Adapterama III: Quadruple-indexed, double/triple-enzyme RADseq libraries (2RAD/3RAD)

Natalia J. Bayona-Vásquez[1,2,3,*], Travis C. Glenn[1,3,4,5,*], Troy J. Kieran[1], Todd W. Pierson[1,8], Sandra L. Hoffberg[4,9], Peter A. Scott[6,10], Kerin E. Bentley[4,11], John W. Finger[1,5,12], Swarnali Louha[3], Nicholas Troendle[4,13], Pindaro Diaz-Jaimes[2], Rodney Mauricio[4] and Brant C. Faircloth[7]

[1] Department of Environmental Health Science, University of Georgia, Athens, GA, United States of America

[2] Unidad Académica de Ecología y Biodiversidad Acuática, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México, Mexico City, Mexico

[3] Institute of Bioinformatics, University of Georgia, Athens, GA, United States of America

[4] Department of Genetics, University of Georgia, Athens, GA, United States of America

[5] Interdisciplinary Toxicology Program, University of Georgia, Athens, GA, United States of America

[6] Department of Biological Sciences, University of Alabama, Tuscaloosa, AL, United States of America

[7] Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA, United States of America

[8] Current affiliation: Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN, United States of America

[9] Current affiliation: Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY, United States of America

[10] Current affiliation: Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA, United States of America

[11] Current affiliation: LeafWorks Inc., Sebastopol, CA, United States of America

[12] Current affiliation: Department of Biological Sciences, Auburn University, Auburn, AL, United States of America

[13] Current affiliation: Department of Natural, Health, and Mathematical Sciences, MidAmerica Nazarene University, Olathe, KS, United States of America

[*] These authors contributed equally to this work.

Corresponding authors
Natalia J. Bayona-Vásquez, njbayona@uga.edu
Travis C. Glenn, travisg@uga.edu

ABSTRACT

Molecular ecologists frequently use genome reduction strategies that rely upon restriction enzyme digestion of genomic DNA to sample consistent portions of the genome from many individuals (e.g., RADseq, GBS). However, researchers often find the existing methods expensive to initiate and/or difficult to implement consistently, especially because it is difficult to multiplex sufficient numbers of samples to fill entire sequencing lanes. Here, we introduce a low-cost and highly robust approach for the construction of dual-digest RADseq libraries that build on adapters and primers designed in *Adapterama I*. Major features of our method include: (1) minimizing the number of processing steps; (2) focusing on a single strand of sample DNA for library construction, allowing the use of a non-phosphorylated adapter on one end; (3) ligating adapters in the presence of active restriction enzymes, thereby reducing chimeras; (4) including an optional third restriction enzyme to cut apart adapter-dimers formed by the phosphorylated adapter, thus increasing the efficiency of adapter ligation to sample DNA, which is particularly effective when only low quantity/quality DNA samples are available; (5) interchangeable adapter designs; (6) incorporating variable-length internal indexes within the adapters to increase the scope of sample indexing, facilitate

pooling, and increase sequence diversity; (7) maintaining compatibility with universal dual-indexed primers and thus, Illumina sequencing reagents and libraries; and, (8) easy modification for the identification of PCR duplicates. We present eight adapter designs that work with 72 restriction enzyme combinations. We demonstrate the efficiency of our approach by comparing it with existing methods, and we validate its utility through the discovery of many variable loci in a variety of non-model organisms. Our 2RAD/3RAD method is easy to perform, has low startup costs, has increased utility with low-concentration input DNA, and produces libraries that can be highly-multiplexed and pooled with other Illumina libraries.

INTRODUCTION

Although next-generation DNA sequencing (NGS) facilitates data collection at low cost, it is not yet economically or computationally feasible for most ecological projects to sequence whole genomes from many individuals or from organisms with large genomes. However, many questions can be addressed with a small fraction of the genome (*DeWoody & DeWoody, 2005*; *Cariou, Duret & Charlat, 2013*; *Pante et al., 2015*; *Andrews et al., 2016*). Thus, researchers have created a variety of strategies to sample a consistent portion of the genome from large numbers of individuals at low cost (*Harvey et al., 2016*; *Heyduk et al., 2016*; *Glenn & Faircloth, 2016*). One of the most popular genome sampling strategies uses restriction enzymes to reduce genome complexity and sequence a set of orthologous loci across individuals (Restriction site Associated DNA sequencing, RADseq; *Miller et al., 2007*; *Baird et al., 2008*). RADseq approaches have several key advantages. First a reference genome is not required (although it is better to have one (e.g., *Shafer et al., 2016*)). Second library preparations are relatively low-cost. Third, RADseq techniques can be applied with minimal modification across a broad spectrum of organisms. Finally, mature software that is updated regularly is available for data analyses (e.g., Stacks; *Catchen et al., 2011*; *Catchen et al., 2013*; pyRAD; *Eaton, 2014*).

Many variants of the general RADseq approach have been developed (*Andrews et al., 2016*; *Salas-Linaza & Oono, 2018*), including but not limited to: the original method (RAD; *Miller et al., 2007*; *Baird et al., 2008*), genotype-by-sequencing (GBS; *Elshire et al., 2011*), 2-enzyme GBS (*Poland & Rife, 2012*), dual-digest RADseq (ddRAD; *Peterson et al., 2012*), 2bRAD (*Wang et al., 2012*), ezRAD (*Toonen et al., 2013*), and quaddRAD (*Franchini et al., 2017*). Although some of these RADseq approaches are in widespread use, they also have well-documented limitations (*Davey et al., 2011*; *Andrews et al., 2014*) which have been thoroughly reviewed in other publications (*Andrews et al., 2016*; *Harvey et al., 2016*; *Heyduk et al., 2016*).

Limitations of methods that pool DNA samples from multiple individuals within putative populations are particularly acute (*Andrews et al., 2016*); thus, we focus on
methods where individual samples are indexed such that individual organisms can be genotyped. Limitations of current individual-based RADseq methods include: (1) high up-front costs for adapters (∼$4,550 USD, for 12 pools of 48 samples; *Peterson et al., 2012*; *Salas-Linaza & Oono, 2018*), which limits experimental flexibility; (2) that adapters are phosphorylated in both ends, and thus, can form adapter dimers; (3) the use of an optional step to incorporate a biotin-containing primer and streptavidin beads to separate correct constructs; (4) the inability to reduce chimera formation; (5) the fact that no sequence diversity is present across the restriction recognition sequence of the resulting libraries, which constrains sequencing options; (6) the need for moderate to high amounts of high-molecular-weight DNA; (7) limited ability to multiplex high numbers of libraries, and thus, high sequencing costs; and (8) workflows of varying complexity. Thus, developing RADseq methods that reduce the cost of adapters and primers, vastly improve sample multiplexing, better accommodate low-concentration DNA samples (low DNA input), improve flexibility, increase consistency, and simplify the workflow would be helpful to many researchers.

Here, we present an alternative approach (2RAD/3RAD) for preparing RADseq libraries, which is similar in spirit to and builds upon the strengths of ddRAD, while also addressing each of the limitations summarized above and described below in our Rationale for methodological approach. In brief, 2RAD/3RAD uses DNA that is digested with multiple restriction enzymes and ligates adapters in the presence of these functional restriction enzymes, followed by PCR with primers developed in *Adapterama I* (*Glenn et al., 2019*) to make fully active quadruple-indexed Illumina libraries that can be highly-multiplexed (Figs. 1 and S1). Although some of our adapter designs and working procedures have been implemented and published during the development of our method (e.g., *Graham et al., 2015*; *Hoffberg et al., 2016*; *Scott, Glenn & Rissler, 2017*), additional designs, design details, flexibility, advantages and disadvantages of the system have yet to be described. Below, we explain the design goals and rationale for our approach, detail how we have implemented the method, demonstrate that large numbers of polymorphic loci can be discovered from a broad array of organisms using just one of the possible variations of our method, and discuss these results and additional work to extend this approach.

