

RESEARCH

Open Access



Novel insight into lepidopteran phylogenetics from the mitochondrial genome of the apple fruit moth of the family *Argyresthiidae*

Abdelhameed Elameen^{1*}, Simo N. Maduna², Melissa H. Mageroy¹, André van Eerde¹, Geir Knudsen¹, Snorre B. Hagen² and Hans Geir Eiken²

Abstract

Background The order Lepidoptera has an abundance of species, including both agriculturally beneficial and detrimental insects. Molecular data has been used to investigate the phylogenetic relationships of major subdivisions in Lepidoptera, which has enhanced our understanding of the evolutionary relationships at the family and superfamily levels. However, the phylogenetic placement of many superfamilies and/or families in this order is still unknown. In this study, we determine the systematic status of the family *Argyresthiidae* within Lepidoptera and explore its phylogenetic affinities and implications for the evolution of the order. We describe the first mitochondrial (mt) genome from a member of *Argyresthiidae*, the apple fruit moth *Argyresthia conjugella*. The insect is an important pest on apples in Fennoscandia, as it switches hosts when the main host fails to produce crops.

Results The mt genome of *A. conjugella* contains 16,044 bp and encodes all 37 genes commonly found in insect mt genomes, including 13 protein-coding genes (PCGs), two ribosomal RNAs, 22 transfer RNAs, and a large control region (1101 bp). The nucleotide composition was extremely AT-rich (82%). All detected PCGs (13) began with an ATN codon and terminated with a TAA stop codon, except the start codon in *cox1* is ATT. All 22 tRNAs had cloverleaf secondary structures, except *trnS1*, where one of the dihydrouridine (DHU) arms is missing, reflecting potential differences in gene expression. When compared to the mt genomes of 507 other Lepidoptera representing 18 superfamilies and 42 families, phylogenomic analyses found that *A. conjugella* had the closest relationship with the Plutellidae family (Yponomeutoidea-super family). We also detected a sister relationship between Yponomeutoidea and the superfamily Tineidae.

Conclusions Our results underline the potential importance of mt genomes in comparative genomic analyses of Lepidoptera species and provide valuable evolutionary insight across the tree of Lepidoptera species.

Keywords *Argyresthia conjugella*, Illumina HiSeq, Lepidoptera, Mitochondrial genome, Yponomeutoidea

*Correspondence:

Abdelhameed Elameen
abdelhameed.elameen@nibio.no

¹ Division of Biotechnology and Plant Health, NIBIO, Norwegian Institute of Bioeconomy Research, Høgskoleveien 7, N-1431 Aas, Norway

² Division of Environment and Natural Resources, NIBIO, Norwegian Institute of Bioeconomy Research, Høgskoleveien 7, N-1431 Aas, Norway

Background

Lepidoptera is the second largest insect order, with > 160,000 species [1]. This order includes both butterflies and moths, many of which are important model organisms in ecology and evolutionary biology [2]. In Lepidoptera, mitochondrial (mt) genomes are widely



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

used to study population genetics, phylogeography, phylogenetics and molecular taxonomy [3–5]. In particular, the mitogenome represents an ideal tool for the analysis of phylogenetic relationships due to its simple structure, maternal inheritance, low recombination, and high conservation over the course of evolution [6, 7]. Mitogenomes may also provide information to identify novel genes that may serve as targets in future research [8]. The mitogenome size of Lepidoptera ranges from 15,000 bp to above 16,000 bp [6, 7], mostly due to the variable length of noncoding regions, particularly the control region [8]. Moreover, lepidopteran mt genomes have a conserved rich adenine and thymine (A+T) region and usually consist of 37 genes, encoding 13 conserved protein-coding genes (PCGs), 22 tRNAs, 2 rRNAs, and a noncoding control region [6, 7, 9]. Until recently, the bulk of Argyresthiidae phylogenetic analyses utilized a common set of 8–11 mitochondrial and nuclear genes [10] and a set of up to 27 protein-coding genes [11–16]. However, inadequate node support hindered research that attempted to unravel relationships among superfamilies, even with over 1500 genes [17–20]. The potential causes and consequences of the competing phylogenetic hypotheses were discovered to be compositional bias and other model violations [21].

Yponomeutoidea is a large superfamily of Lepidoptera with 11 families and 1800 species [16, 22]. Surprisingly, only seven mt genome species of this superfamily are available in databases [23]. For the Argyresthiidae family, which belongs to this superfamily and contains 157 species [16], no mt genomes exist. Thus, obtaining the mt genomes of this family may be warranted for further resolving patterns of genomic evolution and assessing phylogenetic relationships.

The apple fruit moth (*Argyresthia conjugella*, Zeller) has a wide circumpolar distribution [24, 25]. Its main host, rowan (*Sorbus aucuparia*), is a masting species with spatiotemporally synchronized crop output [26]. In heavy intermast years, the apple fruit moth can hatch to find no host material available and will therefore seek secondary hosts, causing serious damage to apple crops [27]. *A. conjugella* is known to have high genetic diversity and a wide distribution in Fennoscandia [28–30], but the lack of complete mitogenomes for *A. conjugella* and the family Argyresthiidae hampers further studies on systematics, population genetics, taxonomy and evolutionary biology.

Our primary aim was to characterize the first entire mitogenome of a species in the Argyresthiidae family using the apple fruit moth in Norway as our study organism. Second, we analysed genome structure, base composition, substitution, and evolutionary rates among superfamilies using previously published Lepidoptera mitogenomes to obtain a better understanding

of the phylogeny of Lepidoptera. We hypothesized that our phylogenetic analysis would recover Argyresthiidae nested with Yponomeutoidea. Furthermore, we evaluated the phylogenetic hypothesis that Argyresthiidae shows a sister-group relationship with Lyonetiidae, i.e., the ‘AL’ clade (Argyresthiidae + Lyonetiidae) of Sohn et al. (2013) [16] obtained based on nuclear genes. Finally, we wanted to provide an up-to-date identification of source taxa of lepidopteran sequences lacking superfamily-, family-, and/or genus-level ID on GenBank using a phylogenetic systematics framework.

Results and discussion

Genome assembly

Except for the variable control region (CR) and NORGAL assemblies, we recovered the same gene order and content using both of our mitogenome assembly strategies. We discovered that NORGAL failed to assemble mitogenome sequences when using the de novo assembly strategy, with no mitogenomic features (PCGs, tRNAs, and rRNAs) found for both assemblies resulting from using default (assembly size = 24,871 bp) and adjusted parameters (assembly size = 29,520 bp, $-m$ 500). However, when run using the baited de novo assembly strategy, NORGAL recovered the same gene order and content as SPADES and GENEIOUS PRIME[®], with the exception of the large ribosomal RNA gene (*rrnL*), which differed in size by 1 bp and sequence from that recovered by SPADES and GENEIOUS PRIME[®] (pairwise p -distance = 0.01). We found that the variations in mitogenome sizes were associated with the properties of the control region (CR), which include variation in the copy number of tandemly repeated sequences and extensive length variation of a variable domain [31, 32]. When using SPADES under the de novo assembly strategy, the nearly complete CR (1101 bp) was recovered. When the baited de novo assembly strategy was used, SPADES recovered a partial CR of 380 bp in which the repetitive sequences could not be assembled. As a result, we present the complete mitogenome sequences of the apple fruit moth from the SPADES de novo assembly, where the mitogenome is a 16,044 bp closed circular molecule (GenBank accession: ON496993; Fig. 1). Interestingly, the mitogenome size of the apple fruit moth was similar to available Yponomeutoid mitogenomes [33–35], which are relatively longer on average compared to other superfamilies of Lepidoptera ($n = 4$, 16092 ± 353 bp, Table 1).

Genome organization and base composition

The gene content of the apple fruit moth mitogenome is similar to that of other Dityrsian insects studied previously, with 22 tRNA genes, 13 PCGs, 2 rRNAs and a non-coding control region. The low-strand codes for 9 PCGs

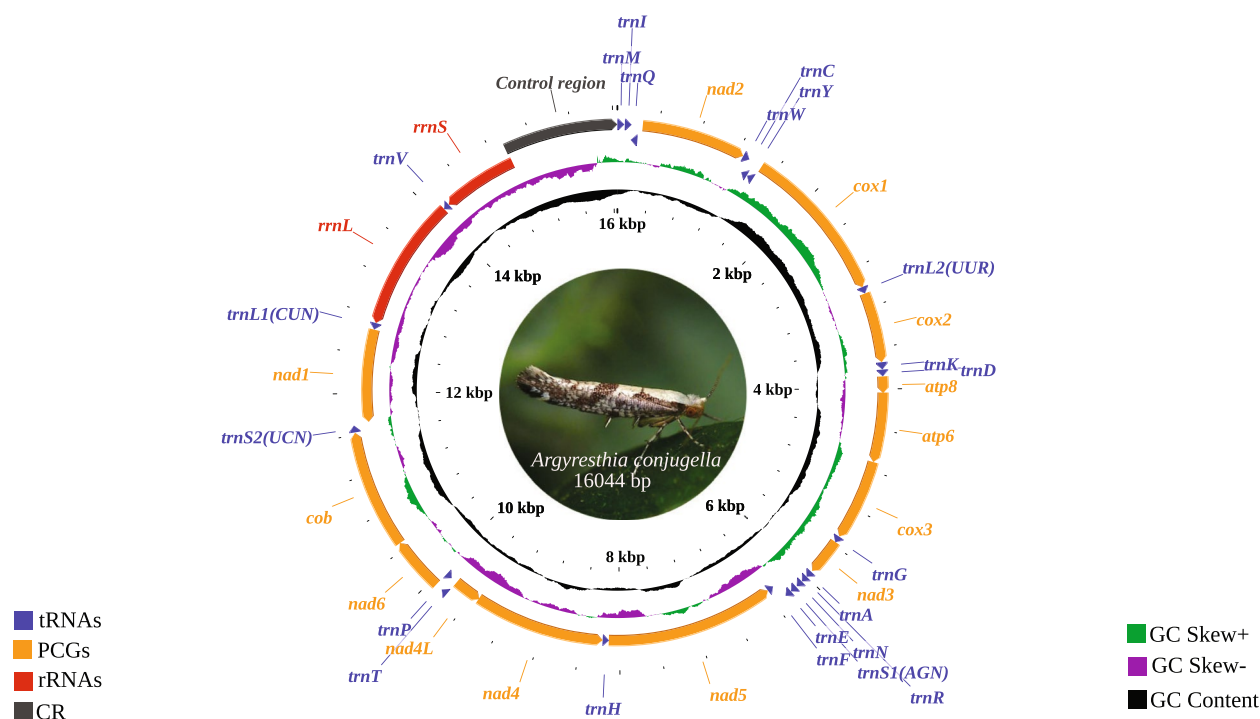


Fig. 1 Circular map of the complete mitogenome of *Argyresthia conjugella* depicting gene order. Labelling of tRNA genes was conducted in accordance with IUPAC-IUB single-letter amino acid codes

(*cob*, *cox1*, *cox2*, *cox3*, *atp6*, *atp8*, *nad2*, *nad3* and *nad6*), 14 tRNAs (*trnM*, *trnI*, *trnW*, *trnL2*, *trnK*, *trnD*, *trnG*, *trnA*, *trnR*, *trnN*, *trnS1*, *trnE*, *trnT* and *trnS2*), 4 PCGs (*nad1*, *nad4*, *nad4L* and *nad5*), 8 tRNAs (*trnC*, *trnF*, *trnH*, *trnL1*, *trnP*, *trnQ*, *trnV*, *trnY*) and two mitochondrial rRNAs (*rrnL* and *rrnS*) (Fig. 1, Table 2). The lengths of the tRNA genes range from 64 to 75 bp (Table 2), which is well within the range of the corresponding tRNA genes of other lepidopterans: *Plutella xylostella* [34], *Parnassius apollo* [36], *Leucoma salicis* [7], *Ephesthia kuehniella* [37] and *Speiredonia retorta* [6]. All 22 tRNAs had cloverleaf secondary structures, except *trnS1*, where one of the dihydrouridine (DHU) arms is missing (Fig. 2). The loss of the DHU arm in tRNAs has been detected in various Lepidoptera species [6, 38, 39]. DHU lacking arm was hypothesized to have evolved in response to recognition signals for seryl-tRNA synthetases, reflecting potential differences in gene expression [40, 41]. The location of *rrnL* is between *trnV* and *trnL1*, while *rrnS* is detected between the control region and *trnV*. These are the same gene positions found in *P. xylostella* [34]. The lengths of *rrnL* and *rrnS* in *A. conjugella* are 1371 bp and 783 bp, while the lengths of these genes are 1371 bp & 783 bp, 1344 bp & 840 bp and 1413 bp & 781 bp in *S. retorta*, *L. salicis* and *P. xylostella*, respectively [6, 7, 34]. The rRNA genes were A+T rich (82%), falling within

the range detected in other Lepidoptera species, including *Agrotis segetum* [42], *Agrotis ipsilon* [43], *Spodoptera frugiperda* [44], and *Papilio machaon* [45]. The rRNA AT and GC skewness values were to be negative in most of the analyzed Lepidoptera mitogenomes in the study, including *A. conjugella*; however, in *Tecia solanivora* [46], *Spilarctia subcarnea* [47] and *S. retorta* [6], these values were positive. In *A. conjugella*, the *cox1* gene starts with ATT, which is different from the start codon in the superfamily Yponomeutoidea members *P. xylostella*, *Leucoptera malifoliella* and *Prays oleae*, where the gene start codon is CGA. The start codon of the *cox1* gene was found to be variable in other Lepidoptera species [48]. The size of this gene (1534) in *A. conjugella* is 3 bp larger than that in these three species (*P. xylostella*, *L. malifoliella* and *P. oleae*) in the same superfamily. The *cox2* gene size (682 bp) is the same size as that of *L. malifoliella* but larger than that found in *P. xylostella* and *P. oleae* (679), while all these species have the size of the *cox3* gene (789 bp). The largest PCG found in *A. conjugella* mitogenomes is *nad5* (1732 bp), and the smallest one is *atp8* (162 bp). These results are widely reported in various insect mitogenomes [49, 50]. Overlap of the alginate sequences of *atp6* and *atp8* in *A. conjugella* (Fig. 3) showed the conserved nucleotide sequence ATG ATA A, which is detected in most lepidopteran species [34, 51].

