# Development of Novel, Exon-Primed Intron-Crossing (EPIC) Markers from EST Databases and Evaluation of their Phylogenetic Utility in Commiphora (Burseraceae)

Authors: Gostel, Morgan R., and Weeks, Andrea

Source: Applications in Plant Sciences, 2(4)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1300098

Applications in Plant Sciences 2014 2(4): 1300098

PRIMER NOTE

# DEVELOPMENT OF NOVEL, EXON-PRIMED INTRON-CROSSING (EPIC) MARKERS FROM EST DATABASES AND EVALUATION OF THEIR PHYLOGENETIC UTILITY IN *COMMIPHORA* (BURSERACEAE)[1]

MORGAN R. GOSTEL[2,3] AND ANDREA WEEKS[2]

[2]George Mason University, 4400 University Drive, MSN 5F2, Fairfax, Virginia 22030-4444 USA

- *Premise of the study:* Novel nuclear exon-primed intron-crossing (EPIC) markers were developed to increase phylogenetic resolution among recently diverged lineages in the frankincense and myrrh family, Burseraceae, using *Citrus*, *Arabidopsis*, and *Oryza* genome resources.
- *Methods and Results:* Primer pairs for 48 nuclear introns were developed using the genome resource IntrEST and were screened using species of *Commiphora* and other Burseraceae taxa. Four putative intron regions (*RPT6A*, *BXL2*, *mtATP Synthase D*, and *Rab6*) sequenced successfully for multiple taxa and recovered phylogenies consistent with those of existing studies. In some cases, these regions yielded informative sequence variation on par with that of the nuclear ribosomal DNA internal transcribed spacer.
- *Conclusions:* The combination of freely available genome resources and our design criteria have uncovered four single-copy nuclear intron regions that are useful for phylogenetic reconstruction of Burseraceae taxa. Because our EPIC primers also amplify *Arabidopsis*, we recommend their trial in other rosid and eudicot lineages.

**Key words:** Burseraceae; *Commiphora*; EPIC markers; shallow-scale phylogenetics.

Resolving phylogenetic relationships among closely related angiosperm species is frequently hindered due to limited variation in currently available markers (Li et al., 2008; Zimmer and Wen, 2012). This challenge is no less problematic in the myrrh genus, *Commiphora* Jacq. (Burseraceae), where complete, species-level resolution has not been achieved despite the use of multiple gene regions (Weeks and Simpson, 2007). We describe the development and evaluation of four novel, exon-primed intron-crossing (EPIC) markers for Burseraceae (Sapindales) using a repository of putative, intron-flanking nuclear unigenes from 43 plant taxa and two complete reference genomes (IntrEST; Ilut and Doyle, 2012). Markers were evaluated for their phylogenetic utility at the species level using a recently radiated lineage of *Commiphora* and a generic-level sampling in Burseraceae. Sequence variation from these novel markers was compared to existing nuclear markers and shows promise for resolving relationships at both shallow and deeper phylogenetic scales.

[1]Manuscript received 19 December 2013; revision accepted 22 January 2014.

This material is based in part upon work supported by a dissertation research fellowship to M.R.G. from the Department of Environmental Science and Policy at George Mason University and by the National Science Foundation under grant number 0919179 to A.W. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation. Publication of this article was funded in part by the George Mason University Libraries Open Access Publishing Fund.

[3]Author for correspondence: mgostel2@gmu.edu

doi:10.3732/apps.1300098

## METHODS AND RESULTS

***Marker development*—**Development of EPIC markers for Burseraceae involved unigene data sets of *Citrus clementina* hort. ex Tanaka and *C. sinensis* (L.) Osbeck (Rutaceae; Sapindales) and two reference genomes available in IntrEST, *Oryza sativa* L. and *Arabidopsis thaliana* (L.) Heynh. We developed 12 primer pairs for putative introns from each of four predicted amplicon size categories (200-bp increments between 400–1200 bp), resulting in 48 total primer pairs. For each size category, six primer pairs were developed from a percent-identity criterion of either 80–89.9% or 90–100% between the unigene and the corresponding reference. We predicted that the lower percent-identity criterion (80–89.9%) might yield more informative variation among closely related species. Half of the primer pairs were generated using *C. clementina* and the other half from *C. sinensis* unigenes. Primer sequences were a consensus between unigene and the corresponding reference genome. Primers were preferentially designed using *A. thaliana*. Primers were designed between 18–30 bp, within 50 bp of putative intron splice regions in the reference genome, having a melting temperature ($T_m$) between 51–74°C, without predicted dimers, and a 35–60% G-C content. Primer characteristics were evaluated using Oligo-Evaluator (Sigma-Aldrich, St. Louis, Missouri, USA). Exceptions for $T_m$ and %GC were made for 18 primers where it was not possible to meet all necessary criteria (Appendix 1). Each primer pair was tested by its ability to amplify a single PCR product from two species of Madagascan *Commiphora* (*C. lamii* H. Perrier and *C. ankaranensis* (J.-F. Leroy) Cheek & Rakot.) and a positive control (*A. thaliana*) and to sequence cleanly.

