MOLECULAR ECOLOGY
RESOURCES
Molecular Ecology Resources (2017) **17**, 418–430 doi: 10.1111/1755-0998.12575

Comparative performance of double-digest RAD sequencing across divergent arachnid lineages

MERCEDES BURNS,* JAMES STARRETT,* SHAHAN DERKARABETIAN,*† CASEY H. RICHART,*† ALLAN CABRERO* and MARSHAL HEDIN*

Department of Biology, San Diego State University, 5500 Campanile Drive, San Diego CA 92182, USA, †Department of Biology, University of California, 900 University Avenue, Riverside CA 92521, USA

Abstract

Next-generation sequencing technologies now allow researchers of non-model systems to perform genome-based studies without the requirement of a (often unavailable) closely related genomic reference. We evaluated the role of restriction endonuclease (RE) selection in double-digest restriction-site-associated DNA sequencing (ddRADseq) by generating reduced representation genome-wide data using four different RE combinations. Our expectation was that RE selections targeting longer, more complex restriction sites would recover fewer loci than RE with shorter, less complex sites. We sequenced a diverse sample of non-model arachnids, including five congeneric pairs of harvestmen (Opiliones) and four pairs of spiders (Araneae). Sample pairs consisted of either conspecifics or closely related congeneric taxa, and in total 26 sample pair analyses were tested. Sequence demultiplexing, read clustering and variant calling were performed in the *pyRAD* program. The 6-base pair cutter *EcoRI* combined with methylated site-specific 4-base pair cutter *MspI* produced, on average, the greatest numbers of intra-individual loci and shared loci per sample pair. As expected, the number of shared loci recovered for a sample pair covaried with the degree of genetic divergence, estimated with cytochrome oxidase I sequences, although this relationship was non-linear. Our comparative results will prove useful in guiding protocol selection for ddRADseq experiments on many arachnid taxa where reference genomes, even from closely related species, are unavailable.

Keywords: Araneomorphae, ddRADseq, genomics, Mygalomorphae, non-model organism, Opiliones

Received 19 November 2015; accepted 23 June 2016

Introduction

New high-throughput molecular sequencing methods continue to revolutionize biological avenues to which they are applied, including diverse fields such as medicine, conservation and ecology (Davey *et al.* 2011; Niedringhaus *et al.* 2011; Georgiou *et al.* 2014; Hoffman *et al.* 2014; Barley *et al.* 2015; Gallego *et al.* 2015; Hess *et al.* 2015; Rius *et al.* 2015). Until recently, many of these applications relied upon reference genomes, which are currently available for a limited number of model organisms, although new genome sequences for microbes, plants and animals are accumulating rapidly (Pruitt *et al.* 2007; Dohm *et al.* 2014; Nossa *et al.* 2014; Worley 2014; Wu *et al.* 2014; Varghese *et al.* 2015). This deficit directed many workers in the fields of evolutionary biology and ecology to seek alternative strategies for acquiring variant data (typically represented in single nucleotide polymorphisms, or SNPs) representative of the whole genome, as the time and cost required to generate a fully sequenced genome remains relatively expensive (Everett *et al.* 2011).

The adaptation of restriction endonucleases (REs) to the acquisition of genomic sequence data has allowed for unprecedented access to genome-wide variation for researchers working with non-model organisms (e.g. Hoffman *et al.* 2014; Barley *et al.* 2015). One such method is restriction-site-associated DNA sequencing (RADseq), which utilizes enzymes with varied sequence specificities to generate a tractable representation of the genome (Baird *et al.* 2008; Andrews *et al.* 2016). Peterson *et al.* (2012) first introduced the double-digest version of this method, utilizing two REs with different site specificities to digest the genome, allowing for the selection of fragments with known 5′ and 3′ ends to be produced. Digestion with enzymes rather than random shearing, coupled with precise size selection, reduces overall genome coverage, but ensures fragment libraries with high repeatability and site-specific coverage (Peterson *et al.* 2012).

Double-digest RADseq (ddRADseq) has been adopted rapidly by workers to address evolutionary and

Correspondence: Mercedes Burns, Fax: 619-594-5676;
E-mail: mercedes.burns@gmail.com

ecological questions at the landscape, population and systematic levels (Jones *et al.* 2013; Leaché *et al.* 2014, 2015a; Zhou *et al.* 2014; Blair *et al.* 2015; Jezkova *et al.* 2015; Mason & Taylor 2015; Meik *et al.* 2015; Recknagel *et al.* 2015; Rittmeyer & Austin 2015; Schield *et al.* 2015a). While this sequencing methodology does not require the development of probes, primers or sequence scaffolds from a reference genome, the lack of a closely related reference limits the inference that can be made regarding the success of ddRADseq projects. Without a closely related reference to, for example, estimate the efficacy of digestion *in silico* (Lepais & Weir 2014), the expected performance of a particular set of enzymes cannot be determined *a priori* (Davey *et al.* 2011). In such cases, it is advisable to perform a pilot study to assess the ability of enzyme combinations to maximize sequence output (and thus, hopefully, genome coverage) and the discovery of genetic variants.

Arachnida is one of many branches in the tree of life that continues to benefit from the advent of high-throughput sequencing technologies (Brewer *et al.* 2014; Ellegren 2014). Transcriptomics and pyrosequencing approaches have been adapted to the study of venoms and silks (Clarke *et al.* 2014, 2015; Haney *et al.* 2014; Sanggaard *et al.* 2014), phylogenomics (Hedin *et al.* 2012a; Sharma *et al.* 2014; Fernández & Giribet 2015; Garrison *et al.* 2015), population biology (Mattila *et al.* 2012; Planas *et al.* 2014) and parasite–vector interactions (Schwarz *et al.* 2013). In spite of these unique applications and diverse systems, few complete chelicerate genomes are available (Grbić *et al.* 2011; Cao *et al.* 2013; Ellegren 2014). For example, only two complete spider genomes have been sequenced, displaying incongruences in genome size and content as to suggest significant variability in the rest of the order (Sanggaard *et al.* 2014). Genomes for other species-rich orders, such as Opiliones (harvestmen), remain unavailable. To date, next-generation sequencing approaches have been leveraged for few population genomics or phylogeographic studies in arachnids (e.g. Hamilton *et al.* 2016), and more of these efforts are anticipated in the near future. As the cost of genomic sequencing continues to drop, we expect the number of sequenced genomes to increase. Until adequate genomic resources are available for all arachnid orders, adaptable and cost-effective reduced representation sequencing approaches scalable to the sample sizes necessary for successful population-level studies are required. RADseq and related strategies potentially fill this need for arachnologists and researchers of other non-model organisms.

For this study, we leveraged the flexibility of the ddRADseq preparation protocol to contrast the quality and volume of short-read data produced with four different RE combinations. Nine sample pairs of arachnids, covering two orders, nine genera and a range of conspecific and congeneric divergences were selected for sequencing and comparison. This sampling structure was intended to elucidate RE combinations likely to be broadly useful in different arachnid lineages, as well as varying levels of evolutionary divergence. Individual and shared read statistics were taken from each experimental condition, including overall loci recovered and an estimate of sequencing efficiency – the proportion of loci recovered to the number of raw sequenced reads. Although we did not expect one RE combination to fit all situations, we predicted that RE combinations targeting longer restriction sites would produce fewer loci within sample pairs, as these sites would be more likely to undergo mutation (Andrews *et al.* 2016), and predicted that sites with fewer repetitive bases would be more likely to be conserved between more distantly related taxa.

Overall, RE combination did not have a pronounced effect on the conversion ratio of raw reads to loci within sequenced individuals. However, within sample pairs, RE selection had a strong effect on the number of loci shared, particularly for sample pairs with greater expected genomic divergence. Specifically, RE combinations including the enzymes *EcoRI* and *MluCI* produced data matrices with the most shared loci and SNPs between sample pairs. We explore possible explanations for this pattern and offer suggestions for future work in arachnids and other non-model taxa.

