



Research Networking Programmes

Short Visit Grant or Exchange Visit Grant

(please tick the relevant box)

Scientific Report

The scientific report (WORD or PDF file – maximum of eight A4 pages) should be submitted online within one month of the event. It will be published on the ESF website.

***Proposal Title:* Maintaining fitness and diversity using genomic measures of coancestry**

***Application Reference N°:* 4579**

1) Purpose of the visit

Conservation programmes aim at optimizing the probability of survival of the population in the programme, which is obtained by maintaining maximum values of genetic diversity [de Cara et al 2011]. However, when molecular data is used to achieve this goal, deleterious variants could be kept in a population, which can have a negative effect on the overall fitness [de Cara et al 2013]. Recently, a measure of coancestry based on shared regions of the genome has been proposed as a compromise to maintain both fitness and genetic diversity when the population in the programme has some inbreeding load [de Cara et al 2013]. These predictions for management based on genealogical, molecular or IBD segments have been tested with simulated data, but have so far not been implemented with real data. The publication of the pig reference genome in 2012 [Groenen et al 2012] created the opportunity to analyze pig genomes in greatest detail. Stretches of shared coancestry are occurring often in the genome of individual pigs

[Bosse et al 2012], and therefore pigs are an excellent model to test the different proposed management strategies. By using re-sequence data of two highly divergent pig populations, we tested the hypothesis that the management strategy based on shared genome segments can maintain both diversity and fitness. This exchange visit enabled a collaboration between Angeles de Cara and Mirte Bosse in which their expertises on in silico population management and pig genomics were combined.

2) Description of the work carried out during the visit

Within this project, we used re-sequence data, genotype data and pedigree information from two very different pig populations: a commercial line from the European Pietrain breed and a zoo population from the endangered species *Sus cebifrons*. Five Cebifrons and 11 Pietrain individuals were re-sequenced to ~10x depth of coverage each and aligned to the *Sus scrofa* reference genome build 10.2 [Groenen et al 2012]. In addition, 47 Pietrain pigs were genotyped on the Illumina porcine 60K iSelect beadchip [Ramos et al 2009].

Before the start of in silico management, we examined the background of both populations for their most important characteristics. The 5 re-sequenced Cebifrons individuals and 11 Pietrain pigs were compared in terms of their nucleotide diversity and distribution of variable sites over the genome. A filtered genotype matrix including ~100.000 variable sites was constructed from the Cebifrons re-sequence data, and for the Pietrain population the genotypes from the SNPchip were used. For all re-sequenced individuals, we checked whether their genome contained deleterious mutations with the Variant Effect Predictor (VEP) and these alleles were added to the genotype matrices. Because sequence data is available for only a subset of the Pietrain dataset, we included deleterious positions in all individuals and assigned randomly the allele at these positions, so that each individual roughly contained the same proportion of deleterious variants. All chromosomes were phased separately with shapeit to reconstruct the haplotypes that were present before the start of the in silico management. The status of both populations before the management started was recorded in terms of observed heterozygosity, mean fitness and shared coancestry.

Since *Sus cebifrons* is an endangered species, we conducted an analysis of their past effective population size in order to gain more insight into their demographic history. The past N_e of the Cebifrons population was examined with two independent methods using the distribution of variation in individual genomes derived from re-sequence data. We used the pairwise sequential Markovian coalescent (PSMC, Li and Durbin 2011) on all Cebifrons individuals and the runs of homozygosity method described by MacLeod et al (2013) in one of the males to compare the performance of the methods.

Because Pietrain is a commercial pig breed, we screened the Pietrain population for signatures of selection. Although maintenance of variation is important in these populations, diversity in some regions should be reduced because haplotypes linked to particular commercial traits are present. We used the R rehh package [Gautier and Vitalis

2013] to check for extended haplotype homozygosity in the Pietrain 60K dataset and used PLINK [Purcell et al 2007] to screen the genomes for runs of homozygosity.

