Annu. Rev. Ecol. Syst. 1999. 30:23–49

USES OF EVOLUTIONARY THEORY IN THE HUMAN GENOME PROJECT

Alan R. Templeton
Department of Biology, Washington University, St. Louis, Missouri 63130-4899; e-mail: temple_a@biodec.wustl.edu

Key Words complex traits, gene trees, genetic disease, genetic architecture, population structure

■ **Abstract** The Human Genome Project (HGP) originally sought to sequence the human genome but excluded studies on genetic diversity. Now genetic diversity is a major focus, and evolutionary theory provides needed analytical tools. One type of diversity research focuses on complex traits. This is often done by screening genetic variation at candidate loci functionally related to a trait followed by gene/phenotype association tests. Linkage disequilibrium creates difficulties for association tests, but evolutionary analyses using haplotype trees can circumvent these problems and result in greater statistical power, better disease risk prediction, the elimination of some polymorphisms as causative, and physical localization of causative variation when combined with an analysis of recombination. The HGP also now proposes to map over 100,000 single nucleotide polymorphisms to test for gene/phenotype associations through linkage disequilibrium in isolated human populations affected by past founder or bottleneck events. This strategy requires prior knowledge of recent human evolutionary history and current population structure, but other evolutionary considerations dealing with disequilibrium and nonrandom mutation pose difficulties for this approach. Studies on population structure also focus upon traits of medical relevance, and an understanding of the evolutionary ultimate cause for the predisposition of some populations to certain diseases is a useful predictor for shaping public health policies. Studies on the genetic architecture of common traits reveal much epistasis and variation in norms of reaction, including drug response. Because of these interactions, context dependency and sampling bias exist in disease association studies that require population information for effective use. Overall, the population thinking of evolutionary biology is an important counterweight to naive genetic determinism in applying the results of the HGP to issues of human health and well-being.

INTRODUCTION

The Human Genome Project (HGP) began in 1990 with the goal of sequencing an entire human genome by the year 2005. This project also included the goal of obtaining genome sequences of several other organisms, many of which have

already been completely sequenced (17). The sequencing of genomes from several species obviously creates opportunities for evolutionary studies in comparative genomics. However, these macroevolutionary studies are outside the scope of this review, which is restricted to studies solely on humans. At first glance, this restriction may seem to preclude any relevance of evolutionary biology to the HGP. This perception is reinforced by the fact that the initial HGP excluded studies on human genetic diversity (16, 17)—the raw material of all evolutionary phenomena at the intraspecific level. Partly to remedy this perceived deficiency of the HGP, a Human Genome Diversity Project was first proposed in 1992, but it never became an integrated, focused project (118). Meanwhile, sequencing technology advanced such that in May 1998, private initiatives were announced to sequence much of the human genome de novo within just three years and at a tenth of the cost of the federally funded HGP (105). Following the announcements of these private initiatives, a major reassessment of the goals of the HGP was undertaken (17). The HGP now has a "new focus on genetic variation" because such variation is the "fundamental raw material for evolution" and "is also the basis for variations in risk among individuals for numerous medically important, genetically complex human diseases" (17). The HGP has finally opened the door to evolutionary biology.

Because of the deliberate exclusion of studies on human genetic diversity until 1998, few of the papers cited in this review were funded by the HGP. Nevertheless, many non-HGP studies have examined human genetic diversity at the molecular level, often with a clinical or medical objective. This review includes such projects because they now fall within the new goals of the HGP and show the contribution of evolutionary principles to such research.

DISEASE ASSOCIATIONS

General Considerations

The primary justification of the HGP is its promise to yield new insights into and possibly cures for a variety of human diseases. Mapping and understanding the "simple Mendelian diseases" of humanity is part of this goal. However, only about 1.25% of human births are affected by such Mendelian diseases (25). In contrast, 65.41% of live births are affected at some time in their life by a disease that has a heritability greater than 0.3 (25). Accordingly, the new goals of the HGP include a major focus upon the "genetically complex human diseases" (17). This review is limited to genetically complex diseases because evolutionary theory and practice can provide critical analytical tools for their study.

The primary impact of molecular genetics upon studies of complex diseases has been a strong shift from unmeasured to measured genotype approaches (126). In the unmeasured approach, individual phenotypes are scored and pedigree information is obtained to define the genetic relationships among the individuals. The

phenotypic and pedigree information is then used to obtain estimates of classical quantitative genetic parameters such as heritabilities. In contrast, measured genotype approaches start with some assay of genetic variation (usually at the molecular level) and test for associations between measured genetic variation and phenotypic variation. Measured genotype approaches predate the HGP (e.g. 45, 127), but the ready availability of molecular markers throughout the human genome has now made measured genotype approaches increasingly common. There are three major measured genotype approaches in human genetics: candidate loci, quantitative trait loci (QTL) mapping through linkage associations within families, and QTL mapping through linkage disequilibrium in populations.

Candidate Loci

With the candidate locus approach, prior information about the basis of the phenotype and the function of known genes is used to identify loci that are likely to contribute to the phenotype of interest (126). For example, after age and sex, various lipid traits are the major risk factors for coronary artery disease (CAD), which accounts for about a third of the total human mortality in developed nations. The etiology of CAD involves the deposition of cholesterol-laden plaque onto the interior walls of the coronary arteries. Many genes have been identified that play critical roles in lipid metabolism, so these are all candidate genes for CAD (6, 126). Our knowledge of the etiology of many diseases and the function of specific genes is increasing at a rapid rate, so the candidate approach will play an increasingly important role in human genome studies.

With this approach, a population is screened for variation at the candidate loci followed by tests for associations between genetic and phenotypic variation. With low-resolution genetic screens such as protein electrophoresis, such tests of association were relatively straightforward by measuring mean phenotypes of genotypes. For example, one of the candidate genes involved in lipid metabolism is *Apoprotein E* (*ApoE*), and protein electrophoresis reveals three alleles that are common in most human populations, the $\varepsilon 2$, $\varepsilon 3$, and $\varepsilon 4$ alleles. In most human populations, the $\varepsilon 2$ allele is associated with lowered total serum and LDL (low density lipoprotein) cholesterol levels, and $\varepsilon 4$ is associated with elevated levels (24, 43, 120, 121, 126). Prospective studies confirm that individuals bearing the $\varepsilon 4$ allele die disproportionately from CAD (130). The marginal effects of *ApoE* alone account for about 7% of the total phenotypic variation in cholesterol levels and about 14% of the total genetic variance (121). Hence, *ApoE* is considered to be a major locus for CAD risk prediction.

As genetic resolution increases, certain difficulties arise. This is illustrated by studies on another candidate gene for CAD, the *lipoprotein lipase* locus (*LPL*): 71 individuals from three populations were sequenced for a 9.7-kb region within this locus (98). Of the 88 polymorphic sites were discovered, and 69 of these sites had their phases determined to define 88 distinct haplotypes (14). Thus, using

only a subset of the known polymorphic sites in just a third of a single gene, a sample of 142 chromosomes reveals 88 "alleles" or haplotypes that in turn define 3916 possible genotypes—a number considerably larger than the sample size. Obviously, the increased genetic resolution now poses some daunting statistical challenges.

One simple and common approach for dealing with such vast arrays of genetic variation is to analyze each polymorphic site separately. However, polymorphic sites within a candidate locus are, virtually by definition, tightly linked and often show strong linkage disequilibrium, as is the case for *LPL* (14, 98). As a consequence, the multiple single-site tests are not independent from one another, making statistical and biological interpretation difficult. Most commonly, such single-site approaches are interpreted as a minilinkage analysis; each polymorphic site is treated as a "marker," and those markers that show the strongest phenotypic associations are regarded as being physically closest to the causative site or being the causative site. For example, *ApoE* is also a candidate gene for Alzheimer's disease, and the *ε4* allele often shows a significant marginal association with sporadic Alzheimer's disease (1, 8, 10, 13, 19, 56, 78, 104, 107), although not in all cases (68, 70, 75). This discovery has led to a plethora of papers trying to explain how this specific allele "causes" Alzheimer's disease (80, 93, 112, 132, 157). However, three other apoprotein loci are tightly linked to the *ApoE* locus (2), and the genetic markers throughout this chromosomal region show strong linkage disequilibria, apparently due to low rates of recombination (135). The question therefore arises: Which of these functionally related loci harbors causative variation?

At this point, it is necessary to consider the evolutionary forces that determine the amount and pattern of linkage disequilibrium. A standard equation from population genetics is:

$$D_t = D_0(1 - r)^t \qquad 1.$$

where D_t is the linkage disequilibrium between two sites at generation t, r is the recombination frequency between the sites, and D_0 is the linkage disequilibrium at the initial generation 0. Attention is often focused upon r in Equation 1, leading to the general belief that the magnitude of linkage disequilibrium can be used as a proxy for recombination frequency. However, when dealing with small regions of DNA, recombination is often extremely rare (but not always; 14). When recombination is rare, the magnitude of disequilibrium is determined primarily by the chromosomal background upon which a mutation initially occurred and its subsequent evolutionary fate (137, 139). Therefore, the magnitude of disequilibrium in many candidate regions reflects the temporal positioning of mutational events over evolutionary time, and not the spatial positioning over a physical region of DNA. Indeed, many studies reveal little to no correlation between linkage disequilibrium and physical distance within small regions of DNA (48, 50, 61, 74, 76, 145, 165, 166). This lack of correlation means that there is no reliable positional information even when a strong association is found only with

a single marker in a candidate region. Hence, it is premature to infer that *ApoE* is the only plausible locus for "causing" Alzheimer's disease in this chromosome region and that the mutation leading to the electrophoretic *ε4* allele is a causative mutation. Multiple single-marker association studies in a candidate region are statistically indefensible because of non-independence and easily misinterpretated biologically unless the evolutionary basis of disequilibrium is taken into account.

One way to eliminate the problems caused by disequilibrium is to organize all the genetic variation in the DNA region into haplotypes and then use haplotypes as the units of analysis (149, 150). However, as noted above for *LPL*, the number of haplotypes is often still so large that even extremely large samples will only provide sparse coverage of the possible genotypic space defined by haplotypes. Evolutionary history can be used to solve this problem. Given a candidate gene region in which there has been either no or little recombination, it is possible to estimate a haplotype tree that reflects the evolutionary history that generated the disequilibrium and the haplotypes (139). The estimation of such haplotype trees can be accomplished by standard phylogenetic inference procedures, but many of these procedures were originally worked out for interspecific phylogenies and make assumptions that are inappropriate for haplotype trees (22, 140). Accordingly, new procedures have been developed for estimating and testing haplotype trees (21, 22, 139, 141).

