The Origin of Oranges: A Multi-Locus Phylogeny of Rutaceae Subfamily Aurantioideae

Author(s): Thomas Schwartz, Stephan Nylinder, Chandrika Ramadugu, Alexandre Antonelli, and Bernard E. Pfeil

Source: Systematic Botany, 40(4):1053-1062.
Published By: The American Society of Plant Taxonomists
URL: http://www.bioone.org/doi/full/10.1600/036364415X690067

Systematic Botany (2015), 40(4): pp. 1053–1062

DOI 10.1600/036364415X690067
Date of publication December 15, 2015

The Origin of Oranges: a Multi-locus Phylogeny of Rutaceae Subfamily Aurantioideae

Thomas Schwartz,[1] Stephan Nylinder,[2] Chandrika Ramadugu,[3] Alexandre Antonelli,[1,4] and Bernard E. Pfeil[1,5,6]

[1]Department of Biological and Environmental Sciences, University of Gothenburg, Box 461, Gothenburg 405 30, Sweden.
[2]Swedish Museum of Natural History, Stockholm, Sweden.
[3]Department of Botany and Plant Sciences, University of California, Riverside, California, U. S. A.
[4]Gothenburg Botanical Garden, Gothenburg, Sweden.
[5]Commonwealth Scientific and Industrial Research Organisation Plant Industry, Canberra, Australia.
[6]Author for correspondence (bernard.pfeil@gu.se)

Communicating Editor: James Smith

Abstract—The phylogeny of Rutaceae subfamily Aurantioideae has previously been estimated only using plastid and repetitive nuclear sequences. We added sequences of two low copy nuclear loci to allow further diagnosis of phenomena that may mislead phylogenetic inference. After testing for patterns expected under recombination, positive selection, and hybridization, we excluded data sets or sequences accordingly and then inferred the species tree using the multispecies coalescent. We then reconstructed the ancestral area using parsimony and the dispersal-extinction-cladogenesis model to test the hypothesis that *Citrus* s. l. may have originated in Australasia and migrated or rafted to Eastern Asia. The ancestral area of *Citrus* s. l. inferred under either method and several models was west of Wallace's line. Therefore, *Citrus* s. l. did not appear to have rafted west on what became the Halmahera Islands (Indonesia). Our findings are also consistent with previously reported ages for the origin of this group that may be too young to have allowed this rafting. The species tree is well resolved and largely consistent with previous molecular phylogenies, especially those using chloroplast sequences.

Keywords—Ancestral area, biogeography, *Citrus*, Malesia, molecular phylogeny, species phylogeny.

Rutaceae subfamily Aurantioideae includes important crops such as oranges, lemons, limes (*Citrus* L.), other well-known plants such as the Curry Tree (*Bergera koenigii* L.), used in cooking, and the ornamental plant Orange Jessamine (*Murraya paniculata* (L.) Jack). The subfamily's center of diversity is in the monsoon region of eastern Australasia, extending west through South Asia into Africa, and eastwards into Polynesia. The first systematic treatment of the family was made by Engler (1897: his table 1, last column), who divided it into six subfamilies: Aurantioideae, Dictyolomoideae, Flindersioideae, Rutoideae, Spathelioideae, and Toddalioideae. Newly discovered or recognized genera have since been added to Engler's Aurantioideae, or other genera have been treated at different ranks, but by and large his original concept of the subfamily remains. Swingle and Reece's (1967) final treatment of the subfamily recognized two tribes, six subtribes, 33 genera, and over 200 species. This circumscription has been widely followed since, although some genera are being treated informally under synonymy as a result of recent molecular evidence (e.g. the Germplasm Resources Information Network, GRIN: www.ars-grin.gov).

Early molecular phylogenetic inferences have been in agreement with the classification of these plants as a natural group (Chase et al. 1999). However, later studies have suggested some discrepancies with Swingle's scheme with respect to the intra-sub-familial classification, e.g. that the affinities of *Feroniella* and *Oxanthera* lie with subtribe Citrinae rather than with subtribes Balsamocitrinae or Triphasiinae, respectively (Bayer et al. 2009), and that *Bergera koenigii* (among others) should not be placed within *Murraya* J. Koenig ex L. (Samuel et al. 2001). Thus far, although the sampling of genera in molecular studies has been nearly complete (e.g. 97% in Bayer et al. 2009), the sampling of linkage groups has been limited. Apart from chloroplast (cp) DNA regions, the only additional linkage group has come from the internal transcribed spacer (ITS) nuclear (n) DNA region (Morton 2009), a multi-copy repeat that generally evolves with concerted evolution homogenizing the repeats to a single sequence type.

Phylogenies based on a single linkage group can be misleading (i.e. not track the species phylogeny) because of positive selection (e.g. Stefanović et al. 2009), mistaken orthology (e.g. Straub et al. 2006), lineage sorting (e.g. Syring et al. 2007), hybridization (e.g. Maureira-Butler et al. 2008), and recombination (e.g. Poke et al. 2006). Although ITS has been examined for Aurantioideae, it comes with its own challenges (Álvarez and Wendel 2003), including the retention of multiple (sometimes deeply coalescing) sequence types if homogenization across the multiple copies is incomplete (Suh et al. 1993). This may be difficult to recover in direct sequencing (Rauscher et al. 2002). Even if homogenization is complete, differential homogenization to each parental lineage can obscure the true relationships in species groups with a common hybrid origin (Wendel et al. 1995).

Generating sequence data from more than one linkage group is an important addition to improve species tree inference (Edwards et al. 2007). However, analyses relying on concatenation of sequence data can be misleading when single linkage group gene trees differ due to incomplete lineage sorting (Kubatko and Degnan 2007). Instead, the coalescent model may provide a more accurate means of producing phylogenetic inferences of species, as opposed to genes (Degnan and Rosenberg 2009). Therefore, low-copy nuclear genes (those perhaps least likely to evolve concertedly) may provide the best source of additional genes not linked to the chloroplast genome to complement existing phylogenetic evidence in Rutaceae. However, there is no replacement for a thorough examination of potential sources of incongruence at each step in an analysis.

Previous hypotheses regarding the origin of *Citrus* s. l. proposed that the group originated in the southeastern part of the current distribution (i.e. eastern Australia) (Swingle and Reece 1967). The group is thought to have then diversified and migrated from there, eventually producing the more palatable species found in the northwestern part of the distribution (Swingle and Reece 1967). Beattie et al. (2006), following Swingle and Reece's (1967) hypothesis, proposed migration

from Australasia to East Asia via the movement of terranes, notably the Halmahera Islands and the eastern Philippines, beginning 37 million years ago (mya).

Around 40 mya, the Australian plate was separated by deep ocean from the Sunda shelf by around 10° latitude, but was rapidly moving northwards (Hall 1998; Lohman et al. 2011). The submerged Sunda shelf margin and the Australian plate eventually came close enough to connect c. 23 mya, beginning the northward rotation of the Sunda region (Lohman et al. 2011). As the Australian plate continued to move northwards, more and more islands and shallow seas were created between the Sunda shelf and the Australian plate as the latter was subducted, forming the island-rich zone between these plates known as Wallacea (Lohman et al. 2011). However, no dry land connection has yet been established between Asia and Australia (Lohman et al. 2011), with the best land and shallow sea connections, based on interpreting reconstructions (Hall 1998), probably existing from 7 mya to the present day.

Halmahera Island, lying in the northeastern part of Wallacea and northwest of New Guinea, is composed of young (western part) and older (eastern part) volcanic rocks, overlain by Miocene carbonates in the eastern part (Hall 1998; Jong 1998; Hall 2002). The parts of the island are thought to have moved from far to the east (lying north east of central New Guinea), past the Bird's Head between 4–8 mya, continuing westward to their current location, with most movement occurring from the early Miocene, about 20–25 mya (Hall 1998; Jong 1998; Hall 2002). Halmahera has, for most of its history, been closer to elements of New Guinea and the eastern Philippines than to other parts of Wallacea, where it is found today (Jong 1998). If *Citrus* s. l. originated in eastern Australia/New Guinea, it could perhaps have used Halmahera as a vehicle to move partway to Asia, by rafting on this land mass as it rapidly shuttled from east to west across the top of New Guinea, subsequently dispersing from there, in accordance with Beattie et al.'s (2006) hypothesis. Alternatively, the volcanic arc and shallows seas that connected Halmahera, the Philippines, and northeastern Sulawesi around 25 mya (Hall 1998) may have allowed stepping stone dispersal from north-eastern New Guinea to central Wallacea and then to mainland Asia to occur. According to Hall's (1998) reconstructions, there were probably fewer opportunities to disperse along this route after that time, because fewer volcanic islands connected the route and because Halmahera was often submerged. In either scenario, an ancestral *Citrus* species would need to have been present in north-eastern New Guinea around 20–25 mya in order to use either opportunity. Alternatively, a southern route, via Sumatra, Java, and the Inner or Outer Banda Arcs to New Guinea may have been increasingly feasible from about 7 mya to the present day, based on Hall's reconstructions (Hall 1998, 2002).

