PROGRAM NOTE

FAMSPHERE: a computer program for parental allocation from known genotypic pools

ANTONIO CARVAJAL-RODRÍGUEZ

Departamento de Bioquímica, Genética e Inmunología. Universidad de Vigo, 36310 Vigo, Spain

Abstract

FAMSPHERE **is a computer program that implements a distance-based method to allocate families based on microsatellite marker information. It formalizes the exclusion method considering genotyping errors or mutations. Contrary to the standard exclusion method,** FAMSPHERE **provides a way to resolve conflicts in the case of multiple paternities. The information linked to the defined distances is similar to that of the transmission probabilities used in the likelihood approaches. However, in** FAMSPHERE **the information of loci is used as a whole, contrary to the likelihood approaches.** FAMSPHERE **is free software and can be downloaded from www.uvigo.es/webs/c03/webc03/XENETICA/XB2/antonio/famsphere/famsphere.htm**

Keywords: exclusion method, genotyping error, microsatellites, parental allocation, stepwise mutation model

Received 31 July 2006; revision accepted 28 September 2006

The importance of paternity analysis based on genetic techniques has become more and more important during the past decade especially with the advent of codominant molecular markers as microsatellites, single nucleotide polymorphism (SNP), allozymes, etc. Currently, several techniques and software allow for parentage analysis in natural and experimental populations (Jones & Ardren 2003; Duchesne *et al*. 2005). A useful and intuitive approach is the exclusion method. This is focused on the offspring genotypes to exclude certain father and/or mother genotypes. However, when multiple parents cannot be excluded then Bayesian and likelihood methods are preferred (Hedrick 2005). In this study, a new method for parental allocation based on microsatellite marker information is presented. Thus, we will define some useful interindividual measurements which will allow us to identify paternity relationships among individuals. The usefulness of the mentioned measurements which can be considered as distances between individuals in the space of genotypes relies on the fact that under the microsatellite evolution stepwise model, allele lengths contain historical information. Thus, two alleles of similar length can be considered to be evolutionary closer than two alleles of very different length and in this way, different useful distances between alleles can be defined (Coulson *et al*. 1998). In the present work we will first consider a kind of special 'numbers' called pairs (the two alleles of a given locus) and define an n-dimensional space for them. In such space, the genotype for n loci of an offspring is considered as an n-dimensional sphere. Accordingly, the centre of such sphere is the diploid genotype of the individual and the radius is a magnitude somewhat related with mutation and genotyping error. The paternity assignment of parental genotypes which are considered as 'points' in the defined space is modelled in terms of the inclusion or exclusion in such spheres. The model when parameterized with radius 0 is just a formalization of the strict paternity exclusion criteria but allowing for the application of some rules to distinguish among multiple parentage candidates.

The model and the assignment criteria are briefly formally described as follows. Let define an n-dimensional space E^n of pairs of values $(a, b) \in E$ and an ordination property $a \leq b$. If the pair (a, b) holds the ordination property we say that $(a, b) \in E_o$. Thus, we also define an ordination function o such that given a nonordered pair $(a, b) \in E$, then $o(a, b) = (c, d) \in E_o$ with $c = \min(a, b)$ and $d = \max(a, b)$. We also define the sum of pairs $(a, b) + (c, d) = (a + c, b + d)$ which holds for both E and E_o.

Correspondence: Antonio Carvajal-Rodríguez, Fax: 34 986812556; E-mail: acraaj@uvigo.es

In the space E_o^n a point $P' = [(p_{11}, p_{12}),(p_{21}, p_{22}), \ldots, (p_{n1}, p_{n2})] = (P_1', P_2', \ldots, P_n') \in E_o^n$ is said to be included in the sphere S of centre $C' = [(c_{11}, c_{12}),(c_{21}, c_{22}), \ldots ,(c_{n1}, c_{n2})] = (C_1', C_2', \ldots, C_n') \in E_o^n$ and radius $R = (r, r) \in E_o$ if, given $D_1(P', C') = (A, B) \in E_o$ then $A \leq 0$ and $B \leq r$ where

$$D_1(P', C') = \sqrt{\Sigma_{i=1}^n o(dparMin(P_i', C_i'))} \quad \text{and}$$

$$dparMin(P_i', C_i') = \Big[\min\Big(|P_{i1}' - C_{i1}'|, |P_{i2}' - C_{i1}'|\Big),$$
$$\min\Big(|P_{i1}' - C_{i2}'|, |P_{i2}' - C_{i2}'|\Big)\Big] = (a, b) \in E.$$

Otherwise, if $A \leq r$ we say that P' is *semi-included* in S, note that inclusion implies the occurrence of semi-inclusion but the reverse is not true. Finally, if $A > r$ then P' is excluded from S.

