

## Online Supplement 1: Misplaced *rbcL* accessions and upgrade to NCBI taxonomy tree

There are a number of *rbcL* sequences stored in gene banks (NCBI GenBank, DDBJ database, EMBL nucleotide database) that are erroneously placed in the NCBI taxonomy tree according to the accepted systematic frameworks such as APG II. A full list of such mislabeled accessions is provided in the following. These types of sequence misidentification and misplacement can seriously bias BS and ML inferences on comprehensive datasets if they remain undetected. However, they affect GRTS-based results to a lesser degree.

*Comprehensive BS and ML analyses* – Using all available *rbcL* data and a taxon-naming scheme that encodes the systematic affiliation, misplaced sequences could be identified directly from the resulting trees. Mislabeled sequences that belonged to a different group (e.g., a Fagales-like *rbcL* stored as a Rosales taxon) were nested within the ‘correct’ clade, i.e., Fagales for the given example. Taxa that corresponded to outgroup-type *rbcL* (e.g., an asterid *rbcL* erroneously labeled as Rosales) were not grouped with any other accession or other mislabeled sequences: Each of the best ML trees based on the four comprehensive matrices contained a moderately to highly supported ‘junk’ clade (e.g. “eurosoid junk clade” in Text-fig. 3). In these cases the *CA-BS* support can also be interpreted as support for the monophyly of the ingroup.

*GRTS analyses* – Based on our data, the effect of most erraneous (misplaced in the *CA-BS* trees) *rbcL* sequences is alleviated by numerous valid (grouped according to their pre-defined groups in the comprehensive ML trees) *rbcL* sequences, because each of the 100 *GRTS-ML* or 10 *GRTS-BS* matrices contains different sets of randomly selected representative taxa of the predefined groups. Naturally, a negative effect can only be observed if the proportion of “wrong” sequences compared to “correct” ones is comparatively high, i.e., if the probability of “wrong” sequences being sampled in GRTS-replicates increases. Thus, the reliability of our method for identification of mislabeled taxa is limited by the size of the original data basis. Thus, the more sequences are available for a group of interest (a TU), the lower the impact of taxonomic (misidentified source), contaminant, or systematic (misplaced source) error. This is encouraging, because it might allow for automated definition of TUs for GRTS, by directly incorporating the lineage information provided in the NCBI GenBank files, rather than the visual inspection of each subalignment and manual identification of aberrant sequences as performed here. In the case of angiosperms, the NCBI taxonomy tree included c. 20 genera with different systematic affiliation (family, order) at the time of download than proposed by APG II, APW, and the *rbcL* phylogeny.

In addition to misplaced sequences, there are some inconsistencies between the NCBI tree at the time of download and the systematic resources (APG II; APW and references given there):

NCBI Taxonomy Tree needing revision	According NCBI taxonomy tree	According APG II (order, family level)/APW/rbcL data (considering order, family, and genus)
<i>Drypetes</i> spp., <i>Sibangea</i> spp. ( <i>Lingelshemia</i> , <i>Putranjiva</i> , no rbcL data)	Malpighiales; Phyllanthaceae; Drypeteae p.p.	Malpighiales; Putranjivaceae
<i>Limeum</i> spp.	Caryophyllales; Molluginaceae	Caryophyllales; Limeaceae
<i>Harungana madagascariensis</i> [198768 AF518396]	Malpighiales; Clusiaceae	Malpighiales; Hypericaceae
<i>Vismia baccifera</i> [198795 AF518382]		
<i>Hypericum perforatum</i> [65561 AF206779]		
<i>Cratoxylum formosum</i> [198761 AF518395]		
<i>Paradrypetes subintegrifolia</i> [314339 AY788184]	Malpighiales; Phyllanthaceae; Drypeteae	Malpighiales; Rhizophoraceae
<i>Centroplassis glaucinus</i> [289629 AY663646]	Malpighiales; Phyllanthaceae; Phyllanthaceae <i>incertae</i>	Malpighiales; Malpighiales <i>incertae sedis</i> [monotypic]
<i>Desbordesia glaucescens</i> [300981 AY663631]	Sapindales; Simaroubaceae	Malpighiales; Irvingiaceae
<i>Irvingia malayana</i> [85215 AF123278/AF206782]		
<i>Klainedoxa gabonensis</i> [289639 AY663630]		
<i>Liquidamber</i> spp.	Saxifragales; Hamamelidaceae	Saxifragales; Altingiaceae
<i>Perrottetia</i> spp.	Celastrales; Celastraceae	Huerteales <i>incertae sedis</i>
<i>Aphloia theiformis</i> [112806 AF206735]	rosids; Aphloiaceae [unplaced family]	Crossosomatales; Aphloiaceae
<i>Geissoloma marginatum</i> [39019 X83990]	rosids; Geissolomataceae [unplaced family]	Crossosomatales; Geissolomataceae
<i>Ixerba brexioides</i> [85557 AF084475]	rosids; Ixerbaceae [unplaced family]	Crossosomatales; Ixerbaceae
<i>Strasburgeria robusta</i> [125056 AJ403007]	rosids; Strasburgeriaceae [unplaced family]	Crossosomatales; Strasburgeriaceae

