

The pig X and Y Chromosomes: structure, sequence, and evolution

Benjamin M. Skinner,^{1,5} Carole A. Sargent,^{1,5} Carol Churcher,^{2,5,6} Toby Hunt,² Javier Herrero,^{3,4} Jane E. Loveland,² Matt Dunn,² Sandra Louzada,² Beiyuan Fu,² William Chow,² James Gilbert,² Siobhan Austin-Guest,² Kathryn Beal,³ Denise Carvalho-Silva,³ William Cheng,² Daria Gordon,² Darren Grafham,^{2,7} Matt Hardy,² Jo Harley,² Heidi Hauser,² Philip Howden,^{1,2} Kerstin Howe,² Kim Lachani,¹ Peter J.I. Ellis,^{1,8} Daniel Kelly,² Giselle Kerry,² James Kerwin,² Bee Ling Ng,² Glen Threadgold,² Thomas Wileman,² Jonathan M.D. Wood,² Fengtang Yang,² Jen Harrow,² Nabeel A. Affara,¹ and Chris Tyler-Smith²

¹Department of Pathology, University of Cambridge, Cambridge CB2 1QP, United Kingdom; ²Wellcome Trust Sanger Institute, Hinxton, Cambridge CB10 1SA, United Kingdom; ³European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, Cambridge CB10 1SD, United Kingdom; ⁴Bill Lyons Informatics Centre, UCL Cancer Institute, University College London, London WC1E 6BT, United Kingdom

We have generated an improved assembly and gene annotation of the pig X Chromosome, and a first draft assembly of the pig Y Chromosome, by sequencing BAC and fosmid clones from Duroc animals and incorporating information from optical mapping and fiber-FISH. The X Chromosome carries 1033 annotated genes, 690 of which are protein coding. Gene order closely matches that found in primates (including humans) and carnivores (including cats and dogs), which is inferred to be ancestral. Nevertheless, several protein-coding genes present on the human X Chromosome were absent from the pig, and 38 pig-specific X-chromosomal genes were annotated, 22 of which were olfactory receptors. The pig Y-specific Chromosome sequence generated here comprises 30 megabases (Mb). A 15-Mb subset of this sequence was assembled, revealing two clusters of male-specific low copy number genes, separated by an ampliconic region including the *HSFY* gene family, which together make up most of the short arm. Both clusters contain palindromes with high sequence identity, presumably maintained by gene conversion. Many of the ancestral X-related genes previously reported in at least one mammalian Y Chromosome are represented either as active genes or partial sequences. This sequencing project has allowed us to identify genes—both single copy and amplified—on the pig Y Chromosome, to compare the pig X and Y Chromosomes for homologous sequences, and thereby to reveal mechanisms underlying pig X and Y Chromosome evolution.

[Supplemental material is available for this article.]

The therian (marsupial and eutherian) sex chromosomes evolved originally from a homologous pair of autosomes (Ohno 1967a) ~170–180 million years ago (Livernois et al. 2012; Cortez et al. 2014) and have become extensively differentiated in terms of structure and sequence content. The gene content and organization of the emergent X Chromosome have been subject to strong conservation across different mammalian species with retention of much of the ancestral X (Ross et al. 2005; Bellott and Page 2009). In contrast, the acquisition of a dominant male sex-determining function and accumulation of male benefit genes (e.g., genes involved in regulating male germ cell differentiation) on the Y

Chromosome has been accompanied by the genetic isolation of much of the Y through suppression of recombination with the emergent X, subsequent degeneration with loss of much of the ancestral Y gene content, and dosage compensation of genes on the X Chromosome to restore equivalence of gene expression between males and females (Graves 2010; Bachtrog 2013). Selection has also acted to retain a strictly X-Y homologous pseudoautosomal region (PAR) that permits X-Y pairing during meiosis and within which there is obligate recombination between the sex chromosomes. The gene and sequence content of the PAR varies between species, reflecting processes of expansion and pruning of the PAR in different mammalian lineages (Otto et al. 2011).

Comparisons of X Chromosome sequences from several mammalian species have confirmed strong conservation of gene sequence and order (Chinwalla et al. 2002; Sandstedt and Tucker 2004). Groenen et al. (2012) published the first assembly of the porcine X Chromosome as part of the initial description of the

⁵Joint first authors

Present addresses: ⁶School of Clinical Medicine, University of Cambridge, Cambridge CB2 0SP, UK; ⁷Sheffield Diagnostic Genetics Service, Sheffield Children's NHS Foundation Trust, Western Bank, Sheffield S10 2TH, UK; ⁸School of Biosciences, University of Kent, Canterbury CT2 7NJ, UK

Corresponding author: cts@sanger.ac.uk

Article published online before print. Article, supplemental material, and publication date are at <http://www.genome.org/cgi/doi/10.1101/gr.188839.114>. Freely available online through the *Genome Research* Open Access option.

© 2016 Skinner et al. This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

pig genome sequence, and again this demonstrated conservation of synteny across the chromosome. Nonetheless, sequence gaps and ambiguities remained within this first assembly, complicating genomic studies within pigs and comparative studies between mammals.

In contrast to the broadly conserved X Chromosomes, the hemizygous nature of the Y Chromosome and suppression of recombination, in combination with normal processes of genome evolution, have led to a gradual degeneration of the chromosome over time, chromosomal rearrangements, and colonization by sequences from the X Chromosome and autosomes. Newly introduced genes will either drift or degenerate, or selection may act on variants to fix new genetic functions on the Y, particularly where these confer a male benefit. The haploid state of the sex chromosomes in males has led to the accumulation of male gametogenesis genes on both X and Y Chromosomes (Vallender and Lahn 2004). A further consequence of the nonrecombining status of the male-specific region of the Y is the relaxation of restraint on sequence amplification, leading to generation of ampliconic regions containing amplified gene and sequence families (Bellott and Page 2009).

The highly repetitive nature of many regions of mammalian Y Chromosomes has impeded the generation of complete chromosome sequences; while there are tens of mammalian genomes sequenced, only a small fraction have a Y assembly. These few assemblies, and several partial sequence assemblies, have permitted the elucidation of chromosome topology and gene order in human (Skaletsky et al. 2003), chimpanzee (Hughes et al. 2010), rhesus macaque (Hughes et al. 2012), wallaby (Murtagh et al. 2012), mouse (Soh et al. 2014), marmoset, rat, and opossum (Bellott et al. 2014), cattle (Elsik et al. 2009), horse (Paria et al. 2011), and cat and dog (Li et al. 2013a). These works show divergence in gene content, order, structure, and sequence between Y Chromosomes from different mammalian species. However, few data are available on the porcine Y Chromosome sequence or gene order, and their relationship to the X, despite the recent sequencing project for the pig genome (Groenen et al. 2012), remains unclear. Much of our knowledge of Y gene order comes from Quilter et al. (2002), who combined radiation hybrid mapping data with physical mapping of BAC clones to generate an ordered gene framework.

The current work presents a second-generation, much improved assembly and gene annotation of the porcine X Chromosome based on the Duroc X Chromosome. We also present the Y Chromosome sequence derived predominantly from Duroc, with some Meishan, which has permitted a first-generation assessment of the Y Chromosome short-arm gene content and order, and analysis of how this compares to other mammalian Y Chromosomes, the evolutionary processes leading to the current Y organization, and the structural relationships between the porcine sex chromosomes.

Results

A second-generation porcine Chromosome X assembly

Sequence statistics

The Chromosome X assembly (http://vega.sanger.ac.uk/Sus_scrofa/Location/Chromosome?r=X-WTSL) comprises 129,927,919 bp of sequence in five contigs, with 13 gaps and an N50 length of 4,824,757 bp. Compared with the previous 10.2 build, many

gaps have been filled and the order of sequences on the chromosome has been updated. Much of this improvement was informed by the use of optical mapping techniques, which helped resolve some of the more repetitive regions of the chromosome; an example can be seen in the short-arm clone CH242-202P13 (see Supplemental Methods Fig. OM7 for details on how the optical mapping approach was used here). Supplemental Figure S3 shows a dot-plot alignment of the 10.2 X with our X assembly, highlighting the regions of the chromosome for which the sequence order has been corrected.

