

Mitochondrial genome evolution in the Diprionidae: Major gene rearrangement in the basal Hymenoptera

Min Li^{1,#}, Gengyun Niu^{1,#}, Min Xu¹, Mengxuan Dai², Xinghong Jiang², Yong Ma², Guanliang Meng³ and Meicai Wei^{1,*}

1 Laboratory of Insect Systematics and Evolutionary Biology, College of Life Sciences, Jiangxi Normal University, Nanchang 330022, China

2 School of Computer Information Engineering, Jiangxi Normal University, Nanchang 330022, China

3 Center of Taxonomy and Evolutionary Research, Zoological Research Museum Alexander Koenig, D-53112 Bonn, Germany

* Correspondence: weim@jxnu.edu.cn

Abstract

In comparison to other non-parasitic basal lineages, Apocrita have consistently demonstrated a greatly accelerated rate of gene rearrangement. A number of mechanisms or correlates have been proposed for this observation, such as oxidative stress tolerated by exposure to the host immune system might lead to a high proportion of rearranged mt-genomes. Our studies reveal that gene rearrangements involving the protein-coding gene are present in the basal Hymenoptera lineage based on enriched sampling. We speculate the processes of diversification of rearrangements in the vicinity of *nad2* involved tRNAs and NCRs by producing the chronogram of Diprionids. Furthermore, we investigated the relationship between rearranged genes and their nucleotide sequences. In conclusion, we demonstrate the great potential of gene order and associated sequence features as phylogenetic markers in the study of Hymenoptera evolution, offering a new perspective on studying organisms that undergo frequent gene rearrangements.

Keywords: gene rearrangement, *nad2*, evolutionary pattern, phylogeny, Ka/Ks

1. Introduction

The studies of diminutive mitochondrial genomes, the fledglings of genomics, have led to many 'textbook descriptions' [1]. Rapid advances in sequencing technology have dramatically expanded those findings and deepened or refreshed our understanding of the typical features, leading us to explore the mechanisms or patterns of evolution that are more explanatory. Genomic architecture is probably the most fruitful Genome-level character, with new findings including two co-existing and divergent mt-molecules in single individuals in vertebrate [2] and invertebrates [3], changes in the transcriptional orientations of genes [2], architectural variation of individual genes [4], etc.. Recent meta-analysis [5,6] reinforced more variable rearrangements than anticipated, and identified convergence events. However, as taxonomy experience shows, evolution patterns differ along ranks [7,8]. That means that patterns revealed by meta-analyses of higher taxa, i.e. phyla or orders, can obscure patterns found within order or families. It may thus be possible to obtain new insights into evolution patterns and molecular mechanisms base on dense sampling for representative lineages.

The mt-genome of Hymenoptera is a good model for investigating gene rearrangements [9–12]. It is large enough to accommodate repetitive patterns and complexity, while the basal taxa beyond the massive radiation are small enough to allow for near-complete genus-level sampling in phylogenetic analyses. Recently, a greater than expected diversity of gene rearrangements has been reported in basal Hymenoptera, including inversion, transposition, inverse transposition, shuffling of tRNA clusters, and even inversions of protein-coding genes [13]. This prompts caution regarding the previous conclusion

that gene rearrangements occur mainly in parasitic taxa [14]. Inversions of protein-coding genes have been observed in the mt-genome of Diprionidae, which forms a monophyletic group with Cimbicidae and is located at the base of Tenthredinidae.

The internal relationships of Diprionidae have not been fully investigated, but traditionally they have been divided into two subfamilies. Diprioninae are multi-generational over the year, while Monocteninae have only one generation per year. Molecular phylogenetic inferences [15] support a paraphyletic Monocteninae, with the representative of Augomonoctenus being separated at the most basal position of the family. Diprionidae is unique in that all of its species have reverted to a lifestyle of feeding on gymnosperms [16]. Did retreating to less prosperous hosts subject them to resulting selective pressures that prevented them from developing their diversity, and how did this astonishing reversal occur during evolution?

Here we provide a robust mt-genomic dataset including 33 sampling represent 11 genera or same level for inferring the evolution of the Diprionidae and reconstructing gene rearrangement pattern and ancestral gene states in a phylogenetic framework. The gene rearrangement diversification was analysed from multiple perspectives, including nucleotide composition (polymorphism analysis), Ka/Ks calculated by sliding window and rearrangement frequencies. We aim to describe and quantify the diversity of the gene rearrangements by tracing evolutionary history, to detect sequence footprints generated by gene rearrangements, and discuss their phylogenetic significance. This will provide new insight for the application of gene rearrangements in the study of the evolutionary history.

2. Materials and Methods

2.1. Mitogenome sequencing, assembly and annotation

In total, 33 Diprions were analyzed across two subfamilies and eleven of twelve genera (Table 1). Among them, 13 are newly sequenced and assembled, and 15 are assembled based on raw data from UCEs, transcriptomes, and Hic libraries. Total genomic DNA of a single specimen was used for library preparation with insert size of 250 bp following the manufacturer's instructions, and then 150 bp PE sequenced on an illumina Nova 6000 platform at Berry Genomics Corporation, Beijing, China. The raw reads were quality checked using FastQC.

Mt-genomes in this study were assembled using two parallel strategies: *de novo* genome assembly by applying the compulsory MitoZ [17] and either NovoPlasty [18] or GetOrganelle [19], and mapping the reads to the reference using Geneious Prime v. 2022.2.1. Secondary verification and assessment of the sequencing coverage were performed for gene rearrangement regions, the absence in *de novo* genome assembly, and other aberrant results. The tRNA reported as absent in this study have been assembled using both the homologous gene and the flanking regions of the closely related species as the reference sequence, and return invalid results. Due to the high uncertainty of CR, the potential for multiple repeats within it was not considered here and only the most parsimonious version was recovered. PCG, tRNA and rRNA were hand annotated according to homology to validate automated annotations in Mitos [20]. For additional copies of tRNA, sequence comparisons with the homologous gene from related species were conducted.

Table 1. Summary information of mitogenomes used in phylogenetic analyses

Family	Species	Accession number	References	Voucher number
Diprionidae	<i>Nesodiprion japonicus</i>	ON964464	[13]	CSCS-Hym-MC0010
	<i>Nesodiprion biremis</i>	ON964465	[13]	CSCS-Hym-MC0055

Family	Species	Accession number	References	Voucher number
	<i>Zadiprion rohweri</i>	/	SRX16688665	/
	<i>Zadiprion townsendi</i>	/	SRX16688666	/
	<i>Neodiprion sertifer</i>	MK994526	[10]	/
	<i>Neodiprion virginiana</i>	/	SRX13257720	/
	<i>Neodiprion fabricii</i>	/	SRX13257719	/
	<i>Neodiprion lecontei</i>	/	SRX13257718	/
	<i>Neodiprion pinetum</i>	/	SRX13257615	/
	<i>Microdiprion pallipes</i>	/	SRX16688689	/
	<i>Prionomeion huizeensis</i>	OP874569	This study	CSCS-Hym-MC0761
	<i>Prionomeion qinghaiicus</i>	ON964471	[13]	CSCS-Hym-MC0198
	<i>Diprion similis</i>	/	SRX13257717	/
	<i>Diprion pini</i>	/	SRX642914	/
	<i>Diprion pini</i>	/	SRX16688808	/
	<i>Diprion wenshanicus</i>	OP874571	This study	CSCS-Hym-MC0174
	<i>Diprion infuscalae</i>	OP574194	This study	CSCS-Hym-MC0327
	<i>Diprion nipponicus</i>	/	SRX16688807	/
	<i>Gilpinia liuae</i>	OP574193	This study	CSCS-Hym-MC0173
	<i>Gilpinia liuae</i>	OP874573	This study	CSCS-Hym-MC0762
	<i>Gilpinia oligomaculata</i>	OP874574	This study	CSCS-Hym-MC0779
	<i>Gilpinia xiaowei</i>	OP580614	This study	CSCS-Hym-MC0509
	<i>Gilpinia maomingia</i>	OP874570	This study	CSCS-Hym-MC0389
	<i>Gilpinia xiai</i>	ON840092	[13]	CSCS-Hym-MC0178
	<i>Hugilpinia tabulaeformis</i>	OP874575	This study	CSCS-Hym-MC0513
	<i>Hugilpinia brevictenidia</i>	OP580615	This study	CSCS-Hym-MC0778
	<i>Macrodiprion lishui</i>	OP574195	This study	CSCS-Hym-MC0488
	<i>Macrodiprion lushana</i>	OP874568	This study	CSCS-Hym-MC0099
	<i>Macrodiprion nemoralis</i>	/	SRX16688646	/
	<i>Monoctenus sp.</i>	/	SRX16688695	/
	<i>Monoctenus obscuratus</i>	/	SRX16688694	/
	<i>Monoctenus juniperi</i>	/	SRX16688693	/
	<i>Augomonoctenus smithi</i>	OP874572	This study	CSCS-Hym-MC0844
Cimbicidae	<i>Corynis zhengi</i>	OL549451	[21]	CSCS-Hym-MC0176
Athaliidae	<i>Dentathalia scutellariae</i>	ON808426	[13]	CSCS-Hym-MC0360
Argidae	<i>Arge bella</i>	MF287761	[22]	CSCS-Hym-MC0008
Blasticotomidae	<i>Blasticotoma minuta</i>	ON964461	[13]	CSCS-Hym-MC0060
Xyelida	<i>Xyela sinicola</i>		This study	CSCS-Hym-MC0404
	<i>Megaxyela euchroma</i>	OL794667	[13]	CSCS-Hym-MC0001

2.2. Comparative analysis of mitogenome and visualisation

The sequences were aligned using Geneious with the default parameter, and manual adjustments were made when necessary. Geneious was also used to calculate consensus, identical sites, and pairwise identities.

