MOLECULAR ECOLOGY

Molecular Ecology (2013) **22**, 856–866 doi: 10.1111/j.1365-294X.2012.05670.x

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*

*Resource Ecology Group, Wageningen UR, P.O. Box 47, 6700AA Wageningen, The Netherlands, †Animal Breeding and Genomics Centre, Wageningen UR, P.O. Box 338, 6700AH Wageningen, The Netherlands, ‡Wildlife Research Institute, Pützchens Chaussee 228, 53229 Bonn, Germany, §School of Biology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece

Abstract

Present-day genetic introgression from domestic pigs into European wild boar has been suggested in various studies. However, no hybrids have been identified beyond doubt mainly because available methods were unable to quantify the extent of introgression and rule out natural processes. Genetic introgression from domestic pigs may have far-reaching ecological consequences by altering traits like the reproduction rate or immunology of wild boar. In this study, we demonstrate a novel approach to investigate genetic introgression in a Northwest (NW) European wild boar data set using a genome-wide single nucleotide polymorphism (SNP) assay developed for domestic pigs. We quantified the extent of introgression using allele frequency spectrum analysis, in silico hybridization simulations and genome distribution patterns of introgressed SNPs. Levels of recent introgression in the study area were expected to be low, as pig farming practices are prevailingly intensive and indoors. However, evidence was found for geographically widespread presence of domestic pig SNPs in 10% of analysed wild boar. This was supported by the identification of two different pig mitochondrial DNA haplotypes in three of the identified hybrid wild boar, suggesting that introgression had occurred from multiple sources (pig breeds). In silico hybridization simulations showed that the level of introgression in the identified hybrid wild boar is equivalent to first-generation hybrids until fifth-generation backcrosses with wild boar. The distribution pattern of introgressed SNPs supported these assignments in four of nine hybrids. The other five hybrids are considered advanced-generation hybrids, resulting from inter-breeding among hybrid individuals. Three of nine hybrids were genetically associated with a different wild boar population than the one in which they were sampled. This discrepancy suggests that genetic introgression has occurred through the escape or release of an already hybridized farmed wild boar stock. We conclude that genetic introgression from domestic pigs into NW European wild boar populations is more recent and more common than expected and that genome-wide SNP analysis is a promising tool to quantify recent hybridization in free-living populations.

Keywords: domestic pig, hybridization, introgression, single nucleotide polymorphism, *Sus scrofa*, wild boar

Received 19 January 2012; revision received 4 May 2012; accepted 9 May 2012

Correspondence: Daniel J. Goedbloed, Fax: +31 317 419000;
E-mail: daniel.goedbloed@wur.nl

Introduction

European and Asian pigs were independently domesticated from wild boar (*Sus scrofa*) (Giuffra *et al.* 2000; Larson *et al.* 2005). Even though the first domestication of European pigs is estimated to have occurred 9000 years ago (Giuffra *et al.* 2000; Larson *et al.* 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 *et al.* 2000; Larson *et al.* 2005, 2007). In contrast, the extent of introgression from domestic pigs into wild boar is largely unknown (Scandura *et al.* 2011). Frequent genetic introgression from domestic pigs may lead to either hybrid vigour or to maladaptation to the natural environment (Verhoeven *et al.* 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 *et al.* 2000; Larson *et al.* 2005) and in 5–10% of wild boar in Europe based on a combination of mitochondrial DNA and microsatellites (Scandura *et al.* 2008). The latter authors consider their estimate to be slightly inflated and report introgression in general to be lower than 5% (Scandura *et al.* 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 *et al.* 2010).

European wild boars have survived Pleistocene ice ages in Mediterranean refugia (Scandura *et al.* 2008). Wild boars in Western Europe are considered to originate from the Iberian refugium and have a chromosome number of $2n = 36$. They differ in their karyotype from domestic pigs and from Balkan refugium wild boar in eastern Europe, both with chromosome number $2n = 38$ (Fang *et al.* 2006). Hybridization can occur, resulting in individuals with chromosome number $2n = 37$ (Scandura *et al.* 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 *et al.* 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 *et al.* 2004). Few studies have used genome-wide SNP sets in nonmodel organisms (e.g. Kraus *et al.* 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 *et al.* 2008; Willing *et al.* 2010; Miller *et al.* 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 *et al.* 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 *et al.* 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 *et al.* 2009) and initially analysed for all 45720 autosomal SNPs. The total genotyping rate was 0.98. During exploration using PLINK v1.06 (Purcell *et al.* 2007), we found that SNPs with a low minor allele frequency ($0.005 < \text{MAF} < 0.030$) 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