A simple example of the use of a haplotype tree is provided by a study of the *ApoAI-CIII-AIV* gene cluster (49). Seven haplotypes (numbered *1* through *7*) defined by restriction site polymorphisms were found in a population of 147 French Canadians scored for total serum cholesterol and other lipid phenotypes. Of these individuals 140 bore at least one copy of haplotype *6* and therefore only one haplotype state varied among these 140 individuals. Thus, the genotypes in this subset could be analyzed as if they were haploid. The haplotype tree in this case is a simple star in which the six rarer haplotypes radiate from the central and common haplotype *6*, with all six of the tip haplotypes differing from haplotype *6* by only a single restriction site. Note that there are 21 different pairwise contrasts among these 7 haplotypes, but there are only 6 degrees of freedom. In general, there are $n-1$ degrees of freedom in a haploid analysis of n haplotypes and $\frac{1}{2}n(n-1)$ pairwise contrasts, so this discrepancy gets worse with increasing haplotype diversity. One of the basic principles of the comparative method of evolutionary biology is that the most meaningful contrasts are between evolutionary neighbors. Applying this principle, the evolutionarily relevant contrasts are those between haplotype *6* and the six others. Accordingly, only these six pairwise contrasts were performed, each using one of the available degrees of freedom, and several significant phenotypic associations were discovered (49). In contrast, no significant phenotypic associations were discovered with a traditional, one-way ANOVA that does not use evolutionary principles to concentrate statistical power upon the most relevant comparisons (49).

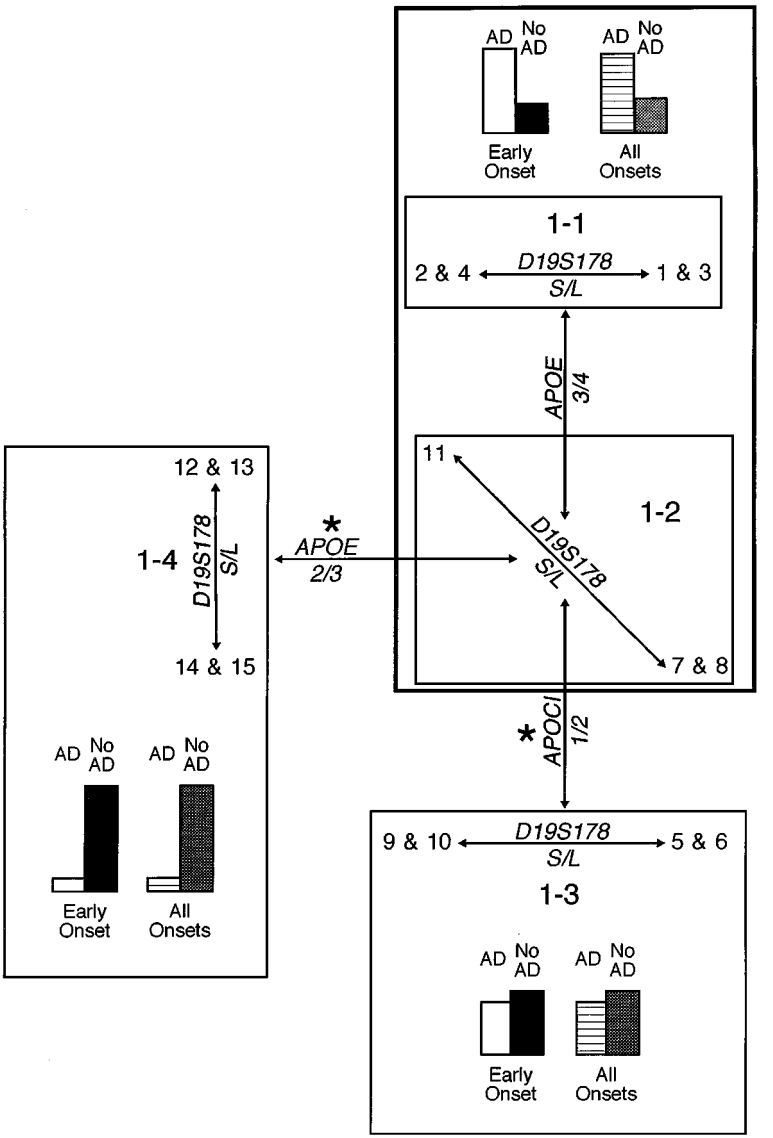
Most haplotype trees deviate from the pure star topology. Still, any haplotype tree of n haplotypes will have $n - 1$ connections in a fully resolved tree, so in theory all haplotype trees could define contrasts across their branches that would fully use all available degrees of freedom and no more. Statistical problems can still exist because many haplotypes, particularly those on the tips of the haplotype tree, are expected to be quite rare in the population (22), resulting in little power. This problem can be diminished by pooling haplotypes into larger categories, particularly tip haplotypes. One method of pooling is to start with tip haplotypes (in general the rarest) and move one mutational step toward the interior of the haplotype tree, pooling together all the tip and interior haplotypes that are reached by such a movement into a "1-step clade" (139). The resulting initial set of 1-step clades on the tips of the haplotype tree are then pruned off, and the same procedure is then repeated upon the remainder of the haplotype tree until all haplotypes have been placed into 1-step clades. One then applies the same nesting algorithm to the tree of 1-step clades to create 2-step clades. This procedure is iterated until all clades are nested into a single category. Special rules of nesting are needed to deal with symmetries in the tree and regions of topological ambiguity (141). This nesting procedure has several advantages. First, nesting categories are determined exclusively by the evolutionary history of the haplotypes and not by a phenotypic pre-analysis, thereby eliminating a major source of potential bias. Second, the clades define a nested design that makes full and efficient use of the available degrees of freedom. Third, statistical power has been enhanced by contrasting pooled clades across many of these branches instead of individual haplotypes (124, 125, 136, 139).

Most human studies involve diploid genotypes. The nested clade statistics were therefore extended to deal with diploid populations through the quantitative genetic device of the average excess of a haplotype (134, 143). The statistical significance of these average excesses can be determined by a variety of methods. For example, a nonparametric nested permutational analysis of the *ApoAI-CIII-AIV* region revealed a significant association with the log of the serum triglyceride level, another risk factor for CAD that was undetectable with a standard nonevolutionary analysis (143). Alternatively, likelihood-ratio tests of linear models of parameterized haplotype effects were used to identify clades of haplotypes in the *Apoprotein B* (*ApoB*) region that explained about 10% of the genetic variance and 5% of the total variance in HDL-cholesterol and triglyceride levels (44). None of the sites used to define these *ApoB* haplotypes showed any significant phenotypic associations when tested individually (44). As these examples show, evolutionary analyses provide greater statistical power than nonevolutionary alternatives.

The evolutionary approach also results in better individual level risk prediction over the single-marker approach. Analyzing one marker at a time is inherently a bivariate analysis, but phenotypes typically come in more than two categories. When this occurs, single-marker analyses are inherently incapable of detecting the full range of phenotypic heterogeneity. For example, a single marker analysis of case-control data concluded that the *ApoE* $\varepsilon 4$ allele and a restriction site marker

in the tightly linked *ApoCI* locus can each divide people into higher and lower Alzheimer's disease risk categories (10). The nested evolutionary analysis of the same data (Figure 1) identifies the *ApoCI* marker as being associated with a significant change in Alzheimer's risk, as is *ε2* (but not *ε4*) at the *ApoE* locus, with the *ε4* marginal effect being due to linkage disequilibrium (135). Thus, both approaches identified the *ApoCI* marker as providing information about Alzheimer's risk that is independent from the *ApoE* locus markers, but the evolutionary analysis results in three risk categories (Figure 1), not two (135). The "high-risk" category in the marginal analysis of the *ApoCI* marker contains those individuals with both lowest risk (clade 1-4 in Figure 1) and highest risk (clades 1-1 and 1-2). Hence, using the *ApoCI* marker by itself would mean that many people with the lowest risk for Alzheimer's disease would incorrectly be advised that they are in a high-risk category. Thus, evolutionary analyses provide greater precision of risk prediction for individuals than nonevolutionary analyses of the same data.

Another advantage of the evolutionary approach is that it provides guidance for studies attempting to find causative variation. For example, a nested clade analysis of restriction-site haplotypes at the *low-density lipoprotein receptor* locus (*LDLR*) localized a significant phenotypic change in CAD risk factors to a branch between two haplotypes (designated *H1* and *H5*) defined by a single *Taq*I restriction site change (46, 47, 124). This *Taq*I site also had a significant marginal association, so a non-evolutionary analysis detects this association as well. However, because there is no reliable relationship between physical distance and degree of association in such small gene regions, the non-evolutionary analysis gives no guidance as to the nature or location of the causative site or sites. Similarly, in the nested clade analysis, the mutational change/s that define the branch in the haplotype tree that is associated with a phenotypic change are not necessarily causative (139). Therefore, neither analysis justifies equating the *Taq*I site to the causative site. However, a comparison of individuals with and without the *Taq*I site would include much more background genetic variation than the more refined comparison of bearers of *H1* versus *H5* (47). Hence, the evolutionary analysis minimizes the differences in background variation in favor of candidate functional variations (47). More detailed genetic surveys can then be performed upon just these two haplotypes to execute a more refined nested analysis to eliminate additional sites as being causative (47, 123). For example, a nested clade analysis of haplotypes in the *ApoB* region revealed a clade of haplotypes defined by an *Xba*I restriction site polymorphism that explained 6.5% of the phenotypic variance in reduction of total cholesterol and 22.3% of the phenotypic variance in reduction of LDL-cholesterol (the "bad" cholesterol for CAD risk) in response to placing 63 male students upon a controlled low-cholesterol diet (38). It was hypothesized that the putative LDL receptor binding region of apo B might determine how individuals respond to low-cholesterol diets, so this region was sequenced in individuals homozygous for the two haplotype categories on either side of the *Xba*I branch in the *ApoB* haplotype tree. The sequence analysis revealed no differences in this region despite these haplotypes being associated with marked differences in


AD No AD
Early Onset
AD No AD
All Onsets
1-1
2 & 4
D19S178
S/L
1 & 3
APOE
3/4
11
1-2
D19S178
S/L
7 & 8
12 & 13
1-4
D19S178
S/L
14 & 15
*
APOE
2/3
*
APOCI
1/2
9 & 10
D19S178
S/L
5 & 6
1-3

dietary response (38). Thus, the causative sites are not in the putative binding region.

