Analyses of pig genomes provide insight into porcine demography and evolution

Citation for published version:

Digital Object Identifier (DOI):
10.1038/nature11622

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
Nature

Publisher Rights Statement:
Available under Open Access.
© 2012 Nature Publishing Group, a division of Macmillan Publishers Limited. All Rights Reserved. partner of AGORA, HINARI, OARE, INASP, ORCID, CrossRef and COUNTER

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Download date: 02. Oct. 2019
Analyses of pig genomes provide insight into porcine demography and evolution

A list of authors and their affiliations appears at the end of the paper

For 10,000 years pigs and humans have shared a complex and close relationship. From domestication to modern breeding practices, humans have shaped the genomes of domestic pigs. Here we present the assembly and analysis of the genome sequence of a female domestic Duroc pig (Sus scrofa) and a comparison with the genomes of wild and domestic pigs from Europe and Asia. Wild pigs emerged in South East Asia and subsequently spread across Eurasia. Our results reveal a deep phylogenetic split between European and Asian wild boars ~1 million years ago, and a selective sweep analysis indicates selection on genes involved in RNA processing and regulation. Genes associated with immune response and olfaction exhibit fast evolution. Pigs have the largest repertoire of functional olfactory receptor genes, reflecting the importance of smell in this scavenging animal. The pig genome sequence provides an important resource for further improvements of this important livestock species, and our identification of many putative disease-causing variants extends the potential of the pig as a biomedical model.

The domestic pig (Sus scrofa) is a eutherian mammal and a member of the Cetartiodactyla order, a clade distinct from rodent and primates, that last shared a common ancestor with humans between 79 and 97 million years (Myr) ago (http://www.timetree.net). Molecular genetic evidence indicates that Sus scrofa emerged in South East Asia during the climatic fluctuations of the early Pliocene 5.3–3.5 Myr ago. Then, beginning ~10,000 years ago, pigs were domesticated in multiple locations across Eurasia (Frantz, L. A. F. et al., manuscript submitted).

Here we provide a high-quality draft pig genome sequence developed under the auspices of the Swine Genome Sequencing Consortium, established using bacterial artificial chromosome (BAC) and whole-genome shotgun (WGS) sequences (see Methods and Supplementary Information). The assembly (Sccrofa10.2) comprises 2.60 gigabases (Gb) assigned to chromosomes with a further 212 megabases (Mb) in unplaced scaffolds (Table 1 and Supplementary Tables 1–3).

Genome annotation
A de novo repeat discovery and annotation strategy (Supplementary Fig. 8) revealed a total of 95 novel repeat families, including: 5 long interspersed elements (LINEs), 6 short interspersed elements (SINEs), 8 satellites and 76 long terminal repeats (LTRs). The relative content of repetitive elements (~40%, Supplementary Figs 9 and 10) is lower than reported for other mammalian genomes. The main repetitive element groups are the LINE1 and glutamic acid transfer RNA (tRNAGlu)-derived SINEs or PRE (porcine repetitive element). The expansion of PRE is specific to the porcine lineage. Phylogenetic analysis of LINE1 and PRE (Supplementary Figs 13 and 14) indicates that only a single lineage of each is currently active and that the main expansion of both LINE1 and PRE occurred in the first half of the Tertiary period. Smaller expansions, particularly in LINE1, have occurred since, but recent activity is very low (Supplementary Information).

Annotation of genes, transcripts and predictions of orthologues was performed using the Ensembl analysis pipeline (Supplementary Information). Twenty-seven of these genes (17%) are known to be actively evolving in mammals because many immune genes were not included in the analysis of 1:1 orthologues, we examined a randomly selected subset of 158 immunity-related pig proteins for evidence of accelerated evolution (Supplementary information).

Table 1 | Assembly and annotation statistics
<table>
<thead>
<tr>
<th>Assembly</th>
<th>Placed</th>
<th>Unplaced</th>
<th>Annotation*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total length</td>
<td>2,596,639,456</td>
<td>211,869,922</td>
<td>21,640 protein-coding genes</td>
</tr>
<tr>
<td>Ungapped length</td>
<td>2,323,671,356</td>
<td>195,490,322</td>
<td>380 pseudogenes</td>
</tr>
<tr>
<td>Scaffolds</td>
<td>5,343</td>
<td>4,562</td>
<td>2,965 ncRNAs†</td>
</tr>
<tr>
<td>Contigs</td>
<td>73,524</td>
<td>168,358</td>
<td>197,675 gene exons</td>
</tr>
<tr>
<td>Scafoid N50</td>
<td>637,332</td>
<td>98,022</td>
<td>26,487 gene transcripts</td>
</tr>
<tr>
<td>Contig N50</td>
<td>80,720</td>
<td>24,232</td>
<td></td>
</tr>
</tbody>
</table>

