Chapter 5

Discussion
5. Discussion

Taxonomic classification of living species based on shared features has been a part of human society for centuries (Vences 2013: 201-244). The quantification of biodiversity, which is increasingly decimated, presents a daunting challenge to taxonomists because it requires discovery and analysis to proceed at a greatly accelerated pace (Smith et al., 2005: 1825-1834). The recognition of subtle anatomical differences between closely related species requires subjective judgment of a taxonomist and also demands in-depth taxonomic knowledge (Brooks et al., 2004: 616-618). According to Hebert et al. (2003a: 313), with the current pace of taxonomic identification of animals, it will take centuries to complete even preliminary ‘Encyclopedia of Life’ of our planet. Thus to speed up the rate of species identification, molecular characterization and identification of animal species from DNA Barcoding was adopted by researchers across the world in the last decade on the basis of its initial success in the identification of fishes, birds and insects of the order Lepidoptera (Hebert et al., 2003a: 313-321; b: S96-S99; 2004a: 14812-14817; b: 312). The majority of molecular characterization studies with large number of insect samples has been reported from new world (American continents) (Hebert et al., 2003a: 313-321; Smith et al., 2007: 4967-4972), however, a large biomass of animals evolved on the Asian continent are still awaiting their identification in Global databases. India is considered one of the biodiversity rich subcontinents in northern hemisphere of the world and insects inhabit this land with more than one million species. In the current study, for the first time insets were collected at this large scale for molecular identification work from all over India. Based on a study conducted by Paul Hebert and his team and the Consortium for the Barcode of Life, the partial DNA sequence of a mitochondrial protein coding gene, cytochrome c oxidase subunit 1 established as a DNA barcode for metazoans (Hebert et al., 2003a: 313-321). Unlike taxonomy this technique is not life stage specific for identification as DNA remains conserved throughout their life cycle. Thus, the present work was initiated in the year 2013 to check the applicability of the concept of DNA Barcoding on various insect species of agricultural importance. Collections of samples were made from a diverse array of insects inhabiting India by applying both mitochondrial and nuclear markers. The primers selected to amplify the targeted loci in the insect DNA were evaluated by guidelines provided by CBOL and BOLDSYSTEMS. Three major criteria were laid down by CBOL and BOLDSYSTEMS in the year 2007 for standardizing the molecular identification of metazoans and development
of their database. These were (i) primers for targeted loci should be robust so that the desired region can be retrieved easily. Universality: using single primer pair desired DNA fragment should be obtained across the species belonging to different Genus or family. (ii) DNA sequences obtained from sequencing machine should be of high quality, the loci used must be amenable for the production of bidirectional (in both forward and reverse directions) sequences without any ambiguous base and manual editing of DNA sequences in the coding frame should be submitted to databases. (iii) The loci which enables recovery of the desired DNA fragment should only be used to infer phylogenetic trees and for BLAST analysis. In the present study, insects representing 8 major insect orders of agricultural importance and 2 orders of non insect species were identified by molecular characterization and their DNA sequences were submitted to NCBI-GenBank, DNA barcodes were developed from BOLDSYSTEMS by uploading specimen and DNA sequence data for each insect species. The details of research framework, methodologies and results are discussed further in detail.

A total of 316 insect species from eight insect orders were collected from 20 states and 2 union territories, all these species were collected from 58 different species of host plants. As type of collection methods varies from insect to insect species different collection methods were followed to collect different kinds of insects. Insect species were identified by morphological taxonomists immediately upon collection.

5.1 Insect specimen: sample collection, identification and documentation

Insects pose great threat to agricultural, horticultural and ornamental crops in India, various types of insect pests feeds on a variety of commercial crops, they affect animals and poultry birds. Several haplotypes and biotypes of insect population feeds upon selective crops, this will result in lower market value of crops, fruits and vegetables. Rapid and timely identification of insect pests may help at some extent to overcome pest outbreak by developing management strategies. Quick identification of insect pest in the field plays a major role in applying any control measures to reduce losses to farmers.

In the present study, insect was collected by different methods such as moths and butterflies of order Lepidoptera were collected by sweeping net and malaise traps (Townes 1972: 239-267), larvae feeding on fruits and leaves were hand picked, beetles and weevils of order Coleoptera were handpicked or were collected from pit traps. Hemipteran bugs,
whiteflies were collected by aspirators by suction, flies of order Diptera were collected by bait trap methods (Butler 1966: 1030-1031). Odonates were collected by sweeping net method, Orthopterans were hand picked, thrips of order Thysnaoptera were collected by aspirator (Weins and Burgess 1972:1557-1558). Non-insect species, viz., spiders were collected by picking infested leaves and mites were collected by hand picking (Glick 1939: 150). All the species collected were identified by full time insect taxonomists available in our country, if not, species were distributed order wise based on their appearance. Collections brought to the laboratory at Bengaluru were linked with a specimen voucher numbers, the same was used while submitting DNA sequence data at NCBI-GenBank.

