Groenen *Genet Sel Evol (2016) 48:23*
DOI 10.1186/s12711-016-0204-2

GSE Genetics Selection Evolution

REVIEW Open Access

A decade of pig genome sequencing: a window on pig domestication and evolution

Martien A. M. Groenen*

Abstract

Insight into how genomes change and adapt due to selection addresses key questions in evolutionary biology and in domestication of animals and plants by humans. In that regard, the pig and its close relatives found in Africa and Eurasia represent an excellent group of species that enables studies of the effect of both natural and human-mediated selection on the genome. The recent completion of the draft genome sequence of a domestic pig and the development of next-generation sequencing technology during the past decade have created unprecedented possibilities to address these questions in great detail. In this paper, I review recent whole-genome sequencing studies in the pig and closely-related species that provide insight into the demography, admixture and selection of these species and, in particular, how domestication and subsequent selection of *Sus scrofa* have shaped the genomes of these animals.

Background

The domestic pig (*Sus scrofa*) is a member of the family of Suidae, a group of pig species from the order of Cetartiodactyla that originated some 20 to 30 million years ago (Mya) [1, 2]. Of this family, *S. scrofa* (wild boars and domestic pigs) is the only species that was domesticated [2]. Sequencing its genome was initiated with the establishment of the Swine Genome Sequencing Consortium (SGSC) in September 2003 [3], following the successful generation of genetic [4] and physical [5] maps for the pig. The strategy by the SGSC was based on hierarchical shotgun Sanger sequencing of bacterial artificial chromosome (BAC) clones representing a minimal tile path across the genome [6], which was supplemented at a later stage with Illumina next-generation sequencing data [7]. These efforts resulted in the assembly and publication of a draft reference genome sequence of *S. scrofa* in 2012 [8]. In addition to this reference genome sequence, which was derived from a female Duroc pig, the SGSC described the genome sequence of another 48 pigs from a variety of breeds and wild boars [8]. Moreover, an independent genome assembly of a Chinese Wuzhishan minipig, based on Illumina short reads, was published simultaneously [9], followed by additional de novo assembled genomes of a Göttingen minipig [10] and a Tibetan wild boar [11] in 2013. Since then, the genomes of hundreds of individual pigs have been re-sequenced to study genome variation, evolution, and selection in this species [11–25], and currently around 350 complete genomes are publically available (Table 1).

The Suidae family represents a diverse group of species that comprises 15 to 17 extant species that are grouped into five genera [1, 2]. Although currently a de novo assembled genome is only available for *S. scrofa*, their close evolutionary relationship to *S. scrofa* allows the use of its reference genome [8] for analysis of these other species as well. Consequently, the genomes of 10 other species of this family have been sequenced and studied by aligning the sequences against the pig reference genome [26, 27]. Particular care has to be taken using such across-species sequence alignments, since aligning sequences against the genome of another species can result in underestimation of sequence variability. Nevertheless, this can be controlled by carefully choosing the sequence alignment program and the genotype-calling algorithm [27].

Review

Suid speciation

Members of the five genera of the Suidae family can be found across Africa, Europe, and Asia, and over the

*Correspondence: martien.groenen@wur.nl
Animal Breeding and Genomics Centre, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, The Netherlands

Table 1 Pig whole-genome resequencing: overview of pig short read Illumina sequences deposited in the European Nucleotide Archive (1-9-2015)

Accession	Species	Pig breeds/populations	Publications	Number of individuals
PRJEB1683	Ssc, Svr, Sba, Sce, Scb, Paf	EWB, AWB, DU, HA, PI, LR, LW, XI, JQ, MS	[8, 12, 14]	77
PRJNA144099	Ssc	WZ	[9]	1
PRJNA41185	Ssc	DU	–	1
PRJNA176189	Ssc	GM	[10]	1
PRJNA231897	Ssc	RC	–	6
PRJNA186497	Ssc	AWB, TWB, PZ, WJ, YN, NJ, JH	[11]	49
PRJNA238851	Ssc	TC	[23]	5
PRJNA260763	Ssc	DU, LR, YM, KWB, LW	[18, 19]	70
PRJEB9115	Ssc	DU[a]	–	1
PRJNA213179	Ssc	AWB, BX, EH, HT, LA, LU, MI, GA, SC, TP, YN, WZ	[16]	69
PRJNA221763	Ssc	BK	[25]	3
PRJNA239399	Ssc	MA, DU	[15]	4
PRJNA190683	Ssc	IB	–	1
PRJNA255085	Ssc	EWB, GC, IB, MP	[22]	4
PRJEB9326	Ssc, Scb	PI	[24]	18
PRJEB9922	Ssc, Svr	EWB, ASW, WS, ZA, LS, AS, BB, BK, BS, CM, CS, CA, CT, GO, HA, LB, LI, MA, MW, IB, NS, TA, RE, JQ, XI	[20]	102
PRJNA281548	Ssc	BK	[21]	10

