Deep ancestral introgression shapes evolutionary history of dragonflies and damselflies

Anton Suvorov[1$], Celine Scornavacca[2], M. Stanley Fujimoto[3], Paul Bodily[3], Mark Clement[3], Keith A. Crandall[4], Michael F. Whiting[5,6], Daniel R. Schrider[1] and Seth M. Bybee[5,6]

[1]Department of Genetics, University of North Carolina at Chapel Hill, Chapel Hill, NC

[2]Institut des Sciences de l'Evolution Université de Montpellier, CNRS, IRD, EPHE CC 064, Place Eugène Bataillon, 34095 Montpellier Cedex 05, France.

[3]Computer Science Department, Brigham Young University, Provo, UT

[4]Computational Biology Institute, Department of Biostatistics and Bioinformatics, Milken Institute School of Public Health, George Washington University, Washington, DC

[5]Department of Biology, Brigham Young University, Provo, UT

[6]M.L. Bean Museum, Brigham Young University, Provo, UT

$Correspondence: antony.suvorov@med.unc.edu

SUMMARY

Introgression is arguably one of the most important biological processes in the evolution of groups of related species, affecting at least 10% of the extant species in the animal kingdom. Introgression reduces genetic divergence between species, and in some cases can be highly beneficial, facilitating rapid adaptation to ever-changing environmental pressures. Introgression also significantly impacts inference of phylogenetic species relationships where a strictly binary tree model cannot adequately explain reticulate net-like species relationships. Here we use phylogenomic approaches to understand patterns of introgression along the evolutionary history of a unique, non-model, relic insect system: dragonflies and damselflies (Odonata). We demonstrate that introgression is a pervasive evolutionary force across various taxonomic levels within Odonata. In particular, we show that the morphologically "intermediate" species of Anisozygoptera (one of the three primary suborders within Odonata besides Zygoptera and Anisoptera), which retain phenotypic characteristics of the other two suborders, experienced high levels of introgression probably coming from zygopteran genomes. Additionally, we found evidence for multiple cases of deep inter-familial ancestral introgression. However, weaker evidence of introgression was found within more recently diverged focal clades: Aeshnidae, Gomphidae+Petaluridae and Libellulidae.

INTRODUCTION

In recent years there has been a large body of evidence amassed demonstrating that multiple parts of the Tree of Life did not evolve according to a strictly bifurcating phylogeny [1, 2]. Instead, many organisms experience reticulate network-like evolution that is caused by an exchange of inter-specific genetic information via various biological processes. In particular, lateral gene transfer, incomplete lineage sorting (ILS) and introgression can result in gene trees that are discordant with the species tree [3, 4]. Lateral transfer and introgression both involve gene flow following speciation, thereby producing "reticulate" phylogenies. Incomplete lineage sorting (ILS), on the other hand, occurs when three or more lineages coalesce in their ancestral population. This often arises in an order that conflicts with that of the true species tree, and thus involves no post-speciation gene flow and therefore does not contribute to reticulate evolution.

58 Thus, phylogenetic species-gene tree incongruence observed in empirical data can provide
59 insight into underlying biological factors that shape the evolutionary trajectories of a set of taxa.
60 The major source of reticulate evolution for eukaryotes is introgression where it affects
61 approximately 25% of flowering plant and 10% of animal species [1, 5]. Introgressed alleles can
62 be fitness-neutral, deleterious [6] or adaptive. For example, adaptive introgression has been
63 shown to provide an evolutionary rescue from polluted habitats in gulf killifish (*Fundulus*
64 *grandis*) [7], yielded mimicry adaptations among *Heliconius* butterflies [8] and archaic
65 introgression has facilitated adaptive evolution of altitude tolerance [9], immunity and
66 metabolism in modern humans [10]. Additionally, hybridization and introgression are important
67 and often overlooked mechanisms of invasive species establishment and spread [11].
68 Odonata, the insect order that contains dragonflies and damselflies, lacks a strongly
69 supported backbone tree to clearly resolve higher-level phylogenetic relationships [12, 13].
70 Current evidence places odonates together with Ephemeroptera (mayflies) as the living
71 representatives of the most ancient insect lineages to have evolved wings and active flight [14].
72 Odonates possess unique anatomical and morphological features such as a specialized body
73 form, specialized wing venation, a distinctive form of muscle attachment to the wing base [15]
74 allowing for direct flight and accessory (secondary) male genitalia that support certain unique
75 behaviors (e.g., sperm competition). They are among the most adept flyers of all animals and are
76 exclusively carnivorous insects relying primarily on vision [16, 17] to capture prey. They spend
77 much of their adult life in flight [18]. Biogeographically, odonates exhibit worldwide dispersal
78 [19] and play crucial ecological roles in local freshwater communities, being a top invertebrate
79 predator [20]. Due to this combination of characteristics, odonates are quickly becoming model
80 organisms to study specific questions in ecology, physiology and evolution [21, 22]. However,
81 the extent of introgression at the genomic scale within Odonata remains primarily unknown.
82 Two early attempts to tackle introgression/hybridization patterns within Odonata were
83 undertaken in [23, 24]. The studies showed that two closely related species of damselflies,
84 *Ischnura graellsii* and *I. elegans,* can hybridize under laboratory conditions and that genital
85 morphology of male hybrids shares features with putative hybrids from *I. graellsii–I. elegans*
86 natural allopatric populations [23]. The existence of abundant hybridization and introgression in
87 natural populations of *I. graellsii* and *I. elegans* has received further support from an analysis of
88 microsatellite data [25]. Putative hybridization events have also been identified in a pair of

89 calopterygoid damselfly species, *Mnais costalis* and *M. pruinosa* based on the analyses of two
90 molecular loci (mtDNA and nucDNA) [26], and between *Calopteryx virgo* and *C. splendens*
91 using 16S ribosomal DNA and 40 random amplified polymorphic DNA (RAPD) markers [27]. A
92 more recent study identified an inter-specific hybridization between two cordulegasterid
93 dragonfly species, *Cordulegaster boltonii* and *C. trinacriae* using two molecular markers
94 (mtDNA and nucDNA) and geometric morphometrics [28].
95 Here we present a comprehensive analysis of transcriptomic data from 83 odonate
96 species. First, we reconstruct a robust phylogenetic backbone using up to 4341 genetic loci for
97 the order and discuss its evolutionary history spanning from the Carboniferous period (~360
98 mya) to present day. Second, we infer signatures of introgression with varying degrees of
99 phylogenetic depth using genomic data across the tree primarily focusing on four focal clades,
100 namely Anisozygoptera ("deep"), Aeshnidae ("shallow"), Gomphidae+Petaluridae
101 ("intermediate") and Libellulidae ("shallow"). We find that introgression is pervasive in Odonata
102 throughout its entire evolutionary history. In particular, we identify a strong signal of deep
103 introgression in the Anisozygoptera suborder, species of which possess traits of both main
104 suborders, Anisoptera and Zygoptera.
105
106 **RESULTS**
107
108 **Overview of Odonata Phylogeny**
109
110 We compiled transcriptomic data for 83 odonate species including 49 new transcriptomes
111 sequenced for this study (Data S1). To assess effects of various steps of our phylogenetic
112 pipeline on species tree inference, we examined different methods of sequence homology
113 detection, multiple sequence alignment strategies, postprocessing filtering procedures and tree
114 estimation methods. Specifically, three types of homologous loci (gene clusters) were used to
115 develop our supermatrices, namely 1603 conserved single-copy orthologs (CO), 1643 all single-
116 copy orthologs (AO) and 4341 paralogy-parsed orthologs (PO) with ≥ 42 (~50%) species present
117 (for more details, see Method Details). To date, our data represents the most comprehensive
118 resource available for Odonata in terms of gene sampling. Each gene cluster was aligned,
119 trimmed and concatenated resulting in five main supermatrices, CO (DNA/AA), AO (AA) and
120 PO (DNA/AA), which included 2,167,861 DNA (682,327 amino acid (AA) sites), 882,417 AA

121 sites, 6,202,646 DNA (1,605,370 AA) sites, respectively. Thus, the largest alignment that we
122 used to infer the odonate phylogeny consists of 4341 loci concatenated into a supermatrix with
123 >6 million nucleotide sites. All supermatrices are summarized in Data S2; the inferred odonate
124 relationships are shown in Figure S1A whereas topologies of all inferred phylogenies are plotted
125 in Figure S1B and topologies of 1603 CO gene trees are shown in Figure S1C. Additionally, we
126 performed nodal dating of the inferred phylogeny using 20 fossil calibration points (Data S3).
127 The inferred ML phylogenetic tree of Odonata using DNA supermatrix of 1603 BUSCO
128 loci (Figure 1) was used as a primary phylogenetic hypothesis throughout this study as it agrees
129 with the majority of relationships inferred by other methods (Figures S1A, S1B). Divergence of
130 Zygoptera and Epiprocta (Anisozygoptera+Anisoptera) from the Most Recent Common Ancestor
131 (MRCA) occurred in the Middle Triassic ~242 mya (Figure 1), which is in line with recent
132 estimates [14, 29]. Comprehensive phylogenetic coestimation of subordinal relationships within
133 Odonata showed that the suborders were well supported (Figure S1A), as they were consistently
134 recovered as monophyletic clades in all analyses. In several previous studies, paraphyletic
135 relationships of Zygoptera had been proposed based on wing vein characters derived from fossil
136 odonatoids and extant Odonata [30], analysis of 12S [31], analysis of 18S, 28S, Histone 3 (H3)
137 and morphological data [32] and analysis of 16S and 28S data [33]. In most of these studies,
138 Lestidae was inferred to be sister to Anisoptera. Functional morphology comparisons of flight
139 systems, secondary male genitalia and ovipositors also supported a non-monophyletic Zygoptera
140 with uncertain phylogenetic placement of multiple groups [34]. Nevertheless, the relationships
141 inferred from these previous datasets seem to be highly unlikely due to apparent morphological
142 differentiation (e.g., eye spacing, body robustness, wing shape) between the suborders and
143 support for monophyletic Anisoptera and Zygoptera from more recent morphological [35],
144 molecular [14, 17, 36, 37] and combined studies using both data types [38]. Our analyses recover
145 Zygoptera as monophyletic consistently (Figure S1A).
146 Divergence time estimates suggest a TMRCA of Anisoptera and Anisozygoptera in the
147 Late Triassic (~205 mya; Figure 1). Epiprocta as well as Anisoptera were consistent with more
148 recent studies and recovered as monophyletic with very high support. We also note here that our
149 divergence time estimates of Anisoptera tend to be younger than those found in ref. [39]. The
150 fossil-calibration approach based on penalized likelihood has been shown to overestimate true
151 nodal age [40] preventing direct comparison between our dates and those estimated in ref. [39].

152 The phylogenetic position of Gomphidae and Petaluridae, both with respect to each other
153 and the remaining anisopteran families, has long been difficult to resolve. Several phylogenetic
154 hypothesis have been proposed in the literature based on molecular and morphological data
155 regarding the placement of Gomphidae as sister to the remaining Anisoptera [41] or to
156 Libelluloidea [42]. Petaluridae has exhibited stochastic relationships with different members of
157 Anisoptera, including sister to Gomphidae [42], sister to Libelluloidea [37], sister to
158 Chlorogomphidae+Cordulegasteridae [38] and sister to all other Anisoptera [30, 43]. The most
159 recent analyses of the major anisopteran lineages using several molecular markers [13] suggest
160 Gomphidae and Petaluridae as a monophyletic group, but without strong support. Here the
161 majority of our supermatrix analyses (Figure S1A) strongly support a sister relationship between
162 the two families, and in our phylogeny (Figure 1) they split from the MRCA ~165 mya in the
163 Middle Jurassic (Figure 1); however, almost all the coalescent-based species trees reject such a
164 relationship with high confidence (Figure S1A). In the presence of incomplete lineage sorting,
165 concatenation methods can be statistically inconsistent [44] leading to an erroneous species tree
166 topology with unreasonably high support [45]. Thus, inconsistency in the recovery of a sister
167 group relationship between Gomphidae and Petaluridae can be explained by elevated levels of
168 incomplete lineage sorting between the families and/or possible introgression events [46] (see
169 below).
170 New zygopteran lineages originated in the Early Jurassic ~191 mya with the early split of
171 Lestoidea and the remaining Zygoptera (Figure 1). Subsequent occurrence of two large
172 zygopteran groups, Calopterygoidea and Coenagrionoidea, was estimated within the Cretaceous
173 (~145-66 mya) and culminated with the rapid radiation of the majority of extant lineages in the
174 Paleogene (~66-23 mya). Our calibrated divergences generally agree with estimates in [14].
175 However, any further comparisons are precluded by the lack of comprehensive divergence time
176 estimation for Odonata in the literature. The backbone of the crown group Calopterygoidea that
177 branched off from Coenagrionoidea ~129 mya in the Early Cretaceous was well supported as
178 monophyletic in most of our inferred phylogenies (Figures 1, S1A). Previous analyses struggled
179 to provide convincing support for the monophyly of the superfamily [12, 37, 38], whereas only
180 11 out of 48 phylogenetic reconstructions rejected Calopterygoidea (Figure S1A).
181 We used quartet sampling [47] to provide additional information about nodal support and
182 investigate biological explanations for alternative evolutionary histories that received some

support. We found that for most odonate key radiations, the majority of quartets (i.e. Frequency > 0.5) support the proposed phylogenetic hypothesis with Quartet Concordance (QC) scores > 0 (Figure S2) across all estimated putative species trees (Figure S1B). The few exceptions consist of the Gomphidae+Petaluridae split and the A+B split, where we have Frequency < 0.5 and QC < 0, which suggests that alternative relationships are possible. Quartet Differential (QD), inspired by *f* and *D* statistics for introgression, provides an indicator of how much the proportions of discordant quartets are skewed (i.e. whether one discordant is more common than the other), suggestive of introgression and/or introgression or substitution rate heterogeneity [47]. Interestingly, we identified skewness (i.e. QD < 0.5) for almost every major radiation (Figure S2), which suggests that alternative relationships can be a result of underlying biological processes (e.g. introgression) rather than ILS alone [47]; however, this score may not be highly informative if the majority of quartets agree with the focal topology (i.e Frequency > 0.5 and QC > 0). However, for both the Gomphidae+Petaluridae and the A+B splits, we have Frequency > 0.5, QC < 0 and QD < 0.5, implying that alternative phylogenetic relationships are plausibly not only due to ILS but also possible ancestral introgression.

