

## Review

## Y and W Chromosome Assemblies: Approaches and Discoveries

Marta Tomaszekiewicz,<sup>1</sup> Paul Medvedev,<sup>2,3</sup> and  
Kateryna D. Makova<sup>1,3,\*</sup>

Hundreds of vertebrate genomes have been sequenced and assembled to date. However, most sequencing projects have ignored the sex chromosomes unique to the heterogametic sex – Y and W – that are known as sex-limited chromosomes (SLCs). Indeed, haploid and repetitive Y chromosomes in species with male heterogamety (XY), and W chromosomes in species with female heterogamety (ZW), are difficult to sequence and assemble. Nevertheless, obtaining their sequences is important for understanding the intricacies of vertebrate genome function and evolution. Recent progress has been made towards the adaptation of next-generation sequencing (NGS) techniques to deciphering SLC sequences. We review here currently available methodology and results with regard to SLC sequencing and assembly. We focus on vertebrates, but bring in some examples from other taxa.

### Why Should We Study SLCs?

Sex chromosomes are present in the majority of vertebrate genomes. Most mammals have male heterogamety (XY males and XX females), and all studied birds have female heterogamety (ZW females and ZZ males). Reptile, amphibian and fish species with genetic (and not environmental) sex determination can have either male or female heterogamety (Table 1). Nevertheless, most vertebrate genome projects have sequenced and assembled homogametic genomes (XX males or ZZ females). For instance, among 98 mammalian species sequenced [1], only 18 are complemented by Y chromosome assemblies (Figure 1, Key Figure). The challenge of assembling the **sex-limited chromosome** (SLC, see Glossary) is due to its haploid and highly repetitive nature (Boxes 1 and 2). The resulting paucity of SLC sequences impedes the development of genetic markers used to study sex-specific dispersal and population genetics, and to evaluate sex bias during hybridization in natural populations [2,3]. More generally, the lack of SLC chromosome sequences prevents our complete understanding of vertebrate genome function and evolution.

Y chromosome fulfills a myriad of important functions that have been understudied outside of model organisms. Present only in males, the Y chromosome (Box 1) is passed uniquely along the paternal lineage. The Y chromosome harbors *SRY*, the **master sex-determining gene** (MSDG) in most mammals, and genes expressed in the testis that are important for spermatogenesis [4]. Human Y chromosome genetic variants are associated with spermatogenic failure and male infertility [5], germline and somatic tumors [6], hearing impairment [7], and coronary artery disease [8]. In fruit flies (*Drosophila melanogaster*), genetic variation on the Y affects the expression of hundreds of X chromosome and autosomal genes [9,10]. The W chromosome (Box 1) is present exclusively in females and, for instance in birds, does not carry the MSDG.

### Trends

NGS technology and advances in assembly algorithms allow researchers to decipher chromosomes Y and W, the most challenging parts of vertebrate genomes to sequence and assemble.

The rapidly accumulating data accelerate discoveries related to the evolution and structure of these chromosomes, suggesting that they are much more varied among vertebrate species than was previously realized.

Y and W chromosome data are being used to trace the divergence of male- and female-specific lineages. Novel technology makes it possible to generate population-scale sequencing data for these chromosomes and to study their diversity.

<sup>1</sup>Department of Biology, Pennsylvania State University, University Park, PA 16802, USA

<sup>2</sup>Department of Computer Science and Engineering, and Department of Biochemistry and Molecular Biology, Pennsylvania State University, University Park, PA 16802, USA

<sup>3</sup>Genome Sciences Institute and Center for Medical Genomics, Huck Institutes of the Life Sciences, Pennsylvania State University, University Park, PA 16802, USA

\*Correspondence: [kdm16@psu.edu](mailto:kdm16@psu.edu) (K.D. Makova).

Table 1. Sex Chromosome Systems in Vertebrates

Sex chromosome systems	Taxa	Refs
♀XX/♂XY (male heterogamety)	Most mammals, some lizards, some turtles, some amphibians, some Rajiformes, Osteoglossiformes, Clupeiformes, Stomiiformes, Myctophiformes, Gobiesociformes fish	[11,93,95]
♀ZW/♂ZZ (female heterogamety)	Birds, snakes, some lizards, some turtles, some amphibians, some Esociformes, Aulopiformes fish	[11,95]
♀XX, ♀WX, ♀WY/♂XY, ♂YY (male and female heterogamety coexisting in the same species)	Platyfish ( <i>Xiphophorus maculatus</i> ), some frogs ( <i>Xenopus tropicalis</i> , <i>Rana rugosa</i> )	[96,97]
♀XO/♂XY	Creeping vole ( <i>Microtus oregoni</i> )	[98]
♀XX/♂XX	Mole voles ( <i>Ellobius tancrei</i> , <i>E. talpinus</i> , <i>E. alaicus</i> )	[68]
♀XO/♂XO	Mole vole ( <i>Ellobius lutescens</i> ), spiny rats ( <i>Tokudaia osimensis</i> , <i>T. tokunoshimensis</i> ), pygmy mouse ( <i>Mus triton</i> )	[64,68,99]
♀XX, ♀XY/♂XY	Wood lemming ( <i>Myopus schistocolor</i> ), collared lemming ( <i>Dicrostonyx torquatus</i> ), South American grass mice ( <i>Akodon</i> spp.), African pygmy mouse ( <i>Mus minutoides</i> )	[100–102]
♀OW/♂OO	Hochstetter's frog ( <i>Leiopelma hochstetteri</i> )	[103]
♀X <sub>1</sub> X <sub>1</sub> X <sub>2</sub> X <sub>2</sub> /♂X <sub>1</sub> X <sub>2</sub> Y	35 species of fish, howler monkeys ( <i>Alouatta</i> spp.), Goeldi's marmoset ( <i>Callimico goeldii</i> ) <sup>a</sup> , Peruvian owl monkey ( <i>Aotus</i> spp.) <sup>a</sup> , leaf litter frog ( <i>Eleutherodactylus maussii</i> ), some lizards	[104–107]
♀X <sub>1</sub> X <sub>1</sub> /♂X <sub>1</sub> Y <sub>1</sub> Y <sub>2</sub>	Fish: <i>Hoplias malabaricus</i> , <i>Ancistrus</i> sp.1, <i>Harttia carvalhoi</i>	[104]
♀X <sub>1</sub> X <sub>1</sub> X <sub>2</sub> X <sub>2</sub> /♂X <sub>1</sub> X <sub>2</sub> Y <sub>1</sub> Y <sub>2</sub>	Some monkeys	[107]
♀Z <sub>1</sub> Z <sub>2</sub> W <sub>1</sub> W <sub>2</sub> /♂Z <sub>1</sub> Z <sub>1</sub> Z <sub>2</sub> Z <sub>2</sub>	Fish: <i>Ancistrus</i> sp.2	[104]
♀ZW <sub>1</sub> W <sub>2</sub> /♂ZZ	Fish: <i>Trachinocephalus myops</i> , <i>Apareiodon affinis</i>	[104]
♀ X <sub>1</sub> X <sub>1</sub> X <sub>2</sub> X <sub>2</sub> X <sub>3</sub> X <sub>3</sub> X <sub>4</sub> X <sub>4</sub> X <sub>5</sub> X <sub>5</sub> /♂ X <sub>1</sub> Y <sub>1</sub> X <sub>2</sub> Y <sub>2</sub> X <sub>3</sub> Y <sub>3</sub> X <sub>4</sub> Y <sub>4</sub> X <sub>5</sub> Y <sub>5</sub>	Platypus ( <i>Ornithorhynchus anatinus</i> ), echidna ( <i>Tachyglossus aculeatus</i> )	[108,109]

<sup>a</sup>Previously named ♀XX/♂XO.

Instead, the best-investigated MSDG candidate (*DMRT1*) is located on the Z in all studied birds, and thus is present in both sexes. Note that the evidence surrounding the role of *DMRT1* in sex determination has been controversial (reviewed in [11]). Most of the W chromosome genes in chicken (*Gallus gallus*) have homologs on the Z and, except for the multicopy *HINTW* [12], have been thought to evolve under relaxed selective constraints [13]. Little is known about W chromosomes outside birds.

SLCs differ in their mutation rates and evolutionary patterns from the rest of the genome. Because many vertebrate males undergo a higher number of germline cell divisions than females, the former are expected to have a higher mutation rate, in other words a male mutation bias. As a result, in species with male (female) heterogamety, more mutations accumulate on the Y (Z) than on the X (W) chromosome and autosomes [14]. SLC sequences for a wide taxonomic range will enable investigations of how male mutation bias is affected by life-history traits [15,16]. SLC sequences should also allow us to study selection that depends on the heterogamety system. Both the Y and W chromosomes have smaller effective population sizes than autosomes and X/Z chromosomes, but are **hemizygous** outside pseudoautosomal regions (PARs, Box 1), leading to a unique selective regime. Furthermore, both Y and W do not recombine other than within PARs, and the whole chromosome becomes a unit of selection.

## Glossary

**Contig:** a long contiguous segment that belongs to the genome, as produced by an assembler.

**Genome-wide Association Study (GWAS):** an investigation of the relationship between genetic variants across genomes of different individuals and a particular phenotypic trait (e.g., association between genetic variants and a particular disease).

**Hemizygous:** a chromosome or chromosomal regions without a homologous pair.

**k-mer:** a short nucleotide sequence of length k.

**Long reads:** reads produced by the Pacific Biosciences (PacBio) technology. These reads are typically 1000–10 000 bp in length, but are significantly more expensive to obtain than short reads.

