Transition from Background Selection to Associative Overdominance Promotes Diversity in Regions of Low Recombination

Kimberly J Gilbert, Fanny Pouyet, Laurent Excoffier, Stephan Peischl

Report

Current Biology

Transition from Background Selection to Associative Overdominance Promotes Diversity in Regions of Low Recombination

Highlights

- AOD (associative overdominance) increases diversity in low-recombination regions
- The strongest AOD is driven by purging recessive mutations of intermediate effects
- AOD may be confounded with balancing selection in low-recombination regions
- 22 regions in the human genome are identified as showing signatures of AOD

Authors

Kimberly J. Gilbert, Fanny Pouyet, Laurent Excoffier, Stephan Peischl

Correspondence

kimberly.gilbert@unil.ch

In Brief

Gilbert, Pouyet et al. find that selection against recessive deleterious mutations in regions of low recombination can lead to elevated levels of diversity, counter to the actions of background selection (BGS), and mimic balancing selection. We derive the conditions for this to occur and identify potential exemplary regions in the human genome.

Gilbert et al., 2020, Current Biology *30*, 101–107
January 6, 2020
https://doi.org/10.1016/j.cub.2019.11.063

CellPress

Current Biology
Report

Transition from Background Selection to Associative Overdominance Promotes Diversity in Regions of Low Recombination

Kimberly J. Gilbert,[1,2,4,6,8,*] Fanny Pouyet,[1,2,5,6] Laurent Excoffier,[1,2,7] and Stephan Peischl[2,3,7]
[1]Institute of Ecology and Evolution, Baltzerstrasse 6, University of Bern, 3012 Bern, Switzerland
[2]Swiss Institute of Bioinformatics, Quartier Sorge-Batiment Amphipole, 1015 Lausanne, Switzerland
[3]Interfaculty Bioinformatics Unit, Baltzerstrasse 6, University of Bern, 3012 Bern, Switzerland
[4]Present address: Department of Computational Biology, University of Lausanne (UNIL), Génopode 2016, 1015 Lausanne, Switzerland
[5]Present address: Sorbonne Université, CNRS, Institut de Biologie Paris-Seine, Laboratory of Computational and Quantitative Biology, 7-9 Quai Saint Bernard, 75005 Paris, France
[6]These authors contributed equally
[7]Senior author
[8]Lead Contact
*Correspondence: kimberly.gilbert@unil.ch
https://doi.org/10.1016/j.cub.2019.11.063

SUMMARY

Linked selection is a major driver of genetic diversity. Selection against deleterious mutations removes linked neutral diversity (background selection [BGS]) [1], creating a positive correlation between recombination rates and genetic diversity. Purifying selection against recessive variants, however, can also lead to associative overdominance (AOD) [2, 3], due to an apparent heterozygote advantage at linked neutral loci that opposes the loss of neutral diversity by BGS. Zhao and Charlesworth [3] identified the conditions under which AOD should dominate over BGS in a single-locus model and suggested that the effect of AOD could become stronger if multiple linked deleterious variants co-segregate. We present a model describing how and under which conditions multi-locus dynamics can amplify the effects of AOD. We derive the conditions for a transition from BGS to AOD due to pseudo-overdominance [4], i.e., a form of balancing selection that maintains complementary deleterious haplotypes that mask the effect of recessive deleterious mutations. Simulations confirm these findings and show that multi-locus AOD can increase diversity in low-recombination regions much more strongly than previously appreciated. While BGS is known to drive genome-wide diversity in humans [5], the observation of a resurgence of genetic diversity in regions of very low recombination is indicative of AOD. We identify 22 such regions in the human genome consistent with multi-locus AOD. Our results demonstrate that AOD may play an important role in the evolution of low-recombination regions of many species.

RESULTS AND DISCUSSION

The interaction of recombination with selective and non-selective processes impacts both functional and neutral diversity across the genome, where a positive correlation between genetic diversity and recombination is generally expected [1, 6]. Yet, Zhao and Charlesworth [3] have recently identified a possible transition from BGS to associative overdominance (AOD) at sites tightly linked to a locus with recessive deleterious mutations, leading to a switch from reducing to increasing diversity in genomic regions of very low recombination. This observation is consistent with empirical analyses in *Drosophila simulans* [7]. However, the prevalence of AOD and its impact on genomic diversity in natural populations is unknown, especially in a genomic context where multiple linked recessive loci co-segregate in regions of low recombination. Here, we investigate the evolutionary dynamics of multi-locus AOD through a combination of simulation and mathematical modeling that allows us to identify empirical signatures of AOD in humans.

Simulations of Linked Selection

Evidence of BGS reducing diversity in regions of low recombination has been observed across many taxa, including *Drosophila* [8–12], humans [13], white-throated sparrows [14], tomatoes [15], and maize [16, 17]. We performed genome-scale simulations to understand how selection (Ns), dominance of deleterious alleles (h), and recombination rate interact to shape patterns of genetic diversity at neutral sites, such as nucleotide diversity (π) and the average derived allele frequency within individuals, computed for polymorphic sites defined over all populations in the sample ($\overline{DAF_i}$) [5]. $\overline{DAF_i}$ is insensitive to differential demography between populations, unlike conventional DAF where frequencies are only computed over sites polymorphic within a given population (see STAR Methods for the exact computation). For co-dominant mutations, our simulations qualitatively agree with the classical BGS result that linked neutral diversity decreases with decreasing recombination (Figures 1A and 1B; see e.g., Equation 8 in [18]).

