nature human behaviour

ARTICLES

https://doi.org/10.1038/s41562-021-01168-8

Check for updates

Genomic evidence consistent with antagonistic pleiotropy may help explain the evolutionary maintenance of same-sex sexual behaviour in humans

Brendan P. Zietsch[1]✉, Morgan J. Sidari[1], Abdel Abdellaoui[2], Robert Maier[3], Niklas Långström[4], Shengru Guo[5], Gary W. Beecham[5], Eden R. Martin[5], Alan R. Sanders[6,7] and Karin J. H. Verweij[2]

Human same-sex sexual behaviour (SSB) is heritable, confers no immediately obvious direct reproductive or survival benefit and can divert mating effort from reproductive opportunities. This presents a Darwinian paradox: why has SSB been maintained despite apparent selection against it? We show that genetic effects associated with SSB may, in individuals who only engage in opposite-sex sexual behaviour (OSB individuals), confer a mating advantage. Using results from a recent genome-wide association study of SSB and a new genome-wide association study on number of opposite-sex sexual partners in 358,426 individuals, we show that, among OSB individuals, genetic effects associated with SSB are associated with having more opposite-sex sexual partners. Computer simulations suggest that such a mating advantage for alleles associated with SSB could help explain how it has been evolutionarily maintained. Caveats include the cultural specificity of our UK and US samples, the societal regulation of sexual behaviour in these populations, the difficulty of measuring mating success and the fact that measured variants capture a minority of the total genetic variation in the traits.

Across human societies, a fraction of men and women (around 2–10%) report engaging in sex with same-sex partners (same-sex sexual behaviour (SSB)), either exclusively or in addition to sex with opposite-sex partners[1–4]. SSB runs in families[5], and is concordant more often in genetically identical twin pairs than in non-identical twin pairs or siblings, which suggests a genetic influence on the trait[4,6]. SSB is also widespread in the animal kingdom, and has been studied across most major clades[7]. Because SSB confers no immediately obvious direct reproductive or survival benefit and can divert mating effort away from reproductive opportunities, its widespread occurrence across the animal kingdom and human cultures raises questions for evolutionary biology[8]. If, in the evolutionary past, SSB has on average been associated with lower rates of reproduction, then in the absence of countervailing forces, genetic variants associated with SSB would be expected to reduce in frequency until the behaviour disappeared from the population (see simple evolutionary simulations in Fig. 1a and details in Supplementary Methods and Results and Supplementary Figs. 1–4). The fact that SSB exists across human and animal populations is thus often discussed as a Darwinian paradox[7,9–17].

One possible explanation is that SSB has not been associated with reduced rates of reproduction. But human studies have found SSB to be associated with much lower reproductive rates in Western countries[18,19], even among heterosexually married individuals[20], and in Indonesia[21] and a traditional Samoan community[22]. In these studies, reproductive rates were reduced whether behaviour/attraction was oriented predominantly to the same or both sexes[19,21,22] (especially see fig. 1b in ref. [19]). Some scholars have concluded, based on these findings and anecdotal accounts in the anthropological literature, that same-sex attraction is cross-culturally associated with a large fitness cost[22]. However, we note that, given that SSB has been subject to varying degrees of legal, medical and societal regulation, existing findings may not accurately reflect associations between reproductive rates and SSB in our evolutionary past.

Another possibility, which we investigate here, is that alleles associated with SSB confer an advantage to individuals who only engage in opposite-sex sexual behaviour (OSB individuals; see Box 1), thereby offsetting the fitness cost of the relevant alleles[13,23]. In evolutionary simulations in the Supplementary Results, we show that such a situation can maintain substantial rates of SSB in the population (Fig. 1b), despite its fitness cost. The general phenomenon of alleles having multiple effects with opposing consequences with respect to fitness is called antagonistic pleiotropy and may be widespread[24]. Using data from identical and non-identical twins, Zietsch et al.[13] provided indirect evidence consistent with the countervailing advantage of SSB genes, in that the genetic factors associated with SSB confered a mating advantage in OSB individuals (in terms of lifetime number of opposite-sex sexual partners; see Supplementary Discussion). A direct test of this hypothesis was not previously possible because the effects of specific genetic variants on SSB had not been characterized (early attempts to identify such effects used sample sizes and methodologies that are now regarded as underpowered for detecting effect sizes typical for complex traits[25,26], and their findings have not consistently replicated).

[1]Centre for Psychology and Evolution, School of Psychology, University of Queensland, Brisbane, Queensland, Australia. [2]Department of Psychiatry, Amsterdam UMC location AMC, University of Amsterdam, Amsterdam, the Netherlands. [3]Department of Genetics, Harvard Medical School, Boston, MA, USA. [4]Department of Medical Epidemiology and Biostatistics, Karolinska Institutet, Stockholm, Sweden. [5]Department of Human Genetics, University of Miami, Miami, FL, USA. [6]Department of Psychiatry and Behavioral Sciences, NorthShore University HealthSystem Research Institute, Evanston, IL, USA. [7]Department of Psychiatry and Behavioral Neuroscience, University of Chicago, Chicago, IL, USA. ✉e-mail: zietsch@psy.uq.edu.au

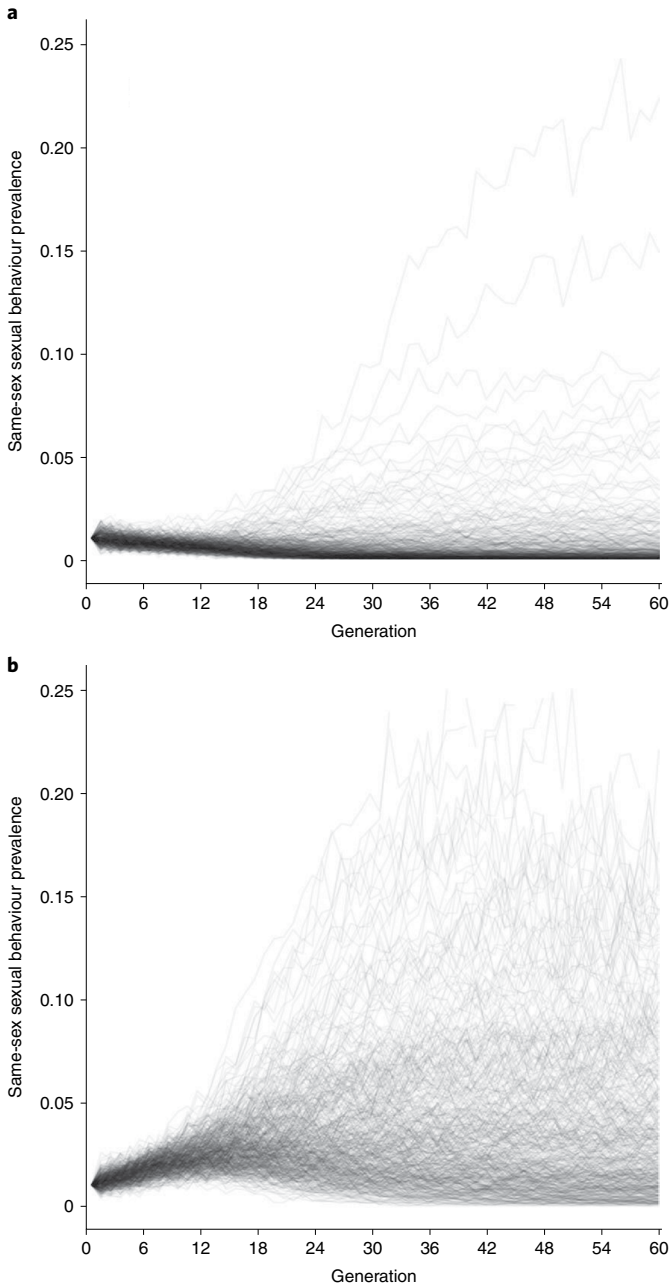

Fig. 1 | Evolutionary simulations. a,**b**, Evolutionary simulations of prevalence of SSB over 60 generations under no antagonistic pleiotropy (**a**) and antagonistic pleiotropy (**b**). See Supplementary Method and Results for more details.

