

Chapter 11

The Adaptations That Have Made Us Human: The Genome



Elena Bosch

Abstract Identifying what makes us humans and understanding the genetic basis of our uniqueness are issues of interest and debate in biology. By analyzing genomic data we can detect and use the footprints that adaptive (positive) natural selection leaves in our genome to identify human adaptations at different timescales, as well as the resulting phenotypical traits that in current populations define us as humans. Although genomics has allowed us to recognize some of these adaptations from the pattern of variation expected under the classical or hard selective sweep model, the field starts to address novel approaches. These new strategies include not only methods to identify polygenic adaptations but also multidisciplinary procedures to understand the relationship between the genomic footprints of selection, the underlying functional genetic variants, the adaptive phenotype they determine, and the selective pressure that has driven each adaptation.

Keywords Human adaptation · Natural selection · Positive selection · Selective sweep · Polygenic selection

11.1 Introduction

Natural selection refers to the gradual, nonrandom process of accumulation of beneficial and heritable characteristics that we observe generation after generation in populations as a result of the existence of heritable variation in the reproductive success and/or survival of individuals. The concept, proposed simultaneously by Charles Darwin and Alfred R. Wallace (Darwin and Wallace 1858), is one of the main mechanisms that explain evolutionary change, but above all it is the only evolutionary force that facilitates the adaptation of populations and species to their changing environments.

E. Bosch (✉)

Department of Medicine and Life Sciences, Institute of Evolutionary Biology (UPF-CSIC),
Universitat Pompeu Fabra, Barcelona, Spain
e-mail: elena.bosch@upf.edu

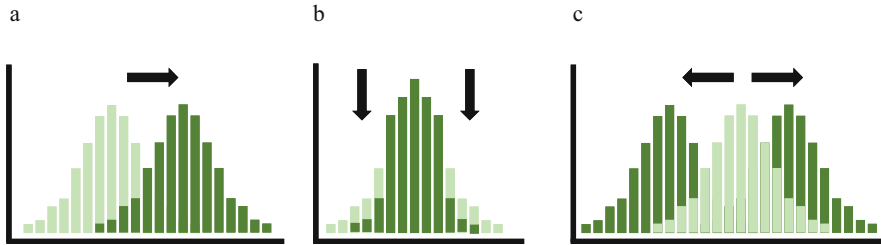


Fig. 11.1 Types of natural selection. (a) Directional (positive or Darwinian) selection. This is a type of selection in which the individuals of the population presenting the features of one of the extremes of the phenotypic distribution for a biological trait are favored. Consequently, the phenotypic distribution of the population will move towards the extreme with the most favored phenotypes. (b) Stabilizing (purifying or negative) selection. The extremes of the distribution of a particular phenotype are selected against when a particular range of values in the distribution is favored. Thus, the phenotypic distribution becomes narrowed. (c) Disruptive or balancing selection. A type of selection in which the individuals from the two extremes of the phenotypic distribution in the population are favored. Consequently, those individuals with intermediate phenotypic features disappear from the population and the distribution splits into two (or more) peaks that will represent the contrasting phenotypes simultaneously favored in that given environment

Thanks to the development of different high-throughput genotyping and sequencing platforms during the last decades and the subsequent abundance of genomic data from different human populations, we can easily test which specific regions of our genome show patterns of variation compatible with different types of natural selection (Fig. 11.1). Moreover, we can do this search without any a priori assumption about which genes or biological functions have facilitated our adaptation to different environments. However, once these candidate regions with patterns of variation compatible with the action of natural selection are recognized, we will then need to identify the true underlying adaptive variants, the adaptive phenotype they determine, as well as the selective pressure that has driven each adaptation. In this review I will present the main strategies for identifying the footprint of adaptive natural selection (also called Darwinian, positive or directional selection; Fig. 11.1a), some examples of well-characterized human adaptations identified with these methods, and the main limitations in the interpretation and understanding of these footprints.

Depending on the evolutionary timeframe in which we want to interrogate the action of natural selection, we will use strategies based on comparing genomes: between the human genome and that of other primate species, between different human populations (and this is the level at which most methods are more robust), or even between particular human populations and different extinct hominids, thanks to the recent availability of the genomes of several Neanderthal and Denisovan individuals. In this way, we will be able to infer species-specific adaptations, selective events geographically restricted in particular human populations, or even adaptive introgressions received from archaic humans.

