Cell Genomics

CellPress
OPEN ACCESS

Perspective

Representing sex chromosomes in genome assemblies

Sarah B. Carey,[1,2] John T. Lovell,[2] Jerry Jenkins,[2] Jim Leebens-Mack,[3] Jeremy Schmutz,[2,4] Melissa A. Wilson,[5] and Alex Harkess[1,2,*]

[1]Department of Crop, Soil, and Environmental Sciences, Auburn University, Auburn, AL 36849, USA
[2]HudsonAlpha Institute for Biotechnology, Huntsville, AL 35806, USA
[3]Department of Plant Biology, University of Georgia, Athens, GA 30602, USA
[4]US Department of Energy Joint Genome Institute, Lawrence Berkeley National Laboratory, Berkeley, CA, USA
[5]School of Life Sciences, Center for Evolution and Medicine, The Biodesign Institute, Arizona State University, Tempe, AZ 85287, USA
*Correspondence: aharkess@hudsonalpha.org
https://doi.org/10.1016/j.xgen.2022.100132

SUMMARY

Sex chromosomes have evolved hundreds of independent times across eukaryotes. As genome sequencing, assembly, and scaffolding techniques rapidly improve, it is now feasible to build fully phased sex chromosome assemblies. Despite technological advances enabling phased assembly of whole chromosomes, there are currently no standards for representing sex chromosomes when publicly releasing a genome. Furthermore, most computational analysis tools are unable to efficiently investigate their unique biology relative to autosomes. We discuss a diversity of sex chromosome systems and consider the challenges of representing sex chromosome pairs in genome assemblies. By addressing these issues now as technologies for full phasing of chromosomal assemblies are maturing, we can collectively ensure that future genome analysis toolkits can be broadly applied to all eukaryotes with diverse types of sex chromosome systems. Here we provide best practice guidelines for presenting a genome assembly that contains sex chromosomes. These guidelines can also be applied to other non-recombining genomic regions, such as *S*-loci in plants and mating-type loci in fungi and algae.

THE HISTORY OF SEX CHROMOSOME ASSEMBLY

Dr. Nettie Stevens made the groundbreaking cytogenetic discovery that male mealworms (*Tenebrio* sp.) possessed a small chromosome that determined sex.[1] Deemed the "heterochromosome," which we now recognize as the male-specific Y chromosome, this small chromosome was never found in eggs. Since then, sex chromosomes have been identified widely across plants, animals, and fungi.[2,3] Sex chromosomes were first discovered using microscopy and today genomic analyses enable their identification, assembly, and subsequent comparative analysis.

The monumental, global effort that produced the first human genome draft published in 2000,[4] involved tiled sequencing of P1 artificial chromosomes (PACs), cosmids, and bacterial artificial chromosomes. The initial X chromosome was highly contiguous with only 14 intractable gaps.[5] It took nearly 20 more years for the human X chromosome[6] and autosomes[7] to be fully assembled, from telomere-to-telomere without any sequence gaps. Whereas substantial progress has been made in assembling the human Y chromosome,[8,9] telomere-to-telomere assembly remains unfinished due to the large heterochromatic segment taking up about two-thirds of the human Y, however, long-read sequencing is poised to resolve the complete sequence of the Y chromosomes soon as well[7,9] (Figure 1). To date, hundreds of plant and animal genomes with sex chromosomes have been sequenced, assembled, and published, with varying degrees of contiguity and completeness.[10,11] As genome sequencing technologies continue to improve with higher-fidelity long-read sequencing, combined with improvement of phased assembly and scaffolding algorithms, we expect that highly contiguous assemblies of sex chromosome pairs will soon become commonplace.

Approximately 95% of animals have separate sexes (called gonochory[12]) and 8% of land plants (called dioecy[10,13]). With several large genome projects in progress, such as the 10,000 Plants Genome Sequencing Project,[14] Earth BioGenome Project,[15] Global Invertebrates Genomics Alliance,[16] Vertebrate Genome Project,[11] and user-driven projects through the Department of Energy Joint Genome Institute (e.g., https://phytozome-next.jgi.doe.gov/ogg/), thousands of genome assemblies containing sex chromosomes will be published in the next decade. It is critical, therefore, that we develop a standard for consistent reporting of sex chromosomes in genome assemblies, if not across all gonochoric and dioecious eukaryotes, then at least for all species within taxa included in comparative analyses (e.g., mammals, birds, flies, flowering plants).

The lack of standard representation of the sex chromosome pair in a genome assembly can be attributed to the immense variation in systems across eukaryotes (Table 1). Consequently, downstream analysis tools are missing rigorous considerations for accommodating the unique nature of sex chromosomes across all eukaryote lineages; indeed, many simply ignore the

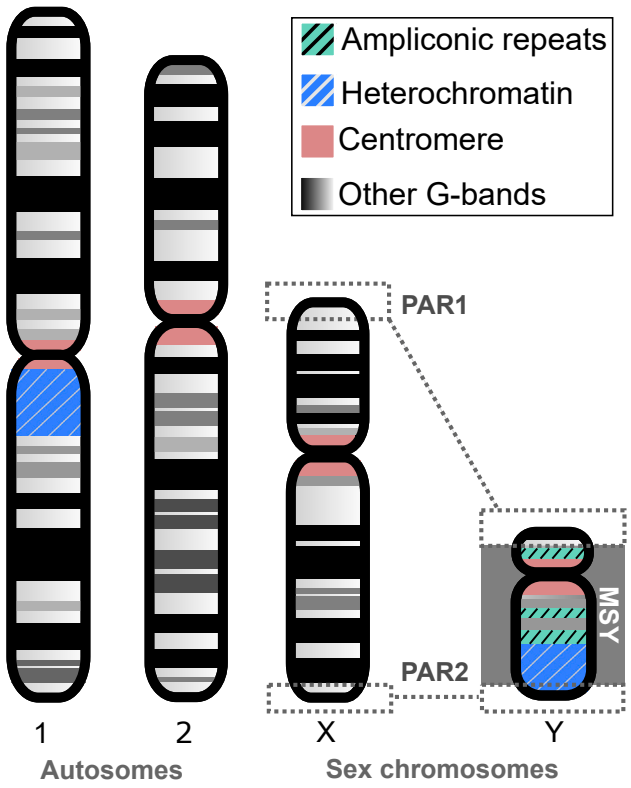

Figure 1. Ideogram of human chromosomes

The human genome reference contains a single haplotype for autosomes (here only chromosomes 1 and 2 are shown, but the logic applies to all 22 autosomes). In contrast, both of the sex chromosomes are represented in a heterogametic assembly, which is important because, although they were once entirely homologous, they are highly diverged across most of their lengths. The male-specific region of the Y (MSY), also called the sex determination region (SDR), in humans has lost most genes and has accumulated many repeats, like in the ampliconic regions where the repeats have high sequence similarity (>99%) and can be found in palindromes or tandem arrays, and it has more heterochromatic regions when compared with the X. In contrast, the pseudoautosomal regions (PAR), which pair and freely recombine during meiosis, share 100% homology and are represented twice.

sex chromosomes all together. Here we outline the key issues with sex chromosome structure that impede genome assembly and describe how current technologies are poised to change these norms. Importantly, we describe key considerations for reporting sex chromosomes in genome assembly releases, encompassing X/Y, Z/W, and U/V sex chromosomes. These considerations will be crucial for ensuring that computational genomic analysis toolkits can be broadly applied to the oncoming deluge of genome assemblies with sex chromosomes, and that there is a consistent and practical format for releasing these genomes in public repositories.