Major Trends in Evolutionary History of Odonata

Investigation of diversification rates in Odonata highlighted two major trends correlated with two mass extinction events in the Permian-Triassic (P-Tr) ~252 mya and Cretaceous-Paleogene (K-Pg) ~66 mya. First, it appears likely that the ancient diversity of Odonata was largely eliminated after the P-Tr mass extinction event (see the temporal distribution of fossil samples in Figure 1) as was also the case for multiple insect lineages [48]. According to the fossil record at least two major odonatoid lineages went extinct (Protodonata and Protanisoptera [49]) and likely many genera from other lineages as well (e.g., Kargalotypidae from Triadophlebimorpha [50]). The establishment of major odonate lineages was observed during the Cretaceous starting ~135 mya (Figure 1, red dot). This coincided with the radiation of angiosperm plants that, in turn, triggered the formation of herbivorous insect lineages [29]. Odonates are exclusively carnivorous insects and their diversification was likely driven by the aforementioned sequence of events. Interestingly, molecular adaptations in the odonate visual systems are coupled with their diversification during the Cretaceous as well [17].

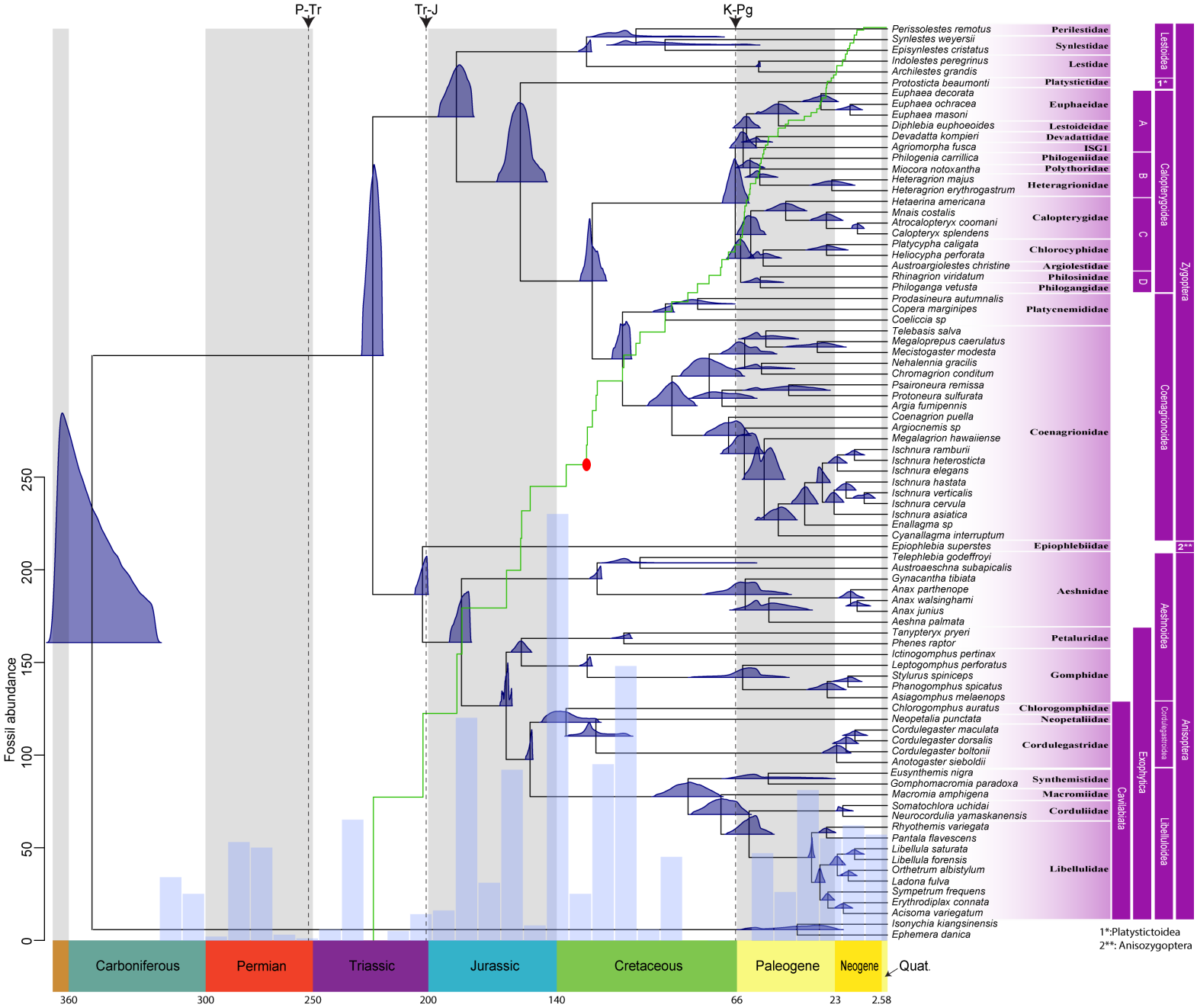

Figure 1. Evolutionary History of Odonata

Fossil calibrated ML phylogenetic tree of Odonata using a DNA supermatrix consisting of 1603 BUSCO genes with a total of 2,167,861 aligned sites. The blue densities at each node represent posterior distributions of ages estimated in MCMCTree using 20 fossil calibration points. Lineage Through Time (LTT, pink solid line) trajectory shows accumulation of odonate lineages through time. The red dot on the LTT line indicates the beginning of establishment for major odonate lineages originating in and spanning the Cretaceous. The histogram (blue bars) represents temporal distribution of fossil Odonatoptera samples. The dashed vertical lines mark major extinction events, namely Permian-Triassic (P-Tr, ~ 251 mya), Triassic-Jurassic (Tr-J, ~ 201.3 mya) and Cretaceous-Paleogene (K-Pg, ~66 mya)

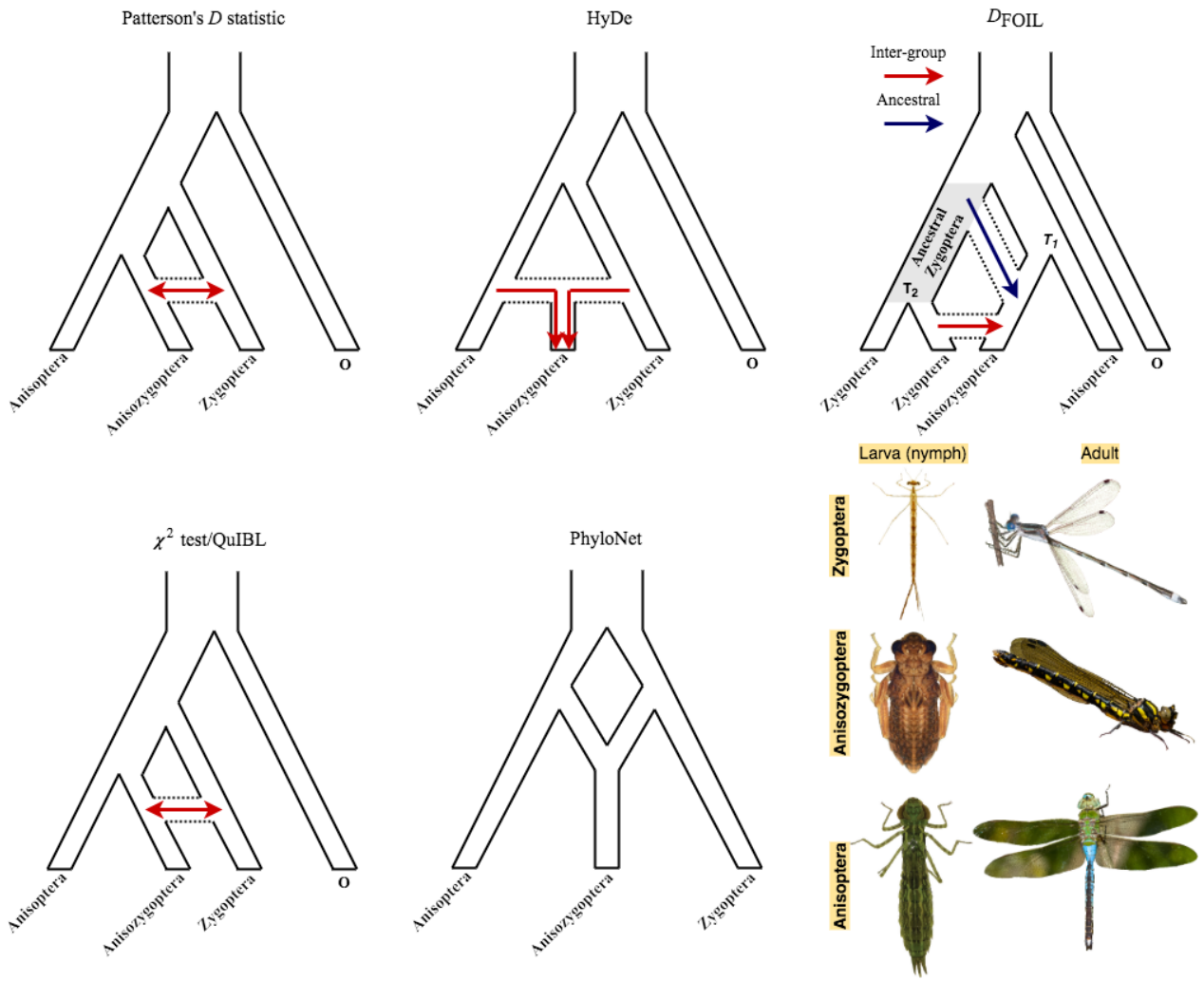

Figure 2. Detection of Introgression/Hybridization Trajectories by Different Methods in Anisozygoptera. Three site pattern-based (*D* statistic, HyDe and D_{FOIL}), gene tree count/branch length-based (χ^2 count test/QuIBL) and network ML inference (PhyloNet) methods were used to test for introgression/hybridization. Arrows denote introgression. The figure panel represents larval and adult stages for three Odonata suborders. Species from top to bottom: *Lestes australis, Epiophlebia superstes* and *Anax junius.* Image credit: *Epiophlebia superstes* adult by Christian Dutto Engelmann; *Lestes australis* and *Anax junius* adults by John Abbott.