***Taxonomic sampling and molecular methods*—**Markers that passed all of the above criteria were evaluated using 14 species of Burseraceae (Appendix 2), including eight *Commiphora* ingroup species and six outgroup species from closely (*Bursera* Jacq. ex L., *Aucoumea* Pierre) and distantly (*Boswellia* Roxb. ex Colebr., *Protium* Burm. f., and *Beiselia* Forman) related genera, respectively (Weeks et al., 2005; Thulin et al., 2008). All ingroup taxa are Madagascan, and seven correspond to one of two species-rich clades in Madagascar. We sampled densely from one clade to test phylogenetic utility at shallow scales. Whole genomic DNA was extracted from specimens using the FastPrep FastDNA Spin Kit (Bio101 Systems, La Jolla, California, USA). All markers were amplified in

 

25-µL PCR reactions including: 0.5 µL forward and reverse primers (5 µM), 0.5 µL spermidine (4 mM), 2 µL total DNA, and 5 µL GoTaq Green Master Mix (Promega Corporation, Madison, Wisconsin, USA). A ramp-up PCR thermocycler protocol followed a 4-min presoak at 94°C with 35 cycles of 30 s at 94°C (denaturation), 1 min at 48–56°C (annealing), and 50 s at 72°C (extension), followed by a 4-min postsoak at 72°C. PCR products were purified prior to sequencing reactions using ExoSAP (USB Corporation, Cleveland, Ohio, USA) and sequenced by Macrogen (Rockville, Maryland, USA) using an ABI 3730XL Analyzer with BigDye Terminator version 3.1 (Applied Biosystems, Foster City, California, USA). Sequencing reactions (10 µL) for both directions included 40 ng/µL template. Products were assembled and edited using Sequencher 4.7 (Gene Codes Corporation, Ann Arbor, Michigan, USA).

***Phylogenetic analyses*—**Multiple sequence alignment (MSA) was performed using MUSCLE version 3.7 (Edgar, 2004). Gap regions in the MSA were treated as missing data. Markers were evaluated using maximum parsimony (MP) and Bayesian inference (BI). MP analyses were conducted using PAUP* 4.0b10 (Swofford, 2002) with a two-step protocol modified from Plunkett et al. (2005). Branch support for internal nodes was inferred by bootstrapping 1000 replicates in PAUP*. BI analyses were performed using MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003). Two runs were performed for each data set using the best-fitting model as determined by jModelTest (Posada, 2008) consisting of four chains each for 10 million generations sampled every 1000 generations; 10% sampling was discarded as burn-in for each run. MSA and BI analyses were performed in the CIPRES Science Gateway (Miller et al., 2010).

***Marker evaluation*—**Fifteen of the 48 EPIC primer pairs (31%) amplified at least one species, and four pairs (8%; 10F-10R, 16F-16R, 39F-39R, and 43F-43R) produced amplicons that sequenced cleanly for multiple taxa. Provisional marker names are provided based on gene ontology categories from reference taxa (Table 1). When searched in BLAST, sequences of putative intron regions for *RPT6A*, *BXL2*, and *Rab6* (Appendix 2) matched gene ontology categories predicted for the *Arabidopsis* and *Oryza* references in IntrEST. Sequence products for *mtATP Synthase D* (Appendix 2) did not BLAST to predicted gene ontology categories. Sequences produced by this study have been deposited into GenBank (Appendix 2). Sequence alignment files are deposited in the Dryad Digital Repository (http://doi.org/10.5061/dryad.382p0; Gostel and Weeks, 2014). Phylogenetic statistics of new EPIC markers are presented in comparison (Table 1) with those from nuclear markers developed for previous phylogenetic studies of Burseraceae (ETS: Weeks and Simpson, 2007; ITS: Gostel and Weeks, unpublished). Phylogenetic analysis of EPIC markers developed in this study recovered well-resolved phylogenies consistent with those from previous studies (Fig. 1). The concatenated set of all four EPIC markers resulted in improved phylogenetic resolution compared to previously developed markers (Fig. 1).

***Critical assessment of primer design criteria*—**Each of the 15 primer pairs that amplified at least one species spanned the range of melting temperatures (51–74°C), differed from their pair by less than 10°C in $T_m$, and were developed from both *Citrus* unigene data sets and both reference genomes. More than half (9/15) of these markers were designed using 80–89.9% identity criteria, yet only two (16F-16R and 43F-43R) sequenced cleanly for multiple taxa. Two of the six primer pairs developed from the 90–100% identity criterion (10F-10R and 39F-39R) sequenced cleanly for multiple taxa and yielded the most informative variation among *Commiphora* species. These results do not support predictions that lower percent identity would provide better shallow-scale phylogenetic resolution, which suggests mutation rates between exon and intron regions are independent.

## CONCLUSIONS

The EPIC markers developed in this study may also be useful for phylogenetic reconstruction in other angiosperm taxa. Most primer pairs amplified *A. thaliana* (Brassicales), and they may work in other rosid or eudicot taxa. Of the four markers, *RPT6A* is most promising for further evaluation. This ca. 400-bp region sequenced cleanly for all Burseraceae taxa and yielded a percentage of phylogenetically informative characters on par with ITS. Our study demonstrates how genomic resources from

TABLE 1. Marker names, primer sequences, and phylogenetic statistics for the novel nuclear EPIC markers and the benchmark nuclear ribosomal DNA ETS and ITS regions.

