Open Access

SCHOLARENA

SAJ Biotechnology

RESEARCH ARTICLE

ISSN: 2375-6713

Mitogenome Sequence of the Curved Back Grasshopper *Arcyptera meridionalis* and the Classification Status of its Related Subfamilies

Rong-Qian Sun[1,2], Xue-Mei Wu[1], Guo-Fang Jiang[1,*], Fang Hong[1] and Wen-Jing Chen[1]

[1]*College of Oceanology and Food Science, Quanzhou Normal University, Quanzhou 362000, China*

[2]*Chengdu Institute of Biology, Chinese Academy of Sciences, Chengdu 610213, China*

[*]**Corresponding Author:** Guo-Fang Jiang, College of Oceanology and Food Science, Quanzhou Normal University, Quanzhou 362000, China, E-mail: cnjgf1208@qztc.edu.cn

Citation: Rong-Qian Sun, Xue-Mei Wu, Guo-Fang Jiang, Fang Hong, Wen-Jing Chen et al. (2024) Mitogenome Sequence of the Curved Back Grasshopper Arcypera meridionalis and the Classification Status of its Related Subfamilies, SAJ Biotechnol 9: 105

Abstract

The complete mitogenome of the curved back grasshopper *Arcyptera (Pararcyptera) meridionalis* was determined. The complete mitogenome is 15,621 bp in length with A+T content 75.23%. All 37 genes are conserved in the position observed in that of other grasshoppers, and the orientation and gene order of these genes are identical to those found in the other analysed species Caelifera. Genes are closely assembled one after the other, leaving a total of 109 bp (excluding the A+T rich region) in intergenic spacers, ranging in size from 1 bp to 21 bp. There are a total of 44 overlapping nucleotides among six genes with overlapping range from 1 to 9 bp. Four types (ATG, ATC, ATT and ATA) of start codons were identified in the 13 protein-coding genes (PCGs), in which TAA and TAG were used in 12 PCGs as stop codons, except COII (T). The 13 PCGs are 11,200 bp in length, 22 tRNA genes are 1,475 bp, the length of two rRNA genes is 2,106 bp, and the length of A+T rich region is 777 bp. The sequences of the PCGs of 18 orthopteran insects were used for phylogenetic analysis by using ML and BI methods. The results indicated that neither of the phylogenetic analyses supported the monophyletic status of the subfamily Gomphocerinae and he family Arcypteridae in the current international taxonomy.

Keywords: Complete mitogenome; *Arcyptera meridionalis;* Gomphocerinae; Arcypteridae; Phylogenetics

Introduction

The mitogenome varies greatly among insect species [1]. The mitogenome is associated with aging, apoptosis, energy metabolism, and disease, and has its own genetic material [2]. Insect mitogenomes are the most widely used genetic markers in insect evolution and population genetics studies. Their genomes are based on their maternal inheritance, lack of recombination, loss of introns and rapid evolutionary speed. It has been widely used as an information molecular marker to identify cryptic species, reconstruct phylogenetic trees and phylogeographic patterns, and infer the molecular evolution of different animal taxa, playing a key role in phylogenetic reconstruction, comparative evolutionary genomics, population genetics, phylogeography, and molecular evolution studies [3-9]. Insect mitogenome contains 13 protein-coding genes, 22 transfer RNA genes, 2 ribosomal RNA genes and standard groups rich in A + T region. At the same time, mitogenomes are becoming more and more popular among scientists in advanced phylogenetic analysis, as it can provide better resolution of deep relationships for single or multi-gene analysis and can be relatively easy to sequence the feasibility of the entire genome [10-12].

Orthoptera, is the only known taxon with significantly enlarged genomes in the class Insecta, and the only group of invertebrates with whole genome size larger than 10 Gb. Currently, there are more than 29,500 living species [13]. Some families of Orthoptera, such as Tettigonidae [30] and Locidae [31], have been found to have a large number of traditional paraphyletic taxa. Orthoptera have diversified, occupying all possible similar terrestrial habitats except the poles, and have collectively evolved similar external morphological characteristics in phylogenetically distinct lineages. These convergent morphological features are highly convincing, and taxonomists use them to formalize tribes, subfamilies and even families, which play an important role in global ecosystems [32]. In the Orthoptera Species File (OSF) online system, Orthoptera is divided into two suborders: Ensifera and Caelifera, in which Acridoidea is the largest superfamily in Caelifera [33-34].

With the maturity of sequencing technologies and the application of universal primers for mitochondrial genes [9, 13], 147 orthopteroid species have now been determined to date. *Locusta migratoria migratoioides* is the first species of Orthoptera [14] with a mitogenome sequenced, and its transposition of the ^{tRNA}Asp—^{tRNA}Lys gene between the *COX 2* and *ATP8* genes is the most prominent feature compared with other insects. Locusts are important pests in agriculture and animal husbandry all over the world, some of which have important economic value. Up to now, 61 species of grasshoppers have been included in the NCBI website in March 2023.

According to the Orthoptera Species File Version 5.0/5.0 (http://Orthoptera. SpeciesFile.org)，*Arcyptera (Pararcyptera) meridionalis* Ikonnikov, 1911 belongs to the subgenus *Pararcyptera* of the genus *Arcyptera* in the subfamily Gomphocerinae of Acrididae. It is mainly distributed in China, Mongolia and Soviet Union. In order to better understand the diversity and evolution of locusts, we sequenced and annotated the *Arcyptera meridionalis* mitogenome and discussed the phylogeny of its related subfamilies.

