This page intentionally left blank
to match pagination of print book

ORIGINS OF MODERN HUMANS

CHAPTER NINE

This chapter presents our current understanding of where and when modern humans arose. In the last two chapters, we saw that the split between humans and our closest living relatives, chimpanzees and **bonobos**, occurred approximately 6–7 MYA, and that some of the important genetic changes on the human lineage could be identified. Here, we will consider events after the chimpanzee–human split, leading to the origin of our species, modern humans, a much more recent event that occurred within the last 200 KY. We will consider evidence of many kinds: for much of the period, the only substantial source is the fossil record, and we will have to confront a plethora of names and opinions in order to extract the crucial points. After ~2.6 MYA, the exquisitely rare fossil finds start to be supplemented by a more abundant record, archaeology, and more detailed inferences about how our ancestors behaved become possible. And since the genetic diversity apparent among humans alive today has mostly accumulated over the past 1 MY or so, after this time we can begin to make use of the ever-expanding genetic information that forms the core of this book. We will see again and again the overwhelming importance of Africa for human evolution, as the place where both our genus *Homo* and species *H. sapiens* originated, and the continent that contains most of our genetic diversity and the roots of most of our genetic lineages. While it is clear that there were multiple dispersals of our genus out of Africa, we will have to consider the extent to which early expansions contributed to contemporary human gene pools, a question debated for decades and initially presented as a choice between a multiregional model, in which the early migrants contributed extensively, and an out-of-Africa model, in which they were completely replaced. A marriage of paleontology with genetics has given rise to the field of ancient DNA analysis, and has provided data supporting limited genetic contributions from both **Neanderthals** (*H. neanderthalensis*) and **Denisovans** to some but not all modern humans, giving us a new way of looking at old questions. We will thus set the stage for Section 4, in which we will see how modern humans were able to colonize the entire planet.

There can be ambiguity about the meaning of the term "human": we know that we are human, and that chimpanzees are not, but where in the continuum of evolutionary ancestors that link us should we draw the line between human and nonhuman? It makes no sense to ask which human baby had a nonhuman mother. Any decision is somewhat arbitrary; here, we will draw the line between the genus *Homo* and other genera, but will sometimes refer to "archaic" and "modern" humans when we need to refer to early and late members of our genus. We will also use the phrase the "human lineage," meaning the lineage that *led to* humans, and includes all species on the human line since the chimpanzee–human split. There is also ambiguity about the meaning of the term

"ape": some would contrast apes with humans as mutually exclusive groups, but here we consider humans as one of the great apes.

Humans differ from our last common ancestor with other apes in several respects:

1. Morphology: the structure of our bodies, including our brains
2. Behavior: from the way we walk to our social organization, complex tool use, and language
3. Genetics: many neutral and some selected changes. Individual changes could have occurred independently, or in packages linked by selection or drift

Two kinds of evidence have been particularly important for understanding the likely times and places of the changes. Fossils and archaeology provide information about the environment, morphology, and, to some extent, the behavior of our ancestors and the related species present at different times. Genetics reveals the history of the lineages that have survived in living humans, and is also starting to include information from more ancient individuals who lived within the last ~100 KY.

The anthropologist Vincent Sarich once contrasted the two by remarking "I know my molecules had ancestors, the paleontologist can only hope that his fossils had descendants."

9.1 EVIDENCE FROM FOSSILS AND MORPHOLOGY

Both extinct and living species more closely related to humans than to chimpanzees are known as **hominins**; the term **hominid** has been used as an alternative to hominin in the past, but now hominid generally includes all fossils of humans, other great apes, and their immediate ancestors (**Section 7.1**). Candidate fossils can be examined for characteristics that differ between humans and other apes (**Box 9.1**) and their likely position in the phylogeny determined. This aim is hindered by the rarity of fossil hominins: our ancestors appear to have existed at low population densities and were seldom fossilized; furthermore, many of the early fossils are likely to have been predated by carnivores, but this became rarer as defenses against carnivores improved. In addition, the fossils that are found are very incomplete: teeth are the most frequent finds, then the relatively tough bones of the head, the **cranium** (skull excluding lower jaw) and

Box 9.1: Human characteristics that can be preserved in rock

- Brain cavity: absolutely and relatively larger in humans compared with other apes; attachment to spinal cord is more centrally located in the base of the skull.
- Teeth: enamel is thicker; canines and incisors are smaller in humans than other apes.
- Chest: more cylindrical in humans; that of other apes widens toward the base to accommodate larger gut.
- Legs longer, feet and pelvis adapted for upright walking in humans.
- Bipedalism can also be recognized from tracks of preserved footprints.
- Hands adapted for grasping with longer thumb in humans.
- Slow development and prolonged childhood in humans, identified from the "age at death" of fossils, measured, for example, from growth lines on teeth.
- Tool use more extensive in humans; complex tools and fire specific to humans.

mandible (lower jaw), but other bones (often collectively called **post-cranial**) are less often preserved, and traces of soft body tissues only in very exceptional circumstances, such as **endocasts** of the brain. Moreover, there are conceptual problems: there would undoubtedly have been variation within each species due to differences between individuals within the same population including **sexual dimorphism**, and also geographical differences, as well as changes over time. A decision has to be made about which fossils should be grouped under the same name (genus and species), and this inevitably has an arbitrary element. The person who discovers and names a new hominin species gains considerable credit and even celebrity, especially if it can be claimed as a direct human ancestor. The cynic might consider that there is thus a danger of excessive splitting and overemphasis of human similarities at the expense of features shared with other apes. The overabundance of names can be very confusing and the reader should be aware that the field is subject to continuous revision; opinions about the number of genera and species, and their relationships, differ considerably between experts, and change over time. Fortunately, images and information about many of the key fossils are readily available, from Websites, for example, http://www.archaeologyinfo.com or http://www.modernhumanorigins.com, or from books.[58]

Some fossils that may represent early hominins from 4–7 MYA are known from Africa

We saw in Chapter 7 that the human and chimpanzee lineages are generally assumed to have diverged about 6–7 MYA. The earliest hominin fossils should therefore originate from this period, but not before (see **Table 9.1** for a summary of physical dating methods). A few candidate fossils of this age are known

TABLE 9.1:
PHYSICAL DATING METHODS FOR FOSSIL SITES

Method	Basis	Useful time span		
		Type[a]	(KY)	Materials used
Isotopic:	radioactive decay	absolute		
Uranium–Thorium (U–Th)			>500	carbonates
Uranium–Lead (U–Pb)			1–500	carbonates/zircon
^{40}K-^{40}Ar, $^{40}Ar/^{39}Ar$			>10	volcanic rocks
^{14}C			<50	organic
Cosmogenic nuclide burial	*in situ* formation of ^{26}Al, ^{10}Be and subsequent decay		<5000	quartz
Trapped electron dating:	electrons caught in defects in crystal lattices	absolute		
Thermoluminescence (TL)			1–200	quartz, feldspar
Optically Stimulated Luminescence (OSL)			1–500	quartz, feldspar
Electron Spin Resonance (ESR)			1–3000	carbonates, silicates
Amino acid racemization	chemical instability of L amino acids	relative	40–200	organic materials
Paleomagnetism	direction of magnetic field (N or S)	correlative	>780	iron-rich rocks
Paleontology	comparison of fossils with other sites	relative	all	fossils

[a] Absolute, provides a date in years; relative, allows comparisons with other sites or materials; correlative, provides findings which may allow refinement of an approximate date determined by other methods.

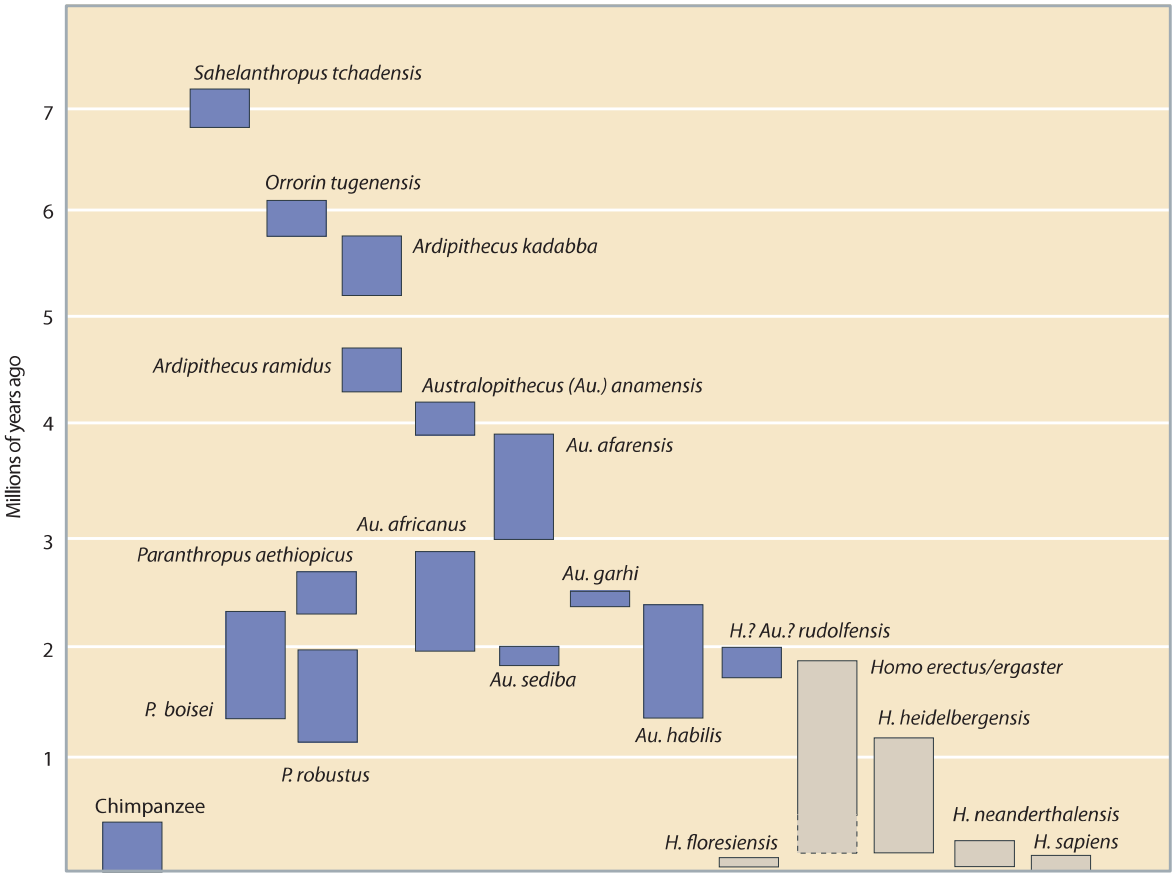

Figure 9.1: Fossil hominins. The time span of each species indicates either the uncertainty in dating or the times of the earliest and latest fossils, whichever is larger. Dotted lines indicate particular uncertainty about the later dates for *Homo erectus*. *Blue*: found only in Africa. *Gray*: found in Africa and elsewhere, or only outside Africa. Many aspects of the classification of these fossils are still debated and are likely to be revised.

(**Figures 9.1** and **9.2**). The oldest of these consists of a nearly complete cranium (Toumai, meaning "hope of life"), jaw fragment, and teeth from Chad, designated *Sahelanthropus tchadensis*,[12] dated to 6.8–7.2 MYA. This date would place it at the upper limit of the divergence time between humans and chimpanzees as estimated from genetic data (**Section 7.3**), and thus the date and relationship of *S. tchadensis* to the hominin lineage are key issues about which there is debate. Despite having a chimpanzee-sized brain, *S. tchadensis* has a number of features that link it to the hominin lineage, including a relatively flat face, attachment of the spinal cord at the bottom of the skull rather than at the back, and intermediate tooth enamel thickness. Unfortunately, an understanding of its mode of locomotion awaits the discovery of post-cranial remains: the position of attachment of the spinal cord hints at an upright stance, but is also consistent with other postures, and interpretation is hindered by the distorted and fragmentary nature of this part of the skull. The environment appears to have had a mosaic structure including forest, **savanna**, desert, and lakes.

Two genera dating to between 5 MYA and 6 MYA have been described: *Orrorin* and *Ardipithecus*. *Orrorin tugenensis*, also called Millennium Man, is represented by 13 fossils dating to about 5.8–6.1 MYA from the Tugen Hills in Kenya, East Africa.[56] These fossils include three fragmentary thigh bones which indicate upright walking, and small thick-enameled molars judged to link *Orrorin* to the human lineage. *Ardipithecus kadabba* is represented by 11 fossils, including a nearly complete foot, from the Middle Awash in Ethiopia, dated to between 5.2 and 5.8 MYA.[20] Despite the presence of thin enamel characteristic of chimpanzees, some researchers consider *Ardipithecus* to lie on the human lineage and also note the similarity of *Ardipithecus kadabba* teeth with those of *Orrorin* and *Sahelanthropus*, suggesting that "it is possible that all of these remains represent specific or subspecific variation within a single genus."[20] However, none of the key features of this group of fossils from 5–7 MYA—tooth morphology,

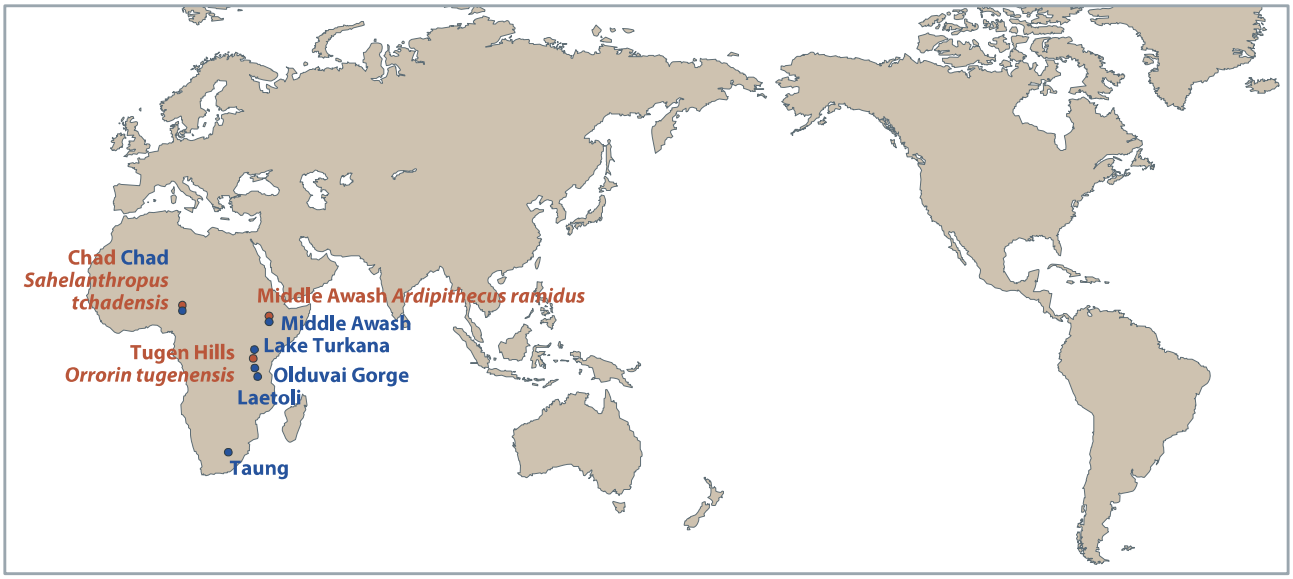

Figure 9.2: Sites of the earliest hominin and gracile australopithecine fossils. Sites of the earliest hominin fossils (7.0–4.5 MYA) in Chad and East Africa (*red*) and gracile australopithecine fossils (4.0–2.0 MYA), spread over more of Africa but absent from the rest of the world (*blue*). In the overlapping locations, the points representing the later sites have been offset to retain visibility.

attachment position of the spinal cord, or foot structure—provides unequivocal evidence for hominin status and after a decade of debate there is still no consensus about their relationships to one another or to chimpanzees and humans.[76]

The period between 5 MYA and 4 MYA is represented by an extensive but crumbling collection of fossils from 36 or more *Ardipithecus ramidus* individuals (4.7–4.3 MYA) from the Middle Awash in Ethiopia, including much of the skull, pelvis, lower arms, and feet from one female, Ardi[70] dated to ~4.4 MYA (see **Opinion Box 7**). Ardi would have weighed around 50 kg when alive and stood ~1.2 m tall: both the pelvis and foot structures provide strong evidence for bipedalism, although with a gait distinguishable from later hominins, and she may have spent substantial time in the surrounding trees. The earliest *Australopithecus* fossils are only slightly younger, and are considered in the next section.

Where the appropriate hominin features can be identified, all of these early fossils appear to represent chimpanzee-sized but apparently upright-walking species that lived in or near a wooded environment. The evidence for **bipedalism** at such an early date is an important finding and raises the question of how the human–chimpanzee common ancestor moved around. The most radical possibility is that bipedalism is the **primitive** trait, present in the common ancestor of all three species, and **knuckle walking** in chimpanzees and gorillas is **derived**, although this possibility requires the nonparsimonious assumption that knuckle walking arose independently in the two great ape lineages (**Section 7.1**). The difficulty of distinguishing between human and chimpanzee lineages at times close to their split should not be a surprise (and see Opinion Box 7), but is made worse by the lack of chimpanzee fossils dating to before 0.5 MYA.

Fossils of australopithecines and their contemporaries are known from Africa

Most hominin fossils dating after about 4.2 MYA and before the appearance of *Homo* are ascribed to the genus *Australopithecus* (*Au.*). The presence of a second genus, *Kenyanthropus,* has also been proposed for a series of 3.5–3.2 MYA fossils,[38] but many researchers consider these specimens to belong to *Au. afarensis*. The earliest known *Australopithecus* species is *Au. anamensis* from Lake Turkana and other sites from Kenya to Ethiopia dating between 3.9 and 4.2 MYA[37, 72] and it is generally considered to be the oldest indisputable hominin. This species is distinguished from the better-known and slightly later *Au. afarensis*[28] mainly by the larger size of the males (weight estimated as ~55 kg

OPINION BOX 7: Shared morphology does not always mean shared recent evolutionary history

The ongoing and sometimes contentious debates about whether this or that fossil taxon is the first or earliest hominin,[70] or the first or earliest member of the genus *Homo*,[8] have much in common. In both cases it boils down to the combination of a matter of principle plus arguments about the presence or absence of a particular morphology in a particular fossil. This opinion box focuses on the matter of principle.

Those who argue in favor of a fossil taxon being the "first" or "earliest" member of a clade do so on the basis that the presence of *any* of the distinctive morphological features seen in later members of that clade "clinches the deal." They assume there is a direct and simple relationship between morphological similarity and genetic relatedness; *all* shared morphology means shared recent ancestry, period.

So why are other researchers[76] skeptical about this principle, especially when it is applied to the hominin fossil record? The main reason is that when it has been "tested" in other mammal groups it has been found wanting. So, for example, when the relationships among the members of a contemporary group that are supported by morphology are compared with the relationships that are supported by molecular evidence, there are discrepancies. And when paleontologists have examined the fossil records of other mammal clades, they also see compelling evidence that similar morphology must have evolved more than once. This has been the case for investigations of bovids, equids, elephantids, carnivores, and Old World monkeys. There is no reason to assume that extinct higher primate lineages that lived at the same time in the same territory were immune from the tendency to adapt in similar morphological and phylogenetically confounding ways to similar ecological challenges. Long ago, the zoologist Ray Lankester suggested the term **homoplasy** should be used for morphology that is seen in sister taxa, but not in their most recent common ancestor. Because homoplasy can be mistaken for shared derived similarity, it complicates attempts to reconstruct relationships. Homoplasies give the impression that two taxa are more closely related than they really are.

One could cope with the confounding effects of homoplasy if the "noise" that it generates was trivial compared with the strength of the phylogenetic "signal." But in some attempts to infer relationships among extant higher primates using skeletal and dental (that is, hard-tissue) data in the form of either traditional non-metrical characters or characters generated from metrical data, the ratio between noise and signal was in the order of 1:2. The results of these analyses were not only frustratingly inconclusive, but when they were compared with the pattern of relationships generated using molecular data, some were found to be misleading. Other researchers suggested this dismal performance was due to the exclusion of character state data from fossil taxa, but this is arguable because soft-tissue characters (for which there are no fossil data) *are* capable of recovering a pattern of relationships among extant higher primates that is consistent with the molecular evidence.[15] It is not just the absence of fossils, it must be something about hard-tissue evidence. Thankfully not all hard-tissue evidence is problematic; it *can* produce results congruent with the relationships generated from molecular data as long as the anatomical regions targeted have a high enough signal to noise ratio and as long as the information about morphology is detailed enough. It is not good news for paleoanthropologists that the type of data the fossil record provides (that is, mostly craniodental hard-tissue morphology) seems to be particularly prone to homoplasy when used at this relatively fine taxonomic level.

The important point is that shared similarities can only take one so far in determining phylogenetic relationships, because homoplasy, as well as uncertainties in determining the polarity of character transformation, have the potential to generate substantial noise that serves to confound attempts to generate reliable hypotheses about relationships. These considerations have clear implications for generating hypotheses about the phylogenetic position of *Ardipithecus* (**Figure 1**) and *Australopithecus sediba*. Even if these taxa share *some* derived features with either later Pliocene hominins or with later *Homo*, it would be rash to simply presume those features are immune from homoplasy, especially when other aspects of their respective phenotypes suggest more distant relationships with, respectively, the hominin clade and later *Homo*.

Bernard Wood, The George Washington University, Department of Anthropology, Washington, DC, USA

Figure 1: *Ardipithecus ramidus* partial skeleton.
Composite image including bones that may come from more than one individual from the same site. [From White TD et al. (2009) *Science* 326, 75. Reproduced with permission from AAAS.]

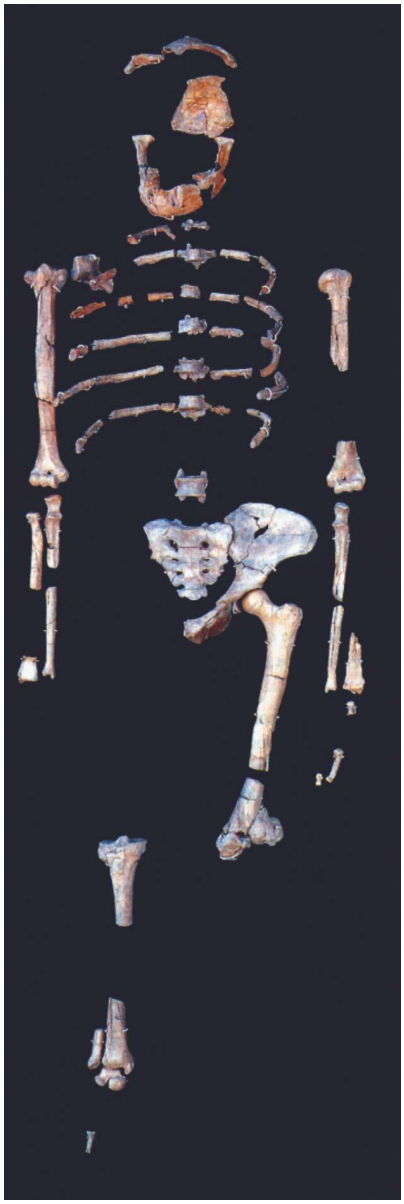

Figure 9.3: Skeleton of *Australopithecus afarensis* (A. L. 288–1 or Lucy). *Au. afarensis* lived in East Africa between 3 and 4 MYA and Lucy, dating to ~3.2 MYA and named after the Beatles' song "Lucy in the sky with diamonds," is its best-known representative. Lucy was a mature adult female when she died, but was only just over 1 m in height. (Photograph by Denis Finnin and Jackie Beckett, © American Museum of Natural History.)

and 45 kg, respectively); the females of the two species were similar (~30 kg). *Au. afarensis* (~3.0–3.9 MYA) is present at a number of sites in East Africa from Ethiopia to Tanzania, and includes the famous partial skeleton Lucy (3.2 MYA; **Figure 9.3**) and probably the Laetoli footprints (3.5 MYA; see Figure 8.2),[36] dramatically illustrating the existence of bipedal locomotion at this time.

The fossil material available from *Au. afarensis* is extensive enough to allow many of its characteristics to be deduced. The species is estimated to have been 1–1.5 m tall (and, as mentioned, bipedal); weight was between 25 and 50 kg, with considerable dimorphism between the sexes. Brain size was 400–500 cc [cubic centimeters, a non-SI unit (1 cc = 1 cm^3) still used in the field and adopted here]: similar, in proportion to body mass, to that of the chimpanzee. The habitat is thought to have been more open than that inhabited by the earlier hominins, perhaps with grassland as well as trees. *Au. africanus*, considered below, was probably a similar species, although with less sexual dimorphism and perhaps more human-like in some ways.

Australopithecus africanus, the first member of the genus to be discovered and named (the Taung Child),[14] is also known from a number of sites, all from the south of the continent, and most of these fossils date to between 2.0 and 2.9 MYA.[24] Later *Australopithecus* is represented in East Africa by fossils from the Middle Awash in Ethiopia designated *Au. garhi* (~2.5 MYA),[3] characterized by relatively large teeth; and in South Africa by *Au. sediba*, well dated at 1.977 MYA.[50] Thus **gracile** (lightly built) australopithecines were present in many areas of Africa from around 4.2 MYA to ~2.0 MYA (Figure 9.2). The relationships between the species mentioned here are unclear: the simplest scheme would consider *Au. africanus* and *Au. garhi* as geographical variants; *Au. afarensis* would be a descendant of *Au. anamensis*, and *Au. africanus/garhi* of *Au. afarensis*; *Au. sediba* is interpreted as a descendant of *Au. africanus*, although the Mrs Ples *Au. africanus* skull from Sterkfontein is virtually contemporary with *Au. sediba* at around 2 MYA. An as yet unclassified third species of *Australopithecus* has been suggested from South Africa in the form of the 2.6–2.2 MYA Little Foot skeleton from Sterkfontein, which has similarities to *Paranthropus* (see below).

Robust (heavily built) hominins with small brains and large jaws and chewing teeth were originally included in the genus *Australopithecus*, but are now commonly placed in a separate genus, *Paranthropus*. *P. aethiopicus* is represented by only a small number of specimens between 2.7 and 2.4 MYA, but these include the Black Skull, a fairly complete ~2.5-MY-old skull from Lake Turkana. *Paranthropus boisei* fossils, including the skull Zinj from Olduvai Gorge, Tanzania,[34] are found mostly in Ethiopia, Tanzania, and Kenya, and span the range 1.4–2.3 MYA. *Paranthropus robustus* remains are known from several sites in South Africa (Swartkrans, Gondolin, Drimolen, Coopers D, Sterkfontein, and Kromdraai B), and have been dated to between ~2.0 and ~1.2 MYA.[25] The robust morphology of these species is thought by some to represent an adaptation to a diet that required heavy chewing, such as low-quality fibrous vegetable food, for example, roots and nuts.

