Robust and window-insensitive mixture model approaches for localizing balancing selection

Xiaoheng Cheng[1,2] and Michael DeGiorgio*[2,3,4]

[1]Huck Institutes of Life Sciences, Pennsylvania State University, University Park, PA, USA
[2]Department of Biology, Pennsylvania State University, University Park, PA, USA
[3]Department of Statistics, Pennsylvania State University, University Park, PA, USA
[4]Institute for CyberScience, Pennsylvania State University, University Park, PA, USA

Abstract

Long-term balancing selection typically leaves narrow footprints of increased genetic diversity, and therefore most detection approaches only achieve optimal performances when sufficiently small genomic regions (*i.e.*, windows) are examined. Such methods are sensitive to window sizes and suffer substantial losses in power when windows are large. This issue creates a tradeoff between noise and power in empirical applications. Here, we employ mixture models to construct a set of five composite likelihood ratio test statistics, which we collectively term B statistics. These statistics are agnostic to window sizes and can operate on diverse forms of input data. Through simulations, we show that they exhibit comparable power to the best-performing current methods, and retain substantially high power regardless of window sizes. Moreover, these methods display considerable robustness to high mutation rates and uneven recombination landscapes, as well as an array of other common confounding scenarios. Further, we applied these statistics on a bonobo population-genomic dataset. In addition to the *MHC-DQ* genes, we uncovered several novel candidate genes, such as *KLRD1*, involved in viral defense, and *NKAIN3*, associated with sexuality. Finally, we integrated the set of statistics into open-source software named `BalLeRMix`, for future applications by the scientific community.

Introduction

Balancing selection maintains polymorphism at selected genetic loci, and can operate through a variety of mechanisms (Charlesworth, 2006). In addition to overdominance (Charlesworth and Charlesworth, 2010), other processes such as sexual selection (Cho et al., 2006), periodical environmental shifts (Bergland et al., 2014), pleiotropy (Andrés, 2001; Mitchell-Olds et al., 2007), meiotic drive (Ubeda and Haig, 2004; Charlesworth and Charlesworth, 2010), and negative frequency-dependent selection (Charlesworth and Charlesworth, 2010) can also maintain diversity at underlying loci. Due to the increasing availability of population level genomic data, in which allele frequencies and genomic density of polymorphisms can be assessed in detail, there is an expanding interest in studying balancing selection and detecting its genomic footprints (*e.g.*, Andrés et al., 2009; Leffler et al., 2013; DeGiorgio et al., 2014; Gao et al., 2015; Hunter-Zinck and Clark, 2015; Sheehan and Song, 2016; Lonn et al., 2017; Sweeney et al., 2017; Guirao-Rico et al., 2017; Siewert and Voight, 2017, 2018; Bitarello et al., 2018; Ye et al., 2018; Cheng and DeGiorgio, 2019). However, despite multiple efforts to design statistics for identifying balanced loci (*e.g.*, DeGiorgio et al., 2014; Siewert and Voight, 2017, 2018; Bitarello et al., 2018; Cheng and DeGiorgio, 2019), performances of existing methods still leave room for improvement.

*mxd60@psu.edu

Early methods applied to this problem evaluated departures from neutral expectations of genetic diversity at a particular genomic region. For example, the Hudson-Kreitman-Aguadé (HKA) test (Hudson et al., 1987) uses a chi-square statistic to assess whether genomic regions have higher density of polymorphic sites when compared to a putative neutral genomic background. In contrast, Tajima's D (Tajima, 1989) measures the distortion of allele frequencies from the neutral site frequency spectrum (SFS) under a model with constant population size. However, these early approaches were not tailored for balancing selection, and have limited power. Recently, novel and more powerful summary statistics (Siewert and Voight, 2017, 2018; Bitarello et al., 2018) and model-based approaches (DeGiorgio et al., 2014; Cheng and DeGiorgio, 2019) have been developed to specifically target regions under balancing selection. In general, the summary statistics capture deviations of allele frequencies from a putative equilibrium frequency of a balanced polymorphism. In particular, the non-central deviation statistic (Bitarello et al., 2018) adopts an assigned value as this putative equilibrium frequency, whereas the β and $\beta^{(2)}$ statistics of Siewert and Voight (2017, 2018) use the frequency of the central polymorphic site instead. On the other hand, the T statistics of DeGiorgio et al. (2014) and Cheng and DeGiorgio (2019) compare the composite likelihood of the data under an explicit coalescent model of long-term balancing selection (Hudson et al., 1987; Hudson and Kaplan, 1988) to the composite likelihood under the genome-wide distribution of variation, which is taken as neutral.

Nevertheless, all extant approaches are limited by their sensitivity to the size of the region that the statistics are computed on (hereafter referred to as the "window size"). Because the footprints of long-term balancing selection are typically narrow (Hudson and Kaplan, 1988; Charlesworth, 2006), small windows with size comparable to that of the theoretical footprint are commonly used in practice to optimize their power, especially for summary statistics. However, such small fixed window sizes lead to increased noise in the estimation of each statistic, and the uneven recombination landscape (Smukowski and Noor, 2011) across chromosomes ensures that a fixed window size is not suitable for all genomic regions. On the other hand, when an overly large genomic region is considered as the window, true signals will likely be overwhelmed by the surrounding neutral regions, diminishing method power as shown by Cheng and DeGiorgio (2019). Available model-based approaches (DeGiorgio et al., 2014; Cheng and DeGiorgio, 2019) could have been made robust to large window sizes if they instead adopted the SFS expected under a neutrally-evolving population of constant size as the null hypothesis, because their model of balancing selection for the alternative hypothesis converges to this constant-size neutral model for large recombination rates. However, this neutral model does not account for demographic factors that can impact the genome-wide distribution of allele frequencies, such as population size changes. To guard against such demographic influences, the model-based T_1 and T_2 statistics (DeGiorgio et al., 2014; Cheng and DeGiorgio, 2019) employ the genome-wide SFS instead, compromising the robustness against large windows. Moreover, Cheng and DeGiorgio (2019) showed that although the power of the T_2 statistic decays much slower than other approaches as window size increase, the loss of power is still substantial.

In this article, we describe a set of composite likelihood ratio test statistics that are based on a mixture model (Figures 1A and B), which allows for the robust and flexible detection of long-term balancing selection regardless of window size. Dependent on the types of data available, we propose a set of five likelihood ratio test statistics termed B_2, $B_{2,\mathrm{MAF}}$, B_1, B_0, and $B_{0,\mathrm{MAF}}$, which respectively accommodate data with substitutions and derived (B_2) or minor ($B_{2,\mathrm{MAF}}$) allele frequency polymorphisms, with substitutions

and polymorphisms with unknown allele frequency (B_1), and with derived (B_0) or minor ($B_{0,\mathrm{MAF}}$) allele frequency polymorphisms only. We comprehensively evaluated their performances under diverse scenarios, including their powers for balancing selection with varying ages, distinct strengths and equilibrium frequencies, robustness against window sizes, and robustness against confounding factors such as demographic history, recombination rate variation, and mutation rate variation. We also compared and discussed their performances with other leading approaches—namely HKA, β, β^*, $\beta^{(2)}$, NCD, T_1, and T_2. We further applied B_2 on bonobo genomic data (Prado-Martinez et al., 2013) to explore long-term balancing selection in the other close relative of humans. Lastly, we developed the software `BalLeRMix` (BALancing selection LikElihood Ratio MIXture models) to implement these novel tests for the convenience of the scientific community.

Theory

In this section, we describe a novel set of likelihood ratio test statistics (B_2, $B_{2,\mathrm{MAF}}$, B_1, B_0, and $B_{0,\mathrm{MAF}}$) to scan genomes for balancing selection. These five new methods have been implemented in the software package `BalLeRMix`, which is written in Python and is available at `http://www.personal.psu.edu/mxd60/ballermix.html`.

Probability distributions given derived allele polymorphisms and substitutions

For n sampled alleles at an informative site (*i.e.*, polymorphism or substitution), denote the number of derived alleles as k, $k = 1, 2, \ldots, n$. Let $\xi_n(k)$ be the total number of informative sites across the whole-genome with k derived alleles observed out of n sampled alleles. The probability of observing such a site is therefore

$$g_n^{(2)}(k) = \frac{\xi_n(k)}{\sum_{j=1}^{n} \xi_n(j)}.$$

When balancing selection maintains an equilibrium frequency of x on the site under selection, the outcomes of observing derived alleles on this site (out of n lineages) can be approximated by a binomial distribution of n trials with a success probability of x. Following this binomial model, the probability of observing the selected site with k observed derived alleles is

$$h_{n,x}^{(2)}(k) = \frac{\mathrm{Bin}(k; n, x)}{\sum_{j=1}^{n} \mathrm{Bin}(j; n, x)}.$$

For an informative site d recombination units away from the presumed site under selection, it can either be linked to the derived (with equilibrium frequency x) or ancestral (with equilibrium frequency $1 - x$) haplotype under balancing selection, resulting in a bimodal distribution (Figure 1C). Therefore, the probability of observing k derived alleles out of n sampled alleles is

$$f_{n,x,A}^{(2)}(k, d) = \alpha_A(d)\left[\frac{1}{2}h_{n,x}^{(2)}(k) + \frac{1}{2}h_{n,1-x}^{(2)}(k)\right] + \left[1 - \alpha_A(d)\right]g_n^{(2)}(k),$$

where $\alpha_A(d) = e^{-Ad}$ and where A is a model parameter that determines the size of the genomic footprint of balancing selection. When allele frequency information is unavailable at polymorphic sites, the probability

of observing a polymorphic site ($k \neq n$) or substitution ($k = n$) would be

$$f^{(1)}_{n,x,A}(k,d) = f^{(2)}_{n,x,A}(n,d)\mathbf{1}_{\{k=n\}} + \left[1 - f^{(2)}_{n,x,A}(n,d)\right]\mathbf{1}_{\{k\neq n\}}.$$

Similarly, when substitutions are not considered or are missing in the data (*i.e.*, only observe derived allele counts $k = 1, 2, \ldots, n-1$), the two mixing components can be normalized as

$$g^{(0)}_n(k) = \frac{\xi_n(k)}{\sum_{j=1}^{n-1} \xi_n(j)},$$

and

$$h^{(0)}_{n,x}(k) = \frac{\mathrm{Bin}(k; n, x)}{\sum_{j=1}^{n-1} \mathrm{Bin}(j; n, x)}.$$

The probability of observing a polymorphic site with k derived alleles out of n sampled alleles is then

$$f^{(0)}_{n,x,A}(k,d) = \alpha_A(d)\left[\frac{1}{2}h^{(0)}_{n,x}(k) + \frac{1}{2}h^{(0)}_{n,1-x}(k)\right] + \left[1 - \alpha_A(d)\right]g^{(0)}_n(k).$$

Probability distributions given minor allele polymorphisms and substitutions

When alleles cannot be confidently polarized, minor allele frequencies are often used instead. For informative sites with n sampled alleles, denote the minor allele count as k, $k = 0, 1, \ldots, \lfloor n/2 \rfloor$, and the total number of such sites in the genome as $\eta_n(k)$. Substitutions are assigned to $\eta_n(0)$, as the minor allele count is zero. The probability of observing a site with k minor alleles out of n sampled alleles in the genome is

$$g^{(2,\mathrm{MAF})}_n(k) = \frac{\eta_n(k)}{\sum_{j=0}^{\lfloor n/2 \rfloor} \eta_n(j)}.$$

Assume the equilibrium minor allele frequency at the locus undergoing long-term balancing selection is $x \in (0, 0.5]$. The probability of observing k minor alleles out of n sampled alleles is then

$$h^{(2,\mathrm{MAF})}_{n,x}(k) = \frac{\mathrm{Bin}(k; n, x) + \mathrm{Bin}(n-k; n, x)\mathbf{1}_{\{k\neq n/2\}}}{\sum_{j=1}^{n} \mathrm{Bin}(j; n, x)}.$$

Hence, for an informative site d recombination units away from the presumed site under selection, the probability of observing k minor alleles out of n sampled alleles is

$$f^{(2,\mathrm{MAF})}_{n,x,A}(k,d) = \alpha_A(d)h^{(2,\mathrm{MAF})}_{n,x}(k) + [1 - \alpha_A(d)]\ g^{(2,\mathrm{MAF})}_n(k),$$

Similarly, when substitutions are not considered or are missing in the data (*i.e.*, only observed minor alleles counts $k = 1, 2, \ldots, \lfloor n/2 \rfloor$), the two mixing components can be normalized as

$$g^{(0,\mathrm{MAF})}_n(k) = \frac{\eta_n(k)}{\sum_{j=1}^{\lfloor n/2 \rfloor} \eta_n(j)}$$

and

$$h^{(0,\mathrm{MAF})}_{n,x}(k) = \frac{\mathrm{Bin}(k; n, x) + \mathrm{Bin}(n-k; n, x)\mathbf{1}_{\{k\neq n/2\}}}{\sum_{j=1}^{n-1} \mathrm{Bin}(j; n, x)}.$$

The probability of observing a polymorphic site with k minor alleles out of n sampled alleles is then

$$f_{n,x,A}^{(0,\mathrm{MAF})}(k,d) = \alpha_A(d) h_{n,x}^{(0,\mathrm{MAF})}(k) + [1-\alpha_A(d)]\; g_n^{(0,\mathrm{MAF})}(k).$$

Composite likelihood ratio tests based on the mixture models

Based on the probability distributions described for the five models, for each model $X \in \{\text{“2”}, \text{“2,MAF”}, \text{“1”}, \text{“0”}, \text{“0,MAF”}\}$, the composite likelihood of a genomic region with L informative sites under the null hypothesis of neutrality is

$$\mathcal{L}_0^{(X)}(\mathbf{n},\mathbf{k}) = \prod_{i=1}^{L} g_{n_i}^{(X)}(k_i),$$

where $\mathbf{n} = [n_1, n_2, \ldots, n_L]$ and $\mathbf{k} = [k_1, k_2, \ldots, k_L]$ are the vectors of sample sizes and derived or minor allele counts, respectively, at the L informative sites in the genomic region. Similarly, the composite likelihood under the alternative hypothesis of model X would be

$$\mathcal{L}_a^{(X)}(x, A\,;\, \mathbf{n},\mathbf{k},\mathbf{d}) = \prod_{i=1}^{L} f_{n_i,x,A}^{(X)}(k_i, d_i),$$

where $\mathbf{d} = [d_1, d_2, \ldots, d_L]$ is the vector of recombination distances between the test site and each of the L informative sites. This likelihood is maximized at

$$(\widehat{x}, \widehat{A}) = \underset{(x,A)}{\arg\max}\, \mathcal{L}_a^{(X)}(x, A\,;\, \mathbf{n},\mathbf{k},\mathbf{d}).$$

Hence, under model $X \in \{\text{“2”}, \text{“2,MAF”}, \text{“1”}, \text{“0”}, \text{“0,MAF”}\}$, the log composite likelihood ratio test statistic for the test site is

$$B_X = 2\Big[\ln \mathcal{L}_a^{(X)}(\widehat{x}, \widehat{A}\,;\, \mathbf{n},\mathbf{k},\mathbf{d}) - \ln \mathcal{L}_0^{(X)}(\mathbf{n},\mathbf{k})\Big].$$

Interpretation of estimated A and x parameters

The likelihood for the alternative model is maximized over the parameters A and x, where x represents the presumed equilibrium minor allele frequency, and A decides the influence of balancing selection on neutral sites of varying distance away from the test site. After optimizing over this parameter space, the parameter values under the optimal likelihood, $\widehat{A}$ and $\widehat{x}$, provide information on the nature of detected genomic footprints. The value of $\widehat{x}$ should reflect the enriched minor allele frequency across the region, whereas $\widehat{A}$ is informative about the impact of the balancing selection. Intuitively, A describes how far balancing selection has impacted the genomic variation at nearby neutral sites, and the smaller the $\widehat{A}$, the wider the footprint would be, and likely the younger the balanced polymorphism. Nonetheless, note that the increase in A values is not only decided by the age of balancing selection, as a large A also reduces the number of informative sites that yield meaningful likelihood ratios, and can thus also occur when data in the examined area fit the alternative model poorly. Therefore, we advise only comparing the $\widehat{A}$ values among regions with reasonably high composite likelihood ratios, and that caution be used when making inferences from these values as they do not map to an explicit evolutionary model.