Materials and Methods

The specimens of *A. meridionalis* were collected from the Heihe Grand Canyon (3420 m above sea level) in Qilian County, Qinghai Province in July, 2016, and stored in 100% ethanol. The total genomic DNA was extracted from the muscle of the specimen's femora using the standard proteinase K and genomic DNA purification kit (Promega, Beijing, China) according to the manufacturer's instructions [35]. At present, *16S rRNA*, *18S rRNA*, *COI* and *COII* genes are used to construct phylogenetic and evolutionary relationship for locust classification [36]. The sequences of the above genes are highly conserved, moderate in size and easy to obtain through cloning sequencing, so they are used for systematic classification and evolutionary analysis. The principle is to distinguish the interspecific relationships according to the highly conserved regions of genes between species, re-

flecting the relatedness between species and the difference regions of sequences [37]. We first amplified partial sequences of *COI*, *Cyt b*, *12S RNA* and *16S RNA* using primers at our laboratory, which were synthesized by Sangon Biotechnology Company in Guangzhou. Gene sequences of related species were found on the NCBI website, and multiple pairs of primers were designed for PCR amplification (see Table S1). Normal PCRs were performed in a 25 μL reaction mixture consisting of 16.3 μL ddH_2O, 2.5 μL 10×PCR Buffer, 2 μL 25 mmol/L MgCl2, 2 μL 2.5 mmol/LdNTP, 1 μL DNA templates, 0.5 μL Primer F, 0.5 μL Primer R, 0.2 μL r-Taq DNA polymerase. A total of 30 cycles were run, each consisting of 94℃ denaturation for 30s, 38-55℃ annealing (according to primer setting) for 30s, 72℃ extension for 30-90s (According to the length of the template, the extension of 500 bp was prescribed for 30s), 94℃ pre-denaturation for 5min before the formal cycle, 72℃ extension for 10 min after the cycle is completed, and 4℃ preservation. PCR products were detected by 1% agarose gel electrophoresis. PCR products with clear and single bands detected by AGAR gel electrophoresis were packed together with 10 uL upper and downstream primers and sent to Sangon Biotechnology Company (Shanghai, China) for sequencing.

Orthoptera is recognized as a monophyletic group in the international classification, which also verifies the suitability of the data set and tree construction method. At present, most studies usually select the first and second sites of 13 PCGs for phylogenetic analysis. However, the phylogenetic signal strength of PCG and rRNA data sets in Orthoptera have been tested [38].

The complete mitogenomes of related species were downloaded from the NCBI database. The software Chromas ( ) was used to view the peak map of sequencing, SeqMan ( ) was used to spliced multiple sequencing sequences, and finally compared with related species to determine the location of each gene sequence. MEGA11 ( ) was used to translate protein sequences and analyze their sequence information.

Table 1: Mitogenomes of some grasshoppers used in this study

Subfamily	Tribe	Species	Accession number	References
Gomphocerinae	Phlaeobini	*Phlaeoba tenebrosa*	NC_029150	[28]
		Phlaeoba albonema	EU370925	[18]
		Phlaeoba infumata	NC_031506	[29]
	Ochrilidiini	*Gonista bicolor*	NC_029205	[27]
	Chrysochraontini	*Euchorthippus fusigeniculatus*	HM583652	[22]
	Gomphocerini	*Gomphocerippus rufus*	GU294759	
		Gomphocerus licenti	GQ180102	[24]
		Gomphocerus sibiricus	NC_021103	
		Gomphocerus sibiricus tibetanus	NC_015478	[25]
	Orinhippini	*Chorthippus chinensisOrinhippus tibetanus*	EU029161NC_023467.1	[16]
Oedipodinae	ArcypteriniParapleurini	*Arcyptera coreanaArcyptera meridionalisCeracris kiangsuCeracris versicolor*	NC_013805This studyGU270284.1NC_025285	[15][17][17]

Pyrgomorphinae	Atractomorphini	*Atractomorpha sinensis*	EU263919	[24]?
Thrinchinae	Thrinchini	*Thrinchus schrenkii*	GU181288	[18]
		Filchnerella beicki	NC_024923	[23]

In this study, 13 PCGs from 18 species were used to explore the taxonomic status of the family Gomphoceridae. Among them, the outgroups consisted of three species, namely *At. sinensis* of Pyrgomorphidae, *Thrinchus schrenkii* and *Filchnerella beicki* of Pamphagidae. Combined with the mitogenome sequences of 17 species (including outgroups), the mitogenome of *A. meridionalis* and the phylogenetic relationships of Arcypteridae and Gomphoceridae were discussed. The nucleotide sequences of 13 protein-coding genes in the 18 grasshopper mitogenomes were clipped, and the constructed protein-coding gene dataset was used for phylogenetic analysis.

Results and Discussion

Mitogenome Characteristics

Genome Structure

The complete mitogenome sequence of *A. meridionalis* was submitted to the NCBI, and the sequence accession number was MF997490. The total length of the splice annotated mitogenome is 15,621 bp, which, like most metazoan mitogenomes, contains 37 typical genes (13 protein-coding genes, *22 tRNAs*, and *2 rRNAs*) and A+T rich region (non-coding region between $tRNA^{Ile}$ and *12S rRNA*)[37] as shown in Table 2. They share genes with most species in the suborder Caelifera. There are overlapping genes in mitogenomes of *A. meridionalis* and other metazoans [10].

Table 2: Annotation of the complete mitogenome of *A. meridionalis*

Gene	CodingStrand	Position	Size (bp)	IntegenicNucleotides
$tRNA^{Ile}$	J	24473	67	
$tRNA^{Gln}$	N	71-138	69	3
$tRNA^{Met}$	J	142-210	69	3
ND2	J	211-1241	1032	0
$tRNA^{Trp}$	J	1233-1300	68	-9
$tRNA^{Cys}$	N	1293-1355	64	-8
$tRNA^{Tyr}$	N	1368-1434	67	12
COI	J	1427-2971	1545	-8
$tRNA^{Leu(UUR)}$	J	2967-3032	66	-5
COII	J	3034-3715	682	1
$tRNA^{Asp}$	J	3716-3780	65	0
$tRNA^{Lys}$	J	3783-3853	71	2
ATP8	J	3868-4029	162	14

