

RESEARCH ARTICLE

Open Access

# A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree

Tiina Särkinen<sup>1,2\*</sup>, Lynn Bohs<sup>3</sup>, Richard G Olmstead<sup>4</sup> and Sandra Knapp<sup>1</sup>

## Abstract

**Background:** The Solanaceae is a plant family of great economic importance. Despite a wealth of phylogenetic work on individual clades and a deep knowledge of particular cultivated species such as tomato and potato, a robust evolutionary framework with a dated molecular phylogeny for the family is still lacking. Here we investigate molecular divergence times for Solanaceae using a densely-sampled species-level phylogeny. We also review the fossil record of the family to derive robust calibration points, and estimate a chronogram using an uncorrelated relaxed molecular clock.

**Results:** Our densely-sampled phylogeny shows strong support for all previously identified clades of Solanaceae and strongly supported relationships between the major clades, particularly within *Solanum*. The Tomato clade is shown to be sister to section *Petota*, and the *Regmandra* clade is the first branching member of the Potato clade. The minimum age estimates for major splits within the family provided here correspond well with results from previous studies, indicating splits between tomato & potato around 8 Million years ago (Ma) with a 95% highest posterior density (HPD) 7–10 Ma, *Solanum* & *Capsicum* c. 19 Ma (95% HPD 17–21), and *Solanum* & *Nicotiana* c. 24 Ma (95% HPD 23–26).

**Conclusions:** Our large time-calibrated phylogeny provides a significant step towards completing a fully sampled species-level phylogeny for Solanaceae, and provides age estimates for the whole family. The chronogram now includes 40% of known species and all but two monotypic genera, and is one of the best sampled angiosperm family phylogenies both in terms of taxon sampling and resolution published thus far. The increased resolution in the chronogram combined with the large increase in species sampling will provide much needed data for the examination of many biological questions using Solanaceae as a model system.

## Background

Divergence times are of major interest for studies of evolutionary biology and historical biogeography, but also to researchers who focus on understanding various types of trait evolution, such as the development of chemical and genetic pathways, climatic niche and geographic range sizes, and morphological, ecological and behavioural characters. With the recent publication of fully annotated genomes in the Solanaceae [1-3], genomic tools now exist for unravelling genetic mechanisms that

control such traits and their development. What is lacking, however, is a robust phylogenetic framework that encompasses species and generic diversity across the family in order to maximise the potential of these new data sources in a wider evolutionary context. Although several studies have focused on understanding evolution of particular characteristics in Solanaceae in a phylogenetic context, including analyses of genome and chromosome evolution [4-6], life history and polyploidy [7-9], floral and fruit morphology [10,11], gene family evolution and sub-functionalization [12-15], and broad-scale biogeographic patterns [16], only a single study has examined character evolution through time [8]. A central problem has been the lack of a robust, densely sampled, dated molecular phylogeny for the entire family.

\* Correspondence: t.sarkinen@rbge.ac.uk

<sup>1</sup>Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK

<sup>2</sup>Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK

Full list of author information is available at the end of the article

Solanaceae are a particularly interesting angiosperm family, not only because they include many major crop species (e.g., potato - *Solanum tuberosum*, tomato - *S. lycopersicum*, eggplant - *S. melongena*, sweet and chili peppers - *Capsicum* spp., tobacco - *Nicotiana* spp.) and ornamentals (e.g., *Petunia* spp., *Solanum* spp.), but also a number of taxa used as biological model systems (e.g. *Nicotiana* spp., *Solanum* spp., *Petunia* spp., *Datura* spp.) [17-20]. A taxonomic and phylogenetic framework is now available for the family at the generic level (see [www.solanaceasource.org](http://www.solanaceasource.org)) [21,22]. Separate studies have explored relationships at tribal [23-26] and generic levels [27-33], and many phylogenetic studies have focused on *Solanum*, a genus that comprises nearly half of the species within the family [34-40]; see [21] for references prior to 2006]. Some of these studies have resulted in taxonomic changes, such as re-circumscription of the previously distinct genera *Lycopersicon*, *Cyphomandra*, *Normania*, and *Triguera* as parts of a monophyletic *Solanum* [35,41,42]. Such changes, although sometimes disruptive in the short term, have helped to both stabilise names and provide a better evolutionary context for future studies.

Although a relatively robust understanding of the major clades within Solanaceae exists, a densely sampled species-level phylogeny is still lacking. The most recent molecular systematic study focused on establishing major relationships within the family, but lacked depth in terms of sampling as it only included 190 (c. 7.3%) of a total c. 2,700 Solanaceae species [22]. A larger phylogenetic analysis with 995 species by Goldberg and colleagues [8] focused on the evolution of breeding systems, and did not discuss details of topology or implications for family-wide systematics. For *Solanum* itself, the most recent phylogeny included only 102 (7.7%) of the total c. 1,325 species in the genus [43]. Because several phylogenetic studies at various taxonomic levels across Solanaceae have since been published (see above), there is now a large quantity of new sequence data available for a wider family-level analysis.

Molecular divergence time analyses do not only depend on the availability of a robust, well-sampled phylogeny, but also require robust fossil calibration points [44,45]. The Solanaceae fossil record has never been fully reviewed, and only a few fossils have been used in molecular studies [8,46,47]. These studies used fossils as calibration points without a careful comparison of fossil morphology in relation to extant diversity. A recent survey of the earliest fossil record of the Asterid clade, including Solanaceae, highlighted the need to re-assess the earliest putative Solanaceae fossils that could provide robust calibration points for the crown or stem node of the family [48].

This study is part of a collaborative approach to studying the taxonomy and phylogeny of the Solanaceae. Here

we present a densely sampled phylogenetic study of the family coupled with a molecular dating analysis with fossil calibrations. We review all known seed fossils in the family, and assess them for identity, age, and phylogenetic position. We then use all available sequences for seven DNA loci found in GenBank with nearly all genera and 1,075 species represented. A dating analysis is run using an uncorrelated relaxed molecular clock model within a Bayesian framework with direct fossil calibrations. The resulting time-calibrated phylogeny offers important insights into the evolution of the family at different taxonomic levels, and a robust platform for future evolutionary studies.

## Results

### Fossil review

A total of 50 fossil records previously assigned to Solanaceae were found in the literature (Table 1, see Additional file 1 for full details). These included 39 seed fossils, one leaf fossil, five flowers, two wood and three pollen fossils (Table 1). None of the leaf or flower fossils showed any distinct morphological characters that allowed us to definitely assign them to the family. Of the two wood fossils, *Solanumxylum paranensis* can be clearly assigned to Solanaceae based on a large number of anatomical characters such as para- and apotracheal axial parenchyma that is diffuse in aggregates, simple perforation plates, bordered and alternate intervessel pits, homocellular rays, fibres that are polygonal and quadrangular in section, and the presence of septate fibers (Table 1) [49]. The other wood fossil shows no specific characters of Solanaceae except those common to Solanaceae and Asteraceae and lacks axial parenchyma; we do not consider this a member of Solanaceae (Table 1) [50]. Of the two pollen records, the classification of *Datura* cf. *discolor* awaits further examination, since no description or illustration of the fossil was provided in the original publication (Table 1). A pollen fossil-taxon from California, based on two poorly preserved specimens of 3-colporate, 5-colpate, prolate shaped grains with striate ornamentation, resembles pollen grains of *Lycium*, *Nolana*, and *Hyoscyamus* [51,52]. Similar characters appear in the pollen of the unrelated genera *Brucea* (Simaroubaceae) and *Skimmia* (Rutaceae) [51,53], and hence we have not assigned this pollen fossil-taxon to Solanaceae for our analysis (Table 1).

The putative Solanaceae seed fossils were analysed using a combination of characteristics known from clades within the family [54]: (1) Seeds flattened, (2) circular to reniform in shape, (3) hilum sub-laterally or laterally positioned, and (4) testa cells sinuate-margined. We assigned seeds with all four of these characters to the subfamily Solanoideae (N = 28), while those with some but not all of these were assigned to the family as

