Polygenic signals of sexually antagonistic selection in contemporary human genomes

Filip Ruzicka^{1*}, Luke Holman^{2,3}, Tim Connallon¹

¹ School of Biological Sciences, Monash University, Clayton, Victoria, Australia

² School of BioSciences, University of Melbourne, Parkville, Victoria, Australia

³ School of Applied Sciences, Edinburgh Napier University, Edinburgh, UK

*Filip Ruzicka

Email: Filip.Ruzicka@monash.edu

This PDF file includes:

Main Text

Figures 1 to 5

Abstract

Mutations that increase fitness in one sex may decrease fitness in the other. Such “sexually antagonistic” (SA) genetic variants can constrain adaptation and increase variability for fitness components (e.g., survival, fertility, and disease susceptibility). However, detecting SA selection in genomes is immensely challenging, as it requires prohibitively large datasets that combine genomic sequences with individual fitness measurements. Here, we use genotypic and reproductive success data from ~250,000 UK Biobank individuals to comprehensively assess the extent of SA genetic variation in humans. We first develop new theoretical models for signals of SA selection spanning a full generational life cycle—including SA polymorphisms affecting survival, reproductive success and overall fitness. Comparing our models with UK Biobank data, we uncover multiple empirical signals of polygenic SA selection, including sex-differential effects of genetic variants on each fitness component, and positive correlations between sex-differential effects and minor allele frequencies. We show that these signals cannot be explained by simple models of sex differences in purifying selection, or by potential confounders such as population structure and sequence mapping errors. We further show that candidate SA sites disproportionately affect functional genomic regions, including polymorphisms associated with quantitative traits and disease. Finally, we examine historical evolutionary processes affecting candidate SA sites, which are consistent with the drift-dominated dynamics predicted by previous theory. Overall, our results support SA genomic variation in humans and highlight its broader functional and evolutionary consequences.

Introduction

Adaptation of a population to its environment requires heritable genetic variation for fitness¹. Although many populations show substantial genetic variation for fitness components²—including life-history traits such as maturation rate, lifespan, mating success, and fertility^{2,3}—genetic trade-offs between components reduce variation for overall fitness, and thereby limit adaptive potential⁴. For example, a mutation that increases the probability of survival to adulthood might simultaneously decrease adult reproductive success (*e.g.*,⁵), weakening the mutation's net fitness effect⁴. In addition to slowing adaptation^{6–8}, genetic trade-offs can increase standing genetic variation^{2,9}, give rise to balancing selection^{10,11}, and favour evolutionary transitions between mating systems^{12,13}, modes of sex determination¹⁴ and genome structures^{15–18}.

Sexually antagonistic (SA) genetic polymorphisms—in which the alleles that benefit one sex are harmful to the other—are a type of genetic trade-off that may be common in sexually reproducing species¹⁹. Theory shows that SA polymorphisms are likely to arise when mutations differentially affect trait expression in each sex, or when mutations similarly affect traits under divergent directional selection between the sexes²⁰. Empirical quantitative genetic studies imply that both conditions are frequently met^{21–24} and, accordingly, that SA polymorphisms contribute to phenotypic variation in a range of plant and animal populations (*e.g.* ^{25–27}), including humans^{28–30}.

Although there is now abundant evidence that SA polymorphisms contribute to phenotypic variation, efforts to identify and characterise SA alleles in genomic data face two formidable challenges³¹. First, methods using explicit fitness measurements to identify SA polymorphisms (*e.g.*, genome-wide association studies (GWAS) of fitness³²) are rarely feasible, because it is challenging to obtain fitness measurements for large numbers of genotyped individuals under natural conditions². Second, methods using allele frequency

differences between adult females and males as genomic signals of SA viability selection (*e.g.*, between-sex F_{ST} estimates^{31,33–40}) are limited in several ways: they have low power to detect individual SA loci, they cannot distinguish SA selection from sex differences in the strength of selection, they are susceptible to artefacts generated by population structure and erroneous mapping of sequence reads to sex chromosomes^{31,38,39,41}, and they neglect fitness components other than viability, such as reproductive success^{31,42}. Previous studies of human genomic data^{33,34,41} have, to varying degrees, been affected by each of these issues^{31,39,40,42}, such that we currently lack robust evidence of SA genomic variation in humans. More generally, these impediments help to explain the limited catalogue of confirmed SA polymorphisms across species^{43–45}, which currently comprises a handful of loci with exceptionally large phenotypic effects (*e.g.*^{46–50}).

Despite these challenges, new datasets and analytical approaches provide opportunities to identify robust genomic signals of SA selection. First, massive ‘biobank’ datasets, which are widely used in human genomics, sometimes include both genotype and offspring number data^{29,51} that can be used to detect loci with SA effects on reproductive components of fitness³¹. Second, estimates of allele frequency differences between sexes—though ill-suited for confidently identifying individual SA loci affecting viability—may nevertheless be amenable to genome-wide tests for *polygenic* SA viability selection^{31,33}. Third, polygenic metrics of sex-differential selection (*e.g.*, between-sex F_{ST}) may include an appreciable proportion of genuine SA loci in the upper tails of their distributions, providing a set of candidate loci that can collectively yield insights into the general properties of SA polymorphisms (*e.g.*, their functional characteristics and evolutionary dynamics), despite uncertainty about individual candidates.

Here, we extend^{31,33} and develop new statistical tests to detect polygenic signals of SA polymorphism affecting viability, reproduction, and total fitness during a full generational

cycle. We apply them to the UK Biobank ⁵¹, a dataset comprising quality-filtered genotype and offspring number data for ~250,000 men and women. The tests, which are based on between-sex F_{ST} metrics and the covariance between F_{ST} metrics and minor allele frequencies (MAF), reveal polygenic signals of SA selection. Specifically: (i) F_{ST} estimates are consistently higher than expected under a null hypothesis in which sex-differential selection is absent; (ii) F_{ST} estimates are positively correlated with MAF across loci, as predicted under SA selection. We support the F_{ST} -based results by: (i) replicating these analyses using mixed-model association tests that explicitly control for any systematic differences in the genetic ancestry of female and male individuals; and (ii) showing that loci exhibiting artefactual signatures of SA selection due to sequence mapping errors ⁴¹ do not occur in our quality-filtered dataset. We further show that candidate SA polymorphisms are preferentially situated in coding and regulatory sequences, and contribute to variation in a range of morphological, behavioural and medically relevant phenotypes. Finally, we use genetic diversity data from global human populations to test for long-term balancing selection, which has been hypothesised to affect SA polymorphisms ¹⁰.

Results

Genomic signals of sex differences in selection: theoretical predictions

To illustrate our analytical approach, consider a large, well-mixed population containing many polymorphic, bi-allelic, autosomal loci. At the start of a generation (*i.e.*, at fertilisation), Mendelian inheritance equalises allele frequencies between the sexes (Fig. 1, left box). Loci with sex-differential effects on zygote-to-adult survival will accumulate between-sex allele frequency differences that become apparent among adults (*e.g.*, the black allele becomes enriched in adult males and deficient in adult females because it improves zygote-to-adult survival in males but reduces it in females; Fig. 1, middle box). Furthermore, loci with sex-

differential effects on the number of offspring produced will have different transmission rates to the next generation from surviving females *versus* surviving males (*e.g.*, the black allele increases transmission to offspring of males but decreases transmission to offspring of females; Fig. 1, right box). Thus, provided allele frequency and offspring number data are available for adult females and males, we can, in principle, estimate sex-differential effects of genetic variation during a full generational life cycle (Fig. 1), even though allele frequencies among zygotes and gametes are inferred and not directly observed.

The UK Biobank provides this opportunity, as it includes genotypes and reported offspring numbers (hereafter “lifetime reproductive success” or LRS, following standard evolutionary biology terminology²⁹) among putatively post-reproductive adults (ages 45-69 after filtering; see Methods). For a focal autosomal locus with alleles A_1 and A_2 , we denote $\hat{p}_f$ and $\hat{p}_m$ ($\hat{q}_f$ and $\hat{q}_m$) the respective estimated frequencies of the A_1 (A_2) allele in adult females and males of the UK Biobank. Using LRS data of UK Biobank individuals, projected paternal and maternal transmission rates of each allele are estimated as:

$$\hat{p}'_m = \frac{M_{11} + \frac{1}{2}M_{12}}{M_{11} + M_{12} + M_{22}} \quad [1],$$

$$\hat{p}'_f = \frac{F_{11} + \frac{1}{2}F_{12}}{F_{11} + F_{12} + F_{22}}$$

where M_{ij} and F_{ij} represent the cumulative LRS of individuals (males and females, respectively) with genotype ij (*e.g.*, M_{11} , M_{12} and M_{22} correspond to genotypes A_1A_1 , A_1A_2 , and A_2A_2).

