Published in final edited form as:

Theor Popul Biol. 2011 May ; 79(3): 55–63. doi:10.1016/j.tpb.2011.01.003.

Match probabilities in a finite, subdivided population

Anna-Sapfo Malaspinas^a, Montgomery Slatkin^a, and Yun S. Song^{b,*}

^aDepartment of Integrative Biology, University of California, Berkeley, CA 94720, USA

^bDepartment of EECS and Department of Statistics, University of California, Berkeley, CA 94720, USA

Abstract

We generalize a recently introduced graphical framework to compute the probability that haplotypes or genotypes of two individuals drawn from a finite, subdivided population match. As in the previous work, we assume an infinite-alleles model. We focus on the case of a population divided into two subpopulations, but the underlying framework can be applied to a general model of population subdivision. We examine the effect of population subdivision on the match probabilities and the accuracy of the product rule which approximates multi-locus match probabilities as a product of one-locus match probabilities. We quantify the deviation from predictions of the product rule by R , the ratio of the multi-locus match probability to the product of the one-locus match probabilities. We carry out the computation for two loci and find that ignoring subdivision can lead to underestimation of the match probabilities if the population under consideration actually has subdivision structure and the individuals originate from the same subpopulation. On the other hand, under a given model of population subdivision, we find that the ratio R for two loci is only slightly greater than 1 for a large range of symmetric and asymmetric migration rates. Keeping in mind that the infinite-alleles model is not the appropriate mutation model for STR loci, we conclude that, for two loci and biologically reasonable parameter values, population subdivision may lead to results that disfavor innocent suspects because of an increase in identity-by-descent in finite populations. On the other hand, for the same range of parameters, population subdivision does not lead to a substantial increase in linkage disequilibrium between loci. Those results are consistent with established practice.

Keywords

Match probability; Population subdivision; Product rule; Match graph

1. Introduction

In forensic science, the analysis of DNA has become increasingly important. The multi-locus genotype, *the DNA profile*, of a biological sample from a crime scene is compared with the DNA profile of one or more suspects and often with the set of profiles in a large database. In the United States, the CODIS (Combined DNA Index System) is comprised of genotypes for 13 tetranucleotide microsatellite loci (<http://www.fbi.gov/hq/lab/html/codis1.htm>).

In the absence of laboratory error, a match between an individual's profile and that from the crime scene can be explained in one of two ways: either the individual is the source of the

* Corresponding author. sapfo@berkeley.edu (A.-S. Malaspinas), slatkin@berkeley.edu (M. Slatkin), yss@eecs.berkeley.edu (Y.S. Song).

crime-scene sample or the individual is not the source but has the same profile by chance alone. The probability of such a match has been defined in different ways that also depend on the underlying population genetic model. In this paper, we define the *match probability* as the probability of a match between two individuals drawn at random. We consider the cases where the two individuals come from the same subpopulation or from different subpopulations. Given that a genetic match is often sufficient to uphold a conviction (Song et al., 2009), the match probability plays an important role in the US and other judicial systems.

In the US courts, the match probability is usually computed by following the recommendations of the second National Research Council Report (Committee on DNA forensic science: an Update, 1996), called NRCII. Recommendation 4.1 of the NRCII advocates for the use of the profile frequencies to compute the match probability. But, if the subpopulation origin of the sample is known but not the allele frequencies for the specific subpopulation, NRCII recommends the use of the equation developed by Balding and Nichols (1994, 1995) (recommendation 4.2). The NRC report has been criticized in the scientific community (see for example Evett and Weir (1998) and Balding (2005)). In particular, it has been argued that the equation of Balding and Nichols (1994, 1995) should be used even in cases where the origin of the sample is unknown. The latter equation describes the conditional probability that given an observed profile, we find another individual with the same profile. This conditional match probability depends on a parameter, θ , which accounts for small deviations from Hardy–Weinberg frequencies caused by population subdivision or other deviations from random mating, and which is equivalent to F_{ST} in the population genetics literature. NRCII recommend a value of θ between 0.01 and 0.03 to account for the observed genotype frequencies in all known human populations, but others have argued that in some minority groups a value of 0.05 may be more appropriate (e.g. Balding (2005)). NRCII recommends that the multi-locus genotypic match probability be computed by multiplying the one-locus match probabilities, which, following convention in the forensics literature, we will call the *product rule*. For the 13 CODIS loci, the match probability of two unrelated individuals computed using the product rule is on the order of 10^{-14} – 10^{-15} (Song et al., 2009).

The use of the product rule remains controversial (Laurie and Weir, 2003; Bhaskar and Song, 2009). Because of the difficulty of analyzing multi-locus models in finite populations, there has been relatively little theoretical work on the conditions under which the product rule provides an accurate approximation to the match probabilities of multi-locus profiles. For a small number of loci, Laurie and Weir (2003) and Song and Slatkin (2007) computed the match probability in a finite randomly mating population and found that the deviations from the predictions of the product rule are small unless mutation rates are unrealistically high. Bhaskar and Song (2009) later generalized the graphical method of Song and Slatkin (2007) to compute the match probability of haploid genotypes for as many as 10 loci in a Wright-Fisher model and 13 loci in a Moran model. They showed that deviations from the predictions of the product rule are sensitive to the assumed mutation rate but, for mutation rates consistent with the observed levels of heterozygosity at CODIS loci, deviations from the predictions of the product rule were relatively small for up to 13 loci.

All of these theoretical studies of match probabilities assume an infinite-alleles model of mutation at each locus (i.e., every mutation leads to a new allele never seen before in the population). They predict the deviation from the product rule resulting from multi-locus identity-by-descent created by genetic drift in a finite population. These models cannot predict the overall deviations from the product rule in practice, however, because the loci used for forensic purposes are short tandem repeat (STR) loci for which the infinite-alleles model is not realistic. The same allele can be created more than once by mutation. A

stepwise mutation model is more appropriate for STR loci but does not lead to a tractable theory of match probabilities even in a single population. The match probabilities calculated from the infinite-alleles model indicate the magnitude of deviations from the product rule expected because of the accumulated identity-by-descent, which is likely to be the primary effect of population subdivision.

Here, we extend the graphical framework derived by Song and Slatkin (2007) to explicitly account for population subdivision. We model two linked loci in two subpopulations that exchange migrants. We derive a system of coupled linear recurrence equations for the match probabilities. The equilibrium match probabilities are then found by assuming stationarity and solving the recurrence equations. As in the previous studies, we assume an infinite-alleles model. Our goals are to determine the effect of population subdivision on the match probabilities, and to study what assumptions about population subdivision lead to two-locus match probabilities that differ substantially from the predictions made using the product rule. The model allows us to address two scenarios of practical interest.

