Tags For Living Beings

THE SAGA OF DNA BARCODE

A. BIJU KUMAR & SANIL GEORGE

Very soon all life forms on Earth could have a tag for identification. Scientists round the world are working on the DNA Barcoding technology that could help distinguish the millions of species that inhabit the planet.

IN the early 1960s, World War II veteran Gene Roddenberry brought to air a now famous science fiction drama, *Star Trek*, in which a handheld 'tricorder' device was used to scan and identify alien life forms. Though this remains a fancy, today's researchers are working on a technology called DNA barcoding that envisages an inexpensive handheld device the size of a mobile phone, a DNA barcode scanner, that provides quick identification of every species in the world and that could transform people's relationship with nature!

Wouldn't it be wonderful to sequence the DNA tag of any organism in our planet from part of a leg of an insect or part of fin of a fish, or even a bit of a leaf or bark using a portable device, Google that name and find out its biological identity!

Classifying Life

Ever since inquisitive human beings started grouping, classifying and naming living beings around them, the amazing diversity of life and the subtle variations that distinguish each species of organism in the planet puzzled them a lot. Over the last 250 years, starting with the period of Carl Linnaeus, taxonomists have so far documented about 1.8 billion organisms on our planet Earth. However, moderate estimates show that the number of species may be around 10 million.

The paradoxical situation is that the trained manpower for identifying biological diversity, the taxonomists, are reducing in number consistently over the last few decades, resulting in what is referred to as "taxonomic impediment". Conversely, the modern world is witnessing the "sixth extinction", that is, extinction of biodiversity at a faster pace than that happened in the geological past.

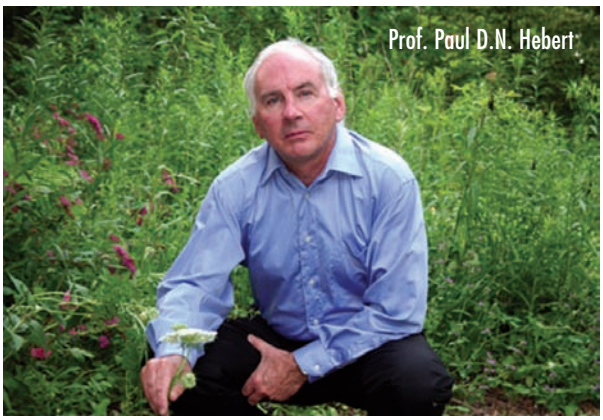
With the armada of technologies now available with molecular biologists is it possible to develop a simple technology with which we could identify species

The sequences obtained as distinct bases

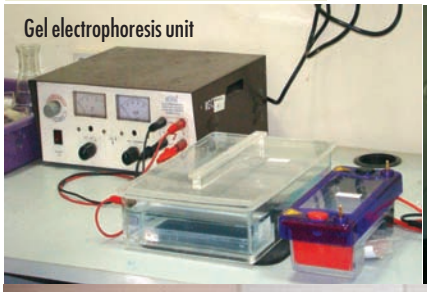
around us at a faster pace? Can it be as simple as the Universal Product Code that is often used for identifying products in super stores? Investigations in this line by scientists have finally resulted in an entirely new technology called DNA Barcoding.

The credit for developing a barcode to distinguish and differentiate species

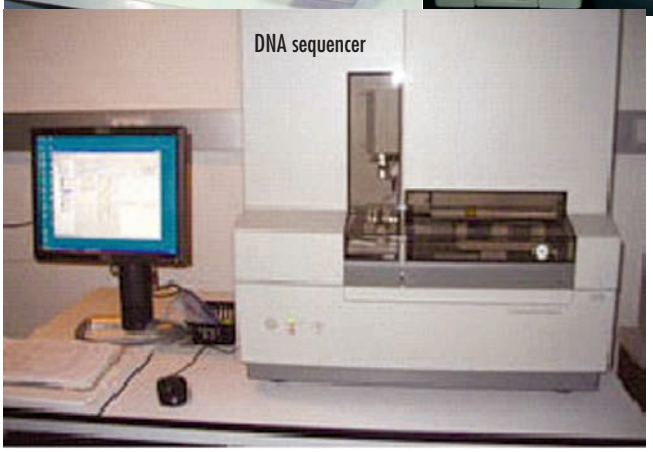
Variations in a short fragment of DNA barcode between human, chimpanzee and gorilla species. Note that in the entire DNA barcode, humans differ from chimpanzees at about 60 locations and gorillas at about 70 locations



