Garland Science

HUMAN EVOLUTIONARY GENETICS

second edition

Jobling · Hollox · Hurles · Kivisild · Tyler-Smith

WHAT HAPPENS WHEN POPULATIONS MEET

CHAPTER FOURTEEN

The previous four chapters presented the evidence for the first settlement of the Old and the New World by anatomically modern humans. In Chapter 9 we learned that the majority of the genetic variation observed outside Africa today derives from an out-of-Africa dispersal event 50–70 KYA with a minor addition of gene flow from archaic hominins, including the Neanderthals. In the succeeding Chapters 10–13 we saw how, during and after the initial colonization of the continents, populations diversified due to genetic drift and accumulation of new mutations. As a result, we can infer the continental origins of individuals easily, even from a small number of carefully chosen genetic polymorphisms. However, populations are not discrete entities: we know from archaeological and historical evidence that they are often in flux, generating hybrid populations and individuals of mixed ancestry. This chapter discusses this process of mixing, known as **admixture**, and how we can detect and measure the extent of admixture from genetic evidence.

14.1 WHAT IS GENETIC ADMIXTURE?

Neighboring populations frequently exchange individuals that contribute to an ongoing process of bidirectional **gene flow** between them. However, a third, hybrid population does not usually result from this kind of exchange. The term admixture is reserved for the formation of a hybrid population from the mixing of ancestral populations that have previously been relatively isolated from one another. The range expansion or migration of one population into a region inhabited by a previously isolated population is one such scenario. Thus, admixture can be thought of as being initiated at a specific point in time, when the populations first came into contact.

As with most studies described in this book, we are almost exclusively limited to examining modern genetic diversity. When we examine modern populations, we detect not simply the proportions of admixture established when the populations first met, but the summation of cumulative gene flow from the time when they first met to the present day. Thus the consequences of admixture and gene flow may be difficult to distinguish. Of course, the imprint of past admixture in modern populations has also been modified by the **drift**, selection, and mutation processes that shape all genetic diversity.

Many different issues of population prehistory can be viewed as questions about admixture. All that is required is that alternative ancestral populations can be differentiated from one another in either time or space. For example, the relative contributions to the modern gene pool of several migrations to the same location can be thought of as admixture between the source populations for each migration. It is in this framework that the relative contributions of **Paleolithic**

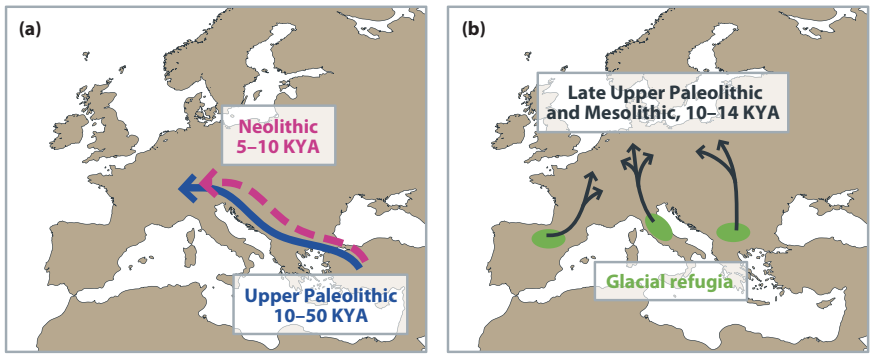


Figure 14.1: Maps showing potential sources of past admixture in Europe. (a) Neolithic and Paleolithic peoples migrated into Europe at different times, although by similar routes, and both are thought to have contributed to the modern European gene pool. (b) Peoples from different glacial refugia (*green*) may have been isolated from one another during the ice age, and as conditions improved northward migration would have presented opportunities for admixture.

hunter-gatherers and **Neolithic** farmers to modern European diversity have been considered (see **Figure 14.1**). Whilst the two ancestral populations spread from largely similar geographical origins in the Near East, they are separated in time by thousands of years. This particular example is explored in greater detail in **Section 12.5**.

The processes of isolation and range expansion that result in subsequent admixture can be driven by environmental changes. During the recent **ice ages**, the environment in more northerly latitudes became uninhabitable. Humans and other plant and animal species found refuge in pockets of more hospitable climate, known as **glacial refugia**. These refugia were often isolated from one another. For example, three major European glacial refugia were the Iberian Peninsula, Italy, and the Balkans/Greece.

After the end of the last ice age, about 14 KYA, many species started the long process of recolonizing the more northerly latitudes from these refugia. During this process, previously isolated populations were often brought back into contact with one another, the genetic consequences of which can be analyzed through a consideration of admixture.

More recent historical events that can be studied through an appreciation of admixture processes include episodes of enforced migration. These episodes have often been motivated by colonialism and/or the creation of a subjugated labor force—in other words, a slave trade. Although the eighteenth century Atlantic slave trade has received most attention, slavery was widespread throughout the ancient world including the Egyptian, Greek, and Roman empires, among Arabs, in Iceland, in the Pacific, and in Africa itself.

Historically, some of the first studies of genetic admixture at the molecular level were those that analyzed the frequencies of different blood group protein alleles in African-Americans, comparing them to European-Americans and Africans.[12] The aim was primarily to quantify European admixture among African-Americans. Both marital records and a supposed lightening of African-American skin color provided external evidence of admixture. A number of studies of different blood groups in different US populations were published throughout the 1950s and 1960s (for example by T. E. Reed[30]). They demonstrated that the extent of European admixture varied considerably among the different regional populations of African-Americans over the 10 generations or so since the peak of the major period of African slavery. The proportion of European genes within different African-American populations was shown to vary from ~4% to ~30%, with southern populations having consistently lower levels of European admixture. More recent studies use DNA polymorphisms rather than protein polymorphisms, but the conclusions remain the same.

Often genetic studies of prehistoric admixture events are initiated when evidence from nongenetic sources indicates that admixture might have occurred. This is because, unsurprisingly, the meeting of previously isolated populations

Box 14.1: The ever-changing terminology of people with mixed ancestry

Societies undergoing admixture have often sought to classify individuals on the basis of their proportions of admixture. Many of these largely historical terms introduced fine gradations of admixture, but the number of terms required to cover all possible proportions doubles each generation, and it quickly becomes impractical to maintain a word for each fraction. As a result many of these words have all but died out, whilst others have been retained in common usage as general terms for people of mixed ancestry.

As with all terminologies associated with contentious societal issues, many of these words have been considered offensive at some point in time. It could be argued that the rapid turnover of names for individuals of mixed ancestry is driven by society's need to find neutral words free from negative connotations. However, as the societal inequalities remain, these new terms attract derogatory associations and fresh terms need to be invented at regular intervals.

Term	Parents	Proportion
Mulatto	Black and White	1/2 Black
Quadroon	Mulatto and White	1/4 Black
Octoroon	Quadroon and White	1/8 Black
Mustifee	Octoroon and White	1/16 Black
Mustifino	Mustifee and White	1/32 Black
Cascos	Mulatto and Mulatto	1/2 Black
Sambo	Mulatto and Black	3/4 Black
Mango	Sambo and Black	7/8 Black
Metisse	White and Native American	General
Mestizo	White and Native American	General
Griffe	Black and Native American	General
Hapa	Asian/Polynesian and White	General

has effects beyond the realm of genetics (see **Box 14.1**). But can genetic admixture be recognized in the absence of corroborative historical or prehistorical information? This question will be addressed in Section 14.3.

Admixture has distinct effects on genetic diversity

The process of admixture shapes genetic diversity in a number of different ways. In this chapter we will explore how seeking these different imprints in modern genetic diversity can lead to the inference that admixture occurred some time in the past. Our ability to detect admixture depends in part on the number of polymorphisms used, and how differentiated the source populations were from one another. As we shall see, the more different the ancestral populations were, the easier it is to detect and quantify admixture.

While admixture at the population level can be detected in a single genetic locus, multiple loci will be required to infer admixed ancestry in a single individual. This raises the additional complication that some alleles may have their ancestry in one parental population while other alleles have their ancestry in another. This is an inevitable consequence of sexual reproduction and **diploidy**. In fact, it becomes increasingly unlikely over time that any individual from an admixed population will be able to trace all their genes to a single source population (see **Figure 14.2** and **Table 14.1**). Different genomes within an admixed population, though, are likely to exhibit differing amounts of admixture. In **panmictic** populations this difference decreases rapidly in time, but admixed populations are not always panmictic. Nonrandom mating could be due to geographic structure or socioeconomic factors, both of which can contribute to the variation in admixture proportions at the individual level. Although an estimate of population admixture is typically presented as an average of the admixture among the individual genomes within it, the range of variation among individuals can inform us both about the time since the admixture event, and the nature of the admixture event. For example, analysis of 181 Mexicans, representing a relatively young admixed population, showed the full range of 0–100% European ancestry.[9] In contrast, analysis of Polynesians, who derive from an admixture

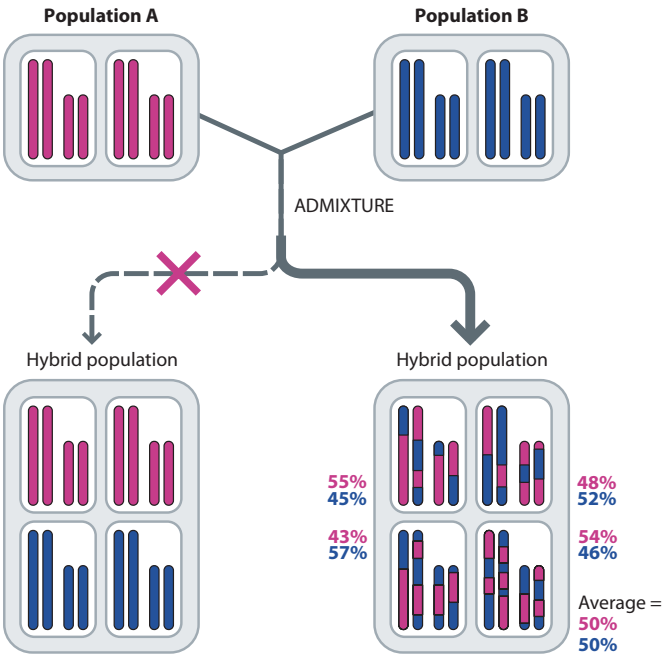


Figure 14.2: Admixture within individual genomes. Diploid genomes comprising two autosomes are shown schematically in ancestral and hybrid populations. There are two individuals (*white rounded rectangles*) in each ancestral population (*gray rounded rectangles*), and four individuals in each hybrid population. Admixture does not result in a population in which individuals can trace the ancestry of their entire genome to one of two ancestral populations, but rather a population in which all individuals have genomes of mixed ancestry.

event more than 100 generations ago, showed a narrow range of 76–84% of Asian contribution[19] (see **Section 13.3** for details).

The complex genetic ancestry of a genome often contrasts with the simplicity of an individual's perceived identity. Because of this, public dissemination of the results from admixture studies needs to be undertaken responsibly and with due care for their potential impact. It is worth remembering that ancestral populations themselves are likely to carry mixed ancestry of genes some of which, as we saw previously (**Section 9.5**), can be traced back to admixture between modern humans and Neanderthals.

In this chapter we will consider a variety of statistical methods that have been used to study this important issue. Then we will examine some case studies that illustrate the power of these methods and demonstrate the diverse outcomes of population encounters in the past. But first, we will consider the impact of admixture on other features of the populations involved.

TABLE 14.1:
PROBABILITY OF ALL GENES OF AN INDIVIDUAL DERIVING FROM A SINGLE ANCESTRAL POPULATION IN A PANMICTIC HYBRID POPULATION

ADMIXTURE 50:50					ADMIXTURE 1:20				
t	A	B	A+B	AB	t	A	B	A+B	AB
0	0.5	0.5	1	0	0	0.05	0.95	1	0
1	0.25	0.25	0.5	0.5	1	0.003	0.903	0.905	0.095
2	0.063	0.063	0.125	0.875	2	6×10^{-6}	0.815	0.815	0.185
3	0.004	0.004	0.008	0.992	3	4×10^{-11}	0.663	0.663	0.337
4	2×10^{-5}	2×10^{-5}	3×10^{-5}	1	4	2×10^{-21}	0.44	0.44	0.56
5	2×10^{-10}	2×10^{-10}	5×10^{-10}	1	5	2×10^{-42}	0.194	0.194	0.806

t, number of generations since admixture. A, B, probability of observing individuals with full ancestry in either of the two source populations A and B; $A_t = (A_{t-1})^2$; $B_t = (B_{t-1})^2$. A+B, probability of observing non-admixed individuals. AB, probability of observing admixed individuals; 1 – (A + B).

14.2 THE IMPACT OF ADMIXTURE

Different sources of evidence can inform us about admixture

Genetic admixture is not the only consequence of the meeting of populations. Such events often impact significantly upon the cultural features of the populations involved. Thus many episodes of prehistoric admixture may well be detected using other records of prehistory, although it should be remembered that there need not necessarily be an archaeological or linguistic correlate for every genetic episode, and vice versa.

In the first few generations of admixture, individuals that descend primarily from one of the ancestral populations are often easily identifiable through their language or appearance. Individuals apportioned to the different ancestries are rarely on an equal footing in the nascent society. For example, the status of African slaves, European settlers, and Native Americans within the Americas was far from equal. Genetics cannot be divorced from these sociological considerations, since they directly influence the nature of the admixture. Rather, integrating genetic evidence with other prehistorical and historical records allows a richer appreciation of population encounters in prehistory.

Consequences of admixture for language

Populations that have been isolated from one another will accumulate linguistic differences relatively rapidly, and perhaps even speak different, mutually unintelligible, languages. What kinds of linguistic changes in a hybrid population might we expect to see as a result of their admixture? Bilingualism can be one short-term outcome, involving little change to either language. Alternative outcomes can be language mixing leading to highly dynamic **pidgin** languages that can be specific to a given location where admixture occurred (for example, Spanglish, referring to different blends of Spanish and English in Central America), or the establishment of a **lingua franca**—a language that would be widely understood in a broader geographic region where many languages meet. For example, during the Renaissance era commerce and diplomacy in the eastern Mediterranean was mediated largely by a mixed language that was based on Romance languages (Italian, Spanish, and French) enriched with specific loan-words from the Arabic, Turkish, and Greek languages.

