

The mitochondrial genomes of palaeopteran insects and insights into the early insect relationships

Nan Song^{1,*}, Xinxin Li¹, Xinming Yin¹, Xinghao Li¹, Jian Yin², Pengliang Pan²

¹College of Plant Protection, Henan Agricultural University, Zhengzhou 450002, China

²Xinyang Agriculture and Forestry University, Xinyang 464001, China

*Corresponding authors, e-mail: songnan@henau.edu.cn

Table S1. Taxa included in this study.

Item	Order	Suborder	Superfamily	Family	Species	Accession number	Voucher numbers for newly sequenced species	DNA conc. (ng/ul)
Outgroup	Collembola	Entomobryomorpha	Isotomoidae	Isotomidae	<i>Cryptopygus antarcticus</i>	EU016194	-	-
	Collembola	Poduromorpha	Poduroidea	Neanuridae	<i>Bilobella aurantiaca</i>	EU084034	-	-
	Diplura	Dicellurata	Japygoidea	Japygidae	<i>Occasjapyx japonicus</i>	JN990600	-	-
	Diplura	Dicellurata	Japygoidea	Japygidae	<i>Japyx solifugus</i>	AY771989	-	-
	Diplura	Rhabdura	Campodeoidea	Campodeidae	<i>Lepidocampa weberi</i>	JN990601	-	-
	Diplura	Rhabdura	Campodeoidea	Campodeidae	<i>Campodea fragilis</i>	DQ529236	-	-
	Diplura	Rhabdura	Campodeoidea	Campodeidae	<i>Campodea lubbocki</i>	DQ529237	-	-
	Archaeognatha	-	-	-	<i>Nesomachilis australica</i>	AY793551	-	-
	Archaeognatha	-	-	-	<i>Allopsontus</i> sp.	KJ754500	-	-
	Archaeognatha	-	-	-	<i>Songmachilis xinxiangensis</i>	JX308221	-	-
	Archaeognatha	-	-	-	<i>Petrobius brevistylis</i>	AY956355	-	-
	Archaeognatha	-	-	-	<i>Trigoniophthalmus</i>	EU016193	-	-
	Archaeognatha	-	-	-	<i>Pedetontus silvestrii</i>	EU621793	-	-
	Zygentoma	-	-	Nicoletiidae	<i>Atelura formicaria</i>	EU084035	-	-
	Zygentoma	-	-	Lepismatidae	<i>Thermobia</i> sp.	MK951657	EMHAU_1 6071107	270.03
	Zygentoma	-	-	Lepidotrichidae	<i>Tricholepidion gertschi</i>	AY191994	-	-
	Ephemeroptera	Pisciforma	-	Baetidae	<i>Cloeon dipterum</i>	MK951672	EMHAU_1 6071111	559.50
	Ephemeroptera	Pisciforma	-	Baetidae	<i>Baetis</i> sp.	GU936204	-	-
	Ephemeroptera	Pisciforma	-	Baetidae	<i>Alainites yixiani</i>	GU479735	-	-
	Ephemeroptera	Pannota	-	Caenidae	<i>Caenis</i> sp. YJ-2009	GQ502451	-	-
	Ephemeroptera	Pannota	-	Ephemerellidae	<i>Vietnamella dabieshanensis</i>	HM067837	-	-
	Ephemeroptera	Scaphodonta	-	Ephemeridae	<i>Ephemer</i> sp.	MK951659	EMHAU_1 6071112	23.73
	Ephemeroptera	Scaphodonta	-	Ephemeridae	<i>Ephemer</i> <i>orientalis</i>	NC_012645	-	-
	Ephemeroptera	Setisura	-	Heptageniidae	<i>Rhithrogena</i> sp.	MK951670	EMHAU_1 6071115	21.57
	Ephemeroptera	Setisura	-	Heptageniidae	<i>Parafronurus</i> sp.	MK951673	EMHAU_1 6082304	327.47