Type of preservation method plays an important role in molecular characterization and DNA Barcoding study as DNA should be of high quality for getting a good quality DNA sequence. All the insect species collected in the current study were primarily preserved in absolute alcohol, however, mealybugs were preserved in RNAlater.

5.2 Insect tissue preservation and DNA extraction

Insects collected in their native state were preserved as dry specimens and immediately stored in -80°C deep freezer and alcohol preserved were kept in -20 °C deep freezer at ICAR-NBAIR, Bengaluru, however, mealybugs were collected in RNAlater than normal alcohol preservation method and kept in -80°C ultra deep freezer till DNA was isolated from them.

In any molecular characterization study high quality DNA is a basic requirement to obtain high quality DNA sequence. Initially to standardize DNA extraction protocol, five DNA extraction protocols were followed, viz., three traditional methods, viz., Lysis method, Chelex® and CTAB method and two commercially available kits (Qiagen DNeasy® Blood and Tissue kit and innuPREP® Forensic Kit) were followed. Traditional methods were followed based on their previous success in insect DNA isolation (Folmer et al., 1994: 297; Wang and Wang, 2012: 421-426; Muspa et al., 2013: c3281; Freitas et al., 2014: 8776-8782), however kit based methods proved successful in achieving high quality DNA on Agarose Gel (Sambrok and Russel 2001: 2100; Boykin et al., 2013: 4; Asghar et al., 2015: 132-138). DNA Barcoding demands high quality genomic DNA in order to identify species at a faster rate, many literatures supports commercial kits over traditional methods for successful DNA
isolation for development of high quality DNA barcode (Brandon-Mong et al., 2015: 1-11; Karthika et al., 2016: 1-13; Boykin et al., 2013: 4-5; Asghar et al., 2015: 132-138).

Out of 316 species of insects, ten species of mealybugs were first preserved in absolute alcohol, which was not suitable for getting a high quality of DNA. However, upon getting assistance from Dr. Thibaut Malausa, UMR Institut Sophia Agrobiotech, France (expert in molecular taxonomy for scale insects) through email, he suggested to preserve mealybugs in RNAlater instead of normal alcohol preservation. Thus mealybugs collected from fields were directly transferred to RNAlater upon collection.

To identify insect species based on their DNA data a universal marker should be applied across the insect lineages, hence, a universal primer set of LCO1490 and HCO2198 designed by Folmer et al., (1994: 294-299) was used extensively for all insect species collected in the current study, however, for species of mealybugs were subjected to same loci with slight variations in their primer sequences.

5.3 Selection of target loci in insect DNA

The loci for amplifying a desired fragment to develop DNA barcodes from insects, DNA could be either from mitochondrial or nuclear genome. These could be from both coding as well as non coding regions. However, to qualify as DNA barcode targeted loci should possess predefined characteristics, for mitochondrial protein coding gene cytochrome c oxidase subunit 1 is defined by Hebert et al., (2003a: 312). An ideal DNA barcode sequence should be short in length (550-800bp), which makes it feasible for amplification by universal primers and sequenced by single pass DNA sequencing. The ideal DNA marker should be variable within species to distinctly identify the gap. Most importantly, it should be easy to recover even from dry or archived specimens and should be easy to align for generating comparable data to distinguish species. The initial studies on the molecular identification of insects through DNA barcodes suggested regions, mainly from mitochondrial genome due to its maternal inheritance (Moore 1995: 718-726). Robust primers enabling 5’ region of cytochrome c oxidase subunit 1 was developed by Folmer et al., (1994: 294-299). In the year 2004, CBOL gave the status of the standard barcode region to cytochrome c oxidase subunit 1 for metazoans. Based on these studies cytochrome c oxidase subunit 1 was used as a preliminary marker for identification of insect species in the
present study. However, to locate genetic variations among population of animal species, researchers recommended to use the nuclear ribosomal region as supplementary marker along with mtCO1 (Hebert et al., 2003: 320; Rubinoff, 2005: 1031).

Internal transcribed spacers divides the subunit of ribosomal DNA, in silico analysis of these spacers have proved their conserved region across animal DNA, hence can be used in finding variations among individuals of species (Moore 1995: 718-726; Coeur D’ Acier et al., 2007: 175-193; Boykin et al., 2013: 1-15). Several studies conducted for species identification of animals using both mtCO1 and ITS regions, the majority of these suggested to use both markers in conjunctions, however many studies have proven inability of ITS region to become an ideal marker for DNA barcodes (Boykin et al., 2013: 1-15). Being non-coding regions does not provide predefined characteristics for recognizing their sequences. Current studies based on ITS region report about the co-amplification of ITS region with insect fungal DNA and the region was difficult to recover from most of insect population and the species in this study. Hence contaminant DNA sequences were discarded completely from the analysis to avoid any confusion.

All the 316 insect species and 5 non insect species were amplified by mtCO1 universal primers, however, mealybugs were amplified by primer designed by Park et al., 2010 from the mtCO1 region as Folmer primers failed to amplify a full length DNA sequence from mealybug species. A total of 46 insect populations covering five major insect orders was analyzed by internal transcribed spacer regions. For internal transcribed spacer region 1 and 2 a total of 6 primer pairs were used, however, none of the ITS primers showed universality across class insecta.

