, Van Wieren SE, Ydenberg RC, Prins HHT: **Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations.***Molecular ecology* 2012, **22**:856–866.
13. Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz L a. F, Schook LB, Crooijmans RPM a., Groenen M a. M: **Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape**. *PLoS Genetics* 2012, **8**:e1003100.
14. Hayes BJ, Visscher PM, McPartlan HC, Goddard ME: **Novel multilocus measure of linkage disequilibrium to estimate past effective population size.***Genome research* 2003, **13**:635–43.
15. De Roos a PW, Hayes BJ, Spelman RJ, Goddard ME: **Linkage disequilibrium and persistence of phase in Holstein-Friesian, Jersey and Angus cattle.***Genetics* 2008, **179**:1503–12.
16. Purfield DC, Berry DP, McParland S, Bradley DG: **Runs of homozygosity and population history in cattle.***BMC genetics* 2012, **13**:70.
17. Amaral AJ, Ferretti L, Megens H-J, Crooijmans RPMA, Nie H, Ramos-Onsins SE, Perez-Enciso M, Schook LB, Groenen MAM: **Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA.***PloS one* 2011, **6**:e14782.
18. Rubin C-J, Megens H-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg O, Jern P, Jorgensen CB, Archibald AL, Fredholm M, Groenen MAM, Andersson L: **Strong signatures of selection in the domestic pig genome**. *Proceedings of the National Academy of Sciences* 2012, **109**:19529–36.
19. Amaral AJ, Megens H-J, Crooijmans RPM a, Heuven HCM, Groenen M a M: **Linkage disequilibrium decay and haplotype block structure in the pig.***Genetics* 2008, **179**:569–79.
20. Tortereau F, Servin B, Frantz L, Megens H-J, Milan D, Rohrer G, Wiedmann R, Beever J, Archibald AL, Schook LB, Groenen MA: **A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content.***BMC genomics* 2012, **13**:586.
21. Wilkinson S, Archibald AL, Haley CS, Megens H-J, Crooijmans RPMA, Groenen MAM, Wiener P, Ogden R: **Development of a genetic tool for product regulation in the diverse British pig breed market.** *BMC genomics* 2012, **13**:580.

22. Ramos AM, Megens HJ, Crooijmans RPMA, Schook LB, Groenen MAM: **Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing.** *Animal genetics* 2011, **42**:613–20.
23. Herrero-Medrano JM, Megens HJ, Crooijmans RP, Abellaneda JM, Ramis G: **Farm-by-farm analysis of microsatellite, mtDNA and SNP genotype data reveals inbreeding and crossbreeding as threats to the survival of a native Spanish pig breed.** *Animal genetics* 2012, **In press**.
24. Badke YM, Bates RO, Ernst CW, Schwab C, Steibel JP: **Estimation of linkage disequilibrium in four US pig breeds.** *BMC genomics* 2012, **13**:24.
25. Uimari P, Tapio M: **Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds.***Journal of animal science* 2011, **89**:609–14.
26. Evanno G, Regnaut S, Goudet J: **Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study.***Molecular ecology* 2005, **14**:2611–20.
27. Sved JA: **Linkage Disequilibrium of Chromosome Segments**. *Theoretical Population Biology* 1971, **141**:125–141.
28. Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R, Lander ES: **Linkage disequilibrium in the human genome.***Nature* 2001, **411**:199–204.
29. Megens H-J, Crooijmans RPM a, Bastiaansen JWM, Kerstens HHD, Coster A, Jalving R, Vereijken A, Silva P, Muir WM, Cheng HH, Hanotte O, Groenen M: **Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromosomes in chicken.***BMC genetics* 2009, **10**:86.
30. Ardlie KG, Kruglyak L, Seielstad M: **Patterns of linkage disequilibrium in the human genome.***Nature reviews. Genetics* 2002, **3**:299–309.
31. Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM: **Recent human effective population size estimated from linkage disequilibrium.***Genome research* 2007, **17**:520–6.
32. Downing T, Imamura H, Decuypere S, Clark TG, Coombs GH, Cotton JA, Hilley JD, Doncker S De, Maes I, Mottram JC, Quail MA, Rijal S, Sanders M, Stark O, Sundar S, Vanaerschot M, Hertz-fowler C, Dujardin J, Berriman M: **Whole genome sequencing of multiple Leishmania donovani clinical isolates provides insights into population structure and mechanisms of drug resistance**. 2011:2143–2156.
33. Gautier M, Laloë D, Moazami-Goudarzi K: **Insights into the genetic history of French cattle from dense SNP data on 47 worldwide breeds.***PloS one* 2010, **5**.

34. Decker JE, Pires JC, Conant GC, McKay SD, Heaton MP, Chen K, Cooper A, Vilkki J, Seabury CM, Caetano AR, Johnson GS, Brenneman RA, Hanotte O, Eggert LS, Wiener P, Kim J-J, Kim KS, Sonstegard TS, Van Tassell CP, Neibergs HL, McEwan JC, Brauning R, Coutinho LL, Babar ME, Wilson GA, McClure MC, Rolf MM, Kim J, Schnabel RD, Taylor JF: **Resolving the evolution of extant and extinct ruminants with high-throughput phylogenomics.***Proceedings of the National Academy of Sciences of the United States of America* 2009, **106**:18644–9.

35. Berthouly C, Maillard JC, Pham Doan L, Nhu Van T, Bed'Hom B, Leroy G, Hoang Thanh H, Laloë D, Bruneau N, Vu Chi C, Nguyen Dang V, Verrier E, Rognon X: **Revealing fine scale subpopulation structure in the Vietnamese H'Mong cattle breed for conservation purposes.** *BMC genetics* 2010, **11**:45.

36. Druml T, Salajpal K, Dikic M, Urosevic M, Grilz-Seger G, Baumung R: **Genetic diversity, population structure and subdivision of local Balkan pig breeds in Austria, Croatia, Serbia and Bosnia-Herzegovina and its practical value in conservation programs.***Genetics, selection, evolution : GSE* 2012, **44**:5.

37. McKay SD, Schnabel RD, Murdoch BM, Matukumalli LK, Aerts J, Coppieters W, Crews D, Dias Neto E, Gill C a, Gao C, Mannen H, Wang Z, Van Tassell CP, Williams JL, Taylor JF, Moore SS: **An assessment of population structure in eight breeds of cattle using a whole genome SNP panel.** *BMC genetics* 2008, **9**:37.

38. Royo LJ, Álvarez I, Fernández I, Pérez-Pardal L, Álvarez-Sevilla, A. Godinho R, Ferrand N, Goyache F: **Genetic characterisation of Celtic-Iberian pig breeds using microsatellites**. In *Proceedings of the 6 th International Symposium on the Mediterranean Pig*. 2007:32–35.

39. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF: **Genomic runs of homozygosity record population history and consanguinity.***PloS one* 2010, **5**:e13996.

40. Wang J: **Estimation of effective population sizes from data on genetic markers.***Philosophical transactions of the Royal Society of London. Series B, Biological sciences* 2005, **360**:1395–409.

41. Hill WG: **Estimation of effective population size from data on linkage disequilibrium**. *Genet. Res.* 1981, **38**:209–216.

42. Park L: **Effective population size of current human population.** *Genetics research* 2011:1–10.

43. Waples RS, England PR: **Estimating contemporary effective population size on the basis of linkage disequilibrium in the face of migration.***Genetics* 2011, **189**:633–44.

44. Pemberton TJ, Absher D, Feldman MW, Myers RM, Rosenberg N a, Li JZ: **Genomic patterns of homozygosity in worldwide human populations.** *American journal of human genetics* 2012, **91**:275–92.

45. Mcquillan R, Leutenegger A, Abdel-rahman R, Franklin CS, Pericic M, Barac-lauc L, Smolej-narancic N, Janicijevic B, Polasek O, Tenesa A, Macleod AK, Farrington SM, Rudan P, Hayward C, Vitart V, Rudan I, Wild SH, Dunlop MG, Wright AF, Campbell H, Wilson JF: **Runs of Homozygosity in European Populations**. *The American Journal of Human Genetics* 2008, **83**:359–372.

46. Célio Alves P, Pinheiro I, Godinho R, Vicente J, Gortázar C, Scandura M: **Genetic diversity of wild boar populations and domestic pig breeds (Sus scrofa) in South-western Europe**. *Biological Journal of the Linnean Society* 2010, **101**:797–822.

47. Esteve-Codina A, Kofler R, Himmelbauer H, Ferretti L, Vivancos AP, Groenen MAM, Folch JM, Rodríguez MC, Pérez-Enciso M: **Partial short-read sequencing of a highly inbred Iberian pig and genomics inference thereof.***Heredity* 2011, **107**:256–64.

48. Stumpf MPH, McVean G a T: **Estimating recombination rates from population-genetic data.***Nature reviews. Genetics* 2003, **4**:959–68.

49. Apollonio M, Randi E, Toso S: **The systematics of the wildboar (Sus scrofa L.) in Italy.***Bollettino Di Zoologia* 1988, **3**:213–221.

50. Ferreira E, Souto L, Soares AMVM, Fonseca C: **Genetic structure of the wild boar population in Portugal: Evidence of a recent bottleneck**. *Mammalian Biology - Zeitschrift für Säugetierkunde* 2009, **74**:274–285.

51. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-J, Li S, Larkin DM, Kim H, Frantz LAF, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S, Bujie Z, et al.: **Analyses of pig genomes provide insight into porcine demography and evolution.***Nature* 2012, **491**:393–8.

52. Ottoni C, Flink LG, Evin A, Geörg C, De Cupere B, Van Neer W, Bartosiewicz L, Linderholm A, Barnett R, Peters J, Decorte R, Waelkens M, Vanderheyden N, Ricaut F-X, Cakirlar C, Cevik O, Hoelzel a R, Mashkour M, Karimlu AFM, Seno SS, Daujat J, Brock F, Pinhasi R, Hongo H, Perez-Enciso M, Rasmussen M, Frantz L, Megens H-J, Crooijmans R, Groenen M, et al.: **Pig domestication and human-mediated dispersal in western Eurasia revealed through ancient DNA and geometric morphometrics.** *Molecular biology and evolution* 2013, **30**:824–32.

53. Larson G, Burger J: **A population genetics view of animal domestication.** *Trends in genetics : TIG* 2013, **29**:197–205.

54. Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, Vigne J-D, Edwards CJ, Schlumbaum A, Dinu A, Balaçsescu A, Dolman G, Tagliacozzo A, Manaseryan N, Miracle P, Van Wijngaarden-Bakker L, Masseti M, Bradley DG, Cooper A: **Ancient DNA, pig domestication, and the spread of the Neolithic into Europe.** *Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**:15276–81.

55. Alves E, Ovilo C, Rodríguez MC, Silió L: **Mitochondrial DNA sequence variation and phylogenetic relationships among Iberian pigs and other domestic and wild pig populations.** *Animal genetics* 2003, **34**:319–24.

56. Van Asch B, Pereira F, Santos LS, Carneiro J, Santos N, Amorim a: **Mitochondrial lineages reveal intense gene flow between Iberian wild boars and South Iberian pig breeds.***Animal genetics* 2012, **43**:35–41.

57. Nei M: **Genetic distance between populations**. *The American Naturalist* 1972, **106**:283–92.

58. Liu K, Muse S V: **PowerMarker: an integrated analysis environment for genetic marker analysis.** *Bioinformatics* 2005, **21**:2128–9.

59. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0.***Molecular biology and evolution* 2007, **24**:1596–9.

60. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.** *Genetics* 2000, **155**:945–59.

61. Earl DA, vonHoldt BM: **STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method**. *Conservation Genetics Resources* 2011, **4**:359–361.

62. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D: **Principal components analysis corrects for stratification in genome-wide association studies.***Nature genetics* 2006, **38**:904–9.

63. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira M a R, Bender D, Maller J, Sklar P, De Bakker PIW, Daly MJ, Sham PC: **PLINK: a tool set for whole-genome association and population-based linkage analyses.** *American journal of human genetics* 2007, **81**:559–75.

64. Barrett JC, Fry B, Maller J, Daly MJ: **Haploview: analysis and visualization of LD and haplotype maps.***Bioinformatics* 2005, **21**:263–5.

5

Whole-genome sequence analysis reveals differences in population management and selection of low-input breeds between European countries

Herrero-Medrano JM[1], Megens HJ[1], Groenen MAM[1], Bosse M[1], Pérez-Enciso M[2,3], Crooijmans RPMA[1]

[1] Animal Breeding and Genomics Centre. Wageningen University. De Elst 16708 WD. Wageningen, The Netherlands. [2] Centre for Research in Agricultural Genomics, Consortium CSIC-IRTA-UAB-UB. Edifici CRAG, Campus Universitat Autonoma Barcelona, 08193 Bellaterra, Spain. [3] Institut Català de Recerca i Estudis Avançats (ICREA), 08010 Barcelona, Spain

Paper in preparation (2013)

Abstract

A major concern in conservation genetics is to maintain the genetic diversity of populations. Genetic variation in livestock species is threatened by the progressive marginalisation of local breeds in benefit of high-output pigs worldwide. We used high-density SNP and re-sequencing data to assess genetic diversity of local pig breeds from Europe. In addition, we re-sequenced pigs from commercial breeds to identify potential candidate mutations responsible for phenotypic divergence among these groups of breeds. Our results point out some local breeds that harbour low genetic diversity, whose genome shows a high proportion of regions of homozygosis and that harbour a large number of potentially damaging mutations. We also observed a high correlation between genetic diversity estimates using high-density SNP data and Next Generation Sequencing (NGS) data. The study of non-synonymous fixed mutations in commercial and local breeds revealed candidate mutations that may underlie differences in the adaptation to the environment but also potentially disadvantageous genotypes in highly productive breeds. This finding may be due to the strong artificial selection in the intensive production systems in pig. Our study highlights the importance of low input breeds as a valuable genetic reservoir for the pig production industry, emphasizing the need to implement conservation programmes to preserve the genomic variability of local breeds.

Key words: conservation genetics, pig, SNP, NGS, genetic diversity

5.1 Introduction

The use of a relatively small number of international high-output or commercials breed largely explains the increase in livestock productivity over the last 50-60 years. In addition, the number of commercial populations is even decreasing due to consolidation of breeding stock and breeding companies [1]. While high productive breeds may not compete with low-input breeds in marginal regions or extensive production, FAO has expressed concern due to the shift from local breeds to high-output animals [2]. Local breeds may be more resistant than high-performance breeds to local diseases, may be better adapted to local climate, and may be adapted to poorer food quality [2, 3]. These characteristics of local breeds are very relevant for humans living in developing countries where local domestic animals form an important source of protein. Local breeds are also appreciated in developed countries for their cultural heritage value, and as producers of traditional and high quality meat products [4]. Increasingly, local heritage breeds are recognized for their potential in sustainable or organic food production systems. Moreover, they represent a yardstick against which to compare highly selected breeds and allowing the detection of genes under selection [5]. Lastly, local breeds are claimed to harbour a large amount of the variation of domesticated species [6, 7], and as such are recognized as important genetic reservoirs that need to be protected for future food security [8].

Despite all those inherent properties of local breeds, the long term survival of many of them is not assured [8]. Inbreeding is particularly relevant in local breeds that have low population numbers [6, 7]. The loss of genetic diversity within a breed due to drift and inbreeding can have direct consequences for reduction of survival, reproduction efficiency and capacity of adaptation to environmental changes [9]. The reduction in reproduction and growth rates is particularly relevant for local livestock breeds as it can directly lead to economic loss. Minimising inbreeding is, therefore, a major goal to guarantee the sustainability and maintenance of domestic populations of livestock species.

