**File S1: Full description of the model**

***Observed and hidden variables*** The data consist of genotyped individuals for genes $k \in \{1, 2, ..., K\}$, each with biallelic sites $t \in \{1, 2, ..., T_k\}$. For each biallelic site, the observed alleles are randomly named $a$ and $b$, such that three genotypes are possible: homozygote for allele $a$ ("$aa$"), heterozygote ("$ab$"), and homozygote for allele $b$ ("$bb$"). At each biallelic site $t$ of gene $k$, and for each individual $i$, $\mathrm{OG}^{tk}_{i_h g}$ is an indicator of the individual having observed genotype $g \in \{1 = aa, 2 = ab, 3 = bb\}$ and sex $h \in \{1, 2\}$. $h = 1$ for females, $h = 2$ males; for convenience, we will write ♀ and ♂. is the number of individuals with sex $h$ and observed genotype $g$; note that the total number of observations (i.e., genotyped individuals) can vary between sites ($N^{kt}$ is the total number of observed genotypes at a site). The vector $\mathbf{OG}^{kt}$ describes all observations at a site.

We seek under which segregation type $S_j$, $j \in 1...7$, these observed genotypes are most likely. These segregation types will be listed in fixed order, and a specific number corresponds to each of them:

1. Diploid autosomal segregation

2. Haploid sequences

3. Paralogs

4. X-hemizygous segregation

5. XY gametologous segregation

6. Z-hemizygous segregation

7. ZW gametologous segregation

For some segregation types, as will be detailed below, several sub-types of segregation have to be specified; these are denoted $A_l$ with $l \in \{1..L\}$. The conditional likelihood of observing the genotypes under each segregation type depends on the allele frequencies and a genotyping error rate.

We introduce a hidden variable $\mathrm{TG}^{kt}_{i_h g'}$ which is an indicator for the true genotype $g' \in \{aa, ab, bb\}$ of an individual $i$ with sex $h$. The conditional probabilities of observing a true genotype for an individual, given the fully specified segregation type, are

$$\mathbf{TG}^{kt}_{i_h} | S_j, A_l \sim \mathcal{M}\left(1; P^{kt}_{1hjl}, P^{kt}_{2hjl}, P^{kt}_{3hjl}\right) .$$

$\mathbf{P}^{kt}_{hjl}$ is the vector of the probabilities $\left(P^{kt}_{1hjl}, P^{kt}_{2hjl}, P^{kt}_{3hjl}\right)$ for each genotype at a site, given the sex of the individual and the segregation type and subtype. The genotype probabilities $P^{kt}_{g'hjl}$ are calculated from the empirical allele frequencies $\widehat{f}^{kt}_{jl}$ using the following population genetic expectations.

1. For autosomal segregation, the Hardy-Weinberg equilibrium should hold in both sexes. Thus,

$$\mathbf{P}^{kt}_{♀,j=1} = \mathbf{P}^{kt}_{♂,j=1} = \begin{pmatrix} (\widehat{f}^{kt}_{j=1})^2 \\ 2\widehat{f}^{kt}_{j=1}\left(1 - \widehat{f}^{kt}_{j=1}\right) \\ \left(1 - \widehat{f}^{kt}_{j=1}\right)^2 \end{pmatrix}$$

where

$$\widehat{f}^{kt}_{j=1} = \frac{2N^{kt}_{aa} + N^{kt}_{ab}}{2N^{kt}} .$$

2. Haploid segregation is modeled by

$$\mathbf{P}^{kt}_{♀,j=2} = \mathbf{P}^{kt}_{♂,j=2} = \begin{pmatrix} \widehat{f}^{kt}_{j=2} \\ 0 \\ 1 - \widehat{f}^{kt}_{j=2} \end{pmatrix}$$

and

$$\widehat{f}^{kt}_{j=2} = \frac{N^{kt}_{aa}}{N^{kt}_{aa} + N^{kt}_{bb}} .$$

3. Paralogy is caused by the mapping of the reads of two more or less recently duplicated genes on one locus in the reference. There is no recombination between the copies, that thus evolve independently. For simplicity, we assume that one of the copies is fixed for one of the alleles. The genotype probabilities depend on which allele is considered fixed in one of the copies, and two sub-types have to be modeled.

(a) First, we consider that allele $a$ is fixed in one of the copies. $\widehat{f}^{kt}_{j=3,l=1}$ is the frequency of allele $b$ in the other copy. In reality, such sites have four copies; thus, the genotypes are $aaaa$, $aaab$, are $aabb$, with frequencies $\left(1 - \widehat{f}^{kt}_{j=3,l=1}\right)^2$, $2\widehat{f}^{kt}_{j=3,l=1}\left(1 - \widehat{f}^{kt}_{j=3,l=1}\right)$ and $(\widehat{f}^{kt}_{j=3,l=1})^2$. Genotypes $aaab$ and $aabb$ will probably be considered as $ab$ by the genotyper that expects only diploids; thus, the genotype probabilities are:

$$\mathbf{P}^{kt}_{\female,j=3,l=1} = \mathbf{P}^{kt}_{\male,j=3,l=1} = \begin{pmatrix} \left(1 - \widehat{f}^{kt}_{j=3,l=1}\right)^2 \\ (\widehat{f}^{kt}_{j=3,l=1})^2 + 2\widehat{f}^{kt}_{j=3,l=1}\left(1 - \widehat{f}^{kt}_{j=3,l=1}\right) \\ 0 \end{pmatrix}$$

