Adapterama III: Quadruple-indexed, triple-enzyme RADseq libraries for about $1USD per Sample (3RAD)

Glenn TC[1-4], Bayona-Vásquez NJ[1,5*], Kieran TJ[1], Pierson TW[1,6], Hoffberg SL[3,†], Scott PA[7,‡], Bentley KE[3], Finger JW Jr.[1,4,§], Watson PR[1], Louha S[2], Troendle N[3,Ψ], Díaz-Jaimes P[5], Mauricio R[3], Faircloth BC[8]

[1]Department of Environmental Health Science, University of Georgia, Athens, GA 30602, USA
[2]Institute of Bioinformatics, University of Georgia, Athens, GA 30602, USA
[3]Department of Genetics, University of Georgia, Athens, GA 30602, USA
[4]Interdisciplinary Toxicology Program, University of Georgia, Athens, GA 30602, USA
[5]Laboratorio de Genética de Organismos Acuáticos, Instituto de Ciencias del Mar y Limnología, Universidad Nacional Autónoma de México. Apdo. Postal 70-305, México D.F., 04510, México
[6]Department of Ecology and Evolutionary Biology, University of Tennessee, Knoxville, TN 37996, USA
[7]Department of Biological Sciences, University of Alabama, Tuscaloosa, AL 35487, USA
[8]Department of Biological Sciences and Museum of Natural Science, Louisiana State University, Baton Rouge, LA 70803, USA
[†] current address: Department of Ecology, Evolution and Environmental Biology, Columbia University, New York, NY 10027, USA
[‡]current address: Department of Ecology and Evolutionary Biology, University of California, Los Angeles, CA 90095, USA
[§]current address: Department of Biological Sciences, Auburn University, Auburn, AL 36849, USA
[Ψ]current address: Department of Science and Mathematics, MidAmerica Nazarene University, 2030 E. College Way, Olathe, KS 66062, USA

Keywords: 3RAD, dual digest RADseq, ddRAD, DNA libraries, multiplexing, NGS.

[*]**Correspondence:** Natalia J. Bayona-Vásquez, Dept. of EHS, Environmental Health Science Bldg., University of Georgia, Athens, GA 30602, USA; Tel: 706-542-3636; fax: 706-542-7472; E-mail: njbayona@uga.edu

Running Title: 3RAD – triple-enzyme RADseq

ABSTRACT

Molecular ecologists have used genome reduction strategies that rely upon restriction enzyme digestion of genomic DNA to sample consistent portions of the genome from individuals being studied (e.g., RADseq, GBS). However, researchers often find the existing methods expensive to initiate and/or difficult to implement consistently. Here, we present a low-cost and highly robust approach for the construction of dual-digest RADseq libraries. 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; 5) integrating adapter designs that can be used interchangeably; 6) incorporating variable-length internal tags within the adapters to increase the scope of sample tagging and facilitate pooling while also increasing sequence diversity; 7) maintaining compatibility with universal dual-indexed primers, which facilitate construction of combinatorial quadruple-indexed libraries that are compatible with standard Illumina sequencing reagents and libraries; and, 8) easy tuning for molecular tagging and PCR duplicate identification. We present eight adapter designs that work with 72 restriction enzyme combinations, and we demonstrate our approach by the use of one set of adapters and one set of restriction enzymes across a variety of non-model organisms to discover thousands of variable loci in each species.

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 large populations of organisms with genome sizes greater than about 0.5 Gbp (i.e., most plants and animals). However, many questions can be addressed with a small fraction of the genome (Andrews, Good, Miller, Luikart, & Hohenlohe, 2016; Cariou, Duret, & Charlat, 2013; DeWoody & DeWoody 2005; Pante *et al*. 2015). Thus, researchers have created a variety of strategies to sample a consistent portion of the genome from large numbers of individuals at low cost (Glenn & Faircloth 2016; Harvey, Smith, Glenn, Faircloth, & Brumfield, 2016; Heyduk, Stephens, Faircloth, & Glenn, 2016). One of the most popular genome sampling strategies uses restriction enzymes (REs) to reduce genome complexity and sequence a set of orthologous loci across individuals (Restriction site Associated DNA sequencing, RADseq; Baird *et al*. 2008; Miller, Dunham, Amores, Cresko, & Johnson, 2007). Key advantages of the RADseq approach include: 1) the lack of a need for a reference genome (although it is better to have one [e.g. Shafer *et al.* 2016]); 2) the relative low-cost of library preparations; 3) the applicability of techniques with minimal modification across a broad spectrum of organisms; and, 4) the availability of suitable software for data analyses (e.g. Stacks; Catchen, Amores, Hohenlohe, Cresko, & Postlethwait, 2011; Catchen, Hohenlohe, Bassham, Amores, & Cresko, 2013; pyRAD; Eaton 2014).

Many variants of the general RADseq approach have been developed (Andrews *et al.* 2016), including but not limited to: the original method (RAD; Baird *et al.* 2008; Miller *et al.* 2007), genotype-by-sequencing (GBS; Elshire *et al.* 2011), 2-enzyme GBS (Poland & Rife 2012), dual-digest RADseq (ddRAD; Peterson, Weber, Kay, Fisher, & Hoekstra, 2012), 2bRAD (Wang, Meyer, McKay, & Matz, 2012), ezRAD (Toonen *et al.* 2013), and high multiplexing with PCR duplicate removal (quaddRAD; Franchini, Monné Parera, Kautt, & Meyer, 2017). Although each of these RADseq approaches is in widespread use, they have well-known limitations

(Davey *et al*. 2011, Andrews *et al*. 2014) which have been thoroughly examined in recent reviews (Andrews *et al*. 2016, Harvey *et al*. 2016, Heyduk *et al.* 2016). Limitations of the 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 tagged such that individual organisms can be genotyped. Limitations of current individual-based RADseq methods include: 1) high up-front costs for adapters, which limits experimental flexibility; 2) the need for moderate to high amounts of high-molecular-weight DNA; 3) variable consistency in loci obtained; 4) limited ability to multiplex RADseq libraries with other types of libraries; and, 5) workflows of varying complexity. Thus, developing RADseq methods that reduce the cost of adapters, reduce the amount of input DNA, improve flexibility, increase consistency, and simplify the workflow would be helpful to many researchers.

Dual-digest RADseq (ddRAD, Peterson *et al.* 2012) is a popular form of RADseq, because: 1) it allows preparation of DNA using only REs (no random shearing, thus reducing the time and cost per sample for sample prep); 2) sequencing can be completed from each end with consistent starting positions (therefore focusing sequencing efforts and increasing depth per dollar invested); 3) the number of loci can be tuned based on the choice of enzymes and the size-range of DNA that is selected for sequencing; and, 4) a combinatorial dual-indexing strategy is used, allowing pooling within sets of 48 samples (half of a 96-well plate) and indexing of up to 12 pools. These strengths have led to many researchers adopting ddRAD. However, limitations of ddRAD include: 1) the cost of synthesizing the adapters (~$4,550 USD, for 12 pools of 48 samples; Peterson *et al.* 2012); 2) that adapters to both ends are phosphorylated, thus each can form adapter dimers and yield incorrect constructs; 3) the use of an optional step to incorporate a biotin-containing primer and streptavidin beads to separate correct library constructs from incorrect constructs (often skipped by resource-limited researchers); 4) the inability to reduce

chimera formation; and, 5) the fact that no sequence diversity is present across the restriction recognition sequence of the resulting libraries, which constrains sequencing options.