Using our own Fortran code based on [de Cara et al 2013], we managed both populations for 10 generations, maintaining the same population size and sex ratio as in the initial population. All management strategies were replicated in silico 100 times for each population. Management was based on optimizing diversity by using optimal contributions based on three different measures of coancestry: 1) genealogical coancestry as obtained from pedigree information, where available (when not available, we assumed those individuals unrelated); 2) molecular coancestry as obtained from IBS status of all markers; 3) coancestry based on measurements of IBD segments between individuals. Recombination events per chromosome were drawn from a Poisson distribution based on the mean recombination rate for each chromosome as described in [Tortereau et al 2012]. After each generation, we measured the observed heterozygosity, expected heterozygosity, average coancestry and the average fitness of the population. Finally, we checked the persistence of selection signatures in the Pietrain after 10 generations of each management strategy.

3) Description of the main results obtained

Both populations showed signatures of inbreeding in their genomes, but referring to different time points. We assessed the demographic history of the Cebifrons individuals with two independent methods (MacLeod method displayed in Figure 1). Two major bottlenecks can be observed, which coincide with past fluctuations in sea level and glaciations [Frantz et al 2013]. Both methods show roughly the same patterns of population expansion and reduction, suggesting that the Cebifrons population has been small for a substantial period of time.

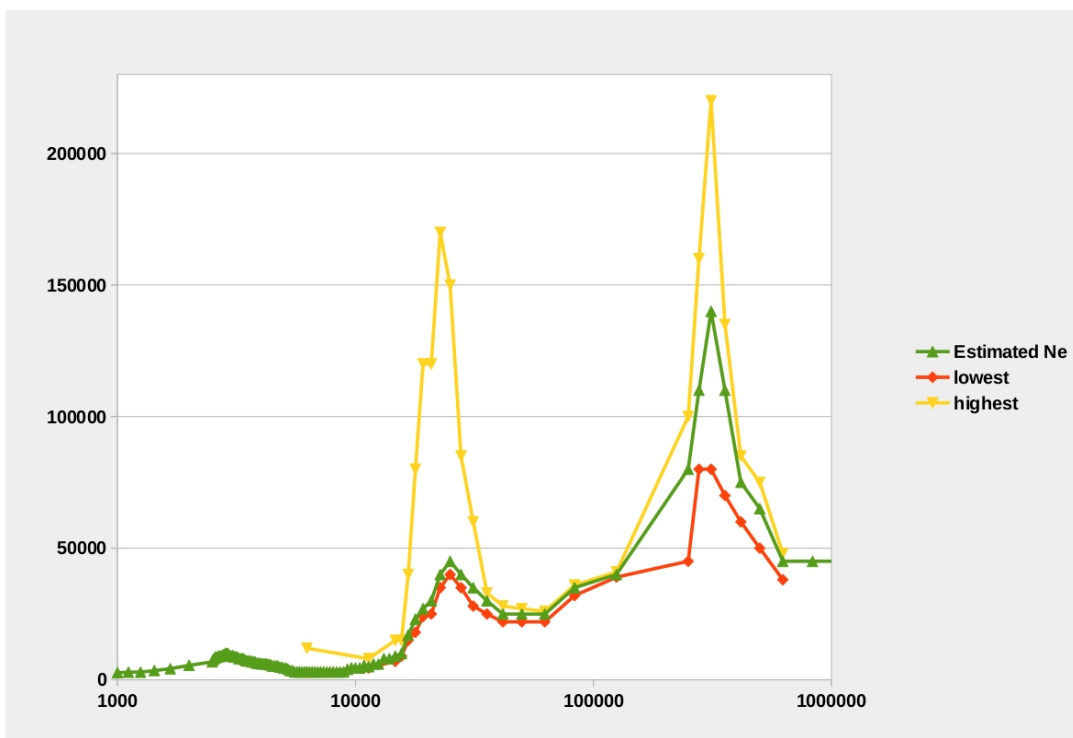


Figure 1. Past effective population size for Cebifrons population. N_e obtained for one male Cebifrons with the ROH-based method as described by MacLeod. The original estimate is in green, confidence intervals are between the red and yellow line. Time (years in the past) is displayed on the x-axis and the y-axis indicates the corresponding effective population size. We assumed a generation time of approximately 5 years.