5.4 Amplification and purification of the targeted loci

The relative usefulness of target loci from both mitochondrial and nuclear region was tested by PCR amplification. A total of 321 species of insects and non-insect species representing 10 arthropod orders were amplified by mtCO1 with a success rate of 99%. Previously at such large scale Hebert et al., (2003a: 313-321) has covered 200 Lepidopteran species of Canada and reported 96% of success rate with mtCO1 gene. However, only 45 individuals of insects could be amplified by ITS1 or ITS2 primers with 70% of success rate in getting a proper DNA sequence, a total of 6 primer pairs were used out of which 2 primers (1
& 5) were used to amplify ITS1 region and 4 primers (2, 3, 4 & 6) were used to amplify ITS2 region (mentioned in Chapter 3 Materials and Methods, Table 3.2). All the ITS primers used in this study resulted in taxa specific, none of them showed universality. Mitochondrial CO1 worked on 45°C of annealing temperature for all the 311 species of arthropods (exception; 10 species of mealybugs worked at 48°C), which is consistent with studies conducted by Hebert et al., 2003a, 2003b, 2004a and 2004b (mealybugs were not studies). However, ITS region amplified on varying annealing temperatures of 45-60°C. Out of 6 primer pairs used in this study, primer pair 1 amplified ITS1 region from Plutella xylostella at 54°C of annealing temperature, which was consistent with work of Ellango et al., (2012: 241-245). A primer pair five could amplify ITS1 region for only 7 fruit flies out of 16 fruit fly specimens from Genus Bactrocera at 60°C of annealing temperature, in a similar study, conducted by Boykin et al., (2013: 1-19) ITS 1 was amplified from 6 species of fruitflies. Primer pair 2 could amplify ITS2 region from only 4 species of ants at 55°C of annealing temperature, however, only one species of ant could be amplified by Jie-Ya et al., (2003: 584) at same temperature. Primer pair 3 amplified all the 8 populations of Chilo partellus at 45°C of annealing temperature, however, the same primer pair amplified all the eight populations of Tea shot hole borer (Euwallacea fornicatus) at 51°C of annealing temperature which is reported first time. Primer pair 4 was used to amplify ITS2 region from fruitflies of Genus Bactrocera but gave unspecific DNA bands at amplification even at a gradient temperature of 45-55°C, which was in the range of melting temperature of the this primer pair. Primer pair 6 amplified ITS2 region from only fours species representing aphids at 55°C of annealing temperature, however, only one species of aphid could be amplified by Jie-Ya et al., (2003: 584). Both full length and minibarcode regions were amplified on 45°C and 55°C annealing temperatures respectively. All the 15 species collected from Museum gave 99% success rate of amplification with full length and minisocode primers which was consistent with studies conducted by Hajibabaei et al., (2006a: 959-964). However, DNA extracted from Qiagen DNeasy® blood and Tissue kit did not produce high quality DNA for museum specimens for recovering minibarcodes, hence, Innuprep® Forensic Kit was used which resulted in high quality PCR products specially to recover ~170bp PCR product (130bp of DNA sequence).

The amplicons produced are generally sequenced directly. For DNA Barcoding direct sequencing of the amplified product is advocated by CBOL (http://www.barcoding.si.edu/protocols.html) as cloning procedure takes additional time and chemicals are quite costly. However, co-amplification, paralogous genes or spacers are
present in the agarose gel than cloning becomes mandatory to obtain a specific product. In the present study all the PCR products were directly sequenced, before sequencing all the PCR products were purified by Gel purification methods, Hebert et al., (2003a: 314) used Qiaex II kit for PCR product purification. To remove unbound primers and excess dNTPs, Qiaquick PCR purification kit (Qiagen CA) was used extensively to purify the PCR products of insects and non insect species for this study. A single band of amplicon was purified in columns provided in the Qiaquick PCR purification kit. For ITS2 region, PCR products were first loaded on to a 2% agarose gel instead of 1.5% to check for co-amplified products, DNA bands of desired length was first cut with the help of a scalpel and then purified using a Qiaquick PCR purification kit which is inconsistent with studies conducted by Kim and Lee, (2007: 510-522).

All the 642 mitochondrial DNA sequences (321 in both directions) representing mtCO1 gene and 92 DNA sequences (46 in both directions) amplified by ITS1 and ITS2 region were sequenced through capillary sequencing technology and checked for quality by licensed software CODONCODE. All the DNA sequences were passed through a set quality parameter (phread score) of 20Q, DNA sequences passed these criteria only were processed further for DNA sequence analysis through BLAST and subsequent submission in GenBank and BOLDSYSTEMS.

5.5 DNA Sequencing and quality check from chromatogram files

Sanger’s di-deoxy chain termination technology has been used extensively for molecular characterization studies (Sanger et al., 1977: 5463-5467). For the current study, modern capillary sequencers with modified chemistry devised by Sanger applied to generate DNA sequences up to 750bp. In modern automated sequencer made by Applied Biosystems (ABI), 3730xl sequencer based on Sanger’s di-deoxy chain termination method can sequence up to 1000bp (1KB) fragment of DNA in a single run (Chan, 2005: 13).

