2019. Journal of Arachnology 47:190–201

Cost effective microsatellite isolation and genotyping by high throughput sequencing

Henrik Krehenwinkel[1,2,3], **Susanne Meese**[4,*], **Christoph Mayer**[5], **Jasmin Ruch**[6], **Jutta Schneider**[6], **Trine Bilde**[7], **Sven Künzel**[8], **James B. Henderson**[3], **Joseph Russack**[3], **Warren Brian Simison**[3], **Rosemary Gillespie**[2] and **Gabriele Uhl**[4]: [1]Department of Biogeography, University of Trier, Trier, Germany; E-mail: krehenwinkel@uni-trier.de; [2]Department of Environmental Science, Policy and Management, University of California, Berkeley, USA; [3]Center for Comparative Genomics, California Academy of Sciences, San Francisco, California, USA; [4]General and Systematic Zoology, Zoological Institute and Museum, University of Greifswald, Germany; [5]Zoologisches Forschungsmuseum Alexander Koenig, Bonn, Germany; [6]Institute of Zoology, Behavioural Biology, Universität Hamburg, Germany; [7]Department of Bioscience, Aarhus University, Denmark; [8]Max Planck Institute for Evolutionary Biology, Plön, Germany

Abstract. High throughput sequencing (HTS) has emerged as a valuable tool for the rapid isolation of genetic markers for population genetics and pedigree analysis. HTS-based SNP (single nucleotide polymorphism) genotyping protocols like RAD (Restriction-site associated DNA) sequencing or hybrid capture allow for the isolation of thousands of markers from any non-model organism. However, these protocols are relatively laborious and expensive and the resulting high marker density is not always necessary. Since HTS technology has also greatly simplified the process of microsatellite marker isolation and genotyping, we develop microsatellite markers as a cost-efficient and simple alternative to SNP genotyping. We present low coverage genome sequencing data from seven distantly related spider species (*Argiope bruennichi* (Scopoli, 1772), *Larinia jeskovi* Marusik, 1987, *Oedothorax retusus* (Westring, 1851), *Pisaura mirabilis* (Clerck, 1757), *Australomisidia ergandros* (Evans, 1995), *Cheiracanthium punctorium* (Villers, 1789), *Theridion grallator* Simon,1900) and show the utility of HTS for microsatellite isolation. We also present a simple Illumina amplicon sequencing protocol to genotype microsatellites from multiplex PCR amplicons in the Hawaiian happy face spider *T. grallator*. We discuss advantages and drawbacks of the use of microsatellites for a range of research questions, and highlight an unexpectedly fast decay and gain of repeat loci for *T. grallator*.

Keywords: Genome, paternity assessment, population genetics, amplicon sequencing

High throughput sequencing is currently revolutionizing molecular ecology and systematics. Thousands of markers can be isolated from any non-model organism with protocols like RAD (Restriction-site associated DNA) sequencing (Peterson et al. 2012) or sequence capture (Smith et al. 2013; Mayer et al. 2016) and whole genome and transcriptome sequencing are now feasible (Ekblom & Galindo 2011; Ellegren 2014). The resulting marker density has greatly contributed to an in-depth understanding of evolutionary and ecological processes, including in the field of arachnology (Brewer et al. 2014). Several spider genomes have recently been sequenced (Sanggaard et al. 2014; Babb et al. 2017; Schwager et al. 2017), the spider tree of life has been tackled using high-density markers (Bond et al. 2014; Fernández et al. 2014, 2018) and genomic and transcriptomic analyses have provided insights into evolutionary divergence in spiders (Croucher et al. 2013; Bechsgaard et al. 2015; Krehenwinkel et al. 2015; Settepani et al. 2017). The recent isolation of Ultra Conserved Elements (Starrett et al. 2017) and application of ddRAD (double digest) sequencing protocols (Burns et al. 2017; Settepani et al. 2017) have additionally contributed a wealth of genetic markers for spider research. However, RAD sequencing (Burns et al. 2017) or sequence capture (Smith et al. 2013; Cotoras et al. 2018) protocols are relatively laborious and expensive and the analysis of such high throughput genotyping data requires considerable computational resources. Depending on the question of research, a density of thousands of SNPs may not always be necessary, and thus a simpler and more cost-efficient approach is desirable for some types of studies. In this regard, microsatellites are noteworthy: They are repeated short sequences of DNA that evolve rapidly and hence are powerful markers for analyzing mating rates and paternity success, for population and conservation genetics, as well as pedigree analyses and recent evolutionary divergence (e.g., Schäfer et al. 2008; Tuni et al. 2012; Krehenwinkel & Tautz 2013; Zimmer et al. 2014; Krehenwinkel et al. 2016b). While their isolation used to be laborious, microsatellites are now routinely isolated from any species by shotgun sequencing of genomic DNA (Castoe et al. 2010; Malausa et al. 2011). Thus, while little used for arachnids in the past (Brewer et al. 2014), microsatellites are becoming available for an increasing number of spider species (Rütten et al. 2001; Bilde et al. 2009; da Silveira & Bonatto 2009; Hataway et al. 2011; Esquivel-Bobadilla et al. 2013; Parmakelis et al. 2013; Planas et al. 2014). High throughput sequencing has not only simplified the isolation of microsatellite markers but is also well-suited for genotyping of microsatellites (Cao et al. 2014; Darby et al. 2016). In particular, Illumina amplicon sequencing provides a rapid, accurate and very cost-efficient alternative to the laborious and expensive capillary electrophoresis protocols, which were commonly used for microsatellite genotyping (Cao et al. 2014; Darby et al. 2016). Large-scale amplicon sequencing of multiplex PCRs can be routinely performed for hundreds of samples in parallel (Fadrosh et al. 2014) and microsatellite alleles can be directly called from amplicon sequencing data by analyzing the read length distribution per specimen. Several software solutions for microsatellite genotyping from Illumina amplicon sequencing data have been published or are under development (Suez et al. 2016; Zhan et

* current address: Department of Environmental Systems Science, ETH Zurich, Switzerland

al. 2017; Henderson, Russack, Krehenwinkel & Simison unpublished data). Against this background, it is worthwhile to consider the utility of microsatellite markers as a simple and cost-efficient alternative to high throughput SNP genotyping.

Here, we provide an assessment of the isolation and genotyping processes of microsatellite markers. Our aims are twofold. We first analyze the feasibility of high throughput sequencing for microsatellite marker isolation. For this, we present low coverage genome sequencing data for seven distantly related spider species from six families. These species are important models in different fields of arachnology, from behavioral ecology, to phylogeography and population genetics. We identify tandem repeat contents in the analyzed spider genomes, isolate markers from the genomic data and present structural similarities of the repeat content between genomes. Based on these results, we discuss the necessary sequencing depth for microsatellite marker discovery in spiders. We then provide a set of established markers for the studied species. Secondly, we present an overview of the workflow for microsatellite genotyping using Illumina amplicon sequencing, from PCR set up, to library preparation and sequencing, and genotyping from raw read data, using the Hawaiian happy face spider *Theridion grallator* Simon, 1900. The phylogeography of the happy face spider has been studied using mitochondrial sequence information and allozyme data (Croucher et al. 2012); thus, by applying the approach to *T. grallator*, we can compare directly the effectiveness of microsatellite markers for recovering genetic differentiation among populations. The well-understood genetic history of the species also offers the potential to investigate the evolution of repeat loci over time, e.g., decay or gain of microsatellites between different populations. Based on our results, we discuss promises and drawbacks of Illumina based microsatellite analyses and highlight unique features of tandem repeat evolution in spiders. Our results demonstrate the feasibility of high throughput sequencing based microsatellite isolation and genotyping, and may help in establishing microsatellites as a more commonly used marker type for genetic studies on spiders.

METHODS

Target species.—We targeted seven distantly related spider species from six families to assess the utility of high throughput sequencing for microsatellite marker isolation. *Argiope bruennichi* (Scopoli, 1772) (Araneidae) serves as a model species in research on the evolution of sexual cannibalism and its consequences for mating systems (Fromhage et al. 2003; Schneider & Andrade 2011; Schneider 2014; Schneider et al. 2015; Uhl et al. 2015) and has recently gained importance as a model for evolutionary divergence during contemporary range expansions (Krehenwinkel & Tautz 2013; Krehenwinkel et al. 2015, 2016a). *Larinia jeskovi* Marusik, 1987 (Araneidae) is remarkable due to its mating behavior, since males mutilate the female genitalia to prevent further insemination by rival males (Mouginot et al. 2015, 2017). *Oedothorax retusus* (Westring, 1851) (Linyphiidae) has received considerable attention due to its mating strategy, in which males perform gustatory courtship and plug the female's genital opening with secretion (Kunz et al. 2012, 2014). *Pisaura mirabilis* (Clerck, 1757) (Pisauridae) is known as the nuptial gift spider, in which males provide females with a prey item prior to mating that results in sexual conflict (Bilde et al. 2007; Albo et al. 2013; Ghislandi et al. 2018). *Cheiracanthium punctorium* (Villers, 1789) (Cheiracanthiidae) is known for being the only medically important spider in Central Europe and is currently rapidly expanding its range (Muster et al. 2008; Krehenwinkel et al. 2016b). The thomisid species *Australomisidia ergandros* (Evans, 1995) shows communal hunting behavior (Ruch et al. 2014, 2015; Dumke et al. 2016) and brood care by which the female provides herself to her offspring as food (Evans 1998). The happy face spider, *Theridion grallator* Simon, 1900 (Theridiidae), is widely-distributed across the rainforests of the Hawaiian Islands of Oahu, Molokai, Maui, and Big Island, and is well known for its conspicuous and eponymous color morphs that occur as a balanced polymorphism in every population (Gillespie & Oxford 1998). The diverse research questions ranging from paternity and relatedness to population structure require fast evolving genetic markers.