The first point to appreciate is that a language is unlike a genome in several important ways. Whereas it is perfectly possible to assemble a fully functioning hybrid genome from several ancestral genomes with no associated costs, the same is not generally true of languages, because they must maintain a certain level of coherence to function adequately. Pidgin languages can arise and be erased over the span of a single generation. When the next generation of children of the pidgin speakers starts speaking it, the hybrid language becomes fixed. Linguists call native languages that represent fixed hybrids of parental languages **creoles**. A number of well-known modern creoles derive from European languages (French, Spanish, Portuguese, and English) and indigenous languages brought together by the actions of European colonial powers over the past few hundred years. For example, the Cajun language of Louisiana is a creole derived from French and languages spoken by African slaves. Creolization is considered to be a relatively rare process in language evolution.

Much more common than the development of creoles is the limited incorporation of certain features of one language into a dominant substrate from another language. These linguistic borrowings can affect different aspects of the language. The simplest example is the incorporation of outside words into a language. For example, among Polynesian languages, the term for the sweet potato (*kumara* in New Zealand Maori) derives from the word *kumar* from the Quechuan languages spoken in South America. As well as words, elements of structure can also be borrowed. For example, the order of subject, verb, and object within a sentence is often different between languages. Some **Austronesian** languages

spoken in areas with neighboring **Papuan** speakers have adopted the "verb last" order of these Papuan languages (subject–object–verb: "he it hit"), as opposed to the more common "verb medial" organization (subject–verb–object: "he hit it") found among closely related Austronesian languages. These types of structural change result from what linguist Malcolm Ross calls:

> ... the natural pressure to relieve the bilingual speakers' mental burden by expressing meanings in parallel ways in both languages.

There are thus a wide variety of possible linguistic consequences of contact between populations speaking different languages, and which particular consequence follows in which situation is—as with the genetic outcomes—largely determined by the social context of this contact.

Spoken languages are not the only linguistic source of evidence for admixture: the names of people (surnames in particular) and of places (together referred to as **onomastic** evidence) are both capable of revealing the hybrid nature of a population. For example, English towns have names derived from Celtic, French (Norman), and Scandinavian languages as well as from Anglo-Saxon. It is worth noting, however, that evidence of past contact from sources such as place names does not imply that significant genetic admixture will be found in the current inhabitants. The Caribbean island populations of today, for example, may contain little genetic input from the original inhabitants.

In societies where surnames follow clear lines of inheritance, they have often been used in population genetic analyses, and admixture studies are no exception. Patterns of surname introgression have been shown to be correlated with levels of admixture in a number of different populations.[7] These conclusions have been reinforced by genetic analysis. Nevertheless, such surname studies have been dubbed the "poor man's population genetics,"[11] and are of real use only where genetic data are unavailable, and when admixture has occurred within the time frame of surname usage: this varies greatly from population to population, and may be very recent. However, if records are sufficiently detailed, surname analysis can reveal how admixture processes may have changed over time.

Archaeological evidence for admixture

The answers that archaeology provides to the question, What happens when populations meet? are primarily cultural in nature. The temporal and spatial distribution of archaeological sites can be used to demonstrate contact between cultures, and subsequent cultural change. However, such evidence only establishes the *potential* for genetic admixture. Before genetic admixture can be inferred, it must be assumed that populations with different **material cultures** are also different genetically and that the movement of artifacts is mirrored by the movement of people. In other words, artifacts are not being distributed by a set of sequential trading exchanges. New cultures can be adopted wholesale, or elements of individual cultures can be combined together in a process of integration. The integration of two cultural traditions may or may not be accompanied by genetic admixture. Similarly, a wholesale replacement of cultural practices may or may not be associated with a similar replacement of genes. With these caveats in mind it is worth noting that the spatiotemporal spread of an archaeological culture does indicate the geographical location of likely ancestral populations. In addition, the precision of archaeological dating provides good estimates for the time-scale of potential admixture processes.

Approaches based on **physical anthropology** have been adopted to seek phenotypic changes associated with genetic admixture in skeletal remains. Such work is often contentious, as human populations can rarely be well differentiated on skeletal evidence alone. In addition, alterations in cultural practices, for example, specific foot-binding or skull-compression traditions, or differences in diet, may cause significant morphological changes through **developmental plasticity**, rather than any change of genes[18] (see **Section 15.2**).

The biological impact of admixture

Our focus in this chapter is on identifying and quantifying past admixture. While in cases of highly differentiated ancestral populations, a small number of variants are sufficient to detect admixture, quantification of admixture proportions requires a larger number of polymorphisms across the whole genome so that specific selection processes can be excluded as explanations for modern patterns of genetic diversity. Nonetheless, all genetic admixtures will lead to a variety of phenotypic effects. Any quantitative trait that is genetically encoded and well differentiated between populations will be altered in admixed populations. Obvious physical examples include pigmentation, body proportions, and stature. In the past, these phenotypic data have been used to calculate admixture proportions, and, indeed, protein-coding genes associated with skin color show F_{ST} (see Box 5.2) values that are higher than the genome average, and can therefore give information about ancestry in admixed populations (Chapter 15).

Disease **prevalences** are often different between ancestral populations (see Chapter 16 for a discussion). An obvious medical consequence of admixture is that the hybrid population is expected to have disease prevalences for **Mendelian** disorders that are intermediate between those of the ancestral populations. When the most frequent diseases differ between the populations, this can lead to an overall lowering of burden of these diseases through a reduction in the probability of having two parents carrying the same deleterious **recessive** allele (**Table 14.2**).

Given the variation in degree of admixture among individuals in an admixed population, the proportion of admixture can be correlated with susceptibility to certain diseases more prevalent in one or other of the ancestral populations. It has been proposed that the prevalence of type 2 diabetes (OMIM 125853) in different Native American populations is positively correlated with the proportion of Native American genes, irrespective of whether this proportion is assessed phenotypically, genealogically, or genetically.[7] In practice, many studies have confirmed that non-diabetic Native Americans have on average significantly higher European admixture in their genomes than those diagnosed with diabetes. However, greater European ancestry also correlates with higher socioeconomic status which can, at least partly, explain the relationship between the disease and ancestry. While it is unlikely that assessing overall levels of individual admixture will have major predictive value of disease for individuals or for drug design (see Box 17.5), inferring the ancestry of particular genomic regions in admixed populations has proved to be a successful method for identifying disease loci (see **Section 14.4**).

For complex diseases, the possibility remains that each ancestral population contains individuals with co-adapted combinations of alleles that will be disrupted by admixture, resulting in a higher burden of disease in the hybrid

TABLE 14.2:
ADMIXTURE CAN REDUCE THE DISEASE BURDEN OF RECESSIVE SINGLE-GENE DISORDERS

Population	Carrier frequency of allele A	Carrier frequency of allele B	Incidence of disease A	Incidence of disease B	Total disease incidence
A	1/10	0	1/400	0	1/400
B	0	1/15	0	1/900	1/900
1:1 admixture	1/20	1/30	1/1600	1/3600	~1/1100

Only a quarter of children with carrier parents will be affected by a recessive disease.

population than in either ancestral population. No example of this **outbreeding depression** has yet been demonstrated in mixed human groups. However, analyses of numbers of children born to Icelandic couples over the past two centuries revealed a positive correlation between kinship and fertility: the highest reproductive success was observed in couples who were related at the level of third and fourth cousins.[15] Given the relatively minor socioeconomic differences among the mostly non-admixed families of Iceland, this correlation may have a biological basis.

There remains another mechanism by which admixture can result in an increased disease burden. Human populations often harbor their own populations of pathogens to which they have previously developed resistance. The release of these pathogens into previously unexposed populations could result in a substantial increase in the incidence and severity of infectious disease. This type of episode is exemplified by the population crashes witnessed in Polynesia and the Americas on first contact with Europeans bearing novel pathogens, such as those causing smallpox and measles. In such cases the resulting selective pressures are expected to result in a substantial bias toward contributions from the resistant ancestral population in the admixed population. In the first generations after admixture this bias would extend toward all genomic loci irrespective of their linkage to the locus conferring disease resistance, although in later generations, as the admixture patterns become more and more fragmented across the genome, this bias would be confined to linked loci.

14.3 DETECTING ADMIXTURE

Methods based on allele frequency can be used to detect admixture

Allele frequency-based methods were among the first tools developed to detect admixture from protein data. The simplest scenario occurs when no alleles are shared between the ancestral populations. Each allele in the hybrid population can then be unambiguously assigned to an ancestral population, and the proportion of admixture calculated by simply counting up the number of alleles assigned to each population. However, an absolute distinction between ancestral populations is rare; more often, alleles are found within many populations at differing frequencies. In principle, it is easy to estimate the proportion of admixture in a hybrid population formed from two ancestral populations (see **Figure 14.3**).

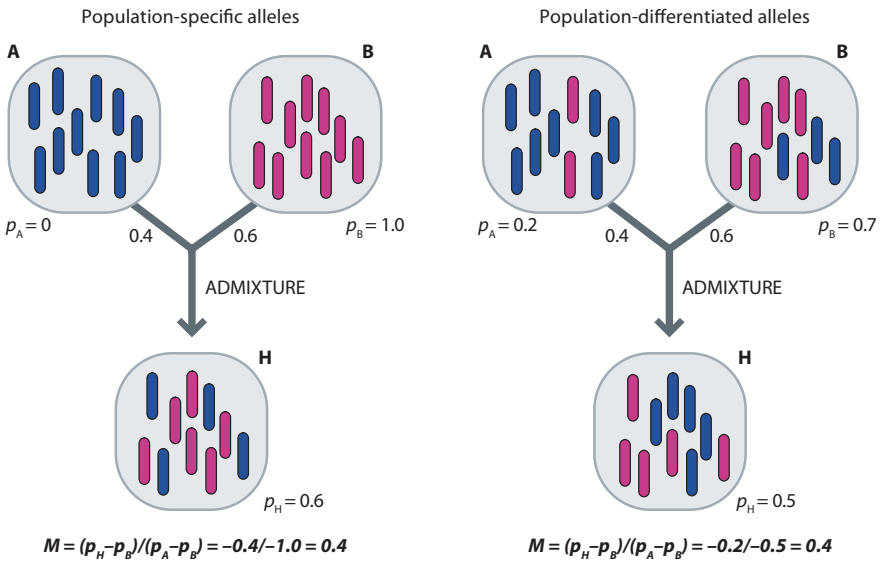


Figure 14.3: Calculating admixture proportions (*M*) when alleles are population-specific, and when they are present in both populations but at different frequencies.
p_A, p_B, and p_H are the frequencies of an allele in the two parental populations A and B, and the hybrid population, H, respectively. The admixture proportions (*M*) are calculated using the equation described in the text, for two different scenarios.

For any given allele, if we know its frequency in the ancestral populations A and B (p_A and p_B) and in the hybrid population (p_H), we can estimate the proportion (M) that ancestral population A contributed to the admixed population by re-arranging the equation[5]

$$p_H = Mp_A + (1 - M)p_B$$

to give

$$M = (p_H - p_B)/(p_A - p_B)$$

Obviously, this approach requires the unambiguous identification of the ancestral populations as well as a number of assumptions, such as a lack of subsequent gene flow, which are either unrealistic or difficult to test in practice. Further complicating questions include:

- If more than one locus is being studied, how should they be averaged?
- What have the effects of genetic drift and selection upon allele frequencies been in all three populations since admixture?
- What if we misidentify the ancestral populations?
- What if there were more than two ancestral populations?
- What if an allele in the admixed population is not found in either ancestral population?

These complications have led to the development of a series of different admixture estimation procedures, which can be classified on the basis of the type of data used (genomewide or locus-specific), the assumed model for admixture, and whether they seek to estimate admixture at the level of the population or the individual. **Figure 14.4** illustrates a number of different admixture models.

Given a set of multi-locus allele frequencies in ancestral and hybrid populations, what is the best way of getting a single estimate of admixture proportions from

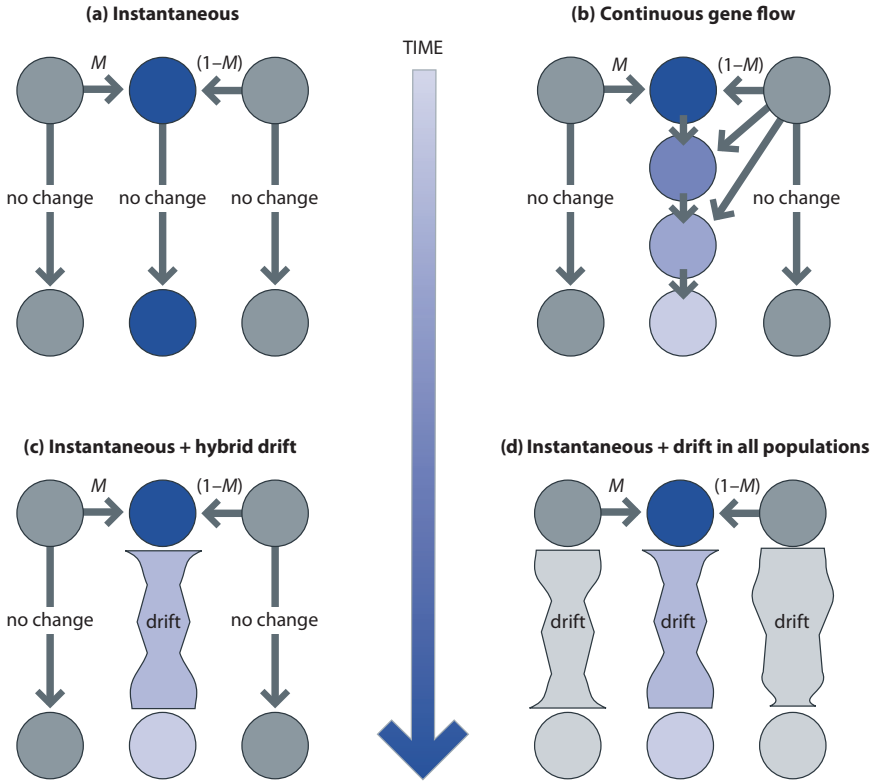


Figure 14.4: Different admixture models of varying complexity.
(a) Instantaneous admixture, (b) cumulative effect of gene flow across many generations, (c) instantaneous admixture allowing for drift in the hybrid population, and (d) instantaneous admixture allowing for drift in all three populations. *Gray* circles, parental populations; *blue* circles, hybrid populations.

these data? Admixture estimates can be calculated for each allele, or locus, individually and then averaged. There are a number of different ways of averaging this information across loci and assessing ancestry from the data.