Extremely large samples of genotyped individuals have now been collected, providing the opportunity to estimate the association of a trait with genetic variants across the entire genome. In a recent paper[19], we analysed genome-wide association study (GWAS) results from 408,995 participants from the UK Biobank (aged 40–69 years) and 55,594 from 23andMe (mean age 51.3 years) who reported about their sexual behaviour. The primary phenotype in terms of SSB was a dichotomous, self-report measure of whether respondents had ever had sex with someone of the same sex (here termed 'SSB individuals'; see Box 1) or had only had sex with opposite-sex others (here termed 'OSB individuals'). We emphasize that this variable captures behaviour and not any specific sexual identities. This variable maximizes effective sample size and statistical power, and thus serves as our primary measure in the current study; however, although any SSB confers no immediately obvious direct reproductive or survival benefit and may divert mating effort away from reproductive opportunities as discussed above, the apparent evolutionary paradox is stronger for higher ratios of SSB to OSB. Therefore, we explore more stringent definitions of the variable in sensitivity analyses. UK Biobank participants also reported their lifetime number of opposite-sex and same-sex partners. In testing for antagonistic pleiotropy in associated genetic variants, we want to test for a mating advantage in OSB individuals, so our primary criterion variable is lifetime number of opposite-sex sexual partners among individuals who have had opposite-sex partners but never a same-sex partner (final $N = 358{,}426$). In the Supplementary Discussion, we argue why number of opposite-sex sexual partners is more useful as a current marker of a historical mating advantage than is number of children, though it still has important limitations. Similarly, our approach is significantly limited by our reliance on data gathered from UK and US participants, the majority of whom were born at a time when homosexuality was to some extent criminalized in their country of birth, or only very shortly after decriminalization, as discussed further in Methods and Discussion.

To test our hypothesis that genetic variants associated with SSB may also be associated with a mating advantage in OSB individuals, we estimated genome-wide single-nucleotide polymorphism (SNP) effects for (a) SSB[19], and (b) number of lifetime opposite-sex partners among OSB individuals, and tested the extent to which these effects correlate between the two traits.

Box 1 | Notes on terminology

We use the terms 'OSB individuals' (for those who reported at least one opposite-sex sexual partner and had never had a same-sex partner) and 'SSB individuals' (for those who have had at least one same-sex sexual partner) for convenience and brevity, not to infer the sexual identities of these individuals, nor to imply that one group is the norm and one group is 'other'. We also wish to make clear that we regard SSB as a normal part of the diversity of human sexual behaviour, not as a disorder or dysfunction. At the same time, SSB is associated with elevated vulnerability to physical and mental health problems, so for the fullest understanding of health we believe it is important to include this behavioural variation in evolutionary and biological analyses of human traits. Answers to frequently asked questions can be found in the Supplementary Information.

Results

We ran GWASs for number of opposite-sex sexual partners in men and women separately and in the combined sample (see Supplementary Figs. 5 and 6 for the Manhattan and QQ plots and Supplementary Table 1 for an overview of significant SNPs). In aggregate, all available SNPs (that is SNP-based heritability[27]) captured 7% (95% CIs 6% to 8%) of the variation for women and 9% (95% CIs 8% to 10%) for men. For SSB, the SNP-based heritability was estimated at 8–25%, where the range reflects differing estimates using different analytic methods (details in ref. [19]). For both variables, this SNP heritability was not driven by a small number of genes of large effect but rather a very large number of genes of very small effect spread across the genome. This finding is evidenced by two types of observations. First, even the most strongly associated SNPs individually accounted for tiny proportions of the total genetic variation. Second, for both variables, there was significant correlation between the length of a chromosome and the heritability it explained, consistent with many small genetic effects spread evenly across the genome (ref. [19] and Fig. 2).

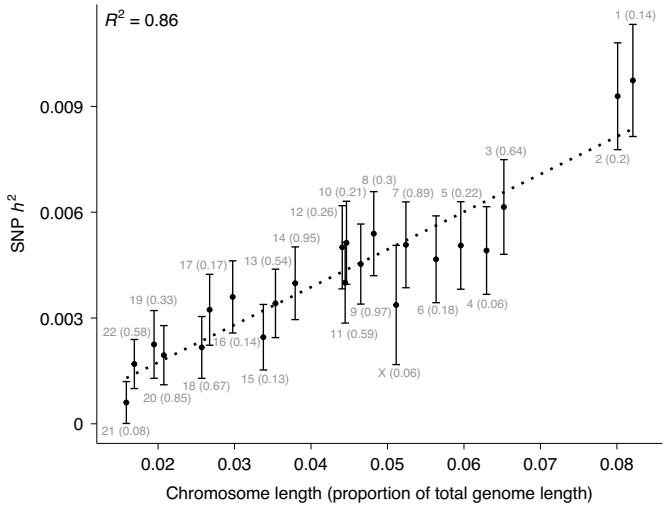

Fig. 2 | Per-chromosome SNP-based heritability. SNP-based heritability estimates for number of opposite-sex sexual partners per chromosome, ordered by chromosome length as percentage of total genome. Error bars indicate 95% confidence intervals. Numbers at each dot indicate the chromosome number and the *P* value for the SNP-based heritability. The dotted line represents the linear regression fitted using the plotted values.

Given that the genetic architecture comprises many small effects spread across the genome, our test of the antagonistic pleiotropy hypothesis does not involve analysis of individual genetic variants but rather analysis of the aggregate effect of all measured variants. We used two different methods to conduct these analyses (see Methods for details). First, using cross-trait linkage disequilibrium (LD) score regression[27], we estimated the genetic correlation between the two traits using the genome-wide SNP effect estimates obtained from the GWASs. Consistent with our hypothesis, there was a significant and positive genetic correlation of SSB with lifetime number of opposite-sex partners in OSB individuals for both men (r_g = 0.31, 95% CI 0.21 to 0.42) and women (r_g = 0.73, 95% CI 0.61 to 0.85; sex difference $\chi^2_{(1)} = 29.84$, $P < 0.001$). (For the corresponding cross-sex genetic correlations, refer to Supplementary Table 2.) In the UK Biobank, sensitivity analyses using more stringent definitions of SSB (Table 1) showed directionally the same effects; the strengths varied with the different definitions, though due to low power in these analyses the effect size estimates were imprecise, and two were nonsignificant.

The second method used the genome-wide SNP effects from UK Biobank to build polygenic scores[28] for SSB in the independent Add Health sample[29] (N = 4,414; Table 2). These polygenic scores are genotype-based scores for each individual that estimate the extent to which their genotype contains variants associated with SSB. We showed that this polygenic score was positively correlated with lifetime number of opposite-sex sexual partners for OSB individuals, in both men and women. Note that effect sizes for these polygenic score associations are not biologically meaningful, as they depend on the size of the discovery sample.