Ingroup	Ephemeroptera	Setisura	-	Heptageniidae	<i>Epeorus</i> sp.	MK951674	EMHAU_1 6082311	245.10
	Ephemeroptera	Setisura	-	Heptageniidae	<i>Paegniodes cupulatus</i>	HM004123	-	-
	Ephemeroptera	Setisura	-	Heptageniidae	<i>Epeorus</i> sp. JZ-2014	KJ493406	-	-
	Ephemeroptera	Setisura	-	Heptageniidae	<i>Epeorus</i> sp. MT-2014	KM244708	-	-
	Ephemeroptera	Setisura	-	Heptageniidae	<i>Parafronurus youi</i>	EU349015	-	-
	Ephemeroptera	Setisura	-	Isonychiidae	<i>Isonychia</i> sp.	MK951658	EMHAU_1 6071109	451.00
	Ephemeroptera	Setisura	-	Isonychiidae	<i>Isonychia ignota</i>	HM143892	-	-
	Ephemeroptera	Pisciforma	-	Siphonuridae	<i>Siphuriscus chinensis</i>	HQ875717	-	-
	Ephemeroptera	Pisciforma	-	Siphonuridae	<i>Siphonurus immanis</i>	FJ606783	-	-
	Odonata	Anisoptera	-	Aeshnidae	<i>Anax imperator</i>	KX161841	-	-
	Odonata	Anisoptera	-	Cordulegastridae	<i>Anotogaster sieboldii</i>	MK951663	EMHAU_1 6071526	247.77
	Odonata	Anisoptera	-	Corduliidae	<i>Somatochlora hineana</i>	MG594801	-	-
	Odonata	Anisoptera	-	Corduliidae	<i>Cordulia aenea</i>	JX963627	-	-
	Odonata	Anisoptera	-	Gomphidae	<i>Davidius lunatus</i>	NC_012644	-	-
	Odonata	Anisoptera	-	Gomphidae	<i>Ictinogomphus</i> sp.	KM244673	-	-
	Odonata	Anisoptera	-	Libellulidae	<i>Orthetrum sabina</i>	KU361234	-	-
	Odonata	Anisoptera	-	Libellulidae	<i>Orthetrum melania</i>	MK951662	EMHAU_1 6071517	28.50
	Odonata	Anisoptera	-	Libellulidae	<i>Orthetrum albistylum</i>	MK951665	EMHAU_1 6071520	110.80
	Odonata	Anisoptera	-	Libellulidae	<i>Sympetrum eroticum</i>	MK951664	EMHAU_1 6071528	187.87
	Odonata	Anisoptera	-	Libellulidae	<i>Brachythemis contaminata</i>	KM658172	-	-
	Odonata	Anisoptera	-	Libellulidae	<i>Hydrobasileus croceus</i>	KM244659	-	-
	Odonata	Anisozygoptera	-	Epiophlebiidae	<i>Epiophlebia superstes</i>	JX050223	-	-
	Odonata	Zygoptera	-	Calopterygidae	<i>Vestalis melania</i>	JX050224	-	-
	Odonata	Zygoptera	-	Calopterygidae	<i>Mnais tenuis</i>	MK951660	EMHAU_1 6071524	81.73
	Odonata	Zygoptera	-	Calopterygidae	<i>Atrocalopteryx atrata</i>	NC_027181	-	-
	Odonata	Zygoptera	-	Calopterygidae	<i>Atrocalopteryx melli</i>	MG011692	-	-
	Odonata	Zygoptera	-	Coenagrionidae	<i>Paracercion malayanum</i>	MK951669	EMHAU_1 5122908	11.27
	Odonata	Zygoptera	-	Coenagrionidae	<i>Agriocnemis femina</i>	MK951667	EMHAU_1 6071508	317.33
	Odonata	Zygoptera	-	Coenagrionidae	<i>Enallagma cyathigerum</i>	MF716899	-	-
	Odonata	Zygoptera	-	Coenagrionidae	<i>Ischnura elegans</i>	MK951668	EMHAU_1 6071502	202.60