To estimate the empirical allele frequency, note that the $ab$ genotype counts that are obtained from the genotyper ($N_{ab}$) will likely be a mixture of $aaab$ and $aabb$. The expected proportions $N_{aaab}$ and $N_{aabb}$ can be calculated depending on the frequency $\widehat{f}^{kt}_{j=3,l=1}$ that we concisely denote $\widehat{f}$ here: $2\widehat{f}\left(1 - \widehat{f}\right)/(\widehat{f}^2 + 2\widehat{f}\left(1 - \widehat{f}\right))$ and $\widehat{f}^2/(\widehat{f}^2 + 2\widehat{f}\left(1 - \widehat{f}\right))$. While in reality, $\widehat{f} = 0.5(N_{aaab} + 2N_{aabb})/(N_{aaaa} + N_{aaab} + N_{aabb})$, we instead calculate

$$\widehat{f} = \frac{1}{2\left(N_{aa} + N_{ab}\right)}\left(\frac{2\widehat{f}\left(1 - \widehat{f}\right)N_{ab}}{\widehat{f}^2 + 2\widehat{f}\left(1 - \widehat{f}\right)} + \frac{2\widehat{f}^2 N_{ab}}{\widehat{f}^2 + 2\widehat{f}\left(1 - \widehat{f}\right)}\right)$$

This yields

$$\widehat{f}^{kt}_{j=3,l=1} = 1 - \sqrt{1 - \frac{N^{kt}_{ab}}{N^{kt}_{aa} + N^{kt}_{ab}}}.$$

(b) Alternatively, allele $b$ could be fixed in one of the copies. $\widehat{f}^{kt}_{j=3,l=2}$ is the frequency of allele $a$ in the other copy. The genotype probabilities and empirical allele frequency are

$$\mathbf{P}^{kt}_{\female,j=3,l=2} = \mathbf{P}^{kt}_{\male,j=3,l=2} = \begin{pmatrix} 0 \\ (\widehat{f}^{kt}_{j=3,l=2})^2 + 2\widehat{f}^{kt}_{j=3,l=2}\left(1 - \widehat{f}^{kt}_{j=3,l=2}\right) \\ \left(1 - \widehat{f}^{kt}_{j=3,l=2}\right)^2 \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=3,l=2} = 1 - \sqrt{1 - \frac{N^{kt}_{ab}}{N^{kt}_{ab} + N^{kt}_{bb}}}.$$

4. For X-hemizygously segregating genes, the males are haploid while the females are diploid.

$$\mathbf{P}^{kt}_{\female,j=4} = \begin{pmatrix} (\widehat{f}^{kt}_{j=4})^2 \\ 2\widehat{f}^{kt}_{j=4}\left(1 - \widehat{f}^{kt}_{j=4}\right) \\ \left(1 - \widehat{f}^{kt}_{j=4}\right)^2 \end{pmatrix} \quad ; \quad \mathbf{P}^{kt}_{\male,j=4} = \begin{pmatrix} \widehat{f}^{kt}_{j=4} \\ 0 \\ 1 - \widehat{f}^{kt}_{j=4} \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=4} = \frac{2N^{kt}_{aa,\female} + N^{kt}_{ab,\female} + N^{kt}_{aa,\male}}{2\left(N^{kt}_{aa,\female} + N^{kt}_{ab,\female} + N^{kt}_{bb,\female}\right) + N^{kt}_{aa,\male} + N^{kt}_{bb,\male}}$$

5. XY gametologous segregation is characterized by the presence of two independent copies in males, and two copies of the X gene in females. We assume that an allele is fixed in at least one of the copies.

(a) X-polymorphism, allele $a$ fixed on Y. $f$ is the frequency of allele $b$ on X.

$$\mathbf{P}^{kt}_{\female,j=5,l=1} = \begin{pmatrix} \left(1 - \widehat{f}^{kt}_{j=5,l=1}\right)^2 \\ 2\widehat{f}^{kt}_{j=5,l=1}\left(1 - \widehat{f}^{kt}_{j=5,l=1}\right) \\ (\widehat{f}^{kt}_{j=5,l=1})^2 \end{pmatrix} \quad ; \quad \mathbf{P}^{kt}_{\male,j=5,l=1} = \begin{pmatrix} 1 - \widehat{f}^{kt}_{j=5,l=1} \\ \widehat{f}^{kt}_{j=5,l=1} \\ 0 \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=5,l=1} = \frac{2N^{kt}_{bb\female} + N^{kt}_{ab\female} + N^{kt}_{ab\male}}{2\left(N^{kt}_{aa\female} + N^{kt}_{ab\female} + N^{kt}_{bb\female}\right) + N^{kt}_{aa\male} + N^{kt}_{ab\male}}$$