The pseudoautosomal region in pig is of a similar size to the PAR in other closely related mammals (e.g., cattle) and has been discussed previously (Skinner et al. 2013). The precise location of the PAR boundary was recently confirmed to be within the gene *SHROOM2* (Das et al. 2013). A 33-kb region containing a lincRNA with homology to the PAR (X-WTSL: 1,840,693–1,874,130) can be found on Xq (X-WTSL: 114,853,327–114,886,764), and most likely arose via duplication and transposition from the PAR onto Xq.

Comparative X alignments

We aligned the X (and available Y) Chromosome sequences of nine mammalian species (Fig. 1). The previously documented high level of conservation of synteny is more apparent with the new pig X, as many of the reported breakpoints from cross-species comparisons to the build 10.2 X were due to errors that have been resolved. Li et al. (2013b) produced a genome assembly of a female Tibetan wild boar and reported regions of the genome with apparent inversions with respect to the Duroc assembly. We compared the X inversions with our new assembly and found that they lie outside the regions that have been resolved from the 10.2 X. That is, these remain as potential inversions between Duroc and Tibetan wild boar that require further investigation.

Gene content of the X

Chromosome X contains complex duplicated gene families such as olfactory receptors (OR) along with pseudogenes, which are hard to discriminate using only automatic pipelines. Since the reference assembly was high quality, the sequence underwent manual annotation to allow resolution of these gene families. Table 1 shows the updated annotation compared to the manual annotation available on the 10.2 X build and comparable statistics for the manual annotation of the Y Chromosome. The full gene annotation is provided in Supplemental Table S5 and is available through the Vega website. The majority (76%) of annotated loci in pig are shared with human. Many genes from the previous build that were disrupted by gaps have now been completed, and the number of long noncoding RNA loci has increased from 33 to 100.

Some genes with updated annotation from build 10.2 stand out as being of particular biological interest. Comparing the human and pig X Chromosomes, 38 coding loci in pig are not found in human. Twenty-two of these coding genes and five novel pseudogenes are in olfactory gene clusters. Pigs are known to have a large olfactory receptor repertoire (Nguyen et al. 2012), and this adds to the reported collection. Supplemental Figure S4 shows the improved assembly and annotation around one of the olfactory region clusters on Xq. The region lies within an inversion in the 10.2 assembly, corrected here and matching the gene order on the human X. The full list of genes present on pig Chromosome X, but not on the human X, is provided in Supplemental Table S7.

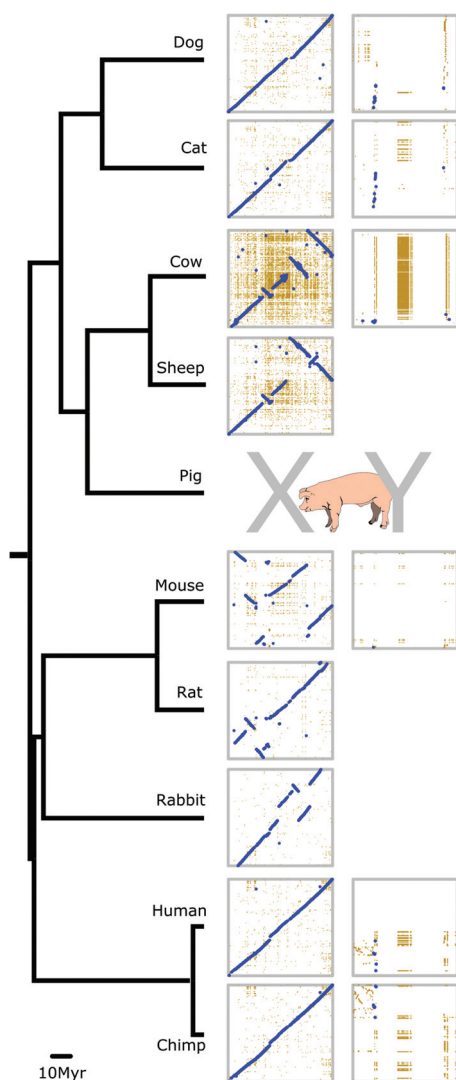


Figure 1. Comparative X and Y map. Sequenced X Chromosomes from nine mammals, plus available Y Chromosomes, aligned to our pig X and Y assemblies. In each dotplot, the pig chromosome is on the horizontal axis, and the subject chromosome is on the vertical axis. The cattle X sequence is plotted in reverse orientation. High-stringency alignments are shown in blue with less stringent alignments in yellow. Human, chimpanzee, cat, and dog retain the ancestral X arrangement. Sheep and cattle show a small number of rearrangements, while rodents and rabbit have a greater rate of chromosomal change. Chromosomes derived from shotgun assemblies are more prone to showing rearrangement and reflect the need for continuous assembly improvement. The Y alignments show highly variable organization, and different ancestral genes have amplified in different lineages (note that the sizes of the Y assemblies are not to scale here; see Supplemental Fig. S8 for larger versions).

A first-generation porcine Chromosome Y sequence assembly

Successfully assembled repetitive portions of Y Chromosomes have been generated only for a limited number of species (see, for example, the human, mouse, or chimpanzee Y) (Skaletsky et al. 2003; Hughes et al. 2010; Soh et al. 2014). Given the highly repetitive nature of the long arm of the pig Y Chromosome, we targeted the short arm, which contains most, if not all, of the single-copy material. The details of sequenced clones are given in Supplemental Table S1; the total sequence produced, including

finished and unfinished clones, without overlaps covers 29,953,871 bp, with 15,592,828 bp annotated and placed in the Vega assembly (http://vega.sanger.ac.uk/Sus_scrofa/Location/Chromosome?r=Y-WTSL). Here, we focus on the assembled and ordered contigs within that sequence.

Organization of the porcine Y Chromosome

The genes on Chromosome Y are organized into two main blocks of low copy number sequences (Fig. 2). These blocks are separated by a region containing highly amplified sequences including ~100 copies of the *HSFY* gene family (Skinner et al. 2015). Our final mapped sequence comprises seven contigs each in the distal and proximal blocks (Supplemental Fig. S5). Contigs that were assembled, but could not be assigned to the physical map, are included in a separate assembly (U_Y) in Vega. Two of these contigs appear to lie on Yp close to the centromere (Supplemental Figs. S6, S7).

Few low-copy loci have been identified on the Y long arm, which seems mostly composed of repetitive sequences (e.g., Quilter et al. 2002). Although we attempted to sequence one repetitive clone, it was not possible to assemble a framework physical or sequence map. The sequences we obtained belong to previously published male-specific pig repeat families (McGraw et al. 1988; Mileham et al. 1988; Thomsen et al. 1992; Pérez-Pérez and Barragán 1998). Metaphase FISH did, however, reveal a small low-copy region at the q terminus (Supplemental Fig. S13).

Gene-related content of the Y Chromosome

As with the X Chromosome, the Y-chromosomal sequence was run through the Otterlace/Zmap analysis pipeline, which performs homology searches, de novo sequence analysis, and gene predictions (Loveland et al. 2012). Repeat masking proved challenging due to the paucity of known pig-specific repeats. Manual annotation resolved this, as homology analysis is routinely run on-the-fly within the annotation tools, without repeat masking, to more accurately elucidate gene structures, especially using known Y-chromosomal genes from other species as targets and

Table 1. Comparison of the HAVANA manually annotated gene content between porcine build 10.2 X, our updated build of the X (X-17), and the new Y annotation (Y-13)

	10.2X	X-17 (WTSL-X)	Y-13 (WTSL-Y)
Total number of genes	632	1033	193
Total number of protein-coding	422	690	75
Known protein-coding	211	437	18
Novel protein-coding	210	245	57
Putative protein-coding	1	8	0
Noncoding genes			
LincRNA	20	77	12
Antisense	11	22	0
Sense intronic	0	1	1
Pseudogenes			
Processed pseudogenes	140	204	29
Unprocessed pseudogenes	2	23	73
Unitary pseudogenes	3	11	0
Transcribed unprocessed pseudogenes	1	4	3
Polymorphic pseudogenes	0	1	0

The number of identified coding genes has been substantially increased and brings the porcine X closer in gene content to other well-sequenced mammalian X Chromosomes (e.g., human and mouse).

