

TITLE: Refugia genomics: Study of the evolution of wild boar through space and time

ESF Research Networking Programme:

Conservation genomics: amalgamation of conservation genetics and ecological and evolutionary genomics

Final report of the exchange visit (Ref. Num.: 3929)

1.1. Purpose of the visit

The major aim of this exchange visit to Prof. Groenen's group at the University of Wageningen included the use of genomic tools to study wild boar populations in Greece and identify evolutionary and conservation units. Recent increase in wild boar populations requires accurate management which should at the same time take into account the genetic structure of natural populations. The genetic structure of European wild boar is considered to be a result of expansion from multiple glacial refugia like Iberia and the Balkans (Scandura *et al.* 2008; 2011). Previous analysis of wild boar populations identified a large number of mitochondrial DNA haplotypes occurring in the Balkans and in particular Greece (Alexandri *et al.* 2012). These findings suggest several distinct populations survived in the Balkans during the last glacial period but not all of them were involved in the post glacial colonization of central and northern Europe. On the contrary, some populations remained isolated, resulting in geographically confined groups. However, human activities and in particular introgression with domestic pigs can be responsible for the erosion of these indigenous patterns of genetic variation, dilution of adaptive genomes (Verhoeven *et al.* 2011) and sometimes disease transfer and outbreaks. Classic genetic approaches are not sensitive enough to identify genes responsible for adaptation, infer recent and past demographic events as well as accurately estimate the extent and sources of domestic pig introgression. On the other hand, genomic approaches provide valuable data that can be used to solve problems in conservation (Pertoldi *et al.*, 2010; Kerstens *et al.*, 2010).

As indicated in the application form, the current proposal for the ESF exchange visit grant 2012 was awarded to use genomic tools in order to estimate the extent of hybridization between wild boars and domestic pigs; identify genomic regions responsible for adaptation in local conditions; identify conservation units and develop and transfer genomic knowledge and approaches in a conservation context between the Wageningen University and Aristotle University.

1.2. Description of the work carried out during the visit

In order to estimate levels of genetic diversity in Greek wild boar populations I analyzed 92 individuals genotyped with the Illumina porcine SNP60 genotyping Beadchip Infinium SNP assay (Ramos *et al.*, 2009) developed for the domestic pig. These individuals came from three different wild boar populations in Greece and Bulgaria previously identified with mtDNA analysis. Moreover, whole sequence genome data, which were available for some of the individuals, were used for comparison. Greek and Bulgarian wild boars were also compared with other wild and domestic European populations.

For the estimation of domestic pig introgression extent and the identification of possible hybrids between wild and domestic individuals an allele spectrum frequency analysis as described by Goedbloed *et al.* (2013) was performed. Moreover, wild boar individuals were compared with domestic pigs from commercial breeds in order to identify the source of the introgression.

Heterozygosity, allele diversity and levels of genetic differentiation between different populations were calculated with ARLEQUIN version 3.5 (Excoffier & Lischer 2010) while population structure was assessed using the SNP data with STRUCTURE 2.3.3 (Pritchard *et al.* 2000; Hubisz *et al.* 2009). Population structure results were compared with previously obtained microsatellite data analysis and different approaches were evaluated. Moreover, a large number of wild boar individuals from other European populations were used to compare different post glacial expansion scenarios.

In order to identify recent demographic changes in the populations, regions of homozygosity (ROHs) across the genome were identified with both SNP and sequence data using the methods described by Bosse *et al.* (2012). Differences of ROHs between different populations were tested with the χ^2 test of proportions and goodness of fit in R v. 2.15.1.

Linkage disequilibrium (LD) patterns between different populations were measured by the correlation between variable bases (r^2) for each population separately. LD (r^2) was estimated for all marker pairs less than 3 Mb apart for each chromosome independently. Effective population sizes for each group were estimated using the equation (Sved 1971): $r^2 = 1/4N_e + 1$, where r^2 is the linkage disequilibrium, c is the distance between markers and N_e the effective population size. Past effective population size at generation T was calculated according to Hayes *et al.* (2003) as $T = 1/2c$. In order to avoid incorrect estimates of past N_e due to different recombination rate across porcine chromosomes (Bosse *et al.* 2012) the average recombination map described by Tortereau *et al.* (2012) was used.

Effective population size estimates were obtained by averaging multiple genomic regions for a better approximation (Stumpf & McVean 2003): chromosomes were divided into 1 Mb bins containing recombination rate information and average r^2 for all SNP pairs included in each bin.

In order to identify genomic regions under selection pairwise F_{ST} and heterozygosity values for each SNP were calculated between all populations. F_{ST} values were plotted against heterozygosities and candidate loci for directional and balancing selection between populations were identified. “Bedtools”, and “closestBed” (<http://bedtools.readthedocs.org/en/latest/content/tools/closest.html>) tool were used in order to find whether the SNPs are located inside specific genes in the wild boar genome.