For the current study all the purified amplified products were sequenced by 3730xl DNA sequencer of Applied Biosystems. Both 16 and 96 capillary system was used for bidirectional sequencing for this study, DNA sequences generated from these sequencers were evaluated by reading their Chromatogram files by CODONCODE™ software. All the trace files were subjected to quality check by recording Phred Score, a minimum of 20Q was
assigned to qualify a base to be called as the actual base (the chance of this base being incorrect is <1%). The Phred score is logarithmically represents error probabilities in base calling, hence this algorithm used by majority of sequence analysis softwares (Ewing and Green, 1998: 186-194). BOLDSYSTEMS accepts a quality score of 20 to qualify a sequence as a DNA barcode in its database, and DNA sequences obtained for this study qualified these criteria (Ratnasingham and Hebert, 2007: 355-364). To assure high quality DNA sequences for development of DNA barcodes each sequence has to be trimmed at both the ends. After trimming of low quality bases from both the sequences (forward and reverse), sequence assembly software (BioEdit 7.0v) assembled the sequence to generate a contig. All the mitochondrial bidirectional sequences (642 in both forward and reverse directions) were assembled into contigs and a total of 321 high quality DNA contigs analyzed by protocols of CBOL and BOLDSYSTEMS and were submitted to both NCBI-GenBank and BOLDSYSTEMS to generate high quality DNA barcodes for 321 arthropod species. Similarly, all the 92 bidirectional DNA sequences (46 populations of insects) produced for ITS regions were also assembled into contigs and submitted to NCBI-GenBank.

All the 316 insects and 5 non insect species were amplified by the universal mitochondrial marker. Primers used for Mitochondrial cytochrome c oxidase subunit 1 could amplify all the arthropods collected in this study. However, primers internal transcribed spacer 1 and 2 could not amplify all the populations in this study and were found to be taxa specific than universal.

5.6 Universality of DNA barcode loci

To determine the universality of the target loci and the primers, the amplification and sequencing success of the same loci were also evaluated in species belonging to non insect orders. The target loci could also be amplified from four species of spiders and one mite species. The criterion for determining universality of the three DNA markers used in this study involved assessment of the targeted regions, which could be routinely recovered and sequenced in the maximum number of analysed insects and non-insects. Hebert et al., 2003a, 2003b, 2004a and 2004b have reported about universality of mtCO1 primers designed by Folmer et al., 1994. Among the tested loci only mitochondrial cytochrome c oxidase subunit 1 (mtCO1) had highest recoverability with amplification success of 99% and sequenced all 321 species in this study. The remaining two loci of ITS region viz., ITS1 and ITS2 though
exhibited low amplification success as compared to mtCO1. To achieve universality, not only primers but PCR reaction and thermal cycler play a major role. Change in PCR cycling condition often and use of different thermal cycler gives different results altogether, hence, same PCR cycling conditions (standradized) and thermal cycler (C1000 of BioRad) was used throughout the study.

All the insect and non insect species collected were identified through BLAST analysis. Insect populations were also identified by internal transcribed spacer region 1 and 2, however, ITS1 and ITS2 were not successful in identifying all the insect populations, instead for some insects viz., *Plutella xylostella*, fruitflies and ants ITS primers showed similarity with fungal DNA upon BLAST analysis.

5.7 Species discrimination rates and evaluating insect DNA barcodes

To obtain clear species resolution, three methods were selected for evaluating DNA barcodes are i) genetic distance, ii) phylogenetic method and iii) BLAST analysis. Genetic distance involves assessment of inter and intra specific variations analyzed by applying K2P method on DNA sequences. BLAST analysis involves % homology to differentiate among species, 97% of threshold limit sets species boundary. Less than 97% of homology between two sequences results into two different species.

In the present study, species were identified by just a fragment of a dipteran fly, which is inconsistent with studies conducted by Wells *et al.*, (2001: 1098-1022) on fragments of a fly from family Sarcophagidae. Molecular characterization studies and BLAST analysis revealed the identity of these fragments obtained in a medicine bottle as *Pollenia rudis* (KT368817). BLAST analysis resulted in 100% homology with GenBank Accession numbers, GQ409351.1 and JF439551.1 respectively. DNA barocode was developed at BOLDSYSTEMS in VETIP project under accession number VETIP006-15, which was later upgraded in versio 4 of BOLDSYSTEMS.

Similarly, two externally similar and commonly misidentified species of lady beetles (Crotch. 1874: 311; Gordon, 1987: 1-46) distributed in Indian subcontinent were identified by molecular characterization and BLAST analysis as *Oenopia sauzeti* Mulsant and *Oenopia mimica* Weise. Blast analysis resulted in 87% of homology for both *O. sauzeti* and *O. mimica*
with *Calvia quatuordecimguttata* (from Germany), this makes both the species first time submission to GenBank database. Pairwise BLAST between DNA sequences of both *O. sauzeti* and *O. mimica* resulted into 89% homology, which clearly indicates that both these species are quite different from each other as a difference of 75 nucleotides between both these species sets them apart. DNA barcodes were evaluated by BOLDSYSTEMS and generated in project AGIMP under accession numbers AGIMP042-15 for *O. sauzeti* and AGIMP043-15 for *O. mimica* (upgraded to version 4).