Ssc, *Sus scrofa;* Svr, *Sus verrucosus*; Sba, *Sus barbatus*; Sce, *Sus celebensis*; Scb, *Sus cebifrons*; Paf, *Phacochoerus africanus*, AS, Angler Sattleschwein; AWB, Asian wild boar; BB, Bunte Bentheimer; BK, Berkshire; BS, British Saddleback; BX, Bamaxiang; CA, Calabrese; CM, Chato Murciano; CS, Cinta Senese; CT, Casertana; DU, Duroc; EH, Erhualian; EWB, European wild boar; GA, Gansu; GC, Guatemala Creole pig; GM, Göttingen minipig; GO, Gloucester Old Spot; HA, Hampshire; HT, Hetao; IB, Iberian; JH, Jinhua; JQ, Jiangquhai; KWB, Korean wild boar; LA, Laiwu; LB, Large Black; LE, Leicoma; LI, Linderodsvin; LR, Landrace; LS, Leping Spotted; LU, Luchuan; MI, Min; LW, Large White; MA, Mangalica; MP, 16th century pig; MS, Meishan; MW, Middle White; NJ, Neijiang; NS, Nera Siciliana; PI, Pietrain; PZ, Penzhou; RC, Rongchang; RE, Retinto; SC, Sichuan; TA, Tamworth; TC, Tongcheng; TP, Tibetan pig; TWB, Tibetan wild boar; WJ, Wujin; WS, Wannan Spotted; WZ, Wuzhishan; XI, Xiang; YM, Yucatan miniature pig; YN, Yannan; ZA, Zang

[a] Short read sequences of the Duroc pig on which the pig reference genome assembly [8] is based

past decade the genomes of several of these species have been sequenced. Among the suids found in Africa, the genomes of three species from two different genera have been sequenced [27]: *Phacochoerus africanus* (common warthog), *Potamochoerus porcus* (red river hog), and *Potamochoerus larvatus* (bush pig), while the genomes of seven species from three genera have been sequenced from Suids in Asia [five *Sus* sp. [26], *Babyrousa babyrussa* [27] and *Porcula salvania* (pygmy hog; MAM Groenen, unpublished results)]. The largest variety of pig species is found on the Islands of South East Asia (ISEA), which is likely a result of the large number of isolated islands separated by a shallow sea that during glacial periods and low sea levels allowed them to be connected by land bridges [26]. Currently, six different species from this area have been sequenced [12, 26, 27]: *S. scrofa* (from Sumatra), *Sus verrucosus* (Javan warty pig), *Sus barbatus* (bearded pig), *Sus celebensis* (Sulawesi warty pig), *Sus cebifrons* (Visayan warty pig), and *B. babyrussa* (pig-deer). Apart from the pygmy hog from Northern India, all the other wild pigs from Asia that have been sequenced are considered a single species (*S. scrofa*), although phylogenetic analysis based on their sequenced genomes shows that several of these populations diverged more than a million years ago (Fig. 1) [28, 29]. The availability of complete genome sequences of these species and sub-species allowed the reconstruction of a well-supported phylogeny (Fig. 1) that shows that some of the *S. scrofa* populations in Eurasia diverged at around the same time as some of the *Sus* species found in ISEA. In spite of the deep phylogenetic split between species of the genus *Sus*, extensive gene flow has occurred between many of these species through natural dispersal as a result of land bridges during glacial periods [26, 30]. To date, there are no detailed analyses of selective sweeps during the speciation of these species, but a comparison of copy number variation (CNV) between the Asian suids [31] showed that olfactory receptor genes and immune-related genes are among the most rapidly evolving genes. The latter likely reflects adaptation to different pathogens in different environments. It has been suggested [32] that the rapid evolution of the olfactory receptor gene repertoire was not only an adaptation to the new environment but also might have acted as a species barrier by affecting mate choice.

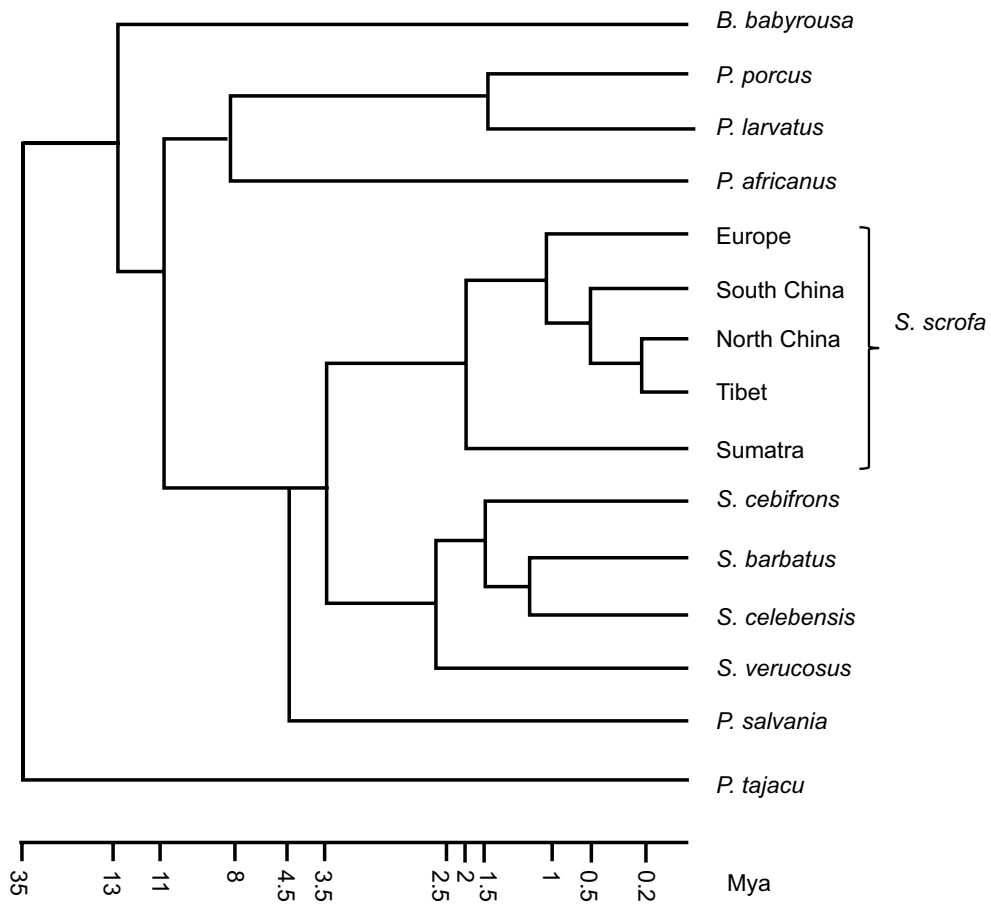

Fig. 1 Schematic phylogenetic tree. The phylogeny diagram summarises the results of phylogenetic studies based on whole-genome sequence data [26–28]. Timing is derived from the estimates described in [26–28, 30]

