Long title: Human demographic history has amplified the effects of background selection across the genome

Short title: Background selection and demography in humans

Raul Torres[1], Zachary A. Szpiech[2], and Ryan D. Hernandez[2,3,4,5*]

[1] Biomedical Sciences Graduate Program, University of California San Francisco, San Francisco, CA, United States of America

[2] Department of Bioengineering and Therapeutic Sciences, University of California San Francisco, San Francisco, CA, United States of America

[3] Institute for Human Genetics, University of California San Francisco, San Francisco, CA, United States of America

[4] Institute for Computational Health Sciences, University of California San Francisco, San Francisco, CA, United States of America

[5] Quantitative Biosciences Institute, University of California San Francisco, San Francisco, CA, United States of America

* Corresponding author

E-mail: ryan.hernandez@ucsf.edu (RDH)

1 Abstract

2 Natural populations often grow, shrink, and migrate over time. Demographic pro-

3 cesses such as these can impact genome-wide levels of genetic diversity. In addition,

4 genetic variation in functional regions of the genome can be altered by natural selection,

5 which drives adaptive mutations to higher frequencies or purges deleterious ones. Such

6 selective processes impact not only the sites directly under selection but also nearby

7 neutral variation through genetic linkage through processes referred to as genetic hitch-

8 hiking in the context of positive selection and background selection (BGS) in the context

9 of purifying selection. While there is extensive literature examining the impact of selec-

10 tion at linked sites at demographic equilibrium, less is known about how non-equilibrium

11 demographic processes impact the effects of hitchhiking and BGS. Utilizing a global

12 sample of human whole-genome sequences from the Thousand Genomes Project and

13 extensive simulations, we investigate how non-equilibrium demographic processes

14 magnify and dampen the consequences of selection at linked sites across the human

15 genome. When binning the genome by inferred strength of BGS, we observe that, com-

16 pared to Africans, non-African populations have experienced larger proportional de-

17 creases in neutral genetic diversity in such regions. We replicate these findings in ad-

18 mixed populations by showing that non-African ancestral components of the genome

19 have also been impacted more severely in these regions. We attribute these differences

20 to the strong, sustained/recurrent population bottlenecks that non-Africans experienced

21 as they migrated out of Africa and throughout the globe. Furthermore, we observe a

22 strong correlation between F_{ST} and inferred strength of BGS, suggesting a stronger rate

23 of genetic drift. Forward simulations of human demographic history with a model of BGS

24   support these observations. Our results show that non-equilibrium demography signifi-

25   cantly alters the consequences selection at linked sites and support the need for more

26   work investigating the dynamic process of multiple evolutionary forces operating in con-

27   cert.

28

Author summary

30   Patterns of genetic diversity within a species are affected at broad and fine

31   scales by population size changes ("demography") and natural selection. From both

32   population genetics theory and observation of genomic sequence data, it is known that

33   demography can alter genome-wide average neutral genetic diversity. Additionally, nat-

34   ural selection can affect neutral genetic diversity regionally across the genome via se-

35   lection at linked sites. During this process, natural selection acting on adaptive or dele-

36   terious variants in the genome will also impact diversity at nearby neutral sites due to

37   genetic linkage. However, less is well known about the dynamic changes to diversity

38   that occur in regions impacted by selection at linked sites when a population undergoes

39   a size change. We characterize these dynamic changes using thousands of human ge-

40   nomes and find that the population size changes experienced by humans have shaped

41   the consequences of linked selection across the genome. In particular, population con-

42   tractions, such as those experienced by non-Africans, have disproportionately de-

43   creased neutral diversity in regions of the genome inferred to be under strong back-

44   ground selection (i.e., selection at linked sites that is caused by natural selection acting

45   on deleterious variants), resulting in large differences between African and non-African

46   populations.

3

47

Introduction

49   Genetic diversity in a species is determined through the complex interplay of mu-

50   tation, demography, genetic drift, and natural selection. These evolutionary forces oper-

51   ate in concert to shape patterns of diversity at both the local scale and genome-wide

52   scale. For example, in recombining species, levels of genetic diversity are distributed

53   heterogeneously across the genome as peaks and valleys that are often correlated with

54   recombination rate and generated by past or ongoing events of natural selection [1].

55   But at the genome-wide scale, average levels of genetic diversity are primarily impacted

56   by population size changes, yielding patterns of diversity that are a function of a popula-

57   tion's demographic history [2]. These patterns of diversity may also yield information for

58   inferring past events of natural selection and population history, giving valuable insight

59   into how populations have evolved over time [3–8]. With recent advances in sequencing

60   technology yielding whole-genome data from thousands of individuals from species with

61   complex evolutionary histories [9,10], formal inquiry into the interplay of demography

62   and natural selection and testing whether demographic effects act uniformly across the

63   genome as a function of natural selection is now possible.

64   In the past decade, population genetic studies have shed light on the pervasive-

65   ness of dynamic population histories in shaping overall levels of genetic diversity across

66   different biological species. For example, multiple populations have experienced major

67   population bottlenecks and founder events that have resulted in decreased levels of ge-

68   nome-wide diversity. Evidence for population bottlenecks exists in domesticated species

69   such as cattle [11], dogs [12], and rice [13], and in natural populations such as *Drosoph-*

4

70   *ila melanogaster* [14–16], rhesus macaque [17], and humans [18,19]. Notably, popula-

71   tion bottlenecks leave long lasting signatures of decreased diversity, which may be de-

72   pressed even after a population has recovered to or surpassed its ancestral size

73   [20,21]. Such examples are evident in humans, where non-African populations exhibit a

74   lower amount of genetic diversity compared to Africans [9], despite the fact that they

75   have been inferred to have undergone a greater population expansion in recent times

76   [22,23].

77   Locally (i.e., regionally) across the genome, the action of natural selection can

78   also lead to distinct signatures of decreased genetic diversity (although some forms of

79   selection, such as balancing selection, can increase genetic diversity [24]). For exam-

80   ple, mutations with functional effects may be removed from the population due to purify-

81   ing selection or fix due to positive selection, thereby resulting in the elimination of genet-

82   ic diversity at the site. But while sites under direct natural selection in the genome rep-

83   resent only a small fraction of all sites genome-wide, the action of natural selection on

84   these selected sites can have far-reaching effects across neutral sites in the genome

85   due to linkage. Under positive selection, genetic hitchhiking [25] causes variants lying

86   on the same haplotype as the selected allele to rise to high frequency during the selec-

87   tion process (note that we will use the term "genetic hitchhiking" here only in the positive

88   selection context of selection at linked sites). Conversely, under purifying selection,

89   background selection (BGS) [26] causes linked neutral variants to decrease in frequen-

90   cy or be removed from the population. Both of these processes of selection at linked

91   sites result in decreased neutral genetic diversity around the selected site. Recombina-

92   tion can decouple neutral sites from selected sites in both cases and neutral diversity

5

93   tends to increase toward its neutral expectation as genetic distance from selected sites

94   increases [27].

95   Evidence for genetic hitchhiking and BGS has been obtained from the genomes

96   of several species, including *Drosophila melanogaster* [28–33], wild and domesticated

97   rice [34,35], nematodes [36,37], humans [3,6,38–42], and others (see [1] for a review).

98   While the relative contributions of genetic hitchhiking and BGS to shaping patterns of

99   human genomic diversity have been actively debated [40,43–45], the data strongly sup-

100   port a large role for BGS in shaping genome-wide patterns of neutral genetic variation

101   [41,42]. Indeed, recent arguments have been made in favor of BGS being treated as the

102   null model when investigating the impact of selection at linked sites across recombining

103   genomes [1,32,45–48], with one study in humans showing that BGS has decreased ge-

104   netic diversity by 19-26% if other modes of selection at linked sites are assumed to be

105   minor [6].

106   Although the effects of selection at linked sites across the genome have been

107   described in a multitude of studies, it is still less obvious whether populations that have

108   experienced different demographic histories, such as African and non-African human

109   populations, should exhibit similar relative effects in those regions. In the context of

110   BGS, early work resulted in the expression $\pi \approx 4f_0 N_e \mu$ [26], which suggests that the ex-

111   pected level of diversity with BGS is simply proportional to the neutral expectation (with

112   proportionality constant f_0 being a function of the rates of deleterious mutation and re-

113   combination). Importantly, while populations with different N_e are expected to have dif-

114   ferent levels of π in the context of BGS, the proportionality constant f_0 is assumed to be

115   independent of population history (so long as the recombination landscape and delete-

6

116   rious mutation rates remain constant). Much of the theory developed in the context of

117   BGS has been developed under the assumption that the population is at demographic

118   equilibrium, and recent work has demonstrated that this assumption likely holds under

119   changing demography if selection is strong enough (or populations are large enough)

120   such that mutation-selection balance is maintained [49,50]. However, strong, sustained

121   population bottlenecks may lead to violations of that assumption. Thus, in humans and

122   other natural populations experiencing non-equilibrium demography, genetic drift may

123   perturb genetic diversity at neutral sites under BGS that are unaccounted for in these

124   models.

125   While little attention has been given to the potential consequences of demogra-

126   phy on patterns of neutral variation in regions experiencing selection at linked sites (but

127   see [51,52] for how selection at linked sites may impact the inference of demography

128   itself), recent studies have suggested that alleles directly under natural selection experi-

129   ence non-linear dynamics in the context of non-equilibrium demography. For the case of

130   purifying selection, the equilibrium frequency of an allele is dependent on its fitness ef-

131   fect, with deleterious alleles having lower equilibrium frequencies than neutral alleles.

132   After a population size change, deleterious alleles tend to change frequency faster than

133   neutral alleles, allowing them to reach their new equilibrium frequency at a faster rate

134   [53,54]. This can result in relative differences in deleterious allele frequencies among

135   populations with different demographic histories. Such effects are especially apparent in

136   populations suffering bottlenecks [55] and have been tested and observed between dif-

137   ferent human populations with founder populations exhibiting a greater proportion of

138   non-synonymous variation relative to synonymous variation [56–58]. We hypothesized

7

139   that these non-equilibrium dynamics could also perturb nearby neutral variants due to

140   the effects of selection at linked sites. In support of our hypothesis, a recent simulation

141   study modeling *Drosophila* observed that population bottlenecks can result in different

142   rates of recovery of neutral genetic diversity depending on the strength of BGS [48].

143   Another recent study [59] analyzed neutral diversity surrounding putatively deleterious

144   loci in domesticated versus wild maize. They found that the extreme domestication bot-

145   tleneck reduced the efficiency of purifying selection, which has resulted in higher rela-

146   tive diversity in the domesticated population compared to the wild population (which has

147   likely experienced a much more stable demographic history). Together, these studies

148   provide further evidence that non-equilibrium demography should have a strong effect

149   on patterns of diversity in the presence of selection at linked sites.

150   To investigate the impact of non-equilibrium dynamics in regions experiencing

151   selection at linked sites, we measure patterns of average pairwise neutral genetic diver-

152   sity (π), or neutral heterozygosity if the population was admixed, as a function of the

153   strength of BGS, B (inferred by Ref. [6]), within a global set of human populations from

154   phase 3 of the Thousand Genomes Project (TGP) [9]. We focus on the ratio of neutral

155   diversity in regions of strong BGS (low B) to regions of weak BGS (high B, the closest

156   proxy available for neutral variation in humans), which we term "relative diversity." While

157   there are many caveats that may plague the direct interpretation of specific B values

158   (e.g., positive selection is not modeled, the distribution of fitness effects are inconsistent

159   with other studies, and the deleterious mutation rate exceeds the per base pair mutation

160   rate of other studies), we argue below that B scores nevertheless provide a decent

8

161   proxy for ranking sites from most closely linked to deleterious loci (low B) to most un-

162   linked from deleterious loci (high B) in humans.

163   We find substantial differences in relative diversity between populations, which

164   we attribute to their non-equilibrium demographics. We confirm that the interplay of de-

165   mography and selection at linked sites can explain the differences of relative diversity

166   across human populations with simulations incorporating a parametric demographic

167   model of human history [7] with and without a model of BGS. We also investigate how

168   genetic differentiation between TGP populations (as measured by F_{ST}) is shaped by se-

169   lection at linked sites by measuring F_{ST} as a function of B. Finally, we demonstrate that

170   back migration from Europeans and Asians into Africa re-introduces sufficient deleteri-

171   ous variation to impact patterns of BGS, leading to decreased relative diversity in Afri-

172   cans. Our results demonstrate the strong impact that changing demography has on per-

173   turbing levels of diversity in regions experiencing selection at linked sites and have im-

174   plications for population genetic studies seeking to characterize linked selection across

175   any species or population that is not at demographic equilibrium.

176

Results

Differential impact of selection at linked sites across human populations

179   We measured mean pairwise genetic diversity (π) in the autosomes (we ignore

180   the sex chromosomes and the mitochondrial genome for all analyses) among the 20

181   non-admixed populations from the phase 3 TGP data set, consisting of 5 populations

182   each from 4 continental groups: Africa (AFR), Europe (EUR), South Asia (SASN), and

183   East Asia (EASN; population labels and groupings reported in S12 Table in Supporting

9

184   information). A set of stringent filters, including the masking of sites inferred to be under

185   selective sweeps, were first applied to all 20 populations to identify a high-quality set of

186   putatively neutral sites in the genome (see Materials and Methods). Sites were then di-

187   vided into quantile bins based on B [6]. For our initial set of analyses, we focused on the

188   bins corresponding to the 1% of sites inferred to be under the strongest amount of BGS

189   (i.e., sites having the lowest inferred B values) and the 1% of sites inferred to be under

190   the weakest amount BGS (i.e., sites having the highest inferred B values). Mean diversi-

191   ty was normalized by divergence with rhesus macaque within these bins for each popu-

192   lation and is shown in Figs 1A-B. As expected, normalized diversity was highest in Afri-

193   can populations and lowest in East Asian populations across both 1% B quantile bins.