**Master sex-determining gene**

**(MSDG):** the major trigger of gonadal differentiation leading to ultimate testicular or ovarian development.

**Non-recombining region on the Y or W (NRY or NRW):** consists of the total chromosome length with pseudoautosomal regions excluded.

**Read:** a short fragment of DNA whose sequence has been determined.

**Read depth of a genomic**

**segment:** the average number of reads covering a nucleotide site in a genomic segment.

**Scaffold:** an ordered and oriented collection of contigs. Scaffolds are built from contigs by incorporating mate-pair read information. Scaffolds routinely have large gaps with undetermined sequence.

**Sex-limited chromosome (SLC):**

sex chromosome (Y or W) that is unique to a heterogametic genome (XY male or ZW female).

**Sexual selection:** a form of reproductive competition within a species according to which the individuals of one sex choose members of another sex to mate. Depending on mating pattern, sexual selection can be more pronounced before or after copulation (mate choice or sperm competition, respectively).

**Short reads:** reads produced by Illumina or Ion Torrent technology, typically 100–300 bp in length.

**Transposable elements (TEs):** repeated DNAs of different lengths

## Box 1. The Structure and Genes of SLCs

## Primate Y Chromosome Structure

The primate Y chromosome is composed of pseudoautosomal region(s) (PAR(s), which recombine between Y and X chromosomes, and non-recombining X-degenerate, ampliconic, X-transposed, and heterochromatic regions [25]. The X-degenerate regions are composed of segments with different levels of sequence identity to X chromosome – in other words the evolutionary strata – reflecting the timing of Y-chromosome inversions [110]. The ampliconic regions consist of repetitive sequences that have >50% identity to each other [25] and contain palindromes – inverted repeats up to several Mb in length whose arms are >99.9% identical to each other [111]. The X-transposed region, detected so far only on the human Y, was transposed from the X to the Y after the human–chimpanzee split [25]. The heterochromatic regions have a high content of satellite repeats organized in arrays [25] that can be interrupted with TEs [112].

## Y Chromosome Genes

Most Y chromosome protein-coding genes are located in either the X-degenerate or ampliconic regions. X-degenerate genes (16 on the human Y) are single-copy, ubiquitously expressed housekeeping genes that have homologs on the X [4]. Ampliconic genes, by contrast, have a complex organization into multigene families (nine families on the human Y) located on palindromes, with the exception of the *TSPY* gene family that is located outside of palindromes (Figure 2A). Many of them are crucial for reproduction because they are expressed only in testis and encode proteins functioning in spermatogenesis. Gene copies whose location is mirrored across the homologous arms of a palindrome are almost identical, while gene copies present on different arms or on different palindromes have lower identity [113]. Owing to this high but variable degree of homology, ampliconic gene sequences and copy number (Figure 2B) have been difficult to assay, and their evolution has been understudied. It has been hypothesized that abundant gene conversion between the palindrome arms can maintain the functionality of Y-chromosomal ampliconic gene duplicates, and thus prevents them from genetic decay [111].

## Chromosome W

In most respects, the organization of chromosome W resembles that of chromosome Y. Similarly to the Y chromosome, the W chromosome is typically relatively small (~7 Mb in flycatcher [17]) and is packed with repeats, including heterochromatin [114]. For instance, ~75% of the chicken W is occupied by two heterochromatin satellite families [115]. Similarly to Y, the W chromosome carries genes homologous to genes on its counterpart, chromosome Z [17]. In contrast to Y, the W chromosome usually does not harbor many ampliconic genes. Indeed, for instance, only one multicopy gene (*HINTW*) was found on the chicken and flycatcher W chromosomes [17,116,117].

that are able to move using different mobilization mechanisms within the genome.

It has been hypothesized that Y chromosome can be directly affected by **sexual selection**, which is more common in males than in females, whereas this is not expected for W chromosome [11]. Moreover, W and Y chromosomes might be affected by selective pressures unique to oogenesis and spermatogenesis, respectively [17]. These hypotheses can be tested directly once Y and W chromosome sequences become available.

Sexually antagonistic selection is especially strong on evolving sex chromosomes, when an allele increases fitness in one sex, but decreases fitness in the other sex, resulting in the suppression of recombination. This process will eventually lead to Y or W chromosome degeneration due to silencing, and deletion of the genes that are not beneficial for male or female fitness, respectively [18–20]. In addition, sexually antagonistic selection is considered to be one of the major forces responsible for sex determination transitions between chromosome pairs and thus between male and female heterogamety [21], and has been linked to dosage compensation [22].

The Y chromosome has undergone rapid loss of functional genes in most mammals, and this process favors the evolution of global dosage compensation [23]. While the gene density on the human X is approximately 7 genes/Mb [24], it is approximately 1 gene/Mb on the human Y [25]. In birds, W chromosome degeneration is not as pronounced as in mammals, and thus dosage compensation is local, affecting only a few genes [22,26]. Indeed, the gene density on the Z chromosome of flycatcher (*Ficedula albicollis*) is almost 11 genes/Mb, whereas it is approximately 7 genes/Mb on its W chromosome [17,27]. An intermediate pattern of dosage compensation is present in the platypus (*Ornithorhynchus anatinus*), where 50–75% of its genes are compensated [28].

Key Figure

Y and W Chromosomes Sequenced at the Whole-Chromosome Level or Assembled from Whole-Genome Sequences

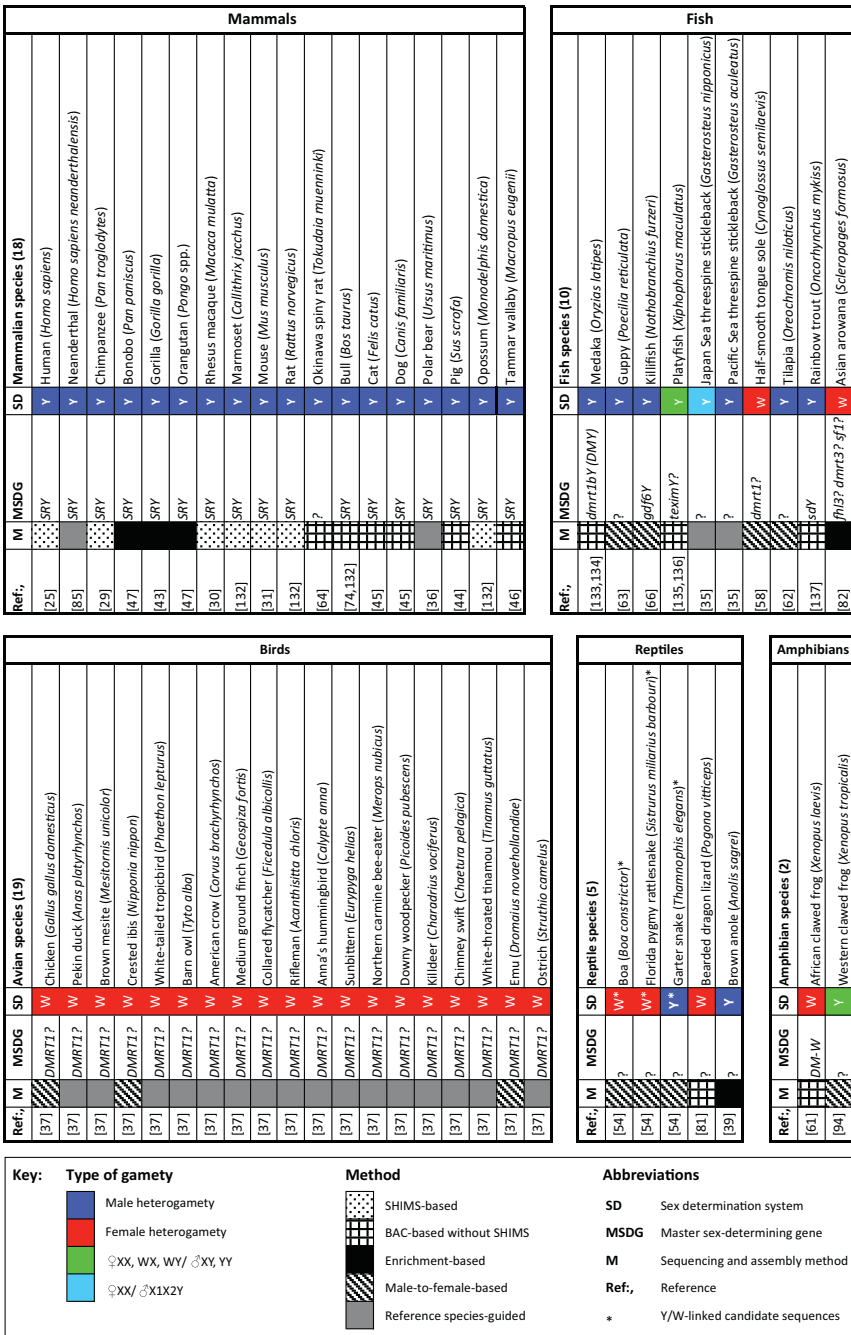


Figure 1. Details from [25,29–31,35–37,39,43–47,54,58,61–64,66,74,81,82,85,94,132–137]. Abbreviations: BAC, bacterial artificial chromosome; SHIMS, single-haplotype iterative mapping and sequencing

## Box 2. Why Are Chromosomes Y and W So Difficult to Sequence and Assemble?

### Assembly Basics

Assembly is a computational technique that uses the overlaps between the sequenced reads to chain them together into longer sequences called contigs. When other proximity data are available (e.g., mate-pairs [118], optical maps [119], Hi-C [120]), the assembler can further group the contigs into scaffolds in which each scaffold contains an ordering and relative orientation of the contigs. Assembly quality is typically measured in terms of completeness (i.e., how much of the genome is included in contigs/scaffolds), contiguity (i.e., how long are the contigs/scaffolds), and accuracy (how many errors do the contigs/scaffolds contain). Long-read technology (i.e., Pacific Biosciences, PacBio) produces reads up to dozens of kb in length. It can improve assembly contiguity, while short-read technology (i.e., Illumina) can improve its accuracy. At the moment, long-read technology is significantly more expensive than short-read technology. For more information we refer the reader to reviews on assembly (e.g., [121]).