Using PAML 4.2 [23], we estimated pairwise non-synonymous substitutions per non-synonymous site (dN), synonymous substitutions per synonymous site (dS), and the ratio of these values. For fast base substitution, a KaKs calculator [24] implementation was developed in Python using a sliding window that moves along every pair of sequences. For each window, point marks are used in order to generate a continuous curve. This curve is then combined for all specified samples for comparison.

Protein structures were predicted by SWISS-MODEL [25] and DeepTMHMM [26], using PyMol (www.schrodinger.com/pymol) for editing and visualization.

2.3. Phylogenetic analysis and ancestral states estimation

DAMBE [27] was used to test for severe substitution saturation in the PCG of mtG. The unsaturated aligned sequences of PCG were concatenated using SequenceMatrix v. 1.7.8 [28] to form a single

supermatrix. To improve the accuracy of the phylogenetic analysis, GTR+I+G was selected as the best-fit model using PartitionFinder and subjected to maximum likelihood (ML) analyses and Bayesian inference (BI) analyses using the standard partitioning schemes "BIC" and "AICc", respectively. ML analysis was performed using the IQ-TREE web server [29], based on a probabilistic modeling approach to calculate the relationships between individual species. Default parameters were used except for a perturbation strength of 0.1 and an IQ-TREE stopping rule of 1,000. BI analyses were performed using MrBayes v. 3.2.2 [30] for Markov chain Monte Carlo (MCMC) simulations based on probabilistic and Bayesian statistical theories. A representative phylogenetic tree was derived by constructing a consensus tree.

We also tried some alternative matrixes to deleted the fragmented sample. The results did not strongly affect the inferences (data not shown); and thus, we only use the result from the matrix 1. We chose to map gene arrangement states onto the tree of matrix, which contains the most OTUs (operational taxonomic units), even if those OTUs from the UCE assembly didn't always yield a definitive gene arrangement state. On terminals, gene arrangements were counted and grouped based on consistency. Then the gene arrangement states on each node were inferred based on the maximum parsimony. Species divergence times were estimated using the PAML package MCMCTreeR with three fossil calibration loci, 237~251 Ma, 185~235 Ma and 165~170 Ma, respectively. Evolution times were aliquoted in units of 5 Ma and the gene rearrangement of all nodes in each interval were quantified using qMGR [31].

3. Results

3.1. Architecture of mitogenomes in Diprionidae

In this study, we assembled a total of 27 diprionid mt-genomes, using various sequencing approaches. Specifically, nine mt-genomes were obtained from UCEs, one from the transcriptome, five from Hic libraries, and 13 from whole genome sequencing (WGS). Although UCE sequencing may be an efficient method for obtaining mitogenomes, our results demonstrate a low degree of assembly completeness. Only two of the nine samples retrieved all 37 genes, while others had varying degrees of gaps, with the shortest mt-genome containing only nine PCGs. The Hic libraries yielded complete mitogenomes for three Nesodiprions, but severely deficient ones for the remaining two Diprions. On the other hand, WGS produced more complete mitogenomes, with nine resulting in circular complete mitogenomes based on the minimal CR protocol, and four others yielding all 37 genes. The transcriptome data source yielded a poorer assembly, with 11 incomplete PCGs but no tRNAs. Additionally, we observed heteroplasmy in *Gilpinia maomingia*, with the majority of variation occurring in the PCGs. We retained degenerate bases in those nucleotide positions in order to generalize their variation.

In addition to the previously reported diprionid mt-genomes [10,13], a comparative study of mt-genomes was conducted on 33 diprionid species from eleven genera (Table 1). Overall, diprionid mt-genomes exhibit a high number of non-coding intergenic spacer (IGS) regions, and unlike most other sawflies, the majority of diprionid mt-genomes possess two major non-coding regions (NCR). Multiple gene rearrangements, including a significant reversed *nad2*, have also occurred in numerous species. The length of the Diprionid mt-genome varies considerably, ranging from 15,344 bp (*Gilpinia Xiai*) to 17,991 bp (*Prionomeion qinghaiicus*), primarily due to the presence of multiple IGSs and major NCRs of varying sizes. Several notable features can be observed in these non-coding regions:

- (1) It is common for Diprionid mt-genome to have two major NCRs between 472bp and 1240bp in length, and NCRs at the same position are highly similar within the genus. Two Nesodiprion species, for example, have NCRs between *trnM* and *trnY* that are 454 and 449 bp, respectively.

Their alignment is 456 bp long and contains 389 identical sites. Similarly, the length of the major NCRs in Mt-genomes of *Neodiprion* ranges from 977 to 1002, with 82.3% of the sites being identical within a 1059 bp alignment.

- (2) *Gilpinia* species showed moderate similarity among the major NCRs at the same position, but the two major NCRs themselves were somewhat consistent. The major NCR located between *nad2* and tRNA cluster *MIY* had 529 identical sites relative to the total length of 1162 alignment (49%), whereas the one located between tRNA cluster *MIY* and *cox1* had 891 identical sites relative to the total length of 1358 alignment (65.6%). Yet the similarity between two NCR consensus is 54.2% with 45.3% of the sites being identical.
- (3) Unlike most sawflies mt-genomes, where the major NCR starts from the 5' end of *rrnS*, the Diprionid mt-genome has only short NCRs, with length less than 70 bases, inferred from the secondary structure of *rrnS*, with some exceptions, such as *Neodiprion sertifer* (650 bp) and *Augomonoctenus smithi* Xiao et Wu (472 bp).
- (4) Genes that have undergone rearrangements in diprionids are surrounded by many IGSs. In the genus *Gilpinia*, for instance, the *trnE* gene has been shifted to the third position from the fifth position in the tRNA cluster between *nad3* and *nad5*, resulting in longer IGSs and a loose arrangement of tRNAs.

Diprionids are characterized by their reversed *nad2* apart from the loosely packed mt-genome. To confirm the reversal, three methods were used: 1. by checking the mapping of reads assembled without reference. 2. by extending both ends by referencing the region between *nad2* and *cox1* and the region between *nad2* and *rrnS*. 3. for some samples, by PCR (unpublished). However, due to the unavailability of raw data, validation for *Neodiprion sertifer* could not be carried out. Generally, in Diprioninae, *trnWC* and *trnIM* as the subsets form a common interval [32] with *nad2*, which undergoes initial complete reversal, followed by various translocations and reversals within or between the small subsets, leading to diverse rearrangements between *rrnS* and *cox1*. Apart from the common interval, other gene rearrangements are also clade-specific.

3.2. Diprionid Phylogeny and rare genomic changes as features of higher taxa

Phylogenetic inference of PCGs was performed using 39 genomes, sampling 6 outgroups and 33 Diprionids taxa (Matrix 1, 11622 bp). The initial set was filtered into WGS source data for 27 Diprionidae taxa (Matrix 2, 11607 bp), and only the Diprioninae were retained (Matrix 3, 11403 bp). We utilized the infinite mixture site heterogeneous model (CAT-GTR) for all matrices, while ML analysis and BI analysis were used for the latter two matrices, respectively.

The monophyly of Diprioninae, *Monoctenus*, and *Augomonoctenus* was consistently recovered in all five analyses. *Augomonoctenus* was found to be the basal clade of the family, while *Monoctenus* was the sister group to Diprioninae (bootstrap frequency [BS] > 95%; posterior probability > 0.95). The relationships within Diprioninae varied across topologies, but the monophyly of each genus, and three clades was consistently confirmed: 1. *Nesodiprion* and *Neodiprion*, 2. *Prionomeion* and *Diprion*, 3. *Macrodiprion* (*Hugilpinia* + *Gilpinia*). Clade 1 and 2 were sister groups, while clade 3 was basal, which was supported by all analyses except for the ML inference for matrix 2. With dense sampling in matrix 1, clade 1 was turned to ((*Zadiprion* + *Neodiprion*) *Microdiprion*) *Nesodiprion* (Figure S1). In order to assess the topologies, we refined several key morphological features. Based on their distribution along the topology, we concluded that the Bayesian inference of matrix 3 under the heterogeneity model was the most reliable, which was then further analyzed.

