Comparative Mitogenomic Analyses of Praying Mantises (Dictyoptera, Mantodea): Origin and Evolution of Unusual Intergenic Gaps

Hong Li Zhang[1†], Fei Ye [2,3,4†*]

1.School of Life Sciences, Datong University, Datong 037009, China;

2. College of Ecology and Evolution, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

3. State Key Laboratory of Biocontrol, Sun Yat-sen University, Guangzhou 510275, China

4. College of Life Sciences, Shaanxi Normal University, Xi'an 710062, China

[†]These authors contributed equally to this work

[*]Corresponding author: feiyeemail@gmail.com
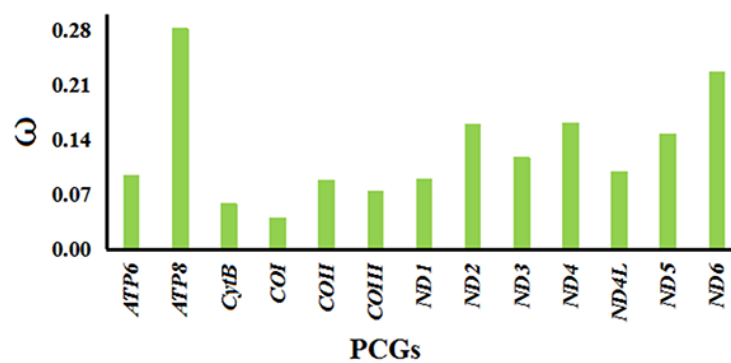
Figure S1. Evolutionary rates of 13 PCGs in the mitogenomes of fifteen Mantodea species. PCGs: protein-coding genes. ω: the ratio of *Ka/Ks* (*Ks*: synonymous nucleotide substitutions; *Ka*: nonsynonymous nucleotide substitutions)
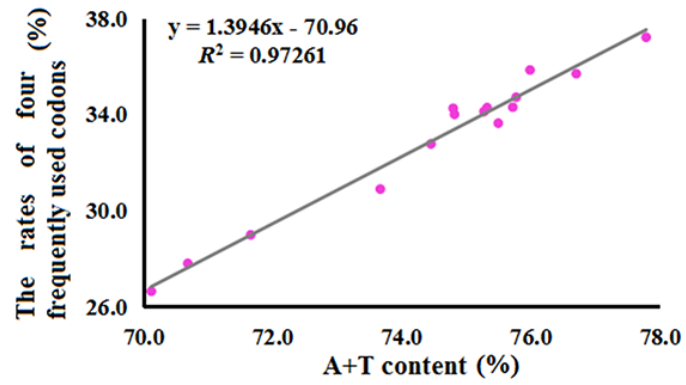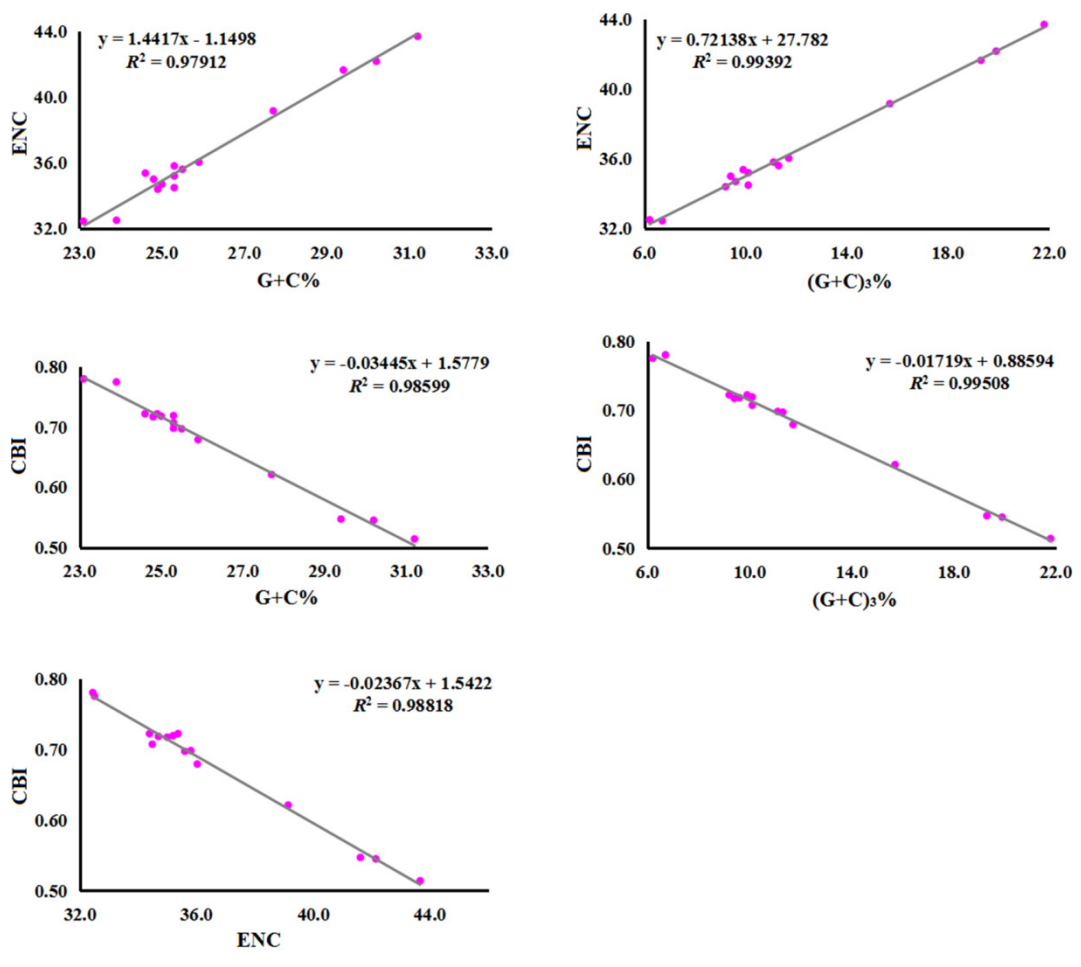
Figure S2. Correlation between the percentage of four most frequently used codons and A+T content of all protein-coding genes.

