Annu. Rev. Anthropol. 1991. 20:97–117

RECENT DEVELOPMENTS IN POPULATION GENETICS

Partha P. Majumder

Anthropometry and Human Genetics Unit, Indian Statistical Institute, Calcutta 700035, India

KEY WORDS: genetic population structure, multigene family, phylogenetic tree, genetic polymorphism, genetic relationship

INTRODUCTION

In a Workshop sponsored by the United States National Institutes of Health in 1983 it was pointed out (51) that "Since 1974, theoretical population genetics has treated increasingly complex models, but sometimes with decreasing biological relevance." With the increasing emphasis and interest in molecular studies, the character of population genetics has changed considerably. "The subject has changed from one that is rich in theory and poor in data to one that is almost the opposite" (13). There has also been a shift from the bitter debates over the selective/neutral dichotomy. With the increasing availability of DNA sequence data, there is unequivocal evidence in favor of the neutral theory. The observed rates and patterns of DNA base substitutions are generally in concordance with the predictions of the neutral theory (35, 36, 42). This, of course, does not imply that all mutations are neutral; evidence for positive Darwinian selection at the molecular level has also recently been found (92). The theoretical and empirical studies of the last two decades have made it abundantly clear that the forces of mutation and random genetic drift are much more important in the creation and maintenance of genetic variability in populations than had previously been envisaged. Many important theoretical population genetic results, incorporating mutation and drift, on ages of alleles, sampling distributions of functions of allele frequencies, statistics for

0084-6570/91/1015-0097$02.00

testing the neutral theory, etc have been derived in the last few years. The recent emphasis in theoretical population genetics has been to study the dynamics of variation of DNA sequences in populations (see e.g. 28, 74, 87, 88, and references therein). Most of these theoretical studies make considerable use of stochastic processes and diffusion approximations. I do not attempt to review these theoretical studies here; interested readers may consult the excellent review by Ewens (20). Here I review studies that are generally empirical in nature. Special emphasis is given to new concepts introduced and to the problems studied rather than to a complete description of results obtained.

THE HARDY-WEINBERG LAW

While much effort is rightly being spent on the stochastic modeling of population genetic phenomena, we are still learning things about the most basic deterministic law of population genetics. In celebration of the 80th anniversary of the Hardy-Weinberg (H-W) law, Li (41) proved that random mating is a sufficient condition, but not a necessary one, for the attainment of H-W proportions. With respect to an autosomal diallelic locus (with alleles A and a occurring in proportions p and q in a population) it is well-known that random mating implies: 1. the equilibrium proportions of genotypes AA, Aa, and aa in the population are p^2, 2pq, and q^2, respectively; 2. correlation between mates is 0; and 3. parent-child correlation = sib-sib correlation = 1/2. The question Li asked is this: Does the satisfaction of all three properties imply that the population is practicing random mating? The answer is no; Li produced an infinite number of counterexamples! The usual statistical tests for random mating are actually only tests for random union of gametes, whatever the mating pattern in the population. This has a profound implication in anthropological genetics: Even when no statistically significant deviation of observed genotype frequencies from H-W expectations is detected at a particular locus, one really cannot be certain that the population is practicing random mating with respect to this locus; this makes the study of mating patterns worthwhile.

POPULATION STRUCTURE

Peculiarities in mating patterns often create genetic subdivisions within a population. In humans, small local subpopulations are often created because of formation of factions and subsequent fission, phenomena that profoundly affect (sub)population size and structure. The effect is an acceleration of genetic microdifferentiation, as has been clearly demonstrated among the Yanomama (79).

Fixation Indexes

The nature and extent of genetic microdifferentiation in a subdivided population are measured by fixation indexes (101, 102), or, more generally, by indexes of gene diversity (54). While Wright defined the fixation indexes using population gene and genotype frequencies, in practice these indexes have to be estimated from sample frequencies. Nei & Chesser (56) have obtained estimators, most of which are unbiased, of parameters that arise in connection with gene diversity analysis. They have also shown that the sampling biases of the estimates are quite small when the number of individuals sampled from each subpopulation exceeds 50. Weir & Cockerham (97) have also studied this problem using the "superpopulation" framework. These authors consider a two-stage sampling process—sampling of subpopulations (from a "superpopulation") and sampling of individuals from subpopulations. Their estimators, therefore, account for biases at both stages of sampling. However, in practice, the subpopulations considered are usually not treated as a sample from a population; data from all identifiable subpopulations of a population are generally collected. The practical relevance of considering the first stage of sampling therefore remains unclear. An alternative way of defining the indexes was proposed by Cockerham (11, 12) using the analysis of variance (ANOVA) technique. This approach was extended to the multiallele, multilocus case by Long (44) using the multivariate ANOVA technique. An application is provided by Smouse & Long (78). However, Long's estimators do not entirely agree with the spirit of Cockerham's estimators and can result in gross numerical differences (8).

Effects of Migration and Mutation

Although all populations are finite, in practice, often no serious error in inference results even when genetic analyses are performed under the assumption of infinite population size. In the study of subdivided populations, however, consideration of the finiteness of the total population and of subpopulations is very important. In a finite population, in the absence of mutation, the entire population eventually becomes homozygous. Further, if the population is subdivided, and there is no migration among the subpopulations, random genetic drift creates genetic divergence of the subpopulations and each subpopulation eventually becomes homozygous, but not necessarily for the same allele. Migration among subpopulations retards genetic divergence. Although there are several models for analyzing the genetic structure of subdivided populations, the island model is the simplest and the most well-studied. For the island model, it is well-known that the equilibrium value of F, the probability that two alleles chosen at random from the same subpopulation are identical by descent is $F \simeq (1 + 4N_em)^{-1}$, where N_e = effective subpopulation size and m = migration rate per genera-

tion. The decrease in F with increase in m is very rapid. Without any migration (m = 0), obviously F = 1. With one migrant every fourth generation (N_em = 0.25), F = 0.5. Thus, even a small amount of migration substantially prevents genetic divergence between subpopulations. In the presence of mutation, certain remarkable things happen. The coefficient of gene differentiation (54), G_{ST}, which measures the extent of reduction in average heterozygosity of the subpopulations relative to the total population heterozygosity, can be shown (14, 40, 90) to be $G_{ST} = (1 + 4N_em)^{-1}$, at equilibrium, if the migration rate is substantially greater than the mutation rate (which is generally true) and if the total effective population size and the number of subpopulations are large. Thus, G_{ST} is independent of the mutation rate and of the number of alleles; provided that the number of subpopulations is large ($\simeq$ 30), it is also independent of the actual number of subpopulations.

Estimation of Migration Rates

Using the idea that unless migration rates among a set of populations is high, the dispersal of rare alleles will be restricted, Slatkin (73) devised a new indirect measure of gene flow. If s denotes the number of subpopulations (or locations of sampling) of a population, then the "conditional average frequency", $\bar{p}(i)$, defined as the average frequency of all alleles found in exactly i of the s subpopulations, was found to be independent of mutation and selection but strongly dependent on the average level of gene flow as measured by N_em in an island or a stepping-stone model. $\bar{p}(1)$ was shown to be a simple function of N_em: $\ln[\bar{p}(1)] \simeq a \cdot \ln(N_em) + b$, where a and b are functions of sample sizes. Inversion of this function provides a simple estimator of N_em: $\hat{N_em} = \exp[\{\ln \bar{p}(1) - b\}/a]$. However, for low levels of gene flow, a long time is required for the conditional average frequencies to reach equilibrium. This method, therefore, is not useful for distinguishing low levels of long-term gene flow from no current gene flow. More recently, a means of inferring gene flow from phylogenies of nonrecombining DNA segments has been devised (76). This measure, too, is insensitive to low levels of gene flow. In humans, slow infusion of genes over a long period from one population to another is the rule. Mass migrations in humans are generally related to infrequent social or physical catastrophies. Thus, the utility of these methods in anthropological genetics seems limited. Some attempts are now being made to devise measures for detecting low levels of gene flow (75).