194   To estimate the impact that selection at linked sites has had on neutral diversity,

195   we calculated a statistic called "relative diversity" (analogous to π/π_0 in the BGS litera-

196   ture; [26,60]) for each population. We define relative diversity as the ratio of normalized

197   diversity in the lowest 1% B bin to normalized diversity in the highest 1% B bin, which

198   should capture the relative impact of selection at linked sites within the genome. Fig 1C

199   shows that relative diversity was lower in non-African populations (0.348-0.365 for non-

200   Africans, 0.396-0.408 for Africans), suggesting that these populations have experienced

201   a greater reduction in diversity due to selection at linked sites than can be explained by

202   demography alone.

203   To characterize these effects across a broader distribution of sites experiencing

204   selection at linked sites, we grouped populations together according to their continental

205   group (i.e., African, European, South Asian, and East Asian, see S12 Table in Support-

206   ing information for a detailed description) and estimated relative diversity at neutral sites

207   for each of the continental groups in bins corresponding to the lowest 1%, 5%, 10%,

208   and 25% quantiles of *B* (note these partitions were not disjoint). As expected, relative

209   diversity increased for all continental groups as the bins became more inclusive (Fig

210   2B), reflecting a reduced impact on the reduction of diversity due to selection at linked

211   sites. We also observed that non-African continental groups consistently had a lower

212   relative diversity compared to African groups, demonstrating that the patterns we ob-

213   served in the most extreme regions experiencing selection at linked sites also held for

214   broader regions. Interestingly, we observed a consistent trend of rank order for relative

215   diversity between the different continental groups for each quantile bin, with the East

216   Asian group experiencing the greatest reduction of relative diversity, followed by the

217   South Asian, European, and African groups. This result suggested a stronger effect for

218   demography on the diversity-reducing effect selection at linked sites for those popula-

219   tions experiencing the strongest bottlenecks. However, the observed differences in rela-

220   tive diversity between non-African and African continental groups became less pro-

221   nounced as the bins became more inclusive (Fig 2B). These effects remained even af-

222   ter we controlled for the effects of GC-biased gene conversion and recombination

223   hotspots (S2 and S4 Figs) or if we did not normalize diversity by divergence (S3 and S5

224   Figs). Patterns of relative diversity in regions of local ancestry (i.e., African, European,

225   or Native American) across admixed TGP populations also largely recapitulated the pat-

226   terns observed in their continental group counterparts across *B* quantile bins, with the

227   largest reductions in relative diversity occurring for the Native American and European

228   ancestral segments (S11 Fig, S1 Supporting information).

11

229   To test if demography has impacted selection at linked sites more recently in

230   time, we also calculated the number of singletons observed per site (normalizing by di-

231   vergence and using the same set of neutral filters as was used for the calculations of π)

232   across the lowest and highest 1% B quantile bins. Singletons are, on average, the

233   youngest variants within the genome and should better capture signals about more re-

234   cent population history. We took the ratio of singletons observed per-site across these

235   extreme B quantile bins to create a statistic called relative singleton density, which we

236   term "ψ/ψ_0." We accounted for differences in population sample size by first projecting

237   down all populations to 2N=170 (Materials and methods). Qualitatively, our measure-

238   ments of ψ/ψ_0 showed patterns in the opposite direction of our calculations of π/π_0, with

239   Africans exhibiting a lower ratio of ψ/ψ_0 when compared to non-Africans (0.665-0.695

240   for Africans, 0.733-0.804 for non-Africans; Fig 3). These patterns suggest that the im-

241   pact of demography on regions experiencing selection at linked sites is transient, with

242   patterns of relative diversity between populations dependent on the time frame in which

243   they are captured (see Discussion).

244

245   **Selection at linked sites has shaped patterns of population differentiation**

246   We have shown that selection at linked sites increases the rate of drift at neutral

247   loci, and that this process is amplified by demographic changes like population bottle-

248   necks. One might then expect that selection could also amplify population differentiation

249   at linked neutral loci. This outcome is obvious in the context of hitchhiking (where linked

250   neutral loci sweep to high frequency) but is also expected with BGS [61,62], since de-

251   creases in diversity can accelerate differentiation. Here we quantify the magnitude of

12

252   the effect of BGS on population differentiation in humans, and find that population dif-

253   ferentiation at neutral loci is indeed highly correlated with B (the inferred strength of

254   BGS; Fig 4 and Table 1). Specifically, we divided the genome into 2% quantile bins

255   based on the genome-wide distribution of B and measured F_{ST} in each bin for all pairs

256   of populations from different continental groups [63]. We then performed simple linear

257   regression using B as an explanatory variable and F_{ST} as our dependent variable with

258   the linear model $F_{ST} = \beta_0 + \beta_1 B + \varepsilon$. We found that across all 150 population compari-

259   sons (i.e., the "Global" estimate in Table 1), B explained 26.9% of the change in F_{ST}

260   across the most extreme B values, was robust to outliers [64] (S6 Table in Supporting

261   information), and dominated the effects of local recombination rate (see Supporting in-

262   formation).

263   **Table 1. Regression coefficient estimates for linear regression of F_{ST} on 2% quan-**
264   **tile bins of B**.

	AFR vs. EASN	AFR vs. EUR	AFR vs. SASN	EUR vs. SASN	EUR vs. EASN	SASN vs. EASN	Global
β_0 ± SEM (*p*)	0.2044 ± 0.0039 (< 1e-04)	0.1716 ± 0.0031 (< 1e-04)	0.1596 ± 0.0029 (< 1e-04)	0.0455 ± 0.0011 (< 1e-04)	0.1216 ± 0.0029 (< 1e-04)	0.0903 ± 0.0023 (< 1e-04)	0.1322 ± 0.0019 (< 1e-04)
β_1 ± SEM (*p*)	-0.0434 ± 0.0046 (< 1e-04)	-0.0358 ± 0.0037 (< 1e-04)	-0.0355 ± 0.0034 (< 1e-04)	-0.0098 ± 0.0013 (< 1e-04)	-0.0173 ± 0.0035 (< 1e-04)	-0.0261 ± 0.0027 (< 1e-04)	-0.0280 ± 0.0022 (< 1e-04)
r ± SEM	-0.8363 ± 0.0295	-0.7441 ± 0.0362	-0.7794 ± 0.0332	-0.3847 ± 0.0414	-0.6220 ± 0.0785	-0.5968 ± 0.0348	-0.1292 ± 0.0098

265
266   The first two rows give the regression coefficients for the linear model $F_{ST} = \beta_0 + \beta_1 B +$
267   ε, where B represents the mean background selection coefficient for the bin being test-
268   ed and F_{ST} is the estimated F_{ST} for all population comparisons within a particular pair of
269   continental groups (given in the column header). The final column, "Global", gives the
270   regression coefficients for the linear model applied to all pairwise population compari-
271   sons (150 total). The correlation coefficient, *r*, between B and F_{ST} for each comparison
272   is shown in the bottom row. Standard errors of the mean (SEM) for β_0, β_1, and *r* were
273   calculated from 1,000 bootstrap iterations (see Materials and Methods). P-values are
274   derived from a two-sided t-test of the t-value for the corresponding regression coeffi-
275   cient.
276

277

Demographic inference in putatively neutral regions of the genome

279   One consequence of BGS and hitchhiking in driving patterns of neutral variation

280   within and between human populations is that demographic inference could be substan-

281   tially biased [51,52,65]. To assess the degree of bias in the context of human data, we

282   fit a 13-parameter demographic model of African, European, and East Asian demogra-

283   phy using only putatively neutral regions of the genome under the weakest effects of

284   selection at linked sites ($B \geq 0.994$) from a subset of TGP individuals with high coverage

285   whole genome sequence data (see Materials and Methods). Our demographic model

286   followed that of Gutenkunst et al. [7], with an ancient human expansion in Africa and a

287   single out-of-Africa bottleneck followed by European- and East Asian-specific bottle-

288   necks, as well as exponential growth in both non-African populations and migration be-

289   tween all populations. To make comparisons to previous studies that have used se-

290   quence data from coding regions or genes [7,22,23], which may be under strong BGS

291   or hitchhiking effects, we also inferred demographic parameters using coding four-fold

292   degenerate synonymous sites. Our inferred parameters for human demography were

293   strikingly different between the two sets of sequence data (S1 Fig, S1 Table in Support-

294   ing information). Notably, inferred effective population size parameters were larger for

295   contemporary population sizes when using four-fold degenerate synonymous sites ver-

296   sus ascertained neutral regions with $B \geq 0.994$, with N_e inferred to be 22%, 23%, and

297   29% larger for AFR, EUR, and EASN populations, respectively. This is despite the fact

298   that the ancestral N_e was inferred to be lower for four-fold degenerate synonymous sites

299   (N_e = 18,449 and 17,118, for neutral regions with $B \geq 0.994$ and four-fold degenerate

14

300   sites, respectively). This result may stem from the expected decrease in N_e going into

301   the past in regions of strong BGS, which can lead to inflated estimates of recent popula-

302   tion growth [52] and has also been shown in simulation studies of synonymous sites

303   under BGS [65]. Put more simply, the skew of the site-frequency spectrum towards rare

304   variants in regions experiencing selection at linked sites mimics a population expansion,

305   thus leading to erroneous inference.

306

307   **Simulations confirm that demographic effects can impact background selection**

308   Using the demographic parameters inferred from neutral regions where $B \geq$

309   0.994, we simulated patterns of neutral diversity with and without the effects BGS (see

310   Materials and Methods). To measure the relative impact of BGS for each population, we

311   then took the ratio of neutral diversity from BGS simulations (π) and neutral diversity

312   from simulations without BGS (π_0) to calculate relative diversity (π/π_0). As expected, we

313   found that BGS reduced relative diversity ($\pi/\pi_0 < 1$) for all three populations in our simu-

314   lations. However, non-African populations experienced a proportionally larger decrease

315   in π/π_0 compared to the African population (π/π_0 = 0.43, 0.42, 0.41 in AFR, EUR, and

316   EASN respectively). These results are comparable to (but not quite as extreme as) the

317   effects we observed in the regions of the genome with the strongest effects of BGS for

318   these population groups (Fig 1C). To understand how this dynamic process occurs, we

319   sampled all simulated populations every 100 generations through time to observe the

320   effect of population size change on π, π_0, and the ratio π/π_0 (Fig 5). We observed that

321   there is a distinct drop in π and π_0 at each population bottleneck experienced by non-

322   Africans, with East Asians (who had a more severe bottleneck) experiencing a larger

15

323 drop than Europeans. Fig 5C shows that the population bottlenecks experienced by

324 non-African populations also reduces π/π_0. Surprisingly, Africans also experienced a

325 large drop in π/π_0 (but less than non-Africans) even though they did not experience any

326 bottlenecks. This was attributable to migration between non-Africans and Africans and

327 this pattern disappeared when we ran simulations using an identical demographic mod-

328 el with BGS but without migration between populations (S7 Fig). This finding highlights

329 an evolutionary role that deleterious alleles can play when they are transferred across

330 populations through migration (see Discussion).

331 We also calculated ψ/ψ_0 across these simulations and again qualitatively ob-

332 served patterns similar to, but not as extreme, as our empirical calculations of ψ/ψ_0. As

333 with pairwise neutral diversity, BGS predictably decreased the density of singletons

334 across all populations (S12 Fig A). However, Africans exhibited a lower ψ/ψ_0 when

335 compared to non-Africans (ψ/ψ_0 = 0.79, 0.89, 0.91 in AFR, EUR, and EASN respective-

336 ly). Calculating ψ and ψ_0 through time showed that the population bottlenecks experi-

337 enced by non-Africans led to strong decreases in both ψ and ψ_0, with recent expansion

338 in these populations then leading to large, rapid recoveries. Strong decreases in ψ/ψ_0

339 after each population bottleneck were also observed, including a slight decrease in ψ/ψ_0

340 in Africans that disappeared in the simulations without migration (S12 Fig B). While ψ/ψ_0

341 for the European/East Asian ancestral population in the simulations with migration re-

342 mained below that of Africans during the course of the Out-of-Africa bottleneck, we ob-

343 served a rapid recovery in ψ/ψ_0 for this population in the simulations without migration

344 (compare bottoms panels, S12 Fig A and B). This suggests that for populations experi-

345 encing a sustained population bottleneck, the response of singletons to changes in the

16

346 weakened intensity of BGS is quite rapid, especially when compared to patterns of π/π_0

347 (compare Fig S7 C to Fig S12 B bottom panel). However, population migration mitigates

348 this pattern. However, regardless of whether migration between populations was simu-

349 lated, BGS seemed to have little effect on singleton density recovery in Europeans and

350 Asians once population expansion occurred.

351   Our simulations were based on the functional density found in a 2 Mb region of

352 the human genome with the lowest B values and, thus, where BGS was inferred to be

353 strongest (chr3: 48,600,000-50,600,000). There, 20.46% of sites were either coding or

354 conserved non-coding (see Materials and Methods). Thus, the fraction of the genome

355 experiencing deleterious mutation in our simulations of strong BGS was 0.2046. The

356 patterns we observed in these simulations likely represent an upper bound on the

357 strength of BGS in the human genome. Since the strength of BGS is dependent upon

358 the density of sites experiencing deleterious mutation within a given region (or more

359 formally, U, which is the product of the per-site deleterious mutation rate and the num-

360 ber of sites experiencing deleterious mutation [66]), we simulated weaker effects of

361 BGS by reducing the fraction of sites experiencing purifying selection (keeping the dis-

362 tribution of selective effects constant, see Materials and Methods). When the fraction of

363 sites experiencing selection was decreased 2-4 fold in our simulations, we continued to

364 observe a stepwise decrease in π/π_0 while maintaining the specific rank order of African,

365 followed by European, and then East Asian populations (S8 Fig). As expected, π/π_0 in-

366 creased for all populations as the fraction of sites that were simulated as deleterious

367 decreased (π/π_0 = 0.641 vs. 0.802, 0.62 vs. 0.777, and 0.611 vs. 0.777 for AFR, EUR,

368 and EASN when the fraction of sites experiencing selection was reduced to 0.1023 and

17

369   0.05115, respectively). These simulations resulted in π/π_0 values much larger than we

370   observed in data (Figs 1C, 2B).

