



# The Visayan Warty Pig (*Sus cebifrons*) Genome Provides Insight Into Chromosome Evolution and Sensory Adaptation in Pigs

Langqing Liu ,<sup>\*,1,2</sup> Hendrik-Jan Megens,<sup>1</sup> Richard P.M.A. Crooijmans,<sup>1</sup> Mirte Bosse,<sup>1</sup> Qitong Huang,<sup>1,3</sup> Linda van Sonsbeek,<sup>4</sup> Martien A.M. Groenen ,<sup>1</sup> and Ole Madsen<sup>1,\*</sup>

<sup>1</sup>Animal Breeding and Genomics, Wageningen University & Research, Wageningen, The Netherlands

<sup>2</sup>Division of Evolutionary Biology, Faculty of Biology, Ludwig-Maximilians-Universität Munich, Munich, Germany

<sup>3</sup>Center for Animal Genomics, Agricultural Genome Institute at Shenzhen, Chinese Academy of Agricultural Sciences, Shenzhen 518124, China

<sup>4</sup>Rotterdam Zoo (diergaarde blijdorp), PO Box 532, 3000AM Rotterdam, The Netherlands

\*Corresponding authors: E-mails: Langqing.Liu@biologie.uni-muenchen.de; ole.madsen@wur.nl.

Associate Editor: Aida Ouangraoua

## Abstract

It is largely unknown how mammalian genomes evolve under rapid speciation and environmental adaptation. An excellent model for understanding fast evolution is provided by the genus *Sus*, which diverged relatively recently and lacks postzygotic isolation. Here, we present a high-quality reference genome of the Visayan warty pig, which is specialized to a tropical island environment. Comparing the genome sequences and chromatin contact maps of the Visayan warty pig (*Sus cebifrons*) and domestic pig (*Sus scrofa*), we characterized the dynamics of chromosomal structure evolution during *Sus* speciation, revealing the similar chromosome conformation as the potential biological mechanism of frequent postdivergence hybridization among *Suidae*. We further investigated the different signatures of adaptive selection and domestication in Visayan warty pig and domestic pig with specific emphasize on the evolution of olfactory and gustatory genes, elucidating higher olfactory diversity in Visayan warty pig and positive and relaxed evolution of bitter and fat taste receptors, respectively, in domestic pig. Our comprehensive evolutionary and comparative genome analyses provide insight into the dynamics of genomes and how these change over relative short evolutionary times, as well as how these genomic differences encode for differences in the phenotypes.

**Key words:** de novo genome assembly, speciation, chromosomal rearrangement, adaptation, domestication.

## Introduction

Changes in the environment can initiate the process of speciation, where exposure to new environmental selective forces directs populations toward new evolutionary trajectories (Templeton 2008; Coyne and Orr 2009). To respond to altered conditions there must be sufficient genetic variation for the population to reach adaptation through natural selection. Revealing the genotype–phenotype relationships of adaptive traits deepen our understanding of aspects like anatomy, metabolism, behavior, and ecology.

Ever since Darwin's time, pairs of closed related insular and continental species have often been used to study adaptive trajectories. The genus *Sus*, pigs and hogs, comprises at least seven morphologically and genetically well-defined species. Except for *S. scrofa*, which inhabit the Eurasian continent, all other *Sus* species are restricted to Islands of Southeast Asia (ISEA). Recent findings showed that these species diverged during the late Pliocene (4–2.5 Ma), as a result of isolation on different islands of

ISEA due to climate fluctuations accompanied by sea-level changes (Liu et al. 1992; Frantz et al. 2013). One of the representative species of ISEA *Sus* is the Visayan warty pig (*Sus cebifrons*), which is named after the islands it lives on, and the three pairs of fleshy warts on the face of boars. The species lives in social groups in dense tropical island forested areas and has an omnivorous diet of wide variety (Clauss et al. 2008; Melletti and Meijaard 2017). Contrary to the endemic occurrence of *S. cebifrons*, the wild boar and domesticated pig (*S. scrofa*) have a widespread geographical distribution throughout Eurasia, with domesticated pigs found on all continents, except polar regions. The widespread distribution has recently been suggested to have been facilitated by admixture events with local pig species during the natural expansion of the species from Southeast Asia to Europe (Liu et al. 1992; Frantz et al. 2013). In addition, *S. scrofa* has also been interacting with humans during the process of domestication. Having been domesticated 10,000–9,000 years ago independently in East Anatolia and in China (Larson et al. 2005; Larson et al. 2010; Frantz et al. 2015), pigs have experienced long-

© The Author(s) 2022. Published by Oxford University Press on behalf of Society for Molecular Biology and Evolution.

This is an Open Access article distributed under the terms of the Creative Commons Attribution-NonCommercial License (<https://creativecommons.org/licenses/by-nc/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is properly cited. For commercial re-use, please contact [journals.permissions@oup.com](mailto:journals.permissions@oup.com)

Open Access

term artificial selection for various traits including production, fertility, and disease resistance. The phenotypic and ecological differences between these two sister species in the genus *Sus* provide an excellent opportunity to study the molecular mechanism of speciation and adaptation by comparative evolutionary genomics.

Since publication of the reference genome of domesticated pig in 2012 (Groenen et al. 2012), the number of *S. scrofa* genomes available has increased dramatically during the last couple of years (Fang et al. 2012; Li et al. 2013, 2017; Vamathevan et al. 2013). However, to date, no high-quality reference genome has been reported for other species in the *Suidae* family. Here we provide a high-quality genome assembly for the Visayan warty pig. By comparative genome analysis, various evolutionary events and mechanisms were studied. Specifically, we investigated genomic changes that provide insight into the evolution of reproductive barriers. Furthermore, we investigated genome differences related to species-specific traits, such as the adaptive evolution of smell and taste.

## Results

### Genome Sequencing, Assembly, and Quality Evaluation

For the Visayan warty pig genome assembly, 10× genomics, Hi-C, and mate-pair sequences were employed. From *k*-mer analyses, the size of the Visayan warty pig genome was estimated to be around 2.25 Gb (supplementary fig. S1, Supplementary Material online). The genome sequence was constructed by a hybrid assembly approach (see Material and Methods), which yielded a genome assembly showing a contig N50 of 159.6 Kb, a scaffold N50 of 141.8 Mb, an assembled genome size of 2.48 Gb, and comprised of 1585 scaffolds (table 1, supplementary fig. 2 and table 1, Supplementary Material online). The assembly of the mitochondrial genome sequence resulted in a 16.56 kb, circular, assembly (supplementary fig. 3, Supplementary Material online). Features of the genome assembly are shown in fig. 1a.

**Table 1.** Assembly and Annotation Statistics of the Draft Genome of Visayan Warty Pig.

<i>Assembly features</i>	
Total length (bp)	2,459,328,531
Scaffold count	1,585
Longest scaffold (bp)	289,860,557
N50 of scaffold (bp)	141,782,568
N50 of contig (bp)	159,795
GC ratio (%)	41.78
<i>Genome annotation</i>	
Number of putative coding genes	21,153
Average gene model length (bp)	2554.18
Average CDS length (bp)	1,640.66
Average gene exon length (bp)	220.89
Average exon number per gene	11.56
Average gene intron length (bp)	4,565.05
Total size of TEs (bp)	1,079,874,097
TEs in genome (%)	43.91

We mapped the previously published Illumina short reads sequences from the same individual to the assembled genome. A total of 99.73% of the short sequence reads covered 99.41% of the genome assembly. The low percentage of homozygous single nucleotide polymorphisms (0.008% of the whole genome) reflects the sequence accuracy of the genome assembly. The Visayan warty pig assembly was also assessed for completeness using Benchmarking Universal Single-Copy Orthologs (BUSCO) (Simão et al. 2015). The BUSCO analysis, based on a mammal-specific gene set, indicated 95.7% of the genes to be complete, suggesting a high completeness of the assembly (supplementary table 2, Supplementary Material online).

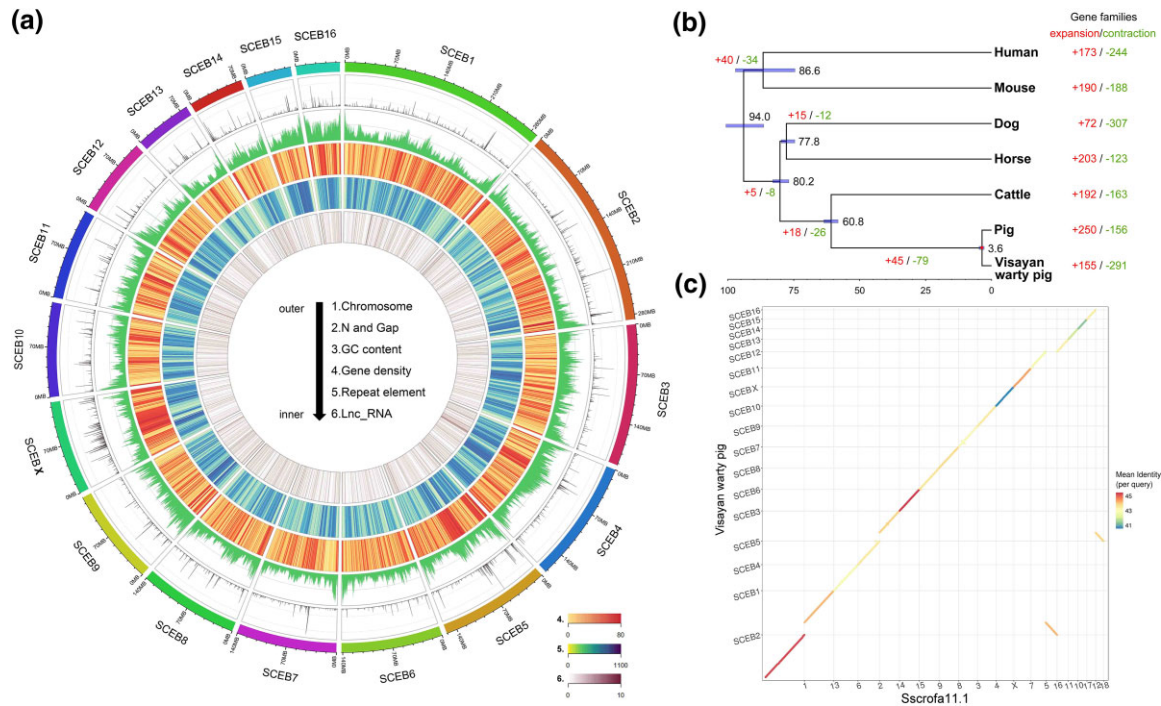
### Repetitive Sequences, Telomeres, and Centromeres

The number of different repeat classes and the divergence distribution of these repetitive elements were identified in the Visayan warty pig genome (supplementary table 3 and fig. 4, Supplementary Material online, respectively). In total, 43.9% of the Visayan warty pig genome consists of repetitive elements (43.8% in domesticated pig and 52.6% in human). The main repetitive transposable elements (TEs) were long interspersed nuclear elements (LINEs) (~511 Mb, covering 20.8% of genome).

