TITLE: Would an RRS by any other name sound as RAD?

AUTHORS

Erin O Campbell[1*], Bryan M T Brunet[1], Julian R Dupuis[2], and Felix A H Sperling[1]

AFFILIATIONS

[1]Department of Biological Sciences, CW405 BioSciences Centre, University of Alberta, Edmonton, Alberta, Canada T6G 2E9

[2]Department of Plant and Environmental Protection Sciences, University of Hawai'i at Mãnoa, Honolulu, Hawai'i, United States of America 96822

[*]CORRESPONDENCE: eocampbe@ualberta.ca

KEYWORDS

RADseq, GBS, reduced representation sequencing, RRL, next-generation sequencing

ABSTRACT

Sampling markers throughout a genome with restriction enzymes emerged in the 2000s as reduced representation shotgun sequencing (RRS). Rapid advances in sequencing technology have since spurred modifications of RRS, giving rise to many derivatives with unique names, such as RADseq. But naming conventions have often been more creative than consistent, with unclear criteria for recognition as a unique method resulting in a proliferation of names characterized by ambiguity. We conducted a literature review to assess methodological and etymological relationships among 36 restriction enzyme-based methods, as well as rates of correct referencing of commonly-used methods. We identify several instances of methodological convergence or misattribution in the literature, and note that many published derivatives have modified only minor elements of parent protocols. We urge greater restraint in naming derivative methods, to strike a better balance between clarity, recognition of scientific innovation, and correct attribution.

INTRODUCTION

Recent advances in next-generation sequencing (NGS) have given researchers access to unprecedented amounts of genomic data. The versatility of NGS, exemplified by its myriad applications to biology (Andrews *et al.* 2016), is arguably one of its greatest assets and has in turn led to more than 400 published methods that use this technology (Hadfield & Retief 2018). While NGS has indisputably spurred rapid innovation across biology, associated names have also proliferated. These names are commonly acronyms meant to clearly identify a methodology or application but, due to their sheer numbers, are now themselves a source of confusion. A set of suggested guidelines for the use of such acronyms was published several years ago (NUAP 2011), but Hadfield & Retief (2018) have recently reignited this conversation, discussing the excess of names for NGS methods but not analyzing their patterns of naming, publication, or citation

Methods that use restriction enzymes to sample genomes represent an informative subset of NGS techniques to explore in this context. These methods provide diverse options for reducing genomic complexity and surveying large numbers of loci across populations or species, and are widely used in ecology and evolutionary biology (Baird *et al.* 2008; Davey *et al.* 2011). Early approaches include reduced representation shotgun sequencing (RRS, Altshuler *et al.* 2000) and Complexity Reduction of Polymorphic Sequences (CRoPs™, van Orsouw *et al.* 2007), which have since served as springboards for derivative techniques, most of them published with unique names. At least 36 of these methods have been published as of December 2017 (Table S1).

Two methods in particular, Restriction site-associated DNA sequencing (RADseq, Baird *et al.* 2008) and Genotyping-by-Sequencing (GBS, Elshire *et al.* 2011), have been modified for diverse work on association and genetic mapping, population structure, and shallow-scale

phylogenetic relationships (Poland *et al.,* 2012; Baird *et al.*, 2013; Narum *et al.,* 2013; Eaton, 2014). They are so popular that 26 methods have explicitly been modified from either RADseq or GBS (Table S1), with their increasing importance demonstrated by recent reviews (Davey *et al.,* 2011; Andrews *et al*., 2016; Jiang *et al.* 2016), as well as debates (e.g. Andrews & Luikart (2014) and Puritz *et al.* (2014), and Lowry *et al.* (2016, 2017), McKinney *et al.* (2017), and Catchen *et al.* (2017)). Although reviews have tried to distinguish these approaches, it is clear from these publications as well as informal discussion on online forums (Table S2) that differences between many techniques are perceived to be minor and subtle. Naming conventions for derivatives have been variable and inconsistent, and literature discussing or employing these techniques has been ambiguous about the origins of techniques as well as which names to use as "catch-all" terms.

For instance, "GBS" is sometimes used to refer to all restriction-based methods collectively (e.g. Franchini *et al.* 2017), while other authors take the opposite approach and use "RADseq" as the generic term (e.g. Hoffberg *et al.* 2016). Two-enzyme, or double digest, adaptations of these techniques are similarly ambiguous; Peterson *et al.* (2012) has been credited with developing this approach (see Andrews *et al.* 2016) and did coin the term ddRAD to expand on the already-popular single-enzyme RADseq approach (Baird *et al*. 2008), but the CRoPS™ protocol (van Orsouw *et al.* 2007) was the first RRS method to do this. Some authors even give a common acronym a new meaning. For example, NextRAD (Fu *et al.* 2017) was developed by authors of the original RADseq papers (Miller *et al.* 2007; Baird *et al.* 2008) but uses RAD in this derived protocol to stand for Reductively-Amplified DNA. Thus, reduced representation genome sampling methods (hereafter referred to as RRS methods after Altshuler *et al.* 2000) exemplify the naming problem that is now typical of NGS methods.

