

Soup to tree: the phylogeny of beetles inferred by mitochondrial metagenomics of a Bornean rainforest sample

Crampton-Platt, A.L., Timmermans, M.J.T.N., Gimmel, M.L., Kutty, S.K., Cockerill, T.D., Chey, V.K., and Vogler, A.P.

Supplementary Data

Table S1.....	1
Table S2.....	2
Figure S1.....	8
Supplementary Information S1.....	9
Supplementary Information S2.....	10
Table S3.....	11
Figure S2.....	12
Figure S3.....	13
Figure S4.....	14
Figure S5.....	15
Supplementary Information S3.....	18
Supplementary Information S4.....	18
Table S4.....	19
Script 1.....	20
Script 2.....	21
Script 3.....	22
Script 4.....	24

Table S1. The number of morphospecies and individual specimens therein, and the number of successful *cox1* barcodes available as bait sequences, grouped by family.

Superfamily	Family	Morphospecies	Individuals	PCR success (morphospecies)	PCR success (individuals)
Archostemata	Cupedidae	1	1	0	0
Caraboidea	Carabidae	9	45	7	24
Scirtoidea	Scirtidae	3	3	3	3
Bostrichoidea	Anobiidae	4	5	2	3
Staphylinoidea	Staphylinidae	27	50	21	32
Byrrhoidea	Ptilodactylidae	6	20	6	17
Buprestoidea	Buprestidae	7	14	6	10
Elateroidea	Eucnemidae	6	8	5	7
Elateroidea	Lycidae	4	4	3	3
Elateroidea	Lampyridae	1	2	1	2
Elateroidea	Elateridae	7	13	6	10
Cleroidea	Cleridae	3	5	3	4
Cleroidea	Malachiidae	1	1	0	0
Cleroidea	Prionoceridae	1	1	1	1
Cucujoidea	Coccinellidae	11	21	9	19
Cucujoidea	Corylophidae	4	6	3	4
Cucujoidea	Discolomatidae	4	8	4	6
Tenebrionoidea	Aderidae	4	4	2	2
Tenebrionoidea	Anthicidae	4	6	4	5
Tenebrionoidea	Ciidae	1	1	1	1
Tenebrionoidea	Melandryidae	1	1	0	0
Tenebrionoidea	Mordellidae	5	6	4	4
Tenebrionoidea	Tenebrionidae	9	44	6	30
Tenebrionoidea	Zopheridae	2	2	2	2
Cucujoidea	Nitidulidae	2	2	2	2
Cucujoidea	Erotylidae	4	5	4	5
Chrysomeloidea	Cerambycidae	6	14	4	12
Chrysomeloidea	Chrysomelidae	19	58	14	38
Cucujoidea	Phalacridae	3	4	2	2
Curculionoidea	Anthribidae	9	17	9	15
Curculionoidea	Attelabidae	4	9	4	7
Curculionoidea	Brentidae	2	2	1	1
Curculionoidea	Curculionidae	33	87	19	51
Histeroidea	Histeridae	2	2	1	1
[unknown]	[unknown]	3	6	2	4

Table S2. The mitochondrial genomes used in this study for filtering mitochondrial sequences (n=245), the mitochondrial tree-of-life for beetles (n=240, 10 protein-coding genes required, unused sequences highlighted in grey), and the tree-of-life under reduced reference taxon sampling (n=34, highlighted in blue).

Suborder/Series/ Superfamily	Family	Subfamily	Genus, species	Voucher (BMNH)	GenBank Accession No.	Geographic origin	Published by	
Archostemata	Cupedidae	Priacminae	<i>Priacma serrata</i>	842702	JX412806	USA		
	Ommatidae	Tetraphalerinae	<i>Tetraphalerus bruchi</i>		EU877953	Argentina	Sheffield et al. (2008)	
	Hydroscaphidae	Hydroscaphinae	<i>Hydroscapha granulum</i>		AM493667	Italy	Pons et al. (2010)	
	Sphaeriusidae	Sphaeriusinae	<i>Sphaerius</i> sp.		NC011322	USA	Sheffield et al. (2008)	
Adephaga	Carabidae	Cicindelinae	<i>Habrodera capensis</i>	838091	JX412824	South Africa	Timmermans et al. (submitted)	
		Harpalinae	<i>Agonum muelleri</i>	832916	JX412835	UK	Timmermans et al. (submitted)	
	Carabidae	Brachyninae	<i>Brachinus crepitans</i>	844234	JX412826	Slovakia	Timmermans et al. (submitted)	
		Carabinae	<i>Calosoma</i> sp.		GU176340	USA	Song et al. (2010)	
	Carabidae	Lebiinae	<i>Calleida angusticollis</i>	838156	JX412855	South Africa	Timmermans et al. (submitted)	
		Carabinae	<i>Carabus mirabilissimus</i>		GQ344500	Korea	Wan et al. (2012)	
	Carabidae	Panagaeinae	<i>Craspedophorus nobilis</i>	838140	JX412738	South Africa	Timmermans et al. (submitted)	
		Carabidae	Pterostichinae	<i>Cyrtomelis</i> sp.	838103	JX412794	South Africa	Timmermans et al. (submitted)
	Carabidae	Carabidae	Ctenodactylinae	<i>Hexagonia terminalis</i>	838155	JX412768	South Africa	Timmermans et al. (submitted)
		Carabidae	Promecognathinae	<i>Promecognathus crassus</i>	743008	JX313665	USA	Timmermans et al. (submitted)
	Carabidae	Carabidae	Trechinae	<i>Trechini</i> gen. sp.	840194	HQ232802	South Africa	Timmermans et al. (2010)
		Trachypachidae	Trachypachinae	<i>Trachypachus holimbergi</i>		EU877954	Canada	Sheffield et al. (2008)
	Hydralephaga	Aspiyridae	Aspiyridinae	<i>Aspidytes niobe</i>		AM493668	South Africa	Pons et al. (2010)
		Dytiscidae	Copelatinae	<i>Copelatus</i> sp.	679508	JX220995	South Africa	Timmermans et al. (submitted)
		Gyrinidae	Gyrininae	<i>Macrogyrus oblongus</i>		NC013249	Australia	Cameron et al. (2010)
			Gyrinidae	Gyrininae	gen. sp.	840223	JX412840	South Africa
Halipidae		Halipinae	<i>Pelodytes quadratus</i>	794142	JX313685	Madagascar	Timmermans et al. (submitted)	
		Halipidae	Halipinae	<i>Halipilus flavicollis</i>	835590	JX313677	UK	Timmermans et al. (submitted)
Meruidae		Meruidae	<i>Meru phyllisae</i>	703458	JX313691	Venezuela	Timmermans et al. (submitted)	
		Noteridae		gen. sp.	838166	JX313666	South Africa	Timmermans et al. (submitted)
Polyphaga		Clambidae	Clambinae	<i>Clambus</i> sp.		JX412725	South Africa	Timmermans et al. (submitted)
			Scirtidae	<i>Cyphon</i> sp.	840362	EU877949	USA	Sheffield et al. (2008)
	Scirtidae	Scirtinae	<i>Prionocyphon serricornis</i>	842693	JX412728	Czech Republic	Timmermans et al. (submitted)	
		Scirtinae	gen. sp.	842692	JX412785	Cameroun	Timmermans et al. (submitted)	
	Staphyliniformia	Agyrtidae	<i>Agyrtodes labralis</i>	842916	JX412779	New Zealand	Timmermans et al. (submitted)	
		Leiodidae	Cholevinae	<i>Nargus velox</i>	833928	JX313674	UK	Timmermans et al. (submitted)
	Staphylinidae	Silphinae	<i>Necrophila americana</i>		GU176343	USA	Song et al. (2010)	
		Staphylinidae	Scaphidiinae	<i>Brachynopus latus</i>	842925	JX412745	New Zealand	Timmermans et al. (submitted)
	Staphylinidae	Staphylinidae	<i>Carpelimus</i> sp.	840291	JX412816	South Africa	Timmermans et al. (submitted)	
		Staphylinidae	Aleocharinae	<i>Chiloporata longitarsis</i>	724053	JX313678	UK	Timmermans et al. (submitted)
Staphylinidae	Staphylinidae	<i>Euplectus karsteri</i>	729943	KM586245	UK	Timmermans et al. (submitted)		
	Staphylinidae	Pselaphinae	<i>Tychus niger</i>	742905	KM586244	UK	Timmermans et al. (submitted)	
Staphylinidae	Staphylinidae	<i>Habrocerus capillaricornis</i>	844248	JX412795	UK	Timmermans et al. (submitted)		
	Staphylinidae	Omalinae	<i>Omalium rivulare</i>	844244	JX412759	UK	Timmermans et al. (submitted)	
Staphylinidae	Aleocharinae	<i>Oxyopa opaca</i>	844246	JX412751	UK	Timmermans et al. (submitted)		

