

## 2.6 Natural selection: II. Positive selection and adaptation

The previous chapter gave an introduction to the basic model of fitness and selection, and the role of purifying selection. Here we explore positive selection in greater detail, illustrated with key examples in humans.

**Positive selection and adaptation.** We now come to the type of selection that is arguably the most interesting and – as we discuss in the next chapter – the most argued-about form of selection: positive selection in favor of advantageous phenotypes and alleles.

Positive selection is the central organizing force of evolutionary change. It drives populations to adapt to their environments, and over longer evolutionary timescales it drives the evolution of new forms and functions at all levels: for example, the emergence of multicellular eukaryotes and of animals; the transition from fish to amphibians that enabled vertebrates to move onto land; the evolution of primates, of apes, and of humans.

Key evolutionary changes in the human lineage include the transition to bipedalism; bigger brains; changes in body size and shape, musculature, body hair and so on; enhanced capacity for language and highly complex social structures. All of these changes are genetically encoded and were presumably driven, at least in part, by positive selection. Moreover, we now have many examples where aspects of these genetic transitions have been elucidated, although there is still much more to be learned.

In more recent human evolution, during the last ~70,000 years, humans have spread around the globe to inhabit nearly all the world's land masses and ecosystems. Prehistoric humans successfully colonized a huge range of **environments**: environments with extreme cold and ice, or extreme heat and humidity; high altitude in Tibet, the Andes, east Africa, and elsewhere; tropical rainforests; deserts. Humans subsist on a wide range of **foods**, and encounter a diversity of infectious **pathogens**. All of these factors must have exerted strong selective pressures on human populations, driving both genetic adaptations—as well as cultural adaptations such as innovations in clothing, hunting and agriculture<sup>268 a</sup>.

Genetic adaptation proceeds mainly through two general types of processes: **selective sweeps** and **polygenic adaptation**. In a selective sweep, selection drives a strongly advantageous allele from low to high frequency in a population. In contrast, polygenic adaptation is driven by small shifts in allele frequencies spread across many loci, and is most relevant for complex traits. In practice, these two models are extremes along a spectrum, and adaptation may often proceed through a mixture of both types of processes.

Here we describe the features of both types of adaptation, as well as a third model, **balancing selection** which can drive both short term direc-



Figure 2.94: **Inuit seal hunter in Noatak, Alaska, 1929.** During the past 70,000 years, humans successfully colonized a wide range of different ecosystems, including arctic regions, deserts, and tropical rainforests. Credit: Edward Curtis 1929. [[Link](#)], Public Domain.

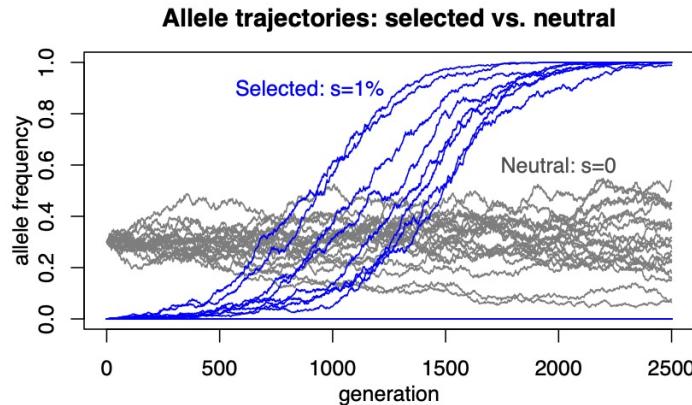
<sup>a</sup> William Clark, of the Lewis and Clark expedition, marveled at the ability of the native Mandan people (in present day North Dakota) to survive extreme cold: "This morning a boy of 13 years of age Came to the fort with his feet frozed, haveing Stayed out all night without fire, with no other Covering than a Small Robe goat skin leagens & a pr. Buffalow Skin mockersons— The Murcery Stood at 72° below the freeing point— Several others Stayed out all night not in the least hurt, This boy lost his Toes only...those people has ancered to bare more Cold than I thought it possible for man to indure."—Jan. 10, 1805.

Credit: Journals of the Lewis and Clark Expedition: [[Link](#)].

tional selection, and long-term stable polymorphism.

**Signatures of sweeps in genome data.** We've already covered basic models of positive selection in the last chapter. But, in practice, how can we find signals of positive selection in data?

The key insight here is that *strong positive selective drives very rapid allele frequency changes that would be extraordinarily unlikely for a neutral allele meandering under the random effects of genetic drift*. The next plot illustrates this for simulated data:

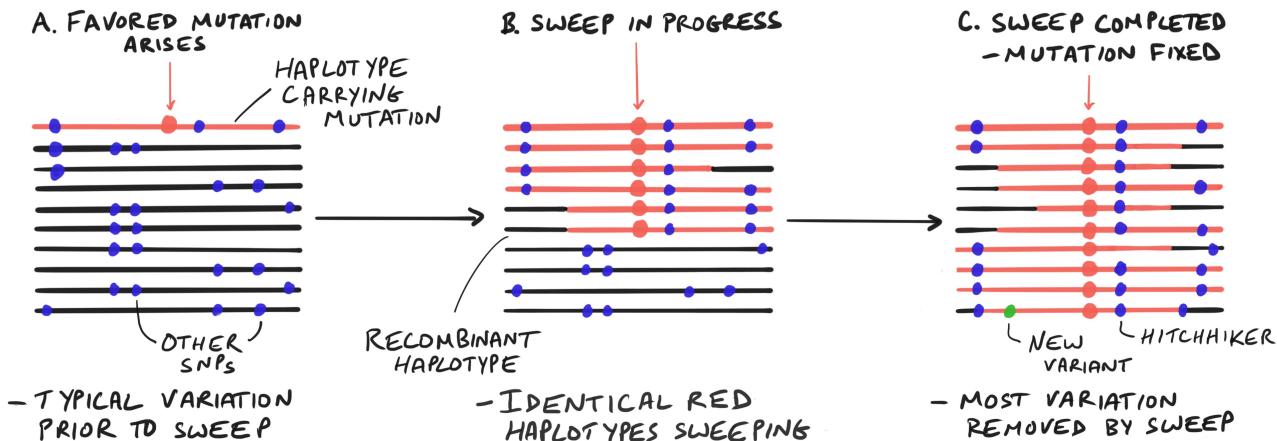


**Figure 2.95: Rapid increase in allele frequency for favored alleles versus neutral.** This simulation compares trajectories for favored alleles with  $s = 1\%$ , starting from new mutations, compared to common neutral alleles.

Simulations: 1000 favored alleles with  $s = 1\%$  each starting at a frequency of  $1/2N$  at time 0, versus 20 neutral alleles, starting at frequency 0.3. Only about 1% of the favored alleles spread to fixation; the remainder are on the  $y = 0$  line. Population size  $N = 10,000$ .

As you see above, the neutral alleles drift along aimlessly, while a favored allele rushes toward fixation<sup>269</sup>.

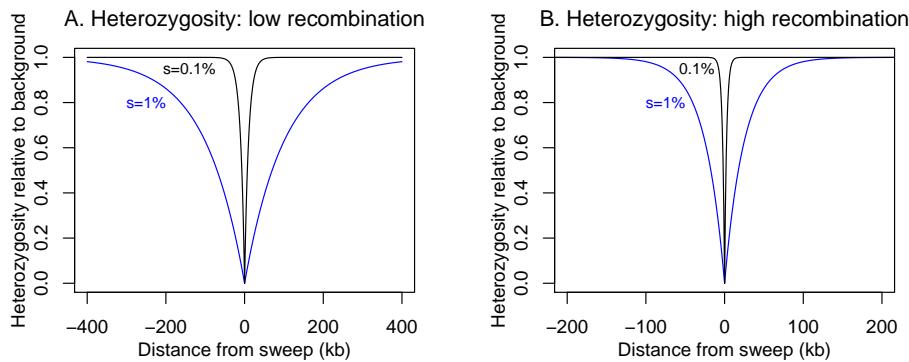
Crucially, this very rapid change in allele frequency distorts patterns of genetic variation in a large region in predictable ways. As the selected variant spreads rapidly through the population, it drags a haplotype up to high frequency along with it. This means that nearby neutral variants on the same haplotype are also dragged up to high frequency, in a process known as **genetic hitchhiking**<sup>270</sup>:



**Figure 2.96: Sweeps reduce variation in a linked region.** When a new favored mutation spreads through the population, it increases frequency very rapidly, and drags a long haplotype to high frequency along with it. Red indicates the haplotype on which the favored mutation occurred, as well as its descendants; to some extent this gets whittled down by recombination (black segments). We cannot observe the red versus black coloring directly, but we can infer this from the haplotype structure.

While the sweep is in progress the favored allele sits on a long, nearly identical haplotype. This contrasts markedly with relatively normal haplotypes carrying the ancestral allele<sup>271</sup>. Next, as the sweep completes, it essentially wipes out variation in a window around it, aside from any rare variants that arose during the sweep.