371

Discussion

373   In our analyses of thousands of genomes from globally distributed human popu-

374   lations, we have confirmed that the processes of demography and selection at linked

375   sites govern neutral variation across the genome. While this observation is not unex-

376   pected, we have characterized the dynamic consequence of non-equilibrium demo-

377   graphic processes in regions experiencing selection at linked sites in humans. We find

378   that demography (particularly population bottlenecks) can amplify the consequences of

379   selection at linked sites. To remove any possible biases that would influence our results,

380   we controlled for functional effects of mutations, variability in mutation along the ge-

381   nome, potential sequencing artifacts, GC-based gene conversion, the potential muta-

382   genic effects of recombination hotspots, and also removed regions of the genome with

383   signatures of positive selection. None of these factors qualitatively affected our results.

384   We do recognize that one caveat of our controls is the fact that divergence itself

385   is not independent of BGS [67], and this may present biases when using divergence to

386   control for variation in mutation rate along the genome. This is because the rate of coa-

387   lescence in the ancestral population of two groups will be faster in regions of strong

388   BGS compared to regions of weak BGS due to the lower N_e of the former, leading to a

389   decrease in overall divergence in those regions. To limit the contribution of such biases

390   in ancestral N_e to divergence, we use rhesus macaque since it is more distantly related

391   to humans than other primate species such as orangutan or chimpanzee (human-

18

392   rhesus divergence: 29.6 MYA; [68]). Furthermore, if B truly captures the effects of BGS,

393   then normalizing by the lower divergence that is characteristic of strong BGS bins and

394   the higher divergence that is characteristic of weak BGS bins should make any differ-

395   ences between the two smaller, not greater. In fact, for our calculations of relative diver-

396   sity in which we skip the normalization step, the differences in diversity between the

397   lowest 1% and highest 1% B value bins are greater and give a lower ratio of relative di-

398   versity (π/π_0 for AFR is 0.373 without the divergence step and 0.402 with the divergence

399   step). A similar pattern is also observed for other continental groups (compare Fig 2 and

400   S5 Fig). More importantly though, we should not expect the potential biases of our di-

401   vergence step to contribute to the differences in relative diversity between each of the

402   continental groups since biases in divergence estimates across the genome are based

403   on the human reference sequence and are therefore identical across all human popula-

404   tions.

405   We also note that the estimates of B by McVicker et al. [6] may be biased by

406   model assumptions concerning mutation rates and the specific sites subject to purifying

407   selection, with the exact values of B unlikely to be precisely inferred. In fact, the B val-

408   ues provided by McVicker et al. range from 0 to 1, suggesting that some regions of the

409   genome should be essentially devoid of diversity (but we do not observe this to be the

410   case). Since our own analyses show that relative diversity has a lower bound at only

411   ~0.35 in humans, the exact value of B itself should not be taken at face value. Rather,

412   our primary motivation for using B was to ascertain regions that should be on the ex-

413   treme ends of the genome-wide distribution of regions experiencing selection at linked

414   sites, for which B should provide a good assessment. A study by Comeron et al. [32]

19

415   that investigated BGS in *Drosophila* and utilized the same model of BGS as McVicker et

416   al. found that biases presented by model assumptions or mis-inference on the exact

417   value of *B* do not significantly change the overall rank order for the inferred strength of

418   BGS across the genome. Thus we, expect McVicker et al.'s inference of *B* to provide

419   good separation between the regions experiencing the weakest and strongest effects of

420   selection at linked sites within the human genome, with model misspecification unlikely

421   to change our empirical results.

422   While the effects of selection at linked sites in humans captured in our analyses

423   could in principle include the consequences of positive selection, such as soft-sweeps

424   and classic selective sweeps, we applied stringent filters to remove any such regions

425   before our analysis (Materials and methods, S1 Appendix). We therefore expect that

426   most of the resulting patterns we observe can be attributed to the consequences of

427   BGS. Recently, a joint model of classic selective sweeps and BGS was applied to *Dro-*

428   *sophila* and predicted that BGS has had a ~1.6 to 2.5-fold greater effect on neutral ge-

429   netic diversity than classic selective sweeps [33], with another study giving similar re-

430   sults [69]. We should expect this magnitude to be even greater for humans, since clas-

431   sic selective sweeps were found to be rare in recent human evolution [41] and adaptive

432   substitutions in the human genome are much less frequent than in *Drosophila* [5,70,71].

433   Nonetheless, we cannot rule out all contributions from hitchhiking to our results. In fact,

434   our own simulations of BGS fail to capture the complete effects of linked selection on

435   reducing π/π_0 in different human populations (compare Figs 1C and 5C) and the addi-

436   tional contribution of hitchhiking to humans, which we did not simulate, may explain this

437   discrepancy.

438   Non-equilibrium demography has also been of recent interest in regards to its

439   impact on patterns of deleterious variation across human populations (often referred to

440   as genetic load), with initial work showing that non-African populations have a greater

441   proportion of non-synonymous deleterious variants that are homozygous compared to

442   synonymous variants [56]. Similar results in human founder populations [57,72],

443   *Arabadopsis* [73], and domesticated species such as dogs [12] and sunflowers [74]

444   further demonstrate the pervasive impact that demography has on influencing the

445   relative amount of deleterious variation across a variety of populations and species.

446   Since BGS is a function of deleterious variation, it is not surprising that we also witness

447   differences in π/π_0 across human populations that have experienced different

448   demographic histories. These effects are likley ubiquitous across other species as well.

449   However, there has been recent contention about whether the previously described

450   patterns of increased deleterious variation are driven by a decrease in the efficacy of

451   natural selection (thus resulting in increased genetic load) or are solely artifacts of the

452   response of deleterious variation to demographic change [58,75–78]. Recently, Koch et

453   al. [55] investigated the temporal dynamics of demography on selected sites within

454   humans and observed that after a population contraction, heterozygosity at selected

455   sites can undershoot its expected value at equilibrium as low-frequency variants are lost

456   at a quicker rate before the recovery of intermediate frequency variants can occur. In

457   the context of both BGS and hitchhiking, which skew the site frequency spectrum of

458   linked neutral mutations towards rare variants [26,66,79,80], we also expect a transient

459   decrease in diversity as low-frequency variants are lost quickly during a population

460   contraction. Indeed, as evident from our simulations of BGS and demography,

21

461   immediatley after a population bottleneck, rapid losses in singleton density can occur,

462   leading to transient decreases in ψ/ψ_0. However, the recovery in singleton density is

463   also quite rapid, while the recovery in π and π/π_0 is quite slow. This is due to the fact that

464   higher frequency variants, which contribute a greater amount to π, take a longer amount

465   of time to recover after a population contraction compared to lower-frequency variants

466   such as singletons. However, Koch et al. also demonstrated that the effect of

467   demography on diversity is only temporary and that long-term diversity at selected sites

468   approaches greater values once equilibrium is reached. We also stress these temporal

469   effects on the patterns of relative neutral diversity that we observe here. In the context

470   of BGS, if no population recovery occurs following a contracation, deleterious variants

471   may behave more neutrally, weaking the overall effect of BGS and causing π/π_0 to

472   become larger than in the ancestral, pre-bottleneck population.

473   The temporal effects of non-equilbrium demographics on patterns of π/π_0 and

474   ψ/ψ_0 may also explain the conflicting results obtained in a similar study of selection at

475   linked sites in teosinte and its domesticated counterpart, maize [59]. In that study, the

476   authors observed that π/π_0 was higher in maize, which underwent a population bottle-

477   neck during domestication (no bottleneck event was inferred for the teosinte population)

478   but that ψ/ψ_0 was also lower. This result is contrary to what we observed qualitatively

479   between non-African and African human populations. However, the demographic mod-

480   els that have been inferred for maize and humans are quite different. Maize is inferred

481   to have had a recent, major domestication bottleneck that was essentially instantaneous

482   and followed by rapid exponential growth [59]. In contrast, demographic models for non-

483   African humans suggest a much more distant bottleneck that was sustained over a

484   longer period of time, and only recently have non-African populations experienced ram-

485   pant growth (coinciding with the advent of agriculture). Given the vastly different demo-

486   graphic models of the two species, it is therefore not surprising that patterns of variation

487   would be different.

488   The greater contemporary N_e of non-Africans could theoretically result in a great-

489   er efficacy of purifying selection and, consequently, a stronger efficacy of BGS, leading

490   to their observed lower π/π_0 relative to Africans. However, it is very unlikely that this

491   explains the observed patterns of relative diversity that we see. The greater contempo-

492   rary population size of non-Africans has transpired only in the very recent past, with ac-

493   celerated growth in Europeans occurring within the last few hundred generations

494   [23,81–84]. Rather, our simulations indicate that the response of π in regions under

495   BGS is driven by the ancient, sustained population contractions humans have experi-

496   enced, with reductions in π/π_0 occurring concomitantly with the out-of-Africa bottleneck

497   and European-East Asian split bottleneck events (Fig 5) and continuing even after the

498   European and East Asian expansion events. In addition, our empirical measurements of

499   ψ/ψ_0 demonstrate that in recent time, BGS has had a lower effect in non-Africans com-

500   pared to Africans. Our simulations further corroborate this result. The greater sensitivity

501   of regions under BGS to population contractions and accelerated genetic drift is a more

502   likely explanation. Broadly, our results show that contemporary patterns of neutral di-

503   versity cannot easily be attributable to contemporary forces of selection but instead may

504   be exhibiting signatures that are still dominated by older demographic events. Interest-

505   ingly though, our simulations reveal an additional factor that can influence the impact of

506   BGS within populations – migration between populations. We observe that the ex-

23

507   change of deleterious variants from populations that have experienced extensive bottle-

508   necks to populations with a more stable demography can magnify the strength of selec-

509   tion at linked sites. In particular, our simulations show that both π/π_0 and ψ/ψ_0 decrease

510   in Africans despite the fact that they are inferred to have been constant in size in their

511   recent evolutionary history (Fig 5B). These patterns disappear when migration is re-

512   moved (Fig S7, S12 Fig B), however more work is needed to definitively test this.

513   While we describe here the differential effects of non-equilibrium demography on

514   neutral diversity in regions under strong and weak BGS, it is worth mentioning that dif-

515   ferences in the reduction of neutral diversity in the genome between different popula-

516   tions have also been investigated at the level of entire chromosomes. In particular,

517   analyses of neutral diversity comparing autosomes to non-autosomes (i.e., sex chromo-

518   somes and the mitochondrial genome [mtDNA]) have been conducted. Interestingly,

519   these studies have shown that population contractions have impacted the relative re-

520   duction of neutral diversity between non-autosomes and autosomes in a similar fashion

521   to what we have observed between regions of strong BGS and weak BGS, with the

522   greatest losses occurring in bottlenecked populations. This was demonstrated in hu-

523   mans [85] and later modeled and shown in other species [86], with the explanation that

524   stronger genetic drift due to the lower N_e of non-autosomes causes diversity to be lost

525   more quickly in response to population size reductions. Recent work in humans has

526   confirmed such predictions by showing that relative losses of neutral diversity in the

527   non-autosomes are greatest for non-Africans [87–89]. These studies, plus others [90],

528   have also shown that there is strong evidence for a more dominant effect of linked se-

529   lection on the sex chromosomes relative to the autosomes in humans.

24

530   Since linked selection is a pervasive force in shaping patterns of diversity across

531   the genomes in a range of biological species [1], it has been provided as an argument

532   for why neutral diversity and estimates of N_e are relatively constrained across species in

533   spite of the large variance in census population sizes that exist [47,91]. However, since

534   population bottlenecks are common among species and have an inordinate influence on

535   N_e [20], demography has also been argued as a major culprit for constrained diversity

536   [2,91–93]. Yet, as we show in humans, it is likely that patterns of neutral diversity are in

537   fact jointly impacted by both of these forces, magnifying one another to deplete levels of

538   diversity beyond what is expected by either one independently. This may play an even

539   larger role in higher N_e species such as *Drosophila*, where the overall distribution of *B*

540   was inferred to be even smaller (i.e., exhibiting stronger BGS) than in humans [32]. In

541   our work, we also identify a potentially substantial role for migration from smaller popu-

542   lations that harbor more strongly deleterious alleles on patterns of linked neutral diversi-

543   ty in large populations. Together, these combined effects may help provide additional

544   clues for the puzzling lack of disparity in genetic diversity among different species [94].

545   Finally, our results also have implications for medical genetics research, since

546   selection may be acting on functional regions contributing to disease susceptibility.

547   Since different populations will have experienced different demographic histories, the

548   action of linked selection may result in disparate patterns of genetic variation (with ele-

549   vated levels of drift) near causal loci. Recent work has already demonstrated that BGS's

550   consequence of lowering diversity impacts power for disease association tests [95]. Our

551   results indicate that this impact may be even further exacerbated by demography in bot-

552   tlenecked populations, leading to potentially larger discrepancies in power between dif-

25

553   ferent populations. Overall, this should encourage further scrutiny for tests and SNP

554   panels optimized for one population since they may not be easily translatable to other

555   populations. It should also further motivate investigators to simultaneously account for

556   demography and linked selection when performing tests to uncover disease variants

557   within the genome [95–97].