Putative telomeres were identified at the proximal ends of Visayan warty pig chromosome assemblies of SCEB1–4, SCEB6, SCEB8, SCEB9, and SCEB10–15 (supplementary fig. 5 and table 4, Supplementary Material online) and putative centromeres were identified in the Visayan warty pig chromosome for SCEB1–10, SCEB12, SCEB13, and SCEB15 (supplementary fig. 5 and table 5, Supplementary Material online), by searching for concentrations of signature repeats. Both chromosomes SCEB7 and SCEB13 harbored two centromeric repeat regions, which consist with observations for the *Duroc* genome (*S. scrofa* chromosomes [SSC]8 and SSC11) (Warr et al. 2020). However, on SCEB6 we only identified one centromeric region, whereas there are two in the homologous SSC15. Chromosomes SCEB1–10, SCEB12, and SCEB13 are metacentric, while chromosome SCEB15 is acrocentric, which are consistent with homologous chromosomes in *Duroc* genome.

### Genome Comparison of Structural Variations Between *S. cebifrons* and *S. scrofa*

Genome structural variations, such as insertions, deletions, inversions, and duplications, are important sources of the genetic diversity that shapes phenotypic variations. We identified 52,827 deletions and 59,299 insertions (INDELS) with a length >100 bp compared with the *S. scrofa* (*Duroc*) assembly. The average lengths of the deletions (374 bp) and insertions (342 bp) are very similar (supplementary fig. 6, Supplementary Material online), and 54.2% of the INDELS are associated with TEs (more than 10 bp overlapping). The TE distribution patterns for insertions and deletions are distinct. Short interspersed



**Fig. 1.** Overview of the assembly characteristics of the Visayan warty pig genome. (a) Circular diagram depicting the characteristics of the Visayan warty pig genome. The tracks from outer to inner circles are indicated in the legend (density of coding gene are shown in track 4). (b) Divergence times and history of orthologous gene families. Numbers on the nodes represent divergence times, with the error range shown in blue bar. The numbers of gene families that expanded (red) or contracted (green) in each lineage after speciation are shown on the corresponding branch. (c) Whole-genome alignment between Visayan warty pig and Duroc pig (Sscrofa11.1).

nuclear elements have the highest proportion associated with INDELs (38.3% for insertion, 30.6% for deletion). LINES have the second-highest proportion of deletions accounting for 27.0%, whereas simple sequence repeats have the largest number of insertions representing ~23.9% (supplementary fig. 7 and table 6, Supplementary Material online).

### Genome Annotation of the Visayan Warty Pig

RNA-seq data from the Visayan warty pig and the proteomes of six additional mammalian species were used for the structural and functional gene annotations. By using an integrative annotation approach (see Material and Methods), we identified 21,153 protein-coding genes (supplementary table 7, Supplementary Material online). Among the 21,153 protein-coding genes, 20,750 (99%) are functionally annotated with Gene Ontology (GO) terms. Gene annotation of mitochondria was performed using MitoZ (Meng et al. 2019) and 13 protein-coding genes as well as 21 tRNA genes are annotated (supplementary table 8, Supplementary Material online).

Of the predicted protein-coding genes in Visayan warty pig genome, 17,175 genes (81.2%) have clear one-to-one orthologs in the Duroc pig genome, and 14,643 genes (69.2%) in the human genome. We compared the gene models of Visayan warty pig with other mammalian species. The structural feature of genes, such as CDS, exon, and intron lengths and exon numbers, is highly conserved

within the mammalian lineages, and Visayan warty pig has very similar gene features as the Duroc pig (supplementary fig. 8 and table 7, Supplementary Material online).

In the Visayan warty pig genome, we predicted 21,467 noncoding RNA transcripts, including 11,767 long noncoding RNAs (lncRNAs). Based on miRBase (<http://www.mirbase.org>), hairpin sequence alignments, 1037 miRNA loci were identified. Based on alignment evidence from Ensembl noncoding RNA sequences (release 100), ~20,515 noncoding RNA genes in Visayan warty pig had a reciprocal best hit (RBH) or second-best hit against the Duroc pig (supplementary table 9, Supplementary Material online).

We estimated the number of expanded and contracted gene families among the six mammalian genomes (fig. 1b, supplementary tables 10–12, Supplementary Material online). In the Visayan warty pig, Duroc pig and the ancestral *Sus* genus branch, the most prominent gene family involved in gene expansion is related to olfaction.

### Phylogenetic and Divergence Time Analysis

We obtained a high-confidence orthologous gene set of Visayan warty pig together with pig, cattle, horse, dog, mouse, and human. From 18,586 gene family clusters we obtained 10,057 high-confidence single-copy gene families which were used for phylogenetics analysis (supplementary figs. 9–10, Supplementary Material online). The phylogenetic tree displays high support values for all

nodes and the time of divergence estimates for the different nonpig outgroups are in line with previous studies (fig. 1b, supplementary fig. 11, Supplementary Material online) (Frantz et al. 2013, 2014). The time tree indicates that Visayan warty pig and the Duroc pig diverged  $\sim 3.6$  Ma and that the evolutionary rate in Visayan warty pig was  $\sim 4.1 \times 10^{-10}$  mutations per generation, which was significantly lower than that of the Duroc pig ( $4.78 \times 10^{-10}$ , Student's *t*-test,  $P < 0.01$ ) (supplementary fig. 12, Supplementary Material online).

### Contiguity and Chromosomal Rearrangements

The chromosome-level Visayan warty pig genome assembly allows a detailed comparison to the assembled domesticated pig genome (Sscrofa11.1; Warr et al. 2020) derived from a Duroc sow, to assess the evolutionary dynamics between these closely related species. We assessed the scaffold orderings of these two genomes through whole-genome sequence alignments (fig. 1c). Our Visayan warty pig assembly reveals a high degree of collinearity with the Duroc pig assembly which is based on PacBio sequencing. Besides the strong overall collinear relationship between the two genomes, previously described chromosomal rearrangements between the two species were confirmed (O'Brien et al. 2007; Musilova et al. 2010) (supplementary fig. 13, Supplementary Material online). The karyotype of Visayan warty pig ( $2n = 34$ ) reveals centric fusions of pig chromosomes SSC13/16 and SSC14/18. In the Visayan warty pig assembly, these fused chromosomes are the largest and third largest chromosomes, respectively. Chromosome contact maps of Visayan warty pig and domesticated pig each were assessed from Hi-C data. The comparison of Hi-C interaction maps for homologous chromosome pairs shows obvious similarities (supplementary fig. 14, Supplementary Material online) and the fused chromosomes in Visayan warty pig retained the same chromosome interaction patterns as the corresponding Duroc pig homologous chromosomes (fig. 2). The interchromosomal interactions in the fused chromosome SCEB1 (SSC13/16) are mainly restricted to within the two chromosome arms, whereas SCEB3 (SSC14/18) shows more interactions between the fused chromosomes. We further quantified the similarity of interaction matrixes between Visayan warty pig and Duroc pig using Phylo-HMRF (Yang et al. 2019). Around 23.5% of the total interactions are categorized as highly or medium conserved (fig. 2, supplementary fig. 15, Supplementary Material online, see Materials and Methods).

### Evolution of Functional Elements and Genes

Conservation of DNA sequences across distantly related species reflects functional constraints. We first generated a multiple genome alignment from Visayan warty pig, Duroc pig, cattle, horse, dog, mouse, and human genomes. Among the aligned mammalian genomes, we predicted 16.6 million highly conserved elements (HCEs) at a resolution of 20 bp or greater, spanning 7.5% of the Visayan

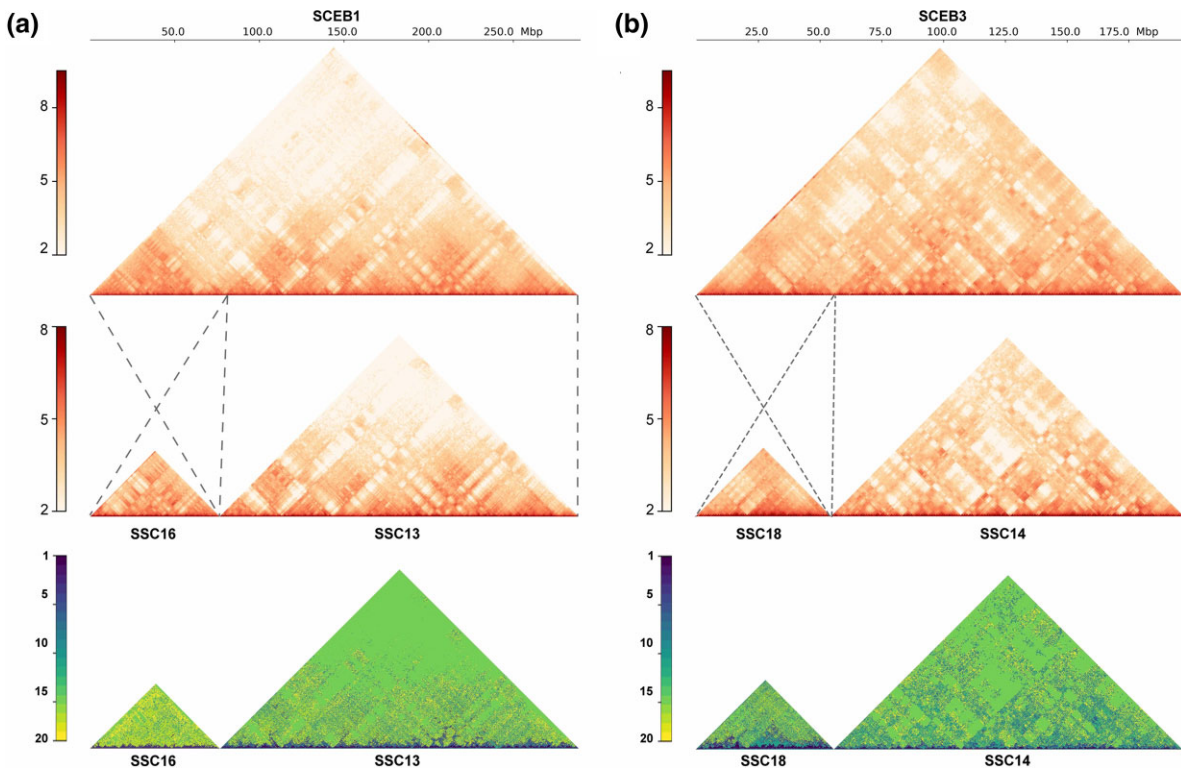
warty pig genome. Functional annotations revealed that  $\sim 3.5\%$  of these HCEs is associated with protein-coding genes, whereas the majority of the remaining HCEs are located in intron and intergenic regions (supplementary fig. 16, Supplementary Material online). We further subset the *Sus* genus-specific HCEs (SHCEs). Noncoding HCEs may play important roles in the regulation of gene expression. We compared the transcription factor binding sites (TFBSes) found in the ENCODE project with these SHCEs. In total, 37.9% (49,846) of the SHCEs contain known TFBSes (supplementary table 13, Supplementary Material online). To investigate evolutionary rate on genes, we identified positively selected genes (PSGs) and rapidly evolving genes (REGs) for the 10,057 one-to-one orthologs. Functional enrichment analyses of the PSGs and REGs in Duroc pig and the ancestral *Sus* genus branch both exhibit enrichment in immune response and energy metabolism functions (supplementary tables 14–21, Supplementary Material online). Specifically, we observed a series of PSGs and REGs in Visayan warty pig involved in the regulation of growth, adaptive thermogenesis, and chromatin organization (supplementary fig. 17 and tables 22–25, Supplementary Material online). In addition, we also compared the identified PSGs in Duroc pig with those reported in previous studies (Groenen et al. 2012; Li et al. 2013) and found very few overlapping. This discordance will be further discussed below.

