RESEARCH NOTE

Are mini DNA-barcodes sufficiently informative to resolve species identities? An *in silico* analysis using *Phyllanthus*

R. SRIRAMA[1,2,3], B. R. GURUMURTHY[4], U. SENTHILKUMAR[1,3], G. RAVIKANTH[1,3]*, R. UMA SHAANKER[1,3] and M. B. SHIVANNA[2]

[1]*School of Ecology and Conservation and Department of Crop Physiology, University of Agricultural Sciences, GKVK, Bangalore 560 065, India*
[2]*Department of Applied Botany, Kuvempu University, Shankaraghatta, Shimoga 577 451, India*
[3]*Ashoka Trust for Research in Ecology and the Environment, Royal Enclave, Srirampura, Jakkur Post, Bangalore 560 064, India*
[4]*University of Agriculture and Horticultural Sciences, Shimoga 577 204, India*

[Srirama R., Gurumurthy B. R., Senthilkumar U., Ravikanth G., Uma Shaanker R. and Shivanna M. B. 2014 Are mini DNA-barcodes sufficiently informative to resolve species identities? An *in silico* analysis using *Phyllanthus*. *J. Genet.* **93**, 823–829]

Introduction

DNA barcodes using certain regions of the genomic DNA have become a popular diagnostic tool to assign species-specific signatures. In plants, a number of chloroplast DNA regions such as *psbA-trnH* of size 400 bp have been shown to successfully discriminate members of various taxa. However, the technique is not always successful, as in the case of museum specimens or samples used in raw drug trade, where the DNA is often degraded. In this context, recent studies have suggested the use of shorter stretches of the region, called mini-barcodes, to resolve species identity. The mini-barcodes are relatively more stable and easily recovered from the degraded DNA. In this study, an attempt has been made to compare the effectiveness of mini-barcodes over full-length DNA barcodes in differentiating 16 species of *Phyllanthus* (Phyllanthaceae) used in the raw drug trade. Using an *in silico* approach, mini-barcodes of varying lengths (50–200 bp) of the region *psbA-trnH* were generated and evaluated for their ability to resolve the 16 *Phyllanthus* species in comparison to the full-length DNA barcode of size 398 bp. Results have been discussed in the light of the overall utility of mini-barcodes in resolving the species identities.

First developed in animal systems, barcodes were used to demonstrate the ability of the mitochondrial gene cytochrome *c* oxidase subunit I (*COI*, 650 bp) to completely resolve over 200 closely related species of Lepidopterans (Herbert *et al.* 2003). Since then, the COI has been effectively used as 'universal DNA barcode' in several animal groups such as birds, butterflies, amphibians and fishes (Hebert *et al.* 2003; Gu *et al.* 2011). However, in plants, the *COI* was found to be ineffective in discriminating the taxa, because the region has a very low divergence rate (Cho *et al.* 2004) and is prone to rapid change in genome structure (Adams and Palmer 2003). Against this failure, the Consortium of Barcode of Life (CBOL) programme evaluated seven other potential DNA barcodes using 907 plants representing 445 angiosperms, 38 gymnosperms and 67 cryptogamic species. The seven potential regions, all from the chloroplast, were *atpF-atpH* spacer, *matK* gene, *rbcL* gene, *rpoB* gene, *rpoC1* gene, *psbK-psbI* spacer and *trnH-psbA* (CBOL 2009). Among these, CBOL recommended the combination of *rbcL* and *matK* as the potential plant DNA barcode (CBOL 2009). However, besides these, several regions such as *trnH-psbA* were also found to be very effective (Kress and Erickson 2007). The latter region is easy to amplify and is one of the most variable intergenic spacers in plants (Shaw *et al.* 2007). This region has been successfully used in many studies (Gonzalez *et al.* 2009; Kress *et al.* 2009, 2010) and can be used as a supplementary barcode for plants.

In a previous study, we showed that the *psbA-trnH* region could effectively resolve 16 species of *Phyllanthus* that were used in raw herbal trade in India (Srirama *et al.* 2010). Thus, using DNA barcodes or DNA signatures, it is possible to effectively discriminate species used in herbal trade and further identify adulterating plant material, if any, in shipments of raw herbal trade material. However, the success of the technique as an identification service depends upon the ease of DNA extraction and its subsequent amplification at the desired gene region. Since, the raw herbal drug material are very dry and not stored under ideal conditions, the DNA is either sheared or not very appropriate for amplification.

*For correspondence. E-mail: gravikanth@atree.org.

Keywords. DNA barcoding; minibarcodes; species identity; *Phyllanthus*.

Failure in amplifying and sequencing the degraded DNA of museum samples has been reported earlier (Whitfield 1999; Hajibabaei *et al.* 2005, 2006; Meusnier *et al.* 2008).