Using F_{ST} ⁵², we partition between-sex allele frequency differentiation over one generation into the differentiation due to: 1) sex-differential variation in survival to adulthood (hereafter “adult F_{ST} ”; see^{31,33,42}); 2) sex-differential variation in adult LRS (hereafter “reproductive F_{ST} ”); and 3) sex-differential variation in overall fitness (hereafter “gametic

135 F_{ST}). Single-locus estimates of adult F_{ST} , reproductive F_{ST} and gametic F_{ST} are defined,
136 respectively, as:

$$\hat{F}_{ST} = \frac{(\hat{p}_f - \hat{p}_m)^2}{4\bar{p}(1 - \bar{p})} \quad [2a],$$

$$\hat{F}_{ST} = \frac{((\hat{p}'_f - \hat{p}_f) - (\hat{p}'_m - \hat{p}_m))^2}{4\bar{p}(1 - \bar{p})} \quad [2b],$$

$$\hat{F}_{ST} = \frac{(\hat{p}'_f - \hat{p}'_m)^2}{4\bar{p}'(1 - \bar{p}')} \quad [2c],$$

137 where $\bar{p} = (\hat{p}_f + \hat{p}_m)/2$ and $\bar{p}' = (\hat{p}'_f + \hat{p}'_m)/2$.

138 In the absence of sex differences in selection (*e.g.*, under neutrality, or under sexually
139 concordant selection of equal magnitude and direction in each sex), with large sample sizes,
140 and excluding single-nucleotide polymorphisms (SNPs) with very low minor allele frequencies
141 (*i.e.*, MAF < 1% in the UK Biobank), we show that adult, reproductive and gametic $\hat{F}_{ST}$
142 converge, respectively, to the following simple null distributions (*SI Appendix*, Section A):

$$\hat{F}_{ST} \approx [1/8N_f + 1/8N_m]X \quad [3a],$$

$$\hat{F}_{ST} \approx [(\sigma_f^2/\mu_f^2)/8N_f + (\sigma_m^2/\mu_m^2)/8N_m]X \quad [3b],$$

$$\hat{F}_{ST} \approx [(1 + \sigma_f^2/\mu_f^2)/8N_f + (1 + \sigma_m^2/\mu_m^2)/8N_m]X \quad [3c].$$

143 where X is a chi-square random variable with one degree of freedom, N_f and N_m denote adult
144 female and male sample sizes, μ_f and μ_m denote mean LRS, and σ_f^2 and σ_m^2 denote the
145 corresponding variances in LRS for sampled adult females and males. All three null
146 distributions are independent of MAF, provided sites with very low MAFs are first removed.

147 Relative to the null distributions in eqs. [3a-c], sex differences in selection alter the
148 distributions of each F_{ST} metric in two important ways. First, sex differences in selection inflate
149 $\hat{F}_{ST}$ metrics relative to null model predictions (*SI Appendix*, Section B). These inflations may
150 arise due to polymorphisms under sex-differential selection *and* neutral polymorphisms that

hitchhike with selected polymorphisms, but linkage disequilibrium (LD) alone cannot inflate genome-wide $\hat{F}_{ST}$ in the absence of genuine selected polymorphisms (*SI Appendix*, Section C). Thus, $\hat{F}_{ST}$ inflations represent reliable signals of sex-differentially selected polymorphism³¹ (if not of the number of independent selected loci), provided population structure and technical artefacts are also controlled (as shown below).

Second, sex differences in selection generate statistical associations between $\hat{F}_{ST}$ and MAF that are indicative of the form of sex-differential selection. Specifically, SA selection can generate a positive covariance between each $\hat{F}_{ST}$ metric and MAF, whereas sexually concordant (and/or sex-limited) selection generates a negative covariance (*SI Appendix*, Section B). The negative covariance under sexually concordant selection arises because purifying selection disproportionately lowers the frequency of large-effect alleles relative to small-effect alleles, which is a general feature of loci evolving under purifying selection (see⁵³). In contrast, a positive covariance can arise under SA selection because large-effect SA alleles reach intermediate frequencies more easily under balancing selection than small-effect SA alleles, and because intermediate-frequency alleles inflate $\hat{F}_{ST}$ more than low-frequency alleles (irrespective of the strength of balancing selection; *e.g.*, even under drift). Thus, while any sex difference in selection can elevate genome-wide $\hat{F}_{ST}$, positive associations between $\hat{F}_{ST}$ and MAF cannot arise under sex differences in purifying selection alone—instead, they indicate the presence of SA variation. These predictions apply when current patterns of sex-differential selection (which affect measures of F_{ST}) are similar to historical patterns of selection (which influence MAF), and are valid under both deterministic and non-deterministic scenarios (*SI Appendix*, Section B). Below, we test for both predicted signals of SA selection in the UK Biobank dataset.

UK Biobank SNP data

The sample size after removing individuals with close relatedness or non-White British ancestry (see Methods) was $N = 250,297$ ($N_m = 116,661$ males and $N_f = 133,636$ females). We removed rare polymorphic sites ($MAF < 1\%$), sites with low genotype or imputation quality, and sites potentially driving artefactually high between-sex allele frequency differences based on criteria identified by Kasimatis et al.⁴¹. These criteria were based on differences in missing rates between sexes, deficits of minor allele homozygotes, and excess heterozygosity (*i.e.*, beyond what can be plausibly explained by sex differences in selection; see *SI Appendix*, Section D). Unless noted, we focus on quality-filtered genotyped SNPs ($N = 544,635$) in the main text. Results from LD-pruned genotyped SNPs and imputed SNPs ($N = 7,844,649$) closely match genotyped data and are presented in the *SI Appendix* (Sections E-H), along with full statistical tables.

Polygenic signals of sex differences in selection in genome-wide $\hat{F}_{ST}$ distributions

We tested for sex differences in selection by calculating adult $\hat{F}_{ST}$, reproductive $\hat{F}_{ST}$ and gametic $\hat{F}_{ST}$ (eqs. [2a-c]) in the UK Biobank, and contrasting these standardised estimates of allele frequency differentiation between sexes against: (i) their respective theoretical null distributions (eqs. [3a-c]), and (ii) empirical null distributions generated by a single random permutation of male and female labels among individuals or, in the case of reproductive $\hat{F}_{ST}$, a single permutation of LRS among individuals of a given sex (see Methods).

All three $\hat{F}_{ST}$ metrics showed significantly greater differentiation than predicted by their theoretical and empirical null distributions, consistent with sex differences in selection contributing to genome-wide allele frequency differentiation between the sexes. Mean adult $\hat{F}_{ST}$ in the observed data was larger than predicted by both null distributions (theoretical null: 2.012×10^{-6} ; permuted null: 2.014×10^{-6} ; observed: 2.107×10^{-6} ; Wilcoxon and Kolmogorov-Smirnov tests, $p < 0.001$; Fig. 2A,D; *SI Appendix*, Section E, Figs. S1-5; Tab. S1), with a 20.4%

excess of SNPs in the top 1% quantile of the theoretical null and a 16.5% excess in the top 1% quantile of the empirical null (χ^2 tests, $p < 0.001$). Mean reproductive $\hat{F}_{ST}$ was larger than predicted by both nulls (theoretical: 8.667×10^{-7} ; permuted: 8.770×10^{-7} ; observed: 9.049×10^{-7} ; Wilcoxon and Kolmogorov-Smirnov tests, $p < 0.001$; Fig. 2B,E; *SI Appendix*, Section E, Figs. S1-5; Tab. S1), with a 17.9% excess of SNPs in the top 1% quantile of the theoretical null and a 11.5% excess in the top 1% quantile of the empirical null (χ^2 tests, $p < 0.001$). Mean gametic $\hat{F}_{ST}$ was larger than predicted by both nulls (theoretical: 2.874×10^{-6} ; permuted: 2.874×10^{-6} ; observed: 2.986×10^{-6} ; Wilcoxon and Kolmogorov-Smirnov tests, $p < 0.001$; Fig. 2C,F; *SI Appendix*, Section E, Figs. S1-5; Tab. S1), with a 17.1% excess of SNPs in the top 1% of the theoretical and empirical nulls (χ^2 tests, $p < 0.001$).

Signals of sex differences in selection in adult, reproductive, and gametic $\hat{F}_{ST}$ were polygenic: although each $\hat{F}_{ST}$ metric was markedly inflated on a genome-wide scale, no individual locus reached a p-value below the Bonferroni-corrected threshold (9.181×10^{-8} ; *SI Appendix*, Section E, Figs. S6-7). For example, the minimum p-values were 3.260×10^{-7} for adult $\hat{F}_{ST}$ (minimum q-value=0.101), 7.036×10^{-7} for reproductive $\hat{F}_{ST}$ (minimum q-value=0.383), and 2.858×10^{-6} for gametic $\hat{F}_{ST}$ (minimum q-value=0.346).