Effects of Ignoring Population Subdivision

A practice not uncommon in genetic data analysis is the pooling of samples. Samples from distinct Mendelian isolates are pooled to inflate sample size, and the pooled sample is treated as a sample from a single panmictic unit if a statistical test fails to detect significant departures of pooled genotype pro-

portions from H-W expectations. Sometimes, a sample is assumed to be from a single panmictic unit when the existence of subdivisions is unknown. What effect does this process of amalgamation have on the various population genetic parameters? This was the question asked in a recent study (10). The major effects of amalgamation were found to be: 1. a large excess (about two-fold in a mixture of 12 populations using 27 loci) in the observed number of alleles per locus compared to that expected under neutral-mutation/drift equilibrium, 2. a statistically significant excess of the number of rare alleles, 3. increase in both excesses with the number of populations amalgamated and also with the average genetic diversity (54) among them, and 4. an insensitivity of average heterozygosity (54) to amalgamation. Thus, the ignoring of population subdivision may lead to incorrect population genetic inferences. The excess of rare alleles observed at some loci in six Asian populations (98) may be because of amalgamation of heterogeneous subpopulations (9). Further, estimates of mutation rates using rare alleles (52, 53) may be seriously affected by population amalgamation.

Estimation of Effective Population Size

The effective population size (N_e) is one of the most basic parameters in population genetics. Dynamics of gene frequency change in a population is directly dependent on N_e, not on the total population size N (15). The demographic profile of the population under consideration determines the relationship between N_e and N. The direct estimation of N_e being difficult and cumbersome, attempts have been made to estimate N_e indirectly (38, 60, 68, 69, 89, 95). The indirect estimation procedure uses the following relations (15): V_t = variance of allele frequency in t-th generation = $p_0(1 - p_0)[1 - \{1 - (1/2N_e)\}^t]$, F = fixation index (101) = $V_t/\{p_0(1 - p_0)\} \simeq t/(2N_e)$, where p_j = frequency of a neutral allele in generation j. Thus, $\hat{N}_e = t/(2F)$. The major problem with the above estimator of N_e is that F needs to be replaced by $\hat{F}$, the sample estimate of F, since the allele frequencies are known only in a sample, and not in the total population. Further, $\hat{F}$ is dependent on the actual process of sampling of individuals for estimation of allele frequencies. In humans, this is generally done by sampling adults before reproduction begins (and then replaced into the population) or after reproduction ceases. (For other sampling schemes and their effects, see 60, 95.) For this sampling scheme, when N_e = N, the estimator of N_e is (60): $\hat{N}_e = (t - 2)/[2\{\hat{F} - 1/(2n_0) - 1/(2n_t)\}]$, where

$$\hat{F} = (1/K) \sum_{i=1}^{K} (p_{oi} - p_{ti})^2/[(p_{oi} + p_{ti})/2 - p_{oi}p_{ti}],$$

K = number of alleles at the locus, n_j = sample size for generation j, p_{ji} =

frequency of allele i in generation j. For the more general case (95) when $N_e \leq N$, $\hat{N}_e = t/[2\{\hat{F} - 1/(2n_0) - 1/(2n_t) + 1/N\}]$. The above estimator is not useful unless the population size is known. If the population size is not known, but a rough estimate of the proportion $N_e/N = r^{-1}$ is known, then the above formula can be modified as (95): $\hat{N}_e = (rt - 2)/[2r\{\hat{F} - 1/(2n_0) - 1/(2n_t)\}]$. A generalized approach to estimation of N_e can be found in Waples (95). From a practical point of view, precision in the estimate of N_e increases with increase in sample size (which should preferably be > 50), the ratio n/N_e (which means that populations with small N_e are studied more effectively by the temporal method), number of alleles, and number of generations between samples. Use of alleles with intermediate frequencies also increases precision. Effects of selection, migration, and variability in effective population size over generations have also been considered (60, 95).

MULTIGENE FAMILIES

Recent molecular data have resulted in some major changes in perspective for the population geneticist. These include the consideration of the dynamics of large regions of a genome encompassing a large number of loci, instead of just one or two loci. This expansion of consideration has resulted in an increase in the number of parameters (e.g. number of loci in the genomic region under consideration, recombination rates and patterns among loci, etc) and in the mathematical complexity of the models used to study the dynamics of genomic regions. The study of large genomic regions, instead of individual loci, has provided us with a better understanding of the nature and extent of genetic diversity and the genetic structure of populations.

Characteristics of a Multigene Family

A multigene family is defined (25) as a group of genes that exhibit four properties— 1. multiplicity, 2. close linkage, 3. sequence homology, and 4. related overlapping phenotypic functions. Multigene families are found in eukaryotes but not in prokaryotes. Gene duplication and subsequent functional differentiation, which have been major features of long-term evolution (62), handle the increased complexity of biological organization. In humans, the most well-studied multigene families are hemoglobin genes and antibody genes. Here I consider only the hemoglobin genes. Located on chromosomes 11 and 16 are, respectively, the β-globin and the α-globin gene families. Products of genes of these two families lead to the formation of the hemoglobin molecule. The organizations of these two gene families are: ζ - $\psi\zeta$ - $\psi\alpha_2$ - $\psi\alpha_1$ - α_2 - α_1 - θ_1 and $\psi\beta_2$ - ϵ - G_γ - A_γ - $\psi\beta_1$ - δ - β. These two gene families obviously satisfy properties 1 and 2 (above) required of a multigene family. The sequence homologies (property 3 above) among

members of both the α- and the β-globin gene families is well documented (16). The functions of the genes of each of these families are related (property 4). The genes are all involved in the production of the hemoglobin molecule but are specialized to express during different periods of development. The ϵ and ζ genes are expressed in the embryo; the G_γ, A_γ and the α genes in the fetus; and the β, δ, and α genes in the adult. The function and the time of expression of the θ_1 gene is unclear (46). The evolution of these gene clusters has been extensively studied (16, 30). Noting that a single-chain globin still occurs in a lower vertebrate (the lamprey), the most plausible model, arrived at by extensive comparisons of amino-acid and nucleotide sequences of globin genes within and across species, is that a gene duplication event occurred early in vertebrate evolution about 500 million years ago, which resulted in the α- and β-globin genes. Thereafter, a series of tandem gene duplications occurred, which at some stage became unlinked to give the separate α and β gene clusters.

