Why barcode? High-throughput multiplex sequencing of mitochondrial genomes for molecular systematics

Timmermans, M.J.T.N., Dodsworth, S., Culverwell, C.L., Bocak, L., Ahrens, D., Littlewood, D.T.J., Pons, J., and Vogler A.P.

Supplementary Material

Supplementary Table 1. Primers for PCR

	Primer	Locus	Primer Sequence 5'=>3'	Anneal Temp. (°C)	Length (kbp)	PCR †
Sanger Sequences	SPatR	*cox1*	GCACTAWTCTGCCATATTAGA	53	0.9	(1)
	Jerry	*cox1*	CAACATTTATTTTGATTTTTTGG			
	SPatR	*cox1*	GCACTAWTCTGCCATATTAGA	51	0.9	(1)
	SJerry_F	*cox1*	CAACATYTATTYTGATTYTTTGG			
	Sytb_F	*cob*	TGAGGNCAAATATCHTTYTGAGG	55	0.5	(1)
	Sytb_R	*cob*	GCAAATARRAARTATCATTCDGG			
	SteveND5_F	*nad5*	CCYATAWANCGAATATCYTG	49	0.6	(1)
	SteveND5_R	*nad5*	GTWTCWTATTGTTTRGTYATTTATT			
454 Sequences	16stev_R2	*rrnL*	ACCTTAGGGATAACAGCGT	53	10	(2)
	SPatF	*cox1*	TCTAATATGGCAGAWTAGTGC			
	16stev_R2	*rrnl*	ACCTTAGGGATAACAGCGT	53	10	(2)
	SJerry_F	*cox1*	CAACATYTATTYTGATTYTTTGG			
	sCOII_R	*cox2*	CARATTTCDGARCATTG	49	2	(3)
	SJerry_F	*cox1*	CAACATYTATTYTGATTYTTTGG			
	stRNAM_F	*tRNA*	AAGCTWNTRGGTTCATACC	48	2	(3)
	SPatR	*cox1*	GCACTAWTCTGCCATATTAGA			
	ND1A	*rrnL*	GGTCCCTTACGAATTTGAATATATCCT	59	2	(3)
	12stevEND_R	*rrnS*	GTGCCAGCARTTGCGGTT			

† PCR conditions provided in Supplementary Table 2

Supplementary Table 2. PCR conditions

	(1) BIOTAQ™ DNA Polymerase		(2) TaKaRa LR Taq™		(3) BIOTAQ™ DNA Polymerase		No. of Cycles
	Temp. (°C)	Time	Temp. (°C)	Time	Temp. (°C)	Time	
Initialisation	94	2 min	94	1 min	94	2 min	1
Denaturation	94	30 sec	98	5 sec	94	30 sec	35
Annealing	*	30 sec	53	30 sec	*	30 sec	
Extension	70	1 min	60	15 min	60	2 min	
Final Extension	72	10 min	72	10 min	72	10 min	1
Final Hold	4	∞	4	∞	4	∞	

