

# SCIENTIFIC REPORTS

OPEN

## Comparative Mitochondrial Genome Analysis of *Eligma narcissus* and other Lepidopteran Insects Reveals Conserved Mitochondrial Genome Organization and Phylogenetic Relationships

Received: 01 February 2016

Accepted: 29 April 2016

Published: 25 May 2016

Li-Shang Dai\*, Bao-Jian Zhu\*, Yue Zhao, Cong-Fen Zhang & Chao-Liang Liu

In this study, we sequenced the complete mitochondrial genome of *Eligma narcissus* and compared it with 18 other lepidopteran species. The mitochondrial genome (mitogenome) was a circular molecule of 15,376 bp containing 13 protein-coding genes (PCGs), 22 transfer RNA (tRNA) genes, two ribosomal RNA (rRNA) genes and an adenine (A) + thymine (T) — rich region. The positive AT skew (0.007) indicated the occurrence of more As than Ts. The arrangement of 13 PCGs was similar to that of other sequenced lepidopterans. All PCGs were initiated by ATN codons, except for the *cytochrome c oxidase subunit 1* (*cox1*) gene, which was initiated by the CGA sequence, as observed in other lepidopterans. The results of the codon usage analysis indicated that *Asn*, *Ile*, *Leu*, *Tyr* and *Phe* were the five most frequent amino acids. All tRNA genes were shown to be folded into the expected typical cloverleaf structure observed for mitochondrial tRNA genes. Phylogenetic relationships were analyzed based on the nucleotide sequences of 13 PCGs from other insect mitogenomes, which confirmed that *E. narcissus* is a member of the Noctuidae superfamily.

The *Ailanthus defoliator* *Eligma narcissus* is a member of the family Noctuidae (superfamily Noctuidae), which has spread all over China. This insect pest completes three or four generations every year. It is an important species that is harmful to many plants of economic importance, such as *Ailanthus altissima*, *Amygdalus persica* L., *Toona sinensis*, and others<sup>1</sup>. Many studies have been conducted on the biology and chemical pest control of *E. narcissus*<sup>1,2</sup>. However, no research has been conducted on its mitochondrial genome.

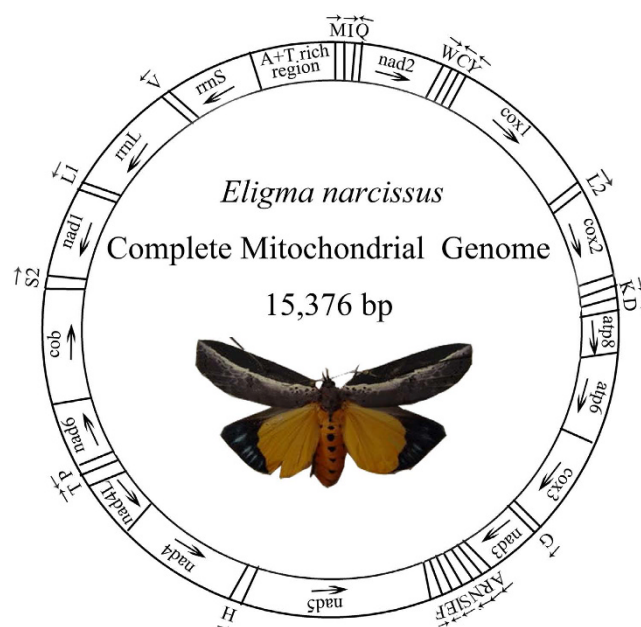
Insect mitochondrial DNA (mtDNA) has been widely used for species identification and for population genetics and molecular evolution studies due to the low or lack of sequence recombination and its maternal inheritance<sup>3</sup>. In most insects, mtDNA is composed of small, double strands: an L-strand, which is biased in favor of As and Cs, and an H-strand, which has an abundance of Ts and Gs<sup>4</sup>. Moreover, the mtDNA is a relatively conserved circular molecule of 14–19 kb in length<sup>5</sup> and includes 13 protein-coding genes (PCGs: *atp6*, *atp8*, *cox1*, *cox2*, *cox3*, *cob*, *nad1*, *nad2*, *nad3*, *nad4*, *nad5*, *nad6*, and *nad4L*), large and small ribosomal RNA (*rrnL* and *rrnS*) genes, 22 transfer RNA (tRNA) genes and a large non-coding element called the A+T-rich region that contains sequences essential for the initiation of transcription and gene replication<sup>6</sup>.

Lepidoptera (moths and butterflies) is the second largest order in Insecta, accounting for more than 155,000 insect species<sup>7</sup>. Although the superfamily Noctuidae is the largest superfamily in Lepidoptera, only few mitogenomes of Noctuidae are available in GenBank (Table 1). In the present study, we determined the complete mitogenome sequence of *E. narcissus* and compared the nucleotide composition, codon usage, tRNA secondary structure and A+T rich region with those of other lepidopteran species. Furthermore, the concatenated

College of Life Sciences, Anhui Agricultural University, 130 Changjiang West Road, Hefei 230036, China. \*These authors contributed equally to this work. Correspondence and requests for materials should be addressed to C.-L.L. (email: cyschx@163.com)

Superfamily	Family	Species	Acc. number	Reference
Noctuoidea	Noctuidae	<i>Eligma narcissus</i>		This study
		<i>Agrotis ipsilon</i>	KF163965	32
		<i>Helicoverpa armigera</i>	GU188273	12
		<i>Helicoverpa assulta</i>	KR149448	unpublished
		<i>Sesamia inferens</i>	NC_015835	unpublished
	Erebidae	<i>Hyphantria cunea</i>	GU592049	33
	Notodontidae	<i>Phalera flavescens</i>	JF440342	34
	Lymantriidae	<i>Lymantria dispar</i>	NC_012893	unpublished
Bombycoidea	Bombycidae	<i>Bombyx mori</i>	NC_002355	unpublished
	Saturniidae	<i>Actias selene</i>	NC_018133	9
	Sphingidae	<i>Manduca sexta</i>	EU286785	35
Pyraloidea	Crambidae	<i>Elophila interruptalis</i>	NC_021756	36
	Pyralidae	<i>Corcyra cephalonica</i>	NC_016866	unpublished
Tortricoidea	Tortricidae	<i>Acleris fimbriana</i>	NC_018754	unpublished
		<i>Adoxophyes orana</i>	NC_021396	37
Gelechioidea	Oecophoridae	<i>Promalactes suzukiella</i>	KM875542	38
Papilionoidea	Papilionidae	<i>Papilio syfanius</i>	NC_023978	39
		<i>Teinopalpus aureus</i>	NC_014398	unpublished
	Nymphalidae	<i>Polyura nepenthes</i>	NC_026073	unpublished

**Table 1. Lepidopteran mitogenomes used in this study.**



**Figure 1.** Circular map of the mitogenome of *E. narcissus*. *cox1*, *cox2* and *cox3* refer to the cytochrome *c* oxidase subunits. *cob* refers to cytochrome *b*. *nad1-nad6* refer to NADH dehydrogenase components. *rrnL* and *rrnS* refer to ribosomal RNAs. tRNAs are denoted as a one-letter symbol according to the IUPAC-IUB single-letter amino acid codes. Gene names with lines indicate that these genes are located on L strand, whereas the others are located on H strand.

nucleotide sequence of 13 PCGs of *E. narcissus* was used to provide insight into the phylogenetic relationships among lepidopteran superfamilies.

