SCIENTIFIC REPORTS

OPEN

Comparative analysis of the *Liriomyza chinensis* mitochondrial genome with other Agromyzids reveals conserved genome features

Jing-Yun Chen[1,2], Ya-Wen Chang[1], Si-Zhu Zheng[2], Ming-Xing Lu[1] & Yu-Zhou Du [1,3]

Liriomyza chinensis is a serious pest of onions in many countries, especially in East Asia. We sequenced the complete mitochondrial genome of this species and compared it with five other Agromyzidae species. The *L. chinensis* mitogenome is a double-stranded 16,175 bp circular molecule with an A + T content of 78.3%. It contains 37 genes and a control region as do the sequenced *Liriomyza* species. The mitogenomes of *L. chinensis* and other Agromyzidae species showed a clear bias in nucleotide composition with a positive AT-skew. Most PCGs used standard ATN as start codons, and TAN as termination codons. The tRNAs exhibited the typical clover-leaf structure, except for tRNASer(AGN) and the two rRNA genes are conserved with those of other Agromyzids. The *L. chinensis* mitogenome control region included several conserved regions, including a poly-T, two (TA)n and one poly-A stretch, which are considered important replication and transcription. The 13 PCGs were used to study the phylogeny of *L. chinensis* and five related Agromyzids. Analysis by maximum likelihood, Bayesian inference and genetic distance suggest congruent phylogenetic relationships in *Liriomyza* spp. and provide a useful supplement to taxonomic classification by morphology.

Mitochondria are involved in energy metabolism, apoptosis, aging, disease and oxidative phosphorylation[1–3]. Arthropod mitochondrial genomes (mtDNA) are generally circular, duplex molecules of 14–19 kb in length[4–6]. Insect mtDNA genomes contain a remarkably conserved set of 37 genes including 22 transfer RNA (tRNA) genes, two ribosomal RNA (rRNA) genes, 13 protein-coding genes (PCGs) and a control region (CR) or A + T-rich non-coding region[6,7]. Mitochondrial genomes have been widely used in phylogenetic studies and comparative and evolutionary genomics of insects and as molecular markers of population genetics and evolution[6,8–10].

Liriomyza chinensis belongs to the group Phytomyzinae, family Agromyzidae and order Diptera and causes significant damage to *Allium* spp.[11]. The damage incited by *L. chinensis* on onions is very similar to other *Liriomyza* spp., the mining of leaves by larvae and puncturing of foliage by females for feeding and oviposition reduces photosynthesis, which leads to lower crop quality and quantity[12–15]. The leafminer *L. chinensis* has become a serious pest of onions in many countries and regions, especially in East Asia[11,16,17]. The taxonomic status of *L. chinensis* is particularly controversial over the past decades, but now its classification status is settled. Kato considered *L. chinensis* as a sub-species of *Dizygomyza cepae* (also known as *L. cepae*)[18], and Hering reported that larvae of *L. chinensis* and *L. cepae* had identical spiracle structures. The male genitalia of *L. chinensis* show similarity to *L. cepae* and both species have characteristically pale wings and solid black scutellum[19], these characteristics are different from the typical form of *Liriomyza*, represented by *L. nietzkei*[11]. Subsequent speciation produced *L. chinensis* in China, Japan and Malaysia, and *L. cepae* in western Europe, which are reproductively isolated from *L. nietzkei*. Consequently, Spencer upgraded the classification of *L. chinensis* from sub-species to species and assigned it to the genus *Liriomyza*[11].

The systematics of Agromyzids is rather poorly understood due to their small size and morphological homogeneity. For molecular phylogenetic study of Agromyzids, Scheffer *et al.* investigated the phylogenetic relationships

[1]School of Horticulture and Plant Protection & Institute of Applied Entomology, Yangzhou University, Yangzhou, 225009, China. [2]Suzhou Customs (formerly Suzhou Entry-Exit Inspection and Quarantine Bureau), Suzhou, 215000, China. [3]Joint International Research Laboratory of Agriculture and Agri-Product Safety, Yangzhou University, Yangzhou, 225009, China. Correspondence and requests for materials should be addressed to Y.-Z.D. (email: yzdu@yzu.edu.cn)
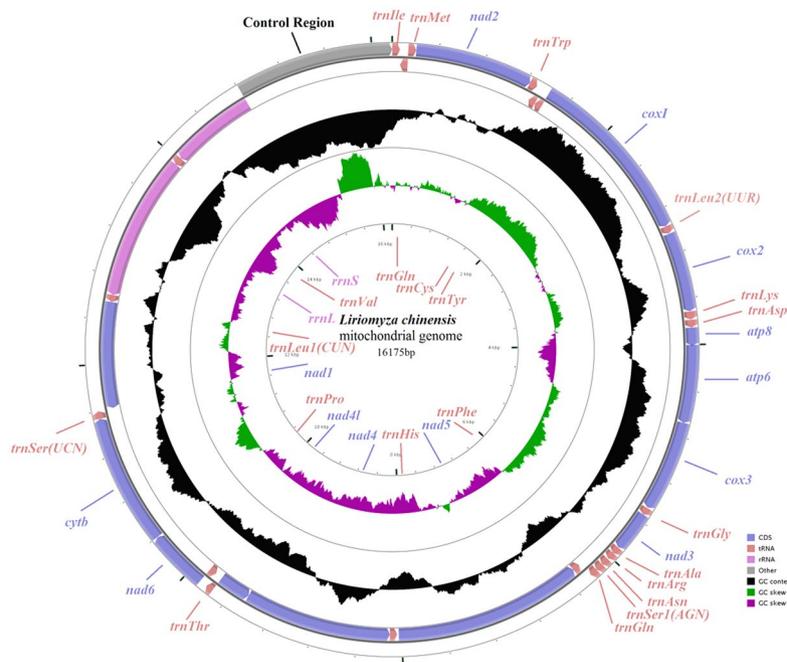
Figure 1. Map of mitochondrial genome of *L. chinensis*. Genes outside the map are transcribed in a clockwise direction (J-strand), whereas those inside the map are transcribed counterclockwise (N-strand). The second circle shows the GC content and the third shows the GC skew. GC content and GC skew are plotted as the deviation from the average value of the entire sequence.

among genera within the Agromyzidae using parsimony and Bayesian analyses of the mitochondrial COI gene, the nuclear ribosomal 28 S gene, and the single copy nuclear CAD gene[20]. But the study on phylogenetic relationship of the genus *Liriomyza* based on whole mitochondrial genome was relatively limited. The development of improved sequencing technology has generated the complete or near complete mitogenomes of five Agromyzids including *L. sativae*[21], *L. trifolii*[22,23], *L. bryoniae*[23], *L. huidobrensis*[24], and *Chromatomyia horticola*[25], which provide the basis for studying the phylogeny of Agromyzid species.

