JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.13129

RESPONSE

Land ahoy? Navigating the genomic landscape of speciation while avoiding shipwreck

A. M. WESTRAM* & M. RAVINET†

*Department of Animal and Plant Sciences, University of Sheffield, Sheffield, UK

†Centre for Ecological and Evolutionary Synthesis, University of Oslo, Oslo, Norway

If there is one certainty about speciation, it is that those who research the process are able to disagree about almost anything. Arguments have raged over how to define species, over the importance of spatial separation and, more recently, over whether peaks of differentiation in genome scans might point to barrier loci. The debate calling attention to the shortcomings of searching for 'genomic islands' (Noor & Bennett, 2009; Cruickshank & Hahn, 2014) has been important to clarify where we are heading before scarce research funding is blown on tortuous journeys (*sensu* Baird, 2017).

With our review (Ravinet *et al.*, 2017), we aimed to clarify some of the issues surrounding interpretations of genome scan data and to suggest a practical way forward in dealing with the confounding factors that might obscure the genomic signal of reproductive isolation (RI). As the commentaries in this issue show, there is clearly still room for discussion and deeper understanding, as well as a penchant for inventive and maritime themed titles. In the interest of space, we will not respond to all the points the commentaries make; instead, our aim here is to highlight, and expand a little, on some of the common themes raised.

The quest to identify barrier loci – chasing the white whale?

> I try all things. I achieve what I can
>
> Ishmael in *Moby-Dick; or, The Whale* by Herman Melville.

Understanding which processes lead to the evolution of RI is central to speciation research. Does RI evolve as a by-product of divergent ecological selection, or do incompatibilities arise from genomic conflict? How does gene flow during divergence affect the evolution of RI? Do incompatible alleles evolve via novel mutation or are old variants from standing genetic variation reused? The relative importance of different processes for speciation across the tree of life is unclear and will vary between taxa. Narrowing down individual barrier loci as much as possible, and characterizing them in detail, will help with answering such questions in a focal species pair, and also more generally (Feder *et al.*, 2017). For example, the functional annotations of barrier loci can be classified, or the evolutionary history of adaptive alleles can be inferred, once barrier loci are known. Information on barrier loci also makes it possible to test whether they cluster in the genome, what effect sizes they have on RI, if coding or regulatory changes are more important (Hoekstra & Coyne, 2007), and how barrier loci and genome structure co-evolve (Burri, 2017a; Feulner & De-Kaye, 2017; Ortiz-Barrientos & James, 2017).

Despite such promise, identifying barrier loci to such a detailed level is challenging. Lohse (2017) and Baird (2017) highlight that explicitly integrating demographic history when identifying barrier loci is crucial. This may require outlier scans accounting for an explicitly modelled demographic history, or models of demographic history directly including among-locus variation in m (to account for lower effective migration at barrier loci) or N_e (to account for the effects of selective sweeps and background selection which are expected to reduce this parameter) (Roux *et al.*, 2014, 2016). Sophisticated modelling aside, identifying and characterizing barrier loci may be infeasible in many systems. The limitations go beyond the difficulties in detecting highly polygenic barriers (see following section). For example, if achieving fine-scale genomic resolution is technically impossible, then so is narrowing down barrier loci to individual nucleotides or structural variants; obtaining unequivocal proof for the role of each candidate locus will be difficult with a lack of power, and distinguishing between different types of barrier loci will be hard if there are limited data beyond the genomic level (Lindkte & Yeaman, 2017). Instead, we may end up with genomic landscapes where we can identify broad patterns of differentiation and diversity, but lack power or resolution to pinpoint and characterize individual barrier loci with confidence.