Signatures of Introgression within Odonata

We tested introgression/hybridization scenarios within Odonata on different taxonomic ranks, such as order, suborder, superfamily (except Aeshnoidea, since it has been recovered as paraphyletic, and Platystictoidea, which contains only one species in our phylogeny; Figure 1) and certain specific focal clades. Specifically, we used five different methods to test for

222 introgression within Odonata, as exemplified for the Anisozygoptera suborder in Figure 2. As
223 our test cases for introgression, we *a priori* selected Aeshnidae and Libellulidae families as
224 possible examples of "shallow" introgression, Gomphidae+Petaluridae as a possible example of
225 "intermediate" introgression and Anisozygoptera as a possible case of "deep" introgression.
226 Since the introgression signatures within Odonata are unknown, we chose these test clades based
227 on possession of characteristic traits of both suborders for Anisozygoptera [35, 51-53], the
228 unstable phylogenetic placement of Gomphidae relative to Petaluridae [41, 42] and the presence
229 of closely related species within Libellulidae and Aeshnidae. For Aeshnidae, Libellulidae and
230 Gomphidae+Petaluridae, we tested all possible introgression scenarios between species within
231 each of these clades using methods outlined in Figure 2. For Anisozygoptera we tested whether
232 the lineage leading to *E. superstes* experienced ancestral introgression with Anisoptera and/or
233 Zygoptera (see Figure 2). Additionally, we searched for introgression within other taxonomic
234 groups, including the remaining suborders (Zygoptera and Anisoptera) and five superfamilies
235 within Zygoptera and Anisoptera (Lestoidea, Calopterygoidea, Coenagrionoidea,
236 Cordulegastroidea and Libelluloidea; Figure 1). The introgression results for the entire Odonata
237 order either comprise of all tests performed within the entire phylogeny (HyDe) or of a union of
238 tests performed within Anisoptera, Zygoptera and Anisozygoptera (D_{FOIL} and χ^2 count
239 test/QuIBL).
240
241
242

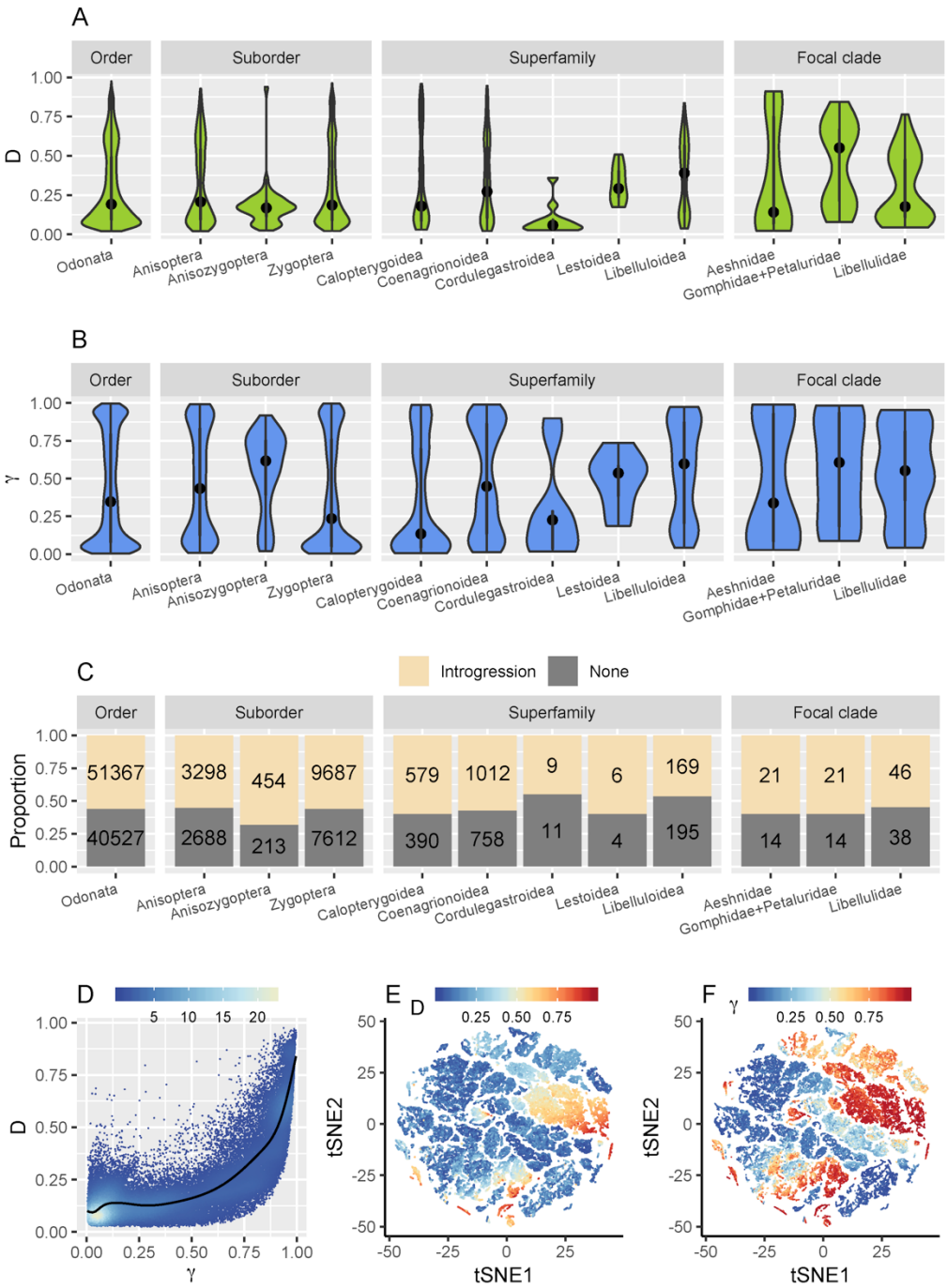

Figure 3. Distributions of the Patterson's *D* Statistic, HyDe γ and Their Relations Across Odonate Taxonomic Levels.
(A) Estimated distributions of significant (Bonferroni corrected $P < 0.05$) *D* statistic from ABBA-BABA site pattern counts for each quartet using HyDe output. Black dots mark medians of violin plots.
(B) Distribution of significant (Bonferroni corrected $P < 10^{-6}$) γ values for each quartet estimated by HyDe. In general, γ values that are not significantly different from 0 (or 1) denote no relation of a putative hybrid species to either of the parental species P_1 (1-γ) or P_2 (γ) in a quartet.
(C) Proportions of quartets that support or reject introgression based on simultaneous significance of *D* statistics and γ.
(D) Non-linear relationships between absolute values of significant *D* statistics and γ. The black line denotes a GAM fit. The color legend shows density of quartets across γ-*D* plane.
(E)-(F) tSNE projections of the 15 site pattern counts derived for each of the 51367 significant quartets from HyDe output (each dot on a tSNE map represents a quartet). Color schemes reflect significant values of *D* statistic (E) and significant values of of γ (F).

243 First, we compared the distributions of *D*-statistic values within different taxa computed

244 from ABBA-BABA [54] site patterns (Figure 3A). The multimodal shape of the *D* distribution
245 for Odonata can be explained by the presence of quartets of older as well as more recently
246 diverged taxa. Although quartets with older divergence times tend to have only slightly
247 decreased values of *D* [55], this effect can be more pronounced for extremely old divergences
248 leading to significantly smaller *D*. Indeed, quartets that were formed within more recently
249 diverged clades of Coenagrionoidea, Aeshnidae (both in Early Cretaceous, ~ 140 mya) and
250 Libellulidae (Oligocene ~ 33 mya) have significantly greater *D* values (Wilcoxon rank-sum test
251 [WRST], $P = 8.366 \times 10^{-10}$) if compared with quartets that were formed from distantly related
252 species of Zygoptera and Anisoptera (Late Triassic ~237 mya). Comparisons of *D* distributions
253 derived from the entire order with those from other odonate monophyletic groups showed that
254 only Coenagrionoidea, Gomphidae+Petaluridae and Libelluloidea exhibit significantly higher *D*
255 statistic values (WRST, all $P < 10^{-4}$), whereas Zygoptera, Anisozygoptera and Cordulegastroidea
256 showed reduced average *D* values (WRST, all $P < 10^{-4}$). Second, we estimated the γ parameter
257 of HyDe's hybrid speciation model [56], where γ and $(1-\gamma)$ corresponds to the parental fractions
258 in a putatively hybrid genome, with values of 1 or 0 indicating the absence of hybridization.
259 Comparisons of γ distributions (Figure 3B) show similar trends except that here Anisozygoptera,
260 Anisoptera and Libellulidae along with Coenagrionoidea, Gomphidae+Petaluridae and
261 Libelluloidea exhibit elevated average γ values (WRST, all $P < 0.00064$) compared to the total
262 distribution of γ. Additionally, we detected an overall excess of quartets with both significant γ
263 and *D* for Anisozygoptera (Fisher exact test [FET], $P = 1.865584 \times 10^{-10}$) and Calopterygoidea
264 (FET, $P = 0.016$) which further support signatures of introgression within these clades (Figure
265 3C). Overall *D* and γ are significantly correlated (Spearman's rank correlation test, $\rho = 0.78$, $P =$
266 0) with a nonlinear relationship, and the majority of quartets with low *D* also have low γ values
267 (brighter region, Figure 3D).
268 There are a variety of other test statistics that have been developed to detect introgression
269 (e.g., [55, 57-59]). Because, like *D* and γ, many of these statistics are computed from different
270 invariants, we attempted to visualize the relationships between all 15 site patterns computed by
271 HyDe using *t*-distributed stochastic neighbor embedding (tSNE; [60]) for dimensionality
272 reduction, along with the corresponding values of *D* and γ (Figures 3 E and F, respectively).
273 Specifically, we showed that quartets with similar signatures of introgression tend to cluster
274 together. Moreover, tSNE maps further demonstrate the correlation between *D* and γ. The

275 clustering of quartets with significant introgression according to a two-dimensional
276 representation of their site patterns may suggest the presence of additional site-pattern signatures
277 of introgression.

278 Further, we tested introgression using an alternative invariant-based method D_{FOIL} (Figure
279 4), which is capable of detecting ancestral as well as inter-group introgression and inferring its
280 polarization (see Method Details). Overall, the distributions of D_{FOIL} statistics (Figure 4A) except
281 for Lestoidea and Aeshnidae were significantly different from the D_{FOIL} distribution within the
282 entire order across all taxonomic levels (WRST, all $P < 3.4\times 10^{-4}$). This observation is indicative
283 of differences in the amount of ancestral introgression, which is determined significance of D_{FOIL}
284 statistics. Further we note that average D_{FOIL} (Figure 4A) was significantly different from 0 (one
285 sample *t*-test [OSTT], all $P < 0.0223$) for all taxa except Lestoidea, Aeshnidae,
286 Gomphidae+Petaluridae and Libellulidae. Together these significant deviations from 0 of D_{FOIL}
287 statistics support the hypothesis of introgression for the tested taxonomic levels [61]. Tests of
288 individual quintets as implemented in D_{FOIL} identified significant cases of introgression within
289 all clades except Lestoidea (Figure 4B). Considering polarization of introgression for
290 Anisozygoptera, almost ubiquitously (1157 out of 1158) cases with positive ancestral
291 introgression quintets suggest that Anisozygoptera is a recipient taxon whereas Zygoptera is a
292 donor taxon (Figure 2). Also, 4473 evaluated quintets with significant inter-group introgression
293 are indicative of one-way introgression from Zygoptera to Anisozygoptera as well. As did D and
294 γ, D_{FOIL} detected an excess of cases that support introgression for Anisozygoptera (FET, $P =$
295 0.04165). The other focal clade, Aeshnidae showed two significant quintets with only inter-

group type introgression (Figure 4B) where *Aeshna palmata* putatively introgressed into *Austroaeschna subapicalis*. For Gomphidae+Petaluridae group D_{FOIL} identified a single case of ancestral introgression where *Phenes raptor* introgressing into ancestral Gomphidae lineage. For Libellulidae, two ancestral and five inter-group introgression scenarios were significant. Specifically, for inter-group introgression we found that *Acisoma variegatum* putatively

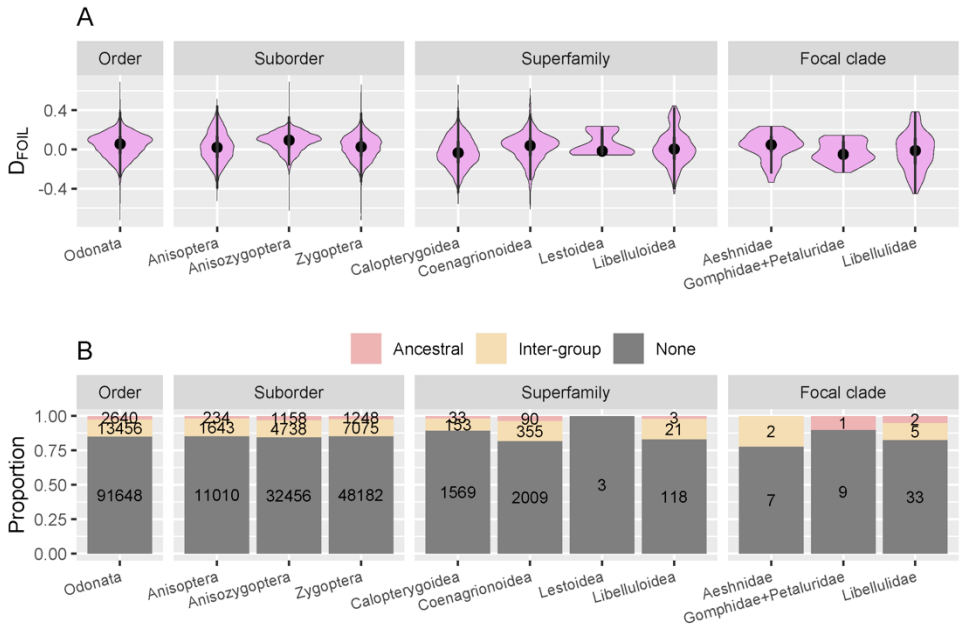

Figure 4. Distributions of the D_{FOIL} Statistics and Their Relations Across Odonate Taxonomic Levels.