Species trees rather than gene trees are required to test hypotheses regarding the properties of species, such as their biogeographic history. However, no credible species tree has yet been proposed for Aurantioid genera that avoids the potential problems raised above regarding single linkage groups or data concatenation. Our aims here are to (1) add new information from low-copy nDNA markers, (2) thoroughly explore and analyze any cases of incongruence, and (3) infer a species tree using a coalescent model that improves on earlier estimations of both topology and clade ages. With this species tree we infer ancestral areas of distribution and examine the results in light of the previous hypotheses for the origin of *Citrus*.

Materials and methods

***Taxon Sampling, Extraction, Amplification, and Sequencing*—**The data used in this study included sequences from four genetic regions ("genes") analyzed individually, with three of them analyzed together in a multispecies coalescent framework. We initially collected data from three different linkage groups: (1) the cpDNA alignment of 88 accessions for > 10 kb previously generated (Bayer et al. 2009), downloaded from TreeBASE (www.treebase.org), of which we used 47 sequences from 25 Aurantioid genera (excluding artificial hybrids and retaining only a few representatives of the heavily sampled Asian species of *Citrus* to maximize overlap with the other genes); (2) part of nDNA malate dehydrogenase (*mdh*) for 24 Aurantioid genera (58 alleles), and (3) part of nDNA beta-hydroxylase (*hyB*) for 15 Aurantioid genera (60 alleles), much of the latter two data sets comprised sequences generated as part of this study. Three outgroup sequences (from three genera) were also used for the cpDNA and *mdh* alignments (Appendix 1). Considering the taxonomy used by GRIN with generic synonymies proposed by Mabberley (1998) and Bayer et al. (2009), which we accept although we use the earlier names here to clarify the original concepts included in our sample, the generic coverage of the subfamily represented by this sample is high (21 of 26 genera). Sequences from 56 individuals are present in two of three data sets; sequences from 16 individuals are present in all three. The two nDNA regions are present in single copy in two *Citrus reticulata* Blanco draft genomes and are found in different assembly scaffolds (http://www.phytozome.net/citrus.php but listed there as *C. clementina* hort.), therefore, we assume that they are unlinked in *Citrus* and likely to be unlinked in other Aurantioid genera. BLASTn searches (http://www.ncbi.nlm.nih.gov) using representative chloroplast sequences to the "*Citrus sinensis*" chloroplast genome produced matches only in a single location for each marker, consistent with the expectation that all of these regions come from the single copy portions of the chloroplast and are therefore not believed to be present in multiple copies.

A fourth data set, a previously published 27-taxon internal transcribed spacer (ITS) data set (Morton 2009), was also examined for comparison with the initial three datasets, because of incongruence seen between that gene tree and those we inferred from the other linkage groups (see Results). In each case, we use a Muscle (Edgar 2004) alignment followed by minor manual modifications.

DNA was extracted using either the EZNA plant DNA miniPrep kit (OMEGA Bio-Tek, Norcross, Georgia) or the DNAzol kit (Invitrogen, Carlsbad, California), as per the manufacturer's instructions. Amplifications by polymerase chain reaction (PCR) were done using HotStar Taq polymerase (Qiagen, Hilden, Germany) with the supplied buffer, a final Mg^{2+} concentration of 2 mM and a final concentration of 0.38 μM per primer, with the following reaction cycles: 95°C 5 min, (94°C 30 sec, 54–55°C 30 sec, 72°C 1–1.5 min) × 30, 72°C 5 min, 4°C hold. The primers used for *hyB* were: F 120: CTGCCGTCATGTCTAGTTTTGG, R 635: GAAAGAGCCCATATGGAACACC. The primers used for *mdh* were: F1: GCTCCTGTGGAAGAGACCC, R1: GCTCCAGAGATGACCAAAC. Amplification products were purified with a QIAquick spin purification procedure and their DNA quality and concentration was assessed with a Pharmacia Biotech Genequant II (Cambridge, UK) before being sequenced by Macrogen (Seoul, South Korea). Some sequences of *mdh* and *hyB* were generated previously (Ramadugu et al. 2013), Appendix 1.

In *mdh* and *hyB* we found some double peaks with approximately equivalent heights in the direct sequence traces. We inferred the phase of the alleles (or gene copies) by comparing these sequences to that of the most closely related other species based on the cpDNA phylogeny published previously (Bayer et al. 2009). Where a double peak shared a base with the closest relative, we assumed that that base belonged to one allele, whereas the other base to a second allele. In this way we have probably overestimated the difference between putative alleles and underestimated the difference between the first allele and the sequence from the closest relative. We then examined whether the alleles from an individual formed a monophyletic group with all alleles from that species (indicating that the assumptions of this allelic assignment will not affect inferences made regarding species relationships). If all alleles from one species did not form a monophyletic group it might indicate that either the allelic assignment was incorrect or that a biological process was in operation that might affect inferences made regarding species relationships (such as deep coalescence or paralogy). We did not observe

triple peaks of equivalent height and so have no direct evidence for more than two alleles per individual.

***Examination of Intra- and Intergenic Sources of Incongruence*—** The three nDNA regions (*mdh*, *hyB*, and ITS) were tested for intra-locus recombination using RDP v. 3beta34 (Martin et al. 2005). All the search alternatives were enabled, Bonferroni correction was used and an overall *p* value of 0.01 was set as the threshold of detection. Siscan and BootScan had their search window size set to 150, otherwise all methods had their default values.

The possible effect of positive selection on our phylogenetic inferences were examined using SplitsTree v.4 (Huson and Bryant 2006) by examining networks produced by codon-encoding characters and comparing the first and second position versus third position networks for any inconsistencies in taxon position. Although first and second position changes do not always result in changes to the encoded amino acids, and third positions sometimes do, by and large these partitions do reflect potential selective changes or neutral changes respectively.

Gene trees were inferred with two replicate runs (of four chains each) in MrBayes v.3.2 (Ronquist and Huelsenbeck 2003), using the GTR + G model in each case, summarizing trees drawn from the stationary phase of a total of five million generations as a majority rule consensus. These models were the closest available in MrBayes to those selected by the Akaike information criterion (AIC: Akaike 1974) in Modeltest (Posada and Crandall 1998) (*hyB* and *mdh* – transversion model (TVM) + gamma distribution (G); cpDNA – transversion model (TVM) + proportion of invariable sites (I) + gamma distribution (G); ITS – general time reversible (GTR) + gamma distribution (G)). The maximum clade credibility tree was determined using TreeAnnotator from BEAST (Drummond and Rambaut 2007). Parsimony bootstrap (BS) analyses were also performed with the following settings: 1,000 BS replicates, within which were 10 heuristic search replicates, step wise and random addition of taxa, holding one tree per replicate, and multrees on. The trees from each locus were compared for instances of supported (≥ 0.95 posterior probability (PP)) incongruence.

To determine whether hybridization and/or lineage sorting might explain the observed incongruences among loci, we used a coalescent-based hybrid detection test (Maureira-Butler et al. 2008), with modifications (Blanco-Pastor et al. 2012) that we include in our synopsis as follows: 500 trees (sampled from 500,000 mcmc generations) were drawn from the end of one of the tree files. Then 100 randomly chosen trees were sampled, to represent the posterior distribution of trees. *HyB* was excluded to maximize overlapping taxa and any remaining taxa that did not overlap between *mdh* and cpDNA were excluded. The resulting 100 trees were then made ultrametric in r8s 1.71 (Sanderson 2003) with the out-group species excluded (as required by r8s). We used cross validation with logarithmic increases of 0.3, ranging from -3 to +3, to determine the optimal smoothing parameter for the penalized likelihood method (Sanderson 2002). The crown age of the Aurantioids (>12.1 Ma) and of *Citrus* s. l. (>2.4 Ma) were drawn from earlier estimates (Pfeil and Crisp 2008). In order to accommodate age uncertainty, we used the youngest times as fixed node ages from the range of estimates from Pfeil and Crisp (2008), their table 1, last column. By using the youngest age estimates, we favor the null hypothesis, namely that lineage sorting can explain gene tree incongruence; thus, if the test rejects the null then we can have a high degree of confidence that another process is required to explain gene tree incongruence.

One hundred simulated trees were made based on each of the 100 trees in Mesquite (Maddison and Maddison 2006) with the "Coalescence Contained within Current Tree" module, assuming population sizes of 4,000 (derived from estimates made in *Citrus*, with the same assumptions: Ramadugu et al. 2013) and 20,000 (the latter as a sensitivity analysis in case the population sizes derived from *Citrus* are too low to be reasonable across ancestral lineages within Aurantioideae). The population size estimates were made using several alleles (mean n = 12) for four species of *Citrus* for three nuclear loci (Ramadugu et al. 2013) but similar data were not available for any other species used here. Because of the difficulty in getting precise effective population size estimates for deeper ancestral branches, we extrapolated the population size for all ancestral branches from the most shallow internal branches connecting the four species of *Citrus* we used, an approximation that has been used in several cases (Maureira-Butler et al. 2008; Blanco-Pastor et al. 2012; Ramadugu et al. 2013).

We then determined the distribution of tree to tree distances between each observed gene tree and the simulated set of trees generated from each. We compared these distributions (representing the null expected under lineage sorting alone) to the distance between the observed gene trees themselves. After finding that the differences were suggestive of hybridization (the observed gene tree distances lay beyond the null distribution values), we excluded some individuals one at a time to examine the magnitude of their effect on the tree to tree distances. We only excluded taxa that had some incongruence upon manual inspection of the gene trees. Some taxa were removed based on the results of this test before estimation of the species tree (see Results).