THE STRUCTURE OF SEX CHROMOSOMES

Over the last century of research into sex chromosome evolution, several key similarities have emerged among many, but not all, sex chromosomes. Sex chromosomes can evolve from an ancestral pair of autosomes, typically forming a region of suppressed recombination between the sex chromosome pair, called a "non-recombining region" or "sex determination region" (SDR) (Figure 1). Whereas the genes that initiate female or male sex determination typically reside in the SDR,[49] there are clear cases where these sex determination genes have translocated to other chromosomes.[50] Instead, for some systems, like in humans, a better way to refer to the non-recombining region is the male-specific region of the Y (MSY; Figure 1). To encompass a wide range of sex chromosome types across kingdoms, which we describe below, for simplicity we will use SDR to refer to the non-recombining region of a sex chromosome. In systems studied to date, the SDR varies in size, ranging from <100 kilobases (Kb) to >100 megabases (Mb), accounting for <1% to nearly 100% of a sex chromosome's length (Figure 2). Flanking these non-recombining regions is the pseudoautosomal region (PAR), which is the homologous sequence of both sex chromosomes that pairs normally at meiosis and can recombine (Figure 1).

The SDR has been shown to evolve in existing regions of low recombination, including centromeres,[51] arise from large-scale mutations that inhibit recombination, including inversions,[52–54] deletions, or translocations, resulting in hemizygosity[18,55,56] (Figure 2), or through the gradual build-up of transposable elements.[57] While some sex chromosome pairs are stable across taxa, having a single origin tens of millions of years ago,[25,45,58,59] others are more labile and frequently transition to a new, non-homologous chromosome pair[55,60] or have a recent, independent origin from a hermaphroditic ancestor.[49]

After their initial evolution, SDRs evolve on different molecular evolutionary trajectories than autosomes and PARs. The lack of recombination reduces the efficacy of natural selection, allowing for substantial changes in the sex chromosome haplotype, such as further structural variation, gene loss, and repeat accumulation.[61] An extreme example is the human XY, where 90% of the ancestral genes have been lost on the Y chromosome relative to the X over its 160 million years of evolution[62] (Figure 1). In other cases, like the flowering plant *Silene latifolia*, the Y chromosome has expanded with repetitive DNA to nearly twice the size of the X chromosome over the past 11 million years, but retains many homologous genes.[63,64] These "degenerative" processes occur at different structural and temporal scales across taxa, creating a kaleidoscope of sex chromosome haplotype variation.[49,65]

Sex chromosomes also have incredibly diverse pairing systems, chromosomal structures, and genes that determine sex. For the purposes of this review, we define three major sexual chromosome systems that most plant and animal species fall into: X/Y, Z/W, and U/V (Figure 2). The differences between X/Y and Z/W systems depend on which sex, male or female, is heterogametic for the sex chromosome pair (i.e., can make gametes containing different sex chromosomes). In X/Y systems, males are typically heterogametic, carrying both an X and Y chromosome as a pair. Females are typically homogametic, carrying two copies of an X chromosome. In ZW systems, females are the heterogametic sex, carrying a Z and W, while males are ZZ. A third system, U/V, is found in haploid-dominant systems, where females inherit a single U chromosome and males a single V^2.

Table 1. Examples of sex chromosome variation across animals and plants

Species	Sex chromosome cytology	Source
Pufferfish	proto-XY	Kamiya et al[17]
Garden asparagus, papaya, green anole	Homomorphic XY	Harkess et al[18]; Liu et al[19]; Alfoldi et al[20]
Mealworm, human, common hop, white campion	Heteromorphic XY	Stevens[1]; Rozen et al[21]; Winge[22]; Westergaard[23]
Japanese hop	XY_1Y_2	Kihara[24]
Platypus	$X_1X_2X_3X_4X_5Y_1Y_2Y_3Y_4Y_5$	Veyrunes et al[25]
Smoky jungle frog	$X_1X_2X_3X_4X_5X_6Y_1Y_2Y_3Y_4Y_5Y_6$	Gazoni et al[26]
Spiny rat, nematodes	XO	Kobayashi et al[27]; Hodgkin[28]
Most spiders	X_1X_2O	Kral[29]
Heartwing sorrel	XY and XY_1Y_2	Smith[30]
Black muntjac deer, *Drosophila miranda*	neo-XY	Zhou et al[31]; Bachtrog and Charlesworth[32]
Strawberries	proto-ZW	Spigler et al[33]
Emu, boa constrictor, red bayberry	Homomorphic ZW	Ellegren[34]; Ohno[35]; Jia et al[36]
Chicken	Heteromorphic ZW	Hirst et al[37]
Marsh marigold moth	ZO	Traut and Marec[38]
Hochstetter's frog	WO	Green et al[39]
Darter characin fish	ZW_1W_2	Filho et al[40]
Ancistrus catfishes	$Z_1Z_2W_1W_2$	de Oliveira et al[41]
Northeast-Asian wood white butterfly	$Z_1Z_2Z_3Z_4Z_5Z_6W_1W_2W_3$	Sichova et al[42]
Western clawed frog, Burtoni cichlid fish	YWZ	Roco et al[43]; Roberts et al[44]
Fire moss	Homomorphic UV	Carey et al., 2021[45]
Common liverwort, *Sphaerocarpos* liverwort	Heteromorphic UV	Yamato et al[46]; Allen[47]
Dilated scalewort	U_1U_2V	Sousa et al[48]

Note that many multiple sex chromosome systems may arise through the formation of neo-sex chromosomes but are not indicated here.

There is also remarkable diversity in sex chromosome cytotypes, including variation in the size of the Y/W compared with the X/Z (i.e., hetero- versus homogametic), dosage systems where one sex chromosome in the pair was lost (e.g., XX/XO or ZZ/ZO sex determination systems known in some species; Table 1), and multiple sex chromosome pairs (e.g., $X_1X_2Y_1Y_2$), as well as diversity within a species or genus, including aneuploidies and those with neo-sex chromosomes (Figure 2; Table 1). Because the non-recombinant SDRs of sex chromosomes evolve on separate evolutionary trajectories from each other and from the autosomes, the SDR haplotypes can diverge rapidly, producing tremendous sequence, structural, and functional variation among populations and species.[66]

CHALLENGES OF SEX CHROMOSOME ASSEMBLY

Because of the complex nature of SDRs, and half of the sequencing coverage relative to autosomes in XY or ZW genotypes, it is far more challenging to generate assemblies of sex chromosomes than for autosomes. Consequently, sex chromosomes have been the most poorly assembled and annotated regions of plant and animal genomes. For example, sex chromosomes in the Vertebrate Genome Project assemblies were typically more fragmented than autosomes.[11] Advances in genome sequencing, assembly, and long-range scaffolding techniques are poised to change this trend. Pacific Biosciences (PacBio) high-fidelity (HiFi) reads are medium sized (15–25 kb) and high accuracy (99%+), enabling the highly contiguous and allele-phased assembly of complex genomes.[67] Oxford Nanopore Technologies reads can reach multi-Mb sizes though with a higher error rate, and were a key tool in scaffolding the first telomere-to-telomere X chromosome in humans.[6]

While genome sequencing techniques have rapidly advanced, a key complication is that genome assembly algorithms are not designed with sex chromosomes in mind. The current generation of PacBio HiFi assembly algorithms, such as hifiasm,[68] IPA (https://github.com/PacificBiosciences/pbbioconda/wiki/Improved-Phased-Assembler), HiCanu,[69] and Flye[70] are designed to phase structurally similar autosomes into separate allelic haplotypes. Sex chromosomes often do not conform to this expectation, given their potentially large heteromorphy that can involve size, gene content, repeat content, and structural variation between the two members of a sex chromosome pair (Figures 1 and 2). In our experience, accurate HiFi assembly of sex chromosomes requires at least two additional analysis processes: Hi-C scaffolding and genetic inference of the identity of contigs belong to the non-recombining region of sex chromosomes. Inference of sex linkage can be aided by identification of sex-specific sequences and sex-biased sequencing coverage in analyses of relatively inexpensive short-read sequence data.[71,72] Integrated analyses of phased PacBio HiFi genome assemblies, Hi-C, and standard short-read data are now enabling the full-length, accurately phased assembly of sex chromosomes,[73] although there are certainly cases where sex chromosome assembly will remain challenging (e.g., large genomes, polyploidy, high repeat content).