(A) Estimated distributions of D_{FOIL} statistics from different site pattern counts for each symmetric quintet. D_{FOIL} allows to determine introgression and its polarization between a donor and recipient taxa by comparison of a sign (+/-/0) for all D_{FOIL} statistics. Black dots mark medians of violin plots. Lestoidea had no significant cases.

(B) Counts of quintets that support ancestral, inter-group or no introgression scenarios based on significance of D_{FOIL} statistics ($P < 0.01$) and their directionality.

introgressed into *Rhyothemis variegate*, *Libellula saturata* into *Ladona fulva*, *Ladona fulva* into *Erythrodiplax connata* and two cases of *Pantala flavescens* into *Sympetrum frequens*. Despite the notion that D_{FOIL} approach exhibits low false positive rate, it requires tree symmetry (Figure 2) [61], thus not all the quintet combinations of taxa can be tested for introgression with this method. For this reason we were unable to form more total quintets for Anisozygoptera, i.e. symmetric trees (((Zygoptera, Zygoptera),(Anisozygoptera, Anisoptera)),Outgroup).

307 Next, we used an approach based on gene tree topology counts to identify introgression
308 within Odonata (Figure 5). Briefly, for a triplet of species, under ILS one would expect equal
309 proportions of gene tree topologies supporting the two topologies disagreeing with the species
310 tree, and any imbalance may suggest introgression. Thus, deviation from equal frequencies of

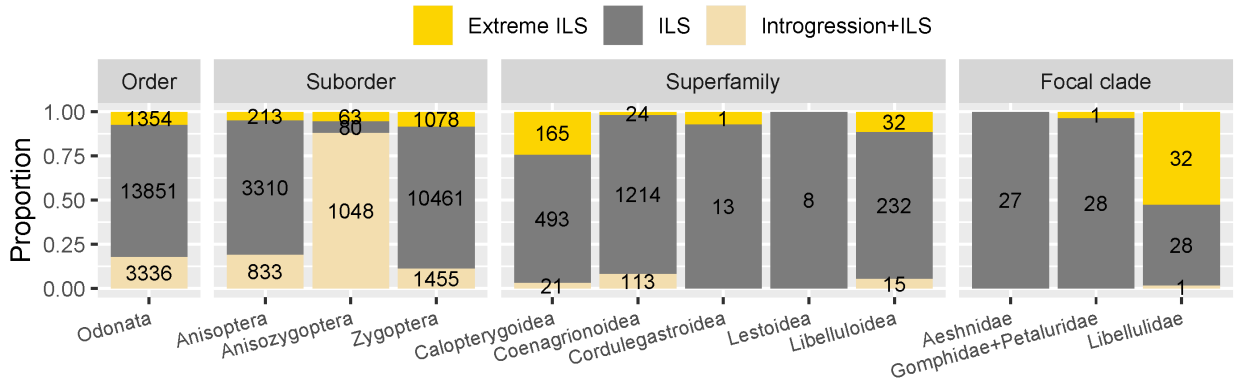

Figure 5. Overview of χ^2 Count Test for Odonate Taxonomic Levels Classification of triplets based on χ^2 test results. Counts of triplets that are concordant with the species tree (FDR < 0.05 and maximum number of gene trees agreeing with a triplet), extreme ILS (FDR > 0.05 for all gene tree counts), ILS (FDR > 0.05 for the discordant gene tree counts) and introgression+ILS (FDR < 0.05 for the discordant gene tree counts)

311 gene tree counts among discordant gene trees can be assessed using a χ^2 test, similar to the
312 method proposed in [62]. With this method, we identified a clear excess of significant triplets
313 that support introgression for Anisozygoptera (FET, $P = 0$), in fact they account for almost 1/3 of
314 all significant cases within Odonata. In agreement with D_{FOIL}, D and γ, we found introgression
315 patterns between Anisozygoptera and Zygoptera. Out of 1048 significant triplets 1021 suggest
316 introgression between Anisozygoptera and Zygoptera. An alternative gene tree branch length-
317 based approach, QuIBL [63], also showed an excess of significant triplets (FET, $P =$
318 9.556092×10^{-6}; Figure S3) compared to the entire order. Strikingly, out of 1042 significant
319 triplets, 1015 suggest introgression between Anisozygoptera and Zygoptera as well. Among
320 other focal clades only Libellulidae has a single triplet (Figure 5) with significant imbalance of
321 discordant gene tree topology counts, which supports introgression of *Ladona fulva* and *Libellula*
322 *forensis*.
323 Finally, we performed phylogenetic network inference in PhyloNet [64, 65] for our four
324 focal clades (Figure 6) specifying a single reticulation scenario. For the Anisozygoptera,

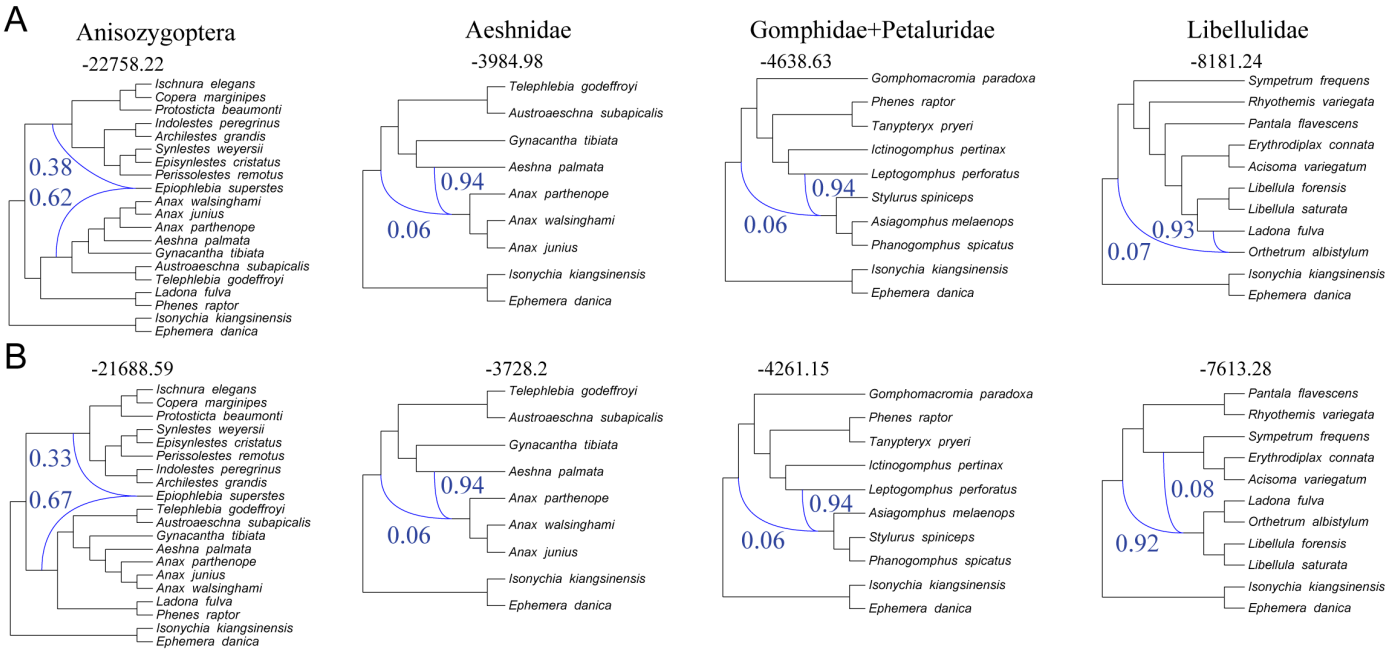

Figure 6. Phylogenetic Networks for Focal Clades Inferred in PhyloNet.

(A) Phylogenetic networks estimated from a set of ML gene trees using maximum pseudolikelihood approach. *Epiophlebia superstes* was specified as a putative hybrid for Anisozygoptera clade. Blue lines indicate a reticulation event with the value of PhyloNet's estimated γ. Numbers above the network indicates the log-likelihood score.

(B) Phylogenetic networks estimated from a set of ML gene trees using maximum likelihood approach. *Epiophlebia superstes* was specified as a putative hybrid for Anisozygoptera clade. Blue lines indicate a reticulation event with the value of PhyloNet's estimated γ. Number above the network indicates the log-likelihood score.

Aeshnidae and Gomphidae+Petaludaridae clades maximum pseudolikelihood (Figure 6A) and full maximum likelihood (Figure 6B) approaches recovered topologically similar networks. In Anisozygoptera pseudolikelihood analysis suggests a reticulation event between Aeshnidae (Anisoptera) and Zygoptera, whereas full likelihood infers reticulation between Anisoptera and Zygoptera suborders. For the Aeshnidae and Gomphidae+Petaludaridae clades both pseudolikelihood and full likelihood approaches both recover the same reticulation. For the Libellulidae clade pseudolikelihood and full likelihood methods did not converge to the same results, however we note significantly better log-likelihood score for the full likelihood network. Overall, we note that the full likelihood approach returned more optimal networks based on the log-likelihood scores. This observation is most likely due to the fact that we performed full likelihood analysis in an exhaustive manner (Data S4) testing every possible reticulation event

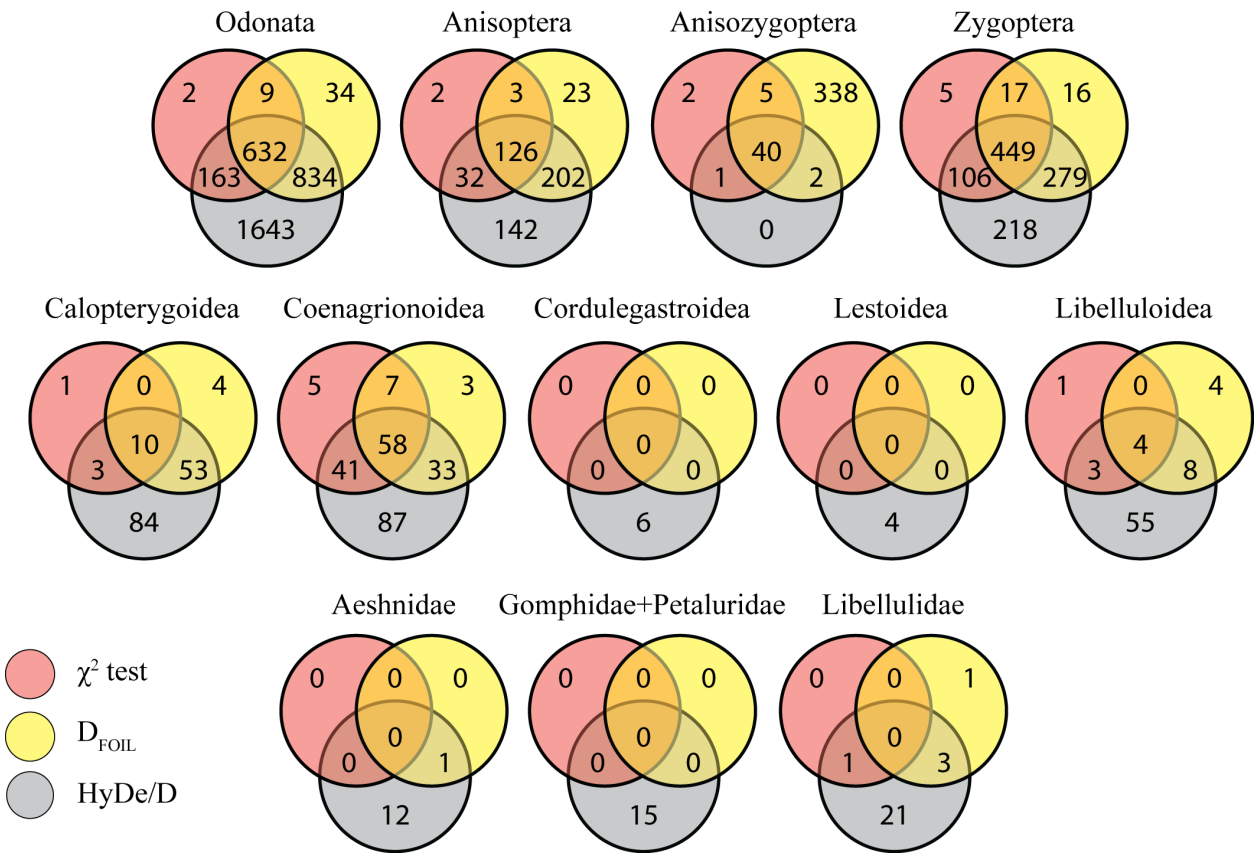

Figure 7. Overlap between Putatively Introgressed Species Pairs Inferred by Hyde/*D*, D_{FOIL} and χ^2.
The numbers within sets represent the number of unique introgressing species pairs identified by a corresponding method.