### Genome Differences Related to Species and Genus Characteristics

In mammals, the chemosensory system mediates many behaviors such as locating food, mating, and shelter. *Suidae* species do not have particularly good vision. Thus, they heavily utilize their olfactory and gustatory sensation (Sutherland-Smith 2015). For example, olfactory receptor (OR) gene duplications are evident in the genomes of domesticated pig, which potentially enhances the adaptations to novel habitats (Truong Nguyen et al. 2012; Paudel et al. 2015). Having distinct evolutionary history, habitat, and diet, a comparison of Visayan warty pig and Duroc pig can provide insight in the selection of genomes leading to species-specific traits. In the meantime, comparing two *Sus* genomes with other mammalian genomes will elucidate the sensory characteristics in the *Sus* genus.

#### Evolution of Olfaction

In the Visayan warty pig genome, we identified 1,163 OR genes of which 921 are functional ORs. This number of functional ORs is approximately equivalent to the number found in the Duroc pig (1,213 ORs; 1,089 functional ORs) and is more than seen in most other mammals (fig. 3a, supplementary table 26, Supplementary Material online). Based on the OR families we estimated the dynamics of ORs and found evidence for OR expansion at the ancestral branch of the *Sus* genus, leading to Visayan warty pig and Duroc pig (supplementary fig. 18, Supplementary Material online). This suggests that an expanded OR repertoire



**FIG. 2.** Chromosome interactions in homologous chromosome pairs. Hi-C contact map of (a) SCEB1 and (b) SCEB3, with signal scale displayed at the left. The units are the chromosomal interaction number on a Log10 scale. Dash lines highlight the homologous region between Visayan warty pig (top panel) and Duroc pig (middle panel). The darker color in the Hi-C contact map represents higher contact frequency. Bottom panel shows cross-species Hi-C contact frequency states identified by Phylo-HMRF. The darker color represents less conserve state between two species.

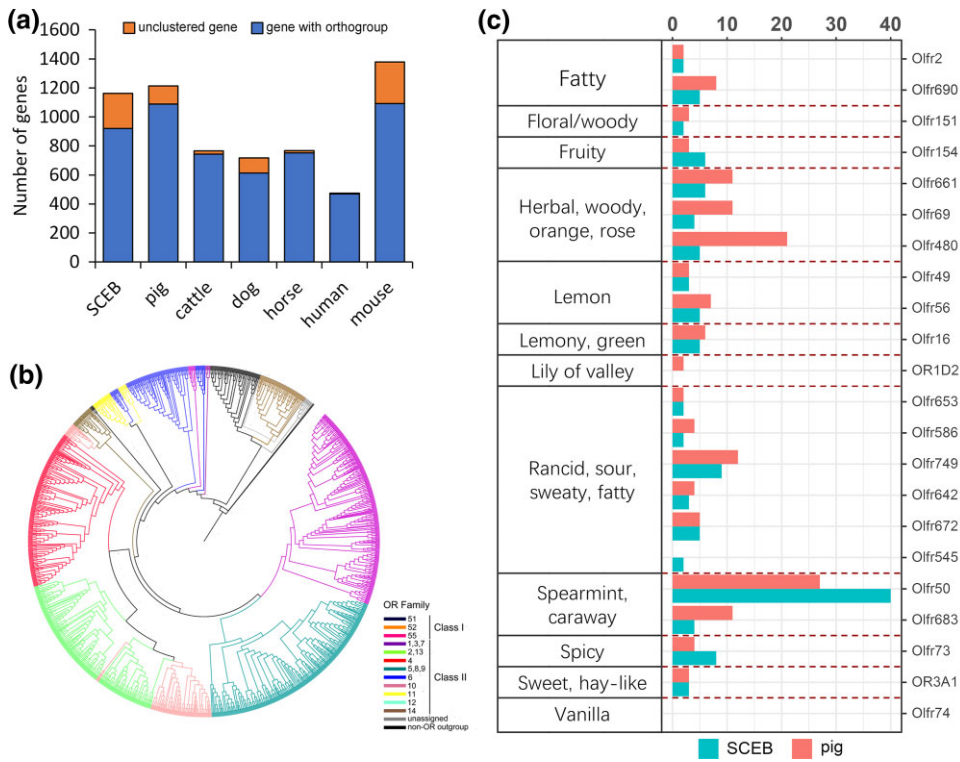
became important already in the ancestral *Sus* species. The diversity of the OR gene repertoire in the Visayan warty pig is represented by different families. Similar to Duroc pig (Truong Nguyen et al. 2012), the identified Visayan warty pig OR genes could be clustered into 2 classes and 17 families, and further grouped into 428 subfamilies, based on phylogenetic analyses and their sequence similarity (fig. 3b).

To identify the potential target specificity of the OR genes in the two pig species, the amino acid sequences of translated Visayan warty pig OR genes and Duroc pig OR genes were compared with those of other species with annotated information on odorant specificity. Using Nguyen et al.'s methods (Truong Nguyen et al. 2012), two sequences were considered to have the same odorant recognition if their identity reached more than 60%. Of the 921 Visayan warty pig functional OR genes, 121 ORs matched to an OR with a known odorant target (fig. 3c). We found that expanded ORs in Visayan warty pig are mainly involved in spearmint, caraway, spicy, and fruit odor detection, whereas Duroc pig is more enriched in herbal, woody, orange, rose, and fatty odor detection. Additionally, we calculated the amino acid diversity of each OR gene family. In general, OR genes in Duroc pig show higher diversity than those in Visayan warty pig (supplementary table 27, Supplementary Material online).

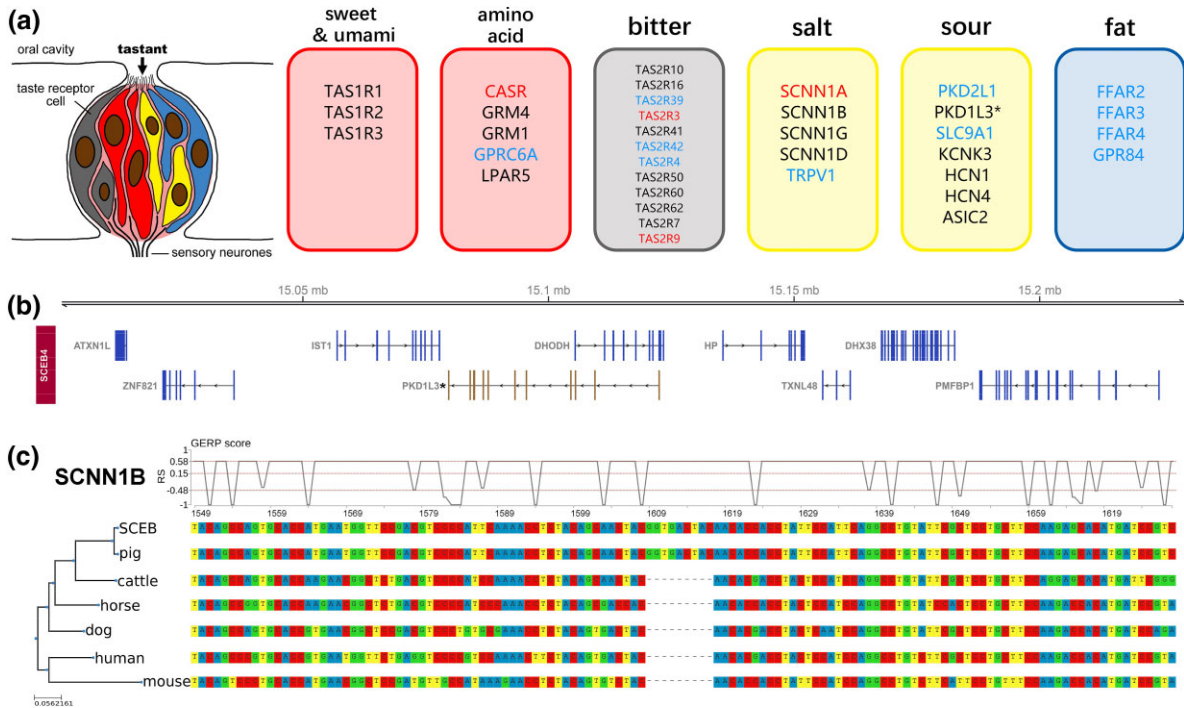
### Evolution of Taste

Taste perception affects diet preference and consequently food/feed intake. The mammalian gustatory system usually discriminates five major basic taste classes: salty, sour, sweet, umami, and bitter. In addition, as fat is an essential ingredient in feed, fat taste is one of the taste types with high potential implications relevant to the swine industry. We did a systematic study on taste transduction-related genes in the two pig species (fig. 4a).

For the 35 well-characterized gustatory genes (Danilova et al. 1999; Glaser et al. 2001; Huang et al. 2006; Ishimaru et al. 2006; Bachmanov and Beauchamp 2007; Horio et al. 2011; Roura et al. 2013; Li and Zhang 2014), we carried out a comprehensive evolutionary analysis (supplementary tables 28 and 29, Supplementary Material online). Among all the bitter taste receptor genes, only *TAS2R42* and *TAS2R39* had significantly elevated dN/dS ratios in the Visayan warty pig, whereas relaxed constrain is found in human and positive selection is found in mouse (supplementary table 28, Supplementary Material online). The cluster of genes sensing amino acids showed a high proportion of members under relaxed selection in Visayan warty pig, Duroc pig, horse, and cattle (supplementary table 29, Supplementary Material online). All fatty acid sensors in the Duroc pig are under relaxed selection, whereas none of them shows the same pattern in other mammalian species.



**FIG. 3.** Evolution of Olfactory gene family. (a) Shared and specific OR genes in Visayan warty pig and six other mammalian species. (b) Phylogenetic tree constructed with OR protein sequences of Visayan warty pig showing classification of ORs in Visayan warty pig. The different family members are colored according to the colors shown in the legend of b. (c) Potential association between olfactory receptor gene and odorant recognition.



**FIG. 4.** Genomic features related to taste transduction. (a) Diagram of the taste bud cell and related taste receptor genes. The red font indicates higher selection pressure present in Duroc pig than Visayan warty pig. The blue font indicates higher selection pressure present in Visayan warty pig than Duroc pig. Black Asterisk indicates a putative pseudogene. (b) Genome track for Visayan warty pig. PKD1L3 is nested with DHODH. Functional genes are shown in blue, whereas the putative pseudogene is shown in brown. (c) Multispecies DNA alignment of SCNN1B, showing the region with *Sus* genus-specific insertion. GERP scores are shown on top of the alignment. Dashed lines represent upper quartile, average, and lower quartile, respectively.

Of the six identified sour taste receptor genes in mammals, *PKD1L3* and *PKD2L1* are crucial in the reception of H<sup>+</sup>. In the Visayan warty pig genome, the annotated *PKD1L3* gene is pseudogenized which is similar to the Duroc pig where the homologous *PKD1L3* gene is also pseudogenized and identified as a lncRNA (*LOC102158108*, as annotated by NCBI) overlapping *DHODH* (fig. 4b).

