Journal of
Cell Science &
Molecular Biology

Open Science Publications

Volume 2, Issue 1 - 2015

www.opensciencepublications.com

DNA Metabarcoding: A New Approach for Rapid Biodiversity Assessment

Review Article

Pavan-Kumar A*, Gireesh-Babu P and Lakra WS

Division of Fish Genetics and Biotechnology, ICAR-Central Institute of Fisheries Education, Versova, Mumbai-61

***Corresponding author:** Dr. Annam Pavan Kumar, Scientist, Division of Fish Genetics and Biotechnology, ICAR-Central Institute of Fisheries Education, Versova, Mumbai-61, India, E-mail: pavanannam@gmail.com

Article Information: **Submission:** 29/11/2014; **Accepted:** 25/02/2015; **Published:** 06/03/2015

Abstract

Biodiversity characterization is important to understand the ecological processes on earth. The recent advancements in molecular techniques have enabled us to identify the species composition more efficiently than the traditional methods. In DNA metabarcoding, the pooled genomic DNA extracted from environmental samples is used to amplify evolutionarily conserved genes by universal primers and sequenced using next generation sequencing technologies. In this brief review, the concept of DNA metabarcoding and its applications, limitations and challenges have been discussed.

Keywords: Biodiversity; DNA metabarcoding; Next Generation Sequencing; Environmental DNA

Introduction

The extant /present biodiversity is a result of several million years of evolution of life on earth. Biodiversity is a key component of ecosystem and plays a major role in proper functioning of the ecosystem. Several factors like climate change, habitat loss and invasive species are disturbing the ecosystem biotic components thereby adversely affecting the function and services of ecosystem [1]. The effective management measures for restoring the degraded ecosystem can be taken if the information (data) about indicator species / abundance and pattern of biological diversity (Species) in that ecosystem is available. Traditional morphological and meristic tools for characterizing and assessing the biodiversity demands high skilled personnel and have limitations in identifying cryptic species. With the advent of molecular biology, DNA based species identification methods have been devised using molecular markers (mitochondrial and nuclear). Since the last decade, taxonomically informative genes have been tested over large groups of organisms (animals: mitochondrial cytochrome c oxidase subunit I [2]; Fungi: Nuclear ribosomal internal transcribed spacer [3]; Plants: two chloroplast genes, rbcL & matK [4]; Bacteria: 16S rRNA & protein coding Chaperonin-60, *cpn60* [5-6] for their efficiency to delimit the species and designated them as barcode genes for respective groups. The success in this approach resulted in creation of huge reference databases that include species taxonomic details along with DNA barcode gene sequences (Fish-BOL, BOLD, MarBOL, QBOL etc). These reference barcode sequence databases are useful in assigning taxon to unknown specimen by comparing the sequence similarity of specimen barcode gene with reference database. Until recently, most of the barcoding studies were aimed at developing reference databases by generating species specific DNA barcodes from individual specimens. However, it is also important to characterize/assess the species diversity and abundance within an ecosystem as a whole to understand the spatial and temporal changes in species diversity [7]. Normal DNA barcoding approach using Sanger sequencing method can identify only one specimen at a time and cannot identify multiple species if the sample contains a mixture of different species. With the advancements in sequencing technology, it is now possible to assess the species composition of ecosystems including environmental samples such as soil, sediment and water at a stretch than screening individual specimens at a time.

DNA Metabarcoding

Taberlet et al. [8] introduced the term DNA metabarcoding to designate high-throughput multispecies identification using the total or typically degraded DNA extracted from an environment sample or from bulk samples of entire organisms. The multispecies identification technique was originally applied to microbial communities [9] in the name of metagenomics, and now it is being applied for eukaryotic

ISSN: 2350-0190

organisms such as fungi [10], invertebrates [11], plants [12] and vertebrates [13-15]. Metabarcoding differs from metagenomics in several ways as metagenomics refers to the study of all genomes within a particular ecosystem whereas metabarcoding aims to study a subset of genes / gene. From methodology point of view, metagenomics approach includes preparation of shotgun (random) libraries for sequencing while metabarcoding is based on amplicon sequencing. Metagenomics approach generally used to get more insights about the interaction between species within an ecosystem (taxonomic and functional information). Metabarcoding approach is mainly used to document / characterize species diversity in the ecosystem and it can have better coverage to identify rare taxa within an ecosystem.

DNA Metabarcoding methodology

Selection of Next Generation Sequencing (NGS) technology

A series of high-throughput sequencing technologies based on different chemistries and detection techniques have been introduced commercially. All these NGS technologies can generate several hundred thousands of millions of sequencing reads in parallel. This massively parallel throughput sequencing capacity can generate sequence reads from fragmented libraries of a specific genome (i.e. genome sequencing) or from a pool of PCR amplified molecules (i.e. amplicon sequencing, [16]). Metabarcoding approach relies on this technology where large number of amplicons of taxonomic informative (barcode) gene can be sequenced without a necessity for cloning [8]. A comparison of currently available NGS platforms is given in Table 1. Till now, most of the DNA metabarcoding studies have used Roche 454 FLX platform due to its ability to produce long read length and relatively short run time. However, this platform cannot read homopolymers accurately and may provide erroneous sequences. This problem has been largely alleviated by implying bioinformtic tools that filter out erroneous sequences [17]. Other sequencing technologies such as Illumina, SoLiD and Ion torrent platforms can read homopolymers accurately but the read length is relatively low. Selection of appropriate sequencing methodology depends on the question to be addressed and length of fragment (amplicon). If amplicon length is short (100-200bp), Illumina and Ion torrent platforms are appropriate whereas for large amplicons Roche GS FLX is more useful.

PCR amplification of DNA

A certain genomic region can be amplified from the DNA extracted from sample that contains a mixture of species. Although any part of the genome can be used to delimit the species, certain features like mutation rate (molecular evolution rate), availability of universal primers, short sequences with sufficient phylogenetic signal and availability of comprehensive taxonomic reference database are some of the important features to consider before selecting a DNA fragment for metabarcoding studies. Some researchers have used short fragments of the nuclear 18S and 28S ribosomal markers for metazoans [18-19], but these regions may underestimate species diversity due to their slow rate of evolution compared to other mitochondrial markers [20-22]. Short fragment of mitochondrial 12S ribosomal gene has successfully been used for delineating metazoans, however, taxonomic reference databases are limited for this marker compared to cytochrome *c* oxidase subunit I [18]. The mitochondrial partial cytochrome *c* oxidase I gene (COI) has been adopted as the standard barcode gene for most of the animal groups [3] and this gene has the most represented taxonomic reference database in public domain. However, through *in silico* [23] and empirical analyses it has been found that the available universal primers for COI gene are not well conserved in certain groups viz., nematodes [22,24], echinoderms [25] and gastropods [26]. In summary, the accuracy of metabarcoding is highly dependent on marker choice, but unfortunately no marker has all the features to be used as a perfect metabarcoding marker and the best marker choice could be study-specific [27]. Ficetola et al. [23] have developed a software ecoPCR (electronic PCR) to test the efficacy of barcode primers. Based on two parameters: barcode coverage (Bc) and barcode specificity (Bs), this method measures the conservation of the primers and the capacity of the amplified region to discriminate between taxa [28]. This software facilitates a preliminary comparison of several DNA regions to identify the most appropriate barcodes. A summary of the genes / markers used for various metabarcoding studies is given in table 2.