Odonata	Zygoptera	-	Coenagrioni dae	<i>Ischnura pumilio</i>	KC878732	-	-
Odonata	Zygoptera	-	Euphaeidae	<i>Euphaea formosa</i>	HM126547	-	-
Odonata	Zygoptera	-	Euphaeidae	<i>Euphaea ornata</i>	NC_026059	-	-
Odonata	Zygoptera	-	Euphaeidae	<i>Euphaea yayeyamana</i>	KF718293	-	-
Odonata	Zygoptera	-	Megapodagri onidae	Mesopodagrion tibetanum	MK951671	EMHAU_1 6071523	249.90
Odonata	Zygoptera	-	Platycnemidi dae	Coeliccia cyanomelas	MK951666	EMHAU_1 6071505	1242.83
Odonata	Zygoptera	-	Platycnemidi dae	Platycnemis phyllopoda	MK951661	EMHAU_1 6071521	102.23
Odonata	Zygoptera	-	Platycnemidi dae	<i>Platycnemis foliacea</i>	KP233804	-	-
Odonata	Zygoptera	-	Pseudolestid ae	<i>Pseudolestes mirabilis</i>	FJ606784	-	-
Odonata	Zygoptera	-	Pseudostigm atidae	<i>Megaloprepus caerulatus</i>	KU958377	-	-
Plecoptera	-	Perloidea	Perlidae	<i>Acroneuria hainana</i>	KM199685	-	-
Mecoptera	-	-	Panorpidae	<i>Neopanorpa pulchra</i>	FJ169955	-	-
Diptera	Nematocera	Culicoidea	Culicidae	<i>Anopheles gambiae</i>	MG753657	-	-
Neuroptera	Hemerobiifo rmia	-	Chrysopidae	<i>Apochrysa matsumurae</i>	AP011624	-	-
Megalopter a	-	-	Corydalidae	<i>Acanthacorydalis orientalis</i>	KF840564	-	-
Siphonapter a	-	-	Vermipsyllid ae	<i>Dorcadia ioffi</i>	MF124314	-	-
Coleoptera	Polyphaga	Chrysomel oidea	Chrysomelid ae	<i>Acanthoscelides obtectus</i>	KX825864	-	-
Raphidiopte ra	-	-	Raphidiidae	<i>Mongoloraphidia harmandi</i>	FJ859902	-	-
Trichoptera	Integripalpia	Phryganeo idea	Phryganeida e	<i>Eubasilissa regina</i>	KF756943	-	-
Lepidoptera	Glossata	Papilionoi dea	Nymphalidae	<i>Abrota ganga</i>	KF590536	-	-
Orthoptera	Caelifera	Tetrigoi da	Tetrigidae	<i>Alulatettix yunnanensis</i>	JQ272702	-	-
Orthoptera	Caelifera	-	Acrididae	<i>Acrida cinerea</i>	GU344100	-	-
Hemiptera	Auchenorrh yncha	Cercopoid ea	Cercopidae	<i>Callitettix biformis</i>	JX844627	-	-
Hemiptera	Heteroptera	Coreoidea	Rhopalidae	<i>Aeschyntelus notatus</i>	EU427333	-	-
Grylloblatto dea	Grylloblatti dae	-	Grylloblattid ae	<i>Grylloblatta sculleni</i>	DQ241796	-	-
Phasmatode a	Anareolatae	-	Phasmatidae	<i>Extatosoma tiaratum</i>	AB642680	-	-
Mantophas matodea	-	-	Mantophasm atidae	<i>Sclerophasma paresisensis</i>	DQ241798	-	-
Mantodea	-	-	Hymenopodi dae	<i>Anaxarcha zhengi</i>	KU201320	-	-
Blattodea	-	Blaberoid ea	Ectobiidae	<i>Blattella bisignata</i>	JX233805	-	-
Blattodea	Blattoidea	Blattoidea	Termitoidae	<i>Aciculitermes aciculatus</i>	KY224548	-	-

Note: Bold indicates the species newly sequenced in this study. "-" indicates no application.

Table S2. (A) The partition schemes and best-fitting models selected by PartitionFinder for dataset of 85taxa_PCG1_{NT}2_{NT}3_{RY}RNA.

85taxa_PCG1 _{NT} 2 _{NT} 3 _{RY} RNA			
Subset	Partition names	Sites	Model
Partition 1	nad5, nad1, nad4l, nad4	4269	GTR+I+G
Partition 2	cob, cox3	1911	GTR+I+G
Partition 3	atp8, nad2, nad6	1659	GTR+I+G
Partition 4	nad3, atp6	1026	GTR+I+G
Partition 5	cox1	1527	GTR+I+G
Partition 6	cox2	678	GTR+I+G
Partition 7	rrna	2013	GTR+I+G
Partition 8	trna	2135	GTR+I+G

Abbreviations used in the BestModel: GTR, General-Time-Reversible model; I, invariant sites; G, discrete Gamma distribution.

Table S2. (B) The partition schemes and best-fitting models selected by PartitionFinder for dataset of 85taxa_PCG1_{NT}2_{NT}RNA.

