Title

Responsible RAD: Striving for best practices in population genomic studies of adaptation.

Journal

Molecular Ecology Resources, 17(3)

Authors

Lotterhos, Katie

Reed, Laura

Antolin, Michael

et al.

Publication Date

2017-05-01

DOI

10.1111/1755-0998.12677

Peer reviewed

Published in final edited form as:

Mol Ecol Resour. 2017 May ; 17(3): 366–369. doi:10.1111/1755-0998.12677.

Responsible RAD: Striving for best practices in population genomic studies of adaptation

David B. Lowry[a,b,*], **Sean Hoban**[c,d], **Joanna L. Kelley**[e], **Katie E. Lotterhos**[f], **Laura K. Reed**[g], **Michael F. Antolin**[h], **Andrew Storfer**[e]

[a]Department of Plant Biology, Michigan State University, East Lansing MI 48824, USA

[b]Program in Ecology, Evolutionary Biology, and Behavior, Michigan State University, East Lansing, MI 48824, USA

[c]The Morton Arboretum, Lisle, IL, USA

[d]National Institute for Mathematical and Biological Synthesis (NIMBioS), Knoxville, TN, USA

[e]School of Biological Sciences, Washington State University, Pullman, WA 99164, USA

[f]Department of Marine and Environmental Sciences, Northeastern University Marine Science Center, 430 Nahant Rd., Nahant, MA 01908, USA

[g]Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, 35406, USA

[h]Department of Biology, Colorado State University, Fort Collins, CO 80523-1878, USA

Keywords

Genotyping by sequencing; genome-environment association; genome scan; local adaptation; outlier analysis

We thank McKinney et al. (2017) and Catchen et al. (2017) for their recent commentaries on Lowry et al (2017), “Breaking RAD: An evaluation of the utility of restriction site associated DNA sequencing scans of adaptation.” Both articles argued that we overstated the limitations of RADseq for studies of local adaptation and McKinney et al. (2017) argued that “RADseq provides unprecedented insights” into mechanisms of evolutionary adaptations. While we agree with some of the comments made by the authors of the two papers, we still believe caution should be employed in RADseq studies that aim to detect loci that contribute to adaptation. In this rebuttal, we evaluate the key points made in these papers, attempt to identify a middle ground, and make suggestions for responsibly conducting future studies to understand the genomic mechanisms of adaptation.

In general, we agree with both of the commentaries that RADseq has been useful in the field of molecular ecology, particularly for understanding population structure and demography, and for conducting Quantitative Trait Locus (QTL) mapping; all applications that we acknowledged in Lowry et al. (2017). However, the key point of contention is not whether

*To whom correspondence should be addressed: Department of Plant Biology, Michigan State University, Plant Biology Laboratories, 612 Wilson Road, Room 166, East Lansing MI 48824, USA, dlowry@msu.edu.

RADseq is a useful tool for molecular ecology (obviously, it is), but whether or not it is a generally useful tool across species for advancing our understanding of the mechanisms of adaptation. For the remainder of this rebuttal, we focus on this key issue.

First of all, we want thank Catchen et al. (2017) for identifying and correcting the error in the model that was included in a recent review by Tiffin & Ross-Ibarra (2014). However, we should note that none of our arguments are based on that model and we made no mention of the model in Lowry et al. (2017). A correction for the Tiffin & Ross-Ibarra model has now been posted online (http://rpubs.com/rossibarra/257207). While correcting this error appears on the surface to improve the situation for RADseq, we urge the community to interpret this development with caution for a number of reasons. First, the model by Tiffin & Ross-Ibarra (2014) only pertains to recent hard sweeps. As we mentioned in Lowry et al. (2017), hard sweeps are just one of multiple ways that adaptation can occur, and if only hard sweeps can be detected, it can bias our view of how adaptation actually occurs. Second, the model was highly simplified, and like our model in Lowry et al. (2017), represents a best case scenario. Numerous other problems arise if you relax the simplifying assumptions of these models. We clearly outlined the assumptions for our model in Lowry et al. (2017), as does Ross-Ibarra for their model (http://rpubs.com/rossibarra/257207).

