Wakeley, J., and Wilton, P.R. (2016) Coalescent and Models of Identity by Descent. In: Kliman, R.M. (ed.), Encyclopedia of Evolutionary Biology. vol. 1, pp. 287–292. Oxford: Academic Press.

Coalescent and Models of Identity by Descent

J Wakeley and PR Wilton, Harvard University, Cambridge, MA, USA

Glossary

Coalescence The occurrence, looking backward in time, of common ancestors of two or more alleles.

Effective population size A standardized measure of population size based on the rate of change of genetic variation as a result of reproduction but not selection.

Gamete Haploid germ cell, for example, sperm or egg.

Gene conversion A recombinational process during meiosis in which a tract of sequence is replaced by corresponding sequence from a different chromatid.

Haplotype Any haploid set of multi-locus genetic variations; more specifically, here, any particular segment of DNA sequence in a genome.

Markov process A statistical model of successive changes in the state of a system in which transition rates or probabilities depend only on the current state.

Monecious Having a single mating type; hermaphroditic.

Path coefficients Weights of factors in linear statistical models, used to assess causation.

Phenotype The physical characteristics of an individual organism.

Poisson process A random process that counts numbers of events, usually as they occur in time, in which the times between events are exponentially distributed.

Identity by descent (IBD) is a property of genetic material in related individuals. Specifically, two alleles – i.e., two pieces of DNA from the same genetic locus, and therefore segregating at meiosis in a diploid organism – are identical by descent if one of them is descended from the other or if both are descended from a third allele which existed at some time in the past. Descent, here, means DNA replication and the transmission of genetic material from parents to offspring. The voluminous literature on IBD diverges as to whether mutations are allowed in descent between IBD alleles, though it is probably more commonly allowed than not. In addition, although it usually denotes pairwise identity, IBD can be extended to more than two alleles. For example, the common ancestral, third allele mentioned above is identical by descent with both of its descendant alleles.

IBD describes an important type of sameness which arises in biological systems. Cotterman (1940) took pains to define IBD clearly along with two other kinds of genetic identity: allelic identity, meaning from the same locus, as above, and functional identity, meaning alleles that are interchangeable without consequence for function or phenotype. Crow (1954) coined the term identical by descent (Cotterman had used 'derivative') and contrasted it with a specific kind of functional identity, namely identity in state, which means having the same DNA sequence. When mutation is understood to preclude IBD, then IBD may be viewed as a special case of identity in state. Similarly, it should be noted that 'alleles' and 'allelic' are used confusingly in the literature, to refer either to sequences from the same locus only, as here, or to functionally different sequences at a locus.

The concept of IBD is general, but it is specifically used to indicate close genetic relationship, in excess of background levels in a population. Because any two individuals, even from different species, may be said to be related if a sufficiently long time frame is considered, detailed definitions of IBD depend on what is chosen to measure close relationship. Notions of IBD have been developed for (1) individuals related by a known family structure or pedigree; (2) alleles descended from a common ancestral allele within a specified time and/or without any mutations between them; and (3) genomic tracts of strong genetic similarity demarcated by recombination events. Here, these are referred to, respectively, as pedigree definitions of IBD, coalescent definitions of IBD, and ancestral-segment definitions of IBD.

The ultimate utility of any population-genetic concept, including IBD, is in the interpretation of genetic variation. Theoretical treatments of IBD have guided analyses of genetic and genomic data, and provided avenues for inference by making connections between patterns of variation and key population-genetic parameters or sources of shared common ancestry.

Pedigree Identity by Descent

Studies of what later became known as IBD began early in population genetics, with efforts to allow some non-independence of alleles in the context of the Hardy–Weinberg Law (Hardy, 1908; Weinberg, 1908). The Hardy–Weinberg Law by itself leaves little room for relatedness. In it each individual receives two alleles independently at random from an essentially infinite population. Alleles are either identical (in state) or they are different. Only if a particular relationship or set of relationships is specified and embedded within the population does it become possible to consider IBD. We may think, for example, of a single parent and its offspring, which share exactly one pair of alleles that are identical by descent, while the other two alleles possessed by these two individuals would represent independent random samples from the population.