336 within a focal clade topology congruent with the species tree (Figure 1), whereas
337 pseudolikelihood analysis used the hill-climbing algorithm to search a network space, which is
338 not guaranteed to retrieve the most optimal solution.
339 We examined the agreement across site- and gene tree-based methods for detecting
340 introgression (i.e. Hyde, *D*, and D_{FOIL} and the topology count test) and summarize the results in
341 Figure 7. Specifically, we compared sets of unique introgressing species pairs that were
342 identified with significant pattern of introgression. Overall, all methods were able to identify
343 abundant introgression within the entire Odonata order and the three suborders with strong
344 agreement. However, there was little intersection among the predictions from these methods
345 within our focal groups (except for Anisozygoptera). Notably, our topology count test produced
346 very few predictions that were not in agreement with Hyde, *D* and/or D_{FOIL}. Further, the analyses

347 of phylogenetic networks suggest that we were not able to identify consistent patterns of
348 introgression for Aeshnidae, Gomphidae+Petaluridae and Libellulidae.
349
350 **DISCUSSION**
351
352 Identification of reticulations within the Tree of Life has become possible for non-model
353 organisms due to cost-effective approaches for obtaining genomic data as well as rapid
354 development of phylogenetic tools for testing models of introgression.
355 Introgression among Odonata has been previously thought to be uncommon [66, 67] as a
356 result of probable reproductive isolation mechanisms such as ethological barriers, phenotypic
357 divergence (e.g., variable morphology of genitalia [68, 69]) and habitat and temporal isolation.
358 Contrary to this statement, we found abundant patterns of ancestral introgression (Figure 7)
359 among distantly related odonate taxa. This observation may be explained by reduced sexual
360 selection pressures in the early stages of odonate evolution that may have inhibited rapid genital
361 divergence [70], which is probably a primary source of reproductive isolation in Odonata [71].
362 Most of the introgression and hybridization research in Odonata has been done within the
363 zygopteran Coenagrionidae family and especially between populations of genus *Ishnura*
364 focusing on mechanisms of reproductive isolation (e.g. [25, 72]). These studies established
365 "hybridization thresholds" for these closely related species and showed positive correlation
366 between isolation and genetic divergence. We found multiple patterns of introgression within
367 Coenagrionoidea superfamily (Figures 3-6, S3) which is perhaps attributed to the signal of
368 introgression within several *Ishnura* species included in our taxon sampling.
369 In particular we found compelling evidence for deep ancestral introgression in
370 Anisozygoptera suborder, that based on the fossil record, became genetically isolated after the
371 Lower Jurassic [73] and species of this suborder exhibit anatomical characteristics of both
372 Anisoptera and Zygoptera suborders. Some general features of Anisozygoptera that relate them
373 to Zygoptera include dorsal folding of wings during perching in adults, characteristic anatomy of
374 proventriculus (a part of digestive system that is responsible for grinding of food particles) and
375 absence of specific muscle groups in the larval rectal gills; whereas abdominal tergite shape, rear
376 wing geometry and larval structures are similar to Anisoptera [51]. More recent studies also
377 revealed that Anisozygoptera ovipositor morphology shares similarity with Zygoptera [52];
378 muscle composition of the head resemble characteristics of both Anisoptera and Zygoptera [53];

379 thoracic musculature of Anisozygoptera larva exhibit similarity between Anisoptera and
380 Zygoptera [35]. Thus Anisozygoptera represent a morphological and behavioral "intermediate"
381 [74], which is supported by our findings where we consistently recovered introgression between
382 Zygoptera and Anisozygoptera (Figures 2-7, S3). Moreover, with the D_{FOIL} method the
383 Anisozygoptera was inferred as a recipient taxon from a zygopteran donor. Strikingly, the
384 average HyDe probability parameter $\gamma \sim 0.7$ (Figure 3) inferred from multiple quartets is very
385 similar to PhyloNet's inheritance probability $\gamma \sim 0.62$. This may suggest that between 30–40% of
386 the Anisozygoptera genome descends from zygopteran lineages.
387 Interestingly, we were not able to detect a robust introgression signal within our other three focal
388 clades, namely Aeshnidae, Gomphidae+Petaluridae and Libellulidae, where we previously
389 suspected introgression due to the reasons outlined in the Results section. Phylogenetic
390 instability of Gomphidae+Petaluridae is most likely explained by ILS rather than introgression.
391 Indeed, monophyly of Gomphidae+Petaluridae has been strongly rejected by the ILS-aware
392 phylogenetic reconstructions in ASTRAL (Figure S1A). Overall, due to limited taxon sampling
393 within Aeshnidae and Libellulidae our methods did not infer a clear introgression signal. We
394 note that most of the PhyloNet's γ*'s* for these clades are ≥ 0.92 (Figure 6) which suggests that the
395 majority of loci have not experienced introgression. Further the phylogenetic networks
396 reconstructed by PhyloNet suggest that these clades may contain species that experienced
397 introgression, but for whom the donor taxa were not sampled. In particular, the topology of a
398 reticulation edge that branches off from the base of the tree for Gomphidae+Petaluridae,
399 Aeshnidae and Libellulidae may be indicative of this scenario. Additionally, the absence of the
400 clear introgression signal within these focal clades can be explained by the fact that a potential
401 donor can exist outside of the focal clades or went extinct.
402 Our findings provide the first insight into global patterns of introgression that were
403 pervasive throughout the evolutionary history of Odonata. Further work based on full genome
404 sequencing data will be necessary to accurately estimate the fraction of the introgressed genome
405 in various species. We also stress the importance of further method development to allow for
406 more accurate and scalable inference of phylogenetic networks and enable researchers to pin-
407 point instances of inter-taxonomic introgression on the phylogeny.
408
409

410 **ACKNOWELEGEMENTS**

411 We thank Gavin Martin, Nathan Lord and Camilla Sharkey for the generation of sequence data.
412 We also thank the DNA Sequencing Center and Fulton Supercomputer Lab (BYU) for
413 assistance. This work was supported by the NSF grant (SMB; DEB-1265714) and NIH grant
414 (DRS; R00HG008696).

415 **AUTHOR CONTRIBUTIONS**

416 AS, CS and SMB conceived the research. AS, CS, MSF and PB analyzed the data. MC, KAC,
417 MFW and DRS supervised the project. AS, CS, DRS and SMB wrote the manuscript.

418

419 **DECLARATION OF INTERESTS**

420 The authors declare no competing interests.

421

422 **REFERENCES**

423

424 1. Mallet, J., Besansky, N., and Hahn, M.W. (2016). How reticulated are species? Bioessays
425 *38*, 140-149.
426 2. Hallstrom, B.M., and Janke, A. (2010). Mammalian evolution may not be strictly
427 bifurcating. Molecular biology and evolution *27*, 2804-2816.
428 3. Degnan, J.H., and Rosenberg, N.A. (2009). Gene tree discordance, phylogenetic
429 inference and the multispecies coalescent. Trends in ecology & evolution *24*, 332-340.
430 4. Posada, D., and Crandall, K.A. (2001). Intraspecific gene genealogies: trees grafting into
431 networks. Trends in ecology & evolution *16*, 37-45.
432 5. Mallet, J. (2005). Hybridization as an invasion of the genome. Trends in ecology &
433 evolution *20*, 229-237.
434 6. Petr, M., Paabo, S., Kelso, J., and Vernot, B. (2019). Limits of long-term selection
435 against Neandertal introgression. Proc Natl Acad Sci U S A *116*, 1639-1644.
436 7. Oziolor, E.M., Reid, N.M., Yair, S., Lee, K.M., Guberman VerPloeg, S., Bruns, P.C.,
437 Shaw, J.R., Whitehead, A., and Matson, C.W. (2019). Adaptive introgression enables
438 evolutionary rescue from extreme environmental pollution. Science *364*, 455-457.
439 8. Heliconius Genome, C. (2012). Butterfly genome reveals promiscuous exchange of
440 mimicry adaptations among species. Nature *487*, 94-98.
441 9. Huerta-Sanchez, E., Jin, X., Asan, Bianba, Z., Peter, B.M., Vinckenbosch, N., Liang, Y.,
442 Yi, X., He, M., Somel, M., et al. (2014). Altitude adaptation in Tibetans caused by
443 introgression of Denisovan-like DNA. Nature *512*, 194-197.
444 10. Gouy, A., and Excoffier, L. (2020). Polygenic Patterns of Adaptive Introgression in
445 Modern Humans Are Mainly Shaped by Response to Pathogens. Molecular biology and
446 evolution *37*, 1420-1433.

11. Perry, W.L., Lodge, D.M., and Feder, J.L. (2002). Importance of hybridization between indigenous and nonindigenous freshwater species: an overlooked threat to North American biodiversity. Syst Biol *51*, 255-275.
12. Dijkstra, K.D., Kalkman, V.J., Dow, R.A., Stokvis, F.R., and Van Tol, J. (2014). Redefining the damselfly families: a comprehensive molecular phylogeny of Zygoptera (Odonata). Syst Entomol *39*, 68-96.
13. Carle, F.L., Kjer, K.M., and May, M.L. (2015). A molecular phylogeny and classification of Anisoptera (Odonata). Arthropod Syst Phylo *73*, 281-301.
14. Thomas, J.A., Trueman, J.W., Rambaut, A., and Welch, J.J. (2013). Relaxed phylogenetics and the palaeoptera problem: resolving deep ancestral splits in the insect phylogeny. Syst Biol *62*, 285-297.
15. Busse, S., Genet, C., and Hornschemeyer, T. (2013). Homologization of the flight musculature of zygoptera (insecta: odonata) and neoptera (insecta). PloS one *8*, e55787.
16. Chauhan, P., Hansson, B., Kraaijeveld, K., de Knijff, P., Svensson, E.I., and Wellenreuther, M. (2014). De novo transcriptome of Ischnura elegans provides insights into sensory biology, colour and vision genes. BMC Genomics *15*, 808.
17. Suvorov, A., Jensen, N.O., Sharkey, C.R., Fujimoto, M.S., Bodily, P., Wightman, H.M., Ogden, T.H., Clement, M.J., and Bybee, S.M. (2017). Opsins have evolved under the permanent heterozygote model: insights from phylotranscriptomics of Odonata. Molecular ecology *26*, 1306-1322.
18. Corbet, P.S. (1999). Dragonflies : behavior and ecology of Odonata, (Ithaca, N.Y.: Comstock Pub. Associates).
19. Troast, D., Suhling, F., Jinguji, H., Sahlen, G., and Ware, J. (2016). A Global Population Genetic Study of Pantala flavescens. PloS one *11*, e0148949.
20. Dijkstra, K.D., Monaghan, M.T., and Pauls, S.U. (2014). Freshwater biodiversity and aquatic insect diversification. Annual review of entomology *59*, 143-163.
21. Cordoba-Aguilar, A. (2008). Dragonflies and damselflies : model organisms for ecological and evolutionary research, (Oxford ; New York: Oxford University Press).
22. Bybee, S., Córdoba-Aguilar, A., Duryea, M.C., Futahashi, R., Hansson, B., Lorenzo-Carballa, M.O., Schilder, R., Stoks, R., Suvorov, A., Svensson, E.I., et al. (2016). Odonata (dragonflies and damselflies) as a bridge between ecology and evolutionary genomics. Frontiers in zoology *13*, 46.
23. Monetti, L., SÁNchez-GuillÉN, R.A., and Rivera, A.C. (2002). Hybridization between Ischnura graellsii (Vander Linder) and I. elegans (Rambur) (Odonata: Coenagrionidae): are they different species? Biological Journal of the Linnean Society *76*, 225-235.
24. SÁNchez-GuillÉN, R.A., Van Gossum, H., and Cordero Rivera, A. (2005). Hybridization and the inheritance of female colour polymorphism in two ischnurid damselflies (Odonata: Coenagrionidae). Biological Journal of the Linnean Society *85*, 471-481.
25. Sanchez-Guillen, R.A., Wellenreuther, M., Cordero-Rivera, A., and Hansson, B. (2011). Introgression and rapid species turnover in sympatric damselflies. BMC evolutionary biology *11*, 210.
26. Hayashi, F., Dobata, S., and Futahashi, R. (2005). Disturbed population genetics: suspected introgressive hybridization between two Mnais damselfly species (Odonata). Zoological science *22*, 869-881.