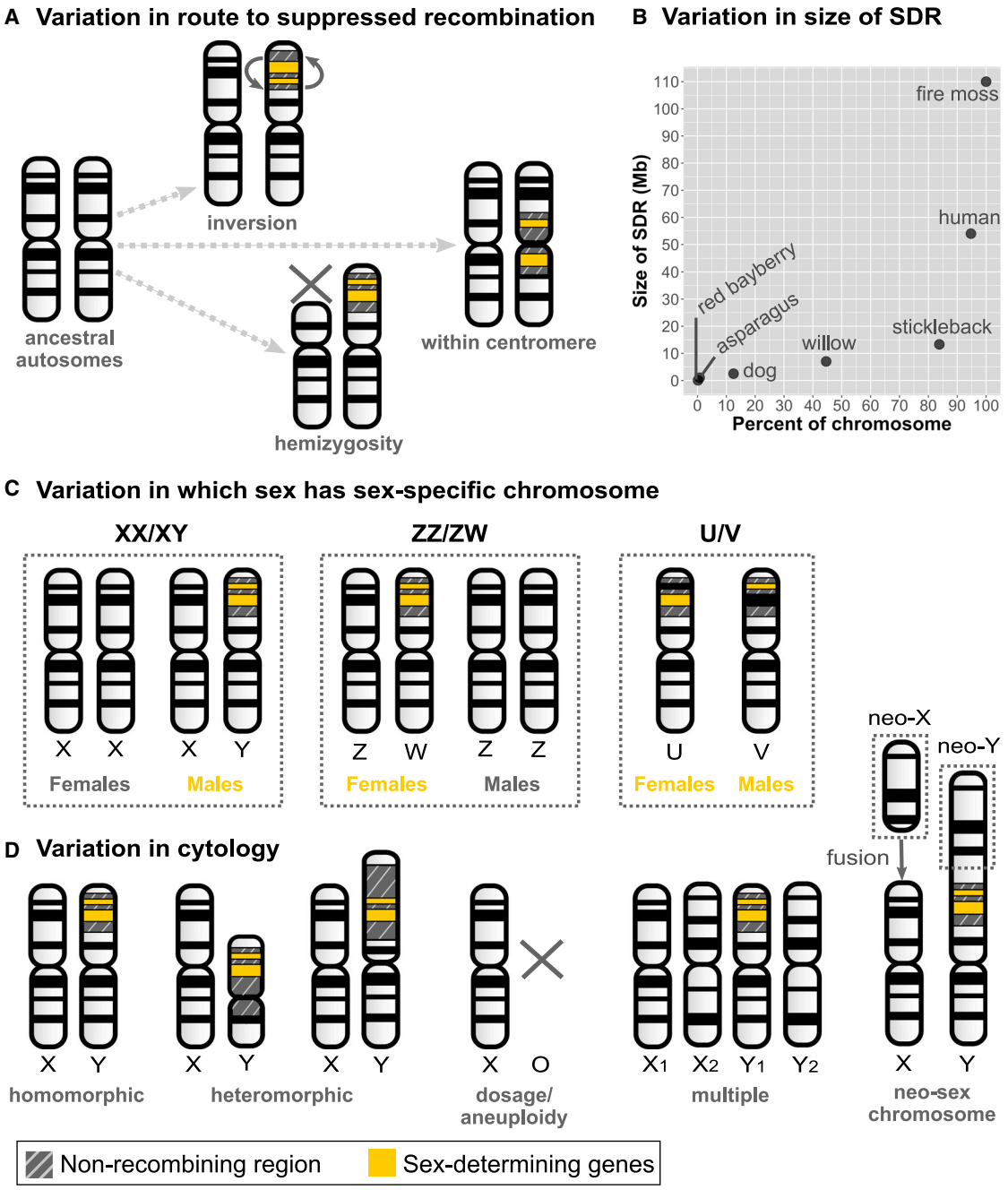

Figure 2. Remarkable variation found across sex chromosomes

(A) Different routes to suppressed recombination have been identified involving inversions or hemizygosity through deletions or translocations. Some SDRs have instead evolved in regions of existing low recombination, such as centromeres.

(B) The size of the SDR varies across species, with some <1 Mb, representing <1% of the sex chromosome, while others are >110 Mb and across the entirety of the sex chromosome.

(C) There are differences in which sex contains the sex-specific chromosome. In XX/XY systems, males are XY, while females are XX. In ZZ/ZW systems, the opposite is true, where females are the heterogametic sex inheriting ZW and males are ZZ. In species that have haploid sex determination, the inheritance of a single U chromosome correlates with females and a single V with males.

(D) There is also cytological variation between the homologous pairs of sex chromosomes. Some are homomorphic, where the X and Y are the same in size, while others are heteromorphic, where either the X or Y is larger. In others, the sex-specific chromosome like the Y has been lost, and dosage of genes on the X determines sex. In other systems, several chromosomes are inherited in a sex-specific fashion, called "multiple" sex chromosomes. Neo-sex chromosomes have also been identified, where a fusion between an autosomal pair and the sex chromosomes has occurred. Examples for each of these sex chromosome types can be found in Table 1.

ISSUES WITH SEX CHROMOSOME INFORMATICS

Most analytical and assembly challenges stem from major sequence differences between the sex chromosomes and unique structural variation absent in autosomes. For example, the human reference genome contains 22 haploid representations of autosomal chromosomes, but a diploid representation of two structurally divergent X and Y chromosomes (Figure 1). While this is appropriate for the non-recombining and diverged regions, the homologous PARs on the ends of the X and Y are represented twice with nearly 100% sequence identity in the state-of-the-art human genome assembly. If not adequately controlled for, this duplicated region will cause erroneous interpretation of output from short-read-based analyses, with reads mapping identically to multiple places, resulting in a map quality score of 0 when both PARs are present in the genome.[74]

In the human genome, these duplicated PARs represent a small amount of the total nuclear genome sequence (0.1%), likely limiting the global effects of potential biases.[8] However, the PARs are far larger in other systems (e.g., 0.7% of total nuclear sequence in *Canis lupus familiaris* and 11% in *Asparagus officinalis*).[18,75] Duplicated, meiotically homologous assemblies of these PARs could introduce major downstream analytical problems, including variant calling, gene and repeat annotation, and gene expression quantification. These issues would be compounded when using the same reference genome assembly representation (i.e., Chr01-22, X, Y, and mitochondria) for all individuals, whether they have a Y chromosome or not.

For the homogametic sex (i.e., XX individuals), and samples that have lost the Y chromosome (as sometimes occurs with aging[76]), a simple solution is to soft or hard mask the Y chromosome completely, thus prohibiting mapping to this reference, but keeping it within the index for downstream analyses. The development of this approach has shown vast improvements in analyses in humans.[74,77] In contrast, for samples with evidence of a Y chromosome, one approach is to soft or hard mask one copy of the PARs (typically on the Y chromosome) prior to downstream analyses.[74] However, *ad hoc* modification of traditional genome analysis pipelines is limited by the lack of a standard for reporting sex chromosome complement-specific reference sequences, and by lack of reporting of important boundary regions of the sex chromosomes for each genome build.