Amplicon multiplexing, library preparation and sequencing

Generally, NGS platforms are intended to sequence whole genome of important model and non-model organisms and during this process they generate thousands of millions of reads per investigation. However, metabarcoding studies aims to sequence short fragment of homologous gene (amplicon sequencing) from different species of many samples. Sequencing each sample (contains many organisms) separately is not economical. Multiplexing of samples with short, identifying sequences (barcodes/ multiplex identifier (MID)/index) is a widely used strategy in which different molecular tags (4-5 nucleotides) are attached to all DNA fragments / amplicons to identify samples. This kind of sequence indexing is used for data sorting after sequencing and to assign the sequence reads to specific samples. The most effective way to produce an indexed amplicons is to amplify the genomic target region using PCR with specific primers that include a sequencing adaptor and a barcode (Figure 1) [29-30]. Other indexing strategies rely on ligation of barcodes or barcoded sequencing adapters to the DNA amplicons [31]. After indexing, amplicons can be pooled (equimolar concentration of each amplicon) and sequencing would be carried out using the pooled barcoded libraries. The number of different samples to be pooled per sequencing run is determined by number of barcodes available. High quality reagents with barcoded adapters and PCR primers are readily available in kits from many vendors. After tagging with MIDs, the amplicon library fragments are clonally amplified onto the bead particles through emulsion PCR. These particles containing sequence clones are deposited on chips/ flow cells for sequencing by NGS platforms. Designing of tags / MID sequences is important as these sequences may lead to PCR bias and several software are available for this purpose (BARCRAWL [32]; oligotag program of OBITools, http://www.gren oble.prabi.fr/trac/OBITools).

Data analysis

The data generated by NGS platforms are huge and traditional computer operating systems (Windows/DOS) are not capable to

Citation: Pavan-Kumar A, Gireesh-Babu P, Lakra WS. DNA Metabarcoding: A New Approach for Rapid Biodiversity Assessment. J Cell Sci Molecul Biol. 2015;2(1): 111.

Table 1: Comparison of available NGS technologies.

Category	Platform	Read length (bp)	Maximum Number of reads /run	Sequencing output	Runtime
PCR based NGS technologies	Roche 454 GS FLX	400-500	1X 10^6	450Mb	10h
	Roche 454 GS FLX+	600-800	1X10^6	700Mb	23h
	Roche 454 GS Junior	400	1X10^5	~35Mb	10h
	Roche 454 GS junior+	~700	1X10^5	~70Mb	18h
	Illumina Hi Seq 2500	100-200	3X10^8	10-300Gb	7-60h
	Illumina Hi Seq 3000	100-200	2X10^9	125-750 Gb	<1-3.5 days
	Illumina Hi Seq 4000	100-200	2X10^9	125-1500Gb	<1-3.5 days
	Illumina Mi Seq	100-300	7X10^6	0.3-15 Gb	5-55h
	AB SOLiD 5500 system	35-75	2.4 X10^9	~100Gb	4 d
	AB SOLiD 5500xl system	35-75	6X10^9	~250 Gb	7-8 d
	Ion Torrent 314 chip	100-200	1X10^6	≥10 Mb	3.5h
	Ion torrent 316 chip	100-200	6 X10^6	≥ 100Mb	4.7 h
	Ion Torrent 318 chip	100-200	11 X 10^6	≥ 1Gb	5.5h
Single Molecule Sequencing technologies	Helicos Heliscope	30-35	1 X10^9	~20-28 Gb	≤ 1d
	Pacific Biosciences System	≥1500	50 X10^3	~60-75 Mb	0.5h

Table 2:

ssDNA fragment	Target species group	Primer sequences	Amplicon length	NGS platofrm	Reference (Primers)	Reference (Study)
Mitochondrial Cytochrome *c* oxidase subunit I	Arthropods	*LCO1490*: 5´-GGTCAACAAATCATAAAGATATTGG-3´ *HCO2198*: 5´-TAAACTTCAGGGTGACCAAAAAATCA-3´	658 bp	Roche GS FLX 454	Folmer et al. [71]	Yu et al. [38]
	Metazoans	*mlCOIintF:* 5'-GGWACWGGWTGAACWGTWTAYCCYCC-3' *mlCOIintR:* 5'-GGRGGRTASACSGTTCASCCSGTSCC-3'	~320 bp	Roche GS FLX 454	Leray et al.[72]	Leray M et al. [72]
	Insect content in Avian diet	*ZBJ-ArtF1c:* 5'-AGATATTGGAACWTTATATTTTATTTTT-GG-3' *ZBJ-ArtR2c:*5'-WACTAATCAATTWCCAAATCCTCC-3'	160 bp	Roche GS Junior	Zeale et al. [73]	Coghlan et al. [74]
	Arthropods	*Fol-degen-for:* 5'-TCNACNAAYCAYAARRAYATYGG-3' *Fol-degen-rev:* 5'-TANACYTCNGGRTGNCCRAARAAYCA-3'	~650 bp	Roche GS FLX 454	Yu et al. [38]	Yu et al.[38] Ji et al. [75] Yang et al. [39]
				Illumina HiSeq 2000		Liu et al. [76]
Mitochondrial 12S ribosomal DNA	Fish species	*12S V5 F: 5'*-ACTGGGATTAGATACCCC-3' *12S V5R:* 5'-TAGAACAGGCTCCTCTAG-3'	106 bp	Illumina Mi Seq	Riaz et al. [77]	Kelly et al. [78]
	Vertebrates		98 bp	Illumina Hi Seq		De Barba *et al.*, [66]
	Avian species	*12S a*: 5'-CTGGGATTAGATACCCCACTAT-3' *12S h*: 5'- CCTTGACCTGTCTTGTTAGC-3'	250 bp	Roche GS Junior	Cooper [79]	Coghlan et al.[74]
	Mammals	*12S a' F:* 5'-CTGGGATTAGATACCCCACTA-3' *12S o R:* 5'-GTCGATTAT AGG ACAGGTTCCTCTA-3'	100 bp	Ion Torrent	Cooper et al.[80]	Murray et al. [81]
Mitochondrial 16S ribosomal DNA	Invertebrates	*16SMAV-F: 5'*-CCAACATCGAGGTCRYAA-3' *16SMAV-R:* 5'-ARTTACYNTAGGGATAACAG-3'	36 bp	Illumina Hi Seq	De Barba et al. [66]	De Barba et al. [66]
	Mammals	*16S mam1F:* 5'- CGGTTGGGGTGACCTCGGA-3' *16S mam2R:* 5'-GCTGTTATCCCTAGGGTAACT-3'	90-100 bp	Roche GS Junior	Taylor [82]	Coghlan et al.[74]
Mitochondrial 16S ribosomal DNA	Fish species	*16S1F-deg: 5'*- GACGAKAAGACCCTA-3' *16S2R-deg: 5'*- CGCTGTTATCCCTADRGTAACT-3'	180-270 bp		Deagle et al. [83]	
	Fish species	*Chord_16S_F:* 5'CGAGAAGACCCTRTGGAGCT-3' *Chord_16S_R_Short:* 5'-CCTNGGTCGCCCCAAC-3'	120 bp	Ion torrent	Deagle et al. [68]	Deagle et al. [68]
	Bacteria	*S-D-Bact-0341-b-S-17:* 5'-CCTACGGGNGGCWGCAG-3' *S-D-Bact-0785-a-A-21*: 5'-GACTACHVGGGTATCTA ATCC-3	400 bp	Roche GS FLX	Herlemann et al. [84]	Klindworth et al.[85]