| Provisional marker name[a] | Locus | Primer sequence (5′–3′) | Ingroup statistics[b] | Family-wide statistics[b] | % Missing data |
|---|---|---|---|---|---|
| *RPT6A* Intron | 10F | CTCCARCACATYCAYGARCTCCAGC | (454, 1.1, 0.95, 0.6, 0.569) | (454, 11.8, 0.91, 0.84, 0.764) | 4.6 |
| | 10R | AGCTGTAAYTCTTCTYTRAGCATCC | | | |
| *BXL2* Intron | 16F | CTTGTGGGAAKCCATCGGAC | (1,049, 0.6, 0.96, 0.625, 0.601) | (1,049, 9.4, 0.916, 0.8, 0.761) | 17.9 |
| | 16R | CGTTGTACATKGCYCTKGCYTCA | | | |
| *mtATP Synthase D* Intron | 39F | TCCTYCCYTACRCMTCTGAGC | (1,600, 0.2, 0.991, 0.8, 0.792) | (1,600, 5.4, 0.976, 0.864, 0.844) | 46.7 |
| | 39R | GTTGATGCKGGAAYKATRACCA | | | |
| *Rab6* Intron | 43F | CCTTCAACAGATACAACAACATGCA | (984, 2.4, 0.974, 0.939, 0.915) | (984, 8.4, 0.979, 0.936, 0.916) | 32.6 |
| | 43R | TCCATGYCCCCACATATGCA | | | |
| ETS | ETS1F | TTCGGTATCCTGTGTTGCTTAC | (389, 4.4, 0.85, 0.4, 0.34) | (389, 13.9, 0.723, 0.6, 0.435) | 2.6 |
| | 18S2R | GAGACAAGCATATGACTACTGGCAGGATCAACCAG | | | |
| ITS | ITSny183 | CCTTATCATTTAGAGGAAGGAG | (850, 1.9, 0.878, 0.52, 0.456) | (850, 11, 0.809, 0.639, 0.517) | 12.2 |
| | ITSny109Com | GWGACACCCAGGCAGACG | | | |

[a] Provisional marker names correspond to the predicted gene ontology category for the reference genome (*Arabidopsis thaliana* or *Oryza sativa*) that most closely matches unigene sequences identified in IntrEST.

[b] Phylogenetic statistics are reported for the ingroup (*Commiphora* spp.) and family-wide (all Burseraceae spp.) sampling and correspond to aligned length, percent parsimony informative characters, consistency index, retention index, and corrected retention index, respectively.

