Restriction enzyme optimization for RADseq with camel spiders (Arachnida: Solifugae)

Authors: Santibáñez-López, Carlos E., Farleigh, Keaka, Cushing, Paula E., and Graham, Matthew R.

Source: The Journal of Arachnology, 48(3) : 346-350

Published By: American Arachnological Society

URL: https://doi.org/10.1636/JoA-S-20-040

2021. Journal of Arachnology 48:346–350

SHORT COMMUNICATION

Restriction enzyme optimization for RADseq with camel spiders (Arachnida: Solifugae)

Carlos E. Santibáñez-López[1], **Keaka Farleigh**[2], **Paula E. Cushing**[3] and **Matthew R. Graham**[1]: [1]Department of Biology, Eastern Connecticut State University, 83 Windham Street, Willimantic, Connecticut 06226, U.S.A.; E-mail: santibanezlopezc@easternct.edu; [2]Department of Biology, Miami University of Ohio, 700 East High Street, Oxford, Ohio 45056, U.S.A.; [3]Department of Zoology, Denver Museum of Nature & Science, 2001 Colorado Boulevard, Denver, Colorado 80205, U.S.A.

Abstract. Phylogenetic relationships and evolutionary patterns within the arachnid order Solifugae are poorly understood and largely unresolved due to conserved morphology and scarce genomic resources. In this study, we evaluated the role of restriction endonuclease (RE) selection in double-digest restriction-site-associated DNA sequencing (ddRADseq) as a methodology for exploring the evolutionary history of solifuges and their responses to changing desert landscapes and climate. We optimized this method by using computer simulations to explore the effect of different enzyme combinations on the process *in silico*. Genome data are not yet available for Solifugae, so we performed the simulations using available spider, scorpion, tick, mite and xiphosuran genomes. Guided by the simulations, we then pioneered ddRADseq in Solifugae by generating data for four samples representing two families and three genera. Our results highlight the utility of simulated data and give us confidence that ddRADseq will be ideal for studying the evolution of solifuges.

Keywords: 3RAD, Illumina, ddRADseqtools, simulations, sun spiders, wind scorpions

https://doi.org/10.1636/JoA-S-20-040

Camel spiders, also referred to as sun spiders or wind scorpions, are neither spiders nor scorpions; instead they comprise the arachnid order Solifugae. Although they are common elements of arid environments around the world, knowledge of camel spider biology is limited, even in well-studied systems such as the arid lands of North America. Lack of information about the biology and taxonomy of this group is partly a result of the challenges inherent in collecting solifuges and maintaining them in the lab. Collecting techniques are improving (Cushing & González-Santillán 2018; Graham et al. 2019) as is our understanding of basic camel spider biology (Bird et al. 2015; Franz-Guess & Starck 2016; Franz-Guess et al. 2016) and evolutionary relationships (Cushing et al. 2015; Botero-Trujillo et al. 2017; Maddahi et al. 2017; Ballesteros et al. 2019).

Our research team is contributing to this initiative by using genomic techniques to revise the North American camel spider family Eremobatidae and to explore their evolutionary history relative to changing landscapes and climates. To address the latter, we are pioneering the use of restriction-site associated DNA sequencing (RADseq) in Solifugae using a three-enzyme approach called 3RAD. Briefly, the method uses pairs of restriction enzymes to cleave genomic DNA. Combinations of indexed adapters are then added to the restriction cut sites, a third enzyme is used to cleave adapter dimers, and the genome is reduced using size selection. The resulting fragments with adapters (indexed libraries) are then sequenced using high-throughput techniques. Bioinformatics pipelines are used to demultiplex and assemble sequences. The resulting single nucleotide polymorphisms (SNPs) can then be used to address questions by employing approaches from phylogenetics, phylogeography, and population genetics.

As a first step toward utilizing the RADseq approach with solifuges, we first wanted to identify the most efficient enzyme combinations. In other words, which restriction enzyme pairs produce the most loci? In a recent study, Burns et al. (2017) assessed the efficacy of various enzyme combinations in non-model arachnids (harvestmen and spiders) by generating RADseq data. The approach has also been applied in genomic studies in scorpions (e.g., Bryson et al. 2016, 2018) and spiders (Graham et al. 2020). The order Solifugae, however, has not been studied using ddRADseq data.