Early notions of IBD trace back to Wright's work on inbreeding coefficients, which may be interpreted in terms of IBD. Wright applied his general method of path coefficients for decomposing correlations to compute inbreeding coefficients

given a pedigree (Wright, 1921a,b, 1922). Specifically, for a pair of individuals (a possible mating pair) who are connected to common ancestors through K different paths, or 'loops' (Cotterman, 1940) in the pedigree, the inbreeding coefficient of an offspring of the pair is

$$f_o = \sum_{i=1}^{K} \left(\frac{1}{2}\right)^{n_i + n'_i + 1} \qquad [1]$$

in which n_i and n'_i are the numbers of generations on path i from each individual back to a common ancestor. Equation [1] neglects the possibility that common ancestors might themselves be inbred, but this is easily remedied if f_a is known for each ancestor (Wright, 1922).

As an illustration, it is well known that Charles Darwin married his first cousin, Emma Wedgewood. Together they had eight children, plus an additional two who did not survive infancy. The relevant family tree, genealogy, or pedigree, is depicted in Figure 1. Two paths must be considered in computing f_o: one in which the two alleles both came from Josiah Wedgewood and one in which they came from Sarah Wedgewood. For each path connecting Charles and Emma Darwin to their grandparents, $n_i = n'_i = 2$. Applying [1] gives

$$f_o = 2\left(\frac{1}{2}\right)^{2+2+1} = \frac{1}{16} = 0.0625 \qquad [2]$$

as the inbreeding coefficient of any child of Charles and Emma Darwin.

Although Wright (1921a,b) derived [1] by considering the decomposition of the correlation of uniting gametes, it can also be interpreted in terms of IBD. Specifically, 1/2 is the probability that a randomly selected allele in an individual came from a particular one of its parents. The $n_i + n'_i$ factors of 1/2 in [1] give the chance that the two alleles which unite to form a gamete came from a given common ancestor. Another factor of 1/2 gives the chance that the two alleles traced back to that common ancestor are derived from the same allele. Finally the sum is taken over all paths.

For Charles and Emma Darwin's children, there are two paths with $n_i = n'_i = 2$. Thus, [2] gives the probability that the two alleles at a locus in one of Charles and Emma Darwin's children are identical by descent. This sort of thinking underlies the general likelihood calculations on pedigrees that have

been of crucial importance in human genetics (Cannings et al., 1978).

Wright (1921b) also used the method of path coefficients, later switching from f to F (Wright, 1931), to establish recursive equations of the type discussed in Section 'Coalescent Identity by Descent,' which relate inbreeding coefficients in the current generation to those in previous generations, and to quantify deviations from Hardy–Weinberg genotype proportions. For two alleles A_1 and A_2 with relative frequencies p and $q = 1 - p$ in the population, Wright described deviations from Hardy–Weinberg proportions using

$$
\begin{aligned}
A_1A_1 : \quad & p^2(1 - F) + pF \\
A_1A_2 : \quad & 2pq(1 - F) \\
A_2A_2 : \quad & q^2(1 - F) + qF
\end{aligned}
\qquad [3]
$$

Then F is the correlation in allelic state of gametes that unite to form a diploid zygote, and ranges from -1 to $+1$. If inbreeding is the source of the correlation, F ranges between 0 and 1 and may be interpreted as the probability of IBD. Equation [3] says that an individual is formed either by sampling two alleles at random (with probability $1 - F$) or by sampling one allele and duplicating it to make a diploid individual (with probability F). Implicit in this interpretation, there is a 'separation of time scales' between a slow population-level process and a fast individual-level or sub-population-level process (Rousset, 2004, p. 57).

The pedigree concept of IBD has proven useful in a number of settings. Malécot (1948) and Kempthorne (1955) employed it to re-derive and extend the calculations of correlations in trait values between relatives which form the basis of quantitative genetics (Fisher, 1918; Falconer and and MacKay, 1996). Another important class of applications uses 'gene dropping' simulations (MacCluer et al., 1986) to account for IBD within a pedigree in generating null distributions of genotypes under selection; for a recent example, see Gao et al. (2015).