1. *Same-subpopulation scenario*: Suppose there are two subpopulations α and β , and the source of the crime-scene sample and the defendant are both from the same subpopulation, say subpopulation α . In this case, the correct thing to do would be to compute the match probability between two individuals taken at random within the subpopulation α . However, the subpopulation origin of the individuals is not known and one considers the match of two individuals taken at random from the entire population assuming no subdivision, i.e., assuming random mating between all individuals of the whole population. This case corresponds to the situation described by Balding and Nichols (1994, 1995).
2. *Cryptic subdivision scenario*: One is interested in the match probability between two individuals randomly drawn from the entire population. He or she computes this probability assuming no population subdivision, but what is thought to be a single randomly mating population actually consists of two subpopulations with some gene flow between them.

This paper is organized as follows. In Section 2, we first describe the Wright–Fisher random mating model for a subdivided population and give a more precise definition of match probabilities. We then describe in Section 3 our extension of the aforementioned graphical framework to incorporate population structure. In Section 4, we present the results for the haplotypic and genotypic match probabilities for one and two loci for various migration rates. We show that the aforementioned same-subpopulation scenario can lead to a substantial underestimation of the match probabilities, while both scenarios considered result in relatively small deviations from the product rule for reasonable mutation rates at the two loci.

2. Random mating model for a subdivided population and match probabilities

In this paper, we consider a population consisting of two subpopulations with migration. Provided below is a brief description of the assumed random mating model. See Table 1 for a summary of notation.

1. The total population is finite and of constant size N ($2N$ gametes).
2. Generations are non-overlapping.
3. The population has two subpopulations, α and β , with constant sizes N_α and N_β . Note that $N = N_\alpha + N_\beta$.

4. We assume a standard Wright–Fisher random union of gametes within each subpopulation as in Laurie and Weir (2003), extended to incorporate migration.
5. Mutations occur at locus i with probability μ_i per generation and are independent of the subpopulation, and of other loci. Each mutation event produces a new allele never seen before in the population, i.e., we assume an infinite-alleles model.
6. Recombination may occur at each generation between loci with probability r per generation for gametes within a subpopulation, at a rate independent of the subpopulation.
7. Forward in time, the probability a particular individual from subpopulation α migrates to subpopulation β is $m_{\beta\alpha}N_\beta/N_\alpha$ per generation. The analogous probability for migration from β to α is $m_{\alpha\beta}N_\alpha/N_\beta$. The parameters $m_{\alpha\beta}$ and $m_{\beta\alpha}$ are *backward* migration fractions.
8. All loci are neutral.
9. Migration is conservative (Nagylaki, 1980), i.e., $m_{\alpha\beta}N_\alpha = m_{\beta\alpha}N_\beta$, or equivalently

$$N_\alpha = \frac{m_{\beta\alpha}}{m_{\alpha\beta} + m_{\beta\alpha}} N \quad \text{and} \quad N_\beta = \frac{m_{\alpha\beta}}{m_{\alpha\beta} + m_{\beta\alpha}} N.$$

A new generation is obtained by drawing individuals from the previous generation. For two loci, a gamete from an unspecified subpopulation is denoted by $\mathbf{x} = x_1x_2$, where x_i denotes the allele at locus i . Further, we denote by $\mathbf{a} = a_1a_2$ and $\mathbf{b} = b_1b_2$ gametes from subpopulation α and β , respectively. Considering only two loci, our assumptions imply the following scheme for the generation of offsprings.

1. Two gametes, $\mathbf{x}^1 = x_1^1x_2^1$ and $\mathbf{x}^2 = x_1^2x_2^2$, are drawn with replacement from within a subpopulation.
2. The two gametes drawn recombine with probability r to create the offspring gamete. With probability $\frac{1}{2}(1-r)$, $\frac{1}{2}(1-r)$, $\frac{1}{2}r$, $\frac{1}{2}r$, the offspring gametes are, respectively, $x_1^1x_2^1$, $x_1^2x_2^2$, $x_1^1x_2^2$, $x_1^2x_2^1$.
3. As described above, each offspring gamete produced in a subpopulation may migrate to the other subpopulation with a certain probability.
4. At each locus i , each offspring gamete may have undergone mutation with probability μ_i , independent of the subpopulation. Every mutation creates a new allele never seen before in the population.

The above procedure is repeated N_α times in subpopulation α and N_β times in subpopulation β .

We are interested in the one-locus and the two-locus random match probability for the haplotypic and genotypic case. We define it as the probability that two individuals drawn at random match at 1 or 2 loci, respectively, (Laurie and Weir, 2003; Song and Slatkin, 2007). We will consider three different cases; the two haploid or diploid individuals come from the same subpopulation (either α or β) or each of them comes from a different subpopulation. We assume that the alleles of each locus for each individual are known. Our definition differs from the match probability defined by some authors (e.g. Weir (2004)) in that it is not a conditional match probability.

3. Description of the graphical framework

Song and Slatkin (2007) derive recursion equations by representing match relations as graphs and by performing operations on them. The four parts in the above-mentioned mating scheme will be captured by four different operations on graphs, namely, vertex-merge, vertex-split, vertex-recolor, and vertex-count, respectively. Our graphical method is the same as theirs', except that vertices are now colored to distinguish gametes from different subpopulations.

3.1. Correspondence between match probabilities and graphs

The graphs we consider are undirected graphs with edge labels and vertex colors. Each vertex represents a gamete and the color of the vertex indicates the subpopulation to which the gamete belongs. Two vertices are joined by an edge (or an arc) labeled i if the associated gametes match at locus i . Three examples of match graphs are shown in Fig. 1. We sometimes omit edge labels for convenience of drawing, in which case we adopt the convention of drawing locus 1 (respectively, locus 2) edges above (respectively, below) the vertices.

As pointed out in Song and Slatkin (2007), two match probabilities are equal under random mating if they are related by some permutation of gamete labels. More formally, two graphs are equivalent if they are isomorphic as vertex-colored edge-labeled graphs. For example in Fig. 1, G_1 and G_2 are isomorphic, but neither G_1 nor G_2 is isomorphic to G_3 . If we permute the superscripts on gametes as $1 \rightarrow 3, 2 \rightarrow 1, 3 \rightarrow 2, 4 \rightarrow 4$, we can transform G_1 into G_2 , but there exists no permutation that can transform G_1 (or G_2) into G_3 .

