DNA barcoding: An exercise in futility or utility?

Taxonomy, the science of naming and classifying organisms, is the foundation of all biology. Unfortunately, over the past few decades, taxonomy is being completely overshadowed by seemingly spectacular and glamorous branches of biology. However, in the last few years some of these subjects such as molecular biology have rejuvenated taxonomy as a fashionable science once again. The advent of powerful DNA-based marker systems for identifying species has brought back the charm of 18th century biology and promises a less painstaking path for identification and discovery of new species on the 'PCR-desk' by 'lay-taxonomists' who need not necessarily be 'naturalists'. DNA barcoding, a tool that obtains species-specific DNA signature, is based on the simple premise that sequence diversity within small stretches of the organism's genome can provide a 'biological barcode' to enable identification of any organism at the species level[1,2]. As a result, scientists are hoping that DNA barcoding will provide a 'universal key' that will allow identification of a species by running unknown DNA sequences through a DNA barcode database.

The use of DNA barcoding is relatively new (with the first publication appearing in 2003)[1,3,4]. Yet in this short span many investigators have reported robustness of this technique through a flurry of papers. It is reported that with the use of a short sequence (~600 bp) of the cytochrome oxidase-I (*COI*) gene, the success rate of identification down to species level can be a remarkable – 98–100% in many organisms, including birds[5], fishes[6] and butterflies[7]. Using this 'barcode', several cryptic species are possible to be described within what had previously been thought to be a single species by conventional taxonomy[8,9]. These exciting developments have fuelled speculation that species could be identified even without the expertise of conventional taxonomy[10]. In addition to providing rapid and accurate identification of species, this technology also promises to uncover the phylogenetic affiliations among different taxa.

This initial success in DNA barcoding led to the formation of the Consortium for the Barcode of Life (CBOL, http://barcoding.si.edu). The Consortium has an ambitious programme of developing DNA barcodes for all species on the planet, including those that are yet to be described. CBOL foresees many applications of this technique, from fundamental research on biodiversity to enforcement of food laws, quarantine and phytosanitary laws and protection of wildlife[11]. That the recent advances in sequencing technology have become even more rapid, accurate and inexpensive, means that the barcoding endeavour appears to be both plausible and worthwhile.

DNA barcoding, in principle, is a diagnostic technique that uses short DNA sequence(s) for identification of species. The basic rationale of using the short DNA sequence (universal molecular yardstick), is that it allows to discriminate among species of a taxon under the assumption that the sequences chosen have relatively lower 'within-taxon' variation than that 'between-taxa'. It involves extraction of genomic DNA from tissue samples collected from an individual organism and using it for targetted amplification of one or several regions (short-listed based on their information content) by polymerase chain reaction (PCR) and sequencing the amplified products. The resulting sequences are used as 'barcodes' for tagging the species.

The use of the technique can be traced to the work of Carl Woese, who first demonstrated the utility of rRNA genes in inferring phylogenetic relationships among microorganisms[12]. Currently this method is being proposed to be used for non-microbes too, with appropriate genes. For example, DNA barcodes based on a 5′, 648 bp fragment of the mitochondrial *COI* gene, developed for a skipper butterfly species, *Astraptes fulgerator*, could be used to differentiate a complex of at least ten related species[8]. The effectiveness of DNA barcodes based on *COI* was also demonstrated in distinguishing 260 species of North American birds[5]. More recently, *COI* was used[7] to discriminate 521 species of Lepidoptera with a resolution of 97.9%. While *COI* has generally been found and also accepted as the standard genomic region for barcoding animal species, for plants various regions, including the chloroplast *rbcL* region, nuclear *ITS* (Internal Transcribed Spacer) region of rRNA genes and plastid nuclear intergenic spacer *trnH–psbA* have been recommended as possible candidate segments[13–15].

DNA barcoding has invited several criticisms. These range from skepticism on the technique as an effective taxonomic tool, to a more moderate view that it could only complement the existing approaches of conventional taxonomy. Mallet and Willmott[16] have argued that DNA barcodes based on a few specific genes may fail to distinguish closely related species because of the persistence of ancestral polymorphism. It is also feared that DNA barcoding exercises may supplant genuine taxonomic projects and merely end up in spewing-out alternate sets of data, without adding meaningful information on the taxa[17]. Some argue that DNA barcoding is gross oversimplification of the science of taxonomy and that DNA barcodes are no substitutes to detailed understanding (morphological, physiological and behavioural attributes) of taxa, practised in conventional taxonomy. Insisting that it as a mandatory step in taxonomic studies has been opposed on the grounds that this may hinder the already slow process of describing new taxa.