Pigs have a limited ability to taste salt (sodium chloride [NaCl]) (Danilova et al. 1999; Hellekant and Danilova 1999). Na<sup>+</sup> taste reception involves the selective epithelial amiloride-sensitive sodium channel, ENaC. In most mammals, there are four ENaC channel subunits,  $\alpha$ ,  $\beta$ ,  $\gamma$ , and  $\delta$ . By comparing the ENaC genes among mammals, we found a specific insertion (S420\_Y421insNYG) in the ENaC channel subunit  $\beta$  (*SCNN1B*) in both Visayan warty pig and in Duroc pig (fig. 4c). In a second candidate NaCl receptor, vanilloid receptor subtype 1 (*TRPV1*), we also found specific deletions in Visayan warty pig and in Duroc pig, respectively (T662del in Visayan warty pig and S661\_R665del in Duroc pig, using human *TRPV1* as reference, supplementary fig. 19, Supplementary material online). These two amino acid variants are located in hydrophobic regions on the protein surface and are likely to affect function (supplementary fig. 20, Supplementary material online).

## Discussion

Recent developments in DNA sequencing have put chromosome-level, high-quality genome assemblies of nonmodel, endangered species, within reach. Such high-quality genomes of ecological and evolutionary distinct species provide opportunities to study speciation at the genome level. In this study, we have produced a hybrid assembly approach for the Visayan warty pig genome using 10 $\times$  Chromium linked-reads to generate long contigs. These contigs were scaffolded using Hi-C chromatin contact map and jump reads. This approach enabled us to produce a highly contiguous genome, with all chromosomes reconstructed as single scaffolds.

### Evolution of *Suidae* Genomes

On the basis of a comparison of the genome assembly of Visayan warty pig with that of the Duroc pig (*Sscrofa11.1*), it was confirmed that the Visayan warty pig genome is of comparable continuity and essentially chromosome level. Short read sequencing technology limits resolving low-complexity repeats (Minoche et al. 2011). This explains the relatively low contig N50 (159.6 kb) and incomplete telomeres and centromeres observed in the Visayan warty pig genome assembly. However, the contig N50 is less likely to affect Hi-C-based genome scaffolding (Zhang et al. 2019; Peart et al. 2021). Thus, the clustering, ordering, and orientation of the final assembly are reliable. We also compared basic genome structural features, including gene lengths, coding regions, and noncoding regions of the Visayan warty pig genome with the Duroc pig, all of which show a high level of similarity.

Using fossil calibrations, we estimated the divergence between Visayan warty pig and Duroc pig between 4.57 and 2.71 Ma (Pliocene), which is reflected in the high genomic collinearity between them. The chromosome number for the Visayan warty pig has been reported as  $2n = 35$  in contrast to the domestic pig (*S. scrofa*) where in most instances  $2n = 38$  (O'Brien et al. 2007). A systematic survey of the karyotype of members of the *Suidae* family indicated that the centric fusion in the Visayan warty pig genome, homologous to SSC 13/16, is more likely to be the ancestral karyotype of *Suidae* (supplementary fig. 21, Supplementary Material online). In the ancestral lineages, *Babyrussa* and Sub-Saharan Suids (Melander and Hansen-Melander 1980; Bosma et al. 1996; Thomsen et al. 1996; Musilova et al. 2010), SSC13/SSC16 are fused. After the diversification of *Suidae* species in Eurasia, Pygmy hog (*Porcula salvania*) and other *Sus* species are all carrying separated SSC13/SSC16 homologous chromosome (Bosma et al. 1983; Bosma et al. 1991; Rothschild and Ruvinsky 2011). We hypothesize that the ancestral chromosomes SSC13/SSC16 split in the common ancestor of *Porcula* and *Sus*, whereas the centric fusion independently occurred in Visayan warty pig. The other centric fusion, homologous to SSC14/18, is only found in Visayan warty pig. Notably, chromosome fusion/fission seems to happen quite frequently among the *Suidae* lineage. For example, the karyotype of Sub-Saharan suids ( $2n = 34$ ) differs from that of the domestic pig by the presence of two fusion chromosomes homologous to SSC13/16 and SSC15/17 (Musilova et al. 2010), and some wild boars living in Western Europe carry a single Robertsonian translocation (centric fusion) between SSC15 and SSC17 (Raudsepp and Chowdhary 2011). It is known that recent Robertsonian translocations do not impede producing fertile offspring, for example, in the case of the European wild boar (McFee et al. 1966). Similarly, fertile hybrid offspring between *S. scrofa* and Visayan warty pig, as well as other ISEA *Sus* (Blouch and Groves 1990; IUCN 1993; Ruvinsky et al. 2011), has been reported. Genome analyses also reveal that *S. scrofa* exhibited extensive gene flow with other *Sus* species, and even with its sister genus *Porcula* (Pygmy hog) during range expansion (Liu et al. 1992; Frantz et al. 2013; Frantz et al. 2014). This indicates limited postzygotic reproductive isolation despite frequent Robertsonian translocation. In meiotic prophase, homologous chromosomes must be organized into compacted loop arrays in order to identify one another, pair along their length and physically link to ensure their accurate recombination and segregation in the meiosis I division (Keeney et al. 2014; Patel et al. 2019). The comparison of chromosome interaction maps suggests that, in the short divergence time between Visayan warty pig and Duroc pig ( $\sim 3$  Ma), the chromosome structure remains the same after chromosome fusion. Moreover, we did not observe excess chromosomal inversions or recurrent deletions or insertions of TEs, which may form genomic islands and contribute to reproductive isolation (Chan et al. 2010; Schwander et al. 2014; Lamichhaney et al. 2015; Saenko et al. 2015;

Kirubakaran et al. 2016). This may explain the absence of a postzygotic species barrier. Even with different chromosome numbers, the synapsis during meiosis is probably not interrupted, leading to the absence of postzygotic reproductive isolation.

### Selective Constraints on Functional Elements and Genes

With the newly generated Visayan warty pig genome assembly, an evolutionary genome analysis was conducted aimed at revealing correlations with specific traits in the *Sus* genus.

Functional enrichment analyses of the PSGs and REGs in the ancestral *Sus* genus branch both exhibit enrichment in immune response and energy metabolism. Although fossil evidence cannot unambiguously resolve the origin and dispersion route of *Sus* species, colonization of most of the Eurasian continent and ISEA during the Pliocene is well supported (Azzaroli 1992; van der Made et al. 2006; Frantz et al. 2013; Pickford 2013). The genomic changes associated with olfaction, the immune system, and metabolism may relate to the wide distribution and local adaptation of the ancestral *Sus* population.

Comparing the genome of Visayan warty pig with domestic pigs further showed species-specific traits and the impact of domestication. In the genome of the Visayan warty pig, we observed a series of PSGs and REGs involved in the regulation of growth and thermoregulation. Like many island species, the Visayan warty pig is relatively small in size with an adult weight of 20–40 kg (Rabor 1977; Oliver et al. 1993; Clauss et al. 2008). Reducing body size might be more favorable to maximize energy allocation to growth and reproduction. In contrast, the adult weight of a Duroc pig shows an up to 10-fold difference (250–300 kg) to that of the Visayan warty pig, whereas Eurasian wild boar shows a 5-fold difference (75–100 kg) (Evans et al. 1946; Commission 2010). We also identified REGs in Visayan warty pig related to thermoregulation. The geographical distribution of Visayan warty pig has always been restricted to the Visayan islands in the Philippine archipelago, which have a tropical rainforest climate. The genomic changes associated with body temperature regulation may trigger adaptation to high annual mean temperature and absence of significant seasonality (Villafuerte et al. 2020). Moreover, based on Bergmann's rule, a smaller body size can also have a collective effect on adapting to a high-temperature condition (Ashton et al. 2015). These results reflect the evolutionary trajectory of the Visayan warty pig toward adaptation of tropical island environment.

Domestication and subsequent long-term selective breeding have profoundly changed the anatomical, physiological, and behavioral characteristics of livestock and left clear genomic signatures consistent with selection. We found functional enrichment of the PSGs and REGs in Duroc pig exhibiting enrichment in neural functions, immune response, energy metabolism functions, growth,

and muscle development, consistent with previous reports of pathways related to pig domestication (Li et al. 2013; Bosse et al. 2014; Leno-Colorado et al. 2017; Zhu et al. 2017). We also compared the PSGs identified in our study with previous reports (Groenen et al. 2012; Li et al. 2013) (supplementary fig. 22, Supplementary Material online). Unexpectedly, there is very little intersection between different studies. The branch-site model used to detect positive selection is based on a likelihood ratio test for a predetermined branch of a phylogenetic tree (Yang 2007). We hypothesized that the different selection of taxa will result in different estimation of evolutionary rate in the background branches. Thus, each of them represents a specific evolutionary history, associating with ecological diversification and genomic adaptation. Nonetheless, we would like to note that without the genome information of the wild boar, we cannot pinpoint when these selection constraints occurred. Future studies, where the genome assembly of wild boar and, ideally, the genomes of early domesticated pigs are available, will probably further reveal the genetic basis underlying the evolution of the pig (*S. scrofa*).

### Evolution of Olfaction and Taste

As two of the basic senses of mammals, olfaction, and taste play a very important role in daily life. The two types of chemical sensors involved, are important for recognizing environmental conditions, and have a positive coactivation (integration) in flavor perception (Murphy and Cain 1980; Veldhuizen et al. 2009).

ORs are largely responsible for odor perception and detection of chemical cues, facilitating the differentiation of tens of thousands of unique odorants. This makes olfaction an important physiological function in recognizing food, mates, offspring, territories, and predators. We identified 921 and 1089 functional ORs in Visayan warty pig and Duroc pig, respectively, more than found in most mammals, but lower than, for example, in mouse (1135 ORs). A large number of OR genes found in the mouse probably reflects its nocturnal lifestyle, urban living conditions, and the ability to identify various odors from complex environments (Restrepo et al. 2004). Olfactory gene families expanded in the ancestral *Sus* branch, and have further, independently, expanded in the two *Sus* lineages. Functional categories of expanded ORs in Visayan warty pig are different from Duroc pig, which suggests that environmental adaptation of the two species is driven by the olfactory system. For example, we found expanded ORs involved in the identification of fatty odor in Duroc pig. Wild *Suiformes* are foraging animals with a diet consisting roughly 90% of plant-derived foods such as fruits, roots, leaves, and grasses. The relative amount of dietary fat is usually low (<10%; Leus 1994; Ballari and Barrios-García 2014; Souron et al. 2015), whereas in the swine industry additional fat is added to feed for a better performance (Lewis and Southern 2000). Fat-related OR duplications could be the consequence of adaptation to a high-fat



diet. Another explanation is that forest nuts, which are rich in fats, form an important component of wildboar diet (Schley and Roper 2003; Baubet et al. 2004). Increasing fat-related OR genes may boost the ability of tracing nuts in the ancestral wildboar. Lastly, diversification of genes within each OR subfamily in Visayan warty pig is lower than in Duroc pig. Visayan warty pig is specialized to the tropical island habitat in Southeast Asia. In contrast, the source population of the domesticated pigs, the wild boar (*S. scrofa*), has a vast distribution range across most of the Eurasian continent (also including ISEA). This may have resulted in wild boar having a broader range of olfactory genes to be able to cope with more diverse environmental factors.