**Table 1 Fossil records of Solanaceae**

Name	Age (Mya)	Epoch	Organ	Fossil strata	Country	Solanaceae	Phylogenetic positioning
<i>Datura cf. stramonium</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Kholmech, Dnieper river	Belarus	Yes	Solanoideae
<i>Hyoscyamus</i> sp.	5.3 - 2.6	Pliocene	seed	Kroskienko	Poland	Yes	Solanoideae
<i>Hyoscyamus niger</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Pont-de-Gail, Cantal	France	Yes	Solanoideae
<i>Physalis alkekengi</i>	7.3 - 3.6	Mio/Pliocene: Messinian-Zanclean	seed	Saugbagger Flora	France	Yes	Solanoideae
<i>Physalis alkekengi</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Kholmech, Dnieper river	Belarus	Yes	Solanoideae
<i>Physalis alkekengi</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Bezirk Halle	Germany	Yes	Solanoideae
<i>Physalis pliocenica</i>	28.4 - 23.0	Oligocene: Chattian	seed	Zentendorf	Germany	Yes	Solanoideae
<i>Physalis pliocenica</i>	5.3 - 2.6	Pliocene	seed	Kroskienko	Poland	Yes	Solanoideae
<i>Physalis pliocenica</i>	11.6 - 7.3	Miocene: Tortonian	seed	Stare Gliwice	Poland	Yes	Solanoideae
<i>Physalis pliocenica</i>	13.7 - 11.6	Miocene: Serravallian	seed	Tongrube Forst Neukollm	Germany	Yes	Solanoideae
<i>Physalis pliocenica</i>	13.7 - 11.6	Miocene: Serravallian	seed	Klettwitz	Germany	Yes	Solanoideae
<i>Physalis alkekengi</i>	7.3 - 3.6	Mio/Pliocene: Messinian-Zanclean	seed	Nochten-Ost	Germany	Yes	Solanoideae
<i>Physalis</i> aff. <i>alkekengi</i>	5.3 - 3.6	Pliocene: Zanclean	seed	Baldevo Formation	Bulgaria	Yes	Solanoideae
<i>Scopolia carniolica</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Bezirk Halle	Germany	Yes	Solanoideae
<i>Solanispermum reniforme</i>	48.0 - 46.0	Eocene: Late Ypresian/Early Lutetian	seed	Lower Bagshot (Arne), Poole Formation	UK	Yes	Solanaceae
<b><i>Solanispermum reniforme</i></b>	<b>48.0 - 46.0</b>	<b>Eocene: Late Ypresian/Early Lutetian</b>	<b>seed</b>	<b>Bournemouth Freshwater Beds, Poole Formation</b>	<b>UK</b>	<b>Yes</b>	<b>Solanaceae</b>
<i>Solanispermum reniforme</i>	44.0 - 40.0	Eocene: Late Lutetian	seed	Boscombe Sand Formation	UK	Yes	Solanaceae
<i>Solanispermum reniforme</i>	40.4 - 37.2	Eocene: Bartonian	seed	Highcliffs sands/Cliff End Beds, Barton Formation	UK	Yes	Solanaceae
<i>Solanispermum reniforme</i>	33.9 - 28.4	Oligocene: Rupelian	seed	Bovey Tracey	UK	Yes	Solanaceae
<i>Solanum arnense</i>	48.0 - 46.0	Eocene: Late Ypresian/Early Lutetian	seed	Lower Bagshot (Arne), Poole Formation	UK	Yes	Solanaceae
<i>Solanum</i> cf. <i>persicum</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Kholmech	Belarus	Yes	Solanoideae
<i>Solanum dulcamara</i>	7.3 - 3.6	Mio/Pliocene: Messinian-Zanclean	seed	Saugbagger Flora	France	Yes	Solanoideae
<i>Solanum dulcamara</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Bezirk Halle	Germany	Yes	Solanoideae
<i>Solanum dulcamara</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Nordhausen	Germany	Yes	Solanoideae
<i>Solanum dulcamara</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Rippersroda	Germany	Yes	Solanoideae
<i>Solanum dulcamara</i>	5.3 - 2.6	Pliocene	seed	Tegelen-Sur-Meuse	Holland	Yes	Solanoideae
<i>Solanum dulcamara</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Pont-de-Gail, Cantal	France	Yes	Solanoideae

**Table 1 Fossil records of Solanaceae (Continued)**

<i>Solanum dulcamara</i>	3.6 - 2.6	Pliocene: Piazencian	seed	Pont-de-Gail, Cantal	France	Yes	Solanoideae
<i>Solanum dulcamara</i>	5.3 - 2.6	Pliocene	seed	Kroskienko	Poland	Yes	Solanoideae
<i>Solanum nigrum</i>	11.6 - 5.3	Miocene: Tortonian-Messinian	seed	North Rhine, Hambach	Germany	Yes	Solanoideae
<i>Solanum</i> sp.	5.3-2.6	Pliocene	seed	Limburg & Prussian border	Holland	Yes	Solanoideae
<b><i>Cantisolanum daturoides</i></b>	<b>55.0 – 50.0</b>	<b>Eocene: early Ypresian</b>	<b>seed</b>	<b>London Clay Formation</b>	<b>UK</b>	<b>No</b>	<b>-</b>
Unknown Solanaceae	5.3-2.6	Pliocene	seed	Limburg & Prussian border	Holland	No	-
<i>Physalis</i> sp.	5.3 - 0.0	Plio/Pleistocene	seed	Torre Picchio section (P11 & P12), Santa Maria di Ciciliano Formation	Italy	? <sup>1</sup>	-
<i>Solanum</i> sp.	1.7 - 1.2*	Pleistocene	seed	Olduvai Gorge	Tanzania	? <sup>1</sup>	-
<i>Physalis</i> sp.	23.0 - 16.0	Miocene: Aquitanian-Burdigalian	seed	Kireevsky Village, Ob River	Russia	? <sup>2</sup>	-
<i>Solanum</i> sp. #1	3.6 - 2.6	Pliocene: Piacenzian	seed	Chernoluch Village, Irtysh River	Russia	? <sup>2</sup>	-
<i>Solanum</i> sp. #2	16.0 - 11.6	Miocene: Langhian-Serravallian	seed	Novonikolsky Village, Irtysh River	Russia	? <sup>2</sup>	-
<i>Solanum</i> sp. #2	16.0 - 11.6	Miocene: Langhian-Serravallian	seed	Eburgulsky Village, Irtysh River	Russia	? <sup>2</sup>	-
<i>Solanumxylon paranensis</i>	16.0 - 11.6	Miocene: Langhian-Serravallian	wood	Paraná Formation	Argentina	Yes	Solanaceae
Solanaceae or Asteraceae	70.6 - 65.5	Cretaceous: Maastrichtian	wood	Panoche Formation, Del Puerto	USA	No	-
Pollen forma C	72.0 - 68.0	Cretaceous: Campanian-Maastrichtian boundary	pollen	San Joaquin Valley (D-1 & D-2), Great Valley Sequence	USA	No	-
<i>Datura</i> cf. <i>discolor</i>	37.2 - 33.9	Eocene: Priabonian	pollen	Florissant Basin	USA	? <sup>1</sup>	-
<i>Solanum</i> sp.	5.3 - 0.0	Plio/Pleistocene	pollen	Olduvai Gorge	Tanzania	? <sup>1</sup>	-
<i>Solandra haeliadum</i>	55.8 -33.9	Eocene	leaf	Salcedo	Italy	No	-
<i>Solanites brongniartii</i>	33.9 - 23.0	Oligocene	flower	Aix-en-Provence	France	No	-
<i>Solanites crassus</i>	55.8 - 40.4	Eocene: Lutetian-Ypresian	flower	Claiborne	USA	No	-
<i>Solanites pusillus</i>	55.8 - 40.4	Eocene: Lutetian-Ypresian	flower	Claiborne	USA	No	-
<i>Solanites saportanus</i>	55.8 - 40.4	Eocene: Lutetian-Ypresian	flower	Claiborne	USA	No	-
<i>Solanites sarachaformis</i>	55.8 - 40.4	Eocene: Lutetian-Ypresian	flower	Claiborne	USA	No	-

List of all known fossil records of Solanaceae. Fossils highlighted in bold were scanned using high-resolution X-ray computed tomography [T.Särkinen, M.Collinson, P.Kenrick, F.Ahmed, unpublished observations]. For full details of the fossil and references to primary sources, see Additional file 1.

<sup>1</sup> No description nor illustration or proper reference to specimen.

<sup>2</sup> Reference not located.

a whole (N = 6) (Table 1) [10]. The currently recognised earliest fossils assignable to Solanaceae include two seed fossils from Eocene Europe: *Solanispermum reniforme* recorded from various beds from southern England [55,56], and *Solanum arnense*, a fossil-taxon described based on a few specimens found from the Lower Bagshot (Table 1) [55]. Neither of these shows the combination of flat seeds with sinuate-margined testa cells, a unique combination that could tie them to the tribe

Solanoideae. The flattened seeds of *Solanispermum reniforme* lack sinuate margined testa cells, and *Solanum arnense* seeds show the characteristic testa cells but seeds are round rather than flattened. Hence, we consider these fossils as earliest evidence of Solanaceae and the presence of the family in Eocene Europe, but do not assign them to any particular clade within the family. Seeds of the fossil-taxon *Cantisolanum daturoides* [57] have previously been cited as the oldest known

Solanaceae fossil by some authors [10] but doubtfully a member of the family by others [48]. Results from a CT-scanning study have shown that this *Cantisolanum* seed is anatropous and does not belong to Solanaceae, but has likely affinities to the monocot family Phylodraceae [T. Särkinen, M. Collinson, P. Kenrick, F. Ahmed, unpublished observations].

### Solanaceae phylogeny

Our final supermatrix had a taxon coverage density of 0.45, and included 1,075 species of Solanaceae, representing all but two genera (the monospecific *Darcyanthus* and *Capsicophysalis*) and 40% of total species within the family, including 34% sampling of species within the large genus *Solanum*. Two plastid regions, *ndhF* and *trnL-F*, were available for all genera except *Darcyanthus* and *Capsicophysalis* and the plastid and nuclear regions ITS, *waxy* and *trnL-F* were the most densely sampled regions at the species level (Table 2). The matrix included a total of 4,576 variable characters, with an aligned length of 10,672 bp (Table 2). A total of 1,902 bp were excluded from analyses due to ambiguous alignment (see Methods section) resulting in a matrix of 8,770 bp (Table 2). Proportionately, *waxy* (33.9%) and *ndhF* (20.6%) contributed most PI (parsimony informative) characters (Table 2). The relatively little-used plastid region *trnS-G* showed a surprising number of PI characters (13.5% of total), considering it had relatively poor taxon coverage density (0.23), compared to *trnL-F* which had a coverage of 0.66 but only 6.6% of total PI characters (Table 2). The final matrix included 54.7% missing data (Table 2). At the species level, there was an average of 58.7% missing data, as measured by number of base pairs, but only 49.9% when measured in terms of PI characters expected from the missing regions.