Given the proliferation of RRS methods and ambiguity of their naming conventions (Jiang *et al.* 2016), we have sought to characterize RRS methods in a literature review and meta-analysis. We asked two main questions. First, what are the trends or criteria for naming new methods? And second, are researchers citing and referring to methods correctly? To answer these questions, we summarized the methodological and etymological relationships of these techniques in a concept map, and then conducted a meta-analysis of citation metrics to investigate patterns of literature referencing.

METHODS

Literature review and concept mapping

We compiled a list of RRS methods published on or before 31 December 2017 (N = 36), and evaluated approaches based on their methodological characteristics (Table S2). We then created a conceptual map of all methods, linking each derived technique to the main protocol that served as the basis for the modification, as specified by the authors (Fig. 1). In several cases a parent protocol was not directly specified, and in these instances, we linked methods based on overall methodological similarity. The subjective construction of this map reflects our experience as typical arms-length users of several of these approaches. Any technique that explicitly altered a protocol was considered a direct modification, and in this conceptualized map, a separate node. We plotted defining characteristics for each derivative along the connecting branches to assess distinctiveness or methodological convergence. Defining characteristics were generally those considered by the authors of the protocol to distinguish the derived method from its parent. To preserve clarity, characteristics that were highly variable across methods (for instance barcode and adaptor design and the overall order of methodological steps in each protocol) were not

plotted on the map unless they were definitive for the method(s). We also downloaded complete citation data from Web of Science® for the 36 methods, and determined the average number of citations per year for each publication. The size of ellipses in Fig. 1 reflects these numbers.

Meta-analysis of commonly used RRS approaches

To assess whether methods are recognized and attributed accurately in the literature, we reviewed all journal articles citing one- and two-enzyme RADseq and GBS (RADseq, Baird *et al.* 2008; GBS, Elshire *et al.* 2011; two-enzyme GBS, Poland *et al.* 2012; and ddRAD, Peterson *et al.* 2012), as well as Sequence-based Genotyping (SBG, Truong *et al.* 2012). RADseq, GBS, two-enzyme GBS and ddRAD are four of the most widely cited RRS approaches, and have been extensively modified to form the basis of many derivative methods. While the SBG protocol of Truong *et al.* (2012) is far less frequently cited, we included it for its methodological and etymological similarity to these methods as well as its date of publication, which occurred between that of Poland *et al.* (2012) and Peterson *et al.* (2012). It is also the subject of U.S. patent 8,815,512 B2, owned by KeyGene, which claims legal ownership and protection of all methods that simultaneously discover and genotype single nucleotide polymorphisms, including RADseq, GBS, two-enzyme GBS and ddRAD (KeyGene 2016).

Complete citation lists were downloaded from Web of Science® on 6 February 2018 for the period up to and including 31 December 2017 for each of Baird *et al.* (2008), Elshire *et al.* (2011), Poland *et al.* (2012), Truong *et al.* (2012), and Peterson *et al.* (2012). The lists were then combined and filtered to remove duplicates (Table S3). Only articles whose titles, abstracts, or keywords contained “GBS”, “SBG” or “RAD” (and all variant search strings in Table S4) were retained for further analysis. Incorrect name usage was defined as any case of an alternate name

being used to refer to a technique (e.g. "RAD" to describe the GBS protocol of Elshire *et al.* (2011)). Strings for "two-enzyme GBS" and "ddRAD" were not searched separately since these were treated as variants of "GBS" and "RAD", respectively. A complete description of the methods used in the literature review is in Table S6.

RESULTS

Construction of RRS conceptual map

Of the 36 RRS methods we examined, those of Baird *et al.* (2008), Peterson *et al.* (2012), and Elshire *et al.* (2011) are the precursors of the greatest number of directly derived methods (Fig. 1), and the most highly cited (Table S2). "RAD" was used in 18 named techniques, while "GBS" was used in 6 and the remaining 12 methods had names that lacked "RAD", "GBS", or any specified name at all (see Sonah *et al.* 2013 and Mascher *et al.* 2013). Many derived methods were named after the protocol they modified (*e.g.*: ddRAD (Peterson *et al.* 2012) from RADseq (Baird *et al.* 2008)), but there were several exceptions (*e.g.*, SBG (Truong *et al.* 2012) from CRoPS^{TM} (van Orsouw *et al.* 2007)). We observed multiple occurrences of methodological convergence across methods, including the use of paired restriction enzymes in double digest methods, sequence capture, bisulfite sequencing, and the use of PCR amplification to create reduced representation libraries, which we discuss below.

Meta-analysis of GBS, SBG and RAD citation accuracy

The number of journal articles that refer to "GBS", "SBG", or "RAD" within their title, abstract or keywords and uniquely cite either Baird *et al.* (2008), Elshire *et al.* (2011), Poland *et al.* (2012), Truong *et al.* (2012), or Peterson *et al.* (2012) has increased rapidly since 2010, with

the greatest number of citations occurring in 2017. Of a total of 788 journal articles, 335 (42.5%) refer only to GBS, 2 (0.2%) refer only to SBG, and 418 (53.1%) refer only to RAD (Fig. 2; Table S5). Two or more of these names ("Multiple(≥2)") are used in only 33 (4.3%) journal articles and these refer only to GBS and RAD, not SBG (Fig. 2; Table S5).