Other informatic issues exist with sex chromosomes where reference genomes contain a mixture of haploid and diploid representations of chromosomes. Any analysis step that uses coverage as a filter, as many variant callers do, will often apply the same read depth filter to the autosomes and sex chromosomes. However, genome coverage on the sex chromosomes in the heterogametic sex, for highly diverged regions, is expected to be approximately half that of autosomes, resulting in systematic biases in variant calling, though this effect has not been directly tested. While some tools focus specifically on analysis of the X chromosome in genome-wide association studies,[78] the sex chromosome pair is often removed from population genetic analyses,[79,80] which is problematic given the important role these genes have been shown to play in development and disease, among other traits.[49,76,81]

Box 1. Proposed suggestions for representing sex chromosomes in genome assemblies

- Report the sex of the genome isolate, and method of discovery (e.g., floral phenotyping or sex chromosome karyotype), or clearly state if unknown. Similarly, note if the species has sex chromosomes or if unknown.
- Generate a genome reference for the heterogametic sex chromosome pair. When possible, attempt phased diploid assembly of the heterogametic sex.
- The chromosome that contains the SDR/PARs should be labeled the sex chromosome pair (e.g., XY, not Chr19).
- Report the genomic location of the SDR and PAR(s) as metadata in the genome release.

THE NEAR FUTURE OF SEX CHROMOSOME REPRESENTATION

In order for downstream (post-assembly) informatics tools to accurately incorporate the sex chromosomes, there needs to be a set of standards for reporting sex chromosomes in a genome assembly that the tools can use as input. As diverse genome sequencing technologies converge on both long and accurate reads, highly contiguous sex chromosome pair assemblies will very soon become the norm. Before this deluge of oncoming genomes, we have several recommendations for how to approach genome assembly projects. Here we discuss different scenarios for presenting and releasing sex chromosome assemblies in the context of the latest genome sequencing and assembly techniques that accommodate the diversity of sex chromosomes in eukaryotes.

The goal of many large-scale genome projects is to provide a single, complete reference haplotype for a species. Ideally, the isolate used for genome sequencing should be of a known sex and this reported in the metadata and repositories in which the assembly is submitted (Box 1). For gonochoristic/dioecious species, publishing the genome sequence of an individual containing the homogametic sex chromosomes (i.e., ZZ or XX) can follow existing practices with reporting chromosomes, by numbering the autosomes and designating the X/Z chromosome. Targeting the homogametic sex also obviates many of the complications that we have discussed, such as the computational challenge of assembling highly diverged sex chromosome haplotypes. However, critically, the reference will not be adequate for ~50% of the individuals in the species (i.e., individuals carrying the Y or W) given the aforementioned immense variation in haplotype that can exist on an SDR. Therefore, it is our strong suggestion that the reference be an individual containing the heterogametic sex chromosome pair (i.e., ZW or XY).

There are several possibilities for representing sex chromosomes in genome assemblies within a heterogametic individual, each with a different set of pros and cons that must be considered (Figure 3; Table 2). Like the human genome, one option is to represent a single haplotype for the autosomes and the full length of both the Y/W and the X/Z chromosomes (Figure 3). A challenge with this approach is that the PAR needs to be demarcated, otherwise there will be two chromosomes with a

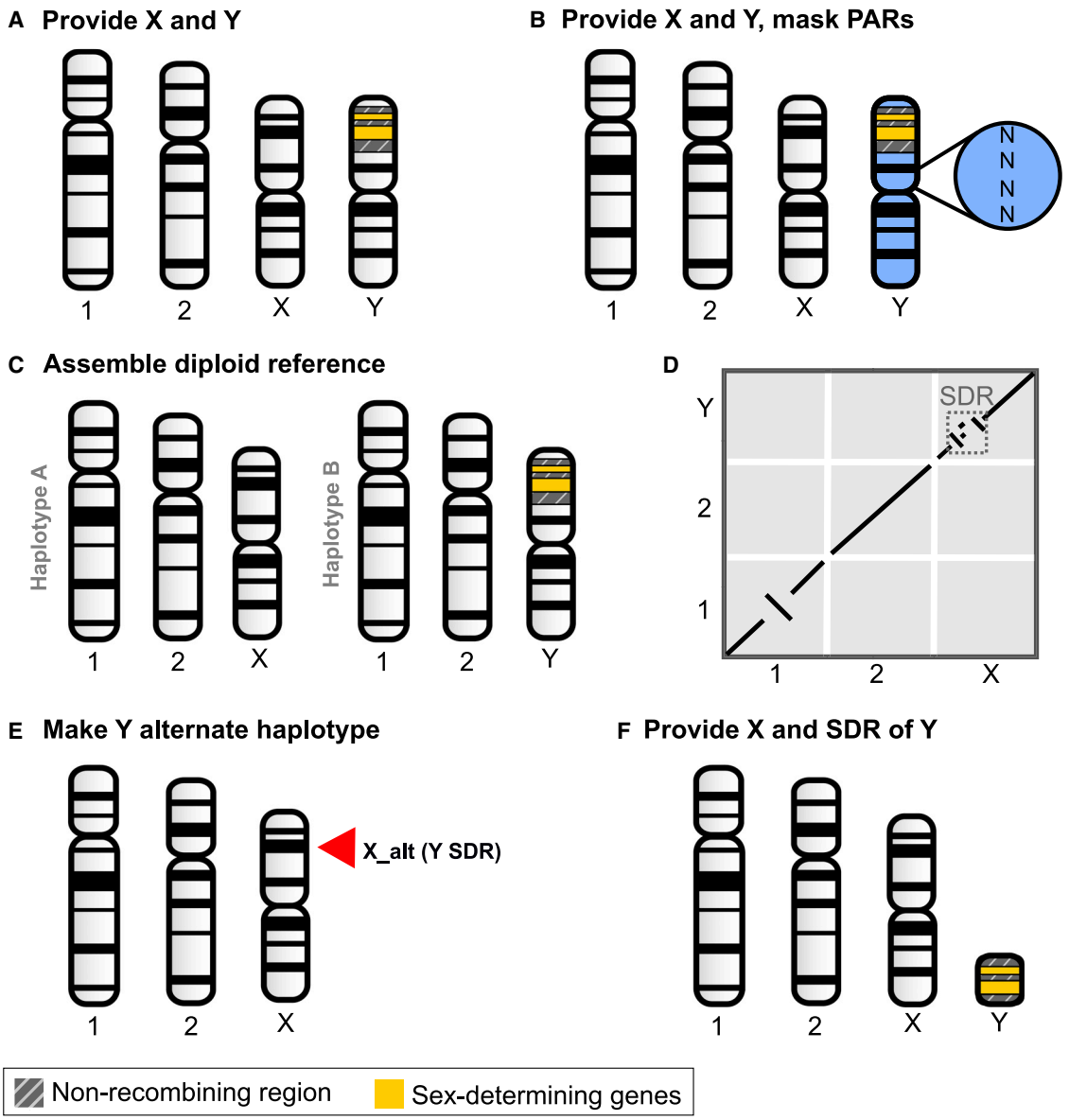

Figure 3. Solutions for representing sex chromosomes in genome assemblies

(A) In the genome release, one option is to provide the primary haplotype for the autosomes and both pairs of the sex chromosomes, like the human reference (see Figure 1).

(B) Because the PARs will be represented twice, causing issues with downstream analyses, a solution is to mask the PARs on the Y chromosome (in blue).