The question as to which of these hominin species is our direct ancestor has attracted considerable attention. It is widely agreed that the *Paranthropus* species form a separate lineage with no surviving descendants. *Au. anamensis* and *Au. afarensis* are good candidates for human ancestors before 3 MYA, but there seems to be no consensus about which fossils represent our ancestors between 3 MYA and the emergence of *Homo* (**Figure 9.4**). Recently, *Au. sediba* has been suggested as such an ancestor due to its mixture of australopithecine and *Homo*-like traits.[50] Its young age at 1.98 MYA might preclude this, but the

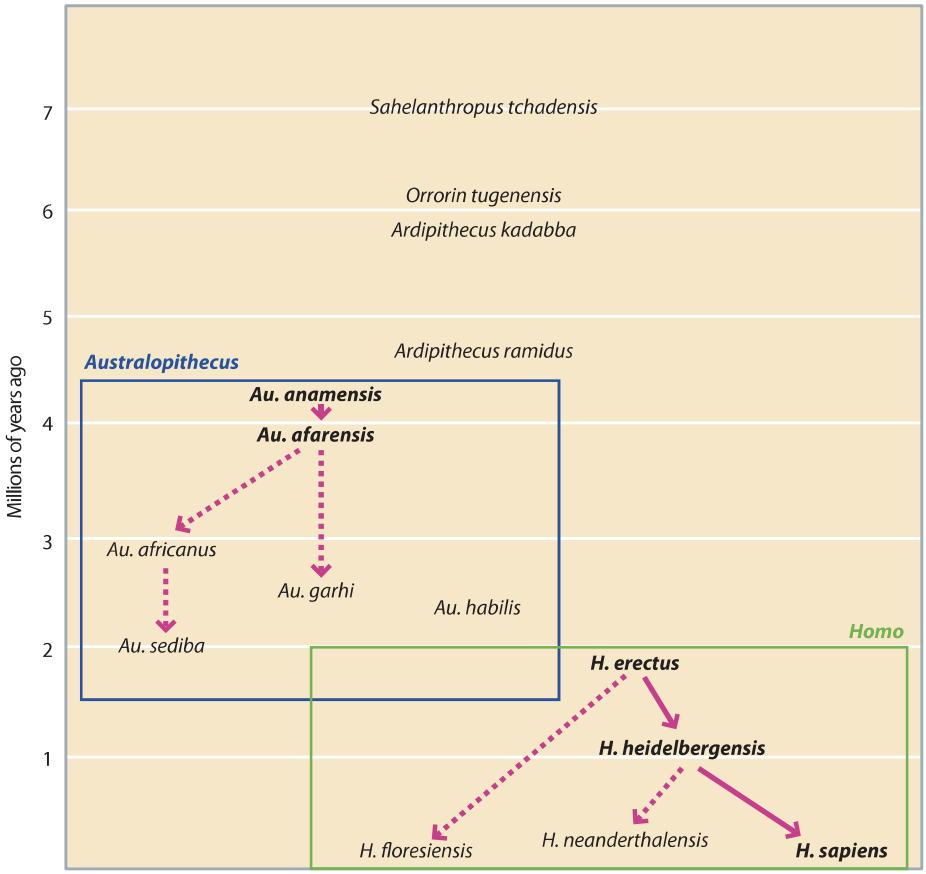

Figure 9.4: Relationships of fossil hominin species, indicating plausible human ancestors. Species relationships are shown by *red* arrows, solid when on the human ancestral line; species in bold are likely human ancestors. Note the uncertainty about the relationships and human line affinities of the early hominins, and the uncertainty about which later australopithecine is ancestral to *Homo*. Compare with Figure 9.1.

type specimen from the site of Malapa may simply represent a late-occurring individual from a species that had existed for some time. Much debate has occurred over whether the large series of fossils (over 500 from Sterkfontein alone) attributed to *Au. africanus* represent two species or whether they simply represent a single highly variable species. This sample of fossils may include earlier specimens of *Au. sediba* that have yet to be identified, or the variability may be temporal with the Sterkfontein Member 4 deposit estimated to have formed over half a million years. If *Au. sediba* were confirmed as the ancestor of *Homo*, this would suggest a southern African, rather than eastern African, origin for our genus.

The genus *Homo* arose in Africa

The reader should not be surprised to learn that there are disagreements about which species should be included within our own genus, *Homo*, and thus about its origin. For many years, the earliest member of the genus was considered to be *H. habilis*, "handy man," named on the basis of a partial skull and jaw, OH 7, from Olduvai Gorge, Tanzania;[35] some *H. habilis* specimens may date back to about 2.3 MYA. However, *H. habilis* has been described as "a mishmash of traits and specimens, whose composition depends upon what researcher one asks" (Kreger, http://archaeologyinfo.com/homo-habilis/); in addition, *H. habilis* does not show the body size and shape, or small teeth, characteristic of humans, while later species do. These features appear shortly after 2 MYA in fossils described as *Homo ergaster* or *erectus*, and it therefore seems reasonable to draw the distinction between *Australopithecus* and *Homo* here;[74, 75] thus *habilis* would be assigned to *Australopithecus* and *erectus/ergaster* would be the first *Homo*. We will therefore refer to "*Au. habilis*" in the following sections.

A small group of fossils from Koobi Fora in northern Kenya, dating to 1.78–2.03 MYA, have been identified with flatter faces and shorter and more rectangular jaws than *Au. habilis*,[39] and assigned to the species *H. rudolfensis*. Debate about

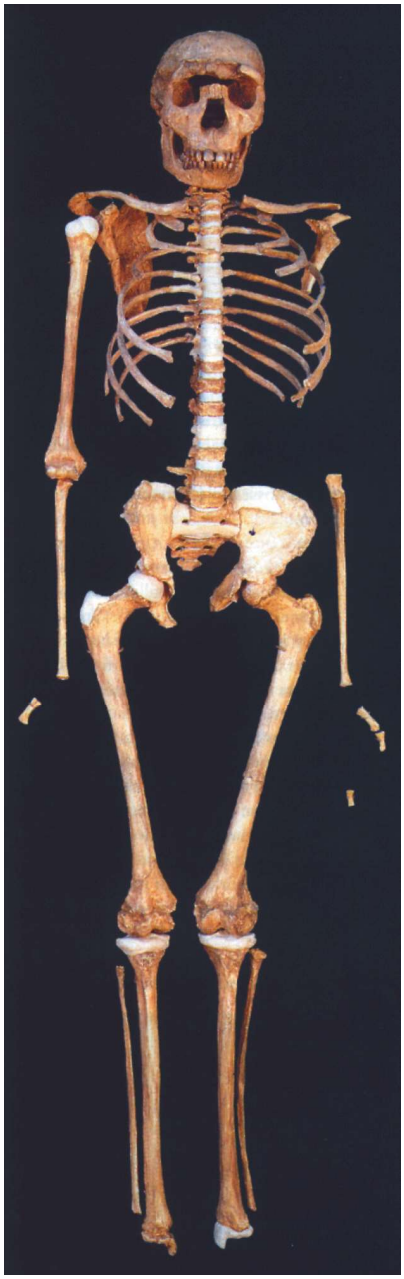

Figure 9.5: Skeleton of *Homo erectus* (WT 15000 or the Nariokotome Boy).
H. erectus (African specimens are sometimes called *H. ergaster*) is known from Africa at ~1.9 MYA and from Asia soon afterward (~1.8 MYA). This specimen, from Lake Turkana, Kenya, dates to about 1.6 MYA. The Nariokotome Boy was an adolescent male when he died, with the body size and shape of modern humans but a smaller brain. (Photograph by Denis Finnin and Jackie Beckett, © American Museum of Natural History.)

their affinities to contemporary hominins and taxonomic status, at the level of both genus and species, continues. In view of the lack of information about their body size and shape, they are assigned below to *Australopithecus*.

There is also debate about the distinctiveness of the two species *H. ergaster* and *H. erectus*: it is difficult to find morphological characteristics that separate them reliably. While one view considers *H. ergaster* to cover African individuals and reserves *H. erectus* for those found outside Africa, an alternative analysis would include all these specimens as a single widespread and variable species, *H. erectus*. The latter view is somewhat strengthened by the finding of a ~1.0-MY-old specimen resembling Asian *H. erectus* in Africa,[2] although this individual could alternatively have migrated back to Africa. Here, the name *H. erectus* will be used for this whole group of fossils. The first examples date to between 1.8 and 1.9 MYA[50] and, like all earlier hominins, are found in Africa, demonstrating an African origin for our genus. *H. erectus* fossils include the outstanding Nariokotome Boy (~1.6 MYA; **Figure 9.5**),[67] the most complete early hominin thus far found, which provides important insights into this species. He is thought to have been in early adolescence when he died, male, tall and thin at about 1.5 m high and weighing 47 kg. As an adult he would probably have reached 1.8 m and 68 kg, common figures for modern humans. His limb proportions and tooth size were also similar to those of modern humans, but his brain size (880 cc, corresponding to 909 cc at maturity) was significantly smaller than the modern human average (around 1450–1500 cc), although just within the modern human range (830–2300 cc).

H. erectus is the earliest hominin to be found outside Africa (**Figure 9.6**), and includes influential fossils discovered in the late nineteenth and early twentieth centuries, such as Java Man (Indonesia, the site of the type specimen Trinil 2 discovered in 1891) and Peking Man (China). The earliest *H. erectus* dates outside Africa, from Dmanisi (Georgia),[43, 59] are ~1.8 MYA,[17] a little younger than the earliest African *H. erectus* from Swartkrans Member 1 and Koobi Fora in Kenya at ~1.9 MYA. It has been suggested that the large body size providing tolerance to heat stress and dehydration, coupled with improved stone toolkits, may have allowed the species to live in a wide range of environments and thus expand out of Africa rapidly.[74] In Java, *H. erectus* may have survived until 135 KYA;[26] if so, they would have been contemporaries of fossils on the East Asian mainland that have in most cases been referred to as archaic *H. sapiens*. The analysis of aDNA from a >50-KY-old finger bone from Denisova Cave in Siberia has rekindled interest in Asian hominins, and we discuss this enigmatic taxon below (**Section 9.5**).

The report of fossils representing a tiny 1-m-tall hominin species, *H. floresiensis*, from the island of Flores in Indonesia in 2004[11] surprised paleontologists so much that the first reaction of some was to think that the story must be a hoax. But these "hobbits," named in tribute to the imaginary characters of J.R.R. Tolkien, were supported by the remains of multiple individuals and archaeological deposits spanning the period 17–74 KYA, including the fairly complete 380 cc cranium and skeleton of the type specimen, LB1. Despite the suggestion that *H. floresiensis* might represent modern humans suffering from **microcephaly** (a neurodevelopmental disorder in which the head is abnormally small), most paleontologists now accept them as distinct hominins, perhaps *H. erectus* descendants, surviving on an isolated island poor in resources since ~1 MYA, resulting in selection for small size.

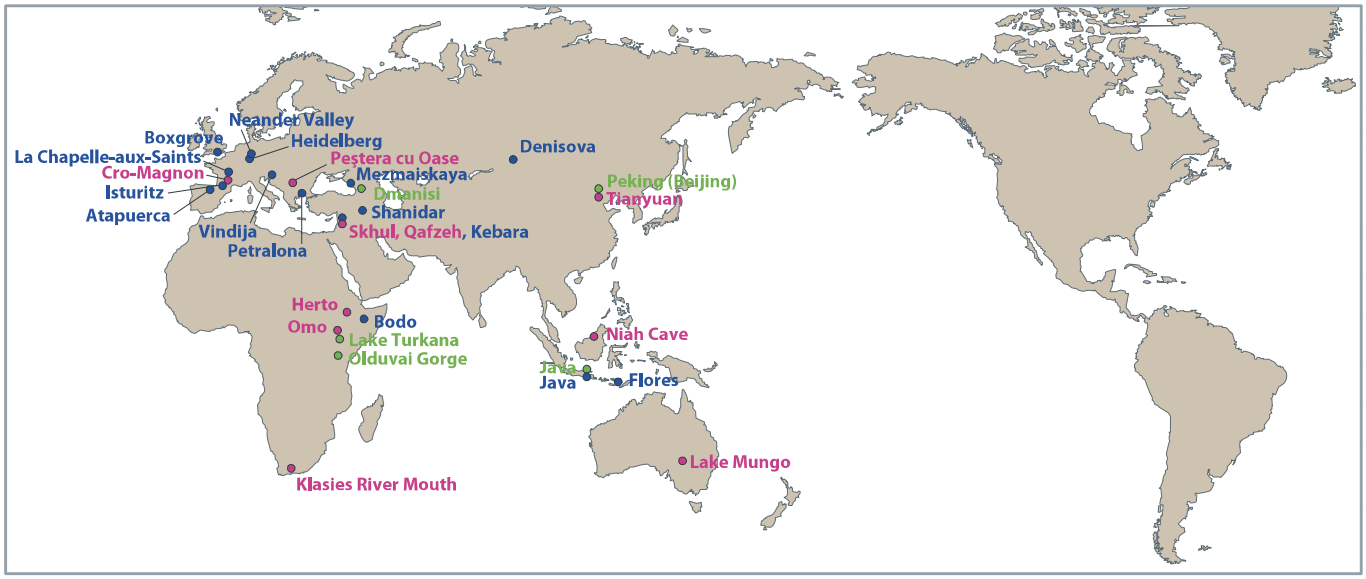

Figure 9.6: Sites of *Homo* fossils.
Early sites (1.9–1.6 MYA) are shown in *green*, later sites (800–12 KYA) for species other than *H. sapiens* in *blue*, and *H. sapiens* in *red*. These sites are spread throughout much of the world, illustrating the extensive spread of *Homo* compared with earlier hominins (Figure 9.2).

Later *Homo* from Africa and Europe (**Figure 9.7**) are less robust and have larger brains (~1200 cc instead of ~900 cc) than early *H. erectus* and are often designated *H. heidelbergensis*, the type specimen of which is the ~609 KYA Heidelberg Jaw from Germany[66] (**Figure 9.8**). Related specimens include the massive Bodo cranium (Ethiopia, ~600 KYA), the tibia (lower leg bone) from Boxgrove (England, ~500 KYA), and the Petralona 1 cranium (Greece, age uncertain but with estimates between 200 and 700 KYA). Many would also place the 1.2–0.80 MYA specimens from Gran Dolina and Sima del Elefante, Spain, designated *Homo antecessor* by their discoverers,[9] within *H. heidelbergensis*. According to this view, *H. heidelbergensis* would have been a widespread and somewhat variable species, perhaps originating from *erectus* in Africa some time prior to 1 MYA and giving rise to more recent *Homo* species, including *H. sapiens* and *H. neanderthalensis*. Other researchers, however, prefer to call the post ~1 MYA African specimens *H. rhodesiensis* after the ~300–125-KY-old Kabwe (or Broken Hill 1) skull from Zambia. In South Africa, other potential specimens assigned to *H. rhodesiensis* include the Saldanha Man skullcap from Elandsfontein dated between 1.1 and 0.6 MYA and the Cave of Hearths material which perhaps dates to between 800 and 400 KYA.[23] In this more complex scenario, not followed here, *H. heidelbergensis* is a European species giving rise to *H. neanderthalensis*, and *H. rhodesiensis* an African species giving rise to *H. sapiens*.

Neanderthals (also spelled "Neandertals"; **Figure 9.9**) form a morphologically distinct group of fossils from Europe and Western Asia between ~250 KYA and ~28 KYA, robust and with large brains (~1400 cc, larger than those of many modern humans) and well-developed brow ridges. Well-known examples include the type specimen Feldhofer 1 ~40 KYA from the Neander valley in Germany, The Old Man of La Chapelle-aux-Saints ~50 KYA from France (a 40–50-year-old individual showing evidence of arthritis, which was not recognized as pathological when the specimen was first described in the early twentieth century, leading to Neanderthals being wrongly stereotyped as "brutish" and "bent-kneed"), Kebara 2 (Israel, ~60 KYA), and Shanidar 4 from Iraq (~60 KYA), sometimes interpreted

Figure 9.7: *Homo heidelbergensis* (Broken Hill 1 or the Kabwe Cranium).
This example was found in a lead and zinc mine in Zambia and its context is uncertain, but a date of 125–300 KYA has been suggested. [(Courtesy of Gerbil under Creative Commons Attribution-Share Alike 3.0 Unported license.]

Figure 9.8: *Homo heidelbergensis* mandible (Mauer 1 or the Heidelberg Jaw).
Opinions differ about which African and European fossils dating from 1100–200 KYA should be ascribed to *H. heidelbergensis*, but this mandible is the type specimen and so must belong to *H. heidelbergensis*. It was found near Heidelberg, Germany, and dates to ~609 KYA. (Reproduced with permission of the Science Photo Library.)

as representing a deliberate burial. In the era of successful aDNA sequencing, bone fragments from Vindija, Croatia (see Figure 9.20), have become well known for the molecular, rather than archaeological, information they have provided. Neanderthals are thought to be descendants of *H. heidelbergensis* and are usually assigned to a distinct species, *Homo neanderthalensis*, but their relationship to modern humans has aroused intense debate, now informed by a **low-coverage** genome sequence, and is considered further below (**Section 9.5**).

Figure 9.9: *Homo neanderthalensis* skull (The Old Man of La Chapelle-aux-Saints).
H. neanderthalensis lived in Europe and Western Asia from ~250 to 28 KYA. This specimen from France dates to ~50 KYA and was derived from a 40–50-year-old man. Note the large brow ridges and small chin. Several pathological features, including arthritis and resorption of the tooth sockets, are also present and contributed to the misinterpretation of Neanderthals as shuffling, brutish cavemen. (Reproduced with permission of the Science Photo Library.)

Needless to say, these classifications are not universally accepted. Indeed, all *Homo* species arising after *H. erectus* and before modern *H. sapiens* have sometimes been referred to collectively as "archaic *H. sapiens*," with specimens definitively assigned to *H. sapiens* (defined strictly) being referred to as **anatomically modern humans** (**AMH**).

The earliest anatomically modern human fossils are found in Africa

The origin of modern humans has probably been the most contentious issue in the field over the last 30 years. We will see that there is an important distinction between **morphology** and **behavior**, and will begin by considering modern human morphology and ask where and when this first appeared. Anatomically modern humans differ from earlier hominins ("archaic humans" or "archaic *H. sapiens*"), but these differences are not easy to define; indeed, it is often pointed out that there is no type specimen for *Homo sapiens*. Paleontologists have focused mainly on cranial features, which can be summarized by two characteristics, derived from a comparison of 100 recent humans, 10 fossils classified as anatomically modern *H. sapiens*, and 9 classified as *H. neanderthalensis* or *H. heidelbergensis*:[42] (1) extent of the globular shape of the skull; (2) degree of retraction of the face (Figure 8.8). This system allows a clear distinction between AMH and archaic humans, with zero overlap, but has the disadvantage that relatively complete specimens are needed; for fragmentary specimens it is necessary to use less reliable criteria.

The earliest accepted fully modern human skull comes from Omo-Kibish (Ethiopia, **Figure 9.10**) and dates to ~195 KYA.[45] Slightly later crania of one child and two adults from Herto (also Ethiopia) date to 154–160 KYA and show many of the features of modern human morphology,[71] yet the authors created a new subspecies *Homo sapiens idaltu* to accommodate them, emphasizing the morphological variation at this time. Despite this, few researchers use this subspecies level classification and they are most often defined as AMHs. The most complete cranium is large (1450 cc) and has the globular braincase of modern humans, but retains some more archaic features such as protruding eyebrows. Interestingly, both adults show evidence of postmortem modification, including cut marks, interpreted as resulting from mortuary practices. Other fragmentary specimens are known from Klasies River Mouth in South Africa at 90–120 KYA, and two sites from Israel dated to between 90 and 130 KYA: the cave at Qafzeh

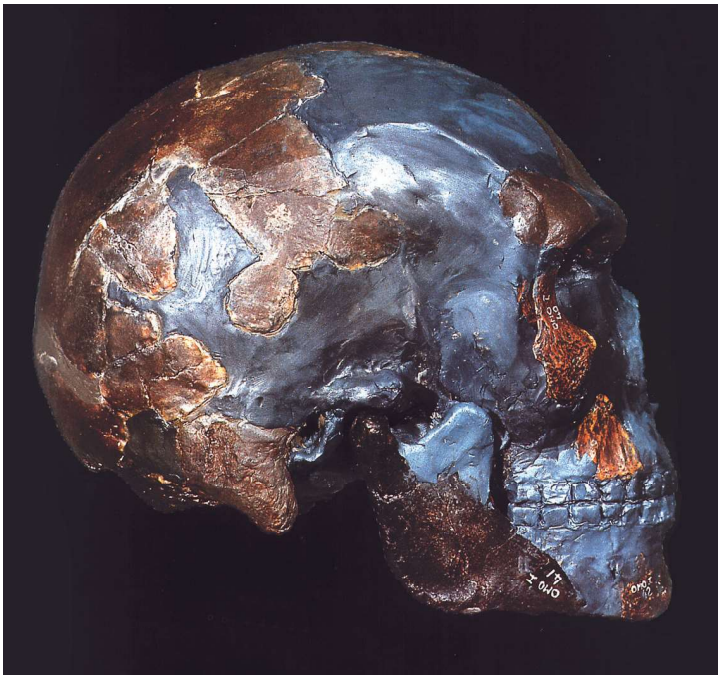

Figure 9.10: The earliest anatomically modern human cranium (Omo I). Modern features include the high forehead and developed chin; note that the gray portions are reconstructed. This specimen from Omo-Kibish in southern Ethiopia provides crucial evidence that modern human anatomy had developed in Africa by ~195 KYA. (Reproduced with permission from Michael Day.)

TABLE 9.2:
THE OLDEST ANATOMICALLY MODERN HUMAN FOSSILS FROM DIFFERENT REGIONS OF THE WORLD

Continent	Location	Remains	Date (KYA)	Reference
Africa	Omo-Kibish, Ethiopia	Omo I skull	195	McDougall I et al. (2005) *Nature* 433, 733.
Africa	Herto, Ethiopia	three crania	154–160	White TD et al. (2003) *Nature* 423, 742.
Africa	Klasies River Mouth, South Africa	multiple fragments	90–120	Royer D et al. (2009) *Am. J. Phys. Anthropol.* 140, 312.
Middle East	Qafzeh and Skhul, Israel	multiple, >30 individuals	90–130	Grün R et al. (2005) *J. Hum. Evol.* 49, 316.
East Asia	Niah Cave, Borneo	cranium and leg bones	34–46	Barker G et al. (2007) *J. Hum. Evol.* 52, 243.
East Asia	Tianyuan Cave near Beijing, China	partial skeleton including mandible	39–42	Shang H et al. (2007) *Proc. Natl Acad. Sci. USA* 104, 6573.
Australia	Lake Mungo	Lake Mungo 3	40 ± 2	Bowler J et al. (2003) *Nature* 421, 837.
Europe	Grotta del Cavallo	two molars	43–45	Benazzi S et al. (2011) *Nature* 479, 525.

with parts of more than 20 skeletons and the rock shelter at Skhul with at least 10 individuals, both including some likely burials. The **Levantine** nonhuman fauna at this time is interpreted as a temporary extension of the African fauna, and thus all of these early human remains, like the animals, can be considered African. Outside Africa (interpreted in this sense), the earliest accepted dates for modern fossils are all <45 KYA,[6] with fossils dating close to 40 KYA known from Europe, East Asia, and Australia (**Table 9.2**).

Evidence for the appearance of modern human behavior will be discussed in **Section 9.2**, and timing of the first modern human presence in different regions of the world will be considered in more detail in Chapters 11 and 13. Here, we note that, despite uncertainties in classification and dating, and the extremely incomplete nature of the fossil record, the earliest dates outside Africa are much more recent than dates inside Africa: it is clear from the fossil evidence that modern human morphology appeared considerably earlier in Africa than elsewhere.

The morphology of current populations suggests an origin in Africa

Morphological variation among present-day populations should also carry information about the origins of modern human anatomy: if this variation is predominantly neutral, the simple expectation is that it should be greatest close to the origin. However, morphology is shaped by selection as well as by neutral forces, so a combination of a large dataset (37 measurements each from 4666 male skulls belonging to 105 worldwide populations) and allowance for the correlations with climate was necessary to detect a signal of the origin. This showed that populations in sub-Saharan Africa were the most variable, and variance fell with distance away from this region, with distance from Africa accounting for 19–25% of the variation,[44] a striking parallel with the genetic pattern (**Section 9.4**). This study was not able to pinpoint a specific area within sub-Saharan Africa as the most likely origin, but interestingly found no evidence for a second origin, thus providing no support from cranial morphology for a **multiregional model** (see **Section 9.3**).

9.2 EVIDENCE FROM ARCHAEOLOGY AND LINGUISTICS

Archaeological evidence may be considered as the preserved signs (other than fossils) of hominin activity, although this definition could be extended to include the activity of nonhuman apes as well. While hominin fossils are very rare, archaeological remains, such as stone tools, are much more common.

Assemblies can be classified and associated with one or more hominin type through rare sites that contain both archaeological remains and fossils, and then allow the presence of these hominins to be inferred elsewhere, albeit with the limitation that there is no one-to-one correspondence between technology and species.

Chimpanzees use a range of tools, including sticks to extract termites and stones to break open nuts,[73] and even manufacture and use sticks for hunting,[51] while orangutans also use tools in a variety of ways, including for seed extraction and autoerotic purposes,[65] so a parsimonious assumption is that our common ancestor used tools as well. Most of these would not be preserved in the archaeological record, but a 4.3 KYA chimpanzee archaeological site containing modified stones carrying starch residues has been recognized in Ivory Coast in West Africa.[46] However, the identification of any tools used by the earliest hominins remains an area for future research, so known archaeology currently begins with the **Oldowan** culture (**Mode 1** technology) starting about 2.6–2.5 MYA (**Figures 9.11** and **9.12a**).

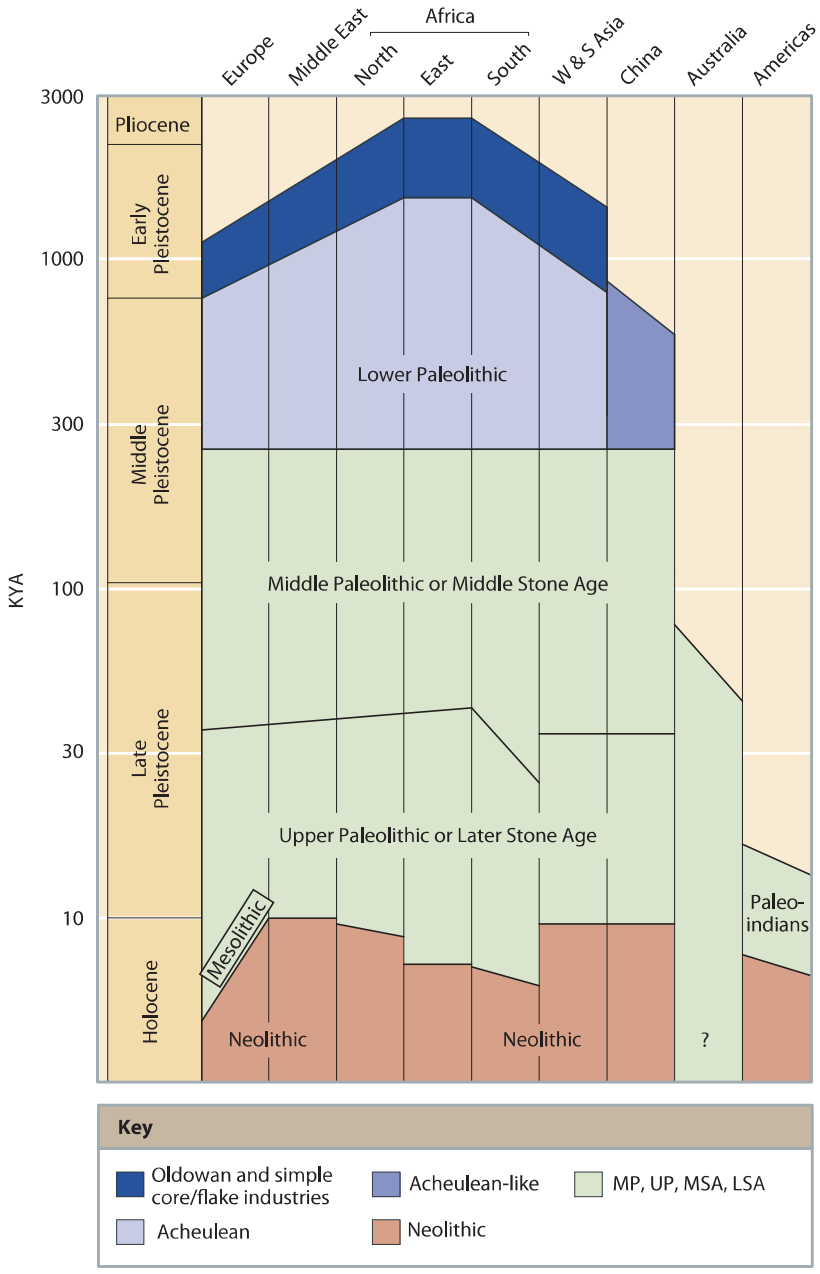

Figure 9.11: Chronology of archaeological stages in different regions of the world.
The first column shows the time-scale (note that it is nonlinear), and the second column gives the geological epoch. Subsequent columns show the archaeological stage in selected world regions. Archaeological remains appear earlier in Africa than elsewhere. Opinions vary about whether or not there was a Neolithic period in Australia. MP, Middle Paleolithic; UP, Upper Paleolithic; MSA, Middle Stone Age; LSA, Later Stone Age.