Genetic characterization of livestock breeds by applying genetic marker technology is needed to enhance breeding and to better direct biodiversity conservation strategies. In pigs, the Porcine SNP60 Bead-array [10] is a commercially available marker system extensively used in genetic studies (e.g. [11, 12]). However, whole-genome re-sequencing has emerged as a tool for assessing genomic variation among pig populations [13]. In contrast to the commercially available SNP chip, the study of the whole genome sequence provides the opportunity of performing unbiased and accurate studies to estimate genetic diversity [14], regions of

homozygosity [15], and scanning the pig genome to detect signatures of selection [16]. The study of entire genomes increases the availability of information on neutral loci, and thereby the accuracy of estimates of demographically important parameters, such as the inbreeding factor (F) [17]. Next generation sequencing (NGS) also allows for direct assessment of polymorphisms in coding regions that could have consequences in selective processes. For instance, genes involved in local adaptation, or alleles responsible for inbreeding depression can be analysed [17].

In this study, we first assess and compare genetic diversity of low-input breeds from Europe by integrating high-density SNP and re-sequencing data. Secondly, we explore the role of local breeds as reservoirs for genetic variation in a domesticated species. Finally, we assess differences between local and commercial populations in terms of functional variation and explore evidences for inbreeding in local breeds that could lead to inbreeding depression.

5.2 Results

We genotyped 12 local populations (Table 5.1) from different countries across Europe with the Porcine SNP60 BeadChip [10]. SNP markers with more than 5% missing genotypes were excluded from the analysis. A total of 48,641 SNPs that could be mapped to autosomes on *Sus scrofa* build 10.2 [13] were finally used for the genetic diversity analysis. In addition, one or two representative genotyped pigs of these populations were re-sequenced to approximately 10x depth of coverage. The number of genomic variants, SNPs, and insertions or deletions (INDELs), varied greatly among the animals studied, ranging from 3.10 million in one Large White pig to 5.77 million in one British Saddleback pig. The number of variants and variability within exonic, intergenic, and intronic regions in all the re-sequenced animals is shown in the supplementary Table S1. In addition, a re-sequenced African Warthog was used as an out-group to deduce which fixed allele is ancestral or derived. Lastly, to understand the distribution of alleles in non-western domestic populations, we made comparisons with a panel consisting of European and Asian Wild Boar and Chinese pigs.

Table 5.1 Sampling information and analysis performed in each pig population.

Breed	Code	Category	Country	N	SNP	NGS
British Saddleback	BS	Local	UK	29	29	2
Gloucester Old Spots	GO	Local	UK	33	33	2
Large Black	LB	Local	UK	30	30	1
Middle White	MW	Local	UK	27	27	2
Tamworth	TA	Local	UK	30	30	2
Chato Murciano	CM	Local	Spain	46	46	2
Iberian Pig	IB	Local	Spain	29	29	2
Cinta Senese	CS	Local	Italy	13	13	1
Cassertana	CT	Local	Italy	15	15	2
NeraSiciliana	NS	Local	Italy	15	15	0
Calabrese	CA	Local	Italy	15	15	1
Mangalica	MA	Local	Hungary	25	0	2
Duroc	DU	Commercial	International	2	0	2
Large White	LW	Commercial	International	2	0	2
Landrace	LR	Commercial	International	2	0	2
Pietrain	PI	Commercial	International	2	0	2
Warthog	-	Wild	-	2	0	2
Wild boar	WB	Wild	China	3	0	3
Wild boar	WB	Wild	The Netherlands	2	0	2

Genetic diversity

To estimate genetic diversity with 60K data, we used the gene diversity parameter computed with Arlequin [18] per population (He_60K). We also estimated individual inbreeding factor averaged in each population (F_60K) (Table 5.2). In addition, NGS data was used to calculate heterozygosity (h_NGS) [14]. The estimation of h_NGS was performed for each pig separately, and, when data from two individuals were available, the average was used as the estimation of h_NGS in the breed. The comparison of genetic diversity derived from 60K and NGS is shown in Table 5.2 and Figure 5.1. All parameters indicated that Mangalica has the lowest genetic diversity (He_60K = 0.19; h_NGS = 5.55E-04) and British Saddleback the highest (He_60K = 0.29; h_NGS = 1.56E-03). The two marker systems also agreed in the low genomic variability of Cinta Senese breed (He_60K = 0.21, h_NGS = 8.23E-04), high variability in Chato Murciano and Middle White (He_60K = 0.27-0.28; h_NGS= 1.30E-03-1.35E-03 respectively) and intermediate levels for Calabrese (He_60K = 0.25; h_NGS = 1.14E-03). Minor disagreements between the genotyping methods were observed in Iberian breed, with a lower estimate of genetic diversity based on NGS than on 60K data. In the English breeds Tamworth and Gloucester Old Spots the genetic diversity was low according to the 60K data (He_60K < 0.21) but intermediate based on the NGS data (h_NGS~ 1.10E-03). The major disagreement between 60K and NGS data was found in the populations Cassertana and Large Black. We observed a proportionally higher diversity in Large Black when NGS data was used and the opposite for the Cassertana breed. We observed that all English breeds and Chato Murciano had higher genetic diversity in the estimates based on NGS than in 60K in those breeds, relative to the fitted line. On the other hand, pigs from Italy, Hungary and Iberian pig showed lower than estimated genetic diversity based on NGS relative to the 60K SNP data. Despite these systematic deviation of the fitted model, the Pearson's correlation coefficient was high and significant between He_60K and h_NGS genetic diversity estimates (0.70, P = 0.02), as well as between F_60K and h_NGS (-0.83, P = 0.003).

In order to allow a direct comparison between genetic diversity using the two marker systems, parameters studied at the individual level – F_60K and h_NGS – were compared (Figure 5.2). The correlation between h_NGS and F_ 60K for the individual comparisons was -0.91 (P = 0.0).

The number of Runs of Homozygosity (ROH) as well as their length varied greatly among populations as estimated from both 60K and NGS. In agreement with the genetic diversity estimates, all the analyses showed that the Mangalica breed had the highest proportion of the genome covered by ROH (Figure S1, supplementary material). The Italian breeds Cassertana and Cinta Senese and the English breeds

Tamworth and Gloucester Old Spots also had a high coverage of ROH (50-55% using NGS data). At the other end of the spectrum, the breed British Saddleback showed the lowest proportion (35%) followed by Calabrese and Chato Murciano (~ 40%).
A high correlation between estimates of ROH was observed between estimates derived from NGS and 60K SNP data, although the 60K SNP data consistently underestimated the proportion of the genome covered by ROH (Figure S1). The comparison between the number and length of ROH using 60K and NGS revealed that 60K data tended to not discover short ROH and to overestimate the length of long ROH (Figure S2, supplementary material). The correlation between length of ROH and genetic diversity estimates was high (~ -0.8, $P < 0.05$) as inferred in Figure S3 (supplementary material). The comparison of F value against the total length of ROH in the populations Calabrese, Chato Murciano, Cassertana and Middle White encompassed pigs with a pattern of negative F values as well as shorter and lower number of ROH (see cut-off, Figure S3).

Table 5.2 Genetic diversity parameters

Breed	N_a*	He*	F*	h**
BS	1.88	0.29	0.13	1,56E-03
NS	1.79	0.26	0.19	NA
MA	1.60	0.19	0.55	5,55E-04
CS	1.70	0.21	0.37	8,23E-04
CA	1.71	0.25	0.19	1,14E-03
MW	1.84	0.27	0.16	1,30E-03
CM	1.95	0.28	0.17	1,35E-03
LB	1.75	0.25	0.23	1,50E-03
IB	1.85	0.24	0.33	9,51E-04
CT	1.91	0.28	0.20	9,16E-04
TA	1.61	0.20	0.38	1,07E-03
GO	1.71	0.21	0.34	1,12E-03

*Estimates of genetic diversity using Porcine 60SNP Beadchip data. He: expected heterozygosity; F: inbreeding coefficient; N_a: Mean number alleles. **Heterozygosity estimated using NGS data. h: observed heterozygosity

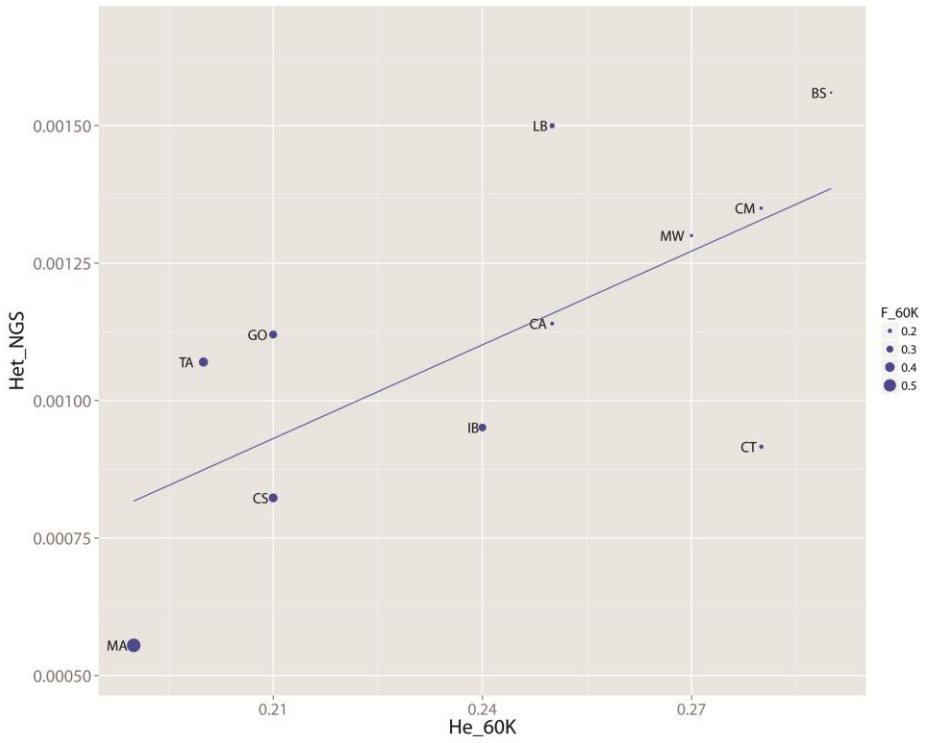

Figure 5.1 Genetic diversity with NGS Vs. Genetic diversity using 60K data. Each dot represents the average value in the populations. The size of the dots is proportional to the inbreeding factor (F_60K) observed in the population.

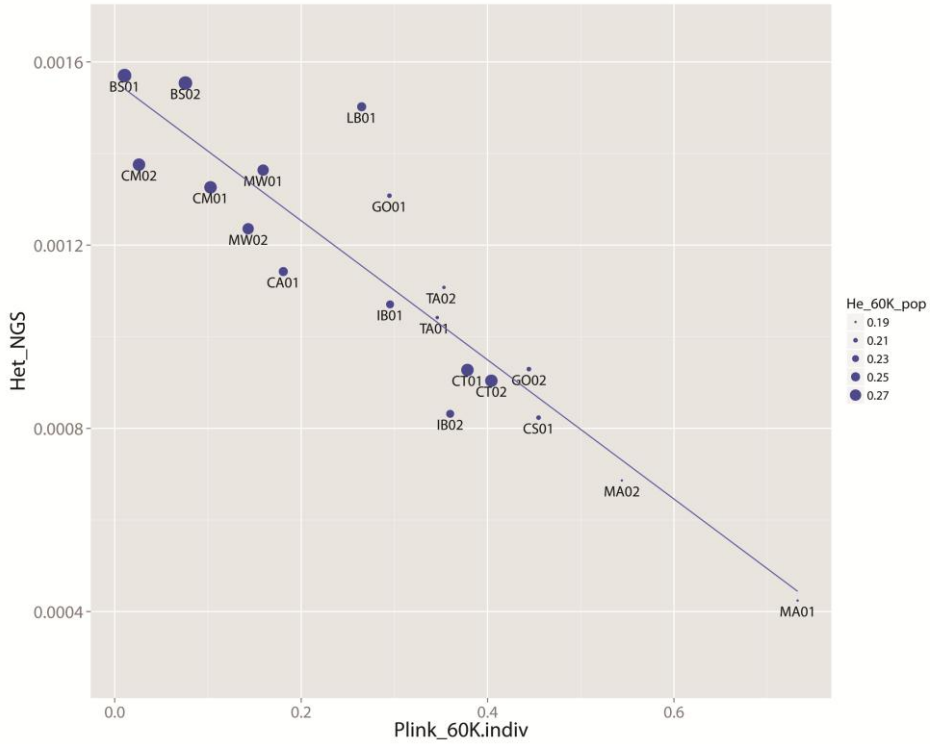

Figure 5.2 Genetic diversity with NGS Vs. Inbreeding factor using 60K data at individual level. The size of the dots is proportional to the He_60K of the population. BS, British Saddleback; GO, Gloucester Old Spots; LB, Large Black; MW, Middle White; TA, Tamworth; CM, Chato Murciano; IB, Iberian Pig; CS, Cinta Senese; CT, Cassertana; CA, Calabrese; MA, Mangalica.

Functional significance of non-synonymous variants

Of all the SNPs discovered by NGS, an average of 0.17% was annotated as non-synonymous variants (Table S1, supplementary material). Considering all individuals, we observed a total of 16,409 different non-synonymous SNPs. All non-synonymous SNPs were analysed with Polyphen2 [19], that classifies mutations as benign and possible/probably damaging. In agreement with the genetic diversity estimations (File S1, supplementary material) a high number of potentially damaging mutations is fixed in the breeds Mangalica, Cinta Senese, Tamworth and Gloucester Old Spots.

In order to find SNPs that potentially explain phenotypic differences between local populations and high-output pigs, we extracted all possible non-synonymous SNPs and we computed F_{st} (File S2). Eight pigs derived from commercial elite lines

(Duroc, Large White, Landrace and Pietrain) were considered as one population and each local breed was used separately to determine F_{st}. We focussed on those non-synonymous SNPs that were fixed in commercial breeds and also in any local breed but with different allele, i.e. Fst = 1. This analysis revealed 99 SNPs with different fixed alleles in commercial and at least one of the local breeds (Table S2, supplementary material). Moreover, we explored the occurrence of ROH and published QTL overlapping these SNPs and the result of the Polyphen2 analysis.

The 99 non-synonymous fixed SNPs affected 65 genes (Table S2). The comparison with a Warthog pig revealed that in 64% of fixed alleles it was the derived allele that was fixed in local pigs and 36% in commercial pigs. Among these 65 genes, we focused on those (i) with a potential phenotypic effect, (ii) with the two alleles –the ancestral and the derived– present in wild populations, (iii) those that were affected by several fixed SNPs and (iv) with a mutation classified by polyphen2 (Figure 5.3). We observed a possible damaging mutation in the gene *AZGP1* in the breeds Mangalica, Cinta Senese and Gloucester Old Spot, as well as in European wild boar. This mutation overlaps with a QTL related with the number of vertebra and occupied a 50 kb genomic region where genetic diversity varied greatly among populations –from 0 to 5 times the averaged genetic diversity in the pig–. We observed two fixed SNPs within the gene *IL12RB2*, with Gloucester Old Spots, Middle White, Tamworth, Calabrese carrying the two ancestral alleles. Other local breeds such as British Saddleback and Chato Murciano were heterozygous at this locus, as were European and Asian wild pigs. This genomic region overlaps with a QTLs for back fat thickness and intramuscular fat content. It also overlaps with ROH or low genetic diversity regions, except in British Saddleback and Large Black. A mutation classified as benign was observed within the gene *STAB1.* This gene codes for a protein involved in defence against bacterial infection by binding to bacteria and inducing phagocytic activity [20–22]. The allele was present in English breeds, Casertana and Asian pigs. *STAB1* overlaps with four QTLs related with CD4 and CD8 leukocyte percentage and ratio. The genetic diversity in this region is low, especially in commercial breeds with seven out of eight commercial breeds overlapping with a ROH. The two animals of the breed Mangalica, Chato Murciano and several English pigs were all homozygous for three derived alleles within the gene *EIF2AK3*. The protein coded by this gene is involved in skeletal system development. The gene overlaps with QTL for feet and leg conformation and Osteochondrosis score. Local pigs carrying the derived allele have a ROH or a low genetic diversity in the 50Kb region overlapping this gene.

