# PhaseWY: A pipeline for haplotype phasing, sex chromosome identification, and extraction of W sequences

Contents

[PhaseWY: A pipeline for haplotype phasing, sex chromosome identification, and extraction of W sequences 1](#_Toc212481545)

[Introduction 2](#_Toc212481546)

[Step 1: Identify callable regions 3](#_Toc212481547)

[Step 2: Identify sex-linked regions based on sex differences in sequence depth 4](#_Toc212481548)

[Step 3: Extract and prepare scaffold specific variants for phasing 4](#_Toc212481549)

[Step 4: Read-based phasing 4](#_Toc212481550)

[Step 5: Statistical phasing 4](#_Toc212481551)

[Step 6. Determine sex-linkage and chromosome type of each haplotype 5](#_Toc212481552)

[Step 7. Re-organise genotypes according to genomic region 6](#_Toc212481553)

[Step 8. Summarise genome with bed-files 7](#_Toc212481554)

[Step 9. Create final vcf-files 8](#_Toc212481555)

[Step 10. Produce pipeline statistics 10](#_Toc212481556)

[5. References 10](#_Toc212481557)

## Introduction

We developed a bioinformatic pipeline – *PhaseWY* v.2024-01-18 – for phasing homologous sex chromosome sequences (<https://github.com/sjellerstrand/PhaseWY>). By aligning read data of several female and male individuals to a homogametic reference genome the pipeline classifies genomic regions as autosomal or sex-linked and extracts female Y/W sequences. *PhaseWY* operates scaffold-wise, running independently on each scaffold through several key steps. We applied the pipeline to two datasets including the Skylark and the Raso lark with a ZW sex determination system (Ellestrand & Hansson; unpublished), which are referenced to throughout this documentation.

Throughout the pipeline, we use a developmental version of *BCFtools* (version 1.17, github commit b7b2a32). This is motivated by the need to accurately handling multiallelic sites during phasing. Both read-based phasing and statistical phasing are only compatible with biallelic sites. Therefore, multiallelic sites in the input-vcf must first be split into biallelic records, which are then reconstructed after phasing. In earlier versions of *BCFtools*, the *norm* command alters phased haplotypes during the reconstruction of multiallelic variants (see <https://github.com/samtools/bcftools/issues/1893>), which is why we primarily use this developmental version. However, in any occurrence of vcf-files being merged, the pipeline temporarily switches to the release version 1.17 due to an introduced error in the otherwise used developmental version (see <https://github.com/samtools/bcftools/issues/1954>).

A diagram of a company

Description automatically generated

Figure . Flowchart of the phasing steps of the PhaseWY pipeline. Blue squares represent pipeline steps. Purple rounded squares represent files that are generated and outputted at various pipeline steps. The connections between pipeline steps and files are indicated by arrows. Hearts represent steps and files that are specific to sex chromosome phasing. Honeycombs represent steps and files that are specific to phasing combining diploids with haploids as reference sets.

### Step 1: Identify callable regions

While the standard variant calling format (vcf) is storage-efficient, it does not indicate whether non-variant regions are callable and monomorphic, or whether they are non-callable and thus missing data. To address this, each bam-file originally used for variant calling is reprocessed to retrieve site-wise depth using *SAMtools* v1.17. Sites are excluded if they fail to meet threshold criteria based on previous variant filtering: minimum depth, mean depths (minimum and maximum) and proportion of missing data. *BEDTools* v.2.31.0 (Quinlan and Hall, 2010) is then used to merge the callable sites into coherent regions as a bed-file. Repeat regions previously identified are subtracted using *BEDTools* v.2.31.0 to yield a final bed-file of callable regions. Scaffolds lacking callable regions are excluded (this primarily affects short scaffolds in fragmented reference genomes).

### Step 2: Identify sex-linked regions based on sex differences in sequence depth

In sex-linked regions where the W chromosome a highly degenerated, W sequences may fail to align to the homologous Z reference (Zhou et al., 2014). Consequently, in such regions, female (ZW) alignment depth is expected to be about half the depth across non-sex-linked regions. To quantify this, sequencing depth at all sites within callable regions is calculated for each bam-file using *SAMtools* v1.17. Depth values are then standardized within individual by dividing the depth at each site by the mean depth of the individual. A custom R script is used to calculate the average standardized depth per site separately for females and males, yielding within-sex standardized depths profiles. The relative sex difference in depth is calculated by dividing the female standardized depth with the male standardized depth at each site. Sites with a relative depth ratio below 0.75 are classified as sex-linked and merged into coherent regions as a bed-file with *BEDTools* v.2.31.0.