(b) X-polymorphism; allele $b$ fixed on Y. $f$ is the frequency of allele $a$ on X :

$$\mathbf{P}^{kt}_{\female,j=5,l=2} = \begin{pmatrix} (\widehat{f}^{kt}_{j=5,l=2})^2 \\ 2\widehat{f}^{kt}_{j=5,l=2}\left(1-\widehat{f}^{kt}_{j=5,l=2}\right) \\ \left(1-\widehat{f}^{kt}_{j=5,l=2}\right)^2 \end{pmatrix} \quad ; \quad \mathbf{P}^{kt}_{\male,j=5,l=2} = \begin{pmatrix} 0 \\ \widehat{f}^{kt}_{j=5,l=2} \\ 1-\widehat{f}^{kt}_{j=5,l=2} \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=5,l=2} = \frac{2N^{kt}_{aa\female} + N^{kt}_{ab\female} + N^{kt}_{ab\male}}{2\left(N^{kt}_{aa\female} + N^{kt}_{ab\female} + N^{kt}_{bb\female}\right) + N^{kt}_{bb\male} + N^{kt}_{ab\male}}$$

(c) Y-polymorphism, allele $a$ fixed on X. $f$ is the frequency of allele $b$ on Y:

$$\mathbf{P}^{kt}_{\female,j=5,l=3} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} \quad ; \quad \mathbf{P}^{kt}_{\male,j=5,l=3} = \begin{pmatrix} 1-\widehat{f}^{kt}_{j=5,l=3} \\ \widehat{f}^{kt}_{j=5,l=3} \\ 0 \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=5,l=3} = \frac{N^{kt}_{ab\male}}{N^{kt}_{aa\male} + N^{kt}_{ab\male}}$$

(d) Y-polymorphism, allele $b$ fixed on X. $f$ is the frequency of allele $a$ on Y:

$$\mathbf{P}^{kt}_{\female,j=5,l=4} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} \quad ; \quad \mathbf{P}^{kt}_{\male,j=5,l=4} = \begin{pmatrix} 0 \\ \widehat{f}^{kt}_{j=5,l=4} \\ 1-\widehat{f}^{kt}_{j=5,l=4} \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=5,l=4} = \frac{N^{kt}_{ab\male}}{N^{kt}_{bb\male} + N^{kt}_{ab\male}}$$

6. Z-hemizygous segregation is similar to X-hemizygous segregation:

$$\mathbf{P}^{kt}_{\female,j=6} = \begin{pmatrix} \widehat{f}^{kt}_{j=6} \\ 0 \\ 1-\widehat{f}^{kt}_{j=6} \end{pmatrix} \quad ; \quad \mathbf{P}^{kt}_{\male,j=6} = \begin{pmatrix} (\widehat{f}^{kt}_{j=6})^2 \\ 2\widehat{f}^{kt}_{j=6}\left(1-\widehat{f}^{kt}_{j=6}\right) \\ \left(1-\widehat{f}^{kt}_{j=6}\right)^2 \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=6} = \frac{2N^{kt}_{aa,\male} + N^{kt}_{ab,\male} + N^{kt}_{aa,\female}}{2\left(N^{kt}_{aa,\male} + N^{kt}_{ab,\male} + N^{kt}_{bb,\male}\right) + N^{kt}_{aa,\female} + N^{kt}_{bb,\female}}$$

7. ZW gametologous segregation is modeled similar to XY gametologous segregation, for both Z and W polymorphism, and two asymmetrical cases for each.

(a) Z-polymorphism, allele $a$ fixed on W. $f$ is the frequency of allele $b$ on Z.

$$\mathbf{P}^{kt}_{\female,j=7,l=1} = \begin{pmatrix} 1-\widehat{f}^{kt}_{j=7,l=1} \\ \widehat{f}^{kt}_{j=7,l=1} \\ 0 \end{pmatrix} \quad ; \quad \mathbf{P}^{kt}_{\male,j=7,l=1} = \begin{pmatrix} \left(1-\widehat{f}^{kt}_{j=7,l=1}\right)^2 \\ 2\widehat{f}^{kt}_{j=7,l=1}\left(1-\widehat{f}^{kt}_{j=7,l=1}\right) \\ (\widehat{f}^{kt}_{j=7,l=1})^2 \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=7,l=1} = \frac{2N^{kt}_{bb\male} + N^{kt}_{ab\male} + N^{kt}_{ab\female}}{2\left(N^{kt}_{aa\male} + N^{kt}_{ab\male} + N^{kt}_{bb\male}\right) + N^{kt}_{aa\female} + N^{kt}_{ab\female}}$$