The resolved Maximum Likelihood topology shows strong support for all previously identified major clades within Solanaceae [22], and increased node support is observed particularly within *Solanum* (Figure 1). Only major clades and their relationships are discussed here due to the fact that our analyses only accounted for incongruence issues amongst data sets between major clades rather than at shallow taxonomic levels. We encourage readers to refer back to available clade-specific studies for detailed species-level phylogenies (see references cited here and in ref. [21] for studies prior to 2006); these studies have incorporated larger sets of markers than used here, incorporate methods that test/account for gene tree – species tree incongruence, and discuss issues that could have led to any detected incongruences between gene trees such as polyploidy and/or hybridisation, and incomplete lineage sorting.

The branching order at the base of Solanaceae is not well defined, similar to the findings of Olmstead et al. [22], and four groups are identified as the first branching taxa: *Schizanthus*, *Duckeodendron*, the previously unplaced *Reyesia*, and the tribe Goetzeoideae (Figure 1). *Reyesia* has been previously associated with *Salpiglossis* [54], but is here placed with Goetzeoideae and *Duckeodendron* (Figure 1, Additional file 2). The previously unsampled genera *Heteranthis*, *Trianaea* and *Schraderanthus* are placed within Schwenckieae, Juanulloae, and Physalinae, respectively (Figure 1, Additional file 2). The informally named X = 12 clade is here recovered with strong support and Nicotianoideae is resolved as sister to the rest of the clade (Figure 1). Within the Physalinae, work is clearly needed to delimit monophyletic genera (Figure 1, Additional file 2, see [58,59]). Two closely related genera, *Larnax* and *Deprea*, are resolved as sister to Withaninae, in agreement with morphology (Figure 1, Additional file 2). These genera have been linked with Iochrominae in some molecular

**Table 2 Supermatrix details**

Region	Genome	Genera	Species	Aligned length <sup>1</sup>	Variable characters	%	PI characters	%	PI of combined matrix (%)	Missing data (%)	Taxon coverage density
<i>ITS</i>	Nuclear ribosomal	51	765	585	461	78.8	345	59.0	10.6	28.8	0.71
<i>matK</i>	Plastid	44	287	1027	419	40.8	262	25.5	8.1	73.3	0.27
<i>ndhF</i>	Plastid	90	472	2109	960	45.5	670	31.8	20.6	56.1	0.44
<i>psbA-trnH</i>	Plastid	30	223	698	340	48.7	227	32.5	7.0	79.3	0.21
<i>trnS-G</i>	Plastid	15	246	1329	694	52.2	439	33.0	13.5	77.1	0.23
<i>trnL-F</i>	Plastid	90	707	905	353	39.0	206	22.8	6.3	34.2	0.66
<i>waxy</i>	Nuclear	41	708	2085	1349	64.7	1104	52.9	33.9	34.1	0.66
<b>Total</b>		90	1075	8738	4576	52.4	3253	37.2	100	54.7	0.45

List of the seven regions used for building the Solanaceae mega-phylogeny, showing level of sampling at specific and generic levels, aligned length, and proportions of parsimony informative (PI) and variable sites per region. Values have been calculated without outgroups.

<sup>1</sup>ingroup only, excluded regions omitted.





(See figure on previous page.)

**Figure 1 Solanaceae phylogeny.** Phylogenetic relationships between major clades of Solanaceae based on a Maximum Likelihood analysis of a 1076 taxon supermatrix (ITS, *waxy*, *ndhF*, *matK*, *psbA-trnH*, *trnS-G*, *trnL-F*) with 10,672 bp of sequence data. Major clades recovered by previous phylogenetic studies [22,43,64] are labelled, as is the M Clade identified for the first time here. Clades with low bootstrap support (60-79%) are shown in pink, while strongly supported clades (bootstrap support 80-100%) are in black. **A.** Major clades of Solanaceae. **B.** Relationships within *Solanum*.

analyses [22], but considered distant outgroups of Iochrominae by others [58,59]. The molecular data support the treatment of *Schraderanthus* as distinct from *Leucophysalis* [60], and *Schraderanthus* is here found as sister to *Brachistus* + *Witheringia* (Figure 1, Additional file 2).

Within *Solanum*, all 12 major clades identified by Weese & Bohs [43] are recovered, with nearly fully resolved relationships among them (Figure 1). The Thelopodium clade is resolved as the first branching group, and the remaining *Solanum* species are divided into two strongly supported clades. Clade I comprises all non-spiny, often herbaceous (e.g., tomatoes, potatoes) species without stellate hairs, but also includes woody climbers (e.g., Dulcamaroids) and some shrubby species (e.g., Morelloids). Clade II comprises species that are often shrubs or small trees (although some are only weakly woody), many with prickles and/or stellate hairs (Figure 1). Within Clade I, which includes a total of c. 525 known species, two clear clades are resolved: (1) the Potato clade, with Regmandra clade as the first branching group, and (2) Clade M, including Morelloid, Dulcamaroid, Archaesolanum, Normania, and the African Non-Spiny clades (Figure 1). Relationships within Clade M are well resolved and highly supported, revealing the position of the African Non-Spiny clade as distinct from and not closely related to the Dulcamaroid clade, despite their morphological similarities such as a twining habit and twisting petioles [61]. Within the Potato clade, relationships are equally well resolved: section *Petota* is resolved as sister to a group comprising the Tomato clade plus a set of smaller early-branching clades (Figure 1). The Regmandra clade, a group of 11 species whose centre of diversity is the hyper-arid Atacama desert, is here resolved as part of the Potato Clade for the first time (Figure 1), a result supported by morphology [62,63].

Relationships within Clade II are less well-resolved. The clade consists of c. 800 mostly woody species, and includes the large Leptostemonum clade known as “spiny solanums”. There is moderate support for *S. clandestinum* + *S. mapiriense* as sister to the rest of Clade II (Figure 1). Relationships within the large Leptostemonum clade remain relatively unresolved, but all 14 major clades found in previous analyses [64] are supported. A set of previously unplaced species, *S. crotonoides*, *S. hayesii*, and *S. multispinum*, are resolved sequentially as sister to the Torva clade (Additional file 2), although on morphological grounds *S. hayesii* would be a member of the Torva clade.

### Molecular dating

The general topology of the Bayesian maximum clade credibility tree matched that of the best scoring Maximum Likelihood tree with similar levels of support for major clades (Additional file 3). The only topological difference, although not a hard incongruence, was observed at the base of Solanaceae: Bayesian analyses resolved Schwenckieae as the first branching group within the family, while the base of the tree remained largely unresolved in the maximum likelihood topology. Results from PATHd8 gave generally similar ages as those from the BEAST analysis (Table 3). A notable trend is that BEAST ages were consistently younger especially towards the early-branching nodes (Table 3). The younger ages obtained from the BEAST analysis reflect that diversification rates across Solanaceae have been non-linear especially towards the base of the tree, and/or that extinction and speciation rates have varied across the tree. We will focus our discussion on the BEAST results, which we consider to be more robust due to the more realistic model assumptions used, including the relaxed molecular clock model that accounts for rate variation across lineages, as well as Birth-Death tree model accounting for extinction [65].

The BEAST results place the stem age of Solanaceae at c. 49 Million years ago (Ma, 95% highest posterior density (HPD): 46–54 Ma), and the crown node at c. 30 Ma (95% HPD 26–34) (Additional files 2 and 3). The crown node of the  $x = 12$  Clade, which is the split between *Nicotiana* and *Solanum*, was estimated to be c. 24 My old (95% HPD 23–26) (Additional files 2 and 3). The Solanoideae began diversifying c. 21 Ma (95% HPD 19–23) (Additional files 2 and 3). *Solanum*, a genus which includes nearly half of the total species diversity in the family, split from *Jaltomata* c. 17 Ma (95% HPD 15–19) and started diversifying c. 16 Ma (95% HPD 13–18) (Additional files 2 and 3). The *Solanum* – *Capsicum* split, corresponding to the most common ancestor of *Solanum* & *Physalis*, occurred c. 19 Ma (95% HPD 17–21) (Additional files 2 and 3). Within *Solanum*, major splits include tomato – potato c. 8 Ma (95% HPD 7–10) and the eggplant – tomato/potato lineages corresponding to the Clade I – Clade II split c. 14 Ma (95% HPD 13–16) (Additional files 2 and 3). Crown node age estimates show that section *Petota*, which includes all cultivated potatoes, started diversifying c. 7 Ma (95% HPD 6–9), section *Lycopersicon*, which includes the cultivated

**Table 3 Molecular age estimates**

Clade	Group members	Dating method	
		BEAST (uncorrelated lognormal relaxed clock model)	PATHd8 (local clock model)
Tomato-potato	<i>S. lycopersicon</i> – <i>S. tuberosum</i>	8.0 (6.6-9.5)	9.1
Eggplant-tomato/potato	<i>S. lycopersicon</i> – <i>S. melongena</i>	14.3 (12.5-16.3)	18.8
Section <i>Lycopersicon</i>	<i>S. lycopersicon</i> – <i>S. peruvianum</i>	2.0 (1.2-2.6)	2.9
Section <i>Petota</i>	<i>S. tuberosum</i> – <i>S. piurae</i>	7.1 (5.9-8.5)	7.5
Eggplant clade	<i>S. melongena</i> – <i>S. macrocarpon</i>	3.4 (2.7-4.1)	2.8
<i>Capsicum</i> cultivated	<i>C. frutescens</i> – <i>C. eximium</i>	3.4 (2.7-4.3)	3.6
<i>Solanum</i> Clade I – Clade II	<i>S. melongena</i> – <i>S. tuberosum</i>	14.3 (12.5-16.3)	18.8
<i>Solanum</i> (crown)	<i>S. melongena</i> – <i>S. thelopodium</i>	15.5 (13.3-17.5)	22.0
<i>Solanum</i> (stem)	<i>Solanum</i> – <i>Jaltomata</i>	17.0 (14.5-18.7)	23.6
<i>Solanum</i> - <i>Capsicum</i>	<i>Solanum</i> – <i>Capsicum</i>	19.1 (17.0-21.0)	25.6
Solanoideae	<i>Solanum</i> – <i>Lycium</i>	21.0 (19.0-23.3)	25.6
<i>Solanum</i> - <i>Nicotiana</i> (i.e., X = 12 clade) (i.e., Solanoideae stem)	<i>Solanum</i> – <i>Nicotiana</i>	24.0 (23.0-25.7)	28.4
Solanaceae (crown)	<i>Solanum</i> – <i>Schizanthus</i>	30.4 (26.3-34.0)	32.1
Solanaceae (stem)	Solanaceae - Convolvulaceae	49.2 (46.2-53.7)	54.8