The mtDNA sequence of *L. chinensis* has not been previously reported and would be valuable in clarifying the taxonomic issues described above. In this paper, we report the complete mitochondrial genome of *L. chinensis* and provide a thorough description of its structural features. The *L. chinensis* mitogenome was compared with mtDNA sequences of five other related species to better understand taxonomy and phylogeny within the Agromyzidae.

Results and Discussion

Genome organization. The complete mitochondrial genome of *L. chinensis* is a circular 16,175 bp molecule (GenBank accession no. MG252777). It includes 37 mitochondrial genes (13 PCGs, 22 tRNA genes and two rRNA genes) and a large non-coding region (control region) (Fig. 1). The gene order in the *L. chinensis* mitochondrial genome is identical to *D. melanogaster*[26], which is the classic structure for Diptera. There are 23 genes located on the J-strand (nine PCGs and 14 tRNAs) and 14 genes on the N-strand (four PCGs, eight tRNAs and two rRNAs). Sixteen intergenic spacers were identified with a total length of 64 bp; these ranged in size from 1–19 bp with the longest intergenic spacer located between tRNA^Glu and tRNA^Phe. There were nine overlapping genes in the mitochondrial genome of *L. chinensis*; the longest overlap was 8 bp and mapped between tRNA^Trp and tRNA^Cys (Table 1).

The mitochondrial genome length of *L. chinensis* was similar to other Agromyzidae family members. Exceptions were *L. sativae* and *C. horticola*; the former contains a short A + T region[21], and the latter lacks an A + T region due to incomplete sequencing[25]. The gene order is identical to the ancestral *trnI–trnQ-trnM* arrangement. There were no gene rearrangements in the six Agromyzidae mitogenomes, which indicates that the mitochondrial gene order is highly conserved in Agromyzidae. Furthermore, the length and position of intergenic spacers was also highly conserved.

Nucleotide Composition. The nucleotide composition of the *L. chinensis* mtDNA showed an obvious bias for A and T. The A + T content of the whole genome was 78.3% (A = 41.3%, T = 37.0%, G = 8.9%, C = 12.8%). The A + T content of isolated PCGs, tRNAs, rRNAs, and the control region exceeded 75%, and the control region had the highest A + T content (89.4%) (Table 2). This strand bias in nucleotide composition is a universal phenomenon in metazoan mitochondrial genomes and is evident by a comparative analysis of AT- and GC-skews[4,6,27]. The Agromyzids mtDNAs showed a positive AT- and negative GC-skew over the entire genome (Table 3). The PCGs, tRNAs and rRNAs of the six Agromyzids mtDNAs show a relatively consistent A + T content and AT-skew (Table 3). The A + T-rich region *L. chinensis* exhibited a lower A + T content (~90%) in comparison to the other

Feature	Region		Direction	Length	Codon		Intergenic Nucleotides
	From	To			Start	Stop	
tRNA^{Ile(I)}	1	67	F	67			
tRNA^{Gln(Q)}	71	139	R	69			3
tRNA^{Met(M)}	139	207	F	69			−1
ND2	208	1228	F	1021	ATT	T	−1
tRNA^{Trp(W)}	1229	1295	F	67			0
tRNA^{Cys(C)}	1288	1351	R	64			−8
tRNA^{Tyr(Y)}	1353	1416	R	64			1
COI	1414	2953	F	1540	ATCA	TAA	−4
tRNA^{Leu(UUR)}	2955	3020	F	66			1
COII	3024	3713	F	690	ATG	TAA	2
tRNA^{Lys(K)}	3715	3785	F	71			1
tRNA^{Asp(D)}	3788	3854	F	67			2
ATP8	3855	4010	F	156	ATT	TAA	0
ATP6	4004	4678	F	675	ATG	TAA	−7
COIII	4678	5469	F	792	ATG	TAA	−1
tRNA^{Gly(G)}	5475	5539	F	65			5
ND3	5540	5893	F	354	ATT	TAA	−1
tRNA^{Ala(A)}	5896	5959	F	64			2
tRNA^{Arg(R)}	5960	6022	F	63			0
tRNA^{Asn(N)}	6029	6094	F	66			6
tRNA^{Ser(AGN)}	6095	6161	F	67			0
tRNA^{Glu(E)}	6162	6228	F	67			0
tRNA^{Phe(F)}	6248	6314	R	67			19
ND5	6315	8028	R	1714	ATT	T	−1
tRNA^{His(H)}	8044	8109	R	66			15
ND4	8110	9448	R	1339	ATG	T	0
ND4L	9449	9738	R	290	ATG	TA	0
tRNA^{Thr(T)}	9741	9804	F	64			2
tRNA^{Pro(P)}	9805	9870	R	66			0
ND6	9873	10397	F	525	ATT	TAA	1
CYTB	10399	11533	F	1135	ATG	T	1
tRNA^{Ser(UCN)}	11534	11599	F	66			0
ND1	11602	12564	R	963	GTG	TAG	2
tRNA^{Leu(CUN)}	12566	12629	R	64			1
16 S rRNA	12630	13952	R	1323			0
tRNA^{Val(V)}	13952	14023	R	72			−1
12 S rRNA	14024	14808	R	785			0
control region	14809	16175		1367			0

Table 1. Annotation of the mitochondrial genome of *L. chinensis*.

Region	Nucleotides Proportions (%)						AT-skew	GC-skew
	A	T	G	C	A + T	G + C		
Whole genome	41.3	37.0	8.9	12.8	78.3	21.7	0.05	−0.18
Protein coding genes	40.1	36.0	9.9	14.0	76.1	23.9	0.05	−0.17
1st codon position	40.0	34.7	11.8	13.5	74.7	25.3	0.07	−0.07
2nd codon position	36.7	35.6	10.7	17.0	72.3	27.7	0.02	−0.23
3rd codon position	43.8	37.6	7.1	11.5	81.4	18.6	0.08	−0.24
tRNA genes	40.3	37.3	10.1	12.2	77.6	22.3	0.04	−0.09
16 S rRNA	43.8	39.5	5.9	10.8	83.3	16.7	0.05	−0.29
12 S rRNA	41.4	39.6	6.9	12.1	81.0	19.0	0.02	−0.27
control region	49.2	40.2	4.5	6.1	89.4	10.6	0.10	−0.15

Table 2. Nucleotide composition of the *L. chinensis* mitogenome in different regions.

Species	Whole A+T%	PCGs A+T%	tRNA A+T%	rRNA A+T%	control region A+T%	AT-skew	GC-skew
L. sativae	77.5	75.7	77.1	82.2	93.0	0.05	−0.18
L. trifolii	78.2	75.6	77.2	82.6	93.6	0.05	−0.19
L. huidobrensis	78.3	75.7	77.4	82.2	93.0	0.04	−0.17
L. bryoniae	79.3	76.7	78.5	82.4	95.5	0.04	−0.19
C. horticola	78.2	76.1	76.9	81.4	N/A	0.05	−0.16

Table 3. Nucleotide composition in regions of Agromyzidae mitogenomes. N/A: data not available.

congener mitogenomes. The underlying mechanism of the A + T bias has been explained by asymmetric mutation and selection pressure during replication and transcription[28]. We tried to determine if there were any relationships between the A + T content and phylogeny but clear patterns were not evident. The A + T nucleotide bias has significance for the study of replication, transcription and rearrangement of the mitochondrial genome.

