

Supplementary Material – Systematic Biology

Yang Liu, Cymon J. Cox, Wei Wang and Bernard Goffinet – Mitochondrial Phylogenomics of Early Land Plants: Mitigating the Effects of Saturation, Compositional Heterogeneity, and Codon-usage Bias

Table S1. List of 60 taxa sampled for the mitochondrial genomic dataset in this study.

Table S2. Characteristics of 41 mitochondrial genes.

Table S3. Characteristics of data matrices and statistics of the best-scoring ML trees inferred from each partitioning strategy.

Table S4. Average GC percentage of each plant group, and difference significance test among groups.

Figure S1. The average pairwise distances of the 41 mitochondrial genes.

Figure S2. A summary of single gene ML tree topologies.

Figure S3-S19. Supplementary trees.

Note: MCMC chain convergence diagnostics for supplementary trees

Various chain convergence diagnostic measures were evaluated. These statistics fall into two categories, namely, 1) those evaluating the effectiveness/quality of a particular analytical run, and 2), those measuring the convergence of two independent analytical runs. Where possible we have tried to follow the recommendations (when given) of the software authors as to the merit of these diagnostics. In the first category all three software used, MrBayes, P4, and PhyloBayes-MPI, report the Effective Sample Size (ESS) values for the sampled parameters. Here we evaluate and report the following ESS values: all values above 300 “excellent”, one value $< 300 > 100$ “good”, one value $< 100 > 50$ “poor”, one value < 50 “bad”. Note that in models with many free parameters, all but one value could be in the 1000's while only one is < 100 and the convergence diagnostic would still be evaluated as “poor”. MrBayes and P4 report free parameter proposal acceptance rates which we consider “good” if $> 10\%$ and $< 40\%$, otherwise “poor”. MrBayes and P4 implement Metropolis-coupled MCMC chains and report the “mixing” of those chains (i.e. numbers of accepted chain swap proposals between the heated and cold chains) which we evaluate as “good” if some swapping occurs between all chain pairs, else it is evaluated as “bad”. MrBayes also reports the Potential Scale Reduction Factor (PSRF) of sampled parameters which approaches 1 when sampling from the posterior density – here we report the mean and standard deviation of the average PSRF value averaged across duplicate chains. By default MrBayes (ver. 3.2.1) implements duplicate parallel MCMC runs each with Metropolis-coupling, whereas P4 and PhyloBayes only implement a single analytical chain. When possible, duplicate analytical runs were performed with both P4 and PhyloBayes. Convergence between duplicate MrBayes and P4 runs with respect to tree topology is measured by an average standard deviation of split support (ASDOSS). An ASDOSS value of ≤ 0.01 was considered sufficiently low to signify convergence between separate runs and the results were combined. Similarly, PhyloBayes reports a maximum difference between similar splits (“maxdiff”) and a mean difference (“meandiff”). A “maxdiff” of < 0.3 was considered acceptable and independent PhyloBayes runs were combined. When relevant, these diagnostics are reported in the legends of the supplemental tree figures.

Table S1. List of 60 taxa sampled for the mitochondrial genomic dataset in this study. Newly sequenced taxa are highlighted in gray. Voucher information and GenBank accession numbers for the newly sampled taxa are listed. The average NGS sequencing coverage was estimated for each taxon based on the assembled mitochondrial genomes or contigs (bp = base pair).

Group	Taxon	Voucher / reference	NGS average sequencing depth	Size of mt genome or contigs (bp)	No. of mt genes sampled (41 total)	GenBank accession number (s)
Algae	<i>Chaetosphaeridium globosum</i>	Turmel, et al. 2002a	-	56,574	32	NC_004118
	<i>Chara vulgaris</i>	Turmel, et al. 2003	-	67,737	35	NC_005255
	<i>Chlorokybus atmophyticus</i>	Turmel, et al. 2007	-	201,763	31	NC_009630
	<i>Mesostigma viride</i>	Turmel, et al. 2002b	-	42,424	28	NC_008240
	<i>Nitella hyalina</i>	No	-	80,193	35	NC_017598
Liverwort	<i>Bazzania trilobata</i>	Goffinet 10577	70×	159,033	37	KC662627, KC662647, KC662667, KC662687, KC662707, KC662727, KC662747, KC662767, KC662786, KC662819, KC662839, KC662859, KC662879, KC662899, KC662937, KC662957, KC662977, KC662997, KC663017, KC663037, KC663057, KC663077, KC663097, KC663117, KC663136, KC663156, KC663194, KC663214, KC663234, KC663254, KC663274, KC663294, KC663314, KC663334, KC663352, KC663371, KC663391
	<i>Blasia pusilla</i>	Goffinet et al. 9754	16×	186,178	37	KC662628, KC662648, KC662668, KC662688, KC662708, KC662728, KC662748, KC662768, KC662787, KC662820, KC662840, KC662860, KC662880, KC662900, KC662919, KC662938, KC662958, KC662978, KC662998, KC663018, KC663038, KC663058, KC663078, KC663098, KC663137, KC663157, KC663176, KC663195, KC663215, KC663235, KC663255, KC663275, KC663295, KC663315, KC663335, KC663372, KC663392
	<i>Marchantia polymorpha</i>	Oda et al., 1992	-	186,609	39	NC_001660
	<i>Pleurozia purpurea</i>	Wang et al., 2009	-	168,526	39	NC_013444
	<i>Scapania nemorea</i>	Goffinet 10589	35×	139,644	37	KC662637, KC662657, KC662677, KC662697, KC662717, KC662737, KC662757, KC662776, KC662796, KC662829, KC662849, KC662869, KC662889, KC662909, KC662947, KC662967, KC662987, KC663007, KC663027, KC663047, KC663067, KC663087, KC663107, KC663126, KC663146, KC663166, KC663204, KC663224, KC663244, KC663264, KC663284, KC663304, KC663324, KC663344, KC663361, KC663381, KC663400
	<i>Treubia lacunosa</i>	Liu et al., 2011	-	151,983	36	NC_016122
	<i>Andreaea rothii</i>	Goffinet 10586	15×	111,911	39	KC662623, KC662643, KC662663, KC662683, KC662703, KC662723, KC662743, KC662763, KC662782, KC662802, KC662815, KC662835, KC662855, KC662875, KC662895, KC662915,
Moss						

<i>Anomodon attenuatus</i>	Goffinet 10590	78×	104,252	40	KC662933, KC662953, KC662973, KC662993, KC663013, KC663033, KC663053, KC663073, KC663093, KC663113, KC663132, KC663152, KC663172, KC663190, KC663210, KC663230, KC663250, KC663270, KC663290, KC663310, KC663330, KC663367, KC663387, KC662624, KC662644, KC662664, KC662684, KC662704, KC662724, KC662744, KC662764, KC662783, KC662803, KC662816, KC662836, KC662856, KC662876, KC662896, KC662916, KC662934, KC662954, KC662974, KC662994, KC663014, KC663034, KC663054, KC663074, KC663094, KC663114, KC663133, KC663153, KC663173, KC663191, KC663211, KC663231, KC663251, KC663271, KC663291, KC663311, KC663331, KC663349, KC663368, KC663388
<i>Anomodon rugelii</i>	Liu et al., 2011	-	104,239	40	NC_016121
<i>Atrichum angustatum</i>	Goffinet 10582	220×	115,146	40	KC662625, KC662645, KC662665, KC662685, KC662705, KC662725, KC662745, KC662765, KC662784, KC662804, KC662817, KC662837, KC662857, KC662877, KC662897, KC662917, KC662935, KC662955, KC662975, KC662995, KC663015, KC663035, KC663055, KC663075, KC663095, KC663115, KC663134, KC663154, KC663174, KC663192, KC663212, KC663232, KC663252, KC663272, KC663292, KC663312, KC663332, KC663350, KC663369, KC663389, KC662626, KC662646, KC662666, KC662686, KC662706, KC662726, KC662746, KC662766, KC662785, KC662805, KC662818, KC662838, KC662858, KC662878, KC662898, KC662918, KC662936, KC662956, KC662976, KC662996, KC663016, KC663036, KC663056, KC663076, KC663096, KC663116, KC663135, KC663155, KC663175, KC663193, KC663213, KC663233, KC663253, KC663273, KC663293, KC663313, KC663333, KC663351, KC663370, KC663390, KC662629, KC662649, KC662669, KC662689, KC662709, KC662729, KC662749, KC662769, KC662788, KC662821, KC662841, KC662861, KC662881, KC662901, KC662920, KC662939, KC662959, KC662979, KC662999, KC663019, KC663039, KC663059, KC663079, KC663099, KC663118, KC663138, KC663158, KC663177, KC663196, KC663216, KC663236, KC663256, KC663276, KC663296, KC663316, KC663336, KC663353, KC663373, KC663393, KC662630, KC662650, KC662670, KC662690, KC662710, KC662730, KC662750, KC662770, KC662789, KC662822, KC662842, KC662862, KC662882, KC662902, KC662921, KC662940, KC662960, KC662980, KC663000, KC663020, KC663040, KC663060, KC663080, KC663100, KC663119, KC663139, KC663159, KC663178, KC663197, KC663217, KC663237, KC663257, KC663277, KC663297, KC663317, KC663337, KC663354, KC663374, KC663394, KC662631, KC662651, KC662671, KC662691, KC662711, KC662731, KC662751, KC662771, KC662790, KC662806, KC662823, KC662843, KC662863, KC662883, KC662903, KC662922,
<i>Bartramia pomiformis</i>	Goffinet 10587	58×	106,198	40	
<i>Bryum argenteum</i>	Jairo Patiño s.n. 2009.11.4	113×	101,277	39	
<i>Buxbaumia aphylla</i>	Goffinet s.n. 2012.3.5	123×	100,725	39	
<i>Climacium americanum</i>	Goffinet 10578	266×	105,048	40	

<i>Dicranum scoparium</i>	Goffinet 10584	106×	106,779	38	KC662941, KC662961, KC662981, KC663001, KC663021, KC663041, KC663061, KC663081, KC663101, KC663120, KC663140, KC663160, KC663179, KC663198, KC663218, KC663238, KC663258, KC663278, KC663298, KC663318, KC663338, KC663355, KC663375, KC663395, KC662632, KC662652, KC662672, KC662692, KC662712, KC662732, KC662752, KC662791, KC662824, KC662844, KC662864, KC662884, KC662904, KC662923, KC662942, KC662962, KC662982, KC663002, KC663022, KC663042, KC663062, KC663082, KC663102, KC663121, KC663141, KC663161, KC663180, KC663199, KC663219, KC663239, KC663259, KC663279, KC663299, KC663319, KC663339, KC663356, KC663376, KC663396
<i>Funaria hygrometrica</i>	Goffinet 5576	1,404×	109,586	40	KC662633, KC662653, KC662673, KC662693, KC662713, KC662733, KC662753, KC662772, KC662792, KC662807, KC662825, KC662845, KC662865, KC662885, KC662905, KC662924, KC662943, KC662963, KC662983, KC663003, KC663023, KC663043, KC663063, KC663083, KC663103, KC663122, KC663142, KC663162, KC663181, KC663200, KC663220, KC663240, KC663260, KC663280, KC663300, KC663320, KC663340, KC663357, KC663377, KC663397, KC662634, KC662654, KC662674, KC662694, KC662714, KC662734, KC662754, KC662773, KC662793, KC662808, KC662826, KC662846, KC662866, KC662886, KC662906, KC662925, KC662944, KC662964, KC662984, KC663004, KC663024, KC663044, KC663064, KC663084, KC663104, KC663123, KC663143, KC663163, KC663182, KC663201, KC663221, KC663241, KC663261, KC663281, KC663301, KC663321, KC663341, KC663358, KC663378, KC663398, KC662635, KC662655, KC662675, KC662695, KC662715, KC662735, KC662755, KC662774, KC662794, KC662809, KC662827, KC662847, KC662867, KC662887, KC662907, KC662926, KC662945, KC662965, KC662985, KC663005, KC663025, KC663045, KC663065, KC663085, KC663105, KC663124, KC663144, KC663164, KC663183, KC663202, KC663222, KC663242, KC663262, KC663282, KC663302, KC663322, KC663342, KC663359, KC663379, KC663399
<i>Hypnum imponens</i>	Goffinet 10576	68×	103,830	40	
<i>Orthotrichum stellatum</i>	Goffinet 10579	319×	104,131	40	
<i>Physcomitrella patens</i>	Terasawa et al., 2007	-	105,340	40	NC_007945
<i>Ptychomnion cygnisetum</i>	Buck 58936	200×	104,480	39	KC662636, KC662656, KC662676, KC662696, KC662716, KC662736, KC662756, KC662775, KC662795, KC662810, KC662828, KC662848, KC662868, KC662888, KC662908, KC662927, KC662946, KC662966, KC662986, KC663006, KC663026, KC663046, KC663066, KC663086, KC663106, KC663125, KC663145, KC663165, KC663184, KC663203, KC663223, KC663243, KC663263, KC663283, KC663303, KC663323, KC663343, KC663360, KC663380, KC662638, KC662658, KC662678, KC662698, KC662718, KC662738, KC662758, KC662777, KC662797, KC662811, KC662830, KC662850, KC662870, KC662890, KC662910, KC662928,
<i>Sphagnum girgensohnii</i>	Goffinet 10588	124×	141,137	40	