The Add Health data also included ratings of physical attractiveness by interviewers, which we analysed because of potential relevance to the mating advantage hypothesized under antagonistic pleiotropy. Among heterosexuals in Add Health, the polygenic score for SSB correlated positively with observer-rated physical attractiveness, though in the sex-specific analyses the *P* value only barely passed the nominal significance threshold in males and was nonsignificant in females (Table 2), so the findings should be treated as tentative until replicated.

Overall, these findings are largely consistent with the antagonistic pleiotropy hypothesis that alleles that are associated with SSB are also associated with a mating advantage in OSB individuals. However, we do not know what underlies this relationship. To test different possibilities, we first identified personality-type traits genetically correlated in at least one sex with both SSB and number of opposite-sex partners among OSB individuals (Table 3A). We then applied genomic structural equation modelling[30] to estimate the residual genetic correlation between SSB and OSB, partialling out these personality traits. We found that the personality variables 'risk-taking propensity' and 'openness to experience' jointly capture part, but not all, of the genetic correlation (r_g decreases from 0.50 (95% CIs 0.41 to 0.59) to 0.30 (95% CIs 0.22 to 0.38); see Table 3B and Methods for further explanation), consistent with the possibility that these two personality traits may contribute to the mating advantage of SSB-associated alleles in OSB individuals.

Table 1 | Sensitivity analyses for association between SSB and number of opposite-sex sexual partners in OSB individuals

(A) In UK Biobank, genetic correlations (r_g) of SSB (using different definitions of the SSB group according to different thresholds of the proportion of number of same-sex partners over total number of lifetime sexual partners) and number of opposite-sex partners in OSB individuals

Proportion of same-sex partners/total lifetime sexual partners	Genetic correlations with number of opposite-sex sexual partners in OSB individuals					
	All		Male		Female	
	r_g (SE)	*P* value	r_g (SE)	*P* value	r_g (SE)	*P* value
Less than one-third same-sex partners	0.64 (0.06)	2.3×10^{-23}	0.54 (0.10)	1.5×10^{-7}	0.74 (0.11)	1.8×10^{-12}
Between one-third and two-thirds same-sex partners	0.44 (0.14)	2.1×10^{-3}	0.14 (0.12)	0.23	0.44 (0.13)	4.1×10^{-4}
More than two-thirds same-sex partners	0.38 (0.13)	2.8×10^{-3}	0.26 (0.10)	0.01	0.35 (0.18)	0.06

(B) Association of polygenic score for exclusively SSB with number of opposite-sex sexual partners among OSB individuals in UK Biobank using a cross-validation approach ($N_{males+females}$ target sample of 40,900)

Sex	Cross-validated β	*P* value
Women	0.027	0.069
Men	0.053	0.058
Men and women combined	0.032	0.013

Note that confidence intervals were not obtained, as this was too computationally demanding.

Table 2 | Associations of polygenic scores for SSB with number of sexual partners among OSB individuals, and with rated attractiveness, in the Add Health sample ($N_{males+females}$ = 4,278)

	Number of sexual partners in OSB individuals					Attractiveness scale				
Sex	*N*	Mean (s.d.)	β (s.e.)	*P* value	ΔR^2 (95% CI)	*N*	Mean (s.d.)	β (s.e.)	*P* value	ΔR^2 (95% CI)
Men + women	4,278	12.40 (19.98)	1.39 (0.317)	1.45×10^{-5}	0.4% (0.2–0.8%)	4,414	3.47 (0.81)	0.02 (0.009)	0.026	0.1% (0.0–0.3%)
Women	2,228	9.34 (12.39)	0.67 (0.215)	1.88×10^{-3}	0.3% (0.0–0.8%)	2,281	3.49 (0.85)	0.01 (0.014)	0.485	0.0% (0.0–0.3%)
Men	2,050	15.72 (25.40)	2.01 (0.621)	1.26×10^{-3}	0.8% (0.3–1.9%)	2,133	3.41 (0.77)	0.01 (0.005)	0.045	0.0% (0.0–0.2%)

Sex-specific analyses are done using the corresponding sex-specific GWAS results for SSB.

Discussion

Our findings are consistent with the existence of a mating advantage for OSB individuals who carry genes that are associated with SSB. We cannot say how much such a mating advantage would have increased overall fitness during evolutionary history. In the populations from which our samples are drawn, modern contraception and in vitro fertilization have decoupled sexual behaviour from reproduction in a manner that was not the case during our evolutionary history, so current associations between number of mates and reproductive success (number of children) are unlikely to be informative (Supplementary Discussion). Furthermore, sexual behaviour has been subject to specific societal constraints in recent history. For example, most of the UK sample were born at a time when SSB was criminalized in that country, and OSB is likely affected by social norms concerning number of partners for women and men. Importantly, the results presented here cannot speak to levels of sexual activity among different groups, nor to any genetic associations with number of same-sex sexual partners. It is also important to note that our mate quantity measure lacks information about relationship lengths and mate quality, which are especially relevant to women's fitness. Nevertheless, our evolutionary reasoning is based on an assumption that the genes associated with number of mates today were under selection in the evolutionary past (Supplementary Discussion).

It is unclear what processes underlie the genetic correlation of SSB with number of mates among OSB individuals. We found evidence consistent with the roles of openness and risk-taking propensity, and in men, physical attractiveness, but replication and extension of these findings is needed. Zietsch et al.[13] previously implicated continuous gender identity: more feminine self-concept in men and more masculine self-concept in women were associated with having at least some attraction to the same-sex and, among OSB individuals, with having had more opposite-sex partners. We do not have access to large-scale GWAS results for continuous gender identity, so could not assess its role here. For the same reason, we could not assess the roles of other potentially relevant traits, such as sex drive, orientation towards short- versus long-term relationships and charisma. It is also unclear whether similar factors underlie the association in males and females; the significantly larger genetic correlation in females than in males suggests that different factors may be involved, but alternatively the same factors may drive a stronger association in females than in males. For example, greater sexual motivation may manifest more in women's number of sexual partners than in men's, because cross-culturally, men are more open to uncommitted sex than are women, meaning women's sexual desires are less constrained by men's willingness than vice versa[31]. Note, in any case, that the antagonistic pleiotropy hypothesis makes no prediction about the relative strength of the association in each sex. For example, if number of mates were more relevant to men's than women's fitness, this would not lead to an expectation of a stronger male genetic correlation between SSB and number of mates in OSB individuals.

There are other limitations and uncertainties. One is that common SNPs account for around 8–25% of the variance in SSB, whereas pedigree analysis of the UK Biobank data suggest that the overall genetic component is 32% (ref. [19]). That is, the additive effects of the measured common variants capture only a portion of the overall heritability, as is typical for complex traits. It follows that our findings only apply to this component of the genetic influence. Rare variants, which are poorly captured in our study, probably contribute to the 'missing' heritability[32] and may be subject to different evolutionary processes. In particular, mutation–selection balance may be involved, whereby fitness-reducing genetic variants are removed from the population by selection at a rate similar to that at which new fitness-reducing variants arise (see Supplementary Discussion for more information about this possibility). Further, though we found that common variants associated with SSB are also associated with a higher number of opposite-sex partners in OSB individuals, it is unlikely that this effect applies to every variant in question. Another issue is that the genetic influence on SSB is heterogeneous across varying ratios of same-sex to opposite-sex partners[19]. To the extent that we could explore that heterogeneity here, it appears that the antagonistic pleiotropy effect may apply to both bisexuality and homosexuality, though there was some variation in significance of the effect across definitions of SSB. More precise characterization of this issue will require larger samples with more nuanced data on same-sex and opposite-sex attraction and behaviours. Another issue is theoretical uncertainty about the extent to which antagonistic pleiotropy can actively maintain genetic variation. Early models showed that it could, in principle, maintain genetic variation in one-locus[33], two-locus[34] and multi-locus[35] systems. Later work suggested that the conditions under which variation can be maintained indefinitely appear quite restrictive[36], though it can clearly slow the removal of alleles by making them (more) selectively neutral[37,38]. These analyses did not model the particular form of antagonistic pleiotropy that we predict for SSB—that is, a polygenic threshold trait that only has negative fitness consequences above the threshold and a pleiotropic advantage that only manifests below the threshold—which is why we provide simple modelling (see Supplementary Methods and Results and Supplementary Figs. 1–4). Finally, we acknowledge the possibility of ascertainment and self-report biases, the potential effects of which are hard to predict or account for. Given these various limitations, our findings cannot provide a complete picture of the evolutionary maintenance of SSB.