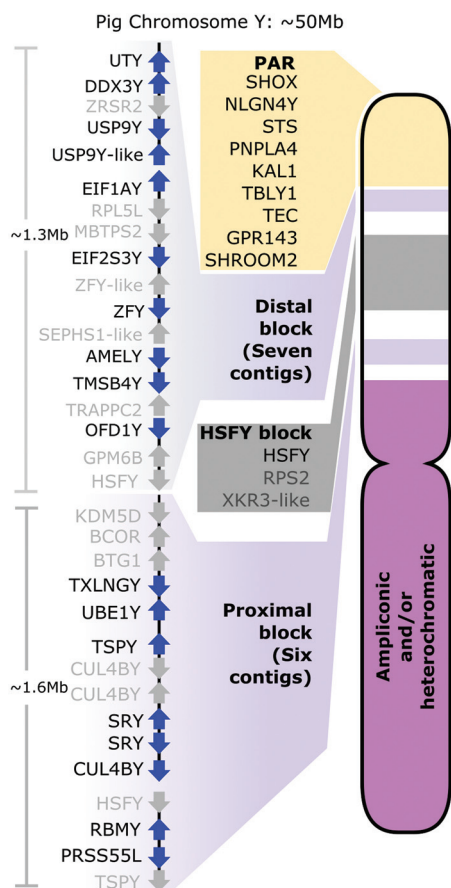


Figure 2. The organization of the pig Y Chromosome. All identified male-specific single-copy genes are on the short arm and split into two blocks by the ampliconic *HSFY* region. Genes (blue) and pseudogenes (gray) are shown within each block. The long and short arms toward the centromere appear to contain ampliconic or other repetitive sequences.

identifying those pig homologs present. Many of the ancestral X-related genes previously reported in other mammalian Y Chromosomes are represented either as active genes or partial sequences, some of the latter with supporting ESTs. There is also evidence for rearrangement of a number of Y-linked genes, which may have rendered them nonfunctional or modified their transcription products. An overview of Y gene loci is given in Table 2, and the full Y gene annotation table is provided as Supplemental Table S6. The contigs here show no evidence for novel pig genes, though these may yet be found in the ampliconic regions of the chromosome.

Ampliconic gene sequences

Although our sequence contigs are limited to the low-copy regions of the chromosome, the FISH data show regions where gene sequence-containing clones are present in multiple copies. An example is the *CUL4BY* gene, which has a partial copy in fosmid WTSI_1061-13A5. Probes detecting this sequence bind multiple targets proximal to *SRY* (and likely proximal to *RBMY*), as well as additional sites toward the centromeric end of the Y short arm (Supplemental Fig. S13F). The sequence supports gene expression from a “full-length” version of the sequence centromeric to *SRY*, and RT-PCR shows expression in testis, kidney, and brain

(Supplemental Fig. S2). A similar pattern of amplification exists for fosmid clones containing *TSPY*.

Regions of X–Y homology

We examined all sequenced Y clones for homology with the X. The overview X–Y comparison for the short arm of the Y is shown in Figure 3. Besides the X–Y homologous genes, a region of 90.25 kb on the distal block encompassing the genes (or gene fragments) *TRAPPC2-OFD1Y-GPM6B* has high sequence identity to the X, even in intronic and intergenic regions (Supplemental Fig. S1). *OFD1Y* is expressed highly in testis with lower expression in kidney and brain (Supplemental Fig. S2), but there is no evidence for expression of the Y copies of *TRAPPC2* or *GPM6B*. Previous work suggested that three X BAC clones around the olfactory receptor genes contain a sequence amplified on Yq; we did not find matches to these clones within our Y sequence data beyond repetitive elements; the potential for olfactory receptor-related sequences on the Y nevertheless remains.

Evolution of the porcine Chromosome Y

Inverted and duplicated blocks of sequence

Inversion and duplication to form palindromes has occurred around both *SRY* and the *CUL4BY* fragments (Fig. 4). The *SRY* gene itself is present in two head-to-head copies. There are unlikely to be more than two *SRY* copies on the chromosome (qPCR) (Supplemental Fig. S10). The pattern of markers at the breakpoint regions reveals that the *SRY* duplication preceded the *CUL4B* duplication (see Fig. 4; Supplemental Fig. S9). Transposable elements at both inversion boundaries are annotated as specific to the *Sus* lineage, suggesting these are relatively recent duplications; two copies of *SRY* could also be present in closely related suid species. The arms of the palindrome have high levels of sequence identity; we found no difference in the *SRY* sequence from clones on one arm versus the other arm. Our sequence contigs do not cover the centers of the palindromes (~20 kb is missing in each), so we do not know if the arms abut—it remains possible that there is a short stretch of unique sequence between them. Prior sequencing of the *SRY* gene has given no evidence for polymorphisms in the recovered *SRY* sequence from any individual, despite there being breed-specific differences—i.e., there are no heterozygous males identified (more than 300 *SRY* sequences are currently deposited in NCBI for *S. scrofa* alone). A further palindromic region lies at the proximal end of *USP9Y*, covering 56 kb, including the final exons 18–43 of the gene (Supplemental Fig. S11). Compared to *SRY* and *CUL4BY*, the breakpoints are less well defined, with sequence identity decreasing over some tens of base pairs.

Structural rearrangements compared to other species

We generated likely pathways of gene-only rearrangements from the ancestral Y Chromosome to pig, using data from Li et al. (2013a) updated with pig and cattle (Elsik et al. 2009) Y Chromosomes (Supplemental Fig. S12). Global alignments of chromosome content with other available Y Chromosome assemblies shown in Figure 1 are presented as larger versions in Supplemental Figure S8. Although little of the Y sequences align outside genic regions, the comparison highlights which ancestral gene families have become amplified in different lineages, e.g., *HSFY* in pigs and *TSPY* in species including dogs, horses, and humans (Paria et al. 2011; Xue and Tyler-Smith 2011; Li et al. 2013a).

Table 2. Genes on pig Chromosome Yp with expression status and supporting evidence from expressed sequence tag (EST) and Sequence Read Archive (SRA) libraries

Gene (Alias)	Accession	Expression	Copy number
<i>DDX3Y (DBY)</i>	OTTSUSG00000005192	Not restricted	1
<i>EIF1AY</i>	OTTSUSG00000005597	Not restricted	1
<i>EIF2S3Y</i>	OTTSUSG00000005598	Not restricted	1
<i>UTY (KDM6A)</i>	OTTSUSG00000002750	Not restricted	1
<i>OFD1Y</i>	OTTSUSG00000005612	Not restricted	1
<i>TMSB4Y</i>	OTTSUSG00000005246	Not restricted	1
<i>TXLNGY</i>	OTTSUSG00000005200	Not restricted	1
<i>USP9Y (DFFRY)</i>	OTTSUSG00000005756	Not restricted	1
<i>ZFY</i>	OTTSUSG00000002751	Not restricted	1
<i>AMELY</i>	OTTSUSG0000000517	Restricted: teeth (Hu et al. 1996)	1
<i>CUL4BY</i>	OTTSUSG00000005721	Restricted: predominantly testis (SRA and EST data)	One known intact expressed copy; potentially multicopy
<i>HSFY</i>	Multicopy	Restricted: predominantly testis (Skinner et al. 2015)	~100
<i>PRSS55L</i>	OTTSUSG00000005712	Restricted: testis (SRA data)	1
<i>RBMY</i>	OTTSUSG00000005714	Restricted: testis (SRA and EST data)	One known intact expressed copy; potentially multicopy
<i>SRY</i>	OTTSUSG00000002747 OTTSUSG00000005707	Restricted: testis (SRA and EST data), genital ridge (Daneau et al. 1996)	2
<i>TSPY</i>	OTTSUSG00000005701	Restricted: testis (SRA and EST data)	One known intact expressed copy; potentially multicopy
<i>UBA1Y (UBE1Y)</i>	OTTSUSG00000005209	Restricted: testis (SRA and EST data)	1

Genes are grouped by pattern of expression. A more comprehensive table may be found in Supplemental Table S4.

Discussion

We present here an updated and substantially improved assembly of the pig Chromosome X and a first-generation assembly of sequence from the pig Chromosome Y. These sequences have also allowed us to confirm the strong conservation of the X and recover information on the evolution of the Y Chromosome and how this relates to sex chromosome evolution in other mammals.

An updated assembly and annotation of the porcine Chromosome X

The picture of mammalian X Chromosomes is one of general structural stability (see Fig. 1). Specific lineages, such as rodents, have many more X rearrangements than others, but these species are characterized by a globally higher number of chromosomal rearrangements (Stanyon et al. 1999). Apparent inversions and translocations in the pig X, relative to the ancestral X, detected in previous builds, are corrected here to an order more reflective of the inferred ancestral X Chromosome. Similar findings may be seen with other mammalian X Chromosomes as the quality of the assemblies improves. It paints a stark contrast to the dynamic and ever-changing mammalian Y Chromosomes that we discuss below.