## Results and Discussion

**Genome organization and base composition.** The complete mitogenome of *E. narcissus* is a circular molecule of 15,376 bp in size (Fig. 1), which is bigger than the genomes of *A. selene* (15,236 bp) and *H. armigera* (15,347 bp) but smaller than the genomes of *P. flavescens* (15,659 bp) and *M. sexta* (15,516 bp). The mitogenome of *E. narcissus* contains a remarkably conserved set of 37 genes, including 13 protein-coding genes (PCGs: *atp6*, *atp8*, *atp9*, *cox1*, *cox2*, *cox3*, *cob*, *nad1*, *nad2*, *nad3*, *nad4*, *nad5*, *nad6*, and *nad4L*), large and small ribosomal RNA (*rnrL* and *rnrS*) genes, 22 transfer RNA (tRNA) genes and a large non-coding element called the A+T-rich region.

Species	Size (bp)	A%	G%	T%	C%	A+T %	ATskewness	GCskewness
Whole genome								
<i>E. narcissus</i>	15,376	40.78	7.68	40.21	11.33	80.99	0.007	−0.192
<i>H. armigera</i>	15,347	40.54	7.69	40.43	11.34	80.97	0.001	−0.192
<i>O. lunifer</i>	15,593	40.09	7.56	37.75	14.60	77.84	0.030	−0.318
<i>P. flavescent</i>	15,659	40.07	7.87	40.80	11.26	80.87	−0.009	−0.177
<i>B. mandarina</i>	15,682	43.11	7.40	38.48	11.01	81.59	0.057	−0.196
<i>A. selene</i>	15,236	38.54	8.05	40.37	13.03	78.91	−0.023	−0.236
<i>M. sexta</i>	15,516	40.67	7.46	41.11	10.76	81.79	−0.005	−0.181
PCG								
<i>E. narcissus</i>	11,190	40.68	8.40	37.72	13.20	78.42	0.038	−0.222
<i>H. armigera</i>	11,203	34.25	10.62	45.18	9.95	79.43	−0.138	0.033
<i>O. lunifer</i>	11,266	32.47	12.08	43.26	12.19	75.73	−0.142	−0.004
<i>P. flavescent</i>	11,206	39.40	8.90	39.56	12.15	78.96	−0.002	−0.154
<i>B. mandarina</i>	11,196	42.83	8.26	37.04	11.87	79.87	0.072	−0.179
<i>A. selene</i>	11,231	37.93	8.74	39.44	13.89	77.37	−0.020	−0.228
<i>M. sexta</i>	11,178	40.40	8.22	39.89	11.49	80.29	0.006	−0.166
tRNA								
<i>E. narcissus</i>	1450	41.52	8.00	40.48	10.00	82.00	0.013	−0.111
<i>H. armigera</i>	1473	41.41	8.15	40.39	10.05	81.81	0.012	−0.104
<i>O. lunifer</i>	1666	41.78	7.32	39.86	11.04	81.63	0.023	−0.202
<i>P. flavescent</i>	1474	41.66	7.80	40.64	9.91	82.29	0.012	−0.119
<i>B. mandarina</i>	1472	41.78	7.81	39.95	10.46	81.73	0.022	−0.145
<i>A. selene</i>	1459	40.37	8.16	40.23	11.24	80.60	0.002	−0.159
<i>M. sexta</i>	1470	41.09	8.16	40.68	10.07	81.77	0.005	−0.145
rRNAs								
<i>E. narcissus</i>	2122	41.89	5.00	42.60	10.51	84.50	−0.008	−0.355
<i>H. armigera</i>	2189	41.75	4.89	43.40	9.96	85.15	−0.019	−0.341
<i>O. lunifer</i>	2157	41.96	4.82	40.19	13.03	82.15	0.022	−0.460
<i>P. flavescent</i>	2198	41.31	4.73	44.04	9.92	85.35	−0.032	−0.354
<i>B. mandarina</i>	2134	43.86	4.78	41.05	10.31	84.91	0.028	−0.366
<i>A. selene</i>	2126	39.93	4.99	43.79	11.29	83.73	−0.046	−0.387
<i>M. sexta</i>	2168	41.37	4.84	44.05	9.73	85.42	−0.031	−0.335
A+T-rich region								
<i>E. narcissus</i>	434	47.47	1.15	49.08	2.30	96.54	−0.017	−0.332
<i>H. armigera</i>	328	44.51	1.22	50.61	3.66	95.12	−0.064	−0.500
<i>O. lunifer</i>	319	44.5	1.6	48.9	5.0	93.4	−0.047	−0.524
<i>P. flavescent</i>	541	42.14	2.22	49.72	5.91	91.87	−0.083	−0.454
<i>B. mandarina</i>	484	46.49	2.69	47.93	2.89	94.42	−0.015	−0.036
<i>A. selene</i>	339	43.07	5.90	44.84	6.19	87.91	−0.020	−0.024
<i>M. sexta</i>	324	45.06	1.54	50.31	3.09	95.37	−0.055	−0.419

**Table 2. Composition and skewness in the lepidopteran mitogenomes.**

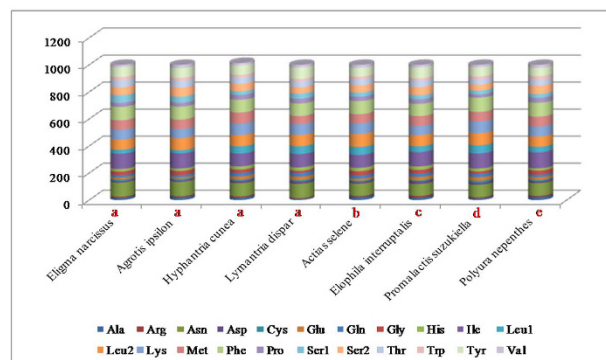
The gene order is the same as in other sequenced Noctuidae mitogenomes<sup>8–10</sup>, with a *trnM-trnI-trnQ* order, which is different from the ancestral *trnI-trnQ-trnM* order<sup>11</sup>. The nucleotide composition (A: 40.78%, T: 40.21%, G: 7.68%, C: 11.33%) of the *E. narcissus* mitogenome is biased toward A+T nucleotides (80.99%), which is a higher percentage than that of six other lepidopterans but a lower percentage than that of *M. sexta* (81.79%) (Table 2). The positive AT skew (0.007) indicates the occurrence of more As than Ts, similar to some lepidopterans such as *H. armigera* (0.001)<sup>12</sup>, *O. lunifer* (0.030)<sup>13</sup>, and the Chinese *B. mandarina* (0.057)<sup>14</sup>.