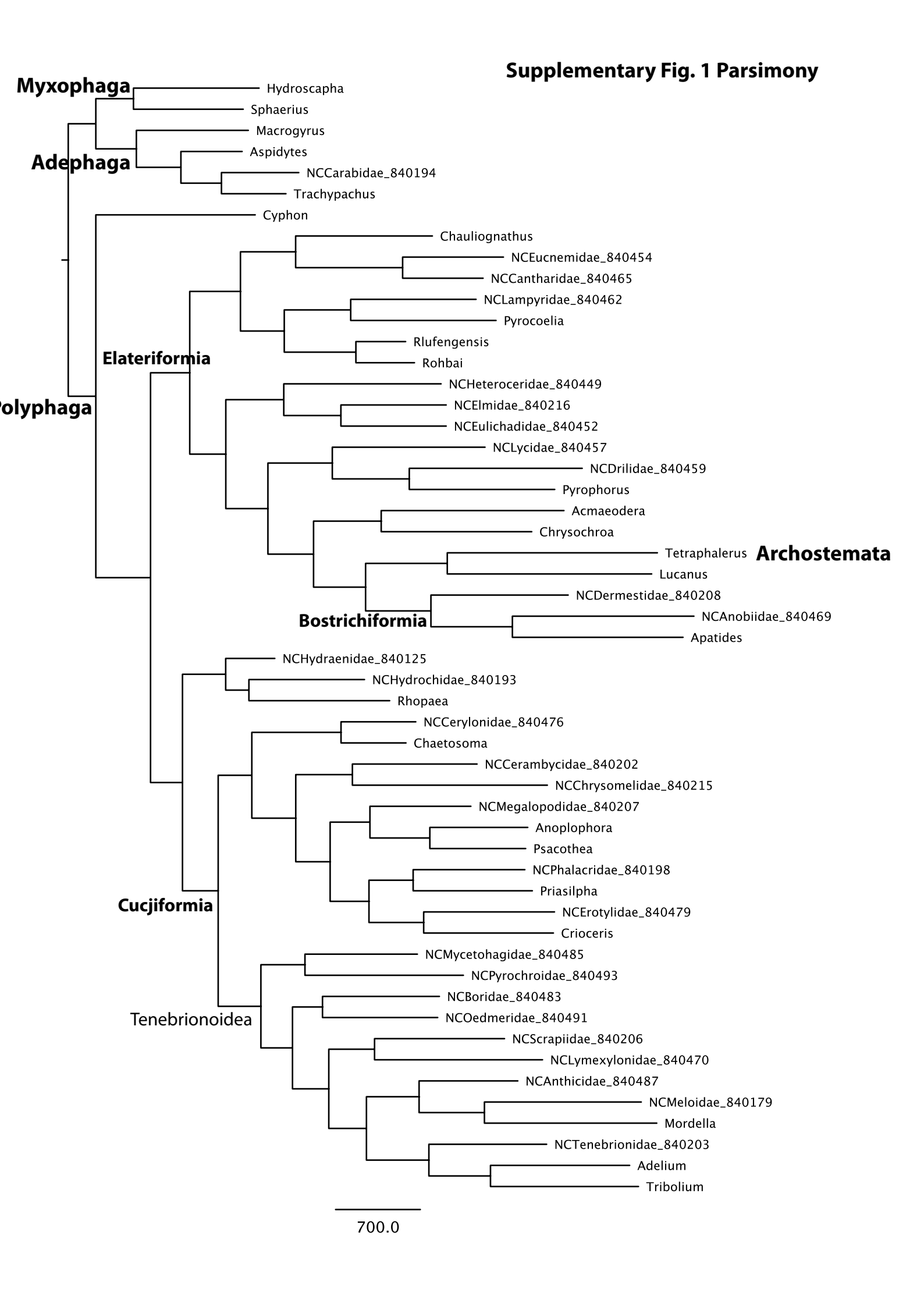
* Annealing temperature varied between each primer pair [see primer table].
In all protocols the central steps of the PCR were repeated 35 times.