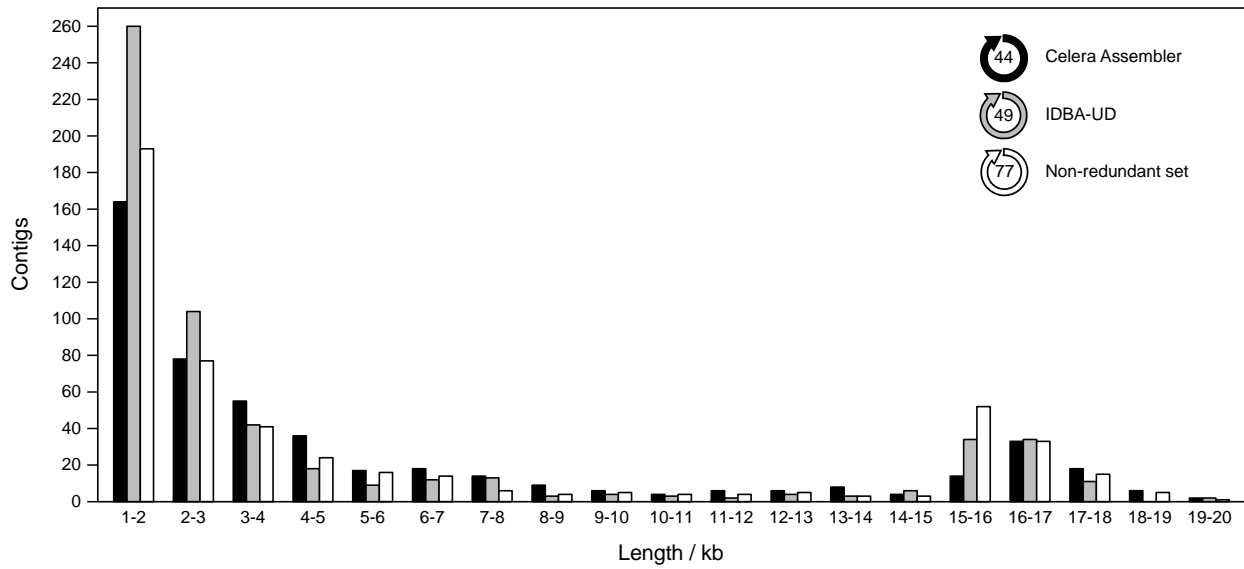
	Staphylinidae	Oxytelinae	<i>Oxytelus</i> sp.	844247	JX412796	UK	Timmermans et al. (submitted)
	Staphylinidae	Paederinae	<i>Hypomedon debilicornis</i>	833953	JX313672	UK	Timmermans et al. (submitted)
Hydrophiloidea	Histeridae	Abraeinae	<i>Abraeus globosus</i>	844242	JX313690	UK	Timmermans et al. (submitted)
	Histeridae	Saprininae	<i>Euspilotus</i> sp.		GU176344	USA	Song et al. (2010)
	Histeridae	Dendrophilinae	<i>Paromalus flavicornis</i>	844243	JX412766	UK	Timmermans et al. (submitted)
	Hydrophilidae	Hydrophilinae	<i>Amphiops globus</i>	838127	JX412726	South Africa	Timmermans et al. (submitted)
	Hydrophilidae	Hydrophilinae	<i>Sternolophus</i> sp.	838163	JX412800	South Africa	Timmermans et al. (submitted)
	Hydrophilidae	Hydrophilinae	<i>Tropisternus</i> sp.		GU176339	USA	Song et al. (2010)
	Hydrophilidae	Hydrochinae	<i>Hydrochus capensis</i>	840193	HQ232801	South Africa	Timmermans et al. (2010)
Scarabaeiformia							
Scarabaeoidea	Geotrupidae	Bolboceratinae	<i>Bolboceratex</i> sp.	838179	JX412746	South Africa	Timmermans et al. (submitted)
	Geotrupidae	Geotrupinae	<i>Anaplotrupes stercorosus</i>	843249	JX412838	Germany	Timmermans et al. (submitted)
	Glaphyridae	Glaphyritinae	<i>Glaphyrus comosus</i>	843250	JX313673	Israel	Timmermans et al. (submitted)
	Glaresidae	Glaresinae	<i>Glaresis</i> sp.	843266	JX412819	South Africa	Timmermans et al. (submitted)
	Hybosoridae	Ceratocanthinae	<i>Ceratocanthus</i> sp.	843270	JX412772	Namibia	Timmermans et al. (submitted)
	Hybosoridae	Hybosorinae	<i>Hybosorus</i> sp.	838175	JX412828	South Africa	Timmermans et al. (submitted)
	Lucanidae	Lucaninae	<i>Dorcus parallelipedus</i>	843271	JX412841	Czech Republic	Timmermans et al. (submitted)
	Lucanidae	Lucaninae	gen. sp.	857087	JX313682	China	Timmermans et al. (submitted)
	Lucanidae	Lucaninae	<i>Lucanus mazama</i>		FJ613419	Not available	Sheffield et al. (2009)
	Lucanidae	Lucaninae	<i>Nigidius</i> sp.	843262	JX412771	South Africa	Timmermans et al. (submitted)
	Passalidae	Passalinae	gen. sp.	843267	JX412836	Peru	Timmermans et al. (submitted)
	Passalidae	Passalinae	gen. sp.	899706	JX412812	Belize	Timmermans et al. (submitted)
	Scarabaeidae	Melolonthini	<i>Schizonychia</i> sp.	843259	JX412739	South Africa	Timmermans et al. (submitted)
	Scarabaeidae	Aphodiini	<i>Aphodius</i> sp.	843252	JX412729	South Africa	Timmermans et al. (submitted)
	Scarabaeidae	Melolonthinae	<i>Asteropholis</i> sp.	838182	JX412755	South Africa	Timmermans et al. (submitted)
	Scarabaeidae	Dynastinae	<i>Cyphoristes vullatus</i>	843257	JX412731	South Africa	Timmermans et al. (submitted)
	Scarabaeidae	Cetoniini	<i>Leucoscelis</i> sp.	843253	JX412740	South Africa	Timmermans et al. (submitted)
	Scarabaeidae	Trichini	<i>Myoderma</i> sp.	843255	JX412847	South Africa	Timmermans et al. (submitted)
	Scarabaeidae	Sericini	<i>Pleophylla</i> sp.	843251	JX412736	South Africa	Timmermans et al. (submitted)
	Scarabaeidae	Anomalini	<i>Popillia</i> sp.	843256	JX412777	South Africa	Timmermans et al. (submitted)
	Scarabaeidae	Melolonthinae	<i>Rhopaea magnicornis</i>		NC013252	Australia	Cameron et al. (2010)
	Scarabaeidae	Scarabaeinae	<i>Sarophorus</i> sp.	843265	JX412735	South Africa	Timmermans et al. (submitted)
	Scarabaeidae	Anomalini	<i>Adoretus</i> sp.	843258	JX412788	South Africa	Timmermans et al. (submitted)
	Trogidae	Troginae	<i>Trox</i> sp.	843260	JX412734	South Africa	Timmermans et al. (submitted)
Bostrichiformia							
Bostrichoidea	Bostrichidae	Bostrichinae	<i>Apatides fortis</i>		NC013582	Not available	Sheffield et al. (2009)
	Bostrichidae	Bostrichinae	<i>Sinoxylon</i> sp.	838164	JX412742	South Africa	Timmermans et al. (submitted)
	Dermestidae	Anthreninae	<i>Anthrenus</i> sp.	840208	HQ232808	South Africa	Timmermans et al. (2010)
	Dermestidae	Anthreninae	<i>Altigenus nottentotus</i>	838157	JX412837	South Africa	Timmermans et al. (submitted)
	Dermestidae	Dermestinae	<i>Dermestes peruvianus</i>	835562	JX313683	UK	Timmermans et al. (submitted)
	Dermestidae	Dermestinae	gen. sp.	838145	JX412839	South Africa	Timmermans et al. (submitted)
	Dermestidae	Trinodinae	<i>Trinodes nirtus</i>	724082	JX313664	UK	Timmermans et al. (submitted)
	Ptinidae	Anobiinae	gen. sp.	838099	JX412730	South Africa	Timmermans et al. (submitted)
	Ptinidae	Xyletininae	<i>Xyletinus</i> sp.	844236	JX313671	Slovakia	Timmermans et al. (submitted)
	Ptinidae	Ptininae	<i>Ptinus rufipes</i>	840469	HQ232819	Czech Republic	Timmermans et al. (2010)
Elateriformia							