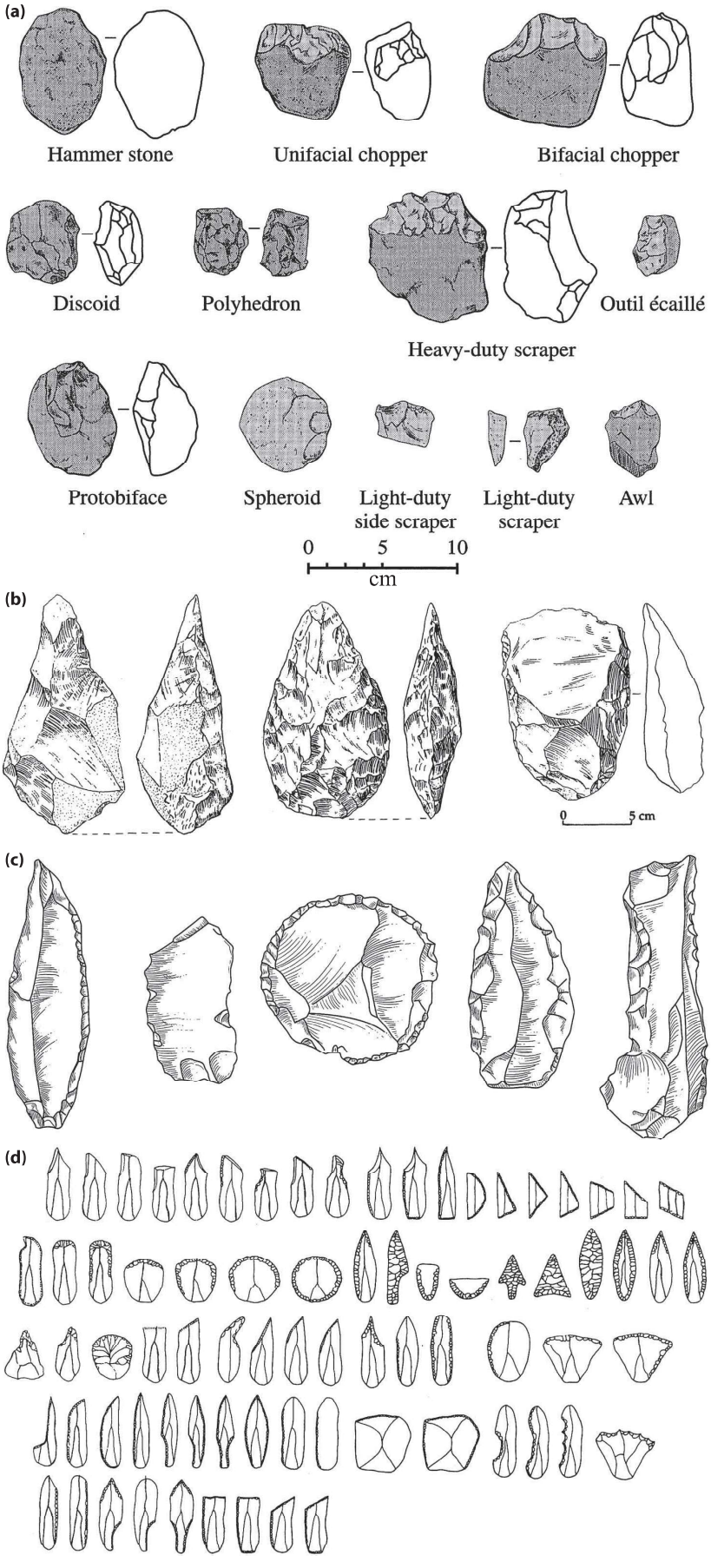

Figure 9.12: Stone tool technologies. (a) The oldest recognized stone tools are Oldowan, manufactured from pebbles and dating back to ~2.6 MYA. (b) After ~1.8 MYA, Acheulean tools are found, including bifaces. They continued to be used until ~150 KYA. (c) Mousterian tools were manufactured by the Levallois technique after ~300 KYA, associated with both Neanderthals and early anatomically modern humans. (d) Upper Paleolithic tools dating after 50 KYA showing the wide range of forms. [From Lewin R (1999) Human Evolution. With permission from John Wiley & Sons Inc.]

Paleolithic archaeology has been studied extensively

Oldowan tools, named after Olduvai Gorge, Tanzania, East Africa, could be recognized as artifacts by characteristics such as:

- **Conchoidal fracture patterns** resulting from striking one stone with another, which differ from the fractures seen in naturally cracked stones
- Transport of stone over several kilometers, so that tools may be made from lavas or quartzite which do not occur naturally at the site
- Concentrations of stone tools, sometimes in association with butchered animal bones

Oldowan tools consist of hammer stones, flakes, and cores, and, as implied above, were probably used to scavenge animal carcasses, including breaking open the bones. Even hyenas cannot crack the thick-walled limb bones of large animals, so tool use would have provided the hominins with a rich and novel food source—bone marrow. It is impossible to be certain of the identity of the toolmakers at Olduvai Gorge, but they are usually assumed to be *Au. habilis*. Other potential toolmakers include *Au. garhi* and *Au. rudolfensis*. In South Africa, the earliest stone tools are found in the same deposits as *Paranthropus robustus* remains, like the much younger Oldowan (<1.8 MYA) from Sterkfontein and Kromdraai.[25] The older Oldowan deposits from Swartkrans Member 1 are associated with both *P. robustus* and *H. erectus* and thus the latter may be responsible for the stone tool manufacture. At Drimolen and Swartkrans bone tools are associated with *P. robustus*.[5] As yet, no stone tools have been associated with the pre-2-MYA *Au. africanus*, but the ~2 MYA remains of *Au. sediba*, while not thus far associated with stone tools, have a hand morphology capable of making and using stone tools.[29]

By 1.76 MYA, strikingly different tools started to be made: symmetrical teardrop-shaped **handaxes**, worked around all or most of either one (unifacial working) or both sides (bifacial working, **bifaces**)[40] (Figure 9.12b). These are called **Acheulean** (also spelled "Acheulian" and referred to as **Mode 2** technology) after the French site St. Acheul, and are often found in association with larger flake tools than the Oldowan, as well as the same Oldowan tools, at least to begin with. The uses of these handaxes are poorly understood, but they have been described as the "Swiss Army knife" of the **Paleolithic** and continued to be used, with little obvious change in overall shape, until around 125 KYA when they are associated with stone tools characteristic of the next technological stage, the **Middle Stone Age** (**MSA**). They were, perhaps, the most successful of all human tools. They are found throughout Africa, in Europe, and in Asia south of the **Movius Line** which runs from the Caucasus mountains to the Bay of Bengal, but are largely absent from Eastern Asia.

The earliest evidence for the Acheulean is found in Africa, where it co-occurs with Oldowan tools near Lake Turkana, Kenya, at 1.76 MYA;[40] the earliest hominin sites outside Africa lack Acheulean tools. Acheulean sites in Israel (Ubeidiya) and India may date to 1.5 MYA,[49] while in East Asia, well-crafted stone tools are known from the Bose basin in Southern China by ~803 KYA, where hominins apparently exploited the rock exposed by a meteorite impact.[77] However, these eastern assemblies are described as Acheulean-like rather than Acheulean. Acheulean technologies are associated with both *H. erectus* and *H. heidelbergensis*, and some have speculated that their construction required advanced mental capacity, including the ability to visualize their shape in advance. Archaeology thus adds significantly to our understanding of this period: the first *H. erectus* to leave Africa apparently did not use Acheulean tools, which spread substantially later, as shown by the Oldowan-like stone tools from the ~1.8 MYA site of Dmanisi in Georgia and 1.2–0.8 MYA sites of Atapuerca in Spain; in Asia the tools reveal an important cultural difference between regions east and west of the Movius Line.

While stone tools dominate the early archaeological record, we would expect that many other materials would have been used, but would seldom have been

preserved. A set of wooden throwing spears from Schöningen in Germany was found in association with butchered horses and is dated to ~400 KYA,[62] providing evidence for use of multiple materials and sophisticated hunting activity at this time.

Between ~800 KYA and 280 KYA a series of innovations occurred with the development of **prepared core technology**, followed by smaller flake-based stone tools that are characteristic of the MSA. The first prepared core technology was developed to create standardized blanks for the construction of handaxes and occurs between ~800 and ~300 KYA in Africa and Israel. By ~550–500 KYA the first blades are seen, and between 500 KYA and 280 KYA the first points occur, along with the potential early occurrence of the **Levallois technique** (**Mode 3**). The Levallois technique is a more complex form of prepared core technology that is generally associated with the MSA, in which the shaping of the tool was accomplished by removing flakes from a core, followed by removal of one final shaped flake which would form the tool itself (Figure 9.12c). The MSA first occurs by at least 280 KYA at Gademotta in Ethiopia and around 250–200 KYA in the Kapthurin Formation of central Kenya. In South Africa there is the suggestion, as at Kapthurin, that all the elements of the MSA may have been in place by 500–400 KYA, either marking the earlier beginnings of the MSA or the occurrence of a transitional industry often called the Sangoan or Fauresmith.[23] In Africa a variety of MSA industries have been defined which range from simplistic small flake-based industries to industries with refined bifacial points (Lupemban and Still Bay) to the early use of microliths (Howieson's Poort; ~65 KYA) that are normally characteristic of the **Later Stone Age** (**LSA**) of Africa or the **Upper Paleolithic** (**UP**) of Europe after 40 KYA.

The Levallois technique was being used in Europe by 200 KYA. Among other Mode 3 **Middle Paleolithic** industries in Europe is the **Mousterian**, characterized by flakes described as side scrapers and points. The human remains associated with Mousterian artifacts are usually Neanderthal, but at Qafzeh and Skhul they are early modern humans, and late Neanderthals may have used non-Mousterian tools, so again there is no simple correspondence between species and technology (**Section 11.2**). Mousterian-like toolkits are found in Asia as far to the east as Lake Baikal, and in southern Asia tools have been labeled as "Mousteroid" because of the high incidence of scrapers. Artifacts classified as MSA are also associated with the earliest modern human remains in Australia.

The Upper Paleolithic in Eurasia and Later Stone Age in Africa are defined by a greater diversity of stone tools and artifacts including microlithic technology and the use of bone and wood (Figure 9.12d). While the earliest potential art forms, such as the ~72 KYA Blombos Cave engraved and shell necklaces, occur within the MSA, the oldest unequivocal art occurs in the form of cave paintings and carved bone and ivory. In the Upper Paleolithic, the predominant (**Mode 4**) tools are described as **blades** instead of flakes; blades are long narrow flakes made from specialized cores and then reworked in a number of ways (Figure 9.12d). In particular, they may be retouched at the end rather than the side. Objects made from other materials, such as wood and bone, become much more abundant in the Upper Paleolithic and unequivocal art is found. The Upper Paleolithic is often associated with modern humans although, as we have seen above, there is no simple correspondence between toolkits and species. Discussion of subsequent developments will be continued in later chapters.

Evidence from linguistics suggests an origin of language in Africa

Languages change rapidly, even within the span of a human lifetime, so it may seem surprising that linguistics can be informative about ancient human origins. The relevant evidence comes not from the study of vocabulary, which turns over quickly, but from the basic units of sound: **phonemes**. A study of 504 diverse languages[4] found that phonemic diversity was highest in Africa and declined with distance from central/southern Africa; after correcting for

population size, distance from Africa accounted for 19% of the variance in phonemic diversity. The parallels with the morphological and genetic patterns are striking (**Sections 9.1** and **9.4**).

9.3 HYPOTHESES TO EXPLAIN THE ORIGIN OF MODERN HUMANS

While many of the fossil discoveries described in the previous parts of this chapter have been made in the last few years, and dates have often been refined or revised, the basic pattern of an early exodus of *H. erectus* from Africa to occupy much of the Old World, followed by a much later appearance and expansion of modern *H. sapiens*, has been clear for decades, and has conditioned the debate that dominated the field during the second half of the twentieth century. This debate can be most easily appreciated by first considering two extreme views (**Figure 9.13**):

- The multiregional model proposed that the transition from *H. erectus* to *H. sapiens* took place in a number of areas of the Old World, with different modern human characteristics arising at different times in different places.
- In contrast, the **out-of-Africa model** proposed that the transition took place in Africa, and that these humans recently (<100 KYA) replaced the hominins already present on other continents.

One way of characterizing the difference is that, according to the multiregional model, our ancestors lived on several continents over the past 1 MY; in contrast, according to the out-of-Africa model, we descend entirely from the ancestors who lived less than a few hundred thousand years ago in Africa; their contemporaries from other continents did not contribute to our ancestry. These models were formulated before the classification of many of the species between the times of *H. erectus* and *H. sapiens* was adopted, and it is not entirely clear how all the additional species would fit into them.

Intermediate models are obviously possible, for example involving a recent origin of most human characteristics in Africa, but also interbreeding with archaic populations inside or outside Africa—a **leaky replacement** model. Fossil, archaeological, and genetic evidence provided little support for an extreme multiregional model, instead generally being interpreted to favor an out-of-Africa model, with or without a low level of admixture. Archaeological evidence, for example, suggested cultural contact between Neanderthals and modern humans in Eurasia before Neanderthals went extinct, and some paleontologists have identified intermediate morphology in some fossils including a ~25 KYA boy's skeleton from the Lapedo Valley in Portugal that was interpreted as a hybrid of Neanderthals and AMHs. We will see that aDNA data have provided new insights into this topic, transforming the debate in ways that had not been anticipated (**Section 9.5**).

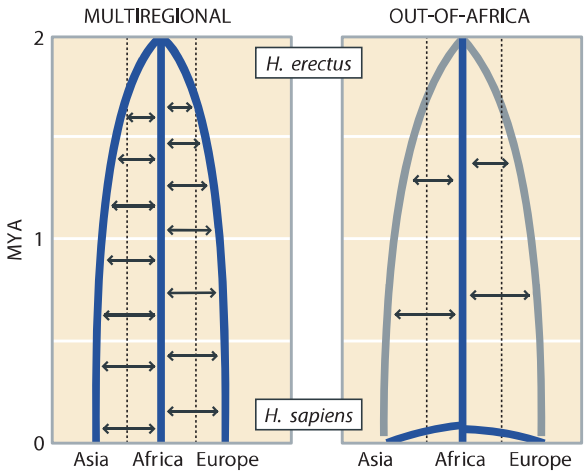

Figure 9.13: Two extreme models for the origins of modern humans. Both models begin with *H. erectus* shortly after 2 MYA and lead to contemporary humans; many intermediate models could also be proposed. Horizontal arrows indicate gene flow between populations on different continents. In the multiregional model, extensive gene flow is required; the out-of-Africa model requires less. *Blue* lines: ancestors of modern humans. *Gray* lines: lineages that are not ancestors of modern humans.

With this background, we will now consider data from present-day samples: patterns of genetic variation in current populations should contain information about modern human origins.

9.4 EVIDENCE FROM THE GENETICS OF PRESENT-DAY POPULATIONS

The scenarios in Figure 9.13 make different predictions about the geographical distribution of genetic variation. According to the multiregional model, there is no strong reason for any one geographical region to show more diversity than another, or be the source of a majority of lineages; in contrast, the out-of-Africa model predicts both greater diversity in Africa, and that Africa would be the root of the majority of genetic lineages. These predictions would remain true for intermediate models close to one or other of these extremes. In order to evaluate such predictions, it is important to use datasets where the results are not significantly influenced by **ascertainment bias**. Suitable data can best be obtained by resequencing, but genotypes consisting of microsatellites (**Section 3.4**), which are variable in all populations, or haplotypes consisting of multiple SNPs, where different sets of common SNPs from a particular genomic region tend to identify the same set of haplotypes, are also suitable. We now have extensive genetic data that can be used to study the geographical distribution of genetic variation.

As explained in Chapter 6, the amount of genetic variation can be assessed in a number of ways, ranging from simple direct measures like the number of variants or nucleotide diversity, to more indirect statistics such as **effective population size** or the extent of **linkage disequilibrium**. We will see that all of these are informative, and identify a consistent pattern.

Genetic diversity is highest in Africa

Genetic diversity can best be evaluated using whole genome sequences, and such data are beginning to become available. The 1000 Genomes Pilot Project resequenced the genomes of population samples originating from Africa, Europe, and East Asia, providing a genomewide and reasonably unbiased view of the variation in each sample. The total numbers of SNPs discovered in the populations, the numbers per individual, and the corresponding numbers of **indels** were all highest in the YRI from Africa, intermediate in the CEU from Europe, and lowest in the CHB+JPT from East Asia[61] (**Table 9.3**).

In this study, the variant ascertainment (including sample size) was similar for the three areas, so it is meaningful to compare these raw numbers. Nucleotide diversity (**Section 6.2**) measured far from genes also shows its highest value in the YRI, an intermediate level in the CEU, and the lowest in the CHB+JPT: 1.3×10^{-4}, 1.0×10^{-4}, and 0.9×10^{-4}, respectively.[22]

Although the genome coverage in this study was close to the maximum possible—the entire **accessible genome**—the number of populations was small. Additional populations were examined in another component of the 1000 Genomes Pilot Project, which sequenced ~700 genes in seven populations.

TABLE 9.3:
GENETIC DIVERSITY, MEASURED AS SNP AND INDEL NUMBERS, FROM AFRICAN, EUROPEAN, AND EAST ASIAN SAMPLES IN THE 1000 GENOMES PILOT PROJECT

Continent	Sample	Total SNPs	SNPs per individual	Total indels	Indels per individual
Africa	YRI	10,938,130	3,335,795	941,567	383,200
Europe	CEU	7,943,827	2,918,623	728,075	354,767
East Asia	CHB+JPT	6,273,441	2,810,573	666,639	347,400

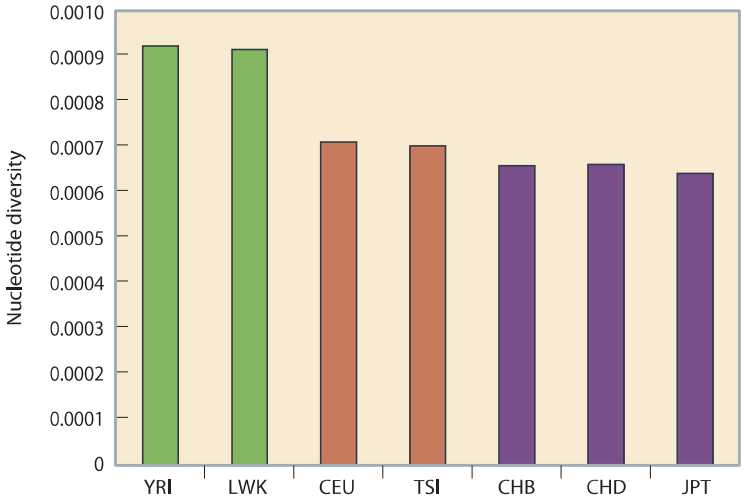

Figure 9.14: Nucleotide diversity of samples from Africa, Europe, and East Asia.
Data are from **fourfold degenerate** (approximately neutral) sites produced by the 1000 Genomes Project exon pilot. The abbreviations refer to population samples that are part of the HapMap Phase III project (see Box 3.6).

Again, there was a clear pattern, with the African samples showing the highest diversity, the European samples intermediate values, and the East Asian samples the lowest (**Figure 9.14**). This shows that such a pattern is a general feature of these geographical regions. Nevertheless, the geographical representation in these large-scale resequencing studies is still very limited (Figure 10.4). To compare levels of genetic variation in larger sets of populations, we need to turn to other datasets.

Microsatellite variation (**Section 3.4**) was typed at 377 autosomal loci in 51 populations from the **HGDP-CEPH panel** (Box 10.2) and the highest heterozygosity values were found in sub-Saharan Africa,[55] a finding confirmed by large-scale SNP genotyping in the same populations.[41] Strikingly, heterozygosity showed a strong linear decrease with distance from East Africa along plausible migration routes ($R^2 = 0.85$, $p < 10^{-4}$, **Figure 9.15**).[52, 53] Y-chromosomal variation in the same panel similarly showed decreases in TMRCA, expansion time, and N_e with distance from East Africa.[57] Such observations emphasize the importance of studying diversity within Africa (**Opinion Box 8**) and suggested a

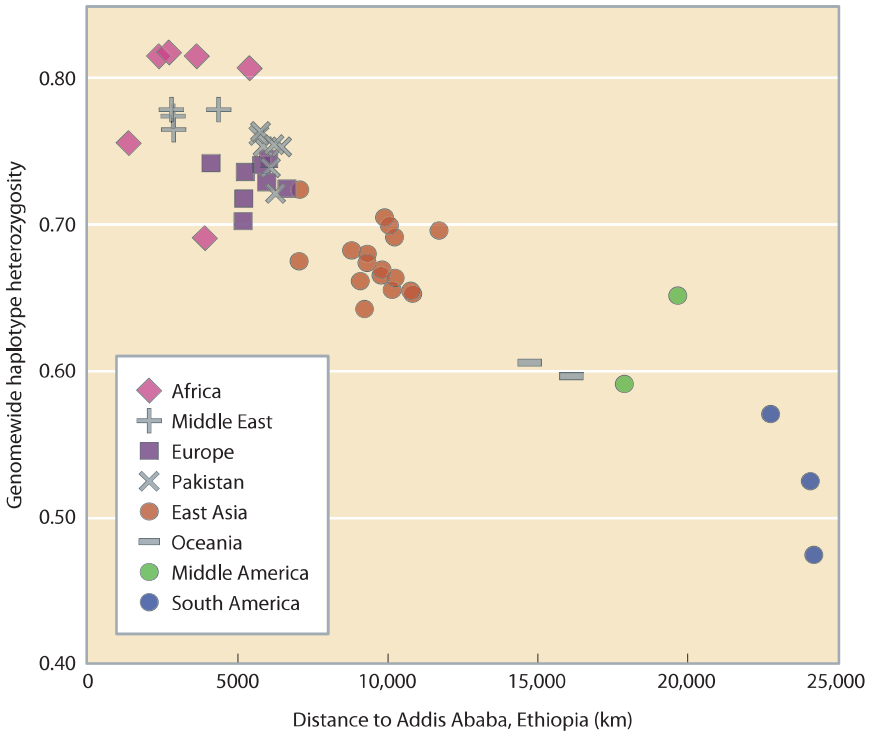

Figure 9.15: Decline of genomewide diversity with distance from East Africa.

Africa is a region of great genetic, linguistic, cultural, and phenotypic diversity. It contains more than 2000 distinct ethno-linguistic groups, speaking nearly a third of the world's languages, and practicing a wide range of subsistence patterns including agriculture, **pastoralism**, and hunting-gathering. Africans live in environments ranging from the world's largest desert and second-largest tropical rainforest to savanna, swamps, and mountain highlands, and these environments have undergone dramatic changes in the past. Differences in diet, climate, and exposure to pathogens among ethnically and geographically diverse African populations are likely to have produced distinct selection pressures, resulting in local genetic adaptations, some of which may play a role in disease susceptibility.

Given this great cultural and environmental variation, several important questions exist: (1) How much genetic structure exists among African populations? (2) How old is that genetic structure? (3) When and where did modern humans originate in Africa? (4) What are the source populations for migration(s) of modern humans out of Africa? and (5) Has introgression from archaic species shaped the African genomic landscape? A study of ~800 microsatellites and ~400 indel polymorphisms genotyped in >2500 Africans indicated high levels of population substructure.[64] Fourteen genetically divergent "ancestral population clusters" correlate with self-described ethnicity and shared cultural and/or linguistic properties. Most African populations have mixed ancestry from these different clusters, reflecting high levels of migration and admixture among ethnically diverse groups. Although some of the inferred ancestral populations likely reflect recent differentiation (for example, eastern and western Niger–Congo speakers which split within the last few thousand years), other structure is likely to be ancient. Indeed, there is evidence of shared common ancestry between several of the major hunter-gatherer populations in Africa who currently reside in Central, Southern, and Eastern Africa (Pygmies, San, Hadza, Sandawe). Analyses of mtDNA and Y-chromosome lineages that differentiated >50 KYA in these populations suggest that the common ancestry could have been quite ancient.

The oldest anatomically modern human fossil, dated to ~195 KYA, was found in southern Ethiopia.[45] However, the most divergent mtDNA, Y chromosome, and autosomal lineages are found in the San hunter-gatherer populations currently residing in southern Africa. Additionally, archaeological data suggest that the earliest modern behavior occurred in both southern and eastern Africa. However, there is a dearth of fossil and archaeological data for modern human origins, particularly from central and western Africa where material is poorly preserved due to the tropical climate. Heterozygosity based on microsatellite and SNP variation is highest in the **click-language**-speaking San, consistent with them being derived from a population ancestral to other populations. Given their current geographic location in southern Africa, Henn et al.[21] argued for a southern African origin of modern humans. However, linguistic data suggest that click languages may have originated in eastern Africa, as far north as Ethiopia. Therefore, it is possible that the San may have originated in eastern Africa and migrated south within the past 10–50 KY.

Furthermore, recent targeted and whole-genome[32] sequencing of African hunter-gatherers suggests that modern humans in Africa may have admixed with archaic populations that were as divergent from modern humans as Neanderthals were. The implication is that introgression from different archaic species occurred across the globe. Thus, although most of the modern human genome originated directly from Africa, some genomic regions have much older lineages that may have originated via non-African groups such as Neanderthals and Denisovans, and as-of-yet unknown archaic African populations. Thus, on a global level, a recent African origin model incorporating low levels of ancestry from local archaic populations is most appropriate. However, within Africa, many questions remain. For example, it is possible that modern human origins could involve multiple locations within the continent, given the closer geographic proximity of populations and opportunities for long-distance gene flow and admixture. Such a model would imply ancient substructure (and hence, ancient lineages) within African genomes (**Figure 1**). Additionally, the transition to modern human morphology could have been gradual, rather than abrupt.[69] Inference of modern human origins within Africa will ultimately require integration of novel paleobiological, archaeological, and whole-genome sequence data from diverse Africans, together with development of sophisticated computational modeling approaches.