### Difficulties in SLC Sequencing and Assembly

(i) *Haploid Nature of SLCs.* Most genome projects have focused on sequencing homogametic genomes (i.e., XX females or ZZ males), leaving SLCs undeciphered. Sequencing a homogametic genome results in a relatively robust assembly of autosomes and the X (or Z) chromosome, while sequencing a heterogametic genome produces a suboptimal assembly of both sex chromosomes (X and Y, or Z and W) because in a heterogametic genome they are haploid and thus have low sequencing depth. To assemble an SLC from whole-genome sequencing reads, one needs to sequence a heterogametic genome to at least twice the depth of a homogametic genome, which at least doubles the project cost.

(ii) *Ampliconic Regions.* Rich in multicopy gene families, these regions can be very abundant – as in human and mouse. The main difficulty is in assembling these regions; however, this is currently possible only using assembled BACs, provided that only a single haplotype is present in the sequencing library (i.e., using SHIMS) [25].

(iii) *Heterochromatin.* The Y chromosome has an unusual structure (Figure 2), frequently with more than half of its sequence covered by heterochromatin (Box 1). Even though the generation of sequencing reads originating from heterochromatin is possible, the main problem is to accurately assemble them. Heterochromatin repeat units have high identity to each other, making it difficult to assemble even when using BACs, either because of high inefficiency of cloning or ambiguous assembly. Both types of repeats (ampliconic regions and heterochromatin) pose challenges for any assembler [121].

## Present-Day Approaches for SLC Sequencing and Assembly

Historically, methods used for sequencing SLCs were based on the use of bacterial artificial chromosomes (BACs) and Sanger sequencing. One such technique – single-haplotype iterative mapping and sequencing (SHIMS) – was used to sequence human, chimpanzee, rhesus macaque, and mouse Y chromosomes [25,29–31]. However, this highly accurate BAC-based approach is both time-consuming and expensive.

Recent technological advances have made sequencing of SLCs faster and cheaper. We review below current approaches used to sequence and assemble SLCs on a chromosome-wide scale using NGS techniques. Sequencing of SLCs on a smaller scale (e.g., RAD-seq [32]) and sequencing of SLC transcriptomes (e.g., [33,34]) might be useful in some situations, for example for deriving sex-specific markers, but are not the subject of this review. Chromosome-level SLC characterization requires sequencing combined with pre- and post-processing, and results in an assembly (Box 2) that is usually performed with generic software. We divide these approaches into three groups, each of which we now describe in detail. The reader is directed to Figure 1 for examples of the species in which each approach has been used.

### Reference Species-Guided Approaches

When the resources for sequencing are restricted, some limited approaches are possible. If low-coverage **reads** from a heterogametic genome of the species of interest are available, they can be aligned to the genome of a closely related species. Reads aligning to the SLC of a closely related species with a significant **read depth** are candidate SLC sequences in the sequenced

species [35]. In a different approach, to identify putative SLC sequences, known SLC sequences (e.g., genes) from a closely related species can be aligned to the heterogametic genome assembly of a species of interest [36]. When the only assembly of a closely related species is from a homogametic sample, some SLC-linked **contigs** can be identified by their distant homology to the non-SLC sex chromosome of the related species [37]. These procedures are limited to detecting homologous sequences and usually cannot locate rearrangements, duplications, or highly diverged sequence. Another shortcoming of these methods is in that they result in sequences that are only putatively SLC-specific, and thus additional validations are necessary to confirm that such sequences belong to SLCs.

#### Enrichment-Based Approaches

The following enrichment-based approaches increase the content of SLC DNA in the sample before sequencing, dramatically reducing sequencing costs.

##### Chromosome Microdissection

This is a microscope-assisted excision of the chromosome of interest from a metaphase spread of one or more cells using a microneedle or a laser microbeam [38]. The microdissected chromosomes are amplified, optionally cloned, and finally sequenced. Microdissection is fairly precise, avoiding much contamination from other chromosomes. However, it has relatively low processivity and requires amplification before sequencing. Amplification can introduce undesirable artifacts that can complicate downstream assembly. Microdissection has been applied to SLC sequencing and assembly in brown anole lizard (*Anolis sagrei*) [39] and the flour moth (*Ephesia kuehniella*) [40].

##### Chromosome Flow-Sorting

This technique separates chromosomes based on their size and GC content by applying a laser charge to a chromosomal suspension [41]. In an ideal scenario, each chromosome forms a unique peak on a fluorescence histogram and can be collected; however, chromosomes that have a similar size and GC content may be sorted together (e.g., [42]). In addition, small chromosomes frequently flow-sort with debris of larger chromosomes, decreasing the efficiency of enrichment [43]. The advantage of flow-sorting is its high levels of automation and throughput. Flow-sorting was recently used to enrich for the Y chromosome of gorilla (*Gorilla gorilla*) [43] and the sex chromosomes of pig (*Sus scrofa*) [44] before assembly, to make probes to identify the Y chromosome BACs of cat (*Felis catus*), dog (*Canis familiaris*) [45], and tammar wallaby (*Macropus eugenii*) [46], to validate the SHIMS-based Y assembly of chimpanzee (*Pan troglodytes*) [29], and to characterize the microchromosomes in *Anolis* lizards [39].

##### Targeted Capture

This is an enrichment approach that uses sequences of an SLC from a closely related species, or partial SLC sequences from the species of interest, as a probe. This approach has been implemented in a recent study of the 2.6–3.8 Mb fragment of the great ape Y chromosomes [47], using human Y sequences as a probe. While its high processivity allows the analysis of a large number of samples, targeted capture based on a probe derived from another species suffers from drawbacks similar to the reference species-guided approach. A modification of targeted capture is to identify BACs specific to an SLC and sequence them with NGS, as was done before assembling the Y chromosomes of dog and cat [45], and the neo-Y of the Okinawa spiny rat (*Tokudaia muenninki*) [48]. This method is laborious, expensive, and might not resolve palindromes (Box 1) in such detail as SHIMS [25], but it is expected to result in a more contiguous assembly than whole-genome NGS.

##### Computational Methods Used in Conjunction with Enrichment Approaches.

Because none of the enrichment approaches are 100% efficient, computational approaches are needed to remove the non-SLC reads before assembly. One approach is to use the reads



aligning to the SLC of a related species; however, this suffers from reference bias and might mistakenly pick up X-linked or autosomal genes [36]. Alternatively, one can remove reads that align to the homogametic genome of the same species. However, assemblies created using this ‘subtraction-based’ approach can only include sequences that are specific to the SLC, leaving out regions that are shared by SLC with non-SLCs [49]. For instance, the human Y chromosome shares some regions with the X – for example PARs – and with autosomes – such as the *DAZ* gene region that is homologous between chromosomes 3 and Y [50]. **Transposable elements** (TEs), which are shared by all chromosomes, will also be excluded from the assembly, and this will lead to overly fragmented assemblies.

An alternative approach, called RecoverY, was recently developed to identify SLC-specific **short reads** from Illumina sequencing of an SLC-enriched flow-sorted dataset. RecoverY selects high-frequency **k-mers** as originating from the SLC because SLC-linked *k*-mers should have much higher frequencies than *k*-mers not from the SLC. RecoverY was applied to sequenced gorilla Y flow-sorted material before assembly [43].

#### Male-to-Female Comparison Approaches

Enrichment-based techniques, while powerful, add an additional level of complexity. A more straightforward approach is to sequence and assemble a heterogametic genome and then to identify which **scaffolds** (Box 2) derive from the SLC. Sometimes SLC sequences come as a byproduct of whole-genome sequencing projects, and these can generate a high-quality heterogametic genome assembly. In such cases, the only task is to identify SLC-linked scaffolds (e.g., for *Drosophila* [51]). In the more common case, when a homogametic genome and reads from a heterogametic genome from the same species are available, one can use the subtraction-based approach described above [49].

Because SLC sequences have different copy counts in males versus females, we can use read depth to assist in predicting SLC scaffolds. After assembling the reads from a heterogametic genome, we can align them back to the generated assembly. Then, the read depth of an SLC scaffold should be about half of that for an autosomal scaffold. The read depth alone, however, is not sufficient to differentiate between scaffolds belonging to an SLC (e.g., the Y) versus a sex chromosome present in both sexes (e.g., the X) because there exists high variation in read depth in diploid and haploid regions for most sequence projects (e.g., 50×). Alternatively, one can additionally sequence a homogametic genome and align its reads to the assembly generated from a heterogametic genome. The Y (YGS) chromosome Genome Scan method [51] is based on the idea that SLC scaffolds should have very low read depth in this case. It was applied to human (*Homo sapiens*), fruit fly (*Drosophila melanogaster*) [51], and the kissing bug (*Rhodnius prolixus*) [52].

