Received: 18 July 2020 | Revised: 4 September 2020 | Accepted: 29 September 2020

DOI: 10.1111/mec.15690

NEWS AND VIEWS

PERSPECTIVE

MOLECULAR ECOLOGY WILEY

It's time to stop sweeping recombination rate under the genome scan rug

Laurie S. Stevison[1] | Suzanne E. McGaugh[2]

[1]Department of Biological Sciences, Auburn University, Auburn, AL, USA

[2]Ecology, Evolution, and Behavior, University of Minnesota, Saint Paul, MN, USA

Correspondence
Laurie S. Stevison, Department of Biological Sciences, Auburn University, Auburn, AL, USA.
Email: lstevison@auburn.edu

Different parts of the genome can vary widely in their evolutionary histories and sequence divergence from other species. Indeed, some of the most interesting biology (e.g., hybridization, horizontal gene transfer, variable mutation rates across the genome) is revealed by the discordant relationships between taxa across the genome. The goal for much of evolutionary genetics is centred on understanding the evolutionary processes by which such varied signatures arise and are maintained. Many evolutionary genetics studies seek to identify signatures of positive selection between two closely related ecotypes or taxa by delineating regions with particularly high divergence relative to a genome-wide average, often termed "divergence outliers." In a From the Cover article in this issue of *Molecular Ecology*, Booker et al. take a major step forward in showing that recombination rate differences are sufficient to create false positive divergence outliers, even under neutrality. They demonstrate that the variance of genome scan metrics is especially high in regions with low recombination rates, consistent with previous work. Furthermore, they show that both relative and absolute measures of divergence (F_{ST} and D_{XY}, respectively) as well as other commonly used statistics in genome scans (e.g., π_W, Tajima's D and H12) all have similar covariance between variance and local recombination rate. Finally, Booker et al. show that low recombination regions will tend to produce more outliers if genome-wide averages are used as cut-offs to define genomic outliers. Booker et al.'s results suggest that recombination rate variation, even under neutral conditions, can shape genome scans for selection, and this important variable can no longer be ignored.

KEYWORDS

divergence outliers, islands of divergence, selection scans

Highly heterogeneous divergence across the genome is commonly observed between taxa, and the relative importance of different evolutionary forces in shaping the genomes of diverging taxa is the subject of intense discussion (e.g., Cruickshank & Hahn, 2014). Prominent explanations for heterogeneous divergence across the genome (e.g., "islands of divergence") include that highly divergent genomic regions are the product of divergent natural selection between taxa or that they contain reproductively incompatible alleles that produce selection against introgression (Haasl & Payseur, 2016). Window-based genome scans are commonly used for identifying differentiation outliers, and these scans extract the extremes of the distributions of summary statistics (Lotterhos, 2019). Genome scans make an underlying assumption that estimates from each marker within a window are identically distributed and independent, making them particularly vulnerable to unknown sources of heterogeneity in the data (Lotterhos, 2019).

The recombination rate landscape is a known source of heterogeneity that can play a major role in creating outlier regions (Cruickshank

& Hahn, 2014; Noor & Bennett, 2009) and may contribute to different evolutionary histories across the genome (Lotterhos, 2019). For instance, in a low recombination region, linkage is more extensive. Following selective sweeps and background (i.e., negative) selection in the population, nucleotide diversity is eroded in longer stretches of the genome than in moderate-to-high recombination regions. This effect accounts for substantial reductions in diversity and more rapid coalescence across low recombination rate regions of the genome. Additionally, lower recombination rates can result in more effective selection against introgression because linked blocks containing incompatibilities can be efficiently removed without recombination to break them up (Schumer et al., 2018). This process leads to a predictable positive relationship between introgression and recombination rate (Martin et al., 2019). Thus, especially over the past decade, recombination landscape variation has become appreciated as an essential component for understanding how evolutionary processes create heterogeneity in genomic divergence (Haasl & Payseur, 2016).