Figure S3. Evaluation of codon bias in the mitogenomes of fifteen Mantodea species. ENC: effective number of codons; CBI: codon bias index; G+C%: GC content of codons; $(G+C)_3$%: GC content of the third codon positions.
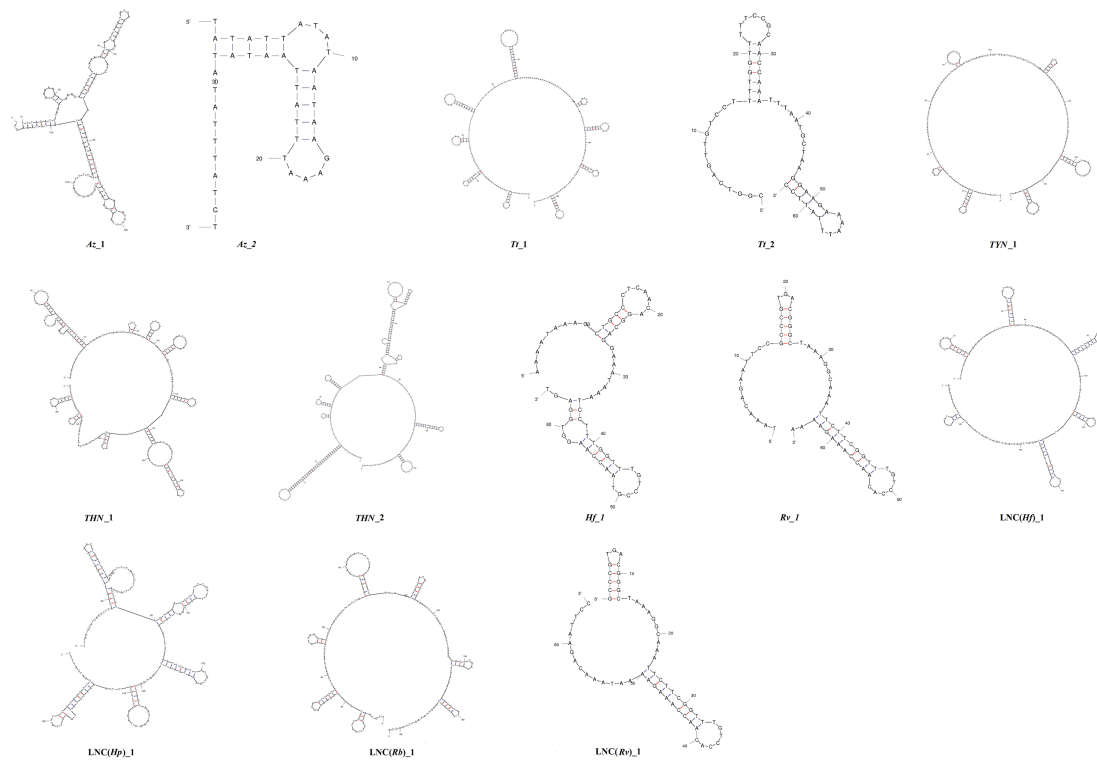
Figure S4. The secondary structure of repetitive unit sequence of TDRs. *Az*: *Anaxarcha zhengi*; *Tt*: *Tamolanica tamolana*; *THN*: *Theopompa* sp.