Exon-Primed Intron-Crossing (EPIC) Markers for Evolutionary Studies of Ficus and Other Taxa in the Fig Family (Moraceae)

Authors: Yao, Xiaohong, Li, Chenhong, and Dick, Christopher W.

Source: Applications in Plant Sciences, 1(10)

Published By: Botanical Society of America

URL: https://doi.org/10.3732/apps.1300037

Applications in Plant Sciences 2013 1(10): 1300037

Primer Note

Exon-primed intron-crossing (EPIC) markers for evolutionary studies of *Ficus* and other taxa in the fig family (Moraceae)[1]

Xiaohong Yao[2,3], Chenhong Li[4], and Christopher W. Dick[3,5,6]

[2]Key Laboratory of Plant Germplasm Enhancement and Specialty Agriculture, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan 430074, Hubei, People's Republic of China; [3]Department of Ecology and Evolutionary Biology, University of Michigan, Ann Arbor, Michigan 48109-1048 USA; [4]Key Laboratory of Exploration and Utilization of Aquatic Genetic Resources, Shanghai Ocean University, Ministry of Education, Shanghai, People's Republic of China; and [5]Smithsonian Tropical Research Institute, P.O. Box 0843-03092, Balboa, Ancón, Republic of Panama

- *Premise of the study:* The genus *Ficus* (fig trees) comprises ca. 750 species of trees, vines, and stranglers found in tropical forests throughout the world. Fig trees are keystone species in many tropical forests, and their relationship with host-specific wasp pollinators has received much attention, although many questions remain unresolved regarding the levels of host specificity, cospeciation, and the role of hybridization in fig and wasp speciation. We developed exon-primed intron-crossing (EPIC) markers to obtain phylogenetic resolution needed to address these questions.
- *Methods and Results:* Expressed sequence tags (ESTs) from *F. elastica* were compared to *Arabidopsis* and *Populus* genomes to locate introns and to design primers in flanking exons. Primer pairs for 80 EPIC markers were tested in samples from divergent clades within *Ficus* and the outgroup *Poulsenia* (Moraceae).
- *Conclusions:* Thirty-one EPIC markers were successfully sequenced across *Ficus*, and 29 of the markers also amplified in *Poulsenia*, indicating broad transferability within Moraceae. All of the EPIC markers were polymorphic and showed levels of polymorphism similar to that of the widely used internal transcribed spacer (ITS).

Key words: exons; *Ficus*; Moraceae; nuclear DNA markers; phylogeny; transcriptome.

Ficus L. (Moraceae) is a pantropical genus comprised of ca. 750 species of trees, epiphytes, shrubs, vines, and stranglers found primarily in humid tropical forests. As a year-round source of calcium-rich fig fruits, *Ficus* trees are often described as keystone species. However, *Ficus* may be best known for their pollination mutualism with small (1–2 mm), short-lived (1–2 d) "fig wasps" in the family Agaonidae (Weiblen, 2002; Herre et al., 2008). Female fig wasps pollinate flowers and oviposit within the enclosed inflorescence (syconium or "fig"), in which the larvae develop before emerging to pollinate and oviposit in the syconia of asynchronously flowering conspecific trees. For sustained reproduction of the figs and the wasps, the wasps must exhibit a high degree of host-specificity, and the host population must provide access to flowers (i.e., figs) throughout the year.

Although the fig-wasp pollination mutualism is one of the tightest known in terms of host-pollinator specificity, there are many exceptions to the one pollinator species/one host species rule. In some cases, two or more wasp species pollinate the same host species in different parts of its geographic range, and multiple wasp species have been found in a single host tree (Herre et al., 2008). Furthermore, in Central America and in South Africa some wasp species have been shown to use more than one fig species in the local fig community (reviewed in Herre et al., 2008). The nonspecificity of some pollinators, in addition to some genetic studies (e.g., Machado et al., 2005), suggests that hybridization is possible.

Most phylogenetic studies of *Ficus* have used chloroplast DNA and/or one or two commonly used nuclear DNA markers (e.g., internal transcribed spacer [ITS]) (e.g., Rønsted et al., 2005). These markers are insufficient in number for studies of introgression, and they do not resolve phylogenies of closely related species or phylogeographic structure in widespread species (C. Dick, unpublished). To address the deficiency in nuclear genomic markers for *Ficus*, we have developed a set of exon-primed intron-crossing (EPIC) markers by comparing an expressed sequence tag (EST)–library for *F. elastica* Roxb. ex Hornem. with the annotated genomes of *Populus trichocarpa* Torr. & A. Gray (Salicaceae) and *Arabidopsis thaliana* (L.) Heynh. (Brassicaceae) using a bioinformatics pipeline developed by Li et al. (2010).