Coalescent Identity by Descent

Wright (1931, p. 107) referred to F as the 'correlation between uniting egg and sperm' and the 'total proportional change of heterozygosis.' He also used path coefficients to study the rate of increase of F over time under various mating schemes,
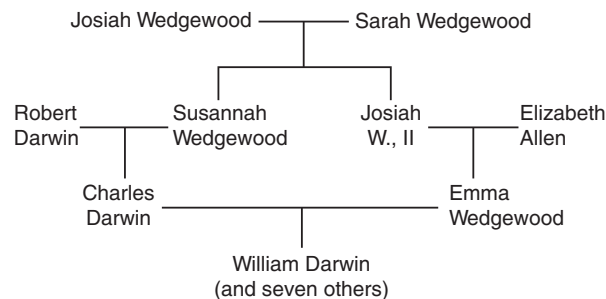
Figure 1   Part of the genealogy of Charles Darwin and Emma Darwin (*née* Wedgewood), who were first cousins. Their common grandparents, Josiah Wedgewood and Sarah Wedgewood, were also third cousins. Using a more complete pedigree, Berra *et al.* (2010) estimate that Charles and Emma Darwin's children had an inbreeding coefficient of 0.0630, which is only a little larger than the 0.0625 computed here, based only on the relationships above.

including what is now well known as the Wright–Fisher model of random mating in a finite population (Fisher, 1930; Wright, 1931). In the simple case of a diploid, monecious organism, over one generation

$$F_g = \frac{1}{2N} + \left(1 - \frac{1}{2N}\right) F_{g-1} \qquad [4]$$

in which g refers to the current generation and $g-1$ to the previous generation. Equation [4] can be rearranged and applied iteratively to obtain

$$1 - F_g = (1 - F_0)\left(1 - \frac{1}{2N}\right)^g \qquad [5]$$

which, with reference to the second line of [3], shows that heterozygosity is lost at rate $1/2N$ in a population of constant size N diploid individuals. More complicated populations behave similarly if they are large, and are described in relation to this monecious case using the concept of effective population size (Wright, 1931).

Malécot (1941, 1946, 1948) called F the 'average coefficient of consanguinity' and clearly understood it as the probability of IBD for two alleles in an individual. He derived recursive equations like [4] and [5] by explicitly considering the occurrence of common ancestors. In the simple example given by [4], which includes the possibility that the individual is produced by self-fertilization, $1/2N$ is the product of the probability that two alleles came from the same parent ($1/N$) and the probability they are descended from the same allele in that parent ($1/2$). When the parents are distinct, the two alleles in an individual descend from two alleles in different individuals in the previous generation, and here Malécot used 'average coefficient of kinship' to refer to the probability of IBD. Because he introduced the notion of tracing allelic lineages back to common ancestors, Malécot is credited with fundamental idea behind the gene-genealogical or coalescent approach to population genetics (Kingman, 1982; Hudson, 1983; Tajima, 1983).

Coalescent theory is reviewed in Hein *et al.* (2005) and Wakeley (2008). In the context of (pairwise) IBD, it is enough to consider the number of generations, G, back to the common ancestor between two alleles. Under the Wright–Fisher model above, G is geometrically distributed:

$$P(G=g) = \frac{1}{2N}\left(1 - \frac{1}{2N}\right)^{g-1}, \quad g = 1, 2, \ldots \qquad [6]$$

If, as typical in coalescent theory, time is rescaled so that it is measured in units of $2N$ generations and N is taken to be very large, technically taking the limit $N \to \infty$ and using $T = G/2N$, then the pairwise coalescence time, T, is exponentially distributed:

$$f_T(t) = e^{-t}, \quad t > 0 \qquad [7]$$

Thus the expected value of the time back to the most recent common ancestor of a pair of alleles is equal to one on the coalescent time scale, or $2N$ when measured in generations.

Two notions of close relationship have been used in gene genealogical or coalescent approaches to population genetics. The first defines IBD relative to an arbitrarily chosen past population, as for example: "The probability of identity by descent is defined as the chance that two genes are descended from the same gene in some ancestral population" (Whitlock and Barton, 1997). Fixing a given generation g in the past and using [6], the probability of IBD would be

$$P(G \le g) = \sum_{i=1}^{g} P(G=g) = 1 - \left(1 - \frac{1}{2N}\right)^g \qquad [8]$$

and for the corresponding coalescence time $t = g/2N$, the probability of IBD would be

$$P(T < t) = \int_0^t f_T(x)dx = 1 - e^{-t} \qquad [9]$$

When g and t are small, these probabilities are both small because it is unlikely for two alleles to be descended from a very recent common ancestor, while both probabilities approach 1 as g and t approach infinity. The occurrence of IBD under this definition is illustrated in Figure 2(a).