(b) Z-polymorphism; allele $b$ fixed on W. $f$ is the frequency of allele $a$ on Z :

$$\mathbf{P}^{kt}_{\female,j=7,l=2} = \begin{pmatrix} 0 \\ \widehat{f}^{kt}_{j=7,l=2} \\ 1-\widehat{f}^{kt}_{j=7,l=2} \end{pmatrix} \quad ; \quad \mathbf{P}^{kt}_{\male,j=7,l=2} = \begin{pmatrix} ()\widehat{f}^{kt}_{j=7,l=2})^2 \\ 2\widehat{f}^{kt}_{j=7,l=2}\left(1-\widehat{f}^{kt}_{j=7,l=2}\right) \\ \left(1-\widehat{f}^{kt}_{j=7,l=2}\right)^2 \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=7,l=2} = \frac{2N^{kt}_{aa\male} + N^{kt}_{ab\male} + N^{kt}_{ab\female}}{2\left(N^{kt}_{aa\male} + N^{kt}_{ab\male} + N^{kt}_{bb\male}\right) + N^{kt}_{bb\female} + N^{kt}_{ab\female}}$$

(c) W-polymorphism, allele $a$ fixed on Z. $f$ is the frequency of allele $b$ on W:

$$\mathbf{P}^{kt}_{\female,j=7,l=3} = \begin{pmatrix} 1 - \widehat{f}^{kt}_{j=7,l=3} \\ \widehat{f}^{kt}_{j=7,l=3} \\ 0 \end{pmatrix} \quad ; \quad \mathbf{P}^{kt}_{\male,j=7,l=3} = \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=7,l=3} = \frac{N^{kt}_{ab\female}}{N^{kt}_{aa\female} + N^{kt}_{ab\female}}$$

(d) W-polymorphism, allele $b$ fixed on Z. $f$ is the frequency of allele $a$ on W:

$$\mathbf{P}^{kt}_{\female,j=7,l=4} = \begin{pmatrix} 0 \\ \widehat{f}^{kt}_{j=7,l=4} \\ 1 - \widehat{f}^{kt}_{j=7,l=4} \end{pmatrix} \quad ; \quad \mathbf{P}^{kt}_{\male,j=7,l=4} = \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix}$$

$$\widehat{f}^{kt}_{j=7,l=4} = \frac{N^{kt}_{ab\female}}{N^{kt}_{bb\female} + N^{kt}_{ab\female}}$$

As we calculate all genotype probabilities for all segregation types at all sites, calculation of $\widehat{f}^{kt}_{jl}$ might lead to division by 0. To avoid this problem, counts that are expected to be 0 under a segregation type are added to the numerator and the denominator.

Genotyping errors (whether they are due to sequencing errors, read mapping errors, or violations of the assumptions of the method for genotyping) cause the observed genotype $g$ to be different from the true genotype $g'$. We define $q_{gg'} = \mathbb{P}\left(\text{OG}^{kt}_{i_h g}|\text{TG}^{kt}_{i_h g'}\right)$, i.e., the probability to observe genotype $g$ when the true genotype is $g'$, and $\mathbf{Q}$ is the matrix of all $q_{gg'}$, such that

$$\mathbf{Q} = \begin{pmatrix} q_{1,1} & q_{1,2} & q_{1,3} \\ q_{2,1} & q_{2,2} & q_{2,3} \\ q_{3,1} & q_{3,2} & q_{3,3} \end{pmatrix}$$

We can now directly calculate the probabilities of the observed genotypes for each segregation type:

$$\begin{aligned} \mathbb{P}\left(\text{OG}^{kt}_{i_h g}|S_j, A_l\right) &= \sum_{g'} \mathbb{P}\left(\text{TG}^{kt}_{i_h g'}|S_j, A_l\right) \mathbb{P}\left(\text{OG}^{kt}_{i_h g}|\text{TG}^{kt}_{i_h g'}\right) \\ &= \sum_{g'} P^{kt}_{g'hjl} q_{gg'} \end{aligned}$$

We rename the quantity $\sum_{g'} P^{kt}_{g'hjl} q_{gg'}$ as $\widetilde{P}^{kt}_{ghjl}$; it is the expected frequency of the observed genotype given the segregation type and a certain genotyping error rate. For each sex, $\mathbf{OG}$ follows a multinomial distribution $\mathcal{M}\left(N^{kt}_h; \widetilde{P}^{kt}_{1hjl}, \widetilde{P}^{kt}_{2hjl}, \widetilde{P}^{k}_{3hjl}\right)$. Thus, the conditional likelihood of the data (given the segregation type) at each site, that we name $M^{kt}_{jl}$, is

$$M^{kt}_{jl} = \mathbb{P}\left(\mathbf{OG}^{kt}|S_j, A_l\right) = \prod_{gh} \left(\widetilde{P}^{kt}_{ghjl}\right)^{N^{kt}_{gh}} \tag{1}$$

***Parameters*** The error rates $q_{gg'}$ depend on one error parameter $e$. We assume all genotyping errors to occur with the same frequency, so $q_{g,g'\neq g} = e$ and $q_{g,g'=g} = 1 - 2e$, which gives the error matrix

$$\mathbf{Q} = \begin{pmatrix} 1 - 2e & e & e \\ e & 1 - 2e & e \\ e & e & 1 - 2e \end{pmatrix}$$