On the phylogenetic tree, gene rearrangement shows step-like changes, resulting in each clade having its

own characteristic pattern (Fig. 1). All Diprionids had their *trnPT* swapped relative to the Hymenoptera ancestral types. After that, *trnAR* was switched in all clades except Augomonocenus, which is the most basal. Once the sub-basal Monoctenus was removed from consideration, a ground pattern of reversal was observed in Diprioninae, with *trnQWC* located upstream of the reversal *nad2*, and *trnIMY* downstream. On this basis, *trnC* translocated to the upstream of *trnQ* in Zadiption in Clade 1. With the exception of incomplete genomes (*Zadiption rohweri* and *Microdiprion pallipes*), the major NCR of all other genomes lies between *trnM* and *trnY*.

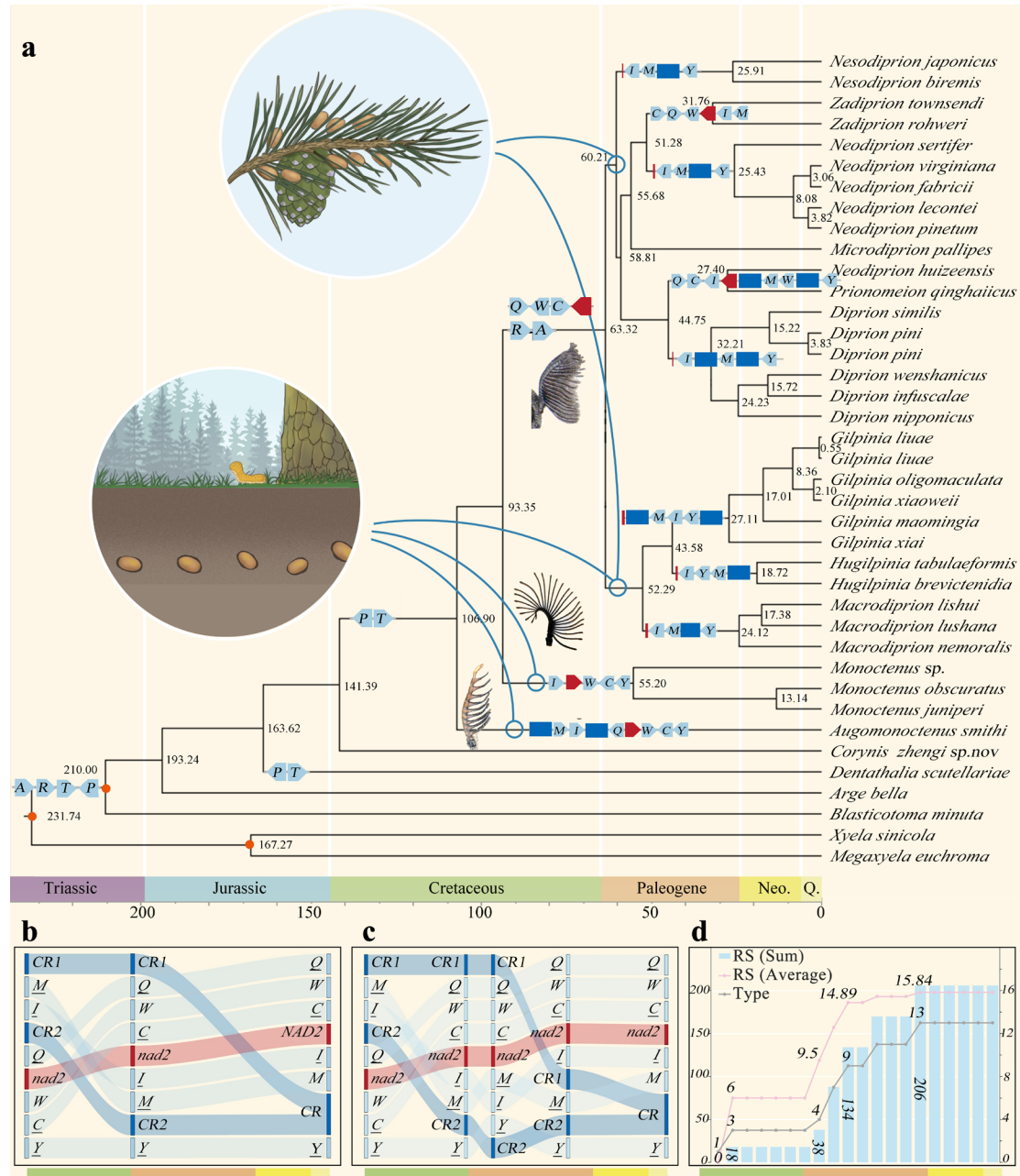


Fig. 1 Timescale of Diprionidae evolution and intensity of gene rearrangement.

- (a) Time-calibrated phylogeny of Diprionidae is based on the Bayesian tree inferred with matrix 1 and divergence times are estimated with MCMCTree. Placements of fossil calibrations are indicated by red spots and gene rearrangement of each clade are showed on the respective nodes; Two different locations of cocoon in Diprionidae are marked on the tree.
- (b) The predicted gene sequences of Nesodiprion and Neodiprion changed over time.
- (c) The predicted gene sequences of Diprion changed over time.
- (d) The observed gene rearrangements in Diprionidae are counted using qMGR based on ancestral gene patterns and counted every five million years from the expected time of first appearance of Diprionidae.

Clade 2 has undergone a striking rearrangement around the *nad2* gene, with *trnW* relocating from the *trnQWC* cluster to the downstream of *nad2*, setting in the middle of *trnM* and *trnY*. Meanwhile, *trnL*, originally flanking *nad2* at the downstream side, has moved upstream and inserted into the middle of *trnC* and *nad2*. In *Prionomeion qinghaiicus*, *trnM* and *trnW* have shifted even further upstream, forming a cluster of five trns positioned upstream of *nad2*. Both *Prionomeion* genomes possess two major NCRs, with the 700bp NCR located upstream of the longer NCR exceeding 1000 bp. These NCRs are located upstream and downstream of *trnI* in *Prionomeion qinghaiicus* and of *trnM-trnW* in *Prionomeion huizeensis*, representitively.

Clade 3 exhibits increased flexibility in the relative positions within *trnIMY* cluster. *Hugilpina* mtgenomes possess only one major NCR located downstream of the *trnIYM*, whereas *Gilpinia* mtgenomes have two major NCRs flanking the *trnMIY*. In addition, the major NCRs of the basal *Macrodipton* are situated within the *trnIMY* cluster.

Aside from the synapomorphic rearrangements around *nad2* observed in each genus, the tRNA cluster between *nad3* and *nad5* is distinctive in *Gilpinia*, with *trnE* relocating from its ancestral position (between *trnS1* and *trnF*) to a derived position (between *trnA* and *trnN*).

3.3. Comparative sequence features in the mitogenomes in *Diprionidae*

Using *Dentathalia scutellariae* as a reference, we calculated Ka/Ks ratios for *Diprionidae*. Across all 13 protein-coding genes (PCGs), the extreme values differed by 0.922, from 1.118 (*atp8*) to 0.195 (*cox1*). We performed the same calculations for *Cimbicidae*, the sister group of *Diprionidae*, and the differences between the genes of the two populations showed approximately the same trend.. (see Figure S2). When calculated pairwise for each species, the values also remained relatively low. Even for *nad2*, which had the high ratio, the comparison ranged from 0.536 (*Prionomeion huizeensis*) to 0.960 (*Diprion nipponicus*), with a mean of 0.754 (Table S1). The Ka/Ks ratios of *atp8*, *nad6*, *nad2* and *nad4l* ranked in the top four. These findings indicate that purifying selection has occurred, both for individual genes and individual species.

However, when we applied a sliding window to perform higher resolution statistics on the sequences, we observed that the sub-regions with Ka/Ks values above 1 emerged (Fig. 2). This observation suggests that although the entire genes followed neutral evolution, specific regions underwent amino acid fixation during evolution, indicating that there is variability in the selection to which different structural and functional protein domains were subjected. Specifically, *nad2* exhibited a clade-specific variations throughout the family, three to five waves appear in the Ka/Ks>1 section, and have a maximum peak at the end of the sequence.

This pattern shows similarity in related taxa and we divided all the species involved in the calculation into five clades. For example, *Diprion* and *Gilpinia* both displaced four waves in similar intervals, while *Monoctenus* and *Neodiprion*+*Neodiprion* have one less in relative terms, located in intervals 20-40 and 87-105, respectively. Similar phenomena were also observed in *nad4l* and *nad6*, where waveform similarities were observed within clades, but variability was seen between clades. This repeatable variability was hardly observed in other PCGs, like *cox1*, which repeats the similar waveform patterns across the entire family.

Considering that the conserved nature of reproducible differences appears in genes with positional correlation, we therefore predicted the structure of the proteins encoded by these genes and also tested for the presence of co-expression between these genes. to identify more precisely whether there is co-evolution in the above sub-regions (Fig. 2). The results show that *nad2* and *nad4l* and *nad6* are closely related and overlap each other, which may indicate that they are functionally and structurally

interdependent and adaptive. In the case of *nad2*, most of the regions with Ka/Ks values higher than 1 are located in the transmembrane region.

