File S1

Contents

A Estimator based on IBS

For clarity of exposition, here we derive results for $\widehat{\text{IBS}}_m$ under a simple model that assumes no genotyping error (a more general result that includes genotyping error can be found in Appendix B, equation (B.9)). The simple model assumes that the IBD state at the t-th locus, IBD_t, is Bernoulli with relatedness parameter $r \in [0, 1]$. Given $\text{IBD}_t = 0$, we assume that $Y_t^{(i)}$ and $Y_t^{(j)}$ are independent Categorical variables with parameter $(f_t(g))_{g \in \mathcal{G}_t}$. Given $\text{IBD}_t = 1$, we assume that $Y_t^{(i)}$ follows a Categorical distribution with parameter $(f_t(g))_{g \in \mathcal{G}_t}$ and that $Y_t^{(j)} = Y_t^{(i)}$ with probability one.

A.1 Expectation of estimator based on IBS

In this section no assumptions are made about dependence between marker loci: equation (A.1) holds under both independence and dependence. The expectation of the estimator $\widehat{\text{IBS}}_m$ conditional

on the frequencies $(f_t(g))_{g\in\mathcal{G}_t}\forall t = 1,\ldots,m$ is

$$\begin{aligned}
\mathbb{E}[\widehat{\text{IBS}}_m] &= \frac{1}{m}\sum_{t=1}^{m}\mathbb{E}\left[\text{IBS}_t\right], \\
&= \frac{1}{m}\sum_{t=1}^{m}\mathbb{P}\left(\text{IBS}_t = 1 \mid \text{IBD}_t = 1\right)\mathbb{P}\left(\text{IBD}_t = 1\right) + \mathbb{P}\left(\text{IBS}_t = 1 \mid \text{IBD}_t = 0\right)\mathbb{P}\left(\text{IBD}_t = 0\right), \\
&= \frac{1}{m}\sum_{t=1}^{m}\{r + \sum_{i=1}^{K_t} f_t(g_i)^2(1-r)\}, \\
&= r + \bar{h}_m(1-r), \\
&= \bar{h}_m + (1-\bar{h}_m)r, \qquad \text{(A.1)}
\end{aligned}$$

where $\bar{h}_m = m^{-1}\sum_{t=1}^{m}\sum_{i=1}^{K_t} f_t(g_i)^2$. Under different observation models, we would still obtain $\mathbb{E}[\widehat{\text{IBS}}_m]$ as a linear function of r; see second line above, where $\mathbb{P}\left(\text{IBS}_t = 1 \mid \text{IBD}_t = 1\right)$ and $\mathbb{P}\left(\text{IBS}_t = 1 \mid \text{IBD}_t = 0\right)$ could be anything as long as these expressions do not involve r.

A.2 Convergence of estimator based on IBS

Here we work under the simplest setting: the measurements $(Y_t^{(i)}, Y_t^{(j)})$ are independent across $t = 1,\ldots,m$. In order to discuss convergence we need to imagine an asymptotic regime where $m \to \infty$. We introduce an infinite sequence $(f_t(g_i))_{t\geq 1}, i = 1,\ldots,K_t$, where each $f_t(g)$ is in $(0,1)$, and we introduce $\bar{h} = \lim_{m\to\infty} m^{-1}\sum_{t=1}^{m}\sum_{i=1}^{K_t} f_t(g_i)^2$, assuming the existence of that limit. To show that $\widehat{\text{IBS}}_m$ is not consistent for r, we show that it is consistent for $\bar{h} + (1-\bar{h})r$, which is different to r unless $r = 1$. Thus we show that $\widehat{\text{IBS}}_m$ satisfies,

$$\widehat{\text{IBS}}_m \xrightarrow[m\to\infty]{\mathbb{P}} \bar{h} + (1-\bar{h})r, \qquad \text{(A.2)}$$

where the arrow is interpreted as "convergence in probability". Since $\mathbb{E}[\widehat{\text{IBS}}_m] = \bar{h}_m + (1-\bar{h}_m)r \to \bar{h} + (1-\bar{h})r$ as $m \to \infty$, we can establish (A.2) by showing that for every $\varepsilon > 0$

$$\mathbb{P}\left(\mid \widehat{\text{IBS}}_m - \mathbb{E}\left[\widehat{\text{IBS}}_m\right] \mid > \varepsilon\right) \to 0 \text{ as } m \to \infty. \qquad \text{(A.3)}$$

We show equation (A.3) by use of Hoeffding's inequality (see Chapter 4 in (Wasserman 2013)). Since $\widehat{\text{IBS}}_m$ is an average of variables IBS_t, which are bounded ($\text{IBS}_t \in \{0,1\}$) and assumed independent, Hoeffding's inequality yields

$$\mathbb{P}\left(\mid \widehat{\text{IBS}}_m - \mathbb{E}\left[\widehat{\text{IBS}}_m\right] \mid \geq \varepsilon\right) \leq 2\exp\left(-2m\varepsilon^2\right). \qquad \text{(A.4)}$$

Since $2\exp\left(-2m\varepsilon^2\right) \to 0$ as $m \to \infty$, equation (A.4) shows that equation (A.3) holds and therefore that equation (A.2) holds. Note that consistency could also be established in the dependent case, for instance via the application of a version of Hoeffding's inequality for dependent processes.

Plots of $\widehat{\text{IBS}}_m$ for data simulated under the independence model (Figure A.1) numerically show for $r = 0$ and 0.5 that $\widehat{\text{IBS}}_m$ concentrates on its expectation (equation (A.1)) as more and more markers ($m = 24$, 96 and 192) are typed.

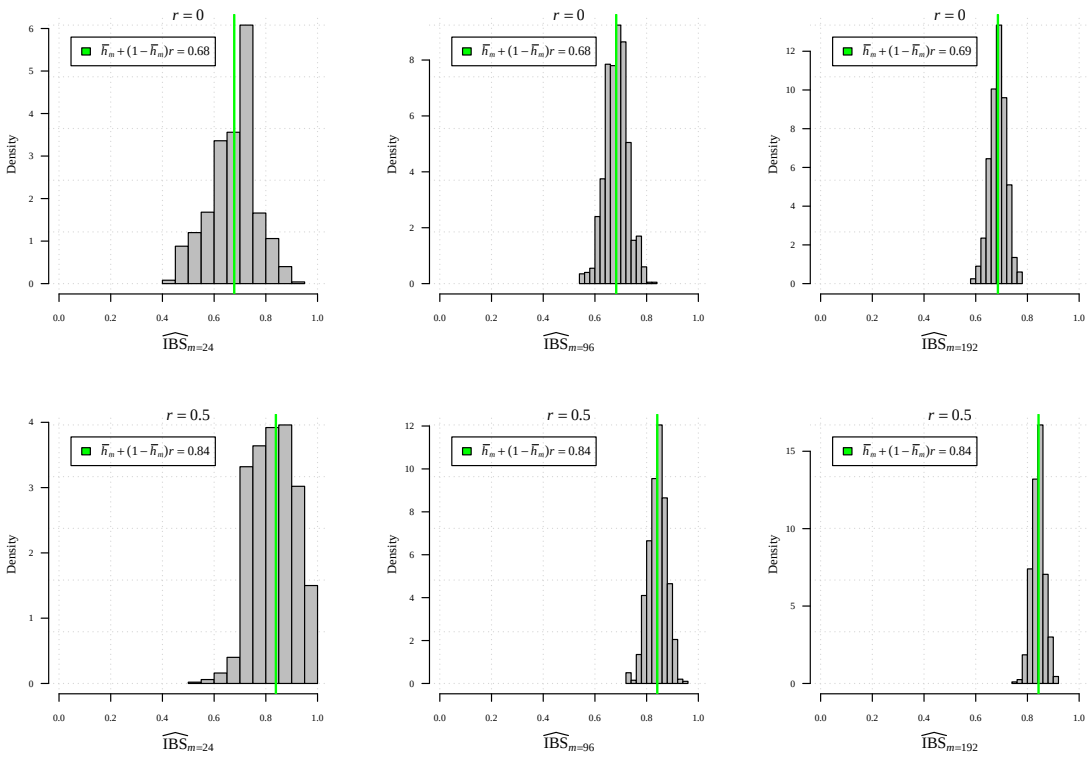


Figure A.1: Distributions of $\widehat{\text{IBS}}_m$ between pairs of biallelic marker data simulated under the independence model with different numbers of markers, m, and relatedness, r. Each distribution is based on 1000 simulated pairs. The green vertical line marks $\bar{h}_m + (1 - \bar{h}_m)r$ which is a function of the allele frequencies (equation 3, main text). Allele frequencies were sampled without replacement from the Thai WGS dataset with probability proportional to minor allele frequency estimates.

A.3 Corrected estimator based on IBS

A corrected version of the estimator $\widehat{\text{IBS}}_m$ could be consistent for r (equation (A.7)) and is similar to existing method of moments estimators (reviewed in Bink *et al.* (2008)), which generally underperform compared to maximum likelihood estimators (Chapter 9 of Wasserman (2013)).