Results

Performances on simulated data

We simulated 50 kilobase (kb) long sequences using SLiM3 (Haller and Messer, 2019), under the three-species demographic model (Figure S1) inspired by the demographic history of great apes (see *Methods*), and extensively evaluated the performances of all five B statistic variants. We also compared the B statistics to the summary statistics β, β^*, HKA, NCD2, and $\beta^{(2)}$, which are respectively analogues to B_0, $B_{0,\mathrm{MAF}}$, B_1, $B_{2,\mathrm{MAF}}$, and B_2, and to the likelihood statistics T_1 and T_2, which are respectively analogues to B_1 and B_2.

Robust high power under varying window sizes

We first examined the robustness of the B statistics to overly large window sizes, under a scenario of strong heterozygote advantage (selective coefficient $s = 0.01$ with dominance coefficient $h = 20$) acting on a mutation that arose 7.5×10^4 generations prior to sampling. Because `BetaScan` (Siewert and Voight, 2017, 2018) (which implements the standardized and nonstandardized β, β^*, and $\beta^{(2)}$ statistics, among which we only consider the standardized) operates on windows of fixed physical length, we adopted window sizes of 1, 1.5, 2.5, 3, 5, 10, 15, 20, and 25 kb for all summary statistics and B statistics. The T statistics were applied on windows with matching expected numbers of informative sites. *Supplementary Note* 1 details the calculation for matching the number of informative sites to physical length of a genomic region.

To reduce potential stochastic fluctuations in the number of true positives when the false positive rate is controlled at a low level, we examined the area under a partial curve with no greater than a 5% false positive rate (hereafter referred to as "partial AUC"). As shown in Figure 2, under optimal window sizes for most other statistics, all variants of B statistics display substantial partial AUCs comparable to that of the respective T statistic variant, which has outperformed other equivalent summary statistics in most previous simulation studies (DeGiorgio et al., 2014; Siewert and Voight, 2017, 2018; Bitarello et al., 2018; Cheng and DeGiorgio, 2019). Most remarkably, as the window size increases, while all other statistics exhibit drastic decays in power, the powers of all variants of the B statistic only show minor decreases. In fact, when comparing the powers under 25 kb windows against those under optimal window sizes for each statistic, the powers of all statistics drop more than twice as much as B_1 and B_2 (Figure 2B). In comparison with each method's optimal performance, most statistics (except all B statistics and T_2, the model-based analog of B_2) lose more than 80% of their optimal power under the largest window size examined (Figure 2C). Although T_2 still retains considerably higher partial AUC compared to all other extant methods, it still decreases to a value substantially lower than that of B_2. Such robustness of B statistics to large windows is reasonable and expected, because the probability distribution of allele frequencies at sites far enough from the test site will match the genome-wide SFS, thereby contributing little to the overall likelihood ratio.

Among all statistics evaluated, we found that those considering polymorphism data only (*i.e.*, B_0 variants and β variants) demonstrated relatively poor robustness to increases in window size. This result indicates that the detectable footprint of balancing selection in polymorphism data by itself may decay faster than other types of information, and that incorporating substitution data may help improve robustness to large windows.

Considering that the powers of all B statistics stabilize at a fixed level as the window size increases

(Figure 2), we permit the B statistics to employ all informative sites on a chromosome. However, to reduce computational load, we only consider sites with mixing proportion $\alpha_A(d) \geq 10^{-8}$ for each value of A considered during optimization, which does not create discernible differences in performance from when all data are considered (Figure S2). However, to ensure that other methods still display considerable power for their comparisons, we applied the summary statistics with their optimal window sizes of one kb, and T statistics with numbers of informative sites expected in a one kb window (see *Methods*), unless otherwise stated.

High power for detecting balancing selection of varying age and selective strength

Next, we explored the powers of B statistics when the selective strength s, equilibrium frequency (controlled by the dominance parameter h), and the age of balancing selection vary. Specifically, we examined scenarios where the selective coefficients were moderate ($s = 0.01$, Figures 3A, C, D, and E) or weak ($s = 10^{-3}$, Figure 3B), and when the equilibrium frequency of the minor allele is 0.5 ($h = 20$, Figures 3A and B), 0.4 ($h = 3$, Figure 3C), 0.3 ($h = 1.75$, Figure 3D), or 0.2 ($h = 1.33$, Figure 3E). Across all scenarios considered, T_2 and β^* show the highest power for old balancing selection. The best-performing B variants, B_2 and $B_{2,\mathrm{MAF}}$, display high power as well, and are often comparable to that of the $\beta^{(2)}$ statistic. The power of B_1 is also similar to HKA, which is its summary statistic analogue. Furthermore, we noticed that B statistics exhibit superior power for younger balanced alleles, particularly when balancing selection is more recent than 2×10^5 generations, and when the equilibrium frequency does not equal to 0.5 (Figure S3). For older selected polymorphisms, although several statistics outperform B statistics, it is important to point out that all previous methods were provided optimal window sizes, whereas B statistics were set to use all sites with considerable $\alpha_A(d)$, under which they show lower power than when window sizes are optimized (Figure 2). This difference in performance between previous methods applied with their optimal window sizes and B statistics can also explain the seemingly inferior performance of the two B_0 variants when compared with the analogous β statistics, as the B_0 variants lose more power than other B variants when computed on extended windows. When applied with the same window size, however, B_0 outperforms β by a large margin (Figure 2). Nevertheless, these results give us confidence that B statistics have generally high power to detect young and old balancing selection, even when adopting large windows.

Robustness to recombination rate variation and elevated mutation rates

Despite their robustness to large windows and high power for detecting balancing selection, model-based methods, such as the T and B statistics, incorporate recombination distances in their inference framework, and can therefore be especially susceptible to potential inaccuracies in input recombination maps. Additionally, because many approaches for detecting balancing selection aim to identify genomic regions with increased genetic diversity, the elevation of mutation rates is also a common and potent confounding factor for detecting balancing selection (Charlesworth, 2006; Siewert and Voight, 2018; Cheng and DeGiorgio, 2019).

To test their robustness to inaccurate recombination rates, we applied B and T statistics on simulated sequences with uneven recombination maps (10^2-fold fluctuations in recombination rates; see *Methods*). When the sequences evolve neutrally, neither approach is misled (Figure S4). When the fluctuation in recombination rate is even more drastic (*e.g.*, 10^4-fold instead of 10^2), all methods tend to report fewer

false signals than they would under a uniform map (Figure S5). This result suggests that the misleading effects of inaccurate recombination maps are limited.

We next simulated a 10 kb mutational hotspot at the center of the 50 kb sequence with a mutation rate five times higher than original and surrounding rate μ, and applied each statistic with parameters derived from the original neutral replicates with constant mutation rate μ across the entire sequence. All methods exhibit considerable robustness against this regional increase of mutation rate (Figure S6).

We further considered an elevated mutation rate of 5μ across the entire 50 kb sequence, and re-examined the performance of each method. As expected, most statistics display substantially inflated proportions of false signals (Figures 4A and D). Among them, the B_2 statistic reports the least proportion of false signals, followed by the B_1 statistic. Meanwhile, at low false positive rates, B_2 and $B_{2,\mathrm{MAF}}$ statistics report higher proportions of false signals than T_2, their coalescence model-based analogue, whereas B_1 outperformed T_1. Additionally, all statistics that consider only polymorphism data, namely the B_0, $B_{0,\mathrm{MAF}}$, β, and β^* statistics, are substantially misled. The $\beta^{(2)}$ statistic, albeit taking substitutions into account, also displays surprisingly high proportions of false signals. Taken together, these results suggest that B statistics are reasonably robust against high mutation rates, relative to their analogous approaches, and that incorporating substitutions and polarized allele frequencies may further buttress the robustness.

Robustness to demographic history

The influence of demographic history was the major motivation for T statistics to adopt the genome-wide SFS instead of the coalescence-based constant-size neutral model as the null hypothesis, despite that the latter being nested under the alternative model for balancing selection used by the T statistics. This trade-off has endowed T statistics with considerable robustness to population size changes (DeGiorgio et al., 2014; Cheng and DeGiorgio, 2019), but has also potentially compromised their robustness to large windows, as shown in *Robust high power under varying window sizes* subsection of the *Results*. For B statistics, however, because their null models both reflect the genome-wide SFS and are nested under the alternative models, they should exhibit considerable robustness to both oversized windows and demographic changes.

To evaluate their performances under recent population expansions and bottlenecks, we considered the demographic histories of contemporary European humans (Terhorst et al., 2017, Figure S7A) and bonobos (Prado-Martinez et al., 2013, Figure S8A; see details in *Methods*), respectively. The former have been extensively characterized (*e.g.,* Lohmueller et al., 2009; Gravel et al., 2011; Terhorst et al., 2017), and therefore can reliably reflect the performance of each method under realistic scenarios. On the other hand, because we intend to apply the B statistics on bonobo genomic data, we are also interested in evaluating their performance under an inferred bonobo demographic model.

As previously described, we applied the B statistics with unlimited window sizes, whereas the other statistics were provided with smaller window sizes matching the theoretical size for a footprint of long-term balancing selection (see *Supplementary Note* 1). Despite being provided disadvantageous window sizes, B statistics still demonstrate comparable to, and often higher power than, current summary statistic approaches, both under the human (Figure S7) and the bonobo (Figure S8) demographic models. Although T_2 has higher power than the B statistics, we note that the T statistics were operating with optimal window sizes, whereas the window sizes for B statistics are identified across a parameter range. When B_1 and B_2 are applied with identical window sizes as T_1 and T_2 (Figures S9 and S10), the margins between their

performances are no longer substantial. Additionally, we noticed that most statistics tend to have higher power for sequences evolving under the bonobo demographic history than under that of the Europeans (notice that the y-axes in Figures S7 and S8 have different scales).

Finding footprints of balancing selection in bonobo genomes

We further applied the B_2 statistic on the variant calls of 13 bonobos (Prado-Martinez et al., 2013) lifted-over to human genome assembly GRCh38/hg38. Only biallelic single nucleotide polymorphisms (SNPs) were considered, and substitutions were called using bonobo panPan2 reference sequence (Prüfer et al., 2012), with the human sequence as the ancestral state. Stringent filters were applied to remove repetitive regions and regions with poor mappability (see *Methods*), with the resulting mask unintendedly removing the majority of the major histo-compatibility (MHC) locus as well. We observed many genomic regions with outstanding B_2 scores (Figure 5), which lend insights on the evolution of several biological features in bonobos, including pathogen defense and neuro-behaviors.

Evidence for balancing selection on pathogen defense and autoimmunity

Despite that most of the MHC region was removed by the mappability filter, we still observed extraordinary signals from the remainder of this region. More specifically, the *HLA-DQA1* and *HLA-DQB1* genes harbor the highest peak across the genome (Figures 5 and S11). These two genes encode the α and β chains of HLA-DQ molecule, which is a surface receptor on antigen-presenting cells (Ball and Stastny, 1984), and has long been known to be highly polymorphic in great apes (Takahata et al., 1992; Prüfer et al., 2012; Teixeira et al., 2015).

In addition to the *HLA-DQ* genes, *KLRD1* also presents prominent B_2 scores (Figure 6) on its first intron. This gene expresses a natural killer (NK) cell surface antigen, also known as CD94. The interaction between KLRD1 (CD94) and NKG2 family proteins can either inhibit or activate the cytotoxic activity of NK cells (Pende et al., 1997; Cantoni et al., 1998; Masilamani et al., 2006), as well as pivot the generation of cell memory in NK cells (Cerwenka and Lanier, 2016). Furthermore, KLRD1 (CD94) has been shown to play an important role in combating viral infections such as cytomegalovirus (CMV; Cerwenka and Lanier, 2016) and influenza (Bongen et al., 2018) in humans, as well as the mousepox virus in mice (Fang et al., 2011). The involvement in viral defense of *KLRD1* presents an especially intriguing case for bonobos, because they have been recently shown to have reduced level of polymorphism in MHC class I genes (Maibach et al., 2017; Wroblewski et al., 2017), and were predicted to have lower ability to bind with viral peptides when compared with chimpanzees (Maibach and Vigilant, 2019). The genes encoding another regulator of MHC class I molecules, the Killer cell Immunoglobin-like Receptors (KIR), were also found to have contracted short haplotypes in bonobos (Rajalingam et al., 2001; Walter, 2014; Wroblewski et al., 2019), and with the lineage III *KIR* genes serving reduced functions (Wroblewski et al., 2019). In this light, the polymorphisms on *KLRD1* may be compensating the reduced diversity in their binding partners.

Moreover, besides pathogen defense, variants in the *KLRD1* gene were also found to be significantly associated with the condition of Small for Gestational Age (SGA; Harmon et al., 2014) pregnancies among African American humans, which is speculated to result from its interaction with the placental HLA-E, working against implantation. The *KLRD1* gene in humans and chimpanzees was also found to be highly conserved (Khakoo et al., 2000; Shum et al., 2002), unlike in bonobos, which we observed high diversity

on this gene. This may suggest differential selective pressures on *KLRD1* in bonobos compared to humans and chimpanzees.

Several other genes in high-scoring regions are also found to be involved in immunity. For one, the highest peak on chromosome 7 encompasses the entire *GPNMB* gene (Figure S12), with elevated B_2 scores particularly on exons. This gene encodes osteoactivin, a transmembrane glycoprotein found on osteoclast cells, macrophages, and melanoblast (Loftus et al., 2009; Yu et al., 2016), and is shown to regulate proinflammatory responses (Ripoll et al., 2007). Aside from its heavy involvement in cancer (Zhou et al., 2012), GPNMB has also been shown to facilitate tissue repair (Li et al., 2010; Rose et al., 2010; Hu et al., 2013), as well as influence iris pigmentation (Bächner et al., 2002; Maric et al., 2013).

Another genomic region showing outstanding B_2 scores is the intergenic region between *BPIFB4* and *BPIFA2* (Figure S13), which encode two Bacterialcidal Permeability-Increasing Fold-containing (BPIF) family proteins. The BPI proteins are strong bacterialcidal proteins secreted in the mucosal epithelia as one of the key members in first defenses against gram negative bacterial infection (Levy, 2000; Schultz and Weiss, 2007). The protein BPIFA2, found mainly in saliva, also carries antibacterial functions as a part of innate immunity (Prokopovic et al., 2014) and has been shown to protect against intestinal inflammation (Gorr et al., 2015). The *BPIFA2* gene region is recently shown to harbor many SNPs significantly associated with enteropathy (Fujimori et al., 2019) as well. The *BPIFB4* gene, however, is better-known by its association with longevity (Villa et al., 2015b; Spinetti et al., 2017; Villa et al., 2018), and is speculated to accomplish this via vascular protection (Villa et al., 2015a; Puca et al., 2016; Spinelli et al., 2017), although a recent study did find its significant association with enteropathy (Fujimori et al., 2019). Considering that the intergenic region between *BPIFB4* and *BPIFA2* where we observe high B_2 scores locates downstream of *BPIFB4* but upstream of *BPIFA2*, we speculate that the polymorphisms in this area function more to regulate the expression of *BPIFA2*, and are thus driven by innate immunity.