Paternity rules

Consider the sphere S of centre $C = (C_1, C_2, \ldots, C_n) \in E^n$ and radius $R \in E_o$ and the points $F = (F_1, F_2, \ldots, F_n)$ and $M = (M_1, M_2, \ldots, M_n) \in E^n$ and the corresponding $C' = o(C)$, $F' = o(F)$ and $M' = o(M) \in E_o^n$. Then the following rules apply:

1 F is a putative father of individual C if F' is included or semi-included into the sphere of centre C' and radius R.

2 M is a putative mother of individual C if M' is included or semi-included into the sphere of centre C' and radius R.

3 Let M and F be, respectively, a putative mother and a putative father of C, then to assign F and M as the family of C let consider in the space E^n each of the n positive planes with axis at each corresponding pair $C_i, I \in \{1, \ldots, n\}$. We say that F and M are compatible parents of C, i.e. they are a family of C, if and only if, for each of the positive planes the translations $C_i - F_i$ and $C_i - M_i$ lay at a distance minor than the radius R for both axes of each plane (Fig. 1).

Multipaternity rules

Let M be a mother of an offspring C, F_1 and F_2, two compatible fathers and let M', C', F_1' and F_2' be the corresponding ordered points belonging to E_o^n such that F_1' and F_2' were both included or semi-included into the sphere of centre C' and radius R. Then, F_1 is chosen as the most probable father of C if:

1 F_1' is the closest to C' concerning the measurement D_1, i.e. F_1' is included or semi-included into the sphere of centre C' and radius $D_1 (F_2', C')$ but F_2' is not included or semi-included in the sphere of centre C' with radius $D_1 (F_1', C')$.

2 F_1' is the closest to C' concerning the measurement D_2. That is, if both spheres from (1) are equal, F_1 will be the most probable father if F_1' is included or semi-included into the sphere of centre C' and radius $D_2 (F_2', C')$ with
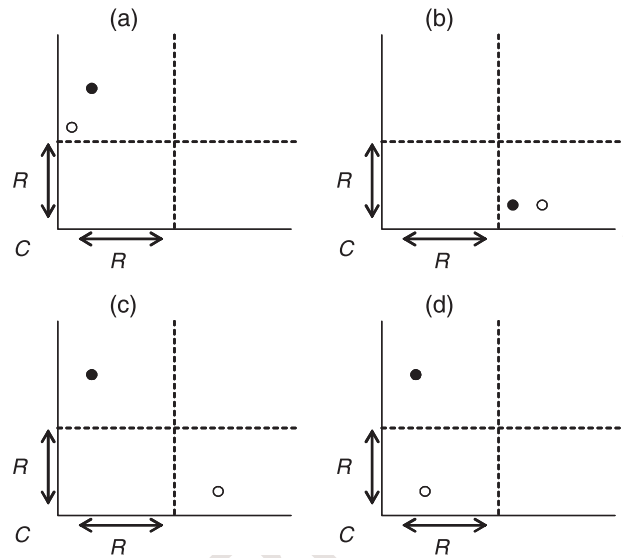

Fig. 1 Family allocation with one locus. The filled point represents the putative mother genotype and the empty point the putative father. C is the genotype of the son and R the margin of genotyping error and mutation assumed. The two axes represent the two alleles. (a) Both genotypes have one allele inside the allowed region for paternity. However, only one axis (allele) is covered. Then, these two parents are not a compatible family. (b) Similar case to A, but concerning the other axis (allele). (c) Both genotypes have one allele inside the allowed region for paternity and both axes are covered. Thus, this is a compatible family for the offspring C. (d) The father genotype has one allele close enough to both axes. The other genotype is close enough from just one axis. Then, this is also a compatible family.

$$D_2(F_2', C') = \sqrt{\Sigma_{i=1}^n o(dparMax(F_{2i}', C_i'))} \quad \text{where}$$

$$dparMax(F_{2i}', C_i') = \Big[\max\Big(|F_{2i1}' - C_{i1}'|, |F_{2i2}' - C_{i1}'|,\Big)$$
$$\max\Big(|F_{2i1}' - C_{i2}'|, |F_{2i2}' - C_{i2}'|\Big)\Big] = (a, b) \in E.$$

As previously stated, the above is a formalization of the exclusion method (Jones & Ardren 2003) taking into account genotyping errors or mutation. Indeed, when multiple-paternity is not excluded the above rules provide a way to assign paternity, which is not present in the standard exclusion method. Furthermore, this new method uses the information of all loci as a whole instead of using each locus individually as in likelihood approaches. Still, the information linked to the comparison of distances D_1 and D_2 is similar to that of the transmission probabilities between heterozygous or homozygous loci used in the likelihood approach.