	<i>Sphagnum palustre</i>	Goffinet 10575	238×	141,276	40	KC662948, KC662968, KC662988, KC663008, KC663028, KC663048, KC663068, KC663088, KC663108, KC663127, KC663147, KC663167, KC663185, KC663205, KC663225, KC663245, KC663265, KC663285, KC663305, KC663325, KC663345, KC663362, KC663382, KC663401, KC662639, KC662659, KC662679, KC662699, KC662719, KC662739, KC662759, KC662778, KC662798, KC662812, KC662831, KC662851, KC662871, KC662891, KC662911, KC662929, KC662949, KC662969, KC662989, KC663009, KC663029, KC663049, KC663069, KC663089, KC663109, KC663128, KC663148, KC663168, KC663186, KC663206, KC663226, KC663246, KC663266, KC663286, KC663306, KC663326, KC663346, KC663363, KC663383, KC663402, KC662640, KC662660, KC662680, KC662700, KC662720, KC662740, KC662760, KC662779, KC662799, KC662832, KC662852, KC662872, KC662892, KC662912, KC662930, KC662950, KC662970, KC662990, KC663010, KC663030, KC663050, KC663070, KC663090, KC663110, KC663129, KC663149, KC663169, KC663187, KC663207, KC663227, KC663247, KC663267, KC663287, KC663307, KC663327, KC663347, KC663364, KC663384, KC663403, KC662641, KC662661, KC662681, KC662701, KC662721, KC662741, KC662761, KC662780, KC662800, KC662813, KC662833, KC662853, KC662873, KC662893, KC662913, KC662931, KC662951, KC662971, KC662991, KC663011, KC663031, KC663051, KC663071, KC663091, KC663111, KC663130, KC663150, KC663170, KC663188, KC663208, KC663228, KC663248, KC663268, KC663288, KC663308, KC663328, KC663365, KC663385, KC663404, KC662642, KC662662, KC662682, KC662702, KC662722, KC662742, KC662762, KC662781, KC662801, KC662814, KC662834, KC662854, KC662874, KC662894, KC662914, KC662932, KC662952, KC662972, KC662992, KC663012, KC663032, KC663052, KC663072, KC663092, KC663112, KC663131, KC663151, KC663171, KC663189, KC663209, KC663229, KC663249, KC663269, KC663289, KC663309, KC663329, KC663348, KC663366, KC663386, KC663405
	<i>Tetraphis pellucida</i>	Goffinet s.n. 2012.2.3	968×	107,736	39	
	<i>Tetraplodon fuegianus</i>	Lewis 998	197×	104,019	39	
	<i>Ulota hutchinsiae</i>	Goffinet 10580	95×	104,608	40	
Hornwort	<i>Megaceros aenigmaticus</i>	Li et al., 2009	-	184,908	20	NC_012651
	<i>Phaeoceros laevis</i>	Xue et al., 2010	-	209,482	20	NC_013765
Lycophyte	<i>Huperzia squarrosa</i>	Liu et al., 2012	-	413,530	33	NC_017755
	<i>Isoetes engelmannii</i>	Grewe et al., 2009	-	186,743	22	FJ010859, FJ176330, FJ390841, FJ536259, FJ628360
Gymnosperm	<i>Cycas taitungensis</i>	Chaw et al., 2008	-	414,903	36	NC_010303
Angiosperm	<i>Arabidopsis thaliana</i>	Unseld et al., 1997	-	366,924	29	NC_001284
	<i>Beta vulgaris</i>	Kubo et al., 2000	-	368,801	28	NC_002511
	<i>Boea hygrometrica</i>	Zhang et al., 2011	-	510,519	27	NC_016741

<i>Brassica napus</i>	Handa, 2003	-	221,853	29	NC_008285
<i>Carica papaya</i>	Magee et al., 2010	-	476,890	37	NC_012116
<i>Citrullus lanatus</i>	Alverson et al., 2010	-	379,236	36	NC_014043
<i>Cucurbita pepo</i>	Alverson et al., 2010	-	982,833	36	NC_014050
<i>Daucus carota</i>	Iorizzo et al., 2012	-	281,132	28	NC_017855
<i>Lotus japonicus</i>	Kazakoff et al., 2012	-	380,861	30	NC_016743
<i>Malus x domestica</i>	Goremykin et al., 2012	-	396,947	30	NC_018554
<i>Millettia pinnata</i>	Kazakoff et al., 2012	-	425,718	30	NC_016742
<i>Mimulus guttatus</i>	Mower et al., 2012	-	525,671	32	NC_018041
<i>Nicotiana tabacum</i>	Sugiyama et al., 2005	-	430,597	33	NC_006581
<i>Oryza sativa</i>	Notsu et al., 2002	-	490,520	29	NC_011033
<i>Phoenix dactylifera</i>	Fang et al., 2012	-	715,001	36	NC_016740
<i>Raphanus sativus</i>	Tanaka et al., 2012	-	258,426	30	NC_018551
<i>Ricinus communis</i>	Rivarola et al., 2011	-	502,773	35	NC_015141
<i>Silene latifolia</i>	Sloan et al., 2010	-	253,413	25	NC_014487
<i>Sorghum bicolor</i>	No	-	468,628	31	NC_008360
<i>Spirodela polyrhiza</i>	Wang et al., 2012	-	228,493	34	NC_017840
<i>Tripsacum dactyloides</i>	No	-	704,100	31	NC_008362
<i>Triticum aestivum</i>	Ogihara et al., 2005	-	452,528	31	NC_007579
<i>Vigna radiate</i>	Alverson et al., 2011	-	401,262	29	NC_015121
<i>Vitis vinifera</i>	Goremykin et al., 2009	-	773,279	36	NC_012119
<i>Zea mays</i>	Clifton et al., 2004	-	569,630	31	NC_007982

Table S2. Data characteristics of 41 mitochondrial genes, including the number of taxa sampled per gene in the data matrix, the number of total aligned characters per gene, and the percentage of gaps or missing data per gene.

Gene	No. of taxa (60 total)	Total aligned characters	Missing %
<i>atp1</i>	59	1,518	1.9%
<i>atp4</i>	57	522	6.1%
<i>atp6</i>	60	744	0.4%
<i>atp8</i>	58	351	5.4%
<i>atp9</i>	60	222	0.0%
<i>ccmB</i>	50	609	21.9%
<i>ccmC</i>	50	690	16.8%
<i>ccmFC</i>	48	1,239	21.3%
<i>ccmFN</i>	47	1,413	24.4%
<i>cob</i>	58	1,143	3.5%
<i>cox1</i>	59	1,557	2.0%
<i>cox2</i>	58	726	3.7%
<i>cox3</i>	60	783	0.0%
<i>matR</i>	27	1,929	56.4%
<i>nad1</i>	59	969	1.8%
<i>nad2</i>	60	1,446	1.1%
<i>nad3</i>	60	351	0.4%
<i>nad4</i>	60	1,482	0.1%
<i>nad4L</i>	60	300	0.2%
<i>nad5</i>	60	1,983	0.2%
<i>nad6</i>	60	558	0.4%
<i>nad7</i>	47	1,176	21.8%
<i>nad9</i>	60	564	0.6%
<i>rpl2</i>	36	1,260	47.4%
<i>rpl5</i>	53	495	13.5%
<i>rpl6</i>	31	291	48.4%
<i>rpl10</i>	43	411	30.3%
<i>rpl16</i>	50	405	16.8%
<i>rps1</i>	46	585	34.6%
<i>rps2</i>	38	624	39.5%
<i>rps3</i>	53	1,131	15.8%
<i>rps4</i>	54	513	11.6%
<i>rps7</i>	46	534	31.4%
<i>rps11</i>	31	330	48.4%
<i>rps12</i>	55	378	8.8%
<i>rps13</i>	49	354	19.9%
<i>rps14</i>	41	297	31.8%
<i>rps19</i>	40	264	35.5%
<i>sdh3</i>	39	393	37.6%
<i>sdh4</i>	41	231	32.0%
<i>tatC</i>	56	696	7.4%

Table S3. Characteristics of data matrices and statistics of the best-scoring ML trees inferred from each partitioning strategy: single (not partitioning); 12-3 (all 1st and 2nd codon as one partition; 3rd codon as one partition); 1-2-3 (each 1st, 2nd and 3rd codon as one partition); gene (partitioned by gene); gene-12-3 (1st and 2nd codon in each gene as one partition, and 3rd in each gene as one partition); gene-1-2-3 (each of the three codon positions in each gene as one partition). The optimal partitioning strategy of each dataset is highlighted in bold.

Matrix	Taxa/ characters	Partitioning strategy	No. of partition	Log-likelihood	Parameters	AIC	Δ AIC	BIC	Δ BIC
nt	60/31,467	single	1	-307,997	10	616,013	13,423	616,011	13,603
		12-3	2	-305,069	20	610,178	7,588	610,174	7,765
		1-2-3	3	-304,539	30	609,137	6,547	609,131	6,722
		gene	41	-305,303	410	611,427	8,837	611,336	8,928
		gene-12-3	82	-300,475	820	602,590	0	602,408	0
		gene-1-2-3	123	-301,454	1230	605,368	2,778	605,095	2,687
Degenerated nt	60/31,467	single	1	-163,305	10	326,630	326,627	5,336	5,606
		12-3	2	-163,096	20	326,232	326,227	4,938	5,206
		1-2-3	3	-162,540	30	325,139	325,133	3,845	4,112
		gene	41	-160,688	410	322,196	322,105	902	1,084
		gene-12-3	82	-160,358	820	322,356	322,174	1,062	1,153
		gene-1-2-3	123	-159,417	1230	321,294	321,021	0	0

Note: AIC= $-2\log\text{-likelihood} + 2K$; BIC= $-2\log\text{-likelihood} + K \log(n)$; n is the taxon number; K is the number of free parameters.

Table S4. Average GC percentage of mt genes in each plant group. The GC percentages were pairwise compared among groups, the significance level were achieved by one-tailed student T-test. Abbreviations: L, liverworts; M, mosses; H, hornworts; V, vascular plants. * significant and ** extremely significant. Note, hornwort have a small sample size (2 taxa only) which may cause the T-test is inaccurate when hornwort is involved.