Methods

Sample preparation

We selected nine total sample pairs (Table 1), including geographically distant conspecific populations of two harvestman suborders (Eupnoi: *Leiobunum manubriatum*; Laniatores: *Speleonychia sengeri*, *Theromaster brunneus*) and two primary spider suborders (Araneomorphae: *Habronattus tarsalis*, *Hypochilus pococki*; Mygalomorphae: *Antrodiaetus riversi*, *Calisoga longitarsis*), as well as congeneric samples from two harvestman suborders (Dyspnoi: *Acuclavella quattuor* and *A. shoshone;* Laniatores: *Sclerobunus nondimorphicus* and *S. idahoensis*). Sample pairs were selected to define best practices for ddRADseq protocols for a wide variety of downstream applications (landscape genetics, population genetics, phylogeography, species delimitation). See Table S1 (Supporting information) for detailed locality and voucher information. Samples were preserved in 100% ethanol and stored at −80°C until extraction commenced. Genomic DNA was extracted from coxal muscle and leg tissue, or whole bodies (*Sclerobunus* spp., *Speleonychia, Theromaster*) using the DNeasy Blood and Tissue Kit

Table 1 Analysis results for all sample pairs by RE combination

Voucher name	Species	# Raw reads	# Reads passed	% Passed	# Loci within sample	Estimated heterozygosity	Error rate	# Polymorphisms	Frequency of polymorphisms	Loci shared	Total # of loci with variable sites	Sampled unlinked SNPs
		SbfI-MspI	**Protocol adjustment: 10 Units *SbfI*, 100 Units *MspI***									
OP1650	*Sclerobunus idahoensis*	156 163	130 194	0.833 706	8427	0.006288	0.001014	521	0.003447	464	709	327
OP3784	*Sclerobunus nondimorphicus*	259 524	215 248	0.829395	46956	0.007434	0.00143	699	0.003132			
OP1851	*Leiobunum manubriatum*	136 208	81 120	0.59556	7430	0.008938	0.001313	286	0.004905	127	225	95
OP1856	*Leiobunum manubriatum*	209 786	128 518	0.612615	11077	0.00547	0.000411	169	0.002931			
OP1618	*Theromaster brunneus*	1 034 745	602 730	0.582491	21365	0.003092	0.000237	550	0.001644	361	736	309
OP1623	*Theromaster brunneus*	373 358	220 796	0.591379	16263	0.003422	0.000258	290	0.001703			
HA1055	*Habronattus tarsalis*	83 787	70 882	0.845978	8693	0.005203	0.002118	434	0.003751	540	905	408
HA1070	*Habronattus tarsalis*	87 834	76 960	0.876198	6397	0.005642	0.001305	495	0.004101			
H508	*Hypochilus pococki*	311 119	269 275	0.865505	12557	0.004095	0.000162	424	0.003734	579	926	418
H595	*Hypochilus pococki*	241 796	210 977	0.872541	22519	0.002465	0.000157	234	0.002217			
AR17	*Antrodiaetus riversi*	264 039	230 437	0.872738	11878	0.005155	0.000308	322	0.002939	800	954	499
AR47	*Antrodiaetus riversi*	365 167	320 988	0.879017	15312	0.006398	0.000324	555	0.00437			
MY4085	*Calisoga longitarsis*	849 405	727 370	0.856329	37333	0.006443	0.001	2572	0.004416	3028	4163	2106
MY4416	*Calisoga longitarsis*	458 939	385 442	0.839855	30954	0.004084	0.001522	1154	0.0024405			
		EcoRI-MspI	**Protocol adjustment: 100 Units *EcoRI*, 100 Units *MspI***									
OP1650	*Sclerobunus idahoensis*	3 396 401	2 988 344	0.879856	143 986	0.008113	0.001436	11 846	0.004025	7347	12 702	5656
OP3784	*Sclerobunus nondimorphicus*	2 250 608	1 970 910	0.875723	138 873	0.005967	0.001259	6281	0.002772			
MMB30	*Leiobunum manubriatum*	3 484 266	2 987 301	0.857369	136 153	0.008594	0.001217	11 959	0.004049	6522	12 617	5073
MMB89	*Leiobunum manubriatum*	4 065 335	3 533 528	0.869185	136 184	0.008384	0.001108	13 713	0.004283			
OP1609		1 099 288	955 127	0.86886	115 011	0.006258	0.001956	7581	0.004084	756	1987	651

Table 1 (Continued)

Voucher name	Species	# Raw reads	# Reads passed	% Passed	# Loci within sample	Estimated heterozygosity	Error rate	# Polymorphisms	Frequency of polymorphisms	Loci shared	Total # of loci with variable sites	Sampled unlinked SNPs
	Theromaster brunneus											
OP1613	*Theromaster brunneus*	976 469	831 567	0.851606	109 932	0.003968	0.002174	2885	0.001795			
OP1683	*Speleonychia sengeri*	4 679 179	4 212 322	0.900227	240 337	0.001984	0.000461	2719	0.000624	45 699	2379	1859
OP1732	*Speleonychia sengeri*	4 342 125	3 909 384	0.900339	227 289	0.001485	0.000468	2957	0.000676			
OP2270	*Acuclavella quattuor*	2 903 412	2 682 337	0.923857	123 175	0.006487	0.00019	4317	0.00417	3849	7060	3028
OP2335	*Acuclavella shoshone*	2 461 053	2 278 677	0.925895	77 055	0.004806	0.000128	3147	0.003297			
HA1055	*Habronattus tarsalis*	969 465	880 218	0.907942	276 014	0.014443	0.00568	5608	0.009117	1429	2116	898
HA1070	*Habronattus tarsalis*	776 173	702 884	0.905576	249 976	0.016503	0.006149	3741	0.010171			
H508	*Hypochilus pococki*	1 007 766	818 385	0.812078	141 908	0.006228	0.000929	10 328	0.005162	9307	16 520	7140
H595	*Hypochilus pococki*	753 477	643 190	0.853629	99 621	0.003686	0.000673	5298	0.002935			
AR17	*Antrodiaetus riversi*	804 285	720 081	0.895306	92 168	0.004877	0.001084	6468	0.003577	11 611	14 747	7842
AR47	*Antrodiaetus riversi*	787 075	709 303	0.901189	90 938	0.006613	0.001045	8735	0.005072			
MY4085	*Calisoga longitarsis*	1 426 615	1 240 406	0.869475	301 835	0.012859	0.00322	11 632	0.007553	5082	7686	3625
MY4416	*Calisoga longitarsis*	1 527 772	1 330 367	0.870789	296 907	0.009393	0.002901	8486	0.00483			
		SphI-MspI	**Protocol adjustment: 10 Units *SphI*, 100 Units *MspI***									
MMB30	*Leiobunum manubriatum*	913 619	829 440	0.907862	39 929	0.009772	0.001521	3743	0.004741	1824	3378	1419
MMB89	*Leiobunum manubriatum*	1 201 799	1 090 338	0.907255	48 767	0.008127	0.00106	3340	0.00397			
OP1683	*Speleonychia sengeri*	3 375 049	3 068 684	0.909227	96 154	0.001523	7.66E−05	1296	0.000591	22 744	2379	861
OP1732		4 484 443	4 023 252	0.897158	119 347	0.001925	6.28E−05	1792	0.000718			

Table 1 (Continued)