By rearranging equation (A.1),

$$r = \frac{1}{(1 - \bar{h}_m)} \left(\mathbb{E}\left[\widehat{\text{IBS}}_m \right] - \bar{h}_m \right), \tag{A.5}$$

we can propose the following corrected estimator of r,

$$\widehat{\text{IBS}}_m^{(c)} = \frac{1}{(1 - \bar{h}_m)} \left(\widehat{\text{IBS}}_m - \bar{h}_m \right), \tag{A.6}$$

whose expectation is precisely r. The corrected estimator $\widehat{\text{IBS}}_m^{(c)}$ is consistent for r, with the same reasoning as in Appendix A.2 assuming independent observations,

$$\widehat{\text{IBS}}_m^{(c)} = \frac{1}{(1-\bar{h}_m)}\left(\widehat{\text{IBS}}_m - \bar{h}_m\right) \xrightarrow[m\to\infty]{\mathbb{P}} \frac{1}{(1-\bar{h})}(\bar{h} + (1-\bar{h})r - \bar{h}) = r. \tag{A.7}$$

Figure A.2 shows a plot of equation (A.6) for different values of $\bar{h}_m \in (0.5, 1)$. The range of $\widehat{\text{IBS}}_m^{(c)}$ includes negative values. Setting negative estimates to zero can considerably improve results Bink *et al.* (2008), but can also introduce bias Huang *et al.* (2015). For the *Plasmodium* datasets considered in the main text, Figure A.3 shows $\widehat{\text{IBS}}_m^{(c)}$ estimates truncated to [0,1].

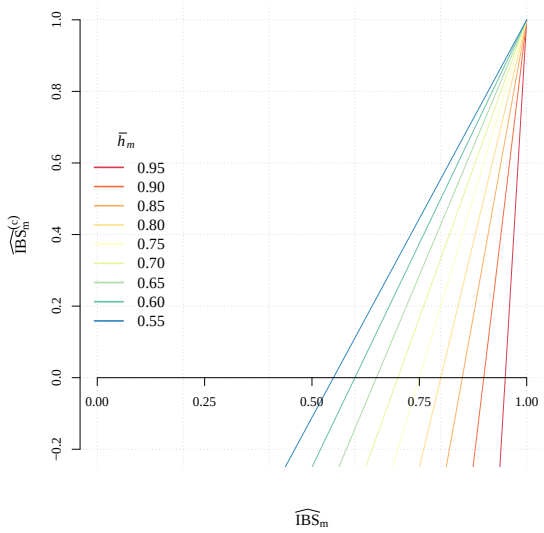


Figure A.2: $\widehat{\text{IBS}}_m^{(c)}$ as a function of $\widehat{\text{IBS}}_m$ for various $\bar{h}_m$ (equation (A.6)).

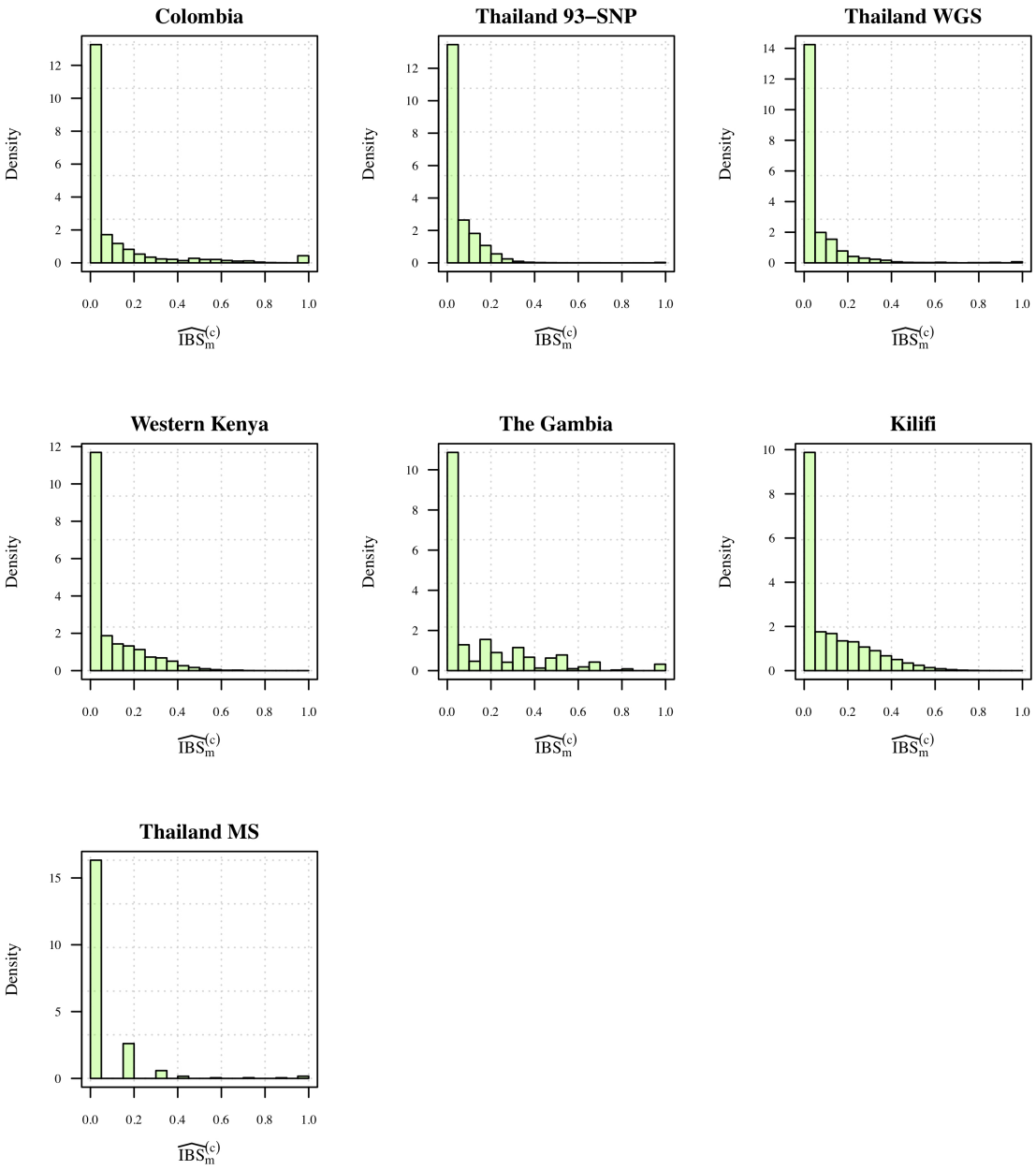


Figure A.3: $\widehat{\mathrm{IBS}}_m^{(c)}$ for several monoclonal *Plasmodium* datasets.

B Model-based estimation of relatedness

B.1 Framework

In this section we describe models that relate the available data to the objects of interest, in a self-contained presentation. The data comprise frequencies of alleles denoted by $(f_t(g))_{g\in\mathcal{G}_t}$, and allele indicators $Y_t^{(i)}$, where the index t denotes a locus on the genome, and the superscript (i) refers to the i-th individual. The index t will run from 1 to m, the number of markers genotyped, and we will be particularly interested in the impact of m and K_t on the precision of the estimators. Note that m cannot be larger than L, the total length of the genome, which will create difficulties in making sense of an asymptotic regime where m goes to infinity, as will be discussed below.

We will consider pairs of individuals, i and j, for which we want to estimate the relatedness denoted by r and taking values in the interval $[0, 1]$. The models below might involve other parameters, and overall the vector of parameters is denoted by θ. We will make the first component of θ represent the relatedness r, so that $r = \theta_1$.

For each pair of individuals, we introduce a sequence of latent binary variables denoted by (IBD_t) for identity-by-descent: $\mathrm{IBD}_t = 1$ indicates identity-by-descent at locus t. We view this sequence as a two-state Markov chain. The case of independent variables for (IBD_t) constitutes a particular case. In any case, the relatedness $r \in [0, 1]$ represents the marginal probability that IBD_t is equal to one, assumed to be identical for all t. While we do not observe (IBD_t), we observe $Y_t^{(i)}$ and $Y_t^{(j)}$ that are related to IBD_t at locus t via an observation model, which can take into account the presence of genotyping errors. Together, the specification of the latent process (IBD_t) and of the observation model fully describes a hidden Markov model, that can be used to estimate r using the data. Complete model specification is deferred to Appendix B.3, after a description of the general estimation procedure and some specific issues arising in the present case.

The estimation procedure is here based on the maximum likelihood approach. The likelihood function can be written as

$$\mathcal{L}_{1:m}(\theta) = \prod_{t=1}^{m} \mathbb{P}(Y_t^{(i)}, Y_t^{(j)} | \mathcal{Y}_{t-1}, \theta),$$

where $\mathcal{Y}_{t-1}$ represents all the observations from locus 1 to locus $t-1$, with the convention that $\mathcal{Y}_0$ is the empty set. We can further write each "incremental likelihood term" as

$$\mathbb{P}(Y_t^{(i)}, Y_t^{(j)} | \mathcal{Y}_{t-1}, \theta) = \sum_{\mathrm{IBD}_t \in \{0,1\}} \mathbb{P}(Y_t^{(i)}, Y_t^{(j)} | \mathrm{IBD}_t, \theta) \mathbb{P}(\mathrm{IBD}_t | \mathcal{Y}_{t-1}, \theta).$$

Since (IBD_t) is a Markov chain, the forward algorithm (Rabiner 1989) can be used to evaluate each incremental likelihood term for $t = 1, \ldots, m$, for a cost of the order of m operations given θ.

