In pursuit of a universal barcode of plants: peril of followers?

G. Ravikanth, R. Srirama, K. N. Ganeshaiah and R. Uma Shaanker

Barcoding: a brief history

In May 2007, during the early days of the DNA bar coding project in India, we published an article, 'DNA barcoding: an exercise in futility or utility'[1]. As the title reflects, we were literally at crossroads, caught between the cross-fire of traditional taxonomists (we think it is disrespectful to call them traditional; they are as much modern as are archaeologists and molecular biologists) and molecular systematists and not knowing which way to go forward. After a reasonable amount of brain-storming that took us through well-trodden criticisms of the DNA barcoding initiative, we concluded that while debates can go on, the tool itself can be effectively used in complementing conventional taxonomic studies and in securing Intellectual Property Rights (IPRs) for important taxa. We also felt at that time that it would 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.

However, the article[1] in 2007 did not discuss the fidelity of the barcode. The 648 bp fragment of the mitochondrial *COI* genome was shown to effectively discriminate North American birds, butterflies, moths, fishes, etc. For plants, in contrast, several regions, including the chloroplast *rbcL* region, nuclear *ITS* (internal transcribed region) region of rRNA gene and plastid nuclear intergenic spacer *trnH-psbA* were recommended as possible candidates. However, no sooner had we rested on the laurels of these barcodes and their potential uses in barcoding the Indian taxa, we are witness to increasing instances of dissatisfaction, helplessness and even frustration.

What was thought to be the Holy Grail of barcoding (the *COI* region) was becoming increasingly vulnerable as more and more scientists were beginning to realize its limitations. For example, in a few systems such as, benthic Cnidarians, two groups of amphibians and some Gastropod species, there were already some signals of breakdown of the barcode. Mitochondrial DNA of Cnidarians has been shown to possess slower mutation rates compared to other taxa and this could impair the resolving power of *COI* in this group. Conversely, high intraspecific variation at *COI* up to 18% was observed in two amphibian groups (mantellid frogs and salamanders), which are shown to be overlapping with interspecific variation, making species discrimination difficult. Studies on the Gastropods, particularly the two subclasses, Heterobranchia and Patellogastropoda, have reported complicated alignment because of insertions and deletions[2–4].

If the animal barcoders had started squirming thanks to the Cnidarians, Amphibians and Gastropods, the plant barcoders were left out in the cold owing to an utter lack of clarity in the universal barcode for plants. After *COI* was announced as the universal barcode for animal taxa, plant scientists were looking out for a similar universal barcode for plants. Unfortunately, *COI* and other mitochondrial genes were not suitable for plants because of their low mutation rates and the rapidly changing structure of its genome[5–7]. By simple logical extension, the search for a region analogous to animal *COI* focused on chloroplast DNA[5,7].

A number of genome regions in the chloroplast were suggested to be potential candidates for barcoding plants (coding regions *accD*, *matK*, *ndhJ*, *rpoB2*, *rpoC1* and *ycf5*[8,9]; coding region *rbcL*[10]; *trnL* intron[11]; Universal Plastid Amplicon (UPA), *rpoB*, *rpoC1*[12] and non-coding spacer *trnH-psbA*[7,10]). However, the initial enthusiasm on most of these chloroplast regions seems to be waning[13]. In a recent effort to arrive at a possible universal plant barcode, nine different regions which include portions of five coding plastid regions (*rbcL*, *matK*, *rpoC1*, *rpoB* and *23S rDNA*), three non-coding plastid intergenic spacer regions (*trnH-psbA*, *atpF–atpH* and *psbK–psbI*) and the mitochondrial *COI* gene (the animal barcode region) were compared to discriminate species but with limited success[14]. In short, although a number of chloroplast regions have been proposed, there was no consensus[15,16].

In a desperate attempt to arrive at an acceptable bar code, besides the chloroplastic region, several other genes and regions were also proposed. The non-coding region *ITS*, located in the nuclear genome of plants, was shown to be promising as a plant DNA barcode[7,17]. It was argued that *nrITS* can be used as a 'local barcode' for those groups where there is a low level of plastid DNA variation[18,19]. However, *nrITS* has been detected to have multiple copies which are functional in some groups of angiosperms[20].

