MOLECULAR ECOLOGY
RESOURCES

Molecular Ecology Resources (2017) **17**, 362–365 doi: 10.1111/1755-0998.12669

COMMENT

Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations

JULIAN M. CATCHEN,* PAUL A. HOHENLOHE,† LOUIS BERNATCHEZ,‡ W. CHRIS FUNK,§ KIMBERLY R. ANDREWS¶ and FRED W. ALLENDORF**

*Department of Animal Biology, University of Illinois at Urbana–Champaign, Urbana, IL 61801, USA, †Department of Biological Sciences, Institute for Bioinformatics and Evolutionary Studies, University of Idaho, 875 Perimeter Drive, Moscow, ID 83844, USA, ‡Institut de Biologie Intégrative et des Systèmes (IBIS), Université Laval, Québec, QC G1V 0A6, Canada, §Department of Biology, Graduate Degree Program in Ecology, Colorado State University, Fort Collins, CO 80523, USA, ¶Department of Fish and Wildlife Sciences, University of Idaho, 875 Perimeter Drive MS 1136, Moscow, ID 83844, USA, **Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA*

Abstract

Recently, Lowry et al. addressed the ability of RADseq approaches to detect loci under selection in genome scans. While the authors raise important considerations, such as accounting for the extent of linkage disequilibrium in a study system, we strongly disagree with their overall view of the ability of RADseq to inform our understanding of the genetic basis of adaptation. The family of RADseq protocols has radically improved the field of population genomics, expanding by several orders of magnitude the number of markers available while substantially reducing the cost per marker. Researchers whose goal is to identify regions of the genome under selection must consider the LD of the experimental system; however, there is no magical LD cutoff below which researchers should refuse to use RADseq. Lowry et al. further made two major arguments: a theoretical argument that modeled the likelihood of detecting selective sweeps with RAD markers, and gross summaries based on an anecdotal collection of RAD studies. Unfortunately, their simulations were off by two orders of magnitude in the worst case, while their anecdotes merely showed that it is possible to get widely divergent densities of RAD tags for any particular experiment, either by design or due to experimental efficacy. We strongly argue that RADseq remains a powerful and efficient approach that provides sufficient marker density for studying selection in many natural populations. Given limited resources, we argue that researchers should consider a wide range of trade-offs among genomic techniques, in light of their study question and the power of different techniques to answer it.

Keywords: genome scan, linkage disequilibrium, RADseq, selection

Received 4 January 2017; revision accepted 9 March 2017

Recently, Lowry *et al.* (2016) addressed the ability of RADseq approaches to detect loci under selection in genome scans. While the authors raise an important consideration for designing studies and interpreting RADseq data, we strongly disagree with their overall view of the ability of RADseq to inform our understanding of the genetic basis of adaptation. RADseq is one of several techniques for population genomic studies, and all of them come with important trade-offs and limitations. Which approach is best depends on the goals of the study, as well as the biology of the organism, including the extent of linkage disequilibrium (LD) across the genome, as Lowry *et al.* (2016) emphasize. However, we believe that RADseq remains well suited for a wide range of systems and questions, including genome scans for adaptive variation. In particular, RADseq protocols have a large degree of flexibility for tailoring sampling and study design for particular systems (Andrews *et al.* 2016), and accounting for factors such as LD, and they have demonstrated their potential to identify genetic signatures of selection in nature.

We do agree with Lowry and coauthors on some points. The family of RADseq protocols has radically improved the field of population genomics. Building on previous marker technologies, such as allozymes, microsatellites and AFLPs, RAD protocols expanded by several orders of magnitude the number of markers available while substantially reducing the cost per

Correspondence: Julian Catchen, Fax: +1-217-244-1224;
E-mail: jcatchen@illinois.edu

marker and the number of person hours required to discover and genotype them. This reduction in labour and cost enabled a significant expansion in experimental sample sizes. In brief, it is not an exaggeration to say that RADseq protocols 'democratized' the field of population genomics. RADseq protocols have been widely applied for studies of phylogeny, phylogeography, hybridization, demography, population assignment and genetic mapping (Narum *et al.* 2013), importantly opening experimental avenues for nonmodel organisms.