***Species Tree Inference*—** We inferred species trees using a multispecies coalescent model implemented in *BEAST (Heled and Drummond 2010) estimating uncorrelated lognormal relaxed clocks for each partition (Drummond et al. 2006). We only used three genes for this analysis, because the ITS was reduced to very short alignments of non-recombined parts (see Results). DNA substitution models used were as above. We calibrated the root node (the Aurantioid crown) using a truncated normal prior as follows: mean = 19.8 million years (Ma), stdev = 3.9, lower = 5.0 and upper = 40.0, which follow ages and uncertainties estimated previously (Pfeil and Crisp 2008), while preventing the sampling of possible (under the normal distribution) but unrealistic ages well above and below two standard deviations. We did not use the fossil leaf described from eastern Africa assigned to *Clausena* (Pan 2010), because the only character that separated this assignment from *Zanthoxylum* (Toddalioideae) was the presence in the fossil of multicellular uniseriate hairs on both ab- and adaxial leaf surfaces, as found in some *Clausena* species, whereas *Zanthoxylum* species examined in that study (Pan 2010) were found to have these hairs only on the abaxial surface. Therefore, we consider the placement of the fossil in a phylogeny as uncertain at this stage.

The gene ploidy value was adjusted for the cpDNA to be only half that of the autosomal nuclear genes, because these plants are generally hermaphrodites. We used uniform prior distributions for the cpDNA and nDNA clock rates ("ucld.mean"), with initial values and means based upon the previous r8s results (9×10^{-10} cpDNA, 3.3×10^{-9} for *hyB*, and 1.9×10^{-9} for *mdh*), large standard deviations (1×10^{-9} in each case) and upper and lower limits one order of magnitude above and below the initial values (based on the r8s results, not shown). We began with starting species and gene trees based on a preliminary run.

After running some test analyses, we discovered a lack of stability in parameters relating to the clock rate and species tree root height (i.e. species coalescent, speciation likelihood, Yule birth rate, ucld mean, and mean rate), as well as population size (i.e. population size likelihood and population mean), the prior, and the posterior, even after 400 M generations. We tried various prior adjustments to attempt to reach stability, as follows: (1) uniform species.yule.birthRate prior between 0 and 1, exponential species.popMean prior with mean = 0.02 and upper of 10, exponential treeModel.rootHeight prior for each gene tree with mean = 8 and offset of 12 (just below the lower 95% of the species tree root height prior). (2) As 1, but with an exponential species.yule.birthRate prior with mean = 0.1 instead of the uniform prior. (3) As 2, but the exponential species.popMean prior with mean = 0.08 and upper of 10. (4) As 3, but restricting the species tree root height prior standard deviation to 1.95 (down from 3.9). (5) As 4, but removing the exponential treeModel .rootHeight prior on the gene trees. (6) As 1, but with mean = 10 on the exponential treeModel.rootHeight prior for each gene tree. (7) As 1, but replacing the exponential treeModel.rootHeight prior for each gene tree with one that links the gene tree root heights to the current value of the species tree root height, with a diffuse normal prior.

Each combination (1–7, above) was run for 400 M generations. The run with a good balance between the best posterior mean value and lowest variance around the mean (run 7) was repeated five additional times to check for convergence and the results summarized (as a maximum clade credibility tree, using Tree Annotator) across all six runs after an appropriate burn-in was discarded. Species tree results were also checked across all runs (1–7, above) to see if the change in prior probabilities had any significant effect on clade or clade age posterior probabilities.

***Ancestral Area Reconstruction*—** The likely ancestral areas of clades were inferred using Fitch parsimony in Mesquite and maximum likelihood (the dispersal-extinction-cladogenesis [DEC] model) in Lagrange (Ree and Smith 2008). Our main interest was to test the hypothesis that the *Citrus* clade originated in Australasia (or more broadly anywhere east of Wallace's line and south of the Philippines). We defined six operational areas based on observed patterns of endemism as follows: (1) 'India' + Pakistan, Sri Lanka, Bangladesh, Western Myanmar, Bhutan, Nepal; (2) 'China' + Taiwan; (3) 'South East Asia' (SEA) including everything east and south of the former two areas, up to Wallace's line + Philippines; (4) 'Malesia', including east of Wallace's line, northern and western New Guinea and south Pacific Islands not included elsewhere; (5) 'Australia' + southeastern New Guinea + New Caledonia; and (6) 'Africa' + any remaining western Asian areas. We also explored an

even simpler coding, namely (1) west of Wallace's line or (2) east of Wallace's line, in a second analysis.

Each species' native distribution was identified by the GRIN database for all Auratioideae further modified by Swingle and Reece (1967), except for *Citrus medica* L., *C. reticulata*, and *Triphasia trifolia* DC., where this information was lacking. These species' ranges were assigned to the South East Asia region. The former two may come from further west or north, respectively (Swingle and Reece 1967), so they should not bias the results with respect to the hypothesis being tested. Genera other than *Citrus* were coded to include the cumulative distribution of all their species (by adding the additional ranges to the most widespread example in our data set), thus minimizing the effect of the lack of complete species-level sampling. In this way, we expect that the precision of ancestral area inference for nodes beyond the *Citrus* s. l. clade may be decreased, but otherwise unbiased. However, the key question for this study hinges on the inference of ancestral areas where we have the most dense sampling of genes and species, as well as the more precise coding of species' distributions.

For the Lagrange analysis, we first assessed several combinations of parameters using the MCC tree. We then checked some of these using 20 trees drawn from the posterior distribution of species trees from the *BEAST run used to create Fig. 1. The parameter combinations were (1) 'East West simple,' areas West and East of Wallace's line only, no range constraints, no dispersal constraints, and estimated dispersal and local extinction rates; (2) 'six areas simple,' as 1, but using six areas as described above; (3) 'isolation by distance,' using six areas, but range changes between adjacent areas = 1 (e.g. China – SEA), between areas one step away (e.g. China – Malesia) = 0.5, and between areas more than one step away = 0.1 (e.g. China – Australia). Barriers to range changes were ignored. (4) 'isolation and barriers,' using six areas, with range change relative probabilities as in 3, but barriers to dispersal and gene flow also considered. Thus, China was considered not adjacent to India nor Africa, because of the Himalayas (thus a single species in these areas must also be present in SEA). China was also considered not adjacent to Australia, because temperate plants need to cross the equator to connect these populations by gene flow, which was treated as unrealistic in this model, unless they were also present in the areas in between. These shifts were also scored as 0 probability in the dispersal constraints, to reflect difficulty in dispersal across these barriers. The dispersal matrices are presented in Table S1. Because the crown age of subfamily Aurantioideae was estimated to be relatively recent (see Results), we did not attempt to modify the cost matrix with respect to changing dispersal probabilities (e.g. reflecting changes in geological settings) in the time period of this group.

Finally, a biogeographical analysis was conducted under parsimony as implemented in Mesquite. We used the "Trace character over Trees" module to optimize areas on all internodes of the 7,500 trees selected from our *BEAST analyses. The results were summarized by computing the relative frequencies of ancestral area reconstructions for each node of the MCC tree.

Results

Majority rule consensus trees based on the Bayesian analyses' posterior distributions are presented for cpDNA (Fig. S1), nuclear *hyB* (Fig. S2), *mdh* (Fig. S3), ITS analyzed with only positions 1–322 (Fig. S4A), and ITS with only positions 323–end (Fig. S4B). We recovered matrices with c. 36% missing sites or gaps (of 1,341 aligned sites) in *hyB* (including for those taxa where alleles were not sampled, which were assumed to represent a single unknown allele in each case), ca. 8% (of 998 aligned sites) in *mdh*, ca. 18% (of 10,161 aligned sites) in the cpDNA, and c. 12% (of 764 aligned sites) in ITS. The cpDNA, *hyB*, and *mdh* data sets can be found in TreeBASE (http://purl.org/phylo/treebase/phylows/study/TB2:S16378), whereas the ITS data is available as a popset GenBank submission (UID: 223470234). The xml file from run 7 is available as a supplementary file.

***Sources of Incongruence*—**Tests of intra-locus recombination found that the *mdh* alignment contained no recombination events, whereas a single weak signal of recombination was found within the *hyB* alignment. As only one of the methods made this identification, this 'event' was not considered significantly supported to warrant removal of the sequence (Poke et al. 2006). Within the ITS alignment, however, significant evidence for recombination was discovered within six taxa, the strongest indications being from all methods for recombination of approximately the front half of ITS from *Merrillia* and/or *Murraya* ($p = 1.3 \times 10^{-6}$ from MaxChi to $p = 3.7 \times 10^{-11}$ from BootScan – the same probabilities for each taxon), with putative parental sequences *Clausena* and *Glycosmis* in both cases. The RDP cannot reliably determine which of the sequences are recombinants and which are parents, so all four sequences should be considered as possible recombinants. The other four possible recombinants were *Balsamocitrus* (5′ terminus), *Glycosmis* (3′ terminus), *Ruta* (5′ terminus), and *Feroniella* (small part of the middle region).

