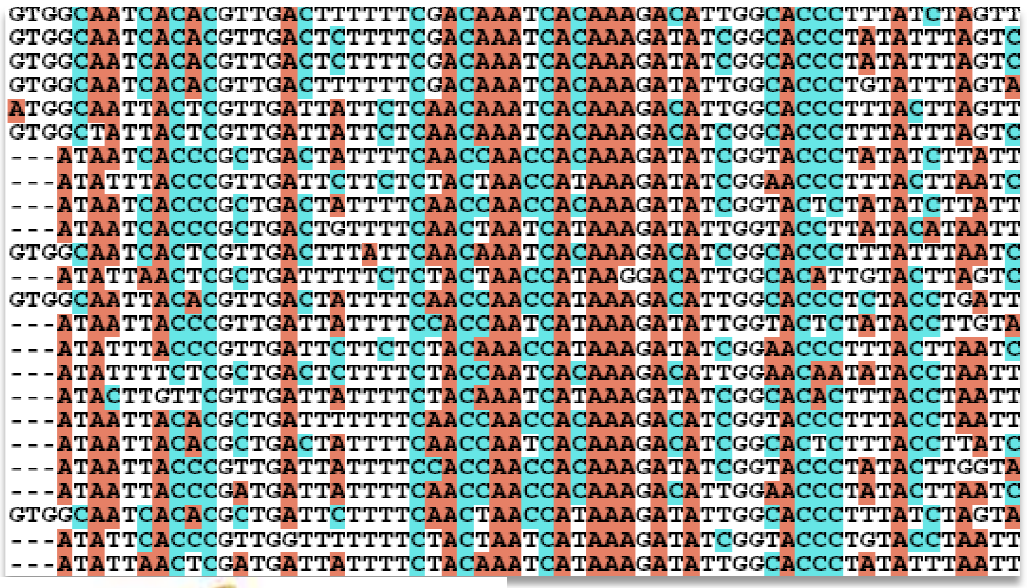
Barcoding anurans of India

Final Technical Report

2010



Principal Investigators

Karthikeyan Vasudevan

Wildlife Institute of India, Dehradun

Ramesh K. Aggarwal

Centre for Cellular and Molecular Biology, Hyderabad

S.K. Dutta

North Orissa University, Orissa

Supported by

Department of Biotechnology

Government of India

Chapter 1

DNA Barcoding

1.1 Introduction

It has been estimated that the earth has around 3.6 million to 100 million species, out of which only 1.5 to 1.8 million species are recorded (Wilson 2000). In the last 15 years the global amphibian number has increased from 4000 to 5400 and the total number of flowering plants (272, 000) may increase to 300, 000 as around 2000 species are added up each year (Wilson 2000). Considering this enormous diversity in the life's form the taxonomist's span of work becomes highly diversified and complicated. Screening all the species carefully for a variety of taxonomic characters poses a huge problem and may take another century to prepare a draft of all the extant species on earth (Blaxter 2003). A rough calculation arrived at a community of around 15000 taxonomists that needs to work continuously for many years to execute this massive exercise. How can this process be hastened? Bracoding of life is proposed by many to achieve this task. However, this initiative has some contentious issues to be resolved before taxonomy can embrace it. Some believe that this cannot become part of classical taxonomy dictated by delimitation of species by morphological characters (Will *et al.* 2003; Lipscomb *et al.* 2003). Some trimly believe that DNA barcoding is here to stay and it will invigorate classical taxonomy (Hebert *et. al.* 2003a, b;Tautz *et al.* 2003;A. Minelli, 2003; E.O.Wilson, 2003; Savolainen 2004; Schindel 2005; Erpenbech *et al.*2006). A review of literature on these view points is presented here to address the basic question, whether taxonomy should embrace barcoding at all?

***Descriptive Taxonomy*:** Advancement in taxonomy typically happens through a hypothesis testing approach where each species constitutes a hypothesis that can be falsified. Though rigorous, it is a tedious process and in many ways this slows down documentation of biological diversity on earth (Tautz *et al.* 2003).The morphological character based taxonomy has problem dealing with homology and homoplasy (Stearns and Hoekstra 2002). This approach overlooks cryptic taxa with poor morphological description (Knowlton 1993; Jarman and Eliot 2000; Hebert *et al.* 2006). The method targets a specific life history stage of the organism and cannot differentiate between all the stages of their life (Hebert *et al.* 2003 a). Literature on taxonomy is scarce, old and published in different languages. Accessing this information is expensive as well as sometimes inaccessible (Minelli 2003). Many of the voucher specimens are stored in the museums of wealthy nations and are inaccessible to taxonomists in "biodiversity rich" poor nations. Taxonomic keys for species identification when used by novices can lead to wrong diagnosis. Description of species is often a long drawn procedure where taxonomists work in isolation with large private collections. These collections are unavailable to other taxonomists and leads to multiple collections raising an ethical question (Daniels 2006). These problems of taxonomy are often flagged not to undermine the science, but it is an indicator of a healthy appetite for identification of species by contemporary fields such as conservation biology, forensic science, biotechnology and invasive species biology. Today, the end user of taxonomy is extremely diverse. If an effective service has to be provided, taxonomy has to create a unified "catalogue of life" that initially provides access to most, if not all the information available on already described species. The tediousness in the progress of descriptive taxonomy has proved its worth. Compromises on the rigor can undermine the objective of understanding the biological diversity on earth. The expectation from descriptive taxonomy somehow has to be packaged into the science to sustain the enormous amount of attention that taxonomy has been able to gain in recent years.