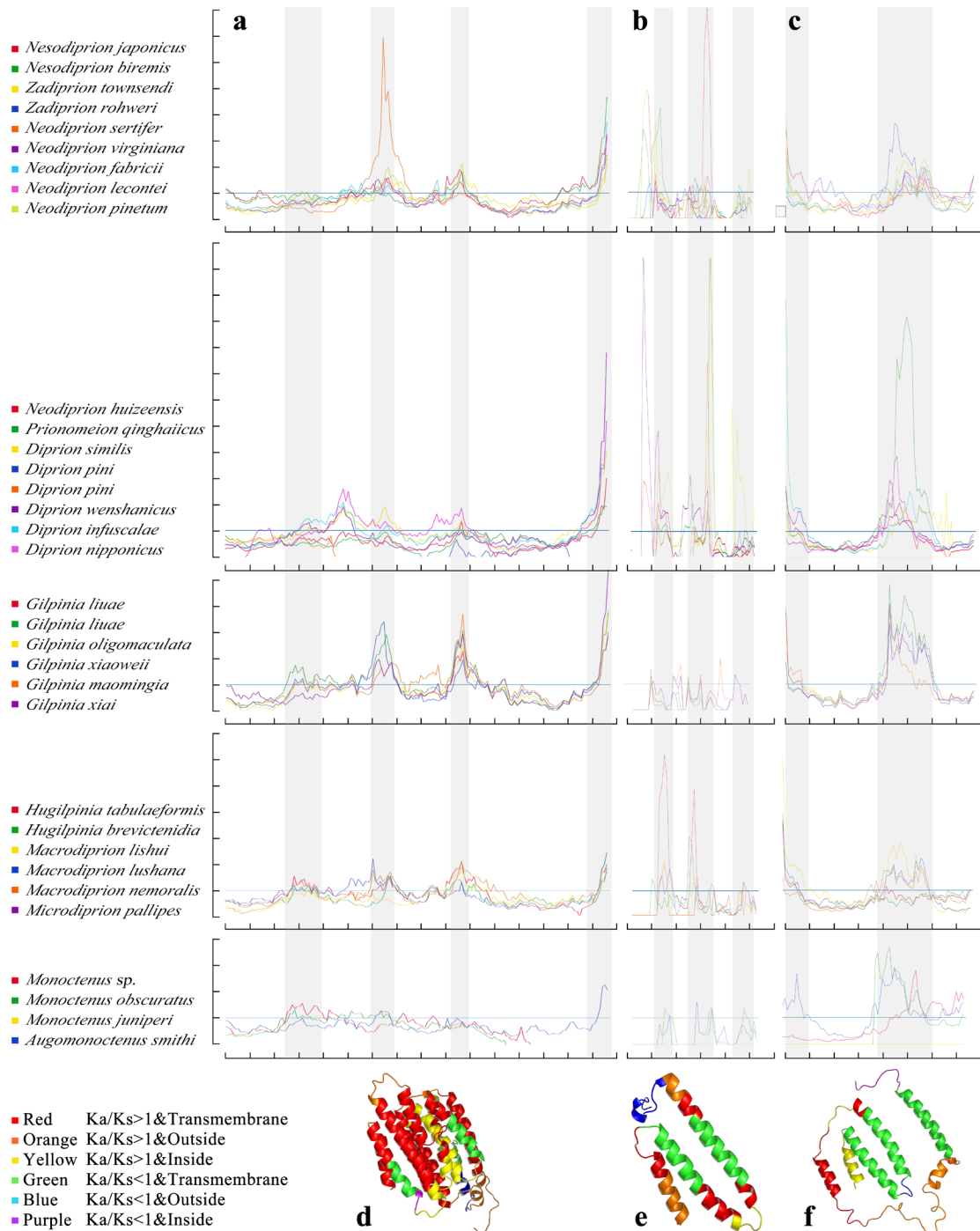


Fig. 2 Sliding window analyses of Ka/Ks and 3D models in Diphronids.

Sliding window analyses of Ka/Ks ratios among (a) *nad2*, (b) *nad4l* and (c) *nad6* of 33 species from Diphronidae, using *Dentathalia scutellariae* as reference genome. The window size used are 132,45 and 90, respectively. Species that displayed similar curves are grouped in a clade. The Y-axis showed the Ka/Ks ratio with each scale representing 1 and the x-axis showed the nucleotide sites at which the alignment sequence begins with each scale representing 10.

Protein coding genes (d) *nad2*, (e) *nad4l* and (f) *nad6* of *Nesodiprion biremis* are predicted by using SWISS-MODEL. The topological structure and Ka/KS values were color-coded by Pymol according to the results of DeepTMHMM.

4. Discussion

4.1. Phylogenetic Framework of Diphronidae

Through proportional sampling of lineages within the clade, we were able to cover 11 of 12 genera or equivalent rank. All analyses produced a consistent framework, indicating that Diprionidae is composed of a diverse, monophyletic Diprioninae (including 8 genera and over 140 species) and sequentially, two poorly diversified basal clades (comprising 3 genera and 16 species). These results are consistent with previous finding based on short fragments of mitochondrial and nuclear genes [15]. As such, the taxonomic system that divides Diprionidae into two subfamilies was rejected, as the subfamily Monocteninae was not found to be monophyletic.

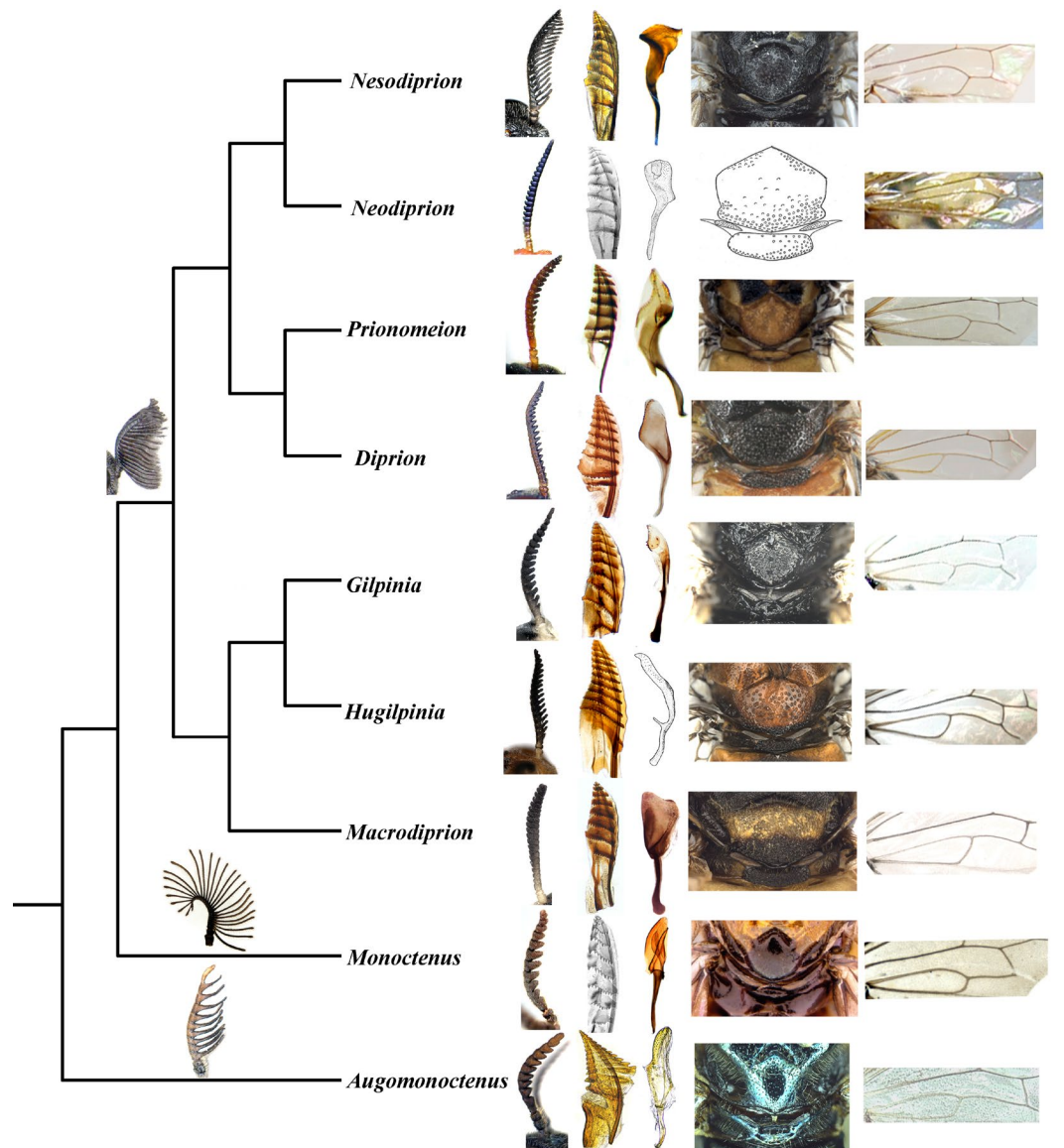


Fig. 3 The morphological characteristics of the Diprionidae: male antennae, penile valve and vein.