Multigene families show two remarkable characteristics: 1. change in the number of repeating genes within the family during evolution, and 2. concerted evolution. An extreme example of characteristic 1 is the number of 5S ribosomal RNA genes in *Xenopus* frog species; the number in *X. laevis* is 24,000 and in *X. mulleri* is 9,000. Although to a much smaller degree, changes in repeating gene numbers are also known for human hemoglobins. Chromosomes carrying three α-globin genes or one α-globin gene have been detected (23). "Concerted evolution" (103), earlier termed "coincidental evolution" (25), is a process by which duplicate genes do not diverge independently after duplication. Some mechanism maintains a close sequence homology between them. Often, the same mutation is found to occur in all members of the family, the probability of which is virtually zero if these mutations are all independent. Examples of concerted evolution related to human globin genes are: 1. the G_γ and A_γ genes on a single chromosome show extreme homology, even though these genes arose by gene duplication about 20–40 million years ago (77); and, 2. the α_1 and α_2 genes are almost identical in sequence even though they arose by gene duplication at least 8 million years ago (43). [Amplification of these concepts, further examples, and references can be found in Dover (18).] Several hypotheses have been propounded to explain the expansion of the size of a multigene family. The most tenable seems to be that of homologous unequal crossing over (see 99, Figure 1). For explaining concerted evolution, a "correction" mechanism called "gene conversion" has been proposed (18).

Population Genetics of Multigene Families

The study of the evolution and maintenance of genetic variability in multigene families from a population genetics standpoint has been pioneered by Ohta (63, 66) and Nagylaki (48, 50, and references in these papers). Mechanisms

considered in these studies are gene duplication, crossing over, gene conversion, mutation, and random genetic drift. Since the mathematical treatment of change in copy number is complex, this phenomenon has been studied mainly by computer simulation (65). It has been shown that starting with one gene, the copy number increases rapidly with the number of generations. There is also an accumulation of pseudogenes and an increase in the number of beneficial alleles. Since the gene families in the population, although different from one another, are mutually related through common ancestors, modeling and analyses of concerted evolution are based mainly on identity coefficients defined as the probability that two genes chosen at random either both from the same gene family or one each from two different families are identical by descent. [The similarity of these identity coefficients to those used in inbreeding studies (15) should be noted.] Homogenization of genetic information by concerted evolution (crossing over and gene conversion) obviously has profound effects on the identity coefficients. In a series of papers, Ohta (see 64, 66 for reviews) has derived the equilibrium values of these identity coefficients. She has also shown that the extent of uniformity of genetic information among members of a multigene family is strongly dependent on the rate of homogenization, which is thought to be evolutionarily adjusted. Some of the predictions have yielded good fits to data (47, 72). However, because the evolution of multigene families is dependent on many biological factors, not all of which are well understood, mathematical modeling has necessarily used a simplified set of assumptions. This has inevitably led to some surprising, and apparently contradictory, results (49). A great deal of experimental work is needed to verify the biological validity of some of the assumptions.

In the modeling of multigene families, selection has largely been ignored. Recently, it has been found (29) that the rate of amino-acid-altering substitutions is higher than that of synonymous substitutions in the antigen recognition site (ARS) in the genes of the major histocompatibility complex (MHC) of humans and mice, although this phenomenon is not observed in the other regions of the same genes. Since the ARS plays the crucial role of recognizing antigens, greater polymorphism in the ARS implies increased ability of recognizing a wider variety of antigens. This has been claimed as evidence favoring overdominant selection. Evidence for positive Darwinian selection has also been found in the immunoglobulin heavy-chain variable-region genes in mammals (92). Thus, natural selection may play a strong role in determining the extent of polymorphism present in multigene families. Selection, therefore, needs to be incorporated in population genetic modeling of evolution of multigene families. [Ohta (63) has discussed some restricted selection schemes. The pattern of polymorphism in the MHC gene cluster (some genes being polymorphic while others are monomorphic) indicates, however, that these schemes may not be appropriate (M. Nei, personal communication).]

Colonization of Pacific Islands as Revealed by Globin Genes

Polymorphism in multigene families has proved to be an extremely useful tool in studies of human population structure, migration, and affinity. A recent study (61) exemplifies this well. Haplotypes of the α-globin gene cluster, which contains at least 17 informative sites, have provided important clues regarding the colonization of the Pacific Islands. In this study, the α-globin haplotype was characterized by seven polymorphic restriction sites and two length polymorphisms. Seven of these $\alpha\alpha$ haplotypes are common and can be classified in five groups I–V. Haplotypes of groups I and II are common in most populations of the world—in Southeast Asia, the combined frequency is about 90%. In Melanesia, the reverse is true; the combined frequency of haplotypes of groups III–V is 93–100%. The haplotype frequencies found in Micronesia and Polynesia are intermediate between those of Southeast Asia and Melanasia. Some of the -α gene variants that appear to have originated in Melanesia are also found in Polynesia (but nowhere else in the world), and some other more common variants that are found in Micronesia seem to be derived from Melanesia. Triplicated-ζ-gene ($\zeta\zeta\zeta$) chromosomes were found in polymorphic frequencies (6–22%) that formed a continuum from Southeast Asia through Micronesia to Polynesia, but largely excluded Melanesia. The α-globin haplotypes of the $\zeta\zeta\zeta$ chromosomes revealed that these chromosomes are all derived from a common ancestor that most probably arose in mainland Southeast Asia. Thus, the analysis of molecular polymorphisms in the globin multigene family indicates that the Pacific Islands were colonized by people who migrated from Southeast Asia through Melanesia to Micronesia and Polynesia. This genetic inference agrees well with linguistic history.

PHYLOGENETIC TREES

Estimating genetic relationships among populations/species has been a major interest of the population geneticist. For reconstructing phylogenetic relationships among populations, the traditional approach has been to estimate allele frequencies at several loci for each of the populations under consideration, estimate pairwise genetic distances among the populations, and then use a clustering algorithm (80). More recently, since nucleotide sequences of the same gene have become available in many populations, devising methods of phylogenetic inference based on sequence data has become important. A tree constructed from sequence data of one gene in several populations—termed a "gene tree" (93)—yields the evolutionary relationships of the same gene in different populations. A gene tree may differ from the corresponding "population tree" because of the presence of polymorphism. A simple illustration is given in Figure 1. Suppose A, B, and C are three extant populations, and C diverged first (that is, became reproductively isolated) so that the

"population tree" is as in Figure 1(*a*). Suppose two genes labelled (a,a'), (b,b'), and (c,c') are chosen from each of the three populations A, B, and C, respectively. Panels (*b*) and (*c*) of Figure 1 present two possible evolutionary scenarios of these genes. Irrespective of which genes are chosen and sequenced, the data will yield a unique gene tree, which coincides with the population tree, if scenario (*b*) holds. For scenario (*c*), however, the structure of the gene tree is dependent on whether gene b or b' is chosen from population B. If gene b is chosen, the gene tree coincides with the population