Minimum age estimates produced based on both PATHd8 and BEAST analyses for Solanaceae and its major clades. Clades refer to crown nodes unless otherwise indicated. Ages in brackets are 95% confidence intervals for BEAST. A densely sampled, time calibrated Solanaceae mega-phylogeny with associated confidence intervals can be found in Additional file 2.

tomato, c. 2 Ma (95% HPD 1–3), the group containing all species of cultivated eggplants (*S. melongena*, *S. anguivi* and *S. macrocarpon*) c. 3 Ma (95% HPD 2–4), and the group (*C. frutescens* – *C. eximium*) including all cultivated pepper species c. 3 Ma (95% HPD 2–4) (Additional files 2 and 3).

## Discussion

### Phylogenetic relationships within Solanaceae

Although individual studies have contributed significantly to a better understanding of the systematics and evolution of the family at generic and tribal levels, our results bring together data from a large number of studies into a single analysis, and present a coherent view on the current systematic knowledge of this diverse family and its major clades. Our analyses support all of the major clades previously identified within Solanaceae [22], *Solanum* [43] and the Leptostemonum clade of *Solanum* [64]. All of these major clades within the family are now strongly supported, and furthermore, our results reveal strongly-supported relationships between the major clades of the mega-diverse genus *Solanum*, strengthening the backbone.

The increased resolution in the current phylogeny can be attributed to both the increased sampling of markers as well as species. In the quest for better resolved phylogenies, studies often seek large amounts of sequence data, but it is now well established that increased species sampling can have an equally positive effect on phylogenetic resolution and accuracy [66–68]. The addition of more species to a data set has the effect of splitting long branches and detecting multiple substitutions, as well as resolving phylogenetic conflict, improving parameter estimation, and making inferences less dependent on particular evolutionary models [68]. In our approach we chose to maximise species sampling, while minimising missing data by choosing only the most densely sampled markers available. This approach generally boosted resolution without introducing any of the significant negative effects that large amounts of missing data can have on phylogeny estimation.

Our study presented here is a significant step forward in working towards a fully sampled species-level phylogeny for Solanaceae. A previous study by Goldberg et al. [8] included 995 species but did not present a fully annotated molecular phylogeny that would allow an



analysis of systematic relationships within the family. With > 1,000 species now covered, the current phylogeny includes 40% of known species and all genera of Solanaceae, except the monospecific and recently segregated *Darcyanthus* and *Capsicophysalis*. This is a substantial improvement on previous studies, and our current phylogeny is one of the best sampled family-level studies in angiosperms e.g., [69-71].

The sampling is now adequate to test for generic monophyly in previously poorly sampled groups. Although the number of genera is becoming stable with 97 currently recognised genera in Solanaceae (recent changes include those documented in refs. [26,60,72]), our analyses support previous results in identifying a set of groups where generic re-evaluation will be necessary, including *Lycianthes/Capsicum*, the genera in the Physalineae (especially *Physalis*) [59], *Deprea/Larnax*, the Iochrominae [58], and the Australian endemics in the Anthocercideae (see Additional file 2); many of these clusters of generic problems have been identified by previous authors.

Broader level relationships within Solanaceae and *Solanum*, as well as generic delimitations and problems identified in previous studies are supported by our species-rich dataset. Relationships between some of the major clades remain unresolved, however, most notably those at the base of the family and within the Solanoideae, and the *Leptostemonum* clade of *Solanum*. Resolving these nodes will be a priority in order to better understand evolution of some particularly complex traits, such as chromosome evolution. For example, resolving the sister group to the X = 12 clade, as well as the first branching taxa within Solanaceae, would allow us to determine the ancestral base chromosome number in the family and to fully understand directionality of chromosome evolution. Despite the increased resolution introduced by the use of more sequence data and higher species-level sampling, our results do not show any improvement in the resolution in these critical nodes. More genes will be needed to resolve these relationships, but the question remains which genes should be used. Highly variable nuclear loci, such as COSII markers already used in Solanaceae [73,74], and the PPR genes used in families within the related Asterid order Lamiales [75,76], present the most promising candidates. The widely sequenced regions ITS, *waxy*, *ndhF*, and *trnSG* are the most variable across the Solanaceae and species-level sampling using these regions should be increased. The traditionally used plastid marker *trnT-F*, which is relatively slowly evolving within Solanaceae, is known to include pseudogenes in *Solanum* [77] and care should be taken when using this region in phylogenetic studies.

#### **Solanaceae fossil record**

A few fossils have been used in previous molecular dating studies of Solanaceae, but without re-evaluation of

fossil morphology and hence their placement within the phylogeny [29,46]. As revealed by our literature review, a relatively large record for the family exists. The most usable evidence comes from fossil Solanaceae seeds, the oldest of which are from Eocene Europe (c. 48–40 Ma), with a sharp increase in the number of seed morphotypes observed towards the Pleistocene. The fossil seeds can be divided into two sets: (1) seeds showing four morphological characters present in the extant members of the Solanoideae, and (2) seeds that bear resemblance to the family in general but cannot be assigned to more specific clades within it because they lack the unique combination of seed flattening and presence of sinuate-margined testa cells. Although some of these fossils have been described with names associated with extant species and/or genera, our morphological review shows that none of them show unique morphological characters that can be used to place them to any extant genera. We consider the placement of these fossils on terminal nodes as has been done by previous authors [29,46] unjustified.

All of the fossils we were able to unambiguously identify as Solanaceae are from Eocene Europe, where none of the first branching lineages of the family occur. South America is the centre of diversity of extant Solanaceae, and all of the early diverging lineages are exclusively found in the New World. This suggests that the fossil record of the family is still far from complete, and that further studies on South American fossils might reveal crucial evidence with respect to the timing of diversification in Solanaceae. A promising avenue for future fossil studies would be to carefully evaluate wood fossil records, especially Cretaceous-Eocene material from the area in which the early-branching lineages all now occur [16,22].

#### **Dates for Solanaceae**

Our study is the largest Bayesian molecular dating analysis executed to date in terms of taxon sampling. Most previous studies have used Bayesian dating methods after pruning their original, large phylogenetic datasets largely due to an a priori assumption that Bayesian methods cannot cope with datasets with >500 terminals e.g., [78-80]. Our study with 1,075 species and >10,000 bp of sequence data demonstrates that large matrices with >500 terminals can be analysed using Bayesian dating methods. Further studies are needed, however, to fully explore best methods for analysing large datasets with the currently available dating methods that implement relaxed molecular clock models required for analyses of diverse clades where rates are expected to vary [81,82]. Such studies should focus on exploring trade-offs between number of taxa, complexity of models and partitions used in order to fully understand limitations and potential error sources in large scale analyses.

In our dating analysis, we followed the recent recommended best practice guidelines for fossil calibration [83] and placed fossil calibrations at stem nodes of the most inclusive extant groups using apomorphy-based morphological assignment. Morphological evidence from the seed fossils only allowed assignment to the broad groups Solanoideae or Solanaceae as a whole. Fossils provide only minimum age estimates for the nodes they are assigned, and hence results from our dating analysis where fossil calibrations were used should be considered as minimum age estimates. We further biased our results towards younger ages by assigning the oldest known fossils of Solanaceae to the stem node of the family rather than to more specific nodes within Solanaceae due to lack of morphological and anatomical characters that could be used to assign them to more specific nodes. There is always a possibility, however, that these seeds represent more specific clades within Solanaceae, which would push back age estimates for the family. Currently, the earliest fossil evidence for the family comes from Eocene Europe, but based on biogeographic analyses, the crown group of Solanaceae is thought to have originated and first diversified in South America [16,22,84]. Total evidence analysis, where fossils are placed as terminal taxa in the dating analysis using both molecular and morphological data matrix, could help in exploring the robustness of fossil placement [85], but as pointed above, the lack of characters in the Solanaceae seed fossils does not currently permit such analyses. The most promising avenue in strengthening the dating analysis would be in finding further fossil records (see Solanaceae fossil record above). This would increase the number of fossil calibration points and allow the use of cross-validation methods [86].

The rate of molecular evolution in plants has been found to correlate with life history traits, whereby longer living species show consistently lower substitution rates compared to shorter living species [81]. Molecular clock models should incorporate such rate variation, especially in groups such as Solanaceae which include a range of growth and life forms. Our dating analyses did not incorporate such models, although the model used in our Bayesian analysis allows rates to vary between lineages independently. The lack of such models in our analyses implies that the age of herbaceous, shorter lived plants (e.g., *Schizanthus* and the Tomato clade of *Solanum*) will be systematically overestimated, while ages in dominantly woody clades (e.g. *Solanum* Clade II) will be consistently underestimated. Future studies should explore how molecular clock models that account for rate variation due to life history traits could be implemented.