11.2 Cross-Species Comparisons

Selection tests based on cross-species comparisons allow us to identify substitutions and genomic changes that are specific to our lineage and that may potentially have contributed to our divergence. Therefore, they allow us to explore adaptations that occurred millions of years ago (for example, about six million years ago when we compare the human genome with that of the chimpanzee) and that characterize us as a species. In this case, one of the most commonly used strategies to identify the footprint of positive selection between species is the comparison of nucleotide substitution rates. Given a region of DNA coding for a protein, we can compare the proportion of nonsynonymous changes (nucleotide changes that imply an amino acid change in the protein) to the proportion of synonymous changes (nucleotide changes that do not imply an amino acid change) between the genomes of two species to calculate the ratio K_a/K_s , also referred to as dN/dS or w depending on the context. Assuming that synonymous changes are neutral (not affected by natural selection), the ratio between the two types of substitutions will measure the possible accumulation of deleterious or beneficial mutations, in a coding region. Usually, mutations at the DNA level that change an amino acid are not well received by natural selection, as they may compromise the function or structure of the encoded protein; therefore, most genes in our genome have a K_a/K_s ratio well below 1. In contrast, values above 1 are exceptional and are interpreted to occur when the accumulation of amino acid changes in a protein has been favored as these changes could represent evolutionary novelties and therefore be potentially adaptive in one of the compared lineages. Moreover, other related methods have been adapted to explore the K_a/K_s ratio in particular lineages or branches in a given phylogenetic tree, as well as, in particular codons or within particular protein domains (Yang and Nielsen 2002; Zhang et al. 2005). In general, while the aforementioned comparative analyses have limited power to detect selection on individual genes, they are powerful strategies to identify which biological functions have changed most significantly in our lineage relative to that of the chimpanzee or other nonhuman primates.

The great sequence similarity found between the human and chimpanzee genomes, and the fact that our proteins are virtually identical, has led us to explore other types of variation in our genomes to identify other possible biological innovations in our lineage. For example, the use of comparative genomic hybridization techniques with cDNA (complementary DNA) probes in *arrays* (Marques-Bonet et al. 2009) and more recently also, the analysis of read coverage (*coverage* or number of unique reads that include a given nucleotide in the reconstructed DNA sequence) differences across genomes after their massively parallel sequencing (Kronenberg et al. 2018) allows to identify human-specific gene duplications and expansions that could represent evolutionary novelties in our lineage, especially when these changes can be related to particular changes in function or to actual differential expression patterns between the compared species.

11.3 Tests Based on Intraspecific Variation

Tests based on the intraspecific variation of our genome (or human polymorphism) will allow to identify adaptations that occurred more recently in our lineage, from a few tens of thousands of years ago to more ancient adaptations depending on the signature they interrogate. Thus, we can study the human adaptative response to the particular selective pressures we experienced during different key points in our evolutionary history such as during the human expansion out of Africa and subsequent colonization of new territories, where the climatic conditions were very different from our ancestral savannah, or after the introduction of farming and pastoralism as a form of subsistence in the Neolithic (Sabeti et al. 2006). In this case, most genome-wide searches (*genome scans*) for positive selection in humans have focused on detecting the footprints of the so-called hard selective sweep model. The model assumes the appearance of a new allele in a given genomic region that provides a selective advantage to its carriers. Consequently, generation after generation, the new allele is expected to rapidly increase its frequency towards fixation, sweeping with it the immediate surrounding alleles in the original chromosome where it arose by mutation and thus eliminating the remaining allele combinations of the population not carrying that new adaptive allele (Fig. 11.2a). We have different statistical tests that allow us to capture different key features of the pattern of variation resulting from this classical or hard selective sweep: (a) frequency spectrum-based methods, which allow us to detect a reduction of diversity, an excess of high-frequency derived alleles and/or a tendency towards an excess of rare alleles; (b) linkage disequilibrium extension-based methods that detect unusually long haplotypes relative to their frequencies in the population; and (c) methods based on population differentiation that capture allelic variants with high levels of population differentiation and that are therefore highly informative when the selective events and subsequent genetic adaptations are geographically restricted.

11.3.1 Tests Based on Frequency Spectrum

According to the classical sweep model, when, in a given genomic region, an adaptive variant becomes fixed sweeping with it the immediate genomic surrounding variation, we will expect to find a drastic reduction of its diversity when compared to other regions of the genome evolving under neutrality, which are expected to present both recent and much older (and more frequent) polymorphisms in the population. Thus, we initially expect a significant loss of diversity around positively selected regions of the genome. Over time, new alleles will appear by mutation, but for these new alleles, probably being most likely neutral, it will take a long time to increase their frequencies and evenly restore the initial diversity. Under this scenario, in the pattern of variation of the genomic regions under positive selection, we expect to find a characteristic signature of low overall genetic diversity and some excess of

