Mitonuclear discordance in wolf spiders: Genomic Evidence for Species Integrity and Introgression

Vladislav Ivanov[1*], Kyung Min Lee[1], Marko Mutanen[1]

[1]*Department of Ecology and Genetics, University of Oulu, PO Box 3000, Oulu, 90014, Finland*

Key words: DNA barcoding; Lycosidae; molecular taxonomy; endosymbionts; interspecific hybridization; RAD sequencing

[*]Corresponding author:

Vladislav Ivanov

Department of Ecology and Genetics, University of Oulu, PO Box 3000, Oulu, 90014, Finland

Tel. +358 41 369 95 53

Fax. +358 8 344 084

vladislav.ivanov@oulu.fi

Mitonuclear discordance in wolf spiders

Abstract

Systematists and taxonomists have benefited greatly from the emergence of molecular methods. Species identification has become straightforward through DNA barcoding and the rapid build-up of massive DNA barcode reference libraries. In animals, mitonuclear discordance can significantly complicate the process of species identification and delimitation. The causes of mitonuclear discordance are either biological (e.g. introgression, incomplete lineage sorting, horizontal gene transfer, androgenesis) or induced by operational factors (e.g. human error with specimen misidentification or incorrect species delimitation). Moreover, endosymbionts may play an important role in promoting fixation of mitochondrial genomes. Here, we study the mitonuclear discordance of wolf-spiders species (Lycosidae) (independent cases from *Alopecosa aculeata* and *Pardosa pullata* groups) that share identical COI DNA barcodes. We approached the case utilizing double-digest Restriction-Site Associated DNA sequencing (ddRADseq) to obtain and analyze genomic-scale data. Our results suggest that the observed cases of mitonuclear discordance are not due to operational reasons but result from biological processes. Further analysis indicated introgression and that incomplete lineage sorting is unlikely to have been responsible for the observed discrepancy. Additional survey of endosymbionts provided ideas on further research and their role in shaping mitochondrial DNA distribution patterns. Thus, ddRADseq grants an efficient way to study the taxonomy of problematic groups with insight into underlying evolutionary processes.

Introduction

During the last two decades, molecular tools, and especially high-throughput sequencing technologies, have become widely adopted in all fields of biological studies, including population genetics, conservation, evolutionary biology and systematics. New methods have allowed insight into complicated histories of species interactions, their evolution and population dynamics (e.g. Wagner *et al.* 2013; Boyer *et al.* 2014; Escudero *et al.* 2014; Takahashi *et al.* 2014; Eaton *et al.* 2015; Saenz-Agudelo *et al.* 2015; Catchen *et al.* 2017). Under different frameworks, the discrepancy between phylogenies of nuclear and cytoplasmic DNA (mitochondrial or plastid DNA) has been a subject for scrutiny for a long time. Such discrepancy is usually referred to as mitonuclear or cytonuclear discordance (Di Candia & Routman 2007; Toews & Brelsford 2012; Li *et al.* 2016; Meyer *et al.* 2016).

Mitonuclear discordance is defined as a significant difference in levels of differentiation between nuclear and mitochondrial markers, where either mitochondrial DNA (mtDNA) is more structured than nuclear DNA (nDNA) or *vice versa* (Toews & Brelsford 2012). It is observed among all major taxa of animals (Toews & Brelsford 2012) and seriously complicates species delimitation and identification (Papakostas *et al.* 2016). Mitonuclear discordance may or may not be linked to geographical structure of populations (phylogeographic discordance) (Toews & Brelsford 2012). Resolving the extent of mitonuclear discordance is particularly relevant within the DNA barcoding framework of the animal kingdom that solely relies on mtDNA. While an average of 95% success in differentiating species is reported for the animal DNA barcoding region (Hebert *et al.* 2016), DNA barcoding has frequently revealed obvious cases of mitonuclear discordance, which in

lack of nuclear data are often referred to as “DNA barcode sharing” or “deep intraspecific DNA barcode divergences” (Hausmann *et al.* 2013; Mutanen *et al.* 2016).

Several reasons that might explain mitonuclear discordance have been proposed, but many of them are merely conjectural and hard to quantify (Toews & Brelsford 2012; Bonnet *et al.* 2017). The currently recognized reasons behind mitonuclear discordance include horizontal gene transfer (HGT e.g. Bergthorsson & Palmer 2003; Soucy *et al.* 2015), androgenesis (Hedtke & Hillis 2011), unresolved phylogenetic polytomy (e.g. Caraballo *et al.* 2012), mitochondrial pseudogenes in nuclear DNA (NUMTs) (Leite 2012; Song *et al.* 2014), incomplete lineage sorting (ILS) and introgression (Toews & Brelsford 2012; Mutanen *et al.* 2016). Mitonuclear discordance may also result from operational factors (Mutanen *et al.* 2016), of which taxonomic oversplitting of species is the most obvious one (Funk & Omland 2003; Hausmann *et al.* 2013; Zahiri *et al.* 2014; Raupach *et al.* 2015). Ross (2014) estimated that approximately 10% of cases of species-level non-monophyly in COI gene trees (i.e. putative cases of mitonuclear discordance) could be explained by operational causes, whereas Mutanen *et al*. (2016) observed a significantly higher rate of 58.6%.

The presence of NUMTs, HGT and androgenesis are supposed to result in elevated or overestimated haplotype diversity (e.g. Song *et al.* 2008). Therefore, we consider ILS and introgression to be more likely causes of the haplotype sharing observed among the studied wolf spiders. Both ILS and introgression may result in presence of two or more distinct haplotypes in a species (=genetic polymorphism) (Melo-Ferreira *et al.* 2012). In cases of young species, however, ILS may result in mitonuclear discordance because not enough time has elapsed for differentiation of the lineages to occur. Introgression may, however, yield

79 exactly the same pattern, making distinguishing these causes from each other difficult (Funk
80 & Omland 2003).

81 Introgression is now recognized as quite a widespread process. Its presence has been
82 detected among various taxa of both vertebrate and invertebrate animals as well as plants
83 (e.g. Eaton *et al.* 2015; Cahill *et al.* 2015; Huang 2016). Additionally, mtDNA can become
84 introgressed even in the absence of significant nuclear genome introgression (Chan & Levin
85 2005). If mitochondrial introgression has been prevalent throughout a species' history,
86 discordance between mitochondrial and nuclear topologies can be expected. Several studies
87 have considered ancient introgression as a potential explanation for mitonuclear
88 discordance (Shaw 2002; Gómez-Zurita & Vogler 2006; Linnen & Farrell 2007; Sota & Vogler
89 2009). One reason for this phenomenon is that introgression events become progressively
90 more difficult to detect as time since hybridization accumulates and geographic signals of
91 introgression (e.g., shared haplotypes in areas of sympatry) are eroded by range changes
92 and mutation (Funk & Omland 2003). Furthermore, in many invertebrate animals,
93 endosymbiotic bacteria may play an important role in shaping variability of mtDNA through
94 rapid spread of introgressed haplotypes and fixation of introgressed haplotypes (Hurst &
95 Jiggins 2000). *Wolbachia* is one of the best studied bacteria in this regard (Jiggins 2003; Hurst
96 & Jiggins 2005; Smith *et al.* 2012) but other species of bacteria might also contribute to the
97 diversity of mtDNA of their hosts (Stefanini & Duron 2012; Curry *et al.* 2015).

98 We attempted to shed light on the causes of mitonuclear discordance in the two sympatric
99 groups of wolf spiders using double-digest Restriction-Site Associated DNA sequencing
100 (ddRADseq). RAD methods represent the reduced-representation methods (Davey *et al.*

2011) and the resulting RAD tags provide a comprehensive overview of the entire genome. Mitonuclear discordance, in form of COI haplotype uniformity across species, was detected between four species in the *Pardosa pullata* group (*P. fulvipes*, *P. prativaga*, *P. riparia*, *P. sphagnicola*) and between *Alopecosa aculeata* and *A. taeniata*. In both groups, the species are clustered together, and the minimum interspecific divergence in COI is significantly smaller than maximum intraspecific divergence, and the species lack reciprocal monophyly (Fig. 1). In general, COI divergences have been reported to be relatively low in *Alopecosa* and *Pardosa* (Sim *et al.* 2014; Blagoev *et al.* 2016).

We hypothesized that if taxonomic oversplitting was involved, data from hundreds of thousands of loci will not provide evidence for the taxonomic integrity of the species. If species represent distinct evolutionary lineages, nDNA data analysis is supposed to show no conflict with the established taxonomical view of the studied groups and therefore, the mitonuclear discordance can be viewed as not being due to the operational factors. The reasons for mitonuclear discordance might then lie in introgression or ILS.

In order to have a preliminary view on the presence of ILS, we used p-distances measured between pairs of individuals in COI and nDNA datasets, and then calculated the ratio of COI p-distance values to those of nDNA. Our prediction was that the ratio values among species with ILS should be similar to the ratio values of species without ILS or introgression, i.e. to species that could be safely identified by COI. In introgressed species, the ratio of COI p-distances to nDNA data should be significantly lower in comparison to species with no introgression, i.e. with ILS or clearly distinct species. We predict that under neutral theory the changes in nuclear and mtDNA are accumulated with different speed, but the ratio

123 between the magnitude of accumulation (based on p-distances in our case) is expected to be
124 the same regardless of divergence time between the lineages. Theoretically, COI could have
125 become fixed in ancestral lineage and retained unchanged in the diverged populations for
126 stochastic reasons or stabilizing selection. However, mtDNA shows reduced effective
127 population size, absence of recombination, accumulated mutation rate and overall tendency
128 for rapid fixation (Hurst & Jiggins 2005; Rubinoff *et al.* 2006), for which reasons we consider
129 this scenario unlikely. Nevertheless, our approach may provide only indirect and hence
130 tentative support for either scenario.

131 We continued by examining the presence of four different species of endosymbiotic bacteria
132 among the study specimens to understand the magnitude of bacterial infections in these
133 species. The presence of endosymbionts is not conclusive *per se*, but it can bring additional
134 insight to their possible influence on mtDNA haplotype distributions as well as help to define
135 the direction of future research.