Positive associations between $\hat{F}_{ST}$ and MAF indicate SA selection

To distinguish polygenic signals of SA selection from other modes of selection that might also give rise to elevated $\hat{F}_{ST}$ (e.g., selection acting only on one sex, or differences in the strength but not the direction of selection between sexes), we examined the association between MAF and each $\hat{F}_{ST}$ metric. This association can be positive under SA selection but should be negative under sex-differential purifying selection, provided patterns of sex-differential selection are relatively consistent over time (see above). Consistent with SA selection, we observed positive correlations between MAF and all three $\hat{F}_{ST}$ metrics (adult $\hat{F}_{ST}$, $\rho = 0.014$, $p < 0.001$; reproductive

$\hat{F}_{ST}$, $\rho=0.010$, $p<0.001$; gametic $\hat{F}_{ST}$, $\rho=0.010$, $p<0.001$; Fig. 3A-C; *SI Appendix*, Section F, Figs. S9-13; Tab. S2), with all correlations stronger in observed than null data (*SI Appendix*, Section F, Fig. 3D-F; Figs. S10,12).

Signals of SA selection persist after controlling for sex-specific population structure

In principle, genome-wide elevations in $\hat{F}_{ST}$ can arise entirely from systematic differences in ancestry (population structure) between sexes and in the absence of genuine sex differences in selection^{31,42}. We therefore replicated our analyses using mixed-model association tests that are analogous to $\hat{F}_{ST}$ but which explicitly correct for population structure.

We first evaluated signals of sex differences in viability selection by performing a GWAS of sex^{31,41,54} using standardised estimates of the log-odds ratio ($\hat{\mathcal{L}}_{ST}$; see Methods). $\hat{\mathcal{L}}_{ST}$ quantifies between-sex allele frequency differences among adults (like adult $\hat{F}_{ST}$) while controlling for population structure by including a kinship matrix of genome-wide relatedness between individuals. As expected, $\hat{\mathcal{L}}_{ST}$ was highly correlated with adult $\hat{F}_{ST}$ (Spearman's $\rho=0.988$; $p<0.001$; *SI Appendix*, Section E, Fig. S8) and was elevated relative to an empirical null distribution generated through permutation of sex labels (null mean $\hat{\mathcal{L}}_{ST}$: 5.000×10^{-7} ; mean observed: 5.188×10^{-7} ; Wilcoxon and Kolmogorov-Smirnov tests, $p<0.001$; *SI Appendix*, Section E, Figs. S1-5; Tab. S1), with a 14.0% excess of SNPs in the top 1% quantile of the empirical null (χ^2 test, $p<0.001$), and a significant SNP-heritability of sex ($h^2_{SNP} \pm SE = 0.014 \pm 0.002$).

We then tested for a signal of sex differences in selection through reproductive fitness components by performing separate GWAS for LRS in females and males (controlled for population structure) and quantifying the difference between female and male effect sizes using a t -statistic ($|t|$; see Methods). As expected, the distribution of $|t|$ was highly correlated to reproductive $\hat{F}_{ST}$ (Spearman's $\rho=0.921$; $p<0.001$; *SI Appendix*, Section E, Fig. S8) and was

elevated relative to an empirical null generated by permutation of LRS within each sex (null mean $|t|=0.797$, observed mean $|t|=0.817$, Wilcoxon and Kolmogorov-Smirnov tests, $p<0.001$; *SI Appendix*, Section E, Figs. S1-5; Tab. S1), with an 17.9% excess of SNPs in the top 1% quantile of the empirical null (χ^2 test, $p<0.001$). Finally, $\hat{\mathcal{L}}_{ST}$ and $|t|$ were both positively correlated with MAF ($\hat{\mathcal{L}}_{ST}$: $\rho=0.014$, $p<0.001$; $|t|$: $\rho=0.010$, $p<0.001$; *SI Appendix*, Section F, Figs. S9-13; Tab. S2), confirming the signals of SA selection detected with $\hat{F}_{ST}$ metrics.

Functional analyses of candidate SA SNPs

If loci with high $\hat{F}_{ST}$ (or related metrics) include genuine SA polymorphisms, they should preferentially be found in functionally important regions in the genome, whereas such enrichment is not expected if the observed $\hat{F}_{ST}$ inflation results from random chance, genotyping errors, or sex-specific population structure. We therefore conducted several enrichment tests, both to support our inference that SA selection is occurring and to explore functional effects of candidate SA loci.

First, we asked whether sites with high sex-differentiation were disproportionately situated in “genic” regions, defined broadly to include coding and regulatory elements (see Methods). This test revealed consistently positive associations between each metric of sex-differential selection and genic status (Binomial GLMs: adult $\hat{F}_{ST}$, $p=0.021$; $\hat{\mathcal{L}}_{ST}$, $p=0.018$; reproductive $\hat{F}_{ST}$, $p=0.0497$; $|t|$, $p=0.025$; gametic $\hat{F}_{ST}$, $p=0.026$; Fig. 4A-C; *SI Appendix*, Section G, Figs. S14-15; Tab. S3). Polygenic patterns of genic enrichment in the imputed data were similar to the genotyped data (*SI Appendix*, Section G, Fig. S16; Tab. S3), though extremely differentiated loci were less often genic in the imputed data, which may reflect higher genotype uncertainty among imputed sites and the consequently higher likelihood of artefacts.

Second, we used summary statistics from the Neale laboratory database of UK Biobank GWAS to explore whether metrics of sex-differential selection were associated with candidate loci for 43 phenotypes, chosen for their medical relevance and/or relationship to phenotypic sex differences. Though few traits were associated after FDR-correction in the genotyped data (*SI Appendix*, Section G, Fig. S17), we found multiple associations in the imputed data, where the number of loci (and thus statistical power) is larger (Fig. 4D). For example, several disease-relevant traits (high blood pressure; red blood cell count; diseases of the musculoskeletal system and connective tissue; heel bone mineral density) and behavioural traits (neuroticism score) were associated with metrics of sex-differential survival and LRS (Mann-Whitney tests; FDR-corrected $p < 0.05$). We also found trait associations that were unique to metrics of sex-differential survival (white blood cell count; body fat percentage) and LRS (standing height; platelet count). Loci affecting testosterone had lower-than-average gametic $\hat{F}_{ST}$.

Evolutionary histories of candidate SA SNPs

To gain insights into the evolutionary dynamics of candidate SA sites, we tested for signals of long-term balancing selection, which can, under certain scenarios of selection and dominance, stably maintain SA polymorphisms over long periods of time¹⁰. We used three metrics of balancing selection: allele age estimates from the Atlas of Variant Age database⁵⁵, between-population $\hat{F}_{ST}$ from four worldwide non-European populations from the 1000 Genomes Project⁵⁶, and three independent sets of candidate loci for balancing selection^{57–59}. In each case, we looked for associations between $\hat{F}_{ST}$ metrics of sex-differential selection and metrics of balancing selection, while controlling for ascertainment bias of intermediate-frequency alleles (which are, on average, older) among highly sex-differentiated sites (*i.e.*, Fig. 3; see Methods). Under long-term balancing selection, we would expect candidate SA alleles (and linked loci) to be older, to exhibit lower between-population $\hat{F}_{ST}$, and to disproportionately co-

localise with previous candidates for balancing selection, compared to non-candidate SA sites with similar allele frequencies in the UK Biobank.

We found no significant associations between metrics of sex-differential selection and metrics of balancing selection, including allele age (adult $\hat{F}_{ST}$, Spearman's $\rho=0.001$, $p=0.353$; reproductive $\hat{F}_{ST}$, Spearman's $\rho=0.000$, $p=0.919$; gametic $\hat{F}_{ST}$: Spearman's $\rho=0.002$, $p=0.092$; Fig. 5; *SI Appendix*, Section H, Figs. S18,21,22; Tab. S4), between-population $\hat{F}_{ST}$ (*SI Appendix*, Section H, Fig. S19; Tab. S5), or previous candidates for balancing selection (*SI Appendix*, Section H, Figs. S20,21,23; Tabs. S6-8). Overall, there is no compelling evidence for long-term balancing selection among candidate SA SNPs.

Discussion

Sex differences in directional selection on phenotypes have been reported in a wide range of animal taxa^{21–23,60}, including post-industrial human populations^{29,30}, yet population genomic signals of SA selection have been extremely difficult to establish. The reason is simple: sexual reproduction equalises autosomal allele frequencies between the sexes every generation, which restricts genetic divergence between males and females and, in effect, precludes the use of tests that are commonly used to detect other forms of selection (*e.g.*, McDonald-Kreitman tests for positive selection and F_{ST} outlier tests for spatially-varying selection^{61–63}). Published studies using human genomic data illustrate the challenges of studying SA polymorphisms, including sample sizes that may be insufficient for detecting polygenic signals of selection^{31,33,34}, lack of controls for population structure or technical artefacts^{33,34}, and/or absence of data concerning reproductive components of fitness^{33,34,41}.