Taste receptors play a fundamental role in survival through the identification of dietary nutrients or potentially toxic compounds. Thus, dietary adaptation through selection on taste sensing is likely a major evolutionary force. *TAS2R42* and *TAS2R39* in Visayan warty pig are under positive selection. Similarly, several *TAS2R* genes are under positive selection in mouse, which is also under free-living conditions. *TAS2Rs* are part of the sensory mechanism to identify potentially toxic compounds by bitter taste (Bachmanov and Beauchamp 2007). This positive selection can be explained by adaptation to avoid undesirable substances, for example, plant and insect toxins. Accordingly, relaxed selection constraints can lead to trait reduction or loss (Lahti et al. 2009; Bely 2010; Wicke et al. 2016). Relaxed selection of fat sensory genes in Duroc pig indicated a weakened fatty taste sensibility, which may be due to captive breeding.

Pigs have a limited ability to taste NaCl (Danilova et al. 1999; Hellekant and Danilova 1999). By comparing the pig genomes to other mammalian species, two INDELs in *SCNN1B* and *TRPV1* were identified in Visayan warty pig and Duroc pig (fig 4c, supplementary fig. 19, Supplementary Material online). Since those INDELs are located in an evolutionary conserved region, and, by modeling the proteins, their existence will change the conformation of the hydrophobic surface (supplementary fig. 20, Supplementary Material online). The observed changes, therefore, could potentially influence transport of ions, resulting in reduced nerve response to salty gustatory stimulation in *Sus* species. Moreover, we observed pseudogenization of an acid taste receptor *PKD1L3*, overlapping with *DHODH*, which is private to Visayan warty pig and Duroc pig. *PKD1L3* is present in most mammalian species and does not overlap with *DHODH*. Nerve pulse recordings during taste stimulation indicate that acids give a distinct taste to the pig (Hellekant and Danilova 1999; Glaser et al. 2001). Further research is needed to validate the association between the absence of *PKD1L3* and the sour preference of pigs.

## Conclusions

Our comprehensive evolutionary and comparative genome analyses provided insight into the genomics of recent

species divergence. Comparing the chromosome-level assembly of Visayan warty pig with domestic Duroc pig revealed genome evolution and reproductive division among *Suidae*. We identified numerous genetic differences correlated with natural selection for local adaptation and artificial selection during domestication, especially on olfaction and taste, shedding new light on the origin of species-specific characteristics. This study provides valuable genomic resources as well as insights into not only the evolution and diversification of Suiform but also our understanding of mammalian genome biology.

## Materials and Methods

### Sample Collection and Sequencing

DNA of a female Visayan warty pig was obtained from tissue samples at the Rotterdam Zoo. Tissues for RNA-seq were dissected from seven organs (liver, spleen, lymph, muscle, lung, frontal lobe, olfactory bulb).

### 10× Genomics Library

Visayan warty pig's genomic DNA was size selected for fragments >40 kb on a BluePippin instrument (Sage Sciences, Beverly, MA, USA) and Illumina sequencing libraries were constructed using the 10× Genomics Chromium Controller instrument with the Genome Reagents Kit v2 chemistry (10× Genomics, Pleasanton, CA, USA) according to the manufacturer's recommendations. The resulting Illumina library was sequenced on a NextSeq500 using a High Output Kit v2 for paired-end, 2 × 151 bp run (Illumina, San Diego, CA, USA).

### Hi-C Library

The Hi-C library was generated by Dovetail Genomics as described by Lieberman-Aiden et al. (2009). Briefly, for each library, chromatin was fixed in place with formaldehyde in the nucleus and then extracted. Fixed chromatin was digested with DpnII, the 5' overhangs filled in with biotinylated nucleotides, and then free blunt ends were ligated. After ligation, crosslinks were reversed, and the DNA purified from protein. Purified DNA was treated to remove biotin that was not linked to ligated fragments. The DNA was then sheared to a mean fragment size of ~350 bp, and sequencing libraries were generated using NEBNext Ultra enzymes and Illumina-compatible adapters. Biotin-containing fragments were isolated using streptavidin beads before PCR enrichment of each library. The libraries were sequenced on an Illumina Hi-seq 4000 sequencing platform (2 × 151 bp).

### Long Jumping Libraries

Briefly, DNA was extracted from the same Visayan warty pig tissues, and used to create two mate-pair libraries, with insert sizes of 10 and 20 kb, respectively. The mate-pair libraries were prepared with Illumina's "Nextera Mate-Pairs Sample Prep kit" followed by the "TruSeq

DNA Sample Prep kit” and sequenced on Illumina HiSeq 2500 platform.

#### RNA-seq

High-quality RNA (1 µg) was used to generate TruSeq Stranded RNA-seq libraries (TruSeq Stranded RNA Sample Preparation Kit, Illumina, San Diego, CA, USA) by the HTS Lab (University of Illinois, Urbana, IL, USA) following standard protocols and sequenced on an Illumina HiSeq 2000 platform.

#### Genome Assembly

Based on linked reads that were sequenced with a standard Illumina system, we used the Supernova assembler (version 1.2.2, 10× Genomics Inc., Pleasanton, CA, USA) with the parameters (run `-maxreads = 1,500,000,000; mkoutput -style = pseudohap2 -minsize = 500`) to construct contigs and scaffolds of the Visayan warty pig.

#### Chromosome Assembly Using Hi-C

For the intermediary de novo assembly, Chicago library reads, and Dovetail Hi-C library reads were used as input data for HiRise ([https://github.com/DovetailGenomics/HiRise\\_July2015\\_GR](https://github.com/DovetailGenomics/HiRise_July2015_GR)), a software pipeline designed specifically for using proximity ligation data to scaffold genome assemblies (Putnam et al. 2016). An iterative analysis was conducted. First, Shotgun and Chicago library sequences were aligned to the draft input assembly using a modified SNAP read mapper. The separations of Chicago read pairs mapped within draft scaffolds were analyzed by HiRise to produce a likelihood model for genomic distance between read pairs, and the model was used to identify and break putative misjoins, to score prospective joins, and make joins above a default threshold. After aligning and scaffolding Chicago data, Dovetail Hi-C library sequences were aligned and scaffolded following the same method. After scaffolding, shotgun sequences were used to close gaps between contigs.

#### Reference Guided and Long Jumping Reads Scaffolding

Since Visayan warty pig and Duroc pig (*S. scrofa*) (Frantz et al. 2013) are closely related, the HiRise assembly was subjected to a secondary assembly with AlignGraph (Bao et al. 2014) using the *Sus scrofa*11.1 reference genome (Warr et al. 2019) as guidance. Briefly, in the guided secondary assembly with AlignGraph the de novo generated scaffolds are aligned to the reference and the long jumping reads (10 and 20 kb insert size) are mapped to the assembled scaffolds and to the reference. This results in a paired-end multi-positional de Bruijn graph from which the scaffolds are extended if possible. The orientation of the chromosomes in *S. cebifrons* assembly is manually adjusted to match the orientation of the orthologous chromosomes in the pig reference genome.

#### Mitochondrial Genome Assembly

We randomly selected 4 Gb paired-end reads from resequencing libraries to assemble the mitochondrial genome sequence using MitoZ v2.3, resulting in a 16.52 kb circular assembly (supplementary fig. 3, Supplementary Material online).

#### Evaluation of Genome Assembly

The Illumina short reads from the 10× libraries were used for *k*-mer frequency analyses of genomes. A *k*-mer count analysis was done using Jellyfish v2.3.0 (Marçais and Kingsford 2011) on the Illumina data. From the paired-end reads, only the first read was truncated to 100 bp to avoid the lower quality part of the read. The second read was omitted from this analysis to avoid counting overlapping *k*-mers. A *k*-mer size of  $k = 21$  was used. After converting the *k*-mer counts into a histogram format, this file was analyzed using Genomescope (Vurture et al. 2017). We used Perl script (<https://github.com/calacademy-research/ccgutils/tree/master/asmstats>) to summarize the statistics of genome assemblies with default options. Any more than ten consecutive *N* should split the scaffolds into contigs.

#### Evaluation of Genome Completeness with BUSCO

BUSCO v3.1.0 (Simão et al. 2015) was used to assess the genome completeness by estimating the percentage of expected single-copy conserved orthologs captured in the Visayan warty pig assembly, applying the *mammalia\_odb9* BUSCO set.

#### Cross-Species Hi-C Processing

Hi-C data for domesticated pig (large white) were downloaded from FR-AgENCODE (<http://www.fragsencode.org/results.html>)86. The Hi-C reads from domesticated pig were mapped to the pig genome *Sscrofa*11.1 and the Visayan warty pig Hi-C was mapped to the Visayan warty pig assembly separately, using *bwa mem* 0.7.5a (Li and Durbin 2009) with parameters “-E50 -L0.” We mapped  $\sim 7.31E+07$  and  $2.26E+07$  Hi-C contact pairs for the Visayan warty pig and Duroc pig genomes, respectively. Next, we aligned the Hi-C contact pairs of Visayan warty pig to the pig genome. We mapped the aligned loci of the two ends of a contact pair in the Hi-C data of the Visayan warty pig from the original genome assembly to the pig genome with reciprocal mapping using the tool *liftOver* (Hinrichs 2006) based on the whole-genome alignment. We used HiCExplorer v3.0 (Ramírez et al. 2018) to create the contact matrices and extract the Hi-C contact maps of each species at the resolution of 50kb. Normalization was performed using Knight–Ruiz matrix balancing (Knight and Ruiz 2013). We applied Phylo-HMRF (Yang et al. 2019) to the multispecies Hi-C data to predict 20 hidden states. For all the autosomes and chromosome X in the Duroc pig genome, we run Phylo-HMRF jointly on the multiple synteny blocks of the chromosomes and identified possible different evolutionary patterns of the Hi-C contact frequencies across

species in a genome-wide manner. We further categorized the 20 estimated hidden states based on their conservation status.

### Identification of Repetitive Elements

We identified repetitive elements by a combination of homology alignment and de novo searches, as follows. We used RepeatMasker (Tarailo-Graovac and Chen 2009) with Repbase (v.16.10) (Jurka et al. 2005) to scan for sequences homologous to annotated repeat sequences in published databases and then used RepeatModeler (Jurka et al. 2005) with the default parameters to predict de novo TEs. We combined the repeat sequences identified by both methods as the final annotated repeat set. We integrated the overlapping TEs and removed those with scores  $< 200$ . Subsequently, tandem repeats were annotated using Tandem Repeat Finder (Benson 1999) (version 4.07b; with the following settings: 2 7 7 80 10 50 2000 -d -h).

### Gene Prediction

Gene prediction was conducted with the BRAKER v2.1.2 (Hoff et al. 2019) pipeline by integrating ab initio, transcriptome-based, and protein homology-based evidence. Ab initio gene predictions were performed with Augustus v3.3.2 (Stanke et al. 2006) and GeneMark-ET v4.33 (Lomsadze et al. 2014). Before starting the annotation process, we used the mammalia\_odb9 BUSCO set to train Augustus. For prediction based on the RNA-seq data, all RNA reads were aligned to the Visayan warty pig genome by HISAT v2.1.0 (Kim et al. 2019). The results were used to identify candidate exon regions, donor, and acceptor sites. Protein sequences of all mammalian species were downloaded from UniProt (UniProt Consortium et al. 2015) as protein homology-based evidence.