Our overall goal was to develop a ddRAD-style method with the following characteristics: (1) a simplified workflow with few consecutive buffer exchanges; (2) sequential or simultaneous digestion of DNA and ligation of adapters; (3) reduced chimera formation; (4) increased library efficiency (i.e., increasing the ratio of sample molecules converted into complete library molecules) through suppressed adapter dimer formation; (5) hierarchical combinatorial indexing to facilitate efficient multiplexing of many samples; (6) reduced costs, both for initial buy-in (i.e., cost of all reagents to start using the method) and per sample prepared; and (7) facilitation of pooling with any other Illumina library type. We built upon the adapter design and methods of *Glenn & Schable (2005)* to achieve the first four goals, whereas we extended the work of *Faircloth & Glenn (2012)* and *Glenn et al. (2019)* to achieve the last three goals (Files S1–S2).
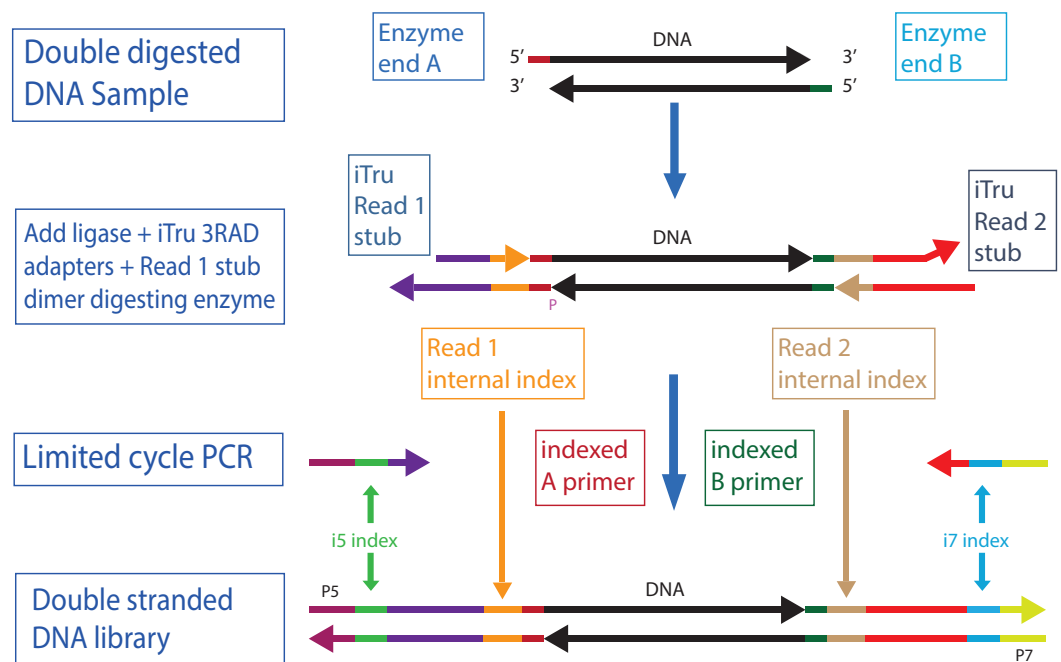
Figure 1 **Overview of 2RAD/3RAD library construction.** Genomic DNA is digested with two restriction enzymes (A and B). Adapters are ligated to the digested DNA, but only the bottom strand has functional adapters. The top strand has shorter, non-functional versions of the adapters. The ligation products are then used in a limited cycle PCR with iTru5 and iTru7 primers to form fully active double-stranded DNA molecules. The color-scheme follows those of *Glenn et al. (2019)* and *Hoffberg et al. (2016)*.

Rationale for methodological approach

To achieve our first three design goals, we use reagents that allow simultaneous digestion of the sample DNA and ligation of the adapters onto the sample DNA (*Glenn & Schable, 2005*). Simultaneous digestion of DNA with multiple restriction enzymes requires that the enzymes are active in the same buffer and at the same temperature. New England Biolabs (NEB; Ipswich, MA, USA) has developed many enzymes that retain high activity in a single buffer (CutSmart) and describes the activity of their enzymes in their other standard buffer formulations (Table 1). However, this approach is transferable to enzymes and buffers from other companies as long as cut-sites and buffers are carefully chosen. Because T4 DNA ligase can be used in the same buffers as most restriction enzymes, if the buffers are supplemented with ATP, researchers can start by digesting DNA, then add ligase and ATP to the digestion reaction and change the temperature to promote ligation. By cycling between temperatures that promote ligation, then digestion, multiple times, reactions can be driven to highly efficient outcomes (i.e., high proportions of the input DNA will be cut and will have adapters ligated onto the ends; Figs. S1–S2). A major distinction between the methods of *Glenn & Schable (2005)* and those needed here is that a single blunt-ended 5′ phosphorylated adapter (Super SNX) was used in the prior work, resulting in identical adapters on each end of the resulting libraries, whereas Illumina libraries require unique adapter sequences on each end of the library molecules (Fig. 1). By focusing on a single

strand of the template, rather than both strands, it is possible to use only one adapter that is phosphorylated (i.e., Read 1 adapter, bottom strand; Fig. 1, Figs. S1–S2) in 2RAD/3RAD, whereas the other adapter can use plain oligonucleotides for both strands (File S3), forming fewer adapter chimeras.

To achieve our fourth design goal, we design custom adapters. During ligation, double-stranded adapters that are modified versions of the TruSeq Read 1 and Read 2 sequences (Table 2) are ligated onto each fragment of DNA. The iTru Read 2 adapters are unphosphorylated on the 5′ end and will not self-ligate to form dimers. The iTru Read 1 adapter is a perfect match to the sticky end of the insert DNA, but the adapter does not have the correct bases to recreate the restriction site used to cut the sample DNA. Because the iTru Read 1 adapter is phosphorylated on the 5′ end, it can and will form dimers.

For 2RAD, we select sets of low-cost, type II restriction enzymes that form unique cohesive-ends (i.e., incompatible sticky-ends). To further achieve goal four, in our 3RAD protocol, we use these two restriction enzymes (e.g., XbaI and EcoRI; Table 1) with a third restriction enzyme (e.g., NheI) that produces a cohesive-end compatible with one of the other restriction enzymes (e.g., XbaI; Fig. 2). We then assigned the two restriction enzymes with compatible cohesive-ends (e.g., XbaI, and NheI) to Illumina Read 1 adapter stub sequences (*Glenn et al., 2019*) and assigned the incompatible restriction enzyme (e.g., EcoRI) to Read 2 adapter stubs. Next, we designed the Read 1 stubs such that if they self-ligated to form Read 1 adapter-dimers, they create the recognition sequence for the third restriction enzyme (e.g., NheI; File S3; *Glenn & Schable, 2005*). Similarly, Read 1 adapters ligated to genomic DNA with third restriction enzyme cut-sites recreate the recognition sequence for the third restriction enzyme. See Fig. 2 for a graphical representation of one example of this design using the restriction enzymes XbaI, EcoRI, and NheI.

As described above, we cycle temperatures in the simultaneous digestion and ligation to allow this third restriction enzyme to cut apart adapter-dimers, which increases the consistency and efficiency of 3RAD library preparation, even with limited amounts of sample DNA. However, while genomic DNA cut by this third restriction enzyme and ligated to Read 1 adapters should be recut by the same restriction enzyme, any remaining molecules of this form are suitable for PCR amplification and may be present in final libraries. 2RAD functions without this third restriction enzyme, and in practice, the differences between 2RAD and ddRAD are: (a) simultaneous digestion and ligation reactions; (b) inclusion of variable-length internal indexes on each end (see below); (c) compatibility with iTru primers from *Adapterama I* that allow for highly-multiplexed libraries to be pooled for sequencing on Illumina platforms (File S3); and (d) potential substitution of the normal iTru5 primer containing specific index for iTru5-8N primer pool with 65,536 indexes, which facilitates identification and removal of PCR duplicates (*Hoffberg et al., 2016*).

Additionally, we construct 2RAD/3RAD adapters so that each double-stranded adapter has one active strand (i.e., the bottom strand as shown in all figures herein) and one unused strand (i.e., the top strand in all figures herein). The dummy strand is simply used for structural support and correct 3D structure of the adapters and constructs through the ligation process. Both strands fit together on each side of the input DNA during ligation, but the nick between the sample DNA and Read 2 adapter top dummy strand is not ligated

Bayona-Vásquez et al. (2019), *PeerJ*, DOI 10.7717/peerj.7724

Table 1 Enzyme combinations and characteristics. Four design sets each for Read1 (R1) and Read2 (R2) are given. For 2RAD, any two-enzyme combination of Read 1 and Read 2 in black can be used. For 3RAD, the third enzyme (in blue) blocks adapter-dimer formation of the Read 1 adapter (File S3). Digestion efficiency is given for three NEB buffers (2.1, 3.1, and CutSmart®), with the best conditions highlighted in green, and poor or important non-standard conditions in red. Sensitivity to methylation in the template sequence is given, as is the optimal temperature for digestion and the number of bases in the recognition sequence. Note: some restriction enzymes are available as high-fidelity (HF, i.e.: NheI-HF, SpeI-HF, and NsiI-HF), all these have 100% efficiency in CutSmart® Buffer.

		Read 1 adapter sets							Read 2 adapter sets					
		NEB buffer			CpG meth	Cut temp	Base cutter			NEB buffer			CpG meth	Cut temp	Base cutter
Set	Enzyme	2.1	3.1	CutSmart®				Set	Enzyme	2.1	3.1	CutSmart®			
R1.A	NheI	100	10	100	+/−	37	6	R2.1	EcoRI-HF	100	10	100	+/−	37	6
	XbaI	100	75	100	−	37	6		MfeI-HF	25	10	100	−	37	6
	SpeI	100	25	100	−	37	6		ApoI	75	100	75	−	50*	6
R1.B	ClaI	50	50	100	+	37	6	R2.2	BamHI-HF	50	10	100	−	37	6
	MspI	100	50	100	−	37	4		BclI	100	100	75	−	50*	6
	TaqαI	75	100	100	−	65	4		BstYI	100	75	100	−	60**	6
R1.C	PstI-HF	75	50	100	−	37	6								
	PstI	75	100	50	−	37	6	R2.3	DdeI	100	100	100	−	37	4
	NsiI	75	100	25	−	37	6								
R1.D	CviQI	100	100	75	−	25	6	R2.4	HindII-HF	100	10	100	−	37	6
	NdeI	100	100	100	−	37	6		HindIII	100	50	50	−	37	6
	MseI	100	75	100	−	37	4								
	AseI	50	100	10	−	37	6								
	BfaI	10	10	100	−	37	4								
Table 2 Example 2RAD/3RAD adapter stub sequences. Groups of four adapters form a balanced set, all eight complete sets are available in File S3. Non-complementary sequences are given in lower case. Tag sequences are in italics. Adapters must be hydrated and annealed prior to use (File S4).