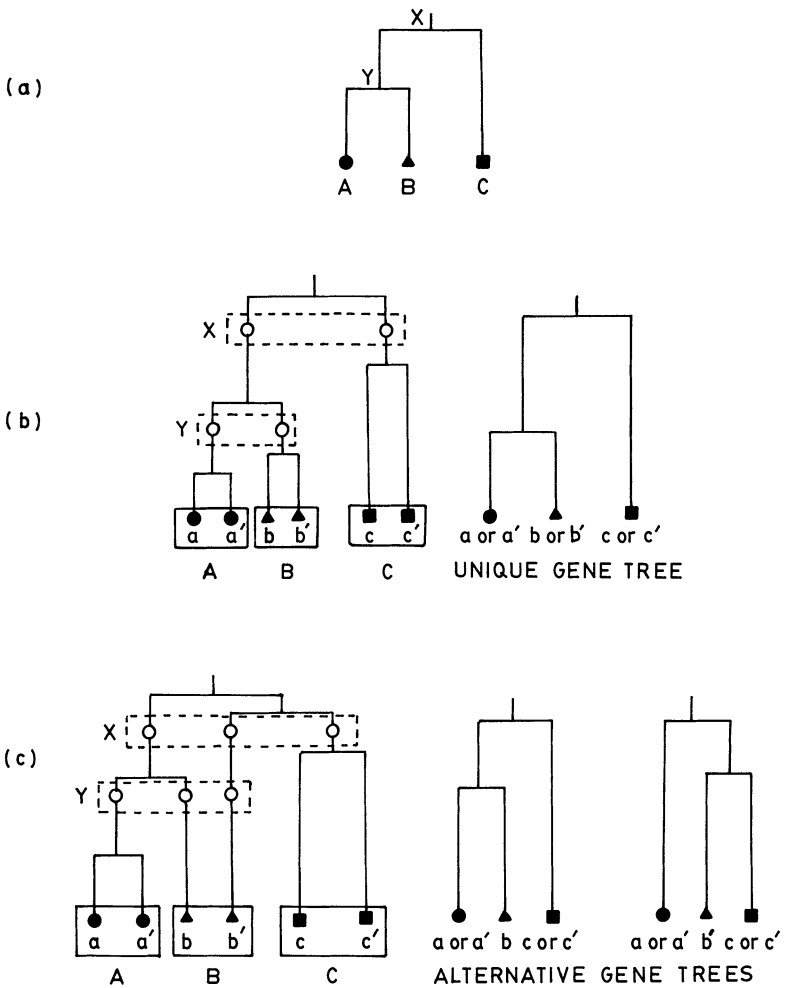

Figure 1 Population trees and gene trees. (*a*) Population tree of three populations. (*b*) The unique gene tree constructed from samples of genes drawn from each population, under one evolutionary scenario. (*c*) Alternative gene trees under another evolutionary scenario.

tree; otherwise not. Thus, the existence of polymorphism in ancestral populations affects phylogenetic inference in respect of populations (54, 67).

Interest is currently high in studying phylogenetic relationships among genes, especially for multigene families. This attention has resulted in a proliferation of statistical methods for constructing phylogenetic trees from DNA sequence data (21, 54). The various methods can be broadly grouped as: 1. distance methods, 2. parsimony methods, and 3. maximum likelihood methods. The assumptions underlying the methods are different; the newer methods generally have less restrictive assumptions (nonconstant nucleotide substitution rates) than the earlier methods [e.g. the average-linkage method (80), which assumes constancy of substitution rates]. Primarily through simulation, the robustness of the various methods of phylogenetic reconstruction is being studied. However, most methods are fairly sensitive to deviations from assumptions regarding substitution rates, patterns of substitution rates among the four nucleotides, etc. The method of choice, therefore, depends on which genes are being compared from which populations—in other words, which assumptions are likely to be satisfied. For recent comparative reviews see Nei (55), Saitou & Imanishi (71) and Felsenstein (21). Finally, to restate the obvious, most methods perform poorly if the number of nucleotides in the sequences under comparison is small. Further, mechanical applications of these methods are likely to lead to incorrect phylogenetic inferences. For example, it is well known that recombination is common in multigene families. It is important to investigate first whether similarities among some of the genes under consideration are likely to be due to past recombinational events. If so, only those genes, or those portions of genes that are not products of recombination, should be analyzed for phylogenetic reconstruction. Some considerations along these lines have been discussed (26, 83, 84).

THE COALESCENT PROCESS

Related to the concept of gene trees is the concept of coalescence. Although this section may seem out of tune with the others in this review, the concept of coalescence introduced by Kingman (37) has proved extremely useful in the study of diverse population genetic problems, including the evolution of multigene families (33, 96). I describe this concept in relation to the study of the age of a gene at a locus. To find out how old a gene is, one can study, moving forward in time, the persistence/extinction of current genes, and then reverse the argument in time, when mathematically permissible (34), to determine the ages of the genes. A more natural approach is to study the process of evolution of genes backward in time, starting from the present. If one samples g genes from a population of G genes ($g \geqslant G$), one can study

their ancestries, since the genes are all related. For example, suppose the ancestries of 5 (= g) genes, labelled 1, 2, 3, 4, and 5 in a population are as depicted in Figure 2. Then, on this genealogy of genes, one can define a stochastic process as follows. At time t_1, the state was $S_1 = \{(1, 2, 3, 4, 5)\}$, meaning that the 5 genes had a common ancestral gene. At time t_2 the state was $S_2 = \{(1, 2, 3),(4, 5)\}$, that is, the genes (1, 2, 3) had one common ancestor and formed an "equivalence class", and the genes (4, 5) had another common ancestor and formed another "equivalence class"; the two equivalence classes being distinguishable from each other. At time t_3, $S_3 = \{(1),(2, 3),(4, 5)\}$; at t_4, $S_4 = \{(1),(2),(3),(4, 5)\}$; and at t_5, $S_5 = \{(1),(2),(3),(4),(5)\}$. At any time, therefore, the relationship among the genes is described by a partition of the set of integers {1, 2,...,g} into a set of equivalence classes. Let E denote the set of all equivalence class partitions of {1, 2,..,g}. For a pair of elements e_1 and e_2 belonging to E, we write $e_1 \rightarrow e_2$ (that is, e_2 can be "reached" from e_1, or, equivalently, e_1 can "move" to e_2) if e_2 can be formed by coalescing two equivalence classes of e_1. [Obviously, not all elements of E can be reached from an element of E [e.g. if $e_1 = \{(1),(2, 3),(4, 5)\}$ and $e_2 = \{(1, 2),(3),(4, 5)\}$, then e_2 cannot be reached from e_1.] Moving backwards in time along the gene genealogy is moving from S_g to S_1 through a random sequence of members of E. The Kingman coalescent process is a continuous time Markov chain, in which the process can move from state S at time t to state S' at time t + dt (provided that $S \rightarrow S'$) with probability $dt + O(dt)^2$, or can remain in S with probability $1 - w \cdot dt + O(dt)^2$, where $w = k(k - 1)/2$, k = number of equivalence classes in S; the probability of any other transition is of order $(dt)^2$. Using this theory, it can be shown that the mean time ($\bar{t}$) when g neutral genes from a locus sampled from a random-mating population coalesce to a single ancestral gene is: $\bar{t} = 4N_e(1 - 1/g)$. The distribution of the coalescent process is completely characterized for many selectively neutral population genetic models (94). Properties of this process under selection and recombination have also been studied (27, 32). I do not dwell further on this process here; important properties have been derived and applications provided (17, 37). I emphasize that the concept of coalescence is very natural from an evolutionary standpoint and that the coalescent process has played an important role in recent theoretical population genetic studies.