Previous studies have produced a wide range of estimates for the stem age of the family, ranging from 34–85 Ma [48,87–89], but none of these studies included

dense sampling within the family nor used robust Solanaceae-specific fossil calibrations. Paape et al. [90] analysed divergence times within Solanaceae but with a small dataset consisting of 29 species only. This study was based on three fossil calibration points without re-assessment or morphological study of the original fossils, and estimated Solanaceae stem age to have diverged 62 Ma (95% HPD 54–70 Ma) [90]. The oldest estimates for the family stem node age come from earlier molecular studies which used calibration points with more simplistic dating methods (65–85 Ma) [87,88], while the most recent molecular dating study of angiosperms by Bell et al. [89] who used 36 fossil calibrations across the tree and a relaxed molecular clock model, estimated the Solanaceae stem node to have diverged c. 59 Ma (95% HPD 49–68 Ma). Our results, which we consider as minimum ages, are broadly consistent with Bell et al. [89] in estimating the stem node of Solanaceae to date back to c. 49 Ma (95% HPD: 46–54).

The age of the major splits within the family has been of interest to various fields, including studies on chromosomal [4] and genome evolution [5,6,91]. Our minimum age estimates for the major splits between tomato – potato (c. 8 Ma, 95% HPD 7–10), eggplant – tomato/potato (c. 14 Ma, 95% HPD 13–16), *Solanum* – *Capsicum* (c. 19 Ma, 95% HPD 17–21), and *Solanum* – *Nicotiana* (c. 24 Ma, 95% HPD 23–26) are consistent with the age estimates produced in previous studies without fossil calibrations using much sparser sampling and more simplistic molecular clock models [4,6,91]. Our results for the *Nicotiana* – *Symonanthus* split (c. 15 Ma, 95% HPD 11–20) corroborate results obtained using island age (c. 15 Ma) [92] and those calculated using paralogy-free subtree analysis (>15 Ma for section *Suaveolentes*) [93]. Our results presented here suggest that the rate of chromosomal and genome evolution within Solanaceae has been marginally slower at least within particular lineages than previously thought. With the densely sampled chronogram presented in this study, a more detailed analysis of chromosomal evolution at the species level could now be performed in the Solanaceae to study rate differences and drivers of chromosomal changes such as environmental or life history factors. Similarly, morphological characters such as fruit type [10] could be analysed in relation to diversification rates to identify whether particular morphological traits are associated with speciation rate shifts in Solanaceae.

## Conclusions

Despite much focus on character and trait evolution within Solanaceae, little has been known about the origin of traits in the family in terms of time. We present here minimum age estimates and associated confidence intervals for the entire Solanaceae using a species-rich dataset comprising almost half of the species diversity

within the family. This densely sampled chronogram will provide the basis for unravelling the tempo and mode of evolution of many of the much-studied and complex traits in this diverse and economically important family such as self-incompatibility, fruit type, cold and salt tolerance, disease resistance, chromosomal re-arrangements, genome size, and gene sub-functionalization.

## Methods

### Fossil study

References to fossil records were compiled from various sources, including Yale Paleobotany Online Catalog (<http://peabody.yale.edu/collections/paleobotany>), the Paleobiology Database (<http://paleodb.org>), InsideWood Database (<http://insidewood.lib.ncsu.edu>), Burke Paleontology Collection Database ([www.washington.edu/burkemuseum/collections/paleontology](http://www.washington.edu/burkemuseum/collections/paleontology)), the Stratigraphy Database ([www.stratigraphy.net](http://www.stratigraphy.net)), Fossil Record 2 [94], and Google searches on terms “Solanaceae” and “fossil”. The morphology of two fossil specimens was analysed using high-resolution X-ray computed tomography (Table 1) [T. Särkinen, M. Collinson, P. Kenrick, F. Ahmed, unpublished observations]. The morphology of other specimens was evaluated using descriptions and illustrations provided in original publications. The numeric ages for fossils were derived by matching the specific strata from which fossils were found with the most recent geochronological stratigraphy found in the literature (see Additional file 1). The oldest fossil specimens assigned to Solanaceae and the Solanoideae stem nodes were then used as calibration points (see below). The younger age brackets of these oldest specimens were used following best practise guidelines [83].

### Supermatrix construction and analysis

Our supermatrix data harvesting and construction largely followed the modified supermatrix method termed ‘mega-phylogeny’ designed for larger datasets by Smith et al. [95]. The mega-phylogeny method has been designed for large datasets, where maximally dense supermatrices are built based on BLAST searches of all genebank sequences limited to the taxonomic rank of interest [95]. This differs from traditional supermatrix approach where no threshold to missing data or taxa is set, and the resulting sparser matrices are built using clustering techniques.

We looked for all orthologous sequence data available in GenBank release 184 using the PhyLoTA Browser [96]. PhyLoTA identifies available sequence clusters based on BLAST searches, where all sequences for the specified taxonomic group are blasted against each other. We explored all phylogenetically informative sequence clusters identified by PhyLoTA for Solanaceae, and chose seven clusters that had the highest taxon sampling both in terms of genera and species. These seven

clusters included data from two nuclear (*waxy* and ITS) and five plastid regions (*matK*, *ndhF*, *trnS-G*, *trnL-F*, *psbA-trnH*) (Table 2). Gaps in generic sampling were identified and sequences for three previously unsampled genera, *Trianaea*, *Heteranthera*, and *Archihyoscyamus*, were generated for *ndhF*, *trnL-F*, and ITS (Additional file 4). Further sequences were generated for poorly sampled genera (*Reyesia*, *Benthamiella*, *Deprea*, and particular clades of *Solanum*) (Additional file 4). The new sequences were joined with the clusters downloaded from PhyLoTA. Each region was aligned using the profile alignment algorithms Muscle [97] and MAFFT [98,99], after which all datasets were manually checked and adjusted to assure high quality alignments. MAFFT produced better quality alignments compared to Muscle for the most complex alignments (ITS and *waxy*) based on visual comparisons. Short multirepeats and ambiguously alignable regions were excluded. For *trnL-F*, a variable repeat region towards the 5' end of the intergenic spacer was removed; this is where putative pseudogenic copies of *trnF* have been found in *Solanum* [77]. Taxon names were checked for synonymy in all matrices. Duplicate sequences for species were pruned out. *Montinia* (Montiniaceae), *Convolvulus* and *Ipomoea* (Convolvulaceae) were added as outgroups representing two of the closely related families of Solanaceae within the order Solanales [100]. Gene regions were analysed individually using MrBayes v. 3.1.2 [101,102] via the Oslo Bioportal [103] in order to visually check for topological incongruence, rogue taxa, and presence of potentially misidentified sequences.

Ten potentially misidentified sequences were detected in the individual analyses and removed prior to supermatrix construction (Additional file 5). No hard incongruences were detected between the individual matrices with respect to the major clades of the Solanaceae. Incongruence issues were not tested at shallower taxonomic levels due to methodological constraints, and hence individual studies cited in the Background section should be referred to for phylogenetic relationships within genera or major clades in *Solanum*. The software AIR-Appender as implemented in the Oslo BioPortal [103] was used to concatenate the individual matrices. We measured missing data in two ways: missing data per gene region and per species. Missing data for each species was calculated using two measures, missing data and missing information. Missing data was measured as the absolute number of missing base pairs, while missing information was measured as the sum of the parsimony informative characters of missing regions. All species with > 90% missing data and/or information were removed prior to analysis.

Before analysis, the matrix was cleaned by pruning rogue taxa, identified as unstable terminals causing artificial lowering of branch support, using the software RogueNaRok [104]. RogueNaRok analyses were based on

trees derived from fast RAxML bootstrap analyses using a 50% majority-rule consensus threshold and support values for optimization with drop setsize set to one. Four iterations were run and rogue taxa were removed after each iteration. Rapid bootstrap analyses were run in RAxML-VI-HPC v2.0.1 [105,106] via the CIPRES Science Gateway [107] applying partitioning for each gene region using a GTR + CAT approximation rate substitution model and the rapid Bootstrap algorithm with 100 replicates [106]. We removed a total of 85 rogue taxa, some of which had a large amount of missing data and/or information (60-90%), but others with nearly complete sampling. The final matrix included 10,672 bp of aligned sequence data of which 1,902 bp were excluded due to ambiguous alignment (Additional files 6, 7, 8). The matrix included a total of 1,075 Solanaceae species and a single outgroup (*Ipomoea*, Convolvulaceae). We minimized outgroup sampling in order to simplify the BEAST analysis, as the number of outgroups significantly affected run time. The final supermatrix was analysed using RAxML-VI-HPC v2.0.1 [105] via the CIPRES Science Gateway applying partitioning for each gene region using GTR + CAT approximation rate substitution model and the rapid Bootstrap algorithm with 1,000 replicates. The resulting trees were used either as input trees or as starting topologies for dating analyses.