Materials and Methods

Focal Taxa

We focused on two groups of Lycosidae spiders exhibiting mitonuclear discordance based on previous observations. These included *Alopecosa taeniata* (C.L. Koch, 1835) and *A. aculeata* (Clerck, 1757) from the *Alopecosa pulverulenta* (Clerck, 1757) group and *Pardosa sphagnicola* (Dahl, 1908), *P. riparia* (C.L. Koch, 1833), *P. pullata* (Clerck, 1757), *P. prativaga* (L. Koch, 1870) and *P. fulvipes* (Collett, 1876) from the *Pardosa pullata* (Clerck, 1757) group (Holm & Kronestedt 1970; Kronestedt 1990). Moreover, the following species of *Pardosa* were included in the analyses: *P. agricola* (Thorell, 1856), *P. amentata* (Clerck, 1757), *P. eiseni* (Thorell, 1875), *P. hyperborea* (Thorell, 1872), *P. lugubris* (Walckenaer, 1802), *P. maisa* Hippa and Mannila, 1982, *P. palustris* (Linnaeus, 1758). *Pisaura mirabilis* (Clerck, 1757) (Pisauridae), representative of the sister group of Lycosidae (Murphy *et al.* 2006), was selected to serve as a primary outgroup for general COI analysis (Fig. 1). We also included *Trochosa spinipalpis* (F. O. P.-Cambridge, 1895) (Lycosidae) as a more closely related outgroup to *Alopecosa*. Altogether 54 specimens were included into the final analysis (20 for the *Alopecosa* case and 34 for *Pardosa*). All specimens for the target groups were collected during the season of 2015. For collection data, please see the Supporting Information (Table S1).

Mitochondrial DNA Sequencing

Before tissue sampling, a 96-well plate was pre-filled with absolute alcohol. Each specimen was taken from the tube with sterile forceps and one leg was detached and put into the well.

158 The specimens were photographed, and the full collection and taxonomic information was
159 entered into the BOLD database in the 'Arachnida of Finland' project. The plate was sent to
160 CCDB (Canadian Centre for DNA barcoding) where DNA was extracted and the mitochondrial
161 COI gene (654 bp) was sequenced using standard protocols (deWaard *et al.* 2008). DNA
162 barcode sequences with full collection, taxonomic and laboratory information are publicly
163 available in the BOLD dataset DS-LYCOSRAD.

164

165 *ddRADseq Library Preparation*

166 Genomic DNA (gDNA) extraction was performed using DNeasy Blood & Tissue Kit (Qiagen)
167 according to the manufacturer's instructions. One to four legs were taken from each
168 specimen. The degradation level of gDNA was checked with gel electrophoresis, and was
169 found to be considerably degraded in several cases. In order to improve the quality of gDNA,
170 we performed whole genome amplification for all samples using REPLI-g Mini Kit (Qiagen)
171 following the standard protocol provided by the manufacturer. Whole genome amplification
172 (WGA) methods in general, and REPLI-g protocols in particular, are known to cause biases.
173 However, these are negligible in most cases when WGA was used for increasing the amount
174 of DNA for analysis (e.g. Barker *et al.* 2004; Pinard *et al.* 2006; Han *et al.* 2012). Moreover
175 REPLI-g kits were successfully used in many RADseq studies with no consequences for the
176 analysis (e.g. Rheindt *et al.* 2014; Blair *et al.* 2015; Burford Reiskind *et al.* 2016).

177 Library preparation was implemented following protocols described in Peterson *et al.* (2012)
178 and DaCosta & Sorenson (2015). In order to perform ddRADseq, gDNA was digested with
179 two restriction enzymes, one for frequent and another for rare cutters. We tested several

enzyme pairs to find the best combination that worked efficiently in the target species. The combination pair *Pst*I and *Mse*I proved to be the most efficient for both target groups and outgroup species. The samples were then ligated to adapters designed for the *Pst*I-*Mse*I pair of restriction enzymes. Then samples were pooled in four sub-pools based on the DNA concentration measurements by PicoGreen Kit (Molecular Probes). Following ligation, size selection was performed by automated size-selection technology, Blue Pippin (Sage Science, 2% agarose cartridge). The size selected library, with a mean of 300 bp, was eluted in 40 µl volumes. The selected fragments were amplified with Phusion High-Fidelity PCR Master Mix (Finnzymes). The PCR products were purified with AMPure XP magnetic beads (Agencourt). The quality, size distribution, and concentration of the sub-pools were measured with MultiNA (Shimadzu), and if an excess of non-target fragments was present, purification steps were repeated additional times. Finally, the sub-pools were combined into one library in equimolar amounts, and the library was sequenced on an Illumina HiSeq 2500 machine in FIMM, 100 PE (Institute for Molecular Medicine Finland). DNA reads from ddRAD sequencing are available at the NCBI Sequence Read Archive (BioProject ID: PRJNA345307).

ddRADseq Data Bioinformatics

The quality control was performed with FastQC (Andrews 2010). Paired-end reads were assembled *de novo* using pyRAD version 3.0.64 (Eaton 2014). Briefly, the pyRAD pipeline allows filtration (VSEARCH) and alignment (MUSCLE) of the reads both within and among species. The pipeline is capable of discarding low quality reads based on Phred scores (< 20 converted to Ns and reads with Ns > 4 were discarded), and number of haplotypes (> 2 for

diploids). As an additional filtering step, consensus sequences with excessive undetermined or heterozygous sites (> 3) were discarded. The most important parameters are minimum depth of coverage (d), clustering threshold (c), and minimum number of individuals per locus (m), with range variation: d = 3; c = 80, 85 and 90; m = 6, 9 and 10. Minimum depth of clusters (d) is a statistical base call at each site in a cluster, i.e. how many reads contain the same base. The higher the value that is chosen, the more data will be discarded. For the ddRADseq data d = 3 is usually sufficient while taking into account other filtering steps. In a trial with d = 6 the amount of recovered loci was not significantly different, thus we used the data set of d = 3 in the subsequent analysis. The clustering threshold (c) is a percentage of similarity between samples, i.e. reads that have a sufficient percentage of similarity will be included, and others are discarded. This step establishes locus homology among individuals. Each locus was aligned and a filter was used to exclude potential paralogs. The paralog filter removes loci with excessive shared heterozygosity among samples. The justification for this filtering method is that shared heterozygous SNPs across species are more likely to represent a fixed difference among paralogs than shared heterozygosity within orthologs among species. We applied a strict filter that allowed a maximum of three samples to show heterozygosity at a given site (p = 3). Finally, minimum number of individuals per locus (m) allows filtering of data and inclusion of only the loci that are shared by a given number of samples. Hence, the lower the number of samples, the more loci will be obtained. We compiled data matrices separately for *Alopecosa* and *Pardosa*. For the subsequent analyses, individuals with low sequence quality were eliminated (Tables 1 and 2 and Table S1, Supporting Information). All strict filtering steps allowed additional control for possible WGA bias.

225

226 *Phylogenetic Analyses*

227 Phylogenetic analyses were conducted to reveal historical relationships among taxa and to
228 test the validity of prevailing species hypotheses. Phylogenetic trees were built for both the
229 COI gene and ddRAD data, the latter separately for both *Alopecosa* and *Pardosa*. Maximum
230 likelihood (ML) trees were inferred in RAxML version 7.2.8 (Stamatakis 2014) with rapid
231 bootstrap support (BS) estimated from 500 replicate searches from random starting trees
232 using the GTR+GAMMA nucleotide substitution model.

233

234 *Four-taxon D-statistics*

235 For analysis of introgression among species in *Alopecosa* and *Pardosa* groups, we used the
236 four-taxon D-statistic (Durand *et al.* 2011). The analysis was performed with pyRAD version
237 3.0.64 (Eaton 2014). The test is based on the assumption of a true four-taxon asymmetric
238 phylogeny (((P1, P2) P3,) O). All sites considered in the alignment of sequences from these
239 taxa must be either mono- or biallelic, with the outgroup defining the ancestral state 'A'
240 relative to the derived state 'B'. In cases when there are two alleles in a site, the possible
241 combinations are ABBA and BABA. The D-statistic compares the occurrence of these two
242 discordant site patterns, representing sites where an allele is derived in P3 relative to the
243 outgroup (O), and is derived in one but not both of the sister lineages P1 and P2. These
244 discordant sites can arise through the sorting of ancestral polymorphisms. In the absence of
245 introgression (possibly lineage sorting under drift), the frequencies of these two outcomes
246 are expected to be equal. We tested *Alopecosa* and *Pardosa* groups separately for all

possible combinations. For the test, 1000 bootstrap replicates were performed to measure the standard deviation of the D-statistics. Significance was evaluated by converting the Z-score (which represents the number of standard deviations from zero from D-statistics) into a two tailed P-value, and using $\alpha = 0.01$ as a conservative cutoff for significance after correcting for multiple comparisons using Holm-Bonferroni correction. A significant Z-score (> 3.55) suggests that introgression might have occurred between species.

Testing Admixture with TreeMix

We used the program TreeMix version 1.12 (Pickrell & Pritchard 2012) to jointly estimate a tree topology with admixture using SNP frequency data. A single biallelic SNP was randomly sampled from each variable locus that contained data across all individuals, assuming that all SNPs were independent, yielding a total of 7,144 biallelic SNPs for *Pardosa*. Support for the tree topology was assessed by means of 1000 bootstrap replicates using a block size of 10 SNPs. Trees were rooted with *P. pullata*. We then built trees allowing for different numbers of migration events (1 to 10). We also ran the TreeMix analysis for the *Alopecosa* species. We used the following parameters: migration events in the tree (-m), 1 to 10, 100, and 1000; linkage disequilibrium (-k), 10 and 50; build the ML tree (-i); rooting (-root) to *A. pulverulenta*; bootstrapping (-bootstrap) with 1000 replicates for judging the confidence in a given tree topology.

Detecting ILS using p-distances

We calculated the p-distances in all possible pairs of species for which we had both COI partial sequences and ddRADseq datasets. The p-distances were calculated in MEGA 7 (Kumar *et al.* 2016). Then, we divided the resulting value of COI p-distances by ddRADseq p-distances for each pair of species compared. The resulting ratios were divided into two groups: with mitonuclear discordance (MD-group, our target species group) and and without it (No-MD-group, the rest of the analysed species, i.e. reference species). The ratios of two groups were compared using the Welch Two Sample t-test in R with 1000 permutations using the custom script. In order to visualize the differences between these two groups we used the 'boxplot' function in RStudio version 1.0.136 (RStudio Team 2015).