Evidence for balancing selection on neuro-behavior

Among the high-scoring genomic regions, we also found a number of genes involved in neurodevelopment. One of them locates on an intron of *NKAIN3* (Figure 7), which has been significantly associated with male sexuality in humans (Drabant et al., 2012; Sanders et al., 2017). This finding echoes with the pervasive socio-sexual behaviors specific in bonobos (de Waal, 1990; Kano, 1992). Encoding a transmembrane cation channel (Gorokhova, 2006), *NKAIN3* is specifically expressed in the brain (Uhlén et al., 2015), and has also been associated with vitamin E bioavailability (Major et al., 2012; Desmarchelier et al., 2015) and adult ADHD (Weiflog et al., 2013). However, the related molecular mechanisms behind the action of *NKAIN3* remain largely elusive. We also uncovered outstanding signals on the pain receptor gene *SCN9A* (Figure S14), which encodes $Na_V1.7$, a voltage-gated sodium channel, with mutations on the gene associated with various pain disorders (Yang et al., 2004; Cox et al., 2006; Reimann et al., 2010). Interestingly, the peak we observe covers the overlapping RNA gene encoding its anti-sense transcript, *SCN1A-AS1*, which regulates the expression of *SCN9A* (Koenig et al., 2015), suggestive of diversified regulation of pain perception in bonobos.

The other two major signals on chromosome 8 sit in *CSMD1* and *CSMD3* (Figures S15 and S16, respectively). Despite that both genes are large in size (approximately two and 1.5 Mb, respectively) and may be more likely to harbor increased densities of polymorphisms by chance, the high peaks we observe

mostly locate on exons, lending increased support for selection to diversify their functions. Both genes have been associated with neurological conditions such as bipolar disorder (Xu et al., 2014), autism (Floris et al., 2008), schizophrenia (Håvik et al., 2011; Kwon et al., 2013), and epilepsy (Shimizu et al., 2003). Furthermore, multiple association studies have linked *CSMD1* to cognitive functions, especially in elderly (Athanasiu et al., 2017; Stepanov et al., 2017), as well as Parkinson's disease (Patel, 2017; Ruiz-Martínez et al., 2017). *CSMD3* has also be shown to promote dendrite growth (Mizukami et al., 2016), potentially explaining its link to epilepsy (Shimizu et al., 2003) and autism spectrum disorders (Floris et al., 2008).

Discussion

In this study, we introduced a novel set of composite likelihood ratio statistics—B_2, $B_{2,\mathrm{MAF}}$, B_1, B_0, and $B_{0,\mathrm{MAF}}$—to robustly detect footprints of balancing selection with high power and flexibility. The B statistics are based on a mixture model creating a proper nested likelihood ratio test, which helps them overcome the common susceptibility to oversized windows held by current methods. We have extensively evaluated their performances on simulated data compared with current state-of-the-art methods, and have demonstrated the superior properties of the B statistics under various scenarios. We further applied B_2 onto the genomic data of bonobos (Prado-Martinez et al., 2013), and uncovered not only the *HLA-DQA1/HLA-DQB1* gene cluster, but also intriguing candidates that are involved in innate immunity.

Evaluating the performance of B statistics through simulations

In our simulation study, the B statistics showed remarkable robustness to large window sizes, with only minor decays in power under oversized windows, whereas other methods exhibited large declines in power. Moreover, even when considering all data available as input (*i.e.*, the most disadvantageous window size) all variants of B statistics still exhibit comparable power to extant methods and displayed satisfactory performance across varying types and strengths of balancing selection. Under scenarios with confounding factors, such as high mutation rate and non-equilibrium demographic history, the B statistics demonstrated satisfactory robustness as well. Overall, we believe our results have comprehensively characterized the promising performance of the B statistics.

Confounding effects of mutation rate or recombination rate variation

In our simulation study, sequences with a central 10 kb mutational hotspot did not mislead methods as much as those with the mutation rate elevated across the entire sequence. This result may seem counter-intuitive at first, as a smaller region of increased mutation rate may better resemble the footprints of long-term balancing selection. However, upon a closer examination of the site frequency spectra and proportions of polymorphic sites (Figure S17), sequences with an extended region of high mutation rate exhibit a greater departure in these features under scenarios with no elevated mutation rate than for scenarios with a central mutational hotspot. Specifically, these sequences have more sites with high derived allele frequencies and a higher proportion of polymorphic sites overall (Figure S17B), likely resulting from the recurrent mutation on sites that were originally substitutions. The increase is also more profound on sites with high derived allele frequency. For example, the proportions of sites with derived allele frequency of 0.96 increased by almost two-fold from approximately 0.00104 to 0.00190, and the proportions of sites with derived allele

frequency of 0.98 increased by almost three-fold from 0.00105 to 0.00273. By contrast, the difference in scale between the proportions of polymorphisms (0.182 versus 0.189) is minor. The larger fold-change in the proportions of high-frequency polymorphisms (*i.e.*, sites with $k = n - 1,\ n - 2$, and $n - 3$ derived alleles) relative to that of substitutions ($k = n$ derived alleles) could explain the more profound inflation in power for the statistics relying only on information at polymorphic sites. Similarly, after folding the SFS, the large changes in the proportions of low-frequency alleles were substantially mitigated, echoing the superior performance of $B_{2,\mathrm{MAF}}$ and β relative to their unfolded counterparts.

Another unexpected result from the simulations of elevated mutation rate is the drastic inflation of false signals reported by β statistics (Figure 4), which can also be observed in the non-standardized β statistics (Figure S18). Although Siewert and Voight (2018) tested their power to detect balancing selection under high mutation rate, it was unexplored whether their β statistics would mis-classify highly mutable neutral sequences as those undergoing balancing selection, and our results show that they could be easily misled. However, we further found that the performances of the standardized β statistics largely improve when provided with the correct mutation rate and divergence time (Figure S18B). This result partly confirms the superiority of standardized β statistics over the unstandardized ones. It also suggests that β statistics are considerably susceptible to the confounding effect of mutation rate elevation, and that their performance relies highly on the accuracy of the provided mutation rate. Instead of using a constant mutation rate for the entire scan, we propose that providing locally-inferred population-scaled mutation rates θ may help improve the robustness of β statistics. Indeed, when we instead estimate θ using the mean pairwise sequence difference $\widehat{\theta}_\pi$ (Tajima, 1983) for each replicate and provided `BetaScan` the respective inferred value as the θ parameter, the standardized statistics no longer report as many false signals (Figure S18C).

On the other hand, because B, T, and β statistics are computed on every informative site, as suggested by Cheng and DeGiorgio (2019), multiple-testing can account for some inflation in their powers because sequences with a higher mutation rate will have a greater number of informative sites. To evaluate the effect of multiple testing for sequences with high mutation rates, we down-sampled the test sites (see *Methods*) such that the number of test scores being computed approximately matches that under the original mutation rate μ. Although all statistics show varying levels of improvements in performance (Figures 4B, C, E, and F), some still report high proportions of false signals, especially all β statistics and $B_{0,\mathrm{MAF}}$. That is, multiple-testing cannot account for all the factors that drive these statistics to mis-identify features of elevated mutation rates as footprints of balancing selection. This result corresponds to the fact that both the SFS and the density of polymorphic sites are altered under scenarios with extended regions of elevated mutation rate (Figure S17), likely due to recurrent mutation.

Furthermore, we observed that both before and after down-sampling, the T statistics report fewer false signals than their respective B statistic analogues. One potential factor behind their marginally superior performance may be that T statistics perform tests on fixed numbers of informative sites, instead of genomic regions measured by physical lengths (as did B statistics and the summary statistics). For T statistics, the size of the genomic region covered by the same number of informative sites would be much narrower under rapidly mutating sequences than in sequences with the original mutation rate. This means that the resulting T scores in either scenario are reflective of the levels of variation for sequences with drastically different lengths. To account for this factor, we provide B_1 and B_2 with informative site-based windows identical to that of T statistics and re-examined their performances (Figure S19). After matching the

windows, B_1 and B_2 variants in turn display higher robustness than T_1 and T_2 to elevated mutation rates, suggesting that B statistics are at least comparably robust to T statistics. Meanwhile, we also matched the window size for B_0 variants and β to gauge the effect of adopting large windows on the proportions of false signals from B_0 variants. When B_0 scans the sequences with one kb windows, though there is an increase in the resulting number of false signals (Figure S19A), at a 1% false positive rate the proportions of false signals for the two B_0 variants only increase by less than 0.1, and are still substantially lower than that of β and β^* (Figure S19B).

In contrast to mutation rate variation, all statistics are robust to recombination rate variation, with B_0 and $B_{0,\mathrm{MAF}}$ reporting substantially fewer false signals than the others (Figure S4). This robustness to recombination rate variation is likely due to the high similarity in the SFS and proportion of polymorphic sites to sequences with uniform recombination rates (Figure S20).

Evidence for balancing selection in bonobos

As one of the two sister species to humans, bonobos (initially known as the pygmy chimpanzees; Prüfer et al., 2012) have been drawing increasing attention from the genomics community (*e.g.*, Prüfer et al., 2012; Prado-Martinez et al., 2013; de Manuel et al., 2016). However, compared with chimpanzees (the other sister species), bonobos are relatively understudied, despite their close relationship to humans and unique social behaviors. For bonobos, one of their most idiosyncratic traits is their high prevalence of sociosexual activities (de Waal, 1990; Kano, 1992; Wrangham, 1993), which serve important non-reproductive functions and include frequent same-sex encounters. As a close relative to humans, their female-dominance, low-aggression, and hypersexual social behaviors contrast fiercely with those of humans and chimpanzees (Kano, 1992; Wrangham, 1993), and has challenged the traditional understanding of intra- and inter-sexual relationships in hominid societies (Parish et al., 2000). In fact, a standing hypothesis postulates self-domestication as a main player in the evolution of bonobos, in that their unique social behavior has driven the selection against aggression and shaped other related traits (Hare et al., 2012). A growing number of recent studies have also characterized the differences in physiological responses between bonobos and chimpanzees behind their social behaviors (Heilbronner et al., 2008; Hohmann et al., 2009; Wobber et al., 2010a; Deschner et al., 2012; Surbeck et al., 2012), yet the genetic component underlying their unique behaviors, however, remains largely elusive.

In this study, we performed the first model-based scan for footprints of balancing selection on bonobo polymorphism data. In addition to the genomic regions previously reported to be under ancient balancing selection in humans and chimpanzees (*e.g.*, the *HLA-DQ* genes at the MHC locus; Leffler et al., 2013; Teixeira et al., 2015; Cheng and DeGiorgio, 2019), we have also uncovered novel candidates such as *KLRD1* and *NKAIN3*, which are related to pathogen defense and sexuality, respectively. Accumulating evidence (de Waal, 1990; Hare et al., 2012; de Groot et al., 2017; Wroblewski et al., 2017; Maibach and Vigilant, 2019) suggests that bonobos host unique features in phenotypes related to these areas, and our results may enhance our understanding of both the genetic basis of their idiosyncrasies and their evolutionary history.

One surprising finding among all outstanding candidates is the signal of balancing selection on *NKAIN3* (Figure 7), a gene significantly associated with male sexuality in humans (Drabant et al., 2012; Sanders et al., 2017). The high B_2 scores are also observed on another sexuality-associated locus, the intergenic region between *SLITRK5* and *SLITRK6* (Sanders et al., 2017, Figure S21). These genomic footprints

echo the high prevalence of same-sex sociosexual interactions in bonobo communities, which suggest a potential balance between the social benefits and possible decrease in fitness of other traits brought by such behaviors. Further, a few other genes involved in neuro-development, such as *CSMD1* and *SCN9A*, also show outstanding B_2 scores. These footprints may correspond with the different stress-coping behaviors in bonobos, as well as the observations of heterochrony in the development of social inhibition learning (Wobber et al., 2010b) and spacial memory (Rosati, 2019) between bonobos and chimpanzees.

In addition to their unique behavior, the immune genes in bonobos are also intriguingly more reduced than in humans and chimpanzees, both in their lower diversity of MHC-I repertoires (de Groot et al., 2017; Maibach et al., 2017; Wroblewski et al., 2017), and in the reduction in the number of both *KIR* genes and their allotypes (Rajalingam et al., 2001). These features are unlikely the natural consequences of demographic factors even after considering the harsher bottlenecks bonobos have undergone compared with chimpanzees; many studies have speculated potential selective sweeps in bonobos on these regions (Prüfer et al., 2012; Walter, 2014; Maibach et al., 2017; Wroblewski et al., 2017, 2019). Specifically, de Groot et al. (2002) suggest that an ancient selective sweep, potentially due to simian immunodeficiency virus (SIV; de Groot et al., 2010; Prüfer et al., 2012), may have occurred in the ancestors of bonobos and chimpanzees around two to three million years ago, reducing the MHC-I diversity in both species. This selective pressure may have relaxed later in bonobos due to the reduction of SIV-carrying monkeys in their diet (Serckx et al., 2015; de Groot et al., 2018), which explains the lack of detectable SIV-infection in bonobos (Van Dooren et al., 2002; Li et al., 2012). de Groot et al. (2018) further proposed that the diversity at the MHC locus in bonobos was likely shaped by strong selective pressures from malaria parasites after their divergence from chimpanzees, as several species of *Plasmodium* parasites have been found in bonobo populations (Krief et al., 2010; Liu et al., 2017), and the few remaining MHC-B subtypes in bonobos have been shown to confer protection against malaria (Hill et al., 1992; Sanchez-Mazas et al., 2017). This hypothesis agrees with our observation that *MHC-DQ* genes, which are often involved in parasite defense, show the highest B_2 scores genome-wide. On the other hand, our results can also relate to the defense of viral pathogens in bonobos. Besides the involvement of *KLRD1* in combating virus infections, *CSMD1* (Figure S15) has also been shown to be associated with antibody response to smallpox vaccines in humans (Ovsyannikova et al., 2012)—intriguingly, bonobos were found to tolerate monkeypox, the simian-borne poxvirus (Inogwabini and Leader-Williams, 2012).

Lastly, we noticed that some candidate genes carry multiple distinct functions, and may have been undergoing balancing selection due to potential evolutionary conflicts between some of their functions. For example, the gene *CSMD1* is not only associated with antibody response to smallpox vaccines (Ovsyannikova et al., 2012), it also plays a role in neuro-development (Kwon et al., 2013; Xu et al., 2014; Athanasiu et al., 2017). Likewise, the gene *GPNMB* plays roles not only in tissue repair (Li et al., 2010), but also in iris pigmentation (Bächner et al., 2002). Another candidate, *PDE1A* gene (Figure S22), encodes a phosphodiesterase that is pivotal to Ca^{2+}- and cyclic nucleotide-signaling (Lefièvre et al., 2002). It is expressed in brain, endocrine tissues, kidneys, and gonads (Uhlén et al., 2015), and has multiple splicing variants. In fact, the high-scoring peak we observed on this gene happens to locate around the exons that are spliced out in some variants (Figure S22). Studies have demonstrated the relation of this gene to brain development (Yan et al., 1994), mood and cognitive disorders (Xu et al., 2011; Martinez and Gil, 2013; Pekcec et al., 2018; Betolngar et al., 2019), and hypertension (Kimura et al., 2017). Meanwhile, the PDE1A protein is

also a conserved component of mammalian spermatozoa (Lefièvre et al., 2002; Vasta et al., 2005), and is involved in the movement of flagella—organelles featured in both animal sperms and some protozoa such as the *Trypanosoma* parasites (Oberholzer et al., 2007), with which the endemic geographical regions are avoided by bonobos (Inogwabini and Leader-Williams, 2012). Though it is difficult to judge for these genes which functions may be subject to selective pressures, they nonetheless indicate that pleiotropy can be an important driver of balancing selection.

Comparing the B statistics with the T statistics

Because the T statistics of DeGiorgio et al. (2014) have previously been the only model-based approach for the detection of long-term balancing selection from polymorphism data in a single species, the comparisons between the model-based B and T approaches is particularly intriguing for researchers with empirical data suitable for the application of either. The T statistics are based on an explicit coalescent model (Hudson and Kaplan, 1988; Kaplan et al., 1988), and have been shown to have superior power to a number of other methods in previous studies (DeGiorgio et al., 2014; Siewert and Voight, 2017, 2018; Bitarello et al., 2018; Cheng and DeGiorgio, 2019), consistent with our simulation results. The B statistics, on the other hand, employ a mixture model, where the component modeling balancing selection is not based on an explicit evolutionary model, but nevertheless shows impressive performance on simulated data, as the shape of the distribution of allele frequencies is similar to what might be expected under balancing selection. The often superior performances of both approaches over summary statistics is understandable, as both utilize the genomic spatial distribution of genetic diversity in their inferences.