### Molecular dating analyses

The Bayesian uncorrelated relaxed clock-model as implemented in BEAST [108,109] was used as a primary dating method because it allows for rate variation across branches and measures for rate autocorrelation between lineages. Topology and node ages are estimated simultaneously in BEAST, hence topological uncertainty is incorporated into node age estimation. The best tree from the RAxML search was used as a starting topology (Additional file 9). Each region was partitioned separately and given its own substitution model (GTR + G) and rate. A Birth-Death tree prior was used, which accounts for both speciation and extinction [110]. The Solanoideae seed fossils were used to constrain the stem node of Solanoideae with a lognormal offset of 23.0 Ma, mean of 0.01, and standard deviation (SD) of 1.0. The age constraint reflects the youngest age bracket of the oldest known fossil seed assignable to the Solanoideae. Similarly, the Solanaceae stem node was constrained with a lognormal offset of 46.0 Ma, mean of 0.01, and SD of 1.0 based on the youngest age estimate of the oldest fossil specimen of Solanaceae type seeds. Priors for the relaxed clock model mean rate and standard deviation were set to 1.0 and 0.3, respectively, based on known substitution rates in plants. The parameter weights of the delta exchange operator were modified to reflect the length of each partition. Default priors were

used for all other parameters. A total of 100 million generations (10 runs with c. 10 million generations each) were run in BEAST v.1.7.4 [108]. Results were combined using LogCombiner and TreeAnnotator (BEAST package).

A second dating analysis was run using PATHd8 [111]. PATHd8 is a local rate smoothing method that estimates node ages by calculating mean path lengths from the node to the tips. Deviations from a strict molecular clock are corrected as suggested by the calibrated nodes. Only simple calibrations are allowed as point estimates of minimum, maximum or mean ages. Because substitution rates are smoothed locally, rather than simultaneously over the whole tree, PATHd8 allows analysis of very large trees. The best tree from the RAxML search was used as the input phylogeny for the PATHd8 analysis (Additional file 10). The stem node of Solanoideae was constrained with the identified Solanoideae seed fossils with minimum age of 23.0 Ma. PATHd8 requires a minimum of one fixed node constraint, and hence the stem node of the family was constrained with a fixed age of 46.0 Ma. Results from both the Maximum Likelihood and Bayesian dating analyses have been deposited in TreeBase (<http://purl.org/phylo/treebase/phylovs/study/TB2:S14458>).

### Additional files

**Additional file 1: Solanaceae fossil record.** Details of all records of Solanaceae with full references to primary sources.

**Additional file 2: Solanaceae time-calibrated phylogeny with tips.** A detailed dated phylogeny of Solanaceae showing mean node ages and 95% confidence intervals for all nodes. Posterior probability branch support values are indicated in branch colours, where red refers to nodes with < 80% support. Major clades are indicated, and studies which include more detailed phylogenies of the particular groups are indicated on the left. These studies should be referred to as primary phylogenetic sources for the particular clades with more up-to-date details of species-level relationships because the individual studies used more markers and discuss specific issues relevant at such low taxonomic levels, including polyploidy, hybridisation, and gene tree – species tree incongruences.

**Additional file 3: Solanaceae time-calibrated phylogeny.** Dated molecular phylogeny of the Solanaceae based on the supermatrix calibrated using fossil data. Major clades are shown with their associated ages and 95% confidence intervals. Thick branches indicate highly supported clades with > 0.9 posterior probability. Clade size is proportional to the number of species sampled in each clade. Associated floral and fruit forms are shown on the right. A more detailed view of this phylogeny is shown in Additional file 2.

**Additional file 4: New sequence data.** Voucher data and GenBank numbers for sequences newly generated as part of the study.

**Additional file 5: New sequence data.** Details of sequences downloaded from GenBank which appeared clearly misidentified or potentially contaminated sequences based on BLAST searches and their position in our preliminary Maximum Likelihood phylogenies.

**Additional file 6: Supermatrix sequence alignment file (RAxML input).** Final supermatrix alignment file in phylip format used as input for the Maximum Likelihood phylogeny estimation in RAxML.

**Additional file 7: Supermatrix partitions (RAxML input).** Details of the partitioning scheme used in the supermatrix for the Maximum Likelihood analysis in RAxML.



**Additional file 8: Supermatrix excluded regions (RAxML input).**

Details of the excluded regions within the supermatrix alignment file in RAxML.

**Additional file 9: BEAST xml input file.** Input file used for Bayesian relaxed molecular clock dating analysis in BEAST.

**Additional file 10: PATHd8 input file.** Input file used for local clock model dating analysis in PATHd8.

**Competing interests**

The authors declare that they have no competing interests.

**Authors' contributions**

TS carried out the molecular sequence alignments and analyses, and the fossil review. LB provided sequence data for missing and poorly sampled genera. RO participated in the sequence alignment. SK participated in the design and coordination of the study. All authors contributed to writing, and read and approved the final manuscript.

**Acknowledgements**

We thank Cody Hinchliff and Chris Barton for helpful guidance on mega-phylogeny methods, Mario dos Reis for advice on Bayesian dating methods, Farah Ahmed, Peta Hayes and Paul Kenrick for assistance in fossil review, and Peter Foster for assistance in running analyses. This work is supported by the National Science Foundation (NSF) grant (to LB & SK) 'PBI *Solanum* – a world treatment' DEB-0316614.

**Author details**

<sup>1</sup>Department of Life Sciences, Natural History Museum, Cromwell Road, London SW7 5BD, UK. <sup>2</sup>Royal Botanic Garden Edinburgh, 20A Inverleith Row, Edinburgh EH3 5LR, UK. <sup>3</sup>Department of Biology, University of Utah, 257 S 1400 E, Salt Lake City, UT 84112, USA. <sup>4</sup>Department of Biology, University of Washington, Seattle, WA 98195, USA.

Received: 28 March 2013 Accepted: 16 September 2013

Published: 30 September 2013

**References**

1. The Potato Genome Sequencing Consortium: **Genome sequence and analysis of the tuber crop potato.** *Nature* 2011, **475**:189–195.
2. The Tomato Genome Consortium: **The tomato genome sequence provides insights into fleshy fruit evolution.** *Nature* 2012, **485**:635–641.
3. Bombarely A, Rosli HG, Vrebalov J, Moffett P, Mueller LA, Martin GB: **A draft genome sequence of *Nicotiana benthamiana* to enhance molecular plant-microbe biology research.** *Mol Plant Microbe Interact* 2012, **25**:1523–1530.
4. Wu F, Tanksley SD: **Chromosomal evolution in the plant family Solanaceae.** *BMC Genomics* 2010, **11**:182.
5. Doganlar S, Fray A, Daunay M-C, Lester RN, Tanksley SD: **A comparative genetic linkage map of eggplant (*Solanum melongena*) and its implications for genome evolution in the Solanaceae.** *Genetics* 2002, **161**:1697–1711.
6. Wang Y, Diehl A, Wu F, Vrebalov J, Giovannoni J, Siepel A, Tanksley SD: **Sequencing and comparative analysis of a conserved syntenic segment in the Solanaceae.** *Genetics* 2008, **180**:391–408.
7. Robertson KA, Goldberg EE, Igic B: **Comparative evidence for the correlated evolution of polyploidy and self-compatibility in Solanaceae.** *Evolution* 2011, **65**:139–155.
8. Goldberg EE, Kohn JR, Lande R, Robertson KA, Smith SA, Igic B: **Species selection maintains self-incompatibility.** *Science* 2010, **330**:459–460.
9. Igic B, Bohs L, Kohn JR: **Ancient polymorphism reveals unidirectional breeding system shifts.** *Proc Natl Acad Sci USA* 2006, **103**:1359–1363.
10. Knapp S: **Tobacco and tomatoes: a phylogenetic perspective on fruit diversity in the Solanaceae.** *J Exp Bot* 2002, **53**:2001–2022.
11. Knapp S: **On 'various contrivances': pollination, phylogeny and flower form in the Solanaceae.** *Phil Trans R Soc B* 2010, **365**:449–460.
12. Rodriguez GR, Munos S, Anderson C, Sim S-C, Michel A, Causse M, McSpadden BB, Francis D, van der Knaap E: **Distribution of SUN, OVATE, LC, and FAS in the tomato germplasm and the relationship to fruit shape diversity.** *Plant Physiol* 2011, **156**:275–285.