New label	Taxon	Published in	According TaxonomyTree	According APW/AF	According <i>rbcL</i> (rosids and eudicots tree)
OxaOx_L14692	<i>Averrhoa carambola</i>	Price & Palmer, Ann MO Bot Gard, 1993	Oxalidales; Oxalidaceae	same	Identical to <i>Wendtia gracilis</i> L14708 (Geraniales; Ledocarpaceae)
CanRo_AF062004	<i>Trema micrantha</i>	Qiu et al., IJPS, 1998	Rosales; Ulmaceae	Rosales; Cannabaceae	Identical to <i>Leitneria floridana</i> AF062003 (Sapindales; Simaroubaceae)
PapFb_AY904389	<i>Andira aubletii</i>	Haston et al., AJB, 2005	Fabales; Fabaceae; Papilionoideae	same	Connaraceae (Oxalidales); highly similar (99%) to <i>Conarus conchocarpus</i> U06798
MyrFa_L01934	<i>Morella</i> [" <i>Myrica</i> "] <i>cerifera</i>	Albert et al., Science, 1992	Fagales; Myricaceae	same	Cannabaceae (Rosales); highly similar (99%) to <i>Celtis sinensis</i> D86309
JugFa_AY147094	<i>Cyclocarya paliurus</i>	Li et al., 2004?	Fagales; Juglandaceae	same	Sapindales, as sister to Simaroubaceae; most similar (97%) to <i>Leitneria</i> , <i>Ailanthus</i> and others
MorRo_L01933	<i>Morus alba</i>	Soltis et al., PNAS, 1990	Rosales; Moraceae	same	Nested among <i>Prunus</i> accessions (Rosales; Rosaceae)
MorRo_AJ390066	<i>Dorstenia psilurus</i>	Richardson et al., AJB, 2000	Rosales; Moraceae	same	Placed within one of the Rhamnaceae subclades including <i>Emmenosperma</i> , <i>Colubrina</i> , among others (97% identity)
CanRo_L12638	<i>Celtis yunnanensis</i>	Qiu et al., Ann MO Bot Gard 1993	Rosales; Ulmaceae	Rosales; Cannabaceae	Nested among <i>Morella</i> accessions (Fagales; Myricaceae)
CanRo_AY263941	<i>Celtis philippensis</i>	Li et al., 2004?	Rosales; Ulmaceae	Rosales; Cannabaceae	Placed between <i>Ulmus</i> and <i>Zelkova</i> (Rosales; Ulmaceae s.str.); 98% identity
BonMa_AJ402990	<i>Ploiariium sp.</i> Cameron s.n.	Savolainen et al., 2000	Malpighiales; Bonnetiaceae	same	Thymelaeaceae (Malvales); next to <i>Daphne</i> (96%), <i>Thymelaea</i> (97% identity), and others
SalMa_AY757086	<i>Gerrardina foliosa</i>	Alford, 2006	Malpighiales; Salicaceae	Huerteales; Gerradinaceae [APW, Mai 2007]	Sister to <i>Petenaee cordata</i> AJ233158, forming an isolated eurosid 2 clade; sequence most similar to some Huerteales (97% identity)
CapBr_AY483279	<i>Tirania purpurea</i>	Hall et al., Syst Bot, 2004	Brassicales; Capparaceae	same	Resedaceae
BraBr_X73284	<i>Sinapsis alba</i>	Capesius, unpublished [2004]	Brassicales; Brassicaceae	same	Papilionoideae (Fabales; Fabaceae)
EmbBr_AJ402949	<i>Emblingia calceoliflora</i>	Savolainen et al., Kew Bull, 2000	Brassicales; Emblingiaceae	same	Not a rosid, but an asterid <i>rbcL</i> ; forms a clade with misplaced <i>Sinopodophyllum hexandrum</i> (AF203487) and <i>Getonia</i> [" <i>Calycopteris</i> "] <i>floribunda</i> (AF281478), as sister clade to asterid clades included; sequence most similar to asterids (lamiids) <i>Gentiana</i> (L14398) and <i>Gentianella</i> (Y11862) (Gentianales; Gentianaceae) and <i>Hedyosmum</i> (AF121364; <i>Chloranthales</i> ; <i>Chloranthaceae</i> ), latter a putative misnamer or misplacer
SapSp_AJ402931	<i>Botteogoa insignis</i>	Savolainen et al., Kew Bull, 2000	Sapindales; Sapindaceae	same	Rutaceae; grouped with <i>Ptaeroxylon obliquum</i> (AF123276; 97%) and <i>Cneorum pulverulentum</i> (AF206752; 98% identity)
SapSp_AY510146	<i>Brucea javanica</i>	Lee et al., local Korean journal, 2004	Sapindales; Simaroubaceae	same	Anacardiaceae; grouped with <i>Rhus</i> , <i>Toxicentron</i> , and others (99% identity)