Citation: Pavan-Kumar A, Gireesh-Babu P, Lakra WS. DNA Metabarcoding: A New Approach for Rapid Biodiversity Assessment. J Cell Sci Molecul Biol. 2015;2(1): 111.

The P6 loop of the chloroplast DNA trnL intron (UAA)	Plants	*g :* 5'-GGGCAATCCTGAGCCAA-3' *h:* 5'-CCATTGAGTCTCTGCACCTATC-3'	143 bp	Roche GS FLX	Taberlet et al. [86]	Yoccoz et al.[52]
				Illumina GA-II		Que´me´re´ E, [87]
				Roche GS Junior		Coghlan et al. [74]
		c: 5'-CGAAATCGGTAGACGCTACG-3' *d:* 5'-GGGGATAGAGGGACTTGAAC-3'	569 bp	Illumina Hi Seq	Taberlet et al. [88]	De Barba et al.[66]
Chloroplast ribulose-bisphosphate carboxylase gene (rbcL)		*rbcLa_f:* 5'-ATGTCACCACAAAC AGAGACTAAAGC-3' *rbcLa_rev:* 5'-GTAAAATCAAGTCCACCRCG-3'	553 bp	Roche GS FLX	CBoL Plant Working Group [4]	Yoccoz et al.[52]
Nuclear Internal Transcribed Spacer seque-nce2 (ITS 2)		*Forward: 5 ' -ATGCGATACTTGGTGTGAAT-3 ' ;* *Reverse: 5 ' -GACGCTTCTCCAGACTACAAT-3 '*	460 bp	Illumina Mi Seq	Chen et al.[89]	Richardson et al.[90]
		ITS2Ros-F: 5'- YCTGCCTGGGCGTCACA-3' *ITS2Ros-R: 5'- CGTKVGYCGCCGAGGAC-3'*	82 bp	Illumina Hi Seq	De Barba et al. [66]	De Barba et al.[66]
Nuclear Internal Transcribed Spacer seque-nce 1 (ITS 1)	Asteraceae Poaceae	*ITS1-F Forward:* GATATCCGTTGCCGAGAGTC *ITS1Ast-R Reverse:* CGGCACGGCATGTGCCAAGG *ITS1Poa-R Reverse:* CCGAAGGCGTCAAGGAACAC	81 bp	Illumina Hi Seq	Baamrane et al. [91]	De Barba et al. [66]
	Fungal diversity	*ITS1F:* 5'-CTTGGTCATTTAGAGGAAGTAA-3' *ITS4R:* 5'-TCCTCCGCTTATTGATATGC-3'	280 bp	Roche GS FLX	White et al. [92]	Geml et al.[93]
Nuclear 18S ribosomal RNA	Eukaryotic species	*All18SF: 5'*-TGGTGCATGGCCGTTCTTAGT-3' *All18SR:* 5'-CATCTAAGGGCATCACAGACC-3'	~200 bp	Roche GS FLX	Hardy et al. [94]	Chariton et al. [44]
	Marine littoral benthos	*SSUF04:* 5'- GCTTGTAAAGATTAAGCC-3' *SSUR22:* 5'-GCCTGCTGCCTTCCTTGGA-3'	450 bp	Roche 454 GSFLX	Blaxter et al.[95]	Creer et al. [96]

handle the data. UNIX operating system has been considered as the standard computing environment for NGS data. Further, most of the bioinformatics software programs/ algorithms are compatible with UNIX operating system. In UNIX, tasks can be performed by writing commands and in personal computers UNIX environment can be provided by installing Linux. Certain operating systems such Mac OSX (Apple Inc) provide UNIX environment that uses both Graphical User Interfaces (GUI) and command mode.

NGS output data consists of DNA sequences (reads) and corresponding quality values (for each nucleotide of each sequence read). All the resulting sequences may not represent the species composition of sample. The sequence data may contain sequencing noise, PCR chimeras [33], contaminant sequences, nuclear mitochondrial pseudogenes (Numts) and PCR errors. Eliminating all these error sequences from final data analysis is prerequisite before assigning taxon to the sequences. In brief, the data analysis consists of three steps viz., data pre-processing (removal of primers, sequencing adaptors and demultiplexing), processing raw reads (denoising, chimera and PCR artefacts removal) and performing analyses (clustering, BLASTing) (Figure 2). Different software / algorithms have been developed to perform the above tasks. Different researchers have designed different pipelines by combining various software to analyse NGS metabarcode data. Among different packages, QIIME (Quantitative Insights Into Microbial Ecology [34]) has been successfully used for different metabarcoding and metagenetics studies [35-37]. QIIME is an open-source bioinformatics pipeline consists of native python code and additionally cover many external applications for data analysis from raw data processing to taxonomic assignment. It is initially developed for microbial community however; QIIME environment is now being used for metazoan and plant metabarcoding studies [38,39].

Likewise, OBITools is another open source package that has been specifically designed for analyzing metabarcoding data. The main advantage of the OBITools is their ability to take into account the taxonomic annotations, ultimately allowing sorting and filtering of sequence records based on the taxonomy. Apart from these software, some other packages like Operational Clustering of Taxonomic Units from Parallel UltraSequencing (OCTUPUS) Bioconductor packages ShortRead [40] and Biostrings (run on R language) are also available for NGS data analysis.