In the mitogenome of *E. narcissus*, there are a total of 34 bp that contain overlapping genes. The 34 bp gene overlap in 11 positions of the *E. narcissus* mitogenome ranges in size from 1 to 8 bp, and the longest overlapping region is present between *trnW* and *trnC*. A total of 192 bp of intergenic spacers are dispersed in 16 regions and range in size from 1 to 52 bp, with the longest spacer present between *trnQ* and *nad2*. The length of the intergenic spacers is considerably longer than that of *A. selene* (137 bp over 13 regions) but shorter than that of *O. lunifer* (371 bp over 20 regions)<sup>9,13</sup>.

**Protein-coding genes and codon usage.** The 13 PCGs of *E. narcissus* are 11,181 bp in length, accounting for 72.72% of the entire mitochondrial genome. Like other lepidopterans, all 13 PCGs in the *E. narcissus* mitogenome have ATN as their start codon, except for the *cox1* gene, which uses CGA instead. The start codon of the *cox1* is rare in insect mtDNA; the canonical codons TTG, ACG, TTA and TTAG are often reported as the *cox1* start codons<sup>15–18</sup>. Ten of the 13 PCGs harbor the complete stop codon TAA, whereas the other three possess the

Gene	Direction	Location	Size	Anticodon	Start codon	Stop codon	Intergenic nucleotides*
<i>trnM</i>	F	1–67	67	CAT	–	–	0
<i>trnI</i>	F	68–132	65	GAT	–	–	–3
<i>trnQ</i>	R	130–198	69	TTG	–	–	52
<i>nad2</i>	F	251–1264	1014	–	ATT	TAA	1
<i>trnW</i>	F	1266–1334	69	TCA	–	–	–8
<i>trnC</i>	R	1327–1394	68	GCA	–	–	2
<i>trnY</i>	R	1397–1461	65	GTA	–	–	2
<i>cox1</i>	F	1464–2997	1534	–	CGA	T	–1
<i>trnL2(UUR)</i>	F	2995–3061	67	TAA	–	–	0
<i>cox2</i>	F	3062–3743	682	–	ATG	T	0
<i>trnK</i>	F	3744–3814	71	CTT	–	–	7
<i>trnD</i>	F	3822–3887	66	GTC	–	–	0
<i>atp8</i>	F	3888–4052	165	–	ATC	TAA	–7
<i>atp6</i>	F	4046–4723	678	–	ATG	TAA	–1
<i>cox3</i>	F	4723–5511	789	–	ATG	TAA	2
<i>trnG</i>	F	5514–5578	65	TCC	–	–	0
<i>nad3</i>	F	5579–5932	354	–	ATT	TAA	42
<i>trnA</i>	F	5975–6039	65	TGC	–	–	2
<i>trnR</i>	F	6042–6104	63	TCG	–	–	6
<i>trnN</i>	F	6111–6175	65	GTT	–	–	5
<i>trnS1(AGN)</i>	F	6181–6251	71	GCT	–	–	–1
<i>trnE</i>	F	6251–6315	65	TTC	–	–	4
<i>trnF</i>	R	6320–6386	67	GAA	–	–	–2
<i>nad5</i>	R	6391–8132	1742	–	ATT	TAA	–2
<i>trnH</i>	R	8131–8197	67	GTG	–	–	–2
<i>nad4</i>	R	8196–9531	1336	–	ATA	TA	–6
<i>nad4L</i>	R	9526–9812	287	–	ATT	TAA	23
<i>trnT</i>	F	9836–9899	64	TGT	–	–	0
<i>trnP</i>	R	9900–9963	64	TGG	–	–	7
<i>nad6</i>	F	9971–10,507	537	–	ATA	TAA	–1
<i>cob</i>	F	10,507–11,654	1148	–	ATG	TAA	3
<i>trnS2(UCN)</i>	F	11,658–11,723	66	TGA	–	–	30
<i>nad1</i>	R	11,754–12,683	930	–	ATT	TAA	4
<i>trnL1(CUN)</i>	R	12,688–12,754	67	TAG	–	–	0
<i>rrnL</i>	R	12,755–14,089	1335	–	–	–	0
<i>trnV</i>	R	14,090–14,155	66	TAC	–	–	0
<i>rrnS</i>	R	14,156–14,942	787	–	–	–	0
CR		14,943–15,376	434	–	–	–	–

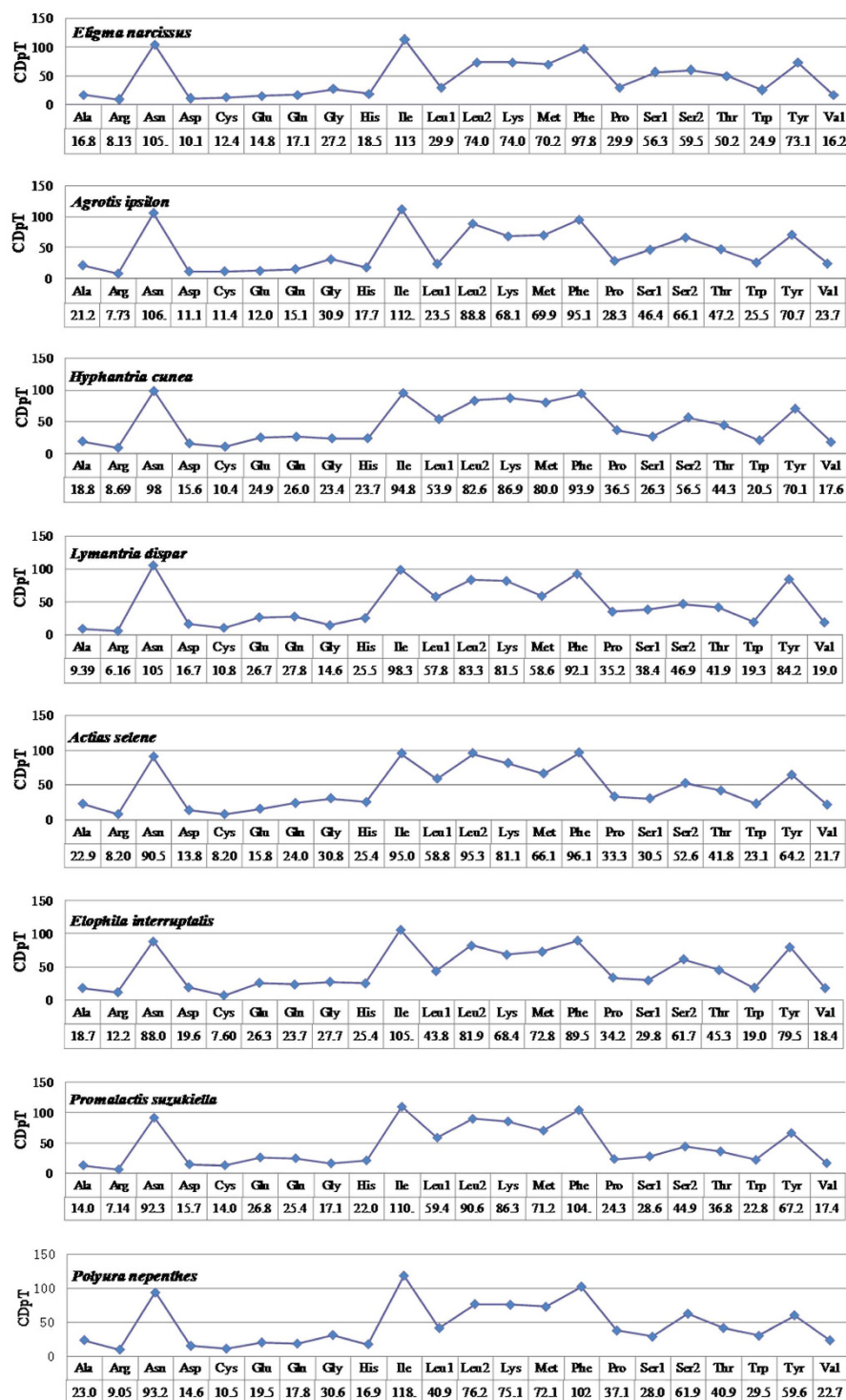
**Table 3.** Annotation and gene organization of the *Eligma narcissus* mitogenome.