Additional analysis

We also used other approaches to investigate mitonuclear discordance in the study groups. We used STRUCTURE v 2.3.1 to estimate the gene flow between the species and SVDquartets to test species hypothesis. The protocol descriptions and results can be found in Supporting information, Figures S1 and S2 correspondingly.

In order to check for the presence of bacteria and to study their strain variability across the study species, we sequenced markers for four different bacteria: *Wolbachia*, *Cardinium*, *Spiroplasma* and *Rickettsia*. Detailed information of PCR protocols and primer sequences can be found in Table S2 and endosymbionts occurrence summary can be found in Table S3. Neighbor-joining trees of the endosymbionts' sequences are represented in Figure S3, (Supporting information).

289 **Results**

290 *Patterns of COI in Lycosidae*

291 The maximum likelihood (ML) tree of the more inclusive dataset and neighbor joining (NJ)
292 tree of all Lycosidae barcoded from Finland supported the presence of widespread DNA
293 barcode sharing among two species of *Alopecosa* and four species of *Pardosa* as observed
294 earlier (Fig. 1). The non-monophyly of each of these species was confirmed with the
295 Monophylizer tool, available at http://monophylizer.naturalis.nl/ (Mutanen *et al*. 2016). In
296 the ML tree, *A. taeniata-aculeata* was split into three well-supported subgroups, one with
297 only specimens of *A. aculeata* and two groups with *A. aculeata-taeniata* intermixed. The
298 maximum COI p-distance between *A. taeniata* and *A. aculeata* was 1.5%, while the COI
299 intraspecific distances were 1.3% for *A. taeniata* and 0.9% for *A. aculeata*. The *Pardosa*
300 *pullata* group was supported to be monophyletic (BS 85%) and was divided into two
301 subgroups: *P. pullata* (BS 91%) and *P. fulvipes-prativaga-riparia-sphagnicola* complex (Pf-
302 complex) (BS 87%) (Fig. 1). *P. pullata* showed 1.9% minimum COI p-distance to the Pf-
303 complex, which in turn had a maximum COI p-distance divergence of 0.7% (Table 4).

304

305 *Optimization of ddRADseq Loci Parameters and Phylogenies*

306 To examine the sensitivity of the phylogenetic inference to the parameters used to identify
307 loci and create nucleotide matrices, we generated two data matrix combinations for
308 *Alopecosa* and six for *Pardosa* by changing the values for minimum number individuals per
309 locus (m) and clustering thresholds (c) with pyRAD. The impact of 'm' was tested with a
310 clustering threshold of c = 80 for *Pardosa*, while the clustering threshold was investigated

311 with m = 6 for both groups (see Table 2). The total number of loci ranged from 23 to 7,258
312 between the six data matrices in *Pardosa*, demonstrating the dramatic effect of parameter
313 selection on the amount of data. Data assemblages that maximized the number of
314 individuals per locus contained relatively few loci and SNPs, but at the same time reduced
315 the proportion of missing data (e.g., c80m18_pardosa; Table 2). The different clustering
316 thresholds had a significant effect on the total number of loci (range 12,641–13,757 loci),
317 variable sites (184,787-207,608) as well as the number of phylogenetically informative sites
318 (62,619-70,437) in *Alopecosa* (Table 2). Resulting data matrices analysed in RAxML produced
319 overall similar tree topologies in most trials (not shown), but 'c80m18_pardosa' (i.e. with no
320 missing data) produced a poorly resolved and deviant tree as a result of scarcity of
321 phylogenetic information. The final ddRAD data of *Alopecosa* from 20 individuals yielded ca.
322 2.4 million base pairs, 12,641 loci, and 12,423 unlinked SNPs with average cluster depth of
323 115.74. For 34 specimens of *Pardosa*, a total of ca. 1.4 million base pairs, 7,157 loci, and
324 7,109 unlinked SNPs from 34 individuals were retrieved with average cluster depth of 102.1
325 (Table 2). Unlike COI, ML trees based on ddRADseq data supported the species integrities in
326 both *Alopecosa* and *Pardosa* with 100% bootstrap support values (Fig. 2-3). *P. pullata* was
327 placed in basal position with respect to the Pf-complex.

328

329 *Four-Taxon D-Statistics*

330 D-statistics was used for detection of introgression. High Z-score values were observed in
331 many tests in both *Alopecosa* and *Pardosa*, suggesting that possible hybridization and
332 resulting introgression has occurred in the past (Table 3). Z-scores with significant p-values

ranged from 4.18 to 14.18 for *Alopecosa* and 3.78 to 12.95 for *Pardosa* (see Table S4, Supporting Information). In the *Alopecosa* group, evidence for introgression was found to have occurred in five cases between *A. aculeata* and *A. taeniata,* and in one case between *A. aculeata* and *A. pulverulenta*. The mean number of loci available for testing *Alopecosa* was 26, and the mean percentage of discordant sites was 0.29. In *Pardosa*, we detected significant values in 38 cases: *P. prativaga* showed the highest degree of introgression with *P. riparia* (12.8% in nSig/n), while *P. pullata* was rarely introgressed with other species (1.4%). The mean number of loci in the comparisons for *Pardosa* was 65, and the mean percentage of discordant sites was 0.21.

TreeMix Analysis

We used ancestry graphs implemented in TreeMix to further identify patterns of divergence and migration within *Pardosa*. Using 7,144 SNPs, we estimated a maximum likelihood tree (Fig. 4) rooted with *P. pullata*, which was chosen because it does not show DNA barcode sharing based on analysis of COI. By sequentially adding migration events, we found significant improvement in fit for up to three events, resulting in the lowest residuals and standard error relative to other trees. For two of these migration events, we inferred significant levels of exchange between species: from *P. sphagnicola* to *P. prativaga* (29.1%), and from *P. sphagnicola* to *P. fulvipes* (16.4%; see Fig. 4). The TreeMix result also supported introgression from *P. fulvipes* to *P. riparia* (10.0%). However, our result did not support any introgression between *P. prativaga* and *P. riparia*, since all of the possible tree topologies

involving these species were insignificant. The TreeMix analysis did not return significant results for gene flow between the *Alopecosa* species (not shown).

Divergence ratio comparison between mitochondria and nuclear loci

The relationship of mean p-distance ratios between COI and ddRADseq data for species with mitonuclear discordance and without it are shown in Figure 5. Within MD-group, the mean p-distance of COI between species ranged from 0.2% to 0.7%, while that of the No-MD-group ranged from 2.1% to 7.3%. In ddRAD, the mean p-distances ranged from 2.7% to 3.5% in MD-group and from 1.9% to 5.0% in No-MD-group . *P. pullata* does not belong to the MD-group , but the species is genetically close to those species in both mitochondrial and genomic data (p-distance 1.9-2.2% in COI, 3.5-3.7% in ddRAD (see Table 4), and revealed signs of introgression with other *Pardosa* species. The ratio of COI p-distances to ddRADseq p-distances in No-MD-group ranged from 0.51 to 2.36 while in MD-group corresponding values ranged from 0.07 to 0.42. The mean values of p-distance ratios for No-MD-group and MD-group were 1.22 and 0.19 respectively. The Welch Two sample t-test with 1,000 permutations revealed significant ($p < 0.01$), which suggests the absence of genome-scale ILS among the studied species.

Discussion

In our study, we focused on two different cases of mitonuclear discordance in wolf spiders, one group having two species, another as many as four. By using *Alopecosa* and *Pardosa* as representative cases, we demonstrated the potential of ddRAD sequencing to provide

evidence of species integrity in the presence of mitonuclear phylogenetic conflicts. We studied the possibility of taxonomic inaccuracy, and then proceeded to examine if introgression had taken place between the taxa. Our results indicate that mitonuclear discordance is likely explained by historical or ongoing introgression, whereas incomplete lineage sorting appears as a less likely cause. Furthermore, taxonomic inaccuracy and other operational causes could be ruled out. As being present, endosymbiotic bacteria may have had a role in fixation of mtDNA across species but this cannot be confirmed at current stage of research.

Role of Operational Causes

Misassignment of specimens into species is clearly not responsible for mitonuclear discordance. Maximum likelihood trees based on nDNA data for both *Alopecosa* and *Pardosa* groups strongly suggest that all the species included in the study can be considered natural, monophyletic lineages with no or limited gene exchange between them. Incongruence between COI and ddRAD data was shown to be true and the possibility of taxonomical oversplitting of species could be rejected. The same idea comes from the results of SVDquartets analysis where all studied groups were estimated as distinct species (Fig. S2). Morphological characteristics can be efficiently used to distinguish between adult specimens, although differences are slight and reliable identification requires deeper taxonomic scrutiny. Moreover, in three cases of *Pardosa*, specimens were initially misidentified by the authors and later blindly validated by an experienced expert. These results suggest that biological rather than operational mechanisms are likely to be responsible for mitonuclear phylogenetic discordance in the studied wolf spiders.

Occurrence of Introgression between the Species

Species with high levels of introgression are supposed to show close overall relatedness and intermixing of species in recently introgressed genes (Eaton *et al.* 2015). Historical or ongoing introgression often revealed by discordant patterns between mitochondrial and nuclear phylogenies (Linnen & Farrell 2007; Papakostas *et al.* 2016; Suchan *et al.* 2017). Unlike the mitochondrial phylogeny, the nDNA phylogeny showed a clear distinction between the target species and did not support the idea of introgression at the genomic level in either *Alopecosa* or *Pardosa*. Additional analysis in STRUCTURE corroborates this idea (Fig. S1). However, four taxon D-statistic tests as well as TreeMix results (for *Pardosa* only) in both *Alopecosa* and *Pardosa* groups returned positive results (see Table 3 and Fig. 4) and suggested introgression had occurred in some loci between the species. All tested species are known to be sympatric, morphologically and ecologically close to each other and there are overlaps in their ranges of areas (Holm & Kronestedt 1970). Hence, the introgression through occasional hybridization has been possible over time. D-statistics suggested that within the *Alopecosa* group the direction of gene flow was from *A. taeniata* to *A. aculeata* and from *A. pulverulenta* to *A. aculeata*, the first case being the strongest based on Z-scores (see Table 3). Such unidirectional introgression has been observed before. For example, in European hares (Lepidae), introgression was observed to have occurred from one species (*Lepus timidus*) to four other species, but not in the opposite direction (Melo-Ferreira *et al.* 2012). The admixture between *A. taeniata* to *A. aculeata* might be one of the reasons we observe the mitonuclear discordance now, because soon after an

introgression event(s) there could have been a fixation of mtDNA haplotypes due to stochastic reasons, adaptive mitochondrial introgression or endosymbionts.