Dispersal and divergence of *S. scrofa* populations across Eurasia

Sus scrofa (hereafter referred to as wild boar or as pig in the case of the domesticated populations) originated in Southeast Asia some 3 to 4 Mya and over the past one million years colonized almost the entire Eurasian mainland. Wild boars are also found on the Northern parts of Sumatra, although this population diverged from the Eurasian wild boar some 1.5 to 2 Mya [26, 27] (Fig. 1). The European and Asian wild boar populations diverged around 1 Mya [8, 26], resulting in very different minor allele frequencies at millions of genomic locations and over a million locations that are fixed for alternative alleles [8]. In agreement with the CNV analysis of the other suids [31], genes involved in sensory perception, immunity, and host defence were among the most rapidly evolving genes [32]. A selective sweep analysis for the split between European and Asian wild boars, using a method that was originally developed to study human Neanderthal divergence and that is based on identification of regions that share fewer derived alleles [33], identified 251 putative selected regions [8]. Surprisingly, these regions displayed an over-representation of genes involved in RNA splicing and RNA processing, suggesting changes in gene expression and gene regulation during adaptation to novel habitats while the species expanded across Eurasia.

Within Asia, there is a clear split between wild boar populations in the North (North China, Tibet, Japan) and the South (South China), with an estimated divergence time of around 0.5 Mya [26, 27]. A selective sweep analysis between these populations (including both wild boar and domestic breeds from these regions), using locus-specific branch-length analyses [34], indicated that genes involved in biological processes that contribute to maintenance of thermostatic status during heat or cold stress, such as hair growth (*DCAF17*) and blood circulation (*VPS13A*), have been under divergent selection [16]. These results confirmed earlier studies on a de novo assembled genome of a Tibetan wild boar [11]. This

latter study also showed possible selection in Tibetan wild boars for adaptation to high altitude. Examples of this adaptation are genes with vitamin B6 binding activity (*ALB*, *SPTLC2*, and *GLDC*) and genes related to hypoxia (*ALB*, *ECE1*, *GNG2*, and *PIK3C2G*). It was also suggested that the Tibetan wild boar exhibited extensive lineage-dependent gene expansion and contraction (e.g. the olfactory receptor genes) [11]. However, these observations are most likely mainly the result of the draft status of the pig reference genome [8, 35] and collapses of repetitive and duplicated regions in the next-generation sequencing-based genome assembly of the Tibetan wild boar [28, 36].

The X chromosome of European and Asian pigs and wild boars shows a remarkable 48-Mb large region of homozygosity [16, 37] that overlaps with a region that shows almost no recombination [38]. Effectively, this region exhibits two different haplotypes, one found in European and North Chinese pigs and wild boar, and one in pigs and wild boar from South China. In addition, a third haplotype is observed in some North Chinese pigs [16] and Meishan pigs (HJ Megens and MAM Groenen, unpublished results), which appears to be a recombinant haplotype between the first two haplotypes. In fact, the South Chinese haplotype clusters with haplotypes observed in the other *Sus* species from ISEA, which indicates that it may have originated from an admixture event between South Chinese wild boar and another, now extinct, suid. One explanation of the current distribution of this haplotype is that it provided an advantage in Northern latitudes and therefore has swept to fixation, and was subsequently transmitted to European wild boar populations as a result of migrations across Eurasia [16]. Gene flow was also observed between *S. scrofa* and other *Sus* species on ISEA and, although it was originally attributed to human-mediated dispersal of pigs across this area [26], subsequent analyses indicated that it was the result of natural dispersal of these pigs in South East Asia [30]. Furthermore, admixture between *S. scrofa* and *S. verrucosus* was shown to be predominantly the result of gene flow from *S. verrucosus* into *S. scrofa*, mostly likely between 0.5 and 1 Mya [30].

Suid demography

An interesting feature of the availability of whole-genome sequence data from a diploid species' individual is that it allows the inference of its demography up to several hundreds of thousands generations ago (e.g. by using pairwise sequential Markovian coalescence; PSMC [39]). Using PSMC and assuming an equal generation time of 5 years and the same mutation rate of 2×10^{-8} for all *Sus* species, indicated that wild boars from Eurasia and most suids from ISEA experienced a strong bottleneck during the Pleistocene period [8, 26]. These declines in population size are in agreement with the reduction of temperature during this period, which resulted in reduced overall forest cover, which is the natural habitat for these species. For many of the species, in particular *S. scrofa*, the population sizes reached their minimum during the last glacial maximum [8], around 20 thousand years ago (Kya). Pig populations in Europe were the most affected by this climate change and it has been proposed that wild boar populations in Europe retreated into three refugia during that period, i.e. Iberia, Italy and the Balkans [40]. In Asia, wild boar populations in Northern China and Tibet were more affected than populations in South China [26, 28]. Interestingly, the population size of *S. celebensis* on Sulawesi Island appears to have increased dramatically towards the end of the interglacial period that preceded the last glacial maximum 20 Kya [26].