What is the jinx in locating the candidate loci? A number of problems. Let us illustrate this using only one of the regions, *psbA-trnH* that was proposed by Kress and Erickson[10]. *psbA-trnH* is one of the variable non-coding regions of the plastid genome in angiosperms and thus promised to return a high level of species discrimination[21]. However, careful examination of this region revealed a bagful of worries. Owing to the high rates of insertions and deletions, there are enormous problems associated with alignment of the sequences.

There is a substantive length variation even among closely related taxa; Kress and Erickson[10] reported a length variation between 119 and >1000 bp in the angiosperm taxa. This spacer is exceedingly short in some groups of plants unlike orchids where this region is much longer as it contains copies of *rpl22* and *rps19* genes[7,8]. In certain groups of plants, such as *Crocus* and *Hordeum*, where both *matK* and *rpoC1* contain more variable positions, the region *psbA-trnH* is not rapidly evolving[22]. Another study on *Scalesia* (Asteraceae) using *psbA-trnH* and *nrITS* regions revealed no variation[22].

Finally and more recently, perhaps exasperated by the successive failures, there has been a suggestion to use a combination of two or more gene regions as promising candidates[7,23,24]. Examples of such combinations include three coding genes *matK*, *rpoB* and *rpoC1*[7] and the spacer region *trnH-psbA* with the coding gene *rbcL*[10,12]. The Consortium of Barcode of Life (CBOL) initiative and the Kew group have proposed six chloroplast regions namely *matK*, *rpoC1*, *rpoB*, *accD*, *YCF5* and *ndhJ* as probable candidates singly or in combination as univer-

sal plant barcode[17]. More recently, the CBOL has proposed *rbcL* + *matK* as the standard plant barcode for land plants after the scrutiny of different chloroplast regions such as *atpF–atpH* spacer, *matK* gene, *rbcL* gene, *rpoB* gene, *rpoC1* gene, *psbK–psbI* spacer and *trnH–psbA* spacer based on four major criteria such as universality, sequence quality, coverage and discrimination[13].

What should India do?

This question is best answered at several levels. For leaders in the field, especially for those dealing with plant systems, the road ahead is not smooth. They would still have to grapple with the idiosyncrasies of plant taxa and look for a barcode that best fits all the taxa. This is clearly an uphill task, beyond the reach of small national groups with small operating budgets and access to limited national or regional taxa. Although the leaders may want to trudge this path in the hope of discovering the elusive universal barcode, at this level, they might benefit from spending a little time, debating and discussing the philosophy of a universal barcode. To what extent is our belief in the existence of a universal barcode backed up by the realities of biology and by the compulsions of a 'pet-theory'. After the Hebert school sowed the seeds of a universal barcode[3], as expected there has been a flight of excitement in generating data from all corners of the world. However, in less than 6 years, the initial excitement has lead to despair (Figure 1).
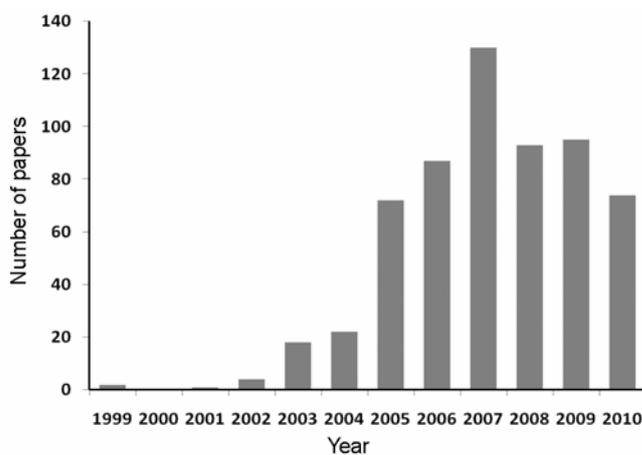
Is something seriously wrong with the theory of barcoding? Has our initial enthusiasm and excitement blinded us to noise that is flying all around? Are we in the same boat as the alchemists in the futile search for the 'elixir of life'? Are we wishfully (much against the reality) trying to confirm the universality of barcodes? Here, we have to admit that the 'DNA barcode' has become a very neat and attractive 'marketing brand' and no one would want to put it to dust, least of all its proponents. Having invested heavily on both resources and intellect, it seems to be caught up in a Concorde fallacy.