Microsatellite isolation by high throughput sequencing.—Genomic DNA from six spider species was extracted from leg muscle tissue using the Qiagen DNeasy blood and tissue kit according to the manufacturer's protocol (Qiagen, Hilden, Germany). The taxonomy and origin of samples are shown in Table 1. The untreated genomic DNA of five species was sequenced, each on1/8th flow cell of a 454 GS FLX+ flow cell in 2010 and 2011. Library preparation and sequencing were performed according to the manufacturer's protocols (Roche, Basel, Switzerland; see Krehenwinkel & Tautz (2013) for more details on the 454 sequencing and analyses). An additional data set of the yellow sac spider *Cheiracanthium punctorium* was generated by sequencing pooled genomic DNA on two full flow cells of an Illumina Miseq in 2014 (Illumina, San Diego, CA, USA). Library preparation and sequencing were performed using the V3 chemistry according to the manufacturer's protocols and sequencing 300 bp paired end reads. The 454 and Illumina sequencing runs were both performed at the Max Planck Institute for Evolutionary Biology in Plön, Germany. The paired end Illumina reads were quality trimmed using PoPoolation (Kofler et al. 2011) with a minimum quality threshold of 20 and adapters removed using Trimmomatic (Bolger et al. 2014). A *de novo* assembly was generated including both Illumina libraries and using CLC genomic workbench at a minimum contig length of 1000 and including a scaffolding step (CLC Bio, Boston, USA). For the happy face spider *Theridion grallator*, we used a subset of ~20 Mb of contigs from a preliminary genome assembly, based on Illumina paired end sequencing data (Croucher, unpublished data). We estimated the GC content of reads and assemblies using UNIX.

We identified tandem repeats in the Illumina assemblies of *C. punctorium* and *T. grallator* and the raw 454 reads of the remaining spiders using MSat Commander (Faircloth 2008) with a minimum length of 10 repeats for mononucleotide repeats, 10 repeats for di-, 6 repeats for tri-, 5 repeats for tetra-, 4 repeats for penta- and hexanucleotide repeats. We counted the number of repeat loci for each repeat class and calculated the genome wide content of different repeat motifs per megabase of sequence data (Table 1, 2). In order to identify the tandem repeat content, we designed primers for

Table 1.—The upper table shows the taxonomy and sampling origin of the sequenced spider species, sequencing platform used, number of sequenced bases, number of sequenced reads and estimates for the genome-wide GC-content for the six species. The lower table presents the assembly statistics for the two assemblies used for marker isolation.

Family	Genus	Species	Origin	Platform	Sequenced bases	No. reads	GC-content (%)
Araneidae	*Argiope*	*bruennichi*	*Germany*	454	$30.29*10^6$	80,031	29
Araneidae	*Larinia*	*jeskovi*	*Poland*	454	$30.39*10^6$	101,616	31
Linyphiidae	*Oedothorax*	*retusus*	*Germany*	454	$24.79*10^6$	86,424	34
Pisauridae	*Pisaura*	*mirabilis*	*Germany*	454	$23.12*10^6$	81,970	34
Theridiidae	*Theridion*	*grallator*	*Hawaii*	-	-	-	27
Thomisidae	*Australomisidia*	*ergandros*	*Australia*	454	$24.18*10^6$	84,722	33
Cheiracanthiidae	*Cheiracanthium*	*punctorium*	*Baltic States*	MiSeq	$18.78*10^9$	19,673,693	34
Cheiracanthiidae	*Cheiracanthium*	*punctorium*	*Mediterranean*	MiSeq	$12.37*10^9$	14,150,027	35

Family	Genus	Species	Origin	Platform	Assembly size	Median contig size (bp)	GC-content (%)
Cheiracanthiidae	*Cheiracanthium*	*punctorium*	*Mediterranean*	MiSeq	$148.60*10^5$	1024	32
Theridiidae	*Theridion*	*grallator*	*Hawaii*	HiSeq	$22.00*10^6$	6290	27

all possible di, tri and tetranucleotide repeat loci using the Primer3 plugin (Rozen & Skaletsky 1999) of MSat Commander. A subset of the microsatellite loci was tested for variability using fluorescently labeled primers and following genotyping as described in Krehenwinkel & Tautz (2013).

Amplicon sequencing for microsatellite genotyping.—We used several Hawaiian populations of the happy face spider *Theridion grallator* to explore the utility of Illumina amplicon sequencing for microsatellite genotyping. Specimens of *T. grallator* were collected in March 2015 on the Hawaiian Islands of Oahu, Maui, Molokai and Hawai'i by beating vegetation and hand collecting from the underside of leaves. Two separate subpopulations per island were sampled (see Table 3). All specimens were stored in 99% ethanol and brought back to the University of California Berkeley for further analyses. DNA extractions were performed on the whole prosoma of each specimen using the Qiagen Puregene Tissue kit according to the manufacturer's protocol (Qiagen, Valencia, USA). The DNA from 96 specimens was quantified using a Qubit Fluorometer (Thermo Scientific, Waltham, USA), diluted to approximately 20 ng/μl and distributed among wells of a 96-well plate. Fifty primer pairs were tested in a subset of three specimens from each island for determining PCR amplification efficiency. The targeted loci contained di-, tri- or tetranucleotide repeats and the targeted amplicons were less than 400 bp long, to achieve a good overlap during read merging. Each primer pair was tested in an annealing temperature gradient from 50–60 °C in increments of 2.5 °C. Gradient PCR were run for each primer pair using the Qiagen Multiplex PCR kit according to the manufacturer's protocol and with 30 cycles. No Q-solution was added. PCR products were screened on a 1.5 % agarose gel. The 25 primer pairs which most consistently amplified specimens from all four islands were chosen for further analyses (see Supplementary Table 1, online at http://dx.doi.org/10.1636/JoA-S-16-017.s1 for details on primer pairs and repeat loci). This selection was done to avoid priming bias and drop out of loci by sequence divergence between populations. PCR reactions with primers of similar optimal annealing temperature were then combined to multiplex reactions. We added a 5′ tail to each primer and amplified the 25 primer pairs in six multiplex PCRs of 30 cycles for all of the 96 specimens and otherwise as described above. PCRs were run in 10 μl volumes, with 0.5 μl of each 10 μM primer and 1 μl of the 20 ng/μl template. The added 5′ tails served as a priming site for a second PCR round of 5 cycles, in which dual indices (short, unique sequences for individual identification) and Illumina Truseq adapters were introduced. The concept of PCR-based library preparation followed that described in Lange et al. (2014). Before the second PCR, all multiplexes were pooled into a single 96-well plate according to specimen and placed for an approximate quantification on a 1.5 % agarose gel. A second PCR round was run with 8 forward times 12 reverse combinations of indexed primers. This indexing PCR was run as described above, but with only 5 cycles at 55 °C annealing temperature, 0.25 μl of each 10 μM

Table 2.—Coverage of different tandem repeat motifs (TR) per megabase (Mb) of sequenced DNA for all spider species studied. The last two columns show the number of primer pairs which could be designed on the recovered tandem repeats per Mb of DNA and the % of recovered repeat motifs on which primers could be designed.

Species	TR motifs/Mb	Mono/Mb	Di/Mb	Tri/Mb	Tetra/Mb	Penta/Mb	Hexa/Mb	Primers/Mb	% motifs with primers
A. bruennichi	340.84	293.89	35.55	7.13	2.48	1.39	0.45	3.93	8.36
L. jeskovi	103.30	78.79	17.67	3.16	1.48	2.01	0.07	1.88	7.71
O. retusus	118.20	102.3	3.15	4.36	5.28	2.22	0.89	2.74	17.23
P. mirabilis	225.90	113.74	94.27	5.02	9.09	3.37	0.43	9.13	8.14
A. ergandros	248.30	200.18	25.93	6.86	13.44	1.61	0.33	8.15	16.92
C. punctorium	464.78	434.22	10.91	9.65	8.08	1.83	0.11	19.18	62.72
T. grallator	218.19	203.22	13.02	0.6	0.75	0.55	0.05	2.5	16.70

Table 3.—Collection sites and sample numbers per population for the eight sampled populations of *Theridion grallator* on the Hawaiian Islands.