558

Materials and methods

560   **Data**

561   2,504 samples from 26 populations in phase 3 of the Thousand Genomes Project

562   (TGP) [9] were downloaded from ftp://ftp.ncbi.nlm.nih.gov/1000genomes/. vcftools

563   (v0.1.12a) [98] and custom python scripts were used to gather all bi-allelic SNP sites

564   from the autosomes of the entire sample set.

565   A subset of TGP samples that were sequenced to high coverage (~45X) by

566   Complete Genomics (CG) were downloaded from

567   ftp://ftp.ncbi.nlm.nih.gov/1000genomes/. After filtering out related individuals via pedi-

568   gree analyses, we analyzed 53 YRI, 64 CEU, and 62 CHS samples (S2 Table). The

569   cgatools (v1.8.0) listvariants program was first used to gather all SNPs from the 179

570   samples using their CG ASM "Variations Files" (CG format version 2.2). Within each

571   population, the number of reference and alternate allele counts for each SNP was then

572   calculated using the cgatools testvariants program and custom python scripts. Only al-

573   lele counts across high quality sites (i.e., those classified as VQHIGH variant quality by

574   CG) were included. Low quality sites (i.e., those with VQLOW variant quality) were

575   treated as missing data. Only autosomes were kept. Non-bi-allelic SNPs and sites vio-

576   lating Hardy-Weinberg equilibrium (HWE) (p-value < 0.05 with a Bonferroni correction

577   for multiple SNP testing) were also removed.

578   We collected 13 whole-genome sequenced KhoeSan samples (sequence-

579   coverage: 2.5-50X, see S3 Table in Supporting information) from 3 studies [99–101]

580   and used the processed vcf files from each of those respective studies to gather all bi-

581   allelic polymorphic SNPs (i.e., the union of variants across all vcf files). SNPs were only

582   retained if they were polymorphic within the 13 samples (i.e., sites called as alternate

583   only within the sample set were ignored).

584

585   **Filtering and ascertainment scheme**

586   Positions in the genome were annotated for background selection by using the

587   background selection coefficient, B, which was inferred by McVicker et al. [6] and down-

588   loaded from http://www.phrap.org/othersoftware.html. B was inferred by applying a clas-

589   sical model of BGS [60], which treats its effects as a simple reduction in N_e at neutral

590   sites as a function of their recombination distance from conserved loci, the strength of

591   purifying selection at those conserved loci, and the deleterious mutation rate. B can be

592   interpreted as the reduced fraction of neutral genetic diversity at a particular site along

593   the genome that is caused by BGS, with a value of 0 indicating a near complete remov-

594   al of neutral genetic diversity due to BGS and a B value of 1 indicating little to no effect

595   of BGS on neutral genetic diversity ($B = \pi/\pi_0 = N_e/N_0$). Positions for B were lifted over

596   from hg18 to hg19 using the UCSC liftOver tool. Sites that failed to uniquely map from

597   hg18 to hg19 or failed to uniquely map in the reciprocal direction were excluded. Sites

598   lacking a B value were also ignored. We focused our analyses on those regions of the

27

599   genome within the top 1%, 5%, 10%, and 25% of the genome-wide distribution of *B* and

600   within the bottom 1% of the genome-wide distribution of *B*. These quantiles correspond

601   to the *B* values 0.095, 0.317, 0.463, 0.691, and 0.994, respectively.

602   A set of 13 filters (referred to as the "13-filter set") were used to limit errors from

603   sequencing and misalignments with rhesus macaque and to remove regions potentially

604   under the direct effects of natural selection and putative selective sweeps (we ignore

605   the linked selection effects of background selection). These filters were applied to all

606   samples in phase 3 TGP (all filters are in build hg19) for all sets of analyses (see S4

607   Table in Supporting information for the total number of Mb that passed the described

608   filters below for each particular *B* quantile):

609   1. Coverage/exome: For phase 3 data, regions of the genome that were part of

610   the   high   coverage   exome   were   excluded   (see

611   ftp://ftp.ncbi.nlm.nih.gov/1000genomes/ftp/technical/reference/exome_pull_do

612   wn_targets/20130108.exome.targets.bed.README). This was done to limit

613   biases due to differing levels of coverage across the genome and to remove

614   likely functional sites within the exome.

615   2. phyloP: Sites with phyloP [102] scores > 1.2 or < -1.2 were removed to limit

616   the effects of natural selection due to conservation or accelerated evolution.

617   Scores   were   downloaded   from

618   http://hgdownload.cse.ucsc.edu/goldenPath/hg19/phyloP46way/.

619   3. phastCons: Regions in the UCSC conservation 46-way track (table:

620   phastCons46wayPlacental) [103] were removed to limit the effects of natural

621   selection due to conservation.

28

4. CpG: CpG islands in the UCSC CpG islands track were removed because of their potential role in gene regulation and/or being conserved.

5. ENCODE blacklist: Regions with high signal artifacts from next-generation sequencing experiments discovered during the ENCODE project [104] were removed.

6. Accessible genome mask: Regions not accessible to next-generation sequencing using short reads, according to the phase 3 TGP "strict" criteria, were removed (downloaded from ftp://ftp.ncbi.nlm.nih.gov/1000genomes/ftp/release/20130502/supporting/acce ssible_genome_masks/StrictMask/).

7. Simple repeats: Regions in the UCSC simple repeats track were removed due to potential misalignments with outgroups and/or being under natural selection.

8. Gaps/centromeres/telomeres: Regions in the UCSC gap track were removed, including centromeres and telomeres.

9. Segmental duplications: Regions in the UCSC segmental dups track [105] were removed to limit potential effects of natural selection and/or misalignments with rhesus macaque.

10. Transposons: Active transposons (HERVK retrotransposons, the AluY subfamily of Alu elements, SVA elements, and L1Ta/L1pre-Ta LINEs) in the human genome were removed.

11. Recent positive selection: Regions inferred to be under hard and soft selective sweeps (using iHS and iHH12 regions from selscan v1.2.0 [106]; S1 Ap-

29

645   pendix) within each phase 3 population were removed.

646   12. Non-coding transcripts: Non-coding transcripts from the UCSC genes track

647   were removed to limit potential effects of natural selection.

648   13. Synteny: Regions that did not share conserved synteny with rhesus macaque

649   (rheMac2) from UCSC syntenic net filtering were removed (downloaded from

650   http://hgdownload.soe.ucsc.edu/goldenPath/hg19/vsRheMac2/syntenicNet/).

651 Additionally, an extra set of filters was applied, but only for those estimates of diversity

652 that controlled for GC-biased gene conversion and recombination hotspots:

653   14. GC-biased gene conversion (gBGC): Regions in UCSC phastBias track [107]

654   from UCSC genome browser were removed to limit regions inferred to be un-

655   der strong GC-biased gene conversion.

656   15. Recombination hotspots: All sites within 1.5 kb (i.e., 3 kb windows) of sites

657   with recombination rates ≥ 10 cM/Mb in the 1000G OMNI genetic maps for

658   non-admixed populations (downloaded from ftp://ftp-

659   trace.ncbi.nih.gov/1000genomes/ftp/technical/working/20130507_omni_reco

660   mbination_rates/) and the HapMap II genetic map [108] were removed. 1.5 kb

661   flanking regions surrounding the center of hotspots identified by Ref. [109]

662   (downloaded from

663   http://science.sciencemag.org/content/sci/suppl/2014/11/12/346.6211.125644

664   2.DC1/1256442_DatafileS1.txt) were also removed, except for the cases in

665   which the entire hotspot site was greater than 3 kb in length (in which case

666   just the hotspot was removed).

667   To generate a set of four-fold degenerate synonymous sites, all polymorphic

668   sites that we retained from the high-coverage Complete Genomic samples were anno-

669   tated using the program ANNOVAR [110] with Gencode V19 annotations. ANNOVAR

670   and Gencode V19 annotations were also used to gather an autosome-wide set of four-

671   fold degenerate sites, resulting in 5,188,972 total sites.

672

673   **Demographic inference**

674   The inference tool dadi (v1.6.3) [7] was used to fit, via maximum likelihood, the 3-

675   population 13-parameter demographic model of Gutenkunst et al. [7] to the 179 YRI,

676   CEU, and CHS samples from the high coverage CG dataset of TGP. This sample set

677   consisted of 53 YRI (African), 64 CEU (European), and 62 CHS (East Asian) samples.

678   The demographic model incorporates an ancient human expansion in Africa and a sin-

679   gle out-of-Africa bottleneck followed by European- and East Asian-specific bottlenecks,

680   as well as exponential growth in both non-African populations and migration between

681   populations. During the inference procedure, each population was projected down to

682   106 chromosomes, corresponding to the maximum number of chromosomes available

683   in the CG YRI population. Sites were polarized with chimpanzee to identify putative an-

684   cestral/derived alleles using the chain and netted alignments of hg19 with panTro4

685   (http://hgdownload.soe.ucsc.edu/goldenPath/hg19/vsPanTro4/axtNet/), and the correc-

686   tion for ancestral misidentification [111] option in dadi was used. The 13-filter set de-

687   scribed previously was applied to the CG data set, and an additional filter keeping only

688   the autosomal sites in the top 1% of B ($B \geq 0.994$) was also applied in order to mitigate

689   potential biases in inference due to BGS [52,65] or other forms of linked selection [51].

690   After site filtering and correction for ancestral misidentification, a total of 110,582 segre-

691 gating sites were utilized by dadi for the inference procedure. For optimization, grid

692 points of 120, 130, and 140 were used, and 15 independent optimization runs were

693 conducted from different initial parameter points to ensure convergence upon a global

694 optimum. An effective sequence length (L) of 7.15 Mb was calculated from the input se-

695 quence data after accounting for the fraction of total sites removed due to filtering. In

696 addition to the 13-filter set, this filtering included sites violating HWE, sites without B

697 value information, sites that did not have at least 106 sampled chromosomes in each

698 population, sites with more than two alleles, sites that did not have tri-nucleotide infor-

699 mation for the correction for ancestral misidentification step, and sites treated as miss-

700 ing data. For calculating the reference effective population size, a mutation rate (μ) of

701 1.66×10^{-8} (inferred from Ref. [112]) was used. Using the optimized θ from dadi after

702 parameter fitting, the equation $\theta = 4N_e\mu L$ was solved for N_e to generate the reference

703 effective population size, from which all other population N_e's were calculated. This

704 same procedure was also used to infer demographic parameters from four-fold degen-

705 erate synonymous sites across the same set of samples. After site filtering (note that B

706 and the 13-filter set were not included in the filtering step for four-fold degenerate syn-

707 onymous sites), 41,260 segregating sites were utilized by dadi for the inference proce-

708 dure, and an effective sequence length of 2.37 Mb was used for calculating the refer-

709 ence effective population size.

710

711 **Simulations**

712   Forward simulations incorporating the results from the demographic inference

713 procedure described above and a model of background selection were conducted using

32

714   SFS_CODE [113]. For the model of background selection, the recombination rate, ρ,

715   and the fraction of the genome experiencing deleterious mutation were calculated using

716   the 2 Mb region of chr3: 48,600,000-50,600,000, which has been subject to the strong-

717   est amount of BGS in the human genome (mean B = 0.002). A population-scaled re-

718   combination rate (ρ) of 6.0443 x 10^{-5} was calculated for this region using the HapMap II

719   GRCh37 genetic map [108]. For ascertaining the fraction of sites experiencing deleteri-

720   ous mutation, the number of non-coding "functional" sites in this region was first calcu-

721   lated by taking the union of all phastCons sites and phyloP sites with scores > 1.2 (indi-

722   cating conservation) that did not intersect with any coding exons. This amount totaled to

723   270,348 base pairs. Additionally, the number of coding sites was calculated by summing

724   all coding exons within this region from GENCODE v19, which totaled to 138,923 base

725   pairs. From these totals, the total fraction of deleterious sites, 0.2046, was generated.

726   The background selection model was simulated using a middle 30 kb neutral re-

727   gion flanked by two 1 Mb regions under purifying selection. From the calculated fraction

728   of deleterious sites described above, 20.46% of sites in the two 1 Mb flanking regions

729   were simulated as being deleterious. The mutation rate in our simulations for the delete-

730   rious sites and for neutral sites were both set to 1.66 x 10^{-8} [112]. Two distributions of

731   fitness effects were used for the deleterious sites, with 66.06% of deleterious sites using

732   the gamma distribution of fitness effects inferred across conserved non-coding regions

733   by Ref. [114] (β = 0.0415, α = 0.00515625) and 33.94% of deleterious sites using the

734   gamma distribution of fitness effects inferred across coding regions by Ref. [5] (β =

735   0.184, α = 0.00040244). The relative number of non-coding "functional" sites and coding

736   exons described above determined the relative number of sites receiving each distribu-

33

737   tion of fitness effects in our simulations. Gamma distribution parameters were scaled to

738   the ancestral population size of the demographic models used in Refs. [5,114]. An ex-

739   ample of the SFS_CODE command for our simulations is in S1 Supporting information.

740   To simulate varying levels of background selection strength, different total fractions of

741   our original calculated deleterious fraction of 0.2046 were used (i.e., 5%, 10%, 25%,

742   50%, and 100% of 0.2046). However, the same relative percentage of non-coding and

743   coding sites and mutation rate were used. These different simulated fractions of delete-

744   rious sites resulted in a reduced total deleterious mutation rate, U, which is the product

745   of the per-site deleterious mutation rate and the total number of sites experiencing dele-

746   terious mutation [66]. Thus, weaker effects of BGS were simulated. To simulate only the

747   effects of demography without background selection, only the 30 kb neutral region was

748   simulated. 2,000 independent simulations were conducted for each particular set of the

749   deleterious site fraction (2,000 x 6 = 12,000 total). Simulations output population genetic

750   information every 100 generations and also at each generation experiencing a popula-

751   tion size change (22,117 total generations were simulated), from which mean pairwise

752   nucleotide diversity (π) and singleton density (ψ) was calculated across the 2,000 simu-

753   lations.