Here, we present a new approach (3RAD) for preparing RADseq libraries, which is similar in spirit and retains and builds upon the strengths of ddRAD, while also addressing each of the limitations listed above. Our method: 1) reduces costs of adapters and sample preparation; 2) uses only a single phosphorylated adapter; 3) does not need an optional step to separate correct constructs; 4) digests chimeras; 5) achieves sequence diversity across the restriction recognition sequence; and, 6) is compatible with other Illumina libraries. Overall, our 3RAD method uses DNA that is digested with multiple REs and ligates on adapters in the presence of functional REs, followed by PCR with primers developed in Adapterama I (Glenn *et al.* 2016) to make fully active quadruple-indexed Illumina libraries (Figs 1 and S1). Although some adapter designs and the working procedures have been published (e.g., Graham *et al.* 2015; Hoffberg *et al.* 2016), 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 then discuss these results and additional work to extend this approach.

Methods

Methodological Objectives

Our overall goal was to develop a ddRADseq-type method with the following characteristics: 1) enabling a simplified workflow with few consecutive buffer exchanges; 2) allowing sequential or simultaneous digestion of DNA and ligation of adapters; 3) reducing chimera formation; 4) suppressing adapter dimer formation to increase library efficiency; 5) enabling hierarchical

combinatorial tagging to facilitate efficient multiplexing of many samples; 6) reducing costs, both for initial buy-in (i.e., cost of all reagents to start using the method) and per sample prepared; and, 7) facilitating pooling of the resulting libraries with any other Illumina library type. We built upon the adapter design and methods of Glenn and Schable (2005) to achieve the first four goals, whereas we extended the work of Faircloth and Glenn (2012) and Glenn *et al.* (2016) to achieve the last three goals (Appendix S1 and Appendix S2).

Methodological Approach

Design goals 1-3 can be achieved by using reagents that allow simultaneous digestion of the sample DNA and ligation of the adapters onto the sample DNA (Glenn & Schable, 2005). A major distinction between methods of Glenn and 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; Supporting Information Figure S1, and S2) in 3RAD, whereas the other adapter can use plain oligonucleotides for both strands (Appendix S3), forming fewer adapter chimeras.

Simultaneous processing of DNA with multiple enzymes requires that the enzymes are active in the same buffer and at the same temperature. New England Biolabs (NEB; Beverly, MA, USA) has developed many enzymes that retain high activity in one buffer (CutSmart®) and NEB has long described activity of their enzymes in their other standard buffer formulations. We sought sets of low-cost type II REs that could be used in groups of three (Table 1): two REs that form unique cohesive-ends (i.e., incompatible sticky-ends) and a third RE that produced a compatible cohesive end with one of the other REs (Fig. 2). We then paired the two REs with

compatible cohesive ends into sets with Illumina Read 1 adapter stub sequences (Glenn *et al.* 2016), and assigned unpaired REs 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 RE, thus achieving design goal 4 (Appendix S3) (Glenn and Schable, 2005). Similarly, Read 1 adapters ligated to genomic DNA with third RE cut sites will recreate the recognition sequence for the third RE. Because T4 DNA ligase can be used in the same buffers as most REs if the buffers are supplemented with ATP, researchers can start by digesting the 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 adapters ligated onto the ends; Supporting Information, Figure S1, and S2).

In addition to the digestion of adapter dimers, we constructed the 3RAD adapters so that each double-stranded adapter has one active strand (the bottom strand as shown in all figures herein) and one “dummy” strand (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 (Fig. S1). Thus, the dummy 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.

Design goals 5-7 were achieved by including variable-length internal indexing tags—also known as "in-line barcodes" (Andrews *et al*. 2016)—within the Read 1 and Read 2 adapter stubs, then making the adapter stubs compatible with the primers of Glenn *et al.* (2016; Figs 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 (Appendix S3). Each adapter stub version includes an internal indexing tag of 5, 6, 7, or 8 nucleotides (nt). The purpose of these internal tags is twofold: (1) combinations of the Read 1 and Read 2 adapters creates 96 (8 x 12) combinations, which facilitates pooling of samples from 96-well plates (Appendix S3), and (2) the variable length of each tag increases base diversity within pools of libraries (Krueger, Andrews, & Osborne, 2011) which is important when sequencing libraries derived from RE digestion (Glenn *et al*. 2016; Mitra, Skrzypczak, Ginalski, & Rowicka, 2015). Finally, the complete 3RAD libraries are compatible with (i.e., can be pooled with) iTru and Illumina®™ TruSeq libraries prior to sequencing on Illumina sequencing instruments because they are constructed using the iTru primers from Adapterama I (Glenn *et al.* 2016).

Methodological Overview

As an example of the 3RAD method (see Figs 2 and S1), consider sample DNA cut with two different REs that leave different sticky ends – e.g., XbaI on the left-hand, Read 1 side and EcoRI-HF on the right-hand, Read 2 side. 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 adapter (compatible with EcoRI-HF) is 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 XbaI restriction site used to cut the sample DNA (Fig. 2). Because the iTru Read 1 adapter is phosphorylated on the 5' end, it can and will form dimers. Read 1 adapter dimers form NheI restriction sites. By

adding a third enzyme (i.e., NheI) that digests such dimers, the concentration of adapter dimers is reduced. This digestion of adapter dimers increases the consistency and efficiency of 3RAD library preparation, even with limited amounts of sample DNA. However, while genomic DNA cut by NheI and ligated to Read 1 adapters should be recut by the same RE, any remaining molecules of this form are suitable for PCR amplification and may be present in final libraries.

After ligating the adapters, full-length libraries are created through reduced cycle PCR using the iTru5 and iTru7 primers of Glenn *et al.* (2016; Figs 1, 3, S1, and S2). Because the 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 PCR amplify individual wells with unique external tags, and pool the resulting amplifications, allowing redundant tagging (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).

3RAD Applied Case Studies

Study systems

We tested the 3RAD protocol on eight example projects: Kinosternidae (turtles), Ixodidae (ticks), *Eurycea bislineata* species complex (salamanders), *Wisteria floribunda* x *Wisteria sinensis* hybrid population (plants), *Rhodnius pallescens* (insects), *Gambusia affinis* (freshwater fishes), *Sphyrna tiburo* (sharks), and *Sphyrna lewini* (sharks). Each dataset comprised 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 enzyme combinations, we tested RE combinations for each of these

species corresponding to adapter sets R1.A and R2.1 (Table 1 and Design 1 in Appendix S3). We used the enzymes XbaI, EcoRI-HF, and NheI for all example projects.