Protein-coding genes.   The nucleotide bias was also reflected in the 13 PCGs, which had a relatively high A + T percentage (~76.1%, Table 2). The average A + T content among PCGs in *L. chinensis* was 76.1%. The A + T content of the third codon position (81.4%) was higher than the first (74.7%) and second codon (72.3%) positions (Table 2), which may suggest that both higher mutation rates and increased A + T content are related and depend on a relaxed selection at the third codon position[21,29].

Eleven PCGs of *L. chinensis* were found to initiate with ATN (five with ATT and six with ATG). However, *ND1* and *COI* started with GTG and the special quadruplet start codon of ATCA (Table 1), respectively, which agreed with *L. trifolii* and *L. sativae*, but differed from other Agromyzidae species[10,21]. These special start codons are converted into typical initiation codons by RNA editing during transcription[30], which can reduce the intergenic spacer and avoid gene overlap[31].*COI* genes generally use nonstandard and varied start codons in insects. Among the six Agromyzids, five *Liriomyza* species all used "ATCA" as special quadruplet start codon[21,23], while for *C. horticola*, which used the "TTG" as nonstandard start codon[25]. Consequently, the use of nonstandard initiation codon in *COI* gene was not unexpected and it was shown to be dependent on the translated amino acid sequence and subsequent sequence alignments[32]. Eight PCGs used the typical termination codons TAA and TAG (*ND1*), whereas *ND2*, *ND5*, *ND4*, and *CYTB* used incomplete stop codons with T as a termination signal (Table 1). *ND4L* used TA as a termination signal, which has been reported in other Agromyzids[10]. Incomplete termination codons are common in metazoan mitochondrial genomes. It has been speculated that the polyadenylation site is generated by adding A to the 3' end of the mRNA transcript, which is then converted into a complete stop codon for termination of transcription[33].

The relative synonymous codon usage (RSCU) values of the *L. chinensis* mitogenome were calculated and illustrated, and the RSCU for Agromyzidae is shown (Fig. 2). The use of anticodons NNA and NNU indicated a preference for A or T in the third nucleotide of PCG anticodons. All possible codons are present in the PCGs of *L. chinensis* and the other four *Liriomyza* spp., whereas GCG was not found in *C. horticola*. Previous research indicates that codons with high G and C content are generally not favored, a phenomenon with low GC content that is found in some insects, such as moths[34,35], stonefly[36], whitefly[37] etc.

The rates of nonsynonymous (Ka) and synonymous substitutions (Ks) and the Ka/Ks ratio were calculated for all PCGs in the six Agromyzidae mtDNA genomes using *D. melanogaster* as a reference sequence (Fig. 3). All Ka values were less than Ks values; consequently, the Ka/Ks ratios were less than 1 (Fig. 3), indicating the likelihood of purifying selection in these species[38]. Agromyzidae species generally show relatively consistent evolutionary rates, which may be related to their relatively constant habitat as larval leafminers[39,40].

tRNA genes.   The tRNA genes of all the six Agromyzidae species contained an A + T content exceeding 76%. Twenty-two complete tRNAs were identified in the *L. chinensis* mtDNA, and 20 were discovered using tRNAscan-SE. tRNA^{Arg} and tRNA$^{Ser(AGN)}$ could not be detected by software, but were instead determined through comparison with published Agromyzidae mitochondrial genomes (Fig. 4). The typical number of tRNA genes in the mtDNA of Agromyzids was 22, but in *L. trifolii* and *L. bryoniae*, there were two additional tRNA genes in the A + T-rich region near 12 S rRNA[10]. Since the anticodons of the four additional tRNAs were atypical, Yang *et al.* suggested that these additional tRNAs were generated by gene duplications that could be folded into tRNA-like secondary structures but were nonfunctional[10]. All *L. chinensis* tRNAs folded into the typical clover-leaf structure except for tRNA$^{Ser(AGN)}$, which lacked the dihydrouridine (DHU) arm (Fig. 4). The DHU arm of tRNA$^{Ser(AGN)}$ formed a large loop instead of the conserved stem-and-loop structure (Fig. 4). Atypical numbers and structures of tRNA have been reported in other insects[8,38,41–43]. The factors that may have led to these truncated tRNAs remain unknown, although truncation may be a result of generalized evolutionary pressures for size reduction in mitogenomes[6,44], but such an explanation requires the existence of compensatory mechanisms. Thus, the abnormal structure of tRNA$^{Ser(AGN)}$ warrants further study. The *L. chinensis* tRNAs ranged from 63 (tRNA$^{Arg(R)}$) to 72 (tRNA-$^{Val(V)}$) nucleotides; the length of the tRNA usually depends on the size of the variable and D-loops[45]. Based on the secondary structure of the tRNAs in the *L. chinensis* mtDNA, there were 19 unmatched nucleotides including 14 G-U and five U-U pairs; these mapped to the amino acid, TψC, and anticodon arms. A total of 27 mismatched bases were reported in the tRNA of *L. sativae*, these included 21 G-U, four U-U, one A-A and one A-C pairs that were located in the AA (8 bp), DHU (10 bp), AC (5 bp) and TψC arms (4 bp), respectively[21].
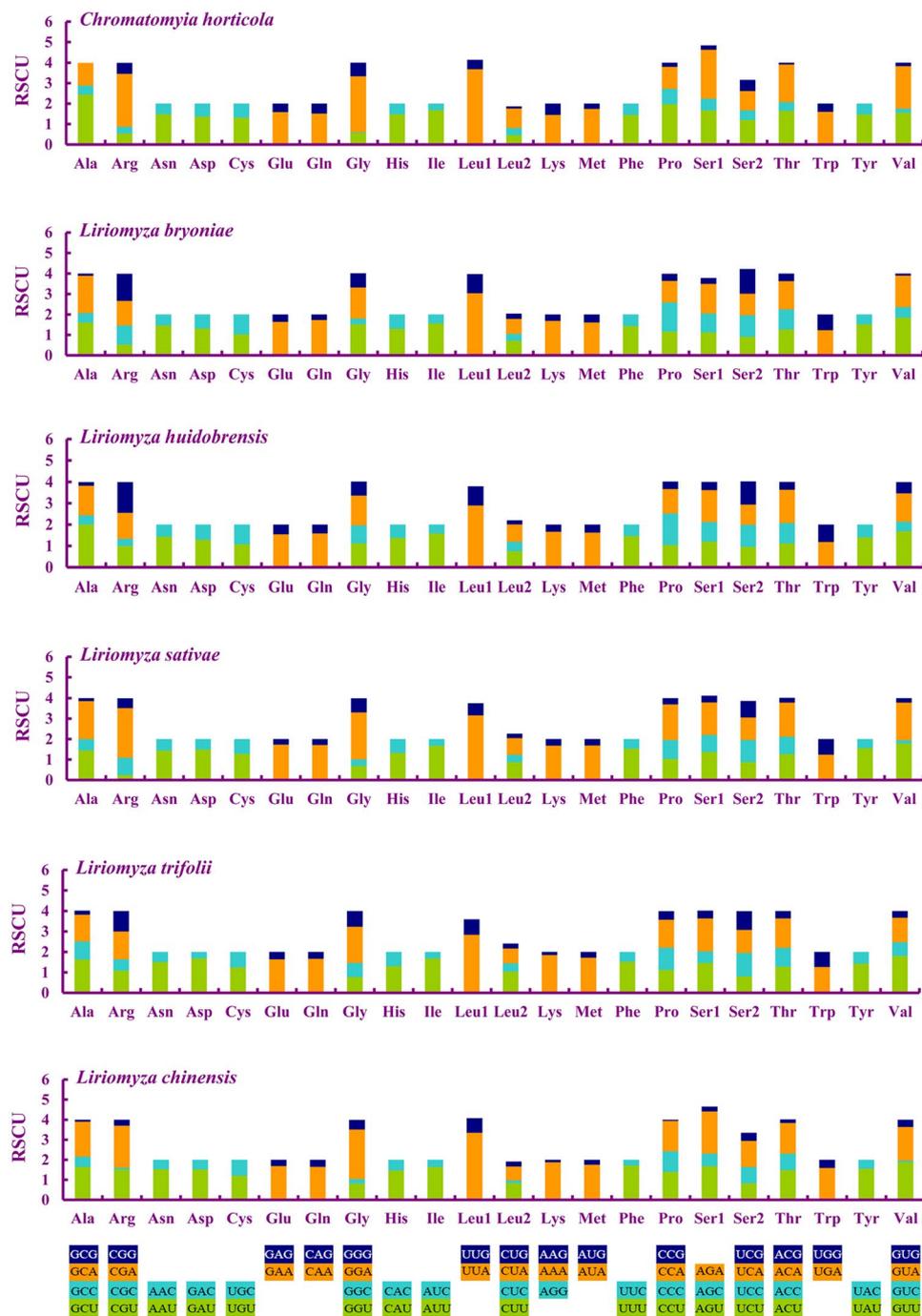
Figure 2. The mitochondrial genome relative synonymous codon usage (RSCU) across six Agromyzidae flies. Codon families are provided on the X axis.