	All codons					1 st codons					2 nd codons					3 rd codons				
	GC%	A	L	M	H	GC%	A	L	M	H	GC%	A	L	M	H	GC%	A	L	M	H
Algae	33.64					43.19					36.99					20.75				
Liverworts	37.25	0.00**				42.80	0.24				38.16	0.03*				30.79	0.00**			
Mosses	35.88	0.01**	0.03*			42.30	0.01**	0.17			37.82	0.03*	0.23			27.52	0.01**	0.01**		
Hornworts	37.43	0.10	0.46	0.24		43.58	0.43	0.37	0.30		39.32	0.21	0.32	0.29		29.40	0.00**	0.14	0.09	
Vascular plants	42.70	0.00**	0.00**	0.00**	0.08	48.20	0.00**	0.00**	0.00**	0.11	42.96	0.00**	0.00**	0.00**	0.15	36.93	0.00**	0.00**	0.00**	0.02*

References

- Alverson A.J., Wei X., Rice D.W., Stern D.B., Barry K., Palmer J.D. 2010. Insights into the evolution of mitochondrial genome size from complete sequences of *Citrullus lanatus* and *Cucurbita pepo* (Cucurbitaceae). *Molecular Biology and Evolution* 27:1436-1448.
- Alverson A.J., Zhuo S., Rice D.W., Sloan D.B., Palmer J.D. 2011. The mitochondrial genome of the legume *Vigna radiata* and the analysis of recombination across short mitochondrial repeats. *PLoS ONE* 6:e16404.
- Chaw S.M., Chun-Chieh Shih A., Wang D., Wu Y.W., Liu S.M. 2008. The mitochondrial genome of the gymnosperm *Cycas taitungensis* contains a novel family of short interspersed elements, Bpu sequences, and abundant RNA editing sites. *Molecular Biology and Evolution* 25:603-615.
- Clifton S.W., Minx P., Fauron C.M.-R., Gibson M., Allen J.O., Sun H., Thompson M., Barbazuk W.B., Kanuganti S., Tayloe C. 2004. Sequence and comparative analysis of the maize NB mitochondrial genome. *Plant Physiology* 136:3486-3503.
- Fang Y., Wu H., Zhang T., Yang M., Yin Y., Pan L., Yu X., Zhang X., Hu S., Al-Mssallem I.S. 2012. A complete sequence and transcriptomic analyses of date palm (*Phoenix dactylifera* L.) mitochondrial genome. *PloS ONE* 7:e37164.
- Goremykin V.V., Salamini F., Velasco R., Viola R. 2009. Mitochondrial DNA of *Vitis vinifera* and the issue of rampant horizontal gene transfer. *Molecular Biology and Evolution* 26:99-110.
- Goremykin V.V., Lockhart P.J., Viola R., Velasco R. 2012. The mitochondrial genome of *Malus domestica* and the import-driven hypothesis of mitochondrial genome expansion in seed plants. *The Plant Journal*.
- Grewe F., Viehoveer P., Weisshaar B., Knoop V. 2009. A trans-splicing group I intron and tRNA-hyperediting in the mitochondrial genome of the lycophyte *Isoetes engelmannii*. *Nucleic Acids Research* 37:5093-5104.
- Handa H. 2003. The complete nucleotide sequence and RNA editing content of the mitochondrial genome of rapeseed (*Brassica napus* L.): comparative analysis of the mitochondrial genomes of rapeseed and *Arabidopsis thaliana*. *Nucleic Acids Research* 31:5907-5916.
- Iorizzo M., Senalik D., Szklarczyk M., Grzebelus D., Spooner D., Simon P. 2012. De novo assembly of the carrot mitochondrial genome using next generation sequencing of whole genomic DNA provides first evidence of DNA transfer into an angiosperm plastid genome. *BMC Plant Biology* 12:61.
- Kazakoff S.H., Imelfort M., Edwards D., Koehorst J., Biswas B., Batley J., Scott P.T., Gresshoff P.M. 2012. Capturing the Biofuel Wellhead and Powerhouse: The Chloroplast and Mitochondrial Genomes of the Leguminous Feedstock Tree *Pongamia pinnata*. *PloS ONE* 7:e51687.
- Kubo T., Nishizawa S., Sugawara A., Itchoda N., Estiati A., Mikami T. 2000. The complete nucleotide sequence of the mitochondrial genome of sugar beet (*Beta*

vulgaris L.) reveals a novel gene for tRNACys (GCA). Nucleic Acids Research 28:2571-2576.

- Li L., Wang B., Liu Y., Qiu Y.L. 2009. The complete mitochondrial genome sequence of the hornwort *Megaceros aenigmaticus* shows a mixed mode of conservative yet dynamic evolution in early land plant mitochondrial genomes. Journal of Molecular Evolution 68:665-678.
- Liu Y., Xue J.Y., Wang B., Li L., Qiu Y.L. 2011. The mitochondrial genomes of the early land plants *Treubia lacunosa* and *Anomodon rugelii*: dynamic and conservative evolution. PLoS ONE 6:e25836.
- Liu Y., Wang B., Cui P., Li L., Xue J.Y., Yu J., Qiu Y.L. 2012. The mitochondrial genome of the lycophyte *Huperzia squarrosa*: The most archaic form in vascular plants. PLoS ONE 7:e35168.
- Magee A.M., Aspinall S., Rice D.W., Cusack B.P., Sémon M., Perry A.S., Stefanović S., Milbourne D., Barth S., Palmer J.D. 2010. Localized hypermutation and associated gene losses in legume chloroplast genomes. Genome Research 20:1700-1710.
- Mower J.P., Case A.L., Floro E.R., Willis J.H. 2012. Evidence against equimolarity of large repeat arrangements and a predominant master circle structure of the mitochondrial genome from a monkeyflower (*Mimulus guttatus*) lineage with cryptic CMS. Genome Biology and Evolution 4:670-686.
- Notsu Y., Masood S., Nishikawa T., Kubo N., Akiduki G., Nakazono M., Hirai A., Kadowaki K. 2002. The complete sequence of the rice (*Oryza sativa* L.) mitochondrial genome: frequent DNA sequence acquisition and loss during the evolution of flowering plants. Molecular Genetics and Genomics 268:434-445.
- Oda K., Yamato K., Ohta E., Nakamura Y., Takemura M., Nozato N., Akashi K., Kanegae T., Ogura Y., Kohchi T. 1992. Gene organization deduced from the complete sequence of liverwort *Marchantia polymorpha* mitochondrial DNA. A primitive form of plant mitochondrial genome. Journal of Molecular Biology 223:1.
- Ogihara Y., Yamazaki Y., Murai K., Kanno A., Terachi T., Shiina T., Miyashita N., Nasuda S., Nakamura C., Mori N. 2005. Structural dynamics of cereal mitochondrial genomes as revealed by complete nucleotide sequencing of the wheat mitochondrial genome. Nucleic Acids Research 33:6235-6250.
- Rivarola M., Foster J.T., Chan A.P., Williams A.L., Rice D.W., Liu X., Melake-Berhan A., Creasy H.H., Puiu D., Rosovitz M. 2011. Castor Bean Organelle genome sequencing and worldwide genetic diversity analysis. PLoS ONE 6:e21743.
- Sloan D.B., Alverson A.J., Štorchová H., Palmer J.D., Taylor D.R. 2010. Extensive loss of translational genes in the structurally dynamic mitochondrial genome of the angiosperm *Silene latifolia*. BMC Evolutionary Biology 10:274.
- Sugiyama Y., Watase Y., Nagase M., Makita N., Yagura S., Hirai A., Sugiura M. 2005. The complete nucleotide sequence and multipartite organization of the tobacco mitochondrial genome: comparative analysis of mitochondrial genomes in higher plants. Molecular Genetics and Genomics 272:603-615.

- Tanaka Y., Tsuda M., Yasumoto K., Yamagishi H., Terachi T. 2012. A complete mitochondrial genome sequence of Ogura-type male-sterile cytoplasm and its comparative analysis with that of normal cytoplasm in radish (*Raphanus sativus* L.). *BMC Genomics* 13:352.
- Terasawa K., Odahara M., Kabeya Y., Kikugawa T., Sekine Y., Fujiwara M., Sato N. 2007. The mitochondrial genome of the moss *Physcomitrella patens* sheds new light on mitochondrial evolution in land plants. *Molecular Biology and Evolution* 24:699-709.
- Turmel M., Otis C., Lemieux C. 2002a. The complete mitochondrial DNA sequence of *Mesostigma viride* identifies this green alga as the earliest green plant divergence and predicts a highly compact mitochondrial genome in the ancestor of all green plants. *Molecular Biology and Evolution* 19:24-38.
- Turmel M., Otis C., Lemieux C. 2002b. The chloroplast and mitochondrial genome sequences of the charophyte *Chaetosphaeridium globosum*: insights into the timing of the events that restructured organelle DNAs within the green algal lineage that led to land plants. *Proceedings of the National Academy of Sciences* 99:11275-11280.
- Turmel M., Otis C., Lemieux C. 2003. The mitochondrial genome of *Chara vulgaris*: insights into the mitochondrial DNA architecture of the last common ancestor of green algae and land plants. *The Plant Cell* 15:1888-1903.
- Turmel M., Otis C., Lemieux C. 2007. An unexpectedly large and loosely packed mitochondrial genome in the charophycean green alga *Chlorokybus atmophyticus*. *BMC Genomics* 8:137.
- Unsold M., Marienfeld J.R., Brandt P., Brennicke A. 1997. The mitochondrial genome of *Arabidopsis thaliana* contains 57 genes in 366,924. *Nature Genetics* 15:57.
- Wang B., Xue J., Li L., Liu Y., Qiu Y.L. 2009. The complete mitochondrial genome sequence of the liverwort *Pleurozia purpurea* reveals extremely conservative mitochondrial genome evolution in liverworts. *Current Genetics* 55:601-609.
- Wang W., Wu Y., Messing J. 2012. The mitochondrial genome of an aquatic plant, *Spirodela polyrhiza*. *PloS ONE* 7:e46747.
- Xue J.Y., Liu Y., Li L., Wang B., Qiu Y.L. 2010. The complete mitochondrial genome sequence of the hornwort *Phaeoceros laevis*: retention of many ancient pseudogenes and conservative evolution of mitochondrial genomes in hornworts. *Current Genetics* 56:53-61.
- Zhang T., Zhang X., Hu S., Yu J. 2011. An efficient procedure for plant organellar genome assembly, based on whole genome data from the 454 GS FLX sequencing platform. *Plant Methods* 7:38.

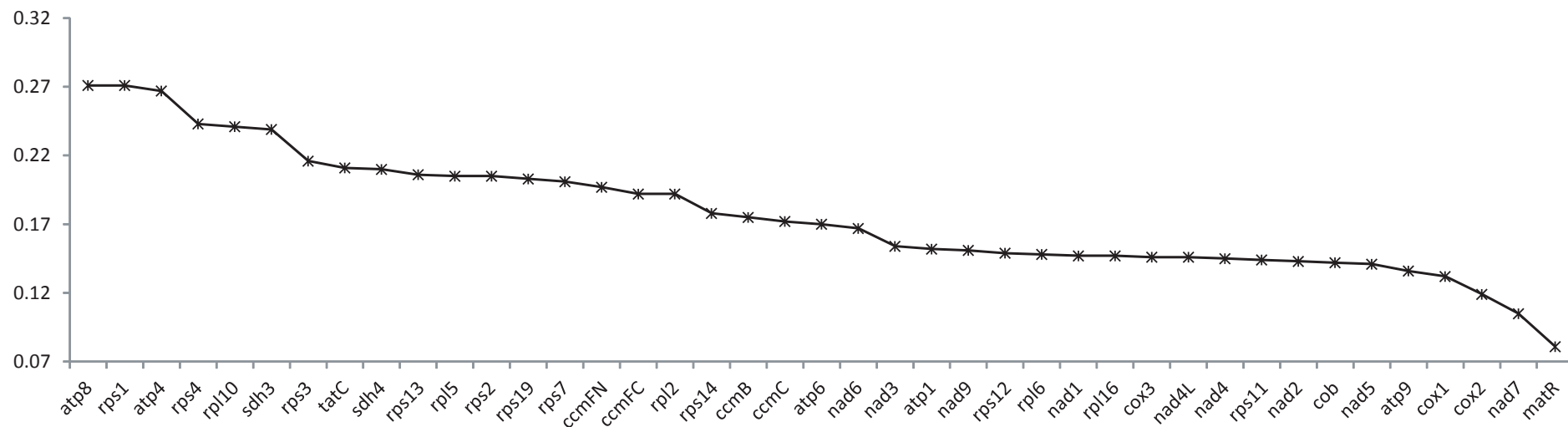


Figure S1 The average pairwise distances of the 41 mitochondrial genes. The genes are shown in the order of highest to lowest average distance.