[1]Manuscript received 12 May 2013; revision accepted 17 June 2013.

Financial support was received from the National Science Foundation (DEB 0640379 to C.W.D.). The authors thank Allen Herre, Adalberto Gomez, and Katrin Heer for help in locating Panamanian *Ficus*, and the Smithsonian Tropical Research Institute and Autoridad Nacional del Ambiente (ANAM) for facilitating research and providing research permits for work in Panama.

[6]Author for correspondence: cwdick@umich.edu

doi:10.3732/apps.1300037

METHODS AND RESULTS

***Selection of taxa*—**Neotropical *Ficus* contains two distinct and phylogenetically distant subgenera, which represent two important neotropical life forms: the free-standing fig trees (subg. *Pharmacosycea* (Miq.) Miq. sect. *Pharmacosycea*)

Table 1. Characterization of 31 EPIC markers designed to amplify broadly across the genus *Ficus*.[a]

Locus[b]	Primer sequences (5′–3′)	Total/intron length (bp) (+range)	No. of polymorphic sites	Nucleotide diversity	GenBank accession no.	Reference locus[c]	Gene abbreviation[d]
FA08190	F: CCAAATTGTTGCGAGGATCT R: TTTAGAGCATCGGTCATGGA	484/435 (+2)	24	0.05053	JQ341915 JQ341916	AT1G08190	ATVAM2
FA02580	F: CTAGATCTTGCACAGCAGCAG R: GCATGGTGTAGTGCCACAAA	487/381 (+2)	22	0.04593	JQ341917 JQ341918	AT4G02580	T10P11_14
FA03310	F: GCGGGTATAAGAAGGGAACC R: GGTGCATTGACCACCTTGAT	740/581 (+2)	43	0.05866	JQ341919 JQ341920	AT3G03310	LCAT3
FA07360a	F: GCTGATAAAATGGTTGCTGCTG R: CCCCTTGATCTTCCCCATTACT	540/287	29	0.05410	JQ341921 JQ341922	AT2G07360	T13E11.13
FA08510	F: TGCTGGACTTCTTGGTGATG R: CAATCACGACGCATACCATTC	893/741	51	0.05724	JQ341923 JQ341924	AT1G08510	FATB
FA11980	F: AGTTGGGCCATGCATCAGA R: ACCCAACGATGTGAATCCAA	851/734 (+4)	35	0.04142	JQ341925 JQ341926	AT5G11980	F14F18_150
FA14000	F: TCCAGTGCTGATCATTTGAAAG R: GGCTGCCTCATAAGGCTCA	443/278 (+7)	23	0.05349	JQ341927 JQ341928	AT1G14000	F7A19_9
FA16180b	F: CGGACTTATGGAACCAGAGTAATTC R: GATGCTCCAGTACAATGACAACAT	417/281 (+3)	21	0.05147	JQ341929 JQ341930	AT4G16180	DL4130C
FA16690b	F: TCACAATTCTCCAGTGGTCATAAT R: TTCTCAGAAAAACTGCAACCTT	964/674 (+3)	40	0.04171	JQ341931 JQ341932	AT5G16690	ATORC3
FA19690*	F: ACTTGGCCTTCTTACTTCATGG R: AGCAATCCCAGACATGATGC	386/258 (+2)	12	0.03158	JQ341933 JQ341934	AT5G19690	STT3A
FA23640*	F: ATTCCTTTTGGTCCTCCACATC R: ACCCCCAATCCAGGGAACTA	1032/821 (+1)	55	0.05478	JQ341935 JQ341936	AT3G23640	HGL1
FA24620a	F: CCTTACAAGGACAGCCTTTTG R: CTCAAGCTCCCAATCATGG	513/323	20	0.04219	JQ341937 JQ341938	AT4G24620	PGI1
FA24620b	F: TGGCTAGATTTCCCATGTTTG R: AGCTGCTGCAGATACCGACT	980/827 (+4)	50	0.05149	JQ341939 JQ341940	AT4G24620	PGI1
FA26990	F: GGAAGCGTACAGGGTGATGT R: CATCAGAGCCATCTTCCTTTTG	476/246	13	0.02760	JQ341941 JQ341942	AT2G26990	FUS12
FA32180	F: TGCTCGAACTAAGGGAAGAATG R: GCTGCAAGAACACCTTCAATAA	741/628 (+13)	38	0.05163	JQ341943 JQ341944	AT4G32180	ATPANK2
FA32910	F: GGTTGGAATTCTTGGAGAAAATAC R: GTGAAGCCAAAACTTGAGCATA	455/284 (+2)	12	0.02655	JQ341945 JQ341946	AT4G32910	F26P21_30
FA36880b	F: GCTGTTGGGACATTGTTGAC R: ATAACCGCTACACTCCCCTTC	1044/896 (+6)	41	0.03958	JQ341947 JQ341948	AT5G36880	F5H8_15
FA45300	F: GGAGGACTTGGTCTTGGTTACTT R: CCATTAGTGCACCACATCTTGT	890/684	41	0.04622	JQ341949 JQ341950	AT3G45300	ATIVD
FA48520*	F: TCATCCATATTTGGTCGGAGAT R: CCACCCATTGTCTTTCACTTG	1059/890 (+4)	71	0.07305	JQ341951 JQ341952	AT5G48520	ATAUG3
FA73180	F: CGGGACTTATCTTCAGACTTTTCA R: GTGCCTTAGAAAGCTCAACTGC	470/235	18	0.03863	JQ341953 JQ341954	AT1G73180	T18K17_15
FP04090b	F: GGAATGCAAGCAATTGATGA R: AGGTCCAGCAACCTCAGCTA	438/275 (+10)	15	0.03529	JQ341955 JQ341956	POPTR_0006s00800	CYP97B3
FP08470	F: GCGATGTGCTGCGTGTATTT R: GGTCCATAAAGACTTGGAGAGG	550/404 (+7)	25	0.04630	JQ341957 JQ341958	POPTR_0017s08470	BGAL9
FP08550	F: CCGCTATCCTTTGGCTGTTA R: CACATGCTTCTGCACGTTCT	741/451 (+5)	36	0.05233	JQ341959 JQ341960	POPTR_0006s08550	F6E21_100
FP09670	F: GCAGCAACGTGGTGATAAGA R: ATCACATTAGCCTCGGGAATATC	642/509	32	0.05016	JQ341961 JQ341962	POPTR_0001s09670	XPB1
FP10430	F: GTGGGATGTCAGTTTGGATTT R: CAGCCCAGGAAAAGTATCCA	1021/658 (+161)	44	0.05176	JQ341963 JQ341964	POPTR_0009s10430	FUT11
FP10550	F: GGTGAAGGTGCAGTTGATCAGT R: GCTTGACAGCCTCTTCATCAGT	473/325 (+1)	24	0.05172	JQ341965 JQ341966	POPTR_0008s10550	ALDH22a1