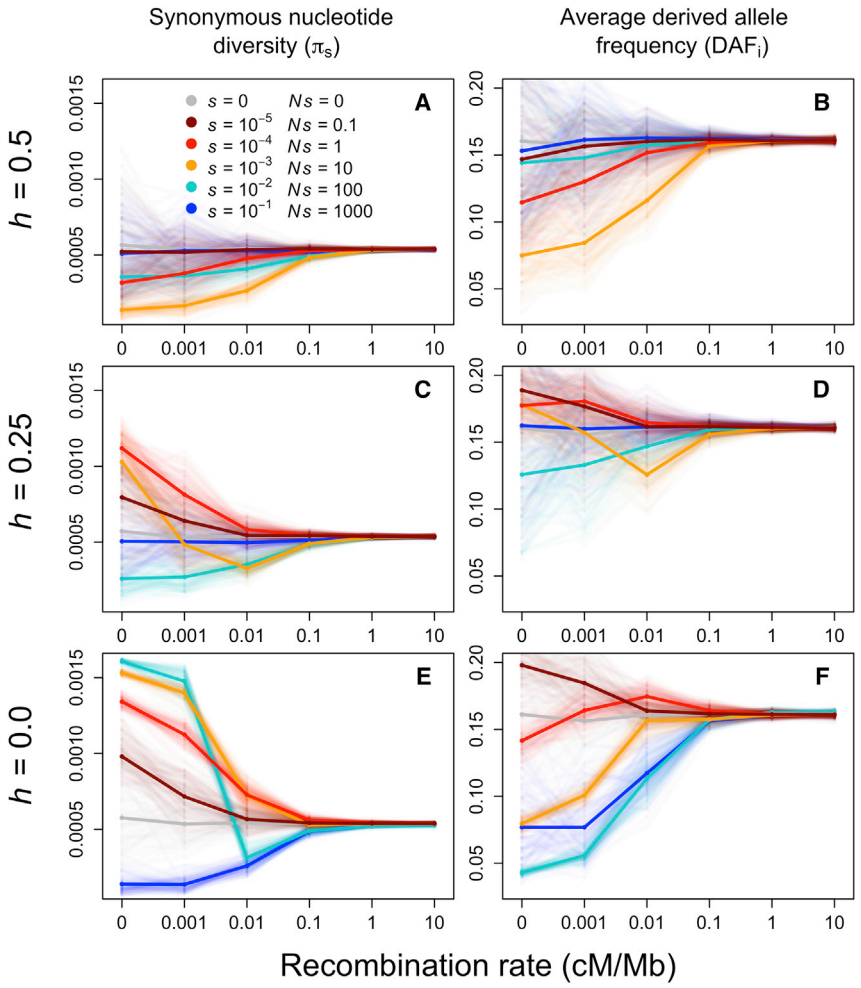

Figure 1. Genomic Diversity at Neutral Sites as Measured by Nucleotide Diversity (π) and Derived Allele Frequency ($\overline{DAF_i}$) from Simulations across Cases of Selection Coefficients (s), Dominance of Deleterious Alleles (h), and Recombination Rate

This and all main text results are for a scenario where 2% of all incoming mutations are deleterious. Means across all replicates are shown in thick lines while all 100 replicates are transparent lines in the background.

See also Figure S4 for results from a higher mutation rate.

In agreement with theoretical expectations (e.g., [19]), BGS has minimal effects on neutral diversity around nearly neutral or strongly deleterious variants (Figures 1A and 1B), because selection is either so weak (Ns = 0.1) that it has little impact on fitness, or so strong (Ns = 1,000) that mutations are rapidly purged before diversity can be reduced (Figure S1).

In the presence of partially (h = 0.25) and fully (h = 0) recessive deleterious alleles, neutral nucleotide diversity (π) exceeds neutral expectation at lower recombination rates (Figures 1C and 1E), except when selection is very strong (Ns = 1,000). In some cases, we observe that nucleotide diversity decreases and then rebounds as recombination decreases (e.g., Ns = 10, orange lines in Figure 1C), potentially reflecting a transition from BGS to AOD. The purging of homozygous genotypes in regions of low recombination is expected to only slightly increase diversity (on the order of a few percent) and be limited to small selection coefficients ($Ns \lesssim 1$ [3], but see [7]). In our simulations, however, we often see more than a 2-fold increase in diversity, even for selection coefficients where single-locus theory predicts a reduction of diversity due to BGS (e.g., for Ns = 100 in Figure 1E). Our multi-locus results thus extend the finding of Zhao and Charlesworth [3] that two tightly linked loci under selection show a stronger impact of AOD over a wider range of selective strengths than predicted by single-locus theory, suggestive of a synergistic effect between linked deleterious sites.

We gain further insight into the potential signal of AOD by contrasting π with $\overline{DAF_i}$. Nucleotide diversity captures the amount of heterozygosity in the population whereas $\overline{DAF_i}$ measures allele frequencies across all polymorphic loci in the sample, and thus the presence of rare heterozygotes affects these two quantities in contrasting ways. We find that π and $\overline{DAF_i}$ respond differently to a decrease in recombination rate for various selection intensities (Figures 1C–1F). For example, in Figure 1E, the largest π and thus greatest impact of AOD is obtained when Ns = 100, h = 0, and no recombination, yet this is equivalently the lowest respective $\overline{DAF_i}$ measure across simulated cases. Contrasting $\overline{DAF_i}$ and π thus suggests that AOD can lead to maintenance of high levels of diversity via many polymorphic sites at either low frequencies (high π, low $\overline{DAF_i}$, e.g., Ns = 100 in Figures 1E and 1F), or high frequencies (high π, high $\overline{DAF_i}$, e.g., Ns = 0.1 in Figures 1E and 1F).

In agreement with this explanation, the site frequency spectra (SFS) show peaks of intermediate frequency variants (Figure 2), with most pronounced peaks for intermediate selection coefficients. As selection coefficients increase, peaks shift to lower classes of the SFS and become more pronounced (Figures 2C and 2D). Clustering analyses suggest that the increased heterozygosity is caused by maintenance of complementary deleterious haplotypes in the population (Figures S1A–S1C). Thus, higher selection coefficients imply a larger number of rare complementary haplotypes across loci and therefore higher π but lower $\overline{DAF_i}$ values. For smaller selection coefficients, both π and $\overline{DAF_i}$ are high because complementary haplotypes are maintained in the population at higher frequencies (Figures 1E, 1F, 2A, and 2B). Meanwhile under our strongest selection case (Ns = 1,000), AOD is not observed, as we explain in the following model.

