Molecular Phylogenetics and Evolution 54 (2010) 23–35

Contents lists available at ScienceDirect

Molecular Phylogenetics and Evolution

journal homepage: www.elsevier.com/locate/ympev

An empirical demonstration of using pentatricopeptide repeat (PPR) genes as plant phylogenetic tools: Phylogeny of Verbenaceae and the *Verbena* complex

Yao-Wu Yuan *, Chang Liu, Hannah E. Marx, Richard G. Olmstead

Department of Biology, University of Washington, Seattle, WA 98195, USA

ARTICLE INFO

Article history:
Received 20 January 2009
Revised 28 August 2009
Accepted 28 August 2009
Available online 3 September 2009

Keywords:
PPR genes
Nuclear gene marker
Phylogeny
Non-model organisms
Verbenaceae
Verbena complex

ABSTRACT

The pentatricopeptide repeat (PPR) gene family, with hundreds of members in land plant genomes, has been recognized as a tremendous resource for plant phylogenetic studies based on publicly available genomic data from model organisms. However, whether this appealing nuclear gene marker system can be readily applied to non-model organisms remains questionable, particularly given the potential uncertainties in designing specific primers to only amplify the locus of interest from the sea of PPR genes. Here we demonstrate empirically the use of PPR genes in the family Verbenaceae and the *Verbena* complex. We also lay out a general scheme to design locus-specific primers to amplify and sequence PPR genes in non-model organisms. Intergeneric relationships within the family Verbenaceae were fully resolved with strong support. Relationships among the closely related genera within the *Verbena* complex and among some species groups within each genus were also well resolved, but resolution among very closely related species was limited. Our results suggest that PPR genes can be readily employed in non-model organisms. They may be best used to resolve relationships in a spectrum from among distantly related genera to among not-so-closely related congeneric species, but may have limited use among very closely related species.

1. Introduction

The pentatricopeptide repeat (PPR) genes form one of the largest gene families in land plants. Recent genome-wide studies have found 103 PPR gene members in *Physcomitrella patens*, 450 in *Arabidopsis thaliana*, 477 in *Oryza sativa* (rice) (O'Toole et al., 2008), about 600 in *Vitis vinifera* (Jaillon et al., 2007) and *Populus trichocarpa* (Tuskan et al., 2006), but only a few or none in nonplant organisms (Lurin et al., 2004) except *Trypanosoma brucei*, a parasitic protozoan, with 28 members (Pusnik et al., 2007). A preliminary survey of *Selaginella moellendorffii* genome sequences also suggests that there could be over 600 PPR genes in this lycophyte genome (Sugita and Hattori, http://wiki.genomics.purdue.edu/index.php/PPR_gene_family). These observations indicate that two major expansions of this gene family have occurred during the evolution of land plants, one at the very early stage of land plant divergence and the other following the divergence of bryophytes and the lineage leading to vascular plants (O'Toole et al., 2008; Sugita and Hattori, pers. comm.). PPR proteins are characterized by 2–26 tandem repeats of a highly degenerate 35 amino acid motif (Small and Peeters, 2000; Lurin et al., 2004) and usually are targeted to mitochondria or plastids (Lurin et al., 2004; Pusnik et al., 2007). Studies of individual PPR proteins indicate they function as sequence-specific RNA-binding proteins in many post-transcriptional processes of organellar transcripts, including splicing, editing, RNA stability, and translation (reviewed in Delannoy et al., 2007; Schmitz-Linneweber and Small, 2008).

Comparative genomic analyses (Walker et al., 2007; O'Toole et al., 2008) have revealed two intriguing characteristics of the PPR genes—(1) despite the enormous number, an extraordinary large proportion of these genes form well supported one-to-one orthologue pairs between rice and *Arabidopsis*, which suggests that most PPR gene loci predate the divergence of monocots and eudicots (tricolpates); (2) approximately 80% of the PPR genes in both rice and *Arabidopsis* genomes are intronless. The first property is instinctively appealing to systematists who are seeking multiple nuclear gene loci for phylogenetic inferences. The second property is a great advantage when resolving phylogenetic relationships at intergeneric or higher levels, as elaborated in a previous study (Yuan et al., 2009) that recognized the great potential of PPR genes as phylogenetic markers.

In this earlier study (Yuan et al., 2009), we have screened 127 loci that all are intronless and have a single orthologue in both rice and *Arabidopsis* by using publicly available plant genomic databases and bioinformatics tools. Comparison of orthologous gene

* Corresponding author. Present address: 4504 Miller Plant Sciences Building, Department of Plant Biology, University of Georgia, Athens, GA 30602, USA. Fax: +1 706 542 1805.
E-mail address: yyuan@plantbio.uga.edu (Y.-W. Yuan).

1055-7903/$ - see front matter © 2009 Elsevier Inc. All rights reserved.
doi:10.1016/j.ympev.2009.08.029

sequences between two *Arabidopsis* species (*A. thaliana* vs. *A. lyrata*) and between three Poaceae genera (*Oryza*, *Zea*, *Sorghum*) at all 127 loci showed that these PPR genes have a rapid rate of evolution, which on average is substantially higher than the commonly used non-coding chloroplast *trn*L-F region and could be comparable to the nuclear ITS region. Based on these results, Yuan et al. (2009) proposed that PPR genes are promising plant phylogenetic markers and can be best used at intergeneric and interspecies levels, as evidenced by phylogenetic analyses of six Poaceae genera and nine diploid *Oryza* species using 13 of the 127 PPR loci, whose sequences were retrieved from public genomic databases. However, whether this compelling nuclear marker system can be readily applied to non-model organisms that lack genomic resources (i.e., whole genome sequences or expressed sequence tags [ESTs]) may remain questionable to some researchers. In particular, whether it is straightforward to design such efficient primers that only amplify the locus of interest during PCR in non-model organisms is perhaps a major concern to empirical systematists. The primary goals of this study, therefore, are to present an empirical demonstration of using PPR genes as phylogenetic markers in a poorly studied family, Verbenaceae, and to lay out a general scheme for primer design in non-model organisms.

Verbenaceae contains 25–34 genera and some 500–1200 species, depending on different authors (Sanders, 2001; Atkins, 2004). The family comprises small trees, shrubs, lianas, and herbs, predominantly distributed in arid, subtropical regions of the New World. In spite of its moderate size, Verbenaceae is an important floristic component in certain ecosystems of North and South America. The bulk of Verbenaceae is composed of two species-rich groups. The *Verbena/Glandularia/Junellia* complex (*Verbena* complex *sensu* Yuan and Olmstead, 2008a) contains ca. 170 species (O'Leary et al., 2007, in press; Peralta et al., 2008; Peralta, 2009), and the *Lantana/Lippia/Aloysia* complex comprises 200–400 species (Sanders, 2001; Atkins, 2004). Two other genera that probably have over 50 species are *Citharexylum* and *Starchytarpheta*, although species number estimates for them vary drastically among different authors (Moldenke, 1958; Sanders, 2001; Atkins, 2004). The remaining 20 or so genera have fewer than 20 species each.

To date, no molecular phylogenetic studies on the infrafamiliar relationships of Verbenaceae have been published. Two recent treatments of this family based on morphological characters classified it into 4 and 6 tribes, respectively (Sanders, 2001; Atkins, 2004), but relationships between and within tribes remain unknown. As part of a larger scale study of Verbenaceae ongoing in our lab, the current study is an important step towards a more comprehensive understanding of phylogenetic relationships within the family.

In order to examine the phylogenetic utility of PPR genes at multiple taxonomic levels simultaneously, our taxon sampling was designed to cover distantly related genera, closely related genera, distantly related species, as well as very closely related species. The rationale of using Verbenaceae as an empirical demonstration is that—if the PPR gene system can be readily employed in a group as Verbenaceae, which do not have any genomic resources whatsoever, it may also be easily transferred to other groups of non-model organisms.

2. Materials and methods

2.1. Taxon sampling

Forty-five taxa were sampled in this study (Table 1), including 20 of the 25–34 genera of Verbenaceae that represent all the tribes defined in the recent taxonomic treatments (Sanders, 2001; Atkins, 2004). Seventeen genera were each represented by 1 or 2 species, whereas the *Verbena* complex was represented by 6 species of *Junellia*, 6 species of *Glandularia*, and 14 species of *Verbena*. Our sampling was informed by a more comprehensive study of the family based on chloroplast DNA data (Marx and Olmstead, unpubl.) and previous studies of the *Verbena* complex (Yuan and Olmstead, 2008a,b).

2.2. Primer design

With the assumption that all of the 127 loci screened in the previous study (Yuan et al., 2009) are good candidates, we arbitrarily drew one locus at each of the five *Arabidopsis* chromosomes from the pool of 127 loci (5 loci in total, see Table 1 for locus name). We then used the following protocol to design primers for amplification and sequencing of each locus:

(1) Retrieve the *Arabidopsis thaliana* coding sequence (CDS) from the TAIR database (http://www.arabidopsis.org/) by searching the locus name (e.g., AT1G09680). To design primers for monocots, the corresponding rice CDS should be retrieved for further blast searches. This can be done easily by searching the POGs/PlantRBP database (Walker et al., 2007; http://plantrbp.uoregon.edu/index.php) using the *Arabidopsis thaliana* locus name, and then following the POGs/PlantRBP link to the TIGR (http://www.tigr.org/) rice database and downloading the rice sequence.