Voucher name	Species	# Raw reads	# Reads passed	% Passed	# Loci within sample	Estimated heterozygosity	Error rate	# Polymorphisms	Frequency of polymorphisms	Loci shared	Total # of loci with variable sites	Sampled unlinked SNPs
	Speleonychia sengeri											
OP2270	*Acuclavella quattuor*	1 699 236	1 543 153	0.908145	93 077	0.005706	0.000197	1743	0.003791	1551	2317	1078
OP2335	*Acuclavella shoshone*	1 750 166	1 612 604	0.921401	84 668	0.004462	0.000192	1549	0.002752			
		Sph1-MluC1	**Protocol adjustment: 10 Units *SphI*, 100 Units *MluCI***									
MMB30	*Leiobunum manubriatum*	1 098 578	1 043 523	0.949885	252 397	0.019415	0.004075	6821	0.007319	2981	5241	2258
MMB89	*Leiobunum manubriatum*	1 220 839	1 154 847	0.945945	262 401	0.014984	0.002759	8233	0.006939			
OP1683	*Speleonychia sengeri*	1 785 830	1 587 291	0.888825	273 225	0.002315	0.000455	2881	0.000986	21 829	1251	861
OP1732	*Speleonychia sengeri*	3 035 408	2 640 106	0.86977	227 609	0.002155	0.0008	2939	0.001034			
OP2270	*Acuclavella quattuor*	651 926	564 703	0.866207	110 326	0.006056	0.000296	1994	0.003941	2641	3296	1615
OP2335	*Acuclavella shoshone*	957 723	857 676	0.895537	162 811	0.005272	0.000523	2465	0.002739			
HA1055	*Habronattus tarsalis*	1 072 483	949 431	0.885264	381 251	0.013859	0.004532	4201	0.007233	1761	2905	1240
HA1070	*Habronattus tarsalis*	634 429	548 908	0.8652	234 725	0.013358	0.004928	2859	0.007884			
H508	*Hypochilus pococki*	1 747 976	1 493 248	0.854273	268 240	0.008107	0.001358	11 552	0.005619	8555	15 686	6652
H595	*Hypochilus pococki*	1 583 573	1 392 569	0.879384	243 969	0.005726	0.001076	6464	0.003498			
AR17	*Antrodiaetus riversi*	921 796	816 143	0.885384	176 584	0.006673	0.001611	3630	0.003693	5790	6986	3731
AR47	*Antrodiaetus riversi*	820 754	725 526	0.883975	182 171	0.008312	0.001769	4594	0.005156			
MY4085	*Calisoga longitarsis*	3 750 501	3 232 616	0.861916	724 583	0.01024	0.003145	28 422	0.006541	13 765	22 081	10 121
MY4416	*Calisoga longitarsis*	2 705 808	2 265 543	0.837289	604 993	0.00788	0.003047	14 615	0.004769			

All statistics from *PyRAD* analyses, including raw read and loci counts, estimated heterozygosity (H_O), number and frequency of polymorphisms (F_{poly}), number of shared loci and unlinked SNPs sampled. Two sample pairs were represented by four individuals spanning similar localities/divergences (*L. manubriatum*, *T. brunneus*) as limited by DNA extract availability.

(Qiagen). Prior to genome digestion, we quantified sample DNA using a Qubit 2.0 fluorometer (Thermo Fisher Scientific) to assure a concentration of at least 12.5 ng/μL.

Four combinations of REs, including *SbfI* (restriction site: 5′-CCTGCAGG-3′), *EcoRI* (5′-GAATTC-3′), *MspI* (5′-CCGG-3′), *SphI* (5′-GCATGC-3′) and *MluCI* (5′-AATT-3′) were used, avoiding combinations with redundant site specificities and producing a range of RE site length and complexity, from most (*SbfI–MspI*) to least (*EcoRI–MspI*) complex, with combinations *SphI–MspI* and *SphI–MluCI* approximately intermediate (Fig. 1). Each sample pair was sequenced with at least two sets of REs (see Figs 1 and 2 for enzyme pairings). Library preparation and sequencing took place in October 2014 (effort 1, testing *SbfI–MspI*), January 2015 (effort 2, testing *SbfI–MspI, EcoRI–MspI, SphI–MluCI* and *SphI–MspI*), and May 2015 (effort 3, testing *SbfI–MspI, EcoRI–MspI* and *SphI–MluCI*). To prepare sequence libraries, we followed a customized ddRADseq protocol adapted from Peterson *et al.* (2012). Genomic DNA (500 ng) from each specimen was digested with 10-100 units each of two restriction endonucleases in CutSmart buffer (New England BioLabs) for a reaction volume of 50 μL, incubated at 37 °C for 4 h. We assessed the completeness of digestion by running subsamples of each enzyme combination alongside aliquots of undigested template DNA on a 1% agarose gel stained with ethidium bromide. Enzymes, buffer and fragments ≤100 base pairs were subsequently removed using 1.5 times the reaction volume of Agencourt AMPure XP (Beckman Coulter) magnetic beads, following manufacturer's protocols. At this stage, we estimated the average concentration of digests for each enzyme protocol and diluted samples to within one standard deviation of the mean.

Eight custom adaptors prepared for each enzyme combination and with barcode sequences designed for each row of a 96-well plate were ligated to genomic fragments using 100 units of T4 Ligase. Samples were incubated at room temperature (23 °C) for 40 min, heat killed at 65 °C for 10 min and cooled by 2 °C per 90 s for 22 cycles. Samples were pooled by column and purified prior to fragment size-selection. We used a Pippin Prep (Sage Sciences) automated size-selection instrument to isolate fragments in a size range of 415–515 bp for sequencing effort 1, and a wider frame of 400–600 bp during sequencing efforts 2 and 3. Pooled fragments were amplified (98 °C for 30 s, 12 cycles of 98 °C for 10 s to 72 °C for 20 s, and a final elongation of 10 min at 72 °C) using the Phusion PCR kit (New England BioLabs) and standard Illumina primers. After purification, sample molarity was determined using an Agilent Bioanalyzer 2100 (Agilent Technologies) and an equimolar pooled sample was prepared. Resulting libraries were sequenced with an Illumina HiSeq 2500 under the 100-bp single-end protocol at the University of California, Riverside, IIGB Genomics Core facility.

Data analysis

Barcode demultiplexing, quality control, within-sample clustering and between-pair variant calling were carried out using *pyRAD* v. 3.0.5 (Eaton 2014). We analysed pairs separately for each RE combination. Only reads with unambiguous barcodes and phred scores ≥20 were retained, and loci with more than one undetermined base were additionally discarded. We set a within-sample and between-sample read clustering threshold at 95% similarity to reduce clustering of paralogs and reads with undetermined sites (Andrews *et al.* 2016). To avoid

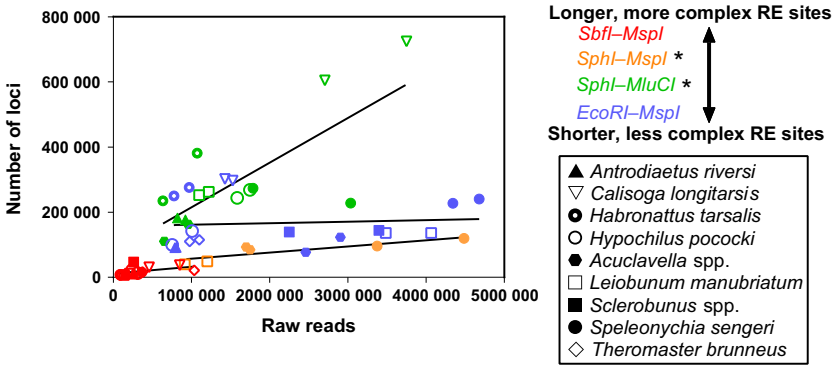

Fig. 1 Regression plot for conversion of raw reads to loci in ddRADseq protocols using four RE combinations. Least-squares regression of loci recovered as a function of raw reads sequenced for all samples. RE combinations, identified by colour, are ordered by relative length and complexity of restriction sites (note that REs *SphI–MspI* and *SphI–MluCI* have approximately equivalent site length and complexity). Bold lines indicate line of best fit; asterisk (*) following enzyme protocol indicates a significantly non-zero slope. Species are identified by symbol shape.