Table 1. Continued.

Locus[b]	Primer sequences (5′–3′)	Total/intron length (bp) (+range)	No. of polymorphic sites	Nucleotide diversity	GenBank accession no.	Reference locus[c]	Gene abbreviation[d]
FP11540b	F: GATTACAACAACCTCTGCCAGT R: AGCATGTGCTTGTACTCATCAAC	661/496 (+4)	28	0.04328	JQ341967 JQ341968	POPTR_0017s11540	MZN14.21
FP12610a	F: GGATGCACTGGTTATGGTCA R: TCGTAAGGAGCACCAGCAAC	362/238	14	0.03889	JQ341969 JQ341970	POPTR_0011s12610	uncharacterized
FP13070	F: GGCACATTTGCTTCCATTCT R: TAATGCATGATTCCTGTTCCAA	844/748 (+2)	38	0.04612	JQ341971 JQ341972	POPTR_0013s13070	uncharacterized
FP17290	F: CTCACATGCCTCACTCATGC R: GTCTCCACACGGTCCTTTCT	781/642 (+2)	33	0.04465	JQ341973 JQ341974	POPTR_0001s17290	F18B13_28
FP35460	F: TCTCTGGTTGTTGCTGATTTTGG R: TGGGGTCTGCTCCTCCAGT	735/634 (+8)	41	0.05840	JQ341975 JQ341976	POPTR_0001s35460	unknown

[a] The locus descriptions (total and intron length, polymorphism) represent comparisons between *Ficus obtusifolia* (sect. *Americana*) and *F. maxima* (sect. *Pharmacocysea*).
[b] The first two letters of the marker name indicate the genomic comparisons (e.g., FA = *Ficus/Arabidopsis*; FP = *Ficus/Populus*) followed by the numerical locus identifier of the reference genome.
[c] Full reference genome locus name.
[d] Abbreviation for the putative gene function.
* Denotes markers that were not transferable to the *Poulsenia armata* (Moraceae) outgroup.