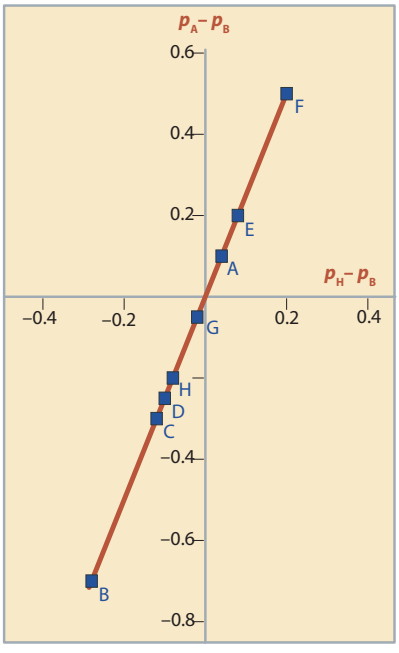
The equation for M given above suggests that estimates of admixture proportions from different alleles should be related linearly. In other words plotting ($p_H - p_B$) against ($p_A - p_B$) for different alleles should give a straight line of gradient M. However, drift, selection, and imprecision of allele frequency estimation can lead to deviations from the linear relationship in real data. **Figure 14.5** shows this property for an idealized admixture situation where all alleles give the same estimate of M.

The plot in Figure 14.5 immediately suggests one method of averaging information from different estimates, namely to plot the least-squares regression line between the points and take its gradient as the multi-locus estimate of admixture.[32] This **estimator** of admixture is often known as $\mathbf{m_R}$.

The above method assumes that the allele frequencies are known without error; it does not take account of the different levels of precision associated with each individual estimate. This can be considered by averaging the different estimates weighted according to their precision, as assessed by their **variances**. These variances depend on the size of the samples. The weighting factor commonly used is the inverse of the variance of the estimate. In other words, the higher the variance of the estimate, the less we are sure that it is accurate and the lower the weight we give it.[6] This weighted average approach takes into account sampling effects on all allele frequency estimates.

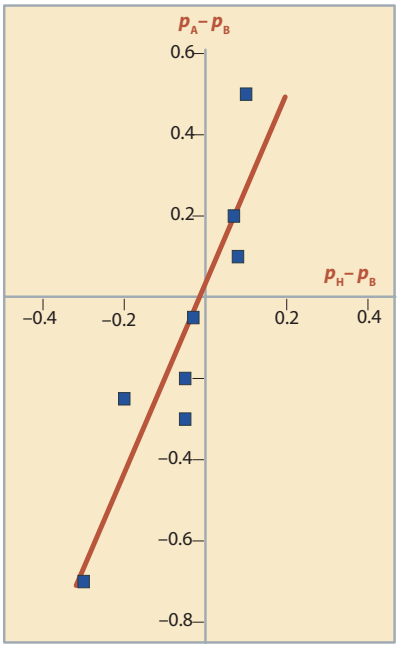
(a) Artificial population

Allele	p_A	p_B	p_H	$p_H - p_B$	$p_A - p_B$
A	0.6	0.5	0.54	0.04	0.1
B	0.1	0.8	0.52	−0.28	−0.7
C	0.3	0.6	0.48	−0.12	−0.3
D	0.45	0.7	0.6	−0.1	−0.25
E	0.35	0.15	0.23	0.08	0.2
F	0.8	0.3	0.5	0.2	0.5
G	0.2	0.25	0.23	−0.02	−0.05
H	0.4	0.6	0.52	−0.08	−0.2



1/gradient = M = 0.4

(b) Real population



1/gradient = M = 0.43

Figure 14.5: Linearity of ($p_H - p_B$) against ($p_A - p_B$).
(a) An idealized case of admixture in which eight alleles (A–H) give exactly the same value for M, and (b) a more realistic set of variant estimates of M from different alleles; M is estimated by fitting a best fit line through the points.

More recently, simulation modeling methods have been developed to infer multiple parameters of population histories in the presence of admixture.[10] However, all methods designed for admixture estimation will produce estimates of *M* even if the ancestral populations have been grossly misidentified, and so care will always be necessary to identify them correctly. This often requires the support of historical, archaeological, and linguistic evidence.

Admixture proportions vary among individuals and populations

A positive estimate of admixture proportions within a hybrid population does not mean that all individuals have the same ancestry ratios. Even if several generations have passed since the admixture event, populations often remain heterogeneous in terms of individual admixture proportions. Similarly, different sampling locations may yield variable estimates if admixture is geographically structured. Besides these inter-individual and interpopulation variations in the amount of *global* genomic estimates of admixture, representing averages over individual genomes, there is also *local* variation at various genes and noncoding loci (**Section 14.4**). Substructure within the admixed population, which can be caused by geographic partitioning, socioeconomic factors, natural selection, and potentially also by **assortative mating**, may reveal itself in unusually large variation in individual admixture estimates as seen, for example, in a number of studies examining admixture in African-Americans (**Table 14.3**). Levels of population admixture can be determined from individual admixture estimates, but not vice versa. How can we investigate this finer-grained admixture?

Calculating individual admixture levels using multiple loci

The actual calculation of individual admixture levels is statistically complex, because information has to be incorporated from many alleles at many loci, and most approaches necessarily have to make simplifying assumptions about past population structure. These include reduction of the number of ancestry groups, ideally down to two parental populations, and reduction of the observed genetic diversity to a minimum number of components. An appreciation of genomewide admixture can only be gained by inferring the ancestry of multiple unlinked loci within that genome, and thus requires substantial genotyping effort. Some alleles may not be particularly well differentiated between different ancestral populations, and thus may not be particularly informative. To reduce the genotyping load many studies have focused on what are called **ancestry informative markers** (AIMs). These are polymorphisms where the allele frequency in the ancestral populations differs considerably (>45%;[37] this work is discussed in a forensic context in **Section 18.2**). Analysis of AIMs has the additional advantage of giving a more accurate estimate of the amount of admixture compared to the same number of less well-differentiated alleles. The high degree of population differentiation exhibited by mtDNA and the nonrecombining portion of the Y chromosome makes them valuable sources of such AIMs. However, both mtDNA and the entire male-specific region of the Y chromosome each represent a single locus, with a single evolutionary history that is not necessarily representative of the rest of the genome.

There are potential problems in focusing solely on AIMs. Because of the low degree of **genetic differentiation** among modern humans, most alleles that exist at moderate to high frequency are not highly population-specific (see **Section 10.2** for more details). Besides drift there are at least two reasons why some alleles do show appreciable population specificity: (1) they result from relatively recent mutations that have not had sufficient time to disperse to any great degree; (2) they result from the action of selection, which influences allele frequency differently in different selective environments. It should also be noted that because of the development of high-throughput genotyping methods there is less need to focus on AIMs and it may be easier and more cost-effective to genotype samples for many polymorphisms genomewide (for example, using a SNP chip) rather than using a small number of custom AIMs.

TABLE 14.3:
ADMIXTURE ESTIMATES IN SOUTH AMERICAN POPULATIONS FOR LOCI WITH DIFFERENT INHERITANCE PATTERNS

Population/locus	% African	% European	% Native American	% Other
Afro-Uruguayan[a]				
Y-chromosomal	30	64	6	
Autosomal	47	38	15	
Mitochondrial	52	19	29	
Brazilian Whites[b]				
Y-chromosomal	3	97	0	
Mitochondrial	28	39	33	
Colombians[c]				
Y-chromosomal	9	79	12	
Autosomal	11	42	47	
Mitochondrial	6	4	90	
US Hispanic[d]				
Y-chromosomal	21	69	8	2
Autosomal	12	61	15	12
Mitochondrial	15	24	49	12
Argentinians[e]				
Y-chromosomal	1	94	5	
Autosomal	4	80	15	
Mitochondrial	2	44	54	

Data from:
[a] Sans M et al. (2002) *Am. J. Phys. Anthropol.* 118, 33.
[b] Alves-Silva J et al. (2000) *Am. J. Hum. Genet.* 67, 444; and Carvalho-Silva DR et al. (2001) *Am. J. Hum. Genet.* 68, 281.
[c] Rojas W et al. (2010) *Am. J. Phys. Anthropol.* 143, 13.
[d] Lao O et al. (2010) *Hum. Mutat.* 31, E1875.
[e] Corach D et al. (2010) *Ann. Hum. Genet.* 74, 65.

Calculating individual admixture levels using genomewide data

High-resolution genomewide genotype data are now available for many populations throughout the world and a number of approaches have been developed to simultaneously assess population structure (**Sections 10.2 and 10.3**) and admixture from such data. One method of reconstructing genetic ancestry from given genotype data, employed by programs such as EIGENSTRAT,[26] is based on **principal component analysis** (PCA; **Section 6.3**). PCA can be used to assess the clustering of individuals or populations at low-dimensional projections of the data. As populations diverge from each other over time by genetic drift, the clusters detected on the low-dimensional scatter plots of PCA are expected to become more and more distinct with decreasing levels of overlap between them. Individuals representing admixture of two distinct ancestral populations would be expected to lie on a cline between the clusters: the exact position of admixed individuals on this cline would be determined by the ancestral contributions of the two parental populations.

Computer simulations show that PCA can also be used to predict contributions to an admixed population from ancestral populations that are either extinct or not available for genotyping for other reasons. **Figure 14.6** illustrates a two-dimensional PCA plot applied to the results of a computer simulation study where hybrid population C was derived from ancestral populations A (extinct) and B (extant) while population D was left to drift without admixture. As a result, individuals from the two non-admixed populations B and D cluster tightly together whereas individuals from the hybrid population C are dispersed between the two ancestral populations. Although population A was not sampled it can be reconstructed as a source of admixture because some individuals in the hybrid group C are still characterized by a high genetic contribution from A. This outcome is important because when a hybrid population is formed from two distinct parental groups, then for a number of generations individuals in the hybrid group will carry different proportions of admixture. This proportion scales linearly with the genetic distance of each individual from their parental populations.

However, accurate and reliable assessment of individual admixture proportions would certainly benefit if data from parental populations were available, particularly in cases where one of them has made a low contribution. Application of PCA methods for inferring admixture to real data should also be carried out with care because alternative demographic models can produce similar patterns: for example, population substructure and higher effective population size of population C in Figure 14.6 could have generated the observed pattern even if ancestral population A never existed. It should be also noted that, when data from only a single hybrid population are available, this method is capable of identifying admixture only where individuals differ from each other in admixture proportions. If admixture is ancient, all individuals are likely to have highly similar admixture proportions and subsequently do not allow admixture detection by PCA.

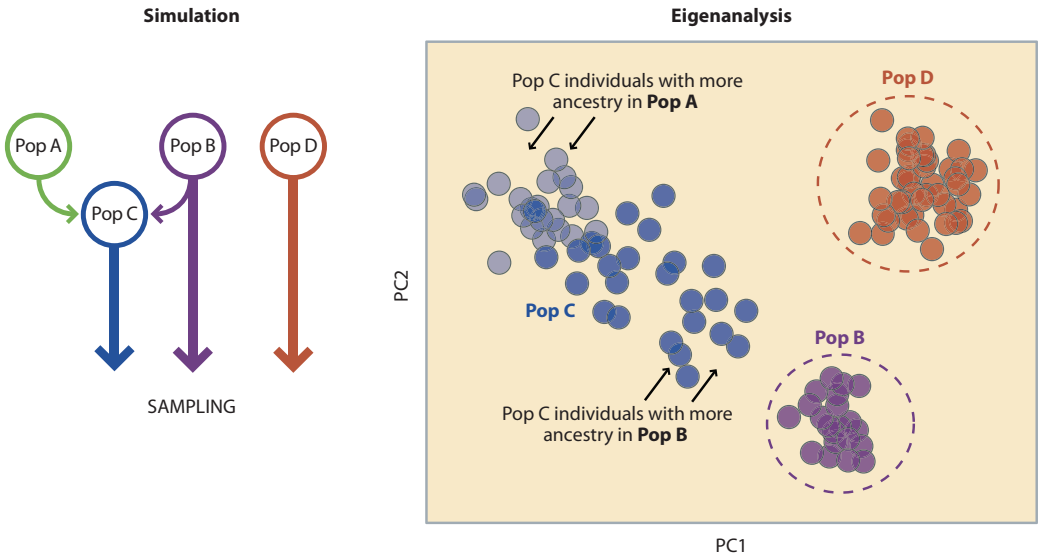


Figure 14.6: Assessment of admixture by PCA.
The plot is based on a computer simulation of multi-locus data (10,000 unlinked variants) where population C was derived by recent admixture from populations A (70%) and B (30%). Individuals from population A were not sampled and included in the PCA. In the PCA each individual is plotted on a scatter plot, with the *x* and *y* coordinates of each individual point representing the values of the first two PCs of that individual's genotype data. Individuals from population C have variable levels of admixture and are dispersed along a line joining the two parental populations. Their position on this line is determined by the admixture proportions and does not change when incorporating population A back in the analyses. [Adapted from Patterson N et al. (2006) *PLoS Genet.* 2, e190. With permission from Public Library of Science.]

Calculating admixture levels from estimated ancestry components

To address admixture events that have occurred many hundreds of generations ago, multi-population ancestry assessment algorithms have been proposed.[31] These methods should be regarded with caution because they rely on specific models of population history, whose assumptions are not always testable. Like the individual ancestry assessment from PCA, these algorithms relate the ancestry proportions in admixed populations to coordinates of a regression line at a low-dimensional projection of the data. As an example, consider PCA of South Asian populations in the context of data from European and East Asian populations (**Figure 14.7a**). While Europeans and East Asians form tight clusters, the individuals from South Asia are dispersed more widely in the plot. Within each population of South Asia the variation is minimal, but when plotted together they form an inverted v-shaped cluster. One arm of this cluster, dubbed the "Indian Cline," stretches out toward European populations. Because North Indian and Pakistani populations appear to be closest to Europe on this cline, one possible interpretation is that these populations have received a higher admixture proportion from a hypothetical Ancestral North Indian (ANI) population that shares its ancestry with populations of Europe and northern Caucasus. In contrast, South Indian populations would be derived largely from a separate, Ancestral South Indian (ASI) source (Figure 14.7b). Assuming such ancestral populations existed, how can we quantify these admixture proportions given that neither of the ancestral groups can be sampled today? One such method,[31] based on the regression line estimated from a four-population statistic, f_4, is illustrated in Figure 14.7c. This method estimated the ancestral ANI contribution in Indian populations as 40–80%.

