Variation in genetic identity within kinships

MARIANNE RASMUSON

Department of Genetics, University of Umeå, S-901 87 Umeå, Sweden

Genetic identity, which may be important for kin recognition, is the fraction of the genome that is identical by descent. It is, except for the parent–offspring relation, governed by probability and its variance depends on the number of segregating units during meiosis. Using the recombination index as an approximation of this number the variance for genetic identity has been estimated for different kinds of kinship.

Keywords: genetic similarity, identity variation, kin recognition, recombination index.

Introduction

Kin recognition, the faculty to distinguish near relatives from non-kin, is important both in models for inclusive fitness and inbreeding avoidance. For litter mates and closely coherent families, environmental conformity might be the cue to lasting recognition, but in many cases the recognition of kin seems to depend on genetic similarity. It is therefore of significance that in diploid, sexually reproducing organisms the gene identity is maximized to 50 per cent, this value being expected for both types of first degree relatives, parent–offspring and full sibs.

There is, however, a difference. In the absence of inbreeding the offspring has exactly half of its genes in common with each parent, disregarding the difference in genetic content of the sex chromosomes. The gene identity between parent and offspring is thus $1/2$ for all autosomal loci. Full sibs also have a probability of $1/2$ for gene identity by descent in autosomal loci but this is the expectation for a random variable. As the contributions from the male and the female parent are independent, both, one or none of the alleles of a locus can be identical, with probabilities $1/4$, $2/4$, and $1/4$, respectively. The total expectation of identity by descent thus becomes $1/2$ in a population without inbreeding, and the variance around this value is $1/8$.

In socially organized species it might be most important to distinguish first-degree relatives from less related or unrelated individuals. How are the genetic cues to be constructed to meet these needs? Apparently a genetic discrimination depending on the variation in one single locus has too high a sampling variance and thus gives an imperfect signal of relatedness. For two full sibs the expected gene identity is $\frac{1}{2}$ but in 25 per cent of the cases there is no identity. For half sibs and other kinships with an average probability of

identity of $1/4$, the expectation of no identity is 50 per cent. Thus these two types of relatedness cannot be easily distinguished from information at one single locus.

An accurate identification may require a large number of variable loci and in species with a sharp distinction capacity a large part of the genome may be involved. Thus the concept of gene identity can be expanded to parts of chromosomes or to whole chromosomes, and it is sufficient to refer to the *genetic identity* between two individuals of a specified relatedness, meaning the *fraction of their genome that is identical by descent*. Except for pairs of monozygous twins, with complete identity barring new mutations, and the parent–offspring relation, the genetic identity is also exposed to random fluctuations, although its variance is smaller than for single polymorphic loci: the exact values depend on many factors.

R. A. Fisher (1949) was the first to mention genetic identity as a random variable. When discussing the inbreeding coefficient in his *Theory of Inbreeding* he observed that there is variation in the extent of homozygosity between individuals with the same coefficient of inbreeding, i.e. with the same probability of two alleles being identical by descent, and also formulated ways to estimate the variance.

Later, estimates of the heterogenic fraction of the genome after inbreeding were obtained by Franklin (1977) from considerations of the joint identity by descent at two loci, and by Stam (1980) from assumptions about the distribution of the number of heterogeneous segments and their length. Both predict certain distributions of the junction sites that result from chromosomal recombinations in successive generations. These results concern variation in genome heterogeneity after inbreeding (mostly sib mating) but their analogy with gene identity between sibs and

various other types of relatives is obvious. In particular they show the dependence of heterogeneity upon the number of chromosomes and their total map length.

The human genome is extremely well studied and may serve as an example for evaluation of the variance around expected genetic identities between relatives. If the genetic identity was concerned only with polymorphic structural genes, which in humans have been estimated (by means of electrophoresis) to be between 6000 and 7000, and all of these segregated independently, the variance around the 50 per cent genetic identity for full sibs would be very small, with a standard deviation of less than 1 per cent.