The study on dietary response demonstrates that the candidate locus approach can be applied to many types of traits, including responses to environmental variables. The idea that a single genotype can give rise to a variety of phenotypes depending upon environmental conditions (phenotypic plasticity) and the idea that different genotypes will display different phenotypic responses to the same environmental change (norm of reaction) have long played an important role in evolutionary biology (115). Although it has long been known that humans display different genetically determined reactions to environmental variables [e.g., the genetic polymorphisms for lactose intolerance (35) or drug response (72, 87, 90, 159)], a widespread appreciation for the importance of these evolutionary concepts is only recent in human genetics. This appreciation was slow in developing because drug and food metabolism are sufficiently redundant and complex as to mask the effects of single-locus variants with classic Mendelian studies (83). However, there are a plethora of candidate loci in this area, so the increased feasibility and power of the candidate locus approach is revolutionizing this field (60, 83). Indeed, "pharmacogenomic" companies have been founded in recent years with the expressed purpose of developing haplotype data at candidate loci that would be amenable to evolutionary lineage analysis in predicting drug response (26).

Any evolutionary analysis of haplotype variation is based upon the presumption that a haplotype tree can be constructed in the first place. If recombination is frequent and uniform in a candidate region, a nested clade analysis would be impossible. However, the presence of a few recombinants can actually enhance biological inference by providing information about physical location of causative sites (141). For example, the *angiotensin-I converting enzyme* (*ACE*) locus is a candidate gene for hypertension and CAD. Two major clades (*A* and *B*) explain 36% of the phenotypic variance for the amount of circulating ace levels (59). A third clade (*C*) was inferred to have arisen from a recombination event between clades *A* and *B*, and this recombinant clade had phenotypic effects indistinguishable from those of clade *B* (59). Given that clade *C* received its 3′ end from clade *B*, the causative sites or sites must be 3′ to the inferred crossover event, thereby

←

Figure 1 Nested haplotype analysis of frequency of Alzheimer's disease in a case (AD)-control (no AD) study. Haplotypes are indicated by the numbers 1–15. Each arrow in the tree indicates a single mutational change, with the nature of the change indicated by the arrow. Microsatellite variation at the marker *D19S178* was simply characterized as short (*S*) or long (*L*) alleles, and transitions between these two allelic classes could not be ordered in the tree. Four 1-step clades are defined by nesting, indicated by 1-1 through 1-4. Two significant transitions were identified in the tree, as indicated by asterisks, that subdivide people into three categories. The relative incidences of early onset and all onset Alzheimer's disease is indicated for each of these three categories. Modified from Templeton (135).

excluding the 5′ region and the *ACE* promoter from harboring the causative variation (59).

A nested clade analysis is also possible in DNA regions with frequent recombination as long as the recombination is concentrated into a hotspot. For example, 30 statistically significant recombination/gene conversion events were inferred to have occurred in the 9.7-kb region of the *LPL* gene sequenced in 71 individuals from three populations, but 24 of these 30 events were concentrated into intron 6 in the *LPL* gene (AR Templeton, AG Clark, KM Weiss, DA Nickerson, E Boerwinkle, CF Sing, —139a). As a consequence, there is much phylogenetic structure in the regions flanking this recombinational hotspot.

Given the improvements in sequencing technology, data sets such as *LPL* should become more common in the future. Such data will provide the abundance of markers that are necessary to infer both recombination events and highly resolved haplotype trees. Although rigorous statistical analyses of haplotype trees are still relatively rare, more informal analyses of haplotype trees that make use of evolutionary relatedness are becoming increasingly common in human disease association studies (3, 53, 54, 71, 84, 151). Hence, the new goals of the HGP should be a boon to evolutionary analysis of candidate gene regions.

QTL Mapping Through Linkage Associations Within Families

Under this approach, the measured genotypes are used as markers of physical location within the genome, and phenotypic associations are detected through linkage. This methodology is not discussed in this review because it is an application of quantitative genetics and not evolutionary genetics per se, although it certainly interfaces with evolutionary genetics. However, such QTL studies do have implications for genetic architecture, as is discussed later.

QTL Mapping Through Linkage Disequilibrium in Populations

One of the major new goals of the HGP is to develop a map of over 100,000 single nucleotide polymorphisms (SNPs) distributed over the entire human genome. Although the new HGP goals include developing technologies for scoring any type of polymorphism (17), SNPs are the initial focus because they are abundant and amenable to rapid and efficient screening of large samples (16, 17, 156). The justification for creating such a map is the expectation that linkage disequilibrium, at least in relatively isolated, homogeneous human populations, should exist between at least one of these SNPs and the common alleles at any locus in the genome.

As pointed out earlier, linkage disequilibrium is commonly found among closely linked polymorphisms in regions with infrequent recombination. The magnitude and physical extent of likely disequilibria values can be augmented by founder events that establish relatively isolated populations. Indeed, the first success in positional cloning of a human genetic disease gene, Huntington's chorea, depended critically upon first identifying a relatively isolated founder population that was polymorphic for this disease (42). Such relatively isolated populations

also diminish the importance of genetic background effects, making it more likely to detect marginal effects of alleles linked to a SNP. However, linkage disequilibrium can be induced by population subdivision even between unlinked loci (73). Hence, if a sample includes individuals drawn from more than one subpopulation, disequilibria could exist with many markers regardless of their linkage relationships, undermining the SNP strategy. However, if the population subdivisions are known a priori and admixture is occurring, this information can be used in disequilibrium mapping (131). It is therefore critical in SNP studies to either sample within a relatively panmictic and homogeneous population or to use a known admixture situation. In either case, analyses of the evolutionary history and genetic structure of human populations are necessary prerequisites in designing SNP studies. Accordingly, human populations thought to have undergone founder events within the last several centuries and that have been relatively isolated since are now extremely valuable resources for association studies of both classical genetic diseases and the more common complex diseases (28, 41, 119). Indeed, such relatively isolated human populations are now valuable commercial entities. For example, a private company, deCODE, is seeking legislation that will give the company a 12-year license to operate a database containing the medical records of the entire population of Iceland and sell access to third parties (32, 41).

Although evolutionary considerations have gone into designing the SNP strategy, other evolutionary considerations undermine the claim that "this strategy should in theory reveal the identity of the gene or genes underlying any phenotype not due to a rare allele" (16). Given that the current strategy has about one SNP per gene, this claim requires that a single SNP marker near or in the locus of interest would show significant disequilibrium with nearly all common alleles at this locus. As pointed out earlier, not all marker pairs display significant disequilibrium within a gene region, and the disequilibrium that exists is not well correlated with physical distance. Therefore, it is doubtful if a single SNP would display disequilibrium with all common polymorphic sites within a gene. Moreover, a significant phenotypic association with a SNP is not a reliable guide to the actual location of the gene causing the phenotypic variation because of the physical clustering of functionally related genes in the genome. This problem is only accentuated when dealing with founder events because chance events (genetic drift) play an important role in creating disequilibrium even between markers that are well separated physically.

Another problem with randomly chosen SNP's is the danger of selecting a highly mutable site. Such sites will show complicated patterns of association due to the fact that identity by state does not reflect identity by descent. It is commonplace in much of the human genetic literature to assume that the infinite sites model (which does not allow multiple hits) is appropriate for nuclear DNA (e.g. 155). At first the case for the infinite sites model seems compelling. For example, only 88 sites out of 9734 are polymorphic in the sequenced region of *LPL* (98)—a figure that seems to be well below saturation. Moreover, almost all these 88 polymorphic sites are biallelic, which seemingly further bolsters the argument against multiple hits (14). However, both of these arguments are based upon the premise that

mutations are equally likely to occur at all sites and, given a mutation, that any of the three nucleotide states are likely to arise. Neither of these premises is justified by the human genetic literature. For example, about a third of all mutations in human nuclear DNA are transitions from 5-methylcytosine to thymine that occur exclusively at CpG dinucleotides, a combination markedly underrepresented in human DNA (55, 64, 81, 110, 116, 164). Mutational hotspots have also been reported for mononucleotide-repeat regions, DNA polymerase α arrest sites, and other rarely occurring sequence motifs in human DNA (63, 91, 147, 148). The pattern of site polymorphism in the *LPL* gene parallels the results of these mutation studies, with 9.6% of the nucleotides in CpG sites being polymorphic, 3.3% of the nucleotides in mononucleotide runs of length 5 or greater, 3.0% of the nucleotides within three base pairs of the polymerase α arrest site motif of TG(A/G)(A/G)GA, and 0.5% at all other sites (AR Templeton, AG Clark, KM Weiss, DA Nickerson, E Boerwinkle, CF Sing, unpublished observations). Altogether, almost half of the polymorphic sites in the sequenced portion of the *LPL* gene were from one of these three highly mutable classes and therefore would be less than ideal choices for a SNP marker. This consideration further dims the likelihood of a randomly chosen SNP providing adequate coverage of a gene.

The nonrandom nature of mutation means that the infinite sites model is not applicable to the human nuclear genome. If the patterns observed at the *LPL* locus hold true for other loci, a model that is a mixture of the finite and infinite sites model would be more appropriate, but such a mixture is qualitatively more similar to a finite sites model (which allows multiple hits) than to an infinite sites model (which forbids multiple hits). This is a serious implication because many of the standard descriptors and test statistics for looking at genetic diversity, effective sizes, recombination, etc, are based upon the infinite sites model, primarily because of its greater mathematical simplicity. Consequently, these human genome studies constitute a major challenge to population geneticists. New analytical and statistical models need to be developed to advance our understanding of the nature of genetic diversity within the human genome.