A more advanced approach measures the depth for both female reads and male reads, separately, and takes the ratio of the two. For Y, autosomal, and X scaffolds, for instance, the ratio should be close to 0, 1, and 2, respectively. This ‘chromosome quotient (CQ) method’ [53] has been applied to numerous species [17,36,37,49,54–56]. Note that, unlike the subtraction-based method, this method is expected to retain many regions shared between SLC and autosomes because they will have a different read depth in male versus female genomes. However, as in the subtraction method above, the chromosome quotient method would not retain high-copy TEs and might classify PARs as autosomal sequences.

In contrast to enrichment-based approaches, male-to-female comparison approaches are widely accessible but require the more-costly sequencing of two (male and female) genomes instead of one. As a result, they have been predominantly used with the cheaper short-read sequencing data, which limits assembly contiguity (Box 2). Indeed, for large genomes, such as

that of gorilla, using **long reads** (e.g., from Pacific Biosciences, PacBio) would add an order of magnitude more costs (Table S17 in [43]), but leaving it out would decrease contiguity substantially [43]. As long-read sequencing costs drop, male-to-female comparison approaches will start to give more contiguous assemblies, but, for the time being, enrichment-based approaches are more affordable when high assembly contiguity is desired. The limitations and future developments of these methods are discussed in Box 3.

Species with old heteromorphic sex chromosomes – for example mammals, whose sex chromosomes are as old as 166–190 million years (Myr) [57], snakes with ~140 Myr-old sex chromosomes [54] or birds whose Z and W are ~100 Myr old [37] – can be studied using all the presented methods: reference species-guided approaches, enrichment-based techniques, and male-to-female comparison approaches. The enrichment-based techniques are especially recommended here because the sex chromosomes are easily distinguishable from autosomes for flow-sorting or chromosome microdissection purposes. Young and/or still homomorphic sex chromosomes in some taxa – for example sex chromosomes in fish (30 Myr old or younger [58]) – cannot be identified with reference species-guided, enrichment-based, or depth-based male-to-female comparison methods. In such cases, differences in SNP density between the two young sex chromosomes and fixation index ( $F_{ST}$ )-based methods can be used to identify sex chromosomes. Such methods were used, for instance, to detect the neo-Y region in Japan stickleback [59] and to study a sex-chromosome-like supergene in the white-throated sparrow, *Zonotrichia albicollis* [60].

#### Box 3. Limitations of Current Methods, Ways To Alleviate Them, and Future Methodology Development

The major limitation of all NGS-based approaches (in contrast to SHIMS) is that they suffer from the collapse of duplicated or repetitive sequences, as well as other potential inaccuracies [122]. On SLCs in particular, NGS is unable to completely resolve the highly repetitive and frequently palindromic nature of sex chromosomes. Even long PacBio reads are still substantially shorter than the arms of many palindromes. The NGS-based assemblies frequently collapse arms of the same palindrome [43], precluding an accurate identification of the copy number for ampliconic genes located within palindromes from the assemblies. This calls for novel methodological developments.

##### Experimental Methods

First, novel methods should be developed to resolve heterochromatin array sequence and length given the role of Y-linked heterochromatin in regulating the functions of the whole genome and because of intraspecific variation in heterochromatin abundance [9,10,123,124]. PacBio reads have a potential to inform the tandem arrangements of heterochromatic repeats on SLCs [55], but are unlikely to span the lengths of long arrays [125]. Oxford Nanopore technology – generating sequencing reads of an average length of 5.5–10 kb [126], with 98 kb being the longest recently reported [127] – is another method to consider.

It will be important that the error rates of these technologies (~16% for PacBio [128] and up to 40% for Oxford Nanopore [127]) continue to decrease because high error rates can confound repeat and amplicon analysis. Similarly, methods should be developed to sequence through and assemble palindromes on SLCs. Methods operating at larger genomic scales – for example optical mapping of DNA labeled at sites recognized by a restriction enzyme [119] or by a nicking endonuclease [129] – might be able to fulfill this purpose.

##### Computational Methods

Computational methods used to assemble SLCs should be developed further as well. Current approaches only take advantage of a few potential data sources. For example, a method could potentially combine enriched male reads, male whole-genome reads, female whole-genome reads, and the female assembly. More sophisticated methods for assembling Y chromosomes from enrichment, subtraction, or differential read depth approaches are also possible but have not received sufficient attention. In particular, current methods use pre- and post-processing to take advantage of generic, out-of-the-box assemblers. However, such assemblers are oblivious to the nature of the data, and a smarter assembler that is designed specifically to assemble SLCs has the potential to significantly improve such assemblies. Methods for computational evaluation of ampliconic gene copy number from read depth should be further developed as well [71].



## Novel Insights from Recent Sequencing of SLCs

The reader is directed to [Figure 1](#) for the methods with which the sex chromosomes mentioned in this section were identified and sequenced.

### Evolution of SLCs

#### Early Stages of SLC Evolution

Recent NGS-based projects have demonstrated that, in some species, sex chromosomes have not yet differentiated morphologically but have undergone some genetic divergence. This is the case for the African clawed frog (*Xenopus laevis*), tilapia (*Oreochromis niloticus*), and guppies (*Poecilia reticulata*) [35,61–63]. In tilapia, a proto-Y region was identified because of its sex-biased gene expression and high divergence from the homologous region on the X [62]. NGS techniques were also used to characterize the early stages of sex chromosome evolution in the Okinawa spiny rat (XO/XO) in which neo-sex chromosomes are so young that they have had no time to diverge even at the cytogenetic level [64] – a rare situation for mammals, in which most sex chromosomes are old. In this species, as well as in the Japan Sea threespine stickleback (*Gasterosteus nipponicus*) [35], neo-sex chromosomes evolved from a fusion of ancestral sex chromosomes to autosomes. Moreover, the Y chromosome of the brown anole (*Anolis sagrei*) also originated from fusions of the ancestral sex chromosomes with autosomes [39], confirming that this is a recurring mechanism of sex chromosome formation in a variety of unrelated species [65].

Initial stages of X–Y chromosome divergence were also deciphered in killifish (*Nothobranchius furzeri*) after its male and female genomes were sequenced [66]. Interestingly, the length of the **non-recombining region on the Y** (NRY) varies by orders of magnitude among killifish strains, suggesting that multiple Y chromosomes can coexist early on [66]. Population-level variation was also observed in the youngest sex chromosome stratum ([Box 1](#)) in guppies, in which sexual selection for color in males drives the expansion of the NRY and leads to a higher Y–X divergence [63].

#### Mutational Rates on Differentiated SLCs

The sheer number of rapidly accumulating SLC sequences enables a detailed analysis of their evolutionary rates. For instance, in donkey–horse whole-genome alignments, the divergence of the Y and the X was higher and lower, respectively, than of the autosomes [67]. W-linked genes have lower divergence than Z-linked sequences in the collared flycatcher [17]. These examples are consistent with male mutation bias [14].