Other evolutionary models of SSB have been proposed. One of the most prominent proposes that the fitness cost of genes predisposing to male SSB are offset by increased fecundity in female carriers, which the theory's proponents measure with number of children[12]. In contrast to this sexually antagonistic selection hypothesis, we found no significant genetic correlation between male SSB and female number of children[39] (r_g = 0.01; 95% CIs −0.14 to 0.16; see Supplementary Table 2). We note, though, that it is unclear to what extent number of children (that is, fertility) in modern societies relates to fecundity (that is, natural ability to reproduce), given contraception-based family planning and the availability of in vitro fertilization. Another

Table 3 | (A) Genetic correlations of SSB and number of opposite-sex sexual partners in OSB individuals with openness to experience and risk-taking behaviour, as estimated in LD score regression

		Genetic correlation with SSB				Genetic correlation with number of opposite-sex sexual partners in OSB individuals			
		Men		Women		Men		Women	
Phenotype	*N*	r_g (s.e.)	*P* value	r_g (s.e.)	*P* value	r_g (s.e.)	*P* value	r_g (s.e.)	*P* value
Openness to experience[54,55]	76,551	0.135 (0.060)	2.37×10^{-2}	0.312 (0.066)	2.72×10^{-6}	0.262 (0.045)	7.14×10^{-9}	0.362 (0.044)	9.77×10^{-17}
Risk-taking behaviour (UK-Biobank)	~440,000	0.224 (0.047)	1.77×10^{-6}	0.402 (0.052)	9.54×10^{-15}	0.557 (0.029)	3.58×10^{-84}	0.490 (0.029)	8.43×10^{-66}

(B) Genetic correlations (95% confidence intervals) between SSB and number of opposite-sex sexual partners in OSB individuals with and without correcting for the genetic contribution of risk-taking and openness to experience. Results obtained from a multivariate regression model using Genomic SEM. The correlations obtained from the adjusted models can be interpreted as semi-partial correlations

	Genetic correlation	Semi-partial genetic correlations, when adjusted for genetic covariance with:		
	Unadjusted	Risk-taking	Openness to experience	Risk-taking and openness to experience
Men	0.31 (0.20–0.41)	0.18 (0.09–0.28)	0.27 (0.16–0.37)	0.18 (0.08–0.28)
Women	0.71 (0.59–0.83)	0.52 (0.42–0.63)	0.60 (0.48–0.72)	0.51 (0.39–0.60)
Men and women	0.50 (0.41–0.59)	0.31 (0.23–0.39)	0.42 (0.33–0.52)	0.30 (0.22–0.38)

line of evidence against the sexually antagonistic selection model pertains to its prediction that any genetic variants influencing SSB will be concentrated on the X chromosome[40,41]: Ganna et al.'s GWAS of 479,420 individuals revealed no significant SNP associations on the X chromosome, and all X-chromosome SNPs in aggregate did not account for any more of the heritability in SSB than did those on the average autosome[19]. Indeed, one mathematical model of sexually antagonistic selection[40] proposes that only two loci (both on the X chromosome) influence SSB, which is clearly contrary to the highly polygenic architecture shown by Ganna et al.[19]. Other evolutionary possibilities, which we cannot assess with our data, are discussed in Supplementary Discussion.

In conclusion, we found genetic evidence consistent with predictions of antagonistic pleiotropy as an explanation for the evolutionary maintenance of SSB in the population. Many uncertainties remain, and caution should be exercised in interpreting these findings. Further research is needed to more fully characterize the genetic effects, to understand what drives the link between SSB and number of mates in OSB individuals and to evaluate additional evolutionary possibilities.

Methods

This study has received ethical approval from the University of Queensland Human Research Ethics Committee (approval number 2017001005). The participants of this study were sourced from the UK Biobank and The National Longitudinal Study of Adolescent to Adult Health (Add Health). UK Biobank has received ethical approval from the National Health Service North West Centre for Research Ethics Committee (reference 11/NW/0382). Add Health GWAS study protocols and consent from human subjects were approved by the Institutional Review Board of the University of North Carolina at Chapel Hill (IRB #12-1479). Informed consent was obtained from all participants.

The hypothesis tested in this paper and our analysis plan were pre-registered at the Open Science Framework (https://osf.io/357tn/). We split the study into two papers: one on the genetics of SSB[19] and the current one focusing on the evolutionary basis of SSB. Details of the pre-registered research plan and deviations from it can be found in the Supplementary Methods.

This study relied on existing, available individual-level datasets and GWAS summary statistics. Sample sizes were therefore not chosen by the authors but determined by UK Biobank and Add Health. For continuous variables, the data distribution was assumed to be normal, but this was not formally tested due to the large sample sizes[42]. All statistical tests performed in the study were two tailed.

Description of data. *UK Biobank.* The UK Biobank resource is a population-based cohort of approximately 500,000 participants born between 1934 and 1971 (aged between 37 and 73 years at assessment) recruited in the United Kingdom between 2006 and 2010 (ref. [43]). Invitations to participate were sent out to approximately 9.2 million individuals who lived within 25 miles of one of the 22 assessment centres in England, Wales and Scotland. The participants provided DNA samples and completed extensive questionnaire data including questions related to sexual behaviour (414,751 individuals with genetic data passed quality control and answered sex-related questions).

The UK Biobank ascertainment strategy was designed to capture sufficient variation in socioeconomic, urban–rural and ethnic background[44]. The participation rate, however, was 5.5% and was consistent with the 'healthy volunteer' effect. Participation was biased towards older, more healthy, female residents, and participants living in less socioeconomically deprived areas than non-participants[45]. Moreover, in this study, we only used data from individuals with European ancestry.

From the UK Biobank we obtained two phenotypes:

1. 'Number of lifetime opposite-sex sexual partners in OSB individuals' was determined with an open question about the total number of sexual partners during the participant's lifetime. Individuals who reported having had sexual intercourse with someone of the same sex (3.3%) or who had never had a sexual relationship (0.9%) were set at missing. We also excluded participants with more than 100 sexual partners, since these outliers could increase phenotypic heterogeneity and have a disproportionate influence on our findings; this reduced our sample by 1035 males and 79 females (0.5% and 0.04%, respectively). Data were available for 162,183 males and 196,243 females. In our final sample, OSB males had on average 21.1 (s.d. 26.1) sexual partners, and OSB females had on average 5.5 (s.d. 6.2) sexual partners.
2. 'SSB' (ever versus never having had sex with someone of the same sex), was based on responses to the question 'Have you ever had sexual intercourse with someone of the same sex?' And if the participant activated the Help button they were shown the message: 'Sexual intercourse includes vaginal, oral or anal intercourse.' Participants who reported to have never had a sexual relationship were set at missing. In the final dataset, data were available for 188,825 males and 220,170 females.