(2) Blast the *Arabidopsis thaliana* CDS against *Mimulus guttatus* (Phrymaceae, Lamiales, *sensu* Beardsley and Olmstead, 2002) whole genome sequence (WGS) trace archives using the "discontiguous megablast" program and default parameters (http://blast.ncbi.nlm.mih.gov/). Sequences with *E-value* < e^{-10} were downloaded as SCF files, and were annotated using Sequencher 4.7 (Gene Codes Corporation, Ann Arbor, MI, USA). Usually this first-round blast can retrieve the largest part of the *Mimulus* sequence from the locus of interest, with part of the 5′ end sequence missing. That is because the PPR gene sequences are most variable towards the 5′ end and the *Arabidopsis* and *Mimulus* orthologous sequences do not share significant similarities in this portion at the nucleotide level. In order to recover the complete CDS for this locus, we did a second-round blast against the *Mimulus* WGS trace archives using the 200 nucleotides at the 5′ end of the annotated partial sequence with the "megablast" program. Sequences whose 3′ ends overlapped with these 200 nucleotides were then downloaded and assembled with the previously annotated partial sequence, yielding a complete *Mimulus* CDS. Complete CDS can be recovered successfully after the second-round blast in most cases. If this is not the case, one can simply reiterate the second-round blast process until the full CDS is recovered. The positions of start and stop codons were determined by comparison with the *A. thaliana* sequence, using the amino acid sequence alignment between these two distantly related taxa as a guide.

The purpose of this step is to retrieve the complete CDS from a "model organism" that has complete genome sequences (including trace archives) available and is most closely related to the group under investigation, which is Verbenaceae in the current study. The NCBI genome projects database (http://www.ncbi.nlm.nih.gov/Genomes/) lists 69 land plants (up to August, 2009) with their genome sequencing projects at various stages, which can be used as a preliminary guide to select "model organisms" that are most closely related to the study groups.

(3) Blast the annotated *Mimulus guttatus* CDS against the TIGR Plant Transcript Assemblies database (http://blast.jcvi.org/euk-blast/plantta_blast.cgi; Childs et al., 2007), which contains a vast number of plant ESTs. The ESTs are organized by species, families, and orders in this database. We blasted the *Mimulus* CDS against all ESTs of Lamiales, Solanales, and Gentinales, which comprise the bulk of euasterids I clade (APG, 2003) where *Mimulus* and Verben-

Table 1
Taxa included in this study and their sampling locality, voucher information, and percentage of intra-individual polymorphic sites at the five PPR loci. Sequence length and percentage of phylogenetically informative sites of each locus are also shown. ''NA'' indicates the value is not available due to the lack of sequence data. The dashes indicate the data are not meaningful for the corresponding entries because of a recent gene duplication at AT2G37230.

Taxon	Locality and voucher information	AT1G09680	AT2G37230	AT3G09060	AT4G01570	AT5G39980
Aloysia virgata (Ruiz & Pav.) Pers.	Misiones, Argentina, RGO 2004-133 (WTU)	0	0.41	0.27	0	0
Bouchea fluminensis (Vell.) Moldenke	Cultivation, H. Rimpler 1141 (FB)	0	4.06	1.86	0	0
Casselia glaziovii (Briq. & Moldenke) Moldenke	Minas Gerais, Brazil, M.A Silva et al. 3630 (US)	NA	0.07	NA	0.32	0.08
Citharexylum ligustrinum van Houtte	Cultivation, Kew 000-69.51235	0	0.83	7.27	0.05	4.70
Citharexylum montevidense (Spreng.) Moldenke	Buenos Aires, Argentina, RGO 2004-102 (WTU)	7.10	1.38	0.09	0.11	0.96
Dipyrena glaberrima Gill & Hook.	Mendoza, Argentina, RGO 2004-179 (WTU)	0.32	0.83	0.29	0.21	0
Duranta serratifolia (Griseb.) Kuntze	Salta, Argentina, RGO 2007-009 (WTU)	0	0.21	0	0.27	0
Glandularia araucana (Phil.) Botta	Mendoza, Argentina, RGO 2004-177 (WTU)	0.08	—	0.29	0	0
Glandularia dissecta (Willd. ex Spreng.) Schnack & Covas	Misiones, Argentina, RGO 2004-122 (WTU)	1.36	—	0.62	0.23	0.08
Glandularia flava (Gillies & Hook.) Schnack & Covas	Mendoza, Argentina, RGO 2004-153 (WTU)	1.60	—	0	0.21	0.08
Glandularia incisa (Hook.) Tronc.	Corrientes, Argentina, RGO 2004-108 (WTU)	0	—	0	0	0
Glandularia microphylla (Kunth) Cabrera	Mendoza, Argentina, RGO 2004-156 (WTU)	0.08	—	2.04	0.16	0
Glandularia subincana Tronc.	Corrientes, Argentina, RGO 2004-117 (WTU)	0.72	—	0	0.23	1.02
Junellia aspera (Gillies & Hook.) Moldenke	Mendoza, Argentina, RGO 2004-163 (WTU)	1.44	0.34	1.77	0	0.08
Junellia asparagoides (Gillies & Hook.) Moldenke	Mendoza, Argentina, RGO 2004-192 (WTU)	0.72	0.69	0.18	0.64	0.08
Junellia scoparia (Gillies & Hook.) Botta	Mendoza, Argentina, RGO 2004-178 (WTU)	0.16	0.41	0.39	0.31	0.08
Junellia seriphioides (Gillies & Hook.) Moldenke	Mendoza, Argentina, RGO 2004-147 (WTU)	0	0.07	0.44	0	0
Junellia spathulata (Gillies & Hook.) Moldenke	Mendoza, Argentina, RGO 2004-190 (WTU)	0.08	0.07	0	0	0
Junellia uniflora (Phil.) Moldenke	Mendoza, Argentina, RGO 2004-155(WTU)	0	0	0.18	0	0
Lampaya castellani Moldenke	Jujuy, Argentina, RGO 2007-063 (WTU)	0	0.14	0	0.53	0.55
Lantana canescens Kunth	Salta, Argentina, RGO 2007-006 (WTU)	1.36	0	0.27	0	0.47
Lippia dulcis Trevir.	Cultivation, RGO 1998-56 (WTU)	0	0	0	0	0
Neosparton ephedroides Griseb.	Catamarca, Argentina, RGO 2007-077 (WTU)	0	0	1.00	0	0.08
Petrea racemosa Nees	RBGK #000.73.17818	0.16	0.14	0.62	0	0.08
Pitraea cuneato-ovata (Cav.) Caro	Mendoza, Argentina, RGO 2004-186 (WTU)	0.40	0.07	0.18	0	0.47
Priva cordifolia Druce	KwaZulu-Natal, South Africa, W. Vos 391 (NU)	0	0	0	0	0
Rhaphithamnus venustus B.L. Rob.	T.F. Stuessy 11855	0	0	0	0	0
Stachytarpheta cayennensis (Rich.) Vahl	Corrientes, Argentina, RGO 2004-113 (WTU)	0	0	0.18	0	0
Stachytarpheta dichotoma (Ruiz & Pav.) Vahl	Hawaii, US, RGO 951 (WTU)	0	0	0.27	0	0
Tamonea boxiana (Moldenke) R.A. Howard	Puerto Rico, US, RGO 2003-12 (WTU)	0	0.07	0.09	0	0
Verbena bonariensis L.	Cultivation, RGO 2003-25 (WTU)	0	—	0.10	0	0.23
Verbena bracteata Lag. & Rodr.	Arizona, US, YYW 2005-11 (WTU)	0	—	0	0	0
Verbena canescens Kunth	Texas, US, YYW 2005-14 (WTU)	0	—	0	0	0
Verbena halei Small	Texas, US, YYW 2005-16 (WTU)	0	—	0	0	0
Verbena hastata L.	Cultivation, RGO 2003-155 (WTU)	0	—	0	0	0
Verbena hispida Ruiz & Pav.	Jujuy, Argentina, V. Soza 1826 (WTU)	0	—	0	0	0
Verbena intermedia Gill. & Hook.	Entre Rios, Argentina, RGO 2004-106 (WTU)	0	—	0.53	0	0.16
Verbena litoralis Kunth	Entre Rios, Argentina, RGO 2004-105 (WTU)	0.16	—	0.36	0.23	0.16
Verbena macdougalii A. Heller	New Mexico, US, YYW 2005-19 (WTU)	0	—	0	0	0
Verbena montevidensis Spreng.	Corrientes, Argentina, RGO 2004-112 (WTU)	0.08	—	0.27	0.23	0.23
Verbena officinalis L.	Cultivation, RGO 2003-156 (WTU)	0.08	—	0	0	0
Verbena orcuttiana L.M. Perry	Baja California, Mexico, YYW 2005-26 (WTU)	0	—	0	0	0
Verbena perennis Wooton	Texas, US, YYW 2005-17 (WTU)	0.16	—	0.09	0.15	0.08
Verbena rigida Spreng.	Corrientes, Argentina, RGO 2004-111 (WTU)	0.08	—	1.25	0.23	0.31
Xeroaloysia ovatifolia (Moldenke) Tronc.	San Luis, Argentina, RGO 2004-184 (WTU)	0.24	0.41	0.09	0.32	0.08
	Average percentage of polymorphic sites (%)	0.37	0.41	0.48	0.10	0.22
	Percentage of phylogenetically informative sites (%)	33.89	28.03	32.00	28.40	30.28
	Sequence length (bp)	1254	1452	1128	1294	1278

aceae belong, as well as ESTs of Asterales, a slightly more distantly related clade. Species hit most frequently by the blast search include *Triphysaria versicolor* (Orobanchaceae), *Antirrhinum majus* (Plantaginaceae), *Solanum lycopersicum* (Solanaceae), *Solanum tuberosum*, and various taxa in the family Asteraceae.

(4) Align the *Mimulus* CDS with all putatively orthologous ESTs retrieved from step (3) and scan the alignment for conserved motifs. Numerous short conserved motifs (15–30 bp) can be readily identified for all 5 loci that we arbitrarily selected, although a few nucleotide sites are often required to be degenerate for most primers (Table 2) designed from these motifs.

2.3. Molecular data collection

Total DNA was extracted from silica-gel dried tissue using the modified CTAB method (Doyle and Doyle, 1987). The following

Table 2
Primers optimized for amplifying and sequencing each locus in Lamiales. "F" and "R" indicate forward and reverse direction, respectively. Numbers indicate the approximate priming positions in the *Arabidopsis thaliana* coding sequence (CDS). The asterisks indicate primers that were not used for the majority of taxa but facilitated amplification in a few difficult taxa.