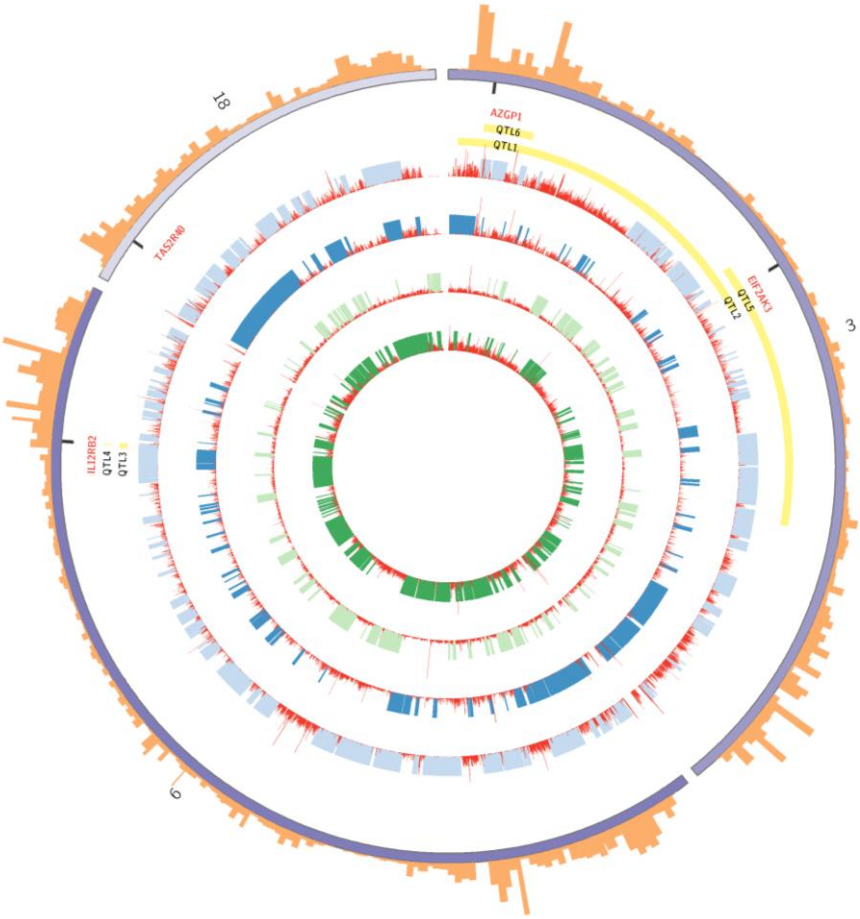

Figure 5.3 Chromosomes are arranged circularly end-to-end using Circos[54]. The four inner rings display ROH (green and blue bars) and genetic diversity (red histograms) in Large White, Landrace, Mangalitza and Tamworth respectively. Some QTLs overlapping any of the four genes studied are represented in yellow (QTL1: Abdominal fat weight; QTL2: Osteochondrosis score; QTL3: Intramuscular fat content; QTL4: Backfat thickness; QTL5: Feet and leg conformation ; QTL6: Vertebra number. The outer ring represents the averaged high-density recombination map described by Tortereau et al. [55].

5.3 Discussion

The advances in sequencing technologies now allows sequencing whole genomes in multiple individuals [17, 23]. However, the cost of this technology is still high, and budgets for conservation genetics research are limited. While high-density SNP panels allow the study of a representative sample size of a population at a much lower cost, there is a concern regarding the ascertainment bias implicit in the use of SNP chips [24]. This concern is even higher for local pig populations since they were not considered in the design of the Porcine SNP60 Beadchip [10]. In this study, we found a high correlation between diversity estimates derived from the Illumina porcine 60SNP Beadchip and NGS data. These results indicate that the Illumina porcine 60SNP Beadchip provides reliable estimates of genomic diversity for comparative studies between European populations, despite the expected bias. Nevertheless, the English breeds and Chato Murciano, showed greater diversity with NGS compared to 60K data than expected compared to expected values derived from all populations combined. These results may highlight the influence of historical breeding practices, whereby Asian pigs were used to improve local English pigs during the late 18^{th} and 19^{th} century [13, 25] which were subsequently used to improve other European breeds such as Chato Murciano [26]. Despite the additional diversity found in English pigs owing to Asian introgression, some English pigs display high levels of ROH. These ROH are the result of recent inbreeding and could indicate that these breeds are prone to inbreeding depression.

SNP variants were annotated and potential deleterious effects were predicted with Polyphen2. Recessive deleterious alleles can be a major cause of inbreeding depression in populations with low genetic diversity [27]. In our study we find the largest number of putative deleterious mutations in those animals that also have the highest percentage of the genome covered by ROH and the lowest genetic diversity, i.e. Mangalica and Cinta Senese breeds, and in the breeds Tamworth and Gloucester Old Spots. Genomic diversity in these breeds was lower than almost all domestic and wild populations from Europe and Asia [15] corroborating the hypothesis that damaging mutations can accumulate, due to drift, in populations with high levels of inbreeding. A similar relation between genetic diversity and proportion of deleterious alleles has been described in human populations [28]. This finding points out the need to develop conservation programs for endangered livestock populations that are very prone to high levels of inbreeding.

We found non-synonymous, high allele frequency differences (fixed for different alleles) at non-synonymous sites to be overrepresented in genes involved in immune response, anatomical development, behaviour, and sensory perception

between commercial and local populations. Local breeds tend to be reared in traditional systems without being subjected to intense artificial selection (e.g. BLUP, GBLUP selection) as applied to commercial pig populations. As a result of years of different selection pressures and environments, genomic variations underlying phenotypic differences can be expected. We have specifically focussed on non-synonymous variants because they will alter the amino acid sequence of gene products, which may result in different phenotypes [29]. Although phenotypic change is expected to a large extent to result from regulation of genes, rather than differences in amino acid sequences, regulatory important variations are currently difficult to predict reliably and were therefore not considered in this study.

The gene *AZGP1* stimulates lipid degradation in adipocytes and subsequently is considered a lipid-mobilizing factor [30]. This gene is linked with obesity in humans and its expression is inversely associated with body weight and percentage of body fat in mice and humans [31, 32]. In pigs, a 20Mb QTL in chromosome 3 [33] for abdominal fat weight overlaps this gene. Mangalica, Cinta Senese and one European wild boar are homozygous for a derived allele annotated as probably damaging. This allele is absent in commercial pigs and also in some local pigs. The inferred status of the allele as 'probably damaging' may, for pig, rather result in having a large effect on the phenotype. Whereas pigs used to be bred for high fat deposition, in modern pig production systems lean meat is desired.

AZGP1 also overlaps with a 8.5Mb QTL for vertebra number [34]. Related to that same phenotype, we found a fixed non-synonymous mutation in the Mangalica breed within the gene *PLAG1* that has been related with stature in humans and cattle [35, 36]. Rubin et al. [16] concluded a strong signature of selection in the domestic pig genome at *PLAG1*. These data suggest that the mutations found in the genes *AZGP1* and *PLAG1* may represent signatures of different selection pressures between local breeds as Mangalica and commercial pigs. Another compelling example of potential differential selection between commercial and local populations are the two mutations found in the bitter taste receptor *TAS2R40*. The high variability within the family of taste receptor genes has been suggested a consequence of adaptation of populations to specific dietary repertoires and environment [37], such as prevention of consumption of plant toxins [38].

It has been observed that selection for economically important traits tends to increase the susceptibility to environmental factors [39, 40]. In our study, ancestral mutations classified as benign in genes involved in immune related genes such as *IL12RB2* and *STAB1*, were observed in several local pigs. The *IL12RB2* subunit plays an important role in Th1 cell differentiation that is critical for an effective immune response against different types of pathogens [41]. The three mutations observed

in this gene overlap with important QTLs in pig production such as back fat thickness and intramuscular fat content [42, 43]. The fact that mutations in *IL12RB2* can lead to a defective IFN-gamma response to microorganisms [44, 45], suggests that disadvantageous genotypes could have been maintained in commercial populations.

The *EIF2AK3* gene overlaps with QTLs for osteochondrosis score [46] and feet and leg conformation [47]. Interestingly, this gene encompass functions of bone mineralization, chondrocyte development insulin secretion and fat cell differentiation and has being related with the Wolcott-Rallison syndrome in humans [48]. Leg weakness is a major concern in growing pigs raised under modern production systems and osteochondrosis is considered to be the primary cause of this syndrome. Indeed, forced selection for high growth capacity predisposes to these disorders due to an imbalance between the development of the skeletal system and muscle [49]. The allelic differences between local and commercial pigs within the *EIF2AK3* gene, could underlie strong directional selection in commercial breeds. The fact that the same alleles are segregating in both wild boar and low-input breeds supports this hypothesis.

The genes discussed above had different fixed alleles for non-synonymous SNPs between commercial and local pigs. The presence of both alleles, the ancestral and the derived, in wild boars indicates that the variation was present before domestication. While differences in allele frequencies of SNPs in genes such as *AZGP1* and *TAS2R40* may underlie a rapid adaptation to different environments, it can also occur due to drift effects in small populations in the absence of selection, or even if the allele is in fact disadvantageous. The fixed alleles in *EIF2AK3* and *IL12RB2* could potentially result in disadvantageous phenotypes in high-output breeds owing to the strong artificial selection for production traits. We demonstrated that genetic variability found in wild populations is also being preserved in local breeds at genomic sites with potential phenotypic effect. This further highlights the importance of preserving local breeds as a source of genomic diversity that could be used in future selection programs of commercial pigs. However, the results presented also highlight high levels of ROHs, inbreeding and potentially damaging mutations that threat the future of local pig breeds, emphasizing the need of implementing conservation programmes to preserve the genomic variability of low-input breeds.

5.4 Conclusion

In this study, we assessed genetic diversity of low-input breeds from different European regions by integrating high-density SNP and re-sequencing data. The comparison of the two marker system estimations provided insights for strategies to the genetic characterization of local breeds. Furthermore, the re-sequenced local pigs were compared with re-sequenced commercial pigs to report candidate mutations responsible for phenotypic divergence among those groups of breeds. We observed that local pig breeds are an important source of genomic variation within-species, and thereby, they represent a genomic stock that could be important for future adaptation to long-term changes in the environment or consumers preferences. However, high levels of inbreeding threaten the long term survival of some of the local breeds studied.

5.5 Material and methods

Animals and sampling and SNP genotyping

Blood samples from 315 unrelated domestic pigs were collected and DNA was extracted by using the QIAamp DNA blood spin kit (Qiagen Sciences). The study included domestic pigs that belonged to 12 local breeds from England, Spain, Italy and Hungary (Table 5.1). Samples were genotyped using the Illumina Porcine 60K iSelect Beadchip [10] per manufacturers protocols. We included only SNPs mapped to one of the 18 autosomes on *Sus scrofa* build 10.2 and that had less than 5% missing genotypes. In addition, 1-2 animals of each local breed were selected for re-sequencing with the exception of the Nera Siciliana breed. We also re-sequenced eight individuals that belonged to the commercial, international pig breeds Duroc, Large White, Landrace and Pietrain. The samples used are detailed in Table 5.1.

Sequencing alignment and SNP discovery

Library construction and re-sequencing of the samples was performed using 1-3 µg of genomic DNA following the Illumina library prepping protocols (Illunima Inc.). The library insert size ranged for 300–500 bp and fragments were sequenced from both sides yielding two times 100bp mated sequences. Short read alignment was done against the *Sus scrofa* genome, build 10.2 [13] using Mosaik. The pigs were sequenced to a depth of approximately 10x. Further details on sequence mapping can be found in [15].

Archives in BAM format generated with the Mosaik Text function were used for the SNPs calling against the *Sus scrofa* genome, build 10.2. The mpileup function implemented in SAMtools v1.4-r985 [50] was used to obtain variant calls. Variations were filtered for a minimum genotype SNP and INDEL quality (20 and 50

respectively). Only variations based on coverage in the range of 5x until twice the genome average were considered.

Data analysis using high-density SNP genotyping

We used ARLEQUIN 3.5 [18] to compute the expected heterozygosity, number of polymorphic sites and mean number of alleles in each population.

The ROHs were defined with PLINK 1.07 as regions of a minimum size of 10 kbp and encompassing 20 homozygous genomic sites, while allowing one heterozygous SNP. We predefined a minimum SNP density of 1 SNP/Mb and a largest possible gap between SNPs of 1Mb to assure that the ROHs were not severely affected by the SNP density. Finally, we computed the Pearson's correlation coefficient between length of ROHs and genetic diversity parameters in each breeds using R (www.r-project.org).

Data analysis using NGS data

Heterozygosity was estimated for each individual as the number of heterozygous sites per 50 Kb-bin, corrected for total number of sites per bin [14]. Only bins that were sufficiently covered (per base at least a sequence depth of 7x and maximum of approximately 2*average coverage) were considered. We obtained the heterozygosity for the population by averaging the individual heterozygosity of all individuals that belonged to that population. Correlations between 60K and NGS genomic diversity estimates were calculated using Pearson's correlations in R environment. Graphics were obtained using the plotting system ggplot2 for R. To estimate the ROH from re-sequencing data, we followed the procedure implemented by Bosse et el. [15], using a 100 Kb sliding window. ROH were defined as a genomic region of at least 10kb where the number of SNPs in an individual is less than expected based on the genomic average. Briefly, if the number of SNPs per bin =< 0.25 the genomic average, and if 10 or more consecutive bins showed a total SNP average lower than the total genomic average, they were extracted as candidates ROH. ANNOVAR [51] was used to obtain the functional annotation (non-synonymous, synonymous, stop codon gain/loss, amino acid changes) of the genomic variants in each animal based on the pig reference genome (Swine Genome Sequencing Consortium Sscrofa10.2) obtained from the UCSC database (http://genome.ucsc.edu). For further analysis, only the non-synonymous sites were considered. The genes that overlap with the non-synonymous mutations were retrieved using Biomart [52].

The F_{st} value for all non-synonymous mutations was calculated using Genepop [53]. For this analysis all the commercial pigs were considered as a single population while each local breed was considered separately. To reduce the number of SNPs to those that most likely represent the genetic basis of the phenotypic differences

between commercial and local breeds, we only included in the study SNPs with F_{st} = 1 between the groups (i.e. fixed differences). Moreover, in order to avoid false positives, we exclusively considered those mutations that were homozygous in at least the two animals of the local breed. In the case of the local breeds that had only one animal re-sequenced or when one of the two animals of the breed showed missing data, the SNP was not considered for the functional analysis regardless its F_{st} value. Those SNPs with missing data in more than three commercial pigs were equally excluded. The sequence of a re-sequenced Warthog was used to ascertain the alleles as ancestral or derived. The genotypes for those SNPs were also obtained from re-sequenced data from two domestic Meishan pigs, one wild boar from South China, two from North China and two European wild boars. The sequencing alignment and SNP discovery of these samples was the same as previously detailed. Finally, we used the Polymorphism Phenotyping (PolyPhen2) algorithm [19] to predict phenotypic consequences of the non-synonymous sites. PolyPhen2 predicts whether a SNP is 'benign', 'possibly damaging' or 'probably damaging' on the basis of evolutionary conservation, structure and sequence information.