ATP6	J	4023-4700	678	-7
COIII	J	4705-5496	792	4
$tRNA^{Gly}$	J	5499-5564	67	2
ND3	J	5565-5918	354	0
$tRNA^{Ala}$	J	5919-5984	66	0
$tRNA^{Arg}$	J	5988-6054	67	3
$tRNA^{Asn}$	J	6057-6124	68	2
$tRNA^{Ser(AGN)}$	J	6125-6191	67	0
$tRNA^{Glu}$	J	6192-6257	66	0
$tRNA^{Phe}$	N	6258-6321	64	0
ND5	N	6322-8040	1719	0
$tRNA^{His}$	N	8056-8120	65	15
ND4	N	8125-9459	1335	4
ND4L	N	9453-9746	294	-7
$tRNA^{Thr}$	J	9749-9817	69	2
$tRNA^{Pro}$	N	9818-9882	65	0
ND6	J	9885-10406	522	2
Cyt b	J	10415-11554	1140	8
$tRNA^{Ser(UCN)}$	J	11563-11632	70	8
ND1	N	11654-12598	945	21
$tRNA^{Leu(CUN)}$	N	12602-12667	66	3
lrRNA	N	12668-13979	1312	0
$tRNA^{Val}$	N	13980-14050	71	0
srRNA	N	14051-14844	794	0
A+T-rich		14845-15621	777	0

Note: J and N refer to the majority and minority strand, respectively.

The order of 37 genes in the *A. meridionalis* mitogenome is same as that of the 37 genes in the *C. versicoloris* mitogenome, is a typical circular DNA molecule. The localization and transcription directions of 37 genes and A+T rich regions are shown in Figure 1.

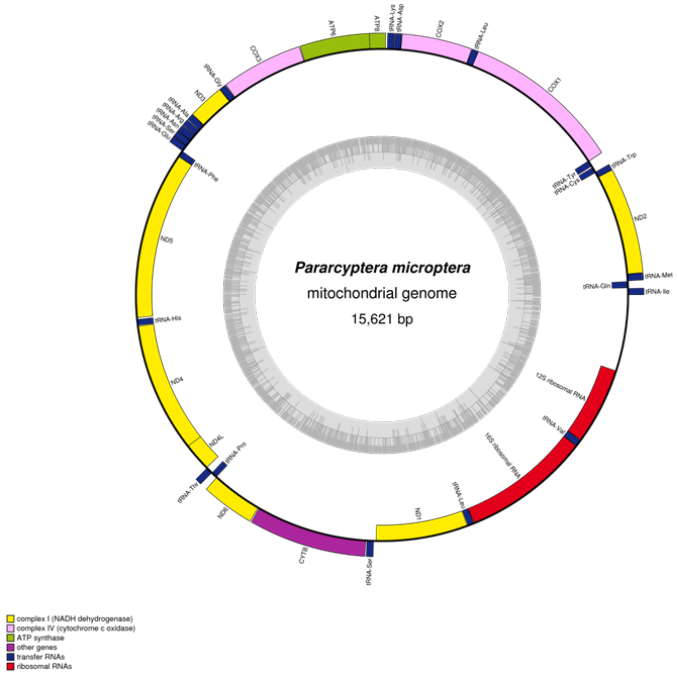

Figure 1: Gene map of the *Arcyptera meridionalis* mitogenome

Base Composition Characteristics

The base composition of mitogenome in *A. meridionalis* is shown in Table 3. It can be seen from Table 3 that the whole A+T content is 75.23%, indicating an obvious AT bias (42.82% of A, 32.4% of T, 14.19% of C, 10.56% of G). Meanwhile, we also calculated the content of G+C, which is 24.77% in *A. meridionalis*. The contents of PCGs, tRNAs, rRNAs and AT in control areas were 74.66%, 72.54%, 75.97% and 85.33%, respectively. The region with the highest content of A+T is the codon site 3 of the J-chain protein coding gene. Nucleotide composition differences were measured using AT and GC skew (AT skew = (A-t)/(A + T), GC skew = (g-c)/(G + C)) [10]. The mitogenome (AT skew = 0.14, GC skew = -0.15) was skewed towards A and C (Table 3).

Table 3: Composition and skewness of the *A. meridionalis* mitogenome

Feature	Size	%T	%C	%A	%G	%A+T	AT Skew	GC Skew
Whole genome	15621	32.41	14.19	42.82	10.56	75.23	0.14	-0.15
PCGs	11200	42.65	12.54	32.01	12.79	74.66	-0.14	0.01
First codon position	3734	35.94	11.81	32.94	19.31	68.88	-0.04	0.24
Second codon position	3733	45.70	20.20	19.77	14.33	65.47	-0.40	-0.17
Third codon position	3733	46.32	5.63	43.32	4.74	89.63	-0.03	-0.09
PCGs J-strand	6907	36.67	14.56	36.83	11.93	73.51	0.002	-0.01
First codon position	2303	29.44	13.89	36.86	19.80	66.30	0.11	0.18
Second codon position	2302	44.18	21.94	20.37	13.51	64.55	-0.37	-0.24
Third codon position	2302	36.40	7.86	53.26	2.48	89.66	0.19	-0.52
PCGs N-strand	4293	52.27	9.29	24.25	14.19	76.52	-0.37	0.21
First codon position	1431	46.40	8.46	26.62	18.52	73.03	-0.27	0.37

Second codon position	1431	48.15	17.40	18.80	15.65	66.95	-0.44	-0.05
Third codon position	1431	62.26	2.03	27.32	8.39	89.59	-0.39	0.61
tRNA genes	1475	33.02	14.64	39.53	12.81	72.54	0.10	-0.07
tRNA genes J-strand	945	33.54	12.59	40.11	13.76	73.65	0.09	0.04
tRNA genes N-strand	530	32.08	18.30	38.49	11.13	70.57	0.09	-0.24
rRNA genes	2106	32.24	14.86	43.73	9.16	75.97	15.13	-0.24
lrRNA	1312	32,93	13.87	44.82	8.38	77.74	15.29	-0.25
srRNA	794	31.11	16.50	41.94	10.45	73.05	0.15	-0.22
Control region	777	39.12	8.75	46.20	5.92	85.33	0.08	-0.19

Codon Usage of Protein-Coding Genes

The total length of 13 PCGs of *A. meridionalis* was 11,200 bp, of which nine (*ND2, COI, COII, ATP8, ATP6, COIII, ND3, ND6* and *CYTb*) were distributed on the main chain (J chain), and the rest (*ND5, ND4, ND4L*, and *ND1*) were distributed on a few strands (N chain).