Locus	Primers (5′–3′)	Usage
AT1G09680	180F: ACCRCCCTWTCTCAAGCCATCCAAA*	PCR & Sequencing
	320F: TCTTCTCTTTCTTCACATGGCT	PCR & Sequencing
	850F: GTTAGTTTCAATACTTTGATGAA	Sequencing
	850R: TTCATCAAAGTATTGAAACTAAC	Sequencing
	1300F: TTGTAAGGARGGAGATYTGGA	Sequencing
	1300R: TCCARATCTCCYTCCTTACAA	Sequencing
	1590R: TAACCGTTCATAAGCACATTGTA	PCR & Sequencing
	1760R: TARTCAAGAACAAGCCCTTTCGCAC*	PCR & Sequencing
AT2G37230	320F: GCCTGGACDACMCGTTTRCAGAA	PCR & Sequencing
	520F: AAGCTBAATCAYGCYAGGTGYAT	Sequencing
	720R: TAYCKCTTGGCCATCATRTACC	Sequencing
	1000F: ACSACSTTGATTAAAGGGTATGT	Sequencing
	1260R: AGCACRTCTGCAGCWGCATC	Sequencing
	1390F: TTGAGAARGATATCATNBTGAG	Sequencing
	1770R: TCRAACAAGCTCTCCATCAC*	PCR & Sequencing
	1800R: GCYGTCTGAACWCSYCCATCYTC	PCR & Sequencing
AT3G09060	930F: AGTGCTYTGATTCATGGGTTGTG	PCR & Sequencing
	1000F: TCACCTGATGCTGTTGTATRT*	PCR & Sequencing
	1310F: GAGGTGTTCTAGATGCTTTTGC	Sequencing
	1310R: GCAAAAGCATCTAGAACACCTC	Sequencing
	1760F: ATGCATAATATTTTGATTCATGG	Sequencing
	1760 R: CCATGAATCAAAATATTATGCAT	Sequencing
	2000R: GATATTCTATTGCAAGAACA*	PCR & Sequencing
	2080R:ACAGCTCKRACAAGTATRTTCCA	PCR & Sequencing
AT4G01570	530F: CAAGTWTGATTCKGCTCTGGAAGT*	PCR & Sequencing
	1070F: TTCTCATWCAGGGKTGTTSTAAAT	PCR & Sequencing
	1070R: ATTTASAACAMCCCTGWATGAGAA*	Sequencing
	1460F: TTGGTRGARGAAATGGAGG	Sequencing
	1460R: CCTCCATTTCYTCYACCAA	Sequencing
	1980F: TGTAARTTATTTGAGATHTTCA	Sequencing
	1990R: ACGCCCATGTCKGTGAAAATCTC	Sequencing
	2420R: GGCGCRCAYCCAKCATCCAACAT	PCR & Sequencing
AT5G39980	550F: CACGGRCTGTTCGACGAAATGCG	PCR & Sequencing
	650F: TCTTGGCTTCAGAAGATGGA	Sequencing
	1030F: TGTAATATAATGATAGATGTKTA	Sequencing
	1030R: GCCCATAMACATCKATCATTAT	Sequencing
	1370F: GGGAAGYTRGATAGAGCAGC	Sequencing
	1620R: AAGACCGTTATRTCCTTGACCTC	Sequencing
	1890R: AGACTCAGCATCTGRAAATGAAC	PCR & Sequencing

PCR procedure was used to amplify the PPR loci: one cycle of 2 min at 94 °C, 4 cycles of 15 s at 94 °C, 10 s at 55 °C, and 2 min at 72 °C, followed by 30 cycles of 15 s at 94 °C, 15 s at 52 °C, and 1.5 min at 72 °C, with a final extension step for 10 min at 72 °C. This procedure worked well for all 5 PPR loci except AT4G01570, for which the annealing temperatures were modified to 57 and 55 °C at the 4-cycle stage and 30-cycle stage, respectively, to generate a single amplicon. Amplified PCR products were purified by precipitation from a 20% polyethylene glycol (PEG) solution and washed in 70% ethanol prior to sequencing. Both strands of the cleaned PCR products were sequenced using the DYEnamic ET Terminator Cycle Sequencing Kit (Amersham Biosciences, Piscataway, New Jersey, USA). Sequences of the five exemplar loci were obtained for all 45 Verbenaceae taxa except *Casselia glaziovii*, for which only 3 loci (AT2G37230, AT4G01570, AT5G39980) were successfully amplified and sequenced. Sequences generated in this study were deposited in GenBank (Accession Nos. FJ549069–FJ549291).

2.4. Phylogenetic analyses

All sequence alignments were conducted manually using the program Se-Al version 2.0a11 (Rambaut, 1996). Besides the 45 Verbenaceae taxa, *Mimulus guttatus*, with its sequences retrieved during primer design, was also included as outgroup in all phylogenetic analyses. We executed both parsimony and Bayesian analyses on the five individual data sets from each locus to generate multiple, presumably independent gene trees, so that we can visually examine congruence and discrepancies between them. The same analyses were also conducted on a concatenated supermatrix that includes all 5 loci. Since all sequences were generated by directly sequencing without cloning, many of them include polymorphic sites due to intra-individual allelic variation. The intra-individual polymorphism was treated as uncertainty in all the analyses.

Parsimony analyses were carried out using PAUP* version 4.0b10 (Swofford, 2002). Heuristic searches were performed with 500 random stepwise addition replicates and tree-bisection-reconnection (TBR) branch swapping with MULTREES on. Clade support was estimated by bootstrap analyses (Felsenstein, 1985) of 200 pseudoreplicates. Markov chain Monte Carlo (MCMC) Bayesian analyses were conducted using default parameters in MrBayes version 3.1.2 (Ronquist and Huelsenbeck, 2003). Modeltest version 3.7 (Posada and Crandall, 1998) was used to determine optimal models according to the Akaike information criterion (AIC; Akaike, 1974). We performed two independent runs of 1,000,000 generations, each with four Markov chains (one cold and three heated chains), sampling 1 tree every 100 generations (10,000 trees were saved). The program Tracer v.1.4 (Rambaut and Drummond, 2007) was used to examine the output parameters from Bayesian analyses to determine stationarity. In all cases, likelihoods of two separate runs reach stationarity before 200,000 generations. Convergence was assessed by the standard deviation of split frequencies, which were <0.01 towards the end of runs in all analyses. Trees prior to reaching stationarity were discarded as burn-in, and the remaining trees were used to compute consensus trees and posterior probabilities.

2.5. Comparison of interspecies sequence divergence and richness of phylogenetic information

To compare the extent of sequence divergence and richness of phylogenetic information between the five PPR loci and previously used non-coding chloroplast DNA (cpDNA) and nuclear *waxy*, *PHOT1*, and *PHOT2* gene loci (Yuan and Olmstead, 2008a,b), we constructed a subset of data including three species pairs (*Verbena canescens* and *V. macdougalii*; *Glandularia microphylla* and *G. incisa*; *Junellia uniflora* and *J. spathulata*). Selection of the species pairs was based on two criteria: (1) availability of sequence data for cpDNA and all eight nuclear loci; (2) minimal intra-individual allelic variation. The six selected species possess little or no intra-individual polymorphism at any of the eight nuclear loci. This is to minimize the effect of "uncertainty" states, as polymorphisms being treated, on estimation of sequence divergence and number of phylogenetically informative sites.

Sequences from the six species were aligned, resulting in nine data sets (a combined cpDNA data set and eight individual nuclear gene data sets). The uncorrected *p*-distance between the two species in each of the three pairs was then calculated using the "Pairwise Base Differences" function implemented in PAUP* version 4.0b10 (Swofford, 2002). Number of phylogenetically informative sites was determined by parsimony analysis of each data set.

3. Results

3.1. Sequence characteristics—length variation and allelic divergence

The average lengths of the amplified region are 1254 bp, 1452 bp, 1128 bp, 1294 bp, and 1278 bp for the locus AT1G09680, AT2G37230, AT3G09060, AT4G01570, and AT5G39980, respectively