Acknowledgements

This project was financially supported by European Research Council under the European Community's Seventh Framework Programme (FP7/2007-2013)/ERC grant #ERC-2009-AdG: 249894 (Sel Sweep project).

Supporting Information

Chapter_5.zip
https://mega.co.nz/#!uZNC2QpD!Q4vf5yj2ydx-TyJ_i2lPnwt30fQpZhGzaT55PMsw_2Q

References

1. Muir WM, Wong GK-S, Zhang Y, Wang J, Groenen M a M, Crooijmans RPM a, Megens H-J, Zhang H, Okimoto R, Vereijken A, Jungerius A, Albers GAA, Lawley CT, Delany ME, MacEachern S, Cheng HH: **Genome-wide assessment of worldwide chicken SNP genetic diversity indicates significant absence of rare alleles in commercial breeds.***Proceedings of the National Academy of Sciences of the United States of America* 2008, **105**:17312–7.
2. FAO: *Intergovernmental Technical Working Group on Animal Genetic Resources for Food and Agriculture. Seventh Session*. Rome: 2012.
3. Ramsay K: **Marketing rare breeds in Sub-Saharan Africa**. In *Incentive measures for sustainable use and conservation of agrobiodiversity. Experiences and lessons from Southern Africa.* Lusaka, Zambia: 2002:61–68.
4. García-González DL, Aparicio R, Aparicio-Ruiz R: **Volatile and amino Acid profiling of dry cured hams from different Swine breeds and processing methods.***Molecules (Basel, Switzerland)* 2013, **18**:3927–47.
5. Wilkinson S, Lu ZH, Megens H-J, Archibald AL, Haley C, Jackson IJ, Groenen M a. M, Crooijmans RPM a., Ogden R, Wiener P: **Signatures of Diversifying Selection in European Pig Breeds**. *PLoS Genetics* 2013, **9**:e1003453.
6. Tapio I, Värv S, Bennewitz J, Maleviciute J, Fimland E, Grislis Z, Meuwissen THE, Miceikiene I, Olsaker I, Viinalass H, Vilkki J, Kantanen J: **Prioritization for conservation of northern European cattle breeds based on analysis of microsatellite data.***Conservation biology : the journal of the Society for Conservation Biology* 2006, **20**:1768–79.
7. Ollivier L: **European pig genetic diversity: a minireview.***Animal : an international journal of animal bioscience* 2009, **3**:915–24.
8. FAO: **Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration.***(available at http://www.fao.org/docrep/010/a1404e/a1404e00.htm).* 2007.
9. Frankham R, Ralls K: **Conservation biology: Inbreeding leads to extinction**. *Nature* 1998, **392**:441–442.
10. Ramos AM, Crooijmans RPM a, Affara N a, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu Z-L, Kerstens HH, Law AS, Megens H-J, Milan D, Nonneman DJ, Rohrer G a, Rothschild MF, Smith TPL, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen M: **Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology.***PloS one* 2009, **4**:e6524.

11. Sahana G, Kadlecová V, Hornshøj H, Nielsen B, Christensen OF: **A genome-wide association scan in pig identifies novel regions associated with feed efficiency trait.***Journal of animal science* 2013.
12. Manunza A, Zidi A, Yeghoyan S, Balteanu VA, Carsai TC, Scherbakov O, Ramírez O, Eghbalsaied S, Castelló A, Mercadé A, Amills M: **A High Throughput Genotyping Approach Reveals Distinctive Autosomal Genetic Signatures for European and Near Eastern Wild Boar**. *PLoS ONE* 2013, **8**:e55891.
13. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-J, Li S, Larkin DM, Kim H, Frantz LAF, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S, Bujie Z, et al.: **Analyses of pig genomes provide insight into porcine demography and evolution.***Nature* 2012, **491**:393–8.
14. Esteve-Codina A, Kofler R, Himmelbauer H, Ferretti L, Vivancos AP, Groenen MAM, Folch JM, Rodríguez MC, Pérez-Enciso M: **Partial short-read sequencing of a highly inbred Iberian pig and genomics inference thereof.***Heredity* 2011, **107**:256–64.
15. Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz L a. F, Schook LB, Crooijmans RPM a., Groenen M a. M: **Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape**. *PLoS Genetics* 2012, **8**:e1003100.
16. Rubin C-JC-J, Megens H-JH-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg O, Jern P, Jorgensen CB, Archibald AL, Fredholm M, Groenen MAM, Andersson L, Martinez Barrio A, Carlborg Ö, Jørgensen CB: **Strong signatures of selection in the domestic pig genome.***Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**:19529–36.
17. Allendorf FW, Hohenlohe P a, Luikart G: **Genomics and the future of conservation genetics.***Nature reviews. Genetics* 2010, **11**:697–709.
18. Excoffier L, Lischer HEL: **Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows.***Molecular ecology resources* 2010, **10**:564–7.
19. Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR: **A method and server for predicting damaging missense mutations.***Nature methods* 2010, **7**:248–9.
20. Adachi H, Tsujimoto M: **FEEL-1, a novel scavenger receptor with in vitro bacteria-binding and angiogenesis-modulating activities.***The Journal of biological chemistry* 2002, **277**:34264–70.

21. Kzhyshkowska J, Gratchev A, Brundiers H, Mamidi S, Krusell L, Goerdt S: **Phosphatidylinositide 3-kinase activity is required for stabilin-1-mediated endosomal transport of acLDL.***Immunobiology* 2005, **210**:161–73.
22. Kzhyshkowska J: **Multifunctional receptor stabilin-1 in homeostasis and disease.***TheScientificWorldJournal* 2010, **10**:2039–53.
23. Lachance J, Vernot B, Elbers CC, Ferwerda B, Froment A, Bodo J-M, Lema G, Fu W, Nyambo TB, Rebbeck TR, Zhang K, Akey JM, Tishkoff S a: **Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers.***Cell* 2012, **150**:457–69.
24. Albrechtsen A, Nielsen FC, Nielsen R: **Ascertainment Biases in SNP Chips Affect Measures of Population Divergence Research article**. 2010, **27**:2534–2547.
25. Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L: **The origin of the domestic pig: independent domestication and subsequent introgression.***Genetics* 2000, **154**:1785–91.
26. Herrero-Medrano JM, Megens HJ, Crooijmans RP, Abellaneda JM, Ramis G: **Farm-by-farm analysis of microsatellite, mtDNA and SNP genotype data reveals inbreeding and crossbreeding as threats to the survival of a native Spanish pig breed.***Animal genetics* 2012, **In press**.
27. Hagenblad J, Olsson M, Parker HG, Ostrander E a, Ellegren H: **Population genomics of the inbred Scandinavian wolf.***Molecular ecology* 2009, **18**:1341–51.
28. Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, Sninsky JJ, White TJ, Sunyaev SR, Nielsen R, Clark AG, Bustamante CD: **Proportionally more deleterious genetic variation in European than in African populations.***Nature* 2008, **451**:994–7.
29. T S, Read A: **Human Molecular Genetics**. In *An overview of mutation, polymorphism, and DNA repair*. 2nd editio. edited by Wiley-Liss New York: 1999.
30. Bao Y, Bing C, Hunter L, Jenkins JR, Wabitsch M, Trayhurn P: **Zinc-alpha2-glycoprotein, a lipid mobilizing factor, is expressed and secreted by human (SGBS) adipocytes.***FEBS letters* 2005, **579**:41–7.
31. Gong F-Y, Zhang S-J, Deng J-Y, Zhu H-J, Pan H, Li N-S, Shi Y-F: **Zinc-alpha2-glycoprotein is involved in regulation of body weight through inhibition of lipogenic enzymes in adipose tissue.***International journal of obesity (2005)* 2009, **33**:1023–30.
32. Mracek T, Ding Q, Tzanavari T, Kos K, Pinkney J, Wilding J, Trayhurn P, Bing C: **The adipokine zinc-alpha2-glycoprotein (ZAG) is downregulated with fat mass expansion in obesity.***Clinical endocrinology* 2010, **72**:334–41.

33. Beeckmann P, Schroffel J, Moser G, Bartenschlager H, Reiner G, Geldermann H: **Linkage and QTL mapping for Sus scrofa chromosome 3**. *Journal of Animal Breeding and Genetics* 2003, **120**:20–27.

34. Harmegnies N, Davin F, De Smet S, Buys N, Georges M, Coppieters W: **Results of a whole-genome quantitative trait locus scan for growth, carcass composition and meat quality in a porcine four-way cross.***Animal genetics* 2006, **37**:543–53.

35. Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson B V, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S, Helgadottir A, Ingason A, Steinthorsdottir V, Olafsdottir EJ, Olafsdottir GH, Jonsson T, Borch-Johnsen K, Hansen T, Andersen G, Jorgensen T, Pedersen O, Aben KK, Witjes JA, Swinkels DW, Den Heijer M, Franke B, Verbeek ALM, Becker DM, Yanek LR, Becker LC, et al.: **Many sequence variants affecting diversity of adult human height.***Nature genetics* 2008, **40**:609–15.

36. Karim L, Takeda H, Lin L, Druet T, Arias JAC, Baurain D, Cambisano N, Davis SR, Farnir F, Grisart B, Harris BL, Keehan MD, Littlejohn MD, Spelman RJ, Georges M, Coppieters W: **Variants modulating the expression of a chromosome domain encompassing PLAG1 influence bovine stature.***Nature genetics* 2011, **43**:405–13.

37. Hayakawa T, Sugawara T, Go Y, Udono T, Hirai H, Imai H: **Eco-geographical diversification of bitter taste receptor genes (TAS2Rs) among subspecies of chimpanzees (Pan troglodytes).***PloS one* 2012, **7**:e43277.

38. Glendinning JI: **Is the bitter rejection response always adaptive?***Physiology & behavior* 1994, **56**:1217–27.

39. Van der Waaij EH: **A resource allocation model describing consequences of artificial selection under metabolic stress.***Journal of animal science* 2004, **82**:973–81.

40. Bloemhof S, Kause A, Knol EF, Van Arendonk JAM, Misztal I: **Heat stress effects on farrowing rate in sows: genetic parameter estimation using within-line and crossbred models.***Journal of animal science* 2012, **90**:2109–19.

41. Koch MA, Thomas KR, Perdue NR, Smigiel KS, Srivastava S, Campbell DJ: **T-bet(+) Treg cells undergo abortive Th1 cell differentiation due to impaired expression of IL-12 receptor β2.***Immunity* 2012, **37**:501–10.

42. Ovilo C, Fernández A, Noguera JL, Barragán C, Letón R, Rodríguez C, Mercadé A, Alves E, Folch JM, Varona L, Toro M: **Fine mapping of porcine chromosome 6 QTL and LEPR effects on body composition in multiple generations of an Iberian by Landrace intercross.***Genetical research* 2005, **85**:57–67.

43. Muñoz G, Ovilo C, Silió L, Tomás a, Noguera JL, Rodríguez MC: **Single- and joint-population analyses of two experimental pig crosses to confirm quantitative**

trait loci on Sus scrofa chromosome 6 and leptin receptor effects on fatness and growth traits.*Journal of animal science* 2009, **87**:459–68.

44. Matsui E, Kaneko H, Fukao T, Teramoto T, Inoue R, Watanabe M, Kasahara K, Kondo N: **Mutations of the IL-12 receptor beta2 chain gene in atopic subjects.***Biochemical and biophysical research communications* 1999, **266**:551–5.
45. Poltorak A, Merlin T, Nielsen PJ, Sandra O, Smirnova I, Schupp I, Boehm T, Galanos C, Freudenberg MA: **A point mutation in the IL-12R beta 2 gene underlies the IL-12 unresponsiveness of Lps-defective C57BL/10ScCr mice.***Journal of immunology (Baltimore, Md. : 1950)* 2001, **167**:2106–11.
46. Laenoi W, Uddin MJ, Cinar MU, Grosse-Brinkhaus C, Tesfaye D, Jonas E, Scholz AM, Tholen E, Looft C, Wimmers K, Phatsara C, Juengst H, Sauerwein H, Mielenz M, Schellander K: **Quantitative trait loci analysis for leg weakness-related traits in a Duroc × Pietrain crossbred population.***Genetics, selection, evolution : GSE* 2011, **43**:13.
47. Uemoto Y, Sato S, Ohnishi C, Hirose K, Kameyama K, Fukawa K, Kudo O, Kobayashi E: **Quantitative trait loci for leg weakness traits in a Landrace purebred population.***Animal science journal = Nihon chikusan Gakkaihō* 2010, **81**:28–33.
48. Mihci E, Türkkahraman D, Ellard S, Akçurin S, Bircan I: **Wolcott-Rallison syndrome due to a novel mutation (R491X) in EIF2AK3 gene.***Journal of clinical research in pediatric endocrinology* 2012, **4**:101–3.
49. Arnbjerg J: **Effect of a low-growth rate on the frequency of osteochondrosis in Danish Landrace pigs (short communication)**. *Arch. Tierz., Dummerstorf* 2007, **50**:105–111.
50. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The Sequence Alignment/Map format and SAMtools.***Bioinformatics (Oxford, England)* 2009, **25**:2078–9.
51. Wang K, Li M, Hakonarson H: **ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data.***Nucleic acids research* 2010, **38**:e164.
52. Haider S, Ballester B, Smedley D, Zhang J, Rice P, Kasprzyk A: **BioMart Central Portal--unified access to biological data.***Nucleic acids research* 2009, **37**:W23–7.
53. Rousset F: **genepop'007: a complete re-implementation of the genepop software for Windows and Linux.***Molecular ecology resources* 2008, **8**:103–6.
54. Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA: **Circos: an information aesthetic for comparative genomics.***Genome research* 2009, **19**:1639–45.

55. Tortereau F, Servin B, Frantz L, Megens H-J, Milan D, Rohrer G, Wiedmann R, Beever J, Archibald AL, Schook LB, Groenen MA: **A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content.***BMC genomics* 2012, **13**:586.

6

General discussion

6.1 Introduction

A basic assumption in conservation genetics is that genetic diversity is directly and positively related with population size [1]. The effect of genetic drift is strong in isolated and small populations, leading to a loss of alleles and resulting in inbreeding via the fixation of deleterious recessive alleles [2]. The loss of genetic variation may result in lower reproduction rates and losses in the capacity of populations to adapt to changing environments [3]. The genetic characterization of populations provide a knowledge base for population management strategies within conservation programs [4]. Advances in molecular genetics and data analysis have increased the number of available neutral markers, providing tools to assess population diversity and structure at a higher accuracy than ever before [5]. Recently, the availability of complete genome sequences makes it possible to study functional genomics and thus, explore the genetic basis of local adaptation and inbreeding depression.

By integrating nuclear and mitochondrial molecular data, I reported the effect of past demographic events -genetic drift, migration, bottlenecks and crossbreeding- on the genetic variation of pig populations. By analyzing a wide representation of pigs that inhabit different environments and have been subjected to different selection histories, it was possible to make comparisons between populations and interpret differences. In the previous chapters different marker systems and statistical methods have been comprehensively integrated to address important questions for the conservation and rational exploitation of domestic and wild pig populations. In this chapter, the main findings and the practical relevance of the results presented in this thesis are discussed.

6.2 Genetic marker systems and demographic history of pig populations

Historical records on populations or breeds often provide inaccurate or biased accounts, but can provide hypotheses that can be tested with genetic studies [6]. Genetic data represents an objective tool to reliably assess population structure, effective population size, migration and admixture (e.g. [4, 7, 8]). Here, my goal was to obtain insight in population genetic processes and demographic history based on mitochondrial DNA (mtDNA), microsatellites, high-density SNP data and Next Generation Sequence (NGS) data.