The maximum parsimony BS tree presented in figure 4 of Morton's paper (2009), using the entire ITS alignment, recovered a clade containing *Glycosmis* + *Merrillia* (MP BS = 98%) that we also recover with those data (but our Muscle alignment) using Bayesian analysis (BA) (PP = 1; tree not shown). Citreae also is not monophyletic in Morton's (2009) and our analyses of ITS, whether *Merrillia* + *Murraya* are included in this tribe or not (cf. clade G in Bayer et al. 2009: their figure 1). However, when reanalyzed with the recombined portion of these taxa excluded (using only positions 323–end; Fig. S4B), we instead recovered a modestly supported clade containing *Merrillia* + *Murraya* (PP = 0.89) that is tentatively supported as sister to the Citreae (PP = 0.91). This clade, (*Merrillia* + *Murraya*) + Citreae, corresponds to clade G in Bayer et al. (2009). Citreae itself is robustly supported, despite the limited remaining ITS data (PP = 0.98), contra Morton's (2009) findings. It should be noted that with MP analysis alone, we do not recover all clades with the same support as Morton (2009), presumably due to the difference between alignments, although the key findings are recovered with BA.

We also reanalyzed the remaining part of the ITS alignment (using only positions 1–322) to examine the effect on *Glycosmis*, which appeared to have a recombinant part in the 3′ end of the sequence. We found that *Glycosmis* + *Merrillia* formed a clade (PP = 0.99) sister to *Murraya* (PP = 0.94; Fig. S4A). The terminal branch lengths of *Merrillia* in both Figs. 4A and 4B are both short, which suggests that *Merrillia* is the recombinant, rather than *Murraya*, and that it may be of recent origin.

When we examined first and second versus third positions for exon coding regions using SplitsTree, we found that fewer than three changes were involved in any differences in taxon placements between partitions within either *mdh* or *hyB* (not shown). This was due to the low variation within exon regions. We conclude that positive selection that might be detected by examining exon coding sequences, if it is operating at all, is unlikely to affect the phylogenetic inferences made here because of the relatively low rate of change in exon versus intron sequences.

Regarding allelic inference, in all but one case, both alleles for an individual accession in the gene trees were monophyletic with respect to all alleles inferred from that species. In the single case of non-monophyly, *hyB* alleles of *Balsamocitrus dawei* Stapf formed a grade with respect to a close relative, *Aegle marmelos* (L.) Corrêa; these species together formed a clade with *Aeglopsis chevalieri* Swingle and *Afraegle paniculata* (Schumach. & Thonn.) Engl. This clade of four species is

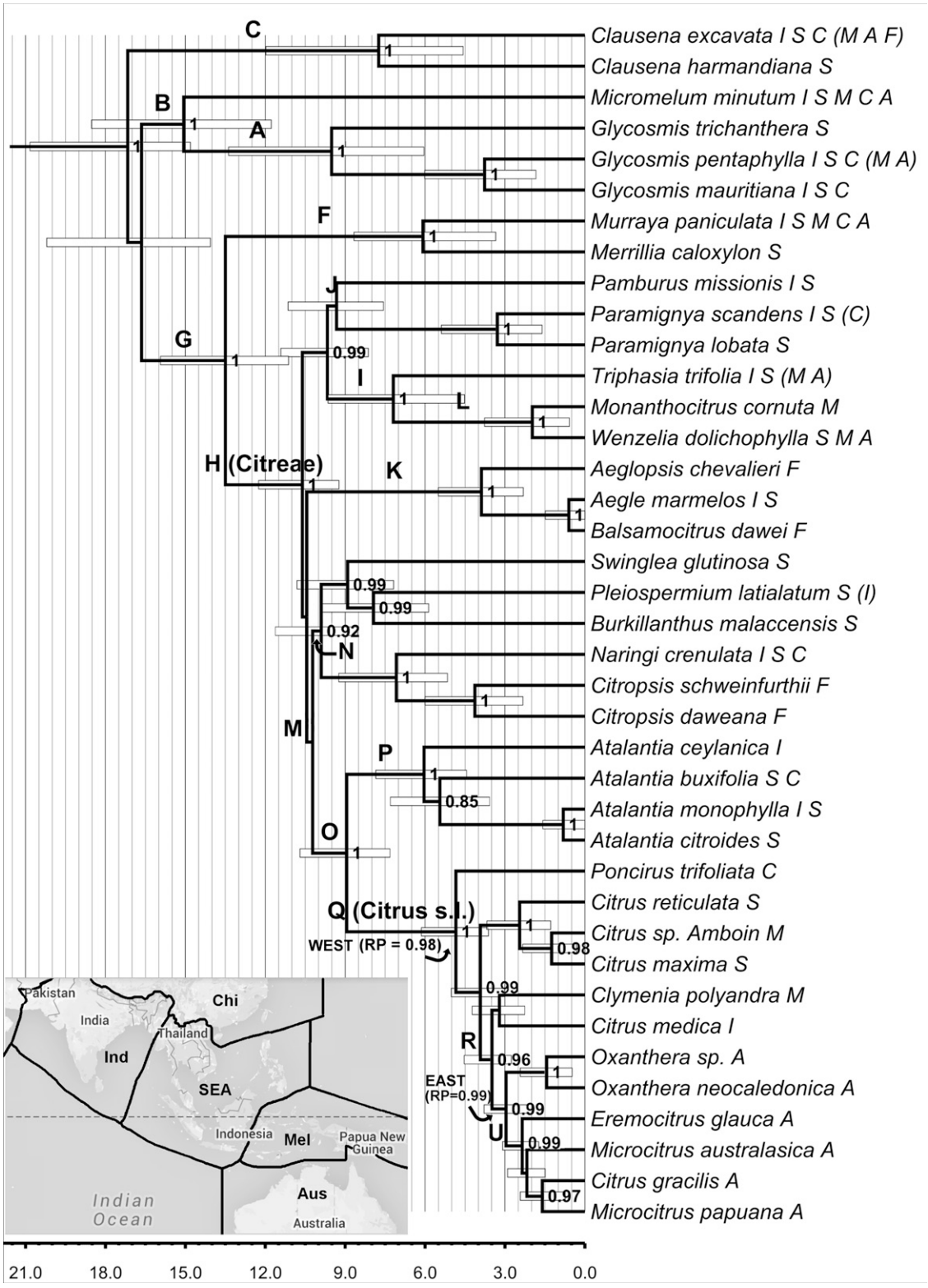

FIG. 1. A species tree based on chloroplast (cpDNA), and nuclear beta-hydroxylase (*hyB*) and malate dehydrogenase (*mdh*) loci using a Bayesian multispecies coalescent analysis in *BEAST. The tree is the maximum clade credibility (MCC) tree resulting from run 7 (see Methods). The areas around Wallace's line, and their boundaries as used here, are shown at the bottom left. The area coding used in the ancestral area inference for each taxon is shown by a single letter abbreviation per area to the right of the taxon name. For non-*Citrus* s. l. species, where the species sample is low, we included areas where other species in each genus are found, listed here in parentheses after the most wide-spread included species name and its area(s). A = Australasia (= Aus in the map); C = China (= Chi in the map); F = Africa (not shown in the map); I = India (= Ind in the map); M = Malesia (= Mel in the map); S = SE Asia (= SEA in the map). The relative probabilities (RP) of the inferred ancestral area for two nodes (based on the DEC analysis, using only two coded areas and this tree) — the crown of *Citrus* s. l. west of Wallace's line and the crown of the Australasian subgroup east of it — are shown with arrows pointing to the relevant node. Numbers to the right of nodes indicate clade posterior probabilities (PP). Scale axis units are millions of years. Clades are labelled by single letters above and/or to the left of branches to match Bayer et al. (2009), except clade N, where an arrow indicates the corresponding branch. Bars indicate 95% HPD of node ages.

found in all three gene trees (PP = 1 in each), but the position of *Balsamocitrus* differs within the clade — it is either sister to *Aegle* (*mdh*) or sister to *Aeglopsis* (cpDNA). Therefore, we interpret the non-monophyly of *Balsamocitrus* alleles in *hyB* as an isolated case of deep coalescence rather than paralogy.

Hybridization Test—We compared the observed gene trees for *mdh* and cpDNA in relation to null distributions generated using lineage sorting alone (under the coalescent). After doing this, we found that the difference between the observed gene trees was much larger than could be explained by lineage sorting alone, given the effective population sizes and generation times we assumed here (Fig. S5). Given that we cannot place too much weight in these assumptions (i.e. because we extrapolated the same effective population sizes to every internal branch in the trees), we used the results of this test as a guide to the presence of taxa affected by hybridization, rather than a definitive test. Several individual accessions were removed one by one and the effect on the tree to tree distances was examined and ranked. The distance between the observed gene trees and the upper value of the null distributions reduced by between no and five steps when any individual was removed (not shown). A hybrid suspected a priori (sweet orange, *Citrus* × *aurantium*) (Scora 1975; Barrett and Rhodes 1976), which also has different cpDNA, mtDNA versus nDNA origins (Bayer et al. 2009; Froelicher et al. 2011; Ramadugu et al. 2013), reduced this difference by the maximum observed five points. Six other individuals reduced the difference by almost as much (four points each). When compared to the sensitivity analysis (using 20,000 effective population size), a similar result was observed, except that only four individuals in common with the first set produced the highest or second highest effect on the test scores. These individuals were *Citrus* × *aurantium* (the suspected hybrid), *Afraegle paniculata*, *Bergera koenigii,* and *Feronia limonia*. All four were excluded as suspected hybrids in order not to violate the assumptions of species tree inference.