Despite the fact that the two tests (D-statistics and TreeMix) did not return identical results regarding the introgression between *Pardosa* species*,* introgression has clearly occurred between some of them. Both tests are congruent in that the direction of gene flow was from *P. sphagnicola* to *P. fulvipes* and from *P. fulvipes* to *P. riparia*. Introgression between *P. sphagnicola* and *P. prativaga* was detected in TreeMix, but not in D-statistics. The TreeMix analysis suggested the strongest admixture between these two species, *P. sphagnicola* being a donor and *P. prativaga* being a recipient, while the D-statistics did not detect any introgression signals between the species. Another incongruent case is *P. pullata*. D-statistics suggested this species was one of the most frequently observed recipients of introgressed genetic material from *P. sphagnicola*, *P. fulvipes* and *P. riparia*, while TreeMix did not support this conclusion. Specific introgression patterns were not always consistent across different analyses. This could be expected given the complex species histories inferred, absence of closely related genomes for alignment, large amount of missing data inherent to RAD data sets (Arnold *et al.* 2013; Davey *et al.* 2013), and limited power to detect introgression after substantial genetic drift in these species complexes (Patterson *et al.* 2012).

Incomplete Lineage Sorting

Because of features characteristic to mitochondrial genome, mtDNA based markers are expected to show rapid lineage sorting permitting efficient discrimination between species (Hebert *et al.* 2003). Rapid lineage sorting is expected to be further promoted by recurrent

444 selective sweeps characteristic to mtDNA (Jiggins 2003). Conversely, incomplete lineage
445 sorting suggests young age of species. Short distances and incomplete lineage sorting in
446 mtDNA should therefore be reflected as short overall distances in the nuclear genome, thus
447 the ratio between COI and nDNA p-distances in species pairs with ILS is expected to be
448 similar to species without ILS. Contrary to that, we observed that species with mitonuclear
449 discordance did not show significantly shorter p-distances in overall nDNA data and the COI
450 to nDNA p-distances ratio was significantly lower in our target species groups in comparison
451 to the reference species (Fig. 5 and Table 4).
452 The p-distances ratio and t-test analysis do not represent a rigorous statistical analysis of the
453 sequence data. Nevertheless, when introgression between the studied species is indicated
454 by the D-statistics and TreeMix analysis, the shared mtDNA haplotypes are more likely to be
455 explained by introgression rather than ILS. If no introgression was detected, the probability
456 of both scenarios would have been equivalent and equivocal (see Battey & Klicka 2017).
457 Moreover, even with the absence of introgression in the nuclear data, shared mtDNA
458 haplotypes can result from the introgression process (Good *et al.* 2015).
459
460 *Conclusions*
461 The two studied species of *Alopecosa* and four species of *Pardosa* with mitonuclear
462 discordance represent distinct evolutionary lineages as confirmed by genome-wide ddRAD
463 data. We could safely rule out any “operational factor” as being responsible for the DNA
464 barcode sharing. Historical or ongoing introgression is responsible for the observed patterns
465 of mitonuclear discordance in these taxa. Since species pairs with mitonuclear discordance
466 showed comparable overall genetic distances to those without it, we consider incomplete

lineage sorting as an unlikely cause. We consider mitochondrial introgression followed by fixation as the most likely evolutionary cause of mitonuclear discordance. This scenario is supported by the observed signs of introgression between several species and the presence of several endosymbionts, although strong evidence for the role of endosymbionts was not obtained. Finally, our study demonstrates that ddRAD sequencing is a powerful tool for providing a genome-wide insight on the evolution and relationships of species when sparcer sampling of DNA markers show contradictory results.

Acknowledgements

We would like to thank five anonymous reviewers who provided very deep and constructive comments. We would like to sincerely thank Ari Kakko and Timo Pajunen for identification help and providing specimens; Gergin Blagoev for his support in sequencing the COI gene in CCDB, Canada; Laura Törmälä for incredibly efficient work in the laboratory; Mikko Pentinsaari for providing specimens and advice; Tvärminne Zoological Station and Helsinki University Biological Station (Kilpisjärvi) staff for their support in collecting samples; Jesper Smærup Bechsgaard for helpful advice concerning degraded gDNA in spiders; Sami Kivelä and Romain Sarremejane for their help with statistical analyses; FIMM staff for their high-quality work; John Danforth and Marianna Teräväinen for proofreading the manuscript. The authors also wish to acknowledge CSC – IT Center for Science, Finland for providing computational resources. We are very grateful to staff at the Biodiversity Institute of Ontario for their continuous help in generating sequences, entering data into BOLD and aiding the curation of this information. The whole study was funded by the Finnish Academy of Sciences, grant # 277984 accorded to Marko Mutanen.

References

Andrews S (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc

Arnold B, Corbett-Detig RB, Hartl D, Bomblies K (2013) RADseq underestimates diversity and introduces genealogical biases due to nonrandom haplotype sampling. *Molecular Ecology*, **22**, 3179–3190.

Barker DL, Hansen MST, Faraqi AF *et al.* (2004) Two methods of whole-genome amplification enable accurate genotyping across a 2320-SNP linkage panel. *Genome Research*, **14**, 901–907.

Battey CJ, Klicka J (2017) Cryptic speciation and gene flow in a migratory songbird species omplex: insights from the Red-Eyed Vireo (*Vireo olivaceus*). *Molecular Phylogenetics and Evolution*, **113**, 67–75.

Bergthorsson U, Palmer JD (2003) Widespread horizontal transfer of mitochondrial genes in owering plants. *Nature*, **763**, 197–201.

Blagoev GA, deWaard JR, Ratnasingham S *et al.* (2016) Untangling taxonomy: A DNA barcode reference library for Canadian spiders. *Molecular Ecology Resources*, **16**, 325–341.

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

Bonnet T, Leblois R, Rousset F, Crochet P-A (2017) A reassessment of explanations for discordant introgressions of mitochondrial and nuclear genomes. *Evolution*, **71**, 2140–2158.

Boyer MC, Muhlfeld CC, Allendorf FW, Johnson E a (2014) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology*, **22**, 3002–3013.

Burford Reiskind MO, Coyle K, Daniels H V. *et al.* (2016) Development of a universal double-digest RAD sequencing approach for a group of nonmodel, ecologically and economically important insect and fish taxa. *Molecular Ecology Resources*, **16**, 1303–1314.

Cahill JA, Stirling I, Kistler L *et al.* (2015) Genomic evidence of geographically widespread effect of gene flow from polar bears into brown bears. *Molecular Ecology*, **24**, 1205–1217.

Di Candia MR, Routman EJ (2007) Cytonuclear discordance across a leopard frog contact zone. *Molecular Phylogenetics and Evolution*, **45**, 564–575.

Caraballo DA, Abruzzese GA, Rossi MS (2012) Diversity of tuco-tucos (*Ctenomys*, Rodentia) in the Northeastern wetlands from Argentina: Mitochondrial phylogeny and chromosomal evolution. *Genetica*, **140**, 125–136.

Catchen JM, Hohenlohe PA, Bernatchez L *et al.* (2017) Unbroken: RADseq remains a powerful tool for understanding the genetics of adaptation in natural populations.

Molecular Ecology Resources, **17**, 362–365.

Chan KMA, Levin SA (2005) Leaky prezygotic isolation and porous genomes: rapid introgression of maternally inherited DNA. *Evolution*, **59**, 720–729.

Curry MM, Paliulis L V, Welch KD, Harwood JD, White J a (2015) Multiple endosymbiont infections and reproductive manipulations in a linyphiid spider population. *Heredity*, **115**, 146–152.

DaCosta JM, Sorenson MD (2015) ddRAD-seq phylogenetics based on nucleotide, indel and presence – absence polymorphisms: Analyses of two avian genera with contrasting histories with contrasting histories. *Molecular phylogenetics and evolution*, **94**, 122–135.

Davey JW, Cezard T, Fuentes-Utrilla P *et al.* (2013) Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology*, **22**, 3151–3164.

Davey JW., Hohenlohe PA, Etter PD *et al.* (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Rev Genet*, **12**, 499–510.

deWaard JR, Ivanova N V, Hajibabaei M, Hebert PDN (2008) Assembling DNA Barcodes. In: *Environmental Genomics* (eds Martin CC, Martin CC), pp. 275–294. Humana Press, Totowa, NJ.

Durand EY, Patterson N, Reich D, Slatkin M (2011) Testing for ancient admixture between closely related populations. *Molecular Biology and Evolution*, **28**, 2239–2252.

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

Eaton DAR, Hipp AL, González-Rodríguez A, Cavender-Bares J (2015) Historical introgression among the American live oaks and the comparative nature of tests for introgression. *Evolution*, **69**, 2587–2601.

Escudero M, Eaton DAR, Hahn M, Hipp AL (2014) Genotyping-by-sequencing as a tool to infer phylogeny and ancestral hybridization: A case study in *Carex* (Cyperaceae). *Molecular Phylogenetics and Evolution*, **79**, 359–367.

Funk DJ, Omland KE (2003) Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annu. Rev. Ecol. Syst.*, **34**, 397–423.

Gómez-Zurita J, Vogler AP (2006) Testing introgressive hybridization hypotheses using statistical network analysis of nuclear and cytoplasmic haplotypes in the leaf beetle *Timarcha goettingensis* species complex. *Journal of Molecular Evolution*, **62**, 421–433.

Good JM, Vanderpool D, Keeble S, Bi K (2015) Negligible nuclear introgression despite complete mitochondrial capture between two species of chipmunks. *Evolution*, **69**, 1961–1972.