The initial codon and termination codon of 13 PCGs in the *A. meridionalis* mitogenome are shown in Table 4. Except *COI*, the initial codons of *ATP8* genes prefer ATC, *ND3* and *ND5* genes prefer ATT, and *ND1* genes are ATA, the other 8 PCGs prefer ATG. By analyzing the mitogenomes of insects that had been detected and uploaded to NCBI, it was found that most insects with ATN as the initial codon of *COI* gene had ATC as the initial codon. In addition, insects with the very few TTA, CCG, TTG, CGA, TCG, and CTG as initial codons are in the minority [40]. In terms of termination codon, complete TAA or TAG was found to be used as termination codon in all 12 PCGs, and only incomplete T-- was found in *COII* gene.

After removing the termination codons of 13 PCGs in *A. meridionalis*, MEGA11 was used to make a statistical analysis of their usage and relative synonymous codon usage (RSCU) respectively (Table 4). If the RSCU value is greater than 1, it indicates that the actual frequency of occurrence of the codon is higher than the expected value, otherwise, it indicates that the actual occurrence of the codon is lower than the expected value[41].The Table 4 showed that the use of degenerate codons is biased towards the position of the third codon, where GS and CS are 2-4 times more common than AS and TS. The most commonly used codons in *A. meridionalis* are AT-rich codons, such as TTA, ATT, ATA, TTT, AAT, and TAT.

Table 4: Codon usage of the mitochondrial PCGs of *A. meridionalis*

Amino acid	Codon	N	RSCU	N+	RSCU	N-	RSCU
Phe (F)	UUU	302	1.71	136	1.5	166	1.93
	UUC	51	0.29	45	0.5	6	0.07
Leu (L)	UUA	356	4.15	191	4.08	165	4.23
	UUG(L)	45	0.52	9	0.19	36	0.92
	CUU(L)	54	0.63	25	0.53	29	0.74
	CUC(L)	4	0.05	4	0.09	0	0
	CUA(L)	51	0.59	48	1.02	3	0.08
	CUG(L)	5	0.06	4	0.09	1	0.03
Ile (I)	AUU(I)	346	1.77	220	1.69	126	1.92

	AUC(I)	45	0.23	40	0.31	5	0.08
Met (M)	AUA(M)	232	1.74	175	1.87	57	1.43
	AUG(M)	35	0.26	12	0.13	23	0.57
Val (V)	GUU(V)	106	2.36	50	1.85	56	3.11
	GUC(V)	3	0.07	1	0.04	2	0.11
	GUA(V)	68	1.51	56	2.07	12	0.67
	GUG(V)	3	0.07	1	0.04	2	0.11
Ser (S)	UCU(S)	126	2.61	26	1.03	100	4.35
	UCC(S)	7	0.15	6	0.24	1	0.04
	UCA(S)	129	2.67	103	4.08	26	1.13
	UCG(S)	2	0.04	2	0.08	0	0
Pro (P)	CCU(P)	58	1.74	29	1.21	29	3.14
	CCC(P)	9	0.27	8	0.33	1	0.11
	CCA(P)	63	1.89	56	2.33	7	0.76
	CCG(P)	3	0.09	3	0.13	0	0
Thr (T)	ACU(T)	55	1.15	26	0.68	29	3.05
	ACC(T)	11	0.23	10	0.26	1	0.11
	ACA(T)	121	2.53	115	3.01	6	0.63
	ACG(T)	4	0.08	2	0.05	2	0.21
Ala (A)	GCU(A)	73	1.76	34	1.14	39	3.32
	GCC(A)	7	0.17	5	0.17	2	0.17
	GCA(A)	84	2.02	79	2.66	5	0.43
	GCG(A)	2	0.05	1	0.03	1	0.09
Tyr (Y)	UAU(Y)	147	1.74	51	1.48	96	1.92
	UAC(Y)	22	0.26	18	0.52	4	0.08
Stop (*)	UAA(*)	10	1.67	8	2	2	1
	UAG(*)	2	0.33	0	0	2	1
His (H)	CAU(H)	56	1.56	41	1.46	15	1.88
	CAC(H)	16	0.44	15	0.54	1	0.13
Gln (Q)	CAA(Q)	56	1.67	50	1.92	6	0.8
	CAG(Q)	11	0.33	2	0.08	9	1.2
Asn (N)	AAU(N)	144	1.75	101	1.68	43	1.91
	AAC(N)	21	0.25	19	0.32	2	0.09
Lys (K)	AAA(K)	69	1.47	57	1.78	12	0.8
	AAG(K)	25	0.53	7	0.22	18	1.2
Asp (D)	GAU(D)	73	1.82	41	1.71	32	2

	GAC(D)	7	0.17	7	0.29	0	0
Glu (E)	GAA(E)	69	1.75	49	1.88	20	1.48
	GAG(E)	10	0.25	3	0.12	7	0.52
Cys (C)	UGU(C)	41	1.95	11	1.83	30	2
	UGC(C)	1	0.05	1	0.17	0	0
Trp (W)	UGA(W)	91	1.82	70	2	21	1.4
	UGG(W)	9	0.18	0	0	9	0.6
Arg (R)	CGU(R)	20	1.45	7	0.8	13	2.6
	CGC(R)	0	0	0	0	0	0
	CGA(R)	33	2.4	27	3.09	6	1.2
	CGG(R)	2	0.15	1	0.11	1	0.2
Ser (S)	AGU(S)	29	0.6	11	0.44	18	0.78
	AGC(S)	2	0.04	1	0.04	1	0.04
	AGA(S)	86	1.78	52	2.06	34	1.48
	AGG(S)	5	0.1	1	0.04	4	0.17
Gly (G)	GGU(G)	99	1.83	29	0.9	70	3.22
	GGC(G)	4	0.07	1	0.03	3	0.14
	GGA(G)	99	1.83	90	2.79	9	0.41
	GGG(G)	14	0.26	9	0.28	5	0.23

N: Total number in all PCGs, N+: total number in J-strand, total number in N-strand, RSCU: relative synonymous codon usage. Values in bold type stand for the most commonly used codon for amino acid.