and the strangler figs (subg. *Urostigma* (Gasp.) Miq. sect. *Americana* Miq.). Sect. *Pharmacocysea* is sister to all the other fig subgenera, and therefore our sect. *Americana* and sect. *Pharmacocysea* samples share a most recent common ancestor that is the base of the entire *Ficus* crown clade, which, based on fossil records, dates back to at least 60 million years before present (Rønsted et al., 2005). All primers were tested on *F. obtusifolia* Kunth (sect. *Pharmacocysea*) and *F. maxima* Mill. (sect. *Americana*), which were collected from the Barro Colorado National Monument (BCNM) in central Panama. The subset of primers that amplified in both *Ficus* species were also tested on *Poulsenia armata* (Miq.) Standl., which is a monotypic genus in the fig family Moraceae (Datwyler and Weiblen, 2004). Botanical vouchers (*Dick and Gomez 234*, *F. obtusifolia*; *Dick and Gomez 240*, *F. maxima*; and *Dick and Gomez 180*, *P. armata*) were deposited at the herbaria of the University of Panama (PMA) and University of Michigan, Ann Arbor (MICH). Genomic DNA was extracted with the cetyltrimethylammonium bromide (CTAB) method of Doyle and Doyle (1987).

Bioinformatics pipeline—Researchers from the United States Department of Agriculture (USDA) previously developed an EST library of *F. elastica* to characterize the genetic basis of rubber biosynthesis (McMahan and Whalen, personal communication). We compared 9289 unique *F. elastica* ESTs from the National Center for Biotechnology Information (NCBI) database with the annotated genomes of *A. thaliana* (Brassicaeae) and *P. trichocarpa* (Salicaceae) using the informatics pipeline developed by Li et al. (2010). Briefly, we (1) retrieved coding sequences (CDS) that were longer than 100 bp from the annotated genomes of *A. thaliana* and *P. trichocarpa*. (2) We compared those CDS with the genome of the same species to identify "single-copy" CDS. (3) The candidate single-copy CDS thus identified were subsequently compared to the EST library of *F. elastica* to find markers that were conserved (identity >80%) among all three species. (4) After locating the single-copy conserved CDS, we screened for CDS flanking small introns, which were smaller than 1000 bp in the compared genomes, to facilitate the subsequent PCR and sequencing steps. Primers based on the *F. elastica* exons were initially designed by eye and subsequently checked with the Primer3 web interface program (Rozen and Skaletsky, 2000).

Primer assays—PCR was performed in a final volume of 20 μL containing 10 mM Tris–HCl (pH 8.4), 50 mM $(NH_4)_2SO_4$, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 0.1 μM each primer, 2 ng of genomic DNA, and 0.5 units of *Taq* polymerase (BioTherm, Gaithersburg, Maryland, USA). The amplification profiles included an initial denaturing at 94°C for 5 min; followed by 35 cycles of 50 s at 94°C, 50 s at 54°C, and 1 min at 72°C; and a final extension step of 10 min at 72°C. PCR products were ligated into the pMD 18-T plasmid vector (Promega Corporation, Madison, Wisconsin, USA) and transformed into *E. coli* strain (DH5α, Promega Corporation). Insert-positive plasmids were isolated using the E.Z.N.A. Plasmid Mini Kit I (Omega Bio-Tek, Norcross, Georgia, USA) and amplified using M13 primers. Forward and reverse strands of each amplicon were sequenced on an ABI 3730xL DNA sequencer (Applied Biosystems, Carlsbad, California, USA) at the University of Michigan Sequencing Core Facility. All *Ficus* insert sequences have been deposited in GenBank (accession numbers JQ341915–JQ341980; also see Table 1). For comparisons with ITS, we also obtained ITS sequences for *F. obtusifolia*, *F. maxima*, and *P. armata* (GenBank accessions JX137113–JX137114) using standard methods.

Data analyses—DNA chromatograms were edited using the SEQUENCHER program (Gene Codes Corporation, Ann Arbor, Michigan, USA). DNA sequences were initially aligned using ClustalX version 1.81 (Thompson et al., 1997) with default settings, and subsequently aligned manually using Se-Al (Rambaut, 1996). We determined number of polymorphic sites, nucleotide diversity (π), and GC content using MEGA 5 software (Kumar et al., 2008).