rRNA genes. The boundaries of rRNA genes were identified by sequence alignment with published Dipteran sequences. The *L. chinensis* mtDNA contained the 16 S rRNA and 12 S rRNA, which mapped between tRNA[Leu(-CUN)]/tRNA[Val] and tRNA[Val]/control region, respectively (Fig. 1). The 16 S and 12 S rRNA genes are 1323 and 785 bp with an A + T content of 83.3 and 81.0%, respectively. The two rRNAs mapped to the same location as described for other Agromyzidae mitogenomes. Both the 16 S and 12 S rRNAs have been widely used for population genetics, molecular phylogeny and species identification[46,47]. However, the rRNA sequences of Agromyzids were relatively conserved[23], and thus would not provide much useful insight regarding population genetics. However, the secondary structure of rRNA genes in *Liriomyza* spp. may contain potentially useful information. For example, the 12 S rRNA of *L. huidobrensis* showed more variability in sequence and structure of the H51-H100 region arm as compared to *L. trifolii*; thus, this region may be a potential marker for identification of *Liriomyza* spp.[24].
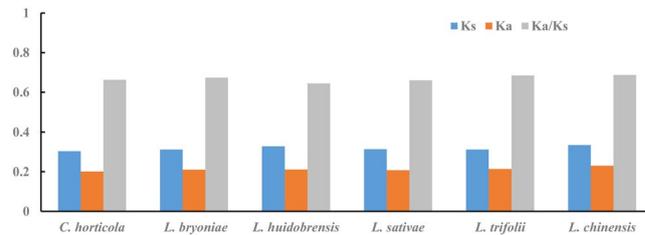
Figure 3. Evolutionary rates of Agromyzidae flies mitochondrial genomes. The number of nonsynonymous substitutions per nonsynonymous site (Ka), the number of synonymous substitutions per synonymous site (Ks), and the ratio of Ka/Ks for each Agromyzidae flies mitochondrial genome are given, using that of *D. melanogaster* as a reference sequence.

Control Region (A + T-rich region). The control region of the *L. chinensis* mtDNA is located between the 12 S rRNA and tRNA^{Ile} genes (Fig. 1); it consists of 1367 nucleotides and has the highest A + T content (89.4%) of the mtDNA genome (Table 1). This region varies greatly in length among insects, ranging from 70 bp to 13 kb[48,49], and it accounts for most of the variation in mtDNA size. The control regions in *L. trifolii*, *L. bryoniae*, *L. huidobrensis*, and *L. sativae* have a high A + T content (>90%), also map between the 12 S rRNA and tRNA^{Ile} genes, and are 1338, 1354, 1416, and 741 bp in length, respectively[21,22]. The A + T-rich region is the fastest evolving region in the mitochondrial genome[50], and a comparative analysis of mtDNA sequences in *Drosophila* shows that divergence in the control region is very significant in most species[51]. However, there are five conserved structural elements have been found in the control region of many insects including a poly-T stretch, a [TA(A)]n-like stretch, a highly conserved stem-and-loop structure, a pair of sequences immediately flanking the stem/loop structure with reiterated TATA and G(A)n T consensus sequences, and a G + A-rich stretch downstream of the secondary structure[49]. We identified several conserved structural elements in the control region of the *L. chinensis* mtDNA; these included one poly-T stretch, two (TA)n stretches and one poly-A stretch (Fig. 5). The A + T-rich region is the largest noncoding region in mtDNA and is associated with replication and transcription, which is why it is named the control region[6,52]. It is highly variable both in content and size due to insertions and deletions, variation in copy numbers of tandem repeats, and extensive change in the length of a variable domain[50,53,54]. Studies have shown that the A + T-rich region harbors sufficient polymorphisms to be a suitable marker for studying population genetics and phylogenetic reconstruction of closely related taxa[55,56].