In summary, while the above criticisms may be valid among themselves, it has been grudgingly accepted that barcodes or barcode-like techniques might after all be useful in specific circumstances and need not be viewed as a panacea for all taxonomic hurdles. The evidence so far however weighs in favour of DNA barcoding as a taxonomic tool and the strength of the technique to reveal cryptic species should make it a valuable tool wherever conventional taxonomy is found wanting[18]. It could also be useful in analysing museum specimens that cannot be easily subjected to conventional taxonomic treatments and taxonomy of groups for which regular expertise is not available.

Against this background, it is obviously important to ask: should India commit itself to the barcoding of its organisms and if yes, to what purpose? Responses to both the questions could be divided, because the concept of barcoding has sown distrust between taxonomists and molecular systematists. In the Indian context, the barcoding exercise should

perhaps be weighed from at least two important angles, namely (a) meeting the taxonomic challenges and providing a robust identification of species and (b) securing intellectual property rights (IPRs) for some of the country's important bioresources.

Survey and description of Indian biota began as a strong programme almost 200 years ago. Despite this, we have not yet completely described even a single group of taxa, barring perhaps birds and butterflies; even mammals are being newly described. Further among those described, perhaps every group is burdened with taxonomic controversies of diverse kinds. Obviously these challenges range from subtle to serious, and require intervention that can successfully resolve the conflicts of classification. Besides the existing species lists, discovery of newer species especially from lower taxa, continuously demands expert taxonomic treatment. For example, in the recent past three new species of frogs, *Philautus nerostagona* from Wayanad District (Kerala) and *Philautus anili* sp. nov. and *Philautus dubois* sp. nov. from Wayanad and Kodaikanal (Tamil Nadu) respectively[19,20], were reported. Further, in a rather spectacular discovery, a new family of frog, Nasikabatrachidae from the Western Ghats[21] and a primate species *Macaca munzala* from the forests of Arunachal Pradesh[22] were reported from the country. In fact, Aravind *et al.*[23] showed that for a number of taxa, species discoveries are yet to attain an asymptote – a suggestion that there might be many more species waiting to be discovered. With three of the megadiversity hotspots in the country, it is not unlikely that we will have more discoveries than we can possibly address. Clearly, it would be to the advantage of taxonomy to complement existing taxonomic tools with DNA barcoding and aim at developing a robust identification scheme for specific groups[24,25]. Such a process can provide useful insights into the subtleties of an already identified taxonomic group (for example, see Hebert *et al.*[8]). Further, DNA barcoding information can also help in developing newer hypotheses relating to the taxonomic position of species and thereby even resolve taxonomic quagmires.

There have been several independent and concerted efforts to use DNA barcoding in taxonomic studies in India. Currently only two institutions from India, namely the Vector Control Research Centre, Pondicherry and St. Peter's College, Mumbai have been listed as members of CBOL (http://barcoding.si.edu/). We have recently initiated efforts to barcode Indian butterflies in our laboratory, especially those that might be economically important, taxonomically intriguing and endemic to the country. Specifically, we have attempted to explore how the barcode of a select group of butterflies of peninsular India places itself against the global barcodes available (Figure 1). Further, we have also shown that a group of butterflies that are otherwise morphologically similar, are clearly differentiated using the barcode (Figure 2).

In the Indian context, barcoding could be useful in contributing to what might be called as 'remote taxonomy'. Lack of type specimens and also access to large collections often held elsewhere in international museums has been among the most important constraints in practising taxonomy in India. Of late, this constraint is partly being overcome by the