The observed pattern, characterized by a sparse genealogy of multiple ancient species at the base and a diversified clade, is supported by a suite of morphological, biological and genomic evidence. Morphologically, each of the three evolutionary clades possesses unique male antenna type (Fig. 3), a key diagnostic feature for higher taxa in the basal Hymenopteran family and subfamilies. Furthermore, three additional plesiomorphic characters are present in the basal two clades are: an ancient pattern of mesoscutellum and metascutellum, a smooth and hardly punctured body surface, and a simple penis valve. Conversely, in Diprioninae, the mesoscutellum and metascutellum are distinctly modified, the body surface is densely and strongly punctate, and the penis valve are specialized. Both basal clades have

typically one generation per year, with *Augomonoctenus smith* requiring up to two years to complete its life history [33], and cocooning 5 to 10 cm below the soil surface. *Augomonoctenus* larvae bore cones of *Cupressus* [34] or *Libocedrus* [35], while *Monoctenus* larvae feed on *Juniperus* needles [36]. The diverse Diprioninae are mostly multi-generational, with *Nesodiprion* even having 5-6 generations per year [37]. Diprioninae larvae are needle-feeders and no longer cocoon in the soil, but instead have shifted to the ground [38], litter [39,40], or even needles [41].

In terms of gene order, the dense taxonomic sampling has filled in the gaps in the otherwise blurred evolutionary history. Table S1 detailed the inferred mtgenomic organizations from nearly complete or complete sequences of 33 Diprionids. However, the UCE and Hic rawdata failed to obtain some sequences, particularly those near *nad2*. While the impact is perhaps minor in genera such as *Diprion* and *Neodiprion*, where no significant difference was found when comparing mtgenomes with the genus. Gene arrangements may yet be underestimated in *Microdiprion* and *Zadiprion*. Moreover, caution should be exercised with genera that have small sample size, such as *Prionomeion*, which displayed gene arrangements divergence in only two samples. At least eleven types of gene arrangement have been observed, with each genus having at least one exclusive rearrangement pattern. The two basal clades retained the basic pattern, while the entire Diprioninae share the derived inversion of *nad2*. Furthermore, additional rearrangements involving tRNA and NCRs led to different arrangement patterns among lineages. Some genera also exhibit characteristics that are distinctive from the general state. The gene rearrangement in the Diprionidae family has several notable features, including: 1. A derived *trnPT* is shared by the family; 2. Two major NCRs are commonly found, including the two basal lineages; 3. Diprioninae share a derived *trnRA*; 4. *trnE* is translocated within the cluster in the *Gilpinia*.

4.2. Pattern and evolution pathway of gene rearrangements in Diprionids

The family Diprionidae provides a good model system for studying gene rearrangements, as it includes: 1. a lineage containing a high frequency of gene rearrangements and other lineages that maintain their ancestral status. 2. intermediate types of progressive rearrangements can be observed within the lineage of high frequency gene rearrangements. 3. rearrangement types are conserved within evolutionary branches. This model enables us to quantify the frequency and intensity of rearrangements over time and infer the basic patterns (Fig. 1).

Our results confirmed previously identified hot spots, such as the most mobile IQM tRNA gene cluster, and the *nad3–nad5* junction [42], but not *trnK* and *trnD* [9]. However, shared gene order within the hot spot may result from different evolutionary pathways. For instance, the *trnI-trnM-NCR-trnY* is shared by *Nesodiprion*, *Neodiprion* and *Macrodipteron*. By tracing the path, we found that if the organization of *Augomonoctenus* is assumed to be the ancestral type of the family, the former two genera underwent only one rearrangement event no later than 63 Ma ago, while the latter underwent at least three changes at varying intervals. It becomes more complex if we consider the NCR. Although NCR are both located between *trnM* and *trnY*, their lengths differ. In *Nesodiprion*, they are approximately 450 bp, whereas in *Neodiprion*, they are around 1000 bp, which is twice as long as in *Nesodiprion*. There appears to be no evidence to suggest that the longer NCR originated from the loss of another ancestral NCR or from the merger of two NCRs.

Observation of gene rearrangements over time reveals that in the Cretaceous, they occurred only at the crown node of the family and then remained silent until the Paleogene, when their frequency increases linearly increased with time. As shown in the Fig. 1, total rearrangement scores (RS) in blue and type statistics in orange showed at least two brief pauses, while average RS scores in purple showed marginal effects soon after the first contemporaneous pause.

The latest rearrangement occurred at 27 Ma and extant species diversity is mainly derived from Neogene and Quaternary from 25 Ma. Therefore, extant species richness not correlated with rearrangements, suggesting that gene rearrangements are a feature of evolutionary branches and do not carry relevant signals for species differentiation within them.

The stepwise scenario prompts a reassessment of convergence [43], previously understood as the convergence of similar rearrangements occurring simultaneously in distant taxa [6] and assigned a phylogenetic significance weighting [44]. We have, however, discovered that rearrangements are frequent [45] and favor hotspots [9], but are random in lineage [46]; if this bias is attributed to molecular mechanisms, the adaptive value of the new organization may depend on specific molecular features, rather than responses to ecological challenges [47]. Unlike the generally mt genomes with tightly arranged gene, Diprionidae has multiple intergenic sequences. The potential homology between IGS and major NCRs suggest that these intergenic regions play a crucial role in regulating the expression of nearby genes [48]. Based on the similarity of major NCR and tRNAs, we can speculate that pseudogenes were once present, and these intervals then shrank over time under deletion pressure to maintain genome compactness [49]. Furthermore, when considering IGS and major NCR, similar arrangements exhibit different degrees of polymorphism. Therefore, assuming those mobile elements underwent neutral evolution and deleting them to reduce noise [6,50] may thus lead to the lose valuable phylogenetic signals or miss relics of the rearrangement event.

Gene rearrangements are rare [51], making them a useful tool for establishing a link between 'adaptive radiation' and its 'key innovations' [52]. Diprionidae, as a model, offers a perspective that underscores the necessity of defining the the unit of evolution beforehand. While gene rearrangements appear to be infrequent in genera, they are widely observed at higher taxonomic ranks. At the family level, the intensity and frequency of gene rearrangements linearly increases over time with non-episodic pauses (Fig. 1). This saltatory (non-clocklike) manner [53] has also been observed in nematodes [54], snails [55], and Salamanders [56], indicating that gene rearrangement is temporally and spatially repeatable. In other words, ecological factors, adaptive landscapes, or evolutionary opportunities may shape similar patterns of gene arrangement.

4.3. Sequence footprints and application of gene arrangements in phylogenetic inference

It was believed that all the variations observed in the mitochondrial genome had to be neutral due to the fatal impact of mutations on functions [57]. As a result, mitochondrial genes were thought to be quantifiable, enabling to estimate the timing of species divergence. In contrast, mitochondrial genes typically perform poorly in genomic analyses aimed at identifying the genetic basis of adaptive changes. Our findings demonstrate that although most PCGs underwent purifying selection, certain regions underwent significant positive selection. Furthermore, the pattern of mutation diverged among different clades (Fig. 1). The sequence footprints left behind during rearrangement events may offer more detailed and reliable evidence for inferring the history of rearrangements and the corresponding mechanisms [58]. Divergence occurs in *nad2*, *nad4l* and *nad6*, suggesting that different genes, or individual loci within genes, underwent various evolutionary pathways, resulting in inconsistent sequence footprints. This phenomenon essentially rejects the homogeneity of similar rearrangement types. It also suggests the possibility that mitochondrial gene rearrangements occur following a series of changes with buffering effects [59]. Rearrangements and footprints may provide an opportunity to understand the nature of selection operating on mitochondria.

It should be noted that all three genes encode subunits of the respiratory chain complex I. However, it is still too early to determine whether there is co-evolution between them.

4.4. Evidence of recombination and rethinking TDRL

Low-resolution taxonomic sampling often leads to an overestimation of the frequency of gene shuffling, which subsequently favors the use of the Tandem Duplication and Random Loss (TDRL) model [60–62] to interpret discontinuous evolution. However, increasing the sample density may provide stepwise evolutionary pathways by filling in gaps. In Diprionidae, rearranged genes can be grouped into several common intervals, while the major non-coding region appears to contain pseudogenes. These observations support the TDRL model, except for the inversion of *trnQ*, which suggests recombination [43,63,64]. Furthermore, multiple hypotheses, such as the Tandem Duplication and Non-Random Loss (TDNL) model [48] and the Dimer-Mitogenome and Non-Random Loss (DMNR) model [49], have been proposed for specific taxa.

Taking a broader perspective of the major eukaryotic groups [65], it is evident that plant mt-genomes, particularly those in vascular plants are frequently rearranged [66–68]. As are those of protists, which also exhibit variability in gene order [69]. Various rearrangements have also been reported in animal mitochondrial genomes, particularly in birds [70], molluscs [1,71] and Insects [45]. Fungi, also are known for their high level of variability [72,73]. And in terms of the mechanism of rearrangement may be explained by the combined effects of non-homologous recombination [74], cumulative duplication [75] (especially in intergenic regions), mobile element dynamics [66], and possibly mitochondrial nuclear gene interactions [76].