In metabarcoding, species are defined operationally as a cluster of similar sequences, and the clusters are known as Operational Taxonomic Unit (OTU). The most critical step in the analysis is to assign taxon to sequences / cluster of sequences (OTU). This can be achieved by comparing each sequence to a reference database that is a subset of public databases (eg. EMBL, NCBI GenBank, SILVIA and BOLD) or a set of sequences specifically produced for the study. The comparison for sequence similarity between queried sequence and reference database sequence can be performed through BLAST search or ecotage [41]. In the case of non availability of reference database, sequences would not be linked to a taxonomic name, however; would be clustered in MOTUs (Molecular Operational Taxonomic units) that can be compared in different studies, for example, comparing the diversity of MOTUs in different localities or under different parameters in the same locality [28]. The Program MEGAN (MEta Genome Analyzer) is useful in representing the species/ taxon composition of sample and for taxonomic binning, even for very large data sets [42].

Citation: Pavan-Kumar A, Gireesh-Babu P, Lakra WS. DNA Metabarcoding: A New Approach for Rapid Biodiversity Assessment. J Cell Sci Molecul Biol. 2015;2(1): 111.

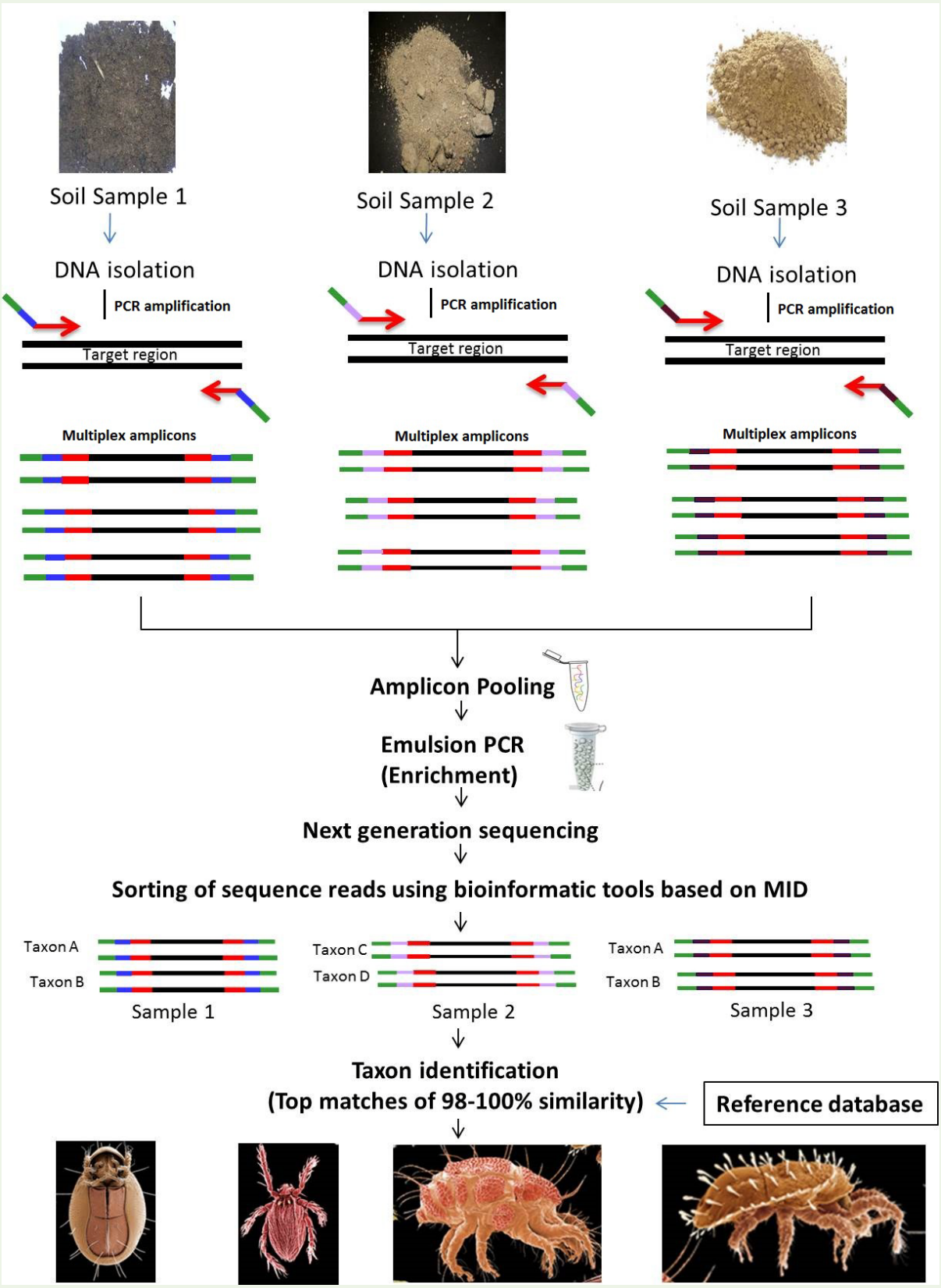

Figure 1: DNA metabarcoding methodology.

Citation: Pavan-Kumar A, Gireesh-Babu P, Lakra WS. DNA Metabarcoding: A New Approach for Rapid Biodiversity Assessment. J Cell Sci Molecul Biol. 2015;2(1): 111.