Faced with these difficulties, many of the commentators doubt whether identifying barrier loci and their function in detail is even necessary (Baird, 2017; Buerkle, 2017; Ellegren & Wolf, 2017; Elmer, 2017; Feder *et al.*, 2017; Jiggins & Martin, 2017; Wagner & Mandeville, 2017). There are interesting questions we can address without this information: What is the history of divergence and how much gene flow has been exchanged? Is gene flow particularly restricted in some parts of the genome, for example on sex chromosomes (Muirhead & Presgraves, 2016)? Genomic landscapes are contributing to answering such questions. However, we should make use of theoretical predictions of the genomic basis for RI (e.g. barrier loci number and

Correspondence: Anja M. Westram, Department of Animal and Plant Sciences, Alfred Denny Building, University of Sheffield, Western Bank, Sheffield S10 2TN, UK.
Tel.: +44 114 222 0113;
e-mail: a.westram@sheffield.ac.uk

distribution, effect size and the role of pleiotropy) under different speciation scenarios (e.g. with vs. without gene flow, sexual vs. natural selection). Building a predictive framework is possible with or without detailed knowledge of barrier loci and is important for moving away from more descriptive research.

'Difficult-to-find' barrier loci

It is not necessarily the case that '*the genic view of speciation assumed here anticipates that loci of particular importance for reducing gene flow leave genomic signatures detectable by genome scans*' (Ellegren & Wolf, 2017). Identifying the limits to our understanding of genomic barriers to gene flow from genome scans was a major motivation for the review (see Step 6 of the Roadmap in Ravinet *et al.* (2017) to see the clearest demonstration of this). The genetic basis of RI may include loci that are 'undetectable' in genome scans even if many of the technical issues mentioned above are solved. These include one-allele barriers that do not leave a signature of genomic differentiation (Ravinet *et al.*, 2017), loci in highly repetitive genomic regions and loci with relatively small individual effects underlying polygenic barriers (Baird, 2017; Lohse, 2017). The 'best detectable' types of barrier loci will mainly be loci with relatively large fitness effects in genomes with generally low differentiation (Wagner & Mandeville, 2017).

The resulting bias is problematic (Baird, 2017; Buerkle, 2017; Jiggins & Martin, 2017). Theoretical work has shown that numerous small-effect loci can drive speciation without the need for clustering in the genome (Barton, 2001; Chevin *et al.*, 2014; Flaxman *et al.*, 2014; Fraïsse *et al.*, 2016). The apparent lack of empirical evidence for such patterns is not surprising, given a focus on genome scans. Alternative methods need to be employed to follow up outlier scans that do point to a polygenic basis (e.g. Riesch *et al.*, 2017) and to test for polygenic variation even in systems where large-effect loci are known to play a role. Existing data may sometimes be sufficient for such analyses. For example, if divergent traits vary continuously in hybrid zones or laboratory crosses, a polygenic basis is likely. Such observations can be formalized and combined with genomic data. The variation in phenotypic traits in crosses (Lande, 1981) and hybrid zones (Rieseberg & Buerkle, 2002) can be used to estimate the number of loci underlying a divergent trait. Some mapping approaches can partition polygenic variation among chromosomes or genomic regions, shifting the focus away from individual SNPs or short markers (Yang *et al.*, 2011). An additional challenge will be to estimate the relative importance of barrier loci detectable with genome scan approaches vs. those with smaller effects when they cannot be detected with the same methodology. Here, genetic manipulation may prove fruitful; for example, knockout of large-effect loci (following localization in a genome scan) will make it possible to measure the remaining polygenic barrier effects.

Even though loci with large effects may be identified via the genome scan route, it might often be difficult to find further support for them with independent data, as suggested in our road map (Ravinet *et al.*, 2017). This is particularly the case for loci involved in complex epistasis or genotype x environment interactions. The role of these loci might be obscured in laboratory crosses, association studies or to genetic manipulation (Buerkle, 2017) due to its dependence on a specific genomic and ecological environment. For example, Arnegard *et al.* (2014) demonstrate a genomic incompatibility that only manifests under (semi-) natural conditions (i.e. not in the laboratory). Given large numbers of loci and environmental factors, testing all combinations is impossible. This is a challenging aspect and progress will require small steps, starting with testing and developing methods in well-characterized systems (e.g. Ono *et al.*, 2017).