Island	Population	Locality	N	H_{obs}	H_{exp}
Big Island	HiBiK15	Saddle Road, Kipuka15	12	0.277	0.351
Big Island	HiBiM21	Saddle Road, Milemarker 21	8	0.265	0.388
Maui	HiMaWA	Waikamaoi, outside preserve	13	0.485	0.676
Maui	HiMaWB	Waikamoi, upper preserve	7	0.424	0.678
Molokai	HiMoKA	Kamakou, TNC cabin	17	0.324	0.508
Molokai	HiMoKB	Kamakou, boardwalk	14	0.318	0.463
Oahu	HiOhPA	Pahole	10	0.369	0.485
Oahu	HiOhPB	Pahole	10	0.332	0.474

primer and 0.5 µl of the PCR template from the first PCR round. After each PCR, the products were purified from remaining primer using AMpure XP beads (Beckman Coulter, Brea, USA). The purified and dual indexed PCR products were quantified using a Qubit, and then 10 ng of each sample were pooled. The sample was sequenced on approximately 1/4 of a MiSeq flow cell using the Illumina V3 chemistry, with 300 bp paired reads and according to the manufacturer's protocol (Illumina, San Diego, USA). The remainder of the flow cell contained microbial 16S and arthropod mitochondrial COI samples. A "spike-in" of 15 % PhiX was added to the run.

Adapter sequences were trimmed from the raw reads using Trimmomatic (Bolger et al. 2014). Paired reads were then merged using PEAR (Zhang et al. 2014) with a minimum overlap of 75 bp and a minimum quality of 30. Only those assembled reads with at least 90 % of bases with Q30 or higher were transformed into fasta files using the fastx toolkit (Gordon & Hannon 2010). All sequences, starting with the forward primer and ending with the reverse primer of each specimen and each of the 25 microsatellite loci were filtered and saved to new files. This step was performed using UNIX and served to demultiplex all separate loci. For allele-calling, we measured the length of each amplicon and counted the abundance of each length per specimen and per locus using the programming language awk. By the previous filtering of sequences, which start and end exactly at the PCR primers, all fragments can be measured in exact relation to each other. We plotted the distribution of different fragment lengths for each specimen and locus using a custom-made software (Henderson, Russack, Krehenwinkel & Simison unpublished data). The approach is very comparable to that implemented in classic fragment length analysis software for dye-labelled allele-calls, e.g., Genemapper (Thermo Scientific, Waltham, USA). However, here we plotted the abundance of reads for different fragment lengths instead of the intensity of dye fluorescence (Fig. 1 A, B). The final version of our simple and user-friendly software solution for allele-calling will perform all tasks from sequence adapter trimming and assembly, to demultiplexing, allele-calling and exporting the called alleles in a user-friendly graphical interface (Henderson, Russack, Krehenwinkel & Simison unpublished data). Alleles were called for each specimen and locus and the fragment size distributions for each locus were manually inspected to edit allele calls. Due to locus and population specific stutter patterns and the possibility of flanking indels contributing to repeat length, this manual curation step proved essential. We used a minimum coverage of 40 reads per locus and specimen to call alleles. We additionally inspected a random subset of unassembled reads for their microsatellite pattern and compared the result with that identified based on assembled reads pairs. This step served to exclude assembly-based artifacts. MEGA (Tamura et al. 2013) was used to visualize repeat containing sequences and to align subsamples of sequences. We visually inspected the sequences of every specimen for the presence of repeat motifs, indels outside of repeat motifs and flanking SNPs.

The genetic structure of *T. grallator* across the Hawaiian Islands was evaluated by a STRUCTURE (Pritchard et al. 2000; Falush et al. 2003) analysis. STRUCTURE was run with an admixture model, k-values from 1–10, 10 replicates per k-value, 150,000 MCMC generations of which 50,000 were removed as burnin. The optimal number of clusters was identified using STRUCTURE HARVESTER (Earl et al. 2012) with the method described in Evanno et al. (2005). Pairwise F_{ST} values between populations were calculated in Genepop (Raymond & Rousset 1995).

RESULTS

Microsatellite isolation by high throughput sequencing.—The results of sequencing and assembly can be found in Table 1. The 454 runs yielded about 100,000 reads and 30,000,000 bp per species, the MiSeq runs between 15–20 million reads. The assembly resulted in 148,602 contigs over 1 kb long, with a median size of 1,024 and a total size of ~232 million bases (Table 1). All spider genomes were characterized by a relatively low GC content between 27–34 %. Depending on the species, we recovered between 103–465 microsatellites per megabase (Mb) of sequenced DNA (Table 2). The repeat content in spider genomes was highly biased towards mononucleotides and was variable between species. As mononucleotide repeats are hard to genotype, we did not design primers for this repeat type. Excluding mononucleotides, only between 15–112 microsatellites could be recovered per Mb. Due to the high AT content of spider genomes (Sanggaard et al. 2014; Krehenwinkel et al. 2015), AT rich repeats dominate in the isolated loci (Supplementary Table 1).

A relatively small subset of the identified repeat loci contained sufficient flanking sequences to design primers. Between 2–19 microsatellite markers could be recovered per sequenced Mb of DNA. This corresponds to between 8–62 % of the total number of recovered repeat loci. Moreover, a considerable proportion of designed primer pairs had to be dropped after initial analyses. On average, we could establish

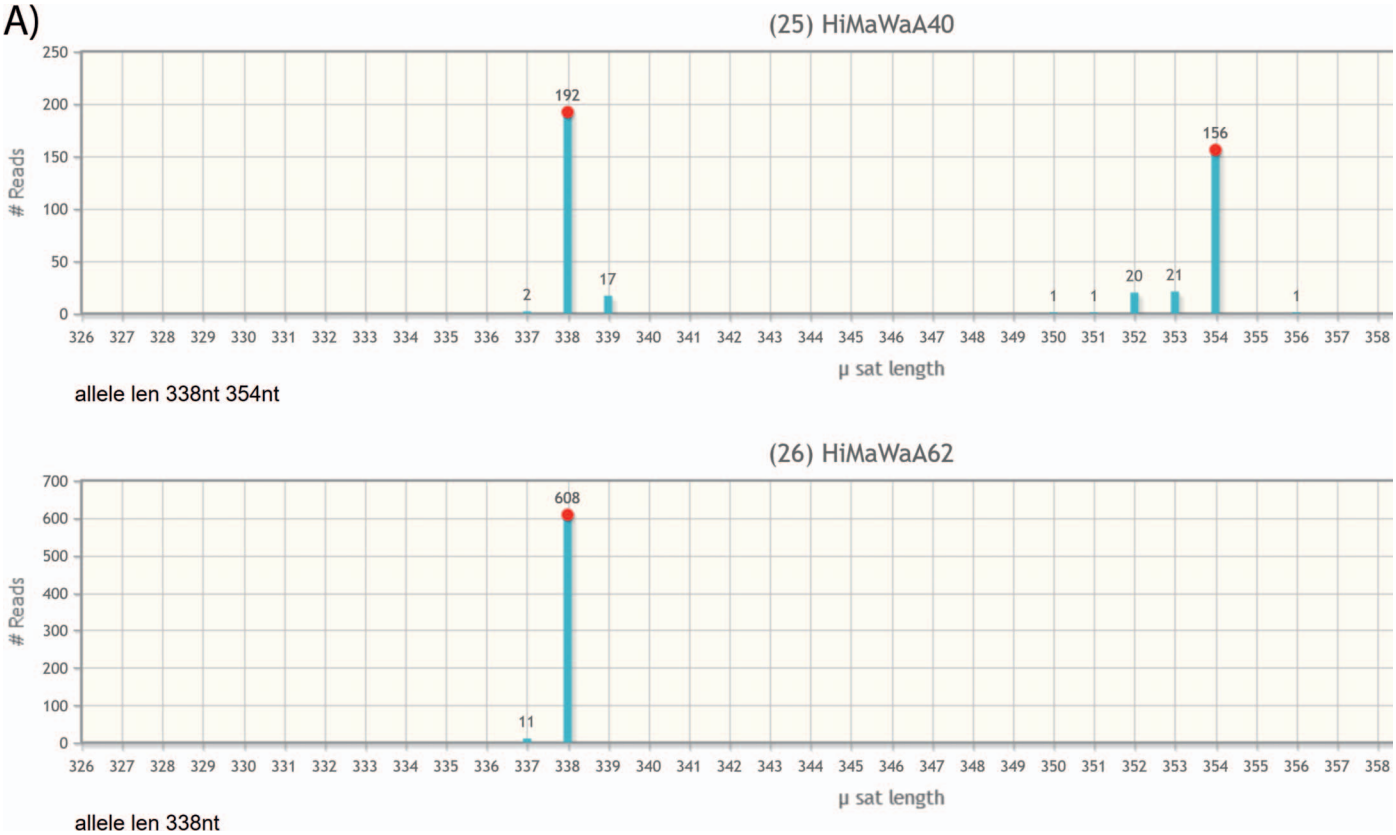

Figure 1.—(A) Subsection of genotyping plots for the allelic distribution of a CT-dinucleotide repeat locus (TG_MS41) for two happy face spider specimens from Maui. The plots show the abundance (Y-axis) of reads of different fragment length (X-axis). Each bar indicates the abundance of one fragment length in a mixture of sequences, with the red dot indicating the called allele length. The upper specimen is heterozygous and the lower is homozygous.