In an attempt to address this constraint and still be able to resolve species with degraded DNA in museum samples, in processed food products and in raw drug trade, several studies have explored the possibility of using short stretches of DNA as 'mini-barcodes' (100–300 bp) to distinguish the species. In contrast to the regular bar codes that range between 300 bp–1 kb, the mini-barcodes are usually stable and easily recovered from degraded DNA (Hajibabaei *et al.* 2006; Meusnier *et al.* 2008; Zimmermann *et al.* 2008). Utility of mini-barcodes have been successfully demonstrated in a number of plants (Sonstebo *et al.* 2010), fishes (Hajibabaei *et al.* 2006; Meusnier *et al.* 2008; Saitoh *et al.* 2008; Baumstegier and Kerby 2009; Ficetola *et al.* 2010; Hajibabaei and McKenna 2012), reptiles (Dubey *et al.* 2010), birds (Lee and Prysjones 2008; Meusnier *et al.* 2008; Patel *et al.* 2009), arthropods (Dean and Ballard 2001; Hajibabaei *et al.* 2006; Meusnier *et al.* 2008; Houdt *et al.* 2009; Smith and Fisher 2009; Rougerie *et al.* 2010; Hajibabaei *et al.* 2011), fungi (Meusnier *et al.* 2008; Houdt *et al.* 2009), mammals (Meusnier *et al.* 2008; Ficetola *et al.* 2010) and insects (Meusnier *et al.* 2008).

In this study, we compared the effectiveness of mini-barcodes over full-length DNA barcodes in discriminating different species of *Phyllanthus* (Phyllanthaceae) used in raw drug trade. The mini-barcodes of varying lengths (50–200 bp) of the region *psbA-trnH* were generated and evaluated using an *in silico* approach for their ability to resolve the 16 *Phyllanthus* species. The results are discussed considering the overall utility of mini-barcodes in resolving species identities.

Materials and methods

In silico analysis

For analysing the effectiveness of mini-barcodes, the DNA barcode (398 bp) of 16 *Phyllanthus* species obtained using the chloroplast intergenic spacer *psbA-trnH* was used (Srirama *et al.* 2010, GenBank accession no. GQ409804-23, GU598536-60; table 1). Using this data set, for each of the species; subsets of DNA barcodes representing mini-barcodes of sizes 50, 75, 100, 150 and 200 bp from the 5′ end were derived. To obtain the 50 bp size fragments, the full length sequence (398 bp) was segmented into seven bits of 50 bases each and numbered according to their position relative to the full-length (5′–3′) barcode region (table 2). Similarly, the full-length sequence was divided into 75, 100, 150 and 200 base fragment barcodes (table 2).

For each such mini-barcodes, the pairwise distances (*p* distance) between all pairs of the 16 species were computed. This distance is the proportion (*p*) of nucleotide sites, at which two sequences being compared are different. It was obtained by dividing the number of nucleotide differences by the total number of nucleotides compared. The *p* distance values between individuals of each species were multiplied by 100 and taken as the per cent intraspecific divergence. Similarly, the *p* distance values between species were also multiplied by 100 and were taken as the per cent interspecific divergence. The computations were carried out in MEGA 5.0 (table 2). The variance in *p* distance for each of the mini-barcode lengths was also computed.

The neighbour-joining trees (NJ trees) were constructed using each mini-barcode in MEGA 5.0 software (http://www.megasoftware.net). The NJ trees developed using the

Table 1. List of *Phyllanthus* species occurring in south India along with their voucher numbers, collection locations and GenBank accession numbers.

	Species name	Voucher no.	Location	GenBank accession no.
1	*P. kozhikodianus* Sivar. & Manilal	SK124A, C, D	Kerala Forest Research Institute, Peechi, Kerala	GQ409804–6
2	*P. rheedi* Wight	SK117A–C	Neliyampathy, Palghat, Kerala	GQ409807–9
3	*P. debilis* Klein ex Willd.	SK105A	Nesari, Kolhapur, Maharashtra	GQ409810
		SK119	Alleppey, Kerala	GQ409811
		SK122B	Ernakulam, Kerala	GQ409812
4	*P. urinaria* L.	SK114A–C	Neliyampathy, Palghat, Kerala	GQ409813–15
5	*P. amarus* Schumach.	SK101A–C	Bangalore, Karnataka	GQ409816–18
		SK104A–B	Pune, Maharashtra	GQ409819–20
6	*P. tenellus* Roxb.	SK116A–C	Bangalore, Karnataka	GQ409821–23
7	*P. reticulatus* Poir.	SK226A, C	Bangalore, Karnataka	GU598539–40
8	*P. lawii* J. Graham	SK265A, SK506 A	Coorg, Karnataka	GU598556–57
9	*P. rotundifolius* Klein ex Willd.	SK414A–C	Kanyakumari, Tamil Nadu	GU598548–50
10	*P. maderaspatensis* L.	SK112A–C	Chennai, Tamil Nadu	GU598536–38
11	*P. missionis* Hook.f.	SK484A, B; SK547A	Courtallum, Tamil Nadu	GU598553–55
12	*P. emblica* L.	SK227B	Bangalore, Karnataka	GU598547
13	*P. indofischeri* Bennet	SK541A–C	BRT Hills, Karnataka	GU598558–60
14	*P. talbotii* Sedgw.	SK554A, B	Moem, Goa	GU598551–52
15	*P. acidus* (L.) Skeels	SK225A–C	Bangalore, Karnataka	GU598541–43
16	*P. polyphyllus* Willd.	SK115A–C	Bangalore, Karnataka	GU598544–46

