JOURNAL OF ZOOLOGICAL SYSTEMATICS AND EVOLUTIONARY RESEARCH

On the shoulder of giants: Mitogenome recovery from non-targeted genome projects for phylogenetic inference and molecular evolution studies

Journal:	*Journal of Zoological Systematics and Evolutionary Research*
Manuscript ID	JZS.202000004.R2
Wiley - Manuscript type:	Original Article
Date Submitted by the Author:	n/a
Complete List of Authors:	Adrián Serrano, Silvia; Universitat de Barcelona Facultat de Biologia, Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals; Universitat de Barcelona Facultat de Biologia, Institut de Recerca de la Biodiversitat (IRBio) Lozano-Fernandez, Jesus; Instituto de Biologia Evolutiva, (CSIC-Universitat Pompeu Fabra); Universitat de Barcelona Facultat de Biologia, Departament de Biologia Evolutiva, Ecología i Ciències Ambientals; Universitat de Barcelona Facultat de Biologia, Institut de Recerca de la Biodiversitat (IRBio) Pons, Joan; Institut Mediterrani d'Estudis Avancats, (CSIC-UIB), Departament de Biodiversitat i Conservació Rozas, Julio; Universitat de Barcelona Facultat de Biologia, Departament de Genètica, Microbiologia i Estadística; Universitat de Barcelona Facultat de Biologia, Institut de Rechercha de la Biodiversitat (IRBio) Arnedo, Miquel; Universitat de Barcelona, Evolutionary biology, ecology and environmental sciences; Universitat de Barcelona, Institut de Rechercha de la Biodiversitat (IRBio)
Keywords:	Araneae, Island colonization, Gene rearrangement, tRNA secondary structure, High-throughput sequencing
Abstract:	The advent of high-throughput sequencing technologies (HTS) has generated an unprecedented amount of genomic and transcriptomic information. A vast amount of these data is not even used in targeted projects but is made available in public repositories. Previous studies have demonstrated that HTS data constitute a valuable resource to recover mitogenomic information, which is most relevant for studies of non-model and undersampled taxa. The spider family Dysderidae exemplifies well this situation: it is a highly diverse group, exceptionally well-suited for evolutionary and ecological research, but with a sparse mitogenomic record. In this study we used public and in-house HTS data to assemble and annotate at no cost 13 complete and 6 partial Dysderidae mitogenomes, as well as 34 partial mitogenomes belonging to other taxa within the Synspermiata clade, to which Dysderidae belongs. The mitogenomic information was further used to interrogate on a diverse array of evolutionary questions posed by the family. Phylogenetic inference clarified the evolutionary scenario of the colonization of the Canary Islands by the genus Dysdera, supporting two

	independent colonizations from the continent. Synteny analyses indicated that gene organization at the mitogenomic level is overall conserved within Dysderidae, the only exceptions being two cave-dwelling species, each of them with a unique putative transposition not described before in spiders. Finally, tRNA secondary structure reconstruction confirmed that the extreme compaction of tRNA is conserved across the family, suggesting that its origin could be traced back to approximately 100 million years ago (Mya). Altogether, this study demonstrates the potential of publicly available HTS data to conduct low-cost evolutionary research at different biological levels.

On the shoulder of giants: Mitogenome recovery from non-targeted genome projects for phylogenetic inference and molecular evolution studies

Silvia Adrián-Serrano[1], Jesus Lozano-Fernandez[1,2], Joan Pons[3], Julio Rozas[4], Miquel A. Arnedo[1]

[1]Departament de Biologia Evolutiva, Ecologia i Ciències Ambientals & Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain

[2]Institut de Biologia Evolutiva (CSIC-Universitat Pompeu Fabra), Passeig Marítim de la Barceloneta, 37-49, 08003 Barcelona, Spain

[3]Departament de Biodiversitat i Conservació, Institut Mediterrani d'Estudis Avançats (CSIC-UIB), Miquel Marquès, 21, 07190 Esporles, Illes Balears, Spain

[4]Departament de Genètica, Microbiologia i Estadística & Institut de Recerca de la Biodiversitat (IRBio), Universitat de Barcelona, Avinguda Diagonal 643, 08028 Barcelona, Spain

Corresponding author: Silvia Adrián, e-mail: silviaadrian@ub.edu. ORCID ID: https://orcid.org/0000-0002-1489-7854

Author contribution statement

MA and J-LF designed the study. JR contributed part of the raw data. SA conducted the bioinformatic, the gene annotation and secondary reconstruction and phylogenetic analyses under the supervision of J-LF, JP and MA, respectively. SA wrote the first draft of the paper. All authors contributed to the writing of the paper through discussions and additions to the text.

Conflict of Interest statement

The authors declare no conflicts of interest.

Key words: Araneae, island colonization, gene rearrangement, tRNA secondary structure, high-throughput sequencing.

ABSTRACT

The advent of high-throughput sequencing technologies (HTS) has generated an unprecedented amount of genomic and transcriptomic information. A vast amount of these data is not even used in targeted projects but is made available in public repositories. Previous studies have demonstrated that HTS data constitute a valuable resource to recover mitogenomic information, which is most relevant for studies of non-model and undersampled taxa. The spider family Dysderidae exemplifies well this situation: it is a highly diverse group, exceptionally well-suited for evolutionary and ecological research, but with a sparse mitogenomic record. In this study we used public and in-house HTS data to assemble and annotate at no cost 13 complete and 6 partial Dysderidae mitogenomes, as well as 34 partial mitogenomes belonging to other taxa within the Synspermiata clade, to which Dysderidae belongs. The mitogenomic information was further used to interrogate on a diverse array of evolutionary questions posed by the family. Phylogenetic inference clarified the evolutionary scenario of the colonization of the Canary Islands by the genus *Dysdera*, supporting two independent colonizations from the continent. Synteny analyses indicated that gene organization at the mitogenomic level is overall conserved within Dysderidae, the only exceptions being two cave-dwelling species, each of them with a unique putative transposition not described before in spiders. Finally, tRNA secondary structure reconstruction confirmed that the extreme compaction of tRNA is conserved across the family, suggesting that its origin could be traced back to approximately 100 million years ago (Mya). Altogether, this study demonstrates the potential of publicly available HTS data to conduct low-cost evolutionary research at different biological levels.

INTRODUCTION

The advent of high-throughput sequencing technologies (HTS), some fifteen years ago (Margulies et al., 2005), has had a major impact on life sciences. Unlike Sanger methods, this technology enables the parallel sequencing of multiple samples at much reduced time and costs (Kulski, 2016; Mardis, 2008). The rapid progress of HTS tools, combined with the emergence of sequencing service providers and a great development of bioinformatics, has fostered the generation of a large amount of non-targeted DNA sequences as part of Whole-Genome Sequencing (WGS) (Lam et al., 2012) and RNA sequencing (RNA-seq) (Wang, Gerstein, & Snyder, 2009) projects from a variety of non-model organisms (Kulski, 2016). HTS data generated in published research is usually made publicly available on on-line platforms, such as the National Center for Biotechnology Information (NCBI) (Geer et al., 2010). The increasing amount of genetic information available in these databases (Sayers et al., 2019) not only allows the validation of results through experimental reproducibility (Resnik & Shamoo, 2017), but it is also a valuable resource that can be "recycled" for its use in novel projects at no cost. In this regard, WGS and RNA-seq data may potentially include as a by-product a relevant yet usually overlooked piece of information, the mitochondrial genome (mitogenome, mtDNA).

Previous studies have shown that HTS data mining has a great potential for the retrieval of *de novo* mtDNA at a low cost, which is especially relevant in organisms with poor mitogenomic information (e.g. non model organisms) (Vieira & Prosdocimi, 2019). While the assembly of complete mitogenomes from WGS data has been proven relatively simple (Vieira & Prosdocimi, 2019), recovering mitochondrial data from RNA-seq data is more challenging. A common procedure during the generation of RNA-seq data is the target enrichment of nuclear messenger RNA (mRNA) sequences –which make up a tiny fraction of the total RNA– usually accomplished by means of poly-A enrichment treatments, which could hinder subsequent mitogenomic mining (Ozsolak & Milos, 2011). Nevertheless, available evidence suggests that polyadenylation is not exclusive of nuclear mRNA. Although with different functionality, some prokaryotes are also capable of adding poly-A tails, not only to mRNA, but also to transfer RNA (tRNA) and ribosomal RNA (rRNA) (Régnier & Marujo, 2013). The former observation in combination with the high abundance and expression levels of mitochondria within the cell (Pisani et al., 2013; Smith, 2016), and the use of proper assembly tools would favor the successful recovery of mitochondrial reads to a variable extent (Plese et al., 2019; Tian & Smith, 2016). In this regard, Plese et al. (2019)

have recently designed *Trimitomics*, a bioinformatic pipeline that simplifies and optimizes the assembly of mtDNA data from RNA-seq sequencing projects by combining freely available software. The pipeline, consisting in the sequential implementation of three assembly methods, namely *Novoplasty*, *Bowtie2* and *Velvet*, has been proven useful for the assembly of complete or partial mitogenomes (Plese et al., 2019), with a level of success depending on factors such as (1) the sequencing depth, (2) the evolutionary distance of the target taxa to the closer available mitogenome, which is used as a reference or seed, and (3) the presence and role of polyadenylation within the target taxa.

The mtDNA has been extensively used as a marker to trace evolutionary history in different fields of biology, including evolutionary biology (Carapelli, Liò, Nardi, Van Der Wath, & Frati, 2007), systematics (Carapelli, Fanciulli, Frati, & Leo, 2019), biogeography (Bauzà-Ribot et al., 2012), population genetics (Kllnç et al., 2018) or conservation (Rubinoff, 2006). In metazoans the mtDNA is haplotypic, generally assumed to lack recombination, and maternally inherited (although there are some remarkable exceptions, see Breton, Beaupré, Stewart, Hoeh, & Blier, 2007; Ladoukakis & Zouros, 2017). It encodes a relatively conserved set of 37 genes, including 13 protein coding genes (PCG), 22 tRNAs and 2 ribosomal subunits, alongside one or more control regions (Boore, 1999; Gissi, Iannelli, & Pesole, 2008).

Several features make the mtDNA data a valuable resource to resolve phylogenetic relationships and tackle a diverse array of evolutionary questions. On the one hand, its high copy number, unambiguous orthology and combination of faster and slower evolving regions make the mtDNA a relevant source of information for phylogenomic inference, especially at the lower taxonomic levels (Pisani et al., 2013). However, its higher substitution rate compared to nuclear genes—due to a greater mutation rate and smaller efficient population size— and strand-specific compositional bias are responsible for distortions in the phylogenetic signal, which may lead to poor or artifactual results when considering relationships at higher taxonomic levels (Pisani et al., 2013). On the other hand, mtDNA can be used to interrogate aspects such as the evolution of gene rearrangement and RNA secondary structure across taxa.

The typical secondary structure of metazoan tRNAs consists in a cloverleaf with four arms, namely the acceptor, the DHU, the anti-codon and the TΨC –except for tRNA-$S1_{GCT}$, that usually lacks a DHU-arm (Jühling, Pütz, Bernt, et al., 2012). Although mitochondrial tRNA sequences evolve at a high rate and can be quite variable –even between close

relatives– their secondary structure is usually conserved, since it determines the tertiary structure and function of these small molecules (Sakurai, Watanabe, Watanabe, & Ohtsuki, 2006). However, several studies have described atypical tRNAs, indicating that deviations from the canonical structure could be more common than previously thought. Non canonical tRNAs are characterized by a shortened encoding sequence which conforms a secondary structure lacking either the DHU arm, the TΨC arm (Masta & Boore, 2008; Wolstenholme, Macfarlane, Okimoto, Clary, & Wahleithner, 1987), or even both (Jühling, Pütz, Florentz, & Stadler, 2012; Pons, Bover, Bidegaray-Batista, & Arnedo, 2019). These atypical tRNAs apparently can fold in the canonical tertiary structure and become fully functional thanks to post-transcriptional editing (Wende et al., 2014).

In spite of constituting the most diverse biological group, arthropod mitogenomes are poorly represented in public databases, where the availability of mtDNA data is highly biased towards vertebrates (Geer et al., 2010). Spiders (order Araneae) exemplify well this pattern. Despite ranking sixth among the most diverse orders of animals (Roskov et al., 2019), with more than 48000 species, 4100 genera and 120 families (World Spider Catalog, 2019), there are only 45 complete spider mtDNA records at the organelle genome database at the NCBI (Wolfsberg, Schafer, Tatusov, & Tatusova, 2001) (Retrieved from https://www.ncbi.nlm.nih.gov/genome/organelle, accessed on July 2019).

The family Dysderidae C. L. Koch,1837 **(Fig. 1)** is a highly diverse group of araneomorph spiders, which currently comprises 564 species (World Spider Catalog, 2019), classified in 24 genera and 3 subfamilies, namely Dysderinae, Harpacteine and Rhodine (Deeleman-Reinhold & Deeleman, 1988). The family is an exceptionally well-suited model system for evolutionary research, and has been used to address evolutionary (Arnedo, Oromí, & Ribera, 2001; Bidegaray-Batista, Ferrández, & Arnedo, 2014; Macías-Hernández, Oromí, & Arnedo, 2008), ecological (Arnedo, Oromí, Múrria, Macías-Hernández, & Ribera, 2007; Macías-Hernández et al., 2018; Řezáč, Pekár, & Lubin, 2008) and biogeographic (Bidegaray-Batista & Arnedo, 2011; Bidegaray-Batista, Macías-Hernández, Oromí, & Arnedo, 2007; Macías-Hernández, Bidegaray-Batista, Emerson, Oromí, & Arnedo, 2013) questions.

Dysderidae has a Western Palearctic distribution, mostly restricted to the Mediterranean basin. Nonetheless, the species-rich genus *Dysdera* Latreille, 1804 (285 species) has managed to further colonize the Macaronesian archipelagos, a group of volcanic islands on the eastern North Atlantic Ocean, where it has undergone a major diversification process (Arnedo et al., 2001). Approximately fifty endemic species have been documented in the

Canary Islands (Macías-Hernández, López, Roca-Cusachs, Oromí, & Arnedo, 2016), providing one of the most extreme cases of island diversification within spiders world-wide (Arnedo, 2009), most likely representing a case of adaptive radiation (Arnedo et al., 2007). Former studies have revealed that most of the archipelago diversity is the result of local diversification processes following two to three independent colonization events (Arnedo et al., 2007, 2001). Canarian endemics form two well-defined clades correlating to the geographic location of the islands. The Eastern clade includes species from Lanzarote and Fuerteventura, the oldest and more eroded islands (closest to the African continent), while most of the endemic species from the central (Gran Canaria, Tenerife) and western islands (La Palma, La Gomera and El Hierro) belong to the Western clade. Finally, the eastern species *D. lancerotensis* Simon, 1907 is more closely related to continental species, and represents a recent independent colonization of the archipelago (Arnedo et al., 2001). However, previous studies have failed to clarify the relationship between the Eastern and Western clades, and the phylogenetic position of two western endemics, namely *D. sibyllina* Arnedo, 2007 and *D. andamanae* Arnedo & Ribera, 1997, remains unresolved, since some evidence suggests they could actually be more closely related to the Eastern clade (Arnedo et al., 2007).

The family Dysderidae is also a suitable model to study tRNA evolution. It is known that spiders have compacted tRNAs, exhibiting a three-arms only structure, with the presence of several miss-pairings and G-U pairs in the acceptor arm (Masta & Boore, 2008). Additionally, Pons et al. (2019) have recently reported extremely compacted tRNAs in the Dysderidae genera *Parachtes* Alicata, 1964 and *Harpactocrates* Simon, 1914, including the shortest arm-less tRNAs ever described in any organism. These extremely compacted structures have been preserved at least since the divergence of these two sister taxa, which dates back to 30 Mya (Bidegaray-Batista & Arnedo, 2011). This raises the question of whether the extreme compaction of these molecules is exclusive of these two genera. Alternatively, it could be shared by other members of the family or even other closely related families, which would trace its origin even further back in time.

Here, 1) we investigate the potential of HTS data generated in non-mitogenomic targeted studies to recover and assemble complete mitogenomes, and 2) we assess the potential of the recovered mitogenomic data to tackle a diverse array of evolutionary questions within the spider family Dysderidae, including a) phylogenetic relationships and

island colonization within Canarian *Dysdera*, and b) evolution of gene arrangement and tRNA secondary structure within the family Dysderidae.