* Numbers refer to the annotation performed by Ensembl (release 67). Results of an independent annotation by the NCBI can be obtained from http://www.ncbi.nlm.nih.gov/mapview/stats/BuildStats.cgi?taxid=9823&build=4&ver=1.
† An improved ncRNA annotation with 3,601 ncRNAs and structured elements is available as a separate track in Ensembl version 70 and for download from http://rth.dk/resources/rnannotator/susscr102. N50, 50% of the genome is in fragments of this length or longer.

Evolution of the porcine genome
Evolution of genes and gene families
To examine the mutation rate and type of protein-coding genes that show accelerated evolution in pigs, we identified ~9,000 as 1:1 orthologues within a group of six mammals (human, mouse, dog, horse, cow and pig). This orthologous gene set was used to identify proteins that show accelerated evolution in each of these six mammalian lineages (Supplementary Information). The observed number of synonymous substitutions per synonymous site (dS) for the pig lineage (0.160) is similar to that of the other mammals (0.138–0.201) except for the mouse (0.458), indicating similar evolutionary rates in pigs and other mammals. The observed dN/dS ratio (ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions) of 0.144 is between those of humans (0.163) and mice (0.116), indicating an intermediate level of purifying selection pressure in the pig. Genes showing increased dN/dS ratios in each lineage were analysed using DAVID to examine whether these rapidly evolving genes were enriched for specific biological processes. Most lineages show different fast-evolving pathways, but some pathways are shared (Fig. 1).

Immune genes are known to be actively evolving in mammals. Because many immune genes were not included in the analysis of 1:1 orthologues, we examined a randomly selected subset of 158 immunity-related pig proteins for evidence of accelerated evolution (Supplementary information). Twenty-seven of these genes (17%) show accelerated evolution in each of the six mammalian lineages (Supplementary Information). The observed number of synonymous substitutions per synonymous site (dS) for the pig lineage (0.160) is similar to that of the other mammals (0.138–0.201) except for the mouse (0.458), indicating similar evolutionary rates in pigs and other mammals. The observed dN/dS ratio (ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions) of 0.144 is between those of humans (0.163) and mice (0.116), indicating an intermediate level of purifying selection pressure in the pig. Genes showing increased dN/dS ratios in each lineage were analysed using DAVID to examine whether these rapidly evolving genes were enriched for specific biological processes. Most lineages show different fast-evolving pathways, but some pathways are shared (Fig. 1).

Immune genes are known to be actively evolving in mammals. Because many immune genes were not included in the analysis of 1:1 orthologues, we examined a randomly selected subset of 158 immunity-related pig proteins for evidence of accelerated evolution (Supplementary information). Twenty-seven of these genes (17%) show accelerated evolution in each of the six mammalian lineages (Supplementary Information). The observed number of synonymous substitutions per synonymous site (dS) for the pig lineage (0.160) is similar to that of the other mammals (0.138–0.201) except for the mouse (0.458), indicating similar evolutionary rates in pigs and other mammals. The observed dN/dS ratio (ratio of the rate of non-synonymous substitutions to the rate of synonymous substitutions) of 0.144 is between those of humans (0.163) and mice (0.116), indicating an intermediate level of purifying selection pressure in the pig. Genes showing increased dN/dS ratios in each lineage were analysed using DAVID to examine whether these rapidly evolving genes were enriched for specific biological processes. Most lineages show different fast-evolving pathways, but some pathways are shared (Fig. 1).

Due to the rapidly evolving nature of immune genes, this analysis focused on all genes, including those that are not usually considered immune genes and those that are not expressed in the immune system. The results indicate that a large number of genes involved in immune response and regulation show accelerated evolution in each of the six mammalian lineages, consistent with the hypothesis that these genes have evolved rapidly in response to the selective pressures of infection and disease. This suggests that the immune system of pigs may have evolved more rapidly in response to environmental factors than in other mammals, potentially providing insights into the evolution of immune responses in humans and other species.
demonstrated accelerated evolution (Supplementary Table 8). A parallel analysis of 143 human and 145 bovine orthologues revealed very similar rates of evolution (18% in human and 12% in cattle, respectively). Using a branch-site analysis, we detected accelerated evolution of amino acids in PRSS12, CD1D and TRAF3 specific to pig (positive selection on pig branch), as well as amino acids in TREM1, IL1B and SCARA5 specific to pig and cow (positive selection on the cetartiodactyl branch).