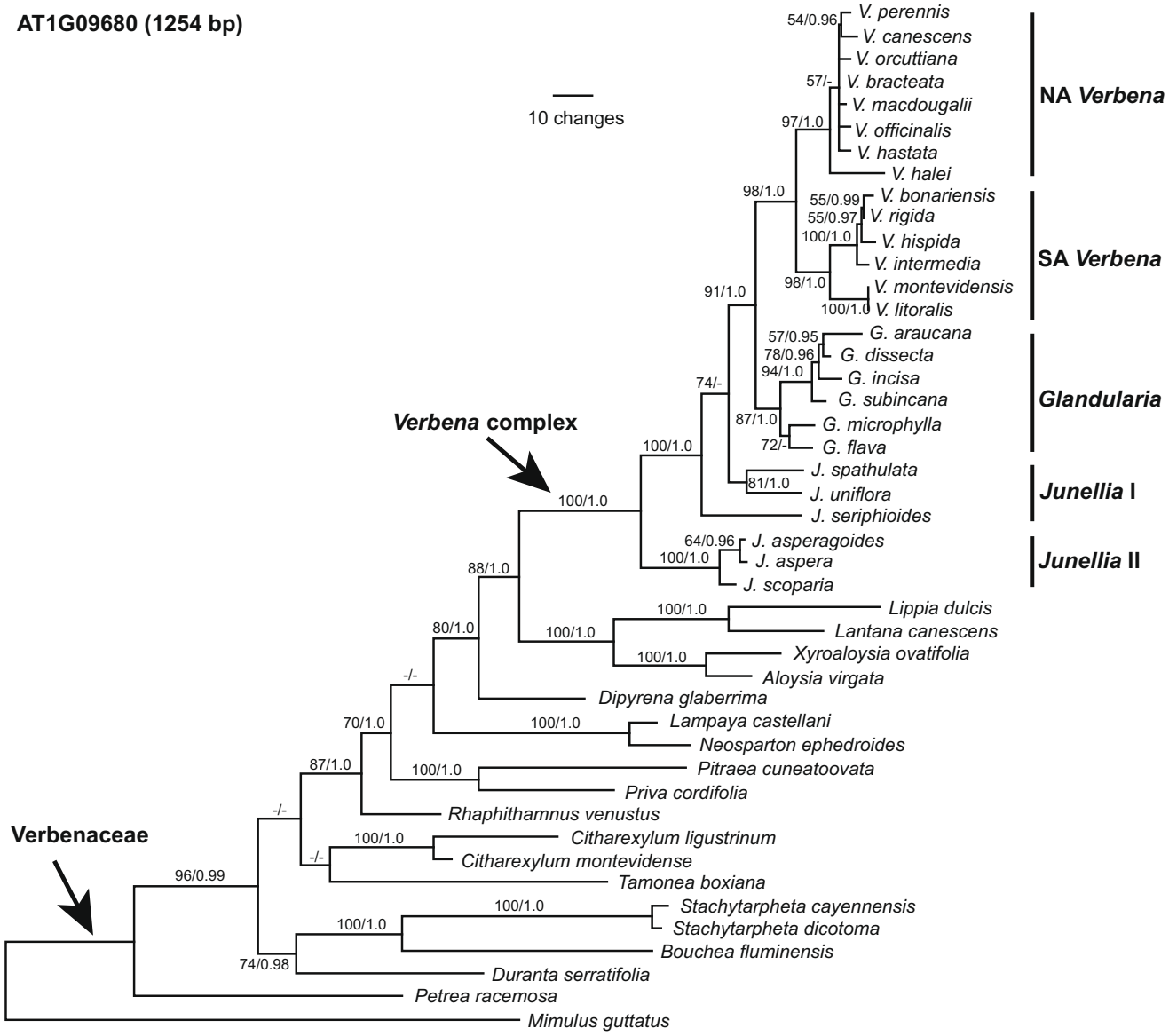

Fig. 1. Phylogram of 1 of 3 maximum parsimony (MP) trees from AT1G09680 (length = 1416; CI = 0.62; RI = 0.77). The topology is equal to the Bayesian consensus tree. Numbers along branches represent clade support by Bootstrap values (BS) and Bayesian posterior probabilities (PP). The hyphens indicate that BS < 50 or PP < 0.95 (BS/PP). Subdivisions within the *Verbena* complex were designated on the right. NA, North American; SA, South American.

(Table 1). In general, sequence length is very conserved within the family Verbenaceae at these loci, three of which (i.e., AT1G09680, AT2G37230, AT5G39980) have no observed variation at all. One 3-bp deletion occurred in *Neosparton ephedroides* at AT3G09060 (see Fig. 3) and seven short insertions/deletions were found at AT4G01570 (see Fig. 4). Two of these seven insertions/deletions are unique to *Dipyrena glaberrima* and *Casselia glaziovii*, respectively. A 18-bp deletion and a 3-bp insertion are potential synapomorphies of the genus *Citharexylum*. A 3-bp insertion defines the clade consisting *Casselia*, *Tamonea*, and *Citharexylum*. A 3-bp insertion and a 9-bp deletion define the monophyly of the *Verbena* complex (Fig. 4).

Although without cloning, the quality of generated PPR sequences is comparable to that of directly sequenced chloroplast DNAs (see some exemplar sequence chromatograms in Fig. 7). Polymorphism due to intra-individual allelic variation can be easily recognized as double peaks in the sequence reads of both the forward and reverse strands. Intra-individual polymorphic sites were identified and counted across all 5 loci and across all taxa (Table 1). Average percentage of polymorphic sites at each locus ranges from 0.10% (AT4G01570) to 0.48% (AT3G09060). The highest percentage of polymorphic sites, 7.27%, was found in *Citharexylum ligustrinum* at AT3G09060, whereas virtually no intra-individual allelic variation was found in *Lippia dulcis*, *Priva cordifolia*, *Rhaphithamnus venustus*, and most North American *Verbena* species (see Figs. 1–6 for geographic distribution of *Verbena* species) at all loci (Table 1). Compared with the average percentage of intra-individual sequence polymorphism at each locus, the percentage of phylogenetically informative sites is much higher. It ranges from 28.03% at locus AT2G37230 to 33.89% at locus AT1G09680 (Table 1). The concatenated supermatrix contains 1952 (30.40%) phylogenetically informative sites among the total 6421 sites.

3.2. Phylogenetic analyses

The GTR+I+G model was determined as the optimal model for each of the five loci as well as the concatenated data. Phylogenetic

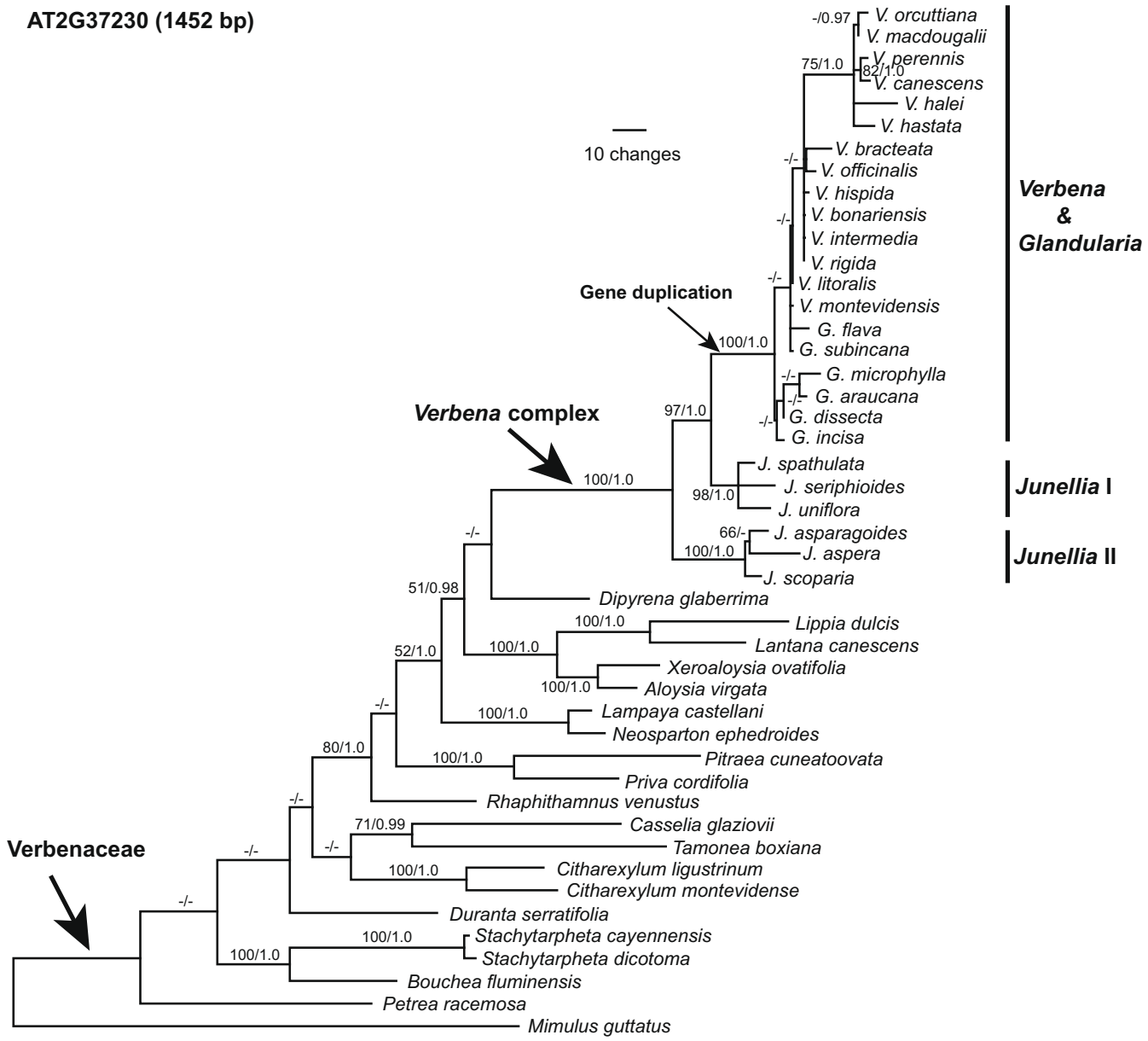

Fig. 2. One of 208 MP trees from AT2G37230 (length = 1474; CI = 0.58; RI = 0.76), shown in the same fashion as Fig. 1. Note that a putative gene duplication occurred in the common ancestor of *Verbena* and *Glandularia*.

trees inferred from the 5 individual data sets are shown in Figs. 1–5. Fig. 6 represents the results from analyses on the concatenated supermatrix. Parsimony and Bayesian analyses gave very similar results for all data sets except AT4G01570 (see Fig. 4). The intergeneric relationships of Verbenaceae resolved by the individual data sets are remarkably similar among all loci, albeit some of the nodes on the phylogenetic backbone are only weakly supported by individual data sets. When combining all data from 5 loci, relationships among the 20 sampled genera were fully resolved, with all nodes leading to different genera supported by high bootstrap (BS, 90–100%) values and Bayesian posterior probabilities (PP, all 1.0) (Fig. 6). This phylogeny inferred from the concatenated data is regarded as our reference phylogeny.