Here, by developing new metrics of SA genomic variation and applying them to data from a quarter-million UK adults, we present robust evidence for polygenic sex-differential selection in humans. Specifically, UK Biobank individuals showed sex differences in allele

frequencies—both among adults and their (projected) offspring—that consistently exceeded expectations defined by our theoretical null models for viability, reproductive and total fitness, and these signals persisted after controlling for sex-specific population structure. Our analyses of reproductive $\hat{F}_{ST}$ and $|t|$ metrics are particularly compelling because they control for allele frequency differences between samples of adults of each sex and rule out factors that might otherwise affect estimated adult allele frequencies in the UK Biobank (*e.g.*, mis-mapping of reads to sex chromosomes, participation bias ⁵⁴; see below). These elevations in reproductive $\hat{F}_{ST}$ and $|t|$ are consistent with previous observations in post-industrial human populations, including variation in female and male LRS ⁶⁴ (a necessary precondition for sex-differential selection), widespread sex differences in the genetic basis of quantitative traits (*e.g.*, in the UK BioBank ⁶⁵), and SA selection on phenotypes (*e.g.*, height ^{29,30} and multivariate trait combinations ⁶⁴), which should collectively lead to SA polymorphisms throughout the genome ²⁰. In particular, our finding that candidate loci for standing height were positively associated with reproductive $\hat{F}_{ST}$ and $|t|$ is in line with earlier reports of SA selection acting on adult height ^{29,30}. Furthermore, the positive associations we report between MAF and $\hat{F}_{ST}$ (and related metrics) are consistent with SA selection and cannot be explained by simple models of sex differences in purifying selection. Though it is conceivable that contemporary patterns of selection (which influence $\hat{F}_{ST}$) are sufficiently different from historical patterns of selection (which influence MAF) that our models do not reflect the evolutionary causes of the positive associations between MAF and $\hat{F}_{ST}$, SA polymorphisms represent a parsimonious explanation for the positive correlation we detect ^{29,30}.

At first glance, the signals of SA viability selection in adult $\hat{F}_{ST}$ and $\hat{L}_{ST}$ metrics may seem surprising given the low rates of infant and juvenile mortality in post-industrial human populations ⁶⁴. However, mortality after fertilisation but before birth is common (on the order of 50% ⁶⁶), giving ample opportunity for sex differences in mortality to generate allele

frequency differences⁶⁷. Another possibility, raised by Pirastu et al.⁵⁴, is that sex differences in adult allele frequencies arise because of sex differences in the genetic basis of individuals' predisposition to take part in the UK Biobank study, which relies upon active participation. To support this interpretation, they reported significant SNP heritability of sex (a polygenic measure of sex differences in allele frequencies) in biobanks relying on active participation (UK Biobank; 23andMe), and non-significant SNP heritability in biobanks using passive participation (Biobank Japan; FinnGen; iPsych). However, this result might instead reflect differences in statistical power, as passive participation studies are considerably smaller than active participation studies. Further work (*e.g.*, parent-offspring trio analyses that control for participation effects⁶⁷; replication studies in other Biobank-scale datasets) will be required to conclusively distinguish the relative contributions of SA selection and participation bias on allele frequency differentiation among male and female adults, and to assess whether SA variants affecting viability tend to increase or decrease LRS within and across sexes.

We also found that candidate SA sites were preferentially found within coding and regulatory elements, and contributed to variation in morphological, physiological, behavioural, and medically relevant phenotypes. These functional analyses provide further evidence that elevated between-sex $\hat{F}_{ST}$ reflects selection rather than technical artefacts. They also lend support to the hypothesis that SA alleles contribute to disease susceptibility⁹. For example, loci associated with bone mineral density and red blood cell count showed elevated signals of SA selection. Bone mineralisation and red blood cell count are sexually dimorphic traits that positively genetically correlate with musculoskeletal disease and high blood pressure⁶⁸. Accordingly, loci associated with musculoskeletal disease and high blood pressure also showed elevated sex-differentiation.

Our analyses do not reveal clear signals of long-term balancing selection among candidate SA sites, which may stem from several factors. First, balancing selection—to the

extent that it maintains SA polymorphisms—may be too recent to generate a clear statistical signal in our analyses⁶⁹. Second, balancing selection at SA loci may be *effectively* weak, owing to relatively small N_e in humans⁷⁰ and the high susceptibility of SA alleles to genetic drift⁷¹. Third, long-term balancing selection may be present, but statistical tests for it may be too weak to stand out from the background noise of false positives in our metrics and the datasets used to quantify balancing selection⁷². Overall, evidence that SA polymorphisms contribute to standing genetic variation—as in our study—is at present stronger than evidence that they are stably maintained by balancing selection (but see³²).

To what extent should we expect the patterns reported here to replicate across studies? Many taxa exhibit greater variance for reproductive success than humans⁷³, generating higher potential for polygenic signals of SA selection. In line with this, polygenic inflation of adult $\hat{F}_{ST}$ has previously been documented in modest samples of pipefish and flycatchers^{31,36,37}, suggesting that sex differences in selection might be stronger in those species than in humans. Yet while polygenic signals of SA selection may replicate across studies, candidate SA polymorphisms may not. One reason is that environmental differences between populations (*e.g.*, cultural differences in family planning between human populations) could alter the set of loci under SA selection. Additionally, the noisiness of polygenic signals of SA selection^{31,42}, along with the near-certainty that most polymorphic loci have small effects on fitness⁷⁴, generates variation in the set of candidate SA polymorphisms *identified* across populations or generations, even if causal SA polymorphisms do not differ. Similar predictions apply to scenarios of polygenic local adaptation with high gene flow⁷⁵, whose population genetic dynamics strongly parallel SA selection^{31,76}.

Finally, the extent to which signals of balancing selection will co-occur with SA polymorphisms awaits further study, but given the pronounced sensitivity of SA balancing selection to genetic drift^{71,77}, we should expect the relationship between signals of SA selection

and balancing selection to vary with N_e . For example, previous work in *Drosophila melanogaster* indicates that candidate SA polymorphisms segregate across worldwide populations and even species³², and in contrast to the results of our study, these candidates are putatively maintained under balancing selection. A parsimonious explanation for these contrasting findings is that the effectiveness of balancing selection is lower in humans than fruit flies due to much smaller N_e .

Methods

Quality control of UK Biobank data

Access to UK Biobank data was granted under project number 52049. We used sample-level information provided by the UK Biobank (see⁵¹ for details) to perform individual-level (phenotypic) quality controls. Specifically, we excluded individuals with high relatedness (3rd degree or closer), non-White British ancestry, high heterozygosity and high missing rates. Individuals whose reported sex did not match their inferred genetic sex, whose age was less than 45 years, and aneuploids were also excluded. In addition to site-level quality-controls implemented by the UK Biobank⁵¹, we used PLINK and PLINK2⁷⁸ to remove sites that were non-diallelic, had MAF<1%, missing rates>5%, and p-values<10⁻⁶ in tests of Hardy-Weinberg equilibrium. These filters were applied to all autosomal genotyped and imputed sites. We further removed imputed sites with INFO score ≤0.8, denoting poor imputation.

Artefact filtering in UK Biobank data

Mis-mapping of autosomal reads to sex chromosomes can generate between-sex allele frequency differences in the absence of sex differences in selection^{31,41} (though note that such artefacts are not expected to inflate metrics of adult reproductive success, since these metrics control for any sex differences in allele frequencies among adults). To remove mis-mapping

artefacts, we excluded: (i) sites with heterozygosity levels that exceeded what could plausibly be expected under SA selection (see below and *SI Appendix*, Section D); (ii) sites with a deficit of minor allele homozygotes; (iii) sites exhibiting large differences in missing rate between sexes. These three patterns have previously been shown to associate with mis-mapping of reads to sex chromosomes ⁴¹.

To remove sites with artificially inflated heterozygosity, we estimated F_{IS} for each SNP, defined as:

$$\hat{F}_{IS} = \frac{P_{Aa}}{2\bar{p}(1 - \bar{p})} - 1$$

where P_{Aa} denotes the frequency of heterozygotes for a given locus and $\bar{p}$ the sex-averaged allele frequency. For a SA locus at polymorphic equilibrium, the distribution of $\hat{F}_{IS}$ is well approximated by a normal distribution with expectation and variance as follows:

$$E[\hat{F}_{IS}] \approx \frac{1}{2n} + \frac{p(1-p)}{4} \left(\frac{s_{\max}}{1 - ps_{\max}} \right)^2$$

$$\text{var}[\hat{F}_{IS}] = \frac{1}{n}$$

where n is total sample size of adults, p the minor allele frequency, and $s_{\max} = \max(s_m, s_f)$ with s_m and s_f representing male and female selection coefficients (*SI Appendix*, Section D). To identify SNPs with excess heterozygosity, we compared $\hat{F}_{IS}$ in the observed data to expected $\hat{F}_{IS}$ under strong SA selection ($s_{\max}=0.2$) by performing a one-tailed Z-test for excess heterozygosity. We thus obtained p-values for each locus, corrected p-values for multiple testing using Benjamini-Hochberg false discovery rates ⁷⁹, and removed sites with FDR q-values below 0.05. Note that this test was designed to eliminate mapping errors, but it will also eliminate sites under strong (but not weak or moderately strong) sex-differential selection. The net effect of applying this filter is a slightly conservative test of sex-differential selection.