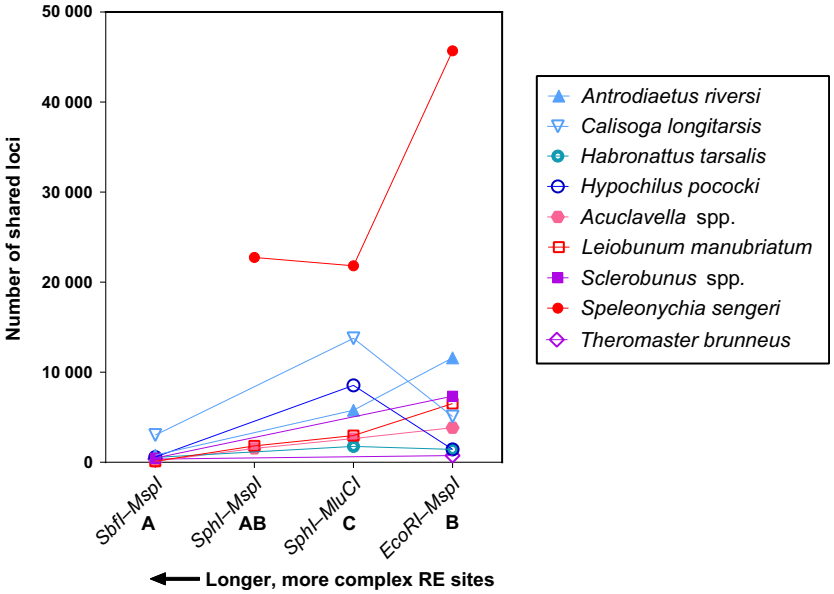

Fig. 2 Comparison of shared loci recovered given RE selection. The number of shared loci for each sample pair is given based on recovery under each of four RE combinations. Spider sample pairs are colour-coded in blue shades; harvestman sample pairs are coloured in red shades. Letters indicate significantly different mean results from ANOVA analysis with multiple comparisons and Bonferroni post hoc correction. [Colour figure can be viewed at wileyonlinelibrary.com]

over-inflation of estimated heterozygosity, we required a minimum of 10 reads for each cluster during consensus base-calling. Up to three shared polymorphic sites per called locus were allowed. The preceding settings constitute conservative standards for variant calling (Eaton 2014; Harvey *et al.* 2015) which were designed to recover strongly supported loci within individuals while allowing for divergence within a sample pair. Each analysis was run on the same computer (Early 2015 MacBook Pro, 2.7 GHz Intel Core i5 with 8 GB RAM) to determine computational duration in real time. For each RE combination, we compiled the number of high-quality reads and loci, estimated heterozygosity, error rate and the number and frequency of polymorphic sites for each sequenced sample, as well as the number of shared loci, variable sites and unlinked SNP counts for each sample pair. To avoid biasing protocol efficacy towards maximum read count, we additionally calculated the ratio of recovered loci to average number of raw reads per sample pair.

We estimated evolutionary divergence between sample pairs using COI mitochondrial sequence data from either the exact individuals sequenced in this study or conspecific specimens from the same collecting localities. In most cases, these data had been previously published – *Sclerobunus* spp. (Derkarabetian *et al.* 2011), *Speleonychia* (SDSU_OP1683 - Derkarabetian *et al.* 2010; Hedin & Thomas 2010), *Antrodiaetus* (Hedin *et al.* 2013), *Calisoga* (Leavitt *et al.* 2015) and *Hypochilus* (Keith & Hedin 2012). COI sequence data were compiled and imported into Geneious Pro v. 8.1.7 (http://www.geneious.com, Kearse *et al.* 2012), where they were aligned using MUSCLE (Edgar 2004) and uncorrected *p*-distances were calculated for each sample pair. For the *Sclerobunus nondimorphicus* and *S. idahoensis* pair, the mean *p*-distance was calculated from pairwise alignments of three divergent samples of each species collected from across the known range (Derkarabetian *et al.* 2011). For the *Acuclavella* spp. pair, the mean *p*-distance was estimated from pairwise alignments of a specimen of the *A. quattuor* locality to another *Acuclavella* sequence that spans the same root node as the pair analysed here (Richart & Hedin 2013).

Results

Quality filtering

For the 22 individuals sequenced in three separate efforts across 2–4 RE profiles, an average of 1 402 832 (SD ± 1 279 033) reads were obtained after demultiplexing (Table 1). Average read count differed across the four protocols, with the *SbfI–MspI* combination yielding the lowest number of raw reads per individual (Fig. 1; mean 345 134 reads). The *SphI–MspI* combination produced the greatest amount of sequence data, with an average of 2 237 385 reads across three sample pairs (*Acuclavella* spp., *Leiobunum* and *Speleonychia*). Enzyme selection, by contrast, had minimal influence on read pass rate; on average, 86% of raw reads passed our quality filter (SD ± 8.48%) and pass rates arranged by RE combination all fell within one standard deviation of the mean. There was an effect of sequencing effort for the *SbfI–MspI* combination (Mann–Whitney $U = 3$; $P < 0.001$); the average pass rate for sequences generated by this enzyme pair was only 60% for the first sequencing effort in October 2014, but rose to 83% for sequences generated in January 2015 and 86% in May 2015.

Analysis of RE combinations

A total of 26 sample pair analyses using four different enzyme combinations (*SbfI–MspI:* N = 7, *EcoRI–MspI:* N = 9, *SphI–MspI:* N = 3, *SphI–MluCI:* N = 7) were completed in *pyRAD*. From sample quality filtering to output of SNP data, analysis time ranged from just over six minutes for *Leiobunum* under the *SbfI–MspI* protocol (*SbfI–MspI* protocol mean = 0:18:52) to over five hours for *Calisoga* with *SphI–MluCI* (*SphI–MluCI* protocol mean = 2:22:32) (Fig. S1, Supporting information). The average computational time across all analyses was 1:21:15. The *SbfI–MspI* combination also yielded the smallest number of loci (mean = 18 369), shared loci (mean = 843) and unlinked SNPs (mean = 595) between sample pairs as compared to the other RE combinations (Figs 1and 2; Table 1). *EcoRI–MspI* produced the greatest number of shared loci (mean = 8646) and SNPs (mean = 3975) (Tables 1 and S1, Supporting information).

We examined whether RE pairs that target shorter/lower complexity restriction sites (*EcoRI–MspI* and *SphI–MluCI*), and theoretically produce larger numbers of raw reads, also tend to have the most loci (thus providing the most useful downstream data). Numbers of loci recovered were plotted against reads sequenced for each individual under all RE conditions to examine the effect of RE combination on locus identification (Fig. 1). Of the four RE combinations surveyed, two had significantly positive, non-zero slopes (Fig. 1; *SphI–MspI*: R^2= 0.75, m = 0.0189, $P < 0.05$; *SphI–MluCI*: R^2= 0.59, m = 0.137, $P < 0.01$), indicating an increase in the number of loci recovered per individual sequenced as raw read count increased. Slopes from enzyme combinations *SbfI–MspI* and *EcoRI–MspI* were not significantly different than zero, demonstrating that loci count was largely invariant with raw read count. In comparison with mean loci number between RE combination using ANOVA with multiple comparisons and a Bonferroni post hoc test correction, we found a significant effect of RE ($F_{3,48}$= 18.55, $P < 0.0001$) with *SphI–MluCI* showing significantly greater locus recovery over all the RE combinations tested (Fig. 2, Table S2, Supporting information). At the level of sample pairs, two taxa, *Habronattus* and *Calisoga*, displayed greater numbers of loci per reads sequenced than other samples for RE combinations *SphI–MluCI* and *EcoRI–MspI* (Fig. 1).