Not considering the *Verbena* complex, the AT1G09680, AT3G09060, and AT5G39980 gene trees have the same topology as the reference phylogeny. The AT2G37230 gene tree differs from the reference phylogeny only in the positions of two taxa, *Duranta* and *Dipyrena* (see Figs. 2 and 6), but neither of these positions is strongly supported (Fig. 2). The AT4G01570 gene trees generated by parsimony analysis (Fig. 4b) and Bayesian inference (Fig. 4a) are slightly different. Except for the position of *Dipyrena*, the Bayesian majority consensus tree is the same as the reference tree in topology. But in the single MP tree, the usually recovered *Casselia/Tamonea/Citharexylum* clade formed a grade near the base of the tree (Fig. 4b).

Within the *Verbena* complex, *Verbena* and *Glandularia* are strongly supported as monophyletic by all PPR loci except AT2G37230, at which phylogenetic signal has been smudged by a putative recent gene duplication that occurred in the common ancestor of *Verbena* and *Glandularia* (see Section 3.3). *Junellia* was resolved as non-monophyletic in all cases and there are two potential clades (*Junellia* I and *Junellia* II) with their relationships varying among different loci (Figs. 1–5). The reference tree suggests that *Verbena* and *Glandularia* are sister groups, *Junellia* I is sister to the

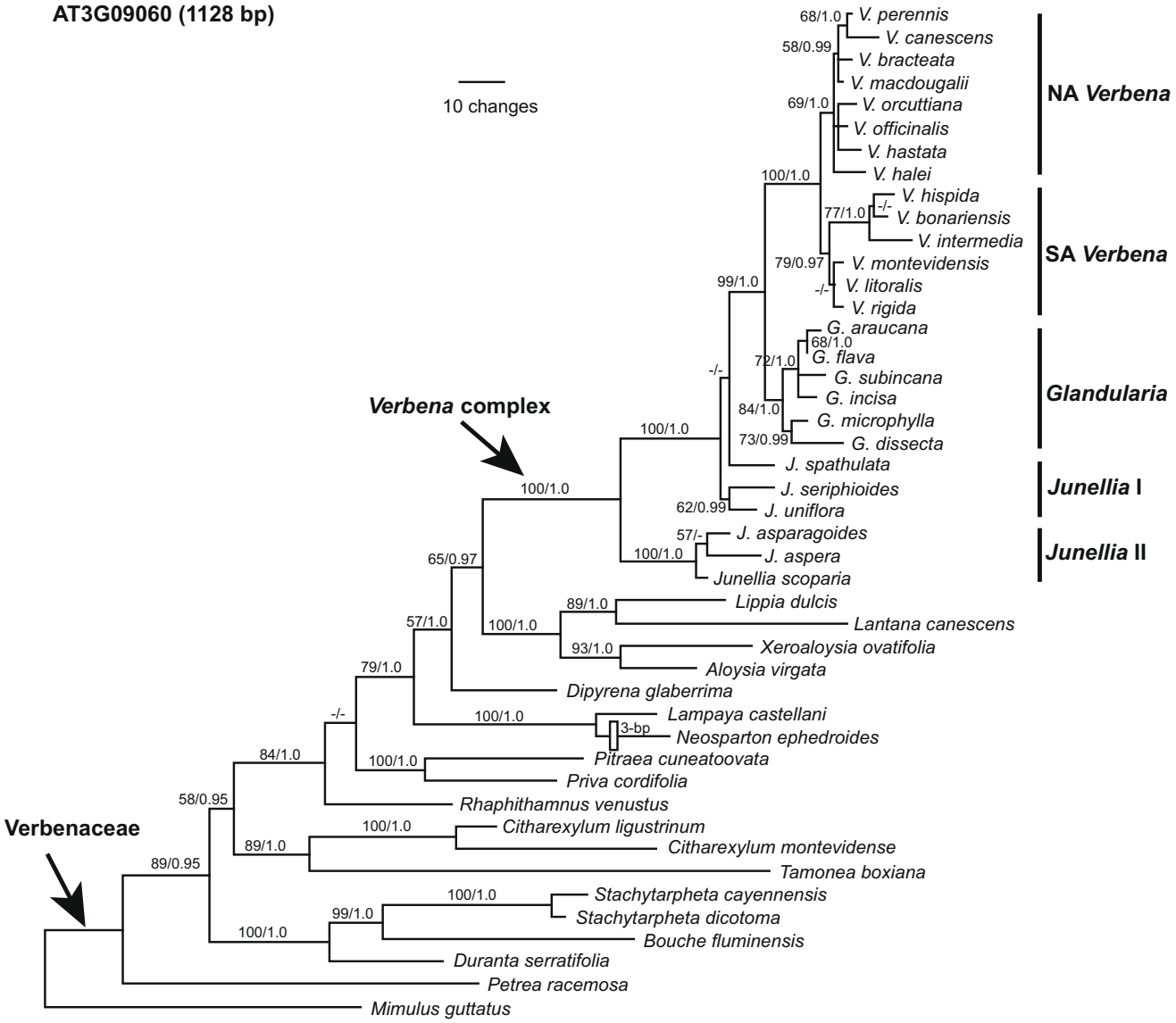

Fig. 3. One of 24 MP trees from AT3G09060 (length = 1216; CI = 0.65; RI = 0.80), shown in the same fashion as Fig. 1. The white bar indicates a 3-bp deletion in *Neosparton ephedroides* at this locus.

Glandularia + *Verbena* clade, and *Junellia* II is sister to this more inclusive clade. Within the genus *Verbena*, the North American clade and South American clade are differentiated by the PPR genes consistently, but within each clade there is little resolution. Within the genus *Glandularia*, interspecific relationships were resolved to some extent by each individual locus, but the relationships are variable between loci. A comparison of relationships among the two *Junellia* clades, *Glandularia*, and the two *Verbena* clades (South vs. North American) across the 5 PPR loci as well as chloroplast DNA and another three nuclear gene loci used in previous studies (Yuan and Olmstead, 2008a,b), is presented in Table 3.

3.3. Recent gene duplication of AT2G37230

Initial evidence for a recent gene duplication of AT2G37230 came from phylogenetic analyses. Whereas the genera *Glandularia* and *Verbena* can be differentiated readily by the other 4 PPR loci, they are intermingled in the AT2G37230 gene tree (Fig. 2), which provides little resolution within the *Glandularia* + *Verbena* clade. Direct sequencing of this locus yielded numerous polymorphic sites shared by all *Verbena* and *Glandularia* species, but not by other taxa. Fig. 7 represents two exemplar sites located at position 186 and 1363 of the AT2G37230 alignment. These polymorphisms are displayed by both forward- and reverse-strand sequences, ruling out the possibility of sequencing error. The most likely explanation to the observed pattern of polymorphisms is that a recent gene duplication of AT2G37230 had occurred in the common ancestor of *Verbena* and *Glandularia*, and therefore, each AT2G37230 sequence of *Verbena* or *Glandularia* generated in this study actually represents a pool of two paralogous sequences. This also explains the "loss" of phylogenetic signal within the *Verbena* + *Glandularia* clade. Separating the multiple sequences from *Verbena* and *Glandularia* individuals via cloning will be a direct means to test this hypothesis. Phylogenetic analyses of the cloned sequences should result in two distinct clades, each represented by a paralogue of AT2G37230.

After identifying this putative gene duplication, we excluded the *Verbena* and *Glandularia* AT2G37230 sequences from the concatenated data and performed phylogenetic analyses on the new concatenated supermatix. The results are essentially the same as