Table 2. Mean intraspecific and interspecific nucleotide divergences and per cent species resolution obtained from different mini-barcodes.

DNA barcode	Length (bp) and position of mini-barcode from 5′ end	Interspecific nucleotide divergence ± SD ($n = 986$)	Intraspecific nucleotide divergence ± SD ($n = 49$)	Per cent species resolution
Full-length barcode	398, 1–398	0.14 ± 0.09	0.004 ± 0.007	100.0
Mini-barcode 50-1	50, 1–50	0.08 ± 0.06	0.007 ± 0.01	56.25
Mini-barcode 50-2	50, 50–100	0.08 ± 0.11	0 ± 0.0	18.75
Mini-barcode 50-3	50, 100–150	0.17 ± 0.16	0.0 ± 0.0	37.5
Mini-barcode 50-4	50, 150–200	0.18 ± 0.18	0.000 ± 0.005	43.75
Mini-barcode 50-5	50, 200–250	0.15 ± 0.16	0.0 ± 0.0	37.5
Mini-barcode 75-1	75, 1–75	0.06 ± 0.04	0.006 ± 0.016	62.5
Mini-barcode 75-2	75, 76–150	0.10 ± 0.08	0 ± 0.004	18.75
Mini-barcode 75-3	75, 151–225	0.17 ± 0.13	0.0 ± 0.0	50.0
Mini-barcode 75-4	75, 226–300	0.19 ± 0.17	0.032 ± 0.011	43.75
Mini-barcode 100-1	100, 1–100	0.08 ± 0.04	0.004 ± 0.010	87.5
Mini-barcode 100-2	100, 50–150	0.19 ± 0.11	0.0 ± 0.003	62.5
Mini-barcode 100-3	100, 100–200	0.14±0.13	0.000 ± 0.003	68.75
Mini-barcode 100-4	100, 150–250	0.19 ± 0.17	0.004 ± 0.008	75.0
Mini-barcode 100-5	100, 200–300	0.29 ± 0.24	0.004 ± 0.015	56.25
Mini-barcode 150-1	150, 1–150	0.19 ± 0.17	0.004 ± 0.008	100.0
Mini-barcode 150-2	150, 50–200	0.14 ± 0.11	0.004 ± 0.003	75.0
Mini-barcode 150-3	150, 101–250	0.19 ± 0.16	0.003 ± 0.010	56.25
Mini-barcode 150-4	150, 150–300	0.19±0.17	0.004 ± 0.008	100.0
Mini-barcode 150-5	150, 200–350	0.31 ± 0.25	0.008 ± 0.018	56.25
Mini-barcode 200-1	200, 1–200	0.13 ± 0.05	0.002 ± 0.006	100.0
Mini-barcode 200-2	200, 50–250	0.17 ± 0.13	0.002 ± 0.006	100.0
Mini-barcode 200-3	200, 100–300	0.19 ± 0.15	0.008 ± 0.024	87.5
Mini-barcode 200-4	200, 150–350	0.3 ± 0.29	0.016 ± 0.055	25.0
Mini-barcode 200-5	198, 200–398	0.30 ± 0.21	0.011 ± 0.043	50.0

mini-barcodes were then compared with that of the full-length DNA barcode (Hajibabaei and McKenna 2012). The robustness of the clades was checked using the boot-strap analysis (1000 replicates). The resolution of the NJ trees, i.e. species resolution developed using the mini-barcodes and full-length DNA barcodes were analysed. Species resolution of a mini-barcode was computed (using MEGA 5.0) by dividing the number of species, whose individuals clustered together into separate clades with the total number of species used in the phylogenetic NJ tree and this value was converted into percentage value (table 2, Tamura *et al.* 2011; Hajibabaei and McKenna 2012). Correlation of *p* distances between the mini-barcode and full-length barcodes and a scatter plot of the pairwise interspecific nucleotide distances between mini-barcode and the full-length DNA barcode was carried out in Microsoft Excel (http://www.office.microsoft.com). For this analysis, after multiple alignments, since only 360 bp of the full-length DNA barcode aligned with each of the mini-barcodes, the analysis was restricted only to 360 bp.