(C) Assembling both haplotypes is the best solution, because the entire genome would be represented twice.

(D) These first three approaches are ideal because the location of the SDR and structural variants are maintained. The hypothetical dot plot between two haplotypes highlights a large inversion on Chr01 and several structural variants in the SDR.

(E and F) If assembling the whole chromosome is not possible, (E) the Y SDR could instead be represented as an alternative haplotype of the X or (F) as a separate contig. There are pros and cons for each of these representations of sex chromosomes in the genome (Table 2), but is imperative regardless of the approach for the SDR and PAR boundaries to be reported in the genome release, so comparative analyses can be undertaken.

complement of meiotically homologous sequence that would severely complicate read mapping, protein mapping, and *ab initio* gene prediction and annotation. Although we recognize the PAR can sometimes be polymorphic within a species,[82,83] obscuring the demarcation of a single boundary, a highly informed boundary within the genome of the sequenced individual is vital. Similarly, representing the Y/W in full, but masking the PAR (i.e., hard mask by replacing sequence with "N" characters or soft mask by converting the sequence to lowercase) in the reference release, or accompanying it, would eliminate these double-mapping issues at the outset, but maintain the context of the SDR within the chromosome (Figure 3).

Table 2. Pros and cons in approaches for representing sex chromosomes in genome assemblies

Approaches for representing the genome	Pro	Con	Solution for cons
Provide both sex chromosomes in fasta reference, but only one copy of each autosome	Both sex chromosome haplotypes are available for mapping Context for each SDR represented	PARs are identical and represented twice Homologous regions in the SDR with low divergence will have mapping issues	Mask PARs Mask SDR for homogametic sex
Provide both sex chromosomes in fasta reference, but mask the PARs	Both sex chromosome haplotypes are available for mapping Context for each SDR represented, but only one PAR is available to map	Some SDRs are very small (<1% of the chromosome) and a chromosome composed nearly entirely of N's would increase computational burden (e.g., storage requirements), while providing other no additional genomic information within these masked regions SDR boundaries can be variable within a species	Maintain a version of genome assembly with and without masking in an accessible database
Provide contig of only SDR	SDR available for mapping	Context for location and structural variation for SDR is lost	Provide coordinates for the homologous region of the SDR
Provide sex-specific chromosome as an alternate haplotype	Genome is represented as haploid (except for any alternate haplotype contigs)	Context for location and structural variation for SDR is lost	Provide coordinates for the homologous region of the SDR
Provide diploid genome assembly	Autosomes and sex chromosomes both represented as diploid	Generating fully phased diploid references currently a challenge Many current analysis tools are not designed for diploid assemblies	Use trio-binning or Hi-C to aid in phasing

While haploid representations have been an integral first step in generating a reference genome, it is clear diploid representations, which contain homologous chromosome pairs for the entire genome, are better reflections of the genetic diversity that exists within a heterozygous individual.[84–87] Producing fully phased diploid representations of genomes, where every chromosome, both autosomes and sex chromosomes, would be represented as a homologous pair, would alleviate many of the bioinformatic complications of combining haploid and diploid chromosomal representations in a single assembly (Figure 3; Table 2). The recent advances in genome sequencing technology and analysis have unlocked the ability to produce phased diploid assemblies,[68,69] including the sex chromosome pair.[73] Further, publication of accurately phased, diploid assemblies would also aid comparative analyses of other non-recombining regions, such as large inversions on autosomal chromosomes and the *S*-locus in self-incompatible plants (Figure 3). However, the generation of phased diploid assemblies creates an additional problem: how should a reference genome that contains a sex chromosome pair be represented in a single fasta file? Phased genome assembly is still in its infancy, and since tools will continue to be built around the notion that phased assemblies will soon be commonplace, we propose that the most versatile path forward for representing sex chromosomes in genome assemblies is to preserve as much information as possible by publishing assemblies for each haplotype in full (Figure 3). In addition, we recommend providing genomic coordinates for the SDR/PAR in the release of these haplotype assemblies to aid in comparative analyses. This gives both the genome producer and users the ability to modify the reference genome to fit any number of bioinformatic scenarios of presenting the sex chromosome pair for a given analysis, such as hard masking PARs (Figure 3).

Despite these advancements in phased diploid assembly, we realize there are biological, technical, and financial realities that limit the ability to produce such references. For example, in species with long stretches of low heterozygosity, phasing maternal and paternal haplotype blocks without high-quality trio bins is still currently difficult, meaning only a single collapsed haplotype can be assembled.[68,88,89] To accommodate situations in which a fully phased diploid assembly is intractable, a different approach for haploid representations of the sex chromosomes is to represent the Y or W as an alternative haplotype of the X or Z in assemblies[90] (Figure 3; Table 2). This may be an especially well-suited option when the SDR is a relatively small fraction of the sex chromosome like in *A. officinalis*, *Morella rubra,* or *C. lupus familiaris.*[18,36,75] In cases where an alternative haplotype cannot be assembled, but the Y or W can still be assembled separately, a similar approach would be to append the contig(s) containing the Y/W SDR to the primary assembly containing the autosomes and X/Z. A notable issue with these alternatives is that all necessary genomic context between the X/Z and Y/W is lost, including the true size of the Y or W chromosome, major structural variations between sex chromosomes of a heterogametic genotype, and the absolute base pair location of the SDR on the hemizygous chromosome. If using these approaches, it is also necessary to provide metadata with the location of the SDR relative to the X or Z to recover these important contexts. While diploid assemblies may be the best path forward for genome references, representing the sex chromosomes as either an alternative haplotype or as a pair in an otherwise

haploid assembly, may be the most broadly applicable approach for most systems in which a fully phased diploid assembly is not feasible.

UV sex chromosomes present a unique set of obstacles. Because UV systems are haploid, where females have a U chromosome and males have a V^2 (Figure 2), both sex chromosomes are sex-specific and there is no heterogametic sex to target for a genome reference. To capture the diversity between the U and V chromosomes, a genome reference will need to be generated for both sexes. This is functionally analogous to generating a phased diploid assembly, though perhaps easier to accomplish given a haploid individual only contains a single haplotype. This makes representing the individual references straightforward, by labeling the autosomes and sex chromosomes within each assembly respectively. Although, similar to diploid systems, UVs are expected to have PARs that should be demarcated on both for downstream analyses. An analogous approach can be extended to mating-type loci found in many algae and fungi.

Because of the diversity of sex chromosomes that we have described, and others yet to be discovered, it is likely no one of these options will fit all scenarios. Regardless, moving toward a form of consistency is imperative, such that comparisons can easily be made across different species. This starts with unfailingly noting the sex of the genome reference, whether sex chromosomes are known in the species, and clearly noting contigs and coordinates for PARs and SDRs as part of the genome release and associated metadata (e.g., within a README file) (Box 1).

FUTURE PROSPECTS OF STUDYING SEX CHROMOSOMES

There are practical outcomes of assembling and properly representing diverse eukaryotic sex chromosomes. This includes the identification of genes and variants that are linked to sex-specific development, disease, breeding, and evolution. A consistent set of genome assembly representation standards that takes into account the unique biology of the species, as well as the quality and type of data available, will enable a powerful comparative framework to explore the veritable smorgasbord of sex chromosome evolution, function, and diversification.

ACKNOWLEDGMENTS

Funding was provided by NSF-IOS #2128196 to A.H. and USDA-NIFA #2022-67012-36818 to S.B.C. The work conducted by the US Department of Energy Joint Genome Institute is supported by the Office of Science of the US Department of Energy under Contract No. DE-AC02-05CH11231. This publication was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award Number R35GM124827 to M.A.W.