Modeling the Transition from BGS to AOD

To better understand the action of linked selection on neutral variants, we developed an analytical model of the multi-locus

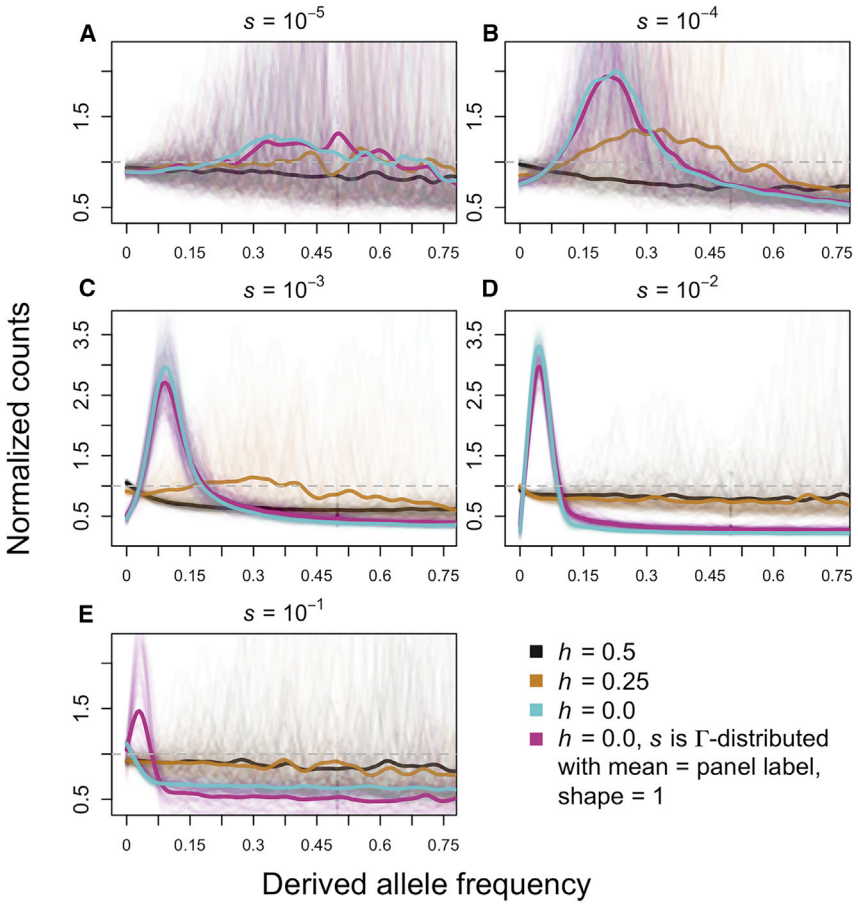

Figure 2. Site Frequency Spectra for Neutral Normalized by Dividing Each Class by Its Expectation in a Stationary Population, Indicated by a Dashed Gray Line, across the Whole Genome (All Recombination Rates)

Data are smoothed with a spline function, average SFS across all SLiM simulations are shown in thick solid lines, and individual replicates are transparent lines in background. The x axis shows only the classes below frequency 0.75 to focus on the changes in peaks of heterozygosity.

See also Figure S1 for haplotype clustering that generates observed heterozygosity.

dynamics of mutation, recombination, and selection acting on fully recessive variants. Our goal is to find the conditions for a transition from BGS to selection maintaining complementary haplotypes. This complementation arises as selection favors multi-locus heterozygotes over individuals homozygous for recessive variants (pseudo-overdominance [4]), creating multi-locus AOD at nearby neutral sites. Pseudo-overdominance occurs only when the wild-type haplotype without any deleterious mutations is lost from the population, that is, when every individual carries at least one deleterious allele and heterozygotes are fitter than any of the homozygotes [20–22]. Frequent recombination would re-create the wild-type haplotype, and we thus expect a transition from BGS to pseudo-overdominance and strong AOD only when recombination is sufficiently weak.

We consider a region with map length r that harbors n equidistant sites at which deleterious mutations occur at rate u per site. We assume that deleterious mutations are fully recessive (additional results for partially recessive variants can be found in Data S1). In Data S1, we show that in an (effectively infinitely) large population the frequency of the wild-type class at mutation-selection balance can be approximated by $P_0 \approx 1 - n\sqrt{\mu} + ((n+1)n/12)\sqrt{\mu}\,\rho$, where $\mu = (u/s)$ and $\rho = (r/s)$ are the scaled strengths of mutation and recombination, respectively. We use this result to predict the conditions for loss of the wild-type haplotype (that is, when $P_0 = 0$) and hence a transition from purifying selection to pseudo-overdominance, which fits well to multi-locus simulations (Figure 3). Figure 3A compares the analytical predictions with 2-locus simulations and shows that the wild-type haplotype is lost if μ is high and ρ is low. Our model fits well with simulations, unless mutation and recombination rates are large. This makes sense, because our derivations assume that haplotypes carry at most a single deleterious mutation at mutation-selection balance, which is only the case if mutation rates and recombination rates are sufficiently low.

Increasing the number of deleterious loci while keeping the per-locus mutation rate constant increases the probability of losing the wild-type haplotype, because the total deleterious mutation rate ($\mu \times n$) increases and because heterozygosity increases if deleterious alleles are scattered across multiple loci, masking them from selection (Figure 3B). Loss of the wild-type haplotype is most likely for intermediate selection coefficients (Figure 3C), explaining why we observe the strongest increase in diversity for intermediate selection coefficients (Figure 1E). When mutations are only slightly deleterious, recombination is too strong relative to selection and the wild-type haplotype is unlikely to be lost (large ρ in Figure 3A). Alternatively, if selection is very strong relative to the mutation rate, haplotypes carrying a deleterious allele will be purged quickly and loss of the wild-type haplotype is unlikely (small μ in Figure 3A), explaining our observation of strictly decreasing diversity and $\overline{DAF_i}$ under the strongest selection coefficient simulated (Figures 1E and 1F, dark blue lines). Once the wild-type haplotype is lost, complementary haplotypes will be maintained by pseudo-overdominance (see Figure S2), and our model predicts that the equilibrium frequency of these haplotypes decreases with increasing selection coefficients. Our model predictions closely match our simulations: nucleotide diversity will generally be high under multi-locus AOD due to pseudo-overdominance that increases heterozygosity. $\overline{DAF_i}$ will, however, only be large if selection coefficients are small enough because the frequency of complementary haplotypes, and hence the per-locus derived allele frequency, decreases with increasing selection intensity. Our analytical results also explain why a strong increase in heterozygosity is not