Non-adaptive mechanisms may contribute to our understanding of complex gene rearrangement patterns. Although recombination events may be detrimental, only a few of them are adaptive. The discontinuous gene order observed in these events results from both diversification and extinction. Microevolutionary mechanisms may explain diversification, but not extinction, which may explain the lack of direct correlation between rearrangement patterns and biological features. As previously mentioned, a clear correlation was not observed between the number of generations, gene rearrangement, and corresponding sequence feature changes in Diprionidae. Instead, adaptive features such as life history and cocoon location in Diprionidae may have a co-innovative relationship with the genome, rather than a one-way causal relationship [77]. Given that Diprionidae hosts have shifted to gymnosperms, mutual adaptation between them may be explained by a coevolutionary arms race model, where some recombinant mtDNA may have gained replicative advantages. This could lead to their overexpression in early heteroplasmies and an increased probability of fixation as the lineage becomes more homogeneous.

5. Conclusions

This study represents the first comprehensive investigation of the mitochondrial genome across a wide range of Diprionid taxa. The unexpectedly discovery of extensive and high-frequency gene rearrangements in Diprionids provides insights beyond the sequence-based approaches for understanding the organisation, genetics and evolution of animal mitochondrial genomes.

Our investigation highlights the derived *nad2* inversion, clade-specific tRNA rearrangements, and repeatable Ka/Ks variation in complex I, all of which provide opportunities for further exploration of the complex relationship between host transfer phenotypes and molecular evolution in Diprionidae.

By examining the changes in gene order at the generic level within a phylogenetic framework and inferring the frequency and timing of gene rearrangements, we propose that genus-level and species-level have distinct evolutionary patterns, but the effects of radiation and natural selection on gene order remain unclear.

Author Contributions:

Conceptualization, M.W.; methodology, G.N. and G.M.; software, M.D., X.J. and M.L.; validation, G.N. and Y.M.; formal analysis, M.L. and M.X.; investigation, G.N. and M.L.; resources, M.W.; data curation, G.N. and M.L.; writing—original draft, G.N. and M.L.; writing—review and editing, M.W.; visualization, M.L.; supervision, M.W.; project administration, M.W.; funding acquisition, G.N. and M.W. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Natural Science Foundation of China (31501885, 31970447) and the Jiangxi Provincial Natural Science Foundation (20202BABL213044).

Data Availability Statement: The data presented in this study are openly available in Figshare at <https://figshare.com/account/home#/projects/149867>

Acknowledgments: The members of the Lab of Insect Systematics and Evolutionary Biology (LISEB) at Jiangxi Normal University are thanked for their contributions to laboratory work. The authors thank Wang Jin for the assistance with figure editing. We thank the anonymous reviewers for their careful reading and many constructive comments.

Conflicts of Interest: The authors declare no conflict of interest.

Reference

- [1] F. Ghiselli, A. Gomes-dos-Santos, C.M. Adema, M. Lopes-Lima, J. Sharbrough, J.L. Boore, Molluscan mitochondrial genomes break the rules, *Philosophical Transactions of the Royal Society B: Biological Sciences*. 376 (2021) 20200159. <https://doi.org/10.1098/rstb.2020.0159>.
- [2] J.R. Macey, S. Pabinger, C.G. Barbieri, E.S. Buring, V.L. Gonzalez, D.G. Mulcahy, D.P. DeMeo, L. Urban, P.M. Hime, S. Prost, A.N. Elliott, N.J. Gemmill, Evidence of two deeply divergent co-existing mitochondrial genomes in the Tuatara reveals an extremely complex genomic organization, *Commun Biol*. 4 (2021) 1–10. <https://doi.org/10.1038/s42003-020-01639-0>.
- [3] L. Schärer, Evolution: Mitochondrial lodgers can take over in hermaphroditic snails, *Current Biology*. 32 (2022) R477–R479. <https://doi.org/10.1016/j.cub.2022.04.039>.
- [4] C.A. Milbury, J.C. Lee, J.J. Cannone, P.M. Gaffney, R.R. Gutell, Fragmentation of the large subunit ribosomal RNA gene in oyster mitochondrial genomes, *BMC Genomics*. 11 (2010) 485. <https://doi.org/10.1186/1471-2164-11-485>.
- [5] P. Montaña-Lozano, M. Moreno-Carmona, M. Ochoa-Capera, N.S. Medina, J.L. Boore, C.F. Prada, Comparative genomic analysis of vertebrate mitochondrial reveals a differential of rearrangements rate between taxonomic class, *Sci Rep*. 12 (2022) 5479. <https://doi.org/10.1038/s41598-022-09512-2>.
- [6] M. Moreno-Carmona, S.L. Cameron, C.F. Prada Quiroga, How are the mitochondrial genomes reorganized in Hexapoda? Differential evolution and the first report of convergences within Hexapoda, *Gene*. 791 (2021) 145719. <https://doi.org/10.1016/j.gene.2021.145719>.
- [7] T.G. Barraclough, A.M. Humphreys, The evolutionary reality of species and higher taxa in plants: a survey of post-modern opinion and evidence, *New Phytol*. 207 (2015) 291–296. <https://doi.org/10.1111/nph.13232>.
- [8] A.M. Humphreys, T.G. Barraclough, The evolutionary reality of higher taxa in mammals, *Proc. R. Soc. B*. 281 (2014) 20132750. <https://doi.org/10.1098/rspb.2013.2750>.
- [9] M. Downton, A.D. Austin, Evolutionary dynamics of a mitochondrial rearrangement “hot spot” in the Hymenoptera., *Molecular Biology and Evolution*. 16 (1999) 298–309.

- <https://doi.org/10.1093/oxfordjournals.molbev.a026111>.
- [10] M.N. Aydemir, E.M. Korkmaz, Comparative mitogenomics of Hymenoptera reveals evolutionary differences in structure and composition, *International Journal of Biological Macromolecules*. 144 (2020) 460–472. <https://doi.org/10.1016/j.ijbiomac.2019.12.135>.
- [11] M. Mao, T. Gibson, M. Downton, Evolutionary Dynamics of the Mitochondrial Genome in the Evaniomorpha (Hymenoptera)—A Group with an Intermediate Rate of Gene Rearrangement, *Genome Biology and Evolution*. 6 (2014) 1862–1874. <https://doi.org/10.1093/gbe/evu145>.
- [12] B. Zheng, Y. Han, R. Yuan, J. Liu, C. van Achterberg, P. Tang, X. Chen, Comparative Mitochondrial Genomics of 104 Darwin Wasps (Hymenoptera: Ichneumonidae) and Its Implication for Phylogeny, *Insects*. 13 (2022) 124. <https://doi.org/10.3390/insects13020124>.
- [13] G. Niu, M. Budak, E. Korkmaz, Ö. Doğan, A. Nel, S. Wan, C. Cai, C. Jouault, L. Min, M. Wei, Phylogenomic Analyses of the Tenthredinoidea Support the Familial Rank of Athaliidae (Insecta, Tenthredinoidea), *Insects*. 13 (2022) 1–25. <https://doi.org/10.3390/insects13100858>.
- [14] P. Tang, J. Zhu, B. Zheng, S. Wei, M. Sharkey, X. Chen, A.P. Vogler, Mitochondrial phylogenomics of the Hymenoptera, *Molecular Phylogenetics and Evolution*. 131 (2019) 8–18. <https://doi.org/10.1016/j.ympev.2018.10.040>.
- [15] T. Malm, T. Nyman, Phylogeny of the symphytan grade of Hymenoptera: new pieces into the old jigsaw(fly) puzzle, *Cladistics*. 31 (2015) 1–17. <https://doi.org/10.1111/cla.12069>.
- [16] T. Nyman, R.E. Onstein, D. Silvestro, S. Wutke, A. Taeger, N. Wahlberg, S.M. Blank, T. Malm, The early wasp plucks the flower: disparate extant diversity of sawfly superfamilies (Hymenoptera: ‘Symphyta’) may reflect asynchronous switching to angiosperm hosts, *Biological Journal of the Linnean Society*. 128 (2019) 1–19. <https://doi.org/10.1093/biolinnean/blz071>.
- [17] G. Meng, Y. Li, C. Yang, S. Liu, MitoZ: a toolkit for animal mitochondrial genome assembly, annotation and visualization, *Nucleic Acids Res.* 47 (2019) e63. <https://doi.org/10.1093/nar/gkz173>.
- [18] N. Dierckxsens, P. Mardulyn, G. Smits, NOVOPlasty: de novo assembly of organelle genomes from whole genome data, *Nucleic Acids Res.* 45 (2017) e18. <https://doi.org/10.1093/nar/gkw955>.
- [19] J.-J. Jin, W.-B. Yu, J.-B. Yang, Y. Song, C.W. dePamphilis, T.-S. Yi, D.-Z. Li, GetOrganelle: a fast and versatile toolkit for accurate de novo assembly of organelle genomes, *Genome Biol.* 21 (2020) 241. <https://doi.org/10.1186/s13059-020-02154-5>.
- [20] M. Bernt, A. Donath, F. Jühling, F. Externbrink, C. Florentz, G. Fritsch, J. Pütz, M. Middendorf, P.F. Stadler, MITOS: Improved de novo metazoan mitochondrial genome annotation, *Molecular Phylogenetics and Evolution*. 69 (2013) 313–319. <https://doi.org/10.1016/j.ympev.2012.08.023>.
- [21] G. Niu, Y. Cheng, C. Cai, G. Meng, Y. Yan, Z. Sun, S. Wan, M. Wei, Phylogenomics illuminates the divergence timeline and the dynamic mitogenome of Cimbicidae, (n.d.).
- [22] S. Du, G. Niu, T. Nyman, M. Wei, Characterization of the mitochondrial genome of *Arge bella* Wei & Du sp. nov. (Hymenoptera: Argidae), *PeerJ*. 6 (2018) e6131. <https://doi.org/10.7717/peerj.6131>.
- [23] Z. Yang, PAML 4: Phylogenetic Analysis by Maximum Likelihood, *Molecular Biology and Evolution*. 24 (2007) 1586–1591. <https://doi.org/10.1093/molbev/msm088>.
- [24] Z. Zhang, J. Li, X.-Q. Zhao, J. Wang, G.K.-S. Wong, J. Yu, KaKs_Calculator: Calculating Ka and Ks Through Model Selection and Model Averaging, *Genomics, Proteomics & Bioinformatics*. 4 (2006) 259–263. [https://doi.org/10.1016/S1672-0229\(07\)60007-2](https://doi.org/10.1016/S1672-0229(07)60007-2).
- [25] A. Waterhouse, M. Bertoni, S. Bienert, G. Studer, G. Tauriello, R. Gumienny, F.T. Heer, T.A.P.