Table 1 General information and nucleotide composition for a subset of 51 representative mitochondrial genomes of the order Lepidoptera and 5 Trichopteran outgroups used in this study

Superfamily	Family	Species	Accession number	Size (bp)	Whole genome composition							
					T%	C%	A%	G%	A + T%	AT skew	GC skew	
Bombycoidea	Bombycidae	<i>Bombyx huttoni</i>	KP216766	15638	38.0	11.1	43.8	7.1	81.8	0.071	-0.220	
	Sphingidae	<i>Ampelophaga rubiginosa</i>	KT153024	15,82	40.1	11.0	41.4	7.5	81.5	0.016	-0.189	
	Eupterotidae	<i>Ganisa cyanogrisea</i>	MF100140	15174	39.8	12.8	39.7	7.7	79.5	-0.001	-0.249	
	Endromidae	<i>Prismosticta fenestrata</i>	MF100145	15772	39.9	11.2	41.2	7.6	81.1	0.016	-0.191	
	Saturniidae	<i>Neoris haraldi</i>	MF664471	15343	40.3	12.4	39.4	7.9	79.7	-0.011	-0.222	
	Brahmaeidae	<i>Brahmaea certhia</i>	MG018026	15429	40.2	11.6	40.7	7.5	80.9	0.006	-0.215	
	Lasiocampidae	<i>Dendrolimus houi</i>	MK952760	15443	38.8	12.5	41.2	7.6	80.0	0.030	-0.244	
	Callidulidae	<i>Pterodecta felderi</i>	MT370823	15340	39.3	12.7	40.0	7.9	79.3	0.009	-0.233	
	Copromorphoidea	<i>Carposina sasakii</i>	HQ840719	15611	39.5	10.7	42.0	7.8	81.5	0.031	-0.157	
	Cossoidea	<i>Zeuzera multistrigata</i>	MF491642	15260	39.2	13.1	39.7	7.9	78.9	0.006	-0.248	
Drepanoidea	Drepanidae	<i>Oreta fuscopurpurea</i>	MG572766	15564	39.2	12.3	40.4	8.1	79.6	0.015	-0.206	
	Epicopeiidae	<i>Epicopeia hainesii</i>	MK033610	15395	39.6	11.8	41.0	7.6	80.6	0.017	-0.216	
	Stathmopodidae	<i>Atrijuglans hetaohei</i>	KT581634	15379	42.1	11.1	39.3	7.6	81.4	-0.034	-0.187	
	Scythrididae	<i>Scythris sinensis</i>	MH230111	15216	42.2	11.4	38.7	7.7	80.9	-0.043	-0.194	
	Coleophoridae	<i>Coleophora theinella</i>	MH473596	15539	41.6	11.8	39.1	7.5	80.7	-0.031	-0.223	
	Oecophoridae	<i>Opisina arenosella</i>	MK467611	15389	41.3	12.0	39.2	7.4	80.5	-0.026	-0.237	
	Cosmopterigidae	<i>Meleonoma mirabilis</i>	MW366996	15268	40.0	14.8	37.1	8.2	77.1	-0.038	-0.287	
	Lecithoceridae	<i>Issikiopteryx taipingensis</i>	ON323501	15234	41.5	12.4	38.4	7.7	79.9	-0.039	-0.234	
	Gelechiidae	<i>Monochroa</i> sp.	ON323502	15299	40.9	12.5	39.0	7.7	79.9	-0.024	-0.238	
	Geometroidea	Uranidae	<i>Acropteryx iphiata</i>	MN093120	15346	37.9	11.5	43.1	7.5	81.0	0.064	-0.211
Hedyloidea	Geometridae	<i>Armaica recuraria</i>	MZ713635	15609	39.8	11.0	41.6	7.6	81.4	0.022	-0.183	
	Hedyliidae	<i>Macrosoma conifera</i>	NC_050856	15344	41.2	10.9	40.5	7.5	81.7	-0.009	-0.185	
	Hepialidae	<i>Endoclita signifer</i>	KT780172	15285	41.4	10.5	40.5	7.6	81.9	-0.011	-0.160	
	Hesperioidea	Hesperidae	<i>Chooaspes benjamini</i>	JX101620	15272	40.7	11.7	40.1	7.5	80.8	-0.007	-0.219
	Noctuoidea	Eutelidae	<i>Eutelia adaltricoidea</i>	KJ185131	15360	40.7	11.3	40.2	7.8	80.9	-0.006	-0.183
	Erebioidea	Nolidae	<i>Sinna extrema</i>	MG872330	15345	40.7	11.4	40.1	7.8	80.8	-0.007	-0.188
		Erebidae	<i>Asota caricae</i>	MZ779033	15409	40.3	12.2	40.0	7.5	80.3	-0.004	-0.239
		Erebidae	<i>Spiralaria casigneta</i>	MZ959068	15402	40.2	12.2	40.2	7.4	80.4	0.000	-0.245
		Noctuidae	<i>Acrionicta rumicis</i>	NC_062117	15388	40.3	11.7	40.4	7.7	80.7	0.001	-0.206
		Notodontidae	<i>Syntypistis chambae</i>	NC_062123	15938	39.4	11.3	41.9	7.4	81.3	0.031	-0.209

Table 1 (continued)

Superfamily	Family	Species	Accession number	Size (bp)	Whole genome composition						
					T%	C%	A%	G%	A + T%	AT skew	GC skew
Papilionoidea	Lycaenidae	<i>Cigaritis/Spindasis takanonis</i>	HQ184266	15349	41.0	10.7	41.3	6.9	82.3	0.004	-0.216
	Riodinidae	<i>Abisara fylloides</i>	HQ259069	15301	41.7	11.3	39.5	7.5	81.2	-0.027	-0.202
	Papilionidae	<i>Graphium chironides</i>	KP159289	15235	40.8	11.9	39.6	7.7	80.4	-0.015	-0.214
	Pteridae	<i>Collas erate</i>	KP715146	15184	41.4	11.2	39.9	7.5	81.3	-0.018	-0.198
Pyraloidea	Nymphalidae	<i>Apatura laverna</i>	MF444860	15187	40.2	12.2	39.7	7.9	79.9	-0.006	-0.214
	Thyrididae	<i>Pyrioides aurea</i>	KT337662	15362	40.0	12.2	40.0	7.9	80.0	0.000	-0.214
	Pyralidae	<i>Orthaga olivacea</i>	MN078362	15174	41.2	12.8	37.8	8.2	79.0	-0.043	-0.219
	Crambidae	<i>Crambus perlellus</i>	NC_061606	15440	41.0	11.2	40.3	7.5	81.3	-0.009	-0.198
	Choreutidae	<i>Choreutis emplecta</i>	MG581932	14718	40.8	11.4	39.9	7.9	80.7	-0.011	-0.181
	Sesiidae	<i>Sesia sinigensis</i>	MN708363	15454	38.6	12.5	41.0	7.9	79.6	0.030	-0.225
Tineoidea	Tineidae	<i>Amorophaga japonica</i>	MH823253	15027	39.4	10.9	42.3	7.5	81.7	0.035	-0.185
	Psychidae	<i>Dahlia ochrostigma</i>	MK890245	15429	39.1	11.1	42.8	7.0	81.9	0.045	-0.227
	Gracillariidae	<i>Gibbovalva kobusi</i>	MK956103	15717	39.5	11.6	41.0	7.9	80.5	0.019	-0.190
	Meessiidae	<i>Eudarcia gwangneungensis</i>	MN413148	15391	39.4	12.7	40.4	7.4	79.8	0.013	-0.264
Tortricoidea	Tortricidae	<i>Eudemis lucina</i>	MK820027	16056	39.9	11.6	40.6	7.8	80.5	0.009	-0.196
	Yponomeutoidea	<i>Plutella xylostella</i>	JF911819	16079	40.5	10.9	40.9	7.7	81.4	0.005	-0.172
Zygaenoidea	Lyonetidae	<i>Leucoptera malifoliella</i>	JN790955	15646	40.6	10.4	41.9	7.0	82.5	0.016	-0.195
	Praydidae	<i>Prays oleae</i>	KM874804	16499	40.1	10.7	41.7	7.5	81.8	0.020	-0.176
	Argyresthiidae	<i>Argyresthia conjugella</i>	ON496993*	16044	41.2	10.6	40.8	7.4	82.0	-0.005	-0.178
	Limacodidae	<i>Iragoides fasciata</i>	MK250437	15645	42.0	10.8	40.0	7.2	82.0	-0.024	-0.200
Hydropsychoidea	Zygaenidae	<i>Amesia sanguiflua</i>	NC_046467	15203	39.8	12.4	40.0	7.8	79.8	0.003	-0.228
	Hydropsychoidea	<i>Cheumatopsyche brevilineata</i>	KX385010 ^a	15302	38.6	14.5	39.2	7.7	77.8	0.008	-0.306
Limnephiloidea	Limnephilidae	<i>Limnephilus hyalinus</i>	NC_044710 ^a	15168	38.4	13.9	39.5	8.1	77.9	0.014	-0.264
	Phryganeidae	<i>Phryganea cinerea</i>	NC_039714 ^a	15039	39.1	14.0	39.2	7.7	78.3	0.001	-0.290
Stenopsychidae	Phryganopsychidae	<i>Phryganopsyche latipennis</i>	NC_043771 ^a	15167	39.5	13.4	39.1	8.1	78.6	-0.005	-0.247
	Stenopsychinae	<i>Stenopsyche angustata</i>	NC_051529 ^a	15371	35.0	15.5	41.6	7.8	76.6	0.086	-0.330

*This study

^aTrichopteran outgroup

Table 2 The organization and characteristics of the complete mitochondrial genome of *Argyresthia conjugella*

Gene	Strand	Position		Size		Codon		Anticodon	IGN
		Start	End	size (bp)	Amino acid	Start	Stop		
<i>trnM</i>	L	1	68	68				AUG	0
<i>trnI</i>	L	72	139	68				AUC	3
<i>trnQ</i>	H	135	205	71				CAA	-5
<i>nad2</i>	L	246	1259	1014	337	ATT	TAA		40
<i>trnW</i>	L	1261	1328	68				UGA	1
<i>trnC</i>	H	1321	1385	65				UGC	-8
<i>trnY</i>	H	1390	1455	66				UAC	4
<i>cox1</i>	L	1464	2997	1534	511	ATT	T		8
<i>trnL2^(UUR)</i>	L	2994	3062	69				UUA	-4
<i>cox2</i>	L	3062	3743	682	227	ATG	TAA		-1
<i>trnK</i>	L	3743	3815	73				AAG	-1
<i>trnD</i>	L	3815	3881	67				GAC	-1
<i>atp8</i>	L	3881	4042	162	53	ATC	TAA		-1
<i>atp6</i>	L	4036	4712	677	225	ATG	TA		-7
<i>cox3</i>	L	4713	5501	789	262	ATG	TAA		0
<i>trnG</i>	L	5503	5570	68				GGA	1
<i>nad3</i>	L	5567	5923	357	118	ATT	TAA		-4
<i>trnA</i>	L	5935	6002	68				GCA	11
<i>trnR</i>	L	6000	6064	65				CGA	-3
<i>trnN</i>	L	6066	6131	66				AAC	1
<i>trnS1^(AGN)</i>	L	6140	6207	68				AGC	8
<i>trnE</i>	L	6206	6274	69				GAA	-2
<i>trnF</i>	H	6311	6379	69				UUC	36
<i>nad5</i>	H	6379	8110	1732	577	ATT	TA		-1
<i>trnH</i>	H	8107	8174	68				CAC	-4
<i>nad4</i>	H	8173	9519	1347	448	ATG	TAA		-2
<i>nad4L</i>	H	9513	9800	288	95	ATG	TAA		-7
<i>trnT</i>	L	9803	9866	64				ACA	2
<i>trnP</i>	H	9866	9933	68				CCA	-1
<i>nad6</i>	L	9938	10468	531	176	ATT	TAA		4
<i>Cob</i>	L	10468	11622	1155	384	ATG	TAA		-1
<i>trnS2^(UCN)</i>	L	11626	11692	67				UCA	3
<i>nad1</i>	H	11709	12656	948	315	ATG	TAA		16
<i>trnL1^(CUN)</i>	H	12654	12728	75				CUA	-3
<i>rrnL</i>	H	12726	14096	1371					-3
<i>trnV</i>	H	14094	14161	68				GUA	-3
<i>rrnS</i>	H	14161	14943	783					-1
Control region		14944	16044	1101					