Prof. Paul D.N. Hebert


Gel electrophoresis unit


PCR machines


DNA sequencer

DNA barcoding also facilitates identifying illegally obtained wildlife species and wildlife-derived products based on fragments and poorly preserved samples.

goes to Prof. Paul D.N. Hebert of Guelph University, Ontario, Canada. In 2003, Hebert and his colleagues read and compared the DNA barcodes from museum specimens of 260 species of North American birds. They found that birds in different species had very different barcodes, whereas birds within the same species did not. Based on barcode analysis, they also surmised that some birds fall into four new species!

What's the Barcode?

DNA sequence analysis of a uniform target gene (genetic marker) in an organism's mitochondrial DNA to enable species identification is called DNA barcoding. Why has this technique opted for mitochondrial DNA? Primarily, it is considerably smaller than the nuclear DNA and this makes sequencing much easier. Mitochondrial DNA (mtDNA) has a relatively fast mutation rate, which results in significant variance in the sequences between species and, in general, a comparatively small variance within species.

Because mitochondria reproduce without sexual recombination, their genes are less prone to insertions, deletions or other large-scale rearrangements that could spoil the barcode and make it harder to read. Further, as the rate of DNA mutation is inversely related to the size of the genome, nuclear DNA undergoes relatively slow mutation compared with mtDNA. Hence, the former requires a much longer nucleotide sequence than is necessary with mtDNA in order to provide a barcode capable of differentiating species.

In animals, mtDNA occurs as a single double-helical circular molecule containing 13 protein-coding genes, 2 ribosomal genes, a non-protein-coding control region, and several tRNAs. Each mitochondrion contains several such circular molecules and, therefore, several complete sets of mitochondrial genes. As several mitochondria occur within each cell, even though the amount of tissue available for examination is limited, sufficient quantity of DNA could be isolated.

Let us now see which region of mtDNA could be used as the genetic marker or barcode. A 648-bp (base pair) region of the cytochrome c oxidase subunit I gene (popularly known as COI or cox 1) is the most popular barcode used worldwide. COI possesses a very slow rate of amino acid change, and is highly conserved at the DNA sequence level, particularly within species. Insertions and deletions are rare in this gene. Moreover, COI is easier to isolate from a wide range of organisms and hence the popularity.

DNA barcode of a frog species

In Hebert's work COI variations between 260 species of birds studied averaged 7.93%, whereas variation within species averaged 0.43%. They found that every single one of the birds studied had a different COI sequence. In these species represented by more than two specimens, COI sequences were either identical or were most similar to sequences of the same species. The specimens that exhibited greater variations in sequences or arrangement of nitrogenous bases were considered as a separate species. Hebert, aptly referred to as the father of DNA barcoding, proposed a standard sequence threshold to define new species; this threshold, the so-called "barcoding gap", was defined as 10 times the mean intraspecific variation for the group under study.

The order of mitochondrial DNA's nitrogenous bases Adenine (A), Thymine (T), Cytosine (C) and Guanine (G) would fill the role taken by numbers in a commercial barcode. This means each species has its own sequence of ATCG in a shorter stretch of mtDNA, as an identification mark!

In primates, each cell has about 3.5 billion base pairs. The COI barcode is only 648 base pairs long. Yet examples taken from humans, chimpanzees and the other great apes harbour enough differences to distinguish the groups. The COI barcode sequences in *Homo sapiens* showed differences only in one or two base pairs out of 648bp in COI. On the other hand,

On the other hand, the importance given to barcoding may result in gross oversimplification of the science of taxonomy, and may reduce the already lesser amounts of grants available for basic taxonomic studies.

