

ESSAY

Unveiling recent and ongoing adaptive selection in human populations

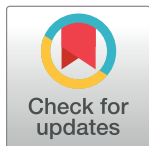
Ziyue Gao *

Department of Genetics, Perelman School of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, United States of America

* ziyuegao@penncmedicine.upenn.edu

Abstract

Genome-wide scans for signals of selection have become a routine part of the analysis of population genomic variation datasets and have resulted in compelling evidence of selection during recent human evolution. This Essay spotlights methodological innovations that have enabled the detection of selection over very recent timescales, even in contemporary human populations. By harnessing large-scale genomic and phenotypic datasets, these new methods use different strategies to uncover connections between genotype, phenotype, and fitness. This Essay outlines the rationale and key findings of each strategy, discusses challenges in interpretation, and describes opportunities to improve detection and understanding of ongoing selection in human populations.



OPEN ACCESS

Citation: Gao Z (2024) Unveiling recent and ongoing adaptive selection in human populations. *PLoS Biol* 22(1): e3002469. <https://doi.org/10.1371/journal.pbio.3002469>

Published: January 18, 2024

Copyright: © 2024 Ziyue Gao. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work is supported by a Research Fellowship (FG-2021-15702) from the Alfred P. Sloan Foundation (<https://sloan.org/>) and a grant (R35GM146810) from the National Institute of General Medical Sciences to ZG. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

Abbreviations: ASMC, ascertained sequentially Markovian coalescent; GWAS, genome-wide association study; IMR, infant mortality rate; MHC, major histocompatibility complex; PGS, polygenic score; SDS, singleton density score.

Introduction

A central query in human evolutionary genetics is to understand the functions and evolutionary history of genes or genomic regions that are under natural selection. Selection favors genetic variants that lead to advantageous phenotypic changes in specific environments, resulting in increases in allele frequency over time and distinctive patterns of genetic variation in present-day populations (Figs 1, 2A and 2B). Beyond unraveling the origin and evolutionary history of these selective genetic changes, it is of immense interest to gauge their contribution to phenotypic diversity in present-day human populations, as well as their impacts on disease risk and overall fitness (Box 1) in contemporary environments. Therefore, recent research endeavors are increasingly shifted towards identifying and characterizing extremely recent and even ongoing selection.

Numerous scans have been carried out in the human genome for targets under selection of intermediate scales (e.g., over 1,000 generations), but it remains a challenging task to demonstrate that selection on the identified targets is still ongoing or to detect selection that started recently. Enabled by the recent availability of population-scale genomic data and the development of efficient algorithms for inferring local genealogical trees, many new methods have been developed in the past 20 years to detect signals of selection from the past few millennia (e.g., [1–4]). Complementary to this approach, ancient DNA data provide direct estimates of past allele frequencies in human populations across time and geography and have refined estimation of the tempo and strength of selection in many instances of selection signals identified

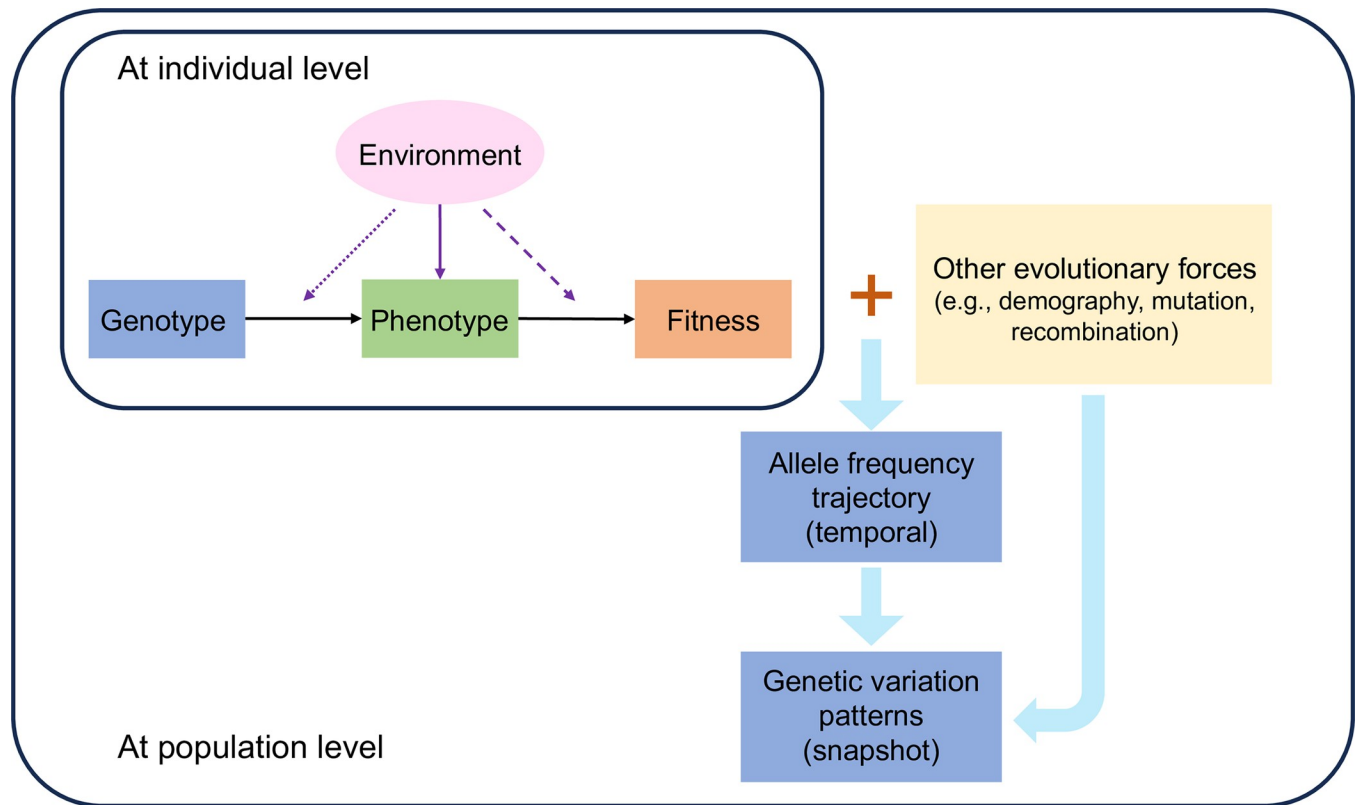


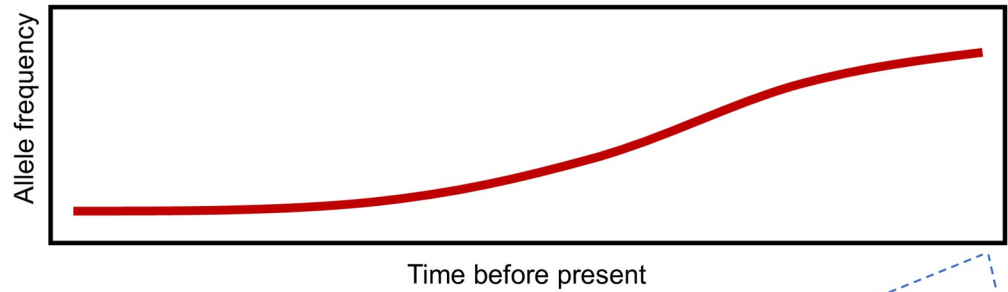
Fig 1. Overall framework of selection bridging population genetics and quantitative genetics models. In this conceptual framework, selection on genotype is mediated by fitness-relevant phenotype and manifests in allele frequency changes and genetic variation patterns. In any specific environment, genotype and environment together shape the phenotype of an individual, which in turn determines the fitness. In addition to its direct effect on the phenotype (solid purple arrow), the environment also modifies the genotype-to-phenotype mapping (i.e., genotype-by-environment interaction; indicated by the dotted purple arrow) and phenotype-to-fitness mapping (dashed purple arrow). Through interactions with other evolutionary forces (indicated by the brown plus sign), natural selection shapes the allele frequency trajectory over time and leaves footprints in genomic variation in present-day populations.

<https://doi.org/10.1371/journal.pbio.3002469.g001>

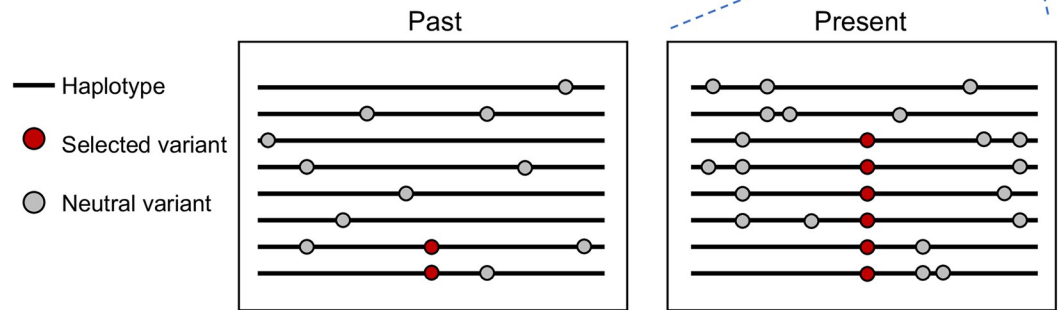
in modern genomes. Most recently, population-scale biobank-style datasets, encompassing genomic information and phenotypic data on reproduction, disease, mortality, and other quantitative traits, have pinpointed variants associated with various fitness components, at times in a sex-specific manner. These findings signify the presence of ongoing selection occurring within just one or a few generations.

