SHORT COMMUNICATION   **Open Access**

A novel recursive algorithm for the calculation of the detailed identity coefficients

Luis Alberto García-Cortés

Abstract

Background: A recursive algorithm to calculate the fifteen detailed coefficients of identity is introduced. Previous recursive procedures based on the generalized coefficients of kinship provided the detailed coefficients of identity under the assumption that the two individuals were not an ancestor of each other.

Findings: By using gametic relationships to include three, four or two pairs of gametes, we can obtain these coefficients for any pair of individuals. We have developed a novel linear transformation that allows for the calculation of pairwise detailed identity coefficients for any pedigree given the gametic relationships. We illustrate the procedure using the well-known pedigree of Julio and Mencha, which contains 20 Jicaque Indians of Honduras, to calculate their detailed coefficients.

Conclusions: The proposed algorithm can be used to calculate the detailed identity coefficients of two or more individuals with any pedigree relationship.

Background

The 15 detailed states of identity were first described by Harris [1] and Gillois [2]. Throughout this paper, they will be referred to identity coefficients as described by Cockerham [3].

To circumvent the absence of recurrence rules to obtain identity coefficients, Karigl [4], following the rules for generalized kinship coefficients [1,3], obtained identity coefficients using a triangular linear transformation. This transformation provided the 9 condensed identity coefficients for any pair of individuals, but the 15 detailed identity coefficients can only be calculated using Karigl's method under the assumption that neither of the two individuals is an ancestor of the other. To obtain generalized kinship coefficients, Lange and Sinsheimer [5] described an alternative way, which can calculate the detailed coefficients of identity without this limitation. Unfortunately, the implementation of the latter method is not a triangular linear transformation.

An alternative to using multiple kinship coefficients is the use of multiple gametic relationships. These relationships, called chromosome pedigrees by Donnelly [6], have been succesfully used to account for dominance in linear models [7-9]. Here, we use these multiple gametic relationships to develop a new method to calculate the 15 detailed identity coefficients.

Generalized gametic relationships

The coancestry between two individuals X and Y, ϕ_{XY}, is usually calculated following simple recurrence rules. These rules can be implemented using tabular methods or languages with recursive function support. To calculate the whole set of coancestries for a given pedigree, only two formulae are required [10]:

$$\phi_{XX} = (1 + \phi_{FM})/2,$$
$$\phi_{XY} = (\phi_{FY} + \phi_{MY})/2, \qquad (1)$$

where F and M are the father and the mother of individual X, respectively. These equations operate successively over pairs (X, Y) where X is assumed to be more recent than Y. Let x and x' be the paternally and maternally inherited copies at a given locus carried by individual X and y and y' the corresponding copies carried by individual Y. The coancestry between X and Y can then be written as:

$$\phi_{XY} = \frac{1}{4}\left[P\left(x \equiv y\right) + P\left(x \equiv y'\right) + P\left(x' \equiv y\right) + P\left(x' \equiv y'\right)\right],$$

where $\equiv$ stands for identical by descent (IBD).

Correspondence: garcia@inia.es
Departamento de Mejora Genética, Instituto Nacional de Investigación Agraria, Ctra. de La Coruña km. 7.5, 28040 Madrid, Spain

Analogous relationships have been described for gametes [6] as $\psi_{ab} = P(a \equiv b)$ and recurrence rules have been developed for their pairwise relationships [6,7]:

$$\psi_{aa} = 1,$$
$$\psi_{ab} = 1/2\left(\psi_{gb} + \psi_{hb}\right), \tag{2}$$

where a and b denote two gametes in the pedigree. Both g and h are the direct ancestral gametes of a, that is, the gametes of the father or the mother if a is a paternal or maternal gamete, respectively. Although Equations (1) and (2) are closely related, they have different interpretations.