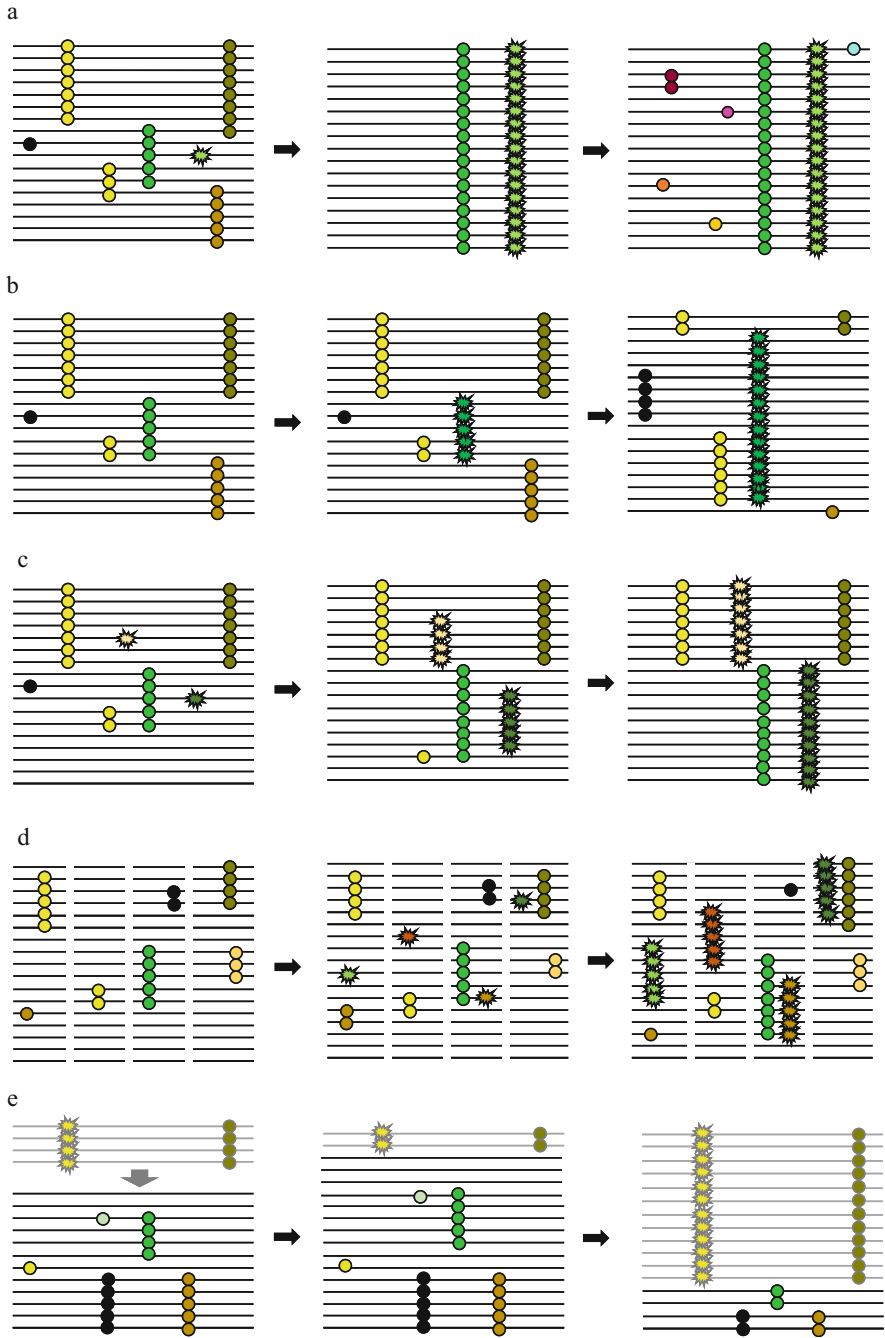


Fig. 11.2 Molecular footprints of positive natural selection or Darwinian selection. (a) Classical or hard sweep. (b) Soft sweep from pre-existing neutral variation in the population that becomes adaptive to an environmental change. (c) Soft sweep from the appearance of more than one adaptive

rare alleles once mutation starts to introduce new variation. The most commonly used tests that allow us to detect this type of deviation in the frequency spectrum are the Tajima's D test (Tajima 1989), and other related neutrality test statistics such as Fu and Li's D (Fu and Li 1993) and Fu's F (Fu 1997).

Derived alleles (i.e., nonancestral alleles) arise by mutation and generally tend to have lower allele frequencies than the ancestral ones. In a selective sweep, however, the derived alleles linked to the beneficial allele can be swept at high frequency in the so-called *hitchhiking effect*. Since many of these derived alleles will not reach complete fixation (as a result of an incomplete sweep or due to the occurrence of a recombination event between their genomic position and the selected allele), positive selection can create another feature in the pattern of variation of the genomic regions under positive selection consisting of an excess of derived alleles at high frequency with respect to neutrality. The most common approaches used to detect this excess are the Fay and Wu H test (Fay and Wu 2000) and the direct search for a high proportion of SNPs (single nucleotide polymorphisms) with a given derived allele frequency threshold, such as above 80% (Walsh et al. 2006).