It is important to point out that gene genealogy is different from allele genealogy. Two randomly chosen genes at a locus from a population may be identical in nucleotide sequence and may have remained as the same allele. The probability of this event is large when the gene has a short nucleotide sequence or when the population size is small. In empirical investigations one is usually interested in the history of various alleles in a population. It is, therefore, important to study the coalescence of alleles (allele genealogy) rather than of genes. The history of the different alleles at a locus starts from the time a new allele is created by mutation or by recombination from the

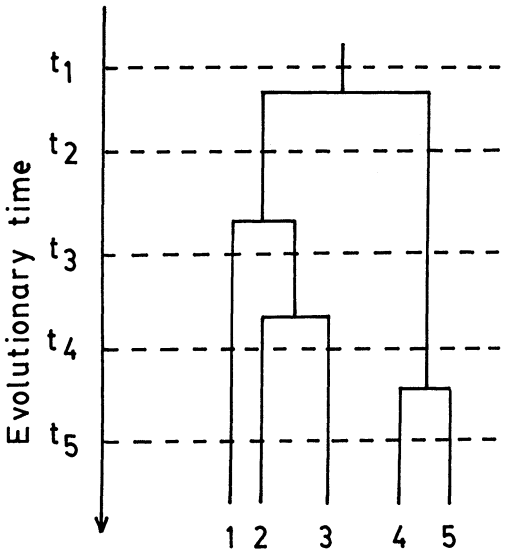

Figure 2 The coalescence of five sampled genes to their ancestral genes

wild-type allele. Unfortunately, the mathematical formulation of allele genealogy seems complicated and remains to be done. In a simulation study (91) it was found that under neutrality the mean coalescence time of different alleles is always smaller than that for genes; the difference in the mean coalescence times decreases with increasing values of $4N_e v$, where v denotes mutation rate. Some simulation results under various selection schemes are also available (91).

GENETIC POLYMORPHISM AND RELATIONSHIPS AMONG HUMAN POPULATIONS

Nuclear Gene Markers

Studying genetic relationships among contemporary human populations has long been of interest. These studies shed light on the origin of modern humans and also on the nature, extent, and causes of genetic differentiation. From various studies it is clear (59) that there is greater genetic variability within major population groups—Caucasoids, Mongoloids, Negroids—than between them. In a recent study (57) it has been shown, using gene frequency data on 186 (61 DNA marker; 152 polymorphic) loci, that Europeans and Asians are genetically closer to each other than to Africans at a statistically significant level. An African/non-African split has also been observed in another large study comprising 42 populations and 42 loci (7). These inferences are consistent with an African origin of modern humans. The non-

African cluster also separates into two major clusters, one corresponding to Caucasoids, East Asians, Arctic populations, and Amerindians, and the other to Southeast Asians, Pacific islanders, New Guineans, and Australians. The average genetic distances between clusters correlate well with the separation times estimated from archeological materials; there is also a remarkable correspondence between the linguistic phyla and genetic clusters (7). Languages evolve more rapidly than genes. Historical invasions have played a role in the replacement of languages in some geographical areas, even if the invaders were sometimes a numerical minority (70). Therefore, although linguistic and genetic distances may correlate well at a broad level, no such correspondence may be found at a narrower level because linguistic changes within language phyla are rather rapid. Apart from physical barriers, language also forms a barrier to migration, and migration is a homogenizing factor from a genetic viewpoint. By studying spatial patterns of gene frequencies at different loci, is it possible to discover the various factors that have affected the genetic structure of populations residing in a particular geographical area? This is the question that Sokal's group has addressed (81 and references therein). The geographical area they have considered is Europe. They have analyzed gene frequency data at 26 loci collected from about 3500 localities. Sophisticated statistical methods (correlograms, "Wombling", etc) were used to study spatial patterns. The major results of these studies (2, 81) are: 1. there is a strong decline in overall genetic similarity with geographic distance, 2. directional patterns caused by past migration are still discernible, 3. of the 33 detectable boundaries of sharp gene-frequency change, 31 coincide with linguistic boundaries, and 4. of these 31 boundaries, 22 are obvious physical barriers. These results indicate that the general pattern is of localized random fluctuations of gene frequencies and patchy patterns consistent with an isolation-by-distance model. On the larger scale, directional patterns caused by past (probably more than 5000 years ago) migration manifest as language boundaries and geographical barriers. It is also clear that the present genetic structure of European populations cannot be explained by a single population genetic model. Local adaptation seems to be a less important factor than obstacles to population admixture in determining abrupt genetic changes.

Mitochondrial DNA Markers

Mitochondrial DNA (mtDNA) is a self-replicating circular DNA molecule of about 16.5 kilobases (1) . Being maternally inherited in a haploid fashion (22), mtDNA morphs of individuals can be identified without family data. The effective population size for mtDNA is about one fourth that for nuclear genes (3). The base substitution rate of mtDNA is between five and ten times that of nuclear DNA (4), and mtDNA evolves without recombination independently of the nuclear DNA. The high substitution rate coupled with a

smaller effective population size makes mtDNA very useful in studying genetic differentiation of human populations. Because it is maternally inherited, phylogenetic trees of mtDNA morphs reflect the maternal histories of populations. There is a great deal of restriction fragment length polymorphism (RFLP) of mtDNA. Further, comparison with the known nucleotide sequence of mtDNA (1) often permits the identification of the exact location and nature of the base substitution responsible for a particular restriction site polymorphism (5). Since the early 1980s, Cavalli-Sforza's group in Palo Alto and Allan Wilson's group in Berkeley have done pioneering studies on mtDNA variation in human populations. The combined data have revealed an amazing degree of mtDNA polymorphism in human populations (6, 31). In fact, with the higher-resolution restriction mapping, it has been shown (6) that the five continents have completely disjoint sets of mtDNA types. In contrast to the inference from nuclear gene polymorphism (58) that only about 10% of the total genetic variation in the world is attributable to between-population variation ($G_{ST} = 0.10$), mtDNA data have shown (85) that the corresponding percentage is about 30 ($G_{ST} = 0.31$). Variation of a comparable degree has also been found within smaller geographic regions of Papua New Guinea (85). Thus mtDNA markers are very useful in studies of genetic differentiation of human populations. Polymorphism in mtDNA has also been used to test two contrasting hypotheses on origin of modern *Homo sapiens*: 1. The Single Replacement hypothesis states that all anatomically modern populations of *H. sapiens* originated in Africa and then replaced with little or no hybridization all older populations of archaic *H. sapiens*. 2. The Multiregional Transition hypothesis states that *H. erectus* originated in Africa and spread to temperate regions of Eurasia and evolved independently to produce the modern, major regional human populations. These hypotheses also involve a temporal dimension not discussed here (see 86 for details). Wilson et al (100) have constructed a phylogenetic tree of 147 mtDNA types sampled from Africa (n = 20), Asia (n = 34), Australia (n = 20), New Guinea (n = 30), and Europe (n = 43). The tree has two primary branches: one African branch comprising 7 African mtDNA types, and the other branch comprising all other mtDNA types in which the remaining African mtDNA types are interspersed as "twigs". This finding has been interpreted to support the Single Replacement hypothesis and to suggest that the common ancestral mother ("Eve") was African (6, 100). There are two major problems with this conclusion. First, the phylogenetic tree of mtDNA types is a gene tree. As we have noted above, in the presence of ancestral polymorphism a gene tree may not coincide with a population tree. There is some evidence that ancestral polymorphism may indeed have been present for mtDNA (24). Second, mtDNA evolves essentially without recombination, implying that the phylogenetic tree constructed is based on data of only one genetic locus, and

hence is subject to large stochastic errors. Cann et al (6) have also calculated the age of our common mother to be about 200,000 years. The methods used to estimate this evolutionary time have been widely criticized. The primary concern has been for the validity of assumptions regarding mtDNA substitution rates. I do not discuss these issues here; interested readers may consult the relevant sources (39, 58, 82). Can father Adam be spatially (and temporally) far removed? To investigate this issue, studies on the human Y chromosome have been initiated (19, 45). Unfortunately the extent of polymorphism of the Y chromosome has been found to be very low. Although the Y chromosome data are compatible with an African origin, the evidence is still extremely weak.