Improved gene annotation of the porcine Chromosome X

The revised gene annotation increases the number of protein-coding genes identified on the pig X to 690, bringing the reported gene content closer to that identified in the X Chromosomes of well-studied species (i.e., humans and mice, with 813 and 940 protein-coding genes, respectively). The majority of X Chromosome genes are shared between species (76% of annotated pig genes shared with human), in accordance with Ohno's law (Ohno 1967b). We have highlighted some specific genes of interest with an updated status from build 10.2 X.

Eleven protein-coding genes present on the human X have been annotated as unitary pseudogenes (also known as loss-of-function genes) on the pig X (Supplemental Table S10). These include: *GUCY2F*, a possible candidate for involvement in X-linked retinitis pigmentosa (Yang et al. 1996); *AWAT1*, an acyl-CoA wax alcohol acyltransferase involved in sebum production (Turkish et al. 2005); *ITIH6*, a trypsin inhibitor; and *RAB41*, a member of the RAS oncogene family (Supplemental Results, section 3).

Other regions of difference lie in the cancer/testis (CT) antigen clusters found in humans and other primates, which are significantly smaller in pig. This is in line with evidence that enlarged CT antigen clusters arose due to a recent amplification in primates (Zhang and Su 2014), perhaps driven by a retrotransposition event. Their potential functions remain unknown, though they may have been involved in primate speciation (Zhang and Su 2014).

Apart from the olfactory receptor gene clusters, we have not found evidence for widespread ampliconic gene families on the pig X. This contrasts with the X Chromosomes of human and mouse, which contain independently amplified gene families, with little overlap between the species (Mueller et al. 2013). Human X Chromosomes contain multiple inverted repeats with high sequence identity, enriched for testis-expressed genes (Warburton et al. 2004). Mice have a greater number of X-ampliconic genes than humans, apparently driven by a genomic conflict between X and Y sequences; the gene *Sly* on mouse Y represses gene expression from sex chromosomes in spermatids, and the copy number of X genes has increased in response to maintain expression of key genes (Ellis et al. 2011). These examples led to an expectation that this might be a general feature of mammalian X Chromosomes and that the pig X would also contain unique ampliconic testis-expressed genes. However, we have no evidence supporting this—either due to a lack of such genes, the paucity of pig-derived evidence (e.g., ESTs, cDNAs, RNA-seq data), or because ampliconic genes lie within the remaining gaps within the assembly.

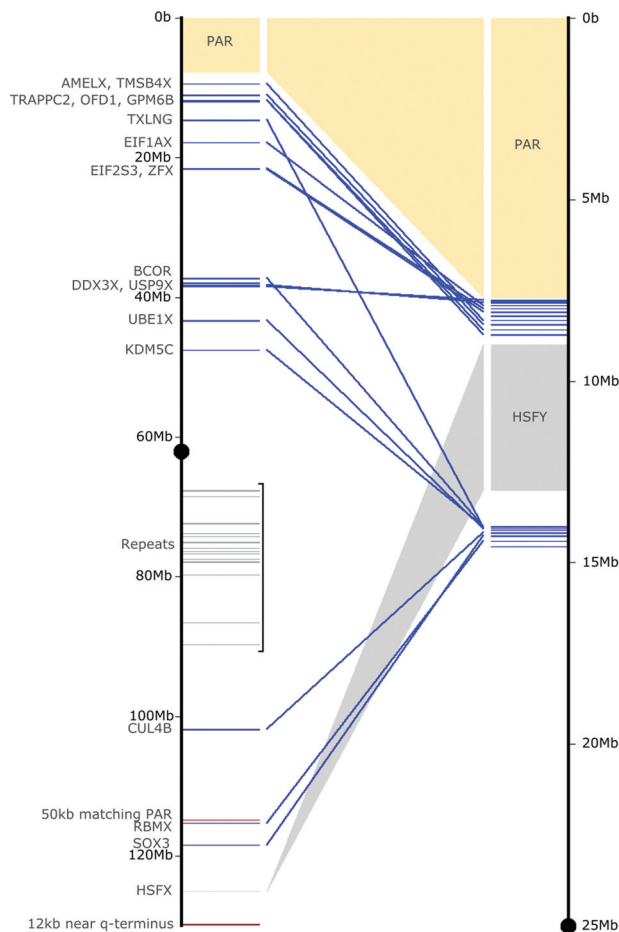


Figure 3. Homology between the X and Y. Outline of X-Y homology regions detected between the pig X and Yp sequences. Centromeres are black dots; the PAR is yellow. Lines for genes are blue, repetitive content is gray, and sequences as yet unplaced on the Y are red. The 50-kb region at X-114 Mb is an X-X transposition from the PAR. The amplified HSFY region is shown in dark gray. Regions of repetitive content correlate well with CGH patterns we found previously (Skinner et al. 2013).

The porcine Chromosome Y

A striking aspect of the Chromosome Y organization is that the known single-copy male-specific genes are found in tight clusters of contigs spanning only 2 or 3 Mb of sequence. This is a pattern observed in some other mammalian Y Chromosomes—for example, mice (Soh et al. 2014), cattle (Elsik et al. 2009), and cats and dogs (Li et al. 2013a). Each lineage appears to have preserved a small region of key ancestral X genes, while the remainder of the chromosome has evolved in a species-specific manner. In contrast, in human and chimpanzee Y Chromosomes (Skaletsky et al. 2003; Hughes et al. 2010), the single-copy material is more widely distributed along the chromosome. Patterns of hybridization from FISH suggest to us that there is additional single-copy sequence on Yq, including near/at the Yq terminus, but we were unable to elucidate the sequences involved.

Organization of the pig Y

Palindromic sequences

A recurring feature of the Y sequences we have assembled is the presence of palindromic regions, each on the order of 120 kb

end to end, and reminiscent of ampliconic structures found on the mouse (Soh et al. 2014), human (Skaletsky et al. 2003), and chimpanzee (Hughes et al. 2010) Y Chromosomes. The palindromes on the pig Y have resulted from duplication and inversion of sequences, and at least three such palindromes are present. Two have very high levels of sequence identity; the inverted structure may facilitate the process of gene conversion by allowing the formation of stem-loop structures, as seen in the palindromes on the human Y Chromosome (e.g., Rozen et al. 2003; Hallast et al. 2013). These palindromic sequences are also reminiscent of the “core duplicons” found in humans and great apes (Marques-Bonet and Eichler 2009). In the case of pig, however, there is no evidence for novel gene structures and functional innovation as a result of the duplications.

The first two of these palindromes are in the proximal gene block. One encompasses the two copies of the male-determining gene *SRY*. Multiple *SRY* copies are found in dog (Li et al. 2013a) and some rodent species (e.g., Lundrigan and Tucker 1997; Murata et al. 2010; Prokop et al. 2013), but there has previously been no suggestion of this being the case in the pig. While most other species with multiple *SRY* copies have identifiable sequence differences between the copies in a single individual, there is also a known example in rabbits of a palindrome encompassing *SRY*, with gene conversion maintaining sequence identity (Gerald et al. 2010). A similar mechanism may maintain the sequence identity between the palindrome arms in pig.

The third palindrome is found in the distal gene cluster. Unlike the previous two palindromes, both breakpoint ends lie within complete transposable elements (TEs). Sequence identity between the palindrome arms is lower around the breakpoints, perhaps indicating that the duplication results from an older event or that homogenizing mechanisms such as gene conversion have been less active. In all three palindromes, the TEs at boundaries are annotated as deriving from within the pig lineage—these are not ancient repeat elements and show the ongoing impact of repetitive content in the genome. Extant suids diverged after ~10–15 Mya, and the copy number of the genes involved across these lineages remains to be identified.

Ampliconic sequences

Most mapped mammalian Y Chromosomes have been found to contain multicopy gene families (e.g., Li et al. 2013a), and the pig is no exception. Outside the palindromes, other sequences have amplified to a much greater extent. There are three gene families of note here, all involved in amplifications in other species and with functions suggesting involvement in spermatogenesis.