Adapter	Oligo name	Sequence (5′ to 3′)
iTru_NheI_R1_A	iTru_NheI_R1_stub_A	ACGACGCTCTTCCGATCT*CCGAAT*G
	iTru_NheI_R1_RCp_A	/5phos/CTAGC*ATTCGG*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
iTru_EcoRI_R2_1	iTru_EcoRI_R2_RC_stub_1	AATTA*CGTTAG*AGATCGGAAGAGCACACGTaatcc
	iTru_EcoRI_R2_1	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*CTAACG*T
iTru_ClaI_R1_B	iTru_ClaI_R1_stub_B	ACGACGCTCTTCCGATCT*TTAGGCA*AT
	iTru_ClaI_R1_RCp_B	/5phos/CGAT*TGCCTAA*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
iTru_BamHI_R2_2	iTru_BamHI_R2_RC_stub_2	GATCG*GTACCGA*AGATCGGAAGAGCACACGTaatcc
	iTru_BamHI_R2_2	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*TCGGTAC*C
iTru_PstI_R1_C	iTru_PstI_R1_stub_C	ACGACGCTCTTCCGATCT*AACTCGTC*CTGCA
	iTru_PstI_R1_RCp_C	/5phos/G*GACGAGTT*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
iTru_DdeI_R2_3	iTru_DdeI_R2_RC_stub_3	TNACC*AACGATC*AGATCGGAAGAGCACACGTaatcc
	iTru_DdeI_R2_3	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*GATCGTTG*G
iTru_CviQI_R1_D	iTru_CviQI_R1_stub_D	ACGACGCTCTTCCGATCT*GGTCTACGT*G
	iTru_CviQI_R1_RCp_D	/5phos/TAC*ACGTAGACC*AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT
iTru_HindIII_R2_4	iTru_HindIII_R2_RC_stub_4	AGCTA*AGTGTAGCT*AGATCGGAAGAGCACACGTaatcc
	iTru_HindIII_R2_4	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT*AGCTACACT*T

(Fig. S1). Thus, the unused strand construct breaks apart during PCR steps, and only those constructs with bottom strands that successfully ligate both kinds of adapters are amplified. This ensures that valid constructs with the correct restriction sites at opposite ends dominate the amplified library pools. Additionally, the oligonucleotides for the top strand (as depicted in Figs. 1, S1, and S2) are not full-length, so they cannot be used as templates for the iTru5 or iTru7 primers. Finally, the top strand of the Read 2 adapter ends in five non-complementary bases so that it cannot serve as an unwanted primer during library amplification.

We achieve our next three design goals (i.e., 5–7) by including variable-length internal indexes—also known as "in-line barcodes" (Andrews et al., 2016)—within the Read 1 and Read 2 adapter stubs and making the adapter stubs compatible with the primers of Glenn et al. (2019; Fig. 1 and S1). For each adapter stub design, we have made eight versions of the Read 1 adapter stub and 12 versions of the Read 2 adapter stub (File S3). Each adapter stub version includes an internal index of 5, 6, 7, or 8 nucleotides (nt). The purpose of these internal indexes is twofold: (1) combinations of the Read 1 and Read 2 adapters create 96 (8 ×12) index combinations, which facilitates pooling of samples from 96-well plates (File S3); and (2) the variable length of indexes increases base diversity within pools of libraries (Krueger, Andrews & Osborne, 2011), which is important when sequencing libraries derived from restriction enzymes digestion (Mitra et al., 2015; Glenn et al., 2019).

To create full-length libraries, after digestion-ligation cycles, we amplify through reduced cycle PCR using the iTru5 and iTru7 primers of Glenn et al. (2019; Figs. 1 and 3, S1 and S2). Because the 2RAD/3RAD adapters already include internal tags that can identify all samples in a 96-well plate, samples can be pooled prior to PCR and externally tagged with the iTru5 and iTru7 primers (to identify multiple plates of samples), or users can
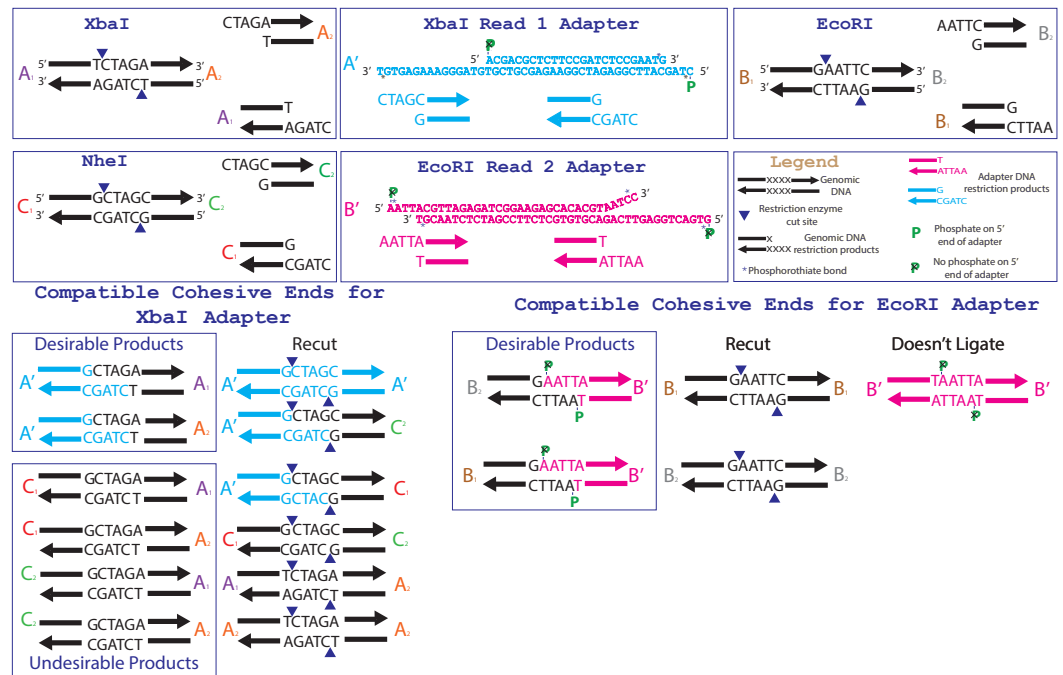
Figure 2 **Specific adapter sequences and products created during the ligation of 3RAD libraries.** The full adapter sequences for the 3RAD enzyme combination NheI, XbaI and EcoRI-HF (Table 1) are given in the top center boxes. The relevant recognition sequences for the three restriction endonucleases are given in the top outer boxes. The products that are formed from ligation of the triple-enzyme digests and adapters are shown at the bottom.

PCR amplify individual wells with unique external tags and pool PCR products, creating redundant indexing (Fig. 3). The resulting libraries are then size-selected and sequenced using the four standard Illumina TruSeq primers, each of which returns a different indexing read (Fig. 4). Because 2RAD/3RAD libraries are constructed using the iTru primers from *Adapterama I* (*Glenn et al., 2019*), they are compatible with (i.e., can be pooled with) iTru and Illumina TruSeq libraries prior to sequencing on Illumina sequencing instruments, achieving our final design goal. This potential for highly-multiplexed pools is especially important with the advent of platforms such as the Illumina NovaSeq, which is capable of producing up to 3,000 Gb of data from one flow cell.

MATERIALS & METHODS

Design of adapters

Oligonucleotide sequences for all versions of all adapter combinations were designed using File S3. We ordered the resulting oligonucleotide sequences in plates from Integrated DNA Technologies (IDT; Coralville, IA, USA) and prepared them following directions from File S4.
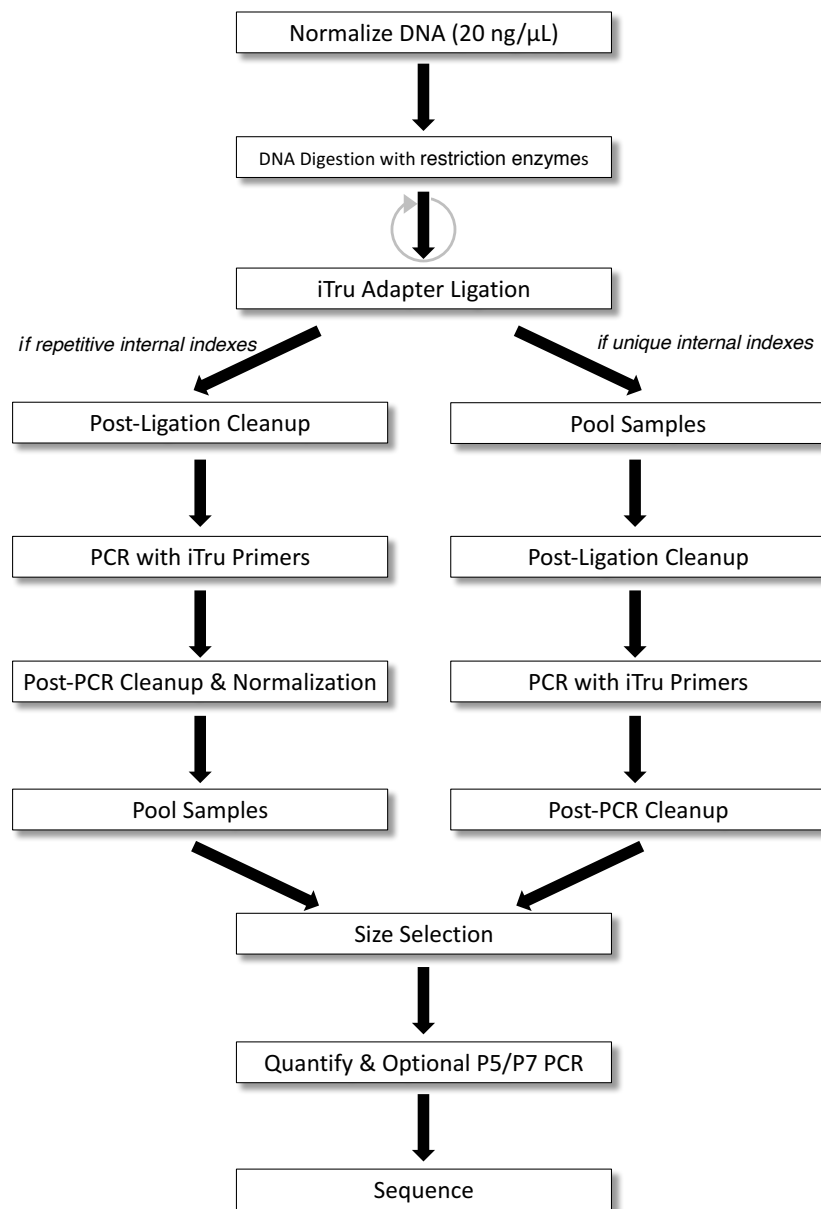
Figure 3 **2RAD/3RAD workflow for samples with unique or repeated adapter indexes.** DNA is normalized, digested with restriction enzymes, and ligated to adapters. If indexes within adapters uniquely identify all samples (right), samples can be pooled before clean-up and PCR. If indexes do not uniquely identify individuals, PCR must be done separately on each sample, and samples must be normalized and cleaned before pooling. Then, samples are size-selected and quantified to determine if a final P5/P7 PCR should be performed before sequencing.