11 highly variable microsatellite markers out of 50 primer pairs, which we designed per species and tested in PCR assays. This low yield was due to two problems. First, some markers showed highly population specific amplification biases, leading to dropouts if genetically distant populations were analyzed. This problem was most evident in the analyses of divergent Eastern and Western Palearctic populations of *A. bruennichi* (Krehenwinkel et al. 2016a). Second, many repeat loci lacked size variation in some spider species, particularly *O. retusus* and *A. ergandros*. We provide a list of established primer sequences, examples for their application and amplification bias or repeat variation issues in Supplementary Table 1.

Amplicon sequencing for microsatellite genotyping.—After trimming and assembly, we recovered 26,425 high-quality sequences per *T. grallator* specimen on average (± 8,787 standard deviation). Four specimens had to be removed from the analysis due to low sequence coverage (12–16 sequences only). We found 1,001 reads on average (± 1,488 standard deviation) per microsatellite locus and specimen. One locus was removed from the analysis because coverage was too low (16 sequences per specimen on average). Our final dataset consisted of 92 specimens and 24 loci. Many recovered microsatellite loci showed variable repeat motifs within and among populations (Fig. 1A, B).

A STRUCTURE analysis of the microsatellite dataset supported k = 4 populations (Fig. 2), each of them corresponding to one of the Hawaiian Islands and in line with recent research by Croucher et al. (2012). An F_{ST} analysis corroborated these results (Table 4). While populations within islands showed little to no differentiation, we found considerable divergence between islands.

Alignments of the microsatellite amplicons allowed more detailed insights into the distribution of repeat patterns, flanking indels and SNPs in different populations of *T. grallator*, and in comparison to the population from Maui, for which the loci were designed (Fig. 3). Repeat sequences were often imperfect, e.g., containing additional bases which break the repeat pattern (70 % of loci; Fig. 3). One of the amplified loci did not contain any tandem repeat, despite being identified as a dinucleotide repeat in Msatcommander. Many microsatellites appeared to be population-specific and were absent in other populations. Specimens from Maui, which were also used as template to design microsatellites, carried repeat motifs for 23 out of 24 analyzed loci. In most other populations, we found an absence of many repeat loci, with about 1/3 of loci not showing the repeat in comparison to Maui (Fig. 4, Table 5). Overall, 13 out of 24 loci showed a missing repeat in at least one population. The decay of repeats was often part of the standing variation, with some specimens in a population carrying repeats and others not. Most of the loci, for which repeats were absent in other populations, carried the loss of the repeat pattern in the standing variation

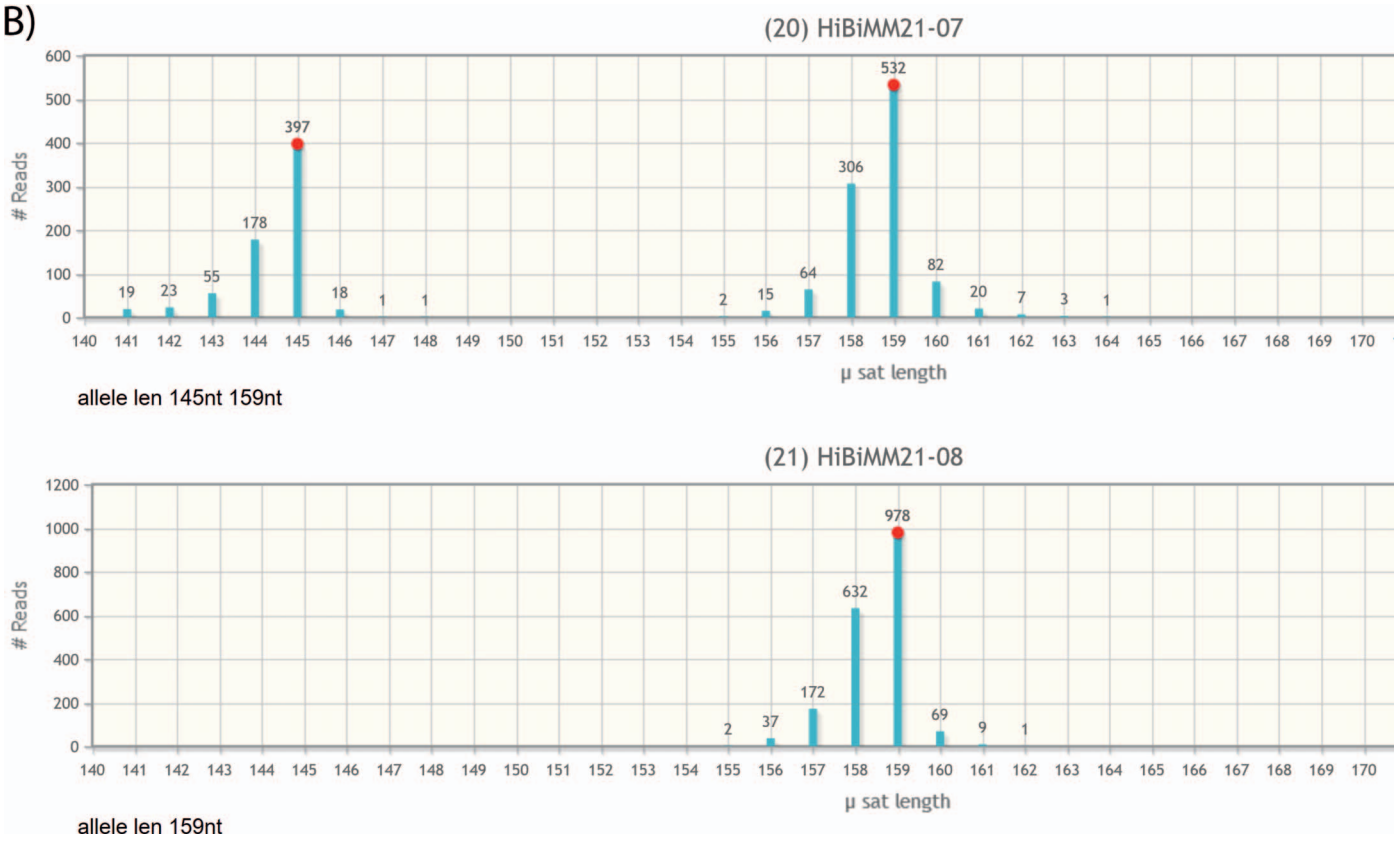

Figure 1.—(B) Subsection of genotyping plots for the allelic distribution of an AT-dinucleotide repeat (TG_MS1) for two happy face specimens from Hawaii Island. The plots show the abundance (Y-axis) of reads of different fragment length (X-axis). Each bar indicates the abundance of one fragment length in a mixture of sequences, with the red dot indicating the called allele length. The upper specimen is heterozygous and the lower is homozygous.

on Maui. Across the phylogeny of the happy face spider (Croucher et al. 2012), our data suggest considerable gains and losses of repeat loci within about 1 million years of inter-island colonization of the species (Table 5, Fig. 4).

Apart from losses of repeats, additional factors contributed to size differences in PCR amplicons (Fig. 3, Table 5). About 25 % of the analyzed loci carried a second tandem repeat motif, often right next to the targeted one. Moreover, indels outside of the targeted tandem repeat pattern had a substantial contribution to amplicon size differences. Depending on the population, up to 90% of the analyzed loci carried flanking indels in the amplicon. Even after a complete loss of the repeat motif, these indels contributed to variable fragment sizes. This was particularly important in populations outside of Maui, where a significant proportion of the repeat motifs decayed.

DISCUSSION

Microsatellite isolation by high throughput sequencing.—Our results show that microsatellite markers can be routinely isolated by low coverage sequencing from any spider genome. Simple high throughput sequencing of untreated genomic DNA will yield sufficient markers for certain types of population studies in spiders. It is important to aim for long reads to provide sufficient flanking regions for primer design. The best combination of read length and high output is currently offered by the Illumina MiSeq system. With its V3

Table 4.—Pairwise F_{ST} between Hawaiian populations of the happy face spider *Theridion grallator* for the microsatellite fragment length dataset generated by Illumina amplicon sequencing. The population names correspond to those in Table 3.