Further analysis of porcine immune genes (Supplementary Table 5) revealed evidence for specific gene duplications and gene-family expansions (Supplementary Tables 6 and 7). The analysis of this second cetartiodactyl genome indicates that some expansions are cetartiodactyl-specific (cathelicidin) whereas others are ruminant/bovine-specific (β-defensins, C-type lyzozymes) or potentially porcine-specific (type I interferon, δ subfamily).

Pigs have at least 39 type I interferon (IFN) genes, which is twice the number identified in humans and significantly more than in mice. We also detected 16 pseudogenes in this family. Cattle have 51 type I IFNs (13 pseudogenes), indicating that both bovine and porcine type I IFN families have undergone expansion. This is particularly important for interferon subtypes δ (IFND), ω (IFNW) and τ (IFNT); pigs and cattle are evolving species-specific subtypes of IFND and IFNT, respectively. Both species are expanding the IFNW family and share many more IFNW isoforms than other species. Thus, expansion of interferon genes is not ruminant-specific as proposed earlier, although duplication within some specific sub-families seems to be either bovine- or porcine-specific.

Within the immunity-related genes annotated, we found evidence for duplication of six immune-related genes: IL1B, CD36, CD68, CD163, CRP and IFIT1, and one non-immune gene, RDH16. The CD36 gene is also duplicated in the bovine genome, whereas the IL1B gene duplication, where evidence for a partial duplication was reported previously, is unique in mammals. Other key immune genes in the major histocompatibility complex, immunoglobulin, T-cell receptor and natural killer cell receptor loci have been characterized in detail (Supplementary Information).

Another significant porcine genome expansion is the olfactory receptor gene family. We identified 1,301 porcine olfactory receptor genes and 343 partial olfactory receptor genes. The fraction of pseudogenes within these olfactory receptor sequences is the lowest observed in any species so far. This large number of functional olfactory receptor genes most probably reflects the strong reliance of pigs on their sense of smell while scavenging for food.

Conservation of synteny and evolutionary breakpoints
Alignment of the porcine genome against seven other mammalian genomes (Supplementary Information) identified homologous synteny blocks (HSBs). Using porcine HSBs and stringent filtering criteria, 192 pig-specific evolutionary breakpoint regions (EBRs) were located. The number of porcine EBRs (146, Supplementary Table 11 and Supplementary Fig. 16) is comparable to the number of bovine-lineage-specific EBRs (100) reported earlier using a slightly lower resolution (500 kilobases (kb)), indicating that both lineages evolved with an average rate of ~2.1 large-scale rearrangements per million years after the divergence from a common cetartiodactyl ancestor ~60 Myr ago. This rate compares to ~1.9 rearrangements per million years within the primate lineage (Supplementary Table 11). A total of 20 and 18 cetartiodactyl EBRs (shared by pigs and cattle) were detected using the pig and human genomes as a reference, respectively.

Pig-specific EBRs were enriched for LTR endogenous retrovirus 1 (LTR-ERV1) transposons and satellite repeats (Supplementary Table 12), indicating that these two families of repetitive sequences have contributed to chromosomal evolution in the pig lineage. Different families of transposable elements seem to have been active in the cetartiodactyl ancestor. The cetartiodactyl EBRs are enriched for LINE1 elements and tRNA<sup>Glu</sup>-derived SINEs. tRNA<sup>Glu</sup>-derived SINEs, previously found over-represented in cetartiodactyl EBRs defined in the bovine genome, originated in the common ancestor of cetartiodactyls. Our observation that these elements are also enriched in porcine EBRs strongly supports the hypothesis that active transposable elements promote lineage-specific genomic rearrangements.

