Comparative phylogeny of few Nymphalid butterfly species distributed across North East India using mitochondrial and nuclear marker genes

Thesis submitted in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Biotechnology

by

Catherine Vanlalruati

Ph.D Registration No: MZU/Ph.D/375 of 30.05.2011

Under the Supervision of

Dr. N. Senthil Kumar

Professor

Department of Biotechnology

Joint- Supervision of

Dr. G. Gurusubramanian

Professor

Department of Zoology

&

Dr. K. Praveen Karanth

Assistant Professor

Centre for Ecological Sciences

Indian Institute of Science, Bangalore

Department of Biotechnology

School of Life Sciences

Mizoram University Aizawl, Mizoram

CERTIFICATE

I certify that the thesis entitled "**Comparative phylogeny of few Nymphalid butterfly species distributed across North East India using mitochondrial and nuclear marker genes**" submitted to the Mizoram University for the award of a degree of Doctor of philosophy in Biotechnology by CATHERINE VANLALRUATI is a record of research work carried out by her during the period from 2011 to 2015 under my guidance and supervision, and that this work has not formed the basis for the award of any degree, diploma, associateship, fellowship or other similar titles in this University or any other University or institution of higher learning.

Signature of the Supervisor

(N.SENTHIL KUMAR)

Signature of the Co-Supervisor

(G.GURUSUBRAHMANIAN)

Declaration of the Candidate

I, Catherine Vanlalruati, a Ph. D. scholar in Biotechnology Department, Mizoram University, Aizawl, hereby declare that the subject matter of this thesis is the record of work done by me, that the contents of this thesis did not form bias of the award of any previous degree to me or to the best of my knowledge to anybody else, and that the thesis has not been submitted by me for any research degree in any other University/Institute.

This is being submitted to the Mizoram University for the degree of Doctor of Philosophy in Biotechnology.

(Catherine Vanlalruati)

Candidate

ACKNOWLEDGMENT

First of all I would like to thank the Almighty God for his guidance throughout my studies. I also want to thank my family for their love, patience, endurance and encouragement during the entire duration of my studies which ultimately led me to the completion of my thesis.

My deepest gratitude goes out to my supervisor Prof. N. Senthil Kumar. Under his guidance I successfully overcame many difficulties and learned a lot. His patience and support helped me overcome many crisis situations and finally complete this dissertation. I also thank my Co-supervisor Prof.G.Gurusubramanian and Dr. Praveen Karanth who have always been there to listen and give me advice. I am deeply grateful to them.

I am also indebted and thankful to all my friends and fellow research Scholars from both the Department of Biotechnology and Department of Zoology who squeezed in time from their busy schedule to help me complete my thesis. Sincere thanks to Dr. Syed Ibrahim, Research Associate and Mrs. BrindhaSenthil Kumar for their invaluable help.

I am also grateful to all the faculties and non-teaching staffs from the Department of Biotechnology for all their helpin whatever way possible.

Many friends have helped me stay sane throughout these difficult years. Their support and care helped me overcome setbacks and stay focused on my study. I greatly value their friendship and deeply appreciate their belief in me.

I take this opportunity to sincerely acknowledge the DBT, Govt. of India, New Delhi for the Twining project and computational facility in the form of Bioinformatics Infrastructure Facility. And I am also thankful to the UGC, New Delhi for the Rajiv Gandhi Fellowship.

CATHERINE VANLALRUATI

CONTENTS

Page no

LIST OF FIGURES

LIST OF TABLES

List of Acronyms

%	:	Percentage
°C	:	Degree Celsius
α	:	Alpha
β	:	Beta
µg	:	Microgram
µl	:	Microlitre
A	:	Adenine
AC	:	Anticodon
asl	:	above sea level
Avg	:	Average
BLAST	:	Basic Local Alignment Search Tool
bp	:	Base Pair
BSA	:	Bovine Serum Albumin
C	:	Cytosine
CBI	:	Codon bias index
Cm	:	centimeter
CO1	:	Cytochrome *c* Oxidase subunit 1
CUB	:	Codon usage bias
DNA	:	Deoxyribo Nucleic Acid
dNTPs	:	Deoxynucleotide Triphosphates
DARwin	:	Dissimilarity analysis and representation for windows
EDTA	:	Ethylene Diamine Tetra Acetic acid