	HiBiK15	HiBiMM21	HiMaWaA	HiMaWaB	HiMoKaA	HiMoKaC	HiOhPeA
HiBiMM21	0.002						
HiMaWaA	0.302	0.277					
HiMaWaB	0.348	0.315	0.000				
HiMoKaA	0.372	0.341	0.250	0.255			
HiMoKaC	0.403	0.375	0.273	0.275	0.000		
HiOhPeA	0.497	0.463	0.337	0.354	0.437	0.460	
HiOhPeB	0.517	0.483	0.346	0.365	0.446	0.467	0.012

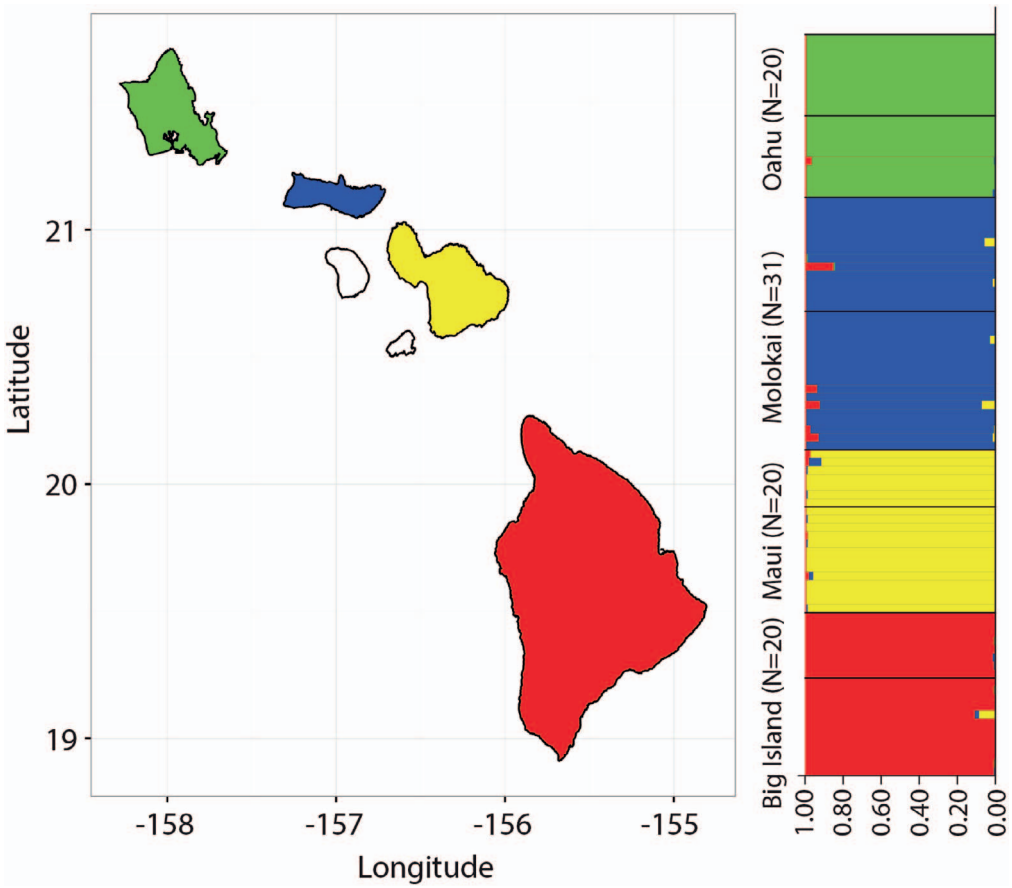

Figure 2.—Results (right) of a STRUCTURE analysis of the microsatellite dataset assuming k = 4 populations. Colors correspond to that of islands in the sampling map (left).

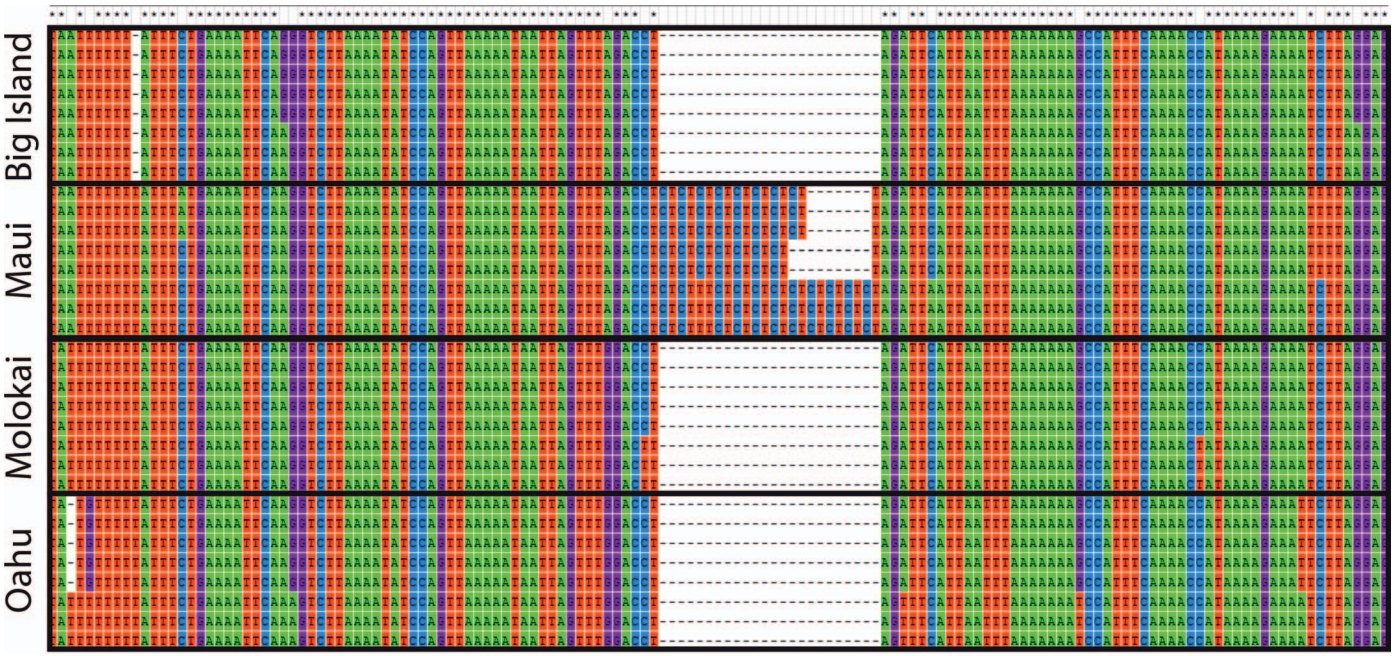

Figure 3.—Alignment of sequences of happy face spiders (*Theridion grallator*) from different Hawaiian Islands for a CT-dinucleotide repeat (MS41). Each island is represented with 8 sequences. The repeat motif is only found in populations on Maui and absent on all other islands. The repeat motif is not perfect, with a Thymine inserted instead of Cytosine in some specimens. Flanking SNPs support the signal of the repeat pattern. A flanking deletion is present in all specimens from Big Island and a subset of those from Oahu.

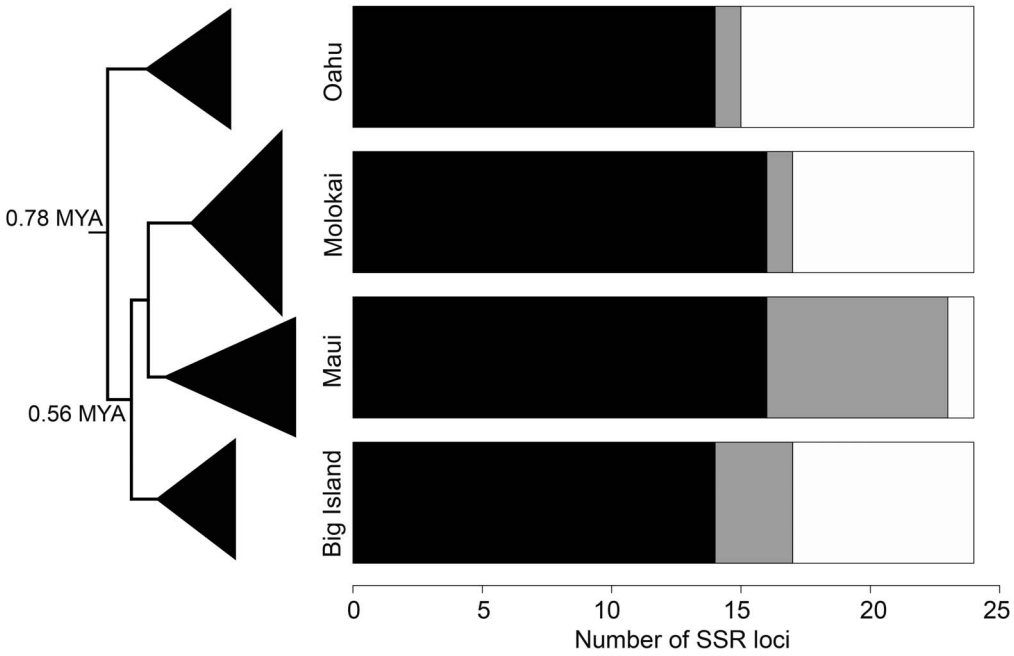

Figure 4.—Losses and gains of tandem repeat loci over the microsatellite-based phylogeny of the four happy face spider populations analyzed. The root of the tree and the divergence ages for the major clades are in accordance with Croucher et al. (2012). The bar plot shows the number of loci with a repeat motif present in all specimens (black), present in only a subset of the population, e.g., part of the standing variation of the population (grey) and absent (white).

chemistry, it is possible to obtain up to 20,000,000 paired reads of 2 × 300 bp per run. About 50 microsatellite loci should provide sufficient resolution for some population genetic applications depending on the question in focus, as well as for paternity analysis. Considering the recovery of 2–19 primer pairs per sequenced Mb of DNA, this translates to a minimum of 25 Mb of DNA that needs to be sequenced for microsatellite recovery. Our analyses in *T. grallator* suggest a very fast turnover of repeat loci in spider genomes. Due to the fast evolution of the repeat content, it is probably not possible to predict the expected repeat content for a target species from other related taxa. The repeat content has to be determined *de novo* for each species. We found a high dropout rate due to the lack of variation in repeat length for some taxa, e.g., many loci could not be used for further analysis in these taxa. Consequently, we recommend a higher sequencing coverage of about 250 Mb per specimen for the development of microsatellite markers. Considering the throughput of an Illumina MiSeq, this still translates to 80 separate spider species, for which microsatellites could be isolated in a single sequencing run. At a cost of about $1,600 USD per run, microsatellites for any spider species can be isolated for about $50 USD, including the cost for library preparation. Using enrichment protocols, the output of isolated loci could be increased substantially (Malausa et al. 2011).