Sarah A. Tishkoff, Departments of Genetics and Biology, University of Pennsylvania, USA

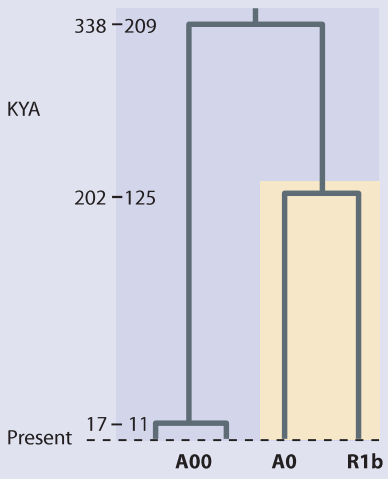

Figure 1: Outstanding questions about Africa Y-chromosomal history.
Until 2012, the first known branch in the Y-chromosomal phylogeny was between haplogroup A0 and the other haplogroups, represented here by R1b (*yellow*). In 2013, a much deeper-rooting haplogroup, A00, was reported as a very rare lineage in African Americans and the Mbo from Cameroon in West Africa (*blue*). Using different calibration scales, the new root could be placed at either 338 or 209 KYA. Does this indicate ancient population structure, archaic introgression or some other complexity of human origins? [Adapted from Mendez FL et al. (2013) *Am. J. Hum. Genet.* 92, 454. With permission from Elsevier.]

serial founder model for human expansion, described below. Although genotyping previously discovered SNPs provides a biased view of variation, haplotypes of ≥20 kb constructed from such genotypes escape this bias and show a similar pattern of highest haplotype diversity in sub-Saharan Africa, decreasing to lowest in South America.[13]

Having established that the highest levels of genetic variability are found in sub-Saharan Africa, it is of interest to ask how precisely the most likely place of origin for modern humans can be pinpointed. This is less straightforward than might be expected, because current genetic patterns in Africa are dominated by the spread of agriculturalists in the last few thousand years, which erased many earlier patterns (**Section 12.6**). Studies based on the HGDP-CEPH samples have pointed to East Africa,[52] while a combination of two studies that incorporated more samples from both hunter-gatherer populations (including the Hadza and Sandawe from Tanzania and ‡Khomani from South Africa) and Ethiopians favored a southern African origin, because the lowest LD values were found in this region.[21, 48] At present, considerable uncertainty remains, and it is possible that modern African populations do not retain sufficient genetic information to reach a clear conclusion about this topic.

Genetic phylogenies mostly root in Africa

The phylogeny of a locus can provide information about the time and place of its origin. This involves some assumptions: (1) that the phylogeny can be reconstructed accurately; and (2) that geographical movement has been limited, so that the modern distribution provides information about the ancient distribution. Molecular phylogenetic information was first applied to the question of human origins when mtDNA data became available. The features of this locus that make it particularly suitable for such studies, and aspects of the nomenclature of clades, are explained in the **Appendix**.

Mitochondrial DNA phylogeny

An mtDNA phylogeny based on the complete mtDNA sequences of 53 individuals of diverse geographical origins, rooted by comparison to a chimpanzee sequence, was constructed by Ingman et al.[27] All sequences were different and 657 variable positions were found, 516 of which were outside the hypervariable control region. Despite the elevated mutation rate of mtDNA compared with nuclear sequences, a robust phylogeny could be obtained from the complete sequence excluding the control region (**Figure 9.16**). This has some striking features:

- Complete separation of African and non-African lineages
- The first three branches lead exclusively to African lineages, while the fourth branch contains both African and non-African lineages

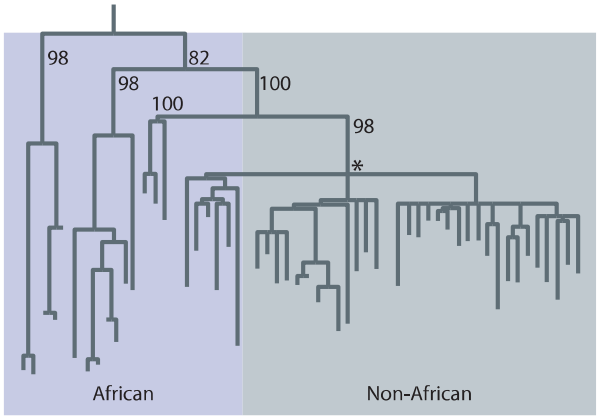

Figure 9.16: Neighbor-joining tree of mtDNA sequences.
The tree was constructed using sequence information from the entire mtDNA genome except the control region. *Blue* shading: African lineages. *Gray* shading: non-African lineages. Numbers indicate the percentage of bootstrap replicates. The asterisk is discussed in the text. [Adapted from Ingman M et al. (2000) *Nature* 408, 708. With permission from Macmillan Publishers Ltd.]

- Deep branches within African lineages, contrasting with **star-like** structure within non-African lineages
- **TMRCA** for the entire phylogeny: 172 ± 50 KY
- TMRCA for the branch containing African plus non-African lineages, marked with an asterisk in Figure 9.16: 52 ± 28 KY
- Expansion time for non-African lineages estimated at 1925 generations or 38.5 KY at 20 years/generation by the authors (48 KY at 25 years/generation)

A study focused specifically on the African lineages investigated 624 complete sequences. It found that the deepest phylogenetic split was between L0 lineages and the rest (L1'5), and proposed that this also corresponded to long-lasting population substructure originating before 90 KYA,[7] thus emphasizing an origin in sub-Saharan Africa and adding more detail to the model.

Y-chromosomal phylogeny

The Y chromosome is also a highly informative locus for such phylogenetic studies (**Appendix**). A Y-chromosomal phylogeny derived from **DHPLC**-based mutation detection in 64 kb of DNA from 43 individuals was, for many years, the largest ascertainment-bias-free global survey.[63] It revealed 56 variants which distinguished 32 lineages that fell into the parsimony tree shown in **Figure 9.17**, again rooted by comparison with other ape sequences. Although less detailed than the mtDNA phylogeny, its structure shows close parallels:

- Complete separation of African and non-African lineages
- The first two branches lead exclusively to African lineages, while the third branch contains both African and non-African lineages
- TMRCA for the entire phylogeny: 59 (40–140) KY, assuming 25 years/generation
- TMRCA for the branch containing African plus non-African lineages, marked with an asterisk in Figure 9.17: 40 (31–79) KY

This **point estimate** (best single estimate, but not taking account of the uncertainty) for the Y phylogeny TMRCA is very recent, and is discussed further in the **Appendix**.

Other phylogenies

Phylogenies from several autosomal and X-chromosomal loci are available. These are potentially complicated by recombination, but by analyzing very closely linked polymorphisms, usually within 10 kb or less, haplotypes showing little recombination can be identified and the effects of recombination minimized or excluded. An examination by Takahata and co-workers[60] in 2001 of

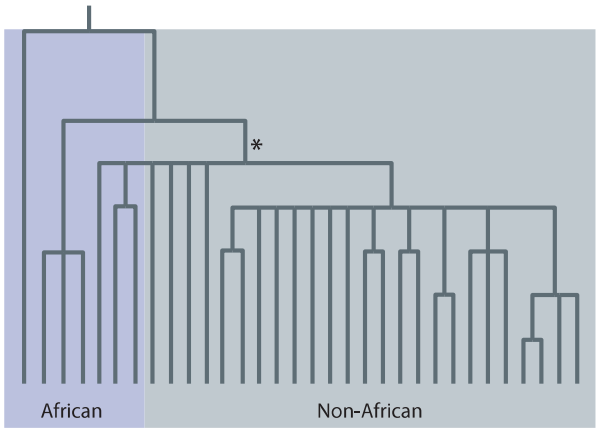

Figure 9.17: Y-chromosomal phylogeny. *Blue* shading: African lineages; *Gray* shading: non-African lineages. The asterisk is discussed in the text. [Adapted from Thomson R et al. (2000) *Proc. Natl Acad. Sci. USA* 97, 7360. With permission from the National Academy of Sciences.]

published data found that it was possible to infer the ancestral origin of a total of 10 loci. Nine of the 10 origins were in Africa, and one was in Asia. The latter, *Glycerol Kinase* (*GK*), was based on a sample size of 10 and a single variable position, so this conclusion might change if more data were available.

Unusual phylogenies including highly divergent non-African haplotypes have been identified in a number of genomic regions, including an inversion polymorphism on chromosome 17q, the microcephalin gene (*MCPH1*), and within the *HLA* region, leading to suggestions that they might represent examples of introgression from other hominins. Available aDNA evidence (see **Section 9.5**) has not supported this scenario for the first two examples, but has led to the remarkable suggestion (awaiting independent confirmation) that more than half of the *HLA-A* ancestry in Europe and Asia derives from archaic hominins.[1] It was proposed that this exceptional ancestry reflects a selective advantage of alleles already adapted to a Eurasian environment for humans migrating out of Africa.

We must remember that these analyses are of *loci*, not populations. If the population size were the same on each continent, the finding of an African origin for at least 9/10 loci would support the out-of-Africa model. In reality, population sizes have not been the same and it is likely that African populations were larger than those on other continents for much of human prehistory, and there is some evidence from phylogeny for ancient non-African contributions to our gene pool. Nevertheless, phylogenetic analyses overwhelmingly point to an African origin for most loci, and the simple conclusion is that most of our ancestors lived in Africa before 60 KYA.

Insights can be obtained from demographic models

It should be possible to formalize the insights from the variation observed in modern DNA into tests that seek to distinguish between explicit alternative models. The strength of such an approach is that it can quantify the likelihood of the alternatives, but the weakness is that available models are grossly oversimplified, and explore only a small proportion of potential models. Two types of model have, however, been used widely.

"Best-fit" demographic models have been sought to model characteristics of ancestral populations that would lead to the observed levels of genetic variation in current populations. **Parameters** have generally included effective population sizes at different times (including bottlenecks, expansions), the order and times at which populations split, and the migration rates between them. Such models of African, European, and East Asian populations have mostly supported an African origin, single exit involving a bottleneck, and large expansions of the European and East Asian populations. Some have suggested a divergence between African and Eurasian populations ~100 KYA[19] (**Figure 9.18a**) or a divergence between Europeans and East Asians as recently as 22.5 KYA[33] (Figure 9.18b). Some models have included archaic admixture as one alternative, and have found support for this, for example 14(2–20)% archaic contribution to Europeans and 1.5(0-5–2.5)% to East Asians[68] (Figure 9.18c).

Serial founder models have sought to capture and suggest explanations for the observed global patterns of genetic variation (**Figure 9.19**): they can incorporate many more populations than the best-fit models above, but use fewer parameters. The underlying observations are that diversity decreases while LD and population differentiation (F_{ST}) increase with distance from Africa,[52, 53] leading to two general conclusions. (1) These trends are linear with migrational (walking) distance, rather than with direct (great circle) distance (**Section 6.8**). (2) Sharp discontinuities, which would imply distinct types of human rather than a continuum, are not seen. A model that explains these observations in a simple and effective way starts from a single source population; a new population is formed by a subset of individuals (that is, founders), and after growth a subset of this subset founds the next population, and so on—this is the "serial" aspect (Figure 9.19).

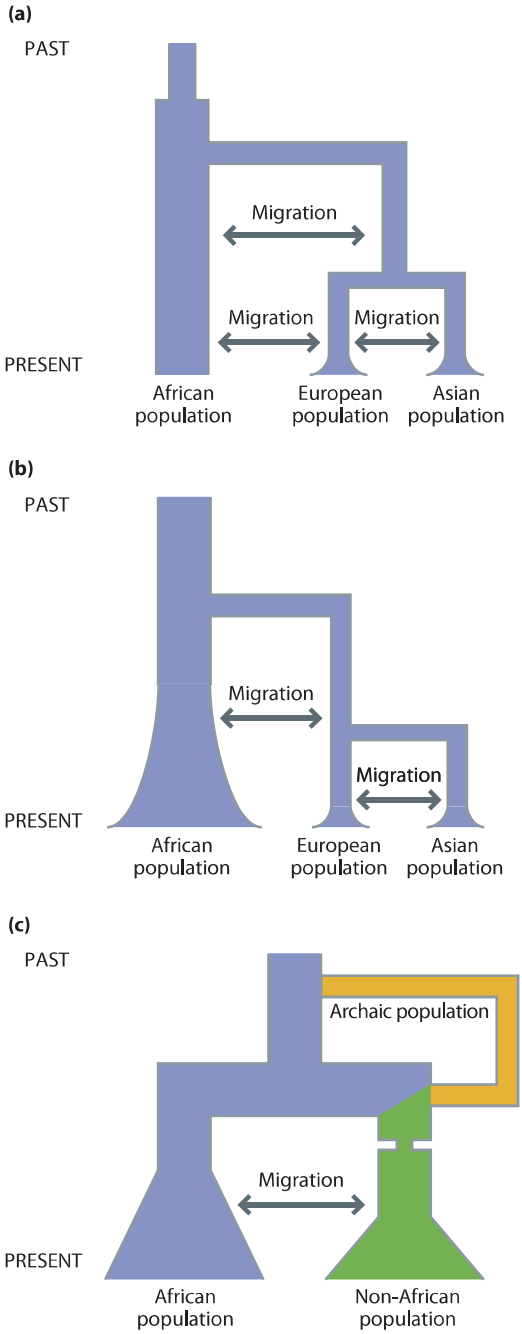

Figure 9.18: Best-fit demographic models.
Modern populations are represented at the bottom of each model, and the past population splits, size changes, and migration events considered by the model are shown. Model (c) includes admixture with archaic hominins (*yellow* and *green*). [a, adapted from Gutenkunst RN et al. (2009) *PLoS Genet.* 5, e1000695. With permission from Public Library of Science. b, adapted from Laval G et al. (2010) *PLoS One* 5, e10284. With permission from Public Library of Science. c, adapted from Wall JD et al. (2009) *Mol. Biol. Evol.* 26, 1823. With permission from Oxford University Press.]

Overall, demographic models support an origin of modern humans in sub-Saharan Africa and a stepwise expansion throughout the rest of the world involving multiple small bottlenecks. But specific details vary so much between models that conclusions drawn from current models need support from independent evidence for credibility.

9.5 EVIDENCE FROM ANCIENT DNA

The analysis of ancient DNA should be an ideal way to distinguish between different hypotheses about the origins of modern humans: by investigating a time series of fossils from any region, it should tell us directly whether there was regional continuity or replacement of early lineages by African ones. Unfortunately, it is impossible to obtain trustworthy DNA sequence data from most fossils: DNA does not survive well, and contaminating DNA, from the

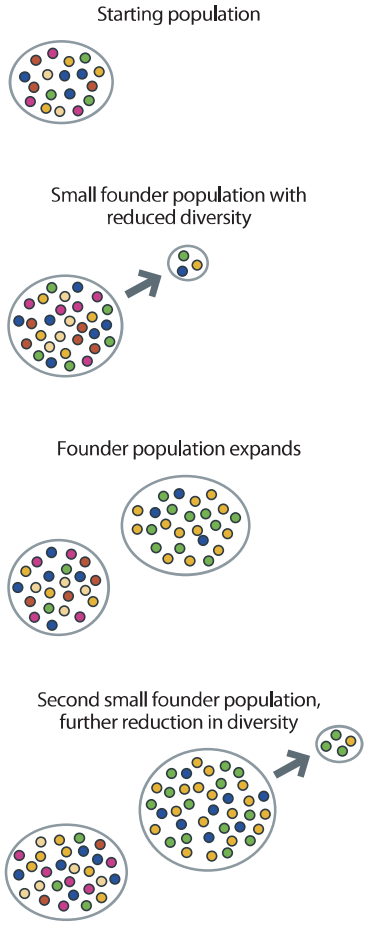

Figure 9.19: Serial founder model.

environment and from people who have handled the fossils or carried out the analysis, can provide a high background (**Section 4.10**). Ancient DNA work is technically demanding and stringent criteria must be met before results can be accepted as authentic. These criteria are discussed in the Opinion Box 2 in Chapter 4. Only with the advent of next-generation sequencing technology (**Section 4.4**) has aDNA fulfilled its potential and transformed our view of human evolution. aDNA analyses of anatomically modern human remains will be discussed in later sections. Here, we will be concerned with aDNA insights from archaic humans: Neanderthals and Denisovans.

Ancient mtDNA sequences of Neanderthals and Denisovans are distinct from modern human variation

Neanderthal mtDNA was an early target for aDNA studies: its generally high copy number in the cell suggested that it might be the easiest aDNA to derive data from, while its phylogenetic informativeness promised insights into the relationship between this extinct branch of hominin and the closely related current modern humans. The determination of a partial mtDNA hypervariable region sequence from the Neanderthal type specimen (called Feldhofer 1 after the cave of origin) in 1997 was thus a major step forward for aDNA workers.[31] Since then, complete mtDNA sequences from six individuals have been determined (**Figure 9.20**),[10] and partial sequences from 10 others. Some general conclusions about Neanderthal mtDNA sequences are possible: (1) Neanderthal mtDNAs are distinct from modern human mtDNAs: for example, a comparison of the six complete Neanderthal mtDNA sequences with 54 present-day humans and one ~30-KY-old early modern human identified an average of 202 substitutions (range 185–220) between Neanderthals and humans, compared with 60 (range 1–106) among this set of modern humans; this corresponds to a divergence time of 466 (321–618) KYA:[30] see **Figure 9.21**. (2) Neanderthal mtDNAs show low diversity, apparently lower even than that within modern humans, and much less than that within other apes: an average of 20 substitutions among the six Neanderthal sequences, corresponding to a coalescence time of ~100 KY.[10] This observation is particularly striking since the six sampling sites range from El Sidron in Spain to Mezmaiskaya in Russia (Figure 9.20), and the fossil dates from ~38 KYA to ~60–70 KYA. (3) Neanderthal mtDNAs are no more similar to Europeans than to other modern humans: for example,[31] Neanderthal–European hypervariable region differences = 28.2 ± 1.9, while Neanderthal–African differences = 27.1 ± 2.2.

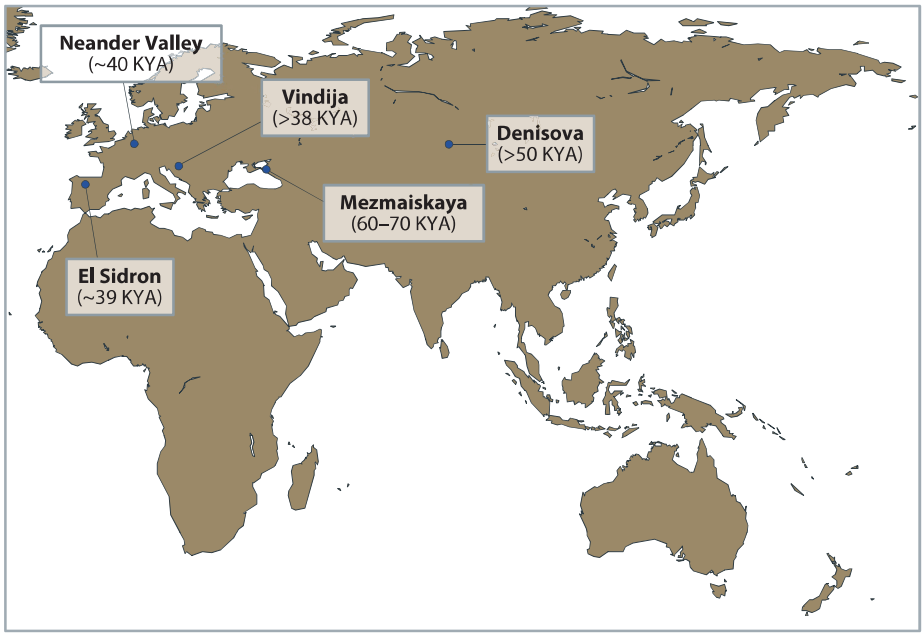

Figure 9.20: Sites of fossils used to generate aDNA data from Neanderthals and Denisovans.

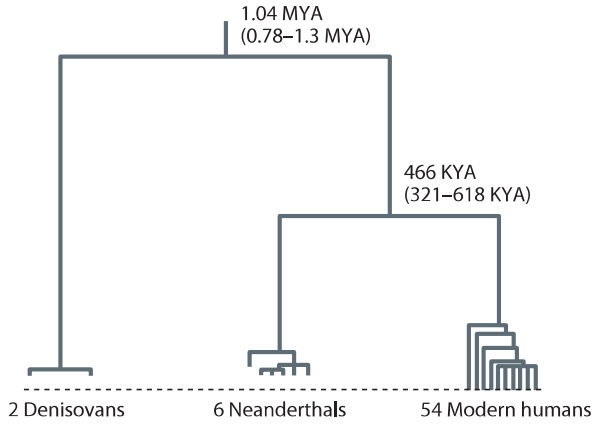

Figure 9.21: Phylogeny of Neanderthal, Denisovan, and modern human mtDNAs.
[Adapted from Krause J et al. (2010) *Nature* 464, 894. With permission from Macmillan Publishers Ltd.]

Indeed, by 2010 aDNA analysis had become such an effective way of characterizing fossils that it was being applied to hominin bones of uncertain affinity in order to identify them. A juvenile hominin "distal manual **phalanx** of the fifth digit" (tip of the little finger) too small to be dated directly but from deposits estimated to date to between 50 and 100 KYA, was excavated from Denisova Cave in the Altai Mountains of Siberia (Figure 9.20), and a 156×-coverage full mtDNA sequence was determined by Illumina GAII™ sequencing.[30] Surprisingly, this sequence did not match either Neanderthals or modern humans, showing on average 385 differences (range 372–396) from 55 modern humans, almost twice as many as from Neanderthals. The new hominins were subsequently designated Denisovans after their place of origin. An mtDNA sequence from a tooth from the same cave differed at just two positions.[54] Assuming a chimpanzee–human split 6 MYA, the Denisovan sequences diverged from the human–Neanderthal mtDNA clade 1.04 (0.78–1.3) MYA (Figure 9.21). The extraordinarily good DNA preservation in the phalanx, partly but not entirely explained by the low temperature in Siberia, allowed a sequence of the entire genome to be determined; this, and the affinity of this enigmatic hominin, is described below.

A Neanderthal draft genome sequence has been generated

The analysis of ancient nuclear DNA is more difficult than for ancient mtDNA because of its much larger size and lower concentration. Studies between 2007 and 2009 investigated individual genes of particular interest involved in skin pigmentation (*MC1R*, **Section 15.3**), speech and language (*FOXP2*, see **Section 8.3**), the ABO blood group (Box 3.1), and phenylthiocarbamide taste perception (*TAS2R38*, **Section 15.5**). The application of new sequencing technologies provided the potential to overcome these limitations, but was initially plagued by contamination, and improved procedures involving tagging of the aDNA library with a short oligonucleotide in the clean room used for extraction, before transfer to the contamination-prone sequencing environment, were developed as a response. Using these, a ~1.3× rough draft sequence of the Neanderthal genome was determined in 2010,[18] providing the basis for the rest of this section; insights into functional variants specific to modern humans are covered in Chapter 8.

The draft Neanderthal sequence was mainly derived from three bones from Vindija Cave in Croatia (Figure 9.20: Vi33.16, Vi33.25, and Vi33.26, yielding 1.2 Gb, 1.3 Gb, and 1.5 Gb, respectively). These came from three different females, although two carried the same mtDNA sequence, and dated to ~45 KYA. Small amounts of additional sequence were generated from additional Neanderthals from El Sidron, Feldhofer, and Mezmaiskaya. The Neanderthal sequence was shown to have <1% contamination with modern human DNA, and thus allowed an initial comparison of the modern human and Neanderthal genomes.

This comparison had to take into account the Neanderthal low coverage and DNA damage, which together resulted in a high error rate: the number of substitutions specific to the Neanderthal lineage was apparently 30× that on the human lineage. Consequently, Neanderthal-specific changes could not be identified with any confidence, but the Neanderthal sequence could be used to identify which changes on the lineage leading to humans arose before the human–Neanderthal split, and which after. Overall, this proportion was 12.7%, corresponding to an average split time ~825 KYA of individual segments of the genome assuming chimpanzee and human DNAs diverged 6.5 MYA. Lineage divergences are always earlier than population divergences, and the ancestral populations of humans and Neanderthals were estimated to have split 270–440 KYA.

The aspect of the Neanderthal genome study that attracted most attention was the comparison of allele sharing between Neanderthals and modern humans from different geographical regions, showing an excess of sharing with all humans outside Africa. SNPs were ascertained by resequencing modern humans and a statistic *D* (here designated Patterson's *D* to distinguish it from other *D* statistics; Table 6.2) was developed to quantify the sharing (see **Section 6.3**). A pair of humans was picked, and a random copy of each binary SNP chosen from each individual. When these differed, and Neanderthals carried the derived allele, the position contributed to *D*, which combined these differences over the genome. If the two humans were equally closely related to Neanderthals, *D* would be zero, but if one were more closely related, *D* would depart from zero. Application of this test led to the following conclusions:

- When the two humans were both from sub-Saharan Africa, or both from outside Africa, *D* was not significantly different from zero.
- However, when any human from inside Africa was compared with any from outside, *D* departed from zero, revealing excess sharing outside Africa.
- Two hypotheses could account for this excess sharing: (1) ancestral population structure, such that humans outside Africa were derived from a source population more closely related to Neanderthals; or (2) ~1–4% gene flow from Neanderthals to non-African humans.
- Green et al. favored hypothesis (2), suggesting mixing in the Middle East soon after the exit from Africa to account for the uniform *D* values in non-African populations. Note that the standard models for the migration out of Africa do not predict contact between humans and Neanderthals (Figure 11.7). An alternative interpretation is presented in **Opinion Box 9**.
- No excess of allele sharing specific to Europeans was found, excluding, at this level of sensitivity, admixture in Europe during the ~10 KY of possible coexistence and contact between ~40 KYA and ~30 KYA.

The possibility of admixture between humans and Neanderthals was particularly intriguing. Although it could not be distinguished from the less exciting explanation of ancestral population structure in this analysis (Opinion Box 9), debate about this issue had hardly begun by the time the Denisovan genome sequence was published later in 2010.

A Denisovan genome sequence has been generated

The extraordinarily good preservation and low contamination level of the Denisovan finger bone described above allowed a 1.9× draft genome sequence[54] to be determined in 2012. This was derived from a female and contained <1% contamination, and the sequence was higher in quality than the Neanderthal genome, in part because of the greater coverage, but also because improvements in aDNA technology allowed enzymatic removal of uracil residues, resulting from damage, from the starting DNA.

Analyses applied to the Neanderthal genome could also be used on the Denisovan sequence. A similar polarization of human lineage variants into ones that arose

The debate on the possible admixture between anatomically modern humans and other hominins has been revolutionized by the recovery of reliable genetic information from ancient specimens. The availability of these genomes opens the possibility of directly testing for localized admixture: in principle, if hybridization only occurred within part of the range of anatomically modern humans, we would expect the hominin genome to be genetically more similar to modern populations in that area than to modern populations in other areas. Indeed, a first analysis of the Neanderthal genome revealed Neanderthals to be genetically more similar to present-day Eurasians than to present-day Africans.[18] This asymmetry has been interpreted as evidence for admixture between Neanderthals and anatomically modern humans during the latter's exit out of Africa. Given that there is no significant difference between Europeans and Asians in their similarity to Neanderthal, it has been argued that such admixture would have had to happen at the very beginning of the out-of-Africa exodus, before the split between these two groups.

A possible complication in interpreting spatial patterns of similarity between any ancient hominin and modern human populations is that, while such differences might arise through recent hybridization, they could also, in principle, be the consequence of population structure in early humans and Neanderthals (that is, a case of **incomplete lineage sorting**—see **Section 7.3**). Because there is both archaeological and genetic evidence for ancient population structure in Africa, the effect of population structure in early humans has to be taken into account. To see how incomplete lineage sorting could generate the observed patterns, let us consider a simple hypothetical scenario. The common ancestor of both anatomically modern humans and Neanderthals would have inhabited the whole of Africa, Europe, and Central Asia at some point in the past, let say half a million years ago. The populations of this ancestor would have most certainly shown isolation by distance, such that populations in the northern part of the African continent would have been more similar to European ones than populations found further south. Approximately 300 KYA, the link between African and European populations was severed by a change in climate, with the European populations differentiating into Neanderthals and the African part of the range eventually becoming anatomically modern humans. It is likely that the population structure found in Africa would have persisted, at least to some extent, thus implying that the northern range of anatomically modern humans in that continent would have been more similar to Neanderthals than the southern range. When anatomically modern humans expanded out of Africa approximately 60–70 KYA, it is then quite likely that populations in the northern part of the range would have contributed most of the colonists due to their proximity to the exit points out of the continent (see **Figure 1**).[16] Thus, it would have been the populations that were more similar to Neanderthals who exited Africa and founded the European and Asian lineages of anatomically modern humans, generating exactly the pattern that we see of equal higher similarity of European and Asians to Neanderthals.