Supplementary Table 3. Results of contig generation in simulated data sets

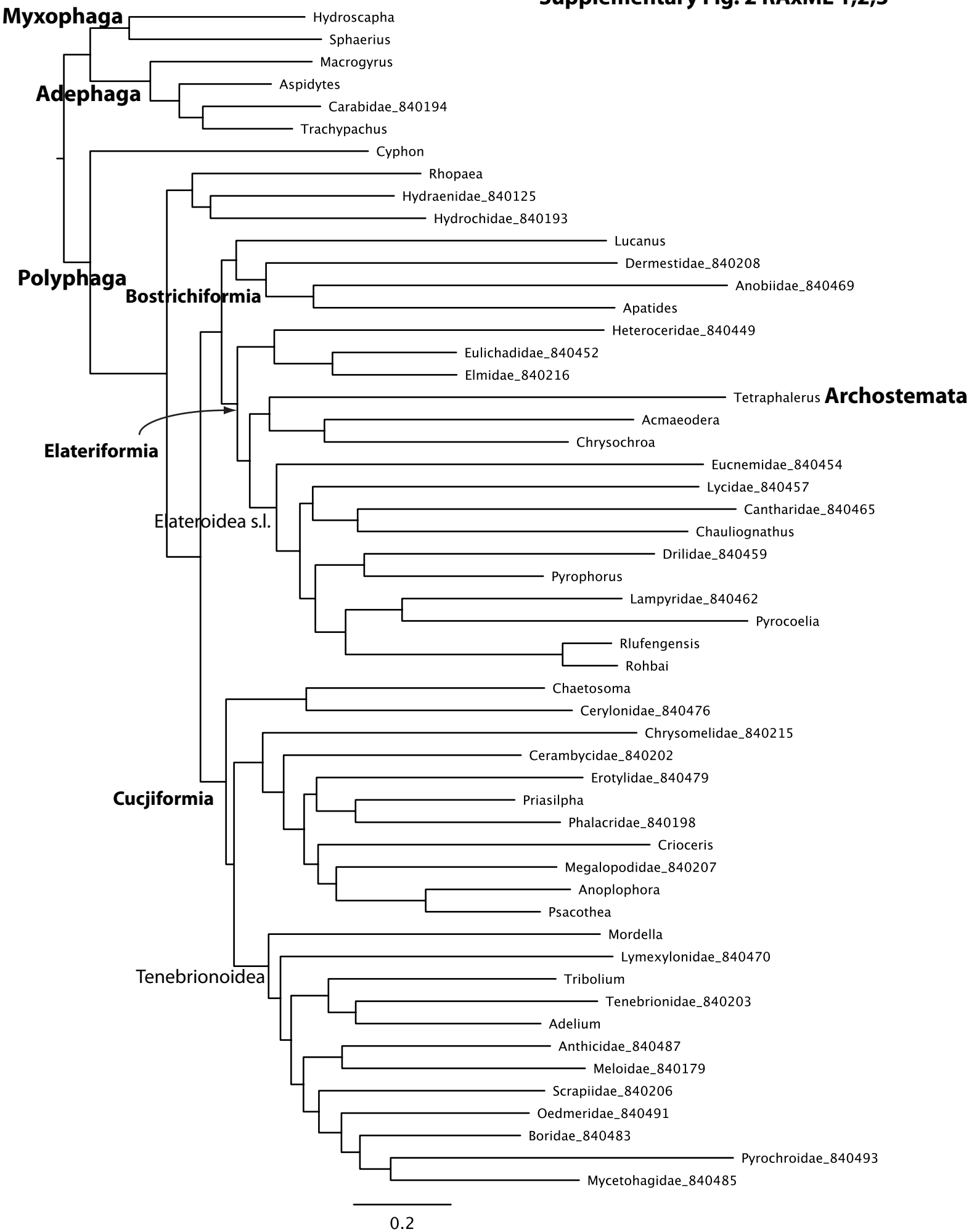
Accession number	Species	Number simulated reads[1)]	Av. read length (bp)	stdev	GenBank (bp)	Assembled (bp)[2)]	Number mis-matches[3)]
NC_013580.1	*Acmaeodera sp.*	1168	279.3	30.4	16217	16103	0
NC_013554.1	*Adelium sp.*	1260	281.6	35.5	16449	16357	7
NC_008221.1	*Anoplophora glabripennis*	1202	288.1	35.1	15774	15728	21*
NC_013582.1	*Apatides fortis*	1185	276.5	30.0	16171	15696	9**
NC_012139.1	*Aspidytes niobe*	1055	286.4	35.8	14257	14188	9
NC_011324.1	*Chaetosoma scaritides*	1170	288.1	37.3	15511	15452	0
NC_013576.1	*Chauliognathus opacus*	1068	285.3	32.3	14893	14774	1
NC_012765.1	*Chrysochroa fulgidissima*	1248	279.0	31.5	15592	15501	14*
NC_003372.1	*Crioceris 12-punctata*	1238	288.1	37.5	15880	15786	6
NC_011320.1	*Cyphon sp.*	1232	287.1	35.8	15919	15859	4
NC_012144.1	*Hydroscapha granulum*	1162	290.4	36.6	15975	15907	2