Genotyping and imputation. We used genotype data from the May 2017 release of imputed genetic data from UK Biobank. The quality control and imputation were done by UK Biobank and have been described elsewhere[19]. Briefly, genotyped variants were filtered based on batch effects, plate effects, departures from Hardy–Weinberg equilibrium, genotype platform and discordance across control replicates. Participant samples were excluded based on missing rate, inconsistencies in reported versus genetic sex and heterozygosity based on a set of 605,876 high-quality autosomal markers. Imputation was performed using IMPUTE4 with the HRC UK10K and 1,000 Genomes phase 3 dataset used as the reference set. For the X chromosome, Hardy–Weinberg equilibrium was calculated on females only; see ref. [19] for how the allele frequency was calculated.

For ethnicity definition, we used *K*-means clustering[19] to identify four clusters on the first four principal components (PCs) of the genetic data provided by UK Biobank. The first four PCs were chosen because the fifth shows substantial spread

within the self-identified British population. The fourth cluster was the basis for identifying individuals as 'White-European'. The 'White' cluster completely contained the white British subset as defined by UK Biobank. Individuals in this cluster who self-reported as different ancestry group were dropped from the analyses.

The National Longitudinal Study of Adolescent to Adult Health (Add Health). Add Health originated as an in-school survey of a nationally representative sample of US adolescents enrolled in grades 7 through 12 during the 1994–1995 school year[46]. Respondents were born between 1974 and 1983, and a subset of the original Add Health respondents has been followed up with in-home interviews. Phenotypes included in this study were obtained in wave 4 (2007–2009), where the mean birth year of respondents is 1979 (s.d. 1.8 years) and the mean age at assessment is 29.0 years (s.d. 1.8 years). Two phenotypes were obtained from Add Health:

1. 'Number of opposite-sex sexual partners in OSB individuals' is a continuous measure of reported lifetime number of sexual partners among individuals who only had opposite-sex partners. Data were available for 2,228 females and 2,050 males.
2. 'Rated attractiveness': interviewers were asked to rate the physical attractiveness of interviewees. Response options included (1) very unattractive, (2) unattractive, (3) about average, (4) attractive and (5) very attractive. We do not know the sex of the interviewer that rated each participant's attractiveness, but the raters' sex was random with respect to gender and sexuality of participants. Overall, 79% of interviewers were women, 18% men and 3% unknown. Other studies have shown high consistency between male and female ratings of attractiveness (for example, ref. [47]; male raters correlated 0.94 with average of all raters, female raters correlated 0.92 with average of all raters). Data were available for 2,281 females and 2,133 males.

Genotyping and imputation. Genotyping was performed at the Expression Analysis Laboratory in Research Triangle Park, NC, using Illumina's Human Omni1-Quad-BeadChip[29]. After imputing the genetic data to the Haplotype Reference Consortium (HRC)[48] using the Michigan Imputation Server[49], only HapMap 3 variants with a call rate above 98% and a minor allele frequency (MAF) above 1% were included. Analyses were limited to individuals of European ancestry, and cryptically related individuals were dropped from analyses.

Use of existing GWAS summary-level data. We used GWAS summary statistics from ref. [19] for two variables (SSB and exclusively same-sex sexual behaviour), and from ref. [39] for number of children.

1. 'SSB' (ever versus never having had sex with someone of the same sex). GWAS results were based on the meta-analytic sample of UK Biobank and 23andMe (N = 477,522 individuals)[19].
   A sample description of the UK Biobank dataset has been provided above. Regarding 23andMe sample: only US individuals of European ancestry were included, and the mean age of the 23andMe sample in the source GWAS was 51.3 years (s.d. 16.0 years). Note that the 23andMe sample had very high rates of SSB, probably due to self-selection of participants to answering questions on sexual behaviour. Individuals who engage in SSB may be more likely to fill out the 'Sexual orientation survey'.
2. 'Exclusively same-sex sexual behaviour', which differentiates between participants who had exclusively had sex with same-sex sexual partners versus those who had exclusively had sex with opposite-sex sexual partners. GWAS results were based on data from the UK-Biobank (N = 1,766 exclusively SSB males and 180,431 OSB males and 693 exclusively SSB females and 214,062 OSB females).
3. 'Number of children' (ever born) was determined based on self-reported number of children[39]. This phenotype was either asked directly (for example 'How many children do you have?') or calculated based on survey questions (such as pregnancy histories and outcomes, number of deliveries). When possible to distinguish, only the number of live-born biological children was included. Only participants older than 55 years (males) or 45 years (females) were included, to capture individuals who were very likely to have completed reproduction.

Please note. Our study is based on data obtained from UK (UK Biobank) and US (23andMe and Add Health) participants, the majority of whom were born at a time when homosexuality was to some extent criminalized in their country of birth, or only very shortly after decriminalization. Also, until recently, homosexuality was still classified as a psychiatric disorder. At that time, but even now, social stigmas are attached to both same-sex behaviour and high numbers of sexual partners in both countries. These historical and social contexts present significant limitations to the conclusions we can draw from these data; our findings may not accurately reflect associations between reproductive rates and SSB in our evolutionary past. Nonetheless, these are currently among the most appropriate large-scale datasets in which we could investigate our research questions.

Analyses. *GWAS analysis.* For the GWAS of number of opposite-sex sexual partners among OSB individuals in UK Biobank, we ran linear mixed models implemented in BOLT-LMM[50] to account for cryptic population structure and relatedness. We adjusted for sex, year of birth, year of birth squared, ten genetic PCs, genetic relatedness, genotype platform and GWAS chip. The genetic relationship matrix only included autosomal genetic variants that were common (MAF > 1%), passed quality control in all 106 batches and were present on both genotyping arrays.

We used the FUMA pipeline[51] to identify independent loci. We used pre-calculated LD structure based on the European 1,000 Genome panel to identify genome-wide significant SNPs independent from each other at $r^2 < 0.6$. We identified independent lead SNPs as those independent from each other at $r^2 < 0.1$. If LD blocks of independent significant SNPs were closely located to each other (<250 kb apart, based on the most right and left SNPs within each LD block), they were merged into one genomic locus.

SNP-based heritability. We used LD score regression[27] to estimate the proportion of variance in number of opposite-sex sexual partners among OSB individuals that could be explained by the aggregate effect of all analysed SNPs (h^2_{SNPs}). The method relies on the idea that an estimated SNP effect includes effects of all SNPs in LD with that SNP. A SNP that tags many other SNPs will, on average, have a higher probability of tagging a causal variant than an SNP that tags only a few other SNPs. Therefore, for polygenic traits, SNPs with a higher LD score have on average stronger effect sizes than SNPs with lower LD scores. When regressing the effect size obtained from the GWAS against the LD score for each SNP, the slope of the regression line provides an estimate of the proportion of variance accounted for by all the SNPs. We included SNPs available in the HapMap 3 reference panel (N = 1,217,312). LD scores were based on the HapMap 3 reference panel, restricted to European populations[27].

Additionally, we estimated the per-chromosome heritability for number of opposite-sex sexual partners per chromosome. We used BOLT-LMM[50] to estimate the per-chromosome heritability and the corresponding standard error for the 22 autosomes and the X chromosome. We tested whether the observed heritability was significantly different from what is expected given the chromosomal length. For more details on this method, see ref. [19].