To identify sites with a deficit of minor allele homozygotes, we compared the observed frequency of minor allele homozygotes to the expected frequency under Hardy-Weinberg equilibrium (p^2 , where p is the frequency of the minor allele) by performing a one-tailed binomial test, removing sites with FDR q-values below 0.05. Tests for excess heterozygosity and deficits of minor allele homozygotes were performed across all individuals (regardless of sex), and also for each sex separately. Sites were removed if they exhibited q-values below 0.05 in any of the three tests (*i.e.*, both sexes combined, females, and males). Finally, to assess differences in missing rate between the sexes, we performed a χ^2 test, removing sites with FDR q-values below 0.05.

Reassuringly, applying these filters eliminated all 8 sites that Kasimatis et al.⁴¹ identified as false-positives for SA viability selection in the genotyped data. One remaining site (rs114928327) appeared in the imputed data but exhibited lower-than-average adult $\hat{F}_{ST}$ and $\hat{L}_{ST}$.

Quantifying polygenic signals of sex differences in selection

$\hat{F}_{ST}$ -based metrics

Allele frequencies in adults of each sex were obtained directly from sequence data and used to calculate adult $\hat{F}_{ST}$ for each polymorphic site (post-filtering). Allele frequency differences between sexes were estimated using the $\hat{F}_{ST}$ metrics presented in eqs. [2a-c]. Data on offspring number were obtained from field 2405 “Number of children fathered”, for males, and field 2734 “Number of live births”, for females (hereafter “lifetime reproductive success” (LRS)). Because some individuals were asked to report offspring number at repeated assessment points, we considered the maximum offspring number reported as the definitive value of LRS for that individual. Though mis-estimation of LRS for each individual cannot be definitively excluded (*e.g.*, due to mis-reporting, or additional reproduction after data collection), we minimised this

possibility by removing individuals: (i) younger than 45 years of age (see “Quality control of UK Biobank data”); (ii) reporting fewer offspring at a later assessment point than at an earlier assessment point; (iii) with 20 or more reported offspring numbers (large offspring numbers often ended in zero—*e.g.*, 20, 30, 50, 100—and were thus considered less reliable). Furthermore, we note that uncounted LRS data adds imprecision but should not systematically bias our analyses. Finally, previous research indicates that offspring number represents a good proxy for LRS in post-industrial human populations, based on observations of positive genetic correlations between offspring and grand-offspring numbers across generations ⁸⁰.

The theoretical null distributions presented in eqs. [3a-c] apply to genome-wide data in which the sample of female and male sequences are constant across loci. For each null, we adjusted for locus-specific missing data (which causes variation in locus-specific sample sizes) by allowing sample sizes to vary across loci (*i.e.*, N_{f_i} and N_{m_i} adults, or $2N_{f_i}$ and $2N_{m_i}$ sequences, per diploid locus i).

Case-control GWAS of sex

To complement the test for sex-differential viability selection based on adult $\hat{F}_{ST}$, we performed a GWAS of sex ^{31,41,54}. By analogy to adult $\hat{F}_{ST}$, loci with sex-differential effects on viability in a GWAS of sex will tend to have relatively large absolute log-odds ratios (corresponding to relatively large allele frequency differences between sexes). Unlike adult $\hat{F}_{ST}$, this approach additionally permits the inclusion of covariates in the GWAS to account for population structure and other possible confounders ³¹.

We used BOLT-LMM to run a mixed-model GWAS ⁸¹, using a kinship matrix to account for population structure. The kinship matrix was constructed from a subset of 241,669 LD-pruned genotyped SNPs (settings ‘--indep-pairwise 50 10 0.2’ in PLINK; MAF>0.01). We added individual age (field 54) and assessment centre (field 21003) as fixed-effect covariates.

To facilitate comparisons with adult $\hat{F}_{ST}$, we standardised the regression coefficients (log-odds ratios) from the GWAS by allele frequency, such that:

$$\hat{\mathcal{L}}_{ST} = (\sqrt{\bar{p}(1 - \bar{p})}\hat{\mathcal{L}})^2$$

where $\hat{\mathcal{L}}$ is the log-odds ratio and $\bar{p}$ is the sex-averaged allele frequency among adults. We also quantified a genome-wide signal of sex-differential viability selection by estimating the SNP-heritability of sex, in BOLT-LMM.

t-statistic for sex differences in GWAS of LRS

To complement the test for sex differences in selection on LRS based on reproductive $\hat{F}_{ST}$, we performed a GWAS of LRS in each sex separately, using a mixed-model GWAS implemented in BOLT-LMM, which allowed us to correct for population structure in effect size estimates. Following ⁸², we quantified differences in male and female effect size estimates by means of a *t*-statistic, defined as:

$$|t| = \left| \frac{\beta_M - \beta_F}{\sqrt{SE_M^2 + SE_F^2 - 2\rho SE_M SE_F}} \right|$$

where β_M and β_F are estimated effect sizes obtained from sex-stratified GWAS, SE_M and SE_F are sex-specific standard errors, and ρ is the between-sex rank correlation among genome-wide LD-pruned loci. Loci with elevated values of $|t|$ in observed relative to permuted data denote candidate loci with sex-differential effects on LRS.

Statistical comparisons of null and observed distributions

Null distributions for each metric of sex differences in selection (*i.e.*, adult $\hat{F}_{ST}$, $\hat{\mathcal{L}}_{ST}$, reproductive $\hat{F}_{ST}$, $|t|$, and gametic $\hat{F}_{ST}$) have been theoretically derived (for the $\hat{F}_{ST}$ metrics; see *SI Appendix*, Section A) and/or obtained empirically through permutation (all five metrics). Empirical null distributions were generated as follows. For adult $\hat{F}_{ST}$, gametic $\hat{F}_{ST}$ and $\hat{\mathcal{L}}_{ST}$, we

performed a single permutation of female and male labels and re-calculated the statistic in the permuted data. For reproductive $\hat{F}_{ST}$ and $|t|$, we performed a single permutation of LRS values within each sex—without permuting sex—and re-calculated the statistic in the permuted data. Permuting LRS without permuting sex allows allele frequencies to differ between sexes (as would happen if, for example, sex-differential viability selection is occurring among juveniles) while randomising effects of genotype on LRS (which ensures that sex-differential selection on LRS cannot contribute to the empirical null). Performing a single permutation for each metric was sufficient for our purposes because we were focussed on testing a cumulative signal of selection across loci, rather than establishing statistical significance at the single-locus level.

To test for elevations of each metric relative to the (theoretical or empirical) nulls, we performed Wilcoxon Rank-Sum and Kolmogorov-Smirnov tests. These tests assess differences in the median and distribution of the observed and null data, respectively. As a complementary way of comparing observed and null data, we quantified enrichment of observed values in the top 1% of each null using a χ^2 test. Finally, we estimated the difference between the mean value of the metric in the observed data and the mean value of the metric in each null, obtaining 95% confidence intervals and empirical p-values through bootstrapping (1,000 replicates; where each replicate consists of the set of relevant genotyped or imputed SNPs, sampled with replacement).

Associations between metrics of sex differences in selection and MAF

To test for associations between metrics of sex differences in selection and MAF, we performed a Spearman's rank correlation between each metric and MAF. We also tested whether the relationship between metrics of sex differences in selection and MAF was more pronounced in the observed data than in the (theoretical or empirical) nulls by estimating the difference between correlations in the observed data and correlations in the null data among 1,000

bootstrap replicates (as above), thereby generating 95% confidence intervals and empirical p-values.

Functional analyses

Following research on candidate locally adapted loci^{83–85}, we tested whether loci exhibiting elevated values of each metric of sex-differential selection were disproportionately situated in putatively functional regions of the genome. We categorised SNPs as either intergenic or non-intergenic based on annotations obtained with SnpEff⁸⁶. We then fitted a logistic regression model to our data, with annotation (genic/intergenic) as the binary response variable and the metric of sex differences in selection under consideration as an independent variable. We also examined whether associations between metrics of sex differences in selection and genic status were more pronounced in the observed data than under (theoretical or empirical) nulls by calculating the difference between the regression coefficients in the observed and null data among 1,000 bootstrap replicates (as above).

We compared our five metrics to previously-documented GWAS hits from Neale laboratory's analyses of all UK Biobank traits⁶⁸. Briefly, we compared the average value of the focal metric (*e.g.*, adult $\hat{F}_{ST}$ or $|t|$) between loci that were and were not associated ($\alpha = 10^{-4}$) with the focal phenotype ($n = 43$ traits) using a Mann-Whitney test, and then FDR-corrected the resulting p-values. These analyses were performed on LD-pruned SNPs to reduce pseudo-replication due to LD between loci. Analyses are fully documented (including code and results) at https://lukeholman.github.io/ukbb_sex_fst.