This Essay aims to highlight growing evidence for very recent and ongoing genetic adaptation in the human genome, with a focus on positive selection and directional selection on polygenic traits, as these modes of selection may potentially contribute to genetic and phenotypic differences across populations. It is important to note that the effects of negative selection (such as purifying selection and background selection; [Box 1](#)) are evident and prevalent in the human genome. However, due to space limitations, this Essay does not discuss the advances made in the past decade in identifying genomic regions and phenotypes subject to recent and ongoing negative and stabilizing selection (e.g., [5–8]). Instead, it only briefly discusses the challenges associated with detecting and interpreting signals of positive and directional selection in the context of pervasive negative selection. The Essay starts with the latest methodological innovations in inference of positive selection at individual genomic loci, and then discusses techniques for detecting aggregate selection signals across genetic loci that collectively influence a quantitative trait. Rather than delving deeply into the technical details, it emphasizes

A. Allele frequency trajectory



B. Genetic variation



C. Methods for selection inference

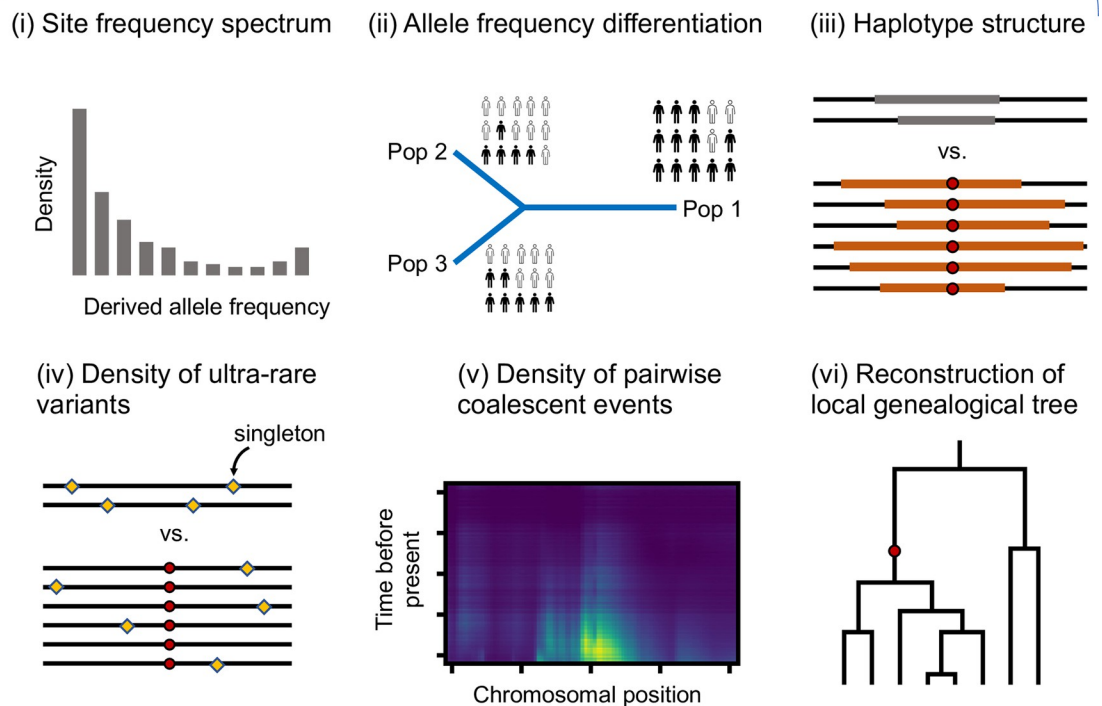


Fig 2. Signals of recent positive selection in genetic variation and corresponding methods for selection inference. (A) The hallmark of positive selection is faster allele frequency increase than would be expected under neutrality. (B) The rapid allele frequency change leaves footprints in the surrounding genomic region, although the specific patterns depend on the strength, tempo, and mode of selection (e.g., selection on standing variation versus on de novo variants). (C) Major methods for detecting positive selection based on present-day genetic variation.

<https://doi.org/10.1371/journal.pbio.3002469.g002>

Box 1. Glossary

Fitness

A measure of how well an individual can survive or reproduce; it consists of multiple components such as viability, mating success, and fecundity.

Positive selection

An evolutionary process in which a genetic variant becomes more common in a population because it increases the fitness of individuals who carry it.

Negative selection

An evolutionary process that weeds out fitness-reducing genetic variants from the population. Purifying selection acts directly on the deleterious variants, whereas background selection affects nearby variants linked to the deleterious variants.

Positive and negative selection

Two inseparable concepts that describe the same phenomenon from different angles. To facilitate communication, population geneticists often adopt either of these terms focusing on the impact of selection on the derived allele, such that positive selection tends to speed up molecular evolution, whereas negative selection decelerates or prevents it. Nonetheless, in many cases, identity of the derived allele is ambiguous or less relevant (e.g., during transient selection), and the direction of selection often refers to the effect of selection on the rare allele (for example, a scenario where the rare allele is beneficial is often considered positive selection, although one could consider the same scenario as negative selection against the more common allele).

Genetic adaptation

The process by which organisms evolve heritable characteristics or traits that help them to better survive and reproduce in their specific environment. In many cases, adaptation is used synonymously with positive selection, but adaptation also encompasses other selection modes such as balancing selection and polygenic adaptation.

Stabilizing selection

A type of natural selection that favors individuals with an intermediate value of a fitness-relevant trait. Individuals with deviation from the optimal trait value are selected against, and the result is a stabilization of the trait around a specific value. Stabilizing selection concerns the relationship between phenotype and fitness, regardless of the genetic basis. Other types of phenotype-focused selection include disruptive selection, which favors individuals with extreme trait values, and directional selection, which favors individuals at only one end of the phenotypic spectrum.

Polygenicity

Polygenicity refers to a scenario in which variation in a trait within a population is contributed to by genetic variants at multiple genes or genomic loci rather than by just one

or a few. Many complex traits in humans, such as height and disease susceptibility, are highly polygenic.

Pleiotropy

Pleiotropy occurs when a single genetic variant (or gene) influences two or more seemingly unrelated phenotypes in an organism. Two traits are pleiotropically related when certain variants exist that simultaneously affect them.

the connection and distinction among “genotype-focused,” “phenotype-focused,” and “fitness-focused” strategies, as well as the advantages and limitations of each (Fig 3). Some major findings stemming from these innovative approaches are discussed, along with challenges in interpretation of the signals.

Positive selection at individual genomic loci

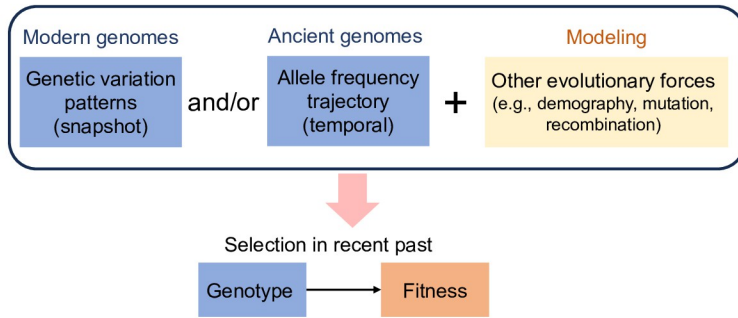
Genomic footprints in present-day genetic variation

Traditional methods for detecting selection take a genotype-focused approach (Fig 3A) by adopting classic population genetics models. Specifically, these models predict changes in allele frequency and patterns of surrounding genomic variation by assuming arbitrary fitness effects of different genotypes at a single genetic locus. The obvious advantage of this modeling approach is that it establishes expectations for genomic signatures of selection while requiring very little phenotypic information, such as how genotypes map to phenotypes or which phenotypes are under selective pressure.

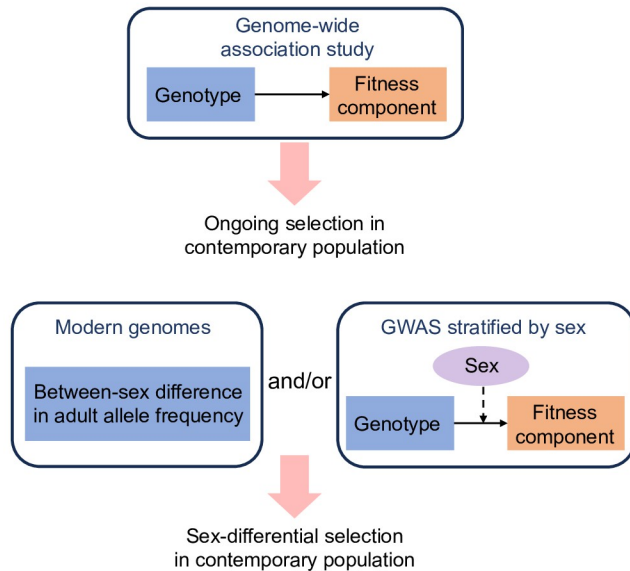
Typical genomic signatures of positive selection include extreme differentiation in allele frequencies across populations, extended haplotypes/linkage disequilibrium, or distortion in the site frequency spectrum of segregating variants (reviewed in [9–11]; Fig 2C(i–iii)). These statistics capture complementary features of genomic variation, but most are powerful in detecting selection on intermediate timescales (i.e., hundreds of generations or longer). More recent methods increase detection power by considering multiple summary statistics jointly. This idea was initially implemented using a few basic summary statistics [12] and later expanded through techniques such as Approximate Bayesian Computation [13] or supervised machine learning (reviewed in [14]). Thanks to the recently available population-scale genomic data and continuous theoretical and methodological developments, genome-wide scans based on population genetic summary statistics have identified thousands of putative targets under selection, largely independently of biological knowledge regarding the corresponding phenotype or selective pressure.