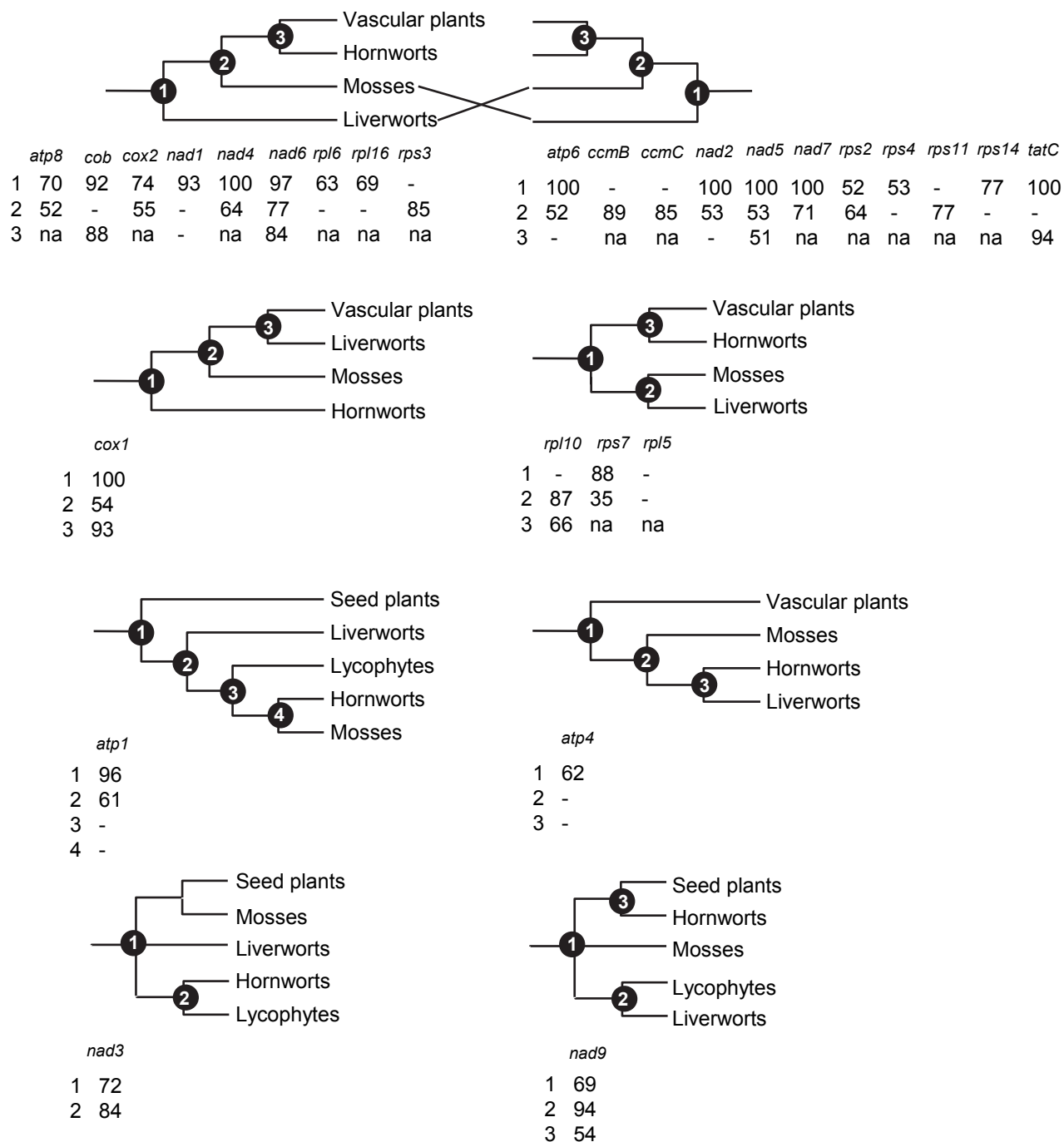


Figure S2 A summary of single gene ML tree topologies. The trees were inferred by RAxML, and bootstrap replicates were performed for 100 times. The bootstrap values of deep nodes of land plants are summarized. “-” indicates bootstrap value < 50. “na” means the node is not available, due to gene lost in a certain whole group. The rest trees are not shown, either because of paraphyletic of one group, or gene lost in outgroups.

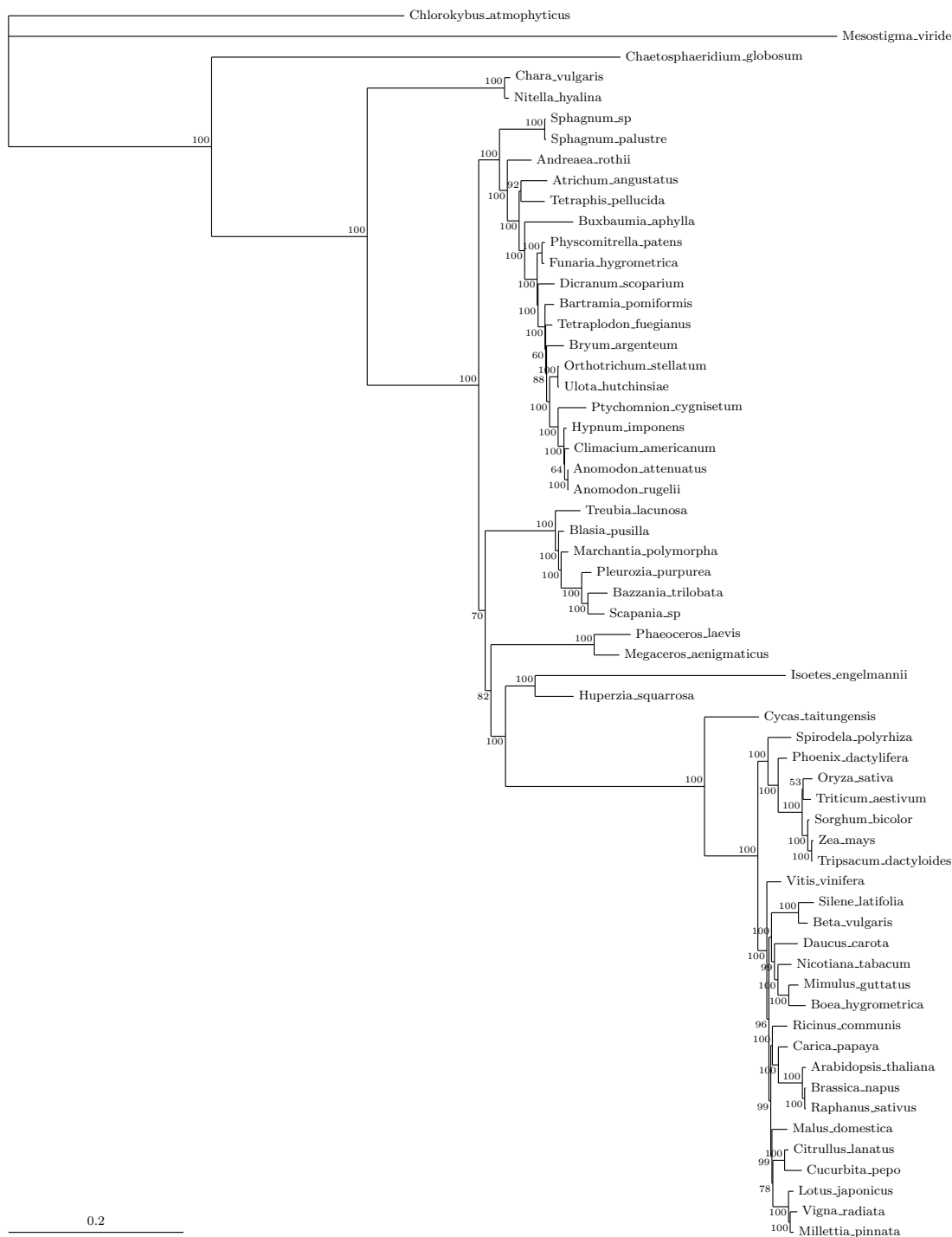


Figure S3 Protein-coding gene data: RAxML ML GTR+I+ Γ ML phylogram inferred from the nt data using RAxML, GTR+I+G model with 100 bootstrap replicates, data partitioned by model gene-12-3 (1st and 2nd codon in each gene as one partition, and 3rd in each gene as one partition). This is the optimal partitioning strategy determined by AIC and BIC (see Table S3).

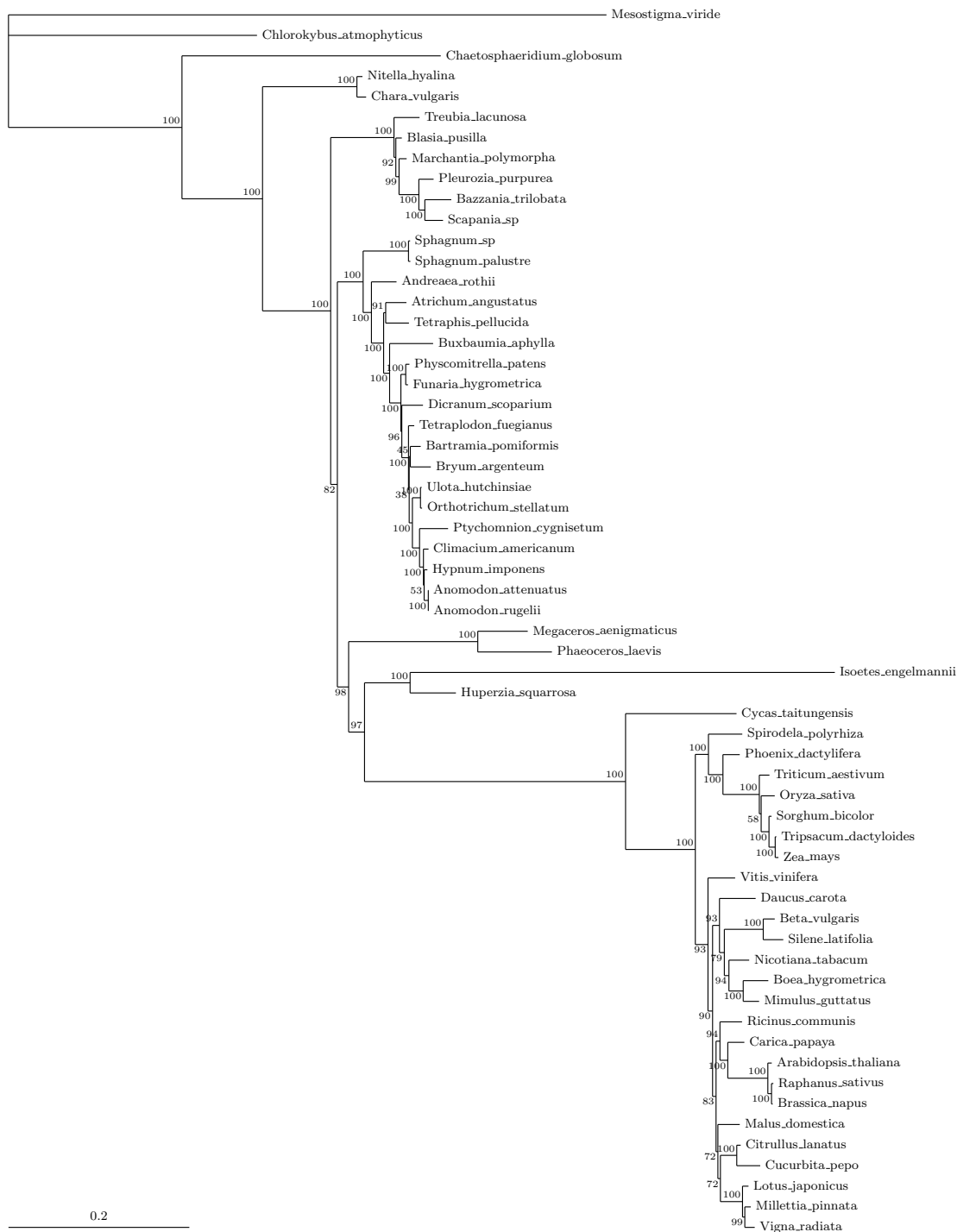


Figure S4 Protein data: RAxML ML stmlREV+Γ ML phylogram inferred from the aa data using RAxML, stmlREV+G model with 100 bootstrap replicates.