-HN; *TYN*: *Theopompa* sp.-YN; *Hf*: *Hierodula formosana*; *Rv*: *Rhombodera valida*; *Hp*: *Hierodula patellifera*; *Rb*: *Rhombodera brachynota*.
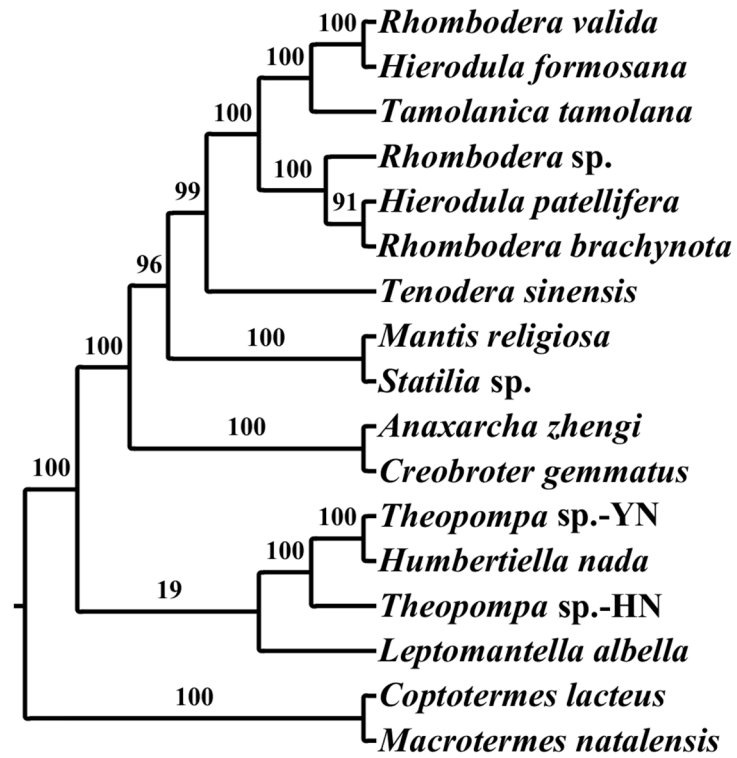
Figure S5. Phylogenetic relationship among fifteen Mantodea species based on mtDNA using ML analyses. The ML bootstrap values were shown close to the nodes.
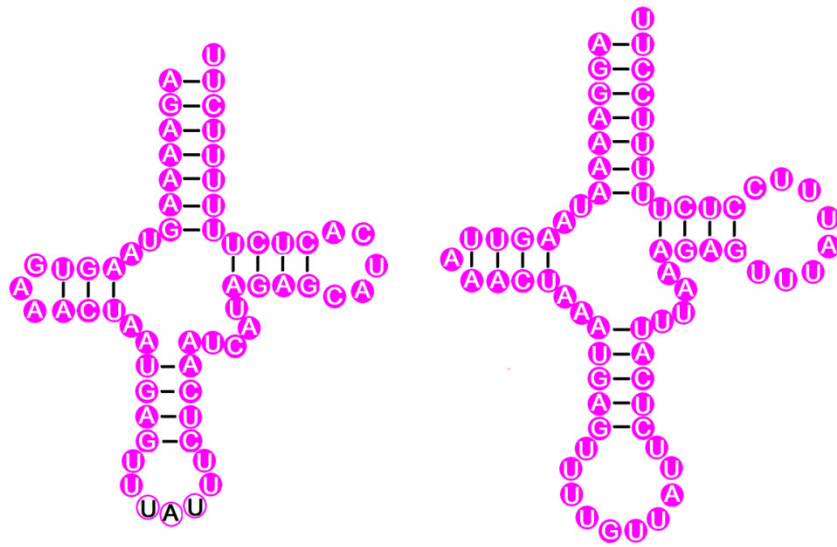
trnM - Like structure in *Rhombodera valida*