IGN values represent intergenic nucleotides and overlapping nucleotides (–)

H-strand Heavy strand, L-strand Low strand

We found that the locations of the *trnM* gene follow the ditrysian type *trnM-trnI-trnQ* [52], which is different from non-ditrysian groups in Lepidoptera and from the ancestral order in which *trnM* is translocated: *trnI-trnQ-trnM* [52–54]. The control region of *A. conjugella* is large (1101 bp), which is a common feature detected

in the superfamily Yponomeutoidea [35]. In comparison, the CR of the olive and diamondback moths were found to be ~1600 bp and ~1081 bp, respectively [34, 35]. We found that the CR is comprised of nonrepetitive sequences, including the motif 'ATAGA' followed by a 20 bp poly-T stretch, dinucleotide microsatellites (AT)₁₈

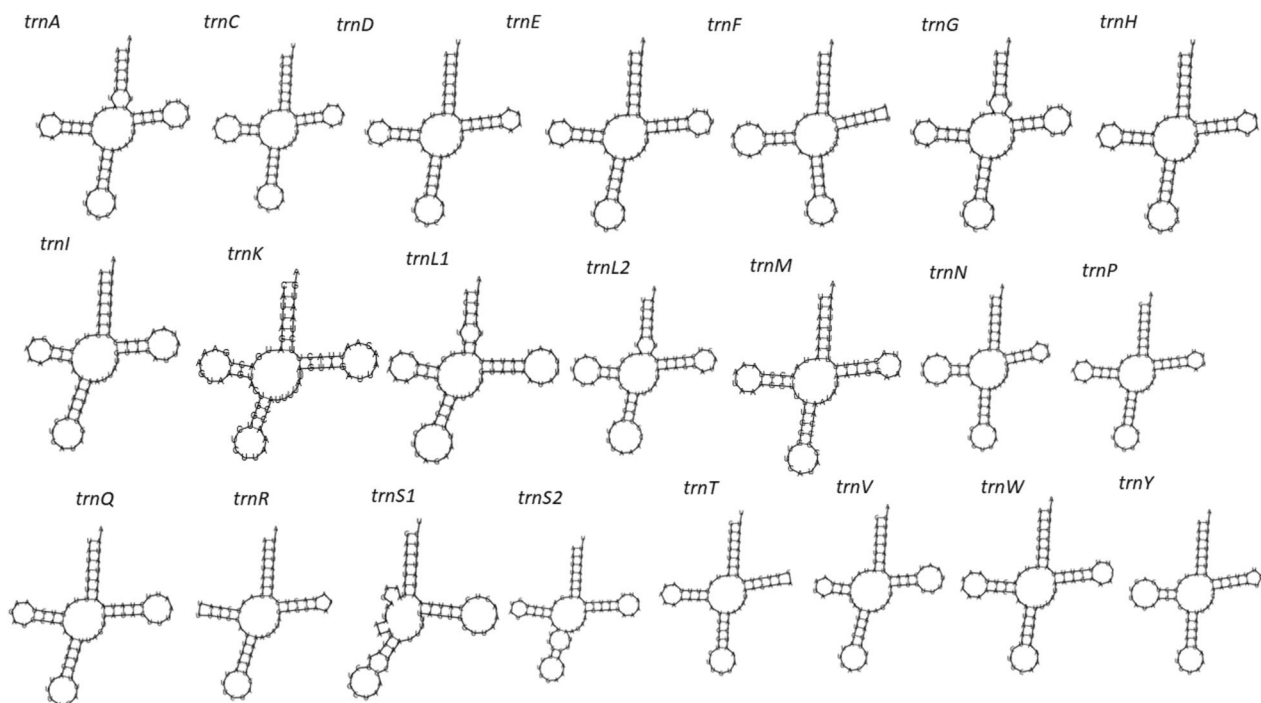


Fig. 2 Predicted secondary structures of the 22 typical tRNA genes in the *A. conjugella* mitogenome

and (AT)₅₃, each flanked by ATTTA motifs, a (TAAA)₄ adjacent to *trnM* instead of the 11 bp poly-A adjacent to tRNAs, and several imperfect repeat elements, indicating that the sequence in the present study may be partial. We found that the nucleotide composition of the CR was highly AT-rich, where the AT content was estimated at 94.3%, (A: 47.6%, T: 46.7%, G: 1.8%, C: 3.9%), where the AT skew was positive and the GC skews was negative, 0.010 and -0.368, respectively. Overall, the nucleotide composition of the apple fruit moth mitogenome was also highly AT-rich, where the AT content was estimated at 82%, (A: 40.8%, T: 41.2%, G: 7.4%, C: 10.6%), and AT and GC skews were negative, -0.005 and -0.178, respectively (Table 1). These results are in agreement with results obtained in *P. xylostella* [34], *L. salicis* [7], *E. kuehniella* [37] and *S. retorta* [6].

The codon usage in *A. conjugella* was compared with twelve Lepidopteran species from different families (Fig. 4). The comparison showed that the pattern of codon usage in the PCGs of the *A. conjugella* mitogenome is very similar to the patterns in these Lepidopteran mitogenomes. Asn, Ile, Leu2, Met and Phe are the most commonly used codon families in all these species, while Cys codons are the rarest (Figs. 4 and 5). The relative synonymous codon usage (RSCU) was analysed for *A. conjugella* and compared with the same set of Lepidopteran insects (Fig. 6). CTG, CTC, AGG and ACG were completely absent in the *A. conjugella* mitogenome

PCGs. Codons with high G and C contents are also rare or absent in the PCGs in other Lepidopteran mitogenomes. Moreover, TTA (Leu2), TCT (Ser2), CGT (Arg), GCT (Ala), and GGA (Gly) are the most frequently used codons and account for 36.41%. These five amino acids are also detected in other Lepidoptera species, such as *Manduca sexta* [55], *Helicoverpa armigera* [56], *P. xylostella* [34], *T. solanivora* [46], *P. machaon* [45], and *Ostrinia nubilalis* [57]. In particular, Leu2 was found to be the most frequently detected amino acid in all Lepidoptera species in the study, and this result is supported by results found in *L. salicis* [7] and *S. retorta* [6].

Phylogenetics

To obtain an overview of *A. conjugella* and its relationships with other Lepidoptera species, our study investigated 18 superfamilies representing 42 families and 507 Lepidoptera species (Tables S1, S2 and Figure S2). This is the first phylogenetic study (using the mt genome) of *A. conjugella* in the Argyresthiidae family, which belongs to the Yponomeutoidea superfamily. Various studies tried to resolve phylogenetic tree of Lepidoptera using mitochondrial genomes, nucleotide alignments, amino acid alignments and transcriptomes and target enrichment approaches [6, 7, 9, 17–21, 58]. However, inadequate node support hindered research that attempted to unravel relationships among superfamilies [17–21]. The challenges are not the lack of data but, how to the

<i>Adoxophyes orana</i>	AAATTGAAAATGATAAATAATTTAT -7
<i>Agrotis ipsilon</i>	AAATTGAAAATGATAAATAATTTAT -7
<i>Agrotis segetum</i>	AAATTGAAAATGATAAGTAATTTAT -7
<i>Amyelois transitella</i>	TTCTTGAAAATGATAACAAACTTAT -7
<i>Argyresthia conjugella</i>	AATTTGAAAATGATAACAAATTTAT -7
<i>Chilo suppressalis</i>	TAACTGAAAATGATAAATAATTTAT -7
<i>Cydia pomonella</i>	TAATTGAAAATGATAAATAACTTAT -7
<i>Ephestia kuehniella</i>	CCTTTGAAAATGATAACAAATTTAT -7
<i>Grapholita molesta</i>	TAACTGAAAATGATAAATAATTTAT -7
<i>Helicoverpa armigera</i>	CAATTGAAAATGATAAGTAATTTAT -7
<i>Helicoverpa zea</i>	CAATTGAAAATGATAAGTAATTTAT -7
<i>Leucoptera malifoliella</i>	TTTTTGAAAATGATAACTAATTTAT -7
<i>Lobesia botrana</i>	TAATTGAAAATGATAAATAATTTAT -7
<i>Lyonetia clerkella</i>	AAACTGAAAATGATAACAAATTTAT -7
<i>Lymantria dispar</i>	TAACTGAAAATGATAAGTAATTTAT -7
<i>Manduca sexta</i>	CATTGAAAATGATAAGAAATTTAT -7
<i>Papilio machaon</i>	AAATTGAAAATGAAAATAATTTAT -7
<i>Parnassius apollo</i>	AAACTGAAAATGATAAGTAATTTAT -7
<i>Pieris rapae</i>	ATCTTGAAAATGATAAATAACTTAT -7
<i>Plodia interpunctella</i>	TAATTGAAAATGATAACAAATTTAT -7
<i>Plutella australiana</i>	TTATTGAAAATGATAACAAATTTAT -7
<i>Plutella xylostella</i>	TATTGAAAATGATAACAAATTTAT -7
<i>Prays oleae</i>	AAACTGAAAATGATAACAAATTTAT -7
<i>Operophtera brumata</i>	ACTGAAAATGATAAGCAATTTAT -7
<i>Ostrinia nubilalis</i>	AATTTGAAAATGATAACTAACTTAT -7
<i>Ostrinia furnacalis</i>	AATTTGAAAATGATAACTAACTTAT -7
<i>Spodoptera frugiperda</i>	TAATTGAAAATGATAAGTAATTTAT -7
<i>Tecia solanivora</i>	AACTTGAAAATGATAACAAATTTAT -7

Fig. 3 Alignment of *atp8* and *atp6* overlap of the selected lepidopteran species in the study, including *A. conjugella*. The green arrow shows the *atp6* start codon, and the red arrow shows the *atp8* stop codon

data analyze, the quality of data and the number of taxon investigated [18, 21]. We constructed a phylogenetic tree using 507 Lepidoptera species (Fig. S2), and the subset data using 51 species (Fig. 7) to understand the position

of *A. conjugella* in Lepidoptera phylogenetic tree. Using the ML approach, analyses of the three datasets (specified in the materials & methods section) resulted in the generation of three topologies. Generally, our study agrees