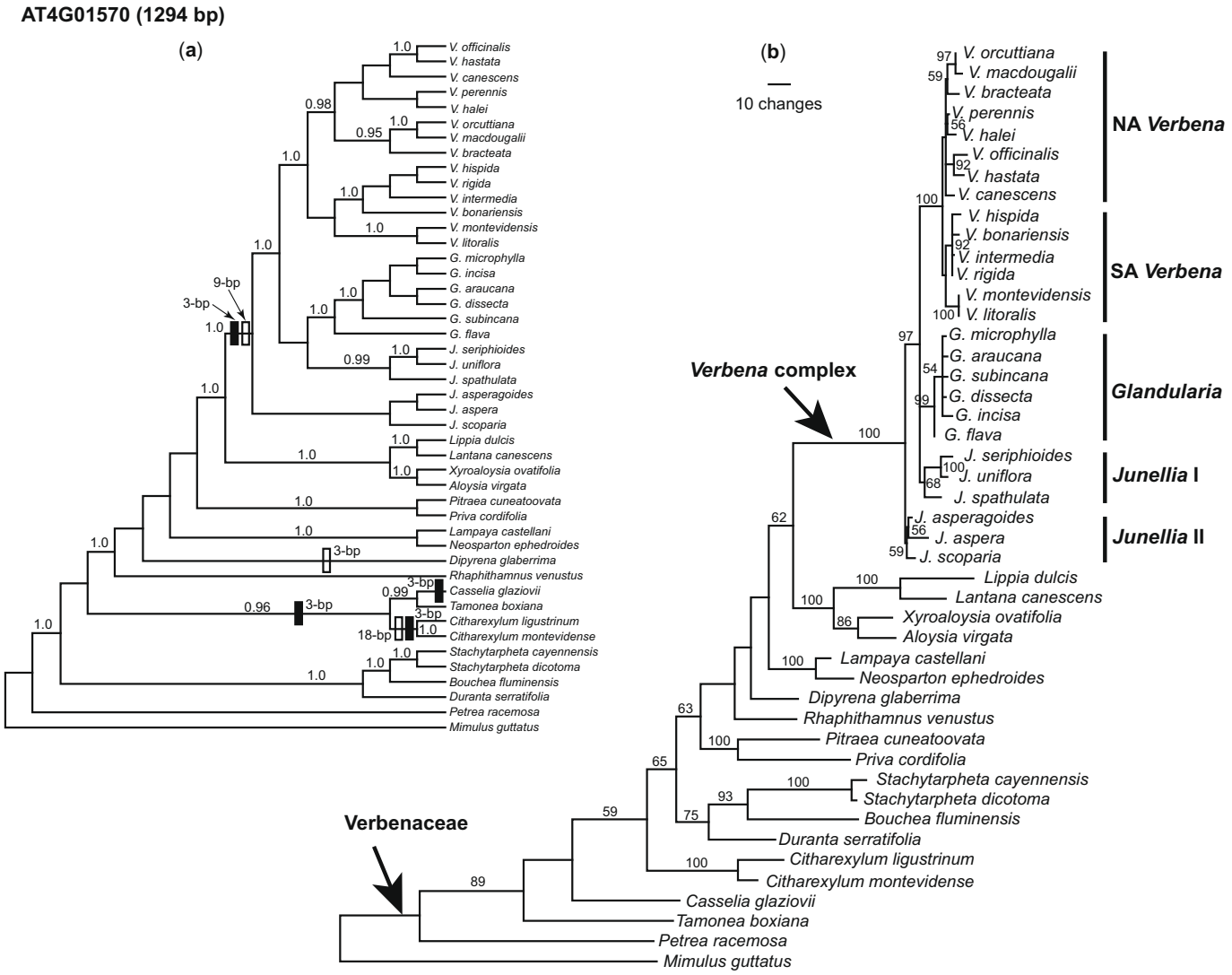

Fig. 4. Phylogenetic trees inferred from AT4G01570 sequences. (a) Bayesian majority-rule consensus tree, with posterior probabilities shown along the branches when >0.95. White and black bars represents deletions and insertions, respectively, with the size of deletions/insertions marked along. (b) Phylogram of the single MP tree (length = 1356; CI = 0.62; RI = 0.75) inferred from the AT4G01570 data set. Clade support values (BS) are shown along the branches when >50, and subdivisions within the *Verbena* complex are designated on the right. Note that topology of the single MP tree differs from the Bayesian consensus tree on the early branched lineages.

our earlier analyses, which is likely due to two major reasons. Firstly, these artifact polymorphisms that are caused by pooling two paralogues together were treated as "uncertainty" states in the analyses. This decrease the information content of the affected sequences, but also decrease the risk of misleading signals from these sequences. Secondly, the duplication was occurred in the most recent common ancestor of *Verbena* and *Glandularia*. It should not affect the rest part of the tree.

3.4. Comparison of sequence divergence and richness of phylogenetic information

Based on the six-species data sets, the PPR gene loci are intermediate between non-coding cpDNA and other nuclear loci that are mainly composed of intronic sequences, in both pair-wise distance and percentage of phylogenetic informative sites (Table 4). From an empirical point of view, percentage of phyogenetic informative sites is more meaningful than pair-wise distance to evaluate phylogenetic utility. The percentage of phylogenetic informative sites for the five PPR gene loci is 1.3–4.1 times that for non-coding cpDNA, whereas that for the other intron-containing nuclear loci is 3.8–6.2 times that for non-coding cpDNA (Table 4).

4. Discussion

The PPR gene family is an appealing phylogenetics marker system due to the combination of three characteristics (Yuan et al., 2009): (1) a large number of loci with established orthology assessment; (2) an absence of introns; (3) a rapid rate of evolution. The first property enables researchers to readily access multiple nuclear loci. The second and third properties allow them to be used best at intergeneric and interspecific levels while circumventing many practical hurdles associated with intron-containing nuclear genes. However, the feasibility of designing primers that can specifically amplify a certain PPR locus is critically important to its application in practice. Considering that all 127 loci screened in our previous study should have a single orthologue in most diploid flowering plants (Yuan et al., 2009), one would expect each of these loci to act as a typical single-copy nuclear gene for primer design. Our

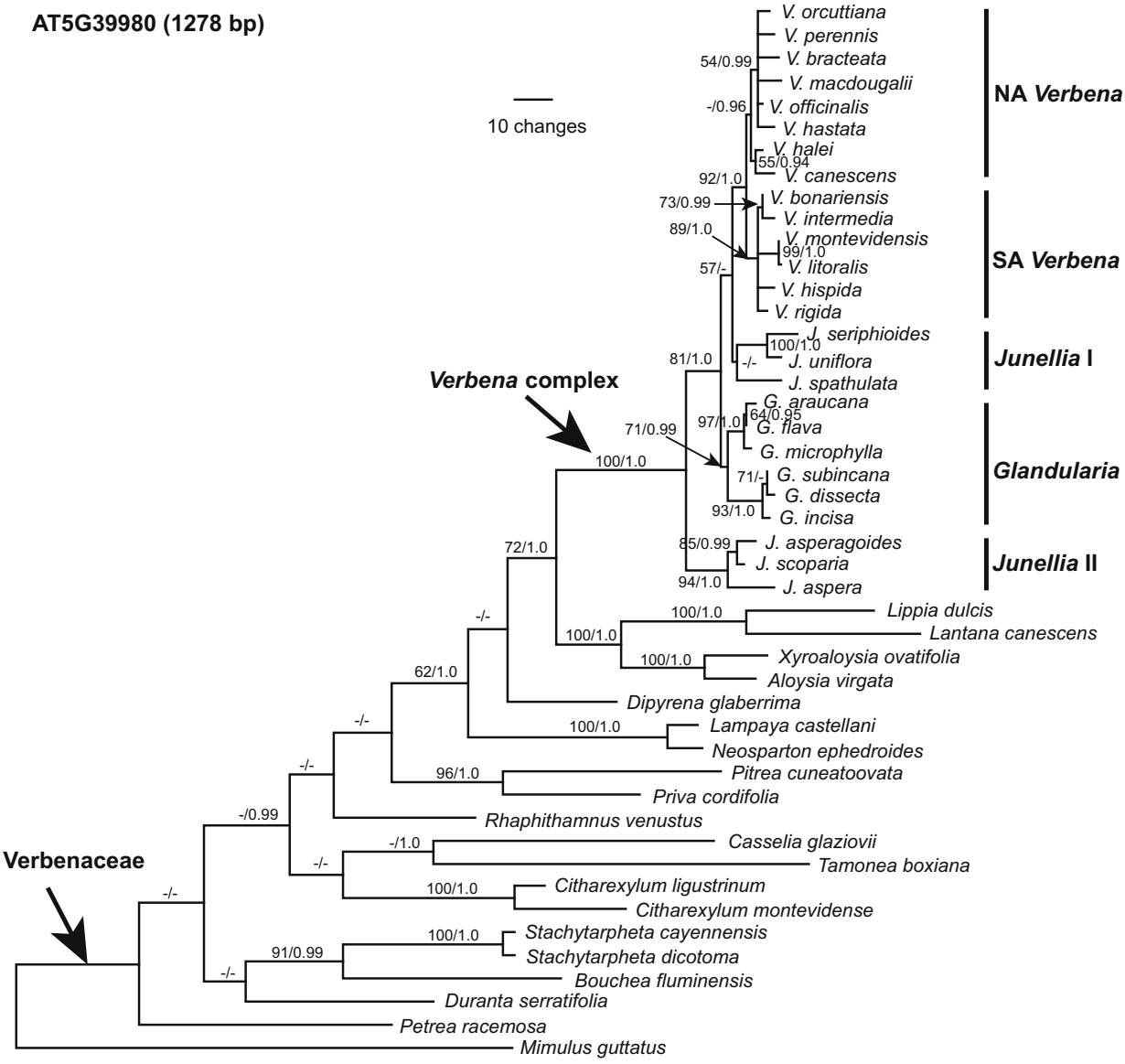

Fig. 5. One of 8 MP trees from AT5G39980 (length = 1431; CI = 0.53; RI = 0.75), shown in the same fashion as Fig. 1.

experience with designing primers for amplifying and sequencing the five arbitrarily selected PPR loci in the family Verbenaceae strongly supports this view. Primers optimized for Lamiales in this study (Table 2) can readily amplify each of the 5 loci for all Verbenaceae taxa except one difficult species, *Casselia glaziovii*. It seems reasonable to assume that following the general scheme presented (see Section 2.2) and taking advantage of the available genomic resources of "model organisms", it should be similarly straightforward to design locus-specific primers for most other angiosperm groups.

Generating PPR gene sequence data required a relatively modest effort compared with most nuclear gene sequencing. All 5 loci were directly sequenced without cloning. Although intra-individual allelic polymorphism was observed in many taxa (see Table 1), they did not confound phylogenetic inferences at the intergeneric level. In addition, alignment of these sequences could hardly be easier. Three of the five loci do not show any length variation within Verbenaceae. The other two loci only have a few short insertions/deletions (3–18 bp), where homology assessment of the adjacent nucleotides can be assisted by the corresponding amino acid alignment.

Intergeneric relationships within the family Verbenaceae are more or less resolved by all individual PPR data sets (Figs. 1–5), and the inferred relationships are consistent from locus to locus, albeit that some of these relationships are only weakly supported by a single locus. Analyses of the concatenated data of all 5 loci (ca. 6.4 kb) generated a fully resolved phylogeny with all intergeneric relationships strongly supported (BS, 90–100%; PP, all 1.0). These results demonstrate that PPR genes can be used as powerful markers to resolve intergeneric relationships in non-model organism systems.