Each name has been used inconsistently to refer to methods described by their parent publications, but to varying degrees: 8% (28/349) of publications that uniquely cite Elshire *et al.* (2011) or Poland *et al.* (2012) refer to SBG or RAD alone or in combination with GBS; 66.7% (2/3) of publications that uniquely cite Truong *et al.* (2012) refer to GBS or RAD alone or in combination with SBG; and 8.3% (36/436) of publications that uniquely cite Baird *et al.* (2008) or Peterson *et al.* (2012) refer to GBS or SBG alone or in combination with RAD (Fig. 2; Tables S3 and S5). Thus, use of ambiguous or incorrect names are apparent in about 8.4% of journal articles citing these five papers.

DISCUSSION

We have applied a review and literature meta-analysis to characterize the naming and use of RRS methods. Our concept map shows that RAD-based methods are more numerous than GBS-based methods. Although derived methods are often given unique names, most follow some of the etymological elements of the parent technique that was modified, even when derived protocols from different camps converge methodologically (Fig. 1). We also identified a rate of ~8.4% ambiguous or incorrect citations for these methods (Fig. 2).

The RAD acronym leads the popularity race when considering citations for RAD-based methods as well as the number of derivative protocols bearing this term; GBS-based methods have fewer overall citations and methodological offspring. While the original RAD or modified

ddRAD methods may simply be more methodologically attractive, unconscious linguistic factors in the naming of derivatives may also be contributing to this trend. Acronyms that form simple, recognizable words are more likely to be remembered (NUAP 2011), so this may explain use of the RAD acronym despite citation of a GBS or SBG publication (Fig. 2). RAD has also proven to be easy to incorporate into memorable titles that improve name recognition and visibility in a rapidly expanding field (*e.g.* "Demystifying the RAD fad" (Puritz *et al.* 2014); "Breaking RAD" (Lowry *et al.* 2016); present study). However, rates of potential misattribution do not appear to be biased toward RAD over GBS (Fig. 2), and so researchers who are unclear or unconvinced of the distinctions between methods may simply be randomly using both terms as synonyms. Methodological convergence by several GBS- and RAD-based techniques (Fig. 1) could contribute to further ambiguity among methods.

"What's in a name?" (Shakespeare 1594-98). Separate publication of a method implies that the authors consider the new method to be substantively different from other published methods, thereby warranting a separate name. But for RRS methods, differences between many techniques are minor, often primarily implementing streamlined library preparation and cost reduction (e.g.: GGRS (Chen *et al.* 2013), ezRAD (Toonen *et al.* 2013)), or adaptations that optimize methods for specific groups of organisms (e.g.: MiddRAD, Yang *et al.* 2016). Many published methods arguably do not meet this criterion for publication (NUAP 2011). And while some RRS methods have made larger methodological changes, for instance the use of sequence capture or bisulfite sequencing, the publication of these methods is problematic for other reasons.

Naming of a method also suggests ownership over that method (NUAP 2011). In cases where only minor changes were made to an existing protocol, the authors of the new method profit from advances made by prior authors, which may comprise the bulk of the methodology.

191 The broad convergence of several methods in Fig. 1 creates an additional layer of complexity, as
192 two separate groups of authors are essentially claiming ownership over similar techniques that
193 have different names. For instance, ddRADseq-ion (Recknagel *et al.* 2015) and GBS for
194 semiconductor sequencing platforms (Mascher *et al.* 2013) have both incorporated double digests
195 and modified adaptors for Ion Torrent sequencing. EpiGBS (van Gurp *et al.* 2016) and bs-
196 RADseq (Trucchi *et al.* 2016) both incorporate bisulfite sequencing, and several methods have
197 employed some form of sequence capture (Spiked GBS (Rife *et al.* 2016); RADcap (Hoffberg *et*
198 *al.* 2016); HyRAD (Suchan *et al.* 2016); Rapture (Ali *et al.* 2016); 3RAD (Graham *et al.* 2015);
199 hyRADx (Schmid *et al.* 2017)).

200 While each technique is prone to distinct biases and technical difficulties (van Dijk *et al.*
201 2014; Flanagan & Jones 2017), we argue that most RRS methods are sufficiently similar that they
202 can be adapted to a user-specified combination of characteristics that suit the needs of an
203 individual study. The recently upheld US KeyGene patent covering these methods also seems to
204 suggest that, at least from a legal standpoint, they are not significantly different from one another
205 (US Patent 8,815,512 B2). So why do we keep naming derivative methods, and how can we work
206 to preserve clarity of communication in discussing them?

207 *"Action is eloquence"* (Shakespeare 1605-08). Rapid sequencing advances may have
208 unwittingly created a sense of momentum among researchers, thereby fostering the proliferation
209 of names for genome-sampling techniques. Almost half of the methods in Fig. 1 and all five of
210 the key methods in Fig. 2 were published in *PLoS ONE*, which has published several RRS
211 derivatives within the same years. Other journals have also published multiple derivative
212 methods, although to a lesser degree. Several researchers have also been involved the naming and
213 publication of more than one method, indicating research groups developing suites of techniques.