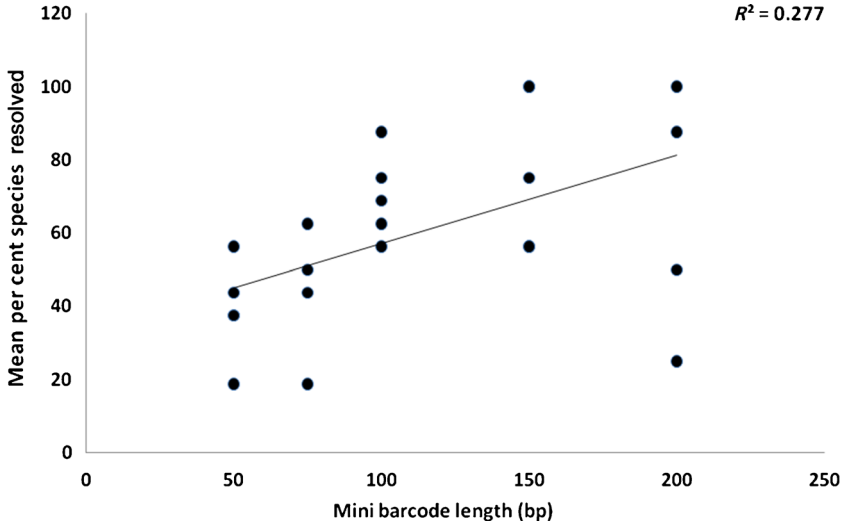


Figure 1. The percentage of species resolved in relation to the length of mini-barcode.

Results and discussion

The mean interspecific and intraspecific divergence using the full-length DNA barcode of the 16 species of *Phyllanthus* was 14.92 and 0.48%, respectively (Srirama *et al.* 2010). In contrast, the mean interspecific divergence distance obtained from the various mini-barcodes ranged from as low as 8.87% to as high as 30% (table 2). The interspecific divergence distance for 50-bp barcodes ranged from 8.87 to 18.33%, 75-bp barcodes from 6.9 to 19.4% and for 100 bp barcodes from 8.91 to 29.14%. Interspecific divergence distance for 150 and 200-bp barcodes ranged from 13 to 30% (table 2).

The mean per cent species resolved increased with the length of the barcode (figure 1). However, it is important to note that for certain mini-barcodes (150 and 200 bp), one of the iterations of the barcodes (e.g., 150-1 and 200-1) could resolve 100% of the species (table 2). This is also reflected in the dendrogram illustrating the species relationships. The dendrogram generated using the full-length barcode sequence differentiated all the species into separate clades and showed 100% species resolution (Srirama *et al.* 2010) (figure 2). A similar clustering pattern of the 16 species was obtained using mini-barcodes of 150 and 200 bp sequence lengths (figure 2). However, for most other short sequence lengths, species were not resolved as compared to the full-length DNA barcodes of *Phyllanthus* species (figure 1). These results were also reflected in the correlation between the *p* distances among various mini-barcodes and the full-length barcode (figure 3). Clearly, the correlations were stronger as the length of the mini-barcode increased