754

755   **Population-specific calculations of diversity and singleton density**

756   Mean pairwise genetic diversity (π) and singleton density (ψ) was calculated as a

757   function of the *B* quantile bins described in "Filtering and ascertainment scheme" for

758   each of the 20 non-admixed populations in phase 3 TGP and, for π, across 4 broad

759   populations that grouped the 20 non-admixed populations together by continent (Afri-

34

760 can, European, South Asian, and East Asian, see S12 Table in Supporting information).

761 Additionally, only regions of the genome passing the 13-filter set were used in the calcu-

762 lations of π and ψ (see S4 Table in Supporting information for total number of Mb used

763 in diversity calculations for each B quantile). When calculating ψ for each non-admixed

764 phase 3 TGP population, the site-frequency spectrum was first projected down to 2N =

765 170 samples (the number of chromosomes in MSL, the smallest phase 3 population

766 sample) using a hypergeometric distribution [7] from each population's full site-

767 frequency spectrum. This allowed for unbiased comparisons of singleton density be-

768 tween all populations. For estimates of diversity controlling for gBGC or recombination

769 hotspots, the additional corresponding filters described in "Filtering and ascertainment

770 scheme" were also used. Only 100 kb regions of the genome with at least 10 kb of di-

771 vergence information with Rhesus macaque were used in π and ψ calculations (see

772 "Normalization of diversity and divergence calculations with Rhesus macaque" below).

773

774 **Normalization of diversity/singleton density and divergence calculations with**

775 **Rhesus macaque**

776   To calculate human divergence with Rhesus macaque, we downloaded the

777 syntenic net alignments between hg19 and rheMac2 that were generated by blastz from

778 http://hgdownload.cse.ucsc.edu/goldenpath/hg19/vsRheMac2/syntenicNet/. We binned

779 the human genome into non-overlapping 100 kb bins and calculated divergence within

780 each bin by taking the proportion of base pair differences between human and Rhesus

781 macaque. Gaps between human and Rhesus macaque, positions lacking alignment in-

782 formation, and positions that did not pass the 13-filter set described in "Filtering and

35

783   ascertainment scheme" were ignored in the divergence estimate. Additionally, a sepa-

784   rate set of divergence estimates were also made using the additional set of filtering cri-

785   teria that removed those regions under gBGC or in recombination hotspots and were

786   used for normalizing diversity in those measurements that controlled for gBGC and

787   hotspots.

788   When normalizing diversity and singleton density by divergence, only 100 kb bins

789   that had at least 10 kb of divergence information were used (21,100 bins total for 13-

790   filter set, 20,935 bins total for the 13-filter set plus the additional gBGC and hotspot fil-

791   ters). Bins with less than 10 kb of divergence information were ignored. To make esti-

792   mates comparable, in those measurements of diversity that did not normalize by diver-

793   gence, diversity was still calculated using the same set of 100 kb bins that had at least

794   10 kb for estimating divergence.

795

796   **Calculations of population differentiation (F_{ST}) and linear regression**

797   F_{ST} calculations were performed as a function of B between every pair of non-

798   admixed phase 3 TGP populations not belonging to the same continental group (150

799   pairs total). We followed the recommendations in Bhatia et al. [63] to limit biases in F_{ST}

800   due to 1) type of estimator used, 2) averaging over SNPs, and 3) SNP ascertainment.

801   Specifically, we 1) used the Hudson-based F_{ST} estimator [115], 2) used a ratio of

802   averages for combining F_{ST} estimated across different SNPs, and 3) ascertained SNPs

803   based on being polymorphic in an outgroup (i.e., the KhoeSan). For ascertaining SNPs

804   in the KhoeSan, we also performed filtering according to the filtering scheme described

805   under "Filtering and ascertainment scheme." For a position to be considered

36

806   polymorphic in the KhoeSan, at least one alternate allele and one reference allele had

807   to be called across the 13 genomes we utilized (see "Data"). These criteria left

808   3,497,105 total sites in the genome in the phase 3 dataset for F_{ST} to be estimated

809   across.

810   F_{ST} was calculated across 2% quantile bins of B (based on the genome-wide

811   distribution of B) for all pairwise comparisons of populations between a specific pair of

812   continental groups (25 pairs total) or across all pairwise comparisons using all continen-

813   tal groups (150 pairs total). Simple linear regression was performed with the model F_{ST}

814   = $\beta_0 + \beta_1 B + \varepsilon$. The mean of the bounds defining each quantile bin was used when de-

815   fining the explanatory variables for the regression. Linear regression, robust linear re-

816   gression [64], and simple correlation were performed using the lm(), rlm(), and cor()

817   functions, respectively, in the R programming language (www. r-project.org). To gener-

818   ate standard errors of the mean, this same procedure was performed on F_{ST} results

819   generated from each of 1,000 bootstrapped iterations of the data.

820

821   **Bootstrapping**

822   **Diversity Estimates.** To control for the structure of linkage disequilibrium and correla-

823   tion between SNPs along the genome, we partitioned the human genome into non-

824   overlapping 100 kb bins (these bins were identical to the 100 kb bins used for estimat-

825   ing divergence) and calculated mean pairwise diversity (π) or heterozygosity within each

826   bin. We also normalized the diversity estimates by divergence within each bin. We then

827   bootstrapped individual genomes by sampling, with replacement, the 100 kb bins until

828   the number of sampled bins equaled the number of bins used for calculating the diversi-

829   ty point estimates (i.e., 21,100 bins or 20,935 bins total, depending on whether filters for

830   gBGC and hotspots were applied). 1,000 total bootstrap iterations were completed and

831   standard errors of the mean were calculated by taking the standard deviation from the

832   resulting bootstrap distribution.

833   **F_{ST}.** For bootstrapping F_{ST}, the human genome was partitioned into non-overlapping

834   100 kb bins and were sampled with replacement until 28,823 bins were selected (the

835   total number of non-overlapping 100 kb bins in the human autosomes). F_{ST} was then

836   calculated genome-wide for the bootstrapped genome as a function of B for every pair-

837   wise comparison of non-admixed phase 3 TGP populations not belonging to the same

838   continental group. 1,000 total bootstrap iterations were completed and standard errors

839   of the mean were calculated by taking the standard deviation from the F_{ST} distribution

840   calculated from all 1,000 iterations.

841

Acknowledgements

847

References

849   1.   Cutter AD, Payseur BA. Genomic signatures of selection at linked sites: unifying
850   the disparity among species. Nat Rev Genet. 2013;14: 262–274.
851   doi:10.1038/nrg3425
852   2.   Ellegren H, Galtier N. Determinants of genetic diversity. Nat Rev Genet. 2016;17:
853   422–433. doi:10.1038/nrg.2016.58
854   3.   Sabeti PC, Reich DE, Higgins JM, Levine HZP, Richter DJ, Schaffner SF, et al.

855   Detecting recent positive selection in the human genome from haplotype
856   structure. Nature. 2002;419: 832–837. doi:10.1038/nature01027.1.
857   4.   Williamson SH, Hernandez R, Fledel-Alon A, Zhu L, Nielsen R, Bustamante CD.
858   Simultaneous inference of selection and population growth from patterns of
859   variation in the human genome. Proc Natl Acad Sci. 2005;102: 7882–7887.
860   doi:10.1073/pnas.0502300102
861   5.   Boyko AR, Williamson SH, Indap AR, Degenhardt JD, Hernandez RD, Lohmueller
862   KE, et al. Assessing the evolutionary impact of amino acid mutations in the
863   human genome. PLoS Genet. 2008;4: e1000083.
864   doi:10.1371/journal.pgen.1000083
865   6.   McVicker G, Gordon D, Davis C, Green P. Widespread genomic signatures of
866   natural selection in hominid evolution. PLoS Genet. 2009;5: e1000471.
867   doi:10.1371/journal.pgen.1000471
868   7.   Gutenkunst RN, Hernandez RD, Williamson SH, Bustamante CD. Inferring the
869   joint demographic history of multiple populations from multidimensional SNP
870   frequency data. PLoS Genet. 2009;5: e1000695.
871   doi:10.1371/journal.pgen.1000695
872   8.   Li H, Durbin R. Inference of human population history from individual whole-
873   genome sequences. Nature. 2011;475: 493–496. doi:10.1038/nature10231
874   9.   Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al.
875   A global reference for human genetic variation. Nature. 2015;526: 68–74.
876   doi:10.1038/nature15393
877   10. Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. A thousand fly
878   genomes: An expanded Drosophila genome nexus. Mol Biol Evol. 2016;33: 3308–
879   3313. doi:10.1093/molbev/msw195
880   11. Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversole KA, Gill CA, et al.
881   Genome-wide survey of SNP variation uncovers the genetic structure of cattle
882   breeds. Science. 2009;324: 528–532. doi:10.1126/science.1167936
883   12. Marsden CD, Ortega-Del Vecchyo D, O'Brien DP, Taylor JF, Ramirez O, Vilà C,
884   et al. Bottlenecks and selective sweeps during domestication have increased
885   deleterious genetic variation in dogs. Proc Natl Acad Sci. 2016;113: 152–157.
886   doi:10.1073/pnas.1512501113
887   13. Caicedo AL, Williamson SH, Hernandez RD, Boyko A, Fledel-Alon A, York TL, et
888   al. Genome-wide patterns of nucleotide polymorphism in domesticated rice. PLoS
889   Genet. 2007;3: 1745–1756. doi:10.1371/journal.pgen.0030163
890   14. Begun DJ, Aquadro CF. African and North American populations of Drosophila
891   melanogaster are very different at the DNA level. Nature. 1993;365: 548–550.
892   doi:10.1038/365548a0
893   15. Haddrill PR, Thornton KR, Charlesworth B, Andolfatto P. Multilocus patterns of
894   nucleotide variability and the demographic and selection history of Drosophila
895   melanogaster populations. Genome Res. 2005;15: 790–799.
896   doi:10.1101/gr.3541005
897   16. Ometto L, Glinka S, De Lorenzo D, Stephan W. Inferring the effects of
898   demography and selection on Drosophila melanogaster populations from a
899   chromosome-wide scan of DNA variation. Mol Biol Evol. 2005;22: 2119–2130.
900   doi:10.1093/molbev/msi207

901   17.   Hernandez RD, Hubisz MJ, Wheeler DA, Smith DG, Ferguson B, Rogers J, et al.
902   Demographic histories and patterns of linkage disequilibrium in Chinese and
903   Indian rhesus macaques. Science. 2007;316: 240–243.
904   doi:10.1126/science.1140462
905   18.   Ramachandran S, Deshpande O, Roseman CC, Rosenberg NA, Feldman MW,
906   Cavalli-Sforza LL. Support from the relationship of genetic and geographic
907   distance in human populations for a serial founder effect originating in Africa. Proc
908   Natl Acad Sci. 2005;102: 15942–15947. doi:10.1073/pnas.0507611102
909   19.   Henn BM, Cavalli-Sforza LL, Feldman MW. The great human expansion. Proc
910   Natl Acad Sci. 2012;109: 17758–17764. doi:10.1073/pnas.1212380109
911   20.   Charlesworth B. Effective population size and patterns of molecular evolution and
912   variation. Nat Rev Genet. 2009;10: 195–205. doi:10.1038/nrg2526
913   21.   Nei M, Maruyama T, Chakraborty R. The bottleneck effect and genetic variability
914   in populations. Evolution (N Y). 1975;29: 1–10. doi:10.2307/2407137
915   22.   Gravel S, Henn BM, Gutenkunst RN, Indap AR, Marth GT, Clark AG, et al.
916   Demographic history and rare allele sharing among human populations. Proc Natl
917   Acad Sci. 2011;108: 11983–11988. doi:10.1073/pnas.1019276108
918   23.   Tennessen JA, Bigham AW, O'Connor TD, Fu W, Kenny EE, Gravel S, et al.
919   Evolution and functional impact of rare coding variation from deep sequencing of
920   human exomes. Science. 2012;337: 64–69. doi:10.1126/science.1219240
921   24.   Charlesworth D. Balancing selection and its effects on sequences in nearby
922   genome regions. PLoS Genet. 2006;2: 379–384.
923   doi:10.1371/journal.pgen.0020064
924   25.   Maynard Smith J, Haigh J. The hitch-hiking effect of a favourable gene. Genet
925   Res. 1974;23: 23–35. doi:10.1017/S0016672308009579
926   26.   Charlesworth B, Morgan MT, Charlesworth D. The effect of deleterious mutations
927   on neutral molecular variation. Genetics. 1993;134: 1289–1303.
928   27.   Kim Y, Stephan W. Joint effects of genetic hitchhiking and background selection
929   on neutral variation. Genetics. 2000;155: 1415–1427.
930   28.   Begun DJ, Aquadro CF. Levels of naturally occurring DNA polymorphism
931   correlate with recombination rates in D. melanogaster. Nature. 1992;356: 519–
932   520. doi:10.1038/356519a0
933   29.   Charlesworth B. Background selection and patterns of genetic diversity in
934   Drosophila melanogaster. Genet Res. 1996;68: 131–149.
935   doi:10.1017/S0016672300034029
936   30.   Andolfatto P. Hitchhiking effects of recurrent beneficial amino acid substitutions in
937   the Drosophila melanogaster genome. Genome Res. 2007;17: 1755–1762.
938   doi:10.1101/gr.6691007
939   31.   Sella G, Petrov DA, Przeworski M, Andolfatto P. Pervasive natural selection in the
940   Drosophila genome? PLoS Genet. 2009;5: e1000495.
941   doi:10.1371/journal.pgen.1000495
942   32.   Comeron JM. Background selection as baseline for nucleotide variation across
943   the drosophila genome. PLoS Genet. 2014;10: e1004434.
944   doi:10.1371/journal.pgen.1004434
945   33.   Elyashiv E, Sattath S, Hu TT, Strutsovsky A, McVicker G, Andolfatto P, et al. A
946   genomic map of the effects of linked selection in Drosophila. PLoS Genet.