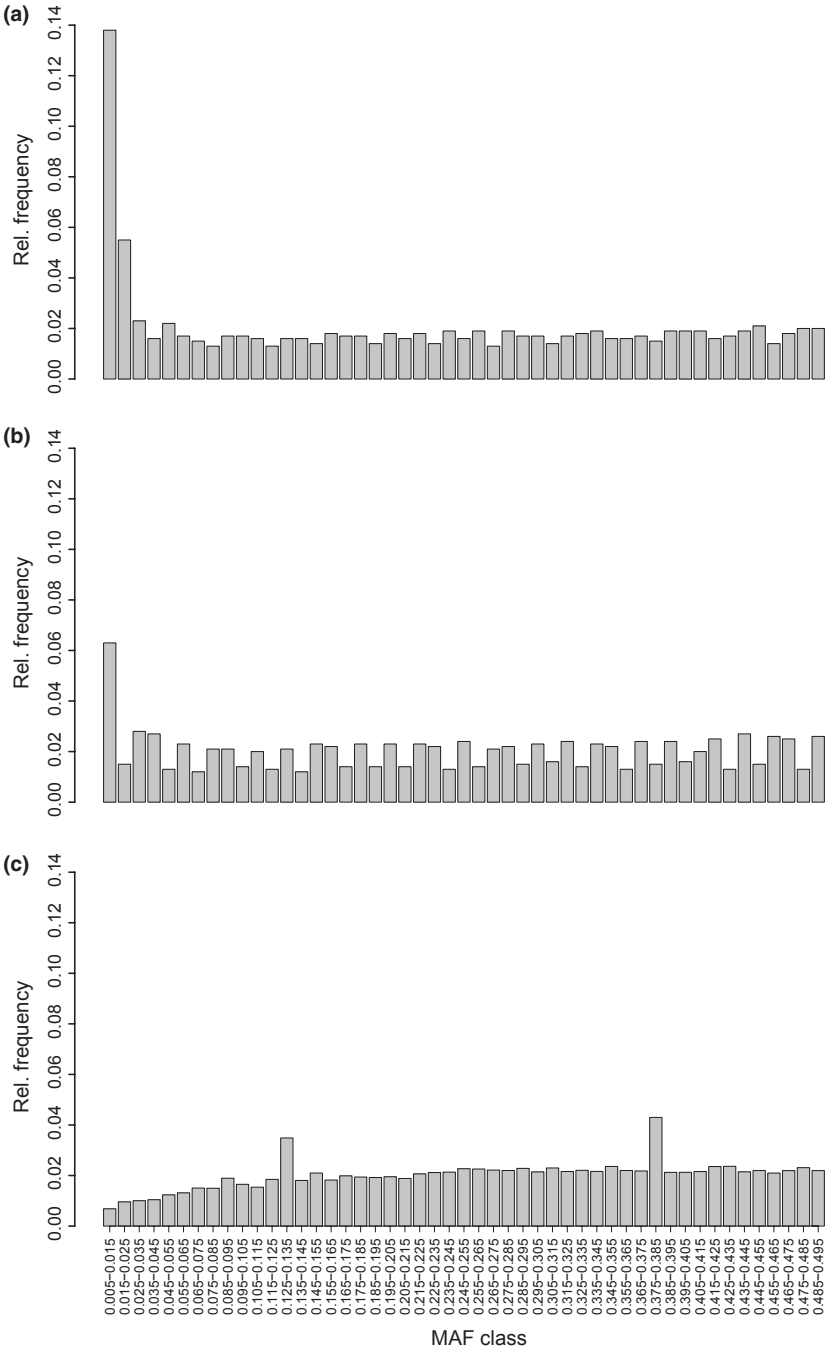

Fig. 1 MAF distribution in (a) the wild boar data set, (b) the wild boar data set without nine putative hybrids and (c) the domestic pig data set. The *x*-axis indicates the MAF class. The *y*-axis indicates the frequency of each MAF class relative to the total number of single nucleotide polymorphisms in the data set.

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, $n = 20$) 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 *et al.* 2006; Price *et al.* 2006). In addition, we performed a population assignment analysis using STRUCTURE 2.3.1 (Pritchard *et al.* 2000) based on 10 runs per number of clusters (K) for $K = 1$–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 (K) was identified using the method of Evanno *et al.* (2005). Observed and expected heterozygosity were calculated in R 2.13.0 using the package Adegenet (Jombart 2008). Individual observed heterozygosity (Table 1, H_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 *et al.* (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 *et al.* 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.2 (Larkin *et al.* 2007) and grouped into haplotypes using the program ALTER (Glez-Pena *et al.* 2010). As not all samples yielded the complete fragment (722 bp), a 624-bp fragment common to most samples was finally used for the analysis. Phylogenetic relationships among the haplotypes were determined with MEGA 5.03 (Tamura *et al.* 2007) using the neighbour joining (NJ) method based on Tamura–Nei model. We included three additional NW European pig breeds: Berkshire, Bunte Bentheimer and Gloucester Old Spot in the mtDNA haplotype analysis (Table S2, Supporting information), as well as three sequences (accession numbers: DQ379224, DQ379100 and DQ379099) from Fang & Andersson (2006). Novel sequences were submitted to GenBank (Table S3, Supporting information).