In the graphical framework, the goal is to relate an offspring graph G^O (corresponding to a match relationship at time t) to a set of parental graphs $G_1^P, G_2^P, \dots$ (corresponding to a set of match relationships at time $t - 1$). Let $\mathcal{A}$ denote the set of all possible ancestries (i.e., a set of mutation, migration, recombination and coalescent events) one-generation back in time for the gametes in G^O . Then, the probability of G^O can be decomposed as

$$\mathbb{P}(G^O) = \sum_{A \in \mathcal{A}} \mathbb{P}(G^O | A) \mathbb{P}(A). \quad (3.1)$$

For many ancestries $A \in \mathcal{A}$, the conditional probability $\mathbb{P}(G^O | A)$ may be zero; our method considers only the set of ancestries $A \in \mathcal{A}$ with positive $\mathbb{P}(G^O | A)$ and sums over that set. We say that an ancestry A is *valid* if $\mathbb{P}(G^O | A) > 0$. Given a particular valid ancestry $A_k \in \mathcal{A}$, there corresponds a parental graph G_k^P such that $\mathbb{P}(G_k^P) = \mathbb{P}(G^O | A_k)$. At stationarity, isomorphic parental and offspring graphs have the same probability, and hence we can construct a closed system of linear equations by repeatedly using (3.1). We assume that mutation (U), migration (M), recombination (R) and coalescence (C) events are independent of each other, and hence we can decompose the probability of a particular one-generation ancestry A_k into four parts as $\mathbb{P}(A_k) = Q_k^U Q_k^M Q_k^R Q_k^C$, with the superscript denoting the type of event.

3.2. Operations on graphs

We now describe a graphical method to find, for a given offspring graph G^O , the set of all valid one-generation ancestries A . Further, we describe the computation of Q_k^U, Q_k^M, Q_k^R , and Q_k^C for each ancestral A_k .

3.2.1. Mutation or vertex-count—Since we consider an infinite-alleles model, if two offspring gametes match at a particular locus, then their parental gametes in the previous generation must also match at that locus. Moreover, no mutation could have happened at that locus in producing the offspring gametes.

If n gametes match at locus i in the offspring graph G^O , an overall factor of $(1 - \mu_i)^n$ will contribute to Q_k^U . More generally, for every valid ancestry A_k , we obtain

$Q_k^U = \prod_{i=1}^2 (1 - \mu_i)^{\delta_i(G^O)}$, where $\delta_i(G^O)$ is the number of vertices in G^O that are involved in a match relation for locus i (i.e., the number of vertices incident with an edge corresponding to locus i). Examples are provided in Fig. 2. This procedure is referred to as a “vertex-count” operation.

The intermediate graph resulting from a mutation event or a vertex-count operation is identical to the original graph. We will denote by G^M , the resulting graph to refer to the fact that the next operation corresponds to a migration event.

3.3. Migration or vertex-recolor

The graphs considered in Song and Slatkin (2007) have uncolored vertices. To take into account the population subdivision, we color the vertices. A migration event corresponds to a “vertex recoloring” operation in the graphical framework (see Fig. 3).

Consider a graph with n_α vertices from subpopulation α and n_β vertices from subpopulation β . Since the migration rates are typically small, we linearize the probability of migration per generation, and therefore $Q_k^M = 1 - n_\alpha m_{\alpha\beta} - n_\beta m_{\beta\alpha}$ if no migration happens, $Q_k^M = m_{\alpha\beta}$ if a gamete from α migrates to β or $Q_k^M = m_{\beta\alpha}$ if a gamete from β migrates to α .

The graph resulting from a recoloring operation is an intermediate graph denoted by G^R , indicating that the next operation is for a recombination event.

3.4. Recombination or vertex-split

Going backward in time, if a recombination event occurs in an offspring gamete, then two parental gametes must be tracked, each contributing a new vertex. Recombination is relevant only if we consider two or more loci.

The operation on graphs that corresponds to recombination is a “vertex-split” operation. Any vertex with at least two incident edges with different labels can undergo a split operation. We can associate a number $\delta(v)$ (the δ -degree) to each vertex v , such that for a given match graph the δ -degree is the number of loci at which the vertex is involved in a match relationship. Denote by n the number of vertices with δ -degree larger than one. Hence, if s out of the n vertices split, then $Q_k^R = r^s (1 - r)^{n-s}$. See Fig. 4 for an example.

The graph obtained from a vertex-split operation is an intermediate graph denoted by G^C . The next operation to consider corresponds to coalescent events.

3.5. Coalescence or vertex-merge

Finally, because parental gametes are drawn with replacement, a parental gamete may be drawn more than once. Going backward in time, this means that two (or more) vertices in G^C may undergo a “merge operation”, provided that the vertices have the same color. Adopting Convention 1 of Song and Slatkin (2007), merging a set of vertices entails removing all edges between the vertices in that set. Any set of vertices in the same

subpopulation can merge. The probability of such a merge is the probability of drawing the same gamete two or more times from the same subpopulation. Such a merged graph will be called a parental graph G^P .

The probability of a merge event given a graph G^C is almost identical to Eq. (2) of Song and Slatkin (2007). Since two gametes from different subpopulations cannot be merged, each subpopulation can be considered independently. Having n_α (respectively, n_β) vertices of type α (respectively, β) before the merge operation and j_α (respectively, j_β) after merging, implies that j_α (respectively, j_β) distinct gametes were drawn. The probability of such an event is given by

$$Q_k^C = f_\alpha(n_\alpha, j_\alpha) f_\beta(n_\beta, j_\beta),$$

where

$$f_\alpha(n_\alpha, j_\alpha) = \frac{2N_\alpha(2N_\alpha - 1) \cdots (2N_\alpha - j_\alpha + 1)}{(2N_\alpha)^{j_\alpha}},$$

and $f_\beta(n_\beta, j_\beta)$ is similarly defined with N_α replaced with N_β .

Note that two different merge operations may produce isomorphic graphs. Isolated vertices are not involved in any match relationship, and hence can be ignored. A graph containing only isolated vertices has probability one. See Fig. 5 for an example.

3.6. Summary: count, recolor, split, and merge

There are four graphical operations corresponding to different evolutionary events: vertex-count, vertex-recolor, vertex-split, and vertex-merge. The first three operations produce intermediate graphs G^M , G^R , and G^C . They are used to relate an offspring graph G^O

representing a match probability at time t to a set of parental match graphs G_k^P at time $t - 1$. A schematic summary (corresponding to Fig. 8 of Song and Slatkin, 2007) is shown in Fig. 6. At stationarity, the match probabilities are independent of t .

3.7. An example of a closed system of equations

Illustrated in Fig. 7 is an example of a closed system of recurrence equations in our graphical framework. It is for the simplest case (i.e., the one-locus haplotypic match), involving three equations and three unknown variables. Because we consider only a single locus, no vertex-split operations are involved. The one-locus match probability $\mathbb{P}_h(x_i \equiv x'_i)$ appears as one of the unknown variables in Fig. 7, and solving the coupled equations allows us to compute that stationary probability.

Note that, when $\mu_i = 0$ for all loci i , every match probability is equal to one, and, therefore, the right-hand side of each equation in Fig. 7 must sum to one. This can be easily verified. Such consistency conditions are useful for checking that the coefficients in the recurrence equations have been determined correctly.