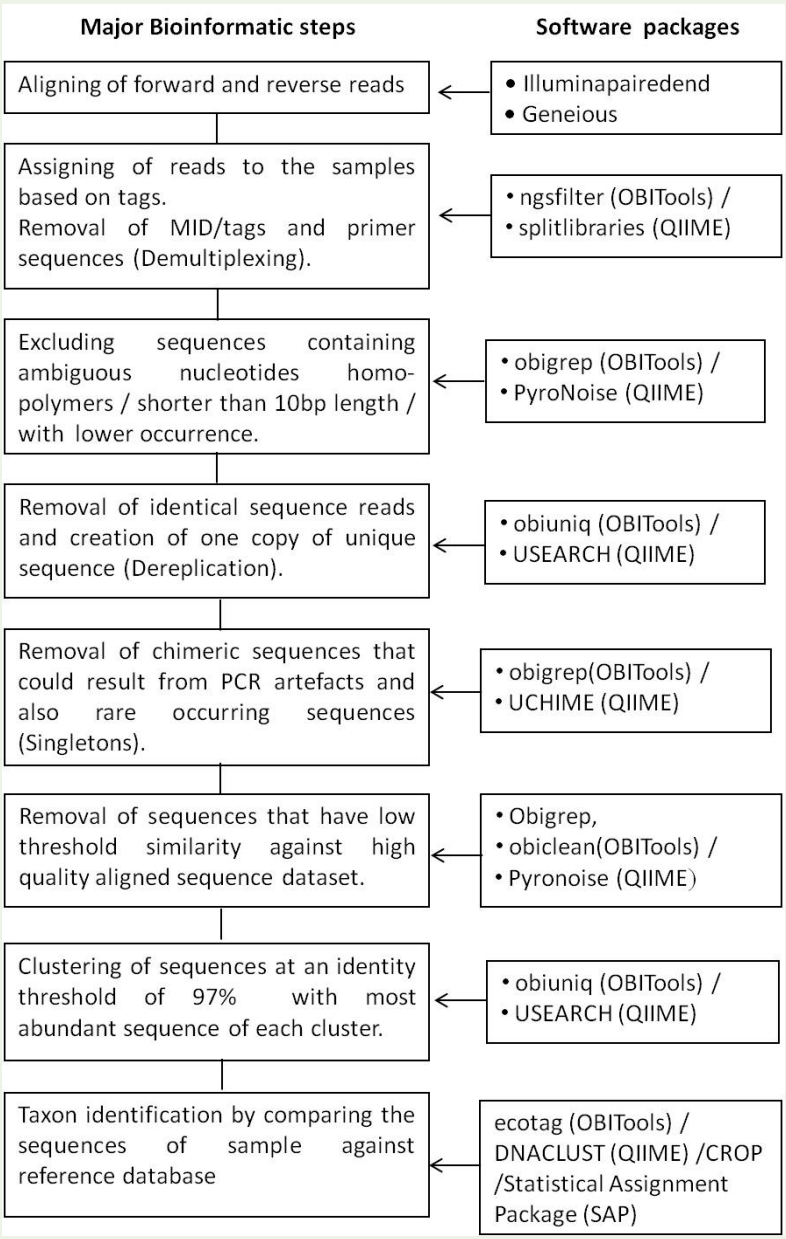

Figure 2: Bioinformatics pipeline for NGS data analysis.
Abbreviations: BOLD: Barcode of Life Database; FISH-BOL: Fish Barcode of Life Initiative ; Mar BOL: Marine Barcode of Life initiative; QBOL: Quarantine organisms Barcode of Life

Applications

Biodiversity studies

In this decade of biodiversity (2011-2020), fund allocation has been increased for biodiversity characterization and these efforts resulted in creation / strengthening of existing taxonomic reference databases such as Global Biodiversity Information Facility (GBIF) and Barcode of Life database (BOLD). These databases have cured reference taxonomic information for about 144,357 Species (BOLD www.boldsystems.org.in [43]) species and are being constantly updating with new taxon information to include all the species on earth. Once the comprehensive database is prepared, DNA metabarcoding approach can be used to analyse the species composition of different types of samples, from soils to sediments, faeces, air and water. Metabarcoding of soil / sediment samples collected from nuclear power plant areas or any other industrial area can be used to compare temporal and spatial species assemblages to assess human impacts on biodiversity [44]. Likewise, water samples can be used to detect the presence of invasive species [45]. Next generation sequencing technology has been used to analyse species composition of sensitive ecosystem (Coral reefs [46]), extreme habitats (acid mines [47]). DNA metabarcoding approach has been successfully used to

Citation: Pavan-Kumar A, Gireesh-Babu P, Lakra WS. DNA Metabarcoding: A New Approach for Rapid Biodiversity Assessment. J Cell Sci Molecul Biol. 2015;2(1): 111.

characterize soil microbial diversity [48-50], fungal diversity [51] and plant diversity [52] using 16S rRNA, ITS and P6 loop of the plastid DNA *trnL* intron amplicons, respectively. Hajibabaei et al. [53] were used short fragments of COI DNA barcodes were used to identify freshwater macro invertebrates from benthic samples.

The effect of climate change on biodiversity could be assessed and species distributions can be predicted for future if data on past distributions together with past climate conditions are available. DNA metabarcoding of soil samples collected at different depths could provide a new source of information about past species distributions [7]. Murray et al. [54] have analysed ancient DNA using metabarcoding approach and identified diverse range of taxa, including endemic, extirpated and previously unrecorded taxa. Haouchar et al. [55] assessed ancient DNA of vertebrate fossils and plants and provided valuable information about past biodiversity of Kangaroo Island, Australia. Haile et al. [56] utilized both 454 pyrosequencing and conventional Sanger sequencing methods in the analysis of ancient DNA recovered from Arctic permafrost cores

Trophic studies

The interaction between predator and prey play an important role in maintaining ecosystem health and stability. Gut content analysis of the species to identify their feeding habits, especially for endangered species will help to formulate conservation measures/ strategies [57]. Traditionally, the diet composition of any species is assessed by macro- or micro histological methods [58] and stable isotopes [59]. However, these methods are time consuming, require highly skilled personnel and cannot identify variably digested food items. To overcome this, DNA isolated from gut content and faeces can be used for the molecular identification of diet composition by metabarcoding. Several studies have used NGS technology for investigating gut microbe ecology and species composition of diet. Some of these studies have included analyses of herbivore diet from gut contents using the plastid trnL sequence [13, 41, 60-61]. Several studies investigated the species composition of diet by analysing prey DNA collected from faeces of Australian fur seal (*Arctocephalus pusillus doriferus* [27]) little penguin (*Eudyptula minor* [62-63]), reptile [15] leopard cat [64]. Recently De Barba et al., [66] analysed plant, vertebrate and invertebrate components of the diet of brown bear by analysing faecal matter through multiplexing strategy.

Limitations

Metabarcoding, like many new technological advances in science, offers new opportunities and at the same time new challenges. Since metabarcoding studies generally include amplicon sequencing, factors like PCR efficiency, primer tags and sequencing efficacy need to be considered to avoid errors [67, 68, 69]. To circumvent primer tag bias, a two stage PCR where template DNA is first amplified using untagged primers and subsequently by tagged primers during the last few PCR cycles has been suggested [53,68]. Another limitation is lack of comprehensively cured reference databases for certain metazoans for assigning taxon to the OTUS. Future studies are needed to improve sampling strategies (selection of season, sampling location within habitat) and to understand the relationship between sequence reads and species density [70]. Further, the integration of knowledge from ecology, taxonomy and evolution is essential for addressing any biodiversity questions using metabarcoding.