The size of the affected region depends on the speed of the sweep versus the local rate of recombination. A very fast sweep (large  $s$ ) carries a large haplotype with it, simply because recombination does not have time to chop it down very far; similarly, the sweep region would also be larger in regions with a low recombination rate. The reduction in heterozygosity as a function of distance  $x$  from a selected site can be approximated by  $1 - e^{-\tau rx}$  where  $\tau = 2\log(2N)/s$ , and  $r$  is the local recombination rate<sup>272</sup>. Notice that the size of the swept region depends on the ratio  $r/s$ :



**Figure 2.97: Reduced genetic diversity around the site of a sweep due to hitch-hiking.** The plots show the expected reduction in heterozygosity relative to the average heterozygosity at neutral sites, assuming a recently completed sweep at position zero. The size of the affected region depends on  $s$  and on the local recombination rate. Parameters:  $r = 0.5 \times 10^{-8} / \text{bp}$  (left plot), and  $2 \times 10^{-8} / \text{bp}$  (right plot);  $N = 10,000$ . Figure inspired by Coop (2020) [Link], Fig. 13.5.

Starting from this intuition about how sweeps impact patterns of variation, there has been a huge amount of work on methods for detecting selective sweeps using genetic data<sup>273</sup>. In short, these methods use a variety of features in the data to detect positive selection:

- *long haplotypes with low genetic variability around a putative selected allele*; these contrast with typical patterns of variation on haplotypes carrying the ancestral allele at the selected site (ongoing sweep);
- *low genetic diversity in a genomic region around a recently fixed site* (recently completed sweep);
- *most variants in a region are young and at low frequency* (recently completed sweep);
- *large allele frequency differences at a selected site, and potentially nearby sites*, between populations where the sweep is occurring compared to control populations, or in time series data from ancient DNA (ongoing or completed sweep)

We illustrate these principles using several examples of recent selection in humans, highlighting some of the key selective pressures as well as signals in the genetic data.

**Selection pressures due to diet.** Diet has been a major driver of selection in recent human evolution. As humans spread around the globe to inhabit virtually all possible ecosystems, they were forced to learn to survive on a wide array of different foods. Further enormous shifts in diet were driven by the transition to **agriculture**, starting in the past 5,000–10,000 years, in many parts of the world. Several potential signals of selection have been hypothesized as relating to diet, including at the FADS locus, which is involved in metabolism of fatty acids<sup>274</sup> and at Amylase1, which is involved in starch digestion<sup>275</sup>.

The clearest diet-related signal is at the **lactase** locus. Lactase is the enzyme that is responsible for digesting the sugar lactose, which is present in milk. Most mammals stop consuming milk (and lactose) after weaning, and expression of the lactase gene is generally turned off in adults.

The first known evidence for dairy farming is in Anatolia (modern day Turkey) in the early Neolithic, about 9,000 years ago. Dairy farming subsequently became important in many places, including in Europe, in India and the Middle East, and in east Africa. This, in turn, provided a strong selective pressure for early humans to be able to digest milk throughout life. Consequently, several different **regulatory mutations** that cause the lactase gene to be expressed throughout life have spread to intermediate or high frequency in different farming populations. These regulatory mutations are often referred to as **lactase persistence** alleles as they cause lactase to persist throughout life.

The strongest signals of selection on lactase are found in Europe. As it happens, there are now extensive genotype data from early European populations, collected from skeletons<sup>b</sup>; this allows a rare opportunity to track the selective spread of an allele directly using allele frequencies in ancient DNA. Analysis by Iain Mathieson shows the remarkable spread of the lactase allele during the past 5,000 years to a modern frequency above 80% in some parts of Europe:

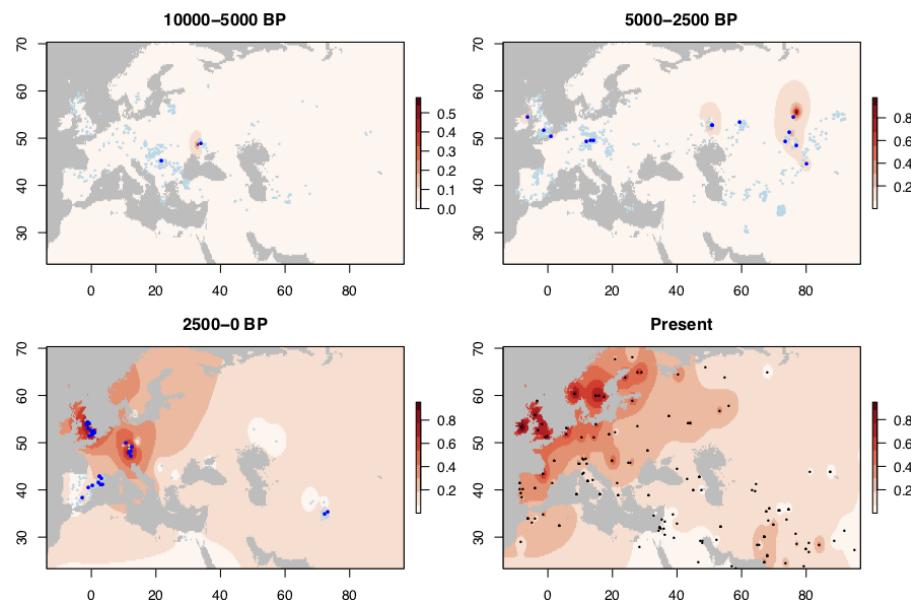


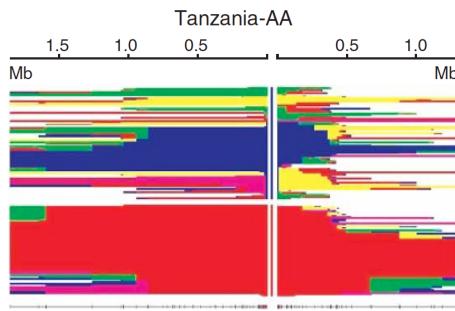
Figure 2.98: **Maasai herder, with cattle.** The lactase locus has been a target of selection in the Maasai and other east African farming populations. Credit: Nicor. [[Link](#)]. CC BY 3.0.

<sup>b</sup> We'll cover ancient DNA in Chapter 3.3.

Figure 2.99: **Increase in allele frequency of the lactase variant in Europe (from ancient DNA).** The maps indicate locations of samples with ancestral alleles (light blue) and derived alleles (dark blue) in each time period (BP = years Before Present); as well as spatially smoothed maps of the derived allele frequency in each period (red color scale beside each panel). Credit: Iain Mathieson blog post (2019) [[Link](#)]. Used with permission of the author.

Recall that the rate of spread of a favored allele depends on its selective advantage  $s$ . Using the data shown above, Mathieson estimated  $s$  for the lactase persistence allele at 2%–3% in central and western Europe, consistent with other estimates using haplotype patterns in modern data<sup>276</sup>. This makes lactase persistence one of most strongly selected traits in recent human evolution<sup>277</sup>.

Signals of selection at lactase are also found in other dairy farming populations. For example, dairy farming was practiced in east Africa by around 6,000 years ago and is a major food source for several east African groups<sup>278</sup>. As you can see below, the lactase locus shows strong signals of a sweeping haplotype in Tanzania: the derived allele sits on a long shared haplotype, in sharp contrast with the much higher diversity on ancestral haplotypes<sup>279</sup>:



**Figure 2.100: Signal of a partial lactase sweep in a Tanzanian population.** Visualization of haplotype sharing around a sweep-in-progress. The derived allele at the likely selected site is indicated in red at the center of the plot, and the ancestral allele in blue. The x-axis indicates location within the region in Mb; rows are haplotypes. Shared blocks of color indicate shared haplotypes extending outward from the selected site. Credit: From Figure 5, Sarah Tishkoff et al (2007) [Link] Used with permission.

The selected variants in African populations are distinct from the European variant, indicating that they arose from independent mutations, rather than being carried in by migration.

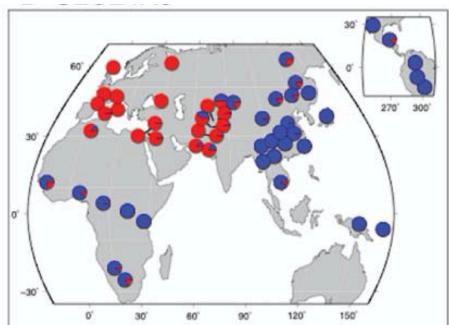
**Selection on pigmentation: SLC24A5.** Another important target of natural selection in human evolution is on **skin pigmentation**, as well as hair and eye coloring. Globally, populations that live close to the equator, including in our ancestral range in Africa, tend to have darker pigmentation. Populations at higher (and lower) latitudes tend to have lighter pigmentation<sup>280</sup>.

Variation in pigmentation seems to have been driven by strong selective pressures<sup>281</sup>. In regions with intense sunlight it is advantageous to have darker skin as this protects against ultraviolet (UV) damage from the sun. In addition to skin damage, excess UV radiation also degrades folic acid, deficiencies of which cause neural tube defects during pregnancy. However, too little UV is also bad, as UV catalyzes Vitamin D production; Vitamin D plays an important role in bone development, reproductive health, and other traits.