### Step 3: Extract and prepare scaffold specific variants for phasing

*BCFtools* v.1.17.b7b2a32 is used to extract variants from each specified scaffold, remove the FORMAT fields, and split multiallelic sites into biallelic records to ensure compatibility with downstream phasing software. *VCFtools* v.0.1.16 is then used to verify that the scaffold contains at least one biallelic sites with more than one heterozygous genotype. If this condition is not met (which can occur on small scaffolds in fragmented reference genomes), the scaffold is excluded from further analyses, as statistically phasing cannot be performed.

### Step 4: Read-based phasing

*BCFtools* v.1.17.b7b2a32 is used to subset the previous vcf-file into individual vcf-files (one per samples). *WhatsHap* v.1.7 (Martin et al., 2016) is then employed for read-based phasing. Each individual bam-file - originally used for variant calling - is used to identify sequencing read pairs spanning at least two heterozygous variants, resulting in phase-sets spanning as many variants as can be physically linked on the same scaffold. These phase-sets are subsequently used to inform statistical phasing. Read-based phasing can improve phasing accuracy in small datasets and can manage to phase singleton alleles when they are part of phase-sets spanning several variants. Only variants present in the input vcf-file are evaluated. The consensus reference genome is also provided to *WhatsHap* to enable its realignment detection algorithm. Finally, *BCFtools* v.1.17 is used to merge all phased individual vcf-files into a single, combined vcf-file.

### Step 5: Statistical phasing

Statistical phasing is performed with *SHAPEIT* v.4.1.3 (Delaneau et al., 2019), incorporating the phase-sets from *WhatsHap*. The recommended expected error rate of 0.01 % and default parameters for sequencing data are used. Notice that *SHAPEIT* also imputes missing genotypes based on haplotype information, resulting in full genotypes in the output. We consider imputation acceptable when the proportion of missing data is low. In our study, no datasets exceeded 5% missing data, which corresponds to at most one out of 18 missing genotypes.

Following phasing, *BCFtools* v.1.17.b7b2a32 is used to reconstruct multiallelic variants. The allele count field (AC) is updated with *vcffixup* from *vcflib* v.1.0.3, and *vcffilter* from *vcflib* v.1.0.3 is used to remove sites that became monomorphic for the reference allele. In rare cases, genotype formatting may change from phased “|” to non-phased “/” during multiallelic reconstruction, and such changes are reset to a phased state to ensure compatibility during following steps.

### Step 6. Determine sex-linkage and chromosome type of each haplotype

If haplotype phasing is successful, haplotypes are expected to cluster into Z- and W-haplotypes, provided sufficient evolutionary divergence has occurred in sex-linked regions. Based on the number of males and females in the dataset, there is a predefined expectation on the number of haplotypes in the Z-haplotype cluster (2\**Nmales* + 1\**Nfemales*) and the W-haplotype cluster (1\**Nfemales*), respectively. This composition allows inferring of whether a region is sex-linked or autosomal. Furthermore, if a region is inferred to be sex-linked, the corresponding W-haplotypes of individual females can also be inferred.

Only a subset of variants is used to infer these regions, as low-frequency alleles provide limited information for clustering haplotypes. In this study, we adopted a conservative approach by excluding singleton alleles. However, an upper frequency threshold can be specified based on the expected frequency of W chromosomes in the dataset. Any excluded variant is subsequently associated with its corresponding haplotype. This filtering step proved particularly important in the case of the Skylark, whose high genetic diversity with many singletons’ obscure signals of sex-linkage, especially in the younger regions of sex chromosomes.