Two more series of parameters are required to model the data; these indicate the proportion of the genome that segregates under each type. There are a maximum of seven segregation types $S_j$, each occupying a proportion $\pi_j$ of the genome, such that $\sum_j \pi_j = 1$. $\boldsymbol{\pi}$ is the vector containing all $\pi_j$. The segregation types $S^{kt}_j$ are distributed multinomially, thus

$$\mathbf{S} \sim \mathcal{M}(1, \boldsymbol{\pi})$$

Several biologically relevant segregation types ($S$) have several "subtypes" ($A$), depending on the fixation of one of the alleles on either of the copies. Thus, for a segregation type $S_j$, there are $L$ subtypes, and each subtype $A_{jl}$ applies to a proportion $\alpha_{jl}$ of the proportion $\pi_j$ of the genome (corresponding to the segregation type $S_j$). For each segregation type with subtypes, $\sum_{l=1}^{L} \alpha_{jl} = 1$, and

$$\mathbf{A}_j | S_j \sim \mathcal{M}(1, \boldsymbol{\alpha}_j)$$

For the paralogs, the subtype depends uniquely on the choice of what allele is called $a$, which is random. Thus, no parameter is needed, and

$$\boldsymbol{\alpha}_3 = \left( \frac{1}{2}, \frac{1}{2} \right)$$

For the XY and ZW types, more sites can be polymorphic on one chromosome than on the other. The proportion of XY or ZW sites that are polymorphic on X or on Z is described by the parameter $\rho_j$, which takes a single value for each segregation type. The (random) choice what allele is called $a$ affects both X (or Z) and Y (or W) polymorphisms, leading to four subtypes

$$\boldsymbol{\alpha}_5 = \left( \frac{\rho_5}{2}, \frac{\rho_5}{2}, \frac{1-\rho_5}{2}, \frac{1-\rho_5}{2} \right)$$

When aggregating the segregation subtypes ($A$) to biologically relevant types ($S$), we get

$$\mathbb{P}\left( \mathbf{OG}^{kt} | S_j \right) = B_j^{kt} = \sum_l \alpha_{jl} M_{jl}^{kt} \tag{2}$$

**Expectation-Maximization algorithm** The full log-likelihood of the model is given by

$$
\begin{aligned}
\log \mathbb{P}\left( \mathbf{OG}, \mathbf{TG}, \mathbf{S}, \mathbf{A} \right) &= \log \mathbb{P}\left( \mathbf{OG} | \mathbf{TG} \right) \\
&+ \log \mathbb{P}\left( \mathbf{TG} | \mathbf{S}, \mathbf{A} \right) \\
&+ \log \mathbb{P}\left( \mathbf{A} | \mathbf{S} \right) \\
&+ \log \mathbb{P}\left( \mathbf{S} \right)
\end{aligned}
$$

This likelihood is maximized through an Expectation-Maximization (EM) algorithm.

**E-step** The posterior segregation types are given by

$$\mathbb{E}\left( \log \mathbb{P}\left( \mathbf{S} \right) | \mathbf{OG} \right) = \sum_{jkt} \mathbb{E}(S_j^{kt} | \mathbf{OG}^{kt}) \log \pi_j$$

with

$$\mathbb{E}(S_j^{kt} | \mathbf{OG}^{kt}) = \widehat{S}_j^{kt} = \frac{\pi_j B_j^{kt}}{\sum_{j'} \pi_{j'} B_{j'}^{kt}} \tag{3}$$

The posteriors for the subtypes are calculated by

$$
\begin{aligned}
\mathbb{E}\left( \log \mathbb{P}\left( \mathbf{A}^{kt} | \mathbf{S}^{k(t)} \right) | \mathbf{OG}^{kt} \right) &= \sum_{jlkt} \mathbb{E}\left( A_l^{kt} S_j^{k(t)} | \mathbf{OG}^{kt} \right) \log \alpha_{jl} \\
&= \sum_{jlkt} \mathbb{E}\left( A_l^{kt} | \mathbf{OG}^{kt}, S_j^{k(t)} \right) \mathbb{E}\left( S_j^{k(t)} | \mathbf{OG}^{k(t)} \right) \log \alpha_{jl} \\
&= \sum_{jlkt} \widehat{A}_{lj}^{kt} \widehat{S}_j^{k(t)} \log \alpha_{jl}
\end{aligned}
$$

$$\widehat{A}_{lj}^{kt} = \frac{\alpha_{jl} M_{jl}^{kt}}{\sum_{l'} \alpha_{jl'} M_{jl'}^{kt}}$$