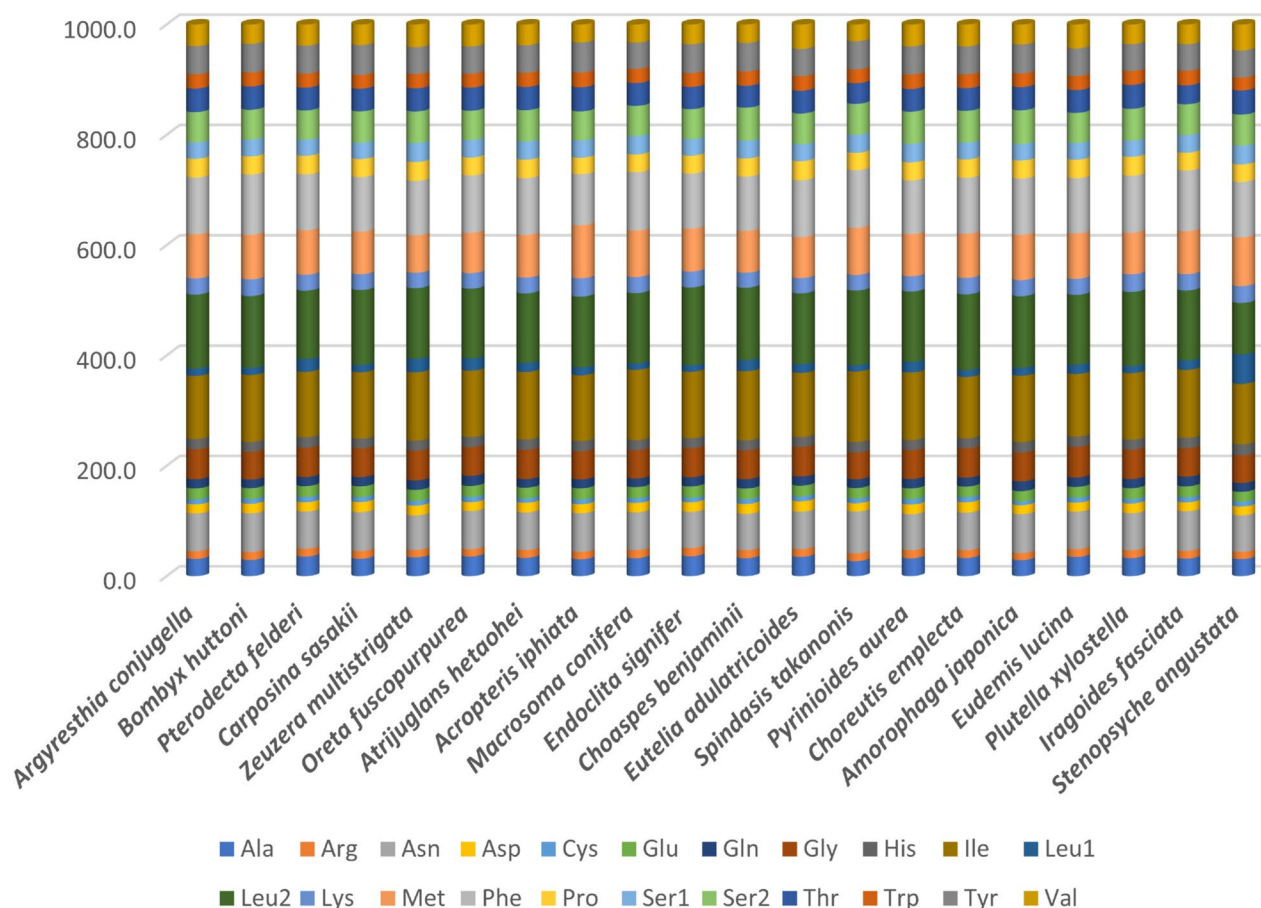


Fig. 4 Comparison of codon usage of the 20 selected mitochondrial genomes of the Lepidoptera species in the study, including *A. conjugella*

with the most updated study Rota et al. (2022) [21], that detected nine main clades superfamilies in a butterfly and moth phylogeny using 331 genes for 200 taxa. Additionally, our phylogenetic analysis supports the previous morphological characterization of the Yponomeutoidea superfamily [16, 59, 60]. The 507 Lepidoptera species showed that some families clustered together, such as Papilionidae & Pieridae, Pyralidae & Tortricidae, Geometridae & Sphingidae, Erebidae & Noctuidae and Gelechiidae & Sphingidae, while other families as Tortricidae and Crambidae clustered alone and separately. Yponomeutoidea was recovered as a well-supported monophyly group and as one of the earliest lepidopteran groups after Tineoidea and the basal Hepialoidea (Fig. 7, Figures S1 and S2). However, the paraphyletic Tineoidea to some extent led to the phylogenetic instability of the monophyly of Yponomeutoidea in cases of Datasets 1 and 2 (Fig. 7, Figure S1), which was fully resolved with dense taxon sampling (Figure S2). Wang et al. (2018) [61], Bao et al. (2019) [38], Jeong et al. (2022) [23] and Zhang et al. (2020) [62], all found similar results for Yponomeutoidea and Tineoidea superfamilies. Furthermore, Boa

et al. (2019) [38] and Jeong et al. (2022) [23] also found that Yponomeutoidea, Tineoidea and Gracillarioidea in Ditrysia have strong phylogenetic relationships. We also detected strong relationships between Yponomeutoidea, Zygaenidae and Tortricidae, findings that are in line with results found by Liu et al. (2016) [48], Zhang et al. (2020) [62], Wang et al. (2018) [61], and Kim et al. (2014) [63]. Only a weak phylogenetic relationship was observed between the superfamilies Yponomeutoidea and Bombycoidea, results that are supported by Liu et al. (2016) [64] and Liu et al. (2017) [65]. Nonetheless, we consistently recovered Argyrethiidae embedded in Yponomeutoidea with a sister-group relationship to Plutellidae (Dataset 1: $SH-aLRT=92$, $UFBoot2=100$; Dataset 2: $SH-aLRT=88$, $UFBoot2=100$; Dataset 3: $SH-aLRT=87$, $UFBoot2=99$). Our phylogenetic tree hypothesis rejects the provisional 'AL' clade (Argyrethiidae + Lyonetiidae) recovered with nuclear gene datasets by Sohn et al. (2013) [16]. We found that Lyonetiidae was unstable, possibly due to its relatively long branch length. We recovered Lyonetiidae as basal to the Yponomeutoidea clade (Figure S1, Dataset 1: $SH-aLRT=99$, $UFBoot2=100$) or as a sister-group



Fig. 5 Relative synonymous codon usage (RSCU) of the 20 selected mitochondrial genomes of Lepidoptera in the study, including *A. conjugella*. Codon are plotted on the x-axis

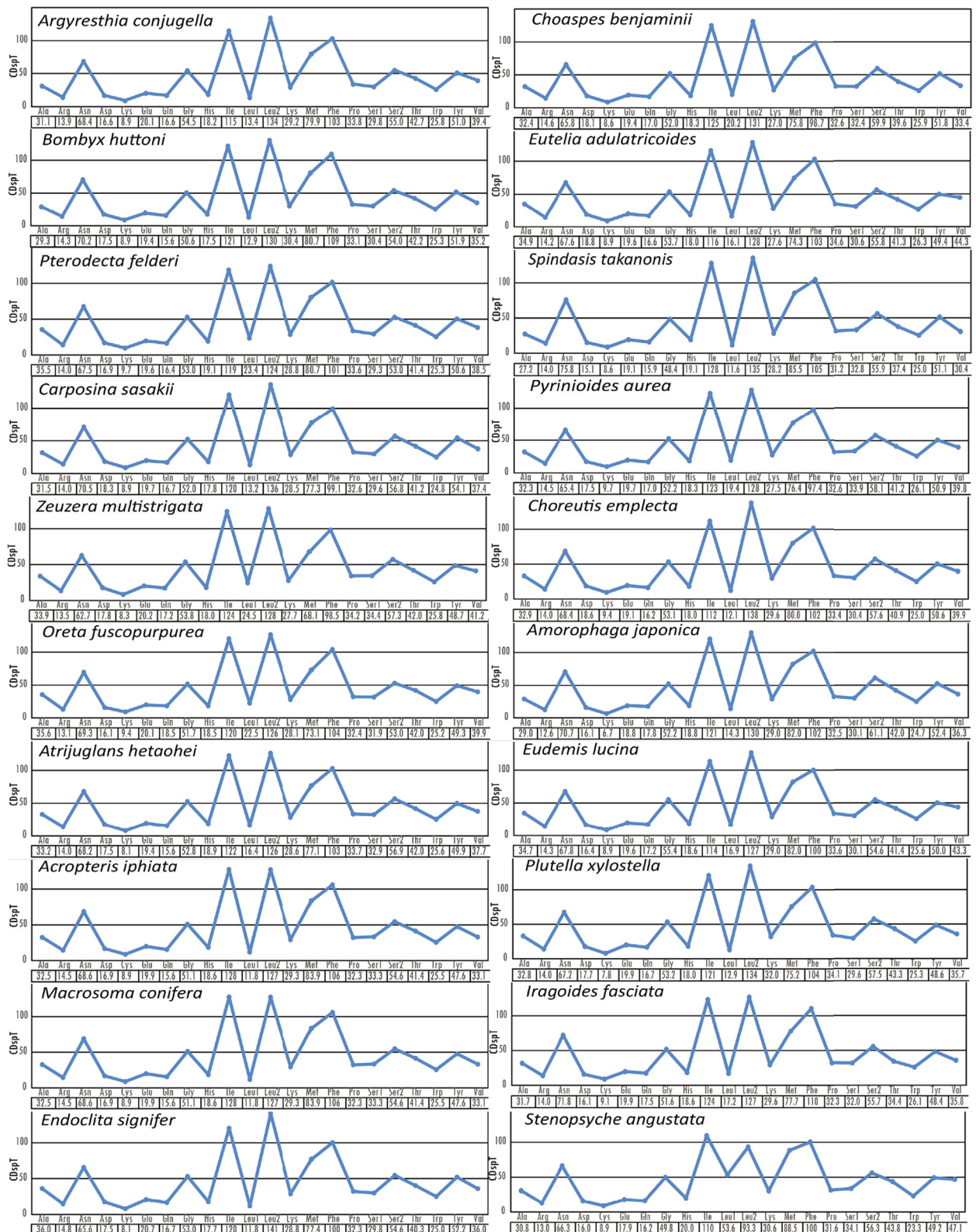


Fig. 6 The distribution of codons among the selected lepidopteran species in the study. CDspT codons per thousand codons

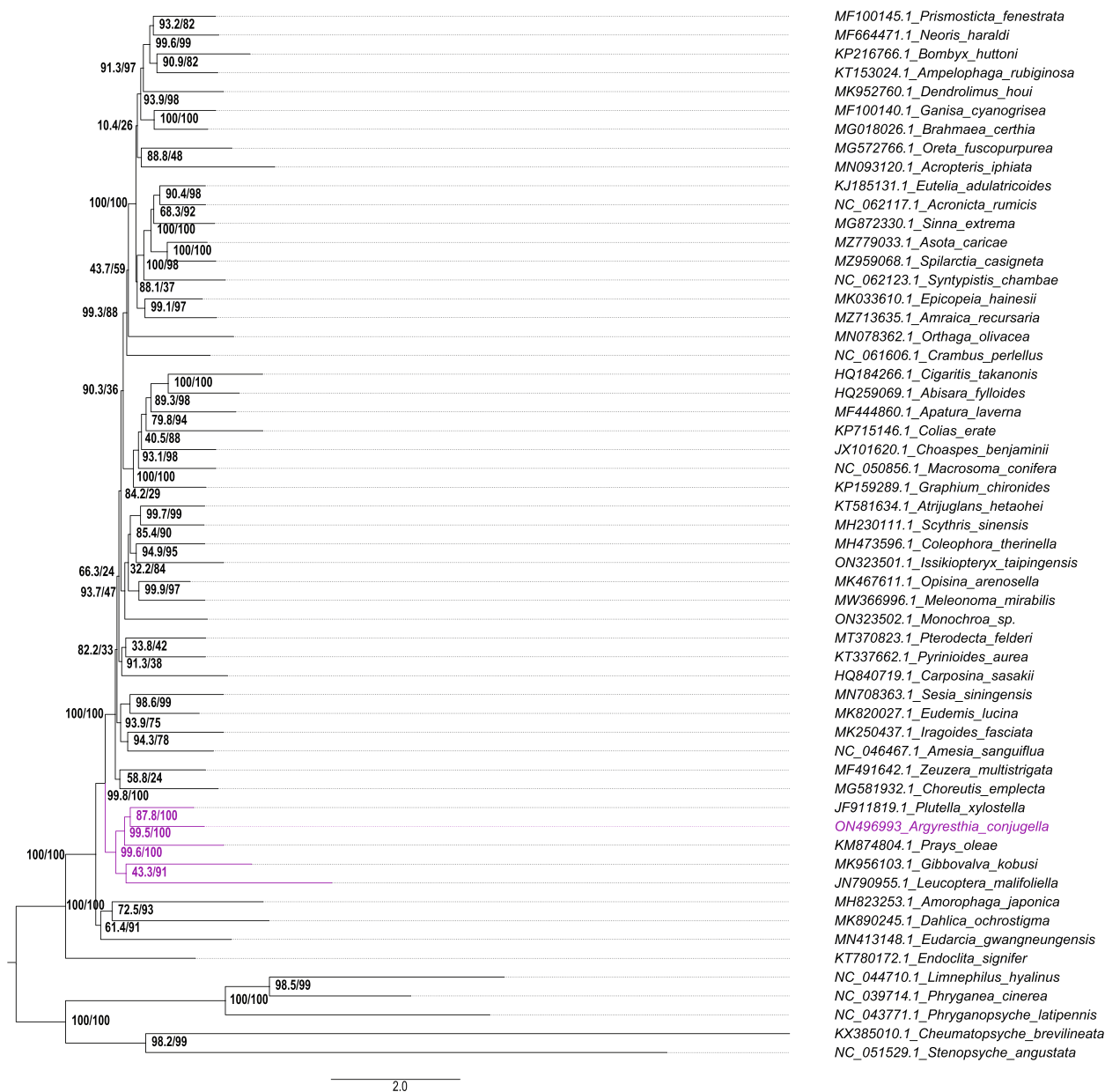


Fig. 7 The phylogenetic tree included *A. conjugella*, constructed using the nearest neighbor interchange (NNI) approach to search for tree topology and for computing branch supports with 1000 replicates of the Shimodaira-Hasegawa approximate likelihood-ratio test *SH-aLRT* [68] and 1000 bootstrapped replicates of the ultrafast bootstrapping (*UFBoot2*) approach [69]. *Phryganea cinerea*, *Phryganopsyche latipennis*, *Cheumatopsyche brevilineata*, *Limnephilus hyalinus*, and *Stenopsyche angustata* were used as controls and outgroup species (Table 1)

to Praydidae with Yponomeutoidea (Figure S2, Dataset 3: *SH-aLRT*=84, *UFBoot2*=100), and as sister-group to Gracillariidae of the order Tineoidea, although with weak support (Fig. 7, Dataset 2: *SH-aLRT*=43, *UFBoot2*=91). With increased taxon sampling, our phylogenetic tree hypotheses strongly supported the basal placement of Lyonetiidae within the Yponomeutoidea clade (Fig. 7, Figure S2, Dataset 2: *SH-aLRT*=98, *UFBoot2*=99).