11.3.2 Tests Based on Population Differentiation

In humans, the levels of population differentiation are mostly determined by genetic drift and migration. However, local adaptation (i.e., that adaptation specific to a given geographical area) can accentuate the levels of differentiation of particular loci in our genome. For example, while the O allele of the *Duffy* blood group is virtually absent outside Africa, the same allele presents very high frequencies in some African populations where it has been locally selected because it confers resistance to malaria. Other known cases of local genetic adaptations identified from unusual patterns of population differentiation include several variants that determine lighter skin pigmentation outside Africa, as well as other genetic variants that confer protection against malaria in Africa. We can quantify the extreme allele frequency differences observed in the *Duffy* system calculating the F_{ST} statistic (the most commonly used population differentiation measure) between African and non-African populations and using all SNPs in the human genome as a reference.

Other related strategies, such as the *Population Branch Statistic* (PBS) (Yi et al. 2010), use the values of F_{ST} obtained among three populations to identify which genomic positions exhibit extreme differentiations along each of the branches of the corresponding three-population tree. This design has been particularly informative for identifying loci involved in local adaptations to extreme conditions such as the arctic climate or high altitude by comparing two recently diverged populations but

Fig. 11.2 (continued) variant at the same locus. (d) Polygenic selection or selection on a complex trait. (e) Adaptive introgression

living under clearly contrasting local selective pressures (extreme cold or hypoxia, respectively, in the two previous cases, compared to normal conditions) with a third population with a deeper divergence time as reference (Yi et al. 2010; Fumagalli et al. 2015; Amorim et al. 2017).

11.3.3 Tests Based on Unusual Haplotype Extension or Linkage Disequilibrium

Selective sweeps create a distinctive footprint in the patterns of linkage disequilibrium that it is not expected under neutrality. When a new allele is favored by positive selection, it increases its frequency rapidly enough for recombination not to have time to break the initial association of the new favored allele with the remaining neighboring alleles on the ancestral chromosome where it appeared by mutation. Thus, recent positive selection produces an unusual combination of alleles at high frequencies that have long-range associations (or linkage disequilibrium) with other alleles present on the same chromosome. We will detect these long-range associations in the form of a haplotype or a particular combination of alleles (where the selected allele originally appeared) as long as recombination does not substantially erase these associations (which will eventually happen with time). On the contrary, since neutral alleles will require a long time to reach high population frequencies by genetic drift (if this is the case), they will never simultaneously present high frequency and associations over long distances. Thus, we have different tests for the detection of recent positive selection events (less than 30,000 years) that explore the relationship between allele frequency and the extent of linkage disequilibrium by directly comparing the length of haplotypes associated with different allele variants in each candidate region or the extent of haplotypic homozygosity.

The *iHS* (*integrated Haplotype Score*) and *XP-EHH* (*Cross-Population Extended Haplotype Homozygosity*) are the most commonly used linkage-based statistics. While *iHS* compares the extent of haplotypic homozygosity between the alleles of a given SNP in a population and is particularly informative for identifying variants under positive selection that have not yet reached fixation, i.e., detecting incomplete or ongoing selective sweeps (Voight et al. 2006), *XP-EHH* compares the extent of haplotypic homozygosity between populations and it is especially useful for detecting recently selected variants after a complete (or almost complete) selective sweep in a given population (Sabeti et al. 2007).

11.4 Verifying the Statistical Significance

The statistical tests described above allow to distinguish the pattern of genetic variation expected under positive natural selection from neutrality. However, particular events in the demographic history of each population (such as bottlenecks and population expansions) may create patterns of variation similar to those expected under positive selection. Thus, the direct inference of a selection event sometimes is complicated, especially when only a small number of loci are studied. However, since demographic events influence the entire genome more or less equally but adaptive events only act on the pattern of variation in the genomic regions under selection, it is possible to identify cases of positive selection by using genome-wide data and determining which regions of the genome have the most extreme deviations from neutrality using any of the statistics described. From the empirical distribution of the statistics values across the whole genome, either calculated by polymorphic position or by overlapping windows comprising several SNPs, we can easily identify which loci (or genomic regions) present extreme values in each distribution and consider them as the most likely candidates to have suffered strong selection pressures (Akey 2009). Although in this type of empirical approach we may obtain false positives, the subsequent identification of functional variants in these candidate regions and the complete understanding of their relationship with an adaptive phenotype and underlying selective pressure may help to convincingly identify new true adaptive cases. In turn, we must be also aware that we are likely to obtain many false negatives due to lack of power when focusing only on the topmost extreme signals of classical selective sweeps.