Quantifying signals of long-term balancing selection

Allele ages

If candidate SA variants experience sufficiently strong and sustained balancing selection relative to the countervailing effects of genetic drift, we expect them to be older than the genome-wide average ⁶⁹. We used the Atlas of Variant Age database to obtain allele age estimates for genome-wide variants ⁵⁵. Estimates of allele age in this database apply to the non-reference (*i.e.*, alternative) allele and are derived from coalescent modelling of the time to the most recent common ancestor using the “Genealogical Estimation of Variant Age” method (see ⁵⁵ for details). Estimates of allele age make use of genomic data from: (i) the 1000 Genomes Project; (ii) the Simons Genome Diversity Project; (iii) both datasets combined. For each site in the UK Biobank, we obtained the median estimate of allele age from the combined dataset (when available), from the 1000 Genomes Project, or the Simons Genome Diversity Project (when neither alternative estimate was available).

Between-population F_{ST}

If candidate SA variants experience sufficiently strong balancing selection maintaining a fixed polymorphic equilibrium, they should exhibit lower-than-average allele frequency differences *between* populations ⁶⁹. We obtained allele frequency data from four non-European populations from the 1000 Genomes Project: Southern Han Chinese (CHS, N=105); Yoruba Nigerians (YRI, N=108); Puerto Ricans (PUR, N=104) and Gujarati Indians (GIH, N=103). For each pairwise combination of populations, we estimated between-population $\hat{F}_{ST}$, as:

$$\hat{F}_{ST} = \frac{(\hat{p}_1 - \hat{p}_2)^2}{4\bar{p}(1 - \bar{p})}$$

where $\hat{p}_1$ and $\hat{p}_2$ are allele frequency estimates in the relevant pair of populations, and $\bar{p} = (\hat{p}_1 + \hat{p}_2)/2$.

Previous candidates for balancing selection

If candidate SA variants experience strong balancing selection, they should disproportionately co-occur with previously identified candidates for balancing selection. We used three independent sets of candidate sites for balancing selection to investigate this possibility ^{57–59}: (i) the dataset of Andrés et al. ⁵⁷, which consists of 64 genes exhibiting elevated polymorphism (as determined using the Hudson-Kreitman-Aguadé test) and/or intermediate-frequency alleles across 19 African-American or 20 European-American individuals; (ii) the dataset of DeGiorgio et al. ⁵⁹, which consists of 400 candidate genes exhibiting elevated T_1 or T_2 statistics among 9 European (CEU) and 9 African (YRI) individuals. T_1 or T_2 statistics quantify the likelihood that a genomic region exhibits levels of neutral polymorphism that are consistent with a linked balanced polymorphism; (iii) the dataset of Bitarello et al. ⁵⁸, which consists of 1859 candidate genes exhibiting elevated values of “Non-Central Deviation” (NCD) statistics. NCD statistics also quantify the likelihood that given genomic regions are situated nearby a balanced polymorphism, using polymorphism data from 50 random individuals from two African (YRI; LWK) and European (GBR; TSI) populations and divergence data from a chimpanzee outgroup.

We assigned each site in the UK Biobank dataset to a gene using SnpEff ⁸⁶ and categorised sites as candidates or non-candidates for balancing selection based on whether they were annotated as belonging to a candidate or non-candidate in each of the three aforementioned datasets.

Statistical associations between metrics of sex-differential selection and balancing selection

To test whether signals of sex differences in selection were associated with signals of balancing selection, we performed Spearman’s rank correlations between alternative allele age (scaled by alternative allele frequency in the UK Biobank, to control for ascertainment bias) and each metric of sex differences in selection. For between-population $\hat{F}_{ST}$, we performed multiple

linear regressions, with metric of sex differences in selection as the independent variable and MAF in the UK Biobank as a fixed-effect covariate (to control for ascertainment bias). For previous candidates for balancing selection, we performed multiple logistic regressions, where candidate/non-candidate status was treated as a binary response variable, metric of sex differences in selection as the independent variable, and UK Biobank MAF as a fixed-effect covariate. We also tested whether associations between metrics of balancing selection and sex differences in selection were more pronounced in the observed data than under (theoretical or empirical) null. This was done by calculating the difference between the summary statistics (*i.e.*, correlation or regression coefficients, as outlined above) in the observed and null data among 1,000 bootstrap replicates.

References

1. Fisher, R. A. *The Genetical Theory of Natural Selection*. *Genetics* (Clarendon Press, 1930). doi:10.1038/158453a0.
2. Charlesworth, B. Causes of natural variation in fitness: evidence from studies of *Drosophila* populations. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 1662–1629 (2015).
3. Visscher, P. M. *et al.* 10 years of GWAS discovery: Biology, function, and translation. *Am. J. Hum. Genet.* **101**, 5–22 (2017).
4. Walsh, B. & Blows, M. W. Abundant Genetic Variation + Strong Selection = Multivariate Genetic Constraints: A Geometric View of Adaptation. *Annu. Rev. Ecol. Evol. Syst.* **40**, 41–59 (2009).
5. Johnston, S. E. *et al.* Life history trade-offs at a single locus maintain sexually selected genetic variation. *Nature* **502**, 93–95 (2013).
6. Connallon, T. & Hall, M. D. Genetic constraints on adaptation: a theoretical primer for the genomics era. *Ann. N. Y. Acad. Sci.* **1422**, 65–87 (2018).
7. Arnqvist, G. & Tuda, M. Sexual conflict and the gender load: Correlated evolution between population fitness and sexual dimorphism in seed beetles. *Proc. R. Soc. B Biol. Sci.* **277**, 1345–1352 (2010).
8. Whitlock, M. C. & Agrawal, A. F. Purging the genome with sexual selection: Reducing mutation load through selection on males. *Evolution* **63**, 569–582 (2009).
9. Gilks, W. P., Abbott, J. K. & Morrow, E. H. Sex differences in disease genetics: Evidence, evolution, and detection. *Trends Genet.* **30**, 453–463 (2014).
10. Kidwell, J. F., Clegg, M. T., Stewart, F. M. & Prout, T. Regions of stable equilibria for models of differential selection in the two sexes under random mating. *Genetics* **85**, 171–183 (1977).
11. Curtsinger, J. W., Service, P. M. & Prout, T. Antagonistic pleiotropy, reversal of

- 657 dominance, and genetic polymorphism. *Am. Nat.* **144**, 210–228 (1994).
- 658 12. Charlesworth, B. & Charlesworth, D. A model for the evolution of dioecy and
659 gynodioecy. *Am. Nat.* **112**, 975–997 (1978).
- 660 13. Olito, C. & Connallon, T. Sexually antagonistic variation and the evolution of
661 dimorphic sexual systems. *Am. Nat.* **193**, 688–701 (2019).
- 662 14. Van Doorn, G. S. & Kirkpatrick, M. Turnover of sex chromosomes induced by sexual
663 conflict. *Nature* **449**, 909–912 (2007).
- 664 15. Charlesworth, D. & Charlesworth, B. Sex differences in fitness and selection for
665 centric fusions between sex-chromosomes and autosomes. *Genet. Res.* **35**, 205–214
666 (1980).
- 667 16. Jordan, C. Y. & Charlesworth, D. The potential for sexually antagonistic
668 polymorphism in different genome regions. *Evolution* **66**, 505–516 (2012).
- 669 17. Kirkpatrick, M. & Guerrero, R. F. Signatures of sex-antagonistic selection on
670 recombining sex chromosomes. *Genetics* **197**, 531–541 (2014).
- 671 18. Patten, M. M. & Haig, D. Maintenance or loss of genetic variation under sexual and
672 parental antagonism at a sex-linked locus. *Evolution* **63**, 2888–2895 (2009).
- 673 19. Bonduriansky, R. & Chenoweth, S. F. Intralocus sexual conflict. *Trends Ecol. Evol.*
674 **24**, 280–288 (2009).
- 675 20. Connallon, T. & Clark, A. G. Evolutionary inevitability of sexual antagonism. *Proc. R.*
676 *Soc. B Biol. Sci.* **281**, 2013–2123 (2014).
- 677 21. Cox, R. M. & Calsbeek, R. Sexually antagonistic selection, sexual dimorphism, and
678 the resolution of intralocus sexual conflict. *Am. Nat.* **173**, 176–187 (2009).
- 679 22. Singh, A. & Punzalan, D. The strength of sex-specific selection in the wild. *Evolution*
680 **72**, 2818–2824 (2018).
- 681 23. De Lisle, S. P., Goedert, D., Reedy, A. M. & Svensson, E. I. Climatic factors and