Population structure

Population structure analysis allows the assignment of the individuals to their population of origin [9], the identification of admixed individuals [10] and the

traceability of meat products in livestock [11]. Within conservation genetics, population structure analyses aid to identify those populations that deserve special attention for conservation purposes [4, 12, 13]. In this thesis, Bayesian methods [14], F_{st} estimations [15], genetic distances and Principal Component Analysis (PCA) were used to assess population structure and substructure of pig populations.

Generally, it is assumed that the availability of a higher number of genetic markers implies higher precision in population structure analysis. Nevertheless, I observed a high agreement between the structure analysis computed with 34 microsatellites and with ~48,000 SNPs (Chapter 3) in the Spanish pig Chato Murciano. In agreement with these findings, microsatellites have shown their usefulness to discern between closely related populations within the same breed [16].

High-density SNP panels suffer from ascertainment bias, distorting population structure inferences in e.g. human populations [17]. Nevertheless, the results presented in this thesis indicate that the Illumina Porcine 60K Beadchip can be reliably used to assign pigs to their population of origin, a finding in agreement with other studies [11]. The results presented in this thesis are particularly noteworthy since a large variety of populations including commercial, wild and local breeds from multiple European regions could be clearly differentiated. The Bayesian structure analysis implemented in the software Structure [14] allowed clustering the individuals to their population of origin. This study at individual level is interesting since it makes it possible to detect potential sampling errors and admixed individuals. For example, in Chapter 4 two Iberian pigs belonging to different variants where not correctly assigned to their putative population of origin, possibly due to a sampling error. Regarding admixture, Manchado de Jabugo and Chato Murciano showed signs of admixture with other pigs in agreement with historical records.

Effective population size

The estimation of effective population size (N_e) from genetic data [18] is of major interest in conservation genetics. Since wild and local populations often have limited, or even no pedigree information available, genetic data is a suitable tool to assess past and present N_e [8, 19, 20]. In Chapter 4, 60K SNP data was used to estimate present and past N_e based on Linkage Disequilibrium (LD), in conjunction with a recombination map of the pig [21]. Previous studies in livestock species used an approximation of 1 Mb = 1cM to compute N_e, but this assumption may lead to incorrect estimates due to the fact that recombination rate varies substantially across and within porcine chromosomes [22]. The results showed good agreement with historical records of the populations and other genetic parameters. Thus, the populations Manchado de Jabugo and Chato Murciano are those categorized as

endangered with extinction according to the records of the Food and Agriculture Organization of the United Nations(FAO; http://dad.fao.org/) showed indeed the lowest current N_e.

It has been stated that LD is highly affected by sample size. Since the estimation of N_e depends on LD [18], the use of a low sample size is expected to decrease the accuracy of N_e estimations. Nevertheless, I observed that populations with low sample size ($N < 15$) showed N_e values consistent with genomic diversity estimations and with historical records, despite the standard deviation being high. The study of the 60K SNP data also allowed us to obtain insights in past population size. The results presented were consistent with known events such as the decrease of N_e in wild pigs during the Last Glacial Maximum. Groenen et al. [23] described a decrease in past N_e of European wild boar using whole-genome sequence data similar to the decrease observed in wild Iberian pigs (Chapter 4). I also observed an increase in population size around 4,000 generations ago exclusively in the Iberian pigs, highly consistent with the domestication timeframe.

Admixture or Crossbreeding

Crossbreeding or admixture refers to the interbreeding between different populations, breeds or varieties. Admixture has been important in shaping genome variation of domestic animals [6]. Despite the fact that new production systems largely avoid crossbreeding between domestic and wild pigs, the existence of wild boars with domestic introgression has been confirmed using nuclear [24] and mitochondrial [25, 26] data. The study of mtDNA haplotypes showed Asian wild pigs from Japan and Taiwan carrying mtDNA haplotypes that are common in European domestic pigs. Likewise wild pigs from Italy and the Netherlands among others carry haplotypes of Asian origin. With regard to domestic pigs, I observed a high frequency of mtDNA haplotypes of Asian origin in English and commercial breeds, which is in agreement with the deliberate introgression of Asian pigs into European pig populations in the late 18th and 19th centuries [23, 25, 27]. Around 25-50% of the pigs belonging to the commercial breeds Pietrain, Large White and Landrace, and 20-90% of English local pigs – depending on the breed- carried Asian haplotypes [28]. Asian haplotypes were also found in European local pigs such as Manchado de Jabugo, and, conversely, haplotypes of European origin were observed, mostly at low frequencies, in Asian breeds. Thus, introgression between pig populations was corroborated using mtDNA. The use of mtDNA to detect crossbreeding has, however, major disadvantages [6]. The fact that mtDNA represents a single locus that is maternally inherited hampers the detection of crossbreeding from the paternal side. This is particularly relevant for a species like the pig, which, in the wild, displays pronounced male-driven dispersal.

Furthermore, in pig trading and breeding, both male and female biased practices may occur. An additional drawback of mtDNA is that timeframe and degree of admixture cannot be precisely determined since mtDNA does not recombine.
The analysis of nuclear DNA corroborated crossbreeding in various pig populations with recorded histories pigs. In Chapter 3, the population structure analysis using 60K SNP data revealed that some Chato Murciano pigs had been introgressed with Duroc pigs in a farm where Duroc and Chato Murciano are reared together. This demonstrates that genetic structure analysis is a suitable tool to infer recent introgression based on the 60K SNP panel. Indirect signs of introgression were detected by the joint analysis of regions of homozygosity (ROH) and inbreeding factor (F) in each animal. Indeed, the four Chato Murciano pigs recently introgressed with Duroc pigs, as discussed in Chapter 3, had a characteristic pattern consisting of short ROH and a low inbreeding. In chapter 4 I observed other local pigs with the same pattern that probably indicates similar introgression events. Goedbloed et al. [24] also observed higher levels of genetic diversity in wild pigs that had been crossed with domestic pigs.

6.3 Genetic marker systems and genetic diversity of pig populations

In this thesis, genetic diversity was estimated using microsatellites, a high-density SNP panel and NGS data. The analysis of 34 microsatellites in the Chato Murciano pigs showed that different farms can have different levels of genetic diversity. These estimations were in good agreement with documented reports regarding the history of the breed and the census of the farms. Estimation of genetic diversity with high-density SNP data was also in agreement with the expected results regarding the history of other breeds. For example, populations with well-known high level of inbreeding like Mangalica had very low genetic diversity. Contrary to the microsatellite data, the high number and even distribution of SNPs using 60K SNP data allowed the estimation of parameters that rely on genome-wide information such as ROH. While the expected heterozygosity is computed at the population level, ROH and also inbreeding factor were calculated in each pig separately, being well correlated with genetic diversity estimates [21]. The individual assessment of genetic diversity has practical implications in a conservation program. We may prioritize breeders that display low inbreeding factor and proportion of genome covered by ROH.

High-density SNP chips are designed on the basis of the genetic variability between relatively small numbers of populations. Furthermore, such SNP discovery panels are usually focusing on a population of prior interest, such as an experimental population or commercially relevant breeds. In pigs, the Porcine SNP60 Beadchip was designed using the breeds Duroc, Large White, Landrace and Pietrain [29]. Although some wild boar information was added to the SNP discovery process, the contribution was very small. The bias originating from the SNP discovery process, also referred as the ascertainment bias, results in under estimation of variation in populations as they are more dissimilar to the SNP discovery panel. Nevertheless, the correlation between genetic diversity computed with the 60K SNP and NGS data was high when the same individual animals were both genotyped and sequenced and also when averaged estimations across each population were compared. These results indicate that the Porcine 60SNP Beadchip provides reliable estimates of genomic diversity for comparative studies between European pig populations despite the inherent bias. However, the inability of the Porcine SNP60 Beadchip to detect part of the genetic diversity in English breeds and also local breeds crossed with English breeds demonstrates the strength of applying whole genome sequencing in conservation studies.

6.4 Genomic variation within pig populations

Since the first pigs were domesticated around 10,000 years ago, selection, in particular artificial selection, for traits such as coat colour, body size, reproduction and behaviour have resulted in hundreds of breeds worldwide [30]. Breeding practices developed in England in the 18^{th} – 19^{th} century resulted in rapid phenotypic changes with the emergence of new improved breeds [27]. Some of the English breeds that emerged from crosses with Asian pigs, were subsequently crossed with traditional breeds [27, 31]. Advances in the livestock industry during the last 50-60 years have accelerated phenotypic change between a hand-full of widely used, highly productive, commercial breeds, and the local breeds. These local breeds have largely escaped such intense selection and, as a consequence, generally show low performance in production traits. However, while commercial pigs have seen a large increase in production traits, they may, simultaneously, have become more sensitive to housing system, food quality, climate and disease [32, 33]. In wild populations, pigs have not been subjected to domestication events but the human influence on this populations still has been remarkable due to hunting and restocking activities, together with crossbreeding with domestic pigs [7, 24, 26]. These disparities in selection pressures are expected to affect the genomic

variability between pig populations. Indeed, genomic regions under selection due to domestication –based on a comparison between wild and domestic pigs– and with breed development –based on a comparison between domestic pig breeds– have been recently pinpointed [34, 35]. Rubin et al. [35] observed signatures of selection at loci related to the number of vertebrae and elongation of the back [27]. Wilkinson et. al. [34] reported evidence for genomic differences between European breeds associated with genes involved in breed standard characteristics such as coat colour and certain production traits. In Chapter 5, the genomic variation underlying phenotypic differences between local and commercial European breeds was explored. The differentiation of allele frequencies among populations is considered to be an appropriate approach to detect genomic regions under selection for local adaptation [5, 36]. The study of non-synonymous SNPs that may have phenotypic consequences, revealed candidate SNPs with a high allele frequency difference between local breeds and commercial pigs (F_{st}=1).

These data demonstrate that local breeds harbour different genomic variants than commercial pigs, even though most of the local breeds did show lower genetic diversity overall. Considering the eight commercial pigs of four different breeds as a single population, I observed 99 homozygous non-synonymous sites for which some local pigs were heterozygous or homozygous for both possible alleles, while all commercial pigs carried the same allele. The lack of variation at certain genomic regions despite the high genetic diversity found in commercial breeds could have resulted from different selection pressures between commercial and local pigs. Artificial selection may have reduced diversity at loci under selection, as well as the neighbouring linked sites [37]. In local pigs, population-wide homozygosity may have been induced due to genetic drift, particularly in those breeds that have become highly marginalized.

Genetic drift effects, however, are likely also important in commercial populations, particularly in boar lines that are generally kept at effective population sizes of a few tens of animals[38]. Low effective population size has been related to the occurrence and maintenance of recessive damaging mutations in livestock species [39]. Thus, disadvantageous genotypes can easily spread in the population due to (partially) recessive genotypes carried by a founder. For example, missense mutations within the gene encoding bovine CD18 and the gene SLC35A3 were associated to immune deficiency and vertebral malformation, respectively, in Holstein-Friesian bulls. These mutations were widely disseminated in the population by a founder used extensively in the Holstein-Friesian breed [40, 41].

According to Allendorf et al. [5], the application of genomics to conservation will allow the identification of loci related with fitness and demographic vital rates, and

thereby, the prediction of population's viability or capacity to adaptation. This novel approach may be incorporated to the conservation management of the populations. Several mutations detailed in Chapter 5 were found in genes that may be involved in sensory perception such as the taste receptor *TAS2R40* [42] or the gene *USH1C* related to hearing function [43]. Other mutations affect genes with immune related and skeletal development functions. These non-synonymous mutations could reflect differences in the adaptation between local and commercial animals. Furthermore, given that many of these mutations –particularly within immune-related genes– were classified in Polyphen analysis as "benign" for the fixed allele carried by local pigs, one may argue advantageous selection in local pigs for harsher environment. However, the lack of detailed phenotypic information of the animals genotyped, and having little information on the environment and artificial selection pressure, preclude accurate interpretation. As presented in Chapter 1, the access to the whole genome sequence led us to a reverse genetics approach –a gene sequence is known, but the exact function is unknown–. Given that the genes presented here can potentially be involved in a better adaptation of local pigs to a harsher environment, further studies are needed to unravel the phenotypic effect of these mutations.

Interestingly, for most of the fixed variants observed in commercial and local pigs, wild populations also preserved both, ancestral and derived alleles. On one hand, this indicates that the majority of the variation presented in domestic pigs is not new but derived from the wild ancestor. On the other hand, it implies that local breeds preserve genetic variants from wild populations that could eventually be used to increase the genomic diversity of commercial animals, highlighting the convenience of conserving local pig populations.

6.5 Phylogeographic diffusion of *Sus scrofa* using mtDNA

Migration implies movement of alleles between populations and thereby can be responsible for changes in allele frequencies [44]. The high reproduction rate and adaptability to different environments of wild boar populations [45] can result in a high dispersal rate. Thus, the study of migration patterns of wild populations is of importance to understand the current genetic structure of *Sus scrofa* worldwide.

Inferences of migration pathways in *Sus scrofa* using genetic data have traditionally been based on the topology of phylogenetic trees or networks, estimated with standard methods that do not explicitly take spatial information into account. In Chapter 1, the migration patterns of wild pigs throughout Eurasia were inferred using a Bayesian approach. This Bayesian method used an asymmetric discrete

diffusion model as implemented in BEAST with a stochastic search variable selection procedure that identifies the parsimonious descriptions of the diffusion process that includes prior geographic information [46] to unravel historical migration patterns. Most of the significant and consistent migration pathways described in Chapter 2 are consistent with fossil records, or with migrations described for other animal species. For instance, the re-colonization of central Europe from the so-called refugia was strongly supported by our analysis and corroborates previous studies [47]. In addition, recently postulated migration events of *Sus scrofa* were apparent also from the mtDNA analysis. Larson et al. [48] noticed the existence of two divergent lineages of mtDNA in South Asia. A hitherto untested migration route from Siberia to India west of the Himalaya would explain this observation. We also found evidence for the migration of wild pigs from North East Asia to Japan through the island of Sakhalin that so far has not been described in pigs, but which was postulated based on fossil evidence [49, 50]. Recently, whole genome sequence data analysis revealed that north Asian pigs migrated to Europe 1.6-0.8 million years ago [23]. The mtDNA analysis, however, does not offer a well-supported migration pattern from Asia to Europe highlighting the apparent limitation of mtDNA to infer this migration event. Furthermore, the poor representation of samples from west Asia could also explain the absence of a significant link between Asia and Europe.

The Bayesian phylogeographic analysis performed in this study has shown its robustness to infer historical migration events in bears [51], beetles [52, 53] and viruses [46, 54]. Despite the fact that our work shed light on migration patterns of *Sus scrofa* dating, at least, of middle-late Pleistocene, we must point out limitations when applied in species with complex demographic history such as *Sus scrofa*. First, human mediated events, particularly domestication, may have a large influence on the results [6] as was observed when the analysis performed in SET_1 and SET_2 were compared. Second, the resampling procedure used identified “key samples” that determined well-supported migration pathways, implying that a representative set of samples is required. Third, an inappropriate delineation of geographic regions will lead in a loss of accuracy in the conclusions. Testing migration between restricted and well-delimited geographic regions may be desirable in particular geographic regions.

6.6 Genetic characterization of local breeds. Insights for conservation genetics

Conservation management of any population requires in depth knowledge of biogeography, population structure and genetic diversity. To prevent the permanent loss of our genetic and cultural heritage, local breeds need to be managed by conservation programs. While governments and farmer associations generally subscribe to this need, they usually lack funds to include the necessary genetic characterization and monitoring of local breeds. Therefore, an ideal conservation program based on genetic data to be implemented in low input populations must find the best trade-off between accuracy and economic affordability. At this point major questions may arise: which marker systems are appropriate for the genetic characterization of local livestock breeds? which analysis better explains the idiosyncrasy of the population? which individuals or populations are worth to prioritize within the conservation program?.