In addition, among all the codons of PCGs in the *A. meridionalis* mitogenome of, five codons rich in A and T, UUA, AUU, UUU, AUA and UAU, were used most frequently and with the highest frequency. RSCU table showed that among the total number of codons, their frequency was as high as 37.2%. However, in the codon rich in G and C, the actual use times (N) is the least and the RSCU value is the lowest, such as the N and RSCU value of GCG is only 2 and 0.05 respectively (Table 4).

After the removal of the termination codon, a total of 3,721 codons were found in 13 PCGs. Based on the base composition of mitogenome (Table 3), it is found that the A+T content is very high, which has a great influence on the codon usage of protein--coding genes and the composition of amino acids. RSCU analysis shows that the number of synonymous codons ending in base A or T is more than other synonymous codons. For example, in the synonymous codon encoding phenylalanine, it was found that the TTC codon was used only 51 times with the RSCU of 0.29, but the TTT codon was used 302 times with the RSCU of 1.71. By analyzing the statistical data in the RSCU table (Table 4), it can be found that there is a strong bias in the codon use of mitochondrial PCGs of *A. meridionalis.*

In this study, detailed statistics were made on the amino acids encoded by the mitogenome protein of *A. meridionalis* (Table 5). The data showed that the contents of leucine (13.84%), isoleucine (10.51%), serine (10.37%) and phenobarbital (9.49%) were much higher than those of other amino acids. The sum of amino acids accounted for 44.21% of the total.

Table 5: Amino acid composition of the *A. meridionalis* mitogenome

Amino acid		Number	Content（%）
Phe	Phenylalanine	353	9.49
Leu	Leucine	515	13.84
Ile	Isoleucine	391	10.51
Met	Methionine	267	7.18
Val	Valine	180	4.84
Ser	Serine	386	10.37
Pro	Proline	133	3.57
Thr	Threonine	191	5.13
Ala	Alanine	166	4.46
Tyr	Tyrosine	169	4.54
His	Histidine	72	1.93
Gln	Glutamine	67	1.80
Leu	leucine	165	4.43
Lys	Lysine	94	2.53
Asp	Aspartic acid	80	2.15
Glu	Glutamic acid	79	2.12
Cys	Cysteine	42	1.13
Trp	Tryptophan	100	2.69
Arg	Arginine	55	1.48
Gly	Glycine	216	5.80

Among the 18 locust species, 13 protein-coding genes mostly started with the typical ATN (ATA, ATC, ATG, and ATT) codon as their initiation codon. In addition, in 17 species of grasshoppers, except for some species with TA-, T- incomplete codon as the termination codon, most species with the usual TAA and TAG as the termination codon.

Transfer RNA (tRNA) Genes

Based on the mitogenome sequence of *Ceracris versicolor*, the locations of 22 tRNA genes of *A. meridionalis* were identified and labeled by MEGA11 software. The data showed that the length of the 22 tRNA genes in the mitogenome of *A. meridionalis* ranged from 64 to 71 bp (see Table 4). In addition, special anticodon wasn't found in mitogenome anticodon, which is consistent with most orthoptera insects such as *C. versicolor*.

Ribosome RNA (rRNA) Genes

Two rRNA genes (lrRNA and srRNA) were found in the mitogenome of *A. meridionalis*, located between $tRNA^{Leu(CUN)}$ and $tRNA^{Val}$, and between $tRNA^{Val}$ and A+T rich regions, respectively, as is the case with most metazoan mitogenomes. Both rRNAs are encoded on a few strands (N strands), where the length of lr RNA is 1312 bp and the length of srRNA is 794 bp. And lr RNA and sr RNA base composition as well as wide wings stooped locust base composition characteristics, the two bases for A

and T have serious bias, A + T content in its base composition percentage were 77.74% and 73.05%, respectively.

A+T Rich Area (D-loop)

The uneven length of A+T rich region (D-loop) is the main reason for the length difference of insect mitogenome sequence, and the length difference of A+T rich region is caused by different number of repeating units. In addition, the A + T rich region contains control elements associated with animal gene replication and transcription[42][43]. Studies have shown that the measured length of A+T rich region in insect mitogenome ranges from 70 bp[44] to 4061 bp[39][45].

The A + T rich area was analyzed by comparison with *C. versicolor*[46]. In the mitogenome of *A. meridionalis*, the A+T rich region is located in the same position as the mitochondrial genes of most insects, between srRNA and tRNAIle genes. Like the mitogenome of arthropods, the A+T rich region of the species in this study also shows high conservation. In the mitogenome of *A. meridionalis*, the length of the A+T non-coding region was 777 bp, the nucleotide composition was 46.20%A, 39.12%T, 5.92G and 8.75%C respectively, and the A+T deviation was 85.33%. In the A + T rich area, the AT slope was 0.08, and the average GC slope was -0.19, indicating that the mitogenome had a significant bias against AT.

Comparision of Mitogenomes in some Grasshoppers

Comparison of Mitogenomes Structure and Sequences between *A. meridionalis* and *C. versicolor*

In order to better compare the structure and gene sequences of the mitogenome of *A. meridionalis* in this study, we selected and downloaded the mitogenome of *C. versicolor* from GenBank (NCBI).

The length of mitogenome in *A. meridionalis* and *C. versicolor* was 15,621 bp and 15,616 bp respectively. Even though the two species are different in sequence length, their genetic composition both include 37 genes and 1 A+T rich region. Among the 37 mitochondrial genes, there are 14 genes encoded on the N chain (a few chains), among which 4 encoding protein genes (ND1, ND4, ND4L and ND5), 8 transporter RNA (tRNA) genes (Gln, Cys, Tyr, Phe, His, Pro, Leu$^{(CUN)}$, and Val) and 2 ribose (rRNA) genes (16S rRNA and 12S rRNA). The remaining 23 genes are all encoded on the J chain.