To determine optimal enzyme pairs for camel spiders, we built on the foundational work of Burns et al. (2017) and explored the effect of enzyme combinations on arachnid genomes *in silico* using ddRADseqTools (Mora-Márquez et al. 2017). A camel spider genome has not yet been produced, or at least made publicly available; therefore, we conducted the analyses using genomes for the following species (Table 1): one tick (*Ixodes scapularis* Say, 1821), one mite (*Varroa jacobsoni* Oudemans, 1904), two spiders [*Parasteatoda tepidariorum* (C.L. Koch, 1841) and *Stegodyphus dumicola* Pocock, 1898], two scorpions [*Centruroides sculpturatus* (Wood, 1863) and *Mesobuthus martensii* (Karsch, 1879)], and two xiphosurans [*Limulus polyphemus* (Linnaeus, 1758) and *Tachypleus tridentatus* (Leach, 1819)].

ddRADseqTools consists of several python scripts to simulate a digestion in a reference genome using user-designated pairs of restriction enzymes (RE) (*rsitesearch.py*); then, it simulates paired-end (PE) Illumina ddRADseq raw reads (*simddradseq.py*). In this study, we explored the performance of six enzyme combinations (Table 1), with fragment sizes of 462 to 638 bp. We simulated PE raw read files with two individuals (per species and enzyme combination), sampling up to 1000 loci and generating 3 million reads. All other parameters were kept as default. Next, we quantified and removed PCR duplicates obtained from our read simulation (*pcrdupremoval.py*), and demultiplexed the reads per individual (*indsdemultiplexing.py*). The demultiplexed raw reads were assembled using ipyrad version 0.9.41 (Eaton & Overcast 2020) with default parameters (except for the cluster threshold = 0.95 and keeping 1 sample per locus for output).

To ground-truth the results from the computer simulations, we generated actual RADseq data with the optimal enzyme combination for four different camel spider species, representing two families and three genera: *Eremocosta bajaensis* (Muma, 1951) (Eremobatidae), *Eremocosta titania* Roewer, 1934 (Eremobatidae), *Hemerotrecha*

Table 1.—The number of reads and loci for simulated and observed RADseq data sets using different enzyme combinations.

Order	Species	Accession number	Reference	Type_data	RE	# Reads	# Loci
Araneae	*Parasteatoda tepidariorum*	GCA_000365465.6	Schwager et al. 2017	Simulations	BamHi-ClaI	87296	2164
					EcoRI - ClaI	535649	13703
					EcoRI-MseI	148284	2289
					SbfI-MspI	12621	278
					SphI - MluCI	37894	540
					SphI - MspI	450394	11474
	Stegodyphus dumicola	GCA_010614865.6	Liu et al. 2019	Simulations	BamHi-ClaI	304793	4938
					EcoRI - ClaI	1255757	28773
					EcoRI-MseI	607821	9087
					SbfI-MspI	43605	1071
					SphI - MluCI	204510	3219
					SphI - MspI	1500547	29644
Parasitiformes	*Ixodes scapularis*	GCA_002892825.4	Miller et al. 2018	Simulations	BamHi-ClaI	321625	5180
					EcoRI - ClaI	597275	11012
					EcoRI-MseI	1497331	23712
					SbfI-MspI	146152	2700
					SphI - MluCI	1499007	25915
					SphI - MspI	1498612	22953
	Varroa jacobsoni	GCA_002532875.9	Techer et al. 2019	Simulations	BamHi-ClaI	90308	2976
					EcoRI - ClaI	221776	7294
					EcoRI-MseI	205294	6731
					SbfI-MspI	19883	662
					SphI - MluCI	155993	5136
					SphI - MspI	381491	12555
Scorpiones	*Centruroides sculpturatus*	GCA_000671375.4	Schwager et al. 2017	Simulations	BamHi-ClaI	117058	2852
					EcoRI - ClaI	462236	11027
					EcoRI-MseI	106841	2073
					SbfI-MspI	17566	286
					SphI - MluCI	36602	603
					SphI - MspI	277278	5912
	Mesobuthus martensii	GCA_000484575.6	Cao et al. 2013	Simulations	BamHi-ClaI	180618	3961
					EcoRI - ClaI	492269	12607
					EcoRI-MseI	110168	2329
					SbfI-MspI	12527	298
					SphI - MluCI	32192	730
					SphI - MspI	248947	5697
Xiphosura	*Limulus polyphemus*	GCA_000517525.6	Lavrov et al. 2000	Simulations	BamHi-ClaI	105867	2788
					EcoRI - ClaI	354870	10269
					EcoRI-MseI	344651	7776
					SbfI-MspI	22878	643
					SphI - MluCI	195206	3843
					SphI - MspI	508135	13122
	Tachypleus tridentatus	GCA_004210375.6	GenBank Submission	Simulations	BamHi-ClaI	114009	2249
					EcoRI - ClaI	414901	9626
					EcoRI-MseI	437395	6409
					SbfI-MspI	27991	524
					SphI - MluCI	251985	2863
					SphI - MspI	509926	11068
Solifugae	*Ammotrechula* sp.	SRR11840038	This study	Empirical	EcoRI - ClaI	3077992	2131
	Eremocosta titania	SRR11840036	This study	Empirical	EcoRI - ClaI	21100460	39261
	Eremocosta bajaensis	SRR11840035	This study	Empirical	EcoRI - ClaI	13836386	6694
	Hemerotrecha branchi	SRR11840037	This study	Empirical	EcoRI - ClaI	19980046	2899