While PSMC allows inference of pig demography up to 1 Mya, it does not provide information on recent (10 Kya to current) demography. However, insight into more recent demography can be obtained from analysis of runs of homozygosity (ROH) in the genome. In particular large ROH are most sensitive to recent population changes. A comparison between European and Asian wild boars showed an on average much larger number of ROH in European wild boars, with clear indications of recent inbreeding [12], which is in agreement with the strong recent bottlenecks that were caused by over-hunting and a decline in suitable habitats in Europe.

The severe population bottleneck during the last glacial period and recent inbreeding resulted in much lower genetic diversity in European wild boar compared to Asian wild boar [8]. This lower genetic diversity might also be explained by a bottleneck due to migration of wild boar from Asia to Europe, although no indication for such a bottleneck is evident from the PSMC analysis [8]. While this difference in genetic diversity is also observed for European and Asian domestic breeds, the recent gene flow from Asian domestic into European domestic breeds [41] resulted in higher nucleotide diversity in European breeds than in the European wild boar (discussed below) [8, 12, 20]. In agreement with these observations, approximate Bayesian computation by [20] showed an expected (modest) decline in effective population size in Asian domestic pig breeds compared to Asian wild boar, while the effective population size in European domestic pig breeds was similar to that of Asian domestic breeds and more than twice as large as that of European wild boar.

Pig domestication

Of all the suids, only wild boars have managed to spread across several continents, which shows that they are extremely adaptable to a wide range of environments and

climates. In that regard, it is interesting that wild boar appears to have been the only suid that was domesticated by humans. The latter might be simply due to its wide distribution across Eurasia, bringing them into frequent direct contact with humans, but it is also possible that the extreme adaptability of this species has contributed to its domestication by humans. Domestication of the pig took place some 9000 to 10,000 years ago independently at two locations: in East Anatolia and in China. Following these initial domestications, pigs accompanied early farmers as they spread from East Anatolia to Europe and throughout China, respectively. The picture that has emerged from analysis of pig mitochondrial genomes [42, 43] and more recently from whole-genome sequence data [20] is that this domestication process was very diffuse, taking many millennia and involving repeated admixture and gene flow from wild boars into the domesticated populations. Therefore, pig domestication should not be considered as a series of fixed events that happened some 10,000 years ago but as a gradual process, in which both animal and humans played their part. Wild boars might have been initially attracted to human settlements as an easy way of accessing food, and it is only after millennia that humans might have actually started to keep pigs as a truly domesticated species. This process occurred in Europe and China at vastly different rates, with pigs being kept in enclosures within human settlements at a relative early stage in China, while in Europe until the late middle ages, pigs were allowed to freely roam the forests as domesticated herds [44]. By the late middle ages, European and Asian domestic pigs were genetically very different for two reasons: (1) they were based on wild boar populations that diverged around 1 Mya ago and differed at over one million positions throughout their genomes [8]; and (2) for thousands of years they were submitted to selection pressures on very different traits. Nevertheless, it is expected that, at the same time, selection acted on similar traits such as behaviour (docility) and morphology (such as coat colour and body size). Indeed, early studies on pooled genome sequence data suggest that selection occurred on genes that affect such traits [37, 45]. Genomic regions and genes that are likely to have been under selection will be discussed in the next section, but there is one other event that played a major role in shaping the genomes of the majority of modern domestic pig breeds. By the late eighteenth- to early nineteenth-century, pig breeding, in particular in the UK, underwent a series of major changes due to growing demands for pig meat as a result of growth of the human population at the time of the industrial revolution. Breeders turned their attention to Asia and imported Chinese pigs to improve their breeding stock [41]. Admixture analysis using D-statistics [46] indicates 35 % Asian contributions to the genomes of modern European commercial breeds [8, 14, 20], while haplotype-based estimates [16, 47] and a partially supervised admixture analysis [48] suggest a 20 % Asian contribution in European domestic pigs. In agreement with historical records, Iberian pig breeds show no signs of admixture with Chinese pigs [13, 48].

Identification of selective sweeps in the pig genome

Domestication and subsequent selection by humans have generated an enormous amount of phenotypic variations that are not seen in the original wild animals. Current pig breeds in Europe and Asia exhibit a vast range of distinct morphological characteristics related to, e.g. body size, coat colour, ear shape, and shape of the skull. Other phenotypes related to reproduction and behaviour of these animals have also changed dramatically compared to their wild progenitors. To understand the molecular mechanisms and to identify the genes that underlie these changes, it is important to distinguish changes during early domestication from more recent changes during breed development and intense selection by breeders. Many of the observed morphological and behavioural changes seen in pigs are also observed in other domestic species [49, 50], which suggests that selection occurred on genes within the same pathways that affect these biological processes in these species. Several of these morphological changes, such as coat colour (spots), floppy ears, and curly tails, have been suggested to be directly correlated with domestication and selection for tameness [50]. Nevertheless, it is likely that many of these characteristics (in particular coat colour) were favoured by some breeders and were strongly selected for in certain breeds after initial domestication. Several mutations with a major effect on coat colour (*KIT* [51], *MC1R* [52]) or lean growth (*IGF2* [53], *RYR1* [54], *PRKAG3* [55]) have been identified. While providing examples of genes that clearly have been under strong selection, the number of genes identified remains small, and they do not provide insight into the changes within the genome during initial domestication of pigs.