References

1. Sala OE, Chapin FS, Armesto JJ, Berlow E, Bloomfield J, et al. (2000) Global biodiversity scenarios for the year 2100. Science 287: 1770-1774.
2. Hebert PDN, Cywinska A, Ball SL, deWaard JR (2003) Biological identifications through DNA barcodes. Proc Biol Sci 270: 313-322.
3. Schoch CL, Seifert KA, Huhndorf S, Robert V, Spouge JL, et al. (2012) Nuclear ribosomal internal transcribed spacer (ITS) region as a universal DNA barcode marker for Fungi. Proc Natl Acad Sci USA 109: 6241-6246.
4. CBoL Plant Working Group (2009) A DNA barcode for land plants. Proc Natl Acad Sci USA 106: 12794-12797.
5. Makarova O, Contaldo N, Paltrinieri S, Kawube G, Bertaccini A, Nicolaisen M (2012) DNA barcoding for identification of "Candidatus phytoplasmas" using a fragment of the elongation factor Tu gene, PLoS ONE.
6. Links MG, Dumonceaux TJ, Hemmingsen SM, Hill JE (2012) The chaperonin-60 universal target is a barcode for bacteria that enables de novo assembly of metagenomic sequence data. PLoS One.
7. Yoccoz GN (2012a) The future of environmental DNA in ecology. Mol Ecol 21: 2031-2038.
8. Taberlet P, Coissac E, Pompanon F, Brochmann C, Willerslev E (2012) Towards next-generation biodiversity assessment using DNA metabarcoding. Mol Ecol 21: 2045-2050.
9. Sogin ML, Morrison HG, Huber JA, Welch DM, Huse SM, et al. (2006) Microbial diversity in the deep sea and the underexplored "rare biosphere.". Proc Natl Acad Sci USA 103: 12115-12120.
10. Fouts DE, Szpakowski S, Purushe J, Torralba M, Waterman RC, et al. (2012) Next Generation Sequencing to Define Prokaryotic and Fungal Diversity in the Bovine Rumen. PLoS One 7: e48289.
11. Porazinska DL, Giblin-Davis RM, Esquivel A, Powers TO, Sung W, et al. (2010) Ecometagenetics confirms high tropical nematode diversity. Mol Ecol 19: 5521-5530.
12. Hiiesalu I, Opik M, Metsis M, Lilje L, Davison J, et al. (2012) Plant species richness belowground: higher richness and new patterns revealed by next-generation sequencing. Mol Ecol 21: 2004-2016.
13. Kowalczyk R, Taberlet P, Coissac E, Valentini A, Mique C et al. (2011) Influence of management practices on large herbivore diet—Case of European bison in Białowieza Primeval Forest (Poland). F Ecology and Management 261: 821-828.
14. Raye´ G, Miquel C, Coissac E, Redjadj C, Loison A, et al. (2011) New insights on diet variability revealed by DNA barcoding and high-throughput pyrosequencing: chamois diet in autumn as a case study. Eco Res 26: 265-276.
15. Brown DS, Jarman SN, Symondson WO (2012) Pyrosequencing of prey DNA in reptile faeces: analysis of earthworm consumption by slow worms. Mol Ecol Res 12: 259-266.
16. Shokralla S, Spall JL, Gibson JF, Hajibabaei M (2012) Next-generation sequencing technologies for environmental DNA research. Mol Ecol 21: 1794-1805.
17. Quince C, Lanzen A, Curtis TP, Davenport RJ, Hall N, et al. (2009) Accurate determination of microbial diversity from 454 pyrosequencing data. Nat Methods 6: 639-641.
18. Machida RJ, Knowlton N (2012) PCR Primers for metazoan nuclear 18S and 28S ribosomal DNA sequences. PLoS One, 7:e46180.
19. Fonseca VG, Carvalho GR, Sung W, Johnson HF, Power DM, et al. (2011) Second-generation environmental sequencing unmasks marine metazoan biodiversity. Nat Commun 1: 98 doi:10.1038/ncomms1095.

Citation: Pavan-Kumar A, Gireesh-Babu P, Lakra WS. DNA Metabarcoding: A New Approach for Rapid Biodiversity Assessment. J Cell Sci Molecul Biol. 2015;2(1): 111.

20. Hillis D, Dixon M (1991) Ribosomal DNA - molecular evolution and phylogenetic inference. Q Rev Biol 66: 411-453.

21. Tautz D, Arctander P, Minelli A, Thomas RH, Vogler AP (2003) A plea for DNA taxonomy. Trends Ecol Evol 18: 70-74.

22. Derycke S, Vanaverbeke J, Rigaux A, Backeljau T, Moens T (2010) Exploring the use of Cytochrome Oxidase c Subunit 1 (COI) for DNA barcoding of free living marine nematodes. PLoS One 5: e13716.

23. Ficetola GF, Coissac E, Zundel S, Riaz T, Shehzad W, et al. (2010) An in silico approach for the evaluation of DNA barcodes. BMC Genomics 11: e434 doi: 10.1186/1471-2164-11-434.

24. Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, et al. (2011) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Mol Ecol 19: 4-20.

25. Hoareau TB, Boissin E (2010) Design of phylum-specific hybrid primers for DNA barcoding: addressing the need for efficient COI amplification in the Echinodermata. Mol Ecol Res 10: 960-967.

26. Meyer CP, Paulay G (2005) DNA Barcoding: Error Rates Based on Comprehensive Sampling. Godfray C, ed. PLoS Biology 3:e422.

27. Deagle BE, Kirkwood R, Jarman SN (2009) Analysis of Australian fur seal diet by pyrosequencing prey DNA in faeces. Mol Ecol 18: 2022-2038.

28. Coissac E (2012) OligoTag a program for designing set of tags for next generation sequencing of multiplexed samples. In: Data Production and Analysis in Population Genomics (eds Bonin A and Pompanon F). Methods in Molecular Biology Series, Humana Press, New York

29. Binladen J, Gilbert MTP, Bollback JP, Panitz F, Bendixen C, Nielsen R, Willerslev E (2007) The use of coded PCR primers enables high throughput sequencing of multiple homolog amplification products by 454 parallel sequencing. PLoS One 2:e197.

30. Hoffmann C, Minkah N, Leipzig J, Wang G, Arens MQ, et al. (2007). DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic Acids Res 35:e91.

31. Meyer M, Stenzel U, Hofreiter M (2008) Parallel tagged sequencing on the 454 platform. Nat Protoc 3: 267-278 .

32. Frank DN (2009) BARCRAWL and BARTAB: software tools for the design and implementation of barcoded primers for highly multiplexed DNA sequencing. BMC Bioinformatics 10: 362.

33. Lenz T, Becker, S (2008) Simple approach to reduce PCR artefact formation leads to reliable genotyping of MHC and other highly polymorphic loci - implications for evolutionary analysis. Gene 427: 117-122.

34. Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, et al. (2010) Qiime allows analysis of high-throughput community sequencing data. Nature Meth 7: 335-336.

35. Brinkman BM, Hildebrand F, Kubica M, Goosens D, Favero JD, et al. (2011) Caspase deficiency alters the murine gut microbiome. Cell Death and Dis 2: e220.