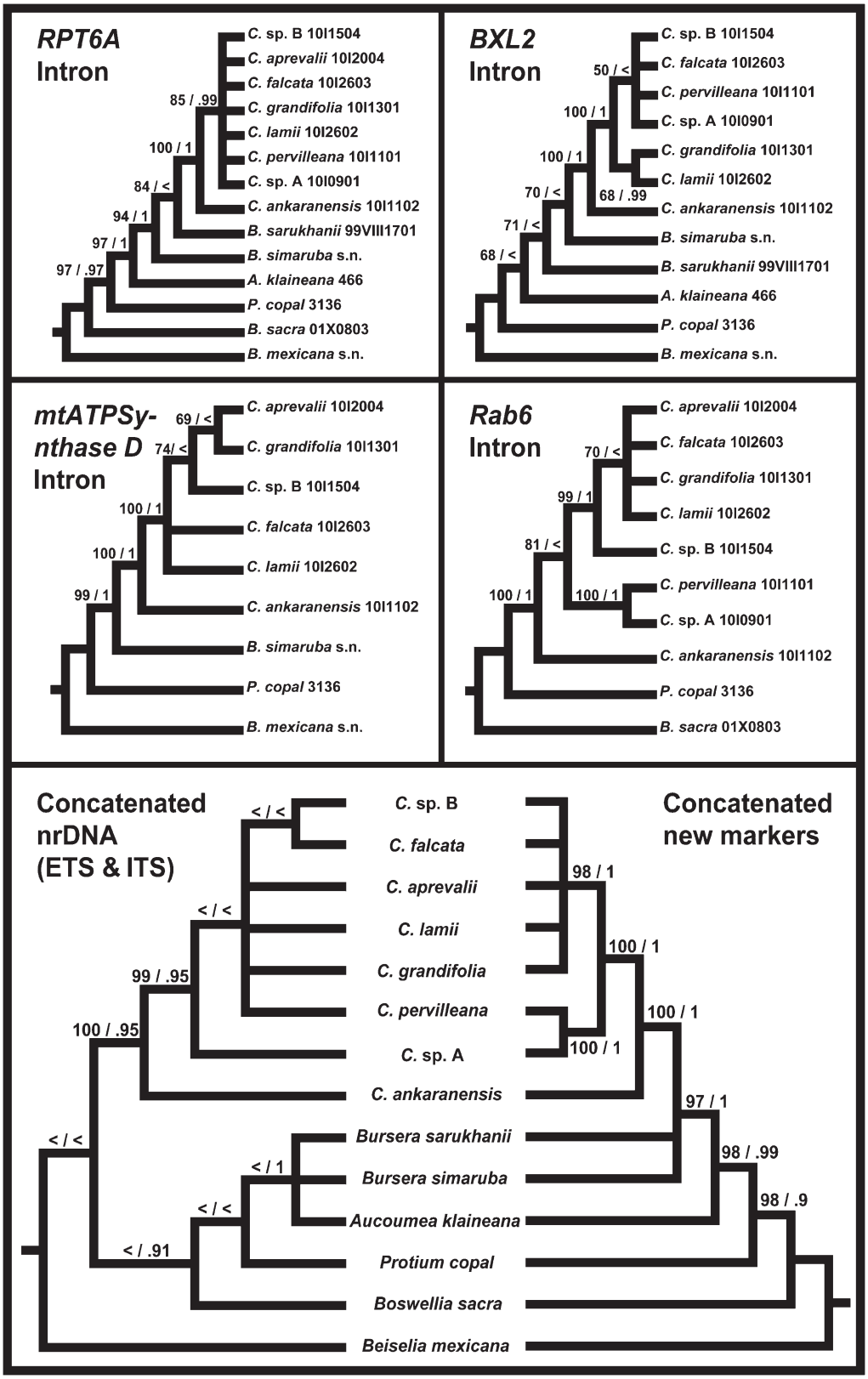


Fig. 1. Phylogeny of 14 representative taxa in the Burseraceae sampled in this study. Values above branches correspond to maximum parsimony bootstrap support values, followed by Bayesian posterior probabilities. "Concatenated new markers" refers to a concatenated data set of all four new markers. Refer to Table 1 for individual marker statistics.

model organisms can be leveraged to advance the phylogenetic systematics of nonmodel organisms.

## LITERATURE CITED

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

Gostel, M. R., and A. Weeks. 2014. Data from: Development of novel, exon-primed intron-crossing (EPIC) markers from EST databases and evaluation of their phylogenetic utility in *Commiphora* (Burseraceae). Dryad Digital Repository. http://doi.org/10.5061/dryad.382p0.

Ilut, D. C., and J. J. Doyle. 2012. Selecting nuclear sequences for fine detail molecular phylogenetic studies in plants: A computation approach and sequence repository. *Systematic Botany* 37: 7–14.

Li, M., J. Wunder, G. Bissoli, E. Scarponi, S. Gazzani, E. Barbaro, H. Saedler, and C. Varotto. 2008. Development of COS genes as universally amplifiable markers for phylogenetic reconstructions of closely related plant species. *Cladistics* 24: 727–745.

Miller, M. A., W. Pfeiffer, and T. Schwartz. 2010. Creating the CIPRES Science Gateway for inference of large phylogenetic trees. *In* Proceedings of the Gateway Computing Environments Workshop (GCE), 14 November 2010, New Orleans, Louisiana, 1–8.

Plunkett, G. M., P. P. Lowry II, D. G. Frodin, and J. Wen. 2005. Phylogeny and geography of *Schefflera*: Pervasive polyphyly in the largest genus of Araliaceae. *Annals of the Missouri Botanical Garden* 92: 202–224.

Posada, D. 2008. jModeltest: Phylogenetic model averaging. *Molecular Biology and Evolution* 25: 1253–1256.

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

Swofford, D. L. 2002: PAUP* 4.0b10. Phylogenetic analysis using parsimony (*and other methods). Sinauer, Sunderland, Massachusetts, USA.

Thulin, M., B.-A. Beier, S. G. Razafimandimbison, and H. I. Banks. 2008. *Ambilobea*, a new genus from Madagascar, the position of *Aucoumea*, and comments on the tribal classification of the frankincense and myrrh family (Burseraceae). *Nordic Journal of Botany* 26: 218–229.

Weeks, A., D. C. Daly, and B. B. Simpson. 2005. The phylogenetic history and biogeography of the frankincense and myrrh family (Burseraceae) based on nuclear and chloroplast sequence data. *Molecular Phylogenetics and Evolution* 35: 85–101.

Weeks, A., and B. B. Simpson. 2007. Molecular phylogenetic analysis of *Commiphora* (Burseraceae) yields insight on the evolution and historical biogeography of an "impossible" genus. *Molecular Phylogenetics and Evolution* 42: 62–79.

Zimmer, E. A., and J. Wen. 2012. Using nuclear gene data for plant phylogenetics: Progress and prospects. *Molecular Phylogenetics and Evolution* 65: 774–785.

Appendix 1. List of all 48 primer pairs developed in this study and their characteristics.