The Position of Feroniella—We found *Feroniella* in the combined cpDNA tree to be placed sister to *Citrus reticulata* (PP = 1). In the *mdh* tree it was contained a corresponding group (PP = 0.99), both results that are in contrast to previous studies. We excluded this accession from the species tree analysis because of a suspected mix up between samples (see Discussion).

Species Tree—Each species tree run (1–7) had overlapping posterior probability distributions. The best run (7) was repeated 5 times, reaching convergence among runs. The combined analysis (1,880 M generations) still had effective sample size values lower than 200 for all parameters related to rates, population mean (but not split population sizes), and divergence times, including the prior and posterior, but well above 200 in the remaining parameters. Despite this, species trees from a selection of runs (1, 3, 5, 6, and 7) were nearly identical in topology and clade posterior probabilities (not shown), but differed somewhat in tree root height (means ca. 15.5–19.5 Ma), one of the unstable parameters.

The species trees from the replicated run (run 7) were also nearly identical, with clade posterior probabilities differing by only a few percentage points among replicates. Presented is the tree based on one of these replicate runs (Fig. 1). In many respects the tree is similar to that based on cpDNA alone (Bayer et al. 2009). For example, Bayer's clade Q (*Citrus* s. l.; with BS/PP = 100%/1.0) matches a clade we found, also with PP = 1 (the same letter codes are used where the clades in Fig. 1 are shared with Bayer's results). Other clades in the species tree in common with Bayer et al.'s (2009) results are clades A, B, C, F, G, H, I, J, K, L, M, N, O, and P (Fig. 1), although J is weakly supported in both their and our analyses.

We found that the Aurantioideae crown is at least 14.8–20.8 Ma (95% HPD), median = 17.2, whereas the *Citrus* s. l. crown is at least 3.6–6.1 Ma (95% HPD), median = 4.8, despite the lack of stability in these parameter estimates. These ages are minimum estimates, because they are ultimately derived from fossil calibrations.

Ancestral Area Reconstruction—The Lagrange ancestral area inference, using only areas east/west of Wallace's line and no area or dispersal restrictions, returned a western origin for *Citrus* s. l. (Fig. 1). Using the MCC tree, the relative probability (RP) was 0.98 (Fig. 1); using 20 trees drawn from the end of the 7,500 trees found in one *BEAST analysis to acount for topology and branch length uncertainty, the RP ranged from 0.96–0.99. The ancestral area shifts to east of Wallace's line with high RP (0.99) at the crown of the Australasian subgroup, consistent with a shift in distribution of the ancestral lineage between these nodes from west to east of Wallace's line between ca. six and two mya (Fig. 1).

Under either the'six areas simple' or the 'isolation by distance' models, using the MCC tree, SEA was returned as the most likely origin of *Citrus* s. l. (RP = 0.69). With India, China and/or SEA (in any combination), the summed RP increased to 0.88. With 20 trees drawn from the posterior distribution, the 'six areas simple' model had a mean RP of 0.69 for SEA and 0.86 when summed for India, China and/or SEA as the origin of *Citrus* s. l. Under the more restrictive 'isolation and barriers' model the RP of a SEA origin increased to 0.73, and with India or China combined, to 0.87. With 20 trees drawn as before to include uncertainty, the mean RP for a SEA origin was also 0.73 and the mean RP including India and China was 0.86.

A parsimony reconstruction of the ancestral area in Mesquite also pointed to west of Wallace's line, and specifically to SEA, as the most reasonable location for the *Citrus* s. l. origin. More than 99% of trees drawn from the BA posterior distribution returned a SEA ancestral area using the six area coding. When coded simply as east/west of Wallace's line, 100% of trees returned a western ancestral area for the *Citrus* s. l. node.

Discussion

The new hypothesis of relationships within Rutaceae subfamily Aurantioideae has several strengths over the existing ones, including the addition of two nDNA markers that are unlinked and in single copy in a single *Citrus* accession, the use of the multispecies coalescent model, and the examination of processes that may confound species tree estimation if undetected (such as recombination and hybridization).

Classification—With respect to the classification of Swingle and Reece (1967), we find no support for tribe Clauseneae, nor subtribe Clauseninae (the only one with more than one genus), as monophyletic groups. However, Swingle's tribe Citreae is supported by our species tree. If we consider the expanded circumscription proposed by the GRIN classification (by June 2011), Citreae s. l. (including an expanded Merrilliinae s. l.: *Merrillia* + *Murraya*) is also supported as monophyletic, as is

Merrilliinae s. l. GRIN's recircumscribed Clauseneae (*Bergera*, *Clausena*, *Glycosmis*, *Micromelum*: Clauseneae s. s. here) is not supported, but nor is it contradicted by the species tree.

Within Citreae, some of Swingle and Reece's (1967) groups can be discerned, although others are contradicted by our results. Subtribe Triphasiinae is supported in the species tree as a clade, if we take a narrower circumscription that excludes *Oxanthera*, although *Luvunga* and *Merope* were not sampled. Citrinae most closely corresponds to our well-supported clade M, but includes *Swinglea* (from Balsamocitrinae) and *Oxanthera* (from Triphasiinae).

Only one subgroup within Swingle and Reece's (1967) subtribes is entirely consistent with our species tree. The included genera of Balsamocitrinae group B form a well-supported group, notwithstanding the exclusion of *Afraegle*, in our species tree analysis. The latter genus is still found within this group in each gene tree. The support in every analysis is high for this group, so there appears to be little doubt that this constitutes a natural group. A second subgroup, Citrinae group C (which includes all closely allied genera of *Citrus* s. s.), can be found in our clade Q and includes all 6 genera placed together by Swingle and Reece (1967). However, *Oxanthera* is clearly also part of this clade, with support in each analysis presented here. This clade corresponds to Bayer et al.'s (2009) clade Q and places genera not universally accepted as part of *Citrus* nested within it, such as *Clymenia* (*Citrus polyandra*), *Eremocitrus* (*Citrus glauca*), *Microcitrus* (e.g. *Citrus australasica*), and *Oxanthera* (*Citrus neocaledonica*), along with one genus that is sister to the remaining taxa, *Poncirus* (*Citrus trifoliata*).

Our results are in contrast to one finding by Morton (2009), namely that the position of *Murraya* "*sister to Pamburus*" renders both Citreae and Clauseneae sensu Swingle paraphyletic. Two aspects need to be addressed here. The first is that the "support" for this relationship is that this clade is present in 96% of most parsimonious trees, although it has MP BS support of < 50% in the combined molecular and morphological analysis (Morton 2009: see Fig. 6 in that paper), and 61% MP BS support in the combined molecular analysis (Morton 2009: see Fig. 5 in that paper). Secondly, if the alternative position for *Murraya* (i.e. sister to *Merrillia*) is affected in some way, then the position of this taxon may become unstable in the tree and lack strong support for any particular position.

We found that the ITS region used by Morton (2009) is affected by recombination involving *Murraya* and *Merrillia* among others. When the recombined part alone (Fig. S4A) or the whole region is used (not shown), *Merrillia* is strongly supported as sister to *Glycosmis*, reflecting the findings of Morton (2009). However, when the non-recombined part (with respect to *Murraya* and *Merrillia*) of this region is used alone (Fig. S4B), a phylogenetic inference consistent with our species tree is found, where *Murraya* is sister to *Merrillia* (although only weakly supported in the fragment of ITS). Our BA analyses differed from Morton's results in one important aspect: whether we analyzed the full ITS region or only the recombined part, *Murraya* was always supported as sister to *Merrillia* + *Glycosmis*.

Our interpretation is that *Merrillia* is the recombinant, and that either the model-based analysis, our alignment, or both, contribute to placing *Murraya* close to *Merrillia* despite the recombination event, rather than "floating" to a spurious and weakly supported position (i.e. sister to *Pamburus*).

Our result with regard to the placement of *Feroniella* is in contrast to earlier findings based on cpDNA (Araújo et al. 2003; Morton 2009; Penjor et al. 2010). In Morton (2009), using maximum parsimony, *Feroniella* was sister to *Feronia* in cpDNA (*trnLF*, 71% BS; *rp*S16, 95% BS; *atp*B-*rbc*L spacer, 94% BS), but together were unresolved within the Aurantioideae. In Araújo et al. (2003), using maximum parsimony, *Feroniella* was poorly resolved within Citreae (clade H on our Fig. 1), but *Feronia* was not sampled. In Penjor et al. (2010), using neighbor joining, these two genera were again sister (75% BS), but even more poorly resolved — the monophyly of Aurantioideae was not supported. Each of these studies found that *Feroniella* was not nested within *Citrus* s. l. and excluded from that possibility by moderately to well-supported nodes, in contrast to our findings. We found *Feroniella* sister to *Citrus reticulata* in both cpDNA and *mdh*, whereas *Feronia* was closer to the *Atalantia* + *Severinia* clade (PP = 0.99) in cpDNA or grouped with *Swinglea* and *Burkillanthus* in *mdh* (PP = 1). This difference is present despite our use of the same germplasm source between our and Morton's (2009) studies (and possibly that of Penjor et al. 2010), and the sequencing of several markers that agree within studies. The other study (Araújo et al. 2003) derived their *Feroniella* material from a different source. When we reanalyzed different parts of the ITS with BA, *Feroniella* was excluded from *Citrus* s. l. by moderately (PP = 0.9 positions 1–322) or strongly (PP = 1 positions 323–end) supported nodes, reaffirming this main difference between ITS and our analyses.