Phylogenetic analysis. We performed phylogenetic analysis of mtDNA using the nucleotide sequences of 13 PCGs in six Agromyzid mitochondrial genome sequences; *D. melanogaster* served as an outgroup. The topology of two phylogenetic trees constructed separately by maximum likelihood (ML) and Bayesian inference (BI) analyses were very similar. *L. sativae* grouped with *L. trifolii*, while *L. huidobrensis* and *L. bryoniae* were in another group, and *L. chinensis* was situated between the other *Liriomyza* spp. and *Chromatomyia* (Fig. 6). Phylogenetic analyses indicated that *L. trifolii*, *L. sativae*, *L. huidobrensis* and *L. bryoniae* are closely related; however, it was difficult to determine which *Liriomyza* spp. was most closely related to *L. chinensis* (Fig. 6).

Interspecific divergence spanned from 8.5% (*L. sativae* and *L. trifolii*) to 20.8% (*C. horticola* and *L. chinensis*) (Table 4). The genetic distance between DNA sequences is an important characteristic for classification and identification[57], and a 2% genetic distance was previously proposed as the threshold between species[58]. In this study, the genetic distance far exceeded 2%, which is consistent with the classification of the six Agromyzids as distinct species. The genetic distance among *Liriomyza* spp. was close to each other, while the genetic distance between *C. horticola* and other *Liriomyza* species was far away. In general, the pattern of genetic distance was consistent with the phylogenetic tree. During the long-term evolution progresses, the highly invasive polyphagous species (such as *L. trifolii*, *L. sativae*, *L. huidobrensis* and *L. bryoniae*) have similar host niches and environmental stress, so the genetic distance was close to each other. The convergence of environmental variation and ecological factors can influence speciation[59]; however, the genetic distance between *L. chinensis* and other *Liriomyza* species was relatively far, which probably resulted from the substantial differences in food preferences of *L. chinensis*[60].

Based on our data, the genus *Liriomyza* has relatively conserved mtDNA genomes and phylogenetic relationships, which conform the assignment of *L. chinensis* to the genus *Liriomyza* and provide a useful supplement to traditional taxonomic classification.

Materials and Methods

Sample and DNA extraction. Specimens of *L. chinensis* were collected from onions at Laiwu (36.12°N, 117.04°E) in Shangdong, China. All specimens were preserved in 100% ethanol and stored at -20 °C until DNA extraction was performed. Genomic DNA was extracted from samples using AxyPrep^{TM} Multisource Genomic DNA Kit (Axygen, California, USA) and then used for PCR.

PCR amplification and sequencing. The mitochondrial genome of *L. chinensis* was amplified from extracted genomic DNA using short, overlapping PCR fragments (<1.2 kb). Twenty-five universal primer pairs specific for Diptera mtDNA[61] were designed using Primer Premier 5.0 software (Supplementary Table S1). Conditions for PCR amplification were as follows: initial denaturation for 5 min at 94 °C, followed by 35 cycles of denaturation for 1 min at 94 °C, annealing for 1 min at 45–55 °C, elongation for 1.5 min at 72 °C, and a final extension step of 72 °C for 10 min. These PCR products were analyzed by 1.0% agarose gel electrophoresis and purified
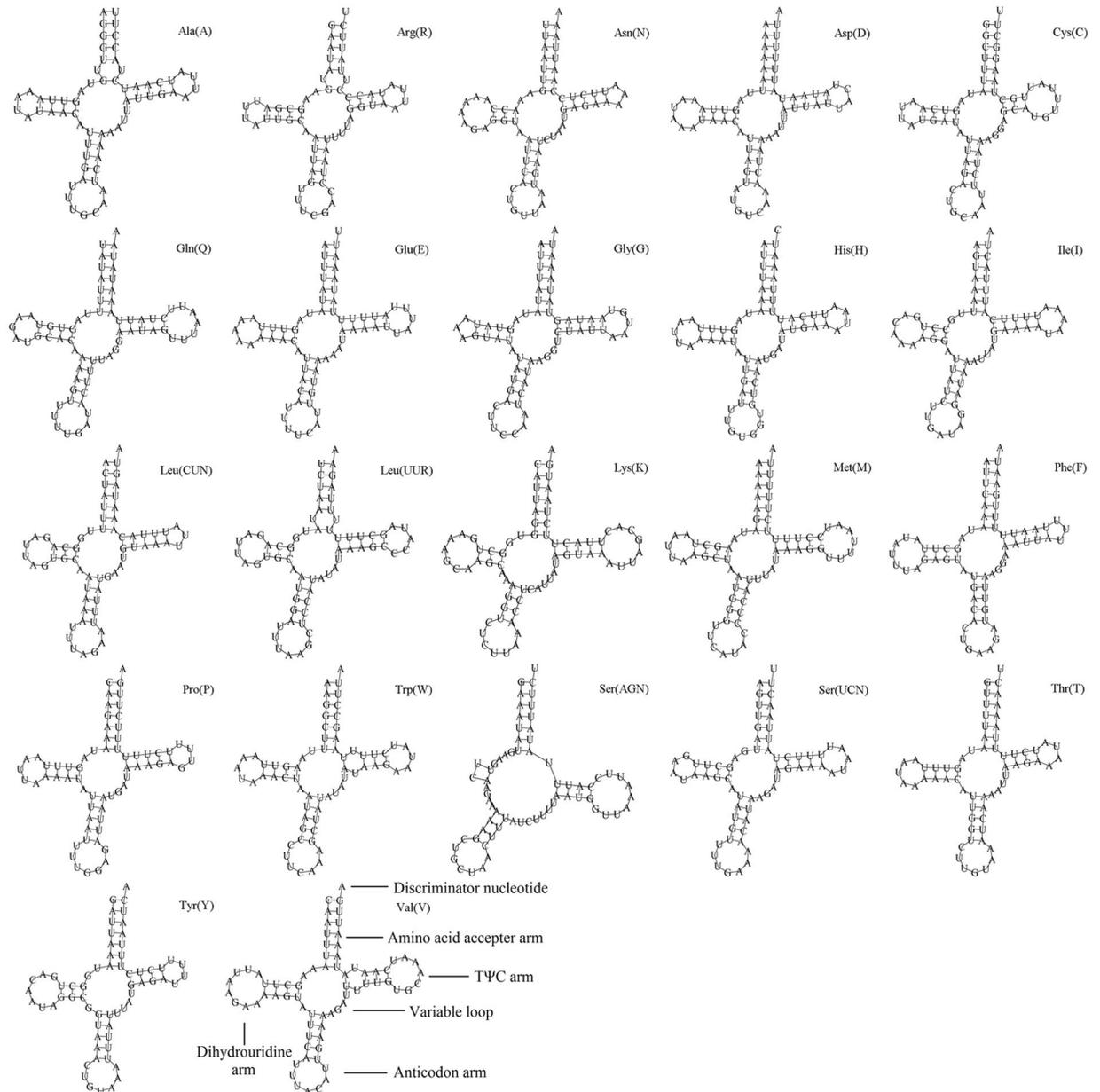
Figure 4. Inferred secondary structures of tRNAs from the *L. chinensis* mitogenome. The tRNAs are labelled with the abbreviations of their corresponding amino acids. Structural elements in tRNA arms and loops are illustrated as for *trnV*.
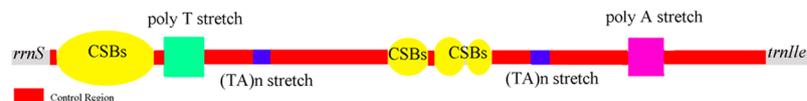