Despite being able to pick up selection signals over the past hundreds or thousands of generations, these scans are limited in power for detecting very recent selection because the narrow time window involved leaves very subtle genetic footprints in the site frequency spectrum or haplotype structure. From the perspective of the local genealogical tree, very recent selection only impacts branches near the leaf nodes but leaves most of the tree unchanged. Realizing this, researchers have developed methods that explicitly leverage features of terminal branches of the local genealogical tree. The singleton density score (SDS) is one such method that detects recent allele frequency changes based on extremely rare variants [15]. Specifically, SDS tests for deficiency of singletons (i.e., variants that appear exactly once in the entire sample) on haplotypes carrying the putatively favored allele, which is indicative of a faster coalescent rate in the recent past (Fig 2C(iv)). Along these lines, another method called ascertained

A. Genotype-focused strategy



B. Fitness-focused strategy



C. Phenotype-focused strategy

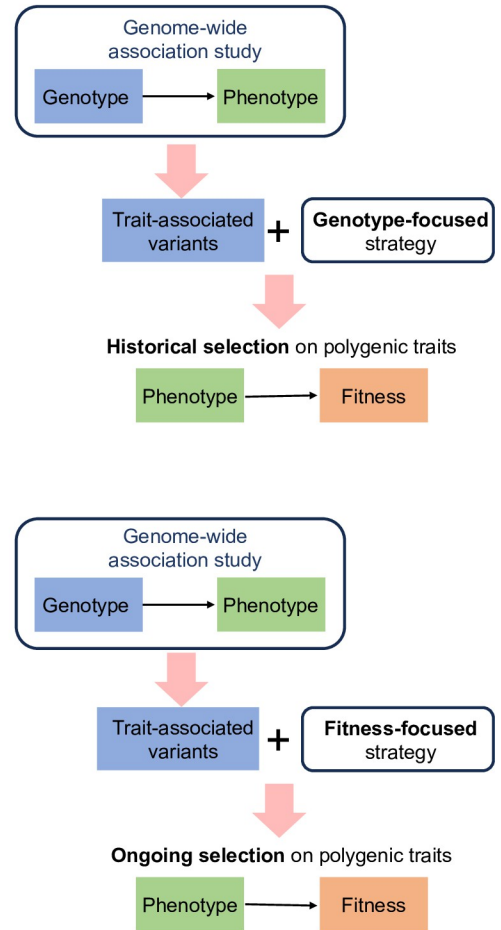


Fig 3. Common strategies for detecting signatures of recent or ongoing selection. (A) A “genotype-focused” strategy focuses on the cumulative effects of historical selection on genetic variation patterns and relies on population genetics modeling to tease apart the influence of other evolutionary forces. Ancient DNA data provide direct information on allele frequency changes, which helps reduce inference uncertainty and confounding by demographic history. (B) A “fitness-focused” strategy focuses on direct association between genotype and fitness component(s) and utilizes allele frequency changes within one generation to detect selection in contemporary populations. As a special case of this strategy, between-sex differences in adult allele frequency or effect size of association to fitness components can be leveraged to detect sex-differential selection. (C) A “phenotype-focused” strategy relies on aggregation of selection signals revealed by genotype-focused or fitness-focused strategies across trait-associated variants identified by genome-wide association studies (GWAS).

<https://doi.org/10.1371/journal.pbio.3002469.g003>

sequentially Markovian coalescent (ASMC) detects targets of recent positive selection by inferring pairwise coalescent times and looking for unusually high densities of coalescent events in the recent past (Fig 2C(v)) [16,17]. When applied to whole-genome sequences of approximately 3,200 individuals of European ancestry, SDS detected selection signals in the past 2,000 to 3,000 years in the major histocompatibility complex (MHC) region and at variants associated with lactose tolerance and pigmentation [15]. In comparison, application of ASMC to over 487,000 British individuals identified signals of selection in the past 1,500 years, including those detected by SDS, as well as several new candidate loci harboring genes related to immune response, tumor growth, and other phenotypes [17].

With the recent development of algorithms for inference of the ancestral recombination graph or its proxies, several tree-based statistics have been developed for detecting positive selection (reviewed in [18]; Fig 2C(vi)). One of these methods, Relate, estimates local genealogy from sequence data and detects selection by searching for rapid propagation of lineages carrying a putatively beneficial allele relative to other lineages, effectively testing for differences in the coalescent rate between haplotypes carrying different alleles [19]. However, this selection metric is calculated on only one point estimate of the local genealogy. By contrast, a likelihood method called CLUES leverages the posterior distribution of local genealogical trees to infer selection coefficients and allele frequency trajectories at individual loci [20]. These new methods have confirmed strong selection on variants associated with lactase persistence, immune response, and pigmentation traits in Europeans in the past few thousand years and some signals in other populations (such as the *EDAR* gene in East Asians), although very few new signals have been detected.

Selection signals in ancient genomes

While modern genomes provide a snapshot of population evolution and allow for indirect inference of past demographic and selective events, genomic sequences from ancient samples enable direct glimpses into the genetic history of human populations. By providing estimates of allele frequencies at multiple time points (Fig 2A and 2B), ancient DNA has shed valuable insights on the evolutionary histories of multiple selected variants in human evolution during the past 15,000 years (reviewed in [21–23]). Analysis based on ancient DNA has also been particularly helpful in detecting candidates under spatially or temporally restricted selection.

Ancient DNA transformed our understanding of selection in humans by resolving complex interactions between selection and demographic history. As recent human history features many episodes of population splits and admixture, signals of selection are often obscured by changes in ancestry [24]. One instance is the evolutionary history of the *FADS* locus, which contains genes encoding enzymes involved in the conversion of long-chain polyunsaturated fatty acids. Using present-day genomic data, studies detected strong selection signals on *FADS* genes in human populations from multiple continents, with different alleles being favored across time and geography [25–29]. However, analysis of ancient DNA showed that the selection signal in Native Americans was largely an artifact driven by parallel selection in European and Asian populations [30]. Another intriguing case is the evolution of pigmentation in west Eurasia in the context of several major admixture events revealed by ancient DNA. The derived alleles associated with lighter skin or eye color at several pigmentation-associated genes exhibited distinct frequencies in different ancestral populations, potentially reflecting differential selective pressures across geography prior to the Mesolithic period (i.e., before 9,000 to 10,000 years ago) [31,32]. Moreover, the observed allele frequencies and ancestry fractions at these pigmentation-associated variants in later admixed populations significantly deviated from neutral expectations, suggesting subsequent selection during the Neolithic, Bronze Age, and historical periods [33–35]. These findings point to continued selective pressure for light pigmentation over the past 2,000 years in west Eurasia and support the concept that admixture may facilitate rapid adaptation by introducing advantageous alleles [34–37].

Ancient DNA data have also refined our knowledge of the onset, duration, and strength of selection events. For example, selection on the variant conferring lactase persistence was initially estimated to begin around 7,500 years ago based on modern genomic data and archeological evidence of dairy production [38]. Surprisingly, ancient DNA data have shown that the selected allele was rare in Bronze Age Europe until 3,000 years ago, suggesting a much later onset of positive selection than was previously inferred [31]. In addition, based on the allele

frequency trajectory in ancient DNA samples, the positive selection for this allele was inferred to be strong 100 to 150 generations ago but drastically reduced in the past 100 generations [39]. Significant variation in selection strength has also been found at several other previously identified selected loci [39]. Overall, ancient DNA studies have confirmed selection signals near multiple genes associated with diet, pigmentation, and immune response revealed in modern genomic data, and have provided fine-resolution insights into the temporal dynamics and geographic distribution of the selected variants and the corresponding selection strengths [26,34,39,40].

With recurrent observations of selection targeting genes in immune pathways, the quest to discern the specific pathogens driving these selective pressures has been immensely captivating. A strategy to link selection signals with the causative pathogens is to search for variants with unusual allele frequency changes during well-documented catastrophic pandemics. A recent investigation scrutinized ancient genomes of roughly 200 individuals who died before, during, and after the Black Death pandemic in the fourteenth century [41]. This study reported an overall enrichment of allele frequency differentiation in immune genes as well as a handful of potential targets under positive selection. However, serious skepticism has been raised towards the findings due to technical concerns [42], and other studies adopting similar designs (though with smaller sample sizes) failed to replicate the selection signals at immune genes overall or at individual candidates [43,44]. These results suggest the selection effects of historical pandemics at individual genomic loci are relatively modest, necessitating expansive sample sizes for detection.

Fitness-focused strategy for detecting selection in contemporary populations

The fitness of an individual consists of several components such as viability, mating success, and fecundity. A genetic variant that influences any of these components is subject to natural selection unless its effects on all components cancel out. Based on this reasoning, one can identify loci under ongoing selection using a fitness-focused approach by performing GWAS on proxies for fitness components (Fig 3B). However, traits closely associated with fitness are expected to have low heritability [45], and fitness-related variants tend to be rare in frequency. Therefore, identification of these variants via association requires exceedingly large sample sizes, which only became feasible in the past decade. It is worth noting that, due to limited power, this association approach is biased towards detecting common variants and does not pick up fitness-influencing variants that are under strong negative selection.

One of the most studied proxies of fertility is the number of children ever born to or fathered by an individual, because it can be easily surveyed and approximates the overall fitness well in modern populations with low mortality. Using data from hundreds of thousands of individuals born in the 1950s to 1970s, dozens of genomic loci have been associated with the number of children [46–48]. Interestingly, among the top associations stands the *FADS* locus, which also harbors strong signals of historical positive selection in both ancient and present-day DNA samples [26,28,29,49]. By contrast, the two most significant association regions lack evidence of historical positive selection but demonstrate signals of balancing selection, possibly due to pleiotropic effects (Box 1) on other fitness components or temporally fluctuating selection [46,50,51].

Besides reproduction, viability is a key component of fitness. In principle, the number of children closely reflects their contribution to the population gene pool of the next generation, but current association studies for this trait include only individuals who survived to completion of their reproductive lifespan, leaving out those who did not reach adulthood. To detect common variants linked to early-life survival, Wu and colleagues performed a clever GWAS

on time- and location-matched infant mortality rate (IMR) for living individuals in the UK Biobank [52]. The rationale is that individuals who survived in tougher environments during infancy, as indexed by a higher local IMR in their birth years, tend have higher “relative viability.” Interestingly, the two genome-wide significant loci identified by this approach, *LCT* and *TLR6-TLR1-TLR10*, are both known targets of recent positive selection in Europeans, with the survival-increasing alleles matching the evolutionarily favored allele [15,26].