Dascilloidea	Dascillidae	Dascillinae	<i>Dascillus cervinus</i>	833042	JQ034414	UK	Timmermans & Vogler (2012)
Buprestoidae	Buprestidae	Polycestinae	<i>Acmaeodera</i> sp.	838158	JX412781	South Africa	Timmermans et al. (submitted)
		Polycestinae	<i>Acmaeodera</i> sp.		NC013580	Not available	Sheffield et al. (2009)
		Agrilinae	<i>Agrilus biguttatus</i>	833027	JX412786	UK	Timmermans et al. (submitted)
		Agrilinae	<i>Agrilus</i> sp.	838102	JX412834	South Africa	Timmermans et al. (submitted)
		Buprestinae	<i>Anthaxia</i> sp.	840209	JX412831	South Africa	Timmermans et al. (submitted)
		Chrysochroinae	<i>Chrysochroa fulgidissima</i>		EU826485	Korea	Hong et al. (2009)
		Chrysochroinae	<i>Perotis lugubris</i>	840447	X412762	Greece	Timmermans et al. (submitted)
Byrrhoidea	Byrrhidae	Byrrhinae	<i>Byrrhus</i> sp.	840448	JQ034419	Slovakia	Timmermans & Vogler (2012)
		Mastiginae	gen. sp.	840216	HQ232810	South Africa	Timmermans et al. (2010)
		Elminae	gen. sp.	842675	JX412724	Czech Republic	Timmermans et al. (submitted)
		Elminae	<i>Elimis aenea</i>	843150	JX412821	Slovakia	Timmermans et al. (submitted)
		Elminae	<i>Limnius perrisi</i>	843148	JX412747	Slovakia	Timmermans et al. (submitted)
		Eulichadinae	<i>Eulichas baeri</i>	840452	HQ232812	Philippines	Timmermans et al. (2010)
		Heterocerinae	<i>Heterocerus fenestratus</i>	840449	HQ232811	Slovakia	Timmermans et al. (2010)
		Limnichinae	<i>Byrrhinus</i> sp.	838097	JX412827	South Africa	Timmermans et al. (submitted)
		Limnichinae	gen. sp.	842697	JQ034416	Indonesia	Timmermans & Vogler (2012)
		Ptilodactylinae	<i>Ptilodactyla</i> sp.	840214	JX412727	South Africa	Timmermans et al. (submitted)
Elateroidea	Artematopodidae	Artematopodinae	<i>Eurypogon</i> sp.	842698	JX412809	Japan	Timmermans et al. (submitted)
		Cantharinae	<i>Malthodes minimus</i>	842708	JX412853	Czech Republic	Timmermans et al. (submitted)
		Cantharinae	<i>Cantharis pellucida</i>	840465	HQ232817	Czech Republic	Timmermans et al. (2010)
		Chauliognathinae	<i>Chauliognathus opacus</i>		NC013576	Not available	Sheffield et al. (2009)
		Elaterinae	<i>Agriotes ustulatus</i>	842673	JX412737	Czech Republic	Timmermans et al. (submitted)
		Elaterinae	<i>Dicronychus</i> sp.	844235	JX412848	Slovakia	Timmermans et al. (submitted)
		Drilinae	<i>Drilus flavescens</i>	840459	HQ232815	Malta	Timmermans et al. (2010)
		Elaterinae	<i>Propsephus</i> sp.	838123	JX412817	South Africa	Timmermans et al. (submitted)
		Agrypninae	<i>Pyrophorus divergens</i>		NC009964	Brasil	Arnoldi et al. (2007)
		Macraulacinae	<i>Formax</i> sp.	899821	JX412858	Belize	Timmermans et al. (submitted)
		Ototretinae	<i>Drilaster</i> sp.	840462	HQ232816	Japan	Timmermans et al. (2010)
		Lampyrinae	<i>Pyrocoelia rufa</i>		AF452048	Korea	Bae et al. (2004)
		Dictyopterinae	<i>Dictyoptera aurora</i>	844238	JX412733	Slovakia	Timmermans et al. (submitted)
		Lycidae	<i>Merolycus dentipes</i>	840457	HQ232814	South Africa	Timmermans et al. (2010)
		Omalisinae	<i>Omalisus fontisbellaquaei</i>	840456	JX412744	Slovakia	Timmermans et al. (submitted)
		Omethinae	<i>Drilonius striatulus</i>	840464	JX412822	Japan	Timmermans et al. (submitted)
		Phengodinae	<i>Mastinocerus</i> s.l. sp.	900076	JX412758	Belize	Timmermans et al. (submitted)
		Rhagophthalmidae	gen. sp.	840461	JX412750	Indonesia	Timmermans et al. (submitted)
		Rhagophthalmidae	<i>Rhagophthalmus lufengensis</i>		NC010969	China	Li et al. (2007)
		Rhagophthalmidae	<i>Rhagophthalmus ohbai</i>		NC010964	Japan	Li et al. (2007)
		Throscidae	<i>Trixagus</i> sp.	840455	JX412793	Czech Republic	Timmermans et al. (submitted)
Cucujiformia							
Cleroidea	Chaetosomatidae	Chaetosomatinae	<i>Chaetosoma scaritides</i>		NC011324	New Zealand	Sheffield et al. (2008)
	Melyridae		gen. sp.	838169	JX412765	South Africa	Timmermans et al. (submitted)
	Melyridae		gen. sp.	840183	JX412770	South Africa	Timmermans et al. (submitted)
	Melyridae		<i>Psilothrix</i> sp.	840200	JX412801	South Africa	Timmermans et al. (submitted)
	Melyridae	Malachiinae	<i>Cianoptilus assimilis</i>	838184	JX412833	Greece	Timmermans et al. (submitted)
	Melyridae	Malachiinae	gen. sp.	838161	JX412799	South Africa	Timmermans et al. (submitted)

	Phioiophilidae	Phioiophilinae	<i>Phioiophilus edwardsi</i>	842910	JX412815	UK	Timmermans et al. (submitted)
	Phycosecidae	Phycosecinae	<i>Phycosecis limbata</i>	842917	JX412820	New Zealand	Timmermans et al. (submitted)
	Trogossitidae	Trogossitinae	<i>Tennoscheila virescens</i>	900004	JX412752	Belize	Timmermans et al. (submitted)
Lymexyloidea	Lymexyloidea	Lymexyloinae	<i>Hylecoetus dermestoides</i>	840470	HQ232820	Czech Republic	Timmermans et al. (2010)
	Aderidae	Aderinae	gen. sp.	838114	JX412776	South Africa	Timmermans et al. (submitted)
	Anthicidae	Anthicinae	<i>Formicomus</i> sp.	842678	JX412763	Cameroon	Timmermans et al. (submitted)
	Anthicidae	Anthicinae	<i>Omonadus floralis</i>	840182	JX412857	South Africa	Timmermans et al. (submitted)
	Boridae	Borinae	<i>Boros schneideri</i>	840483	HQ232825	Czech Republic	Timmermans et al. (2010)
	Çiidae	Çiidae	gen. sp.	840483	HQ232823	Slovakia	Timmermans et al. (2010)
	Melandryidae	Melandryinae	<i>Mikadonius gracilis</i>	838173	JX412846	South Africa	Timmermans et al. (submitted)
	Melandryidae	Melandryidae	<i>Ospya bipunctata</i>	842714	JX412823	Japan	Timmermans et al. (submitted)
	Meloidae	Meloidae	gen. sp.	835592	JX313675	UK	Timmermans et al. (submitted)
	Meloidae	Meloidae	<i>Mylabris</i> sp.	840179	HM486073	South Africa	Timmermans et al. (submitted)
	Mordellidae	Mordellinae	<i>Mordella atrata</i>	838144	JX412732	South Africa	Timmermans et al. (submitted)
	Mordellidae	Mordellinae	gen. sp.	FJ859904	FJ859904	USA	Cameron et al. (2010)
	Mycteridae	Hemipeplinae	<i>Hemipeplus</i> sp.	838118	JX412844	South Africa	Timmermans et al. (submitted)
	Mycetophagidae	Mycetophaginae	<i>Mycet. quadripustulatus</i>	900019	JX412852	Belize	Timmermans et al. (submitted)
	Mycetidae	Laconotinae	<i>Sitiponotus mexicanus</i>	840485	HQ232824	Slovakia	Timmermans et al. (2010)
	Oedemeridae	Oedemerinae	<i>Ischnomera cyanea</i>	899983	JX412811	Belize	Timmermans et al. (submitted)
	Oedemeridae	Oedemerinae	<i>Oedemera virescens</i>	840494	JX412790	Czech Republic	Timmermans et al. (submitted)
	Prostomidae	Prostominae	<i>Prostomis</i> sp.	840491	HQ232826	Czech Republic	Timmermans et al. (2010)
	Pyrochroidae	Ischaliinae	<i>Ischalia</i> sp.	842709	JX412787	Indonesia	Timmermans et al. (submitted)
	Scrautiidae	Anaspidinae	gen. sp.	840493	HQ232827	Laos	Timmermans et al. (2010)
	Scrautiidae	Anaspidinae	<i>Anaspis</i> sp.	840252	JX412791	South Africa	Timmermans et al. (submitted)
	Scrautiidae	Anaspidinae	<i>Anaspis</i> sp.	840206	HQ232806	South Africa	Timmermans et al. (2010)
	Scrautiidae	Anaspidinae	<i>Anaspis</i> sp.	838162	JX412856	South Africa	Timmermans et al. (submitted)
	Scrautiidae	Anaspidinae	<i>Anaspis</i> sp.	840490	JX412814	Czech Republic	Timmermans et al. (submitted)
	Scrautiidae	Scrautiinae	<i>Scraptia</i> sp.	838150	JX412851	South Africa	Timmermans et al. (submitted)
	Scrautiidae	Scrautiinae	<i>Scraptia</i> sp.	838154	JX412825	South Africa	Timmermans et al. (submitted)
	Scrautiidae	Scrautiinae	<i>Scraptia</i> sp.	840489	JX412798	Indonesia	Timmermans et al. (submitted)
	Tenebrionidae	Lagriinae	<i>Adelium</i> sp.	NC013554	NC013554	Not available	Sheffield et al. (2009)
	Tenebrionidae	Aleculinae	gen. sp.	838142	JX412775	South Africa	Timmermans et al. (submitted)
	Tenebrionidae	Aleculinae	gen. sp.	840222	JX412818	South Africa	Timmermans et al. (submitted)
	Tenebrionidae	Lagriinae	<i>Chrysolagria</i> sp.	838101	JX412760	South Africa	Timmermans et al. (submitted)
	Tenebrionidae	Lagriinae	<i>Eutrapeia ruficollis</i>	840203	HQ232805	South Africa	Timmermans et al. (2010)
	Tenebrionidae	Aleculinae	<i>Paramarygmus</i> sp.	838115	JX412808	South Africa	Timmermans et al. (submitted)
	Tenebrionidae	Lagriinae	<i>Paratenetus cf. tropicalis</i>	899913	JX412774	Belize	Timmermans et al. (submitted)
	Tenebrionidae	Diaperinae	<i>Platydera</i> sp.	838098	JX412842	South Africa	Timmermans et al. (submitted)
	Tenebrionidae	Coleometopinae	<i>Strongylium suspicax</i>	838137	JX412780	South Africa	Timmermans et al. (submitted)
	Tenebrionidae	Tenebrioninae	<i>Tribolium castaneum</i>	NC003081	NC003081	USA	Friedrich and Mugim (2003)
Cucujoidae	Coccinellidae	Coccidulinae	<i>Coccidula rufa</i>	832905	JX412767	UK	Timmermans et al. (submitted)
	Coccinellidae	Coccinellinae	<i>Coccinella septempunctata</i>	JQ321839	JQ321839	Korea	Kim et al. (2012)
	Cerylonidae	Ceryloninae	<i>Cerylon histeroideus</i>	840476	HQ232821	Slovakia	Timmermans et al. (2010)
	Corylophidae	Corylophinae	<i>Gloeosoma</i> sp.	842694	JX412843	Indonesia	Timmermans et al. (submitted)
	Cryptophagidae	Cryptophaginae	<i>Micrambina</i> sp.	842921	JX313663	New Zealand	Timmermans et al. (submitted)
	Cucujidae	Cucujinae	<i>Cucujus clavipes</i>	GU176341	GU176341	USA	Song et al. (2010)