The power of comparative analyses

Many of the commentaries point out the advantages of extending analyses to a wider range of taxa, either closely related to a focal system, or across the tree of life (Burri, 2017a; Ellegren & Wolf, 2017; Elmer, 2017; Feder *et al.*, 2017; Wagner & Mandeville, 2017). Burri (2017a) and Ellegren & Wolf (2017) suggest sampling multiple taxon pairs to account for genomic variation in recombination rate and gene density. If these confounding factors are conserved across the sampled phylogeny, their effects can be indirectly inferred from patterns of diversity and differentiation shared among taxa, and they can be taken into account when trying to identify regions under divergent selection in a focal taxon pair (e.g. Vijay *et al.*, 2016; Dutoit *et al.*, 2017).

However, there are caveats to this approach. First, it relies on recombination/gene density landscapes being more conserved across the phylogeny than patterns of divergent selection. In birds, where synteny and recombination rates are strongly conserved, this approach may be justified. However, recombination landscapes may themselves evolve, sometimes quickly so (Feulner & De-Kaye, 2017; Ortiz-Barrientos & James, 2017). There is even some limited evidence that this may be the case in hybrid bird species (Elgvin *et al.*, 2017); outside of birds, conservation of recombination landscapes across large phylogenetic distances may be uncommon. Clearly, broader sampling is needed to test this (Ellegren & Wolf, 2017). On the other hand, shared regions of high differentiation may actually be involved in RI repeatedly, for example due to parallel evolution (Elmer & Meyer, 2011; Baird, 2017; Elmer, 2017), genomic constraints (Conte *et al.*, 2012) or genomic conflict (Presgraves, 2010).

Even if regions of high differentiation shared across multiple comparisons are mainly driven by shared genomic features, this does not preclude them from also containing loci involved in RI. There is a danger of favouring background selection as an explanation for patterns of high differentiation without conclusively ruling out alternatives. Excluding such regions from sets of candidate loci may mean losing potentially important loci (Burri, 2017b). Tests for positive selection may be informative (Burri *et al.*, 2015), but ultimately modelling is likely to be necessary to test whether background selection alone can explain observed patterns of diversity and differentiation.

Another main motivation for a comparative approach is to get closer to one of the larger goals of speciation research – relating observed genomic patterns to the factors that explain them (Feder *et al.*, 2017). One can contrast taxon pairs that are phylogenetically close (i.e. have similar genomic backgrounds and constraints), but differ in aspects that might be important for speciation – for example, in the extent of gene flow between diverging populations (Martin *et al.*, 2013; Riesch *et al.*, 2017). Such studies will help in testing specific predictions, for example a shift towards large-effect loci or clusters with high gene flow (Yeaman & Whitlock, 2011). As Elmer (2017) discusses, the consistency of the genomic basis across 'replicates' of speciation (parallel evolution) may give key insights into the genomic basis, constraints and repeatability of speciation. However, more theoretical work is needed to predict expected genomic patterns where the same loci are involved in divergence – as Fig. 4 in the Target Review (Ravinet *et al.*, 2017) demonstrates, genomic patterns around barrier loci may be highly stochastic; because each instance of parallel divergence is subject to such stochasticity, identifying shared barrier loci may be challenging.

Studying taxon pairs at various genetic distances (i.e. pairs along the 'speciation continuum') might help identifying types and patterns of barrier loci that contribute to speciation in the long term, rather than being ephemeral (Feder *et al.*, 2017; Lindkte & Yeaman, 2017; Wagner & Mandeville, 2017). Nonetheless, we caution that ephemeral contributions may play an important role during the progression towards speciation. Finally, Baird (2017) emphasizes that broadly sampling the tree of life is necessary to have a representative understanding of how species evolve, and studying genomic landscapes will ultimately not answer general questions if we focus on a biased subset of taxa. Clearly, our focus needs to extend to understudied, nonmodel systems.

Conclusions

Understanding the evolution of reproductive isolation will not always be plain sailing. We may need information from diverse, sometimes challenging approaches including theory, modelling, experiments, field surveys and new sequencing technologies. However, the commentaries identify several key areas for future research, many of them 'in parallel'. We hope these, alongside our own perspective and road map, will provide inspiration, basis for discussion and some sense of direction in a complex but fascinating research field.