The possibility to sequence the genome of multiple individuals has enabled the decision to start to address this in more detail. This has resulted in several studies that used whole-genome sequence data to identify selective sweeps in pigs and numerous methods have been developed to identify such signatures of selection [56]. It is important to realize that different models underlie these methods and, thus, interpretation of the results must be done with caution. Genomic evidence for selection based on these analyses is often suggestive rather than conclusive, and it is often difficult to distinguish drift from true selection. One of the first studies that used next-generation sequence data from populations

towards this aim was based on reduced representation sequencing data that represented only 2 % of the genome [45]. The results from that study suggested selection on genes that affect coat colour, growth, muscle development, olfaction, immunity, and brain development. Among the identified regions, were those that harbour potential candidate genes affecting behaviour, such as the *PPP1R1B* gene on pig chromosome (SSC for *S. scrofa*) 12 and the *LRRTM* gene on SSC2. However, this study lacked sufficient resolution to unequivocally identify specific genes due to the low coverage of the sequence data. The first study that used sequence data covering the complete genome of multiple individuals compared genome sequence data that were generated on pools of European domestic and European wild boars [37]. This approach identified several strong candidate genes that have been under selection and that affect body size (*NR6A1*, *PLAG1*, and *LCORL*) or body composition (*OSTN*). Furthermore, these authors also identified 72 derived nonsynonymous substitutions that approached fixation in domestic pigs, including the Pro192Leu missense mutation in *NR6A1*, which is most likely the causative mutation that resulted in the increased number of vertebrae in domestic pigs. These authors also revealed the staggering complexity of multiple duplications around the *KIT* gene and its potential regulatory sequences, which are responsible for different coat colour phenotypes, such as dominant white, patch, and belt. This provides further evidence that structural changes, many of which affect *cis*-acting regulatory sequences, underlie the observed rapid evolution in domestic animals [57].

Using a combination of genotyping data for 60 K SNPs on a large number of individuals and genome sequence data for a limited number of the individuals, Wilkinson et al. [58] were able to identify a region on SSC1 with evidence of differential selection in domestic breeds compared to wild boar. This selection signal is within a region that contains the *THSB2* and *SMOC2* genes, which have been suggested to affect skull development in dogs [59, 60], thus, they are potential candidates for genes that have been under selection at an early stage during domestication. A number of studies identified selective sweeps in specific breeds, such as the Berkshire breed [21, 25] and the Yucatan minipig [19], and suggested that selection on genes that affect growth and fatness traits occurred in these breeds. However, it is likely that all these represent genes that have been under selection very recently during the formation of these breeds. Furthermore, it remains unclear how many of these genes have only been under selection in the specific populations studied or whether these signals are the result of drift in these populations.

The different origin of the Chinese and European breeds has resulted in many fixed sequence differences in their genomes and distinct sequence variants have been selected for in the European versus the Asian breeds [8]. Examples of selective sweeps of European variants are the sweeps around the *NR6A1*, *PLAG1* and *LCORL* genes [37] described above. A number of studies specifically analysed selective sweeps in Asian pigs [16, 17, 23]. In a study [18] that included the Korean native pig breed, selective sweeps were observed around the *CLDN1* and *TWIST1* genes, which may affect fertility and fatness, respectively. Asian breeds are well known for their high prolificacy, fertility and fatness. Strong selective sweep signals at genes with a potential effect on fertility (*GPR149* and *JMJD1C*) and on genes that are involved in fatness traits were identified in the Chinese Tongcheng breed [23]. The same authors reported that the *MTF* and *EDNRB* genes were responsible for the coat colour phenotype of this breed, which is sometimes referred to as the Huazhong Two-End Black pig. These findings confirm earlier ones on the involvement of the *EDNRB* gene in the coat colour phenotype of Chinese spotted pigs [16] and of the old English breed Gloucester Old Spot [58]. The *EDNRB* variant that was identified in the Gloucester Old Spot breed was shown to be of Chinese origin and to result from the introgression of Chinese pigs into European breeds during the early nineteenth century. Another Chinese variant that was selected for in European breeds after its introgression is the region around the *LEMD3* gene [58], which shows a strong selection signature that is associated with ear morphology (floppy vs. upright ears). Bosse et al. [14, 61] subsequently performed a selective sweep analysis that was specifically designed to identify Asian haplotypes that have been under strong selection after their introgression into European breeds in the early nineteenth century. In agreement with the distinct features that differentiate Asian and European pigs with regard to a number of morphological traits, fertility and fatness, selective sweeps in regions that harbour genes affecting these traits were identified. This specific approach allowed the authors to identify Asian-derived non-synonymous mutations in the *AHR* gene on SSC9 that is associated with increased litter size [14]. These studies also highlighted regions that specifically lacked introgressed Asian haplotypes in the Large White population studied [61], several of which contained genes that are known to have been under selection for European variants (*MC1R*, *KIT*).

With the possible exception of the selective sweep around the *THSB2* and *SMOC2* genes, the genes described above were most likely under selection long after the initial stages of domestication. The traits that are most likely to have been under selection during early domestication are behaviour and increased tameness. Therefore, genes that are involved in brain and neuronal

development are likely candidates to have been under selection during this process. Indeed, several studies [20, 23, 25, 45, 62] reported overrepresentation of genes with GO (gene ontology) terms related to neuronal development and neurological regulation, although care must be taken to avoid over-interpreting the results of such GO analyses. Moreover, the selective sweeps identified in these studies concerned different locations and genes and there was no gene that was consistently identified in multiple studies. Nevertheless, this could also be related to the complex genetic background of traits such as behaviour and increased tameness, as suggested in a study that addressed domestication genes in the rabbit [63]. These authors suggested that in rabbit and other domesticated animals, selection for these traits involved allele frequency shifts at many loci, rather than at a few major domestication loci. Frantz et al. [20] investigated whether parallel selective sweeps at the same loci had occurred in Asian and European domestic pigs. Whereas they did find several such loci, the majority were attributed to the introgression of Asian alleles into European breeds in the nineteenth century. Nevertheless, they identified one locus around position 82.37 Mb on SSC4 (build10.2) that had been independently selected for in European and Asian pigs, although it is located 70 kb from the nearest annotated gene (*PENK*). The lack of annotated genes in the pig genome at this location does not appear to be attributable to the draft state of the pig reference genome because the homologous region in the human genome is also a gene desert, at least with respect to protein-coding genes.