We also appreciate Lowry and colleagues' attention to the ability of RADseq to detect loci under selection in genome scans, given the density of markers and the extent of linkage disequilibrium (LD). Researchers whose goal is to identify regions of the genome under selection must consider the LD of the experimental system. However, there is no magical LD cutoff below which researchers should refuse to use RADseq to address questions related to selection or adaptation. Rather, results should be presented in the context of the experimental characteristics known about the system (including LD), given the available data. It is worth noting that the extent of LD can be empirically measured using RADseq if a reference genome assembly or dense genetic map is available; that is, researchers can directly estimate their power to detect adaptive loci with a given marker density (Leitwein *et al.* 2016). Such direct estimates of LD are not possible with either of the alternative methods recommended by Lowry and colleagues – Pool-Seq (Schlotterer *et al.* 2014), which sacrifices individual-level genotypes, or whole-genome sequencing, which is typically limited to a relatively small number of individuals.

Regardless, even with small estimates of LD, RADseq may detect targets of selection. For example, the *Eda* locus in threespine stickleback has been repeatedly identified as a strong target of divergent selection in many independent RADSeq studies (Hohenlohe *et al.* 2010; Roesti *et al.* 2012; Ferchaud & Hansen 2016) – using less frequent cutters (SbfI, 8 bp) than Lowry *et al.* deem workable. Other factors, such as the presence of structural variants, may create linkage block outliers and provide a clear signal of selection (Corbett-Detig & Hartl 2012; Roesti *et al.* 2015). In addition, the study goal may not be to identify most or all loci under selection across the genome; rather, a common goal is often to test whether there is any evidence for adaptive differentiation within the genomic regions tested, and what the geographic distribution of such variation is (Funk *et al.* 2012; White *et al.* 2013; Pavey *et al.* 2015; Ferchaud & Hansen 2016). Even when the goal is to find most or all adaptive loci, the LD issue is not a limiting factor for genome scans using RADseq in many systems (see McKinney *et al.* (2016) for a detailed discussion). In particular, this includes many vertebrates and species of conservation concern that have high LD as a result of small effective population sizes, where RADseq provides an attractive option because no prior genomic information is required. In the end, Lowry and colleagues' suggestion that 'only recent hard sweeps from new mutations can realistically be detected' with RADseq is unwarranted, and in fact, this statement is in sharp contrast with the recent empirical literature (Bernatchez 2016).

The basis of Lowry and colleagues' conclusions rests on two arguments, theoretical and anecdotal. In an earlier, broader work by many of the same authors (Hoban *et al.* 2016), the evidence with respect to the effectiveness of RAD in sampling a genome is based on a simulation carried out by Tiffin & Ross-Ibarra (2014) (their box 2). The authors show that detecting sweeps given variable strengths of selection and recombination rates was very difficult for RADseq, with only the densest set of SNPs labelling enough haplotype blocks to have an appreciable chance of detecting sweeps. Unfortunately, Tiffin and Ross-Ibarra made a fundamental error in their calculations (Fig. S1, Supporting information) placing their simulations off by two orders of magnitude in the worst case. This error simulated the likelihood of finding selective sweeps that were less than 20 nucleotides in length – a challenge for any technology. Correcting this technical error to simulate more realistic sweep lengths of 200 to 200K nucleotides in length brightens the outlook for RADseq, providing a number of selection and recombination rate combinations that will provide good power to detect haplotype blocks influenced by selection. As Lowry *et al.* strongly relied on Tiffin & Ross-Ibarra (2014) to support their claim that LD patterns present 'major pitfalls' for RADseq, perhaps the corrected simulations will encourage them to moderate their view. The authors also did additional modelling to show how much of the genome is captured in RAD vs. other technologies, but as McKinney *et al.* (2016) have already addressed this subject, we will not comment further.