We use the *three-way* (ψ_{abc}), the *four-way* (ψ_{abcd}) and the *two-pair* ($\psi_{ab,cd}$) gametic relationships as counterparts of the conventional generalized kinship coefficients [4]. These generalized gametic relationships correspond to the probability of three or four gametes to be IBD. Note that these multiple gametic relationships correspond to multiple gametic identities, regardless of the identity by descent with other gametes. For instance, for individuals X and Y, whose paternally and maternally inherited gametes are described above,

$$\psi_{xx'y} = P\left(x \equiv x' \equiv y\right) = P\left(x \equiv x' \equiv y \equiv y'\right) + P\left(x \equiv x' \equiv y \neq y'\right).$$

$\psi_{ab,cd}$ is the probability that gametes a and b are IBD and simultaneously c and d are also IBD. For instance, for individuals X and Y,

$$\psi_{xx',yy'} = P\left(x \equiv x' \equiv y \equiv y'\right) + P\left(x \equiv x' \neq y \equiv y'\right).$$

The recursive formulae for the whole set of multiple gametic relationships are

$$\psi_{aa} = 1,$$
$$\psi_{ab} = \frac{1}{2}\left(\psi_{gb} + \psi_{hb}\right),$$
$$\psi_{aaa} = 1,$$
$$\psi_{aab} = \psi_{ab},$$
$$\psi_{abc} = \frac{1}{2}\left(\psi_{gbc} + \psi_{hbc}\right),$$
$$\psi_{aaaa} = 1,$$
$$\psi_{aaab} = \psi_{ab},$$
$$\psi_{aabc} = \psi_{abc}, \tag{3}$$
$$\psi_{abcd} = \frac{1}{2}\left(\psi_{gbcd} + \psi_{hbcd}\right),$$
$$\psi_{aa,aa} = 1,$$
$$\psi_{aa,ab} = \psi_{ab},$$
$$\psi_{ab,ac} = \psi_{abc},$$
$$\psi_{aa,bc} = \psi_{bc},$$
$$\psi_{ab,cd} = \frac{1}{2}\left(\psi_{gb,cd} + \psi_{hb,cd}\right).$$

For an easier implementation, Equation (3) can be summarized in a simple set of rules

1. In any *n-way* relationship, merge groups sharing a given gamete, that is, $\psi_{ab,ac} = \psi_{abac}$

2. In any *n-way* relationship, discard repeated gametes, i.e. $\psi_{aabc} = \psi_{abc}$ or $\psi_{aa,bc} = \psi_{a,bc}$.
3. Given that the probability of a gamete to be IBD to itself is 1, discard groups of identity including a single gamete, i.e. $\psi_{a,bc} = \psi_{bc}$.
4. Identities with a single gamete are 1 and identities with two or more founder gametes at the same group are 0.
5. Calculate $\psi_{a\theta} = \frac{1}{2}\left(\psi_{g\theta} + \psi_{h\theta}\right)$, where θ stands for any identity pattern and a for a gamete of the youngest individual. For instance $\psi_{ab,cd} = \frac{1}{2}\left(\psi_{gb,cd} + \psi_{hb,cd}\right)$ or $\psi_{abc} = \frac{1}{2}\left(\psi_{gbc} + \psi_{hbc}\right)$.

As long as these rules are correct and regardless of the number of gametes involved, they can be used to calculate identities involving more than two individuals.

The detailed identity coefficients

The 15 detailed coefficients of identity for individuals X and Y can be calculated from the generalized gametic relationships by a simple linear transformation. We will consider the 15 partitions or identity states as described by Gillois [2] and Jacquard [11], i.e.,