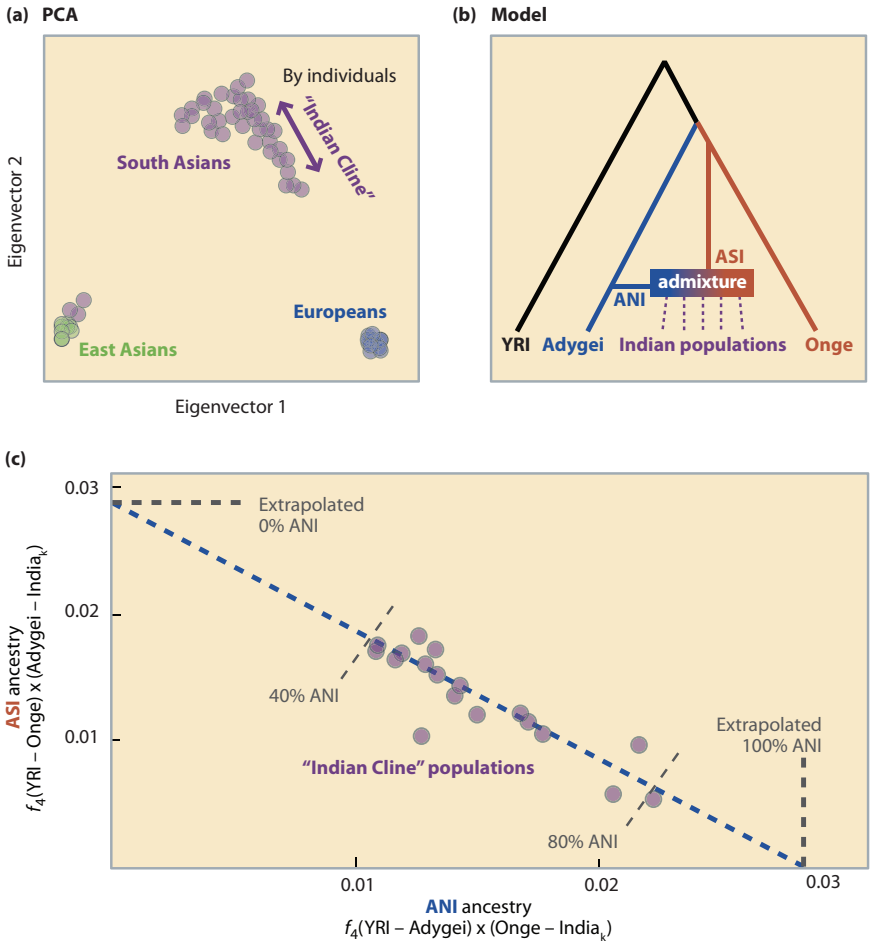


Figure 14.7: Ancestry estimation along the Indian Cline.
(a) PCA plot based on genomewide SNP data assessed in HapMap CEU, CHB, and 22 populations from South Asia. The "Indian Cline" refers to the decreasing genetic distances from the CEU cluster observed from south to northwest Indian and Pakistani populations. (b) Demographically explicit model explaining the extent of admixture of the populations on the "Indian Cline." The split between Onge and Ancestral South Indian (ASI) component is expected to be earlier than the split between Adygei and Ancestral North Indian (ANI). The horizontal line leading to ANI reflects admixture. (c) The proportions of ANI and ASI ancestry components in the Indian and Pakistani populations assessed using f_4 statistics which assesses allele frequency differences between pairs of populations: Yoruban Africans (YRI), Adygei, Onge (Andaman Islands), and a number of South Asian populations tested for admixture. The projections of the Indian and Pakistani population on the regression line are informative about their relative admixture proportions in ANI and ASI. [Adapted from Reich D et al. (2009) *Nature* 461, 489. With permission from Macmillan Publishers Ltd.]

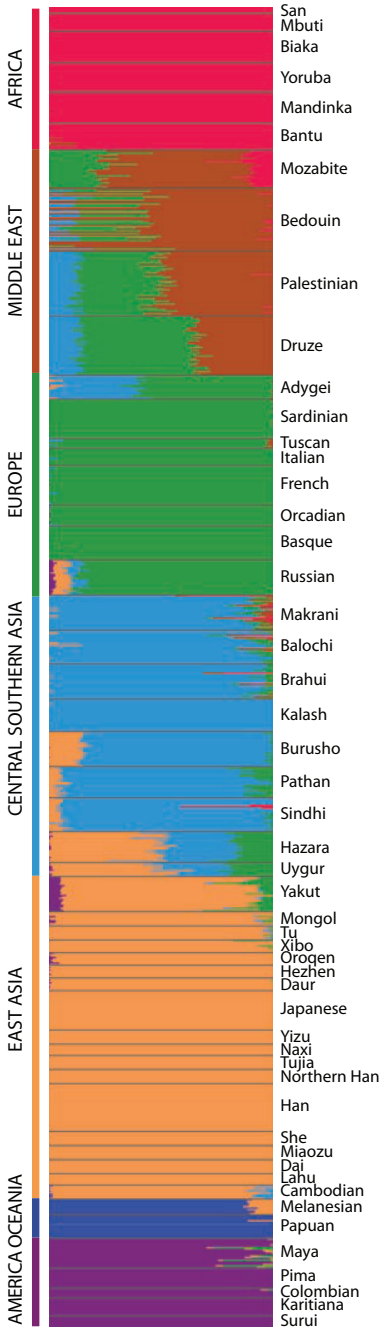


Figure 14.8: Global ancestry profiles of individuals in the HGDP-CEPH panel.
Ancestry as inferred using FRAPPE program at $K = 7$, that is, assuming there are seven clusters. The plot is based on 938 individuals from the HGDP-CEPH panel genotyped for 650,000 SNPs over the genome. Each individual is represented by a horizontal line partitioned into colored segments whose lengths correspond to the ancestry coefficients in the seven ancestry components. Population labels have been added after individual ancestry assessment. [From Li JZ et al. (2008) *Science* 319, 1100. With permission from AAAS.]

In principle, it should be possible in the future to go beyond identification of ancestral populations on PCA plots, and to reconstruct with a certain probability the ancestral genomes themselves from the fragments that survive in modern populations. However, this approach will be feasible only for populations that have gone extinct through processes involving admixture, as is the case, for example, for Tasmanians and many Native American populations.

Clustering methods can take multi-locus genotypes of individuals from several populations and apportion them into well-resolved clusters that are clearly differentiated from one another. One fundamental problem is deciding upon the most likely number of clusters. A number of model-based clustering methods (STRUCTURE, FRAPPE, ADMIXTURE) have been devised that determine the cluster number, the frequency of any given allele in each cluster, and the proportion of each individual's genome that owes ancestry to each cluster.[1, 28, 39] Thus, under the assumption that all loci are in Hardy–Weinberg equilibrium in all populations, STRUCTURE-like approaches (**Section 6.3**) are capable of calculating individual admixture proportions from genetic clusters estimated from the data themselves, rather than from allele frequencies in sampled populations, whose definitions may often rely on external evidence (for example, a shared language or nationality). An example of such a plot where 938 individuals from the 51 populations of the **HGDP-CEPH panel** (Table 10.1, Box 10.3) are allocated to seven ancestry components is shown in **Figure 14.8**. Simulated evolution of two and three populations in the absence of admixture has shown that the clusters formed correspond to the populations themselves and that individuals owe all their ancestry to a single cluster. By contrast, when admixture is included within the simulation, again these clusters are formed but many individuals show some ancestry in between them (**Figure 14.9**).

Although diversity and fine-scale population structure exist both within Africa and Europe, and within Native American populations, it is fairly straightforward to estimate the admixture proportions of each of the three ancestral continental sources in present-day American populations. Earlier, in Chapter 9, we saw how even admixture with Neanderthals is embedded in the genomes of all non-African populations. Thus, a consideration of individual genetic diversity can also reveal cryptic **population structures** that were not previously known to exist. Each individual genome can, in a sense, be considered to be like a palimpsest of multilayered history of successive periods of drift and admixture events in multiple ancestral populations, each of which, in turn, may have had its own multifaceted history. The identification of cryptic population structure is important for other applications. For example, if undetected it can cause spurious associations when hunting disease genes (see Chapter 17) and unreliable match probabilities in forensic situations (see Chapter 18). **Table 14.4** lists software for estimating admixture proportions.

Problems of measuring admixture

Any analysis of admixture is, however, an oversimplification of the population history because of the large number of parameters that are used, some of which need to be fixed at assumed values. Even in cases where it seems that we have succeeded in reducing the ancestry of present populations down to two or a few ancestral populations, it would be naive and wrong to consider

the reconstructed ancestral groups as pure types akin to the nineteenth-century concept of "race." Most clustering algorithms for ancestry detection have been developed to correct for population stratification as a confounding factor in association studies. For such purposes it is useful to detect hidden structuring of the data among **cases and controls** regardless of the meaning of the revealed ancestry components. Their interpretation in terms of demographic histories of populations should be considered with caution.[43] Different opinions have been expressed about whether the ancestry components revealed by the clustering algorithms really reflect past population structure, or are due to sampling at discrete points within an underlying clinal space of variation of human genetic diversity.[14, 34, 36] Some of the admixture components shown in Figure 14.8, for example, would not be compatible with any demographic scenario that would be supported by historical or archaeological evidence: for example, the Native American component detected in Russians.

Natural selection can affect the admixture proportions of individual genes

If all tested loci are evolving neutrally, admixture should affect allele frequencies to an equal degree, since all depend upon the same parameter (*M*, introduced at the beginning of this section). However, selection can bias the frequency of alleles in an admixed population, which often inhabits an environment exerting different selective pressures from that inhabited by either ancestral population. A change in selective pressures will cause a change in allele frequencies at those loci irrespective of any admixture event. Consider an allele that was previously maintained in one of the ancestral populations by **balancing selection** that might now be present in an admixed population in which it has no **heterozygote advantage**. For example, the sickle-cell disease (OMIM 603903) allele (Hb^S) in heterozygotes protects against malaria in Africa (**Section 16.4**), but in the admixed population of African-Americans in the USA, where malaria is absent, is simply deleterious in homozygotes. Consequently, the frequency of

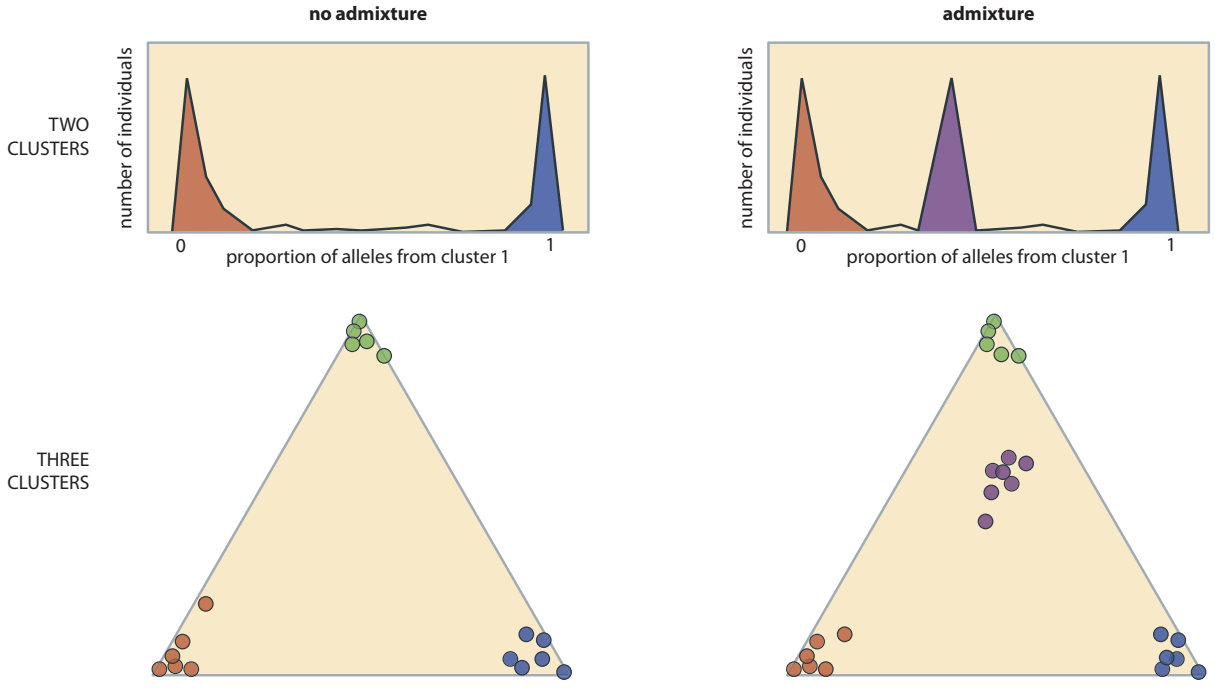


Figure 14.9: Clustering identifies "cryptic" admixture. Schematic graphical representations of how much ancestry each individual traces from each of two or three ancestral populations (clusters). In the absence of admixture the majority of alleles in each individual can be traced to a single cluster. Admixture can be identified when groups of individuals are found to fall between clusters.

TABLE 14.4:
SOFTWARE FOR ADMIXTURE ANALYSIS

Method	Software	URL
Gene identity	ADMIX95	http://www.genetica.fmed.edu.uy/software.htm
Bayesian individual admixture	ADMIXMAP	http://homepages.ed.ac.uk/pmckeigu/admixmap/
Identification of population structure	STRUCTURE	http://pritch.bsd.uchicago.edu/
	FRAPPE	http://med.stanford.edu/tanglab/software/frappe.html
	ADMIXTURE	http://www.genetics.ucla.edu/software/admixture/publications.html
	FINESTRUCTURE	http://www.maths.bris.ac.uk/~madjl/finestructure/chromopainter_info.html
Principal component analysis	EIGENSTRAT	http://genepath.med.harvard.edu/~reich/EIGENSTRAT.htm
Population trees with admixture	TREEMIX	http://pritch.bsd.uchicago.edu/
Coalescent-based simulation of population histories with admixture	SPLATCHE	http://www.splatche.com/
Admixture mapping of complex traits	MALDSOFT	http://pritch.bsd.uchicago.edu/
Identification of ancestry segments in admixed individuals	ANCESTRYMAP	http://genepath.med.harvard.edu/~reich/Software.htm
	SABER	http://med.stanford.edu/tanglab/software/
	HAPMIX	http://www.stats.ox.ac.uk/~myers/software.html
	CHROMOPAINTER	http://www.maths.bris.ac.uk/~madjl/finestructure/chromopainter_info.html

Hb^S in African-Americans is much closer to that in European-Americans than we might otherwise expect. Admixture proportions calculated solely on the basis of this allele give much higher estimates for the contribution of European genes to African-Americans than do other, neutral, loci. Similarly, HLA (Box 5.3) alleles in Puerto Ricans reflect an excess of African and deficiency of European ancestry when compared with global genomewide averages, suggesting adaptive advantage of the African alleles (**Figure 14.10**).