branchi Muma, 1986 (Eremobatidae), and *Ammotrechula* sp. (Ammotrechidae). We began by extracting genomic DNA from cheliceral muscles using a DNeasy Tissue Kit (Qiagen, Valencia, CA, USA). We then generated dual-digest, quadruple indexed RADseq libraries following procedures outlined in Hoffberg et al. (2016), but with enzyme combinations that performed well *in silico* (Table 1). Libraries were checked for quality using Bioanalyzer and then sequenced using 2x150 PE sequencing on a full lane of an Illumina HiSeq X at Admera Health (South Plainfield, NJ). Raw reads were demultiplexed using IPYRAD.

Computer simulations confirm that certain enzyme combinations can produce better results than others among different arachnid orders. The only exception, however, were the results for Parasitiformes (Fig. 1A), which yielded similar numbers of loci among the different RE pairs. Our results were concordant with Burns et al. (2017), finding that the RE combination with the smallest number of loci for scorpions, spiders and horseshoe crabs was SphI-MluCI,

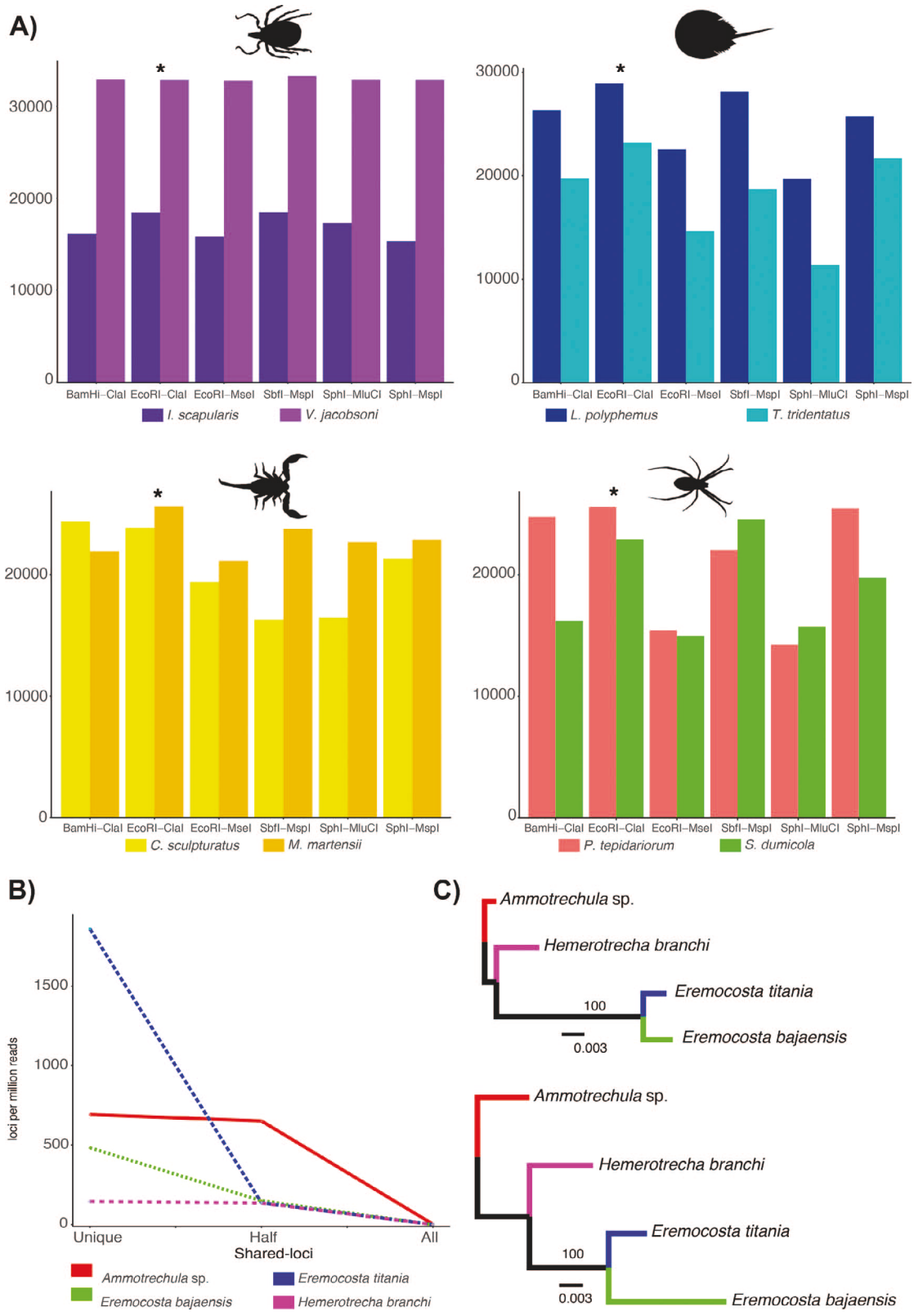

Figure 1.—(A) Simulated numbers of loci per million reads for different restriction enzyme (RE) combinations using the genomes of two Parasitiformes (top left corner), two Xiphosura (top right corner), two Scorpiones (bottom left corner) and two Araneae (bottom right corner). Asterisks identify the optimal enzyme pair for each arachnid group (based on mean values for each species