DNA barcodes can help expand our knowledge by exploring many more species rapidly and inexpensively. Once widespread, this system will revolutionize access to biological information and affect research and many other areas in which societies interact with biodiversity.

we differ from chimpanzees at about 60 locations and gorillas at about 70 locations. If large variations are observed within the species, it may point towards the presence of another species. A similar COI barcode sequence study resulted recently in the recognition of two species of orangutan.

In plants the use of the COI sequence is not appropriate for most species (especially in higher plants) because of a much slower rate of COI gene evolution than in animals. In flowering plants nuclear internal transcribed spacer region and the plastid trnH-psbA intergenic spacer regions are identified as potential DNA barcodes. Of late, the chloroplast genes matK and rbcL are recommended by scientists as the best barcode for land plants. Since 2003, studies by researchers on different living kingdoms have shown that DNA barcoding could be used as the 'master key' for revealing the identity of a wonderful array of organisms within it.

In general, the sequences used thus far for molecular barcoding, in addition to COI, are the mitochondrial genes such as cyotchrome a, cytochrome b, nuclear small subunit ribosomal RNA gene (SSU, also known as 16S in prokaryotes, and 18S in most eukaryotes), the nuclear large-subunit ribosomal RNA gene (LSU, also known as 23S and 28S), the highly variable internal-transcribed spacer section of the ribosomal RNA cistron (ITS, separated by the 5S ribosomal RNA gene into ITS1 and ITS2 regions), and the chloroplast ribulose bisphosphate carboxylase large subunit (rbcL) gene.

How It's Done?

The various steps involved in DNA barcoding include preservation of tissues, isolation of DNA, amplification of DNA, sequencing of target DNA, comparison of the sequence with the databases available, and identification of species.

Specimens are the raw material for any barcoding programme. The specimens collected with ecological notes such as habitat, geographical location, photograph in fresh condition, etc. and date of collection would always prove ideal. While fresh or freshly frozen tissues are ideal for analysis, the samples for DNA extraction may also be fixed in 95% ethanol, not in other fixatives such as formalin. For most studies, samples as small as 25 mg would be more than sufficient for DNA extraction, as there is provision for amplification of these genes.

There are two methods for DNA isolation: DNA release and DNA extraction. DNA release protocols aim to rapidly release DNA into solution, making it accessible for subsequent applications such as Polymerized Chain Reaction (PCR). Release-based methods also enable DNA isolation from samples without their physical disruption. In this case, the entire specimen can be removed after DNA isolation, allowing the retention of specimens.

The more popular method is DNA extraction, which results in obtaining more pure DNA often by binding it to a membrane (e.g. silica) or by chemical fractionation. A variety of protocols and commercial kits are now available for DNA extraction.

The DNA samples are extracted in various media such as agar through electrophoresis. Agarose gel electrophoresis is a method used to separate DNA, molecule by size. This is achieved by moving negatively charged nucleic acid molecules through an agarose matrix with an electric field (electrophoresis). Shorter molecules move faster and migrate farther than longer ones. DNA is negatively charged and hence attracted towards the positive end of the gel. The shorter fragments move faster than the longer fragments. DNA is separated on the basis of size (molecular weight).

The extracted DNA samples are now amplified (copied several times) through a process referred to as DNA amplification, with the help of an instrument called Polymerase Chain Reaction (PCR) machine. PCR is a technique to amplify a single or few copies of a piece of DNA across several orders of magnitude, generating thousands to millions of copies of a particular DNA sequence.

PCR products are visualized on 1.2% agarose gels and visible products are selected for sequencing. The sequencing of amplified DNA barcodes can be done using sequencing machines. The amplified segment of the COI or any selected gene is sequenced and this sequence is referred to as the "barcode". The machine will provide data on the arrangement of nitrogenous bases (each with distinct colour) in the DNA and good quality sequences having no background noise (by examining peaks in the electropherogram) are selected for editing using software such as Bio-Edit or Sequencher and analysed.