- species range position predict sexually antagonistic selection across taxa. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170415 (2018).
24. Mackay, T. F. C. The genetic architecture of quantitative traits. *Annu. Rev. Genet.* **35**, 303–339 (2001).
25. Delph, L. F. *et al.* Environment-dependent intralocus sexual conflict in a dioecious plant. *New Phytol.* **192**, 542–552 (2011).
26. Svensson, E. I., McAdam, A. G. & Sinervo, B. Intralocus sexual conflict over immune defense, gender load, and sex-specific signaling in a natural lizard population. *Evolution* **63**, 3124–3135 (2009).
27. Chippindale, A. K., Gibson, J. R. & Rice, W. R. Negative genetic correlation for adult fitness between sexes reveals ontogenetic conflict in *Drosophila*. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 1671–1675 (2001).
28. Stulp, G., Kuijper, B., Buunk, A. P., Pollet, T. V & Verhulst, S. Intralocus sexual conflict over human height. *Biol. Lett.* **8**, 976–8 (2012).
29. Sanjak, J. S., Sidorenko, J., Robinson, M. R., Thornton, K. R. & Visscher, P. M. Evidence of directional and stabilizing selection in contemporary humans. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 151–156 (2017).
30. Stearns, S. C., Govindaraju, D. R., Ewbank, D. & Byars, S. G. Constraints on the coevolution of contemporary human males and females. *Proc. R. Soc. B Biol. Sci.* **279**, 4836–4844 (2012).
31. Ruzicka, F. *et al.* The search for sexually antagonistic genes: Practical insights from studies of local adaptation and statistical genomics. *Evol. Lett.* **4**, 398–415 (2020).
32. Ruzicka, F. *et al.* Genome-wide sexually antagonistic variants reveal long-standing constraints on sexual dimorphism in fruit flies. *PLoS Biol.* **17**, e3000244 (2019).
33. Cheng, C. & Kirkpatrick, M. Sex-specific selection and sex-biased gene expression in

- humans and flies. *PLoS Genet.* **12**, e1006170 (2016).
34. Lucotte, E. A., Laurent, R., Heyer, E., Ségurel, L. & Toupance, B. Detection of allelic frequency differences between the sexes in humans: a signature of sexually antagonistic selection. *Genome Biol. Evol.* **8**, 1489–1500 (2016).
35. Wright, A. E. *et al.* Sex-biased gene expression resolves sexual conflict through the evolution of sex-specific genetic architecture. *Evol. Lett.* **2**, 52–61 (2018).
36. Flanagan, S. P. & Jones, A. G. Genome-wide selection components analysis in a fish with male pregnancy. *Evolution* **71**, 1096–1105 (2017).
37. Dutoit, L. *et al.* Sex-biased gene expression, sexual antagonism and levels of genetic diversity in the collared flycatcher (*Ficedula albicollis*) genome. *Mol. Ecol.* **27**, 3572–3581 (2018).
38. Bisseger, M., Laurentino, T. G., Roesti, M. & Berner, D. Widespread intersex differentiation across the stickleback genome – The signature of sexually antagonistic selection? *Mol. Ecol.* **29**, 262–271 (2019).
39. Mank, J. E., Shu, J. J. & Wright, A. E. Signature of sexual conflict is actually conflict resolved. *Mol. Ecol.* (2020) doi:10.1111/mec.15311.
40. Cheng, C. & Kirkpatrick, M. The signal of sex-specific selection in humans is not an artefact: Reply to Mank *et al.* *Mol. Ecol.* (2020) doi:10.1111/mec.15420.
41. Kasimatis, K. R. *et al.* Evaluating human autosomal loci for sexually antagonistic viability selection in two large biobanks. *Genetics* **217**, 1–10 (2021).
42. Kasimatis, K. R., Ralph, P. L. & Phillips, P. C. Limits to genomic divergence under sexually antagonistic selection. *G3* **9**, 3813–3824 (2019).
43. Mank, J. E. Population genetics of sexual conflict in the genomic era. *Nat. Rev. Genet.* **18**, 721–730 (2017).
44. Rowe, L., Chenoweth, S. F. & Agrawal, A. F. The genomics of sexual conflict. *Am.*

- 732 *Nat.* **192**, 274–286 (2018).
- 733 45. Kasimatis, K. R., Nelson, T. C. & Phillips, P. C. Genomic signatures of sexual
734 conflict. *J. Hered.* **108**, 780–790 (2017).
- 735 46. Roberts, R. B., Ser, J. R. & Kocher, T. D. Sexual conflict resolved by invasion of a
736 novel sex determiner in Lake Malawi cichlid fishes. *Science* **326**, 998–1001 (2009).
- 737 47. Rostant, W. G., Kay, C., Wedell, N. & Hosken, D. J. Sexual conflict maintains
738 variation at an insecticide resistance locus. *BMC Biol.* **13**, 34 (2015).
- 739 48. Barson, N. J. *et al.* Sex-dependent dominance at a single locus maintains variation in
740 age at maturity in salmon. *Nature* **528**, 405–408 (2015).
- 741 49. VanKuren, N. W. & Long, M. Gene duplicates resolving sexual conflict rapidly
742 evolved essential gametogenesis functions. *Nat. Ecol. Evol.* **2**, 705–712 (2018).
- 743 50. Pearse, D. E. *et al.* Sex-dependent dominance maintains migration supergene in
744 rainbow trout. *Nat. Ecol. Evol.* **3**, 1731–1742 (2019).
- 745 51. Bycroft, C. *et al.* The UK Biobank resource with deep phenotyping and genomic data.
746 *Nature* (2018) doi:10.1038/s41586-018-0579-z.
- 747 52. Wright, S. The genetical structure of species. *Ann. Eugen.* **15**, 323–354 (1951).
- 748 53. Eyre-Walker, A. Genetic architecture of a complex trait and its implications for fitness
749 and genome-wide association studies. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 1752–1756
750 (2010).
- 751 54. Pirastu, N. *et al.* Genetic analyses identify widespread sex-differential participation
752 bias. *bioRxiv* (2020).
- 753 55. Albers, P. K. & McVean, G. Dating genomic variants and shared ancestry in
754 population-scale sequencing data. *PLoS Biol.* **18**, (2020).
- 755 56. The 1000 Genomes Project Consortium. A global reference for human genetic
756 variation. *Nature* **526**, 68–74 (2015).

- 757 57. Andrés, A. M. *et al.* Targets of balancing selection in the human genome. *Mol. Biol.*
758 *Evol.* **26**, 2755–2764 (2009).
- 759 58. Bitarello, B. D. *et al.* Signatures of long-term balancing selection in human genomes.
760 *Genome Biol. Evol.* **10**, 939–955 (2018).
- 761 59. DeGiorgio, M., Lohmueller, K. E. & Nielsen, R. A model-based approach for
762 identifying signatures of ancient balancing selection in genetic data. *PLoS Genet.* **10**,
763 e1004561 (2014).
- 764 60. Morrissey, M. B. Meta-analysis of magnitudes, differences and variation in
765 evolutionary parameters. *J. Evol. Biol.* **29**, 1882–1904 (2016).
- 766 61. Vitti, J. J., Grossman, S. R. & Sabeti, P. C. Detecting natural selection in genomic
767 data. *Annu. Rev. Genet.* **47**, 97–120 (2013).
- 768 62. Hoban, S. *et al.* Finding the genomic basis of local adaptation: Pitfalls, practical
769 solutions, and future directions. *Am. Nat.* **188**, 379–397 (2016).
- 770 63. Booker, T. R., Jackson, B. C. & Keightley, P. D. Detecting positive selection in the
771 genome. *BMC Biol.* **15**, (2017).
- 772 64. Stearns, S. C., Byars, S. G., Govindaraju, D. R. & Ewbank, D. Measuring selection in
773 contemporary human populations. *Nat. Rev. Genet.* **11**, 611–622 (2010).
- 774 65. Bernabeu, E. *et al.* Sex differences in genetic architecture in UK Biobank. *Nat. Genet.*
775 **53**, 1283–1289 (2021).
- 776 66. Benagiano, G., Farris, M. & Grudzinskas, G. Fate of fertilized human oocytes. *Reprod.*
777 *Biomed. Online* **21**, 732–741 (2010).
- 778 67. Lucotte, E. A. *et al.* Detection of sexually antagonistic transmission distortions in trio
779 datasets. *bioRxiv* (2020) doi:10.1101/2020.09.11.293191.
- 780 68. <http://www.nealelab.is/uk-biobank/>.
- 781 69. Fijarczyk, A. & Babik, W. Detecting balancing selection in genomes: Limits and

prospects. *Mol. Ecol.* **24**, 3529–3545 (2015).

70. Li, H. & Durbin, R. Inference of human population history from individual whole-genome sequences. *Nature* **475**, 493–496 (2011).

71. Connallon, T. & Clark, A. G. A general population genetic framework for antagonistic selection that accounts for demography and recurrent mutation. *Genetics* **190**, 1477–1489 (2012).

72. Connallon, T. & Clark, A. G. Antagonistic versus nonantagonistic models of balancing selection: characterizing the relative timescales and hitchhiking effects of partial selective sweeps. *Evolution* **67**, 908–917 (2013).

73. Janicke, T., Häderer, I. K., Lajeunesse, M. J. & Anthes, N. Darwinian sex roles confirmed across the animal kingdom. *Sci. Adv.* **2**, e1500983 (2016).

74. Eyre-Walker, A. & Keightley, P. D. The distribution of fitness effects of new mutations. *Nat. Rev. Genet.* **8**, 610–618 (2007).

75. Yeaman, S. Local adaptation by alleles of small effect. *Am. Nat.* **186**, S74–S89 (2015).

76. Connallon, T., Débarre, F. & Li, X. Y. Linking local adaptation with the evolution of sex differences. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170414 (2018).