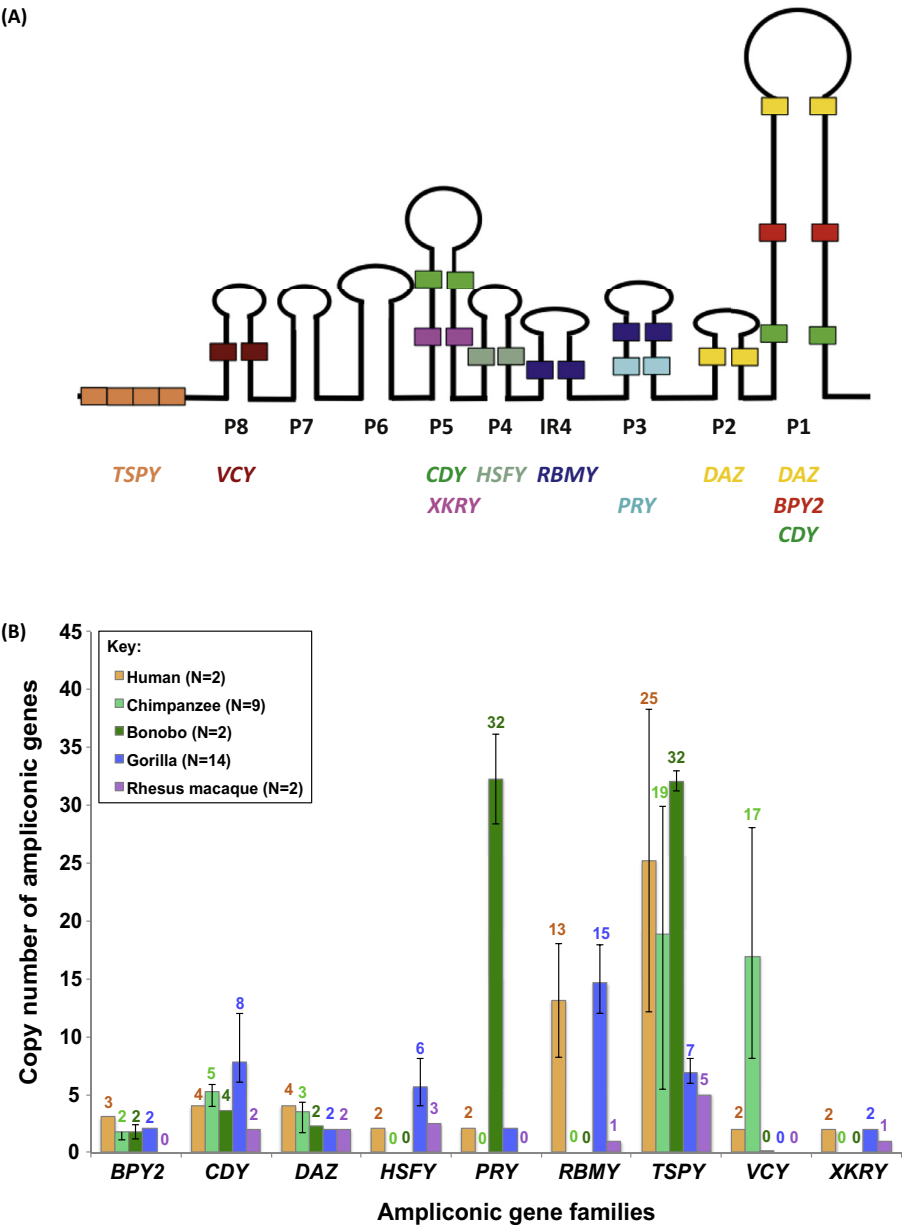
### Sex Chromosome Loss

New technology also provides glimpses into sex chromosome loss. One notable example is the recently sequenced Transcaucasian mole voles, *Ellobius lutescens* (XO/XO) and *E. talpinus* (XX/XX), which lost their Y chromosomes independently from each other [68]. It was previously shown that *Sry* was lost from their Y chromosomes [69], and now this finding was confirmed by sequencing [68]. Sequencing also identified that the homologs of their previously Y-linked genes, including *Zfy*, *Eif2s3y*, *Usp9y*, and *Ssty*, are currently present on autosomes [68]. In fact, one of these Y-linked genes, along with *Sry*, is required and sufficient for successful male meiosis in rodents [70]. The genome assemblies provide a fertile ground for finding the MSDG(s) in these *Ellobius* species.

### Genes on SLCs

#### Ampliconic Genes

NGS data and specialized computational or experimental techniques revealed substantial intra- and interspecific variation in ampliconic gene copy number in humans, other great apes, and monkeys ([Figure 2B](#); also see [Box 3](#)) [43,71–73]. Because of the role of these genes in



Trends in Genetics

**Figure 2. Ampliconic Genes.** (A) Palindromes (P1–P8) and ampliconic genes on the human Y chromosome (based on the data from [25]). Ampliconic genes belonging to one family are shown with the same color. The *TSPY* family is the only ampliconic gene family located outside of palindromes. (B) Intra- and interspecific variation in ampliconic gene copy number among ape males. Ampliconic gene copy number is challenging to obtain from an assembly because of the high sequence similarity of palindrome arms (Box 3), but can be determined experimentally with droplet digital PCR [130] or quantitative PCR, or computationally using a *k*-mer based approach [71] or based on read depth [131]. Below we summarize published data for human [43], chimpanzee [71], bonobo [71], gorilla [43], and rhesus macaque [30,72]. For ampliconic gene families with intraspecific size variation, size averages (numbers above bars) and ranges (error bars) are shown

reproduction, such variability might be under sex-specific selection. In papionin monkeys, ampliconic gene families evolve faster than autosomal multigene families [72].

Extreme cases of ampliconic gene expansion were identified on the bull (*Bos taurus*) and mouse (*Mus musculus*) Y chromosomes using BAC-based sequencing [31,74]. Largely because of expansion of six ampliconic gene families, the bull Y contains 1274 genes in its NRY, making it the most gene-rich chromosome both in the bull genome and among other known Y chromosomes (compared with 78 protein-coding genes present on the human Y [25]). Three bull Y ampliconic gene families transposed from autosomes, and the other three have homologs on the X [74]. Similarly, the recently sequenced mouse NRY contains as many as 700 protein-coding genes, with all but 45 belonging to three recently acquired ampliconic gene families that have homologs on the X [31]. Note that some of the genes on bull and mouse Y chromosomes might be pseudogenes. Nevertheless, the insights from these SLCs put into question the usual narrative of 'gene-poor' mammalian Y chromosomes.

### W Chromosome Genes

All 46 genes on the recently deciphered flycatcher non-recombining region on the W (NRW) have homologs on the Z [17]. This pattern is in contrast to that of mammalian Y chromosome that acquired several genes from autosomes (e.g., *DAZ* [50]). The flycatcher putative W-linked genes are not enriched in functions related to female reproduction [17]. On this chromosome, selection likely acted to retain genes whose expression could complement the expression of their Z chromosome homologs in a dosage-sensitive manner [17]. An exception is the multi-copy *HINTW* that evolves under positive selection and is expressed in female gonads before sexual differentiation [12]. Interestingly, chicken (*Gallus gallus domesticus*), duck (*Anas platyrhynchos*), and turkey (*Meleagris gallopavo*) W chromosome genes with homologs on the Z evolve under purifying selection [75], and chicken W chromosome gene expression appears to be driven by sex-specific selection [76]. Thus, the data are contradictory about whether chromosome W is associated with female-specific function, indicating that this might differ among bird species. Less is known about genes on other W chromosomes, except that, notably, the W chromosome of the African clawed frog harbors sex(female)-determining gene *DM-W* (a W-linked DM-domain gene) [61].

### Sex Chromosome Structure

#### PAR Characterization

PAR has been extensively studied in eutherian mammals, but is absent in marsupials. In marsupials the absence of PAR does not cause disruption of proper segregation at meiosis which is mediated by the formation of the dense plate [77]. By contrast, the presence of PAR in placental mammals is essential for sex chromosome pairing at male meiosis and for the control of regular segregation [78]. New studies confirm that the length of PAR (Box 1) varies among eutherian mammals [79], with mouse PAR being the shortest (~700 kb) and bovine, pig, cat, and dog PARs being relatively long (5–9 Mb). PAR length appears to be much more variable in birds—it ranges from 1% to 66% of the length of the chromosome Z in the white-throated tinamou (*Tinamus guttatus*) and ostrich (*Struthio camelus*), respectively [37]. A short (630 kb) PAR is sufficient for homologous recombination and segregation during meiosis in the flycatcher, whose Z and W chromosomes are morphologically distinguishable (with Z being ~60 Mb in length and W only ~7 Mb [80]).

#### Stratum Identification

Genomic sequences open doors for comprehensive evolutionary stratum (Box 1) identification based on a gradient of alignment identities (driven by multiple recombination suppression events). In birds, there is a high diversity in the presence/absence of different strata among

species, with some similarities due to the common evolutionary history of the lineages [37,75]. Evolutionary strata were also identified in SLCs of snakes [54], killifish [66], and guppies [63].

### Repeats

NGS techniques provide a view of sex chromosome TE content. A dramatic example is the recently discovered TE density on the flycatcher NRW [17]. At 48.5%, it is much higher than that on the Z chromosome and autosomes (8.8% and 5.9%, respectively), illustrating the accumulation of repeats in non-recombining regions of SLCs [20]. Similarly, the recently obtained W-chromosome sequences in the bearded dragon lizard (*Pogona vitticeps*) and Asian arowana (*Scleropages formosus*), a teleost fish, are repeat-rich [81,82]. Intriguingly, not all vertebrate Y chromosomes are heterochromatin-rich (Box 1). The mouse Y, for instance, is 99.9% euchromatic [31].

### Studies of Male and Female Lineage Divergence

SLC sequences can be used to investigate sex-specific divergence patterns. Recently, a comparison of human, chimpanzee, and gorilla Y chromosomes [43] indicated that the gorilla Y is more similar to the human Y [25] than to the chimpanzee Y [29] in terms of shared palindrome sequences, TE content, and gene repertoire. In addition, a higher proportion of the gorilla Y aligned to human Y than to chimpanzee Y. The gorilla and human Y chromosomes share all but one (VCY) protein-coding gene families. By contrast, compared to the gorilla or human Y, the chimpanzee Y has lost 25% of X-degenerate, and ~33% of ampliconic, gene families. Thus, the chimpanzee Y chromosome appears to be an outlier. This is particularly surprising because the alignment of gorilla, chimpanzee, and human Y chromosome sequences resulted in a higher identity of human and chimpanzee Y sequences than between human and gorilla, or gorilla and chimpanzee, Y chromosomes, as expected due to the fact that human and chimpanzee share the most recent common ancestor (MRCA). Thus, the data are inconsistent with incomplete lineage sorting. High conservation between the human and gorilla Y chromosomes might be explained by similarities in their mating patterns. Human and gorilla females mate with a single male during the same periovulatory cycle [83]. The sequence divergence between the human and chimpanzee Y chromosomes has been previously suggested to be explained by differences in mating patterns and levels of sperm competition [29,84]. Chimpanzee females mate with multiple males [83], and this mating pattern is associated with sperm competition and enhanced selection. If selection was acting on the chimpanzee Y, it was likely accompanied by genetic hitchhiking driving gene pseudogenization.

The nucleotide diversity studied in the 2.6–3.8 Mb fragment of the NRY in apes also reflected their mating patterns [47]. Phylogenetic analysis of the data informed the time to the MRCA of major ape species and subspecies, and discovered a potential correlation between generation time and the branch lengths on the Y chromosome tree [47]. Partial Y chromosome (a total of 120 kb) was sequenced from a Neanderthal male (*Homo sapiens neanderthalensis*) [85] and compared with the modern human Y. The resulting estimate of the time to the MRCA of the two lineages was similar to that obtained using autosomal sequences, and several nonsynonymous differences between the two Y chromosomes were identified. The analysis suggested that genetic incompatibilities of the Y with the rest of the genome contributed to the loss of Neanderthal Y chromosome in modern humans [85].

Two highly divergent Y chromosome lineages were identified in polar bears (*Ursus maritimus*) [36] and compared with those of black (*U. americanus*) and brown bears (*U. arctos*). The divergence dates of these lineages were consistent with the fossil record of American black bears.