A stringent set of porcine to human one-to-one orthologues using the MetaCore database revealed that porcine EBRs and adjacent intervals are enriched for genes involved in sensory perception of taste (P < 8.9 × 10<sup>-6</sup>; FDR < 0.05), indicating that taste phenotypes may have been affected by events associated with genomic rearrangements. Pigs have a limited ability to taste NaCl<sup>22</sup>. SCNN1B, a gene encoding a sodium channel involved in the perception of salty tastes, is located in a porcine-specific EBR. Another gene, ITPR3, encoding a receptor for inositol triphosphate and a calcium channel involved in the perception of umami and sweet tastes, has been affected by the insertion of several porcine-specific SINE mobile elements into its 3′ untranslated region (3′ UTR), consistent with our observation of a higher density of transposable elements in EBRs. In addition to 8 bitter taste receptor genes annotated by Ensembl and which were used in the gene enrichment analysis, we identified 9 intact genes, to give a total number of 17 TAS2R receptors in the pig (Supplementary Table 13). This compares to 18 intact bitter taste receptors in cattle, 19 in horse, 15 in dog and 25 in humans<sup>23,24</sup>. Of the 14 bitter taste receptor genes that were mapped to 18 intact bitter taste receptors in cattle, 19 in horse, 15 in dog and 25 in humans<sup>23,24</sup>, 10 were found near 2 EBRs on SSC5 and SSC18 (Supplementary Tables 13 and 15). We also found that at least four taste receptor genes (TAS1R2, TAS2R1, TAS2R40 and TAS2R39) have been under relaxed selection (Supplementary Information). Pigs are not sensitive to bitter tastes and tolerate higher concentrations of bitter compounds than humans<sup>25,26</sup>. Thus, pigs can eat food that is unpalatable to humans. A review of the porcine taste transduction network (Supplementary Fig. 17) revealed additional genes affected by rearrangements that affect ‘apical and taste
receptor cell’ processes. Together with the observed over-representation of genes related to ‘adrenergic receptor activity’ and ‘angiotensin and other binding’ categories in the pig EBRs (Supplementary Fig. 18), our data indicate that chromosomal rearrangements significantly contributed to adaptation in the suid lineage.

Population divergence and domestication
Divergence between Asian and European wild boar
We investigated the evolution within *Sus scrofa* in Eurasia by sequencing ten individual unrelated wild boars from different geographical areas. In total, 17,210,760 single nucleotide polymorphisms (SNPs) were identified among these ten wild boars. The number of SNPs segregating in the four Asian wild boars (11,472,192) was much higher than that observed in the six European wild boars (6,407,224) with only 2,212,288 shared SNPs. This higher nucleotide diversity was visible in the distribution of heterozygous sites of the Asian compared to the European wild boar genomes (Fig. 2). Phylogenomic analyses of complete genome sequences from these wild boars and six domestic pigs revealed distinct Asian and European lineages (Supplementary Fig. 23) that split during the mid-Pleistocene 1.6–0.8 Myr ago (Calabrian stage, Frantz, L. A. F. et al., manuscript submitted). Colder climates during the Calabrian glacial intervals probably triggered isolation of populations across Eurasia. Admixture analyses (Supplementary Information) within Eurasian *Sus scrofa* disclosed gene flow between the northern Chinese and European populations consistent with pig migration across Eurasia, between Europe and northern China, throughout the Pleistocene. Our demographic analysis on the whole-genome sequences of European and Asian wild boar genomes (Fig. 3) revealed an increase in the European population after pigs arrived from China. During the Last Glacial Maximum (LGM; ~20,000 years ago)26, however, Asian and European populations both suffered population bottlenecks. The drop in population size was more pronounced in Europe than Asia (Fig. 3), suggesting a greater impact of the LGM in northern European regions and probably resulting in the observed lower genetic diversity in modern European wild boar.

The deep phylogenetic split between European and Asian wild boars is further supported by our observation of 1,272,737 fixed differences between the six European and four Asian wild boars, 1,706 of which are non-synonymous mutations in 1,191 different genes. Genes involved in sensory perception, immunity and host defence were among the most rapidly evolving genes (Supplementary Table 28), further strengthening the conclusions from our analysis of immunity-related pig proteins. This conclusion is further supported by our observation that these genes are also enriched in porcine segmental duplications (Supplementary Information).

Figure 2 | Distribution of heterozygosity for individual pig genomes. Shown is the distribution of the heterozygosity as the log2(SNPs) per 10k bin.