This suggests a preoccupation with name recognition as a means to increase the visibility of research. We recognize that catchy titles are not inherently negative or irresponsible, and that this practice can beneficially increase the impact of research. But the recent "modify, name, and publish" trend seems more likely to be driven by efforts to increase citations, which dilutes the eloquence of acronyms. This system further confers risk to researchers who choose not to name an adapted technique, by leaving a door open for someone else to employ the same change and name it, taking the credit. Because academic success is so closely tied to citation metrics, there is little incentive to take the high road.

It is instructive to compare reliance on easy-to-digest acronyms to online clickbait headlines in academic publishing and research. An example may be the recent publication of an incendiary essay presumably to increase the impact of a journal despite the article not passing peer review (Flaherty 2017). Academic metrics do not distinguish between "good" and "bad" citations (Gallien & Roelofs 2017), and we are incentivized to market our research beyond the merit of the research itself. Sequencing technology will undoubtedly continue to advance (Goodwin *et al.* 2016), and new RRS approaches will continue to evolve. Exploring the utility and limitations of these approaches has resulted in a wealth of biological knowledge that has been hitherto out of reach. At this level, Shakespeare's immortal phrase got it right: "a rose by any other name would smell as sweet" (Shakespeare 1591-94). But Linnaeus may have disagreed with this sentiment – names *do* matter because they serve to communicate and organize the world around us.

We add our voices to those of Hadfield & Retief (2018); our scientific community would be better served by greater restraint in naming new techniques, except for indisputably large methodological innovations. Continued adaptation of methods is clearly beneficial, but the

publication of new names for minor changes in existing NGS methodologies is a symptom of a larger cultural shift in academia. And the responsibility for righting that course lies with us as researchers, editors, and publishers.

AUTHOR CONTRIBUTIONS

EOC, BMTB, JRD, and FAHS contributed to the study design and draft revisions. EOC and BMTB conducted the literature review and meta-analysis, and wrote the manuscript.

ACKNOWLEDGEMENTS

This work was supported in part by an NSERC Discovery Grant to FAH Sperling (RES0016259).

COMPETING INTERESTS

The authors have no competing interests.

250 REFERENCES

251 Acquadro, A., Barchi, L., Portis, E., Carrasquilla-Garcia, N., Tirone, M., Lanteri, S., & Comino,
252 C. (2016) RAD2seq: an efficient protocol for plant genotyping by sequencing. *Acta*
253 *Horticulturae*, **1147**, 1-8.
254 Ali, O. A., O'Rourke, S. M., Amish, S. J., Meek, M. H., Luikart, G., Jeffres, C., & Miller, M. R.
255 (2016) RAD Capture (Rapture): Flexible and Efficient sequence-based genotyping.
256 *Genetics*, **202**, 389-400.
257 Altshuler, D., Pollara, V. J., Cowles, C. R., Van Etten, W. J., Baldwin, J., Linton, L., & Lander,
258 E. S. (2000) An SNP map of the human genome generated by reduced representation
259 shotgun sequencing. *Nature*, **407**, 513-516.
260 Andolfatto, P., Davison, D., Erezyilmaz, D., Hu, T. T., Mast, J., Sunayama-Morita, T., & Stern,
261 D. L. (2011) Multiplexed shotgun genotyping for rapid and efficient genetic mapping.
262 *Genome Research*, **21**, 610-617.
263 Andrews, K. R., Good, J. M, Miller, M. R., Luikart, G., & Hohenlohe, P. A. (2016) Harnessing
264 the power of RADseq for ecological and evolutionary genomics. *Nature Reviews*
265 *Genetics*, **17**, 81-92.
266 Andrews, K. R., & Luikart, G. (2014) Recent novel approaches for population genomics data
267 analysis. *Molecular Ecology*, **23**, 1661-1667.
268 Baird, N. A., Etter, P. D., Atwood, T. S., Currey, M. C., Shiver, A. L., Lewis, Z. A., Selker, E.
269 U., Cresko, W. A., & Johnson, E. A. (2008) Rapid SNP discovery and genetic mapping
270 using sequenced RAD markers. *PLoS One*, **3**, e3376.
271 https://doi.org/10.1371/journal.pone.0003376
272 Bayerl, H., Kraus, R.H.S., Nowak, C., Foerster, D.W., Fickel, J., & kuehn, R. (2017) Fast and
273 cost-effective single nucleotide polymorphism (SNP) detection in the absence of a
274 reference genome using semideep next-generation Random Amplicon Sequencing
275 (RAMseq). *Molecular Ecology Resources*, https://doi.org/10.1111/1755-0998.12717
276 Catchen, J. M., Hohenlohe, P. A., Bernatchez, L., Funk, W. C., Andrews, K. R., & Allendorf, F.
277 W. (2017) Unbroken: RADseq remains a powerful tool for understanding the genetics of
278 adaptation in natural populations. *Molecular Ecology Resources*, **17**, 362-365.
279 Chen, Q., Ma, Y., Yang, Y., Chen, Z., Liao, R., Xie, X., Wang, Z., He, P., Tu, Y., Zhang, X.,
280 Yang, C., Yang, H., Yu, F., Zheng, Y., Zhang, Z., Wang, Q., & Pan, Y. (2013)
281 Genotyping by Genome Reducing and Sequencing for Outbred Animals. *PLoS One* **8**,
282 e67500. https://doi.org/10.1371/journal.pone.0067500
283 Davey, J. W., Hohenlohe, P. A., Etter, P. D., Boone, J. Q., Catchen, J. M., & Blaxter, M. L.
284 (2011) Genome-wide genetic marker discovery and genotyping using next-generation
285 sequencing. *Nature Reviews Genetics*, **12**, 499-510
286 Eaton, D. A. R. (2014) PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses.
287 *Bioinformatics*, **30**, 1844-1849.
288 Elshire, R.J., Glaubitz, J. C., Poland, J. A., Kawamoto, K., Buckler, E. S., & Mitchell, S. E.
289 (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity
290 species. *PLoS One*, **6**, e19379. https://doi.org/10.1371/journal.pone.0019379
291 Flaherty, M. (2017) Resignations at 'Third World Quarterly'. *Inside Higher Ed.* Retrieved from
292 https://www.insidehighered.com/news/2017/09/20/much-third-world-quarterlys-editorial-
293 board-resigns-saying-controversial-article