The mitogenome sequence of *A. meridionalis* was consistent with that of most metazoans (Figure 1). There were exactly the same in the sequence and transcription direction of 37 genes in *A. meridionalis* and *C. versicolor*. In the mitogenome sequences, both of them have a common feature --KD translocation phenomenon, that is, DK arrangement (tRNA$^{Asp\ (D)}$ gene in front of tRNA$^{lys\ (K)}$ gene). Apart from that this difference, there were in exactly the same order in the rest of the genes. However, the typical arthropod mitogenome shows KD sequence in gene arrangement, such as that of the mitogenome of Leptodonta[44].

Genomic Gene Spacer and Overlap Area

There are a total of 18 gene spacer regions in the mitogenome of *A. meridionalis*. The length of the spacer regions varies from 1 to 21 bp, and the total length of the gene spacer sequences is 109 bp. The longest gene spacer was found between tRNA$^{Ser(UCN)}$ and ND1, with a sequence length of 21 bp. The characteristic of large gene spacing between tRNA$^{Ser(UCN)}$ and ND1 genes can also be found in mitogenomes of other orthopteran insects [41].

Gene overlaps in the *A. meridionalis* mitogenome occurred 6 times in total, and the total length of overlapping sequences was 44bp, with the size ranging from 5 to 9 bp. These overlaps exist between *ND2* and *tRNA^{Trp}*, *tRNA^{Trp}* and *tRNA^{Cys}*, *tRNA^{Tyr}* and *COI*, *COI* and *tRNA$^{Leu\ (UUR)}$*, *ATP8* and *ATP6*, and *ND4* and *ND4L*. The longest overlapping sequence was 9 bp in length, and its position was between *ND2-tRNA^{Trp}*. In other insects that have been measured, the sum of overlapping sequence degrees found ranges from 20 bp (*Bombyx mori*) [15] to 152 bp (*Anopheles quadrimaculatus*) [47].

Phylogenetic Relationships

Based on the Maximum Likelihood (ML) tree (Figure 2), we found, firstly, *Gomphocerus licenti* (belonging to the subfamily Gomphocerinae) and *G. sibiricus* formed a single branch, then it clustered with *G. sibiricus tibetanus*. Secondly, they clustered with *Gomphocerippus rufus* (belonging to Gomphocerinae) forming a branch, then they clustered with *Chorthippus chinensis* (belonging to the subfamily Oedipodinae) forming a larger branch. Thirdly, this branch clustered with *Euchorthippus fusigeniculatus* (belonging to Gomphocerinae), then with *A. meridionalis* formed a branch.

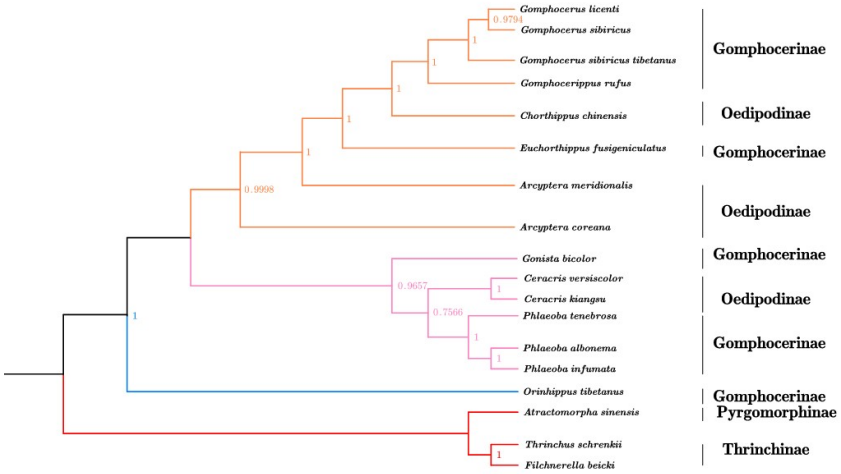

Figure 2: ML phylogenetic tree based on the 13 PCGs dataset

In terms of morphological classification, the most important difference between the Gomphocerinae and the Arcypteridae is that the Gomphocerinae grasshoppers have antenna clavate similar to butterflies. Based on Figure 2, the species of the genus *Phlaeoba, Phlaeoba infumata, P. albonema* and *P. tenebrosa* (belonging to subfamily Gomphocerinae), formed a monophyletic group, then clustered with the bamboo locusts *Ceracris versicolor* and *C. kiangsu. Finally,* they clustered with *Gonista bicolor* of Gomphocerinae. Thus, this again suggests that the subfamily Gomphocerinae is not a monophyletic group. This is consistent with previous research results [13].

Bayesian tree (Figure 3) showed that, 1) P. albonema and C. chinensis clustered to one group, which is not consistent with tranditional taxonomy; 2) G. rufus and G. sibiricus tibetanus merged to one group, which is not consistent with taxonomy. 3) Other relationships were consistent with ML trees.