pair). (B) Unique and shared loci recovered from ddRADseq with four Solifugae species. (C) Maximum likelihood phylogenetic tree recovered using IQ-Tree and the loci found in at least one species (top) and loci shared by the four species (bottom). Numbers above branches indicate ultrabootstrap support values.

followed by EcoRI-MseI (Fig. 1A). In *C. sculpturatus*, SbfI-MspI had a similar performance as SphI-MluCI; however, this was not the case for *M. martensii* which had a better performance than BamHI-ClaI (Fig. 1A). The other four RE combinations (SbfI-MspI, SphI-MspI, BamHI-ClaI and EcoRI-ClaI) yielded an average of 19,396 loci per million reads for Araneae, Scorpiones and Xiphosura, with EcoRI-ClaI generating slightly more loci (Fig. 1A). In contrast, all RE combinations yielded similar numbers in both species of Parasitiformes, with more loci recovered from the whole shotgun genome of *V. jacobsoni* (Fig. 1A). Based on these results, EcoRI-ClaI should be the best enzyme combination to use for ddRADseq studies of arachnids, of the RE combinations we assayed.

We were uncertain if genome size would influence the number of loci recovered. In the EcoRI-ClaI simulations, however, the number of loci recovered was similar among species with and without genome duplication [*V. jacobsoni* = 2N; spiders and scorpions = 4N; xiphosurans = 8N (Schwager et al. 2017)]. Interestingly, for *I. scapularis*, which is diploid, we obtained half as many loci as we did for *V. jacobsoni*.