Remarkably, around ten different genes involved in pigmentation of skin, hair, or eyes show either clear or suggestive signals of selection in some part of the world<sup>282</sup>. In the case of related traits such as blue eyes and blond hair with signals of selection in Europe, it's unclear if UV, or some other factor such as sexual selection, was the main driver of selection.

Among the most striking pigmentation signals is a *missense variant at the*

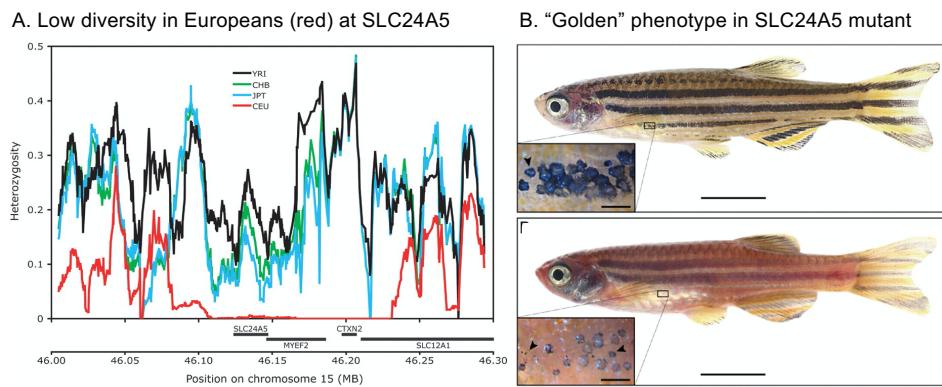
gene *SLC24A5*. The derived allele (in red), which causes lighter pigmentation, has swept to high frequencies throughout western Eurasia.



**Figure 2.101: Selective sweep in western Eurasia at *SLC24A5*.** The derived missense allele is shown in red. The populations, for example in the Americas, are representatives of indigenous populations. From Graham Coop et al 2009, Figure 2B. CC BY 4.

The allele frequency differences between populations at *SLC24A5* are among the most significant frequency differences anywhere in the genome, reflecting very strong selection for the derived allele<sup>283</sup>.

As expected for a recently completed sweep, this event has swept away genetic variation in a large region around *SLC24A5* in Europeans (red line in panel A); this contrasts sharply with more typical levels of variation in other populations. The role of *SLC24A5* in pigmentation is supported by human association data, and a zebrafish knockout:



**Figure 2.102: Selective sweep at the pigmentation gene *SLC24A5*.** A. Near-zero genetic diversity in a European population (red, CEU) near *SLC24A5*. An African (YRI) and two east Asian (CHB, JPT) populations have more normal patterns of genetic diversity across this region. B. Mutations in *SLC24A5* cause changes in melanophore coloring, as seen in a zebrafish mutant ("golden") at bottom, versus the wild type at top. Heterozygosity in Panel A is measured at pre-ascertained SNPs. From Figures 5A, 1A, B. Rebecca Lamason et al (2005) [Link] Used with permission.

Together, the lactase and *SLC24A5* examples illustrate classic features of selective sweeps: rapid changes in allele frequencies at selected sites; large sweeping haplotypes for sweeps in progress (lactase); removal of variation in regions around completed sweeps (*SLC24A5*).

But as we shall see next, not all sweeps show these characteristics.

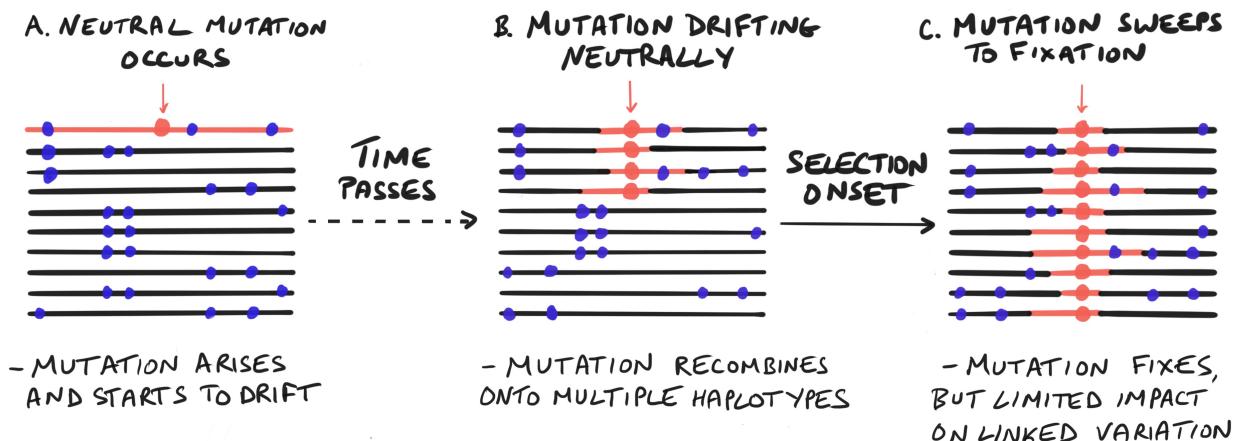
**Soft sweeps.** What would happen if a mutation is not immediately favored: instead it drifts along for a while, and only becomes favored some time later, after an environmental change? In this case, during the drift phase, the favored allele potentially has time to recombine onto multiple haplotype backgrounds. Then, when it *does* sweep, it carries multiple haplotypes with it, and the overall footprint of selection is greatly

reduced. This scenario may seem contrived, but there are many phenotypes that are favored only in specific environments, for example in the presence of a pathogen, a specific food source, or at high altitude – and otherwise neutral<sup>284 c</sup>.

A second scenario that could reduce the sweep signal is if multiple functionally equivalent mutations arise at about the same time and sweep together. These would likely occur on different haplotypes and the sweep signal would be greatly reduced.

In terminology developed by Pleuni Pennings and Joachim Hermisson in 2005, both of these scenarios are described as **soft sweeps**<sup>285</sup>. This term evokes the image of a mutation, or mutations, that sweep to fixation, without greatly disturbing the variation at nearby sites. This contrasts with the classical sweep model described above, which we now refer to as a **hard sweep**.

<sup>c</sup> For more about high altitude adaptation see EPAS1 in Chapters 2.4 and 3.3.



**Figure 2.103: Soft sweeps have minimal impact on variation in the linked region.** In one type of soft sweep, the mutation is initially neutral, and drifts to low or intermediate frequency in the population. During this time, recombination shuffles it onto multiple haplotypes. Selection then turns on, driving the allele to fixation, but without a strong hitchhiking effect.

**Selection for malaria resistance: Duffy.** One likely soft sweep occurred at the **DARC gene**, which encodes a cell-surface protein named **Duffy**, found on the surface of red blood cells. Duffy serves as a cell surface receptor for a class of chemokines, a type of signaling molecule<sup>286</sup>.

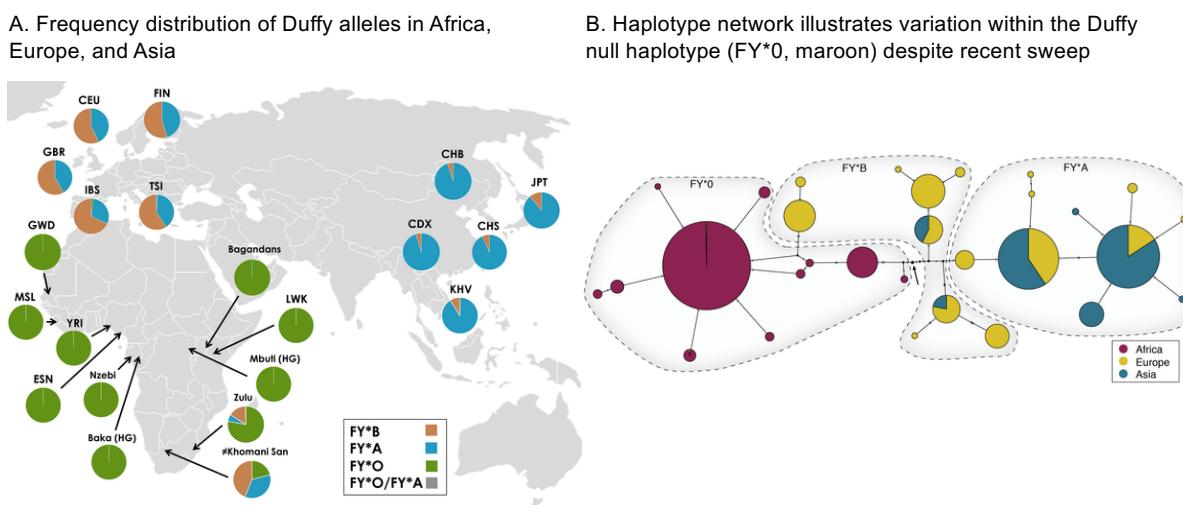
Duffy also plays a critical role in **malaria** infection and resistance. One species of malaria, **Plasmodium vivax**, binds the Duffy protein on the surface of red blood cells and uses it to enter cells. This property leads to a truly remarkable story of selection at the Duffy locus.