Hybridization simulations between domestic pigs and wild boar were performed in Excel 2010 using only monomorphic and rare SNPs with MAF < 0.030 in the wild boar data set. We used genetic data from the Veluwe population in the central Netherlands (Fig. 3, indicated by circles, $n = 23$) as the wild boar parent population. Analysis of shared polymorphisms (Table 3) and mtDNA haplotypes (Table 2) led us to specifically use the LW and the BS pig breed ($n = 20$ per breed) as parent pig populations for the hybridization simulations. LW shared most putative introgressed SNPs (80%) with the identified hybrid wild boar (Table 3) and harboured the observed pig haplotype HP8 (Table 2). BS shared 72% of putative introgressed SNPs with the identified hybrid wild boar and harboured the observed pig haplotype HP110. LW displayed 13879 SNPs with a nonwild boar allele and BS displayed 11989. The first-generation hybridization (F1)

Table 1 The number of putative introgressed single nucleotide polymorphisms (SNPs), observed heterozygosity (H_o) based on 26505 SNPs with MAF > 0.05 and mtDNA haplotype per individual hybrid wild boar. The numbering of individuals corresponds to Figs 2 and 3

Individual	Rare SNPs	H_o	MtDNA haplotype
1	256	0.226	HP165
2	1192	0.328	HP110
3	1086	0.325	HP110
4	129	0.202	HP8
5	580	0.207	HP19
6	1137	0.241	HP164
7	2435	0.354	HP164
8	1207	0.305	HP19
9	648	0.260	HP164

Table 2 Observed heterozygosity (H_o), expected heterozygosity (H_e) and mtDNA haplotype counts of the wild boar clusters, the group of hybrid wild boar and the six domestic pig breeds

Group	n	H_o*	H_e*	HP164	HP165	HP19	HP110	HP8	HPother
Veluwe	23	0.182	0.191	19	0	4	0	0	0
Meinweg	24	0.160	0.160	1	0	23	0	0	0
Kirchhellen	24	0.177	0.170	0	24	0	0	0	0
Germany	11	0.202	0.208	7	0	4	0	0	0
Hybrids	9	0.268	–†	2	1	3	2	1	0
Large White	20	0.333	0.353	2	0	1	0	1	16
Landrace	20	0.329	0.356	2	0	2	0	1	15
Pietrain	20	0.350	0.354	6	0	0	0	0	14
Brit. Saddleback	20	0.337	0.337	1	0	0	11	0	8
Duroc	20	0.335	0.342	6	0	1	0	0	13
Tamworth	20	0.339	0.324	0	0	0	8	0	12

*Standard errors are 0.001 or smaller.
†Not calculated as the hybrids do not constitute a population.

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

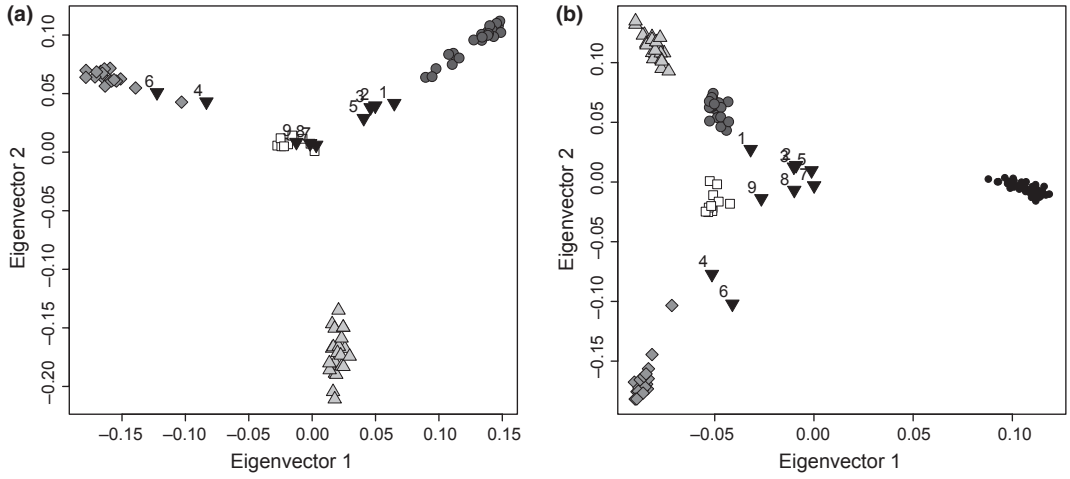

Fig. 2 (a) Principal component analysis (PCA) plot based on 26505 single nucleotide polymorphisms (SNPs) with MAF > 0.05. Four wild boar populations as inferred by STRUCTURE are indicated by different symbols. The nine individuals with putative introgressed SNPs are labelled and numbered explicitly (black inverted triangles). The first two eigenvectors explain 18% of variance in the data set. (b) PCA plot including a sample of all six domestic pig breeds considered in this study (small black dots) in the PCA analysis.