Adapter preparation

Oligonucleotide sequences for all versions of all adapter designs are given in Supplementary File 3. Oligonucleotides were ordered in plates from Integrated DNA Technologies (Coralville, IA). Adapters were solubilized in a solution of 10 mM Tris pH 8, 0.1 mM EDTA, and 100 mM NaCl, mixed in equal-molar amounts, annealed, diluted and aliquoted (Appendix S4).

Library preparation

Similar methods were employed for all samples, but some details varied among species, as the protocols were developed and implemented over the course of multiple years. In general, we focus on the version of our protocol that does not include molecular ID tags (Appendix S1 and Appendix S2), except for *Wisteria* samples where molecular ID tags were used (see Hoffberg *et al.* 2016, Appendix S5, and Appendix S6 for protocols that employ molecular ID tags).

All enzymes and reagents were obtained from NEB. We digested sample DNA for 1 hr at 37 °C in a reaction mix that consisted of: 1.5 µL 10x CutSmart® Buffer, 5.0 µL dH_2O, 0.5 µL of XbaI at 20 U/µL, 0.5 µL of EcoRI-HF at 20 U/µL, 0.5 µL of NheI at 20 U/µL, 1 µL 5 µM double-stranded iTru R1.A adapter, 1 µL 5 µM double-stranded iTru R2.1 adapter, and 5 µL sample DNA. After digestion, we added 2.0 µL dH_2O, 1.5 µL ATP (10 µM), 0.5 µL 10x Ligase Buffer, 1.0 µL T4 DNA Ligase (100 units/µL, NEB M0202L buffer diluted 1:3 in NEB B8001S enzyme dilution buffer) to each reaction, and we incubated the digestion/adapter-ligation mixtures in a thermal cycler with the following conditions: 22 °C for 20 min. and 37 °C for 10 min. for two cycles followed by a single cycle of 80 °C for 20 min. After adapter ligation, we

pooled all *Wisteria*, *Gambusia*, and Kinosternidae individuals by project; we maintained *Eurycea*, *Rhodnius*, *Sphyrna*, and Ixodidae samples individually. We proceeded immediately to a pre-PCR clean-up (Appendix S1) to remove remaining reagents and unincorporated adapters, using Sera-Mag Speed-Beads (Thermo-Scientific, Waltham, MA, USA; see Glenn *et al.* 2016 for preparation methods) at a ratio of 1.2:1 Speed-Beads to DNA, washing and re-suspending DNA in 20 µL of TLE.

To generate full-length library constructs, we combined 5.0 µL of ligated-cleaned up DNA with 5.0 µL Kapa HiFi Buffer (Kapa Biosystems, Wilmington, MA), 0.75 µL dNTPs (10 µM), 8.75 µL dH_2O, 0.5 µL Kapa HiFi DNA Polymerase (1 unit/µL), 2.5 µL iTru5 primer (5 µM), and 2.5 µL iTru7 primer (5 µM), and we amplified the adapter-ligated DNA in a thermal cycler with the following thermal profile: 95 °C for 2 min.; then, 14 cycles of 98 °C for 20 sec., 60 °C for 15 sec., 72 °C for 30 sec.; 72 °C for 5 min.; hold at 15 °C. To validate that the library preparation process was successful, we ran 5 µL of PCR product with 2 µL loading dye on a 1.5% agarose gel for 45 minutes at 90 volts along with Hi-Lo DNA Marker (Minnesota Molecular, Minneapolis, MN). A smear of evenly distributed and bright DNA around ~300-800 bp, without noticeable bands in this target size zone, indicated successful library preparation. After validation, we cleaned the remaining PCR reaction volume with Speed-Beads in at least 1:1 (Speed-Beads:DNA) ratio, and we eluted cleaned DNA in 20 µL of TLE.

We quantified cleaned libraries using either a Qubit Fluorimeter (Life Technologies, Inc.) or SYBR Green assay with a plate reader, and for those projects with samples not already pooled after ligation, we pooled equal amounts prior to size-selection. Pools contained up to 96 samples and totaled 1-1.8 µg. We size-selected pooled libraries using a Pippin Prep (Sage Science, Beverly, MA) with a 1.5% dye-free Marker K agarose gel cassette (CDF1510) set to capture fragments at 550 bp +/-10%. If < 1 ng/µL of DNA was recovered following size-selection, the libraries were amplified with P5 and P7 primers (Glenn *et al.* 2016) with 6-12 cycles of 98 °C for

20 sec., 60 °C for 15 sec., 72 °C for 30 sec.; followed by 72 °C for 5 min. and holding at 15 °C. We cleaned the PCR reaction with Speed-Beads in at least 1:1 (Speed-Beads:DNA) ratio, and we eluted cleaned DNA in 25µL of TLE.

Sequencing and data analyses

We combined each sample pool with individually tagged samples from other projects. Most pools presented here were sequenced using independent runs making use of multiple Illumina sequencers 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. All projects presented here received approximately 1 million reads per sample, which facilitates comparison of the 3RAD method among species with varying genome sizes, except in the Ixodidae project, where the average number of reads per sample is 4 million reads. More detailed analyses using additional samples from individuals of the same species, with varying numbers of reads, will be presented in other manuscripts for each species.

Each set of samples from each species or group was run independently in the software Stacks v1.42 or v.1.44 (Catchen *et al.* 2013; Catchen *et al*. 2011; see Appendix S6). For the *Wisteria* project, molecular ID tags were used to facilitate PCR duplicate removal using the module *clone_filter* from Stacks (Catchen *et al.* 2013; Hoffberg *et al*. 2016). We used the *process_radtags* program to demultiplex and/or clean and trim the sequence data. We removed any read with an uncalled base (-c) and discarded reads with low quality scores (-q) with a default sliding window of 15% of the length of the read and raw Phred score of 10, which means if the score drops below 90% probability of being correct, the read is discarded. XbaI and EcoRI were set as REs, and we rescued sequence tags (internal tags) and RAD-tags (enzyme over-hang) within 2bp of their expected sequence (-r); otherwise, reads were discarded. We truncated (-t)

PE150 reads to 140 nt and PE75 reads to 64 nt to have equal length among all reads with different barcodes.

We parallel-merged the mates of paired-end reads (using the 'paste –d' Unix command). We ran the Stacks *denovo_map* pipeline on each species or group with the following settings: the minimum number of identical raw reads required to create a stack (-m) to 3, the maximum distance between stacks (-M) to 3, the number of mismatches allowed between sample tags when generating the catalog (-n) to 4. The coverage, number of loci, and number of SNPs recovered were scored for each species and compared to genome size and sequencing read length (PE75 or PE150). Next, we ran the Stacks *populations* program with the minimum number of populations to consider a locus (-p) set to between 60 and 75% of the total of the localities considered for each species and the minimum percentage of individuals within a population required to process a locus (-r) set to 75% for all species.