Example

A multipaternity rules application with two loci example is considered. The offspring genotype for the first locus is 122 130 and for the second locus is 120 184. The selected

(a)

SONS	MOTHERS	FATHERS
1	1003 (NON-TOTAL MATCH)	2002 (TOTAL MATCH)
	1019 (TOTAL MATCH)	2013 (NON-TOTAL MATCH))
	1035 (TOTAL MATCH)	2017 (NON-TOTAL MATCH))
	1040 (NON-TOTAL MATCH)	
7	1005 (TOTAL MATCH)	2004 (TOTAL MATCH)
	1014 (TOTAL MATCH)	2008 (TOTAL MATCH))
	1015 (TOTAL MATCH)	2014 (TOTAL MATCH))
	1016 (TOTAL MATCH)	
	1022 (TOTAL MATCH)	
	1026 (TOTAL MATCH)	
	1042 (TOTAL MATCH)	
10	1016 (TOTAL MATCH)	2002 (TOTAL MATCH)
		2010 (TOTAL MATCH)

(b)

3 sons with assigned family ($R = 0$):

SON	MOTHER	FATHER
1	1035	2002
7	1042	2008
10	1016	2010

Fig. 2 FAMSPHERE output files. (a) The *parent.out* file with a list of putative parents. Mothers are identified with a number lower than 2000 and fathers with a number higher than 2000. (b) The *family.out* file with a list of offspring with the selected candidate family. *R*, radius.

mother genotype for the first and the second loci is 122 166 and 120 184, respectively. That is, this mother provides the allele 122 at the first locus but could provide both alleles at the second locus. Consider a putative father genotype 122 130 and 144 184 for the first and the second loci, respectively. Given the selected mother, this father has to provide the allele 130 at the first locus and the allele 184 for the second one. However, consider now a second father 130 142 for the first locus and 184 184 for the second locus. Concerning to the first locus, this father provides the allele 130 with the same probability than the first father. For the second locus, this second father is homozygous and provides the allele 184 with a higher probability than the first putative father. In terms of the spheres model it could be said that both fathers are equal concerning D_1, i.e. D_1 is of the kind of $(0, y)$, but the second father has a lower D_2 distance than the first one and, in consequence, the model will select the second parent to conform the family of the analysed offspring.

The method above is implemented in a simple program called FAMSPHERE, which can be freely downloaded at www.uvigo.es/webs/c03/webc03/XENETICA/XB2/antonio/famsphere/famsphere.htm.

Thus, FAMSPHERE is a C + + computer program and is available as executable for Windows, Mac and Unix-like operative systems.

When using FAMSPHERE, the user is only asked to input the radius to be considered, the default value is zero (strict

paternity exclusion criterium). FAMSPHERE uses as input two different files. The first file includes the progenitor genotypes, and it should be called *parents.PAR*. The second file refers to the offspring genotypes and it is called *sons.PRO*. The format of the two input files is very similar to that of PROBMAX program (Danzmann 1997) but simpler since it is not necessary to indicate any mating groups because FAMSPHERE will consider all combinations among the putative parents. Each pair of columns represents the two alleles of a locus and rows represent individuals. In the progenitor file, the last column represents the individual identification number, which must be less than 2000 for mothers and equal, or higher than 2000 for fathers. This column is not present at the offspring file. More details and example input files are provided at the program web page.

After the analysis, the user obtains two output files (Fig. 2). The first file is called *parent.out* with putative list of fathers and mothers for each offspring, independently of their compatibility to conform a family (see Fig. 2a). The second file is called *family.out* with putative most plausible family (mother + father) assigned by FAMSPHERE for each offspring following the rules described above (see Fig. 2b).

As a conclusion, FAMSPHERE should be faster and more precise than other programs based on the exclusion method. However, there are some other more sophisticated programs which incorporate a simulation step to detect

the most convenient set of loci for allocation. FAMSPHERE does not incorporate any simulation tool however, having the advantage of its fastness and simplicity of using. Therefore, important facts of FAMSPHERE are its two-step method (first compute putative parents and then compute families) and second, the use of a based distance method with the application of multiple paternity decision rules.

References

Coulson TN, Pemberton JM, Albon SD *et al.* (1998) Microsatellites reveal heterosis in red deer. *Proceedings of the Royal Society of London. Series B, Biological Sciences*, **265**, 489–495.

Danzmann RG (1997) PROBMAX: a computer program for assigning unknown parentage in pedigree analysis from known genotypic pools of parents and progeny. *Journal of Heredity*, **88**, 333.

Duchesne P, Castric T, Bernatchez L (2005) PASOS (parental allocation of singles in open systems): a computer program for individual parental allocation with missing parents. *Molecular Ecology Notes*, **5**, 701–704.

Hedrick PW (2005) *Genetics of Populations*, 3rd edn. Jones. & Bartlett Publishers, Boston, Massachusetts.

Jones AG, Ardren WR (2003) Methods of parentage analysis in natural populations. *Molecular Ecology*, **12**, 2511–2523.