Moreover, we consistently recovered the previously described pairing of Yponomeutoidea and Gracillariidae as internested subclades [16, 22]. At a higher level, our phylogenetic tree hypothesis recovers some fundamental and uncontroversial lepidopteran clades that agree with the majority of mitogenomic phylogenies as well as those that included both mitochondrial and/or nuclear markers. The analyses found that *A. conjugella* had the

closest relationship with *P. xylostella*, *L. malifoliella* and *P. oleae*, which belong to the Plutellidae, Lyonetiidae and Praydidae families, respectively (Fig. 7, Figure S2). Wei et al. (2013) [34], Sohn et al. (2013) [16], Liu et al. (2016) [48], Yang et al. (2020) [66], Jeong et al. (2021) [67] and Jeong et al. (2022) [23] all found that *P. xylostella*, *L. malifoliella* and *P. oleae* are closely related.

In our study, Tineodia superfamily was represented by four species (*Amorophaga japonica*, *Dahlia ochrostigma*, *Gibbovalva kobusi* and *Eudarcia gwangneungensis*) with relatively high nodal support (Fig. 7, Figure S2). This superfamily is known to have high genetic diversity and has three different lineages [21]. The cross-talk of the complexity and the relationships among Tineidae group and the disagreements within the superfamily Gelechioidea, Carposinoidea, and Pterophoroidea remain unresolved issues. Both this study and that of Rota et al. (2022) [21], detected a sister relationship between Yponomeutoidea and the superfamily Tineidae, and the sub-clades Gelechioidea, Tortricoidea, Zygaenoidea are clustered together in the same clade. Rota et al. (2022) [21], found Gelechioidea clustered at different positions, when different analyses were performed with different datasets, these may be explained by high amount of compositional heterogeneity, or the limited materials used in the study (five species). While our study showed, the 20 species from Gelechioidea superfamily were clearly clustered together using both datasets and data two analyses (EME and NJ), but surprisingly, one single species (*Periacma orthiodes*) belonging to the superfamily Noctuoidea was clustered together with this family. This might be misidentification of the taxon of the mt genome found in the genebank. Our study showed, Gelechioidea grouped together with Pyralidae, these results are in agreement with the results of [17]. Pyaloidea was also sister to Carposinoidea; and Calliduloidea, Pterophoroidea, Gelechioidea and Thyridioidea are recovered in the same part of the tree, but with Thyridioidea sister to Macroheterocera [21]. Previously, Pterophoroidea was reported as a sister group with a monophyletic Papilionoidea, included Hedyloidea and Hesperioidea. In the same study, Choreutoidea and Immoidea were recovered as sister to Tortricoidea [21]. However, when 50 genes were removed, Choreutoidea were recovered as sister to Urodoidea and Pterophoroidea [21]. One phylogenetic study reported Pterophoroidea within the clade Obectomera, [19] but a more recent study showed results contrary to these findings [21]. The position of Pterophoroidea is highly dependent on the dataset. This superfamily is recovered in the same clade with Urodoidea regardless of the alignment analysed, whereas it's recovered in the clade with Gelechioidea, Calliduloidea and Thyridioidea is dependent on which datasets are analysed

[21]. Pterophoroidea can also be recovered as sister to Papilionoidea and Noctuoidea, when different datasets were used [70]. It should also be noted that, using software with systematic errors and alignment issues can persist with regard to detecting homologies due to use of designed to assess the alignment quality using a threshold of alignment scores, [71].

Comprehensive analyses of insect mitogenomes provide important phylogenetic information to identify potentially novel genes that may serve as valuable targets in future research efforts. Further investigations of the whole genome of *A. conjugella* along with other genomes of Lepidoptera species will facilitate the understanding of the taxonomy and evolutionary process acting on the Ditrysia natural group.

Materials and methods

Specimen collection and DNA extraction

During August 2016, we collected a single female apple fruit moth larva from an infested rowan berry in the field in Skiftenes (N 6471746 and E 472502) in southern Norway. To confirm species identification of the larva, we employed both morphological [24, 65] and molecular methods [28] using microscopy and STR markers, respectively. We placed the apple fruit moth larva on rolls of corrugated cardboard until it entered pupal diapause, and then we stored it at -80 °C until DNA was extracted. DNA was extracted from the apple fruit moth pupal tissue using the DNeasy Blood and Tissue Kit (Qiagen, Tokyo) following a modified version of the manufacturer's instructions [28].

Mitogenome sequencing and assembly

We outsourced the whole genome sequencing of the apple fruit moth to the Norwegian Sequencing Centre (Oslo, Norway), where the whole genome library was prepared (insert size=350 bp) and sequenced on one lane of the Illumina HiSeq 4000 platform (Illumina, USA) with paired end (PE) sequencing (2×150 bp). A total of 820,368,162 raw reads (of which 820,365,390 were paired in sequencing, *i.e.*, 410,182,695 PE read clusters) were generated. We evaluated the quality of the Illumina sequencing run using MULTIQC v.2.31 [72]. Then, we used ADAPTERREMOVAL v.2.1.3 to search for and remove adapter sequences and to trim low-quality bases from the 3' end of reads following adapter removal [73, 74]. After quality control (QC), we used the cleaned PE reads for mitogenome assembly by means of two assembly strategies: (i) *de novo* and (ii) 'baited' *de novo*.

For *de novo* sequence assembly (i), we employed the programs NORGAL v.1.0 [75] and SPADES v.3.15.3 [76]. The NORGAL assembler was executed with (1) default parameters and (2) a *k*-mer range of 21–255 with an

Table 3 The substitution model MODELFINDER [101] was used to reconstruct phylogenies based on the maximum-likelihood (ML) criterion in IQ-TREE

Full data (516 taxa)		Subset data (56 taxa)	
Partition	Model	Partition	Model
<i>(a) Dataset 1 (13PCGs_NT)</i>			
<i>nad2_pos1</i>	GTR+F+R9	<i>nad2_pos1</i>	GTR+F+R5
<i>nad2_pos2</i>	GTR+F+R6	<i>nad2_pos2</i>	GTR+F+I+I+R4
<i>nad2_pos3</i>	GTR+F+ASC+R8	<i>nad2_pos3</i>	K3Pu+F+R6
<i>cox1_pos1</i>	GTR+F+I+G4	<i>cox1_pos1</i>	GTR+F+I+I+R4
<i>cox1_pos2_atp6_pos2</i>	GTR+F+I+I+R8	<i>cox1_pos2</i>	TVM+F+I+I+R3
<i>cox1_pos3_nad3_pos3</i>	GTR+F+I+I+R20	<i>cox1_pos3_cox2_pos3_cox3_pos3_nad6_pos3_cob_pos3</i>	TVM+F+R8
<i>cox2_pos1</i>	GTR+F+I+G4	<i>cox2_pos1</i>	GTR+F+I+I+R5
<i>cox2_pos2</i>	GTR+F+I+G4	<i>cox2_pos2_atp6_pos2</i>	GTR+F+I+I+R3
<i>cox2_pos3_atp6_pos3</i>	GTR+F+R20	<i>atp8_pos1_nad6_pos1</i>	TIM2+F+I+G4
<i>atp8_pos1</i>	GTR+F+I+G4	<i>atp8_pos2_nad6_pos2</i>	GTR+F+R5
<i>atp8_pos2</i>	GTR+F+I+G4	<i>atp8_pos3</i>	K3Pu+F+R3
<i>atp8_pos3</i>	GTR+F+I+G4	<i>atp6_pos1_cob_pos1</i>	GTR+F+I+I+R5
<i>atp6_pos1_cox3_pos1</i>	GTR+F+R10	<i>atp6_pos3_nad3_pos3</i>	TPM2u+F+I+I+R6
<i>cox3_pos2</i>	GTR+F+I+G4	<i>cox3_pos1</i>	GTR+F+I+G4
<i>cox3_pos3</i>	GTR+F+ASC+G4	<i>cox3_pos2_nad3_pos2_cob_pos2</i>	GTR+F+I+I+R4
<i>nad3_pos1</i>	GTR+F+I+G4	<i>nad3_pos1</i>	GTR+F+I+I+R3
<i>nad3_pos2</i>	GTR+F+I+G4	<i>nad5_pos1_nad1_pos1</i>	GTR+F+I+I+R5
<i>nad5_pos1_nad1_pos1</i>	GTR+F+R9	<i>nad5_pos2_nad1_pos2</i>	GTR+F+I+I+R4
<i>nad5_pos2_nad1_pos2</i>	GTR+F+I+I+R7	<i>nad5_pos3</i>	GTR+F+I+I+R6
<i>nad5_pos3</i>	GTR+F+I+G4	<i>nad4_pos1_nad4L_pos1</i>	GTR+F+R4
<i>nad4_pos1</i>	GTR+F+I+G4	<i>nad4_pos2</i>	GTR+F+R5
<i>nad4_pos2</i>	GTR+F+I+G4	<i>nad4_pos3</i>	TVM+F+I+I+R6
<i>nad4_pos3</i>	GTR+F+ASC+G4	<i>nad4L_pos2</i>	TIM+F+I+I+R3
<i>nad4L_pos1</i>	GTR+F+I+G4	<i>nad4L_pos3</i>	TVM+F+R4
<i>nad4L_pos2</i>	GTR+F+I+G4	<i>nad1_pos3</i>	TIM+F+R5
<i>nad4L_pos3</i>	GTR+F+ASC+G4		
<i>nad6_pos1</i>	GTR+F+I+G4		
<i>nad6_pos2</i>	GTR+F+I+G4		
<i>nad6_pos3</i>	GTR+F+I+G4		
<i>cob_pos1</i>	GTR+F+I+G4		
<i>cob_pos2</i>	GTR+F+I+G4		
<i>cob_pos3</i>	GTR+F+ASC+G4		
<i>nad1_pos3</i>	GTR+F+ASC+G4		
<i>(b) Dataset 2 (13PCGs_rRNAs_NT)</i>			
<i>nad2_pos1</i>	GTR+F+R9	<i>nad2_pos1</i>	GTR+F+R5
<i>nad2_pos2</i>	GTR+F+I+I+R7	<i>nad2_pos2</i>	GTR+F+I+I+R4
<i>nad2_pos3</i>	GTR+F+ASC+R9	<i>nad2_pos3</i>	GTR+F+R5
<i>cox1_pos1</i>	GTR+F+I+I+R7	<i>cox1_pos1</i>	GTR+F+I+I+R5
<i>cox1_pos2_atp6_pos2</i>	GTR+F+R7	<i>cox1_pos2</i>	TVM+F+I+I+R3
<i>cox1_pos3_cox2_pos3</i>	GTR+F+I+I+R20	<i>cox1_pos3_cox2_pos3_nad6_pos3_cob_pos3</i>	TVM+F+R6
<i>cox2_pos1</i>	GTR+F+I+I+R6	<i>cox2_pos1</i>	GTR+F+I+I+R4
<i>cox2_pos2</i>	GTR+F+I+I+R6	<i>cox2_pos2_atp6_pos2</i>	GTR+F+I+I+R3
<i>atp8_pos1</i>	TIM3+F+I+I+R5	<i>atp8_pos1_nad6_pos1</i>	TIM2+F+I+G4
<i>atp8_pos2</i>	GTR+F+I+I+R5	<i>atp8_pos2_nad6_pos2</i>	GTR+F+R5
<i>atp8_pos3</i>	K3Pu+F+R6	<i>atp8_pos3</i>	K3Pu+F+R3
<i>atp6_pos1_cox3_pos1</i>	GTR+F+R10	<i>atp6_pos1_cob_pos1</i>	GTR+F+I+I+R4

