Genome-wide single nucleotide polymorphism analysis reveals recent genetic introgression from domestic pigs into Northwest European wild boar populations

D. J. GOEDBLOED*, H.J. MEGENS†, P. VAN HOOFT*, J. M. HERRERO-MEDRANO†, W. LUTZ‡, P. ALEXANDRI§, R. P. M. A. CROOIJMANS†, M. GROENEN†, S. E. VAN WIEREN*, R. C. YDENBERG* and H. H. T. PRINS*

* E , , , . . B 47, 6700AA , , †A B C , , , . . B 338, 6700A , , ‡ , C 228, 53229 B , , § B , , A , 54124 ,

Abstract

Present-day genetic introgression from domestic pigs into European wild boar has been

Introduction

European and Asian pigs were independently domesticated from wild boar ( ) (Giuffra . 2000; Larson . 2005). Even though the first domestication of European pigs is estimated to have occurred 9000 years ago (Giuffra . 2000; Larson . 2005), European wild boar are still fully capable of hybridizing with domestic pigs. The process of domestication and later introgression of genetic elements from wild boar into the domestic pig genome is well studied (Giuffra . 2000; Larson . 2005, 2007). In contrast, the extent of introgression from domestic pigs into wild boar is largely unknown (Scandura . 2011). Frequent genetic introgression from domestic pigs may lead to either hybrid vigour or to maladaptation to the natural environment (Verhoeven . 2011). In addition, regular intimate contact between pigs and wild boar may increase the risk of disease transfer and outbreaks. The extent of genetic introgression is thus a relevant parameter for wild boar conservation management and disease risk management. Genetic signs of introgression have been reported in up to 2% of wild boar in Eurasia based on mitochondrial DNA (Giuffra . 2000; Larson . 2005) and in 5–10% of wild boar in Europe based on a combination of mitochondrial DNA and microsatellites (Scandura . 2008). The latter authors consider their estimate to be slightly inflated and report introgression in general to be lower than 5% (Scandura . 2011). Another study using mtDNA D-loop sequences reports only 1.6% Asian haplotypes in wild boar vs. 29% in the European domestic population (Alves . 2010).

European wild boars have survived Pleistocene ice ages in Mediterranean refugia (Scandura . 2008). Wild boars in Western Europe are considered to originate from the Iberian refugium and have a chromosome number of 2 = 36. They differ in their karyotype from domestic pigs and from Balkan refugium wild boar in eastern Europe, both with chromosome number 2 = 38 (Fang . 2006). Hybridization can occur, resulting in individuals with chromosome number 2 = 37 (Scandura . 2011). Admixture between different wild boar populations may locally introduce new alleles.

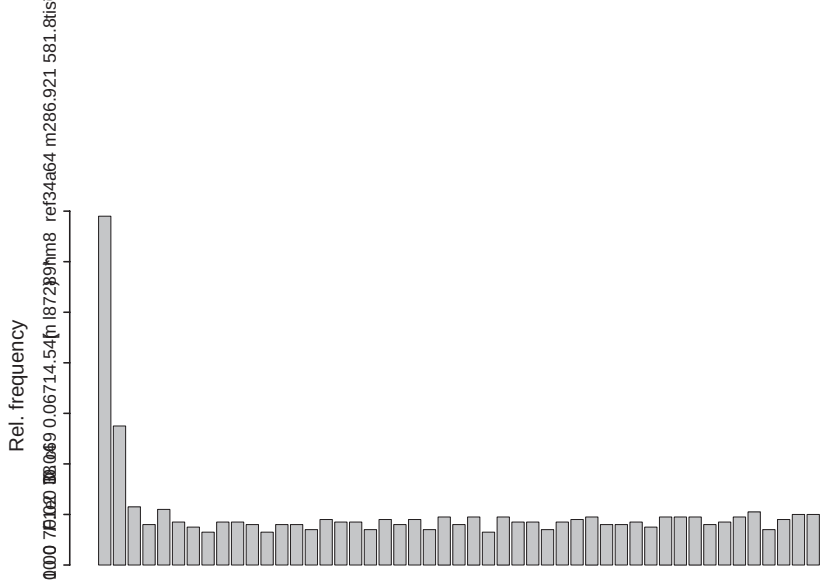
Single nucleotide polymorphism (SNP) genetic markers are found throughout any genome and represent the largest source of genetic variation (Vignal . 2002). Models for the mutation rate of SNPs are well established, and high-throughput genotyping methods are becoming increasingly efficient. These characteristics make SNPs a popular choice of marker for population genetic research (Morin . 2004). Few studies have used genome-wide SNP sets in nonmodel organisms (e.g. Kraus . 2011), as this technology is still relatively new. However, in some cases, a SNP set developed for a model species can be used effectively to study closely related nonmodel species (Narum . 2008; Willing . 2010; Miller . 2011).