We write $\ell_{1:m}(\theta) = \log \mathcal{L}_{1:m}(\theta)$, and $\ell_t(\theta) = \log \mathbb{P}(Y_t^{(i)}, Y_t^{(j)} | \mathcal{Y}_{t-1}, \theta)$. We denote the first and second derivatives of $\ell_t(\theta)$ by $\ell_t'(\theta)$ (a vector) and $\ell_t''(\theta)$ (a matrix) respectively. We will use the maximum likelihood estimator to approximate r, and we define it as

$$\hat{\theta}_m = \mathrm{argmax}_\theta \, \ell_{1:m}(\theta).$$

We next review some asymptotic properties of the maximum likelihood estimator (MLE) and detail how the present setting differs from the one usually considered in asymptotic studies.

B.2 Distribution of the MLE

B.2.1 Standard asymptotic theory

We first recall what the usual asymptotic reasoning is for the distribution of the MLE in HMMs (Douc and Moulines 2012), in informal terms.

The first step is to imagine that the variables indexed by t (such as IBD_t, $Y_t^{(i)}$, $Y_t^{(j)}$, etc.) are part of infinite sequences of variables indexed by $t \geq 1$. This allows us to consider a regime where the number of loci considered m can go to ∞. In Appendix B.2.2 we will discuss issues arising when applying this asymptotic reasoning in the present context of genetic data.

We observe that the log-likelihood and its derivatives are sums of m terms. Dividing by m yields averages, which might converge to limiting values as m grows large. For instance, the scaled log-likelihood might satisfy

$$\forall \theta \quad m^{-1}\ell_{1:m}(\theta) \xrightarrow[m\to\infty]{\mathbb{P}} \bar{\ell}(\theta),$$

where the arrow is to be interpreted as "convergence in probability", the left hand side of it being random if we consider the data to be random. Under some assumptions, the maximizer $\hat{\theta}_m$ of $\theta \mapsto m^{-1}\ell_{1:m}(\theta)$ converges to the maximizer $\theta^\star$ of the limiting function $\theta \mapsto \bar{\ell}(\theta)$. By the Taylor expansion of $\ell'_{1:m}(\hat{\theta}_m)$ at $\theta^\star$ we have

$$\ell'_{1:m}(\hat{\theta}_m) = \ell'_{1:m}(\theta^\star) + \ell''_{1:m}(\theta^\star)(\hat{\theta}_m - \theta^\star) + \text{rest}. \tag{B.1}$$

At the MLE $\hat{\theta}_m$, the derivative of the log-likelihood cancels: $\ell'_{1:m}(\hat{\theta}_m) = 0$, at least if the MLE is in the interior of the parameter space; extra care is required when the MLE is on the boundary of the parameter space, which occurs in the present setting where $\hat{r}_m$ can be exactly zero or one. Therefore we obtain

$$\begin{aligned} 0 &\approx \ell'_{1:m}(\theta^\star) + \ell''_{1:m}(\theta^\star)(\hat{\theta}_m - \theta^\star), \\ \Leftrightarrow (\hat{\theta}_m - \theta^\star) &\approx -\ell''_{1:m}(\theta^\star)^{-1}\ell'_{1:m}(\theta^\star), \qquad \text{(B.2)} \\ \Leftrightarrow \sqrt{m}(\hat{\theta}_m - \theta^\star) &\approx \left(-m^{-1}\ell''_{1:m}(\theta^\star)\right)^{-1} m^{-1/2}\ell'_{1:m}(\theta^\star), \qquad \text{(B.3)} \end{aligned}$$

where $\Leftrightarrow$ means "equivalently".

We will rely on the two following convergence results (see Chapter 13 in Douc *et al.* (2014)),

$$m^{-1/2}\ell'_{1:m}(\theta^\star) \xrightarrow[m\to\infty]{\mathrm{d}} \mathcal{N}\left(0, V^\star\right), \tag{B.4}$$

$$-m^{-1}\ell''_{1:m}(\theta^\star) \xrightarrow[m\to\infty]{\mathbb{P}} J^\star, \tag{B.5}$$

for some matrices $V^\star, J^\star$, assumed to be both semi-definite positive and symmetric. The first line above describes a convergence "in distribution" and can follow from a central limit theorem for the first derivative of the log-likelihood. The second line can follow from a law of large numbers applied to the second derivatives, as in Chapter 13 of Douc *et al.* (2014). We can combine these two convergence results using Slutsky's lemma to obtain the asymptotic normality of the MLE:

$$\sqrt{m}(\hat{\theta}_m - \theta^\star) \xrightarrow[m\to\infty]{\mathrm{d}} \mathcal{N}\left(0, (J^\star)^{-1}V^\star(J^\star)^{-1}\right). \tag{B.6}$$

This key result can be used for sample size determination and for the construction of confidence intervals, provided that we can approximate $\theta^\star$, $V^\star$ and $J^\star$ based on data. The asymptotic variance

$(J^\star)^{-1}V^\star(J^\star)^{-1}$ is sometimes called the sandwich formula, and can be estimated based on samples; see Doucet and Shephard (2012) in the setting of hidden Markov models. If we assume that the model is well-specified, i.e. that the data actually are generated from the model with the parameter $\theta^\star$, then it can be shown that $J^\star = V^\star$ under regularity conditions (Chapter 13 of Douc *et al.* (2014)). In this case, the asymptotic variance in (B.6) simplifies to $(J^\star)^{-1}$. The matrix $J^\star$ is often termed the Fisher Information Matrix at $\theta^\star$.

We briefly discuss the numerical obtention of $\hat{\theta}_m = \text{argmax}_\theta \ell_{1:m}(\theta)$. The log-likelihood function $\theta \mapsto \ell_{1:m}(\theta)$ can be plugged in a numerical optimizer, such as that implemented in the `optim` function of `R`. Evaluations of the log-likelihood function require runs of the forward algorithm on the data, for a cost of the order of m operations. Alternatively, one can also run an expectation-maximization algorithm, which involves calculating expectations with respect to the distribution of the latent process (IBD_t) using the forward-backward algorithm (Cappé *et al.* 2005), also called Baum-Welch in the context of HMMs (Rabiner 1989). If the parameter is small-dimensional, e.g. one or two-dimensional, a simple way of approximating the MLE consists in evaluating the likelihood (using the forward algorithm) on a grid of parameter values, and selecting the parameter associated with the highest likelihood.

The matrix $J^\star$ can be estimated by $-m^{-1}\ell''_{1:m}(\hat{\theta}_m)$, itself computed via numerical differentiation of the log-likelihood function at $\hat{\theta}_m$. The estimation of $V^\star$ is more complicated and has been the topic of a rich literature in time series analysis; see for instance Doucet and Shephard (2012) and references therein.

B.2.2 Applicability of the standard asymptotic theory

The law of large numbers and central limit theorems usually employed to carry out the above reasoning, i.e. to establish (B.4) and (B.5) leading to the asymptotic normality of the MLE in (B.6), might not be meaningful in the present context. Indeed they usually apply to stationary processes observed over increasingly long periods of time. In such asymptotic setting, one eventually observes a realization of a stationary stochastic process over an infinitely long time horizon, which is enough to learn the invariant distribution of the process. We refer to this setting as standard asymptotics. Recall that our primary object of interest is the parameter r, which characterizes indeed the invariant distribution of the Markov chain (IBD_t).

In the present setting where data comprise genetic sequences, increasing m means considering more loci on the genome. The m considered loci are located within the genome whose length is, however, fixed. Therefore increasing m amounts to increasing the subsampling frequency at which data are observed. In other words it decreases the distance between successive observed loci. We refer to this as subsampling asymptotics. To see where this differs from standard asymptotics, consider a simpler context where (IBD_t) would not be hidden but directly observed. In the limit $m \to \infty$ in subsampling asymptotics, we would observe a continuous trajectory of (IBD_t), switching from state 0 to state 1 and back again, over a fixed interval. The maximum likelihood estimate of r for such a model would be the proportion of time that the trajectory would spend in state 1 (Bladt and Sørensen 2005). However this would not be exactly equal to r, even if the trajectory was sampled from the Markov model given r, because the fully-observed realization of (IBD_t) would still be of a finite length; this is well-known, see (Hill and Weir 2011) on the impact of the genome length on relatedness estimates under Mendelian sampling. On the other hand, in the standard asymptotics $m \to \infty$ we would observe an infinitely long trajectory of the Markov chain, for which the maximum likelihood estimator of the transition matrix is consistent. The difference between

the two regimes is illustrated in Figure B.1.