The effect of each management strategy was similar in both populations. Using molecular coancestry in the management maintained the most diversity, while managing based on genealogical coancestry maintained the least diversity (example for Cebifrons in Figure

2). The segment-based coancestry management resulted in intermediate levels of heterozygosity in the population, with longer segments containing less variation than short segments. When diversity levels are compared between the Cebifrons and Pietrain populations, it can be concluded that the decay of variation using the genealogical coancestry method is much stronger in Cebifrons than in Pietrain. This result is expected because the Pietrain has a larger effective population size.

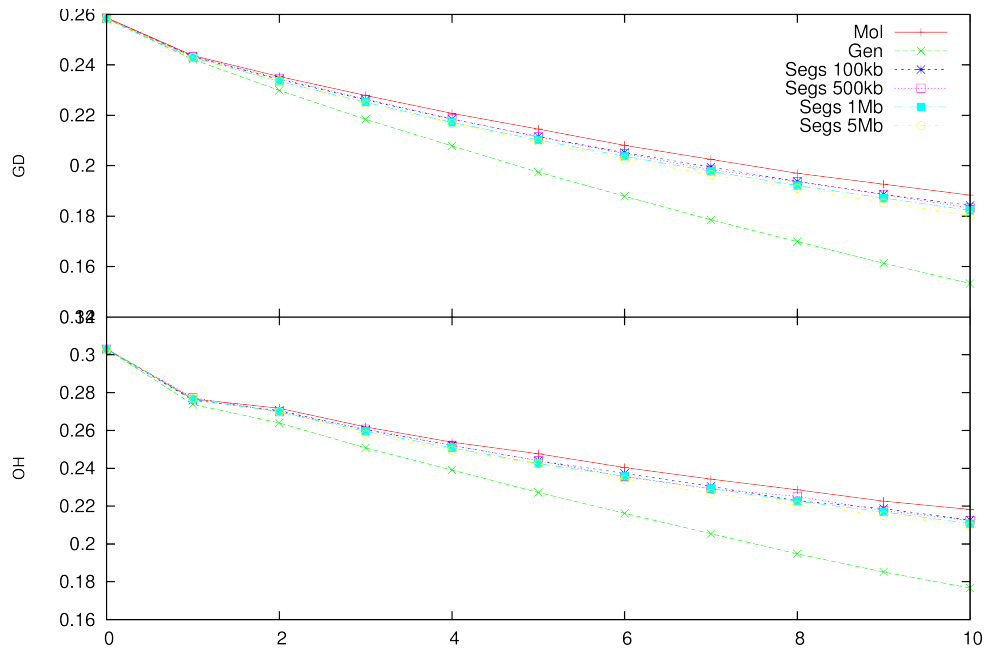


Figure 2. Variation in the Cebifrons population during three management strategies. Genetic diversity and heterozygosity in the Cebifrons population during three management for 10 generations: Molecular-based management, genealogical-based management and management based on segments of 4 different lengths.

The results indicate that management based on molecular coancestry will maintain the highest diversity in the population. However, this strategy does not incorporate the possible negative effects from deleterious alleles that are maintained. When management is applied based on avoiding long segments of coancestry, the diversity is higher than based on genealogical coancestry, but the decrease of fitness should not be as severe as using molecular coancestry in the management. This is illustrated in the decrease of fitness over 10 generations in the Cebifrons populations, when the different management strategies are applied (Figure 3).