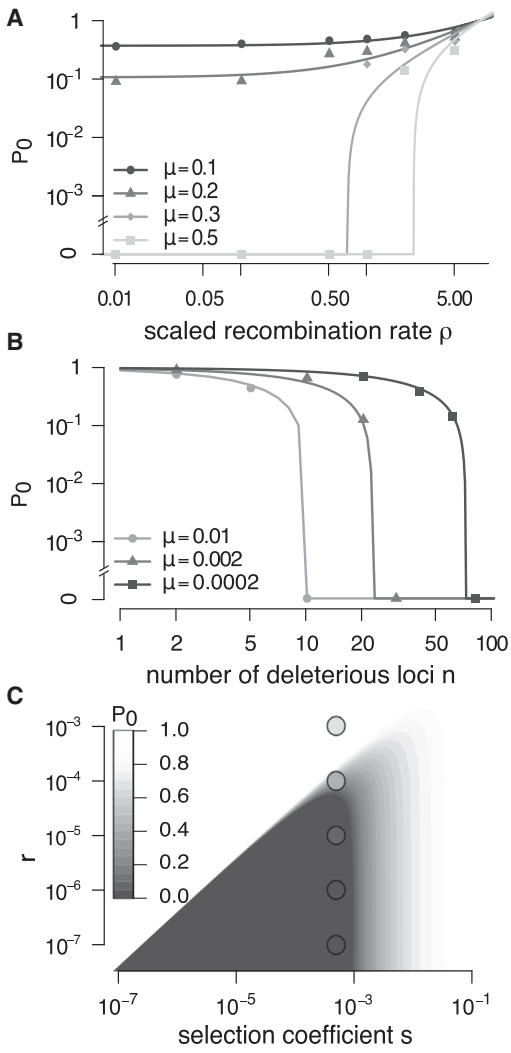

Figure 3. Conditions for a Transition from BGS to AOD

Solid lines show the approximate analytical solution for P_0 at equilibrium, and dots show results from simulations.

(A) Comparison of analytical prediction with results from individual-based simulations for a two-locus model with parameters $s = 0.001$, $h = 0$, $N = 10{,}000$ diploid individuals, and μ and ρ as specified in the figure.

(B) Average frequency of the wild-type haplotype without recombination ($\rho = 0$) as a function of the number of deleterious loci, n, with μ as specified in the figure.

(C) Combinations of r and s for which a transition from BGS to AOD is predicted. Parameters are $\mu = 10^{-6}$ and number of loci $n = 30$. Colored areas show the predicted equilibrium frequency of the wild-type haplotype. White areas correspond to parameter combinations where our approximation no longer holds (that is, where $P_0 > 1$ is predicted). Filled circles indicate results from individual-based simulations with $N = 10{,}000$ diploid individuals. The parameter r measures the recombination rate between the first and last deleterious loci of the simulated region.

See also Figure S2 and Data S1.

necessarily linked to an increase in $\overline{DAF_i}$ if selection is strong (e.g., $Ns = 100$ in Figures 1E and 1F).

Empirical Inference of AOD

Our theoretical results indicate that multi-locus AOD generally increases π in regions of low recombination. Based on this result, we scanned 100 human genomes from ten 1,000G populations to find regions subject to multi-locus AOD (see STAR Methods for details). Overall, we identified 22 regions (38 2-Mb windows, some partially overlapping) where for each population, the difference in π between low and medium recombination rates exceeds two SDs from the chromosomal mean of that population (Data S2A; Table S1). Interestingly, these 22 regions overlap with 355 out of 6,225 previously identified and hypothesized to be under long-term balancing selection based on a statistic extracting information from the SFS [23]. Distinguishing "classical" balancing selection (single-locus overdominance) from pseudo-overdominance is difficult as both processes leave similar signatures in low-recombination regions (Figure S3). Strong cases of AOD, e.g., our instances of $h = 0$ and $Ns = 1$–100, can be distinguished from balancing selection due to the combination of higher π and lower $\overline{DAF_i}$; this is not observed in our human data (Table S1), but this signal could be useful to detect AOD in other species. Our scan of the human genome detects the major histocompatibility complex (MHC) as a candidate region for AOD (Figure 4A), a region known to be under long-term balancing selection and not subject to pseudo-overdominance [25], in line with the possibility that some of our regions detected as subject to AOD could be cases of classical overdominance.

We identify several promising candidate regions for multi-locus AOD, two of which we focus on in Figures 4B and 4C. These two regions contain few genes relative to the MHC locus (29 and 21, respectively, over 2-Mb windows versus 194 in the 3-Mb MHC window; Table S2). Interestingly, the region on chromosome 2 contains functionally important genes such as the protein-coding gene *EYA* or transcription factors of the SIX homeobox family. Both *EYA* and SIX are essential for the development of several organs in mammals, including the kidney and inner ear [26]. The region on chromosome 8 contains genes such as musculin (*ABF-1*), a DNA-binding transcription receptor that may play a role in regulating antigen-dependent B cell differentiation [27]. The presence of these functionally important genes suggests that genetic diversity in these regions may be driven by pseudo-overdominance rather than overdominance, as mutations in these genes are likely to be deleterious and their (partially) recessive effects should be masked if in the heterozygous state. AOD is thus a plausible explanation for the counterintuitive observation that many variants found in scans for balancing selection are associated with severe consequences in protein-coding genes [28].

Additional support for multi-locus AOD in humans comes from the fact that the SFS show strong peaks at intermediate frequencies across populations in the candidate regions compared to the surrounding regions (Figures 4D–4F, Data S2B). While such peaks could be indicative of a mapping issue, we excluded such potentially problematic genomic regions by using the most stringent genomic quality masks (see STAR Methods). We also find noticeable associations of derived states within the same haplotypes across sampled individuals, particularly in non-African populations, suggesting the presence of complementary haplotypes (Data S2C). This matches our predictions of masked recessive mutations as seen in the simulations (Figure S1).

Conclusions

Our results show that multi-locus AOD resulting from pseudo-overdominance of deleterious variants has the potential to