We readily acknowledge that many outlier loci likely do reflect signals of adaptation. We have published genomic scan studies of our own where we argued that detected outliers are potentially adaptive loci (Epstein et al. 2016; Yeaman et al. 2016; Gould et al. 2017). These studies include our recent paper on Tasmanian devils (Epstein et al. 2016), which McKinney et al. (2017) held up as an example of success. RADseq was effective in this study for two important reasons: the authors were careful to develop a large number of RAD-tags relative to the scale of linkage disequilibrium (LD), and the availability of a reference genome assisted in interpretation of results. Tasmanian devils are highly inbred with extended LD (>200kb), which facilitates identification of candidate genes in the vicinity of outlier loci. RADseq is predicted to sample the genome well in this scenario by the simple model in Lowry et al. (2017). Further, an annotated reference genome for Tasmanian devils allowed the linkage relationships among RAD-tags to be assessed and aided in identifying putative targets of selection. Catchen et al. (2017) also correctly point out that large-effect adaptation genes, like *EDA* in sticklebacks, have repeatedly been identified by RADseq outlier studies. We can be confident of outliers like *EDA* because there is additional data supporting the role of this gene in adaptation, as it was originally identified by QTL cloning (Colosimo et al. 2005). Further, RADseq outliers that co-localize with *EDA* are also believable because sticklebacks are a model organism with well-characterized patterns of LD (Jones et al. 2012). For many non-model study systems, we do not know the average length of LD and so cannot reasonably assess whether marker density is sufficient.

The argument by McKinney et al. (2017) that the LD situation for RADseq is improved by varying recombination across the genome is unsubstantiated because the dramatic variation in recombination rates, especially when poorly characterized, is a hindrance to effective genome scan studies. Low rates of recombination can lead to genomic regions with above average genetic differentiation, which “are often found in centromeric or rearranged regions where recombination is reduced” (Cruickshank & Hahn 2014). Further, poorly characterized

variation in LD among populations can inhibit interpretation of genome scan studies. Based on this uncertainty, McKinney et al. (2017) argued that “until actual LD values are known, we should not assume RAD genome scans are destined to fail.” Likewise, until actual LD values are known, we also should not assume that a given RADseq genomes scan study is destined to succeed. To alleviate concerns, researchers should aim to establish patterns of LD before engaging in any reduced representation-based genome scan study.

With sufficient marker density across the genome, RADseq can be an effective method. But, what does sufficient marker density mean? As we mentioned in Lowry et al. (2017), LD is not uniform throughout haplotype blocks. Instead, LD decays over the linear distance of chromosomes. The rough estimates that we made with our model (Lowry et al. 2017) are thus minimum estimates of the numbers of markers needed to conduct a study, while far more are likely needed to saturate the genome. Further, as genome size increases, the number of RAD-tags needed and the sequencing coverage needed for those tags become less cost-effective in terms of number of individuals per lane of sequencing compared to exon capture or transcriptome sequencing (Lowry et al 2017). Catchen et al. (2017) argued that very large genomes are currently “intractable” with available methods. However, a recent study by Yeaman et al. (2016), which conducted population genomic studies of trees with ~20 Gb genomes, suggests that exome capture is an effective solution for larger genomes. Further, although RADseq is now a “go-to” solution for many investigators, other methods of ascertainment can be as cost-effective and potentially result in better data for studying adaptation. For example, Gould et al. (2017) recently conduct a whole genome pooled-sequencing outlier study, which cost $6712 for library preparation and sequencing and resulted in 29,693,578 useful single nucleotide polymorphisms (4424 SNPs per $1).

McKinney et al. (2017) argued that if all of the species in Lowry et al (2017) Table S1 had an LD of 100 kb or greater, then 100% of the genome would be covered by 5000 RAD-tags. However, the length of LD has not actually been established for many of these species and at least 40% of the studies listed in Table S1 had fewer than 5000 polymorphic RAD-tags. Obviously, RADseq protocols can be altered to increase the number of markers, a point made by both McKinney et al. (2017) and Catchen et al. (2017). Yet, many studies still do not achieve the necessary numbers of polymorphic RAD-tags to sample a large fraction of the genome. In addition, the published literature may represent a biased sample of all attempted RADseq studies, as species with long haplotype blocks are more likely to sample a large portion of the genome with fewer RAD-tags and have successful results. The number of unpublished RADseq genome scans, i.e. failure to detect adaptive loci due to short LD, is unknown.