Mitochondrial COI divergence

We estimated COI divergence for seven sample pairs, including *Leiobunum*, *Speleonychia*, *Acuclavella* spp., *Calisoga*, *Antrodiaetus*, *Hypochilus* and *Sclerobunus* spp. COI sequence divergence ranged from nearly invariant (0.1%, *Speleonychia*) to high intraspecific variability (15.2%, *Hypochilus*). COI sequences for one *Speleonychia* (acquired using methods from Derkarabetian *et al.* 2010) and two *Leiobunum* (acquired using methods from Hedin *et al.* 2012b) have been newly submitted to GenBank (voucher # SDSU_OP1732, accession # KX550439; voucher # MMB-30: accession # KX570871, voucher # MMB-89: accession # KX570872).

Visualization of mean shared loci count by COI sequence divergence indicates a roughly non-linear decay relationship (Fig. 3; AICc = 346.2 as compared to AICc = 366.7 for linear model through origin; Speiss & Neumeyer 2010) where a precipitous drop in shared loci was seen as COI sequences differentiate. This effect is largely due to *Speleonychia,* which averages more than double the amount of shared loci of all other sample pairs analysed (Fig. 2; mean = 30 090), although the *Hypochilus* specimens sequenced had the second highest mean (11 044) despite having highly divergent COI sequences. The remaining pairs examined had between 3000 and 7500 loci in common.

Discussion

Restriction endonuclease selection matters

Analysis of statistics from RE selection experiments indicates pronounced differences in both the number of loci

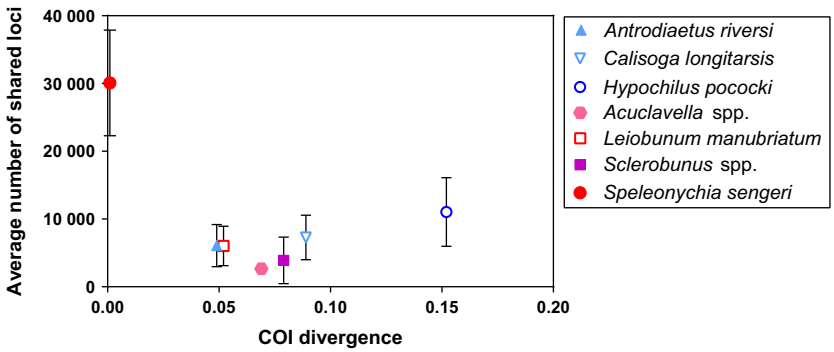

Fig. 3 Mean shared loci recovered for sample pairs as a function of COI divergence. For all sample pairs with available COI sequence data, *y*-values indicate the mean number of shared loci (+ standard error) across all RE combinations with which the sample pair was sequenced. Spider sample pairs are colour-coded in blue shades; harvestman sample pairs are coloured in red shades. [Colour figure can be viewed at wileyonlinelibrary.com]

produced per reads sequenced and the number of loci individuals shared with a congener or conspecific. On average, the *EcoRI–MspI* combination produced the greatest number of shared loci and unlinked SNPs per sample pair (Table 1). The expectation that RE combinations with shorter, less complex restriction sites, such as *SphI–MluCI* and *EcoRI–MspI*, would produce more raw reads and account for more loci was met (Fig. 1) – these RE combinations were also responsible for producing the greatest numbers of shared loci between sample pair investigated (Fig. 2). We did observe a significant effect of enzyme selection on the conversion of raw reads to within-individual loci in two enzyme pairs (Fig. 1), suggesting a lack of genome coverage saturation, as we would expect the number of loci recovered to eventually stabilize with increasing sequence coverage such that all loci predicted are successfully identified (DaCosta & Sorenson 2014). However, RE combinations with non-zero read-to-loci slopes include *SphI–MspI*, which was only used for six specimens (Table 1; three pairs) and thus may be under-sampled, and *SphI–MluCI*, which has points with evidence of outlier behaviour (ESD method: specimen MY4085, $Z = 2.52$, $P < 0.05$). *Calisoga* specimens displayed extremely large numbers of loci recovered per reads sequenced for *SphI–MluCI* (Fig. 1), such that removal of these specimens also eliminates the significantly positive correlation between raw reads and loci recovered. For the *SphI–MluCI* and *EcoRI–MspI* combinations, we observed that both *Calisoga* and *Habronattus* specimens had larger numbers of loci given sequencing effort (Fig. 1). The reasons for this are unclear but could be attributed to these species having larger genomes than the others we investigated (in support of this hypothesis, estimates for *Habronattus* suggest the genome size is over twice that of the spider average; see Gregory & Shorthouse 2003), greater error associated with locus discovery, or more AATT nucleotide motifs.

Increased locus recovery with the *SphI–MluCI* and *EcoRI–MspI* combinations was consistent among all sample pairs investigated (Fig. 2), leading us to postulate that the effect of enzyme combination on ddRADseq results may be rooted in the evolutionary history of the species we examined. Our results indicate that the enzymes *EcoRI* and *MluCI*, which both recognize some subset of a AATT nucleotide motif and leave a 4-base 5′ overhang, had the greatest impact on sequencing statistics (Table 1; Fig. 1). The RE combinations *EcoRI–MspI* and *SphI–MluCI* produced the highest conversion of raw reads to loci for the majority of sample pairs that we examined (Fig. 1). These results are unlikely to be due to star activity (i.e. non-specific RE cleavage); although *EcoRI* can exhibit non-site specific cutting, we utilized a high-fidelity enzyme with buffer designed to inhibit this action (New England BioLabs). The utility of enzymes with AT-rich restriction sites may potentially owe their effectiveness to biased nucleotide composition in Arachnida. Previous studies attempting to improve analytical methods for arachnid phylogenetics have described shifts in nucleotide composition and asymmetrical exchanges in favour of amino acids with AT-biased degeneracy, such as isoleucine. Such AT bias appears frequently in arthropods (Foster & Hickey 1999; Brewer *et al.* 2014; Sanggaard *et al.* 2014) and is often heralded by a decrease in genetic recombination and concomitant lack of gene conversion, the process of which may be GC-dependent (Stensrud *et al.* 2007). We are unaware of studies implicating specific decreases in gene recombination for arachnids, although they are well defined in some model arthropods (Kliman & Hey 1993; Comeron *et al.* 2012).

We found variation in conversion of raw reads to loci between protocols for samples that were sequenced with multiple RE combinations (Figs 1 and 2; Table 1). As genomic DNA from the same individuals was used in each protocol comparison, we might expect conversion to be similar across protocol unless there are differences in restriction site mutation rate (Andrews *et al.* 2016). One enzyme used frequently in ddRADseq studies and in several of our RE combinations, *MspI*, is unable to cleave sites with methylation at the 5′ cytosine (Schield *et al.* 2015b). Although cytosine methylation is known to be limited in arthropods (Regev *et al.* 1998; Lyko & Maleszka 2011; Lechner *et al.* 2013), virtually no literature on DNA methylation has included arachnid taxa. It is possible that, similar to mutation of the restriction site, site methylation operates as an epigenetic contribution to allelic dropout (Kerkel *et al.* 2008). However, the restriction site of *MspI* is both short and repetitive, perhaps ameliorating any effect of methylation in the sheer number of genomic sites with which it should be able to interact.

The results of our investigation into the effects of RE selection on read count and shared loci indicate an effect of sequencing date. We saw a 30% increase in quality control pass rate for fragment libraries prepared with *SbfI–MspI* in January 2015 as compared to those from October 2014. This improvement may be related to changes we made to the fragment size-selection window, which we widened by 100 base pairs in later sequencing efforts (January and May 2015) to increase genomic coverage in the face of unpredictable enzyme cutting frequencies. The window increase allowed the inclusion of larger fragments and may have improved uptake of fragments with properly annealed barcode sequences, culminating in higher pass rates for effort 2. It does not seem that the improvements in sequencing effort 2 stem from an overall increase in raw sequenced reads, which might be due to quality or size differences in fragment libraries,

because the average read count for samples sequenced with *SbfI–MspI* during efforts 2 and 3 was actually lower than for sequencing effort 1.