Library preparation

Step-by-step protocols can be found in our extended material (File S1, Fig. 1). Several variants of the protocol are possible; thus we present specific examples that demonstrate this range (File S1). Briefly, we digested sample DNA with one of the combinations of restriction
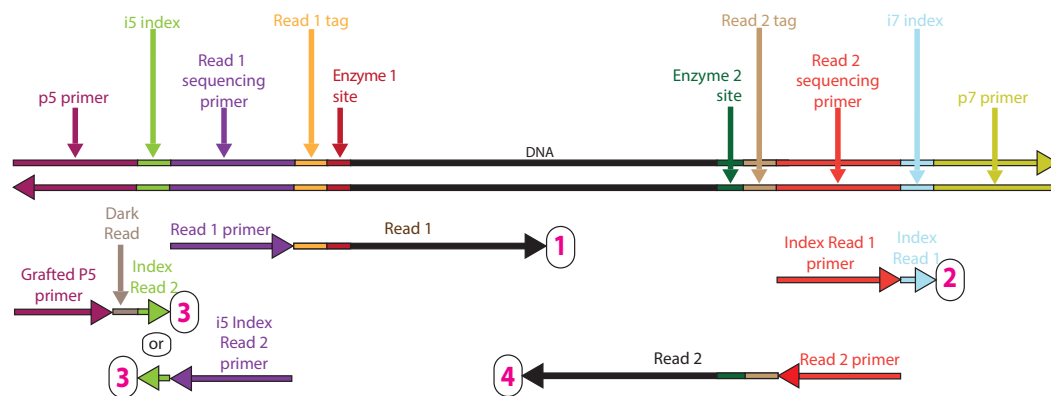
Figure 4 **Sequencing reads that can be obtained from full length 2RAD/3RAD library molecules.** The top double stranded molecule shows a 2RAD/3RAD library molecule prepared as described in the text (File S1). The horizontal arrows beneath the library molecule indicate Illumina sequencing primers (binding to the complementary strand of the library molecules). The tip of the arrowhead indicates the 3′ end of the primer and the direction of elongation for sequencing. Four sequencing reads are shown for each library prepared molecule, with one read for each index and each strand of the genomic DNA, including internal indexes. Reads are arranged 1 to 4 (numbered in magenta) from top to bottom, respectively. The arrow immediately 3′ of the primers, indicates the data that are obtained from that primer, with coloring that is consistent with 2RAD/3RAD library molecule.

enzymes (Table 1), ligated a Read1 and Read2 adapter pair (File S3) and simultaneously digested dimers and chimeras with two alternating cycles of ligation-digestion. Next, we pooled libraries that had unique internal indexes (i.e., samples within 96-well plates), and purified ligation products (Fig. 3); note: this step is optional. To generate full-length libraries, we performed a PCR using iTru primers from *Adapterama I* that contained unique indexes (i.e., external indexes) to further differentiate individual samples (i.e., to indicate which 96-well plate the samples were from). The resulting libraries were purified and quantified. If samples were not pooled after the ligation step, iTru primers with unique external indexes were used, and then the resulting libraries were pooled. Library pools were size-selected to capture fragments in a Pippin Prep (Sage Science, Beverly, MA, USA) with a 1.5% dye-free Marker K agarose gel cassette (CDF1510) set to capture fragments at 550 bp ±10%. Resulting libraries were quantified, and in some cases, we performed a limited-cycle PCR with P5 and P7 primers to increase library concentration before sequencing.

3RAD efficiency

We tested and compared our 2RAD and 3RAD protocols with the traditional ddRAD protocol using the same restriction enzymes, adapters, and primers. To simplify the comparison between protocols, we used the pUC19 vector, which contains XbaI cut-site at position 423 and EcoRI cut-site at position 396, as template DNA. First, we amplified an approximately 500 bp fragment within the vector using the primers pUC19-215F-AAGGAGAAAATACCGCATCAGG and pUC19-774R-TAACCGTATTACCGCCTTTGAG. Then, we made four 10-fold dilution series. We used these five products (one stock and four dilutions) plus a negative control as input

for 2RAD, 3RAD, and ddRAD (i.e., 2RAD with sequential digestion and ligation) libraries (File S5).

3RAD applied case studies for validation of the method

We tested our 3RAD protocol on eight example projects focused on diverse taxa: Kinosternidae (turtles), Ixodidae (ticks), *Eurycea bislineata* species complex (salamanders), *Wisteria floribunda* ×*Wisteria sinensis* hybrid population (plants), *Rhodnius pallescens* (insects), *Gambusia affinis* (freshwater fish), *Sphyrna tiburo* (shark), and *Sphyrna lewini* (shark). The University of Tennessee Knoxville IACUC provided full approval for the study on salamanders (2352-0318) and the field permit was provided by Tennessee Wildlife Resources Agency (Permit #3840). The University of Alabama IACUC provided full approval for the study on turtles (11-357) and field permits were provided by the Alabama Conservation (License #2011-2013000110468680) and the US Fish and Wildlife Service (Permit #TE088913-0). Field sample collection for the *Sphyrna tiburo* project was approved by the US Fish and Wildlife Service and Florida Fish & Wildlife Conservation Commission during FWC-FWRI Fisheries-Independent Monitoring Program.

Each dataset consisted of 12 to 24 samples. These projects span a broad diversity of organisms (e.g., in taxonomic classification, population size, heterozygosity level, and genome size), motivating evolutionary questions, and associated methods (i.e., from population genetics to phylogenetics; Table 3). After preliminary examination of several restriction enzyme combinations, we used XbaI, EcoRI-HF, and NheI and adapter sets R1.A and R2.1 (Design 1 in File S3) for all projects. We prepared all libraries using similar methods to those mentioned above, but some details varied among projects (Files S1, S6 and S7). We used a modification of our 3RAD method to incorporate molecular ID tags for the *Wisteria* project (see *Hoffberg et al., 2016*; Files S6 and S7). We included these datasets to validate the utility of our method and present limited results in File S6; more detailed population genetic or phylogenetic results for each project are beyond the scope of this article.

Sequencing and data analyses

We sequenced libraries on multiple Illumina platforms in multiple core labs. We used the Illumina NextSeq 500 platform to generate PE75 data for the *Rhodnius*, *Gambusia*, Kinosternidae, and *Wisteria* projects and Illumina HiSeq 2500 or NextSeq 500 platforms to generate PE150 data for the Ixodidae, *Sphyrna,* and *Eurycea* projects. We planned all sequencing runs to produce approximately one million reads per sample, which facilitates comparison of the 3RAD results among species with varying genome sizes, with the exception of the Ixodidae project, where four million reads were targeted per sample.

We assembled data from each project independently using Stacks v1.42 (*Catchen et al., 2013*; *Catchen et al., 2011*; see File S6). For the *Wisteria* project, we used molecular ID tags to facilitate PCR duplicate removal with the module *clone_filter* in Stacks (*Catchen et al., 2013*; *Hoffberg et al., 2016*). We describe detailed parameters and software specifications for each project in File S6. Briefly, for most projects, we used the *process_radtags* program to demultiplex and/or clean and trim the sequence data. We parallel-merged the mates

Bayona-Vásquez et al. (2019), *PeerJ*, DOI 10.7717/peerj.7724

*Peer*J

Table 3 3RAD example projects. Classification and genome size of taxa, number of samples tested for each, Illumina read length (nt), number of loci obtained after the assembly method, number loci and SNPs obtained after filtering by only polymorphic loci shared in at least 75% of samples, and the average coverage among loci and individuals. The number of loci can be quite large and certainty of homology variable with distantly related samples, particularly if they have large genomes.