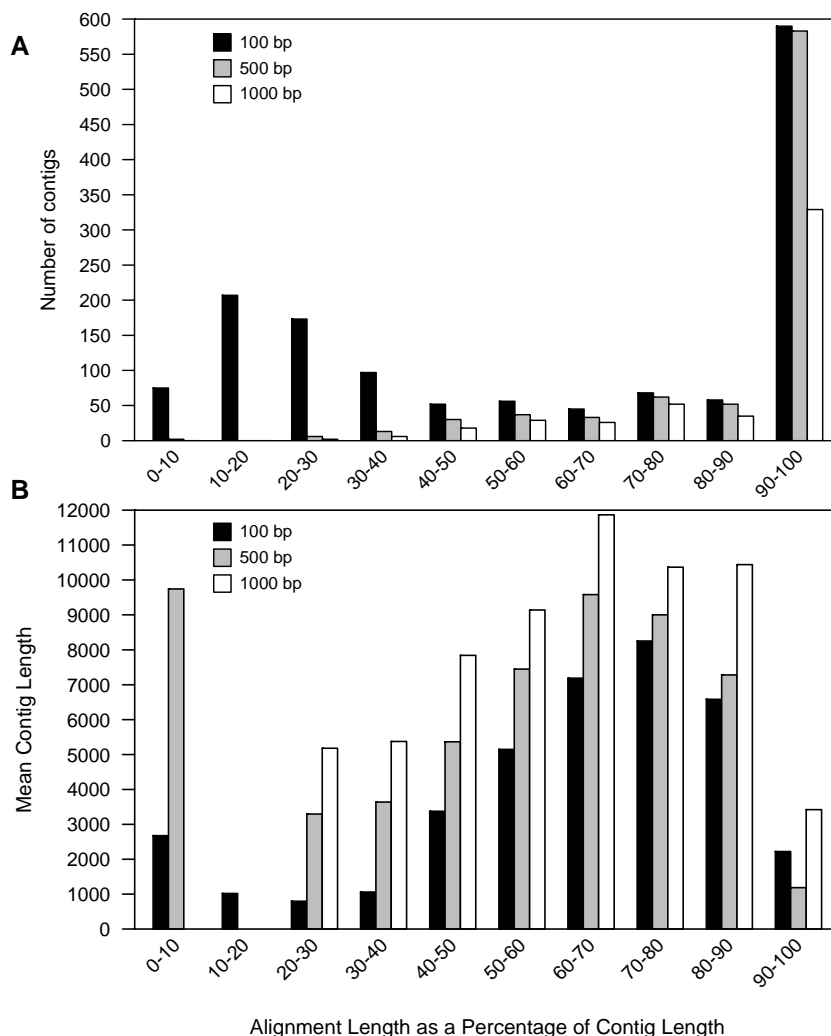
	Discolomatidae	Discolomatinae	gen. sp.	842682	JX412748	Philippines	Timmermans et al. (submitted)
	Erotylidae	Xenoscelinae	<i>Loberontha olivascens</i>	842922	JX412784	New Zealand	Timmermans et al. (submitted)
	Erotylidae	Erotylinae	<i>Tritoma bipustulata</i>	840479	HQ232822	Czech Republic	Timmermans et al. (2010)
	Endomychidae	Endomychinae	<i>Endomychus coccineus</i>	835578	JX313667	UK	Timmermans et al. (submitted)
	Kateretidae	Kateretinae	<i>Kateretes pedicularius</i>	724182	JX313689	UK	Timmermans et al. (submitted)
	Latridiidae	Latridiinae	<i>Elicmus brevicornis</i>	834305	JX313681	UK	Timmermans et al. (submitted)
	Nitidulidae	Cryptarchinae	<i>Glischrochilus hortensis</i>	832900	JX412778	UK	Timmermans et al. (submitted)
	Priasilphidae	Priasilphinae	<i>Priasilpha obscura</i>	EU877952		New Zealand	Sheffield et al. (2008)
	Phalacridae		gen. sp.	840198	HQ232803	South Africa	Timmermans et al. (2010)
	Propalticidae	Propalticinae	<i>Propalticus</i> sp.	844251	JX412792	Australia	Timmermans et al. (submitted)
	Sphindidae	Sphindinae	<i>Sphindus dubius</i>	833440	JX412803	UK	Timmermans et al. (submitted)
	Cerambycidae	Disteniinae	<i>Distenia fimbriata</i>	668016	JX221000	Belize	Timmermans et al. (submitted)
	Cerambycidae	Lepturinae	<i>Necydalis ulmi</i>	704364	JX220989	France	Timmermans et al. (submitted)
	Cerambycidae	Lepturinae	<i>Rhagium mordax</i>	832872	JX412743	UK	Timmermans et al. (submitted)
	Cerambycidae	Lepturinae	<i>Stenurella nigra</i>	704360	JX220998	Spain	Timmermans et al. (submitted)
	Cerambycidae	Lamiinae	<i>Psacotha hillaris</i>		NC013070	Korea	Kim et al. (2009)
	Cerambycidae	Lamiinae	<i>Anoplophora glabripennis</i>		NC008221	China	An et al. (unpublished)
	Cerambycidae	Lamiinae	gen. sp.	838149	JX412849	South Africa	Timmermans et al. (submitted)
	Cerambycidae	Prioninae	gen. sp.	744987	JX220991	Madagascar	Timmermans et al. (submitted)
	Chrysomelidae	Bruchinae	gen. sp.	838117	JX412832	South Africa	Timmermans et al. (submitted)
	Chrysomelidae	Cassidinae	<i>Arescus labiatus</i>	704488	JX220988	French Guiana	Timmermans et al. (submitted)
	Chrysomelidae	Cassidinae	<i>Imatidium capense</i>	704496	JX220993	French Guiana	Timmermans et al. (submitted)
	Chrysomelidae	Cassidinae	<i>Laccoptera ruginosa</i>	838107	JX412753	South Africa	Timmermans et al. (submitted)
	Chrysomelidae	Chrysomelinae	<i>Phaedon armoraciae</i>	704385	JX220992	UK	Timmermans et al. (submitted)
	Chrysomelidae	Chrysomelinae	<i>Timarcha goettingensis</i>		JX412789	UK	Timmermans et al. (submitted)
	Chrysomelidae	Chlamysinae	<i>Exema canadensis</i>	704405	JX220990	USA	Timmermans et al. (submitted)
	Chrysomelidae	Clytrinae	<i>Peplopera acromialis</i>	840215	HQ232809	South Africa	Timmermans et al. (submitted)
	Chrysomelidae	Criocerinae	<i>Criocera duodecimpunctata</i>		NC003372	Canada	Stewart & Beckenbach (2003)
	Chrysomelidae	Criocerinae	<i>Lema</i> sp.	838112	JX412769	South Africa	Timmermans et al. (submitted)
	Chrysomelidae	Cryptocephalinae	<i>Cryptocephalus</i> sp.	838109	JX412804	South Africa	Timmermans et al. (submitted)
	Chrysomelidae	Donaciinae	<i>Donacia distincta</i>	704419	JX220999	Canada	Timmermans et al. (submitted)
	Chrysomelidae	Eumolpinae	<i>Pseudocolaspis</i> sp.	838113	JX412756	South Africa	Timmermans et al. (submitted)
	Chrysomelidae	Galerucinae	<i>Chapuisia ellenbergeri</i>	838108	JX412757	South Africa	Timmermans et al. (submitted)
	Chrysomelidae	Galerucinae	<i>Fleutiauxia armata</i>	846535	JX412797	China	Timmermans et al. (submitted)
	Chrysomelidae	Galerucinae	<i>Sphenoraria micans</i>	846596	JX412854	China	Timmermans et al. (submitted)
	Chrysomelidae	Spilopyrinae	<i>Spilopyra sumptuosa</i>	704500	JX220997	Australia	Timmermans et al. (submitted)
	Chrysomelidae	<i>Sphaerotherix</i> sp.	unrecognized name	838106	JX412783	South Africa	Timmermans et al. (submitted)
	Cerambycidae	Cerambycinae	<i>Closteromerus claviger</i>	840202	HQ232804	South Africa	Timmermans et al. (2010)
	Orsodacnidae	Orsodacninae	<i>Orsodacne lineola</i>	679237	JX220994	UK	Timmermans et al. (submitted)
	Megalopodidae	Megalopodinae	<i>Zeugophora</i> sp.	840207	HQ232807	South Africa	Timmermans et al. (2010)
	Anthribidae	Anthribinae	<i>Chirotenon longimanus</i>	838143	JX412830	South Africa	Timmermans et al. (submitted)
	Anthribidae	Anthribinae	<i>Platystomos albinus</i>	847772	JN163968	France	Haran et al. (2013)
	Anthribidae	Urodontinae	<i>Urodontus glabratus</i>	840197	JX412859	South Africa	Timmermans et al. (submitted)
	Atelabidae	Atelabinae	<i>Apoderus coryli</i>	847768	JN163966	France	Haran et al. (2013)
	Atelabidae	Rhynchitinae	<i>Byctiscus populi</i>	847767	JN163965	France	Haran et al. (2013)
	Atelabidae	Rhynchitinae	<i>Deporaus betulae</i>	832883	JN163945	UK	Haran et al. (2013)

Figure S1. Raw contig length distribution for the Celera Assembler and IDBA-UD assemblies, and the non-redundant set generated by merging these in Geneious. The number of contigs longer than 15 kb that could be circularised in each case is shown within the corresponding circular arrow.