trn?(M) - Like structure in *Rhombodera* sp.

Figure S6. tRNA-like structure in LNCs of *Rhombodera valida* and *Rhombodera* sp. Anticodon is indicated with black nucleotides within purple circles.
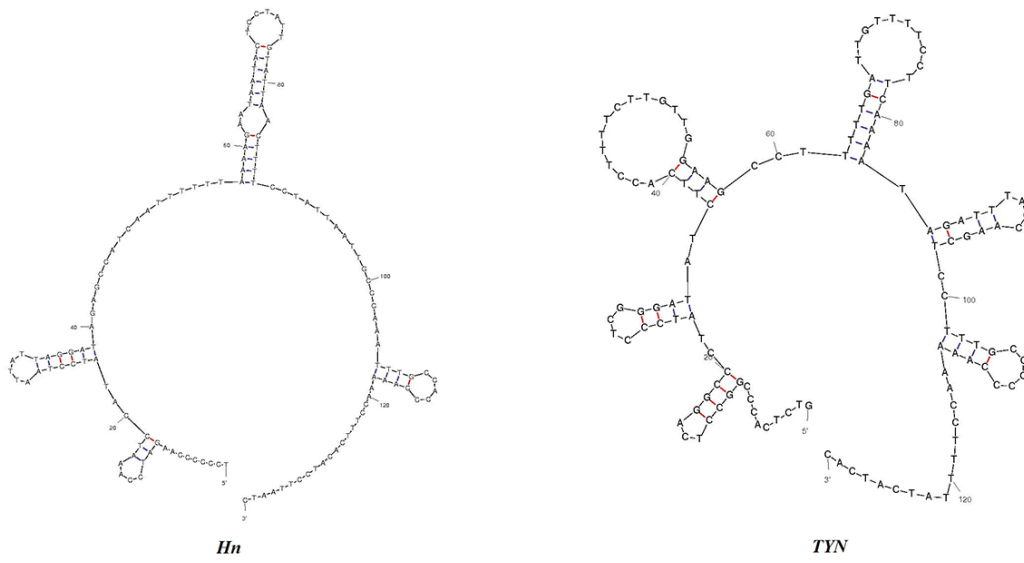
Figure S7. The secondary structure of G1 in *Humbertiella nada* and *Theopompa* sp.-YN. G1: the gap sequence between *trnM* and *trnI*; *Hn*: *Humbertiella nada*; *TYN*: *Theopompa* sp.-YN.
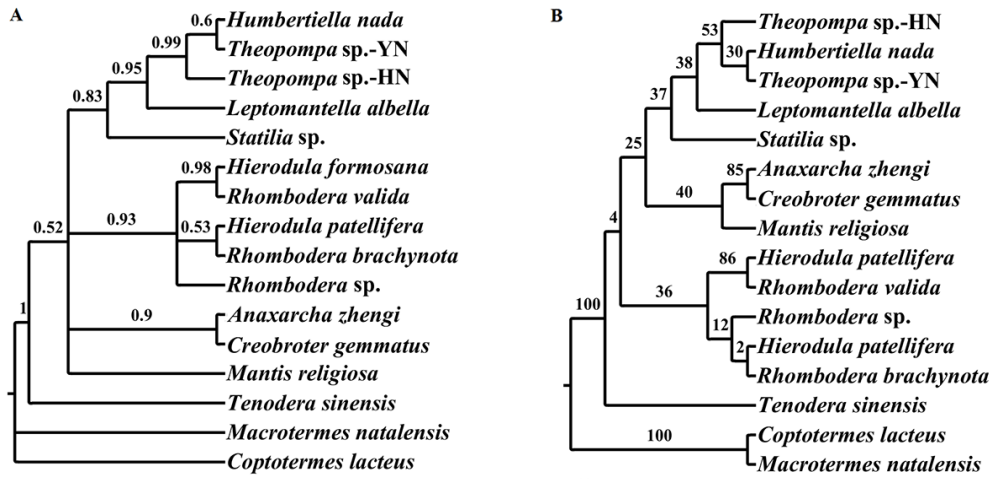
Figure S8. Phylogenetic relationship among fourteen Mantodea species based on CCR using Bayesian inference (A) and Maximum Likelihood (B) analyses. The Bayesian posterior probabilities (BPP) and ML bootstrap values were shown close to the nodes; CCR: the relatively conserved region of CRs in each family.