491 27. Tynkkynen, K., Grapputo, A., Kotiaho, J.S., Rantala, M.J., Väänänen, S., and Suhonen, J.
492 (2008). Hybridization in Calopteryx damselflies: the role of males. Animal behaviour *75*,
493 1431-1439.
494 28. Solano, E., Hardersen, S., Audisio, P., Amorosi, V., Senczuk, G., and Antonini, G.
495 (2018). Asymmetric hybridization in Cordulegaster (Odonata: Cordulegastridae):
496 Secondary postglacial contact and the possible role of mechanical constraints. Ecology
497 and evolution *8*, 9657-9671.
498 29. Misof, B., Liu, S., Meusemann, K., Peters, R.S., Donath, A., Mayer, C., Frandsen, P.B.,
499 Ware, J., Flouri, T., Beutel, R.G., et al. (2014). Phylogenomics resolves the timing and
500 pattern of insect evolution. Science *346*, 763-767.
501 30. Trueman, J.W.H. (1996). A preliminary cladistic analysis of odonate wing venation.
502 Odonatologica *25*, 59-72.
503 31. Saux, C., Simon, C., and Spicer, G.S. (2003). Phylogeny of the dragonfly and damselfly
504 order odonata as inferred by mitochondrial 12S ribosomal RNA sequences. Annals of the
505 Entomological Society of America *96*, 693-699.
506 32. Ogden, T.H., and Whiting, M.F. (2003). The problem with "the Paleoptera Problem:"
507 sense and sensitivity. Cladistics *19*, 432-442.
508 33. Hasegawa, E., and Kasuya, E. (2006). Phylogenetic analysis of the insect order Odonata
509 using 28S and 16S rDNA sequences: a comparison between data sets with different
510 evolutionary rates. Entomological Science *9*, 55-66.
511 34. Pfau, H.K. (1991). Contributions of functional morphology to the phylogenetic
512 systematics of Odonata. Advances in Odonatology *5*, 109-141.
513 35. Busse, S., Helmker, B., and Hornschemeyer, T. (2015). The thorax morphology of
514 Epiophlebia (Insecta: Odonata) nymphs--including remarks on ontogenesis and
515 evolution. Scientific reports *5*, 12835.
516 36. Kim, M.J., Jung, K.S., Park, N.S., Wan, X., Kim, K.-G., Jun, J., Yoon, T.J., Bae, Y.J.,
517 Lee, S.M., and Kim, I. (2014). Molecular phylogeny of the higher taxa of Odonata
518 (Insecta) inferred from COI, 16S rRNA, 28S rRNA, and EF1-α sequences. Entomological
519 Research *44*, 65-79.
520 37. Carle, F.L., Kjer, K.M., and May, M.L. (2008). Evolution of Odonata, with special
521 reference to Coenagrionoidea (Zygoptera). Arthropod Systematics and Phylogeny *66*, 37-
522 44.
523 38. Bybee, S.M., Ogden, T.H., Branham, M.A., and Whiting, M.F. (2008). Molecules,
524 morphology and fossils: a comprehensive approach to odonate phylogeny and the
525 evolution of the odonate wing. Cladistics *24*, 477-514.
526 39. Letsch, H., Gottsberger, B., and Ware, J.L. (2016). Not going with the flow: a
527 comprehensive time-calibrated phylogeny of dragonflies (Anisoptera: Odonata: Insecta)
528 provides evidence for the role of lentic habitats on diversification. Molecular ecology *25*,
529 1340-1353.
530 40. Britton, T., Anderson, C.L., Jacquet, D., Lundqvist, S., and Bremer, K. (2007).
531 Estimating divergence times in large phylogenetic trees. Syst Biol *56*, 741-752.
532 41. Blanke, A., Greve, C., Mokso, R., Beckmann, F., and Misof, B. (2013). An updated
533 phylogeny of Anisoptera including formal convergence analysis of morphological
534 characters. Syst Entomol *38*, 474-490.

42. Misof, B., Rickert, A.M., Buckley, T.R., Fleck, G., and Sauer, K.P. (2001). Phylogenetic signal and its decay in mitochondrial SSU and LSU rRNA gene fragments of Anisoptera. Molecular biology and evolution *18*, 27-37.
43. Rehn, A.C. (2003). Phylogenetic analysis of higher-level relationships of Odonata. Syst Entomol *28*, 181-240.
44. Roch, S., and Steel, M. (2014). Likelihood-based tree reconstruction on a concatenation of aligned sequence data sets can be statistically inconsistent. Theoretical population biology *100C*, 56-62.
45. Kubatko, L.S., and Degnan, J.H. (2007). Inconsistency of phylogenetic estimates from concatenated data under coalescence. Syst Biol *56*, 17-24.
46. Maddison, W.P. (1997). Gene Trees in Species Trees. Syst Biol *46*, 523-536.
47. Pease, J.B., Brown, J.W., Walker, J.F., Hinchliff, C.E., and Smith, S.A. (2018). Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. Am J Bot *105*, 385-403.
48. Labandeira, C.C., and Sepkoski, J.J., Jr. (1993). Insect diversity in the fossil record. Science *261*, 310-315.
49. Grimaldi, D.A., and Engel, M.S. (2005). Evolution of the insects, (Cambridge, UK ; New York, NY: Cambridge University Press).
50. Nel, A., Bechly, G., Martinez-Delclos, X., and Fleck, G. (2001). A new family of Anisoptera from the Upper Jurassic of Karatau in Kazakhstan (Insecta: Odonata: Juragomphidae n. fam.). Stuttgarter Beitraege zur Naturkunde Serie B (Geologie und Palaeontologie) *314*, 1-9.
51. Asahina, S., Asahina, S., and Nihon Gakujutsu, S. (1954). A Morphological Study of a Relic Dragonfly Epiophlebia Superstes Selys: (Odonata, Anisozygoptera), (Japan Society for the Promotion of Science).
52. Matushkina, N.A. (2008). The ovipositor of the relic dragonfly Epiophlebia superstes: a morphological re-examination (Odonata: Epiophlebiidae). International Journal of Odonatology *11*, 71-80.
53. Blanke, A., Beckmann, F., and Misof, B. (2013). The head anatomy of Epiophlebia superstes (Odonata: Epiophlebiidae). Organisms Diversity & Evolution *13*, 55-66.
54. Patterson, N., Moorjani, P., Luo, Y., Mallick, S., Rohland, N., Zhan, Y., Genschoreck, T., Webster, T., and Reich, D. (2012). Ancient admixture in human history. Genetics *192*, 1065-1093.
55. Martin, S.H., Davey, J.W., and Jiggins, C.D. (2015). Evaluating the use of ABBA-BABA statistics to locate introgressed loci. Molecular biology and evolution *32*, 244-257.
56. Meng, C., and Kubatko, L.S. (2009). Detecting hybrid speciation in the presence of incomplete lineage sorting using gene tree incongruence: a model. Theor Popul Biol *75*, 35-45.
57. Durand, E.Y., Patterson, N., Reich, D., and Slatkin, M. (2011). Testing for ancient admixture between closely related populations. Molecular biology and evolution *28*, 2239-2252.
58. Green, R.E., Krause, J., Briggs, A.W., Maricic, T., Stenzel, U., Kircher, M., Patterson, N., Li, H., Zhai, W., Fritz, M.H., et al. (2010). A draft sequence of the Neandertal genome. Science *328*, 710-722.

59. Kubatko, L.S., and Chifman, J. (2019). An invariants-based method for efficient identification of hybrid species from large-scale genomic data. BMC evolutionary biology *19*, 112.
60. van der Maaten, L., and Hinton, G. (2008). Visualizing Data using t-SNE. Journal of Machine Learning Research *9*, 2579-2605.
61. Pease, J.B., and Hahn, M.W. (2015). Detection and Polarization of Introgression in a Five-Taxon Phylogeny. Syst Biol *64*, 651-662.
62. Degnan, J.H., and Rosenberg, N.A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. Trends in ecology & evolution *24*, 332-340.
63. Edelman, N.B., Frandsen, P.B., Miyagi, M., Clavijo, B., Davey, J., Dikow, R.B., Garcia-Accinelli, G., Van Belleghem, S.M., Patterson, N., Neafsey, D.E., et al. (2019). Genomic architecture and introgression shape a butterfly radiation. Science *366*, 594-599.
64. Than, C., Ruths, D., and Nakhleh, L. (2008). PhyloNet: a software package for analyzing and reconstructing reticulate evolutionary relationships. BMC bioinformatics *9*, 322.
65. Wen, D., Yu, Y., Zhu, J., and Nakhleh, L. (2018). Inferring Phylogenetic Networks Using PhyloNet. Syst Biol *67*, 735-740.
66. Tennessen, K. (1982). Review of reproductive isolating barriers in Odonata. Advances in Odonatology *1*, 251-265.
67. Lowe, C., Harvey, I., Thompson, D., and Watts, P. (2008). Strong genetic divergence indicates that congeneric damselflies Coenagrion puella and C. pulchellum (Odonata: Zygoptera: Coenagrionidae) do not hybridise. Hydrobiologia *605*, 55-63.
68. Hosken, D.J., and Stockley, P. (2004). Sexual selection and genital evolution. Trends in ecology & evolution *19*, 87-93.
69. Barnard, A.A., Fincke, O.M., McPeek, M.A., and Masly, J.P. (2017). Mechanical and tactile incompatibilities cause reproductive isolation between two young damselfly species. Evolution; international journal of organic evolution *71*, 2410-2427.
70. Eberhard, W.G. (2004). Rapid divergent evolution of sexual morphology: comparative tests of antagonistic coevolution and traditional female choice. Evolution; international journal of organic evolution *58*, 1947-1970.
71. Cordero Rivera, A., Andres, J.A., Cordoba-Aguilar, A., and Utzeri, C. (2004). Postmating sexual selection: allopatric evolution of sperm competition mechanisms and genital morphology in calopterygid damselflies (Insecta: Odonata). Evolution; international journal of organic evolution *58*, 349-359.
72. Sanchez-Guillen, R.A., Cordoba-Aguilar, A., Cordero-Rivera, A., and Wellenreuther, M. (2014). Genetic divergence predicts reproductive isolation in damselflies. Journal of evolutionary biology *27*, 76-87.
73. Carle, F. (2012). A new Epiophlebia (Odonata: Epiophlebioidea) from China with a review of epiophlebian taxonomy, life history, and biogeography. Arthropod Systematics & Phylogeny *70*, 75–83.
74. Liu, J., Yu, L., Arnold, M.L., Wu, C.H., Wu, S.F., Lu, X., and Zhang, Y.P. (2011). Reticulate evolution: frequent introgressive hybridization among Chinese hares (genus lepus) revealed by analyses of multiple mitochondrial and nuclear DNA loci. BMC evolutionary biology *11*, 223.
75. Grabherr, M.G., Haas, B.J., Yassour, M., Levin, J.Z., Thompson, D.A., Amit, I., Adiconis, X., Fan, L., Raychowdhury, R., Zeng, Q., et al. (2011). Full-length