AUTHOR CONTRIBUTIONS

All authors conceived, wrote, and edited the final manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

1. Stevens, N.M. (1905). Studies in Spermatogenesis (Carnegie Institution of Washington).
2. Bachtrog, D., Mank, J.E., Peichel, C.L., Kirkpatrick, M., Otto, S.P., Ashman, T.-L., Hahn, M.W., Kitano, J., Mayrose, I., Ming, R., et al. (2014). Sex determination: why so many ways of doing it? PLoS Biol. *12*, e1001899.
3. The Tree of Sex Consortium (2014). Tree of sex: a database of sexual systems. Sci. Data *1*, 140015.
4. Venter, J.C., Adams, M.D., Myers, E.W., Li, P.W., Mural, R.J., Sutton, G.G., Smith, H.O., Yandell, M., Evans, C.A., Holt, R.A., et al. (2001). The sequence of the human genome. Science *291*, 1304–1351.
5. Ross, M.T., Grafham, D.V., Coffey, A.J., Scherer, S., McLay, K., Muzny, D., Platzer, M., Howell, G.R., Burrows, C., Bird, C.P., et al. (2005). The DNA sequence of the human X chromosome. Nature *434*, 325.
6. Miga, K.H., Koren, S., Rhie, A., Vollger, M.R., Gershman, A., Bzikadze, A., Brooks, S., Howe, E., Porubsky, D., Logsdon, G.A., et al. (2020). Telomere-to-telomere assembly of a complete human X chromosome. Nature *585*, 79–84.
7. Nurk, S., Koren, S., Rhie, A., Rautiainen, M., Bzikadze, A.V., Mikheenko, A., Vollger, M.R., Altemose, N., Uralsky, L., Gershman, A., et al. (2021). The complete sequence of a human genome. Preprint at bioRxiv. https://doi.org/10.1101/2021.05.26.445798.
8. Skaletsky, H., Kuroda-Kawaguchi, T., Minx, P.J., Cordum, H.S., Hillier, L., Brown, L.G., 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.
9. Jain, M., Olsen, H.E., Turner, D.J., Stoddart, D., Bulazel, K.V., Paten, B., Haussler, D., Willard, H.F., Akeson, M., and Miga, K.H. (2018). Linear assembly of a human centromere on the Y chromosome. Nat. Biotechnol. *36*, 321–323.
10. Carey, S., Yu, Q., and Harkess, A. (2021). The diversity of plant sex chromosomes highlighted through advances in genome sequencing. Genes *12*, 381.
11. Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., Chow, W., Fungtammasan, A., Kim, J., et al. (2021). Towards complete and error-free genome assemblies of all vertebrate species. Nature *592*, 737–746.
12. Jarne, P., and Auld, J.R. (2006). Animals mix it up too: the distribution of self-fertilization among hermaphroditic animals. Evolution *60*, 1816–1824.
13. Renner, S.S. (2014). The relative and absolute frequencies of angiosperm sexual systems: dioecy, monoecy, gynodioecy, and an updated online database. Am. J. Bot. *101*, 1588–1596.
14. Cheng, S., Melkonian, M., Smith, S.A., Brockington, S., Archibald, J.M., Delaux, P.-M., Li, F.-W., Melkonian, B., Mavrodiev, E.V., Sun, W., et al. (2018). 10KP: a phylodiverse genome sequencing plan. Gigascience *7*, 1–9.
15. Lewin, H.A., Robinson, G.E., Kress, W.J., Baker, W.J., Coddington, J., Crandall, K.A., Durbin, R., Edwards, S.V., Forest, F., Gilbert, M.T.P., et al. (2018). Earth BioGenome project: sequencing life for the future of life. Proc. Natl. Acad. Sci. U S A *115*, 4325–4333.
16. Voolstra, C.R., GIGA community of scientists (COS); Wörheide, G., and Lopez, J.V. (2017). advancing genomics through the global invertebrate genomics alliance (GIGA). Invertebr. Syst. *31*, 1–7.
17. Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., et al. (2012). A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, Takifugu rubripes (fugu). PLoS Genet. *8*, e1002798.
18. Harkess, A., Zhou, J., Xu, C., Bowers, J.E., Van der Hulst, R., Ayyampalayam, S., Mercati, F., Riccardi, P., McKain, M.R., Kakrana, A., et al. (2017). The asparagus genome sheds light on the origin and evolution of a young Y chromosome. Nat. Commun. *8*, 1279.

19. Liu, Z., Moore, P.H., Ma, H., Ackerman, C.M., Ragiba, M., Yu, Q., Pearl, H.M., Kim, M.S., Charlton, J.W., Stiles, J.I., et al. (2004). A primitive Y chromosome in papaya marks incipient sex chromosome evolution. Nature *427*, 348–352.

20. Alföldi, J., Di Palma, F., Grabherr, M., Williams, C., Kong, L., Mauceli, E., Russell, P., Lowe, C.B., Glor, R.E., Jaffe, J.D., et al. (2011). The genome of the green anole lizard and a comparative analysis with birds and mammals. Nature *477*, 587–591.

21. Rozen, S., Skaletsky, H., Marszalek, J.D., Minx, P.J., Cordum, H.S., Waterston, R.H., Wilson, R.K., and Page, D.C. (2003). Abundant gene conversion between arms of palindromes in human and ape Y chromosomes. Nature *423*, 873–876.

22. Winge, Ö. (1929). On the nature of the sex chromosomes in Humulus. Hereditas *12*, 53–63.

23. Westergaard, M. (1946). Aberrant Y chromosomes and sex expression in Melandrium album. Hereditas *32*, 419–443.

24. Kihara, H. (1928). On the chromosomes of Humulus japonicus. Bot. Mag. Tokyo *42*, 237–238.

25. Veyrunes, F., Waters, P.D., Miethke, P., Rens, W., McMillan, D., Alsop, A.E., Grützner, F., Deakin, J.E., Whittington, C.M., Schatzkamer, K., et al. (2008). Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. Genome Res. *18*, 965–973.

26. Gazoni, T., Haddad, C.F.B., Narimatsu, H., Cabral-de-Mello, D.C., Lyra, M.L., and Parise-Maltempi, P.P. (2018). More sex chromosomes than autosomes in the Amazonian frog Leptodactylus pentadactylus. Chromosoma *127*, 269–278.

27. Kobayashi, T., Yamada, F., Hashimoto, T., Abe, S., Matsuda, Y., and Kuroiwa, A. (2007). Exceptional minute sex-specific region in the X0 mammal, Ryukyu spiny rat. Chromosome Res. *15*, 175–187.

28. Hodgkin, J. (1987). Primary sex determination in the nematode C. elegans. Development *101* (*Suppl*), 5–16.

29. Král, J. (2007). Evolution of multiple sex chromosomes in the spider genus Malthonica (Araneae: agelenidae) indicates unique structure of the spider sex chromosome systems. Chromosome Res. *15*, 863–879.

30. Smith, B.W. (1964). The evolving karyotype of rumex hastatulus. Evolution *18*, 93–104.

31. Zhou, Q., Wang, J., Huang, L., Nie, W., Wang, J., Liu, Y., Zhao, X., Yang, F., and Wang, W. (2008). Neo-sex chromosomes in the black muntjac recapitulate incipient evolution of mammalian sex chromosomes. Genome Biol. *9*, R98.

32. Bachtrog, D., and Charlesworth, B. (2002). Reduced adaptation of a non-recombining neo-Y chromosome. Nature *416*, 323–326.