Supplementary Information S1. Choice of BLAST alignment length cut-off when filtering contigs against MitoDB.

The 1 kb threshold was chosen as a good balance between the correct retention of contigs for the inclusion in later analyses, and the loss of shorter contigs that are truly mitochondrial but do not contribute to the final dataset. Determining what is a “true” mitochondrial contig without a closely related reference sequence can be challenging and we aim to include only the most plausible and informative contigs in the final dataset. To illustrate this we present an analysis based on the blastn results for the Celera Assembler contigs against MitoDB. Here we look at the alignment length as a proportion of contig length, with the assumption that true mitochondrial contigs are likely to have proportionately longer hits to the MitoDB sequences than non-mitochondrial contigs. Three different alignment length cut-offs are tested (100 bp; 500 bp; 1 kb). Figure A indicates that a short alignment length cutoff of 100 bp retains large numbers of contigs that are unlikely to be mitochondrial as the BLAST alignments to the database are short relative to their total length (0-40%). From Figure B we can see that those contigs are also, on average, quite short (<3kb). Including these contigs will likely increase downstream processing complexity with minimal gain in aligned data as most, if not all, will be non-mitochondrial. In contrast, Figure A indicates that alignment length cutoffs of 500bp and 1 kb give very similar alignment proportion profiles, except that with a 1 kb cutoff you retain overall a lower number of contigs, especially in the 90-100% aligned category. This suggests that there are a large number of contigs retained by the 500 bp cutoff that have long alignments relative to their lengths (i.e. are probably mitochondrial) but are <1kb in total length. This is reinforced by the low average length of the contigs in this category in Figure B. We can therefore estimate there are at least 250 mitochondrial contigs of 0.5-1kb from this assembly that we have missed from our analysis (this set has a mean length of 713 bp). Simultaneously, the 500 bp cutoff retains two contigs of 9kb with alignment proportions of <10%. BLASTing these on the NCBI website indicates that they are bacterial but there are no highly similar matches. Using a 1 kb cutoff has correctly filtered these out, but at the cost of losing a large number of short contigs. Given that our aim is to retain contigs that can ultimately be included in the phylogenetic analysis, the loss of these short contigs is not problematic for this study as they would not have met the 2 kb (including *nad4l*) criterion.



Supplementary Information S2. The effect of quality control on the IDBA-UD assembly.

The IDBA-UD assembly statistics and mitochondrial contig length distributions following two different quality control procedures are outlined in the tables below. In addition, we made comparisons between the circular genomes from the two assemblies. Out of 45 mitogenomes found following QC-2, 39 are >99.9% identical to ones from QC-1. The remaining 6 are circular versions of >15kb linear contigs from QC-1 (at >99% identity i.e. they assemble slightly better under QC-2). 9 of the 10 remaining circular contigs from QC-1 are recovered >10kb at >99% identity under QC-2. The final one was recovered at 100% identity within the PCGs but there are differences in the control region. The contigs from QC-1 were retained for analysis.

Assembly statistics

IDBA-UD Assembly	Contigs	N50	Maximum contig length	Mean contig length	Total assembly length
QC-1 ^a	12588	794	19201	830	10457969
QC-2 ^b	10851	837	19044	867	9412984

Mitochondrial contig length distribution

IDBA-UD Assembly	1-5 kb	5-10 kb	10-15 kb	>15 kb (circular)
QC-1 ^a	424	41	18	81 (49)
QC-2 ^b	421	45	22	74 (45)

^a Prinseq-lite: minimum length 150 bp; trim 3' bases below Q20; minimum mean quality Q25; no Ns; both reads per pair must pass QC. Total number of pairs passing QC: Short insert library 846,156; Long insert library 1,260,119.

^b Prinseq-lite: minimum length 150 bp; trim 5' and 3' bases below Q30; minimum mean quality Q30; no Ns; both reads per pair must pass QC. Total number of pairs passing QC: Short insert library 690,388; Long insert library 1,131,850.

Table S3. Statistics derived from mapping quality controlled reads to the non-redundant contigs (n=429) with five programs using default settings for paired reads. Additionally SMALT was tested with a requirement for 98% identity under otherwise default parameters.

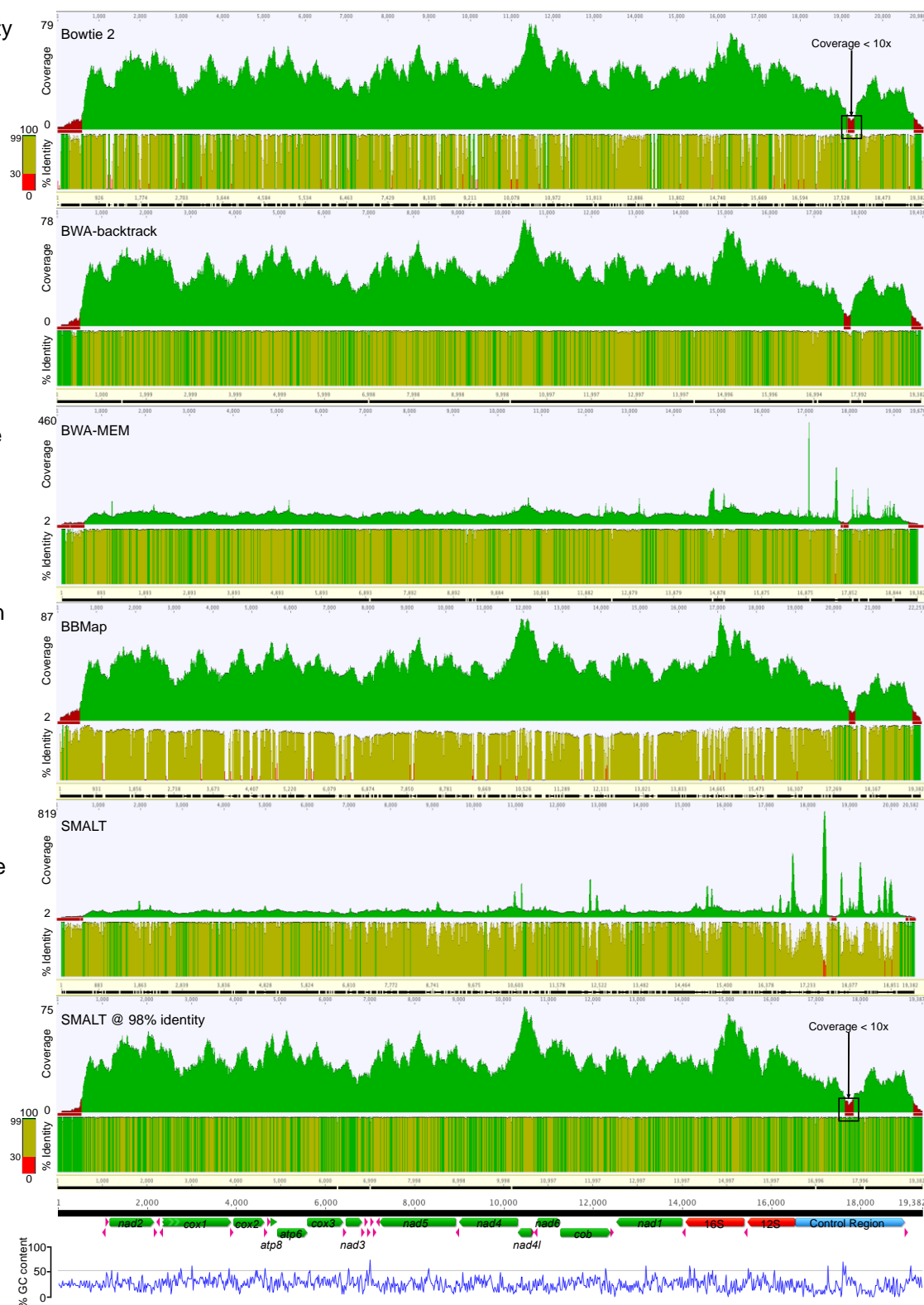
Program	Minimum mean coverage (s.d.)	Maximum mean coverage (s.d.)	Mean coverage (s.d.)	% bases at 0x	% bases at 3x	% bases at 5x	% bases at 10x	No. contigs where $\sigma > \mu$
Bowtie	1.5 (1.0)	351.2 (133.9)	32.3 (49.0)	0.28	96.49	89.81	70.06	6
BWA aln	0.8 (0.8)	340.2 (131.1)	30.6 (46.8)	0.54	95.74	88.4	68.24	4
BWA mem	3.6 (2.4)	557.4 (3139.1)	43.4 (265.6)	0.09	98.31	93.9	78.24	200
bbmap	0.3 (0.5)	389.7 (81.5)	34.9 (51.0)	1.18	95.86	89.89	71.13	24
SMALT	3.9 (2.9)	469.9 (1255.8)	57.5 (279.3)	0.08	98.6	95.09	81.82	303
SMALT (98%)	1.55 (0.80)	313.1 (126.0)	27.7 (42.5)	0.7	94.59	86.1	64.91	3

Minimum % mismatch	Maximum % mismatch	Mean % mismatch	% GC ^b	Mapping Quality	No. reads mapped	No. reads mapped (proper pairs)	No. reads mapped (singletons)
0.1	6.6	1.2	23.56	32.95	387,045	233,488	26,485
0	3.1	0.5	23.61	48.96	379,632	356,642	10,520
1	8	1.1	20.06	45.05	1,370,236	1,251,539	76,361
0.2	55.9 ^a	3.7	23.23	39.19	404,752	362,374	34,632
0.7	15.1	5.5	16.63	35.79	2,151,320	615,390	872,796
0	1.3	0.3	23.63	48.39	330,544	280,428	50,116

^a Setting maxindel=0 reduces this to 11.7%

^b GC content of contigs: 23.9%

Figure S2. The variability in mapping quality for a single contig. Quality-controlled reads were mapped against all non-redundant contigs. This example shows the coverage and mean pairwise identity of the reads at each position for the longest contig (19,382 bp). All programs introduced gaps into the contig and coverage always decreases at the edges. In all cases there is a region of <10x coverage in the control region. This may indicate misassembly, although both assemblers reconstructed this region identically. BWA-MEM and SMALT map the largest number of reads but the coverage is highly uneven. For the other four examples the coverage maps are similar and more even, although Bowtie 2 and BMAP introduce a large number of gaps into the contig and many positions have low pairwise identity between reads (i.e. many disagreements). BWA-backtrack and SMALT (mapping at 98% identity) give a very similar result and are likely to be the best representation of the assembly, although the number of reads mapped is likely to be conservative.