4. Results on match probabilities

We are interested in computing two quantities: the two-locus match probability (denoted by $\mathbb{P}_h$ in the haplotypic case and by $\mathbb{P}_g$ in the genotypic case) computed without assuming independence of the loci, and the ratio (denoted by R_h for the haplotypic case and by R_g for the genotypic case) of that probability to the approximate two-locus match probability obtained by multiplying the one-locus match probabilities. Under linkage equilibrium, the multi-locus match probability is given by the product of the one-locus match probabilities at the loci involved. Therefore, assuming linkage equilibrium the ratios R_h and R_g are equal to one. If the ratios are greater than 1, the actual match probabilities exceed those obtained by assuming the product rule because of linkage disequilibrium created between unlinked loci by the interaction of mutation and genetic drift.

To examine the effect of population subdivision on the above quantities, we present below some numerical results for both haplotypic and genotypic cases. We implemented our graphical method in *Python* to generate the system of equations, and then used *Mathematica* to solve the system. For simplicity, we present results for the case of the same mutation rate for all loci, i.e., $\mu_i = \mu$ for all i .

4.1. Haplotypic match probability

4.1.1. One-locus haplotypic match—The system of equations shown in Fig. 7 can be used to obtain the haplotypic match probability $\mathbb{P}_h(x_i \equiv x'_i)$ for locus i . Note that the ratio R_h is equal to 1 by definition.

Numerical values for one-locus haplotypic match probabilities are given in Table 2. The population size is set to $N = 10,000$, which is an approximate long-term effective population size of humans (Nei and Graur, 1984; Harding et al., 1997; Harpending et al., 1998). The two mutation rates used, 10^{-4} and 10^{-3} , correspond to expected heterozygosities in a randomly mating population of $4N\mu / (1 + 4N\mu) = 0.8$ and 0.976 . In the case of symmetric migration, the range of migration rates from 10^{-6} to 10^{-2} respectively correspond to 0.02 – 200 migrants per generation. Shown in this and later tables are (1) the F_{ST} value, (2) the match probability for two different individuals sampled randomly from the same subpopulation, and (3) the match probability of two different individuals chosen at random without regard to subpopulation. The last probability, denoted by $\mathbb{P}^{\text{random}}$, is a weighted average defined as follows:

$$\mathbb{P}_h^{\text{random}}(x \equiv x') = \frac{N_\alpha^2}{N^2} \mathbb{P}_h(a \equiv a') + \frac{2N_\alpha N_\beta}{N^2} \mathbb{P}_h(a \equiv b') + \frac{N_\beta^2}{N^2} \mathbb{P}_h(b \equiv b'), \quad (4.2)$$

where a and a' (respectively, b and b') denote alleles from subpopulation α (respectively, β). The version of F_{ST} we adopt is

$$F_{ST} = \frac{\frac{N_\alpha}{N} \cdot \mathbb{P}_h(a_i \equiv a'_i) + \frac{N_\beta}{N} \cdot \mathbb{P}_h(b_i \equiv b'_i) - \mathbb{P}_h^{\text{random}}(x \equiv x')}{1 - \mathbb{P}_h^{\text{random}}(x \equiv x')}.$$

In the asymmetric case, the within-subpopulation match probability decreases as the subpopulation size increases. In the symmetric case, the more isolated the subpopulations, the higher the within-subpopulation and the smaller the between-subpopulations match

probabilities. The effect of subdivision on the match probabilities is further discussed below in Section 4.2.2.

4.1.2. Two-locus haplotypic match—The two-locus system has a total of 26 match relations. The corresponding graphs are illustrated in Fig. 8. Given two gametes $\mathbf{x} = x_1x_2$ and $\mathbf{x}' = x'_1x'_2$, the haplotypic match probability is $\mathbb{P}_h(\mathbf{x} \equiv \mathbf{x}')$. If the loci are independent, then the product rule holds, yielding $\mathbb{P}_h(x_1 \equiv x'_1)\mathbb{P}_h(x_2 \equiv x'_2)$ for the match probability. To quantify the departure from the product rule, we consider the following ratio:

$$R_h(\mathbf{x} \equiv \mathbf{x}') = \frac{\mathbb{P}_h(\mathbf{x} \equiv \mathbf{x}')}{\mathbb{P}_h(x_1 \equiv x'_1)\mathbb{P}_h(x_2 \equiv x'_2)}.$$

Both the match probability $\mathbb{P}_h$ and the ratio R_h depend on the subpopulations to which the gametes $\mathbf{x}$ and $\mathbf{x}'$ belong. There are three different cases. (1) They both belong to subpopulation α . (2) They both belong to subpopulation β . (3) One belongs to subpopulation α , while the other belongs to subpopulation β . Graphical representations of $\mathbb{P}_h$ for the three cases are illustrated in Fig. 9.

Numerical values of match probabilities $\mathbb{P}_h$ and ratios R_h are provided in Table 3. The same parameters as in the one-locus case were used. In all cases, R_h is not much greater than 1, even for the smallest migration rate we considered. Therefore, the product rule provides an adequate approximation for computing the match probabilities.

4.2. Genotypic match probability

4.2.1. One-locus genotypic match probability—Consider two unordered pairs $\{x_i, \bar{x}_i\}$ and $\{x'_i, \bar{x}'_i\}$ of alleles. The probability of genotypic match at locus i is denoted by $\mathbb{P}_g(\{x_i, \bar{x}_i\} \equiv \{x'_i, \bar{x}'_i\})$. There are two possible ways to have a one-locus genotypic match: (1) $x_i \equiv x'_i$ and $\bar{x}_i \equiv \bar{x}'_i$, or (2) $x_i \equiv \bar{x}'_i$ and $\bar{x}_i \equiv x'_i$. However, these possibilities are not mutually exclusive, and as explained in Song and Slatkin (2007), the probability of a genotypic match can be computed using the inclusion–exclusion principle. In the graphical framework, the probability of a one-locus genotypic match is as shown in Fig. 10. The system of equations for the one-locus genotypic match case involves 18 variables (and 18 equations). Numerical values are reported in Table 4.

The conclusions are similar to those for the one-locus haplotypic case. One noticeable difference is that the match probabilities are smaller than for haplotypic matches.

4.2.2. Two-locus genotypic match probability—Consider two unordered pairs $\{x_i, \bar{x}_i\}$ and $\{x'_i, \bar{x}'_i\}$ of gametes. The probability of two-locus genotypic match is denoted by $\mathbb{P}_g(\{x_i, \bar{x}_i\} \equiv \{x'_i, \bar{x}'_i\})$. The match probability under the product rule is $\mathbb{P}_g(\{x_1, \bar{x}_1\} \equiv \{x'_1, \bar{x}'_1\}) \times \mathbb{P}_g(\{x_2, \bar{x}_2\} \equiv \{x'_2, \bar{x}'_2\})$. To quantify the departure from the product rule, we compute the following ratio:

$$R_g(\{x_i, \bar{x}_i\} \equiv \{x'_i, \bar{x}'_i\}) = \frac{\mathbb{P}_g(\{x_i, \bar{x}_i\} \equiv \{x'_i, \bar{x}'_i\})}{\mathbb{P}_g(\{x_1, \bar{x}_1\} \equiv \{x'_1, \bar{x}'_1\})\mathbb{P}_g(\{x_2, \bar{x}_2\} \equiv \{x'_2, \bar{x}'_2\})}.$$

There are four possible ways of having a two-locus genotypic match. But, again, those possibilities are not mutually exclusive and the probability of a two-locus genotypic match can be computed using the inclusion–exclusion principle. The three possible two-locus match probabilities, depending on the subpopulations of the gametes, are shown in Fig. 11. The three ratios, R_g , can be computed depending on the origin of the gametes. The closed system of equation has 1463 variables (or equations). Numerical values are reported in Table 5.