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 *et al.* (2005) was $K = 3$ followed by $K = 4$ (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 $K = 3$ was not geographically coherent. German individuals were divided over the Meinweg and the Veluwe clusters with dubious assignment probabilities (Table S1, Supporting information). We suspect that this may be caused by a relatively low sample size of the German cluster ($n = 11$ vs. $n = 21$, 23 and 24) as well as its wide geographic spread, resulting in high internal variation and lack of Hardy–Weinberg equilibrium. The STRUCTURE partitioning $K = 4$ matches fully to geographic and PCA distributions, and we therefore consider $K = 4$ to be the most biologically meaningful structure of this data set.

We investigated some possible sources of SNP introgression by quantifying the presence of the 5038 putative introgressed SNPs of the wild boar data set in six domestic pig breeds ($n = 20$ per breed) as well as a sample of wild boar from the Balkans ($n = 20$, Table 3). The LW domestic pig breed scored best, sharing approximately 80% of the putative introgressed SNPs. However, differences with other pig breeds were relatively small. Commercial pig farmers commonly use breed hybrids. Therefore, we included some combinations of two breeds ($n = 40$ per combination) in Table 3, which increased the percentage of putative introgressed SNPs explained to 86%. The percentage of shared putative introgressed SNPs between hybrid wild boar from NW Europe and wild boar from eastern Europe was only 20%.

The wild boar in our data set mostly displayed one of three common wild boar mtDNA haplotypes (HP164, HP165 and HP19), with three notable exceptions. These exceptions are individuals with putative introgressed SNPs, which had a mtDNA haplotype not normally observed in wild boar (HP110 and HP8, Table 1). Haplotype HP110 is a rare haplotype among European pigs, because it has an Asian origin (Fig. S5, Supporting information). The British heritage pig breeds and Pietrain are the only breeds in NW Europe that display this haplotype: Berkshire at a frequency of 5%, BS at 54%, Gloucester Old Spot at 40%, Tamworth at 43%, and Pietrain at 1.9% ($n = 593$, Table S2, Supporting information). Haplotype HP8 is typical for a number of mainland Europe pig breeds, including Landrace and

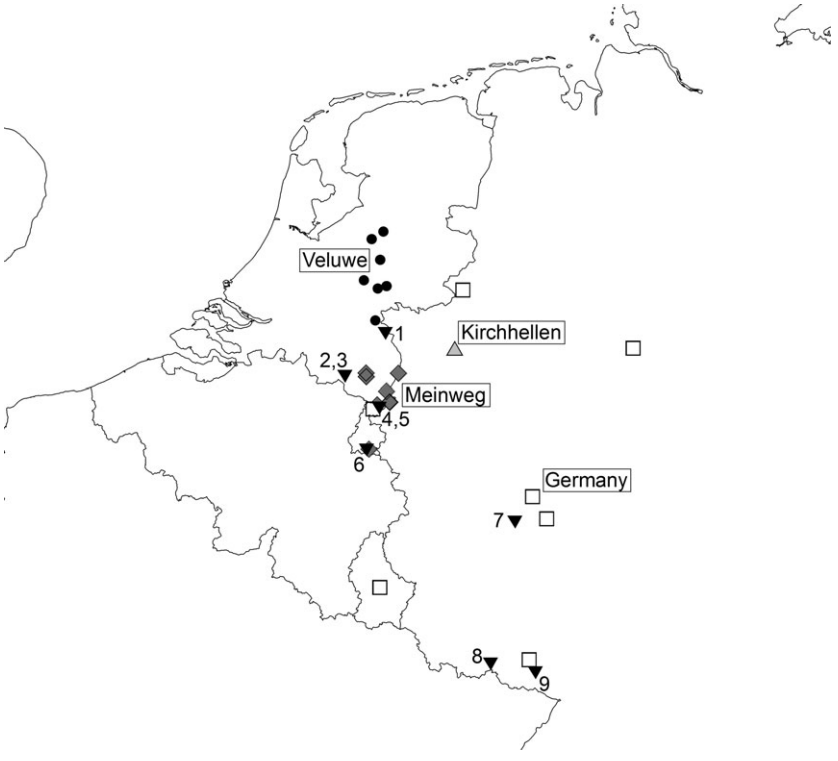

Fig. 3 Geographic sample locations. Symbols and numbering correspond to the principal component analysis (Fig. 2). Multiple samples may originate from one sampling location.

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.