Concluding remarks

A large variety of genes have been identified and implicated as being under strong selection in pigs. Some of these genes exhibit a strong effect and were identified in many studies. Variants of these genes have also been implicated as being under strong selection in other species, which further supports their involvement during selection in pigs. Still, distinguishing true selection from drift remains a challenge that requires further studies. However, so far these studies have clearly shown the importance of a well-designed comparison that includes well-characterized populations and takes historical information on the populations into account. A good example is the well-known and well-documented introgression of Asian haplotypes into European pigs in the late eighteenth- to early nineteenth-century.

However, analysis of DNA samples from current animals will always be limited in what it can teach us about past events. Recent progress in obtaining whole-genome sequence data from ancient DNA samples [22] extracted from fossils, will undoubtedly further revolutionize our ability to more directly provide insight into historical events [43, 64] and to reconstruct selective sweeps in these populations at high resolution.

Although the availability of a pig reference genome has enabled analysis at a resolution not possible before, the draft status of the reference genome hinders or even prevents analysis at many loci in the pig genome [35]. Furthermore, a large number of selective sweeps has been identified at regions that lack annotated genes and many of these regions most likely harbour important regulatory sequences that affect nearby genes. The draft status and incomplete annotation of build 10.2 of the pig reference genome are also highlighted by the recently published improved assembly of the porcine X and Y chromosomes [65]. A comparison between build 10.2 and the new chromosome X assembly reveals many rearrangements in the former and the total number of annotated genes increased from 632 in build 10.2 to 1033 in the new build. Clearly, to fully benefit from the wealth of whole-genome sequence data that continues to be generated, an improved reference genome and improved annotation [66] for all pig chromosomes are essential prerequisites to fully capitalize on this wealth of information.

Acknowledgements

This paper is part of the collection 'ISAFG2015' (6th International Symposium on Animal Functional Genomics, 27–29 July 2015, Piacenza, Italy). The publication of the papers in this collection was partly sponsored by OECD Co-operative Research Programme: Biological Resource Management for Sustainable Agricultural Systems (CRP). Martien Groenen's participation in ISAFG2015 was financed by the OECD Co-operative Research Programme. The opinions expressed and arguments employed in this paper are the sole responsibility of the author and do not necessarily reflect those of the OECD or of the governments of its Member countries.

Competing interests

The author declares that he has no competing interests.

Received: 13 October 2015 Accepted: 16 March 2016
Published online: 29 March 2016

References

1. Ruvinsky A, Rothschild MF, Larson G, Gongora J. Systematics and evolution of the pig. In: Rothschild MF, Ruvinsky A, editors. The genetics of the pig. 2nd ed. Wallingford: CABI; 2011. p. 1–13.
2. Frantz L, Meijaard E, Gongora J, Haile J, Groenen MAM, Larson G. The evolution of Suidae. Ann Rev Anim Biosci. 2016;4:61–85.
3. Schook LB, Beever JE, Rogers J, Humphray S, Archibald A, Chardon P, et al. Swine Genome Sequencing Consortium (SGSC): a strategic roadmap for sequencing the pig genome. Comp Funct Genomics. 2005;6:251–5.
4. Groenen MAM, Schook LB, Archibald AL. Pig genomics. In: Rothschild MF, Ruvinsky A, editors. The genetics of the pig. 2nd ed. Wallingford: CABI; 2011. p. 179–99.
5. Raudsepp T, Chowdhary BP. Cytogenetics and chromosome maps. In: Rothschild MF, Ruvinsky A, editors. The genetics of the pig. 2nd ed. Wallingford: CABI; 2011. p. 134–78.
6. Humphray SJ, Scott CE, Clark R, Marron B, Bender C, Camm N, et al. A high utility integrated map of the pig genome. Genome Biol. 2007;8:R139.