HUMAN POPULATION STRUCTURE

As discussed above, the genetic structuring of human populations is now medically and commercially valuable knowledge that is essential for meeting the new goals of the HGP. Obviously, the SNP surveys themselves will greatly augment our knowledge of human population structure. Therefore, the study of human population structure and recent evolutionary history is and will remain a major focus of the HGP. There is also a great need to understand the patterns of differentiation among human populations for clinically relevant phenotypes and the candidate loci or chromosomal regions associated with them. For example, the incidence of CAD varies tenfold among populations (129). As previously noted, the *ε4* allele of the *ApoE* locus is a major predictor for individual CAD-risk and cholesterol levels within several human populations despite much variation in the population

incidence of CAD (43). The *ε4* allele frequency varies from 0.07 to 0.20 and explains 75% of the variation in CAD mortality rate among the populations, with higher *ε4* allele frequencies being found in those populations with higher CAD (129). However, discrepancies exist between the inter- versus intrapopulational estimates of the effects of the *ε4* allele (129), and studies that integrate the intra- and interpopulation components of the sort already performed in evolutionary genetics (152–154, 160) are obviously going to be needed in examining the dual roles of genetic variation as sources of phenotypic variability within and among populations.

As the above studies illustrate, human populations differ greatly in the incidence of common diseases that have a strong underlying genetic component (158). This component of human population structure has obvious public health importance (52, 117). Evolutionary biology is an important tool for predicting which human populations are at greatest risk for certain diseases because it can provide insights into the ultimate reason why a particular disease is common (158). For example, the insight that sickle-cell anemia (associated with the *S* allele at the *β-globin* locus in humans) is an adaptation to malaria immediately leads to the prediction that *S* alleles should be found in high frequencies in populations that live now or in the recent past in malarial regions. Although sickle cell is commonly portrayed as a disease associated with one "race" of humanity, the evolutionary prediction is far more accurate because the *S* allele is found in high frequency in some, but not all, sub-Saharan African populations, populations in the Mediterranean and Middle East, and populations in India (7, 31, 100, 109, 114). Sickle cell is not cleanly associated with any "race" but rather is more strongly associated with the presence of malaria, its selective agent. Similar considerations hold for the more common diseases. For example, certain human populations, scattered widely throughout the world, have extremely high incidences of adult-onset diabetes (158). Neel (94) proposed an evolutionary explanation for this complex pattern; namely, that populations exposed to prolonged low-calorie diets or periodic famines would be selected for "thrifty genotypes," and these same thrifty genotypes when placed into an environment with a high-calorie diet would be predisposed to diabetes. Recent studies have supported the thrifty genotype hypothesis (5, 29, 94a, 106); with increasing knowledge about the genes responsible for adult-onset diabetes (27, 88, 106, 146, 149), this hypothesis can be tested with greater rigor than hitherto possible. Evolutionary explanations have been proposed for many other diseases in humans (51, 94a, 96, 117, 158), and studies testing these evolutionary hypotheses are needed in order to exploit more fully the public health predictions possible through an evolutionary explanation of ultimate causation.

GENETIC ARCHITECTURE

The genetic architecture of a trait refers in part to the number of loci and their genomic positions, and the number of functional alleles per locus that influence the trait. Genetic architecture also includes the patterns of dominance, epistasis, pleiotropy, and gene-by-environment interactions that characterize the transition

from genotype to phenotype. The relative frequency of different types of genetic architecture constitutes one of the most longstanding debates within evolutionary genetics. Fisher (34) argued that mutations of large phenotypic effect would mostly be deleterious and thereby would be rapidly eliminated from populations. Fisher also argued that the underlying genetic architecture of most traits should consist of a large number of alleles and loci of small phenotypic effect with epistasis playing an unimportant role. Wright (161) also regarded genetic architecture as generally involving multiple loci, but he argued that the effect of any single allele is highly dependent upon its genetic and environmental context, thereby making the distinction between minor polygenes and "major" genes (66) a false one. Wright also regarded epistasis and pleiotropy as "universal" phenomena and emphasized the importance of gene-by-environment interactions and the dynamics of gene expression throughout development and the aging process (162). This debate has continued to the present day (33), in part because the nature of genetic architecture plays an important role in evolutionary models of adaptation (34, 161) and speciation (133).

The new HGP research focus on complex, common traits contributes directly to the debate about genetic architecture. In turn, the resolution of this evolutionary debate has profound implications for the HGP. All of the approaches discussed in the previous section can provide information on genetic architecture. The linkage and linkage-disequilibrium marker approaches can identify some of the loci that influence the trait of interest. These studies clearly show that not all loci are minor in their phenotypic contributions (9, 12, 15, 18, 20, 27, 39, 65, 86, 88, 146) and that the marginal effects of specific chromosomal regions are highly context dependent (to be discussed shortly).

Although epistasis can create difficulties for the linkage marker approach (30), epistasis can be tested in QTL studies (11, 67, 111, 163). However, several biases exist against the detection of epistasis. First, many QTL studies only search for epistasis among markers or chromosomal regions that first show significant marginal effects, but significant epistasis can exist between marked chromosomal regions with no marginal phenotypic effects (142). Second, most searches for epistasis only deal with pairwise interactions, but additional epistasis can exist at higher levels (142). Third, and most importantly, epistasis is not detected because no effort is made to discover it (37). Despite these biases, there are already several QTL studies that report significant epistasis for some common human phenotypes (12, 15, 20).

In principle, the biases against discovering epistasis are less severe when dealing with the candidate locus approach because the loci are chosen on the basis of prior knowledge rather than a significant marginal effect. However, in practice most candidate loci studies do not look for epistasis (37), and those that do are generally limited to pairwise interactions among candidate loci having significant marginal effects. Nevertheless, much epistasis has been detected with the candidate loci approach for many common human phenotypes (4, 40, 57, 58, 62, 69, 77, 79, 89, 92, 99, 101–103, 108, 113, 122, 126, 144).

The study that has best used the potential of the candidate loci approach to avoid bias in the detection of epistasis is that of Nelson et al (95), although even this study retains the bias of only examining pairwise interactions. Nelson et al (95) surveyed genetic variation at 18 markers in six candidate gene regions in a human population scored for the trait of log-transformed triglyceride level, a risk predictor of CAD. Nelson et al (95) used an extensive computer search called the combinatorial partitioning method to identify the pairwise combinations of markers that explained a significant proportion ($\alpha = 0.01$) of the observed phenotypic variance. Because of the immense number of tests, the significant effects identified by the initial search were subjected to tenfold cross-validation. This process was carried out with no a priori model of gene action. The pairwise markers that best explained phenotypic variation in females and males are shown in Figure 2, along with the percent of the phenotypic variation explained by the pair jointly and their single-locus marginal effects. Obviously, epistasis is of critical importance for this lipid trait. The included loci would clearly be overlooked in a single-locus analysis in both females and males.

As more studies of this nature are performed under the HGP, this long-standing debate over genetic architecture should become increasingly resolved. It is already clear that interactions among loci are important in many clinically relevant

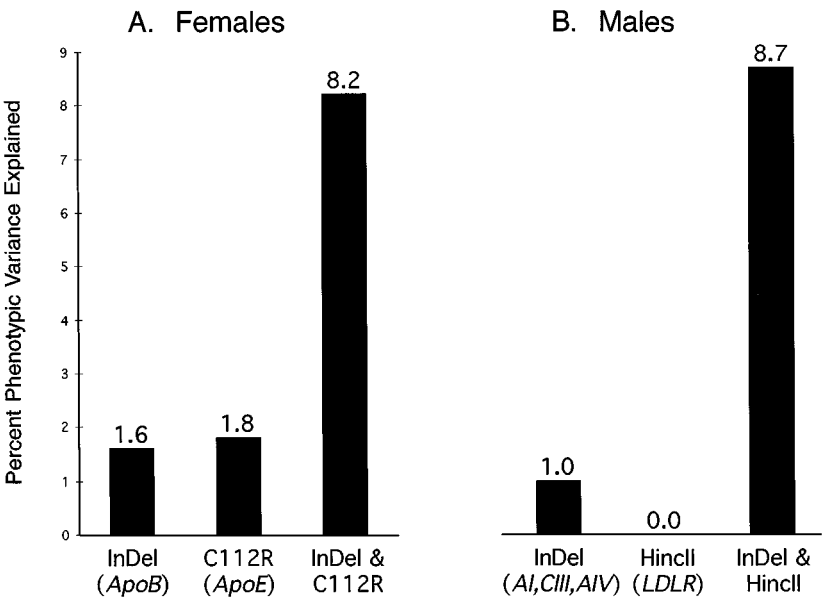


Figure 2 Percent of phenotypic variance in the natural logarithm of serum triglyceride levels in females (*a*) and males (*b*) that explained best by pairs of markers in candidate genes as well as the single marker contributions. Modified from Nelson et al (95).

phenotypes (37). Also, as pointed out before, interactions between genes and environments are important. Together, these interactions imply that the effects of any single locus could display extreme context dependency. This context dependency has profound implications for the study of disease and health under the HGP.

There are now many cases in which one research group identified a particular locus or DNA region as having a strong association with a medically important phenotype only to have another group find no association with that locus or region (23). Such "disappearing" genes occur too frequently to be explained by multiple false positives (23). Context dependency, due to either epistasis or gene-by-environment interactions, provides an explanation for this phenomenon. To illustrate this, consider an example based upon the strong documented epistasis between the *ApoE* and *LDLR* loci such that individuals tend to have elevated serum cholesterol levels only if they have the *ε4* allele at *ApoE* and are homozygous A_2A_2 at the *LDLR* locus (102, 103). Figure 3(*a*) shows a quantitative genetic partitioning of the phenotypic variance in this case using allele frequencies typical of Western European populations: 0.152 for the *ε4* allele and 0.78 for A_2 (138). Note that most of the epistasis in this system is converted to "additive variance" at *ApoE*, a type of conversion expected in epistatic systems (11). Figure 3(*b*) shows the results from the same calculations, but now using allele frequencies of 0.95 for *ε4* and 0.5 for A_2. Although the genotype/phenotype relationships are identical in 3(*a*) and 3(*b*), *ApoE* emerges as a "major" gene and *LDLR* is a "minor" gene in 3(*a*), but the reverse is true in 3(*b*). This illustrates a well-known but little appreciated property of complex systems in which interactions are the true agents of causation and not individual components; namely, there is a confoundment between the frequency of an element and apparent marginal causation at the population level. In particular, the rarer components of the interactive system appear to have the stronger marginal effects, as illustrated in Figure 3. Rareness and commonness are properties of populations, not genes or genotypes. Hence, no matter how elegant the molecular biology, disease associations must always be interpreteted at the population level.