Table 3 (continued)

Full data (516 taxa)		Subset data (56 taxa)	
Partition	Model	Partition	Model
<i>atp6_pos3</i>	TIM2+F+R15	<i>atp6_pos3_nad3_pos3</i>	TPM2u+F+I+I+R5
<i>cox3_pos2</i>	TVM+F+I+I+R7	<i>cox3_pos1</i>	GTR+F+I+G4
<i>cox3_pos3</i>	GTR+F+ASC+R9	<i>cox3_pos2_nad3_pos2_cob_pos2</i>	GTR+F+R4
<i>nad3_pos1</i>	GTR+F+I+I+R6	<i>cox3_pos3</i>	K3Pu+F+R6
<i>nad3_pos2</i>	TVM+F+I+I+R6	<i>nad3_pos1</i>	GTR+F+I+I+R3
<i>nad3_pos3</i>	GTR+F+I+I+R10	<i>nad5_pos1_nad1_pos1</i>	GTR+F+R5
<i>nad5_pos1</i>	TVM+F+I+I+R10	<i>nad5_pos2_nad4_pos2</i>	GTR+F+R5
<i>nad5_pos2_nad1_pos2</i>	GTR+F+I+I+R7	<i>nad5_pos3</i>	K3Pu+F+R5
<i>nad5_pos3_nad4L_pos3</i>	TIM+F+I+I+R20	<i>nad4_pos1_nad4L_pos1</i>	GTR+F+R4
<i>nad4_pos1</i>	TVM+F+R8	<i>nad4_pos3</i>	TIM3+F+I+I+R6
<i>nad4_pos2</i>	GTR+F+R7	<i>nad4L_pos2</i>	TIM+F+I+I+R3
<i>nad4_pos3</i>	GTR+F+ASC+R9	<i>nad4L_pos3</i>	TPM2u+F+R4
<i>nad4L_pos1</i>	GTR+F+I+I+R6	<i>nad1_pos2</i>	TVM+F+I+I+R4
<i>nad4L_pos2</i>	GTR+F+I+I+R5	<i>nad1_pos3</i>	TPM2u+F+R5
<i>nad6_pos1</i>	TIM2+F+R9	<i>rrnL</i>	GTR+F+I+I+R4
<i>nad6_pos2</i>	TVM+F+I+I+R7	<i>rrnS</i>	GTR+F+R4
<i>nad6_pos3</i>	GTR+F+ASC+R9		
<i>cob_pos1</i>	GTR+F+I+I+R8		
<i>cob_pos2</i>	TVM+F+R6		
<i>cob_pos3</i>	TIM2+F+ASC+R11		
<i>nad1_pos1</i>	GTR+F+R7		
<i>nad1_pos3</i>	TIM+F+ASC+R10		
<i>rrnL</i>	GTR+F+I+I+R7		
<i>rrnS</i>	GTR+F+R7		
(c) Dataset 3 (13PCGs_AA)			
<i>nad2_atp6_nad3</i>	mtMet+F+I+I+R10	<i>nad2_cox1_cox2_nad4L_cob_nad1</i>	mtART+F+I+I+R5
<i>cox1_cox3_cob</i>	mtInV+R9	<i>atp8_atp6_cox3_nad3_nad5</i>	mtART+F+R5
<i>cox2</i>	mtMet+I+I+R8	<i>nad4_nad6</i>	mtART+F+I+I+R5
<i>atp8</i>	mtMet+F+I+I+R7		
<i>nad5_nad4_nad4L_nad1</i>	mtMet+F+I+I+R10		
<i>nad6</i>	mtMet+F+R10		

interval of 28 and a contig length threshold of 500 bp. We executed SPADES for all k -mer sizes from 21 to 127 ($-k$ 21, 33, 55, 77, 99, 127), with *-careful* option to minimize number of mismatches in the final contigs. For the baited de novo assembly strategy (ii), we first constructed the FM-index for the mitogenome reference sequence of the olive moth *P. oleae* (Bernard, 1788; Lepidoptera: Yponomeutoidea: Praydidae) (NCBI accession number NC_025948.1; [35] using the index command of the BWA v.0.7.17 aligner [77]. Additionally, we also used the mitogenome reference sequence of the diamondback moth *P. xylostella* [34] (Linnaeus, 1758; Lepidoptera: Yponomeutoidea: Plutellidae) (NCBI accession number JF911819.1). We selected the mitogenomes of the olive and diamondback moths as references due to their completeness and

taxonomic and phylogenetic placement in Yponomeutoidea and the reliability of the PCR-based amplification method used to sequence these mitogenomes, 14 segments of 1.2–2.4 kb and nine segments of variable size, respectively. Second, we aligned the cleaned PE reads of the apple fruit moth separately to the indexed reference genome of the ‘model’ moths using the BWA-MEM algorithm of BWA, excluding reads with a minimum quality score of < 30, and then used the SAMTOOLS v.1.9 suite [78] to convert the SAM to BAM alignment file. Third, we sorted and indexed the BAM alignment file using the *sort* and *index* commands, respectively, from SAMTOOLS. Fourth, we obtained the QC statistics for the sorted and indexed alignment using BAMQC as implemented in QUALIMAP v.2.2.20 [79]. Fifth, we extracted

reads that mapped properly as pairs using SAMTOOLS. Finally, we used the mitochondrial filtered reads for de novo mitochondrial genome assembly using NORGAL, SPADES and GENEIOUS PRIME® v.2022.1.1 (Biomatters Ltd., Auckland, New Zealand; [80]).

Mitogenome annotation and visualization

We conducted a preliminary annotation of the mitogenome assembly referring to the results of the MITOS2 webserver (<http://mitos2.bioinf.uni-leipzig.de/index.py>; [81] Donath et al. 2019) by assessing the location of protein coding (PCGs), transfer RNA (tRNAs), and ribosomal RNA genes (rRNAs). Then, we confirmed gene boundaries for PCGs and rRNAs manually using BLASTn, SMART BLAST, BLASTp, and ORF Finder as implemented at the National Center for Biotechnology Information (NCBI) database [82]. Subsequently, we also validated that coding sequences were translated in the correct reading frame and confirmed the initiation and termination codons in GENEIOUS PRIME® using the published mitochondrial genome sequences of other moths as references, including the olive moth. We then used the program ARWEN [83] to detect the tRNA genes of the apple fruit moth and finally predicted the secondary structures of tRNAs using MITFI [84] as implemented in MITOS2 and TRNASCAN-SE v.2.0 [85]. We also annotated the control (A+T-rich) region (CR) of the apple fruit moth by screening for structural elements characteristic of the region, which include (i) tandem repeats, identified using TANDEM REPEATS FINDER v.4.10 [86] using default settings, and (ii) the motif 'ATAGA' and poly-T stretch. We produced the annotated circular map of the complete mitochondrial genome of the apple fruit moth using the beta version of the CGVIEW server (<http://cgview.ca>; [87]). The secondary structure of tRNAs was predicted using tRNAscan-SE-2.0 [88].

Comparative mitogenomics of Lepidoptera

We conducted a systematic and comprehensive search for complete mitochondrial genomes of Lepidopteran species published in the NCBI nucleotide database using the following keywords: ("lepidoptera"[Organism] OR "lepidoptera"[All Fields]) AND "complete mitochondrial genome"[All Fields] AND mitochondrion[filter] (10 May 2022: 842 hits). We downloaded and processed the full GenBank files in Geneious Prime® to (i) obtain taxonomy metadata, (ii) remove least recently modified duplicates, (iii) remove nonlepidopteran species, and (iv) remove mitogenomes with >90% missing annotations (retained 507 species). To ensure that the taxonomic status of all species was the latest, we verified all the species names against 60 taxonomic databases, including the Catalogue

of Life and the Integrated Taxonomic Information System (ITIS), using the R package *taxize* v.9.94.91 [89]. Then, we corrected any misspellings and used the *classification* function implemented in *taxize* to retrieve the taxonomic ranks of individual species. We included seven species of Trichoptera, representing five families and four superfamilies, to serve as outgroups.

We compared the assembled *A. conjugella* mitochondrial genome with the mitochondrial genomes of 507 other Lepidoptera obtained from GenBank, representing 18 superfamilies and 42 families (Supplementary Table S2). We included only one representative per valid species (longest mitogenome sequence) when more than one known sequence was available in GenBank. We calculated the overall composition of individual mitogenomes based on the proportion of A+T out of the total (%AT content) using MEGA v.11.0.11 [90]. To measure the base composition skewness of nucleotide sequences, we used the formulae of Perna and Kocher (1995) [91]: $AT\text{-skew} = [A-T]/[A+T]$ and $GC\text{-skew} = [G-C]/[G+C]$.

Sequence alignment and phylogenetic reconstruction

We produced codon-aware multiple sequence alignments for each of the 13 PCGs using MACSE v.2.01 [92]. We inspected and manually trimmed each set of alignments using MEGA, and any remaining ambiguously aligned sites were then further trimmed using BMGE v.1.12.1, with a sliding window size of 3 and maximum entropy of 0.5 [93]. We aligned rRNA genes using the online version of MAFFT v.7.299 [94, 95] and removed ambiguously aligned sites using BMGE. Before phylogenetic analysis, we produced two concatenated mitogenomic datasets from (i) the aligned individual PCG datasets (Dataset 1: *13PCGs_NT* dataset) and (ii) the 13 PCGs plus the large and small mitochondrial ribosomal RNA (rRNA) genes (*rrnL* and *rrnS*) (Dataset 2: *13PCGs_rRNAs_NT* dataset) with the R package *concatipede* v1.0.1 [96]. We derived the third mitogenomic dataset by translating the *13PCGs_NT* dataset in MEGA (Dataset 3: *13PCGs_AA* dataset). Furthermore, we used DAMBE v.7.2.141 [97] to conduct two-tailed tests of substitution saturation [98] for each codon position of the 13 PCGs, taking into account the proportion of invariant sites as recommended by Xia and Lemey (2009) [99]. According to the observed index of substitution saturation (I_{SS}), all codon positions showed little saturation ($I_{SS} < I_{SScSym}$ (assuming a symmetrical topology) and $I_{SS} < I_{SScAsym}$ (assuming an asymmetrical topology); see Supplementary Table S2). Likewise, visual inspection of nucleotide saturation for each codon position of the 13 PCGs with DAMBE by plotting transitions and transversions against Kimura two-parameter [100] distances showed little saturation in all codon positions. Therefore, none of the codon

positions were excluded, and the 13 PCG nucleotide (Dataset 1) and protein (Dataset 3) datasets were initially gene-by-codon partitioned (39 partitions) and gene partitioned (13 partitions), respectively. For Dataset 2, we designated two partitions for the rRNA genes (*rrnS* and *rrnL*, treated each as a single partition) and 39 partitions covering the three codon positions in each of the 13 protein-coding genes.

We used MODELFINDER [101] to select the best-fitting partitioning scheme and models of evolution using the corrected Akaike Information Criterion (AICc) and the edge-linked proportional partition model [102] as implemented in IQ-TREE v. 2.2.0.3 [103]. We applied the new model selection procedure (*-m MF+MERGE*), which additionally implements the FreeRate heterogeneity model inferring the site rates directly from the data instead of being drawn from a gamma distribution (*-cmax 20*; [104]). To reduce the computational burden, the top 30% partition merging schemes were inspected using the relaxed clustering algorithm (*-rcluster 30*), as described in [105].

We reconstructed phylogenies based on the maximum likelihood (ML) criterion in IQ-TREE, where we used the substitution models indicated by MODELFINDER (Table 3). We used the nearest neighbor interchange (NNI) approach to search for tree topology and for computing branch supports with 1000 replicates of the Shimodaira-Hasegawa approximate likelihood-ratio test *SH-aLRT* [68] and 1000 bootstrapped replicates of the ultrafast bootstrapping (*UFBoot2*) approach [69]. We abided by the advice that clades with *UFBoot2* ≥ 95 and *SH-aLRT* ≥ 80 can be regarded as being well supported [106].

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12864-023-09905-1>.

Additional file 1: Table S1. List of mitochondrial genomes of Lepidoptera (including *A. conjugella* mitochondrial genome) investigated in the study, representing 18 superfamilies and 42 families (Supplementary Figure S2), including outgroups species (*Phryganea cinerea*, *Phryganopsyche latipennis*, *Cheumatopsyche brevilineata*, *Limnephilus hyalinus*, and *Stenopsyche angustata*).