77. Mullon, C., Pomiankowski, A. & Reuter, M. The effects of selection and genetic drift on the genomic distribution of sexually antagonistic alleles. *Evolution* **66**, 3743–3753 (2012).

78. Chang, C. C. *et al.* Second-generation PLINK: Rising to the challenge of larger and richer datasets. *Gigascience* **4**, (2015).

79. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300 (1995).

80. Zietsch, B. P., Kuja-Halkola, R., Walum, H. & Verweij, K. J. H. Perfect genetic correlation between number of offspring and grandoffspring in an industrialized

- human population. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 1032–1036 (2014).
81. Loh, P.-R., Kichaev, G., Gazal, S., Schoech, A. & Price, A. Mixed model association for biobank-scale data sets. *Nat. Genet.* **50**, 906–908 (2018).
82. Randall, J. C. *et al.* Sex-stratified genome-wide association studies including 270,000 individuals show sexual dimorphism in genetic loci for anthropometric traits. *PLoS Genet.* **9**, e1003500 (2013).
83. Coop, G. *et al.* The role of geography in human adaptation. *PLoS Genet.* **5**, e1000500. (2009).
84. Barreiro, L. B., Laval, G., Quach, H., Patin, E. & Quintana-Murci, L. Natural selection has driven population differentiation in modern humans. *Nat. Genet.* **40**, 340–345 (2008).
85. Key, F. M., Fu, Q., Romagne, F., Lachmann, M. & Andres, A. M. Human adaptation and population differentiation in the light of ancient genomes. *Nat. Commun.* **7**, 10775 (2016).
86. Cingolani, P. *et al.* A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff. *Fly (Austin)*. **6**, 80–92 (2012).
87. Connallon, T., Sharma, S. & Olito, C. Evolutionary consequences of sex-specific selection in variable environments: Four simple models reveal diverse evolutionary outcomes. *Am. Nat.* **193**, 93–105 (2019).
88. Crow, J. F. & Kimura, M. *An Introduction to Population Genetics Theory*. (Harper and Row, 1970). doi:10.2307/1529706.
89. Smith, S. R. T. & Connallon, T. The contribution of the mitochondrial genome to sex-specific fitness variance. *Evolution* **71**, 1417–1424 (2017).
90. Morrow, E. H. & Connallon, T. Implications of sex-specific selection for the genetic basis of disease. *Evol. Appl.* **6**, 1208–1217 (2013).

91. Connallon, T. & Jordan, C. Y. Accumulation of deleterious mutations near sexually antagonistic genes. *G3* **6**, 2273–84 (2016).

Acknowledgements

We thank Tom Allison, Isobel Beasley, Isobel Booksmythe, Katja Kasimatis and three anonymous reviewers for helpful comments on previous versions of the manuscript. This work was supported by the Monash Bioinformatics Platform, the UK Biobank (project number 52049 to FR and TC), and funds from the Australian Research Council (to TC).

Author contributions

Conceptualization: FR, LH, TC; Methodology: FR, LH, TC; Investigation: FR, LH, TC; Visualization: FR, LH, TC; Funding acquisition: TC; Project administration: FR, TC; Supervision: FR, LH, TC; Writing – original draft: FR, TC; Writing – review & editing: FR, LH, TC.

Competing Interests

Authors declare no competing interests.

Data accessibility

Within the restrictions of the UK Biobank MTA agreement, metadata will be made available in the Supplementary Materials. Further metadata will be made available on public data sharing platforms, and code will be made available on public github repositories upon publication ([/filruzicka/polygenic_SA_selection_in_the_UK_biobank/](https://github.com/filruzicka/polygenic_SA_selection_in_the_UK_biobank/) and [/lukeholman/ukbb_sexfst/](https://github.com/lukeholman/ukbb_sexfst/)).

Figure 1. Partitioning signals of sex differences in selection among fitness components. A pair of autosomal alleles at a locus are represented by black and white dots, representing male-beneficial and female-beneficial alleles, respectively; $\hat{p}_f$, $\hat{p}_m$, $\hat{p}'_f$, and $\hat{p}'_m$ depict sex-specific frequency estimates for a given allele at different stages of the life cycle (see main text for details). Autosomal allele frequencies are equalised between sexes at fertilisation (left box; females, top; males, bottom), resulting in negligible allele frequency differentiation at this stage of the life cycle. Differentiation between sexes can arise in the sample of adults (middle box) due to sex differences in viability selection among juveniles (orange arrow), and in the projected gametes (right box) due to sex differences in lifetime reproductive success (LRS) among adults (green arrow). Data on sex-specific allele frequencies and LRS thus allow the estimation of sex-differential effects of genetic variants on each fitness component (including overall fitness; purple arrow), despite the absence of allele frequency data among zygotes (left box) and gametes (right box), which are inferred and not directly observed.

Figure 2. Polygenic signals of SA selection: inflation in $\hat{F}_{ST}$ metrics relative to their nulls.

A-C. Proportion of sites (coloured, observed; grey, permuted) falling into each of 100 quantiles of the theoretical null distributions of adult $\hat{F}_{ST}$ (A), reproductive $\hat{F}_{ST}$ (B) and gametic $\hat{F}_{ST}$ (C). Theoretical null data (x-axes) were generated by simulating $\hat{F}_{ST}$ values (nSNPs = 544,635) from the relevant theoretical null distributions (*i.e.*, eqs. [3a-c] for adult, reproductive and gametic $\hat{F}_{ST}$, respectively). In the absence of sex differences in selection, approximately 1% of SNPs should fall into each quantile of the relevant theoretical null (dashed line). **D-F.** Difference between the mean of each metric (adult, reproductive and gametic $\hat{F}_{ST}$, respectively) in observed and empirical null data (coloured histograms) and in observed and theoretical null data (white histograms), across 1,000 bootstrap replicates. Vertical line intersects zero (no difference between observed and null data). Differences are greater than zero across comparisons and metrics for all bootstrap replicates, so all p-values are $p < 0.001$.

Figure 3. Polygenic signals of SA selection: positive associations between $\hat{F}_{ST}$ metrics and MAF. A-C. Mean MAF (coloured, observed; grey, permuted; black, theoretical) among 100 quantiles of adult $\hat{F}_{ST}$ (D), reproductive $\hat{F}_{ST}$ (E) and gametic $\hat{F}_{ST}$ (F). Quantiles for theoretical null data were obtained as in Fig. 2A-C; quantiles for observed and permuted data were obtained from observed and permuted $\hat{F}_{ST}$ distributions for each respective metric. D-F. Difference between the Spearman's rank correlation in the observed and empirical null data (coloured histograms) and between observed and theoretical null data (white histograms), across 1,000 bootstrap replicates, where the correlation is between each $\hat{F}_{ST}$ metric (adult, reproductive and gametic $\hat{F}_{ST}$, respectively) and MAF. Vertical line intersects zero (no difference between observed and null data). Differences are greater than zero across comparisons and metrics for all bootstrap replicates, so all p-values are $p < 0.001$.

Figure 4. Candidate SA polymorphisms are more likely to be functional and contribute to trait variation. A-C. Genic enrichment (coloured, observed; grey, permuted; black, theoretical) in each of 100 quantiles of adult $\hat{F}_{ST}$ (A), reproductive $\hat{F}_{ST}$ (B) and gametic $\hat{F}_{ST}$ (C). Quantiles for theoretical null data, permuted data and observed data were obtained as in Fig. 3A-C. Genic enrichment is defined as the proportion of sites in a given quantile that are genic divided by the proportion of all sites that are genic. Loess curves ($\pm$ SE) are presented for visual emphasis. D. Relationships between five metrics of sex-differential selection and candidate loci for various UK Biobank phenotypes (as analysed by the Neale laboratory), using an LD-pruned set of imputed SNPs. Each tile compares the mean value of the relevant metric of sex-differential selection between variants that are associated and not associated with the

908 focal phenotype, with colours denoting the ratio of means (with ratios >1 indicating *e.g.* higher
 909 $\hat{F}_{ST}$ among associated than non-associated variants), and statistical significance assessed using
 910 a Mann-Whitney test (*, FDR-adjusted $p < 0.05$; ~, unadjusted $p < 0.05$).
 911

Figure 5. No evidence that candidate SA polymorphisms are older than average. Mean age of the alternative (*i.e.*, non-reference) allele (coloured, observed; grey, permuted; black, theoretical) across 100 quantiles of adult $\hat{F}_{ST}$ (A), reproductive $\hat{F}_{ST}$ (B) and gametic $\hat{F}_{ST}$ (C). Quantiles for theoretical null data, permuted data and observed data were obtained as in Fig. 3A-C. All panels correct for ascertainment bias of alternative allele frequencies. For visualisation purposes, this was done by averaging, in each $\hat{F}_{ST}$ quantile, mean allele age across 20 quantiles of alternative allele frequency in the UK Biobank (such that alternative allele frequency is approximately equal across $\hat{F}_{ST}$ quantiles).

922

923