Kinship and Population Subdivision

Henry Harpending
University of Utah

The coefficient of kinship between two diploid organisms describes their overall genetic similarity to each other relative to some base population. For example, kinship between parent and offspring of *1/4* describes gene sharing in excess of random sharing in a random mating population. In a subdivided population the statistic F_{st} describes gene sharing within subdivisions in the same way. Since F_{st} among human populations on a world scale is reliably 10 to 15%, kinship between two individuals of the same human population is equivalent to kinship between grandparent and grandchild or between half siblings. The widespread assertion that this is small and insignificant should be reexamined.

KEY WORDS: coefficient of kinship; coefficient of relationship; inclusive fitness.

COEFFICIENT OF KINSHIP

It is easy to understand why parental care evolved in many lineages: parents and offspring share genes so that parental effort devoted to offspring is in fact effort devoted to the parent's own genes. Hamilton (1964) formalized this insight and extended it to arbitrary degrees of relationship. When Hamilton and others described the theory they often spoke in terms of gene identity by descent, thinking for example of the one half of the nuclear genes the in a diploid offspring that are identical to those in the parent. Many authors also spoke of shared genes. Neither of these descriptions is completely accurate. I may share many genes with, say, an onion, but this

gene sharing is not relevant to the evolution of social behavior within humans.

A better way to think of kinship, relationship, and Hamilton's theory is to think of gene sharing in excess of random gene sharing. A parent shares many more than half his genes with an offspring, but in a random mating population half those genes are surely identical because they came from the parent, while gene sharing with the other half of the child's genome is just what is shared with any random member of the population.

While Hamilton wrote his theory in terms of the coefficient of relationship, most population geneticists reason instead with the coefficient of kinship. Once kinship is known, relationship follows immediately from a simple formula (Bulmer, 1994).

Here is the definition of kinship between person x and person y: pick a random gene at a locus from x and let the population frequency of this gene be p. Now pick a gene from the same locus from y. The probability that the gene in y is the same as the gene picked from x, p_y is

$$p_y = F_{xy} + (1 - F_{xy})p.$$

An interpretation of this is that with probability F_{xy} the genes are the same, with probability $1 - F_{xy}$ they are different, in which case the probability of identity is just the population frequency p (Harpending, 1979). Rearrangement gives the definition of the coefficient of kinship:

$$F_{xy} = (p_y - p)/(1 - p) \tag{1}$$

Kinship coefficients in a random mating diploid population are simple and well known. For example, pick a gene from me, then pick another gene from the same locus from me. With probability 1/2 we picked the same gene, while with probability 1/2 we picked the other gene at that locus. Therefore the probability that the second gene is the same as the first is just $1/2 + p/2$, and substitution of this conditional frequency in the formula for kinship shows that my kinship with myself is just 1/2. The same reasoning leads to the well known values of 1/4 with my child, 1/8 with my grandchild, my half-sib, or my nephew, and so on.

It is very important that the coefficient of kinship not be confused with the coefficient of relationship. These are conceptually and numerically different creatures. The coefficient of relationship can be thought of as "fraction of shared genes" between two organisms. This coefficient is familiar to many biologists since W. D. Hamilton developed his famous theory of kin selection in terms of the coefficient of relationship. However

most subsequent development of the theory has been in terms of kinship coefficients.

In a random mating diploid population the relationship between the two co-efficients is simple: the coefficient of relationship is just twice the coefficient of kinship. This simple rule of thumb breaks down as soon as any complications like inbreeding or population structure are introduced. The best general definition of the coefficient of relation R_{xy} between individuals x and y is (Bulmer, 1994)

$$R_{xy} = F_{xy}/F_{xx}.$$

where F_{xy} is the kinship between x and y and F_{xx} is the kinship of x with himself. This has the interesting property that it is not necessarily symmetric: R_{xy} is not in general equal to R_{yx}.

POPULATION SUBDIVISION

Most of the applications of Hamilton's theory in biology have used kinship and relationship derived from genealogical relationships. For example, parental care evolves, we think, because parents and offspring share genes. But gene sharing (in excess of random gene sharing, always) can arise in other situations. In a subdivided population, individuals share genes with other members of the same deme, and these shared genes are fuel for evolution by inclusive fitness effects in exactly the same way that pedigree relationships like that between parent and child are fuel for evolution by inclusive fitness effects.

I derive here the relationship between population subdivision and kinship in a very simple case, but the formulae apply much more generally than this simple derivation implies. At this point I must mention that these derivations apply to large populations. In the case of small groups ("trait groups," as D. S. Wilson calls them) I would have to consider that if we pick a gene from an individual, the frequency of that gene in the rest of the deme gene pool is slightly reduced. An exact treatment of small demes leads to annoying algebraic terms of order $1/n$ where n is the deme size. I am concerned with large groups and I ignore these terms.

Consider a population made up of two demes of exactly the same size and a genetic locus with exactly two alleles. The conclusion of the algebra below is that the familiar statistic that describes population subdivision, F_{ST}, is precisely kinship between members of the same deme. In other words genetic differences between demes imply genetic similarity within demes, and F_{ST} is just the coefficient of kinship between members of the same deme

POPULATION AND ENVIRONMENT

due to the population structure. For example F_{ST} among human populations is about 1/8, and this is just the coefficient of kinship in a single population between grandparent and grandchild, uncle and nephew, or half-sibs. In a diverse world, members of the same population are related to each other to the same degree that grandparents and grandchildren are related to each other in a single population.