| Locus[a] | Primer sequence (5′–3′) | $T_m$ | %GC | Nbases | Reference taxon[b] | EST[c] | %ID[d] | *Citrus* sp.[e] |
|---|---|---|---|---|---|---|---|---|
| 1F† | GATCWGARATCGCMGARGAAGTYCGC | 60.7 | 46.2 | 26 | *Arabidopsis* | AT3G12290.1 | 85 | sinensis |
| 1R† | TCAGCRCAMGCYTTYCTCTTCRTRYTC | 74 | 37.1 | 27 | *Arabidopsis* | AT3G12290.1 | 85 | sinensis |
| 2F† | TGCAARTCTCTYKTTGCYGG | 65.3 | 40 | 20 | *Arabidopsis* | AT4G01100.1 | 83 | sinensis |
| 2R† | TCCARWGGWGCAACAGCA | 61.1 | 50 | 18 | *Arabidopsis* | AT4G01100.1 | 83 | sinensis |
| 3F | TACACRTATGCWAGRTGCAC | 63.6 | 40 | 20 | *Arabidopsis* | AT1G48410.2 | 87 | sinensis |
| 3R | GCTGCWAGRTGKGCATARTATGC | 65.9 | 43.5 | 23 | *Arabidopsis* | AT1G48410.2 | 87 | sinensis |
| 4F† | ATTCTYGTCYTGTCCGSWAGAGA | 61 | 39.2 | 23 | *Arabidopsis* | AT4G21960.1 | 83 | clementina |
| 4R† | CCATCTCTYCTTCCTGTCTT | 51.2 | 45 | 20 | *Arabidopsis* | AT4G21960.1 | 83 | clementina |
| 5F | GCTGGAYTMACSSTYGAYCATCC | 64.3 | 39.2 | 23 | *Arabidopsis* | AT1G22410.1 | 84 | clementina |
| 5R | TCATGRGAWGTCCAGAASTC | 67.7 | 40 | 20 | *Arabidopsis* | AT1G22410.1 | 84 | clementina |
| 6F | ATGCTKTTTGGTGCWRTTGGGAC | 74 | 43.5 | 23 | *Arabidopsis* | AT5G27150.1 | 82 | clementina |
| 6R | GAACTGWRTWACACCTAGWGATATGA | 57.3 | 34.7 | 26 | *Arabidopsis* | AT5G27150.1 | 82 | clementina |
| 7F | GCAKCACCRAAGATGYTGAAC | 62.4 | 42.9 | 21 | *Arabidopsis* | AT1G34130.1 | 91 | sinensis |
| 7R | CAGCTCWCCRAAYCKRTARTAKSATA | 60 | 27 | 26 | *Arabidopsis* | AT1G34130.1 | 91 | sinensis |
| 8F | GCTGTTGCSYTGAARCAGGC | 64.8 | 50 | 20 | *Arabidopsis* | AT4G13930.1 | 92 | sinensis |
| 8R | CTTGTTTYGCRTAGRCCTTGAA | 60 | 36.4 | 22 | *Arabidopsis* | AT4G13930.1 | 92 | sinensis |
| 9F† | ATATCARGGTGCYTACAAGA | 57.1 | 35 | 20 | *Arabidopsis* | AT5G50850.1 | 90 | sinensis |
| 9R† | CTCAGGRCCATATTTCTCCAA | 58.1 | 42.9 | 21 | *Arabidopsis* | AT5G50850.1 | 90 | sinensis |
| 10F† | CTCCARCACATYCAYGARCTCCAGC | 55.6 | 48 | 25 | *Arabidopsis* | AT5G19990.1 | 93 | clementina |
| 10R† | AGCTGTAAYTCTTCTYTRAGCATCC | 61.7 | 36 | 25 | *Arabidopsis* | AT5G19990.1 | 93 | clementina |
| 11F | ACGMCTYGAYATGGATTACGTYGA | 54.8 | 37.5 | 24 | *Arabidopsis* | AT1H04690.1 | 90 | clementina |
| 11R* | TTCATCGCCCTCACMGTCTCYTC | 77 | 52.2 | 23 | *Arabidopsis* | AT1H04690.1 | 90 | clementina |
| 12F† | CGTYGAYGTKMTYTATTGYCACA | 57.7 | 30.5 | 23 | *Arabidopsis* | AT1G04690.1 | 90 | clementina |
| 12R† | TCATCGCCCTCACMGTCTC | 67.1 | 57.9 | 19 | *Arabidopsis* | AT1G04690.1 | 90 | clementina |
| 13F† | ACAAGCCWCCTGAAGATGC | 62 | 52.7 | 19 | *Arabidopsis* | AT4G37510.1 | 87 | sinensis |
| 13R† | GTCCAAGTTCRATRTTYCTWGCTTC | 54.5 | 36 | 25 | *Arabidopsis* | AT4G37510.1 | 87 | sinensis |
| 14F† | TTAYTCAATGTTCAACAGA | 57.6 | 26.4 | 19 | *Arabidopsis* | AT4G02580.1 | 86 | sinensis |
| 14R† | CACGKAYCATRCAAGGTGTTGTGCC | 62.7 | 48 | 25 | *Arabidopsis* | AT4G02580.1 | 86 | sinensis |
| 15F | GCTYTWCCTTCRGAKACTGGTC | 57.3 | 45.5 | 22 | *Arabidopsis* | AT5G37850.2 | 88 | sinensis |
| 15R | GTACWGARTKGATTGGATCCAC | 58.5 | 41 | 22 | *Arabidopsis* | AT5G37850.2 | 88 | sinensis |
| 16F† | CTTGTGGGAAKCCATCGGAC | 66.3 | 55 | 20 | *Arabidopsis* | AT1G02640.1 | 84 | clementina |
| 16R† | CGTTGTACATKGCYCTKGCYTCA | 64.9 | 43.5 | 23 | *Arabidopsis* | AT1G02640.1 | 84 | clementina |
| 17F† | CAAGARGCKKTTTGTCGCC | 65.8 | 47.4 | 19 | *Arabidopsis* | AT1G62050.1 | 83 | clementina |