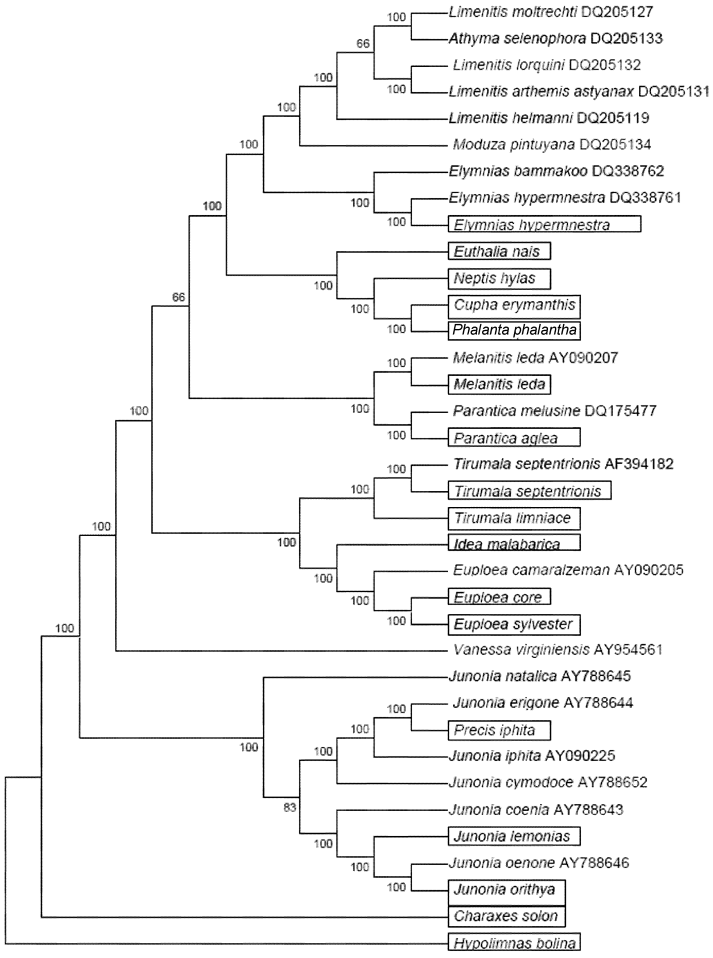

Figure 1. Consensus maximum parsimony tree of Indian and global Nymphalids. Species in box are from India. Numbers attached to species name indicate GenBank IDs. *Idea malabarica*, *Cupha erymanthis* and *Parantica aglea* were collected from Sringeri, Chikamagalur District, Karnataka, while the rest were collected from Botanical Garden, University of Agricultural Sciences, Bangalore. DNA was extracted from thoracic region of each individual using the method of Vandewoestijne and Baguette[32]. The extracted DNA was used for amplification of a 648 bp fragment of *COI* sequence using the primer pair: LepF (5′-ATTCAACCAATCATAAAGATATTGG-3′) and LepR (5′-TAAACTTCTGGATGTCCAAAAAATCA-3′)[8]. The amplified product was purified using Eppendorf gel cleanup kit (Perfectprep® Gel Cleanup) and sequenced. Sequences were edited and aligned using CLUSTALW software. Kimura's two-parameter model[33] of base substitution was used to calculate genetic distances; consensus maximum parsimony trees were obtained using MEGA 3.1 software. DNA barcodes were able to clearly distinguish the various species. Butterfly species from different geographical areas (whose sequences were obtained from GenBank) clustered with their respective members (species/genus) from India, indicating that the DNA barcodes could be used to develop species IDs.