The study of mtDNA aided to unravel parts of the demographic history of *Sus scrofa*. However, major questions in conservation genetics such as genetic diversity and population structure could not be accurately assessed using mitochondrial data. Microsatellites and High-density SNP panels showed high reliability to perform population structure analysis and to estimate genetic diversity. High-density SNP panels in particular are highly suitable for conservation genetics purposes. The cost of genotyping a high number of microsatellites is lower than genotype animals with high-density SNP panels although the possibility to automate SNPs will provide steadily higher panels to be used at low cost. In this thesis, the cost for genotyping 35 microsatellites and the 60K SNP panel were ~30€ and ~100€ respectively per animal. However, the high number of SNP markers widely distributed across the genome allowed the estimation, not only of genetic diversity and population structure, but also other highly informative parameters. For instance, high accuracy in the estimation of LD, N_e and ROH can only be obtained with high-density marker systems. Therefore, genotyping a smaller number of individuals with a higher number of markers can result in higher accuracy of the parameters estimated. Another important disadvantage of microsatellites is the complexity of comparative studies across laboratories, while results obtained from commercially available SNP panels can be readily compared.

Even though, for a fixed budget, SNP panels offer more accurate estimates of relevant parameters, the sampling of animals becomes very important. This is due to the fact that a smaller number of animals can be assayed for the same budget. In pigs, as in other livestock species, sampling must focus on the founders whose

genetic material is spread across the population. Typically, livestock breeds are reared in separate farms. The exchange of animals between farms can vary from total isolation to a continuous flow which may have consequences on the genetic patterns of the population. In Chapter 3, the local breed Chato Murciano was studied in detail. This endangered breed has been affected by inbreeding and crossbreeding, which are among the most important threats to local livestock breeds [30]. This breed, therefore, represents a good model for conservation genetic studies, further enhanced by a) the possibility of genetically characterizing the majority of the breeding animals and b) the low number of farms (n=15) where these pigs are bred. The estimation of population structure and genetic diversity of Chato Murciano revealed relatively high genetic diversity of the whole population, but also a strong substructure by farms. Several farms showed high levels of inbreeding. Similar results have been observed in other studies [55] revealing that in a situation of strong substructure of the population, genetic diversity estimates must be assessed in each subpopulation separately [56]. Population genetic studies generally consider each breed as a single unit. While this may be appropriate to study wild populations of the same geographic regions, this is completely inappropriate for endangered breeds such as Chato Murciano. Chapter 3 highlights the necessity to consider farms separately and assess the substructure within the breeds as an initial step to characterize livestock populations.

This thesis demonstrates that estimations of structure, inbreeding and ROH at the individual level using genome-wide SNP assays allows the detection of animals with a high level of inbreeding and of recent admixture. Admixture or crossbreeding is a major issue in conservation genetics [5]. Crossbreeding between different divergent populations may be advantageous due to the increase of genetic variation [57]. In fact, fragmentation of a population into subpopulations has been proposed as a strategy to manage populations, as long as a minimum number of migrations between subpopulations is carried out to maximize the genetic diversity [58]. This would be the case for the Chato Murciano breed, were gene flow between farms may reduce the erosion of genetic diversity of the breed as a whole. Crossbreeding between populations can also be undesirable, for instance, when results in out-breeding depression reduces the fitness of the population as may be the case between domesticated and wild animals [57]. Crossbreeding furthermore poses a threats to the integrity of the gene pool of a valuable population [59], and thereby to the perceived heritage value of the breed. Detection of recent admixture is of particular interest for those livestock breeds whose products are highly priced and appreciated by the consumer, such as for example Iberian pig to produce the famous Iberico ham [60]. Furthermore, information on admixture can

be used to prioritise pure breed animals or discard hybrid individuals when so dictated by the management program. In must be considered that the admixed animals are likely to have far higher levels of heterozygosity (as demonstrated in Chapters 3 and 4 of this thesis), and therefore, discarding them from the breeding program effectively increases the overall rate of inbreeding.

An ascertainment-bias free estimate of genetic diversity parameters and ROH can be made from whole-genome sequences as showed in Chapter 5. Moreover, NGS offers the possibility to carry out functional genomics approaches by identifying loci involved in adaptation to local environments [5] and deleterious alleles related to inbreeding depression [39]. This is of particularly interest for local breeds since they may have been adapted to harsh environments. At the same time, prolonged inbreeding in such populations may have led to an accumulation of relatively high-frequency deleterious alleles as a consequence of drift effects as observed in Chapter 5 [30]. Despite the advantages of using genome-wide data sets, the cost per pig is still high – ~1500€ for a 10-12x coverage at the time of writing – and the budget available to characterize a low-input population typically will allow to re-sequence a few individuals, at most. In this thesis, one or two animals per breed were re-sequenced while on average around 20 pigs of the same population were genotyped using the 60K SNP panel. Valuable conclusions can be obtained from these analyses. First, the high correlation between 60K SNP and NGS estimates points out the usefulness of 60K SNP data to infer genomic diversity in spite of the ascertainment bias implicit in high-density SNP chips [17], as discussed in section 6.2. Second, genetic diversity estimations from two re-sequenced individuals were highly correlated with the estimations obtained from the analysis of ~20 individuals genotyped with the 60K SNP panel. This shows that a small number of individuals that is re-sequenced in comparison with high-density SNP, is compensated by the very larger number of genomic sites studied [61]. Therefore, the selection of population-representative animals to be re-sequenced is a major issue, especially when the available budget only allows sequencing a small number of individuals. The results in this thesis clearly show that implementation of NGS to characterize local population adds valuable information to the high-density SNP marker systems. Nevertheless, due to the high cost and analysis and computational challenges, nowadays NGS is not yet a suitable tool for long-term genetic monitoring of local populations [5, 62], although this may change rapidly in the coming years as price of sequencing is projected to drop further.

6.6 Conclusions

This thesis provided both theoretical and practical insights for addressing the rational management and exploitation of local genetic resources. Those insights stress the need to assess genetic diversity at all possible levels, from breed, to subpopulations and to individual. At the same time, 60K SNP data allows discarding animals that may not represent the population very well due to recent crossbreeding and aids in prioritizing founders with lower level of inbreeding. NGS data of at least two representative animals of the population can provide additional information that cannot be accurately obtained using other marker systems. For example, NGS enables functional genomics for adaptation to the environment and assessment of genetic load due to inbreeding. Finally, the large variety of populations, marker systems, and analyses to assess genomic diversity described, represents a valuable source of information for comparative purposes. Future studies that aim to assess genomic diversity in livestock species will be able to perform direct comparisons with the results presented in this thesis.

References

1. Ouborg NJ: **Integrating population genetics and conservation biology in the era of genomics.***Biology letters* 2010, **6**:3–6.
2. Lynch M: **The Genetic Interpretation of Inbreeding Depression and Outbreeding Depression**. *Evolution* 1991, **45**:622–629.
3. Keller L: **Inbreeding effects in wild populations**. *Trends in Ecology & Evolution* 2002, **17**:230–241.
4. Bowden R, MacFie TS, Myers S, Hellenthal G, Nerrienet E, Bontrop RE, Freeman C, Donnelly P, Mundy NI: **Genomic tools for evolution and conservation in the chimpanzee: Pan troglodytes ellioti is a genetically distinct population.***PLoS genetics* 2012, **8**:e1002504.
5. Allendorf FW, Hohenlohe P a, Luikart G: **Genomics and the future of conservation genetics.***Nature reviews. Genetics* 2010, **11**:697–709.
6. Larson G, Burger J: **A population genetics view of animal domestication.***Trends in genetics : TIG* 2013, **29**:197–205.
7. Scandura M, Iacolina L, Apollonio M: **Genetic diversity in the European wild boar Sus scrofa: phylogeography, population structure and wild x domestic hybridization**. *Mammal Review* 2011, **41**:125–137.
8. Tenesa A, Navarro P, Hayes BJ, Duffy DL, Clarke GM, Goddard ME, Visscher PM: **Recent human effective population size estimated from linkage disequilibrium.***Genome research* 2007, **17**:520–6.
9. Negrini R, Nicoloso L, Crepaldi P, Milanesi E, Colli L, Chegdani F, Pariset L, Dunner S, Leveziel H, Williams JL, Ajmone Marsan P: **Assessing SNP markers for assigning individuals to cattle populations.***Animal genetics* 2009, **40**:18–26.
10. Herrero-Medrano JM, Megens HJ, Crooijmans RP, Abellaneda JM, Ramis G: **Farm-by-farm analysis of microsatellite, mtDNA and SNP genotype data reveals inbreeding and crossbreeding as threats to the survival of a native Spanish pig breed.***Animal genetics* 2012, **In press**.
11. Ramos AM, Megens HJ, Crooijmans RPMA, Schook LB, Groenen MAM: **Identification of high utility SNPs for population assignment and traceability purposes in the pig using high-throughput sequencing.***Animal genetics* 2011, **42**:613–20.
12. Scandura M, Iacolina L, Cossu A, Apollonio M: **Effects of human perturbation on the genetic make-up of an island population: the case of the Sardinian wild boar.***Heredity* 2011, **106**:1012–20.

13. Fabuel E, Barragán C, Silió L, Rodríguez MC, Toro M a: **Analysis of genetic diversity and conservation priorities in Iberian pigs based on microsatellite markers.***Heredity* 2004, **93**:104–13.
14. Pritchard JK, Stephens M, Donnelly P: **Inference of population structure using multilocus genotype data.***Genetics* 2000, **155**:945–59.
15. Weir BS, Cockerham C. C: **Estimating F-Statistics for the Analysis of Population Structure**. *Evolution* 1984, **38**:1358–1370.
16. Martínez a M, Delgado J V, Rodero a, Vega-Pla JL: **Genetic structure of the Iberian pig breed using microsatellites.***Animal genetics* 2000, **31**:295–301.
17. Albrechtsen A, Nielsen FC, Nielsen R: **Ascertainment Biases in SNP Chips Affect Measures of Population Divergence Research article**. 2010, **27**:2534–2547.
18. Hayes BJ, Visscher PM, McPartlan HC, Goddard ME: **Novel multilocus measure of linkage disequilibrium to estimate past effective population size.***Genome research* 2003, **13**:635–43.
19. Flury C, Tapio M, Sonstegard T, Drögemüller C, Leeb T, Simianer H, Hanotte O, Rieder S: **Effective population size of an indigenous Swiss cattle breed estimated from linkage disequilibrium.***Journal of animal breeding and genetics = Zeitschrift für Tierzüchtung und Züchtungsbiologie* 2010, **127**:339–47.
20. Uimari P, Tapio M: **Extent of linkage disequilibrium and effective population size in Finnish Landrace and Finnish Yorkshire pig breeds.***Journal of animal science* 2011, **89**:609–14.
21. Bosse M, Megens H-J, Madsen O, Paudel Y, Frantz L a. F, Schook LB, Crooijmans RPM a., Groenen M a. M: **Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape**. *PLoS Genetics* 2012, **8**:e1003100.
22. Tortereau F, Servin B, Frantz L, Megens H-J, Milan D, Rohrer G, Wiedmann R, Beever J, Archibald AL, Schook LB, Groenen MA: **A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content.***BMC genomics* 2012, **13**:586.
23. Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-J, Li S, Larkin DM, Kim H, Frantz LAF, Caccamo M, Ahn H, Aken BL, Anselmo A, Anthon C, Auvil L, Badaoui B, Beattie CW, Bendixen C, Berman D, Blecha F, Blomberg J, Bolund L, Bosse M, Botti S, Bujie Z, et al.: **Analyses of pig genomes provide insight into porcine demography and evolution.***Nature* 2012, **491**:393–8.
24. Goedbloed DJ, Megens HJ, Van Hooft P, Herrero-Medrano JM, Lutz W, Alexandri P, Crooijmans RPM a, Groenen M, Van Wieren SE, Ydenberg RC, Prins HHT: **Genome-wide single nucleotide polymorphism analysis reveals recent**

genetic introgression from domestic pigs into Northwest European wild boar populations.*Molecular ecology* 2012, **22**:856–866.

25. Giuffra E, Kijas JM, Amarger V, Carlborg O, Jeon JT, Andersson L: **The origin of the domestic pig: independent domestication and subsequent introgression.***Genetics* 2000, **154**:1785–91.
26. Scandura M, Iacolina L, Crestanello B, Pecchioli E, Di Benedetto MF, Russo V, Davoli R, Apollonio M, Bertorelle G: **Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable?***Molecular ecology* 2008, **17**:1745–62.
27. Darwin C: *The Variation of Animals and Plants under Domestication*. John Murry. London: 1868.
28. Fang M, Andersson L: **Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication.***Proceedings. Biological sciences / The Royal Society* 2006, **273**:1803–10.
29. Ramos AM, Crooijmans RPM a, Affara N a, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu Z-L, Kerstens HH, Law AS, Megens H-J, Milan D, Nonneman DJ, Rohrer G a, Rothschild MF, Smith TPL, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen M: **Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology.***PloS one* 2009, **4**:e6524.
30. FAO: **Global Plan of Action for Animal Genetic Resources and the Interlaken Declaration.***(available at http://www.fao.org/docrep/010/a1404e/a1404e00.htm).* 2007.
31. Porter V: *Pigs. A Handbook to the Breeds of the World*. Mountfield, East Sussex: Helm Information, Ltd.; 1993.
32. Van der Waaij EH: **A resource allocation model describing consequences of artificial selection under metabolic stress.***Journal of animal science* 2004, **82**:973–81.
33. Bloemhof S, Kause A, Knol EF, Van Arendonk JAM, Misztal I: **Heat stress effects on farrowing rate in sows: genetic parameter estimation using within-line and crossbred models.***Journal of animal science* 2012, **90**:2109–19.
34. Wilkinson S, Lu ZH, Megens H-J, Archibald AL, Haley C, Jackson IJ, Groenen M a. M, Crooijmans RPM a., Ogden R, Wiener P: **Signatures of Diversifying Selection in European Pig Breeds**. *PLoS Genetics* 2013, **9**:e1003453.
35. Rubin C-JC-J, Megens H-JH-J, Barrio AM, Maqbool K, Sayyab S, Schwochow D, Wang C, Carlborg O, Jern P, Jorgensen CB, Archibald AL, Fredholm M, Groenen

MAM, Andersson L, Martinez Barrio A, Carlborg Ö, Jørgensen CB: **Strong signatures of selection in the domestic pig genome.***Proceedings of the National Academy of Sciences of the United States of America* 2012, **109**:19529–36.