294 Flanagan, S. P., & Jones, A. G. (2017) Substantial differences in bias between single-digest and
295 double-digest RAD-seq libraries: a case study. *Molecular Ecology Resources*,
296 https://doi.org/10.1111/1755-0998.12734
297 Franchini, P. Parera, D. M., Kautt, A. F., & Meyer, A. (2017) quaddRAD: a new high-
298 multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary
299 insights in a nonradiating cichlid lineage. *Molecular Ecology*, **28**, 2783-2795.
300 Fu, Z., Epstein, B., Kelley, J. L., Zheng, Q., Bergland, A. O., Carrillo, C. I. C., Jensen, A. S.,
301 Dahan, J., Karasev, A. V., & Snyder, W. E. (2017) Using NextRAD sequencing to infer
302 movement of herbivores among host plants. *PLoS One*, **12**, e0177742.
303 https://doi.org/10.1371/journal.pone.0177742
304 Gallien, M. and Roelofs, P. (2017) What the success of a journal article supporting colonialism
305 says about academia in a clickbait era. *Scroll.in*. Retrieved from
306 https://scroll.in/article/851215/after-the-media-the-lure-of-clickbait-has-taken-over-
307 academia
308 Goodwin, S., McPherson J. D., & McCombi, W. R. (2016) Coming of age: ten years of next-
309 generation sequencing technologies. *Nature Reviews Genetics***, 27**, 333-351.
310 Graham, C. F., Glenn, T. C., McArthur, A. G., Boreham, D. R., Kieran, T., Lance, S., Manzon, R.
311 G., Martino, J. A., Pierson, T., Rogers, S. M., Wilson, J. Y., & Somers, C. M. (2015)
312 Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADseq),
313 *Molecular Ecology Resources*, **15**, 1304-1315.
314 Guo, Y., Fang, D., Song, L., Liu, Y., Liu, Y., Wu, L., Yu, J., Li, Z., Xu, X., & Zhang, H. (2014)
315 An improbed 2b-RAD approach (I2B-RAD) offering genotyping tested by a rice (*Oryza*
316 *sativa* L.) F2 population. *BMC genomics*, **15**, 956.
317 Hadfield, J., & Retief, J. (2018) A profusion of confusion in NGS methods naming. *Nature*
318 *Methods*, **15**, 7-8.
319 Hilario, E., Barron, L., Deng, C. H., Datson, P. M., De Silva, N., Davy, M. W., & Storey, R. D.
320 (2015). Random tagging genotyping by sequencing (rtGBS), an unbiased approach to
321 locate restriction enzyme sites across the target genome. *PLoS One*, **10**, e0143193.
322 https://doi.org/10.1371/journal.pone.0143193
323 Hoffberg, S. L., Kieran, T. J., Catchen, J. M., Devault, A. Faircloth, B. C., Mauricio, R., &
324 Glenn, T. C. (2016) RADcap: sequence capture of dual-digest RADseq libraries with
325 identifiable duplicates and reduced missing data. *Molecular Ecology Resources*, **16**, 1264-
326 1278.
327 Jiang, Z., Wang, H., Michal, J. J., Zhou, X., Liu, B., Solberg Woods, L. C., & Fuchs, R. A.
328 (2016) Genome wide sampling sequencing for SNP genotyping: methods, challenges, and
329 future development. *Int. J. Biol. Sci.*, **12**, 100-108.
330 KeyGene (2016). *KeyGene's SBG patent upheld by the USPTO after ex parte reexamination*.
331 [online] Available at: http://www.keygene.com/wp-content/uploads/2016/03/Press-
332 release-KeyGenes-SBG-patent-upheld-by-the-USPTO-after-ex-parte-reexamination.pdf
333 [Accessed 20 Nov. 2017].
334 Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer,
335 A. (2016) Breaking RAD: an evaluation of the utility of restriction site-associated DNA
336 sequencing for genome scans of adaptation. *Molecular Ecology Resources*, **17**, 142-152.

Lowry, D. B., Hoban, S., Kelley, J. L., Lotterhos, K. E., Reed, L. K., Antolin, M. F., & Storfer, A. (2017) Responsible RAD: Striving for best practices in population genomic studies of adaptation. *Molecular Ecology Resources,* **17**, 366-369.