This finding can be turned on its head to provide a means for identifying selection acting upon specific alleles. Heterogeneity among admixture estimates derived from the frequencies of different alleles can be used to pinpoint those alleles whose admixture estimates deviate significantly from those of most others (see **Figure 14.11**).

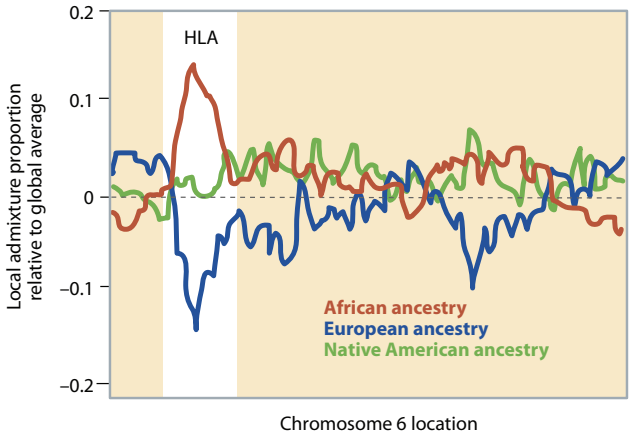


Figure 14.10: Deviations of admixture proportions due to natural selection on HLA alleles.
An excess of African and deficiency of European ancestry at the HLA locus (*white* bar) on chromosome 6. [Adapted from Tang H et al. (2007) *Am. J. Hum. Genet.* 81, 626. With permission from Elsevier; and Oleksyk TK et al. (2010) *Philos. Trans. R. Soc. Lond. B Biol. Sci.* 365, 185. With permission from Royal Society Publishing.]

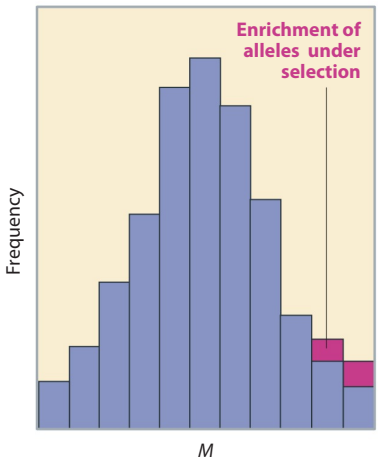


Figure 14.11: Deviant estimates of admixture (*M*) reveal outlier alleles under selection.
A change of selective environments between ancestral and hybrid populations can lead to selection pressures on some alleles resulting in estimates of *M* that are outliers when compared with neutral alleles.

However, selection is not the only evolutionary process that can distort allele frequencies and the admixture estimates derived from them. **Genetic drift** also influences these estimates. Distinguishing between systematic biases in admixture estimates resulting from selection, on the one hand, and random biases introduced by sampling effects and genetic drift, on the other, is far from easy. When searching for outliers of *M* from empirical distributions, as shown in Figure 14.11, caution is required because we do not know what proportion of the genome has been affected by selection and how much variation is due to drift.

14.4 LOCAL ADMIXTURE AND LINKAGE DISEQUILIBRIUM

As we saw in the previous section, admixture proportions can vary in populations due to inter-individual differences, and in addition the genome of each individual can be a composite of clusters of segments that differ in admixture contributions because both selection and drift affect the frequencies of alleles at unlinked loci independently. Over time, the signal of admixture will be divided by recombination into smaller and smaller segments of the genome. **Figure 14.12** illustrates the distinction of the global and local patterns of admixture in the example of the Uyghur population from Central Asia, which has been estimated to derive from an admixture event involving East and West Eurasian sources more than 100 generations ago.

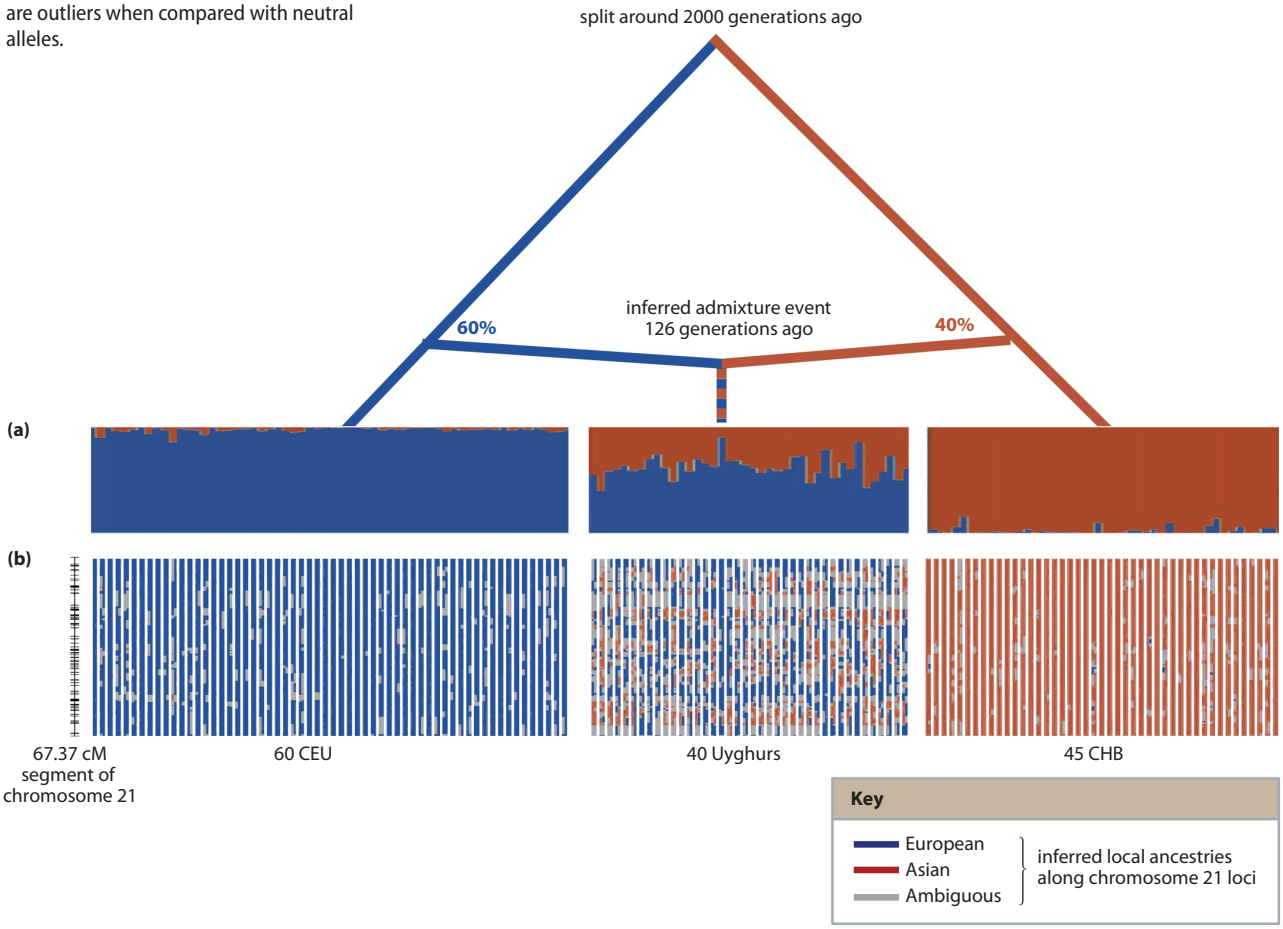


Figure 14.12: Local and global admixture in the Uyghur population.
(a) The global, genomewide admixture contribution from the European source was estimated in the Uyghur population to be around 60% on average. Individuals show a range of variation between 40 and 85%. (b) The local pattern of admixture profile is illustrated on a 67.37 cM (~67 Mbp) segment of chromosome 21. [Adapted from Xu S et al. (2008) *Am. J. Hum. Genet.* 82, 883. With permission from Elsevier.]

How does admixture generate linkage disequilibrium?

In Chapters 3 and 5 we encountered the phenomenon of **linkage disequilibrium** (LD), whereby alleles at different loci tend to be co-inherited more often than we might otherwise expect. Admixture events generate LD between all loci for which differences in allele frequency exist between the two ancestral populations (see **Figure 14.13**). Over the first few generations, LD can be detected even between physically unlinked loci, for example on different chromosomes. However, the detectable association between unlinked loci (self-contradictory though this may sound) dissipates rapidly over a few generations as a result of **chromosomal assortment**. LD at physically linked loci decays more slowly due to recombination events. As a result, recently admixed populations should exhibit LD over greater genetic distances than non-admixed populations. This makes them of potential use for mapping traits and disease genes.

A number of factors affect the extent of LD exhibited by an admixed population. These include:

- The time since admixture
- The admixture dynamics, for example, instantaneous or continuous gene flow
- The relative contributions of different ancestral populations
- The allele frequency differences between ancestral populations
- The pattern of recombination in the human genome

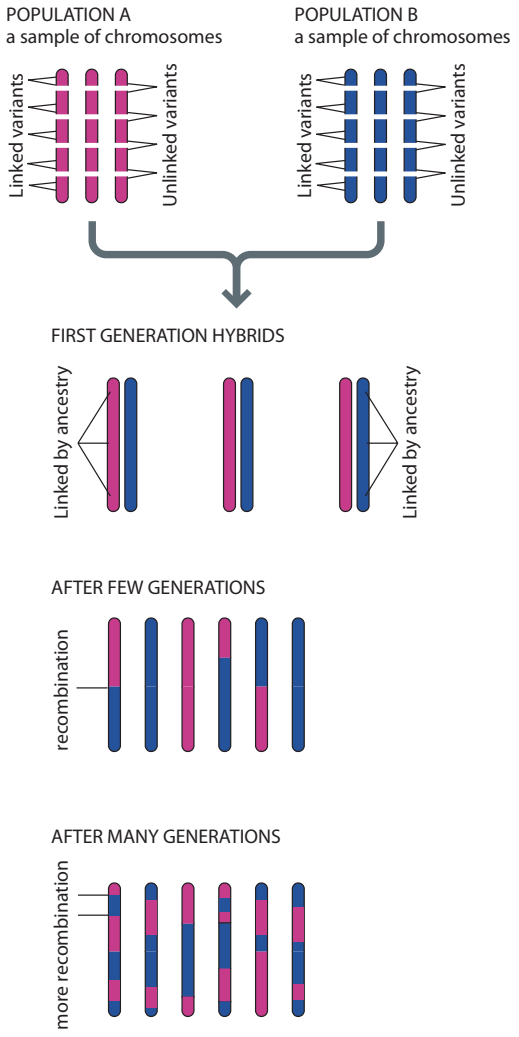


Figure 14.13: Admixture generates linkage disequilibrium (LD). Decline in extent of LD over the generations after admixture. Two ancestral populations (A and B) contribute a sample of three chromosomes to a hybrid population. Three first-generation hybrids show complete linkage of all polymorphic variants on the same chromosome and because of different allele frequencies in A and B the ancestry of each chromosome can be wholly traced back to only one ancestral population. Over time the LD breaks down due to recombination events and the ancestry becomes fragmented, until only closely positioned alleles on the same chromosome are in LD.

Admixture mapping

The approach of mapping genes underlying phenotypic traits by assessing their association with LD caused by admixed ancestry is generally referred to as Mapping Admixture by LD (MALD).[8, 38] A number of MALD algorithms have been developed as statistically powerful alternatives to genome-wide association studies (GWAS) (Table 14.4). The idea behind the approach is that genes associated with a phenotypic trait (or disease) that displays a substantial frequency difference between two parental populations are expected to show pronounced admixture LD among the cases sampled from the hybrid population (**Figure 14.14**). This approach can be successfully applied even as a cases-only study design with a few thousand AIMs and has the power to detect associations with modest odds ratios with sample sizes of only few thousands of individuals.[22, 25] Because the length of the haplotypes resulting from admixture decays

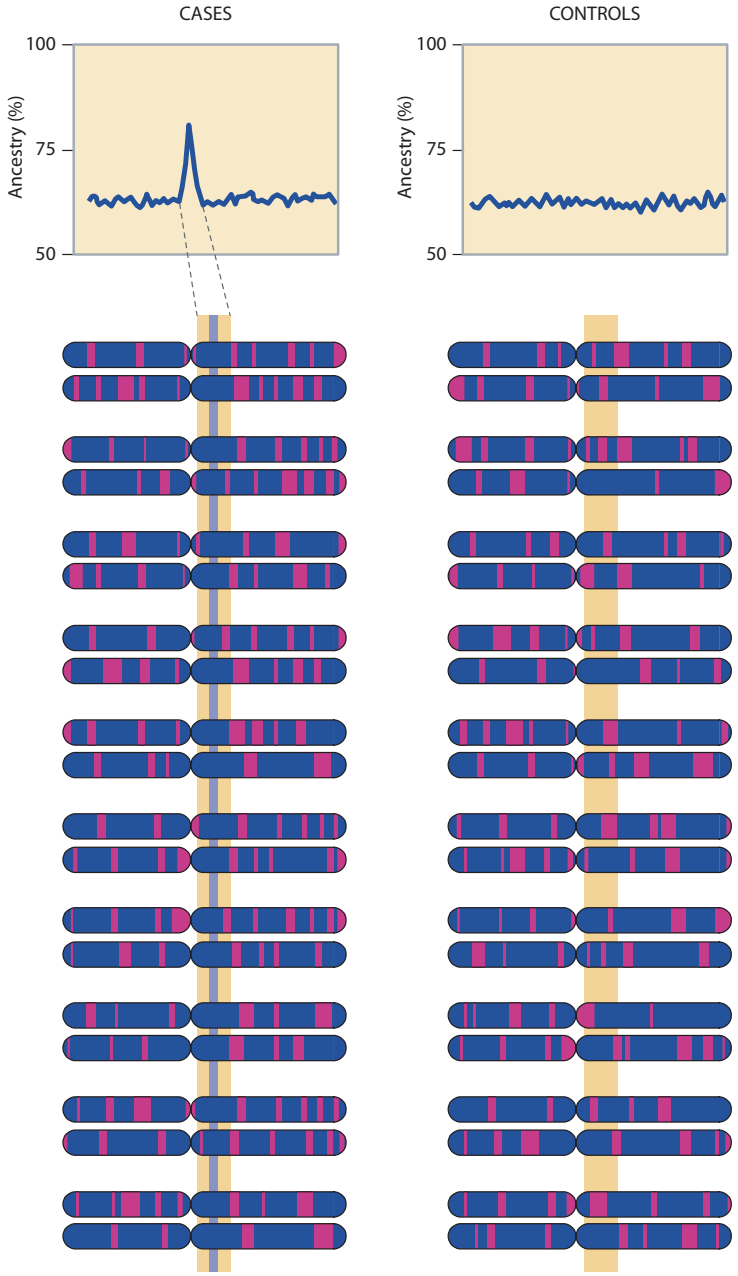


Figure 14.14: Mapping admixture by LD (MALD).
Admixture mapping will highlight genetic loci that show allele frequency differences between two parental populations in association with a disease. Ancestry in the population with the higher disease prevalence is shown in *dark blue*. Individuals with the disease (cases) show higher ancestry at associated loci whereas randomly taken controls from the same hybrid population display an ancestry profile consistent with the genome-average admixture proportions. [From Winkler CA et al. (2010) *Annu. Rev. Genomics Hum. Genet.* 11, 65. With permission from Annual Reviews.]