947   2016;12: e1006130. doi:10.1371/journal.pgen.1006130

948   34.   Flowers JM, Molina J, Rubinstein S, Huang P, Schaal BA, Purugganan MD.
949   Natural selection in gene-dense regions shapes the genomic pattern of
950   polymorphism in wild and domesticated rice. Mol Biol Evol. 2012;29: 675–687.
951   doi:10.1093/molbev/msr225

952   35.   Xu X, Liu X, Ge S, Jensen JD, Hu F, Li X, et al. Resequencing 50 accessions of
953   cultivated and wild rice yields markers for identifying agronomically important
954   genes. Nat Biotechnol. 2012;30: 105–111. doi:10.1038/nbt.2050

955   36.   Andersen EC, Gerke JP, Shapiro JA, Crissman JR, Ghosh R, Bloom JS, et al.
956   Chromosome-scale selective sweeps shape Caenorhabditis elegans genomic
957   diversity. Nat Genet. 2012;44: 285–290. doi:10.1038/ng.1050

958   37.   Cutter AD, Payseur BA. Selection at linked sites in the partial selfer
959   Caenorhabditis elegans. Mol Biol Evol. 2003;20: 665–673.
960   doi:10.1093/molbev/msg072

961   38.   Reed FA, Akey JM, Aquadro CF. Fitting background-selection predictions to
962   levels of nucleotide variation and divergence along the human autosomes.
963   Genome Res. 2005;15: 1211–1221. doi:10.1101/gr.3413205

964   39.   Voight BF, Kudaravalli S, Wen X, Pritchard JK. A map of recent positive selection
965   in the human genome. PLoS Biol. 2006;4: 0446–0458.
966   doi:10.1371/journal.pbio.0040072

967   40.   Cai JJ, Macpherson JM, Sella G, Petrov DA. Pervasive hitchhiking at coding and
968   regulatory sites in humans. PLoS Genet. 2009;5: e1000336.
969   doi:10.1371/journal.pgen.1000336

970   41.   Hernandez RD, Kelley JL, Elyashiv E, Melton SC, Auton A, McVean G, et al.
971   Classic selective sweeps were rare in recent human evolution. Science.
972   2011;331: 920–924. doi:10.1126/science.1198878

973   42.   Lohmueller KE, Albrechtsen A, Li Y, Kim SY, Korneliussen T, Vinckenbosch N, et
974   al. Natural selection affects multiple aspects of genetic variation at putatively
975   neutral sites across the human genome. PLoS Genet. 2011;7: e1002326.
976   doi:10.1371/journal.pgen.1002326

977   43.   Alves I, Šrámková Hanulová A, Foll M, Excoffier L. Genomic data reveal a
978   complex making of humans. PLoS Genet. 2012;8: e1002837.
979   doi:10.1371/journal.pgen.1002837

980   44.   Granka JM, Henn BM, Gignoux CR, Kidd JM, Bustamante CD, Feldman MW.
981   Limited evidence for classic selective sweeps in African populations. Genetics.
982   2012;192: 1049–1064. doi:10.1534/genetics.112.144071

983   45.   Enard D, Messer PW, Petrov DA. Genome-wide signals of positive selection in
984   human evolution. Genome Res. 2014;24: 885–895. doi:10.1101/gr.164822.113

985   46.   Bank C, Ewing GB, Ferrer-Admettla A, Foll M, Jensen JD. Thinking too positive?
986   Revisiting current methods of population genetic selection inference. Trends
987   Genet. 2014;30: 540–546. doi:10.1016/j.tig.2014.09.010

988   47.   Corbett-Detig RB, Hartl DL, Sackton TB. Natural selection constrains neutral
989   diversity across a wide range of species. PLoS Biol. 2015;13: e1002112.
990   doi:10.1371/journal.pbio.1002112

991   48.   Comeron JM. Background selection as null hypothesis in population genomics:
992   insights and challenges from Drosophila studies. Philos Trans R Soc B. 2017;372:

993   20160471. doi:10.1098/rstb.2016.0471
994   49.   Zeng K. A coalescent model of background selection with recombination,
995   demography and variation in selection coefficients. Heredity (Edinb). 2013;110:
996   363–371. doi:10.1038/hdy.2012.102
997   50.   Nicolaisen LE, Desai MM. Distortions in genealogies due to purifying selection
998   and recombination. Genetics. 2013;195: 221–230.
999   doi:10.1534/genetics.113.152983
1000 51.   Schrider DR, Shanku AG, Kern AD. Effects of linked selective sweeps on
1001   demographic inference and model selection. Genetics. 2016;204: 1207–1223.
1002   doi:10.1534/genetics.116.190223
1003 52.   Ewing GB, Jensen JD. The consequences of not accounting for background
1004   selection in demographic inference. Mol Ecol. 2016;25: 135–141.
1005   doi:10.1111/mec.13390
1006 53.   Pennings PS, Kryazhimskiy S, Wakeley J. Loss and recovery of genetic diversity
1007   in adapting populations of HIV. PLoS Genet. 2014;10: e1004000.
1008   doi:10.1371/journal.pgen.1004000
1009 54.   Brandvain Y, Wright SI. The limits of natural selection in a nonequilibrium world.
1010   Trends Genet. 2016;32: 201–210. doi:10.1016/j.tig.2016.01.004
1011 55.   Koch E, Novembre J. A temporal perspective on the interplay of demography and
1012   selection on deleterious variation in humans. G3. 2017;7: 1027–1037.
1013   doi:10.1534/g3.117.039651
1014 56.   Lohmueller KE, Indap AR, Schmidt S, Boyko AR, Hernandez RD, Hubisz MJ, et
1015   al. Proportionally more deleterious genetic variation in European than in African
1016   populations. Nature. 2008;451: 994–997. doi:10.1038/nature06611
1017 57.   Casals F, Hodgkinson A, Hussin J, Idaghdour Y, Bruat V, de Maillard T, et al.
1018   Whole-exome sequencing reveals a rapid change in the frequency of rare
1019   functional variants in a founding population of humans. PLoS Genet. 2013;9:
1020   e1003815. doi:10.1371/journal.pgen.1003815
1021 58.   Simons YB, Sella G. The impact of recent population history on the deleterious
1022   mutation load in humans and close evolutionary relatives. Curr Opin Genet Dev.
1023   2016;41: 150–158. doi:10.1016/j.gde.2016.09.006
1024 59.   Beissinger TM, Wang L, Crosby K, Durvasula A, Hufford MB, Ross-Ibarra J.
1025   Recent demography drives changes in linked selection across the maize genome.
1026   Nat Plants. 2016;2: 16084. doi:10.1038/nplants.2016.84
1027 60.   Nordborg M, Charlesworth B, Charlesworth D. The effect of recombination on
1028   background selection. Genet Res. 1996;67: 159–174.
1029   doi:10.1017/S0016672300033619
1030 61.   Charlesworth B, Nordborg M, Charlesworth D. The effects of local selection,
1031   balanced polymorphism and background selection on equilibrium patterns of
1032   genetic diversity in subdivided populations. Genet Res. 1997;70: 155–174.
1033   doi:10.1017/S0016672397002954
1034 62.   Hu XS, He F. Background selection and population differentiation. J Theor Biol.
1035   2005;235: 207–219. doi:10.1016/j.jtbi.2005.01.004
1036 63.   Bhatia G, Patterson N, Sankararaman S, Price AL. Estimating and interpreting
1037   FST: The impact of rare variants. Genome Res. 2013;23: 1514–1521.
1038   doi:10.1101/gr.154831.113

1039   64.   Yu C, Yao W. Robust linear regression: A review and comparison. Commun Stat -
1040   Simul Comput. 2017;46: 6261–6282. doi:10.1080/03610918.2016.1202271
1041   65.   Messer PW, Petrov DA. Frequent adaptation and the McDonald-Kreitman test.
1042   Proc Natl Acad Sci. 2013;110: 8615–8620. doi:10.1073/pnas.1220835110
1043   66.   Charlesworth B. The effects of deleterious mutations on evolution at linked sites.
1044   Genetics. 2012;190: 5–22. doi:10.1534/genetics.111.134288
1045   67.   Phung TN, Huber CD, Lohmueller KE. Determining the effect of natural selection
1046   on linked neutral divergence across species. PLoS Genet. 2016;12: e1006199.
1047   doi:10.1371/journal.pgen.1006199
1048   68.   Burgess R, Yang Z. Estimation of hominoid ancestral population sizes under
1049   Bayesian coalescent models incorporating mutation rate variation and sequencing
1050   errors. Mol Biol Evol. 2008;25: 1979–1994. doi:10.1093/molbev/msn148
1051   69.   Campos JL, Zhao L, Charlesworth B. Estimating the parameters of background
1052   selection and selective sweeps in Drosophila in the presence of gene conversion.
1053   Proc Natl Acad Sci. 2017;114: E4762–E4771. doi:10.1073/pnas.1619434114
1054   70.   Eyre-Walker A, Keightley PD. Estimating the rate of adaptive molecular evolution
1055   in the presence of slightly deleterious mutations and population size change. Mol
1056   Biol Evol. 2009;26: 2097–2108. doi:10.1093/molbev/msp119
1057   71.   Galtier N. Adaptive protein evolution in animals and the effective population size
1058   hypothesis. PLoS Genet. 2016;12: e1005774. doi:10.1371/journal.pgen.1005774
1059   72.   Lim ET, Würtz P, Havulinna AS, Palta P, Tukiainen T, Rehnström K, et al.
1060   Distribution and medical impact of loss-of-function variants in the Finnish founder
1061   population. PLoS Genet. 2014;10: e1004494. doi:10.1371/journal.pgen.1004494
1062   73.   Cao J, Schneeberger K, Ossowski S, Günther T, Bender S, Fitz J, et al. Whole-
1063   genome sequencing of multiple Arabidopsis thaliana populations. Nat Genet.
1064   2011;43: 956–963. doi:10.1038/ng.911
1065   74.   Renaut S, Rieseberg LH. The accumulation of deleterious mutations as a
1066   consequence of domestication and improvement in sunflowers and other
1067   compositae crops. Mol Biol Evol. 2015;32: 2273–2283.
1068   doi:10.1093/molbev/msv106
1069   75.   Balick DJ, Do R, Cassa CA, Reich D, Sunyaev SR. Dominance of deleterious
1070   alleles controls the response to a population bottleneck. PLoS Genet. 2015;11:
1071   e1005436. doi:10.1371/journal.pgen.1005436
1072   76.   Do R, Balick D, Li H, Adzhubei I, Sunyaev S, Reich D. No evidence that selection
1073   has been less effective at removing deleterious mutations in Europeans than in
1074   Africans. Nat Genet. 2015;47: 126–131. doi:10.1038/ng.3186
1075   77.   Simons YB, Turchin MC, Pritchard JK, Sella G. The deleterious mutation load is
1076   insensitive to recent population history. Nat Genet. 2014;46: 220–224.
1077   doi:10.1038/ng.2896
1078   78.   Gravel S. When is selection effective? Genetics. 2016;203: 451–462.
1079   doi:10.1534/genetics.115.184630
1080   79.   Braverman JM, Hudson RR, Kaplan NL, Langley CH, Stephan W. The hitchhiking
1081   effect on the site frequency spectrum of DNA polymorphisms. Genetics.
1082   1995;140: 783–796.
1083   80.   Stephan W. Genetic hitchhiking versus background selection: the controversy and
1084   its implications. Philos Trans R Soc B. 2010;365: 1245–1253.

1085   doi:10.1098/rstb.2009.0278

1086 81. Coventry A, Bull-Otterson LM, Liu X, Clark AG, Maxwell TJ, Crosby J, et al. Deep
1087   resequencing reveals excess rare recent variants consistent with explosive
1088   population growth. Nat Commun. 2010;1: 131. doi:10.1038/ncomms1130

1089 82. Keinan A, Clark AG. Recent explosive human population growth has resulted in
1090   an excess of rare genetic variants. Science. 2012;336: 740–743.
1091   doi:10.1126/science.1217283

1092 83. Nelson MR, Wegmann D, Ehm MG, Kessner D, St. Jean P, Verzilli C, et al. An
1093   abundance of rare functional variants in 202 drug target genes sequenced in
1094   14,002 people. Science. 2012;337: 100–104. doi:10.1126/science.1217876

1095 84. Gazave E, Ma L, Chang D, Coventry A, Gao F, Muzny D, et al. Neutral genomic
1096   regions refine models of recent rapid human population growth. Proc Natl Acad
1097   Sci. 2014;111: 757–762. doi:10.1073/pnas.1310398110

1098 85. Fay JC, Wu CI. A human population bottleneck can account for the discordance
1099   between patterns of mitochondrial versus nuclear DNA variation. Mol Biol Evol.
1100   1999;16: 1003–1005. doi:10.1093/oxfordjournals.molbev.a026175

1101 86. Pool JE, Nielsen R. Population size changes reshape genomic patterns of
1102   diversity. Evolution (N Y). 2007;61: 3001–3006. doi:10.1111/j.1558-
1103   5646.2007.00238.x

1104 87. Gottipati S, Arbiza L, Siepel A, Clark AG, Keinan A. Analyses of X-linked and
1105   autosomal genetic variation in population-scale whole genome sequencing. Nat
1106   Genet. 2011;43: 741–743. doi:10.1038/ng.877

1107 88. Arbiza L, Gottipati S, Siepel A, Keinan A. Contrasting X-linked and autosomal
1108   diversity across 14 human populations. Am J Hum Genet. 2014;94: 827–844.
1109   doi:10.1016/j.ajhg.2014.04.011

1110 89. Wilson Sayres MA, Lohmueller KE, Nielsen R. Natural selection reduced diversity
1111   on human Y chromosomes. PLoS Genet. 2014;10: e1004064.
1112   doi:10.1371/journal.pgen.1004064

1113 90. Hammer MF, Woerner AE, Mendez FL, Watkins JC, Cox MP, Wall JD. The ratio
1114   of human X chromosome to autosome diversity is positively correlated with
1115   genetic distance from genes. Nat Genet. 2010;42: 830–831. doi:10.1038/ng.651

1116 91. Leffler EM, Bullaughey K, Matute DR, Meyer WK, Ségurel L, Venkat A, et al.
1117   Revisiting an old riddle: What determines genetic diversity levels within species?
1118   PLoS Biol. 2012;10: e1001388. doi:10.1371/journal.pbio.1001388

1119 92. Vucetich JA, Waite TA, Nunney L. Fluctuating population size and the ratio of
1120   effective to census population size. Evolution (N Y). 1997;51: 2017–2021.
1121   doi:10.2307/2411022

1122 93. Coop G. Does linked selection explain the narrow range of genetic diversity
1123   across species? bioRxiv. 2016; doi:10.1101/042598

1124 94. Lewontin RC. The genetic basis of evolutionary change. New York and London:
1125   Columbia University Press; 1974.