$$S_1 : x \equiv x' \equiv y \equiv y' \quad S_2 : x \equiv x' \equiv y \neq y' \quad S_3 : x \equiv x' \equiv y' \neq y$$
$$S_4 : x \equiv y \equiv y' \neq x' \quad S_5 : x' \equiv y \equiv y' \neq x \quad S_6 : x \equiv x' \neq y \equiv y'$$
$$S_7 : x \equiv x' \neq y \equiv y' \quad S_8 : x \neq x' \equiv y \equiv y' \quad S_9 : x \equiv y \neq x' \equiv y'$$
$$S_{10} : x \equiv y \neq x' \neq y' \quad S_{11} : x \neq y \neq x' \equiv y' \quad S_{12} : x \equiv y' \neq x' \equiv y$$
$$S_{13} : x \equiv y' \neq x' \neq y \quad S_{14} : x \neq y' \neq x' \equiv y \quad S_{15} : x \neq y' \neq x' \neq y$$
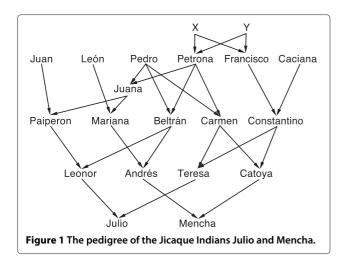
The detailed identity coefficients δ_i are the probabilities of each state S_i for a given pair of individuals.

Generalized kinships [4] are a linear transformation of the pairwise detailed identity coefficients. For instance, $\psi_{xx'yy'}$ requires the state S_1 to be true, then we can write $\psi_{xx'yy'} = \delta_1$. Similarly, $\psi_{xx'y}$ requires either S_1 or S_2 to be true and therefore $\psi_{xx'y} = \delta_1 + \delta_2$, and so on for all ψ. A set of 15 equalities is defined to set up the following linear system of equations.

$$
\begin{bmatrix}
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 1 & 1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
1 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
1 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1
\end{bmatrix}
\begin{bmatrix}
\delta_1 \\ \delta_2 \\ \delta_3 \\ \delta_4 \\ \delta_5 \\ \delta_6 \\ \delta_7 \\ \delta_8 \\ \delta_9 \\ \delta_{10} \\ \delta_{11} \\ \delta_{12} \\ \delta_{13} \\ \delta_{14} \\ \delta_{15}
\end{bmatrix}
=
\begin{bmatrix}
\psi_{xx'yy'} \\ \psi_{xx'y} \\ \psi_{xx'y'} \\ \psi_{xyy'} \\ \psi_{x'yy'} \\ \psi_{xx',yy'} \\ \psi_{xx'} \\ \psi_{yy'} \\ \psi_{xy,x'y'} \\ \psi_{xy} \\ \psi_{x'y'} \\ \psi_{xy',x'y} \\ \psi_{xy'} \\ \psi_{x'y} \\ 1
\end{bmatrix}
\tag{4}
$$

Figure 1 The pedigree of the Jicaque Indians Julio and Mencha.

The coefficient matrix in Equation (4) is triangular and is equivalent to formula (8) in [4]. Nevertheless, the right hand side here includes the generalized gametic relationships instead of their kinship counterparts.

An example: the detailed identity coefficients of the Jicaque Indians

We reanalyzed the pedigree of the Jicaque indians Julio and Mencha [12] presented in Figure 1 and previously analyzed in [4,11,13]. Previous analyses have focused on the identity coefficients between Julio and Mencha and between two of their progeny. Table 1 shows these results and those of two other pairs: Julio vs. one of their progeny and Mencha vs. one of their progeny.

Program source codes written in Fortran 90 and examples are available at https://github.com/agarcor21/IdentityCoefficients.

Discussion

The identity by descent of two gametes [6-9] is conceptually simpler than the kinship between individuals because it avoids any random sampling of genes. Although the number of gametes is larger than the number of individuals, its simple definition results in more intuitive recursive formulae and easier calculations.

Here, we have provided a novel method to derive the detailed identity coefficients [1,2]. Karigl derived a simple transformation to calculate these coefficients from the multiple kinship coefficients, which provided the 15 identity coefficients when the individuals were not ancestor-descendant related. Lange and Sinsheimer [5] first provided exact equations for these 15 coefficients without any assumption. Depending on the purpose of the study, researchers have used 7 or 9 condensed identity coefficients. For instance, in order to apply the identity coefficients to dominance models, 7 condensed coefficients can be used [8,9]. To our knowledge, the algorithm proposed in [9] has never been used to calculate the 15 detailed coefficients, but the formulas presented in their paper could be used to obtain them in a direct way.