Figure 5. Predicted structural elements in the control region of *L. chinensis*. The genes flanking the control region, 12 S RNA and tRNA[Ile(I)], are represented in gray boxes; the red-shaded rectangles indicate A + T-rich regions; the purple/green box indicates conserved poly A/T structures; yellow boxes indicate conserved sequence blocks with other leafminer species; and blue box indicates (TA) n stretches by using the Tandem Repeats.

with an Axygen DNA Gel Extraction Kit (Axygen Biotechnology, Hangzhou, China). All amplified products were sequenced in both directions. If the sequenced result was bimodal, fragments were cloned into pGEM-T easy and re-was sequenced after cloning.
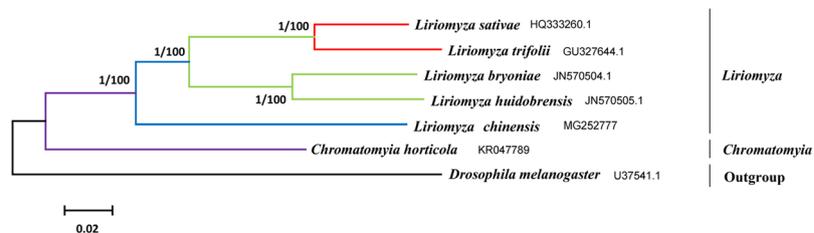
Figure 6. Inferred phylogenetic relationships among Agromyzidae based on nucleotide sequences of 13 protein-coding genes using Bayesian inference (BI) and maximum likelihood (ML). Numbers at each node indicate bootstrap support; percentages of ML bootstrap support values (first value) and Bayesian posterior probabilities (second value), respectively. *D. melanogaster* was used as an outgroup[59]. The scale bar indicates the number of substitutions per site.

	L. bryoniae	*L. huidobrensis*	*L. sativae*	*L. trifolii*	*L. chinensis*	*C. horticola*
L. bryoniae		0.096	0.147	0.156	0.179	0.189
L. huidobrensis	0.096		0.145	0.157	0.186	0.192
L. sativae	0.147	0.145		0.085	0.182	0.188
L. trifolii	0.156	0.157	0.085		0.183	0.194
L. chinensis	0.179	0.186	0.182	0.183		0.208
C. horticola	0.189	0.192	0.188	0.194	0.208	

Table 4. Interspecies average divergence of Agromyzidae based on the Kimura-2-parameters model.

Sequence Assembly, Annotation and Analysis. Protein-coding genes (PCGs) and rRNA genes in the *L. chinensis* mtDNA were identified by comparative analysis with other Agromyzidae family members. PCGs were aligned using Clustal X version 2.0[62] and the boundaries of individual genes were confirmed with ORF finder (https://www.ncbi.nlm.nih.gov/orffinder/). The mitogenomic map was depicted with CG View Server (http://stothard.afns.ualberta.ca/cgview_server/), and PCG nucleotide sequences (lacking start and termination codons) were translated using MEGA v. 6.0[63]. Both the A + T content and codon usage were calculated using MEGA v. 6.0. Skew analysis was carried out with formulas AT-skew = [A−T]/[A + T] and GC-skew = [G−C]/[G + C][64]. The software package DnaSP v. 5.10[65] was used to calculate synonymous substitution (Ks) and non-synonymous substitution rates (Ka). Most tRNAs were recognized by tRNAscan-SE v. 1.21 (http://lowelab.ucsc.edu/tRNAscan-SE/), and tRNAs that could not be identified using tRNAscan-SE were confirmed by sequence comparison with other Dipteran insects[66]. The tandem repeats in the putative control region were analyzed with Tandem Repeats Finder (http://tandem.bu.edu/trf/trf.advanced.submit.html).

Phylogenetic Analysis. Phylogenetic analyses were based on nucleotide sequence of 13 PCGs derived from *L. chinensis* and five other Agromyzidae - *L. sativae*, *L. trifolii*, *L. bryoniae*, *L. huidobrensis*, and *C. horticola* mitogenomes available from GenBank (GenBank accession nos. HQ333260.1, GU327644.1, JN570504.1, JN570505.1 and KR047789, respectively). The mitogenome of *Drosophila melanogaster* (U37541.1) was used as the outgroup[26]. The nucleotide sequences of the 13 PCGs were initially aligned with Clustal X, translated into amino acids using default settings, and then analyzed with MEGA v. 6.0. Alignments of individual genes were concatenated using default settings, and the stop codon was excluded. Phylogenetic analysis was conducted using maximum likelihood (ML) and Bayesian inference (BI), which were conducted with MEGA v. 6.0 and MrBayes v. 3.1.2[67]. The ML method was used to infer phylogenetic trees with 1000 bootstrap replicates. BI analyses were conducted under the following conditions: 1,000,000 generations, four chains (one cold chain and three hot chains) and a burn-in step for the first 10,000 generations. The confidence values of the BI tree were expressed as Bayesian posterior probabilities in percentages. Simultaneously, interspecific genetic divergence was calculated by MEGA v. 6.0 using the Kimmura-2-parameter model[68].

Accession codes. Sequence data used in this study was deposited in GenBank (accession number MG252777).

References

1. Brand, M. D. Regulation analysis of energy metabolism. *J. Exp. Biol.* **200**, 193–202 (1997).
2. Graeber, M. B. & Müller, U. Recent developments in the molecular genetics of mitochondrial disorders. *J. Neurol. Sci.* **153**, 251–263 (1998).
3. Kroemer, G., Dallaporta, B. & Rescherigon, M. The mitochondrial death/life regulator in apoptosis and necrosis. *Annu. Rev. Physiol.* **60**, 619–642 (1998).
4. Wolstenholme, D. R. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.* **141**, 173–216 (1992).
5. Boore, J. L. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**, 1767–1780 (1999).
6. Cameron, S. L. Insect mitochondrial genomics: Implications for evolution and phylogeny. *Annu. Rev. Entomol.* **59**, 95–117 (2014).
7. Shadel, G. S. & Clayton, D. A. Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* **66**, 409–435 (1997).