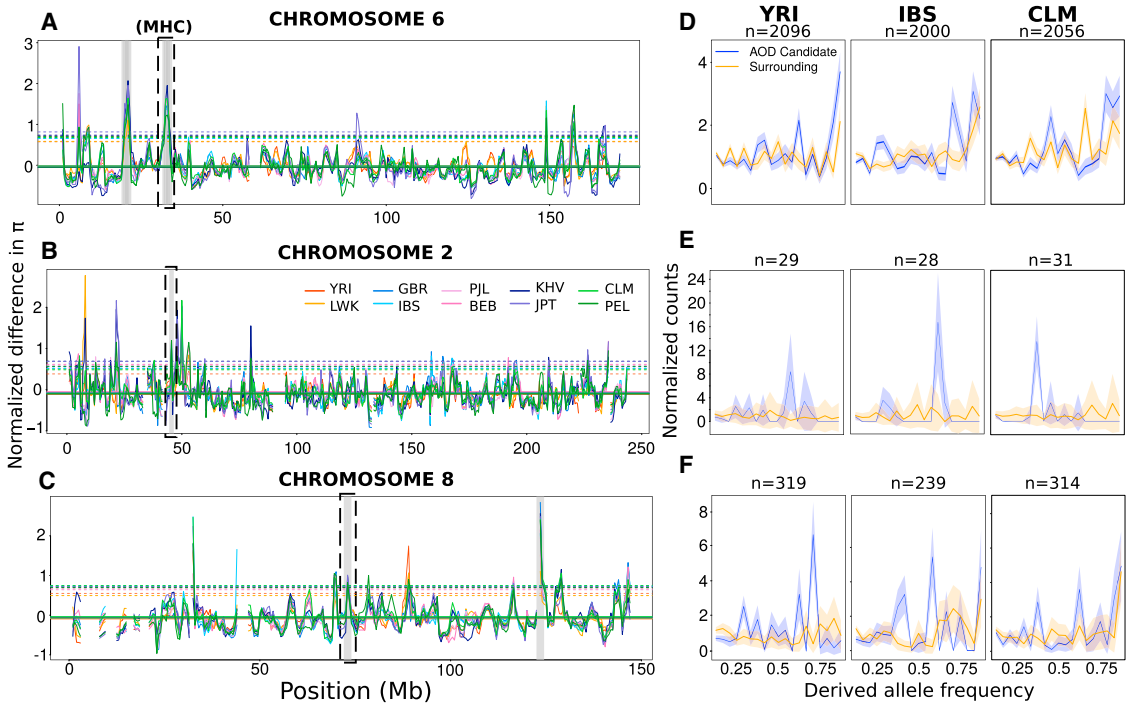

Figure 4. Three Candidate Regions Showing Potential Signatures of AOD in Humans

Left panels present genome scans of the normalized difference in π between low and medium recombination on chromosomes 6 (A), 2 (B), and 8 (C). Shaded gray areas represent candidate windows for AOD where the normalized difference is higher than two SDs (dashed lines) from the mean for each population (see STAR Methods for details). The dashed boxes indicate the outliers presented in the right panels (Chr2: 44,500,001–46,500,000 bp; Chr6: 31,500,001–34,500,000 bp; Chr8: 72,000,001–74,000,000). (B) Observed SFS normalized by dividing each class by its expectation in a stationary population [24] across three populations for low-recombination (LR) SNPs in the candidate region (blue) versus LR SNPs in the surrounding 5Mb region (orange). Numbers of polymorphic LR SNPs in the candidate region within that population are indicated above each SFS. Confidence intervals (shaded areas) are estimated from 100 bootstrap replicates over this same number of SNPs.

See also Figure S3 for distinguishing balancing selection from AOD and all outlier regions listed in Table S1 or shown in Data S2.

impact neutral genomic diversity in regions of low recombination. Importantly, we find that the resulting increase in diversity due to multi-locus AOD is much stronger than previously identified from one- or two-locus models and can lead to a form of selection maintaining complementary heterozygous haplotypes. Moreover, simulations show that AOD can also prevail over BGS in the presence of strongly deleterious mutations ($1 < Ns < 100$), and not only in presence of slightly deleterious mutations ($Ns < 1$), making AOD a potentially more general phenomenon than previously anticipated. The increase in diversity due to AOD is furthermore notable as it will persist in the population over time, unlike temporary increases in diversity due to the potential occurrence of simultaneous selective sweeps resulting from positive selection [29], where one allele will eventually drift to fixation over the other(s). The conditions under which we expect a transition from BGS to multi-locus AOD depend on the dominance of deleterious alleles, the strength of selection, and the recombination rate. The rate of influx of new deleterious mutations also plays a critical role, since there is a balance between strength of selection, the rate of recombination and the speed at which these variants are removed (Figure 3; Data S2C). These interactions thus determine whether deleterious variants are favored in the heterozygous state or if selection can efficiently remove them before they accumulate in heterozygotes across the genome, creating a resurgence of genetic diversity in regions of low recombination.

Multi-locus AOD within the genome is differentially detected by π and $\overline{DAF_i}$ depending on the strength of linked selection and the degree of dominance. Contrasting π and $\overline{DAF_i}$ could thus open avenues for statistical inference of the joint distribution of selection and dominance coefficients. Empirical investigation into the human genome finds 22 regions likely impacted by multi-locus AOD. Distinguishing the signatures of AOD from classic balancing selection requires further investigation, as both these processes increase diversity, and may present a suitable challenge for machine learning methods. Importantly, our results highlight the point that many regions previously identified as under balancing selection might actually be cases of multi-locus AOD at low recombination rates.

Variable demography and genetic drift may impact the dynamics of AOD in the genome. For instance, the transition from BGS to AOD through loss of the wild-type haplotype and increase in frequency of recessive deleterious variants could become more likely due to pulses of increased genetic drift, such as bottlenecks and range expansions. Along these lines, all 22 candidate regions show a clear excess of variants at intermediate frequencies across all populations regardless of their demographic history (Figures 4D–4F, Data S2B). Generally, we

find stronger signals of multi-locus AOD in non-African populations, both in the SFS with distinct frequency peaks (Figures 4D–4F) as well as more pronounced clustering of (potentially complementary) haplotypes (Data S2B and S2C).

Understanding the causes, consequences, and conditions leading to BGS and AOD is a major step in understanding the drivers of diversity across the genome. Our results inform how selection against deleterious variants can impact the evolution of linked neutral variation and lead to unexpected patterns of diversity in regions of low recombination. Inference methods using SFS data may thus be biased differently by AOD than by BGS if regions of low recombination are significant contributors to measured diversity. Negative selection in regions of low recombination may thus create more complex patterns of diversity than previously expected, emphasizing the need for a better understanding of the impacts of linked selection in the genome. For instance, multi-locus AOD may lead to misidentification in scans for balancing selection, interfere with the detection of introgressed or inverted regions of the genome where large regions of heterozygosity are expected, or interfere with the detection of local adaptation using F_{ST} outlier scans (see also [30]). This study opens promising avenues for further empirical and theoretical research into the interaction of recombination with drivers of genomic diversity, including the efficiency of selection against introgressed variation [31], evolutionary dynamics of mutation load and inbreeding depression, or the evolution of recombination modifiers [32].

STAR★METHODS

Detailed methods are provided in the online version of this paper and include the following:

- KEY RESOURCES TABLE
- LEAD CONTACT AND MATERIALS AVAILABILITY
- METHOD DETAILS
  - Simulations
  - Mathematical Model
  - Human Analyses
- QUANTIFICATION AND STATISTICAL ANALYSIS
  - Summary Statistics
  - Identification of AOD in Humans
- DATA AND CODE AVAILABILITY

SUPPLEMENTAL INFORMATION

Supplemental Information can be found online at https://doi.org/10.1016/j.cub.2019.11.063.