Influence of evolutionary divergence on RE selection

Using COI Sanger sequence data, we estimated the genetic divergence between selected sample pairs and explored the effect of increased divergence on ddRADseq analyses. Because allelic dropout is hypothesized to increase as the genetic divergence of samples increases (Gautier *et al.* 2013; Eaton 2014; Leaché *et al.* 2015b), we expected shared loci between sample pairs to decrease as divergence increased. With the exception of *Hypochilus*, our results are generally consistent with this hypothesis. We found that our *Hypochilus* samples had highly divergent COI sequences, in spite of geographically adjacent collection sites (Table S1, Supporting information), and yet had many more shared loci than would be expected given an exponential decay model (Fig. 3). Genomic fragmentation of these samples with RE combinations *EcoRI–MspI* and *SphI–MluCI* yielded similarly high numbers of SNPs and shared loci. These incongruent results may be explained by the underlying population structure of the species: described by some workers as 'living fossils', *Hypochilus* species are known to be extreme habitat specialists with limited dispersal (Hedin, 2001; Hedin & Wood 2002; Keith & Hedin 2012). Large mtDNA distances have previously been described for other southern Appalachian *Hypochilus* species, with reciprocal monophyly of populations within the closely related species, *Hypochilus thorelli*, even at geographic distances of less than 5 kilometres (Hedin & Wood 2002; Keith & Hedin 2012). However, analyses based on coalescent theory or assuming a nested clade hierarchy led the authors of these studies to suggest the presence of some restricted but non-zero gene flow via male dispersal. Our findings of higher than expected numbers of shared ddRADseq loci alongside a background of distant mtDNA sequences for *H. pococki* would be in line with this hypothesis of population structure in *H. thorelli*.

Restriction endonuclease combination had a pronounced effect on the ratio of shared loci to read count for sample pairs with greater estimated genomic divergence, likely in part due to the low vagility of the majority of species we investigated (Pinto-da-Rocha *et al.* 2007; Keith & Hedin 2012; Hedin *et al.* 2013; Leavitt *et al.* 2015). Few studies have examined absolute divergence time estimates for the arachnid species we sequenced, but Derkarabetian *et al.* (2011) used reliable biogeographic evidence (Graham 1999; Brunsfeld *et al.* 2001) for a speciation interval of 2–5 million years (not including 95% HPD) between *Sclerobunus idahoensis* and *S. nondimorphicus*. *SbfI* had the longest restriction site (and thus, predicted to have the rarest cutting frequency) of the REs we used in digestions, and the combination of *SbfI–MspI* appears to have yielded few but consistent genomic regions for pairs of greater estimated genetic divergence, such as *Calisoga* (Figs 1 and 2, Table 1). In prior studies where ddRADseq was employed for the purpose of phylogenetic inference, researchers expressly chose enzymes with very different restriction specificities and cutting frequencies, such as *MspI* combined with *SbfI* or *PstI* (Jones *et al.* 2013; Streicher *et al.* 2014; Leaché *et al.* 2015b; Meik *et al.* 2015). The speciation or introgression events of interest in these phylogenetic studies were estimated between 1 and 5 mya, but accurate interspecies relationships for clades as old as 60 mya have been presented using SNPs from RADseq techniques (Rubin *et al.* 2012; Cariou *et al.* 2013; Leaché *et al.* 2015b).

Among the sample pairs we studied, only *Speleonychia* had an estimated COI divergence less than 5% and thus it is unclear whether its placement in Fig. 3 could be considered 'outlying'. This cave-obligate species was represented by individuals from separate lava tubes in the Indian Heaven lava flow system of southern Washington (Briggs 1974), but the high number of shared loci and near identical COI sequences of these individuals may indicate gene flow mediated by previously undocumented interconnectivity of lava tubes. The similarity of the individuals sampled suggests future studies of the species using a ddRADseq approach may benefit from the use of endonucleases with short restriction sites, as the risks of allelic dropout decrease with closely related targets, while read depth may remain comparatively high (Gautier *et al.* 2013; Eaton 2014; Leaché *et al.* 2015b). Removing *Speleonychia* from consideration, shared loci count remains fairly invariable at ~7000 loci for COI divergences of 5% to 10%.

Our findings in this study come with some caveats. Although major instrumentation did not change throughout the three sequencing periods, only *Leiobunum* was sequenced across all four RE protocols, so some care must be taken when comparing the efficacy of protocols to each other. Furthermore, no samples were sequenced doubly, so it is unclear how reproducible our fragment libraries might be within or between sequencing efforts (Puritz *et al.* 2014). Such a study would be useful in determining saturation points for different RE combinations (i.e. the point at which increasing read count no longer increases locus recovery). Even so, this pilot experiment provides an initial framework for the exploration of RE effects on ddRADseq results as recommended by Davey *et al.* (2011). It should also be reiterated that the analysis parameters we employed for variant calling were quite conservative and suited especially for population-level analyses where large numbers of shared loci are expected and the consequences of

inaccuracy may be more severe. With reasonable relaxation of consensus base-calls and polymorphism allowances, even samples of considerable divergence may produce more robust data sets less prone to the over-splitting of loci (Harvey *et al*. 2015). Given the phylogenetic breadth covered by our sample, we expect these results to be of utility to arachnologists and workers on non-model systems with a variety of research goals related to molecular ecology. Also our general approach should be helpful in circumstances where digest simulations are unreliable and/or an adequate reference genome is lacking.

Acknowledgements

Specimen collection assistance was provided by J. Barklage, D. Carlson, E. Ciaccio, D. Elias, D. Leavitt, M. Lowder, N. Richart, J. Satler, S. Thomas and N. Tsurusaki. A. Gottscho assisted with library preparation in October 2014. The Reeder, Dinsdale and Rohwer labs at SDSU provided equipment access. National Science Foundation funding was provided to MB (DBI 1401110), AC (DBE 1321850) and MH (DEB 1354558). The manuscript was improved with comments from four anonymous reviewers.

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.

Baird NA, Etter PD, Atwood TS, *et al*. (2008) Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PLoS ONE*, **3**, e3376.

Barley AJ, Monnahan PJ, Thomson RC, Grismer LL, Brown RM (2015) Sun skink landscape genomics: assessing the roles of micro-evolutionary processes in shaping genetic and phenotypic diversity across a heterogeneous and fragmented landscape. *Molecular Ecology*, **24**, 1696–1712.

Blair C, Campbell CR, Yoder AD (2015) Assessing the utility of whole genome amplified DNA for next-generation molecular ecology. *Molecular Ecology Resources*, **15**, 1079–1090.

Brewer MS, Cotoras DD, Croucher PJP, Gillespie RG (2014) New sequencing technologies, the development of genomics tools, and their applications in evolutionary arachnology. *Journal of Arachnology*, **42**, 1–15.

Briggs T (1974) Troglobitic harvestmen recently discovered in North American lava tubes (Travuniidae, Erebomastridae, Triaenonychidae: Opiliones). *Journal of Arachnology*, **1**, 205–214.

Brunsfeld SJ, Sullivan J, Soltis DE, Soltis PS (2001) Comparative phylogeography of northwestern North America: a synthesis. In: *Integrating Ecological and Evolutionary Processes in a Spatial Context*(eds Silvertown J., Antonovics J.), pp. 319–339. Blackwell Science, Oxford.

Cao Z, Yu Y, Wu Y, *et al*. (2013) The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. *Nature Communications*, **4**, 2602.

Cariou M, Duret L, Charlat S (2013) Is RADseq suitable for phylogenetic inference? An *in silico* assessment and optimization. *Ecology and Evolution*, **3**, 846–852.

Clarke TH, Garb JE, Hayashi CY, *et al*. (2014) Multi-tissue transcriptomics of the black widow spider reveals expansions, co-options, and functional processes of the silk gland gene toolkit. *BMC Genomics*, **15**, 365.