Additional file 2: Table S2. Results of two-tailed tests of substitution saturation for each codon position of the 13 PCGs. The following abbreviations were used: index of substitution saturation (ISS), critical value of ISS supposing symmetrical cladogenesis (ISS.CSym.), critical value of ISS supposing asymmetrical cladogenesis (ISS.CAsym.).

Additional file 3: Figure S1. Maximum Likelihood phylogenetic tree based on 13 PCGs of 56 mitogenomes including outgroups species (*Phryganea cinerea*, *Phryganopsyche latipennis*, *Cheumatopsyche brevilineata*, *Limnephilus hyalinus*, and *Stenopsyche angustata*).

Additional file 4: Figure S2. Maximum Likelihood phylogenetic tree based on 13 PCGs + 2 rRNAs compared *A. conjugella* mitochondrial genome with the mitochondrial genomes of 507 Lepidoptera obtained from GenBank, representing 18 superfamilies and 42 families (Supplementary Table S1), including outgroups species (*Phryganea cinerea*,

Phryganopsyche latipennis, *Cheumatopsyche brevilineata*, *Limnephilus hyalinus*, and *Stenopsyche angustata*).

Acknowledgements

The Norwegian Institute for Bioeconomy Research (NIBIO) financed this work. The authors would like to thank Dr. Arne Hermansen for his interest in and support for the project. The authors would also like to thank Toril Sagen Eklo for collecting the materials of *A. conjugella*.

Authors' contributions

AE, SNM and HGE conceived and designed the study. GK collected the sample, and SNM and AVE analyzed the data. AE wrote the first draft of the manuscript, and AE and SNM wrote the draft manuscript with input from HGE, SBH, AVE, MHM, and GK. All authors have read, revised and approved the final manuscript.

Funding

This project was funded by The Norwegian Institute for Bioeconomy Research (NIBIO).

Availability of data and materials

A. conjugella. mitochondrial genome has been deposited in GenBank under accession: ON496993; Fig. 1 (https://github.com/Simo-N-Maduna/Mito-Phylo-genomics/tree/main/Mitophylogenomics_PartII_Lepidoptera). The 507 mitogenomes from the study were downloaded from GenBank. Their accession numbers and references are listed in Table S1. Other supporting results are included within the article and its additional files.

Declarations

Ethics approval and consent to participate

Our study was conducted as a component of a forecasting program for apple fruit moth attacks in Norway, which is being coordinated by the Norwegian Institute of Bioeconomy (NIBIO). NIBIO is authorized to collect apple fruit moth specimens.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 June 2023 Accepted: 14 December 2023

Published online: 02 January 2024

References

- van Nieuwerkerken E. Order lepidoptera linnaeus, Zootaxa 3148. Magnolia Press 1758; 2011.
- Regier JC, Zwick A, Cummings MP, Kawahara AY, Cho S, Weller S, Roe A, Baixeras J, Brown JW, Parr C. Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. *BMC Evol Biol.* 2009;9(1):1–21.
- Xu C, Pan Z-Q, Nie L, Hao J-S. The complete mitochondrial genome of *Issoria eugenia* (Lepidoptera: Nymphalidae: Heliconiinae). *Mitochondrial DNA Part B.* 2019;4(1):1662–3.
- Chen D-B, Zhang R-S, Jin X-D, Yang J, Li P, Liu Y-Q. First complete mitochondrial genome of *Rhodinia* species (Lepidoptera: Saturniidae): genome description and phylogenetic implication. *Bull Entomol Res.* 2022;112(2):243–52.
- Liu X, Qi M, Xu H, Wu Z, Hu L, Yang M, Li H. Nine Mitochondrial Genomes of the Pyraloidea and Their Phylogenetic Implications (Lepidoptera). *Insects.* 2021;12(11):1039.
- Sun Y, Huang H, Liu YD, Liu SS, Xia J, Zhang K, Geng J. Organization and phylogenetic relationships of the mitochondrial genomes of *Spireodonia retorta* and other lepidopteran insects. *Sci Rep.* 2021;11(1):2957.

7. Sun YX, Wang L, Wei GQ, Qian C, Dai LS, Sun Y, Abbas MN, Zhu BJ, Liu CL. Characterization of the complete mitochondrial genome of *Leucoma salicis* (Lepidoptera: Lymantriidae) and comparison with other lepidopteran insects. *Sci Rep*. 2016;6:39153.
8. Singh D, Kabiraj D, Sharma P, Chetia H, Mosahari PV, Neog K, Bora U. The mitochondrial genome of Muga silkworm (*Antheraea assamensis*) and its comparative analysis with other lepidopteran insects. *PLoS One*. 2017;12(11):e0188077.
9. Wei ZX, Sun G, Shiu JY, Fang Y, Shi QH. The complete mitochondrial genome sequence of *Dodona eugenes* (Lepidoptera: Riodinidae). *Mitochondrial DNA B Resour*. 2021;6(3):816–8.
10. Wahlberg N, Wheat CW. Genomic outposts serve the phylogenomic pioneers: designing novel nuclear markers for genomic DNA extractions of Lepidoptera. *Syst Biol*. 2008;57(2):231–42.
11. Cho S, Zwick A, Regier JC, Mitter C, Cummings MP, Yao J, Du Z, Zhao H, Kawahara AY, Weller S. Can deliberately incomplete gene sample augmentation improve a phylogeny estimate for the advanced moths and butterflies (Hexapoda: Lepidoptera)? *Syst Biol*. 2011;60(6):782–96.
12. Kawahara AY, Ohshima I, Kawakita A, Regier JC, Mitter C, Cummings MP, Davis DR, Wagner DL, De Prins J, Lopez-Vaamonde C. Increased gene sampling strengthens support for higher-level groups within leaf-mining moths and relatives (Lepidoptera: Gracillariidae). *BMC Evol Biol*. 2011;11(1):1–14.
13. Zwick A, Regier JC, Mitter C, Cummings MP. Increased gene sampling yields robust support for higher-level clades within Bombycoidea (Lepidoptera). *Syst Entomol*. 2011;36(1):31–43.
14. Regier JC, Mitter C, Zwick A, Bazinet AL, Cummings MP, Kawahara AY, Sohn J-C, Zwickl DJ, Cho S, Davis DR. A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS One*. 2013;8(3):e58568.
15. Regier JC, Mitter C, Davis DR, Harrison TL, SOHN JC, Cummings MP, Zwick A, Mitter KT. A molecular phylogeny and revised classification for the oldest ditrysian moth lineages (Lepidoptera: Tineoidea), with implications for ancestral feeding habits of the mega-diverse Ditrysia. *Syst Entomol*. 2015;40(2):409–32.
16. Sohn JC, Regier JC, Mitter C, Davis D, Landry JF, Zwick A, Cummings MP. A molecular phylogeny for Yponomeutoidea (Insecta, Lepidoptera, Ditrysia) and its implications for classification, biogeography and the evolution of host plant use. *Plos One*. 2013;8(1):e55066.
17. Kawahara AY, Breinholt JW. Phylogenomics provides strong evidence for relationships of butterflies and moths. *Proc Royal Society B: Biol Sci*. 2014;281(1788):20140970.
18. Breinholt JW, Earl C, Lemmon AR, Lemmon EM, Xiao L, Kawahara AY. Resolving relationships among the megadiverse butterflies and moths with a novel pipeline for anchored phylogenomics. *Syst Biol*. 2018;67(1):78–93.
19. Kawahara AY, Plotkin D, Espeland M, Meusemann K, Toussaint EF, Donath A, Gimmich F, Frandsen PB, Zwick A, Dos Reis M. Phylogenomics reveals the evolutionary timing and pattern of butterflies and moths. *Proc Natl Acad Sci*. 2019;116(45):22657–63.
20. Mayer C, Dietz L, Call E, Kukowka S, Martin S, Espeland M. Adding leaves to the Lepidoptera tree: capturing hundreds of nuclear genes from old museum specimens. *Syst Entomol*. 2021;46(3):649–71.
21. Rota J, Twort V, Chiochio A, Peña C, Wheat CW, Kaila L, Wahlberg N. The unresolved phylogenomic tree of butterflies and moths (Lepidoptera): Assessing the potential causes and consequences. *Syst Entomol*. 2022;47(4):531–50.
22. Mitter C, Davis DR, Cummings MP. Phylogeny and evolution of Lepidoptera. *Annu Rev Entomol*. 2017;62:265–83.
23. Jeong JS, Park JS, Sohn J-C, Kim MJ, Oh HK, Kim I. The first complete mitochondrial genome in the family Attevidae (*Atteva aurea*) of the order Lepidoptera. *Biodiversity Data J*. 2022;10:e89982.
24. Ahlberg O. Ronnbarsmalen, *Argyresthia conjugella* Zell. En redogörelse för undersökningar åren 1921–1926. Meddel Nr 324 från Centralanstalten för forskningsväsendet på jordbruksområdet 1927.
25. Agassiz D. British argyresthiinae and yponomeutinae. In: *Proceedings and Transactions of the British Entomological and Natural History Society*. 1987. p. 1987.
26. Kobro S, Sørøide L, Dønne E, Rafoss T, Jaastad G, Witzgall P. Masting of rowan *Sorbus aucuparia* L. and consequences for the apple fruit moth *Argyresthia conjugella* Zeller. *Population Ecology*. 2003;45(1):25–30.
27. Bengtsson M, Jaastad G, Knudsen G, Kobro S, Bäckman AC, Pettersson E, Witzgall P. Plant volatiles mediate attraction to host and non-host plant in apple fruit moth. *Argyresthia conjugella* *Entomologia Experimentalis et Applicata*. 2006;118(1):77–85.
28. Elameen A, Eiken HG, Floystad I, Knudsen G, Hagen SB. Monitoring of the Apple Fruit Moth: Detection of Genetic Variation and Structure Applying a Novel Multiplex Set of 19 STR Markers. *Molecules*. 2018;23(4):14.
29. Elameen A, Eiken HG, Knudsen GK. Genetic Diversity in Apple Fruit Moth Indicate Different Clusters in the Two Most Important Apple Growing Regions of Norway. *Diversity-Basel*. 2016;8(2):12.
30. Elameen A, Klutsch CFC, Floystad I, Knudsen GK, Tasin M, Hagen SB, Eiken HG. Large-scale genetic admixture suggests high dispersal in an insect pest, the apple fruit moth. *PLoS ONE*. 2020;15(8):24.
31. Zhang D-X, Szymura JM, Hewitt GM. Evolution and structural conservation of the control region of insect mitochondrial DNA. *J Mol Evol*. 1995;40:382–91.
32. Zhang D-X, Hewitt GM. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem Syst Ecol*. 1997;25(2):99–120.
33. Wu YP, Zhao JL, Su TJ, Li J, Yu F, Chesters D, Fan RJ, Chen MC, Wu CS, Zhu CD. The Complete Mitochondrial Genome of *Leucoptera malifoliella* Costa (Lepidoptera: Lyonetiidae). *DNA Cell Biol*. 2012;31(10):1508–22.
34. Wei SJ, Shi BC, Gong YJ, Li Q, Chen XX. Characterization of the Mitochondrial Genome of the Diamondback Moth *Plutella xylostella* (Lepidoptera: Plutellidae) and Phylogenetic Analysis of Advanced Moths and Butterflies. *DNA Cell Biol*. 2013;32(4):173–87.
35. van Asch B, Bliedch I, Pereira-Castro I, Rei FT, da Costa LT. The mitochondrial genome of *Prays oleae* (Insecta: Lepidoptera: Praydidae). *Mitochondrial DNA Part A*. 2016;27(3):2108–9.
36. Chen YH, Huang DY, Wang YL, Zhu CD, Hao JS. The complete mitochondrial genome of the endangered Apollo butterfly, *Parnassius apollo* (Lepidoptera: Papilionidae) and its comparison to other Papilionidae species. *Journal of Asia-Pacific Entomology*. 2014;17(4):663–71.
37. Lammermann K, Vogel H, Traut W. The mitochondrial genome of the Mediterranean flour moth, *Ephesia kuehniella* (Lepidoptera: Pyralidae), and identification of invading mitochondrial sequences (numts) in the W chromosome. *European Journal of Entomology*. 2016;113:482–8.
38. Bao L, Zhang YH, Gu X, Gao YF, Yu YB. The complete mitochondrial genome of *Eterusia aede* (Lepidoptera, Zygaenidae) and comparison with other zygaenid moths. *Genomics*. 2019;111(5):1043–52.
39. Park B, Hwang UW. The complete mitochondrial genome of the woodwasp *Euxiphidria potanini* (Hymenoptera, Xiphidriidae) and phylogenetic implications for symphytans. *Sci Rep*. 2022;12(1):17677.
40. Bessa MH, de Re FC, de Moura RD, Loreto EL, Robe LJ. Comparative mitogenomics of Drosophilidae and the evolution of the *Zygothrica* genus group (Diptera, Drosophilidae). *Genetica*. 2021;149(5–6):267–81.
41. Watanabe Y, Suematsu T, Ohtsuki T. Losing the stem-loop structure from metazoan mitochondrial tRNAs and co-evolution of interacting factors. *Front Genet*. 2014;5:109.
42. Wu QL, Cui WX, Du BZ, Gu Y, Wei SJ. The complete mitogenome of the turnip moth *Agrotis segetum* (Lepidoptera: Noctuidae). *Mitochondrial DNA*. 2014;25(5):345–7.
43. Wu QL, Cui WX, Wei SJ. Characterization of the complete mitochondrial genome of the black cutworm *Agrotis ipsilon* (Lepidoptera: Noctuidae). *Mitochondrial DNA*. 2015;26(1):139–40.
44. Liu TT, Li ZF. Phylogenetic and taxonomic study of the complete mitochondrial genome of *Spodoptera frugiperda*. *Mitochondrial DNA Part B-Resources*. 2019;4(2):2759–61.
45. Pan ZQ, Xu C, Nie L, Hao JS. The complete mitochondrial genome of the *Papilio machaon annae* Gistel (Lepidoptera: Papilionidae: Papilioninae). *Mitochondrial DNA Part B-Resources*. 2019;4(1):1945–6.
46. Ramírez-Ríos V, Franco-Sierra ND, Alvarez JC, Saldamando-Benjumea CI, Villanueva-Mejía DF. Mitochondrial genome characterization of *Tecia solanivora* (Lepidoptera: Gelechiidae) and its phylogenetic relationship with other lepidopteran insects. *Gene*. 2016;581(2):107–16.
47. Xin Z-Z, Liu Y, Zhang D-Z, Wang Z-F, Tang B-P, Zhang H-B, Zhou C-L, Chai X-Y, Liu Q-N. Comparative mitochondrial genome analysis of *Spilarctia subcarnea* and other noctuid insects. *Int J Biol Macromol*. 2018;107:121–8.