This assumption, however, is obviously incorrect. At the formation of gametes it is the chromosomes that segregate, or more accurately, parts of the chromosomes. Segments of the chromosomes are transferred as units and the number of segregating units depends on the number of chromosomes and exchanges within the bivalents.

The number which should best correspond to the segregating units is the *recombination index* (*RI*), defined by Darlington (1939) as the haploid number of chromosomes plus the total number of chiasmata (*TC*). Thus $RI = n + TC$. The *RI* for human males is around 75 and for females it is estimated to be above 100.

There may of course be a certain lack of correspondence between the observed number of chiasmata and the exact number of exchanges. Chiasma movements and terminalizations may reduce the number of observable recombination sites whilst double exchanges may be unrecognizable. Furthermore, as recombination sites are not fixed, the chromosome segments will vary in length and also in number following different meiotic transmissions; this is important when following an ancestral genome through several transmissions as it will be broken down into an increasing number of smaller segments. It may also be argued that adjacent chromosomal parts, which are separated by a recombination event, are obligated to be included in different gametes, and thus not randomly distributed. This is, however, compensated for by the equally great probability that the gamete will contain one of the unbroken chromatids which have not taken part in the exchange.

Analysis of the total genome shows that the genetic identity for full sibs amounts to 0.50, but with variance of $\frac{1}{2} \times \frac{1}{2} \times 1/N$. Barash *et al.* (1978) discusses the influence of the number of chromosomes on the variance in relatedness among full sibs, but *N* should rightly be equated to the number of segregating units, i.e. $2RI$. Approximating *RI* to 100 would give a standard deviation of 0.035.

The question of variability in genetic identity within human sib pairs has been examined in detail in two

papers. Suarez *et al.* (1979) made simulations of the crossover distribution among male and female meiotic products and used the chiasma number and chromosome sizes reported for human males to estimate the standard deviation. Risch & Lange (1979) applied a model of recombination distribution and calculated the variance analytically. The standard deviations that they obtained from their calculations were 0.056 and 0.040, respectively. According to Risch & Lange, this difference is due to their assumption of a uniform distribution of independent chiasma positions, whereas Suarez *et al.* (1979) used a more localized chiasma distribution in their simulations.

The variance in genetic identity for other types of kinship has not been analysed in detail. Unlike full sibs, which involve a bilineal relationship, half sibs, ancestor-descendant and many other types of kinship are unilineal where only half the genome is derived from common ancestors. In the absence of inbreeding the genetic identity derived by descent cannot exceed 0.5. Using 100 as an average number of segregating units per gamete, and not taking the sex difference into consideration, the mean and variance for the number of shared units in half sibs becomes $\mu_s = \frac{1}{2} \times 100 = 50$ and $V_s = \frac{1}{2} \times \frac{1}{2} \times 100 = 25$. The genetic identity, defined as $r = s/N$ gives $\mu_r = 50/200 = 0.25$ and $V_r = 25/(200)^2 = 0.000625$ (Barash *et al.*, 1978).

For ancestor-descendant relatedness the same considerations are valid. The A-D2 (ancestor-second generation descendant) situation is equal to that of half sibs; A-D3 gives $\mu_s = 0.25 \times 100 = 25$ and $V_s = 0.25 \times 0.75 \times 100 = 18.75$. Thereafter, the probability of identity is halved in each successive generation. Uncle-niece and single first cousins are equivalent to A-D2 and second cousins to A-D5. Double first cousins represent a bilineal relationship where the probability of shared units is 0.25, giving $\mu_s = 0.25 \times 200 = 50$ and $V_s = 0.25 \times 0.75 \times 200 = 37.5$. The variance and standard deviation for genetic identity thus become higher than for the unilineal relationships with equal genetic identity. The expected values for genetic identity and the approximate standard deviations around that are given in Table 1.