*VCFtools* v.0.1.16 and *vcffilter* from *vcflib* v.1.0.3 is used to retain biallelic alleles above a certain minor allele frequency. A custom R-script is used to determine sex-linkage and chromosome type of each haplotype. The R-package *vcfR* v.1.10.0 (Knaus and Grünwald, 2017) is used to import the phased and filtered vcf-file. Each individual haplotype along the scaffold is split into the “left” and “right” side of the genotype pipe (“|”). Because phasing can introduce switch-errors, sex-linkage is evaluated in sliding windows across the genome. In this study, we used a window size of 1,000 bp and a step of 250 bp. The R-package *data.table* v.1.12.8 (<https://Rdatatable.gitlab.io/data.table>) is used to efficiently extract all variant sites within each window. Within each window, a distance matrix is calculated between all haplotypes based on the variants present. If at least one variant is present, haplotypes are divided into two clusters using k-means (Hartigan and Wong, 1979), applying 10 random sets. Since there are always fewer W than Z chromosomes when both sexes are present, the smallest cluster is inferred as containing putative W haplotypes if certain criteria are met; otherwise, the window is classified as autosomal. The criteria, evaluated in order, are: (i) only female haplotypes are present in the smallest cluster, (ii) only one W haplotype per female is allowed, and (iii) the smallest cluster corresponds to the expected number of W haplotypes. If a window is determined to be sex-linked, the W haplotype for each female is inferred as the haplotype present in the smallest cluster. If a continuous series of sex-linked regions contain a switch in W haplotype for any female, this is likely due to a phasing switch-error, and such occurrences are noted.

Once the entire scaffold has been evaluated, consecutive autosomal or sex-linked windows are merged into coherent regions. To ensure a conservative classification, some regions are designated as “unknown genome type” at this stage. This occurs when an autosomal region borders a sex-linked window (or vice versa), or when a switch in W haplotype is detected for any female within a series of uninterrupted sex-linked windows. Because a sliding window approach is applied, all windows overlapping the region of concern are classified as unknown.

Per-window information is output alongside a bed-file summarising the coherent regions. Additionally, two bed-files are generated for each female, specifying the regions were their “left” and “right” haplotypes, respectively, are inferred as W haplotypes. Notice that for any scaffold containing at least one sex-linked region, autosomal regions are reported as “pseudoautosomal”. Due to re-classification of bordering regions as “unknown”, a scaffold can contain pseudoautosomal and unknown regions in the final output, without any coherent region classified as sex-linked. However, the window-specific output retains all underlying information.

Finally, some variants within sex-linked regions may be considered problematic or unreliable. For example, if females carry homozygotic genotypes for both alleles at a site within a sex-linked region, this may indicate that the site represents a Z-linked polymorphism, with no W allele represented at the site (see Step 2). Alternatively, such a pattern could reflect incomplete lineage sorting between Z and W, though this is less likely if they diverged a long time ago. *BCFtools* v.1.17.b7b2a32 is used to retain only females and all variants within sex-linked regions, and *vcffilter* from *vcflib* v.1.0.3 is applied to retain only polymorphic sites. Genotype frequencies are retrieved using *VCFtools* v.0.1.16, and any site where both alleles occur as homozygotes are recorded in a bed-file. These sites are flagged but not removed from the vcf-file at this stage.

### Step 7. Re-organise genotypes according to genomic region

Autosomal regions are retained without modification. However, if a scaffold contains at least one sex-linked region - identified either by sex differences in alignment depth (step 2) or by haplotype clustering (step 6) - genotypes must be extracted and re-organised according to genomic region. For regions identified by sequencing read depth, female homozygous genotypes are interpreted as representing haploid Z genotypes, while heterozygous genotypes are assumed to be erroneous. In regions that have been successfully phased, the inferred W haplotype is extracted as a female haploid genotype, along with the corresponding Z haplotype.

If sex-linked regions have only been identified through sex-specific differences in alignment depth, thay may appear as autosomal, pseudoautosomal, or unknown in step 6. To address that, *BEDTools* v.2.31.0 is used to subtract regions with sex-specific depth differences (identified in step 2) from autosomal regions in step 6. *BCFtools* v.1.17.b7b2a32 is then used to extract the remaining variants into an autosomal-specific vcf-file. Variants within regions of sex-specific depth differences are converted to haploid genotypes using a custom python script *diploid2haploid.py*, which replaces heterozygous genotypes to missing data. The sex-specific vcf-files are then merged with *BCFtools* v.1.17, and the allele count field (AC) is updated using *vcffixup* from *vcflib* v.1.0.3. Sites that are monomorphic for the reference allele are filtered out with *vcffilter* from *vcflib* v.1.0.3.