Because the Kinosternidae project comprises samples of multiple genera and species, we conducted a *de novo* locus assembly using pyRAD1.0.4 (Eaton, 2014; details in Appendix S6). *Gambusia affinis* was the only species with a reference genome available (http://gambusia.genetics.uga.edu/). Thus, we also analyzed the *Gambusia* sequences using the *ref_map* program from Stacks v. 1.44 (Catchen *et al*. 2013). We used the Burrow-Wheeler aligner (bwa) 0.7.10 to index the genome and align independent paired-end reads from each sample against the reference (bwa-mem, unpublished). The resulting paired-reads files for each sample were merged (using the 'cat' Unix command). Then, we used these files as input in the *ref_map* pipeline from Stacks using a minimum number of identical reads to create a stack (-m) of three, similar to our *de novo* approach (more details in Appendix S6).

For population level datasets, we calculated F-statistics, and performed preliminary Structure v. 2.3.4 (Pritchard, Stephens, & Donelly, 2000) analyses using burn-in and sampling lengths of 10,000 and 100,000 MCMC repeats, respectively. We did this with an admixture

model with correlated allele frequencies and with three iterations for each value of *K* to obtain the posterior probability of the number of clusters for each data set (Appendix S6). In Stacks or pyRAD, we generated VCF files for each data set (http://dx.doi.org/10.5061/dryad.285j2).

Finally, we estimated the prevalence and impact of loci with third RE cut sites in our data. We estimated the proportion of this third RE cut site relative to the first RE cut site (i.e., intended cut site) for five of the projects presented here. To evaluate the effect of these loci in downstream analyses, we reanalyzed two of our projects (i.e., both *Sphyrna* projects) after removing third RE loci from the datasets. To do this, we reran Stacks v.1.44 (Catchen *et al.* 2013; Catchen *et al.* 2011) using *process_radtags* two independent times: 1) 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, 2) using the previous step's output as input, checking only for exact, intended RE cut sites (i.e., XbaI and EcoRI). From this output, we parallel-merged the mates of paired-end reads, and ran the *denovo_map*, *populations*, and Structure v. 2.3.4 (Pritchard *et al.* 2000) programs as described above.

Results

We developed four sets of adapters each with eight variants for the Read 1 adapters and 12 variants for the Read 2 adapters (Appendix S3). The iTru_R2_5 tag sequence was modified for BamHI because this tag creates a BamHI recognition site in the adapter; otherwise, all adapters use a universal set of tag sequences. The cost of synthesizing oligonucleotides for the adapters varies with synthesis scale, but starts as low as ~$350 (US) per design set (IDTDNA.com), with a recommended scale (100 nmol) costing ~$500 (US) per design set when synthesized into 96-well plates, useful for up to ~4800 sample libraries.

3RAD libraries can be constructed routinely with approximately 12 hours 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 (Appendix S3, "Library_prep_costs" Sheet). If samples are amplified individually, this adds > $1 per sample, but if the ligations are pooled, 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, the total cost for library preparation is about $1.35 per sample, when samples are amplified in pools of 96.

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. Except for one sample in the Ixodidae project, we always recovered a high percentage of retained reads (78.9- 99.7%) after cleaning and filtering steps (Supporting Information, Table S1). Average coverage per locus varied from 6x for *Eurycea* to 70x for *Gambusia* (Fig. 5, Table 3, Supporting Information Table S1).

The number of loci obtained for each dataset varied from 18,629 in *Gambusia* to 425,729 in *Eurycea*. After filtering to retain only polymorphic loci found in at least 75% of individuals within each population in each dataset, we recovered between 30 (*Eurycea*) to 19,843 (Ixodidae) loci containing between 360 and 69,518 SNPs, respectively (Table 3). As expected for RADseq protocols, the number of loci obtained in the initial steps is proportional to the genome size, with more loci generally in organisms with larger genomes. The final number of loci recovered is dependent upon the intrinsic genetic variability of the organism (e.g., evolutionary distance between samples in dataset) and sequencing coverage (Supporting Information, Figure S3). Detailed results for each project can be found in Appendix S6.

Third RE loci were present in all datasets, comprising an average of 20.5% (sd = 14.1%) of all reads. The percentage of reads from third RE loci varied both among and within datasets, showing a nonrandom pattern with respect to R1 adapter index used (Figure S4). Removing third RE loci from our raw reads increased mean coverage of remaining loci (Tables S2 and S3)

and reduced the size of our final datasets from 7183 to 6738 loci in *Sphyrna tiburo* and 5263 to 4807 loci in *Sphyrna lewini*. Estimates of F_{ST} and results from Structure were qualitatively similar in analyses including and excluding third RE loci (Table S4).

Discussion

We have developed a flexible low-cost system for preparing dual-digest RADseq libraries that is efficient and easy to implement. Our method is built on the foundational approach of Adapterama I (Glenn *et al.* 2016) that utilizes universal stubs (adapters) and primers, both with indexed tags, to allow a single Illumina run to be shared among quadruple-indexed libraries from many individuals. To illustrate the utility of our method, we presented summary statistics from analyses of eight small example projects, representing diversity in taxonomy and scientific objectives. For each project, we obtained tens to thousands of loci containing SNPs for downstream population genetic and phylogenetic analyses, and among 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 in studying variation among populations of *S. depressus* and across relatively deep evolutionary time (~55 Myr; Appendix S6). 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, increasing evidence supports the utility of these large and sparse data matrices for systematic-level questions (e.g. Hosner, Faircloth, Glenn, Braun, & Kimball, 2016; Streicher, Schulte, & Wiens 2015). For detailed discussion of the analyses for each dataset, see Appendix S6.

By using the same set of enzymes on DNA from a diverse set of organisms, we have demonstrated that our 3RAD method produces reasonable numbers of loci and SNPs from organisms with varying genome characteristics. As expected, when we use the same set of enzymes (and thus similar expected frequency of cutting within the different genomes) and the same size selection criteria for organisms with varying genome size, the average sequencing coverage per locus decreases as genome size increases (Fig. 5). For example, our dataset generated here from *Eurycea* (genome size = ~24 Gbp) produced few loci meeting our coverage criteria, but higher sequencing coverage can remedy this (data not shown). The number of loci in the resulting libraries can, however, be tuned by varying which enzymes are used as well as varying the size-selection criteria (Peterson *et al.* 2012). Initial testing of all four design sets from the 3RAD method on chicken DNA showed that the number of loci varied as expected (i.e., using REs with shorter recognition sequences yielded more loci; data not shown). Additionally, using broad size-distributions allows larger numbers of loci to be selected and gives more size tolerance among alleles, whereas narrow distributions yields more limited numbers of 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. However, most researchers are targeting loci without size variation, so this bias should be small.