Given that several labs received or derived material independently of one another and that the DNA extractions and PCRs were performed independently, it is likely that a sample mix up involving *Feroniella* can explain the discordant results shared by us and Bayer et al. (2009), as we used the same DNA extract. To test this, we reacquired this material from the USDA, re-extracted DNA and sequenced the *rpS16* and *trnLF* regions in a different lab (UC Riverside). We then compared these sequences to those submitted previously and found that our new sequences were nearly identical to those of Morton (2009). Using neighbor net, these sequences alone place *Feroniella* close to *Citrus*, but not within it (not shown). However, the remaining cpDNA sequences used in Bayer et al. (2009) recovered *Feroniella* strongly nested within *Citrus* s. l. Therefore, we consider our initial *Feroniella* result, and that of Bayer et al. (2009), to be spurious. The position based on the available confirmed sequences is not well resolved, but not placed within *Citrus* s. l.

Morton (2009) presented a matrix of 20 phenotypic characters in subfamily Aurantioideae and concluded that almost none contained uncontradicted synapomorphic states for any taxon described previously within this group. However, the presence of carbazole (a crystalline substance) in Clauseneae s. s. alone has already been reported (Samuel et al. 2001). Although that clade did not occur in our species tree, the poor support for the alternative position of *Glycosmis* + *Micromelum* means that the monophyly of Clauseneae s. s. is not rejected here, thus this character may still turn out to be an uncontradicted synapomorphy. Morton (2009) listed *Murraya* as being polymorphic for this state and did not list *Bergera* separately in her table. If we consider these genera separately, as did Samuel et al. (2001), then the presence of the aromatic heterocyclic compound — carbazole — is highly consistent with our species tree, apart from the poorly supported clade already mentioned. With respect to the

remaining characters presented in Morton (2009), we concur that in general phenotypic characters do not track the molecular phylogeny well at all. However, the presence of spines might be a synapomorphy for clade H in our species tree (Fig. 1), although it too must have been subsequently lost in a few cases (e.g. *Clymenia*).

Origin of Citrus s. l.—Two previous studies mention hypotheses regarding the origin of *Citrus*: Swingle and Reece (1967), using a narrowly defined *Citrus* (i.e. 16 species from primarily south and east Asia, and west Malesia), thought the group probably arose in east Asia, although a New Guinea – Melanesian origin was also thought possible. However, when the discussion includes all species in *Citrus* s. l. (to include *Clymenia*, *Eremocitrus*, *Fortunella*, *Microcitrus*, *Oxanthera,* and *Poncirus*), Swingle and Reece (1967) suggested a trend from southeast to northwest, culminating in the most "*highly developed*" species of subgenus *Citrus*. Beattie et al. (2006), elaborating on Swingle and Reece (1967), took their proposed timing of evolutionary progression and suggested a scheme by which an Australasian origin was followed by migration to East Asia assisted by the movement of terranes over the last 37 Ma, especially those that formed the Halmahera Islands and the eastern Philippines.

The crown age of *Citrus* s. l. was previously examined and found to be most probably too young to be consistent with an east to west migration via terrane movement (Pfeil and Crisp 2008), namely at least 2.4–13.6 Ma. The results of our analyses here concur and suggest an even younger minimum crown age of this clade (3.6–6.1 Ma), probably due to the use of a multispecies coalescent model, which can infer speciation ages that are younger than the divergence times of included gene lineages. However, the inherent limitations of fossil calibrated age estimates means that older ages cannot be excluded absolutely. In this case, though, positive evidence for older ages that we consider reliable is currently not known for the subfamily. For example, Pan (2010) described a fossil leaf from eastern Africa dated to around 27.2 Ma that was assigned to *Clausena*. If this assignment was accepted and applied to the stem node of our sample of *Clausena* (= crown node of Aurantioideae), we might expect that the *Citrus* s. l. crown age could be twice as high as what we found here (e.g. simply rescaling that node to equal 27.2 Ma yielded an upper age of *Citrus* s. l. of under 10 Ma). This would still not change the inconsistency that we observe between the inferred age estimate of *Citrus* s. l. and terrane movement in Malesia, where a *Citrus* ancestor would probably have needed to be in northeastern New Guinea by at least 20 mya in order to use the Halmahera route to disperse to Asia.

As a complement to age estimation alone to determine the origin of *Citrus* s. l., we used ancestral area reconstruction. By either maximum likelihood or parsimony-based approaches and based on available data, we found that the original location of the *Citrus* s. l. clade was most probably west of Wallace's line. The time frame we infer is significant, because it marks the beginning of the collision between the Australian and the Sundaland margin, the emergence of New Guinea, and the reduction of the distance across deep ocean that still separates these areas (Hall 1998, 2002; Lohman et al. 2011). The inferred age, ancestral area reconstruction, and lack of plausible vicariant breaks are incompatible with the earlier hypotheses regarding the origin of *Citrus* s. l. and movement from Australasia to Asia via the Halmahera route. Available evidence instead suggests that *Citrus* s. l. is a young group that has radiated from an East Asian origin, most probably in the late Miocene or Pliocene, and dispersed across deep ocean gaps at least once to cross from western to eastern Malesia. The timing is consistent with the increasingly close connections formed between SE Asia and New Guinea via the Banda Arc from about 7 mya (Hall 1998, 2002), where these island chains could have been used as a series of stepping stones to aid dispersal. This dispersal was also accompanied by diversification, especially within Australia. There are other examples of other plant groups that have also migrated and then radiated across Wallace's Line in the same direction within the last 20 Ma or so, on the basis of dated molecular phylogenies and explicit ancestral area inferences. These include *Rhododendron* section *Vireya*, Ericaceae (Webb and Ree 2012), *Aglaieae*, Meliaceae (Muellner et al. 2008), Livistoninae, Arecaceae (Bacon et al. 2013), and Isonandreae, Sapotaceae (Richardson et al. 2014). *Citrus* s. l. is certainly not unique in this regard and provides another case of the generation of diversity when new areas were accessed via long distance dispersal.

Acknowledgments. We thank Cathy Miller and Ish Sharma at CSIRO, Canberra, for their assistance and provision of materials. We also thank Henrik Zetterlund (Gothenburg Botanical Garden) for access to material and cultivating Rutaceae for us. We also thank the reviewers and editors for their constructive comments. BEP is supported by grants from VR (Swedish Research Council), KVA (Royal Swedish Academy of Sciences), Lars Hiertas Minne fund, The Royal Physiographic Society in Lund, Helge Ax:son Johnsons fund, and the Lundgrenska fund.

Literature Cited

Akaike, H. 1974. A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19: 716–723.

Álvarez, I. and J. F. Wendel. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular Phylogenetics and Evolution* 29: 417–434.

Araújo, E. F. d., L. P. d. Queiroz, and M. A. Machado. 2003. What is *Citrus*? Taxonomic implications from a study of cpDNA evolution in the tribe Citreae (Rutaceae subfamily Aurantioideae). *Organisms, Diversity & Evolution* 3: 55–62.

Bacon, C. D., F. Michonneau, A. J. Henderson, M. J. McKenna, A. M. Milroy, and M. P. Simmons. 2013. Geographic and taxonomic disparities in species diversity: Dispersal and diversification rates across Wallace's Line. *Evolution* 67: 2058–2071.

Barrett, H. C. and A. M. Rhodes. 1976. A numerical taxonomic study of affinity relationships in cultivated *Citrus* and its close relatives. *Systematic Botany* 1: 105–136.

Bayer, R. J., D. J. Mabberley, C. Morton, C. H. Miller, I. K. Sharma, B. E. Pfeil, S. Rich, R. Hitchcock, and S. Sykes. 2009. A molecular phylogeny of the orange subfamily (Rutaceae: Aurantioideae) using nine cpDNA sequences. *American Journal of Botany* 96: 668–685.

Beattie, G. A. C., P. Holford, D. J. Mabberley, A. M. Haigh, R. J. Bayer, and P. Broadbent. 2006. Aspects and insights of Australia-Asia collaborative research on Huanglongbing. Pp. 47–64 in *Procceedings of the International Workshop for the Prevention of Citrus Greening Disease in Severely Infected Areas*. Tokyo: International Research Division, 46 Agriculture Forestry Fisheries Research Council Secretariat, Ministry of Agriculture, Forestry and Fisheries, Tokyo, Japan.

Blanco-Pastor, J. L., P. Vargas, and B. E. Pfeil. 2012. Coalescent simulations reveal hybridization and incomplete lineage sorting in Mediterranean *Linaria*. *PLoS One* 7: e39089.