over time, different polymorphism densities are required for different admixed populations. In the case of African-Americans, for example, where admixture dates mostly to <10 generations ago, admixture LD extends on average for 17 cM,[25] while in Uyghurs (Figure 14.12), where admixture dates to >100 generations ago, the blocks extend only 2–4 cM on average.[45]

Admixture mapping can also be applied to cases involving more than two ancestral populations. Admixture in Latin America is often three-way (as already shown in Figure 14.10) as it involves mixtures of Native American, European, and African ancestries. Individuals from the "colored" population of South Africa can represent five-way admixture combining ancestries from Bantu, Khoe-San, European, South Asian, and Indonesian sources.[44] The greater the number of parental populations and the more ancient the admixture event, the more polymorphisms will be needed to distinguish the ancestry profiles in MALD. Extension of the existing AIM sets is likely to result from the application of new sequencing technologies, for example in the **1000 Genomes Project** (Box 3.2). With current high-throughput genotyping methods it is feasible to genotype cost-effectively sufficient numbers of SNPs across the genome to allow the recovery of even short-range haplotypes for admixture mapping.

A number of genes have been successfully mapped using the MALD approach. These include variation in the *DARC* gene promoter associated with white blood cell counts. The $FY*B^{ES}$ allele is almost fixed in sub-Saharan Africans while its frequency is close to zero in Europe. Without the European-African admixture in the Americas, the mapping of the locus to a 30 cM region on chromosome 1 would have been challenging when it was achieved[20] in 2000. While MALD is most effective in mapping diseases involving a single gene or a few genes, it has also proven successful in mapping candidate regions for some complex disorders, for example prostate cancer. The lack of MALD success so far in mapping genes associated with hypertension and aggressive breast cancer in African-Americans, despite the existence of significant disease prevalence differences among Africans and Europeans, has been attributed to the involvement of many genes with minor impact and/or the effect of nongenetic factors.[44]

Admixture dating

As we saw in the preceding sections, admixture events leave behind a specific imprint on the genomes of the hybrid population in the form of long-range LD patterns. Because the break-up of LD is time-dependent, it is possible to infer with some accuracy the date of admixture from the lengths of the "migrant tracts," that is, chromosome segments with recent migrant ancestry.[27] Several methods have been developed to further examine the relation between time since admixture and the extent of LD decay, while also considering the confounding factor of the background LD that existed in the two parental groups before the admixture. One such method, ROLLOFF, assesses the decline of LD between SNPs at distances greater than 0.5 cM.[23] It examines the decline of correlation between allele frequency differences and LD in specified parental populations for pairs of polymorphisms at increasing genetic distances. The date estimates are obtained by fitting the exponential distribution to these correlations and solving it as a function of the number of generations and genetic distance.[23] Using this method, the 1–3% of African admixture observed throughout southern Europe was dated to around 55 generations, consistent with the historically attested African slave trade practiced by the Roman Empire. Slightly older, 72-generations-old African admixture was detected by the same method among eight Jewish groups.[23] Computer simulations suggest that the method is accurate up to a time-depth of 300 generations (~9 KY) in cases of single and discrete events of admixture. Other methods have been developed to estimate the time since admixture, such as the sliding window wavelet decomposition method for assessing recombination breakpoints that have occurred on chromosomes after admixture.[29] These methods appear to give accurate dates of admixture up to a few hundred generations, after which the LD signal becomes inaccurate for dating. When gene flow between populations occurs continuously

over time or in multiple pulses, these methods are likely to yield date estimates that are more recent than the actual initiation of admixture.

14.5 SEX-BIASED ADMIXTURE

What is sex-biased admixture?

A phenomenon known as **sex-biased admixture** refers to cases where sex-specific loci give different estimates of admixture proportions. Males or females may contribute disproportionate amounts of admixture. In the extreme case, admixture may be restricted to one sex only, so-called **sex-specific admixture**. Sex-biased admixture can result from a sex bias in the makeup of one of the ancestral populations. There are clear examples of this from recent colonial admixture. The European explorers, traders, and missionaries who traveled the world over the past 500 years were predominantly male. As a consequence, admixed populations resulting from these contacts are likely to exhibit male-biased admixture. We saw an unambiguous example of this in Chapter 13, where the paternally inherited Y chromosomes of Cook Islanders in the Pacific are one-third European, but no European admixture appears among their maternally inherited mitochondrial DNA.[16] Sex-biased admixture is not restricted to the scenario of significant sex biases in one ancestral group as outlined above; it may also result from admixture between ancestral populations, neither of which is sex-biased (see **Figure 14.15**). As emphasized in previous sections, we cannot divorce genetic admixture from the wider social context in which it occurs. Ancestral populations rarely have equal status when they encounter one another for the first time. The colonial situation illustrates this clearly. In addition, human populations rarely exhibit random mating, especially across perceived "racial" or socioeconomic boundaries. Such boundaries are often more permeable to one sex than the other, an imbalance that is sometimes dependent on whether the individual is mating "above" or "below" themselves in social status terms. For example, in the Indian caste system, it is easier for a woman to marry a man from a higher caste than vice versa. Directional mating is also apparent in Western societies. In England, the frequency of marriages between white females and African-Caribbean males is greater than that between African-Caribbean females and white males.[21] The social treatment of mixed marriages is an important factor: how are offspring from unions across

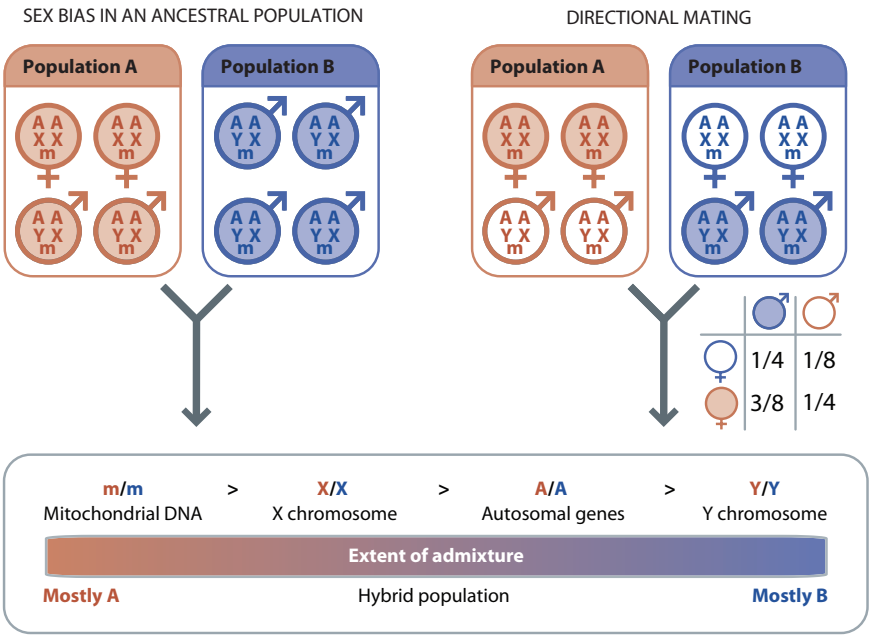


Figure 14.15: Scenarios under which sex-biased admixture may occur. Two different scenarios that lead to sex-biased admixture are shown. Each individual is represented by a pair of autosomes (AA), a pair of sex chromosomes (XX or YX), and a mitochondrial genome (*circle*) colored according to their population affiliation. In each scenario two ancestral populations (A and B) contribute to the admixed population. In the first scenario, population B consists only of males and so contributes no mitochondrial genomes, many Y chromosomes, and intermediate levels of autosomes and X chromosomes to the admixed population. In the second scenario, mating of males from population B with females from population A represents three-quarters of the mixed matings that contribute to the admixed population, and outweighs matings of females from population B with males from population A, leading to a sex bias in the contributions to the admixed population.

status boundaries incorporated into the population? Mixed unions can be stigmatized and therefore incorporated into the lower status group, or ostracized from all groups. All these factors potentially skew the contributions of males and females to all admixed populations. Thus sex-biased admixture may be a relatively common feature of admixture, but how can it be detected?

Detecting sex-biased admixture

Sex-biased admixture will cause admixture estimates from loci with different patterns of inheritance to differ markedly. Thus, admixture estimates are not pooled from all loci, but are compared between loci with different patterns of inheritance. There are four different modes of inheritance in the human genome:

- Exclusively maternal inheritance—mitochondrial DNA
- Exclusively paternal inheritance—Y chromosome
- 2:1 female-biased inheritance—X chromosome
- Equal, **biparental inheritance**—autosomes

If females contributed more than males to an admixed population, estimates of admixture would lie on a gradient (Figure 14.15):

mitochondrial DNA > X chromosome > autosomes > Y chromosome

By contrast, if males contributed a greater proportion, the gradient would be reversed:

Y chromosome > autosomes > X chromosome > mitochondrial DNA

In principle, comparisons between any two of these types of loci can reveal sex-biased admixture. However, for historical and technical reasons, the most common comparisons made in practice have been between either mtDNA and autosomal estimates, or mtDNA and Y-chromosomal estimates. Given the often substantial variance around admixture estimates described above, it makes sense to compare loci at which the greatest differences in admixture estimates should be expected. It is important to take drift into account when comparing mtDNA and Y-chromosomal admixture estimates, which, due to the small effective population sizes of these loci, are prone to **stochastic** fluctuations.

As with all admixture studies, an additional factor to be considered when contrasting admixture estimates from different loci is the allele frequency difference between ancestral populations; the larger this difference, the more power admixture estimation has to detect sex bias. We saw in Chapter 10 that the level of population differentiation differs between loci with different inheritance patterns, with Y-chromosomal polymorphisms exhibiting by far the highest levels of differentiation. This means that, on average, Y-chromosomal polymorphisms will exhibit the greatest difference in frequency between ancestral populations. This makes the inclusion of these polymorphisms particularly attractive for studying sex-biased admixture.

Care must always be taken when comparing loci with different patterns of inheritance because the polymorphisms being analyzed often have different mutational dynamics. Any differences between them may result from the mutation-rate differences rather than the inheritance differences. For example, comparisons are commonly made between mtDNA sequences and autosomal microsatellites (an example is Seielstad et al.[35]), which by virtue of dissimilar mutation dynamics may be more unreliable than comparisons of more similar loci.

Sex-biased admixture resulting from directional mating

A more complex pattern of admixture than the simple model described above is revealed when examining the origins of South American populations. Three major ancestral groups have contributed to modern genetic diversity on this continent: Native Americans, European colonists, and African slaves. Contributions occurred at different times and were from populations of different

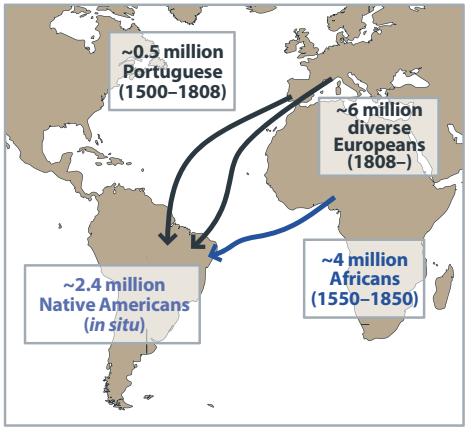


Figure 14.16: Map of genetic sources of Brazilian populations.

sizes. For example, the Native American population of Brazil at the time of its "discovery" by the Portuguese in ~1500 AD was thought to number ~2.4 million. Some half a million European colonists, predominantly male Portuguese, had arrived by 1808. The next two centuries saw the arrival of ~6 million diverse settlers, of which 70% came from Portugal and Italy (other sources included Spain, Germany, Syria, Lebanon, and Japan). Meanwhile, in the 300 years between the mid-sixteenth and nineteenth centuries, some 4 million African slaves were imported into the country. These movements are summarized in **Figure 14.16**.

The present-day populations of Brazil, and the rest of South America, are far from homogeneous. There is substantial population structure, with groups tracing predominant ancestry to different source populations, each with a very different socioeconomic status. A predominantly white middle class tends to occupy a position within society "above" groups with more apparent ancestry from Native Americans or Africans, or both. However, levels of segregation differ greatly between different South American countries. Genetic studies reveal a striking picture of sex-biased admixture in all groups, with directional mating between European males and Native American and African females. Having said that, the male-biased demography of the earliest settlers is also likely to have played a role in establishing the sex-biased admixture among modern populations.

Table 14.3 gives admixture estimates from Y-chromosomal, mitochondrial, and autosomal loci in South and North American populations: an Afro-Uruguayan population claiming predominantly African ancestry, a countrywide Brazilian "white" population, and average countrywide estimates for Colombian and Argentinian populations and for individuals describing themselves as Latinos from the USA. Admixture estimates in these populations were obtained by various methods, including the gene identity method, allele counting, the weighted least-squares approach, STRUCTURE-like analyses, and summary assignments of mtDNA and Y chromosome by their continental affiliations.