If any sex-linked region has been identified through haplotype clustering, *BEDTools* v.2.31.0 is used to remove regions with sex difference in depth (from step 2) from autosomal regions identified in step 6, and *BCFtools* v.1.17.b7b2a32 is used to extract the remaining variants into an autosomal-specific vcf-file. *BEDTools* v.2.31.0 is used to merge sex-linked regions identified through sex-specific depth differences with those identified by haplotype clustering (step 6), creating a bed-file representing all sex-linked regions across the scaffold. All genotypes within regions identified through haplotype clustering are extracted with *BCFtools* v.1.17.b7b2a32 and separated into male- and female-specific vcf-files. Female genotypes are then split into two vcf-files containing the “left” and the “right” haplotypes using a custom python *script split\_phase.py*. For each female, the left and right haplotypes are extracted as Z and W based on their classification in step 6, using *BCFtools* v.1.17.b7b2a32. These are then concatenated into individual female Z- and W-specific vcf-files with *BCFtools* v.1.17.b7b2a32. The individual W haplotypes are merged to create a haploid, female W-specific vcf-file using *BCFtools* v.1.17. The allele count field (AC) is updated with *vcffixup* from *vcflib* v.1.0.3, and sites monomorphic for the reference allele are removed with *vcffilter* from *vcflib* v.1.0.3. If sex-linked regions identified by sex-specific depth differences overlap with those identified by haplotype clustering, these sites are considered unreliable for W-sequence extraction and are excluded from the W-specific vcf-file using *VCFtools* v.0.1.16. The individual female Z haplotypes identified through haplotype clustering are then merged to create a haploid, female Z-specific vcf-file using *BCFtools* v.1.17. Female Z genotypes within regions identified by haplotype clustering are retained, and any additional variants identified by sex-specific depth difference are extracted from the phased vcf-file from step 5 using *BCFtools* v.1.17.b7b2a32 and *VCFtools* v.0.1.16. These are then converted to haploid genotypes with a custom python script *diploid2haploid.py*, converting heterozygous genotypes to missing. These two files containing Z haplotypes are concatenated using *BCFtools* v.1.17.b7b2a3, creating a single vcf-file with all female haploid Z genotypes. These are merged with the male diploid Z genotypes using *BCFtools* v.1.17 to form a single vcf-file containing both male diploid Z genotypes and female haploid Z genotypes. The allele count field (AC) is updated with *vcffixup* from *vcflib* v.1.0.3, and sites monomorphic for the reference allele are removed with *vcffilter* from *vcflib* v.1.0.3.

Finally, some variants within sex-linked regions may be considered problematic or unreliable. If the same alleles segregate on both the Z and W chromosome, this is likely indicative of unsuccessful phasing. Alternatively, it could reflect incomplete lineage sorting between Z- and W-sequences, althoung this would not be expected if they have been diverging for a long time. *BEDTools* v.2.31.0 and *BCFtools* v.1.17.b7b2a32 are used to extract sites represented at both the Z- and W-specific vcf-files. Any site that is not monomorphic in at least one lineage is recorded in a bed-file. These sites are not filtered from the vcf-file at this stage.

### Step 8. Summarise genome with bed-files

*BEDTools* v.2.31.0 is used to concatenate all per-scaffold bed-files into files representing the full genome. The file names follow the format <Project name>\_<Suffix>.bed, where <Project name> is the user-defined project name and <Suffix> corresponds to the type of genomic region or feature, as defined in the list below:

* hetgam\_dropout.bed – Callable regions with a sex difference in depth in step 2.
* unknown\_phase.bed – Regions with unknown sex-linkage (hetgam\_dropout.bed excluding unreliable classification based on clustering of haplotypes in step 6).
* target\_region.bed – All callable regions (i.e., excluding repeat regions and unknown\_phase.bed).
* missing\_region.bed – All uncallable regions (i.e., everything not in target\_region.bed).
* target\_region\_phased.bed – Callable regions excluding scaffolds that have not been phased (i.e., not in the list of scaffolds to phase, or scaffolds lacking variation needed for statistical phasing).
* phase\_windows.txt – Per-window information from haplotype clustering in step 6.
* phase\_info.bed – Classification of coherent regions from haplotype clustering in step 6.
* phase\_sex\_linked.bed – Sex-linked regions based on haplotype clustering in step 6.
* target\_phase\_sex\_linked.bed – Callable, sex-linked regions based on haplotype clustering in step 6.
* target\_phase\_hetgam\_dropout.bed – Callable, phased regions overlapping regions with a sex-specific depth difference in step 2.
* sex\_linked.bed – All callable sex-linked regions based on sex-specific depth difference (step 2) and haplotype clustering (step 6).
* homogametic.bed – All callable and sex-linked regions spanning the homogametic sex chromosome (identical to sex\_linked.bed).
* heterogametic.bed – All callable and sex-linked regions spanning the heterogametic sex chromosome.
* autosomal.bed – All callable regions spanning the autosomal regions, including the PAR.
* sex\_linked\_ILS1.bed – Sites in sex-linked regions from haplotype clustering (step 6) with signs of heterogametic allele drop out or incomplete lineage sorting. These sites have polymorphic alleles segregating as homozygous genotypes on both the homogametic and heterogametic sex chromosomes.
* sex\_linked\_ILS1\_not\_hetgam\_dropout.bed – Sites in sex-linked regions from haplotype clustering (step 6) with signs of incomplete lineage sorting. These sites have polymorphic alleles segregating as homozygous genotypes on both the homogametic and heterogametic sex chromosomes. Since they do not occur in regions with a depth differences (step 2), they are unlikely to result from drop out of the heterogametic allele.
* sex\_linked\_ILS2.bed – Sites in sex-linked regions from haplotype clustering (step 6) with signs of unsuccessful phasing or incomplete lineage sorting. These sites have polymorphic alleles segregating on both the homogametic and heterogametic sex chromosomes.

If a list of scaffolds to phase is provided, all non-phased scaffolds are assumed to be autosomal. However, any variants within such regions will be stored in a separate vcf-file.

### Step 9. Create final vcf-files

All per-scaffold vcf-files generated are concatenated into full-genome files. Concatenation and filtering are performed with *BCFtools* v.1.17.b7b2a32, *VCFtools* v.0.1.16, and *vcffilter* from *vcflib* v.1.0.3. The file names are in the format <Project name>\_<suffix>.vcf(.gz), where <Project name> is the user-defined project identifier, and <suffix> describes the content, as detailed below.

Two files containing all phased variants are generated. These are useful for further investigating the PAR border, as they span both the sex-linked and PAR with individual haplotype information. However, these files do not specify which haplotype is linked to which sex chromosome.

* phased\_all\_variants.vcf – All phased variants (excluding variants in scaffolds with insufficient variation for statistical phasing; only includes scaffolds in the phase list if provided).
* phased\_all\_variants\_filtered.vcf.gz - All phased variants, excluding unreliable classifications from haplotype clustering (step 6) and sites in identified sex-linked regions with signs of unsuccessful phasing or incomplete lineage sorting (see Summarise genome with bed-files: sex\_linked\_ILS2.bed).

Files containing variants per region are generated, i.e. autosomal, homogametic, and heterogametic. Autosomal genotypes are always diploid. Homogametic variants contain diploid genotypes for the homogametic sex (XX females and ZZ males) and haploid genotypes for the heterogametic sex (X males and Z females) and are provided as one file per sex (because some downstream software’s are not compatible with mixed ploidy). Heterogametic variants only include the heterogametic sex (Y males or W females) and are given as haploid.

* autosomal.vcf.gz – All variants spanning the autosomal regions, including the PAR.
* homogametic.vcf.gz – All individuals and all variants spanning the shared X/Z chromosome. Homogametes are coded as diploid, and heterogametes are coded as haploid.
* homogametic\_homogametes.vcf.gz – All homogametic individuals and all variants spanning the shared X/Z chromosome. Homogametes are coded as diploid.
* homogametic\_heterogametes.vcf.gz – All heterogametic individuals and all variants spanning the shared X/Z chromosome. Heterogametes are coded as haploid.
* heterogametic.vcf.gz – All heterogametic individuals and variants spanning the sex-limited Y/W chromosome. Heterogametes are coded as haploid.

Variants that did not pass to the final output are stored in additional files (for ILS, see Step 8. Summarise genome with bed-filesheterogametic\_ILS1.vcf.gz).

* heterogametic\_ILS2.vcf.gz – All heterogametic individuals and variants identify as ILS2. Heterogametes are coded as haploid.
* homogametic\_ILS2.vcf.gz – All homogametic individuals and variants identify as ILS2. Homogametes are coded as diploid.
* nonphased\_variants.vcf.gz – Variants located on scaffolds that were not phased (i.e., insufficient variation for statistical phasing).
* unknown\_phase.vcf.gz – Variants in regions with unknown sex-linkage (hetgam\_dropout.bed excluding unreliable classification from haplotype clustering in step 6).