However, within the T statistic framework, the model for the null hypothesis (neutrality) is not nested in the alternative hypothesis (balancing selection). Although the T_1 and T_2 statistics could have adopted nested models by employing the standard neutral coalescent as the model for the null hypothesis, doing so would increase susceptibility to demographic factors, which can also alter the genome-wide SFS. To better account for these factors, DeGiorgio et al. (2014) instead employed the genome-wide distribution of genetic variation to compute probabilities under the null hypothesis of neutrality. This explains the substantial decay in power for both T statistics as the window size increases (Figure 2), as well as its robust performance under varying sized demographic models (DeGiorgio et al., 2014; Cheng and DeGiorgio, 2019, Figures S7 and S8). In contrast to the T statistics, the null model for B statistics (which also employs the genome-wide SFS) is nested within the alternative, due to their mixture model framework. This feature mitigates the biases introduced by sites far from the test site, while simulataneously accounting for demographic factors. Consequently, the B statistics display robust performance under oversized windows and realistic demographic models in our simulations (Figures 2, S7, and S8).

Another advantage of the B statistics over the T statistic approach, especially for B_2 compared with T_2, is the computational load. Because the probability distribution of allele frequencies under the Kaplan-Darden-Hudson (Kaplan et al., 1988) model is difficult to compute, the T_2 statistic relies on previously-generated sets of simulated site frequency spectra over a grid of equilibrium frequencies $x \in \{0.05, 0.10, \ldots, 0.95\}$ for each distinct sample size n and recombination distance d. Generation of such frequency spectra is computationally intensive, and the load increases substantially with the increase in sample size, thereby limiting the application of T_2 to datasets with larger sample sizes. However, this is not a limitation of B_2, as the SFS under balancing selection is determined simply as a mixture of the

given genome-wide distribution of allele frequencies and a statistical distribution with closed-form solutions whose computational cost is minor, and only increases linearly with the sample size. Moreover, the rapid computation of this spectrum permits a finer grid of equilibrium frequencies x to be interrogated.

Concluding remarks

Extant methods for detecting long-term balancing selections are constrained by the pliability of their inferences as a function of genomic window size. In this study, we presented B statistics, a set of composite likelihood ratio statistics based on nested mixture models. We have comprehensively evaluated their performances through simulations and demonstrated their robust high performances over varying window sizes in uncovering genomic loci undergoing balancing selection. Moreover, we showed that even when applied with the least optimal window sizes, the B statistics still exhibit high power comparable to current methods, which operated under optimal window sizes, in uncovering balancing selection of varying age and selection parameters, as well as robust performance under confounding scenarios such as elevated mutation rates, variable recombination rates, and population size changes. We further applied the B_2 statistic on the whole-genome polymorphism data of bonobos, and discovered not only the well-established *MHC-DQ* genes, but also novel candidates such as *KLRD1*, *NKAIN3*, *PDE1A*, *SCN9A*, and *CSMD1*, with functional implications ranging from pathogen defense to neuro-behavior. We have implemented these statistics in the open source software `BalLeRMix`, which can be accessed at `http://www.personal.psu.edu/mxd60/ballermix.html` along with the empirical scan results for balancing selection in bonobos.

Methods

In this section, we discuss sets of simulations used to evaluate the performances of the B statistics relative to previously-published state-of-the-art approaches (Hudson et al., 1987; DeGiorgio et al., 2014; Siewert and Voight, 2017, 2018; Bitarello et al., 2018). Finally, we describe the application of our B statistics to an empirical bonobo dataset (Prado-Martinez et al., 2013).

Evaluating methods through simulations

We employed the forward-time genetic simulator SLiM (version 3.2; Haller and Messer, 2019) to generate sequences of 50 kb in length evolving with or without balancing selection. Based on the respective levels in humans and other great apes, we assumed a mutation rate of $\mu = 2.5 \times 10^{-8}$ per base per generation (Nachman and Crowell, 2000), and a recombination rate of $r = 10^{-8}$ per base per generation (Payseur and Nachman, 2000). In scenarios with constant population sizes, we set the diploid effective population size as $N = 10^4$. To create baseline genetic variation, each replicate simulation was initiated with a burn-in period of $10N = 10^5$ generations. To speed up simulations, we applied the scaling parameter λ to the number of simulated generations, population size, mutation rate, recombination rate, and selection coefficient, which allows for the generation of the same levels of variation with a speed up in computational time by a factor λ^2. For scenarios based on a model of constant population size, we used $\lambda = 100$. For the demographic models of European humans and bonobos, we used $\lambda = 20$.

We simulated data from two other diverged species, under the demographic history inspired by that of humans, chimpanzees (Kumar et al., 2005), and gorillas (Scally et al., 2012). Specifically, the closer and farther outgroups diverged 2.5×10^5 and 4×10^5 generations ago, respectively, which correspond to five million and eight million years ago, assuming a generation time of 20 years.

To evaluate the power of each method to detect balancing selection with varying selective coefficient s, dominance coefficient h, and age, for each combination of s and h, we considered 15 time points at which the selected allele was introduced, ranging from 5×10^4 to 6.5×10^5 generations prior to sampling with time points separated by intervals of 5×10^4 generations. Assuming a generation time of 20 years, these time points are equivalent to $1, 2, 3, \ldots, 15$ million years before sampling. For each scenario, a single selected mutation was introduced at the center of each sequence at the assigned time point, and we only considered simulations where the introduced allele was not lost.

Accelerated mutation rate

To evaluate whether the B statistics are robust to high mutation rates, we applied the methods on simulated sequences evolving neutrally along the same demographic history (Figure S1), but instead with a five-fold higher mutation rate of $5\mu = 1.25 \times 10^{-7}$ per site per generation. To generate sequences with regional increases in mutation rate, we simulated 50 kb sequences with a five-fold higher mutation rate of $5\mu = 1.25 \times 10^{-7}$ per site per generation at the central 10 kb of the sequence, and the surrounding region with the original rate μ.

Recombination rate estimation error

For evaluating the robustness to erroneous estimation of recombination rates, we simulated sequences with uneven recombination maps, and applied the model-based methods with the assumption that the recombination rate is uniform. In particular, we divided the 50 kb sequence into 50 regions of one kb each, and in turns inflate or deflate the recombination rate of each region by m fold, such that the recombination rates of every pair of neighboring regions have a m^2-fold difference. We tested $m = 10$ and $m = 100$ in this study.

Demographic history

To examine the performance of methods under realistic demographic parameters, we considered the demographic histories of a central European human population (CEU; Terhorst et al., 2017) and of bonobos (Prado-Martinez et al., 2013). For the human population, we adopted the history of population size changes inferred by SMC++ (Terhorst et al., 2017) that spans 10^5 generations, assuming a mutation rate of 5×10^{-9} per site per generation and a scaling effective size of 10^4 diploids. To account for recombination rate variation, we allowed each simulated replicate to have a uniform recombination rate drawn uniformly at random between 5×10^{-9} and 1.5×10^{-8} per site per generation. We also simulated an additional population that split from the human population 2.5×10^5 generations ago, which is identical to the outgroup (named O1) in the demographic model depicted in Figure 3A, with an effective size of 10^4 diploid individuals.

For the bonobo population history, we scaled the PSMC history inferred from the genome of individual A917 (Dzeeta; sample SRS396202) by Prado-Martinez et al. (2013) with a mutation rate of 2.5×10^{-8} per

site per generation, identical to the simulations on the three-population demographic history (Figure 3A). Because the inferred PSMC model provides a specific ratio of the mutation and recombination rates, we set the recombination rate to 2.84×10^{-9} per site per generation. To be consistent with the three-population demographic history, we set the population size prior to 71,640 generations ago, which is the maximum time covered by the PSMC inference, to 10^4 diploid individuals, and had the outgroup split 2.5×10^5 generations ago with the same diploid population size, identical to the outgroup O1 in the three-population demographic history (Figure 2A).

Application to empirical data

We obtained the genotype calls of 13 bonobos from the Great Ape Project (Prado-Martinez et al., 2013), which were mapped to human genome assembly NCBI36/hg18. We lifted over the variant calls to human genome assembly GRCh38/hg38, and polarized the allele frequencies with the bonobo genome assembly panPan2, with the sequence in hg38 considered as the ancestral allele. Only genomic regions mappable across hg38 and panPan2 were considered for further analyses. For mappable polymorphic sites, we only considered bi-allelic SNPs. For mappable sites without variant calls in bonobo, we assumed these sites were monomorphic for the panPan2 reference genome sequence, and called substitutions based on whether the panPan2 reference allele was different from the hg38 reference allele.

To circumvent potential artifacts, we performed one-tailed Hardy-Weinberg equilibrium tests on each site and removed sites with an excess of heterozygotes ($p < 0.01$). This p-value was determined by the distribution of the p-values of all polymorphic sites across the genome, such that 0.035% of such sites are outliers. We also applied the RepeatMasker, segmental duplication, simple repeat, and interrupted repeat filters (all downloaded from UCSC Genome Browser) to remove repetitive regions. To assess the mappability of each genomic region, we employed the mappability module GEMtools (Derrien et al., 2012), and computed the mappability tracks of 50-mers. Regions with 50-mer mappability scores lower than 0.9 were removed. Because `BalLeRMix` employs a pre-specified grid of A values to accompany the distances d in centi-Morgans (cM), when applying the method, we assumed a uniform recombination rate of 10^{-6} cM per site, which is the approximate recombination rate in humans (Payseur and Nachman, 2000).

Acknowledgments

This work was funded by National Institutes of Health grant R35GM128590, by National Science Foundation grant DEB-1753489, and by the Alfred P. Sloan Foundation. Portions of this research were conducted with Advanced CyberInfrastructure computational resources provided by the Institute for CyberScience at Pennsylvania State University.

References

A. M. Andrés. *Balancing Selection in the Human Genome.* John Wiley and Sons, Ltd, 2001.

A. M. Andrés, M. J. Hubisz, A. Indap, D. G. Torgerson, J. D. Degenhardt, A. R. Boyko, R. N. Gutenkunst, T. J. White, E. D. Green, C. D. Bustamante, A. G. Clark, and R. Nielsen. Targets of balancing selection in the human genome. *Molecular Biology and Evolution*, 26(12):2755, 2009.

L. Athanasiu, S. Giddaluru, C. Fernandes, A. Christoforou, I. Reinvang, A. J. Lundervold, L.-G. Nilsson, K. Kauppi, R. Adolfsson, E. Eriksson, et al. A genetic association study of csmd1 and csmd2 with cognitive function. *Brain, Behavior, and Immunity*, 61:209–216, 2017.

D. Bächner, D. Schröder, and G. Gross. mRNA expression of the murine glycoprotein (transmembrane) nmb (Gpnmb) gene is linked to the developing retinal pigment epithelium and iris. *Gene Expression Patterns*, 1(3-4):159–165, 2002.

E. Ball and P. Stastny. Antigen-specific HLA-restricted human T-cell lines. II. A GAT-specific T-cell line restricted by a determinant carried by an HLA-DQ molecule. *Immunogenetics*, 20(5):547564, 1984. ISSN 0093-7711. doi: 10.1007/bf00364357.

A. O. Bergland, E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in drosophila. *PLoS Genetics*, 10(11): e1004775, 2014.

D. B. Betolngar, É. Mota, A. Fabritius, J. Nielsen, C. Hougaard, C. T. Christoffersen, J. Yang, J. Kehler, O. Griesbeck, L. R. Castro, et al. Phosphodiesterase 1 bridges glutamate inputs with no-and dopamine-induced cyclic nucleotide signals in the striatum. *Cerebral Cortex*, 2019.

B. D. Bitarello, C. de Filippo, J. C. Teixeira, J. M. Schmidt, P. Kleinert, D. Meyer, and A. M. Andrs. Signatures of long-term balancing selection in human genomes. *Genome Biology and Evolution*, 10(3): 939–955, 2018.

E. Bongen, F. Vallania, P. J. Utz, and P. Khatri. Klrd1-expressing natural killer cells predict influenza susceptibility. *Genome Medicine*, 10(1):45, 2018.

A. Burri, T. Spector, and Q. Rahman. Common genetic factors among sexual orientation, gender nonconformity, and number of sex partners in female twins: Implications for the evolution of homosexuality. *The Journal of Sexual Medicine*, 12(4):1004–1011, 2015.

C. Cantoni, R. Biassoni, D. Pende, S. Sivori, L. Accame, L. Pareti, G. Semenzato, L. Moretta, A. Moretta, and C. Bottino. The activating form of cd94 receptor complex: Cd94 covalently associated with the kp39 protein that represents the product of the nkg2-c gene. *European Journal of Immunology*, 28(1): 327–338, 1998.

A. Cerwenka and L. L. Lanier. Natural killer cell memory in infection, inflammation and cancer. *Nature Reviews Immunology*, 16(2):112, 2016.

B. Charlesworth and D. Charlesworth. *Elements of evolutionary genetics*. Roberts and Company Publishers Greenwood Village, 2010.

D. Charlesworth. Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genetics*, 2(4):1–6, 04 2006.

X. Cheng and M. DeGiorgio. Detection of shared balancing selection in the absence of trans-species polymorphism. *Molecular Biology and Evololution*, 36:177–199, 2019.

S. Cho, Z. Y. Huang, D. R. Green, D. R. Smith, and J. Zhang. Evolution of the complementary sex-determination gene of honey bees: Balancing selection and trans-species polymorphisms. *Genome Research*, 16(11):1366–1375, 2006.

A. C. Ciani, U. Battaglia, and G. Zanzotto. Human homosexuality: a paradigmatic arena for sexually antagonistic selection? *Cold Spring Harbor Perspectives in Biology*, 7(4):a017657, 2015.

A. C. Ciani, U. Battaglia, L. Cesare, G. C. Ciani, and C. Capiluppi. Possible balancing selection in human female homosexuality. *Human Nature*, 29(1):14–32, 2018.

J. J. Cox, F. Reimann, A. K. Nicholas, G. Thornton, E. Roberts, K. Springell, G. Karbani, H. Jafri, J. Mannan, Y. Raashid, et al. An scn9a channelopathy causes congenital inability to experience pain. *Nature*, 444(7121):894, 2006.

N. G. de Groot, N. Otting, G. G. Doxiadis, S. S. Balla-Jhagjhoorsingh, J. L. Heeney, J. J. van Rood, P. Gagneux, and R. E. Bontrop. Evidence for an ancient selective sweep in the mhc class i gene repertoire of chimpanzees. *Proceedings of the National Academy of Sciences*, 99(18):11748–11753, 2002.

N. G. de Groot, C. M. Heijmans, Y. M. Zoet, A. H. de Ru, F. A. Verreck, P. A. van Veelen, J. W. Drijfhout, G. G. Doxiadis, E. J. Remarque, I. I. Doxiadis, et al. Aids-protective hla-b* 27/b* 57 and chimpanzee mhc class i molecules target analogous conserved areas of hiv-1/sivcpz. *Proceedings of the National Academy of Sciences*, 107(34):15175–15180, 2010.

N. G. de Groot, C. M. Heijmans, P. Helsen, N. Otting, Z. Pereboom, J. M. Stevens, and R. E. Bontrop. Limited mhc class i intron 2 repertoire variation in bonobos. *Immunogenetics*, 69(10):677–688, 2017.

N. G. de Groot, J. M. Stevens, and R. E. Bontrop. Does the MHC Confer Protection against Malaria in Bonobos? *Trends in Immunology*, 39(10):768 – 771, 2018. ISSN 1471-4906.