1126 95. Uricchio LH, Torres R, Witte JS, Hernandez RD. Population genetic simulations of
1127   complex phenotypes with implications for rare variant association tests. Genet
1128   Epidemiol. 2015;39: 35–44. doi:10.1002/gepi.21866

1129 96. Maher MC, Uricchio LH, Torgerson DG, Hernandez RD. Population genetics of
1130   rare variants and complex diseases. Hum Hered. 2012;74: 118–128.

1131   doi:10.1159/000346826

1132 97. Uricchio LH, Zaitlen NA, Ye CJ, Witte JS, Hernandez RD. Selection and explosive
1133   growth alter genetic architecture and hamper the detection of causal rare variants.
1134   Genome Res. 2016;26: 863–873. doi:10.1101/gr.202440.115

1135 98. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The
1136   variant call format and VCFtools. Bioinformatics. 2011;27: 2156–2158.
1137   doi:10.1093/bioinformatics/btr330

1138 99. Henn BM, Botigué LR, Peischl S, Dupanloup I, Lipatov M, Maples BK, et al.
1139   Distance from sub-Saharan Africa predicts mutational load in diverse human
1140   genomes. Proc Natl Acad Sci. 2016;113: E440-449.
1141   doi:10.1073/pnas.1510805112

1142 100. Kidd JM, Sharpton TJ, Bobo D, Norman PJ, Martin AR, Carpenter ML, et al.
1143   Exome capture from saliva produces high quality genomic and metagenomic
1144   data. BMC Genomics. 2014;15: 262. doi:10.1186/1471-2164-15-262

1145 101. Kim HL, Ratan A, Perry GH, Montenegro A, Miller W, Schuster SC. Khoisan
1146   hunter-gatherers have been the largest population throughout most of modern-
1147   human demographic history. Nat Commun. 2014;5: 5692.
1148   doi:10.1038/ncomms6692

1149 102. Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. Detection of nonneutral
1150   substitution rates on mammalian phylogenies. Genome Res. 2010;20: 110–121.
1151   doi:10.1101/gr.097857.109

1152 103. Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, et al.
1153   Evolutionarily conserved elements in vertebrate, insect, worm, and yeast
1154   genomes. Genome Res. 2005;15: 1034–1050. doi:10.1101/gr.3715005

1155 104. Bernstein BE, Birney E, Dunham I, Green ED, Gunter C, Snyder M, et al. An
1156   integrated encyclopedia of DNA elements in the human genome. Nature.
1157   2012;489: 57–74. doi:10.1038/nature11247

1158 105. Bailey JA, Yavor AM, Massa HF, Trask BJ, Eichler EE. Segmental duplications:
1159   Organization and impact within the current human genome project assembly.
1160   Genome Res. 2001;11: 1005–1017. doi:10.1101/gr.187101

1161 106. Szpiech ZA, Hernandez RD. selscan: An efficient multithreaded program to
1162   perform EHH-based scans for positive selection. Mol Biol Evol. 2014;31: 2824–
1163   2827. doi:10.1093/molbev/msu211

1164 107. Capra JA, Hubisz MJ, Kostka D, Pollard KS, Siepel A. A model-based analysis of
1165   GC-biased gene conversion in the human and chimpanzee genomes. PLoS
1166   Genet. 2013;9: e1003684. doi:10.1371/journal.pgen.1003684

1167 108. Frazer KA, Ballinger DG, Cox DR, Hinds DA, Stuve LL, Gibbs RA, et al. A second
1168   generation human haplotype map of over 3.1 million SNPs. Nature. 2007;449:
1169   851–861. doi:10.1038/nature06258

1170 109. Pratto F, Brick K, Khil P, Smagulova F, Petukhova G V, Camerini-Otero RD.
1171   Recombination initiation maps of individual human genomes. Science. 2014;346:
1172   1256442. doi:10.1126/science.1256442

1173 110. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic
1174   variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38:
1175   e164. doi:10.1093/nar/gkq603

1176 111. Hernandez RD, Williamson SH, Zhu L, Bustamante CD. Context-dependent

1177 mutation rates may cause spurious signatures of a fixation bias favoring higher
1178 GC-content in humans. Mol Biol Evol. 2007;24: 2196–2202.
1179 doi:10.1093/molbev/msm149
1180 112. Palamara PF, Francioli LC, Wilton PR, Genovese G, Gusev A, Finucane HK, et
1181 al. Leveraging distant relatedness to quantify human mutation and gene-
1182 conversion rates. Am J Hum Genet. 2015;97: 775–789.
1183 doi:10.1016/j.ajhg.2015.10.006
1184 113. Hernandez RD. A flexible forward simulator for populations subject to selection
1185 and demography. Bioinformatics. 2008;24: 2786–2787.
1186 doi:10.1093/bioinformatics/btn522
1187 114. Torgerson DG, Boyko AR, Hernandez RD, Indap A, Hu X, White TJ, et al.
1188 Evolutionary processes acting on candidate cis-regulatory regions in humans
1189 inferred from patterns of polymorphism and divergence. PLoS Genet. 2009;5:
1190 e1000592. doi:10.1371/journal.pgen.1000592
1191 115. Hudson RR, Slatkin M, Maddison WP. Estimation of levels of gene flow from DNA
1192 sequence data. Genetics. 1992;132: 583–589.
1193
1194

Supporting information

1195

S1 Appendix. Soft sweep detection and implementation in selscan v1.2.0.

1196

S1 Supporting information.

1197

S1 Fig. Inference models inferred from TGP CG high B neutral regions and coding four-fold degenerate sites. Solid lines are the inference results from running dadi on 53 YRI (African), 64 CEU (European), and 62 CHS (East Asian) TGP CG samples (projected down to 106 chromosomes during inference procedure) across neutral regions in the highest 1% B bin ($B \geq 0.994$). Broken lines represent the inference results using the same CG samples, but with sequence data only from coding four-fold degenerate synonymous sites.

S2 Fig. Diversity for TGP non-admixed populations while controlling for GC-biased gene conversion and recombination hotspots. (A) Normalized diversity (π/divergence) measured across the lowest 1% B quantile bin (strong BGS). (B) Normalized diversity measured across the highest 1% B quantile bin (weak BGS). (C) Rela-

1209 tive diversity: the ratio of normalized diversity for the lowest 1% B bin to normalized di-

1210 versity for the highest 1% B bin (π/π_0). Error bars represent ±1 SEM calculated from

1211 1,000 bootstrapped datasets.

1212 **S3 Fig. Diversity for TGP non-admixed populations without normalizing by diver-**

1213 **gence with Rhesus macaque.** (A) Diversity (π) measured across the lowest 1% B

1214 quantile bin (strong BGS). (B) Diversity measured across the highest 1% B quantile bin

1215 (weak BGS). (C) Relative diversity: the ratio of diversity for the lowest 1% B bin to diver-

1216 sity for the highest 1% B bin (π/π_0). Error bars represent ±1 SEM calculated from 1,000

1217 bootstrapped datasets.

1218 **S4 Fig. Diversity for TGP continental groups while controlling for GC-biased gene**

1219 **conversion and recombination hotspots.** (A) Normalized diversity (π/divergence)

1220 measured across the lowest 1%, 5%, 10% and 25% B quantile bins (strong BGS) and

1221 the highest 1% B quantile bin (weak BGS). (B) Relative diversity (π/π_0) for the lowest

1222 1%, 5%, 10%, and 25% B bins. Error bars represent ±1 SEM calculated from 1,000

1223 bootstrapped datasets.

1224 **S5 Fig. Diversity for TGP continental groups without normalizing by divergence**

1225 **with Rhesus macaque.** (A) Diversity (π) measured across the lowest 1%, 5%, 10% and

1226 25% B quantile bins (strong BGS) and the highest 1% B quantile bin (weak BGS). (B)

1227 Relative diversity (π/π_0) for the lowest 1%, 5%, 10%, and 25% B bins. Error bars repre-

1228 sent ±1 SEM calculated from 1,000 bootstrapped datasets.

1229 **S6 Fig. F_{ST} measured across joint bins of B and recombination rate for different**

1230 **TGP continental groups**. The left panels of S6 Figs A-E show F_{ST} measured as a func-

1231 tion of 25 4% recombination rate quantile bins conditional on three 2% B quantile bins

1232   (note log scale of x-axis for recombination rate). The right panels of S6 Figs A-E show

1233 F_{ST} measured as a function of 25 4% B quantile bins conditional on three 2% recombi-

1234   nation rate quantile bins. The following continental group comparisons are shown for

1235   each plot: (A) African vs. European, (B) African vs. East Asian, (C) European vs. South

1236   Asian, (D) European vs. East Asian, (E) South Asian vs. East Asian. Smaller transpar-

1237   ent points and lines show the F_{ST} estimates and corresponding lines of best fit (using

1238   linear regression) for each of the pairwise population comparisons within a particular

1239   pair of continental groups (25 comparisons total). Larger opaque points are mean F_{ST}

1240   estimates across all pairwise comparisons within a particular pair of continental groups

1241   (with bold lines showing their corresponding lines of best fit).

1242   **S7 Fig. Simulations of diversity and relative diversity under BGS using a human**

1243   **demographic model without migration.** (A) Inferred demographic model from Com-

1244   plete Genomics TGP data. The demographic model used for the simulations in S7 Fig

1245   are identical to those used for Fig 5, except that migration parameters between all popu-

1246   lations are set to 0. (B) Simulated diversity at neutral sites across populations as a func-

1247   tion of time under our inferred demographic model without BGS (π_0 - dashed colored

1248   lines) and with BGS (π - solid colored lines). (C) Relative diversity (π/π_0) measured by

1249   taking the ratio of diversity with BGS (π) to diversity without BGS (π_0) at each time point.

1250   Note that the x-axes in all three figures are on the same scale. Time is scaled using a

1251   human generation time of 25 years per generation. Simulation data was sampled every

1252   100 generations.

1253   **S8 Fig. Simulations of diversity and relative diversity under BGS using various**

1254   **fractions of sites experiencing deleterious mutation.** Values for the deleterious site

1255   fraction are provided in the header for each set of plots. Left column plots show results

1256   of simulations under a demographic model with migration between all human popula-

1257   tions. Right column plots show results of simulations under a demographic model with

1258   no migration. Colored lines represent different populations though time and are identical

1259   to those in Fig 5 and S7 Fig The demographic model used is also identical to that in Fig

1260   5 (for simulations with migration) and S7 Fig (for simulations without migration). Simula-

1261   tion data was sampled every 100 generations.

1262   **S9 Fig. F_{ST} is not correlated with recombination rate**.

1263 F_{ST} measured across 2% recombination rate quantile bins. The right panel of S9 Fig

1264   displays a narrower range of recombination rates to show detail. Smaller transparent

1265   points and lines show the estimates and corresponding lines of best fit (using linear re-

1266   gression) for F_{ST} between every pairwise population comparison for a particular pair of

1267   continental groups (25 pairwise comparisons each). Larger opaque points and lines are

1268   mean F_{ST} estimates and lines of best fit across all Thousand Genomes Project (TGP)

1269   population comparisons between a particular pair of continental groups. Error bars rep-

1270   resent ±1 SEM calculated from 1,000 bootstrapped datasets.

1271   **S10 Fig. F_{ST} between African (AFR) and South Asian (SASN) populations jointly**

1272   **across *B* and recombination rate**.

1273   (A) F_{ST} as a function of 25 recombination rate bins (4% quantile bins) conditional on

1274   three different 2% *B* quantile bins (note log scale of x-axis for recombination rate). (B)

1275 F_{ST} as a function of 25 *B* bins (4% quantile bins) conditional on three different 2% re-

1276   combination rate quantile bins. Smaller transparent points and lines show the F_{ST} esti-

1277   mates and corresponding lines of best fit (using linear regression) for each of the pair-

49

1278 wise comparisons of AFR vs. SASN Thousand Genomes Project (TGP) populations (25

1279 comparisons total). Larger opaque points are mean F_{ST} estimates across all pairwise

1280 comparisons of AFR vs. SASN TGP populations (with bold lines showing their corre-

1281 sponding lines of best fit).

1282 **S11 Fig. Comparing patterns of diversity between local ancestry segments of ad-**

1283 **mixed samples and continental groups**.

1284 (A) Normalized diversity (heterozygosity/divergence) and (B) Relative diversity: the ratio

1285 of normalized diversity in the lowest B quantile bins (strong BGS) in (A) to normalized

1286 diversity in the highest 1% B quantile bin (weak BGS). Local ancestry segments include

1287 African, European, and Native American ancestries. Continental groups include African,

1288 European, and East Asian populations. Error bars represent ±1 SEM calculated from

1289 1,000 bootstrapped datasets.