In our own country, the DNA barcoding programme at least with respect to plants seems to be caught in this quagmire[25]. What is the strategy to be followed in the light of the uncertainty regarding the universal barcode for plants? Several comments and suggestions are in place. First and foremost, we should ask ourselves 'why are we barcoding?' We might be doing this as a member-state of CBOL and therefore to contribute the DNA barcode of our taxa to the global dataset. Second, we might wish to have a DNA barcode of the country's taxa, so as to have unique identifier of our taxa that might find potential application in academic studies, conservation, commerce, etc.[26]. On the basis of which of the above two questions we align with, we can relate our future strategies and therefore proceed from there. If we are barcoding our taxa for the first of the two reasons, then clearly we have to await instructions

from the international community of scientists; else the project will have no global application and meaning. However, if our option is to have a unique identifier of our taxa, we can downscale the rigour and carry on with barcoding knowing well that we might not be working with a universal barcode. Here, the imperatives are to work with a barcode or a proxy that might serve our taxa well as a unique identifier and be applicable for a variety of uses. However, beset in this suggestion is the *déjà vu* – how then is this exercise different from the classical molecular systematics/phylogenetics approach?

Barcode creation versus creator's barcode

Barcodes (*not DNA barcodes*) were invented for uniquely identifying man-made products, such as those in supermarkets and bookshops. These barcodes are characterized by their simplicity, universality, low cost and above all unambiguity. All this was possible because the creation of the unique identity lies within the powers of the inventor. Biologists who adopted this concept, however dealt with one major and vital difference – that is, instead of assigning an externally developed barcode, they have tried to identify the barcode from within the system (in some sense, we might refer this as 'reverse engineering of the barcode'). Such an exercise, with all good intentions, is prone to failure. Consider a hypothetical situation where we have been asked to barcode a few thousand books in Higginbothams – not by externally assigning a barcode to them, but by eliciting a barcode from each of the books based on some previously decided 'universal barcode'. By any stretch of imagination, one can easily conclude that this exercise is fraught with improbabilities. Are we entering such a scenario with respect to DNA barcoding? Leaders in this field need to sit and ponder this seriously.

Where do we go from here? For followers in the field, the road ahead is slippery. As in any branch of science, followers normally work in 'straightjackets' and neither have the time nor the resources to ask brave questions. That in itself is not damaging, as their work would have contributed to strengthening an existing paradigm. However, it does

Figure 1. Number of papers published on DNA barcoding of plants and animals between 1999 and 2010 ($n = 598$). There seems to be a perceptible decline following the initial excitement. The figure is based on data obtained from http://ibol.org/resources/scientific-publications/#older-publications.

become damaging when the basic super-structure of the work itself (as it now appears in the DNA barcoding exercise) is itself slippery. This therefore seems to be the plight in which smaller labs and groups trying to emulate the DNA barcoding programme find themselves in. Therein lies the peril of followers.