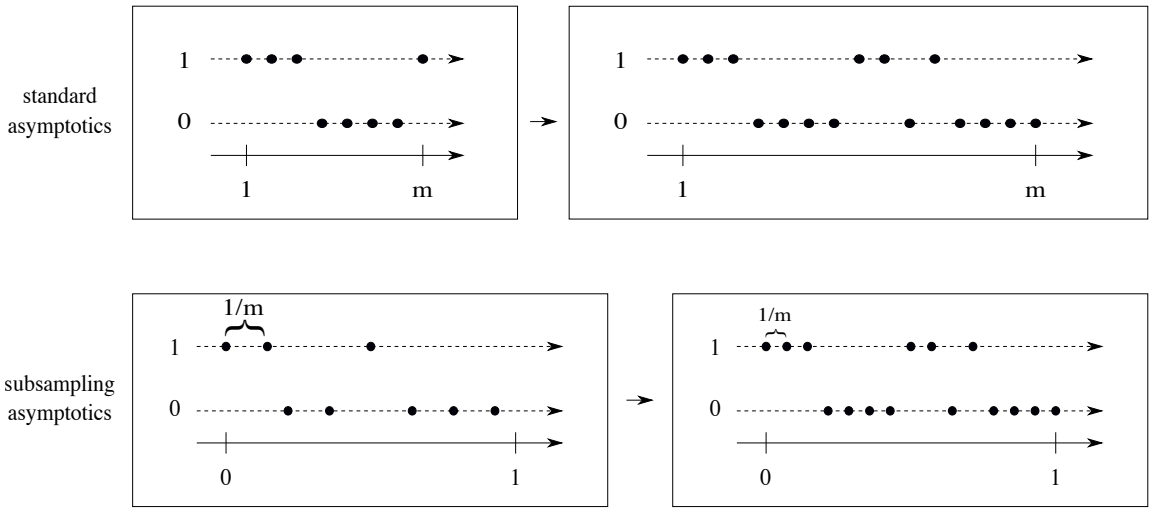


Figure B.1: Two different ways of increasing m: in the top row, m refers to the length of the observation period, while the observations are separated by one unit of time. In the bottom row, the length of the observation interval is fixed to one, and the observations are placed at distance $1/m$ of one another; thus an increase in m means that successive observations are closer to one another, but the length of the observation period is fixed.

The difference in asymptotic regimes has consequences on the estimability of r. In the subsampling asymptotics, it is impossible to arbitrarily decrease the error of $\hat{r}_m$ by increasing m: there is only so much information that can be gathered about r by increasing the number of loci under consideration; hence the distinction between expected IBD and realised IBD in (Speed and Balding 2015). A result such as the asymptotic normality with a $\sqrt{m}$ rate of convergence, as in (B.6), is in fact unlikely to hold. The numerical experiments indeed suggest that the root mean squared error associated with $\hat{r}_m$ does not decrease beyond a certain point, no matter how large m is. The subsampling asymptotic regime has been formally studied with various applications to financial econometrics (Aït-Sahalia 2002; Barndorff-Nielsen *et al.* 2006), but we are not aware of similar results for hidden Markov models such as the ones considered here.

Despite the standard asymptotic results not holding, we do observe that the distribution of $\widehat{r}_m$ is approximately Normal for m large enough (Figure B.2). This can be partially explained by the fact that normality of the MLE depends entirely on the log-likelihood being approximately quadratic (Geyer 2013), which itself does not have to follow from standard asymptotic arguments. Since the log-likelihood function is observed to be approximately quadratic providing $\widehat{r}_m$ is not close to the boundaries (Figure B.3), we can still quantify the precision of the MLE by considering the second derivative of the log-likelihood at its maximum. Thus we will rely on the Fisher Information Matrix as a proxy for the precision of the MLE, in particular for the study of the effect of K_t in Appendix B.3.4.

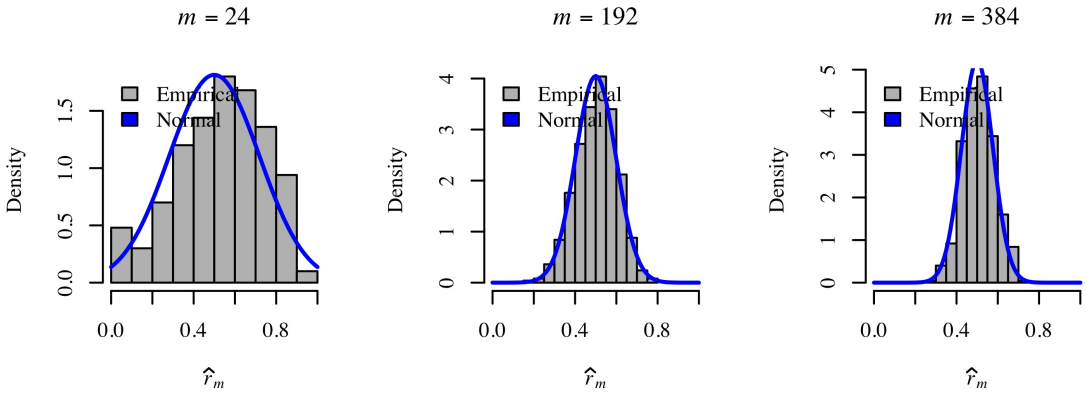


Figure B.2: Empirical distributions of $\hat{r}_m$ for different numbers of markers, m. Each distribution is based on 500 estimates of r given data simulated and analyzed under the HMM with $r = 0.5$, $k = 8$, $K_t = 2\ \forall t = 1 \ldots, m$ and $\varepsilon = 0.001$.

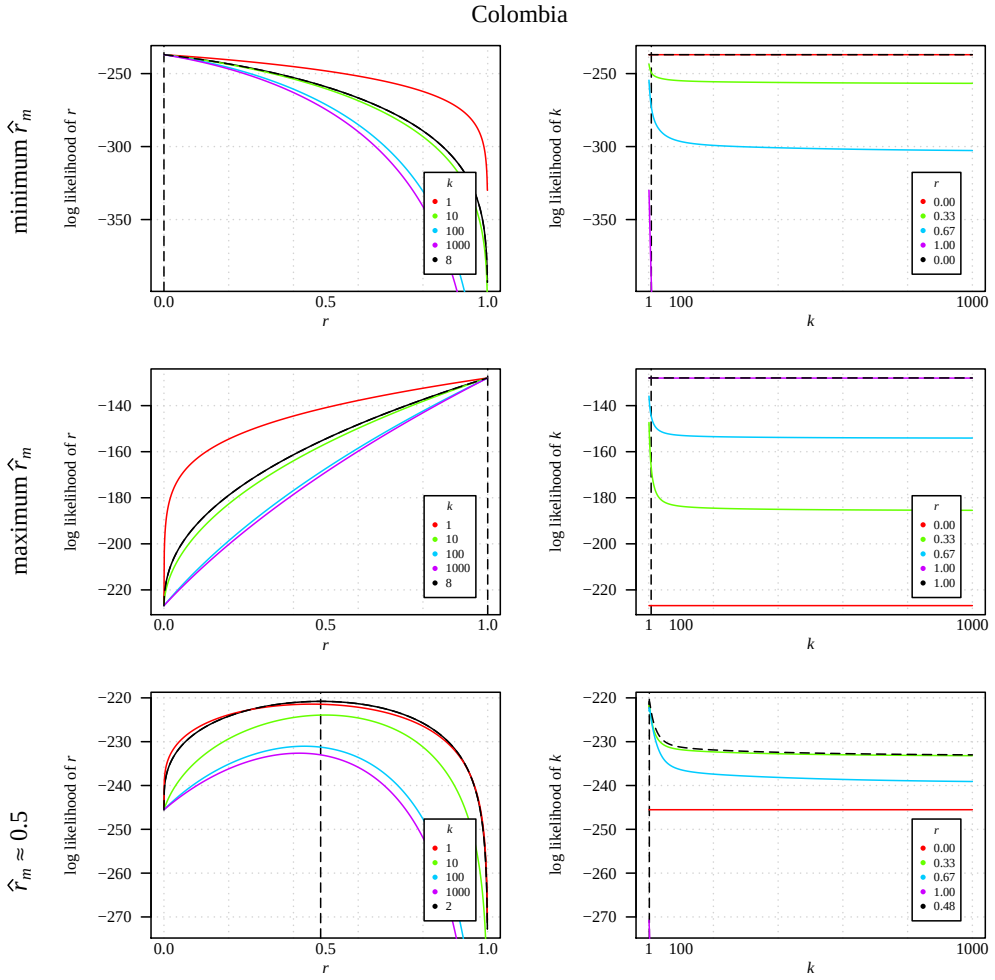


Figure B.3: The log-likelihoods of r for different k (left column) and k for different r (right column) for three different example sample pairs from the Colombian dataset: a sample pair with minimum $\widehat{r}_m$ (top row, $m = 248$), a sample pair with maximum $\widehat{r}_m$ (middle row, $m = 246$), and a sample pair with $\widehat{r}_m \approx 0.5$ (bottom row, $m = 245$). Differences in m are due to missing genotype calls in the data. Vertical black dashed lines mark $\widehat{r}_m$ (left column) and $\widehat{k}_m$ (right column). Black dashed function lines show the log-likelihood of $\widehat{r}_m$ given $\widehat{k}_m$ (left column) and of $\widehat{k}_m$ given $\widehat{r}_m$ (right column). Coloured function lines show the log-likelihood of $\widehat{r}_m$ given values of $k \neq \widehat{k}_m$ (left column) and of $\widehat{k}_m$ given values of $r \neq \widehat{r}_m$ (right column). Where the likelihood of k given $\widehat{r}_m$ is flat, the numerical optimizer, `optim`, returns the initial value 8 as $\widehat{k}_m$.