Genetic correlation between traits. We used cross-trait LD score regression[52] to estimate the genetic covariation between traits using GWAS summary statistics. The genetic covariance is estimated using the slope from the regression of the product of z scores from two GWAS studies on the LD score. The estimate represents the genetic correlation between the two traits based on all polygenic effects captured by SNPs. We used standard LD scores as provided by Bulik-Sullivan et al.[52] based on the 1,000 Genomes reference set, restricted to European populations.

With cross-trait LD score regression, we estimated the genetic correlation of (1) SSB and number of opposite-sex sexual partners in OSB individuals, (2) SSB and number of children and (3) SSB and number of opposite-sex sexual partners in OSB individuals with personality traits (openness to experience and risk-taking behaviour).

Cross-trait polygenic prediction. We performed polygenic score analyses to test whether genetic association with SSB is associated in OSB individuals with having more opposite-sex sexual partners and being rated as more physically attractive. Based on the results from the GWAS of SSB[19], we generated polygenic scores for SSB in the Add Health sample. Polygenic scores are constructed by multiplying the number of copies of the effect allele at each SNP by the effect size of this SNP as obtained from the GWAS on SSB, and summing across SNPs. Accordingly, a polygenic score is a single quantitative estimate of genetic association with a given trait.

The polygenic scores were constructed in LDpred[28], a method shown to have greater predictive accuracy than the conventional risk prediction approach involving LD pruning followed by P value thresholding. LDpred considers the genetic architecture by accounting for LD among the SNPs in creating the polygenic scores. We used the genotyped data from the Add Health prediction cohort to create the LD reference file. We used only HapMap 3 variants with call rate > 98% and MAF > 1% to construct the polygenic scores, and we limited the analyses to individuals with European ancestry. Polygenic scores were calculated with expected fraction of causal genetic markers set at 100%. In total, we used 1,177,001 HapMap 3 variants to construct the polygenic scores. We then used Plink[53] to multiply the genotype probability of each variant by the corresponding LDpred posterior mean over all variants. We created polygenic scores for SSB in the combined dataset (males and females), as well as separately for males and females, based on the corresponding sex-specific summary statistics. We then determined the association of the polygenic score for SSB with number of opposite-sex sexual partners in OSB individuals and with their rated attractiveness. Prediction accuracy was based on an ordinary least-squares regression of the outcome phenotype on the polygenic score and a set of standard controls, which include birth year, sex, an interaction between birth year and sex, and the first ten genetic PCs. Variance explained by the polygenic scores was calculated

in regression analyses as the R^2 change, that is, the R^2 of the model including polygenic scores and covariates minus the R^2 of the model including only covariates. The 95% confidence intervals around all R^2 values are bootstrapped with 1,000 repetitions each.

We also tested the association between polygenic scores for exclusively same-sex sexual behaviour in the UK Biobank sample and number of opposite-sex partners in OSB individuals. We used a polygenic risk score approach rather than genetic correlation based on LD score regression to improve power, given the low number of exclusively homosexual individuals. To avoid overfitting, we used a tenfold cross-validation strategy. We divided the dataset into ten random sub-samples and used nine of those to perform a GWAS for exclusively same-sex sexual behaviour. We repeated the process ten times, each time excluding a different sub-sample, which resulted in ten sets of summary statistics. We then computed ten polygenic scores, applying each set of summary statistics to the corresponding sub-sample that was not included in the GWAS. We considered only SNPs with MAF > 1% and imputation INFO score > 0.8, and we excluded SNPs in the major histocompatibility complex region. We sacrificed one of the ten sets to test which P value threshold resulted in the highest variance explained; the highest variance explained was achieved using $P \leq 0.7$ in males and $P \leq 0.2$ in females. We calculated the polygenic scores in each of the remaining nine datasets. We then tested the association of the scores with number of opposite-sex partners in OSB individuals, adjusting for the effects of age, year of birth and ten genetic PCs. The associations were then meta-analysed across the nine datasets. Note that we did not use LDpred to calculate the scores because that would have been computationally infeasible.

Genomic SEM analysis to test for mediating effects of personality traits. We found a substantial genetic correlation between SSB and number of opposite-sex sexual partners in OSB individuals. There are various potential explanations for this observed genetic correlation. There is a possibility that personality characteristics associated with both traits can explain the genetic correlation; for example, with LD score regression, we found that both openness to experience and risk-taking behaviour are positively genetically correlated with both SSB and number of sexual partners in OSB individuals (Table 3A). While we are not well powered to model a causal influence of these characteristics on our traits of interest, we can use Genomic SEM[30] to test the extent to which accounting for these personality traits attenuates the genetic correlation between SSB and number of sexual partners in OSB individuals.

Genomic SEM can be used to fit a structural equation model based on the genetic covariance among a set of traits using GWAS summary statistics. Here, we used the summary statistics for SSB[19] and number of opposite-sex sexual partners in OSB individuals, as well as summary statistics for openness to experience and risk-taking behaviour[54,55]. We filtered out SNPs with low imputation quality (INFO < 0.6) and MAF < 0.01 (for traits for which these metrics were available). We only included HapMap 3 SNPs to retain a set of high-quality SNPs. We used Genomic SEM's 'munge' function to estimate the genetic covariance matrix and its parameter variance covariance matrix. Using Genomic SEM's 'usermodel' function, we then fitted a structural equation model where the two correlated outcomes, SSB and number of sexual partners in OSB individuals, were both regressed on openness to experience and risk-taking behaviour. We allowed the personality traits to be correlated. We estimated the residual correlations between SSB and number of opposite-sex sexual partners in OSB individuals, after partialling out the correlations with openness to experience, risk-taking and both traits simultaneously (Supplementary Fig. 7 depicts the Genomic SEM model we fitted).

Evolutionary simulations. We developed empirical simulations to demonstrate the feasibility of antagonistic pleiotropy maintaining SSB in the population. The simulation involved a population of simulated individuals whose likelihood of reproducing is affected by their SSB (presence or absence) and their mating success (continuous variable). These traits were influenced by multiple loci, each with two alleles, and by 'environmental' variation (modelled as random error). We tested the frequency (percentage of simulations) with which SSB was maintained in the population when pleiotropy was present and absent. The simulations were written in the R language using various packages for data manipulation and visualization (see Supplementary Methods for details).

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

This research was conducted using data from the UK Biobank resource (application number 25995). UK Biobank data can be accessed on request once a research project has been submitted and approved by the UK Biobank committee (https://www.ukbiobank.ac.uk/researchers/). Data from The National Longitudinal Study of Adolescent to Adult Health (Add Health) can also be applied for (see https://www.cpc.unc.edu/projects/addhealth/documentation for details). GWAS summary statistics of the number of opposite sex sexual partners among OSB individuals in UK-Biobank are available at GWAS Catalog (https://www.ebi.ac.uk/gwas/downloads/summary-statistics, study accession IDs: GCST90026480, GCST90026481, and GCST90026482). Source data are provided with this paper.

Code availability

Custom code used for statistical analyses is available from the corresponding author upon request. Source data are provided with this paper.