Under this time-based notion of IBD, as under the pedigree definition of IBD in Section 'Pedigree Identity by Descent,' IBD has been defined alternately to require or not require identity in state. For example, Crow (1954, p. 544) considered that two alleles are identical by descent if both are 'derived from a single gene in some common ancestor' but added that they 'may be unlike in state if there has been a mutation since their common origin.' Similarly, Cotterman (1940, p. 171) considered that all eight combinations of his three kinds of sameness/difference are possible. In contrast, studying
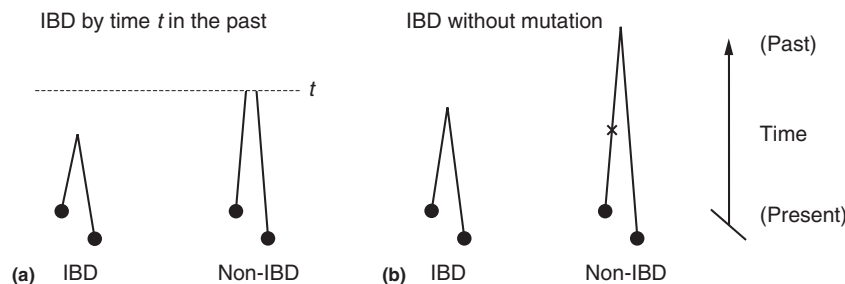

Figure 2   Two different definitions of IBD in gene-genealogical or coalescent models: (a) two alleles are identical by descent if they are descended from a common ancestral allele by time t (or $g=2Nt$ generations) in the past, and (b) two alleles are identical by descent if they are descended from a common ancestral allele without mutation (x).

F in the face of mutation with probability u per allele per generation (so that $(1-u)^2$ gives the probability that neither allele underwent mutation in the previous generation) and writing

$$F_g = \frac{(1-u)^2}{2N} + (1-u)^2\left(1 - \frac{1}{2N}\right)F_{g-1} \qquad [10]$$

Malécot (1946) implicitly assumed that IBD precludes mutation. Nagylaki (1989), interpreting Malécot, states: "Two homologous genes are identical by descent if and only if they are derived from the same gene or one is derived from the other (in both cases without mutation)."

The second coalescent definition of IBD follows Malécot's logic, and can be restated under the backward-time view of coalescent theory (Ewens, 1990) as follows. The probability of IBD is equal to the probability that coalescence, rather than mutation, is the first event encountered when the ancestry of two alleles is traced back into the past. Equations [6] and [7] offer two ways of computing the probability that no mutation occurs on either allele's lineage back to their common ancestor:

$$\sum_{g=1}^{\infty} P(G=g)(1-u)^{2g} = \frac{(1-u)^2}{(2N-1)(2-u)u+1} \qquad [11]$$

$$\int_0^{\infty} f_T(t)e^{-\theta t}dt = \frac{1}{\theta+1} \qquad [12]$$

in which $\theta = 4Nu$ is the usual population mutation rate parameter in coalescent theory. A graphical illustration of this definition of IBD is given in Figure 2(b). Equation [11] is also the solution to [10] for $F_{g-1} = F_g = F$. Malécot (1946) obtained [12] as an approximation to this solution, and noted its accuracy for most biologically reasonable values of N and u, which is to say when N is large and u is small.

The coalescent definitions of IBD described in this section apply to allelic variation at a single locus without recombination. Important extensions include considerations of subdivided populations (Wright, 1951; Slatkin, 1991; Rousset, 2004) and patterns of identity and difference in samples of more than two alleles (Ewens, 1972; Thompson, 2013).

Ancestral-Segment Identity by Descent

With the advent of DNA sequencing, and now whole-genome sequencing and genotyping, single-locus concepts of IBD have given way to a genomic perspective which has yielded new insights, particularly in the field of human population genetics (Thompson, 2013). In this context, it is the joint action of coalescence and recombination that determines close relationship, and IBD refers to segments of genomes which descend from recent common ancestors.