While the above logic cannot disprove admixture, it invalidates current tests for admixture, bringing us back to where we were before the Neanderthal genome was sequenced. It should be noted that the issues described above are a possible complication for any attempt to use geographic patterns in similarity between anatomically modern humans and ancient hominins, irrespective of the hominin in question, and which measure of similarity is used. Simple demographic models (for example, using two populations to represent African structure) are unlikely to capture the subtle patterns generated by the fine-grained structure that is likely to have existed across the whole continent. The only real solution to properly investigate admixture will be to obtain sequences of a number of genomes from ancient hominins (and ideally ancient anatomically modern humans), such that population structure in both sets of populations can be reconstructed. Only then will it be possible to show quantitatively whether differential similarity in certain populations is truly a sign of admixture, or whether population structure is a simpler and more parsimonious explanation.

Andrea Manica and Anders Eriksson,
Department of Zoology, University of Cambridge, UK

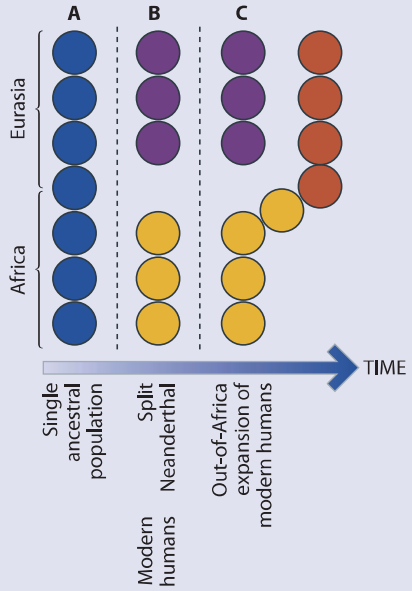

Figure 1: Schematic representation of the relationship between Neanderthal and anatomically modern humans.
A chain of connected populations spanning Africa and Eurasia represents the common ancestor (*blue*). Following the split (B), the northern part of the range became Neanderthals (*purple*), while the African part of the range eventually became modern humans (*orange*). From this African range, modern humans later expand to colonize Eurasia (*red*, C). It is likely that this expansion would have received a large contribution from the northern populations in Africa, which would be more similar to Neanderthals than other African populations due to their proximity to Eurasia before the split. [From Eriksson A & Manica A (2012) *Proc. Natl Acad. Sci. USA* 109, 13956. With permission from Andrea Manica, Cambridge University, UK.]

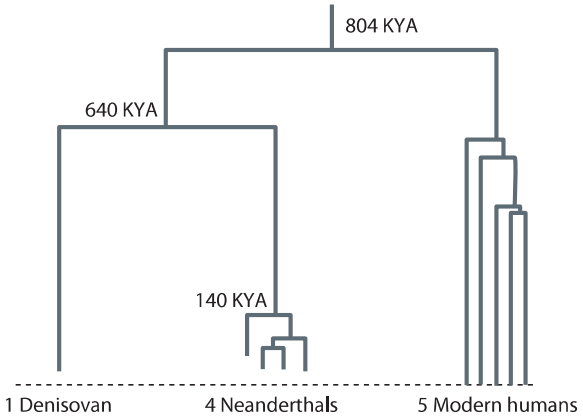

Figure 9.22: Average autosomal sequence divergence times of Neanderthals, Denisovans, and modern humans.
Average autosomal divergences were calculated from low-coverage whole-genome sequences and converted into time estimates assuming that humans and chimpanzees diverged 6.5 MYA. [Data from Green RE et al. (2010) *Science* 328, 710, Reich D et al. (2010) *Nature* 468, 1053.]

before the split with Denisova and after indicated that the lineage split occurred ~760 KYA, not significantly different from the time of the Neanderthal split from the human lineages. Indeed, the same analysis applied to the Neanderthal and Denisova sequences suggested a split between these extinct hominins ~640 KYA. Thus the autosomal analysis indicated shared ancestry between the Neanderthal and Denisovan lineages after the split from humans (**Figure 9.22**), which contrasts with the mtDNA phylogeny (Figure 9.21). This difference could in principle be explained by drift in a large ancestral population, or by introduction of the Denisovan mtDNA lineage from another hominin; in view of the small effective population size inferred for Denisovans as described below, the former explanation appears unlikely. Further comparison of the Neanderthal and Denisovan sequences also suggested a Neanderthal-specific bottleneck, with a divergence date of the Vindija and Mezmaiskaya autosomal DNA sequences 140 ± 30 KYA (Figure 9.22).

Allele sharing was also examined using Patterson's *D* statistic. In contrast to the Neanderthal comparison, no excess sharing between Denisova and all non-Africans was detected. However, excess sharing was seen with Papuan and Melanesian samples, suggesting that 4.8 ± 0.5% of their genomes derive from Denisovans. Excess allele sharing between Denisovans and the ancestors of populations from Oceania cannot be explained by any simple model of ancestral population structure, so this conclusion is most readily accounted for by admixture, but does raise the question of where such admixture might have occurred, since Denisova Cave, in Siberia, lies far away from Oceania. Admixture between modern humans and Denisovans, in turn, makes the possibility of admixture between modern humans and Neanderthals more plausible.

It was not possible to identify, from the morphology of the Denisova finger bone, the source hominin taxon. A tooth (upper molar, either third or second) from the same cave, however, was potentially more informative. As described earlier in this section, this tooth yielded an mtDNA sequence differing at just two positions from the Denisovan reference, and thus is Denisovan. The tooth itself is large, outside the range of Neanderthal and early modern human variation, and, although within the size range of *H. erectus*, still distinguishable from the few known Chinese examples of *H. erectus* second and third molars. It therefore reinforces the distinction between Denisovans, Neanderthals, and modern humans, but leaves the morphological relationships between Denisovans and other hominin taxa in Asia 50–100 KYA unresolved. The discoverers have, with admirable restraint, declined to propose a Linnaean species name for Denisovans.

Subsequent improvements in the construction of libraries for sequencing, in particular the use of single-stranded instead of double-stranded aDNA as the starting material, allowed a higher coverage sequence to be generated from the

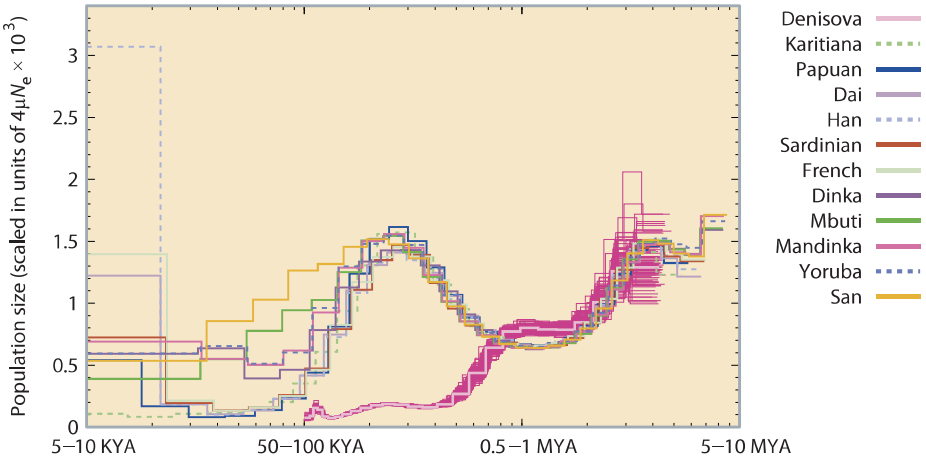

Figure 9.23: Changes in human and Denisovan inferred population size over time.
Past population sizes were estimated from high-coverage whole-genome sequences using the PSMC approach (Chapter 6). PSMC, pairwise sequential Markov coalescence. [Adapted from Meyer M et al. (2012) *Science* 338, 222. With permission from AAAS.]

same finger bone.[47] This more accurate sequence provided several additional insights:

- The Denisovan genome had accumulated 1.16% fewer inferred nucleotide substitutions since the chimpanzee–human common ancestor than 11 present-day humans, suggesting that if this ancestor had lived 6.5 MYA, the Denisovan individual died 74–82 KYA, consistent with the age estimates for the fossil.
- Denisovan heterozygosity was low (2.2×10^{-5}), about one-fifth of the level seen in a present-day African genome. Inference of the change in population size over time (Chapter 6, **Section 6.6**) suggested a demography shared with modern *H. sapiens* ancestors before 1–2 MYA and a decline in numbers after 400–800 KYA, coincidentally or not as modern human ancestors were increasing (**Figure 9.23**). As expected from a small effective population size, the Denisovan genome was enriched for slightly deleterious variants such as nonsynonymous changes.
- Comparison with the human genome led to the identification of 111,812 single-nucleotide changes and 9499 indels where the humans examined were fixed for the derived state and the Denisovan was ancestral; 260 of these coded for amino acid substitutions, including in *CNTNAP2*, a gene regulated by FOXP2 and implicated in language disorders. Further analyses of these fixed differences should yield rich insights into the genetic basis of human uniqueness (Chapter 8).

aDNA studies have thus increased the hominin taxa known from the period 50–150 KYA from four (*H. neanderthalensis*, *H. sapiens*, *H. erectus*, *H. floresiensis*) to five, and provided evidence for low levels of gene flow from both Neanderthals and Denisovans into modern humans. While any level of admixture excludes an extreme out-of-Africa model, the current data suggest a 2.5 ± 0.6% contribution to all non-African populations from Neanderthals and an additional 4.8 ± 0.5% to some Pacific populations from Denisovans, still supporting a predominantly African origin of all modern humans.

SUMMARY

- Information about modern human origins is provided by fossils, archaeological remains, and studies of both present-day human genetic variation and ancient DNA.
- Interpretations of almost all sources of evidence are hotly debated and it is difficult to identify a consensus view about many topics.
- Fossils that date to the approximate time of the chimpanzee–human split, about 5–7 MYA, have been described from three locations in Africa. They

show some features that place them on the human line, including likely upright walking, but their status as possible human ancestors (particularly for the earliest of them, *Sahelanthropus*) is one of the points that remain contentious.

- Several hominin species belonging to the genus *Australopithecus* are known from Africa between about 4.2 MYA and the appearance of *Homo* just after 2 MYA. These specimens are bipedal but have small brains and retain some form of arboreal adaptation.
- The oldest stone tool industry, the Oldowan, occurs after ~2.6 MYA and is associated with slightly larger-brained australopithecines such as *Au. habilis*, *Au. rudolfensis*, and *Au. garhi*.
- The first *Homo* species, *H. erectus*, appeared around 1.9 MYA in Africa, and exhibited a height and weight similar to modern humans, but a smaller brain. *H. erectus* was the first hominin to leave Africa and did so initially with simple Oldowan technology. *H. erectus* is known from Southeast Asia by 1.7 MYA, but hominins have not been found in Europe until after 1.2 MYA.
- After 1.8 MYA, *H. erectus* is associated with the more complex Acheulean technology that is found in Israel and India by 1.5 MYA but not in Europe until after 900 KYA. This technology is absent from Eastern Asia.
- Several later *Homo* taxa are known, including *H. heidelbergensis*, *H. neanderthalensis*, *H. floresiensis*, Denisovans, and *H. sapiens*, although their relationships are still debated. *H. heidelbergensis* was associated initially with Acheulean technology and later Middle Paleolithic or Middle Stone Age technology. While Mousterian technology is generally associated with Neanderthals, it is associated with modern humans and perhaps Denisovans at some sites.
- Modern human morphology is first found in Africa at about 200 KYA, but only much later (after 45 KYA) in other parts of the world.
- These observations led to the development of several hypotheses about the origins of modern humans, two extremes being the multiregional and out-of-Africa models.
- Genetic diversity is higher in Africa than on other continents, which is consistent with a longer period of evolution in Africa, and/or a larger population size.
- Most phylogenies of individual loci show a root in Africa and a subset of lineages in other parts of the world; this is seen particularly clearly in the well-resolved phylogenies of mtDNA and the Y chromosome. Such results imply an African origin for most of our ancestors.
- DNA cannot be extracted from most fossils, but ancient DNA analysis has been successful in generating a draft of Neanderthal genome sequence and a high-coverage Denisovan sequence. Analyses suggested that their genetic lineages diverged from those of modern humans about 800 KYA. However, there is likely to have been a small amount of subsequent mixing as modern humans expanded out of Africa and encountered these species.
- The current consensus view is therefore that an out-of-Africa model with minor archaic admixture explains the fossil, and modern morphological, linguistic, and genetic data most effectively.

QUESTIONS

Question 9–1: Human genetic diversity is generally highest in Africa and decreases with distance from Africa. What explanations could you suggest for finding the following exceptions to this pattern:

(a) An African population with low diversity?

(b) An American population with high diversity?

Question 9–2: To what extent is the assumption that *Sahelanthropus* was a member of the human lineage compatible with genetic data for human origins?

Question 9–3: A haplotype present in many modern humans is also found in Neanderthals. In the light of current interpretations of the Neanderthal genome sequence, what different evolutionary explanations could account for this observation and how might they be distinguished?

Question 9–4: You have managed to generate a draft "hobbit" genome sequence with low contamination, equivalent to the low-coverage Denisova sequence. Given the following genomewide comparisons and calculation of *D*, how would you interpret:

(a) *D* value of zero when comparing a Yoruban and a Japanese genome?

(b) A nonzero *D* value when comparing a Yoruban and a Melanesian genome?

Question 9–5: In a comparison of human, Neanderthal and Denisovan mtDNA sequences, the Denisovan is an outlier, while in an autosomal sequence comparison, the outlier is human (compare Figures 9.21 and 9.22). What explanations can you suggest for these patterns and what additional analyses and datasets could help to distinguish them?

REFERENCES

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

1. **Abi-Rached L, Jobin MJ, Kulkarni S et al.** (2011) The shaping of modern human immune systems by multiregional admixture with archaic humans. *Science* **334**, 89–94.
2. **Asfaw B, Gilbert WH, Beyene Y et al.** (2002) Remains of *Homo erectus* from Bouri, Middle Awash, Ethiopia. *Nature* **416**, 317–320.
3. **Asfaw B, White T, Lovejoy O et al.** (1999) *Australopithecus garhi*: a new species of early hominid from Ethiopia. *Science* **284**, 629–635.
4. **Atkinson QD** (2011) Phonemic diversity supports a serial founder effect model of language expansion from Africa. *Science* **332**, 346–349.
5. **Backwell L & d'Errico F** (2008) Early hominid bone tools from Drimolen, South Africa. *J. Arch. Sci.* **35**, 2880–2894.
6. **Barker G, Barton H, Bird M et al.** (2007) The 'human revolution' in lowland tropical Southeast Asia: the antiquity and behavior of anatomically modern humans at Niah Cave (Sarawak, Borneo). *J. Hum. Evol.* **52**, 243–261.
7. **Behar DM, Villems R, Soodyall H et al.** (2008) The dawn of human matrilineal diversity. *Am. J. Hum. Genet.* **82**, 1130–1140.
8. **Berger LR, de Ruiter DJ, Churchill SE et al.** (2010) *Australopithecus sediba*: a new species of *Homo*-like australopith from South Africa. *Science* **328**, 195–204.
9. **Bermudez de Castro JM, Arsuaga JL, Carbonell E et al.** (1997) A hominid from the lower Pleistocene of Atapuerca, Spain: possible ancestor to Neandertals and modern humans. *Science* **276**, 1392–1395.
10. **Briggs AW, Good JM, Green RE et al.** (2009) Targeted retrieval and analysis of five Neandertal mtDNA genomes. *Science* **325**, 318–321.
11. **Brown P, Sutikna T, Morwood MJ et al.** (2004) A new small-bodied hominin from the Late Pleistocene of Flores, Indonesia. *Nature* **431**, 1055–1061.
12. **Brunet M, Guy F, Pilbeam D et al.** (2002) A new hominid from the Upper Miocene of Chad, Central Africa. *Nature* **418**, 145–151.
13. **Conrad DF, Jakobsson M, Coop G et al.** (2006) A worldwide survey of haplotype variation and linkage disequilibrium in the human genome. *Nat. Genet.* **38**, 1251–1260.
14. **Dart R** (1925) *Australopithecus africanus*: the man-ape of South Africa. *Nature* **115**, 195–199.
15. **Diogo R & Wood B** (2011) Soft-tissue anatomy of the primates: phylogenetic analyses based on the muscles of the head, neck, pectoral region and upper limb, with notes on the evolution of these muscles. *J. Anat.* **219**, 273–359.
16. **Eriksson A & Manica A** (2012) Effect of ancient population structure on the degree of polymorphism shared between modern human populations and ancient hominins. *Proc. Natl Acad. Sci. USA* **109**, 13956–13960.
17. **Ferring R, Oms O, Agusti J et al.** (2011) Earliest human occupations at Dmanisi (Georgian Caucasus) dated to 1.85–1.78 Ma. *Proc. Natl Acad. Sci. USA* **108**, 10432–10436.
18. **Green RE, Krause J, Briggs AW et al.** (2010) A draft sequence of the Neandertal genome. *Science* **328**, 710–722.
19. **Gutenkunst RN, Hernandez RD, Williamson SH & Bustamante CD** (2009) Inferring the joint demographic history of multiple populations from multidimensional SNP frequency data. *PLoS Genet.* **5**, e1000695.
20. **Haile-Selassie Y, Suwa G & White TD** (2004) Late Miocene teeth from Middle Awash, Ethiopia, and early hominid dental evolution. *Science* **303**, 1503–1505.
21. **Henn BM, Gignoux CR, Jobin M et al.** (2011) Hunter-gatherer genomic diversity suggests a southern African origin for modern humans. *Proc. Natl Acad. Sci. USA* **108**, 5154–5162.
22. **Hernandez RD, Kelley JL, Elyashiv E et al.** (2011) Classic selective sweeps were rare in recent human evolution. *Science* **331**, 920–924.
23. **Herries AI** (2011) A chronological perspective on the acheulian and its transition to the Middle Stone Age in southern Africa: the question of the Fauresmith. *Int. J. Evol. Biol.* **2011**, 961401.
24. **Herries AI, Hopley PJ, Adams JW et al.** (2010) Letter to the editor: Geochronology and palaeoenvironments of Southern African hominin-bearing localities—A reply to Wrangham et al., 2009. "Shallow-water habitats as sources of fallback foods for hominins". *Am. J. Phys. Anthropol.* **143**, 640–646.
25. **Herries AIR, Curnoe D & Adams JW** (2009) A multi-disciplinary seriation of early *Homo* and *Paranthropus* bearing palaeocaves in southern Africa. *Quatern. Int.* **202**, 14–28.

26. **Indriati E, Swisher CC 3rd, Lepre C et al.** (2011) The age of the 20 meter Solo River terrace, Java, Indonesia and the survival of *Homo erectus* in Asia. *PLoS One* **6**, e21562.

27. **Ingman M, Kaessmann H, Pääbo S & Gyllensten U** (2000) Mitochondrial genome variation and the origin of modern humans. *Nature* **408**, 708–713.

28. **Johanson DC & White TD** (1979) A systematic assessment of early African hominids. *Science* **203**, 321–330.

29. **Kivell TL, Kibii JM, Churchill SE et al.** (2011) *Australopithecus sediba* hand demonstrates mosaic evolution of locomotor and manipulative abilities. *Science* **333**, 1411–1417.

30. **Krause J, Fu Q, Good JM et al.** (2010) The complete mitochondrial DNA genome of an unknown hominin from southern Siberia. *Nature* **464**, 894–897.

31. **Krings M, Stone A, Schmitz RW et al.** (1997) Neandertal DNA sequences and the origin of modern humans. *Cell* **90**, 19–30.

32. **Lachance J, Vernot B, Elbers CC et al.** (2012) Evolutionary history and adaptation from high-coverage whole-genome sequences of diverse African hunter-gatherers. *Cell* **150**, 457–469.

33. **Laval G, Patin E, Barreiro LB & Quintana-Murci L** (2010) Formulating a historical and demographic model of recent human evolution based on resequencing data from noncoding regions. *PLoS ONE* **5**, e10284.

34. **Leakey LSB** (1959) A new fossil skull from Olduvai. *Nature* **184**, 491–493.

35. **Leakey LSB, Tobias PV & Napier JR** (1964) A new species of genus *Homo* from Olduvai Gorge. *Nature* **202**, 7–9.

36. **Leakey MD & Hay RL** (1979) Pliocene footprints in the Laetolil Beds at Laetoli, northern Tanzania. *Nature* **278**, 317–323.

37. **Leakey MG, Feibel CS, McDougall I & Walker A** (1995) New four-million-year-old hominid species from Kanapoi and Allia Bay, Kenya. *Nature* **376**, 565–571.

38. **Leakey MG, Spoor F, Brown FH et al.** (2001) New hominin genus from eastern Africa shows diverse middle Pliocene lineages. *Nature* **410**, 433–440.

39. **Leakey MG, Spoor F, Dean MC et al.** (2012) New fossils from Koobi Fora in northern Kenya confirm taxonomic diversity in early *Homo*. *Nature* **488**, 201–204.

40. **Lepre CJ, Roche H, Kent DV et al.** (2011) An earlier origin for the Acheulian. *Nature* **477**, 82–85.

41. **Li JZ, Absher DM, Tang H et al.** (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104.

42. **Lieberman DE, McBratney BM & Krovitz G** (2002) The evolution and development of cranial form in *Homo sapiens*. *Proc. Natl Acad. Sci. USA* **99**, 1134–1139.

43. **Lordkipanidze D, Jashashvili T, Vekua A et al.** (2007) Postcranial evidence from early *Homo* from Dmanisi, Georgia. *Nature* **449**, 305–310.

44. **Manica A, Amos W, Balloux F & Hanihara T** (2007) The effect of ancient population bottlenecks on human phenotypic variation. *Nature* **448**, 346–348.

45. **McDougall I, Brown FH & Fleagle JG** (2005) Stratigraphic placement and age of modern humans from Kibish, Ethiopia. *Nature* **433**, 733–736.

46. **Mercader J, Barton H, Gillespie J et al.** (2007) 4300-year-old chimpanzee sites and the origins of percussive stone technology. *Proc. Natl Acad. Sci. USA* **104**, 3043–3048.

47. **Meyer M, Kircher M, Gansauge MT et al.** (2012) A high-coverage genome sequence from an archaic Denisovan individual. *Science* **338**, 222–226.

48. **Pagani L, Kivisild T, Tarekegn A et al.** (2012) Ethiopian genetic diversity reveals linguistic stratification and complex influences on the Ethiopian gene pool. *Am. J. Hum. Genet.* **91**, 83–96.

49. **Pappu S, Gunnell Y, Akhilesh K et al.** (2011) Early Pleistocene presence of Acheulian hominins in South India. *Science* **331**, 1596–1599.

50. **Pickering R, Dirks PH, Jinnah Z et al.** (2011) *Australopithecus sediba* at 1.977 Ma and implications for the origins of the genus *Homo*. *Science* **333**, 1421–1423.

51. **Pruetz JD & Bertolani P** (2007) Savanna chimpanzees, *Pan troglodytes verus*, hunt with tools. *Curr. Biol.* **17**, 412–417.

52. **Prugnolle F, Manica A & Balloux F** (2005) Geography predicts neutral genetic diversity of human populations. *Curr. Biol.* **15**, R159–160.

53. **Ramachandran S, Deshpande O, Roseman CC et al.** (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl Acad. Sci. USA* **102**, 15942–15947.

54. **Reich D, Green RE, Kircher M et al.** (2010) Genetic history of an archaic hominin group from Denisova Cave in Siberia. *Nature* **468**, 1053–1060.

55. **Rosenberg NA, Pritchard JK, Weber JL et al.** (2002) Genetic structure of human populations. *Science* **298**, 2381–2385.

56. **Senut B, Pickford M, Gommery D et al.** (2001) First hominid from the Miocene (Lukeino Formation, Kenya). *C. R. Acad. Sci.* **332**, 137–144.

57. **Shi W, Ayub Q, Vermeulen M et al.** (2010) A worldwide survey of human male demographic history based on Y-SNP and Y-STR data from the HGDP-CEPH populations. *Mol. Biol. Evol.* **27**, 385–393.

58. **Stringer C & Andrews P** (2011) The Complete World of Human Evolution. Thames and Hudson.

59. **Swisher CC 3rd, Curtis GH, Jacob T et al.** (1994) Age of the earliest known hominids in Java, Indonesia. *Science* **263**, 1118–1121.

60. **Takahata N, Lee SH & Satta Y** (2001) Testing multiregionality of modern human origins. *Mol. Biol. Evol.* **18**, 172–183.

61. **The 1000 Genomes Project Consortium** (2010) A map of human genome variation from population-scale sequencing. *Nature* **467**, 1061–1073.

62. **Thieme H** (1997) Lower Palaeolithic hunting spears from Germany. *Nature* **385**, 807–810.

63. **Thomson R, Pritchard JK, Shen P et al.** (2000) Recent common ancestry of human Y chromosomes: evidence from DNA sequence data. *Proc. Natl Acad. Sci. USA* **97**, 7360–7365.

64. **Tishkoff SA, Reed FA, Friedlaender FR et al.** (2009) The genetic structure and history of Africans and African Americans. *Science* **324**, 1035–1044.

65. **van Schaik CP, Ancrenaz M, Borgen G et al.** (2003) Orangutan cultures and the evolution of material culture. *Science* **299**, 102–105.

66. **Wagner GA, Krbetschek M, Degering D et al.** (2010) Radiometric dating of the type-site for *Homo heidelbergensis* at Mauer, Germany. *Proc. Natl Acad. Sci. USA* **107**, 19726–19730.

67. **Walker A & Leakey REF** (eds) (1993) The Nariokotome *Homo erectus* Skeleton. Harvard University Press.

68. **Wall JD, Lohmueller KE & Plagnol V** (2009) Detecting ancient admixture and estimating demographic parameters in multiple human populations. *Mol. Biol. Evol.* **26**, 1823–1827.

69. **Weaver TD** (2012) Did a discrete event 200,000–100,000 years ago produce modern humans? *J. Hum. Evol.* **63**, 121–126.

70. **White TD, Asfaw B, Beyene Y et al.** (2009) *Ardipithecus ramidus* and the paleobiology of early hominids. *Science* **326**, 75–86.

71. **White TD, Asfaw B, DeGusta D et al.** (2003) Pleistocene *Homo sapiens* from Middle Awash, Ethiopia. *Nature* **423**, 742–747.

72. **White TD, WoldeGabriel G, Asfaw B et al.** (2006) Asa Issie, Aramis and the origin of Australopithecus. *Nature* **440**, 883–889.

73. **Whiten A, Goodall J, McGrew WC et al.** (1999) Cultures in chimpanzees. *Nature* **399**, 682–685.

74. **Wood B** (1996) Human evolution. *BioEssays* **18**, 945–954.

75. **Wood B & Collard M** (1999) The human genus. *Science* **284**, 65–71.

76. **Wood B & Harrison T** (2011) The evolutionary context of the first hominins. *Nature* **470**, 347–352.

77. **Yamei H, Potts R, Baoyin Y et al.** (2000) Mid-Pleistocene Acheulean-like stone technology of the Bose basin, South China. *Science* **287**, 1622–1626.

SECTION 4

HOW DID HUMANS COLONIZE THE WORLD?