FINAL REMARKS

As mentioned in the Introduction above, empirical studies and data-oriented modeling have rightly gained importance in population genetics. With the increasing simplification of molecular genetic techniques, theoretical results in population genetics can now be empirically tested with greater ease. Collaborations among the theoretical population geneticists and experimental geneticists have become common. The compromising attitudes of both neutralists and selectionists have resulted in healthy growth of population genetics. There are, of course, many phenomena (e.g. speciation) on which opinions differ. The "molecular clock" (average substitution rate per unit of evolutionary time) does not seem to be ticking at as constant a rate as was previously thought. There are also differences of opinion on the causes of the observed nonconstancy of evolutionary rates along certain evolutionary lineages. In spite of such differences of opinion on many basic issues, the last decade has witnessed tolerance and healthy growth. Why have advances in molecular techniques been so important to the growth of population genetics? The primary reason is that population genetics modeling uses parameters that could earlier be estimated only indirectly but these now, with the advent of molecular technology, can be estimated directly. For example, to estimate mutation rates, specifically nucleotide substitution rates, one had to rely primarily on frequencies of protein variants detected by electrophoresis. This was an indirect procedure because electrophoresis detects only those substitutions that cause a charge change. Thus substitution rates had to be estimated indirectly from the proportion of mutations that caused a charge change. Nucleotide sequences can now be directly compared and mutations counted. For the same reason, while older techniques detected only a fraction of the variation present at the DNA level, the current techniques allow one to detect genetic polymorphism much more accurately. A wide variety of interesting results have been obtained in recent years. The study of mating

patterns is important even when there are no significant deviations from Hardy-Weinberg proportions. Identification of subdivisions of a population is crucial for a correct understanding of population structure. Therefore, in empirical population genetic studies among humans, matrimonial relationships and migrations need to be carefully recorded. Equilibrium values of fixation indexes for subdivided populations indicate that when the number of subpopulations is large, even with a small amount of migration among the subpopulations, the finite n-subpopulation island model with mutation becomes equivalent to the infinite island model with no mutation. This finding is remarkable. Considerable progress has been made in understanding patterns of gene flow. It is remarkable that ancient large-scale gene flow can still be detected from current gene frequency patterns by using appropriate and sensitive statistical methodology. However, methods of detecting low levels of gene flow need to be considerably refined. Substantial progress has been made in the procedures for estimating the effective size of a population using data on temporal changes of gene frequencies. This indirect approach is important because it is difficult to gather the detailed demographic information necessary for direct estimation of effective population size. The clarification of differences between gene trees and population trees has been fundamental to population genetic thinking. The anthropologist is more interested in population trees than gene trees. The concordance between a gene tree and the corresponding population tree increases with increase in the number of genes sampled from each of the populations under consideration. Inferences regarding evolutionary relationships among populations should, therefore, be based on a large sample of genes from each population. Consideration of the dynamics of multigene families has yielded a lot of valuable information on the processes of genome evolution. Molecular characterization of members of important multigene families in humans has also been helpful to anthropogenetic studies, as exemplified by the studies on colonization of the Pacific Islands. Certain major areas of inquiry have not been touched on in this review. First, many genetic studies have been conducted to resolve whether humans are evolutionarily closer to chimpanzees than to gorillas or whether chimpanzees and gorillas are closer to each other than to humans. No unequivocal resolution of this trichotomy has yet been obtained, even by the use of various types of molecular data (DNA-DNA hybridization, nucleotide and amino acid sequences) and statistical techniques. Second, important results on the population genetics of quantitative characters have been obtained in the last decade. Third, in genome evolution and expansion, transposable elements, which are DNA sequences that can change their location within a genome and often increase in copy number, seem to play an important role. The dynamics of change in copy number of transposable elements vis-à-vis its effects on the fitness of the host has been a major recent

interest. However, many of the assumptions used in these theoretical studies must still be empirically tested. Fourth, progress has been made in studying the dynamics of disease genes in populations. The fact that population genetic ideas (e.g. linkage disequilibrium) are being increasingly used to identify locations of disease genes on chromosomes is noteworthy. It is hoped that these issues will be reviewed in the near future.

ACKNOWLEDGMENTS

Professor M. Nei provided clarifications of many issues discussed in the present article; I am very grateful to him. I thank Prof. A. Basu, Dr. A. Chakravarti, Dr. R. Deka, Dr. R. Gupta, and Prof. C. C. Li for reading an earlier draft and for providing valuable comments and suggestions.

Literature Cited

1. Anderson, S., Banitier, A. T., Barrell, B. G., De Bruijn, M. H. L., Coulson, A. R. et al. 1981. Sequence and organization of the human mitochondrial genome. *Nature* 290:457–65
2. Barbujani, G., Sokal, R. R. 1990. Zones of sharp genetic change in Europe are also linguistic boundaries. *Proc. Natl. Acad. Sci. USA* 87:1816–19
3. Birky, C. W., Maruyama, T., Fuerst, P. 1983. An approach to population and evolutionary genetic theory for genes in mitochondria and chloroplasts, and some results. *Genetics* 103:513–27
4. Brown, W. M. 1985. The mitochondrial genome of animals. In *Molecular Evolutionary Genetics,* ed. R. J. MacIntyre, pp. 95–130. New York: Plenum
5. Cann, R. L., Brown, W. M., Wilson, A. C. 1984. Polymorphic sites and the mechanism of evolution in human mitochondrial DNA. *Genetics* 106:479–99
6. Cann, R. L., Stoneking, M., Wilson, A. C. 1987. Mitochondrial DNA and human evolution. *Nature* 325:31–36
7. Cavalli-Sforza, L. L., Piazza, A., Menozzi, P., Mountain, J. 1988. Reconstruction of human evolution: bringing together genetic, archaeological, and linguistic data. *Proc. Natl. Acad. Sci. USA* 85:6002–6
8. Chakraborty, R. 1988. Analysis of genetic structure of a population and its associated statistical problems. *Sankhyā* 50(B):327–49
9. Chakraborty, R. 1990. Mitochondrial DNA polymorphism reveals hidden heterogeneity within some Asian populations. *Am. J. Hum. Genet.* 47:87–94
10. Chakraborty, R., Smouse, P. E., Neel, J. V. 1988. Population amalgamation and genetic variation: observations on artificially agglomerated tribal populations of Central and South America. *Am. J. Hum. Genet.* 43:709–25
11. Cockerham, C. C. 1969. Variance of gene frequencies. *Evolution* 23:72–84
12. Cockerham, C. C. 1973. Analysis of gene frequencies. *Genetics* 74:679–700
13. Crow, J. F. 1987. Population genetics history: a personal view. *Annu. Rev. Genet.* 21:1–22
14. Crow, J. F., Aoki, K. 1984. Group selection for a polygenic behavioral trait: estimating the degree of population subdivision. *Proc. Natl. Acad. Sci. USA* 81:6073–77
15. Crow, J. F., Kimura, M. 1970. *An Introduction to Population Genetics Theory.* New York: Harper and Row
16. Dayhoff, M. O. 1972. *Atlas of Protein Sequence and Structure.* Silver Spring, MD: Natl. Biomed. Res. Found.
17. Donnelly, P., Tavare, S. 1986. The ages of alleles and a coalescent. *Adv. Appl. Probab.* 18:1–19
18. Dover, G. 1982. Molecular drive: a cohesive mode of species evolution. *Nature* 299:111–17
19. Ellis, N., Taylor, A., Bengtsson, B. O., Kidd, J., Rogers, J. et al. 1990. Population structure of the human pseudoautosomal boundary. *Nature* 344:663–65
20. Ewens, W. J. 1990. Population genetics theory—the past and the future. In *Mathematical and Statistical Developments of Evolutionary Theory,* ed. S. Lessard, pp. 177–227. Dordrecht: Kluwer Academic
21. Felsenstein, J. 1988. Phylogenies from molecular sequences: inference and