In the present investigations, mtCO1 has exhibited highest species discrimination rate. The species resolution of 99% with BLAST analysis, makes mtCO1 as a standard DNA barcode marker for metazoans. However, ITS primer has shown only 70% of success rate in the current study in identifying species through BLAST analysis. Mitochondrial cytochrome c oxidase subunit 1 primer used in this study has identified species collected from Salt Lake of The Great Rann of Kutch (GRK), in Gujarat. The unidentified specimens collected from GRK were identified by mtCO1. The present study revealed that 80% of specimens collected were Dipterans and 10% each from order Hymenoptera and Coleoptera. Species discrimination was based on BLAST analysis, ten species showed 77-99% homology in the GenBank database, three species with lowest %homology were identified upto family level only. This study is the first report for molecular characterization of insects from GRK in Gujarat (Ojha *et al.*, 2016a: 166-170), previously the similar study was conducted by Clark and Nguyen (2014: 65) from the Great Salt Lake, Utah in the United States of America.

Use of mtCO1 in identifying putative species applied to insects collected from KAAS plateau in Western Ghats. BLAST analysis of mtCO1 DNA sequences showed 85-99% homology in the GenBank database. All the specimens collected from KAAS plateau, showed <96% homology was identified up to order and family level. Out of 15 insect specimens observed 7 were identified up to order and family level and 8 specimens were designated as a species. Insects from KAAS plateau were first time molecularly characterized and 50% among them were designated as species and 50% could be new species for India (Ojha *et al.*, 2016b: 2170-2174).

As in many studies ITS primers have been used as a secondary primer as it is difficult to recover from the majority of metazoan taxa, however, few studies suggests its use in analyzing genetic variation and phylogenetic analysis along with the mtCO1 gene (Moore,
Interordinal, ordinal and interspecies phylogenetic tree was constructed by applying Neighbor-Joining (NJ) tree and Kimura’s 2 parameter applied throughout analysis, however, for *Pollenia rudis*, maximum parsimony (MP) method was employed instead of NJ tree and maximum likelihood (ML) method was used for populations of *Plutella xylostella*. Phylogenetic tree for interordinal across ten insect orders were constructed by the neighbour-joining method. Phylogenetic tree for order Thysanoptera could not be constructed as minimum four DNA sequences are required for constructing phylogeentic tree in MEGA 6.0v software.

### 5.8 Phylogenetic and genetic variation analysis using both mitochondrial and nuclear regions

Phylogenetic analysis was conducted on different groups of insects, both interspecies and inter-ordinal phylogenetic trees were constructed by applying Neighbor-Joining model as advocated by Hebert *et al.*, (2003a: 316). The phylogenetic tree constructed by mtCO1 gene has determined monophyletic tree for all the insect species in this study. A fragmented dipteran fly identified in this study was used to construct phylogenetic trees with existing similar DNA sequences from GenBank, maximum parsimony tree showed monophyly among individual DNA sequences in an interspecies tree of *P. rudis* from this study and another DNA sequence of *P. rudis* submitted from Germany. Nine species representing order Diptera, Coleoptera and Hymenoptera collected from GRK, in Gujarat formed a monophyletic clade, where species from the order Hymenoptera formed outgroup. *Australospesis niveipennis* has been recorded from different states of India (Ozerov 2005: 14-59; Chakraborty *et al.*, 2014: 6). However, this is the first time it has been recorded from Salt Lake of Rann in Gujarat. *Microchironomus* sp. (? tener) has been recorded from China, Japan, Australia and Sweden (Yan and Wang 2006: 53-68) and usefulness of COX1 gene to improve species resolution and its potential for implementation in monitoring programmes for Chironomidae collected from Baltic Sea (Brodin *et al.*, 2013: 996-1004) has been demonstrated.

Phylogenetic tree was constructed by applying NJ model on butterflies of Bengaluru for the first time, a total of 15 species of butterflies were identified by mtCO1 gene. NJ
clustering analysis resulted in a single monophyletic tree, however, DNA sequences for same species obtained from NCBI-GenBank were from large geographic distances. Deep interspecies and intraspecies divergences were observed as nodes with similar species were branching out with 100% of bootstrap support. Long distance dispersal of butterflies as studied by Craft et al., (2010: 5041-5046) strongly supported by the present study as even conspecific butterfly species from distant geographic locations formed monophyletic clades with our sequences. Hence this study supports the utility of DNA Barcoding in the identification of butterflies as this technique helps in better resolution of butterflies, even when the DNA sequences are from large geographic localities. Our studies are also consistent with studies conducted on butterflies collected from Western Ghats of India by Gaikwad et al., (2012: 2375-2383).