A more direct approach for identifying variants that affect viability is by looking for shifts in allele frequency across individuals of different ages [2]. Limited by the age distribution of participating individuals in current cohorts, this method is underpowered to detect allele frequency changes in early life, when selective pressure is expected to be strong. However, in humans, even variants that exclusively affect viability late in life may be under selection, due to late male reproduction, intergenerational resource transfer, and other reasons [53,54]. By testing for changes in allele frequency with age, a study found and replicated two genome-wide significant signals in 2 independent datasets: one overlaps with the *APOE* $\epsilon 4$ allele that is associated with reduced lifespan and increased risk of Alzheimer’s disease and cardiovascular diseases [55,56]; the other locus contains variants that are close to a nicotine receptor gene *CHRNA3* and associated with increased smoking quantity [57]. Intriguingly, the relatively common frequencies of these survival-reducing variants in present-day populations suggest that they were not under strong negative selection in the recent past. The authors interpreted the lack of abundant associations as evidence for purifying selection against variants with large effects on late-onset disease and speculated that the *APOE* and *CHRNA3* loci were found because their deleterious effects have recently increased in humans due to environmental changes.

Fitness-focused strategy for detecting sex-differential selection

The extraordinary level of sexual dimorphism in many animal species, including humans, reflects sex-specific phenotypic effects and sex differences in the fitness landscape. The fitness effect of a genetic variant may differ between sexes in magnitude or sometimes in direction. Such sex-differential selection is challenging to study because mendelian inheritance equalizes autosomal allele frequencies between the 2 sexes at fertilization in each generation. Nevertheless, the special case of sex-differential selection on viability is expected to leave a distinctive signature in population genetic variation: allele frequency differences between adult females and males (Fig 3B, right). An early study seeking this signature reported signals at hundreds of genetic regions and an enrichment of signals on the X chromosome compared to autosomes [58]. Unfortunately, these findings turned out to be largely false positives driven by random noise, sex-biased genotyping error, and biases due to hemizyosity of the X chromosome in males. Later studies on much larger biobank datasets failed to detect robust signals at any autosomal loci [59] or enrichment on the X chromosome [60].

While signals of sex-differential viability selection are expected to be exceptionally weak at individual loci [61,62], subtle between-sex allele frequency differences across many variants may be detectable in aggregation. Leveraging the genomic and reproductive history data of approximately 250,000 adults in the UK Biobank, Ruzicka and colleagues developed new metrics to measure between-sex allele frequency differentiation over different stages of a life cycle. They found significant shifts in the genome-wide distributions of these metrics, which is consistent with effects of sex-differential selection on survival, reproductive success, and overall fitness [4].

Limitations of the fitness-focused strategy for selection detection

One curious observation from the studies described above is the limited overlap between fitness-associated variants in contemporary populations and targets under historical positive

selection. As the 2 approaches (i.e., fitness-focused and genotype-focused) capture selection signals of very different timescales, one explanation is a highly dynamic selection landscape during recent human evolution. However, the fitness-associated variants identified in biobank-style datasets need to be taken with a grain of salt for several technical reasons.

First, the effect measured by association likely does not reflect the actual fitness effect. Fitness effects that are “visible” to natural selection may be too subtle to be picked up by association studies given current sample sizes, so many targets of ongoing selection might be missed. On the other hand, proxy traits only capture certain aspects of fitness components, so the measured effect of a variant may be greater than its effect on overall fitness in the presence of antagonistic pleiotropy (Box 1). In other words, there may be weaker or even no ongoing positive selection on variants with opposite effects on different fitness components.

Second, as for all GWAS in general, uncorrected population stratification remains a concern for fitness-associated variants, especially for those with highly differentiated frequencies across populations. For example, the lactase persistence variant near *LCT*, the top selection target identified by the IMR GWAS, is among the most differentiated variants across European populations [63]. Despite the authors’ best effort in correcting for population structure, it is still possible that the IMR association signal in UK Biobank data is driven by residual stratification, so the claim of ongoing selection on this variant remains to be validated in independent datasets or by family-based approaches [64].

A related yet different issue applies to analysis based on allele frequency differences between sex. In addition to sex-biased viability selection, between-sex allele frequency differences can also be interpreted as the result of subtly different population structures between sexes or sex-biased participation [65]. The UK Biobank requires active participation, and the participants are not representative of the general population in various sociodemographic and health-related characteristics [66]. Should a genetic variant affect participation inclination in men and women differentially, subtle allele frequency difference between sexes is expected. Consistent with this hypothesis, a “GWAS of sex” performed in 5 biobank-style datasets found significant positive autosomal single-nucleotide polymorphism heritability in those that require active participation (including UK Biobank) but not those with relatively passive recruitment, although this contrast is confounded by differences in sample size across datasets [65]. Therefore, an important future step will be to replicate the findings in more population-representative datasets or family-based studies to rule out or quantify the contribution of sex-differential participation bias.

Directional selection on quantitative traits

Integration of GWAS results with genetic variation patterns

GWAS have provided unprecedented insights into the genetic architecture of human phenotypes, revealing significant heritability and high polygenicity (Box 1) of most traits, as well as unexpectedly small effect sizes for most associated variants. These observations are surprisingly close to the assumptions of classical quantitative genetics models [67]. In the context of adaptation, the measurable heritability means that at least a portion of the phenotypic variation within a population is attributed to existing genetic polymorphisms, which, in response to changes in selective pressure on the phenotype, offer the materials for genetic adaptation without having to await new mutations. In turn, the high polygenicity and tiny effect sizes of most variants suggest that the selective pressure on any individual alleles may be too small to leave discernible genomic footprints but may be detectable in aggregate. These considerations point to the importance of examining polygenic signals of selection on traits during human evolution via a phenotype-focused strategy (Fig 3C) [68].

If all or most trait-influencing variants can be identified in an unbiased manner, signals at these loci can be interrogated jointly to uncover selection on the trait. The most straightforward idea for detecting polygenic adaptation is to directly combine GWAS results and population genetic summary statistics (e.g., some in Fig 2C) [15,19,33–35,69]. Common approaches include tests for shifts in distribution of single-locus summary statistics indicative of selection (e.g., F_{st}) at GWAS hits [69] or correlation between GWAS summary statistics (such as effect direction, magnitude, and significance level) and population genetic summary statistics. This approach has been applied to both present-day and ancient DNA data, and several studies explicitly leveraged population admixture events in recent human history to gain insights into the timing of selection [15,19,33–35,70]. Overall, these studies found consistent evidence of selection on variants underlying anthropometric, pigmentation, and immune-related trait variation in human populations in the past 10,000 years.

Rooted in the classic quantitative genetics model, more direct tests for polygenic adaptation have been devised around the concept of “genotypic value” (also known as the “breeding value” in quantitative genetics when nonadditive genetic effects are ignored) that describes the total contribution of all genetic variants of an individual to their phenotypic value. The polygenic score (PGS)—the sum of allele effect sizes across all independent GWAS loci—provides a proxy for the genotypic value that can be applied at the individual or population level. In addition to empirical comparison of observed PGS to a null distribution based on sets of matching variants [71,72], formal tests for polygenic signals of selection on quantitative traits have been developed in the population genetics framework [73,74]. In a way, these tests are analogous to tests for single-locus selection, but instead of rapid change or differentiation of allele frequency, signals for polygenic adaptation come from unexpected changes or overdispersion of PGS in the history of one or multiple populations [73,74].

GWAS results have also been explicitly incorporated into the coalescent framework. By combining GWAS effect sizes and inferred local genealogical tree at GWAS loci, Edge and Coop developed methods for reconstructing the trajectory of population-mean PGS over time [75]. They applied these methods to test polygenic signals of selection for increased height in the British population but only found very weak signals concordant with prior reports [15,70,73,74]. Taking a different approach, Stern and colleagues extended their method CLUES to estimate selection intensity on a polygenic trait by considering the allele frequency trajectories of GWAS loci conditional on the inferred local coalescent trees [76]. Contrary to the conclusion of prior studies, this method detected no signal of recent directional selection on height or body mass index, but replicated some other traits previously reported to be under recent selection, such as pigmentation traits, age at first birth, glycosylated hemoglobin, and educational attainment. By combining theory of quantitative genetics and population genetics and incorporating empirical GWAS findings, these new methods unveiled many signals of selection on quantitative traits during recent human evolution and are paving the way for many more future findings.

Correlation between phenotypes and fitness components

Analogous to the fitness-focused approach for detecting ongoing selection at individual loci, selection effects on a polygenic trait can be estimated based on phenotypic or genetic correlations between the trait and a proxy of fitness component [77]. Approaches include regression of a measure of reproductive success on PGSs for traits of interest [78,79] or estimation of genetic correlation between traits of interest and proxies for fitness [80,81]. Partially consistent with previous epidemiological studies, these studies found selection in contemporary human populations for genetic variants underlying earlier age at first birth and shorter stature in

females, as well as for those underlying increased body mass index and reduced educational attainment in both sexes.

Hypothesizing that variants influencing polygenic traits may be under sexually antagonistic selection on viability, Zhu and colleagues developed a test based on between-sex allele frequency differences and sex-specific phenotypic effect sizes from GWAS. They found suggestive signals of selection on testosterone levels [82], which is consistent with the recent findings of positive correlation between testosterone level and mortality in females and an inverse relationship in males [83]. Nonetheless, because the model makes some strong assumptions, such as allele frequency under equilibrium and selection coefficient proportional to phenotypic effect size, it remains questionable whether the detected signal is specific to sexually antagonistic selection or can also reflect the effects of other evolutionary processes.

Challenges in validating and interpreting polygenic signals of selection on quantitative traits

Despite significant progress in detecting polygenic adaptation in the past decade, serious concerns quickly emerged regarding the validity and interpretation of the reported signals of polygenic adaptation, for both technical and conceptual reasons [84]. First, technical biases in GWAS may lead to false positive signals or biased effect size estimates at individual loci. For example, the strong signals of selection on height in Europeans were found to largely result from uncorrected population stratification and weakened considerably with effect size estimates from GWAS of less-structured samples [75,85,86]. The inherent ascertainment bias and limited portability of GWAS results cast additional uncertainty on the reliability of selection signals when applying GWAS summary statistics from a study group to selection tests in other groups [84,87,88]. Furthermore, although being intuitive and powerful for combining information across sites, PGSs, especially those constructed with variants that do not reach genome-wide significance, further exacerbate biases of GWAS results due to residual population stratification [89].