13. Van der Knaap E, Tanksley SD: **The making of a bell pepper-shaped tomato fruit: identification of loci controlling fruit morphology in Yellow Stuffer tomato.** *Theor Appl Genet* 2003, **107**:139–147.
14. Lippman ZB, Cohen O, Alvarez JP, Abu-Abied M, Pekker I, Paran I, Eshed Y, Zamir D: **The making of a compound inflorescence in tomato and related nightshades.** *PLoS Biol* 2008, **6**(Pekker I):e288.
15. Chitwood DH, Headland LR, Filiault DL, Kumar R, Jimenez-Gomez JM, Schrage AV, Park DS, Peng J, Sinha NR, Maloof JN: **Native environment modulates leaf size and response to simulated foliar shade across wild tomato species.** *PLoS One* 2012, **7**:e29570.
16. Olmstead RG: **Phylogeny and biogeography in Solanaceae, Verbenaceae and Bignoniaceae: a comparison of continental and intercontinental diversification patterns.** *Bot J Linn Soc* 2013, **171**:80–102.
17. Chitwood DH, Headland LR, Ranjan A, Martinez CC, Braybrook SA, Koenig DP, Kuhlmeier C, Smith RS, Sinha NR: **Leaf asymmetry as a developmental constraint imposed by auxin-dependent phyllotactic patterning.** *Plant Cell* 2012, **24**:2318–2327.
18. Kallenbach M, Bonaventure G, Gilardoni PA, Wissgott A, Baldwin IT: ***Empoasca* leafhoppers attack wild tobacco plants in a jasmonate-dependent manner and identify jasmonate mutants in natural populations.** *Proc Natl Acad Sci U S A* 2012, **109**:E1548–E1557.
19. Weinhold A, Baldwin IT: **Trichome-derived O-acetyl sugars are a first meal for caterpillars that tags them for predation.** *Proc Natl Acad Sci USA* 2011, **108**:7855–7859.
20. Schmidt DD, Kessler A, Kessler D, Schmidt S, Lim M, Gase K, Baldwin IT: ***Solanum nigrum*: A model ecological expression system and its tools.** *Mol Ecol* 2004, **13**:981–995.
21. Olmstead RG, Bohs L: **A summary of molecular systematic research in Solanaceae: 1982–2006.** In *6th International Solanaceae Conference*. Edited by Spooner DB, Bohs L, Giovannoni J, Olmstead RG, Shibata D. Acta Horticulturae 745; 2007:255–268.
22. Olmstead RG, Bohs L, Migid HA, Santiago-Valentin E, Garcia VF, Collier SM: **A molecular phylogeny of the Solanaceae.** *Taxon* 2008, **57**:1159–1181.
23. Garcia VF, Olmstead RG: **Phylogenetics of Tribe Anthocercideae (Solanaceae) based on *ndhF* and *trnL/F* sequence data.** *Syst Bot* 2003, **28**:609–615.
24. Clarkson JJ, Knapp S, Aoki S, Garcia VG, Olmstead RG, Chase MW: **Phylogenetic relationships in *Nicotiana* (Solanaceae) inferred from multiple plastid DNA regions.** *Mol Phylogenet Evol* 2004, **33**:75–90.
25. Yuan Y-W, Zhang Z-Y, Chen ZD, Olmstead RG: **Tracking ancient polyploids: A retroposon reveals an extinct diploid ancestor in the polyploid ancestry of *Belladonna*.** *Mol Biol Evol* 2006, **23**:2263–2267.
26. Levin RA, Bernardello G, Whiting C, Miller JS: **A new generic circumscription in tribe Lycieae (Solanaceae).** *Taxon* 2011, **60**:681–690.
27. Chen S, Matsubara K, Omori T, Kokubun H, Kodama H, Watanabe H, Hashimoto G, Marchesi E, Bullrich L, Ando T: **Phylogenetic analysis of the genus *Petunia* (Solanaceae) based on the sequence of the *Hf1* gene.** *J Plant Res* 2007, **120**:385–397.
28. Miller RJ, Mione T, Phan H-L, Olmstead RG: **Color by numbers: Nuclear gene phylogeny of *Jaltomata* (Solanaceae), sister genus to *Solanum*, supports three clades differing in fruit color.** *Syst Bot* 2011, **36**:153–162.
29. Tu T, Dillon MO, Sun H, Wen J: **Phylogeny of *Nolana* (Solanaceae) of the Atacama and Peruvian deserts inferred from sequences of four plastid markers and the nuclear *LEAFY* second intron.** *Mol Phylogenet Evol* 2008, **49**:561–573.
30. Tate JA, Acosta MC, McDill J, Moscone EA, Simpson BB, Cocucci AA: **Phylogeny and character evolution in *Nierembergia* (Solanaceae): Molecular, morphological, and cytogenetic evidence.** *Syst Bot* 2009, **34**:198–206.
31. Filipowicz N, Renner SS: ***Brunfelsia* (Solanaceae): A genus evenly divided between South America and radiations on Cuba and other Antillean islands.** *Mol Phylogenet Evol* 2012, **64**:1–11.
32. Fregonezi JN, Freitas LB, Bonatto SL, Semir J, Stehmann JR: **Infrageneric classification of *Calibrachoa* (Solanaceae) based on morphological and molecular evidence.** *Taxon* 2012, **61**:120–130.
33. Fregonezi JN, Turchetto C, Bonatto SL, Freitas LB: **Biogeographical history and diversification of *Petunia* and *Calibrachoa* (Solanaceae) in the Neotropical pampas grassland.** *Bot J Linn Soc* 2013, **71**:140–153.
34. Poccai P, Hyvönen J, Symon DE: **Phylogeny of kangaroo apples (*Solanum* subg. *Archaeosolanum*, Solanaceae).** *Mol Biol Rep* 2011, **38**:5243–5259.
35. Bohs L: **Phylogeny of the *Cyphomandra* clade of the genus *Solanum* (Solanaceae) based on ITS sequence data.** *Taxon* 2007, **56**:1012–1026.



36. Stern S, Bohs L: An explosive innovation: Phylogenetic relationships of *Solanum* section *Gonatotrichum* (Solanaceae). *PhytoKeys* 2012, **8**:83–98.
37. Fajardo D, Spooner DM: Phylogenetic relationships of *Solanum* series *Conicibaccata* and related species in *Solanum* section *Petota* inferred from five conserved ortholog sequences. *Syst Bot* 2011, **36**:163–170.
38. Ames M, Spooner DM: Phylogeny of *Solanum* series *Piurana* and related species in *Solanum* section *Petota* based on five conserved ortholog sequences. *Taxon* 2010, **59**:1091–1101.
39. Rodriguez F, Spooner DM: Nitrate reductase phylogeny of potato (*Solanum* sect. *Petota*) genomes with emphasis on the origins of the polyploid species. *Syst Bot* 2009, **34**:207–219.
40. Tepe E, Farruggia FT, Bohs L: A 10-gene phylogeny of *Solanum* section *Herpystichum* (Solanaceae) and a comparison of phylogenetic methods. *Am J Bot* 2011, **98**:1356–1365.
41. Bohs L, Olmstead RG: A reassessment of *Normania* and *Triguera* (Solanaceae). *Plant Systemat Evol* 2001, **228**:33–48.
42. Spooner DM, Anderson GJ, Jansen RK: Chloroplast DNA evidence for the interrelationships of tomatoes, potatoes, and pepinos (Solanaceae). *Am J Bot* 1993, **80**:676–688.
43. Weese TL, Bohs L: A three-gene phylogeny of the genus *Solanum* (Solanaceae). *Syst Bot* 2007, **32**:445–463.
44. Quental TB, Marshall CR: Diversity dynamics: molecular phylogenies need the fossil record. *Trends Ecol Evol* 2010, **35**:434–441.
45. Morlon H, Parsons TD, Plotkin JB: Reconciling molecular phylogenies with the fossil record. *Proc Natl Acad Sci U S A* 2011, **108**:16327–16332.
46. Dillon MO, Tu T, Xie L, Quipuscoa Silvestre V, Wen J: Biogeographic diversification in *Nolana* (Solanaceae), a ubiquitous member of the Atacama and Peruvian Deserts along the western coast of South America. *J Syst Evol* 2009, **47**:457–476.
47. Tu T, Volis S, Dillon MO, Sun H, Wen J: Dispersals of Hyoscyameae and Mandragoreae (Solanaceae) from the New World to Eurasia in the early Miocene and their biogeographic diversification within Eurasia. *Mol Phylogenet Evol* 2010, **57**:1226–1237.
48. Martínez-Millán M: Fossil record and age of the Asteridae. *Bot Rev* 2010, **76**:83–135.
49. Franco MJ, Brea M: Leños fósiles de la formación Paraná (Mioceno Medio), Toma Vieja, Paraná, Entre Ríos, Argentina: registro de bosques estacionales mixtos. *Ameghiniana* 2008, **45**:699–717.
50. Page VM: Dicotyledonous wood from the Upper Cretaceous of central California II. *J Arnold Arbor* 1980, **61**:723–748.
51. Erdtman G: *Pollen morphology and plant taxonomy – Angiosperms*. Stockholm: Almqvist & Wiksell; 1952.
52. Bernardello L, Lujan MC: Pollen morphology of tribe Lycieae: *Grabowskia*, *Lycium*, *Phrodus* (Solanaceae). *Rev Palaeobot Palynol* 1997, **96**:305–315.
53. Fukuda T, Naiki A, Nagamasu H: Pollen morphology of the genus *Skimmia* (Rutaceae) and its taxonomic implications. *J Plant Res* 2008, **121**:463–471.
54. Hunziker AT: *Genera Solanacearum: The genera of Solanaceae illustrated, arranged according to a new system*. Ruggell, Liechtenstein: Gantner; 2001.
55. Chandler MEJ: *The Lower Tertiary Floras of Southern England II Flora of the Pipe-Clay Series of Dorset (Lower Bagshot)*. London: British Museum (Natural History); 1962.
56. Chandler MEJ: The Oligocene Flora of the Bovey Tracey Lake Basin, Devonshire. *Bull Brit Mus (Nat Hist) Geol* 1957, **3**:71–123.
57. Reid C, Chandler MEJ: *The London clay flora*. London: British Museum (Natural History); 1933.
58. Smith S, Baum DA: Phylogenetics of the florally diverse Andean clade *lochrominae* (Solanaceae). *Am J Bot* 2006, **93**:1140–1153.
59. Whitson M, Manos PS: Untangling *Physalis* (Solanaceae) from the Physaloids: A two-gene phylogeny of the Physalidinae. *Syst Bot* 2005, **30**:216–230.
60. Averett JE: *Schraderanthus*, a new genus of Solanaceae. *Phytologia* 2009, **91**:54–61.
61. Knapp S: A revision of the Dulcamaroid clade of *Solanum* L. (Solanaceae). *PhytoKeys* 2013, **22**:1–428.
62. Bennett JR: Revision of *Solanum* section *Regmandra* (Solanaceae). *Edinb J Bot* 2008, **65**:69–112.
63. Peralta IE, Spooner DM, Knapp S: Taxonomy of wild tomatoes and their relatives (*Solanum* sect. *Lycopersicoides*, sect. *Juglandifolia*, sect. *Lycopersicon*; Solanaceae). *Syst Bot Monogr* 2008, **84**:1–186.
64. Stern S, de Fatima Agra M, Bohs L: Molecular delimitation of clades within New World species of “spiny solanums” (*Solanum* subg. *Leptostemonum*). *Taxon* 2011, **60**:1429–1441.
65. dos Reis M, Inoue J, Hasegawa M, Asher RJ, Donoghue PCJ, Yang Z: Phylogenomic datasets provide both precision and accuracy in estimating the timescale of placental mammal phylogeny. *Proc R Soc B* 2012, **279**:3491–3500.
66. Philippe H, Brinkmann H, Lavrov D, Littlewood DTJ, Manuel M, Wörheide G, Baurain D: Resolving difficult phylogenetic questions: Why more sequences are not enough. *PLoS Biol* 2011, **9**:e1000602.
67. Driskell A, Ane C, Burleigh JG, McMahon MM, O’Meara BC, Sanderson MJ: Prospects for building the Tree of Life from large sequence databases. *Science* 2004, **306**:1172–1174.
68. Heath TA, Hedtke SM, Hillis DM: Taxon sampling and the accuracy of phylogenetic analyses. *J Syst Evol* 2008, **46**:239–257.
69. The Legume Phylogeny Working Group: Legume phylogeny and classification in the 21st century: Progress, prospects and lessons for other species-rich clades. *Taxon* 2013, **62**:217–248.
70. Hinchliff CE, Roalson EH: Using supermatrices for phylogenetic inquiry: an example using the sedges. *Syst Biol* 2013, **62**:205–219.
71. Baker WJ, Savolainen V, Asmussen-Lange CB, Chase MW, Dransfield J, Forest F, Harley MM, Uhl NW, Wilkinson M: Complete generic-level phylogenetic analyses of palms (Arecaceae) with comparisons of supertree and supermatrix approaches. *Syst Biol* 2009, **58**:240–256.
72. Averett JE, Martínez M: *Capsicophysalis*: a new genus of Solanaceae (Physaleae) from Mexico and Central America. *J Bot Res Inst Texas* 2009, **3**:71–75.
73. Levin RA, Whelan A, Miller JS: The utility of nuclear conserved ortholog set II (COSII) genomic regions for species-level phylogenetic inference in *Lycium* (Solanaceae). *Mol Phylogenet Evol* 2009, **53**:881–890.
74. Tepe EJ, Bohs L: A molecular phylogeny of *Solanum* sect. *Pteroides* (Solanaceae) and the utility of COSII markers in resolving relationships among closely related species. *Taxon* 2010, **59**:733–743.
75. Yuan Y-W, Liu C, Marx HE, Olmstead RG: An empirical demonstration of using pentatricopeptide repeat (PPR) genes as plant phylogenetic tools: Phylogeny of Verbanaceae and the Verbena complex. *Mol Phylogenet Evol* 2010, **54**:23–35.
76. Yuan Y-W, Liu C, Marx HE, Olmstead RG: The pentatricopeptide repeat (PPR) gene family, a tremendous resource for plant phylogenetic studies. *New Phytol* 2009, **182**:272–283.
77. Poccai P, Hyvönen J: Identification and characterization of plastid *trnF* (GAA) pseudogenes in four species of *Solanum* (Solanaceae). *Biotechnol Lett* 2011, **33**:2317–2323.
78. Couvreur TLP, Forest F, Baker WJ: Origin and global diversification patterns of tropical rain forests: inferences from a complete genus-level phylogeny of palms. *BMC Biol* 2011, **9**:44.
79. Bininda-Emonds ORP, Cardillo M, Jones KE, MacPhee RDE, Beck RMD, Grenyer R, Price SA, Vos RA, Gittleman JL, Purvis A: The delayed rise of present-day mammals. *Nature* 2007, **446**:507–512.
80. Hunt T, Bergsten J, Levkancova Z, Papadopoulou A, St John O, et al: A comprehensive phylogeny of beetles reveals the evolutionary origins of a superradiation. *Science* 2007, **318**:1913–1916.
81. Smith SA, Donoghue MJ: Rates of molecular evolution are linked to life history in flowering plants. *Science* 2008, **322**:86–89.
82. Andreassen K, Baldwin BG: Unequal evolutionary rates between annual and perennial lineages of checker mallows (*Sidalcea*, Malvaceae): evidence from 18S-26S rDNA internal and external transcribed spacers. *Mol Biol Evol* 2001, **18**:936–944.
83. Parham JF, Donoghue PCJ, Bell CJ, Calway TD, Head JJ, et al: Best practices for justifying fossil calibrations. *Syst Biol* 2012, **61**:346–359.
84. Olmstead RG, Palmer JD: A chloroplast DNA phylogeny of the Solanaceae: subfamilial relationships and character evolution. *Ann Mo Bot Gard* 1992, **79**:346–360.
85. Ronquist F, Klopfstein S, Vilhelmsen L, Schulmeister S, Murray D, et al: A total-evidence approach to dating with fossils, applied to the early radiation of the Hymenoptera. *Syst Biol* 2012, **61**:973–999.
86. Near TJ, Bolnick DI, Wainwright PC: Fossil calibrations and molecular divergence time estimates in centrarchid fishes (Teleostei: Centrarchidae). *Evolution* 2005, **59**:1768–1782.
87. Wikström N, Savolainen V, Chase MW: Evolution of the angiosperms: calibrating the family tree. *Proc R Soc Lond B* 2001, **268**:2211–2220.
88. Bremer K, Friis EM, Bremer B: Molecular phylogenetic dating of Asterid flowering plants shows early Cretaceous diversification. *Syst Biol* 2004, **53**:496–505.