MATERIALS & METHODS

Data acquisition

Mitogenome assembly was performed over two types of HTS data, genomic (WGS) and transcriptomic (RNA-seq), obtained from two different sources. First, we used HTS data generated from ongoing in-house projects, specifically 13 genomes and 9 transcriptomes belonging to 17 different Dysderidae species. The taxonomic sampling included 13 species of the genus *Dysdera*, 10 of them from the Canary Islands (8 from the western and central islands and 2 from the eastern ones); 2 species of *Parachtes*, which together with *Dysdera* belong to the Dysderinae subfamily; and one representative of each of the other two Dysderidae subfamilies, namely Harpacteinae and Rhodinae **(Appendix 1)**. Second, we downloaded HTS data from most genomic and transcriptomic records belonging to the spider clade Synspermiata, to which the Dysderidae family belongs (Michalik & Ramírez, 2014), and one of its sister lineages, the family Hypochilidae Marx, 1888 (Fernández et al., 2018; Garrison et al., 2016; Shao & Li, 2018), available at the NCBI database. The transcriptomic dataset consisted of 33 records, representing the family Hypochilidae and 13 out of the 17 Synspermiata families. Regarding genomic data, only one record, belonging to the family Sicariidae Keyserling, 1880 was available at the time this study was conducted **(Appendix 2)**. NCBI files were downloaded in SRA format, and converted to FASTQ format with fastq-dump from SRAtoolkit.2.9.2 (Leinonen, Sugawara, & Shumway, 2011). In addition, for the phylogenetic analyses we downloaded the 45 complete spider mitogenomes available at the NCBI mitogenome database **(Appendix 3)**.

For tRNA annotation and reconstruction we selected a subset of seven Dysderidae species spanning the largest evolutionary diversity within the family, namely the Dysderinae species *D. iguanensis* Wunderlich, 1987, *D. silvatica* Schmidt 1981 and *D. sibyllina*, from the western Canaries; *D. longa* Wunderlich, 1992, from the eastern Canaries, and the continental *D. asiatica* Nosek 1905, the Harpacteinae species *Stalagtia hercegovinensis* (Nosek, 1905) and the Rhodinae species *Parastalita stygia* (Joseph, 1882) **(Fig. 1)**. The species *Pholcus phalangiodes* J.K. Füssli, 1775, which is the closest complete mitogenome available, was also included for outgroup comparisons.

Mitogenome recovery

The assembly of mtDNA reads from genomic and transcriptomic data was performed following the *Trimitomics* pipeline (Plese et al., 2019) originally designed for optimizing the extraction of mitochondrial information from RNA-seq data, but also suitable for WGS. The *Trimitomics* workflow consists of three different methods, implemented stepwise depending on the success of the previous step. The first step was the assembly of raw reads against a full mitogenome, using different k-mer distributions, with the organelle assembler NOVOPlasty v3.2 (Dierckxsens, Mardulyn, & Smits, 2017). This step usually recovers complete mitogenomes from WGS data (Vieira & Prosdocimi, 2019), especially when sequences from closely related species are available to be used as a seed. In the second step the raw reads were cleaned from adapters with Trimmomatic 0.32 (Bolger, Lohse, & Usadel, 2014), mapped to a reference mitogenome with Bowtie2 (Langmead & Salzberg, 2012), and further assembled with Trinity 2.2.0 (Grabherr et al., 2011). Finally, the third step consisted in assembling either the whole transcriptome with Trinity or the whole genome with Velvet 1.2.10 (Zerbino & Birney, 2008) and subsequently extracting mitochondrial contigs with BLASTN (Altschul, Gish, Miller, Myers, & Lipman, 1990). When the complete mitogenome could not be recovered in a single contig, we manually combined the contigs obtained in all three steps with the help of the software Geneious v11.1.2 (Kearse et al., 2012) (**Fig. 2**).

Mitogenome annotation

PCGs and rRNAs were automatically annotated in both complete and partial recovered mitogenomes using the new MITOS2 version of Mitoswebserver (Bernt et al., 2013) (available at http://mitos2.bioinf.uni-leipzig.de/index.py), with default parameters. For the PCGs, the accuracy of the automatic annotation was assessed by checking the reading frame and manually curating to fine-tune the boundaries of the genes. We identified putative start and stop codons and corroborated them by aligning each gene sequence with the sequences of all spider mitogenomes available in NCBI. We used a similar comparative framework to determine the boundaries of the two rRNA sequences.

The annotation of tRNA genes was performed combining the results obtained with two different annotation programs, MITOS2 and ARAGORN (Laslett & Canback, 2004). To improve the detection, automatic annotation was performed using only sequences unassigned to protein coding genes as a template, allowing an overlapping of 50 base pairs (bp) (Morrison, 2010). The two programs were prone to both false positives and false negatives, and in combination were capable to annotate the full complement of tRNAs in only three

species (*D. longa*, *S. hercegovinensis* and *P. stygia*) (**Table S1**). Nevertheless, thanks to the overall conservation of the gene order and the anticodon arm sequence, we were able to manually identify and annotate all the missing tRNAs by aligning and searching for the anticodon motif.

Phylogenetic analyses

Because of the variable levels of success in recovering the full mitogenome among samples (from 10 to 100% of the total length) (see results), we decided to explore the impact of missing data on the phylogenetic inference. We built three different datasets of increasing amounts of missing data: (1) The first set, named S1, included all the complete mitogenomes downloaded from the NCBI database and mitogenomes in which we recovered over 70% of the total length from the in-house Dysderidae data (62 terminals, 0.8% of missing data in the matrix), (2) S2 included the data from S1 and remaining mitogenomes with over 30% recovered length (76 terminals, 9.4% of missing data), and (3) S3 included all the available mitogenomic data, independently of the recovery success (93 terminals, 22.63% of missing data) **(Fig. 2, Appendix 4 & Table S3)**. In addition, we investigated the impact of rRNA data in the phylogenetic inference by constructing datasets exclusively formed by PCGs and datasets including both the PCGs and the rRNAs for each of the three taxa datasets (S1, S2 and S3), adding up to 6 final matrices. Additional information on the final matrices are summarized in **Table 1**.

Alignments of the 13 PCGs were built with the *Translation alignment* option in Geneious. The ribosomal genes were aligned using the online version of the program MAFFT v. 7 (Katoh, Asimenos, & Toh, 2009, available at http://mafft.cbrc.jp). We tried different alignment strategies, and finally selected the G-INS-I algorithm, with default values, based on the similarity of the trees recovered to those obtained from the PCGs. Ambiguous or incomplete alignment ends of each gene were trimmed. Individual gene alignments were concatenated in a supermatrix using Geneious, and non-retrieved fragments were scored as missing data. Finally, we translated the "only PCGs" datasets S1 and S3 to perform analyses at the amino acid level.

Phylogenetic inference on the concatenated matrix was conducted under two different model-based approaches, maximum likelihood and Bayesian inference, to evaluate systematic error, e.g. the sensitivity of the results to changes in methodological assumptions (Wheeler, 1995). The best-fit partitioning scheme and nucleotide substitution models were selected with Partition Finder v2.1.1 (Lanfear, Frandsen, Wright, Senfeld, & Calcott, 2017),

assuming linked branch lengths under the Bayesian Information Criterion (BIC). Further details about the tested partitions are detailed in **tables S5 and S6**.

Maximum likelihood analyses were run with the software IQtree v1.6.2 (Nguyen, Schmidt, Von Haeseler, & Minh, 2015) with 1000 replicates of complete non-parametric bootstrap and edge equal partition model (Chernomor, Von Haeseler, & Minh, 2016). Bayesian inference was conducted using two different programs, MrBayes v3.2.6 (Ronquist & Huelsenbeck, 2003) and PhyloBayes mpi 1.8 (Lartillot, Rodrigue, Stubbs, & Richer, 2013). MrBayes analyses consisted of two Markov Chain Monte Carlo (MCMC) runs of 50 million generations sampling trees and parameters every 1000 generations, each with 8 chains and a "heating temperature" of 0.15. The burn-in was set to the 25% first generations and the supports of the clades were estimated as posterior probabilities. PhyloBayes analyses were run under the site-heterogeneous model CAT-GTR + Γ on both the nucleotide and amino acid complement of the PCGs datasets S1 and S3. The CAT model has been proven convenient for overcoming long-branch attraction (LBA) artifacts, and particularly useful when working with mitogenome data dominated by genes with a relatively high substitution rate and saturation (Lartillot, Brinkmann, & Philippe, 2007; López-López & Vogler, 2017; Talavera & Vila, 2011; Timmermans et al., 2016). The analyses consisted of two independent runs of 10000 cycles. Convergence was assessed using the maximum difference in the bipartitions of the chains, as reported by bpcomp, and assessing the effective sample sizes (ESS) and relative differences for all parameters with tracecomp, both programs part of the PhyloBayes package (Convergence statistics are reported in **Figs. S14-S17**). Although all PhyloBayes analyses converged in the tree space (bpcomp), few parameters in the amino acid analyses showed EES values below 50 (tracecomp). Nevertheless, it has been shown that convergence of the CAT chains towards the correct topology is rapid, and does not require ESS > 50 for every parameter if the topology of the chains is almost identical (post-publication material Simion et al., 2017). Based on this suggestion, we assessed the resulting topology of each chain individually finding nearly indistinguishable topologies.

Phylogenetic analyses were run remotely at the CIPRES Science Gateway (Miller, Pfeiffer, & Schwartz, 2010) and on a local cluster. The trees obtained were visualized with FigTree v1.4.3 (Rambaut, 2016).

Reconstruction of the tRNA structure

The reconstruction of the tRNA secondary structure was first performed automatically with MITOS2 and with ARAGORN. Nevertheless, in several cases these programs were not

able to generate coherent reconstructions (**Table S1**), either because the reconstructions failed to meet the structural requirements or because motifs were not conserved across taxa. In the former cases, we used manual reconstruction by identifying candidate structures that were tested in a comparative framework, using a multiple alignment that included the well annotated sequences of the closely related genera *Parachtes* and *Harpactocrates* available from Pons et al. (2019). Candidate structures for a given tRNA included: (1) suitable structures found with the automatic algorithms for any of the species, (2) secondary structures inferred by Pons et al. (2019) for *Parachtes* and *Harpactocrates;* and (3) structures manually inferred based on conserved complementary DNA sequence regions that could correspond to arms in either canonical, three-armed, o arm-less structures. Finally, when possible, we selected the most optimal and conserved structure across all the species.

RESULTS

Mitogenome recovery

Mitogenome recovery performance with the *Trimitomics* pipeline was variable and dependent upon the taxon and the source and type of sequence data **(Fig. 3, Appendix 4 & Tables S2-S3)**.

The use of NOVOPlasty, the first step in the *Trimitomics* pipeline, already recovered the complete mitogenomes from all the 13 Dysderidae genomes sequenced at low coverage in the lab. Six of them were circularized, while the rest resulted in almost complete mitogenomes after concatenation of 2 to 4 long assembled contigs with Geneious. We tried different k-mer values (25, 39, 45 and 51), 39 being the optimal value in most cases. We used the mitogenomes of *P. teruelis* (Kraus, 1955)—NCBI accession MN052921—and *P. ignavus* (Simon, 1882)—NCBI accession MN052920—as a seed for the first round of analyses, which worked well for most samples. The circularized mitogenome of *D. asiatica* obtained in the first round was further used in a second round as a seed for three samples (*D. sibyllina, D. silvatica and D. yguanirae* Arnedo & Ribera, 1997) that were not fully recovered in the first place. The only problematic sample of this set was *D. sibyllina*, for which the combination of two long contigs obtained in different analyses generated an ambiguous region affecting the *atp6* and *coxIII* genes, which was coded as missing data. Because of format incompatibility errors, the NOVOPlasty analysis could not be conducted on the only genomic sample downloaded from NCBI. On the other hand, NOVOPlasty performed poorly with transcriptomic data. We tested this software with the transcriptomes of three *Dysdera* species from different sources (NCBI and in-house), and the results were sparse even in the

best-case scenario. For instance, we failed to recover the mitogenome of *D. gomeriensis* from transcriptomic data even when using as a seed the full mitogenome of the same species, previously assembled from genomic data.

The second step, consisting in the serial combination of *Trimmomatic*, *Bowtie2* and *Trinity*, managed to recover more than 90% of the total mitogenomic data (including all PCGs and rRNAs) from in-house deep coverage transcriptomic data (>200 millions of reads) by using mitogenomes from the same species as reference, which had been previously assembled from genomic data. The output consisted of many small contigs (≈ 200-2000nt in length) that were manually checked and selected by aligning them against a reference mitogenome. After trimming PolyA tails, we assembled the good quality contigs with Geneious. Nevertheless, we discarded these data for subsequent phylogenetic analyses, since we had already recovered the complete mitogenomes from genomic data of the same species. The performance of the *Trimmomatic*, *Bowtie2* and *Trinity* pipeline with the rest of the transcriptomic data and the NCBI genome was poor, with recovered lengths ranging from 0 to 17%, and most of them below 5%. This final recovered length was the combination of the results obtained by mapping the cleaned reads using five different reference mitogenomes, including all the Synspermiata genera available in our dataset, namely *Dysdera* and *Parachtes* (family Dysderidae), *Pholcus* Walckenaer, 1805 and *Mesavolivar* González-Sponga, 1998 (family Pholcidae), and the outgroup *Hypochylus* Marx, 1888 (family Hypochilidae). This approach was implemented to improve the recovery, as many of the taxa were too distantly related to all the available reference sequences.

The third step of the *Trimitomics* pipeline was the most time-consuming, as it required the *de novo* assembly of the complete transcriptomes. Once again, we combined the results obtained using five different taxa as a seed to improve the recovery performance. The success rate was again variable, but significantly better than in the previous step **(Fig. 3 & Table S2)**. In general, the closer the taxon was to the seed mitogenomes, the more complete were the results.

The combination of all of the contigs obtained in the different steps of the *Trimitomics* pipeline slightly improved the results for some of the samples, since each method recovered a different region of the mitogenome. Three taxa—*Filistata insidiatrix* (Forskoel, 1775), *Hypochilus gertschi* Hoffman, 1963, and *Opopaea* sp.—were discarded for further analysis, as the output contigs turned out to be contamination and it was not possible to safely assign any of the sequences to the corresponding taxa **(Fig. 3, Appendix 4 & Tables S2-S3).**

Finally, we annotated the PCGs and rRNAs in all mitogenomes and checked the PCG reading frames. Surprisingly, we detected two nucleotide insertions affecting the reading frame of the *atp6* of *D. sibyllina*. These insertions were close to the ambiguous region previously discarded. Multiple alignment with all available *Dysdera atp6* genes showed that the rest of the sequence was very conserved, indicating that they were most probably sequencing errors. Therefore, we decided to manually remove both insertions. We did not attempt to annotate the long AT-rich control region, which is located between genes *trnQ* and *trnM* (Pons et al., 2019). This region contains high numbers of short repeated units (Lunt, Whipple, & Hyman, 1998; Wolstenholme, 1992) causing automatic assemblers perform poorly. The only way to circumvent this problem is by validating the assembled sequence either by Sanger or long-read (i.e. PacBio) sequencing (Pollard, Gurdasani, Mentzer, Porter, & Sandhu, 2018). Therefore, the sequence obtained for the AT-rich control region should be regarded with caution, and we decided not to use it in downstream analyses.

Phylogenetic inference

For all PCG matrices, the best partition scheme consisted in 39 partitions (by gene and codon position), while for the PCG+rRNA matrices the preferred scheme consisted in 40 partitions (by gene and codon position in PCG plus a single partition for both rRNA genes). Further information on partition scores are summarized in **Tables S5-S6**.