A key advance of the 3RAD workflow (Appendix S1 and S2) is the combination of enzymes and adapters used during digestion, ligation, and amplification 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 are useable with multiple enzymes that leave compatible, cohesive ends (Table 2). Thus, there are at least 72

different enzyme combinations possible with the current adapter sets. Although this flexibility is desirable, having adapters compatible with multiple enzymes means that it was necessary to name the adapters based on the third enzyme, which can be confusing because the third enzyme is *not* the desired restriction site in the sample libraries (and their resulting reads). The design spreadsheet (Appendix 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 enzyme combinations work well with some species, but not others; one cause of this may be the occurrence of selected restriction sites within repetitive elements. Thus, our standard strategy is to empirically determine what enzyme combination is best for any particular organism in a few representative samples (Table 1). We start by looking at the distribution of post-PCR library DNA run through an agarose gel, and exclude enzyme combinations that don't produce good smears in the desired size range or have dense bands in the target size-range. We then size-select and sequence libraries from one or two enzyme combinations from this small batch of samples (along with other samples) on an Illumina®™ instrument (HiSeq or NextSeq). Data are then demultiplexed and analyzed to verify that (or determine which of) the enzyme combination(s) works well by determining how many variable loci are obtained.

Our results show that XbaI, EcoRI-HF and NheI provide a suitable combination of REs for a wide range of organisms, including plants, vertebrates, and invertebrates. We have used all of the adapter designs and many other enzyme combinations from the 3RAD enzyme list (Table 1) to survey SNPs in a variety of organisms. We have obtained correct sequencing results from all designs, but we reiterate that not all combinations work well in all organisms (data not shown). Most of the organisms we have studied to date work well with these three REs, and most that do not work with the design 1 adapters and primary enzymes, perform better with the

design 2 adapters and primary enzymes (MspI, ClaI, and BamHI; data not shown). Although viable, we only rarely use the adapters pairs from designs 3 and 4 (R1.C, R1.D, R2.3 and R2.4).

In our standard protocols (Appendix S1 and S5), we cut DNA with three different REs—two enzymes to create sticky ends for adapter ligation (similar to ddRAD and 2-enzyme GBS) and the third enzyme to digest a recognition site formed by self-ligation of the phosphorylated adapters (Fig. 2). Although the third enzyme facilitates creation of libraries with very little input DNA (≤0.1 ng; data not shown), it does come with a cost. The third RE 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 RE. To encourage this, our digestion/ligation cycling ends with a digestion step to cleave as many of these products as possible. 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 the intended RE 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 enzyme is not required for this procedure and further investigation into to trade-offs of including this enzyme 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, using a method known as 2RAD. Unfortunately, the 2RAD 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 2RAD adapters is warranted.

The 3RAD approach is similar to other dual-digest RADseq approaches (e.g., ddRAD and 2-enzyme GBS), and most of the advantages of the general approach have been described previously. The major advantages of the specific adapters described here include: 1) the buy-in cost of the adapters is much lower than other approaches, 2) with the additional third enzyme, the

phosphorylated adapters that ligate together get cut apart, thus much lower amounts of input DNA are necessary and the protocol is tolerant of varying amounts of input DNA, especially small amounts of input DNA; 3) multiple enzymes are compatible with many of the adapters (Table 1); 4) the TruSeq-compatible stubs facilitate combination of these libraries with dual-indexed Illumina TruSeqHT adapters in widespread use, as well as the much larger number of indexed primers given by Glenn *et al.* (2016); 5) because there are four different indexing positions, the total number of oligos needed to uniquely tag large numbers of samples is minimized; 6) all tags used herein conform a minimum edit distance of 3 (Faircloth & Glenn 2012); 7) limited PCR-cycles reduces PCR bias, increasing the efficiency of the library; and 8) as mentioned below, easy incorporation of molecular ID tags that facilitates the detection of PCR duplicates in downstream analyses.

As detailed in Hoffberg *et al.* 2016, our 3RAD approach it is easy to modify to include that allow the removal of PCR duplicates (Appendix S5), through a process comparable to quaddRAD (Franchini *et al*. 2017). Our molecular ID tag protocol, that 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, and reduce missing data, and remove PCR duplicates.

In summary, our 3RAD method produces results similar to others reported in the literature but it is easier to perform, costs less to initiate in a lab, yields consistent results among samples, can produce usable libraries from much less input DNA, and leads to libraries that can either be run alone or pooled with other Illumina libraries.

Acknowledgements

Portions of the oligonucleotide sequences are © 2007-2016 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. We thank: Julian Catchen for modifying Stacks and Isaac Overcast and Deren Eaton for modifying ipyrad to deal with the variable-length dual internal tags, as well as incorporating additional enzymes to facilitate data processing of 3RAD data; Erin Lipp for generously sharing laboratory space and equipment; Jane Stewart, Marin Talbot Brewer and Bryan Carstens for suffering with 2RAD development; our colleagues at the Georgia Genomics Facility for conducting Illumina sequencing; 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 logistics and sample collection for example projects included in this paper. This work was partially supported by DEB-1242241, DEB-1242260, DEB-1136626, DEB-1146440, DEB-1405599, DGE-0903734, DGE-1452154, and OISE 0730218 from the National Science Foundation.

Literature Cited

Andrews, K.R., Good, J.M., Miller, M.R., Luikart, G., & Hohenlohe, P.A. (2016). Harnessing the power of RADseq for ecological and evolutionary genomics. *Nature Reviews Genetics*, **17**, 81-92. doi: 10.1038/nrg.2015.28

Andrews, K.R., Hohenlohe, P.A., Miller, M.R., Hand, B.K., Seeb, J.S., & 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

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., … Johnson, E.A. (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, J., Hohenlohe, P.A., Bassham, S., Amores, A., & Cresko, W.A. (2013). Stacks: an analysis tool set for population genomics. *Molecular Ecology*, ***22***, 3124-3140. doi: 10.1111/mec.12354

Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W., & Postlethwait, J.H. (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

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., & Blaxter, M.L. (2011). Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Reviews Genetics*, ***12***, 499-510. doi: 10.1038/nrg3012

DeWoody, Y.D., & DeWoody, J.A. (2005). On the estimation of genome-wide heterozygosity using molecular markers. *Journal of Heredity*, ***96***, 85-88. doi: 10.1093/jhered/esi017

Eaton, D.A. (2014). PyRAD: assembly of de novo RADseq loci for phylogenetic analyses. *Bioinformatics*, ***30***: 1844-1849. doi: 10.1093/bioinformatics/btu121

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., & Mitchell, S.E. (2011). A robust, simple genotype-by-sequencing approach for high diversity species. *PLoS ONE*, ***6***, e19379. doi: 10.1371/journal.pone.0019379

Faircloth, B.C., & Glenn, T.C. (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, A.F., & 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.

Glenn, T.C., & Faircloth, B.C. (2016). Capturing Darwin's dream. *Molecular Ecology Resources*, ***16***, 1051-1058. doi: 10.1111/1755-0998.12574

Glenn, T.C., Faircloth, B.C., Nilsen, R., Kieran, T.J., Finger, J.W., Pierson, T.W., … Faircloth, B.C. (2016). Adapterama I: Universal stubs and primers for thousands of dual-indexed Illumina libraries (iTru & iNext). *BioRxiv*, doi: 10.1101/049114.