Most sub-Saharan Africans carry a derived variant near the Duffy locus that disrupts a DNA binding site for the transcription factor GATA1. GATA1 plays an important role in erythroid (red blood cell) development, and loss of this particular binding site eliminates Duffy expression specifically in erythrocytes while maintaining Duffy expression in other cell types<sup>287</sup>. This variant is known as the *Duffy null* allele (abbreviated FY\*0). The lack of Duffy expression in the null allele has the crucial effect

of blocking entry of vivax malaria into red blood cells. Thus Duffy null individuals are resistant to vivax malaria<sup>288 289</sup>.

The next striking fact about Duffy null is that it shows extreme frequency divergence between populations: it is essentially fixed in most sub-Saharan African populations, and essentially absent outside Africa. The null variant has the largest population differentiation of any high-frequency African allele, anywhere in the genome<sup>290</sup> (green allele in Panel A, below).

This might make you think that Duffy has been the target of a completed sweep in most of Africa: a sweep that presumably started after the major out-of-Africa migrations around 70,000 years ago. However studies of genetic variation show something surprising: namely that the Duffy null allele is carried on two distinct major haplotypes, with additional low frequency variants (maroon circles in Panel B):



**Figure 2.104: Soft sweep of the Duffy null allele in sub-Saharan Africa.** **A.** The Duffy null allele, shown here in green, swept to near fixation in many African populations subsequent to the out-of-Africa migrations. Non-African populations have a mix of A and B alleles (brown and red), corresponding to alternate missense variants. The B allele is ancestral. **B.** Visualization of SNP variation seen on each of the three allelic backgrounds in a 5kb region around the FY\*0 site. Each circle represents a different haplotype, and its area is proportional to frequency. Line segments between the haplotype circles indicate similarity. Notice that haplotype variation on the FY\*0 haplotype is nearly as high as on the A and B backgrounds, arguing against a recent hard sweep at this locus.

Credit: Figs 1, 2a from Kimberly McManus et al (2017) [[Link](#)]. CC BY 4.0

Instead, the high level of variation on the Duffy null background suggests that this is a soft sweep. One study estimated that two major Duffy null haplotypes were drifting at low frequency (0.1%) prior to the onset of strong selection around 45,000 years ago, at which point the Duffy null allele swept to fixation<sup>291</sup>.

There's one more surprising twist in this story, namely that Plasmodium vivax malaria is mainly found in Asia and Latin America – not in Africa! (The main malaria parasite in Africa is a different species, P. falciparum.) So why was there such strong selection at the Duffy locus, specifically in Africa?

Recent work shows that P. vivax is currently found in African chimpanzees

and gorillas<sup>292</sup>. So one plausible model is that vivax malaria jumped into humans around 45,000 years ago, and drove strong selection on pre-existing variation at the Duffy locus. Subsequent fixation of the Duffy null allele provided such powerful disease resistance that the human adaptation ultimately eliminated vivax malaria from human populations in Africa!

**Infectious diseases as major drivers of selection.** Malaria has long been a major cause of global disease and mortality<sup>293</sup>. As such, malaria has exerted strong selective effects on several other genes in addition to Duffy, including  $\alpha$ -globin,  $\beta$ -globin, and G6PD, coming up next.

Beyond malaria, infectious diseases in general are potent agents of natural selection. Like all species, we are continually barraged by a range of pathogens – viruses, bacteria, fungi, protists – that evolve rapidly to outwit our inbuilt defenses. During active infections, our bodies combat pathogens using a mixture of so-called innate and adaptive immune systems. While outside the scope of this book, it's interesting to note that during an infection our **adaptive immune systems** harness the principles of evolution, including genome modification and proliferation, to rapidly evolve B and T cells that recognize the infectious agents. This allows our own immune systems to adapt on the same timescales as rapidly evolving pathogens.

Moreover, the biological systems that combat infections are, themselves, finely tuned by evolution. As a result, several examples of selection in humans relate to pathogens and immunity: in addition to the malaria examples these include the MHC system which we discuss in the next chapter, the Toll-like receptor complex involved in innate immunity, and others.

**Balancing selection.** So far we have been focusing on positive selection that always favors one allele over the other. But what happens if the heterozygote fitness is higher than both homozygotes, or lower than both homozygotes?

To be more precise, recall our fitness model from the previous chapter, where the three genotypes have fitness 1,  $1 + hs$ , and  $1 + s$ , respectively. We have been looking at models where  $h$  is in the range  $[0, 1]$ . But if  $h$  is outside this range, then the heterozygotes are either better, or worse, than both homozygotes, leading to some very interesting models.

To understand this, recall from the previous chapter that we computed the expected change in allele frequency,  $\Delta_p$ , from one generation to the next:

$$\Delta_p = \frac{pq[s(1-h) + qh]}{\bar{w}}. \quad (2.85)$$

$\Delta_p$  tells us how allele frequency  $p$  changes over time. If  $h$  is in the range of  $[0, 1]$ , then we get simple directional selection:  $\Delta_p$  is always increasing (if  $s$  is positive) or always decreasing (if  $s$  is negative). But if  $h$  is out-

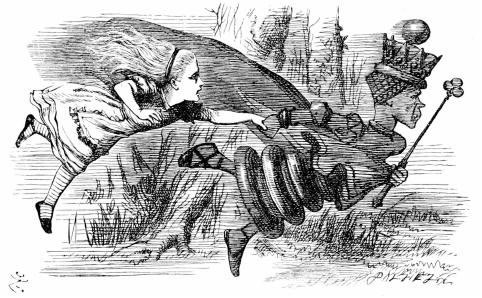


Figure 2.105: **The Red Queen Hypothesis** proposes that pathogens and their hosts are constantly evolving molecular systems to counter each in infection and defense. The name alludes to Lewis Carroll's Red Queen in *Through The Looking Glass* who tells Alice that "here, you see, it takes all the running you can do, to keep in the same place." Credit: Illustration by Sir John Tenniel, 1871. Public Domain.

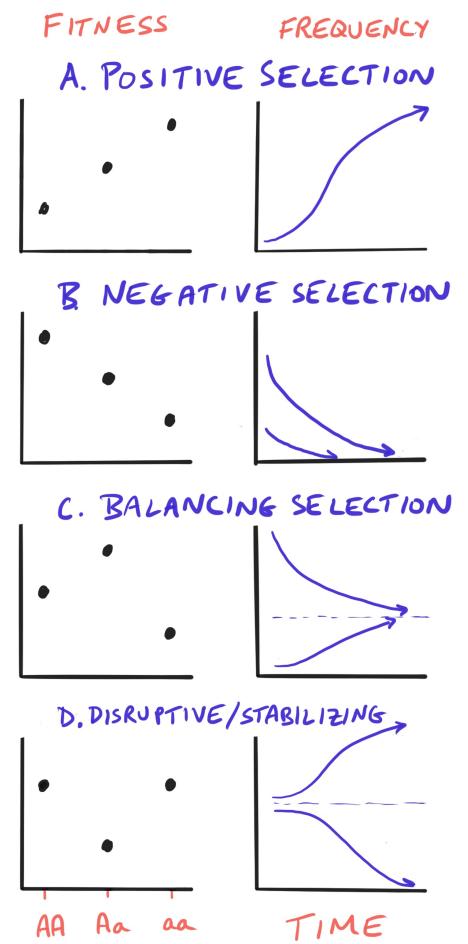
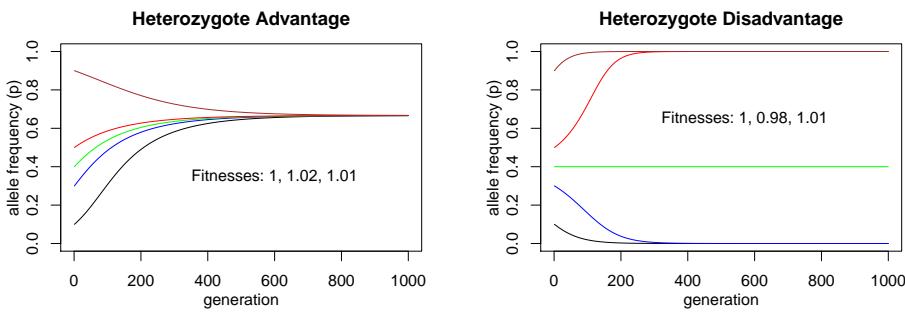


Figure 2.106: **Overview of models.** Left: Relative genotype fitnesses (y-axis), for AA, Aa, and aa genotypes. Right: Simplified frequency trajectories (y-axis) of the 'a' allele over time.

side this range, then the direction of change depends on allele frequency. There are two main scenarios, shown here:



**Figure 2.107: Allele frequency trajectories from different starting points.** **Left:** With heterozygote advantage, the allele trajectories converge to a stable equilibrium. **Right:** With heterozygote disadvantage, the trajectories diverge away to fixation or loss depending on the starting frequency. In the absence of drift, the green trajectory stays precisely at the unstable equilibrium. Parameters:  $s = 0.01, h = 2, \hat{p} = 2/3$  (left) and  $s = 0.01, h = -2, \hat{p} = 0.4$  (right).