Mascher, M., Wu, S., St. Amand, P., Stein, N., & Poland, J. (2013) Application of genotyping-by-sequencing on semiconductor sequencing platforms: a comparison of genetic and reference-based marker ordering in barley. *PLoS One*, **8**, e76925. https://doi.org/10.1371/journal.pone.0076925

McKinney G. J., Larson, W. A., Seeb, L. W., & Seeb, J. E. (2017) RADseq provides unprecedented insights into molecular ecology and evolutionary insights: comment on Breaking RAD by Lowry *et al.* (2016). *Molecular Ecology Resources*, **17**, 356-361.

Miller, M. R., Dunham, J. P., Amores, A., Cresko, W. A., & Johnston, E. A. (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research*, **17**, 240-248.

Narum, S. R., Buerkle, C. A., Davey, J. W., Miller, M. R., & Hohenlohe, P. A. (2013) Genotyping-by-sequencing in ecological and conservation genomics. *Molecular Ecology*, **22**, 2841-2847.

NUAP (no unnecessary acronyms please). (2011) *Nature Methods*, **8**, 521.

Ott, A., Liu, S., Schnable, J.C., Yeh, C.-T., Wang, K.-S., & Schnable, P.S. (2017) tGBS® genotyping-by-sequencing enables reliable genotyping of heterozygous loci. *Nucleic Acids Research*, **45**, e178. https://doi.org/10.1093/nar/gkx853

Peterson, B. K., Weber, J. N., Kay, E. H., Fisher, H. S., & Hoekstra, H. E. (2012) Double digest RADseq: an inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PLoS One*, **7**, e37135. https://doi.org/10.1371/journal.pone.0037135

Poland, J. A., Brown, P. J., Sorrells, M. E., & Jannink, J. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *PLoS One*, **7**, e32253. https://doi.org/10.1371/journal.pone.0032253

Puritz, J. B., Matz, M. V., Toonen, R. J., Weber, J. N., Bolnick, D. I., & Bird, C. E. (2014) Demystifying the RAD fad. *Molecular Ecology*, **23**, 5937-5942.

Recknagel, H. Jacobs, A., Herzyk., P., & Elmer, K. R. (2015) Double-digest RAD sequencing using ion proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. *Molecular Ecology Resources*, **15**, 1316-1329.

Rife, T. W., Wu, S., Bowden, R. L., & Poland, J. A. (2015) Spiked GBS: a unified, open platform for single marker genotyping and whole-genome profiling. *BMC Genomics*, **16**, 248.

Schield, D.R., Walsh, M.R., Card, D.C., Andrew, A.L., Adams, R.H., Castoe, T.A., & Bunce, M. (2016) EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing. *Methods in Ecology and Evolution*, **7**, 60-69.

Schmid, S., Genevest, R., Gobet, E., Suchan, T., Sperisen, C., Tinner, W., & Alvarez, N. (2017) hyRAD-X, a versatile method combining exome capture and RAD sequencing to extract genomic information from ancient DNA. *Methods in Ecology and Evolution*, https://doi.org/10.1111/2041-210X.12785

Shakespeare, W. 1591-94. "Romeo and Juliet" *The complete works of William Shakespeare comprising his plays and poems*, The World Syndicate Publishing Company, 1930, pp. 901.

Shakespeare, W. 1605-08. "Coriolanus" *The complete works of William Shakespeare comprising his plays and poems*, The World Syndicate Publishing Company, 1930, pp. 703.

Sonah, H., Bastien, M., Iquira, E., Tardivel, A., Légaré, G., Boyle, B., Normandeau, É., Larose, S., Jean, M., & Belzile, F. (2013) An improved Genotyping by Sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *PLoS One*, **8**, e54603. https://doi.org/10.1371/journal.pone.0054603

Stolle, E., & Moritz, R. F. A. (2013) RESTseq - Efficient benchtop population genomics with RESTriction Fragment SEQuencing. *PLoS One*, **8**, e63960. https://doi.org/10.1371/journal.pone.0063960

Suchan, T., Pitteloud, C., Gerasimova, N. S., Kostikova, A., Schmid, S., Arrigo, N., Pajkovic, M., Ronikier, M., & Alvarez, N. (2016) Hybridization capture using RAD probes (hyRAD), a new tool for performing genomic analyses on collection specimens. *PLoS One*, **11**, e0151651. https://doi.org/10.1371/journal.pone.0151651

Sun, X., Liu, D., Zhang, X., Li, W., Liu, H., Hong, W., Jiang, C., Guan, N., Ma, C., Zeng, H., Xu, C., Song, J., Huang, L., Wang, C., Shi, J., Wang, R., Zheng, X., Lu, C., Wang, X., & Zheng, H. (2013) SLAF-seq: An efficient method of large-scale De Novo SNP discovery and genotyping using high-throughput sequencing. *PLoS One*, **8**, e58700. https://doi.org/10.1371/journal.pone.0058700

Suyama, Y., & Matsuki, Y. (2015) MIG-seq: an effective PCR-based method for genome-wide single nucleotide polymorphism genotyping using the next-generation sequencing platform. *Scientific Reports*, **5**, 16963. https://doi.org/10.1038/srep16963

Toonen, R. J., Puritz, J. B., Forsman, Z. H., Whitney, J. L., Fernandez-Silva, I., Andrews, K. R., & Bird, C. E. (2013) ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ*, **1**, e203. https://doi.org/10.7717/peerj.203

Trucchi, E., Mazzarella, A. B., Gilfillan, G. D., Lorenzo, M. T., Schönswetter, P., & Paun, O. (2016) BsRADseq: screening DNA methylation in natural populations of non-model species. *Molecular Ecology*, **25**, 1697-1713.