1290 **S12 Fig. Simulations of singleton density and relative singleton density**

1291 A) Results of simulations under a demographic model with migration between all human

1292 populations. B) Results of simulations under a demographic model with no migration.

1293 The second row of A) and B) shows measurements of singleton density (i.e., number of

1294 singletons observed per site) from simulations without BGS (ψ_0 - dashed colored lines)

1295 and with BGS (ψ - solid colored lines). The bottom row of A) and B) shows correspond-

1296 ing relative singleton density (ψ/ψ_0) measured by taking the ratio of singleton density

1297 with BGS (ψ) to singleton density without BGS (ψ_0) at each sampled generation time

1298 point. The simulation data used for these measurements is identical to that of Fig 5 (for

1299 simulations with migration) and S7 Fig (for simulations without migration).
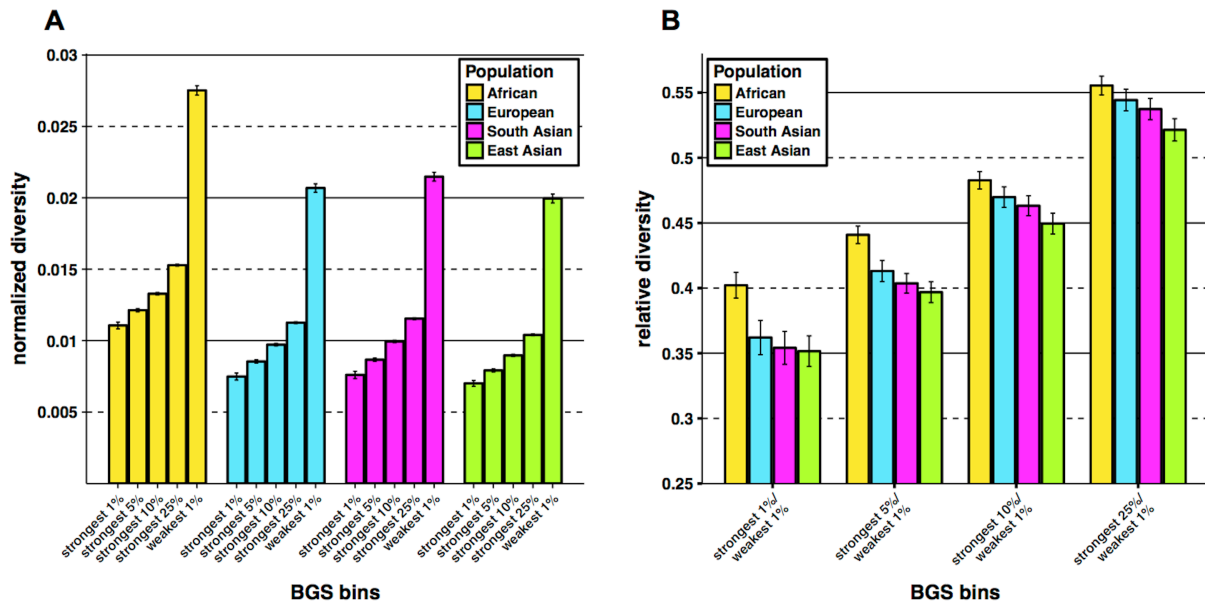
1300

50

1301
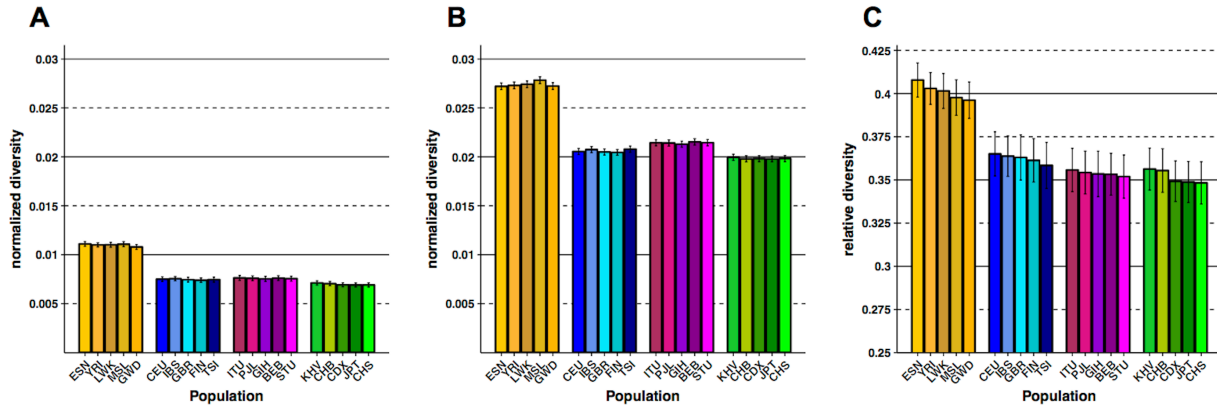
Fig 1. Normalized diversity and relative diversity for non-admixed populations of the Thousand Genomes Project (TGP).

(A) Normalized diversity (π/divergence) measured across the lowest 1% B quantile bin (strong BGS). (B) Normalized diversity measured across the highest 1% B quantile bin (weak BGS). (C) Relative diversity: the ratio of normalized diversity in the lowest 1% B bin to normalized diversity in the highest 1% B bin (π/π_0). TGP population labels are indicated below each bar (see S12 Table in Supporting information for population label descriptions), with African populations colored by gold shades, European populations colored by blue shades, South Asian populations colored by violet shades, and East Asian populations colored by green shades. Error bars represent ±1 SEM calculated from 1,000 bootstrapped datasets.

Fig 2. Normalized and relative diversity for Thousand Genomes Project (TGP) continental groups.

(A) Normalized diversity (π/divergence) measured across the lowest 1%, 5%, 10% and 25% *B* quantile bins (strong BGS) and the highest 1% *B* quantile bin (weak BGS). (B) Relative diversity: the ratio of normalized diversity in the lowest *B* quantile bins (strong BGS) in (A) to normalized diversity in the highest 1% *B* quantile bin (weak BGS). Error bars represent ±1 SEM calculated from 1,000 bootstrapped datasets.

1324
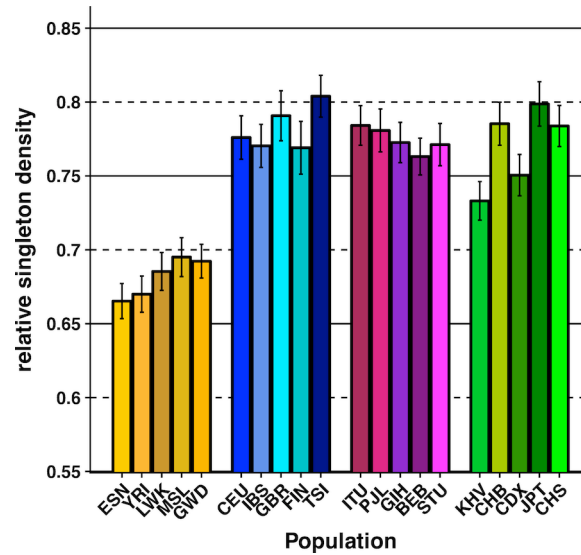
**Fig 3. Relative singleton density for non-admixed populations of the Thousand Genomes Project (TGP).**

Relative singleton density measured by taking the ratio of singleton density in the lowest 1% *B* quantile bin to singleton density in the highest 1% *B* quantile bin (ψ/ψ_0). Singleton density was normalized by divergence with Rhesus macaque. TGP population labels are indicated below each bar (see S12 Table in Supporting information for population label descriptions), with African populations colored by gold shades, European populations colored by blue shades, South Asian populations colored by violet shades, and East Asian populations colored by green shades. Error bars represent ±1 SEM calculated from 1,000 bootstrapped datasets.

1335
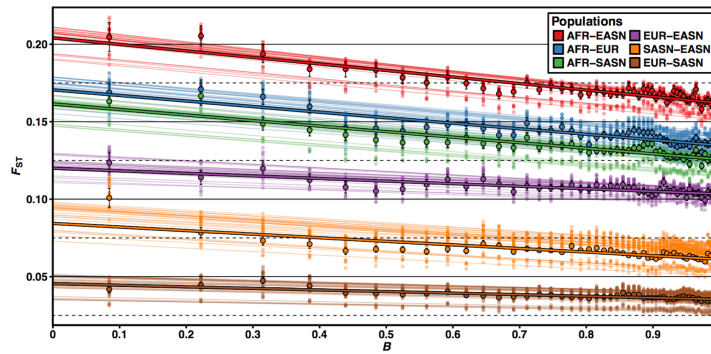
1336

Fig 4. F_{ST} is correlated with B

F_{ST} measured across 2% quantile bins of B. Smaller transparent points and lines show the estimates and corresponding lines of best fit (using linear regression) for F_{ST} between every pairwise population comparison for a particular pair of continental groups (25 pairwise comparisons each). Larger opaque points and lines are mean F_{ST} estimates and lines of best fit across all Thousand Genomes Project (TGP) population comparisons between a particular pair of continental groups. Error bars represent ±1 SEM calculated from 1,000 bootstrapped datasets.

1337

1338

1339

1340

1341

1342

1343

1344

1345
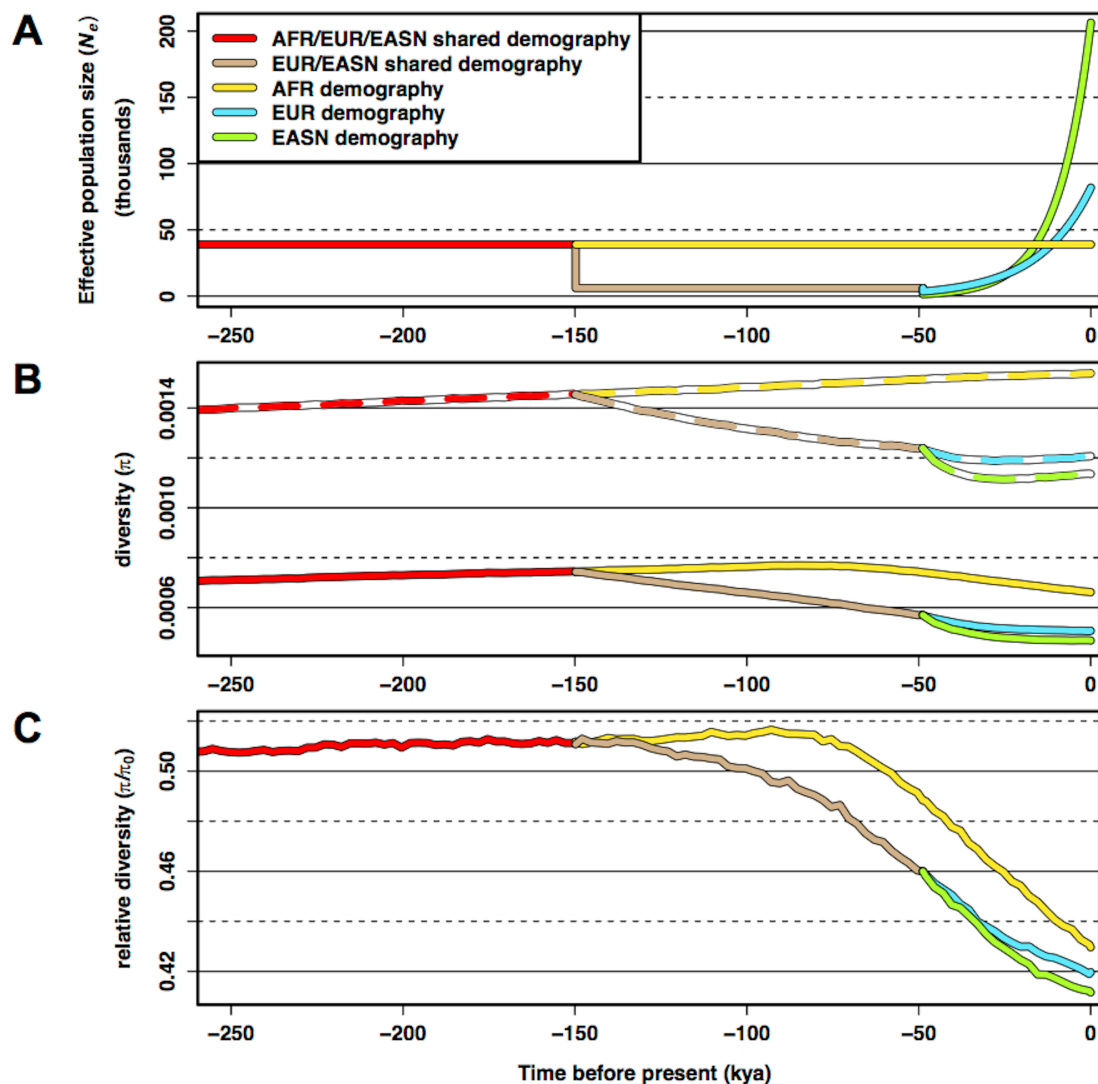
1346

Fig 5. Simulations confirm that demographic events shape the impact of background selection (BGS).

(A) Inferred demographic model from Complete Genomics Thousand Genomes Project (TGP) data showing population size changes for Africans (AFR), Europeans (EUR), and East Asians (EASN) as a function of time that was used for the simulations of BGS. (B) Simulated diversity at neutral sites across populations as a function of time under our inferred demographic model without BGS (π_0 - dashed colored lines) and with BGS (π -

1354   solid colored lines). (C) Relative diversity (π/π_0) measured by taking the ratio of diversity

1355   with BGS (π) to diversity without BGS (π_0) at each time point. Note that the x-axes in all

1356   three figures are on the same scale. Time is scaled using a human generation time of

1357   25 years per generation. Simulation data was sampled every 100 generations.

1358