**Figure 2.** Comparison of codon usage in mitochondrial genomes in *Lepidoptera*. Lowercase letters (a–e) above the species name represent the superfamily that the species belong to ((a) Noctuidae, (b) Bombycoidea, (c) Pyraloidea, (d) Gelechioidea, (e) Papilionoidea).

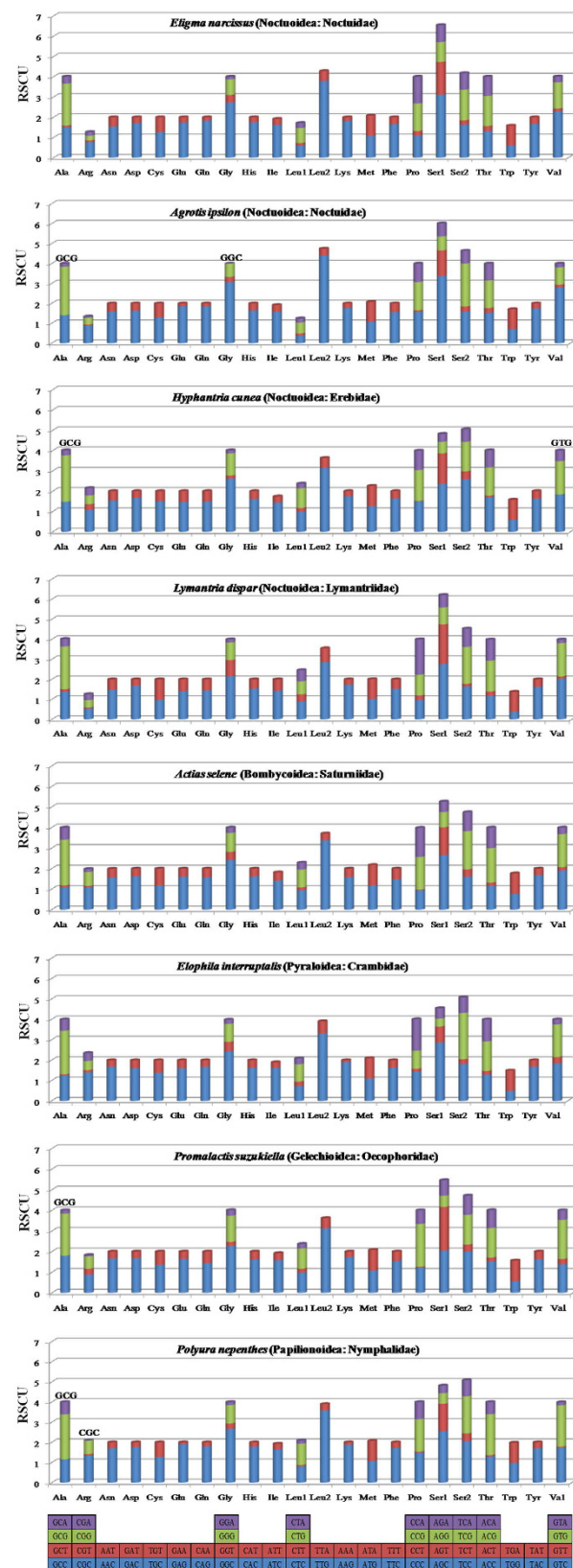
incomplete termination codon T for *cox1* and *cox2*, and TA for *nad4* (Table 3). Incomplete stop codons are commonly observed in lepidopteran species<sup>9,19</sup>.

The 13 PCGs of the *E. narcissus* mitogenome contain 3727 codons in total, which is within the range of 3687 in *H. armigera* and 3742 in *Agrotis ipsilon*. The complete nucleotide sequences of seven lepidopteran insects were