Table S1. Specimen information for the four sequenced mitogenomes in the present study.

Species	Locality	Sampling time	Voucher number
Hierodula patellifera	Chang'an, Xi'an, Shaanxi	2015.09	Mantodea-E068
Rhombodera brachynota	Mengla, Yunnan	2015.09	Mantodea-E067
Rhombodera valida	Mengla, Yunnan	2015.06	Mantodea-E071
Rhombodera sp.	Mengla, Yunnan	2015.06	Mantodea-E072

Table S2. Primer sequences used for amplification of the four mitogenomes in the present study.

Species	Primer	Primer sequence (5' -3')	Gene	Reference
Hierodula patellifera	HpF1	ACACCAAAATGTAGGAAAGT	*trnQ-trnM*	This study
	HpR1	AGTTGAGAATAATCATCGTT	*COI*	This study
Rhombodera brachynota	RbF1	AGGAGCTATTGGTGGATTAA	*ND2*	This study
	RbR1	AGAGGGTGGAAGAAGTCAAA	*COI*	This study
	RbF2	ACCCTTCTAATTTTCTCCA	*CR*	This study
	RbR2	GGAAAAAATCCTAAAAAGGAGG	*ND2*	This study
	RbF3	TACATTACCCATAAATACTT	*COII*	This study
	RbR3	TACTACATCACGTCATCACT	*COIII*	This study
	RbF4	CAGTTTTATTAGCATCAGGTAT	*COIII*	This study
	RbR4	ACATTGAAGCTGTTATTGAGAT	*trnF*	This study
	RbF5	TAACGTAACATCCCCACTTA	*ND5*	This study
	RbR5	GCTCATGTAGAAGCTCCTGT	*ND4*	This study
Rhombodera sp.	RspF1	AACATTCTACTCAATATCACTAA	*CR*	This study
	RspR1	TTCTACTACAGATGGTAATCATC	*ND2*	This study
	RspF2	CGTGATAGGATAAATAATG	*trnI*	This study
	RspR2	ATAACTTTGAAGGATATTAGTTT	*trnW*	This study
	RspF3	ATTCTTTCATTTTCATCAA	*ND2*	This study
	RspR3	AAAAGCGTGAGCAGTTACA	*COI*	This study
	RspF4	TTTCTTATGTCCTTCTCTT	*ND6*	This study
	RspR4	GTAATTACTGTAGCTCCTC	*CytB*	This study
Rhombodera valida	RvF1	CTGTAGAAAGTGGAGCAGGAA	*COI*	This study
	RvR1	ATCAGCTTTAACTCCGATGC	*COII*	This study
	RvF2	TAGAAGAAAACACCCATAAA	*ND4L*	This study
	RvR2	TACTGTAGCTCCTCAAAAGG	*CytB*	This study
Universal	TM-N200	ACCTTTATAARTGGGGTATGARCC	*trnM*	[19]
	C1-J1709	AATTGGWGGWTTYGGAAAYTG	*COI*	[19]
	C1-N2353	GCTCGTGTATCAACGTCTATWCC	*COI*	[19]
	C1-N2776	GGTAATCAGAGTATCGWCGNGG	*COI*	[19]
	C2-N3665	CCACAAATTTCTGAACATTG	*COII*	[19]
	N3-N5731	TTAGGGTCAAATCCRCAYTC	*ND3*	[19]
	TN-J6155	TTTAATTGAARCCAAAAAGAGG	*trnN*	[19]
	N4-J9172	CGCTCAGGYTGRTACCCYCA	*ND4*	[19]
	N4L-N9629	GTTTGTGAGGGWGYTTTRGG	*ND4L*	[19]
	CB-N11010	TATCTACAGCRAATCCYCCYCA	*CytB*	[19]
	LR-J12888	CCGGTCTGAACTCARATCATGTA	*rrnL*	[19]
	SR-J14197	GTACAYCTACTATGTTACGACTT	*rrnS*	[19]
	SR-N14220	ATATGYACAYATCGCCCGTC	*rrnS*	[19]