Moreover, most current methods fundamentally test for deviation from neutrality (i.e., no selection on any trait or variant at all), so the detected signals may reflect effects of other modes of selection. Despite the debate on the prevalence of polygenic adaptation, there is a consensus that GWAS variants with large effect sizes are under negative selection, indicated by the strong negative correlation between variant effect size and minor allele frequency (beyond the expectation under detection bias) [90–92]. This phenomenon is consistent with the action of stabilizing selection (Box 1) on quantitative traits: For a population centered around the phenotypic optimum, mutations that affect fitness-relevant phenotypes tend to shift the population away from the optimum and thus be deleterious [93]. The prevalence of stabilizing selection leads to challenges in detection and interpretation of population differences in PGS. First, under stabilizing selection, adaptive genetic changes do not always mirror shifts in the phenotypic optimum. Environmental changes can alter not only the optimal trait level (Fig 1; dashed purple arrow) but also the mean environmental contribution to the phenotype (Fig 1; solid purple arrow), which induces “genetic compensation” in the opposite direction [84,94]. Second, although stabilizing selection around the same trait optimum constrains phenotypic differentiation between populations, it accelerates genetic differentiation at trait-influencing loci. This counterintuitive effect of stabilizing selection, combined with incomplete and biased ascertainment of GWAS loci, inflates differences in PGS between populations and may even yield spurious signals of polygenic adaptation [95]. These considerations underscore the importance of regarding stabilizing selection (with a constant trait optimum throughout time and space) as a null model for devising and interpreting tests for polygenic adaptation, especially those reliant on inter-population comparisons.

Even when the selection signals are technically sound and effects of stabilizing selection are adequately considered, it remains a formidable challenge to tell which traits are directly under selection, given the prevalent pleiotropy (Box 1) across human complex traits [96–98]. Aware of this issue, researchers developed methods that aim to disentangle effects of selection on genetically correlated traits and found evidence of indirect selection (e.g., signals of selection on educational attainment due to selection on other traits) and opposing selection (e.g., selection for increased type 2 diabetes and decreased glycosylated hemoglobin), which helps with the rejection of the hypothesis that a certain trait is under direct selection [76]. Yet, this study only tested for correlated response of 137 pairs of traits and may have missed signals driven by multi-way pleiotropy or unmeasured traits [96]. In other words, given current data and methods, one can at best conclude selection on variants associated with certain trait(s) but not selection on the trait(s).

Conclusion and future directions

Rapid growth in genomic datasets and advances in computational techniques have enabled identification of parts of the human genome under very recent or ongoing adaptive selection. Early genome-wide selection scans relied on the cumulative effects of selection over relatively long timescales, but statistical innovations have enabled efficient computation using large numbers of modern genomes to study selection over narrower time frames. The utilization of ancient DNA data has further reduced inference uncertainty and confounding due to demographic history, providing valuable insights into the temporal dynamics and geographic distribution of the selected variants. We now have lists of candidate targets with compelling evidence of selection during the past 15,000 years, along with partial information regarding variation in the selection strength.

As selection on genotypes is mediated by differences in fitness-relevant phenotypes (Fig 1), a complete understanding of selective events involves not only the causal variants but also the relevant phenotypes and selective forces [99]. Integrating rich phenotypic data with genomic information in population-scale datasets has facilitated the establishment of associations between variants and phenotypes. Following numerous association studies conducted for both organismal and molecular phenotypes, it is increasingly clear that pleiotropy is widespread across human traits [97,98,100]. It is possible that many inferred selective variants will be associated with multiple phenotypes in future GWAS, so the new questions will become: which of these phenotypes, if any, is mediating selection; where does the selection pressure come from; and is selection still ongoing in present-day populations?

The expanding biobank datasets will be pivotal in addressing these questions. First, they offer an opportunity to directly identify individual or groups of variants associated with fitness components. The partial overlap between fitness-associated variants and those targeted by historical positive selection may arise from limited power to detect subtle fitness effects, antagonistic effects on different fitness components (and/or between sex), or spatial or temporal variation in fitness effects. With the anticipation of long-term longitudinal data, possibly spanning from birth to death, becoming available in the next few decades, it will be possible to develop new statistics that better approximate various fitness components and integrate them throughout a complete life cycle, thus enhancing power to identify variants that influence the overall fitness. It is important to note that since such discoveries are associations in nature, replication in additional biobank datasets or by family-based studies will be crucial.

Second, the rich phenotype data, coupled with theoretical advancements, can potentially distinguish between traits directly or indirectly under selection. Although it remains uncertain which and how many pleiotropically related traits collectively shape the fitness landscape,

emerging evidence suggests that, at least for some traits, a model featuring many traits under stabilizing selection aligns well the empirical GWAS results [3]. These considerations strongly advocate for incorporating pleiotropy alongside stabilizing selection in future models and simulations that characterize genetic signatures of polygenic adaptation [101,102]. Findings from such models, combined with variant-level pleiotropic effect size estimates from empirical association studies, may unveil clearer adaptation signals and help differentiate between traits directly or indirectly influenced by selection.

Lastly, given the emerging evidence of sex differences in phenotypic and fitness effects of the same variant [4,82], along with varying prediction accuracy of PGSs across different contexts (e.g., age, sex, income level) [88], more context-dependent effects will likely be unmasked. These findings may imply gene-by-environment interactions on phenotype and fitness, hinting at the environmental conditions that exert selective pressure. This information, when combined with archeological data about past environments, diets, and lifestyles of human populations, may aid in rejecting and formulating new hypotheses regarding recent selective forces that have shaped the human genomic and phenotypic variation.

Acknowledgments

Thanks to Iain Mathieson for helpful discussion and critical feedback on the manuscript.

Author Contributions

Conceptualization: Ziyue Gao.

Funding acquisition: Ziyue Gao.

Visualization: Ziyue Gao.

Writing – original draft: Ziyue Gao.

Writing – review & editing: Ziyue Gao.

References

1. Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection in the human genome. *PLoS Biol.* 2006; 4(3):e72. Epub 20060307. <https://doi.org/10.1371/journal.pbio.0040072> PMID: 16494531; PubMed Central PMCID: PMC1382018.
2. Mostafavi H, Berisa T, Day FR, Perry JRB, Przeworski M, Pickrell JK. Identifying genetic variants that affect viability in large cohorts. *PLoS Biol.* 2017; 15(9):e2002458. Epub 20170905. <https://doi.org/10.1371/journal.pbio.2002458> PMID: 28873088; PubMed Central PMCID: PMC5584811.
3. Simons YB, Bullaughey K, Hudson RR, Sella G. A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biol.* 2018; 16(3):e2002985. Epub 20180316. <https://doi.org/10.1371/journal.pbio.2002985> PMID: 29547617; PubMed Central PMCID: PMC5871013.
4. Ruzicka F, Holman L, Connallon T. Polygenic signals of sex differences in selection in humans from the UK Biobank. *PLoS Biol.* 2022; 20(9):e3001768. Epub 20220906. <https://doi.org/10.1371/journal.pbio.3001768> PMID: 36067235; PubMed Central PMCID: PMC9481184.
5. Cassa CA, Weghorn D, Balick DJ, Jordan DM, Nusinow D, Samocha KE, et al. Estimating the selective effects of heterozygous protein-truncating variants from human exome data. *Nat Genet.* 2017; 49(5):806–10. Epub 20170403. <https://doi.org/10.1038/ng.3831> PMID: 28369035; PubMed Central PMCID: PMC5618255.
6. Karczewski KJ, Francioli LC, Tiao G, Cummings BB, Alfoldi J, Wang Q, et al. The mutational constraint spectrum quantified from variation in 141,456 humans. *Nature.* 2020; 581(7809):434–43. Epub 20200527. <https://doi.org/10.1038/s41586-020-2308-7> PMID: 32461654; PubMed Central PMCID: PMC7334197.
7. Gardner EJ, Neville MDC, Samocha KE, Barclay K, Kolk M, Niemi MEK, et al. Reduced reproductive success is associated with selective constraint on human genes. *Nature.* 2022; 603(7903):858–63. Epub 20220323. <https://doi.org/10.1038/s41586-022-04549-9> PMID: 35322230.