Similar conclusions can be drawn as in the two-locus haplotypic case. The main difference is that genotypic match probabilities are considerably smaller (by about one order of magnitude). Moreover, for small migration rates, the ratio R_g is not as elevated as R_h . For example, in the case of symmetric migration with $m_{\alpha\beta} = m_{\beta\alpha} = 10^{-6}$, the between-subpopulation ratio R_g ($\{a, \bar{a}\} \equiv \{b, \bar{b}\}$) is equal to 1.229 for $\mu = 10^{-3}$ and 1.007 for $\mu = 10^{-4}$, while the corresponding values for the haplotypic ratio R_h are 5.113 and 1.050, respectively (see Tables 3 and 5). The ratio R_g is close to 1 for $\mu = 10^{-4}$, suggesting that in the genotypic case, the product rule is accurate in the two-locus case for small mutation rates.

To examine the effect of population subdivision on the match probability computation, we plot in Fig. 12 $\mathbb{P}_h^{\text{random}}(x \equiv x')$ and $\mathbb{P}_h(a \equiv a')$ (respectively, $\mathbb{P}_g^{\text{random}}(\{x_i, \bar{x}_i\} \equiv \{x'_i, \bar{x}'_i\})$ and $\mathbb{P}_g(\{a_i, \bar{a}_i\} \equiv \{a'_i, \bar{a}'_i\})$) as a function of F_{ST} for the two-locus haplotypic (respectively, genotypic) case. The one-locus cases are qualitatively similar and therefore not shown. We compare the no subdivision case, the symmetric case, and the asymmetric case of Tables 3 and 5 for $\mu = 1 \times 10^{-4}$. The results in our tables labeled “Within α ” and “Within β ” are relevant for the same-subpopulation scenario, i.e. where the two individuals are drawn from one subpopulation versus drawn from the entire population ignoring subdivision. The “Within α ” results assume population α is that subpopulation and the “Within β ” results assume population β is that subpopulation. “Random” are relevant for the cryptic subdivision scenario, in which two partially isolated populations are treated as a single population.

As has been shown before (e.g. Balding and Nichols (1994, 1995)), ignoring subdivision can lead to underestimation of the match probabilities for the same-subpopulation scenario, i.e. when the two individuals belong to the same subpopulation (subplots on the right of Fig. 12). We see this effect for both the asymmetric and the symmetric case even for small F_{ST} values. On the other hand, for the cryptic subdivision scenario, the asymmetric and symmetric “Random” match probabilities agree closely with the no subdivision match probabilities. In this case, ignoring subdivision leads to significant underestimates of the match probability only when the mutation rate is higher (say, $\mu = 1 \times 10^{-3}$; plots not shown). This effect is qualitatively the same for all cases considered.

5. Discussion and conclusion

Given the increasing use of DNA evidence in criminal investigation in the United States and elsewhere, it is important to determine whether the one-locus match probability equations are adequate and whether the product rule provides accurate estimates of multi-locus match probabilities. In this paper, we showed that migration can be easily incorporated into the previously proposed graphical framework for computing exact match probabilities at stationarity. We computed the match probabilities for two loci in a population with two subpopulations that exchange migrants.

One of the advantages of our approach is that it relies on an explicit population subdivision model. That is, if one has an idea of the actual subdivision, the appropriate scenario can be investigated. Our results show that, for the same-subpopulation scenario, even with small F_{ST} values for the case of two subpopulations with symmetric or asymmetric migration, ignoring subdivision can lead to important underestimates of the match probability. This result had been shown before by others, e.g. Balding and Nichols (1994, 1995), for a different subdivision model. For the cryptic subdivision scenario, underestimation of the match probability occurs only for high mutation rates (μ) of 10^{-3} , not necessarily consistent with reasonable levels of heterozygosity. Similarly, in all the results, noticeable deviations from the product rule are obtained only for the largest mutation rate (μ) of 10^{-3} , and even for that mutation rate, the predictions of the product rule are close to the correct values. Under the infinite-alleles mutation model, a total population size of 10,000 and a mutation rate of 10^{-3} predict a heterozygosity larger than $4N\mu/(1 + 4N\mu) = 0.976$ —how much larger depends on the migration rates. Under a generalized stepwise mutation model, the predicted heterozygosity is somewhat smaller, of the order of 0.9 (Di Rienzo et al., 1994). Published heterozygosities for CODIS loci are in the range of 0.75-0.9 (Budowle et al., 2001), which suggests that mutation rates at CODIS are lower than 10^{-3} and that is consistent with the observed mutation rates at other STR loci (Ellegren, 2000). Therefore, for two loci, we conclude that the population subdivision of the kind modeled in this paper does lead to underestimates of the match probabilities but does not lead to substantial deviations from the predictions of the product rule. This conclusion is consistent with established practice; see e.g. Balding (2005) and Committee on DNA forensic science: an Update (1996). Nevertheless, the effect of more loci and more subpopulations remains an open question. Our intuition, based on the results of higher mutation rate, is that the effects might be more important for more than two loci and more than two subpopulations. Finally, the method we present in this paper is relevant, not only for the forensic sciences, but in a general context of identity-by-descent calculation.

Acknowledgments

ASM thanks Gerasimos Mataragkas, Weiwei Zhai, Philip Johnson, and Anand Patil for helpful discussions, and Anand Bhaskar for independently checking the results presented here. This research is supported in part by a Janggen-Poehn fellowship to ASM, an NIH grant R01-GM40282 to MS, and a Packard Fellowship for Science and Engineering to YSS.