Truong, H. T., Ramos, A. M., Yalcin, F., de Ruiter, M., van der Poel H. J. A., Huvenaars, K. H. J., Hogers, R. C. J., van Enckevort, L. J. G., Janssen, A., van Orsouw, N. J., & van Eijk, J. T. (2012) Sequence-based genotyping for marker discovery and co-dominant scoring in germplasm and populations. *PLoS One*, **7**, e37565. https://doi.org/10.1371/journal.pone.0037565

van Dijk, E. L., Jaszczyszyn, Y., & Thermes, C. (2014) Library preparation methods for next-generation sequencing: Tone down the bias. *Experimental Cell Research,* **32**, 12-20.

van Gurp, T. P., Wagemaker, N. C. A. M., Wouters, B., Vergeer, P., Ouborg, J. N. J., & Verhoeven, K. J. F. (2016) epiGBS: reference-free reduced representation bisulfite sequencing. *Nature Methods*, **13**, 322-328.

van Orsouw, N. J., Hogers, R. C. J., Janssen, A., Yalcin, F., Snoeihers, S., Verstege, E., Schneiders, H., van der Poel, H., van Oeveren, J., Verstegen, H., & van Eijk, M. J. T. (2007) Complexity reduction of polymorphic sequences (CRoPs^{TM}): a novel approach for large-scale polymorphism discovery in complex genomes. *PLoS One*, **2**, e1172. https://doi.org/10.1371/journal.pone.0001172

Van Tassell, C. P., Smith, T. P. L., Matukumalli, L. K., Taylor, J. F., Schnabel, R. D., Lawley, C. T., Haudenschild, C. D., Moore, S. S., Warren, W. C., & Sonstegard, T. S. (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nature Methods*, **5**, 247-252.

Wang, S., Meyer, E., McKay, J. K., & Matz, M. V. (2012) 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods*, **9**, 808-812.

Yang, G., Chen, Y., Wang, J., Guo, C., Zhao, L., Wang, X., Guo, Y., Li, L., Li, D., & Guo, Z. (2016) Development of a universal and simplified ddRAD library preparation approach for SNP discovery and genotyping in angiosperm plants. *Plant Methods*, **12**, 39.

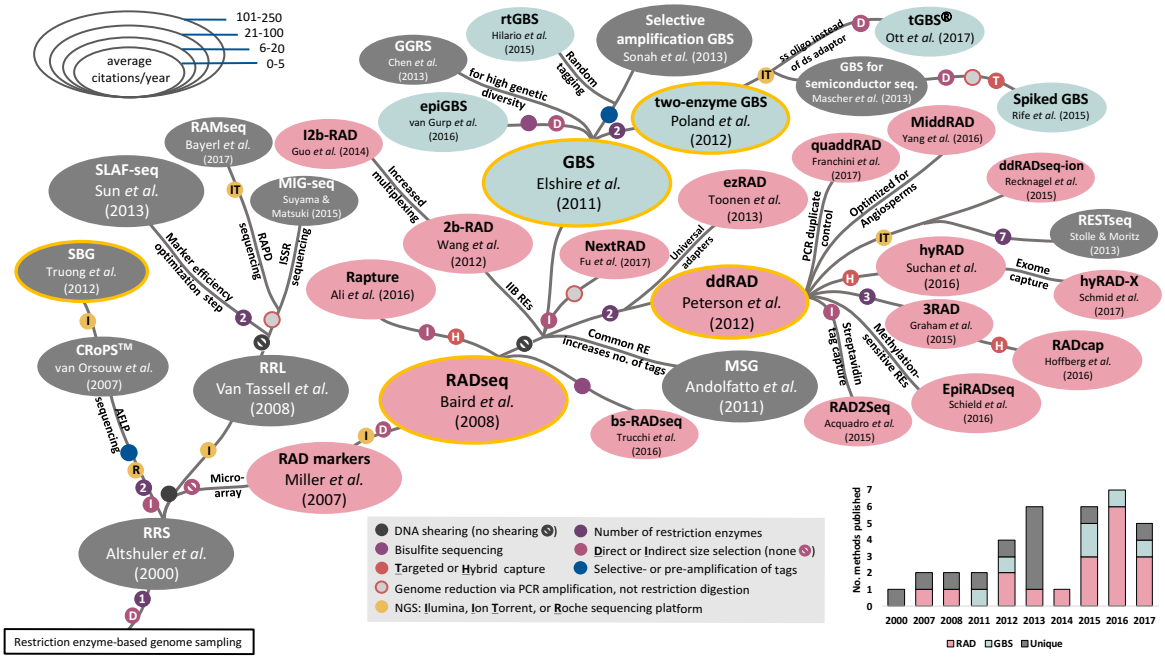

Figure 1. Concept map displaying methodological and etymological relationships among 36 reduced representation genome-sampling methods and their derivatives. Branches connect derived methods to their inferred parent protocols, and significant differences between protocols are indicated by coloured circles on branches. Variations that originate only once are indicated by text along branches. Red ellipses indicate named methods using the "RAD" acronym, blue ellipses indicate names derived from "GBS", and methods with unique names, or lacking names altogether, are in grey. The five methods used to assess attribution rates in an accompanying literature review (Fig. 2) are indicated by ellipses with a gold outline. Inset histogram shows accumulation of methods by year. See Table S1 for a summary of each method.