Table S3. Nucleotide composition of different molecules in the fifteen Mantodea mitogenomes.

Species name	AT %					AT-skew						GC-skew					
	Mito	PCGs	rRNAs	tRNAs	CR	Mito	PCGs-J	PCGs-N	rRNAs	tRNAs	CR	Mito	PCGs-J	PCGs-N	rRNAs	tRNAs	CR
Anaxarcha zhengi	77.8	76.9	80.2	77.4	79.3	0.068	-0.049	-0.235	-0.104	0.016	0.057	-0.158	-0.068	0.247	0.317	0.127	-0.195
Creobroter gemmatus	76.0	75.4	77.4	75.6	79.9	0.039	-0.078	-0.216	-0.053	0.004	0.046	-0.206	-0.136	0.304	0.350	0.125	-0.338
Tenodera sinensis	75.5	74.7	77.8	75.5	78.9	0.058	-0.064	-0.240	-0.099	0.014	0.009	-0.217	-0.135	0.319	0.373	0.116	-0.383
Tamolanica tamolana	75.3	74.7	77.0	76.4	74.7	0.059	-0.071	-0.240	-0.109	0.018	0.072	-0.236	-0.166	0.314	0.416	0.086	-0.195
Hierodula formosana	75.8	75.2	78.3	76.7	78.0	0.047	-0.074	-0.220	-0.073	-0.006	0.093	-0.205	-0.154	0.285	0.399	0.174	-0.081
Hierodula patellifera	74.5	74.1	76.6	76.4	77.9	0.052	-0.070	-0.239	-0.105	0.015	0.055	-0.239	-0.191	0.318	0.420	0.159	-0.195
Rhombodera brachynota	74.8	75.1	77.6	76.2	80.3	0.043	-0.084	-0.225	-0.091	0.016	0.032	-0.204	-0.154	0.303	0.390	0.155	-0.167
Rhombodera sp.	75.7	75.0	77.6	76.3	79.8	0.037	-0.092	-0.226	-0.084	0.010	0.007	-0.209	-0.152	0.299	0.377	0.154	-0.074
Rhombodera valida	74.8	74.6	77.7	77.5	72.5	0.065	-0.062	-0.240	-0.088	0.012	0.122	-0.234	-0.177	0.316	0.404	0.145	-0.152
Mantis religiosa	76.7	76.1	77.7	76.0	81.1	0.027	-0.113	-0.216	-0.063	0.014	0.027	-0.188	-0.111	0.274	0.350	0.096	-0.273
Statilia sp.	75.3	74.7	77.6	75.4	76.2	0.024	-0.105	-0.212	-0.069	0.012	-0.038	-0.175	-0.120	0.273	0.325	0.097	-0.252
Humbertiella nada	70.1	68.8	72.9	73.1	73.7	0.011	-0.109	-0.203	-0.009	-0.021	-0.038	-0.262	-0.229	0.274	0.400	0.107	-0.272
Theopompa sp.-YN	70.7	69.8	74.4	72.7	70.8	0.010	-0.121	-0.216	-0.037	0.011	-0.018	-0.276	-0.228	0.303	0.438	0.089	-0.362
Theopompa sp.-HN	71.7	70.6	74.4	75.3	71.2	0.037	-0.067	-0.227	-0.049	0.004	-0.093	-0.219	-0.244	0.289	0.389	0.084	-0.133
Leptomantella albella	73.7	72.3	76.8	75.0	81.9	0.020	-0.107	-0.209	-0.023	0.011	0.050	-0.207	-0.170	0.224	0.377	0.108	-0.161

AT skew: [A−T]/[A+T]. GC skew: [G−C]/[G+C]. Mito: the majority strand of mitogenome. PCGs-J: PCGs encoded by the majority strand. PCGs-N: PCGs encoded by the minority strand.