8. Chen S, Francioli LC, Goodrich JK, Collins RL, Kanai M, Wang Q, et al. A genome-wide mutational constraint map quantified from variation in 76,156 human genomes. *bioRxiv*. 2022. <https://doi.org/10.1101/2022.03.20.485034>
9. Nielsen R, Hellmann I, Hubisz M, Bustamante C, Clark AG. Recent and ongoing selection in the human genome. *Nat Rev Genet*. 2007; 8(11):857–868. <https://doi.org/10.1038/nrg2187> PMID: [17943193](https://pubmed.ncbi.nlm.nih.gov/17943193/); PubMed Central PMCID: PMC2933187.
10. Hernandez M, Perry GH. Scanning the human genome for “signatures” of positive selection: Transformative opportunities and ethical obligations. *Evol Anthropol*. 2021; 30(2):113–21. Epub 20210331. <https://doi.org/10.1002/evan.21893> PMID: [33788352](https://pubmed.ncbi.nlm.nih.gov/33788352/).
11. Fan S, Hansen ME, Lo Y, Tishkoff SA. Going global by adapting local: A review of recent human adaptation. *Science*. 2016; 354(6308):54–59. <https://doi.org/10.1126/science.aaf5098> PMID: [27846491](https://pubmed.ncbi.nlm.nih.gov/27846491/); PubMed Central PMCID: PMC5154245.
12. Pavlidis P, Jensen JD, Stephan W. Searching for footprints of positive selection in whole-genome SNP data from nonequilibrium populations. *Genetics*. 2010; 185(3):907–22. Epub 20100420. <https://doi.org/10.1534/genetics.110.116459> PMID: [20407129](https://pubmed.ncbi.nlm.nih.gov/20407129/); PubMed Central PMCID: PMC2907208.
13. Peter BM, Huerta-Sanchez E, Nielsen R. Distinguishing between selective sweeps from standing variation and from a de novo mutation. *PLoS Genet*. 2012; 8(10):e1003011. Epub 20121011. <https://doi.org/10.1371/journal.pgen.1003011> PMID: [23071458](https://pubmed.ncbi.nlm.nih.gov/23071458/); PubMed Central PMCID: PMC3469416.
14. Schrider DR, Kern AD. Supervised Machine Learning for Population Genetics: A New Paradigm. *Trends Genet*. 2018; 34(4):301–12. Epub 20180110. <https://doi.org/10.1016/j.tig.2017.12.005> PMID: [29331490](https://pubmed.ncbi.nlm.nih.gov/29331490/); PubMed Central PMCID: PMC5905713.
15. 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. Epub 20161013. <https://doi.org/10.1126/science.aag0776> PMID: [27738015](https://pubmed.ncbi.nlm.nih.gov/27738015/); PubMed Central PMCID: PMC5182071.
16. Palamara PF, Terhorst J, Song YS, Price AL. High-throughput inference of pairwise coalescence times identifies signals of selection and enriched disease heritability. *Nat Genet*. 2018; 50(9):1311–7. Epub 20180813. <https://doi.org/10.1038/s41588-018-0177-x> PMID: [30104759](https://pubmed.ncbi.nlm.nih.gov/30104759/); PubMed Central PMCID: PMC6145075.
17. Nait Saada J, Kalantzis G, Shyr D, Cooper F, Robinson M, Gusev A, et al. Identity-by-descent detection across 487,409 British samples reveals fine scale population structure and ultra-rare variant associations. *Nat Commun*. 2020; 11(1):6130. Epub 20201130. <https://doi.org/10.1038/s41467-020-19588-x> PMID: [33257650](https://pubmed.ncbi.nlm.nih.gov/33257650/); PubMed Central PMCID: PMC7704644.
18. Hejase HA, Dukler N, Siepel A. From Summary Statistics to Gene Trees: Methods for Inferring Positive Selection. *Trends Genet*. 2020; 36(4):243–58. Epub 20200115. <https://doi.org/10.1016/j.tig.2019.12.008> PMID: [31954511](https://pubmed.ncbi.nlm.nih.gov/31954511/); PubMed Central PMCID: PMC7177178.
19. Speidel L, Forest M, Shi S, Myers SR. A method for genome-wide genealogy estimation for thousands of samples. *Nat Genet*. 2019; 51(9):1321–9. Epub 20190902. <https://doi.org/10.1038/s41588-019-0484-x> PMID: [31477933](https://pubmed.ncbi.nlm.nih.gov/31477933/); PubMed Central PMCID: PMC7610517.
20. Stern AJ, Wilton PR, Nielsen R. An approximate full-likelihood method for inferring selection and allele frequency trajectories from DNA sequence data. *PLoS Genet*. 2019; 15(9):e1008384. Epub 20190913. <https://doi.org/10.1371/journal.pgen.1008384> PMID: [31518343](https://pubmed.ncbi.nlm.nih.gov/31518343/); PubMed Central PMCID: PMC6760815.
21. Skoglund P, Mathieson I. Ancient Genomics of Modern Humans: The First Decade. *Annu Rev Genomics Hum Genet*. 2018; 19:381–404. Epub 20180425. <https://doi.org/10.1146/annurev-genom-083117-021749> PMID: [29709204](https://pubmed.ncbi.nlm.nih.gov/29709204/).
22. Dehasque M, Avila-Arcos MC, Diez-Del-Molino D, Fumagalli M, Guschanski K, Lorenzen ED, et al. Inference of natural selection from ancient DNA. *Evol Lett*. 2020; 4(2):94–108. Epub 20200318. <https://doi.org/10.1002/evl3.165> PMID: [32313686](https://pubmed.ncbi.nlm.nih.gov/32313686/); PubMed Central PMCID: PMC7156104.
23. Marciniak S, Perry GH. Harnessing ancient genomes to study the history of human adaptation. *Nat Rev Genet*. 2017; 18(11):659–74. Epub 20170911. <https://doi.org/10.1038/nrg.2017.65> PMID: [28890534](https://pubmed.ncbi.nlm.nih.gov/28890534/).
24. Souilmi Y, Tobler R, Johar A, Williams M, Grey ST, Schmidt J, et al. Admixture has obscured signals of historical hard sweeps in humans. *Nat Ecol Evol*. 2022; 6(12):2003–15. Epub 20221031. <https://doi.org/10.1038/s41559-022-01914-9> PMID: [36316412](https://pubmed.ncbi.nlm.nih.gov/36316412/); PubMed Central PMCID: PMC9715430.
25. Mathias RA, Fu W, Akey JM, Ainsworth HC, Torgerson DG, Ruczinski I, et al. Adaptive evolution of the FADS gene cluster within Africa. *PLoS ONE*. 2012; 7(9):e44926. Epub 20120919. <https://doi.org/10.1371/journal.pone.0044926> PMID: [23028684](https://pubmed.ncbi.nlm.nih.gov/23028684/); PubMed Central PMCID: PMC3446990.
26. Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Roodenberg SA, et al. Genome-wide patterns of selection in 230 ancient Eurasians. *Nature*. 2015; 528(7583):499–503. Epub 20151123. <https://doi.org/10.1038/nature16152> PMID: [26595274](https://pubmed.ncbi.nlm.nih.gov/26595274/); PubMed Central PMCID: PMC4918750.

27. Fumagalli M, Moltke I, Grarup N, Racimo F, Bjerregaard P, Jorgensen ME, et al. Greenlandic Inuit show genetic signatures of diet and climate adaptation. *Science*. 2015; 349(6254):1343–1347. <https://doi.org/10.1126/science.aab2319> PMID: 26383953.
28. Buckley MT, Racimo F, Allentoft ME, Jensen MK, Jonsson A, Huang H, et al. Selection in Europeans on Fatty Acid Desaturases Associated with Dietary Changes. *Mol Biol Evol*. 2017; 34(6):1307–1318. <https://doi.org/10.1093/molbev/msx103> PMID: 28333262; PubMed Central PMCID: PMC5435082.
29. Mathieson S, Mathieson I. FADS1 and the Timing of Human Adaptation to Agriculture. *Mol Biol Evol*. 2018; 35(12):2957–2970. <https://doi.org/10.1093/molbev/msy180> PMID: 30272210; PubMed Central PMCID: PMC6278866.
30. Mathieson I. Limited Evidence for Selection at the FADS Locus in Native American Populations. *Mol Biol Evol*. 2020; 37(7):2029–2033. <https://doi.org/10.1093/molbev/msaa064> PMID: 32145021; PubMed Central PMCID: PMC7306688.
31. Allentoft ME, Sikora M, Sjogren KG, Rasmussen S, Rasmussen M, Stenderup J, et al. Population genomics of Bronze Age Eurasia. *Nature*. 2015; 522(7555):167–172. <https://doi.org/10.1038/nature14507> PMID: 26062507.
32. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive migration from the steppe was a source for Indo-European languages in Europe. *Nature*. 2015; 522(7555):207–11. Epub 20150302. <https://doi.org/10.1038/nature14317> PMID: 25731166; PubMed Central PMCID: PMC5048219.
33. Ju D, Mathieson I. The evolution of skin pigmentation-associated variation in West Eurasia. *Proc Natl Acad Sci U S A*. 2021; 118(1). <https://doi.org/10.1073/pnas.2009227118> PMID: 33443182; PubMed Central PMCID: PMC7817156.
34. Davy T, Ju D, Mathieson I, Skoglund P. Hunter-gatherer admixture facilitated natural selection in Neolithic European farmers. *Curr Biol*. 2023; 33(7):1365–71 e3. Epub 20230323. <https://doi.org/10.1016/j.cub.2023.02.049> PMID: 36963383; PubMed Central PMCID: PMC10153476.
35. Le MK, Smith OS, Akbari A, Harpak A, Reich D, Narasimhan VM. 1,000 ancient genomes uncover 10,000 years of natural selection in Europe. *bioRxiv*. 2022. Epub 20220826. <https://doi.org/10.1101/2022.08.24.505188> PMID: 36052370; PubMed Central PMCID: PMC9435429.
36. Hamid I, Korunes KL, Beleza S, Goldberg A. Rapid adaptation to malaria facilitated by admixture in the human population of Cabo Verde. *Elife*. 2021;10. Epub 20210104. <https://doi.org/10.7554/eLife.63177> PMID: 33393457; PubMed Central PMCID: PMC7815310.
37. Norris ET, Rishishwar L, Chande AT, Conley AB, Ye K, Valderrama-Aguirre A, et al. Admixture-enabled selection for rapid adaptive evolution in the Americas. *Genome Biol*. 2020; 21(1):29. Epub 20200207. <https://doi.org/10.1186/s13059-020-1946-2> PMID: 32028992; PubMed Central PMCID: PMC7006128.
38. Itan Y, Powell A, Beaumont MA, Burger J, Thomas MG. The origins of lactase persistence in Europe. *PLoS Comput Biol*. 2009; 5(8):e1000491. Epub 20090828. <https://doi.org/10.1371/journal.pcbi.1000491> PMID: 19714206; PubMed Central PMCID: PMC2722739.
39. Mathieson I, Terhorst J. Direct detection of natural selection in Bronze Age Britain. *Genome Res*. 2022; 32(11–12):2057–67. Epub 20221031. <https://doi.org/10.1101/gr.276862.122> PMID: 36316157; PubMed Central PMCID: PMC9808619.
40. Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, Unterlander M, et al. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. *Proc Natl Acad Sci U S A*. 2014; 111(13):4832–7. Epub 20140310. <https://doi.org/10.1073/pnas.1316513111> PMID: 24616518; PubMed Central PMCID: PMC3977302.
41. Klunk J, Vilgalys TP, Demeure CE, Cheng X, Shiratori M, Madej J, et al. Evolution of immune genes is associated with the Black Death. *Nature*. 2022; 611(7935):312–9. Epub 20221019. <https://doi.org/10.1038/s41586-022-05349-x> PMID: 36261521; PubMed Central PMCID: PMC9580435.
42. Barton AR, Santander CG, Skoglund P, Moltke I, Reich D, Mathieson I. Insufficient evidence for natural selection associated with the Black Death. *bioRxiv*. 2023. Epub 20230315. <https://doi.org/10.1101/2023.03.14.532615> PMID: 36993413; PubMed Central PMCID: PMC10055098.
43. Gopalakrishnan S, Ebenesersdottir SS, Lundstrom IKC, Turner-Walker G, Moore KHS, Luisi P, et al. The population genomic legacy of the second plague pandemic. *Curr Biol*. 2022; 32(21):4743–51 e6. Epub 20221001. <https://doi.org/10.1016/j.cub.2022.09.023> PMID: 36182700; PubMed Central PMCID: PMC9671091.
44. Hui R, Scheib CL, D'Atanasio E, Inskip SA, Cessford C, Biagini SA, et al. Medieval social landscape through the genetic history of Cambridgeshire before and after the Black Death. *bioRxiv*. 2023. Epub March 06, 2023. <https://doi.org/10.1101/2023.03.03.531048>
45. Mousseau TA, Roff DA. Natural selection and the heritability of fitness components. *Heredity (Edinb)*. 1987; 59(Pt 2):181–197. <https://doi.org/10.1038/hdy.1987.113> PMID: 3316130.