Our assemblies recovered using actual ddRADseq with the EcoRI/ClaI combination yielded more than 2,100 loci in each of the four solifuge species studied (Table 1). The number of shared loci in the four species ranged from 9 (when forcing the presence of the four species per locus) to about 2,000 (two species per locus) (Fig. 1B). To assess the phylogenetic utility of our data, we performed Maximum Likelihood phylogenetic reconstructions with IQ-Tree (Nguyen et al. 2015), using ModelFinderPlus (Kalyaanamoorthy et al. 2017) and ultrabootstrap support (Hoang et al. 2018). The two matrices constructed, using loci present in at least one species (m1 = 9,680,468 bp, constant sites = 99.8%; parsimony informative sites = 42; distinct site patterns = 409; model = TVM+F, lnL = -13200280.8) and those shared by all species (m2 = 2,376 bp, constant sites = 96.7%; parsimony informative sites = 42; distinct site patterns = 69; model HKY+F, lnL = -3768.2), recovered the phylogenetic relationship of the four species in agreement with previous topologies (Cushing et al. 2015). Specifically, our topology was compatible with the monophyly of genus *Eremocosta* Roewer, 1934 (*E. bajaensis* + *E. titania*), and the monophyly of family Eremobatidae Kraepelin, 1899 (*Eremocosta* + *Hemerotrecha*) with high support (Fig. 1C). Furthermore, the EcoRI/ClaI combination yielded from 28,700 single nucleotide polymorphisms (SNPs) and 4,525 unlinked SNPs (uSNPS), when half of the species were present per locus. A total of 93 SNPs and 9 uSNPs were recovered when all species were represented at each locus.

Ultimately, we recommend using the *Eco*RI/*Cla*I RE combination for ddRADseq projects with arachnids. In addition, as we demonstrated with camel spiders, *in silico* simulations can be used to optimize studies of non-model taxa that lack published genomes by analyzing those of related species. Moving forward, we now feel confident that we can use ddRADseq to effectively explore the evolutionary history of solifuges and contribute to our understanding of these enigmatic arachnids and the arid landscapes they inhabit.

ACKNOWLEDGMENTS

We thank Jack Brookhart, Erika Garcia, and R. Ryan Jones for assistance in the field. Funding for this project was provided by NSF grant DEB-1754030 awarded to MRG and NSF grants DEB-1754587 and DEB-0640245 awarded to PEC.

LITERATURE CITED

Ballesteros, J.A., C.E. Santibáñez-López, L. Kovác, E. Gavish-Regev & P.P. Sharma. 2019. Ordered phylogenomic subsampling enables diagnosis of systematic errors in the placement of the enigmatic arachnid order Palpigradi. Proceedings of the Royal Society B 286:20192426.

Bird, T., R.A. Wharton & L. Prendini. 2015. Cheliceral morphology in Solifugae (Arachnida): primary homology, terminology, and character survey. Bulletin of the American Museum of Natural History 394:1–355.

Botero-Trujillo, R., R. Ott & L.S. Carvalho. 2017. Systematic revision and phylogeny of the South American sun-spider genus *Gaucha* Mello-Leitão (Solifugae: Mummuciidae), with description of four new species and two new generic synonymies. Arthropod Systematics & Phylogeny 75:3–44.

Bryson Jr., R.W., W.E. Savary, A.J. Zellmer, R. Bruce Bury & J.E. McCormack. 2016. Genomic data reveal ancient microendemism in forest scorpions across the California Floristic Province. Molecular Ecology 25:3731–3751.

Bryson Jr., R.W., D.A. Wood, M.R. Graham, M.E. Soleglad & J.E. McCormack. 2018. Genome-wide SNP data and morphology support the distinction of two new species of *Kovarikia* Soleglad, Fet & Graham, 2014 endemic to California (Scorpiones, Vaejovidae). Zookeys 739:79–106.

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, Z., Y. Yu, Y. Wu, P. Hao, Z. Di, Y. He et al. 2013. The genome of *Mesobuthus martensii* reveals a unique adaptation model of arthropods. Nature Communications 4:2602.