For the true expected true genotypes, we calculate

$$
\begin{aligned}
\mathbb{E}\left( \log \mathbb{P}(\mathbf{TG} | \mathbf{S}, \mathbf{A}) | \mathbf{OG} \right) &= \sum_{(kt)(jl)(i_h g')} \mathbb{E}\left( S_j^{kt} A_{jl}^{kt} \mathrm{TG}_{i_h g'}^{kt} | \mathbf{OG}_{i_h}^{kt} \right) \log P_{g'hjl}^{kt} \\
&= \sum_{(kt)(jl)(i_h g')} \mathbb{E}\left( \mathrm{TG}_{i_h g'}^{kt} | \mathbf{OG}_{ih}^{kt}, S_j^{kt}, A_{jl}^{kt} \right) \mathbb{E}\left( S_j^{kt}, A_{jl}^{kt} | \mathbf{OG} \right) \log P_{g'hjl}^{kt} \\
&= \sum_{(kt)(jl)(i_h g')} \mathbb{E}\left( \mathrm{TG}_{i_h g'}^{kt} | \mathbf{OG}_{ih}^{kt}, S_j^{kt}, A_{jl}^{kt} \right) \widehat{A}_{lj}^{kt} \widehat{S}_j^{kt} \log P_{g'hjl}^{kt}
\end{aligned}
$$

$$\widehat{\text{TG}}^{kt}_{i_h g' jl} = \frac{P^{kt}_{g' hjl} \prod_g q_{gg'}^{\text{OG}^{kt}_{i_h g}}}{\sum_{g''} P^{kt}_{g'' hjl} \prod_g q_{gg''}^{\text{OG}^{kt}_{i_h g}}}$$

As individuals are defined uniquely by their sex and their observed genotype, $\widehat{\text{TG}}^{kt}_{i_h g' jl}$ is the same for two individuals having the same sex and genotype. Thus, we write $\widehat{\text{TG}}'^{kt}_{h g g' jl} = \frac{P^{kt}_{g' hjl} q_{gg'}}{\sum_{g''} P^{kt}_{g'' hjl} q_{gg''}}$.

Finally, the conditional likelihood of the observed genotypes is given by

$$
\begin{aligned}
\mathbb{E}\left(\log \mathbb{P}\left(\mathbf{OG}|\mathbf{TG}\right)|\mathbf{OG}\right) &= \sum_{(kt)(jl)(i_h g')g} \text{OG}^{kt}_{i_h g} \mathbb{E}\left(\text{TG}^{kt}_{i_h g'}|\mathbf{OG}^{kt}\right) \log q_{gg'} \\
&= \sum_{(kt)(jl)(i_h g')g} \text{OG}^{kt}_{i_h g} \widehat{\text{TG}}^{kt}_{i_h g' jl} \widehat{A}^{kt}_{lj} \widehat{S}^{kt}_j \log q_{gg'}
\end{aligned}
$$

**M-step** The key quantity used in the M-step is the conditional expectation of the complete-data likelihood:

$$
\begin{aligned}
\mathbb{E}\left(\log \mathbb{P}\left(\mathbf{OG},\mathbf{TG},\mathbf{S},\mathbf{A}\right)|\mathbf{OG}\right) &= \sum_{(kt)(jl)(i_h g')g} \text{OG}^{kt}_{i_h g} \widehat{\text{TG}}^{kt}_{i_h g' jl} \widehat{A}^{kt}_{lj} \widehat{S}^{k(t)}_j \log q_{gg'} \\
&+ \sum_{(kt)(jl)(i_h g')} \widehat{\text{TG}}^{kt}_{i_h g' jl} \widehat{A}^{kt}_{lj} \widehat{S}^{k(t)}_j \log P^{kth}_{g' jl} \\
&+ \sum_{(kt)(jl)} \widehat{A}^{kt}_{lj} \widehat{S}^{k(t)}_j \log \alpha_{jl} + \sum_{(k(t))j} \widehat{S}^{k(t)}_j \log \pi_j \\
&= \sum_{(kt)(hg)} N^{kth}_g \left( \sum_{(jl)} \widehat{A}^{kt}_{lj} \widehat{S}^{k(t)}_j \left( \sum_{g'} \widehat{\text{TG}}'^{kt}_{hg' jl} \left( \log q_{gg'} + \log P^{kth}_{g' jl} \right) \right) \right) \\
&+ \sum_{(kt)(jl)} \widehat{A}^{kt}_{lj} \widehat{S}^{k(t)}_j \log \alpha_{jl} + \sum_{(k(t))j} \widehat{S}^{k(t)}_j \log \pi_j
\end{aligned}
$$

Parameters to estimate are $\pi, \alpha$ and error rate $e$. These parameters only involve

$$
\begin{aligned}
\mathbb{E}\left(\log \mathbb{P}\left(\mathbf{OG}|\mathbf{TG}\right)|\mathbf{OG}\right) &= \sum_{(kt)} \sum_{(jl)(i_h)} \text{OG}^{kt}_{i_h g} \widehat{\text{TG}}^{kt}_{i_h g' jl} \widehat{A}^{kt}_{lj} \widehat{S}^{k(t)}_j \log q_{gg'} \\
&= \sum_{(kt)} \sum_{(jl)(gh)} N^{kth}_g \widehat{\text{TG}}'^{kt}_{hg' jl} \widehat{A}^{kt}_{lj} \widehat{S}^{k(t)}_j \log q_{gg'}
\end{aligned}
$$