NitSp_DQ267159	<i>Nitraria sphaerocarpa</i>	Zhang et al., unpublished [2006]	Sapindales; Nitrariaceae	same	Peganaceae-subclade (Sapindales; same family according APG II); sequence identity with other Peg'ceae 99-100%
ComMy_AF281478	<i>Calycopteris</i> [ <i>Getonia</i> ] <i>floribunda</i>	Tan et al., unpublished [2001]	Myrtales; Combretaceae	same	not a rosid, groups with asterids (see <i>Emblingia calceoliflora</i> AJ402949 above); sequence most similar to asterids (lamiids) <i>Teucrium</i> (U78715; 97%) and <i>Teucrium</i> (L14411; 97%)
ComMy_AF281477	<i>Conocarpus erectus</i>	Tan et al., unpublished [2001]	Myrtales; Combretaceae	same	Malvaceae (Malvales); grouped with <i>Conocarpus</i> (AF281477), <i>Hermannia</i> (AJ233130), and <i>Lavatera</i> (AJ233122; 97% sequence similarity)
OnaMy_AY841634	<i>Mkilua fragrans</i>	Chatrou et al., unpublished [2005]	Myrtales; Onagraceae	Magnoliales; Annonaceae	apparently not an eudicot (member of "junk clade"); most similar (98%) to Annonaceae <i>Cymbopetalum</i> (AY743442) and <i>Trigynaea</i> (AY841660)
AizCa_AJ235778	<i>Delosperma echinatum</i>	Savolainen et al., Syst Biol, 2000	Caryophyllales; Aizoaceae	same	within Nyctaginaceae-Sarcobataceae-Phytolaccaceae clade, grouped with <i>Phytolacca</i> (DQ006112; M62567; 98% identity)
PgnCa_M77702	<i>Rheum x hybridum</i>	Giannasi et al., Syst. Bot., 1992	Caryophyllales; Polygonaceae	same	Plumbaginaceae, virtually identical (100%, shorter sequence) to <i>Plumbago</i> (Y16905, Y16906, 100%), both grouped with <i>Dyerophytum</i> (AJ312253, 98% identity)
EbeEr_AJ402968	<i>Lissocarpa benthamii</i>	Savolainen et al., Kew Bull, 2000	Ericales; Ebenaceae	same	Rutaceae (Sapindales), in a subclade including <i>Pleiospermium</i> (AF066821, 95% identity), <i>Clausena</i> (AF066813, 94%), <i>Severinia</i> (AF066806, 94%), and others
TstEr_AY380342	<i>Archytaea multiflora</i>	Davis & Chase, AJB, 2004	Ericales; Ternstroemiaceae	Malpighiales; Bonnetiaceae	placed as sister taxon to <i>Bonnetia</i> (96% identity); most similar (97%) to Clusiaceae <i>Mesua</i> (AY625024), <i>Poeciloneuron</i> (AY625023), and <i>Calophyllum</i> (Z75672)
MsnEr_L12598	<i>Ardisia crenata</i>	Kron & Chase, Ann MO Bot Gard, 1993	Ericales; Myrsinaceae	same	Ericaceae, in a subclade including <i>Arbutus</i> (AF419813, L12597; 97% identity); <i>Arctostaphylos</i> (AF419814, 95%), and <i>Comarostaphylis</i> (AF419819, 96%)
PplEr_AJ402986	<i>Pentaphylax euryoides</i>	Savolainen et al., Kew Bull, 2000	Ericales; Pentaphylacaceae	same	member of "eudicot junk clade"; most similar (98%) to Orchidaceae (Aspergales) such as <i>Megastylis</i> (AY381125) and <i>Chiloglottis</i> (AF074124), among others
PplEr_AJ428891	<i>Pentaphylax euryoides</i>	Bremer et al., MPE, 2002	Ericales; Pentaphylacaceae	same	exact duplicate of <i>Pelliciera rhizophorae</i> (AJ428893, same authors)
GroSf_AF299092	<i>Quintinia verdonis</i>	Xiang & Soltis, unpublished [2000]	Saxifragales; Grossulariaceae	euasterid II	grouped with asterid clades included; most similar (97%) to <i>Linconia</i> (AY490993; Bruniaceae), <i>Pennantia</i> (AJ494842, AJ494843; Apiales; Pennantiaceae), <i>Viburnum</i> (X87398; [Sphenostemonaceae])
GroSf_X87394	<i>Quintinia verdonis</i>	Gustafsson et al., PSE, 1996	Saxifragales; Grossulariaceae	Quintiniaceae	Dipsacales; Adoxaceae and other campanulids