W chromosome sequences can be used to trace maternal lineages and are expected to cosegregate with maternally inherited mitochondrial DNA. Indeed, a population genomics study of

96 female genomes of four flycatcher species (collared flycatcher *Ficedula albicollis*, pied flycatcher *F. hypoleuca*, semi-collared flycatcher *F. semitorquata*, and Atlas flycatcher *F. speculigera*) resulted in an identical phylogenetic topology between these two genetic markers, proving the validity of the identification of the W-linked sequences with NGS [17].

### Diversity of SLCs

Novel technologies and the availability of a high-quality reference sequences allow sequencing of SLCs on a population scale. The human Y in particular was sequenced in hundreds of individuals representing different populations [73,86–88]. This sequence collection allowed calibration of Y chromosome phylogeny and the study of male migration patterns and Y chromosome diversity worldwide. On the Y, the nucleotide diversity is lower than expected given autosomal diversity in Africans and Europeans [89]. The levels of nucleotide diversity on the human Y are low, a pattern inconsistent with neutral evolution [89]. Echoing this result, sequencing 96 female flycatchers resulted in very low levels of diversity on chromosome W [17], in agreement with diversity surveys of the domesticated chicken W [13]. Sexual selection is frequently used to explain low levels of polymorphism observed on the human Y [90], and might be also the reason behind a similarly low diversity on the W in birds, although such an explanation for birds has been controversial [13].

### Perspectives and Conclusions

There are several directions that, in our view, are most promising for future SLC research. Some of these are methodological (Box 3), while those discussed below relate to future SLC sequence and assembly analyses (see Outstanding Questions).

First, SLC assemblies and other genomic resources are expected to accelerate the identification of MSDGs which are still unknown for many vertebrates. Previously such genes were discovered via laborious cloning and sex-linkage techniques. NGS data can accelerate such research, as was recently shown for some fish species in which MSDG candidates were detected [66,82]. Of particular interest will be the identification of MSDGs in organisms with recent sex chromosome loss (e.g., *Ellobius*), with one type of sex chromosomes (e.g., X or W; Table 1), or with competing sex chromosomes present in the same species (X, Y, and W; Table 1). For some of these, the MSDG might be outside SLCs, requiring whole-genome sequence information.

Second, the availability of SLC sequences should allow one to include them in **genome-wide association studies** (GWASs). SLCs encode many important functions but are rarely incorporated in GWASs owing to the lack of sequence data. GWASs including the Y are commencing for human (e.g., [91]), and we predict that SLCs will soon be also included in GWASs for other species. Note that one needs to take the small effective population size of the Y into account when conducting such studies because the Y chromosome-linked markers might be strongly affected by population structure [92]. Third, relatedly, novel SLC data are expected to precipitate functional analysis of SLCs in many species. With new sequences and assemblies accumulating rapidly, we anticipate that research on sex chromosome epigenetics, transcriptomics, and proteome analysis will gain new momentum. Fourth, to fully understand the biology of human male infertility disorders, functional analyses of Y chromosomal genes and non-genic regulatory sequences are needed. The recent availability of the Y chromosome sequences for macaque, a commonly used biomedical model organism whose Y chromosome is similar to that of human [30], might open doors for direct fertility studies *in vivo*.

Fifth, assemblies of SLCs from multiple species will allow the reconstruction of their complex evolutionary trajectories. An immediate goal is to unequivocally determine the SLC homology both among themselves and with other chromosomes in vertebrate genomes. It has been

observed that the same MSDGs are recruited multiple times in vertebrate evolution (Figure 1) and that some regions of amniote genomes carrying different MSDGs share synteny across taxonomic groups [93,94]. New data will allow us to determine whether these result from ancient super-sex chromosomes or are a consequence of a predisposition of particular genomic sequences to have a sex-determining role [93].

### Acknowledgments

The authors acknowledge support from the National Science Foundation (awards DBI-ABI 0965596 to K.D.M.; DBI-1356529, IIS-1453527, IIS-1421908, and CCF-1439057 to P.M.), the Penn State Clinical and Translational Sciences Institute, and the Pennsylvania Department of Health using Tobacco Settlement Funds (the Department specifically disclaims responsibility for any analyses, interpretations or conclusions).

### References

- National Center for Biotechnology Information (2016) Eukaryotic genomes annotated at NCBI. Updated October 7, 2016. [http://www.ncbi.nlm.nih.gov/genome/annotation\\_euk/all/](http://www.ncbi.nlm.nih.gov/genome/annotation_euk/all/)
- Roos, C. *et al.* (2011) Nuclear versus mitochondrial DNA: evidence for hybridization in colobine monkeys. *BMC Evol. Biol.* 11, 77
- Tosi, A.J. *et al.* (2005) Y-chromosomal markers suitable for noninvasive studies of guenon hybridization. *Int. J. Primatol.* 26, 685–696
- Vallender, E.J. and Lahn, B.T. (2004) How mammalian sex chromosomes acquired their peculiar gene content. *BioEssays* 26, 159–169
- Dhanoo, J.K. *et al.* (2016) Y-chromosomal genes affecting male fertility: a review. *Vet. World* 9, 783–791
- Kido, T. and Lau, Y.-F.C. (2015) Roles of the Y chromosome genes in human cancers. *Asian J. Androl.* 17, 373–380
- Wang, Q. *et al.* (2013) Genetic basis of Y-linked hearing impairment. *Am. J. Hum. Genet.* 92, 301–306
- Charchar, F.J. *et al.* (2012) Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome. *The Lancet* 379, 915–922
- Jiang, P.P. *et al.* (2010) Y not a dead end: epistatic interactions between Y-linked regulatory polymorphisms and genetic background affect global gene expression in *Drosophila melanogaster*. *Genetics* 186, 109–118
- Lemos, B. *et al.* (2010) Epigenetic effects of polymorphic Y chromosomes modulate chromatin components, immune response, and sexual conflict. *Proc. Natl. Acad. Sci. U. S. A.* 107, 15826–15831
- Ellegren, H. (2011) Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat. Rev. Genet.* 12, 157–166
- Ceplitis, H. (2003) Adaptive molecular evolution of HINTW, a female-specific gene in birds. *Mol. Biol. Evol.* 21, 249–254
- Berlin, S. and Ellegren, H. (2005) Fast accumulation of non-synonymous mutations on the female-specific W chromosome in birds. *J. Mol. Evol.* 62, 66–72
- Wilson Sayres, M.A. and Makova, K.D. (2011) Genome analyses substantiate male mutation bias in many species. *BioEssays* 33, 938–945
- Wilson Sayres, M.A. (2011) Do variations in substitution rates and male mutation bias correlate with life-history traits? A study of 32 mammalian genomes. *Evolution* 65, 2800–2815
- Amster, G. and Sella, G. (2016) Life history effects on the molecular clock of autosomes and sex chromosomes. *Proc. Natl. Acad. Sci. U. S. A.* 113, 1588–1593
- Smeds, L. *et al.* (2015) Evolutionary analysis of the female-specific avian W chromosome. *Nat. Commun.* 6, 7330–7310
- Charlesworth, B. (1978) Model for evolution of Y chromosomes and dosage compensation. *Proc. Natl. Acad. Sci. U. S. A.* 75, 5618–5622
- Charlesworth, B. (1991) The evolution of sex chromosomes. *Science* 251, 1030–1033
- Charlesworth, B. and Charlesworth, D. (2000) The degeneration of Y chromosomes. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 355, 1563–1572
- Kirkpatrick, M. (2016) The evolution of genome structure by natural and sexual selection. *J. Hered.* 108, 3–11
- Mank, J.E. (2009) The W, X, Y and Z of sex-chromosome dosage compensation. *Trends Genet.* 25, 226–233
- Charlesworth, B. (1998) Sex chromosomes: evolving dosage compensation. *Curr. Biol.* 8, R931–933
- Ross, M.T. *et al.* (2005) The DNA sequence of the human X chromosome. *Nature* 434, 325–337
- Skaletsky, H. *et al.* (2003) The male-specific region of the human Y chromosome is a mosaic of discrete sequence classes. *Nature* 423, 825–837
- Itoh, Y. *et al.* (2007) Dosage compensation is less effective in birds than in mammals. *J. Biol.* 6, 2
- Kawakami, T. *et al.* (2014) A high-density linkage map enables a second-generation collared flycatcher genome assembly and reveals the patterns of avian recombination rate variation and chromosomal evolution. *Mol. Ecol.* 23, 4035–4058
- Deakin, J.E. *et al.* (2008) The status of dosage compensation in the multiple X chromosomes of the platypus. *PLoS Genet.* 4, e1000140
- Hughes, J.F. *et al.* (2010) Chimpanzee and human Y chromosomes are remarkably divergent in structure and gene content. *Nature* 463, 536–539
- Hughes, J.F. *et al.* (2012) Strict evolutionary conservation followed rapid gene loss on human and rhesus Y chromosomes. *Nature* 483, 82–86
- Soh, Y.Q.S. *et al.* (2014) Sequencing the mouse Y chromosome reveals convergent gene acquisition and amplification on both sex chromosomes. *Cell* 159, 800–813
- Gamble, T. and Zarkower, D. (2014) Identification of sex-specific molecular markers using restriction site-associated DNA sequencing. *Mol. Ecol. Resour.* 14, 902–913
- Cortez, D. *et al.* (2014) Origins and functional evolution of Y chromosomes across mammals. *Nature* 508, 488–493
- Muyle, A. *et al.* (2016) SEX-DETECTOR: a probabilistic approach to study sex chromosomes in non-model organisms. *Genome Biol. Evol.* 8, 2530–2543
- Yoshida, K. *et al.* (2014) Sex chromosome turnover contributes to genomic divergence between incipient stickleback species. *PLoS Genet.* 10, e1004223
- Bidon, T. *et al.* (2015) Genome-wide search identifies 1.9 Mb from the polar bear Y chromosome for evolutionary analyses. *Genome Biol. Evol.* 7, 2010–2022
- Zhou, Q. *et al.* (2014) Complex evolutionary trajectories of sex chromosomes across bird taxa. *Science* 346, 1246338–1246338
- Zhou, R.-N. and Hu, Z.-M. (2007) The development of chromosome microdissection and microcloning technique and its applications in genomic research. *Curr. Genomics* 8, 67–72

### Outstanding Questions

How do vertebrates switch from utilizing one master sex-determining gene to utilizing another one? How quickly and why does this occur?