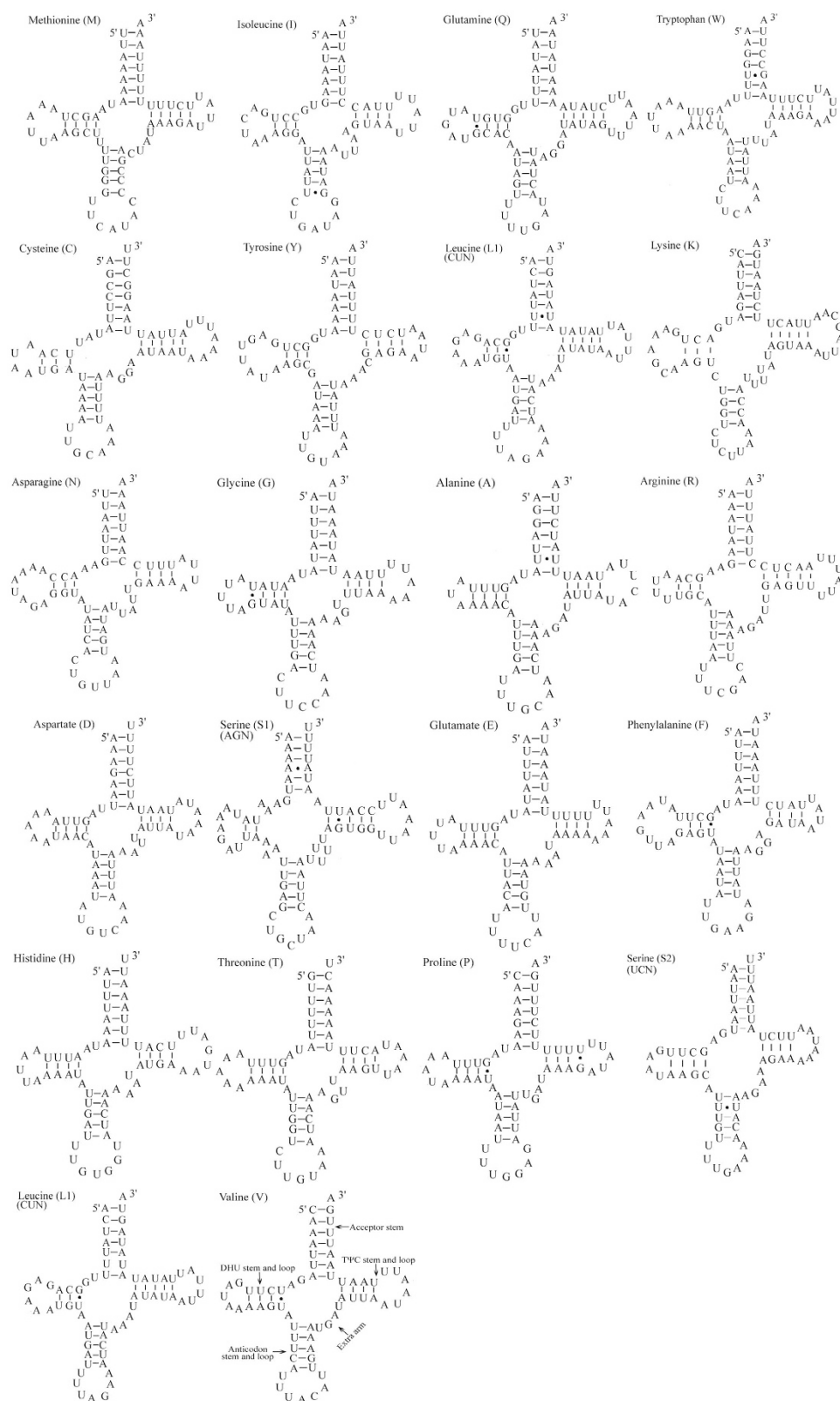
**Figure 3.** Codon distribution in Lepidoptera. CDspT, codons per thousand codons.

downloaded from GenBank to investigate the codon usage among lepidopterans. These mitogenomes are divided into five superfamilies: four species belong to Noctuidae, and the others belong to Bombycoidea, Pyraloidea, Gelechioidea, and Papilionoidea (Fig. 2). We looked at the behavior of codon families in the PCGs and found that *Asn*, *Ile*, *Leu*, *Phe*, and *Tyr* are the most abundant amino acids in the *E. narcissus* mitogenome (Fig. 3). RSCU for Lepidoptera is shown in Fig. 4. All possible codons are present in the PCGs of the *E. narcissus* mitogenome, whereas some codons, such as GCG, GGC, GTG and CGC, are not found in four other species. Previous research indicates that codons with high G and C content are likely not to be favored, a phenomenon that is found in some lepidopteran insects<sup>7,20</sup>.



**Figure 4. The mitochondrial genome relative synonymous codon usage (RSCU) across five superfamilies in Lepidoptera.** Codon families are provided on the X axis. Codons indicated above the bar are not present in the mitogenome.





**Figure 5.** Predicted secondary cloverleaf structures of *E. narcissus* tRNA genes.

**Ribosomal RNA genes and transfer RNA genes.** As with all other insect mitogenome sequences, there were two rRNAs in *E. narcissus* with a total length of 2122 bp. The large ribosomal gene (*rrnL*), located between *trnL1*(CUN) and *trnV*, had a length of 1335 bp, whereas the small one (*rrnS*), located between *trnV* and the A+

*rrnS* - 14,943 -

ATTTATATGTAATTTTTTTTAAATAGAAATTTTTTTTTTTTTTTTATATTAATATATTTATATAT  
TATTTAAATACATATATATATATATATAATATAATATTAATAATTAATATAATTTATTTATTGT  
ATATTAAAATATTTAATATAAAATTTAAATTTTAAATAATTTCTTTTTCCTTTTTCATAATAA  
TAAAATTAATAATTTAAATTCATTAATAAACCAATGTATATTAATTTAAATAATAAAATTAATA  
TAATTAATATTAATTTTTTTTAAATAAAATATAGTATTATAAAATATAAATTCATTTAAATTTTAAT  
ATATATATATAAATATAATTTATTTTATAATTTTAAATAAAATTAATTTATTTATATATTA  
AACCAATTTTAAAAAAATTACATATAAATAAAAAA - 15,376 - *trnM*

**Figure 6.** Features present in the A+T-rich region of *E. narcissus*. The motif, poly-T stretch, microsatellite T/A repeat sequences and poly-A stretch are colored in red, green, blue and purple, respectively.

	<i>rrnS</i>		<i>D-loop region</i>	
<i>E. narcissus</i>	5' ATTTATA---	ATAGA	TTTTTTTTTTTTTTTTTTT	ATATTAATA-3'
<i>H. armigera</i>	5' ATATGTA---	ATAGA	TTTTTTTTTTTTTTTTTTT	ATATTAATA-3'
<i>H. cunea</i>	5' TTATATA---	ATAGA	TTTTTTTTTTTTTTTTTTT	ATATTTAAA-3'
<i>P. flavescentis</i>	5' TTATATA---	ATAGA	TTTTTTTTTTTTTTTTTTT	AATTTTAAT-3'
<i>L. dispar</i>	5' TTATAAT---	ATAGA	TTTTTTTTTTTTTTTTTTT	ATATATTAA-3'
<i>B. mori</i>	5' ATTTAAT---	ATAGA	TTTTTTTTTTTTTTTTTTT	ATATTAATT-3'
<i>M. sexta</i>	5' AATAGTT---	ATAGA	TTTTTTTTTTTTTTTTTTT	ATATTAATA-3'
<i>A. fimbriana</i>	5' ATACAAA---	ATAGA	TTTTTTTTTTTTTTTTTTT	ATATTGAAA-3'
<i>P. suzukiella</i>	5' TTATATA---	ATAGA	TTTTTTTTTTTTTTTTTTT	ATATAAAAA-3'
<i>P. nepenthes</i>	5' CTATTTA---	ATAGA	TTTTTTTTTTTTTTTTTTT	ATATTAATA-3'

**Figure 7.** Sequence alignment of the partial *D-loop* region of 10 moth species. The boxed nucleotides indicate the 'ATAGA' conserved motif.