While genome-wide marker sparseness is a major concern for RADseq studies, there are additional important issues that apply to all types of genome scan studies that should not be overlooked. In a given genome scan study, where success is claimed because outliers are detected, how do we know that a given outlier is adaptive? Further, how do we assess what a set of outliers from a given study can tell us about the process of adaptation? Each outlier locus detected by any type of genome scan study is a hypothesis for local adaptation, but multiple alternative hypotheses may also account for the detected allele frequency differences among populations (Hoban et al 2016). For instance, because genomes

are heterogeneous, local differences in recombination rates and background selection across the genome can lead to differentiation of allele frequencies that may look like adaptive divergence (Charlesworth 2012). Other neutral, purely demographic processes, such as allele surfing (Excoffier et al. 2009), can also result in false positive outliers. From a statistical standpoint, there will always be outliers, but the challenge is to identify definitively the biological processes producing those outliers. Any type of genome outlier study, RADseq or otherwise, is just one of many types of studies that are necessary to understand the genetics of adaptation in a given system. This includes field experimentation (Clausen et al. 1940; Hereford 2009; Ågren & Schemske 2012; Savolainen et al. 2013) and follow-up functional molecular studies, which are necessary to prove that any outlier allele causes an adaptive phenotypic change (Colosimo et al. 2005). Without follow-up functional studies, individual outlier loci remain hypotheses and cannot advance our general understanding of the mechanisms of adaptation. Major scientific debates, such as the involvement of “coding vs. regulatory regions” (Catchen et al. 2017) in adaptation, are unlikely to be resolved by population genomic analyses alone.

In conclusion, we do not deny that RADseq is a useful tool. Rather, we advocate that it be used responsibly. Responsible usage of RADseq for outlier studies of adaptation means that researchers should follow a few basic principles. Many of these principles also apply to genome scans generally.

1. Acknowledge and report the potential study limitations. If a good estimate of LD exists for the study species, then researchers should report the proportion of the genome likely covered with polymorphic RAD-tags. If there is not a good estimate of LD, then try to maximize the number of polymorphic markers to help alleviate concerns. This is a major point of agreement with Catchen et al. (2017): “researchers should be explicit in their expectations of the extent of LD in their system,” and “results should be presented in the context of the experimental characteristics known about the system (including LD).”

2. Focus resources first on developing a reference genome or linkage map, if one does not exist. RADseq outlier studies conducted without a reference genome or linkage map will result in lists of anonymous candidate loci that do not alone advance our understanding of the mechanisms of adaptation. Researchers should be aware that new technologies have considerably reduced the expense of building reference genomes for non-model species (e.g. VanBuren et al. 2015).

3. Include field-based or lab-based evidence that demonstrate local adaptation. RADseq is one of many tools in the toolkit for understanding adaptation. Outliers generated by any type of genome scan study are hypotheses and are not definitive proof of adaptation. Functional molecular studies are still required to prove that any given allele has a true phenotypic effect. Authors should take care not to overstate the results of outlier loci from RADseq or any other ascertainment scheme. Moreover, field experiments are still the best way to demonstrate local adaptation (Hereford 2009; Ågren & Schemske 2012). This has been the case since Clausen, Keck, and Hiesey conducted their pioneering field experiments (Clausen et al. 1940).

4. Use a model, even a simple one, to evaluate the feasibility of a planned experiment and sampling scheme. If conservative assumptions are used, a model can help the investigator understand the trade-offs between different sampling and sequencing strategies. Simple models can be used for planning purposes to calculate the number of RAD-tags required to cover a genome and whether it will be cost-effective to obtain that number. Models can then be used to demonstrate to reviewers that a plan will be feasible, or used by reviewers to assess a given study. Neither McKinney et al. (2017) nor Catchen et al. (2017) criticized the simple model that we presented in Lowry et al. (2017) to compare experimental designs, and we urge readers to use that code to compare experimental designs for their study organisms before embarking on genomic scan studies.

Finally, we would like to note that Lowry et al. (2017) was not intended to attack or offend any investigator or their research program. While “Breaking RAD” was a controversial title, our aim was to encourage researchers to think critically about the hypothesis being tested, and to consider, among possible experimental designs, which design best tests a hypothesis based on the specific details of a system. We did not and do not advocate for any ascertainment method across all scenarios, only that investigators responsibly assess different ascertainment designs including RADseq, whole-genome pooled sequencing, RNA-seq, and sequence capture in the context of a study question, genome size, and expected patterns of LD. We do advocate for careful consideration of experimental designs and acknowledgement of errors when they occur. We acknowledge that we made errors in the original draft of Lowry et al. (2017) and we appreciate all those who brought those errors to our attention.

ACKNOWLEDGEMENTS

We would like to thank Jesse Lasky, Billie Gould, Jeffery Ross-Ibarra, Paul Hohenlohe, Shawn Narum, Angus Davison, and two anonymous reviewers for providing excellent comments that helped to improve the manuscript. This work was conducted as a part of the Computational Landscape Genomic working group at the National Institute for Mathematical and Biological Synthesis, sponsored by the National Science Foundation through NSF Award #DBI-1300426, with additional support from The University of Tennessee, Knoxville. In addition, SH was supported during this working group as a postdoctoral fellow at NIMBioS. Partial support for individuals provided by NSF DEB-1316549 to AS, NIH R01GM098856 to LKR, Army Research Office W911NF-15–1-0175 and NSF IOS-1557795 to JLK, by Michigan State University to DBL, and by Northeastern University to KEL.