ACKNOWLEDGMENTS

We would like to thank Brian Charlesworth, Vitor Sousa, Claudia Bank, Peter Ralph, Hannes Becher, Ben Jackson, and Lei Zhao for useful insights, discussion, and feedback on the topics of this paper. We also thank two anonymous reviewers for their critical assessment and helpful comments. F.P. and K.J.G. were funded by a Swiss NSF grant (310030B-166605 to L.E.).

AUTHOR CONTRIBUTIONS

K.J.G., F.P., L.E., and S.P. conceived the study. K.J.G. performed all simulation analyses. S.P. developed all analytic theory. F.P. performed all human analyses. All authors contributed to writing the manuscript.

DECLARATION OF INTERESTS

The authors declare no competing interests.

Received: August 26, 2019
Revised: October 16, 2019
Accepted: November 21, 2019
Published: December 19, 2019

REFERENCES

1. Charlesworth, B., Morgan, M.T., and Charlesworth, D. (1993a). The effect of deleterious mutations on neutral molecular variation. Genetics *134*, 1289–1303.
2. Ota, T. (1971). Associative overdominance caused by linked detrimental mutations. Genet. Res. *18*, 277–286.
3. Zhao, L., and Charlesworth, B. (2016). Resolving the conflict between associative overdominance and background selection. Genetics *203*, 1315–1334.
4. Ohta, T., and Kimura, M. (1970). Development of associative overdominance through linkage disequilibrium in finite populations. Genet. Res. *16*, 165–177.
5. Pouyet, F., Aeschbacher, S., Thiéry, A., and Excoffier, L. (2018). Background selection and biased gene conversion affect more than 95% of the human genome and bias demographic inferences. eLife *7*, e36317.
6. Barton, N.H., and Etheridge, A.M. (2004). The effect of selection on genealogies. Genetics *166*, 1115–1131.
7. Becher, H., Jackson, B.C., and Charlesworth, B. (2019). Patterns of genetic variability in genomic regions with low rates of recombination. Curr. Biol. *30*. Published online December 19, 2019. https://doi.org/10.1016/j.cub.2019.10.047.
8. Begun, D.J., and Aquadro, C.F. (1992). Levels of naturally occurring DNA polymorphism correlate with recombination rates in D. melanogaster. Nature *356*, 519–520.
9. Aquadro, C.F., Begun, D.J., and Kindahl, E.C. (1994). Selection, recombination, and DNA polymorphism in Drosophila. Non-neutral evolution (Springer), pp. 46–56.
10. Kreitman, M., and Wayne, M.L. (1994). Organization of genetic variation at the molecular level: lessons from Drosophila. Molecular Ecology and Evolution: Approaches and Applications (Birkhäuser, Basel), pp. 157–183.
11. Shapiro, J.A., Huang, W., Zhang, C., Hubisz, M.J., Lu, J., Turissini, D.A., Fang, S., Wang, H.-Y., Hudson, R.R., Nielsen, R., et al. (2007). Adaptive genic evolution in the Drosophila genomes. Proc. Natl. Acad. Sci. USA *104*, 2271–2276.
12. Kulathinal, R.J., Bennett, S.M., Fitzpatrick, C.L., and Noor, M.A. (2008). Fine-scale mapping of recombination rate in Drosophila refines its correlation to diversity and divergence. Proc. Natl. Acad. Sci. USA *105*, 10051–10056.
13. Hellmann, I., Prüfer, K., Ji, H., Zody, M.C., Pääbo, S., and Ptak, S.E. (2005). Why do human diversity levels vary at a megabase scale? Genome Res. *15*, 1222–1231.
14. Huynh, L.Y., Maney, D.L., and Thomas, J.W. (2010). Contrasting population genetic patterns within the white-throated sparrow genome (Zonotrichia albicollis). BMC Genet. *11*, 96.
15. Stephan, W., and Langley, C.H. (1998). DNA polymorphism in lycopersicon and crossing-over per physical length. Genetics *150*, 1585–1593.
16. Tenaillon, M.I., Sawkins, M.C., Long, A.D., Gaut, R.L., Doebley, J.F., and Gaut, B.S. (2001). Patterns of DNA sequence polymorphism along chromosome 1 of maize (Zea mays ssp. mays L.). Proc. Natl. Acad. Sci. USA *98*, 9161–9166.
17. Tenaillon, M.I., Sawkins, M.C., Anderson, L.K., Stack, S.M., Doebley, J., and Gaut, B.S. (2002). Patterns of diversity and recombination along chromosome 1 of maize (Zea mays ssp. mays L.). Genetics *162*, 1401–1413.

18. Hudson, R.R., and Kaplan, N.L. (1995). Deleterious background selection with recombination. Genetics *141*, 1605–1617.

19. Nordborg, M., Charlesworth, B., and Charlesworth, D. (1996). The effect of recombination on background selection. Genet. Res. *67*, 159–174.

20. Charlesworth, D., Morgan, M.T., and Charlesworth, B. (1993b). Mutation accumulation in finite outbreeding and inbreeding populations. Genet. Res. *61*, 39–56.

21. Pamilo, P., and Pálsson, S. (1998). Associative overdominance, heterozygosity and fitness. Heredity *81*, 381–389.

22. Pálsson, S., and Pamilo, P. (1999). The effects of deleterious mutations on linked, neutral variation in small populations. Genetics *153*, 475–483.

23. Bitarello, B.D., de Filippo, C., Teixeira, J.C., Schmidt, J.M., Kleinert, P., Meyer, D., and Andrés, A.M. (2018). Signatures of long-term balancing selection in Human genomes. Genome Biol. Evol. *10*, 939–955.

24. Lapierre, M., Lambert, A., and Achaz, G. (2017). Accuracy of demographic inferences from the Site Frequency Spectrum: the case of the Yoruba population. Genetics *206*, 439–449.

25. Meyer, D., and Thomson, G. (2001). How selection shapes variation of the human major histocompatibility complex: a review. Ann. Hum. Genet. *65*, 1–26.

26. Patrick, A.N., Cabrera, J.H., Smith, A.L., Chen, X.S., Ford, H.L., and Zhao, R. (2013). Structure-function analyses of the human SIX1-EYA2 complex reveal insights into metastasis and BOR syndrome. Nat. Struct. Mol. Biol. *20*, 447–453.