Context dependency is particularly important in medical research because the populations studied are often highly nonrandom samples from the general population. For example, in families with four or more cases of breast cancer, certain mutations at the *BRCA 1* locus are predictive of 52% of the cases of breast cancer, and this increases to 76% of the cases in those families that have both male and female breast cancer (36). When one ascertains families by their having only two cases of breast cancer before the age of 45, *BRCA 1* mutants are only predictive of 7.2% of the cases (82), and when one looks at sporadic cases (the bulk of the disease in the general population), only 1.4% of women with breast cancer have *BRCA 1* mutants (97). Obviously, recommendations for prophylactic mastectomies based on *BRCA 1* testing need to be made in a highly context-dependent fashion (see also 108a).

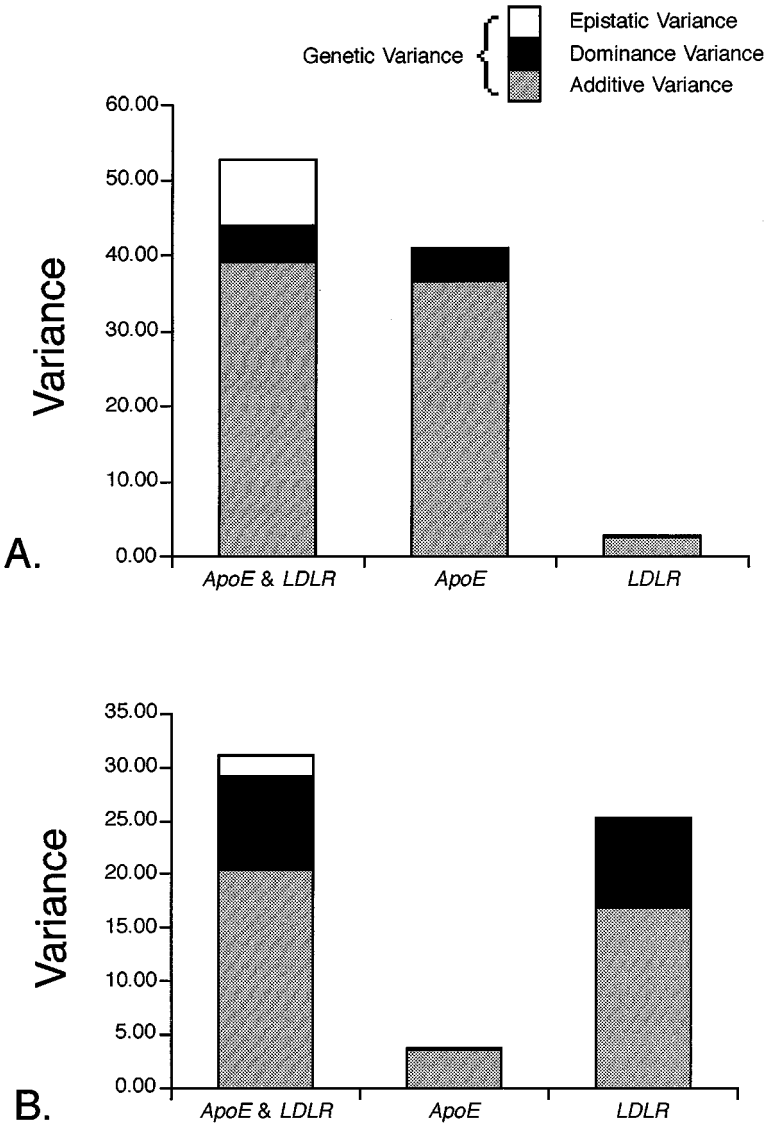


Figure 3 Partitioning of genetic variance for the phenotype of total serum cholesterol level using a constant genotype-to-phenotype mapping of the *ApoE* and *LDLR* loci in two hypothetical populations with differing allele frequencies, as described in the text. The first column shows the decomposition into additive, dominance, and epistatic variance for the two-locus system, and the next two columns shows the marginal decomposition into additive and dominance variance for the two loci considered separately. [From Templeton (138).]

OVERVIEW

The *BRCA 1* story illustrates the practical importance of population thinking in applying the results of the HGP to human health and well-being. Indeed, population thinking (85) may represent the single most important contribution of evolutionary biology to the HGP. Population thinking is central to the interpretation and application of virtually all phases of the diversity research proposed under the new goals of the HGP. Population thinking draws a clear distinction between association and causation, a distinction that is often ignored in human research as illustrated by the earlier discussion of Alzheimer's disease. This distinction is difficult to make because of the evolutionary implications of context-dependent sampling, linkage disequilibrium, and the clustering of functionally related genes within the genome.

The distinction between association and causation is even more critical in light of claims that we will soon be able to "cure" many diseases by altering "defective genes"—a "new eugenics" that "could, at least in principle, be implemented on a quite individual basis" (128). Obviously, such treatments would depend upon an absolute inference of causation, but evolutionary considerations indicate such an absolute inference will be exceedingly difficult. Moreover, population thinking undercuts the idea of a "defective" gene for common diseases. For example, is the *ε4* allele at the *ApoE* locus a "defective gene" and *ϵ2* a "good gene" for CAD because of their marginal associations? Population thinking makes it clear that a marginal association does not always translate well to the individual level—the proposed treatment level of the "new eugenics." For example, considering only the pairwise interactions between *ApoE* genotype and serum cholesterol level upon the odds of CAD, the highest incidence of CAD is found in bearers of the "good" *ε2* allele who have high cholesterol (123). The "defective" *ε4* allele is only associated with the highest incidence of CAD among genotypes when cholesterol levels are closer to average (123). Moreover, given that cholesterol levels can be altered by life style changes (diet, exercise, etc) or drugs, a genetic treatment that takes these interactions into account may still not benefit a particular individual because the individual's phenotypes is temporally dynamic. Indeed, the extensive interactions with environmental factors referred to earlier indicates that the most probable and practical medical applications of genetic knowledge will not be through gene therapy but rather, ironically, by augmenting the importance of the environment; treatment of patients will be individualized by taking into account their norms of reaction to diet, drugs, and other life style variables.

Population thinking therefore serves as an important counterweight to the naive genetic determinism inherent in the almost daily proclamations of the discovery of some new important "defective" gene. Perhaps the greatest use of evolutionary theory in the HGP will be to foster an appreciation of human diversity and its clinical implication that treatment must focus upon an individual with a disease and not upon the disease of the individual.

LITERATURE CITED

1. Albert SM, Gurland B, Maestre G, Jacobs DM, Stern Y, et al. 1995. *ApoE* genotype influences functional status among elderly without dementia. *Am. J. Med. Genet.* 60:583–87
2. Allan CM, Walker D, Segrest JP, Taylor JM. 1995. Identification and characterization of a new human gene (*ApoC4*) in the *apolipoprotein E, C-I*, and *C-II* gene locus. *Genomics* 28:291–300
3. Angelicheva D, Calafell F, Savov A, Jordanova A, Kufardjieva A, et al. 1997. Cystic fibrosis mutations and associated haplotypes in Bulgaria: a comparative population genetic study. *Hum. Genet.* 99: 513–20
4. Berg K. 1990. Risk factor variability and coronary heart-disease. *Acta Geneticae Medicae Et Gemellologiae* 39:15–24
5. Bindon JR, Baker PT. 1997. Bergmann's rule and the thrifty genotype. *Am. J. Phys. Anthropol.* 104:201–10
6. Boerwinkle E, Ellsworth DL, Hallman DM, Biddinger A. 1996. Genetic analysis of atherosclerosis: a research paradigm for the common chronic diseases. *Hum. Mol. Genet.* 5:1405–10
7. Boletini E, Svobodova M, Divoky V, Baysal E, Curuk MA, et al. 1994. Sickle-cell-anemia, sickle-cell beta-thalassemia, and thalassemia major in Albania–characterization of mutations. *Hum. Genet.* 93:182–87
8. Cai XG, Fallin D, Stanton J, Scibelli P, Duara R, et al. 1997. *ApoE* is linked to Alzheimer's disease in a large pedigree. *Am. J. Med. Genet.* 74:365–69
9. Cardon LR, Smith SD, Fulker DW, Kimberling WJ, Pennington BF, et al. 1994. Quantitative trait locus for reading disability on chromosome 6. *Science* 266:276–79
10. Chartier-Harlin M, Parfitt M, Legrain S, Pérez-Tur J, Brousseau T, et al. 1994. *Apolipoprotein E, ε4* allele as a major risk factor for sporadic early and late-onset forms of Alzheimer's disease: analysis of the 19q13.2 chromosomal region. *Human Mol. Genet.* 3:569–74
11. Cheverud JM, Routman EJ. 1995. Epistasis and its contribution to genetic variance components. *Genetics* 139:1455–61
12. Cho JH, Nicolae DL, Gold LH, Fields CT, Labuda MC, et al. 1998. Identification of novel susceptibility loci for inflammatory bowel disease on chromosomes 1q, 3q, and 4q: evidence for epistasis between 1p and Ibd1. *Proc. Natl. Acad. Sci. USA* 95:7502–7
13. Clair DS, Rennie M, Slorach E, Norrman J, Yates C, et al. 1995. *Apolipoprotein E epsilon-4* allele is a risk factor for familial and sporadic presenile Alzheimer's disease in both homozygote and heterozygote carriers. *J. Med. Genet.* 32:642–44
14. Clark AG, Weiss KM, Nickerson DA, Taylor SL, Buchanan A, et al. 1998. Haplotype structure and population genetic inferences from nucleotide sequence variation in human *lipoprotein lipase*. *Am. J. Hum. Genet.* 63:595–612
15. Cloninger CR, Vaneerdewegh P, Goate A, Edenberg HJ, Blangero J, et al. 1998. Anxiety proneness linked to epistatic loci in genome scan of human personality traits. *Am. J. Med. Genet.* 81:313–17
16. Collins FS, Guyer MS, Chakravarti A. 1997. Variations on a theme: cataloging human DNA sequence variation. *Science* 278:1580–81
17. Collins FS, Patrinos A, Jordan E, Chakravarti A, Gesteland R, et al. 1998. New goals for the US human genome project—1998–2003. *Science* 282:682–89
18. Comuzzie AG, Allison DB. 1998. The search for human obesity genes. *Science* 280:1374–77

19. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, et al. 1993. Gene dose of *Apoliprotein E* type *4* allele and the risk of Alzheimer's disease in late onset families. *Science* 261:921–23
20. Cornelis F, Faure S, Martinez M, Prudhomme JF, Fritz P, et al. 1998. New susceptibility locus for rheumatoid arthritis suggested by a genome-wide linkage study. *Proc. Natl. Acad. Sci. USA* 95:10746–50
21. Crandall KA. 1994. Intraspecific cladogram estimation: accuracy at higher levels of divergence. *Syst. Biol.* 43:222–35
22. Crandall KA, Templeton AR. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics* 134:959–69
23. Crow TJ. 1997. Current status of linkage for schizophrenia—polygenes of vanishingly small effect or multiple false positives. *Am. J. Med. Genet.* 74:99–103
24. Cummings AM, Robertson FW. 1984. Polymorphism at the *Apoprotein-E* locus in relation to risk of coronary disease. *Clinical Genet.* 25:310–13
25. Czeizel A. 1989. Application of DNA analysis in diagnosis and control of human diseases. *Biol. Zentrabl.* 108:295–301
26. Davidson B. 1998. Corporate profile Genaissance Pharmaceuticals. *Genet. Engin. News* 18:41,62
27. Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, et al. 1994. A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 371:130–36
28. de la Chapelle A, Wright FA. 1998. Linkage disequilibrium mapping in isolated populations: the example of Finland revisited. *Proc. Natl. Acad. Sci. USA* 95:12416–23
29. Diamond JM. 1992. Diabetes running wild. *Nature* 357:362–63
30. Eaves LJ. 1994. Effect of genetic architecture on the power of human linkage studies to resolve the contribution of quantitative traitloci. *Heredity* 72:175–92
31. El-Hazmi M 1990. *Beta globin* gene haplotypes in the Saudi sickle-cell-anemia patients. *Hum. Hered.* 40:177– 86
32. Enserink M. 1998. Human genetics—opponents criticize Iceland's database. *Science* 282:859
33. Fenster CB, Galloway LF, Chao L. 1997. Epistasis and its consequences for the evolution of natural populations. *Trends Evol. Ecol.* 12:282–86
34. Fisher RA. 1930. *The Genetical Theory of Natural Selection*. Oxford: Clarendon
35. Flatz G. 1992. Lactase deficiency biologic and medical aspects of the adult human lactase polymorphism. In *Genetic Basis of Common Diseases*, ed. RA King, JI Rotter, AG Motulsky, pp. 305–25. Oxford: Oxford Univ. Press
36. Ford D, Easton DF, Stratton M, Narod S, Goldgar D, et al. 1998. Genetic heterogeneity and penetrance analysis of the *BRCA1* and *BRCA2* genes in breast cancer families. *Am. J. Hum. Genet.* 62:676–89
37. Frankel WN, Schork NJ. 1996. Who's afraid of epistasis? *Nat. Genet.* 14:371–73
38. Friedlander Y, Berry EM, Eisenberg S, Stein Y, Leitersdorf E. 1995. Plasma lipids and lipoproteins response to a dietary challenge:analysis of four candidate genes. *Clin. Genet.* 47:1–12
39. Ginns EI, Jean PS, Philibert RA, Galdzicka M, Damschroder-Williams P, et al. 1998. A genome-wide search for chromosomal loci linked to mental health wellness in relatives at high risk for bipolar affective disorder among the Old Order Amish. *Proc. Natl. Acad. Sci. USA* 95:15531–36
40. Gloria-Bottini F, Borgiani P, Amante A, Lucarelli P, Bottini E. 1995. Genetic interactions and the environment: a study of *ADA* and *ACP1* systems in the Sardinian population. *Hum. Hered.* 45:129–34
41. Gulcher J, Stefansson K. 1998. Population genomics laying the groundwork for

genetic disease modeling and targeting. *Clin. Chem. Lab. Med.* 36:523–27
42. Gusella JF, Wexler NS, Conneally PM, Naylor SL, Anderson MA, et al. 1983. A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306:324–28
43. Hallman DM, Boerwinkle E, Saha N, Sandholzer S, Menzel HJ, et al. 1991. The *apolipoprotein E* polymorphism: a comparison of allele frequencies and effects in nine populations. *Am. J. Hum. Genet.* 49:338–49
44. Hallman DM, Visvikis S, Steinmetz J, Boerwinkle E. 1994. The effect of variation in the *apolipoprotein B* gene on plasmid lipid and apolipoprotein B levels. I. A likelihood-based approach to cladistic analysis. *Ann. Hum. Genet.* 58:35–64
45. Harris H. 1966. Enzyme polymorphisms in man. *Proc. R. Soc. Lond. B* 164:298–310
46. Haviland MB, Betard C, Kessling AM, Davignon J, Sing CF. 1991. An application of cladistics to the low density lipoprotein receptor (*LDLR*) gene to identify functional mutations that influence lipid and lipoprotein variation among familial hypercholesterolemia (FH) patients with a 10 kb deletion. *Am. J. Hum. Genet.* 49 (Suppl):471
47. Haviland MB, Ferrell RE, Sing CF. 1997. Association between common alleles of the *low-density lipoprotein receptor* gene region and interindividual variation in plasma lipid and apolipoprotein levels in a population-based sample from Rochester, Minnesota. *Hum. Genet.* 99:108–14
48. Haviland MB, Kessling AM, Davignon J, Sing CF. 1991. Estimation of Hardy-Weinberg and pairwise disequilibrium in the *apolipoprotein AI-CIII-AIV* gene cluster [published erratum appears in *Am. J. Hum. Genet.* 1992 Dec; 51(6):1457]. *Am. J. Hum. Genet.* 49:350–65
49. Haviland MB, Kessling AM, Davignon J, Sing CF. 1995. Cladistic analysis of the *apolipoprotein AI-CIII-AIV* gene cluster using a healthy French Canadian sample. I. Haploid analysis. *Ann. Hum. Genet.* 59:211–31
50. Heizmann C, Kirchgessner T, Kwiterovich PO, Ladias JA, Derby C, et al. 1991. DNA polymorphism haplotypes of the human *lipoprotein-lipase* gene: possible association with high-density-lipoprotein levels. *Hum. Genet.* 86:578–84
51. Hill AVS, Motulsky AG. 1999. Genetic variation and human disease: the role of natural selection. In *Evolution in Health and Disease*, ed. SC Stearns, pp. 50–61. Oxford: Oxford Univ. Press
52. Hill AVS, Sanchez-Mazas A, Barbujani G, Dunston G, Excoffier L, et al. 1999. Human genetic variation and its impact on public health and medicine. In *Evolution in Health and Disease*, ed. SC Stearns, pp. 50–61. Oxford:Oxford Univ. Press
53. Hofmann S, Jaksch M, Bezold R, Mertens S, Aholt S, et al. 1997. Population genetics and disease susceptibility: characterization of Central European haplogroups by mtDNA gene mutations, correlation with D loop variants and association with disease. *Human Mol. Genet.* 6:1835–46
54. Jeunemaitre X, Inoue I, Williams C, Charru A, Tichet J, et al. 1997. Haplotypes of *angiotensinogen* in essential hypertension. *Am. J. Hum. Genet.* 60:1448–60
55. Jones PA, Rideout WMd, Shen JC, Spruck CH, Tsai YC. 1992. Methylation, mutation and cancer. *BioEssays* 14:33–36
56. Kamboh MI. 1995. *Apolipoprotein E* polymorphism and susceptibility to Alzheimer's disease. *Hum. Biol.* 67:195–215
57. Kamboh MI, Sanghera DK, Aston CE, Bunker CH, Hamman RF, et al. 1997. Gender-specific nonrandom association between the *alpha-1-antichymotrypsin* and *apolipoprotein E* polymorphisms in the general population and its implication for the risk of Alzheimer's-disease. *Gen. Epidemiol.* 14:169–80
58. Kamboh MI, Sanghera DK, Ferrell RE, Dekosky ST. 1995. *ApoE*-4* associated