Single exon genes (SEG) were identified as mono-exonic genes containing full-length open reading frames, as specified by the gene prediction pipeline. The same approach was applied to the human, mouse, dog, horse, cattle, and pig genomes. Lists of nonredundant SEG from all six genomes were obtained and aligned to the Visayan warty pig genome using TBLASTN (Camacho et al. 2009) ( $E$ -value cutoff:  $1E-5$ ). We aligned homologous genomic sequences against matching proteins to define gene models using GeneWise v2.4.1 (Birney and Durbin 2000). The function of the predicted OR genes was determined as described in Truong Nguyen et al. (2012). Gene annotation of mitochondria was performed using MitoZ software. EvidenceModeler v1.1.1 (Haas et al. 2008) was used to integrate the genes predicted by BRAKER and SEG approaches and generate a consensus gene set.

### Functional Annotation

We annotated the functions of the protein-coding genes using BLASTP ( $E$ -value cutoff:  $1E-5$ ) (Camacho et al. 2009) based on entries in the Swiss-Prot and TrEMBL databases (UniProt Consortium et al. 2015). The motifs and domains were annotated using InterProScan v5.35 (Jones

et al. 2014) by searching against publicly available databases, including ProDom, PRINTS, Pfam, SMART, PANTHER, and PROSITE. The GO IDs for each gene were assigned according to the corresponding InterPro entry.

### Noncoding Gene Element Prediction

We also predicted gene structures of tRNAs, rRNAs, and other noncoding RNAs. A total of 37,019 tRNAs were predicted using t-RNAscan-SE v2.0 (Lowe and Eddy 1996). Because rRNA genes are highly evolutionarily conserved, we choose human rRNA sequence as references and then predicted rRNA genes using Blast tool with default parameters. Small nuclear and nucleolar RNAs were annotated using the Infernal v1.1.2 (Nawrocki and Eddy 2013). lncRNAs were detected using FEELnc v0.1.1 (Wucher et al. 2017) with default parameters and the output transcripts of the HISAT alignments.

### Whole-Genome Alignments and Structural Variation

The Nucmer program from Mummer v3.23 (Haas et al. 2008) was used with `-maxgap 50` and `-breaklen 400` for comparing the genomes between Visayan warty pig to Duroc pig to obtain the synteny blocks. Structural variations between Visayan warty pig and Duroc pig were called using svmu as described by Chakraborty et al. (2019).

For multiple genome alignments from Visayan warty pig, Duroc pig, cattle, horse, dog, mouse, and human, we first aligned all genomes to the human genome using LAST v980 (Kielbasa et al. 2011). Then, we used MULTIZ v11.2 (Blanchette et al. 2004) to merge the pairwise alignments into multiple genome alignments using the human genome as the reference. Approximately 241 Mb syntenic sequence is shared by the seven mammalian species.

### Gene Family, Phylogenetic Analyses, and Divergence Time Calibration

We used the Orthofinder v2.3.3 to identify gene families/clusters (Emms and Kelly 2019). Genome sequences and annotations for Duroc pig, cattle, horse, dog, mouse, and human were downloaded from Ensembl 100 (<https://www.ensembl.org>). The longest proteins of each gene were aligned to one and another. The species-specific gene families were determined according to the presence or absence of genes for a given species. The shared family expansion and contraction analysis were conducted with CAFÉ v3.1 (De Bie et al. 2006).

Phylogenetic relationships were resolved based on the maximum-likelihood (ML) method as implemented in RAxML v8.2.3 (Stamatakis 2014) using the best-fitting model of substitutions, identified by jModelTest2 (Darriba et al. 2012). 10,057 high-quality 1:1 single-copy orthologous genes were used.

We estimated divergence times using an approximate likelihood method as implemented in MCMCtree (Yang 2007), with an independent relaxed clock and birth–death sampling. We used a float prior and a maximum bound age, with a scale parameter of  $c = 2$ . For the root

divergence, we set the before ( $tU = 2$  [100 Ma],  $P = 0.1$ ,  $c = 2$ ). For MRCA of *Sus*, we used the same fossil calibration as in [Frantz et al. \(2013\)](#) ( $tL = 0.2$  [2 Ma],  $P = 0.1$ ,  $c = 0.5$ ).

DNA evolutionary rates in each species, the common ancestor of *Sus* and *Artiodactyla* was calculated using r8s ([Sanderson 2003](#)) based on the whole-genome alignments.

### Identification of HCEs

PHAST v1.5 ([Hubisz et al. 2011](#)) was used to identify HCEs based on the multiple-genome sequence alignments. Firstly, we extracted the 4-fold degenerate sites from the multiple-genome alignments generated above and used phyloFit (in PHAST v1.5 package) to estimate a neutral phylogenetic model (also considered as the nonconserved model in PhastCons). Then we ran PhastCons (in PHAST v1.5 package) to estimate conserved and nonconserved models and predicted conserved elements and conservation scores based on the conserved and nonconserved models. We identified 1,162,748 HCEs, covering 137.2 Mb of the human genome. Further, we identified SHCEs, which are conserved in *Sus* species, but not in other mammalian species. We used phyloP (in PHAST v1.5 package, with the parameters: phyloP –method LRT –mode CONACC) under the nonconserved model. Finally, we identified 49,846 SHCEs, covering 13.9 Mb.

Noncoding conserved regions usually contain regulatory elements, that is, TFBSes. We downloaded the human TFBS data from the UCSC data set generated by the ENCODE project (<ftp://hgdownload.cse.ucsc.edu/goldenPath/hg38/encRegTfbsClustered/encRegTfbsClusteredWithCells.hg38.bed.gz>) to compare the conservation degrees of SHCEs. We found that of 4,379,357 human TFBSes used in the analysis, 71,908 (1.6%) were conserved in *Sus*.

### Identification of Selective Constraint in Genes

#### *Positively Selected Genes*

The strength of positive selection on each codon of each orthologous gene along a specific targeted lineage of a phylogenetic tree was estimated with the branch-site model using the Codeml program of the PAML package v4.9. To determine to what degree these codon sequences, along the targeted lineage fit the branch-site model (including positive selection) better than the one containing neutral selection or negative selection. An alternative branch-site model (Model = 2, NSsites = 2, and Fix = 0) and a null branch-site model (Model = 2, NSsites = 2, Fix = 1, and Fix  $\omega = 1$ ) were combined to calculate log-likelihood values for each model using likelihood ratio tests. The log-likelihood values generated were used to assess the model fit, using the  $\chi^2$  test with one degree of freedom. Genes with a  $P$ -value  $< 0.05$  were treated as candidates under positive selection.

#### *Rapidly Evolved Genes*

Branch model in Codeml was used to identify evolutionary rate among orthologous genes. To determine whether these codon sequences, along the targeted lineage, are

evolving under a different evolutionary rate than other lineages, an alternative branch-site model (Model = 2) and a null branch-site model (Model = 0) were combined to calculate log-likelihood values for each model using likelihood ratio tests. Genes with a  $P$ -value  $< 0.05$  and a higher  $\omega$  value for the foreground than the background branches were considered as evolving at a faster rate.

#### *Relaxed Selected Genes*

RELAX ([Wertheim et al. 2015](#)) analyses were also performed to evaluate whether selective constraints are stronger in foreground branches compared with background branches. An input file that contains the codon sequence alignment and the RAxML tree was provided, and the foreground branches (test branches) and background branches (reference branches) were then indicated before running the analysis. A likelihood ratio test was also performed to evaluate if selection varied between test and reference branches. A significant result ( $P$ -value  $< 0.05$ ) indicates a stronger level of selection if  $K$  is  $> 1$  or a weaker selection or relaxed constraint if  $K$  is  $< 1$ .

### Functional, Pathway, and Interaction Enrichment Analysis

KOBAS (version 3.0) ([Wu et al. 2006](#)) was applied to perform GO analyses (including cellular composition, molecular function, and biological process terms), Reactome pathway analyses, and Kyoto Encyclopedia of Genes and Genomes pathway enrichment analyses. False discovery rate was performed to adjust  $P$ -values using the Benjamini and Hochberg method. An adjusted  $P$ -value of  $< 0.05$  was used as the cutoff criterion.

### Interspecies Conservation of Taste Receptor Genes

We used genomic evolutionary rate profiling (GERP++) ([Davydov et al. 2010](#)) to estimate the interspecies conservation of taste receptor genes. GERP++ provides conservation scores through the quantification of position-specific constraint in multiple-species alignments. We calculated and attributed the GERP scores for each nucleotide site.

### Protein Structure Modeling

The three-dimensional homology models of proteins were obtained using Swiss-Model (<https://swissmodel.expasy.org/>) ([Waterhouse et al. 2018](#)) to search for templates and build the models. The predicted protein structures were estimated based on target–template alignment using ProMod3 on SWISS-MODEL server (<https://swissmodel.expasy.org/>). Models were built based on the target–template alignment using ProMod3. The global and per-residue model quality was assessed using the QMEAN scoring function.

## Supplementary Material

Supplementary data are available at *Molecular Biology and Evolution* online.

## Acknowledgements

The authors thank Willem Schaftenaar (diergaarde blijdorp) for technical assistance. L.L. was financially supported by the China Scholarship Council (Grant No. 201707720055). M.B. was financially supported by NWO grant 016.Veni.181.050. This project was financially supported by a European Research Council grant (ERC-2009-AdG: 249894). Genetic material of Visayan warty pig (medical euthanasia) was obtained from tissue samples collected by veterinarians at the Rotterdam Zoo, The Netherlands, together with researcher from Wageningen University, according to national legislation. Tissue samples obtained are derived from animals culled within the wildlife management program of the Rotterdam Zoo.

## Author Contributions

M.A.M.G., O.M., and L.L. designed the scope of the study. L.L. analyzed the data; L.L. and Q.H. performed the Hi-C analyses; L.L., H.-J.M., M.B., M.A.M.G., and O.M. discussed the results; L.L. wrote the manuscript; H.-J.M., M.B., M.A.M.G., and O.M. provided valuable suggestion and comments to improve the manuscript. The author(s) read and approved the final manuscript.

## Data Availability

The sequencing data have been deposited in the ENA database as Bioproject PRJEB38928, as well as in Open Science Framework project page (<https://osf.io/6yznw>). Scripts and code used for data generation and analyses are archived on Github ([https://github.com/llq0325/Sus\\_cebifrons\\_genome](https://github.com/llq0325/Sus_cebifrons_genome)).