27. Massari, M.E., Rivera, R.R., Voland, J.R., Quong, M.W., Breit, T.M., van Dongen, J.J., de Smit, O., and Murre, C. (1998). Characterization of ABF-1, a novel basic helix-loop-helix transcription factor expressed in activated B lymphocytes. Mol. Cell. Biol. *18*, 3130–3139.

28. Key, F.M., Teixeira, J.C., de Filippo, C., and Andrés, A.M. (2014). Advantageous diversity maintained by balancing selection in humans. Curr. Opin. Genet. Dev. *29*, 45–51.

29. Chevin, L.-M., Billiard, S., and Hospital, F. (2008). Hitchhiking both ways: effect of two interfering selective sweeps on linked neutral variation. Genetics *180*, 301–316.

30. Cruickshank, T.E., and Hahn, M.W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol. Ecol. *23*, 3133–3157.

31. Harris, K., and Nielsen, R. (2016). The genetic cost of Neanderthal introgression. Genetics *203*, 881–891.

32. Berdan, E.L., Blanckaert, A., Butlin, R.K., and Bank, C. (2019). Muller's ratchet and the long-term fate of chromosomal inversions. bioRxiv. https://doi.org/10.1101/606012.

33. Auton, A., Brooks, L.D., Durbin, R.M., Garrison, E.P., Kang, H.M., Korbel, J.O., Marchini, J.L., McCarthy, S., McVean, G.A., and Abecasis, G.R.; 1000 Genomes Project Consortium (2015). A global reference for human genetic variation. Nature *526*, 68–74.

34. Frazer, K.A., Ballinger, D.G., Cox, D.R., Hinds, D.A., Stuve, L.L., Gibbs, R.A., Belmont, J.W., Boudreau, A., Hardenbol, P., Leal, S.M., et al.; International HapMap Consortium (2007). A second generation human haplotype map of over 3.1 million SNPs. Nature *449*, 851–861.

35. Haller, B.C., and Messer, P.W. (2019). SLiM 3: Forward genetic simulations beyond the Wright–Fisher model. Mol. Biol. Evol. *36*, 632–637.

36. Quinlan, A.R. (2014). BEDTools: The Swiss-Army tool for genome feature analysis. Curr. Protoc. Bioinformatics *47*, 11.12.1–11.12.34.

37. Danecek, P., Auton, A., Abecasis, G., Albers, C.A., Banks, E., DePristo, M.A., Handsaker, R.E., Lunter, G., Marth, G.T., Sherry, S.T., et al.; 1000 Genomes Project Analysis Group (2011). The variant call format and VCFtools. Bioinformatics *27*, 2156–2158.

38. Smedley, D., Haider, S., Durinck, S., Pandini, L., Provero, P., Allen, J., Arnaiz, O., Awedh, M.H., Baldock, R., Barbiera, G., et al. (2015). The BioMart community portal: an innovative alternative to large, centralized data repositories. Nucleic Acids Res. *43* (W1), W589-98.

39. Scally, A. (2016). The mutation rate in human evolution and demographic inference. Curr. Opin. Genet. Dev. *41*, 36–43.

STAR★METHODS

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Deposited Data		
Raw human variants	[33]	http://hgdownload.cse.ucsc.edu/gbdb/hg19/1000Genomes/phase3/
Filtered human variants used in this study	This paper	Mendeley Data Repository: https://doi.org/10.17632/rgp5xss3x9.1
Human accessible genome masks	[33]	ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/release/20130502/supporting/accessible_genome_masks/
Recombination map	[34]	ftp://ftp.1000genomes.ebi.ac.uk/vol1/ftp/technical/working/20130507_omni_recombination_rates/
Derived alleles frequencies per individual per population	[5]	https://datadryad.org/resource/doi:10.5061/dryad.t76fk80.2/16.2
Software and Algorithms		
SLiM v3.2	[35]	https://messerlab.org/slim/
Mathematica	Wolfram	http://www.wolfram.com/mathematica/
bedtools	[36]	https://github.com/arq5x/bedtools2/releases/download/v2.28.0/bedtools-2.28.0.tar.gz
vcftools	[37]	https://sourceforge.net/projects/vcftools/files/
BioMart	[38]	http://grch37.ensembl.org/biomart/martview/1a32c1634d968649752622cb3c48081d
Other		
Simulation analysis files (R scripts)	This paper	Mendeley Data Repository: https://doi.org/10.17632/rgp5xss3x9.1
Mathematica code/scripts	This paper	Mendeley Data Repository: https://doi.org/10.17632/rgp5xss3x9.1
Human genome analysis scripts	This paper	Mendeley Data Repository: https://doi.org/10.17632/rgp5xss3x9.1

LEAD CONTACT AND MATERIALS AVAILABILITY

Further information and requests for resources should be directed to and will be fulfilled by the Lead Contact, Kimberly J. Gilbert (kimberly.gilbert@unil.ch).

METHOD DETAILS

Simulations

We conduct forward-time, individual-based simulations to clarify the processes of BGS and AOD occurring in the genome, using SLiM v3.2 [35]. We model a census population size of 10,000 individuals in one panmictic, constant-sized population, and sample 100 individuals from this population at generation 6*N*. The per-bp mutation rate in all simulations is $1.45e^{-8}$, matched to latest estimates from humans [39], and we prevented mutations from stacking (see SLiM manual). Only mutations unconditionally deleterious to fitness or neutral are considered, *i.e.*, we do not include the impact of beneficial mutations or potential selective sweeps that would be associated with such variants.

Individuals consist of six chromosomes of 7.5Mb each, making each individual's genome a total size of 45Mb. Each chromosome is subject to only one recombination rate (0, 0.001, 0.01, 0.1, 1, and 10 cM/Mb), and each replicate simulation only exhibits deleterious mutations of one constant, fixed selection coefficient *s* (0.00001, 0.0001, 0.001, 0.01, 0.1; where fitness is defined as 1, 1 - *hs*, and 1 - *s*). An additional set of simulations modeled a gamma distribution for deleterious fitness effects with mean *s* equal to the above range and shape parameter 1, *i.e.*, an exponential distribution. Mutations are only either neutral or deleterious, and we varied the proportion of deleterious mutations to be either 2% or 20% (Figure S4) of the total mutations occurring at any given point in time. Across replicates we also varied the dominance of deleterious mutations ($h = 0.5, 0.25, 0$; additive, partially recessive, and fully recessive, respectively).