For the second, anecdotal argument, Lowry *et al.* present a table of recent experimental results from which they calculate the average number of RAD tags per megabase (4.08 tags/Mb, Table S1). However, it is well known that the number of RAD loci obtained for a given species can vary widely based on numerous aspects of experimental design including the restriction enzyme(s) used, the width and accuracy of the size selection, the amount of sequencing effort and the filtering criteria employed (e.g. minimum depth and minimum allele frequency cutoffs). This is clearly illustrated by the three studies of threespine stickleback listed in their table; the number of RAD loci obtained for these three studies ranges from 1879 to 166 711. Lowry and colleagues have merely shown that it is possible to get widely divergent densities of RAD tags for any particular experiment, either by design or due to experimental efficacy. Therefore, averaging locus densities across studies provides little meaningful information regarding the maximum possible density obtainable for a given species using RADseq.

Obtaining a full understanding of the genetic basis of adaptation is exceedingly difficult, and all genomic techniques face limitations. RADseq and other reduced representation approaches, such as RNAseq or sequence capture, necessarily do not sample a large proportion of the genome. The trade-off among them is that RADseq provides a random sample of the genome, which will include a subsample of coding, noncoding and regulatory regions, while transcriptome sequencing or exon capture focus on coding regions, and as such minimally inform about potentially important evolutionary change in regulatory regions. While the relative contribution of coding vs. regulatory regions still remains an open question (Hoekstra & Coyne 2007), biasing the genomic sampling a priori simply cannot address this question and may actually provide a biased view of the genomic determinants of evolutionary change. Furthermore, without neutral loci to define a null expectation, it is impossible to identify loci with 'outlier' behaviour that may be under divergent or stabilizing selection. These methods also differ widely in how much genomic information is required a priori; unless working in a model organism, sequence capture requires the researcher to identify and design the capture baits, while RNAseq is reliant on whatever genes were expressed in the samples collected. Whole-genome resequencing samples a much smaller number of individuals for a given total sequencing effort, so while it can sample all LD blocks, it relies heavily on assumptions that the individuals sampled are representative of the populations under study (e.g. Jones *et al.* (2012)). Pooled sequencing of various library types, while cost-efficient, carries inherent risks and limitations, particularly in the absence of a well-characterized genome (Schlotterer *et al.* 2014; Andrews *et al.* 2016). No matter the technology, the field of population genomics is cursed with the presence of many *intractable* genomes – those that are exceptionally large or complex, or from organisms that are difficult to obtain DNA samples or hard to experimentally manipulate – which present challenges to all of the above methods for the foreseeable future.

In the end, Lowry and coauthors do not provide convincing evidence in favour of these alternatives. In our opinion, the biggest factor affecting this area of science is simply economics. It is funding that provides access to bigger sample sizes, denser SNP discovery and personnel to handle the complex bioinformatics required to synthesize both. Given limited resources, we argue that researchers should consider a wide range of trade-offs among genomic techniques, in the light of their study question and the power of different techniques to answer it. The extent of LD is certainly one of these considerations, but it varies by orders of magnitude across taxa. Researchers should be explicit in their expectations of the extent of LD in their system, which can be either estimated directly from RADseq or other genomic data, or estimated based on related taxa, knowledge of demographic history or other biological information.

Conclusions from population genomic studies should always be tempered based on their power to detect effects, but we strongly argue that RADseq remains a powerful and efficient approach that provides sufficient marker density for studying selection in many natural populations. Funding work in nonmodel organisms has always been difficult and as researchers, we should support any method that can provide new data on systems that were not previously tractable – even if those data are not perfect. We do not have the liberty of methodological partisanship. Instead of focusing on sterile technical debates, we should pay more attention to the conceptual and theoretical basis that is needed to interpret any genome-wide data sets and ask the most relevant questions (Allendorf 2017).

Acknowledgements

We thank Garrett Mckinney, Wesley Larson, Lisa Seeb and Jim Seeb for their useful comments and discussion on this manuscript. We thank Eric Johnson for assistance with the selective sweep simulations.