References

- Balding, DJ. Weight-of-Evidence for Forensic DNA Profiles. John Wiley and Sons Ltd.; Chichester, England: 2005.
- Balding DJ, Nichols RA. DNA profile match probability calculation: how to allow for population stratification, relatedness, database selection and single bands. *Forensic Sci Int.* 1994; 64(2–3):125–140. [PubMed: 8175083]
- Balding DJ, Nichols RA. A method for quantifying differentiation between populations at multi-allelic loci and its implications for investigating identity and paternity. *Genetica.* 1995; 96(1–2):3–12. [PubMed: 7607457]
- Bhaskar, A.; Song, Y. Multi-locus match probability in a finite population: a fundamental difference between the moran and wright-fisher models. *Bioinformatics; Proceedings of the 17th Annual International Conference on Intelligent Systems for Molecular Biology, ISMB; 2009.* p. i187-i195.
- Budowle B, Shea B, Niezgoda S, Chakraborty R. CODIS STR loci data from 41 sample populations. *J Forensic Sci.* 2001; 46(3):453–489. [PubMed: 11372982]
- Committee on DNA Forensic Science: An Update, NRC. *The Evaluation of Forensic DNA Evidence.* National Academy Press; 1996.

- Di Rienzo A, Peterson A, Garza J, Valdes A, Slatkin M, Freimer N. Mutational processes of simple-sequence repeat loci in human populations. *Proc Natl Acad Sci USA*. 1994; 91(8):3166–3170. [PubMed: 8159720]
- Ellegren H. Microsatellite mutations in the germline: implications for evolutionary inference. *Trends Genet*. 2000; 16(12):551–558. [PubMed: 11102705]
- Evett, I.; Weir, B. *Interpreting DNA evidence*. Sunderland: 1998.
- Harding RM, Fullerton SM, Griffiths RC, Bond J, Cox MJ, Schneider JA, Moulin DS, Clegg JB. Archaic African and Asian lineages in the genetic ancestry of modern humans. *Am J Hum Genet*. 1997; 60:772–789. [PubMed: 9106523]
- Harpending HC, Batze MA, Gurven M, Jorde LB, Rogers AR, Sherry ST. Genetic traces of ancient demography. *Proc Natl Acad Sci*. 1998; 95:1961–1967. [PubMed: 9465125]
- Laurie C, Weir BS. Dependency effects in multi-locus match probabilities. *Theor Popul Biol*. 2003; 63:207–219. [PubMed: 12689792]
- Nagylaki T. The strong-migration limit in geographically structured populations. *J Math Biol*. 1980; 9(2):101–114. [PubMed: 7365330]
- Nei M, Graur D. Extent of protein polymorphism and the neutral mutation theory. *Evol Biol*. 1984; 17:73–118.
- Song YS, Patil A, Murphy EE, Slatkin M. Average probability that a cold hit in a DNA database search results in an erroneous attribution. *J Forensic Sci*. 2009; 54(1):22–27. [PubMed: 19040670]
- Song YS, Slatkin M. A graphical approach to multi-locus match probability computation: revisiting the product rule. *Theor Popul Biol*. 2007; 72(1):96–110. [PubMed: 17239909]
- Weir BS. Matching and partially-matching DNA profiles. *J Forensic Sci*. 2004; 49:1009–1014. [PubMed: 15461102]

Fig. 1.

Example of graphical representation of match probabilities. Filled circles represent gametes from subpopulation α , while open circles represent gametes from subpopulation β . The

graph G_1 corresponds to the probability $\mathbb{P}(a_1^1 \equiv a_1^2, a_1^1 \equiv b_1^3, a_1^2 \equiv b_1^3, a_2^1 \equiv a_2^4)$ that three gametes a^1 , a^2 and b^3 match at locus 1, and that two gametes a^1 and a^4 match at locus 2.

The graph G_2 corresponds to $\mathbb{P}(a_1^1 \equiv b_1^2, a_1^1 \equiv a_1^3, b_1^2 \equiv a_1^3, a_2^3 \equiv a_2^4)$ and the graph G_3 corresponds to $\mathbb{P}(b_1^1 \equiv a_1^2, b_1^1 \equiv a_1^3, a_1^2 \equiv a_1^3, b_2^1 \equiv a_2^4)$. Note that the superscripts 1, 2, 3, and 4 allow us to refer to specific vertices, but otherwise they have no meaning. In particular, two graphs obtained by permuting superscripts are equivalent. For example, graphs G_1 and G_2 are isomorphic, while G_1 and G_3 are not.

Fig. 2.

Graphical representation of two-locus match probabilities. From left to right, Q^U is equal to $(1 - \mu_1)^2(1 - \mu_2)^2$, $(1 - \mu_2)^2$, and $(1 - \mu_1)^2(1 - \mu_2)^2$, respectively.

Fig. 3. Illustration of vertex-recolor operations on match graphs for two loci. In the top figure, vertex a^1 is recolored and the label changes to b^1 ; this event occurs with probability $m_{\alpha\beta}$. In the bottom figure, vertex b^2 is recolored (and the label changes to a^2) with probability $m_{\beta\alpha}$.

Fig. 4.

Illustration of vertex-split operations on match graphs for two loci. In, G_1^R vertices a^1 and b^2 each have δ -degree 2, so there can be up to two split operations. The top figure corresponds to there being no split. In the middle figure only vertex a^1 is split, while both a^1 and b^2 are split in the bottom figure.

Fig. 5.

Illustration of vertex-merge operations on match graphs for two loci. In the top figure, there are two different vertex-merge operations that can transform G_1^C into G_1^P , explaining the factor of 2 in the associated probability. Specifically, a^1 and a^4 can be merged into a single vertex, or a^3 and a^4 can be merged. In the bottom figure, a^1, a^3 and a^4 are merged into a single vertex. Vertices from different subpopulations cannot be merged.

Fig. 6.

Schematic summary of the graphical approach of Song and Slatkin (2007) modified to account for population subdivision. Our notation reflects the operation to be performed: the vertices of G^M will be recolored to account for migration, the vertices of G_i^R will be split to account for recombination and finally the vertices of G_i^C will be merged to account for coalescence. The offspring graph G^O on the left can be written as a linear combination of the resulting parental graphs, G_i^P , on the right. The coefficients correspond to all possible single generation evolutionary histories, starting with the offspring graph.

$$\begin{aligned} \dot{p}_{ii}^{(1)} &= (1-m) \left[(1-m) p_{ii}^{(1)} + (1-m) p_{ii}^{(1)} \right] \\ \dot{p}_{ii}^{(2)} &= (1-m) \left[(1-m) p_{ii}^{(2)} + (1-m) p_{ii}^{(2)} \right] \\ \dot{p}_{ii}^{(3)} &= (1-m) \left[(1-m) p_{ii}^{(3)} + (1-m) p_{ii}^{(3)} \right] \end{aligned}$$

Fig. 7. Graphical representation of the closed system of equations involving one-locus match probabilities for locus i . Recall that the probability of migration is linearized.

Fig. 8.
The match graphs involved in the two-locus haplotypic match probability. These probabilities form a closed system of equations comprising 26 variables and 26 equations.

Fig. 9.