	Class	Genome size (c-value)[a]	Groups	Indiv.	PE Read Length (nt)	Loci	Final Loci	SNPs	Mean Coverage (x)
Kinosternidae (*De Smet, 1981*; *Olmo, 1976*)	Reptilia	2.9	7	24	75	233,072	4,034[b]	27,881	12
Ixodidae (*Geraci et al., 2007*; *Palmer et al., 1994*; *Ullmann et al., 2005*)	Arachnida	2.7	4	16	150	332,057	4,484	13,136	36
Eurycea (*Olmo, 1974*; *Licht & Lowcock, 1991*; *Bachmann, 1970*)	Amphibia	25.4	1	21	150	425,729	30	360	7
Wisteria	Magnoliopsida	?	1	24	75	30,029	1,669	5,820	44
Sternotherus depressus (*Olmo, 1976*)	Reptilia	2.7	3	12	75	103,240	16,695	25,578	11
Amblyomma americanum (*Geraci et al., 2007*; *Palmer et al., 1994*)	Arachnida	2.4	2	7	150	128,899	19,843	69,518	36
Rhodnius pallescens (*Gomez-Palacio et al., 2012*)	Insecta	0.7	5	16	75	92,687	7,779	12,099	23
Gambusia affinis (*Cui, Ren & Yu, 1991*; *Lamatsch et al., 2000*; *Ojima & Yamamoto, 1990*; *Tiersch et al., 1989*)	Actinopterygii	0.9	5	24	75	18,629	2,140	5,429	54
Sphyrna tiburo (*Hinegardner, 1976*)	Chondrichthyes	3.9	6	24	150	42,705	7,183	17,555	18
Sphyrna lewini (*Hinegardner, 1976*; *Hardie & Hebert, 2003*; *Hardie & Hebert, 2004*; *Schwartz & Maddock, 1986*; *Stingo et al., 1989*)	Chondrichthyes	3.6	7	15	150	44,125	5,263	12,272	27

Notes.
[a]Genome sizes are approximations from *Gregory* (*2018*, November 20). Animal Genome Size Database. Retrieved from http://www.genomesize.com. We could find no published genome size for *Wisteria* in the literature, so we omitted it. For all other examples, we averaged reported genome sizes for that species or its closest available relatives; for examples including multiple species (e.g., Kinosternidae), we weighted this average dependent upon the taxonomic composition of the sample.
[b]From pyRAD assembly of homologous loci across all Kinosternidae.

of paired-end reads. We used the *denovo_map* program to assemble reads *de novo* and to calculate coverage, number of loci, and number of SNPs recovered for each project; we compared these data to genome size and sequencing read length (PE75 or PE150). Finally, we used the *populations* program to export loci shared in at least in 60–75% of localities and individuals to VCF files. Because there exists a reference genome for *Gambusia affinis* (*Hoffberg et al., 2018*; NCBI: NHOQ01000000; details in File S6), we also assembled data from this project against the reference. For population-level datasets, we calculated *F*-statistics and performed preliminary population clustering analyses in Structure v2.3.4 (*Pritchard, Stephens & Donelly, 2000*; File S6). For the Kinosternidae project, we conducted a *de novo* locus assembly using pyRAD v1.0.4 (*Eaton, 2014*; details in File S6).

Finally, we estimated the prevalence and impact of loci with third restriction enzyme cut-sites in our data. We estimated the proportion of these third restriction enzyme cut-sites relative to the first restriction enzyme cut-site (i.e., intended cut-site) for five of the projects, and we evaluated variation among adapters and projects using ANOVA in R v3.5.1 (*R Core Team, 2018*). To evaluate the effect of these loci in downstream analyses, we reanalyzed data from three of our projects (i.e., both *Sphyrna* and *Amblyomma americanum*) after removing third restriction enzyme loci from the datasets. To do this, we reassembled data in Stacks v1.44 (*Catchen et al., 2011*; *Catchen et al., 2013*) using *process_radtags* two independent times: first, "rescuing barcodes", cleaning, and trimming the raw sequence data as before, but disabling rad check (–disable_rad_check) to leave the cut-sites intact; and second, using the previous step's output as input, checking only for exact, intended restriction enzyme cut-sites (i.e., XbaI and EcoRI). From this output, we assembled and analyzed data similar to above, as detailed in File S6.

RESULTS

We developed four sets of adapters, each with eight versions of Read 1 adapters and 12 versions of Read 2 adapters (File S3). We modified the iTru_R2_5 index sequence for BamHI because this index creates a BamHI recognition site in the adapter; otherwise, all adapters use a universal set of index sequences. The cost of synthesizing oligonucleotides for the adapters varies with synthesis scale, but it starts as low as ~$350 (US) per design set, with a recommended scale (100 nmol) costing ~$500 (US) per design set when synthesized into 96-well plates, which are sufficient for up to ~4,800 sample libraries.

3RAD libraries can be constructed routinely with approximately 12 h of hands-on time over the course of 2–3 days, with some variation depending mostly upon the step at which samples are pooled. The initial cost of restriction digestion and ligation is ~$0.85 per sample (File S3, "Library_prep_costs" Sheet). If PCRs are conducted individually, this adds >$1 per sample, but if the ligations are pooled before PCR, then this cost averages to $0.06 per sample. Size-selection using the Pippin Prep adds $0.12 per sample, assuming user access to the equipment. A total of $0.25 per sample is required for tips, plates and tubes. Thus, when samples are pooled prior to PCR, the total cost for library preparation is about $1.35 per sample.

In the efficiency test, we demonstrate that 3RAD libraries have fewer adapter-dimers and a higher concentration of library constructs than 2RAD or ddRAD libraries, which is
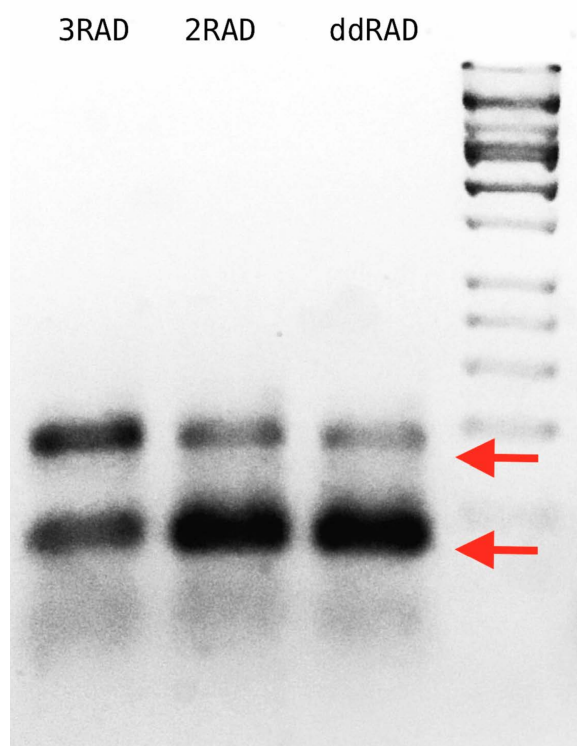
Figure 5 **Agarose gel with 3RAD, 2RAD, and ddRAD library products performed on pUC19 vector with an input quantity of 0.5 ng.** The band close to the 200 bp size standard (arrow above) is that corresponding to a proper library construct. The band below the 100 pb size standard (arrow below) corresponds to adapter-dimers (File S5). The gel indicates that 3RAD libraries outperformed the other two types of libraries tested by decreasing the adapter-dimers and therefore increasing the quantity of desired library constructs.

particularly important when libraries are constructed with low input DNA concentrations (Fig. 5; File S5).

In the Ixodidae example project, we obtained between 1.8–7.3 (mean = 4.2) million reads per sample, and for all other projects, we obtained between 0.6–3.6 (mean = 1.3) million reads per sample. With the exception of one sample from the Ixodidae project, we always recovered a high percentage of retained reads (78.9–99.7%) after cleaning and filtering steps (Table S1). Average coverage per locus varied from 6× for *Eurycea* to 70× for *Gambusia* (Table 3; Table S1; Fig. 6).

Our initial assemblies contained between 18,629 loci (*Gambusia*) and 425,729 loci (*Eurycea*). After filtering to retain only polymorphic loci found in at least 75% of individuals within each population, we recovered between 30 loci (*Eurycea*) and 19,843 loci (Ixodidae) containing between 360 and 69,518 SNPs, respectively (Table 3). As expected for RADseq protocols, the number of loci we obtained in the initial steps was proportional to the genome size, with more loci recovered in organisms with larger genomes. Due to the manner in which we filtered these loci, the final number of loci recovered is dependent
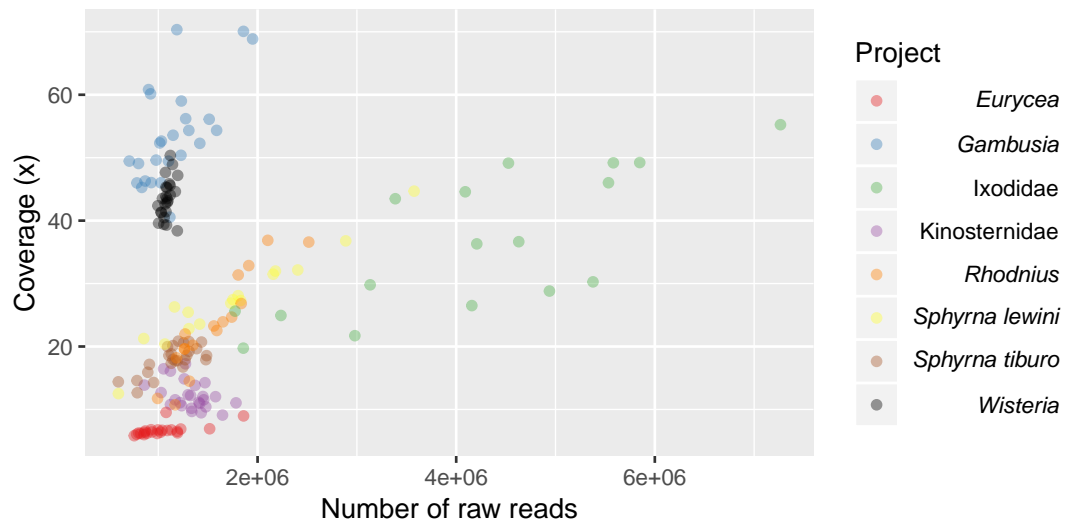
Figure 6 **Scatterplot of the average coverage of all loci (polymorphic and fixed) for each sample relative to sequencing depth of each sample.** *Eurycea* have the largest genome size and therefore the lowest average coverage per locus with approximately 1,000,000 reads. Average coverage increases as the genome size decreases (Fig. S3).

upon the intrinsic genetic variability of the organism, the scale of sampling, and sequencing coverage (Fig. S3). Detailed results for each project can be found in File S6.