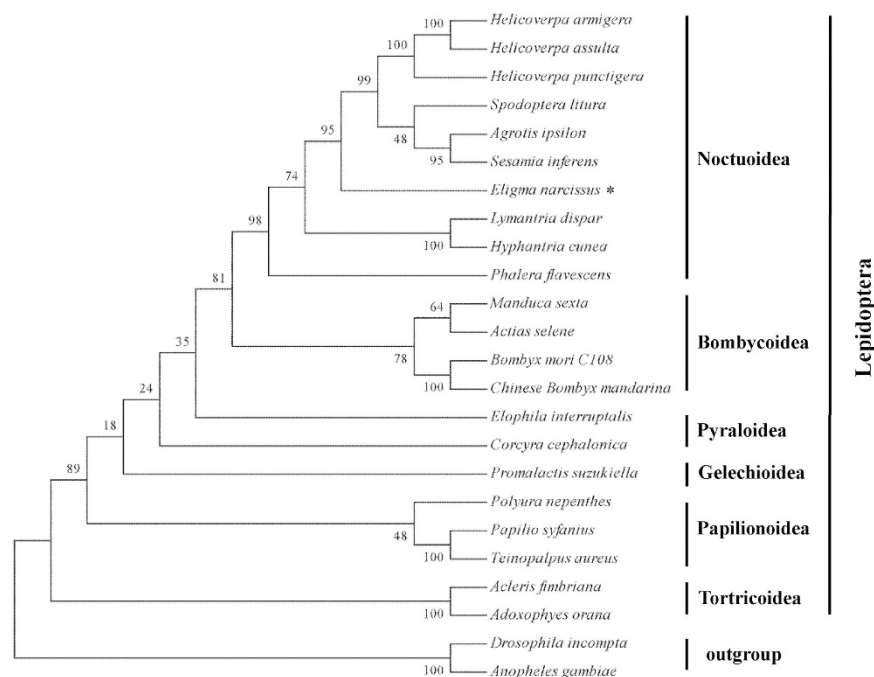
T-rich region, had a length of 786 bp. The negative AT skew ( $-0.335$ ) indicated the occurrence of more Ts than As. The A+T content of the two rRNA genes was 84.50%, which was within the range of 82.15% in *O. lunifer* and 85.42% in *M. sexta* (Table 2).

The *E. narcissus* mitogenome contained 22 tRNAs interspersed throughout the entire genome and ranging in size from 63 to 71 bp, which comprised 1450 bp of the total mitogenome overall. Of these genes, 14 were encoded by the H-strand, and eight were encoded by the L-strand. The predicted structures of the tRNAs are shown in Fig. 5. The A+T content of the 22 tRNAs was 82.00%, with a positive AT skew (0.013). A total of 11 mismatched base pairs were identified in the tRNAs of *E. narcissus*, three of them being mismatched base pairs (1 A-A and 2 U-U) and eight being G-U wobble pairs. In many insect mitogenomes, the *trnS1* (AGN) gene has an unusual secondary structure lacking a stable stem-loop structure in the DHU arm<sup>9,19</sup>; however, we found that all the tRNA genes in *E. narcissus* could be folded into the expected typical cloverleaf secondary structure observed in mitochondrial tRNA genes. All of the secondary structures were drawn using the RNA structure program.

**The A+T-rich region.** The A+T-rich region, known for the initiation of replication in vertebrates and invertebrates, is located between the *rrnS* and *trnM* genes in *E. narcissus*. This region is 434 bp in length, which is longer than the regions in *A. selene* (339 bp) and *M. sexta* (324 bp) but shorter than the regions in Chinese *B. mandarina* (484 bp) and *P. flavescentis* (541 bp) (Table 2). The region contains the highest A+T content (96.54%) in the mitogenome. There are some conserved structures observed in the *E. narcissus* A+T-rich region, including the motif 'ATAGA' followed by an 19 bp poly-T stretch, the motif 'ATTTA' followed by a microsatellite-like (AT) motif<sup>6</sup> and a poly-A element upstream of the *trnM* gene (Fig. 6). The poly-T is located upstream of the *rrnS* 5'-end and is preceded by the 'ATAGA' motif, a structural feature that is found in the majority of lepidopteran insects (Fig. 7). Previous research indicates that the poly-T element may be involved in controlling transcription and replication initiation<sup>21</sup>. Two repeat elements are found in the A+T-rich region of the *E. narcissus* mitogenome. Some lepidopteran insects also have repeat elements in the A+T-rich region. The A+T-rich region of *S. longistyla* harbors two repeat elements<sup>22</sup>, whereas the *C. medinalis* and *C. suppressalis* mitogenomes contain a duplicated 25 bp repeat element and a duplicated 36 bp repeat element, respectively<sup>20</sup>.

**Phylogenetic analyses.** In this study, the mitogenomes of 18 lepidopteran species representing six lepidopteran superfamilies (Noctuoidea, Bombycoidea, Pyraloidea, Tortricidae, Geometrioidea, Papilionoidea) were downloaded from GenBank. The phylogenetic relationships among the superfamilies of Lepidoptera were reconstructed based on concatenated nucleotide sequences of 13 PCGs by using the maximum likelihood (ML) method (Fig. 8). The phylogenetic analyses show that *E. narcissus* was within Noctuidae. Noctuidae is closely related to Bombycoidea. Although these results do not conflict with other published trees<sup>7,23</sup>, more studies of a variety of species are needed to provide further insights into the relationships among Noctuidae species.





**Figure 8. Phylogenetic analysis of lepidopteran insects.** The phylogenetic tree was constructed using the maximum likelihood (ML) method, and bootstrap values (1000 repetitions) of the branches were indicated. *D. incompta* (NC\_025936) and *A. gambiae* (NC\_002084) were used as outgroups.

Primer	Sequence (5' → 3')
F1	ATAGAATTAACTATATCCTATA
R1	ATCTAAGAGATGTTCTACTAT
F2	TACAATTTATCGCTTATAACTCA
R2	TATTAGGAGTTATATATGAGTC
F3	TGAAGTTATGAATATTCAGATT
R3	ATTATTTAGGTGTCGAATTAAA
F4	AGCTGCTAACTTAATTTTAGT
R4	CTGTTTCAGCTTAGTTCATTCA
F5	TTACCATTAAAATTATAACATCT
R5	AGATTAATTAATAGATTAATTAG
F6	TCAAATTATTCTAAAATCAATCT
R6	CAAATCAAAGAATAGTTAATA
F7	CGAACTAACTCTCTCTCACTC
R7	ATATGTACATATTGCCGTCGCT
F8	TAGAAACACTTTCCAGTACCTC
R8	ATTTTAAATTATTAGGTGAAATT
F9	ATGAACATAAATACCGCCAAAT
R9	ATTAATATTGATGAGTTGATTAT

**Table 4. List of primers used to amplify the mitogenome of *Eligma narcissus*.**

## Materials and Methods

**Sample collection and DNA extraction.** No specific permits were required for the insect collection necessary for this study. *E. narcissus* larvae were collected in Hefei city, China. Specimens identified as *E. narcissus* were preserved in 100% ethanol and stored at  $-80^{\circ}\text{C}$ . Total genomic DNA was extracted with the Aidlab Genomic DNA Extraction Kit (Aidlab Co., Beijing, China) according to the manufacturer's instructions. The DNA was examined using 1% agarose gels and was then used for PCR amplification of the complete mitogenome.