36. Flores GE, Bates ST, Knights D, Lauber CL, Stombaugh J, et al. (2011) Microbial biogeography of public restroom surfaces. PLoS One 6: e28132.

37. Kostka JE, Prakash O, Overholt WA, Green SJ, Freyer G, et al. (2011) Hydrocarbon degrading bacteria and the bacterial community response in gulf of mexico beach sands impacted by the deepwater horizon oil spill. Appl Environ Microbiol 77: 7962-7974.

38. Yu DW, Ji YQ, Emerson BC, Wang XY, Ye CX, et al. (2012) Biodiversity Soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. Methods Ecol Evol 3: 613-623.

39. Yang C, Wang X, Miller JA, Blécourt M, Ji Y, et al. (2014) Using metabarcoding to ask if easily collected soil and leaf-litter samples can be used as a general biodiversity indicator. Ecol Indic 46: 379-389.

40. Morgan M, Anders S, Lawrence M, Aboyoun P, Pages H, et al. (2009) ShortRead: a Bioconductor package for input, quality assessment and exploration of high-throughput sequence data. Bioinformatics 25: 2607-2608.

41. Pegard A, Miquel C, Valentini A, coissac E, Bouvier F, et al. (2009) Universal DNA-based methods for assessing the diet of grazing livestock and wildlife from faeces. J Agric Food Chem 57: 5700-5706.

42. Huson DH, Mitra S, Ruscheweyh HJ, Weber N, Schuster SC (2011) Integrative analysis of environmental sequences using MEGAN4. Genome Res 21: 1552-1560.

43. Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System. Mol Ecol Notes 7: 355-364.

44. Chariton AA, Court LN, Hartley DM, Coloff MJ, Hardy CM (2010) Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. Front in Ecol and Envi 8: 233-238.

45. Ficetola GF, Miaud C, Pompanon F, Taberlet P (2008) Species detection using environmental DNA from water samples. Biol Lett 4: 423-425.

46. Wegley L, Edwards R, Rodriguez-Brito B, Liu H, Rohwer F (2007) Metagenomic analysis of the microbial community associated with the coral Poritesastreoides. Envi Micro 9: 2707-2719.

47. Edwards RA, Rodriguez-Brito B, Wegley L, Matthew Haynes, Mya Breitbart, et al. (2006) Using pyrosequencing to shed light on deep mine microbial ecology. BMC Genom 7: 57.

48. Roesch LFW, Fulthorpe RR, Riva A, Casella G, Hadwin AK, et al. (2007) Pyrosequencing enumerates and contrasts soil microbial diversity. The ISME Journal 1: 283-290.

49. Rousk J, Baath E, Brookes PC, Lauber CL, Lozupone C, et al. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. ISME J 4: 1340-1351.

50. Nacke H, Thurmer A, Wollherr A, Will C, Hodac L, et al. (2011) Pyrosequencing based assessment of bacterial community structure along different management types in German forest and grassland soils. PLoS ONE 6: e17000.

51. Jumpponen A, Jones KL, Blair J (2010) Vertical distribution of fungal communities in tall grass prairie soil. Mycologia 102: 1027-1041.

52. Yoccoz NG, Brathen KA, Gielly L, Haile J, Edwards ME et al. (2012b) DNA from soil mirrors plant taxonomic and growth form diversity. Mol Ecol 21: 3647-3655.

53. Hajibabaei M, Shokralla S, Zhou X, Singer GAC, Baird DJ (2011) Environmental barcoding: a next-generation sequencing approach for biomonitoring applications using river benthos. PLoS ONE 6: e17497.

54. Murray DC, Haile J, Dortch J, White NE, Haouchar D, et al. (2013) Scrapheap Challenge: A novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. Sci Rep 3: 3371.

55. Haouchar D, Haile J, McDowell MC, Murray DC, White NE, et al. (2014) Thorough assessment of DNA preservation from fossil bone and sediments excavated from a late Pleistocenee- Holocene cave deposit on Kangaroo Island, South Australia. Quat Sci Rev 84: 56-64.

56. Haile J, Froese DG, MacPhee RDE, Roberts RG, Arnold LJ et al. (2009) Ancient DNA reveals late survival of mammoth and horse in interior Alaska. Proc Natl Acad Sci USA 106: 22352-22357.

57. Marrero P, Oliveira P, Nogales M (2004) Diet of the endemic Madeira Laurel Pigeon Columba trocaz in agricultural and forest areas: implications for conservation. Bird Conser Inter 14: 165-172.

58. McInnis ML, Vavra M, Krueger WC (1983) A comparison of 4 methods used to determine the diets of large herbivores. J of Ran Man 36: 700-709.

59. Fry B (2006) Stable Isotope Ecology. Springer-Verlag, New York City, New York.

60. Soininen EM, Valentini A, Coissac E, Miquel C, Gielly L, et al. (2009) Analysing diet of small herbivores: the efficiency of DNA barcoding coupled

Citation: Pavan-Kumar A, Gireesh-Babu P, Lakra WS. DNA Metabarcoding: A New Approach for Rapid Biodiversity Assessment. J Cell Sci Molecul Biol. 2015;2(1): 111.

with high-throughput pyrosequencing for deciphering the composition of complex plant mixtures. Front Zool 6: 16.

61. Valentini A, Miquel C, Nawaz MA, Bellemain E, Coissac E, et al. (2009) New perspectives in diet analysis based on DNA barcoding and parallel pyrosequencing: the trnL approach. Mol Ecol Resour 9: 51-60.

62. Deagle BE, Chiaradia A, McInnes J, Jarman SN (2010) Pyrosequencing faecal DNA to determine diet of little penguins: is what goes in what comes out? Conserv Genet 11: 2039-2048.

63. Murray DC, Bunce M, Cannell BL, Oliver R, Houston J, et al. (2011) DNA-Based Faecal Dietary Analysis: A Comparison of qPCR and High Throughput Sequencing Approaches. PLoS ONE 6(10): e25776.

64. Shehzad W, Riaz T, Nawaz MA, Miquel C, Poilot C, et al. (2012) Carnivore diet analysis based on next generation sequencing: application to the leopard cat (Prionailurusbengalensis) in Pakistan. Mol Ecol 21: 1951-1965.

65. Deagle BE, Jarman SN, Coissac E, Pompanon F, Taberlet P (2014) DNA metabarcoding and the cytochrome c oxidase subunit I marker: not a perfect match. Biol Lett 10: 0562.

66. De Barba M, Miquel C, Boyer F, Mercier C, Rioux D, et al. (2013) DNA metabarcoding multiplexing and validation of data accuracy for diet assessment: application to omnivorous diet. Mol Ecol Resour 14: 306-323.

67. Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, Gu Y (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics, 13: 341.