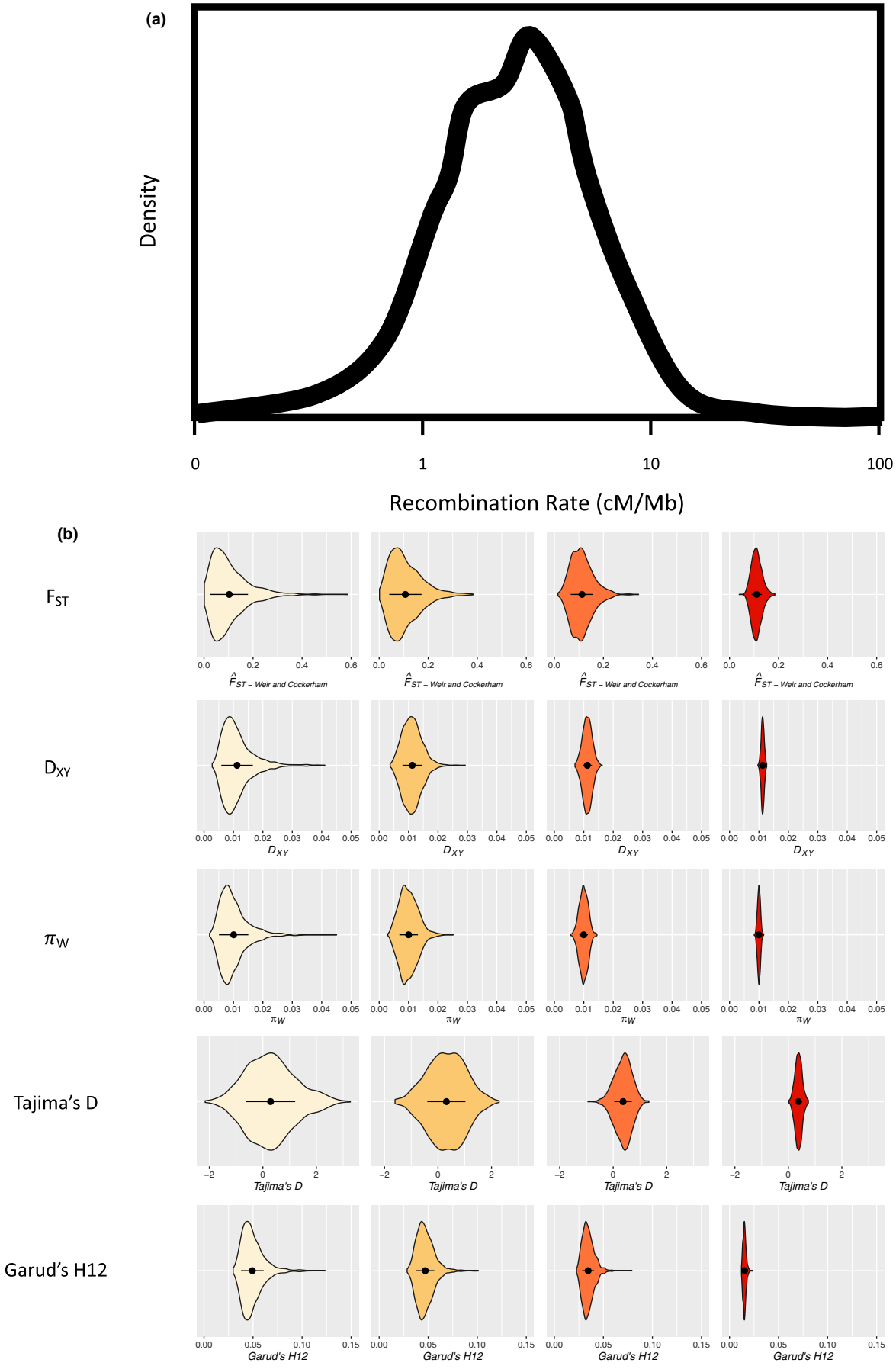
In this issue of *Molecular Ecology*, Booker et al. (2020) show that recombination rate variation impacts regional diversity and divergence estimates even without background selection and/or positive selection impacting the genome (which also result in signatures of selection at linked sites). While previous work has demonstrated that the variance of the number of segregating sites depends on recombination rate (Hudson, 1983), Booker et al. conduct explicit simulations of commonly used genome scan statistics to highlight the impact of recombination rate on their statistical distributions. An important result from their study is that, despite similar mean values across various recombination rates, F_{ST} exhibits increased variance in regions of low recombination and has a long right-hand tail—specifically where "outliers" may be identified (Figure 1). They demonstrate that outliers are enriched in areas of low recombination both with simulated data and re-analyses of a previous empirical data set (Reinhardt et al., 2014). This artefact uncovered by Booker et al. (2020) may be obscuring true signals of selection outliers while exacerbating false positives.