The match graphs corresponding to the two-locus haplotypic match probability for two gametes from subpopulation α , two from subpopulation β and one from subpopulation α and one from β .

$$\begin{aligned}
 \mathbb{P}_g(\{a_i, \bar{a}_i\} \equiv \{a'_i, \bar{a}'_i\}) &= 2 \times \text{Diagram 1} \\
 \mathbb{P}_g(\{b_i, \bar{b}_i\} \equiv \{b'_i, \bar{b}'_i\}) &= 2 \times \text{Diagram 2} \\
 \mathbb{P}_g(\{a_i, \bar{a}_i\} \equiv \{b_i, \bar{b}_i\}) &= 2 \times \text{Diagram 3}
 \end{aligned}$$

Fig. 10.
The one-locus genotypic match probabilities for two pairs of alleles at locus i .

$$\begin{aligned} \text{Pr}(A_1A_2|A_1A_2) &= \frac{1}{4} \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \right) \\ \text{Pr}(A_1A_2|A_1a_2) &= \frac{1}{4} \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \right) \\ \text{Pr}(A_1A_2|a_1A_2) &= \frac{1}{4} \left(\frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \right) \end{aligned}$$

Fig. 11.
The two-locus genotypic match probabilities for two pairs of two-locus gametes.

Fig. 12.

Dependency of the two-locus match probabilities on F_{ST} for F_{ST} values from 1% to 5%. The plots on the left correspond to the cryptic subdivision case where the two individuals sampled (e.g. the defendant and the crime-scene sample) are from the entire population. The plots on the right are for the same-subpopulation scenario where the two individuals sampled are from the same subpopulation, N_α . We fix the mutation rate $\mu = 1 \times 10^{-4}$ and the total population size $N = 10,000$. The top (respectively, bottom) plots are for the haplotypic (respectively, genotypic) match probabilities. The relevant column for the cryptic subdivision scenario (respectively, same-subpopulation scenario) is the “Random” (respectively, “Within α ”) column from Tables 3 and 5. For the no subdivision case (“No sub”) the solid lines are the same for the two scenarios for the haplotypic case (respectively, genotypic case), i.e. the “No sub” row of Table 3 (respectively, Table 5). For the asymmetric case, we fix the migration rate $m_{\beta\alpha} = 1 \times 10^{-4}$, so the subpopulation size N_α increases as F_{ST} increases from 1% to 5%. The plots for the one-locus case are similar and therefore not shown.

Table 1

Summary of notation.

Notation	Signification
$2N$	The total number of gametes in each generation.
α, β	Subpopulation labels.
$2N_\alpha, 2N_\beta$	Number of gametes in subpopulations α and β , respectively.
μ_i	Mutation rate at locus i per generation, per gamete (independent of the subpopulation label).
$m_{\alpha\beta}, m_{\beta\alpha}$	Backward migration rate from α to β , and from β to α , respectively, per generation, per gamete.
x_i	Allele at locus i of a gamete from an unspecified subpopulation.
$\mathbf{x}$	A haplotypic sequence $\mathbf{x} = x_1x_2$.
a_i	Allele at locus i of a gamete from subpopulation α .
$\mathbf{a}$	A haplotypic sequence $\mathbf{a} = a_1a_2$.
b_i	Allele at locus i of a gamete from subpopulation β .
$\mathbf{b}$	A haplotypic sequence $\mathbf{b} = b_1b_2$.
$x_i', \mathbf{x}'$	For locus i , alleles x_i and x_i' match. For all loci i , allele x_i of $\mathbf{x}$ and allele x_i' of $\mathbf{x}'$ match.

Table 2

One-locus haplotype match probabilities for $N = 10,000$.

μ	$m_{a\beta}$	$m_{\beta a}$	N_a	N_β	F_{ST}	$\mathbb{P}_H(x_i = x'_i)$		
							Within α	Within β Random
Sym	1×10^{-4}	1×10^{-2}	1×10^{-2}	5000	0.001	2.01×10^{-1}	2.01×10^{-1}	2.00×10^{-1}
	1×10^{-4}	1×10^{-4}	1×10^{-4}	5000	0.077	2.50×10^{-1}	2.50×10^{-1}	1.87×10^{-1}
	1×10^{-4}	1×10^{-6}	1×10^{-6}	5000	0.197	3.31×10^{-1}	3.31×10^{-1}	1.67×10^{-1}
Asym	1×10^{-4}	1×10^{-3}	1×10^{-4}	9091	0.018	3.30×10^{-1}	1.99×10^{-1}	1.97×10^{-1}
	1×10^{-4}	1×10^{-5}	1×10^{-4}	9091	0.064	2.06×10^{-1}	5.97×10^{-1}	1.90×10^{-1}
	No sub	1×10^{-4}	–	10,000	–		2.00×10^{-1}	
Sym	1×10^{-3}	1×10^{-2}	1×10^{-2}	5000	0.001	2.54×10^{-2}	2.54×10^{-2}	2.43×10^{-2}
	1×10^{-3}	1×10^{-4}	1×10^{-4}	5000	0.020	4.38×10^{-2}	4.38×10^{-2}	2.39×10^{-2}
	1×10^{-3}	1×10^{-6}	1×10^{-6}	5000	0.024	4.75×10^{-2}	4.75×10^{-2}	2.38×10^{-2}
Asym	1×10^{-3}	1×10^{-3}	1×10^{-4}	9091	0.011	1.26×10^{-1}	2.54×10^{-2}	2.41×10^{-2}
	1×10^{-3}	1×10^{-5}	1×10^{-4}	9091	0.019	2.65×10^{-2}	2.00×10^{-1}	2.39×10^{-2}
	No sub	1×10^{-3}	–	10,000	–		2.44×10^{-2}	

The probability of mutation per generation per gamete is assumed to be μ for all loci. Subpopulations α and β have $2N_a$ and $2N_\beta$ gametes, respectively, and $m_{a\beta}$ denotes the backward migration fraction from α to β ; $m_{\beta a}$ is similarly defined. “Sym” indicates symmetric migration; “Asym” indicates asymmetric migration, and “No sub” indicates a single randomly mating population. “Within α ” and “Within β ” indicate the match probabilities for two different individuals drawn randomly from the same subpopulation. “Random” indicates the match probability for two different individuals drawn randomly from the entire combined population, computed as described in (4.2).

Two-locus haplotypic match probabilities and their ratios R_h to the approximate two-locus haplotypic match probabilities obtained using the product rule.