Table 5.—Proportion of loci, out of 24 sequenced loci, for *Theridion grallator* populations from four islands, which (1) contain the tandem repeat motif targeted by primer design, (2) contain an additional tandem repeat motif, (3) contain indels outside of the tandem repeat motif, (4) carry variable SNPs.

	Big Island	Maui	Molokai	Oahu
Repeat motif present	0.71	0.96	0.71	0.63
Second repeat motif present	0.25	0.29	0.25	0.25
InDels present	0.67	0.92	0.83	0.75
SNPs present	0.96	0.96	1.00	1.00

Microsatellite genotyping using Illumina amplicon sequencing.—Microsatellite amplicons can be rapidly sequenced and genotyped using paired-end Illumina sequencing of multiplex PCRs. High throughput sequencing approaches come with many advantages over traditional capillary electrophoresis of dye-labelled amplicons. At the same time, the results of high throughput sequencing approaches are highly congruent with those of capillary electrophoresis (Vartia et al. 2016; Zhan et al. 2017). The high coverage of Illumina sequencing avoids the need to balance amplification of every single marker in a multiplex reaction. While the fluorescence of a single, overrepresented marker can outshine all other loci in a multiplex, the read count information for each locus is independent from that of the others. Even without previous optimization of multiplexes, we recover a very high number of loci (24 out of 25) and specimens (92 out of 96) in our analysis. There is also no need to select non-overlapping groups of marker sizes for multiplex reactions, as is typical for dye-labelled reactions. The only limit of an Illumina based approach is the current maximum size of paired read lengths of 2 × 300 bp. Assuming that each sequence is supposed to read through the tandem repeat, a large overlap is desired for read merging. Thus, amplicons should not exceed 400bp. Our approach of plotting read length abundance profiles per locus (Henderson, Russack, Krehenwinkel & Simison unpublished data) is comparable to classic allele-calling software. Recent work suggests automated calling options (Suez et al. 2016) or a

mixture of automated and manual curation of such datasets (Zhan et al. 2016).

For the genotyping of small to moderate numbers of markers, sequencing of microsatellite amplicons will be cheaper and require less laboratory experience than current SNP genotyping protocols, e.g., RAD sequencing. To generate the 24-locus dataset for each specimen, we required six multiplex PCRs, two clean up reactions and one indexing PCR, adding to $4 USD. Sequencing required another $4 USD per specimen. However, after further optimization of our multiplex PCRs, the number of PCRs and necessary sequencing coverage and processing cost could be considerably reduced. Recovering a higher genetic diversity and having a better population assignment power at low marker densities (Yang et al. 2011), microsatellites might even be preferable over SNPs for small-scale population studies. A combination of microsatellite isolation and genotyping by high throughput sequencing could allow the rapid and cost-efficient analysis of paternity or population structure.

While microsatellites can be valuable markers, microsatellite repeats can show highly complex and unpredictable patterns of evolution (Ellegren 2000) making modelling difficult. This can hamper the interpretation of microsatellite data and requires careful evaluation, e.g., by manual curation of the data before downstream analysis. Furthermore, it should be investigated whether more traditional enrichment marker design protocols as described in Nolte et al. (2005) and Leese et al. (2008) could be combined with HTS techniques, in order to further increase the effectiveness of the procedure.

Implications of a rapid turnover of microsatellite repeat loci in spiders.—The decay and possible gain of tandem repeat loci is a prominent pattern in our data, with almost half of the investigated loci not showing the repeat in at least one population of *T. grallator*. Only populations from Maui, on which the microsatellite loci used in this study were designed, consistently show all repeat loci. Recent molecular analyses suggest a stepping stone mode of colonization of the spiders from Oahu to Maui Nui (Maui and Molokai) and on to the Big Island within the past million years (Croucher et al. 2012). With only a few thousand years of coalescence time, the happy face spider is a young taxon. As an annual species, this corresponds to only a few thousand generations. A turnover of almost half of all repeat loci in this timeframe suggests a rapid decay and gain of repeat loci. The fast turnover may also be the reason for the lack of variation we found in many microsatellite sequences for some spider species, particularly *O. retusus* and *A. ergandros*. The lack of variation may also be the result of ascertainment bias (Ellegren et al. 1995). By choosing relatively long repeat motifs in our primer design, the maximum repeat motif length for some loci may be reached already. It is then likely by chance that evolution will decrease the size of repeat motifs in distant groups.

A study on interspecific decay rates of repeat motifs in Diptera and Hymenoptera, suggest losses of motifs in 30–40 % of loci after 2 million generations (Stolle et al. 2013). Spiders might thus show a faster rate of repeat motif evolution than other arthropod taxa. In contrast to Diptera and Hymenoptera, all spiders have large genome sizes (Gregory 2001). This might result in more genomic regions with little functional constraint, allowing for a faster random evolution and decay of repetitive sequences. Moreover, spider genomes generally show a low GC content (Sanggaard et al. 2014; Krehenwinkel et al. 2015). A low GC content in tandem repeats and their flanking sequence has been suggested to possibly contribute to an increased mutation rate (Schlötterer 2000). Microsatellites are often associated with transposable elements (Ellegren 2004; Meglécz et al. 2007). Their emergence by repeat mobilization could explain the rapid appearance and losses of repeats in *T. grallator*. However, little is currently known about microsatellite evolution in invertebrates (Chapuis et al. 2015), and future studies will have to explore this topic in more detail.

Practically, the rapid decay and gain of repeat loci means that called allele sizes for microsatellites in distant populations are often only based on flanking indels and not the evolving repeat. This urges care in cross-species amplification, still a popular approach (Moodley et al. 2015). This problem is probably not specific for spiders. In particular, distances based on models of evolution for the actual repeat motif might be biased by repeat loss. Another practical issue with the lack of size variation in repeat motifs is the need to explore large numbers of microsatellite loci to identify sets of variable markers for some spider species. However, due to the simplicity of multiplex PCRs and the high throughput of current amplicon sequencing protocols, the scoring of large numbers of microsatellite markers is still a worthwhile approach for studying population genetics in spiders.

Current high throughput sequencing technology allows the rapid and cost-efficient isolation of large numbers of microsatellite markers from spiders as we have shown in seven distantly related spider species. Moreover, Illumina amplicon sequencing is well suited for genotyping of microsatellite markers. As we highlight in an exemplary species, amplicon sequencing based microsatellite genotyping offers a greatly simplified workflow over currently used capillary electrophoresis-based protocols. Provided that microsatellite data are carefully analyzed, our results demonstrate that microsatellite markers can be a useful alternative to SNP genotyping for population genetics and pedigree analyses.

ACKNOWLEDGMENTS

We cordially thank Thoomke Brüning, Sarah Frehse, Nadja Gogrefe and Nicole Thomsen for assistance during lab work. We thank Andrew Rominger, Ellie Armstrong and Nate Yuen for help collecting happy face spider specimens. Anna Sellas and the California Academy of Sciences' Center for Comparative Genomics, are acknowledged for support during sequencing. Jun Ying Lim kindly provided a map of the Hawaiian Archipelago. We acknowledge the Nature Conservancy and the Hawaiian authorities for providing the necessary permits to collect happy face spiders. Natalie Graham provided helpful comments on the manuscript. Diethard Tautz provided helpful discussion and access to high throughput sequencing machines. HK was funded by a PhD fellowship of the Studienstiftung des Deutschen Volkes and by a postdoctoral fellowship of the German Research Foundation (DFG). The research was supported by NSF grant DEB 1241253 to RG and by the DFG to GU (Uh/87-7,

RTG 2010). TB was supported by The Danish Council for Independent Research grant number 4002-00328B.

LITERATURE CITED

Albo, M.J., T. Bilde & G. Uhl. 2013. Sperm storage mediated by cryptic female choice for nuptial gifts. Proceedings of the Royal Society of London B 280:20131735, doi: 10.1098/rspb.2013.1735.

Babb, P.L., N.F. Lahens, S.M. Correa-Garhwal, D.N. Nicholson, E.J. Kim, J.B. Hogenesch et al. 2017. The *Nephila clavipes* genome highlights the diversity of spider silk genes and their complex expression. Nature Genetics 49:895.

Bechsgaard, J., B. Vanthournout, P. Funch, S. Vestbo, R.A. Gibbs, S. Richards et al. 2015. Comparative genomic study of arachnid immune systems indicates loss of beta-1, 3-glucanase-related proteins and the immune deficiency pathway. Journal of Evolutionary Biology 29:277–291.