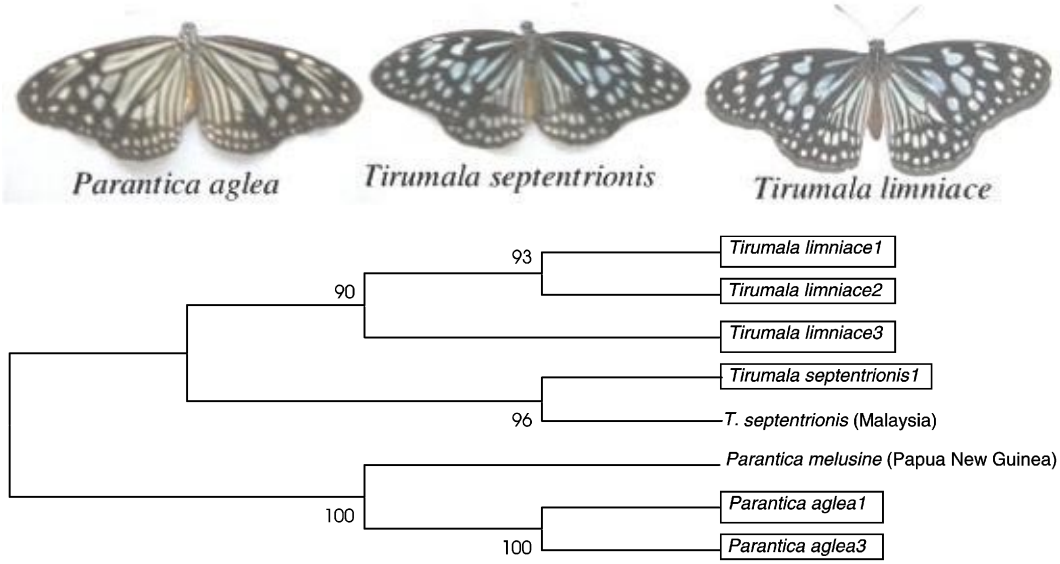

Figure 2. Consensus maximum parsimony tree of blue tigers. Species in box are from India. Numbers attached to species name indicate the individual IDs. Individuals belonging to each species were collected from Botanical Garden, University of Agricultural Sciences, Bangalore; *Parantica aglea* was collected from Sringeri, Chikamagalur District, Karnataka. DNA extraction, amplification, sequencing and analysis were done using the protocol mentioned in Figure 1. DNA barcodes were able to distinguish morphologically similar blue tigers, which is evident from the distinct cohesive clustering of individuals belonging to each species. As seen from the figure there was a clear differentiation of the blue tigers by the DNA barcode signatures, despite their apparent morphological similarities. Further, analysis based on the sequence data from GenBank (www.ncbi.nlm.nih.gov) for *T. septentrionis* (AF394182) from Malaysia and *Parantica melusine* (DQ175477) from Papua New Guinea, showed distinct clustering of the latter species with their respective species/genus from India. This result indicates that DNA barcodes across regions can be comparable to generate a larger taxonomic picture.

introduction of cyber-taxonomy or digital-taxonomy[26,27]. 'Barcoding taxonomy' can further complement the efforts of cyber-taxonomy for specific groups, where DNA technique-based interventions could further resolve the group identities.

DNA barcoding information could help in providing a correct species identification tool, especially of those in which biologically important properties or molecules with IPR potential have been identified. For example, the *Phyllanthus* (Euphorbiaceae) group of species is well known for its multifarious medicinal use ranging from antifungal to antiviral. In recent years, the group has drawn global attention due to its anti-hepatitis property and the associated IPR claims[28,29]. In India, it constitutes one of the most important component in raw drug trade, wherein dry plant material is sold for export and domestic use. Unfortunately because the *Phyllanthus* group has been historically plagued by taxonomic controversies[30,31] and also because of the relative difficulty of identifying the species, it is common to find mixtures of species in the raw drug market. The consequences of such mixtures can range from diluting the efficacy of the drug that is eventually going to be extracted from the herbal mix, to lowering the value of trade. It is estimated that about 33 herbaceous species (depending on the geographical area of collection) can, in theory, be in the mix of dry plant material. At our own laboratory, DNA analysis of raw materials collected from the Bangalore market indicated at least four different species of *Phyllanthus* (Deepali *et al.*, unpublished). Keeping in view India's role in the raw drug trade globally, it is imperative that for such specific and complex groups, DNA barcodes are developed which can then provide unambiguous identification of species. This information would be useful in not only providing diagnostics for rapid and easy identification of species in mixtures in the raw drug trade, but also in drawing specific regulations to protect the national markets.

Besides the above-mentioned example, barcoding could potentially be useful in also identifying species in other groups with high potential of staking IPR claims, such as medicinal leeches (for their anticoagulant property) or parasitoid wasps such as Trichogrammatidae (for their biocontrol uses) or orchids that might have immense commercial value.

In conclusion, while the relative costs and benefits of DNA barcoding technique continue to be debated, it is clear that the tool can be effectively used in complementing conventional taxonomic studies and in securing IPRs for important taxa. Viewed from this context, it will be important for the country to develop skills and infrastructure to undertake barcoding of at least some of the important taxa, both for conservation and commerce.

1. Hebert, P. D. N., Cywinska, A., Ball, S. L. and deWaard, J. R., *Proc. R. Soc. London, Ser. B*, 2003, **270**, 313–321.
2. Marshall, E., *Science*, 2005, **307**, 1037.
3. Hebert, P. D. N., Ratnasingham, S. and deWaard, J. R., *Proc. R. Soc. London, Ser. B* (*Suppl. 1*), 2003, **270**, S96–S99.
4. Blaxter, M., *Nature*, 2004, **421**, 122–124.
5. Hebert, P. D. N., Stoeckle, M. Y., Zemlak, T. S. and Francis, C. M., *PLoS Biol.*, 2004, **2**, e312.
6. Ward, R. D., Zemlak, T. S., Innes, B. H., Last, P. R. and Hebert, P. D. N., *Philos. Trans. R. Soc. London, Ser. B*, 2005, **360**, 1847–1857.