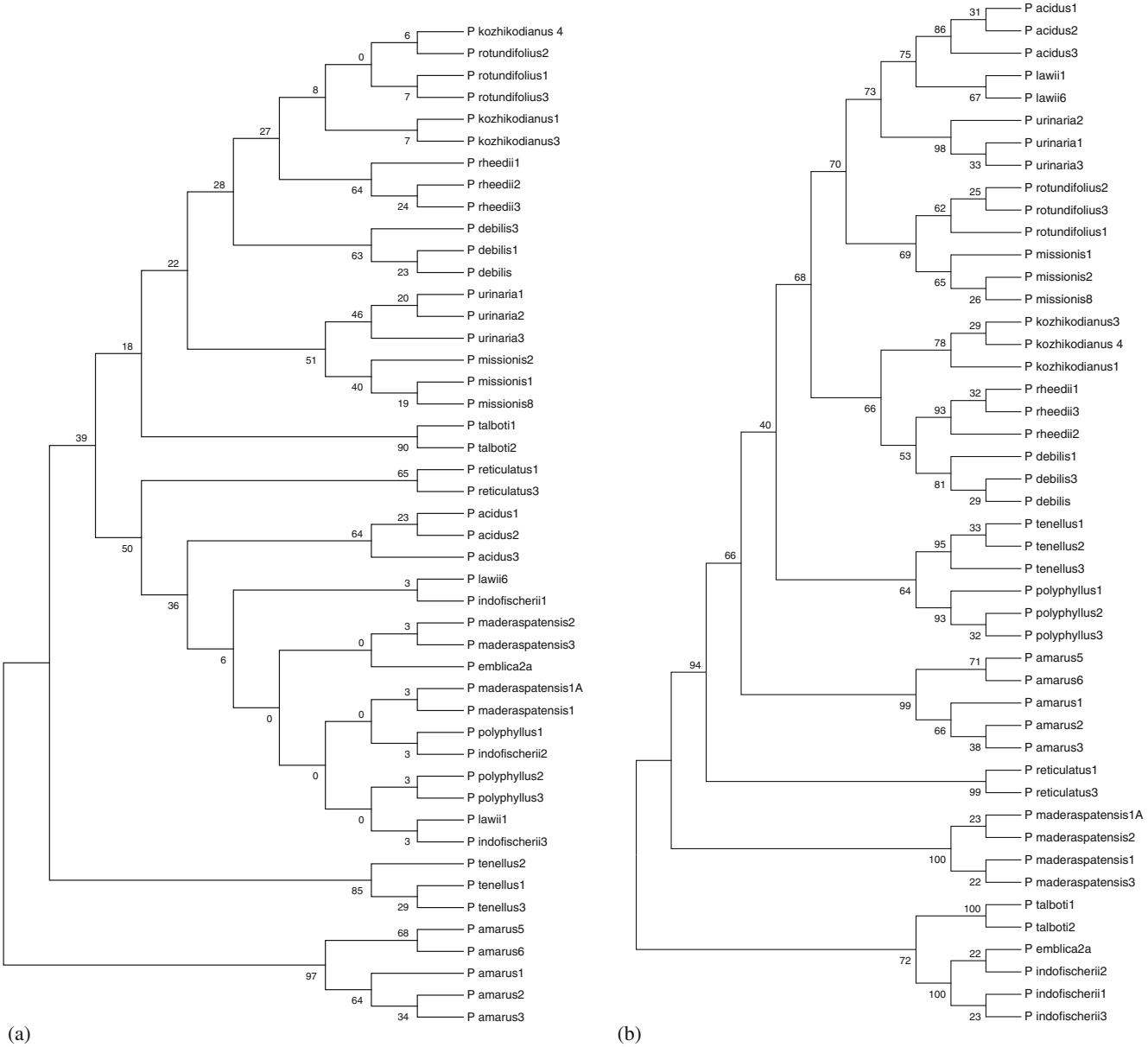
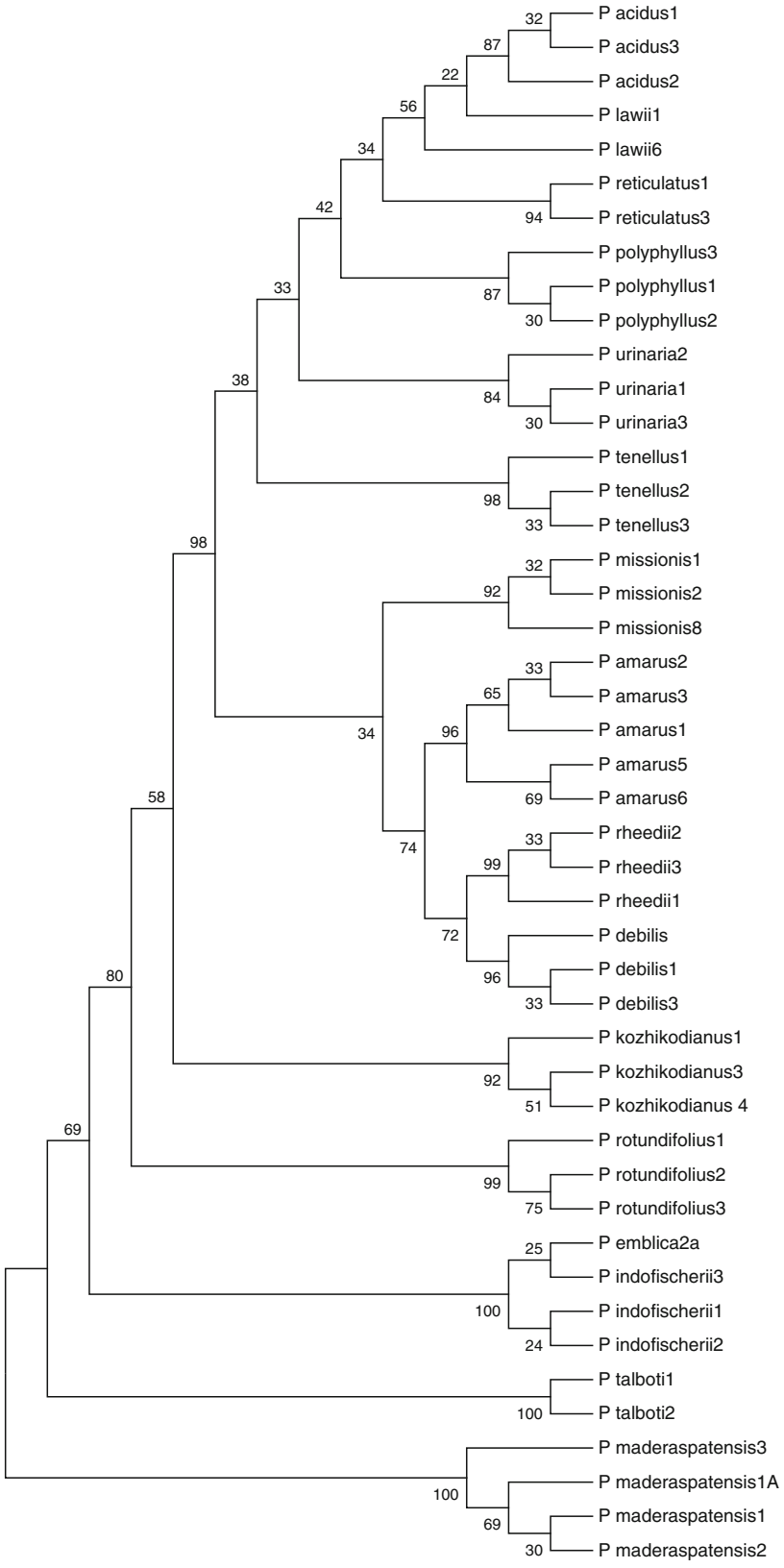