Why does the emergence of new sex chromosomes occur more often in some species than in other species? Do sex chromosomes promote speciation?

Is there an evolutionary advantage to using and reusing particular genomic regions as sex-determining regions in unrelated vertebrates?

How is sex determined in mammals that recently lost their sex chromosomes?

Do mating patterns influence the evolution of SLCs?

Do Y and W chromosomes fulfill functions unrelated to sex determination in different vertebrate species?

Do phylogenies based on Y and W chromosome sequences concur with those obtained from autosomal sequences, or do they differ, reflecting sex-specific evolutionary history of the species studied?

Why do Y and W chromosomes frequently have such low diversity?



39. Kichigin, I.G. *et al.* (2016) Evolutionary dynamics of *Anolis* sex chromosomes revealed by sequencing of flow sorting-derived microchromosome-specific DNA. *Mol. Genet. Genomics* 291, 1955–1966
40. Traut, W. *et al.* (2013) High-throughput sequencing of a single chromosome: a moth W chromosome. *Chromosome Res.* 21, 491–505
41. Doležal, J. *et al.* (2012) Chromosomes in the flow to simplify genome analysis. *Func. Integr. Genomics* 12, 397–416
42. Brinkrolf, K. *et al.* (2013) Chinese hamster genome sequenced from sorted chromosomes. *Nat. Biotechnol.* 31, 694–695
43. Tomaszewicz, M. (2016) A time- and cost-effective strategy to sequence mammalian Y chromosomes: an application to the *de novo* assembly of gorilla Y. *Genome Res.* 26, 530–540
44. Skinner, B.M. *et al.* (2016) The pig X and Y chromosomes: structure, sequence, and evolution. *Genome Res.* 26, 130–139
45. Li, G. *et al.* (2013) Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. *Genome Res.* 23, 1486–1495
46. Murtagh, V.J. *et al.* (2012) Evolutionary history of novel genes on the tammar wallaby Y chromosome: implications for sex chromosome evolution. *Genome Res.* 22, 498–507
47. Hallast, P. *et al.* (2016) Great ape Y chromosome and mitochondrial DNA phylogenies reflect subspecies structure and patterns of mating and dispersal. *Genome Res.* 26, 427–439
48. Murata, C. *et al.* (2015) Initiation of recombination suppression and PAR formation during the early stages of neo-sex chromosome differentiation in the Okinawa spiny rat, *Tokudaia muenninki*. *BMC Evol. Biol.* 15, 234
49. Chen, N. *et al.* (2012) Identification of avian W-linked contigs by short-read sequencing. *BMC Genomics* 13, 183
50. Saxena, R. *et al.* (2000) Four DAZ genes in two clusters found in the AZFc region of the human Y chromosome. *Genomics* 67, 256–267
51. Carvalho, A.B. and Clark, A.G. (2013) Efficient identification of Y chromosome sequences in the human and *Drosophila* genomes. *Genome Res.* 23, 1894–1907
52. Koerich, L.B. *et al.* (2016) First report of Y-linked genes in the kissing bug *Rhodnius prolixus*. *BMC Genomics* 17, 100
53. Hall, A.B. *et al.* (2013) Six novel Y chromosome genes in *Anopheles* mosquitoes discovered by independently sequencing males and females. *BMC Genomics* 14, 273
54. Vicoso, B. *et al.* (2013) Comparative sex chromosome genomics in snakes: differentiation, evolutionary strata, and lack of global dosage compensation. *PLoS Biol.* 11, e1001643
55. Hall, A.B. *et al.* (2016) Radical remodeling of the Y chromosome in a recent radiation of malaria mosquitoes. *Proc. Natl. Acad. Sci. U. S. A.* 113, E2114–E2123
56. Vicoso, B. and Bachtrog, D. (2015) Numerous transitions of sex chromosomes in diptera. *PLoS Biol.* 13, e1002078
57. Graves, J.A.M. (2016) Evolution of vertebrate sex chromosomes and dosage compensation. *Nat. Rev. Genet.* 17, 33–46
58. Chen, S. *et al.* (2014) Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat. Genet.* 46, 253–260
59. Yoshida, K. *et al.* (2016) Accumulation of deleterious mutations on the neo-Y chromosome of Japan Sea stickleback (*Gasterosteus nipponicus*). *J. Hered.* 1, 63–68
60. Tuttle, E.M. *et al.* (2016) Divergence and functional degradation of a sex chromosome-like supergene. *Curr. Biol.* 26, 344–350
61. Mawaribuchi, S. *et al.* (2016) Sex chromosome differentiation and the W- and Z-specific loci in *Xenopus laevis*. *Dev. Biol.* Published online June 11, 2016. <http://dx.doi.org/10.1016/j.ydbio.2016.06.015>
62. Gammerdinger, W.J. (2014) Structure and decay of a proto-Y region in Tilapia, *Oreochromis niloticus*. *BMC Genomics* 15, 975
63. Wright, A.E. *et al.* (2017) Convergent recombination suppression suggests role of sexual selection in guppy sex chromosome formation. *Nat. Commun.* 8, 14251
64. Murata, C. *et al.* (2016) Ancestral Y-linked genes were maintained by translocation to the X and Y chromosomes fused to an autosomal pair in the Okinawa spiny rat *Tokudaia muenninki*. *Chromosome Res.* 24, 407–419
65. Yoshida, K. and Kitano, J. (2012) The contribution of female meiotic drive to the evolution of neo-sex chromosomes. *Evolution* 66, 3198–3208
66. Reichwald, K. *et al.* (2015) Insights into sex chromosome evolution and aging from the genome of a short-lived fish. *Cell* 163, 1527–1538
67. Bertolini, F. *et al.* (2015) Next generation semiconductor based sequencing of the donkey (*Equus asinus*) genome provided comparative sequence data against the horse genome and a few millions of single nucleotide polymorphisms. *PLoS One* 10, e0131925
68. Mulugeta, E. *et al.* (2016) Genomes of *Ellobius* species provide insight into the evolutionary dynamics of mammalian sex chromosomes. *Genome Res.* 26, 1202–1210
69. Just, W. *et al.* (1995) Absence of Sry in species of the vole *Ellobius*. *Nat. Genet.* 11, 117–118
70. Yamauchi, Y. *et al.* (2016) Two genes substitute for the mouse Y chromosome for spermatogenesis and reproduction. *Science* 351, 514–516
71. Oetjens, M.T. *et al.* (2016) Y-chromosome structural diversity in the bonobo and chimpanzee lineages. *Genome Biol. Evol.* 8, 2231–2240
72. Ghenu, A.-H. *et al.* (2016) Multicopy gene family evolution on primate Y chromosomes. *BMC Genomics* 17, 157
73. Wei, W. *et al.* (2015) Copy number variation in the human Y chromosome in the UK population. *Hum. Genet.* 134, 789–800
74. Chang, T.-C. *et al.* (2013) Male-specific region of the bovine Y chromosome is gene rich with a high transcriptomic activity in testis development. *Proc. Natl. Acad. Sci. U. S. A.* 110, 12373–12378
75. Wright, A.E. *et al.* (2014) Independent stratum formation on the avian sex chromosomes reveals inter-chromosomal gene conversion and predominance of purifying selection on the W chromosome. *Evolution* 68, 3281–3295
76. Moghadam, H.K. *et al.* (2012) W chromosome expression responds to female-specific selection. *Proc. Natl. Acad. Sci. U. S. A.* 109, 8207–8211
77. Page, J. (2002) The pairing of X and Y chromosomes during meiotic prophase in the marsupial species *Thylamys elegans* is maintained by a dense plate developed from their axial elements. *J. Cell Sci.* 116, 551–560
78. Waters, P.D. *et al.* (2007) Sex chromosomes of basal placental mammals. *Chromosoma* 116, 511–518
79. Raudsepp, T. and Chowdhary, B.P. (2016) The eutherian pseudoautosomal region. *Cytogenet. Genome Res.* 147, 81–94
80. Smeds, L. *et al.* (2014) Genomic identification and characterization of the pseudoautosomal region in highly differentiated avian sex chromosomes. *Nat. Commun.* 5, 5448–5447
81. Ezaz, T. *et al.* (2005) The dragon lizard *Pogona vitticeps* has ZZ/ZW micro-sex chromosomes. *Chromosome Res.* 13, 763–776
82. Bian, C. *et al.* (2016) The Asian arowana (*Scleropages formosus*) genome provides new insights into the evolution of an early lineage of teleosts. *Sci. Rep.* 6, 24501
83. Moller, A.P. (1988) Ejaculate quality, testes size and sperm competition in primates. *J. Hum. Evol.* 17, 479–488
84. Hughes, J.F. *et al.* (2005) Conservation of Y-linked genes during human evolution revealed by comparative sequencing in chimpanzee. *Nature* 437, 100–103
85. Mendez, F.L. *et al.* (2016) The divergence of Neandertal and modern human Y chromosomes. *Am. J. Hum. Genet.* 98, 728–734
86. Poznik, G.D. *et al.* (2013) Sequencing Y chromosomes resolves discrepancy in time to common ancestor of males versus females. *Science* 341, 562–565
87. Poznik, G.D. *et al.* (2016) Punctuated bursts in human male demography inferred from 1,244 worldwide Y-chromosome sequences. *Nat. Genet.* 48, 593–599
88. Karmin, M. *et al.* (2015) A recent bottleneck of Y chromosome diversity coincides with a global change in culture. *Genome Res.* 25, 459–466