M. de Manuel, M. Kuhlwilm, P. Frandsen, V. C. Sousa, T. Desai, J. Prado-Martinez, J. Hernandez-Rodriguez, I. Dupanloup, O. Lao, P. Hallast, J. M. Schmidt, J. M. Heredia-Genestar, A. Benazzo, G. Barbujani, B. M. Peter, L. F. K. Kuderna, F. Casals, S. Angedakin, M. Arandjelovic, C. Boesch, H. Kühl, L. Vigilant, K. Langergraber, J. Novembre, M. Gut, I. Gut, A. Navarro, F. Carlsen, A. M. Andrés, H. R. Siegismund, A. Scally, L. Excoffier, C. Tyler-Smith, S. Castellano, Y. Xue, C. Hvilsom, and T. Marques-Bonet. Chimpanzee genomic diversity reveals ancient admixture with bonobos. *Science*, 354(6311):477–481, 2016. doi: 10.1126/science.aag2602.

F. B. de Waal. Sociosexual behavior used for tension regulation in all age and sex combinations among bonobos. In *Pedophilia*, pages 378–393. Springer, 1990.

M. DeGiorgio, K. E. Lohmueller, and R. Nielsen. A Model-Based Approach for Identifying Signatures of Ancient Balancing Selection in Genetic Data. *PLoS Genetics*, 10(8):e1004561, 2014.

T. Derrien, J. Estellé, S. M. Sola, D. G. Knowles, E. Raineri, R. Guigó, and P. Ribeca. Fast computation and applications of genome mappability. *PLoS One*, 7(1):e30377, 2012.

T. Deschner, B. T. Fuller, V. M. Oelze, C. Boesch, J.-J. Hublin, R. Mundry, M. P. Richards, S. Ortmann, and G. Hohmann. Identification of energy consumption and nutritional stress by isotopic and elemental analysis of urine in bonobos (pan paniscus). *Rapid Communications in Mass Spectrometry*, 26(1):69–77, 2012.

C. Desmarchelier, F. Tourniaire, M. Nowicki, R. Bott, and P. Borel. The interindividual variability in vitamin e bioavailability in healthy male adults is significantly explained by a combination of snps in genes involved in vitamin e metabolism. *Atherosclerosis*, 241(1):e133–e134, 2015.

E. Drabant, A. Kiefer, N. Eriksson, J. Mountain, U. Francke, J. Tung, D. Hinds, and C. Do. Genome wide association study of sexual orientation in a large, web-based cohort, 2012.

W. Ewens. A note on the sampling theory for infinite alleles and infinite sites models. *Theoretical Population Biology*, 6(2):143–148, 1974.

M. Fang, M. T. Orr, P. Spee, T. Egebjerg, L. L. Lanier, and L. J. Sigal. Cd94 is essential for nk cell-mediated resistance to a lethal viral disease. *Immunity*, 34(4):579–589, 2011.

C. Floris, S. Rassu, L. Boccone, D. Gasperini, A. Cao, and L. Crisponi. Two patients with balanced translocations and autistic disorder: Csmd3 as a candidate gene for autism found in their common 8q23 breakpoint area. *European Journal of Human Genetics*, 16(6):696, 2008.

S. Fujimori, K. Fukunaga, A. Takahashi, T. Mushiroda, M. Kubo, R. Hanada, M. Hayashida, T. Sakurai, K. Iwakiri, and C. Sakamoto. Bactericidal/permeability-increasing fold-containing family b member 4 may be associated with nsaid-induced enteropathy. *Digestive Diseases and Sciences*, 64(2):401–408, 2019.

Z. Gao, M. Przeworski, and G. Sella. Footprints of ancient-balanced polymorphisms in genetic variation data from closely related species. *Evolution*, 69(2):431–446, 2015.

S. Gorokhova. Nkains: A novel family of membrane proteins that interact with the na, k-atpase and affect membrane excitability. 2006.

S.-U. Gorr, H. Hirt, and S. R. Nandula. The salivary protein PSP/Bpifa2 protects against intestinal inflammation. *The FASEB Journal*, 29(1_supplement):LB684, 2015.

S. Gravel, B. M. Henn, R. N. Gutenkunst, A. R. Indap, G. T. Marth, A. G. Clark, F. Yu, R. A. Gibbs, C. D. Bustamante, . G. Project, et al. Demographic history and rare allele sharing among human populations. *Proceedings of the National Academy of Sciences*, 108(29):11983–11988, 2011.

S. Guirao-Rico, A. Sánchez-Gracia, and D. Charlesworth. Sequence diversity patterns suggesting balancing selection in partially sex-linked genes of the plant silene latifolia are not generated by demographic history or gene flow. *Molecular Ecology*, 26(5):1357–1370, 2017.

B. C. Haller and P. W. Messer. Slim 3: Forward genetic simulations beyond the wrightfisher model. *Molecular Biology and Evolution*, 36:632–637, 2019.

B. Hare, V. Wobber, and R. Wrangham. The self-domestication hypothesis: evolution of bonobo psychology is due to selection against aggression. *Animal Behaviour*, 83(3):573–585, 2012.

Q. E. Harmon, S. M. Engel, M. C. Wu, T. M. Moran, J. Luo, A. M. Stuebe, C. L. Avery, and A. F. Olshan. Polymorphisms in inflammatory genes are associated with term small for gestational age and preeclampsia. *American Journal of Reproductive Immunology*, 71(5):472–484, 2014.

B. Håvik, S. Le Hellard, M. Rietschel, H. Lybæk, S. Djurovic, M. Mattheisen, T. W. Mühleisen, F. Degenhardt, L. Priebe, W. Maier, et al. The complement control-related genes csmd1 and csmd2 associate to schizophrenia. *Biological Psychiatry*, 70(1):35–42, 2011.

S. R. Heilbronner, A. G. Rosati, J. R. Stevens, B. Hare, and M. D. Hauser. A fruit in the hand or two in the bush? divergent risk preferences in chimpanzees and bonobos. *Biology Letters*, 4(3):246–249, 2008.

A. V. Hill, J. Elvin, A. C. Willis, M. Aidoo, C. E. Allsopp, F. M. Gotch, X. M. Gao, M. Takiguchis, B. M. Greenwood, A. R. Townsend, et al. Molecular analysis of the association of hla-b53 and resistance to severe malaria. *Nature*, 360(6403):434, 1992.

G. Hohmann, R. Mundry, and T. Deschner. The relationship between socio-sexual behavior and salivary cortisol in bonobos: tests of the tension regulation hypothesis. *American Journal of Primatology: Official Journal of the American Society of Primatologists*, 71(3):223–232, 2009.

X. Hu, P. Zhang, Z. Xu, H. Chen, and X. Xie. GPNMB enhances bone regeneration by promoting angiogenesis and osteogenesis: potential role for tissue engineering bone. *Journal of Cellular Biochemistry*, 114(12):2729–2737, 2013.

R. R. Hudson and N. L. Kaplan. The coalescent process in models with selection and recombination. *Genetics*, 120(3):831–840, 1988.

R. R. Hudson, M. Kreitman, and M. Aguadé. A test of neutral molecular evolution based on nucleotide data. *Genetics*, 116(1):153–159, 1987.

H. Hunter-Zinck and A. G. Clark. Aberrant time to most recent common ancestor as a signature of natural selection. *Molecular Biology and Evolution*, 32(10):2784–2797, 2015.

F. Iemmola and A. C. Ciani. New evidence of genetic factors influencing sexual orientation in men: Female fecundity increase in the maternal line. *Archives of Sexual Behavior*, 38(3):393–399, 2009.

B.-I. Inogwabini and N. Leader-Williams. Effects of epidemic diseases on the distribution of bonobos. *PLoS One*, 7(12):e51112, 2012.

T. Kano. *The last ape: Pygmy chimpanzee behavior and ecology*, volume 155. Stanford University Press Stanford, 1992.

N. L. Kaplan, T. Darden, and R. R. Hudson. The coalescent process in models with selection. *Genetics*, 120(3):819–829, 1988.

S. I. Khakoo, R. Rajalingam, B. P. Shum, K. Weidenbach, L. Flodin, D. G. Muir, F. Canavez, S. L. Cooper, N. M. Valiante, L. L. Lanier, et al. Rapid evolution of nk cell receptor systems demonstrated by comparison of chimpanzees and humans. *Immunity*, 12(6):687–698, 2000.

M. Kimura, Y. Tamura, C. Guignabert, M. Takei, K. Kosaki, N. Tanabe, K. Tatsumi, T. Saji, T. Satoh, M. Kataoka, et al. A genome-wide association analysis identifies pde1a— dnajc10 locus on chromosome 2 associated with idiopathic pulmonary arterial hypertension in a japanese population. *Oncotarget*, 8 (43):74917, 2017.

J. Koenig, R. Werdehausen, J. E. Linley, A. M. Habib, J. Vernon, S. Lolignier, N. Eijkelkamp, J. Zhao, A. L. Okorokov, C. G. Woods, et al. Regulation of nav1. 7: a conserved scn9a natural antisense transcript expressed in dorsal root ganglia. *PLoS One*, 10(6):e0128830, 2015.

S. Krief, A. A. Escalante, M. A. Pacheco, L. Mugisha, C. André, M. Halbwax, A. Fischer, J.-M. Krief, J. M. Kasenene, M. Crandfield, et al. On the diversity of malaria parasites in african apes and the origin of plasmodium falciparum from bonobos. *PLoS Pathogens*, 6(2):e1000765, 2010.

S. Kumar, A. Filipski, V. Swarna, A. Walker, and S. B. Hedges. Placing confidence limits on the molecular age of the human-chimpanzee divergence. *Proceedings of the National Academy of Sciences*, 102(52): 18842–18847, 2005.

E. Kwon, W. Wang, and L. Tsai. Validation of schizophrenia-associated genes CSMD1, C10orf26, CACNA1C and TCF4 as miR-137 targets. *Molecular psychiatry*, 18(1):11, 2013.

E. M. Leffler, Z. Gao, S. Pfeifer, L. Ségurel, A. Auton, O. Venn, R. Bowden, R. Bontrop, J. D. Wall, G. Sella, P. Donnelly, G. McVean, and M. Przeworski. Multiple instances of ancient balancing selection shared between humans and chimpanzees. *Science*, 339(6127):1578–1582, 2013.

L. Lefièvre, E. de Lamirande, and C. Gagnon. Presence of cyclic nucleotide phosphodiesterases pde1a, existing as a stable complex with calmodulin, and pde3a in human spermatozoa. *Biology of Reproduction*, 67(2):423–430, 2002.

L. Lefièvre, K. N. Jha, E. de Lamirande, P. E. Visconti, and C. Gagnon. Activation of protein kinase a during human sperm capacitation and acrosome reaction. *Journal of Andrology*, 23(5):709–716, 2002.

O. Levy. A neutrophil-derived anti-infective molecule: bactericidal/permeability-increasing protein. *Antimicrobial Agents and Chemotherapy*, 44(11):2925–2931, 2000.

B. Li, A. P. Castano, T. E. Hudson, B. T. Nowlin, S.-L. Lin, J. V. Bonventre, K. D. Swanson, and J. S. Duffield. The melanoma-associated transmembrane glycoprotein gpnmb controls trafficking of cellular debris for degradation and is essential for tissue repair. *The FASEB Journal*, 24(12):4767–4781, 2010.

Y. Li, J.-B. Ndjango, G. H. Learn, M. A. Ramirez, B. F. Keele, F. Bibollet-Ruche, W. Liu, J. L. Easlick, J. M. Decker, R. S. Rudicell, et al. Eastern chimpanzees, but not bonobos, represent a simian immunodeficiency virus reservoir. *Journal of Virology*, 86(19):10776–10791, 2012.

W. Liu, S. Sherrill-Mix, G. H. Learn, E. J. Scully, Y. Li, A. N. Avitto, D. E. Loy, A. P. Lauder, S. A. Sundararaman, L. J. Plenderleith, et al. Wild bonobos host geographically restricted malaria parasites including a putative new laverania species. *Nature Communications*, 8(1):1635, 2017.

S. K. Loftus, A. Antonellis, I. Matera, G. Renaud, L. L. Baxter, D. Reid, T. G. Wolfsberg, Y. Chen, C. Wang, N. C. S. Program, et al. Gpnmb is a melanoblast-expressed, MITF-dependent gene. *Pigment Cell & Melanoma Research*, 22(1):99–110, 2009.

K. E. Lohmueller, C. D. Bustamante, and A. G. Clark. Methods for human demographic inference using haplotype patterns from genomewide single-nucleotide polymorphism data. *Genetics*, 182(1):217–231, 2009.

E. Lonn, E. Koskela, T. Mappes, M. Mokkonen, A. M. Sims, and P. C. Watts. Balancing selection maintains polymorphisms at neurogenetic loci in field experiments. *Proceedings of the National Academy of Sciences*, page 201621228, 2017.

V. Maibach and L. Vigilant. Reduced bonobo mhc class i diversity predicts a reduced viral peptide binding ability compared to chimpanzees. *BMC Evolutionary Biology*, 19(1):14, 2019.

V. Maibach, J. B. Hans, C. Hvilsom, T. Marques-Bonet, and L. Vigilant. Mhc class i diversity in chimpanzees and bonobos. *Immunogenetics*, 69(10):661–676, 2017.

J. M. Major, K. Yu, C. C. Chung, S. J. Weinstein, M. Yeager, W. Wheeler, K. Snyder, M. E. Wright, J. Virtamo, S. Chanock, et al. Genome-wide association study identifies three common variants associated with serologic response to vitamin e supplementation in men. *The Journal of Nutrition*, 142(5): 866–871, 2012.

G. Maric, A. A. Rose, M. G. Annis, and P. M. Siegel. Glycoprotein non-metastatic b (GPNMB): A metastatic mediator and emerging therapeutic target in cancer. *OncoTargets and therapy*, 6:839, 2013.

A. Martinez and C. Gil. *Phosphodiesterase Inhibitors as a New Therapeutic Approach for the Treatment of Parkinsons*. Royal Society of Chemistry, 2013.

M. Masilamani, C. Nguyen, J. Kabat, F. Borrego, and J. E. Coligan. Cd94/nkg2a inhibits nk cell activation by disrupting the actin network at the immunological synapse. *The Journal of Immunology*, 177(6):3590–3596, 2006.

T. Mitchell-Olds, J. H. Willis, and D. B. Goldstein. Which evolutionary processes influence natural genetic variation for phenotypic traits? *Nature Reviews Genetics*, 8(11):845–56, 11 2007. Copyright - Copyright Nature Publishing Group Nov 2007; Last updated - 2014-03-31.

T. Mizukami, T. Kohno, and M. Hattori. Cub and sushi multiple domains 3 regulates dendrite development. *Neuroscience Research*, 110:11–17, 2016.

M. W. Nachman and S. L. Crowell. Estimate of the mutation rate per nucleotide in humans. *Genetics*, 156(1):297–304, 2000.

M. Oberholzer, P. Bregy, G. Marti, M. Minca, M. Peier, and T. Seebeck. Trypanosomes and mammalian sperm: one of a kind? *Trends in Parasitology*, 23(2):71 – 77, 2007. ISSN 1471-4922.

I. G. Ovsyannikova, R. B. Kennedy, M. OByrne, R. M. Jacobson, V. S. Pankratz, and G. A. Poland. Genome-wide association study of antibody response to smallpox vaccine. *Vaccine*, 30(28):4182–4189, 2012.

A. R. Parish, F. B. M. de Waal, and D. Haig. The other closest living relative: How bonobos (*Pan paniscus*) challenge traditional assumptions about females, dominance, intra- and intersexual interactions, and hominid evolution. *Annals of the New York Academy of Sciences*, 907(1):97–113, 2000.

M. Patel. Parkinson disease: Csmd1 gene mutations can lead to familial parkinson disease. *Nature Reviews Neurology*, 13(11):641, 2017.

B. A. Payseur and M. W. Nachman. Microsatellite variation and recombination rate in the human genome. *Genetics*, 156(3):1285–1298, 2000.