624 transcriptome assembly from RNA-Seq data without a reference genome. Nature
625 biotechnology *29*, 644-652.
626 76. Haas, B.J., Papanicolaou, A., Yassour, M., Grabherr, M., Blood, P.D., Bowden, J.,
627 Couger, M.B., Eccles, D., Li, B., Lieber, M., et al. (2013). De novo transcript sequence
628 reconstruction from RNA-seq using the Trinity platform for reference generation and
629 analysis. Nature protocols *8*, 1494-1512.
630 77. Buchfink, B., Xie, C., and Huson, D.H. (2015). Fast and sensitive protein alignment
631 using DIAMOND. Nat Methods *12*, 59-60.
632 78. Fu, L., Niu, B., Zhu, Z., Wu, S., and Li, W. (2012). CD-HIT: accelerated for clustering
633 the next-generation sequencing data. Bioinformatics *28*, 3150-3152.
634 79. Simao, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., and Zdobnov, E.M.
635 (2015). BUSCO: assessing genome assembly and annotation completeness with single-
636 copy orthologs. Bioinformatics *31*, 3210-3212.
637 80. Li, L., Stoeckert, C.J., Jr., and Roos, D.S. (2003). OrthoMCL: identification of ortholog
638 groups for eukaryotic genomes. Genome Res *13*, 2178-2189.
639 81. Yang, Y., and Smith, S.A. (2014). Orthology inference in nonmodel organisms using
640 transcriptomes and low-coverage genomes: improving accuracy and matrix occupancy
641 for phylogenomics. Molecular biology and evolution *31*, 3081-3092.
642 82. Eddy, S.R. (2011). Accelerated Profile HMM Searches. PLoS computational biology *7*,
643 e1002195.
644 83. Fujimoto, M.S., Suvorov, A., Jensen, N.O., Clement, M.J., and Bybee, S.M. (2016).
645 Detecting false positive sequence homology: a machine learning approach. BMC
646 bioinformatics *17*, 101.
647 84. Fujimoto, M.S., Suvorov, A., Jensen, N.O., Clement, M.J., Snell, Q., and Bybee, S.M.
648 (2016). The OGCleaner: filtering false-positive homology clusters. Bioinformatics.
649 85. Mirarab, S., Nguyen, N., Guo, S., Wang, L.S., Kim, J., and Warnow, T. (2015). PASTA:
650 Ultra-Large Multiple Sequence Alignment for Nucleotide and Amino-Acid Sequences.
651 Journal of computational biology : a journal of computational molecular cell biology *22*,
652 377-386.
653 86. Nguyen, L.T., Schmidt, H.A., von Haeseler, A., and Minh, B.Q. (2015). IQ-TREE: a fast
654 and effective stochastic algorithm for estimating maximum-likelihood phylogenies.
655 Molecular biology and evolution *32*, 268-274.
656 87. Loytynoja, A., and Goldman, N. (2008). Phylogeny-aware gap placement prevents errors
657 in sequence alignment and evolutionary analysis. Science *320*, 1632-1635.
658 88. Loytynoja, A. (2014). Phylogeny-aware alignment with PRANK. Methods in molecular
659 biology *1079*, 155-170.
660 89. Misof, B., and Misof, K. (2009). A Monte Carlo approach successfully identifies
661 randomness in multiple sequence alignments: a more objective means of data exclusion.
662 Syst Biol *58*, 21-34.
663 90. Mirarab, S., Reaz, R., Bayzid, M.S., Zimmermann, T., Swenson, M.S., and Warnow, T.
664 (2014). ASTRAL: genome-scale coalescent-based species tree estimation. Bioinformatics
665 *30*, i541-548.
666 91. Lanfear, R., Frandsen, P.B., Wright, A.M., Senfeld, T., and Calcott, B. (2016).
667 PartitionFinder 2: New Methods for Selecting Partitioned Models of Evolution for
668 Molecular and Morphological Phylogenetic Analyses. Molecular biology and evolution.

92. Aberer, A.J., Kobert, K., and Stamatakis, A. (2014). ExaBayes: massively parallel bayesian tree inference for the whole-genome era. Molecular biology and evolution *31*, 2553-2556.
93. Yi, H., and Jin, L. (2013). Co-phylog: an assembly-free phylogenomic approach for closely related organisms. Nucleic acids research *41*, e75.
94. Yang, Z. (2007). PAML 4: phylogenetic analysis by maximum likelihood. Molecular biology and evolution *24*, 1586-1591.
95. Puttick, M.N. (2019). MCMCtreeR: functions to prepare MCMCtree analyses and visualize posterior ages on trees. Bioinformatics *35*, 5321-5322.
96. Blischak, P.D., Chifman, J., Wolfe, A.D., and Kubatko, L.S. (2018). HyDe: A Python Package for Genome-Scale Hybridization Detection. Syst Biol *67*, 821-829.
97. Hellmuth, M., Wieseke, N., Lechner, M., Lenhof, H.P., Middendorf, M., and Stadler, P.F. (2015). Phylogenomics with paralogs. Proc Natl Acad Sci U S A *112*, 2058-2063.
98. Wickett, N.J., Mirarab, S., Nguyen, N., Warnow, T., Carpenter, E., Matasci, N., Ayyampalayam, S., Barker, M.S., Burleigh, J.G., Gitzendanner, M.A., et al. (2014). Phylotranscriptomic analysis of the origin and early diversification of land plants. Proc Natl Acad Sci U S A *111*, E4859-4868.
99. Lanfear, R., Calcott, B., Kainer, D., Mayer, C., and Stamatakis, A. (2014). Selecting optimal partitioning schemes for phylogenomic datasets. BMC evolutionary biology *14*, 82.
100. Minh, B.Q., Nguyen, M.A., and von Haeseler, A. (2013). Ultrafast approximation for phylogenetic bootstrap. Molecular biology and evolution *30*, 1188-1195.
101. Gadagkar, S.R., Rosenberg, M.S., and Kumar, S. (2005). Inferring species phylogenies from multiple genes: concatenated sequence tree versus consensus gene tree. J Exp Zool B Mol Dev Evol *304*, 64-74.
102. Lakner, C., van der Mark, P., Huelsenbeck, J.P., Larget, B., and Ronquist, F. (2008). Efficiency of Markov chain Monte Carlo tree proposals in Bayesian phylogenetics. Syst Biol *57*, 86-103.
103. Brooks, S.P., and Gelman, A. (1998). General Methods for Monitoring Convergence of Iterative Simulations. Journal of Computational and Graphical Statistics *7*, 434-455.
104. Lanfear, R., Hua, X., and Warren, D.L. (2016). Estimating the Effective Sample Size of Tree Topologies from Bayesian Phylogenetic Analyses. Genome biology and evolution *8*, 2319-2332.
105. Sayyari, E., and Mirarab, S. (2016). Fast Coalescent-Based Computation of Local Branch Support from Quartet Frequencies. Molecular biology and evolution *33*, 1654-1668.
106. dos Reis, M., and Yang, Z. (2011). Approximate likelihood calculation on a phylogeny for Bayesian estimation of divergence times. Molecular biology and evolution *28*, 2161-2172.

713 **METHOD DETAILS**

714

715 **Taxon sampling and RNA-seq**

716 In this study, we used distinct 85 species (83 ingroup and 2 outgroup taxa). 35 RNA-seq libraries
717 were obtained from NCBI (see Data S1). The remaining 58 libraries were sequenced in Bybee
718 Lab (some species have several RNA-seq libraries; Data S1). Total RNA was extracted for each
719 taxon from eye tissue using NucleoSpin columns (Clontech) and reverse-transcribed into cDNA
720 libraries using the Illumina TruSeq RNA v2 sample preparation kit that both generates and
721 amplifies full-length cDNAs. Prepped mRNA libraries with insert size of ~200bp were
722 multiplexed and sequenced on an Illumina HiSeq 2000 producing 101-bp paired-end reads by the
723 Microarray and Genomic Analysis Core Facility at the Huntsman Cancer Institute at the
724 University of Utah, Salt Lake City, UT, USA. Quality scores, tissue type and other information
725 about RNA-seq libraries are summarized in Data S1.

726 **Transcriptome assembly and CDS prediction**

727 RNA-seq libraries were trimmed and de novo assembled using Trinity [75, 76] with default
728 parameters. Then only the longest isoform was selected from each gene for downstream analyses
729 using Trinity utility script. In order to identify potentially coding regions within the
730 transcriptomes, we used TransDecoder with default parameters specifying to predict only the
731 single best ORF. Each predicted proteome was screened for contamination using DIMOND
732 BLASTP [77] with an E-value cutoff of 10^{-10} against custom protein database. Non-arthropod
733 hits were discarded from proteomes (amino acid, AA sequences) and corresponding CDSs. To
734 mitigate redundancy in proteomes and CDSs, we used CD-HIT [78] with the identity threshold
735 of 0.99. Such a conservative threshold was used to prevent exclusion of true paralogous
736 sequences; thus, reducing possible false positive detection of 1:1 orthologs during homology
737 searches.

738 **Homology assessment**

739 In the present study three types of homologous loci (gene clusters) namely conserved single-
740 copy orthologs (CO), all single-copy orthologs (AO) and paralogy-parsed orthologs (PO)
741 identified by BUSCO v1.22 [79], OrthoMCL [80] and Yang's [81] pipelines respectively were
742 used in phylogenetic inference.

743 BUSCO arthropod Hidden Markov Model Profiles of 2675 single-copy orthologs were
744 used to find significant COs matches within CDS datasets by HMMER's *hmmersearch* v3 [82]
745 with group-specific expected bit-score cutoffs. BUSCO classifies loci into complete [duplicated]
746 and fragmented. Thus, only complete single-copy loci were extracted from CDS datasets and
747 corresponding AA sequences for further phylogenetic analyses. Since loci were identified as true
748 orthologs if they score above expected bit-score and complete if their lengths lie within ~95% of
749 BUSCO group mean length, many partial erroneously assembled sequences were filtered out.
750 OrthoMCL v2.0.9 [80] was used to compute AOs in all species using predicted AA
751 sequences by TransDecoder. AA sequences were used in an all-vs-all BLASTP with an E-value
752 cutoff of 10^{-10} to find putative orthologs and paralogs. The Markov Cluster algorithm (MCL)
753 inflation point parameter was set to 2. Only 1:1 orthologs were used in further analyses. In order
754 to exclude false-positive homology clusters identified by OrthoMCL, we applied machine
755 learning filtering procedure [83] implemented in OGCleaner software v1.0 [84] using a
756 metaclassifier with logistic regression.
757 Finally, to identify additional clusters, we used Yang's tree-based orthology inference
758 pipeline [81] that was specifically designed for non-model organisms using transcriptomic data.
759 Yang's algorithm is capable of parsing paralogous gene families into "orthology" clusters that
760 can be used in phylogenetic analyses. It has been shown that paralogous sequences encompass
761 useful phylogenetic information [97]. First, the Transdecoder-predicted AA sequences were
762 trimmed using CD-HIT with the identity threshold of 0.995. Then, all-vs-all BLASTP with an E-
763 value cutoff of 10^{-5} search was implemented. The raw BLASTP output was filtered by hit
764 fraction of 0.4. Then, MCL clustering was performed with inflation point parameter of 2. Each
765 cluster was aligned using iterative algorithm of PASTA [85] and then was used to infer a
766 maximum-likelihood (ML) gene tree using IQ-TREE v1.5.2 [86] with an automatic model
767 selection. Tree tips that were longer than relative and absolute cutoffs of 0.4 and 1 respectively
768 were removed. Mono- and paraphyletic tips that belonged to the same species were masked as
769 well. To increase quality of homology clusters realignment, tree inference and tip masking steps
770 were iterated with more stringent relative and absolute masking cutoffs of 0.2 and 0.5
771 respectively. Finally, POs (AA sequences and corresponding CDSs) were extracted by rooted
772 ingroups (RI) procedure using *Ephemera danica* as an outgroup (for details see [81]).
773

774 **Cluster alignment, trimming and supermatrix assembly**

775 For most of the analyses only clusters with ≥ 42 (~50%) species present were retained. In total,
776 we obtained five cluster types, namely DNA (CDS) and AA COs, AA AOs and DNA and AA
777 POs. Each cluster was aligned using PASTA [85] for the DNA and AA alignments and PRANK
778 v150803 [87, 88] for the codon alignments and alignments with removed either 1st and 2nd or 3rd
779 codon positions. In order to reduce the amount of randomly aligned regions, we implemented
780 ALISCORE v2.0 [89] trimming procedure (for PASTA alignments) followed by masking any
781 site with ≥ 42 gap characters (for both PASTA and PRANK alignments). Also, sequence
782 fragments with >50% gap characters were removed from clusters that were subjected to
783 ASTRAL v4.10.12 [90] estimation since fragmentary data may have a negative effect on
784 accuracy of gene and hence species tree inference [98]. For each of the cluster type, we
785 assembled supermatrices from trimmed gene alignments. Additionally, completely untrimmed
786 supermatrices were generated from DNA and AA COs with ≥ 5 species present.

787

788 **Phylogenetic tree reconstruction**

789 Four contrasting tree building methods (ML:IQTREE, Bayesian:ExaBayes, Supertree:ASTRAL,
790 Alignment-Free (AF): Co-phylog) were used to infer odonate phylogenetic relationships using
791 different input data types (untrimmed and trimmed supermatrices, codon supermatrices, codon
792 supermatrices with 1st and 2nd or 3rd positions removed, gene trees and assembled
793 transcriptomes). In total we performed 48 phylogenetic analyses and compared topologies to
794 identify stable and conflicting relationships (Data S2).

795 We inferred phylogenetic ML trees from each supermatrix using IQTREE implementing
796 two partitioning schemes: single partition and those identified by PartitionFinder v2.0 (three
797 GTR models for DNA and a large array of protein models for AA) [91] with relaxed hierarchical
798 clustering option [99]. In the first case, IQTREE was run allowing model selection and assessing
799 nodal support with 1000 ultrafast bootstrap (UFBoot) [100] replicates. In the second case,
800 IQTREE was run with a given PartitionFinder partition model applying gene and site resampling
801 to minimize false-positives [101] for 1000 UFBoot replicates.