8. Chen, Z. T. & Du, Y. Z. First mitochondrial genome from *Nemouridae* (Plecoptera) reveals novel features of the elongated control region and phylogenetic implications. *Int. J. Mol. Sci.* **18**, 996 (2017a).
9. Cao, S. S., Yu, W. W., Sun, M. & Du, Y. Z. Characterization of the complete mitochondrial genome of *Tryporyza incertulas*, in comparison with seven other Pyraloidea moths. *Gene* **533**, 356–365 (2014).
10. Yang, F., Du, Y. Z., Cao, J. M. & Huang, F. N. Analysis of three leafminers' complete mitochondrial genomes. *Gene* **529**, 1–6 (2013a).
11. Spencer, K. A. A (Diptera) of economic importance. **9**: Series Entomologica, ed. by E. Schimitschek, Gottingen. Bath: The Hague Publishers. 162–166 (1973).
12. Johnson, M. W., Welter, S. C., Toscano, N. C., Tingi, P. & Trumble, J. T. Reduction of tomato leaflet photosynthesis rates by mining activity of *Liriomyza sativae* (Diptera: Agromyzidae). *J. Econ. Entomol.* **76**, 1061–1063 (1983).
13. Parrella, M. P., Jones, V. P., Youngman, R. R. & Lebeck, L. M. Effect of leaf mining and leaf stippling of *Liriomyza* spp. on photosynthetic rates of chrysanthemum. *Ann. Entomol. Soc. Am.* **78**, 90–93 (1985).
14. Chang, Y. W. *et al.* Cloning and expression of genes encoding heat shock proteins in *Liriomyza trifolii* and comparison with two congener leafminer species. *PLoS ONE* **12**, e0181355 (2017a).
15. Gao, Y. L., Reitz, S., Xing, Z. L., Ferguson, S. & Lei, Z. R. A decade of leafminer invasion in China: lessons learned. *Pest. Manag. Sci.* **73**, 1775–1779 (2017).
16. Tran, D. H., Ridland, P. M. & Takagi, M. Effects of temperature on the immature development of the stone leek leafminer *Liriomyza chinensis* (Diptera: Agromyzidae). *Environ. Entomol.* **36**, 40–45 (2007).
17. Chen, W. L., Li, Z. L., Gu, D. & Liu, Q. Y. Species summary of the genus *Liriomyza* in China (Diptera: Agromyaidae). *J. Southwest Univ.* **29**, 154–158 (2007).
18. Kato, S. A. preliminary report on a survey of agricultural pests in Chahar, Suiyuan and northern Shansi. *Peking Natural History* 11–36 (1949).
19. Hering, E. M. D. L. der Agromyziden (Diptera) III. Tijd-schr. *Ent Leiden* **100**, 73–94 (1957).
20. Scheffer, S. J., Winkler, I. S. & Wiegmann, B. M. Phylogenetic relationships within the leaf-mining flies (Diptera: Agromyzidae) inferred from sequence data from multiple genes. *Mol. Phylogenet. Evol.* **42**, 756–775 (2007).
21. Yang, F., Du, Y. Z., Wang, L. P., Cao, J. M. & Yu, W. W. The complete mitochondrial genome of the leafminer *Liriomyza sativae* (Diptera: Agromyzidae): great difference in the A + T-rich region compared to *Liriomyza trifolii*. *Gene* **485**, 7–15 (2011).
22. Wang, S. Y., Lei, Z. R., Wang, H. H., Dong, B. X. & Ren, B. Z. The complete mitochondrial genome of the leafminer *Liriomyza trifolii*, (Diptera: Agromyzidae). *Mol. Biol. Rep.* **38**, 687–692 (2011).
23. Yang, F., Du, Y. Z., Cao, J. M. & Huang, F. N. Analysis of three leafminers' complete mitochondrial genomes. *Gene* **529**, 1–6 (2013b).
24. Wang, S. Y. *et al.* The complete mitochondrial genome of *Liriomyza huidobrensis* and comparison with *L. trifolii* and *L. sativae* (Diptera: Agromyzidae). *Mitochondrial DNA Part A* **25**, 104–105 (2014).
25. Wang, Y. Z., Jin, G. H., Zhu, J. Y. & Wei, S. J. The mitochondrial genome of the garden pea leafminer *Chromatomyia horticola* (Goureau, 1851) (Diptera: Agromyzidae). *Mitochondrial DNA Part A* **27**, 2653–2655 (2016).
26. Lewis, D. L., Farr, C. L. & Kaguni, L. S. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. *Insect. Mol. Biol.* **4**, 263–278 (1995).
27. Sheffield, N. C., Song, H. J., Cameron, S. L. & Whiting, M. F. Nonstationary Evolution and Compositional Heterogeneity in Beetle Mitochondrial Phylogenomics. *Systematic Biol.* **58**, 381–394 (2009).
28. Zhang, N. X. *et al.* The complete mitochondrial genome of *Delia antiqua* and its implications in Dipteran phylogenetics. *PLoS ONE* **10**, e0139736 (2015).
29. Jermiin, L. S., Dan, G., Lowe, R. M. & Crozier, R. H. Analysis of directional mutation pressure and nucleotide content in mitochondrial cytochrome b genes. *J. Mol. Evol.* **39**, 160–173 (1994).
30. Clary, D. O. & Wolstenholme, D. R. The mitochondrial DNA molecule of *Drosophila yakuba*: Nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* **22**, 252–271 (1985).
31. Guo, Z. L. & Yuan, M. L. Research Progress of Mitochondrial genomes of Hemiptera insects. *Sci China.* **46**, 151–166 (2016).
32. Spanos, L., Koutroumbas, G., Kotsyfakis, M. & Louis, C. The mitochondrial genome of the mediterranean fruit fly, *Ceratitis capitata*. Insect. *Mol. Biol.* **9**, 139–144 (2010).
33. Yokobori, S. I. & Pääbo, S. Polyadenylation creates the discriminator nucleotide of chicken mitochondrial tRNA Tyr 1. *J. Mol. Biol.* **265**, 95–99 (1997).
34. Lu, H. F., Su, T. J., Luo, A. R., Zhu, C. D. & Wu, C. S. Characterization of the complete Mitochondrion genome of diurnal moth *Amata emma* (Butler) (Lepidoptera: Erebidae) and its phylogenetic implications. *PLoS ONE* **8**, e72410 (2013).
35. Chai, H. N., Du, Y. Z. & Zhai, B. P. Characterization of the complete mitochondrial genomes of *Cnaphalocrocis medinalis* and *Chilo suppressalis* (Lepidoptera: Pyralidae). *Int. J. Biol. Sci.* **8**, 561–579 (2012).
36. Chen, Z. T. & Du, Y. Z. Complete mitochondrial genome of *Capnia zijinshana* (Plecoptera: Capniidae) and phylogenetic analysis among stoneflies. *J. Asia-Pac. Entomol.* **20**, 305–312 (2017b).
37. Chen, Z. T., Mu, L. X., Wang, J. R. & Du, Y. Z. Complete mitochondrial genome of the citrus spiny whitefly *Aleurocanthus spiniferus* (Quaintance) (Hemiptera: Aleyrodidae): implications for the phylogeny of whiteflies. *PLoS ONE* **11**, e0161385 (2016).
38. Song, S. N., Tang, P., Wei, S. J. & Chen, X. X. Comparative and phylogenetic analysis of the mitochondrial genomes in basal hymenopterans. *Sci Rep.* **6**, 20972 (2016).
39. Chang, Y. W. *et al.* Selection and validation of reference genes for quantitative real-time PCR analysis under different experimental conditions in the leafminer *Liriomyza trifolii* (Diptera: Agromyzidae). *PLoS ONE* **12**, e0181862 (2017b).
40. Tang, X. T. *et al.* Population genetic structure and migration patterns of *Liriomyza sativae* in China: moderate subdivision and no Bridgehead effect revealed by microsatellites. *B. Entomol. Res.* **106**, 114–123 (2016).
41. Dai, L. S., Zhu, B. J. & Zhao, Y. Comparative mitochondrial genome analysis of *Eligma narcissus* and other Lepidopteran insects reveals conserved mitochondrial genome organization and phylogenetic relationships. *Sci Rep.* **6**, 26387 (2016).
42. Qian, Y. H., Wu, H. Y., Ji, X. Y., Yu, W. W. & Du, Y. Z. Mitochondrial genome of the stonefly *Kamimuria wangi* (Plecoptera: Perlidae) and phylogenetic position of plecoptera based on mitogenomes. *PLoS ONE* **9**, e86328 (2014).
43. Li, X. *et al.* Mitochondrial genomes provide insights into the phylogeny of Lauxanioidea (Diptera: Cyclorrhapha). *Int. J. Mol. Sci.* **18**, 773 (2017).
44. Rand, D. M. Thermal habit, metabolic rate and the evolution of mitochondrial DNA. *Trends Ecol Evol.* **9**, 125–131 (1994).
45. Perret, V., Florentz, C., Puglisi, J. D. & Giegé, R. Effect of conformational features on the aminoacylation of tRNAs and consequences on the permutation of tRNA specificities. *J. Mol. Biol.* **226**, 323–333 (1992).
46. Cawthorn, D. M., Steinman, H. A. & Witthuhn, R. C. Evaluation of the 16S and 12S rRNA genes as universal markers for the identification of commercial fish species in South Africa. *Gene* **491**, 40–48 (2012).
47. Ketchum, H. R., Teel, P. D., Coates, C. J., Strey, O. F. & Longnecker, M. T. Genetic variation in 12S and 16S mitochondrial rDNA genes of four geographically isolated populations of *Gulf Coast ticks*. *J. Med. Entomol.* **46**, (482–489. Ixodidae), Acari, 2009).
48. Zhou, Z. Z., Huang, Y. H. & Shi, F. S. The mitochondrial genome of *Ruspolia dubia* (Orthoptera: Conocephalidae) contains a short A + T-rich region of 70 bp in length. *Genome* **50**, 855–866 (2007).
49. Zhang, D. X. & Hewitt, G. M. Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.* **25**, 99–120 (1997).
50. Zhang, D. X., Szymura, J. M. & Hewitt, G. M. Evolution and structural conservation of the control region of insect mitochondrial DNA. *J. Mol. Evol.* **40**, 382–391 (1995).