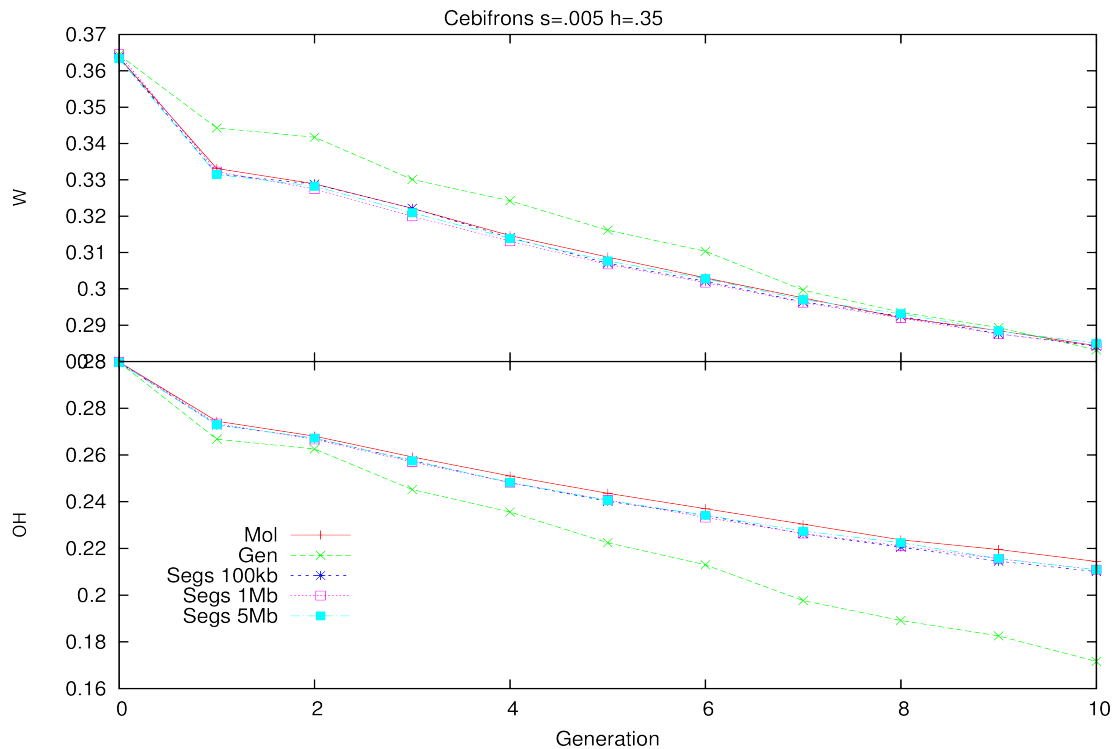


Figure 3. Decay of fitness and observed heterozygosity in the Cebifrons population. Average fitness (W) and observed heterozygosity (OH) is measured in the original population and per generation after management with the molecular, genealogical and segment-based method.

The Pietrain individuals contained some extended regions of homozygosity within their genome. In some instances, this pointed towards signatures of selection as inferred from the extended haplotype homozygosity test (Figure 4). However, we observed that the signatures of selection as inferred from clustering of long haplotypes are reduced in the Pietrain population after management.