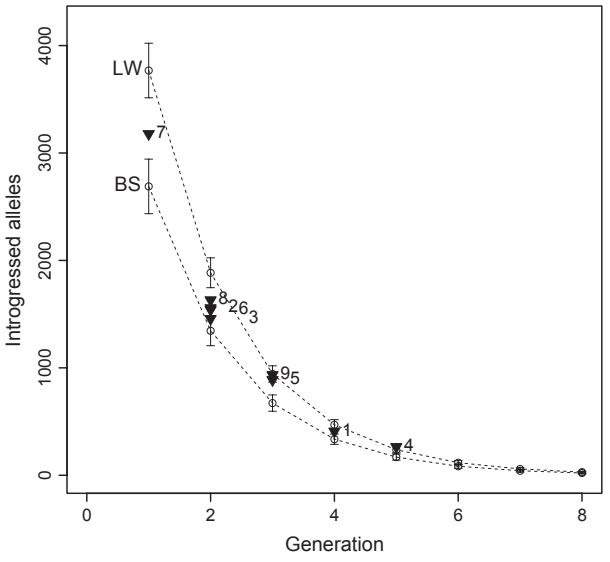

Fig. 4 The open circles connected by dotted lines indicate the simulated mean number of introgressed pig alleles per individual (±SD) per generation of hybridization with Large White or British Saddleback pigs and subsequent backcrossing with wild boar. The number of putative introgressed alleles for each of the nine hybrids in our empirical data set is indicated by inverted triangles. The numbering of hybrids is consistent with Figs 2 and 3.

Discussion

Rare SNPs indicate genetic introgression from domestic pig in wild boar populations

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 clusters and the domestic pig cluster (Fig. 2b). The two observed typical domestic pig mtDNA haplotypes in three of these nine individuals further support a scenario of introgression from domestic pigs.

The proportion of hybrid wild boar in this data set is 10% (Wilson Score 95% Confidence Interval: 5–19%). This is at least as high as previously reported figures (5–10%) for introgression in European wild boar (Scandura *et al.* 2008). High levels of recent introgression in the study area were not expected a priori as intensive indoor pig farming is prevailing in the last decades, and opportunities for direct contact between pigs and wild boar are considered to be minimal. Opportunities for contact between pigs and wild boar were expected to be more prominent in parts of eastern and Mediterranean Europe, where free-ranging pig production in semi-wild conditions is still a common practice (Scandura *et al.* 2008).

Hybridization simulations and genomic distributions of introgressed alleles indicate the level of introgression

The results from the hybridization simulation study indicate that the detected cases of introgression are equivalent to F1 hybrids until F5 backcrosses with wild boar (Fig. 4). The LW hybridization simulation resulted in slightly higher numbers of introgressed alleles, while