REFERENCES

Ågren J, Schemske DW (2012) Reciprocal transplants demonstrate strong adaptive differentiation of the model organism Arabidopsis thaliana in its native range. New Phytologist, 194,1112–1122. [PubMed: 22432639]

Catchen JM, Hohenlohe PA, Bernatchez L, Funk WC, Andrews KR, Allendorf FW (2017) Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations.Molecular Ecology Resources, In Press.

Charlesworth B (2012) The effects of deleterious mutations on evolution at linked sites. Genetics, 190, 5–22. [PubMed: 22219506]

Clausen J, Keck DD, Hiesey WH (1940) Experimental studies on the nature of species I. Carnegie Inst. Wash. Publ. 520, Washington, DC, USA.

Colosimo PF, Hosemann KE, Balabhadra S, Villarreal G, Dickson M, Grimwood J, Schmutz J, Myers RM, Schluter D, Kingsley DM (2005) Widespread parallel evolution in sticklebacks by repeated fixation of ectodysplasin alleles. Science, 307, 1928–1933. [PubMed: 15790847]

Cruickshank TE, Hahn MW (2014) Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Molecular Ecology, 23, 3133–3157. [PubMed: 24845075]

Epstein B, Jones M, Hamede R, Hendricks S, McCallum H, Murchison EP, Schönfeld B, Wiench C, Hohenlohe P, Storfer A (2016) Rapid evolutionary response to a transmissible cancer in Tasmanian devils. Nature Communications, 7, 12684.

Excoffier L, Foll M, Petit RJ. Genetic consequences of range expansions. Annual Review of Ecology, Evolution, and Systematics, 40, 481–501.

Gould BA, Chen Y, Lowry DB (2017) Pooled ecotype sequencing reveals candidate genetic mechanisms for adaptive differentiation and reproductive isolation. Molecular Ecology, In Press.

Hereford J. 2009. A quantitative survey of local adaptation and fitness trade-offs. The American Naturalist, 173, 579–588.

Hoban S, Kelley JL, Lotterhos KE, Antolin MF, Bradburd G, Lowry DB, Poss ML, Reed LK, Storfer A, Whitlock MC (2016) Finding the genomic basis of local adaptation: pitfalls, practical solutions, and future directions. The American Naturalist, 188, 379–97.

Jones FC, Grabherr MG, Chan YF, Russell P, Mauceli E, Johnson J, Swofford R, Pirun M, Zody MC, White S, Birney E (2012) The genomic basis of adaptive evolution in threespine sticklebacks. Nature, 484, 55–61. [PubMed: 22481358]

Lowry DB, Hoban S, Kelley JL, Lotterhos KE, Reed LK, Antolin MF, Storfer A (2017) Breaking RAD: an evaluation of the utility of restriction site-associated DNA sequencing for genome scans of adaptation. Molecular Ecology Resources, In press

McKinney GJ, Larson WA, Seeb LW, Seeb JE (2017) RADseq provides unprecedented insights into molecular ecology and evolutionary genetics: comment on Breaking RAD by Lowry et al. (2017). Molecular Ecology Resources, In Press.

Savolainen O, Lascoux M, and Merilä J. (2013) Ecological genomics of local adaptations. Nature Reviews Genetics, 14, 807–820.

Tiffin P, Ross-Ibarra J (2014) Advances and limits of using population genetics to understand local adaptation. Trends in Ecology & Evolution, 29, 673–680. [PubMed: 25454508]

VanBuren R, Bryant D, Edger PP, Tang H, Burgess D, Challabathula D, Spittle K, Hall R, Gu J, Lyons E, Freeling M. (2015). Single-molecule sequencing of the desiccation-tolerant grass Oropetium thomaeum. Nature, 527, 508–511. [PubMed: 26560029]

Yeaman S, Hodgins KA, Lotterhos KE, Suren H, Nadeau S, Degner JC, Nurkowski KA, Smets P, Wang T, Gray LK, Liepe KJ, Hamann A, Holliday JA, Whitlock MC, Rieseberg LH, Aitken SN (2016) Convergent local adaptation to climate in distantly related conifers. Science, 353, 1431–1433. [PubMed: 27708038]