The trees inferred for each dataset (S1, S2 and S3) under different analytical procedures (Maximum likelihood and Bayesian inference with MrBayes) and character sets (with or without rRNA data), showed high topological congruence. Trees and supports of these alternative analyses are summarized in **Fig. 4, Fig. 5 & Figs. S1-S13**. Overall, Bayesian trees showed higher support values, while there were no remarkable differences in the supports obtained from data with or without rRNA information. The trees derived from the datasets with lower level of missing data, S1 (**Fig. 4)** and S2 (**Fig. S1**) (**Table 1**), were better resolved, most of the nodes obtained high support by both inference methods, and most taxonomic groups were recovered monophyletic. However, analyses including rRNA data, most remarkably S2, failed to recover the sister relationship between the superfamily Scytodoidea and the Lost Tracheae Clade, supported by previous phylotranscriptomic studies (Fernández et al. 2018). The trees obtained from the matrix with the highest proportion of missing data (S3) (**Fig. 5**) also supported the monophyly of all the families, but above family-level relationships were less resolved, particularly in clades that involved the taxa with higher levels of missing data. Additionally, the position of some taxa in S3 trees was conflictive,

albeit with low support: neither the superfamily Scytodoidea nor the Lost Tracheae Clade were recovered as monophyletic due to the position of *Calponia harrrisonfordi* Platnick, 1993, *Ochyrocera* sp. Simon, 1892, and the clade conformed by *Drymusa* sp. Simon, 1892 and *Periegops suteri* (Urquhart, 1892); Synspermiata was recovered as paraphyletic due to the position of *Kukulcania hibernalis* Hentz, 1842 as sister group of the superfamily Dysderoidea; and, surprisingly, the three individuals of *Pholcus phalangioides* were recovered as paraphyletic with high support. When considering all three datasets, some relationships were ambiguous at the genus level, that is the case of *Epeus* Peckham & Peckham, 1886, *Telamonia* Thorell, 1887 and *Plexipus* C. L. Koch, 1896, within the family Salticidae Blackwall, 1841, and *Araneus* Clerck, 1757, *Neoscona* Simon, 1864 and *Argiope* Audouin, 1827 within the family Araneidae Simon, 1895. In this last case, the different matrices recovered conflicting relationships: S1 recovered *Argiope* as sister to *Neoscona*, while the rest of the analyses allied *Argiope* and *Araneus*, both relationships supported in some Bayesian analyses.

Relationships within the family Dysderidae (**Fig. 6**) were congruent and well resolved under all analytical conditions. The subfamilies Rhodinae, represented by *Parastalita stygia*, and Harpacteine, represented by *Stalagtia hercegovinensis*, were supported as sister lineages in most of the trees, and the Dysderinae subfamily, the genera *Parachtes* and *Dysdera*, and the Canarian *Dysdera* species were recovered as monophyletic. Finally, the relationships within the Canarian clade were congruent, fully resolved and highly supported in all trees, and confirmed the paraphyly of the western group, since *D. sibyllina* was supported as sister to the Eastern clade.

The trees obtained with PhyloBayes were congruent and showed no obvious differences with those obtained with IQtree and Mr. Bayes (**Figs. S14-S17**).

Mitogenomic organization

All recovered mitogenomes coded for the standard 13 PCGs and 2 rRNAs commonly found in most metazoan species. Similarly, the 8 species in which the tRNA complement was fully annotated, showed the standard set of 22 tRNAs**.** The plus strand encoded 22 genes (9 PCGs and 13 tRNAs), while the minus strand encoded 15 genes (4 PCGs, 9 tRNAs, and 2 rRNAs) **(Fig. 7)**. The gene arrangement in *Dysdera, Harpactea* and *Stalagtia* was the same recently described in *Parachtes* (Pons et al., 2019), except for the cave-dwelling species *D. sibyllina*, which seems to have suffered a translocation involving the region that includes *trnN*, *trnA*, *trnS1*, *trnR* and *trnE* **(Fig. 8)**, and *S. hercegovinensis*, with a translocation

affecting the *trnL2*. While *trnL2* is located immediately after *nad3* and before a short control region in most species (Pons et al., 2019), in *S. hercegovinensis* it is placed after this control region and immediately before the *trnN*. As previously reported by Pons et al. (2019), the typical arrangement found in Dysderidae is derived, and differs from the closely related Pholcidae in the position of the *trnI*, constituting a putative synapomorphy for the family.

Reconstruction of the tRNA

Only a handful of the secondary structures inferred by automatic algorithms met the requirements to be considered candidate structures **(Table S1)**, most likely as a result of the lack of conservation of the typical cloverleaf structure across spiders. Although MITOS2 was more prone to find three-armed or armless structures than ARAGORN, it usually generated aberrant structures. After manually assessing the alternative candidate structures for each tRNA, we found that none of them could be fold in a cloverleaf secondary structure that was conserved among all the species. Ten tRNAs showed a three-armed structure, eight of them lacked the TΨC arm (D, F, G, H, I, L1, T and W), and the two others lacked the DHU arm (Q, S2). Four tRNAs were found to be armless (A, R, S1 and Y), as we did not find any structure with conserved DHU or TΨC arms. Finally, for the remaining eight tRNAs (C, E, K, L2, M, N, P and V), we found two possible structures, namely arm-less or TΨC-less **(Fig. 9 & Figs. S18a-S18g)**. In the latter cases we were not able to choose among both alternatives because the TΨC-less structure showed a very conserved DHU arm, while the acceptor arm was more conserved in the armless option **(Fig. 10)**. Nevertheless, the TΨC-less structure seems to be more plausible, since it is more similar to the canonical structure. Overall, the 22 tRNA secondary structures inferred were very conserved not only within *Dysdera*, but also within Dysderidae **(Fig. 10)**, albeit with some exceptions—e.g. Both tRNA-T and tRNA-W show a TΨC arm conserved between *Stalagtia* and *Parastalita* that was not recovered in *Dysdera* **(Figs. S18a-S18g)**.

DISCUSSION

Mitogenome recovery

In this study we used HTS data to assemble thirteen complete and forty partial mitogenomes, the latter with a wide range of recovery success (from 98% to 10% of the total length). Our results confirm that HTS data, including low coverage WGS and RNAseq, is a rich source of mitogenomic information. However, the efficiency of the recovery depends on several factors, including (1) the data type (e.g. genomic or transcriptomic), (2) the

phylogenetic distance of the seed sequence to the mined specimens, (3) the sequencing coverage (i.e. depth) and (4) the data source. With regard to the data type, as expected, WGS data yielded the best results. Complete mitogenomes were recovered from all but one of the genomic samples. Conversely, the recovery success with transcriptomic data was variable and relied heavily on additional factors (see below). Concerning the phylogenetic distance of the mined specimen to the seed, the recovery success was greater when the mitogenome from a closely related species was available to use as a seed. This was the case of the species belonging to Dysderidae, Pholcidae and Hypochilidae families, for which mitogenomes of the same genus/family were available, **(Fig. 3, Appendix 4 & Tables S2-S3)**. As regards the sequencing depth, the data coverage seemed largely irrelevant when using WGS data. However, it did have a major impact on the performance of the second step of *Trimitomics* when using transcriptomic data. Transcriptomes with deeper coverage (> 200 million reads) yielded near complete mitogenomes (around 95%), while recovery success dropped below 17% when using lower coverage data (< 50 million reads). Finally, the mitogenome recovery success also seemed dependent on the original lab source of the HTS data **(Table S2)**. The explanation is most probably related to the use of alternative methodologies and protocols to generate the data—i.e. mRNA capture, sequencing kits, etc…—which would determine the amount of mitochondrial information present in the raw reads.

The most successfully recovered mitogenome regions from RNA-seq data were the PCGs and rRNAs, while the tRNAs and intergenic regions were mostly absent. The recovery profile agrees with the pattern of mitochondrial expression found by Plese et al. (2019) in arthropods, with a higher expression of proteins of the respiratory complex IV (COXI, COXII and COXIII), ATP6 and ATP8—which were usually recovered as a single contig—and the two ribosomal genes **(Table S3)**.

Phylogenetic relationships

Mitogenomic data successfully resolved relationships at most phylogenetic levels, especially when the proportion of missing data was minimized. As expected, the higher the level of missing data, the poorest the phylogenetic resolution, including low supports and artifactual placement of taxa. The impact of the rRNA data on the results also seemed to be related to the amounts of missing data. rRNA data improved resolution when missing data levels were low, but it had a negative effect when adding partial mitogenomes to the matrix. This was most probably a methodological artifact resulting from the mitogenome mining. Unlike the PCGs, which quality can be assessed by aligning and checking their reading

frame, assessing the quality of the recovered rRNA is challenging. Also, many of the output contigs recovered as rRNAs with *Bowtie2* consisted in single short fragments that could have been easily misassigned.

Our results (**Figs. 4-5 & Fig. S1**) were overall congruent with those obtained in recent studies aimed to resolve phylogenetic relationships within spiders, using either a Sanger target gene approach (Wheeler et al., 2017) or comparative transcriptomics (Fernández et al., 2018). As exceptions to that general agreement, all our mitogenomic trees recovered the sister relationship between the Mygalomorph families Dipluridae (curtain-web spiders) and Theraphosidae (tarantula spiders), while previous studies had recovered (Wheeler et al., 2017) or robustly supported (Fernández et al., 2018) Theraphosidae as sister to the trap-door family Nemesidae. Similarly, the lynx spiders (Oxyopidae) were supported as sister to the wolf spiders (Lycosidae). This relationship was recovered with low support by Wheeler et al. (2017), but rejected by transcriptomics, which supported Oxyopidae as sister to the crab spiders (Thomisidae) (Fernández et al., 2018). The tree with higher levels of missing data (S3) (**Fig. 5**) recovered the crevice-weaver Filistatidae as sister to the superfamily Dysderoidea, while most previous studies supported its sister relationship to the lampshade spiders Hypochilidae (Fernández et al., 2018; Wheeler et al., 2017). Some of the mitogenomic trees failed to recover the sister relationship between Scytodoidea and the Lost Tracheae Clade (Fernández et al., 2018; Wheeler et al., 2017). The position of Caponiidae was also conflictive, it is included within Scytodoidea in this study, but was previously recovered (Wheeler et al., 2017) or supported (Fernández et al., 2018) as sister to Dysderoidea. Finally, the supported paraphyly of *Pholcus phalangioides in our trees* is most likely the result of a specimen misidentification.

Some relationships at the family level within the superfamily Dysderoidea were also ambiguous, most likely due to an incomplete taxon sampling and high levels of missing data. Previous studies agreed in supporting Segestriidae as the sister taxa to the remaining families, (Fernández et al., 2018; Wheeler et al., 2017). However, most mitogenomic trees in this study supported Segestriidae as sister to Oonopidae. In addition, within Dysderidae, phylotranscriptomics (Fernández et al., 2018) supported Harpacteinae as sister to Dysderinae, while our mitogenomic trees supported Harpacteinae as sister to Rhodinae

Ambiguous relationships between the mitogenomic trees obtained from matrices with different levels of completeness are most likely the result of the increasing amounts of missing data. Conversely, conflictive relationships recovered across all our matrices are most probably due to either poor taxon sampling or phylogenetic artifacts derived from the

accelerated rate of mitochondrial genome evolution (Pisani et al., 2013). The congruent results obtained using the more complex model CAT, in PhyloBayes (**Figs S14-S17**) prove that the supported nodes in our trees are robust and not sensitive to the use of alternative models or inference methods. Increasing the taxonomic sampling will certainly improve our results by shortening long branches and improving alignments (i.e. homology statements), which may avoid long branch attraction problems. But dealing with compositional and mutation biases will require the development of more realistic and sophisticated evolutionary models for mitochondrial evolution.

The colonization of the Canary Islands by Dysdera *spiders*

The mitogenomic data fully resolved, with high support, the phylogenetic relationships within the genus *Dysdera*, providing insights into the evolutionary scenario for the colonization of the Canary Islands by this group **(Fig. 6)**. All trees supported the monophyly of Canarian species, suggesting a single origin for the Eastern and the Western species groups, relationship that had been contentious in former studies (Arnedo et al., 2007, 2001; Macías-Hernández, Bidegaray-Batista, Oromí, & Arnedo, 2012; Macías-Hernández et al., 2008). These results support the hypothesis of the double colonization of the archipelago, first by the ancestor of most Canarian endemics and later by the ancestor of the single endemic *D. lancerotensis* (Arnedo et al., 2001; Bidegaray-Batista et al., 2007). Additionally, our results reject the monophyly of the species from the western islands, since the western species *D. sibyllina* was recovered as sister to the Eastern clade. This result was already advanced by Arnedo et al. (2007) based on two mitochondrial genes and morphological data. Under the new topology, the most parsimonious scenario is that the eastern Canaries were colonized from western-central islands ancestors. Alternatively, and taking into consideration the older age of the eastern islands, the western-central islands would have been colonized twice independently from the eastern ones. A more thorough sampling of Canarian endemics would be required to decide among these two alternative scenarios.

Mitochondrial rearrangements in Dysderidae

The length of the circularized mitogenome contigs ranged from 14150 to 15037 bp **(Table S4)**, being similar to the average size of other spider mitogenomes available at the NCBI (14330 ± 240 bp). This size is among the smallest reported within arthropods, and it is most likely the result of the great compaction, involving a reduction of the non-coding intergenic regions and slight gene overlapping (Pisani et al., 2013; Pons et al., 2019).

We detected three different synteny arrangements within Dysderidae, one in *S. hercegovinensis,* one in *D. sibyllina*, and one shared across the rest of the dysderids included. The mitochondrial gene organization found in most Dysderidae specimens analyzed here agreed with that described by Pons et al. (2019) for the Dysderidae genera *Parachtes* and *Harpactocrates* **(Fig. 7 & Fig. 8)**. This typical Dysderidae arrangement differs from that observed in the remaining araneomorph spiders—including Pholcidae (*P. phalangioides*), the closest family with mitogenomic information available—in the position of *trnI* (Pons et al., 2019). On the other hand, the gene arrangements found in *D. sibyllina and S. hercegovinensis*, involving the transposition of the tRNA group t*rnN*, *trnA*, *trnS1*, *trnR* and *trnE*, and *trnL2* respectively, are described here for the first time **(Fig. 8)**.

As already stated by other authors (Pons et al., 2019; Stokkan, Jurado-Rivera, Juan, Jaume, & Pons, 2016), the generation of new gene arrangements in metazoans at the lower taxonomic levels seems to be primarily driven by the transposition of tRNAs, with PCG and rRNAs remaining in the same relative position. Particularly, the most common rearrangements in arthropod mtDNAs consist on translocations of tRNA genes surrounding the AT-rich Control Region or the cluster *trnA-trnR-trnS1-trnE-trnF* (Gissi et al., 2008). The latter is related to the presence of a short non coding region considered as the putative minor strand origin of replication (Boore, 1999; Pons et al., 2019)

tRNA detection and reconstruction

Our results revealed that the compacted tRNA secondary structures found by Pons et al. (2019) in the Dysderinae genera *Parachtes* and *Harpactocrates* are conserved across the Dysderidae family **(Fig. 9, Fig. 10 & Fig. S18)**. Based on the current estimates of the split of Dysderidae from its sister taxa (Fernández et al., 2018; Magalhaes, Azevedo, Michalik, & Ramírez, 2019) we suggest that the origin of the extreme tRNA compaction in these spiders could trace back to approximately 100 million years ago.

As anticipated, the identification and, particularly, the reconstruction of the secondary structure of Dysderidae tRNAs was a very time consuming and difficult process, even with the help of available mitochondrial searching algorithms specifically designed for tRNAs (Morrison, 2010). Most spider tRNAs are atypical, lacking one or both DHU and TΨC arms and bearing several G-T wobble pairs and miss-pairings in the acceptor arm (Masta & Boore, 2008). Since searching algorithms are designed to detect tRNAs with the canonical cloverleaf structure (Masta & Boore, 2008), they failed to identify and correctly reconstruct most Dysderidae tRNAs (**Table S1**). The automatic annotation retrieved different and divergent

tRNA structures across closely related taxa, which required that the conserved secondary structures had to be mostly manually constructed. In this context, a full comparative analysis based on the multiple alignment of closely related sequences, has been proven to be the most convenient annotation strategy for the mtDNA tRNAs (Morrison, 2010). Additionally, as already pointed out in Pons et al. (2019), the development of covariation models that account for atypical arm-less structures and multiple mispairings in the acceptor arm would greatly ease the retrieval of atypical structures, reducing the impact of subjective decisions made as a result of the manual inference.

