

Supplementary Material

Here we show that it is indeed sufficient to replace $CM(m|g)$ by $CM_{off}(m|g)$ in all related steps of the algorithm. $CM_{off}(m|g)$ is defined below.

Denote by h_{off} the set of haplotypes that can be constructed by phasing the observed offspring genotypes g^{obs} , using all possibilities. If an offspring genotype is missing for a variant, all possible genotypes are used for this variant to construct the set h_{off} .

The algorithm of Horvath et al. proceeds using Steps 1-5 and computes the conditional distribution of the vectors of offspring genotypes under the null hypothesis, where the conditioning is on the sufficient statistic for all nuisance parameters in the model. The corresponding objects are denoted by $CM(m|g), \gamma, L_1, L_1^*, L_2, L_2^*$. $CM(m|g)$ is defined as the set of all phased mating types that are compatible with g and m . Below, we argue why it is sufficient to replace $CM(m|g^{obs})$ and $CM(m|g)$ by $CM_{off}(m|g^{obs})$ and $CM_{off}(m|g)$ in Step 1, 2a, 2b and 4 of the algorithm.

$CM_{off}(m|g)$, with $CM_{off}(m|g) \subseteq CM(m|g)$, is defined as the set of all compatible phased parental mating types where we used only haplotypes from h_{off} to complete the mating type. That means, if a parental genotype is missing, we use only haplotypes from h_{off} for this parent.

In general, a compatible phased parental mating type $M_k \in CM(m|g^{obs})$ can contain haplotypes that are not in h_{off} , but to be compatible with g^{obs} , both parents must have at least 1 haplotype from h_{off} .

If $M_k \in CM(m|g^{obs})$ and $M_k \notin CM_{off}(m|g^{obs})$, this implies that M_k contains at least one “unnecessary” haplotype regarding the compatibility with g^{obs} and m . By $M'_k \in CM_{off}(m|g^{obs})$ we

denote the phased parental mating type where we replaced these haplotypes by the other haplotype of the corresponding parent that is in h_{off} .

Let γ_{off} be the minimal set of offspring genotypes which is consistent with all phased compatible mating types in $CM_{off}(m|g^{obs})$, analogously defined as γ for $CM(m|g^{obs})$ in Horvath et al. (2004).

Claim 1: $\gamma_{off} = \gamma$.

Proof: Since $CM_{off}(m|g^{obs}) \subseteq CM(m|g^{obs})$ and γ as well as γ_{off} are defined as the intersection of the sets of possible offspring genotypes for all corresponding mating types, we obviously have $\gamma \subseteq \gamma_{off}$.

It remains to show that $\gamma_{off} \subseteq \gamma$. Assume that there is an element $\gamma^* \in \gamma_{off}$ with $\gamma^* \notin \gamma$. Since $\gamma^* \notin \gamma$, there must a mating type $M_k \in CM(m|g^{obs})$ such that γ^* is not a possible offspring genotype for M_k . Assume that $M_k \notin CM_{off}(m|g^{obs})$. M'_k (defined above) is in $CM_{off}(m|g^{obs})$ and the possible offspring genotypes for M'_k are a subset of the possible offspring genotypes for M_k . But then, we conclude that γ^* is a possible offspring genotype for M'_k and thus it must be a possible offspring genotype for M_k . A contradiction.

From claim 1, we know that L_1^* gives the same set of all possible offspring genotypes as before, since $\gamma_{off} = \gamma$. Step 2b of the algorithm generates L_1 from L_1^* by removing each vector of offspring genotypes g for which $CM(m|g^{obs}) \neq CM(m|g)$. Therefore, we need to show that we can replace the roles of $CM(m|g^{obs})$ and $CM(m|g)$ by $CM_{off}(m|g^{obs})$ and $CM_{off}(m|g)$.

Claim 2: In step 2 of the algorithm:

$$CM(m | g^{obs}) = CM(m | g) \Leftrightarrow CM_{off}(m | g^{obs}) = CM_{off}(m | g).$$

Proof:

$$1.) CM(m | g^{obs}) = CM(m | g) \Rightarrow CM_{off}(m | g^{obs}) = CM_{off}(m | g)$$

$M_k \in CM_{off}(m | g^{obs})$ leads to $M_k \in CM(m | g^{obs})$ and thus $M_k \in CM(m | g)$ by assumption.

Since m is fixed, we have $M_k \in CM_{off}(m | g)$, by construction. The same argumentation holds in the

other direction and shows overall, that $CM_{off}(m | g^{obs}) = CM_{off}(m | g)$ if

$$CM(m | g^{obs}) = CM(m | g).$$

$$2.) CM(m | g^{obs}) = CM(m | g) \Leftarrow CM_{off}(m | g^{obs}) = CM_{off}(m | g)$$

Choose $M_k \in CM(m | g^{obs})$. If $M_k \in CM_{off}(m | g^{obs})$, the result follows immediately. Therefore,

assume $M_k \notin CM_{off}(m | g^{obs})$. Then, $M'_k \in CM_{off}(m | g^{obs})$ and therefore $M'_k \in CM_{off}(m | g)$. But

if $M'_k \in CM_{off}(m | g)$, then $M_k \in CM(m | g)$.

For the exchanged roles of g and g^{obs} , we fix a $M_k \in CM(m | g)$. The cases of no parental genotypes

observed and both parental genotypes observed are straightforward. We consider the scenario where

only one parental genotype is observed and can assume that $CM_{off}(m | g^{obs}) = CM_{off}(m | g) \neq \emptyset$,

otherwise there is nothing to compute. From this we can derive that we can construct a $M_k^* \in$

$CM_{off}(m | g) = CM_{off}(m | g^{obs})$, analogously to the M'_k construction. But then, we can also conclude

that $M_k \in CM(m | g^{obs})$.

This shows that $CM(m | g^{obs}) = CM(m | g)$ if $CM_{off}(m | g^{obs}) = CM_{off}(m | g)$.

Step 3 computes the list L_2^* and the matrix A . In Step 4, the algorithm identifies the constant rows in the matrix A . For an element of L_2^* , the corresponding row in the matrix A describes the ratio of conditional probabilities for this genotype configuration, given all phased compatible mating types in $CM(m | g^{obs})$. We need to show:

Claim 3: A row in the matrix A of the original algorithm is constant if and only if it is constant along all phased compatible mating types in $CM_{off}(m | g^{obs})$.

Proof: Obviously, if the row is constant along all phased compatible mating types in $CM(m | g^{obs})$, it is also constant along all phased compatible mating types in $CM_{off}(m | g^{obs})$, as a subset of $CM(m | g^{obs})$.

Assume the ratio is equal to constant c along all phased compatible mating types in $CM_{off}(m | g^{obs})$ and there is a mating type $M_k \in CM(m | g^{obs})$, with $M_k \notin CM_{off}(m | g^{obs})$, where the ratio is not equal to this c .

We know that the ratio is equal to c for the modified mating type M_k' , since $M_k' \in CM_{off}(m | g^{obs})$.

But the conditional probability of a specific genotype configuration given M_k and given M_k' can only differ by the same factor in the numerator and denominator of the ratio by Mendel's laws and is not equal to zero by construction, a contradiction.