Notably, simply switching statistics to absolute measures of divergence did little to ameliorate recombination rate's impact on genomic divergence (Figure 2). Booker et al. (2020) incorporated several commonly used statistics, both relative (F_{ST}) and absolute (D_{XY}) measures of divergence, haplotype-based statistics (H12), and within-population diversity (π_W; Tajima's D). Their simulation results of sliding windows using all five statistics reveal that the variance of these statistics covaries with local recombination rate in genome scans (Figure 1b). This suggests that the impact of the recombination landscape on neutral divergence is not just a problem with F_{ST}, as noted by other authors (Cruickshank & Hahn, 2014; Noor & Bennett, 2009). Specifically, the variance of the distribution of these statistics is much higher when recombination rate is low, such that applying a genome-wide cutoff, as is often done in selection studies, would oversample regions of low recombination rate (Figure 1). Interestingly, although only the variance of most test statistics is different across various recombination rates, their data suggest that the mean of H12 is affected by recombination rate. This result might warrant further investigation of haplotype statistics in general.

One can more greatly appreciate these results by reflecting on the remarkable recombination rate variation that is typically observed across the genome (Figure 1a). For example, recombination rates are ubiquitously reduced near centromeres and reduced in telomeric regions in some sexes and species (Stapley et al., 2017). Thus, researchers should be especially mindful that false positive outliers may be aberrantly concentrated in these low recombination areas (e.g., Cruickshank & Hahn, 2014; Noor & Bennett, 2009). Recombination rate also varies dramatically at a finer scale. An extreme example is that in many primates and mice, recombination is governed by a trans-acting factor (PRDM9) that binds to DNA and induces a double strand break, leading to short recombination "hotspots" punctuated by long stretches of recombination desert. Hotspots experience recombination rates up to several thousand times greater than background recombination levels (Grey et al., 2018). In species that lack a functional *Prdm9*, such as dogs, yeast, birds and monkey flowers, recombination is localized to transcription start sites, introducing another important caveat to genome scans. As transcription start sites are then enriched for higher recombination rates (Grey et al., 2018; but see He et al., 2017), they may have lower rates of outlier detection as indicated by Booker et al.'s (2020) simulations. Furthermore, recombination maps are stable over long-term evolutionary scales in some of these species (Lichten, 2015), suggesting that the bias introduced from recombination rate variation is maintained, and the impacts on genomic diversity and divergence could be compounded over evolutionary time.

If recombination is such an important parameter in evolutionary studies, why have few incorporated it? Until recently, fine-scale, pedigree-based measures of recombination data were needed to make evolutionary inferences. Although population-averaged recombination rates can be inferred with whole genome sequences from 10 or so individuals, these estimates can carry the signatures of past selection and demography, rendering their use in understanding selection circular (Lotterhos, 2019). Fine-scale, directly estimated recombination maps remain costly and are limited to a few species, and even fewer studies assay recombination rate variability within an ecotype or species (Peñalba & Wolf, 2020). Notably, Booker et al. (2020) cite two new papers with promising advances in surmounting these challenges by using 10X Genomics Linked Reads technology to leverage phased haplotypes to directly identify crossovers. Despite discontinuation of the Linked Reads product line by 10X Genomics, other companies (e.g., Universal Sequencing Technology) and methods (TELL-seq, single cell sequencing, haplotagging, single-tube long-fragment read sequencing) are filling the niche for direct identification of

FIGURE 1 Simulation results of commonly used genome scan summary statistics, F_{ST} and D_{XY}. Here, the resulting distributions were plotted against each other separately for four different recombination rates: 0, 1, 10 and 100 cM/Mb, respectively. Data used for this figure come directly from Booker et al.'s (2020) simulation results on Github