Figure 2 (*continues*)



(c)

Figure 2. NJ tree of 16 *Phyllanthus* species using (a) 50-bp mini-barcode and (b) 200-bp mini-barcode and (c) full-length DNA barcode. The values at the tree nodes are the bootstrap values.

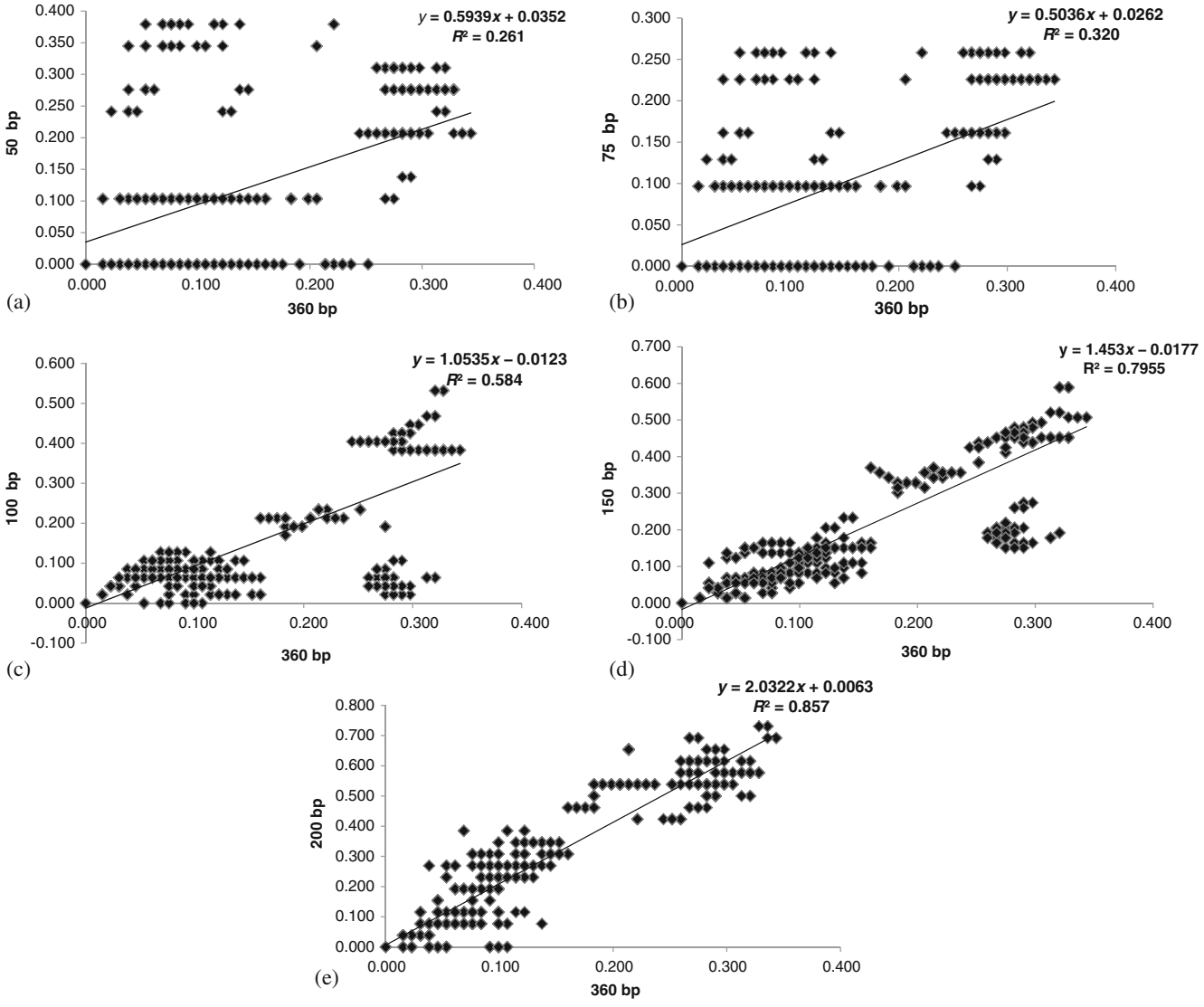


Figure 3. Scatter plot of the pair-wise interspecific nucleotide distances between mini-barcode and the full-length DNA barcode. a, 50 bp; b, 75 bp; c, 100 bp; d, 150 bp; e, 200 bp.

from 50 to 200 bp. In summary, the results indicate that while mini-barcodes may resolve species identities as much as is accomplished by full-length barcodes, they are associated with a relatively high degree of variance. In other words, the resolution offered by mini-barcodes may only be sufficiently informative but not necessarily exhaustive. In this regard, due caution may have to be exercised while using mini-barcodes to unravel species identities pertaining to museum samples or those used in raw drug trade. Though not explicitly stated, this has been echoed earlier by Hajibabaei *et al.* (2006).

Hajibabaei *et al.* (2006) compared the percentage of intraspecific and intrageneric divergences between the mini-barcodes (109 and 218 bp from the 5′ end) and full-length DNA barcode (654 bp) in 204 fish and 61 Lepidopteran species. The identification of fish or Lepidoptera was as accurate with mini-barcodes as with the full-length DNA barcodes (Hajibabaei *et al.* 2006). However, they also showed that the results were sensitive to the choice of length and position of the mini-barcodes (Hajibabaei *et al.* 2006).

Acknowledgements

The work was supported by grants from the Department of Biotechnology, Government of India to G. Ravikanth and R. Uma Shaanker.

References

Adams K. L. and Palmer J. D. 2003 Evolution of mitochondrial gene content: gene loss and transfer to the nucleus. *Mol. Phylogenet. Evol.* **29**, 380–395.