EMR	:	Effective multiplex ratio
ENC	:	Effective number of codon
G	:	Guanine
GTR	:	General Time Reversal
HKY	:	Hasegawa Kishino and Yano
m	:	meter
M	:	Molar
MCMC	:	Metropolis-coupled Markov Chain Monte Carlo
MEGA	:	Molecular Evolutionary Genetic Analysis
$MgCl_2$	:	Magnesium Chloride
Min	:	minute
MI	:	Marker index
ml	:	Millilitre
ML	:	Maximum likelihood
mm	:	Millimeter
mM	:	Millimolar
MP	:	Maximum parsimony
Mya	:	Million years ago
MZU	:	Mizoram University
NaCl	:	Sodium Chloride
NaOH	:	Sodium hydroxide
NCBI	:	National Centre for Biotechnology Information
ND1	:	NADH dehydrogenase subunit 1

NTSYS	:	Numerical Taxonomy system
ORF	:	Open Reading Frame
PAST	:	PAleontological Statistics
PAUP	:	Phylogenetic Analysis Using Parsimony
PCR	:	Polymerization Chain Reaction
pH	:	Negative logarithm of the hydrogen ion concentration/ the power of hydrogen
PIC	:	Polymorphic information content
pM	:	picoMolar
PP	:	Posterior Probability
RAPD	:	Random Amplified polymorphic DNA
rDNA	:	ribosomal Deoxyribo Nucleic Acid
RFLP	:	Restriction fragment length polymorphism
RNA	:	Ribo Nucleic Acid
rpm	:	Rotation per minute
RP	:	Resolving Power
rRNA	:	ribosomal Ribo Nucleic Acid
RSCU	:	Relative synonymous codon usage
SCUO	:	Synonymous codon usage bias
SDS	:	Sodium Dodecyl Sulfate
T	:	Thymine
TAE	:	Tris Acetate EDTA
TBE	:	Tris borate EDTA

Taq	:	Thermusaquaticas
TBR	:	Tree Bisection Reconnection
TCA	:	Trichloroacetic acid
TE	:	Tris EDTA
Tris-HCl	:	Tris Hydrochloric acid
tRNA	:	transfer Ribo Nucleic Acid
ts	:	transition
tv	:	transversion
U	:	Units
U	:	Uracil
UPGMA	:	Unweighted pair group method with arithmetic mean

I. INTRODUCTION AND REVIEW OF LITERATURE

Insects comprise more than half of the world's known animal species (Wilson, 1992) of which the second largest and more diverse order is Lepidoptera of class Insecta (Benton, 1995). The name Lepidoptera, derived from the Greek words *"lepido"* for scale and *"ptera"* for wings, refers to the flattened hairs (scales) that cover the body and wings of most adults.

Classification of Butterflies

Kingdom	Animalia (Animals)
Phylum	Arthropoda (Arthropods)
Subphylum	Hexapoda (Hexapods)
Class	Insecta (Insects)
Order	Lepidoptera (Butterflies and Moths)

Butterflies (Order: Lepidoptera) offer good opportunities for studies on population and community ecology. Many species are strictly seasonal, preferring only a particular set of habitats. They are good indicators of climatic conditions as well as seasonal and ecological changes; they can help in formulating strategies for conservation. Butterflies are included in biodiversity studies and biodiversity conservation prioritization programs (Gadgil, 1996).

Several ecological characteristics make Butterflies promising biodiversity indicators: (i) their short (typically annual) life cycle makes them more sensitive to changes in their habitats compared to other groups of insect (Thomas *et al.,* 2004; Van Swaay *et al.,* 2006) (ii) breeding even in small habitat patches, they are likely to reflect changes occurring at a fine scale (iii) they may be expected to be representative for a wide range of terrestrial

habitats (Van Swaay *et al.*, 2006) and more importantly, to be adequate indicators for many groups of terrestrial insects (Thomas, 2005), which themselves constitute the predominant fraction of biodiversity.