Alzheimer's disease risk is modified by *alpha-1-antichymotrypsin* polymorphism. *Nat. Genet.* 10:486–88
59. Keavney B, McKenzie CA, Connell JMC, Julier C, Ratcliffe PJ, et al. 1998. Measured haplotype analysis of the *angiotensin-I converting enzyme* gene. *Human Mol. Genet.* 7:1745–51
60. Kleyn PW, Vesell ES. 1998. Genetic variation as a guide to drug development. *Science* 281:1820–21
61. Klitz W, Stephens JC, Grote M, Carrington M. 1995. Discordant patterns of linkage disequilibrium of the *peptide-transporter* loci within the *HLA class II* region. *Am. J. Hum. Genet.* 57:1436–44
62. Ko YL, Ko YS, Wu SM, Teng MS, Chen FR, et al. 1997. Interaction between obesity and genetic polymorphisms in the *apolipoprotein CIII* gene and *lipoprotein lipase* gene on the risk of hypertriglyceridemia in Chinese. *Hum. Genet.* 100:327–33
63. Krawczak M, Cooper DN. 1991. Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. *Hum. Genet.* 86:425–41
64. Krawczak M, Reitsma PH, Cooper DN. 1995. The mutational demography of protein C deficiency. *Hum. Genet.* 96:142–46
65. Krushkal J, Xiong MM, Ferrell R, Sing CF, Turner ST, et al. 1998. Linkage and association of adrenergic and dopamine receptor genes in the distal portion of the long arm of chromosome 5 with systolic blood pressure variation. *Human Mol. Genet.* 7:1379–83
66. Lande R. 1983. The response to selection on major and minor mutations affecting a metrical trait. *Heredity* 50:47–65
67. Lander ES, Schork NJ. 1994. Genetic dissestion of complex traits. *Science* 265:2037–48
68. Lannfelt L, Lilius L, Nastase M, Viitanen M, Fratiglioni L, et al. 1994. Lack of association between *apolipoprotein-E* allele *epsilon-4* and sporadic Alzheimer's disease. *NeuroScience Letters* 169:175–78
69. Lehmann DJ, Johnston C, Smith AD. 1997. Synergy between the genes for *butyrylcholinesterase K* variant and *apolipoprotein $\epsilon 4$* in late-onset confirmed Alzheimer's-disease. *Hum. Mol. Genet.* 6: 1933–36
70. Lendon CL, Martinez A, Behrens IM, Kosik KS, Madrigal L, et al. 1997. *E280a Ps-1* mutation causes Alzheimer's-disease but age of onset is not modified by *ApoE* alleles. *Hum. Mutation* 10:186–95
71. Levo A, Partanen J. 1997. Mutation-haplotype analysis of *steroid 21-hydroxylase* (*Cyp21*) deficiency in Finland: implications for the population history of defective alleles. *Hum. Genet.* 99:488–97
72. Levy G. 1998. Predicting effective drug concentrations for individual patients. Determinants of pharmacodynamic variability *Clin. Pharmacokinetics* 34:323–33
73. Li W, Nei M. 1974. Stable linkage disequilibrium without epistasis in subdivided populations. *Theor. Pop. Biol.* 6:173–83
74. Litt M, Jorde LB. 1986. Linkage disequilibria between pairs of loci within a highly polymorphic region of chromosome 2q. *Am. J. Hum. Genet.* 39:166–78
75. Liu L, Forsell C, Lilius L, Axelman K, Corder EH, et al. 1996. Allelic association but only weak evidence for linkage to the *apolipoprotein E* locus in late-onset Swedish Alzheimer families. *Am. J. Med. Genet.* 67:306–11
76. Lobos EA, Rudnick CH, Watson MS, Isenberg KE. 1989. Linkage disequilibrium study of RFLPs detected at the human *muscle nicotinic acetylcholine receptor subunit* genes. *Am. J. Hum. Genet.* 44:522–33
77. Lucarini N, Nicotra M, Gloria-Bottini F, Borgiani P, Amante A, et al. 1995. Interaction between ABO blood groups and *ADA* genetic polymorphism during intrauterine life. A comparative analysis of couples with habitual abortion and normal

puerperae delivering a live-born infant. *Hum. Genet.* 96:527–31

78. Lucotte G, Visvikis S, Leiningermuler B, David F, Berriche S, et al. 1994. Association of *apolipoprotein E* allele *epsilon-4* with late-onset sporadic Alzheimer's disease. *Am. J. Med. Genet.* 54:286–88
79. Ludwig EH, Borecki IB, Ellison RC, Folsom AR, Heiss G, et al. 1997. Associations between candidate loci *angiotensin-converting enzyme* and *angiotensinogen* with coronary heart disease and myocardial infarction—the NHLBI Family Heart Study. *Ann. Epidemiol.* 7:3–12
80. Ma J, Yee A, H. B. Brewer J, Das S, Potter H. 1994. Amyloid-associated proteins α_1-antichymotrysin and apolipoprotein E promote assembly of Alzheimer β-protein into filaments. *Nature* 372:92–94
81. Magewu AN, Jones PA. 1994. Ubiquitous and tenacious methylation of the CpG site in codon 248 of the p53 gene may explain its frequent appearance as a mutational hot spot in human cancer. *Mol. Cell. Biol.* 14:4225–32
82. Malone KE, Daling JR, Thompson JD, Obrien CA, Francisco LV, et al. 1998. *BRCA1* mutations and breast cancer in the general population:analyses in women before age 35 years and in women before age 45 years with first-degree family history. *JAMA: J. Am. Med. Assoc.* 279:922–29
83. Marshall A. 1997. Laying the foundations for personalized medicines. *Nat. Biotech.* 15:954–57
84. Matsumoto T, Imamura O, Yamabe Y, Kuromitsu J, Tokutake Y, et al. 1997. Mutation and haplotype analyses of the Werner's-syndrome gene based on its genomic structure: genetic epidemiology in the Japanese population. *Hum. Genet.* 100:123–30
85. Mayr E. 1974. The challenge of diversity. *Taxon* 23:3–9
86. McClearn GE, Johansson B, Berg S, Pedersen NL, Ahern F, et al. 1997. Substantial genetic influence on cognitive abilities in twins 80 or more years old. *Science* 276:1560–63
87. Meyer UA. 1999. Medically relevant genetic variation of drug effects. In *Evolution in Health and Disease*, ed. SC Stearns, pp. 41–49. Oxford: Oxford Univ. Press
88. Morahan G, Huang D, Tait BD, Colman PG, Harrison LC. 1996. Markers on distal chromosome 2q linked to insulin-dependent diabetes mellitus. *Science* 272:1811–13
89. Morgan K, Morgan L, Carpenter K, Lowe J, Lam L, et al. 1997. Microsatellite polymorphism of the *alpha(1)-antichymotrypsin* gene locus associated with sporadic Alzheimer's disease. *Hum. Genet.* 99:27–31
90. Motolsky AG. 1957. Drug reactions enzymes, and biochemical genetics. *J. Am. Med. Assoc.* 165:835–37
91. Nakagawa H, Koyama K, Miyoshi Y, Ando H, Baba S, et al. 1998. Nine novel germline mutations of *STK11* in ten families with Peutz-Jeghers-Syndrome. *Hum. Genet.* 103:168–72
92. Nakauchi Y, Suehiro T, Yamamoto M, Yasuoka N, Arii K, et al. 1996. Significance of *angiotensin I-converting enzyme* and *angiotensin II type 1 receptor* gene polymorphisms as risk factors for coronary heart disease. *Atherosclerosis* 125:161–69
93. Nathan BP, Bellosta S, Sanan DA, Weisgraber KH, Mahley RW, et al. 1994. Differential effects of apolipoproteins E3 and E4 on neuronal growth *in vitro*. *Science* 264:850–52
94. Neel JV. 1962. Diabetes mellitus: a "thrifty genotype" rendered detrimental by "progress". *Am. J. Hum. Genet.* 14:353–62

94a. Neel JV, Weder AB, Julius S. 1998. Type II diabetes, essential hypertension, and obesity as "syndromes of impaired genetic homeostasis": the "thrifty genotype" enters the 21st century. *Perspect. Biol. Med.* 42:44–74

95. Nelson MR, Kardia SLR, Ferrell RE, Sing CF. 1998. *A combinatorial optimization*

approach to identify multilocus genotypic classes associated with quantitative trait variability. Presented at Annu. Meet. Am. Soc. Hum. Genet., 48th, Denver

96. Nesse RM, Williams GC. 1998. Evolution and the origins of disease. *Sci. Am.* Nov.:86–93
97. Newman B, Mu H, Butler LM, Millikan RC, Moorman PG, et al. 1998. Frequency of breast cancer attributable to *BRCA1* in a population-based series of American women. *J. Am. Med. Assoc.* 279:915–21
98. Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, et al. 1998. DNA sequence diversity in a 9.7-kb region of the human *lipoprotein lipase* gene. *Nat. Genet.* 19:233–40
99. Okuizumi K, Onodera O, Namba Y, Ikeda K, Yamamoto T, et al. 1995. Genetic association of the *very low density lipoprotein* (*VLDL*) *receptor* gene with sporadic Alzheimer's disease. *Nat. Genet.* 11:207–9
100. Oner C, Dimovski AJ, Olivieri NF, Schiliro G, Codrington JF, et al. 1992. *Beta-S* haplotypes in various world populations. *Hum. Genet.* 89:99-104
101. Peacock RE, Temple A, Gudnason V, Rosseneu M, Humphries SE. 1997. Variation at the *lipoprotein lipase* and *apolipoprotein AI-CIII* gene loci are associated with fasting lipid and lipoprotein traits in a population sample from Iceland: interaction between genotype, gender, and smoking status. *Gen. Epidemiol.* 14:265–82
102. Pedersen JC, Berg K. 1990. Gene-gene interaction between the *low-density-lipoprotein receptor* and *apolipoprotein-E* loci affects lipid levels. *Clinical Genet.* 38:287–94
103. Pedersen JD, Berg K. 1989. Interaction between *Low-Density Lipoprotein Receptor* (*LDLR*) and *Apolipoprotein-E* (*ApoE*) alleles contributes to normal variation in lipid level. *Clinical Genet.* 35:331–37
104. Pereztur J, Campion D, Martinez M, Brice A, Tardieu S, et al. 1995. Evidence for *apolipoprotein E epsilon-4* association in early-onset Alzheimer's patients with late-onset relatives. *Am. J. Med. Genet.* 60:550–53
105. Pickering L. 1998. Perkin-Elmer and Incyte to map the human genome. *Genet. Engin. News* 18:18, 41, 51
106. Pillay TS, Langlois WJ, Olefsky JM. 1995. The genetics of non-insulin-dependent diabetes mellitus. *Adv. Genet.* 32: 51–98
107. Poduslo SE, Riggs D, Schwankhaus J, Osborne A, Crawford F, et al. 1995. Association of apolipoprotein E but not B with Alzheimer's disease. *Hum. Genet.* 96:597–600
108. Porkka KV, Taimela S, Kontula K, Lehtimaki T, Aaltosetala K, et al. 1994. Variability gene effects of DNA polymorphisms at the *ApoB, ApoAI/CIII* and *ApoE* loci on serum-lipids:the cardiovascular risk in young Finns study. *Clin. Genet.* 45:113–21

108a. Rahman N, Stratton MR. The Genetics of breast cancer susceptibility. *Annu. Rev. Genet.* 32:95–121

109. Reddy PH, Modell B. 1996. Reproductive behaviour and natural selection for the sickle gene in the Baiga tribe of Central India–the role of social parenting. *Ann. Hum. Genet.* 60:231–36
110. Rideout WMd, Coetzee GA, Olumi AF, Jones PA. 1990. 5-Methylcytosine as an endogenous mutagen in the human *LDL receptor* and *p53* gene. *Science* 249:1288–90
111. Routman EJ, Cheverud JM. 1997. Gene effects on a quantitative trait: two-locus epistatic effects measured at microsatellite markers and at estimated QTL. *Evolution.* 51:1654–62
112. Sandbrink R, Hartmann T, Masters CL, Beyreuther K. 1996. Genes contributing to Alzheimer's disease. *Mol. Psychol.* 1:27–40
113. Sanghera DK, Aston CE, Saha N,