452

453

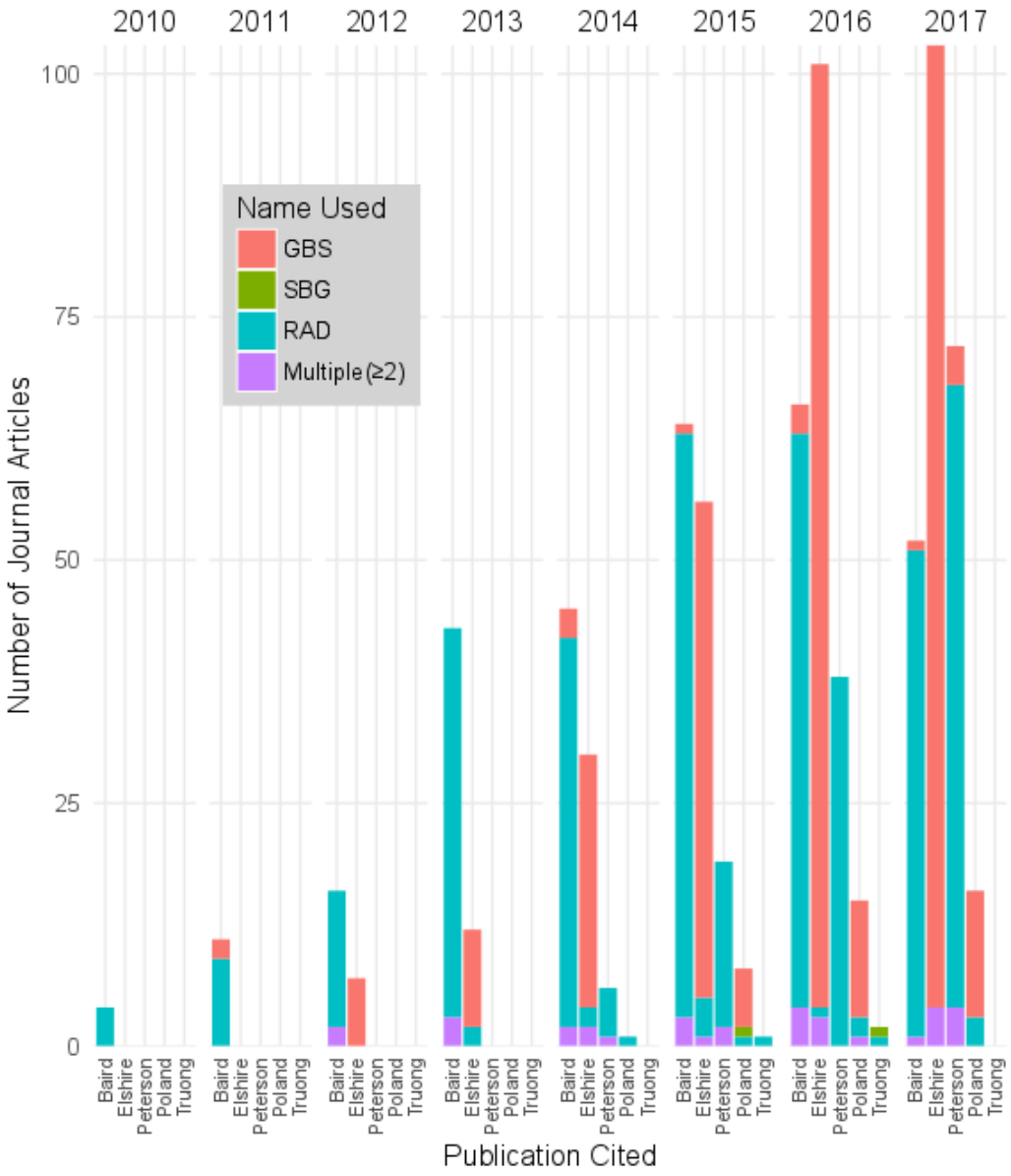

Figure 2. Trends in the use of the names "GBS" (from Elshire *et al.* (2011) and Poland *et al.* (2012)), "SBG" (from Truong *et al*. (2012)), and "RAD" (from Baird *et al.* (2008) and Peterson *et al.* (2012)) in the title, abstract, or keywords of journal articles that cite either Baird *et al.* (2008), Elshire *et al.* (2011), Poland *et al.* (2012), Truong *et al.* (2012) or Peterson *et al.* (2012). Bars indicate the number of journal articles citing each publication, while colours indicate the number referring to each name. About 8.4% of papers use an ambiguous or incorrect name in reference to the cited method (e.g. ~4% of papers uniquely citing Baird *et al*. (2008) in 2017 refer specifically to GBS or SBG alone or in combination with RAD, despite neither name being used in that paper).