85taxa_PCG1 _{NT} 2 _{NT} RNA			
Subset	Partition names	Sites	Model
Partition 1	nad4, nad4l, nad1, nad5	2846	GTR+I+G
Partition 2	nad3, cob, atp6	1438	GTR+I+G
Partition 3	atp8, nad2, nad6	1106	GTR+I+G
Partition 4	cox3, cox1	1538	GTR+I+G
Partition 5	cox2	452	GTR+I+G
Partition 6	rrna	2013	GTR+I+G
Partition 7	trna	2135	GTR+I+G

Table S2. (C) The partition schemes and best-fitting models selected by PartitionFinder for dataset of 79taxa_PCG1_{NT}2_{NT}3_{RY}RNA.

79taxa_PCG1 _{NT} 2 _{NT} 3 _{RY} RNA			
Subset	Partition names	Sites	Model
Partition 1	nad4l, nad4, nad1, nad5	4269	GTR+I+G
Partition 2	cob, cox3, atp6	2586	GTR+I+G
Partition 3	atp8, nad2, nad6	1659	GTR+I+G
Partition 4	nad3	351	GTR+I+G
Partition 5	cox1	1527	GTR+I+G
Partition 6	cox2	678	GTR+I+G
Partition 7	rrna	2013	GTR+I+G
Partition 8	trna	2135	GTR+I+G

Table S2. (D) The partition schemes and best-fitting models selected by PartitionFinder for dataset of 79taxa_PCG1_{NT}2_{NT}RNA.

79taxa_PCG1 _{NT} 2 _{NT} RNA			
Subset	Partition names	Sites	Model
Partition 1	nad4, nad4l, nad1, nad5	2846	GTR+I+G
Partition 2	nad3, cob, atp6	1438	GTR+I+G
Partition 3	atp8, nad2, nad6	1106	GTR+I+G
Partition 4	cox3, cox1	1538	GTR+I+G
Partition 5	cox2	452	GTR+I+G
Partition 6	rrna	2013	GTR+I+G
Partition 7	trna	2135	GTR+I+G

Table S3. Summary of the phylogenetic studies on the Palaeoptera problem.

References	Hypothesis supported	Samples included			Data type	Phylogenetic reconstruction method	Model used	Ref. number in this study
		Odo	Eph	Neo				
Zhang et al. 2008	Eph + (Odo + Neo)	1	1	7	mitogenome data	Maximum parsimony analyses of nucleotide and protein sequences of 12 mitochondrial PCGs (PAUP), ML analyses of protein sequences of 12 mitochondrial PCGs (PhyML), Bayesian analyses of nucleotide and protein sequences of 12 mitochondrial PCGs (MrBayes).	MtArt+I+G (ML), GTR+I+G for nucleotide and MtRev+I+G for protein (BI)	36
Comandi et al. 2009	Odo + Neo	1	0	6		Bayesian inference (MrBayes) of protein sequences of 13 PCGs.	MtPan+I	28
Lin et al. 2010	Odo + (Neo including Eph)	3	2	8		Maximum parsimony analyses of nucleotide and protein sequences of 13 PCGs (PAUP), Maximum likelihood analyses of nucleotide (RAxML) and protein (PhyML) sequences of 13 PCGs, Bayesian analyses of nucleotide and protein sequences of 13 PCGs (MrBayes).	GTRCAT+I+G (ML-nucleotide), MtArt+I+G (ML-protein), complex model (BI-nucleotide), MtRev+G+F (BI-protein)	31
Zhang et al. 2010	Odo + (Neo including Eph)	1	1	10		Bayesian inference using protein sequences of 13 PCGs (MrBayes).	-	79

Li et al. 2014	Odo + (Eph + Neo)	3	4	16		Bayesian (MrBayes) and maximum likelihood (RAxML) analyses based on nucleotide sequences of 13 PCGs (1st codon and 2nd codon positions).	GTR+I+G for both BI and ML analyses	70
Rutschmann et al. 2017	Eph + (Odo + Neo) or (Odo + Eph) + Neo	10	19	51		Bayesian (MrBayes) and maximum likelihood (RAxML) analyses based on protein sequences of 13 PCGs.	MTZOA+G+I and MTZOA+G+I+F both BI and ML analyses	68
Cai et al. 2018	Eph + (Odo + Neo)	27	20	16		Bayesian (MrBayes) and maximum likelihood (PhyML) Inferences using nucleotide sequences - of 13 PCGs.		27
Present study	(Odo + Eph) + Neo	31	18	20		Maximum likelihood (IQ-TREE) and Bayesian analyses (PhyloBayes) based on the nucleotide sequences of 37 mitochondrial genes.	GTR for ML, CAT-GTR for BI analyses	-
Misof et al. 2007	Odo + (Eph + Neo)	44	16	230	18S rRNA	Bayesian analyses (PHASE) based on 18S rRNA sequences.	complex RNA models	12
Ishiwata et al. 2010	(Odo + Eph) + Neo	2	2	38	three nuclear protein-coding genes	ML analyses (RAxML) and Bayesian inferences (MrBayes) based on three nuclear protein-coding gene sequences.	WAGCAT+F and WAG+G+F (ML), WAG+G (BI)	7
Simon et al. 2009	Odo + (Eph + Neo)	1	1	12	EST	ML analyses (RAxML) and Bayesian inferences (MrBayes) based on data from expressed sequence tag (125 genes, and 31,643 amino acid positions).	WAG (ML), WAG+I+G (BI)	26