### Step 10. Produce pipeline statistics

For each generated vcf-file, summary statistics are produced and visualized to enable evaluation of pipeline performance. Using *VCFtools* v.0.1.16, distributions of various variant characteristics are generated, including, proportion of missing data per site, and minor allele frequency (folded Site Frequency Spectrum). In addition, distributions of individual-specific characteristics are generated, such as inbreeding coefficient (FIS), and proportion of missing data per individual. Note that some statistics, such as the inbreeding coefficient, are not applicable to haploid genotypes and are therefore excluded where applicable. *Plink* v.1.90b4.9 (Purcell et al., 2007) is used to linkage-prune the data using 50 kb windows, a step size of 10 kb, and an r2 value of 0.1. Principal components are then calculated on the remaining independent variants. Additionally, the distribution of the allelic balance (i.e., the frequency of the minor allele vs. the major allele) at heterozygote genotype calls is calculated for each individual using a slightly modified version of the script *checkHetsIndvVCF.sh* (<https://github.com/speciationgenomics/scripts/blob/master/checkHetsIndvVCF.sh>).

To evaluate the effectiveness of retrieving autosomal sites by removing sex-linked variants, summary statistics of the autosomal output can be compared with the full dataset (phased\_all\_variants.vcf.gz). Similarly, the homogametic file can be examined for patterns that indicate whether heterogametic variants have been successfully separated from homogametic variants (See Sex chromosome identification and extraction of W-sequences). For example, individuals might cluster by sex on the first principal components (Figure S1). Two peaks in the site mean depth are often observed, with the smaller peak corresponding to degenerated regions of the sex chromosome, reflected by lower average depth in the heterogametic sex. An inflated peak may appear in the folded Site Frequency Spectrum corresponding to the frequency of the heterogametic sex chromosome in the dataset (Figure S2). Furthermore, there might be a sex-bias in the inbreeding coefficient, as the heterogametic individuals are typically more heterozygous and thus have a low inbreeding coefficient; Figure S3).

Addition, a genome-wide bed-file is produced, classifying each region as “Autosomal”, “Sex depth difference”, “Sex phase difference”, “Sex phase & depth difference”, or ”Missing data”. This file is particularly useful for visualising genome features across the entire genome (Figure 1). To facilitate comparative analysis with the genome of another species - such as one with a chromosome-level assembly - this classification file can be used in a lift-over procedure (for an example, see Sex chromosome identification and extraction of W-sequences and <https://github.com/sjellerstrand/Lark_sex_chrom_genomics/blob/main/Skylark_2021/process_data/coordinate_lift_over.sh>).

# 5. References

Delaneau, O., Zagury, J.-F., Robinson, M.R., Marchini, J.L., Dermitzakis, E.T., 2019. Accurate, scalable and integrative haplotype estimation. Nat Commun 10, 5436. https://doi.org/10.1038/s41467-019-13225-y

Hartigan, J.A., Wong, M.A., 1979. A *K*-Means Clustering Algorithm. Applied Statistics 28, 100–108. https://doi.org/10.2307/2346830

Knaus, B.J., Grünwald, N.J., 2017. VCFR: a package to manipulate and visualize variant call format data in R. Molecular Ecology Resources 17, 44–53. https://doi.org/10.1111/1755-0998.12549

Martin, M., Patterson, M., Garg, S., O Fischer, S., Pisanti, N., Klau, G.W., Schöenhuth, A., Marschall, T., 2016. WhatsHap: fast and accurate read-based phasing. bioRxiv. https://doi.org/10.1101/085050

Purcell, S., Neale, B., Todd-Brown, K., Thomas, L., Ferreira, M.A.R., Bender, D., Maller, J., Sklar, P., de Bakker, P.I.W., Daly, M.J., Sham, P.C., 2007. PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage Analyses. The American Journal of Human Genetics 81, 559–575. https://doi.org/10.1086/519795

Quinlan, A.R., Hall, I.M., 2010. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics 26, 841–842. https://doi.org/10.1093/bioinformatics/btq033

Zhou, Q., Zhang, J., Bachtrog, D., An, N., Huang, Q., Jarvis, E.D., Gilbert, M.T.P., Zhang, G., 2014. Complex evolutionary trajectories of sex chromosomes across bird taxa. Science 346, 1246338. https://doi.org/10.1126/science.1246338