Glenn, T.C., & Schable, N.A. (2005). Isolating microsatellite DNA loci. *Methods in Enzymology*, ***395***, 202-222. doi: 10.1016/S0076-6879(05)95013-1

Graham, C.F., Glenn, T.C., McArthur, A.G., Boreham, D.R., Kieran, T., Lance, S., … Somers, C.M. (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, T.R. (2016, December 16). Animal Genome Size Database. Retrieved from http://www.genomesize.com

Harvey, M.G., Smith, B.T., Glenn, T.C., Faircloth, B.C., & Brumfield, R.T. (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, J.D., Faircloth, B.C., & Glenn, T.C. (2016). Targeted DNA region re-sequencing. In: *Field Guidelines for Genetic Experimental Designs in High-Throughput Sequencing* (eds. Aransay AM, Trueba JLL), pp. 43-68. Springer International Publishing, Switzerland. doi: 10.1007/978-3-319-31350-4_3

Hoffberg, S.L., Kieran, T.J., Catchen, J.M., Devault, A., Faircloth, B.C., Mauricio, R., & Glenn, T.C. (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

Hosner, P.A., Faircloth, B.C., Glenn, T.C., Braun, E.L., & Kimball, R.T. (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, S.R., & Osborne, C.S. (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

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., & Johnson, E.A. (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

Pante, E., Abdelkrim, J., Viricel, A., Gey, D., France, S.C., Boisselier, M.C., & Samadi, S. (2015). Use of RAD sequencing for delimiting species. *Heredity* ***114***, 450-459. doi: 10.1038/hdy.2014.105

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. doi: 10.1371/journal.pone.0037135

Poland, J.A., & Rife, T.W. (2012). Genotyping-by-Sequencing for plant breeding and genetics. *Plant Genome*, ***5***, 92-102. doi: 10.3835/plantgenome2012.05.0005

Pritchard, J.K., Stephens, M., & Donelly, P. (2000). Inference of population structure using multilocus genotype data. *Genetics*, ***155***, 945-959.

Shafer, A., Peart, C.R., Tusso, S., Maayan, I., Brelsford, A., Wheat, C.W., & Wolf, J.B. (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.

Streicher, J.W., Schulte, J.A., & Wiens, J.J. (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

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. doi: 10.7717/peerj.203

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***, 808-810. doi: 10.1038/nmeth.2023

Data Accessibility

Sequence reads generated for study are deposited in the NCBI SRA (SUB2631727). All scripts used to analyze data and VCF files produced are available from Dryad (http://dx.doi.org/10.5061/dryad.285j2). A video presentation giving an overview of the 3RAD system is available on YouTube (https://www.youtube.com/watch?v=ZOmwOtfP3N4).

Author Contributions

TCG and BCF conceived Adapterama and 3RAD; TCG designed the 3RAD adapters; NB, TJK, TWP, SLH, PAS, PRW, and BCF conducted lab work and analyzed data; TCG, NB and BCF drafted most text; NB, TJK, TWP, PAS, JWF Jr., KEB, SL, NT, and SLH made figures, tables, and supporting materials; PD-J and RM provided samples, funding and additional data analysis; all authors edited and commented on the manuscript.

Conflicts of Interest

The authors declare competing interests. The EHS DNA lab provides oligonucleotide aliquots and services at cost, including the adapters presented herein. The information presented herein allows all researchers to synthesize the oligonucleotides at any vendor of their choice and freely publish results with proper attribution to this paper and Illumina®™.

Table 1. Enzyme combinations and characteristics. Four design sets each for Read1 (R1) and Read2 (R2) are given. The required enzyme blocks adapter-dimer formation and is not actually required, but is used for naming each adapter set (supplementary file 3). 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.

Read 1 Adapter Sets								Read 2 Adapter Sets							
Set	Enzyme	NEB Buffer			CpG meth	Cut Temp	Base Cutter	Set	Enzyme	NEB Buffer			CpG meth	Cut Temp	Base Cutter
		2.1	3.1	CutSmart®						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	-	37	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								

* 50% activity at 37°C, ** 30% activity at 37°C

Table 2. Example 3RAD adapter stub sequences. All sequences are given in 5' to 3' orientation. Tag sequences are in italics. Groups of 4 adapters form a balanced set, all eight complete sets are available in Appendix S3. Non-complementary sequences are given in lower case. Adapter stub oligonucleotides must be hydrated and annealed prior to use (Appendix 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	TNAC*CAACGATC*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

Table 3. The classification and genome size of taxa included in this study, number of samples tested for each. The 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 big genomes.

	Class	*Genome size (Gb)*[*]	*# Species / Populations*	*# Indiv.*	*PE Read Length (nt)*	*# Loci*	*# Final Loci*	*# SNPs*	*Mean Coverage (x)*
Kinosternidae	Reptilia	2.6	7	24	75	233,072	4,034[Ψ]	27,881	12
Ixodidae	Arachnida	3.3	4	16	150	332,057	4,484	13,136	36
Eurycea	Amphibia	24.5	1	21	150	425,729	30	360	7
Wisteria	Magnoliopsida	?	1	24	75	30,029	1,669	5,820	44
Sternotherus depressus	Reptilia	2.6	3	12	75	103,240	16,695	25,578	11
Amblyomma americanum	Arachnida	3.3	2	7	150	128,899	19,843	69,518	36
Rhodnius pallescens	Insecta	0.68	5	16	75	92,687	7,779	12,099	23
Gambusia affinis	Actinopterygii	0.8	5	24	75	18,629	2,140	5,429	54
Sphyrna tiburo	Chondrichthyes	3.8	6	24	150	42,705	7,183	17,555	18
Sphyrna lewini	Chondrichthyes	3.6	7	15	150	44,125	5,263	12,272	27

[*]Genome sizes are approximations from Gregory T.R. (2016, December 16). Animal Genome Size Database. Retrieved from http://www.genomesize.com

[Ψ]From pyRAD assembly of homologous loci across all Kinosternidae.

Figure 1. Overview 3RAD library construction. Genomic DNA from the study organism is digested with two enzymes (A and B), which is shown at the top. 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 subjected to limited cycle PCR with iTru5 and iTru7 primers (Glenn et al. 2016) to form the fully active double-stranded DNA molecules. The color-scheme follows those of Glenn *et al*. (2016) and Hoffberg *et al* (2016).