CONCLUSIONS

Our results confirm that HTS data are a rich and cost-efficient resource to recover mtDNA information. However, the data mining efficiency depends on several factors, including data type, the phylogenetic relatedness of the seed, the sequencing coverage and the data source. The ability of mtDNA data to resolve phylogenetic questions is sensitive to both the amount of missing data and the taxon sampling but may provide information at multiple phylogenetic levels. In this study, we assembled thirteen complete and forty partial mitogenomes belonging to the spider clade Synspermiata, that helped us to investigate a diverse array of evolutionary questions posed by the family Dysderidae. On the one hand, our analyses confirm the monophyly of the endemic Canary Islands species of the genus *Dysdera*—with the exception of *D. lancerotensis*—supporting two independent colonizations of the archipelago. We also revealed that western-central islands species are paraphyletic with regard to the eastern Canarian species, suggesting a colonization from the former islands to the later. Alternatively, and taking into account the older age of the eastern Canaries, the western-central islands could have been colonized twice from the eastern ones. On the other hand, our results indicate that the mitogenomic gene arrangement is mostly conserved across the Dysderidae family, with the only exceptions of two cave-dwelling species, *S. hercegovinensis* and *D. sibyllina*, each of them with a unique ordering described here for the first time. Finally, we showed that the armless tRNA secondary structures are conserved across the family Dysderidae, suggesting that they could have originated at least as early as the family split from its sister group, estimated approximately 100 Mya.

Increasing the taxon sampling will further help to clarify the phylogenetic structure of the whole family Dysderidae and, specifically, of the remarkable radiation of *Dysdera* in the Canarian archipelago, and to more accurately map the origin and trends of mtDNA compaction within spiders.

ACKNOWLEDGEMENTS

We would like to acknowledge the following colleagues that provided specimens for sequencing: Adrià Bellvert, Luís C. Crespo, Salva de la Cruz, Carla Díaz, Nuria Macías-Hernández, Pedro Oromí, Martina Pavlek, Tin Rožman. We are in debt with Alejandro Sánchez-Gracia, Jose Francisco Sánchez-Herrero and Joel Vizueta for sharing genome and RNAseq information of several species of Canarian *Dysdera*.

ABBREVIATIONS

atp6 and *atp8*: genes for ATP synthase subunits 6 and 8

cob: gene for cytochrome b

cox1, *cox2* and *cox3*: genes for cytochrome c oxidase subunits 1, 2 and 3

nad1, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5* and *nad6*: mitochondrial genes for NADH dehydrogenase subunits 1-6 and 4L

rrnS and *rrnL*: genes for small (12S) and large (16S) subunits of ribosomal RNA

genes for transfer RNAs: *trnA* (alanine), *trnC* (cysteine), *trnD* (aspartic acid), *trnE* (glutamic acid), *trnF* (phenylalanine), *trnG* (glycine), *trnH* (histidine), *trnI* (isoleucine), *trnK* (lysine), *trnL1* (leucine anticodon NAG), *trnL2* (leucine anticodon YAA), *trnM* (methionine), *trnN* (asparagine), *trnP* (proline), *trnQ* (glutamine), *trnR* (arginine), *trnS1* (serine anticodon NCU), *trnS2* (serine anticodon NGA), *trnT* (threonine), *trnV* (valine), *trnW* (tryptophan), *trnY* (tyrosine)

FUNDING

This work was funded by the Ministry of Economy and Competitiveness through projects CGL2016-80651-P and the Catalan Government through grant 2017SGR83 (to MA). JL-F was supported by a Juan de la Cierva Fellowship (FJCI-2015-23723). Additional funding was provided by the Systematics Association with a Systematic Research Fund to JL-F.

DATA AVAILABILITY

Mitogenome sequences, mitogenome annotations, data matrices and trees are archived online: https://doi.org/10.5281/zenodo.3763320

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. http://doi.org/10.1016/S0022-2836(05)80360-2

Arnedo, M. A. (2009). Spiders. In R. G. Gillespie & D. A. Clague (Eds.), *Encyclopedia of Islands* (pp. 861–865). Berkeley: University of California Press.

Arnedo, M. A., Oromí, P., Múrria, C., Macías-Hernández, N., & Ribera, C. (2007). The dark side of an island radiation: Systematics and evolution of troglobitic spiders of the genus Dysdera Latreille (Araneae:Dysderidae) in the Canary Islands. *Invertebrate Systematics*, *21*(6), 623–660. http://doi.org/10.1071/IS07015

Arnedo, M. A., Oromí, P., & Ribera, C. (2001). Radiation of the spider genus Dysdera (Araneae, Dysderidae) in the Canary Islands: Cladistic assessment based on multiple data sets. *Cladistics*, *17*(4), 313–353. http://doi.org/10.1006/clad.2001.0168

Bauzà-Ribot, M. M., Juan, C., Nardi, F., Oromí, P., Pons, J., & Jaume, D. (2012). Mitogenomic phylogenetic analysis supports continental-scale vicariance in subterranean thalassoid crustaceans. *Current Biology*, *22*(21), 2069–2074. http://doi.org/10.1016/j.cub.2012.09.012

Bernt, M., Donath, A., Jühling, F., Externbrink, F., Florentz, C., Fritzsch, G., … Stadler, P. F. (2013). MITOS: Improved de novo metazoan mitochondrial genome annotation. *Molecular Phylogenetics and Evolution*, *69*(2), 313–319. http://doi.org/10.1016/j.ympev.2012.08.023

Bidegaray-Batista, L., & Arnedo, M. A. (2011). Gone with the plate: The opening of the Western Mediterranean basin drove the diversification of ground-dweller spiders. *BMC Evolutionary Biology*, *11*(1), 317. http://doi.org/10.1186/1471-2148-11-317

Bidegaray-Batista, L., Ferrández, M. Á., & Arnedo, M. A. (2014). Winter is coming: Miocene and Quaternary climatic shifts shaped the diversification of Western-Mediterranean Harpactocrates (Araneae, Dysderidae) spiders. *Cladistics*, *30*(4), 428–446. http://doi.org/10.1111/cla.12054

Bidegaray-Batista, L., Macías-Hernández, N., Oromí, P., & Arnedo, M. A. (2007). Living on the edge: Demographic and phylogeographical patterns in the woodlouse-hunter spider Dysdera lancerotensis Simon, 1907 on the eastern volcanic ridge of the Canary

727 Islands. *Molecular Ecology*, *16*(15), 3198–3214. http://doi.org/10.1111/j.1365-
728 294X.2007.03351.x

729 Bolger, A. M., Lohse, M., & Usadel, B. (2014). Trimmomatic: A flexible trimmer for
730 Illumina sequence data. *Bioinformatics*, *30*(15), 2114–2120.
731 http://doi.org/10.1093/bioinformatics/btu170

732 Boore, J. L. (1999). Animal mitochondrial genomes. *Nucleic Acids Research*, *27*(8), 1767–
733 1780. http://doi.org/10.1093/nar/27.8.1767

734 Breton, S., Beaupré, H. D., Stewart, D. T., Hoeh, W. R., & Blier, P. U. (2007). The unusual
735 system of doubly uniparental inheritance of mtDNA: isn't one enough? *Trends in*
736 *Genetics*. http://doi.org/10.1016/j.tig.2007.05.011

737 Carapelli, A., Fanciulli, P. P. P., Frati, F., & Leo, C. (2019). Mitogenomic data to study the
738 taxonomy of Antarctic springtail species (Hexapoda: Collembola) and their adaptation
739 to extreme environments. *Polar Biology*, *42*(4), 715–732.
740 http://doi.org/10.1007/s00300-019-02466-8

741 Carapelli, A., Liò, P., Nardi, F., Van Der Wath, E., & Frati, F. (2007). Phylogenetic
742 analysis of mitochondrial protein coding genes confirms the reciprocal paraphyly of
743 Hexapoda and Crustacea. *BMC Evolutionary Biology*, *7*(SUPPL. 2).
744 http://doi.org/10.1186/1471-2148-7-S2-S8

745 Chernomor, O., Von Haeseler, A., & Minh, B. Q. (2016). Terrace Aware Data Structure for
746 Phylogenomic Inference from Supermatrices. *Systematic Biology*, *65*(6), 997–1008.
747 http://doi.org/10.1093/sysbio/syw037

748 Darty, K., Denise, A., & Ponty, Y. (2009). VARNA: Interactive drawing and editing of the
749 RNA secondary structure. *Bioinformatics*, *25*(15), 1974–1975.
750 http://doi.org/10.1093/bioinformatics/btp250

751 Deeleman-Reinhold, C. L., & Deeleman, P. R. (1988). Revision des Dysderinae (Araneae,
752 Dysderidae), les espèces mediterranéennes occidentales exceptées. *Tijdschrift Voor*
753 *Entomologie*, *131*, 141–269. Retrieved from http://biostor.org/reference/57263

754 Dierckxsens, N., Mardulyn, P., & Smits, G. (2017). NOVOPlasty: De novo assembly of
755 organelle genomes from whole genome data. *Nucleic Acids Research*, *45*(4), e18.
756 http://doi.org/10.1093/nar/gkw955

757 Fernández, R., Kallal, R. J., Dimitrov, D., Ballesteros, J. A., Arnedo, M. A., Giribet, G., &

Hormiga, G. (2018). Phylogenomics, diversification dynamics, and comparative transcriptomics across the spider tree of life. *Current Biology*, *28*(9), 1489–1497. http://doi.org/https://doi.org/10.1016/j.cub.2018.03.064

Garrison, N. L., Rodriguez, J., Agnarsson, I., Coddington, J. A., Griswold, C. E., Hamilton, C. A., … Bond, J. E. (2016). Spider phylogenomics: untangling the Spider Tree of Life. *PeerJ*, *4*, e1719-35. http://doi.org/10.7717/peerj.1719

Geer, L. Y., Marchler-Bauer, A., Geer, R. C., Han, L., He, J., He, S., … Bryant, S. H. (2010). The NCBI BioSystems database. *Nucleic Acids Research*, *38*(SUPPL.1), D492-6. http://doi.org/10.1093/nar/gkp858

Gissi, C., Iannelli, F., & Pesole, G. (2008). Evolution of the mitochondrial genome of Metazoa as exemplified by comparison of congeneric species. *Heredity*. http://doi.org/10.1038/hdy.2008.62

Grabherr, M. G., Haas, B. J., Yassour, M., Levin, J. Z., Thompson, D. A., Amit, I., … Regev, A. (2011). Full-length transcriptome assembly from RNA-Seq data without a reference genome. *Nature Biotechnology*, *29*(7), 644–652. http://doi.org/10.1038/nbt.1883

Greiner, S., Lehwark, P., & Bock, R. (2019). OrganellarGenomeDRAW (OGDRAW) version 1.3.1: expanded toolkit for the graphical visualization of organellar genomes. *Nucleic Acids Research*, *47*(W1), W59–W64. http://doi.org/10.1093/nar/gkz238

Jühling, F., Pütz, J., Bernt, M., Donath, A., Middendorf, M., Florentz, C., & Stadler, P. F. (2012). Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Research*, *40*(7), 2833–2845. http://doi.org/10.1093/nar/gkr1131

Jühling, F., Pütz, J., Florentz, C., & Stadler, P. F. (2012). Armless mitochondrial tRNAs in enoplea (nematoda). *RNA Biology*, *9*(9), 1161–1166. http://doi.org/10.4161/rna.21630

Kahle, D., & Wickham, H. (2013). ggmap: Spatial visualization with ggplot2. *R Journal*, *5*(1), 144–161. http://doi.org/10.32614/rj-2013-014

Katoh, K., Asimenos, G., & Toh, H. (2009). Multiple alignment of DNA sequences with MAFFT. *Methods in Molecular Biology*, *537*, 39–64. http://doi.org/10.1007/978-1-59745-251-9_3

789 Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., …
790 Drummond, A. (2012). Geneious Basic: An integrated and extendable desktop
791 software platform for the organization and analysis of sequence data. *Bioinformatics*,
792 *28*(12), 1647–1649. http://doi.org/10.1093/bioinformatics/bts199

793 Kllnç, G. M., Kashuba, N., Yaka, R., Sümer, A. P., Yüncü, E., Shergin, D., …
794 Götherström, A. (2018). Investigating Holocene human population history in North
795 Asia using ancient mitogenomes. *Scientific Reports*, *8*(1), 8969.
796 http://doi.org/10.1038/s41598-018-27325-0

797 Kulski, J. K. (2016). Next-Generation Sequencing — An Overview of the History, Tools,
798 and "Omic" Applications. In J. K. Kulski (Ed.), *Next Generation Sequencing -*
799 *Advances, Applications and Challenges*. IntechOpen. http://doi.org/10.5772/61964

800 Ladoukakis, E. D., & Zouros, E. (2017). Evolution and inheritance of animal mitochondrial
801 DNA: Rules and exceptions. *Journal of Biological Research Thessaloniki (Greece)*.
802 http://doi.org/10.1186/s40709-017-0060-4

803 Lam, H. Y. K., Clark, M. J., Chen, R., Chen, R., Natsoulis, G., O'Huallachain, M., …
804 Snyder, M. (2012). Performance comparison of whole-genome sequencing platforms.
805 *Nature Biotechnology*, *30*(1), 78–82. http://doi.org/10.1038/nbt.2065

806 Lanfear, R., Frandsen, P. B., Wright, A. M., Senfeld, T., & Calcott, B. (2017).
807 Partitionfinder 2: New methods for selecting partitioned models of evolution for
808 molecular and morphological phylogenetic analyses. *Molecular Biology and*
809 *Evolution*, *34*(3), 772–773. http://doi.org/10.1093/molbev/msw260

810 Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nature*
811 *Methods*, *9*(4), 357–359. http://doi.org/10.1038/nmeth.1923

812 Lartillot, N., Brinkmann, H., & Philippe, H. (2007). Suppression of long-branch attraction
813 artefacts in the animal phylogeny using a site-heterogeneous model. In *BMC*
814 *Evolutionary Biology* (Vol. 7). http://doi.org/10.1186/1471-2148-7-S1-S4

815 Lartillot, N., Rodrigue, N., Stubbs, D., & Richer, J. (2013). Phylobayes mpi: Phylogenetic
816 reconstruction with infinite mixtures of profiles in a parallel environment. *Systematic*
817 *Biology*, *62*(4), 611–615. http://doi.org/10.1093/sysbio/syt022

818 Laslett, D., & Canback, B. (2004). ARAGORN, a program to detect tRNA genes and
819 tmRNA genes in nucleotide sequences. *Nucleic Acids Research*, *32*(1), 11–16.

http://doi.org/10.1093/nar/gkh152

Leinonen, R., Sugawara, H., & Shumway, M. (2011). The sequence read archive. *Nucleic Acids Research*, *39*(SUPPL. 1), D19–D21. http://doi.org/10.1093/nar/gkq1019

López-López, A., & Vogler, A. P. (2017). The mitogenome phylogeny of Adephaga (Coleoptera). *Molecular Phylogenetics and Evolution*, *114*, 166–174. http://doi.org/10.1016/j.ympev.2017.06.009

Lunt, D. H., Whipple, L. E., & Hyman, B. C. (1998). Mitochondrial DNA variable number tandem repeats (VNTRs): Utility and problems in molecular ecology. *Molecular Ecology*. http://doi.org/10.1046/j.1365-294x.1998.00495.x

Macías-Hernández, N., Athey, K., Tonzo, V., Wangensteen, O. S., Arnedo, M., & Harwood, J. D. (2018). Molecular gut content analysis of different spider body parts. *PLoS ONE*, *13*(5), e0196589. http://doi.org/10.1371/journal.pone.0196589

Macías-Hernández, N., Bidegaray-Batista, L., Emerson, B. C., Oromí, P., & Arnedo, M. (2013). The imprint of geologic history on within-island diversification of woodlouse-hunter spiders (Araneae, Dysderidae) in the canary islands. *Journal of Heredity*. http://doi.org/10.1093/jhered/est008

Macías-Hernández, N., Bidegaray-Batista, L., Oromí, P., & Arnedo, M. A. (2012). The odd couple: Contrasting phylogeographic patterns in two sympatric sibling species of woodlouse-hunter spiders in the Canary Islands. *Journal of Zoological Systematics and Evolutionary Research*, *51*(1), 29–37. http://doi.org/10.1111/jzs.12008

Macías-Hernández, N., López, S. de la C., Roca-Cusachs, M., Oromí, P., & Arnedo, M. A. (2016). A geographical distribution database of the genus Dysdera in the Canary Islands (Araneae, Dysderidae). *ZooKeys*, *2016*(625), 11–23. http://doi.org/10.3897/zookeys.625.9847

Macías-Hernández, N., Oromí, P., & Arnedo, M. A. (2008). Patterns of diversification on old volcanic islands as revealed by the woodlouse-hunter spider genus Dysdera (Araneae, Dysderidae) in the eastern Canary Islands. *Biological Journal of the Linnean Society*, *94*(3), 589–615. http://doi.org/10.1111/j.1095-8312.2008.01007.x