Third restriction enzyme loci were present in all datasets, comprising an average of 20.5% (sd = 14.1%) of all reads. The percentage of reads from third restriction enzyme loci varied significantly both among datasets ($p < 0.01$) and among adapters with different indexes ($p < 0.01$; Fig. S4). Removing third restriction enzyme loci from our raw reads increased mean coverage of remaining loci (Tables S2, S3 and S4) and reduced the size of our final datasets from 7,183 to 6,738 loci in *Sphyrna tiburo*, from 5,263 to 4,807 loci in *Sphyrna lewini*, and from 69,518 loci to 39,605 in *Amblyomma americanum*. Estimates of F_{ST} were qualitatively similar in analyses including and excluding third restriction enzyme loci (Table S5).

DISCUSSION

We present an efficient, flexible, and low-cost system for preparing dual-digest RADseq libraries. Our method uses the iTru primers from *Adapterama I* (*Glenn et al., 2019*) and modifies the adapter stub for RADseq libraries by adding the appropriate overhang bases, as well as variable-length internal indexes which facilitates a single Illumina lane to be shared by many quadruple-indexed libraries. To illustrate the utility of our method, we present summary statistics from analyses of eight small example projects, representing diversity in taxonomy and scientific objectives. For each project, we obtained thousands of loci containing SNPs for downstream population genetic and phylogenetic analyses. Among the example projects, we highlighted the role of genome size, genetic variation, and sequencing coverage in determining the quality and quantity of data recovered. For example, when
processed differently, we recovered large numbers of homologous loci both among species in the family Kinosternidae and within a single representative species, *Sternotherus depressus*, from the same libraries and sequencing reads. These data are informative both for studying variation among populations of *S. depressus* and across relatively deep evolutionary time (∼55 Myr; File S6; *Scott, Glenn & Rissler, 2017*). We note that when datasets span deeper evolutionary time (e.g., *Eurycea* and Kinosternidae), we recover more loci in initial steps, but fewer of these loci are shared among individuals. However, many studies support the utility of these large and sparse data matrices for phylogenetic studies (e.g., *Streicher, Schulte & Wiens, 2015*; *Hosner et al., 2016*). For detailed discussion of the analyses for each dataset, see File S6.

The improved performance of 3RAD libraries when low concentration DNA is used as input is a consequence of maximizing the efficiency of the ligation of the adapters to library fragments by the use of the third-enzyme. However, it is important to highlight that: (1) our experiment qualitatively (not quantitatively) evaluated the library performance and (2) our experiment was simplified by using vector DNA that has the restriction sites. We believe the better performance on high complex and diverse samples will be maintained because our mechanism suppresses adapter-dimer formation, but this scenario was not tested here. Similarly, *Graham et al. (2015)* concluded that our 3RAD protocol performs well even with moderately degraded DNA samples.

By using the same set of restriction enzymes on DNA from a diverse set of organisms, we have demonstrated that our 3RAD method recovers suitable numbers of loci and SNPs from organisms with varying genome characteristics. As expected, when we use the same set of restriction enzymes (and thus similar expected frequency of cut-sites) and the same size-selection criteria for organisms with varying genome size, the average sequencing coverage per locus decreases as genome size increases (Fig. S3). For example, our dataset generated here from *Eurycea* (*C*-value = ∼25; Table 3) produced few loci meeting our coverage criteria, but higher sequencing coverage can remedy this. Alternatively, the number of loci in libraries can be tuned by changing restriction enzymes or size-selection criteria (*Peterson et al., 2012*). Additionally, using broad size-distributions in the size-selection step retains more loci and allows for greater size tolerance among alleles, whereas narrow distributions yield fewer loci. We suspect that use of narrow size-distributions excludes alleles at loci with significant size variation and may lead to increased levels of incorrect genotype calls due to the missing alleles outside of the selected size-range, but because most researchers are targeting loci without size variation, this bias should be small. A more significant issue related to narrow size-distributions is that a significant proportion of loci will be near the size cut-off and coverage decreases for these loci because some molecules of the targeted size will not be among the fragments retained (e.g., simply due to variance in migration through a gel).

A key advance of the 3RAD workflow (Files S1–S2) is the combination of enzymes and adapters used during digestion, ligation, and PCR steps to create the desired construct while minimizing the presence of dimers, chimeras, and improperly formed library molecules (those lacking restriction sites at both ends). Sequential and simultaneous digestion and ligation without buffer exchange increases the efficiency of the lab workflow and decreases

the amount of input DNA required. The 3RAD adapters we designed function with multiple restriction enzymes that leave compatible, cohesive ends (Table 2). Thus, there are at least 72 different restriction enzyme combinations possible with the current adapter sets. Although this flexibility is desirable, having adapters compatible with multiple restriction enzymes means that it was necessary to name the adapters based on the third RE, which can be confusing because the third restriction enzyme is *not* the desired restriction site in the sample libraries (and their resulting reads). The design spreadsheet (File S3, Sheets: Design_1-4) can be easily modified to accommodate other restriction enzymes to create additional designs. These sheets also incorporate cost calculations so that researchers may easily change the costs to reflect updated pricing from their supplier(s).

It should be noted that certain restriction enzyme combinations work well with some species, but not with others, primarily due to the presence of restriction sites within repetitive elements. Thus, our standard strategy is to empirically determine what restriction enzyme combination is best for any particular organism in a few representative samples. We start by looking at the distribution of post-PCR library DNA run through an agarose gel and exclude restriction enzyme combinations that don't produce even smears or that have dense bands in the desired size range. Sometimes, we then size-select and sequence libraries from one or two restriction enzyme combinations from this small batch of samples to determine which combination produces the most variable loci.

Our results show that XbaI, EcoRI-HF and NheI provide a suitable combination of restriction enzymes for a wide range of organisms, including plants, vertebrates, and invertebrates. We have used all of the adapter designs and many other restriction enzyme combinations from the enzyme list (Table 1) to survey SNPs in a variety of organisms. While not all combinations work well in all organisms, most of the organisms we have studied to date work well with restriction enzymes from Design 1 or Design 2 adapters. Although viable, we only rarely use Designs 3 and 4 (R1.C, R1.D, R2.3, and R2.4).

In our standard protocols (Files S1 and S5), we digest DNA with two different restriction enzymes (2RAD) or three different restriction enzymes (3RAD) to create sticky ends for adapter ligation (similar to ddRAD and 2-enzyme GBS). In 3RAD, the third restriction enzyme digests a recognition site formed by self-ligation of the phosphorylated adapters (Fig. 2). Although the third restriction enzyme facilitates creation of libraries with very little input DNA (≤ 0.1 ng; File S5), it does come with a cost. The third restriction enzyme also cuts genomic DNA that can be ligated to the R1 adapter, but the adapter:DNA ligation product is susceptible to re-cleavage by the restriction enzyme. To encourage this, our digestion/ligation cycling ends with a digestion step to cleave as many of these products as possible. Still, an average of 20% of loci in our assemblies had these third restriction enzyme cut-sites, and the nonrandom relationship between R1 adapter index and the prevalence of these loci suggests that the index has an effect upon the efficiency of the third restriction enzyme in cleaving the re-created recognition site. These loci are, in principle, suitable for downstream analyses, but because the protocol is designed to minimize their retention, they should have lower coverage than those with the intended restriction enzyme cut-site. A high prevalence of these off-target loci can require additional sequencing (and thus, increase costs), but these reads can easily be filtered and removed for all downstream

analyses if desired. The third restriction enzyme is not required for this procedure and further investigation into to trade-offs of including this restriction enzyme (i.e., 2RAD vs. 3RAD) and its optimal concentration are warranted. Alternatively, it is possible to engineer adapters that can ligate to genomic DNA, but not self-ligate (e.g., using 3′ dideoxycytidine), which we have done. Unfortunately, the adapters are significantly more expensive and were not stable when stored for ≥ ~6 months, both of which make the method impractical. Further research into other modifications or storage solutions for these adapters is warranted.

As detailed in *Hoffberg et al. (2016)*, our 3RAD approach is easy to modify to include tags that allow the removal of PCR duplicates (File S7) through a process comparable to quaddRAD (*Franchini et al., 2017*) and to *Schweyen, Rozenberg & Leese (2014)*. Our molecular ID tag protocol, which uses an iTru5-8N primer, can be used to make libraries for different types of downstream processes and/or modified for other types of libraries. For example, *Hoffberg et al. (2016)* used the iTru5-8N primer and sequence capture to focus sequencing on informative loci, reduce missing data, and remove PCR duplicates.