India possesses about 1641 species of butterflies representing roughly 9.5% of the total world species (Varshney, 2006). Presently, butterflies are classified into two superfamilies, of which Hesperioidea has all the sippers, while Papilionoidea has four families: Papilionidae (Swallowtails), Pieridae (Whites and Yellows), Nymphalidae (Brush-footed) and Lycaenidae (blues). Among them, Nymphalidae are the largest group of butterfly families comprising 12 sub-families, 40 tribes and 6152 species (NSG, 2009) out of which 521 species have been described in India. Nymphalidae are also arguably the most utilized lepidopteran family in biological studies as they are distributed in various habitats worldwide and include many model species for ecological, conservation, evolutionary and developmental studies (Wu *et al.,* 2014). They are usually medium sized to large butterflies. Most species have a reduced pair of forelegs covered with long hairs and appear like brushes. The forelegs being so small are useless and nymphalids use only four of their six legs to perch and to walk and many hold their colorful wings flat when resting. They are also called brush-footed butterflies or four-footed butterflies (Kehimkar, 2008).

India separated from Gondwanaland c. 195 Ma, and finally collided with Asia in the Late Eocene. Thus, India could have acted as a raft, carrying taxa from Africa to Asia, which could spread over Southeast Asia (Roy and Karanth, 2009). During its rift it came in close contact with still northward moving Sumatra, which means that an earlier exchange of floral and faunal elements could have taken place. Possibly, during the close contact between

Sumatra and India, India became populated by Southeast Asian elements, still existing in the forests of Kerala and Sri Lanka (Turner *et al*., 2001).

Southeast Asia is an important region in terms of global biodiversity containing four of the 25 global biodiversity hotspots (Myers *et al*., 2000). These are the Island regions of Sundaland, Wallacea and the Philippines, and the mainland region of Indo-Burma. Together these regions contain an estimated 9.7% of the world's known endemic plant species and 8.3% of the known endemic vertebrate species, the majority of which are concentrated within tropical forest habitat (Brook *et al*., 2003). Research on Oriental biodiversity has been neglected relative to that of other regions, and consequently little is understood of the processes underlying the generation of diversity across the region (Sodhi *et al.,* 2004). The butterfly species are rich within the Oriental Region and occupy a wide variety of ecological niches (Zakharov *et al*., 2004; Zachariah*et al*., 2008). The biogeography of Oriental region is complicated by the fact that this region borders Australian as well as Palearctic realms. Thus, both Australian and Palearctic elements have influenced the biotic composition of the Oriental region. Additionally, a significant component of Oriental biota also has Gondwanan origin (Karanth, 2006).

The biota of tropical Asia has been placed in the Oriental realm by Wallace in 1876 and in more recent times this biogeographic zone is called the Indomalayan region (Corbet and Hill, 1992). The Indomalayan region has been subdivided into various sub regions. For example, mainland tropical Asia (Indomalayan region) is further divided into Indian and Indochinese sub regions with the borderline between these sub regions passing through Bangladesh and Northeast India. Recent studies suggest that species from Indian and Indochinese sub-regions constitute separate and independent radiations (Karanth *et al.,* 2008;

Bansal and Karanth, 2010). Additionally, the river Brahmaputra in Bangladesh and Northeast India appears to form the Eastern and Western boundaries of Indian and Indochinese sub regions respectively. Northeast India, particularly the state of Assam, forms the transition zone between these two sub regions and harbors species from both the sub regions (Mani, 1974). Interestingly, in some cases hybridization between taxa from Indian and Indochinese sub regions have been reported in Assam (Karanth, 2008). Thus, it would be interesting to determine if intraspecific phylogenies of species distributed in India as well as Southeast Asia are concordant with the sub region designation proposed by Wallace.

The Himalayan mountain range harbors the major share of the Indian butterfly diversity (Haribal, 1992). Due to richness in vegetation, north-east India is home to rich diversity of butterflies among other insects. Northeast India is one of the richest biomes of the world, high in endemism and rare species which is now under constant threat. This region is (22-30 degree N and 89-97 degree E) spread over 2, 62,379 sq. km and represents the transition zone between the Indian and Indochinese sub regions of the Oriental biogeographic region (Mani, 1974). Additionally, some elements from the Himalayan Mountains and peninsular India are also distributed here. It was the part of the northward moving 'Deccan Peninsula' that first touched the Asian landmass after the breakup of Gondwanaland in the early Tertiary Period. Northeast India is thus the geographical 'gateway' for much of India's flora and fauna. It is in this lowland-highland transition zone that the highest diversity of biomes or ecological communities can be found, and species diversities within these communities are also extremely high.