A. Pekcec, N. Schülert, B. Stierstorfer, S. Deiana, C. Dorner-Ciossek, and H. Rosenbrock. Targeting the dopamine d1 receptor or its downstream signalling by inhibiting phosphodiesterase-1 improves cognitive performance. *British Journal of Pharmacology*, 175(14):3021–3033, 2018.

D. Pende, S. Sivori, L. Accame, L. Pareti, M. Falco, D. Geraghty, P. Le Bouteiller, L. Moretta, and A. Moretta. Hla-g recognition by human natural killer cells. involvement of cd94 both as inhibitory and as activating receptor complex. *European Journal of Immunology*, 27(8):1875–1880, 1997.

J. Prado-Martinez, P. H. Sudmant, J. M. Kidd, H. Li, J. L. Kelley, B. Lorente-Galdos, K. R. Veeramah, A. E. Woerner, T. D. OConnor, G. Santpere, et al. Great ape genetic diversity and population history. *Nature*, 499(7459):471–475, 2013.

V. Prokopovic, M. Popovic, U. Andjelkovic, A. Marsavelski, B. Raskovic, M. Gavrovic-Jankulovic, and N. Polovic. Isolation, biochemical characterization and anti-bacterial activity of BPIFA2 protein. *Archives of Oral Biology*, 59(3):302–309, 2014.

K. Prüfer, K. Munch, I. Hellmann, K. Akagi, J. R. Miller, B. Walenz, S. Koren, G. Sutton, C. Kodira, R. Winer, J. R. Knight, J. C. Mullikin, S. J. Meader, C. P. Ponting, G. Lunter, S. Higashino, A. Hobolth, J. Dutheil, E. Karako, C. Alkan, S. Sajjadian, C. R. Catacchio, M. Ventura, T. Marques-Bonet, E. E. Eichler, C. Andr, R. Atencia, L. Mugisha, J. Junhold, N. Patterson, M. Siebauer, J. M. Good, A. Fischer, S. E. Ptak, M. Lachmann, D. E. Symer, T. Mailund, M. H. Schierup, A. M. Andrs, J. Kelso, and S. Pbo. The bonobo genome compared with the chimpanzee and human genomes. *Nature*, 486(7404):527, 2012.

A. A. Puca, G. Spinetti, R. Vono, C. Vecchione, and P. Madeddu. The genetics of exceptional longevity identifies new druggable targets for vascular protection and repair. *Pharmacological Research*, 114: 169–174, 2016.

R. Rajalingam, M. Hong, E. J. Adams, B. P. Shum, L. A. Guethlein, and P. Parham. Short kir haplotypes in pygmy chimpanzee (bonobo) resemble the conserved framework of diverse human kir haplotypes. *Journal of Experimental Medicine*, 193(1):135–146, 2001.

F. Reimann, J. J. Cox, I. Belfer, L. Diatchenko, D. V. Zaykin, D. P. McHale, J. P. Drenth, F. Dai, J. Wheeler, F. Sanders, et al. Pain perception is altered by a nucleotide polymorphism in scn9a. *Proceedings of the National Academy of Sciences*, 107(11):5148–5153, 2010.

V. M. Ripoll, K. M. Irvine, T. Ravasi, M. J. Sweet, and D. A. Hume. Gpnmb is induced in macrophages by IFN-γ and lipopolysaccharide and acts as a feedback regulator of proinflammatory responses. *The Journal of Immunology*, 178(10):6557–6566, 2007.

A. G. Rosati. Heterochrony in chimpanzee and bonobo spatial memory development. *American Journal of Physical Anthropology*, 2019.

A. A. Rose, M. G. Annis, Z. Dong, F. Pepin, M. Hallett, M. Park, and P. M. Siegel. ADAM10 releases a soluble form of the GPNMB/Osteoactivin extracellular domain with angiogenic properties. *PLoS One*, 5(8):e12093, 2010.

J. Ruiz-Martínez, L. J. Azcona, A. Bergareche, J. F. Martí-Massó, and C. Paisán-Ruiz. Whole-exome sequencing associates novel csmd1 gene mutations with familial parkinson disease. *Neurology Genetics*, 3(5):e177, 2017.

A. Sanchez-Mazas, V. Černỳ, D. Di, S. Buhler, E. Podgorná, E. Chevallier, L. Brunet, S. Weber, B. Kervaire, M. Testi, et al. The hla-b landscape of africa: Signatures of pathogen-driven selection and molecular identification of candidate alleles to malaria protection. *Molecular Ecology*, 26(22):6238–6252, 2017.

A. R. Sanders, G. W. Beecham, S. Guo, K. Dawood, G. Rieger, J. A. Badner, E. S. Gershon, R. S. Krishnappa, A. B. Kolundzija, J. Duan, et al. Genome-wide association study of male sexual orientation. *Scientific Reports*, 7(1):16950, 2017.

A. Scally, J. Y. Dutheil, L. W. Hillier, G. E. Jordan, I. Goodhead, J. Herrero, A. Hobolth, T. Lappalainen, T. Mailund, T. Marques-Bonet, S. McCarthy, S. H. Montgomery, P. C. Schwalie, Y. A. Tang, M. C. Ward, Y. Xue, B. Yngvadottir, C. Alkan, L. N. Andersen, Q. Ayub, E. V. Ball, K. Beal, B. J. Bradley, Y. Chen, C. M. Clee, S. Fitzgerald, T. A. Graves, Y. Gu, P. Heath, A. Heger, E. Karakoc, A. Kolb-Kokocinski, G. K. Laird, G. Lunter, S. Meader, M. Mort, J. C. Mullikin, K. Munch, T. D. O'Connor, A. D. Phillips, J. Prado-Martinez, A. S. Rogers, S. Sajjadian, D. Schmidt, K. Shaw, J. T. Simpson, P. D. Stenson, D. J. Turner, L. Vigilant, A. J. Vilella, W. Whitener, B. Zhu, D. N. Cooper, P. de Jong, E. T. Dermitzakis, E. E. Eichler, P. Flicek, N. Goldman, N. I. Mundy, Z. Ning, D. T. Odom, C. P. Ponting, M. A. Quail, O. A. Ryder, S. M. Searle, W. C. Warren, R. K. Wilson, M. H. Schierup, J. Rogers, C. Tyler-Smith, and R. Durbin. Insights into hominid evolution from the gorilla genome sequence. *Nature*, 483(7388): 169–75, 2012.

H. Schultz and J. P. Weiss. The bactericidal/permeability-increasing protein (bpi) in infection and inflammatory disease. *Clinica Chimica Acta*, 384(1-2):12–23, 2007.

A. Serckx, H. S. Kühl, R. C. Beudels-Jamar, P. Poncin, J.-F. Bastin, and M.-C. Huynen. Feeding ecology of bonobos living in forest-savannah mosaics: Diet seasonal variation and importance of fallback foods. *American Journal of Primatology*, 77(9):948–962, 2015.

S. Sheehan and Y. S. Song. Deep learning for population genetic inference. *PLoS Computational Biology*, 12(3):e1004845, 2016.

A. Shimizu, S. Asakawa, T. Sasaki, S. Yamazaki, H. Yamagata, J. Kudoh, S. Minoshima, I. Kondo, and N. Shimizu. A novel giant gene csmd3 encoding a protein with cub and sushi multiple domains: a candidate gene for benign adult familial myoclonic epilepsy on human chromosome 8q23. 3–q24. 1. *Biochemical and Biophysical Research Communications*, 309(1):143–154, 2003.

B. P. Shum, L. R. Flodin, D. G. Muir, R. Rajalingam, S. I. Khakoo, S. Cleland, L. A. Guethlein, M. Uhrberg, and P. Parham. Conservation and variation in human and common chimpanzee cd94 and nkg2 genes. *The Journal of Immunology*, 168(1):240–252, 2002.

K. M. Siewert and B. F. Voight. Detecting long-term balancing selection using allele frequency correlation. *Molecular Biology and Evolution*, 32(11):2996–3005, 2017.

K. M. Siewert and B. F. Voight. Betascan2: Standardized statistics to detect balancing selection utilizing substitution data. *bioRxiv*, 2018. doi: 10.1101/497255.

C. Smukowski and M. Noor. Recombination rate variation in closely related species. *Heredity*, 107(6):496, 2011.

C. C. Spinelli, A. Carrizzo, A. Ferrario, F. Villa, A. Damato, M. Ambrosio, M. Madonna, G. Frati, S. Fucile, M. Sciaccaluga, et al. Lav-bpifb4 isoform modulates enos signalling through ca2+/pkc-alpha-dependent mechanism. *Cardiovascular Research*, 113(7):795–804, 2017.

G. Spinetti, E. Sangalli, C. Specchia, F. Villa, C. Spinelli, R. Pipolo, A. Carrizzo, S. Greco, C. Voellenkle, C. Vecchione, et al. The expression of the BPIFB4 and CXCR4 associates with sustained health in long-living individuals from Cilento-Italy. *Aging (Albany NY)*, 9(2):370, 2017.

V. Stepanov, A. Marusin, K. Vagaitseva, A. Bocharova, and O. Makeeva. Genetic variants in csmd1 gene are associated with cognitive performance in normal elderly population. *Genetics Research International*, 2017, 2017.

M. Surbeck, T. Deschner, G. Schubert, A. Weltring, and G. Hohmann. Mate competition, testosterone and intersexual relationships in bonobos, pan paniscus. *Animal Behaviour*, 83(3):659–669, 2012.

C. G. Sweeney, J. M. Rando, H. N. Panas, G. M. Miller, D. M. Platt, and E. J. Vallender. Convergent balancing selection on the mu-opioid receptor in primates. *Molecular Biology and Evolution*, 34(7): 1629–1643, 2017.

F. Tajima. Evolutionary relationship of dna sequences in finite populations. *Genetics*, 105(2):437–460, 1983.

F. Tajima. Statistical method for testing the neutral mutation hypothesis by dna polymorphism. *Genetics*, 123(3):585–595, 1989.

N. Takahata, Y. Satta, and J. Klein. Polymorphism and balancing selection at major histocompatibility complex loci. *Genetics*, 130(4):925–938, 1992.

J. C. Teixeira, C. de Filippo, A. Weihmann, J. R. Meneu, F. Racimo, M. Dannemann, B. Nickel, A. Fischer, M. Halbwax, C. Andre, et al. Long-term balancing selection in lad1 maintains a missense trans-species polymorphism in humans, chimpanzees, and bonobos. *Molecular Biology and Evolution*, 32(5):1186–1196, 2015.

J. Terhorst, J. A. Kamm, and Y. S. Song. Robust and scalable inference of population history from hundreds of unphased whole genomes. *Nature Genetics*, 49(2):303, 2017.

F. Ubeda and D. Haig. Sex-specific meiotic drive and selection at an imprinted locus. *Genetics*, 167(4): 2083–2095, 2004.

M. Uhlén, L. Fagerberg, B. M. Hallström, C. Lindskog, P. Oksvold, A. Mardinoglu, Å. Sivertsson, C. Kampf, E. Sjöstedt, A. Asplund, et al. Tissue-based map of the human proteome. *Science*, 347 (6220):1260419, 2015.

S. Van Dooren, W. M. Switzer, W. Heneine, P. Goubau, E. Verschoor, B. Parekh, W. De Meurichy, C. Furley, M. Van Ranst, and A.-M. Vandamme. Lack of evidence for infection with simian immunodeficiency virus in bonobos. *AIDS Research and Human Retroviruses*, 18(3):213–216, 2002.

V. Vasta, W. K. Sonnenburg, C. Yan, S. H. Soderling, M. Shimizu-Albergine, and J. A. Beavo. Identification of a new variant of pde1a calmodulin-stimulated cyclic nucleotide phosphodiesterase expressed in mouse sperm. *Biology of Reproduction*, 73(4):598–609, 2005.

F. Villa, A. Carrizzo, C. C. Spinelli, A. Ferrario, A. Malovini, A. Maciag, A. Damato, A. Auricchio, G. Spinetti, E. Sangalli, et al. Genetic analysis reveals a longevity-associated protein modulating endothelial function and angiogenesis. *Circulation Research*, 117(4):333–345, 2015a.

F. Villa, A. Malovini, A. Carrizzo, C. C. Spinelli, A. Ferrario, A. Maciag, M. Madonna, R. Bellazzi, L. Milanesi, C. Vecchione, et al. Serum BPIFB4 levels classify health status in long-living individuals. *Immunity & Ageing*, 12(1):27, 2015b.

F. Villa, E. Ciaglia, A. Maciag, F. Montella, A. Ferrario, M. Cattaneo, and A. Puca. Longevity associated variant of *BPIFB4* mitigates monocyte mediated acquired immune response. *Innovation in Aging*, 2 (Suppl 1):884, 2018.

L. Walter. Immunogenetics of nk cell receptors and mhc class i ligands in non-human primates. In *Natural Hosts of SIV*, pages 269–285. Elsevier, 2014.

G. A. Watterson. On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, 7(2):256–276, 1975.

L. Weiflog, C.-J. Scholz, C. P. Jacob, T. T. Nguyen, K. Zamzow, S. Gro-Lesch, T. J. Renner, M. Romanos, D. Rujescu, S. Walitza, S. Kneitz, K.-P. Lesch, and A. Reif. Kcnip4 as a candidate gene for personality disorders and adult adhd. *European Neuropsychopharmacology*, 23(6):436 – 447, 2013. ISSN 0924-977X. Making IMPACT on adult ADHD recent advances in neurobiology and treatment.

V. Wobber, B. Hare, J. Maboto, S. Lipson, R. Wrangham, and P. T. Ellison. Differential changes in steroid hormones before competition in bonobos and chimpanzees. *Proceedings of the National Academy of Sciences*, 107(28):12457–12462, 2010a.

V. Wobber, R. Wrangham, and B. Hare. Bonobos exhibit delayed development of social behavior and cognition relative to chimpanzees. *Current Biology*, 20(3):226–230, 2010b.

R. W. Wrangham. The evolution of sexuality in chimpanzees and bonobos. *Human Nature*, 4(1):47–79, Mar 1993. ISSN 1936-4776.

E. E. Wroblewski, L. A. Guethlein, P. J. Norman, Y. Li, C. M. Shaw, A. S. Han, J.-B. N. Ndjango, S. Ahuka-Mundeke, A. V. Georgiev, M. Peeters, et al. Bonobos maintain immune system diversity with three functional types of mhc-b. *The Journal of Immunology*, 198(9):3480–3493, 2017.

E. E. Wroblewski, P. Parham, and L. A. Guethlein. Two to tango: Co-evolution of hominid natural killer cell receptors and mhc. *Frontiers in Immunology*, 10, 2019.

W. Xu, S. Cohen-Woods, Q. Chen, A. Noor, J. Knight, G. Hosang, S. V. Parikh, V. De Luca, F. Tozzi, P. Muglia, et al. Genome-wide association study of bipolar disorder in canadian and uk populations corroborates disease loci including syne1 and csmd1. *BMC Medical Genetics*, 15(1):2, 2014.

Y. Xu, H.-T. Zhang, and J. M. ODonnell. Phosphodiesterases in the central nervous system: implications in mood and cognitive disorders. In *Phosphodiesterases as Drug Targets*, pages 447–485. Springer, 2011.

C. Yan, J. K. Bentley, W. K. Sonnenburg, and J. A. Beavo. Differential expression of the 61 kda and 63 kda calmodulin-dependent phosphodiesterases in the mouse brain. *Journal of Neuroscience*, 14(3): 973–984, 1994.

Y. Yang, Y. Wang, S. Li, Z. Xu, H. Li, L. Ma, J. Fan, D. Bu, B. Liu, Z. Fan, et al. Mutations in scn9a, encoding a sodium channel alpha subunit, in patients with primary erythermalgia. *Journal of Medical Genetics*, 41(3):171–174, 2004.

C. J. Ye, J. Chen, A.-C. Villani, R. E. Gate, M. Subramaniam, T. Bhangale, M. N. Lee, T. Raj, R. Raychowdhury, W. Li, et al. Genetic analysis of isoform usage in the human anti-viral response reveals influenza-specific regulation of erap2 transcripts under balancing selection. *Genome Research*, 28(12): 1812–1825, 2018.