| 17R† | CCAAGCKRARCGGTGGTGA | 57.6 | 52.7 | 19 | *Arabidopsis* | AT1G62050.1 | 83 | clementina |
| 18F | TTGAGTTRTCTCSWGAAGC | 57.2 | 36.9 | 19 | *Arabidopsis* | AT3G07160.1 | 87 | clementina |
| 18R | GCARTGCAATRTCARCAGC | 55.5 | 42.2 | 19 | *Arabidopsis* | AT3G07160.1 | 87 | clementina |
| 19F | TGAYCTYCTTGATGCRTTGGAC | 62.8 | 41 | 22 | *Arabidopsis* | AT5G11170.2 | 95 | sinensis |
| 19R | GCATATWGARGGRAAATTGCATTC | 55 | 33.4 | 24 | *Arabidopsis* | AT5G11170.2 | 95 | sinensis |
| 20F | AGTTTRCTCTCTGTTGATCCRAC | 51.9 | 39.2 | 23 | *Arabidopsis* | AT2G25660.1 | 93 | sinensis |
| 20R | GCTGMACTTCAACTTCYGTWCCA | 56.8 | 43.5 | 23 | *Arabidopsis* | AT2G25660.1 | 93 | sinensis |
| 21F | GGAAYTWAGGGAAGAATGC | 57.6 | 42.2 | 19 | *Arabidopsis* | AT4G32180.3 | 90 | sinensis |
| 21R | GCATCAASAAAYTGGAAYTC | 67.8 | 30 | 20 | *Arabidopsis* | AT4G32180.3 | 90 | sinensis |
| 22F | GATGGCTCGTGAAAGTGCTC | 65 | 55 | 20 | *Arabidopsis* | AT2G27600.1 | 91 | clementina |
| 22R | CCACGYKKACCACAYAARGAATC | 56.6 | 39.2 | 23 | *Arabidopsis* | AT2G27600.1 | 91 | clementina |
| 23F | GATGCRTTGGACTTYAATCAA | 58.6 | 33.4 | 21 | *Arabidopsis* | AT5G11170.2 | 95 | clementina |
| 23R | GACATKCCAGARTGGATGCATA | 57.7 | 41 | 22 | *Arabidopsis* | AT5G11170.2 | 95 | clementina |
| 24F | AGCTTYTAGCSGACAATGC | 59.8 | 42.2 | 19 | *Arabidopsis* | AT5G26680.2 | 90 | clementina |
| 24R | GTAAATGCTCATGCTAGCATCAA | 62.9 | 39.2 | 23 | *Arabidopsis* | AT5G26680.2 | 90 | clementina |
| 25F | GACAAGGTTCTCATGGARAGC | 54.4 | 47.7 | 21 | *Arabidopsis* | AT5G54160.1 | 80 | sinensis |
| 25R | CCACCWTCAAGMAYTGCATC | 69.9 | 45 | 20 | *Arabidopsis* | AT5G54160.1 | 80 | sinensis |
| 26F | TGGACACTTCGAGGRCTTTG | 60.4 | 50 | 20 | *Arabidopsis* | AT1G67060.1 | 86 | sinensis |
| 26R | ACCCATATKACRGCGAGGA | 56.1 | 47.4 | 19 | *Arabidopsis* | AT1G67060.1 | 86 | sinensis |
| 27F | CTGTAAYCARGACAACCGCGTYAC | 57.8 | 45.9 | 24 | *Arabidopsis* | AT5G21090.1 | 87 | sinensis |
| 27R | AGRTTTGAATTWCCCAAATC | 54.8 | 30 | 20 | *Arabidopsis* | AT5G21090.1 | 87 | sinensis |
| 28F | TGGCTKGGWCARAAYCAGRTTC | 54.8 | 41 | 22 | *Arabidopsis* | AT5G11480.1 | 86 | clementina |
| 28R | CATCWAYTTGTCGCATTTKGTGAA | 66.7 | 33.4 | 24 | *Arabidopsis* | AT5G11480.1 | 86 | clementina |
| 29F | CTGGTTGTTGCTGATGA | 56.3 | 47.1 | 17 | *Arabidopsis* | AT3G54300.2 | 88 | clementina |
| 29R | TCCTTGACYCGTTCGAGA | 59.7 | 50 | 18 | *Arabidopsis* | AT3G54300.2 | 88 | clementina |
| 30F | TCGTYATWGCCTCCCTCGACGTTC | 64.7 | 54.2 | 24 | *Arabidopsis* | AT4G16600.1 | 83 | clementina |
| 30R | CACYACYTTWGCTCCATCTTCYTSTTC | 71.8 | 37.1 | 27 | *Arabidopsis* | AT4G16600.1 | 83 | clementina |
| 31F | GATGCTTTTGAATTCATTGTA | 56.8 | 28.6 | 21 | *Arabidopsis* | ATMG00285.1 | 94 | sinensis |
| 31R | CATGGCAATTAAATCATRAGCCGA | 62.4 | 37.5 | 24 | *Arabidopsis* | ATMG00285.1 | 94 | sinensis |
| 32F | GATCAGGTYCGTGGTGTMATGGA | 55.9 | 47.9 | 23 | *Oryza* | 13101.m04144 | 94 | sinensis |
| 32R | CATTTGGCTYTCYCCATA | 56.1 | 38.9 | 18 | *Oryza* | 13101.m04144 | 94 | sinensis |
| 33F† | GTCGGCAAYCTCGAYCCCCA | 71.1 | 60 | 20 | *Oryza* | 13110.m02788 | 93 | sinensis |
| 33R† | TCCCAWAGTARCTCCTCMGWAA | 51.4 | 41 | 22 | *Oryza* | 13110.m02788 | 93 | sinensis |
| 34F* | TCTCCAGAATACCGCAGGCAGCAAC | 74.5 | 56 | 25 | *Arabidopsis* | AT1G03150.1 | 97 | clementina |
| 34R | CACAAAGTARGCYTTATCAA | 55.4 | 30 | 20 | *Arabidopsis* | AT1G03150.1 | 97 | clementina |
| 35F | GTTGGSTGGTAYCAYTCACA | 68.5 | 40 | 20 | *Oryza* | 13104.m05825 | 91 | clementina |