46. Day FR, Helgason H, Chasman DI, Rose LM, Loh PR, Scott RA, et al. Physical and neurobehavioral determinants of reproductive onset and success. *Nat Genet.* 2016; 48(6):617–23. Epub 20160418. <https://doi.org/10.1038/ng.3551> PMID: 27089180; PubMed Central PMCID: PMC5238953.
47. Barban N, Jansen R, de Vlaming R, Vaez A, Mandemakers JJ, Tropf FC, et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat Genet.* 2016; 48(12):1462–72. Epub 20161031. <https://doi.org/10.1038/ng.3698> PMID: 27798627; PubMed Central PMCID: PMC5695684.
48. 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. *Nat Hum Behav.* 2023; 7(5):790–801. Epub 20230302. <https://doi.org/10.1038/s41562-023-01528-6> PMID: 36864135.
49. Ye K, Gao F, Wang D, Bar-Yosef O, Keinan A. Dietary adaptation of FADS genes in Europe varied across time and geography. *Nat Ecol Evol.* 2017; 1:167. Epub 20170526. <https://doi.org/10.1038/s41559-017-0167> PMID: 29094686; PubMed Central PMCID: PMC5672832.
50. Siewert KM, Voight BF. Detecting Long-Term Balancing Selection Using Allele Frequency Correlation. *Mol Biol Evol.* 2017; 34(11):2996–3005. <https://doi.org/10.1093/molbev/msx209> PMID: 28981714; PubMed Central PMCID: PMC5850717.
51. Bitarello BD, de Filippo C, Teixeira JC, Schmidt JM, Kleinert P, Meyer D, et al. Signatures of Long-Term Balancing Selection in Human Genomes. *Genome Biol Evol.* 2018; 10(3):939–955. <https://doi.org/10.1093/gbe/evy054> PMID: 29608730; PubMed Central PMCID: PMC5952967.
52. Wu Y, Furuya S, Wang Z, Nobles JE, Fletcher JM, Lu Q. GWAS on birth year infant mortality rates provides evidence of recent natural selection. *Proc Natl Acad Sci U S A.* 2022; 119(12):e2117312119. Epub 20220314. <https://doi.org/10.1073/pnas.2117312119> PMID: 35290122; PubMed Central PMCID: PMC8944929.
53. Pavard S, Coste CFD. Evolutionary demographic models reveal the strength of purifying selection on susceptibility alleles to late-onset diseases. *Nat Ecol Evol.* 2021; 5(3):392–400. Epub 20210104. <https://doi.org/10.1038/s41559-020-01355-2> PMID: 33398109.
54. Blekhman R, Man O, Herrmann L, Boyko AR, Indap A, Kosiol C, et al. Natural selection on genes that underlie human disease susceptibility. *Curr Biol.* 2008; 18(12):883–889. <https://doi.org/10.1016/j.cub.2008.04.074> PMID: 18571414; PubMed Central PMCID: PMC2474766.
55. Corder EH, Saunders AM, Strittmatter WJ, Schmechel DE, Gaskell PC, Small GW, et al. Gene dose of apolipoprotein E type 4 allele and the risk of Alzheimer's disease in late onset families. *Science.* 1993; 261(5123):921–923. <https://doi.org/10.1126/science.8346443> PMID: 8346443.
56. Christensen K, Johnson TE, Vaupel JW. The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet.* 2006; 7(6):436–448. <https://doi.org/10.1038/nrg1871> PMID: 16708071; PubMed Central PMCID: PMC2726954.
57. Tobacco and Genetics Consortium. Genome-wide meta-analyses identify multiple loci associated with smoking behavior. *Nat Genet.* 2010; 42(5):441–7. Epub 20100425. <https://doi.org/10.1038/ng.571> PMID: 20418890; PubMed Central PMCID: PMC2914600.
58. Lucotte EA, Laurent R, Heyer E, Segurel L, Toupance B. Detection of Allelic Frequency Differences between the Sexes in Humans: A Signature of Sexually Antagonistic Selection. *Genome Biol Evol.* 2016; 8(5):1489–500. Epub 20160602. <https://doi.org/10.1093/gbe/evw090> PMID: 27189992; PubMed Central PMCID: PMC4898804.
59. Kasimatis KR, Abraham A, Ralph PL, Kern AD, Capra JA, Phillips PC. Evaluating human autosomal loci for sexually antagonistic viability selection in two large biobanks. *Genetics.* 2021; 217(1):1–10. <https://doi.org/10.1093/genetics/iyaa015> PMID: 33683357; PubMed Central PMCID: PMC8045711.
60. Ruzicka F, Connallon T. An unbiased test reveals no enrichment of sexually antagonistic polymorphisms on the human X chromosome. *Proc Biol Sci.* 2022; 289(1967):20212314. Epub 20220126. <https://doi.org/10.1098/rspb.2021.2314> PMID: 35078366; PubMed Central PMCID: PMC8790371.
61. Kasimatis KR, Ralph PL, Phillips PC. Limits to Genomic Divergence Under Sexually Antagonistic Selection. *G3 (Bethesda).* 2019; 9(11):3813–24. Epub 20191105. <https://doi.org/10.1534/g3.119.400711> PMID: 31530636; PubMed Central PMCID: PMC6829153.
62. Cheng C, Kirkpatrick M. Sex-Specific Selection and Sex-Biased Gene Expression in Humans and Flies. *PLoS Genet.* 2016; 12(9):e1006170. Epub 20160922. <https://doi.org/10.1371/journal.pgen.1006170> PMID: 27658217; PubMed Central PMCID: PMC5033347.
63. Bersaglieri T, Sabeti PC, Patterson N, Vanderploeg T, Schaffner SF, Drake JA, et al. Genetic signatures of strong recent positive selection at the lactase gene. *Am J Hum Genet.* 2004; 74(6):1111–20. Epub 20040426. <https://doi.org/10.1086/421051> PMID: 15114531; PubMed Central PMCID: PMC1182075.