In the left plot, the heterozygote is fitter than both homozygotes, and the allele frequencies converge toward a fixed stable point. This is known as **heterozygote advantage** and leads to **balancing selection**.

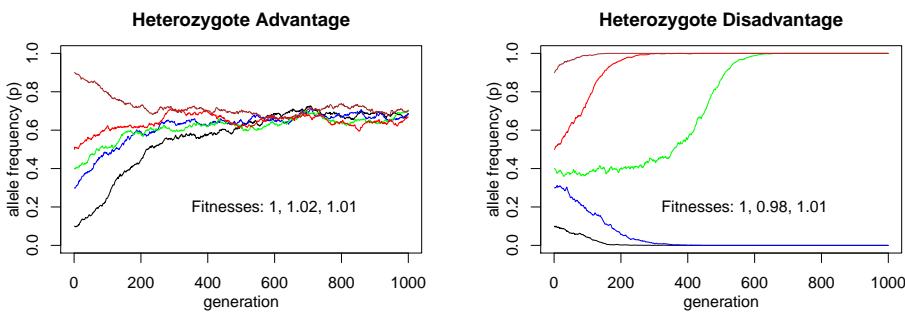
When the heterozygote is less fit than both homozygotes, the population evolves toward fixation of one allele or the other, depending on the starting point. This is called **heterozygote disadvantage** or **disruptive selection**.

In both cases we can solve for an **equilibrium** frequency – i.e., a value of  $p$  for which the allele frequencies don't change under this model<sup>294</sup>:

$$\hat{p} = \frac{h}{2h-1}. \quad (2.86)$$

(You can see here that when  $h$  is inside the range  $[0, 1]$  – i.e., directional selection – this equation does not produce meaningful allele frequencies between 0 and 1.)

We can also look at this with drift. With balancing selection (left), the population converges to the equilibrium and stays there. But when the heterozygote is less fit (right), we see that  $\hat{p}$  is an **unstable equilibrium**. As soon as the green trajectory drifts slightly away from  $\hat{p}$ , selection pushes it rapidly toward fixation or loss.



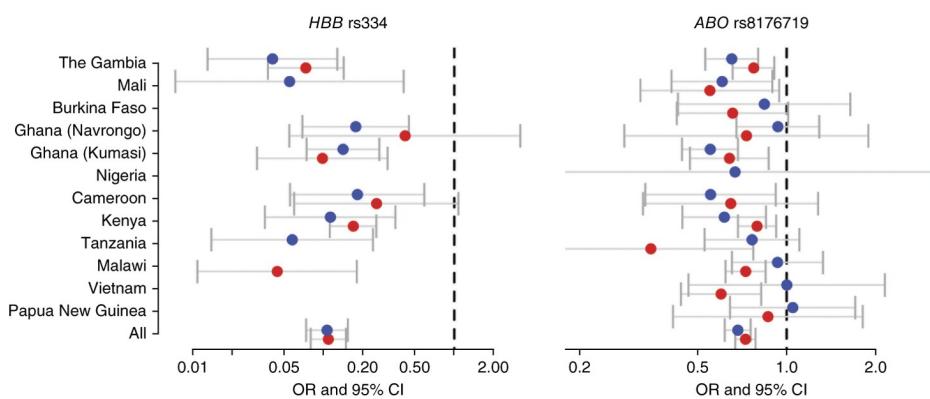
**Figure 2.108: Simulated frequency trajectories with drift.** **Left:** even with drift, heterozygote advantage allows stable balanced polymorphism that can persist for long timescales. **Right:** Drift causes trajectories to move away from the unstable equilibrium; in this case both fixation or loss are possible for the green trajectory. Parameters as above, and  $2N = 20,000$ .

To summarize: if the heterozygote has higher fitness than both homozygotes, then this leads to a stable polymorphic equilibrium. In practice this can last for many millions of years.

But if the heterozygote is worse than both homozygotes, then there is an unstable polymorphic equilibrium, which can result in fixation of either allele depending on the starting allele frequency. This type of selection is usually difficult to detect in practice<sup>295</sup>. However, a similar model will become important when we study **stabilizing selection** later in the book<sup>d</sup>. In that case the unstable equilibrium is at 0.5 and selection acts against minor alleles<sup>296</sup>.

**Balancing selection examples.** Perhaps the best-known example of balancing selection is for **sickle cell disease**. Sickle cell disease is caused by a missense mutation in the  $\beta$ -globin gene (also known as *HBB*), which encodes a major subunit of hemoglobin, the molecule that red blood cells use to transport oxygen<sup>297 298</sup>. In heterozygotes, the sickle cell mutation provides strong defense against both vivax and falciparum malaria without major side effects. However, individuals who are homozygotes for the missense mutation suffer devastating symptoms including hemolytic crisis, severe pain, kidney disease, and stroke.

The protective nature of the sickle allele was first documented in the 1950s<sup>299</sup>. A recent global study confirms that sickle heterozygotes benefit from extremely strong protection against malaria (odds ratio of 0.14, p-value=10<sup>-225</sup>), left panel below<sup>300</sup>:



<sup>d</sup> Stabilizing selection is likely the main form of selection acting against variation in gene expression and many complex traits (Chapter 4.8).

The sickle cell allele is common in most of central/western Africa, peaking at around 15% frequency in Angola<sup>301</sup>. Based on this we can estimate the frequency of sickle cell disease (i.e., sickle homozygotes) using the Hardy-Weinberg rule, as  $\sim 2.25\%$  (i.e.,  $0.15^2 \times 100$ ).

We can use Equation 2.86 to estimate  $h$ . Rearranging that expression gives us  $h = p/(2p - 1)$ : hence  $h = -0.21$ . Assuming  $s \approx -1$ , this implies a heterozygote fitness of 1.21, which is an extraordinarily large fitness effect for humans<sup>302</sup>.

Several other alleles are protective against malaria in heterozygotes but cause disease in homozygotes. In addition to the “classic” sickle cell missense mutation, a variety of other rarer mutations affect either the  $\text{ff}$ -globin or  $\text{fi}$ -globin genes, and produce varying levels of sickle cell-like disease. These diseases are referred to as  **$\alpha$ - or  $\beta$ -thalassemia**. Like the sickle cell mutation, they are mainly found in individuals with ancestry

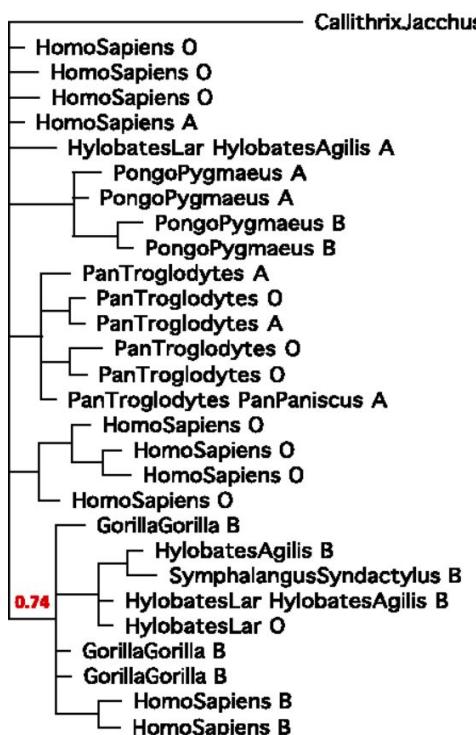
**Figure 2.109: Protective effects of HBB sickle and ABO type-O against malaria.** The x-axes show estimated odds ratios for the risk of cerebral malaria (red) and severe malarial anemia (blue), for samples from different countries. Estimates to the left of the vertical dashed lines indicate protective effects. The odds-ratio is (approximately) a measure of the risk for individuals with this genotype compared to controls; values < 1 indicate protection relative to controls. Grey bars show confidence intervals. Credit: From Figure 1, Malaria Genomic Epidemiology Network (2014) [Link] Used with permission.

from malaria-endemic regions, and are also likely spread by heterozygote advantage. Lastly, mutations in the enzyme **G6PD**, which plays a role in glucose metabolism, are also protective against malaria in heterozygotes but cause pathologies in homozygotes. Balancing selection likely maintains variation in all these genes<sup>303</sup>.

**Ancient trans-species balancing selection.** There's one more unique feature of balancing selection: if the selective pressures are stable, they can maintain polymorphism for extremely long times.

One of the best examples of this is the **ABO gene**, which is responsible for the ABO blood groups. ABO is an enzyme that modifies sugar attachments on cell surface proteins called glycoproteins. Two functional alleles, A and B, differ by a pair of missense variants that lead to different glycoprotein modifications. The third allele, O, carries a frameshift mutation that obliterates enzyme activity. You are likely familiar with the ABO system in the context of blood donations, as some combinations of blood types are incompatible donors and recipients: this is because unfamiliar glycoproteins can trigger immune reactions against donor blood cells. As shown in the figure above, the O allele at ABO is protective against malaria; more broadly the ABO alleles are associated with many different traits.