Baumstegier J. and Kerby J. L. 2009 Effectiveness of salmon carcass tissue for use in DNA extraction and amplification in conservation genetic studies. *N. Am. J. Fish Manag.* **29**, 40–49.

CBOL Plant Working Group 2009 A DNA barcode for land plants. *Proc. Natl. Acad. Sci. USA* **106**, 12794–12797.

Cho Y., Mower J. P., Qiu Y.-L. and Palmer J. D. 2004 Mitochondrial substitution rates are extraordinarily elevated and variable within a genus of flowering plants. *Proc. Natl. Acad. Sci. USA* **101**, 17741–17746.

Dean M. D. and Ballard J. W. 2001 Factors affecting mitochondrial DNA quality from museum preserved *Drosophila simulans*. *Entomol. Exp. Appl.* **98**, 279–283.

Dubey B., Meganathan P. R. and Haque I. 2010 DNA mini-barcoding: an approach for forensic identification of some endangered snake species. *Forensic Sci. Int. Genet.* **5**, 181–184.

Ficetola G. F., Coissac E., Zundel S., Riaz T., Shehzad W., Bessiere J. *et al.* 2010 An *in silico* approach for the evaluation of DNA barcodes. *BMC Genomics* **11**, 434.

Gonzalez M. A., Baraloto C., Engel J., Mori S. A., Pétronelli P. *et al.* 2009 Identification of Amazonian trees with DNA barcodes. *PLoS One* **4**, e7483.

Gu J., Su J.-X., Lin R.-Z., Li R.-Q. and Xiao P.-G. 2011 Testing four proposed barcoding markers for the identification of species within *Ligustrum* L. (Oleaceae). *J. Syst. Evol.* **49**, 213–224.

Hajibabaei M. and McKenna C. 2012 DNA mini-barcodes. In *DNA barcodes: methods and protocols, methods in molecular biology* (ed. W. J. Kress and D. L. Erickson) vol. 858, pp. 339–353. University of Hertfordshire, Hertfordshire, UK.

Hajibabaei M., deWaard J. R. and Ivanova N. V. *et al.* 2005 Critical factors for assembling a high volume of DNA barcodes. *Philos. Trans. R. Soc. London* [*B. Biol. Sci.*] **360**, 1959–1967.