The reference tree inferred from the concatenated data provides a robust phylogenetic framework of the family Verbenaceae, and it suggests that current classifications of this family are problematic. Neither of the two recent classification schemes (Sanders, 2001; Atkins, 2004) based on morphological characters closely reflects this phylogenetic framework (Fig. 6). Two of the four tribes circumscribed by Sanders (2001), Citharexyleae and Verbeneae, are poly-

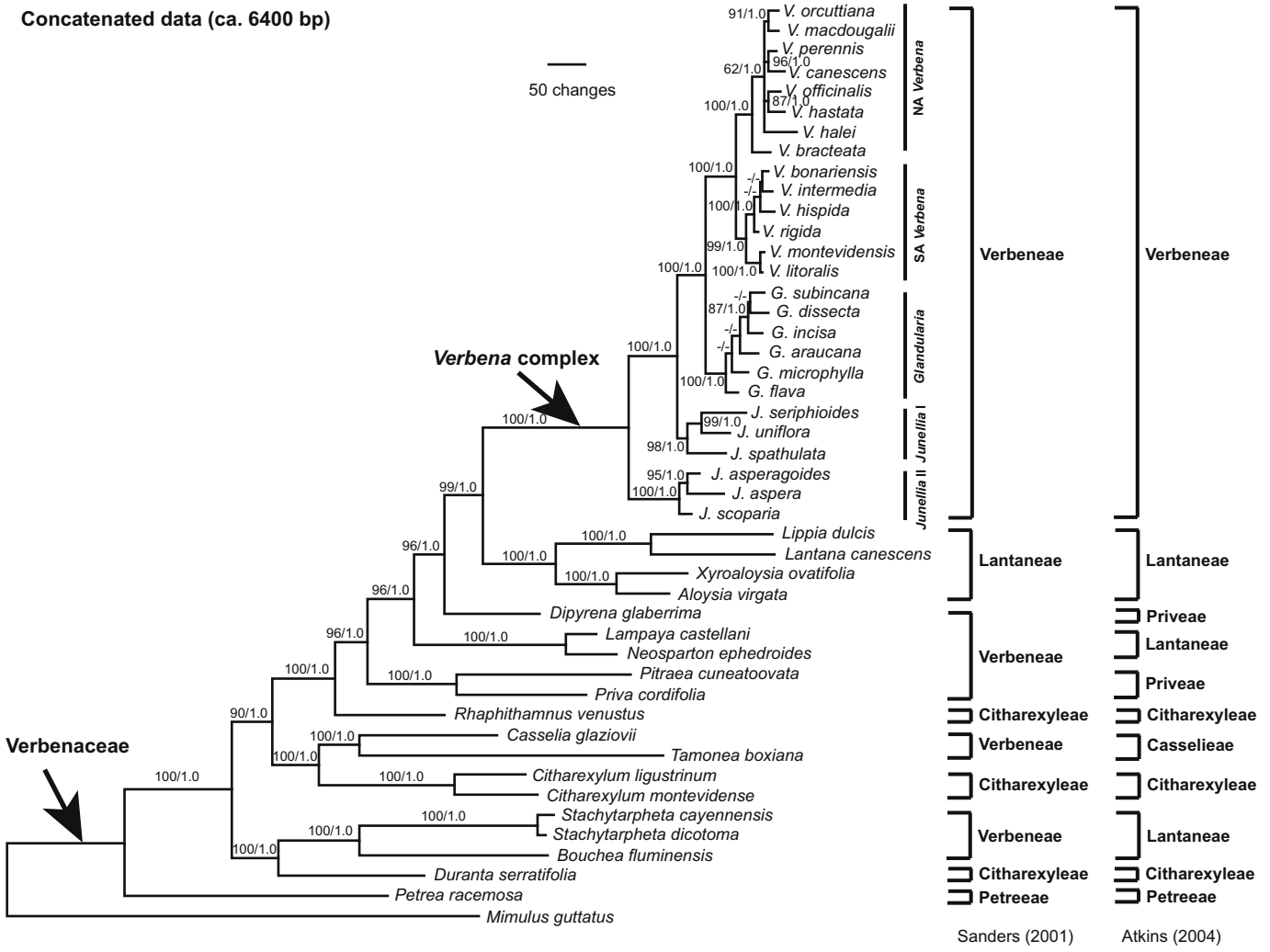

Fig. 6. One of 2 MP trees inferred from the concatenated data of all 5 loci (length = 6957; CI = 0.59; RI = 0.75). Clade support and subdivision designation in the *Verbena* complex are shown in a similar fashion as Fig. 1. Tribe circumscriptions from Sander's (2001) and Atkins' (2004) treatments are shown on the right.

phyletic. Similarly, three of the six tribes delimited by Atkins (2004), Citharexyleae, Lantaneae, and Priveae, are polyphyletic. This molecular phylogenetic framework (Fig. 6) indicates that a phylogenetic classification and re-evaluation of morphological characters used in traditional taxonomic treatments is imperative to shed light on the evolution of this poorly studied, but ecologically important group. A comprehensive sampling of genera in the family and a denser sampling of species in each genus, especially the species-rich ones (e.g., the *Lantana/Lippia/Aloysia* complex, *Stachytarpheta*, *Citharexylum*), are necessary to meet this goal.

The *Verbena* complex is a recent, rapidly diversifying group (Yuan and Olmstead, 2008a), consisting of three closely related genera, *Verbena*, *Glandularia*, and *Junellia*. Both *Verbena* and *Glandularia* are strongly supported as monophyletic groups by the concatenated data as well as all the individual PPR loci except AT2G37230 (Figs. 1–6), which has been involved in a recent gene duplication event (see Section 3.3). The genus *Junellia*, on the other hand, is undoubtedly not monophyletic. Each individual PPR locus recovers *Junellia* either as two distinct clades (Figs. 2, 4, and 5) or a basal grade within the *Verbena* complex (Figs. 1 and 3). The phylogenetic signal supporting two clades of *Junellia* is apparently dominant in the concatenated data set, leading to the high support values in the reference tree for the *Junellia* I and II clades, with the former sister to the *Verbena* + *Glandularia* clade and latter sister to this more inclusive clade (Fig. 6). These results are consistent with previous results based on chloroplast DNA and other nuclear loci (Yuan and Olmstead, 2008a,b; see Table 3).

At the interspecific level, some species groups (e.g., North vs. South American *Verbena*) can be unambiguously differentiated by the PPR genes. In some other species groups (e.g., *J. spathulata*/*J. uniflora*/*J. seriphioides*), there is good resolution to their relationships, but the relationships may be incongruent between different loci, most likely due to incomplete lineage sorting. Some more recently diversified species groups (e.g., *Glandularia*) exhibit limited resolution and species relationships differ between loci. This spectrum can be extended to such an extreme that some very closely related species groups (e.g., North American *Verbena*) can barely be resolved by PPR genes at all. These results from the *Verbena* complex suggest that PPR loci may also be good candidate markers to resolve relationships among closely related genera or not-so-closely related congeneric species, but may have limited use among very closely related species.

5. Conclusion

In this study, we demonstrate empirically the use of PPR genes as powerful nuclear gene markers in the family Verbenaceae and

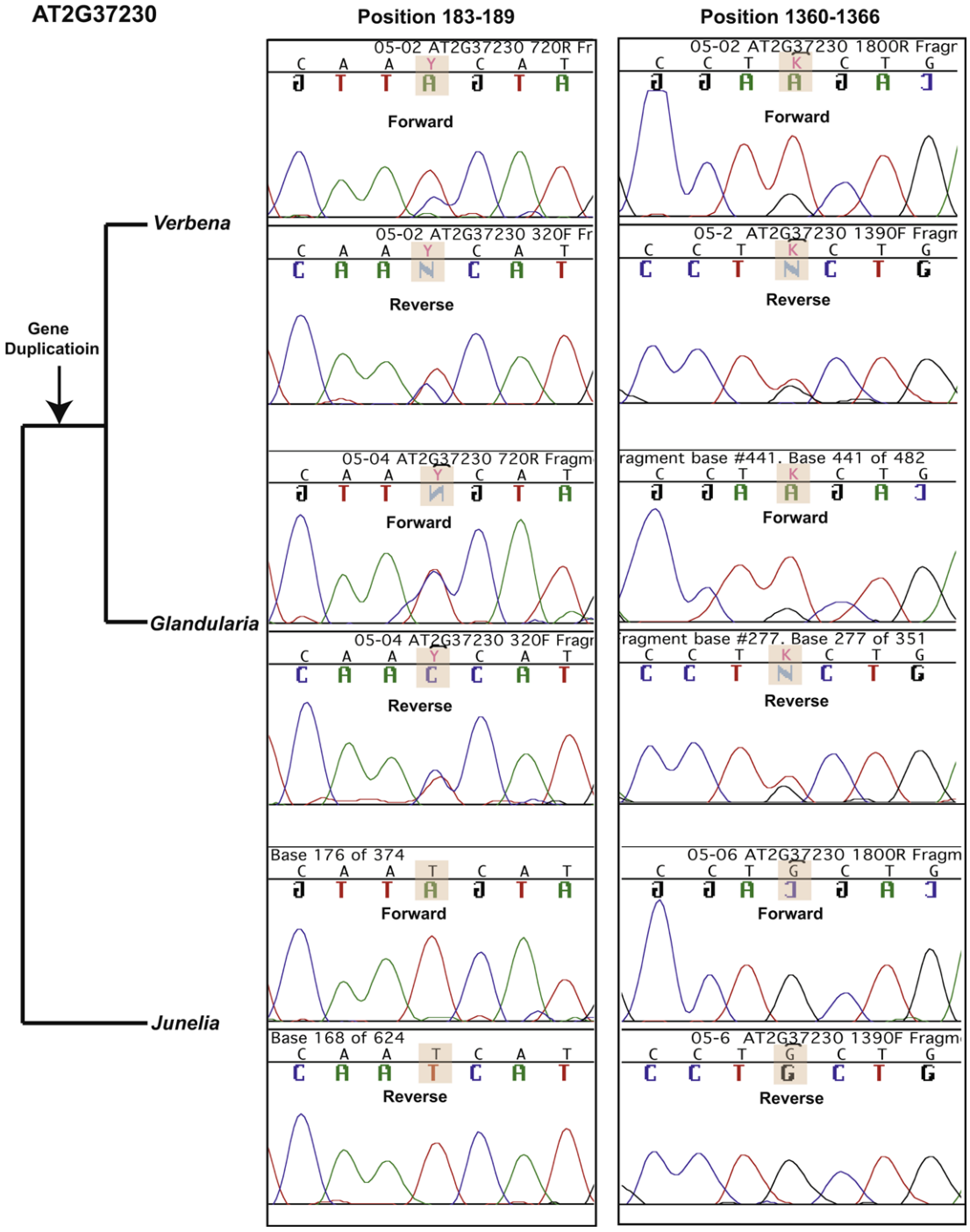

Fig. 7. Portions of chromatograms from the original sequence reads of AT2G37230, showing the polymorphism pattern caused by a recent gene duplication occurred in the common ancestor of *Verbena* and *Glandularia*. The polymorphic sites are highlighted and represented by ambiguity codes, and confirmed by both forward- and reverse-strand sequences. Note that the polymorphic sites are shared by all *Verbena* and *Glandularia* species but not by *Junellia* or other taxa beyond the *Verbena* complex.

the *Verbena* complex. Locus-specific PCR and sequencing primers were optimized for 5 exemplar loci using a general and straightforward scheme for primer design, which we hope will serve as a preliminary guide for empirical systematists who are interested in employing PPR genes in their research. Intergeneric relationships within the family Verbenaceae were fully resolved with strong support, and this first molecular phylogenetic framework should serve as an important step toward a more comprehensive understanding of phylogenetic relationships of the family. Relationships among major lineages within the *Verbena* complex were also well resolved and largely consistent with previous phylogenetic studies on this complex, but relationships among very closely related species were poorly resolved. Taking these results together, the PPR genes can be readily employed in non-model organisms and should be best used to resolve relationships in a spectrum from among distantly related genera to among not-so-closely related congeneric species, but may have limited use among very closely related species.