To investigate further whether specific regions in the genome of European and Asian wild boar have been under positive selection, a selective sweep analysis was performed on the ten wild boar genome sequences using an approach similar to that recently described in the comparison of Neanderthal and *Homo sapiens* genomes27. Regions in the genome under strong positive selection after the divergence of these two populations are expected to share fewer derived alleles. Using stringent criteria (Supplementary Information), we identified a total of 251 putative selective sweep regions, with an average size of 111,269 base pairs (bp), together comprising around 1% of the genome and harbouring 365 annotated protein-coding genes (Supplementary Table 26). Many of these regions (112) do not contain any currently annotated protein-coding exons. In contrast, the 10 putative selective sweep regions located between positions 39–43 Mb on SSC6 together harbour 93 genes. This SSC6 region (Supplementary Fig. 25) is directly adjacent to the centromere and exhibits low recombination rates28. Low recombining regions have been shown to be more prone to selective sweeps and meiotic drive29,30. Although similar large putative selective sweep regions close to the centromere were only observed on SSC6 between positions 56.2–57.5 Mb, on most chromosomes selective sweep regions tended to cluster in the central part of chromosomes, thus exhibiting a clear correlation with regions of low recombination (Supplementary Fig. 27). As expected, regions with the highest nucleotide differentiation between European and Asian wild boars were observed in high recombination regions towards the end of the chromosomes on both metacentric and acrocentric chromosomes31.

The putative selective sweep regions displayed significant over-representation of genes involved in RNA splicing and RNA processing, indicating possible changes in the regulation of genes at the level of RNA processing (Supplementary Table 27). Several of these genes (*CELF1*, *CELF6*, *WDR83*, *RBM39*, *RBM6*, *HNRNPA1*, *HNRNPM*) are involved in alternative splicing, and, small differences in expression might affect alternatively spliced transcripts of specific genes. Evolution of regulatory splicing factors such as heterogeneous ribonucleoprotein particle (hnRNP) proteins has been proposed as an evolutionary model for alternative splicing32, and genetic variation in such factors can affect alternative splicing and result in different phenotypes or disease33. Our observation that specific genes involved in splicing show accelerated evolution in the pig lineage (Fig. 1) supports this hypothesis. Of particular interest is the selective sweep region observed at position 26 Mb on SSC3 around the *ERI2* gene (Fig. 4), which encodes ERI1 exononuclelease family member 2. Different gene variants have...
been fixed in European and Asian wild boar coding for proteins that differ at two amino acid positions: Cys52Arg and His358Leu encoded by exons 3 and 9 of the ER12 gene, respectively. The precise function of ER12 is unknown but the ER1 exoribonuclease family members have been shown to be involved in mRNA decay32 and in Caenorhabditis elegans ERI-1 has been shown to be involved in the degradation of microRNAs (miRNAs)33.

Figure 4 | Putative selective sweep region around the ER12 gene on SSC3. The y axis shows the log-transformed ratio of the observed/expected derived allele frequency using a sliding window at a bin size of 50,000 bp. The x axis shows the position on SSC3 in base pairs.

Independent domestication and admixture events in domestic breeds
A phylogenetic tree constructed using four European wild boar and domestic pigs and six East Asian wild boar and domestic pigs revealed a clear distinction between European and Asian breeds, thus substantiating the hypothesis that pigs were independently domesticated in western Eurasia and East Asia1. An admixture analysis revealed European influence in Asian breeds, and a ~35% Asian fraction in European breeds (Supplementary Table 24). These results are consistent with the known exchange of genetic material between European and Asian pig breeds35. We also observed that European breeds form a paraphyletic clade, which cannot be solely explained by varying degrees of Asian admixture (Supplementary Information). Within each continent, our analysis revealed different degrees of relatedness between breeds and their respective wild relatives (Supplementary Table 20).

During domestication, pigs were often allowed to roam in a semi-managed state and recurrent admixture between wild and domesticated individuals was not uncommon, especially in Europe35. Thus, the most likely explanation for the paraphyletic pattern seen in domestic individuals is a long history of genetic exchange between wild and domestic pigs.

The pig as a biomedical model
The pig is an important biomedical model and the ability to generate transgenics and knockouts in combination with somatic nuclear cloning procedures has resulted in a number of models for specific human diseases36. Naturally occurring mutations also offer opportunities to use pigs as biomedical models37,38. To explore the potential for natural models further, predicted porcine protein sequences were compared with their human orthologues. We observed 112 positions where the porcine protein has the same amino acid that is implicated in a human disease (Supplementary Table 29). Most of these changes in humans have been shown to increase risk in multifactorial traits such as obesity (ADRB3, SDC3) and diabetes (PPP1R8A, SLC30A8, ZNF615) or shown to result in relatively mild phenotypes (for example, dyslexia: KIAA0319) or late-onset diseases such as Parkinson’s disease (LRRK2, SNCA) and Alzheimer’s disease (TUBD1, BLMH, CEPI92, PLAU). These porcine variants are of interest, as they will allow detailed characterization in an experimental model organism whose physiology is very similar to that of human.