Kamboh MI. 1999. DNA polymorphisms in two paraoxonase genes (*PON1* and *PON2*) are associated with the risk of coronary heart disease. *Am. J. Hum. Genet.* 62:36–44

114. Schiliro G, Spena M, Giambelluca E, Maggio A. 1990. Sickle haemoglobinpathies in Sicily. *Am. J. Hematol.* 33:81–85
115. Schmalhausen II. 1949. *Factors of Evolution*. Philadelphia: Blakiston Company. 327 pp.
116. Schmutte C, Jones PA. 1998. Involvement of DNA methylation in human carcinogenesis. *Biol. Chem.* 379:377–88
117. Schork NJ, Cardon LR, Xu X. 1998. The future of genetic epidemiology. *Trends Genet.* 14:266–72
118. Schull WJ, Annas GJ, Arnheim N, Blangero J, Chakravarti A, et al. 1997. *Evaluating Human Genetic Diversity.* Washington, DC: Natl. Acad. 91 pp.
119. Sheffield VC, Stone EM, Carmi R. 1998. Use of isolated inbred human populations for identification of disease genes. *Trends Genet.* 14:391–96
120. Sing CF, Boerwinkle EA. 1987. Genetic architecture of inter-individual variability in apolipoprotein, lipoprotein and lipid phenotypes. *Ciba Found. Symp.* 130:99–127
121. Sing CF, Davignon J. 1985. Role of the *Apolipoprotein E* polymorphism in determining normal plasma lipid and lipoprotein variation. *Am. J. Hum. Genet.* 37:268–85
122. Sing CF, Haviland MB, Reilly SL. 1996. Genetic architecture of common multifactorial diseases. In *Variation in the Human Genome*, ed. G Cardew, pp. 211–32. Chichester: Wiley
123. Sing CF, Haviland MB, Templeton AR, Reilly SL. 1995. Alternative genetic strategies for predicting risk of atherosclerosis. In *Atherosclerosis X. Excerpta Medica International Congress Series*, ed. FP Woodford, J Davignon, AD Sniderman, pp. 638–44. Amsterdam: Elsevier Sci. B.V.
124. Sing CF, Haviland MB, Templeton AR, Zerba KE, Reilly SL. 1992. Biological complexity and strategies for finding DNA variations responsible for interindividual variation in risk of a common chronic disease, coronary artery disease. *Ann. Med.* 24:539–47
125. Sing CF, Haviland MB, Zerba KE, Templeton AR. 1992. Application of cladistics to the analysis of genotype-phenotype relationships. *Eur. J. Epidemiol.* 8:3–9
126. Sing CF, Moll PP. 1990. Genetics of atherosclerosis. *Annu. Rev. Genet.* 24: 171–88
127. Sing CF, Orr JD. 1976. Analysis of genetic and environmental sources of variation in serum cholesterol in Tecumseh, Michigan. III. Identification of genetic effects using 12 polymorphic genetic blood marker systems. *Am. J. Hum. Genet.* 28:453–64
128. Sinsheimer RG. 1969. The prospect of designed genetic change. *Engin. Sci.* 32:8–13
129. Stengard JH, Weiss KM, Sing CF. 1998. An ecological study of association between coronary heart disease mortality rates in men and the relative frequencies of common allelic variations in the gene coding for Apolipoprotein E. *Hum. Genet.* 103:234–41
130. Stengard JH, Zerba KE, Pekkanen J, Ehnholm C, Nissinen A, et al. 1995. *Apolipoprotein E* polymorphism predicts death from coronary heart disease in a longitudinal study of elderly Finnish men. *Circulation* 91:265–69
131. Stephens JC, Briscoe D, O'Brien SJ. 1994. Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *Am. J. Hum. Genet.* 55:809–24
132. Strittmatter WJ, Weisgraber KH, Goedert M, Saunders AM, Huang D, et al. 1994. Hypothesis–microtubule instability and

paired helical filament formation in the Alzheimer disease brain are related to apolipoprotein E genotype. *Exp. Neuro.* 125:163–71

133. Templeton AR. 1982. Genetic architectures of speciation. In *Mechanisms of Speciation*, ed. C Barigozzi, pp. 105–21. New York: Liss
134. Templeton AR. 1987. The general relationship between average effect and average excess. *Genet. Res.* 49:69–70
135. Templeton AR. 1995. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping or DNA sequencing. V. Analysis of case/control sampling designs: Alzheimer's disease and the *Apoprotein E* locus. *Genetics* 140:403–9
136. Templeton AR. 1996. Cladistic approaches to identifying determinants of variability in multifactorial phenotypes and the evolutionary significance of variation in the human genome. In *Variation in the Human Genome*, ed. G Cardew, pp. 259–83. Chichester: Wiley
137. Templeton AR. 1998. Linkage mapping versus the candidate gene approach. *Proceedings of the 6th World Congress on Genetics Applied to Livestock Production* 26:175–82
138. Templeton AR. 1999. Epistasis and complex traits. In *Epistasis and the Evolutionary Process*, ed. M Wade, E Brodie III, J Wolf. Oxford: Oxford Univ. Press. In press
139. Templeton AR, Boerwinkle E, Sing CF. 1987. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. I. Basic theory and an analysis of Alcohol Dehydrogenase activity in *Drosophilaz. Genetics* 117:343–51

139a. Templeton AR, Clark AG, Weiss KM, Nickerson DA, Boerwinkle E, Sing CF. 1999. Recombinational and mutational hotspots within the human *Lipoprotein Lipase* gene. *Am. J. Hum. Genet.* In press

140. Templeton AR, Crandall KA, Sing CF. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* 132:619–33
141. Templeton AR, Sing CF. 1993. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. IV. Nested analyses with cladogram uncertainty and recombination. *Genetics* 134:659–69
142. Templeton AR, Sing CF, Brokaw B. 1976. The unit of selection in *Drosophila mercatorum*. I. The interaction of selection and meiosis in parthenogenetic strains. *Genetics* 82:349–76
143. Templeton AR, Sing CF, Kessling A, Humphries S. 1988. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping. II. The analysis of natural populations. *Genetics* 120:1145–54
144. Thomas AE, Green FR, Lamlum H, Humphries SE. 1995. The association of combined *alpha* and *beta fibrinogen* genotype on plasma fibrinogen levels in smokers and non-smokers. *J. Med. Genet.* 32:585–89
145. Thompson EA, Deeb S, Walker D, Motulsky AG. 1988. The detection of linkage disequilibrium between closely linked markers:RFLPs at the *AI-CIII apolipoprotein* genes. *Am. J. Hum. Genet.* 42:113–24
146. Todd JA. 1995. Genetic analysis of type I diabetes using whole genome approaches. *Proc. Natl. Acad. Sci. USA* 92:8560–65
147. Todorova A, Danieli GA. 1997. Large majority of single-nucleotide mutations along the *dystrophin* gene can be explained by more than one mechanism of mutagenesis. *Hum. Mut.* 9:537–47
148. Tvrdik T, Marcus S, Hou SM, Falt S,

Noori P, et al. 1998. Molecular characterization of two deletion events involving ALU-sequences, one novel base substitution and two tentative hotspot mutations in the *Hypoxanthine Phosphoribosyltransferase* (*HPRT*) gene in five patients with Lesch-Nyhan-Syndrome. *Hum. Genet.* 103:311–18

149. Valdes AM, McWeeney S, Thomson G. 1997. HLA class II DR-DQ amino acids and insulin-dependent diabetes mellitus:application of the haplotype method. *Am. J. Hum. Genet.* 60:717–28
150. Valdes AM, Thomson G. 1997. Detecting disease predisposing variants: the haplotype method. *Am. J. Hum. Genet.* 60:703–16
151. Virtaneva K, Damato E, Miao IM, Koskiniemi M, Norio R, et al. 1997. Unstable minisatellite expansion causing recessively inherited Myoclonus Epilepsy, EPM1. *Nat. Genet.* 15:393–96
152. Wade MJ. 1996. Adaptation in subdivided populations: kin selection and interdemic selection. In *Adaptation*, pp. 381–405. New York: Academic Press
153. Wade MJ, Goodnight CJ. 1991. Wright's shifting balance theory: an experimental study. *Science* 253:1015–18
154. Wade MJ, Griesemer JR. 1998. Populational heritability empirical studies of evolution in metapopulations. *Am. Nat.* 151:135-47
155. Wakeley J, Hey J. 1997. Estimating ancestral population parameters. *Genetics* 145:847–55
156. Wang DG, Fan JB, Siao CJ, Berno A, Young P, et al. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077–82
157. Weisgraber KH, Mahley RW. 1996. Human apolipoprotein E: the Alzheimer's disease connection. *FASEB J.* 10:1485–94
158. Weiss KM. 1993. *Genetic Variation and Human Disease:Principles and Evolutionary Approaches*. Cambridge, UK: Cambridge Univ. Press. 354 pp.
159. West WL, Knight EM, Pradhan S, Hinds TS. 1997. Interpatient variability:genetic predisposition and other genetic factors. *J. Clin. Pharmacol.* 37:635-48
160. Whitlock MC, Phillips PC, Wade MJ. 1993. Gene interaction affects the additive genetic variance in subdivided populations with migration and extinction. *Evol.* 47:1758–69
161. Wright S. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution. *Proc. Sixth Int. Cong. Genet* 1:356–66
162. Wright S. 1934. Physiological and evolutionary theories of dominance. *Am. Nat.* 68:25–53
163. Wu WR, Li WM. 1994. A new approach for mapping quantitative trait loci using complete genetic marker linkage maps. *Theor. Appl. Genet.* 89:535–39
164. Yang AS, Gonzalgo ML, Zingg JM, Millar RP, Buckley JD, et al. 1996. The rate of CpG mutation in Alu repetitive elements within the *p53 tumor suppressor* gene in the primate germline. *J. Mol. Biol.* 258:240–50
165. Zerba KE, Ferrell RE, Sing CF. 1998. Genetic structure of five susceptibility gene regions for coronary artery disease:disequilibria within and among regions. *Hum. Genet.* 103:346–54
166. Zerba KE, Kessling AM, Davignon J, Sing CF. 1991. Genetic structure and the search for genotype-phenotype relationships: an example from disequilibrium in the *Apo B* gene region. *Genetics* 129:525–33