reliability. *Annu. Rev. Genet.* 22:521–65
22. Giles, R. E., Blanc, H., Cann, H. M., Wallace, D. C. 1980. Maternal inheritance of human mitochondrial DNA. *Proc. Natl. Acad. Sci. USA* 77:6715–19
23. Goossens, M., Dozy, A. M., Embury, S. H., Zachariades, Z., Hadjiminas, M. G. et al. 1980. Triplicated alpha-globin loci in humans. *Proc. Natl. Acad. Sci. USA* 77:518–21
24. Harihara, S., Saitou, N., Hirai, M., Gojobori, T., Park, K. S. et al. 1988. Mitochondrial DNA polymorphism among five Asian populations. *Am. J. Hum. Genet.* 43:134–43
25. Hood, L., Campbell, J. H., Elgin, S. C. R. 1975. The organization, expression and evolution of antibody genes and other multigene families. *Annu. Rev. Genet.* 9:305–53
26. Hudson, R. R., Kaplan, N. L. 1985. Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* 111:147–64
27. Hudson, R. R., Kaplan, N. L. 1988. The coalescent process in models with selection and recombination. *Genetics* 120:831–40
28. Hudson, R. R., Kreitman, M., Aguade, M. 1987. A test of neutral molecular evolution based on nucleotide data. *Genetics* 116:153–59
29. Hughes, A. L., Nei, M. 1989. Nucleotide substitution at major histocompatibility complex class II loci: evidence for overdominant selection. *Proc. Natl. Acad. Sci. USA* 86:958–62
30. Jeffreys, A. 1982. Evolution of globin genes. In *Genome Evolution*, ed. G. A. Dover, R. B. Flavell, pp. 157–76. New York: Academic
31. Johnson, M. J., Wallace, D. C., Ferris, S. D., Rattazzi, M. C., Cavalli-Sforza, L. L. 1983. Radiation of human mitochondrial DNA types analyzed by restriction endonuclease cleavage patterns. *J. Mol. Evol.* 19:255–71
32. Kaplan, N. L., Darden, T., Hudson, R. R. 1988. The coalescent process in models with selection. *Genetics* 120:819–29
33. Kaplan, N. L., Hudson, R. R. 1987. On the divergence of genes in multigene families. *Theor. Pop. Biol.* 31:178–94
34. Kelly, F. P. 1979. *Reversibility and Stochastic Networks*. New York: John Wiley and Sons
35. Kimura, M. 1983. *The Neutral Theory of Molecular Evolution*. Cambridge: Cambridge Univ. Press
36. Kimura, M. 1990. Some models of neutral evolution, compensatory evolution and the shifting balance process. *Theor. Pop. Biol.* 37:150–58
37. Kingman, J. F. C. 1982. The coalescent. *Stoch. Proc. Appl.* 13:235–48
38. Krimbas, C. B., Tsakas, S. 1971. The genetics of *Dacus oleae*. V. Changes of esterase polymorphism in a natural population following insecticide control—selection or drift? *Evolution* 25:454–60
39. Kruger, J., Vogel, F. 1989. The problem of our common mitochondrial mother. *Hum. Genet.* 82:308–12
40. Latter, B. D. H. 1973. The island model of population differentiation: a general solution. *Genetics* 73:147–57
41. Li, C. C. 1988. Pseudo-random mating populations. In celebration of the 80th anniversary of the Hardy-Weinberg law. *Genetics* 119:731–37
42. Li, W.-H., Luo, C.-C., Wu, C.-I. 1985. Evolution of DNA sequences. In *Molecular Evolutionary Genetics*, ed. R. J. MacIntyre, pp. 1-94. New York: Plenum
43. Liebhaber, S. A., Goossens, M., Kan, Y. W. 1981. Homology and concerted evolution at the α_1, and α_2 loci of human α-globin. *Nature* 290:26–29
44. Long, J. C. 1986. The allelic correlation structure of Gainj and Kalam speaking people. I. The estimation and interpretation of Wright's F statistics. *Genetics* 112:629–47
45. Lucotte, G., Guerin, P., Halle, L., Loirat, F., Hazout, S. 1989. Y chromosome DNA polymorphisms in two African populations. *Am. J. Hum. Genet.* 45:16–20
46. Marks, J., Shaw, J.-P., Shen, C.-K. J. 1986. Sequence organization and genomic complexity of a primate θ_1 gene, a novel α-globin-like gene. *Nature* 321:785–88
47. Matsuo, Y., Yamazaki, T. 1989. Nucleotide variation and divergence in the histone multigene family of *Drosophila melanogaster*. *Genetics* 122:87–97
48. Nagylaki, T. 1984. The evolution of multigene families under intrachromosomal gene conversion. *Genetics* 106:529–48
49. Nagylaki, T. 1988. Gene conversion, linkage, and the evolution of multigene families. *Genetics* 120:291–301
50. Nagylaki, T., Petes, T. D. 1982. Intrachromosomal gene conversion and the maintenance of sequence homogeneity among repeated genes. *Genetics* 100: 315–37
51. National Institutes of Health. 1984. Proceedings of a workshop on population genetics. *Genetics* 106(No.4, April 1984):i–ii