Mathematical Model

We consider n biallelic loci and assume these loci are equidistant from each other with respect to recombination distance. We denote the recombination rate between the first and last locus as r. Thus, the recombination distance between two adjacent loci is approximately $r/(n-1)$. Derived alleles are fully recessive and deleterious with fitness effect s in the homozygous state. We assume that fitness is multiplicative across loci, *i.e.*, there is no epistasis. Mutations occur at each locus with rate u per generation and individual, and $U_d = n\,u$ is the total genome-wide deleterious mutation rate. We assume that haplotypes segregating at mutation-selection balance carry at most one deleterious allele. If the number of loci is large and the mutation rate is small relative to selection ($|u/s| << 1$) haplotypes carrying multiple mutations are vanishingly rare and we can ignore them (see Data S1). Further, at mutation-selection balance, all haplotypes carrying the same number of deleterious alleles will be at equal frequency. Thus the evolution of the wild-type frequency follows

$$\frac{dP_0}{d_t} = \left(\overline{W} - W_0\right)P_0 - U_d\,P_0 + \sum_{i=1}^{n}\sum_{j=i+1}^{n}(j-i)r/(n-1)P_1^2,$$

where P_1 denotes the frequency of haplotypes carrying a single deleterious allele, and $\overline{W}$ and W_0 are the mean fitness of the population and the marginal fitness of a wild-type haplotype, respectively. The first term captures the change in frequency of P_0 due to selection, the second term the loss of wild-type haplotypes due to mutation, and the last term describes the rate at which recombination between different single-mutation haplotypes creates wild-type haplotypes. In Data S1 we show that for low mutation and recombination rates, the frequency of P_0 at equilibrium can be approximated by $P_0 \approx 1 - n\sqrt{\mu} + (n(n+1)/12)\sqrt{\mu}\rho$, where $\mu = (u/s)$ and $\rho = (r/s)$ are the scaled strengths of mutation and recombination, respectively.

Human Analyses

We use a genotype table containing 100 individuals from ten 1000G populations [33] from [5]. Population labels are according to The 1000G Project Consortium nomenclature: YRI = Yoruba in Idaban Nigeria, LWK = Luhya in Webuye Kenya, GBR = British from England and Scotland UK, IBS = Iberian populations in Spain, BEB = Bengali in Bangladesh, PJL = Punjabi in Lahore Pakistan, KHV = Kinh in Ho Chi Minh City Vietnam, JPT = Japanese in Tokyo Japan, CLM = Colombian in Medellin Colombia, and PEL = Peruvian in Lima Peru (names of individuals are available with the archived Mendeley Data Repository). We use bedtools [36] to focus on genomic regions with very low negative and positive false discovery rates from next generation sequencing methods by using the 1000G strictMask map. This map keeps genomic regions where the depth of coverage (summed across all 1000G samples) is within 50% of the average, where no more than 0.1% of reads have a mapping quality of zero and where the mapping quality is > 56 [33]. This generates a total of 13,385,820 SNPs. We use the LD-based Yoruba-specific recombination map [34] to define recombination rates across the genome.

QUANTIFICATION AND STATISTICAL ANALYSIS

Summary Statistics

To understand the evolutionary dynamics across our combinations of simulated parameters, we compare several summary statistics from polymorphic sites in the simulated data. We calculated the average derived allele frequency across all individuals at sites polymorphic at the level of all populations in the sample (following [5]). For the i-th individual, $DAF_i = n_i/(2S)$, where n_i is the number of derived alleles observed in the i-th individual and S is the number of polymorphic sites among all individuals considered in the analysis. It has been shown that $\overline{DAF_i}$ is independent of demographic history and would only be affected by variations in mutation rates or selection [5]. If mutation rate is constant across loci, $\overline{DAF_i}$ is equal in expectation to the ratio of the average coalescence time between two chromosomes of an individual over the average total length of the coalescent tree relating all chromosomes in the sample. It should thus be sensitive to BGS and AOD, which affect branch lengths in the tree. We generate the unfolded SFS from the sample of 100 diploid individuals, across the whole genome (i.e., including sites subject to all recombination levels). This SFS is normalized by dividing each class by its expectation in a stationary population [24], creating an SFS where deviation from 1 indicates an enrichment or deficit of variants in that SFS class.

Identification of AOD in Humans

We perform a window-based approach with window sizes of 2Mb and a step increase of 500kb to scan all chromosomes of the human genome per population. Within each window, we use a subset of sites occurring in regions of low recombination (LR: $0 < RR \leq 0.05$ cM/Mb) and medium recombination (MR: $0.1 \leq RR < 0.5$ cM/Mb). We only retained windows where LR and MR regions contained at least 50 SNPs each (still defined as polymorphic at the level of all populations). We compute $\overline{DAF_i}$ across the 10 populations in each of these regions, and we use vcftools [37] to estimate nucleotide diversity (π) per site and average it within the LR and MR regions per window and per population. For each window i, we estimate the difference in nucleotide diversity between LR and MR per population:

$$\frac{\pi_{pop,LRi} - \pi_{pop,\,MRi}}{\pi_{pop,MRi}}.$$

These windows are averaged across the chromosome per population, and outliers are identified as those exceeding two standard deviations of this mean, only when exceeded for all ten populations. We ranked our candidate regions using p values of a one-sample *t*-test on the difference in nucleotide diversity between LR and MR regions, such that the region with the smallest *p*-value should show the strongest signal across all 10 populations (see Table S1). Chromosomes 6, 8, and 14 contain the top three candidates, but we exclude the region on chromosome 14 due to differences between reference panels of hg19 (see note in Table S1). For comparison, we also include a focal candidate from chromosome 2 which only has a small number of polymorphic sites (*i.e.*, < 100, Table S1). We check the overlap between our outlier regions and the putative regions of balancing selection from [23] in one East-African population (LWK) with p-value < 0.05, using bedtools [36]. Finally, we used biomart [38] to identify the gene functions present in these regions, as an analysis of AOD due to balancing selection versus pseudo-overdominance (Table S2). All windows analyses are contained in Data S2.

DATA AND CODE AVAILABILITY

No new data was generated in this study. Analysis scripts and input simulation code along with the filtered human dataset are archived on Mendeley here: https://doi.org/10.17632/rgp5xss3x9.1. Scripts that regenerate the SLiM inputs for simulation and analyze the output from these are provided, the Mathematica notebook for the analytic model is provided, and the scripts that analyze the human genomic data are provided.