B. Yu, G. R. Sondag, C. Malcuit, M.-H. Kim, and F. F. Safadi. Macrophage-associated Osteoactivin/GPNMB mediates mesenchymal stem cell survival, proliferation, and migration via a cd44-dependent mechanism. *Journal of Cellular Biochemistry*, 117(7):1511–1521, 2016.

L. Zhou, F. Liu, Y. Li, Y. Peng, Y. Liu, and J. Li. Gpnmb/osteoactivin, an attractive target in cancer immunotherapy. *Neoplasma*, 59(1):1–5, 2012.

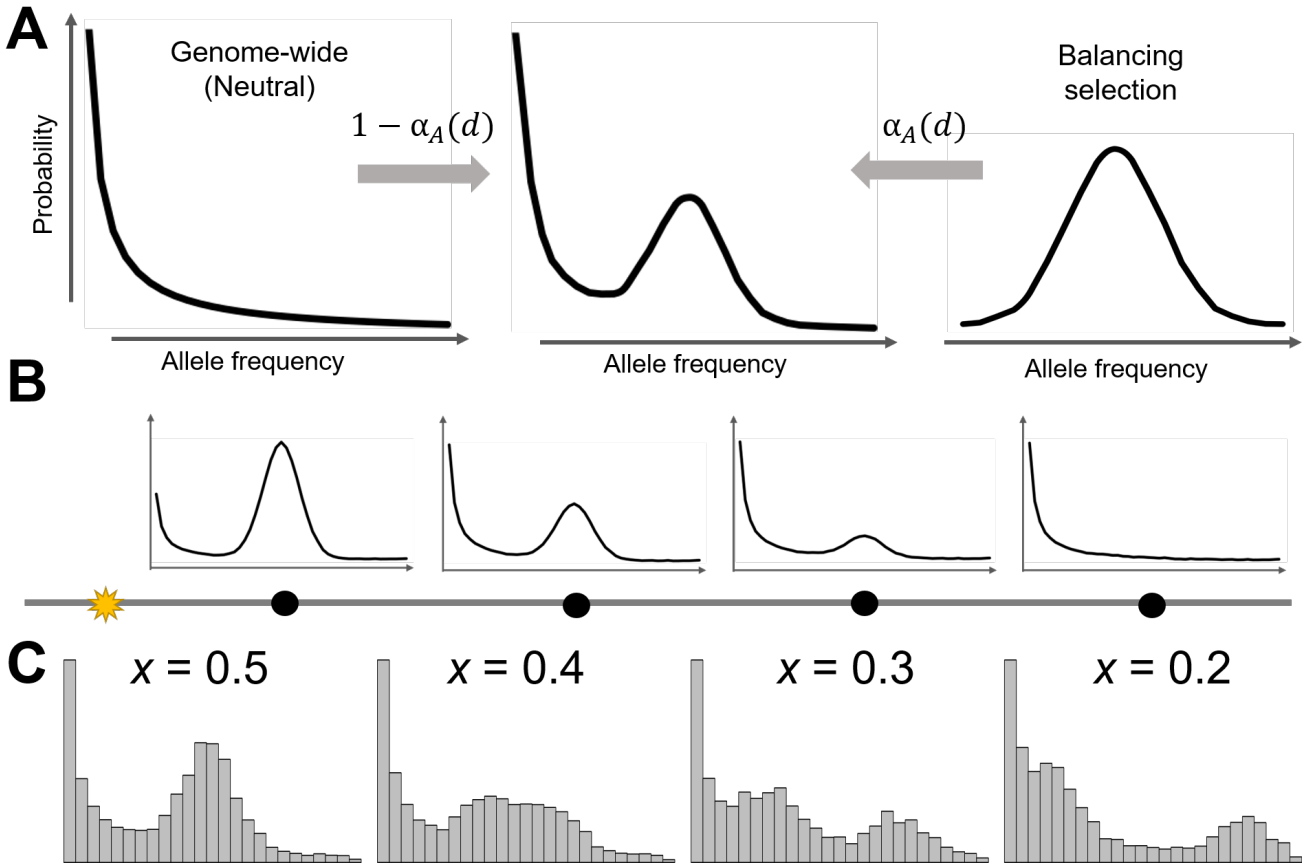

Figure 1: Schematic of the mixture model underlying the B statistics. (A) The model for the alternative hypothesis is a mixture of the distribution of allele frequencies under balancing selection at proportion $\alpha_A(d)$, modeled by either a binomial or beta-binomial distribution, and the distribution under neutrality at proportion $1 - \alpha_A(d)$, modeled by the genome-wide site frequency spectrum. Here, $\alpha_A(d)$ decays as a function of recombination distance d, and so sites close to (*i.e*, small d) the putative selected site will be modeled mostly by the distribution expected under balancing selection, whereas sites far from (*i.e.*, large d) the selected site will be modeled mostly by the distribution expected under neutrality. (B) Distributions of allele frequencies at neutral sites (black dots) under the mixture model at varying distances d from the putative selected site (yellow star). (C) Distributions of allele frequencies from simulated sequences when balancing selection maintains the equilibrium frequency of $x = 0.2$, 0.3, 0.4, or 0.5.

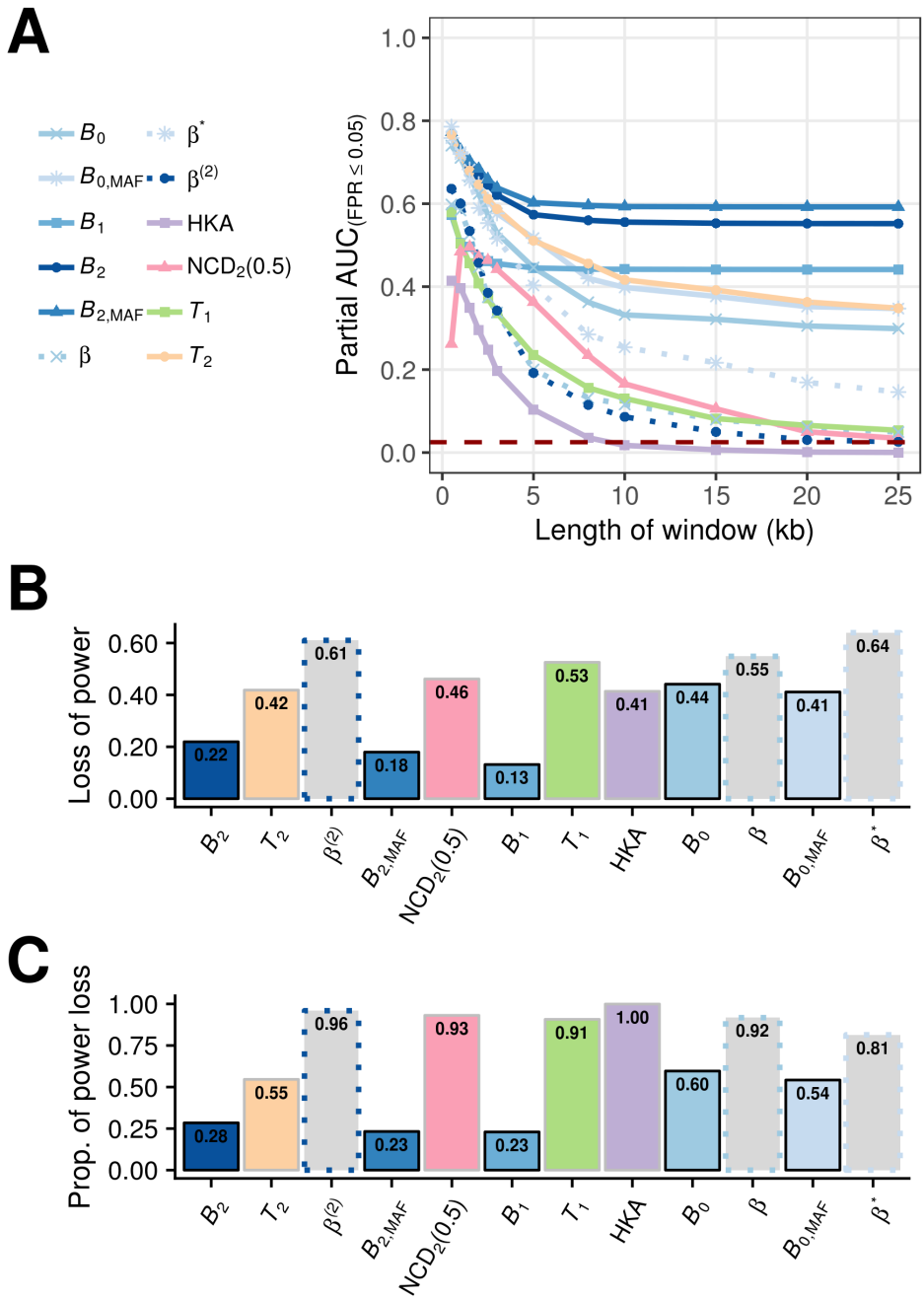

Figure 2: Partial area under the curve (AUC) conditioned on false positive rates (FPRs) less than or equal to 5% (defined such that the maximal value is 1) as a function of window size measured in kilobases (kb) for B statistics (varying shades of blue), β statistics (dotted line with varying shades of blue), T_2 (orange), T_1 (green), HKA (purple), and $\text{NCD}_2(0.5)$ (pink), under a scenario in which a mutation undergoing ancient balancing selection (selective coefficient $s = 0.01$ and dominance coefficient $h = 20$) arose 15 million years ago (assuming a generation time of 20 years). Statistics that consider the same input type share the same point shape. The dark red dashed line marks the level of partial AUC expected at the $y{=}x$ line, or the baseline of randomly choosing between balancing selection and neutrality. (B) The amount of partial AUC lost, and (C) the proportion of the AUC loss as compared with the optimal value for each statistic when the window size increased from the optimum to 25 kb (*e.g.*, largest evaluated).

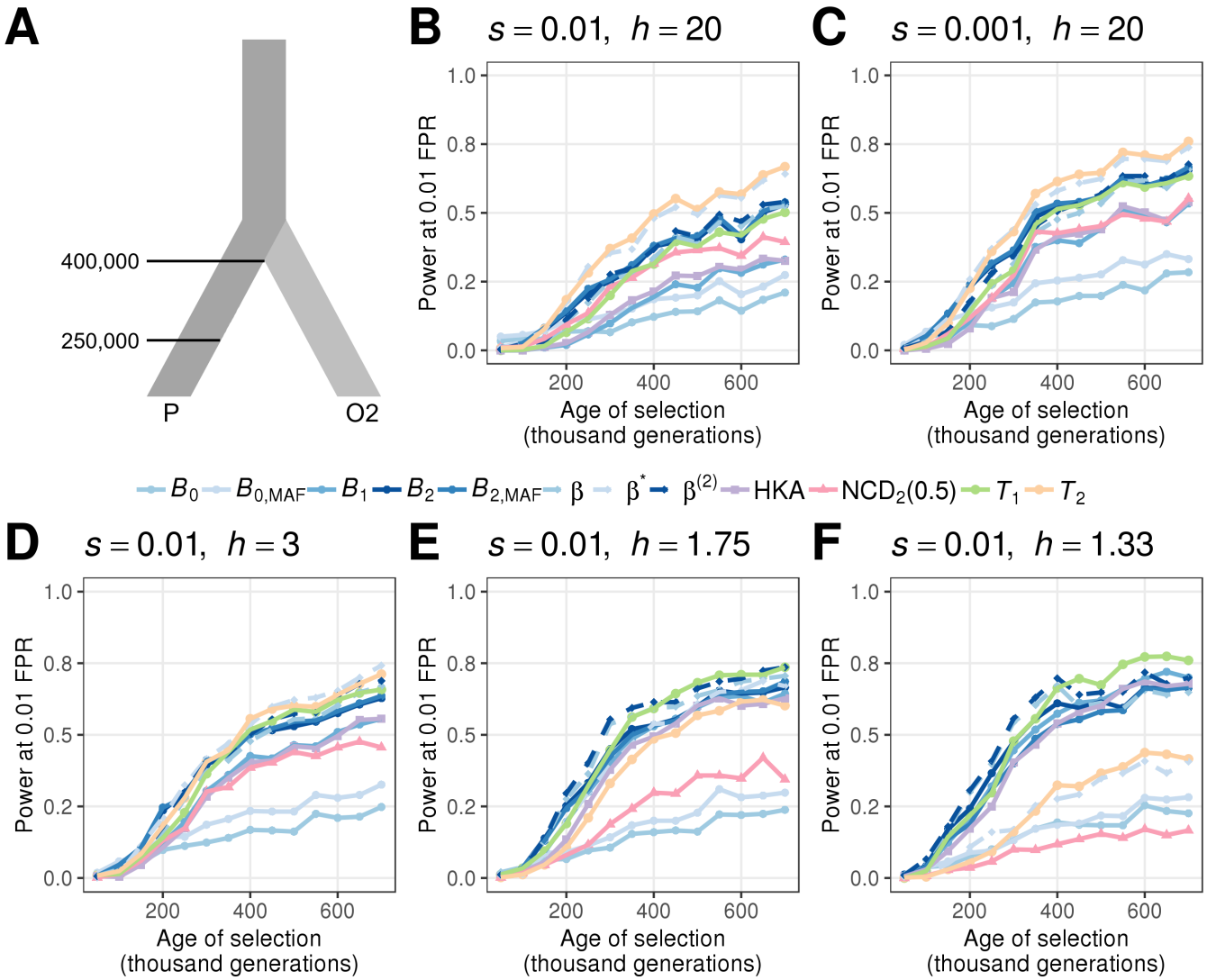

Figure 3: Ability to detect balancing selection for different heterozygote advantage scenarios. (A) Demographic model relating the ingroup (P) and outgroup (O2) populations, with one sample from O2 used as the outgroup sequence. (B-F) Powers at a 1% false positive rate (FPR) for each statistic as a function of age of the allele undergoing balancing selection for different selection (s) and dominance (h) coefficients. The scenarios considered are (B) $s = 0.01$ with $h = 20$, (C) $s = 0.001$ with $h = 20$, (D) $s = 0.01$ with $h = 3$, (E) $s = 0.01$ with $h = 1.75$, and (F) $s = 0.01$ with $h = 1.33$. Note that the equilibrium frequencies for panels D, E, and F are 0.4, 0.3, and 0.2, respectively.