Catchen, J., P.A. Hohenlohe, S. Bassham, A. Amores & W.A. Cresko. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology 22:3121–3140.

Cushing, P.E. & E. González-Santillán. 2018. Capturing the elusive camel spider (Arachnida: Solifugae): effective methods for attracting and capturing solifuges. Journal of Arachnology 46:384–387.

Cushing, P.E., M.R. Graham, L. Prendini & J.O. Brookhart. 2015. A multilocus molecular phylogeny of the endemic North American camel spider family Eremobatidae (Arachnida: Solifugae). Molecular Phylogenetics and Evolution 92:280–293.

Eaton, D.A.R. & I. Overcast. 2020. ipyrad: Interactive assembly and analysis of RADseq datasets. Bioinformatics (early access btz966).

Franz-Guess, S. & J.M. Starck. 2016. Histological and ultrastructural analysis of the respiratory tracheae of *Galeodes granti* (Chelicerata: Solifugae). Arthropod Structure & Development 45:452–461.

Franz-Guess, S., B.J. Klußmann-Fricke, C.S. Wirkner, L. Prendini & J.M. Starck. 2016. Morphology of the tracheal system of camel spiders (Chelicerata: Solifugae) based on micro-CT and 3D-reconstruction in exemplar species from three families. Arthropod Structure & Development 45:440–451.

Graham, M.R., M.B. Pinto & P.E. Cushing. 2019. A test of the light attraction hypothesis in camel spiders of the Mojave Desert (Arachnida: Solifugae). Journal of Arachnology 47:293–296.

Graham, M.R., C. E. Santibáñez-López, S. Derkarabetian & B. Hendrixson. 2020. Pleistocene persistence and expansion in tarantulas on the Colorado Plateau and the effects of missing data on phylogeographical inferences from RADseq. Molecular Ecology. https://doi.org/10.1111/mec.15588

Hoang, D. T., O. Chernomor, A. von Haeseler, B.Q. Minh & L.S. Vinh. 2018. UFBoot2: Improving the ultrafast bootstrap approximation. Molecular Biology and Evolution 35:518–522.

Hoffberg, S. L., T.J. Kieran, J.M. Catchen, A. Devault, B.C. Faircloth, R. Mauricio et al. 2016. RAD cap: sequence capture of dual-digest RAD seq libraries with identifiable duplicates and reduced missing data. Molecular Ecology Resources 16:1264–1278.

Kalyaanamoorthy, S., B.Q. Minh, T.K.F. Wong, A. von Haeseler & L.S. Jermiin. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nature Methods 14:587–589.

Lavrov, D.V., J.L. Boore & W.M. Brown. 2000. The complete

mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. Molecular Biology and Evolution 17:813–824.

Liu, S., A. Aageaard, J. Bechsgaard & T. Bilde. 2019. DNA methylation patterns in the social spider, *Stegodyphus dumicola*. Genes 10(2):E137.

Maddahi, H., M. Khazanehdari, M. Aliabadian, H.G. Kami, A. Mirshamsi & O. Mirshamsi. 2017. Mitochondrial DNA phylogeny of camel spiders (Arachnida: Solifugae) from Iran. Mitochondrial DNA Part A 28:909–919.

Miller, J.R., S. Koren, K.A. Dilley, D.M. Harkins, T.B. Stockwell, R.S. Shabman et al. 2018. A draft genome sequence for the *Ixodes scapularis* cell line, ISE6. F1000Res 7:297.

Mora-Márquez, F., V. García-Olivares, B.C. Emerson & U. López de Heredia. 2017. ddradseqtools: a software package for *in silico* simulation and testing of double-digest RAD seq experiments. Molecular Ecology Resources 17:230–246.

Nguyen, L.-T., H.A. Schmidt, A. Haeseler & B.Q. Minh. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Molecular Biology and Evolution 32:268–274.

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

Techer, M.A., R.V. Rane, M.L Grau, J.M. Roberts, S.T. Sullivan, I. Liachko et al. 2019. Divergent evolutionary trajectories following speciation in two ectoparasitic honey bee mites. Communications Biology 2:357.

Manuscript received 15 May 2020, revised 24 July 2020.