DNA taxonomy is now used in harmony but in addition to other classical morphological data to delimit species (Tautz *et al.* 2002). Although it has been well accepted that DNA taxonomy can solve many taxonomic problems but still it has not got a central role in it. Presently scientists are working on phylogeny and phylogeography of different species using the DNA as the central theme of their analysis. Although the morphological attributes are going to play the major role in the taxonomic description, DNA can be given a better position than what it has today. We believe the best way to give DNA its fair chance in taxonomy will be to implement "DNA barcoding" as an international unit for identification of species.

***Barcoding of Life*:** International units for the identification for the commercial products, having electronic barcodes, are used by the tradesman to get some information about the concerned product. This barcode acts as an id for an article and it varies for each individual item that we purchase. This is actually the "universal product code" method (Savolainen *et al.* 2005), which is known as a barcode in the retail business. Building upon this idea Paul Hebert, from University of Gulep, in Canada developed the use of part of the mitochondrial gene as a universal 'identification' marker for living organisms (Savolainen *et al.* 2005). The whole notion of product barcode is based on the arrangement of 10 alternate numbers in 11 positions to create 100 billion unique numbers which can then be used as an individual product id (Hebert *et al.* 2003a). Genomic DNA can be used in the same manner but the problem here is we only have 4 bases to work with. This problem is solved when we look at the enormous size of the available DNA in the animal and plant cells. Even the minutest microbes have enough amount of DNA which can be utilized for this purpose. It has been calculated by Hebert *et al.* (2003 a) that just taking 15 sites of nucleotide positions can create a possibility of 4^{15} codes, which is huge compared to the artificial barcode system. Some of the sequences are very highly conserved and other regions provide diversity to be checked for at least intraspecific levels.

Although these short sequences have been advocated by Hebert et al. (2003 a,b) in their classic paper, but still there is confusion and chaos among the existing taxonomist and those who are in the process of being trained into one. This conception of the DNA barcoding started budding in a conference on DNA Taxonomy Workshop in Staatsammlung in Munich in April 2002, funded by German Science association (DFG) (Svolainen 2003). Then the idea never looked back and many conferences and workshop has been organized hitherto. Recently Government of India has taken the initiative for barcoding this year and the Department of Biotechnology has invited proposals for barcoding some of the focal taxa. It is interesting to note that more than a dozen proposals were submitted to the steering committee for barcoding these focal taxa. The Botanical Survey of India since 1890 has inventoried about 60% and the Zoological Survey of India since 1916 has inventoried about 35% of India's geographical area (Pushpagandan and Nair 2001). This clearly indicates the enormity of task that is at hand.

In the year 2004 the Sloan Foundation has provided substantial amount for the establishment of a secretariat for the "Barcode of Life" program, based at the Smithsonian National Museum of Natural History in Washington, USA (Savolainen *et al.* 2005). The consortium for the barcode of Life (COBL) was created (www.barcoding.si.edu).There has been many debates on the concerned topic (www.conferences.uiuc.edu/peet/video.html) and there have been many publications which provide the pros and cons of this method (Mortitz *et al.* 2004; Marshall 2005). The main benefits from this huge program would undoubtedly include the following benefits. Would lower the burden of identification of the existing species for the taxonomists so that they can spend more time and energy on identifying new taxa and resolving

problems in the existing taxa. Pair up various life stages of the same species, as all the life phases of a species cannot be distinguished from morphological data (e.g. larval stages of a lepidopteran species or different stages of a amphibian species). Provide a tool for novices and the general public who are interested in the natural history of a particular place, as the classical taxonomic data are meant only for the experts in the field rather than other end users.