**PCR amplification, cloning and sequencing.** To amplify the entire mitogenome of *E. narcissus*, nine pairs of primers were designed based on the known mitogenomes of lepidopteran species (Beijing Sunbiotech Co., Ltd., Beijing, China) (Table 4). PCR reactions were carried out in a  $50\mu\text{l}$  reaction volume, including  $5\mu\text{l}$  of

10× long Taq buffer (Mg<sup>2+</sup> plus), 5 µl of dNTP (20 mM), 1.5 µl of DNA template from a single specimen, 2 µl of each primer (10 µM), 35 µl of sterilized distilled water and 0.5 µl (1 unit) of long Taq (Aidlab Co., Beijing, China). The conditions for PCR amplification were as follows: an initial denaturation for 4 min at 94 °C, followed by 35 cycles of 30 s at 94 °C, 40 s at 46–57 °C (depending on primer combination) and 1–3 min (depending on putative length of the fragments) at 72 °C, as well as a final extension step of 72 °C for 10 min. All PCR reactions were performed in a BIO-RAD thermal cycler.

The above PCR products were resolved by agarose gel electrophoresis (1% w/v) and purified using a DNA gel extraction kit (TaKaRa Co., Dalian, China). The purified PCR fragments were ligated into the T-vector (TaKaRa Co., Dalian, China) and then transformed into competent *Escherichia coli* DH5α. The positive recombinant clone with an insert was sequenced at least three times (Invitrogen Co., Ltd., Shanghai, China).

**Genome assembly and gene annotation.** The *E. narcissus* final consensus mtDNA sequence was performed using the Lasergene software package (DNASTAR Inc. Madison, USA). Sequence annotation was performed using the Online Blast Tool of the NCBI web site (<http://blast.ncbi.nlm.nih.gov/Blast>). The overlapping regions and intergenic spacers between genes were counted manually. The base composition of nucleotide sequences was described by skewness and was measured according to the following formulas: AT skew =  $[A - T] / [A + T]$ , GC skew =  $[G - C] / [G + C]$ . The A+T content and relative synonymous codon usage (RSCU) values were calculated using MEGA 5.0<sup>24</sup>. Transfer RNA genes were identified using the tRNAscan-SE program software available online at <http://lowelab.ucsc.edu/tRNAscan-SE/><sup>25</sup>. The secondary structures of tRNA genes were analyzed by comparison with the nucleotide sequences of other insect tRNA sequences. The nucleotide sequences of the PCGs were translated on the basis of the invertebrate mtDNA genetic code. Alignments of PCGs from other lepidopteran mitogenome sequences were performed using ClustalX software<sup>26</sup>. The entire A+T-rich region was subjected to a search for tandem repeats using the Tandem Repeats Finder program (<http://tandem.bu.edu/trf/trf.html>)<sup>27</sup>.

**Phylogenetic analysis.** Along with the *E. narcissus* mitochondrial genome, 21 available insect mitogenomes were downloaded from GenBank to illustrate the phylogenetic relationships among lepidopteran insects. The mitogenomes of *Drosophila incompta* (NC\_025936)<sup>28</sup> and *Anopheles gambiae* (NC\_002084)<sup>29</sup> were downloaded and used as outgroups. The nucleotide and putative amino acid regions for each of the 13 PCGs were aligned using ClustalW, as implemented in the program MEGA. To select the conserved regions of the putative amino acids, all alignments were analyzed with the program Gblock 0.91b using default settings<sup>30</sup>. Phylogenetic analysis was conducted using the maximum likelihood (ML) method, as implemented in the MEGA 5.0 program<sup>31</sup>. This method was used to infer phylogenetic trees with 1000 bootstrap replicates. Substitution model selection was also conducted based on the lowest BIC scores (Bayesian Information Criterion) using MEGA 5.0. The mtREV24 + G + F model was the appropriate models for the amino acid sequence dataset.

## References

- Shao, Y. L., Sun, X. Q., Gu, L. J. & Tan, X. H. The study in *Eligma narcissus* about living habits and prevention. *Anhui Agri. Sci. Bull.* **18**, 93–94 (2012).
- Li, S. M., Li, M. L. & Wang, F. H. Comparative Analysis of Fatty Acids in Larva and Pupa of *Eligma narcissus*. *Journal of Northwest Forestry University*. **21**, 105–106 (2006).
- Lin, C. P. & Danforth, B. N. How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. *Mol. Phylogenet. Evol.* **30**, 686–702 (2004).
- Hassanin, A., Léger, N. & Deutsch, J. Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. *Syst. Biol.* **54**, 277–298 (2005).
- Boore, J. L. Animal mitochondrial genomes. *Nucleic. Acids. Res.* **27**, 1767–1780 (1999).
- Shadel, G. S. & Clayton, D. A. Mitochondrial DNA maintenance invertebrates. *Annu. Rev. Biochem.* **66**, 409–435 (1997).
- Lu, H. F., Su, T. J., Luo, A. R., Zhu, C. D. & Wu, C. S. Characterization of the complete mitochondrial genome of diurnal Moth *Amata emma* (Butler) (Lepidoptera: Erebididae) and its phylogenetic implications. *PLoS One* **8**, e72410 (2013).
- Dai, L. S., Zhu, B. J., Liu, Q. N., Wei, G. Q. & Liu, C. L. Characterization of the complete mitochondrial genome of *Bombyx mori* strain H9 (Lepidoptera: Bombycidae). *Gene* **519**, 326–334 (2013).
- Liu, Q. N., Zhu, B. J., Dai, L. S., Wei, G. Q. & Liu, C. L. The complete mitochondrial genome of the wild silkworm moth, *Actias selene*. *Gene* **505**, 291–299 (2012).
- Liu, Q. N., Zhu, B. J., Dai, L. S. & Liu, C. L. The complete mitogenome of *Bombyx mori* strain Dazao (Lepidoptera: Bombycidae) and comparison with other lepidopteran insects. *Genomics* **101**, 64–73 (2013).
- Dai, L. S. *et al.* The complete mitochondrial genome of the diamondback moth, *Plutella xylostella* (Lepidoptera: Plutellidae). *Mitochondrial DNA* **27**, 1512–1513 (2016).
- Yin, J., Hong, G. Y., Wang, A. M., Cao, Y. Z. & Wei, Z. J. Mitochondrial genome of the cotton bollworm *Helicoverpa armigera* (Lepidoptera: Noctuidae) and comparison with other lepidopterans. *Mitochondrial DNA* **21**, 160–169 (2010).
- Salvato, P., Simonato, M., Battisti, A. & Negrisol, E. The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera: Notodontidae). *BMC Genomics* **9**, 331 (2008).
- Pan, M. H. *et al.* Characterization of mitochondrial genome of Chinese wild mulberry silkworm, *Bombyx mandarina* (Lepidoptera: Bombycidae). *Sci. China C. Life Sci.* **51**, 693–701 (2008).
- Ogoh, K. & Ohmiya, Y. Complete mitochondrial DNA sequence of the sea-firefly, *Vargula hilgendorffii* (Crustacea, Ostracoda) with duplicate control regions. *Gene* **327**, 131–139 (2004).
- Sabine, L. B., Timo, S. & Stefan, P. Different methods to determine length heteroplasmy within the mitochondrial control region. *Int. J. of Legal. Med.* **118**, 274–281 (2004).
- Yamauchi, M. M., Miya, M. U. & Nishida, M. Complete mitochondrial DNA sequence of the Japanese spiny lobster, *Panulirus japonicus* (Crustacea: Decapoda). *Gene* **295**, 89–96 (2002).
- Lee, E. S. *et al.* The mitochondrial genome of the smaller tea tortix *Adoxophyes honmai* (Lepidoptera: Tortricidae). *Gene* **373**, 52–57 (2008).
- Zhu, B. J. *et al.* Characterization of the complete mitochondrial genome of *Diaphania pyralis* (Lepidoptera: Pyralidae). *Gene* **527**, 283–291 (2013).