To simplify notations, let us denote

$$
\begin{aligned}
\widehat{U}_{gg'} &= \sum_{(kt)} \sum_{(jl)(i_h)} \text{OG}^{kt}_{i_h g} \widehat{\text{TG}}^{kt}_{i_h g' jl} \widehat{A}^{kt}_{lj} \widehat{S}^{k(t)}_j \\
&= \sum_{(kt)} \sum_{(jl)(h)} N^{kth}_g \widehat{\text{TG}}'^{kt}_{hg' jl} \widehat{A}^{kt}_{lj} \widehat{S}^{k(t)}_j
\end{aligned}
$$

Thus,

$$
\mathbb{E}\left(\log \mathbb{P}\left(\mathbf{OG}|\mathbf{TG}\right)|\mathbf{OG}\right) = \sum_{(gg')} \widehat{U}_{gg'} \log q_{gg'}
$$

We find the new values of $e$ by $\frac{\partial \mathbb{E}(\log \mathbb{P}(\mathbf{OG}|\mathbf{TG})|\mathbf{OG})}{\partial e} = 0$, which gives:

$$
\begin{aligned}
\widehat{e} &= \frac{\widehat{U}_{12} + \widehat{U}_{13} + \widehat{U}_{21} + \widehat{U}_{23} + \widehat{U}_{31} + \widehat{U}_{32}}{2(\widehat{U}_{12} + \widehat{U}_{22} + \widehat{U}_{32} + \widehat{U}_{13} + \widehat{U}_{23} + \widehat{U}_{33} + \widehat{U}_{11} + \widehat{U}_{21} + \widehat{U}_{31})} \\
&= \frac{\sum_{(g \neq g')} \widehat{U}_{gg'}}{2 \sum_{(gg')} \widehat{U}_{gg'}}
\end{aligned}
$$

Similarly, we calculate

$$\widehat{\rho} = \frac{\sum_{kt} \widehat{S}_3^{kt} \left( \widehat{A}_{13}^{kt} + \widehat{A}_{23}^{kt} \right)}{\sum_{kt} \widehat{S}_3^{kt}}$$

$$\widehat{\pi}_j = \frac{\sum_{kt} \widehat{S}_j^{kt}}{\sum_{kt} 1}$$

**Monitoring and convergence**     The likelihood of the data in the model is

$$\log \mathbb{P} \left( \mathbf{OG} \right) = \sum_{kt} \log \left( \sum_{jl} \left( \mathbb{P} \left( A_l^{kt} \right) \mathbb{P} \left( S_j^{kt} \right) \prod_i \mathbb{P} \left( \mathrm{OG}_{i_{h^g}}^{kt} | S_j^{kt}, A_l^{kt} \right) \right) \right)$$

$$= \sum_{kt} \log \left( \sum_{jl} \pi_j \alpha_l M_{jl}^{kt} \right)$$

Convergence is evaluated as a function of the relative change in parameter value estimations. Optimization is halted when the largest relative change of all parameters has been less than $10^{-4}$ for 10 iterations, except for the error rate parameter, which is not considered for convergence.

There are $J - 1$ free parameters for the proportion $\pi_j$ of the segregation types, one parameter $\rho$ for each of the XY and ZW types, and one parameter for the error rate. $\hat{\theta}$ is the set of optimized parameters. If the number of parameters is $\xi$, we calculate the Bayesian Information Criterion (BIC) as follows:

$$\mathrm{BIC} = \log \mathbb{P} \left( \mathbf{OG}; \hat{\theta} \right) - \frac{1}{2} \log \left( \sum_{kt} 1 \right) \xi$$

The model with the lowest BIC has the best fit.

***Site- and contig-wise probabilities***     The posterior probabilities per site, as given in Equation 3, are calculated using the priors $\pi_j$, which are the estimated proportions of each segregation type in the genome. The smaller $\pi_j$, the higher the conditional likelihood $B_j^{kt}$ should be to produce a high posterior probability. For the sex-linked segregation types $\pi_j$ can easily be very small. If, say, 0.1% of the sites are inferred as gametologous and 99.9% as autosomal, the conditional likelihood for the gametologous segregation types should be 1000 times higher than the one for autosomal segregation to obtain comparable posterior likelihoods with this formula. In order to avoid excessive biases against rare segregation types, for inference purposes at the end of the optimization, we calculate the posterior probabilities without priors, which amount to using a uniform prior. Thus, for the output, we compute

$$\widehat{S}_j^{kt} = \frac{B_j^{kt}}{\sum_{j'} B_{j'}^{kt}} \tag{4}$$

At the contig level, the goal is to estimate the posterior probability to be sex-linked, autosomal, or not informative (*i.e.,* haploid or paralogous), given the observed data for each of its sites and the optimal parameter values. This probability is the expectation of each segregation type, $\widehat{S}_j^k$, which we calculate from the site-wise probabilities. As sites are treated as unlinked, which they are obviously not within a contig, especially when they are sex-linked, calculating the product of the site likelihoods would lead to ignoring the dependence induced by linkage and to overestimating the effective number of independent observations. This is thus expected to inflate the posterior probability contrasts between alternative hypotheses (segregation types) for a given contig. Instead, we take the geometric mean, which reduces this effect:

$$\widehat{S}_{\mathrm{N}j}^k = \frac{\mathrm{GM} \left( \widehat{S}_j^{kt} \right)}{\sum_{j'} \mathrm{GM} \left( \widehat{S}_{j'}^{kt} \right)} = \frac{\mathrm{GM} \left( B_j^{kt} \right)}{\sum_{j'} \mathrm{GM} \left( B_{j'}^{kt} \right)} = \frac{\exp \left( \frac{1}{T_k} \sum_t \log B_j^{kt} \right)}{\sum_{j'} \exp \left( \frac{1}{T_k} \sum_t \log B_{j'}^{kt} \right)} \tag{5}$$