Appendix 1. Continued.

| Locus[a] | Primer sequence (5′–3′) | $T_m$ | %GC | Nbases | Reference taxon[b] | EST[c] | %ID[d] | *Citrus* sp.[e] |
|---|---|---|---|---|---|---|---|---|
| 35R | CAATRCCYGAWARCCAGCATC | 51.5 | 42.9 | 21 | *Oryza* | 13104.m05825 | 91 | clementina |
| 36F | ACCGGTGTCAAGAGRYTSTA | 62.4 | 40 | 20 | *Oryza* | 13111.m02571 | 93 | clementina |
| 36R | GTGACAGAGTCATTGACATTGA | 60.2 | 41 | 22 | *Oryza* | 13111.m02571 | 93 | clementina |
| 37F | TACAAGCTTWYKGGCATCAAG | 58.3 | 38.1 | 21 | *Arabidopsis* | AT2G38110.1 | 85 | sinensis |
| 37R | ACCACAGGRTCKARAACRGTGC | 60.1 | 45.5 | 22 | *Arabidopsis* | AT2G38110.1 | 85 | sinensis |
| 38F | AGGGTYAAGAATCCAGAATGG | 55.4 | 42.9 | 21 | *Arabidopsis* | AT5G13430.1 | 82 | sinensis |
| 38R | GCATTWGGYAARGGRATGCACC | 54.4 | 45.5 | 22 | *Arabidopsis* | AT5G13430.1 | 82 | sinensis |
| 39F† | TCCTYCCYTACRCMTCTGAGC | 65.3 | 47.7 | 21 | *Arabidopsis* | AT5G47030.1 | 81 | sinensis |
| 39R† | GTTGATGCKGGAAYKATRACCA | 57.1 | 36.4 | 22 | *Arabidopsis* | AT5G47030.1 | 81 | sinensis |
| 40F | GAATTYGTGATCTCYAARKTSGATG | 53.6 | 28 | 25 | *Arabidopsis* | AT5G11770.1 | 88 | clementina |
| 40R | CATRGGCCAGATSGAKCCGSKACGA | 64.9 | 48 | 25 | *Arabidopsis* | AT5G11770.1 | 88 | clementina |
| 41F | GAAGAYTCKGTYAGGGTYAAGAA | 54.2 | 34.8 | 23 | *Arabidopsis* | AT5G13430.1 | 82 | clementina |
| 41R | CAGCATTWGGYAARGGRATGCACC | 58 | 45.9 | 24 | *Arabidopsis* | AT5G13430.1 | 82 | clementina |
| 42F | GCTGAAATYGCTKCTGGAAGTGA | 57.7 | 43.5 | 23 | *Arabidopsis* | AT5G47840.1 | 81 | clementina |
| 42R | TCAGGKACCAAYTGTCCTTTCTCCA | 71.5 | 44 | 25 | *Arabidopsis* | AT5G47840.1 | 81 | clementina |
| 43F† | GAACAAAACTGATCTTGTKGACAA | 58.5 | 33.4 | 24 | *Oryza* | 13107.m03172 | 93 | sinensis |
| 43R† | CCAGCYTTRGCACTRGTYTCAA | 64.9 | 41 | 22 | *Oryza* | 13107.m03172 | 93 | sinensis |
| 44F† | CCTTCAACAGATACAACAACATGCA | 66.7 | 40 | 25 | *Arabidopsis* | AT3G57670.1 | 96 | sinensis |
| 44R† | TCCATGYCCCCACATATGCA | 64.7 | 50 | 20 | *Arabidopsis* | AT3G57670.1 | 96 | sinensis |
| 45F | GCGAGARAARTCAGCTGAYCCA | 59.5 | 45.5 | 22 | *Oryza* | 13102.m04682 | 95 | sinensis |
| 45R | GCAGTCCAYTTRAATATGTTKGAATC | 59.4 | 30.8 | 26 | *Oryza* | 13102.m04682 | 95 | sinensis |
| 46F | AGGCAAGTMTCMATAGAGGA | 55 | 40 | 20 | *Oryza* | 13107.m03172 | 92 | clementina |
| 46R | CCAGCYTTRGCACTRGTYTCAA | 64.9 | 41 | 22 | *Oryza* | 13107.m03172 | 92 | clementina |
| 47F | TGAGACAGGGTGTWCTTGGYATYAA | 52.7 | 40 | 25 | *Oryza* | 13103.m04131 | 92 | clementina |
| 47R | GGATKGTTACRAGATCMGGYAGAG | 53.9 | 41.7 | 24 | *Oryza* | 13103.m04131 | 92 | clementina |
| 48F | CAGCTGAYCCAGAYATYCARTTA | 54 | 34.8 | 23 | *Oryza* | 13102.m04682 | 95 | clementina |
| 48R | GCAGTCCAYTTRAATATGTTKGAATC | 59.4 | 30.8 | 26 | *Oryza* | 13102.m04682 | 95 | clementina |

*Note*: %GC = percent G-C content; Nbases = number of nucleotides that comprise the primer; $T_m$ = melting temperature.