To illustrate, consider a focal site which is identical by descent under the pedigree definition of IBD. If there were no recombination, every site would have the same ancestry as the focal site and the entire chromosome would be identical by descent. Recombination decouples the ancestries of different sites and allows IBD to vary along the chromosome because
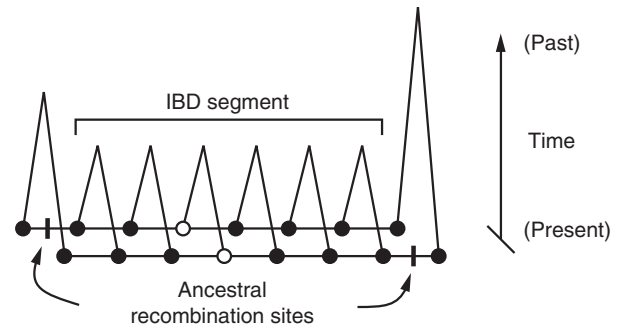

Figure 3 An IBD segment defined by two sites of recombination in the ancestry of two chromosomes. Every site in the IBD segment has the same most recent common ancestor and hence the same time until most recent common ancestry. The focal site is denoted by the open circles.

each recombination event chops the chromosome into maternal and paternal pieces which then may have different ancestries. The result is that IBD will occur in segments, as depicted in Figure 3, such that some number of sites on either side of the focal site will share the same ancestry and be identical by descent.

Under this ancestral-segment view of IBD, two homologous segments in an alignment of two chromosomes are identical by descent if they share a most recent common ancestor, and hence the same coalescence time, at every site. Here these are referred to as IBD segments, but they could also be called IBD tracts or IBD blocks. Note that this definition of IBD implies that every pair of chromosomes is IBD everywhere, because at every site there is a most recent common ancestor which is the same for some number of adjacent sites. As IBD is meant to indicate close relationship, interest is focused on recently co-inherited IBD segments. Younger IBD segments tend to be longer due to the limited opportunity for recombination in their ancestries. Thus, ancestral-segment IBD is typically defined in terms of a minimum length cutoff for segments.

Detection of IBD segments involves finding long haplotypes that are identical in state, although some mismatches are allowed due to the presence of genotyping or sequencing error, mutations that have occurred since the most recent common ancestor defining the IBD segment, and even short gene conversion events that effectively incorporate mutations from other haplotypes. Some IBD detection methods use dense genotype data (Purcell et al., 2007; Gusev et al., 2009; Browning and Browning, 2011), allowing reliable detection of IBD segments as short as two centiMorgans (cM). Recently, other methods have been developed to take advantage of sequence data, enabling reliable detection of IBD segments as short as 0.2 cM (Browning and Browning, 2013; Tataru et al., 2014). Tataru et al. (2014) present comparisons of current IBD detection methods.

Fisher (1949, Chapter 3) made the first theoretical examination describing the lengths and counts of IBD segments (bordered by what he termed 'junctions') in regular mating systems, for example, full-sib or parent-offspring mating. Stam (1980) considered the same quantities in the context of random-mating populations, and Chapman and Thompson (2003) studied IBD tracts in subdivided

populations. These studies attempt to simultaneously model many IBD segments across the genome, which is a very difficult problem.

More recently, coalescent theory has been employed to model IBD-segment distributions. The foundation for this work is the coalescent with recombination, which describes the genetic ancestry of recombining chromosomes as an 'ancestral recombination graph' (Griffiths and Marjoram, 1997), allowing different ancestries at different loci. Wiuf and Hein (1999) provided a formulation of this model that proceeded along the chromosomes rather than backward in time, laying the groundwork for practical applications of coalescent theory to IBD segments. McVean and Cardin (2005) proposed a simplified Markov process to approximate this model, immensely improving its computational efficiency, with a subsequent improvement by Marjoram and Wall (2006). These latter models are referred to as sequentially Markov coalescent (SMC) models.