A phylogeny is a branching tree diagram showing the course of evolution in a group of organisms (Felsenstein, 1983). Phylogeny is derived from a combination of Greek words.

"Phylon" means stem and "genesis" means origin. Phylogenetic analysis is being increasingly used to address research questions in tropical ecology (Moritz *et al*., 2000). With the growing wealth of DNA sequence data at hand, molecular phylogenies of extant taxa offer the opportunity to examine the tempo and mode of speciation (Pagel, 1998). These analyses can enhance our understanding of the evolution in groups for which fossil data are lacking (Paradis, 1998), as is the case for most tropical rainforest taxa (Moritz *et al*., 2000).

The classical way of estimating the relationship between species is to compare their morphological characters (Linnaeus, 1758).Taxonomy is still largely based on morphology. Taxonomy is the science of naming, classifying and describing organisms. It groups together different organisms into taxa depending on their biological similarities. Systematics takes taxonomy one step further by using new methods and theories that can be used to classify species. This classification is based on similarity traits and possible mechanism of evolution. In the 1950s, William Hennig, a German biologist, proposed that systematics should reflect the known evolutionary history of lineages, an approach he called phylogenetic systematics. Therefore, phylogenetic systematics is the field that deals with identifying and understanding the evolutionary relationships among many different kinds of organisms. Whether morphological or molecular approach is preferable for any particular evolutionary question has been hotly debated during the last ten years (Patterson *et al.,* 1993). However, the used of molecular data for inferring phylogenetic trees has now gained considerable interest among biologists of different disciplines, and it is often used in addition to morphological data to study relationship in further detail. The increasingly available molecular information, such as nucleotide or amino acid sequences can also be used to infer phylogenetic relationships. Evolutionary relationships among genes and organisms can be illustrated using a phylogeny; it shows which genes or organisms are more closely related to each other (Lemey *et al.,*

2009).In molecular phylogeny, the relationships among organisms or genes are studied by comparing homologues DNA or protein sequences. Dissimilarities among the sequences indicate genetic divergence as a result of molecular evolution during the course of time.

Today, phylogenies are used in almost every branch of biology. It is extremely useful tools, not only for establishing genealogical relationships among a group of organisms or their parts (e.g. genes), but also for a variety of research once the phylogenies are estimated. In a recent review, Pagel (1999) outline a number of uses for phylogenetic information from discovery of drug resistance to reconstructing the common ancestor to all of life. Phylogenies have been used to predict future trends in infectious disease (Bush *et al.,* 1999) and have even been offered as evidence in a court of law (Vogel, 1997). Besides representing the relationships among species on the tree of life, phylogenies are used to describe histories of populations, the evolutionary and epidemiological dynamics of pathogens, the genealogical relationship of somatic cells during differentiation and cancer development and the evolution of language (Yang and Ranala, 2012).

Till recently taxonomists have largely used morphological characters for identification and classification of animals, plants, insects etc. It has now been established that classification of closely related lepidopteran species based on morphological features can pose several difficulties on account of attributes that can change as a function of environment and prevalence of several biotypes. These factors make morphological criteria not a preferred way for a very accurate differentiation of these species (Linares *et al*., 2009) and basic knowledge on the distribution and dynamics of genetic diversity is mainly lacking. Recently developed molecular marker techniques provide an important tool that ease the assessment of

genetic diversity and facilitate genotyping, classification, inventorying and molecular phylogenetic studies (Silva *et al.*, 2010).

Random amplified polymorphic DNA (RAPD) is a PCR based technique for identifying genetic variation. It involves the use of a single arbitrary primer in a PCR reaction, resulting in the amplification of many discrete DNA products. The technique was developed independently by two different laboratories (Williams *et al.,* 1990; Welsh and McClelland, 1990) and called as RAPD and AP-PCR (Arbitrary primed PCR) respectively. This procedure detects nucleotide sequence polymorphisms in a DNA amplification-based assay using only a single primer of arbitrary nucleotide sequence. In this reaction, a single species of primer binds to the genomic DNA at two different sites on opposite strands of the DNA template. If these priming sites are within an amplifiable distance of each other, a discrete DNA product is produced through thermocylic amplification. The polymorphisms between individuals result from sequence differences in one or both of the primer binding sites, and are visible as the presence or absence of a particular RAPD band. Such polymorphisms thus behave as dominant genetic markers.The standard RAPD technology (William *et al.,* 1990) utilizes short synthetic oligonucleotides (usually 10 bases long) of random sequences as primers to amplify nanogram amounts of total genomic DNA under low annealing temperatures. Amplification products are generally separated on agarose gels (Bardakci, 2001).