Guerrero RF, Hahn MW (2017) Speciation as a sieve for ancestral polymorphism. *Molecular Ecology*, 5362–5368.

565 Han T, Chang C-W, Kwekel JC *et al.* (2012) Characterization of whole genome amplified
566 (WGA) DNA for use in genotyping assay development. *BMC Genomics*, **13**, 217.

567 Hausmann A, Godfray CH, J. *et al.* (2013) Genetic patterns in European geometrid moths
568 revealed by the Barcode Index Number (BIN) system. *PLoS ONE*, **8**, e84518.

569 Hebert PDN, Hollingsworth PM, Hajibabaei M (2016) From writing to reading the
570 encyclopedia of life. *Philosophical Transactions of the Royal Society of London B:*
571 *Biological Sciences*, **371**, 20150321.

572 Hedtke SM, Hillis DM (2011) The potential role of androgenesis in cytoplasmic-nuclear
573 phylogenetic discordance. *Systematic Biology*, **60**, 87–109.

574 Holm Å, Kronestedt T (1970) A taxonomic study of the wolf spiders of the *Pardosa pullata*-
575 group (Araneae, Lycosidae). *Acta entomologica bohemoslovaca*, **67**, 408–428.

576 Huang J-P (2016) Parapatric genetic introgression and phenotypic assimilation: testing
577 conditions for introgression between Hercules beetles (*Dynastes*, Dynastidae).
578 *Molecular Biology and Evolution*, **25**, 5513–5526.

579 Hurst GDD, Jiggins FM (2000) Male-killing bacteria in insects: mechanisms, incidence, and
580 implications. *Emerg Infect Dis*, **6**, 329–336.

581 Hurst GDD, Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population,
582 phylogeographic and phylogenetic studies: the effects of inherited symbionts.
583 *Proceedings of the Royal Society B: Biological Sciences*, **272**, 1525–1534.

584 Jiggins FM (2003) Male-killing *Wolbachia* and mitochondrial DNA: Selective sweeps, hybrid
585 introgression and parasite population dynamics. *Genetics*, **164**, 5–12.

586 Kronestedt T (1990) Separation of two species standing as *Alopecosa aculeata* (Clerck) by
587 morphological, behavioural and ecological characters, with remarks on related species
588 in the pulverulenta group (Araneae, Lycosidae). *Zoologica Scripta*, **19**, 203–205.

589 Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis
590 Version 7.0 for Bigger Datasets. *Molecular biology and evolution*, **33**, 1870–1874.

591 Leite LAR (2012) Mitochondrial pseudogenes in insect DNA barcoding: differing points of
592 view on the same issue. *Biota Neotropica*, **12**, 301–308.

593 Li G, Davis BW, Eizirik E, Murphy WJ (2016) Phylogenomic evidence for ancient hybridization
594 in the genomes of living cats (Felidae). *Genome Research*, **26**, 1–11.

595 Linnen CR, Farrell BD (2007) Mitonuclear discordance is caused by rampant mitochondrial
596 introgression in *Neodiprion* (Hymenoptera: Diprionidae) sawflies. *Evolution*, **61**, 1417–
597 1438.

598 Melo-Ferreira J, Boursot P, Carneiro M *et al.* (2012) Recurrent introgression of mitochondrial
599 DNA among hares (*Lepus spp.*) revealed by species-tree inference and coalescent
600 simulations. *Systematic Biology*, **61**, 367–381.

601 Meyer BS, Matschiner M, Salzburger W (2016) Disentangling incomplete lineage sorting and

602 introgression to refine species-tree estimates for Lake Tanganyika cichlid fishes.
603 *Systematic Biology*, **66**, 531–550.

604 Murphy NP, Framenau VW, Donnellan SC *et al.* (2006) Phylogenetic reconstruction of the
605 wolf spiders (Araneae: Lycosidae) using sequences from the 12S rRNA, 28S rRNA, and
606 NADH1 genes: Implications for classification, biogeography, and the evolution of web
607 building behavior. *Molecular Phylogenetics and Evolution*, **38**, 583–602.

608 Mutanen M, Kivelä SM, Vos RA *et al.* (2016) Species-level para- and polyphyly in DNA
609 barcode gene trees: Strong operational bias in European Lepidoptera. *Systematic*
610 *Biology*, **65**, 1024–1040.

611 Papakostas S, Michaloudi E, Proios K *et al.* (2016) Integrative taxonomy recognizes
612 evolutionary units despite widespread mitonuclear discordance: Evidence from a rotifer
613 cryptic species complex. *Systematic Biology*, **65**, 508–524.

614 Patterson N, Moorjani P, Luo Y *et al.* (2012) Ancient admixture in human history. *Genetics*,
615 **192**, 1065–1093.

616 Peterson BK, Weber JN, Kay EH, Fisher HS, Hoekstra HE (2012) Double digest RADseq: An
617 inexpensive method for de novo SNP discovery and genotyping in model and non-
618 model species. *PLoS ONE*, **7**, e37135.

619 Pickrell JK, Pritchard JK (2012) Inference of population splits and mixtures from genome-
620 wide allele frequency data. *PLoS Genetics*, **8**, e1002967.

621 Pinard R, de Winter A, Sarkis GJ *et al.* (2006) Assessment of whole genome amplification-
622 induced bias through high-throughput, massively parallel whole genome sequencing.
623 *BMC genomics*, **7**, 216.

624 Raupach MJ, Barco A, Steinke D *et al.* (2015) The application of DNA barcodes for the
625 identification of marine crustaceans from the North Sea and adjacent regions. *PLoS*
626 *ONE*, **10**, e0139421.

627 Rheindt FE, Fujita MK, Wilton PR, Edwards S V. (2014) Introgression and phenotypic
628 assimilation in zimmerius flycatchers (Tyrannidae): Population genetic and phylogenetic
629 inferences from genome-wide SNPs. *Systematic Biology*, **63**, 134–152.

630 Ross HA (2014) The incidence of species-level paraphyly in animals: A re-assessment.
631 *Molecular Phylogenetics and Evolution*, **76**, 10–17.

632 RStudio Team (2015) RStudio: Integrated Development for R. RStudio.

633 Rubinoff D, Cameron S, Will K (2006) A genomic perspective on the shortcomings of
634 mitochondrial DNA for “barcoding” identification. *Journal of Heredity*, **97**, 581–594.

635 Saenz-Agudelo P, Dibattista JD, Piatek MJ *et al.* (2015) Seascape genetics along
636 environmental gradients in the Arabian Peninsula: insights from ddRAD sequencing of
637 anemonefishes. *Molecular Ecology*, **24**, 6241–6255.

638 Shaw KL (2002) Conflict between nuclear and mitochondrial DNA phylogenies of a recent
639 species radiation: What mtDNA reveals and conceals about modes of speciation in

Hawaiian crickets. *Proceedings of the National Academy of Sciences*, **99**, 16122–16127.

Sim KA, Buddle CM, Wheeler TA (2014) Species boundaries of *Pardosa concinna* and *P. lapponica* (Araneae: Lycosidae) in the northern Nearctic: Morphology and DNA barcodes. *Zootaxa*, **3884**, 169–178.

Smith MA, Bertrand C, Crosby K *et al.* (2012) *Wolbachia* and DNA barcoding insects: Patterns, potential, and problems. *PLoS ONE*, **40**, e36514.

Song H, Buhay JE, Whiting MF, Crandall KA (2008) Many species in one: DNA barcoding overestimates the number of species when nuclear mitochondrial pseudogenes are coamplified. *Proceedings of the National Academy of Sciences*, **105**, 13486–13491.

Song H, Moulton MJ, Whiting MF (2014) Rampant nuclear insertion of mtDNA across diverse lineages with in Orthoptera (Insecta). *PLoS ONE*, **9**, 41–43.

Sota T, Vogler AP (2009) Incongruence of mitochondrial and nuclear gene trees in the carabid beetles *Ohomopterus*. *Systematic Biology*, **50**, 39–59.

Soucy SM, Huang J, Gogarten JP (2015) Horizontal gene transfer: Building the web of life. *Nature Reviews Genetics*, **16**, 472–482.

Stamatakis A (2014) RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, **30**, 1312–1313.

Stefanini A, Duron O (2012) Exploring the effect of the *Cardinium* endosymbiont on spiders. *Journal of Evolutionary Biology*, **25**, 1521–1530.

Suchan T, Espíndola A, Rutschmann S *et al.* (2017) Assessing the potential of RAD-sequencing to resolve phylogenetic relationships within species radiations: The fly genus *Chiastocheta* (Diptera: Anthomyiidae) as a case study. *Molecular Phylogenetics and Evolution*, **114**, 189–198.

Takahashi T, Nagata N, Sota T (2014) Application of RAD-based phylogenetics to complex relationships among variously related taxa in a species flock. *Molecular Phylogenetics and Evolution*, **80**, 77–81.

Toews DPL, Brelsford A (2012) The biogeography of mitochondrial and nuclear discordance in animals. *Molecular Ecology*, **21**, 3907–3930.

Wagner CE, Keller I, Wittwer S *et al.* (2013) Genome-wide RAD sequence data provide unprecedented resolution of species boundaries and relationships in the Lake Victoria cichlid adaptive radiation. *Molecular Ecology*, **22**, 787–798.

Zahiri R, Lafontaine JD, Schmidt BC *et al.* (2014) A transcontinental challenge - A test of DNA barcode performance for 1,541 species of Canadian Noctuoidea (Lepidoptera). *PLoS ONE*, **9**, e92797.

Data Accessibility

The full collection and taxonomic data as well as DNA barcodes can be found in the public DS-LYCOSRAD dataset in BOLD. DNA reads from ddRAD sequencing are available at the NCBI Sequence Read Archive (Accession SRR4343322-4343375).

Authors' Contributions

VI collected and identified specimens, partly carried out the labwork, participated in data analysis, participated in the design of the study and drafted the manuscript; KML carried out the labwork, performed the bioinformatics and data analysis, and drafted the manuscript; MM conceived of the study, designed the study, coordinated the study and helped draft the manuscript. All authors gave final approval for publication.

Competing Interests

The authors declare no competing financial interests.