Palamara et al. (2012) derived the IBD-segment length distribution as follows. Consider a single focal site in two aligned chromosomes, and assume that the two copies at that site last shared a common ancestor g generations ago. If every ancestral recombination event defines a new IBD segment, which is equivalent to assuming the SMC model of McVean and Cardin (2005), the length of an IBD segment can be modeled by considering the nearest ancestral recombination events on either side of the focal site. This SMC assumption, that a recombination event invariably terminates an IBD segment, is reconsidered below.

There are $2g$ meioses separating the two chromosomes since the most recent shared ancestor at the focal site. If recombination occurs without interference and the distance between recombination sites is measured in Morgans, then in each meiosis recombination can be modeled as a Poisson process along the chromosome with mean equal to 1. Recombination across all $2g$ meioses is then a Poisson process with mean $2g$, and the length L of the IBD segment surrounding the focal site can be described as the sum of two independent exponential random variables with rate $2g$, each representing the distance along the chromosome in one direction away from the focal site.

Given g, the length L is gamma distributed (Palamara et al., 2012):

$$f_{L|G}(l|g) = 4g^2 e^{-2gl} \qquad [13]$$

However, in most contexts the age g is not an observable quantity, so it is integrated out to give the overall IBD-segment length distribution. Using an exponential distribution with rate $1/2N$ to approximate the geometric distribution in [6],

$$f_L(l) = \int_0^\infty f_G(g) f_{L|G}(l|g) dg$$
$$= \int_0^\infty \frac{1}{2N} e^{-g/2N} (2g)^2 l e^{-2gl} dg = \frac{32N^2 l}{(1+4Nl)^3} \qquad [14]$$

which is a power-law distribution, with infinite mean and variance.

Equation [14] gives the distribution of L when IBD segments are sampled by selecting a position (the focal site)

uniformly at random over a long chromosome. Each IBD segment is effectively weighted by its length. It is perhaps more intuitive and more useful to consider the distribution of IBD segments that arises from sampling entire segments uniformly at random rather than weighting them by their lengths. The difference between the two distributions is often referred to as the inspection paradox. If the length of a uniformly sampled IBD segment is denoted S, the density function of S can be derived by weighting [14] by $1/l$ and normalizing, which gives

$$f_S(s) = \frac{\frac{1}{s} \frac{32N^2 s}{(1+4Ns)^3}}{\int_0^\infty \frac{1}{l} \frac{32N^2 l}{(1+4Nl)^3} dl} = \frac{8N}{(1+4Ns)^3} \qquad [15]$$

This is also a power-law distribution, with mean $1/4N$ and infinite variance. Note that deriving [15] without first deriving [14] would require knowing the distribution of coalescence times at IBD-segment edges. The derivation proceeds as above because the distribution of coalescent times at a fixed site arises more naturally from coalescent theory.

The density [15] suggests that in a diploid population of size $N = 10\,000$, the fraction of all IBD segments that are longer than 0.2 cM is $\sim 1.5 \times 10^{-4}$, while the fraction longer than 2 cM is $\sim 1.5 \times 10^{-6}$. Viewed from the other perspective, the density [14] indicates that the mean fraction of the genome contained in IBD segments larger than 0.2 cM is ~ 0.025, while the mean fraction contained in IBD segments larger than 2 cM is 0.0025.

The shapes of these distributions change under variable population sizes or when migration occurs between subpopulations, allowing data on IBD to be used for demographic inference about the recent past. Palamara et al. (2012) used this to infer recent population size changes in the Ashekenzi Jewish and Maasai populations, later generalizing to include migration between two subpopulations (Palamara and Pe'er, 2013). Independently, Ralph and Coop (2013) used IBD sharing to characterize the recent geographic and demographic structuring of Europe.

Common to all of these studies is the assumption that recombination events always terminate IBD segments. In fact one-third of all recombination events do not create a new IBD segment (see Theorem 2.4 in Griffiths and Marjoram, 1997) because they are 'healed' by coalescent events in which the two ancestral lineages separated by recombination coalesce back together. These events were ignored in the original formulation of the SMC, but were included in the subsequent SMC' model (Marjoram and Wall, 2006). IBD-segment length distributions can be derived under the SMC' in the manner above, and predictions for overall IBD length distributions under the SMC' match simulations under the full ancestral recombination graph very well, which is not true of the SMC (Carmi et al., 2014). The resulting equations are cumbersome and involve special mathematical functions. Furthermore, for recent IBD segments there is little opportunity for back coalescence, and the differences between the SMC and SMC' predictions are negligible. However, for older, or shorter, IBD segments (smaller than ~ 0.5–1 cM) and for populations that have a small recent effective population size, it is important to base calculations on the SMC'.