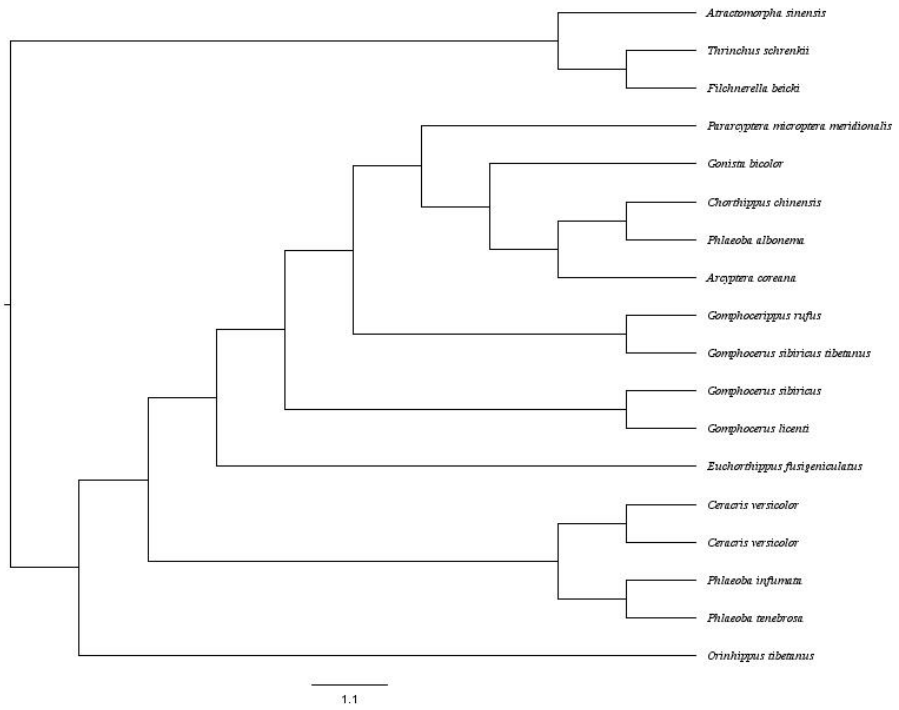

Figure 3: Bayesian tree based on the 13 PCGs gene dataset

The analysis of Bayesian system tree found that the position of individual species in the phylogenetic tree had changed significantly, and some species could not be accurately classified in terms of family classification. Moreover, the results were quite different from those of ML tree and current taxonomy. Therefore, the results of Bayesian system tree were not reliable and were not adopted in this study.

It can be seen from the comparison of the construction methods of the two phylogenetic analyses that in order to discuss the detailed taxonomic status of Orthoptera, a large amount of molecular data needs to be accumulated and analyzed, so that it is possible to obtain reliable taxonomic status system data. Most importantly, the results from partial genotyping should be balanced with those from whole genotyping, a classification system on which the current taxonomic status depends.

Conclusion

In the construction of phylogenetic relationships, it can be found that the classification of Gomphocerinae is not very clear in this study. Monophylies of the Arcypteridae and the Gomphocerinae are not supported, because in the phylogenetic trees, two species always intersect.

Acknowledgment

The specimens used in this study were provided by Professor Jia-Sheng Hao at Anhui Normal University. This project was supported by the Natural Science Foundation of China (Nos. 31970449 and 32470444).

References

1. Dong LJ (2012) Molecular Phylogeny of locusts based on COI, COII and COIII genes. Baoding: Hebei University.

2. Wei SJ, Chen XX (2011) Progress in research on the comparative mitogenomics of insects. Chinese Journal of Applied Entomology, 48: 1573-85.

3. Avise JC (2009) Phylogeography: retrospect and prospect, J Biogeogr, 36: 3-15.

4. Cameron SL (2014) Insect mitochondrial genomics: implications for evolution and phylogeny. Annu. Rev. Entomol, 59: 95–117.

5. Osigus HJ, Eitel M, Bernt M, Donath A, Schierwater B (2013) Mitogenomics at the base of Metazoa. Mol. Phylogenet. Evol, 69: 339-51.

6. Simon C, Buckley TR, Frati F, Stewart JB, Beckenbach AT (2006) Incorporating molecular evolution into phylogenetic analysis, and a new compilation of conserved polymerase chain reaction primers for animal mitochondrial DNA. Annu. Rev. Ecol. Evol. Syst, 37: 545–79.

7. Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. International Review of Cytology, 141: 173.

8. Boore JL (1999) Animal mitogenomes. Nucleic. Acids. Res, 27: 1767-80.

9. Castro LR, Ruberu K, Dowton M (2006) Mitogenomes of *Vanhornia eucnemidarum* (Apocrita: Vanhorniidae) and *Primeuchroeus* spp. (Aculeata: Chrysididae): evidence of rearranged mitogenomes within the Apocrita (Hymenoptera). Genome, 49: 752-66.

10. Flook PK, Rowell CH, Gellissen G (1995) The sequence, organization, and evolution of the *Locusta migratoria* mitogenome. J. Mol. Evol. 41: 928-41.

11. Saccone C, De Giorgi C, Gissi C, Pesole G, Reyes A (1999) Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene, 238: 195–209.

12. Chintauan-Marquier I, Legendre F, Hugel S, Robillard T, Grandcolas P et al. (2016) Laying the foundations of evolutionary and systematic studies in crickets (Insecta, Orthoptera): a multilocus phylogenetic analysis. Cladistics, 32: 54–81.

13. Huang Y, Liu N, Lu HM (2010) Research progress in mitochondrial genomes of the Orthoptera insects. Acta Entomologica Sinica, 53: 58-6.

14. Avz B, Desalle R (1994) Practical and theoretical considerations for choice of a DNA sequence region in insect molecular systematics, with a short review of published studies using nuclear gene regions. Ann. Entomol. Soc. Am, 87: 702-16.

15. Liu N (2006) Sequencing and analysis of the complete mitochondrial genome of *Arcyptera coreana*. Shaanxi Normal University.

16. Ma C, Liu C, Yang P et al. (2009) The complete mitogenomes of two band-winged grasshoppers, *Gastri margus marmoratus* and *Oedaleus asiaticus*. BMC Genom, 10: 156.

17. Xu Q, Hao Y, Mei K et al. (2016) The complete mitogenome of Ceracris versicolor (Orthoptera: Acridoidea: Arcypteridae). Mitochondrial DNA Part A DNA Mapping Sequencing & Analysis, 27: 512.

18. Zhang D, Zhi Y, Yin H et al. (2011) The complete mitogenome of *Thrinchus schrenkii*, (Orthoptera: Caelifera, Acridoidea, Pamphagidae). Mol. Biol. Rep. 38: 611-9.

19. Li XJ, Zhi YC, Liu GJ et al. (2015) The complete mitogenome of Asiotmethis jubatus (Uvarov, 1926) (Orthoptera: Acridoidea: Pamphagidae). Mitochondrial DNA, 26: 785-6.