Magalhaes, I. L. F., Azevedo, G. H. F., Michalik, P., & Ramírez, M. J. (2019). The fossil record of spiders revisited: implications for calibrating trees and evidence for a major faunal turnover since the Mesozoic. *Biological Reviews of the Cambridge*

Philosophical Society, *128*(317), 98–34. http://doi.org/10.1111/brv.12559

Mardis, E. R. (2008). The impact of next-generation sequencing technology on genetics. *Trends in Genetics*. http://doi.org/10.1016/j.tig.2007.12.007

Margulies, M., Egholm, M., Altman, W. E., Attiya, S., Bader, J. S., Bemben, L. A., … Rothberg, J. M. (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, *437*(7057), 376–380. http://doi.org/10.1038/nature03959

Masta, S. E., & Boore, J. L. (2008). Parallel evolution of truncated transfer rna genes in arachnid mitochondrial genomes. *Molecular Biology and Evolution*, *25*(5), 949–959. http://doi.org/10.1093/molbev/msn051

Michalik, P., & Ramírez, M. J. (2014). Evolutionary morphology of the male reproductive system, spermatozoa and seminal fluid of spiders (Araneae, Arachnida) - Current knowledge and future directions. *Arthropod Structure and Development*. http://doi.org/10.1016/j.asd.2014.05.005

Miller, M. A., Pfeiffer, W., & Schwartz, T. (2010). Creating the CIPRES Science Gateway for inference of large phylogenetic trees. In *Proceedings of the Gateway Computing Environments Workshop (GCE)* (pp. 1–8). http://doi.org/10.1109/GCE.2010.5676129

Morrison, D. A. (2010). How and where to look for tRNAs in Metazoan mitochondrial genomes, and what you might find when you get there. *ArXiv.Org*, (1), 1–27. Retrieved from http://arxiv.org/abs/1001.3813

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: A fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. http://doi.org/10.1093/molbev/msu300

Ozsolak, F., & Milos, P. M. (2011). RNA sequencing: advances, challenges and opportunities. *Nature Reviews Genetics*. http://doi.org/10.1038/nrg2934

Pisani, D., Carton, R., Campbel, L. I., Akanni, W. A., Mulville, E., & Rota-Stabelli, O. (2013). An Overview of Arthropod Genomics, Mitogenomics, and the Evolutionary Origins of the Arthropod Proteome. In A. Minelli, G. Boxshall, & G. Fusco (Eds.), *Arthropod Biology and Evolution: Molecules, Development, Morphology* (pp. 41–61). Berlin, Heidelberg: Springer-Verlag. http://doi.org/10.1007/978-3-642-36160-9

Plese, B., Rossi, M. E., Kenny, N. J., Taboada, S., Koutsouveli, V., & Riesgo, A. (2019).

Trimitomics: An efficient pipeline for mitochondrial assembly from transcriptomic reads in nonmodel species. *Molecular Ecology Resources*, *19*(April), 1–10. http://doi.org/10.1111/1755-0998.13033

Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T., & Sandhu, M. S. (2018). Long reads: their purpose and place. *Human Molecular Genetics*. http://doi.org/10.1093/hmg/ddy177

Pons, J., Bover, P., Bidegaray-Batista, L., & Arnedo, M. A. (2019). Arm-less mitochondrial tRNAs conserved for over 30 millions of years in spiders. *BMC Genomics*, *20*(1), 665. http://doi.org/10.1186/s12864-019-6026-1

Rambaut, A. (2016). FigTree v1.4.3. Retrieved from http://tree.bio.ed.ac.uk/software/figtree/

Régnier, P., & Marujo, P. E. (2013). Polyadenylation and Degradation of RNA in Prokaryotes. Retrieved September 10, 2019, from https://www-ncbi-nlm-nih-gov.sire.ub.edu/books/NBK6253/

Resnik, D. B., & Shamoo, A. E. (2017). Reproducibility and Research Integrity. *Accountability in Research*, *24*(2), 116–123. http://doi.org/10.1080/08989621.2016.1257387

Řezáč, M., Pekár, S., & Lubin, Y. (2008). How oniscophagous spiders overcome woodlouse armour. *Journal of Zoology*, *275*(1), 64–71. http://doi.org/10.1111/j.1469-7998.2007.00408.x

Ronquist, F., & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics*, *19*(12), 1572–1574. http://doi.org/10.1093/bioinformatics/btg180

Roskov, Y., Ower, G., Orrell, T., Nicolson, D., Bailly, N., Kirk, P. M., … Penev, L. (2019). Species 2000 & ITIS Catalogue of Life, 25th March 2019. Retrieved September 9, 2018, from www.catalogueoflife.org/col

Rubinoff, D. (2006). Utility of mitochondrial DNA barcodes in species conservation. *Conservation Biology*. http://doi.org/10.1111/j.1523-1739.2006.00372.x

Sakurai, M., Watanabe, Y. I., Watanabe, K., & Ohtsuki, T. (2006). A protein extension to shorten RNA: Elongated elongation factor-Tu recognizes the D-arm of T-armless tRNAs in nematode mitochondria. *Biochemical Journal*, *399*(2), 249–256.

913 http://doi.org/10.1042/BJ20060781

914 Sayers, E. W., Cavanaugh, M., Clark, K., Ostell, J., Pruitt, K. D., & Karsch-Mizrachi, I.
915 (2019). GenBank. *Nucleic Acids Research*, *47*(D1), D94–D99.
916 http://doi.org/10.1093/nar/gky989

917 Shao, L., & Li, S. (2018). Early Cretaceous greenhouse pumped higher taxa diversification
918 in spiders. *Molecular Phylogenetics and Evolution*, *127*(July 2017), 146–155.
919 http://doi.org/10.1016/j.ympev.2018.05.026

920 Simion, P., Philippe, H., Baurain, D., Jager, M., Richter, D. J., Di Franco, A., … Manuel,
921 M. (2017). A Large and Consistent Phylogenomic Dataset Supports Sponges as the
922 Sister Group to All Other Animals. *Current Biology*, *27*(7), 958–967.
923 http://doi.org/10.1016/j.cub.2017.02.031

924 Smith, D. R. (2016). The past, present and future of mitochondrial genomics: Have we
925 sequenced enough mtDNAs? *Briefings in Functional Genomics*.
926 http://doi.org/10.1093/bfgp/elv027

927 Stokkan, M., Jurado-Rivera, J. A., Juan, C., Jaume, D., & Pons, J. (2016). Mitochondrial
928 genome rearrangements at low taxonomic levels: Three distinct mitogenome gene
929 orders in the genus Pseudoniphargus (Crustacea: Amphipoda). *Mitochondrial DNA*,
930 *27*(5), 3579–3589. http://doi.org/10.3109/19401736.2015.1079821

931 Talavera, G., & Vila, R. (2011). What is the phylogenetic signal limit from mitogenomes?
932 the reconciliation between mitochondrial and nuclear data in the Insecta class
933 phylogeny. *BMC Evolutionary Biology*, *11*(1). http://doi.org/10.1186/1471-2148-11-
934 315

935 Tian, Y., & Smith, D. R. (2016). Recovering complete mitochondrial genome sequences
936 from RNA-Seq: A case study of Polytomella non-photosynthetic green algae.
937 *Molecular Phylogenetics and Evolution*, *98*, 57–62.
938 http://doi.org/10.1016/j.ympev.2016.01.017

939 Timmermans, M. J. T. N., Barton, C., Haran, J., Ahrens, D., Culverwell, C. L., Ollikainen,
940 A., … Vogler, A. P. (2016). Family-level sampling of mitochondrial genomes in
941 Coleoptera: Compositional heterogeneity and phylogenetics. *Genome Biology and*
942 *Evolution*, *8*(1), 161–175. http://doi.org/10.1093/gbe/evv241

943 Vieira, G. A., & Prosdocimi, F. (2019). Accessible molecular phylogenomics at no cost:

944 Obtaining 14 new mitogenomes for the ant subfamily Pseudomyrmecinae from public
945 data. *PeerJ*, *7*(1), e6271. http://doi.org/10.7717/peerj.6271

946 Wang, Z., Gerstein, M., & Snyder, M. (2009). RNA-Seq: a revolutionary tool for
947 transcriptomics. *Nature Reviews Genetics*, *10*, 57–63. http://doi.org/10.1038/nrg2484

948 Wende, S., Platzer, E. G., Jühling, F., Pütz, J., Florentz, C., Stadler, P. F., & Mörl, M.
949 (2014). Biological evidence for the world's smallest tRNAs. *Biochimie*, *100*(1), 151–
950 158. http://doi.org/10.1016/j.biochi.2013.07.034

951 Wheeler, W. C. (1995). Sequence alignment, parameter sensitivity, and the phylogenetic
952 analysis of molecular data. *Systematic Biology*, *44*(3), 321–331.
953 http://doi.org/10.1093/sysbio/44.3.321

954 Wheeler, W. C., Coddington, J. A., Crowley, L. M., Dimitrov, D., Goloboff, P. A.,
955 Griswold, C. E., … Zhang, J. (2017). The spider tree of life: phylogeny of Araneae
956 based on target-gene analyses from an extensive taxon sampling. *Cladistics*, *33*(6),
957 574–616. http://doi.org/10.1111/cla.12182

958 Wolfsberg, T. G., Schafer, S., Tatusov, R. L., & Tatusova, T. A. (2001). Organelle genome
959 resources at NCBI. *Trends in Biochemical Sciences*. http://doi.org/10.1016/S0968-
960 0004(00)01773-4

961 Wolstenholme, D. R. (1992). Animal Mitochondrial DNA: Structure and Evolution.
962 *International Review of Cytology*, *141*(C), 173–216. http://doi.org/10.1016/S0074-
963 7696(08)62066-5

964 Wolstenholme, D. R., Macfarlane, J. L., Okimoto, R., Clary, D. O., & Wahleithner, J. A.
965 (1987). Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of
966 nematode worms. *Proceedings of the National Academy of Sciences of the United*
967 *States of America*, *84*(5), 1324–1328. http://doi.org/10.1073/pnas.84.5.1324

968 World Spider Catalog. (2019). World Spider Catalog Version 19.0.
969 http://doi.org/10.24436/2

970 Zerbino, D. R., & Birney, E. (2008). Velvet: Algorithms for de novo short read assembly
971 using de Bruijn graphs. *Genome Research*, *18*, 821–829.
972 http://doi.org/10.1101/gr.074492.107

975 **FIGURE LEGENDS**

976 **Figure 1.** Representatives of the family Dysderidae: A) *Dysdera longa* (Dysderinae), photo
977 credit Pedro Oromí; B) *Dysdera tilosensis* (Dysderinae), photo credit Pedro Oromí; C)
978 *Dysdera silvatica* (Dysderinae), photo credit Pedro Oromí; D) *Dysdera sibyllina*
979 (Dysderinae), photo credit Pedro Oromí; E) *Parastalita stygia* (Rhodinae), photo credit
980 Fulvio Gasparo; F) *Stalagtia hercegovinensis* (Harpacteinae), photo credit Jana Bedek.

981 **Figure 2.** Pipeline workflow for the recovery of mitogenomes from NGS data and
982 constitution of the final datasets for phylogenetic inference. r. l.= recovered length. See text
983 for further details.

984 **Figure 3.** *Trimmitomics* performance. Graph showing the percentage of recovered
985 mitogenome length (calculated over the average length of all the spider mitogenomes
986 available at the NCBI public database), both in total (black bars) and in each step of the
987 pipeline, for each mined specimen. Data type codified as **G** (genomic) and **T**
988 (transcriptomic). The plus sign (+) mark deep coverage genomes (>15X) and transcriptomes
989 (>200 million reads).

990 **Figure 4.** Summary phylogenetic tree from dataset S1 (lower levels of missing data).
991 Bayesian inference (BI) tree obtained from the concatenated analysis of mitochondrial
992 protein coding genes (PCGs) and ribosomal RNAs (rRNAs), summarizing maximum
993 likelihood (ML) and BI (MrBayes) supports from the analyses based in PCG and the analyses
994 based in PCG and rRNA. Clade support: Left column: ML, PCG data (above), PCG + rRNA
995 data (below); right column: BI, PCG data (above), PCG + rRNA data (below). Black square,
996 node supported by ML bootstrap > 80%, and posterior probability > 0,95; grey square, clade
997 recovered but support < 80 or 95%, respectively; white square, clade not recovered. LTC
998 stands for Lost Tracheae Clade and SC for Scytodoidea. Assembled and annotated
999 mitogenomes coded as * (Taxa from S1). Curated mitogenomes downloaded from NCBI
1000 without asterisk.

1001 **Figure 5.** Summary phylogenetic tree from dataset S3 (higher levels of missing data).
1002 Bayesian inference tree obtained from the concatenated analysis of mitochondrial protein
1003 coding genes and ribosomal rRNAs, summarizing maximum likelihood and Bayesian
1004 inference (MrBayes) supports from both the analyses based in PCG and the analyses based
1005 in PCG and rRNA. Clade supports coded as in Fig. 4. Taxa with a taxonomically incongruent

placement are highlighted in red. Assembled and annotated mitogenomes coded as * (taxa from S1), ** (taxa from S2), ***(taxa from S3). Curated mitogenomes downloaded from NCBI without asterisk.

Figure 6. Detail of the family Dysderidae showing the phylogenetic relationships and geographic distribution within the Canarian *Dysdera*. The tree corresponds to the family Dysderidae as recovered with S1 **(Fig. 4)**. Supports coded as in Fig. 4. Map drawn with the ggmap package (Kahle & Wickham, 2013).

Figure 7. Map of the complete *Dysdera silvatica* mitogenome obtained with OGDRAW v1.3.1 (Greiner, Lehwark, & Bock, 2019)**,** representing PCGs, rRNAs and tRNAs. Genes inside the circle are transcribed clockwise, while genes outside the circle are transcribed counter-clockwise. The inner circle shows the GC content graph.

Figure 8. Comparison of the mitochondrial gene organization (PCG, tRNAs and rRNAs) found in the eight species analyzed in this study and *P. romandiolae* from Pons et al. (2019) within a phylogenetic framework. Loci rearrangements are highlighted in yellow and orange, with the dark color tone marking the derived condition. Cave adapted taxa are indicated with a cave symbol. *P. phalangioides* (Pholcidae); *P. stygia* (Rhodinae, Dysderidae); *S. hercegovinensis* (Harpacteinae, Dysderidae); *P. romandiolae, D. asiatica, D. sibyllina, D. longa, D. silvatica,* and *D. iguanensis* (Dysderinae, Dysderidae). Gene arrangements drawn with OGDRAW v1.3.1.

Figure 9. Secondary structures predicted for the set of 22 tRNA of *Dysdera silvatica*. tRNAs with two alternative secondary structures are highlighted in boxes. Structures drawn with VARNA.

Figure 10. Diagram showing two alternative secondary structures predicted for tRNA-M in all the species included in the study within a phylogenetic framework. Cave adapted taxa are indicated with a cave symbol. Structures drawn with VARNA (Darty, Denise, & Ponty, 2009).

LIST OF SUPPORTING INFORMATION

Table S1. tRNA automatic annotation and reconstruction performance.

1035 **Table S2.** Summary table showing the performance of the pipeline *Trimitomics*.

1036 **Table S3.** Mitogenome recovery success by gene (PCG and rRNA).

1037 **Table S4.** Total length of fully recovered mitogenomes.

1038 **Table S5.** Partition scheme scores for each dataset (S1, S2 and S3) with only PCG
1039 information.

1040 **Table S6.** Partition scheme scores for each dataset (S1, S2 and S3) with PCG + rRNA
1041 information.

1042 **Figure S1.** Summary phylogenetic tree from dataset S2 (intermediate levels of missing data).

1043 **Figure S2.** Maximum likelihood tree obtained from the concatenated analysis of
1044 mitochondrial PCG an rRNAs from dataset S1 (lower levels of missing data).

1045 **Figure S3.** Bayesian inference tree obtained from the concatenated analysis of mitochondrial
1046 PCG and rRNAs from dataset S1 (lower levels of missing data).

1047 **Figure S4.** Maximum likelihood tree obtained from the concatenated analysis of
1048 mitochondrial PCG from dataset S1 (lower levels of missing data).

1049 **Figure S5.** Bayesian inference tree obtained from the concatenated analysis of mitochondrial
1050 PCG from dataset S1 (lower levels of missing data).