7. Hajibabaei, M., Janzen, D. H., Burns, J. M., Hallwachs, W. and Hebert, P. D. N., *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 968–971.
8. Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H. and Hallwachs, W., *Proc. Natl. Acad. Sci. USA*, 2004, **101**, 14812–14817.
9. Smith, M. A., Woodley, N. E., Janzen, D. H., Hallwachs, W. and Hebert, P. D. N., *Proc. Natl. Acad. Sci. USA*, 2006, **103**, 3657–3662.
10. Dasmahapatra, K. K. and Mallet, J., *Heredity*, 2006, **97**, 254–255.
11. Pennisi, E., *Science*, 2003, **300**, 1692.
12. Woese, C. R. and Fox, G. E., *Proc. Natl. Acad. Sci. USA*, 1977, **74**, 5088–5090.
13. Chase, M. W., Salamin, N., Wilkinson, M., Dunwell, J. M., Kesanakurthi, R. P., Haidar, N. and Savolainen, V., *Philos. Trans. R. Soc. London, Ser. B*, 2005, **360**, 1889–1895.
14. Kress, W. J., Wurdack, K. J., Zimmer, E. A., Weigt, L. A. and Janzen, D. H., *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 8369–8374.
15. Newmaster, S. G., Fazekas, A. J. and Ragupathy, S., *Can. J. Bot.*, 2006, **84**, 335–341.
16. Mallet, J. and Willmott, K., *TREE*, 2003, **18**, 57–59.
17. Ebach, M. C. and Holdrege, C., *BioScience*, 2005, **55**, 822–823.
18. Greenstone, M. H., *Bull. Entomol. Res.*, 2006, **96**, 1–13.
19. Biju, S. D. and Bossuyt, F., *Curr. Sci.*, 2005, **88**, 175–178.
20. Biju, S. D. and Bossuyt, F., *Amphibia–Reptilia*, 2006, **27**, 1–9.
21. Biju, S. D. and Bossuyt, F., *Nature*, 2003, **425**, 711–714.
22. Sinha, A., Datta, A., Madhusudan, M. and Mishra, C., *Int. J. Primatol.*, 2005, **26**, 977–989.
23. Aravind, N. A., Manjunath, J., Dinesh Rao, Ganeshaiah, K. N., Uma Shaanker, R. and Vanaraj, G., *Curr. Sci.*, 2005, **88**, 258–265.
24. Rubinoff, D., *Conserv. Biol.*, 2006, **20**, 1548–1549.
25. DeSalle, R., *Conserv. Biol.*, 2006, **20**, 1545–1547.
26. Godfray, H. C. J., *Nature*, 2002, **417**, 17–19.
27. Ganeshaiah, K. N., Rajanikanth, G., Mohan, G. S., Nanditha Mahadev and Uma Shaanker, R., Biodiversity Informatics 2004, Taxonomic Databases Working Group, Annual Meeting, University of Christchurch, New Zealand, 11–17 October 2004.
28. Nair, M. D., *Plant Genet. Res.: Char. Util.*, 2005, **3**, 314–319.
29. Venkateswaran, P. S., Millman, I. and Blumberg, B. S., *Proc. Natl. Acad. Sci. USA*, 1987, **84**, 274–278.
30. Ganeshaiah, K. N., Uma Shaanker, R., Ganeshan, R. and Meera, C., *Amruth*, 1998, **8**, 3–8.
31. Webster, G. L., *J. Arnold Arb.*, 1957, **38**, 295–373.
32. Vandewoestijne, S. and Baguette, M., *Heredity*, 2002, **89**, 439–445.
33. Kimura, M., *J. Mol. Evol.*, 1980, **16**, 111–120.

ACKNOWLEDGEMENTS. The work is supported by grants from the Department of Biotechnology (DBT), New Delhi to R.U.S., K.N.G., K.C. and A.R.V. The work on butterflies reported in the paper is part of the DBT-sponsored Butterfly Park Project.

Received 4 September 2006; revised accepted 26 December 2006

K. ARAVIND[1]
G. RAVIKANTH[1,5]
R. UMA SHAANKER[1,2,5,6,*]
K. CHANDRASHEKARA[1,4]
A. R. V. KUMAR[1,4]
K. N. GANESHAIAH[1,3,5,6,7]

[1]*School of Ecology and Conservation,*
[2]*Department of Crop Physiology,*
[3]*Department of Genetics and Plant Breeding,*
[4]*Department of Entomology,*
University of Agricultural Sciences, GKVK,
Bangalore 560 065, India
[5]*Ashoka Trust for Research in Ecology and the Environment,*
Hebbal,
Bangalore 560 024, India
[6]*Jawaharlal Nehru Center for Advanced Scientific Research,*
Jakkur, Bangalore 560 065, India
[7]*National Institute of Advanced Studies, IISc Campus,*
Bangalore 560 012, India
**For correspondence.*
e-mail: rus@vsnl.com