36. Storz JFAYF: **Using genome scans of DNA polymorphism to infer adaptive population divergence.***Molecular ecology* 2005, **14**:671–88.
37. Tenaillon MI, U'Ren J, Tenaillon O, Gaut BS: **Selection versus demography: a multilocus investigation of the domestication process in maize.***Molecular biology and evolution* 2004, **21**:1214–25.
38. Welsh CS, Stewart TS, Schwab C, Blackburn HD: **Pedigree analysis of 5 swine breeds in the United States and the implications for genetic conservation.***Journal of animal science* 2010, **88**:1610–8.
39. Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, Carta E, Dardano S, Dive M, Fasquelle C, Frennet J-C, Hanset R, Hubin X, Jorgensen C, Karim L, Kent M, Harvey K, Pearce BR, Simon P, Tama N, Nie H, Vandeputte S, Lien S, Longeri M, Fredholm M, Harvey RJ, Georges M: **Highly effective SNP-based association mapping and management of recessive defects in livestock.***Nature genetics* 2008, **40**:449–54.
40. Shuster DE, Kehrli ME, Ackermann MR, Gilbert RO: **Identification and prevalence of a genetic defect that causes leukocyte adhesion deficiency in Holstein cattle.***Proceedings of the National Academy of Sciences of the United States of America* 1992, **89**:9225–9.
41. Thomsen B, Horn P, Panitz F, Bendixen E, Petersen AH, Holm L-E, Nielsen VH, Agerholm JS, Arnbjerg J, Bendixen C: **A missense mutation in the bovine SLC35A3 gene, encoding a UDP-N-acetylglucosamine transporter, causes complex vertebral malformation.***Genome research* 2006, **16**:97–105.
42. Go Y, Satta Y, Takenaka O, Takahata N: **Lineage-specific loss of function of bitter taste receptor genes in humans and nonhuman primates.***Genetics* 2005, **170**:313–26.
43. Lentz JJ, Jodelka FM, Hinrich AJ, McCaffrey KE, Farris HE, Spalitta MJ, Bazan NG, Duelli DM, Rigo F, Hastings ML: **Rescue of hearing and vestibular function by antisense oligonucleotides in a mouse model of human deafness.***Nature medicine* 2013, **19**:345–50.
44. Star B, Spencer HG: **Effects of genetic drift and gene flow on the selective maintenance of genetic variation.***Genetics* 2013, **194**:235–44.
45. Bieber C, Ruf T: **Population dynamics in wild boar Sus scrofa: ecology, elasticity of growth rate and implications for the management of pulsed resource consumers**. *Journal of Applied Ecology* 2005, **42**:1203–1213.

46. Lemey P, Rambaut A, Drummond AJ, Suchard M a: **Bayesian phylogeography finds its roots.***PLoS computational biology* 2009, **5**:e1000520.

47. Hewitt G: **The genetic legacy of the Quaternary ice ages.***Nature* 2000, **405**:907–13.

48. Larson G, Cucchi T, Fujita M, Matisoo-Smith E, Robins J, Anderson A, Rolett B, Spriggs M, Dolman G, Kim T-H, Thuy NTD, Randi E, Doherty M, Due RA, Bollt R, Djubiantono T, Griffin B, Intoh M, Keane E, Kirch P, Li K-T, Morwood M, Pedriña LM, Piper PJ, Rabett RJ, Shooter P, Van den Bergh G, West E, Wickler S, Yuan J, et al.: **Phylogeny and ancient DNA of Sus provides insights into neolithic expansion in Island Southeast Asia and Oceania.***Proceedings of the National Academy of Sciences of the United States of America* 2007, **104**:4834–9.

49. Kawamura Y: **Immigration of mammals into the Japanese islands during the Quaternary**. *The Quaternary Research* 1998, **37**:251–257. [in Japanese with English abstract].

50. Takahashi K: **Nihon Retto no Senshin-Koshinsei ni okeru rikusei honyu dobutsu-so no keisei katei (The formative history of the terrestrial mamalian fauna of the Japanese Islands during the Plio-Pleistocene).***Kyusekki Kenkyu (Palaeolithic Research)* 2007, **3**:5–14.

51. Edwards CJ, Suchard M a, Lemey P, Welch JJ, Barnes I, Fulton TL, Barnett R, Connell TCO, Coxon P, Monaghan N, Valdiosera CE, Lorenzen ED, Willerslev E, Baryshnikov GF, Rambaut A, Thomas MG, Bradley DG, Shapiro B, Tables S, O'Connell TC: **Ancient hybridization and an Irish origin for the modern polar bear matriline.***Current biology : CB* 2011, **21**:1251–8.

52. Marske KA, Leschen R a B, Buckley TR: **Reconciling phylogeography and ecological niche models for New Zealand beetles: Looking beyond glacial refugia.***Molecular phylogenetics and evolution* 2011, **59**:89–102.

53. Marske KA, Leschen RAB, Buckley TR: **Concerted versus independent evolution and the search for multiple refugia: comparative phylogeography of four forest beetles.***Evolution; international journal of organic evolution* 2012, **66**:1862–77.

54. De Bruyn A, Villemot J, Lefeuvre P, Villar E, Hoareau M, Harimalala M, Abdoul-Karime AL, Abdou-Chakour C, Reynaud B, Harkins GW, Varsani A, Martin DP, Lett J-M: **East African cassava mosaic-like viruses from Africa to Indian ocean islands: molecular diversity, evolutionary history and geographical dissemination of a bipartite begomovirus.***BMC evolutionary biology* 2012, **12**:228.

55. Nishimaki T, Ibi T, Tanabe Y, Miyazaki Y, Kobayashi N, Matsuhashi T, Akiyama T, Yoshida E, Imai K, Matsui M, Uemura K, Watanabe N, Fujita T, Saito Y, Komatsu T, Yamada T, Mannen H, Sasazaki S, Kunieda T: **The assessment of genetic diversity**

within and among the eight subpopulations of Japanese Black cattle using 52 microsatellite markers.*Animal science journal = Nihon chikusan Gakkaiho* 2013.

56. Ji Y-Q, Wu D-D, Wu G-S, Wang G-D, Zhang Y-P: **Multi-locus analysis reveals a different pattern of genetic diversity for mitochondrial and nuclear DNA between wild and domestic pigs in East Asia.***PloS one* 2011, **6**:e26416.
57. Verhoeven KJF, Macel M, Wolfe LM, Biere A: **Population admixture, biological invasions and the balance between local adaptation and inbreeding depression.***Proceedings. Biological sciences / The Royal Society* 2011, **278**:2–8.
58. Fernández J, Toro MA, Caballero A: **Management of subdivided populations in conservation programs: development of a novel dynamic system.***Genetics* 2008, **179**:683–92.
59. Berthouly-Salazar C, Thévenon S, Van TN, Nguyen BT, Pham LD, Chi CV, Maillard J-C: **Uncontrolled admixture and loss of genetic diversity in a local Vietnamese pig breed.***Ecology and evolution* 2012, **2**:962–75.
60. García-González DL, Aparicio R, Aparicio-Ruiz R: **Volatile and amino Acid profiling of dry cured hams from different Swine breeds and processing methods.***Molecules (Basel, Switzerland)* 2013, **18**:3927–47.
61. Lynch M: **Estimation of nucleotide diversity, disequilibrium coefficients, and mutation rates from high-coverage genome-sequencing projects.***Molecular biology and evolution* 2008, **25**:2409–19.
62. Engelsma KA: **Use of SNP markers to conserve genome-wide genetic diversity in livestock**. 2012.

Summary

Summary

The developments in sequencing technologies, high density SNP panels and data analysis have provided an increasing number of neutral markers to study genetic variation. This allows better estimation of relevant parameters in conservation genetic studies such as genetic diversity, population structure and demographic history. The recent availability of whole-genome sequences makes it possible to further improve and complement these studies with novel functional genomics approaches. The identification of loci involved in adaptation to the environment and inbreeding depression may have a large impact on conservation genetics, providing additional knowledge to design effective population genetic management of livestock species. The aim of the study presented in this thesis was to explore the genetic diversity and demographic history of local pig populations by integrating various genetic marker systems.

A novel Bayesian phylogeographic approach was implemented in **Chapter 2** to infer the historical dispersal patterns of wild boar populations across Eurasia using mitochondrial DNA (mtDNA). This study aimed to obtain further understanding of the past events that shaped the current genetic structure of *Sus scrofa.* We observed dispersal events of wild pig populations consistent with fossil records and with migration events described in other species. Moreover, statistically significant routes were detected between wild pig populations that were not previously corroborated. In **Chapter 3**, microsatellites, mtDNA and high-density SNP panel were jointly used for the genetic characterization of the local Spanish breed Chato Murciano. This breed represents a good model for endangered livestock populations because it has a small population size, there is no studbook, it is prone to inbreeding and prone to being crossbred with commercial breeds to improve specific traits. The analysis of pigs that likely contributed to the genetic stock of Chato Murciano revealed that the entire breeding stock is genetically distinguishable from other breeds. While genetic diversity of the breed was similar to other European local pig breeds, the independent analysis on farm level showed a high level of substructure and high levels of inbreeding in some farms. However, the farm-by-farm approach allowed the identification of farms with signs of recent crossbreeding with other breeds. Therefore, identifying farm-based management practices and farm-based breeding stocks, seems to be an accurate approach for the development of a sustainable breeding program for minority livestock breeds, particularly when pedigree information is absent.

To study population relationships, inbreeding and demographic history, domestic and wild populations from the Iberian Peninsula were genotyped with the 60K SNP

panel (**Chapter 4)**. The integrated data analyses were based on allele frequency differences, linkage disequilibrium (LD) and regions of homozygosity (ROH). Analysis of population structure and persistence of LD phase indicated a close relationship between two variants of the Iberian breeds -Retinto and Negro Iberico- while the variant Manchado de Jabugo showed signs of recent crossbreeding in agreement with historical records. The study of ROH revealed signs of a recent population bottleneck in the Chato Murciano breed, resulting in part of the pigs having a high proportion of the genome covered by ROH. The study of past effective population size revealed a sharp decrease in effective population size approximately 10,000-15,000 generations ago exclusively in wild populations, probably as a result of the last glacial maximum. Genetic signs of domestication in Europe were solely observed in Iberian pig. The low level of substructure of Iberian wild populations from different geographic regions, together with the pattern of ROH, suggest that migration of wild boar across the Iberian Peninsula may be important for the maintenance of low levels of inbreeding. A broader study of local pig breeds from Europe is presented in **Chapter 5**. High-density SNP and Next Generation Sequencing (NGS) data were applied for 12 local pigs from England, Spain, Italy and Hungary to assess genomic diversity. We observed that pigs from local breeds with high levels of inbreeding also had large number of potentially damaging mutations. Genetic diversity was computed based on high-density SNP data and NGS data. There was a high correlation between these marker systems. NGS data from commercial breeds was included in the study to explore potential candidate mutations responsible for phenotypic divergence among local and commercial breeds. This study revealed fixed non-synonymous genomic variants that may underlie differences in commercial and local breeds for adaptation to the environment. This study highlights the importance of the low input breeds as a valuable genetic reservoir for the pig production industry. Therefore, emphasis is needed to preserve the genomic variability of local breeds in conservation programmes. In **Chapter 6**, the relevant findings of this thesis are discussed as well as the strengths and limitations of the methods used. I put special attention on the practical implications of the results in conservation genetics management of livestock species. The use of 60K SNP data proved to be a suitable tool for the study of relevant parameters in conservation genetics of European pig populations, concretely, genetic diversity, population structure, admixture and past effective population size. This, in combination with NGS data of at least two representative animals of a population, provided important information to assess the ascertainment bias of the high-density SNP chip and to perform functional genomics related to adaptation to the environment.

Samenvatting

Recente ontwikkelingen in de technologie om DNA basen volgorde te bepalen, het zogenaamde “Next Generation Sequencing (NGS)”, hebben geresulteerd in het beschikbaar komen van een toenemend aantal genetische merkers ten bate van de varkensgenetica. Dit heeft onder meer geleid tot de beschikbaarheid van gestandaardiseerde, genoom-wijde merker analyses om variatie te bestuderen, de zogenaamde “High-Density SNP analyses” waarmee zo’n 60 duizend potentieel variabele genomische posities tegelijk kunnen worden bepaald, ook wel SNP chip genaamd. Hierdoor is het mogelijk een betere schatting te maken van genetische parameters die relevant zijn voor het begrijpen van diversiteit, populatie structuur en geschiedenis, en demografie. De recente beschikbaarheid van een referentie genoom voor het varken maakt het verder mogelijk om deze parameters nog nauwkeuriger te schatten en bovendien aan te vullen met functionele analyses, bijvoorbeeld door te kijken naar potentiele veranderingen in de functies van specifieke genen. Dit kan leiden tot het identificeren van loci die bijvoorbeeld betrokken zijn bij aanpassing aan de omgeving of inteelt problemen. Dit kan vervolgens van grote betekenis zijn bij het optimaal behouden van genetische variatie middels het ontwerpen van een optimaal fokprogramma. Het doel van het onderzoek zoals beschreven in dit proefschrift was om de genetische diversiteit en demografische geschiedenis van lokale varkenspopulaties te verkennen middels het integreren van verschillende genetische merker systemen.

Hoofdstuk 2 beschrijft het gebruik van een nieuwe Bayesiaanse fylogeografische aanpak voor het verklaren van de historische verspreidingspatronen van wild zwijn populaties over het Euraziatische super continent aan de hand van mitochondriaal DNA (mtDNA). Deze studie had als doel om inzicht te krijgen in hoe versprijdingspatronen uit het verleden de huidige genetische structuur in de soort *Sus scrofa* kunnen verklaren. De geschatte migratie patronen – op basis van genetische data - van wilde varkenspopulaties zijn consistent met paleontologische data, en zijn verder consistent met fylogeografische patronen zoals die voor andere soorten zijn beschreven. Bovendien werden statistisch significante migratieroutes ontdekt tussen wild zwijn populaties die voorheen niet werden bevestigd. Deze studie legt een belangrijke basis voor het interpreteren van de variatie in de gehele soort – zowel wild als gedomesticeerd.

Vervolgens werden verschillende genetische merkersystemen – microsatellieten, mitochondriaal DNA, en High-Density SNP analyses (de 60K SNP chip) – gezamenlijk gebruikt voor de genetische karakterisering van het lokale Spaanse ras Chato Murciano (**Hoofdstuk3**). Dit ras is een uitstekend model voor bedreigde huisdier

rassen, omdat het een klein aantal dieren betreft, het ras slechts op acht boerderijen word gefokt, er geen stamboek is, er een hoog risico is op inteelt en er een hoog risico bestaat van kruising met commerciële rassen - bijvoorbeeld om specifieke eigenschappen te verbeteren. De analyse van het grootste deel van alle dieren die binnen het ras voor de fokkerij worden gebruikt, laat zien dat het ras op dit moment nog duidelijk onderscheidbaar is van andere varkensrassen. Hoewel de genetische diversiteit in zijn totaliteit vergelijkbaar is met andere Europese lokale varkensrassen, blijken er interessante verschillen tussen individuele boerderijen waar Chato Murciano word gefokt. Enkele boerderijen laten een veel hoger inteelt niveau zien dan anderen, terwijl op sommige boerderijen beslist sprake is van kruisingen met andere rassen. Het identificeren van boerderij-specifiek beheer en boerderij-specifiek fokmateriaal lijkt daarmee een belangrijke strategie voor het genetisch beheer van bedreigde huisdier rassen, met name wanneer een stamboek niet voorhanden is.

De inzichten zo verkregen van het Spaanse Chato Murciano ras worden in **Hoofdstuk 4** in breder perspectief geplaatst door demografische geschiedenis, inteelt en populatie structuur van een aantal wilde en gedomesticeerde varkenspopulaties van het Iberische Schiereiland te bestuderen middels genetische merkertechnologie (60K SNP chip). Deze studie is gebaseerd op een integratie van Linkage Disequilibrium (LD) analyse, allel frequentie verschillen, en het voorkomen van grote gebieden zonder variatie in individuele genomen (Regions Of Homozygosity, ofwel ROH). De analyse van demografie en overeenkomsten in patronen van LD laten nauwe verwantschappen zien tussen de twee varianten van het zogenaamde Iberico ras, de roodgekleurde Retinto en de zwarte Negro Iberico. De Iberico variant Manchado de Jabugo daarentegen vertoont tekenen van recente kruising met andere rassen, in overeenstemming met de opgetekende geschiedenis van deze variant. Het Chato Murciano ras vertoont de sporen van een recente populatie contractie, waardoor in een deel van de varkens een grote fractie van het genoom geen variatie aanwezig is (“Regions of Homozygosity”; ROH). Het Chato Murciano ras blijkt verder ook geen nauwe verwantschap te hebben met de Iberico varianten, wat eveneens in lijn is met wat uit historische bronnen bekend was. De schatting van de effectieve populatiegrootte laat een scherpe daling van de effectieve populatiegrootte ,ongeveer 10.000-15.000 generaties geleden, zien in de wilde populaties, waarschijnlijk het gevolg van de laatste ijstijd. De lage mate van populatie substructuur tussen de verschillende Iberische wilde populaties en de lage mate van recente inteelt zoals gesuggereerd door analyse van ROH, geven aan dat natuurlijke migratie op het Iberisch schiereiland een belangrijke component is

in het behoud van natuurlijke variatie. Genetische tekenen van domesticatie in Europa worden uitsluitend waargenomen in Iberische varkens.