The *CUL4B* fragments: Cullin fragments are found proximal to *SRY* and the active *CUL4B* gene and appear to form part of an ampliconic region located toward the centromere. Ubiquitylation pathways are an important part of gamete development. *CUL4BX* gene defects lead to reduced oocyte survival (Yu et al. 2013) and reduced testis volume in human males (Tarpey et al. 2007).

The *TSPY* fragments: These appear to be interspersed in the region closer to the centromere, but it is not clear how they are arranged. *TSPY* is an ampliconic gene in many mammalian species, from artiodactyls to primates (Xue and Tyler-Smith 2011); the genes are involved in spermatogenesis (Xue and Tyler-Smith 2011) and, in cattle, copy number variation of this gene has been linked with fertility in bulls (Hamilton et al. 2012).

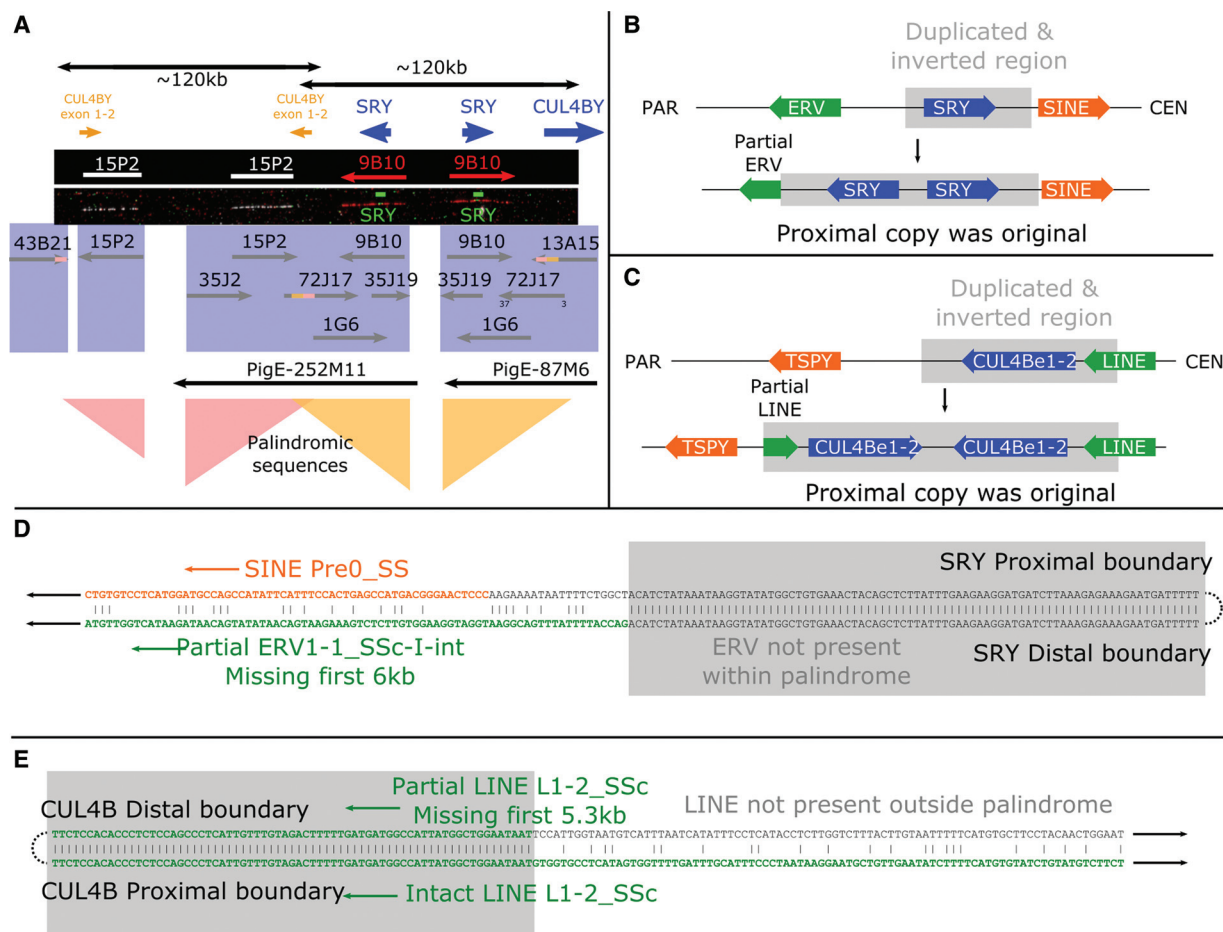


Figure 4. The pig *SRY* region. The Yp proximal block of genes contains two overlapping palindromes of ~120 kb each. These surround the duplicated sequences *CUL4BY* exons 1–2 and *SRY*. (A) FISH results from Y fosmid clones and probes for the *SRY* gene are shown with the BAC and fosmid clone sequences found mapping to the region. The inversion boundaries are both identifiable; the *CUL4BY* inversion runs from the last 3 kb of 43B21 to within 72J17; the *SRY* inversion begins also within 72J17 and runs to 13A15. A schematic view is also shown of the regions surrounding the *SRY* (B) and the *CUL4B* duplications (C). The *SRY* duplication disrupts an ERV element, revealing the proximal copy to be ancestral. The *CUL4B* duplication copies part of a LINE element, again revealing the proximal copy to be ancestral. The sequence alignments across the inversion breakpoints are shown in more detail for *SRY* (D) and *CUL4B* (E). The order of events was therefore a duplication of *SRY*, including the first two exons of *CUL4B*, followed by duplication of the region around the *CUL4B* copy.

The *HSFY* family: These genes are predominantly found in a block between the proximal and distal low-copy gene clusters and show evidence for recent amplification in the *Sus* lineage to ~100 copies, with independent amplification in other suid species (Skinner et al. 2015) and further independent amplification in cattle (Chang et al. 2013). *HSFY* is expressed in pig testis and may have a role in spermatogenesis, though biological function remains to be established.

Other amplified sequences

Yq is dominated by repeat sequences (as demonstrated by the painting pattern from FISH using several BAC and fosmid clones). These clones are composed almost entirely of sequences related to male-specific (or enriched) repeats described previously for pig (e.g., Mileham et al. 1988), and thus more detailed study is needed to understand the organization of this arm of the chromosome. There is some evidence that related sequences are expressed in testis-derived transcripts; however, expression is not exclusive to this tissue, or to males, with sequence homologies also detected in

transcripts from pig uterus. FISH experiments suggest that there is single-copy sequence on Yq, including at or near the Yq terminus, but we were unable to isolate these sequences.

We did not find evidence for exclusively testis-expressed ampliconic sequences. However, ampliconic sequences cause difficulties for assembly that our physical mapping approach was not able to overcome. It is likely that other amplified sequences of biological interest remain to be discovered on the pig Y Chromosome, both on the short and long arms.

Comparative chromosome organization and gene order between mammals

Previous work from primates, mouse, cat, and dog has reconstructed a putative ancestral eutherian Y Chromosome (Li et al. 2013a) based on gene order. We incorporated our pig gene order into this and added information from the cattle Y sequence assembly (Supplemental Fig. S12; Elsik et al. 2009). One group of genes stands out from the comparison: *USP9Y-DDX3Y-UTY* is the only ancestral cluster of genes that have retained their proximity to each other in all the studied species. There may be a selective

disadvantage to disrupting this organization, as has been proposed for other conserved syntenic blocks in general (Larkin et al. 2009) and for these genes in particular. Both *USP9Y* and *DDX3Y* have been implicated as important in human spermatogenesis, though they may not be essential in all great apes (Tyler-Smith 2008).

TRAPPC2P–OFDIY–GPM6B: a potential transposition from the X

Outside the PAR, there are regions of homologous sequence between the X and Y Chromosomes. Most of the X-Y homologies could be attributed to known X-Y homologous genes or to repetitive sequences, such as endogenous retroviral (ERV) families enriched on the sex chromosomes. An exception was the 90.25-kb region *TRAPPC2-OFDI-GPM6B*, which retains 87% sequence identity across the region and is interrupted only by recent insertion of transposable elements in the Y. The orthologous region has been subject to transposition onto the Y from the X Chromosome in dogs (Li et al. 2013a), and a similar transposition affecting the *RAB9A–SEDL–OFDIY* genes has occurred in the primate lineage (Chang et al. 2011). Such a large region of homology, including the introns and intergenic regions, argues against mechanisms such as gene conversion, and the proximal end of the region lies within a transposon; consequently, we consider this as evidence suggesting a transposition of this region in the pig (see Supplemental Fig. S1).