Figure S5 Protein-coding gene data: MCMC MrBayes GTR+I+Γ Marginal likelihood: $-\bar{L}_h = 307760.6634$. 5,000,000 generations, 10,000 samples, 4000 discarded as burn-in. Mixing poor; acceptances low; ASDOSS 0.000182; PSRF mean 1.013 stdev 0.0107; ESS bad.

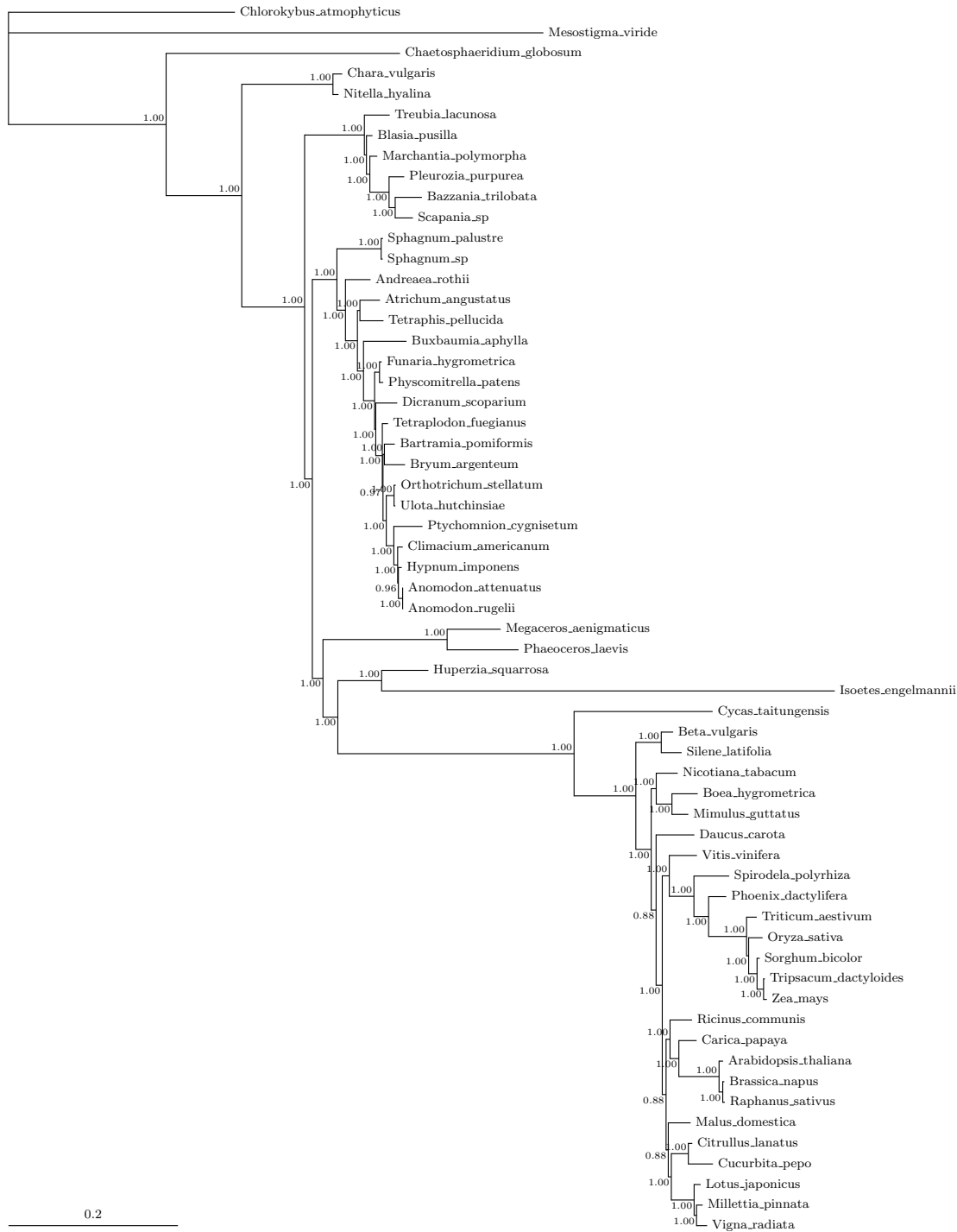


Figure S6 Protein data: MCMC MrBayes JTT+Γ Model composition set to JTT model values. Marginal likelihood: $-\bar{L}_h = 174483.856679$. 1,000,000 generations, 10,000 samples, 4000 discarded as burn-in. Mixing poor; acceptances low; ASDOSS 0.01678; PSRF 1.0034 std dev 0.0034; ESS excellent.

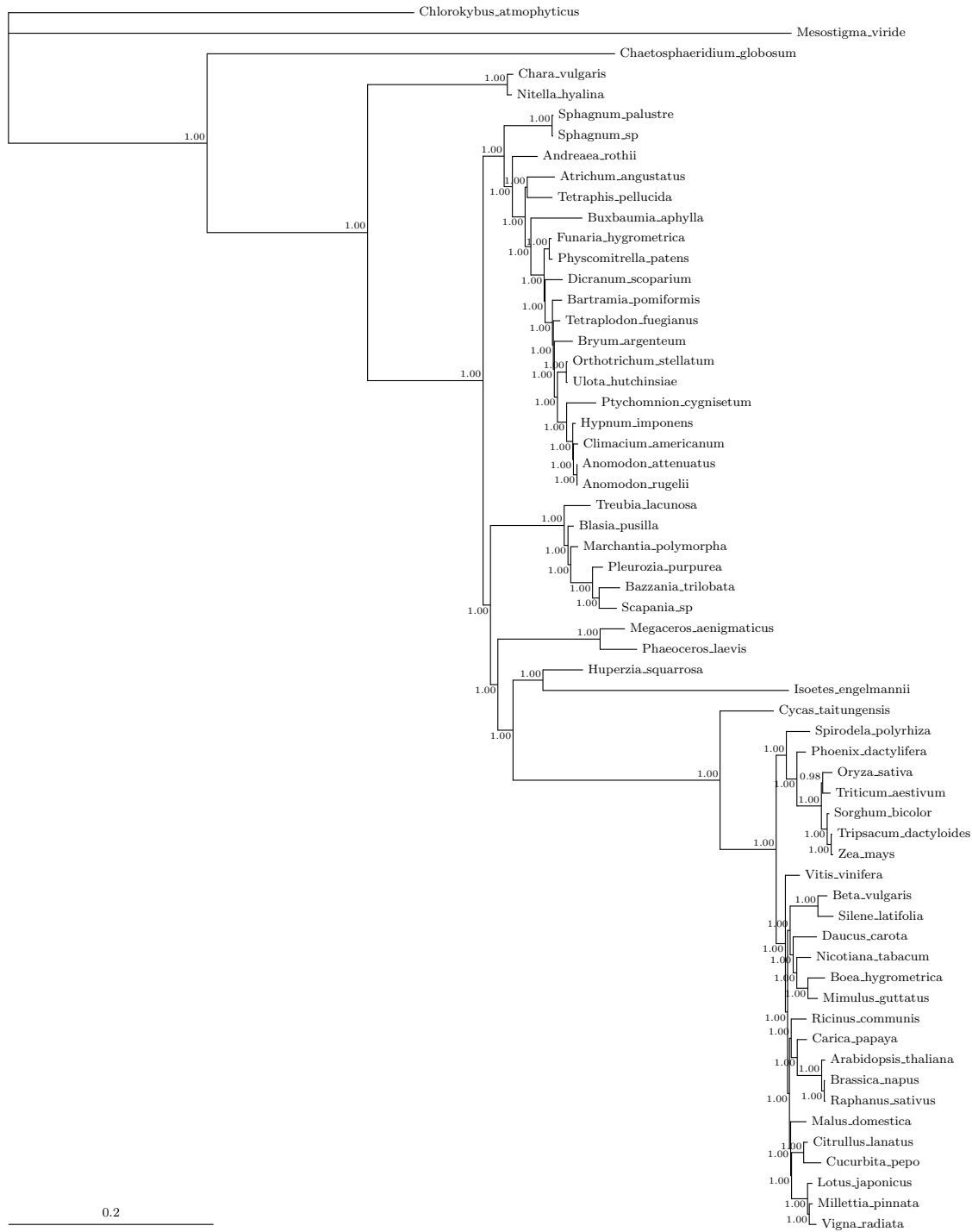


Figure S7 Protein-coding gene data: MCMC P4 GTR+I+Γ+PP Combined marginal likelihood: $-\bar{L}_h = 307761.6732$. ASDOSS 0.0014. Run1: Marginal likelihood: $-\bar{L}_h = 307761.6943$. 3,000,000 generations, 15,000 samples, 10,000 discarded as burn-in. Posterior predictive simulations of X^2 : original statistic = 9598.5261, sample distribution = 17.0272 to 204.0255, $p = 0.0$. Mixing good; acceptances good; ESS excellent. Run2: Marginal likelihood: $-\bar{L}_h = 307761.6522$. 3,000,000 generations, 15,000 samples, 10,000 discarded as burn-in. Posterior predictive simulations of X^2 : original statistic = 9598.5261, sample distribution = 17.4861 to 172.3217, $p = 0.0$. Mixing good; acceptances good; ESS excellent.

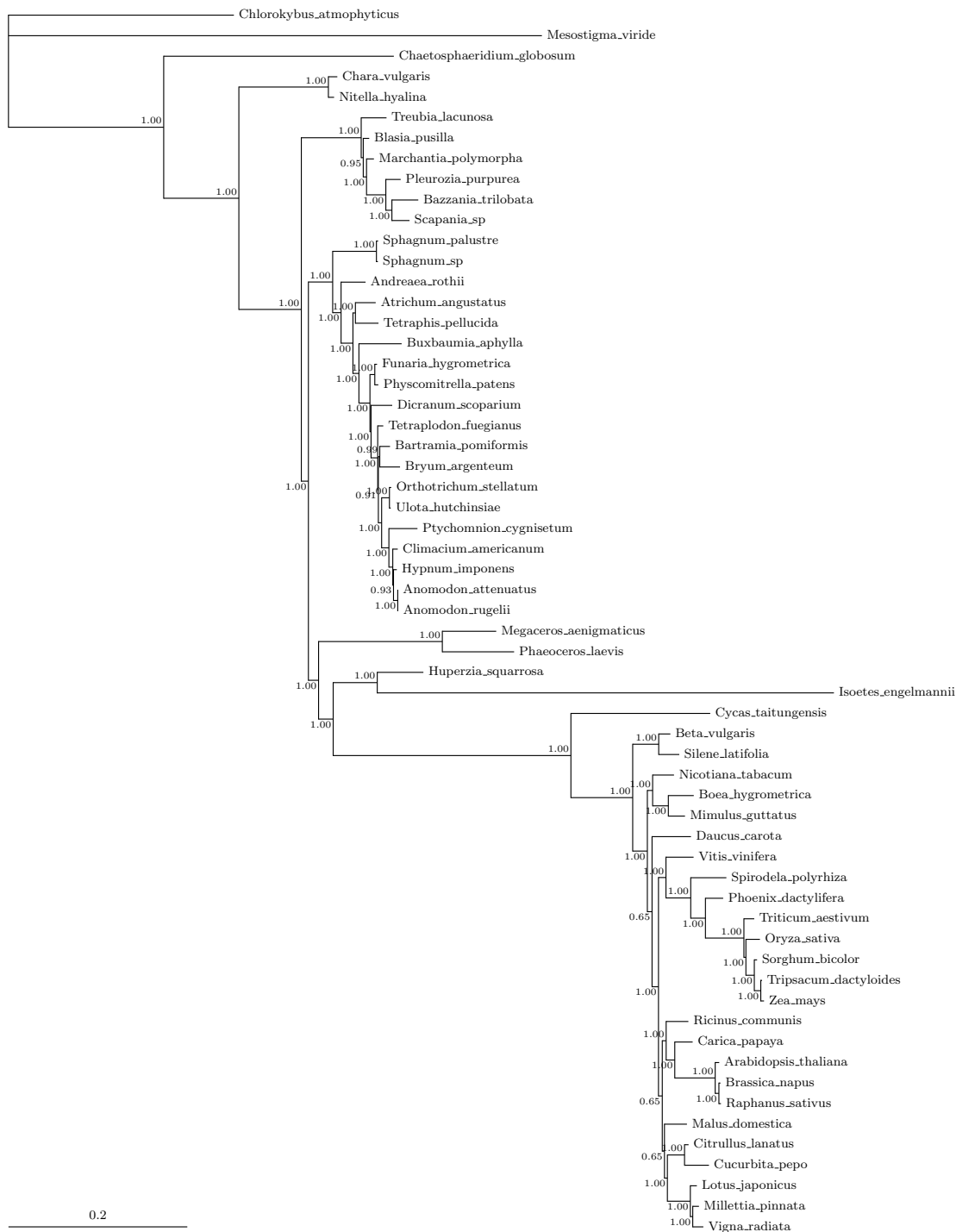


Figure S8 Protein data: MCMC P4 JTT+ Γ +PP Model composition a free parameter. Marginal likelihood: $-\bar{L}_h = 171867.9048$. Posterior predictive simulations of X^2 : original statistic = 5457.1415, sample distribution = 187.8886 to 579.4935, $p = 0.0$. 2,000,000 generations, 20,000 samples, 12,000 discarded as burn-in. Mixing poor; acceptances good; ESS good.