Through simulations we could improve the identification of *outliers* by applying a neutral model and/or considering different selection models incorporating, when appropriate, the specific demography of the populations used, the genome recombination map, and/or possible biases in the type of variation analyzed either by the use of SNP genotyping data or by the use of sequencing data with low coverage. Undoubtedly, if the neutral variation of the genome were perfectly modeled in these simulated scenarios, we could simply identify the variants that are more extreme in comparison with this neutral distribution as candidates for selection or we could even test whether they match a particular simulated selection regime given according to a particular demography and a given selection coefficient. However, we often do not know the real shape of this neutral distribution, neither how it has been affected by the particular demographic histories of human populations. On the other hand, and especially for populations with complex admixture histories in their past, when genomic information from several populations is available, the use of *admixture* plots has also proven to be a quite effective strategy to identify strong selection events and to be able to associate these to changes in allele frequencies in particular branches of their corresponding tree (Refoyo-Martínez et al. 2019). Similarly, estimating genome-wide genealogies for thousands of samples at a time can now be effectively used to estimate the dispersal rate of a given lineage relative to

alternative lineages over a given time period and to investigate different forms of selection (Speidel et al. 2019).

Finally, other approaches employ composite strategies to try to combine the specific footprints that each statistic detects in order to obtain greater sensitivity in the selection analysis, to directly identify the specific variants under selection or even to be able to classify the regions detected under selection according to whether they have been affected by recent/old and incomplete/complete selective sweeps in different human populations (Grossman et al. 2010; Pybus et al. 2015; Szpak et al. 2018). Since each selection statistic captures different aspects of the selective sweep, composite strategies help to avoid the particular false positives of individual statistics and provide greater statistical power to detect selection.

11.5 Beyond the Classic Selective Sweep

Although most studies of positive selection in humans have focused on capturing the footprints of classical selective sweeps, other models of selection such as polygenic selection (where multiple loci contribute to a given adaptive phenotype) and so-called soft selective sweeps (when different novel variants at a particular loci are selected at the same time or when selection favors a pre-existing polymorphic variant in the population that becomes adaptive due to a change in the environment) may have contributed equally to human adaptation (Pritchard et al. 2010). In these cases, the footprints that appear in the pattern of variation in regions under selection are clearly more subtle and difficult to capture with the standard methods developed to detect the so-called hard or classical selective sweeps (Fig. 11.2).

In soft sweeps, since the favored variants will be present in different haplotypes or chromosome types (Fig. 11.2b, c), when these increase in frequency and become fixed, we will not be able to detect the strong signatures of classical selective sweep, where a single variant that has just appeared by mutation in a single haplotype is selected. In this case, statistics such as nSL (Ferrer-Admetlla et al. 2014), H12 and H2/H1 (Garud and Rosenberg 2015), or SDS (Field et al. 2016) have been proposed as strategies that would improve the detection of these type of weaker signatures. Similar to iHS, the nSL statistic was designed to capture the increase in haplotypic homozygosity that positive selection creates when favoring a given segregating variant in the population. In this case, nSL measures the extent of homozygosity from the number of segregating sites with shared alleles across the haplotypes that accompany the corresponding ancestral and derived alleles of a given polymorphic position without requiring a genetic map (Ferrer-Admetlla et al. 2014). On the other hand, the H12 and H2/H1 statistics, also based on haplotypic homozygosity, allow considering, depending on the case, all the haplotypes of the population, the most abundant haplotype or the two most frequent haplotypes present in the sample to capture and distinguish, thanks to their combination, both hard (where we expect a unique major haplotype favored) and soft (with two or more haplotypes selected at the same time) sweeps (Garud and Rosenberg 2015). The SDS statistic (*Singleton*

Density Score; Field et al. 2016) uses whole genome sequence data to infer changes in SNP allele frequencies from the distortion that recent selection creates on the genealogies of the sampled haplotypes, resulting in shorter terminal branches in the case of favored alleles. The method allows the detection of both classical sweeps and cases of recent positive selection from previously pre-existing neutral variation.