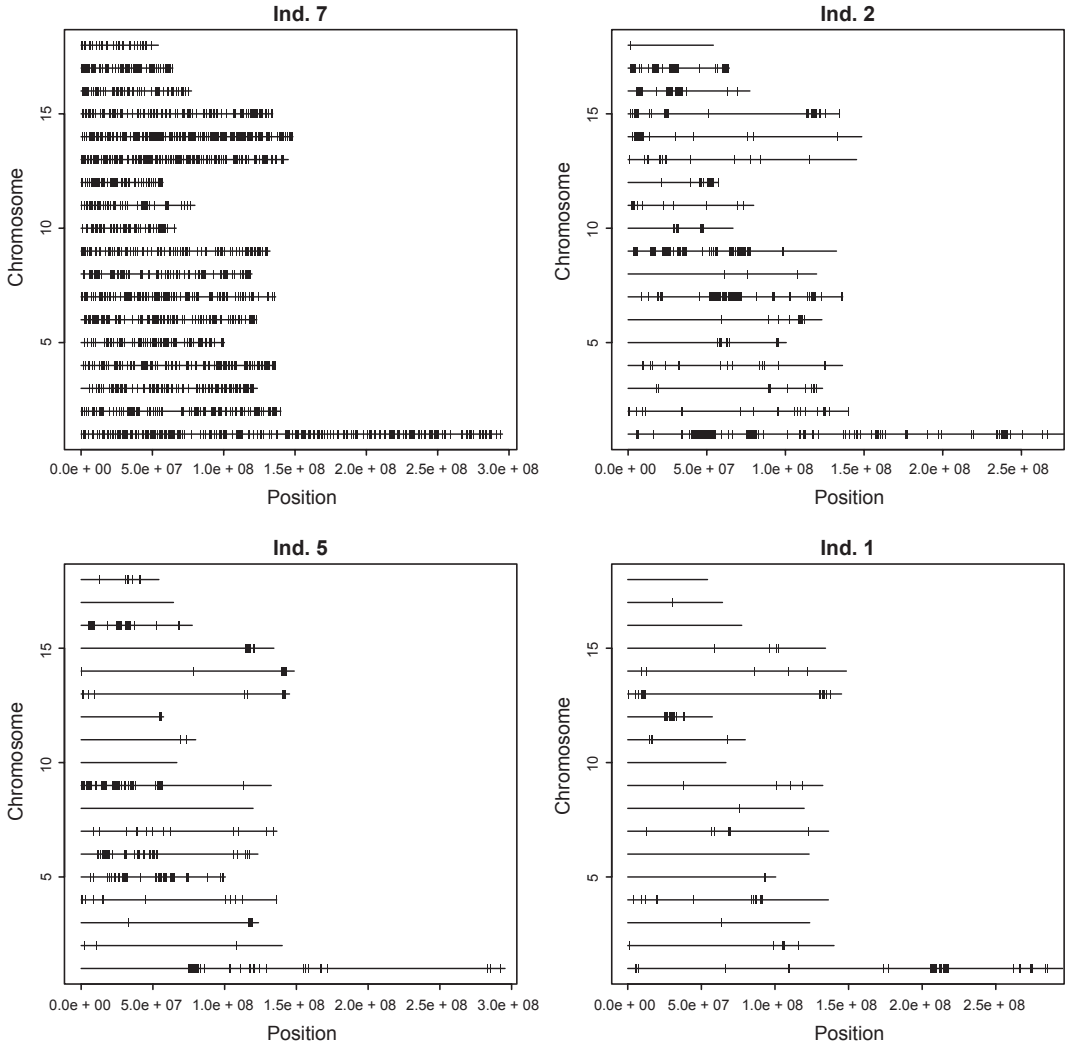

Fig. 5 Chromosomal positions of introgressed single nucleotide polymorphisms. Individual 7 was assigned as an F1 hybrid, individual 2 as an F2 backcross with wild boar, individual 5 as an F3 backcross and individual 1 as an F4 backcross. A complete overview for all identified hybrids is given Fig. S6 (Supporting information).

the BS simulation resulted in slightly lower numbers of introgressed alleles (Fig. 4). This difference is most likely caused by different levels of outbreeding and polymorphism in these breeds, leading to different amounts of nonwild boar alleles that can potentially introgress. Contributions of multiple breeds to the genetic introgression in NW European wild boar populations may have contributed to the observed numbers of introgressed alleles per hybrid wild boar.

Mendelian inheritance and recombination (crossing over) result in the inheritance of chromosomal segments from each parent. In a scenario of hybridization followed by backcrossing with wild boar, one would expect pig alleles to be found only in the chromosomal segments that originate from the parent with domestic pig ancestry. The clustered patterns of introgressed SNPs in individuals 1, 2, 3 and 5 fit this expectation (Fig. 5) and support their assignments as recent hybrids by the hybridization simulation study. Considering a generation time of 1 year for wild boar, we can put these hybridization events in the last few years before sampling in 2008. Clustered patterns of introgressed genetic markers resulting from recent hybridization have to the authors' knowledge not been previously described from natural populations.

Hybrid individuals 4, 6, 8 and 9 display a more widespread distribution of introgressed SNPs across the genome (Fig. S6, Supporting information). This suggests a more complex scenario of reproduction among hybrids (hybrid × hybrid). These individuals are therefore only equivalent to the assigned generations in the hybridization simulation. The actual wild x domestic hybridization may have taken place a number of generations further back in time followed by interbreeding among hybrids,

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 consider the time frame of introgression for these advanced-generation hybrids to be uncertain.

Wild boar number 7 is assigned as a first-generation hybrid. Intuitively one would expect to find a first-generation hybrid at the equidistance between wild and domestic in a PCA. However, one has to keep in mind that in PCA a mean centring procedure is applied. This leads to a gravitation of intermediate individuals (i.e. hybrids) to the origin (0, 0) of the PCA plot, which explains the position of wild boar number 7 at the centre of Fig. 2 rather than of the equidistance between wild and domestic.

We show that genome-wide SNP analysis can reveal the level of introgression (F1–F5 hybrids or equivalent) by identifying putative introgressed SNPs based on allele frequency spectrum analysis, followed by a comparative analysis of the simulated number of introgressed SNPs per individual and the observed number of introgressed SNPs per individual (Fig. 4). Assignments of generations (F1–F5 or advanced-generation hybrids) can be further validated by the identification of introgressed chromosomal segments. These methodologies can be applied to all study systems where large numbers of genome-wide genetic markers are shared between the study taxon and the source of introgression. The growing use of high-density SNP sets has a promising potential to lead to important insights in the processes of hybridization and genetic introgression.

Mechanisms and sources of introgression

The putative introgressed SNPs found in wild boar are by definition polymorphic in domestic pig, because the Illumina porcine SNP60 genotyping beadchip was ascertained on four domestic pig breeds (Duroc, Pietrain, LW and Landrace) and a small sample of wild boar (Ramos *et al.* 2009). A relatively small data set of six domestic pig breeds ($n = 20$ per breed) already accounted for 89% of the additional SNPs found.

The domestic pig breeds included in our analysis shared relatively similar proportions of putative introgressed SNPs (Table 3). Only Duroc and Tamworth displayed lower amounts of shared SNPs and are deemed unlikely to have been involved in the identified cases of introgression. These findings suggest that introgression was not a singular event, but that it occurred on multiple occasions originating from multiple sources or pig breeds. The presence of two distinct pig mtDNA haplotypes that are not found together in any domestic pig breed (Table S2, Supporting information) confirms that multiple sources of introgression were involved.