Chase, M. W., C. C. Morton, and J. A. Kallunki. 1999. Phylogenetic relationships of Rutaceae: A cladistic analysis of the subfamilies using evidence from rbcL and atpB sequence variation. *American Journal of Botany* 86: 1191–1199.

Degnan, J. H. and N. A. Rosenberg. 2009. Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends in Ecology & Evolution* 24: 332–340.

Drummond, A. J. and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evolutionary Biology* 7: 214.

Drummond, A. J., S. Y. W. Ho, M. J. Phillips, and A. Rambaut. 2006. Relaxed phylogenetics and dating with confidence. *PLoS Biology* 4: e88, doi: 10.1371/journal.pbio.0040088.

Edgar, R. C. 2004. MUSCLE: Multiple sequence alignment with high accuracy and high throughput. *Nucleic Acids Research* 32: 1792–1797.

Edwards, S. V., L. Liu, and D. K. Pearl. 2007. High-resolution species trees without concatenation. *Proceedings of the National Academy of Sciences USA* 104: 5936–5941.

Engler, A. 1897. Rutaceae. Pp. 95–201 in *Die natürlichen Pflanzenfamilien*, ed. A. Engler and K. Prantl. Leipzig: Wilhelm Engelmann.

Froelicher, Y., W. Mouhaya, J.-B. Bassene, G. Costantino, M. Kamiri, F. Luro, R. Morillon, and P. Ollitrault. 2011. New universal mitochondrial PCR markers reveal new information on maternal citrus phylogeny. *Tree Genetics & Genomes* 7: 49–61.

Hall, R. 1998. The plate tectonics of Cenozoic SE Asia and the distribution of land and sea. Pp. 99–131 in *Biogeography and geological evolution of SE Asia*, eds. R. Hall and J. D. Holloway. Leiden: Backhuys Publishers.

Hall, R. 2002. Cenozoic geological and plate tectonic evolution of SE Asia and the SW Pacific: computer-based reconstructions, model and animations. *Journal of Asian Earth Sciences* 20: 353–431.

Heled, J. and A. J. Drummond. 2010. Bayesian inference of species trees from multilocus data. *Molecular Biology and Evolution* 27: 570–580.

Huson, D. H. and D. Bryant. 2006. Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution* 23: 254–267.

Jong, R. d. 1998. Halmahera and Seram: Different histories, but similar butterfly fauna. Pp. 315–325 in *Biogeography and geological evolution of SE Asia*, eds. R. Hall and J. D. Holloway. Leiden: Backhuys.

Kubatko, L. S. and J. H. Degnan. 2007. Inconsistency of phylogenetic estimates from concatenated data under coalescence. *Systematic Biology* 56: 17–24.

Lohman, D. J., M. d. Bruyn, T. Page, K. v. Rintelen, R. Hall, P. K. L. Ng, H.-T. Shih, G. R. Carvalho, and T. v. Rintelen. 2011. Biogeography of the Indo-Australian Archipelago. *Annual Review of Ecology Evolution and Systematics* 42: 205–226.

Mabberley, D. J. 1998. Australian Citreae with notes on other Aurantioideae (Rutaceae). *Telopea* 7: 333–344.

Maddison, W. P. and D. R. Maddison. 2006. Mesquite: A modular system for evolutionary analysis, v. 1.12.

Martin, D. P., C. Williamson, and D. Posada. 2005. RDP2: Recombination detection and analysis from sequence alignments. *Bioinformatics* 21: 260–262.

Maureira-Butler, I. J., B. E. Pfeil, A. Muangprom, T. C. Osborn, and J. J. Doyle. 2008. The reticulate history of *Medicago* (Fabaceae). *Systematic Biology* 57: 466–482.

Morton, C. 2009. Phylogenetic relationships of the Aurantioideae (Rutaceae) based on the nuclear ribosomal DNA ITS region and three noncoding chloroplast DNA regions, *atpB-rbcL* spacer, *rps16*, and *trnL-trnF*. *Organisms, Diversity & Evolution* 9: 52–68.

Muellner, A. N., C. M. Pannell, A. Coleman, and M. W. Chase. 2008. The origin and evolution of Indomalesian, Australasian and Pacific island biotas: Insights from Aglaieae (Meliaceae, Sapindales). *Journal of Biogeography* 35: 1769–1789.

Pan, A. D. 2010. Rutaceae leaf fossils from the Late Oligocene (27.23 Ma) Guang River flora of northwestern Ethiopia. *Review of Palaeobotany and Palynology* 159: 188–194.

Penjor, T., T. Anai, Y. Nagano, R. Matsumoto, and M. Yamamoto. 2010. Phylogenetic relationships of *Citrus* and its relatives based on *rbcL* gene sequences. *Tree Genetics & Genomes* 6: 931–939.

Pfeil, B. E. and M. D. Crisp. 2008. The age and biogeography of *Citrus* and the orange subfamily (Rutaceae: Aurantioideae). *American Journal of Botany* 95: 1621–1631.

Poke, F. S., D. P. Martin, D. A. Steane, R. E. Vaillancourt, and J. B. Reid. 2006. The impact of intragenic recombination on phylogenetic reconstruction at the sectional level in *Eucalyptus* when using a single copy nuclear gene (cinnamoyl CoA reductase). *Molecular Phylogenetics and Evolution* 39: 160–170.

Posada, D. and K. A. Crandall. 1998. Modeltest: Testing the model of DNA substitution. *Bioinformatics* 14: 817–818.

Ramadugu, C., B. E. Pfeil, K. L. Manjunath, R. F. Lee, I. J. Maureira-Butler, and M. L. Roose. 2013. Coalescence simulation testing of hybridization versus lineage sorting in *Citrus* (Rutaceae) using six nuclear genes. *PLoS One* 8: e68410.

Rauscher, J. T., J. J. Doyle, and A. H. D. Brown. 2002. Internal transcribed spacer repeat-specific primers and the analysis of hybridization in the *Glycine tomentella* (Leguminosae) polyploid complex. *Molecular Ecology* 11: 2691–2702.

Ree, R. H. and S. A. Smith. 2008. Maximum likelihood inference of geographic range evolution by dispersal, local extinction, and cladogenesis. *Systematic Biology* 57: 4–14.

Richardson, J. E., A. M. Bakar, J. Tosh, K. Armstrong, J. Smedmark, A. A. Anderberg, F. Slik, and P. Wilkie. 2014. The influence of tectonics, sea-level changes and dispersal on migration and diversification of Isonandreae (Sapotaceae). *Botanical Journal of the Linnean Society* 174: 130–140.

Ronquist, F. and J. P. Huelsenbeck. 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* 19: 1572–1574.

Samuel, R., F. Ehrendorfer, M. W. Chase, and H. Greger. 2001. Phylogenetic analyses of Aurantioideae (Rutaceae) based on non-coding plastid DNA sequences and phytochemical features. *Plant Biology* 3: 77–87.

Sanderson, M. J. 2002. Estimating absolute rates of molecular evolution and divergence times: A penalized likelihood approach. *Molecular Biology and Evolution* 19: 101–109.

Sanderson, M. J. 2003. r8s; Inferring absolute rates of evolution and divergence times in the absence of a molecular clock. *Bioinformatics* 19: 301–302.

Scora, R. W. 1975. On the history and origin of citrus. *Bulletin of the Torrey Botanical Club* 102: 369–375.

Stefanović, S., B. E. Pfeil, J. D. Palmer, and J. J. Doyle. 2009. Relationships among phaseoloid legumes based on sequences from eight chloroplast regions. *Systematic Botany* 34: 115–128.

Straub, S. C. K., B. E. Pfeil, and J. J. Doyle. 2006. Testing the polyploid past of soybean using a low-copy nuclear gene—Is *Glycine* (Fabaceae: Papilionoideae) an auto- or allopolyploid? *Molecular Phylogenetics and Evolution* 39: 580–584.

Suh, Y., L. B. Thien, H. E. Reeve, and E. A. Zimmer. 1993. Molecular evolution and phylogenetic implications of internal transcribed spacer sequences of ribosomal DNA in Winteraceae. *American Journal of Botany* 80: 1042–1055.

Swingle, W. T. and P. C. Reece. 1967. The botany of *Citrus* and its wild relatives of the orange subfamily. Pp. 190–430 in *The Citrus industry*, vol. 1, ed. W. Reuther, H. J. Webber, and L. D. Bachelor. Berkeley: University of California Press.

Syring, J., K. Farrell, R. Businsky, R. Cronn, and A. Liston. 2007. Widespread genealogical nonmonophyly in species of *Pinus* subgenus *Strobus*. *Systematic Biology* 56: 163–181.

Webb, C. O. and R. H. Ree. 2012. Historical biogeography inference in Malesia. Pp. 191–215 in *Biotic evolution and environmental change in Southeast Asia*, eds. D. J. Gower, K. G. Johnson, J. E. Richardson, B. R. Rosen, L. Rüber, and S. T. Williams. Cambridge: Cambridge University Press.

Wendel, J. F., A. Schnabel, and T. Seelanan. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences USA* 92: 280–284.