(a)
Density
0
1
10
100
Recombination Rate (cM/Mb)
(b)
F_ST
D_XY
π_W
Tajima's D
Garud's H12
F̂_ST – Weir and Cockerham
D_XY
π_W
Tajima's D
Garud's H12

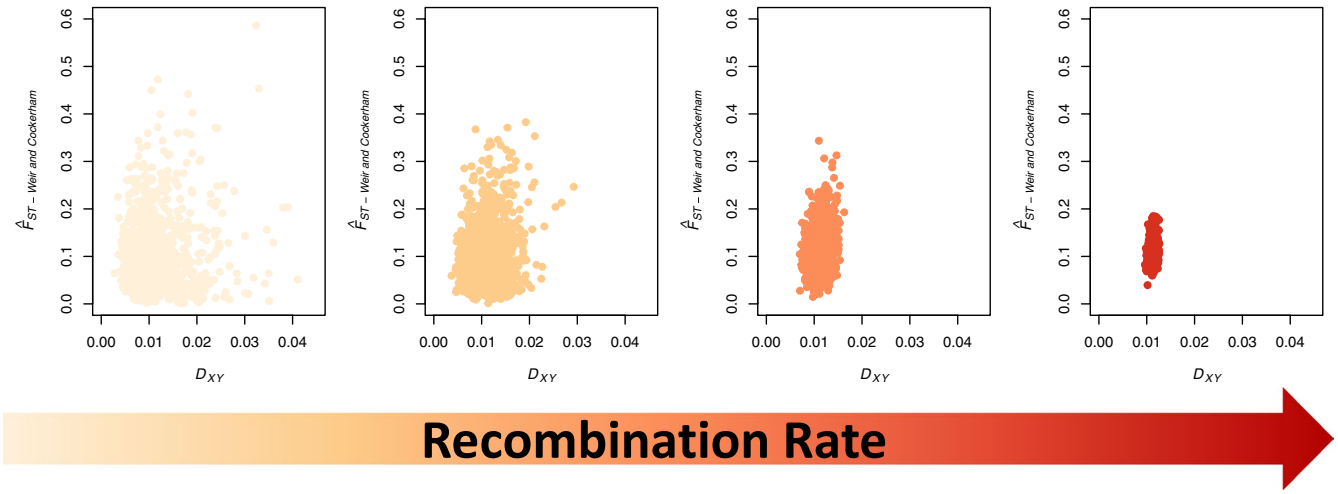


FIGURE 2 Recombination rates vary in natural organisms over several magnitudes throughout the genome (a). Using neutral simulations, Booker et al. (2020) show that the distribution of five commonly used summary statistics for genome scans vary based on recombination rate (b). Their results indicate that outliers in genome scans can be attributed to statistical artefacts due to covariation between recombination rate and the variance of commonly used genome scan summary statistics. Outliers tend to be selected from the right ends of the distributions of these summary statistics, which is unevenly represented in regions of low recombination across statistics. Data used for this figure come directly from Booker et al.'s (2020) simulation results on Github

crossovers, especially in gametes. Thus, as estimating recombination rates becomes less of an obstacle (Peñalba & Wolf, 2020), we hope to see a dramatic expansion of studies that incorporate recombination rate variation into estimation of outliers.

Once recombination maps are obtained, two solutions proposed by Booker et al. are (i) to use recombination rate as a way to define sliding windows as opposed to physical distance, and/or (ii) to use a different threshold for "outliers" from summary statistics based on background recombination rates. Additionally, we recommend that an integrative methodology that jointly estimates divergence while accounting for recombination rate be adopted more broadly in studies of adaptive divergence (Aeschbacher et al., 2017).

The work by Booker et al. (2020) shows unequivocally that this lack of recombination data obscures our understanding of evolutionary processes, and it will be essential to incorporate the often unmeasured covariate of recombination to have confidence in genome scan statistics.

DATA AVAILABILITY STATEMENT

All data used to make figures were from Booker et al., 2020, and are available as supplemental materials in Booker et al., 2020 and on their Github repository: https://github.com/TBooker/Recombination_Fst.