Acknowledgments

We thank Mike Ritchie for giving us the opportunity to write the Target Review and this response. We are grateful to all the authors who took the time to contribute the stimulating commentaries that led to this response. We also thank all co-authors of the Target Review for discussion and their insights on the commentaries. We would particularly like to thank Roger Butlin for his thoughts on the commentaries and for providing feedback on an earlier draft.

References

Arnegard, M.E., McGee, M.D., Matthews, B., Marchinko, K.B., Conte, G.L., Kabir, S. *et al.* 2014. Genetics of ecological divergence during speciation. *Nature* **51**: 307–311.

Baird, S. 2017. The impact of high throughput sequencing technology on speciation research: maintaining perspective. *J. Evol. Biol.*, **30**: 1482–1487.

Barton, N.H. 2001. The role of hybridization in evolution. *Mol. Ecol.* **10**: 551–568.

Buerkle, C.A. 2017. Inconvenient truths in population and speciation genetics point toward a future beyond allele frequencies. *J. Evol. Biol.*, **30**: 1498–1500.

Burri, R. 2017a. Dissecting differentiation landscapes: a linked selection's perspective. *J. Evol. Biol.*, **30**: 1501–1505.

Burri, R. 2017b. Interpreting differentiation landscapes in the light of long-term linked selection. *bioRxiv*. http://doi.org.10.1101/131243.

Burri, R., Nater, A., Kawakami, T., Mugal, C.F., Olason, P.I., Smeds, L. *et al.* 2015. Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res.* **25**: 1656–1665.

Chevin, L.-M., Decorzent, G. & Lenormand, T. 2014. Niche dimensionality and the genetics of ecological speciation. *Evolution* **68**: 1244–1256.

Conte, G.L., Arnegard, M.E., Peichel, C.L. & Schluter, D. 2012. The probability of genetic parallelism and convergence in natural populations. *Proc. Biol. Sci.* **279**: 5039–5047.

Cruickshank, T.E. & 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.

Dutoit, L., Vijay, N., Mugal, C.F., Bossu, C.M., Burri, R., Wolf, J. *et al.* 2017. Covariation in levels of nucleotide diversity in homologous regions of the avian genome long after completion of lineage sorting. *Proc. Biol. Sci.* **284**: 20162756.

Elgvin, T.O., Trier, C.N., Tørresen, O.K., Hagen, I.J., Lien, S., Nederbragt, A.J. *et al.* 2017. The genomic mosaicism of hybrid speciation. *Sci. Adv.* **3**: e1602996.

Ellegren, H. & Wolf, J.B.W. 2017. Parallelism in genomic landscapes of differentiation, conserved genomic features and the role of linked selection. *J. Evol. Biol.*, **30**: 1516–1518.

Elmer, K.A. 2017. Barrier loci and progress toward evolutionary generalities. *J. Evol. Biol.*, **30**: 1491–1493.

Elmer, K.R. & Meyer, A. 2011. Adaptation in the age of ecological genomics: insights from parallelism and convergence. *Trends Ecol. Evol.* **26**: 298–306.

Feder, J., Nosil, P., Gompert, Z., Flaxman, S. & Schiling, M. 2017. Barnacles, barrier loci, and the systematic building of species. *J. Evol. Biol.*, **30**: 1494–1497.

Feulner, P.G.D. & De-Kayne, R. 2017. Genome evolution, structural rearrangements and speciation. *J. Evol. Biol.*, **30**: 1488–1490.

Flaxman, S.M., Wacholder, A.C., Feder, J.L. & Nosil, P. 2014. Theoretical models of the influence of genomic architecture on the dynamics of speciation. *Mol. Ecol.* **23**: 4074–4088.

Fraïsse, C., Gunnarsson, P.A., Roze, D., Bierne, N. & Welch, J.J. 2016. The genetics of speciation: insights from Fisher's geometric model. *Evolution* **70**: 1450–1464.