Clarke TH, Garb JE, Hayashi CY, Arensburger P, Ayoub NA (2015) Spider transcriptomes identify ancient large-scale gene duplication event potentially important in silk gland evolution. *Genome Biology and Evolution*, **7**, 1856–1870.

Comeron JM, Ratnappan R, Bailin S (2012) The many landscapes of recombination in *Drosophila melanogaster*. *PLoS Genetics*, **8**, e1002905.

DaCosta JM, Sorenson MD (2014) Amplification biases and consistent recovery of loci in a double-digest RAD-seq protocol. *PLoS ONE*, **9**, e106713.

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.

Derkarabetian S, Steinmann DB, Hedin M (2010) Repeated and time-correlated morphological convergence in cave-dwelling harvestmen (Opiliones, Laniatores) from montane Western North America. *PLoS ONE*, **5**, e10388.

Derkarabetian S, Ledford J, Hedin M (2011) Genetic diversification without obvious genitalic morphological divergence in harvestmen (Opiliones, Laniatores, *Sclerobunus robustus*) from montane sky islands of western North America. *Molecular Phylogenetics and Evolution*, **61**, 844–853.

Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F (2014) The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*). *Nature*, **505**, 546–549.

Eaton DAR (2014) PyRAD: assembly of *de novo* RADseq loci for phylogenetic analyses. *Bioinformatics*, **30**, 1844–1849.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research*, **32**, 1792–1797.

Ellegren H (2014) Genome sequencing and population genomics in non-model organisms. *Trends in Ecology & Evolution*, **29**, 51–63.

Everett MV, Grau ED, Seeb JE (2011) Short reads and nonmodel species: exploring the complexities of next-generation sequence assembly and SNP discovery in the absence of a reference genome. *Molecular Ecology Resources*, **11**, 93–108.

Fernández R, Giribet G (2015) Unnoticed in the tropics: phylogenomic resolution of the poorly known arachnid order Ricinulei (Arachnida). *Royal Society Open Science*, **2**, 150065.

Foster PG, Hickey DA (1999) Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. *Journal of Molecular Evolution*, **48**, 284–290.

Gallego JC, Shirts BH, Bennette CS, *et al*. (2015) Next-generation sequencing panels for the diagnosis of colorectal cancer and polyposis syndromes: a cost-effectivess analysis. *Journal of Clinical Oncology*, **33**, 2084–2091.

Garrison NL, Rodriguez J, Agnarsson I, *et al*. (2015) Spider phylogenomics: untangling the spider tree of life. *Peer Journal*, **4**, e1719.

Gautier M, Gharbi K, Cezard T, *et al*. (2013) The effect of RAD allele dropout on the estimation of genetic variation within and between populations. *Molecular Ecology*, **22**, 3165–3178.

Georgiou G, Ippolito GC, Beausang J, Busse CE, Wardemann H, Quake SR. (2014) The promise and challenge of high-throughput sequencing of the antibody repertoire. *Nature Biotechnology*, **32**, 158–168.

Graham A (1999) *Late Cretaceous and Cenozoic History of North American Vegetation, North of Mexico*. Oxford University Press, New York.

Grbić M, Van Leeuwen T, Clark RM, *et al*. (2011) The genome of *Tetranychus urticae* reveals herbivorous pest adaptations. *Nature*, **479**, 487–492.

Gregory TR, Shorthouse DP (2003) Genome sizes of spiders. *Journal of Heredity*, **94**, 285–290.

Hamilton CA, Hendrixson BE, Bond JE (2016) Taxonomic revision of the tarantula genus *Aphonopelma* Pocock, 1901 (Araneae, Mygalomorphae, Theraphosidae) within the United States. *ZooKeys*, **560**, 1–340.

Haney RA, Ayoub NA, Clarke TH, Hayashi CY, Garb JE (2014) Dramatic expansion of the black widow toxin arsenal uncovered by multi-tissue transcriptomics and venom proteomics. *BMC Genomics*, **15**, 366.

Harvey MG, Judy CD, Seeholzer GF, Maley JM, Graves GR, Brumfield RT (2015) Similarity thresholds used in DNA sequence assembly from short reads can reduce the comparability of population histories across species. *PeerJ*, **3**, e895.

Hedin MC (2001) Molecular insights into species phylogeny, biogeography, and morphological stasis in the ancient spider genus *Hypochilus*

(Araneae: Hypochilidae). *Molecular Phylogenetics and Evolution*, **18**, 238–251.

Hedin M, Thomas SM (2010) Molecular systematics of eastern North American Phalangodidae (Arachnida: Opiliones: Laniatores), demonstrating convergent morphological evolution in caves. *Molecular Phylogenetics and Evolution*, **54**, 107–121.

Hedin M, Wood DA (2002) Genealogical exclusivity of geographically proximate populations of *Hypochilus thorelli* Marx (Araneae, Hypochilidae) on the Cumberland Plateau of North America. *Molecular Ecology*, **11**, 1975–1988.

Hedin M, Starrett J, Akhter S, Schönhofer AL, Shultz JW (2012a) Phylogenomic resolution of paleozoic divergences in harvestmen (Arachnida, Opiliones) via analysis of next-generation transcriptome data. *PLoS ONE*, **7**, e42888.

Hedin M, Tsurusaki N, Macías-Ordóñez R, Shultz JW (2012b) Molecular systematics of sclerosomatid harvestmen (Opiliones, Phalangioidea, Sclerosomatidae): geography is better than taxonomy in predicting phylogeny. *Molecular Phylogenetics and Evolution*, **62**, 224–236.

Hedin M, Starrett J, Hayashi C (2013) Crossing the uncrossable: novel trans-valley biogeographic patterns revealed in the genetic history of low-dispersal mygalomorph spiders (Antrodiaetidae, *Antrodiaetus*) from California. *Molecular Ecology*, **22**, 508–526.

Hess JE, Campbell NR, Docker MF, *et al.* (2015) Use of genotyping by sequencing data to develop a high-throughput and multifunctional SNP panel for conservation applications in Pacific lamprey. *Molecular Ecology Resources*, **15**, 187–202.

Hoffman JI, Simpson F, David P, *et al.* (2014) High-throughput sequencing reveals inbreeding depression in a natural population. *PNAS*, **111**, 3775–3780.

Jezkova T, Riddle BR, Card DC, Schield DR, Eckstut ME, Castoe TA. (2015) Genetic consequences of postglacial range expansion in two codistributed rodents (genus *Dipodomys*) depend on ecology and genetic locus. *Molecular Ecology*, **24**, 83–97.

Jones JC, Fan S, Franchini P, Scharti M, Meyer A (2013) The evolutionary history of *Xiphophorus* fish and their sexually selected sword: a genome-wide approach using restriction site-associated DNA sequencing. *Molecular Ecology*, **22**, 2986–3001.

Kearse M, Moir R, Wilson A, *et al.* (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*, **28**, 1647–1649.

Keith R, Hedin M (2012) Extreme mitochondrial population subdivision in southern Appalachian paleoendemic spiders (Araneae: Hypochilidae: *Hypochilus*), with implications for species delimitation. *Journal of Arachnology*, **40**, 167–181.

Kerkel K, Spadola A, Yuan E, *et al.* (2008) Genomic surveys by methylation-sensitive SNP analysis identify sequence-dependent allele-specific DNA methylation. *Nature Genetics*, **40**, 904–908.

Kliman RM, Hey J (1993) Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Molecular Biology and Evolution*, **10**, 1239–1258.

Leaché AD, Fujita MK, Minin VN, Bouckaert RR (2014) Species delimitation using genome-wide SNP data. *Systematic Biology*, **63**, 534–542.

Leaché AD, Chavez AS, Jones LN, Grummer JA, Gottscho AD, Linkem CW. (2015a) Phylogenomics of phrynosomatid lizards: conflicting signals from sequence capture versus restriction site associated DNA sequencing. *Genome Biology and Evolution*, **7**, 706–719.