48. Liu QN, Chai XY, Bian DD, Zhou CL, Tang BP. The complete mitochondrial genome of *Plodia interpunctella* (Lepidoptera: Pyralidae) and comparison with other Pyraloidea insects. *Genome*. 2016;59(1):37–49.
49. Kabiraj D, Chetia H, Nath A, Sharma P, Mosahari PV, Singh D, Dutta P, Neog K, Bora U. Mitogenome-wise codon usage pattern from comparative analysis of the first mitogenome of *Blepharipa* sp. (Muga uzifly) with other Oestroid flies. *Sci Rep*. 2022;12(1):7028.
50. Yi JQ, Wu H, Liu JB, Li JH, Lu YL, Zhang YF, Cheng YJ, Guo Y, Li DS, An YX. Novel gene rearrangement in the mitochondrial genome of *Anastatus fulloi* (Hymenoptera Chalcidoidea) and phylogenetic implications for Chalcidoidea. *Sci Rep*. 2022;12(1):1351.
51. Chen SC, Zhao FH, Jiang HY, Hu X, Wang XQ. The complete mitochondrial genome of the bagworm from a tea plantation in China, *Eumeta variegata* (Lepidoptera: Psychidae). *Mitochondrial DNA Part B-Resources*. 2021;6(3):875–7.
52. Boore JL, Daehler LL, Brown WM. Complete sequence, gene arrangement, and genetic code of mitochondrial DNA of the cephalochordate *Branchiostoma floridae* (Amphioxus). *Mol Biol Evol*. 1999;16(3):410–8.
53. Wang YL, Peng CM, Yao QL, Shi QH, Hao JS. The complete mitochondrial genome of *Gonepteryx rhamni* (Lepidoptera: Pieridae: Coliadinae). *Mitochondrial DNA*. 2015;26(5):791–2.
54. Cao YQ, Ma CA, Chen JY, Yang DR. The complete mitochondrial genomes of two ghost moths, *Thitarodes renzhiensis* and *Thitarodes yunnanensis*: the ancestral gene arrangement in Lepidoptera. *BMC Genomics*. 2012;13:276.
55. Cameron SL, Whiting MF. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta*, (Insecta : Lepidoptera : Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene*. 2008;408(1–2):112–23.
56. Yin JA, Hong GY, Wang AM, Cao YZ, Wei ZJ. Mitochondrial genome of the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and comparison with other Lepidopterans. *Mitochondrial DNA*. 2010;21(5):160–9.
57. Zhou N, Dong YL, Qiao PP, Yang ZF. Complete mitogenomic structure and phylogenetic implications of the genus *Ostrinia* (Lepidoptera: Crambidae). *Insects*. 2020;11(4):232.
58. Liao C-Q, Yagi S, Chen L, Chen Q, Hirowatari T, Wang X, Wang M, Huang G-H. Higher-level phylogeny and evolutionary history of nonditrysians (Lepidoptera) inferred from mitochondrial genome sequences. *Zool J Linn Soc*. 2023;198(2):476–93.
59. Kyrki J. The Yponomeutoidea: a reassessment of the superfamily and its suprageneric groups (Lepidoptera). *Insect Syst Evol*. 1984;15(1):71–84.
60. Kyrki J. Tentative reclassification of holarctic Yponomeutoidea (Lepidoptera). *Nota Lepidopterologica*. 1990;13(1):28–42.
61. Wang ZH, Yao S, Zhu XY, Hao JS. The complete mitochondrial genome of *Pidorus atratus* (Lepidoptera: Zygaenoidea: Zygaenidae). *Mitochondrial DNA Part B-Res*. 2018;3(1):448–4.
62. Zhang XY, Tang L, Chen J, You P. The complete mitochondrial genome of *Amesia sanguiflua* (Lepidoptera, Zygaenidae). *Mitochondrial DNA Part B-Res*. 2020;5(1):988–9.
63. Kim MJ, Wang AR, Park JS, Kim I. Complete mitochondrial genomes of five skippers (Lepidoptera: Hesperioidea) and phylogenetic reconstruction of Lepidoptera. *Gene*. 2014;549(1):97–112.
64. Liu GQ, Bi GQ, Du QW, Zhao EZ, Yang JQ, Zhang Z, Shang EL. Complete mitochondrial genome of *Plodia interpunctella* (Lepidoptera: Pyralidae). *Mitochondrial DNA Part A*. 2016;27(6):4538–9.
65. Liu T, Wang S, Li H. Review of the genus *Argyresthia* Hübner, [1825] (Lepidoptera: Yponomeutoidea: Argyresthiidae) from China, with descriptions of forty-three new species. *Zootaxa*. 2017;4292(1):1–135.
66. Yang LL, Dai JJ, Gao QP, Yuan GZ, Liu J, Sun Y, Sun YX, Wang L, Qian C, Zhu BJ, et al. Characterization of the complete mitochondrial genome of *Orthaga olivacea* Warre (Lepidoptera Pyralidae) and comparison with other Lepidopteran insects. *Plos One*. 2020;15(3):e0227831.
67. Jeong SY, Park JS, Kim MJ, Kim S-S, Kim I. The complete mitochondrial genome of *Monopis longella* Walker, 1863 (Lepidoptera: Tineidae). *Mitochondrial DNA Part B*. 2021;6(8):2159–61.
68. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative. *Syst Biol*. 2006;55(4):539–52.
69. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2018;35(2):518–22.
70. Regier JC, Mitter C, Zwick A, et al. A large-scale, higher-level, molecular phylogenetic study of the insect order Lepidoptera (moths and butterflies). *PLoS One*. 2013;8(3):e58568.
71. Naser-Khdour S, Minh BQ, Zhang W, Stone EA, Lanfear R. The Prevalence and Impact of Model Violations in Phylogenetic Analysis. *Genome Biol Evol*. 2019;11(12):3341–52.
72. Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. *Bioinformatics*. 2016;32(19):3047–8.
73. Lindgreen S. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes*. 2012;5(1):1–7.
74. Schubert M, Lindgreen S, Orlando L. AdapterRemoval v2: rapid adapter trimming, identification, and read merging. *BMC Res Notes*. 2016;9(1):1–7.
75. Al-Nakeeb K, Petersen TN, Sicheritz-Pontén T. Norgal: extraction and de novo assembly of mitochondrial DNA from whole-genome sequencing data. *BMC Bioinformatics*. 2017;18:1–7.
76. Pribelski A, Antipov D, Meleshko D, Lapidus A, Korobeynikov A. Using SPAdes de novo assembler. *Curr Protoc Bioinformatics*. 2020;70(1):e102.
77. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2010;26(5):589–95.
78. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Subgroup GPP. The sequence alignment/map format and SAMtools. *Bioinformatics*. 2009;25(16):2078–9.
79. Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. *Bioinformatics*. 2016;32(2):292–4.
80. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics*. 2012;28(12):1647–9.
81. Donath A, Jühling F, Al-Arab M, Bernhart SH, Reinhardt F, Stadler PF, Middendorf M, Bernt M. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. *Nucleic Acids Res*. 2019;47(20):10543–52.
82. Benson D, Karsch-Mizrachi I, Lipman D, Ostell J, Wheeler D. GenBank. *Nucleic Acids Res*. 2005;1:33.
83. Laslett D, Canback B. ARWEN: a program to detect tRNA genes in metazoan mitochondrial nucleotide sequences. *Bioinformatics*. 2008;24(2):172–5.
84. Jühling F, Putz J, Bernt M, Donath A, Middendorf M, Florentz C, Stadler PF. Improved systematic tRNA gene annotation allows new insights into the evolution of mitochondrial tRNA structures and into the mechanisms of mitochondrial genome rearrangements. *Nucleic Acids Res*. 2012;40(7):2833–45.
85. Lowe TM, Chan PP. tRNAscan-SE On-line: integrating search and context for analysis of transfer RNA genes. *Nucleic Acids Res*. 2016;44(W1):W54–7.
86. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573–80.
87. Grant JR, Stothard P. The CGView Server: a comparative genomics tool for circular genomes. *Nucleic Acids Res*. 2008;36:W181–4.
88. Chan PP, Lin BY, Mak AJ, Lowe TM. tRNAscan-SE 2.0: improved detection and functional classification of transfer RNA genes. *Nucleic Acids Res*. 2021;49(16):9077–96.
89. Chamberlain S, Szöcs E, Foster Z. Taxize-taxonomic search and retrieval in R. *F1000Research*. 2013;2:191.
90. Tamura K, Stecher G, Kumar S. MEGA11: molecular evolutionary genetics analysis version 11. *Mol Biol Evol*. 2021;38(7):3022–7.
91. Perna NT, Kocher TD. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J Mol Evol*. 1995;41:353–8.
92. Ranwez V, Douzery EJP, Cambon C, Chantret N, Delsuc F. MACSE v2: toolkit for the alignment of coding sequences accounting for frameshifts and stop codons. *Mol Biol Evol*. 2018;35(10):2582–4.
93. Criscuolo A, Gribaldo S. BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evol Biol*. 2010;10:210.

94. Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol.* 2013;30(4):772–80.
95. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform.* 2019;20(4):1160–6.
96. Vecchi M, Bruneaux M. Concatpede: an R package to concatenate fasta sequences easily. 2021. <https://doi.org/10.5281/zenodo.5130603>
97. Xia X, Xie Z, Salemi M, Chen L, Wang Y. An index of substitution saturation and its application. *Mol Phylogenet Evol.* 2003;26:1–7.
98. Xia X. DAMBE7: New and Improved Tools for Data Analysis in Molecular Biology and Evolution. *Mol Biol Evol.* 2018;35(6):1550–2.
99. Xia X, Lemey P. Assessing substitution saturation with DAMBE. *The phylogenetic handbook: a practical approach to DNA and protein phylogeny.* 2009;2:615–630
100. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *J Mol Evol.* 1980;16(2):111–20.
101. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods.* 2017;14(6):587–+.
102. Chernomor O, von Haeseler A, Minh BQ. Terrace aware data structure for phylogenomic inference from supermatrices. *Syst Biol.* 2016;65(6):997–1008.
103. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol.* 2020;37(5):1530–4.
104. Soubrier J, Steel M, Lee MSY, Sarkissian CD, Guindon S, Ho SYW, Cooper A. The influence of rate heterogeneity among sites on the time dependence of molecular rates. *Mol Biol Evol.* 2012;29(11):3345–58.
105. Lanfear R, Calcott B, Kainer D, Mayer C, Stamatakis A. Selecting optimal partitioning schemes for phylogenomic datasets. *BMC Evol Biol.* 2014;14:82.
106. Minh BQ, Trifinopoulos J, Schrempf D, Schmidt H, Lanfear R. IQTREE version 2.0: tutorials and manual phylogenomic software by maximum likelihood. 2019. <http://www.iqtree.org>.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