The sequence information can then be uploaded and matched with existing barcodes or material from voucher specimens. The barcode data of a large number of species from all parts of the world are now available in open source databases such as GenBank and Barcode of Life Database (BOLD). Remember, you can confirm the identity only when the barcode data of the species under study is available in the databases. If you are an expert in taxonomy, you can upload your barcode data of species concerned with details, which can then be used for other workers for comparison.

Potential Uses

As a standardized and high-tech identification tool for species identification, DNA barcoding will have broad scientific applications. Since this method is faster and relatively cheaper, the identification of species can be done at a faster pace. Hence, it will be of great utility in biodiversity surveys and in conservation biology. This may relieve the enormous burden of identifications from taxonomists, so that they can focus on delimiting taxa, resolving their relationships and discovering and describing new species.

Barcodes help to identify cryptic and polymorphic species, the separation of which is one of the Herculean tasks faced by traditional taxonomists. Potential cryptic species identified using DNA barcoding include, among others, butterflies, flies, birds, arachnids, springtails and parasitic worms. Further, with this technique, it is possible to make identification from any life history stages of organisms such as egg,

pupa, instars and larva. The applications of DNA barcodes, however, are not limited to identifying species and solving taxonomic ambiguities.

Ten Reasons for Barcoding Life

1. Works with fragments. Barcoding can identify a species from bits and pieces, including undesirable animal or plant material in processed foodstuffs and morphologically unrecognizable products derived from protected or regulated species.

2. Works for all stages of life. Barcoding can identify a species in its many forms, from eggs and seed, through larvae and seedlings, to adults and flowers.

3. Unmasks look-alikes. Barcoding can distinguish among species that look alike, uncovering dangerous organisms masquerading as harmless ones and enabling a more accurate view of biodiversity.

4. Reduces ambiguity. A barcode provides an unambiguous digital identifying feature for identification of species, supplementing the more analog gradations of words, shapes and colors.

5. Makes expertise go further. Scientists can equip themselves with barcoding to speed identification of known organisms and facilitate rapid recognition of new species.

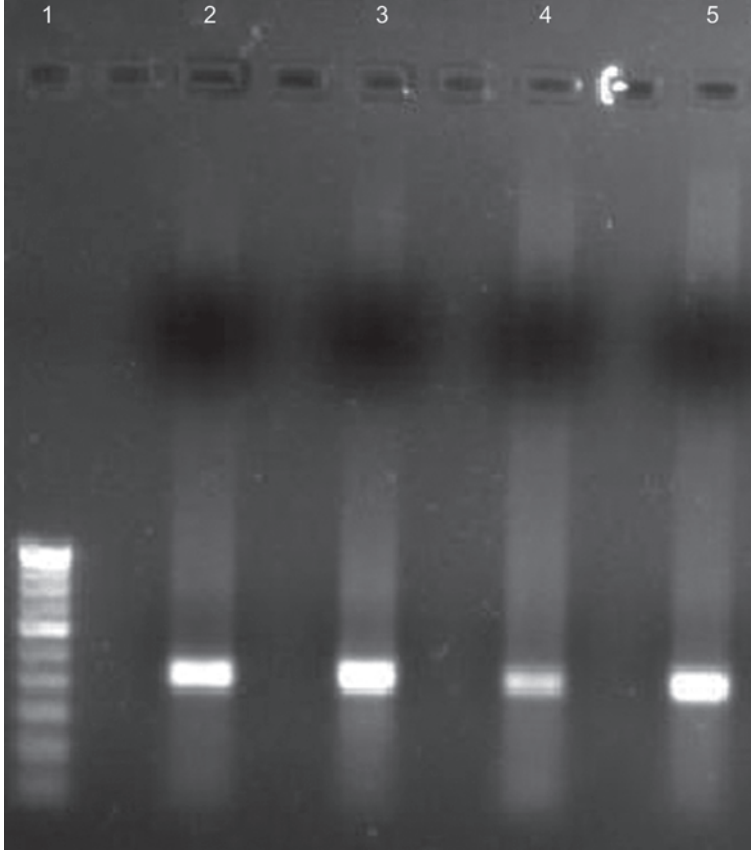
6. Democratizes access. A standardized library of barcodes will empower many more people to call by name the species around them.