Random Amplified Polymorphic DNA (RAPD) has also been widely used for studying genetic polymorphisms and genetic diversity in populations in insects (Lery *et al.,* 2003), plants (Bolaric *et al.,* 2005), butterfly (Sharma *et al.,* 2006; 2010; Galluser *et al.,* 2004; Zakharov, 2001), mosquito (Gupta and Preet, 2014; Santos *et al.,* 2003; Dezfouli *et al*.,

2002). RAPD analysis presents some advantages over techniques targeting mitochondrial or microsatellite DNA with specific primers. Due to the use of short primers and to the low stringency at the hybridization step, the RAPD analysis requires no prior knowledge of the genome under studies, work with almost all types of DNA and usually results in several bands of different molecular weights (Williams *et al*., 1990). There being practically no limits to the number of different arbitrary primers which can be used, a wealth of information is easily obtained, such that it appears possible to refine DNA typing down to the individual level (Lery *et al*., 2003).

In order to better understanding of the genetic relationship between and within species, it is important to assess the pattern of genetic polymorphism and divergence. A suitable tool is the RAPD technique, which provides a virtually unlimited number of neutral DNA markers (Williams *et al*., 1990) and is therefore an appropriate method for initial, overall analysis of variation between populations.

RAPD-PCR technique has been used because it is cost effective, take less time, the results can be directly inferred from the gel and it reveals large amount of genetic variations, so it finds various entomological applications (Hunt and Page, 1992; Dowdy and Mcgaughey, 1996). It has been used in tracking the origin of introduced insect pests (Williams *et al*., 1994) and in differentiating ecotypes (Pornkulwat *et al*., 1998). Genetic distance was determined using intraspecific variation of RAPD alleles between geographically distinct *A. gemmatalis* populations (Gomez, 2004). RAPD markers can easily detect differences among populations and species of different organisms, including plants(Ayres *et al.*, 1999; Bartish *et al*., 1999; Bussell, 1999; Guadagnuolo *et al.*, 2001a, b),vertebrates (Vucetich *et al.*, 2001). In the identification of genetic variation and phylogenetic relationships in invertebrates (Moya

et al., 2001; Sharma *et al.,* 2010; Vanlalruati *et al.,* 2011; Pachuau *et al*., 2012; Murthy *et al*., 2014), inter-specific and intraspecific hybridization between and within species in Lepidoptera (Zakharov, 2001; Galluser *et al*., 2004; Sharma *et al.,* 2006) and differentiation of invasive mosquito species *A.albopictus* (Gupta and Preet, 2013). RAPD has also been successfully applied to study the genetic structure of endangered populations(Vandewoestijne and Baguette, 2002) and gene flow between populations (Hoole *et al.,*1999).

But there are two major and often mentioned drawbacks of RAPDs markers which is their lack of reproducibility and the loss of complete genotypic information, because of the fact that most RAPD bands are dominantly inherited (Galluser *et al.,* 2004). However, the problem of non-reproducible fragments can be highly reduced by using only high-quality DNA and by careful optimization of the PCR conditions (Wiesing *et al.,* 1995).

With the advent of modern molecular tools and data analysis programs, taxonomists are increasingly using molecular data. DNA barcoding is one such system that is designed to provide rapid and accurate species identification by using short standardized gene regions as internal species tags (Hebert *et al*., 2003). It involves rapid sequencing of one or few genes from several representatives of a species, as well as comparisons of these sequences within and between species. The method has revealed the examples of cryptic species diversity in various taxa (Hebert *et al.,* 2004a; Blaxter, 2003). DNA Barcoding holds promises especially in the identification of arthropods, the most species-rich animal phylum in terrestrial ecosystems. Identification of arthropods is often extremely time-consuming and requires taxonomic specialists for any given groups. Therefore, arthropods deserve to be considered the yard-stick for the usefulness of barcoding approaches and it is not surprising that several recent studies had applied DNA barcoding in arthropods (Barrett and Hebert 2005; Hebert *et*

al., 2004; Smith *et al*., 2005; Hajibabaei *et al*., 2006). Mitochondrial DNA has been used more frequently than Nuclear DNA due to the ease of PCR amplification and its perceived suitability, e.g. due to maternal inheritance (shorter time for coalescence than nuclear DNA (nDNA) because of smaller Ne), lack of recombination and relatively high mutation rate (Wahlberg *et al.*, 2009).