Hoekstra, H.E. & Coyne, J.A. 2007. The locus of evolution: evo devo and the genetics of adaptation. *Evolution* **61**: 995–1016.

Jiggins, C. & Martin, S. 2017. Glittering gold and the quest for Isla de Muerta. *J. Evol. Biol.*, **30**: 1509–1511.

Lande, R. 1981. The minimum number of genes contributing to quantitative variation between and within populations. *Genetics* **99**: 541–553.

Lindkte, D. & Yeaman, S. 2017. Identifying the loci of speciation: the challenge beyond genome scans. *J. Evol. Biol.*, **30**: 1478–1481.

Lohse, K. 2017. Come on feel the noise - from metaphors to null models. *J. Evol. Biol.*, **30**: 1506–1508.

Martin, S.H., Dasmahapatra, K.K., Nadeau, N.J., Salazar, C., Walters, J.R., Simpson, F., *et al.* 2013. Genome-wide evidence for speciation with gene flow in Heliconius butterflies. *Genome Res.* **23**: 1817–1828.

Muirhead, C.A. & Presgraves, D.C. 2016. Hybrid incompatibilities, local adaptation, and the genomic distribution of natural introgression between species. *Am. Nat.* **187**: 249–261.

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 (Edinb)* **103**: 439–444.

Ono, J., Gerstein, A.C. & Otto, S.P. 2017. Widespread genetic incompatibilities between first-step mutations during parallel adaptation of *Saccharomyces cerevisiae* to a common environment. *PLoS Biol.* **15**: e1002591.

Ortiz-Barrientos, D. & James, M.E. 2017. Evolution of recombination rates and the genomic landscape of speciation. *J. Evol. Biol.*, **30**: 1519–1521.

Presgraves, D.C. 2010. The molecular evolutionary basis of species formation. *Nat. Rev. Genet.* **11**: 175–180.

Ravinet, M., Faria, R., Butlin, R.K., Galindo, J., Bierne, N., Rafajlović, M. *et al.* 2017. Interpreting the genomic landscape of speciation: a road map for finding barriers to gene flow. *J. Evol. Biol.*, **30**: 1450–1477.

Riesch, R., Muschick, M., Lindtke, D., Villoutreix, R., Comeault, A.A., Farkas, T.E. *et al.* 2017. Transitions between phases of genomic differentiation during stick-insect speciation. *Nat. Ecol. Evol.* **1**: 82.

Rieseberg, L.H. & Buerkle, C.A. 2002. Genetic mapping in hybrid zones. *Am. Nat.* **159**: S36–S50.

Roux, C., Fraïsse, C., Castric, V., Vekemans, X., Pogson, G.H. & Bierne, N. 2014. Can we continue to neglect genomic variation in introgression rates when inferring the history of speciation? A case study in a *Mytilus* hybrid zone. *J. Evol. Biol.* **27**: 1662–1675.

Roux, C., Fraisse, C., Romiguier, J., Anciaux, Y., Galtier, N. & Bierne, N. 2016. Shedding light on the grey zone of speciation along a continuum of genomic divergence. *PLoS Biol.* **14**: e2000234.

Vijay, N., Bossu, C.M., Poelstra, J.W., Weissensteiner, M.H., Suh, A., Kryukov, A.P. *et al.* 2016. Evolution of heterogeneous genome differentiation across multiple contact zones in a crow species complex. *Nat. Commun.* **7**: 13195.

Wagner, C.E. & Mandeville, E.G. 2017. Speciation, species persistence, and the goals of studying genomic barriers to gene flow. *J. Evol. Biol.*, **30**: 1512–1515.

Yang, J., Manolio, T.A., Pasquale, L.R., Boerwinkle, E., Caporaso, N., Cunningham, J.M. *et al.* 2011. Genome partitioning of genetic variation for complex traits using common SNPs. *Nat. Genet.* **43**: 519–525.

Yeaman, S. & Whitlock, M.C. 2011. The genetic architecture of adaptation under migration–selection balance. *Evolution* **65**: 1897–1911.

Received 31 May 2017; accepted 4 June 2017