Table S4. The usage rate of codons ending with different nucleotides.

Feature	NNU	NNC	NNA	NNG
PCGs	45.0%	7.7%	42.9%	4.4%
PCGs-J	42.6%	10.9%	44.0%	2.4%
PCGs-N	48.7%	2.6%	41.1%	7.6%

PCGs-J: PCGs encoded by the majority strand. PCGs-N: PCGs encoded by the minority strand.

Table S5. The usage rate of four most frequently used codons.

Species	UUU(F)	UUA(L)	AUU(I)	AUA(M)	SUM	Total	Rate (%)
Anaxarcha zhengi	299	471	338	272	1380	3709	37.2
Creobroter gemmatus	284	427	355	262	1328	3705	35.8
Tenodera sinensis	276	417	314	239	1246	3705	33.6
Tamolanica tamolana	287	424	316	237	1264	3705	34.1
Hierodula formosana	292	427	318	249	1286	3705	34.7
Hierodula patellifera	278	395	308	233	1214	3705	32.8
Rhombodera brachynota	289	428	318	234	1269	3705	34.3
Rhombodera valida	280	410	327	243	1260	3706	34.0
Rhombodera sp.	287	422	328	234	1271	3706	34.3
Mantis religiosa	292	453	335	243	1323	3707	35.7
Statilia sp.	290	420	332	228	1270	3705	34.3
Humbertiella nada	229	273	291	194	987	3706	26.6
Theopompa sp.-YN	252	303	283	192	1030	3706	27.8
Theopompa sp.-HN	258	307	283	226	1074	3706	29.0
Leptomantella albella	258	352	306	229	1145	3707	30.9

Table S6. The identical percentage of nucleotides in each tRNA family and different components in the secondary structure of tRNAs.

tRNA	INUC (%)	AA (%)	stem DHU (%)	stem AC (%)	stem TψC (%)	stem DHU (%)	loop AC (%)	loop Variable (%)	loop TψC (%)	loop
trnI	76.4	93.3	83.3	80.0	100	75.0	85.7	80.0	50.0	
trnQ	67.1	53.3	100	80.0	30.0	40.0	85.7	50.0	100	
trnM	58.7	73.3	100	40.0	60.0	25.0	71.4	75.0	*	
trnW	72.2	80.0	100	70.0	87.5	50.0	100	100	*	
trnC	56.5	80.0	100	30.0	66.7	*	100	50.0	*	
trnY	38.7	80.0	16.7	70.0	10.0	*	100	*	*	
trnL^{UUR}	66.2	86.7	100	30.0	90.0	50.0	71.4	25.0	*	
trnK	68.1	53.3	66.7	87.5	75.0	75.0	100	55.6	42.9	
trnD	61.6	66.7	100	50.0	80.0	16.7	71.4	75.0	*	
trnG	59.4	80.0	75.0	50.0	100	20.0	85.7	25.0	*	
trnA	75.0	60.0	100	100	70.0	50.0	71.4	100	25.0	
trnR	65.7	80.0	66.7	80.0	66.7	28.6	100	20.0	11.1	
trnN	75.0	86.7	100	80.0	75.0	62.5	85.7	60.0	*	
trnS^{AGN}	52.8	26.7	-	80.0	83.3	18.2	100	40.0	76.9	
trnE	70.6	66.7	100	100	50.0	66.7	100	75.0	*	
trnF	76.1	86.7	100	80.0	60.0	66.7	71.4	75.0	*	
trnH	61.8	53.3	100	80.0	62.5	60.0	85.7	50.0	*	
trnT	66.2	93.3	87.5	100	50.0	33.3	100	*	*	
trnP	62.1	66.7	62.5	90.0	62.5	20.0	71.4	50.0	*	
trnS^{UCN}	77.5	92.9	100	80.0	100	57.1	71.4	50.0	57.1	
trnL^{CUN}	49.3	46.7	66.7	30.0	37.5	50.0	85.7	50.0	14.3	
trnV	64.3	60.0	66.7	70.0	87.5	71.4	100	16.7	71.4	
AVE	64.6	71.2	85.3	70.8	68.4	42.6	87.0	51.0	20.4	

AVE: the average identical percent; -: the DHU arm was lost in *trnS^{AGN}*; *: identical percentage of nucleotides < 10.0%.