[a] Primer names with asterisks (*) indicate primers that did not meet the necessary melting temperature criteria; † indicates primer pairs that were able to successfully amplify in at least one specimen.

[b] The model organism reference in IntrEST from which the primer was developed.

[c] The expressed sequence tag record number that was used to develop the marker.

[d] Percent shared identity between the reference taxon and *Citrus* species.

[e] Species of *Citrus* (*C. sinensis* or *C. clementina*) that was used to develop the primer.

Appendix 2. List of species, vouchers, and geographic origin with GenBank accession numbers for all putative gene regions. GPS coordinates were not included with some of the herbarium vouchers, which is reflected when no coordinate is given.

*Beiselia mexicana* Forman; Pell s.n. (TEX). Mexico. ETS: FJ233929, ITS: JF919030, *RPT6A* Intron: KF906106, *BXL2* Intron: KF906094, *mtATP Synthase D* Intron: KF906084.

*Protium copal* (Schltdl. & Cham.) Engl.; Killeen 3136 (MO). Mexico. 15°15′S, 067°00′W. ETS: AY964612, ITS: KF906073, *RPT6A* Intron: KF906108, *BXL2* Intron: KF906095, *mtATP Synthase D* Intron: KF906085, *Rab6* Intron: KF906120.

*Aucoumea klaineana* Pierre; Walters 466 (MO). Gabon. 00°07′12″S, 011°42′57″E. ETS: KF906082, *RPT6A* Intron: KF906105, *BXL2* Intron: KF906093.

*Boswellia sacra* Flueck.; Weeks 01-X-08-03 (TEX). N.E. Africa. ETS: AF445957, ITS: AF455880, *RPT6A* Intron: KF906107, *Rab6* Intron: KF906119.

*Bursera sarukhanii* Guevara & Rzed.; Weeks 00-VIII-18-06 (TEX). Mexico. ETS: AY315051, ITS: AF445820, *RPT6A* Intron: KF906109.

*B. simaruba* (L.) Sarg.; Goldman s.n. (TEX). Florida, U.S.A. ETS: GQ378038, ITS: GQ378104, *RPT6A* Intron: KF906110, *BXL2* Intron: KF906097, *mtATP Synthase D* Intron: KF906086.

*Commiphora ankaranensis* (J.-F. Leroy) Cheek & Rakot.; Weeks 10-I-11-02 (GMUF). Ankarana, Madagascar. 12°18′53″S, 49°20′18″E. ETS: KF035010, ITS: KF906081, *RPT6A* Intron: KF906118, *BXL2* Intron: KF906104, *mtATP Synthase D* Intron: KF906092, *Rab6* Intron: KF906128.

*Commiphora aprevalii* (Baill.) Guillaumin; Weeks 10-I-20-04 (GMUF). Toliara, Madagascar. 22°57′16″S, 44°20′39″E. ETS: KF034992, ITS: KF906075, *RPT6A* Intron: KF906112, *mtATP Synthase D* Intron: KF906088, *Rab6* Intron: KF906122.

*Commiphora falcata* Capuron; Weeks 10-I-26-03 (GMUF). Toliara, Madagascar. 23°01′29″S, 43°36′60″E. ETS: KF034994, ITS: KF906076, *RPT6A* Intron: KF906113, *BXL2* Intron: KF906099, *mtATP Synthase D* Intron: KF906089, *Rab6* Intron: KF906123.

*Commiphora grandifolia* Engl.; Weeks 10-I-13-01 (GMUF). Ankarana, Madagascar. 12°34′49″S, 49°27′31″E. ETS: KF034996, ITS: KF906077, *RPT6A* Intron: KF906114, *BXL2* Intron: KF906100, *mtATP Synthase D* Intron: KF906090, *Rab6* Intron: KF906124.

*Commiphora lamii* H. Perrier; Weeks 10-I-26-02 (GMUF). Toliara, Madagascar. 23°01′29″S, 43°36′60″E. ETS: KF034998, ITS: KF906078, *RPT6A* Intron: KF906115, *BXL2* Intron: KF906101, *mtATP Synthase D* Intron: KF906091, *Rab6* Intron: KF906125.

*Commiphora pervilleana* Engl.; Weeks 10-I-11-01 (GMUF). Ankarana, Madagascar. 12°18′53″S, 49°20′18″E. ETS: KF035005, ITS: KF906079, *RPT6A* Intron: KF906116, *BXL2* Intron: KF906102, P43: KF906126.

*Commiphora* sp. A. Weeks 10-I-09-01 (GMUF). Ankarana, Madagascar. 12°14′11″S, 49°21′18″E. ETS: KF035009, ITS: KF906080, *RPT6A* Intron: KF906117, *BXL2* Intron: KF906103, *Rab6* Intron: KF906127.

*Commiphora* sp. B. Weeks 10-I-15-04 (GMUF). Ankarana, Madagascar. 12°34′49″S, 49°27′31″E. ETS: KF0906087, ITS: KF906074, *RPT6A* Intron: KF906111, *BXL2* Intron: KF906098, *mtATP Synthase D* Intron: KF906087, *Rab6* Intron: KF906121.