The transformation of humans from a rare African species into a numerous one with a worldwide distribution is an unprecedented biological phenomenon, and is central to understanding why humans are genetically so similar to one another, and explaining the small, but appreciable, geographical differences that do exist among human populations. We continue along a path that is approximately chronological, discussing the early movements of modern humans out of Africa before considering the major effects that have followed the subsequent introduction of farming and the meeting of populations.

CHAPTER 10
THE DISTRIBUTION OF DIVERSITY

Studying human diversity raises important ethical and methodological questions, and we begin this chapter with these. Next, we see that most human genetic variation is found within any individual population, except for a few loci affected by natural selection.

CHAPTER 11
THE COLONIZATION OF THE OLD WORLD AND AUSTRALIA

A key event was the development, in Africa, of modern human *behavior* 100,000–60,000 years ago. A single expansion soon afterward peopled most of Asia, Australia, and Europe by 40,000 years ago, and included mixing with Neanderthals and Denisovans along the way.

CHAPTER 12
AGRICULTURAL EXPANSIONS

When the climate warmed and stabilized 10,000 years ago, agriculture appeared independently in several locations. Agriculture led to an enormous increase in the number of people, but did the farmers themselves expand, or did the neighbors learn farming and then expand themselves?

CHAPTER 13
INTO NEW-FOUND LANDS

In this chapter we consider the last great regions to be inhabited. In the Americas, most of the current indigenous gene pool may date back to one migration before 15,000 years ago. Most of the Pacific region was uninhabited until 3500 years ago, and the major migration was from the west.

CHAPTER 14
WHAT HAPPENS WHEN POPULATIONS MEET

This chapter considers the mixing of populations, or admixture. This is usually sex-biased, affecting the mtDNA, autosomes, and Y chromosome differentially, and establishing linkage disequilibrium: a legacy that persists for many generations.

THE DISTRIBUTION OF DIVERSITY

CHAPTER TEN

In the last chapter, we saw that our species *Homo sapiens* originated in Africa very recently on an evolutionary time-scale, <200 KYA. Now, humans have a worldwide distribution. How did this transformation from a rare tropical population to one with seven billion people inhabiting every continent come about? The key starting point is an understanding of patterns of genetic variation in different populations and that is the subject of this chapter. We have the molecular and statistical tools to generate and interpret such data. Yet, since we want to investigate the full range of human diversity and focus on the differences between populations, we often need the consent of some of the most disadvantaged people in the world to carry out these studies, and may produce findings that are not value-free, may conflict with the sample donors' beliefs, and are always open to misuse and misinterpretation. We therefore begin by considering the distressing history of this field, and the lessons we can learn about how best to carry out such studies in an appropriate way. We will discuss how to sample human genetic diversity, and encounter the leading international projects in this field. These will reveal both how neutral variants are distributed in the world, and how departures from these general patterns can inform us about natural selection and unusual demographic events.

10.1 STUDYING HUMAN DIVERSITY

Observations and inquiries in the area of human diversity are probably as old as humanity, and have in the past been linked to both blatant and subtle forms of **racism**. We will begin by considering some of the historical aspects of the field, ethical issues raised by such work, and the ways in which modern studies have learned from this early history.

The history and ethics of studying diversity are complex

Should we study human genetic diversity at all, or is this an area of work where the potential for misuse (see **Figure 10.1** and **Box 10.1**) outweighs the potential benefits to such an extent that it should not be pursued? Such studies already have a long history, and a pragmatic answer to this question is that information on human genetic diversity is needed, and is therefore generated, for medical and forensic applications, and thus is already available whatever evolutionary geneticists decide to do, so we must be ready to consider its implications and consequences. Furthermore, genetic information, in fact, refutes any scientific basis for racism as the existence of discrete human groups. It can therefore be used to argue that racism—the belief that discrimination between apparent groups is justifiable—is an entirely social construct. In addition, genetic information is of enormous intrinsic interest to many people, not just scientists.

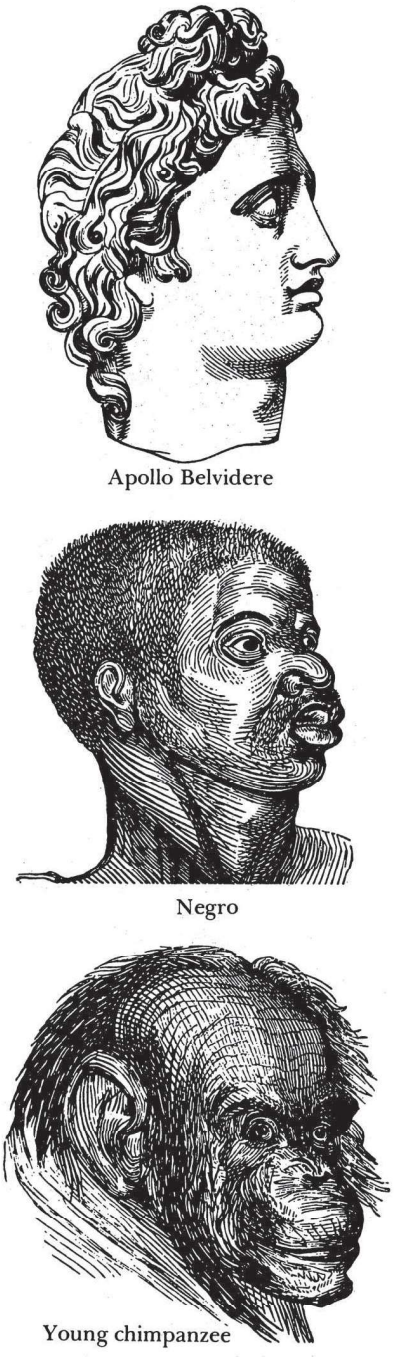

Figure 10.1: A racist view of humanity. Note not only the hierarchy, but also the falsification of Negro and chimpanzee skulls. [From Nott JC and Gliddon GR (1868) Indigenous Races of the Earth. Ayer Co Publishing Inc.]

Linnaeus' classification of human diversity

Our biological classification system originates with Linnaeus (1707–1778), who subdivided humans into two species (*diurnus* and *nocturnus*; see **Figure 10.2**) and a total of seven categories:

- *diurnus* (also referred to as *Homo sapiens* by Linnaeus)
  - *americanus*: red, with black hair and a scanty beard, obstinate, free, painted with fine red lines, regulated by customs
  - *europeus*: white, long flowing hair, blue eyes, sanguine, muscular, inventive, covered with tight clothing, governed by laws
  - *asiaticus*: yellow, melancholy, black hair and brown eyes, severe, haughty, stingy, wears loose clothing, governed by opinions
  - *afer* (that is, African): black, cunning, phlegmatic, black curly hair, women without shame and lactate profusely, anointed with grease, ruled by impulse
  - *monstrosus*: a miscellaneous collection including dwarfs and large, lazy Patagonians
- *nocturnus*
  - *troglodytes*: nocturnal, hunts only at night, lives underground

In some classifications he also included *ferus*: wild, hairy, runs about on all fours. Apart from the misleading notion that humans can be categorized into a small number of such groups and the language that now sounds offensive, this classification is notable for its inclusion of imaginary categories, and mixing of physical, intellectual, and cultural characteristics.

Galton's "Comparative worth of different races"

In *Hereditary Genius: an Inquiry Into its Laws and Consequences* published in 1869, Francis Galton (1822–1911) established a grading system, A, B, C, and so forth, for people within each race, and then compared the grades between races. At the top of the racial hierarchy were the ancient Greeks; Galton was slightly critical of his own race, the English, "the calibre of whose intellect is easily gauged by a glance at the contents of a railway book-stall," and placed them two grades below the Greeks, although "the average standard of the Lowland Scotch and the English North-country men is decidedly a fraction of a grade superior to that of the ordinary English." Inevitably, "the average intellectual standard of the negro race is some two grades below our own" and "the Australian type is at least one grade below the African negro." We see here and in other attempts to identify discrete categories of people how the numbers of categories and criteria used to define them differ substantially.[2, 7]

Modern attitudes to studying diversity

If genetic diversity studies of humans are an acceptable part of science, what are the prerequisites for such studies? There is a fundamental requirement, enshrined in the Declaration of Helsinki and recognized by international law, that research on humans can only be undertaken with **informed consent** from the subject. This means that, with some exceptions for forensic investigations and limited medical circumstances, individuals must not only freely agree to the research before it is undertaken, but must do this on the basis of an understanding of the nature and purpose of the research, its risks and benefits, and the potential outcomes/information produced (see **Opinion Box 10**). The risks associated with the physical procedures of donating saliva, cheek cells, hair, or blood to provide DNA are minimal; debate has focused on the risks associated with the use of the information obtained.[4] The results may have implications for:

- Health and, in some countries, health insurance: what if the donor is found to have a high chance of developing a particular disease later in life?

Box 10.1: Race and racism

> On May 7, 1876, Truganini, the last full-blood Black person in Tasmania, died at seventy-three years of age. Her mother had been stabbed to death by a European. Her sister was kidnapped by Europeans. Her intended husband was drowned by two Europeans in her presence, while his murderers raped her.
>
> It might be accurately said that Truganini's numerous personal sufferings typify the tragedy of the Black people of Tasmania as a whole. She was the very last. 'Don't let them cut me up,' she begged the doctor as she lay dying. After her burial, Truganini's body was exhumed, and her skeleton, strung upon wires and placed upright in a box, became for many years the most popular exhibit in the Tasmanian Museum and remained on display until 1947. Finally, in 1976—the centenary year of Truganini's death—despite the museum's objections, her skeleton was cremated and her ashes scattered at sea.

From *Black War: the Destruction of the Tasmanian Aborigines* by Runoko Rashidi, http://www.cwo.com/~lucumi/tasmania.html

The genocide of the Tasmanian Aborigines by the European settlers in the nineteenth century provides one of the worst of many examples of racism, and is notable for the "anthropological" justification of the public exhibition of Truganini's, and other, remains. Yet racism does not consist only of such crude episodes: racist thinking penetrates deeply into Western, and perhaps all, culture, evolutionary thought, and genetics. Consider these two quotations:

> I advance it, therefore, as a suspicion only, that the blacks, whether originally a distinct race, or made distinct by time and circumstance, are inferior to the whites in the endowment both of body and mind.
>
> We hold these truths to be self-evident: that all men are created equal.

Both are from Thomas Jefferson, and the next is Charles Darwin, writing about the gap between humans and apes after an anticipated future extinction of gorillas and "Hottentots":

> The break will then be rendered wider, for it will intervene between man in a more civilized state, as we may hope, than the Caucasian, and some ape as low as a baboon, instead of as at present between the negro or Australian and the gorilla.

Hierarchies of humans and apes, such as that illustrated in Figure 10.1, were common in anthropological and biological literature. They were usually based on a small number of visible characteristics such as skin color, hair color and morphology, and facial features.

These characteristics are influenced by both environmental and genetic factors, but even if allowance is made for the environment, the genes affecting these phenotypes have probably been subject to particular selection pressures and perhaps sexual selection (Section 15.3). Thus they are unrepresentative of the majority of the genome. Inevitably, the compilers put their own group at the top and those they wanted to exploit at the bottom (see Figure 10.1). Another notable feature of these schemes was that the number of "races" identified varied greatly between authors.

"Race," in addition to its everyday usages, is a biological term with a clear meaning: it refers to a group of individuals who can be cleanly distinguished from other groups of the same species. Of course, this requires that we specify what is meant by "cleanly," and an F_{ST} value ≥0.25 is commonly used: that is, 25% or more of the variation needs to be found between groups for these groups to be classified as "races." Some species are divided into races; the question of whether or not humans are such a species is an empirical one. We will see that the answer is "no."

- Stigmatization: what if the donor carries a trait judged to be undesirable, or is assigned to a group that experiences discrimination because of its identity? Related to this, what if a particular trait or disease becomes associated with a population, and the entire population is stigmatized as a result?
- Commercial applications: what if a **cell line** or DNA sequence leads to a patentable or saleable product?

Additional novel questions are raised by genetic research because we share DNA variants with our relatives, so study of one individual provides information about other members of their family and population. Therefore **group informed consent** is required in some situations and it would be unethical to sample consenting individuals from a group that had not given consent. The appropriate authority to provide such group consent, if there is one, can only be determined on a case-by-case basis for each population.

Benefits from genetic diversity studies are: (1) increased understanding of genetic history and relationships; (2) medical advances such as the identification

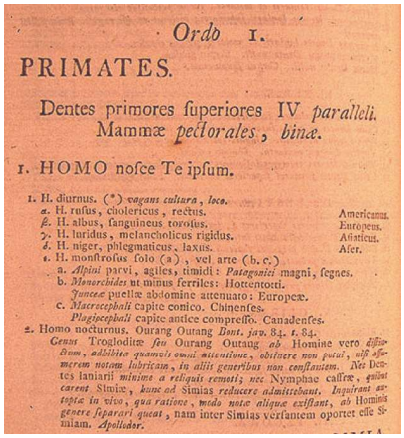
Ordo I.
PRIMATES.
Dentes primores superiores IV paralleli.
Mammæ pectorales, binæ.
1. HOMO nosce Te ipsum.
1. H. diurnus. (*) vagans cultura, loco.
α. H. rufus, cholericus, rectus. Americanus.
β. H. albus, sanguineus torosus. Europæus.
γ. H. luridus, melancholicus rigidus. Asiaticus.
δ. H. niger, phlegmaticus, laxus. Afer.
ε. H. monstrosus solo (a), vel arte (b. c.)
a. Alpini parvi, agiles, timidi: Patagonici magni, segnes.
b. Monorchides ut minus ferriles: Hottentotti.
Junceæ puellæ abdomine attenuato: Europeæ.
c. Macrocephali capite conico. Chinenses.
Plagiocephali capite antice compresso. Canadenses.
2. Homo nocturnus. Ourang Outang Dant. jun. 84. t. 84.
Genus Troglodytæ seu Ourang Outang ab Homine vero differentem, adhibito quamvis omni attentione, obtinere non potui, nisi observarem notam lubricam, in aliis generibus non conflantem. Nec Dentes laniarii minime a reliquis remoti; nec Nymphae callosæ, quibus carent Simiæ, hanc ad Simias reducere admittebant. Inquirant autem in vivo, qua ratione, modo notæ aliquæ existant, ab Hominis genere separari queat, nam inter Simias versantem oportet esse Similam. Apollodor.

Figure 10.2: Linnaeus' 1756 classification of humans.

OPINION BOX 10: Engagement with the indigenous populations of Peninsular Malaysia

The indigenous populations that occupy Peninsular Malaysia are locally known as *Orang Asli*, or the "original peoples." Collectively, they represent approximately 0.6% of Malaysians, and have been classified into three distinct groups—the Negrito, Senoi, and Proto-Malay—based on differences in languages, socio-cultural practices, physical appearance, and habitats. Each group can be further subdivided into six subgroups (**Table 1**, **Figure 1**). Based on the studies of their genetic history, the *Orang Asli* are thought to be the descendants of humans who arrived in Southeast Asia some 60 KYA.[9]

There are 869 recorded *Orang Asli* settlements in Malaysia. Only 1.4% of these are in or close to urban centers, whilst the majority are located in the rural and forest areas. The past decade has witnessed the relocation of some forest fringe communities into government resettlement schemes closer to urban areas. Whilst a significant proportion of Proto-Malays and Senoi work in orchards, plantations, and the fishing industry, the majority of Negrito communities still depend largely on foraging and collection of jungle produce for sale.

The health of most Malaysians has improved in the twentieth century, but communicable diseases abound in *Orang Asli* communities, especially in rural areas with poor access to health care. As a result of increased resettlement and adoption of a sedentary lifestyle, non-communicable diseases, including hypertension, obesity, and cardiovascular diseases, have also increased.

Fieldwork and genetic research amongst the *Orang Asli* require approval from the Malaysian Department of Indigenous Development for each project, community, and study time frame. It is important to include anthropologists in any research team so that the researchers appreciate the different cultural systems in different communities. For example, the Temuan people have firm hierarchies, and the headman "Tok Batin" and his elders must first be consulted during a customary courtesy visit, before any fieldwork can begin. Only with their consent can other members of the community be recruited. Translators are often necessary, as the elderly *Orang Asli* only speak their own languages. Many settlements are remote and only accessible by dirt tracks or by boats in good weather. These settlements may also lack electricity and running water. As a consequence of the history of resettlement and the loss of native lands, some *Orang Asli* can be hostile toward outsiders.

To build trust with the community, it is mandatory to pay courtesy calls before researching any community to explain to the elder members of the tribes the rationale of the study and how samples would be collected. Our sampling activities also include the sharing of direct benefits, addressing the communities' requests for sundry provisions and clothing. We conduct health screening and the physicians from our team conduct basic clinical examinations for any member of the community who requests it, regardless of whether or not they participate in the research project. In addition, we revisit the villages annually and it has been possible to provide health reports to the relevant individuals and health officers in charge for follow ups and treatment in the local government health centers.

TABLE 1:
MAJOR GROUPS AND SUBGROUPS OF *ORANG ASLI* AND THEIR POPULATION SIZES

Major groups	Subgroups	Population size
Negrito	Kensiu	240
	Kintak	132
	Lanoh	349
	Jahai	2072
	Mendriq	216
	Bateq	1542
Senoi	Temiar	25,233
	Semai	43,505
	Semoq Beri	3629
	Che Wong	665
	Jahut	5082
	Mah Meri	2858
Proto-Malay	Temuan	22,819
	Semelai	6584
	Jakun	29,263
	Kanaq	87
	Kuala	3716
	Seletar	1431

(From data collected in 2004 by the Jabatan Kemajuan Orang Asli, Malaysia.)

Boon-Peng Hoh, Institute of Medical Molecular Biotechnology, Universiti Teknologi MARA, Selangor, Malaysia
Maude E. Phipps, Jeffrey, Cheah School of Medicine and Health Sciences, Monash University Sunway Campus, Selangor, Malaysia

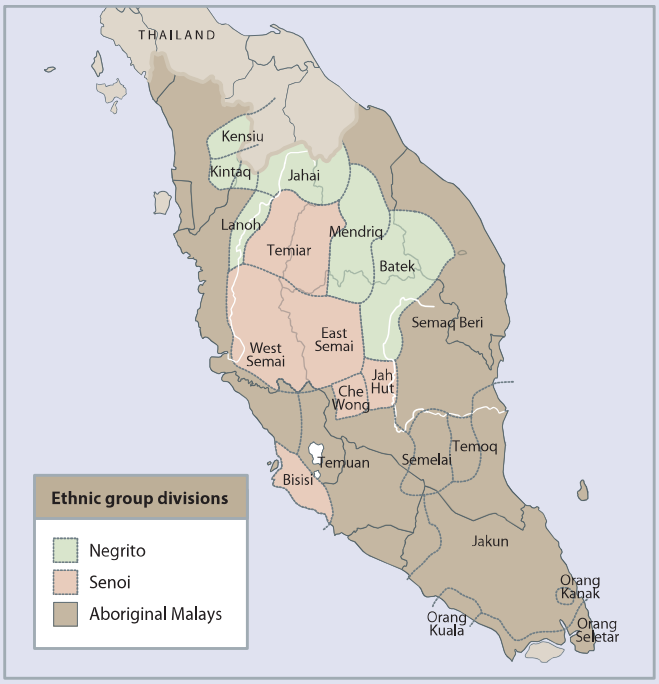

Figure 1: Locations of major groups and subgroups of *Orang Asli* in Malaysia.

of genes predisposing to disease (see Chapter 16); (3) accurate paternity testing, victim and assailant identification, and other forensic applications (see Chapter 18); (4) sometimes, immediate benefits to the population, such as medical advice or treatment. However, complications also arise, for example because the people who receive most of the long-term benefits may not be the donors.

Outstanding issues that have not been fully resolved include:

- Is informed consent from members of cultures that do not ascribe to Western scientific values truly "informed"? Indeed, can even leading geneticists such as Jim Watson and Craig Venter, who have volunteered to have their whole genomes sequenced and made public (Chapter 18), appreciate the full implications when these may only become apparent in the future as research reveals the medical implications of DNA variants?
- How much information about the donor should accompany a cell line or DNA sample, so that the privacy of the donor is not infringed?
- Can samples collected with no written consent many years ago, or perhaps decades ago, still be used?
- Can samples collected for one study be used in another?
- Can an individual give broad consent for all future studies, which may involve techniques that do not yet exist and have implications that are not currently understood (related to the first point, above)?

It is difficult to give general answers to many of the ethical questions that diversity studies raise; indeed, the possibility of ever more comprehensive genetic studies is one of the driving forces in the field of medical ethics. Answers may emerge more satisfactorily through the consideration of individual cases than through prior reasoning based on principles. We will encounter examples of such cases throughout this book.

Who should be studied?

The starting point for any study of human diversity is a set of humans, and this raises the question, Who should be studied? Sampling always creates problems: is the sample appropriate and representative? If not, conclusions drawn from the sample may not be applicable to the rest of the population that was sampled. Analyzing everyone would avoid the complications introduced by sampling, and some have argued that it is fairer, but at present it is impractical for DNA studies and even for DNA-free genetics based, for example, on phenotypic traits. This is likely to remain true for the foreseeable future, so the issues raised by sampling must be addressed.

Although human genetic diversity has been investigated for a long time, the early studies using pre-DNA polymorphisms aroused little controversy or public interest. Attitudes changed with the launch of the **Human Genome Diversity Project** (**HGDP**) in 1991, and all subsequent large-scale projects, including the **HapMap**, **Genographic**, and **1000 Genomes Projects**, have been influenced by this legacy.

A few large-scale studies of human genetic variation have made major contributions to human evolutionary genetics

The HGDP was announced in a paper published by Cavalli-Sforza, Wilson, and others[5] (http://hsblogs.stanford.edu/morrison/human-genome-diversity-project/). The authors called for the collection of "material to record human ethnic and geographic diversity," particularly from populations "that have been isolated for some time [and] are likely to be linguistically and culturally distinct." The planned scale was large: ~25 individuals each from around 500 populations: 12,500 in all. However, the project failed to raise major funding, instead attracting criticism, including from several of the indigenous peoples it aimed to involve. Although it did not achieve its aims, it did establish a panel of 1064 cell

Box 10.2: The HGDP-CEPH samples

The Human Genome Diversity Project (HGDP) did not achieve all of its aims (see text), but has left a legacy that has transformed the fields of human genetic variation and human evolutionary genetics: a panel of 1064 diverse DNA samples derived from **lymphoblastoid cell lines**, representing 51 populations.[3] The cells are held by the Centre d'Etude du Polymorphisme Humain (CEPH) in Paris, hence the name HGDP-CEPH, and DNA samples are distributed on a cost-recovery basis. The availability of cell lines makes the amount of DNA available inexhaustible. Before this panel was available, individual labs had to go to enormous effort to collect or, for the vast majority, assemble by collaboration, a smaller and less representative set of samples. Now, any lab can genotype their variant(s) of interest in this collection, and also benefit from the large amount of data already available on this common resource, resulting from the work of over 100 investigators (http://www.cephb.fr/en/hgdp/diversity.php/).

The samples (Figure 10.3) include representation of indigenous populations from Africa, Europe, Asia, and the Americas. There are, inevitably, limitations and drawbacks. Several important areas of the world are missing, such as Australia, North America, India, and most of Northern Asia. The cell lines themselves cannot be distributed, so cellular phenotypes cannot be investigated; in addition, DNA cannot be sent to commercial companies. There is no information about the phenotypes of the donors except the sex of the individual, and population and geographic origin, and donors cannot be re-contacted. The informed consent provided by the donors was appropriate for the twentieth century when the panel was established, but may not meet all the criteria expected in the twenty-first century.

Nevertheless, the importance of this collection has been immense. It provides the only shared resource for investigating indigenous human diversity, and genomewide short tandem repeat (STR)[24] and SNP[15] genotypes are available both for further analysis and as a reference panel for comparison with new data. Our understanding of the distribution of diversity,[15, 24] and developments such as the serial founder model[20, 21] for the spread of humans, derive directly from it. Human evolutionary geneticists owe a great debt to the large numbers of unnamed donors and sample collectors, as well as to the scientists who established this panel.

lines[3] that has become a standard resource in the field and provided the basis for many of the studies we will describe in later chapters (**Box 10.2**). Why did the HGDP fail to attract more widespread support and funding? The lack of direct benefits to the participating communities, the perceived risks of **biopiracy**, and the proposal to establish immortal cell lines from populations on the brink of extinction were all factors. The contrast with another project proceeding at the same time, the human genome sequencing project, may also be informative. Despite initial doubts about its implications for human genetics and the scientific community, the public sequencing project's policy of making data freely and immediately available to all, instead of just to the labs generating the data, seems to have been a key ingredient in its success. The medical relevance of the sequencing project was obviously another important factor, and its concentration in scientifically advanced countries simplified its organization, but perhaps a more "open" diversity project would have been more successful.

Subsequent human variation projects provide some contrasts to the HGDP, and indeed to one another (**Table 10.1**, **Figure 10.3**). The HapMap Project (Box 3.6; http://hapmap.ncbi.nlm.nih.gov/), initiated a decade after the HGDP, established an exemplary consent process including extensive community involvement and feedback, and has escaped most of the criticisms aimed at the HGDP. Its choice of mainly urban populations and explicit goal of benefiting human health have contributed to this outcome. Many of its accomplishments lie outside the scope of this book, but we will see how it has become central to our understanding of many aspects of human diversity and evolution. The 1000 Genomes Project (Box 3.2; http://www.1000genomes.org/) can be seen as its successor, beginning in 2008 as new sequencing technologies became available, but involving many of the same scientists and initially using HapMap samples. Its stated main objective was to develop a public resource of genetic variation to support the next generation of medical association studies, specifically by finding all accessible variants at a frequency of ≥1% across the genome and down to 0.1–0.5% in gene regions. However, secondary objectives included

TABLE 10.1:
LARGE-SCALE PROJECTS PROVIDING INFORMATION ON HUMAN GENETIC VARIATION

Project	Launch date	Primary aims	Sample size	Populations	Genetic analyses	Cell lines	Data release	Website (Key reference) Further information
HGDP	1991	collection of isolated population samples	1064	51, worldwide	chosen by investigator	yes	on publication	http://www.cephb.fr/en/hgdp/diversity.php/ [Cann HM et al. (2002) *Science* 296, 261.] Box 10.2
HapMap	2002	haplotype map for medical genetics	1184	11, Africa, Europe, South and East Asia	SNP genotyping	yes	full public release	http://hapmap.ncbi.nlm.nih.gov/index.html.en [The International HapMap Project (2003) *Nature* 426, 789.] Box 3.6
Genographic	2005	elucidate migration history	~500,000	many, worldwide; including public participation	Y-SNP and Y-STR genotyping, mtDNA HVSI sequencing	no	on publication	https://genographic.nationalgeographic.com/genographic/lan/en/index.html [Wells, Deep Ancestry: Inside the Genographic Project (2006) National Geographic Books]
1000 Genomes	2008	discover variants at ≥1% frequency for medical genetics	2500	27, Africa, Europe, South Asia, East Asia, Americas	whole-genome sequencing	yes	full public release	http://www.1000genomes.org/ [The 1000 Genomes Project Consortium (2010) *Nature* 467, 1061.] Box 3.2

investigating questions of evolutionary interest. The Genographic Project (https://genographic.nationalgeographic.com/genographic/index.html) illustrates an alternative way of exploring human diversity. Funded by National Geographic, IBM, and the Waite Family Foundation, as well as by public participants who each contributed $99, it focused on ancestry information from mtDNA and the Y chromosome (**Appendix**) and explicitly avoided studying variants of medical significance, or establishing cell lines. With this model, it achieved a sample size more than 100-fold greater than the other projects discussed here, and many of these samples can potentially be studied by further genotyping or sequencing for those participants who have consented appropriately. Along these lines, a Geno 2.0 project has begun (https://genographic.nationalgeographic.com/about/), which is genotyping large numbers of ancestry-informative, medically uninformative, SNPs.