Phylogenetic tree constructed for 10 species of mealybugs, these mealybugs were found infesting on mango trees in mango orchards of Attur farm, Yelhanka, and IIHR, Bengaluru. The tree showed monophyly among mealybugs as six species from 3 major Genus of mealybugs clustered together and this study was found consistent with the studies conducted by Park et al., 2010 and studies conducted on DNA Barcoding of mealybugs in China by Wang et al., (2016: 438-446).

Insects of veterinary importance are less studied groups in India, in the current study insects collected from veterinary sheds and poultry farms in and around Bengaluru were subjected to molecular characterization through mtCO1 gene and phylogenetic tree construction. A monophyletic NJ tree was obtained upon phylogenetic analysis showed interspecies relationship between veterinary important insects from order Diptera. Species belonging to family Sarchophagidae and Muscidae grouped with their conspecific species and species of family Ceratopogonidae formed an out group to species belonging to family Sarcophagidae, Muscidae and Fannidae. Similar studies conducted on European blowflies by Vincet et al., (2000: 820-823), however, they used cytochrome-b (Cyt-b) marker for distinguising blowfly species.

A Neighbor-Joining tree constructed for insects collected from KAAS plateau formed a phylogenetic tree, which resulted into two major clades, showed interordinal relationship among insets representing order Coleoptera, Hemiptera, Diptera, Orthoptera, Lepidoptera and Hymenoptera in first clade and 2nd clade showed interspecies relationship among thrips of
order Thysanoptera. Insect species of KAAS plateau are less studies invertebrate, thus this study suggests more detailed investigation of this landscape to quantify diversity of this plateau. *Adoretus duvauceli* was recorded by Chandra and Gupta (2013: 4660-4671), *A. florea* by Khan *et al.*, (2004: 79-82), *B. lateralis* by Kumar and Yadav (1987: 1192-1193), *C. bissellata* by Kapur (1962: 479-492), *E. excentrica*, *Gynaikothrips* (Denmark *et al.*, 2014) and *E. torus* (Cock, 2015). One specimen could be identified up to genus level, *viz.*, *Cicindela* sp., one up to family level - Pentatomidae and six up to order level only, *i.e.*, Coleoptera (1 specimen), Lepidoptera (1 specimen), Diptera (1 specimen), Orthoptera (1 specimen) and Thysanoptera (2 specimens) (Ojha *et al.*, 2016b: 2170-2174).

As per inter-ordinal analysis through Neighbor-Joining method with General time reversible model, with parameters set to include gamma variations as well, phylogenetic tree constructed to show insect orders evolved in monophyly as all the insect orders clustered were separated from non insect individuals (spiders and mite) of the class Arachnida. Further, each order studied in detail. Order wise phylogenetic tree construction employed on eight insect orders and a combined tree constructed for class Arachnida (including order Araneae and Ixodida). Orderwise insect phylogenetic analysis is first study of its kind conducted for Indian insects.

Intraspecies relationship among populations of *Plutella xylostella* (Linnaeus), was identified by both mtCO1 and Internal Transcribed Spacer region 1 and 2. These moths damage fields of cruciferous crops and cover a distance of over 3000 km in continuous flight for several days. Mitochondrial cytochrome c oxidase subunit 1 revealed maximum variation of seven nucleotides between populations collected from ~12250000km² in India, in this study determined 9 haplotypes by using mtCO1 gene (Ojha *et al.*, 2015: 1-7). A similar study was conducted where one to seven nucleotides of sequence divergence was observed for *Bombyx mori* (Kim *et al.*, 2000: 155-170) for population of *Lycoriella mali* (Bae *et al.*, 2001: 451-457) in an another study populations of *Plutella xylostella* were studied by Li *et al.*, (2006: 605-611). Internal transcribed spacer 1 primers used to construct a NJ tree showed intraspecies sequence divergence among populations of *Plutella xylostella*. The ratio of transition to transversion ‘R’ was quite lesser than obtained by mtCO1 gene. The shift in R may be due to lesser number of population size for ITS1, however, apart from DNA sequences generated from this study no other country have reported this region for *P. xylostella* in NCBI-GenBank, in the absence of ITS1 DNA sequences in NCBI-GenBank.
database from elsewhere, tree was constructed with existing ITS1 sequences. Long distance migration among populations of *P. xylostella* has been reported by Lorimer, (1981:108-109) and Chapman *et al.*, (2002: 641-650). The high genetic variability observed in this study among populations would have helped this species to adapt and evolve at greatly accelerated pace to different environments (Ojha *et al.*, 2015: 1-7).

Populations of *Chilo partellus* (Swinhoe), was studied by using both mitochondrial and nuclear markers, this pest species affects majorly to maize and sorghum fields in India (Divya *et al.*, 2009: 70-74), for the first time *C. partellus* have been studied with internal transcribed spacer 2 region of ribosomal DNA for the present study. DNA sequences submitted for mtCO1 from other countries in NCBI-GenBank are not full length and does not begin with the predefined characters of cytochrome c oxidase subunit 1, hence, no outgroup could be taken for this species as that will mislead phylogenetic analysis for the current study. Mitochondrial marker showed higher rates of base substitution as compared to nuclear gene, which clearly indicates towards intra-specific divergence among the populations of *C. partellus*.