The commercial LW and Landrace breeds seemed most likely to have contributed to the introgression, as they shared the highest number of SNPs with the nine hybrid wild boar (Table 3). However, these breeds were well represented in the ascertainment pool of the Illumina porcine SNP60 genotyping beadchip. Overestimation of the contributions of these breeds vs. breeds not included in the ascertainment pool is therefore possible. Still, these breeds share far more putative introgressed SNPs with the nine hybrid wild boar than some other breeds included in the ascertainment pool (Duroc and Pietrain). The observed mtDNA haplotype HP8 most likely entered the NW Europe wild boar gene pool through the LW or Landrace breeds, which are the most common commercial breeds in the study area. The observed Asian mtDNA haplotype HP110 most likely originated from one of the traditional British pig breeds, as these are the only breeds in this part of the world that display significant levels of this mtDNA haplotype (Table S2, Supporting information).

Possible mechanisms for introgression are (i) cross-breeding with escaped or field-reared domestic pigs, or (ii) escape/release of already hybridized (farmed) wild boar stock. Farmed wild boar are often cross-bred to a certain extent with a number of domestic pig breeds to increase litter size and piglet growth rates (Goulding 2001). In certain areas of Europe, the documented occurrence of escaped farmed wild boar is substantial (Scandura *et al.* 2011).

Three wild boar (individuals 2, 3 and 5) were hybrids between domestic pigs and wild boar from the Veluwe (Fig. 2), but their geographic sampling locations fell within the range of the Meinweg population (Fig. 3). This finding suggests that the second mechanism, escape/release of hybrid farmed wild boar, has occurred at different places. The observed mtDNA haplotypes of individuals 2, 3 and 5 (HP110 and HP19) suggest that a hybridized farmed wild boar stock with ancestry in the Veluwe wild boar population and British traditional pig breeds is present in NW Europe and that this hybrid farmed wild boar stock has introgressed into some free-living wild boar populations.

The route by which mtDNA haplotype HP8 has entered the wild boar gene pool, which represents a separate hybridization event, remains uncertain. However, the genomic distribution pattern of introgressed SNPs in

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 *et al.* 2008). Recent genetic contributions from eastern European wild boar into the study area are considered to be negligible.

Possible effects of introgression

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 *et al*. 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 growth in the last decades.

Acknowledgements

We thank Jörg Brün, the Animal Health Service Deventer in the Netherlands, as well as the Landesuntersuchungsamt of Rhineland-Palatinate in Germany for sample contributions. Thanks also go to Robert Kraus for the technical support. Finally, we thank the Royal Dutch Hunters Association (KNJV) for the financial support.

References

Alves PC, Pinheiro I, Godinho R *et al.* (2010) Genetic diversity of wild boar populations and domestic pig breeds (*Sus scrofa*) in South-western Europe. *Biological Journal of the Linnean Society*, **101**, 797–822.

Bieber C, Ruf T (2005) 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*, **42**, 1203–1213.

Bonfield JK, Smith KF, Staden R (1995) A new DNA sequence assembly program. *Nucleic Acids Research*, **23**, 4992–4999.

Briedermann L (1990) *Schwarzwild*. VEB Deutscher Landwirtschaftsverlag, Berlin, Germany.

Evanno G, Regnaut S, Goudet J (2005) Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular Ecology*, **14**, 2611–2620.

Fang MY, Andersson L (2006) Mitochondrial diversity in European and Chinese pigs is consistent with population expansions that occurred prior to domestication. *Proceedings of the Royal Society B-Biological Sciences*, **273**, 1803–1810.

Fang M, Berg F, Ducos A, Andersson L (2006) Mitochondrial haplotypes of European wild boars with 2n=36 are closely related to those of European domestic pigs with 2n=38. *Animal Genetics*, **37**, 459–464.

Geisser H, Reyer HU (2005) The influence of food and temperature on population density of wild boar *Sus scrofa* in the Thurgau (Switzerland). *Journal of Zoology*, **267**, 89–96.

Gethoffer F, Sodeikat G, Pohlmeyer K (2007) Reproductive parameters of wild boar (*Sus scrofa*) in three different parts of Germany. *European Journal of Wildlife Research*, **53**, 287–297.

Giuffra E, Kijas JMH, Amarger V *et al.* (2000) The origin of the domestic pig: independent domestication and subsequent introgression. *Genetics*, **154**, 1785–1791.

Glez-Pena D, Gomez-Blanco D, Reboiro-Jato M, Fdez-Riverola F, Posada D (2010) ALTER: program-oriented conversion of DNA and protein alignments. *Nucleic Acids Research*, **38**, W14–W18.

Goulding MJ (2001) Possible genetic sources of free-living Wild Boar (*Sus scrofa*) in southern England. *Mammal Review*, **31**, 245–248.

Jombart T (2008) adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics*, **24**, 1403–1405.