In this study, we aimed to identify the occurrence, time frame and possible sources of genetic introgression from domestic pig into Northwest (NW) European wild boar. We used a high-density genome-wide SNP assay developed for domestic pig, the Illumina porcine SNP60 genotyping beadchip (Ramos . 2009), for the genetic analysis of 88 wild boar from the Netherlands, Luxembourg and Western parts of Germany. This assay provided 26505 SNPs that segregated in the wild boar data set and which were distributed across all autosomes. This amounted to a substantially higher genome coverage than commonly seen in molecular ecology studies (Seeb . 2011). We identified genetic introgression based on an increased abundance of rare alleles. Results from a mitochondrial (mt) DNA haplotype study were used to independently verify cases of introgression. The level of introgression from domestic pig was identified using a hybridization simulation study and the genomic distribution patterns of introgressed SNPs.

Methods

In 2008, we collected 88 wild boar blood samples from the Netherlands, Luxembourg and Western parts of Germany. Sample collection was opportunistic and without bias towards age, sex or sampling location (Table S1, Supporting information).

DNA isolation was performed following the Gentra PureGene Blood kit protocol. Samples were genotyped using the Illumina porcine SNP60 genotyping beadchip Infinium SNP assay (Ramos . 2009) and initially analysed for all 45720 autosomal SNPs. The total genotyping rate was 0.98. During exploration using PLINK v1.06 (Purcell . 2007), we found that SNPs with a low minor allele frequency ($0.005 < MAF < 0.030$) were highly abundant in the wild boar data set (Fig. 1a). This allele frequency spectrum was compared with that of a domestic pig data set consisting of 20 individuals per breed for six breeds: British Saddleback (BS), Duroc, Landrace, Large White (LW), Pietrain and Tamworth (Fig. 1b). These breeds were selected on the basis of occurrence in NW Europe and the availability of sufficient SNP data. MAF was in all cases calculated separately for the wild boar and domestic pig data sets. After allele frequency spectrum assessment, we excluded nonpolymorphic sites and potential genotyping errors by applying a rigorous MAF threshold of 0.05 using PLINK, as a standard procedure. This procedure therefore excluded the highly abundant rare alleles for further analysis, making sure that population genetic



inferences were not influenced by potential artefacts. The procedure left 26505 segregating autosomal SNPs for population genetic analysis in the wild boar data set.

The 7083 highly abundant rare SNPs in the wild boar data set (0.005 < MAF < 0.030) were analysed separately and revealed 5038 putative introgressed SNPs, which were private to just nine of 88 wild boar. These putative introgressed SNPs were also analysed for their allelic state in the domestic pig data set and a sample of wild boar from the Balkans (northern Greece and Bulgaria, = 20) to assess the origin of the putative introgressed SNPs.

To identify genetic clustering in the wild boar data set, we performed principal component analysis (PCA) using the eigenvector method as implemented in EIGENSOFT 3.0 (Patterson . 2006; Price . 2006). In addition, we performed a population assignment analysis using STRUCTURE 2.3.1 (Pritchard . 2000) based on 10 runs per number of clusters ( ) for = 1–10 at 1 000 000 iterations and a burn in of 800 000. Putative hybrids

were excluded from these analyses to achieve convergence between runs. The most supported partitioning ( ) was identified using the method of Evanno . (2005). Observed and expected heterozygosity were calculated in R 2.13.0 using the package Adegenet (Jombart 2008). Individual observed heterozygosity (Table 1, $_{o}$) was calculated as the number of heterozygous SNPs divided by the total number of SNPs.