64. Mills MC, Mathieson I. The challenge of detecting recent natural selection in human populations. *Proc Natl Acad Sci U S A*. 2022; 119(15):e2203237119. Epub 20220330. <https://doi.org/10.1073/pnas.2203237119> PMID: 35353603; PubMed Central PMCID: PMC9169803.
65. Pirastu N, Cordioli M, Nandakumar P, Mignogna G, Abdellaoui A, Hollis B, et al. Genetic analyses identify widespread sex-differential participation bias. *Nat Genet*. 2021; 53(5):663–71. Epub 20210422. <https://doi.org/10.1038/s41588-021-00846-7> PMID: 33888908; PubMed Central PMCID: PMC7611642.
66. Fry A, Littlejohns TJ, Sudlow C, Doherty N, Adamska L, Sprosen T, et al. Comparison of Sociodemographic and Health-Related Characteristics of UK Biobank Participants With Those of the General Population. *Am J Epidemiol*. 2017; 186(9):1026–1034. <https://doi.org/10.1093/aje/kwx246> PMID: 28641372; PubMed Central PMCID: PMC5860371.
67. Fisher R. The correlation between relatives on the supposition of mendelian inheritance. *Trans R Soc Edinb*. 1918; 53:399–433.
68. Pritchard JK, Pickrell JK, Coop G. The genetics of human adaptation: hard sweeps, soft sweeps, and polygenic adaptation. *Curr Biol*. 2010; 20(4):R208–R215. <https://doi.org/10.1016/j.cub.2009.11.055> PMID: 20178769; PubMed Central PMCID: PMC2994553.
69. Lohmueller KE, Mauney MM, Reich D, Braverman JM. Variants associated with common disease are not unusually differentiated in frequency across populations. *Am J Hum Genet*. 2006; 78(1):130–6. Epub 20051116. <https://doi.org/10.1086/499287> PMID: 16385456; PubMed Central PMCID: PMC1380210.
70. Turchin MC, Chiang CWK, Palmer CD, Sankararaman S, Reich D, Hirschhorn JN. Evidence of widespread selection on standing variation in Europe at height-associated SNPs. *Nat Genet*. 2012; 44(9):1015–1019. <https://doi.org/10.1038/ng.2368> PMID: 22902787
71. Robinson MR, Hemani G, Medina-Gomez C, Mezzavilla M, Esko T, Shakhbazov K, et al. Population genetic differentiation of height and body mass index across Europe. *Nat Genet*. 2015; 47(11):1357–62. Epub 20150914. <https://doi.org/10.1038/ng.3401> PMID: 26366552; PubMed Central PMCID: PMC4984852.
72. Zoledziewska M, Sidore C, Chiang CWK, Sanna S, Mulas A, Steri M, et al. Height-reducing variants and selection for short stature in Sardinia. *Nat Genet*. 2015; 47(11):1352–6. Epub 20150914. <https://doi.org/10.1038/ng.3403> PMID: 26366551; PubMed Central PMCID: PMC4627578.
73. Berg JJ, Coop G. A population genetic signal of polygenic adaptation. *PLoS Genet*. 2014; 10(8):e1004412. Epub 20140807. <https://doi.org/10.1371/journal.pgen.1004412> PMID: 25102153; PubMed Central PMCID: PMC4125079.
74. Racimo F, Berg JJ, Pickrell JK. Detecting Polygenic Adaptation in Admixture Graphs. *Genetics*. 2018; 208(4):1565–84. Epub 20180118. <https://doi.org/10.1534/genetics.117.300489> PMID: 29348143; PubMed Central PMCID: PMC5887149.
75. Edge MD, Coop G. Reconstructing the History of Polygenic Scores Using Coalescent Trees. *Genetics*. 2019; 211(1):235–62. Epub 20181102. <https://doi.org/10.1534/genetics.118.301687> PMID: 30389808; PubMed Central PMCID: PMC6325695.
76. Stern AJ, Speidel L, Zaitlen NA, Nielsen R. Disentangling selection on genetically correlated polygenic traits via whole-genome genealogies. *Am J Hum Genet*. 2021; 108(2):219–39. Epub 20210112. <https://doi.org/10.1016/j.ajhg.2020.12.005> PMID: 33440170; PubMed Central PMCID: PMC7895848.
77. Stearns SC, Byars SG, Govindaraju DR, Ewbank D. Measuring selection in contemporary human populations. *Nat Rev Genet*. 2010; 11(9):611–22. Epub 20100803. <https://doi.org/10.1038/nrg2831> PMID: 20680024.
78. Beauchamp JP. Genetic evidence for natural selection in humans in the contemporary United States. *Proc Natl Acad Sci U S A*. 2016; 113(28):7774–9. Epub 20160711. <https://doi.org/10.1073/pnas.1600398113> PMID: 27402742; PubMed Central PMCID: PMC4948342.
79. Kong A, Frigge ML, Thorleifsson G, Stefansson H, Young AI, Zink F, et al. Selection against variants in the genome associated with educational attainment. *Proc Natl Acad Sci U S A*. 2017; 114(5):E727–E32. Epub 20170117. <https://doi.org/10.1073/pnas.1612113114> PMID: 28096410; PubMed Central PMCID: PMC5293043.
80. Tropf FC, Stulp G, Barban N, Visscher PM, Yang J, Snieder H, et al. Human fertility, molecular genetics, and natural selection in modern societies. *PLoS ONE*. 2015; 10(6):e0126821. Epub 20150603. <https://doi.org/10.1371/journal.pone.0126821> PMID: 26039877; PubMed Central PMCID: PMC4454512.
81. Sanjak JS, Sidorenko J, Robinson MR, Thornton KR, Visscher PM. Evidence of directional and stabilizing selection in contemporary humans. *Proc Natl Acad Sci U S A*. 2018; 115(1):151–6. Epub 20171218. <https://doi.org/10.1073/pnas.1707227114> PMID: 29255044; PubMed Central PMCID: PMC5776788.

82. Zhu C, Ming MJ, Cole JM, Edge MD, Kirkpatrick M, Harpak A. Amplification is the primary mode of gene-by-sex interaction in complex human traits. *Cell Genom.* 2023; 3(5):100297. Epub 20230406. <https://doi.org/10.1016/j.xgen.2023.100297> PMID: 37228747; PubMed Central PMCID: PMC10203050.
83. Wang J, Fan X, Yang M, Song M, Wang K, Giovannucci E, et al. Sex-specific associations of circulating testosterone levels with all-cause and cause-specific mortality. *Eur J Endocrinol.* 2021; 184(5):723–732. <https://doi.org/10.1530/EJE-20-1253> PMID: 33690154.
84. Novembre J, Barton NH. Tread Lightly Interpreting Polygenic Tests of Selection. *Genetics.* 2018; 208(4):1351–1355. <https://doi.org/10.1534/genetics.118.300786> PMID: 29618592; PubMed Central PMCID: PMC5886544.
85. 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. Epub 20190321. <https://doi.org/10.7554/eLife.39725> PMID: 30895923; PubMed Central PMCID: PMC6428572.
86. Sohail M, Maier RM, Ganna A, Bloemendal 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. Epub 20190321. <https://doi.org/10.7554/eLife.39702> PMID: 30895926; PubMed Central PMCID: PMC6428571.
87. Martin AR, Gignoux CR, Walters RK, Wojcik GL, Neale BM, Gravel S, et al. Human Demographic History Impacts Genetic Risk Prediction across Diverse Populations. *Am J Hum Genet.* 2017; 100(4):635–49. Epub 20170330. <https://doi.org/10.1016/j.ajhg.2017.03.004> PMID: 28366442; PubMed Central PMCID: PMC5384097.
88. Mostafavi H, Harpak A, Agarwal I, Conley D, Pritchard JK, Przeworski M. Variable prediction accuracy of polygenic scores within an ancestry group. *Elife.* 2020;9. Epub 20200130. <https://doi.org/10.7554/eLife.48376> PMID: 31999256; PubMed Central PMCID: PMC7067566.
89. Zaidi AA, Mathieson I. Demographic history mediates the effect of stratification on polygenic scores. *Elife.* 2020;Epub 20201117:9. <https://doi.org/10.7554/eLife.61548> PMID: 33200985; PubMed Central PMCID: PMC7758063.
90. Gazal S, Loh PR, Finucane HK, Ganna A, Schoech A, Sunyaev S, et al. Functional architecture of low-frequency variants highlights strength of negative selection across coding and non-coding annotations. *Nat Genet.* 2018; 50(11):1600–7. Epub 20181008. <https://doi.org/10.1038/s41588-018-0231-8> PMID: 30297966; PubMed Central PMCID: PMC6236676.
91. Schoech AP, Jordan DM, Loh PR, Gazal S, O'Connor LJ, Balick DJ, et al. Quantification of frequency-dependent genetic architectures in 25 UK Biobank traits reveals action of negative selection. *Nat Commun.* 2019; 10(1):790. Epub 20190215. <https://doi.org/10.1038/s41467-019-08424-6> PMID: 30770844; PubMed Central PMCID: PMC6377669.
92. Zeng J, de Vlaming R, Wu Y, Robinson MR, Lloyd-Jones LR, Yengo L, et al. Signatures of negative selection in the genetic architecture of human complex traits. *Nat Genet.* 2018; 50(5):746–53. Epub 20180416. <https://doi.org/10.1038/s41588-018-0101-4> PMID: 29662166.
93. Bulmer MG. The genetic variability of polygenic characters under optimizing selection, mutation and drift. *Genet Res.* 1972; 19(1):17–25. <https://doi.org/10.1017/s0016672300014221> PMID: 5024710.
94. Harpak A, Przeworski M. The evolution of group differences in changing environments. *PLoS Biol.* 2021; 19(1):e3001072. Epub 20210125. <https://doi.org/10.1371/journal.pbio.3001072> PMID: 33493148; PubMed Central PMCID: PMC7861633.
95. Yair S, Coop G. Population differentiation of polygenic score predictions under stabilizing selection. *Philos Trans R Soc Lond B Biol Sci.* 2022; 377(1852):20200416. Epub 20220418. <https://doi.org/10.1098/rstb.2020.0416> PMID: 35430887; PubMed Central PMCID: PMC9014188.
96. Lande R, Arnold SJ. The Measurement of Selection on Correlated Characters. *Evolution.* 1983; 37(6):1210–1226. <https://doi.org/10.1111/j.1558-5646.1983.tb00236.x> PMID: 28556011.
97. Bulik-Sullivan B, Finucane HK, Anttila V, Gusev A, Day FR, Loh PR, et al. An atlas of genetic correlations across human diseases and traits. *Nat Genet.* 2015; 47(11):1236–41. Epub 20150928. <https://doi.org/10.1038/ng.3406> PMID: 26414676; PubMed Central PMCID: PMC4797329.
98. Jordan DM, Verbanck M, Do R. HOPS: a quantitative score reveals pervasive horizontal pleiotropy in human genetic variation is driven by extreme polygenicity of human traits and diseases. *Genome Biol.* 2019; 20(1):222. Epub 20191025. <https://doi.org/10.1186/s13059-019-1844-7> PMID: 31653226; PubMed Central PMCID: PMC6815001.
99. Szpak M, Xue Y, Ayub Q, Tyler-Smith C. How well do we understand the basis of classic selective sweeps in humans? *FEBS Lett.* 2019; 593(13):1431–48. Epub 20190611. <https://doi.org/10.1002/1873-3468.13447> PMID: 31116407.

100. Stearns FW. One hundred years of pleiotropy: a retrospective. *Genetics*. 2010; 186(3):767–773. <https://doi.org/10.1534/genetics.110.122549> PMID: 21062962; PubMed Central PMCID: PMC2975297.
101. Mathieson I. Human adaptation over the past 40,000 years. *Curr Opin Genet Dev*. 2020; 62:97–104. Epub 20200801. <https://doi.org/10.1016/j.gde.2020.06.003> PMID: 32745952; PubMed Central PMCID: PMC7484260.
102. Koch EM, Sunyaev SR. Maintenance of Complex Trait Variation: Classic Theory and Modern Data. *Front Genet*. 2021; 12:763363. Epub 20211112. <https://doi.org/10.3389/fgene.2021.763363> PMID: 34868244; PubMed Central PMCID: PMC8636146.