33. Spigler, R.B., Lewers, K.S., Main, D.S., and Ashman, T.-L. (2008). Genetic mapping of sex determination in a wild strawberry, Fragaria virginiana, reveals earliest form of sex chromosome. Heredity *101*, 507–517.

34. Ellegren, H. (2000). Evolution of the avian sex chromosomes and their role in sex determination. Trends Ecol. Evol. *15*, 188–192.

35. Ohno, S. (2013). Sex Chromosomes and Sex-Linked Genes (Springer Science & Business Media).

36. Jia, H.-M., Jia, H.-J., Cai, Q.-L., Wang, Y., Zhao, H.-B., Yang, W.-F., Wang, G.-Y., Li, Y.-H., Zhan, D.-L., Shen, Y.-T., et al. (2019). The red bayberry genome and genetic basis of sex determination. Plant Biotechnol. J. *17*, 397–409.

37. Hirst, C.E., Major, A.T., and Smith, C.A. (2018). Sex determination and gonadal sex differentiation in the chicken model. Int. J. Dev. Biol. *62*, 153–166.

38. Traut, W., and Marec, F. (1997). Sex chromosome differentiation in some species of Lepidoptera (Insecta). Chromosome Res. *5*, 283–291.

39. Green, D.M., Zeyl, C.W., and Sharbel, T.F. (1993). The evolution of hypervariable sex and supernumerary (B) chromosomes in the relict New Zealand frog, Leiopelma hochstetteri. J. Evol. Biol. *6*, 417–441.

40. Filho, O.M., Bertollo, L.A.C., and Junior, P.M.G. (1980). Evidences for a multiple sex chromosome system with female heterogamety in apareiodon affinis (pisces, parodontidae). Caryologia *33*, 83–91.

41. de Oliveira, R.R., Feldberg, E., dos Anjos, M.B., and Zuanon, J. (2008). Occurrence of multiple sexual chromosomes (XX/XY1Y2 and Z1Z1Z2Z2/Z1Z2W1W2) in catfishes of the genus Ancistrus (Siluriformes: loricariidae) from the Amazon basin. Genetica *134*, 243–249.

42. Šíchová, J., Ohno, M., Dincă, V., Watanabe, M., Sahara, K., and Marec, F. (2016). Fissions, fusions, and translocations shaped the karyotype and multiple sex chromosome constitution of the northeast-Asian wood white butterfly, Leptidea amurensis. Biol. J. Linn. Soc. Lond. *118*, 457–471.

43. Roco, Á.S., Olmstead, A.W., and Degitz, S.J. (2015). Coexistence of Y, W, and Z sex chromosomes in Xenopus tropicalis. Proc. Natl. Acad. Sci. *112*, E4752–E4761.

44. Roberts, N.B., Juntti, S.A., Coyle, K.P., Dumont, B.L., Stanley, M.K., Ryan, A.Q., Fernald, R.D., and Roberts, R.B. (2016). Polygenic sex determination in the cichlid fish Astatotilapia burtoni. BMC Genomics *17*, 835.

45. Carey, S.B., Jenkins, J., Lovell, J.T., Maumus, F., Sreedasyam, A., Payton, A.C., Shu, S., Tiley, G.P., Fernandez-Pozo, N., Healey, A., et al. (2021). Gene-rich UV sex chromosomes harbor conserved regulators of sexual development. Sci. Adv. *7*, eabh2488.

46. Yamato, K.T., Ishizaki, K., Fujisawa, M., Okada, S., Nakayama, S., Fujishita, M., Bando, H., Yodoya, K., Hayashi, K., Bando, T., et al. (2007). Gene organization of the liverwort Y chromosome reveals distinct sex chromosome evolution in a haploid system. Proc. Natl. Acad. Sci. U S A *104*, 6472–6477.

47. Allen, C.E. (1917). A chromosome difference correlated with sex differences in SPHAeROCARPOS. Science *46*, 466–467.

48. Sousa, A., Schubert, V., and Renner, S.S. (2020). Centromere organization and UU/V sex chromosome behavior in a liverwort. Plant J. https://doi.org/10.1111/tpj.15150.

49. Renner, S.S., and Müller, N.A. (2021). Plant sex chromosomes defy evolutionary models of expanding recombination suppression and genetic degeneration. Nat. Plants *7*, 392–402.

50. Tennessen, J.A., Wei, N., Straub, S.C.K., Govindarajulu, R., Liston, A., and Ashman, T.-L. (2018). Repeated translocation of a gene cassette drives sex-chromosome turnover in strawberries. Plos Biol. *16*, e2006062.

51. Yu, Q., Hou, S., Hobza, R., Feltus, F.A., Wang, X., Jin, W., Skelton, R.L., Blas, A., Lemke, C., Saw, J.H., et al. (2007). Chromosomal location and gene paucity of the male specific region on papaya Y chromosome. Mol. Genet. Genomics *278*, 177–185.

52. Wang, J., Na, J.-K., Yu, Q., Gschwend, A.R., Han, J., Zeng, F., Aryal, R., VanBuren, R., Murray, J.E., Zhang, W., et al. (2012). Sequencing papaya X and Yh chromosomes reveals molecular basis of incipient sex chromosome evolution. Proc. Natl. Acad. Sci. U S A *109*, 13710–13715.

53. Peichel, C.L., McCann, S.R., Ross, J.A., Naftaly, A.F.S., Urton, J.R., Cech, J.N., Grimwood, J., Schmutz, J., Myers, R.M., Kingsley, D.M., et al. (2020). Assembly of the threespine stickleback Y chromosome reveals convergent signatures of sex chromosome evolution. Genome Biol. *21*, 177.

54. Lemaitre, C., Braga, M.D.V., Gautier, C., Sagot, M.-F., Tannier, E., and Marais, G.A.B. (2009). Footprints of inversions at present and past pseudoautosomal boundaries in human sex chromosomes. Genome Biol. Evol. *1*, 56–66.

55. Yang, W., Wang, D., Li, Y., Zhang, Z., Tong, S., Li, M., Zhang, X., Zhang, L., Ren, L., Ma, X., et al. (2020). A general model to explain repeated turnovers of sex determination in the Salicaceae. Preprint at bioRxiv. https://doi.org/10.1101/2020.04.11.037556.

56. Harkess, A., Huang, K., van der Hulst, R., Tissen, B., Caplan, J.L., Koppula, A., Batish, M., Meyers, B.C., and Leebens-Mack, J. (2020). Sex determination by two Y-linked genes in garden Asparagus. Plant Cell *32*, 1790–1796.

57. Almeida, P., Proux-Wera, E., Churcher, A., Soler, L., Dainat, J., Pucholt, P., Nordlund, J., Martin, T., Rönnberg-Wästljung, A.-C., Nystedt, B., et al.

(2020). Genome assembly of the basket willow, Salix viminalis, reveals earliest stages of sex chromosome expansion. BMC Biol. *18*, 78.