Regardless of the different methods being used, a number of general conclusions can be drawn from these results:

- All populations show some level of ancestry from at least three continental sources.
- As expected, autosomal values for admixture lie between mitochondrial and Y-chromosomal estimates.
- European ancestry is found to be consistently greater among Y chromosomes than among mtDNAs.
- African and Native American ancestry is greater among mtDNAs than among Y chromosomes.

Thus all the Latin American groups exhibit the same pattern of directional mating, despite their different socioeconomic status within society. Studies of African-American populations from the USA also demonstrate that European admixture is biased toward males. More surprisingly, low levels of female-biased Native American admixture have been identified in these populations, revealing them to have three ancestral populations, rather than simply being a European-African mix. The events of the past 500 years have clearly had a much more detrimental effect on the frequency of Native American paternal lineages than upon the frequency of their maternal lineages.

The effect of admixture on our genealogical ancestry

As we learned in previous sections, when populations meet, the common outcome is that they admix. Because even low levels of gene flow over generations can lead to substantial gene exchange, it is likely that two individuals even from distant corners of the world can share some genetic variants by descent because of shared relatives within the past few generations. But exactly how closely related are we? Theoretical predictions are that in a panmictic population the individual genealogies coalesce to at least one common ancestor $T \approx \log_2 N$ generations ago. Assuming that the historical population size of Yorkshire and County Durham in the UK was between 1 million and 10 million, everyone from this region is expected to share a common ancestor with Kate Middleton (now Duchess of Cambridge) within the last 20–23 generations (~600–700 years) because her paternal and maternal ancestors were from that area. By the same formula, anyone having relatives in Britain is expected to be related to everyone in Britain, including the British Royal Family, via at least a single connection within the last 26 generations (~800 years). But the human species is certainly not a panmictic population: before the era of air and rail travel people were more sedentary and the proportion of international and intercontinental marriages was low. A computer simulation study attempting to model past gene flow in the world through a complex network of intra- and intercontinental migration rates inferred from historical and archaeological evidence estimated that the most recent common ancestor of all humans, considering all individual genealogies, may have lived as recently as 1415 BC—only 114 generations ago.[33] This estimate suggests that for any individual living today anywhere in the world, including the Tasmanians who were physically isolated from the rest of the world for ~12 KY but experienced European admixture over the past six generations, there would be a common ancestor only a couple of thousand years ago, shared with everyone else in the world. On the other hand, how much genetic information have we inherited from such a distant common ancestor? The answer is not much: any one particular ancestor who lived 20 generations ago was one out of 2^{20} ancestors, which means that he or she has passed down to us any particular gene with a probability less than one in a million. Given the limited number of recombination events over this short time period there is also a substantial variance in the amount of genome contributed by the set of ancestors. It is quite likely that even relatively recent ancestors can have left no genomic trace in an individual genome.

14.6 TRANSNATIONAL ISOLATES

When populations meet, the extent of admixture measurable from allele frequency patterns can vary quite substantially. In contrast to the populations that were the focus of the above sections, **isolated populations** are those that by virtue of their geography, history, and/or culture have experienced little gene flow with surrounding populations. Isolation due to geographic barriers has led some island populations from the Indian Ocean, for example the Andaman Islanders, to develop and maintain their unique allele frequency and phenotypic characteristics.[31, 40] In the case of mainland isolates, the term "isolation" is relative. No threshold of per-generation gene flow has been set that defines a population isolate; rather, a population is isolated when its surrounding populations more readily exchange genes with one another than with the isolate. This

isolation can be revealed by unusual allele frequencies within the population, compared with surrounding populations, and is often associated with linguistic and geographical boundaries. For example, Basques and Finns are often regarded as population isolates within Europe. As a result of their isolation, population isolates often have unusually high frequencies of some typically rare genetic diseases. The Finnish genetic **disease heritage** is discussed in greater detail in Chapter 16.

Population isolates are commonly restricted to defined geographical regions. Luba Kalaydjieva has coined the term **transnational isolates** to refer to groups which, despite a widespread geographical distribution, remain isolated, largely through the social practice of **endogamy**. While the word was first used to describe the particular situation of European Roma, here it is extended to other groups. As with traditionally defined population isolates, the genetic coherence of transnational isolates is readily apparent in their common disease heritage.

The paradox between the genetic coherence and geographical dispersal of these transnational isolates could result from one of two processes: either coherence is actively maintained through mating over large distances, but within the group, or coherence results from a recent migration from a common point of origin, but is decaying over time. As we shall see, the latter process is the more frequent explanation.

Roma and Jews are examples of widely spread transnational isolates

European Roma

European Roma, often called Gypsies, represent a population of about 8 million spread over the European continent. They are found in highest concentrations in southeastern Europe and the Iberian Peninsula. Historical records indicate they entered Europe about 1000 YA, gradually spreading across the continent from the southeast (see **Figure 14.17**). The Roma speak a variety of Romani dialects although some have adopted the languages of surrounding populations. The linguistic affinities of Romani languages indicate an origin somewhere on the Indian subcontinent.

The social structure of the Roma is orientated around small, endogamous groups, often associated with a specific trade and religion. However, these religions differ greatly between groups. Islam, Roman Catholicism, Protestantism, and the Eastern Orthodox Church are all represented among European Roma.

A wide variety of Mendelian disorders are found among the Roma, such as Lom type of motor and sensory neuropathy (OMIM 601455) and spinal muscular atrophy (OMIM 253300). These disorders are characterized by homogeneity, with single mutations underlying most cases. These mutations each tend to exist on a single **haplotypic background**, thus indicating a single common and recent origin. Some of these disorders are common in other European populations, while others are specific to the Roma. Some of the Roma-specific mutations are found only in certain groups, whereas others are spread across all European Roma.[17] Thus, while there is an obvious **founder effect** resulting from a common origin, there is also substantial heterogeneity between groups resulting in considerable internal diversity. Genetic distances between Roma groups both within and between countries are typically larger than those between the surrounding European populations. Thus individual Roma groups can be thought of as isolates within a larger isolate. Three processes could cause this population differentiation:

- High levels of drift due to endogamous practices in small populations
- Different levels of admixture between Roma groups and the surrounding populations
- Original substructure in the ancestral Roma population, maintained over time

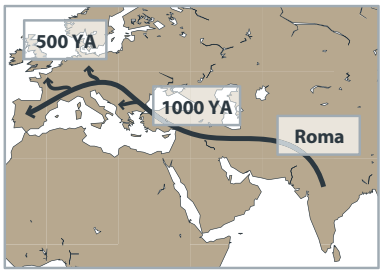


Figure 14.17: The Roma diaspora. The recent Indian origins of the Roma are supported by historical, linguistic, and genetic evidence. Approximate dates derive from historical records.

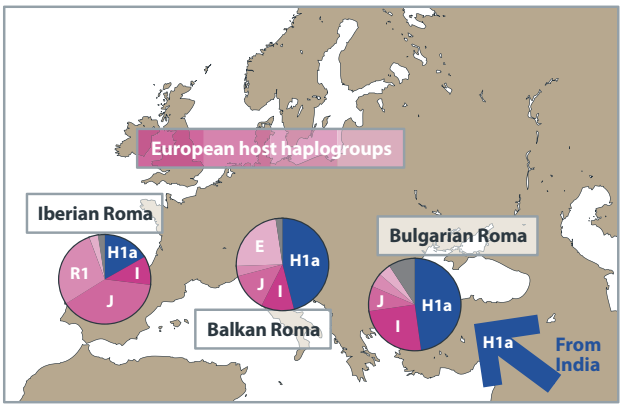


Figure 14.18: Frequencies of Indian and European-specific Y-chromosomal lineages among different Roma groups. [Data from Gresham D et al. (2001) *Am. J. Hum. Genet.* 69, 1314; Gusmao A et al. (2008) *Ann. Hum. Genet.* 72, 215; Regueiro M et al. (2011) *Am. J. Phys. Anthropol.* 144, 80.]

As a consequence of these issues, population genetic studies have tended to focus on similarities between the different Roma groups and their admixture with surrounding European populations after dispersal from their common source. Genetic evidence from classical, Y-chromosomal, and mtDNA studies has shown that the Roma share alleles and lineages with populations on the Indian subcontinent that are not found in other European populations. Nevertheless, it has not been possible to identify a single likely ancestral population. This inability severely hampers our ability to perform the kinds of quantitative admixture analyses examined previously in this chapter.

However, by observing the frequencies of European-specific Y-chromosomal lineages among different Roma populations, some conclusions can be drawn about the nature of the admixture (**Figure 14.18**). First, the degree of admixture is highly variable between different populations, while showing a general decline of the otherwise Indian-specific haplogroup H1a frequency from southeast to western Europe. Second, the admixed lineages reflect the lineage distributions within surrounding populations. It can therefore be inferred that multiple independent admixture events have occurred in the different populations, and that admixture has played a significant role in population differentiation among the different Roma groups.

The Jews

A common religion, language, and traditions unite the Jewish people. Historical and linguistic evidence attests to their Bronze Age origins in the Middle East. The ~14 million modern Jews reside mostly in the USA (~6 million) and Israel (~5 million). Despite the maternal inheritance of "Jewishness" prescribed by religious law, the practice of endogamy ensures a degree of both paternal and maternal genetic continuity; the Jewish religion does not seek converts with the same enthusiasm as some other faiths. Jews are generally classified into three groups on the basis of their ancestral migrations (see **Figure 14.19**).

Ashkenazi Jews ("Ashkenaz" is the medieval Hebrew name for the land around the Rhine Valley) had migrated from the Middle East into Central Europe during the early Middle Ages and subsequently moved within northern Europe, often attempting to avoid persecution. During the nineteenth and twentieth centuries many Ashkenazi Jews left Europe for the Americas, Australia, and South Africa, and as a consequence they now make up ~90% of the US Jewish population.

Sephardic Jews ("Sepharad" in Hebrew meaning "Spanish") had resided in the Iberian Peninsula for centuries prior to being persecuted by the Spanish Inquisition during the fifteenth century. This led to their dispersal to mainly Mediterranean countries (Italy, the Balkans, North Africa, Turkey, and Lebanon), Syria, and the Americas.

Middle-Eastern (Oriental) Jews remained in the Levant (lands on the Eastern edge of the Mediterranean Sea) and surrounding countries (Iran, Iraq, and the Arabian Peninsula).

Based on the analyses of genomewide SNP allele frequencies and autosomal population structure, most Jewish populations carry a composite of ancestry components characteristic of the Near East and their historical host populations (Figure 14.19) while in a fine-resolution PCA all Jewish groups, except for Ethiopian and Indian Jews (additional small groups not listed above), formed a tight cluster overlapping with the Druze and Cypriot populations.[4]

The Mendelian disorders of Jewish people have been investigated in great depth. A plethora of genetic diseases have been identified which typically result from one or two common founder mutations. As with the Roma, some of these mutations are common to many Jewish groups whereas others are specific to certain populations. For example, Tay-Sachs disease (OMIM 272800) is prominent only among Ashkenazi Jews where the responsible mutation reaches a heterozygote carrier frequency of 1/25, whereas the carrier frequency of alleles responsible for familial Mediterranean fever (OMIM 249100) is between 1/10 and 1/5 in populations from all three major Jewish groups. The likelihood that an Ashkenazi Jew is a carrier for one of the eight most common disease alleles is 1/4. It has been suggested that either the common mutations derive from mutational events at different times during the Jewish diaspora, or that genetic drift has resulted in the loss of the disease alleles from some populations, perhaps as a result of founder effects. The dating of different disease alleles to different times supports the former explanation.[24] However, as an additional complication, some disease alleles appear to have been recent introductions via admixture, as they are common in surrounding populations, but not in other Jewish groups. The utility of population isolates in identifying disease-related genes is explored in greater detail in **Section 16.2**.

Many haplotypes have been identified that are shared by all three Jewish populations but not with their surrounding populations. This provides considerable

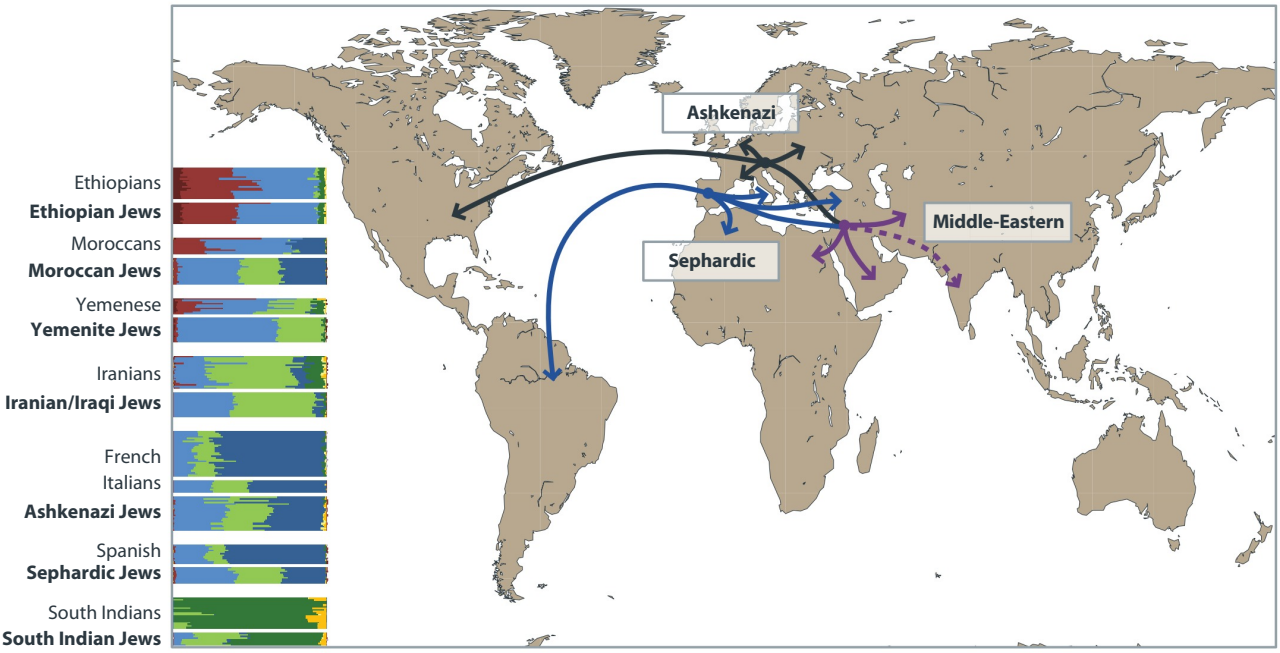


Figure 14.19: The Jewish diaspora.
The tripartite division of Jewish peoples based on their history of migrations is indicated by arrows. Population structure inferred with the ADMIXTURE program at $K = 8$ is shown at the bottom-left corner of the figure. *Light blue* and *light green* colored components, characteristic to Near Eastern populations, can be found, to various degrees, being detected in most Jewish groups. [Adapted from Behar DM et al. (2010) *Nature* 466, 238. With permission from Macmillan Publishers Ltd.]