SaxSf_AF206766	<i>Fendlera rupicola</i>	Soltis et al., unpublished [2003]?	Saxifragales; Saxifragaceae	Cornales; Hydrangaceae	Hydrangaceae; sister to <i>Jamesia americana</i> (AF323201, 98% identity)
PdsSf_AJ403018	<i>Whittonia guanensis</i>	Savolainen et al., Kew Bull, 2000	Saxifragales; Peridiscaceae	same	Malpighiales, as sister taxon to Malpighiaceae clade; most similar (97%) to various Malpighiaceae
CelCe_AY935723	<i>Bhesa robusta</i>	Zhang & Simmons, Syst Bot, 2006	Celastrales; Celastraceae	Malpighiales <i>incertae sedis</i>	Malpighiales, forming a subclade of ambiguous systematic position; most similar (96%) to <i>Vantanea</i> (Z75679; Humiriaceae), <i>Philyra</i> (AY794920; Euphorbiaceae), <i>Lindackeria</i> , and <i>Kiggelaria</i> (AJ418800, AY788180; member of "eudicot junk clade"; most similar to Calycanthaceae (Laurales) such as <i>Calycanthus floridus</i> var. <i>glauca</i> (AJ428413; 94%), <i>Idiospermum</i> (L12651; 93%), and <i>Chimonanthus</i> (L12639; 93%)
CelCe_AY935722	<i>Bhesa paniculata</i>	Zhang & Simmons, Syst Bot, 2006	Celastrales; Celastraceae	same	member of "eudicot junk clade"; most similar to <i>Taxodium distichum</i> AF119185 (same author!; Coniferales: Cupressaceae)
CelCe_AY935740	<i>Stackhousia monogyne</i>	Zhang & Simmons, Syst Bot, 2006	Celastrales; Celastraceae	same	member of "eudicot junk clade"; most similar to <i>Taxodium distichum</i> AF119185 (same author!; Coniferales: Cupressaceae)
NysCo_AF119178	<i>Nyssa sylvatica</i>	Wang & Lincoln, unpublished? [2000]	Cornales; Nyssaceae	same	Huerteales, grouped with <i>Tapiscia</i> (Tapiscaceae; 98% identity)
StaCs_AY646109	<i>Huetea glandulosa</i>	Simmons & Simpson, unpublished? [2005]	Crossosomatales; Staphyleaceae	Huerteales; Tapisciaceae	sister to <i>Gerrardina foliosa</i> AY757086, forming an isolated eurosid 2 subclade; sequence most similar to some Huerteales (95% identity)
PetMv_AJ233158	<i>Petenaee cordata</i>	Bayer et al., BotJLS, 1999	Oxalidales; Elaeocarpaceae	Malvales <i>incertae sedis</i>	identical to <i>Trochodendron aralioides</i> (L01958)
EicOx_AJ233158	<i>Petenaee cordata</i>	Bayer et al., BotJLS, 1999	Oxalidales; Elaeocarpaceae	Malvales <i>incertae sedis</i>	identical to <i>Trochodendron aralioides</i> (L01958)
PlaPr_L01943	<i>Platanus occidentalis</i>	Albert et al., Science, 1992	Proteales; Platanaceae	same	identical to <i>Trochodendron aralioides</i> (L01958)
BerRa_AF203487	<i>Sinopodophyllum</i> [" <i>Podophyllum</i> "] <i>hexandrum</i>	Xiang et al., unpublished? [2003]	Ranunculales; Berberidaceae	same	not a basal eudicot, but an asterid (see <i>Emblingia calceoliflora</i> AJ402949 above); most similar (98%) to the lamiid <i>Campsis radicans</i> (several accessions; Lamiales; Bignoniaceae)
RanRa_AF093729	<i>Placospermum coriaceum</i>	Hoot et al., Ann MO Bot Gar, 1999	Ranunculales; Ranunculaceae	Proteales; Proteaceae	Proteaceae, sister taxon of <i>Persoonia lanceolata</i> U79178 (97% identity)
ZygZy_DQ267165	<i>Tribulus terrestris</i>	Zhang et al., unpublished [2006]	Zygophyllales; Zygophyllaceae	Zygophyllales; Zygophyllaceae	Peganaceae (Sapindales); highly similar (99%) to <i>Malacocarpus</i> (U39280) and <i>Peganum</i> (DQ267164)
ZygZy_DQ267163	<i>Tetraena mongolica</i>	Zhang et al., unpublished [2006]	Zygophyllales; Zygophyllaceae	Zygophyllales; Zygophyllaceae	<i>Nitraria</i> (Sapindales; Nitrariaceae s.str.; 99% identity)