B.3 Models

We now describe a Markov chain model for (IBD_t), followed by observation models for $Y_i^{(t)}$ and $Y_j^{(t)}$ given IBD_t.

B.3.1 Hidden Markov model

We write the transition probabilities of (IBD_t) at a locus t,

$$A(t) = \begin{pmatrix} a_{00}(t) & a_{01}(t) \\ a_{10}(t) & a_{11}(t) \end{pmatrix} = \begin{pmatrix} 1 - r(1 - \exp(-k\rho d_t)) & r(1 - \exp(-k\rho d_t)) \\ (1-r)(1 - \exp(-k\rho d_t)) & 1 - (1-r)(1 - \exp(-k\rho d_t)) \end{pmatrix}.$$

In the above, $a_{j\ell}(t)$ refers to the probability of $\text{IBD}_t = \ell$ given that $\text{IBD}_{t-1} = j$.

In the above expression, the relatedness is denoted by r; d_t denotes a genetic distance in base pairs (bp) between loci $t-1$ and t; $k > 0$ parametrizes the switching rate of the Markov chain and ρ is the recombination rate, assumed fixed across both haploid genotypes with value 7.4×10^{-7}M bp^{-1} for *P. falciparum* parasites (Miles *et al.* 2016).

We can check that, if $\mathbb{P}(\text{IBD}_{t-1} = 1) = r$, then

$$\mathbb{P}(\text{IBD}_t = 1) = \mathbb{P}(\text{IBD}_{t-1} = 1)a_{11}(t) + \mathbb{P}(\text{IBD}_{t-1} = 0)a_{01}(t) = r,$$

and thus the invariant marginal distribution of the chain is given by $\mathbb{P}(\text{IBD}_t = 1) = r$.

The above transition probabilities are at the core of many HMMs of relatedness (e.g. Leutenegger *et al.* (2003), where $k \times \rho = a$ and genetic distance $d_t = t_k$ is measured in cM, plus many subsequent models (see (Brown *et al.* 2012)), including Schaffner *et al.* (2018), where $r = \pi_1$ and $1 - r = \pi_2$.

We can check that, as the distance increases to infinity, the probabilities in $A(t)$ simplify and correspond to the independence Bernoulli model where IBD_t is equal to one with probability r, independently for each locus t. In other words, if loci are distant enough, we expect the HMM and the independence model to give similar results. This will happen in particular when m is small and when the loci under consideration are well-spread across the genome.

B.3.2 Observation model

The observations $Y_t^{(i)}, Y_t^{(j)}$ are related to (IBD_t) only through IBD_t at locus t. The observation model introduces some true genotypes $G_t^{(i)}, G_t^{(j)}$ given IBD_t, and then some genotyping error model defining the distribution of $Y_t^{(i)}, Y_t^{(j)}$ given $G_t^{(i)}, G_t^{(j)}$.

First, the variables $G_t^{(i)}, G_t^{(j)}$ given IBD_t are defined as follows. If $\text{IBD}_t = 0$, then $G_t^{(i)}$ is independent of $G_t^{(j)}$ and both follow a Categorical distribution: for a set of values $\mathcal{G} = \{g(1), \ldots, g(K_t)\}$ and probabilities $\{f_t(g)\}$ for $g \in \mathcal{G}$, we have $\mathbb{P}(G_t^{(i)} = g) = f_t(g)$, and likewise for $G_t^{(j)}$. If there are only two types (e.g. the case for biallelic SNPs) then it is a Bernoulli distribution. If $\text{IBD}_t = 1$, then $\mathbb{P}(G_t^{(i)} = g) = f_t(g)$ and $G_t^{(j)} = G_t^{(i)}$ with probability one. Overall we can write the model as

$$\mathbb{P}(G_t^{(i)} = g^{(i)}, G_t^{(j)} = g^{(j)} | \text{IBD}_t = 0) = f_t(g^{(i)}) f_t(g^{(j)})$$
$$\mathbb{P}(G_t^{(i)} = g^{(i)}, G_t^{(j)} = g^{(j)} | \text{IBD}_t = 1) = f_t(g^{(i)}) \mathbb{1}(g^{(i)} = g^{(j)}).$$

Next, we assume that genotyping errors occur independently for both individuals. This differs to the typical 'all-or-none' diploid setting (e.g. (Leutenegger *et al.* 2003; Brown *et al.* 2012)), since

haploid genotypes in monoclonal parasite samples are genotyped separately. If they occur, we do not observe $Y_t^{(i)} = G_t^{(i)}$ but instead we observe another genotype taken uniformly among the other possible values (by assumption); and likewise for the other individual j. This can be written

$$\mathbb{P}(Y_t^{(i)} = g^{(i)} | G_t^{(i)} = g) = \begin{cases} 1 - (K_t - 1)\varepsilon & \text{if } g^{(i)} = g, \\ \varepsilon & \text{if } g^{(i)} \neq g. \end{cases}$$

In the above expression K_t refers to the number of possibilities, which could be different for different loci t, and ε refers to a parameter such that the error rate is $(K_t - 1)\varepsilon$. This is suited to microsatellites in the sense that the error rate scales with K_t (Hoffman and Amos 2005). For biallelic SNPs, it amounts to a simple miscall.

Overall we can thus think of the observation model as the combination of a model for $(Y_t^{(i)}, Y_t^{(j)})$ given $(G_t^{(i)}, G_t^{(j)})$ and a model for $(G_t^{(i)}, G_t^{(j)})$ given IBD_t. We can integrate $G_t^{(i)}, G_t^{(j)}$ out to obtain directly the probabilities of $(Y_t^{(i)}, Y_t^{(j)})$ given IBD_t:

$$\mathbb{P}(Y_t^{(i)} = g^{(i)}, Y_t^{(j)} = g^{(j)} | \text{IBD}_t) \tag{B.7}$$

$$= \sum_{g, g' \in \mathcal{G}} \mathbb{P}(Y_t^{(i)} = g^{(i)} | G_t^{(i)} = g) \mathbb{P}(Y_t^{(j)} = g^{(j)} | G_t^{(j)} = g') \mathbb{P}(G_t^{(i)} = g, G_t^{(j)} = g' | \text{IBD}_t). \tag{B.8}$$

The cost of evaluating this expression is quadratic in the cardinality of $\mathcal{G}$.

This observation model is the same (besides notation) as that for within-population samples under the HMM of hmmIBD (Schaffner *et al.* 2018) and, if $K_t = 2$, the same as that of isoRelate (Henden *et al.* 2018). Mutations do not feature in it. However, any that do occur can be absorbed as errors, as they are considered to be in Wang (2004). That said, it does not take into account microsatellite mutations in the sense that they scale with both motif size and repeat number (McDew-White *et al.* 2019), nor their inherent ordinal nature or the bias with regards to their amplification (Messerli *et al.* 2017). Bespoke adaptations could be made for specific data types.

Digression: expection of fraction IBS considering error Equation (B.8) means that

$$\mathbb{P}(Y_t^{(i)} = Y_t^{(j)} | \text{IBD}_t = 1) = (1 - (K_t - 1)\varepsilon)^2 + \varepsilon^2 (K_t - 1),$$
$$\mathbb{P}(Y_t^{(i)} = Y_t^{(j)} | \text{IBD}_t = 0) = (1 - (K_t - 1)\varepsilon)^2 h_t + \varepsilon^2 (K_t - 2 + h_t) + 2\varepsilon(1 - (K_t - 1)\varepsilon)(1 - h_t),$$

where $h_t = \sum_{g \in \mathcal{G}} f_t(g)^2$. Consequently, under the present observation model,

$$\begin{aligned} \mathbb{E}[\widehat{\text{IBS}}_m] = \frac{1}{m} \sum_{t=1}^{m} \big\{ & r \left((1 - (K_t - 1)\varepsilon)^2 + \varepsilon^2 (K_t - 1) \right) + \\ & (1 - r) \left((1 - (K_t - 1)\varepsilon)^2 h_t + \varepsilon^2 (K_t - 2 + h_t) + 2\varepsilon(1 - (K_t - 1)\varepsilon)(1 - h_t) \right) \big\}, \\ = r & \left((1 - (K_t - 1)\varepsilon)^2 + \varepsilon^2 (K_t - 1) \right) + \\ & (1 - r) \left((1 - (K_t - 1)\varepsilon)^2 \bar{h}_m + \varepsilon^2 (K_r - 2 + \bar{h}_m) + 2\varepsilon(1 - (K_t - 1)\varepsilon)(1 - \bar{h}_m) \right), \end{aligned} \tag{B.9}$$

where $\bar{h}_m = \frac{1}{m} \sum_{t=1}^{m} h_t$. Equation (B.9) reduces to (A.1) when $\varepsilon = 0$.