Aphids are generally difficult to identify through morphological analysis, molecular characterization with both mitochondrial and nuclear genes was followed in the current study to discriminate 26 different species of aphids. Mitochondrial CO1 could identify all the 26 species, however all these species were first morphologically identified by the taxonomist, phylogenetic analysis revealed monophyly among aphid species. Colour morphs of one aphid species *Aphis craccivora* could be identified as a single species and formed single subclade in the NJ tree constructed in this study, phylogenetic studies conducted on aphids in the present work is consistent with studies conducted by Rebijith *et al.*, (2013: 601-610), however, we could not trace any cryptic genera of aphids as reported by Rebijith *et al.*, (2013: 601-610). Internal transcribed spacer region 2 could amplify only four species of aphids and found to be taxa specific, hence species specific ITS primers needs to be designed to study populations of aphid species.

Ants are important ecosystem engineers, it is estimated that there are more species of ants in a square kilometer of Brazilian forests than all the lions and elephants in Africa (Folgarait 1998: 1221-1244). In India approx 660 species of ants from 87 genera are reported, out of which family Myrmicinae forms 45% of ant biomass while family Formicinae forms
25% of *Camponotus* and *Polyrhachis* species (Bharti 2010: 172). In the present study mtCO1 was used to identify ant species, phylogenetic analysis of 16 species of ants resulted in a monophyletic NJ tree. The majority of species was identified from family Fomicinae formed a single clade consisting eight species. Family Myrmicinae showed a relationship with species representing families Ponerinae and Dolichoderinae, the phylogentic tree is consistent with the traditional phylogeny of ants studies by Urbani *et al.*, (1992: 301-329). No similar study was conducted in India, hence, this study reported species of ants in the project Ants of India (ANIND) at BOLDSYSTEMS (upgraded to version 4). Internal transcribed spacer could identify only 4 species and found to be taxa specific.

A polyphagous shot hole borer, *Euwallacea fornicatus*, collected from 2 states of South India infesting tea wood. Eight populations of *E. fornicatus* were identified by both mtCO1 and ITS2 genes. Mitochondrial CO1 could identify both male and female population collected from Valparai, Tamil Nadu, however, ITS2 formed different clusters for male and female collected from Valparai. Both the markers found successful in identifying populations of *E. fornicatus*, hence, both these markers in conjunction can be applied to resolve the phylogeny for this species.

Fruitflies of genus *Bactrocera* are polyphagous pests, feeds on variety of fruits. The majority of the fruitfly species collected for this study was found damaging mango orchards of GKVK and Attur Farm in Bengaluru and pumpkin in Jorhat, Assam. Phylogenetic analysis conducted on sixteen species of fruitflies by using mtCO1 DNA sequences resulted in a monophyletic tree. Phylogenetic tree constructed with mtCO1 for fruitflies showed a close relationship among *dorsalis* and *carambolae* group. However, the phylogenetic tree constructed with mtCO1 in this study showed contradictory results as compared to the studies conducted by Boykin *et al.*, (2013: 11-13), species representing *carambolae* group forms a separate cluster to *dorsalis* group. Internal transcribed spacer 1 could amplify only seven species of fruitflies, phylogenetic tree constructed based of spacer 1 region could show somewhat similarity with studies of Boykin *et al.*, (2013: 1-19), however, for better resolution of phylogeny in the genus *Bactrocera* large sample has to be assembled and both mitochondrial and nuclear genes has to be concatenated (joined) for phylogenetic reconstruction through bayesian method (Degnan and Salter, 2005: 24-37; Kubatko and Degnan, 2007: 17-24).
Phylogenetic studies conducted in this thesis accept the universal marker status of mtCO1 gene to distinguish species, however, markers of internal transcribed spacer region does not show universality, hence should not be used for species discriminations or phylogenetic studies (Venkatesan et al., 2016: 130-137), however, these primers can be used for taxa specific population studies as these regions are good to study interspecies/ interpopulation relationship.

A total of 15 Museum specimens collected in the current study were subjected to both full length and minibarcode region amplification. For DNA minibarcode Innuprep® Forensic kit, DNA isolation method was used extensively as this kit could recover even minute amount of DNA left in the insect tissue. All the 15 species were identified successfully and DNA barcodes were developed for both regions. All 30 (both full length and minibarcode) DNA sequences were submitted to NCBI-GenBank and accession numbers obtained for the same.