Software	Mean Coverage (s.d)	No. reads mapped	% Mismatch	No. gaps in contig
Bowtie 2	39.60 (12.80)	3240	0.9	1493
BWA-backtrack	38.77 (12.75)	3284	0.5	10
BWA-MEM	45.06 (22.88)	7001	0.6	49
BMAP	42.98 (NaN)	3342	6.3	2773
SMALT	52.69 (47.19)	10324	2.6	1057
SMALT (98%)	35.36 (12.36)	2884	0.3	5

Figure S3. Coverage plots derived from mapping quality controlled reads with SMALT at 98% identity. A. Percentage of bases in non-redundant contigs for the given coverage (maximum of 50X shown). B. Percentage of bases in non-redundant contigs with at least the given coverage (maximum of 50X shown – 13% of bases have >50X coverage). C. Mean coverage of mitochondrial contigs from the CA assembly plotted against contig length. D. Mean coverage of mitochondrial contigs from the IDBA-UD assembly plotted against contig length. E. Mean coverage of mitochondrial contigs from the non-redundant set plotted against contig length.

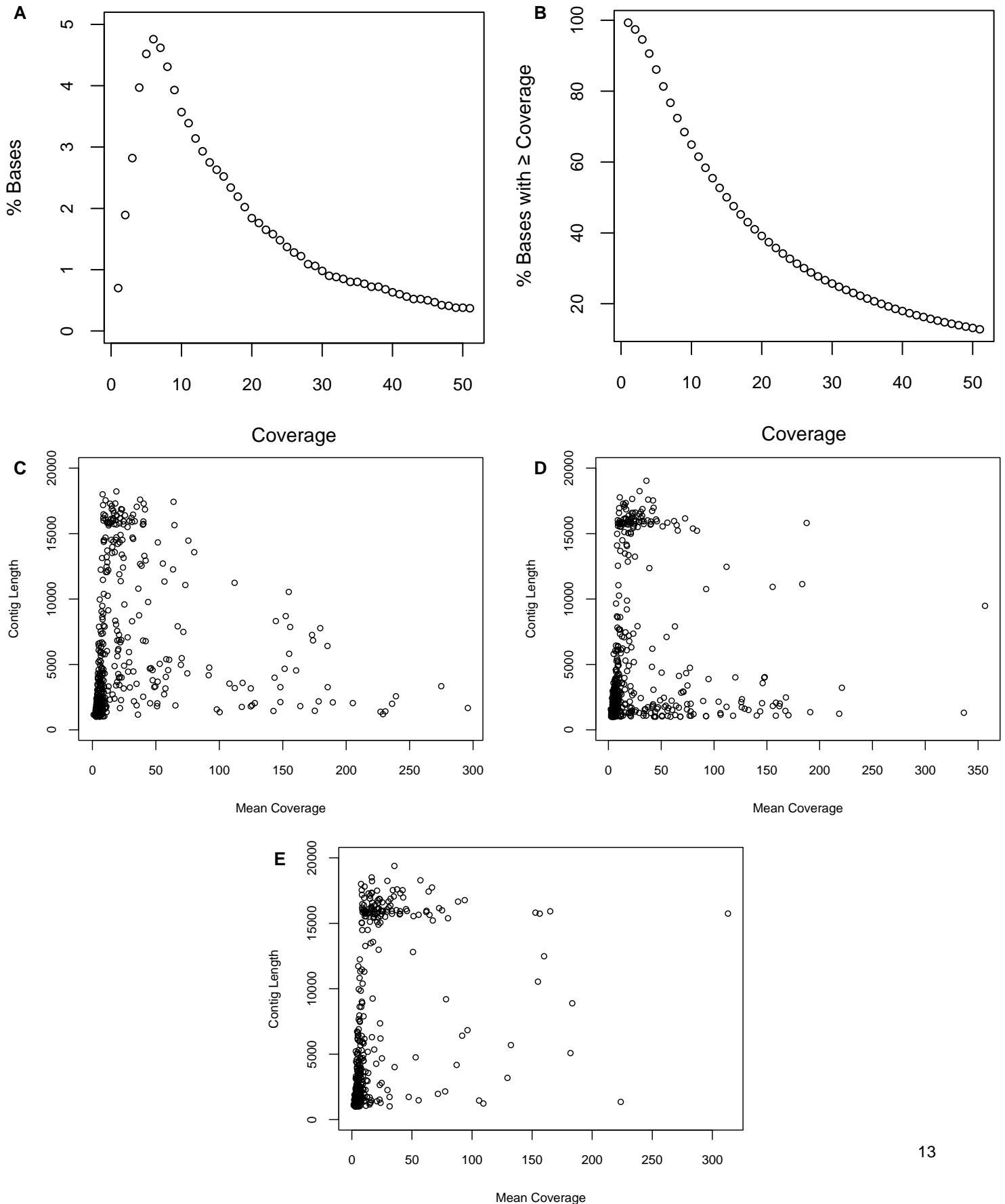


Figure S4. The number of protein-coding genes contributed to the final alignments per contig, for all 429 contigs in the non-redundant set. Approximately the same number of contigs contributed a single gene sequence (120) as the full complement of 13 mitochondrial protein-coding genes (112). Overall, contigs contributing 1-3 protein-coding genes made up 15.65% of the nucleotide matrix whilst those contributing 11-13 made up 70.32%.