20. Shi QY, Zheng JY, Zhi YC et al. (2015) The complete mitogenome of *Sinotmethis brachypterus* Zheng & Xi, 1985 (Orthoptera: Acridoidea: Pamphagidae: Prionotropisinae). DNA Sequence, 27: 4063-4.

21. Zhang DP (2020) Systematic study of Haplotropidini Sergeev from China (Orthoptera: Acridoidea: Pam- phagidae: Pamphaginae) . Shandong Agricultural University, 2020.

22. Zhang HL, Zeng HH, Huang Y et al. (2013) The complete mitogenomes of three grasshoppers, *Asiotmethis zacharjini*, *Filchnerella helanshanensis* and *Pseudotmethis rubimarginis* (Orthoptera: Pamphagidae. Gene 2013, 517: 89-98.

23. Li XJ, Zhi YC, Lang L et al. (2013) The complete mitogenome of *Filchnerella beicki* Ramme,1931 (Orthoptera: Acridoidea: Pamphagidae). Mitochondrial DNA, 25: 348-9.

24. Yin H, Zhi Y Jiang H et al. (2012) The complete mitogenome of *Gomphocerus tibetanus* Uvarov, 1935 (Orthoptera: Acrididae: Gomphocerinae). Gene, 494: 214-8.

25. Zhang HL, Zhao L, Zheng ZM et al. (2013) Complete mitogenome of *Gomphocerus sibiricus* (Orthoptera: Acrididae) and comparative analysis in four Gomphocerinae mitogenomes. Zoolog Sci, 30: 192-204.

26. Song W, Zhi Y, Liu G, et al. (2014) The complete mitogenome of *Orinhippus tibetanus* Uvarov, 1921 (Orthoptera: Acridoidea: Gomphoceridae). Mitochondrial DNA, 27: 76.

27. Qin Z, Guo C, Yuan H (2015) The complete mitogenome of *Gonista bicolor* (Haan) (Orthoptera: Acrididae). Mitochondrial DNA, 27: 1.

28. Song W, Ye B, Cao X et al. (2016) The complete mitogenome of *Phlaeoba tenebrosa* (Orthoptera: Acridoidea: Acrididae). Mitochondrial DNA, 27: 409.

29. Liu F, Qiu Z (2016) The complete mitogenome of *Phlaeoba infumata* (Orthoptera: Acrididae). Mitochondrial DNA, 1: 1-2.

30. Mugleston JD, Naegle M, Song H, Whiting MF (2018) A comprehensive phylogeny of Tettigoniidae (Orthoptera: Ensifera) reveals extensive ecomorph convergence and widespread taxonomic incongruence. Insect Syst. Diver, 2: 1–27.

31. Song H, Mariño-Pérez, R, Woller DA, Cigliano MM (2018) Evolution, diversification, and biogeography of grasshoppers (Orthoptera: Acrididae). Insect Syst. Diver, 2: 1-25.

32. Zahid S, Mariño-P Rez R, Song H (2021) Molecular phylogeny of the grasshopper family Pyrgomorphidae (Caelifera, Orthoptera) reveals rampant paraphyly and convergence of traditionally used taxonomic characters. Zootaxa, 4969: 101118.

33. Yuan H, Huang Y, Mao Y, Zhang N, Nie Y et al. (2021) The evolutionary patterns of genome size in Ensifera (Insecta: Or-

thoptera). Front. Genet, 12: 693541.

34. Alfsnes K, Leinaas HP, Hessen DO (2017) Genome size in arthropods: different roles of phylogeny, habitat and life history in insects and crustaceans. Ecol. Evol. 2017, 7, 5939–47.

35. Sambrook J, Russell DW (2001) Molecular Cloning. 3rd edn. pp. 1949. New York, Cold Spring Harbor Laboratory Press.

36. Xu MY (2009) Phylogeny of Locust based on 18S rRNA, 16S rRNA, COⅠ and COⅡ genes. Hebei University.

37. Yang K (2019) Application of molecular biology to locust research. Agriculture and Technology, 39: 18-19+143.

38. Lessinger AC, Martins Junqueira AC, Lemos TA, Kemper EL, da Silva FR et al. (2000) The mitogenome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera： Calliphoridae). Insect Mol. Biol, 9: 521-9.

39. Zhang DX, Hewitt GM (1997) Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. Biochem. Syst. Ecol, 25: 99-120.

40. Boore JL (2006) The use of genome-level characters for phylogenetic reconstruction. Trends Ecol. Evol, 21: 439-46.

41. Gao J, Chen CH, Huang Y (2009) Analysis of Complete mitochondrial genome sequence of *Aeropus licenti* Chang. Zool. Res, 30: 603-12.

42. Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitogenomes. J. Mol. Evol, 41: 353-8.

43. Lu HJ, Huang Y (2012) Phylogenetic relationship among some groups of orthopteran based on complete sequences of the mitochondrial COI gene. Zool. Res, 33: 319-28.

44. Mitchell SE, Cockburn AF, Seawright JA (1993) The mitogenome of *Anopheles quadrimaculatus* speies: A complete nucleotide sequence and gene organization. Genome, 36: 1058-73.

45. Boore JL, Boore JL (1999) Animal mitogenomes. Nucleic. Acids. Res, 27: 1767-80.

46. Zhou ZJ, Huang Y, Shi FM (2007) The mitogenome of the *Ruspolia dubia* (Orthoptera: Conocephalidae): A short A+T-rich region with 70 bp in length. Genome, 50: 855-66.

47. Shi HW, Ding FM, Huang Y (2008) Complete mtDNA sequence of *Phlaeoba albonema* Zheng (Orthoptera， Acrididae). Chn J Biochem. Mol. Biol, 24: 604-11.

48. Yukuhiro K, Sezutsu H, Itoh M, Shimizu K, Banno Y (2002) Significant levels of sequence divergence and gene rearrangements have occurred between the mitogenomes of the wild mulberry silkmoth, Bombyx mandarina, and its close relative, the domesticated silkmoth, *Bombyx mori*. Mol. Biol. Evol,19: 1385-9.

49. Wang HM (2016) Mitogenome codon bias and its evolution in Orthoptera insects. Shaanxi Normal University.