Regier et al. 2010	(Odo + Eph) + Neo	2	2	6	Bayesian (MrBayes), maximum likelihood (GARLI) and parsimony (PAUP) analyses based on nucleotide and protein sequences from 62 single-copy nuclear protein-coding genes.	GTR+I+G for nucleotide and JTT+G for protein	8
--------------------	-------------------	---	---	---	--	--	---

					genome-scale data		
Misof et al. 2014	(Odo + Eph) + Neo	3	4	114	ML analyses (ExaML) based on nucleotide and amino acid sequences from 1478 single-copy nuclear genes generated from genomes and transcriptomes.	nucleotide applying a site-specific rate model, and amino acid applying a protein domain–based partitioning scheme to improve the biological realism of the applied evolutionary models	10

Blanke et al. 2012	(Odo + Eph) + Neo	8	4	19	morphological data	Bayesian (MrBayes), maximum likelihood (RAxML) and parsimony (TNT) analyses based on head characters.	Mk for BI analyses, Mkv for ML analyses	16
--------------------	-------------------	---	---	----	--------------------	---	---	----

Kjer et al. 2006	(Odo + Eph) + Neo	7	7	101	combining data of molecular and morphological characters	Bayesian (MrBayes) and parsimony (PAUP) analyses based on combined data including a fragment of the 28S (D1–D8) and complete sequences for the 18S, histone (H3), EF-1 α , COI, COII, the 12S and 16S plus the intervening tRNA, and 170 morphological characters.	GTR+I+G model for nucleotides, and MK model for the morphology	9
------------------	-------------------	---	---	-----	--	---	--	---

Note: Odo, Odonata; Eph, Ephemeroptera; Neo, Neoptera. "-" indicates that the model is not presented in the reference, or the item is not applied.

Figure legends for supplementary figures

Fig. S1 Phylogenetic tree based on maximum likelihood bootstrap analysis of the nucleotide sequence dataset of 85taxa_PCG1_{NT}2_{NT}3_{RY}RNA, under the GTR model. Values at nodes are bootstrap scores (Left: the values from dataset of 85taxa_PCG1_{NT}2_{NT}3_{RY}RNA, right: the values from dataset of 85taxa_PCG1_{NT}2_{NT}RNA). “--” indicates the node not being recovered by the dataset of 85taxa_PCG1_{NT}2_{NT}RNA. Scale bar represents substitutions/site. The meaning of color is as Fig. 2.

Fig. S2 Phylogenetic tree based on maximum likelihood bootstrap analysis of the nucleotide sequence dataset of 79taxa_PCG1_{NT}2_{NT}3_{RY}RNA, under the GTR model. Values at nodes are bootstrap scores (Left: the values from dataset of 79taxa_PCG1_{NT}2_{NT}3_{RY}RNA, right: the values from dataset of 79taxa_PCG1_{NT}2_{NT}RNA). “--” indicates the node not being recovered by the dataset of 79taxa_PCG1_{NT}2_{NT}RNA. Scale bar represents substitutions/site. The meaning of color is as Fig. 2.

Fig. S3 Phylogenetic tree based on Bayesian inference of the nucleotide sequence dataset of 79taxa_PCG1_{NT}2_{NT}3_{RY}RNA, under the CAT+GTR model. Values at nodes are Bayesian posterior probability support (Left: the values from dataset of 79taxa_PCG1_{NT}2_{NT}3_{RY}RNA, right: the values from dataset of 79taxa_PCG1_{NT}2_{NT}RNA). “--” indicates the node not being recovered by the dataset of 79taxa_PCG1_{NT}2_{NT}RNA. Scale bar represents substitutions/site. The meaning of color is as Fig. 2.