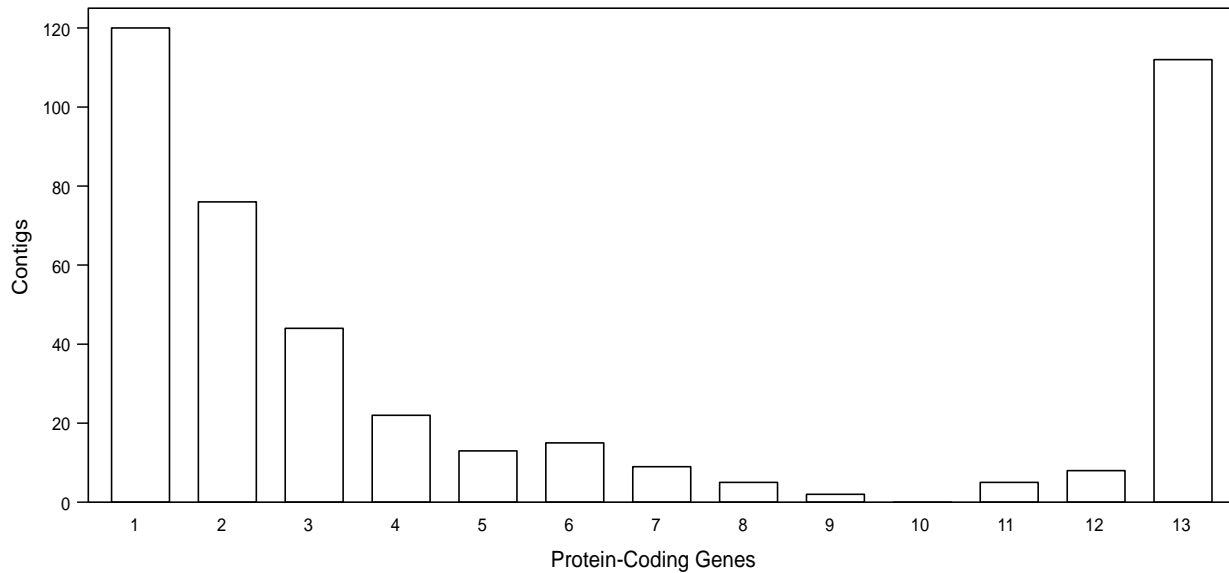
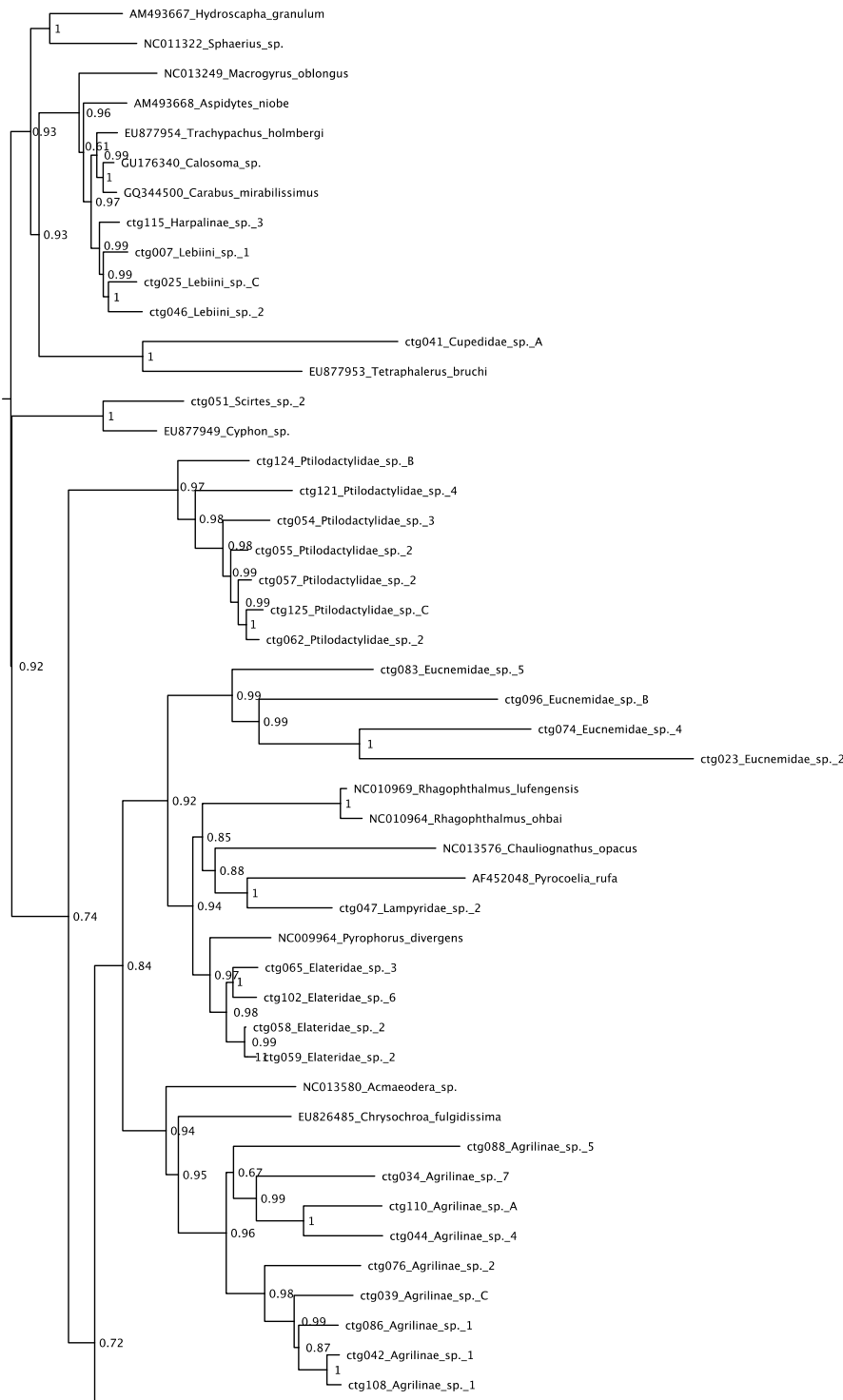
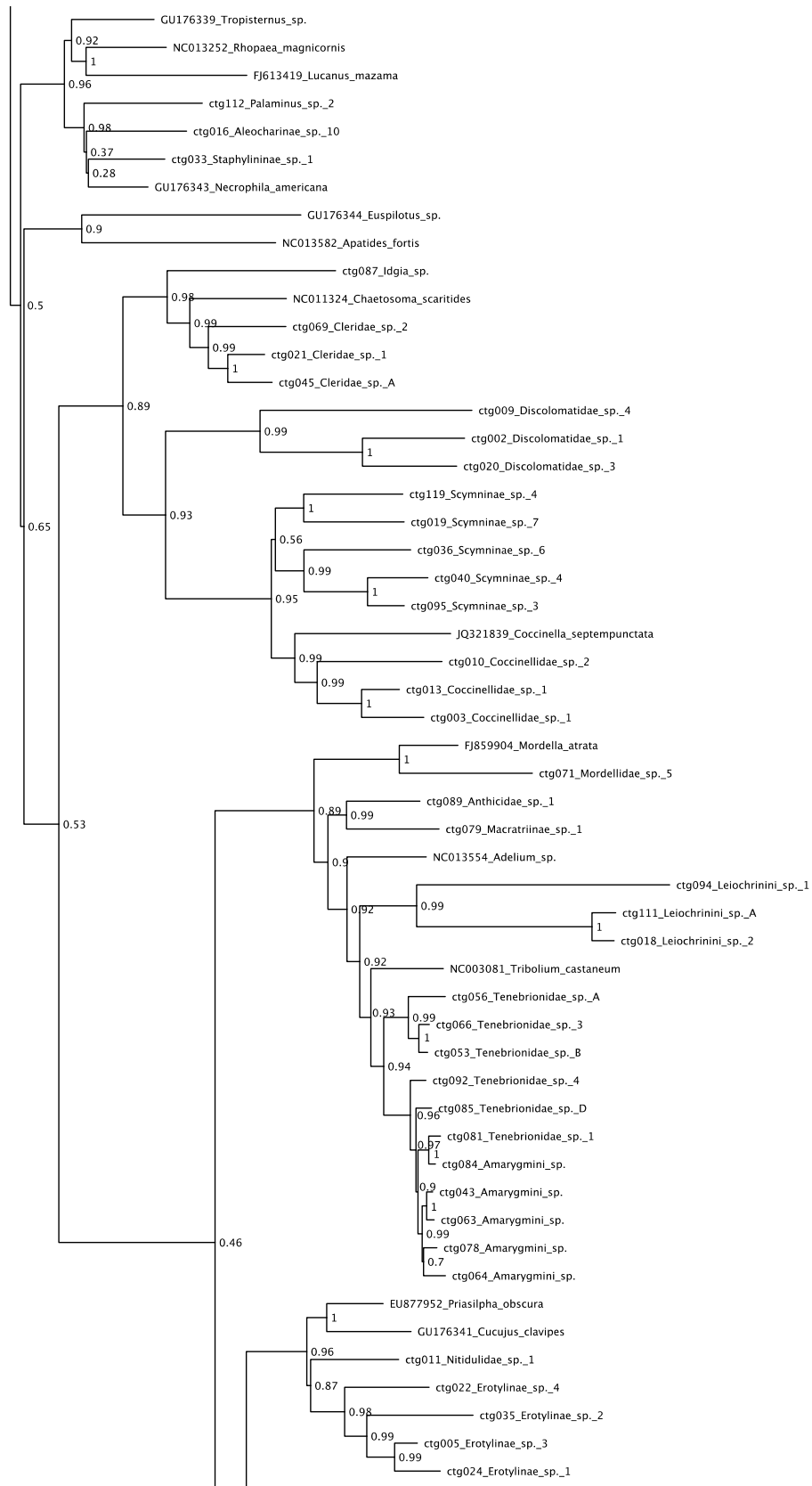


Figure S5. The mitochondrial tree-of-life for beetles under reduced taxon sampling. Maximum clade credibility consensus tree from Phylobayes analysis of 34 complete mitochondrial genomes published on GenBank (listed in Supplementary Table S2) and 124 novel 10+ protein-coding gene contigs from this study. Reference sequences are labelled with accession number and species name. Novel contigs are labelled with the most precise available taxonomic identification – species labels with numbers (and any below family-level without numbers) derive from morphospecies identifications based on bait sequences, whilst species labels with letters derive from contig position in the full beetle phylogeny (Figure 2), requiring monophyly with named reference sequences. Node labels are posterior probability values.







Supplementary Information S3. Failure to assign two contigs to correctly to family under requirement for monophyly with the Mito240 set.

The reliability of phylogenetic assignment relative to the fully identified Mito240 sequences was tested using the 96 10+ contigs that could be linked to specimens and family-level identification via the *cox1* barcodes. In the analysis with Mito240, 73 contigs were correctly assigned to family and a further 23 contigs to superfamily. In all but one of the latter cases, the failure to make a taxon assignment to family level was due to limited representations of small families (e.g. in Cleroidea). The exception to this was a single contig linked to Anthicidae via a bait sequence, which was placed as sister to Mycetophagidae (and therefore correctly classified as Tenebrionoidea under the requirement for strict monophyly) in spite of the formation of an otherwise monophyletic Anthicidae clade containing two reference sequences and one baited contig containing all protein-coding genes. One remaining contig could only conservatively be assigned to suborder as it was not placed within a named clade, but both its position and bait identification were consistent with assignment to the phylogenetically isolated superfamily Scirtoidea.

Supplementary Information S4. PCR primers and conditions used to amplify DNA barcodes (*cox1-5'*). For specimens that did not amplify at the first attempt (with COIF2/R2), and second attempt was made with LCO/HCO.

PCR primers

Attempt	Primer	Primer sequence 5' → 3'
1	COIF2	TCTACYAATCATAAAGATATTGGTAC
	COIR2	ACTTCTGGATGACCAAAGAATCA
2	LCO 1490	GGTCAACAAATCATAAAGATATTGG
	HCO 2198	TAAACTTCAGGGTGACCAAAAATCA

PCR conditions with Promega GoTaq Green Master Mix

Step	Temp. (°C)	Time	No. of cycles
Initialisation	95	2 min	1
Denaturation	95	30 sec	
Annealing	40	30 sec	37
Extension	72	45 sec	
Final Extension	72	5 min	1
Final Hold	4	∞	

Table S4. Programs and settings used for mitogenome assembly.