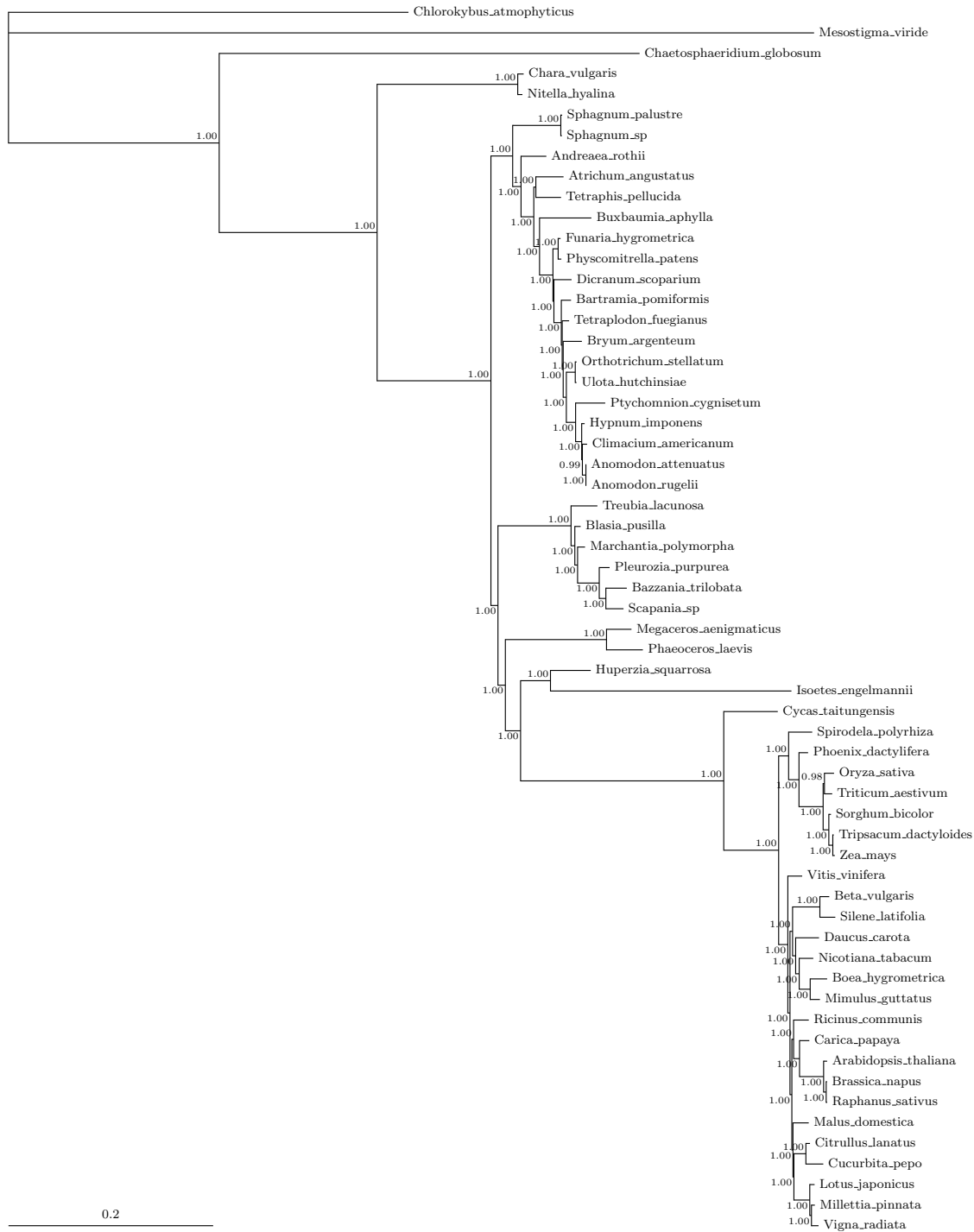


Figure S9 Protein-coding gene data: MCMC MrBayes codon site-specific 3*(GTR+I+Γ). Marginal likelihood: $-\bar{L}_h = 303725.3433$. 5,000,000 generations, 10,000 samples, 5000 discarded as burn-in. Mean partition rates by codon position: first = 0.7785, second = 0.6994, third=1.5220. Mixing poor; acceptances low; ASDOSS 0.000096 ; PSRF mean 1.0051 stdev 0.0065; ESS good.



Figure S10 Protein data: MCMC P4 stmtREV+I+PP Marginal likelihood: $-\bar{L}_h = 166198.2549$. Posterior predictive simulations of X^2 : original statistic = 5457.1415, sample distribution = 195.8218 to 600.1831 p = 0.0. 2,000,000 generations, 20,000 samples, 10,000 discarded as burn-in. Mixing poor; acceptances good; ESS good.

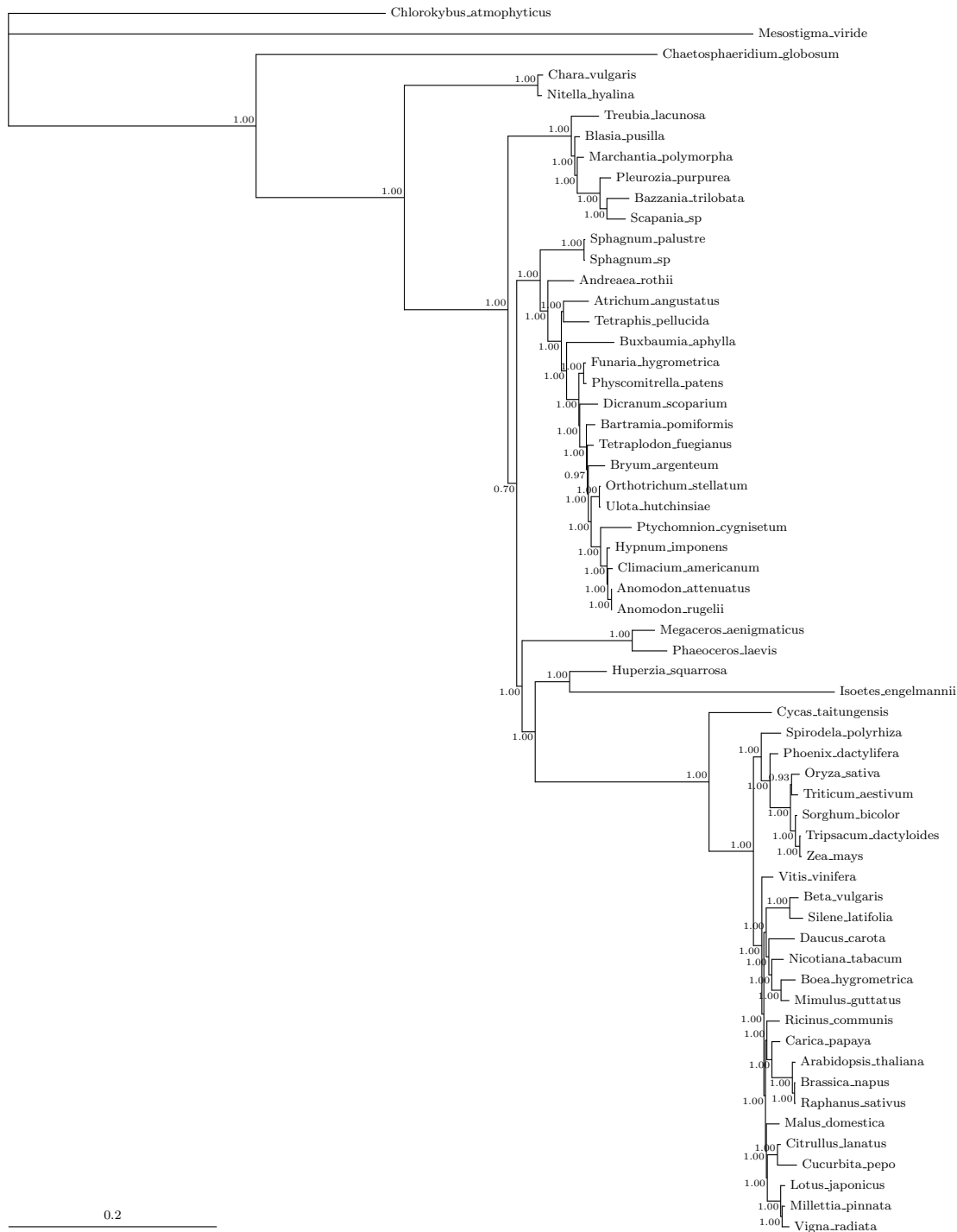


Figure S11 Protein-coding gene data: MCMC P4 GTR+I+Γ+CV2+PP Non-stationary composition MCMC - 2 composition vectors. Combined marginal likelihood: $-\bar{L}_h = 304402.97652$. ASDOSS 0.0100. Run 1: Marginal likelihood: $-\bar{L}_h = 304399.7207$. Posterior predictive simulations of X^2 : original statistic = 9598.5262, sample distribution = 13269.1503 to 18692.2755, $p = 1.0$. 2,000,000 generations, 10,000 samples, 6,000 discarded as burn-in. Mixing good; acceptances good; ESS poor. Run 2: Marginal likelihood: $-\bar{L}_h = 304404.6237$. 2,000,000 generations, 10,000 samples, 5,000 discarded as burn-in. Posterior predictive simulations of X^2 : original statistic = 9598.5262, sample distribution = 13084.5398 to 19231.0647, $p = 1.0$. 2,000,000 generations, 10,000 samples, 5,000 discarded as burn-in. Mixing good; acceptances good; ESS poor.