Medical-genetic studies involving genomewide genotyping often involve >100,000 participants (Chapter 17) and so provide additional large datasets that can potentially be used for evolutionary studies as well, although ethical and bureaucratic factors usually limit the availability of datasets, even when these have been published in the scientific literature. In addition, the choice of populations sampled will restrict their relevance to many evolutionary questions.

Individuals may choose to make their own genotypes and genome sequences freely available, and as costs decrease, these are likely to become an increasingly important source of data (**Section 18.4**). An early project in this area has been the Personal Genome Project (PGP), founded by its self-described "guinea pig #1," George Church, and with >1800 participants by 2012[1]

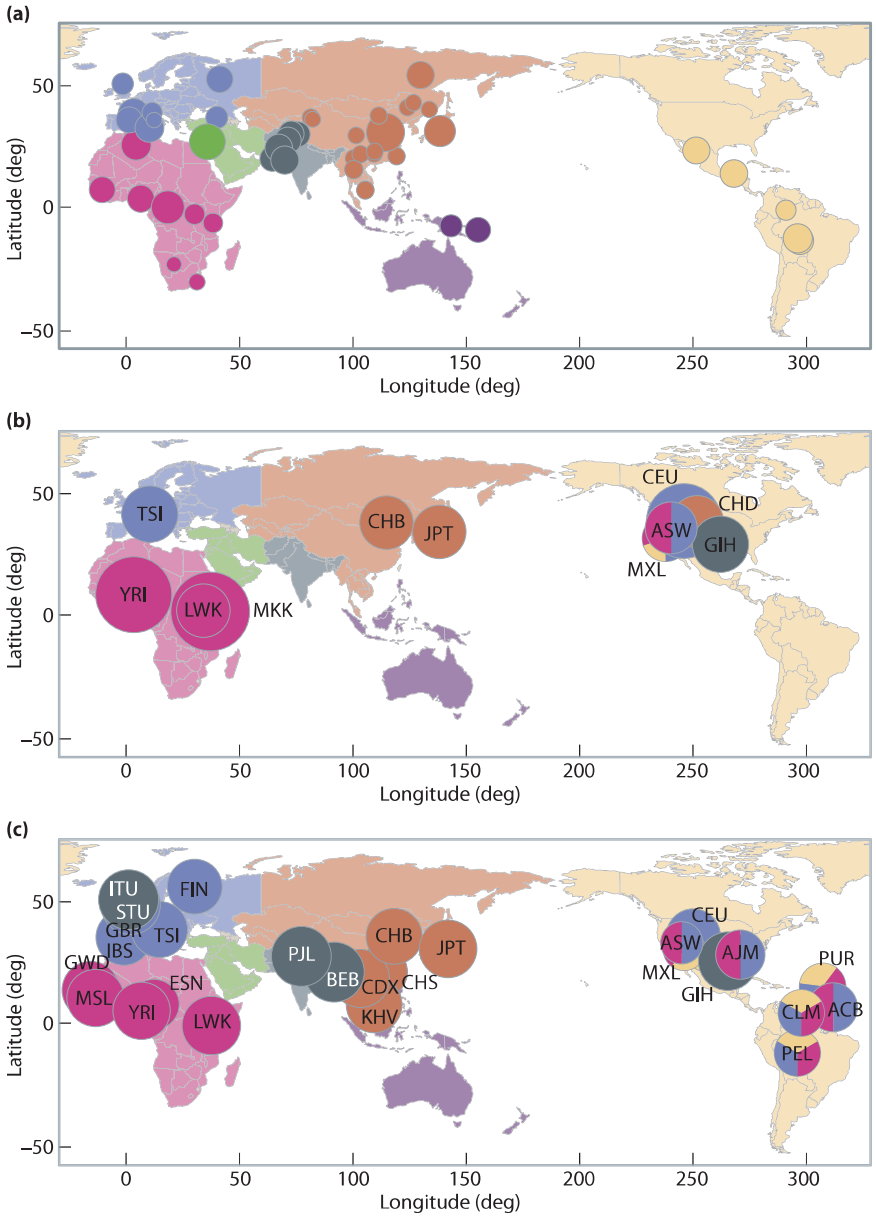

Figure 10.3: Samples included in three large-scale sample collections. (a) HGDP, (b) HapMap, (c) 1000 Genomes. Samples are represented by circles, with an area proportional to sample size (smaller circles in the second panel, for example TSI = 100), placed in the sampling location. Broad regions of the world are assigned colors, and these colors are used to indicate the geographical ancestry of the samples. While sampling location and ancestry often coincide, there are several discrepancies due to sampling migrant and mixed populations in the HapMap and 1000 Genomes collections. Note also the poor representation of indigenous populations from many regions of the world, such as North America or Australia. [Adapted from Colonna V et al. (2011) *Genome Biol.* 12, 234. With permission from BioMed Central Ltd.]

(http://www.personalgenomes.org/). Volunteers share their genotypes and genome sequence, and, crucially, other personal information such as health and medical data with the scientific community and the general public. They also donate tissue specimens which may be used in many ways, including for studies of gene expression and transformation into somatic cell-derived stem cells (that is, **induced pluripotent stem cells** or **iPS cells**). Thus far, to our knowledge, no harm to an individual from participating in genetic research of this kind has yet been documented.

What is a population?

The phrase "human population" is used widely, including in this book, but we now need to examine more carefully what is meant by the term "a population" (see also **Section 5.3**). When sets of individuals are phenotypically distinct and seldom interbreed, it is easy to distinguish populations: for example eastern mountain and eastern lowland gorillas (**Section 7.4**). However, although humans from anywhere in the world are potentially capable of interbreeding, they clearly do not form one worldwide randomly mating (**panmictic**) population, but exhibit structure. What criteria can be used to identify this structure?

Among the ways in which we can decide whether or not people belong to the same population are:

- Geographical proximity: individuals from the same population must be able to meet
- A common language: they must be able to communicate with each other
- Shared ethnicity, culture, religion: they are more likely to intermarry if they share history and values

None of these criteria is an absolute: we do not say that someone belongs to a different population if they move to a different country, if grandparents and grandchildren speak different languages, or if someone converts to a new religion. Nevertheless, after several generations, such changes could lead to the establishment of a new population. If these criteria were used alone, each individual might be considered to belong to many populations (geographical, linguistic, and so forth) defined in different ways, and these might change with time. The extent to which such memberships are correlated between individuals is unclear. If they were highly correlated, we could meaningfully identify distinct populations that would summarize the relationships between individuals, but if they were poorly correlated, it would be difficult to identify populations, reflecting the fact that, as we have seen earlier, human groups are not discrete, but are social constructs. The criteria used in the studies reported in this book are not always consistent. One common practice is to use self-determination: a person is a member of the group they identify with. The Czech writer Karl Deutsch described a nation as "a group of people united by a mistaken view of the past and a hatred of their neighbors"; although cynical, this encapsulates the ambiguity in defining "a population."

How many people should be analyzed?

A single individual can sometimes provide key evolutionary insights. We have seen the importance of the Neanderthal mtDNA and genome sequences, and the Denisovan genome (**Section 9.5**), and will encounter more examples in the next chapters, including Aboriginal Australian (**Section 11.4**) and Paleo-Eskimo genomes (**Section 13.2**). But these insights were possible because these sequences could be compared with many others: the analyses were not based entirely on a single genome. Another approach was based on the insight that a single individual's genome actually contains a "population" of regions with independent evolutionary histories because of recombination; these maternal and paternal copies of regions **coalesce** at different times and the distribution of these coalescence times provides insights into past demography[14] (**Section 9.5, Figure 9.23**).

Nevertheless, for most studies in population and evolutionary genetics, we need to analyze multiple individuals, and different projects have used very different sample sizes (Table 10.1). We thus need to consider questions such as, How many individuals need to be examined in order to address a particular question? How should these be distributed among different populations? There are no simple answers and in future sections we will see a range of strategies employed.

A related issue is what kind of weighting scheme should be used in choosing numbers of individuals and populations. It is generally agreed that, except in forensic investigations, weighting according to current population size is a poor option because it biases strongly toward recent expansions and these are not usually the main focus of interest in evolutionary studies. A geographical scheme is probably the most common, and aims to sample roughly equally from each geographical area; the projects listed in Table 10.1 do approximately this. A third possibility is to use linguistic criteria, on the grounds that the major linguistic divisions pre-date recent demographic expansions and thus sample according to the population structure that existed millennia ago (although such a scheme would complicate investigations of correlations between genetics

and linguistics). For example, a study of Xq diversity[10] examined 69 individuals distributed among 16 out of 17 major language phyla.

In practice, sample availability is often the major criterion. Throughout this book, we will see that sample sizes of 20–50 per population are common in evolutionary DNA analyses, and 100, used in the 1000 Genomes Project, is quite large. Studies often examine a few hundred individuals in total. Advances in technology (Chapter 4) should allow much larger studies to be undertaken in the near future, and current medically motivated sequencing studies such as those in the UK (UK10K, http://www.uk10k.org/), The Netherlands (GoNL, http://www.nlgenome.nl/), and Sardinia (SardiNIA, http://genome.sph.umich.edu/wiki/SardiNIA) provide a foretaste of the datasets that may yield evolutionary insights as a by-product. Overall, the influence of sampling strategy on conclusions from genetic studies remains an under-appreciated and under-studied aspect of the field.

10.2 APPORTIONMENT OF HUMAN DIVERSITY

How distinct, genetically, are different populations? When information on the variation of a set of **classical polymorphisms** (Box 3.1) became available from different populations, it was possible to investigate how diversity was apportioned between human populations or groups of populations, and a pioneering study was presented by Lewontin in 1972.[13] Despite limitations in technology and dated terminology, this work identified the basic view that is still current, more than 40 years later, and so we present it in some detail.

The apportionment of diversity shows that most variation is found within populations

Lewontin[13] used 17 loci (blood groups, serum proteins, and red blood cell enzymes) for which variation had been detected by immunological or electrophoretic methods (see Box 3.1), and had allele frequency data available for several populations. The populations were classified into seven "races" termed **Caucasians**, Black Africans, Mongoloids, South Asian Aborigines, Amerinds, Oceanians, and Australian Aborigines, based on morphological, linguistic, historical, and cultural criteria. Diversity for each locus was measured by H, the Shannon information measure:

$$H = -\sum_{i=1}^{n} p_i \ln 2p_i$$

where p_i is the frequency of the ith allele; H is a somewhat similar measure to Nei's gene diversity (**Section 6.2**) and in this study ranged from 0 (no variation) to 1.9 (high variation). It was calculated at three levels for each locus:

- H_{pop}, the value for each individual population averaged over populations within a "race"
- H_{race}, the value for each "race," calculated from the average gene frequency over all populations within that "race," averaged over all "races"
- $H_{species}$, calculated from the frequency averaged over all populations within the species

These values were then used to apportion the diversity:

- Within populations = $H_{pop}/H_{species}$
- Between populations within races = $(H_{race} - H_{pop})/H_{species}$
- Between races = $(H_{species} - H_{race})/H_{species}$

The proportion of variation within populations ranged from 63.6% for the **Duffy blood group** (**Section 17.4**) to 99.7% for Xm, although only four populations were available for the latter marker; the mean proportion within populations was 85.4%. On average, 8.3% (range, 2.1–21.4%) corresponded to differences

between populations within "races," and only 6.3% (range, 0.2–25.9%) was found between "races."

The overwhelming conclusion was that most variation lies within populations, and that "races" had no genetic reality, a conclusion reinforced by many subsequent analyses using independent population samples and DNA polymorphisms. Lewontin concluded:

> Human racial classification is of no social value and is positively destructive of social and human relations. Since such racial classification is now seen to be of virtually no genetic or taxonomic significance either, no justification can be offered for its continuance.

The apportionment of diversity can differ between segments of the genome

Subsequent studies differ in that they used DNA polymorphisms of several kinds, often analyzed data by F_{ST} or **AMOVA** (Box 5.2; **Section 6.3**), and referred to "continental groups" rather than "races," but the conclusions are strikingly similar: ~83–88% of autosomal variation is found within populations and ~9–13% between continental groups (**Table 10.2**). It is important to realize that such values depend on the frequency of the polymorphisms, and would be even lower if rarer variants were used.

Results from mtDNA and the Y chromosome are somewhat different, with less of the variation within populations and more between groups, as might be

TABLE 10.2:
EARLY STUDIES OF THE APPORTIONMENT OF HUMAN DIVERSITY

Locus	Variation (%)			Reference[a]
	Within population	Between populations within groups	Between groups	
Autosomal				
17 classical polymorphisms	85.4	8.3	6.3	Lewontin RC (1972) *Evol. Biol.* 6, 381.
30 microsatellites	84.5	5.5	10.0	Barbujani G et al. (1997) *Proc. Natl Acad. Sci. USA* 94, 4516.
79 RFLPs	84.5	3.9	11.7	Barbujani G et al. (1997) *Proc. Natl Acad. Sci. USA* 94, 4516.
60 microsatellites	87.9	1.7	10.4	Jorde L et al. (2000) *Am. J. Hum. Genet.* 66, 979.
30 SNPs	85.5	1.3	13.2	Jorde L et al. (2000) *Am. J. Hum. Genet.* 66, 979.
21 *Alu* insertions	82.9	8.2	8.9	Romualdi C et al. (2002) *Genome Res.* 12, 602.
mtDNA				
RFLPs	75.4	3.5	21.1	Excoffier L et al. (1992) *Genetics* 131, 479.
RFLPs	81.4	6.1	12.5	Seielstad M et al. (1998) *Nat. Genet.* 20, 278.
HVSI	72.0	6.0	22.0	Jorde L et al. (2000) *Am. J. Hum. Genet.* 66, 979.
Y chromosome				
22 binary polymorphisms	35.5	11.8	52.7	Seielstad M et al. (1998) *Nat. Genet.* 20, 278.
30 polymorphisms, several types	59	25	16	Santos F et al. (1999) *Am. J. Hum. Genet.* 64, 619.
6 microsatellites	83.3	18.5	–1.8	Jorde L et al. (2000) *Am. J. Hum. Genet.* 66, 979.
14 binary polymorphisms	42.5	17.4	40.1	Romualdi C et al. (2002) *Genome Res.* 12, 602.

[a] Reference for apportionment analysis, which may use data first published elsewhere.

expected from their smaller effective population sizes and hence greater drift (**Appendix**). The latter is particularly marked for the Y chromosome, which may be partly explained by **patrilocality** (**Section 5.5**) although there were large differences between studies. One detected no variation between groups, probably because the polymorphisms used were a small set of rapidly mutating microsatellites, while another actually found that most of the variation (53%) was between groups (Table 10.2).

A comparison of the distribution of diversity in humans with that found in other species of large-bodied mammals provides a useful perspective.[27] Humans, despite their worldwide distribution, show a low F_{ST} value comparable to that seen in either waterbuck or impala, each from a limited geographical region, Kenya; this reflects our recent expansion out of Africa starting from a much smaller and more geographically restricted population (Chapter 9). Species with long-established wider ranges, such as coyotes from North America or gray wolves from Eurasia, have higher values (**Figure 10.4**).

Patterns of diversity generally change gradually from place to place

Having established that there *are* small but nonzero genetic differences between populations, we can ask what patterns are found, and how we can explain their origins and maintenance. The most important observation is that variant frequencies as a rule change gradually from place to place. Sharp changes in frequency over small distances are unusual, and generally have a special explanation, such as a barrier to migration. For example, populations separated by the Strait of Gibraltar [just 14.3 km (8.9 miles) of sea between Spain in Europe and Morocco in Africa] differ substantially in their frequencies of autosomal, mtDNA, and Y loci. This is because these populations lie at the western end of the Mediterranean Sea and both originated from the east, following parallel expansion paths through Europe to the north of the Mediterranean and Africa to the south, respectively, with little gene flow across the sea. They accumulated gradual but independent east-to-west differences, which are therefore maximal at the western extremity: Gibraltar.

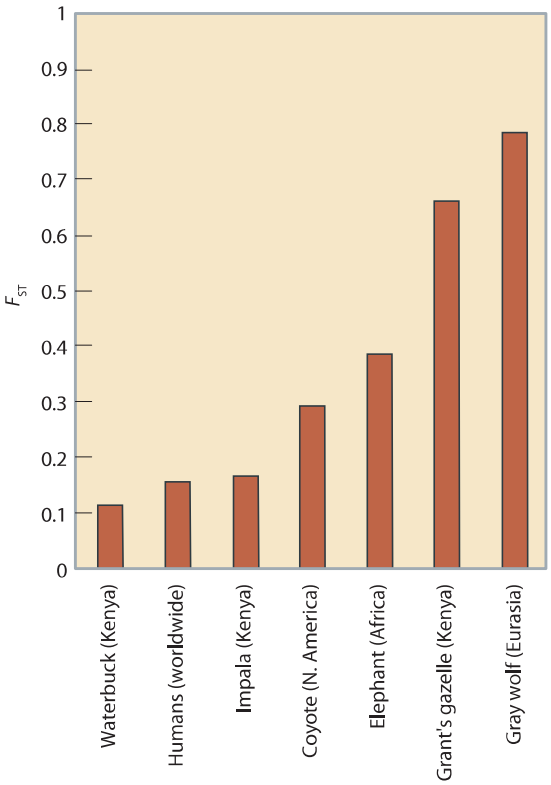

Figure 10.4: F_{ST} values for humans and other large-bodied mammals. Note the low human F_{ST}, typical of species with a restricted geographical distribution, despite the wide distribution of humans. [Data from Templeton AR (1999) *Am. J. Anthropol.* 100, 632.]

In **Section 9.4** and Figure 9.19, we saw that broad global patterns of diversity, highest in Africa and decreasing with distance from Africa, could be explained by the **serial founder model**. This model accounts for the origin of the gradual, **clinal** patterns of variation found throughout the world. An **isolation by distance model**, in which migrants tend to move short distances (**Section 6.8**), but large numbers of people seldom move long distances, explains how such patterns could persist for tens of thousands of years.

Thus we can interpret the pattern of human diversity as clinal. However, in the next section and other chapters, we will see that clustering methods, which assume that humans belong to discrete groups, also provide useful insights into human diversity, and are widely used.[24] So is human genetic variation better described as clinal or clustered?[26] One way of thinking about this question is to consider how the sampling scheme can influence the conclusion. Evenly spaced sampling from a continuous distribution produces a clinal pattern, while sampling groups of populations from locations separated by gaps, from the same distribution, produces clusters (**Figure 10.5**). Is this a good model for the distribution of human variation? Once sampling has been taken into account, some striking discontinuities can still be recognized in human genetic data, often associated with geographical barriers such as the Himalaya Mountains or the Strait of Gibraltar mentioned above, or social conventions such as those associated with the **caste system** in India. A follow-up study re-investigating worldwide patterns of variation concluded that while patterns within a continent might be largely clinal, there were robust clusters corresponding to different continents, arising from small discontinuous jumps in genetic distance for most population pairs on opposite sides of geographic barriers.[23]

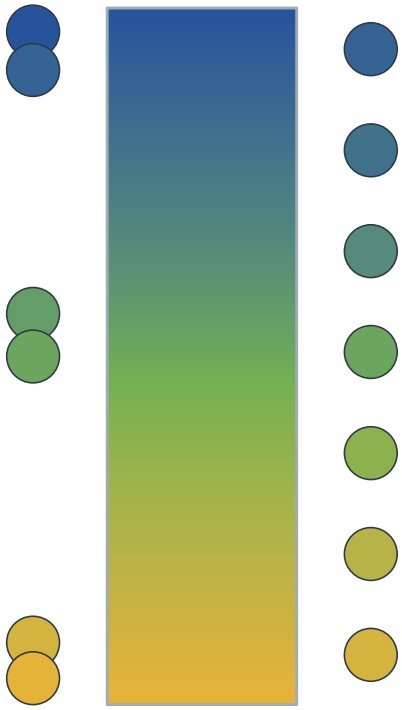

Figure 10.5: Clusters or clines of human diversity?
The way in which samples are chosen from a continuous distribution (*central rectangle*) determines whether discrete clusters are seen (*circles at left*) or a more clinal pattern (*circles at right*).

The origin of an individual can be determined surprisingly precisely from their genotype

The small proportion of variation that lies between populations or continents can still be informative about their origin. A question is therefore how precise an origin can be specified. With genomewide datasets of hundreds or thousands of microsatellites or SNPs, analyzed using methods that summarize the information from multiple polymorphisms, or model ancestry using **STRUCTURE-like methods** (Table 6.1), the answer is that surprising precision can be obtained.

As we saw in the previous chapter, migration distance from East Africa explains ~85% of the genomewide heterozygosity level observed in an individual, meaning that heterozygosity alone can predict distance from East Africa with ~85% accuracy. In **Figure 10.6** and Figure 18.8, we see that even more detailed information can be obtained. PC plots based on genomewide SNP analyses show that world populations from the HGDP-CEPH collection cluster according to continent[15] (Figure 10.6a), European populations cluster approximately according to country[17] (Figure 18.8), and most remarkably of all, individuals from three Scottish Isles, or three Italian valleys, cluster according to the individual isle or valley[18] (Figure 10.6b). Thus these genomes contain detailed information about geographical ancestry, a finding with important implications for personal genomics and **forensic genetics**, discussed further in Chapter 18.

How can the finding that it is possible, with a sufficiently large set of polymorphisms, to deduce so much about the population of origin of an individual be reconciled with the earlier conclusion that most variation exists within populations, not between them? Particular alleles are usually not continent- or population-specific, although specific searches for these have had some success (**Section 14.2**, **Section 18.2**, and below); extraction of the geographical information generally requires analysis of large numbers of loci. If it is possible to identify genetic groups within humans, are these groups then "races"? No, because, as we have seen, the groups identified by current genetic techniques do not correspond to traditional races, and the differences between them are too small to justify being called races, which would require more than 25–30%

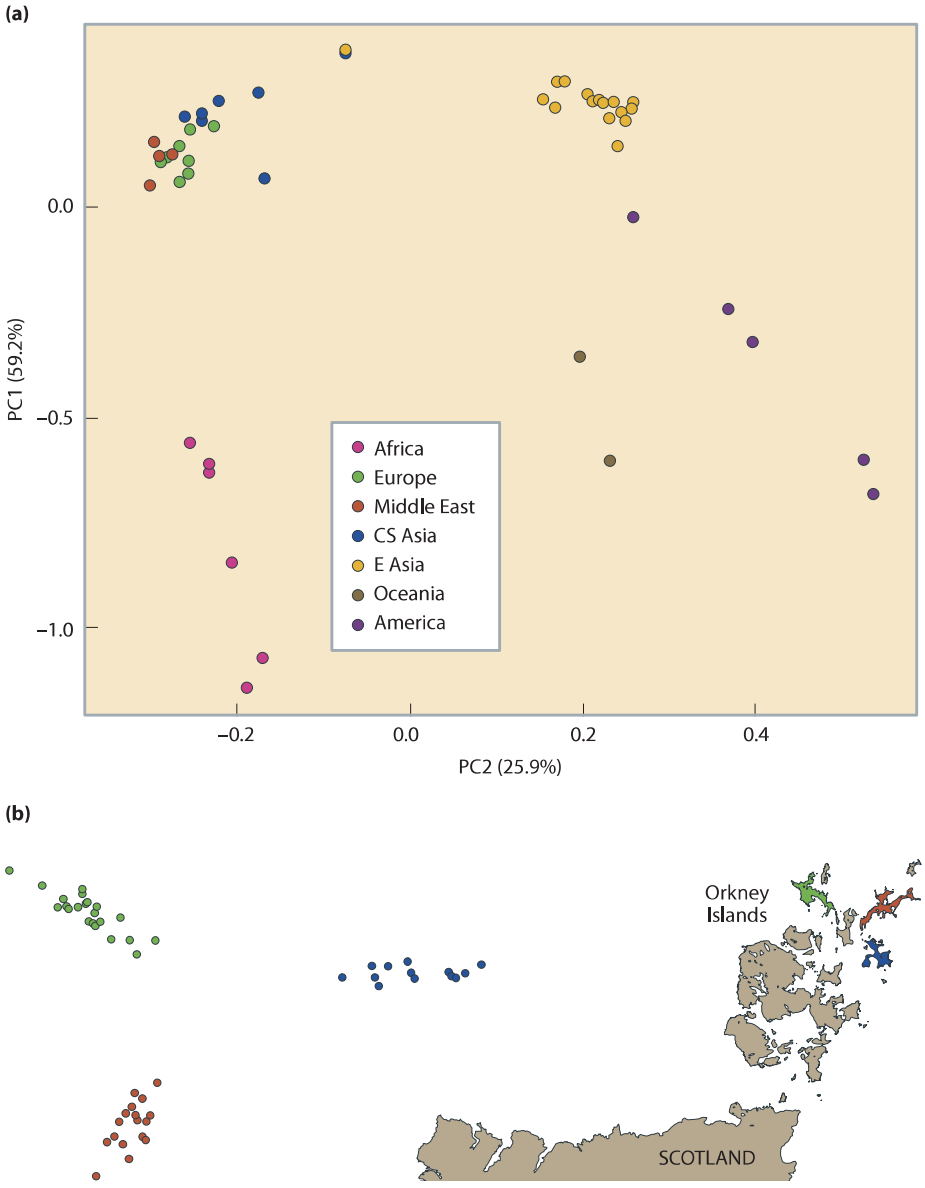

Figure 10.6: Prediction of geographical ancestry from genomewide SNP data. Each panel shows a PC analysis based on genomewide SNP genotypes. (a) Worldwide samples from the HGDP panel. Each dot represents a population and is colored according to the region of origin. (b) Samples from three Orkney Isles off the coast of Scotland (Sanday, *red*; Stronsay, *blue*; and Westray, *green*). [a, from Li JZ et al. (2008) *Science* 319, 1100. With permission from AAAS. b, from O'Dushlaine C et al. (2010) *Eur. J. Hum. Genet.* 18, 1269. With permission from Macmillan Publishers Ltd.]

genomewide difference between groups.[28] Irrespective of the genetic data, any idea that individuals should be treated by society according to their perceived ancestry or "race" is a social construct, not a biological one.