*Curiously, it turns out that the ABO alleles are actually shared among different species of apes and old world monkeys. Analysis shows that the alleles from the different species are actually shared in a single ancient coalescent tree, reaching back at least 20 million years, and probably longer<sup>304</sup>! This type of deep, ancestral sharing of alleles is known as a **trans-species polymorphism** and is extremely rare in the human genome overall:*



**Figure 2.110: Trans species polymorphism in the ABO blood group system.** This shows a phylogenetic tree of sequences from different ape species for exon 7 of the ABO gene which determines A/B/O blood type. Notice that most of the B alleles from different species cluster together in the lower part of the tree, indicating that they descend from a shared ancestral mutation. A and O mutations are shuffled together suggesting that O alleles may have arisen repeatedly (although the tree structure cannot be confidently determined in this clade. Species names: Callithrix (marmoset monkey); Homo (human); Hylobates (gibbon); Pongo (orangutan); Pan (chimpanzee); Gorilla (gorilla); Symphalangus (siamang). Credit: Figure 3A from Laure Ségurel et al 2012. [Link]

The extreme age of this polymorphic system indicates that it cannot be neutral, and must instead be preserved by some form of balancing selection. However, the precise explanation remains mysterious, as there is no obvious advantage to heterozygotes. We do know that cell surface molecules such as the glycoproteins modified by ABO, are frequently used as cellular entry points for pathogens (as for Duffy), and this may explain why blood group O provides some degree of malaria protection. It's possible that the different alleles provide protection against distinct pathogens, and that this creates selection pressure for maintaining diverse alleles. The details remain to be discovered<sup>305 306</sup>.

*So far, we have been describing examples where individual alleles are strongly selected. We close this chapter by considering a different mode of adaptation that depends on the joint action of many variants across the genome.*

**Polygenic adaptation.** These examples of strong positive selection are biologically important, and illustrate essential concepts. But they represent a rather special class of selective events: all of these are variants that – on their own – exert major effects on specific traits.

The most dramatic examples of positive selection are usually associated with genes that play some critical, unique role in a selected process. For example, lactase is *the* critical enzyme involved in digestive breakdown on the main sugar in milk. Duffy is *the* critical receptor involved in vivax entry into erythrocytes. SLC24A5 is one of a handful of genes with strong impact on pigmentation and minimal pleiotropic effects.

However, this situation where a single gene plays a central role in a specific trait without major unintended consequences is the exception rather than the rule. Aside from rare genetic diseases, **most phenotypes are highly polygenic**<sup>e</sup>. The inheritance of most traits is due to thousands of variants across the genome, each with only tiny effects on the trait.

This includes most traits that vary in populations, including for example: morphometrics such as height, weight, and body shape; molecular and cellular traits such as hormone levels, lipid levels, or blood cell counts; risk for most diseases, including cardiovascular disease, diabetes, psychiatric conditions; and even behavioral traits.

For these traits, a person's expected phenotype can be modeled as a sum of thousands of pluses and minuses, depending on their alleles at every contributing variant: this is known as a **polygenic score**.

When selection acts on a polygenic trait, the effect of selection is to increase polygenic scores in the population. This occurs mainly through small shifts in allele frequencies, spread across thousands of variants<sup>307</sup>:

<sup>e</sup> We'll cover the genetics of polygenic traits in much more detail starting in Chapter 4.4.

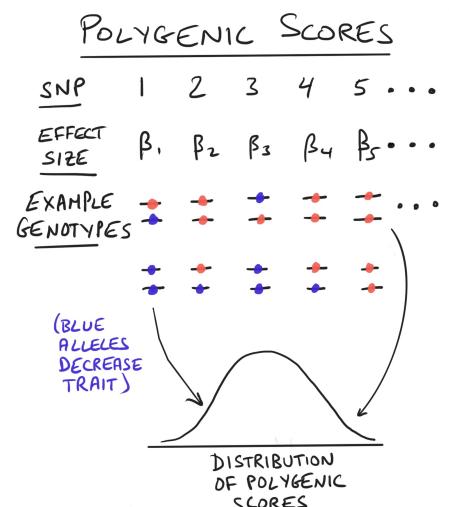
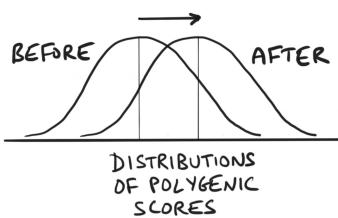
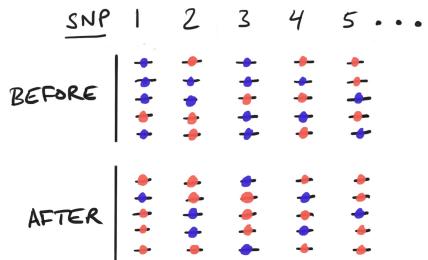


Figure 2.111: The genetics for many traits can be modeled using polygenic scores. Here, blue alleles decrease, and red alleles increase the expected phenotype by some amount  $\beta$  per allele. An individual's polygenic score is a sum across the relevant allelic effects, and predicts the genetic component of their phenotype.

### A. SELECTION ON A POLYGENIC TRAIT



### B. ADAPTATION THROUGH ALLELE FREQUENCY SHIFTS



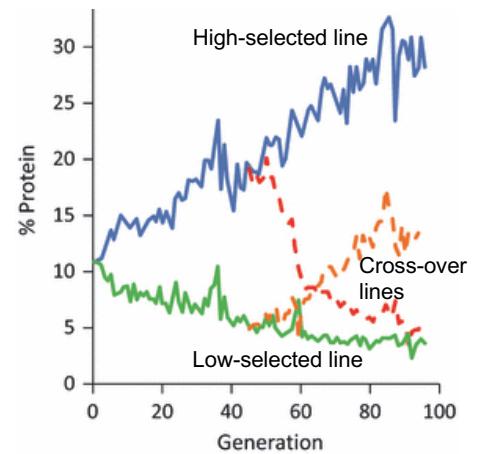
**Figure 2.112: Polygenic adaptation.** **A.** Selection on a polygenic trait drives the mean polygenic score up or down, depending on the direction of selection. **B.** The shift in polygenic scores mainly occurs through small changes in allele frequencies at thousands of sites that contribute to the trait (here, increasing the number of red alleles at each site).

One important feature of polygenic adaptation is that it proceeds extremely rapidly compared to conventional sweeps. This is because for conventional sweeps it can take hundreds of generations for a suitable, favored mutation to reach intermediate frequencies; in contrast, for a polygenic trait there is often a great deal of genetic variation present at the onset of selection (Chapter 4.4).

For this reason, polygenic adaptation is the main mechanism underlying the enormous responses to artificial selection that are commonly seen in plant and animal breeding. Farm animals including meat and dairy cattle, pigs, and chickens; as well as crop plants such as maize and soy, have undergone enormous improvements in yield due to artificial selection on polygenic traits. One example is shown at right, based on a remarkable study of maize, conducted at the University of Illinois continuously since 1896. As you can see, this study observed huge phenotypic changes within just 100 generations of artificial selection<sup>308</sup>.

It seems certain that complex phenotypes must be under a constant assault of selective pressures in one direction or another, though not necessarily in consistent directions in time and space. But despite the likely importance of polygenic adaptation as a mechanism, it has been difficult to detect clear signals in human data: the frequency shifts at most individual variants are very small and cannot be detected by traditional methods for detecting sweeps. There has been progress with alternative approaches, but these are still a work in progress<sup>309</sup>.

*Well done! In this chapter we have covered some of the main mechanisms for positive selection and adaptation, with examples. Next we examine the overall extent of different forms of selection.*



**Figure 2.113: Polygenic adaptation: the Illinois Maize experiment.** Starting in 1896, lines were selected for either high, or low, protein content. In the cross-over lines, the direction of selection was reversed partway through the experiment. Credit: Modified from Figure 4 of Andrew Hendry et al 2011. [[Link](#)]

## Notes and References.

<sup>268</sup>Fan S, Hansen ME, Lo Y, Tishkoff SA. Going global by adapting local: A review of recent human adaptation. *Science*. 2016;354(6308):54-9

<sup>269</sup>The average fixation time for a strongly selected allele is  $4\ln(2N)/s$ , compared to  $4N$  for a neutral allele: see Equation 10.30 in Coop (2020); also see simulations in Teshima and Przeworski (2006)

Coop G. *Population and Quantitative Genetics*; 2020

Teshima KM, Coop G, Przeworski M. How reliable are empirical genomic scans for selective sweeps? *Genome research*. 2006;16(6):702-12

<sup>270</sup>This term was coined in a classic 1974 paper

Maynard Smith J, Haigh J. The hitch-hiking effect of a favourable gene. *Genetics Research*. 1974;23(1):23-35

<sup>271</sup>Sabeti PC, Reich DE, Higgins JM, Levine HZ, Richter DJ, Schaffner SF, et al. Detecting recent positive selection in the human genome from haplotype structure. *Nature*. 2002;419(6909):832-7

Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS biology*. 2006;4(3):e72

<sup>272</sup>A detailed derivation is beyond our scope, but the key idea is that  $\tau$  gives the fixation time in the deterministic model, so  $\tau_{rx}$  measures the ability for recombination to chop up the region at distance  $x$  within the course of the sweep. For more on this see Coop (2020), Chapter 13. For a very nice application to detecting sweeps, and further helpful citations see

Nielsen R, Williamson S, Kim Y, Hubisz MJ, Clark AG, Bustamante C. Genomic scans for selective sweeps using SNP data. *Genome research*. 2005;15(11):1566-75.