In the case of polygenic selection (Fig. 11.2d), the resulting genomic footprints may only involve very subtle coordinated changes of allele frequencies at the multiple loci involved in each particular adaptive function or trait. As most human phenotypes are likely to have a complex genetic architecture and are not the result of just one or a few strong effect variants, several approaches and strategies are being devised to recognize the molecular footprint of polygenic adaptation. Although the footprints of selection on each of the genes involved in these complex phenotypes may often not be sufficient to stand out individually in a genome-wide context, an efficient approach to capture this type of scenario is to compare the evidence of adaptation across the entire set of genes associated with particular phenotypes or complex traits at once (Daub et al. 2013). However, it is often the case that a complete knowledge of all the genes involved in a complex phenotype is not always available, nor it is easy to model the really small effects that some of these loci have on these complex traits. Especially in native populations, usually not considered in large association studies or GWAS consortia, the use of *polygenic risk scores* based on reference populations to model the effects of each of these loci could be problematic. Other approaches to detect the more subtle adaptive patterns expected in the case of polygenic selection focus on correlating allele frequency changes with different environmental variables, such as temperature, humidity, solar radiation, or nutrient availability, among others. In this case, it is important to correct for the genetic structure of the populations used in the analysis, but when using sequencing data, the resulting environmental associations identified allow to directly relate specific adaptive genetic variants with a causal selective pressure. This strategy has successfully identified functional human polymorphisms associated with tolerance to cold, radiation, and pigmentation (Hancock et al. 2011).

Finally, genetic introgression from close species could, in exceptional cases, facilitate our adaptation through a pre-existing adaptive variant in these other species (Fig. 11.2e). Different specific statistical approaches are available to detect introgression which have allowed to describe several examples of adaptive variants that different human populations would have acquired from introgression with Neanderthals or Denisovans (Racimo et al. 2015). This is the case, for instance, of the TLR6-TLR1-TLR10 receptors involved in the recognition of external pathogens in European populations (Dannemann et al. 2016) or of the *EPAS1* gene in Tibetan populations, where it would facilitate their adaptation to hypoxia (Huerta-Sánchez et al. 2014).

11.6 Limitations and Future of Selection Studies

Regardless of the mode of selection operating, once any signature of positive selection has been identified in the genome, we need to understand its biological relevance as a case of human genetic adaptation. We need to go from the putative genotype under selection to the adaptive phenotype. This includes not only deciphering the molecular action and underlying adaptive phenotype of the selected variants, but also identifying the selective pressure at play in each case. Although the number of fully characterized cases of positive selection is still limited (Rees et al. 2020), they have allowed us to identify and understand different human adaptations that have undoubtedly contributed to our survival and colonizing success as a species (see examples related to our defense against pathogens, diet and nutrient availability, ultraviolet radiation, and altitude in Table 11.1). The bottleneck in the studies of positive selection is not the identification of the candidate regions per se but precisely the huge difficulty we have in understanding, without a priori hypothesis or the appropriate experimental design to investigate particular selective pressures, the relationship between the genotypes of the candidate regions found under selection and the corresponding adaptive phenotypes these determine.

The first examples of successfully characterized cases of selection began with prior knowledge of an adaptive phenotype, for which genetic evidence for selection

Table 11.1 Examples of genetic adaptations in humans

Adaptation	Genes	Population or geographic region
Language	<i>FOXP2</i>	Human lineage
Malaria	<i>HBB, HBA, HPA, GYPA, GYPB, GYPC, G6PD, FY</i>	Sub-Saharan Africa
<i>Trypanosoma</i>	<i>APOL1</i>	Africa
Flagellated bacteria	<i>TLR5</i>	Africa
General immune response	<i>TLR1-TLR6-TLR10, HLA group, CD5</i>	Multiple populations
UV radiation	<i>SLC24A5, SLC45A2, OCA2, MC1R, TYRP1, DCT</i>	European, Asian
Hair morphology	<i>WWTP</i>	Asian
Altitude	<i>EPAS1, ELGN1, PPARA</i>	Tibetans
	<i>VAV3, CBARA1, ARNT2</i>	Ethiopian Highlands
	<i>ELGN1, EDNRA, NOS2A, BRINP3</i>	Andean Population
Lactase persistence	<i>TBI</i>	European and African
Diet—Zinc	<i>SLC30A, SLC39A8</i>	East Asia and Africa
Diet—Selenium	<i>DI2, GPX1, GPX3, CELF1, SPS2, SEPSECS</i>	China
Diet—Iron	<i>HFE</i>	Europeans
High-fat diet	<i>FADS</i>	Inuit, Native Americans
Arsenic toxicity	<i>AS3MT</i>	Argentina