7. Opens the way for an electronic handheld field guide. Barcoding links biological identification to advancing frontiers in DNA sequencing, electronics, and information science, paving the way for handheld devices for species identification.

8. Sprouts new leaves on the tree of life. Barcoding the similarities and differences among the estimated 10 million species of animals and plants will help show where their leaves belong on the tree of life.

9. Demonstrates value of collections. Compiling the library of barcodes begins with the multimillions of specimens in museums, herbaria, zoos, and gardens, and other biological repositories, thus highlighting their ongoing efforts to preserve and understand Earth's biodiversity.

10. Speeds writing the encyclopedia of life. A library of barcodes linked to named specimens will enhance public access to biological knowledge, helping to create an on-line encyclopedia of life on Earth.



Amplified DNA in agarose gel; 1 marker; 2-5 amplified regions of the DNA seen as white bands

Ever since inquisitive human beings started grouping, classifying and naming living beings around them, the amazing diversity of life and the subtle variations that distinguish each species of organism in the planet puzzled them a lot.

In population genetics investigations, DNA barcodes can provide a first signal of the extent and nature of population divergences and will facilitate comparative studies of population diversity in many species. The barcode data will also provide important insights into evolutionary processes and could be used to strengthen evolutionary biology studies. The barcodes of various related species can be grouped together using appropriate software for finding out evolutionary distances and relationships among species. Studies initiated so far in this direction show that genetic distances among CO1 barcodes are largely congruent with understanding developed through traditional taxonomy, suggesting a library of barcodes will help evolutionary study.

DNA barcoding also facilitates identifying illegally obtained wildlife species and wildlife-derived products based on fragments and poorly preserved samples. This facility is now increasingly used for finding out illegal trade of endangered flora and fauna. Trade of endangered timber species may also be halted by using barcodes to identify processed wood and lumber products. In India also barcoding has been used to identify illegally transported meat of wildlife and damaged carcasses of endangered fauna.

This technique can be used for assessing the past biodiversity of the earth. For example, researchers have sequenced the CO1 gene of a group of extinct moas (an extinct bird species from New Zealand) using 26 sub-fossil bones, and detected six moa species. Enzyme cocktails are now available that repair DNA damaged by ultra-violet light, oxidation, heat and other factors. Further, techniques are also developed to isolate DNA from formalin-preserved and dry specimens in the museum. Some of the internationally reputed museums in the world have now drawn up specific programmes to barcode their museum deposits.

In marine biology the potential barcode data could be used for getting more reliable identification of catch and by-catch on commercial vessels and at the dock, for better understanding of the food chain through analysis of gut contents and for improved fish stock assessments, based on identification of larvae as well as juveniles and adults.

In environmental studies, barcoding is used for identifying insects and other invertebrates in rivers and streams that are critical indicators of environmental quality. In bio-security (particularly for regulating international transport of living organisms, including pathogens) and agricultural quarantine issues (especially in pest management), this facility could also be effectively employed with great success.

In plants the use of the COI sequence is not appropriate for most species (especially in higher plants) because of a much slower rate of COI gene evolution than in animals.

This is critically important in recent times as biological invasion is considered a major threat to biodiversity, and bioterrorism a greater risk to humanity.

The U.S. Federal Aviation Authority and U.S. Air Force are supporting bird barcoding as it may help reduce bird-aircraft collisions. Today, DNA-based methods are more frequently employed for food authentication. Restaurant owners and consumers could check fish to be sure what they are buying is what is advertised.

Criticisms

The greater applications of DNA barcoding in biological studies notwithstanding, this cannot be projected as the final word for species identification. There are many grey areas as well. In organisms where mtDNA genes are maternally inherited, one species with more than one mtDNA sequence, in cases of hybridisation, male-killing microorganisms, cytoplasmic incompatibility-inducing symbionts, horizontal gene transfer, etc there are chances of errors.

Another argument is that the importance given to barcoding may result in gross oversimplification of the science of taxonomy, and may reduce the already lesser amounts of grants available for basic taxonomic studies. However, many of these criticisms stem from the extravagant claims that barcoding will supersede or radically transform traditional taxonomy.