Leaché AD, Banbury BL, Felsenstein J, Nieto-Montes de Oca A, Stamatakis A (2015b) Short tree, long tree, right tree, wrong tree: new acquisition bias corrections for inferring SNP phylogenies. *Systems Biology*, **64**, 1032–1047.

Leavitt DH, Starrett J, Westphal MF, Hedin M (2015) Multilocus sequence data reveal dozens of putative cryptic species in radiation of endemic Californian mygalomorph spiders (Araneae, Mygalomorphae, Nemesiidae). *Molecular Phylogenetics and Evolution*, **91**, 56–67.

Lechner M, Marz M, Ihling C, Sinz A, Stadler PF, Krauss V. (2013) The correlation of genome size and DNA methylation rate in metazoans. *Theory in Biosciences*, **132**, 47–60.

Lepais O, Weir JT (2014) SimRAD: an R package for simulation-based prediction of the number of loci expected in RADseq and similar genotyping by sequencing approaches. *Molecular Ecology Resources*, **14**, 1314–1321.

Lyko F, Maleszka R (2011) Insects as innovative models for functional studies of DNA methylation. *Trends in Genetics*, **27**, 127–131.

Mason NA, Taylor SA (2015) Differentially expressed genes match bill morphology and plumage despite largely undifferentiated genomes in a Holarctic songbird. *Molecular Ecology*, **24**, 3009–3025.

Mattila TM, Bechsgaard JS, Hansen TT, Schierup MH, Bilde T (2012) Orthologous genes identified by transcriptome sequencing in the spider genus *Stegodyphus*. *BMC Genomics*, **13**, 70.

Meik JM, Streicher JW, Lawing AM, Flores-Villela O, Fujita MK (2015) Limitations of climatic data for inferring species boundaries: insights from speckled rattlesnakes. *PLoS ONE*, **10**, e0131435.

Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE (2011) Landscape of next generation sequencing technologies. *Analytical Chemistry*, **83**, 4327–4341.

Nossa CW, Havlak P, Yue J, Lv J, Vincent KY (2014) Joint assembly and genetic mapping of the Atlantic horseshoe crab genome reveals ancient whole genome duplication. *GigaScience*, **3**, 9.

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.

Pinto-da-Rocha R, Machado G, Giribet G (eds.) (2007) *Harvestmen: The Biology of Opiliones*. Harvard Univ. Press, Cambridge, MA.

Planas E, Bernaus L, Ribera C (2014) Development of novel microsatellite markers for the spider genus *Loxosceles* (Sicariidae) using next-generation sequencing. *Journal of Arachnology*, **42**, 315–317.

Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. *Nucleic Acids Research*, **35**, D61–D65.

Puritz JB, Matz MV, Toonen RJ, Weber JN, Bolnick DI, Bird CE. (2014) Demystifying the RAD fad. *Molecular Ecology*, **23**, 5937–5942.

Recknagel H, Jacobs A, Herzyk P, Elmer KR (2015) Double-digest RAD sequencing using Ion Proton semiconductor platform (ddRADseq-ion) with nonmodel organisms. *Molecular Ecology Research*, **15**, 1316–1329.

Regev A, Lamb MJ, Jablonka E (1998) The role of DNA methylation in invertebrates: developmental regulation or genome defense? *Molecular Biology and Evolution*, **15**, 880–891.

Richart C, Hedin M (2013) Three new species in the harvestmen genus *Acuclavella* (Opiliones, Dyspnoi, Ischyropsalidoidea), including description of male *Acuclavella quattuor* Shear, 1986. *ZooKeys*, **311**, 19–68.

Rittmeyer EN, Austin CC (2015) Combined next-generation sequencing and morphology reveal fine-scale speciation in crocodile skinks (Squamata: Scincidae: *Tribolonotus*). *Molecular Ecology*, **24**, 466–483.

Rius M, Bourne S, Hornsby HG, Chapman MA (2015) Applications of next-generation sequencing to the study of biological invasions. *Current Zoology*, **61**, 488–504.

Rubin BER, Ree RH, Moreau CS (2012) Inferring phylogenies from RAD sequence data. *PLoS ONE*, **7**, e33394.

Sanggaard KW, Bechsgaard JS, Fang X, *et al.* (2014) Spider genomes provide insight into composition and evolution of venom and silk. *Nature Communications*, **5**, 3765.

Schield DR, Card DC, Adams RH, *et al.* (2015a) Incipient speciation with biased gene flow between two lineages of the Western Diamondback Rattlesnake (*Crotalus atrox*). *Molecular Phylogenetics and Evolution*, **83**, 213–223.

Schield DR, Walsh MR, Card DC, Andrew AL, Adams RH, Castoe TA. (2015b) EpiRADseq: scalable analysis of genomewide patterns of methylation using next-generation sequencing. *Methods in Ecology and Evolution*, **7**, 60–69.

Schwarz A, von Reumont BM, Erhart J, Chagas AC, Ribeiro JMC, Kotsyfakis M. (2013) *De novo Ixodes ricinus* salivary gland transcriptome analysis using two next-generation sequencing methodologies. *FASEB Journal*, **27**, 4745–4756.

Sharma PP, Kaluziak ST, Pérez-Porro AR, *et al.* (2014) Phylogenomic interrogation of Arachnida reveals systemic conflicts in phylogenetic signal. *Molecular Biology and Evolution*, **31**, 2963–2984.

Speiss AN, Neumeyer N (2010) An evaluation of R^2 as an inadequate measure for nonlinear models in pharmacological and biochemical research: a Monte Carlo approach. *BMC Pharma*, **10**, 6.

Stensrud Ø, Schumacher T, Shalchian-Tebrizi K, Bjorvand Svegården IB, Kauserud H (2007) Accelerated nrDNA evolution and profound AT bias in the medicinal fungus *Cordyceps sinensis*. *Mycological Research*, **111**, 409–415.

Streicher JW, Devitt TJ, Goldberg CS, Malone JH, Blackmon H, Fujita MK. (2014) Diversification and asymmetrical gene flow across time and space: lineage sorting and hybridization in polytypic barking frogs. *Molecular Ecology*, **23**, 3273–3291.

Varghese NJ, Mukherjee S, Ivanova N, *et al.* (2015) Microbial species delineation using whole genome sequences. *Nucleic Acids Research*, **43**, 6761–6771.

Worley KC, The Marmoset Genome Sequencing and Analysis Consortium (2014) The common marmoset genome provides insight into primate biology and evolution. *Nature Genetics*, **46**, 850–857.

Wu GA, Prochnick S, Jenkins J, *et al.* (2014) Sequencing of diverse mandarin, pummelo and orange genomes reveals complex history of admixture during citrus domestication. *Nature Biotechnology*, **32**, 656–662.

Zhou X, Xia Y, Ren X, *et al.* (2014) Construction of a SNP-based genetic linkage map in cultivated peanut based on large scale marker development using next-generation double-digest restriction-site-associated DNA sequencing (ddRADseq). *BMC Genomics*, **15**, 351.

M.B., J.S. and M.H. designed the research. M.B., J.S., S.D., C.H.R. and A.C. performed the research. M.B. and M.H. contributed reagents and analytical tools. Data were analysed by M.B., J.S., S.D, C.H.R. and A.C. M.B., J.S., S.D., C.H.R., A.C. and M.H. wrote the paper.

Data accessibility

DNA sequences: GenBank accessions JN547482, HM056740, HM056741, EU162815, GQ200413, GQ870663, HM056759, JN547512, GQ870668, JN547540, AF303513, JQ974855, JX951931, JX951911, KR182625, KR182673, HM056727, GQ200415, KX550439, KX570871, KX570872; NCBI SRA SRP078623.

Supporting Information

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

Table S1 Specimen locality information and sequencing date

Table S2 ANOVA and multiple comparisons table for loci counts as produced by four RE combinations