Results—We identified 200 ESTs that satisfied our criterion of 80% exon identity with the published genomes. Based on intron length, we selected a subset of 80 ESTs for further marker development, of which 31 amplified successfully in *Ficus* species from both subgenera, 16 amplified in one species only, and 33 did not amplify in either species. The 31 cross-amplifying primer pairs were further tested in *P. armata*, of which 29 amplified successfully (Table 1). The number of polymorphic sites in *F. obtusifolia* and *F. maxima* comparisons ranged from 12 to 71 (mean = 32), whereas nucleotide diversity ranged from 0.02655 to 0.07305 (mean = 0.0470) (Table 1). In comparison, there were 45 variable sites in ITS between *F. obtusifolia* and *F. maxima*, falling within the range of the EPIC marker variation.

CONCLUSIONS

The 31 EPIC markers that amplified between the two *Ficus* subgenera indicate that these markers might be useful across the full phylogenetic breadth of the >60 Ma genus and its >750 species. The markers that transfer to *Poulsenia* indicate an even broader phylogenetic utility within the Moraceae (ca. 40 genera and 1000 species), which probably originated in the Cretaceous. These markers should therefore be extremely useful for phylogenetic analysis at the family level and potentially beyond. The markers show a level of intron divergence that is of a similar magnitude as ITS, which is one of the most informative and broadly used markers in plant molecular systematics. These EPIC loci should be useful for analyzing recent divergences in which incomplete lineage sorting and/or introgression may be factors, including recent speciation, hybridization, and comparative phylogeography. In combination with EPIC markers developed for chalcid wasps (Lohse et al., 2011), it should now be possible to jointly analyze wasp and host plant phylogenies to study coevolution at both population and phylogenetic scales.

LITERATURE CITED

Datwyler, S. L., and G. D. Weiblen. 2004. On the origin of the fig: Phylogenetic relationships of Moraceae from *ndhF* sequences. *American Journal of Botany* 91: 767–777.

Doyle, J., and J. L. Doyle. 1987. Genomic plant DNA preparation from fresh tissue—CTAB method. *Phytochemical Bulletin* 19: 11–15.

Herre, E. A., K. C. Jander, and C. A. Machado. 2008. Evolutionary ecology of figs and their associates: Recent progress and outstanding puzzles. *Annual Review of Ecology Evolution and Systematics* 39: 439–458.

Kumar, S., M. Nei, J. Dudley, and K. Tamura. 2008. MEGA: A biologist-centric software for evolutionary analysis of DNA and protein sequences. *Briefings in Bioinformatics* 9: 299–306.

Li, C., J. M. Riethoven, and L. Ma. 2010. Exon-primed intron-crossing (EPIC) markers for non-teleost fishes. *BMC Evolutionary Biology* 10: 90.

Lohse, K., B. Sharanowski, M. Blaxter, J. A. Nicholls, and G. N. Stone. 2011. Developing EPIC markers for chalcidoid Hymenoptera from EST and genomic data. *Molecular Ecology Resources* 11: 521–529.

Machado, C. A., N. Robbins, M. T. P. Gilbert, and E. A. Herre. 2005. Critical review of host specificity and its coevolutionary implications in the fig/fig-wasp mutualism. *Proceedings of the National Academy of Sciences, USA* 102: 6558–6565.

Rambaut, A. 1996. Se-Al; Sequence Alignment Editor. Available at http://tree.bio.ed.ac.uk/software/ [accessed 11 September 2013].

Rønsted, N., G. D. Weiblen, J. M. Cook, N. Salamin, C. A. Machado, and O. Savolainen. 2005. 60 million years of co-divergence in the fig-wasp symbiosis. *Proceedings of the Royal Society of London. Series B. Biological Sciences* 272: 2593–2599.

Rozen, S., and H. J. Skaletsky. 2000. Primer3 on the WWW for general users and for biologist programmers. *In* S. Misener and S. A. Krawetz [eds.], Methods in molecular biology, vol. 132: Bioinformatics methods and protocols, 365–386. Humana Press, Totowa, New Jersey, USA.

Thompson, J. D. T. J., F. Gibson Plewniak, F. Jeanmougin, and D. G. Higgins. 1997. The CLUSTAL-X Windows interface: Flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Research* 25: 4876–4882.

Weiblen, G. D. 2002. How to be a fig wasp. *Annual Review of Entomology* 47: 299–330.