1051 **Figure S6.** Maximum likelihood tree obtained from the concatenated analysis of
1052 mitochondrial PCG an rRNAs from dataset S2 (intermediate levels of missing data).

1053 **Figure S7.** Bayesian inference tree obtained from the concatenated analysis of mitochondrial
1054 PCG and rRNA from dataset S2 (intermediate levels of missing data).

1055 **Figure S8.** Maximum likelihood tree obtained from the concatenated analysis of
1056 mitochondrial PCG from dataset S2 (intermediate levels of missing data).

1057 **Figure S9.** Bayesian inference tree obtained from the concatenated analysis of mitochondrial
1058 PCG from dataset S2 (intermediate levels of missing data).

1059 **Figure S10.** Maximum likelihood tree obtained from the concatenated analysis of
1060 mitochondrial PCG and rRNA from dataset S3 (higher levels of missing data).

1061 **Figure S11.** Bayesian inference tree obtained from the concatenated analysis of
1062 mitochondrial PCG and rRNAs from dataset S3 (higher levels of missing data).

1063 **Figure S12.** Maximum likelihood tree obtained from the concatenated analysis of
1064 mitochondrial PCG from dataset S3 (higher levels of missing data).

1065 **Figure S13.** Bayesian inference tree obtained from the concatenated analysis of
1066 mitochondrial PCG from dataset S3 (higher levels of missing data).

1067 **Figure S14.** PhyloBayes Bayesian inference tree derived from the CAT-GTR + Γ analysis
1068 of mitochondrial PCG from dataset S1 (lower levels of missing data).

1069 **Figure S15.** PhyloBayes Bayesian inference tree derived from the CAT-GTR + Γ analysis
1070 of mitochondrial PCG at the amino acid level from dataset S1 (lower levels of missing data).

1071 **Figure S16.** PhyloBayes Bayesian inference tree derived from the CAT-GTR + Γ analysis
1072 of mitochondrial PCG from dataset S3 (higher levels of missing data).

1073 **Figure S17.** PhyloBayes Bayesian inference tree derived from the CAT-GTR + Γ analysis
1074 of mitochondrial PCG at the amino acid level from dataset S3 (higher levels of missing data).

1075 **Figure S18.a-S18.g.** Secondary structures predicted for the set of 22 tRNA of all seven
1076 species analysed. For some tRNAs there is more than one plausible structure.

1077 **TABLES**

1078 **Table 1. Final alignments information.** Information on the size (number of terminals and
1079 characters) and percentage of missing data (%MD) of the six matrices used for phylogenetic
1080 inference analysis.

Dataset	S1		S2		S3	
Markers	**PCG**	**PCG+rRNA**	**PCG**	**PCG+rRNA**	**PCG**	**PCG+rRNA**
# Terminals	62	62	76	76	93	93
# Characters	10473	12492	10527	12549	10562	12586
% MD	0.66	0.80	8.98	9.40	21.97	22.63

APPENDICES

Appendix 1. Dysderidae genomic (a) and transcriptomic (b) data generated in-house. Source coded as MA lab (Miquel Arnedo lab) and JR lab (Julio Rozas lab). Coverage (Cov.) coded as deep (genomes >15X, transcriptomes >200 million reads), and low (genomes <15X, transcriptomes <200 million reads). Institution (Inst.) coded as UB (University of Barcelona) and CBSS (Croatian Biospeleological Society).

a. Genomic

SUBFAMILY: Species name	**Genbank accession #**	**Source**	**Cov.**	**Bio Sample accession #**	**Locality**	**Inst.**
DYSDERINAE:						
Dysdera asiatica Nosek, 1905	SRR12105779	MA lab	Low	SAMN14852306	Fener Mahallesi, Antalya, Turkey	UB
Dysdera bandamae Schmidt, 1973	SRR12115710	JR lab	Deep	SAMN14839776	Gran Canaria, western Canary Islands, Spain	UB
Dysdera gomeriensis Strand, 1911	SRR12114572	JR lab	Deep	SAMN14839773	La Gomera, western Canary Islands, Spain	UB
Dysdera iguanensis Wunderlich, 1987	SRR12106847	MA lab	Low	SAMN14839779	Tenerife, western Canary Islands, Spain	UB
Dysdera sibyllina Arnedo, 2007	SRR12106848	MA lab	Low	SAMN14839778	Tenerife, western Canary Islands, Spain	UB
Dysdera silvatica Schmidt, 1981	SRR7340408	JR lab	Deep	SAMN09381544	La Gomera, western Canary Islands, Spain	UB
Dysdera tilosensis Wunderlich, 1992	SRR12115711	JR lab	Deep	SAMN14839775	Gran Canaria, western Canary Islands, Spain	UB
Dysdera verneaui Simon, 1883	SRR12115483	JR lab	Deep	SAMN14839774	Tenerife, western Canary Islands, Spain	UB
Dysdera yguanirae Arnedo & Ribera, 1997	SRR12106849	MA lab	Low	SAMN14839777	Gran Canaria, western Canary Islands, Spain	UB
Parachtes riberai (2) Bosmans, 2017	SRR12105362	MA lab	Low	SAMN14839780	Mallorca, Balearic Islands, Spain	UB
Parachtes teruelis (2) (Kraus, 1955)	SRR12105366	MA lab	Low	SAMN14839781	Guadalajara, Spain	UB
HARPACTEINAE:						
Stalagtia hercegovinensis (Nosek, 1905)	SRR12105800	MA lab	Low	SAMN14839782	Fuzine, Croatia	CBSS
RHODINAE:						
Parastalita stygia (Joseph, 1882)	SRR12106476	MA lab	Low	SAMN14839783	Dubrovnik, Croatia	CBSS

b. Transcriptomic

SUBFAMILY/ Species name	Genbank accession #	Source	Cov.	Bio Sample accession #	Locality	Inst.
DYSDERINAE:						
Dysdera bandamae Schmidt, 1973	SRR6820569-73, SRR6820584-85	JR lab	Deep	SAMN08667720-26	Gran Canaria, western Canary Islands, Spain	UB
Dysdera coiffaiti Denis, 1962	SRR12105326	MA lab	Low	SAMN14839772	Madeira, Portugal	UB
Dysdera fustigans Alicata, 1966	SRR12105371	MA lab	Low	SAMN14852305	Italy	UB
Dysdera gomeriensis Strand, 1911	SRR6820576-79	JR lab	Deep	SAMN08667716-19	El Hierro, western Canary Islands, Spain	UB
Dysdera longa Wunderlich, 1992	SRR1210536	MA lab	Low	SAMN14839770	Fuerteventura, eastern Canary Islands, Spain	UB
Dysdera nesiotes Simon, 1907	SRR12105351	MA lab	Low	SAMN14839771	Lanzarote, eastern Canary Islands, Spain	UB
Dysdera silvatica Schmidt, 1981	SRR3203326, SRR3203350, SRR3203363-64	JR lab	Deep	SAMN04527047-50	La Gomera, western Canary Islands, Spain	UB
Dysdera tilosensis Wunderlich, 1992	SRR6820566-68, SRR6820574-75, SRR6820586-91	JR lab	Deep	SAMN08667727-36	Gran Canaria, western Canary Islands, Spain	UB
Dysdera verneaui Simon, 1883	SRR6820580-83	JR lab	Deep	SAMN08667712-15	Tenerife, western Canary Islands, Spain	UB

1087

1088 **Appendix 2. Synspermiata genomic and transcriptomic data downloaded from GenBank.** Institution (Inst.) coded as BCM (Baylor College
1089 of Medicine), AB (Auburn University), HU (Harvard University), RF (Rosa Fernandez), and CAS (Chinese Academy of Sciences).

a. Genomic

FAMILY: Species name	Genbank accession #	# of spots	# of bases	Size	Bio Sample accession #	Institution
SICARIIDAE:						
Loxosceles reclusa (g) Gertsch & Mulaik, 1940	SRX494491	106.911.238	21.6G	13.7Gb	SAMN02671248	BCM

b. Transcriptomic

FAMILY: Species name	Genbank accession #	# of spots	# of bases	Size	Bio Sample accession #	Institution
CAPONIIDAE:						
Calponia harrisonfordi Platnick, 1993	SRR3144089	59.921.781	6G	3.7Gb	SAMN04453334	AU
DIGUETIDAE:						
Diguetia sp. Simon, 1895	SRR3144093	19.838.746	4.4G	1.9Gb	SAMN04453337	AU
DRYMUSIDAE:						
Drymusa sp. Simon, 1892	SRR6997739	77.728.210	23.3G	11.1Gb	SAMN08436905	RF
DYSDERIDAE:						
Dysdera crocata (1) C. L. Koch, 1838	SRR3144095	22.746.728	5G	2.2Gb	SAMN04453338	AU
Dysdera crocata (2) C. L. Koch, 1838	SRR1328258	20.060.716	6G	3.2Gb	SAMN02837098	HU
FILISTATIDAE:						
Filistata insidiatrix (Forsskål, 1775)	SRR6997865	30.407.368	9.1 G	4.1 Gb	SAMN08436908	RF
Kukulcania hibernalis (Hentz, 1842)	SRR1514878	42.693.292	4.3 G	2.6 Gb	SAMN02837058	AU
HYPOCHILIDAE:						
Hypochilus gertschi Hoffman, 1963	SRR6997860	20.891.420	6.3 G	3 Gb	SAMN08436914	RF
Hypochilus pococki Platnick, 1987	SRR1514889	25.747.925	2.6 G	1.7 Gb	SAMN02837061	AU
OCHYROCERATIDAE:						
Ochyrocera sp. Simon, 1892	SRR7028536	11.833.027	3.5G	1.7Gb	SAMN08436938	RF

1090

FAMILY: Species name	Genbank accession #	# of spots	# of bases	Size	Bio Sample accession #	Institution
OONOPIDAE:						
Ischnothyreus sp. Simon, 1893	SRR6997859	18.550.258	5.6G	2.7Gb	SAMN08436915	RF
Opopaea cornuta Yin & Wang, 1984	SRR6425918	23.323.737	5.8G	2.9Gb	SAMN08241469	CAS
Opopaea sp. Simon, 1892	SRR6998659	21.857.441	6.6G	3Gb	SAMN08436940	RF
ORSOLOBIDAE:						
Gen. sp. Cooke,1965	SRR6998651	9.269.339	2.8G	1.2Gb	SAMN08436941	RF
Maoriata sp. Forster & Platnick, 1985	SRR6997868	29.155.015	8.7G	4.1Gb	SAMN08436922	RF
PERIEGOPIDAE:						
Periegops suterii (Urquhart, 1892)	SRR6998656	10.563.192	3.2G	1.4Gb	SAMN08436946	RF
PHOLCIDAE:						
Pholcus manueli (1) Gertsch, 1937	SRR6425913	25.653.145	6.4G	3.1Gb	SAMN08241462	CAS
Pholcus manueli (2) Gertsch, 1937	SRR1365208	22.570.061	6.8G	3.6Gb	SAMN02848470	HU
Pholcus phalangioides (1) (Fuesslin, 1775)	SRR3144082	58.019.035	5.8G	3.4Gb	SAMN04453342	AU
Pholcus phalangioides (2) (Fuesslin, 1775)	SRR1514900	24.861.584	2.5G	1.5Gb	SAMN02837063	AU
PSILODERCIDAE:						
Althepus christae Wang & Li, 2013	SRR6425911	30.078.398	9G	2.8Gb	SAMN08241483	CAS
Flexicrurum sp. Tong & Li, 2007	SRR6425916	23.694.710	5.9G	3Gb	SAMN08241467	CAS
SCYTODIDAE:						
Scytodes globula Nicolet, 1849	SRR6998911	47.256.101	14.2G	6.1Gb	SAMN08436955	RF
Scytodes sp. Latreille, 1804	SRR6425915	25.239.370	6.3G	3Gb	SAMN08241464	CAS
Scytodes thoracica (Latreille, 1802)	SRR1514872	30.924.460	6.2G	2.7Gb	SAMN02837050	AU
Stedocys sp. Ono, 1995	SRR6425909	29.003.517	8.7G	3.1Gb	SAMN08241484	CAS
SEGESTRIIDAE:						
Segestria bavarica C. L. Koch, 1843	SRR6425927	25.857.888	6.5G	2.9Gb	SAMN08241472	CAS
Segestria sp. Latreille, 1804	SRR3144084	38.407.502	8.4G	3.7Gb	SAMN04453344	AU
SICARIIDAE:						
Loxosceles deserta Gertsch, 1973	SRR3144077	61.963.685	6.2G	3.6Gb	SAMN04453345	AU
Loxosceles reclusa (1) Gertsch & Mulaik, 1940	SRR1824530	26.556.146	5.4G	3.6Gb	SAMN03353442	BCM
Loxosceles reclusa (2) Gertsch & Mulaik, 1940	SRR1824531	22.846.862	4.6G	3.1Gb	SAMN03353443	BCM
Loxosceles reclusa (3) Gertsch & Mulaik, 1940	SRR1824532	27.587.079	5.6G	3.7Gb	SAMN03353441	BCM
Loxosceles rufescens (Dufour, 1820)	SRR6425917	36.245.511	9.1G	4.1Gb	SAMN08241466	CAS

1091 **Appendix 3. Araneae mitogenomic data downloaded from GenBank.**

FAMILY: Species name	Genbank accesion #	Lenght	Type
AGELENIDAE:			
Agelena silvatica Oliger, 1983	NC_033971.1	14776	circular
ARANEIDAE:			
Araneus angulatus Clerck, 1757	NC_032402.1	14205	circular
Araneus ventricosus Uyemura, 1961	NC_025634.1	14617	circular
Argiope amoena L. Koch, 1878	NC_024282.1	14121	circular
Argiope bruennichi (Scopoli, 1772)	NC_024281.1	14063	circular
Cyclosa argenteoalba Bösenberg & Strand, 1906	NC_027682.1	14575	circular
Cyrtarachne nagasakiensis Strand, 1918	NC_028077.1	14402	circular
Hypsosinga pygmaea (Sundevall, 1831)	NC_028078.1	14193	circular
Neoscona adianta (Walckenaer, 1802)	NC_029756.1	14161	circular
Neoscona nautica (L. Koch, 1875)	NC_029755.1	14049	circular
Neoscona theisi (Walckenaer, 1841)	NC_026290.1	14156	circular
Trichonephila clavate (L. Koch, 1878)	NC_008063.1	14436	circular
DICTYNIDAE:			
Argyroneta aquatica (Clerck, 1757)	NC_026863.1	16000	circular
DIPLURIDAE:			
Phyxioschema suthepium Raven & Schwendinger, 1989	NC_020322.1	13931	circular
DYSDERIDAE:			
Harpactocrates appenicola Simon, 1914	MN052924	14213	circular
Parachtes limbarae (Kraus, 1955)	MN052922	14111	linear
Parachtes riberai (1) Alicata, 1964	MN052919	14632	linear
Parachtes romandiolae (Caporiacco, 1949)	MN052923	14220	circular
Parachtes teruelis (1) (Kraus, 1955)	MN052921	13850	linear
Parachtes igvanus (Simon, 1882)	MN052920	14667	linear
HYPOCHILIDAE:			
Hypochilus thorelli Marx, 1888	NC_010777.1	13991	circular

LIPHISTIIDAE:			
Liphistius erawan Schwendinger, 1996	NC_020323.1	14197	circular
Songthela hangzhouensis (Chen, Zhang & Zhu, 1981)	NC_005924.1	14215	circular
LYCOSIDAE:			
Pardosa laura Karsch, 1879	NC_025223.1	14513	circular
Pirata subpiraticus (Bösenberg & Strand, 1906)	NC_025523.1	14528	circular
Wadicosa fidelis (O. Pickard-Cambridge, 1872)	NC_026123.1	14741	circular
NEMESIIDAE:			
Calisoga longitarsis (Simon, 1891)	NC_010780.1	14070	circular
OXYOPIDAE:			
Oxyopes sertatus L. Koch, 1878	NC_025224.1	14442	circular
PHOLCIDAE:			
Mesabolivar sp. (1) González-Sponga, 1998	NC_040859.1	14941	circular
Mesabolivar sp. (2) González-Sponga, 1999	NC_040860.1	14845	circular
Mesabolivar sp. (3) González-Sponga, 2000	NC_040861.1	14727	circular
Pholcus phalangioides (Fuesslin, 1775)	NC_020324.1	14459	circular
Pholcus sp. Walckenaer, 1805	KJ782458.1	14279	circular
SALTICIDAE:			
Carrhotus xanthogramma (Latreille, 1819)	NC_027492.1	14563	circular
Cheliceroides longipalpis Zabka, 1985	NC_041120.1	14334	circular
Epeus alboguttatus (Thorell, 1887)	NC_042829.1	14625	circular
Habronattus oregonensis (Peckham & Peckham, 1888)	NC_005942.1	14381	circular
Plexippus paykulli (Audouin, 1826)	NC_024877.1	14316	circular
Telamonia vlijmi Prószyński, 1984	NC_024287.1	14601	circular
SELENOPIDAE:			
Selenops bursarius Karsch, 1879	NC_024878.1	14272	circular
SICARIIDAE:			
Loxosceles similis Moenkhaus, 1898	NC_042902.1	14683	circular
TETRAGNATHIDAE:			
Tetragnatha maxillosa Thorell, 1895	NC_025775.1	14414	circular
Tetragnatha nitens (Audouin, 1826)	NC_028068.1	14639	circular

THERAPHOSIDAE:			
Cyriopagopus schmidti (von Wirth, 1991)	NC_005925.1	13874	circular
THOMISIDAE:			
Oxytate striatipes L. Koch, 1878	NC_025557.1	14407	circular

1093 **Appendix 4. Matrices composition for phylogenetic analyses.** Total information for each
1094 terminal expressed in base pairs (bp) and as a percentage over the average length for the
1095 complete concatenated set of loci (13 PCGs + 2 rRNAs) (%). Type of data coded as G
1096 (Genomic) and T (Transcriptomic). Taxa highlighted in grey indicate 0% recovery success,
1097 and exclusion for subsequent analysis. Lateral bars indicate taxa included in datasets S1, S2
1098 and S2.