Een nog bredere en gedetailleerdere studie van lokale varkensrassen uit Europa wordt gepresenteerd in **Hoofdstuk 5** . Een combinatie van High-Density SNP en Next Generation Sequencing (NGS) gegevens werd toegepast voor 12 lokale varkensrassen uit Engeland , Spanje , Italië en Hongarije om genoom-wijde diversiteit te karakteriseren. Er is een sterke correlatie tussen schattingen van genetische diversiteit gebaseerd op de 60K SNP Chip en NGS data. De NGS data is verder gebruikt om direct inzicht te krijgen in mutaties die potentieel verantwoordelijk zijn voor veranderingen in het fenotype. Deze studie identificeert niet-synonieme genomische varianten in genen die bepaalde verschillen tussen lokale en commerciële varkensrassen kunnen verklaren, en die bijvoorbeeld ten grondslag liggen aan aanpassingen aan de lokale omgeving. Eén van de bevindingen is dat in varkens van lokale rassen waarbij een hoge mate van inteelt is geschat op basis van neutrale genetische merkers, ook een relatief groot aantal potentieel schadelijke mutaties werd gevonden. Deze studie toont het belang van de oude, lokale rassen als een waardevol genetisch reservoir voor toekomstige fokprogramma's, en de noodzaak deze genetische variabiliteit te behouden middels beschermingsprogramma's.

In **Hoofdstuk 6** worden de relevante bevindingen van dit proefschrift besproken alsook de sterke punten en beperkingen van de gebruikte methoden. Ik besteed speciaal aandacht aan de praktische implicaties van de resultaten in genetisch beheer van diersoorten . Het gebruik van 60K SNP data blijkt een geschikt instrument voor de studie van relevante parameters ten bate van genetisch beheer van Europese varkenspopulaties, met name door het schatten van genetische diversiteit , de effectieve populatie-grootte, demografie, en het kruisen met andere rassen. In combinatie met NGS gegevens van ten minste twee representatieve dieren van een populatie kan een gedetailleerde inschatting van de functionele genomica van aanpassing aan de omgeving worden gemaakt, alsook de functioneel genomische implicaties van beheersmaatregelen, zoals inteelt, helpen voorspellen.

Acknowledgements

This research project reflects the collaborative research within and across research groups from different countries. The work performed in Spain and The Netherlands has given me the possibility to work with colleagues that have enriched my professional life. At the same time I have met new friends that, together with my old friends and family, have contributed greatly to my personal life.

I would like to express my gratitude to my supervisors Prof. Dr. Martien Groenen, Dr. Richard Crooijmans and Dr. Hendrik-Jan Megens for making this research possible. Their supervision, advice as well as their effort in reading the drafts are greatly appreciated. I thank Prof. Dr. Johan van Arendonk for accepting me as a PhD at Wageningen University and making it possible to extend my stay in Wageningen. Furthermore I would like to thank Dr. Guillermo Ramis and Dr. Antonio Muñoz for introducing me to this Dutch adventure and for their continuous support and guidance. It is being an immensely rewarding experience.

Special thanks also to all my friends at Animal Breeding and Genomics group. I thank my “molecular genomics” friends – Yogesh, Mirte, Laurent– and also to Gus Rose (Mr. R) for their help during my learning process in bioinformatics. I have learnt a lot from all of you. Deepest gratitude to my “quantitative genetics” friends who have greatly contributed to my happiness in Wageningen.

I would like to thank my loved friends from Murcia, especially Joel, Ross, Maria Moreno, Maria Marti, Juani y Aida who have always supported me. Last but not least, I would like to thank my mother for her unconditional support throughout my career.

Curriculum Vitae

About the author

Juan Manuel Herrero Medrano was born in Alicante (Spain) on 16th January 1983. He graduated in Veterinary Medicine at the University of Murcia (Spain) in 2007. Immediately after the completion of his Bachelor, he worked in the Genetic Area of the Animal Production Department (Murcia University, Spain) for five years. During this initial stage of his career, Juan Manuel acquired a strong background in research with focus on animal health and molecular genetics. He worked in several research projects and, at the same time, he completed his Master in Biotechnology of Porcine Reproduction (thesis: "In vitro Production of pig embryos: current status. *Review*", 2008) and his Master in Porcine Livestock (thesis: "Comparative study of the incidence of enteric pathogens at slaughter from pigs vaccinated against *E. coli*", 2008) at Murcia University.

Juan Manuel started his first PhD at Murcia University in 2009, when he received a grant for the genetic characterization of an endangered pig breed, The Chato Murciano Pig. This grant included several visits to the Animal Breeding and Genomics Group (Wageningen University, The Netherlands) where he greatly improved his skills in analysing genetic data. The knowledge acquired in both institutions resulted in his PhD thesis entitled "Genetic Characterization of the Chato Murciano Pig" (2012, Murcia). In that summer of 2012, Juan Manuel moved to The Netherlands to continue his active collaboration with the Animal Breeding and Genomics Group, which resulted in this thesis. Currently, he is continuing his scientific career as a postdoctoral researcher at the TOPIGS Research Center IPG (The Netherlands).

List of publications

Publications under review or in preparation

Herrero-Medrano JM, Megens HJ, Groenen MAM, Ramis G, Bosse M, Crooijmans RP. Conservation genomic analysis of domestic and wild pig populations from the Iberian Peninsula. *BMC genetics*. *Manuscript Submitted.*

Herrero-Medrano JM, Megens HJ, Crooijmans RP, Groenen MAM. Whole-genome sequence analysis reveals differences in population management and selection of low-input breeds between European countries. *Manuscript Submitted.*

Herrero-Medrano JM, Nidup K, Larson, G, Megens HJ, Shapiro B, Ramis G, Groenen MAM, Crooijmans RP. Migration patterns of worldwide pig populations assessed using mitochondrial DNA analysis. *Manuscript in preparation.*

Peer-reviewed publications

Optimization of cytotoxicity assay by real-time, impedance-based cell analysis. Ramis G, Martínez-Alarcón L, Quereda JJ, Mendonça L, Majado MJ, Gomez-Coelho K, Mrowiec A, **Herrero-Medrano JM**, Abellaneda JM, Pallares FJ, Ríos A, Ramírez P, Muñoz A. *Biomed Microdevices* (2013). (In press).

Herrero-Medrano JM, Megens HJ, Crooijmans RP, Abellaneda JM, Ramis G. Farm-by-farm analysis of microsatellite, mtDNA, and SNP genotype data reveals inbreeding and crossbreeding as threats to the survival of a native Spanish pig breed. *Animal Genetics* (2013). 44:259-66.

Goedbloed DJ, Megens HJ, Van Hooft P, Herrero-Medrano JM, Lutz W, Alexandri P, Crooijmans RP, Groenen M, Van Wieren SE, Ydenberg RC, Prins HH. Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Molecular Ecology* (2013). 22:856-66.

Quereda JJ, **Herrero-Medrano JM**, Abellaneda JM, García-Nicolás O, Martínez-Alarcón L, Pallarés FJ, Ramírez P, Muñoz A, Ramis G. Porcine Endogenous Retrovirus

Copy Number in Different Pig Breeds is not Related to Genetic Diversity. *Zoonoses and Public Health*. (2012) 59:401-7.

Ramis G, Martínez-Alarcón L, Majado MJ, Quereda JJ, Mendonça L, **Herrero-Medrano JM**, Abellaneda JM, Ríos A, López-Navas A, Ramírez P, Muñoz A. Assessment of *in vitro* heparin complement regulation capacity during real-time cell analyzer antibody-mediated cytolysis assay: compatibility studies for pig-to-baboon xenotransplantation. *Transplantation Proceedings* (2012). 44:1584-8.

Abellaneda JM, Ramis G, Martínez-Alarcón L, Majado MJ, Quereda JJ, **Herrero-Medrano JM**, Mendonça L, García-Nicolás O, Reus M, Insausti C, Ríos A, López-Navas A, González MR, Pallarés FJ, Muñoz A, Ramírez P, Parrilla P. Generation of Human-to-Pig Chimerism to Induce Tolerance Through Transcutaneous In Utero Injection of Cord Blood-Derived Mononuclear Cells or Human Bone Marrow Mesenchymals Cells in a Preclinical Program of Liver Xenotransplantation: Preliminary Results. *Transplantation Proceedings* (2012). 44:1574-1578.

Ramis G, Martínez-Alarcón L, Majado MJ, Quereda JJ, Mendonça L, **Herrero-Medrano JM**, Abellaneda JM, Gomes-Coelho K, López-Navas A, Ríos A, Ramírez P, Muñoz A. Donor-graft compatibility tests in pig-to-primate xenotransplantation model: serum versus plasma in real-time cell analyzer trials. *Transplantation Proceedings* (2011). 43:249-53

Martínez-Alarcón L, Ríos A, Ramis G, Quereda JJ, **Herrero JM**, Muñoz A, Parrilla P, Ramírez P. Are veterinary students in favor of xenotransplantation? An opinion study in a Spanish university with a xenotransplantation program. *Transplantation Proceedings* (2010). 42:2130-3.

Martínez-Alarcón L, Ramis G, Majado MJ, Quereda JJ, **Herrero-Medrano JM**, Ríos A, Ramírez P, Muñoz A. ABO and rh1 blood group phenotyping in pigs (*Sus Scrofa*) using microtyping cards. *Transplantation Proceedings* (2010). 42:2146-8.

Martínez-Alarcón L, Quereda JJ, **Herrero-Medrano JM**, Majado MJ, Mendonça L, Pallarés FJ, Ríos A, Ramírez P, Muñoz A, Ramis G. Design of a real-time quantitative polymerase chain reaction to assess human complement regulatory protein gene expression in polytransgenic xenograft pigs. *Transplantation Proceedings* (2010). 42:3235-8.

Quereda JJ, Martínez-Alarcón L, Mendoça L, Majado MJ, **Herrero-Medrano JM**, Pallarés FJ, Ríos A, Ramírez P, Muñoz A, Ramis G. Validation of xcelligence real-time cell analyzer to assess compatibility in xenotransplantation with pig-to-baboon model. *Transplantation Proceedings* (2010). 42:3239-43.

Conference proceedings

Mendonça L, Martínez Alarcón L, Rios A, Ramis G, Quereda JJ, Abellaneda J, Sáez-Acosta A, **Herrero-Medrano JM**, Muñoz A, Parrilla P, Ramírez P. Are veterinary students in favour of xenotransplantation? Comparative opinion study in a Spanish and a Brazilian university. XI Congresso Luso Brasileiro de Transplantacao. Coimbra, Portugal (2012).

Mendonça L, **Herrero-Medrano JM** ,Abellaneda J, Quereda JJ, García-Nicolás O, Pallares FJ, Muñoz A, Ramis G. Cytokines gene expression in intestinal epithelial cells at slaughter from pigs vaccinated against Escherichia coli. ECMIS symposium: E. coli and the Mucosal Immune system. Ghent, Belgium (2011).

Ramis G, **Herrero-Medrano JM**, Pallarés FJ, Quereda JJ, Mendonça L, Abellaneda J, Muñoz A. Use of microsatellites and SNP markers in swine. XV Congresso da Abraves. Fortaleza, Brazil (2011).

Maiques E, Quereda JJ, Mendonça L, Abellaneda JM, **Herrero-Medrano JM**, García-Nicolás O, Ramis G, Pallarés FJ, Muñoz A. Serological study in piglets vaccinated and not vaccinated born from vaccinated and not vaccinated sows against PCV2. 3rd European Symposium on Porcine Health Management. Espoo, Finland (2011).

Quereda JJ, Gómez S, Seva J, Pallarés FJ, Ramis G, **Herrero-Medrano JM**, García-Nicolás O, Cerón JJ, Muñoz A. Haptoglobin serum concentration is negatively correlated with lymphoid depletion in PMWS affected pigs. 6th International Symposium on Emerging and Re-emerging Pig Diseases. Barcelona, Spain (2011).

Ramis G, Sánchez JJ, Garrido A, Pallarés FJ, Mendonça L, **Herrero-Medrano JM**, Sanjosé E, Quereda JJ, Toledo M, González JM, Muñoz A. Vaccination of piglets against Escherichia coli: effects on nursery performances. 2nd European Symposium on Porcine Health Management. Hannover, Germany (2010).

Ramis G, Mendonça L, Pallarés FJ, **Herrero-Medrano JM**, Quereda JJ, Sánchez JJ, Garrido A, Sanjosé E, Toledo M, González JM, Muñoz A. Vaccination of piglets against Escherichia coli: effects on finishing performances. 2nd European Symposium on Porcine Health Management. Hannover, Germany (2010).

Ramis G, Garrido A, Pallarés FJ, Sanjosé E, Toledo M, **Herrero-Medrano JM**, González JM, Muñoz A. Vaccination with Circovac® in Iberian sows and gilts: effects on farrowing results. 21st International Pig Veterinary Society (IPVS) Congress. Vancouver, Canada (2010).

Ramis G, Martínez Alarcón L, Majado M, Quereda JJ, Mendonça L, **Herrero JM**, Abellaneda JM, Lopes K, Ríos A, Ramírez P, Muñoz A. Donor-graft compatibility tests in pig-to-primate xenotransplantation model: serum versus plasma in RTCA trials. IX Congress Luso Brazileiro de transplantaçao. Oporto, Portugal (2010).

Martinez-Alarcon L, Ramis G, Majado MJ, Quereda JJ, **Herrero-Medrano JM**, Mauri J, Rios A, Ramirez P, Muñoz A. ABO blood group phenotyping in pigs using microtyping cards. International pancreas and islet transplant association-international xenotransplantation association congress (IPITA-IXA). Venetia, Italy (2009).

Ramis G, Martinez-Alarcon L, Majado MJ, Quereda JJ, **Herrero-Medrano JM**, Mauri J, Rios A, Ramirez P, Muñoz A. Lewis, Duffy and Kidd system prevalence in baboon subspecies and pigs. International pancreas and islet transplant association-international xenotransplantation association congress (IPITA-IXA) Venetia, Italy (2009).

Ramis G, Martinez-Alarcon L, Majado MJ, Quereda JJ, **Herrero-Medrano JM**, Mauri J, Rios A, Ramirez P, Muñoz A. Pig anti-baboon antibodies tested at different pig's ages. International pancreas and islet transplant association- international xenotransplantation association congress (IPITA-IXA). Venetia, Italy (2009).

Colophon

The research of this thesis was financially supported by Animal Breeding and Genomics Centre and The graduate school Wageningen Institute of Animal Sciences (WIAS), Wageningen University, The Netherlands.

The cover of this thesis was designed by Francisco Julian Martinez Cano and Juan Manuel Herrero Medrano.

This thesis was printed by GVO drukkers & vormgevers B.V. | Ponsen and Looijen, Ede, The Netherlands.