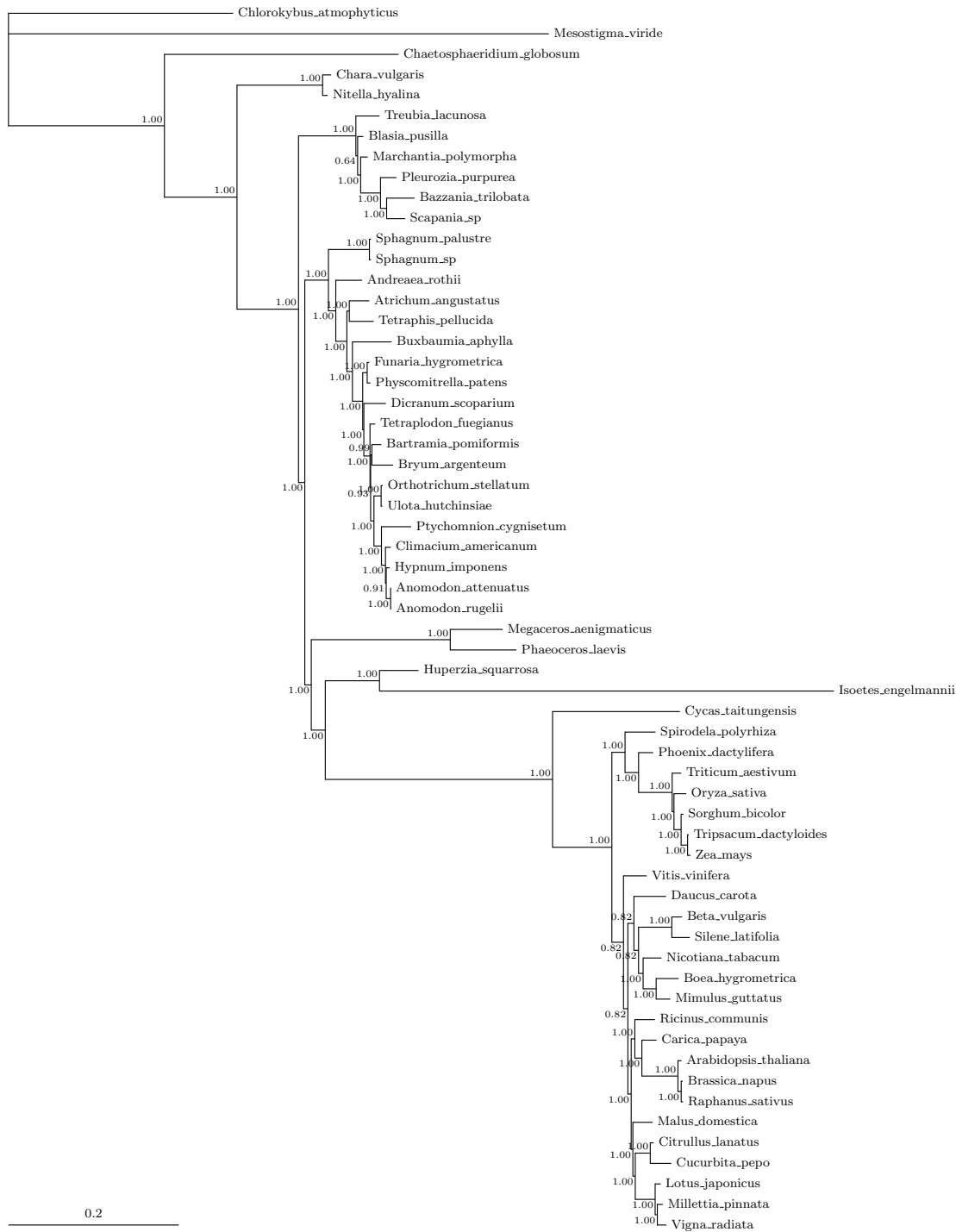


Figure S12 Protein data: MCMC P4 stmtREV+ Γ +CV2+PP Non-stationary composition heterogeneous: 2 composition vectors. Combined marginal likelihood: $-\bar{L}_h = 164325.0067$ Run 1: Marginal likelihood: $-\bar{L}_h = 164327.961414$. Posterior predictive simulations of X^2 : original statistic = 5457.1415, sample distribution = 4428.2591 to 7835.4688, $p = 0.8832$. 2,000,000 generations, 20,000 samples, 14,000 discarded as burn-in. Mixing good; acceptances good; ESS excellent. Run 2: Marginal likelihood: $-\bar{L}_h = 164323.7090$. Posterior predictive simulations of X^2 : original statistic = 5457.1415, sample distribution = 4379.9707 to 8131.0608, $p = 0.8906$. 2,000,000 generations, 20,000 samples, 8,000 discarded as burn-in. Mixing poor; acceptances good; ESS good.



Figure S13 Protein-coding gene data: MCMC Phylobayes CAT-GTR+ Γ Marginal likelihood: $-\bar{L}_h = 266716.8110$. Run 1: 13,780 samples, 3,770 discarded as burn-in, $-\bar{L}_h = 266724.8638$. Run 2: 13,066 samples, 5,310 discarded as burn-in, $-\bar{L}_h = 266687.923676$. Maxdiff 0.184451; mean diff 0.0135794; ESS poor.

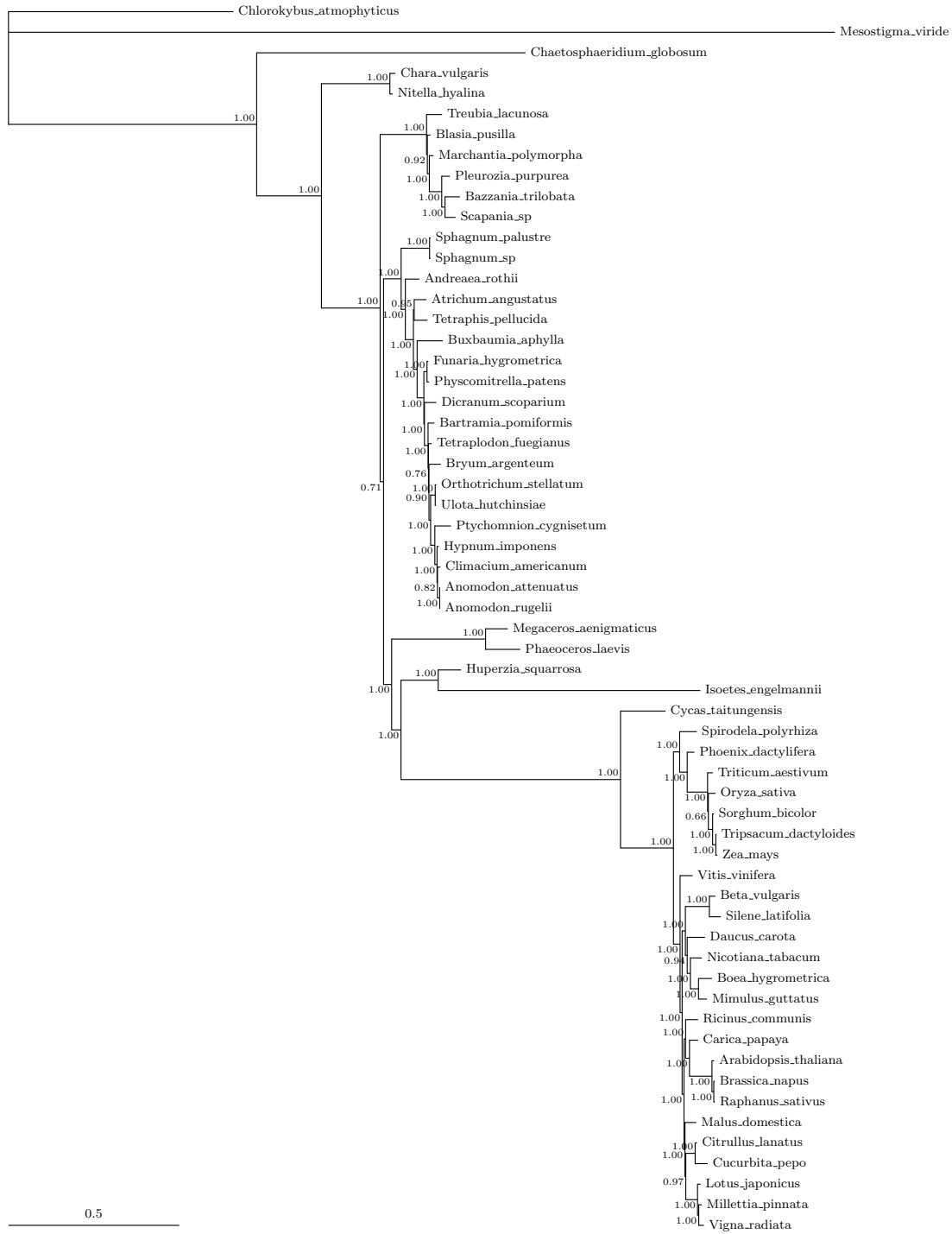


Figure S14 Protein data: MCMC Phylobayes CAT+GTR+ Γ_4 Marginal likelihood: $-\bar{L}_h = 130006.7891$. Run 1: 18741 samples, 10,000 discarded as burn-in, $-\bar{L}_h = 130014.9413$. Run 2: 19522 samples, 10,000 discarded as burn-in, $-\bar{L}_h = 129999.5307$. Maximum difference in clade support: 0.1153, mean difference in clade support: 0.0036.

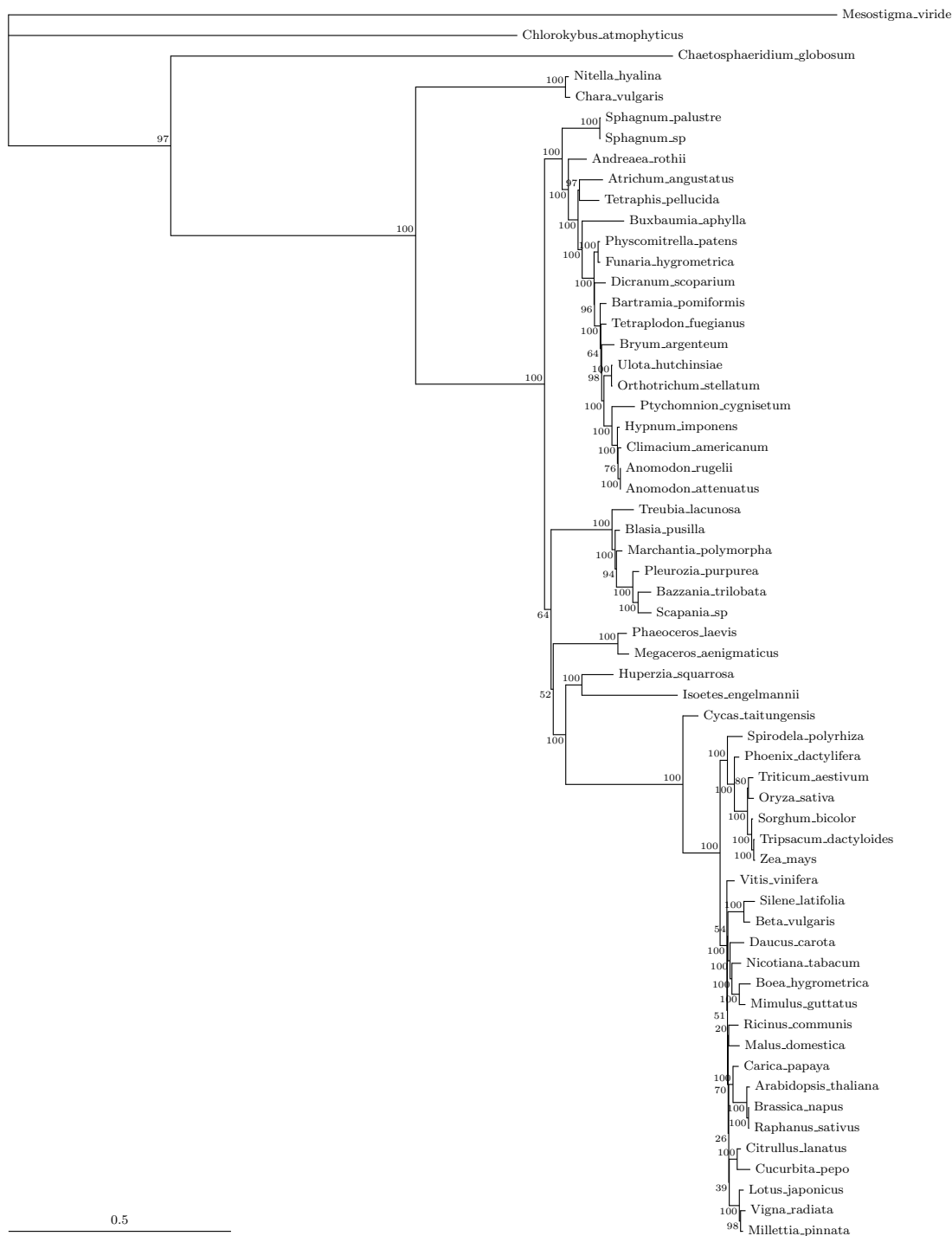


Figure S15 Protein-coding gene data: RAxML ML 3rd positions only GTR+I+Γ 100 bootstrap replicates, data partitioned by each gene.