Source	Type	Family	Species	Total length (bp)	Total length (%)	S1	S2	S3
In-house	G	Dysderidae	*Parastalita stygia*	12443	100			
	G	Dysderidae	*Dysdera gomerensis*	12417	100			
	T	Dysderidae	*Dysdera longa*	12416	100			
	G	Dysderidae	*Dysdera yguanirae*	12400	100			
	G	Dysderidae	*Stalagtia hercegovinensis*	12541	100			
	G	Dysderidae	*Dysdera silvatica*	12399	100			
	G	Dysderidae	*Parachtes teruelis*	12385	100			
	G	Dysderidae	*Dysdera verneaui*	12380	100			
	G	Dysderidae	*Dysdera bandamae*	12375	100			
	G	Dysderidae	*Dysdera tilosensis*	12367	100			
	G	Dysderidae	*Dysdera iguanensis*	12365	100			
	G	Dysderidae	*Dysdera asiatica*	12354	100			
	G	Dysderidae	*Parachtes riberai*	12347	100			
	G	Dysderidae	*Dysdera sibyllina*	12151	98			
	T	Dysderidae	*Dysdera coiffaiti*	11825	95			
	T	Dysderidae	*Dysdera fustigans*	11543	93			
	T	Dysderidae	*Dysdera nesiotes*	9439	76			
NCBI	T	Pholcidae	*Pholcus phalangioides* (1)	11466	92			
	T	Pholcidae	*Pholcus manueli* (2)	11451	92			
	T	Hypochilidae	*Hypochilus pococki*	9999	81			
	T	Pholcidae	*Pholcus manueli* (1)	9114	74			
	T	Dysderidae	*Dysdera crocata* (1)	8066	65			
	T	Segestriidae	*Segestria bavarica*	5362	43			
	T	Scytodidae	*Stedocys sp.*	4935	40			
	T	Dysderidae	*Dysdera crocata* (2)	4747	38			
	T	Sicariidae	*Loxosceles reclusa* (1)	4596	37			
	T	Pholcidae	*Pholcus phalangioides* (2)	4493	36			
	T	Sicariidae	*Loxosceles reclusa* (3)	4326	35			
	T	Psilodercidae	*Flexicrurum sp.*	4132	33			
	T	Sicariidae	*Loxosceles reclusa* (2)	4077	33			
	T	Oonopidae	*Opopaea cornuta*	4000	32			
	T	Filistatidae	*Kukulcania hibernalis*	3510	28			
	T	Segestriidae	*Segestria sp.*	3352	27			
	T	Sicariidae	*Loxosceles rufescens*	3260	26			
	T	Orsolobidae	Gen. sp.	3100	25			
	T	Sicariidae	*Loxosceles deserta*	3049	25			
	T	Psilodercidae	*Althepus christae*	2526	20			
	T	Ochyroceratidae	*Ochyrocera* sp.	2318	19			
	G	Sicariidae	*Loxosceles reclusa* (g)	2239	18			
	T	Orsolobidae	*Maoriata* sp.	2178	18			
	T	Scytodidae	*Scytodes sp.*	2166	17			
	T	Scytodidae	*Scytodes thoracica*	1736	14			
	T	Caponiidae	*Calponia harrisonfordi*	1530	12			

T	Diguetidae	*Diguetia sp.*	1526	12
T	Drymusidae	*Drymusa* sp.	1521	12
T	Scytodidae	*Scytodes globula*	1521	12
T	Oonopidae	*Ischnothyreus* sp.	1513	12
T	Periegopidae	*Periegops suterii*	1427	12
T	Filistatidae	*Filistata insidiatrix*	0	0
T	Hypochilidae	*Hypochilus gertschi*	0	0
T	Oonopidae	*Opopaea* sp.	0	0

1099

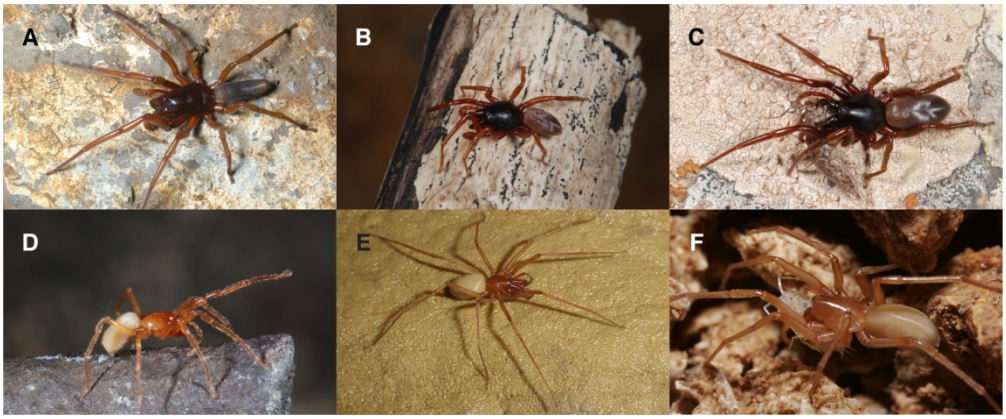
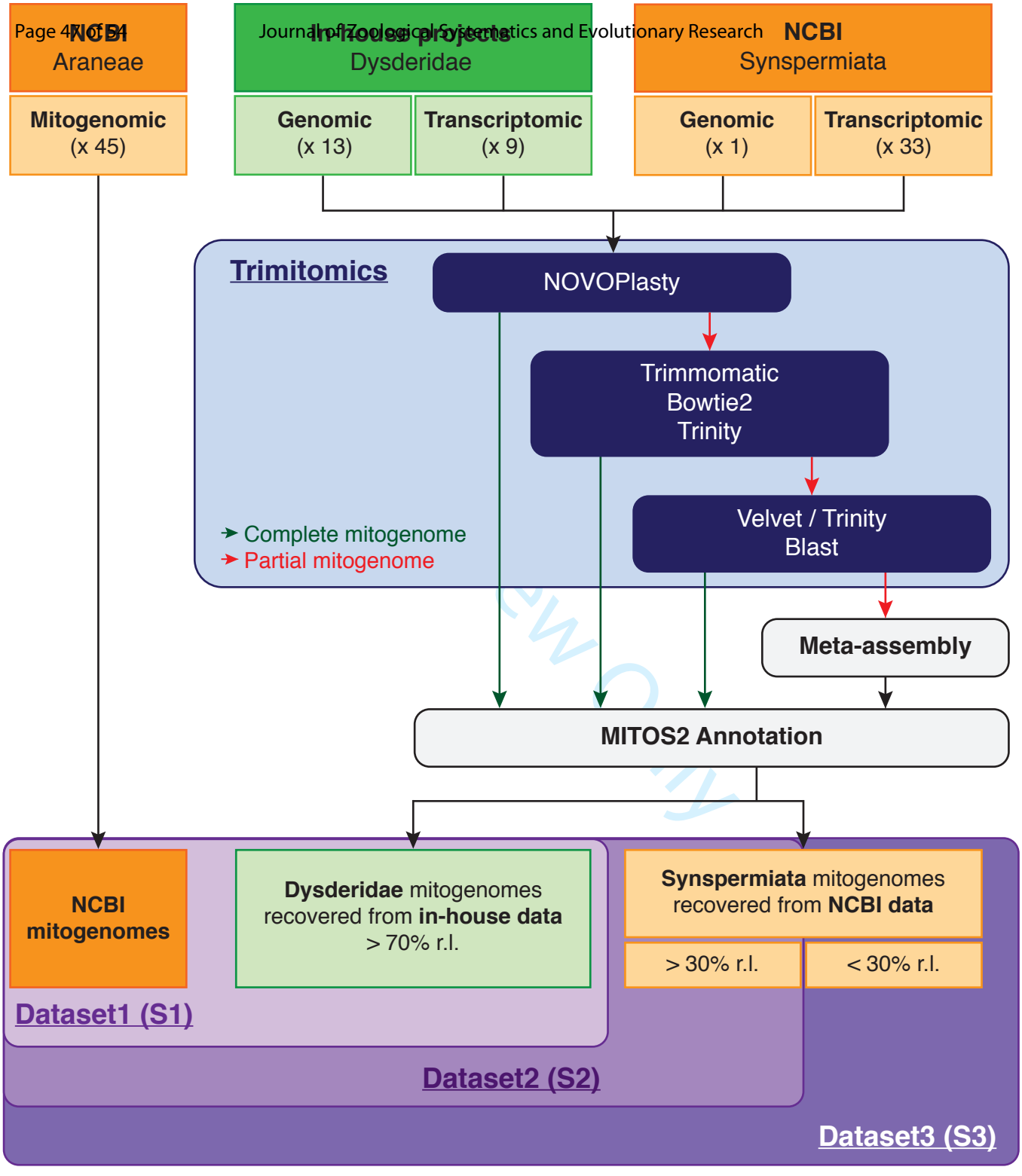
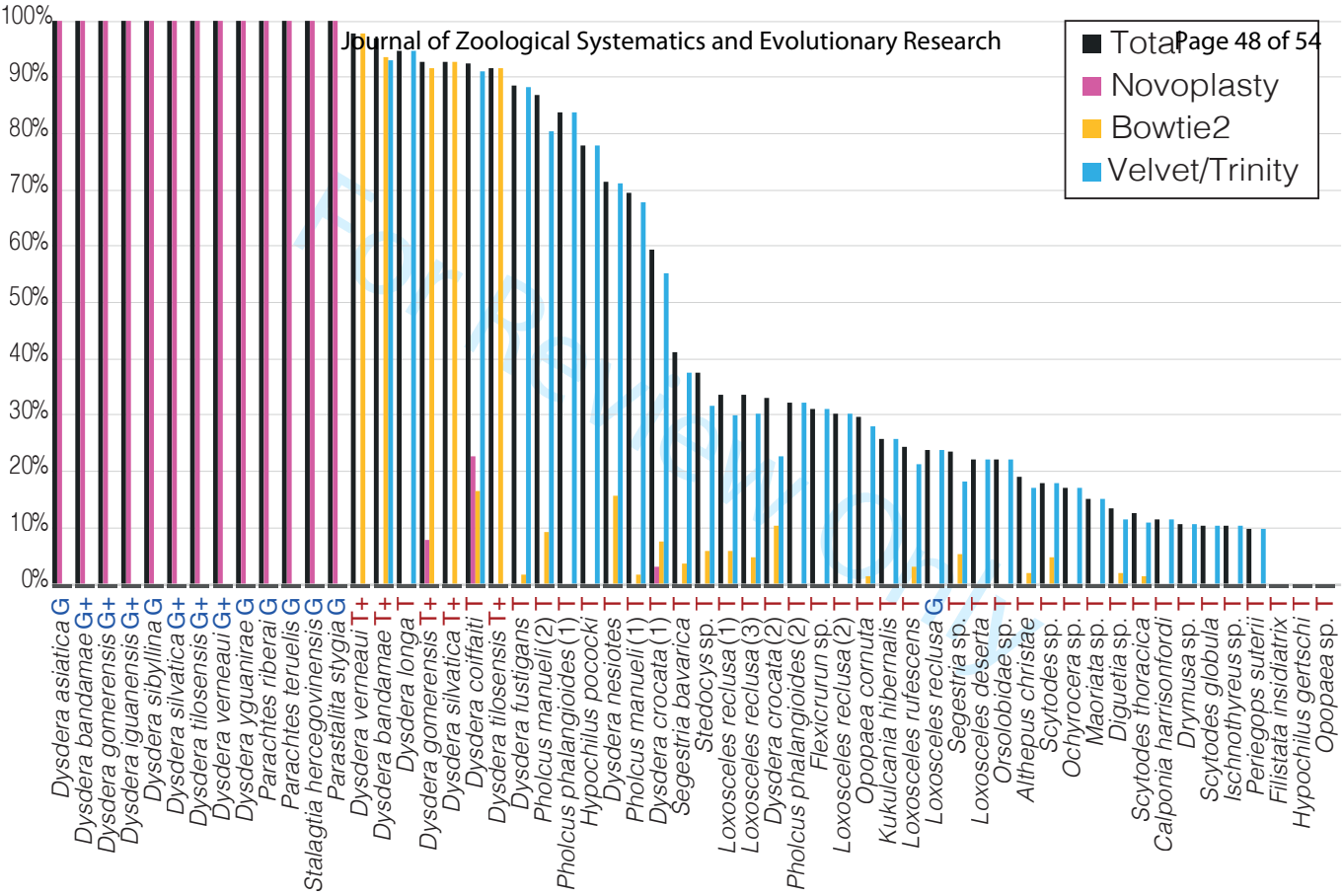
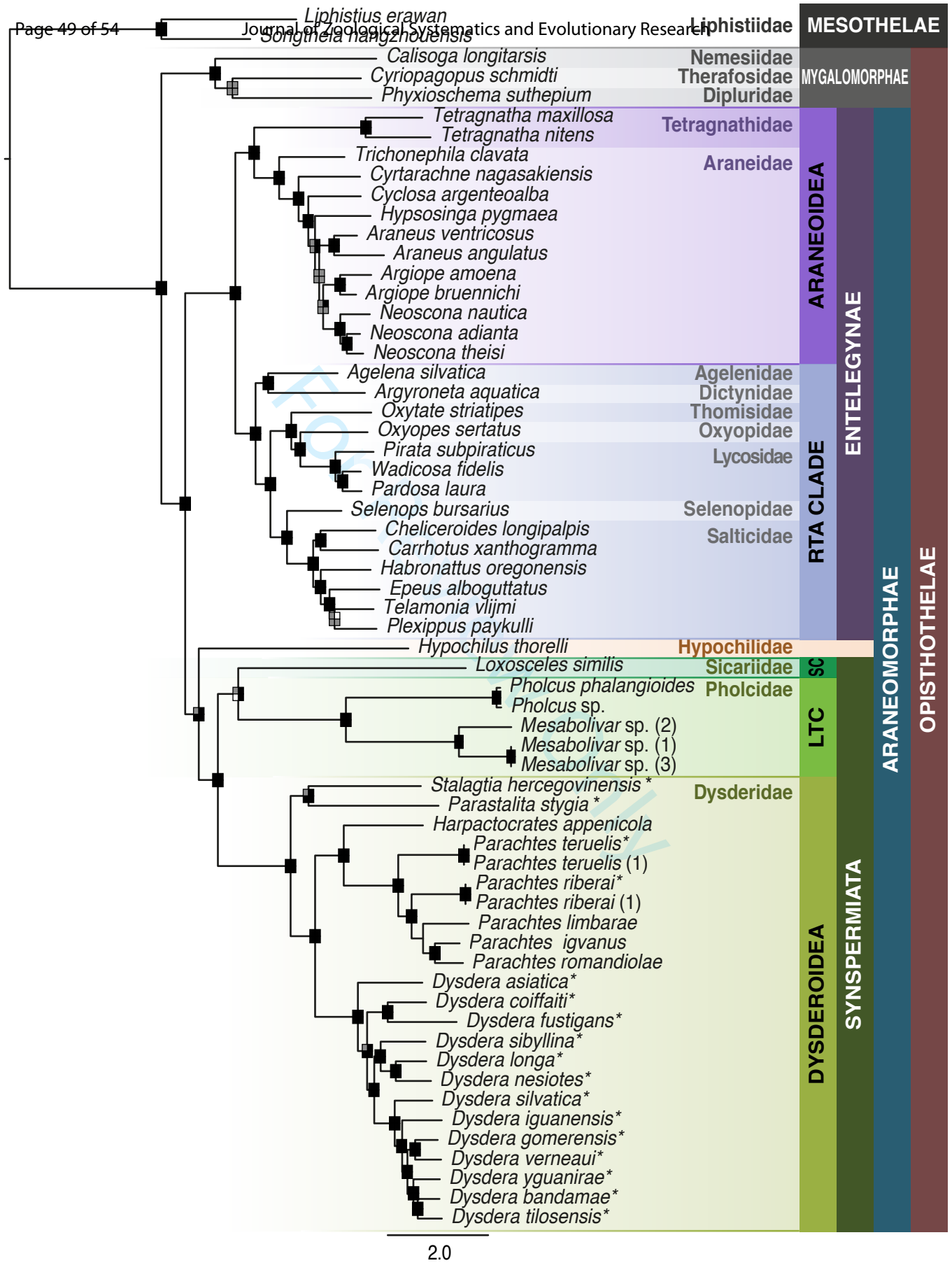
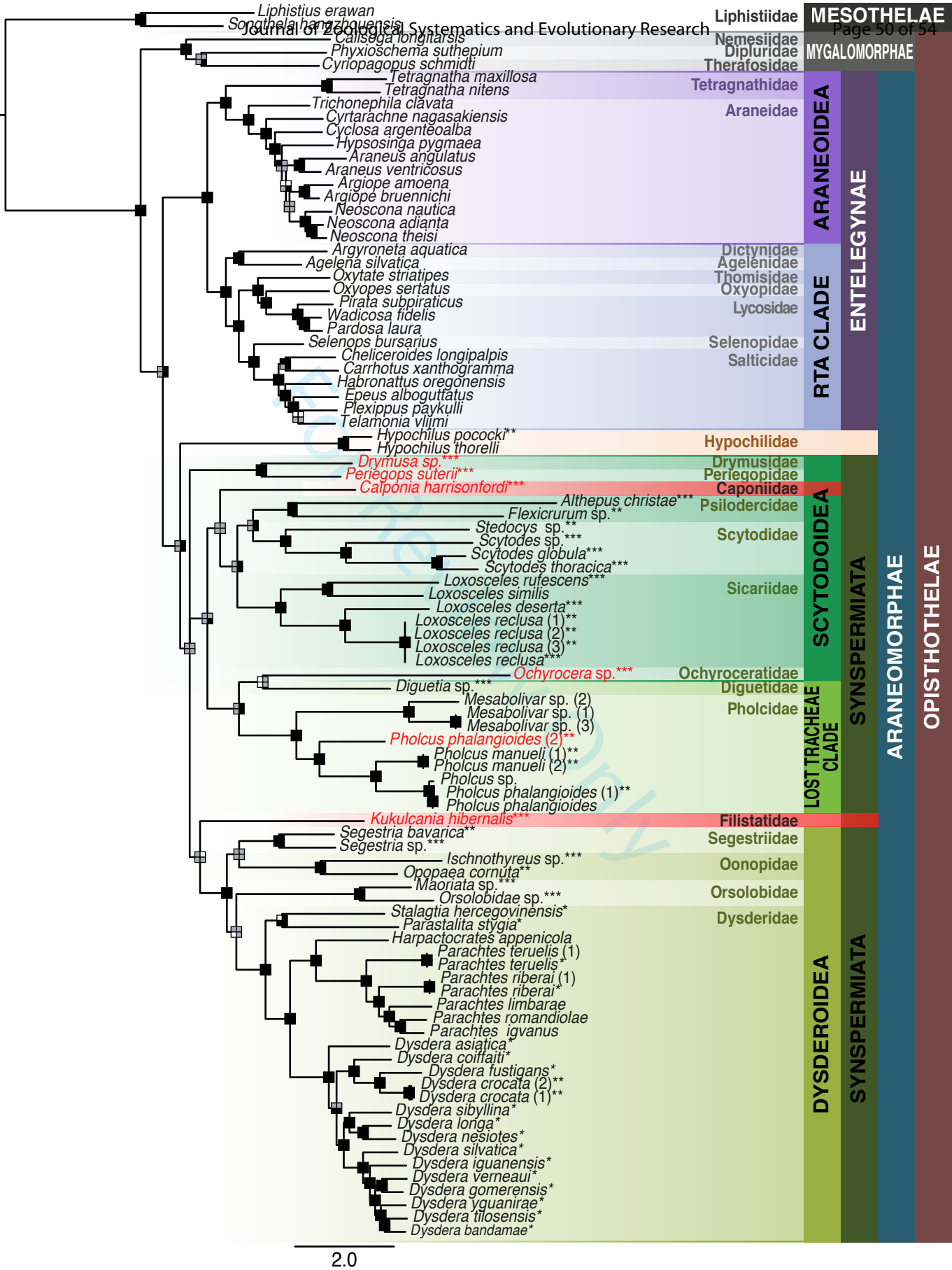