Table 3

μ	$m_{a\beta}$	$m_{\beta a}$	N_a	N_β	F_{ST}	$\mathbf{P}_b(x \equiv x')$		$\mathbf{R}_b(x \equiv x')$			
						Within α	Within β	Random	Within α	Within β	Random
Sym	1×10^{-4}	1×10^{-2}	5000	5000	0.001	4.03×10^{-2}	4.03×10^{-2}	3.99×10^{-2}	1.001	1.001	1.000
	1×10^{-4}	1×10^{-4}	5000	5000	0.077	6.25×10^{-2}	6.25×10^{-2}	3.91×10^{-2}	1.000	1.000	1.000
	1×10^{-4}	1×10^{-6}	5000	5000	0.197	1.10×10^{-1}	1.10×10^{-1}	5.48×10^{-2}	1.000	1.000	1.000
Asym	1×10^{-4}	1×10^{-3}	909	9091	0.018	1.09×10^{-1}	3.98×10^{-2}	3.91×10^{-2}	1.002	1.000	1.000
	1×10^{-4}	1×10^{-5}	9091	909	0.064	4.25×10^{-2}	3.57×10^{-1}	3.93×10^{-2}	1.000	1.001	1.000
No sub	1×10^{-4}	–	10,000	–	–	4.00×10^{-2}			1.000		
Sym	1×10^{-3}	1×10^{-2}	5000	5000	0.001	6.78×10^{-4}	6.78×10^{-4}	6.09×10^{-4}	1.048	1.048	1.027
	1×10^{-3}	1×10^{-4}	5000	5000	0.020	1.95×10^{-3}	1.95×10^{-3}	9.81×10^{-4}	1.016	1.016	1.016
	1×10^{-3}	1×10^{-6}	5000	5000	0.024	2.29×10^{-3}	2.29×10^{-3}	1.14×10^{-3}	1.013	1.013	1.013
Asym	1×10^{-3}	1×10^{-3}	909	9091	0.011	1.61×10^{-2}	6.65×10^{-4}	7.08×10^{-4}	1.012	1.027	1.024
	1×10^{-3}	1×10^{-5}	9091	909	0.019	7.19×10^{-4}	4.01×10^{-2}	9.27×10^{-4}	1.025	1.003	1.017
No sub	1×10^{-3}	–	10,000	–	–	6.09×10^{-4}			1.027		

See the caption of Table 2 for notational convention. The total population size is assumed to be $N = 10,000$ and the loci are assumed to be unlinked (i.e., $r = \frac{1}{2}$).

Table 4

One-locus genotypic match probabilities for $N = 10,000$.

μ	$m_{a\beta}$	$m_{\beta a}$	N_a	N_β	F_{ST}	$\mathbb{P}_g(\{x_i, \bar{x}_i\} \equiv \{x'_i, \bar{x}'_i\})$		
						Within α	Within β	Random
Sym	1×10^{-4}	1×10^{-2}	1×10^{-2}	5000	0.001	6.72×10^{-2}	6.72×10^{-2}	6.65×10^{-2}
	1×10^{-4}	1×10^{-4}	1×10^{-4}	5000	0.077	1.00×10^{-1}	1.00×10^{-1}	6.48×10^{-2}
	1×10^{-4}	1×10^{-6}	1×10^{-6}	5000	0.197	1.65×10^{-1}	1.65×10^{-1}	8.26×10^{-2}
Asym	1×10^{-4}	1×10^{-3}	1×10^{-4}	909	0.018	1.65×10^{-1}	6.64×10^{-2}	6.53×10^{-2}
	1×10^{-4}	1×10^{-5}	1×10^{-4}	909	0.064	7.05×10^{-2}	4.46×10^{-1}	6.53×10^{-2}
	1×10^{-4}	–	10,000	–	–	6.67×10^{-2}		
Sym	1×10^{-3}	1×10^{-2}	1×10^{-2}	5000	0.001	1.26×10^{-3}	1.26×10^{-3}	1.16×10^{-3}
	1×10^{-3}	1×10^{-4}	1×10^{-4}	5000	0.020	3.67×10^{-3}	3.67×10^{-3}	1.85×10^{-3}
	1×10^{-3}	1×10^{-6}	1×10^{-6}	5000	0.024	4.31×10^{-3}	4.31×10^{-3}	2.15×10^{-3}
Asym	1×10^{-3}	1×10^{-3}	1×10^{-4}	909	0.011	2.82×10^{-2}	1.26×10^{-3}	1.33×10^{-3}
	1×10^{-3}	1×10^{-5}	1×10^{-4}	909	0.019	1.37×10^{-3}	6.66×10^{-2}	1.68×10^{-3}
	1×10^{-3}	–	10,000	–	–	1.16×10^{-3}		

See the caption of Table 2 for notational convention.

Table 5

Two-locus genotypic match probabilities and their ratios R_g to the approximate two-locus genotypic match probabilities obtained using the product rule.

μ	$m_{a\beta}$	m_{ba}	N_a	N_β	F_{ST}	$\mathbb{P}_g(\{x_i, \bar{x}_i\} \equiv \{x'_i, \bar{x}'_i\})$		$R_g(\{x_i, \bar{x}_i\} \equiv \{x'_i, \bar{x}'_i\})$			
						Within α	Within β	Random	Within α	Within β	Random
Sym	1×10^{-4}	1×10^{-2}	5000	5000	0.001	4.51×10^{-3}	4.51×10^{-3}	4.43×10^{-3}	1.001	1.001	1.000
	1×10^{-4}	1×10^{-4}	5000	5000	0.077	1.00×10^{-2}	1.00×10^{-2}	5.45×10^{-3}	1.001	1.001	1.001
	1×10^{-4}	1×10^{-6}	5000	5000	0.197	2.71×10^{-2}	2.71×10^{-2}	1.36×10^{-2}	1.000	1.000	1.000
Asym	1×10^{-4}	1×10^{-3}	909	9091	0.018	2.72×10^{-2}	4.41×10^{-3}	4.36×10^{-3}	1.002	1.000	1.000
	1×10^{-4}	1×10^{-5}	9091	909	0.064	4.97×10^{-3}	1.99×10^{-1}	5.83×10^{-3}	1.000	1.001	1.001
	No sub	1×10^{-4}	10,000	–	–	4.44×10^{-3}			1.000		
Sym	1×10^{-3}	1×10^{-2}	5000	5000	0.001	1.67×10^{-6}	1.67×10^{-6}	1.39×10^{-6}	1.049	1.049	1.030
	1×10^{-3}	1×10^{-4}	5000	5000	0.020	1.37×10^{-5}	1.37×10^{-5}	6.84×10^{-6}	1.017	1.017	1.017
	1×10^{-3}	1×10^{-6}	5000	5000	0.024	1.88×10^{-5}	1.88×10^{-5}	9.41×10^{-6}	1.013	1.013	1.013
Asym	1×10^{-3}	1×10^{-3}	909	9091	0.011	8.05×10^{-4}	1.64×10^{-6}	8.03×10^{-6}	1.015	1.027	1.017
	1×10^{-3}	1×10^{-5}	9091	909	0.019	1.92×10^{-6}	4.45×10^{-3}	3.84×10^{-5}	1.025	1.004	1.005
	No sub	1×10^{-3}	10,000	–	–	1.38×10^{-6}			1.027		

See the caption of Table 2 for notational convention. The total population size is assumed to be $N = 10,000$ and the loci are assumed to be unlinked (i.e., $r = \frac{1}{2}$).