## References

- Ashton KG, Tracy MC, De Queiroz A. 2015. Is Bergmann's rule valid for mammals? *Am Nat.* **156**:390–415.
- Azzaroli A. 1992. Suids of the early villafranchian of villafranca d'asti and china. *Rend Lincei.* **3**:109–124.
- Bachmanov AA, Beauchamp GK. 2007. Taste receptor genes. *Annu Rev Nutr.* **27**:389–414.
- Ballari SA, Barrios-García MN. 2014. A review of wild boar *Sus scrofa* diet and factors affecting food selection in native and introduced ranges. *Mammal Rev.* **44**:124–134.
- Bao E, Jiang T, Girke T. 2014. AlignGraph: algorithm for secondary de novo genome assembly guided by closely related references. *Bioinformatics* **30**:i319–i328.
- Baubet E, Bonenfant C, Brandt S. 2004. Diet of the wild boar in the French Alps. *Galemys Boletín Inf la Soc Española para la Conserv y Estud los mamíferos* **16**:101–113.
- Bely AE. 2010. Evolutionary loss of animal regeneration: pattern and process. *Integr Comp Biol.* **50**:515–527. Oxford Academic.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**:573–580.
- Birney E, Durbin R. 2000. Using GeneWise in the Drosophila annotation experiment. *Genome Res.* **10**:547–548.
- Blanchette M, Kent WJ, Riemer C, Elnitski L, Smit AFA, Roskin KM, Baertsch R, Rosenbloom K, Clawson H, Green ED, et al. 2004. Aligning multiple genomic sequences with the threaded blockset aligner. *Genome Res.* **14**:708–715.
- Blouch RA, Groves CP. 1990. Naturally occurring suid hybrid in Java. *Z Säugetierkd.* **55**:270–275.
- Bosma AA, De Haan NA, Blouch RA, Macdonald AA. 1991. Comparative cytogenetic studies in *Sus verrucosus*, *Sus celebensis* and *Sus scrofa vittatus* (Suidae, Mammalia). *Genetica* **83**:189–194.
- Bosma AA, Mellink CHM, Yerle M, Zijlstra C. 1996. Chromosome homology between the domestic pig and the babirusa (family Suidae) elucidated with the use of porcine painting probes. *Cytogenet Genome Res.* **75**:32–35.
- Bosma AA, Oliver WLR, Macdonald AA. 1983. The karyotype, including G- and C-banding patterns, of the pigmy hog *Sus* (*Porcula*) *salvanus* (Suidae, Mammalia). *Genetica* **61**:99–106.
- Bosse M, Megens H-J, Frantz LAF, Madsen O, Larson G, Paudel Y, Duijvesteijn N, Harlizius B, Hagemeyer Y, Crooijmans RPMA, et al. 2014. Genomic analysis reveals selection for Asian genes in European pigs following human-mediated introgression. *Nat Commun.* **5**:4392.
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinform.* **10**, 421.
- Chakraborty M, Emerson JJ, Macdonald SJ, Long AD. 2019. Structural variants exhibit widespread allelic heterogeneity and shape variation in complex traits. *Nat Commun.* **10**:1–11.
- Chan YF, Marks ME, Jones FC, Villarreal G, Shapiro MD, Brady SD, Southwick AM, Absher DM, Grimwood J, Schmutz J, et al. 2010. Adaptive evolution of pelvic reduction in sticklebacks by recurrent deletion of a pitxl enhancer. *Science* **327**:302–305.
- Clauss M, Nijboer J, Loermans JHM, Roth T, Van der Kuilen J, Beynen AC. 2008. Comparative digestion studies in wild suids at Rotterdam Zoo. *Zoo Biol.* **27**:305–319.
- Commission E. 2010. *Farmer's hand book on pig production food and agriculture organization of the United Nations.*
- Coyne JA, Orr HA. 2009. Speciation: a catalogue and critique of species concepts. *Philos Biol Anthol.* 272–292.
- Danilova V, Roberts T, Hellekant G. 1999. Responses of single taste fibers and whole chorda tympani and glossopharyngeal nerve in the domestic pig, *Sus scrofa*. *Chem Senses* **24**:301–316.
- Darriba D, Taboada GL, Doallo R, Posada D. 2012. JModelTest 2: more models, new heuristics and parallel computing. *Nat Methods* **9**:772.
- Davydov EV, Goode DL, Sirota M, Cooper GM, Sidow A, Batzoglou S, Wasserman WW. 2010. Identifying a high fraction of the human genome to be under selective constraint using GERP++. *PLoS Comput Biol.* **6**:e1001025.
- De Bie T, Cristianini N, Demuth JP, Hahn MW. 2006. CAFE: a computational tool for the study of gene family evolution. *Bioinformatics* **22**:1269–1271.
- Emms DM, Kelly S. 2019. OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol.* **20**:238.
- Evans BR, Evans GG, et al. 1946. *The story of Durocs, the truly American breed of swine.*
- Fang X, Mou Y, Huang Z, Li Y, Han L, Zhang Y, Feng Y, Chen Y, Jiang X, Zhao W, et al. 2012. The sequence and analysis of a Chinese pig genome. *Gigascience* **1**:16.
- Foissac S, Djebali S, Munyard K, Vialaneix N, Rau A, Muret K, Esquerré D, Zytnicki M, Derrien T, Bardou P, et al. 2019. Multi-species annotation of transcriptome and chromatin structure in domesticated animals. *BMC Biol.* **17**:108.
- Frantz LAF, Madsen O, Megens HJ, Groenen MAM, Lohse K. 2014. 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.* **23**:5566–5574.

- Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Bosse M, Paudel Y, Semiadi G, Meijaard E, Li N, Crooijmans RPMA, et al. 2013. Genome sequencing reveals fine scale diversification and reticulation history during speciation in *Sus*. *Genome Biol.* **14**:R107.
- Frantz LAF, Schraiber JG, Madsen O, Megens H-J, Cagan A, Bosse M, Paudel Y, Crooijmans RPMA, Larson G, Groenen MAM. 2015. Evidence of long-term gene flow and selection during domestication from analyses of Eurasian wild and domestic pig genomes. *Nat Genet.* **47**:1141–1148.
- Glaser D, Wanner M, Tinti JM, Nofre C. 2001. Pig responses to taste stimuli. In *Chemical signals in vertebrates*. Vol. 9. Boston (MA): Springer. p. 423–431.
- Groenen MAM, Archibald AL, Uenishi H, Tuggle CK, Takeuchi Y, Rothschild MF, Rogel-Gaillard C, Park C, Milan D, Megens H-J et al. 2012. Analyses of pig genomes provide insight into porcine demography and evolution. *Nature* **491**:393–398.
- Haas BJ, et al. 2008. Automated eukaryotic gene structure annotation using EVIDENCEModeler and the Program to Assemble Spliced Alignments. *Genome Biol.* **9**:1–22.
- Hellekant G, Danilova V. 1999. Taste in domestic pig. *Sus scrofa*. *J Anim Physiol Anim Nutr (Berl)* **82**:8–24.
- Hinrichs AS. 2006. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res.* **34**:D590–D598.
- Hoff KJ, Lomsadze A, Borodovsky M, Stanke M. 2019. Whole-genome annotation with BRAKER. *Methods Mol Biol.* **1962**:65–95.
- Horio N, Yoshida R, Yasumatsu K, Yanagawa Y, Ishimaru Y, Matsunami H, Ninomiya Y, Meyerhof W. 2011. Sour taste responses in mice lacking pkd channels. *PLoS One* **6**:e20007.
- Huang AL, Chen X, Hoon MA, Chandrashekar J, Guo W, Tränkner D, Ryba NJP, Zuker CS. 2006. The cells and logic for mammalian sour taste detection. *Nature* **442**:934–938.
- Hubisz MJ, Pollard KS, Siepel A. 2011. Phastand Rphast: phylogenetic analysis with space/time models. *Brief Bioinform.* **12**:41–51.
- Ishimaru Y, Inada H, Kubota M, Zhuang H, Tominaga M, Matsunami H. 2006. Transient receptor potential family members PKD1L3 and PKD2L1 form a candidate sour taste receptor. *Proc Natl Acad Sci U S A.* **103**:12569–12574.
- Jones P, Binns D, Chang H-Y, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, et al. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics* **30**:1236–1240.
- Jurka J, Kapitonov VV, Pavlicek A, Klonowski P, Kohany O, Walichiewicz J. 2005. Repbase update, a database of eukaryotic repetitive elements. *Cytogenet Genome Res.* **110**:462–467.
- Keeney S, Lange J, Mohibullah N. 2014. Self-organization of meiotic recombination initiation: general principles and molecular pathways. *Annu Rev Genet.* **48**:187–214.
- Kielbasa SM, Wan R, Sato K, Horton P, Frith MC. 2011. Adaptive seeds tame genomic sequence comparison. *Genome Res.* **21**:487–493.
- Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. 2019. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol.* **37**:907–915.
- Kirubakaran TG, Grove H, Kent MP, Sandve SR, Baranski M, Nome T, De Rosa MC, Righino B, Johansen T, Otterå H, et al. 2016. Two adjacent inversions maintain genomic differentiation between migratory and stationary ecotypes of Atlantic cod. *Mol Ecol.* **25**:2130–2143.
- Knight PA, Ruiz D. 2013. A fast algorithm for matrix balancing. *IMA J Numer Anal.* **33**:1029–1047.
- Lahti DC, Johnson NA, Ajie BC, Otto SP, Hendry AP, Blumstein DT, Coss RG, Donohue K, Foster SA. 2009. Relaxed selection in the wild. *Trends Ecol Evol.* **24**:487–496.
- Lamichhaney S, Fan G, Widemo F, Gunnarsson U, Thalmann DS, Hoepfner MP, Kerje S, Gustafson U, Shi C, Zhang H, et al. 2015. Structural genomic changes underlie alternative reproductive strategies in the ruff (*Philomachus pugnax*). *Nat Genet.* **48**:84–88.
- Larson G, Dobney K, Albarella U, Fang M, Matisoo-Smith E, Robins J, Lowden S, Finlayson H, Brand T, Willerslev E, et al. 2005. Worldwide phylogeography of wild boar reveals multiple centers of pig domestication. *Science* **307**:1618–1621.
- Larson G, Liu R, Zhao X, Yuan J, Fuller D, Barton L, Dobney K, Fan Q, Gu Z, Liu X-H, et al. 2010. Patterns of East Asian pig domestication, migration, and turnover revealed by modern and ancient DNA. *Proc Natl Acad Sci U S A.* **107**:7686–7691.
- Leno-Colorado J, Hudson NJ, Reverter A, Pérez-Enciso M. 2017. A pathway-centered analysis of pig domestication and breeding in Eurasia. *G3 Genes|Genomes|Genetics* **7**:2171–2184.
- Leus K. 1994. *Foraging behaviour, food selection and diet digestion of *Babyroura babyrussa* (Suidae, Mammalia)*. The University of Edinburgh, College of Medicine and Veterinary Medicine, Royal (Dick) Veterinary School.
- Lewis AJ, Lee Southern L. 2000. *Swine nutrition*. 2nd ed. CRC Press.
- Li M, Chen L, Tian S, Lin Y, Tang Q, Zhou X, Li D, Yeung CKL, Che T, Jin L, et al. 2017. Comprehensive variation discovery and recovery of missing sequence in the pig genome using multiple de novo assemblies. *Genome Res.* **27**:865–874.
- Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**:1754–1760.
- Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung CKL, Chen L, Ma J, et al. 2013. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet.* **45**:1431–1438.
- Li D, Zhang J. 2014. Diet shapes the evolution of the vertebrate bitter taste receptor gene repertoire. *Mol Biol Evol.* **31**:303–309.
- Lieberman-Aiden E, van Berkum NL, Williams L, Imakaev M, Ragoczy T, Telling A, Amit I, Lajoie BR, Sabo PJ, Dorschner MO et al. 2009. Comprehensive mapping of long-range interactions reveals folding principles of the human genome. *Science* **326**:289–293.
- Liu L, Bosse M, Megens H-J, Frantz LAF, Lee Y-L, Irving-Pease EK, Narayan G, Groenen MAM, Madsen O. 2019. Genomic analysis on pygmy hog reveals extensive interbreeding during wild boar expansion. *Nat Commun.* **10**(1):1–9.
- Lomsadze A, Burns PD, Borodovsky M. 2014. Integration of mapped RNA-Seq reads into automatic training of eukaryotic gene finding algorithm. *Nucleic Acids Res.* **42**:e119–e119.
- Lowe TM, Eddy SR. 1996. TRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
- Marçais G, Kingsford C. 2011. A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. *Bioinformatics* **27**:764–770.
- McFee AF, Banner MW, Rary JM. 1966. Variation in chromosome number among European wild pigs. *Cytogenet Genome Res.* **5**:75–81.
- Melander Y, Hansen-Melander E. 1980. Chromosome studies in African wild pigs (Suidae, Mammalia). *Hereditas* **92**:283–289.
- Melletti M, Meijaard E. 2017. *Ecology, conservation and management of wild pigs and peccaries*. p. 1–448.
- Meng G, Li Y, Yang C, Liu S. 2019. MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization. *Nucleic Acids Res.* **47**:e63.
- Minoche AE, Dohm JC, Himmelbauer H. 2011. Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems. *Genome Biol.* **12**(R112):1–15.
- Murphy C, Cain WS. 1980. Taste and olfaction: independence vs interaction. *Physiol Behav.* **24**:601–605.
- Musilova P, Kubickova S, Hornak M, Cernohorska H, Vahala J, Rubes J. 2010. Different fusion configurations of evolutionarily conserved segments in karyotypes of *Potamochoerus porcus* and *Phacochoerus africanus*. *Cytogenet Genome Res.* **129**:305–309.
- Nawrocki EP, Eddy SR. 2013. Infernal 1.1: 100-fold faster RNA homology searches. *Bioinformatics* **29**:2933–2935.
- O'Brien SJ, Menninger JC, Nash WG. 2007. Atlas of mammalian chromosomes. *Choice reviews online*. Vol. 44. John Wiley & Sons.
- Oliver WLR. 1993. *Status survey and conservation action plan. Pigs, Peccaries and Hippos*. IUCN.
- Oliver WLR, Cox R, Groves C. 1993. The Philippine Warty pigs (*Sus philippensis* and *S. cebifrons*). *Pigs Peccaries Hippos*:145–155.
- Patel L, Kang R, Rosenberg SC, Qiu Y, Raviram R, Chee S, Hu R, Ren B, Cole F, Corbett KD. 2019. Dynamic reorganization of the genome