Table S7. The percent nucleotide identity in each helix of the secondary structure of *rRNAs* in fifteen Mantodea species.

Domain	Helix	Total site (nt)	Conserved site (nt)	Percentage of conserved site (%)	Domain	Helix	Total site (nt)	Conserved site (nt)	Percentage of conserved site (%)
rrnL_I	H183	4	0	0	*rrnL_V*	H2455	35	22	62.9
	H235	21	10	47.6		H2507	27	26	96.3
	H461	4	0	0		H2520	31	8	25.8
	H533	6	1	16.7		H2547	10	10	100
	H563	8	6	75.0		H2588	15	12	80.0
rrnL_II	H579	12	7	58.3	*rrnL_VI*	H2646	26	15	57.7
	H589	18	4	22.2		H2675	10	5	50.0
	H671	28	21	75.0		H2735	10	2	20.0
	H687	25	14	56.0					
	H736	17	8	47.1	*rrnS_I*	H9	8	6	75.0
	H777	6	5	83.3		H17	6	6	100
	H812	12	7	58.3		H27	22	10	45.5
	H822	10	7	70.0		H39	17	9	52.9
	H837	30	4	13.3		H47	64	16	25.0
	H946	20	9	45.0		H367	23	18	78.3
	H991	21	6	28.6		H500	10	5	50.0
	H1057	16	8	50.0		H511	26	18	69.2
	H1087	9	8	88.9	*rrnS_II*	H567	8	2	25.0
	H1196	20	2	10.0		H577	36	5	13.9
rrnL_IV	H1648	39	9	23.1		H673	32	5	15.6
	H1764	19	5	26.3		H769	30	24	80.0
	H1775	6	6	100		H885	24	19	79.2
	H1792	16	11	68.8	*rrnS_III*	H921	35	30	85.7
	H1830	8	4	50.0		H939	10	10	100
	H1835	20	15	75.0		H944	25	21	84.0
	H1906	12	11	91.7		H960	8	7	87.5
	H1925	2	2	100		H984	14	11	78.6
	H1935	12	8	66.7		H1047	39	29	74.4
rrnL_V	H2023	10	8	80.0		H1068	12	1	8.3
	H2043	24	20	83.3		H1074	6	4	66.7
	H2064	25	25	100		H1113	10	4	40.0
	H2077	31	12	38.7		H1241	4	4	100
	H2246	8	5	62.5		H1303	16	6	37.5
	H2259	12	4	33.3		H1350	14	10	71.4
	H2347	13	6	46.2		H1399	34	30	88.2
	H2395	8	3	37.5		H1506	20	14	70.0

Table S8. The size of the CRs and the position of the three common elements in the CRs.

Family	Speices	CR size (bp)	CR(-TDR) size (bp)	Poly T	CBS1	CBS2
Hymenopodidae	*Anaxarcha zhengi*	1761	639	214-221	449-473	523-530
	Creobroter gemmatus	708	708	305-313	508-531	581-588
Mantidae	*Tenodera sinensis*	706	706	294-303	513-537	587-594
	Tamolanica tamolana	954	523	-	-	-
	Hierodula formosana	840	779	397-406	598-621	673-680
	Hierodula patellifera	674	674	283-291	485-508	559-566
	Rhombodera brachynota	669	669	287-295	489-512	563-570
	Rhombodera valida	959	768	379-388	583-606	658-665
	Rhombodera sp.	672	672	280-288	481-504	556-563
	Mantis religiosa	639	639	254-262	450-474	524-531
	Statilia sp.	652	652	258-272	459-482	534-541
Liturgusidae	*Humbertiella nada*	866	866	282-291	483-506	556-563
	Theopompa sp.-YN	1436	904	321-328	522-545	595-602
	Theopompa sp.-HN	1775	879	314-320	517-540	590-597
Tarachodidae	*Leptomantella albella*	684	684	269-277	475-498	549-556

CR: control region; -TDR: redundant tandem repeats were removed; CBS: conserved block sequence.