We have proposed using multiple gametic relationships to calculate the 15 detailed identity coefficients. The linear transformation in this paper is similar to equation 8 of Karigl's approach [4]. Reordering the rows in Equation (4) to follow Karigl's pattern [4], both formulae only differ on the right hand side. Note that the meaning of both formulae is rather different.

The procedure presented here has been successfully implemented on small pedigrees. In large data sets, the number of calls to the recursive function will depend on the number of generations and the structure of the pedigree. Implementing the procedure in such scenarios

Table 1 Detailed coefficients of identity for four pairs involving Julio, Mencha and their progeny

	Julio-Mencha	Progeny-progeny	Julio-progeny	Mencha-progeny
δ_1	0.01025	0.06580	0.03467	0.03467
δ_2	0.02393	0.04333	0.08252	0.00000
δ_3	0.02490	0.04333	0.00000	0.08252
δ_4	0.02393	0.04333	0.06665	0.06665
δ_5	0.02490	0.04333	0.08228	0.08228
δ_6	0.00708	0.00729	0.00000	0.00000
δ_7	0.05103	0.02383	0.00000	0.00000
δ_8	0.05103	0.02383	0.00000	0.00000
δ_9	0.05127	0.24713	0.08228	0.06665
δ_{10}	0.10937	0.15900	0.29248	0.00000
δ_{11}	0.16992	0.15900	0.00000	0.29248
δ_{12}	0.00708	0.00729	0.06665	0.08228
δ_{13}	0.05103	0.02383	0.00000	0.29248
δ_{14}	0.05103	0.02383	0.29248	0.00000
δ_{15}	0.34326	0.08582	0.00000	0.00000

is beyond our goal, but path counting or graph theory-based methods [14,15] have been developed to improve the computing efficiency of this calculation.

References
1. Harris DL. Genotypic covariances between inbred relatives. Genetics. 1964;50:1319–48.
2. Gillois M. La relation d'identité en génétique. Ann Inst Henri Poincaré. 1964;B2:1–94.
3. Cockerham CC. Higher order probability functions of identity of alleles by descent. Genetics. 1971;69:235–46.
4. Karigl G. A recursive algorithm for the calculation of identity coefficients. Ann Hum Genet. 1981;45:299–305.
5. Lange K, Sinsheimer JS. Calculation of genetic identity coefficients. Ann Hum Genet. 1992;56:339–46.
6. Donnelly KP. The probability that related individuals share some section of genome identical by descent. Theor Popul Biol. 1983;23:34–63.
7. Smith SP, Allaire FR. Efficient selection rules to increase non-linear merit: application to selection. Genet Sel Evol. 1985;17:387–406.
8. de Boer IJM, Hoeschele I. Genetic evaluation methods for populations with dominance and inbreeding. Theor Appl Genet. 1993;86:245–258.
9. Lo LL, Fernando RL, Cantet RJC, Grossman M. Theory for modelling means and covariances in a two-breed population with dominance inheritance. Theor Appl Genet. 1995;90:49–62.
10. Emik L, Terrill CE. Systematic procedures for calculating inbreeding coefficients. J Hered. 1949;40:51–55.
11. Jacquard A. The genetic structure of populations. New York: Springer; 1973.
12. Chapman AM, Jacquard AM. Un isolat d'Amérique Centrale: Les indiens Jicaques du Honduras. I.N.E.P., editor. Génétique et Populations. Presses Universitaires de France; 1971.
13. Nadot R, Vaysseix G. Apparentement et identité, algorithme de calcul des coefficients d'identité. Biometrics. 1973;29:347–59.
14. Cheng E, Elliott B, Ozsoyoglu ZM. Efficient computation of kinship and identity coefficients on large pedigrees. J Bioinform Comput Biol. 2009;7:429–53.
15. Abney M. A graphical algorithm for fast computation of identity coefficients and generalized kinship coefficients. Bioinformatics. 2009;25:1561–3.