<sup>273</sup>For example see Voight et al (2006), Fan et al 2016, and

Oleksyk TK, Smith MW, O'Brien SJ. Genome-wide scans for footprints of natural selection. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 2010;365(1537):185-205

Sabeti PC, Schaffner SF, Fry B, Lohmueller J, Varilly P, Shamovsky O, et al. Positive natural selection in the human lineage. *science*. 2006;312(5780):1614-20

Booker TR, Jackson BC, Keightley PD. Detecting positive selection in the genome. *BMC biology*. 2017;15:1-10

<sup>274</sup>Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jørgensen ME, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science*. 2015;349(6254):1343-7

Mathieson S, Mathieson I. FADS1 and the timing of human adaptation to agriculture. *Molecular biology and evolution*. 2018;35(12):2957-70

Mathieson I, Day FR, Barban N, Tropf FC, Brazel DM, eQTLGen Consortium, et al. Genome-wide analysis identifies genetic effects on reproductive success and ongoing natural selection at the FADS locus. *Nature human behaviour*. 2023;7(5):790-801

<sup>275</sup>Perry GH, Dominy NJ, Claw KG, Lee AS, Fiegler H, Redon R, et al. Diet and the evolution of human amylase gene copy number variation. *Nature Genetics*. 2007;39(10):1256-60

Usher CL, Handsaker RE, Esko T, Tukey MA, Weedon MN, Hastie AR, et al. Structural forms of the human amylase locus and their relationships to SNPs, haplotypes and obesity. *Nature Genetics*. 2015;47(8):921-5

Pajic P, Pavlidis P, Dean K, Neznanova L, Romano RA, Garneau D, et al. Independent amylase gene copy number bursts correlate with dietary preferences in mammals. *Elife*. 2019;8:e44628

<sup>276</sup>Mathieson I. Estimating time-varying selection coefficients from time series data of allele frequencies. *bioRxiv*. 2020

Segurel L, Guarino-Vignon P, Marchi N, Lafosse S, Laurent R, Bon C, et al. Why and when was lactase persistence selected for? Insights from Central Asian herders and ancient DNA. *PLoS biology*. 2020;18(6):e3000742

<sup>277</sup>Until recently it has been difficult to do similar analyses for other selected variants, or in other parts of the world, as we have less dense sampling of ancient DNA outside Europe. However, this is now changing: for an application in east Asia see

Cong PK, Bai WY, Li JC, Yang MY, Khederzadeh S, Gai SR, et al. Genomic analyses of 10,376 individuals in the Westlake BioBank for Chinese (WBBC) pilot project. *Nature Communications*. 2022;13(1):2939. Furthermore, we have little data before ~10,000 years ago, limiting the aDNA approach to sweeps that are recent.

<sup>278</sup>Bleasdale M, Richter KK, Janzen A, Brown S, Scott A, Zech J, et al. Ancient proteins provide evidence of dairy consumption in eastern Africa. *Nature communications*. 2021;12(1):632

<sup>279</sup>Tishkoff SA, Reed FA, Ranciaro A, Voight BF, Babbitt CC, Silverman JS, et al. Convergent adaptation of human lactase persistence in Africa and Europe. *Nature genetics*. 2007;39(1):31-40

Schlebusch CM, Sjödin P, Skoglund P, Jakobsson M. Stronger signal of recent selection for lactase persistence in Maasai than in Europeans. *European Journal of Human Genetics*. 2013;21(5):550-3

<sup>280</sup>Crawford NG, Kelly DE, Hansen ME, Beltrame MH, Fan S, Bowman SL, et al. Loci associated with skin pigmentation identified in African populations. *Science*. 2017;358(6365):eaan8433

<sup>281</sup>Jones P, Lucock M, Veysey M, Beckett E. The vitamin D-folate hypothesis as an evolutionary model for skin pigmentation: an update and integration of current ideas. *Nutrients*. 2018;10(5):554

Jablonski NG. The evolution of human skin pigmentation involved the interactions of genetic, environmental, and cultural variables. *Pigment Cell & Melanoma Research*. 2021;34(4):707-29

<sup>282</sup>Nielsen et al (2005),

Pickrell JK, Coop G, Novembre J, Kudaravalli S, Li JZ, Absher D, et al. Signals of recent positive selection in a worldwide sample of human populations. *Genome research*. 2009;19(5):826-37

Field Y, Boyle EA, Telis N, Gao Z, Gaulton KJ, Golan D, et al. Detection of human adaptation during the past 2000 years. *Science*. 2016;354(6313):760-4

Ju D, Mathieson I. The evolution of skin pigmentation-associated variation in West Eurasia. *Proceedings of the National Academy of Sciences*. 2021;118(1):e2009227118

<sup>283</sup>One question is why the SLC24A5 variant is not found in east Asia. It appears that the SLC24A5 variant arose after the separation of west and east Eurasian populations, and that to some extent east Asians adapted to higher latitudes via mutations in different genes.

<sup>284</sup>For reviews see e.g.,

Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Current Biology*. 2010;20(4):R208-15,

Messer PW, Petrov DA. Population genomics of rapid adaptation by soft selective sweeps. *Trends in ecology & evolution*. 2013;28(11):659-69

Hermisson J, Pennings PS. Soft sweeps and beyond: understanding the patterns and probabilities of selection footprints under rapid adaptation. *Methods in Ecology and Evolution*. 2017;8(6):700-16

and for a classic example in sticklebacks see

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, et al. The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*. 2012;484(7392):55-61.

<sup>285</sup>Orr HA, Betancourt AJ. Haldane's sieve and adaptation from the standing genetic variation. *Genetics*. 2001;157(2):875-84

Hermisson J, Pennings PS. Soft sweeps: molecular population genetics of adaptation from standing genetic variation. *Genetics*. 2005;169(4):2335-52

Przeworski M, Coop G, Wall JD. The signature of positive selection on standing genetic variation. *Evolution*. 2005;59(11):2312-23

<sup>286</sup>Langhi DM, Orlando Bordin J. Duffy blood group and malaria. *Hematology*. 2006;11(5-6):389-98

<sup>287</sup>Tournamille C, Colin Y, Cartron JP, Le Van Kim C. Disruption of a GATA motif in the Duffy gene promoter abolishes erythroid gene expression in Duffy-negative individuals. *Nature genetics*. 1995;10(2):224-8

<sup>288</sup>Spencer HC, Miller LH, Collins WE, Knud-Hansen C, McGinnis MH, Shiroishi T, et al. The Duffy blood group and resistance to Plasmodium vivax in Honduras. *The American Journal of Tropical Medicine and Hygiene*. 1978;27(4):664-70

<sup>289</sup>A similar mechanism exists for HIV, which uses the CCR5 cell surface protein to enter CD4+ T cells. Individuals who are homozygotes for the CCR5 null allele (about 1% of Europeans) are HIV resistant.

Dean M, Carrington M, Winkler C, Huttley GA, Smith MW, Allikmets R, et al. Genetic restriction of HIV-1 infection and progression to AIDS by a deletion allele of the CCR5 structural gene. *Science*. 1996;273(5283):1856-62

<sup>290</sup>Pioneering work on Duffy by Martha Hamblin and Anna Di Rienzo in 2000 and 2002 showed, surprisingly, that Duffy did not show the expected signals of a hard sweep. Instead they proposed that the two major null haplotypes likely pre-dated the onset of selection. My text relies on updated population genetic analysis, including Fst analysis and model estimates by Kimberly McManus et al (2017); Coop 2009 for genome-wide measures; estimated selection coefficient from Hodgson et al (2014):

Hamblin MT, Di Rienzo A. Detection of the signature of natural selection in humans: evidence from the Duffy blood group locus. *The American Journal of Human Genetics*. 2000;66(5):1669-79

Hamblin MT, Thompson EE, Di Rienzo A. Complex signatures of natural selection at the Duffy blood group locus. *The American Journal of Human Genetics*. 2002;70(2):369-83

McManus KF, Taravella AM, Henn BM, Bustamante CD, Sikora M, Cornejo OE. Population genetic analysis of the DARC locus (Duffy) reveals adaptation from standing variation associated with malaria resistance in humans. *PLoS genetics*. 2017;13(3):e1006560

Coop G, Pickrell JK, Novembre J, Kudaravalli S, Li J, Absher D, et al. The role of geography in human adaptation. *PLoS genetics*. 2009;5(6):e1000500