B.3.3 The likelihood under the independence model

This model assumes independent random variables IBD_t across loci $t \in \{1, \dots, m\}$. It is a particular case of the above HMM when all $d_t = \infty$. Given a relatedness parameter $r \in [0, 1]$, IBD_t is assumed Bernoulli with parameter r. Next, we define an observation model: given $\text{IBD}_t = 0$, we assume that $Y_t^{(i)}$ and $Y_t^{(j)}$ are independent Categorical variables with parameter $f_t(g)$. Given $\text{IBD}_t = 1$, we assume that $Y_t^{(i)}$ follows a Categorical distribution with parameter $f_t(g)$ and that $Y_t^{(j)} = Y_t^{(i)}$ with probability one. This defines the observation model. The associated likelihood at locus t is

$$\mathbb{P}\left(Y_t^{(i)} = g^{(i)}, Y_t^{(j)} = g^{(j)} | r\right) = \sum_{\text{IBD}_t \in \{0,1\}} \mathbb{P}\left(Y_t^{(i)} = g^{(i)}, Y_t^{(j)} = g^{(j)} | \text{IBD}_t\right) \mathbb{P}\left(\text{IBD}_t \mid r\right).$$

At this point we can define, for all t,

$$\begin{aligned}
a_t = \sum_{g,g' \in \mathcal{G}} & \left\{\mathbb{1}(g^{(i)} = g)(1 - (K_t - 1)\varepsilon) + \mathbb{1}(g^{(i)} \neq g)\varepsilon\right\} \times \\
& \left\{\mathbb{1}(g^{(j)} = g')(1 - (K_t - 1)\varepsilon) + \mathbb{1}(g^{(j)} \neq g')\varepsilon\right\} \times \\
& \left\{f_t(g)\mathbb{1}(g = g')\right\}, \\
b_t = \sum_{g,g' \in \mathcal{G}} & \left\{\mathbb{1}(g^{(i)} = g)(1 - (K_t - 1)\varepsilon) + \mathbb{1}(g^{(i)} \neq g)\varepsilon\right\} \times \\
& \left\{\mathbb{1}(g^{(j)} = g')(1 - (K_t - 1)\varepsilon) + \mathbb{1}(g^{(j)} \neq g')\varepsilon\right\} \times \\
& \left\{f_t(g) f_t(g')\right\},
\end{aligned}$$

so that the likelihood reads $\mathcal{L}_t(r) = a_t r + b_t(1 - r)$.

The full log-likelihood can be simply written as

$$\ell_{1:m}(r) = \sum_{t=1}^{m} \ell_t(r) = \sum_{t=1}^{m} \log\left\{a_t r + b_t(1 - r)\right\}.$$

The gradient of the log-likelihood reads

$$\ell'_{1:m}(r) = \sum_{t=1}^{m} \ell'_t(r) = \sum_{t=1}^{m} \left\{\frac{a_t - b_t}{a_t r + b_t(1 - r)}\right\}.$$

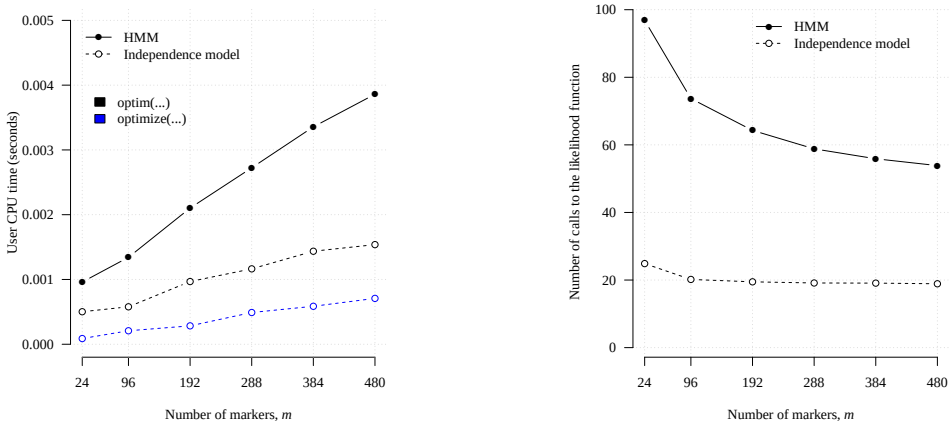
The second-order derivative of the log-likelihood reads

$$\ell''_{1:m}(r) = \sum_{t=1}^{m} \ell''_t(r) = -\sum_{t=1}^{m} \left\{\frac{(a_t - b_t)^2}{(a_t r + b_t(1 - r))^2}\right\}. \tag{B.10}$$

Since both numerator and denominator of each term are positive, $\ell''_{1:m}(r)$ is strictly negative for all $r \in (0, 1)$, and thus the function $r \mapsto \ell_{1:m}(r)$ is concave on $(0, 1)$.

For the HMM model, the form of the likelihood is not explicit, but the likelihood function can still be computed via the forward algorithm (Rabiner 1989). In terms of the computational efforts

involved in obtaining the MLEs, optimizing the likelihood for the independence model is faster (see Figure B.4). Overall, optimizing the log-likelihood in both models comes at a comparable cost, scaling linearly in m in both cases.



(a) Computational time

(b) Computational convergence using `optim`

Figure B.4: Computational time and convergence averaged over 500 simulated pairs under the HMM and the independence model. Computation time scales linearly with m.

B.3.4 Maximizing Fisher information

We focus on a single locus t, which we suppress from the notation. Let us denote the log-likelihood by ℓ and recall the formula

$$\log \mathbb{P}\left(Y^{(i)}, Y^{(j)}; r\right) = \ell(r) = \log\left(ar + b(1-r)\right), \quad \ell''(r) = -\frac{(a-b)^2}{\left(ar + b(1-r)\right)^2}.$$

Assume there is no genotyping error for simplicity. Then $a = f(Y^{(i)})\mathbb{1}(Y^{(i)} = Y^{(j)})$ and $b = f(Y^{(i)})f(Y^{(j)})$. From there the Fisher Information Matrix (FIM) is obtained as

$$\text{FIM} = \mathbb{E}\left[-\ell''(r)\right] = \sum_{y^{(i)}, y^{(j)}} \frac{\left(f(y^{(i)})\mathbb{1}(y^{(i)} = y^{(j)}) - f(y^{(i)})f(y^{(j)})\right)^2}{f(y^{(i)})\mathbb{1}(y^{(i)} = y^{(j)})r + f(y^{(i)})f(y^{(j)})(1-r)}.$$

It is a function of r and of the allele frequencies. The FIM is proportional to the inverse of the asymptotic variance of the MLE, thus if we want precise estimators of r, we want a large FIM. This leads to the idea of maximizing FIM with respect to f for all r, to see which allele frequencies lead to the most accurate estimation of r. We can split the sum into the case for which $y^{(i)} = y^{(j)}$ and

the case for which $y^{(i)} \neq y^{(j)}$; for simplicity we denote $f(y^{(i)})$ by f_i, which leads to

$$\begin{aligned}\mathrm{FIM}(f,r) =& \sum_{i=1}^{K} \frac{f_i^2\,(1-f_i)^2}{f_i r + f_i^2(1-r)} + \sum_{i=1}^{K}\sum_{j\neq i}^{K} \frac{f_i^2 f_j^2}{f_i f_j (1-r)},\\ =& \sum_{i=1}^{K} \frac{f_i\,(1-f_i)^2}{r + f_i(1-r)} + \sum_{i=1}^{K}\sum_{j\neq i}^{K} \frac{f_i f_j}{(1-r)},\end{aligned}$$

where we recall that K denotes the number of possible alleles. We note that $\sum_{j\neq i} f_j = 1 - f_i$ because $\sum_{i=1}^{K} f_i = 1$, therefore we obtain

$$\sum_{i=1}^{K}\sum_{j\neq i}^{K} \frac{f_i f_j}{(1-r)} = \sum_{i=1}^{K} \frac{f_i(1-f_i)}{(1-r)} = \frac{1}{1-r} - \frac{\sum_{i=1}^{K} f_i^2}{1-r},$$

and thus the simpler form for the FIM:

$$\mathrm{FIM}(f,r) = \frac{1}{1-r} + \sum_{i=1}^{K} \left\{ \frac{f_i\,(1-f_i)^2}{r + f_i(1-r)} - \frac{f_i^2}{1-r} \right\}.$$

The notation $\mathrm{FIM}(f,r)$ reflects our consideration of the FIM as a function of f and r. We now wonder how to maximize FIM over the vector $f = (f_1, \ldots, f_K)$, for any r. This is a constrained and nonlinear optimization problem since f has to be made of non-negative entries and sums to one (thus f is in the simplex of dimension K). We restrict our attention to $r \in (0,1)$, that is $r \neq 0$ and $r \neq 1$, since the interpretation of FIM as a measure of the precision of the maximum likelihood estimator is only valid when r is away from the boundaries of the parameter space $[0,1]$. For $r \in (0,1)$, the function $f \mapsto \mathrm{FIM}(f,r)$ is finite and continuous, on the simplex which is a compact set, thus it attains a maximum according to the extreme value theorem.