7. Archibald AL, Bolund L, Churcher C, Fredholm M, Groenen MA, Harlizius B, et al. Pig genome sequence–analysis and publication strategy. BMC Genomics. 2010;11:438.
8. Groenen MA, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, et al. Analyses of pig genomes provide insight into porcine demography and evolution. Nature. 2012;491:393–8.
9. Fang X, Mou Y, Huang Z, Li Y, Han L, Zhang Y, et al. The sequence and analysis of a Chinese pig genome. Gigascience. 2012;1:16.
10. Vamathevan JJ, Hall MD, Hasan S, Woollard PM, Xu M, Yang Y, et al. Minipig and beagle animal model genomes aid species selection in pharmaceutical discovery and development. Toxicol Appl Pharmacol. 2013;270:149–57.
11. Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, et al. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. Nat Genet. 2013;45:1431–8.
12. Bosse M, Megens HJ, Madsen O, Paudel Y, Frantz LA, Schook LB, et al. Regions of homozygosity in the porcine genome: consequence of demography and the recombination landscape. PLoS Genet. 2012;8:e1003100.
13. Esteve-Codina A, Paudel Y, Ferretti L, Raineri E, Megens HJ, Silio L, et al. Dissecting structural and nucleotide genome-wide variation in inbred Iberian pigs. BMC Genomics. 2013;14:148.
14. Bosse M, Megens HJ, Frantz LA, Madsen O, Larson G, Paudel Y, et al. Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. Nat Commun. 2014;5:4392.
15. Molnár J, Nagy T, Stéger V, Tóth G, Marincs F, Barta E. Genome sequencing and analysis of Mangalica, a fatty local pig of Hungary. BMC Genomics. 2014;15:761.
16. Ai H, Fang X, Yang B, Huang Z, Chen H, Mao L, et al. Adaptation and possible ancient interspecies introgression in pigs identified by whole-genome sequencing. Nat Genet. 2015;47:217–25.
17. Miura N, Kucho K, Noguchi M, Miyoshi N, Uchiumi T, Kawaguchi H, et al. Comparison of the genomic sequence of the microminipig, a novel breed of swine, with the genomic database for conventional pig. In Vivo. 2014;28:1107–11.
18. Choi JW, Chung WH, Lee KT, Cho ES, Lee SW, Choi BH, et al. Whole-genome resequencing analyses of five pig breeds, including Korean wild and native, and three European origin breeds. DNA Res. 2015;22:259–67.
19. Kim H, Song KD, Kim HJ, Park W, Kim J, Lee T, et al. Exploring the genetic signature of body size in Yucatan miniature pig. PLoS One. 2015;10:e0121732.
20. Frantz LAF, Schraiber JG, Madsen O, Megens HJ, Cagan A, Bosse M, et al. Analyses of Eurasian wild and domestic pig genomes reveal long-term gene-flow and selection during domestication. Nat Genet. 2015;47:1141–8.
21. Jeong H, Song KD, Seo M, Caetano-Anollés K, Kim J, Kwak W, et al. Exploring evidence of positive selection reveals genetic basis of meat quality traits in Berkshire pigs through whole genome sequencing. BMC Genet. 2015;16:104.
22. Ramírez O, Burgos-Paz W, Casas E, Ballester M, Bianco E, Olalde I, et al. Genome data from a sixteenth century pig illuminate modern breed relationships. Heredity (Edinb). 2015;114:175–84.
23. Wang C, Wang H, Zhang Y, Tang Z, Liu B. Genome-wide analysis reveals artificial selection on coat colour and reproductive traits in Chinese domestic pigs. Mol Ecol Resour. 2015;15:414–24.
24. Bosse M, Megens HJ, Madsen O, Crooijmans RP, Ryder OA, Austerlitz F, et al. Using genome-wide measures of coancestry to maintain diversity and fitness in endangered and domestic pig populations. Genome Res. 2015;25:970–81.
25. Li M, Tian S, Yeung CK, Meng X, Tang Q, Niu L, et al. Whole-genome sequencing of Berkshire (European native pig) provides insights into its origin and domestication. Sci Rep. 2014;4:4678.
26. Frantz LA, Schraiber JG, Madsen O, Megens HJ, Bosse M, Paudel Y, et al. Genome sequencing reveals fine scale diversification and reticulation history during speciation in Sus. Genome Biol. 2013;14:R107.
27. Frantz LA. Speciation and domestication in Suiformes: a genomic perspective. Ph.D. thesis, Wageningen University; 2015.
28. Frantz LA, Madsen O, Megens HJ, Schraiber JG, Paudel Y, Bosse M, et al. Evolution of Tibetan wild boars. Nat Genet. 2015;47:188–9.
29. Gongora J, Cuddahee RE, do Nascimento FF, Palgrave CJ, Lowden S, Ho SYW, et al. Rethinking the evolution of extant sub-Saharan African suids (*Suidae, Artiodactyla*). Zool Scr. 2011;40:327–3ww5.
30. Frantz LA, Madsen O, Megens HJ, Groenen MA, Lohse K. Testing models of speciation from genome sequences: divergence and asymmetric admixture in Island South-East Asian *Sus* species during the Plio-Pleistocene climatic fluctuations. Mol Ecol. 2014;23:5566–74.
31. Paudel Y, Madsen O, Megens HJ, Frantz LA, Bosse M, Crooijmans RP, et al. Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. BMC Genomics. 2015;16:330.
32. Paudel Y, Madsen O, Megens HJ, Frantz LA, Bosse M, Bastiaansen JW, et al. Evolutionary dynamics of copy number variation in pig genomes in the context of adaptation and domestication. BMC Genomics. 2013;14:449.
33. Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, et al. A draft sequence of the Neandertal genome. Science. 2010;328:710–22.
34. Shriver MD, Mei R, Parra EJ, Sonpar V, Halder I, Tishkoff SA, et al. Large-scale SNP analysis reveals clustered and continuous patterns of human genetic variation. Hum Genomics. 2005;2:81–9.
35. Warr A, Robert C, Hume D, Archibals AL, Deeb N, Watson M. Identification of low-confidence regions in the pig reference genome (Sscofa10.2). Front Genet. 2015;6:338.
36. Pérez-Enciso M, Burgos-Paz W, Ramos-Onsins SE. On genetic differentiation between domestic pigs and Tibetan wild boars. Nat Genet. 2015;47:190–2.
37. Rubin CJ, Megens HJ, Martinez Barrio A, Maqbool K, Sayyab S, Schwochow D, et al. Strong signatures of selection in the domestic pig genome. Proc Natl Acad Sci USA. 2012;109:19529–36.
38. Ma J, Iannuccelli N, Duan Y, Huang W, Guo B, Riquet J, et al. Recombinational landscape of porcine X chromosome and individual variation in female meiotic recombination associated with haplotypes of Chinese pigs. BMC Genomics. 2010;11:159.
39. Li H, Durbin R. Inference of human population history from individual whole-genome sequences. Nature. 2011;475:493–6.
40. Scandura M, Iacolina L, Crestanello B, Pecchioli E, Di Benedetto MF, Russo V, et al. Ancient vs. recent processes as factors shaping the genetic variation of the European wild boar: are the effects of the last glaciation still detectable? Mol Ecol. 2008;17:1745–62.
41. 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.
42. Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, et al. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. Science. 2005;307:1618–21.
43. Larson G, Liu R, Zhao X, Yuan J, Fuller D, Barton L, et al. Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. Proc Natl Acad Sci USA. 2010;107:7686–91.
44. White S. From globalized pig breeds to capitalist pigs: a study in animal cultures and evolutionary history. Environ Hist. 2011;16:94–120.
45. Amaral AJ, Ferretti L, Megens HJ, Crooijmans RP, Nie H, Ramos-Onsins SE, et al. Genome-wide footprints of pig domestication and selection revealed through massive parallel sequencing of pooled DNA. PLoS One. 2011;6:e14782.
46. Durand EY, Patterson N, Reich D, Slatkin M. Testing for ancient admixture between closely related populations. Mol Biol Evol. 2011;28:2239–52.
47. Bosse M, Megens HJ, Madsen O, Frantz LA, Paudel Y, Crooijmans RP, et al. Untangling the hybrid nature of modern pig genomes: a mosaic derived from biogeographically distinct and highly divergent *Sus scrofa* populations. Mol Ecol. 2014;23:4089–102.
48. Bianco E, Soto HW, Vargas L, Pérez-Enciso M. The chimerical genome of Isla del Coco feral pigs (Costa Rica), an isolated population since 1793 but with remarkable levels of diversity. Mol Ecol. 2015;24:2364–78.
49. Price OP. Behavioral development in animals undergoing domestication. Appl Anim Behav Sci. 1999;65:245–71.
50. Trut L, Oskina I, Kharlamova A. Animal evolution during domestication: the domesticated fox as a model. Bioessays. 2009;31:349–60.
51. Andersson L, Plastow G. Molecular genetics of coat colour variation. In: Rothschild MF, Ruvinsky A, editors. The genetics of the pig. 2nd ed. Wallingford: CABI; 2011. p. 38–50.