NC_013578.1	*Lucanus mazama*	1106	281.5	32.2	16261	15105	4
NC_013249.1	*Macrogyrus oblongus*	1204	288.1	36.4	16643	16607	4
NC_013254.1	*Mordella atrata*	1182	281.9	31.9	15540	15460	2
NC_011326.1	*Priasilpha obscura*	1216	288.3	36.6	16603	16098	1
NC_013070.1	*Psacothea hilaris*	1265	288.0	34.7	15856	14885	0
NC_003970.1	*Pyrocoelia rufa*	1268	285.9	35.2	17739	16312	9**
NC_009964.1	*Pyrophorus divergens*	1191	278.9	30.9	16120	16081	2
NC_010969.1	*Rhagophthalmus lufengensis*	1219	288.5	36.3	15982	15901	4
NC_010964.1	*Rhagophthalmus ohbai*	1243	288.9	36.6	15704	14625	10
NC_013252.1	*Rhopaea magnicornis*	1335	283.5	33.5	17522	17408	0
NC_011322.1	*Sphaerius sp.*	1178	287.2	34.8	15121	15045	5
NC_011328.1	*Tetraphalerus bruchi*	1202	275.5	31.1	15689	15648	3
NC_011329.1	*Trachypachus holmbergi*	1222	285.7	36.1	15722	15626	25
NC_003081.2	*Tribolium castaneum*	1181	282.3	32.3	15881	15824	1
Average		**1200**	**284.6**	**34.2**	**15960.8**	**15679.0**	**5.7**