After plotting the contours of the function FIM on the simplex and for different values of r (and perhaps noticing that $f \mapsto \mathrm{FIM}(f,r)$ is symmetric with respect to the center of the simplex), we gather that the maximizer might be $f^\star = (K^{-1}, \ldots, K^{-1})$, irrespective of the value of r. We now prove that this is indeed the case. We do so by considering an f such that $f_1 < f_2$. We will see that we can increase $\mathrm{FIM}(f,r)$ by modifying f as follows: define $\tilde{f}$ as $\tilde{f}_1 = f_1 + \epsilon$, $\tilde{f}_2 = f_2 - \epsilon$ and $\tilde{f}_j = f_j$ for all $j \in \{3, \ldots, K\}$ (if $K \geq 3$). We will see that there exists an $\epsilon > 0$ such that $\mathrm{FIM}(\tilde{f}, r) > \mathrm{FIM}(f,r)$. Since this holds for all f with a pair of non-equal entries, we will be able to conclude that the unique maximizer of FIM is $f^\star = (K^{-1}, \ldots, K^{-1})$.

So let us consider f with $f_1 < f_2$. We start by noting that, for all $x \in (0,1)$,

$$\psi(x+\epsilon) := \frac{(x+\epsilon)\,(1-(x+\epsilon))^2}{r + (x+\epsilon)(1-r)} - \frac{(x+\epsilon)^2}{1-r}$$

can be expanded as $\epsilon \to 0$ as

$$\begin{aligned}&\frac{x\,(1-x)^2}{r + x(1-r)} + \epsilon \left\{ \frac{1-x}{r+(1-r)x} \left(1 - 3x - \frac{(1-r)x(1-x)}{r+(1-r)x} \right) \right\} + \mathcal{O}\left(\epsilon^2\right) - \frac{(x + 2\epsilon x + \epsilon^2)}{1-r}\\ =&\psi(x) + \epsilon \left\{ \frac{1-x}{r+(1-r)x} \left(1 - 3x - \frac{(1-r)x(1-x)}{r+(1-r)x} \right) - \frac{2x}{1-r} \right\} + \mathcal{O}\left(\epsilon^2\right),\end{aligned}$$

where $\mathcal{O}(\epsilon^2)$ refers to terms which behave as ϵ^2 when $\epsilon \to 0$ and thus are negligible in front of the term in ϵ. From this we deduce that $\psi(x+\epsilon) - \psi(x) = \epsilon h(x) + \mathcal{O}(\epsilon^2)$ with

$$h(x) := \frac{1-x}{r+(1-r)x}\left(1 - 3x - \frac{(1-r)x(1-x)}{r+(1-r)x}\right) - \frac{2x}{1-r}.$$

We now show that $x \mapsto h(x)$ is decreasing in x over $[0,1]$. We do so by bruteforce differentiation, yielding

$$\frac{d}{dx}h(x) = -\frac{2r}{(1-r)\left(r+(1-r)x\right)^3}.$$

We see that the above expression is strictly negative for all r and x so that $x \mapsto h(x)$ is strictly decreasing.

The fact that $x \mapsto h(x)$ is strictly decreasing allows us to conclude the proof. Indeed, combined with the assumption $f_1 < f_2$, we have $h(f_1) > h(f_2)$. Therefore,

$$\begin{aligned}\mathrm{FIM}(\tilde{f}, r) - \mathrm{FIM}(f, r) &= \psi(f_1+\epsilon) - \psi(f_1) + \psi(f_2-\epsilon) - \psi(f_2) \\ &= \epsilon\left(h(f_1) - h(f_2)\right) + \mathcal{O}(\epsilon^2),\end{aligned}$$

from which we deduce that there is an $\epsilon > 0$ small enough so that $\mathrm{FIM}(\tilde{f}, r) - \mathrm{FIM}(f, r) > 0$. To summarize, if f is such that one of its components is strictly greater than another component, then we can increase the objective function FIM. We deduce that the function $f \mapsto \mathrm{FIM}(f, r)$ is uniquely maximized at $f^\star = (K^{-1}, \ldots, K^{-1})$, for which no component is greater than another one.

C Comparable studies

Dataset/s and citation/s: study goal/s	Related analyses and comparable results
Colombia (Echeverry *et al.* 2013): The goal of (Echeverry *et al.* 2013) was to characterise the population structure of Colombian *P. falciparum* parasites ahead clinical trials of antimalarial drugs and genotype-phenotype association studies.	Among a suite of different genetic analyses, the fraction IBS (called "proportion of alleles shared (ps)" in (Echeverry *et al.* 2013)) was calculated for all monoclonal parasite sample pairs. The overall distribution of the fraction IBS was not summarized directly (fractions IBS were used to cluster parasites into multilocus genotypes (MLGs), which in turn were used to calculate genotypic richness, a measure of the proportion of unique MLGs). However, in the discussion of (Echeverry *et al.* 2013) expected heterozygosities (HEs) of four populations identified from the data using STRUCTRUE (Pritchard *et al.* 2000) were reported: 0.25, 0.21, 0.27, 0.34. The maximum of these numbers is the same as $1-\bar{h}_{m_{\max}} = 0.34$ for the Colombian dataset (Table 1, main text).
Thailand 93-SNP (Nkhoma *et al.* 2013; Taylor *et al.* 2017): The goals of (Nkhoma *et al.* 2013) and (Taylor *et al.* 2017) were to explore population genetic correlates of *P. falciparum* transmission decline, and signal of *P. falciparum* genetic connectivity, respectively, on the Thailand-Myanmar border.	Among a suite of different genetic analyses in (Nkhoma *et al.* 2013), the "number of alleles shared (ps)" was calculated for all monoclonal parasite pairs and used as in (Echeverry *et al.* 2013). HEs comparable to $1-\bar{h}_{m_{\max}} = 0.42$ (Table 1, main text) were reported: 0.427 and 0.429 for early and late time periods respectively. Distributions of IBD-based estimates reported in (Taylor *et al.* 2017) (Figure Q S2 Text) are comparable to those reported here. They were calculated using hmmIBD (Schaffner *et al.* 2018).
Thailand WGS (Cerqueira *et al.* 2017; Taylor *et al.* 2017): The goal of (Cerqueira *et al.* 2017) was to identify genetic signals of antimalarial drug resistance using longitudinal genomic surveillance. The goal of (Taylor *et al.* 2017) is stated above. Both studies estimated IBD-based relatedness using hmmIBD (Schaffner *et al.* 2018).	IBD between sample pairs was used to assess the impact of recent shared ancestry on the identifiablity of variants under drug selection (Cerqueira *et al.* 2017). It was deemed modest besides in 2014 (distributions were plotted, but they are not directly comparable to those reported here due to time partitions and omission of zero valued estimates). Distributions of IBD-based estimates reported in (Taylor *et al.* 2017) (Figure Q S2 Text) are comparable to those reported here.
Thailand MS (Taylor *et al.* 2018): The goal of (Taylor *et al.* 2018) was to infer the state (relapse, reinfection or recrudescence) of recurrent *P. vivax* infections on the Thailand-Myanmar border.	Although this study uses the concept of IBD to infer states, estimates of r were not directly generated: within an intermediate step of a genetic model, the probability of the data given a proposed genetic relationship (stranger, sibling or clone) is calculated by summing over IBD states. There is thus no direct comparator for the results reported here.
The Gambia (Omedo *et al.* 2017a), Kilifi (Omedo *et al.* 2017a) and Western Kenya (Omedo *et al.* 2017b): These datasets derive from two studies whose goals were to characterise the spatial genetic structure of *P. falciparum* parasites collected at a sub-national scale in select sites in East and West Africa.	Among a suite of different genetic analyses, SNP differences were calculated for all parasite sample pairs. SNP differences were used to explore within-site spatial-temporal trends in parasite relatedness and to summarise within-site parasite diversity. In Western Kenya, 15.272 of 83 SNPs (reported fraction 0.184) were different on average. In the Gambia, 2.867 of 33 SNPs (derived fraction 0.087) were different on average. In Kilifi, 3.229 of the same 33 SNPs (derived fraction 0.098) were different on average. Values of $1-\bar{h}_{m_{\max}} = 0.27, 0.22, 0.13$ for Western Kenya, The Gambia and Kilifi (Table 1, main text) are not comparable because they are based on different data (59, 31 and 127 SNPs, respectively) due to different SNP and sample filters described in main text.

Table C.1: A summary of how primary analyses of *Plasmodium* monoclonal datasets compare with analyses reported here. For full details of sample collection and data generation see citations above and references therein. For additional steps taken to process the data for current use see main text.