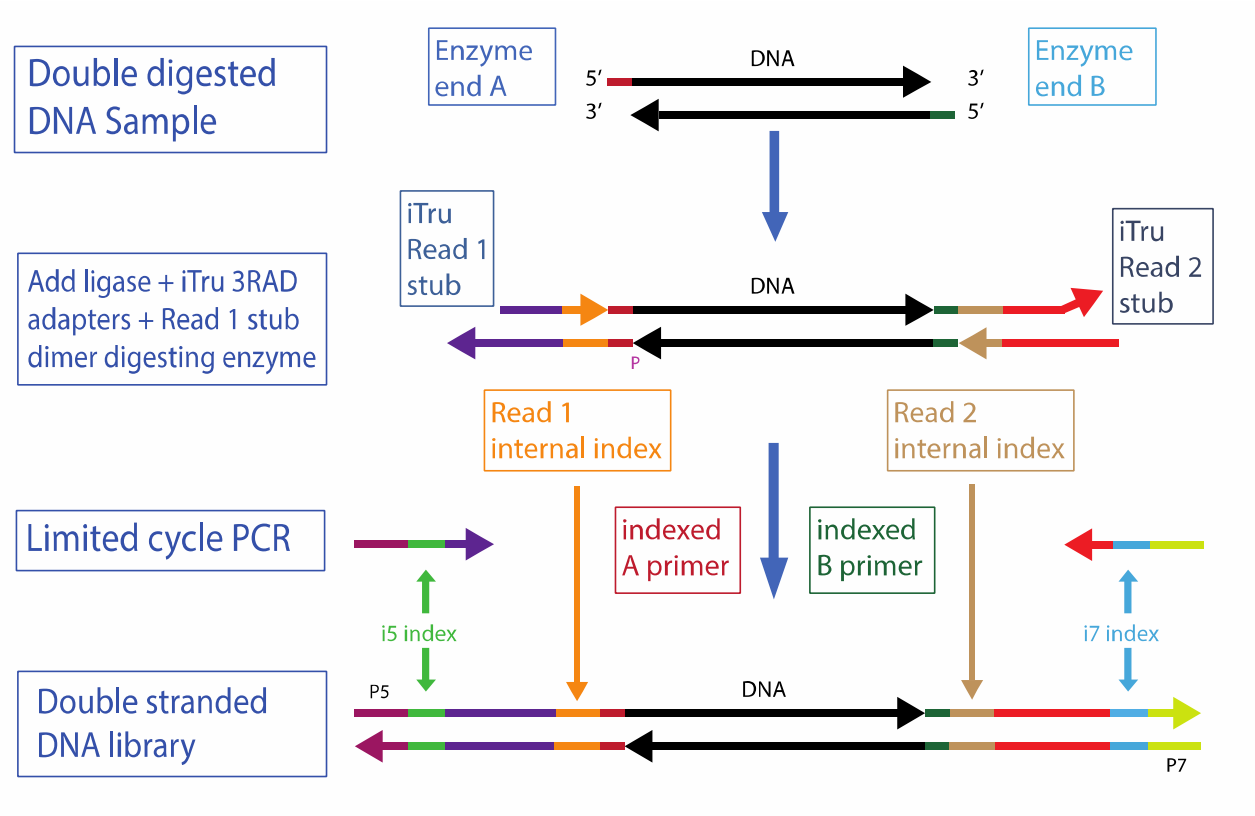

Figure 2. Specific adapter sequences and products created during the ligation of 3RAD libraries. The full adapter sequences for NheI and EcoRI-HF (Table 1) are given in the top center boxes. The relevant recognition sequences for the three restriction endonucleases are given in the top outer boxes. The products that are formed from ligation of the triple-enzyme digests and adapters are shown at the bottom.

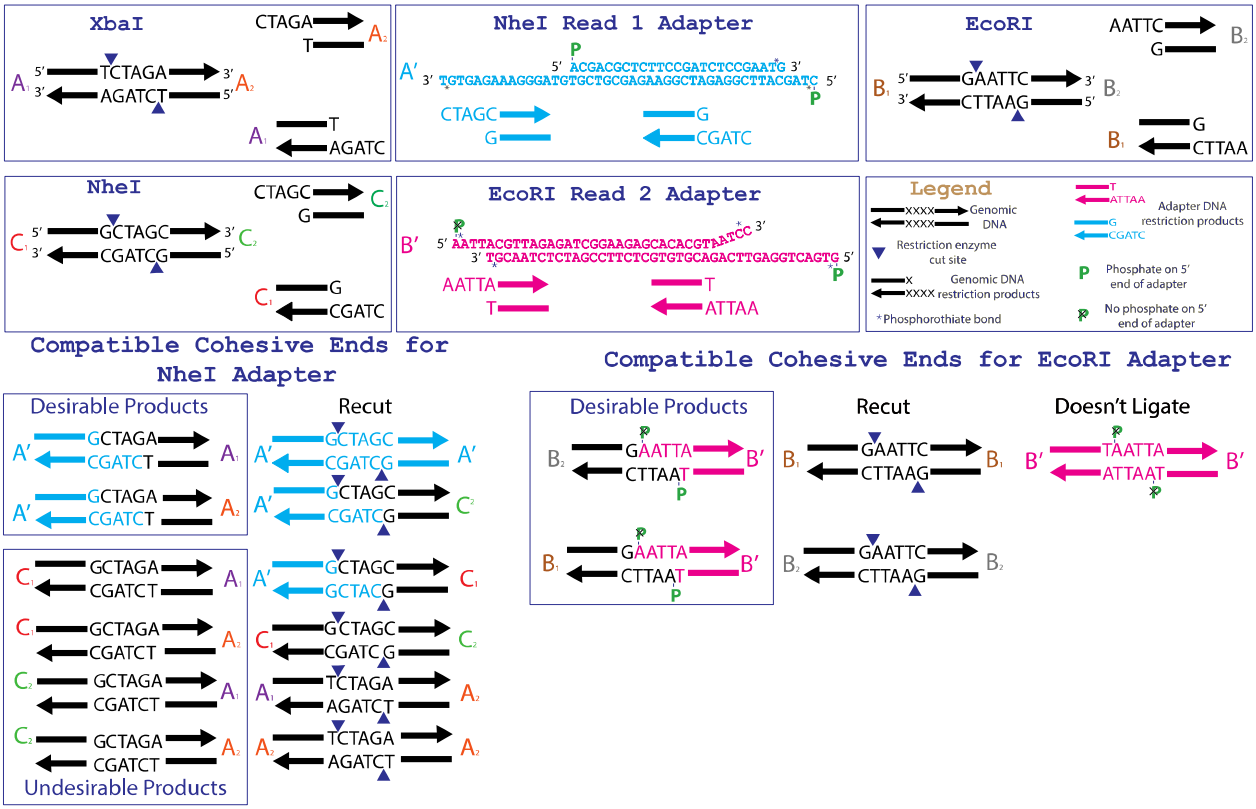

Figure 3. 3RAD workflow for samples with unique or repeated adapter tags. DNA is normalized, digested with REs, and adapters are ligated onto library molecules. If tags within adapters uniquely identify all samples (right), samples can be pooled before clean-up and PCR. If tags do not uniquely identify individuals, PCR must be done separately on each sample, and samples must be normalized and cleaned before pooling. After cleaning, samples must be size-selected and quantified to determine if a recovery P5/P7 PCR should be performed before sequencing.