TABLE 14.5:
GENETIC DIVERSITY AMONG DIVERSE JEWISH POPULATIONS AND THEIR HOST, NON-JEWISH, POPULATIONS

Jewish population	Host population	Less genetic diversity in Jewish compared with host population?	
		Y-chromosomal	Mitochondrial
Ashkenazi	German	no	yes
Moroccan	Berber	yes	yes
Iraqi	Syrian	no	yes
Georgian	Georgian	yes	yes
Bukharan	Uzbekistani	yes	yes
Yemeni	Yemeni	no	yes
Ethiopian	Ethiopian	no	yes
Indian	Hindu	no	yes

Data from Thomas MG et al. (2002) *Am. J. Hum. Genet.* 70, 1411.

support for the common origin of these groups and the partial maintenance of their genetic integrity through endogamous practices over the past 2 KY. One specific Y-chromosomal haplotype that appears to have been co-inherited with the paternally inherited Cohanim priesthood[13, 41] is discussed in Chapter 18.

For calculating admixture estimates for Jewish populations, one ancestral population, that of the Middle East, can be assigned with relative confidence. Given the different migratory histories of Ashkenazi and Sephardic Jews, other genetic contributions could come from a number of other ancestral host populations. The maternal inheritance of "Jewishness" suggests that admixture might be lower in maternal lineages than in paternal lineages, and thus it is of interest to compare admixture estimates for Y-chromosomal and mtDNA polymorphisms. A comparison of genetic diversity among diverse Jewish and non-Jewish host populations showed that whereas Y-chromosomal diversity was not generally lower than that in the host population, mtDNA diversity was frequently significantly lower in Jewish populations[42] (see **Table 14.5**). This suggests either (1) that founder effects and other causes of genetic drift were stronger in maternal lineages; or (2) that admixture has been consistently male-biased, resulting in the introduction of paternal lineage diversity. As we saw in **Section 10.2**, higher drift in maternal than in paternal lineages is an unusual finding; globally, the Y chromosome shows the greatest genetic differentiation of all loci. Ashkenazi Jews, unlike Sephardic or Oriental Jews, are indeed characterized by severe founder effect(s) as evidenced by their characteristically high frequency of mtDNA haplogroup K.[2, 3]

SUMMARY

- Genetic admixture is the process by which a hybrid population is formed from contributions by two or more parental, or ancestral, populations. It is likely that every human population has been influenced by admixture, and many important issues in recent human evolution can be considered to be questions of admixture.
- Past admixture events can be identified in the historical, linguistic, and archaeological records as well as through their impact on patterns of genetic diversity.

- The genetic contribution of an ancestral population to the hybrid population can be estimated by considering the frequencies of a given allele in the hybrid population and both ancestral populations. Admixture estimates require the correct identification of the ancestral populations and are most accurate when an allele is present at very different frequencies in the two ancestral populations.
- A number of different methods for estimating admixture proportions have been devised to best combine information from multiple alleles into a single estimate, and that take into account some of the confounding factors, such as genetic drift in both ancestral and hybrid populations since the admixture event.
- In an admixed population, individual genomes are themselves admixed to a greater or lesser degree. By typing many polymorphisms, levels of individual admixture can be estimated and compared between individuals and regions of the genome.
- Through admixture mapping, different genes and genomic regions can be studied to reveal associations with phenotypes and disease.
- Admixture results in elevated levels of linkage disequilibrium, which decays over time. This relationship between the extent of LD and time can be used for dating admixture events.
- Under a number of different admixture scenarios, the contributions of males and females from ancestral populations may not be equal. This sex-biased admixture reveals itself in discrepant admixture estimates from loci with different patterns of inheritance (that is, mtDNA, Y chromosomes, X chromosomes, and autosomes). Sex-biased admixture is commonly observed in the formation of admixed populations, for example, as a result of the Atlantic slave trade.
- Admixture cannot be divorced from its social context. Social practices such as endogamy restrict admixture with other populations, and the relative socioeconomic status of the different ancestral populations can lead to directional mating between males and females of the two groups.
- Populations that live and breed in isolation due to geographic, cultural, or socioeconomic barriers experience less admixture than others. Transnational isolates (for example, the Roma and the Jews) are populations that maintain genetic coherence over vast geographical distances as a result of recent dispersal from a common origin and endogamous mating practices that restrict admixture.

QUESTIONS

Question 14–1: Consider a situation where a group of individuals from population A moves into the territory of population B with much larger population size and remains relatively isolated there. Each generation 95% of the individuals in the isolated group will mate with each other and only 5% will mate with individuals from population B and remain in the isolated community. What will be the expected ancestry proportions of A and B in that deme after (a) 10 generations, and (b) 100 generations?

Question 14–2: Smokers with A/A genotype (rs762551) in the *CYP1A2* gene metabolize caffeine 1.6 times faster than other genotypes. The A allele frequency is 70% in Europe and 50% in India. You collect samples from second-generation Indians from a number of European cities and determine their A-allele frequency to be 52% on average. What would be your estimate of admixture? If, instead, the estimated allele frequency in your sample of European Indians was 48%, how would you explain the result?

Question 14–3: What are the key intrinsic and technology-related limitations to admixture inference and dating?

Question 14–4: Discuss the benefits and limitations of admixture mapping.

Question 14–5: Using the HapMap browser (http://hapmap.ncbi.nlm.nih.gov/ choose Phase 1, 2, and 3 data merged) and the formula for determining admixture rate M (Section 14.3) estimate the proportion of European admixture in African Americans (ASW) and Mexicans (MEX) on the basis of allele frequency data of the following two SNPs, rs1426654 and rs2227282. For parental sources consider data from the CEU, YRI, and CHB samples.

Question 14–6: Below are mtDNA and Y-chromosome haplogroup (hg) data for populations from three places: Greenland, Madagascar and the Cook Islands. Use information in

the Appendix to deduce: a) which population is which, and b) whether the data suggest a history of admixture, and if so, of what nature.

	mtDNA haplogroup %										Y haplogroup %									
	A	B	E	F	L0/ L1	L2	L3	M	Q	R	B	C	E	I	J	K	L	O	Q	R
Pop. A		80							20			50		10		12				28
Pop. B	100											3	1	29	1				39	27
Pop. C		27	4	4	8	14	15	26		1	8		65		2		1	21		3

REFERENCES

The references highlighted in purple are considered to be important (for this chapter) by the authors.

1. **Alexander DH, Novembre J & Lange K** (2009) Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664.
2. **Behar DM, Metspalu E, Kivisild T et al.** (2006) The matrilineal ancestry of Ashkenazi Jewry: portrait of a recent founder event. *Am. J. Hum. Genet.* **78**, 487–497.
3. **Behar DM, Metspalu E, Kivisild T et al.** (2008) Counting the founders: the matrilineal genetic ancestry of the Jewish Diaspora. *PLoS One* **3**, e2062.
4. **Behar DM, Yunusbayev B, Metspalu M et al.** (2010) The genome-wide structure of the Jewish people. *Nature* **466**, 238–242.
5. **Bernstein F (**1931) Comitato Italiano per o Studio dei Problemi Della Populazione, pp 227–243. Roma: Instituto Poligrafico dello Stato.
6. **Cavalli-Sforza LL & Bodmer WF** (1971) The Genetics of Human Populations, pp 227–243. W.H. Freeman.
7. **Chakraborty R** (1986) Gene admixture in human populations: models and predictions. *Yearbk Phys. Anthropol.* **29**, 1–43.
8. **Chakraborty R & Weiss KM** (1988) Admixture as a tool for finding linked genes and detecting that difference from allelic association between loci. *Proc. Natl Acad. Sci. USA* **85**, 9119–9123.
9. **Choudhry S, Coyle NE, Tang H et al.** (2006) Population stratification confounds genetic association studies among Latinos. *Hum. Genet.* **118**, 652–664.
10. **Cornuet JM, Santos F, Beaumont MA et al.** (2008) Inferring population history with DIY ABC: a user-friendly approach to approximate Bayesian computation. *Bioinformatics* **24**, 2713–2719.
11. **Crow JF** (1983) Surnames as markers of inbreeding and migration. Discussion. *Hum. Biol.* **55**, 383–397.
12. **Glass B & Li CC** (1953) The dynamics of racial intermixture; an analysis based on the American Negro. *Am. J. Hum. Genet.* **5**, 1–20.
13. **Hammer MF, Behar DM, Karafet TM et al.** (2009) Extended Y chromosome haplotypes resolve multiple and unique lineages of the Jewish priesthood. *Hum. Genet.* **126**, 707–717.
14. **Handley LJ, Manica A, Goudet J & Balloux F** (2007) Going the distance: human population genetics in a clinal world. *Trends Genet.* **23**, 432–439.
15. **Helgason A, Palsson S, Gudbjartsson DF et al.** (2008) An association between the kinship and fertility of human couples. *Science* **319**, 813–816.
16. **Hurles ME, Irven C, Nicholson J et al.** (1998) European Y-chromosomal lineages in Polynesians: a contrast to the population structure revealed by mtDNA. *Am. J. Hum. Genet.* **63**, 1793–1806.
17. **Kalaydjieva L, Gresham D & Calafell F** (2001) Genetic studies of the Roma (Gypsies): a review. *BMC Med. Genet.* **2**, 5.
18. **Kaplan B** (1954) Environment and human plasticity. *Am. Anthropologist* **56**, 781–799.
19. **Kayser M, Lao O, Saar K, et al.** (2008) Genome-wide analysis indicates more Asian than Melanesian ancestry of Polynesians. *Am. J. Hum. Genet.* **82**, 194–198.
20. **Lautenberger JA, Stephens JC, O'Brien SJ & Smith MW** (2000) Significant admixture linkage disequilibrium across 30 cM around the FY locus in African Americans. *Am. J. Hum. Genet.* **66**, 969–978.
21. **Model S & Fisher G** (2002) Unions between blacks and whites: England and the US compared. *Ethnic Racial Stud.* **25**, 728–754.
22. **Montana G & Pritchard JK** (2004) Statistical tests for admixture mapping with case-control and cases-only data. *Am. J. Hum. Genet.* **75**, 771–789.
23. **Moorjani P, Patterson N, Hirschhorn JN et al.** (2011) The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* **7**, e1001373.
24. **Ostrer H** (2001) A genetic profile of contemporary Jewish populations. *Nat. Rev. Genet.* **2**, 891–898.
25. **Patterson N, Hattangadi N, Lane B et al.** (2004) Methods for high-density admixture mapping of disease genes. *Am. J. Hum. Genet.* **74**, 979–1000.
26. **Patterson N, Price AL & Reich D** (2006) Population structure and eigenanalysis. *PLoS Genet.* **2**, e190.
27. **Pool JE & Nielsen R** (2009) Inference of historical changes in migration rate from the lengths of migrant tracts. *Genetics* **181**, 711–719.
28. **Pritchard JK, Stephens M & Donnelly P** (2000) Inference of population structure using multilocus genotype data. *Genetics* **155**, 945–959.
29. **Pugach I, Matveyev R, Wollstein A et al.** (2011) Dating the age of admixture via wavelet transform analysis of genome-wide data. *Genome Biol.* **12**, R19.
30. **Reed TE** (1969) Caucasian genes in American Negroes. *Science* **165**, 762–768.

31. **Reich D, Thangaraj K, Patterson N et al.** (2009) Reconstructing Indian population history. *Nature* **461**, 489–494.

32. **Roberts DF & Hiorns RW** (1962) The dynamics of racial intermixture. *Am. J. Hum. Genet.* **14**, 261–277.

33. **Rohde DL, Olson S & Chang JT** (2004) Modelling the recent common ancestry of all living humans. *Nature* **431**, 562–566.

34. **Rosenberg NA, Mahajan S, Ramachandran S et al.** (2005) Clines, clusters, and the effect of study design on the inference of human population structure. *PLoS Genet.* **1**, e70.

35. **Seielstad MT, Minch E & Cavalli-Sforza LL** (1998) Genetic evidence for a higher female migration rate in humans. *Nat. Genet.* **20**, 278–280.

36. **Serre D & Pääbo S** (2004) Evidence for gradients of human genetic diversity within and among continents. *Genome Res.* **14**, 1679–1685.

37. **Shriver MD, Smith MW, Jin L et al.** (1997) Ethnic-affiliation estimation by use of population-specific DNA markers. *Am. J. Hum. Genet.* **60**, 957–964.

38. **Stephens JC, Briscoe D & O'Brien SJ** (1994) Mapping by admixture linkage disequilibrium in human populations: limits and guidelines. *Am. J. Hum. Genet.* **55**, 809–824.

39. **Tang H, Peng J, Wang P & Risch NJ** (2005) Estimation of individual admixture: analytical and study design considerations. *Genet. Epidemiol.* **28**, 289–301.

40. **Thangaraj K, Chaubey G, Kivisild T et al.** (2005) Reconstructing the origin of Andaman Islanders. *Science* **308**, 996.

41. **Thomas MG, Skorecki K, Ben-Ami H et al.** (1998) Origins of Old Testament priests. *Nature* **394**, 138–140.

42. **Thomas MG, Weale ME, Jones AL et al.** (2002) Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. *Am. J. Hum. Genet.* **70**, 1411–1420.

43. **Weiss KM & Long JC** (2009) Non-Darwinian estimation: my ancestors, my genes' ancestors. *Genome Res.* **19**, 703–710.

44. **Winkler CA, Nelson GW & Smith MW** (2010) Admixture mapping comes of age. *Annu. Rev. Genomics Hum. Genet.* **11**, 65–89.

45. **Xu S, Huang W, Qian J & Jin L** (2008) Analysis of genomic admixture in Uyghur and its implication in mapping strategy. *Am. J. Hum. Genet.* **82**, 883–894.