- de Beer, C. Rempfer, L. Bordoli, R. Lepore, T. Schwede, SWISS-MODEL: homology modelling of protein structures and complexes, *Nucleic Acids Research*. 46 (2018) W296–W303. <https://doi.org/10.1093/nar/gky427>.
- [26] J. Hallgren, K. Tsirigos, M.D. Pedersen, J.J. Almagro Armenteros, P. Marcatili, H. Nielsen, A. Krogh, O. Winther, DeepTMHMM predicts alpha and beta transmembrane proteins using deep neural network, (2022). <https://doi.org/10.1101/2022.04.08.487609>.
- [27] X. Xia, DAMBE6: New Tools for Microbial Genomics, Phylogenetics, and Molecular Evolution, *Journal of Heredity*. 108 (2017) 431–437. <https://doi.org/10.1093/jhered/esx033>.
- [28] G. Vaidya, D.J. Lohman, R. Meier, SequenceMatrix: concatenation software for the fast assembly of multi-gene datasets with character set and codon information, *Cladistics*. 27 (2011) 171–180. <https://doi.org/10.1111/j.1096-0031.2010.00329.x>.
- [29] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: A Fast and Effective Stochastic Algorithm for Estimating Maximum-Likelihood Phylogenies, *Molecular Biology and Evolution*. 32 (2015) 268–274. <https://doi.org/10.1093/molbev/msu300>.
- [30] F. Ronquist, M. Teslenko, P. van der Mark, D.L. Ayres, A. Darling, S. Höhna, B. Larget, L. Liu, M.A. Suchard, J.P. Huelsenbeck, MrBayes 3.2: Efficient Bayesian Phylogenetic Inference and Model Choice Across a Large Model Space, *Systematic Biology*. 61 (2012) 539–542. <https://doi.org/10.1093/sysbio/sys029>.
- [31] J. Zhang, X. Kan, G. Miao, S. Hu, Q. Sun, W. Tian, qMGR: A new approach for quantifying mitochondrial genome rearrangement, *Mitochondrion*. 52 (2020) 20–23. <https://doi.org/10.1016/j.mito.2020.02.004>.
- [32] M. Bernt, D. Merkle, K. Ramsch, G. Fritzsche, M. Perseke, D. Bernhard, M. Schlegel, P.F. Stadler, M. Middendorf, CREx: inferring genomic rearrangements based on common intervals, *Bioinformatics*. 23 (2007) 2957–2958. <https://doi.org/10.1093/bioinformatics/btm468>.
- [33] Z. Zhang, S. Zhou, Chinese Diprionid Sawflies I . Distribution, host species, damage and life cycle, *Journal of Beijing Forestry University (English Ed.)*. 5 (1996) 83–89.
- [34] C. Zeng, M. Jin, M. Liu, X. Su, On the bionomics and control of *Augomonoctenus Smith* Xiao et Wu (Hymenoptera, Diprionidae), *Scientia Silvae Sinicae*. 20 (1984) 332–335. <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFD7984&filename=LYKE198403013&v=>.
- [35] S.A. Rohwer, New sawflies of the subfamily Diprioninae (Hym.), *Proceedings of the Entomological Society of Washington*. 20 (1918) 79–90.
- [36] D.R. Smith, G. Sánchez-Martínez, S. Ordaz-Silva, A New *Monoctenus* (Hymenoptera: Diprionidae) Damaging *Juniperus Flaccida* (Cupressaceae) in San Luis Potosí, Mexico, *Proceedings of the Entomological Society of Washington*. 112 (2010) 444–450. <https://doi.org/10.4289/0013-8797.112.3.444>.
- [37] R.A. Beaver, D. Laosunthorn, The biology and control of the pine sawfly, *Nesodiprion biremis* (Konow) (Hymenoptera, Diprionidae), in northern Thailand, *Bull. Entomol. Res.* 65 (1975) 117–128. <https://doi.org/10.1017/S0007485300005824>.
- [38] Y. Zheng, X. Cui, H. Chen, Q. Zhou, Preliminary Study on Biological Characters of *Gilpinia lipuensis*, *Forest Research*. 16 (2003) 651–653. <https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2003&filename=LYKX200305021&v=>.
- [39] H. Ying, Z. Luo, Biological characteristics and control of *Neodiprion Huizeensis*, Yunnan

- Forestry Science and Technology. (2002) 65–67.
<https://doi.org/10.16473/j.cnki.xblykx1972.2002.01.014>.
- [40] Y. Zhang, W. Ye, Y. Li, Bionomics of the Pine Sawfly *Diprion Jingyuanensis*, Journal of Jishou University (Natural Sciences Edition). 24 (2003) 55–61.
<https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFD2003&filename=JS DN200303014&v=>.
- [41] W. Fang, A preliminary study on the biology and control of Jiangchuan pine sawfly (*Microdiprion sp.*), Journal of Southwest Forestry College. 11 (1991) 79–82.
<https://kns.cnki.net/KCMS/detail/detail.aspx?dbcode=CJFD&dbname=CJFD9093&filename=Y NLX199101011&v=>.
- [42] M. Downton, L.R. Castro, S.L. Campbell, S.D. Bargon, A.D. Austin, Frequent mitochondrial gene rearrangements at the hymenopteran nad3-nad5 junction, J Mol Evol. 56 (2003) 517–526.
<https://doi.org/10.1007/s00239-002-2420-3>.
- [43] A. Kurabayashi, M. Sumida, H. Yonekawa, F. Glaw, M. Vences, M. Hasegawa, Phylogeny, Recombination, and Mechanisms of Stepwise Mitochondrial Genome Reorganization in Mantellid Frogs from Madagascar, Molecular Biology and Evolution. 25 (2008) 874–891.
<https://doi.org/10.1093/molbev/msn031>.
- [44] D. Zhang, W.X. Li, H. Zou, S.G. Wu, M. Li, I. Jakovlić, J. Zhang, R. Chen, G. Wang, Homoplasy or plesiomorphy? Reconstruction of the evolutionary history of mitochondrial gene order rearrangements in the subphylum Neodermata, Int J Parasitol. 49 (2019) 819–829.
<https://doi.org/10.1016/j.ijpara.2019.05.010>.
- [45] S.L. Cameron, Insect Mitochondrial Genomics: Implications for Evolution and Phylogeny, Annu. Rev. Entomol. 59 (2014) 95–117. <https://doi.org/10.1146/annurev-ento-011613-162007>.
- [46] M. Babbucci, A. Basso, A. Scupola, T. Patarnello, E. Negrisol, Is It an Ant or a Butterfly? Convergent Evolution in the Mitochondrial Gene Order of Hymenoptera and Lepidoptera, Genome Biology and Evolution. 6 (2014) 3326–3343. <https://doi.org/10.1093/gbe/evu265>.
- [47] D.L. Stern, The genetic causes of convergent evolution, Nat Rev Genet. 14 (2013) 751–764.
<https://doi.org/10.1038/nrg3483>.
- [48] 3 Characterization of intergenic regions and gene definition, Nature. (2019) nature28172.
<https://doi.org/10.1038/nature28172>.
- [49] R.P. Sugino, H. Innan, Natural Selection on Gene Order in the Genome Reorganization Process After Whole-Genome Duplication of Yeast, Molecular Biology and Evolution. 29 (2012) 71–79.
<https://doi.org/10.1093/molbev/msr118>.
- [50] K. Tyagi, R. Chakraborty, S.L. Cameron, A.D. Sweet, K. Chandra, V. Kumar, Rearrangement and evolution of mitochondrial genomes in Thysanoptera (Insecta), Sci Rep. 10 (2020) 1–16.
<https://doi.org/10.1038/s41598-020-57705-4>.
- [51] A. Rokas, P.W.H. Holland, Rare genomic changes as a tool for phylogenetics, Trends in Ecology & Evolution. 15 (2000) 454–459. [https://doi.org/10.1016/S0169-5347\(00\)01967-4](https://doi.org/10.1016/S0169-5347(00)01967-4).
- [52] H. Li, J.M. Leavengood, E.G. Chapman, D. Burkhardt, F. Song, P. Jiang, J. Liu, X. Zhou, W. Cai, Mitochondrial phylogenomics of Hemiptera reveals adaptive innovations driving the diversification of true bugs, Proceedings of the Royal Society B: Biological Sciences. 284 (2017) 20171223. <https://doi.org/10.1098/rspb.2017.1223>.
- [53] J.L. Boore, S.I. Fuerstenberg, Beyond linear sequence comparisons: the use of genome-level characters for phylogenetic reconstruction, Phil. Trans. R. Soc. B. 363 (2008) 1445–1451.