5.9 DNA minibarcoding of archival insect specimens

DNA mini-barcodes are fragments of 100-150bp region first proposed by Hajibabaei et al., (2006: 959-964). These barcodes were developed in order to identify archival or museum specimens as recovery of full length barcode from these samples are complicated due to degradation of good quality DNA (Hajibabaei et al., 2006: 959-964). A total of 15 different species of insects was obtained from the national insect repository located in Bengaluru, due to the stringency over the sharing of full specimens we could collect only one specimen of each species. A fragment of each specimen was used for DNA isolation from both Qiagen DNeasy Kit and Innuprep forensic kit, however, forensic kit produced good quality DNA as compared to Qiagen DNeasy™ blood and tissue kit due to use of Dithiothreitol (DTT). DTT degrades insect proteins faster in the presence of buffers supplied with the kit help to isolate DNA from cells and exclude thilated DNA. A total of 12 species representing order Diptera aged 2-5 years were amplified with Uniminibarcode primers. Two species representing order Hemiptera aged 6-7 years and one moth species of the order Lepidoptera aged 5years could also produced 130bp product from Uniminibarcode primer sets. Studies conducted by Meusnier et al., (2008: 1-4) could show amplification of both full length DNA barcodes from orders Lepidoptera, Diptera and Hymenoptera, however in the present study both barcodes were recovered from Lepioptera, Hemiptera and Diptera. Species from other orders could not be obtained from Museums due to restricted permissions. To
standradize this technique, samples need to be collected across 31 insect orders in order to trace variations among archival and live insects.

Phylogenetic analysis conducted by using both full length and mini barcode DNA sequences we found full length barcode DNA sequences were successful in resolving both interordinal and interspecies relationship. A fragment of 658bp product showed interordinal relationship among insect orders Diptera, Lepidoptera and Hemiptera, however, among Dipterans it could resolve species tree by separating *Spathuina ochroleuca* as out group to fruitfly group, however, minibarcode DNA sequences of 130bp showed *Spathulina ochroleuca* closely related to fruitfly group which is contradictory to the cladastic analysis of fruitflies. Hence, use of minibarcoding demands careful examination of insect taxa, however, minibarcode can be used to identify insect species, but further analysis to explain relationships among individuals of class insect, needs a strict watch to utilize this technique for phylogenetic assessments.

All 762 DNA sequences (bidirectional) representing the mtCO1 region (642 DNA sequences), ITS1 and ITS2 region (90 DNA sequences) and minibarcodes (30 DNA sequences) were checked through licensed software for quality, however, few DNA sequences obtained from ITS region were found to be of low quality and fungal DNA contaminant, hence could not be taken for further analysis. Nuclear copies of mitochondrial DNA could not be detected from any of the DNA sequences obtained for this study, however, careful examination of DNA sequences is must for DNA barcode studies.

### 5.10 Limitations of DNA barcoding

The present study has used the technique of molecular characterization DNA Barcoding extensively on insects in India. However, there are some limitations of this technique as well. DNA Barcoding is a very sophisticated tool, this requires expertise, and one must be trained on the analytical part of it. One should have proper knowledge about different kinds of bioinformatic tools and will be able to analyze DNA sequence through the tools. NUMTs Nuclear mitochondrial pseudogenes are one of the major obstacles in discriminating species on the mitochondrial DNA basis. NUMTs arise both with and without RNA intermediates. Their integration into the nuclear genome was originally associated with transposable elements or short dispersed repeats, but close examination of many different
NUMTs loci reveals lack of common features at integration sites (Bensasson et al., 2001: 315-318).

In the current study mtCO1 DNA Barcoding was employed extensively to identify arthropod species representing 10 major orders of agricultural importance, however, there are several other applications of DNA Barcoding applied in a holistic manner will fetch fruitful results in identification and management of biological resources of our planet.

5.11 Applications of DNA barcoding in species identification

DNA Barcoding is a new tool for species identification, however, it began in the year 2003 but its applicability is yet to be tested on a larger domain of life on our planet. There are several applications of this technique has already been tested. In the year 2009, Jurado-Rivera et al., (2009: 1264) identified host plant by the sequencing stomach content of an insect’s body. In another study conducted by Greenstone et al., (2005: 3247-3266), stomach contents of a carabid beetle and spider was sequenced, which helped in biological control studies. A comprehensive DNA barcode library will help in identification of insect pest at the port of entry. It will help in identifying disease vectors, it will be useful in controlling illegal trade of insects, plants and Fishes. It will help in water quality monitoring by studying organisms inhabiting major water resources. DNA barcoding will help in protecting endangered species, law enforcement can track illegal meat trade. Routine analysis of natural health products can be monitored by DNA barcoding.

The availability of DNA tools for diversity assessment will greatly facilitate and complement taxonomic studies. The combination of DNA sequencing data coupled with traditional taxonomy is a model that can be applied across disciplines and will allow analytical needs to be scaled to match the enormity of the current biodiversity crisis. DNA barcoding and molecular based identification of species will help in the identification and conservation of the evolutionary processes that generate and preserve biodiversity.

In this study, we showed that DNA Barcoding allows the rapid identification of important functional units of hyperdiverse arthropods in the rapid manner needed by conservation groups responding to habitat destruction and degradation. Insect diversity, measured via DNA Barcoding in collaboration with taxonomists, should provide the essential fine-scale maps for assessing biodiversity at a scale at which conservation decisions are made (Ojha et al., 2014: 317-330).