The fact is that both traditional taxonomy and molecular taxonomy using DNA barcodes are complementary, and each one could be used for strengthening the other! DNA barcodes should make species recognition in the field much easier and relatively error-free, especially where traditional methods are not practical. In addition, species identification should become more reliable, particularly for non-experts.

Promising Future

Initially referred to as DNA typing or profiling, the DNA barcoding initiative has taken this step forward, and several taxa have now been surveyed in their natural habitats using this technique. A complete DNA-based inventory of the Earth's present biota using large-scale high-throughput DNA barcoding is an ambitious proposal rivaling even the Human Genome Project.

In order to take advantage of the various potentials offered by barcoding technology to biological sciences and to mankind, several international initiatives have been launched in many parts of the world.

Barcode of Life Initiative (BoLI) is an international movement of researchers, research organizations, and users who are dedicated to developing DNA barcoding as a global standard for species identification. The basic objectives of this initiative are: (i) Creating a global reference library of diagnostic barcode sequences for identifying species; (ii) Engaging with government agencies, private companies, and other potential users of DNA barcode data to launch new barcoding projects; (iii) Creating tools for a cost-effective, rapid system of species identification that can be used by non-specialists; (iv) Adding 21st century diagnostic molecular data to the defining characteristics (morphology, ecology, behaviour, geography) now used to identify biological species; and (v) Expanding the toolkit used by taxonomic researchers as they improve our understanding of global biodiversity.

In 2005, the Consortium for the Barcode of Life (CBOL) was also created and joined by many natural history museums and herbaria, research organizations and private partners. BOLD (www.barcodinglife.org) is an informatics workbench aiding the acquisition, storage, analysis and publication of DNA barcode records. BOLD is freely available to any researcher with interests in DNA barcoding. BOLD now has over 688,000 records from more than 63,000 species in animal kingdom. Each of these records contains the species name, barcode sequence, collection location, links to the voucher specimen, photographs and other biological data.

The goal of CBOL is fostering development of international research alliances needed to build, over the next 20 years, a barcode library for all eukaryotic life. It has already initiated the first campaigns with a global sweep; they seek to deliver barcode coverage for all species of birds and fishes by 2012.

India is also partner to these international initiatives and several research centers and academic institutions have started barcoding species occurring in our geographical area. Department of Biotechnology, Government of India has also made DNA-based taxonomy as a thrust area for future research. The Kerala State Council for Science, Technology & Environment has initiated the process of establishing a separate centre at Thiruvananthapuram for barcoding of the rich biodiversity of the Western Ghats.

As the technical aspects of large-scale production of molecular barcodes become more refined and a wide array of new uses for the barcode data become increasingly apparent, the future of this technology offers more promises. More research is needed on standardization of the markers, DNA banking and proper taxonomic vouchering. There is also a need for creating DNA and tissue banks.

DNA barcoding is an accurate, rapid, cost-effective, and universally accessible DNA-based system for species identification. DNA barcodes can help expand our knowledge by exploring many more species rapidly and inexpensively. Once widespread, this system will revolutionize access to biological information and affect research, policy, pest and disease control, food safety, resource management, conservation, education, recreation, and many other areas in which societies interact with biodiversity.

Prof. Alan Wildeman, Vice-President (Research), Univer-sity of Guelph remarked: "Ever since I became aware of the Barcode of Life project, I felt that I was an observer on the deck of the Beagle as it was going for-ward. Unlike traditional voyages that were gov-erned by compasses and maps, this one is guided by computers, and genomics and partnerships. But the outcome may fundamentally change how we look at species on our planet."

Dr A. Biju Kumar is with the Department of Aquatic Biology and Fisheries, University of Kerala, Thiruvananthapuram-695581, Kerala. Email: abiju@rediffmail.com

Dr Sanil George is with the Chemical Biology Group, Rajiv Gandhi Centre for Biotechnology, Thiruvananthapuram-695014, Kerala. Email: sgeorge@rgcb.res.in