1. Aravind, K., Ravikanth, G., Uma Shaanker, R., Chandrashekara, K., Kumar, A. R. V. and Ganeshaiah, K. N., *Curr. Sci.*, 2007, **92**(9), 1213–1216.
2. Remigio, E. A. and Hebert, P. D. N., *Mol. Phylogenet. Evol.*, 2003, **29**, 641–647.
3. Hebert, P. D. N., Ratnasingham, S. and deWaard, J. R., *Proc. R. Soc. London B*, 2003, **270**, S596–S599.
4. Vences, M., Thomas, M., Bonett, R. M. and Vieites, D. R., *Philos. T. R. Soc. B*, 2005, **360**(1462), 1859–1868.
5. Adams, K. L. and Palmer, J. D., *Mol. Phylogenet. Evol.*, 2003, **29**, 380–395.
6. Cho, Y., Mower, J. P., Qiu, Y. L. and Palmer, J. D., *Proc. Natl. Acad. Sci. USA.*, 2004, **101**, 17741–17746.
7. Kress, J. W., Wurdack, J. K., Zimmer, A. E., Weigt, A. L. and Janzen, H. D., *Proc. Natl. Acad. Sci. USA*, 2005, **102**, 8369–8374.
8. Chase, M. W. *et al.*, *Taxon*, 2007, **56**(2), 295–299.
9. Lahaye, R. *et al.*, *Proc. Natl. Acad. Sci. USA*, 2008, **105**(8), 2923–2928.
10. Kress, J. W. and Erickson, D. L., *PLOS ONE*, 2007, **2**(6), e508.
11. Taberlet, P. *et al.*, *Nucleic Acids Res.*, 2007, **35**(3), e14.
12. Newmaster, S. G., Fazekas, A. J., Steeves, A. D. and Janovec, J., *Mol. Ecol. Notes*, 2008, **8**, 480–490.
13. CBOL, *Proc. Natl. Acad. Sci. USA*, 2009, **106**(31), 12,794–12,797.
14. Fazekas, A. J. *et al.*, *Mol. Ecol. Resour.*, 2009, **9**, 130–139.
15. Pennisi, E., *Science*, 2007, **318**, 190–191.
16. Ledford, H., *Nature*, 2008, **451**, 616.
17. Sass, C., Little, D. P., Stevenson, D. W. and Specht, C. D., *PLOS ONE*, 2007, **2**(11), e1154.
18. Cowan, R. S., Chase, M. W., Kress, J. and Savolainen, V., *Taxon*, 2006, **55**(3), 611–616.
19. Chen, S. *et al.*, *PLOS ONE*, 2010, **5**(1), e8613.
20. Rapini, A., Chase, M. W. and Konno, T. U. P., *Taxon*, 2006, **55**, 119–124.
21. Shaw, J., Lickey, E. B., Schilling, E. E. and Small, R. L., *Am. J. Bot.*, 2007, **94**, 275–288.
22. Seberg, O. and Petersen, G., *PLOS ONE*, **4**(2), e4598.
23. Rubinoff, D., Cameron, S. and Kipling, W., *Trends Ecol. Evol.*, 2006, **21**, 1–2.
24. Vijayan, K. and Tsou, C. H., *Curr. Sci.*, 2010, **99**(11), 1530–1541.
25. Roy, S. *et al.*, *PLOS ONE*, 2010, **5**(10), e13674.
26. Srirama, R. *et al.*, *J. Ethnopharmacol.*, 2010, **130**, 208–215.

The authors are in the School of Ecology and Conservation, University of Agricultural Sciences, GKVK, Bangalore 560 065, India; G. Ravikanth, K. N. Ganeshaiah and Uma Shaanker are in the Ashoka Trust for Research in Ecology and the Environment, Royal Enclave, Srirampura, Jakkur Post, Bangalore 560 064, India; K. N. Ganeshaiah is also in the Department of Forestry and Environmental Science, University of Agricultural Sciences, GKVK, Bangalore 560 065, India; R. Uma Shaanker is also in the Department of Crop Physiology, University of Agricultural Sciences, GKVK, Bangalore 560 065, India.
e-mail: umashaanker@gmail.com

Population rise and growing water scarcity in India – revised estimates and required initiatives

Sharad K. Jain

Preliminary results of the 2011 census released recently by the Government of India show that the current population is higher than the earlier projections. As the requirement of water chiefly depends upon population, earlier estimates have been revised in view of the revised population projections. Initiatives to overcome the impending water scarcity have also been suggested.

Demand for water depends on several factors such as population, income level or lifestyle and industrialization. Demand for water increases with the population as water is needed to sustain life, for sanitation, agriculture, generate energy, run industries, etc. As income rises, people tend to use more water. City-dwellers consume more water than those living in rural areas. The affluent section of the society consumes more water for cleaning, maintaining gardens/lawns, etc.

Preliminary results of the 2011 census released by the Government of India (http://censusindia.gov.in/) have estimated the current population of the country to be 1210 million. The population numbers indicate a rapid growth as the earlier estimates had projected the population to be around 1189 million by 2010 (ref. 1). Revised projections of the population are also available from different sources, including the United Nations (UN). The medium variant of the UN population projection shows that the population of India is expected to stabilize at a level of about 1718 million by 2065 against the earlier estimates which had projected

the population to stabilize at about 1580 million by 2050. Thus, the updated estimates show that the population will stabilize at a higher value and at a later date.

Table 1 presents the projections of population for some selected years according to the UN medium variant.

Water resources are one of the main components of the infrastructure sector which will be facing increasing stress on account of growing population because the demand for water for various uses largely depends on the population. Earlier, an exhaustive assessment of the