With this minor caveat, it is clear that IBD segment-based techniques offer novel and practical ways of using genetic data to learn about the very recent past. IBD approaches to inference depend heavily on accurate detection of IBD segments, emphasizing the importance of recent improvements in IBD detection (Browning and Browning, 2013; Tataru *et al.*, 2014). Considering these developments, along with the recent theoretical advances outlined above, there is promise that IBD-segment methods will continue to reveal new insights into recent demographic processes in ways that complement other traditional approaches of population genetics.

> *See also*: Effective Population Size. Genetic Drift, Models of Random

References

Berra, T.M., Alvarez, G., Ceballos, F.C., 2010. Was the Darwin/Wedgwood dynasty adversely affected by consanguinity? BioScience 60, 376–383.

Browning, B.L., Browning, S.R., 2011. A fast, powerful method for detecting identity by descent. American Journal of Human Genetics 88, 173–182.

Browning, B.L., Browning, S.R., 2013. Detecting identity by descent and estimating genotype error rates in sequence data. American Journal of Human Genetics. 93, 840–851.

Cannings, C., Thompson, E.A., Skolnick, M.H., 1978. Probability functions on complex pedigrees. Advances in Applied Probability 10, 26–61.

Carmi, S., Wilton, P.R., Wakeley, J., Pe'er, I., 2014. A renewal theory approach to IBD sharing. Theoretical Population Biology 97, 35–48.

Chapman, N.H., Thompson, E.A., 2003. A model for the length of tracts of identity by descent in finite random mating populations. Theoretical Population Biology 64, 141–150.

Cotterman, C.W., 1940. A calculus for statistico-genetics. PhD thesis, Ohio State University. Also available in Ballonoff, P. (Ed.), 1974. Genetics and Social Structure. Stroudsburg, PA: Dowden, Hutchinson & Ross, pp. 157–272.

Crow, J.F., 1954. Breeding structure of populations. II. Effective population number. In: Kempthorne, O., Brancroft, T.A., Gowen, J.W., Lush, J.L. (Eds.), Statistics and Mathematics in Biology. Ames, IA: Iowa State University Press, pp. 543–556.

Ewens, W.J., 1972. The sampling theory of selectively neutral alleles. Theoretical Population Biology 3, 87–112.

Ewens, W.J., 1990. Population genetics theory – The past and the future. In: Lessard, S. (Ed.), Mathematical and Statistical Developments of Evolutionary Theory. Amsterdam: Kluwer Academic Publishers, pp. 177–227.

Falconer, D.S., MacKay, T.F.C., 1996. Introduction to Quantitative Genetics, fourth edn Essex: Prentice Hall.

Fisher, R.A., 1918. The correlation between relatives on the supposition of Mendelian inheritance. Transactions of the Royal Society of Edinburgh 52, 399–433.

Fisher, R.A., 1930. The Genetical Theory of Natural Selection. Oxford: Clarendon.

Fisher, R.A., 1949. The Theory of Inbreeding. Edinburgh: Oliver and Boyd.

Gao, Z., Waggoner, D., Stephens, M., Ober, C., Przeworski, M., 2015. An estimate of the average number of recessive lethal mutations carried by humans. Genetics 199, 1243–1254.

Griffiths, R.C., Marjoram, P., 1997. An ancestral recombination graph. In: Donnelly, P., Tavare, S. (Eds.), *Progress in Population Genetics and Human Evolution*, Volume 87 *of IMA Volumes in Mathematics and Its Application*. New York, NY: Springer-Verlag, pp. 257–270.

Gusev, A., Lowe, J.K., Stoffel, M., *et al.*, 2009. Whole population, genome-wide mapping of hidden relatedness. Genome Research 19, 318–326.

Hardy, G.H., 1908. Mendelian proportions in a mixed population. Science 18, 49–50.

Hein, J., Schierup, M.H., Wiuf, C., 2005. Gene Genealogies, Variation and Evolution: A Primer in Coalescent Theory. Oxford: Oxford University Press.