[1)] Sequences were generated on linearized full mt genome sequences, with mean fragment size 284 (parameter 2: 50). Other settings were default.
[2)] Length of mt genome from MIRA assembly from simulated data set; the difference from the original sequence ('GenBank') was due to missing nucleotides near the ends of the linearized mt genomes only, except for those marked with ***.
[3)] Number of mismatched nucleotide positions between the GenBank sequence and the simulated MIRA assembly. *) Assembly produced two (*Anoplophora*) or three (*Chrysochroa*) contigs with partly overlapping sequences that each contain several errors against the GenBank entry in the region overlapping. **) Tandem repeat unit omitted from AT rich region.

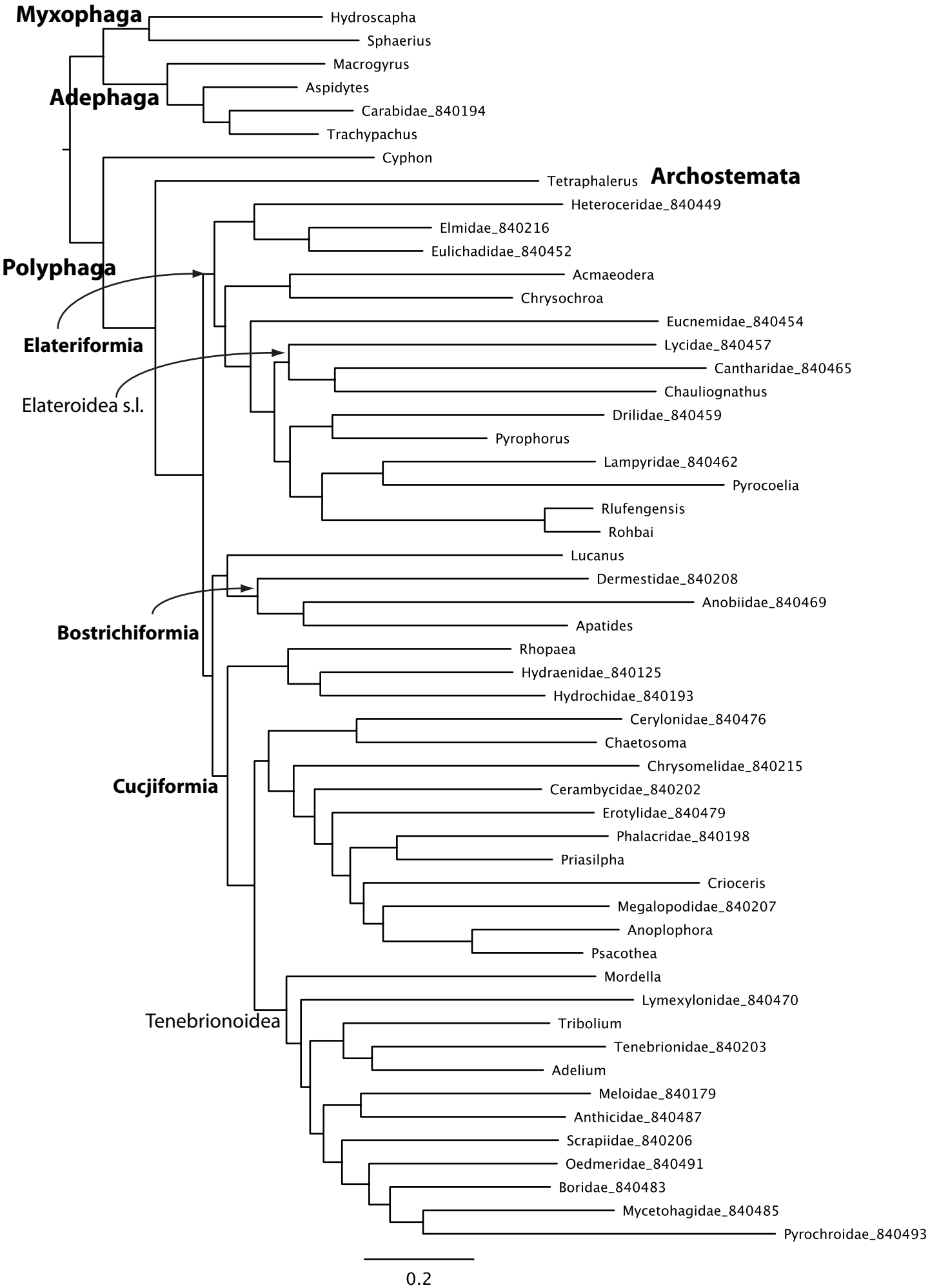
Supplementary Figures S1-S5: Phylogenetic trees obtained from parsimony and model-based analyses.


Supplementary Fig. 1 Parsimony
Myxophaga
Hydroscapha
Sphaerius
Macrogyrus
Adephaga
Aspidytes
NCCarabidae_840194
Trachypachus
Cyphon
Chauliognathus
NCEucnemidae_840454
NCCantharidae_840465
NCLampyridae_840462
Pyrocoelia
Rlufengensis
Elateriformia
Rohbai
Polyphaga
NCHeteroceridae_840449
NCElmidae_840216
NCEulichadidae_840452
NCLycidae_840457
NCDrilidae_840459
Pyrophorus
Acmaeodera
Chrysochroa
Tetraphalerus
Archostemata
Lucanus
NCDermestidae_840208
Bostrichiformia
NCAnobiidae_840469
Apatides
NCHydraenidae_840125
NCHydrochidae_840193
Rhopaea
NCCerylonidae_840476
Chaetosoma
NCCerambycidae_840202
NCChrysomelidae_840215
NCMegalopodidae_840207
Anoplophora
Psacothea
NCPhalacridae_840198
Priasilpha
Cucjiformia
NCErotylidae_840479
Crioceris
NCMycetohagidae_840485
NCPyrochroidae_840493
NCBoridae_840483
Tenebrionoidea
NCOedmeridae_840491
NCScrapiidae_840206
NCLymexylonidae_840470
NCAnthicidae_840487
NCMeloidae_840179
Mordella
NCTenebrionidae_840203
Adelium
Tribolium
700.0