Part of the D-loop region of the mitochondrial DNA (mtDNA) was amplified by polymerase chain reaction (PCR) using the primers described by Luetkemeier . (2010) (L-strand 5′CTCCGCCATCAGCACCCAAAG3′ and H-strand 5′GCACCTTGTTTGGATTRTCG3′) yielding a 772-bp fragment. The PCR amplicons were purified and sequenced for both strands on an ABI 3130® DNA sequencer (Applied Biosystems, USA). Genome Assembly Program (GAP4, Bonfield . 1995) 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

Table 3 Shared single nucleotide polymorphisms (SNPs) between pig breeds (= 20 per breed) and the nine wild boar carrying putative introgressed SNPs. Six two-breed combinations (= 40) with a high amount of shared SNPs are also included, as well as a sample of wild boar from the Balkans (= 20)

Breed/combination	Shared SNPs	Percentage
Large White	4028	80
Landrace	3994	79
Pietrain	3868	77
British Saddleback	3647	72
Duroc	2876	57
Tamworth	1946	39
Large White*Landrace	4310	86
Large White*British Saddleback	4306	86
Large White*Pietrain	4267	85
Landrace*Pietrain	4267	85
Landrace*British Saddleback	4252	84
Pietrain*British Saddleback	4247	84
Balkan wild boar	1002	20

Percentages are calculated relative to the total amount of putative introgressed SNPs in our wild boar data set (5038).

was followed by seven generations of backcrossing with the parent wild boar population. We assumed Mendelian inheritance, meaning that the probability of inheritance for a typical pig allele (absent in nonhybrid wild boar) is 0.5 and 1, respectively, for a heterozygous and homozygous SNP in the pig parent. Inheritance of a pig allele leads by definition to a heterozygous SNP in the next generation of hybrids. Each introgressed pig allele theoretically has a 50% probability to be inherited at each subsequent generation of backcrossing with the parent wild boar population, resulting in a halving of the total number of introgressed SNPs each generation. The standard deviation of the number of introgressed SNPs per individual for each generation was estimated on basis of 200 simulated genotypes per generation.

Genomic positions of putative introgressed SNPs were analysed based on build 9 of the pig genome published by the International Swine Genome Sequencing Consortium in release 66 of the Ensembl database as Sscrofa9 (http://www.ensembl.org/Sus_scrofa/Info).

Results

The wild boar and domestic pig allele frequency spectra (Fig. 1a,c, respectively) differ dramatically at the lower end of the spectrum. In both cases, we expected a more or less uniform distribution of SNPs across the allele frequency range based on random genetic drift and random mating. However, in the wild boar data, we observed a clear excess of rare SNPs (0.005 < MAF < 0.030, Fig. 1a). A large proportion (69%, 5038 SNPs) of these rare SNPs were private to just nine wild boar. These putative introgressed SNPs (all heterozygous in those wild boar) almost correspond to the surplus in this MAF range, which in a uniform distribution would be expected to hold approximately 2250 SNPs rather than the observed 7083 SNPs. The nine wild boar with putative introgressed SNPs displayed higher overall levels of observed heterozygosity ($_o$, Table 1) compared with other wild boar (Table 2).

Principal component analysis separated the wild boar data set into four genetic clusters (Fig. 2a), with the nine putative hybrid individuals scattered across three of these clusters (inverted triangles). The inclusion of a sample of domestic pigs in the PCA provided extra

resolution, and clearly positioned these nine putative hybrid wild boar separately from the wild boar clusters, trailing off in the direction of the domestic pig (Fig. 2b). The geographic origin of six of them (Fig. 3) corresponded to their association with a particular genetic cluster. However, three putative hybrid wild boar (2, 3 and 5) clustered genetically with the Veluwe population (Fig. 2, circles) but were sampled geographically in the Meinweg population in the South of the Netherlands (Fig. 3, diamonds).

The most supported STRUCTURE partitioning of the data following the method of Evanno . (2005) was = 3 followed by = 4 (Fig. S4, Supporting information). However, this method is known to favour only the first level of structure in a given data set. In addition, the assignment of clusters for = 3 was not geographically coherent. German individuals were divided over the Meinweg and the Veluwe clusters with dubious assignment probabilities (Table S1, Supporting information).

LW. Haplotypes HP110 and HP8 were not found in any of the 79 wild boar without putative introgressed SNPs.

The number of putative introgressed SNPs in each of the nine wild boar is indicated in Table 1. These numbers are decreasing (or increasing) more or less stepwise by a factor of two at each putatively assigned generation of backcrossing. This suggested a scenario of introgression followed by backcrossing with a wild boar gene pool theoretically halving the number of introgressed alleles at every generation of backcrossing.