The geometric mean has the further advantage to give more weight to informative sites, for which the probabilities for each segregation type are very different (say, 0.1 and $10^{-5}$), than to sites with less information (say, 0.4 and 0.6). Thus, a site with all females heterozygous and all males homozygous, which would produce a much higher likelihood to be sex-linked than to be autosomal, has more weight than a site with one female heterozygous and all other individuals homozygous, a pattern compatible with both sex-linkage and autosomal segregation.

For completeness (*e.g.* to allow additional calibration by expert users), we provide two other ways to calculate the posterior probabilities per contig. First, we provide the posterior probability as the geometric mean of the site-wise probabilities calculated using the estimated genome proportions $\pi_j$ as priors (as in Equation 3):

$$\widehat{S}_{\mathrm{G}j}^k = \frac{\pi_j \mathrm{GM} \left( B_j^{kt} \right)}{\sum_{j'} \pi_{j'} \mathrm{GM} \left( B_{j'}^{kt} \right)} = \frac{\pi_j \exp \left( \frac{1}{T_k} \sum_t \log B_j^{kt} \right)}{\sum_{j'} \pi_{j'} \exp \left( \frac{1}{T_k} \sum_t \log B_{j'}^{kt} \right)} \tag{6}$$

Second, we provide the arithmetic mean of the expectations per site, $\widehat{S}_j^{kt}$ from Equation 3:

$$\widehat{S}_j^k = \frac{1}{T_k} \sum_t \widehat{S}_j^{kt} = \frac{1}{T_k} \sum_t \frac{\pi_j B_j^{kt}}{\sum_{j'} \pi_{j'} B_{j'}^{kt}} \tag{7}$$

We recommend that inferences of segregation types should be based on the posterior probabilities that were calculated without the priors, *i.e.* Equation 4 for sites and Equation 5 for contigs.

**Population genetic predictions** From the allele frequencies and segregation subtypes, it is possible to calculate the expected diversity and divergence of the gametologous copies. For each site, the frequency of allele $a$ on chromosome $v \in \{W, X, Y, Z\}$ is

$$\widehat{f}_v^{kt} = \widehat{A}_{j,l=1}^{kt} \left(1 - \widehat{f}_{j,l=1}^{kt}\right) + \widehat{A}_{j,l=2}^{kt} \widehat{f}_{j,l=2}^{kt} + \widehat{A}_{j,l=3}^{kt} \quad \text{for} \quad v \in \{X, Z\},$$

$$\widehat{f}_v^{kt} = \widehat{A}_{j,l=1}^{kt} + \widehat{A}_{j,l=3}^{kt} \left(1 - \widehat{f}_{j,l=3}^{kt}\right) + \widehat{A}_{j,l=4}^{kt} \widehat{f}_{j,l=4}^{kt} \quad \text{for} \quad v \in \{W, Y\}.$$

A different way of predicting the allele frequency on both sex chromosomes is to assign it to be the frequency corresponding to the most probable subtype.

This information can be used to infer the consensus sequences of the X and Y sequences. For a given contig (that can be chosen on the basis of $\widehat{S}_{\mathrm{N}j}^k$, but not necessarily if we have other reasons to believe the contig is sex-linked), each polymorphic site can be considered fixed for an allele if $\widehat{f}_X$ or $\widehat{f}_Y$ are above a threshold $U_f$ ($0.5 \leq U_f \leq 1$) or below $1 - U_f$. A further threshold can be applied to genotype non-fixed sites: if $\widehat{f}_X$ or $\widehat{f}_Y$ are above a threshold $u_f$ ($0.5 \leq u_f \leq U_f$) or below $1 - u_f$.

Nucleotide diversity (here denoted as $d$ because the usual symbol $\pi$ is already used for a model parameter) can be calculated as

$$d_v^k = \frac{1}{\tau_k} \sum_t 2\widehat{f}_v^{kt} \left(1 - \widehat{f}_v^{kt}\right)$$

where $\tau_k \geq T_k$ is the total length of the contig $k$, including monoallelic sites. The divergence is

$$D_{XY}^k = \frac{1}{\tau_k} \sum_t \left(\widehat{f}_X^{kt} \left(1 - \widehat{f}_Y^{kt}\right) + \widehat{f}_Y^{kt} \left(1 - \widehat{f}_X^{kt}\right)\right)$$

in the XY case; extension to ZW chromosomes is trivial.