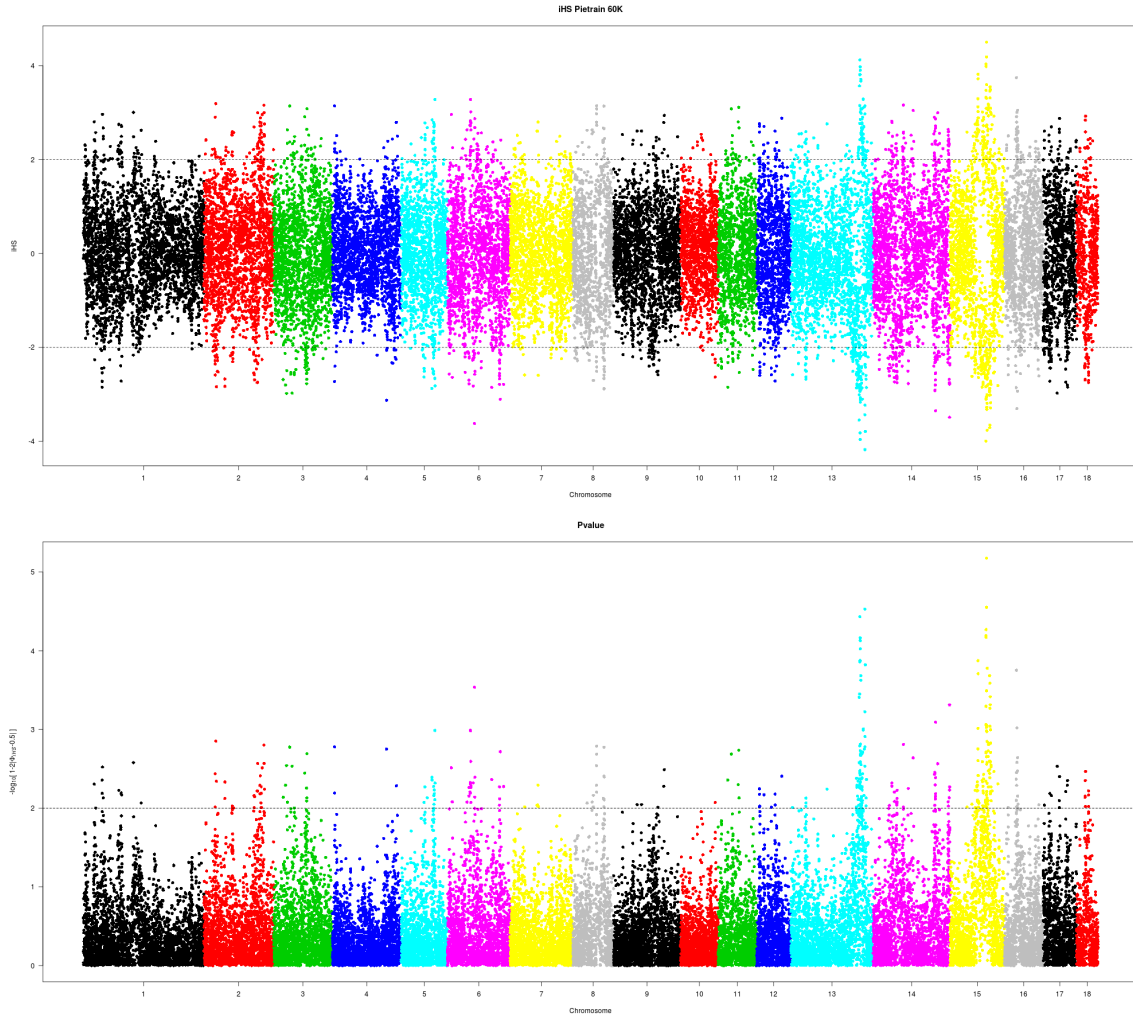


Figure 4. Extended haplotype homozygosity per chromosome for the Pietrain population. 4A displays the raw iHS signal before management over all chromosomes in the Pietrain population. 4B shows the p-value of the iHS signal before management for each marker. Values >2 are defined to be significant.

When artificially selected variants should be maintained in particular regions of the genome, the management strategy should be adapted to the requirements of the breeding goal.

4) **Future collaboration with host institution (if applicable)**

We expect exchanging visits between both labs in order to complete this specific project and depending on available data from the breeding companies, to pursue a detailed study of the demographic and selective history of some breeds.

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 aim is to submit our research paper to a high-impact journal within a few weeks after the exchange. Naturally, the ESF will be acknowledged for this travel grant in the resulting publication.

6) **Other comments (if any)**

References

Bosse M, Megens H-J, Madsen O et al. (2012). Regions of Homozygosity in the Porcine Genome: Consequence of Demography and the Recombination Landscape. *PLoS Genetics* 8(11): e1003100. doi:10.1371/journal.pgen.1003100

de Cara et al. (2013) Using genomic tools to maintain diversity and fitness in conservation programmes *Mol Ecol* 6091–6099, December 2013

de Cara, Fernandez, Toro and Villanueva (2011), "Using genome-wide information to minimise the loss of diversity in conservation programmes", *J Animal Breeding and Genetics* 128:456-464.

Frantz L, Schraiber J, Madsen O et al. (2013). Genomic sequencing provides fine scale inference of evolutionary history. *Genome Biology* 14:R107.

Gautier, Vitalis. rehh: An R package to detect footprints of selection in genome-wide SNP data from haplotype structure

Groenen MAM, Archibald AL, Uenishi H et al. (2012) Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* 491: 393–398.

Li H, Durbin R (2011) Inference of human population history from individual whole-genome sequences. *Nature* 475: 493-496.

MacLeod I et al. (2013) Inferring Demography from Runs of Homozygosity in Whole-Genome Sequence, with Correction for Sequence Errors. *Mol Biol Evol.* Sep 2013; 30(9): 2209–2223.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR et al. (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *Am J Hum Genet* 81: 559-575.

Ramos AM, Crooijmans RPMA, 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.

Tortereau F, Servin B, Frantz L.A.F 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 doi:10.1186/1471-2164-13-586