Hudson, R.R., 1983. Testing the constant-rate neutral allele model with protein sequence data. Evolution 37, 203–217.

Kempthorne, O., 1955. The theoretical values of correlations between relatives in random mating populations. Genetics 40, 153–167.

Kingman, J.F.C., 1982. On the genealogy of large populations. Journal of Applied Probability 19A, 27–43.

MacCluer, J.W., VandeBerg, J.L., Read, B., Ryder, O.A., 1986. Pedigree analysis by computer simulation. Zoo Biology 5, 147–160.

Malécot, G., 1941. Etude mathematique des populations Mendelienne. Annales de l'Université de Lyon. Sciences A 4, 45–601.

Malécot, G., 1946. La consaguinite dans une population limitee. Comptes Rendus de l'Académie des Sciences, Paris 222, 841–843.

Malécot, G., 1948. Les Mathematiques de l'Heredite. Paris: Masson. 1969. Extended translation: The Mathematics of Heredity. San Francisco, CA: W.H. Freeman.

Marjoram, P., Wall, J.D., 2006. Fast "coalescent" simulation. BMC Genetics 7, 16.

McVean, G.A.T., Cardin, N.J., 2005. Approximating the coalescent with recombination. Philosophical Transactions of the Royal Society B 360, 1387–1393.

Nagylaki, T., 1989. Gustave Malécot and the transition from classical to modern population genetics. Genetics 122, 253–268.

Palamara, P.F., Lencz, T., Darvasi, A., Pe'er, I., 2012. Length distributions of identity by descent reveal fine-scale demographic history. American Journal of Human Genetics 91, 809–822.

Palamara, P.F., Pe'er, I., 2013. Inference of historical migration rates via haplotype sharing. Bioinformatics 29, i180–i188.

Purcell, S., Neale, B., Todd-Brown, K., *et al.*, 2007. PLINK: A tool set for whole-genome association and population-based linkage analyses. American Journal of Human Genetics 81, 559–575.

Ralph, P., Coop, G., 2013. The geography of recent genetic ancestry across Europe. PLOS Biology 11, e1001555.

Rousset, F., 2004. Genetic Structure and Selection in Subdivided Populations. Princeton, NJ: Princeton University Press.

Slatkin, M., 1991. Inbreeding coefficients and coalescence times. Genetics Research, Cambridge 58, 167–175.

Stam, P., 1980. The distribution of the fraction of the genome identical by descent in finite random mating populations. Genetics Research 35, 131–155.

Tajima, F., 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105, 437–460.

Tataru, P., Nirody, J.A., Song, Y.S., 2014. diCal-IBD: Demography-aware inference of identity-by-descent tracts in unrelated individuals. Bioinformatics 30, 3430–3431.

Thompson, E.A., 2013. Identity by descent: Variation in meiosis, across genomes, and in populations. Genetics 194, 301–326.

Wakeley, J., 2008. Coalescent Theory: An Introduction. Greenwood Village, CO: Roberts & Company Publishers.

Weinberg, W., 1908. Uber den Nachweis der Vererbung beim Menschen. Jaresh. Verein. f. vaterl. Naturk. Wiirtem. 64, 368–382; 1963. Human Genetics (S.H. Boyer (Ed.) Trans.). Englewood Cliffs, NJ: Prentice Hall, pp. 4–15; 1977. Evolutionary Genetics (D.L. Jameson (Ed.) Trans.). Stroudsburg, PA: Dowden, Hutchinson, and Ross, pp. 115–125.

Whitlock, M.C., Barton, N.H., 1997. The effective size of a subdivided population. Genetics 146, 427–441.

Wiuf, C., Hein, J., 1999. Recombination as a point process along sequences. Theoretical Population Biology 55, 248–259.

Wright, S., 1921a. Correlation and causation. Journal of Agricultural Research 20, 557–585.

Wright, S., 1921b. Systems of mating. Genetics 6, 111–178.

Wright, S., 1922. Coefficients of inbreeding and relationship. American Naturalist 56, 330–338.

Wright, S., 1931. Evolution in Mendelian populations. Genetics 16, 97–159.

Wright, S., 1951. The genetical structure of populations. Annals of Eugenics 15, 323–354.