Study goal and citation	Comparable result
Relatedness inference for close relatives using poor quality samples (Natesh *et al.* 2018)	100 SNPs identified individuals and close relatives
Parentage inference in diploids using likelihood ratio test and numeric approximation of false positive and negative rates for different numbers of loci and genotyping error rates (Anderson and Garza 2006)	60-100 SNPs sufficient
Parentage and sibship inference in diploids (Baetscher *et al.* 2018) using method of (Anderson and Garza 2006)	96 microhaplotype loci
Ancestry assignment and coefficient inference in diploids via inverse expected Fisher information matrix (Rosenberg *et al.* 2003)	4-125,000 biallelic SNPs, depending on allele frequencies and required precision
Relatedness inference in diploids using a variety of estimators and sub-sampling of empirical data on 86 MSs, each with 2 to 19 alleles (Bink *et al.* 2008)	"In this study a set of 34 polymorphic loci seemed to be a good balance between performance of estimators and marker genotyping costs"
Relatedness inference in autopolyploids using a variety of estimators and simulation (Huang *et al.* 2015)	Approximately 200 markers, each with 10 alleles, for 95% confidence interval of $r \pm 0.05$ around diploids
Joint parentage and sibship inference of polyploids whose genotypes are transformed into "pseudodiploid-dominant genotypes" to enable application of likelihood methods designed for diploids, using both simulated and empiric data (Wang and Scribner 2014)	10-20 MSs each having 10 alleles
Connectivity between malaria parasite populations based on relatedness between monoclonal *P. falciparum* parasite samples (Taylor *et al.* 2017)	96 SNPs sufficient to recover comparable spatio-temporal trends to those obtained using WGS, assessed by sub-sampling data
Joint sibship inference in diploids and haplodiploids using maximum likelihood methods and simulation (Wang 2004)	approx. 6-10 markers each with 10 alleles, or approx. 30-40 biallelic markers, depending on family size and error inclusion
Relatedness inference for zebra finch and pigs reviewed in (Speed and Balding 2015)	More than 771 SNPs (for zebra finch) and 2000 SNPs (for pigs)

Table C.2: A non-exhaustive selection of studies in which numbers of loci for relatedness and associated inference are reported. Most of the above studies assume independence between markers.

References

Aït-Sahalia, Y., 2002 Maximum likelihood estimation of discretely sampled diffusions: a closed-form

approximation approach. Econometrica **70**: 223–262.

Anderson, E. C. and J. C. Garza, 2006 The power of single-nucleotide polymorphisms for large-scale parentage inference. Genetics **172**: 2567–2582.

Baetscher, D. S., A. J. Clemento, T. C. Ng, E. C. Anderson, J. C. Garza, *et al.*, 2018 Microhaplotypes provide increased power from short-read DNA sequences for relationship inference. Molecular Ecology Resources **18**: 296–305.

Barndorff-Nielsen, O. E., S. E. Graversen, J. Jacod, and N. Shephard, 2006 Limit theorems for bipower variation in financial econometrics. Econometric Theory **22**: 677–719.

Bink, M. C., A. D. Anderson, W. E. Van De Weg, and E. A. Thompson, 2008 Comparison of marker-based pairwise relatedness estimators on a pedigreed plant population. Theoretical and Applied Genetics **117**: 843–855.

Bladt, M. and M. Sørensen, 2005 Statistical inference for discretely observed Markov jump processes. Journal of the Royal Statistical Society: Series B (Statistical Methodology) **67**: 395–410.

Brown, M. D., C. G. Glazner, C. Zheng, and E. A. Thompson, 2012 Inferring coancestry in population samples in the presence of linkage disequilibrium. Genetics **190**: 1447–1460.

Cappé, O., E. Moulines, and T. Rydén, 2005 *Inference in Hidden Markov Models*. Springer.

Cerqueira, G. C., I. H. Cheeseman, S. F. Schaffner, S. Nair, M. McDew-White, *et al.*, 2017 Longitudinal genomic surveillance of Plasmodium falciparum malaria parasites reveals complex genomic architecture of emerging artemisinin resistance. Genome Biology **18**: 78.

Douc, R. and E. Moulines, 2012 Asymptotic properties of the maximum likelihood estimation in misspecified hidden Markov models. The Annals of Statistics **40**: 2697–2732.

Douc, R., E. Moulines, and D. Stoffer, 2014 *Nonlinear time series: Theory, methods and applications with R examples*. Chapman and Hall/CRC.

Doucet, A. and N. Shephard, 2012 Robust inference on parameters via particle filters and sandwich covariance matrices. University of Oxford, Department of Economics .

Echeverry, D. F., S. Nair, L. Osorio, S. Menon, C. Murillo, *et al.*, 2013 Long term persistence of clonal malaria parasite Plasmodium falciparum lineages in the Colombian Pacific region. BMC Genetics **14**.

Geyer, C. J., 2013 Asymptotics of maximum likelihood without the LLN or CLT or sample size going to infinity. In *Advances in Modern Statistical Theory and Applications: A Festschrift in honor of Morris L. Eaton*, pp. 1–24, Institute of Mathematical Statistics.

Henden, L., S. Lee, I. Mueller, A. Barry, and M. Bahlo, 2018 Identity-by-descent analyses for measuring population dynamics and selection in recombining pathogens. PLoS genetics **14**: e1007279.

Hill, W. and B. Weir, 2011 Variation in actual relationship as a consequence of mendelian sampling and linkage. Genetics Research **93**: 47–64.

Hoffman, J. and W. Amos, 2005 Microsatellite genotyping errors: detection approaches, common sources and consequences for paternal exclusion. Molecular Ecology **14**: 599–612.

Huang, K., S. T. Guo, M. R. Shattuck, S. T. Chen, X. G. Qi, *et al.*, 2015 A maximum-likelihood estimation of pairwise relatedness for autopolyploids. Heredity **114**: 133–142.

Leutenegger, A.-L., B. Prum, E. Génin, C. Verny, A. Lemainque, *et al.*, 2003 Estimation of the Inbreeding Coefficient through Use of Genomic Data. The American Journal of Human Genetics **73**: 516–523.

McDew-White, M., X. Li, S. C. Nkhoma, S. Nair, I. Cheeseman, *et al.*, 2019 Mode and tempo of microsatellite length change in a malaria parasite mutation accumulation experiment. bioRxiv .

Messerli, C., N. E. Hofmann, H.-P. Beck, and I. Felger, 2017 Critical evaluation of molecular monitoring in malaria drug efficacy trials and pitfalls of length-polymorphic markers. Antimicrobial agents and chemotherapy **61**: e01500–16.

Miles, A., Z. Iqbal, P. Vauterin, R. Pearson, S. Campino, *et al.*, 2016 Indels, structural variation and recombination drive genomic diversity in Plasmodium falciparum. Genome Research **26**: 1288–1299.

Natesh, M., R. W. Taylor, N. Truelove, E. A. Hadly, S. Palumbi, *et al.*, 2018 Empowering conservation practice with efficient and economical genotyping from poor quality samples using mPCRseq. bioRxiv .

Nkhoma, S. C., S. Nair, S. Al-Saai, E. Ashley, R. McGready, *et al.*, 2013 Population genetic correlates of declining transmission in a human pathogen. Molecular Ecology **22**: 273–285.

Omedo, I., P. Mogeni, T. Bousema, K. Rockett, A. Amambua-Ngwa, *et al.*, 2017a Micro-epidemiological structuring of Plasmodium falciparum parasite populations in regions with varying transmission intensities in Africa. Wellcome Open Research **2**.

Omedo, I., P. Mogeni, K. Rockett, A. Kamau, C. Hubbart, *et al.*, 2017b Geographic-genetic analysis of Plasmodium falciparum parasite populations from surveys of primary school children in Western Kenya. Wellcome Open Research **2**.

Pritchard, J. K., M. Stephens, and P. Donnelly, 2000 Inference of population structure using multilocus genotype data. Genetics **155**: 945–959.

Rabiner, L. R., 1989 A tutorial on hidden Markov models and selected applications in speech recognition. Proceedings of the IEEE **77**: 257–286.

Rosenberg, N. A., L. M. Li, R. Ward, and J. K. Pritchard, 2003 Informativeness of Genetic Markers for Inference of Ancestry *. Am. J. Hum. Genet **73**: 1402–1422.

Schaffner, S. F., A. R. Taylor, W. Wong, D. F. Wirth, and D. E. Neafsey, 2018 HmmIBD: Software to infer pairwise identity by descent between haploid genotypes. Malaria Journal **17**: 10–13.

Speed, D. and D. J. Balding, 2015 Relatedness in the post-genomic era: Is it still useful? Nature Reviews Genetics **16**: 33–44.

Taylor, A. R., S. F. Schaffner, G. C. Cerqueira, S. C. Nkhoma, T. J. Anderson, *et al.*, 2017 Quantifying connectivity between local plasmodium falciparum malaria parasite populations using identity by descent. PLoS genetics **13**: e1007065.

Taylor, A. R., J. A. Watson, C. S. Chu, K. Puaprasert, J. Duanguppama, *et al.*, 2018 Estimating the probable cause of recurrence in plasmodium vivax malaria: relapse, reinfection or recrudescence? bioRxiv .

Wang, J., 2004 Sibship Reconstruction from Genetic Data with Typing Errors. Genetics **166**: 1963–1979.

Wang, J. and K. T. Scribner, 2014 Parentage and sibship inference from markers in polyploids. Molecular Ecology Resources **14**: 541–553.

Wasserman, L., 2013 *All of statistics: a concise course in statistical inference*. Springer Science & Business Media.