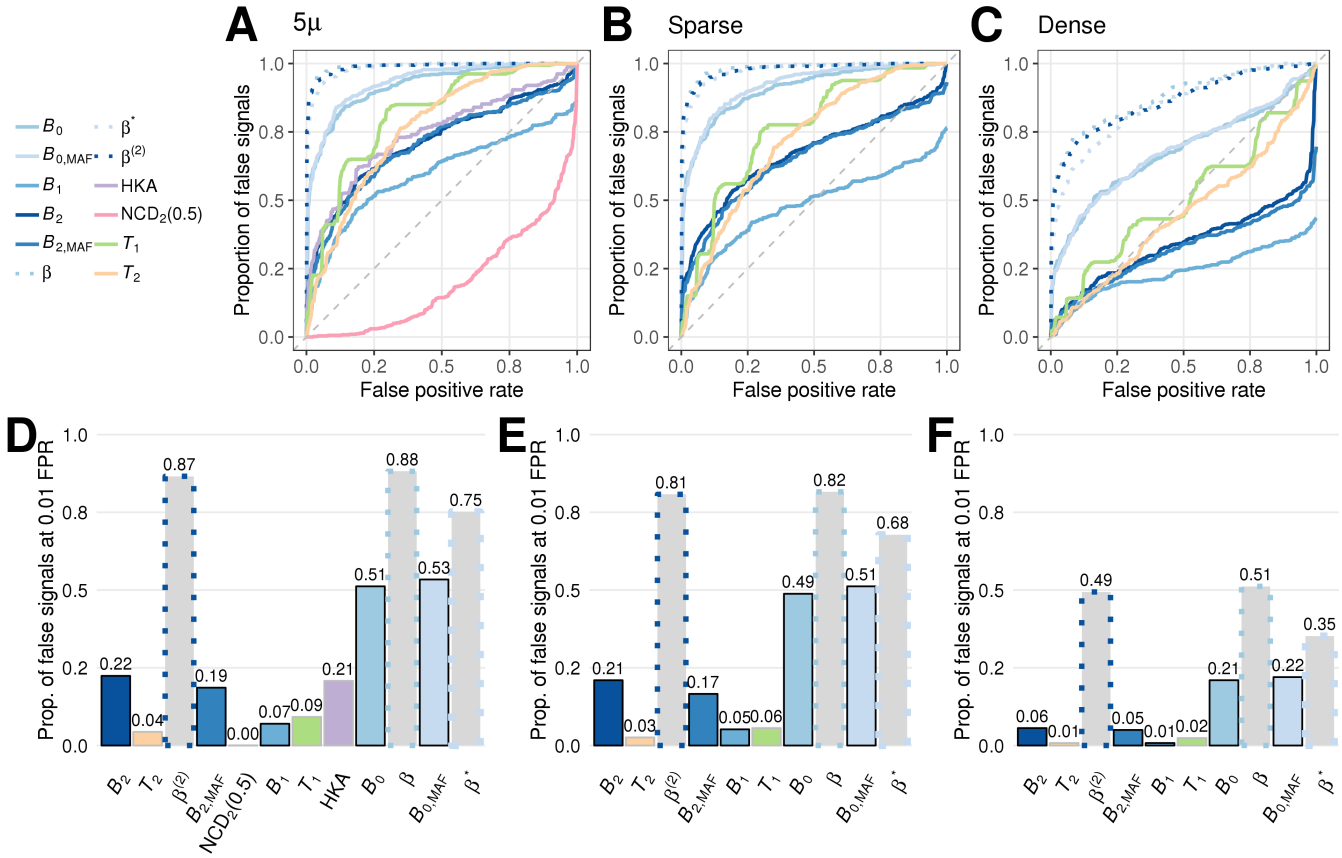

Figure 4: Robustness of each statistic evaluated when the mutation rate is elevated five-fold across an extended genomic region. (A-C) Proportions of false signals as a function of false positive rate (FPR) when (A) all test sites were considered to compute proportions of false signals and when (B, C) the number of test sites were down-sampled to match the number of sites expected on sequences with the original mutation rate μ. In panel B, five informative sites flanking either side of a test site considered were skipped ("sparse" down-sampling), whereas in panel C only the first 1200 informative sites were considered for calculating T, B_2, and B_1 statistics, and the first 240 polymorphic sites were considered for B_0 and β statistics ("dense" down-sampling). (D-F) Proportions of false signals at a 1% FPR for each statistic for (D) all test sites, (E) sparse down-sampling, and (F) dense down-sampling. Bars for all B statistics are bordered by black lines, and all β statistics are bordered by gray lines in panels D-F. HKA and $\mathrm{NCD}_2(0.5)$ are applied with a fixed step size, and are therefore not considered when down-sampling in panels B, C, E, and F.

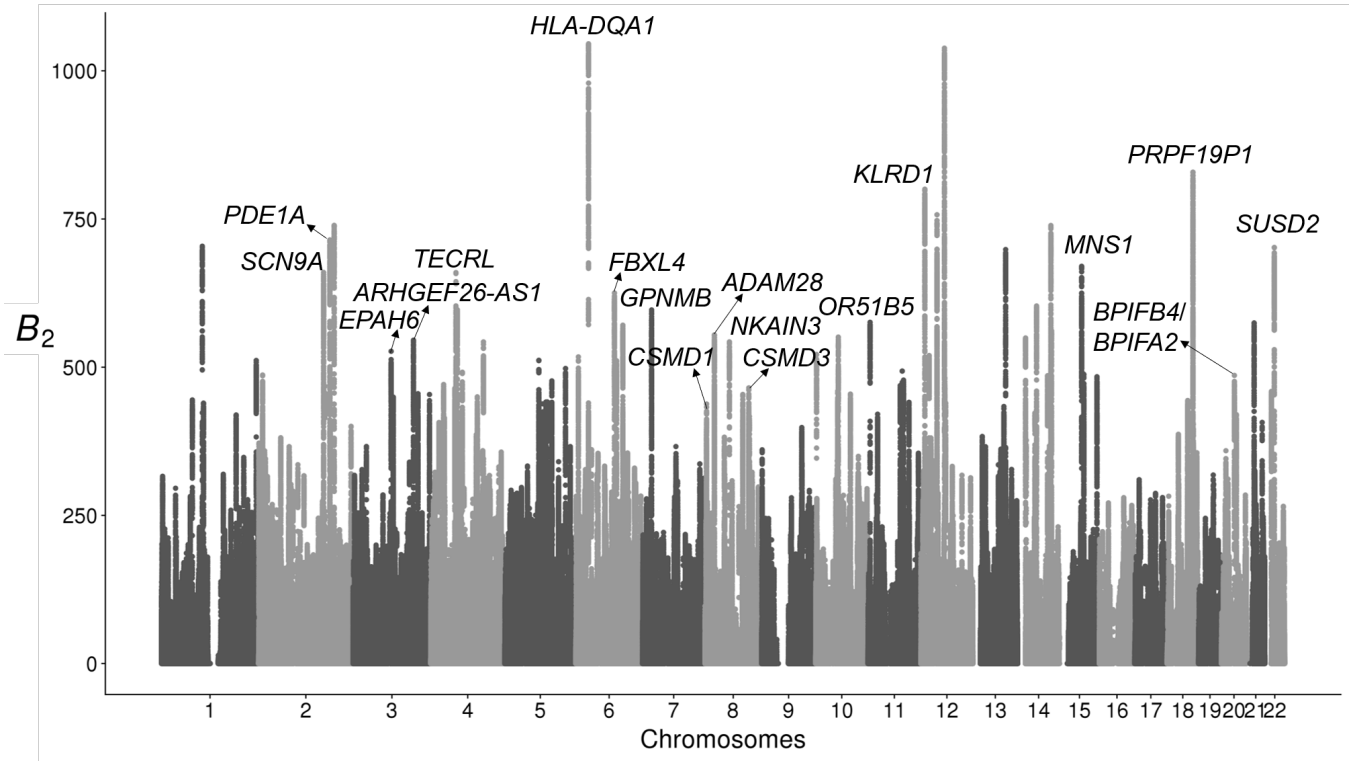

Figure 5: Manhattan plot displaying B_2 scores across the 22 human autosomes for which the genomic data were mapped, with the top candidates annotated. Outstanding peaks ($B_2 > 600$) without annotations do not have neighboring protein-coding regions within a 500 kb radius.

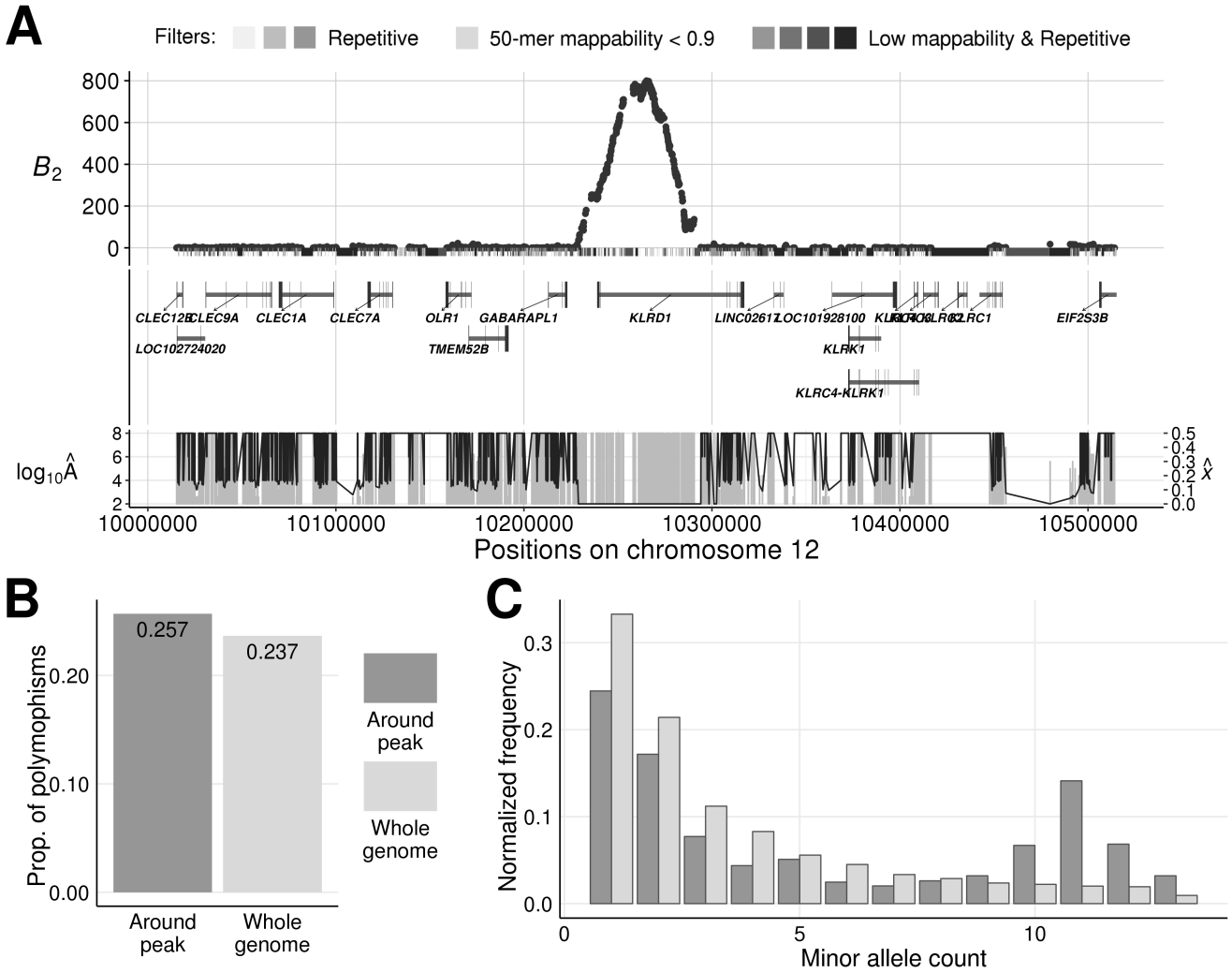

Figure 6: Evidence of balancing selection on *KLRD1*. (A) B_2 scores across the 500 kb genomic region on chromosome 12 centered on the peak. The gray bars directly under the B_2 scores represent the masked regions, as well as the features in these regions. The darker the shade, the greater number of types of repetitive sequences (*e.g.*, RepeatMasker mask, segmental duplication, simple repeats, or interrupted repeats) overlapping the region. Vertical gray bars below display the estimated equilibrium minor allele frequency $\widehat{x}$ for each maximum likelihood ratio B_2, and the black line traces the value for the respective inferred footprint size $\log_{10}(\widehat{A})$. (B) Proportion of informative sites that are polymorphic in the 500 kb region centered on the peak compared with the whole-genome average. (C) Minor allele frequency distribution in the 500 kb region centered on the peak compared with the whole-genome average.

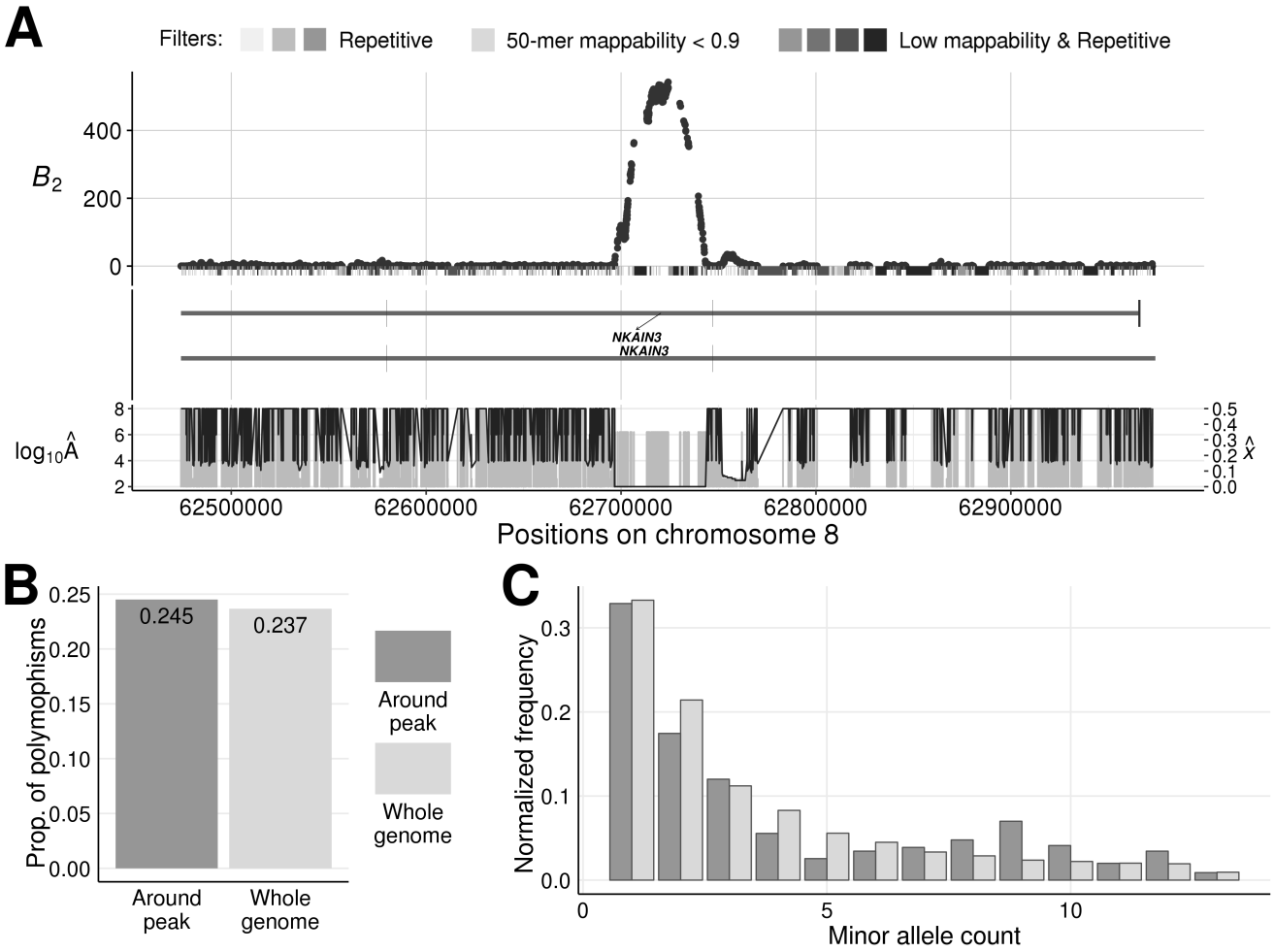

Figure 7: Evidence of balancing selection on *NKAIN3*. (A) B_2 scores across the 500 kb genomic region on chromosome 8 centered on the peak. The gray bars directly under the B_2 scores represent the masked regions, as well as the features in these regions. The darker the shade, the greater number of types of repetitive sequences (*e.g.*, RepeatMasker mask, segmental duplication, simple repeats, or interrupted repeats) overlapping the region. Vertical gray bars below display the estimated equilibrium minor allele frequency $\widehat{x}$ for each maximum likelihood ratio B_2, and the black line traces the value for the respective inferred footprint size $\log_{10}(\widehat{A})$. (B) Proportion of informative sites that are polymorphic in the 500 kb region centered on the peak compared with the whole-genome average. (C) Minor allele frequency distribution in the 500 kb region centered on the peak compared with the whole-genome average.