Bilde, T., C. Tuni, A. Cariani, A. Santini, C. Tabarroni, F. Garoia et al. 2009. Characterization of microsatellite loci in the subsocial spider *Stegodyphus lineatus* (Araneae: Eresidae). Molecular Ecology Resources 9:128–130.

Bilde, T., C. Tuni, R. Elsayed, S. Pekár & S. Toft. 2007. Nuptial gifts of male spiders: sensory exploitation of the female's maternal care instinct or foraging motivation? Animal Behaviour 73:267–273.

Bolger, A.M., M. Lohse & B. Usadel. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30:2114–2120.

Bond, J.E., N.L. Garrison, C.A. Hamilton, R. Godwin, M. Hedin & I. Agnarsson. 2014. Phylogenomics resolves a spider backbone phylogeny and rejects a prevailing paradigm for orb web evolution. Current Biology 24:1765–1771.

Brewer, M.S., D.D. Cotoras, P.J. Croucher & R.G. Gillespie. 2014. New sequencing technologies, the development of genomics tools, and their applications in evolutionary arachnology. Journal of Arachnology 42:1–15.

Burns, M., J. Starrett, S. Derkarabetian, C.H. Richart, A. Cabrero & M. Hedin. 2017. Comparative performance of double-digest RAD sequencing across divergent arachnid lineages. Molecular Ecology Resources 17:418–430.

Cao, M. D., E. Tasker, K. Willadsen, M. Imelfort, S. Vishwanathan, S. Sureshkumar et al. 2014. Inferring short tandem repeat variation from paired-end short reads. Nucleic Acids Research 42:e16.

Castoe, T.A., A.W. Poole, W. Gu, A.P. Jason de Koning, J.M. Daza, E.N. Smith et al. 2010. Rapid identification of thousands of copperhead snake (*Agkistrodon contortrix*) microsatellite loci from modest amounts of 454 shotgun genome sequence. Molecular Ecology Resources 10:341–347.

Chapuis, M.P., C. Plantamp, R. Streiff, L. Blondin & C. Piou 2015. Microsatellite evolutionary rate and pattern in *Schistocerca gregaria* inferred from direct observation of germline mutations. Molecular Ecology 24:6107–6119.

Cotoras, D.O., K. Bi, M.S. Brewer, D.R. Lindberg, S. Prost & R.G. Gillespie. 2018. Co-occurrence of ecologically similar species of Hawaiian spiders reveals critical early phase of adaptive radiation. BMC Evolutionary Biology 18:100.

Croucher, P.J., G.S. Oxford, A. Lam, N. Mody & R.G. Gillespie. 2012. Colonization history and population genetics of the color-polymorphic Hawaiian happy-face spider *Theridion grallator* (Araneae, Theridiidae). Evolution 66:2815–2833.

Croucher, P.J., M.S. Brewer, C.J. Winchell, G.S. Oxford & R.G. Gillespie. 2013. De novo characterization of the gene-rich transcriptomes of two color-polymorphic spiders, *Theridion grallator* and *T. californicum* (Araneae: Theridiidae), with special reference to pigment genes. BMC Genomics 14:862.

da Silveira, L.C.T. & S.L. Bonatto. 2009. Isolation and characterization of 12 dinucletiotide microsatellite loci in *Paratrechalea galianoae* (Araneae, Trechaleidae), a nuptial gift-spider. Molecular Ecology Resources 9:539–541.

Darby, B.J., S.F. Erickson, S.D. Hervey & S.N. Ellis-Felege. 2016. Digital fragment analysis of short tandem repeats by high-throughput amplicon sequencing. Ecology and Evolution 6:4502–4512.

Dumke, M., M.E. Herberstein & J.M. Schneider. 2016. Scrounging or producing: individual feeding tactics change with group size in a communally foraging spider. Proceedings of the Royal Society of London B 283:20160114.

Earl, D.A. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. Conservation Genetics Resources 4:359–361.

Ekblom, R. & J. Galindo. 2011. Applications of next generation sequencing in molecular ecology of non-model organisms. Heredity 107:1–15.

Ellegren, H. 2000. Microsatellite mutations in the germline: implications for evolutionary inference. Trends in Genetics 16:551–558.

Ellegren, H. 2004. Microsatellites: simple sequences with complex evolution. Nature Reviews Genetics 5:435–445.

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

Ellegren, H., C.R. Primmer & B.C. Sheldon. 1995. Microsatellite 'evolution': directionality or bias? Nature Genetics 11:360.

Esquivel-Bobadilla, S., O.A. Lozano-Garza, F.J. García-De-León, I.D.L.A. Barriga-Sosa & M.L. Jiménez. 2013. Development and characterization of 14 microsatellite loci in the beach wolf spider (*Arctosa littoralis*), using next-generation sequencing. Conservation Genetics Resources 5:261–263.

Evanno, G., S. Regnaut & J. Goudet. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. Molecular Ecology 14:2611–2620.

Evans, T.A. 1998. Offspring recognition by mother crab spiders with extreme maternal care. Proceedings of the Royal Society of London B 265:129–134.

Fadrosh, D.W., B. Ma, P. Gajer, N. Sengamalay, S. Ott, R.M. Brotman & J. Ravel 2014. An improved dual-indexing approach for multiplexed 16S rRNA gene sequencing on the Illumina MiSeq platform. Microbiome 2:1.

Faircloth, B.C. 2008. Msatcommander: detection of microsatellite repeat arrays and automated, locus-specific primer design. Molecular Ecology Resources 8:92–94.

Falush D., M. Stephens & J.K. Pritchard. 2003. Inference of population structure using multilocus genotype data linked loci and correlated allele frequencies. Genetics 164:1567–1587.

Fernández, R., G. Hormiga & G. Giribet. 2014. Phylogenomic analysis of spiders reveals nonmonophyly of orb weavers. Current Biology 24:1772–1777.

Fernández, R., R.J. Kallal, D. Dimitrov, J.A. Ballesteros, M.A. Arnedo, G. Giribet & G. Hormiga. 2018. Phylogenomics, diversification dynamics, and comparative transcriptomics across the Spider Tree of Life. Current Biology 28:1489–1497.

Fromhage, L., G. Uhl & J.M. Schneider. 2003. Fitness consequences of sexual cannibalism in female *Argiope bruennichi*. Behavioral Ecology and Sociobiology 55:60–64.

Ghislandi, P.G., S. Pekár, M. Matzke, S. Schulte-Doinghaus, T. Bilde & C. Tuni. 2018. Resource availability, mating opportunity, and sexual selection intensity influence the expression of male alternative reproductive tactics. Journal of Evolutionary Biology 31:1035–1046.

Gillespie, R.G. & G.S. Oxford. 1998. Selection on the color polymorphism in Hawaiian happy-face spiders: evidence from genetic structure and temporal fluctuations. Evolution 52:775–783.

Gordon, A. & G.J. Hannon. 2010. Fastx-toolkit. FASTQ/A short-reads preprocessing tools (unpublished) Online at http://hannonlab.cshl.edu/fastx_toolkit

Gregory, T.R. 2001. Animal genome size database. 2001. Online at http://www.genomesize.com/

Hataway, R.A., D.H. Reed & B.P. Noonan. 2011. Development of 10

microsatellite loci in the wolf spider *Arctosa sancterosae* (Araneae: Lycosidae). Conservation Genetics Resources 3:271–273.

Kofler, R., P. Orozco-terWengel, N. De Maio, R.V. Pandey, V. Nolte, A. Futschik et al. 2011. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PloS ONE 6:e15925.

Krehenwinkel, H. & D. Tautz. 2013. Northern range expansion of European populations of the wasp spider *Argiope bruennichi* is associated with global warming–correlated genetic admixture and population-specific temperature adaptations. Molecular Ecology 22:2232–2248.

Krehenwinkel, H., M. Graze, D. Rödder, K. Tanaka, Y.G. Baba, C. Muster et al. 2016a. A phylogeographical survey of a highly dispersive spider reveals eastern Asia as a major glacial refugium for Palaearctic fauna. Journal of Biogeography 43:1583–1594.

Krehenwinkel, H., D. Rödder, M. Năpăruş-Aljančič & M. Kuntner. 2016b. Rapid genetic and ecological differentiation during the northern range expansion of the venomous yellow sac spider *Cheiracanthium punctorium* in Europe. Evolutionary Applications 9:1229–1240.

Krehenwinkel, H., D. Rödder & D. Tautz. 2015. Eco-Genomic analysis of the poleward range expansion of the wasp spider *Argiope bruennichi* shows rapid adaptation and genomic admixture. Global Change Biology 21:4320–4332.

Kunz, K., S. Garbe & G. Uhl. 2012. The function of the secretory cephalic hump in males of the dwarf spider *Oedothorax retusus* (Linyphiidae: Erigoninae). Animal Behaviour 83:511–517.

Kunz, K., M. Witthuhn & G. Uhl. 2014. Do the size and age of mating plugs alter their efficacy in protecting paternity? Behavioural Ecology and Sociobiology 68:1321–1328.

Lange, V., I. Böhme, J. Hofmann, K. Lang, J. Sauter, B. Schöne et al. 2014. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. BMC Genomics 15:1.

Leese, F., C. Mayer & C. Held. 2008. Isolation of microsatellites from unknown genomes using known genomes as enrichment templates. Limnology & Oceanography: Methods 6:412–426.