89. Bell CD, Soltis DE, Soltis PS: **The age and diversification of the angiosperms re-revisited.** *Am J Bot* 2010, **97**:1296–1303.
90. Paape T, Igic B, Smith SD, Olmstead R, Bohs L, *et al*: **A 15-Myr-old genetic bottleneck.** *Mol Biol Evol* 2008, **25**:655–663.
91. Blanc G, Wolfe KH: **Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes.** *Plant Cell* 2004, **16**:1667–1678.
92. Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR: **Long-term diploidization in allopolyploid *Nicotiana* section *Repandae*.** *New Phytol* 2005, **168**:214–252.
93. Ladiges PY, Marks CE, Nelson G: **Biogeography of *Nicotiana* section *Suaevolentes* (Solanaceae) reveals geographical tracks in arid Australia.** *J Biogeogr* 2011, **38**:2066–2077.
94. Benton MJ: *The fossil record 2*. London: Chapman & Hall; 1993.
95. Smith SA, Beaulieu JM, Donoghue MJ: **Mega-phylogeny approach for comparative biology: an alternative to supertree and supermatrix approaches.** *BMC Evol Biol* 2009, **9**:37.
96. Sanderson MJ, Boss D, Chen D, Cranston KA, Wehe A: **The PhyLoTA Browser: processing GenBank for molecular phylogenetics research.** *Syst Biol* 2008, **57**:335–346.
97. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32**:1792–1797.
98. Katoh K, Kuma K, Toh H, Miyata T: **MAFFT version 5: improvement in accuracy of multiple sequence alignment.** *Nucleic Acids Res* 2005, **33**:511–518.
99. Katoh K, Toh H: **Recent developments in the MAFFT multiple sequence alignment program.** *Brief Bioinform* 2008, **9**:286–298.
100. The Angiosperm Phylogeny Group: **An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG III.** *Bot J Linn Soc* 2009, **161**:105–121.
101. Huelsenbeck JP, Ronquist F: **MRBAYES: Bayesian inference of phylogeny.** *Bioinformatics* 2003, **17**:754–755.
102. Ronquist F, Huelsenbeck JP: **MRBAYES 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2005, **19**:1572–1574.
103. Kumar S, Skjæveland Å, Orr R, Enger P, Ruden T, Mevik B-H, Burki F, Botnen A, Shalchian-Tabrizi K: **AIR: A batch-oriented web program package for construction of supermatrices ready for phylogenomic analyses.** *BMC Bioinforma* 2009, **10**:357.
104. Aberer AJ, Krompass D, Stamatakis A: **Pruning rogue taxa improves phylogenetic accuracy: An efficient algorithm and webservice.** *Syst Biol* 2013, **62**:162–166.
105. Stamatakis A: **RAXML-VI-HPC: maximum likelihood-1 based phylogenetic analyses with thousands of taxa and mixed models.** *Bioinformatics* 2006, **22**:2688–2690.
106. Stamatakis A, Hoover P, Rougemont J: **A rapid bootstrap algorithm for the RAXML Web servers.** *Syst Biol* 2008, **57**:758–771.
107. Miller MA, Pfeiffer W, Schwartz T: *Creating the CIPRES Science Gateway for inference of large phylogenetic trees.* New Orleans, LA: Proceedings of the Gateway Computing Environments Workshop (GCE); 2010:1–8.
108. Drummond AJ, Rambaut A: **BEAST: Bayesian evolutionary analysis by sampling trees.** *BMC Evol Biol* 2007, **7**:214.
109. Drummond AJ, Ho SWY, Phillips MJ, Rambaut A: **Relaxed phylogenetics and dating with confidence.** *PLoS Biol* 2006, **4**:e88.
110. Gernhard T: **The conditioned reconstructed process.** *J Theor Biol* 2008, **253**:769–778.
111. Britton T, Anderson CL, Jacquet D, Lundqvist S, Bremer K: **Estimating divergence times in large phylogenetic trees.** *Syst Biol* 2007, **56**:741–752.

doi:10.1186/1471-2148-13-214

**Cite this article as:** Särkinen *et al.*: A phylogenetic framework for evolutionary study of the nightshades (Solanaceae): a dated 1000-tip tree. *BMC Evolutionary Biology* 2013 **13**:214.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
[www.biomedcentral.com/submit](http://www.biomedcentral.com/submit)