Received: 9 December 2019; Accepted: 18 June 2021;
Published online: 23 August 2021

References

1. ACSF investigators. AIDS and sexual behaviour in France. *Nature* **360**, 407 (1992).
2. Melbye, M. & Biggar, R. J. Interactions between persons at risk for AIDS and the general population in Denmark. *Am. J. Epidemiol.* **135**, 593–602 (1992).
3. Semenyna, S. W., VanderLaan, D. P., Petterson, L. J. & Vasey, P. L. Familial patterning and prevalence of male androphilia in Samoa. *J. Sex. Res.* **54**, 1077–1084 (2017).
4. Bailey, J. M. et al. Sexual orientation, controversy, and science. *Psychol. Sci. Public Interest* **17**, 45–101 (2016).
5. Pillard, R. C. & Bailey, J. M. Human sexual orientation has a heritable component. *Hum. Biol.* **70**, 347–365 (1998).
6. Langstrom, N., Rahman, Q., Carlstrom, E. & Lichtenstein, P. Genetic and environmental effects on same-sex sexual behavior: a population study of twins in Sweden. *Arch. Sex. Behav.* **39**, 75–80 (2010).
7. Bailey, N. W. & Zuk, M. Same-sex sexual behavior and evolution. *Trends Ecol. Evol.* **24**, 439–446 (2009).
8. Clive, J., Flintham, E. & Savolainen, V. Understanding same-sex sexual behaviour requires thorough testing rather than reinvention of theory. *Nat. Ecol. Evol.* **4**, 784–785 (2020).
9. Hutchinson, G. E. A speculative consideration of certain possible forms of sexual selection in man. *Am. Nat.* **93**, 81–91 (1959).
10. McKnight, J. *Straight Science?: Homosexuality, Evolution and Adaptation* (Routledge, 1997).
11. Wilson, E. O. *Sociobiology: The New Synthesis* (Harvard Univ. Press, 1975).
12. Camperio-Ciani, A., Corna, F. & Capiluppi, C. Evidence for maternally inherited factors favouring male homosexuality and promoting female fecundity. *Proc. R. Soc. Lond. Ser. B* **271**, 2217–2221 (2004).
13. Zietsch, B. P. et al. Genetic factors predisposing to homosexuality may increase mating success in heterosexuals. *Evol. Hum. Behav.* **29**, 424–433 (2008).
14. Vasey, P. L., Pocock, D. S. & VanderLaan, D. P. Kin selection and male androphilia in Samoan fa'afafine. *Evol. Hum. Behav.* **28**, 159–167 (2007).
15. Rice, W. R., Friberg, U. & Gavrilets, S. Homosexuality as a consequence of epigenetically canalized sexual development. *Q. Rev. Biol.* **87**, 343–368 (2012).
16. Hoskins, J. L., Ritchie, M. G. & Bailey, N. W. A test of genetic models for the evolutionary maintenance of same-sex sexual behaviour. *Proc. R. Soc. B* **282**, 20150429 (2015).
17. Monk, J. D., Giglio, E., Kamath, A., Lambert, M. R. & McDonough, C. E. An alternative hypothesis for the evolution of same-sex sexual behaviour in animals. *Nat. Ecol. Evol.* **3**, 1622–1631 (2019).
18. Schwartz, G., Kim, R. M., Kolundzija, A. B., Rieger, G. & Sanders, A. R. Biodemographic and physical correlates of sexual orientation in men. *Arch. Sex. Behav.* **39**, 93–109 (2010).
19. Ganna, A. et al. Large-scale GWAS reveals insights into the genetic architecture of same-sex sexual behavior. *Science* **365**, eaat7693 (2019).
20. Bell, A. P. & Weinberg, M. *Homosexualities: A Study of Diversity among Men and Women* (Simon and Schuster, 1978).
21. Nila, S., Barthes, J., Crochet, P.-A., Suryobroto, B. & Raymond, M. Kin selection and male homosexual preference in Indonesia. *Arch. Sex. Behav.* **47**, 2455–2465 (2018).
22. Vasey, P. L., Parker, J. L. & VanderLaan, D. P. Comparative reproductive output of androphilic and gynephilic males in Samoa. *Arch. Sex. Behav.* **43**, 363–367 (2014).
23. Miller, E. M. Homosexuality, birth order, and evolution: toward an equilibrium reproductive economics of homosexuality. *Arch. Sex. Behav.* **29**, 1–34 (2000).
24. Carter, A. J. & Nguyen, A. Q. Antagonistic pleiotropy as a widespread mechanism for the maintenance of polymorphic disease alleles. *BMC Med. Genet.* **12**, 160 (2011).
25. Hamer, D. H., Hu, S., Magnuson, V. L., Hu, N. & Pattatucci, A. M. L. A linkage between DNA markers on the X-chromosome and male sexual orientation. *Science* **261**, 321–327 (1993).
26. Sanders, A. R. et al. Genome-side association study of male sexual orientation. *Sci. Rep.* **7**, 16950 (2017).
27. Bulik-Sullivan, B. K. et al. LD score regression distinguishes confounding from polygenicity in genome-wide association studies. *Nat. Genet.* **47**, 291–295 (2015).

28. Vilhjálmsson, B. J. et al. Modeling linkage disequilibrium increases accuracy of polygenic risk scores. *Am. J. Hum. Genet.* **97**, 576–592 (2015).
29. McQueen, M. B. et al. The national longitudinal study of adolescent to adult health (Add Health) sibling pairs genome-wide data. *Behav. Genet.* **45**, 12–23 (2015).
30. Grotzinger, A. D. et al. Genomic structural equation modelling provides insights into the multivariate genetic architecture of complex traits. *Nat. Hum. Behav.* **3**, 513–525 (2019).
31. Schmitt, D. P. Sociosexuality from Argentina to Zimbabwe: a 48-nation study of sex, culture, and strategies of human mating. *Behav. Brain Sci.* **28**, 247–275 (2005).
32. Wainschtein, P. et al. Recovery of trait heritability from whole genome sequence data. Preprint at *bioRxiv* https://www.biorxiv.org/content/10.1101/588020v1 (2019).
33. Rose, M. R. Antagonistic pleiotropy, dominance, and genetic variation. *Heredity* **48**, 63–78 (1982).
34. Gimelfarb, A. Additive variation maintained under stabilizing selection: a two-locus model for pleiotropy for two quantitative characters. *Genetics* **112**, 717–725 (1986).
35. Zhivotovsky, L. A. & Gavrilets, S. Quantitative variability and multilocus polymorphism under epistatic selection. *Theor. Popul. Biol.* **42**, 254–283 (1992).
36. Hedrick, P. W. Antagonistic pleiotropy and genetic polymorphism: a perspective. *Heredity* **82**, 126–133 (1999).
37. Connallon, T. & Clark, A. G. Antagonistic versus nonantagonistic models of balancing selection: characterising the relative timescales and hitchhiking effects of partial selective sweeps. *Evolution* **67**, 908–917 (2013).
38. Simons, Y. B., Bullaughey, K., Hudson, R. R. & Sella, G. A population genetic interpretation of GWAS findings for human quantitative traits. *PLoS Biol.* **16**, e2002985 (2018).
39. Barban, N. et al. Genome-wide analysis identifies 12 loci influencing human reproductive behavior. *Nat. Genet.* **48**, 1462–1472 (2016).
40. Camperio Ciani, A., Cermelli, P. & Zanzotto, G. Sexually antagonistic selection in human male homosexuality. *PLoS ONE* **3**, e2282 (2008).
41. Gavrilets, S. & Rice, W. R. Genetic models of homosexuality: generating testable predictions. *Proc. R. Soc. B* **273**, 3031–3038 (2006).
42. Altman, D. G. & Bland, J. M. Statistics notes: the normal distribution. *Br. Med. J.* **310**, 298 (1995).
43. Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* **562**, 203–209 (2018).
44. Sudlow, C. et al. UK Biobank: an open access resource for identifying the causes of a wide range of complex diseases of middle and old age. *PLoS Med.* **12**, e1001779 (2015).
45. Fry, A. et al. Comparison of sociodemographic and health-related characteristics of UK Biobank participants with those of the general population. *Am. J. Epidemiol.* **186**, 1026–1034 (2017).
46. Harris, K. M., Halpern, C. T., Haberstick, B. C. & Smolen, A. The national longitudinal study of adolescent health (Add Health) sibling pairs data. *Twin Res. Hum. Genet.* **16**, 391–398 (2013).
47. Lee, A. J. et al. Genetic factors that increase male facial masculinity decrease facial attractiveness of female relatives. *Psychol. Sci.* **25**, 476–484 (2014).
48. McCarthy, S. et al. A reference panel of 64,976 haplotypes for genotype imputation. *Nat. Genet.* **48**, 1279–1283 (2016).
49. Das, S. et al. Next-generation genotype imputation service and methods. *Nat. Genet.* **48**, 1284–1287 (2016).
50. Loh, P. R. et al. Efficient Bayesian mixed-model analysis increases association power in large cohorts. *Nat. Genet.* **47**, 284–290 (2015).
51. Watanabe, K., Taskesen, E., van Bochoven, A. & Posthuma, D. Functional mapping and annotation of genetic associations with FUMA. *Nat. Commun.* **8**, 1826 (2017).
52. Bulik-Sullivan, B. et al. An atlas of genetic correlations across human diseases and traits. *Nat. Genet.* **47**, 1236 (2015).
53. Purcell, S. et al. PLINK: A tool set for whole-genome association and population-based linkage analyses. *Am. J. Hum. Genet.* **81**, 559–575 (2007).
54. de Moor, M. H. M. et al. Meta-analysis of genome-wide association studies for personality. *Mol. Psychiatry* **17**, 337–349 (2012).
55. Lo, M. T. et al. Genome-wide analyses for personality traits identify six genomic loci and show correlations with psychiatric disorders. *Nat. Genet.* **49**, 152–156 (2017).