52. Fang M, Larson G, Ribeiro HS, Li N, Andersson L. Contrasting mode of evolution at a coat color locus in wild and domestic pigs. PLoS Genet. 2011;5:e1000341.
53. Van Laere AS, Nguyen M, Braunschweig M, Nezer C, Collette C, Moreau L, et al. A regulatory mutation in *IGF2* causes a major QTL effect on muscle growth in the pig. Nature. 2003;425:832–6.
54. Fujii J, Otsu K, Zorzato F, de Leon S, Khanna VK, Weiler JE, et al. Identification of a mutation in porcine ryanodine receptor associated with malignant hyperthermia. Science. 1991;253:448–51.
55. Milan D, Jeon JT, Looft C, Amarger V, Robic A, Thelander M, et al. A mutation in *PRKAG3* associated with excess glycogen content in pig skeletal muscle. Science. 2000;288:1248–51.
56. Vitti JJ, Grossman SR, Sabeti PC. Detecting natural selection in genomic data. Annu Rev Genet. 2013;47:97–120.
57. Andersson L. How selective sweeps in domestic animals provide new insight into biological mechanisms. J Intern Med. 2012;271:1–14.
58. Wilkinson S, Lu ZH, Megens HJ, Archibald AL, Haley C, Jackson IJ, et al. Signatures of diversifying selection in European pig breeds. PLoS Genet. 2013;9:e1003453.
59. Bannasch D, Young A, Myers J, Truvé K, Dickinson P, Gregg J, et al. Localization of canine brachycephaly using an across breed mapping approach. PLoS One. 2010;5:e9632.
60. Quilez J, Short AD, Martínez V, Kennedy LJ, Ollier W, Sanchez A, et al. A selective sweep of >8 Mb on chromosome 26 in the Boxer genome. BMC Genomics. 2011;12:339.
61. Bosse M, Lopes MS, Madsen O, Megens HJ, Crooijmans RP, Frantz LA, Harlizius B, Bastiaansen JW, Groenen MA. Artificial selection on introduced Asian haplotypes shaped the genetic architecture in European commercial pigs. Proc Biol Sci. 2015;282:pii:20152019. doi:10.1098/rspb.2015.2019.
62. Moon S, Kim TH, Lee KT, Kwak W, Lee T, Lee SW, et al. A genome-wide scan for signatures of directional selection in domesticated pigs. BMC Genomics. 2015;16:130.
63. Carneiro M, Rubin CJ, Di Palma F, Albert FW, Alföldi J, Barrio AM, et al. Rabbit genome analysis reveals a polygenic basis for phenotypic change during domestication. Science. 2014;345:1074–9.
64. Larson G, Albarella U, Dobney K, Rowley-Conwy P, Schibler J, Tresset A, et al. Ancient DNA, pig domestication, and the spread of the Neolithic into Europe. Proc Natl Acad Sci USA. 2007;104:15276–81.
65. Skinner BM, Sargent CA, Churcher C, Hunt T, Herrero J, Loveland JE, et al. The pig X and Y Chromosomes: structure, sequence, and evolution. Genome Res. 2016;26:130–9.
66. Andersson L, Archibald AL, Bottema CD, Brauning R, Burgess SC, Burt DW, et al. FAANG Consortium: coordinated international action to accelerate genome-to-phenome with FAANG, the Functional Annotation of Animal Genomes project. Genome Biol. 2015;16:57.