To investigate the individual levels of introgression, we simulated hybrid genotypes using genotypes from the Veluwe wild boar population (Fig. 3) and either of two domestic pig breeds: LW and BS. The number of putative introgressed alleles per individual wild boar observed in this study corresponded to expectations according to the hybridization simulations (Fig. 4). Wild boar individual seven was identified as equivalent to a first-generation (F1) hybrid, wild boar individuals 2, 3, 6 and 8 were identified as equivalent to a second-generation (F2) backcross to wild boar, individuals 9 and 5 were equivalent to a third-generation (F3) backcross, individual 1 was equivalent to a fourth-generation (F4) backcross and individual 4 was equivalent to a fifth-generation backcross (Fig. 4).

The chromosomal positions of the introgressed SNPs are indicated for some of the identified hybrids in Fig. 5. Individual 7 displays a wide array of introgressed alleles, resulting in a high prevalence of heterozygous SNPs across the entire genome. This pattern of genome-wide heterozygosity corresponds to expectations for an F1 hybrid. Individuals 2, 5 and 1 represent subsequent generations of backcrossing with wild boar according to our hybridization simulation. The number of introgressed alleles is clearly diluted over the generations, and the chromosomal positions show a clear clustering pattern that is distinct for each individual.

Discussion

The data presented here reveal recent hybridization and widespread genetic introgression from domestic pigs into European wild boar populations. We identified introgression by analysing the wild boar allele frequency spectrum, which showed an excess of rare polymorphisms (Fig. 1a). These putative introgressed SNPs were exclusive to just nine individuals of 88 sampled wild boar, from dispersed geographical origins (Fig. 3). The nine putative hybrid wild boar also displayed elevated levels of observed heterozygosity (Table 1) compared with other wild boar (Table 2). When we included a sample of domestic pigs in a PCA, these nine individuals were positioned between the wild boar

eee h813959 / F T6.52 c13959 / F T78 571 4 832.25.121

which kept the number of introgressed SNPs per individual relatively high over an extended time frame. For example, a third-generation hybrid × third-generation hybrid cross would result in offspring with on average the same number of introgressed alleles as their parents, but it would be the fourth generation since the hybridization event. Sexual reproduction and recombination between different hybrid genomes with distinct individual patterns of introgressed SNP clustering will result in more widespread distribution of introgressed SNPs at every generation of reproduction among hybrids. We

the hybrid with this haplotype (individual 4) suggests an advanced-generation hybrid similar to individuals 6, 8 and 9. The most likely scenario seems to be escape or release of a hybrid wild boar stock influenced by LW or Landrace pigs, which resulted from an older hybridization event followed by interbreeding among hybrids.

The relatively low number of shared introgressed SNPs between the nine identified hybrids and wild boar from the Balkans (Table 3) indicates that natural introgression of alleles from eastern European wild boar cannot explain our observations. We consider the low number of shared introgressed SNPs in Balkan wild boar to reflect a history of free-ranging pig farming practices with associated exchange of genetic material between domestic pigs and wild boar in Mediterranean Europe (Scandura . 2008). Recent genetic contributions from eastern European wild boar into the study area are considered to be negligible.

The domestic pig breeds that are possibly involved in the identified introgression (LW, Landrace, BS, etc.) carry dominant white spotting alleles. This could lead to deviating coat colour in hybrids, particularly in the first generation. Although no phenotypic details were recorded in this study, all wild boar samples were taken from animals identified in the field as true wild boar, and therefore, strong deviations in coat colour are unlikely. If the identified hybrids originate from a hybrid farmed wild boar stock as suggested in some cases by discrepancies in genetic association and geographic distribution, these animals may have been subject to artificial selection against the domestic phenotype during their farm history. Anecdotal reports of wild boar with deviating coat colour in NW Europe are very rare.

Farmed wild boar are often cross-bred to a certain extent with domestic pigs to increase piglet growth rate and litter size (Goulding 2001). Geographic differences in wild boar litter size have been previously reported in Western Germany (Gethoffer . 2007). These may be a result of local differences in the level of genetic introgression from domestic pig through the escape or release of hybrid farmed wild boar.

Wild boar numbers have increased markedly in Europe since the 1960s (Saezroyuela & Telleria 1986; Briedermann 1990; Geisser & Reyer 2005). This population growth and accompanying range expansion has been associated with mild winters and increased food availability through augmented mast frequency and changes in agriculture (Bieber & Ruf 2005; Geisser & Reyer 2005). In some areas, genetic introgression from domestic pigs may have added to the rapid population

Larson G, Albarella U, Dobney K . (2007) Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. *A* , 104, 15276–15281.

pigdg
r Phylogenetiecsi and Evoluation