References

Allendorf FW (2017) Genetics and the conservation of natural populations: allozymes to genomes. *Molecular Ecology*, **26**, 420–430.

Andrews KR, Good JM, Miller MR, Luikart G, Hohenlohe PA (2016) Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81–92.

Bernatchez L (2016) On the maintenance of genetic variation and adaptation to environmental change: considerations from population genomics in fishes. *Journal of Fish Biology*, **89**, 2519–2556.

Corbett-Detig RB, Hartl DL (2012) Population genomics of inversion polymorphisms in *Drosophila melanogaster*. *PLoS Genetics*, **8**, e1003056, doi:10.1371/journal.pgen.1003056.

Ferchaud AL, Hansen MM (2016) The impact of selection, gene flow and demographic history on heterogeneous genomic divergence: threespine sticklebacks in divergent environments. *Molecular Ecology*, **25**, 238–259.

Funk WC, McKay JK, Hohenlohe PA, Allendorf FW (2012) Harnessing genomics for delineating conservation units. *Trends in Ecology & Evolution*, **27**, 489–496.

Hoban S, Kelley JL, Lotterhos KE *et al.* (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. *American Naturalist*, **188**, 379–397.

Hoekstra HE, Coyne JA (2007) The locus of evolution: evo devo and the genetics of adaptation. *Evolution*, **61**, 995–1016.

Hohenlohe PA, Bassham S, Etter PD, Stiffler N, Johnson EA, Cresko WA (2010) Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS Genetics*, **6**, e1000862, doi:10.1371/journal.pgen.1000862.

Jones FC, Grabherr MG, Chan YF *et al.* (2012) The genomic basis of adaptive evolution in threespine sticklebacks. *Nature*, **484**, 55–61.

Leitwein M, Guinand B, Pouzadoux J, Desmarais E, Berrebi P, Gagnaire P-A (2017) A dense brown trout (Salmo trutta) linkage map reveals recent chromosomal rearrangements in the Salmo genus and the impact of selection on linked neutral diversity. *G3: Genes, Genomes, Genetics*, 2160–1836, doi: 10.1534/g3.116.038497.

Lowry DB, Hoban S, Kelley JL *et al.* (2017) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. *Molecular Ecology Resources*, **17**, 142–152, doi:10.1111/1755-0998.12635.

McKinney GJ, Larson WA, Seeb LW, Seeb JE (2016) RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2017). *Molecular Ecology Resources*, doi: 10.1111/1755-0998.12649

Narum SR, Buerkle CA, Davey JW, Miller MR, Hohenlohe PA (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841–2847.

Pavey SA, Gaudin J, Normandeau E *et al.* (2015) RAD sequencing highlights polygenic discrimination of habitat ecotypes in the panmictic American eel. *Current Biology*, **25**, 1666–1671.

Roesti M, Hendry AP, Salzburger W, Berner D (2012) Genome divergence during evolutionary diversification as revealed in replicate lake-stream stickleback population pairs. *Molecular Ecology*, **21**, 2852–2862.

Roesti M, Kueng B, Moser D, Berner D (2015) The genomics of ecological vicariance in threespine stickleback fish. *Nature Communications*, **6**, 8767, doi:10.1038/ncomms9767

Schlotterer C, Tobler R, Kofler R, Nolte V (2014) Sequencing pools of individuals – mining genome-wide polymorphism data without big funding. *Nature Reviews Genetics*, **15**, 749–763.

Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. *Trends in Ecology & Evolution*, **29**, 673–680.

White TA, Perkins SE, Heckel G, Searle JB (2013) Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. *Molecular Ecology*, **22**, 2971–2985.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Fig. S1 A. Probability of finding 50% of 10 sweeps under different parameter values. Recreated Figure 1B, Tiffin & Ross-Ibarra 2014. Based on the values of *s* and *c*, we have overlayed the size of the simulated sweeps. B. Corrected recombination rate: Probability of finding 50% of 10 sweeps under different parameter values. C. Original, public R Code to simulate the sweeps. D. Modified R Code to simulate the sweeps.