89. Wilson Sayres, M.A. (2014) natural selection reduced diversity on human Y chromosomes. *PLoS Genet.* 10, e1004064
90. Dorit, R.L. *et al.* (1995) Absence of polymorphism at the ZFY locus on the human Y chromosome. *Science* 268, 1183–1185
91. Lu, C. *et al.* (2016) Y chromosome haplogroups based genome-wide association study pinpoints revelation for interactions on non-obstructive azoospermia. *Sci. Rep.* 6, 33363
92. McCarroll, S.A. (2008) Extending genome-wide association studies to copy-number variation. *Hum. Mol. Genet.* 17, R135–R142
93. Ezaz, T. *et al.* (2016) Origin of amniote sex chromosomes: an ancestral super-sex chromosome, or common requirements? *J. Hered.* 108, 63–68
94. Furman, B.L.S. and Evans, B.J. (2016) Sequential turnovers of sex chromosomes in african clawed frogs (*Xenopus*) suggest some genomic regions are good at sex determination. *Genetics* 6, 3625–3633
95. Ezaz, T. *et al.* (2006) Relationships between vertebrate ZW and XY sex chromosome systems. *Curr. Biol.* 16, R736–R743
96. Roco, Á.S. *et al.* (2015) Coexistence of Y, W, and Z sex chromosomes in *Xenopus tropicalis*. *Proc. Natl. Acad. Sci. U. S. A.* 112, E4752–E4761
97. Kallman, K.D. (1984) A new look at sex determination in poeciliid fishes. In *Evolutionary Genetics of Fishes* (Turner, B.J., ed.), pp. 95–171, Springer
98. Charlesworth, B. and Dempsey, N.D. (2001) A model of the evolution of the unusual sex chromosome system of *Microtus oregoni*. *Heredity* 86, 387–394
99. Jotterand-Bellomo, M. (1988) Chromosome analysis of five specimens of *Mus bufo-triton* (Muridae) from Burundi (Africa): three cytogenetic entities, a special type of chromosomal sex determination, taxonomy, and phylogeny. *Cytogenet. Genome Res.* 48, 88–91
100. Veyrunes, F. *et al.* (2010) A novel sex determination system in a close relative of the house mouse. *Proc. Biol. Sci.* 277, 1049–1056
101. Fredga, K. (1976) Fertile XX- and XY-type females in the wood lemming *Myopus schisticolor*. *Nature* 261, 225–227
102. Fredga, K. (1988) Aberrant chromosomal sex-determining mechanisms in mammals, with special reference to species with XY females. *Philos. Trans. R Soc. Lond. B Biol. Sci.* 322, 83–94
103. Green, D.M. (1988) Cytogenetics of the endemic New Zealand frog, *Leiopelma hochstetteri*: extraordinary supernumerary chromosome variation and a unique sex-chromosome system. *Chromosoma* 97, 55–70
104. Kitano, J. and Peichel, C.L. (2011) Turnover of sex chromosomes and speciation in fishes. *Environ. Biol. Fish.* 94, 549–558
105. Schmid, M. *et al.* (2002) Chromosome banding in amphibia. XXVI. Coexistence of homomorphic XY sex chromosomes and a derived Y–autosome translocation in *Eleutherodactylus maussii* (Anura, Leptodactylidae). *Cytogenet. Genome Res.* 99, 330–343
106. Bigoni, F. *et al.* (1997) Fluorescence *in situ* hybridization establishes homology between human and silvered leaf monkey chromosomes, reveals reciprocal translocations between chromosomes homologous to human Y/5, 1/9, and 6/16, and delineates an X1X2Y1Y2/X1X1X2X2 sex-chromosome system. *Am. J. Phys. Anthropol.* 102, 315–327
107. Steinberg, E. *et al.* (2014) Multiple sex chromosome systems in howler monkeys (*Platyrrhini, Alouatta*). *Comparative Cytogen.* 8, 43–69
108. Grützner, F. *et al.* (2004) In the platypus a meiotic chain of ten sex chromosomes shares genes with the bird Z and mammal X chromosomes. *Nature* 432, 913–917
109. Rens, W. *et al.* (2007) The multiple sex chromosomes of platypus and echidna are not completely identical and several share homology with the avian Z. *Genome Biol.* 8, R243
110. Lahn, B.T. and Page, D.C. (1999) Four evolutionary strata on the human X chromosome. *Science* 286, 964–967
111. Rozen, S. *et al.* (2003) Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. *Nature* 423, 873–876
112. Miga, K.H. *et al.* (2014) Centromere reference models for human chromosomes X and Y satellite arrays. *Genome Res.* 24, 697–707
113. Bhowmick, B.K. *et al.* (2007) The origin and evolution of human ampliconic gene families and ampliconic structure. *Genome Res.* 17, 441–450
114. International Chicken Genome Sequencing Consortium (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* 432, 695–716
115. Solari, A.J. and Dresser, M.E. (1995) High-resolution cytological localization of the XhoI and EcoRI repeat sequences in the pachytene ZW bivalent of the chicken. *Chromosome Res.* 3, 87–93
116. Backstrom, N. (2005) Gene conversion drives the evolution of HINTW, an ampliconic gene on the female-specific avian W chromosome. *Mol. Biol. Evol.* 22, 1992–1999
117. Hori, T. *et al.* (2000) Wpkci, encoding an altered form of PKCI, is conserved widely on the avian W chromosome and expressed in early female embryos: implication of its role in female sex determination. *Mol. Biol. Cell* 11, 3645–3660
118. van Heesch, S. *et al.* (2013) Improving mammalian genome scaffolding using large insert mate-pair next-generation sequencing. *BMC Genomics* 14, 257
119. Neely, R.K. *et al.* (2011) Optical mapping of DNA: single-molecule-based methods for mapping genomes. *Biopolymers* 95, 298–311
120. van Berkum, N.L. *et al.* (2010) Hi-C: a method to study the three-dimensional architecture of genomes. *J. Vis. Exp.* Published online May 6, 2010. <http://dx.doi.org/10.3791/1869>
121. Simpson, J.T. and Pop, M. (2015) The theory and practice of genome sequence assembly. *Annu. Rev. Genom. Hum. Genet.* 16, 153–172
122. Alkan, C. *et al.* (2010) Limitations of next-generation genome sequence assembly. *Nat. Methods* 8, 61–65
123. Wei, K.H.C. *et al.* (2014) Correlated variation and population differentiation in satellite DNA abundance among lines of *Drosophila melanogaster*. *Proc. Natl. Acad. Sci. U. S. A.* 111, 18793–18798
124. Francisco, F.O. and Lemos, B. (2014) How do Y-chromosomes modulate genome-wide epigenetic states: genome folding, chromatin sinks, and gene expression. *J. Genomics* 2, 94–103
125. Rhoads, A. and Au, K.F. (2015) PacBio sequencing and its applications. *Genomics, Proteomics & Bioinformatics* 13, 278–289
126. Feng, Y. *et al.* (2015) Nanopore-based fourth-generation DNA sequencing technology. *Genomics, Proteomics & Bioinformatics* 13, 4–16
127. Laver, T. *et al.* (2015) Assessing the performance of the Oxford Nanopore Technologies MinION. *Biomolecular Detection and Quantification* 3, 1–8
128. Laehnemann, D. *et al.* (2016) Denoising DNA deep sequencing data-high-throughput sequencing errors and their correction. *Brief. Bioinformatics* 17, 154–179
129. Hastie, A.R. *et al.* (2013) Rapid genome mapping in nanochannel arrays for highly complete and accurate *de novo* sequence assembly of the complex *Aegilops tauschii* genome. *PLoS One* 8, e55864
130. Hindson, C.M. *et al.* (2013) Absolute quantification by droplet digital PCR versus analog real-time PCR. *Nature Methods* 10, 1003–1005
131. Sudmant, P.H. *et al.* (2010) Diversity of human copy number variation and multicopy genes. *Science* 330, 641–646
132. Bellott, D.W. *et al.* (2014) Mammalian Y chromosomes retain widely expressed dosage-sensitive regulators. *Nature* 508, 494–499

133. Matsuda, M. *et al.* (2002) *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417, 559–563
134. Nanda, I. *et al.* (2002) A duplicated copy of *DMRT1* in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proc. Natl. Acad. Sci. U. S. A.* 99, 11778–11783
135. Froschauer, A. *et al.* (2002) Construction and initial analysis of bacterial artificial chromosome (BAC) contigs from the sex-determining region of the platyfish *Xiphophorus maculatus*. *Gene* 295, 247–254
136. Tomaszewicz, M. *et al.* (2014) A multicopy Y-chromosomal SGNH hydrolase gene expressed in the testis of the platyfish has been captured and mobilized by a Helitron transposon. *BMC Genet.* 15, 44
137. Phillips, R.B. *et al.* (2013) Characterization of the OmyY1 region on the rainbow trout Y chromosome. *Int J Genomics* 2013, 261730