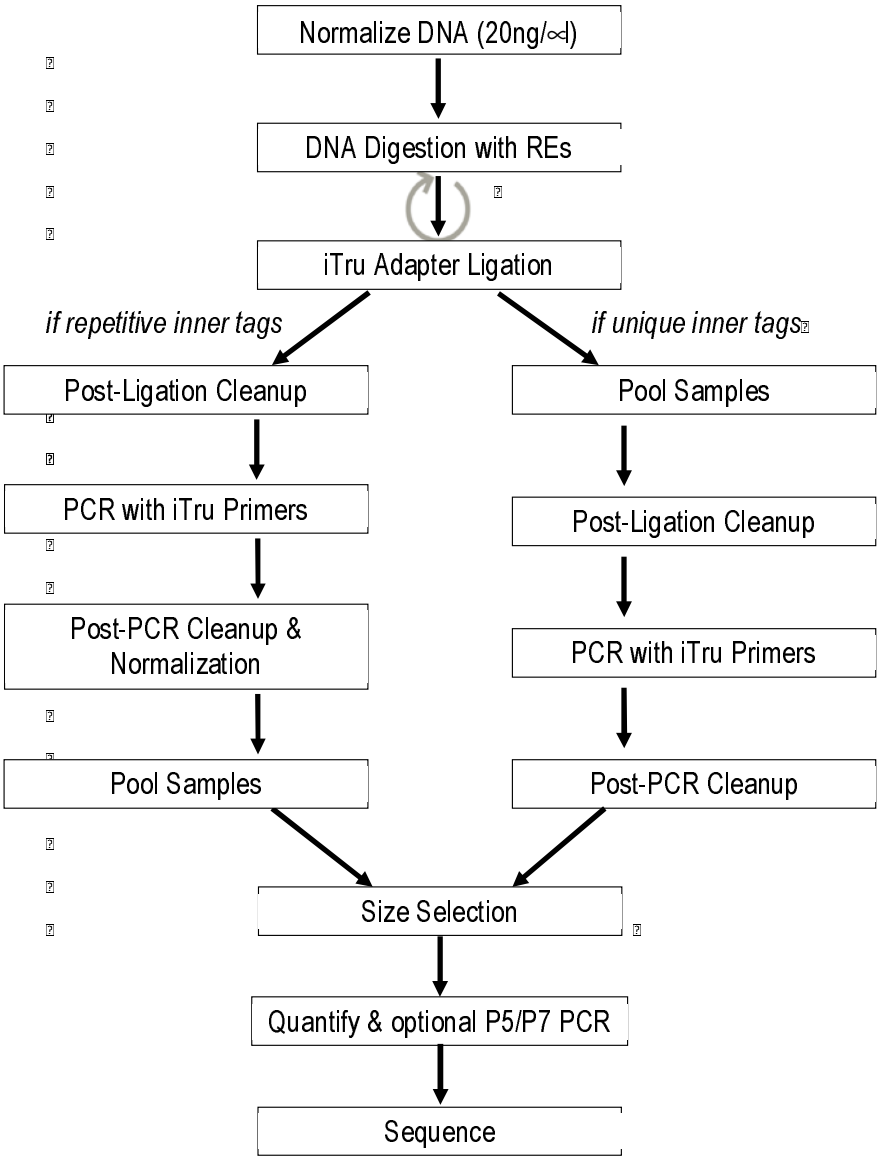

Figure 4. Sequencing reads that can be obtained from full length 3RAD library molecules. The top double stranded molecule shows a 3RAD library molecule prepared as described in the text (Appendix S1). The horizontal arrows beneath the library molecule indicate Illumina sequencing primers (binding to 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 3RAD library molecule.

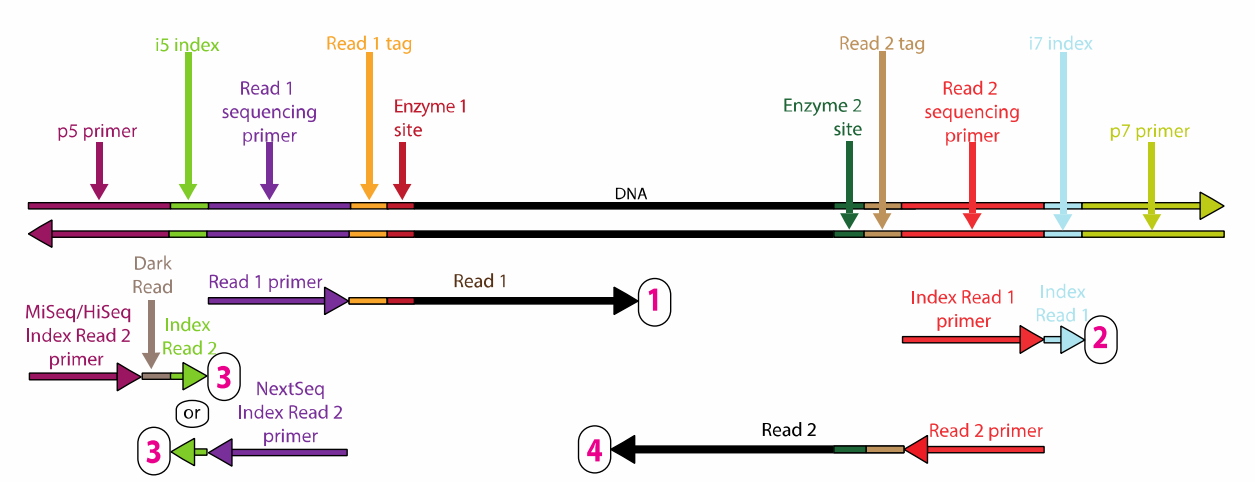

Figure 5. Scatterplot of the average coverage of all loci (polymorphic and fixed) for each sample relative to sequencing depth of each sample. Salamanders have the largest genome size, and therefore the lowest average coverage per locus when approximately 1,000,000 reads are obtained. The average coverage for all loci increases as the genome size decreases.

Appendix S1. High-Throughput 3RAD Protocol. Step by step library construction for 3RAD libraries considering both, samples pooled after ligation because all adapter combinations are used and, samples that are pooled after PCR.

Appendix S2. 3RAD video presentation, what is happening inside the tube? This presentation demonstrates the key features of 3RAD, including how the adapters perform during ligation and PCR, how the adapters and primers can be used combinatorically, and both the desired and undesired ligation products that can be formed during 3RAD. This presentation is also available at: https://www.youtube.com/watch?v=ZOmwOtfP3N4

Appendix S3. 3RAD iTru adapter designs, indexed iTru primers, and detailed library prep costs per sample. Excel workbook containing sheets for each of the four pairs of adapter designs. The designs are paired within the workbook to save space, but each design is independent, thus design 1 Read 1 adapters can be used with design 2 Read 2 adapters. Each adapter requires two oligonucleotides to form a partially complementary pair. Also, we present an example of the iTru primers with incorporated indexes to use in 3RAD libraries. And finally, tab with costs per sample for library prep, considering each step in the process.

Appendix S4. How to handle plates with 3RAD adapter aliquots. Document with instructions of how to reconstitute and anneal 3RAD dried adapters to further use in 3RAD libraries.

Appendix S5. 3RAD Libraries with Molecular ID Tags Protocol. Step by step protocol with instructions to build 3RAD libraries with molecular ID tags to detect PCR duplicates, using iTru5 8N primer (See Hoffberg *et al*. 2016).

Appendix S6. Supplementary methods and results for 3RAD example datasets. A detailed guide through the methods, results, and discussion of sequence analyses from 3RAD data generated for each example dataset presented in this manuscript.

724