Kraus RHS, Kerstens HHD, Van Hooft P *et al.* (2011) Genome wide SNP discovery, analysis and evaluation in mallard (*Anas platyrhynchos*). *BMC Genomics*, **12**, article no. 150.

Larkin MA, Blackshields G, Brown NP *et al.* (2007) Clustal W and clustal X version 2.0. *Bioinformatics*, **23**, 2947–2948.

Larson G, Dobney K, Albarella U *et al.* (2005) Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science*, **307**, 1618–1621.

Larson G, Albarella U, Dobney K *et al.* (2007) 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*, **104**, 15276–15281.

Luetkemeier ES, Sodhi M, Schook LB, Malhi RS (2010) Multiple Asian pig origins revealed through genomic analyses. *Molecular Phylogenetics and Evolution*, **54**, 680–686.

Miller JM, Poissant J, Kijas JW, Coltman DW (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular Ecology Resources*, **11**, 314–322.

Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. *Trends in Ecology & Evolution*, **19**, 208–216.

Narum SR, Banks M, Beacham TD *et al.* (2008) Differentiating salmon populations at broad and fine geographical scales with microsatellites and single nucleotide polymorphisms. *Molecular Ecology*, **17**, 3464–3477.

Patterson N, Price AL, Reich D (2006) Population structure and eigenanalysis. *PLoS Genetics*, **2**, e190.

Price AL, Patterson NJ, Plenge RM *et al.* (2006) Principal components analysis corrects for stratification in genome-wide association studies. *Nature Genetics*, **38**, 904–909.

Pritchard JK, Stephens M, Donnelly P (2000) Inference of population structure using multilocus genotype data. *Genetics*, **155**, 945–959.

Purcell S, Neale B, Todd-Brown K *et al.* (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. *American Journal of Human Genetics*, **81**, 559–575.

Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald 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.

Saezroyuela C, Telleria JL (1986) The increased population of the wild boar (*Sus-Scrofa* L) in Europe. *Mammal Review*, **16**, 97–101.

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.

Scandura M, Iacolina L, Apollonio M (2011) Genetic diversity in the European wild boar *Sus scrofa*: phylogeography, population structure and wild × domestic hybridization. *Mammal Review*, **41**, 125–137.

Seeb JE, Carvalho G, Hauser L *et al.* (2011) Single-nucleotide polymorphism (SNP) discovery and applications of SNP genotyping in nonmodel organisms. *Molecular Ecology Resources*, **11**(Suppl 1), 1–8.

Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

Verhoeven KJ, Macel M, Wolfe LM, Biere A (2011) Population admixture, biological invasions and the balance between local adaptation and inbreeding depression. *Proceedings of the Royal Society B-Biological Sciences*, **278**, 2–8.

Vignal A, Milan D, SanCristobal M, Eggen A (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics Selection Evolution*, **34**, 275–305.

Willing EM, Bentzen P, van Oosterhout C *et al.* (2010) Genome-wide single nucleotide polymorphisms reveal population history and adaptive divergence in wild guppies. *Molecular Ecology*, **19**, 968–984.

This research is a result of a collaboration between the Resource Ecology Group and the Animal Breeding and Genomics Centre of Wageningen University as well as the Wildlife Research Institute in Bonn. D.J.G., P.v.H., S.E.v.W., R.C.Y. and H.H.T.P. are ecologists with interests in wildlife behavior and natural population processes. H.J.M., J.M.H.M., R.P.M.A.C. and M.G. are animal breeding scientists with interests in animal genomics and production traits. P.A. and W.L. are wildlife biologists with interests in wildlife behavior and conservation.

Data accessibility

The 688 mtDNA D-loop sequences used in this study were submitted to GenBank, accession numbers ranging JQ238239–JQ273541. For more detailed information on these mtDNA D-loop GenBank accession numbers, see Table S3 (Supporting information).

The 45720 autosomal SNP genotypes for 88 wild boars and 120 domestic pigs (PLINK and STRUCTURE file format) were deposited in the Dryad data repository: doi:10.5061/dryad.v6f1g.

Supporting information

Additional supporting information may be found in the online version of this article.

Table S1 Information on individual wild boar samples including sex, age in months, sampling location (national park or municipality), sample source (collecting person or institute), PCA clustering with hybrid identification, STRUCTURE population assignment ($K = 4$) and population assignment probabilities of STRUCTURE $K = 3$.

Table S2 MtDNA haplotypes of pig breeds and wild boar populations.

Table S3 Sample details and GenBank accession numbers of the mtDNA D-loop sequences that formed the basis of the mtDNA haplotypes used in this study (Table S2, Supporting information).

Fig. S1 L(K) indicated by points with standard deviation bars and Delta(K) indicated by triangles connected by a solid line, per K of the performed STRUCTURE runs following the method of Evanno *et al.* (2005).

Fig. S2 Neighbour Joining tree of swine mtDNA D-loop haplotypes.

Fig. S3 Chromosomal positions of introgressed SNPs for all nine identified hybrids.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.