Due to the release of cutting-edge technologies with higher throughput, such as the Illumina NovaSeq 6000, which can generate > 2 billion paired-end reads in a single run, the ability to multiplex samples becomes paramount to reduce sequencing costs per sample. Our system, introduced in *Adapterama I* (*Glenn et al., 2019*), and from which 2RAD/3RAD extends, includes two sets of 384 indexed PCR primers, allowing for the pooling of up to 147,456 dual-indexed samples. Furthermore, the adapters designed in *Adapterama III* can create quadruple-indexed libraries, enabling the potential multiplexing of up to ~14 million libraries.

Our 2RAD/3RAD methods are similar to other dual-digest RADseq methods (e.g., ddRAD and 2-enzyme GBS), and most of the advantages of our general approach have been described previously (*Andrews et al., 2016*; *Harvey et al., 2016*; *Heyduk et al., 2016*). Our 2RAD/3RAD method achieves the seven methodological design objectives. We have demonstrated that the third restriction enzyme increases ligation efficiency by reducing adapter-dimers; thus, much less input DNA is necessary. In addition, multiple restriction enzymes are compatible with many of the adapters (Table 1), all indexes used herein conform a minimum edit distance of 3 (*Faircloth & Glenn, 2012*), we use limited PCR-cycles of pooled ligations to reduce PCR bias, and our method facilitates the easy incorporation of molecular ID tags to detect PCR duplicates in downstream analyses (*Hoffberg et al., 2016*).

Sequence reads generated for our study are deposited in NCBI PRJNA378762 (https://www.ncbi.nlm.nih.gov/bioproject/378762). All scripts used to analyze data and VCF files produced are available from Dryad (10.5061/dryad.tr87dh0). A video presentation giving an overview of the 3RAD system is available on YouTube (https://www.youtube.com/watch?v=ZOmwOtfP3N4). Also, protocols and preprints are available at http://baddna.uga.edu/.

CONCLUSIONS

In summary, 2RAD/3RAD protocols function similar to other dual-digest RADseq methods but are easier to perform, have lower startup costs, have increased utility with low-concentration input DNA, and produce libraries that can be highly-multiplexed and pooled with other Illumina libraries.

ACKNOWLEDGEMENTS

Portions of the oligonucleotide sequences are ©2007–2019 Illumina, Inc., all rights reserved. Derivative works, such as the oligonucleotides presented here, are authorized for use with Illumina instruments and products only. All other uses are strictly prohibited. The information we present allows all researchers to synthesize the oligonucleotides at any vendor of their choice, follow or modify the library preparation techniques we have included, and freely publish results simply with proper attribution of this paper and Illumina®. We thank Julian Catchen for modifying Stacks and Isaac Overcast and Deren Eaton for modifying pyRAD to deal with the variable-length dual internal indexes, as well as for incorporating additional enzymes to facilitate data processing of 3RAD data. We thank Erin Lipp for generously sharing laboratory space and equipment. We thank Jane Stewart, Marin Talbot Brewer and Bryan Carstens for suffering through the primal 2RAD development with modified bases in the adapters, our colleagues at the Georgia Genomics and Bioinformatics Core for conducting Illumina sequencing, the Georgia Advanced Computing Resource Center and our colleagues there for support, and Chad Locklear of Integrated DNA Technologies for his assistance with and support for adapter oligonucleotide development and synthesis. We also thank our colleagues involved in the logistics and sample collection for example projects included in this paper.

Funding

This work was supported by DEB-1242241, DEB-1242260, DEB-1136626, DEB-1146440, DEB-1353815, DEB-1405599, DGE-0903734, DGE-1452154, and OISE 0730218 from the U.S. National Science Foundation. Data collection for *Sphyrna lewini* was supported by the DGAPA (PAPIIT IG201215). There was no additional external funding received for this study. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
U.S. National Science Foundation: DEB-1242241, DEB-1242260, DEB-1136626, DEB-1146440, DEB-1353815, DEB-1405599, DGE-0903734, DGE-1452154, OISE 0730218.
DGAPA: PAPIIT IG201215.

Competing Interests

Brant C. Faircloth is an Academic Editor for PeerJ. Kerin E. Bentley is currently affiliated with LeafWorks, Inc.

The EHS DNA lab (http://baddna.uga.edu) provides oligonucleotide aliquots and library preparation services at cost, including some oligonucleotides and services that make use of the adapters and primers.

Author Contributions

- Natalia J. Bayona-Vásquez, Brant C. Faircloth and Travis C. Glenn conceived and designed the experiments, performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Troy J. Kieran, Todd W. Pierson and Sandra L. Hoffberg conceived and designed the experiments, performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Peter A. Scott performed the experiments, analyzed the data, prepared figures and/or tables, authored or reviewed drafts of the paper, approved the final draft.
- Kerin E. Bentley performed the experiments, prepared figures and/or tables, approved the final draft.
- John W. Finger prepared figures and/or tables, approved the final draft.
- Swarnali Louha analyzed the data, contributed reagents/materials/analysis tools, approved the final draft.
- Nicholas Troendle performed the experiments, analyzed the data, approved the final draft.
- Pindaro Diaz-Jaimes and Rodney Mauricio contributed reagents/materials/analysis tools, approved the final draft.

Animal Ethics

The following information was supplied relating to ethical approvals (i.e., approving body and any reference numbers):

The University of Tennessee Knoxville IACUC provided full approval for the study on salamanders here presented (2352-0318). The University of Alabama IACUC provided full approval for the study on turtles here presented (11-357).

Field Study Permissions

The following information was supplied relating to field study approvals (i.e., approving body and any reference numbers):

Field studies in salamanders were approved by the Tennessee Wildlife Resources Agency (#3840).

Field studies for turtles were approved by The Alabama Department of Conservation and Natural Resources and US Fish and Wildlife Service (License #2011-2013000110468680).

Field sample collection for *Sphyrna tiburo* was approved by the US Fish and Wildlife Service and Florida Fish & Wildlife Conservation Commission - Fish & Wildlife Research Institute (#TE088913-0).

Data Availability

The 3RAD data is available at NCBI: PRJNA378762.

All scripts used to analyze 3RAD data and VCF files produced are available from Dryad DOI: 10.5061/dryad.tr87dh0.

Supplemental Information

Supplemental information for this article can be found online at http://dx.doi.org/10.7717/peerj.7724#supplemental-information.

REFERENCES

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 DOI 10.1038/nrg.2015.28.

Andrews KR, Hohenlohe PA, Miller MR, Hand BK, Seeb JS, Luikart G. 2014. Trade-offs and utility of alternative RADseq methods: reply to Puritz et al. 2014. *Molecular Ecology* **23**:5943–5946 DOI 10.1111/mec.12964.

Bachmann K. 1970. Feulgen slope determinations of urodele nuclear DNA amounts. *Histochemie* **22**:289–293.

Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLOS ONE* **3**:e3376 DOI 10.1371/journal.pone.0003376.

Cariou M, Duret L, Charlat S. 2013. Is RAD-seq suitable for phylogenetic inference? An in silico assessment and optimization. *Ecology and Evolution* **3**:846–852 DOI 10.1002/ece3.512.

Catchen JM, Amores A, Hohenlohe P, Cresko W, Postlethwait JH. 2011. Stacks: building and genotyping loci de novo from short-read sequences. *G3: Genes, Genomes, Genetics* **1**:171–182 DOI 10.1534/g3.111.000240.

Catchen J, Hohenlohe PA, Bassham S, Amores A, Cresko WA. 2013. Stacks: an analysis tool set for population genomics. *Molecular Ecology* **22**:3124–3140 DOI 10.1111/mec.12354.

Cui J, Ren X, Yu Q. 1991. Nuclear DNA content variation in fishes. *Cytologia* **56**:425–429 DOI 10.1508/cytologia.56.425.

Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics* **12**:499–510 DOI 10.1038/nrg3012.

De Smet WHO. 1981. The nuclear Feulgen-DNA content of the vertebrates (especially reptiles), as measured by fluorescence cytophotometry, with notes on the cell and chromosome size. *Acta Zoologica et Pathologica Antverpiensia* **76**:119–167.

DeWoody YD, DeWoody JA. 2005. On the estimation of genome-wide heterozygosity using molecular markers. *Journal of Heredity* **96**:85–88 DOI 10.1093/jhered/esi017.

Eaton DA. 2014. PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics* **30**:1844–1849 DOI 10.1093/bioinformatics/btu121.

Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotype-by-sequencing approach for high diversity species. *PLOS ONE* **6**:e19379 DOI 10.1371/journal.pone.0019379.

Faircloth BC, Glenn TC. 2012. Not all sequence tags are created equal: designing and validating sequence identification tags robust to indels. *PLOS ONE* **7**:e42543 DOI 10.1371/journal.pone.0042543.

Franchini P, Monné Parera D, Kautt AF, Meyer A. 2017. quaddRAD: a new high-multiplexing and PCR duplicate removal ddRAD protocol produces novel evolutionary insights in a nonradiating cichlid lineage. *Molecular Ecology* **26**:2783–2795 DOI 10.1111/mec.14077.

Geraci NS, Johnston JS, Robinson JP, Wikel SK, Hill CA. 2007. Variation in genome size of argasid and ixodid ticks. *Insect Biochemistry and Molecular Biology* **37**:399–408 DOI 10.1016/j.ibmb.2006.12.007.