58. Meisel, R.P., Delclos, P.J., and Wexler, J.R. (2019). The X chromosome of the German cockroach, Blattella germanica, is homologous to a fly X chromosome despite 400 million years divergence. BMC Biol. *17*, 100.
59. Xu, L., Auer, G., Peona, V., Suh, A., Deng, Y., Feng, S., Zhang, G., Blom, M.P.K., Christidis, L., Prost, S., et al. (2019). Dynamic evolutionary history and gene content of sex chromosomes across diverse songbirds. Nat. Ecol. Evol. *3*, 834–844.
60. Vicoso, B. (2019). Molecular and evolutionary dynamics of animal sex-chromosome turnover. Nat. Ecol. Evol. *3*, 1632–1641.
61. Bachtrog, D. (2013). Y-chromosome evolution: emerging insights into processes of Y-chromosome degeneration. Nat. Rev. Genet. *14*, 113–124.
62. Wilson Sayres, M.A., and Makova, K.D. (2013). Gene survival and death on the human Y chromosome. Mol. Biol. Evol. *30*, 781–787.
63. Bergero, R., and Charlesworth, D. (2011). Preservation of the Y transcriptome in a 10-million-year-old plant sex chromosome system. Curr. Biol. *21*, 1470–1474.
64. Krasovec, M., Chester, M., Ridout, K., and Filatov, D.A. (2018). The mutation rate and the age of the sex chromosomes in Silene latifolia. Curr. Biol. *28*, 1832–1838.e4.
65. Furman, B.L.S., Metzger, D.C.H., Darolti, I., Wright, A.E., Sandkam, B.A., Almeida, P., Shu, J.J., and Mank, J.E. (2020). Sex chromosome evolution: so many exceptions to the rules. Genome Biol. Evol. *12*, 750–763.
66. Almeida, P., Sandkam, B.A., Morris, J., Darolti, I., Breden, F., and Mank, J.E. (2021). Divergence and remarkable diversity of the Y chromosome in guppies. Mol. Biol. Evol. *38*, 619–633.
67. Wenger, A.M., Peluso, P., Rowell, W.J., Chang, P.-C., Hall, R.J., Concepcion, G.T., Ebler, J., Fungtammasan, A., Kolesnikov, A., Olson, N.D., et al. (2019). Accurate circular consensus long-read sequencing improves variant detection and assembly of a human genome. Nat. Biotechnol. *37*, 1155–1162.
68. Cheng, H., Concepcion, G.T., Feng, X., Zhang, H., and Li, H. (2021). Haplotype-resolved de novo assembly using phased assembly graphs with hifiasm. Nat. Methods *18*, 170–175.
69. Nurk, S., Walenz, B.P., Rhie, A., Vollger, M.R., Logsdon, G.A., Grothe, R., Miga, K.H., Eichler, E.E., Phillippy, A.M., and Koren, S. (2020). HiCanu: accurate assembly of segmental duplications, satellites, and allelic variants from high-fidelity long reads. Genome Res. *30*, 1291–1305.
70. Kolmogorov, M., Yuan, J., Lin, Y., and Pevzner, P.A. (2019). Assembly of long, error-prone reads using repeat graphs. Nat. Biotechnol. *37*, 540–546.
71. Muyle, A., Käfer, J., Zemp, N., Mousset, S., Picard, F., and Marais, G.A. (2016). SEX-DETector: a probabilistic approach to study sex chromosomes in non-model organisms. Genome Biol. Evol. *8*, 2530–2543.
72. Akagi, T., Henry, I.M., Tao, R., and Comai, L. (2014). A Y-chromosome–encoded small RNA acts as a sex determinant in persimmons. Science *346*, 646–650.
73. Xue, L., Gao, Y., Wu, M., Tian, T., Fan, H., Huang, Y., Huang, Z., Li, D., and Xu, L. (2021). Telomere-to-telomere assembly of a fish Y chromosome reveals the origin of a young sex chromosome pair. Genome Biol. *22*, 203.
74. Webster, T.H., Couse, M., Grande, B.M., Karlins, E., Phung, T.N., Richmond, P.A., Whitford, W., and Wilson, M.A. (2019). Identifying, understanding, and correcting technical artifacts on the sex chromosomes in next-generation sequencing data. Gigascience *8*, giz074.
75. Li, G., Davis, B.W., Raudsepp, T., Pearks Wilkerson, A.J., Mason, V.C., Ferguson-Smith, M., O'Brien, P.C., Waters, P.D., and Murphy, W.J. (2013). Comparative analysis of mammalian Y chromosomes illuminates ancestral structure and lineage-specific evolution. Genome Res. *23*, 1486–1495.
76. Wilson, M.A. (2021). The Y chromosome and its impact on health and disease. Hum. Mol. Genet. *30*, R296–R300.
77. Olney, K.C., Brotman, S.M., Andrews, J.P., Valverde-Vesling, V.A., and Wilson, M.A. (2020). Reference genome and transcriptome informed by the sex chromosome complement of the sample increase ability to detect sex differences in gene expression from RNA-Seq data. Biol. Sex Differ. *11*, 42.
78. Gao, F., Chang, D., Biddanda, A., Ma, L., Guo, Y., Zhou, Z., and Keinan, A. (2015). XWAS: a software toolset for genetic data analysis and association studies of the X chromosome. J. Hered. *106*, 666–671.
79. Accounting for sex in the genome. Nat. Med. *23*, 1243.
80. Wise, A.L., Gyi, L., and Manolio, T.A. (2013). eXclusion: toward integrating the X chromosome in genome-wide association analyses. Am. J. Hum. Genet. *92*, 643–647.
81. Hughes, J.F., and Page, D.C. (2015). The biology and evolution of mammalian Y chromosomes. Annu. Rev. Genet. *49*, 507–527.
82. Lappin, F.M., Medert, C.M., Hawkins, K.K., Mardonovich, S., Wu, M., and Moore, R.C. (2015). A polymorphic pseudoautosomal boundary in the Carica papaya sex chromosomes. Mol. Genet. Genomics *290*, 1511–1522.
83. Cotter, D.J., Brotman, S.M., and Wilson Sayres, M.A. (2016). Genetic diversity on the human X chromosome does not support a strict pseudoautosomal boundary. Genetics *203*, 485–492.
84. Zhang, X., Wu, R., Wang, Y., Yu, J., and Tang, H. (2020). Unzipping haplotypes in diploid and polyploid genomes. Comput. Struct. Biotechnol. J. *18*, 66–72.
85. Lovell, J.T., Bentley, N.B., Bhattarai, G., Jenkins, J.W., Sreedasyam, A., Alarcon, Y., Bock, C., Boston, L.B., Carlson, J., Cervantes, K., et al. (2021). Four chromosome scale genomes and a pan-genome annotation to accelerate pecan tree breeding. Nat. Commun. *12*, 1–12.
86. Zhou, Q., Tang, D., Huang, W., Yang, Z., Zhang, Y., Hamilton, J.P., Visser, R.G.F., Bachem, C.W.B., Robin Buell, C., Zhang, Z., et al. (2020). Haplotype-resolved genome analyses of a heterozygous diploid potato. Nat. Genet. *52*, 1018–1023.
87. Ebert, P., Audano, P.A., Zhu, Q., Rodriguez-Martin, B., Porubsky, D., Bonder, M.J., Sulovari, A., Ebler, J., Zhou, W., Serra Mari, R., et al. (2021). Haplotype-resolved diverse human genomes and integrated analysis of structural variation. Science *372*. https://doi.org/10.1126/science.abf7117.
88. Kronenberg, Z.N., Rhie, A., Koren, S., Concepcion, G.T., Peluso, P., Munson, K.M., Porubsky, D., Kuhn, K., Mueller, K.A., Low, W.Y., et al. (2021). Extended haplotype-phasing of long-read de novo genome assemblies using Hi-C. Nat. Commun. *12*, 1935.
89. Rhie, A., Walenz, B.P., Koren, S., and Phillippy, A.M. (2020). Merqury: reference-free quality, completeness, and phasing assessment for genome assemblies. Genome Biol. *21*, 1–27.
90. Church, D.M., Schneider, V.A., Steinberg, K.M., Schatz, M.C., Quinlan, A.R., Chin, C.-S., Kitts, P.A., Aken, B., Marth, G.T., Hoffman, M.M., et al. (2015). Extending reference assembly models. Genome Biol. *16*, 13.