1.3. Description of the main results obtained

Domestic pig introgression analysis was based entirely upon SNP data since the observation of allele frequency spectrum can give important details regarding the extent and the source of introgression. The total number of introgressed individuals was found to be slightly higher than previously reported figures for European wild boar populations (Scandura *et al.* 2008; Goedbloed *et al.* 2013). These individuals come from different areas within the wild boar distribution in Greece, while more than one domestic pig breeds were recognized to share alleles with them, suggesting this genetic introgression had multiple sources.

Population structure and linkage disequilibrium patterns analysis supported the existence of three different groups with separate demographic histories which reflect different evolutionary units. The separation of the three groups was estimated to date back to the last glacial period and involves isolation to different glacial refugia. Hybrid zones were recognized consistent with secondary contact of different groups after the post glacial expansion. Comparison of SNP data with those from other European wild boar populations further confirmed the refugia hypothesis.

Regions of homozygosity analyses results provided important insight in recent wild boar migration events. These events took place during the last decades in the context of a management effort in some areas of Greece and had controversial results: wild boar numbers in the areas involved were increased, but the new wild boar populations were highly inbred.

With the identification of loci under selection it became possible to obtain a list with genes that are likely to be subjected to selection, either balancing or directional (depending on whether the observed F_{ST} value is significantly higher or lower, respectively, than the expected value under

neutrality). The online BioMart tool from the Ensembl database (<http://www.ensembl.org/biomart/martview/fbac9ac357832658d3b1fc7cb6d1761f>) was used to obtain more information, such as description of the genes, the names of proteins they code for and associated Gene Ontology definitions .

1.4 Future collaboration with host institution

We will work closely with Prof. Groenen's group to further explore genomic patterns of diversity in the wild boar genome. Future collaboration will also include analysis of genome wide data obtained from several European wild boar populations as well as the conservation genomics of the low input Greek Black Pig breed.

1.5 Projected publications / articles resulting or to result from the grant (ESF must be acknowledged in publications resulting from the grantee's work in relation with the grant)

The results generated in this study are already used for the preparation of an article dealing with the effects of the last glaciation in European wild boar evolution, the identification of multiple refugia within a refugium and the impact of human activity on genetic diversity patterns. ESF will be acknowledged in the scientific paper in relation with the grant.

1.6 Other comments

It is important to say that during my stay in Prof. Groenen's group I developed considerable knowledge and expertise in the field of genomics. I increased my bioinformatics knowledge by becoming familiar with batch programming in R and Perl scripting languages as well as shell scripting in a Linux environment. I will be able to implement these skills in further genomic studies in the coming years.

References cited

- Alexandri P, Triantafyllidis A, Papakostas S, *et al.* (2012) The Balkans and the colonization of Europe: the post-glacial range expansion of the wild boar, *Sus scrofa*. *Journal of Biogeography* **39**, 713-723.
- Goedbloed DJ, Megens HJ, Van Hooft P, *et al.* (2013) Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations. *Molecular Ecology* **22**, 856-866.
- Bosse M, Megens HJ, Madsen O, *et al.* (2012) Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. *Plos Genetics* **8**, e1003100.
- Excoffier L, Lischer HEL (2010) Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular Ecology Resources* **10**, 564-567.
- Hayes BJ, Visscher PM, McPartlan HC, Goddard ME (2003) Novel multilocus measure of linkage disequilibrium to estimate past effective population size. *Genome Research* **13**, 635-643.
- Hubisz MJ, Falush D, Stephens M, Pritchard JK (2009) Inferring weak population structure with the assistance of sample group information. *Molecular Ecology Resources* **9**, 1322-1332.
- Kesterns H, Crooijmans RPMA, Veenendaal A, *et al.* (2009) Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey *BMC Genomics* **10**, 479.
- Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945-959.
- Pertoldi C, Wójcik JM, Tokarska M., *et al.* (2010) Genome variability in European and American bison detected using the BovineSNP50 BeadChip. *Conservation Genetics* **11**, 627-634.
- Ramos AM, Crooijmans RP, Affara NA, *et al.* (2009) Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. *Plos One* **4**, e6524.
- Scandura M, Iacolina L, Apollonio M (2011) Genetic diversity in the European wild boar *Sus scrofa*: phylogeography, population structure and wild x domestic hybridization. *Mammal Review* **41**, 125-137.
- Scandura M, Iacolina L, Crestanello B, *et al.* (2008) 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* **17**, 1745-1762.
- Stumpf MP, McVean GA (2003) Estimating recombination rates from population-genetic data. *Nature Reviews Genetics* **4**, 959-968.
- Sved JA (1971) Linkage disequilibrium and homozygosity of chromosome segments in finite populations. *Theoretical Population Biology* **2**, 125-141.
- Tortereau F, Servin B, Frantz L, *et al.* (2012) A high density recombination map of the pig reveals a correlation between sex-specific recombination and GC content. *Bmc Genomics* **13**, 586.
- Verhoeven KJ, Macel M, Wolfe LM, Biere A (2011) Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proc Biol Sci* **278**, 2-8.