- <https://doi.org/10.1098/rstb.2007.2234>.
- [54] H. Zou, I. Jakovlić, R. Chen, D. Zhang, J. Zhang, W.-X. Li, G.-T. Wang, The complete mitochondrial genome of parasitic nematode *Camallanus cotti*: extreme discontinuity in the rate of mitogenomic architecture evolution within the Chromadorea class, *BMC Genomics*. 18 (2017) 840. <https://doi.org/10.1186/s12864-017-4237-x>.
- [55] J.-G. Wang, D. Zhang, I. Jakovlić, W.-M. Wang, Sequencing of the complete mitochondrial genomes of eight freshwater snail species exposes pervasive paralogy within the Viviparidae family (Caenogastropoda), *PLOS ONE*. 12 (2017) e0181699. <https://doi.org/10.1371/journal.pone.0181699>.
- [56] R.L. Mueller, J.L. Boore, Molecular Mechanisms of Extensive Mitochondrial Gene Rearrangement in Plethodontid Salamanders, *Molecular Biology and Evolution*. 22 (2005) 2104–2112. <https://doi.org/10.1093/molbev/msi204>.
- [57] M. Ingman, H. Kaessmann, S. Pääbo, U. Gyllensten, Mitochondrial genome variation and the origin of modern humans, *Nature*. 408 (2000) 708–713. <https://doi.org/10.1038/35047064>.
- [58] M. Al Arab, M. Bernt, C. Höner zu Siederdisen, K. Tout, P.F. Stadler, Partially local three-way alignments and the sequence signatures of mitochondrial genome rearrangements, *Algorithms for Molecular Biology*. 12 (2017) 22. <https://doi.org/10.1186/s13015-017-0113-0>.
- [59] V. Nagarajan-Radha, I. Aitkenhead, D.J. Clancy, S.L. Chown, D.K. Dowling, Sex-specific effects of mitochondrial haplotype on metabolic rate in *Drosophila melanogaster* support predictions of the Mother’s Curse hypothesis, *Philosophical Transactions of the Royal Society B*. (2020). <https://doi.org/10.1098/rstb.2019.0178>.
- [60] J.L. Boore, The Duplication/Random Loss Model for Gene Rearrangement Exemplified by Mitochondrial Genomes of Deuterostome Animals, in: D. Sankoff, J.H. Nadeau (Eds.), *Comparative Genomics*, Springer Netherlands, Dordrecht, 2000: pp. 133–147. https://doi.org/10.1007/978-94-011-4309-7_13.
- [61] T. Hartmann, M. Bernt, M. Middendorf, EqualTDRL: illustrating equivalent tandem duplication random loss rearrangements, *BMC Bioinformatics*. 19 (2018) 192. <https://doi.org/10.1186/s12859-018-2170-x>.
- [62] D. San Mauro, D.J. Gower, R. Zardoya, M. Wilkinson, A hotspot of gene order rearrangement by tandem duplication and random loss in the vertebrate mitochondrial genome, *Mol Biol Evol*. 23 (2006) 227–234. <https://doi.org/10.1093/molbev/msj025>.
- [63] A. Rokas, E. Ladoukakis, E. Zouros, Animal mitochondrial DNA recombination revisited, *Trends in Ecology & Evolution*. 18 (2003) 411–417. [https://doi.org/10.1016/S0169-5347\(03\)00125-3](https://doi.org/10.1016/S0169-5347(03)00125-3).
- [64] M. Mao, A.D. Austin, N.F. Johnson, M. Dowton, Coexistence of Minicircular and a Highly Rearranged mtDNA Molecule Suggests That Recombination Shapes Mitochondrial Genome Organization, *Molecular Biology and Evolution*. 31 (2014) 636–644. <https://doi.org/10.1093/molbev/mst255>.
- [65] C.E. Bullerwell, M.W. Gray, Evolution of the mitochondrial genome: protist connections to animals, fungi and plants, *Curr Opin Microbiol*. 7 (2004) 528–534. <https://doi.org/10.1016/j.mib.2004.08.008>.
- [66] J.D. Palmer, K.L. Adams, Y. Cho, C.L. Parkinson, Y.L. Qiu, K. Song, Dynamic evolution of plant mitochondrial genomes: mobile genes and introns and highly variable mutation rates, *Proc Natl Acad Sci U S A*. 97 (2000) 6960–6966. <https://doi.org/10.1073/pnas.97.13.6960>.
- [67] N. Galtier, The intriguing evolutionary dynamics of plant mitochondrial DNA, *BMC Biol*. 9 (2011)

61. <https://doi.org/10.1186/1741-7007-9-61>.
- [68] J. Apitz, A. Weihe, F. Pohlheim, T. Börner, Biparental inheritance of organelles in *Pelargonium*: evidence for intergenomic recombination of mitochondrial DNA, *Planta*. 237 (2013) 509–515. <https://doi.org/10.1007/s00425-012-1768-x>.
- [69] M.W. Gray, B.F. Lang, R. Cedergren, G.B. Golding, C. Lemieux, D. Sankoff, M. Turmel, N. Brossard, E. Delage, T.G. Littlejohn, I. Plante, P. Rioux, D. Saint-Louis, Y. Zhu, G. Burger, Genome structure and gene content in protist mitochondrial DNAs, *Nucleic Acids Res.* 26 (1998) 865–878. <https://doi.org/10.1093/nar/26.4.865>.
- [70] D.P. Mindell, M.D. Sorenson, D.E. Dimcheff, Multiple independent origins of mitochondrial gene order in birds, *Proc Natl Acad Sci U S A.* 95 (1998) 10693–10697. <https://doi.org/10.1073/pnas.95.18.10693>.
- [71] B. Zhao, S. Gao, M. Zhao, H. Lv, J. Song, H. Wang, Q. Zeng, J. Liu, Mitochondrial genomic analyses provide new insights into the “missing” *atp8* and adaptive evolution of Mytilidae, *BMC Genomics*. 23 (2022) 738. <https://doi.org/10.1186/s12864-022-08940-8>.
- [72] B. Paquin, M.J. Laforest, L. Forget, I. Roewer, Z. Wang, J. Longcore, B.F. Lang, The fungal mitochondrial genome project: evolution of fungal mitochondrial genomes and their gene expression, *Curr Genet.* 31 (1997) 380–395. <https://doi.org/10.1007/s002940050220>.
- [73] G. Aguileta, D.M. de Vienne, O.N. Ross, M.E. Hood, T. Giraud, E. Petit, T. Gabaldón, High Variability of Mitochondrial Gene Order among Fungi, *Genome Biol Evol.* 6 (2014) 451–465. <https://doi.org/10.1093/gbe/evu028>.
- [74] S. Portugez, W.F. Martin, E. Hazkani-Covo, Mosaic mitochondrial-plastid insertions into the nuclear genome show evidence of both non-homologous end joining and homologous recombination, *BMC Evol Biol.* 18 (2018) 162. <https://doi.org/10.1186/s12862-018-1279-x>.
- [75] S. Song, F. Jiang, J. Yuan, W. Guo, Y. Miao, Exceptionally high cumulative percentage of NUMTs originating from linear mitochondrial DNA molecules in the *Hydra magnipapillata* genome, *BMC Genomics*. 14 (2013) 447. <https://doi.org/10.1186/1471-2164-14-447>.
- [76] Y.-F. Yang, W. Cao, S. Wu, W. Qian, Genetic Interaction Network as an Important Determinant of Gene Order in Genome Evolution, *Mol Biol Evol.* 34 (2017) 3254–3266. <https://doi.org/10.1093/molbev/msx264>.
- [77] M.J. Donoghue, M.J. Sanderson, Confluence, synnovation, and depauperons in plant diversification, *New Phytologist*. 207 (2015) 260–274. <https://doi.org/10.1111/nph.13367>.