Step	Aim	Software/process	Parameters	Additional requirements	Source
1	Quality assessment	FastQC 0.10.1	NA	NA	http://www.bioinformatics.babraham.ac.uk/project/s/fastqc/
2	Adapter removal	Trimmomatic 0.30	ILLUMINACLIP:TruSeq3-PE.fa:2:30:10	TruSeq3-PE.fa (packaged with Trimmomatic)	http://www.usadellab.org/cms/index.php?page=trimmomatic
3	Filter for putative mitochondrial reads	BLAST 2.2.27+ Extract paired reads	-task blastn -evalue 1e-5 -max_target_seqs 1 -outfmt 6 -dust no NA	Mitogenome reference database for target taxon NA	http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=E=BlastDocs&DOC_TYPE=Download Script 1, Supplementary Data
4	Assembly 1	Celera Assembler 7.0	doToggle=1 toggleUnitigLength=500 unitigger=bogart -min_len 150 -min_qual_score 20 -min_qual_mean 25 -ns_max_n 0 -trim_qual_right 20 NA	frg file & spec file (containing parameters and path to frg file) NA	http://sourceforge.net/apps/mediawiki/wgs-assembler/index.php?title=Main_Page http://prinseq.sourceforge.net
5	Assembly 2	Quality control with prinseq-lite.pl Re-pair reads IDBA-UD	--mink 80 --maxk 230 --similar 0.98 -task blastn -evalue 1e-5 -max_target_seqs 1 -outfmt 6 awk '{(\$4>=1000){print}'	NA NA Compile IDBA-UD to allow 250bp reads & kmers Mitogenome reference database for target taxon NA	Script 2, Supplementary Data http://i.cs.hku.hk/~aise/hkubrg/projects/idba_ud/ http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE_TYPE=E=BlastDocs&DOC_TYPE=Download
6	Filter Assemblies 1 & 2 for mitochondrial contigs	BLAST 2.2.27+ Filter BLAST hits for alignment length > 1kb Extract mitochondrial contigs with cdbfasta/cdbbyank COVE 2.4.4	NA	Filtered BLAST hit table from previous step tRNA covariance models	http://sourceforge.net/projects/cdbfasta/ http://selab.janelia.org/software.html
7	tRNA annotation	Generate annotated GenBank format files Add terminal tRNA annotations	NA	NA	Script 3, Supplementary Data
8	Extract inter-tRNA regions	FeatureExtract 1.2	feature type: tRNA; extract intergenic regions Iterate 3 times; maximum mismatches per read 50%; search more thoroughly for poor matching reads	NA	Script 4, Supplementary Data http://www.cbs.dtu.dk/services/FeatureExtract/
9	Identify genes by mapping	Geneious 6.1		Mitogenome reference e.g. <i>Tribolium castaneum</i> [GenBank:NC_003081]	http://www.geneious.com

Script 1. `FastqExtract.pl` – extract paired FASTQ reads based on BLAST output (i.e. filter FASTQ file using BLAST results).

Use concatenated BLAST results (tabular format) from R1 and R2 to run the script twice, once for each.

```
print "what is the name of the BLAST output file?\n";
$blastIn = <STDIN>;
chomp $blastIn;
print "what is the name of the raw reads file?\n";
$filename = <STDIN>;
chomp $filename;
print "what is the read identifier (first 6 header characters, not inc @)?\n";
$idf = <STDIN>;
chomp $idf;

#INPUT:

$fastq = $filename;
open(FASTQ, "<$fastq") or print "could not open file $fastq";

#HEADERS:

$headers = $blastreader;
chomp $headers;
open(HEADERS, "$headers") or print "could not open file $headers";
@headers = <HEADERS>;
close HEADERS;

foreach $head(@headers){
if($head =~ m/\w/){
@head = split(" ", $head);

#add to hash
$headers{$head[0]}=1}
}

$outfile = "$fastq\.out";

open(OUTFILE, ">$outfile") or print "could not open file $outfile";

#####

open(FASTQ, "<$fastq");
while(<FASTQ>){
$line = $_;

if($flag == 1){$flag = 2; print OUTFILE $line;}
elsif($flag == 2){$flag = 3; print OUTFILE $line;}
elsif($flag == 3){$flag = 0; print OUTFILE $line;}

    elsif($line =~ m/$idf:/){
        @line = split(" ", $line);
        $line[0] =~ s/\@//ig;

            if(exists $headers{$line[0]}){
                ++$counter; $flag = 1;
                print "$counter\:\t\t\t$line";
                print OUTFILE "$line";
            }
        }
    }
}
close OUTFILE;
```

Script 2. Re-pair quality controlled reads from Prinseq.

Source: <http://www.bigsa.org.au/node/92> (Last accessed 02/09/2013)

Requires cdbfasta from <http://sourceforge.net/projects/cdbfasta/>

Where in_R1.QC.fastq and in_R2.QC.fastq are quality controlled FASTQ files from Prinseq:

```
#concatenate R1 and R2
cat in_R[12].QC.fastq > in.QC.fq

#convert IDs to older Illumina format (needed for the awk script below)
cat in.QC.fq | awk '{if (NR % 4 == 1) {split($1, arr, ":"); printf
"%s_%s:%s:%s:%s:%s#0/%s\n", arr[1], arr[3], arr[4], arr[5], arr[6], arr[7], substr($2, 1,
1), $0} else if (NR % 4 == 3){print "+"} else {print $0} }' > in.QC.nwblnr.fq ;

#use cdbfasta to index the files
cdbfasta -Q in.QC.nwblnr.fq

#retrive list of IDs
cdbyank in.QC.nwblnr.fq.cidx -l > in.QC.ids

#retrieve list of IDs for paired reads
awk -v sep="/" '{ if ((sep_i=index($0,sep)) > 0) { name=substr($0,1,sep_i-1);
suffix=substr($0,sep_i); } else { name=$0; } if (r[name]) { print name r[name]; print $0;
delete r[name]; } else { r[name]=suffix; }}' in.QC.ids > in.QC.paired.ids

#make an interleaved FASTQ file for paired reads
cdbyank in.QC.nwblnr.fq.cidx < in.QC.paired.ids > in.QC.paired.fq
```


Script 3. cove_output.pl – generate tRNA annotated GenBank format files from COVE output files.

```

use Bio::SeqIO;

#Remove some (old) files first
@tempfiles = glob("*.tmp");
foreach $tempfile(@tempfiles){unlink $tempfile;}

#The COVE output files to process
@files = glob("*.out");

#Directory to store resulting GenBank formatted output files
mkdir(GB);

foreach $file(@files){
    open(FILE, "$file") or print "could not open $file";
    @data = <FILE>;
    close FILE;

    foreach $line(@data){

#new model -> hash,sort and start again
        if($line =~ m/Model/){
            @allKeys = keys(%hash);
            @allKeys = sort{ $a <=> $b} @allKeys;

#select one with highest value, but only if above ##40##
            $highest = pop(@allKeys);
            $value = $hash{$highest}; if($highest > 40){
                $order{$value} = ("Model\t$highest");
            }

#select penultimate one for Ser and Leu ##two tRNAs for each##
            if($model =~ m/Leu/){
                $highest = pop(@allKeys);
                $value = $hash{$highest}; if($highest > 40){
                    print "Model\t$highest\t$value\n";
                    $order{$value} = ("Model\t$highest");
                }
            }
            if($model =~ m/Ser/){
                $highest = pop(@allKeys);
                $value = $hash{$highest}; if($highest > 40){
                    $order{$value} = ("Model\t$highest");
                }
            }
            %hash = ();
            @line = split("\:", $line);
            $model = $line[1]; chomp $model;
            print "Model\n";
        }
        elsif($line =~ m/version/){next;}
        elsif(($line =~ m/^ /) || ($line =~ m/^1/)){
            $line =~ s/^1/^ 1/ig; #waarden, vul de hash
            ++$counter; @line = split(" {1,8}", $line);

#tRNA positions, add counter to make unique
            $position = "$line[2]\_$line[3]\_$counter";
            $hash{"$line[1]"} = "$position";
        }
        else{next;}
    }

    @allKeys = keys(%hash);
    @allKeys = sort{ $a <=> $b} @allKeys;
    $highest = pop(@allKeys);
    $value = $hash{$highest}; if($highest > 40){
        $order{$value} = ("Model\t$highest");
    }
    %hash = ();
    @allKeys = keys(%order);
    @allKeys = sort{ $a <=> $b} @allKeys;
    $file2 = $file;

```

```

$file2 =~ s/out/tmp/ig; #temporary file
$file3 = $file;
$file3 =~ s/\.out//ig; #base name of genbank flat file

#change to GenBank flatfile format
$seq_in = Bio::SeqIO->new('-file' => "$file3", '-format' => "fasta");
$seq_out = Bio::SeqIO->new('-file' => ">>$file2", '-format' => "genbank");
while($inseq = $seq_in->next_seq){$seq_out->write_seq($inseq);
  open(OUTFILE, ">>GB\/$file3.gb");

open(FILE, $file2);
@data = <FILE>;
close FILE;

#Add the tRNA annotations to the GenBank flatfile format
foreach $line(@data){
  if($line =~ m/FEATURES/){ print $line; print OUTFILE $line;

      foreach $key(@allKeys){
        if($key == "_"){next;}else{
          @key = split("_", $key); $trna = $order{$key}; $trna =~ s/
//ig; @trna = split('.cm', $trna);

          if($key[0]<$key[1]){print OUTFILE "      tRNA
$key[0]..$key[1]\n          \/gene=\"$trna[0]\"\\n\";}

          if($key[1]<$key[0]){print OUTFILE "      tRNA      complement\"($key[1]..$key[0])\\n
\/gene=\"$trna[0]\"\\n\";}

        }}else{print OUTFILE $line;}}

close OUTFILE;

#

system("dos2unix -u GB\/$file3.gb -o");

      %order = ();
      $counter = 0;
}

```

Script 4. `add_start_stop.pl` – add a 1bp tRNA annotation to the start and end of each GenBank formatted sequence.

```
@files = glob("*.gb");
foreach $file(@files){
open(FILE, "$file");
@data = <FILE>;
close FILE;

$file =~ s/ gb/genbank/ig;
open(OUTFILE, ">$file");

foreach $line(@data){

if($line =~ m/LOCUS/){
print OUTFILE $line;
@line = split(" ", $line);
$end = ($line[2]);}

elseif($line =~ m/FEATURES/){
print OUTFILE $line;

print OUTFILE "      tRNA          1..1\n          tRNA          $end..$end\n          \\/gene=\"startmarker\"\\n";

$end = 0;}

else{print OUTFILE $line;}

}
}
```