691 **Table 1** Specimens of *Alopecosa* and *Pardosa* analysed in this study and a summary of the
692 ddRAD data.

Groups	Sample ID	Species	Total reads (million)	Clusters at 80%†	Mean depth (*d*=3)	Mean depth (*d*=6)	Retained loci ‡	Recovered loci (*d*=3)	Recovered loci (*d*=6)
Alopecosa	UUI0002	*A. aculeata*	4.3	80675	16.4	21.3	51451	9530	20176
	UUI0003	*A. aculeata*	4.4	83752	16.3	19.9	52460	9179	20360
	UUI0004	*A. aculeata*	3.4	5251	392.1	373.5	2407	950	1241
	UUI0005	*A. aculeata*	0.9	27593	12.4	16.2	17601	5496	8625
	UUI0007	*A. aculeata*	2.0	21412	43.5	44.8	13241	5196	7861
	UUI0008	*A. aculeata*	4.2	15851	145.3	153.5	9444	4023	5756
	UUI0009	*A. aculeata*	2.9	57723	16.3	21.9	37662	9213	17186
	UUI0010	*A. aculeata*	3.4	20330	49.7	92.8	11660	4879	7362
	UUI0091	*A. aculeata*	21.0	10807	1343.3	1091.5	4000	1629	2196
	UUI0092	*A. aculeata*	2.8	86964	11.0	12.9	51655	9495	19972
	UUI0011	*A. taeniata*	0.5	12305	25.5	28.4	7662	1317	1731
	UUI0012	*A. taeniata*	2.7	25375	54.9	60.7	16613	6368	9650
	UUI0013	*A. taeniata*	0.6	28139	8.4	12.9	17507	5446	8476
	UUI0014	*A. taeniata*	1.3	13931	43.8	61.3	8178	3387	4882
	UUI0015	*A. taeniata*	0.6	10742	23.3	30.6	5558	1854	2550
	UUI0016	*A. taeniata*	2.4	15927	58.4	90.4	8906	3178	4579
	UUI0017	*A. taeniata*	4.2	77820	18.2	22.2	49412	9078	18165
	UUI0018	*A. taeniata*	1.1	38213	9.9	13.4	24608	6573	10564
	UUI0021	*A. pulverulenta*	0.8	26784	15.0	15.9	16809	4304	5754
	UUI0076	*Trochosa spinipalpis*	1.4	39150	11.1	15.9	22020	981	978
		AVERAGE	**3.3**	**34937**	**115.7**	**110.0**	**21443**	**5104**	**8903**
Pardosa	UUI0025	*P. sphagnicola*	2.9	4527	370.2	367.0	2005	647	702
	UUI0029	*P. pullata*	1.0	4455	132.5	139.1	2517	849	767
	UUI0030	*P. pullata*	0.5	3215	84.3	101.1	1681	487	448
	UUI0033	*P. fulvipes*	2.6	25211	40.5	30.3	14621	4425	6217
	UUI0034	*P. fulvipes*	0.6	4691	53.1	51.5	1880	619	727
	UUI0035	*P. fulvipes*	1.5	36865	13.4	17.7	21337	4020	6285
	UUI0036	*P. fulvipes*	2.0	45183	13.3	23.3	23510	5132	8231
	UUI0037	*P. fulvipes*	4.1	38747	44.4	35.8	21386	5528	8776
	UUI0038	*P. fulvipes*	0.8	5092	79.9	80.7	2383	431	433
	UUI0039	*P. fulvipes*	2.3	10368	85.1	97.7	5887	2227	2629
	UUI0040	*P. fulvipes*	2.9	15180	99.7	118.7	8777	3058	3991
	UUI0042	*P. sphagnicola*	10.7	5216	675.7	948.7	2280	750	809
	UUI0043	*P. sphagnicola*	0.7	20187	13.5	14.9	12667	4141	5483
	UUI0046	*P. sphagnicola*	2.0	15676	39.0	60.8	8424	2677	3186
	UUI0049	*P. sphagnicola*	0.9	18873	14.6	20.7	11659	3680	5077
	UUI0050	*P. sphagnicola*	2.1	7983	146.3	140.5	4068	1383	1534
	UUI0051	*P. riparia*	1.1	8342	42.0	55.0	4666	1501	1641
	UUI0084	*P. riparia*	1.9	8868	78.7	127.3	4134	1338	1430
	UUI0086	*P. riparia*	0.3	8336	12.8	17.9	3923	1269	1495

UUI0089	*P. riparia*	0.1	2951	25.0	19.1	907	199	186
UUI0090	*P. riparia*	0.9	8610	49.3	54.7	3628	1289	1362
UUI0054	*P. prativaga*	1.4	24805	15.8	14.3	12832	3952	4772
UUI0056	*P. pullata*	1.8	8026	130.4	146.3	3217	977	939
UUI0058	*P. prativaga*	0.2	3938	23.3	23.6	1682	589	555
UUI0061	*P. prativaga*	0.4	10923	11.4	27.9	5151	1619	4903
UUI0062	*P. prativaga*	16.8	9188	1061.6	19.1	3596	1324	1694
UUI0060	*P. pullata*	1.6	26095	23.9	973.0	13711	4022	1374
UUI0066	*P. hyperborea*	1.3	34434	15.6	19.1	19444	2542	3767
UUI0068	*P. lugubris*	0.9	34101	12.4	15.2	15848	1611	2064
UUI0069	*P. amentata*	0.4	12554	14.1	7.6	5360	1115	1304
UUI0070	*P. palustris*	0.5	24187	8.8	12.8	5248	832	2647
UUI0071	*P. maisa*	1.5	28150	18.1	15.2	14257	2066	2883
UUI0072	*P. eiseni*	1.1	29154	15.1	19.7	15326	2459	3463
UUI0074	*P. agricola*	0.5	27895	7.8	8.7	15371	2078	2957
	AVERAGE	**2.1**	**16824**	**102.1**	**112.5**	**8629**	**2083**	**2786**

693 ǂClusters that passed filtering for 3x minimum coverage.

694 ‡ Loci retained after passing coverage and paralog filters.

695

Table 2 Summary of ddRADseq data exploration for *Alopecosa* and *Pardosa*. The data matrices in bold were chosen for phylogenetic analysis

Matrix	*n*	Loci	Unlinked SNPs	Consensus sequences (bp)	VAR (%)	PIS (%)	Missing (%)
Alopecosa							
c80m6_alopecosa	**20**	**12,641**	**12,423**	**2,420,062**	**207,608 (8.6)**	**70,437 (2.9)**	**59.8**
c85m6_alopecosa	20	13,757	13,503	2,609,577	184,787 (7.1)	62,619 (2.4)	60.1
Pardosa							
c80m6_pardosa	**34**	**7,157**	**7,109**	**1,417,596**	**161,708 (11.4)**	**51,793 (3.7)**	**71.0**
c80m9_pardosa	34	2,455	2,438	484,594	52,883 (10.9)	18,059 (3.7)	62.4
c80m12_pardosa	34	601	595	119,013	13,152 (11.0)	4,774 (4.0)	53.0
c80m15_pardosa	34	118	114	23,783	3,345 (14.1)	1,197 (5.0)	43.1
c80m18_pardosa	34	23	20	4,638	619 (13.3)	188 (4.1)	32.7
c85m6_pardosa	34	7,258	7,203	1,427,328	146,133 (10.2)	45,842 (3.2)	71.1

Note: *n*, number of individuals; VAR, number of variable sites; PIS, number of parsimony informative sites.

Table 3 Patterson's four-taxon D-statistic test results showing significant replicates for introgression in *Alopecosa* and *Pardosa*

Test	P1	P2	P3	O	Range Z	nSig/n	nSig (%)
Alopecosa							
1.1	acul	acul	taen	Tspin	**(0.0 - 14.2)**	**5/359**	**1.4**
1.2	acul	acul	pulv	Tspin	**(0.0 - 4.2)**	**1/44**	**2.3**
Pardosa							
2.1	fulv	fulv	spha	O7	**(0.0 - 12.9)**	**4/179**	**2.2**
2.2	fulv	fulv	ripa	O7	**(0.0 - 5.4)**	**7/179**	**3.9**
2.3	fulv	fulv	prat	O7	**(0.0 - 5.1)**	**3/143**	**2.1**
2.4	fulv	fulv	pull	O7	**(0.0 - 5.9)**	**2/143**	**1.4**
2.5	spha	spha	fulv	O7	**(0.0 - 7.3)**	**4/89**	**4.5**
2.6	spha	spha	ripa	O7	**(0.0 - 5.2)**	**3/49**	**6.1**
2.7	ripa	ripa	fulv	O7	**(0.0 - 11.2)**	**2/89**	**2.2**
2.8	ripa	ripa	spha	O7	**(0.0 - 5.2)**	**2/49**	**4.1**
2.9	ripa	ripa	prat	O7	**(0.0 - 8.2)**	**5/39**	**12.8**
2.10	prat	prat	ripa	O7	**(0.0 - 8.0)**	**1/29**	**3.4**
2.11	pull	pull	fulv	O7	**(0.0 - 4.3)**	**1/53**	**1.9**
2.12	pull	pull	spha	O7	**(0.0 - 5.9)**	**2/29**	**6.9**
2.13	pull	pull	ripa	O7	**(0.0 - 4.7)**	**2/29**	**6.9**

Note: Each test was repeated over all possible four-sample replicates (n), with a range of Z-scores reported, and the number of significant replicates shown (nSig). P1, P2, and P3: for *Alopecosa* group, acul: *A. aculeata*, taen: *A. taeniata*, pulv: *A. pulverulenta*; for *Pardosa* group, fulv: *P. fulvipes*, spha: *P. sphagnicola*, ripa: *P. riparia*, prat: *P. prativaga*, pull: *P. pullata*; O: outgroup 'Tspin' represents *Trochosa spinipalpis* for *Alopecosa* group; for *Pardosa* group, 'O7' consists of all individuals from the seven species *P. agricola, P. amentata, P. eiseni, P. hyperborea, P. lugubris, P. maisa*, and *P. palustris*.