52. Neel, J. V., Rothman, E. D. 1978. Indirect estimates of mutation rates in tribal Amerindians. *Proc. Natl. Acad. Sci. USA* 75:5585–88
53. Nei, M. 1977. Estimation of mutation rate from rare protein variants. *Am. J. Hum. Genet.* 29:225–32
54. Nei, M. 1987. *Molecular Evolutionary Genetics*. New York: Columbia Univ. Press
55. Nei, M. 1990. Relative efficiencies of different tree-making methods for molecular data. In *Recent Advances in Phylogenetic Studies of DNA Sequences*, ed. M. Miyamoto, J. Cracraft. Oxford: Oxford Univ. Press. In press
56. Nei, M., Chesser, R. K. 1983. Estimation of fixation indices and gene diversities. *Ann. Hum. Genet.* 47:253–59
57. Nei, M., Livshits, G. 1989. Genetic relationships of Europeans, Asians and Africans and the origin of modern *Homo sapiens*. *Hum. Hered.* 39:276–81
58. Nei, M., Livshits, G. 1990. Evolutionary relationships of Europeans, Asians and Africans at the molecular level. In *Population Biology of Genes and Molecules*, ed. N. Takahata, J. F. Crow, pp. 251–65. Tokyo: Baifukan
59. Nei, M., Roychoudhury, A. K. 1982. Genetic relationship and evolution of human races. *Evol. Biol.* 14:1–59
60. Nei, M., Tajima, F. 1981. Genetic drift and estimation of effective population size. *Genetics* 98:625–40
61. O'Shaughnessy, D. F., Hill, A. V. S., Bowden, D. K., Weatherall, D. J., Clegg, J. B. et al. 1990. Globin genes in Micronesia: origins and affinities of Pacific Island peoples. *Am. J. Hum. Genet.* 46:144–55
62. Ohno, S. 1970. *Evolution by Gene Duplication*. Berlin: Springer-Verlag
63. Ohta, T. 1980. *Evolution and Variation of Multigene Families*. Berlin: Springer-Verlag
64. Ohta, T. 1983. On the evolution of multigene families. *Theor. Pop. Biol.* 23: 216–40
65. Ohta, T. 1988. Further simulation studies on evolution by gene duplication. *Evolution* 42:375–86
66. Ohta, T. 1990. How gene families evolve. *Theor. Pop. Biol.* 37:213–19
67. Pamilo, P., Nei, M. 1988. Relationships between gene trees and species trees. *Mol. Biol. Evol.* 5:568–83
68. Pamilo, P., Varvio-Aho, S. 1980. On the estimation of population size from allele frequency changes. *Genetics* 95: 1055
69. Pollak, E. 1983. A new method for estimating the effective population size from allele frequency changes. *Genetics* 104:531–48
70. Renfrew, C. 1987. *Archeology and Linguistics*. Cambridge: Cambridge Univ. Press
71. Saitou, N., Imanishi, T. 1989. Relative efficiencies of the Fitch-Margoliash, maximum parsimony, maximum likelihood, minimum evolution, and neighbor-joining methods of phylogenetic tree construction in obtaining the correct tree. *Mol. Biol. Evol.* 6:514–25
72. Seperack, P., Slatkin, M., Arnhein, N. 1988. Linkage disequilibrium in human ribosomal genes: implications for multigene family evolution. *Genetics* 119: 943–49
73. Slatkin, M. 1985. Rare alleles as indicators of gene flow. *Evolution* 39:53–65
74. Slatkin, M. 1987. The average number of sites separating DNA sequences drawn from a subdivided population. *Theor. Pop. Biol.* 32:42–49
75. Slatkin, M. 1989. Detecting small amounts of gene flow from phylogenies of alleles. *Genetics* 121:609–12
76. Slatkin, M., Maddison, W. P. 1989. A cladistic measure of gene flow inferred from the phylogenies of alleles. *Genetics* 123:603–13
77. Slightom, J. L., Blechl, A. E., Smithies, 0. 1980. Human fetal G_γ- and A_γ-globin genes: complete nucleotide sequences suggest that DNA can be exchanged between these duplicated genes. *Cell* 21:627–39
78. Smouse, P. E., Long, J. C. 1988. A comparative F-statistic analysis of the genetic structure of human populations from lowland South America and highland New Guinea. In *Proceedings of the Second International Conference on Quantitative Genetics*, ed. B. S. Weir, G. Eisen, M. M. Goodman, G. Namkoong, pp. 32–47. Sunderland, MA: Sinauer Associates
79. Smouse, P. E., Vitzthum, V. J., Neel, J. V. 1981. The impact of random and lineal fission on the genetic divergence of small human groups: a case study among the Yanomama. *Genetics* 98: 179–97
80. Sneath, P. H. A., Sokal, R. R. 1973. *Numerical Taxonomy*. San Francisco: Freeman
81. Sokal, R. R., Harding, R. M., Oden, N. L. 1989. Spatial patterns of human gene frequencies in Europe. *Am. J. Phys. Anthropol.* 80:267–94
82. Spuhler, J. N. 1989. Raymond Pearl memorial lecture, 1988. Evolution of mitochondrial DNA in human and other

organisms. *Am. J. Hum. Biol.* 1:509–28
83. Stephens, J. C. 1985. Statistical methods of DNA sequence analysis: detection of intragenic recombination or gene conversion. *Mol. Biol. Evol.* 2:539–56
84. Stephens, J. C., Nei, M. 1985. Phylogenetic analysis of polymorphic DNA sequences at the Adh locus in *Drosophila melanogaster* and its sibling species. *J. Mol. Evol.* 22:289–300
85. Stoneking, M., Jorde, L. B., Bhatia, K., Wilson, A. C. 1990. Geographic variation in human mitochondrial DNA from Papua New Guinea. *Genetics* 124:717–33
86. Stringer, C. B., Andrews, P. 1988. Genetic and fossil evidence for the origin of modern humans. *Science* 39:1263–68
87. Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. *Genetics* 105:437–60
88. Tajima, F. 1990. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123: 585–95
89. Tajima, F., Nei, M. 1984. Note on genetic drift and estimation of effective population size. *Genetics* 106:569–74
90. Takahata, N., Nei, M. 1984. F_{ST} and G_{ST} statistics in the finite island model. *Genetics* 109:441–57
91. Takahata, N., Nei, M. 1990. Allelic genealogy under overdominant and frequency-dependent selection and polymorphism of major histcompatibility complex loci. *Genetics* 124:967–78
92. Tanaka, T., Nei, M. 1989. Positive Darwinian selection observed at the variable-region genes of immunoglobulins. *Mol. Biol. Evol.* 6:447–59
93. Tateno, Y., Nei, M., Tajima, F. 1982. Accuracy of estimated phylogenetic trees from molecular data. I. Distantly related species. *J. Mol. Evol.* 18:387–404
94. Tavare, S. 1984. Line-of-descent and genealogical processes, and their applications in population genetic models. *Theor. Pop. Biol.* 26:119–64
95. Waples, R. S. 1989. A generalized approach for estimating effective population size from temporal changes in allele frequency. *Genetics* 121:379–91
96. Watterson, G. A. 1989. Allele frequencies in multigene families. II. Coalescent approach. *Theor. Pop. Biol.* 35:161–80
97. Weir, B. S., Cockerham, C. C. 1984. Estimating F-statistics from the analysis of population structure. *Evolution* 38: 1358–70
98. Whittam, T. S., Clark, A. G., Stoneking, M., Cann, R. L., Wilson, A. C. 1986. Allelic variation in human mitochondrial genes based on patterns of restriction site polymorphism. *Proc. Natl. Acad. Sci. USA* 83:9611–15
99. Williams, S. M. 1990. The opportunity for natural selection on multigene families. *Genetics* 124:439–41
100. Wilson, A. C., Stoneking, M., Cann, R. L., Prager, E. M., Ferris, S. D. et al. 1987. Mitochondrial clans and the age of our common mother. In *Human Genetics*, ed. F. Vogel, K. Sperling, pp. 158–64. Berlin:Springer-Verlag
101. Wright, S. 1951. The genetical structure of populations. *Ann. Eugen.* 15:323–54
102. Wright, S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. *Evolution* 19:395–420
103. Zimmer, E. A., Martin, S. A., Beverley, S. M., Kan, Y. W., Wilson, A. C. 1980. Rapid duplication and loss or genes coding for the α chains of hemoglobin. *Proc. Natl. Acad. Sci. USA* 77:2158–62