The distribution of rare variants differs from that of common variants

It may seem surprising to find a separate section focusing specifically on rare variants, but rare variants differ in some important ways from common ones. Common variants are invariably old, because it takes time for alleles to increase in frequency, even in the occasional cases where they are selectively advantageous (**Sections 6.6 and 6.7**). In contrast, rare variants *can* be old, but the vast majority of them are young. This has two major consequences: first, there has been less time for purifying selection to act on them, so they are enriched for deleterious variants (including medically relevant ones: Chapters 16 and 17) compared with common variants; second, there has been less time for them to spread geographically, so they tend to be restricted to a single population or a group of nearby populations. This is illustrated strikingly by the patterns of sharing of the variants that were found just twice in the 1092 individuals from the

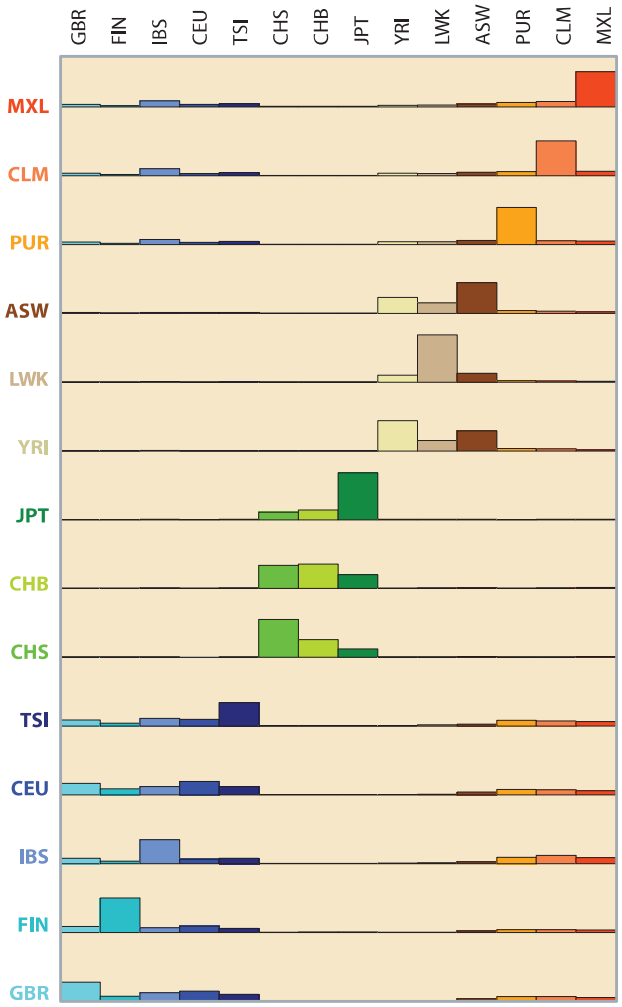

Figure 10.7: Distribution of rare variants within and between human populations.
Each row shows the sharing of variants detected twice in the 2184 chromosomes (1092 individuals) in the 1000 Genomes Project Phase 1 study between the population designated on the left, and all the other populations in the study (*top*). Note both the tendency for these rare variants to be shared within a population or with nearby populations, and the wider sharing in the American populations because of their admixture within historical times. [From The 1000 Genomes Project Consortium (2012) *Nature* 491, 56. With permission from Macmillan Publishers Ltd.]

14 populations included in Phase 1 of the 1000 Genomes Project[30] (**Figure 10.7**). These are the rarest shared variants in the study (with a frequency of <0.1%), and so should show any unusual pattern in its most extreme form. In every population, the second copy is most likely to be found within the same population. When it is found in a different population, the second population usually comes from the same continent. The main exception to this pattern is seen in the American populations, which reveal widespread sharing with both African and European populations, reflecting their history of admixture within historical times (Chapter 14). In contrast, common variants would generally be shared between all populations.

10.3 THE INFLUENCE OF SELECTION ON THE APPORTIONMENT OF DIVERSITY

In the previous sections of this chapter, we have considered studies conducted using variants that were often chosen for historical reasons (data were available when the study was carried out), or because they were considered neutral. We saw in **Section 10.2** that, although the mean between-population component of diversity in an early study was 14.6%, there was variation among the loci used: it ranged from as little as 0.3% for Xm to as much as 36.4% for the Duffy blood group. How much variation in these values is generally found? What effects does selection have? (see also **Section 6.7**). In the current section, we will encounter two measures of population differentiation: F_{ST} and ΔDAF (**Sections 6.3 and 6.7**). Both vary between 0 and 1, with 0 corresponding to no differentiation between populations, and 1 corresponding to complete differentiation; they are highly correlated.

The distribution of levels of differentiation has been studied empirically

Under neutral conditions, variation in allele frequencies between populations is determined by drift, which affects all loci in the same population equally, so all are expected to show the same F_{ST} value, although there will be a spread around this value because of stochastic factors. The observed distribution of F_{ST} values from the 1000 Genomes Project[30] is shown in **Figure 10.8**. The mean value for 35.6 million SNPs ascertained by sequencing 1092 individuals from Africa, Europe, East Asia, and the Americas was 0.05. The distribution, however, is highly asymmetric and has a long tail toward higher values, the maximum being over 0.8.

Low differentiation can result from balancing selection

Balancing selection, which can act through **heterozygote advantage** or **frequency-dependent selection**, for example (**Section 6.7**), favors the maintenance of two or more alleles in the population and thus high diversity, but if the same alleles are favored in all populations, F_{ST} will be low. Among the classical polymorphisms, some **HLA** alleles show low F_{ST} (**Table 10.3**) despite the very high overall diversity at this locus (see Box 5.3 for an introduction to the HLA locus).

With the benefit of whole-genome sequences, it is possible to identify genes with low F_{ST} in a comprehensive way, at least in the limited populations for which the sequence information is currently available. Since F_{ST} depends on allele frequency, low values are inevitable for rare alleles, and are only of interest for common alleles, and are most unexpected (and thus most interesting) for alleles with a derived allele frequency around 50%. Table 10.3 lists the 40–60% frequency SNPs within protein-coding genes that were identified by the 1000 Genomes Pilot Project[29] as showing the lowest differentiation in comparisons of African, European, and East Asian samples.

The HLA variant rs1129740 (Table 10.3) is found in DQA1*01 alleles (DQA1*01:01 to DQA1*01:07) and one of these (DQA1*01:03) is associated with both protection against some conditions such as chronic hepatitis C,[8] and susceptibility to others such as leprosy[22] and AIDS.[12] While there is no evidence that these diseases themselves have been the selective forces in human evolutionary

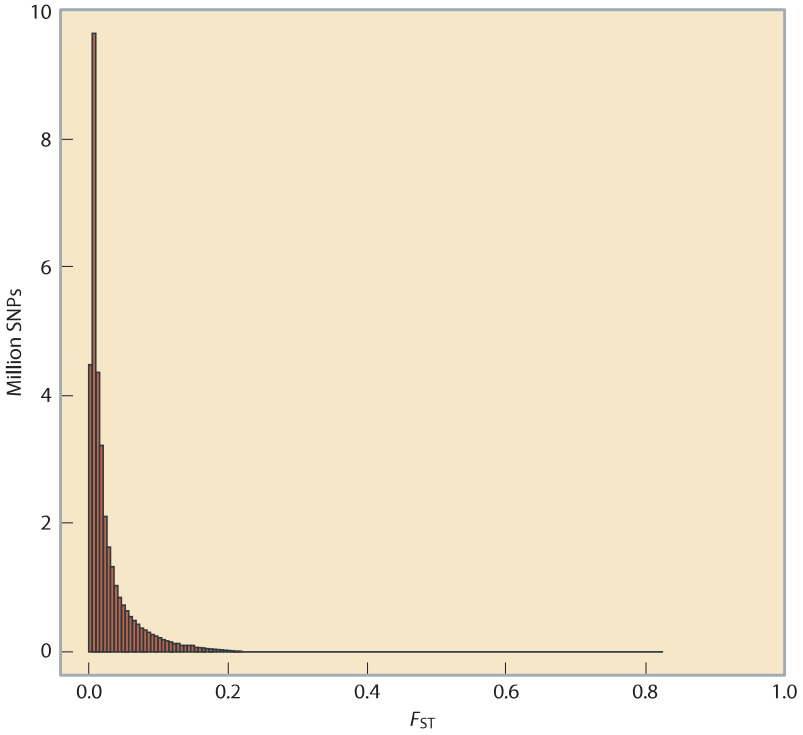

Figure 10.8: Distribution of human F_{ST} values.
Worldwide F_{ST} was calculated for 36.5 million SNPs discovered by Phase 1 of the 1000 Genomes Project. Note the low average value and skewed distribution with a few high values. (Unpublished analysis provided by Vincenza Colonna.)

TABLE 10.3:
THE LEAST-DIFFERENTIATED COMMON CODING SNPs IDENTIFIED BY THE 1000 GENOMES PILOT PROJECT

Chr	Position (GRCh37)	rs_id[a]	Gene	YRI_DAF[b]	CEU_DAF	CHB + JPT_DAF	DAF difference	Change to amino acid	Possible evolutionary consequence
11	32,874,926	rs11032025	*PRRG4*	0.40	0.40	0.40	0.000	synonymous	unknown
7	100,395,588	rs67377634	*MUC3A*	0.51	0.51	0.51	0.000	synonymous	unknown
14	68,042,574	rs11158685	*PLEKHH1*	0.46	0.47	0.46	0.007	nonsynonymous	unknown
20	18,445,963	rs6035051	*C20orf12*	0.47	0.47	0.46	0.007	synonymous	unknown
20	61,048,549	rs41305803	*GATA5*	0.42	0.41	0.42	0.007	synonymous	unknown
22	50,480,108	rs9628315	*TTLL8*	0.44	0.44	0.45	0.007	nonsynonymous	unknown
16	70,954,774	rs1798529	*HYDIN*	0.57	0.55	0.55	0.013	nonsynonymous	unknown
19	1,052,005	rs3764652	*ABCA7*	0.41	0.41	0.39	0.013	synonymous	unknown
22	50,658,424	rs11703226	*TUBGCP6*	0.41	0.39	0.41	0.013	nonsynonymous	unknown
6	32,609,105	rs1129740	*HLA-DQA1*	0.51	0.51	0.49	0.013	nonsynonymous	defense against infection
11	5,373,114	rs5006888	*HBG2/ HBE1*	0.44	0.45	0.43	0.013	nonsynonymous	unknown
11	5,373,646	rs5024041	*OR51B6*	0.44	0.45	0.43	0.013	synonymous	unknown
7	135,406,176	rs4596594	*SLC13A4*	0.53	0.55	0.55	0.013	synonymous	unknown
21	15,481,365	rs7278737	*LIPI*	0.49	0.49	0.47	0.013	nonsynonymous	unknown
6	146,755,140	rs2942	*GRM1*	0.50	0.52	0.52	0.013	synonymous	unknown
20	44,238,741	rs2245898	*WFDC9*	0.58	0.58	0.60	0.013	nonsynonymous	unknown
3	183,558,402	rs3732581	*PARL*	0.49	0.50	0.48	0.013	nonsynonymous	heart condition?

[a] rs_id is a unique identifying number given to each variant
[b] DAF, derived allele frequency

history that maintained both alleles of rs1129740 in the population, the link to several disease-related phenotypes suggests the more general possibility of such a cause. Biological explanations for the even frequencies of the other variants listed are lacking, and many may just be due to chance sampling, showing more variation in other populations. However, the association of the rs3732581 C-allele in the *PARL* gene with both a protective effect in reducing artery wall thickness and a susceptibility effect in increasing the risk of coronary artery disease[19] hints at possible balancing selection here as well.

High differentiation can result from directional selection

In contrast to balancing selection, directional selection acting in a subset of populations will lead to different alleles being at high frequencies in different populations (**Section 6.7**). Diversity within any one population may be low, but worldwide F_{ST} or ΔDAF will be high. The availability of whole-genome sequences from the 1000 Genomes Pilot Project[29] allows us to identify the regions of the genome that show the largest differences between populations. **Table 10.4** lists the five SNPs within protein-coding genes showing the largest ΔDAF values in pairwise analyses between the African (YRI), European (CEU), and East Asian (CHB+JPT) samples included in this project. We can draw

TABLE 10.4:
THE MOST HIGHLY DIFFERENTIATED CODING SNPs IDENTIFIED BY THE 1000 GENOMES PILOT PROJECT

Comparison	Chr	Position (GRCh37)	Frequency difference	Gene	Change to amino acid sequence	Possible evolutionary consequence
CEU vs CHB+JPT	15	46,213,776	0.992	*SLC24A5*	nonsynonymous	light skin color in Europeans
CEU vs CHB+JPT	5	33,987,450	0.968	*SLC45A2*	nonsynonymous	light skin color in Europeans
CEU vs CHB+JPT	4	38,475,043	0.939	*TLR1*	nonsynonymous	altered innate immunity (?)
CEU vs CHB+JPT	17	18,937,740	0.894	*AC007952.8*	nonsynonymous	unknown
CEU vs CHB+JPT	2	108,880,033	0.892	*EDAR*	nonsynonymous	hair and tooth morphology in East Asians
CEU vs YRI	15	46,213,776	0.979	*SLC24A5*	nonsynonymous	light skin color in Europeans
CEU vs YRI	5	33,987,450	0.979	*SLC45A2*	nonsynonymous	light skin color in Europeans
CEU vs YRI	8	145,610,489	0.972	*AF205589.2*	nonsynonymous	unknown
CEU vs YRI	9	126,302,623	0.966	*NR5A1*	nonsynonymous	sexual development, reproduction (?)
CEU vs YRI	4	38,475,043	0.937	*TLR1*	nonsynonymous	altered innate immunity (?)
CHB+JPT vs YRI	2	72,561,382	0.981	*EXOC6B*	synonymous	unknown
CHB+JPT vs YRI	15	39,936,798	0.978	*SPTBN5*	nonsynonymous	unknown
CHB+JPT vs YRI	17	60,251,208	0.976	*AC103810.1*	nonsynonymous	unknown
CHB+JPT vs YRI	8	145,610,489	0.972	*AF205589.2*	nonsynonymous	unknown
CHB+JPT vs YRI	16	46,815,699	0.968	*ABCC11*	nonsynonymous	dry earwax in East Asians

a number of conclusions from these findings. (1) These are the genes showing the most extreme geographical differentiation in our genomes, yet none of the frequency differences quite reach a value of one, which would correspond to complete fixation of one allele in one population and the other allele in the other population: even among these outliers, we cannot find fixed differences between populations. (2) Among the 11 genes identified, we know so little about the function of five that we cannot even speculate plausibly about the likely evolutionary implications (so these could be interesting avenues for future research). (3) For the other six genes, we can suggest either a general class of selective force (related to innate immunity, *TLR1*; or sexual development/reproduction, *NR5A1*), or more specific selection influencing earwax type (*ABCC11*), hair or tooth morphology (*EDAR*; Figure 15.2), or skin pigmentation (*SLC24A5*, *SLC45A2*). However, only for the last two genes can we be reasonably confident that the phenotype mentioned was the direct target of selection (see **Section 15.3**). (4) The skin pigmentation, hair/tooth, and earwax phenotypes were all noted and studied extensively before the era of genomic analyses, illustrating both the non-neutrality of some of the traits favored by anthropologists, and the effectiveness of these early scientists in identifying several of the most highly differentiated genes in humans.

Positive selection at EDAR

The *EDAR* (ectodysplasin A receptor) gene, carrying one of the most highly differentiated SNPs in the genome (Table 10.4), illustrates both the significant evolutionary insights that can be obtained by studying such unusual diversity patterns, and some of the complexities in interpretation that can arise. *EDAR* came to the attention of evolutionary geneticists via the HapMap1 Project, where it stood out because it carries a Val370Ala nonsynonymous variant

(rs3827760) with a low (zero) frequency of the derived allele in the African (YRI) and European (CEU) samples, but a very high frequency in the East Asian (CHB) sample. Subsequent studies confirmed the high level of geographical differentiation in a larger sample[31] (**Figure 10.9a**), and identified a strong signal of

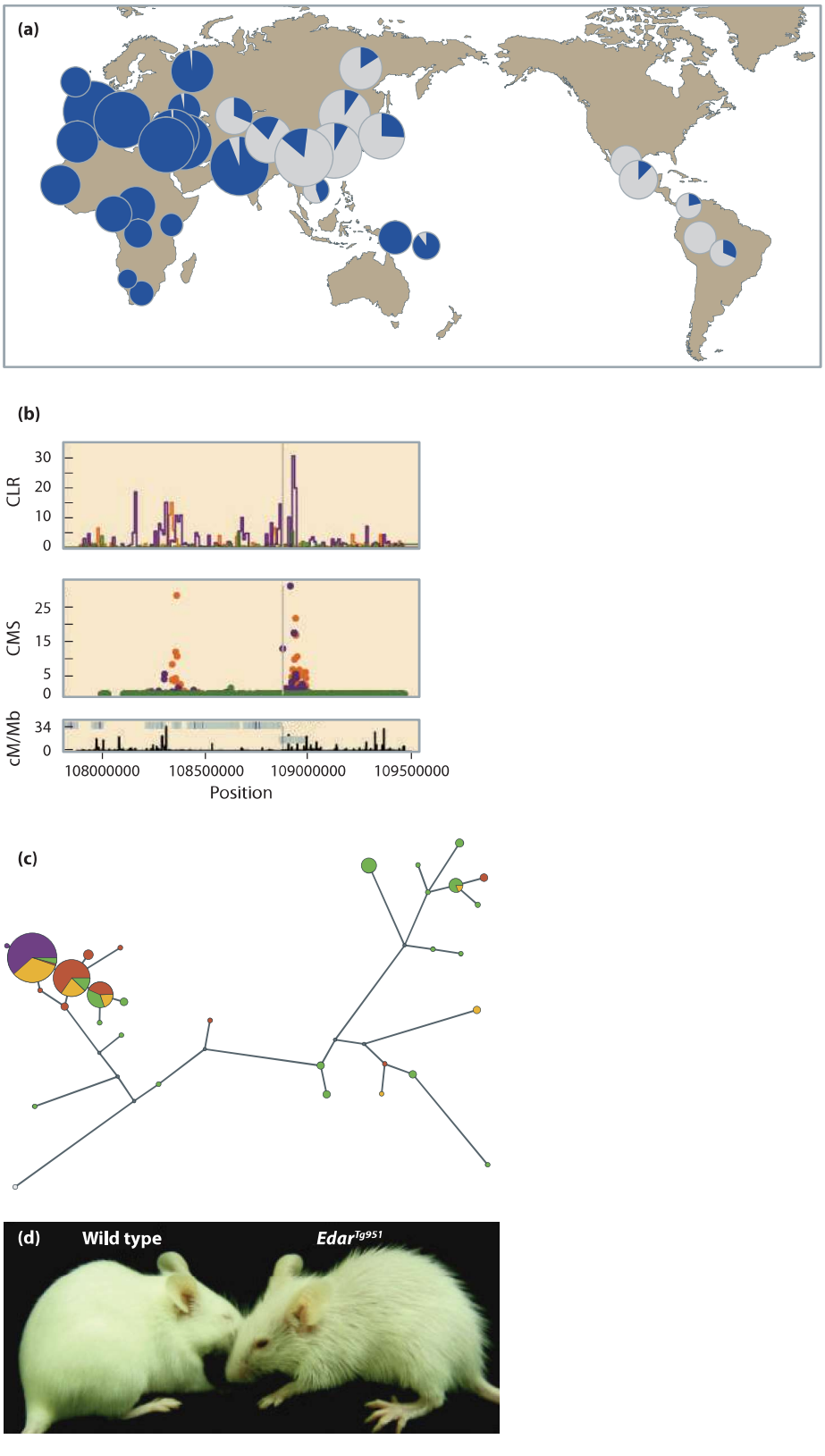

Figure 10.9: Positive selection at the *EDAR* gene.
(a) Geographical distribution of the ancestral (*blue*) and derived (*gray*) alleles of the Val370Ala variant of the *EDAR* gene. Note the 100% frequency of the ancestral (Val) allele in the west, contrasted with the high frequency of the derived (Ala) allele in the east. (b) Signals of positive selection in the vicinity of *EDAR*. CLR, composite likelihood ratio test for selection; CMS, composite of multiple signals; cM/Mb, centimorgans/megabase (recombination rate); *green*, selection in YRI (Africa, no signal); *orange*, selection in CEU (moderate signal, Europe); *purple*, selection in CHB+JPT (strong signal, East Asia). Note that the strongest signal is displaced from the Val370Ala substitution (*vertical gray line*). (c) Signals of positive selection in a 5-kb region surrounding the Val370Ala substitution. Selection is illustrated in these networks by the large circle size. The haplotype carrying the Ala-allele is selected in East Asians and Hispanics (*purple*, *yellow*) while a different haplotype carrying the Val-allele is selected in Europeans (*orange*). (d) Modeling the human thick hair phenotype in mice. In humans, the Ala allele is associated with thick hair, but in transgenic mice a similar phenotype was seen when the Val allele was overexpressed. [a and c, from Xue Y et al. (2009) *Genetics* 183, 1065. With permission from Genetics Society of America. b, from The 1000 Genomes Project Consortium (2010) *Nature* 467, 1061. With permission from Macmillan Publishers Ltd. d, from Mou C et al. (2008) *Hum. Mutat.* 29, 1405. With permission from John Wiley & Sons, Inc.]

positive selection (Figure 10.9b and c) using both haplotype-based tests (**Section 6.7**)[25] and allele frequency spectrum-based tests (**Section 6.7**).[31] Moreover, the gene had long been known to medical geneticists because its inactivation led to anomalies of skin, hair, teeth, and sweat glands, a combination known as hypohidrotic ectodermal dysplasia (OMIM 224900). It therefore seemed natural to suggest that the Val370Ala substitution might lead to a milder change in the same parts of the body, and indeed an association was found with both the thick hair characteristic of East Asians[6] and the shovel-shaped incisors (Figure 15.2) found in these and other populations.[11] It is therefore possible to propose models of positive selection based on these traits: perhaps the hair phenotype was considered attractive and selected sexually, for example, or a different pattern of sweating was advantageous. It is difficult to confirm any such model, and it remains possible that some trait that has not yet been identified was the true target of selection.

In addition, there are more substantial complexities:

- The strongest signal of positive selection in East Asians does not correspond to the Val370Ala substitution; instead it lies about 60 kb away within an intron of the same gene (Figure 10.9b). Modeling shows that a signal of selection *can* be located at such a distance from the target of selection, but the observation nevertheless raises the question of whether the Val370Ala substitution is really the main target of selection.
- Transgenic mice carrying multiple copies of mouse *Edar* (which has the ancestral Val-allele) show increased *Edar* expression and thick hair characteristic of human Ala-allele carriers[16] (Figure 10.9d). In humans, the Ala-allele is expressed at a higher level than the Val-allele, raising the possibility that the underlying cause of the human phenotype might be increased expression rather than the amino acid substitution.
- Europeans also show evidence of positive selection at *EDAR*,[31] with a signal in a similar location to East Asians (Figure 10.9b and c). The Val370Ala allele is absent from Europeans, so cannot be the target of selection, and both the target variant and selected phenotype in Europeans are entirely unknown.

Thus even for one of the best-studied and compelling examples of high population differentiation and positive selection, much remains to be understood.

SUMMARY

- Studies of human diversity raise important ethical questions which need to be addressed before each individual project is carried out.
- The sampling strategy will influence the conclusions, and strategies based on current population size and geographical or linguistic criteria have all been used in different studies.
- The definition of a "population" is not simple, but reflects a combination of geographical proximity, a common language, and shared ethnicity, culture, and religion.
- Most (at least 85%) autosomal variation is found within populations and less than 10% between different continents; more geographical differentiation is seen for mtDNA and Y-chromosomal sequences.
- With a large number of polymorphisms, considerable information about the population of origin of an individual can be obtained, sometimes down to the level of an individual country or even village.
- Selection influences the apportionment of diversity for a few loci: balancing selection can lead to low population differentiation levels (for example, *HLA* DQA1*01), while directional selection can lead to high levels (for example, *EDAR*, *SLC24A5*, *TLR1*).

QUESTIONS

Question 10–1: Professor Pangloss' reorganization of a lab freezer has led to the loss of labels on 150 tubes of DNA. However, on the box is written that it contains 50 DNA samples each from a Chinese, Nigerian and Irish population. Undaunted, and with his access to modern genotyping facilities and statistical methods, Pangloss says that he can assign each of the 150 samples to its population of origin in a short time, with moderate effort and expense. How?

Question 10–2: A population previously unknown to scientists has been contacted in the Amazon region. Would it be ethical to carry out genetic analyses on them? If not, why not? If so, of what kind and in what ways?

Question 10–3: It is proposed to sequence the genomes of the entire population of Erewhon and make the sequences publicly available. What benefits and risks would this pose for the Erewhonese and what advice would you offer to the Erewhon Genome Management Committee about how to maximize the benefits and minimize the risks?

Question 10–4: A company has developed an improved sequencing methodology and has offered to generate near-perfect sequences of 100,000 people. Who should be sequenced, why, with what accompanying details, and with what safeguards on the information?

Question 10–5: The geographical ancestry of a DNA sample can readily be deduced from its sequence. Does this prove that human races exist?

Question 10–6: What evidence suggests that the *EDAR* gene has experienced positive selection in humans? What additional analysis or experiment would you perform to identify the target(s) of selection and the selective force?

REFERENCES

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

1. **Ball MP, Thakuria JV, Zaranek AW et al.** (2012) A public resource facilitating clinical use of genomes. *Proc. Natl Acad. Sci. USA* **109**, 11920–11927.
2. **Barbujani G & Colonna V** (2010) Human genome diversity: frequently asked questions. *Trends Genet.* **26**, 285–295.
3. **Cann HM, de Toma C, Cazes L et al.** (2002) A human genome diversity cell line panel. *Science* **296**, 261–262.
4. **Caulfield T, Fullerton SM, Ali-Khan SE et al.** (2009) Race and ancestry in biomedical research: exploring the challenges. *Genome Med.* **1**, 8.
5. **Cavalli-Sforza LL, Wilson AC, Cantor CR et al.** (1991) Call for a worldwide survey of human genetic diversity: a vanishing opportunity for the Human Genome Project. *Genomics* **11**, 490–491.
6. **Fujimoto A, Kimura R, Ohashi J et al.** (2008) A scan for genetic determinants of human hair morphology: EDAR is associated with Asian hair thickness. *Hum. Mol. Genet.* **17**, 835–843.
7. **Gould SJ** (1981) The Mismeasure of Man. Penguin Books.
8. **Hohler T, Gerken G, Notghi A et al.** (1997) MHC class II genes influence the susceptibility to chronic active hepatitis C. *J. Hepatol.* **27**, 259–264.
9. **HUGO Pan-Asian SNP Consortium, Abdulla MA, Ahmed I et al.** (2009) Mapping human genetic diversity in Asia. *Science* **326**, 1541–1545.
10. **Kaessmann H, Heissig F, von Haeseler A & Pääbo S** (1999) DNA sequence variation in a non-coding region of low recombination on the human X chromosome. *Nat. Genet.* **22**, 78–81.
11. **Kimura R, Yamaguchi T, Takeda M et al.** (2009) A common variation in EDAR is a genetic determinant of shovel-shaped incisors. *Am. J. Hum. Genet.* **85**, 528–535.
12. **Kroner BL, Goedert JJ, Blattner WA et al.** (1995) Concordance of human leukocyte antigen haplotype-sharing, CD4 decline and AIDS in hemophilic siblings. Multicenter Hemophilia Cohort and Hemophilia Growth and Development Studies. *AIDS* **9**, 275–280.
13. **Lewontin RC** (1972) The apportionment of human diversity. *Evol. Biol.* **6**, 381–398.
14. **Li H & Durbin R** (2011) Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496.
15. **Li JZ, Absher DM, Tang H et al.** (2008) Worldwide human relationships inferred from genome-wide patterns of variation. *Science* **319**, 1100–1104.
16. **Mou C, Thomason HA, Willan PM et al.** (2008) Enhanced ectodysplasin-A receptor (EDAR) signaling alters multiple fiber characteristics to produce the East Asian hair form. *Hum. Mutat.* **29**, 1405–1411.
17. **Novembre J, Johnson T, Bryc K et al.** (2008) Genes mirror geography within Europe. *Nature* **456**, 98–101.
18. **O'Dushlaine C, McQuillan R, Weale ME et al.** (2010) Genes predict village of origin in rural Europe. *Eur. J. Hum. Genet.* **18**, 1269–1270.
19. **Powell BL, Wiltshire S, Arscott G et al.** (2008) Association of PARL rs3732581 genetic variant with insulin levels, metabolic syndrome and coronary artery disease. *Hum. Genet.* **124**, 263–270.
20. **Prugnolle F, Manica A & Balloux F** (2005) Geography predicts neutral genetic diversity of human populations. *Curr. Biol.* **15**, R159–160.
21. **Ramachandran S, Deshpande O, Roseman CC et al.** (2005) Support from the relationship of genetic and geographic distance in human populations for a serial founder effect originating in Africa. *Proc. Natl Acad. Sci. USA* **102**, 15942–15947.