Barcodes are also advocated as a means of identifying new taxa (Hebert *et al.* 2003b; Hajibabaei et al. 2006) but there are doubts about its correctness (Will *et al.* 2004) .Recently it has been reported that even short sequences can be used for barcoding specimens whose DNA has been degraded to a certain extent due to preservation for a long time (Hajibabaei *et al.* 2006).Other probable advantages are dependent on the "Universal Barcoder", which is being developed for the purpose of identification. Although this is a notion and research is going on but once we have such handy tool we can use it for the purpose of the identification of specimens in the field itself i.e. it can serve as a very powerful field guide. Now even newer technologies, microarrays, are proposed to be amalgamated with the barcoding to give a pace to identification of organisms (Summerbell *et al.* 2005). With time and newer ideas it may also lead to some other unseen advantages. The expertise to identify and recognize a species remains with a very few people which goes with them when they retire (Tautz et al.2003) and it has been proposed that to make data more accessible and easy DNA barcode should be applied as the measuring stick of new taxonomic system along with morphological data. COI (Cytochrome Oxidase subunit1, also know as cox1) sequences has been tested as potential barcoders for many groups of animal and have found it to be a very effective tool for majority of them except the Cnidarians (Hebert et al. 2003 b). Some other authors have proposed for other sequences too such as control region of mitochondria, mitochondrial Cytochrome b and ND6 (NADH Dehydrogenase subunit 6) etc. for DNA barcoding in animals (Tautz et al. 2003). Some have advocated criteria on sequence similarity, assuming a cut-off value for maximum within-species variations (Lambert et al. 2005). Some times the COI may not be the most important tool for being the DNA barcode other sequences particularly the ITS region is very effective for the plant biologists (Kress et al. 2005) other sequences as the 16S and 18S rDNA has been taken into account for amphibians (Vences et al. 2005a) and insects (Pons et al. 2006), but so far the best sequences available for use in the barcoding of the animal diversity has been the COI (Hebert et al. 2003 a,b; Erpenbeck et al.). There are potential problem in the using COI as the only sequence in certain groups as amphibians where the universal primers cannot be used as the regions are highly variable (Vences et al. 2005 b). Further research is going on to find out more suitable regions of DNA which can be used for the purpose of barcoding. Although there has been a lot of hue and cry by many scientists (Will et al. 2003; Lipscomb et al. 2003), about the application of barcoding in taxonomy, this is going to influence taxonomy and the biodiversity conservation immensely.

It is in the above background that the present collaborative project was realized, with the main focus on the anuran diversity in India. In specific, the projected objectives of the project and the responsibilities of the collaborating institutes were:

- ✓ **To document anuran diversity in 'hotspots' and in biogeographically important areas in India (WII, NOU);**
- ✓ **To create an interactive digital library of photographs, calls and DNA barcode of known amphibian species in India (WII, NOU, CCMB).**
- ✓ **Check barcoding gaps and describe cryptic anuran species (WII, NOU, CCMB).**

- ✓ **Create a web-enabled database providing the above information resource on frog taxa of India with retrievable DNA based/other descriptors (WII).**

Accordingly, the respective work was undertaken at all the three centers, except for the 4th objective for which some efforts were initiated at CCMB but could not be completed due to various reasons (which were informed during the last work progress presentation at DBT, Delhi with a request for extension of the project/computer hardware; the request was not sanctioned by the Task Force). The other objectives were completed with varying success.

In brief, the work to document the anuran diversity across various biogeographically important areas like Eastern Ghats, Western Himalayas, North East, Rajasathan and Madhya Pradesh, were undertaken. More than 350 samples were collected from about 80 different localities for barcoding work (Figure 1). Representative samples from the collected specimen were used for DNA analysis/barcoding work, which involved:

- Primer designing
- Validation and optimization of designed primers
- Sample processing
- DNA barcode generation
- Data analysis

The very first need of any barcode effort is the need for universal primers, which can be used across wide range of species. Therefore Primer designing becomes a critical step and minor adjustments in the sequence can have large impact on barcode recovery. No such universal primers were available in the published literature for the anurans. Therefore, extensive exploratory work was undertaken to design/develop universal primers (if feasible) or group specific primers, which would help in exploring anuran diversity. For the purpose, efforts were made to identify Potential priming sites across a few mitochondrial as well as nuclear genes, viz., 12S/16S ribosomal RNA, CO1, CO2, CO3, Cytochrome-*b,* ND1, RAG1 and Rhodopsin, by comparative *in-silico* analysis of sequences available in the public domain. More than 90 new primer pairs were designed, which were then systematically used for standardization of PCR conditions followed by validatation studies for their universal applicability using a panel of 24 frog species representing the known anuran diversity from India. Following validation more than 37 primers were narrowed down to be suitable for Barcode recovery of Indian anurans, which were further used to develop DNA barcodes from about **200** samples collected under the project.

Figure 1: Location of *sampling* sites under the project.