was subsequently identified, usually targeting plausible candidate genes (Vitti et al. 2013). The search of footprints of selection at the genomic level has the great potential to allow us to infer new hypotheses to test new cases of selection without any bias or prior assumption regarding which genomic regions might facilitate adaptation in each selective context. However, we should be aware that most of these genome-wide searches often provide lists of candidate adaptive regions with no indication of the possible functional variation or specific adaptive phenotypes that these regions might have facilitated. Thus, their interpretation and understanding are not straightforward and follow-up functional studies are usually required to validate the potential adaptive variants and phenotypes of each case in relation to particular selective pressures. Moreover, most signatures of positive selection usually span large regions of the genome, including multiple functional variants, and many different genes often with more than one function each and thus potentially affecting many different phenotypes. The work by Wang et al. (2020) illustrates the difficulty in identifying the variants and adaptive phenotypes behind a candidate region when investigating one of the strongest footprints of positive selection detected in humans: the 2p21.3 locus. The locus contains the lactase gene and the selection signals found in European populations have been associated with the derived allele of a regulatory variant that allows the persistence of lactase in adulthood and therefore the digestion of milk. In this case, it was classically inferred that, in populations where fresh milk is used as a food resource, carriers of the allele variant that facilitates the persistence of lactase would have increased their survival chances in past episodes of famine. However, by investigating the genomic content of the entire region under selection, as well as some of the phenotypes associated with the region in other species, the authors identified a microRNA (miR-128-1) in the European favored haplotype, directly involved in metabolic processes of energy expenditure that could also represent an adaptation to periods of famine in our past.

Those genomic scans of positive selection based on the comparison of populations living in different environments, or even in extreme conditions, have allowed to infer with some success a more direct causal relationship between the potential selective forces considered a priori and the candidate regions for selection identified in the genome. This type of design applied to diverse human populations living at high altitude in the mountains of Tibet, the Andes, or in Ethiopia has allowed to identify, for example, different human genetic adaptations, some of them convergent, related to hypoxia (Yi et al. 2010; Huerta-Sánchez et al. 2013; Borda et al. 2020). In other cases, combining genome-wide genetic data with phenotypic data related to the adaptive hypothesis also provides a promising strategy when trying to identify the footprints of positive selection related to particular adaptive phenotypes, such as short stature in pygmy populations from Africa (Jarvis et al. 2012; Mendizabal et al. 2012) or the Andaman Islands (Mondal et al. 2016).

Undoubtedly, the growing number of functional annotations in the coding and regulatory regions of the human genome, the increasing availability of genome-wide association studies for several complex phenotypes, together with the adoption of new experimental strategies to interrogate the relationship between genotype and phenotype, and the incorporation of other selection models will not only improve our

understanding of the various adaptations that have occurred in our evolutionary history and thus define us as humans, but also identify functional variants in our genome with potential biomedical relevance.

Acknowledgments This review has been possible thanks to the support of the “Dirección General de Investigación, Generalidad de Cataluña” (2017SGR702), the “Ministerio de Ciencia e Innovación,” and the “Agencia Estatal de Investigación (AEI)” through the project PID2019-110933GB-E00/AEI/10.13039/501100011033 and the “Unidad de Excelencia María de Maeztu” funded by the MCIN and the AEI (Doi: 10.13039/501100011033; Ref: CEX2018-000792-M).

References

- Akey JM (2009) Constructing genomic maps of positive selection in humans: where do we go from here? *Genome Res* 19:711–722
- Amorim G, Nunes CE, Meyer K et al (2017) Genetic signature of natural selection in first Americans. *Proc Natl Acad Sci U S A* 114:2195–2199
- Borda V, Alvim I, Mendes M et al (2020) The genetic structure and adaptation of Andean highlanders and Amazonians are influenced by the interplay between geography and culture. *Proc Natl Acad Sci U S A* 117:32557–32565
- Dannemann M, Andrés AM, Kelso J (2016) Introgression of Neandertal- and Denisovan-like haplotypes contributes to adaptive variation in human toll-like receptors. *Am J Hum Genet* 98:22–33
- Darwin C, Wallace A (1858) On the tendency of species to form varieties; and on the perpetuation of varieties and species by natural means of selection. *J Proc Linn Soc London Zool* 3:45–62
- Daub JT, Hofer T, Cutivet E et al (2013) Evidence for polygenic adaptation to pathogens in the human genome. *Mol Biol Evol* 30:1544–1558
- Fay JC, Wu C (2000) Hitchhiking under positive Darwinian selection. *Genetics* 155:1405–1413
- Ferrer-Admetlla A, Liang M, Korneliussen T, Nielsen R (2014) On detecting incomplete soft or hard selective sweeps using haplotype structure. *Mol Biol Evol* 31:1275–1291
- Field Y, Boyle EA, Telis N et al (2016) Detection of human adaptation during the past 2000 years. *Science* 354:760–764
- Fu YX (1997) Statistical tests of neutrality of mutations against population growth, hitchhiking and background selection. *Genetics* 147:915–925
- Fu YX, Li WH (1993) Statistical tests of neutrality of mutations. *Genetics* 133:693–709
- Fumagalli M, Moltke I, Grarup N et al (2015) Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science* 349:1343–1347
- Garud NR, Rosenberg NA (2015) Enhancing the mathematical properties of new haplotype homozygosity statistics for the detection of selective sweeps. *Theor Popul Biol* 102:94–101
- Grossman SR, Shlyakhter I, Shylakhter I et al (2010) A composite of multiple signals distinguishes causal variants in regions of positive selection. *Science* 327:883–886
- Hancock AM, Witonsky DB, Alkorta-Aranburu G et al (2011) Adaptations to climate-mediated selective pressures in humans. *PLoS Genet* 7:e1001375
- Huerta-Sánchez E, DeGiorgio M, Pagani L et al (2013) Genetic signatures reveal high-altitude adaptation in a set of Ethiopian populations. *Mol Biol Evol* 30:1877–1888
- Huerta-Sánchez E, Jin X, Asan et al (2014) Altitude adaptation in Tibetans caused by introgression of Denisovan-like DNA. *Nature* 512:194–197
- Jarvis JP, Scheinfeldt LB, Soi S et al (2012) Patterns of ancestry, signatures of natural selection, and genetic association with stature in Western African pygmies. *PLoS Genet* 8:e1002641
- Kronenberg ZN, Fiddes IT, Gordon D et al (2018) High-resolution comparative analysis of great ape genomes. *Science* 360:eaar6343