Hajibabaei M., Smith M. A., Janzen D. H., Rodriguez J. J., Whitfield J. B. and Hebert P. D. N. 2006 A minimalist barcode can identify a specimen whose DNA is degraded. *Mol. Ecol. Notes* **6**, 959–964.

Hajibabaei M., Shokralla S., Zhou X., Singer G. A. C. and Baird D. J. 2011 Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river Benthos. *PLoS One* **6**, e17497.

Hebert P. D. N., Cywinska A., Ball S. L. and deWaard J. R. 2003 Biological identifications through DNA barcodes. *Proc. R. Soc. Lond. B. Biol. Sci.* **270**, 313–321.

Houdt J. K. J., Breman F. C., Virgilio M. and Meyer M. D. 2009 Recovering full DNA barcodes from natural history collections of Tephritid fruit flies (Tephritidae, Diptera) using mini-barcodes. *Mol. Ecol. Resour.* **10**, 459–465.

Kress W. J. and Erickson D. L. 2007 A two-locus global DNA barcode for land plants: the coding *rbcL* gene complements the non-coding *trnH-psbA* spacer region. *PLoS One* **2**, e508.

Kress W. J., Erickson D. L., Jones F. A., Swenson N. G., Perez R., Bermingham E. and Sanjur O. 2009 Plant DNA barcodes and a community phylogeny of a tropical forest dynamics plot in Panama. *Proc. Natl. Acad. Sci. USA* **106**, 18621–18626.

Kress W. J., Erickson D. L., Swenson N. G., Thompson J., Uriarte M. and Zimmerman J. K. 2010 Advances in the use of DNA barcodes to build a community phylogeny for tropical trees in a Puerto Rican forest dynamics plot. *PLoS One* **5**, e15409.

Lee P. L. M. and Prysjones R. P. 2008 Extracting DNA from museum bird eggs, and whole genome amplification of archive DNA. *Mol. Ecol. Resour.* **8**, 551–560.

Meusnier I., Singer G. A. C., Landry J. F., Hickey D. A., Hebert P. D. N. and Hajibabaei M. 2008 A universal DNA mini-barcode for biodiversity analysis. *BMC Genomics* **9**, 214.

Patel S., Waugh J., Millar C. D. and Lambert D. M. 2009 Conserved primers for DNA barcoding historical and modern samples from New Zealand and Antarctic birds. *Mol. Ecol. Resour.* **10**, 431–438.

Rougerie R., Smith A. M. and Fernandez-Triana J. 2010 Molecular analysis of parasitoid linkages (MAPL): gut contents of adult parasitoid wasps reveal larval hosts. *Mol. Ecol.* **20**, 179–186.

Saitoh K., Uehara S. and Tega T. 2008 Genetic identification of fish eggs collected in Sendai Bay and off Johban, Japan. *Icthyol. Res.* **56**, 200–203.

Shaw J., Lickey E. B., Schilling E. E. and Small R. L. 2007 Comparison of whole chloroplast genome sequences to choose noncoding regions for phylogenetic studies in angiosperms: the tortoise and the hare III. *Am. J. Bot.* **94**, 275–288.

Smith M. A. and Fisher B. L. 2009 Invasions, DNA barcodes and rapid biodiversity assessment using ants of Mauritius. *Front. Zool.* **6**, 31.

Sonstebo J. H., Gielly K., Brysting A. K., Elven R., Edwards M., Haile J. *et al.* 2010 Using next-generation sequencing for molecular reconstruction of past arctic vegetation and climate. *Mol. Ecol. Resour.* **10**, 1009–1018.

Srirama R., Senthilkumar U., Sreejayan N., Ravikanth G., Gurumurthy B. R., Shivanna M. B. *et al.* 2010 Assessing species admixtures in raw drug trade of *Phyllanthus*, a hepatoprotective plant using molecular tools. *J. Ethnopharmacol.* **130**, 208–215.

Tamura K., Peterson D., Peterson N., Stecher G., Nei M. and Kumar S. 2011 MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Mol. Biol. Evol.* **28**, 2731–2739.

Whitfield J. B. 1999 Destructive sampling and information management in molecular systematic research: an entomological perspective. In *Managing the modern herbarium: an interdisciplinary approach* (ed. S. Byers and D. Metsger), pp. 384. Society for Preservation of Natural History Collections and Royal Ontario Museum, Toronto, Canada.

Zimmermann J., Hajibabaei M., Blackburn D. C., Hanken J., Cantin E., Posfai J. *et al.* 2008 DNA damage in preserved specimens and tissue samples: a molecular assessment. *Front. Zool.* **5**, 18.

Received 15 June 2013, in final revised form 5 April 2014; accepted 3 June 2014
Unedited version published online: 26 June 2014
Final version published online: 5 November 2014