OFDI is involved in cilia formation, with gene defects affecting multiple tissues (Thauvin-Robinet et al. 2013), and ciliopathies have been implicated in fertility issues (Fry et al. 2014). It is likely that the testis-expressed *OFDIY* has repeatedly acquired a function in sperm development in different mammalian lineages. Notably, the X copies of *OFDI* and also *CUL4B* have been found to be substantially down-regulated in teratozoospermic men (Platts et al. 2007).

Conclusion

This work presents an improvement to the pig Chromosome X assembly and gene annotation, and the first assembly of sequence for the pig Chromosome Y. The assemblies we have generated have allowed new insights into the content and evolution of the pig sex chromosomes and provide an important resource for the pig genomics community.

Methods

Library construction and sequencing

Chromosome X clones (BAC clones from CHORI 242 library) were sequenced previously under the auspices of the Swine Genome Sequencing Consortium (Groenen et al. 2012).

Phytohaemagglutinin-stimulated peripheral blood culture from a Duroc boar was used to prepare chromosomes for flow sorting. Flow-sorted Y Chromosomes were purified, and 30- to 50-kb-sized fragments were cloned into the pCC1Fos vector (library WTSL_1061: <http://www.ncbi.nlm.nih.gov/clone/library/genomic/330/>) (Supplemental Methods, sections 1–3). Clones for sequencing were targeted by minimal overlapping clones on a fingerprint contig (FPC) map. The targeted 897 clones for the Y Chromosome were sequenced using a combination of three different sequencing platforms: capillary, Illumina, and 454 (Roche). Clones were assembled using a combination of four assembly scripts to produce de novo assemblies. Manual alignment of clone

sequences was used to build the clone map, expanding from clones containing known genes. These contigs were oriented and ordered using fiber-FISH on single DNA-molecule fibers (Supplemental Methods, sections 4, 5).

Molecular combing and FISH

Single-molecule DNA fibers were prepared by molecular combing (Michalet et al. 1997). Purified fosmid DNA was amplified and labeled as described previously (Gribble et al. 2013). Fluorescence in situ hybridization followed standard protocols (Supplemental Methods, section 6). Probes were detected with fluorescently conjugated antibodies. Slides were mounted with SlowFade Gold mounting solution containing 4',6-diamidino-2-phenylindole (Molecular Probes/Invitrogen) and visualized on a Zeiss AxioImager D1 microscope. Digital image capture and processing were carried out using the SmartCapture software (Digital Scientific UK).

X and Y gene annotation, sequence content, and chromosomal evolution

Manual annotation on the pig X and Y Chromosomes was performed using the Otterlace/Zmap suite of annotation tools (Loveland et al. 2012) following previously established annotation protocols (Harrow et al. 2012; Dawson et al. 2013). The assembled chromosomes were run through an annotation pipeline (Searle et al. 2004), aligning EST, mRNA, and protein libraries against the chromosomes with all annotated gene structures (transcripts) supported by at least one form of this transcriptional evidence. The HUGO Gene Nomenclature Committee (HGNC) (Seal et al. 2011) naming convention was used whenever possible for all pig genes; otherwise, HAVANA naming conventions (<http://www.sanger.ac.uk/research/projects/vertebrategenome/havana>) were followed.

RepeatMasker (Smit et al. 1996) was used to identify repetitive elements within Y contigs. Targeted resequencing was performed across specific genes to confirm their structure (primers given in Supplemental Table S4). Regions of X-Y homology were identified by comparing the repeat-masked X assembly to all sequenced repeat-masked Y clones (mapped and unmapped) using LASTZ (Harris 2007) with default parameters. Evolutionary analyses between X and Y gene pairs were conducted using the Nei-Gojobori model (Nei and Gojobori 1986) in MEGA5 (Tamura et al. 2011). For each pair, positions containing gaps and missing data were eliminated. Reconstruction of ancestral Y Chromosome organizations was performed using the Multiple Genomes Rearrangement (MGR) program (Bourque and Pevzner 2002) to calculate optimal rearrangement pathways between each species, as previously described (Skinner and Griffin 2012). Full information is given in Supplemental Methods, section 7.

Gene expression

RT-PCR was used to confirm expression status of selected genes in five tissues (brain, liver, kidney, side muscle, testis), obtained from the same boar from which blood cultures were derived. Samples were taken from tissues stored in RNAlater (Qiagen) and homogenized in TRIzol. Nucleic acids were extracted with phenol-chloroform and DNase I treated. RNA was precipitated with isopropanol and stored at 1 µg/µl in ddH₂O at –80°C. RT-PCR was carried out using a OneStep RT-PCR kit (Qiagen) on 25 ng of total RNA (Supplemental Methods, section 8). Primer sequences are given in Supplemental Table S13.

Copy-number estimation of *SRY* by qPCR

Primers were designed to amplify a 1447-bp region across the majority of the *SRY* ORF and UTRs (F:TAATGGCCGAAAGGAAAGG; R:TGGCTAATCACGGGAACAAC), and products were generated using a MyTaq Red kit (Bioline) using the following profile: 95°C for 3 min, 35 cycles of 95°C/53°C/72°C for 15 sec/15 sec/2 min, with a final extension of 72°C for 10 min. Two female Duroc gDNAs were spiked with dilutions of the purified *SRY* product to give a standard curve of four copies *SRY* per genome to 0.25 copies per genome (assuming diploid genome size of 6 Gb) (Animal Genome Size Database; Gregory 2006). qPCR was performed using a SYBR-FAST qPCR kit (Kapa Biosystems) on the spiked females and on five male Duroc gDNAs with primers for *SRY* and the autosomal (SSC10) gene *NEK7* (Supplemental Table S6). Annealing temperature was optimized at 57°C. Cycling conditions were 95°C for 3 min, followed by 40 cycles of 95°C/57°C/72°C for 10 sec/20 sec/30 sec. The fluorescent signal threshold crossing point (C_t) was normalized to the average signal from *NEK7* to produce a normalized ΔC_t . The data from spiked female gDNA was used to construct a standard curve relating *SRY* signal to *NEK7* signal; from this, an estimate of the absolute *SRY* copy number in the male gDNA samples was produced (see Supplemental Methods, section 9).

Data access

All sequence and annotation is available via the Vega Genome Browser (<http://vega.sanger.ac.uk/index.html>), and the complete chromosomal assemblies can be accessed by FTP (<ftp://ftp.sanger.ac.uk/pub/vega/pig/>). The pseudoautosomal region of X/Y homology between the X and Y Chromosomes is represented on the X Chromosome only in Vega and Ensembl. It is marked as an assembly exception in both chromosomes, but the underlying genomic sequence and annotation is that of Chromosome X. Only the unique regions of Chromosome Y are stored and annotated. The complete Y Chromosome is represented by filling the “gaps” with the PAR regions from the X Chromosome. Raw sequence data have been submitted to the European Nucleotide Archive (ENA; <http://www.ebi.ac.uk/ena/>) under accession number ERP001277.

Acknowledgments

We thank Genus for providing the Duroc boar from which a cell line was established and tissue samples obtained. We also thank the Wellcome Trust Sanger Institute core teams for fingerprinting, mapping, archiving, library construction, sequence improvement, and sequencing which underpins this work. This work was funded by BBSRC grant BB/F021372/1. The Flow Cytometry and Cytogenetics Core Facilities at the Wellcome Trust Sanger Institute and Sanger investigators are funded by the Wellcome Trust (grant number WT098051). K.B., D.C.-S., and J.H. acknowledge support from the Wellcome Trust (WT095908), the BBSRC (BB/I025506/1), and the European Molecular Biology Laboratory. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007–2013) under grant agreement no. 222664 (“Quantomics”). This publication reflects only the authors' views and the European Community is not liable for any use that may be made of the information contained herein.

References

Bachtrog D. 2013. Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. *Nat Rev Genet* **14**: 113–124.