- Marques-Bonet T, Kidd JM, Ventura M et al (2009) A burst of segmental duplications in the genome of the African great ape ancestor. *Nature* 457:877–881
- Mendizabal I, Marigorta UM, Lao O, Comas D (2012) Adaptive evolution of loci covarying with the human African pygmy phenotype. *Hum Genet* 131:1305–1317
- Mondal M, Casals F, Xu T et al (2016) Genomic analysis of Andamanese provides insights into ancient human migration into Asia and adaptation. *Nat Genet* 48:1066–1070
- Pritchard JK, Pickrell JK, Coop G (2010) The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol* 20:R208–R215
- Pybus M, Luisi P, Dall’Olio GM et al (2015) Hierarchical boosting: a machine-learning framework to detect and classify hard selective sweeps in human populations. *Bioinformatics* 31:3946–3952
- Racimo F, Sankararaman S, Nielsen R, Huerta-Sánchez E (2015) Evidence for archaic adaptive introgression in humans. *Nat Rev Genet* 16:359–371
- Rees JS, Castellano S, Andrés AM (2020) The genomics of human local adaptation. *Trends Genet* 36:415–428
- Refoyo-Martínez A, Da Fonseca RR, Halldórsdóttir K et al (2019) Identifying loci under positive selection in complex population histories. *Genome Res* 29:1506–1520
- Sabeti PC, Schaffner SF, Fry B et al (2006) Positive natural selection in the human lineage. *Science* 312:1614–1620
- Sabeti PC, Varilly P, Fry B et al (2007) Genome-wide detection and characterization of positive selection in human populations. *Nature* 449:913–918
- Speidel L, Forest M, Shi S, Myers SR (2019) A method for genome-wide genealogy estimation for thousands of samples. *Nat Genet* 51:1321–1329
- Szpak M, Mezzavilla M, Ayub Q et al (2018) FineMAV: prioritizing candidate genetic variants driving local adaptations in human populations. *Genome Biol* 19:5
- Tajima F (1989) Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* 123:585–595
- Vitti JJ, Grossman SR, Sabeti PC (2013) Detecting natural selection in genomic data. *Annu Rev Genet* 47:97–120
- Voight BF, Kudaravalli S, Wen X, Pritchard JK (2006) A map of recent positive selection in the human genome. *PLoS Biol* 4:e72
- Walsh EC, Sabete P, Hutcheson HB et al (2006) Searching for signals of evolutionary selection in 168 genes related to immune function. *Hum Genet* 119:92–102
- Wang L, Sinnott-Armstrong N, Wagschal A et al (2020) A MicroRNA linking human positive selection and metabolic disorders. *Cell* 183:684–701
- Yang Z, Nielsen R (2002) Codon-substitution models for detecting molecular adaptation at individual sites along specific lineages. *Mol Biol Evol* 19:908–917
- Yi X, Liang Y, Huerta-Sánchez E et al (2010) Sequencing of 50 human exomes reveals adaptation to high altitude. *Science* 329:75–78
- Zhang J, Nielsen R, Yang Z (2005) Evaluation of an improved branch-site likelihood method for detecting positive selection at the molecular level. *Mol Biol Evol* 22:2472–2479