Table 3
Comparison of phylogenetic relationships resolved by different loci among major lineages within the *Verbena* complex. The chloroplast DNA (cpDNA) and nuclear *waxy* gene results are from Yuan and Olmstead (2008a), and the nuclear *PHOT1* and *PHOT2* gene results are from Yuan and Olmstead (2008b). The dashes indicate that the corresponding results are not available due to a recent gene duplication of AT2G37230.

Relationships	Locus									
	Concatenated 5 PPR loci	cpDNA	*waxy*	*PHOT1*	*PHOT2*	AT1G09680	AT2G37230	AT3G09060	AT4G01570	AT5G39980
Both *Verbena* and *Glandularia* are monophyletic	YES	NO	YES	YES	YES	YES	—	YES	YES	YES
Junellia form two clades with relationship (*Junellia* II, (*Junellia* I, (*Glandularia*, *Verbena*)))	YES	YES	YES	NO	NO	NO	YES	NO	YES	NO
Junellia form two clades with relationship (*Junellia* II, (*Glandularia*, (*Junellia* I, *Verbena*)))	NO	NO	NO	NO	YES	NO	NO	NO	NO	YES
Junellia form a basal grade	NO	NO	NO	YES	NO	YES	NO	YES	NO	NO
Both South American and North American *Verbena* are monophyletic	YES	YES	NO	NO	NO	YES	—	YES	YES	YES

Table 4
Comparison of sequence divergence and richness of phylogenetic information across the five PPR loci and previously used cpDNA and nuclear loci (Yuan and Olmstead, 2008a,b). The comparison is based on a subset of data including three species pairs. The dashes indicate excluded results that are not meaningful due to a recent gene duplication of AT2G37230.

		cpDNA	*waxy*	*PHOT1*	*PHOT2*	AT1G 09680	AT2G 37230	AT3G 09060	AT4G 01570	AT5G 39980
Uncorrected *p*-distances (%)	*Verbena canescens* vs *V. macdougalii*	0.23	2.38	2.01	0.75	0.56	—	0.86	0.93	0.94
	Glandularia microphylla vs *G. incisa*	0.81	5.34	4.84	1.44	1.99	—	1.08	0.39	1.50
	Junellia uniflora vs *J. spathulata*	0.70	3.74	4.36	3.02	2.31	1.03	1.77	1.24	1.88
Phylogenetically informative (PI) sites	Percentage (%)	0.68	3.84	2.57	4.20	2.78	—	2.39	2.16	0.86
	No. of PI sites	35	124	110	123	35	—	27	27	11
	No. of total sites	5115	3226	4273	2933	1257	—	1128	1252	1278

Acknowledgments

The authors thank two anonymous reviewers for comments on the paper. This research was supported by an NSF Grant (DEB-0542493) to R.G.O. and an NSF Doctoral Dissertation Improvement Grant (DDIG) (DEB-0710026) to R.G.O. for the first author's dissertation research.

References

Akaike, H., 1974. A new look at the statistical model identification. IEEE Trans. Automat. Control. 19, 716–723.
The Angiosperm Phylogeny Group, 2003. An update of the Angiosperm Phylogeny Group classification for the orders and families of flowering plants: APG II. Bot. J. Linn. Soc. 141, 399–436.
Atkins, S., 2004. Verbenaceae. In: Kubitzki, K., (Ed.), The families and genera of vascular plants, vol. 7, pp. 449–468.
Beardsley, P.M., Olmstead, R.G., 2002. Redefining Phrymaceae: the placement of *Mimulus*, tribe Mimuleae and *Phryma*. Am. J. Bot. 89, 1093–1102.
Childs, K.L., Hamilton, J.P., Zhu, W., Ly, E., Cheung, F., Wu, H., Rabinowicz, P.D., Town, C.D., Buell, C.R., Chan, A.P., 2007. The TIGR plant transcript assemblies database. Nucleic Acids Res. 35, D846–D851.
Delannoy, E., Stanley, W.A., Bond, C.S., Small, I.D., 2007. Pentatricopeptide repeat (PPR) proteins as sequence-specificity factors in post-transcriptional processes in organelles. Biochem. Soc. T. 35, 1643–1647.
Doyle, J.J., Doyle, J.L., 1987. A rapid DNA isolation procedure for small quantities of fresh leaf tissue. Phytochem. Bull. 19, 11–15.
Felsenstein, J., 1985. Confidence limits on phylogenies – an approach using the bootstrap. Evolution 39, 783–791.
Jaillon, O, Aury, J.M., Noel, B., et al. (56 co-authors), 2007. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449, 463–467.
Lurin, C., Andres, C., Aubourg, S., Bellaoui, M., Bitton, F., Bruyere, C., Caboche, M., Debast, C., Gualberto, J., Hoffmann, B., et al., 2004. Genome-wide analysis of *Arabidopsis* pentatricopeptide repeat proteins reveals their essential role in organelle biogenesis. Plant Cell 16, 2089–2103.
Moldenke, H., 1958. Materials toward a monograph of the genus Citharexylum. I. Phytologia 6, 242–256.
O'Leary, N., Múlgura, M.E., Morrone, O., 2007. Revisión taxonómica de las especies del género *Verbena* (Verbenaceae): serie *Pachystachyae*. Ann. Mo. Bot. Gard. 94, 571–622.
O'Leary, N, Múlgura, M.E., Morrone, O., in press. Revisión taxonómica de las especies del género *Verbena* (Verbenaceae): serie *Verbena*. Ann. Mo. Bot. Gard.
O'Toole, N., Hattori, M., Andres, C., Iida, K., Lurin, C., Schmitz-Linneweber, C., Sugita, M., Small, I., 2008. On the expansion of the pentatricopeptide repeat gene family in plants. Mol. Biol. Evol. 25, 1120–1128.
Peralta, P., 2009. Ph.D. Dissertation. Universidad de La Plata, Argentina.
Peralta, P., Múlgura, M.E., Denham, S.S., Botta, S., 2008. Revisión del género *Junellia* (Verbenaceae). Ann. Mo. Bot. Gard. 95, 338–390.
Posada, D., Crandall, K.A., 1998. Modeltest: testing the model of DNA substitution. Bioinformatics 14, 817–818.
Pusnik, M., Small, I., Read, L.K., Fabbro, T., Schneider, A., 2007. Pentatricopeptide repeat proteins in *Trypanosoma brucei* function in mitochondrial ribosomes. Mol. Cell. Biol. 27, 6876–6888.
Rambaut, A., 1996. Se-Al: Sequence Alignment Editor. Oxford, University of Oxford. Available at: <http://evolve.zoo.ox.ac.uk/>.
Rambaut, A., Drummond, A.J., 2007. Tracer v1.4. Available from: <http://tree.bio.ed.ac.uk/software/tracer/>.
Ronquist, F., Huelsenbeck, J.P., 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574.
Sanders, R.W., 2001. The genera of Verbenaceae in the Southeastern United States. Harv. Paper Bot. 5, 303–358.
Schmitz-Linneweber, C., Small, I., 2008. Pentatricopeptide repeat proteins: a socket set for organelle gene expression. Trends Plant Sci. 13, 663–670.
Small, I., Peeters, N., 2000. The PPR motif – a TPR-related motif prevalent in plant organellar proteins. Trends Biochem. Sci. 25, 46–47.
Swofford, D.L., 2002. PAUP*: phylogenetic analysis using parsimony (* and other methods), version 4b10. Sinauer Associates, Inc, Sunderland, Massachusetts, USA.
Tuskan, G.A., DiFazio, S., Jansson, S., et al. (110 co-authors), 2006. The genome of black cottonwood, *Populus trichocarpa* (Torr. and Gray). Science 313, 1596–1604.
Walker, N.S., Stiffler, N., Barkan, A., 2007. POGs/PlantRBP: a resource for comparative genomics in plants. Nucleic Acids Res. 35, D852–D856.

Yuan, Y.W., Olmstead, R.G., 2008a. A species-level phylogenetic study of the *Verbena* complex (Verbenaceae) indicates two independent intergeneric chloroplast transfers. Mol. Phylogenet. Evol. 48, 23–33.

Yuan, Y.W., Olmstead, R.G., 2008b. Evolution and phylogenetic utility of the *PHOT* gene duplicates in the *Verbena* complex (Verbenaceae): dramatic intron size variation and footprint of ancestral recombination. Am. J. Bot. 95, 1166–1176.

Yuan, Y.W., Liu, C., Marx, H.E., Olmstead, R.G., 2009. The pentatricopeptide repeat (PPR) gene family, a tremendous resource for plant phylogenetic studies. New Phytol. 182, 272–283.