51. Clary, D. O. & Wolstenholme, D. R. Drosophila mitochondrial DNA: conserved sequences in the A + T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. *J. Mol. Evol.* **25**, 116–125 (1987).
52. Taylor, M. F., Mckechnie, S. W., Pierce, N. & Kreitman, M. The lepidopteran mitochondrial control region: structure and evolution. *Mol. Biol. Evol.* **10**, 1259–1272 (1993).
53. Duenas, J. C. R., Panzetta-Dutari, G. M., Blanco, A. & Gardenal, C. N. Restriction fragment-length polymorphism of the mtDNA A + T-rich region as a genetic marker in *Aedes aegypti* (Diptera: Culicidae). *Ann. Entomol. Soc. Am.* **95**, 352–358 (2002).
54. Wei, S. J., Pu, T., Zheng, L. H., Shi, M. & Chen, X. The complete mitochondrial genome of *Evania appendigaster* (Hymenoptera: Evaniidae) has low A + T content and a long intergenic spacer between atp8 and atp6. *Mol. Biol. Rep.* **37**, 1931–1942 (2010).
55. Grabowski, M., Grater, W. D. & Stuck, K. C. A novel polymorphic mtDNA marker for population studies of the pink shrimp, *Farfantepenaeus duorarum* (Crustacea, Penaeidae). *Oceanologia* **46**, 147–151 (2004).
56. Wan, X. *et al.* Population genetic structure of the migratory rice leaf roller, *Cnaphalocrocis medinalis* (Lepidoptera: Pyralidae), inferred from the mitochondrial A + T-rich region and nuclear ITS2 sequences. *Genet. Mol. Res.* **10**, 273–294 (2011).
57. Johns, G. C. & Avise, J. C. A comparative summary of genetic distances in the vertebrates from the mitochondrial cytochrome b gene. *Mol. Biol. Evol.* **15**, 1481 (1998).
58. Hebert, P. D. N., Ratnasingham, S. & Waard, J. R. D. Barcoding animal life: cytochrome c oxidase subunit 1 divergences among closely related species. *Proc. Biol. Sci.* **270**(Suppl), 96–99 (2003).
59. Meier, R., Shiyang, K., Vaidya, G. & Ng, P. K. DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. *Systematic. Biol.* **55**, 715–728 (2006).
60. Wang, L., Zhang, Y. & Jiang, Y. Bionomics of *Liriomyza chinensis* (Kato). *Entomol Knowledge.* **37**, 214–217 (2000).
61. Zhang, N. X., Zhang, Y. J., Yu, G. & Chen, B. Structure characteristics of the mitochondrial genomes of Diptera and design and application of universal primers for their sequencing. *Acta Entomol Sin.* **56**, 398–407 (2013).
62. Larkin, M. A. *et al.* Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948 (2007).
63. Tamura, K., Stecher, G., Peterson, D., Filipski, A. & Kumar, S. MEGA6: Molecular evolutionary genetics analysis Version 6.0. *Mol. Biol. Evol.* **30**, 2725–2729 (2013).
64. Perna, N. T. & Kocher, T. D. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. *J. Mol. Evol.* **41**, 353–358 (1995).
65. Librado, P. & Rozas, J. DnaSPv5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451–1452 (2009).
66. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A program for improved detection of transfer RNA genes in genomic sequence. *Nucleic. Acids. Res.* **25**, 955–964 (1997).
67. Huelsenbeck, J. P. & Ronquist, F. MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755 (2001).
68. Kimura, M. A simple method for estimating evolutionary rate of base substitution through comparative studies of nucleotide sequences. *J. Mol. Evol.* **16**, 111–120 (1979).

Acknowledgements

Author Contributions

J.Y.C., Y.W.C. and Y.Z.D. designed the experiments; Y.W.C. and S.Z.Z. performed the experiments; J.Y.C., M.X.L and Y.W.C. analyzed the data and wrote the paper. All authors have read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-27213-7.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.