Table 4 Mean p-distances of ddRAD data (below diagonal) and COI barcode sequences (above diagonal). Species of *P. pullata*-group with mitonuclear discordance are in bold. Values for *Alopecosa* group are given in the text.

	spha	fulv	prat	ripa	pull	agri	amen	eise	hype	lugu	mais	palu
P. sphagnicola	na	**0.2**	**0.5**	**0.4**	2.1	5.2	4.4	5.9	5.9	3.8	5.1	5.5
P. fulvipes	**2.7**	na	**0.6**	**0.6**	2.2	5.2	4.5	6.1	6.0	3.9	5.2	5.6
P. prativaga	**3.0**	**3.1**	na	**0.7**	2.2	5.0	4.1	5.9	5.7	3.8	5.3	5.5
P. riparia	**3.3**	**3.3**	**3.5**	na	1.9	5.2	4.5	6.0	5.6	3.8	5.0	5.5
P. pullata	3.5	3.5	3.6	3.7	na	5.0	5.0	6.2	5.5	4.4	5.0	5.3
P. agricola	4.7	4.8	4.8	5.0	4.9	na	4.9	6.1	3.8	4.4	5.6	2.1
P. amentata	4.4	4.4	4.5	4.9	4.5	3.9	na	5.5	5.2	4.1	5.3	5.8
P. eiseni	4.4	4.4	4.6	4.6	4.3	3.0	3.4	na	5.6	5.9	5.5	7.3
P. hyperborea	4.7	4.9	4.7	4.9	5.0	3.0	3.6	2.8	na	4.6	4.9	5.0
P. lugubris	4.4	4.4	4.5	4.4	4.4	3.6	3.1	3.4	3.9	na	5.5	4.6
P. maisa	4.5	4.6	4.6	4.5	4.4	3.5	3.8	3.3	3.5	3.7	na	6.2
P. palustris	4.7	4.7	4.9	4.6	4.8	1.9	3.7	3.1	3.0	3.8	3.6	na

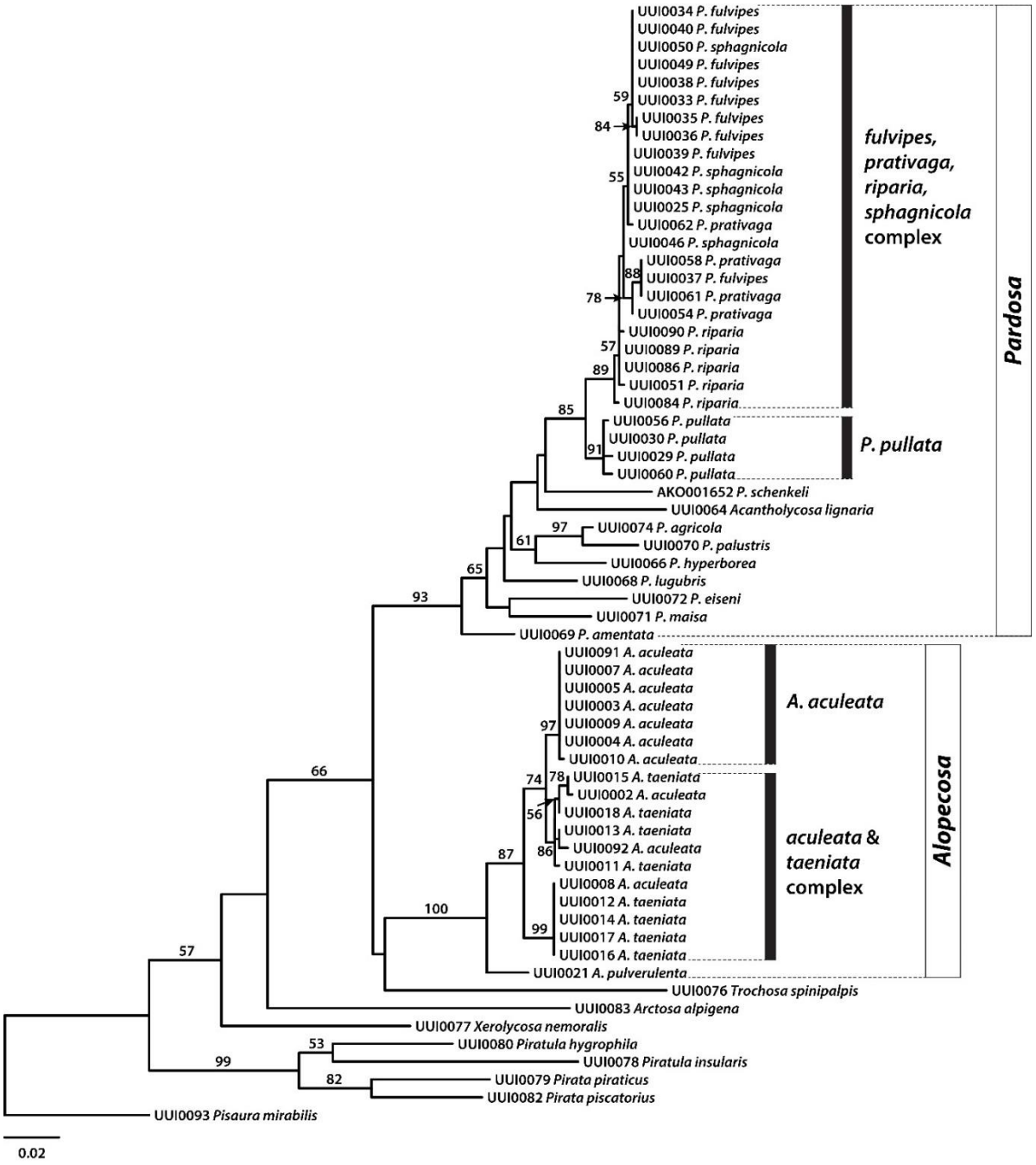

Figure 1 Maximum likelihood tree based on mitochondrial COI sequences. Values indicate node bootstrap support values (only shown for nodes supported with more than 50% BS). The two studied groups with mitonuclear discordance are indicated with black vertical bars. *P. pullata* was considered here as not showing mitonuclear discordance, although being closely related to the complex of four species of *Pardosa* with mitonuclear discordance.

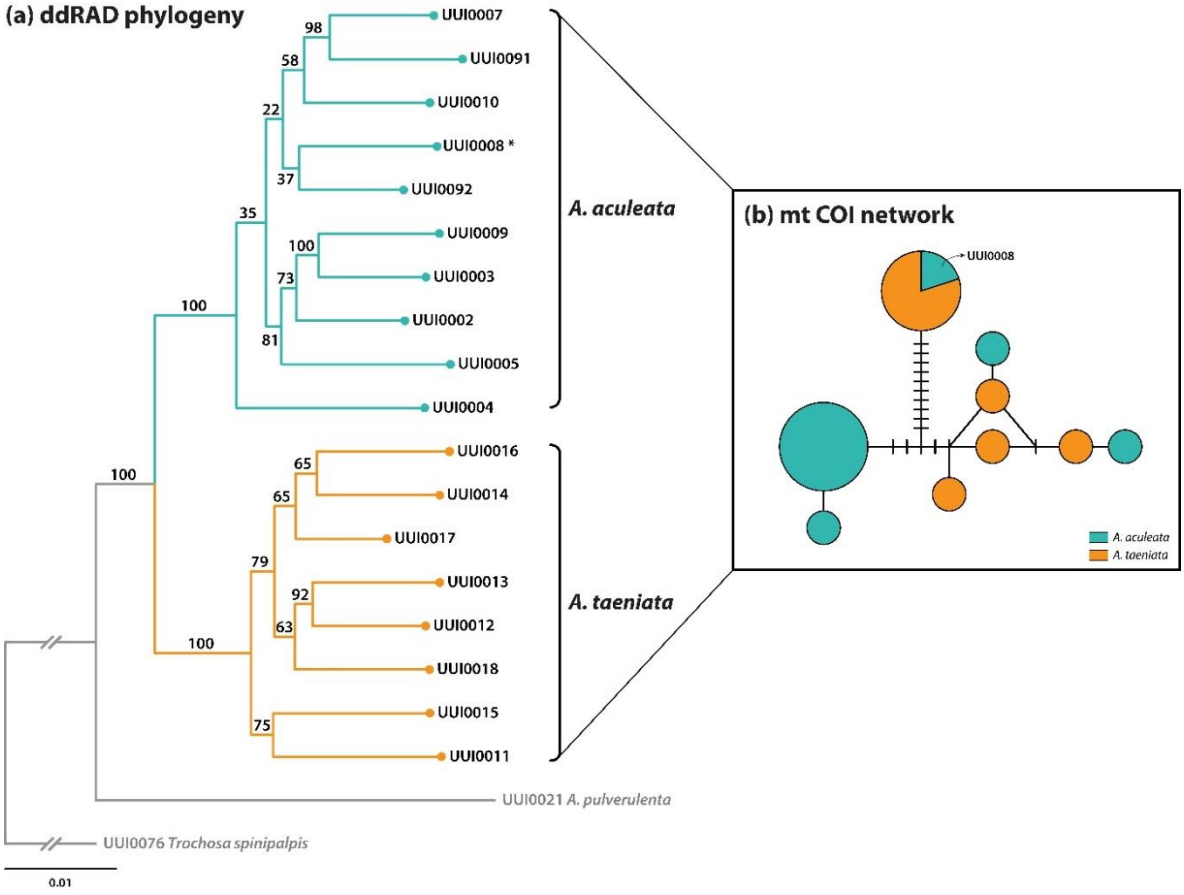

Figure 2 a) Maximum likelihood tree of *Alopecosa* group, including two specimens of *A. pulverulenta* and *Trochosa spinipalpis* as outgroup, based on the data matrix with of 12,423 unlinked SNPs in 2,420,062 bp and b) mitochondrial COI haplotype network. Each circle represents a haplotype and circle size is proportional to strain frequency. Lines between haplotypes are single mutational steps and short solid lines indicate missing haplotypes (either extinct or not sampled). Node confidence values were estimated based on 500 bootstrap replicates. Bootstrap values are indicated near the nodes.