20. 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).
21. Jiang, S. T. *et al.* Characterization of the complete mitochondrial genome of the giant silkworm moth, *Eriogyna pyretorum* (Lepidoptera: Saturniidae). *Int. J. Biol. Sci.* **5**, 351–365 (2009).
22. Chen, Z. T. & Du, Y. Z. Comparison of the complete mitochondrial genome of the stonefly *Sweltsa longistyla* (Plecoptera: Chloroperlidae) with mitogenomes of three other stoneflies. *Gene* **558**, 82–87 (2015).
23. Chai, H. N. & Du, Y. Z. The Complete Mitochondrial Genome of the Pink Stem Borer, *Sesamia inferens*, in Comparison with Four Other Noctuid Moths. *Int. J. Biol. Sci.* **13**, 10236–10256 (2012).
24. Tamura, K. *et al.* MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. *Mol Biol Evol* **28**, 2731–2739 (2011).
25. 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).
26. Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. The CLUSTAL\_X windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res.* **25**, 4876–4882 (1997).
27. Benson, G. Tandem repeats finder: A program to analyze DNA sequences. *Nucleic Acids Res.* **27**, 573–580 (1999).
28. Re, F. C. D., Wallau, L. & Robe, L. J. Characterization of the complete mitochondrial genome of flower-breeding *Drosophila incompta* (Diptera, Drosophilidae). *Genetica* **142**, 525–535 (2014).
29. Beard, C. B., Mills, D. & Collins, F. H. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.* **2**, 103–124 (1993).
30. Castresana, J. Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. *Mol. Biol. Evol.* **17**, 540–552 (2000).
31. Tamura, K. *et al.* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739 (2011).
32. Wu, Q. L., Cui, W. X. & Wei, S. J. Characterization of the complete mitochondrial genome of the black cutworm *Agrotis ipsilon* (Lepidoptera: Noctuidae). *Mitochondrial DNA* **26**, 139–140 (2015).
33. Liao, F. *et al.* The complete mitochondrial genome of the fall webworm, *Hyphantria cunea* (Lepidoptera: Arctiidae). *Int. J. Biol. Sci.* **6**, 172–186 (2010).
34. Sun, Q. Q. *et al.* Complete sequence of the mitochondrial genome of the Japanese buff-tip moth, *Phalera flavescens* (Lepidoptera: Notodontidae). *Genet. Mol. Res.* **11**, 4213–4225 (2012).
35. Cameron, S. L. & Whiting, M. F. The complete mitochondrial genome of the tobacco hornworm, *Manduca sexta* (Insecta: lepidoptera: Sphingidae), and an examination of mitochondrial gene variability within butterflies and moths. *Gene* **408**, 112–113 (2008).
36. Park, J. S., Kim, M. J., Kim, S. S. & Kim, I. Complete mitochondrial genome of an aquatic moth, *Elophila interruptalis* (Lepidoptera: Crambidae). *Mitochondrial DNA* **25**, 275–277 (2014).
37. Wu, Q. L., Liu, W., Shi, B. C., Gu, Y. & Wei, S. J. The complete mitochondrial genome of the summer fruit tortrix moth *Adoxophyes orana* (Lepidoptera: Tortricidae). *Mitochondrial DNA* **24**, 214–216 (2013).
38. Park, J. S., Kim, S. S., Kim, K. Y. & Kim, I. Complete mitochondrial genome of Suzuki's Promolactis moth *Promolactis suzukiella* (Lepidoptera: Oecophoridae). *Mitochondrial DNA* doi: 10.3109/19401736.2014.982572 (2013).
39. Dong, Y. *et al.* Complete mitochondrial genome of *Papilio syfanius* (Lepidoptera: Papilionidae). *Mitochondrial DNA* **27**, 403–404 (2016).

## Acknowledgements

This work was supported by the earmarked fund for modern Argo-industry Technology Research System (CARS-22 SYZ10), the Anhui High Schools Natural Science Foundation (KJ2013B320), the National 863 plans projects of China (2011AA100306), the Sericulture Biotechnology Innovation Team (2013xkdt-05), and Ph.D. programs in Biochemistry and Molecular Biology (xk2013042).

## Author Contributions

C.-L.L., B.-J.Z. and L.-S.D. designed the research. L.-S.D. performed the research. L.-S.D., Y.Z. and C.-F.Z. analyzed the data. L.-S.D., B.-J.Z., C.-F.Z. and C.-L.L. contributed reagents/materials/analysis tools. L.-S.D., B.-J.Z. and C.-L.L. wrote the paper with other authors.

## Additional Information

**Competing financial interests:** The authors declare no competing financial interests.

**How to cite this article:** Dai, L.-S. *et al.* Comparative Mitochondrial Genome Analysis of *Eligma narcissus* and other Lepidopteran Insects Reveals Conserved Mitochondrial Genome Organization and Phylogenetic Relationships. *Sci. Rep.* **6**, 26387; doi: 10.1038/srep26387 (2016).



This work is licensed under a Creative Commons Attribution 4.0 International License. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in the credit line; if the material is not included under the Creative Commons license, users will need to obtain permission from the license holder to reproduce the material. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>