Appendix 1. List of vouchers and sequence references or GenBank accession numbers in the following order: species name and authority; voucher; cpDNA GenBank (ID#), *hyB* GenBank (ID#), *mdh* GenBank (ID#). The ID numbers correspond to the sequences used in the gene trees (Supplementary Figs.). Where two sequences are listed for one locus, these are alleles from the same individual. Sequence references 1 = Ramadugu et al. (2013); 2 = Bayer et al. (2009); 3 = likely mistaken sample, therefore not submitted to GenBank; "-" = not sequenced for that locus; PI = USDA Germplasm Index; CANB = Australian National Herbarium; GB = University of Gothenburg Herbarium; Merbein = CSIRO Citrus collection, Merbein. Synonymy in *Citrus* (where applicable) is given only for the first specimen of a species listed.

Aurantioideae: ***Aegle marmelos*** (L.) Corrêa ex Roxb.; PI 539142; 2 (1), JN612932 and JN612933 (1a, 1b), JN612891 (1). ***Aeglopsis chevalieri*** Swingle; PI 539143; 2 (2), JN612934 and JN612935 (2a, 2b), JN612892 (2). ***Afraegle paniculata*** (Schumach. & Thonn.) Engl.; PI 103107; 2 (67), JN676193 (67), JN676196 (67). ***Atalantia buxifolia*** (Poir.) Oliv.; PI 539793; 2 (3), -, JN612927 (3). ***A. ceylanica*** (Arn.) Oliv.; 2 (4), 1 (4a, 4b), 1 (4). ***A. citroides*** Pierre ex Guillaumin; PI 539145; 2 (5), -, JN612893 (5). ***A. monophylla*** (L.) DC.; PI 109613; 2 (6), JN612936 and JN612937 (6a, 6b), JN612894 (6). ***Balsamocitrus dawei*** Stapf; PI 539147; 2 (7), JN612938 and JN612939 (7a, 7b), JN612895 (7). ***Bergera koenigii*** L.; PI 539745; 2 (8), JN612940 and JN612941 (8a, 8b), JN612896 (8). ***B. koenigii*** L.; CANB 743217; -, -, JN612897 (9). ***Burkillanthus malaccensis*** (Ridl.) Swingle; FRIM TreeNo 380 (see 2); 2 (10), -, JN612898 (10). ***Citropsis daweana*** Swingle and M. Kell.; PI 247137; 2 (11), -, JN612899 (11). ***C. schweinfurthii***

(Engl.) Swingle and M. Kell.; PI 231240; 2 (12), -, JN612900 (12). ***Citrus gracilis*** Mabb.; CANB 644758; 2 (18), JN612942 and JN612943 (18a, 18b), JN612904 (18). ***C. maxima*** (Burm.) Merrill cv. Kao Pan; PI 539357; -, 1 (20), 1 (20). ***C. maxima*** (Burm.) Merrill cv. Kao Panne; PI 539362; -, 1 (21a, 21b), 1 (21). ***C. maxima*** (Burm.) Merrill cv. Mato Buntan; PI 539398; -, 1 (22a, 22b), 1 (22). ***C. maxima*** (Burm.) Merrill; CANB 743219; 2 (23), -, -. ***C. medica*** L. "South Coast Field Station"; PI 539435; -, 1 (24), 1 (24). ***C. medica*** L. "Arizona 861"; PI 600651; -, 1 (25a, 25b), 1 (25). ***C. medica*** L.; CANB 743231; 2 (26), -, -. ***C. reticulata*** Blanco cv. Tien Chieh; PI 539495; -, 1 (29c, 29d), 1 (29a, 29b). ***C. reticulata*** Blanco cv. Scarlet Emperor; PI 539505; -, 1 (30a, 30b), 1 (30). ***C. reticulata*** Blanco cv. Encore; PI 539511; -, 1 (31c, 31d), 1 (31). ***C. reticulata*** Blanco cv. King; PI 539457; -, 1 (32c, 32d), 1 (32a, 32b). ***C. reticulata*** Blanco cv. Cleopatra; PI 539492; -, 1 (33c, 33d), 1 (33a, 33b). ***C. reticulata*** Blanco cv. Emperor; Merbein CO007; 2 (34), -, -. ***C.* sp. "Amboin, New Guinea"**; Merbein CO054; 2 (13), -, JN612901 and JN612902 (13a, 13b). ***C.* × *aurantium*** L.; Merbein Valencia Clone 2; 2 (71), -, -. ***Citrus* × *aurantium*** L. cv. Washington Navel; PI 539560; -, 1 (66a, 66b), 1 (66c, 66d). ***Clausena excavata*** Burm. f.; PI 235419; 2 (40), -, JN612906 (40). ***C. harmandiana*** (Pierre) Guillaumin; PI 600640; 2 (41), JN612944 and JN612945 (41a, 41b), JN612907 (41). ***Clymenia polyandra*** (Tanaka) Swingle [=*Citrus polyandra* Tanaka]; PI 263640; 2 (28), -, JN612908 (28). ***Eremocitrus glauca*** (Lindl.) Swingle [=*Citrus glauca* (Lindl.) Burkill]; PI 539717; 2 (17), JN612946 (17), JN612903 (17). ***Feronia limonia*** (L.) Swingle; PI 236991; 2 (68), -, KM245160 (68). ***Feroniella oblata*** Swingle (=Citrus); PI 539720; 2 (42), -, 3 (42). ***Glycosmis mauritiana*** (Lam.) Tanaka; PI 600641; 2 (43), -, JN612909 and JN612910 (43a, 43b). ***G. pentaphylla*** Corrêa; CANB 782574; 2 (44), -, JN612911 (44). ***G. trichanthera*** Guillaumin; PI RRUT-12 (see 2); 2 (45), -, JN612912 (45). ***Merrillia caloxylon*** (Ridl.) Swingle; PI 539733; 2 (46), JN612947 and JN612948 (46a, 46b), JN612913 (46). ***Microcitrus australasica*** (F.Muell.) Swingle [=*Citrus australasica* F.Muell.]; CANB 782571; 2 (16), JN612949 (15), -. ***M. australasica*** (F.Muell.) Swingle; PI 312872; -, 1 (14c, 14d), 1 (14a, 14b). ***M. australis*** (Planch.) Swingle [=*Citrus australis* (Sweet) Planch.]; CANB 743271; 2 (69), -, -. ***M. inodora*** (Bailey) Swingle [=*Citrus inodura* F.M.Bailey]; PI 539741; 2 (19), JN612950 (19), -. ***M. papuana*** *Winters [=Citrus wintersii* Mabb.]; PI 410943; 2 (39), -, JN612905 (39). ***Micromelum minutum*** (G.Forst.) Wight and Arn.; PI 600637; 2 (47), -, JN612914 (47). ***Monanthocitrus cornuta*** (Lauterb.) Tanaka; TJ Hoe s.n. (specimen CANB); 2 (48), JN612951 and JN612952 (48a, 48b), JN612915 (48). ***Murraya paniculata*** (L.) Jack; CANB 743224; -, JN612954 and JN612955 (49a, 49b), JN612916 (49). ***M. paniculata*** (L.) Jack; Merbein CR013; 2 (51), -, JN612917 (51). ***M. paniculata*** (L.) Jack; Antonelli 830 (GB); -, JN612953 (50), JN612918 (50). ***Naringi crenulata*** (Roxb.) Nicolson; PI 539748; 2 (52), 1 (52a, 52b), JN612920 (52). ***Oxanthera neocaledonica*** (Guillaumin) Tanaka [=*Citrus neocaledonica* Guillaumin]; PI 539671; 2 (27), JN612956 (27), JN612921 (27). ***O.* sp. nov.** sensu TG Hartley; Veillon 7758 (specimen CANB); 2(35), -, JN612922 (35). ***Pamburus missionis*** (Wall. ex Wight) Swingle; PI 095350; 2 (54), JN612957 and JN612958 (53a, 53b), JN612923 (53). ***Paramignya lobata*** Burkill; PI 600642; 2 (55), JN612959 and JN612960 (55a, 55b), JN612924 (55). ***P. scandens*** (Griff.) Craib; PI 109758; 2 (56), -, JN612925 (56). ***Pleiospermium latialatum*** Swingle; PI 600643; 2 (57), -, JN612926 (57). ***Poncirus trifoliata*** (L.) Raf. [=*Citrus trifoliata* L.] cv. Flying Dragon; PI 539768; -, 1 (36c, 36d), 1 (36a, 36b). ***P. trifoliata*** (L.) Raf. cv. Pomeroy; -, 1 (37c, 37d), 1 (37a, 37b). ***P. trifoliata*** (L.) Raf.; CANB 743205; 2 (38), -, -. ***Swinglea glutinosa*** (Blanco) Merr.; PI 142571; 2 (59), 1 (58a, 58b), JN612928 (58). ***Triphasia trifolia*** (Burm. f.) P. Wils.; PI 539800; 2 (60), -, JN612929 (60). ***Wenzelia dolichophylla*** (Lauterb and K.Schum.) Tanaka; PI 277441; 2 (61), -, JN612930 (61).

Outgroups: ***Choisya ternata*** Kunth; CANB 743233; 2 (62), -, JN676195 (62). ***Flindersia australis*** R.Br.; CANB 743207; 2 (63), -, JN676194 (63). ***Ruta graveolens*** L.; Antonelli 831 (GB); -, -, JN612931 (64). ***R. graveolens*** L.; Bayer, R. J. GH06-114 (CANB living collection); 2 (65), -, -.