ORCID

Laurie S. Stevison https://orcid.org/0000-0001-7900-5266
Suzanne E. McGaugh https://orcid.org/0000-0003-3163-3436

REFERENCES

Aeschbacher, S., Selby, J. P., Willis, J. H., & Coop, G. (2017). Population-genomic inference of the strength and timing of selection against gene flow. *Proceedings of the National Academy of Sciences of the United States of America*, *114*(27), 7061–7066. https://doi.org/10.1073/pnas.1616755114

Booker, T., Yeaman, S., & Whitlock, M. C. (2020). Variation in recombination rate affects detection of outliers in genome scans under neutrality. *Molecular Ecology*, *29*, 4274–4279. https://doi.org/10.1111/mec.15501. [Epub ahead of print].

Cruickshank, T. E., & Hahn, M. W. (2014). Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. *Molecular Ecology*, *23*, 3133–3157. https://doi.org/10.1111/mec.12796

Grey, C., Baudat, F., & de Massy, B. (2018). PRDM9, a driver of the genetic map. *PLoS Genetics*, *14*(8), e1007479. https://doi.org/10.1371/journal.pgen.1007479

Haasl, R. J., & Payseur, B. A. (2016). Fifteen years of genomewide scans for selection: Trends, lessons and unaddressed genetic sources of complication. *Molecular Ecology*, *25*(1), 5–23. https://doi.org/10.1111/mec.13339

He, Y., Wang, M., Dukowic-Schulze, S., Zhou, A., Tiang, C.-L., Shilo, S., & Pawlowski, W. P. (2017). Genomic features shaping the landscape of meiotic double-strand-break hotspots in maize. *Proceedings of the National Academy of Sciences*, *114*(46), 12231–12236. https://doi.org/10.1073/pnas.1713225114

Hudson, R. R. (1983). Properties of a neutral allele model with intragenic recombination. *Theoretical Population Biology*, *23*(2), 183–201. https://doi.org/10.1016/0040-5809(83)90013-8

Lichten, M. (2015). Putting the breaks on meiosis. *Science*, *350*(6263), 913. https://doi.org/10.1126/science.aad5404

Lotterhos, K. E. (2019). The effect of neutral recombination variation on genome scans for selection. *G3: Genes, Genomes, Genetics*, *9*(6), 1851–1867. https://doi.org/10.1534/g3.119.400088

Martin, S. H., Davey, J. W., Salazar, C., & Jiggins, C. D. (2019). Recombination rate variation shapes barriers to introgression across butterfly genomes. *PLOS Biology*, *17*(2), e2006288. https://doi.org/10.1371/journal.pbio.2006288

Noor, M. A. F., & Bennett, S. M. (2009). Islands of speciation or mirages in the desert? Examining the role of restricted recombination in maintaining species. *Heredity*, *103*(6), 439–444. https://doi.org/10.1038/hdy.2009.151

Peñalba, J. V., & Wolf, J. B. W. (2020). From molecules to populations: Appreciating and estimating recombination rate variation. *Nature Reviews Genetics*, *21*(8), 476–492. https://doi.org/10.1038/s41576-020-0240-1

Reinhardt, J. A., Kolaczkowski, B., Jones, C. D., Begun, D. J., & Kern, A. D. (2014). Parallel geographic variation in Drosophila melanogaster. *Genetics*, *197*(1), 361–373. https://doi.org/10.1534/genetics.114.161463

Schumer, M., Xu, C., Powell, D. L., Durvasula, A., Skov, L., Holland, C., & Przeworski, M. (2018). Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science*, *360*(6389), 656–660. https://doi.org/10.1126/science.aar3684

Stapley, J., Feulner, P. G. D., Johnston, S. E., Santure, A. W., & Smadja, C. M. (2017). Variation in recombination frequency and distribution across eukaryotes: Patterns and processes. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *372*(1736), 20160455. https://doi.org/10.1098/rstb.2016.0455

How to cite this article: Stevison LS, McGaugh SE. It's time to stop sweeping recombination rate under the genome scan rug. *Mol Ecol*. 2020;29:4249–4253. https://doi.org/10.1111/mec.15690