Glenn TC, Faircloth BC. 2016. Capturing Darwin's dream. *Molecular Ecology Resources* **16**:1051–1058 DOI 10.1111/1755-0998.12574.

Glenn TC, Nilsen RA, Kieran TJ, Sanders JG, Bayona-Vásquez NJ, Finger JW, Pierson TW, Bentley KE, Hoffberg SL, Louha S, García-De León FJ, del Río Portilla MA, Reed KD, Anderson JL, Meece JK, Aggrey SE, Rekaya R, Alabady M, Bélanger M, Winker K, Faircloth BC. 2019. Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). *PeerJ* **7**:e7755 DOI 10.7717/peerj.7755.

Glenn TC, Schable NA. 2005. Isolating microsatellite DNA loci. *Methods in Enzymology* **395**:202–222 DOI 10.1016/S0076-6879(05)95013-1.

Gomez-Palacio A, Jaramillo-O N, Caro-Riano H, Diaz S, Monteiro FA, Perez R, Panzera F, Triana O. 2012. Morphometric and molecular evidence of intraspecific biogeographical differentiation of Rhodnius pallescens (HEMIPTERA: REDUVIIDAE: RHODNIINI) from Colombia and Panama. *Infection, Genetics and Evolution* **12**:1975–1983 DOI 10.1016/j.meegid.2012.04.003.

Graham CF, Glenn TC, McArthur AG, Boreham DR, Kieran T, Lance S, Manzon RG, Martino JA, Pierson T, Rogers SM, Wilson JY, Somers CM. 2015. Impacts of degraded DNA on restriction enzyme associated DNA sequencing (RADSeq). *Molecular Ecology Resources* **15**:1304–1315 DOI 10.1111/1755-0998.12404.

Gregory TR. 2018. Animal genome size database. *Available at http://www.genomesize.com* (accessed on 20 November 2018).

Hardie DC, Hebert PDN. 2003. The nucleotypic effects of cellular DNA content in cartilaginous and ray-finned fishes. *Genome* **46**:683–706 DOI 10.1139/g03-040.

Hardie DC, Hebert PDN. 2004. Genome-size evolution in fishes. *Canadian Journal of Fisheries and Aquatic Sciences* **61**:1636–1646 DOI 10.7717/peerj.7755.

Harvey MG, Smith BT, Glenn TC, Faircloth BC, Brumfield RT. 2016. Sequence capture versus restriction site associated DNA sequencing for shallow systematics. *Systematic Biology* **65**:910–924 DOI 10.1093/sysbio/syw036.

Heyduk K, Stephens JD, Faircloth BC, Glenn TC. 2016. Targeted DNA region re-sequencing. In: Aransay AM, Trueba JLL, eds. *Field guidelines for genetic experimental designs in high-throughput sequencing.* Switzerland: Springer International Publishing, 43–68 DOI 10.1007/978-3-319-31350-4_3.

Hinegardner R. 1976. The cellular DNA content of sharks, rays and some other fishes. *Comparative Biochemistry and Physiology* **55B**:367–370.

Hoffberg SL, Kieran TJ, Catchen JM, Devault A, Faircloth BC, Mauricio R, Glenn TC. 2016. RADcap: Sequence capture of dual-digest RADseq libraries with identifiable duplicates and reduced missing data. *Molecular Ecology Resources* **16**:1264–1278 DOI 10.1111/1755-0998.12566.

Hoffberg SL, Troendle NJ, Glenn TC, Mahmud O, Louha S, Chalopin D, Bennetzen JL, Mauricio R. 2018. A high-quality reference genome for the invasive mosquitofish *Gambusia affinis* using a Chicago library. *G3: Genes, Genomes, Genetics* **8**:1855–1861 DOI 10.1534/g3.118.200101.

Hosner PA, Faircloth BC, Glenn TC, Braun EL, Kimball RT. 2016. Avoiding missing data biases in phylogenomic inference: an empirical study in the landfowl (Aves: Galliformes). *Molecular Biology and Evolution* **33**:1110–1125 DOI 10.1093/molbev/msv347.

Krueger F, Andrews SR, Osborne CS. 2011. Large scale loss of data in low-diversity Illumina sequencing libraries can be recovered by deferred cluster calling. *PLOS ONE* **6**:e16607 DOI 10.1371/journal.pone.0016607.

Lamatsch DK, Steinlein C, Schmid M, Schartl M. 2000. Noninvasive determination of genome size and ploidy level in fishes by flow cytometry: detection of triploid Poecilia formosa. *Cytometry* **39**:91–95 DOI 10.1002/(SICI)1097-0320(20000201)39:2<91::AID-CYTO1>3.0.CO;2-4.

Licht LE, Lowcock LA. 1991. Genome size and metabolic rate in salamanders. *Comparative Biochemistry and Physiology* **100B**:83–92.

Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* **17**:240–248 DOI 10.1101/gr.5681207.

Mitra A, Skrzypczak M, Ginalski K, Rowicka M. 2015. Strategies for achieving high sequencing accuracy for low diversity samples and avoiding sample bleeding using Illumina platform. *PLOS ONE* **10**:e0120520 DOI 10.1371/journal.pone.0120520.

Ojima Y, Yamamoto K. 1990. Cellular DNA contents of fishes determined by flow cytometry. *La Kromosomo II* **57**:1871–1888.

Olmo E. 1974. Further data on the genome size in the urodeles. *Bollettino di Zoologia* **41**:29–33 DOI 10.1080/11250007409430082.

Olmo E. 1976. Genome size in some reptiles. *Journal of Experimental Zoology* **195**:305–310 DOI 10.1002/jez.1401950215.

Palmer MJ, Bantle JA, Guo X, Fargo WS. 1994. Genome size and organization in the ixodid tick Amblyomma americanum (L.). *Insect Molecular Biology* **3**:57–62.

Pante E, Abdelkrim J, Viricel A, Gey D, France SC, Boisselier MC, Samadi S. 2015. Use of RAD sequencing for delimiting species. *Heredity* **114**:450–459 DOI 10.1038/hdy.2014.105.

Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE. 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *PLOS ONE* **7**:e37135 DOI 10.1371/journal.pone.0037135.

Poland JA, Rife TW. 2012. Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome* **5**:92–102 DOI 10.3835/plantgenome2012.05.0005.

Pritchard JK, Stephens M, Donelly P. 2000. Inference of population structure using multilocus genotype data. *Genetics* **155**:945–959.

R Core Team. 2018. R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. *Available at* https://www.R-project.org/.

Salas-Linaza R, Oono R. 2018. Double-digest RADseq loci using standard Illumina indexes improve deep and shallow phylogenetic resolution of *Lophodermium*, a widespread fungal endophyte of pine needles. *Ecology and Evolution* **8**:6638–6651 DOI 10.1002/ece3.4147.

Schwartz FJ, Maddock MB. 1986. Comparisons of karyotypes and cellular DNA contents within and between major lines of elasmobranchs. In: Uyeno T, Arai R, Taniuchi T, Matsuura K, eds. *Indo-Pacific fish biology.* Tokyo: Ichthyological Society of Japan, 148–157.

Schweyen H, Rozenberg A, Leese F. 2014. Detection and removal of PCR duplicates in population genomic ddRAD studies by addition of a degenerate base region (DBR) in sequencing adapters. *The Biological Bulletin* **227(2)**:146–160 DOI 10.1086/BBLv227n2p146.

Scott PA, Glenn TC, Rissler LJ. 2017. Resolving taxonomic turbulence and uncovering cryptic diversity in the musk turtles (*Sternotherus*) using robust demographic modeling. *Molecular Phylogenetics and Evolution* **120**:1–15 DOI 10.1016/j.ympev.2017.11.008.

Shafer A, Peart CR, Tusso S, Maayan I, Brelsford A, Wheat CW, Wolf JB. 2016. Bioinformatic processing of RAD-seq data dramatically impacts downstream population genetic inference. *Methods in Ecology and Evolution* **8**:907–917 DOI 10.1111/2041-210X.12700.

Stingo V, Capriglione T, Rocco L, Improta R, Morescalchi A. 1989. Genome size and A-T rich DNA in selachians. *Genetica* **79**:197–205 DOI 10.1007/BF00121513.

Streicher JW, Schulte JA, Wiens JJ. 2015. How should genes and taxa be sampled for phylogenomic analyses with missing data? An empirical study in iguanian lizards. *Systematic Biology* **69**:929–939 DOI 10.1093/sysbio/syv058.

Tiersch TR, Chandler RW, Wachtel SS, Elias S. 1989. Reference standards for flow cytometry and application in comparative studies of nuclear DNA content. *Cytometry* **10**:706–710 DOI 10.1002/cyto.990100606.

Toonen RJ, Puritz JB, Forsman ZH, Whitney JL, Fernandez-Silva I, Andrews KR, Bird CE. 2013. ezRAD: a simplified method for genomic genotyping in non-model organisms. *PeerJ* **1**:e203 DOI 10.7717/peerj.203.

Ullmann AJ, Lima CMR, Guerrero FD, Piesman J, Black WC. 2005. Genome size and organization in the blacklegged tick, Ixodes scapularis and the Southern cattle tick, Boophilus microplus. *Insect Molecular Biology* **14**:217–222.

Wang S, Meyer E, McKay JK, Matz MV. 2012. 2b-RAD: a simple and flexible method for genome-wide genotyping. *Nature Methods* **9**:808–810 DOI 10.1038/nmeth.2023.