802 For Bayesian analyses implemented in ExaBayes [92], we used highly trimmed (retaining
803 sites only with occupancy of ≤ 5 gap characters) and original DNA and AA CO supermatrices
804 assuming a single partition. We initiated 4 independent runs with 4 Markov Chain Monte Carlo

805 (MCMC) coupled chains sampling every 500th iteration. Due to high computational demands of
806 the procedure, only the GTR and JTT substitution model priors were applied to DNA and AA
807 CO supermatrices respectively with the default topology, rate heterogeneity and branch lengths
808 priors. However, all supported protein substitution models as a prior was specified for the
809 trimmed AA CO supermatrix. For convergence criteria an average standard deviation of split
810 frequencies (ASDSF) [102], a potential scale reduction factor (PSRF) [103] and an effective
811 sample size (ESS) [104] were utilized. Values of 0% < ASDSF <1% and 1% < ASDSF < 5%
812 indicate excellent and acceptable convergence respectively; ESS >100 and PSRF ~ 1 represent
813 good convergence (see ExaBayes manual, [92]).
814 ASTRAL analyses were conducted using two input types: (i) gene trees obtained by IQ-
815 TREE allowing model selection for fully trimmed DNA and AA clusters and (ii) gene trees
816 obtained from the alignment-tree coestimation process in PASTA. Nodal support was assessed
817 by local posterior probabilities [105]. ASTRAL is a statistically consistent supertree method
818 under the multispecies coalescent model with better accuracy than other similar approaches [90].
819 In addition to standard phylogenetic inferential approaches, we applied an alignment-free
820 (AF) species tree estimation algorithm using Co-phylog [93]. Raw Transdecoder CDS outputs
821 were used in this analysis using k-mer size of 9 as the half context length required for Co-phylog.
822 Bootstrap replicate trees were obtained by running Co-Phylog with the same parameter settings
823 on each subsampled with replacement CDS Transdecoder libraries and were used to assess nodal
824 support.
825
826 **Assessment of phylogenetic support via quartet sampling**
827 As an additional phylogenetic support, we implemented quartet sampling (QS) approach [47].
828 Briefly, it provides three scores for internal nodes: (i) the quartet concordance (QC) score gives
829 an estimate of how sampled quartet topologies agree with the putative species tree; (ii) quartet
830 differential (QD) estimates frequency skewness of the discordant quartet topologies, which can
831 be indicative of introgression if a skewed frequency is observed and (iii) quartet informativeness
832 (QI) quantifies how informative sampled quartets are by comparing likelihood scores of
833 alternative quartet topologies. Finally, QS provides a quartet fidelity (QF) score for terminal
834 nodes that measures a taxon "rogueness". We performed QS analysis with all 48 putative species
835 phylogenies using SuperMatrix_50BUSCO_dna_pasta_ali_trim supermatrix, specifying an IQ-

836 TREE engine for quartet likelihood calculations with 100 replicates (i.e. number of quartet draws
837 per focal branch).
838
839 **Fossil dating**
840 A Bayesian algorithm of MCMCTREE [94] with approximate likelihood computation was
841 implemented to estimate divergence times within Odonata using 20 crown group fossil
842 constraints with corresponding prior distributions (Data S3). First, we estimated branch length by
843 ML and then the gradient and Hessian matrix around these ML estimates in MCMCTREE using
844 SuperMatrix_50BUSCO_dna_pasta_ali_trim supermatrix. Second, we used the gradient and
845 Hessian matrix which construct0 an approximate likelihood function by Taylor expansion [106]
846 to perform fossil calibration in MCMC framework. For this step we specified GTR+Γ
847 substitution model with 4 gamma categories; birth, death and sampling parameters of 1, 0.5 and
848 0.01, respectively. To ensure convergence, the analysis was run independently 10 times for 10^6
849 generations, logging every 50th generation and then removing 10% as burn-in. Visualization of
850 the calibrated tree was performed in R using MCMCtreeR package [95].
851
852 **Analyses of introgression**
853 In order to address the scope of possible reticulate evolution across odonate phylogeny, we used
854 various methods of introgression detection such as HyDe/*D* [96], D_{FOIL} [61], χ^2 goodness-of-fit
855 test, QuIBL [63] and PhyloNet [64, 65].
856 The HyDe framework allows detection of hybridization events which relies on
857 quantification of phylogenetic site patterns (invariants). HyDe estimates whether a putative
858 hybrid population (or taxon) H is sister to either population P1 with probability γ or to P2 with
859 probability 1-γ in a 4-taxon (quartet) tree ((P1,H,P2),O), where O denotes an outgroup. Then, it
860 conducts a formal statistical test of H_0: $\gamma = 0$ vs. H_1: $\gamma > 0$ using Z-test, where $\gamma = 0$ (=1) is
861 indicative of non-significant introgression. We applied HyDe to the concatenated supermatrix
862 SuperMatrix_50BUSCO_dna_pasta_ali_trim of 1603 BUSCO genes under default parameters
863 specifying *Ephemera danica* as an outgroup. Under this setup HyDe evaluates all possible taxa
864 quartets. Since HyDe only allows indication of a single outgroup taxon (i.e. *Ephemera danica*),
865 we excluded all quartets that contained the *Isonychia kiangsinensis* outgroup from the HyDe
866 output. Additionally, we calculated Patterson's *D* statistic [54] for every quartet from the

867 frequency (p) of ABBA-BABA site patterns estimated by HyDe as $D = \frac{p_{ABBA} - p_{BABA}}{p_{ABBA} + p_{BABA}}$. To test
868 significance of D statistics we used a χ^2 test to assess whether the proportions p_{ABBA} and p_{BABA}
869 were significantly different. To minimize effect of false positive cases (type I error) in the
870 output, we first applied a Bonferroni correction to the P values derived from Z- and χ^2 tests and
871 then filtered the results based on a significance level of 10^{-6} for both D and γ. Also, we noticed
872 that negative values of D were always associated with nonsignificant γ values, thus all of the D
873 values remained positive after filtering was applied.

874 D_{FOIL} is an alternative site pattern-based approach that detects introgression using
875 symmetric 5-taxon (quintet) trees, i.e. (((P1,P2),(P3,P4)),O). D_{FOIL} represents a collection of
876 similar to D statistics applied to quintet trees and, if considered simultaneously, they provide a
877 powerful approach to identify introgression including ancestral as well as donor and recipient
878 taxa (i.e. introgression directionality). Moreover, D_{FOIL} exhibits exceptionally low false positive
879 rates [61]. Since the number of possible quintet topologies for a phylogeny of 85 taxa is >
880 32×10^6, for analysis we extracted them only for every odonate suborder individually with the
881 custom R scripts. Note, that for Anisozygoptera we only considered quintets that can be formed
882 between Anisozygoptera, Anisoptera, and Zygoptera taxonomic groups. Analogously to HyDe,
883 we applied D_{FOIL} to the concatenated supermatrix SuperMatrix_50BUSCO_dna_pasta_ali_trim
884 of 1603 BUSCO genes under default parameters specifying *Ephemera danica* as an outgroup.
885 Also, since D_{FOIL} requires that every quintet has a symmetric topology we considered only those
886 quintets within our phylogeny that met this criterion (Figure 1). Additionally, D_{FOIL} requires that
887 the divergence time of P3 and P4 precedes divergence of P1 and P2, thus we switched species
888 assignments for those cases that violated this assumption, i.e. from (((P1,P2),(P3,P4)),O) to
889 (((P3,P4),(P1,P2)),O).

890 As an alternative test for introgression we performed a simple yet conservative χ^2
891 goodness-of-fit test on the gene tree count values for each triplet. First, we tested the three count
892 values to see whether their proportions significantly deviate from 1/3, and if the null hypothesis
893 is *not* rejected it will be indicative of extreme levels of ILS. Second, we tested the two count
894 values that originated from discordant gene trees to see whether their proportions significantly
895 deviate from 1/2, and if the null hypothesis is rejected, it will suggest a presence of introgression.
896 In order to correct the P values resulted from the χ^2 test for multiple testing, we applied the
897 Benjamini-Hochberg procedure at a false discovery rate cutoff (FDR) of 0.05.

898 QuIBL is based on the analysis of branch length distributions across gene trees to infer
899 putative introgression patterns. Briefly, under coalescent theory internal branches of rooted gene
900 trees for a set of 3 taxa (triplet) can be viewed as a mixture of two distributions with the
901 underlying parameters. Each mixture component generates branch lengths corresponding to
902 either ILS or introgression/speciation. Thus, estimated mixing proportions (π_1 for ILS and π_2 for
903 introgression/speciation; $\pi_1 + \pi_2 = 1$) of those distribution components show what fraction of the
904 gene trees were generated through ILS or non-ILS processes. For a given triplet, QuIBL
905 computes frequency of gene trees that support three alternative topologies. Then for every
906 alternative topology QuIBL estimates mixing proportions along with other relevant parameters
907 via Expectation-Maximization and computes Bayesian Information Criterion (BIC) scores for
908 ILS-only and introgression models. For concordant topologies elevated values of π_2 are expected
909 whereas for discordant ones π_2 can vary depending on the severity of ILS/intensity of
910 introgression. In extreme cases when the gene trees were generated exclusively under ILS, π_2
911 will approach to zero and the expected gene tree frequency for each alternative topology of a
912 triplet will be approximately 1/3. To identify significant cases of introgression here we used a
913 stringent cutoff of ΔBIC < -30 [63]. We ran QuIBL on every triplet individually under default
914 parameters with number of steps (`numsteps` parameter) is equal to 50 and specifying one of
915 the Ephemeroptera species for triplet rooting. For computational efficiency we extracted triplets
916 only from the odonate superfamilies in a similar manner as we did for D_{FOIL} (see above). For this
917 analysis we used 1603 ML gene trees estimated from CO orthology clusters.
918 In order to estimate phylogenetic networks from the 1603 ML gene trees estimated from
919 CO orthology clusters, we used pseudolikelihood (InferNetwork_MPL) and likelihood
920 (CalGTProb) approaches implemented in PhyloNet. For both analyses we only selected gene
921 trees that had at least one of the outgroup species (*Isonychia kiangsinensis* and *Ephemera*
922 *danica*) and at least three ingroup taxa. We performed network searches for our focal clades,
923 namely Anisozygoptera, Aeshnidae, Gomphidae+Petaluridae and Libellulidae. For
924 pseudolikelihood analysis, we ran PhyloNet on every clade individually allowing a single
925 reticulation event, with the starting tree that corresponds to the species phylogeny (`-s` option),
926 100 iterations (`-x` option), 0.9 bootstrap threshold for gene trees (`-b` option) and optimization of
927 branch lengths and inheritance probabilities on the inferred networks (`-po` option). To ensure
928 convergence, the network searches were repeated 3 times for every clade. For the

929 Anisozygoptera, we explicitly indicated *Epiophlebia superstes* as a putative hybrid (-h option).
930 For the full likelihood estimation, we fixed the topology of a focal clade (equivalent to the
931 species tree topology) and calculated likelihood scores for possible networks with a single
932 reticulation (generated with a custom script) using CalGTProb. Additionally, to asses
933 significance of networks, we used difference of BIC scores (ΔBIC) derived from network
934 without reticulation (i.e. tree) and a network with a reticulation.
935
936 **Dimensionality reduction and visualization**
937 To uncover and visualize complex relationships between site pattern frequencies and Patterson's
938 D statistic and Hyde γ parameter we implemented a dimensionality reduction technique t-
939 distributed stochastic neighbor embedding (tSNE) [60] under default parameters in R.
940 Specifically we estimated tSNE maps from counts of 15 quartet site patterns calculated by HyDe
941 ("AAAA" ,"AAAB" ,"AABA" , "AABB" , "AABC" , "ABAA" , "ABAB", "ABAC" , "ABBA"
942 , "BAAA" ,"ABBC" , "CABC" , "BACA" , "BCAA" , "ABCD").
943
944 **QUANTIFICATION AND STATISTICAL ANALYSES**

945 All statistical analyses were performed in R (https://www.R-project.org/). Statistical significance
946 was tested using the Wilcoxon Rank Sum test, Fisher Exact test and/or χ^2 test.

947
948 **DATA AND SOFTWARE AVAILABILITY**
949
950 All raw RNA-seq read libraries generated for this study have been uploaded to the Short Read
951 Archive (SRA) (SRR12086708-SRR12086765). All assembled transcriptomes, alignments,
952 inferred gene and species trees, fossil calibrated phylogenies (including paleontological record
953 assessed from PaleoDB used to plot the fossil distribution in Figure 1), PhyloNet results, QS
954 results and introgression results are available on figshare
955 (https://doi.org/10.6084/m9.figshare.12518660). Custom scripts that were used to create various
956 input files as well as the analysis pipeline are available on GitHub
957 https://github.com/antonysuv/odo_introgression.