When involving the entire genome there is thus no confusion of the genetic likeness between full and half sibs, the minimum value for genetic identity between full sibs being above 40 per cent whereas the maximum for half sibs stays below 30 per cent. Overlapping between different categories of relatedness is possible where the degree of genetic identity is lower, e.g. the genetic identity between two cousins may be as large as between an uncle and his niece. In more remote kinships the genetic identity is so low that there is a substantial chance of a zero identity, which means there is

Table 1 Expected genetic identity and its approximate standard deviation, assuming a genome with a recombination index of 100

Type of relatedness	Genetic identity (%)	s.d. (%)
Full sibs	50	3.53
Half sibs	25	2.50
Ancestor-descendant		
A-D1	50	0
A-D2	25	2.50
A-D3	12.5	2.17
A-D4	6.25	1.65
A-D5	3.125	1.21
Uncle-niece	25	2.50
Double first cousins	25	3.06
Single first cousins	12.5	2.17
Second cousins	3.125	1.21

no increased similarity over that of random pairs in the population.

Pairs of half sibs, grandparent-grandchild, uncle-niece, and double first cousins all have an expected genetic identity of 0.25, but differ in the number of transmissions that separate the members of a related pair. This may bias the estimated standard deviations because each meiotic recombination will decrease the certainty as to the origin of the chromosomes and therefore influence the variance of genetic identity among relatives. For grandparent-grandchild there is only one meiotic event where the ancestral genome can be divided into segregating units of a homologous or non-homologous kind, whereas double first cousins are separated from each other by 12 meioses. It seems obvious that such varying types of relationship also differ in variability around their expected genetic identity.

As the variance around the expected genetic identity depends on chromosomal and meiotic mechanisms it is species-specific. Applied estimates therefore have to be restricted to specified species and founded on knowledge of their karyotype. Species with fewer chromo-

somes and less recombination will have a higher variance. An extreme example of high chromosomal coherence is the dipterans with few chromosomes and no recombination in the male meiosis. The recombination index among vertebrates is usually higher. A survey of mammals (Burt & Bell, 1987) gives a male mean *RI* of 47 ± 3.2 for 32 species.

It can be assumed that phenotypic resemblance is correlated to genetic identity but thus far we do not know how much of the genome is involved in kin recognition. The ability to recognize multiple cues to genetic similarity will make use of different perceptive senses: vision, scent, hearing, which probably demand a well-equipped cognitive system in order to analyse them. One hypothesis is that the more complex the social organization of a species, the greater the contribution its genome makes towards the phenotypic characters that are used to evaluate genetic identity, thus making a more exact discrimination of kinship possible.

References

- BARASH, D. P., HOLMES, W. G. AND GREENE, P. J. 1978. Exact versus probabilistic coefficients of relationship: some implications for sociobiology. *Am. Nat.*, **112**, 355-363.
- BURT, A. AND BELL, G. 1987. Mammalian chiasma frequencies as a test of two theories of recombination. *Nature*, **326**, 803-805.
- DARLINGTON, C. D. 1939. *The Evolution of Genetic Systems*. Cambridge University Press, Cambridge.
- FISHER, R. A. 1965. *The Theory of Inbreeding*. Oliver and Boyd Ltd, London.
- FRANKLIN, I. R. 1977. The distribution of the proportion of the genome which is homozygous by descent in inbred individuals. *Theor. Pop. Biol.*, **11**, 60-80.
- RISCH, N. AND LANGE, K. 1979. Application of a recombination model in calculating the variance of sib pair genetic identity. *Ann. Hum. Genet.*, **43**, 177-186.
- STAM, P. 1980. The distribution of the fraction of the genome identical by descent in finite random mating populations. *Genet. Res. Camb.*, **35**, 131-155.
- SUAREZ, B. K., REICH, T. AND FISHMAN, P. M. 1979. Variability in sib pair genetic identity. *Hum. Hered.*, **29**, 37-41.