There are two demes of equal size labelled A and B. At a locus the frequency of a gene is p_A in deme A and p_B in deme B. The frequencies in the two demes of the alternate allele are q_A and q_B. The overall mean frequencies are simply p and q. It is convenient to use a slightly different notation to describe the gene frequencies:

$$p_A = p + \delta$$
$$p_B = p - \delta$$

so of course

$$q_A = q - \delta$$
$$q_B = q + \delta$$

Now imagine that we pick a gene at random from the population, then pick another gene from the same locus from the same deme. What is the coefficient of kinship within demes? In order to find this we use the formula (1) above.

With probability 1/2 we pick someone from population A initially, and with probability p_A we pick the allele whose frequency is p_A. With probability $q_A = 1 - p_A$ we pick the alternate allele. Putting these possibilities into equation (1) we have

$$F = (1/2)p_A(p_A - p)/q + (1/2)p_B(p_B - p)/q + (1/2)q_A(q_A - q)/$$
$$p + (1/2)q_B(q_B - q)/p$$

Using the substitutions above, this becomes

$$F = \{(p + \delta)(\delta) + (p - \delta)(-\delta)\}/2q + \{q - \delta)(-\delta) + (q + \delta)(\delta)\}/2p$$
$$= 2\delta^2/2q + 2\delta^2/2p$$

and since $p + q = 1$

$$F = 4\delta^2/4pq$$

$$= \delta^2/pq$$

This is simply the F_{ST} genetic distance between the two populations—the variance of the gene frequency divided by the mean gene frequency multiplied by its complement. When F_{ST} is reported for a collection of populations, it is essentially an average of all the pairwise population F_{ST} statistics. The statistic is computed for each allele at each locus, then averaged over all loci.

Many studies agree that F_{ST} in world samples of human populations is between ten and fifteen percent. If small long-isolated populations are included, the figure is usually somewhat higher. A conservative general figure for our species is $F_{ST} \approx 0.125 = 1/8$. This number was given by Cavalli-Sforza in 1966, and a widely cited paper by Lewontin (1972) argued at length that this is a small number implying that human population differences are trivial. An alternative perspective is that kinship between grandparent and grandchild, equivalent to kinship within human populations, is not so trivial. For further discussion see Klein and Takahata (2002, pp. 387–390).

Kinship in a Subdivided Population

Equation 1 and its derivation shows that if we pick a gene at random from a population of two demes and find that that its overall frequency is p, then the frequency of that gene in the same deme is on average

$$p_{same} = p + (1 - p)F_{ST}$$

while the frequency of that gene in the other deme is on average

$$p_{other} = p - (1 - p)F_{ST}.$$

Using equation 1 and these relations we can derive kinship and relationship coefficients within and between demes easily.

An individual's coefficient of kinship with someone from his own deme is just F_{ST} while his kinship with someone from the other deme is $-F_{ST}$. What about kinship with oneself in a subdivided population? Pick a gene from an individual, then pick another at random from the same individual: with probability 1/2 we picked the same gene and with probability 1/2 we picked the other one, in which case the probability it is the same is

POPULATION AND ENVIRONMENT

$p + (1 − p)F_{ST}$. Therefore

$$p_{self} = 1/2(1 + p + (1 − p)F_{ST})$$

Using equation 1, we find that

$$F_{self} = 1/2(1 + F_{ST})$$

rather than the simple 1/2 kinship with self in a single random mating population. It is simple to derive familiar family kinship coefficients in the same way: for example kinship with a child when the other parent is from the same deme is

$$F_{child} = 1/4 + 3F_{ST}/4$$

and so on. In general, if the kinship in a random mating population with a relative is 1/x, then in a subdivided population the kinship with that same relative is

$$F_{relative\ of\ degree\ x} = 1/x + (1 − x)F_{ST}/x \qquad (2)$$

What about kinship with a relative who is a hybrid between the populations? Consider, for example, a child whose other parent is from the other deme. Pick a gene from the parent: the probability of picking the same gene from the child is 1/4, the probability of picking a gene from the child not identical to the first but from the same deme as the parent is 1/4, and the probability of picking a gene from the other deme is 1/2. Putting these together, the probability of the picking the same gene is

$$p_{hybrid\ offspring} = 1/4 + 1/4(p + (1 − p)F_{ST}) + 1/2(p − (1 − p)F_{ST}).$$

Using equation 1, this becomes

$$F_{hybrid\ offspring} = 1/4 − F_{ST}/4.$$

In general the same derivations shows that kinship with a hybrid relative of degree x, meaning a relative with whom kinship in a random mating population would be x, is

$$F_{hybrid\ relative\ of\ degree\ x} = 1/x − F_{ST}/x. \qquad (3)$$

The difference between equations 2 and 3 is just F_{ST}, the difference between kinship with an intra-demic relative and a hybrid relative. Notice also that as x becomes large, equation 2 shows that kinship with a random member of the same deme is F_{ST} and kinship with an otherwise unrelated hybrid offspring is 0.

REFERENCES

Bulmer, M. (1994). *Theoretical Evolutionary Ecology*. Sunderland, Massachusetts: Sinauer.

Cavalli-Sforza, L. L. (1966). Population structure and human evolution. *Proceedings of the Royal Society Series B, 164*, 362–379.

Hamilton, W. D. (1964). The genetic evolution of social behavior, parts 1 and 2. *Journal of Theoretical Biology, 7*, 1–51.

Harpending, H. (1979). The population genetics of interactions. *American Naturalist, 113*, 622–630.

Klein, J., & Takahata, N. (2002). *Where Do We Come From: The Molecular Evidence For Human Descent*. Berlin: Springer.

Lewontin, R. C. (1972). The apportionment of human diversity. *Evolutionary Biology, 6*, 381–398.