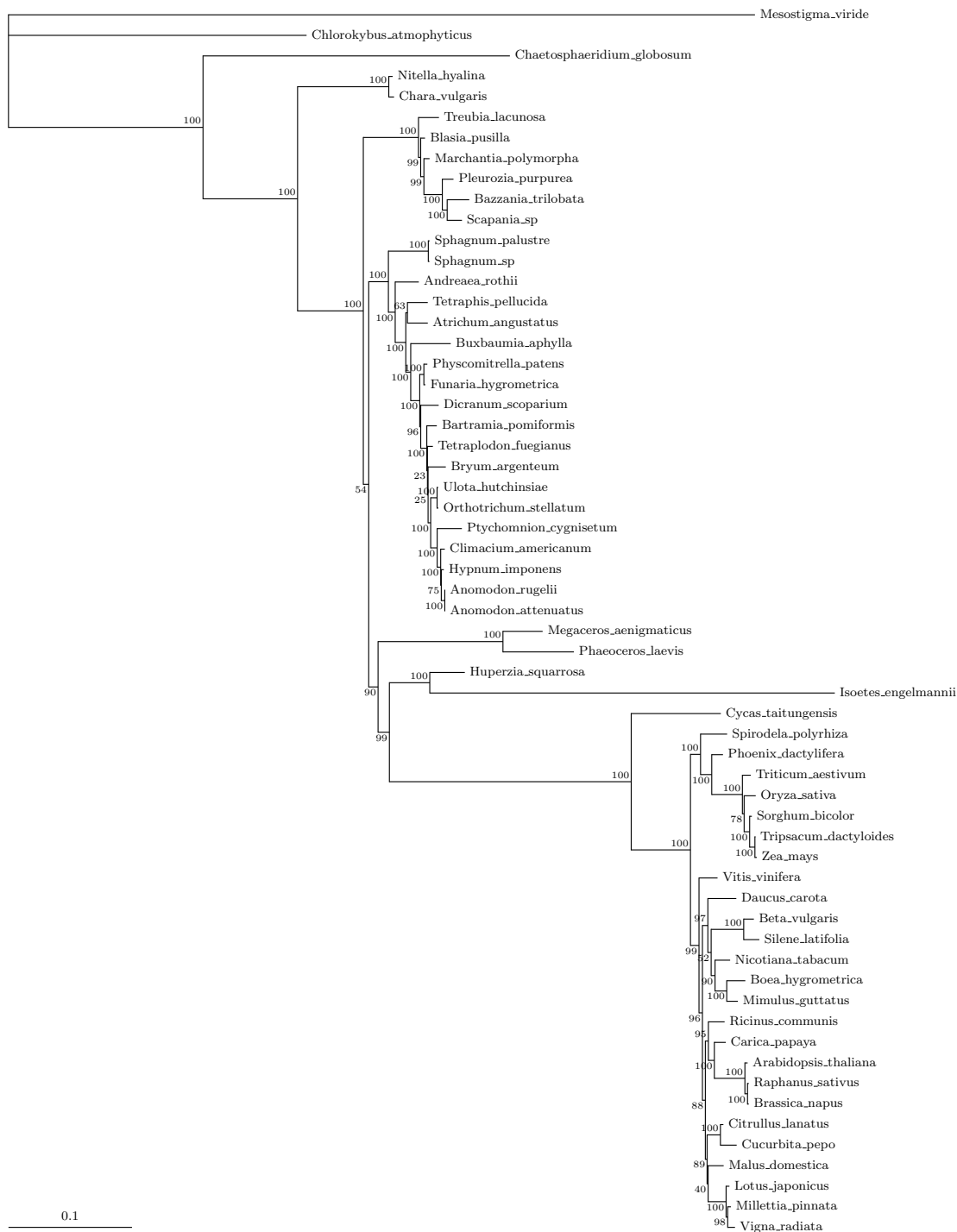


Figure S16 Protein-coding gene data: RAxML ML 1st & 2nd codon positions only GTR+I+Γ 100 bootstrap replicates, data partitioned by each gene.

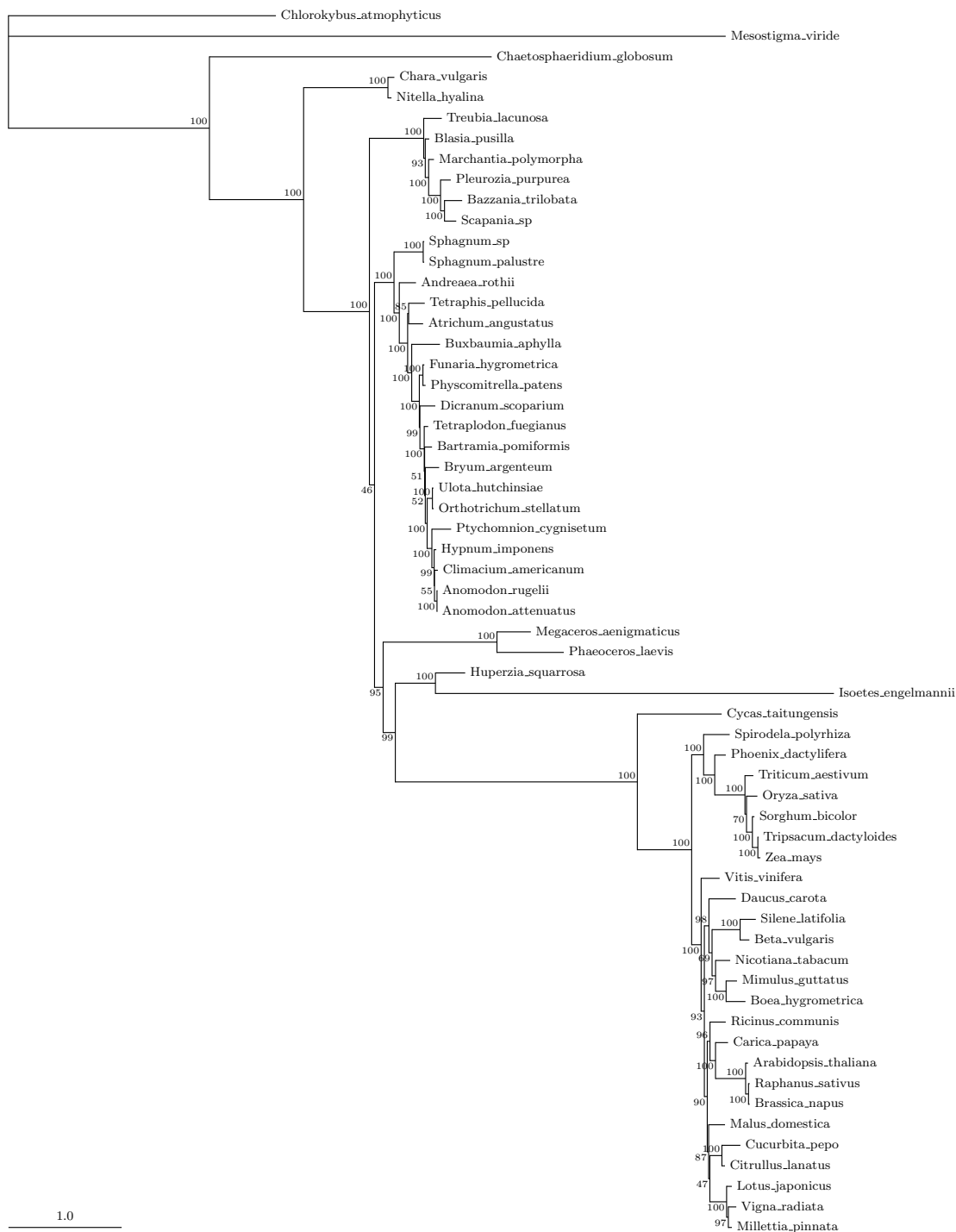


Figure S17 Codon-degenerated protein-coding gene data: RAxML ML GTR+I+Γ ML phylogram inferred from the codon-degenerated nt data. 100 bootstrap replicates, data partitioned by model gene-1-2-3 (each of the three codon positions in each gene as one partition). This was the optimal partitioning strategy determined by AIC and BIC (see Table S3).



Figure S18 Codon-degenerated protein-coding gene data: MCMC MrBayes GTR+I+ Γ Marginal likelihood: $-\bar{L}_h$ = 163345.0345. ASDOSS 0.000687. 5,000,000 generations, 10,000 samples, 2000 discarded as burn-in. Mixing poor; acceptances poor; PSRF mean 1.0026 stdev 0.0032; ESS good.

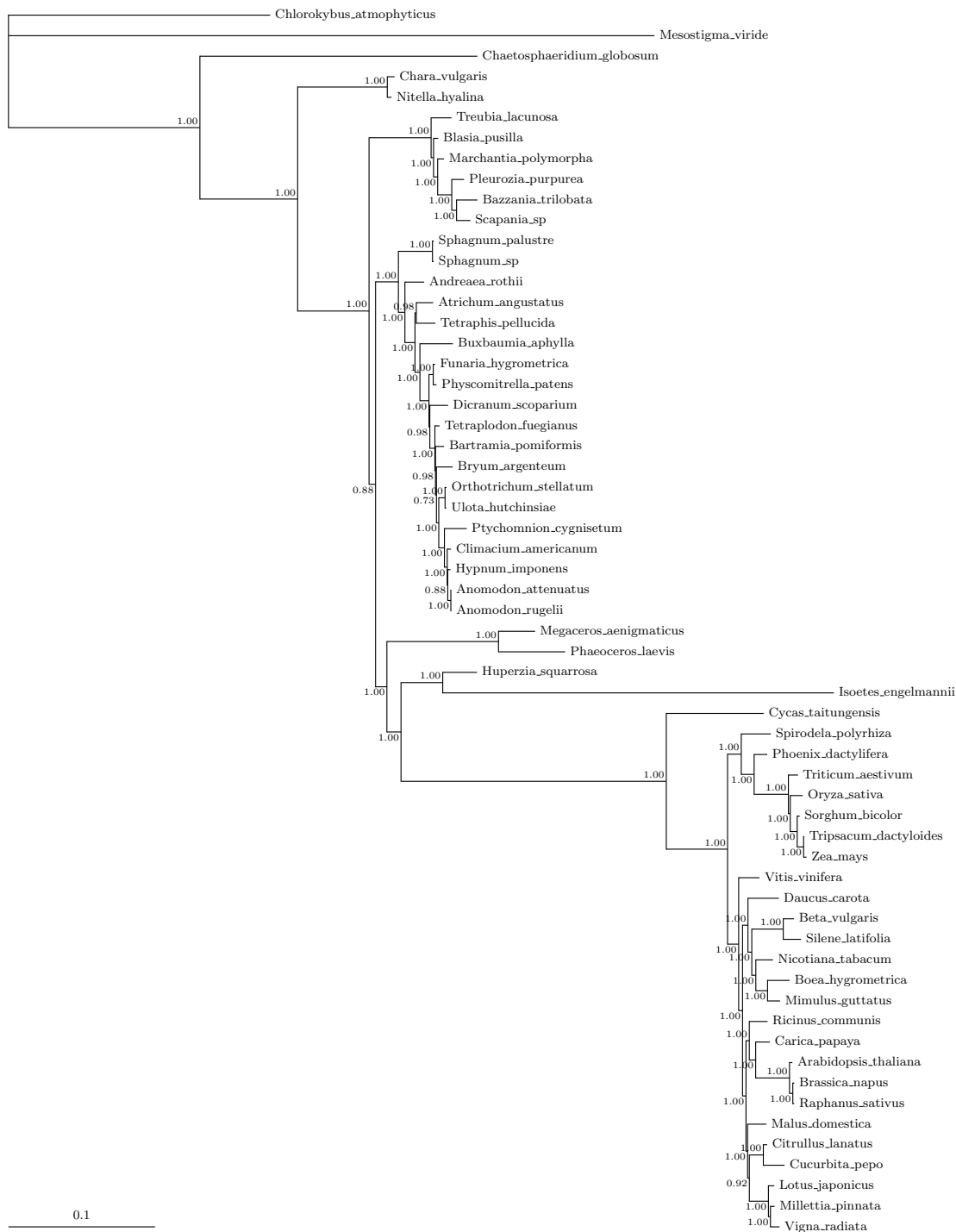


Figure S19 Codon-degenerated protein-coding gene data: MCMC P4 GTR+I+Γ+PP Marginal likelihood: $-\bar{L}_h = 163346.9570$. 2,000,000 generations, 10,000 samples, 7,000 discarded as burn-in. Posterior predictive simulations of X^2 : original statistic = 5452.9031, sample distribution = 13.8715 to 152.9760, $p = 0.0$. Mixing good; acceptances good; ESS poor.