Hodgson JA, Pickrell JK, Pearson LN, Quillen EE, Prista A, Rocha J, et al. Natural selection for the Duffy-null allele in the recently admixed people of Madagascar. *Proceedings of the Royal Society B: Biological Sciences*. 2014;281(1789):20140930

<sup>291</sup>McManus et al (2017)

<sup>292</sup>Reservoir populations of *P vivax* can be found in African great apes:

Prugnolle F, Rougeron V, Becquart P, Berry A, Makanga B, Rahola N, et al. Diversity, host switching and evolution of *Plasmodium vivax* infecting African great apes. *Proceedings of the National Academy of Sciences*. 2013;110(20):8123-8

Liu W, Li Y, Shaw KS, Learn GH, Plenderleith LJ, Malenke JA, et al. African origin of the malaria parasite *Plasmodium vivax*. *Nature communications*. 2014;5(1):3346

<sup>293</sup>Globally in 2016 there were 216 million reported cases of malaria, and 445,000 deaths: [Link]

<sup>294</sup>To identify values of  $p$  for which  $\Delta_p = 0$ , with  $h$  and  $s$  fixed, we set

$$\frac{pqs[p(1-h) + qh]}{\bar{w}} = 0 \quad (2.87)$$

Noting that  $q = 1 - p$ , and assuming that  $\bar{w} > 0$  for sensible biological parameters, we see immediately that two trivial solutions are

$$\hat{p} = 0 \quad (2.88)$$

$$\hat{p} = 1. \quad (2.89)$$

Next, let's consider the cases where  $p \neq 0$  and  $p \neq 1$ . We further assume that  $s \neq 0$ . (The  $1, 1 + hs, 1 + s$  parameterization used in this book has a slight oddity in that it does not allow the heterozygote to have a fitness different than 1 if  $s = 0$ .) Then we can divide both sides by  $p, q$ , and  $s$ , and multiply by  $\bar{w}$ , to yield

$$p(1-h) + qh = 0 \quad (2.90)$$

$$p(1-h) + (1-p)h = 0 \quad (2.91)$$

$$\hat{p} = \frac{h}{2h-1} \quad (2.92)$$

Note that this equilibrium for  $p$  is outside  $[0, 1]$  and thus not relevant for an allele frequency, unless either  $h < 0$  or  $h > 1$ . This is a stable equilibrium (i.e. balanced polymorphism) if  $hs > 0$  and otherwise an unstable equilibrium.

<sup>295</sup>There are some good examples of disruptive selection in nature: for example, in the evolution of bird beak sizes

Hendry AP, Huber SK, De Leon LF, Herrel A, Podos J. Disruptive selection in a bimodal population of Darwin's finches. *Proceedings of the Royal Society B: Biological Sciences*. 2009;276(1657):753-9.

<sup>296</sup>Simons YB, Bullaughey K, Hudson RR, Sella G. A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS biology*. 2018;16(3):e2002985

Simons YB, Mostafavi H, Smith CJ, Pritchard JK, Sella G. Simple scaling laws control the genetic architectures of human complex traits. *bioRxiv*. 2022:2022-10

<sup>297</sup>Rees DC, Williams TN, Gladwin MT. Sickle-cell disease. *The Lancet*. 2010;376(9757):2018-31

<sup>298</sup>Under normal conditions, two units of the  $\beta$ -globin protein, along with two units of  $\alpha$ -globin, join together to form the hemoglobin molecule, which is responsible for carrying oxygen in red blood cells. In individuals who are homozygous for the  $\beta$ -globin mutation, especially under low oxygen conditions, their hemoglobin molecules can stick together to form polymers. This in turn leads the red blood cells to change shape from a disc-like shape to a sickle-like shape. The sickling reduces oxygen-carrying capacity, and blocks blood vessels, leading a variety of severe symptoms. In individuals who are heterozygotes, only half of the  $\beta$ -globin proteins carry the mutation, and the tendency for red blood cells to sickle is greatly reduced under normal conditions. Importantly however, infection by the malaria parasite causes low oxygenation within the cell and causes sickling specifically of the infected cells. These can then be removed by the spleen, thereby helping to clear infection. Prior to modern medicine these children had very low survival rates. In recent years, treatment options have greatly improved, giving new hope for this devastating disease, although treatment is expensive and equitable access remains highly problematic.)

<sup>299</sup>Allison AC. Protection afforded by sickle-cell trait against subtropical malarial infection. *British medical journal*. 1954;1(4857):290

<sup>300</sup>The Malaria Genomic Epidemiology Network (2014) reported a huge reduction in severe malaria among sickle heterozygotes compared to non-sickle controls (odds ratio of 0.14, p-value=10<sup>-225</sup>).

MalariaGen. Reappraisal of known malaria resistance loci in a large multicenter study. *Nature genetics*. 2014;46(11):1197-204

MalariaGen. Insights into malaria susceptibility using genome-wide data on 17,000 individuals from Africa, Asia and Oceania. *Nature communications*. 2019;10(1):5732

<sup>301</sup>Piel et al (2013) used spatial smoothing to estimate allele frequencies on global maps, as local sample sizes are often small. Their highest estimate at any location was 18% in northern Angola, but with high uncertainty, while they are more confident in estimates around 15%:

Piel FB, Patil AP, Howes RE, Nyangiri OA, Gething PW, Dewi M, et al. Global epidemiology of sickle haemoglobin in neonates: a contemporary geostatistical model-based map and population estimates. *The Lancet*. 2013;381(9861):142-51

<sup>302</sup>A recent study of selection at sickle uses a slightly lower allele frequency and concludes the following: "If we take the 21% HbAS average prevalence in Gabon, it translates to a HbS frequency  $p = 0.105$  and to a selection coefficient  $s = 0.12, \dots$  a figure comparable to that of 0.11 found by Cavalli-Sforza and Bodmer"

Elguero E, Délicat-Loembet LM, Rougeron V, Arnathau C, Roche B, Becquart P, et al. Malaria continues to select for sickle cell trait in Central Africa. *Proceedings of the National Academy of Sciences*. 2015;112(22):7051-4

<sup>303</sup>More on G6PD:

Tishkoff SA, Varkonyi R, Cahinhinan N, Abbes S, Argyropoulos G, Destro-Bisol G, et al. Haplotype diversity and linkage disequilibrium at human G6PD: recent origin of alleles that confer malarial resistance. *Science*. 2001;293(5529):455-62

<sup>304</sup>There's a similar polymorphism in old world monkeys and it's likely that the origin goes back even further, to the ancestor of apes and monkeys.

Ségurel L, Thompson EE, Flutre T, Lovstad J, Venkat A, Margulis SW, et al. The ABO blood group is a trans-species polymorphism in primates. *Proceedings of the National Academy of Sciences*. 2012;109(45):18493-8

<sup>305</sup>It has been suggested that other types of pressures, such as gut pathogen interactions may also be important in maintaining the system. For discussion of selective pressures see

Segurel L, Gao Z, Przeworski M. Ancestry runs deeper than blood: the evolutionary history of ABO points to cryptic variation of functional importance. *Bioessays*. 2013;35(10):862-7.

<sup>306</sup>For more examples of ancient balancing selection see

Leffler EM, Gao Z, Pfeifer S, Ségurel L, Auton A, Venn O, et al. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*. 2013;339(6127):1578-82

Fortier AL, Pritchard JK. Ancient Trans-Species Polymorphism at the Major Histocompatibility Complex in Primates. *bioRxiv*. 2022:2022-06

<sup>307</sup>Pritchard et al (2010);

Hayward LK, Sella G. Polygenic adaptation after a sudden change in environment. *Elife*. 2022;11:e66697

<sup>308</sup>Illinois Maize study lab website: [\[Link\]](#);

Moose SP, Dudley JW, Rocheford TR. Maize selection passes the century mark: a unique resource for 21st century genomics. *Trends in plant science*. 2004;9(7):358-64

Hendry AP, Kinnison MT, Heino M, Day T, Smith TB, Fitt G, et al. Evolutionary principles and their practical application. *Evolutionary Applications*. 2011;4(2):159-83

<sup>309</sup>Most promising, there has been interesting work on detecting polygenic shifts for specific traits, but these are still challenging to apply in practice:

Berg JJ, Harpak A, Sinnott-Armstrong N, Joergensen AM, Mostafavi H, Field Y, et al. Reduced signal for polygenic adaptation of height in UK Biobank. *Elife*. 2019;8:e39725

Sohail M, Maier RM, Ganna A, Bloemberdal A, Martin AR, Turchin MC, et al. Polygenic adaptation on height is overestimated due to uncorrected stratification in genome-wide association studies. *Elife*. 2019;8:e39702

Cox SL, Ruff CB, Maier RM, Mathieson I. Genetic contributions to variation in human stature in prehistoric Europe. *Proceedings of the National Academy of Sciences*. 2019;116(43):21484-92

Chen M, Sidore C, Akiyama M, Ishigaki K, Kamatani Y, Schlessinger D, et al. Evidence of polygenic adaptation in Sardinia at height-associated loci ascertained from the Biobank Japan. *The American Journal of Human Genetics*. 2020;107(1):60-71