Malausa, T., A. Gilles, E. Meglecz, H. Blanquart, S. Duthoy, C. Costedoat et al. 2011. High-throughput microsatellite isolation through 454 GS-FLX Titanium pyrosequencing of enriched DNA libraries. Molecular Ecology Resources 11:638–644.

Mayer, C., M. Sann, A. Donath, M. Meixner, L. Podsiadlowski, R.S. Peters et al. 2016. BaitFisher: A software package for multispecies target DNA enrichment probe design. Molecular Biology and Evolution 33:1875–1886. Online at https://doi.org/10.1093/molbev/msw056

Meglécz, E., S.J. Anderson, D. Bourguet, R. Butcher, A. Caldas, A. Cassel-Lundhagen et al. 2007. Microsatellite flanking region similarities among different loci within insect species. Insect Molecular Biology 16:175–185.

Moodley, Y., J.F. Masello, T.L. Cole, L. Calderon, G.K. Munimanda, M.R. Thali et al. 2015. Evolutionary factors affecting the cross-species utility of newly developed microsatellite markers in seabirds. Molecular Ecology Resources 15:1046–1058.

Mouginot, P., J. Prügel, U. Thom, P.O.M. Steinhoff, J. Kupryjanowicz & G. Uhl. 2015. Securing paternity by mutilating female genitalia in spiders. Current Biology 25:1–5.

Mouginot, P., G. Uhl & L. Fromhage. 2017. Evolution of external female genital mutilation: why do males harm their mates? Royal Society Open Science 4:171195.

Muster, C., A. Herrmann, S. Otto & D. Bernhard. 2008. Zur Ausbreitung humanmedizinisch bedeutsamer Dornfinger-Arten *Cheiracanthium mildei* und *C. punctorium* in Sachsen und Brandenburg (Araneae: Miturgidae). Arachnologische Mitteilungen 35:13–20.

Nolte, A.W., K.C. Stemshorn & D. Tautz. 2005. Direct cloning of microsatellite loci from *Cottus gobio* through a simplified enrichment procedure. Molecular Ecology Notes 5:628–636.

Parmakelis, A., K. Balanika, S. Terzopoulou, F. Rigal, R.R. Beasley, K.L. Jones et al. 2013. Development of 28 polymorphic microsatellite markers for the endemic Azorean spider *Sancus acoreensis* (Araneae, Tetragnathidae). Conservation Genetics Resources 5:1133–1134.

Peterson, B.K., J.N. Weber, E.H. Kay, H.S. Fisher & H.E. Hoekstra 2012. Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. PLoS ONE 7:e37135.

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

Pritchard, J.K., M. Stephens & P. Donnelly. 2000. Inference of population structure using multilocus genotype data. Genetics 155:945–959.

R Core Team 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Online at https://www.R-project.org/

Raymond, M. & F. Rousset. 1995. GENEPOP on the Web (Version 3.4). Online at http://wbiomed.curtin.edu.au/genepop/ Updated from Raymond & Rousset.

Rozen, S.H. & H. Skaletsky. 1999. Primer3 on the WWW for general users and for biologist programmers. Pp. 365–386. *In* Bioinformatics Methods and Protocols. Vol. 132: Methods in Molecular Biology. (S. Misener & S.A. Krawetz eds.). Humana Press, New York.

Ruch, J., M. Dumke & J.M. Schneider. 2015. Social network structure in group-feeding spiders. Behavioral Ecology & Sociobiology 69:1429–1436.

Ruch, J., M.E. Herberstein & J.M. Schneider. 2014. Offspring dynamics affect food provisioning, growth and mortality in a brood-caring spider. Proceedings of the Royal Society of London B 281:20132108.

Rütten, K.B., I. Schulz, K. Olek & G. Uhl. 2001. Polymorphic microsatellite markers in the spider *Pholcus phalangioides* isolated from a library enriched for CA repeats. Molecular Ecology Notes 1:255–257.

Sanggaard, K.W., J.S. Bechsgaard, X. Fang, J. Duan, T.F. Dyrlund, V. Gupta et al. 2014. Spider genomes provide insight into composition and evolution of venom and silk. Nature Communications 5:3765.

Schäfer, M.A., B. Misof & G. Uhl. 2008. Effects of body size of both sexes and female mating history on male behaviour and paternity success in a spider. Animal Behaviour 76:75–86.

Schlötterer, C. 2000. Evolutionary dynamics of microsatellite DNA. Chromosoma 109:365–371.

Schneider, J.M. 2014. Sexual cannibalism as a manifestation of sexual conflict. *In* Sexual Conflict. (B. Rice & S. Gavrilets eds.). Cold Spring Harbor Perspectives in Biology. doi: 10.1101/cshperspect.a017731.

Schneider, J.M. & M.C.D. Andrade. 2011. Mating behaviour and sexual selection. Pp. 215–275. *In* Spider Behaviour: Variability and Versatility. (M.E. Herberstein, ed.). Cambridge University Press, Cambridge UK.

Schneider, J., G. Uhl & M.E. Herberstein. 2015. Cryptic female choice within the genus *Argiope*: A comparative approach. Pp. 55–77. *In* Cryptic Female Choice in Arthropods: Patterns, Mechanisms and Prospects. (A.V. Peretti & A. Aisenberg, eds.). Springer, Heidelberg.

Schwager E.E., P.P. Sharma, T. Clarke, D.J. Leite, T. Wierschin, M. Pechmann et al. 2017. The house spider genome reveals an ancient whole-genome duplication during arachnid evolution. BMC Biology 15:62.

Settepani, V., M.F. Schou, M. Greve, L. Grinsted, J. Bechsgaard & T. Bilde. 2017. Changes in mating system and life history traits with the evolution of sociality leads to depleted genomic diversity at

both population and species level. Molecular Ecology 26:4197–4210.

Smith, B.T., M.G. Harvey, B.C. Faircloth, T.C. Glenn & R.T. Brumfield. 2013. Target capture and massively parallel sequencing of ultraconserved elements (UCEs) for comparative studies at shallow evolutionary time scales. Systematic Biology 63:83–95.

Starrett, J., S. Derkarabetian, M. Hedin, R.W. Bryson Jr., J.E. McCormack & B.C. Faircloth. 2017. High phylogenetic utility of an ultraconserved element probe set designed for Arachnida. Molecular Ecology Resources 17:812–823.

Stolle, E., J.H. Kidner & R.F. Moritz. 2013. Patterns of evolutionary conservation of microsatellites (SSRs) suggest a faster rate of genome evolution in Hymenoptera than in Diptera. Genome Biology and Evolution 5:151–162.

Suez, M., A. Behdenna, S. Brouillet, P. Graça, D. Higuet & G. Achaz. 2016. MicNeSs: genotyping microsatellite loci from a collection of (NGS) reads. Molecular Ecology Resources 16:524–533.

Tamura, K., G. Stecher, D. Peterson, A. Filipski. & S. Kumar. 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Molecular Biology and Evolution 30:2725–2729.

Tuni, C., S. Goodacre, J. Bechsgaard & T. Bilde. 2012. Moderate multiple parentage and low genetic variation reduces the potential for genetic incompatibility avoidance despite high risk of inbreeding. PLOSOne, 7(1): e29636. doi:10.1371/journal.pone.0029363

Uhl, G., S.M. Zimmer, D. Renner & J.M. Schneider. 2015. Exploiting a moment of weakness: male spiders escape sexual cannibalism by copulating with molting females. Scientific Reports 5:16928, DOI: 10.1038/srep16928.

Vartia, S., J.L. Villanueva-Cañas, J. Finarelli, E.D. Farrell, P.C. Collins, G.M. Hughes et al. 2016. A novel method of microsatellite genotyping-by-sequencing using individual combinatorial barcoding. Royal Society Open Science 3:150565.

Yang, X., Y. Xu, T. Shah, H. Li, Z. Han, J. Li & J. Yan. 2011. Comparison of SSRs and SNPs in assessment of genetic relatedness in maize. Genetica 139:1045–1054.

Zhan, L., I.G. Paterson, B.A. Fraser, B. Watson, I.R. Bradbury, P. Nadukkalam Ravindran et al. 2017. MEGASAT: automated inference of microsatellite genotypes from sequence data. Molecular Ecology Resources 17:247–256.

Zhang, J., K. Kobert, T. Flouri & A. Stamatakis. 2014. PEAR: a fast and accurate Illumina Paired-End reAd mergeR. Bioinformatics 30:614–620.

Zimmer, S.M., H. Krehenwinkel & J.M. Schneider. 2014. Rapid range expansion is not restricted by inbreeding in a sexually cannibalistic spider. PLoS ONE 7:e95963.

Manuscript received 6 March 2016, revised 7 January 2019.

SUPPLEMENTAL MATERIAL

The following data sets are available in the Dryad Digital repository (Online at doi:10.5061/dryad.c4d7cc0)

1. 454 reads and Assemblies for all studied species
2. Isolated primer sequences for all studied spider species
3. Primer sequences, which have already been tested for variability and PCR amplification success.
4. Illumina reads for microsatellite genotyping of Hawaiian happy face spider populations.