Acknowledgements

This research was conducted using the UK Biobank Resource under applications 25995. We thank all cohort participants for making this study possible. A.R.S. received funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development specifically to investigate the genetics of sexual orientation: R01HD041563 (A.R.S. PI) and R21HD080410 (A.R.S., E.R.M. MPI). E.R.M., G.W.B. and S.G. are also supported by R21HD080410. No other member of the group received funding specifically for this study, but members of our team received salary funding from organizations as well as their own universities. B.P.Z. received funding from The Australian Research Council (FT160100298). A.A. is supported by ZonMw grant 849200011 from The Netherlands Organisation for Health Research and Development. Study-specific acknowledgements: This research uses data from Add Health, a programme project directed by K.M. Harris (PI) and designed by J.R. Udry, P.S. Bearman and K.M. Harris at the University of North Carolina at Chapel Hill, and funded by grant P01-HD031921 from the Eunice Kennedy Shriver National Institute of Child Health and Human Development, with cooperative funding from 23 other federal agencies and foundations. Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). This research uses Add Health GWAS data funded by Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD) grants R01 HD073342 to K. M. Harris (PI) and R01 HD060726 to K.M. Harris, J.D. Boardman and M.B. McQueen (MPIs). The authors thank A. Ganna and M. Nivard for analysis work, A. Ganna, J. Perry, M. Nivard, B. Neale, R. Wedow and M. Keller for extensive discussion, and B. von Hippel, F. Barlow, F. Sathirapongsasuti, A. Auton and J. McCreight for comments on manuscript drafts. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Author contributions

B.P.Z. and K.J.H.V. conceived and designed the study. M.S., A.A. and R.M. analysed the data and produced the figures. B.P.Z., M.S. and K.J.H.V. wrote the manuscript. A.A., R.M., J.N.L., S.G., G.W.B., E.R.M. and A.R.S. provided significant feedback on the analyses and the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41562-021-01168-8.

Correspondence and requests for materials should be addressed to B.P.Z.

Peer review information Primary Handling Editors: Charlotte Payne and Stavroula Kousta.

nature research

Corresponding author(s): Brendan Zietsch

Last updated by author(s): Jun 2, 2020

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see Authors & Referees and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
☐	☒ The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐ A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒ The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☐	☒ A description of all covariates tested
☐	☒ A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒ A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒ For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☒	☐ For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐ For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☐	☒ Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used
Data analysis	R e.g. version 3.6

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Information on how to obtain the Add Health data files is available on the Add Health website (http://www.cpc.unc.edu/addhealth). Data and code for same-sex sexual behaviour GWAS can be obtained per the following: The code is available through GitHub (https://github.com/andgan/sexual_orientation_GWAS), archived at Zenodo (48), and the GWAS summary statistics of the UK Biobank sample (and the top 10,000 independent SNPs from the meta-analysis including 23andMe data) are available at GWAS Catalog (www.ebi.ac.uk/gwas/downloads/summary-statistics). Access to the full summary statistics of the 23andMe sample (for all SNPs) can be obtained by qualified researchers through a data transfer agreement with 23andMe that protects the privacy of the 23andMe participants. Researchers interested in the full meta-analysis summary statistics containing 23andMe data must also apply to 23andMe. Please visit (https://research.23andme.com/dataset-access) for more information and to apply to access the data. Access to individual level data from the UK Biobank can be obtained by bona fide scientists through application with UK Biobank (www.ukbiobank.ac.uk/researchers).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	Data used here was collected by UK Biobank, and the sample size was determined without regard to our particular study or its hypotheses.
Data exclusions	For the new GWAS on number of opposite-sex partners among those who had never had a same-sex partner, individuals who reported having had sexual intercourse with someone of the same-sex (3.3%) or that had never had a sexual relationship (0.9%) were set at missing. We also excluded participants with more than 100 sexual partners, since these outliers could increase phenotypic heterogeneity and have a disproportionate influence on our findings; this reduced our sample by 1035 males and 79 females (0.5% and 0.04%, respectively).
Replication	We replicated the main finding in an independent sample, AddHealth.
Randomization	NA
Blinding	Data used here was collected by UK Biobank without regard to our particular study or its hypotheses.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒	☐ Antibodies
☒	☐ Eukaryotic cell lines
☒	☐ Palaeontology
☒	☐ Animals and other organisms
☐	☒ Human research participants
☒	☐ Clinical data

Methods

n/a	Involved in the study
☒	☐ ChIP-seq
☒	☐ Flow cytometry
☒	☐ MRI-based neuroimaging

Human research participants

Policy information about studies involving human research participants

Population characteristics	We used data from 191,737 males and 223,014 females aged from 40 to 70 who lived within 25 miles of one of the 22 assessment centers in England, Wales, and Scotland (participation rate for the baseline assessment 5.5%) and who answered the questions relating to sexual behavior and whose genotype data passed quality control.
Recruitment	Recruitment procedures for the UK Biobank study - not done in this study - is described in Bycroft, C. et al. The UK Biobank resource with deep phenotyping and genomic data. Nature 562, 203-209, doi:10.1038/s41586-018-0579-z (2018).
Ethics oversight	The UK Biobank ethics committee The University of Queensland Human Research Ethics Committee

Note that full information on the approval of the study protocol must also be provided in the manuscript.