- Bellott DW, Page DC. 2009. Reconstructing the evolution of vertebrate sex chromosomes. *Cold Spring Harb Symp Quant Biol* **74**: 345–353.
- Bellott DW, Hughes JF, Skaletsky H, Brown LG, Pyntikova T, Cho T-J, Koutseva N, Zaghul S, Graves T, Rock S, et al. 2014. Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* **508**: 494–499.
- Bourque G, Pevzner PA. 2002. Genome-scale evolution: reconstructing gene orders in the ancestral species. *Genome Res* **12**: 26–36.
- Chang T-C, Klabnik JL, Liu W-S. 2011. Regional selection acting on the OFD1 gene family. *PLoS One* **6**: e26195.
- Chang T-C, Yang Y, Retzel EF, Liu W-S. 2013. Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development. *Proc Natl Acad Sci* **110**: 12373–12378.
- Chinwalla AT, Cook LL, Delehaunty KD, Fellwag LA, Fulton RS, Graves TA, Hillier LW, Mardis ER, McPherson JD, et al. 2002. Initial sequencing and comparative analysis of the mouse genome. *Nature* **420**: 520–562.
- Cortez D, Marin R, Toledo-Flores D, Froidevaux L, Liechti A, Waters PD, Grütznér F, Kaessmann H. 2014. Origins and functional evolution of Y chromosomes across mammals. *Nature* **508**: 488–493.
- Daneau I, Ethier JF, Lussier JG, Silversides DW. 1996. Porcine *SRY* gene locus and genital ridge expression. *Biol Reprod* **55**: 47–53.
- Das PJ, Mishra DK, Ghosh S, Avila F, Johnson GA, Chowdhary BP, Raudsepp T. 2013. Comparative organization and gene expression profiles of the porcine pseudoautosomal region. *Cytogenet Genome Res* **141**: 26–36.
- Dawson HD, Loveland JE, Pascal G, Gilbert JGR, Uenishi H, Mann KM, Sang Y, Zhang J, Carvalho-Silva D, Hunt T, et al. 2013. Structural and functional annotation of the porcine immunome. *BMC Genomics* **14**: 332.
- Ellis PJI, Bacon J, Affara NA. 2011. Association of Sly with sex-linked gene amplification during mouse evolution: a side effect of genomic conflict in spermatids? *Hum Mol Genet* **20**: 3010–3021.
- Elsik CG, Tellam RL, Worley KC. 2009. The genome sequence of taurine cattle: a window to ruminant biology and evolution. *Science* **324**: 522–528.
- Fry AM, Leaper MJ, Bayliss R. 2014. The primary cilium: guardian of organ development and homeostasis. *Organogenesis* **10**: 62–68.
- Geraldes A, Rambo T, Wing RA, Ferrand N, Nachman MW. 2010. Extensive gene conversion drives the concerted evolution of paralogous copies of the *SRY* gene in European rabbits. *Mol Biol Evol* **27**: 2437–2440.
- Graves JAM. 2010. Review: sex chromosome evolution and the expression of sex-specific genes in the placenta. *Placenta* **31**: S27–S32.
- Gregory TR. 2006. Animal genome size database. <http://www.genomesize.com>.
- Gribble SM, Wiseman FK, Clayton S, Prigmore E, Langley E, Yang F, Maguire S, Fu B, Rajan D, Sheppard O, et al. 2013. Massively parallel sequencing reveals the complex structure of an irradiated human chromosome on a mouse background in the Tc1 model of down syndrome. *PLoS One* **8**: e60482.
- 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.
- Hallast P, Balaesque P, Bowden GR, Ballereau S, Jobling MA. 2013. Recombination dynamics of a human Y-chromosomal palindrome: rapid GC-biased gene conversion, multi-kilobase conversion tracts, and rare inversions. *PLoS Genet* **9**: e1003666.
- Hamilton CK, Verduzco-Gómez AR, Favetta LA, Blondin P, King WA. 2012. Testis-specific protein Y-encoded copy number is correlated to its expression and the field fertility of Canadian Holstein bulls. *Sex Dev* **6**: 231–239.
- Harris RS. 2007. “Improved pairwise alignment of genomic DNA.” PhD thesis, The Pennsylvania State University, University Park.
- Harrow J, Frankish A, Gonzalez JM, Tapanari E, Diekhans M, Kokocinski F, Aken BL, Barrell D, Zadissa A, Searle S, et al. 2012. GENCODE: the reference human genome annotation for The ENCODE Project. *Genome Res* **22**: 1760–1774.
- Hu CC, Bartlett JD, Zhang CH, Qian Q, Ryu OH, Simmer JP. 1996. Cloning, cDNA sequence, and alternative splicing of porcine amelogenin mRNAs. *J Dent Res* **75**: 1735–1741.
- Hughes JF, Skaletsky H, Pyntikova T, Graves TA, van Daalen SKM, Minx PJ, Fulton RS, McGrath SD, Locke DP, Friedman C, et al. 2010. Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* **463**: 536–539.
- Hughes JF, Skaletsky H, Page DC. 2012. Sequencing of rhesus macaque Y chromosome clarifies origins and evolution of the DAZ (Deleted in AZoospermia) genes. *Bioessays* **34**: 1035–1044.
- Larkin DM, Pape G, Donthu R, Auville L, Welge M, Lewin HA. 2009. Breakpoint regions and homologous synteny blocks in chromosomes have different evolutionary histories. *Genome Res* **19**: 770–777.
- Li G, Davis BW, Raudsepp T, Wilkerson AJP, Mason VC, Ferguson-Smith M, O'Brien PC, Waters PD, Murphy WJ. 2013a. Comparative analysis of

- mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Res* **23**: 1486–1495.
- Li M, Tian S, Jin L, Zhou G, Li Y, Zhang Y, Wang T, Yeung CKL, Chen L, Ma J, et al. 2013b. Genomic analyses identify distinct patterns of selection in domesticated pigs and Tibetan wild boars. *Nat Genet* **45**: 1431–1438.
- Livernois AM, Graves JA, Waters PD. 2012. The origin and evolution of vertebrate sex chromosomes and dosage compensation. *Heredity* **108**: 50–58.
- Loveland JE, Gilbert JGR, Griffiths E, Harrow JL. 2012. Community gene annotation in practice. *Database (Oxford)* **2012**: bas009.
- Lundrigan BL, Tucker PK. 1997. Evidence for multiple functional copies of the male sex-determining locus, Sry, in African murine rodents. *J Mol Evol* **45**: 60–65.
- Marques-Bonet T, Eichler EE. 2009. The evolution of human segmental duplications and the core duplication hypothesis. *Cold Spring Harb Symp Quant Biol* **74**: 355–362.
- McGraw RA, Jacobson RJ, Akamatsu M. 1988. A male-specific repeated DNA sequence in the domestic pig. *Nucleic Acids Res* **16**: 10389.
- Michalet X, Ekong R, Fougerousse F, Rousseaux S, Schurra C, Hornigold N, van Slegtenhorst M, Wolfe J, Povey S, Beckmann JS, et al. 1997. Dynamic molecular combing: stretching the whole human genome for high-resolution studies. *Science* **277**: 1518–1523.
- Mileham AJ, Siggens KW, Plastow GS. 1988. Isolation of a porcine male specific DNA sequence. *Nucleic Acids Res* **16**: 11842.
- Mueller JL, Skaletsky H, Brown LG, Zaghul S, Rock S, Graves T, Auger K, Warren WC, Wilson RK, Page DC. 2013. Independent specialization of the human and mouse X chromosomes for the male germ line. *Nat Genet* **45**: 1083–1087.
- Murata C, Yamada F, Kawachi N, Matsuda Y, Kuroiwa A. 2010. Multiple copies of SRY on the large Y chromosome of the Okinawa spiny rat, *Tokudaia muenninki*. *Chromosome Res* **18**: 623–634.
- Murtagh VJ, O'Meally D, Sankovic N, Delbridge ML, Kuroki Y, Boore JL, Toyoda A, Jordan KS, Pask AJ, Renfree MB, et al. 2012. Evolutionary history of novel genes on the tammar wallaby Y chromosome: Implications for sex chromosome evolution. *Genome Res* **22**: 498–507.
- Nei M, Gojobori T. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol Biol Evol* **3**: 418–426.
- Nguyen DT, Lee K, Choi H, Choi M, 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.
- Ohno S. 1967a. Conservation of the original X and homology of the X-linked genes in placental mammals. In *Sex chromosomes and sex-linked genes: monographs on endocrinology*, pp. 46–73. Springer, Berlin/Heidelberg.
- Ohno S. 1967b. *Sex chromosomes and sex-linked genes: monographs on endocrinology*, Vol. 1. Springer, Berlin/Heidelberg.
- Otto SP, Pannell JR, Peichel CL, Ashman T-L, Charlesworth D, Chippindale AK, Delph LF, Guerrero RF, Scarpino SV, McAllister BF. 2011. About PAR: the distinct evolutionary dynamics of the pseudoautosomal region. *Trends Genet* **27**: 358–367.
- Paria N, Raudsepp T, Pearks Wilkerson AJ, O'Brien PCM, Ferguson-Smith MA, Love CC, Arnold C, Rakestraw P, Murphy WJ, Chowdhary BP. 2011. A gene catalogue of the euchromatic male-specific region of the horse Y chromosome: comparison with human and other mammals. *PLoS One* **6**: e21374.
- Pérez-Pérez J, Barragán C. 1998. Isolation of four pig male-specific DNA fragments by RDA. *Anim Genet* **29**: 157–158.
- Platts AE, Dix DJ, Chemes HE, Thompson KE, Goodrich R, Rockett JC, Rawe VY, Quintana S, Diamond MP, Strader LF, et al. 2007. Success and failure in human spermatogenesis as revealed by teratozoospermic RNAs. *Hum Mol Genet* **16**: 763–773.
- Prokop JW, Underwood AC, Turner ME, Miller N, Pietrzak D, Scott S, Smith C, Millsted A. 2013. Analysis of Sry duplications on the *Rattus norvegicus* Y-chromosome. *BMC Genomics* **14**: 792.
- Quilter CR, Blott SC, Mileham AJ, Affara NA, Sargent CA, Griffin DK. 2002. A mapping and evolutionary study of porcine sex chromosome genes. *Mamm Genome* **13**: 588–594.
- Ross MT, Grafham DV, Coffey AJ, Scherer S, McLay K, Muzny D, Platzer M, Howell GR, Burrows C, Bird CP, et al. 2005. The DNA sequence of the human X chromosome. *Nature* **434**: 325–337.
- Rozen S, Skaletsky H, Marszalek JD, Minx PJ, Cordum HS, Waterston RH, Wilson RK, Page DC. 2003. Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* **423**: 873–876.
- Sandstedt SA, Tucker PK. 2004. Evolutionary strata on the mouse X chromosome correspond to strata on the human X chromosome. *Genome Res* **14**: 267–272.
- Seal RL, Gordon SM, Lush MJ, Wright MW, Bruford EA. 2011. genenames.org: the HGNC resources in 2011. *Nucleic Acids Res* **39**: D514–D519.
- Searle SMJ, Gilbert J, Iyer V, Clamp M. 2004. The otter annotation system. *Genome Res* **14**: 963–970.
- Skaletsky H, Kuroda-Kawaguchi T, Minx PJ, Cordum HS, Hillier L, Brown LG, Repping S, Pyntikova T, Ali J, Bieri T, et al. 2003. The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* **423**: 825–837.
- Skinner BM, Griffin DK. 2012. Intrachromosomal rearrangements in avian genome evolution: evidence for regions prone to breakpoints. *Heredity* **108**: 37–41.
- Skinner BM, Lachani K, Sargent CA, Affara NA. 2013. Regions of XY homology in the pig X chromosome and the boundary of the pseudoautosomal region. *BMC Genet* **14**: 3.
- Skinner BM, Lachani K, Sargent CA, Yang F, Ellis P, Hunt T, Fu B, Louzada S, Churcher C, Tyler-Smith C, et al. 2015. Expansion of the HSFY gene family in pig lineages. *BMC Genomics* **16**: 442.
- Smit A, Hubley R, Green P. 1996. *RepeatMasker Open-3.0*. <http://www.repeatmasker.org>.
- Soh YQS, Alföldi J, Pyntikova T, Brown LG, Graves T, Minx PJ, Fulton RS, Kremitzki C, Koutseva N, Mueller JL, et al. 2014. Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* **159**: 800–813.
- Stanyon R, Yang F, Cavaña P, O'Brien PC, Bagga M, Ferguson-Smith MA, Wienberg J. 1999. Reciprocal chromosome painting shows that genomic rearrangement between rat and mouse proceeds ten times faster than between humans and cats. *Cytogenet Cell Genet* **84**: 150–155.
- Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol Biol Evol* **28**: 2731–2739.
- Tarpey PS, Raymond FL, O'Meara S, Edkins S, Teague J, Butler A, Dicks E, Stevens C, Tofts C, Avis T, et al. 2007. Mutations in CUL4B, which encodes a ubiquitin E3 ligase subunit, cause an X-linked mental retardation syndrome associated with aggressive outbursts, seizures, relative macrocephaly, central obesity, hypogonadism, pes cavus, and tremor. *Am J Hum Genet* **80**: 345–352.
- Thauvin-Robinet C, Thomas S, Sinico M, Aral B, Burglen L, Gigot N, Dollfus H, Rossignol S, Raynaud M, Philippe C, et al. 2013. OFD1 mutations in males: phenotypic spectrum and ciliary basal body docking impairment. *Clin Genet* **84**: 86–90.
- Thomsen PD, Hindkjaer J, Christensen K. 1992. Assignment of a porcine male-specific DNA repeat to Y-chromosomal heterochromatin. *Cytogenet Cell Genet* **61**: 152–154.
- Turkish AR, Henneberry AL, Cromley D, Padamsee M, Oelkers P, Bazzi H, Christiano AM, Billheimer JT, Sturley SL. 2005. Identification of two novel human acyl-CoA wax alcohol acyltransferases: members of the diacylglycerol acyltransferase 2 (DGAT2) gene superfamily. *J Biol Chem* **280**: 14755–14764.
- Tyler-Smith C. 2008. An evolutionary perspective on Y-chromosomal variation and male infertility. *Int J Androl* **31**: 376–382.
- Vallender EJ, Lahn BT. 2004. How mammalian sex chromosomes acquired their peculiar gene content. *Bioessays* **26**: 159–169.
- Warburton PE, Giordano J, Cheung F, Gelfand Y, Benson G. 2004. Inverted repeat structure of the human genome: the X-chromosome contains a preponderance of large, highly homologous inverted repeats that contain testes genes. *Genome Res* **14**: 1861–1869.
- Xue Y, Tyler-Smith C. 2011. An exceptional gene: evolution of the TSPY gene family in humans and other great apes. *Genes* **2**: 36–47.
- Yang R-B, Fülle H-J, Garbers DL. 1996. Chromosomal localization and genomic organization of genes encoding guanylyl cyclase receptors expressed in olfactory sensory neurons and retina. *Genomics* **31**: 367–372.
- Yu C, Zhang Y-L, Pan W-W, Li X-M, Wang Z-W, Ge Z-J, Zhou J-J, Cang Y, Tong C, Sun Q-Y, et al. 2013. CRL4 complex regulates mammalian oocyte survival and reprogramming by activation of TET proteins. *Science* **342**: 1518–1521.
- Zhang Q, Su B. 2014. Evolutionary origin and human-specific expansion of a cancer/testis antigen gene family. *Mol Biol Evol* **31**: 2365–2375.

Received December 19, 2014; accepted in revised form November 9, 2015.



The pig X and Y Chromosomes: structure, sequence, and evolution

Benjamin M. Skinner, Carole A. Sargent, Carol Churcher, et al.

Genome Res. 2016 26: 130-139 originally published online November 11, 2015

Access the most recent version at doi:[10.1101/gr.188839.114](https://doi.org/10.1101/gr.188839.114)

Supplemental Material <http://genome.cshlp.org/content/suppl/2015/12/08/gr.188839.114.DC1>

References This article cites 68 articles, 15 of which can be accessed free at:
<http://genome.cshlp.org/content/26/1/130.full.html#ref-list-1>

Open Access Freely available online through the *Genome Research* Open Access option.

Creative Commons License This article, published in *Genome Research*, is available under a Creative Commons License (Attribution 4.0 International), as described at <http://creativecommons.org/licenses/by/4.0/>.

Email Alerting Service Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or [click here](#).

A banner advertisement for a webinar. The background is dark blue with a grid pattern. On the left, the word "Webinar" is written in white. In the center, the text "Automation-friendly full-length scRNA-seq" is written in white. On the right, there is a green circular logo with the text "that's GOOD science" and the Takara logo, which includes the text "Takara" and "Genetech Takara celisartis" below it.

To subscribe to *Genome Research* go to:
<http://genome.cshlp.org/subscriptions>