68. Deagle BE, Thomas AC, Shaffer AK, Trites AW, Jarman SN (2013) Quantifying sequence proportions in a DNA-based diet study using Ion Torrent amplicon sequencing: which counts count?. Mol Ecol Resour 13: 620-633.

69. Berry D, Mahfoudh KB, Wagner M, Loy A (2011) Barcoded primers used in multiplex amplicon pyrosequencing bias amplification. Appl Environ Microbiol 77: 7846-7849.

70. Herder JE, Valentini A, Bellemain E, Dejean T, van Delft CWJJ, et al. P(2014). Environmental DNA-a review of the possible applications for the detection of invasive species. Stitching RAVON, Nijmegen. Report 2013-14.

71. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. Mol Mar Biol Biotechnol 3: 294-299.

72. Leray M, Yang JY, Meyer CP, Mills SC, Agudelo N, et al. (2013) A new versatile primer set targeting a short fragment of the mitochondrial COI region for metabarcoding metazoan diversity: application for characterizing coral reef fish gut contents. Front Zool 10: 34.

73. Zeale MRK, Butlin RK, Barker GL, Lees DC, Jones G (2011) Taxon-specific PCR for DNAbarcoding arthropod prey in bat faeces. Mol Ecol Resour 11: 236-244.

74. Coghlan ML, White NE, Murray DC, Houston J, Rutherford W, et al. (2013) Metabarcoding avian diets at airports: implications for birdstrike hazard management planning. Investigative Genetics 4: 27.

75. Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, et al. (2013) Reliable, verifiable and efficient monitoring of biodiversity via metabarcoding. Ecol Lett 16: 1245-1257.

76. Liu S, Li Y, Lu J, Su X, Tang M, et al. (2013) SOAPBarcode:revealing arthropod biodiversity through assembly of Illumina shotgun sequences of PCR amplicons. Methods Ecol Evol 4: 1142-1150.

77. Riaz T, Shehzad W, Viari A, Pompanon F, Taberlet P, et al. (2011) ecoPrimers: inference of new DNA barcode markers from whole genome sequence analysis. Nucleic Acids Res 39: e145.

78. Kelly RP, Port JA, Yamahara KM, Crowder LB (2014) Using Environmental DNA to Census Marine Fishes in a Large Mesocosm. PLoS ONE 9(1): e86175.

79. Cooper A: DNA from museum specimens. New York: Springer; 1994.

80. Cooper A, Lalueza-Fox C, Anderson S, Rambaut A, Austin J, et al. (2001) Complete mitochondrial genome sequences of two extinct moas clarify ratite evolution. Nature 409: 704-707.

81. Murray DC, Haile J, Dortch J, White NE, Haouchar D, et al. (2013) Scrapheap challenge: a novel bulk-bone metabarcoding method to investigate ancient DNA in faunal assemblages. Sci Rep 3: 1-8.

82. Taylor PG (1996) Reproducibility of ancient DNA sequences from extinct Pleistocene fauna. Mol Biol Evol 13: 283-285.

83. Deagle BE, Gales NJ, Evans K, Jarman SN, Robinson S, et al. (2007) Studying Seabird Diet through Genetic Analysis of Faeces: A Case Study on Macaroni Penguins (Eudyptes chrysolophus). PLoS ONE 2(9): e831.

84. Herlemann DPR, Labrenz M, Juergens K, Bertilsson S, Waniek JJ, et al. (2011) Transition in bacterial communities along the 2000 km salinity gradient of the Baltic Sea. ISME J 5: 1571-1579.

85. Klindworth A, Pruesse E, Schweer T, Peplies J, Quast C, et al. 2013) Evaluation of general 16S ribosomal RNA gene PCR primers for classical and next-generation sequencing-based diversity studies. Nucleic Acids Res 41:e1.

86. Taberlet P, Coissac E, Pompanon F et al. (2007) Power and limitations of the chloroplast trnL (UAA) intron for plant DNA barcoding. Nucleic Acids Res 35: e14.

87. Quéméré E, Hibert F, Miquel C, Lhuillier E, Rasolondraibe E, et al. (2013) A DNA Metabarcoding Study of a Primate Dietary Diversity and Plasticity across Its Entire Fragmented Range. PLoS ONE 8(3): e58971.

88. Taberlet P, Gielly L, Pautou G, Bouvet J (1991) Universal primers for amplification of three non-coding regions of chloroplast DNA. Plant Mor Bio 17: 1105-1109.

89. Chen S, Yao H, Han J, Liu C, Song J, et al. (2010) Validation of the ITS2 region as a novel DNA barcode for identifying medicinal plant species. PLoS One. 5:e8613.

90. Richardson RT, Lin CH, Sponsler DB, Quijia JO, Goodell K, et al. (2015) Application Of Its2 Metabarcoding To Determine The Provenance Of Pollen Collected By Honey Bees In An Agroecosystem. Appl Plant Sci 3(1). pii: apps.1400066.

91. Baamrane MAA, Shehzad W, Ouhammou A, Abbad A, Naimi M, et al. (2012) Assessment of the food habits of the Moroccan dorcas gazelle in M'Sabih Talaa, West central Morocco, using the trnL approach. PLoS ONE 7: e35643.

92. White TJ, Bruns T, Lee S, Taylor JW (1990)Amplification and direct sequencing of fungal ribosomal RNA genes for phylogenetics. In PCR protocols: a guide to methods and applications. Edited by Innis MA, Gelfand DH, Sninsky JJ, White TJ, Innis MA, Gelfand DH, Sninsky JJ, White TJ. New York, N.Y: Academic Press, Inc; 315-322.

93. Geml J, Gravendeel B, van der Gaag KJ, Neilen M, Lammers Y, et al. (2014) The Contribution of DNA Metabarcoding to Fungal Conservation: Diversity Assessment, Habitat Partitioning and Mapping Red-Listed Fungi in Protected Coastal Salix repens Communities in the Netherlands. PLoS ONE 9(6): e99852.

94. Hardy CM, Krull ES, Hartley DM, Oliver R (2010) Carbon source accounting for fish using combined DNA and stable isotope analyses in a regulated lowland river weir pool. Mol Ecol 19: 197-212.

95. Blaxter ML, De Ley P, Garey JR , Liu LX, Scheldeman P, et al. (1998) A molecular evolutionary framework for the phylum Nematoda. Nature 392: 71-75.

96. Creer S, Fonseca VG, Porazinska DL, Giblin-Davis RM, Sung W, et al. (2010) Ultrasequencing of the meiofaunal biosphere: practice, pitfalls and promises. Mol Ecol 19: 4-20.

Citation: Pavan-Kumar A, Gireesh-Babu P, Lakra WS. DNA Metabarcoding: A New Approach for Rapid Biodiversity Assessment. J Cell Sci Molecul Biol. 2015;2(1): 111.