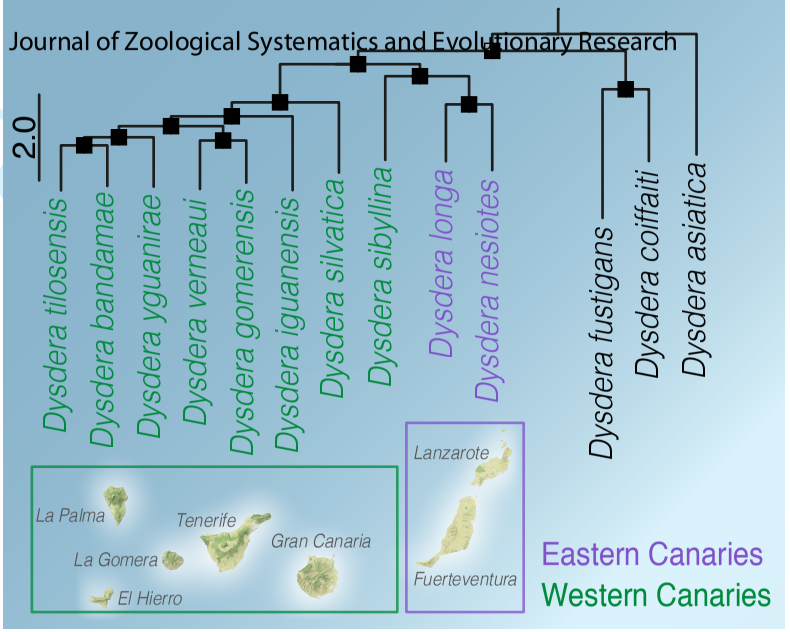
Figure 1. Representatives of the family Dysderidae: A) *Dysdera longa* (Dysderinae), photo credit Pedro Oromí; B) *Dysdera tilosensis* (Dysderinae), photo credit Pedro Oromí; C) *Dysdera silvatica* (Dysderinae), photo credit Pedro Oromí; D) *Dysdera sibyllina* (Dysderinae), photo credit Pedro Oromí; E) *Parastalita stygia* (Rhodinae), photo credit Fulvio Gasparo; F) *Stalagtia hercegovinensis* (Harpacteinae), photo credit Jana Bedek.


NCBI
Araneae
Mitogenomic
(x 45)
In-House project
Dysderidae
Genomic
(x 13)
Transcriptomic
(x 9)
NCBI
Synspermiata
Genomic
(x 1)
Transcriptomic
(x 33)
Trimitomics
NOVOPlasty
Trimmomatic
Bowtie2
Trinity
Velvet / Trinity
Blast
Complete mitogenome
Partial mitogenome
Meta-assembly
MITOS2 Annotation
NCBI
mitogenomes
Dysderidae mitogenomes
recovered from in-house data
> 70% r.l.
Synspermiata mitogenomes
recovered from NCBI data
> 30% r.l.
< 30% r.l.
Dataset1 (S1)
Dataset2 (S2)
Dataset3 (S3)


Total
Novoplasty
Bowtie2
Velvet/Trinity
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Dysdera asiatica G
Dysdera bandamae G+
Dysdera gomerensis G+
Dysdera iguanensis G+
Dysdera sibyllina G
Dysdera silvatica G+
Dysdera tilosensis G+
Dysdera verneaui G+
Dysdera yguanirae G
Parachtes riberai G
Parachtes teruelis G
Stalagtia hercegovinensis G
Parastalita stygia G
Dysdera verneaui T+
Dysdera bandamae T+
Dysdera longa T
Dysdera gomerensis T+
Dysdera silvatica T+
Dysdera coiffaiti T
Dysdera tilosensis T+
Dysdera fustigans T
Pholcus manueli (2) T
Pholcus phalangioides (1) T
Hypochilus pococki T
Dysdera nesiotes T
Pholcus manueli (1) T
Dysdera crocata (1) T
Segestria bavarica T
Stedocys sp. T
Loxosceles reclusa (1) T
Loxosceles reclusa (3) T
Dysdera crocata (2) T
Pholcus phalangioides (2) T
Flexicrurun sp. T
Loxosceles reclusa (2) T
Opopaea cornuta T
Kukulcania hibernalis T
Loxosceles rufescens T
Loxosceles reclusa G
Segestria sp. T
Loxosceles deserta T
Orsolobidae sp. T
Althepus christae T
Scytodes sp. T
Ochyrocera sp. T
Maoriata sp. T
Diguetia sp. T
Scytodes thoracica T
Calponia harrisonfordi T
Drymusa sp. T
Scytodes globula T
Ischnothyreus sp. T
Periegops suteriii T
Filistata insidiatrix T
Hypochilus gertschi T
Opopaea sp. T


Liphistius erawan
Songthela hangzhouensis
Calisoga longitarsis
Cyriopagopus schmidti
Phyxioschema suthepium
Tetragnatha maxillosa
Tetragnatha nitens
Trichonephila clavata
Cyrtarachne nagasakiensis
Cyclosa argenteoalba
Hypsosinga pygmaea
Araneus ventricosus
Araneus angulatus
Argiope amoena
Argiope bruennichi
Neoscona nautica
Neoscona adianta
Neoscona theisi
Agelena silvatica
Argyroneta aquatica
Oxytate striatipes
Oxyopes sertatus
Pirata subpiraticus
Wadicosa fidelis
Pardosa laura
Selenops bursarius
Cheliceroides longipalpis
Carrhotus xanthogramma
Habronattus oregonensis
Epeus alboguttatus
Telamonia vlijmi
Plexippus paykulli
Hypochilus thorelli
Loxosceles similis
Pholcus phalangioides
Pholcus sp.
Mesabolivar sp. (2)
Mesabolivar sp. (1)
Mesabolivar sp. (3)
Stalagtia hercegovinensis *
Parastalita stygia *
Harpactocrates appenicola
Parachtes teruelis*
Parachtes teruelis (1)
Parachtes riberai*
Parachtes riberai (1)
Parachtes limbarae
Parachtes igvanus
Parachtes romandiolae
Dysdera asiatica*
Dysdera coiffaiti*
Dysdera fustigans*
Dysdera sibyllina*
Dysdera longa*
Dysdera nesiotes*
Dysdera silvatica*
Dysdera iguanensis*
Dysdera gomerensis*
Dysdera verneaui*
Dysdera yguanirae*
Dysdera bandamae*
Dysdera tilosensis*
Liphistiidae
Nemesiidae
Therafosidae
Diplurididae
Tetragnathidae
Araneidae
Agelenidae
Dictynidae
Thomisidae
Oxyopidae
Lycosidae
Selenopidae
Salticidae
Hypochilidae
Sicariidae
Pholcidae
Dysderidae
MESOTHELAE
MYGALOMORPHAE
ARANEOIDEA
RTA CLADE
ENTELEGYNAE
SC
LTC
DYSDEROIDEA
SYNSPERMIATA
ARANEOMORPHAE
OPISTHOTHELAE
2.0


Liphistius erawan
Songthela hangzhouensis
Callisoga longitarsis
Phyxioschema suthepium
Cyriopagopus schmidti
Tetragnatha maxillosa
Tetragnatha nitens
Trichonephila clavata
Cyrtarachne nagasakiensis
Cyclosa argenteoalba
Hypsosinga pygmaea
Araneus angulatus
Araneus ventricosus
Argiope amoena
Argiope bruennichi
Neoscona nautica
Neoscona adianta
Neoscona theisi
Argyroneta aquatica
Agelena silvatica
Oxytate striatipes
Oxyopes sertatus
Pirata subpiraticus
Wadicosa fidelis
Pardosa laura
Selenops bursarius
Cheliceroides longipalpis
Carrhotus xanthogramma
Habronattus oregonensis
Epeus alboguttatus
Plexippus paykulli
Telamonia vlijmi
Hypochilus pococki**
Hypochilus thorelli
Drymusa sp.***
Periegops suteri***
Calponia harrisonfordi***
Althepus christae***
Flexicrurum sp.**
Stedocys sp.**
Scytodes sp.***
Scytodes globula***
Scytodes thoracica***
Loxosceles rufescens***
Loxosceles similis
Loxosceles deserta***
Loxosceles reclusa (1)**
Loxosceles reclusa (2)**
Loxosceles reclusa (3)**
Loxosceles reclusa***
Ochyrocera sp.***
Diguetia sp.***
Mesabolivar sp. (2)
Mesabolivar sp. (1)
Mesabolivar sp. (3)
Pholcus phalangioides (2)**
Pholcus manueli (1)**
Pholcus manueli (2)**
Pholcus sp.
Pholcus phalangioides (1)**
Pholcus phalangioides
Kukulcania hibernalis***
Segestria bavarica**
Segestria sp.***
Ischnothyreus sp.***
Opopaea cornuta**
Maoriata sp.***
Orsolobidae sp.***
Stalagtia hercegovinensis*
Parastalita stygia*
Harpactocrates appenicola
Parachtes teruelis (1)
Parachtes teruelis*
Parachtes riberai (1)
Parachtes riberai*
Parachtes limbarae
Parachtes romandiolae
Parachtes igvanus
Dysdera asiatica*
Dysdera coiffaiti*
Dysdera fustigans*
Dysdera crocata (2)**
Dysdera crocata (1)**
Dysdera sibyllina*
Dysdera longa*
Dysdera nesiotes*
Dysdera silvatica*
Dysdera iguanensis*
Dysdera verneaui*
Dysdera gomerensis*
Dysdera yguanirae*
Dysdera tilosensis*
Dysdera bandamae*
Liphistiidae
Nemesiidae
Dipluridae
Theraphosidae
Tetragnathidae
Araneidae
Dictynidae
Agelenidae
Thomisidae
Oxyopidae
Lycosidae
Selenopidae
Salticidae
Hypochilidae
Drymusidae
Periegopidae
Caponiidae
Psilodercidae
Scytodidae
Sicariidae
Ochyroceratidae
Diguetidae
Pholcidae
Filistatidae
Segestriidae
Oonopidae
Orsolobidae
Dysderidae
MESOTHELAE
MYGALOMORPHAE
ARANEOIDEA
RTA CLADE
ENTELEGYNAE
SCYTODOIDEA
LOST TRACHEAE CLADE
SYNSPERMIATA
DYSDEROIDEA
SYNSPERMIATA
ARANEOMORPHAE
OPISTHOTHELAE
2.0


2.0
Dysdera tilosensis
Dysdera bandamae
Dysdera yguanirae
Dysdera verneaui
Dysdera gomerensis
Dysdera iguanensis
Dysdera silvatica
Dysdera sibyllina
Dysdera longa
Dysdera nesiotes
Dysdera fustigans
Dysdera coiffaiti
Dysdera asiatica
La Palma
Tenerife
La Gomera
Gran Canaria
El Hierro
Lanzarote
Fuerteventura
Eastern Canaries
Western Canaries