- shapes the recombination landscape in meiotic prophase. *Nat Struct Mol Biol.* **26**:164–174.
- Paudel Y, Madsen O, Megens H-J, Frantz LAF, Bosse M, Crooijmans RPMA, Groenen MAM. 2015. Copy number variation in the speciation of pigs: a possible prominent role for olfactory receptors. *BMC Genomics* **16**:1–4.
- Pearl CR, Williams C, Pophaly SD, Neely BA, Gulland FMD, Adams DJ, Ng BL, Cheng W, Goebel ME, Fedrigo O, et al. 2021. Hi-C scaffolded short- and long-read genome assemblies of the California sea lion are broadly consistent for syntenic inference across 45 million years of evolution. *Mol Ecol Resour.* **21**(7):2455–2470.
- Pickford M. 2013. Suids from the Pleistocene of Naungkwe Taung, Kayin State, Myanmar. *Paleontol Res.* **16**:307–317.
- Putnam NH, O'Connell BL, Stites JC, Rice BJ, Blanchette M, Calef R, Troll CJ, Fields A, Hartley PD, Sugnet CW, et al. 2016. Chromosome-scale shotgun assembly using an in vitro method for long-range linkage. *Genome Res.* **26**:342–350.
- Rabor DS. 1977. *Philippine birds and mammals*. UP Science Education Center.
- Ramírez F, Feo T, Harvey TA, Prum RO. 2018. High-resolution TADs reveal DNA sequences underlying genome organization in flies. *Nat Commun.* **9**:1–15.
- Raudsepp T, Chowdhary BP. 2011. Cytogenetics and chromosome maps. *Genet Pig Second Ed.*:134–178.
- Restrepo D, Arellano J, Oliva AM, Schaefer ML, Lin W. 2004. Emerging views on the distinct but related roles of the main and accessory olfactory systems in responsiveness to chemosensory signals in mice. *Horm Behav.* **46**:247–256.
- Rothschild MF, Ruvinsky A. 2011. *The genetics of the pig: second edition*. CABI.
- Roura E, Baldwin MW, Klasing KC. 2013. The avian taste system: potential implications in poultry nutrition. *Anim Feed Sci Technol.* **180**:1–9.
- Ruvinsky A, Rothschild MF, Larson G, Gongora J. 2011. *Systematics and evolution of the pig. The genetics of the pig: second edition*.
- Saenko SV, Lamichhaney S, Barrio AM, Rafati N, Andersson L, Milinkovitch MC. 2015. Amelanism in the corn snake is associated with the insertion of an LTR-retrotransposon in the OCA2 gene. *Sci Rep.* **5**:17118.
- Sanderson MJ. 2003. r8s: inferring absolute rates of molecular evolution and divergence times in the absence of a molecular clock. *Bioinformatics* **19**:301–302.
- Schley L, Roper TJ. 2003. Diet of wild boar *Sus scrofa* in Western Europe, with particular reference to consumption of agricultural crops. *Mamm Rev.* **33**:43–56.
- Schwander T, Libbrecht R, Keller L. 2014. Supergenes and complex phenotypes. *Curr Biol.* **24**:R288–R294.
- Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics* **31**:3210–3212.
- Souron A, Merceron G, Blondel C, Brunetière N, Colyn M, Hofman-Kamińska E, Boissierie J-R. 2015. Three-dimensional dental microwear texture analysis and diet in extant Suidae (Mammalia: Cetartiodactyla). *Mammalia* **79**:279–291.
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**:1312–1313.
- Stanke M, Keller O, Gunduz I, Hayes A, Waack S, Morgenstern B. 2006. AUGUSTUS: ab initio prediction of alternative transcripts. *Nucleic Acids Res.* **34**:W435–W439.
- Sutherland-Smith M. 2015. Suidae and Tayassuidae (Wild Pigs, Peccaries). In *Fowler's zoo and wild animal medicine*, Vol. 8, Elsevier, p. 568–584.
- Tarailo-Graovac M, Chen N. 2009. Using RepeatMasker to identify repetitive elements in genomic sequences. *Curr Protoc Bioinform.* **25**:4.10.1–4.10.14.
- Templeton AR. 2008. The reality and importance of founder speciation in evolution. *BioEssays* **30**:470–479.
- Thomsen PD, Høyheim B, Christensen K. 1996. Recent fusion events during evolution of pig chromosomes 3 and 6 identified by comparison with the babirusa karyotype. *Cytogenet Genome Res.* **73**:203–208.
- Truong Nguyen D, Lee K, Choi H, Choi M-K, Le MT, Song N, Kim J-H, Seo HG, Oh J-W, Lee K, et al. 2012. The complete swine olfactory subgenome: expansion of the olfactory gene repertoire in the pig genome. *BMC Genomics* **13**:584.
- UniProt Consortium. 2015. UniProt: a hub for protein information. *Nucleic Acids Res.* **43**:D204–D212.
- Vamathevan JJ, Hall MD, Hasan S, Woollard PM, Xu M, Yang Y, Li X, Wang X, Kenny S, Brown JR, et al. 2013. Minipig and beagle animal model genomes aid species selection in pharmaceutical discovery and development. *Toxicol Appl Pharmacol.* **270**:149–157.
- van der Made J, Morales J, Montoya P. 2006. Late Miocene turnover in the Spanish mammal record in relation to palaeoclimate and the Messinian Salinity Crisis. *Palaeogeogr Palaeoclimatol Palaeoecol.* **238**:228–246.
- Veldhuizen MG, Shepard TG, Wang MF, Marks LE. 2009. Coactivation of gustatory and olfactory signals in flavor perception. *Chem Senses.* **35**:121–133.
- Villafuerte MQ, Macadam I, Daron J, Katzfey J, Cinco TA, Ares ED, Jones RG. 2020. Projected changes in rainfall and temperature over the Philippines from multiple dynamical downscaling models. *Int J Climatol.* **40**:1784–1804.
- Vurtture GW, Sedlazeck FJ, Nattestad M, Underwood CJ, Fang H, Gurtowski J, Schatz MC, Berger B. 2017. GenomeScope: fast reference-free genome profiling from short reads. *Bioinformatics* **33**:2202–2204.
- Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Flicek P, et al. 2019. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience* **9**:giaa051.
- Warr A, Affara N, Aken B, Beiki H, Bickhart DM, Billis K, Chow W, Eory L, Finlayson HA, Flicek P, et al. 2020. An improved pig reference genome sequence to enable pig genetics and genomics research. *Gigascience* **9**:1–14.
- Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, et al. 2018. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**:W296–W303.
- Wertheim JO, Murrell B, Smith MD, Pond SLK, Scheffler K. 2015. RELAX: detecting relaxed selection in a phylogenetic framework. *Mol Biol Evol.* **32**:820–832.
- Wicke S, Müller KF, dePamphilis CW, Quandt D, Bellot S, Schneeweiss GM. 2016. Mechanistic model of evolutionary rate variation en route to a nonphotosynthetic lifestyle in plants. *Proc Natl Acad Sci U S A.* **113**:9045–9050.
- Wu J, Mao X, Cai T, Luo J, Wei L. 2006. KOBAS server: a web-based platform for automated annotation and pathway identification. *Nucleic Acids Res.* **34**:w720–w724.
- Wucher V, Legeai F, Hedan B, Rizk G, Lagoutte L, Leeb T, Jagannathan V, Cadieu E, David A, Lohi H, Cirera S et al. 2017. FEELnc: a tool for long non-coding RNA annotation and its application to the dog transcriptome. *Nucleic Acids Res.* **45**:e57.
- Yang Z. 2007. PAML 4: phylogenetic analysis by maximum likelihood. *Mol Biol Evol.* **24**:1586–1591.
- Yang Y, Zhang Y, Ren B, Dixon JR, Ma J. 2019. Comparing 3D genome organization in multiple species using Phylo-HMRF. *Cell Syst.* **8**:494–505.e14.
- Zhang X, Zhang S, Zhao Q, Ming R, Tang H. 2019. Assembly of allele-aware, chromosomal-scale autopolyploid genomes based on Hi-C data. *Nat Plants.* **5**:833–845.
- Zhu Y, Li W, Yang B, Zhang Z, Ai H, Ren J, Huang L. 2017. Signatures of selection and interspecies introgression in the genome of Chinese domestic pigs. *Genome Biol Evol.* **9**:2592–2603.