Among 32,548 non-synonymous mutations identified by sequencing 48 individual pigs, representing 8 different European and Asian breeds and wild boars39, 6 protein variants implicated in human disease were identified (Supplementary Table 30). In addition, another 157 nonsense mutations in 142 genes were identified, 11 of which have also been implicated in human disease (Supplementary Table 31). Most of these 11 variants were only observed in a heterogeneous state and those for which homozygous individuals were observed probably result in either a mild phenotype (ASS1, mild form of citrullinaemia in humans) or in phenotypes unlikely to affect the fitness of wild boars (RBBP8, pancreatic carcinomas). Our estimate for the average number of nonsense mutations per individual (~30) is smaller than that observed in humans40 despite the observed threefold higher nucleotide diversity in pigs39. This is in agreement with the higher effective population size in the pig compared to that for the human population, which exhibited a strong bottleneck followed by an exponential increase in size during recent history41.

When considering pig-to-human xenotransplantation, porcine endogenous retroviruses (PERVs) pose a risk of zoonotic infection. The pig genome contains fewer endogenous retroviruses than many vertebrates, including humans and mice, and most PERVs were characterized as defective. However, the potential risk posed by activation of rare replication-competent PERVs and defective PERVs by recombination remains, as shown for murine ERVs (XMRV)42. Most PERVs consist of γ and γ-like groups (68%), with β-retroviral ERVs comprising the second most abundant group (Supplementary Fig. 15). Our phylogetic study shows a particularly close relationship between the most intact γ group of PERVs (γ1) and murine γ-ERVs, suggesting a potential recent instance of murine-to-porcine transmission of γ1 ERVs (Supplementary Fig. 15). We identified 20 almost intact PERV γ1 loci (Supplementary Table 10), none of which contained a complete set of gag, pol or env open reading frames, indicating that these proviruses are not replicable. We also identified four β-retroviral PERVs, each containing defects, primarily in env. These were distantly related to intracisternal type A particle (IAP) proviruses of mice and the mouse mammary tumour virus (MMTV)-like (HML) proviruses of humans. None of the above loci was shared in more than 120 pigs tested, indicating considerable PERV polymorphisms.

Conclusion
The draft pig genome sequence reported here has illuminated the evolution of Sus scrofa and confirmed its speciation in South East Asia and subsequent domestication at multiple regions across Eurasia. The high-quality annotated reference genome sequence has already proven to be a critical framework for comparing individual genomes39,43,44 and its value is further illustrated in associated papers published elsewhere (http://www.biomedcentral.com/series/swine). The genome sequence also provides a valuable resource enabling effective uses of pigs both in agricultural production and in biomedical research.

METHODS SUMMARY
Assembly. We constructed a hybrid de novo assembly based primarily on sequences from BAC clones sequenced clone-by-clone and supplemented with Illumina whole-genome shotgun (WGS) reads. BAC clones were selected from the high-resolution physical (BAC contig) map with CHORI-242 library clones prepared from DNA from a single Duroc sow (Duroc 2-14) chosen preferentially. The WGS sequence data were generated using DNA isolated from the same animal. The BAC-derived sequence data were assembled into sequence contigs using Phrap on a clone-by-clone basis and subsequently independently assembled WGS contigs (Supplementary Information) were used to extend BAC clone-derived sequence contigs and to close gaps between clone-derived contigs. Further details and other methods are described in Supplementary Information.


The final assembly (Sscrofa10.2) is deposited in the public sequence databases (GenBank/EMBL/DDBJ) under accession number Sscrofa10.2. The primary source of the Sscrofa10.2 assembly is the NCBI ftp site (ftp://ftp.ncbi.nih.gov/genbank/genomes/Eukaryotes/vertebrates_mammals/sus_scrofa/Sscrofa10.2/). The chromosomes are CM000812–CM000830 and CM001155. They are built from 5,343 placed scaffolds, with Genbank accession numbers CM000812–CM001155 and HJ108403–HJ118999. The final assembly (Sscrofa10.2) has been deposited in the reference assemblies GLR92100–GLR966682 and HJ118403–HJ118999. The final assembly is available at www.nature.com/reprints. The authors declare competing financial interests. Readers are welcome to comment on the paper. The authors declare competing financial interests. Readers are welcome to comment on the paper.

The authors declare competing financial interests. Readers are welcome to comment on the paper.

The authors declare competing financial interests. Readers are welcome to comment on the paper.

The authors declare competing financial interests. Readers are welcome to comment on the paper.