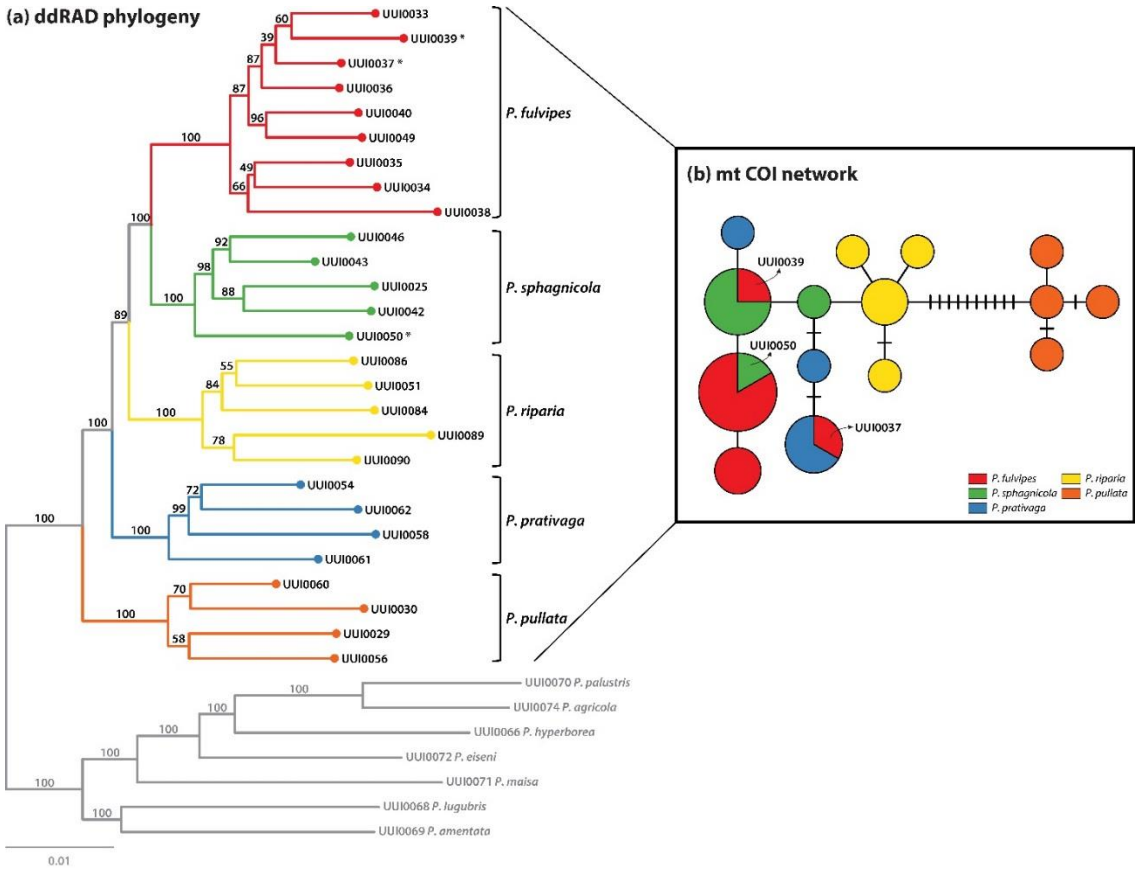

Figure 3 a) Maximum likelihood tree of *Pardosa* group based on the data matrix with 7,109 unlinked SNPs in 1,417,596 bp and b) mitochondrial COI haplotype network. Node colours correspond to the haplotype network indicating distinct species. Each circle represents a haplotype and circle size is proportional to strain frequency. Lines between haplotypes are single mutational steps and short solid lines indicate missing haplotypes (either extinct or not sampled). Node confidence values were estimated based on 500 bootstrap replicates. Bootstrap values are indicated above the nodes.

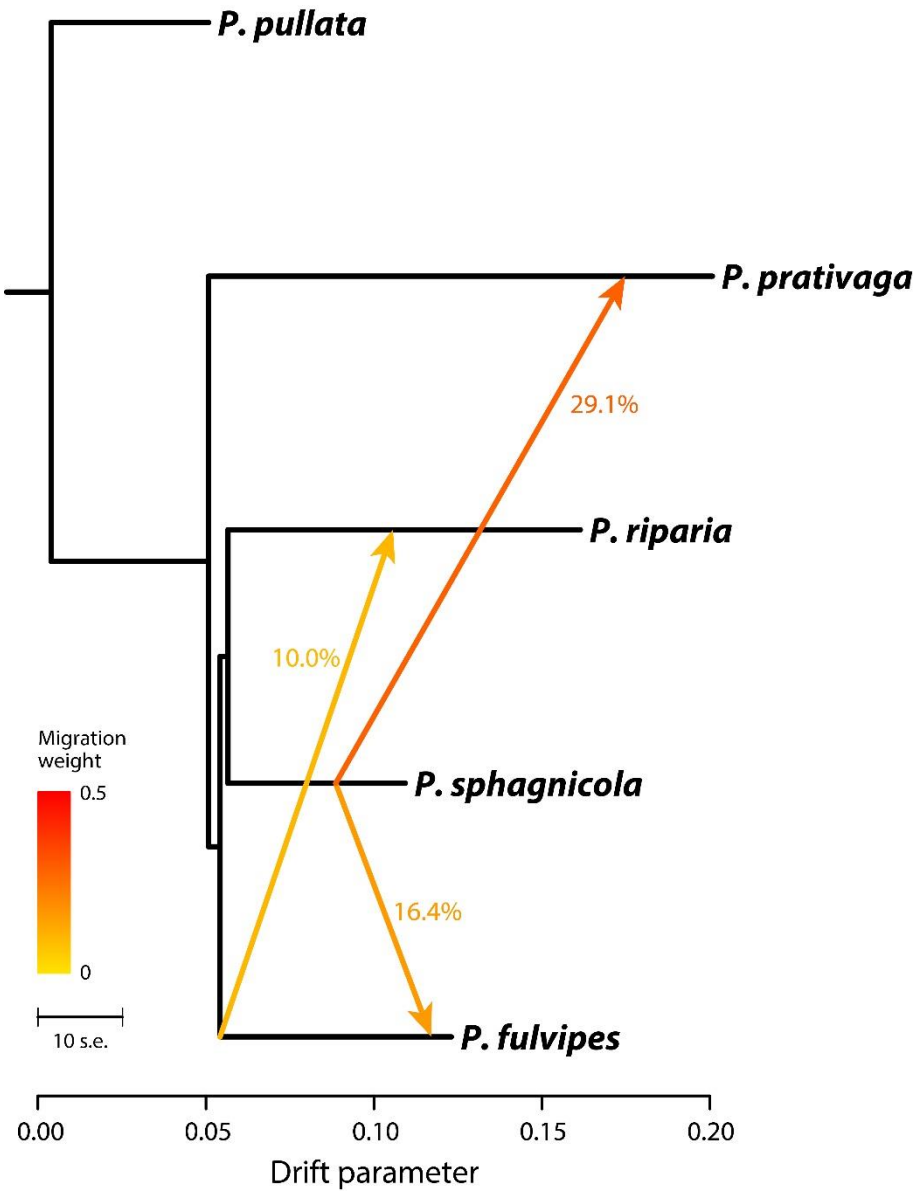

Figure 4 Maximum likelihood tree, generated by TreeMix, showing the relationship among five species of *Pardosa*. The scale bar shows ten times the average standard error (s.e.) of the estimated entries in the sample covariance matrix. Migration arrows are coloured according to their weight. The migration weight represents the fraction of ancestry derived from the migration edge.

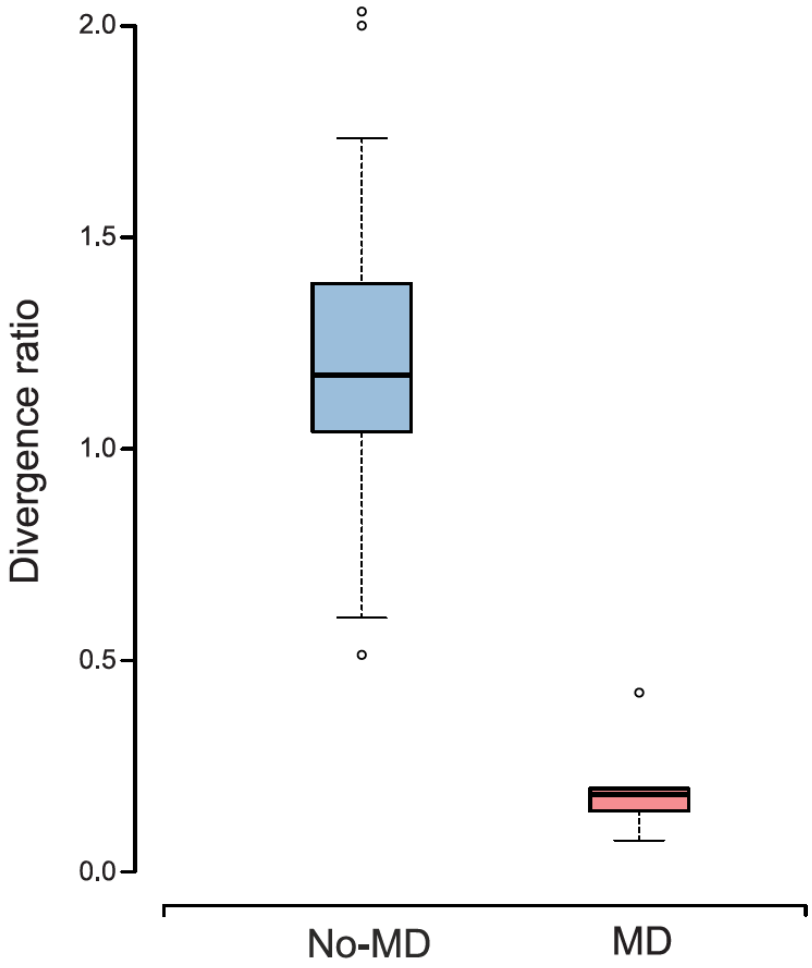

Figure 5 Boxplots depicting median, first and third quartile, and standard deviation of COI to ddRADseq p-distance ratios. The blue boxplot (No-MD) includes ratios of species with no introgression or ILS. The red boxplot (MD) represents the ratios in *P. pullata* and *A. pulverulenta* groups of species with detected DNA barcode sharing.