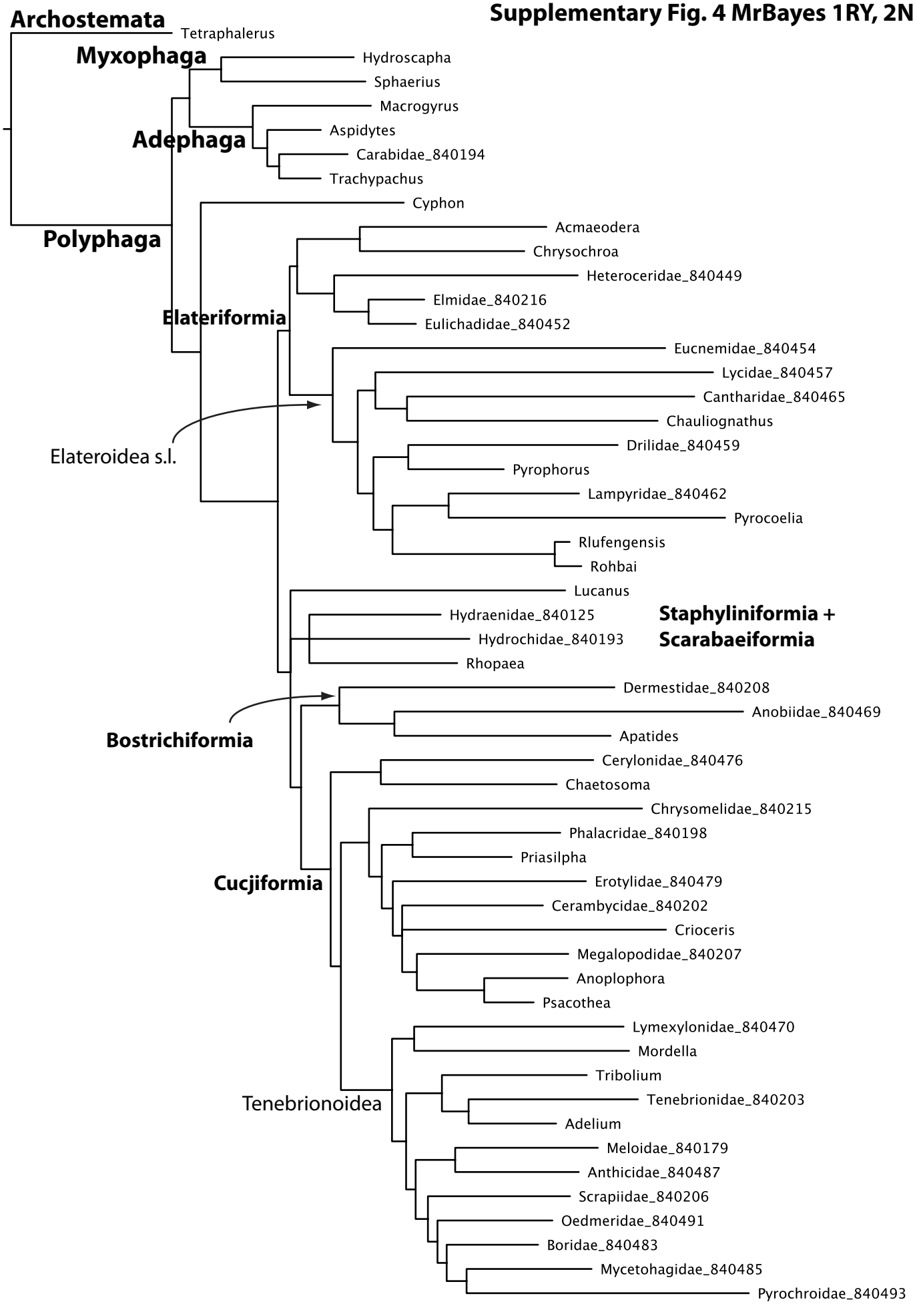
Supplementary Fig. 2 RAxML 1,2,3



Supplementary Fig. 3 MrBayes 1, 2, 3



Supplementary Fig. 4 MrBayes 1RY, 2N



Supplementary Fig. 5 Phylobase Nucl.