Mitochondria are key energy generators in most eukaryotic cells. Research on mitochondria primarily focused on the process of ATP generation, phylogeny and evolutionary origins. Mitochondria are believed to have evolved in eukaryotes through a process called serial endosymbiosis from an unknown microbial ancestor. An alternate theory proposed by Gray *et al*. (1999) suggests that mitochondria arose from a common ancestral extinct eukaryote, and evolved concurrently with the nucleus. Although the common mitochondrial ancestor is yet to be identified, several studies have suggested a very close relationship with endosymbionts belonging to α-Proteobacteria such as *Rickettsia* spp., *Anaplasma* spp. and *Ehrlichia* (Gray *et al.*, 1999; 2001).Irrespective of their origins, mtDNA in general appear to have lost genes and have retained identical genes (coding and non-coding) by a process commonly referred to as reductive evolution.The mitochondria have their own DNA, entirely separated from the nuclear DNA. Unlike nuclear DNA which is inherited equally from both the parents, mtDNA is inherited solely from the mother (of course some exceptions are there).The mtDNA is a histone-free, double-stranded circular molecule.

Insect mitochondrial DNA (mtDNA) consists of a circular molecule of 13–19 kb in size, with 13 protein-coding genes (PCGs) or polypeptides which are all involved in oxidative phosphorylation process: ND1-4, -4L, -5 and -6 of complex I (NADH dehydrogenase);

cytochrome b (CytB) of complex III (bcl complex); COX 1-3 of complex IV (cytochrome c oxidase) and ATP 6 and -8 of complex V (ATP synthase), two ribosomal RNA genes and 22 tRNA genes (Wolstenholme, 1992; Boore, 1999). Additionally, it contains a major non-coding area, i.e., the control region or the A+T-rich region, which regulates the transcription and replication of the mitochondrial genome (Boore, 1999; Taanman, 1999).

Mitochondrial genomes have been used for the wide array of research goals as individual mitochondrial genes, including molecular systematics (at both deep and shallow taxonomic scales), population genetics/phylogeography (Ma *et al.,* 2012), diagnostics (Nelson *et al.,* 2012), and molecular evolutionary studies (Castro *et al.*, 2002; Salvato *et al.*, 2008 and Shao *et al.,* 2003).And it is also used to study animal, bird, butterfly and human migration. In addition, wholegenome sequencing allows the study of comparative and evolutionary genomics questions, such as the frequency and type of gene rearrangements (Cameron *et al.,* 2011; Dowton *et al.,* 2009) and the evolution of genome size.

Cytochrome oxidase is one of a superfamily of proteins which act as the terminal enzymes of respiratory chains. It is the component of the respiratory chain that catalyzes the reduction of oxygen to water. Subunits 1-3 form the functional core of the enzyme complex. CO I is the catalytic subunit of the enzyme. Electrons originating in cytochrome c are transferred via.the copper A center of subunit 2 and heme A of subunit 1 to the bimetallic center formed by heme A3 and copper B.

Hebert *et al*. (2003) proposed that a DNA barcoding system for animal life could be based upon sequence diversity in cytochrome c oxidase subunit 1 (COI). The "DNA barcode" itself, as proposed by Hebert *et al* (2003a), is a small gene region and consists of 648 bp from

the 5'-end of the *cytochrome c oxidase 1* (COI) mitochondrial DNA gene was sufficient to reliably place species into higher taxonomic categories from phyla to orders (Brandao *et al.*, 2009). They also found that diversity in nucleotide sequences of the same gene region regularly permitted the discrimination of closely allied species of lepidopteran, a group with modest rates of molecular evolution and high species diversity. As such, these insects provided a challenging test for the ability of COI diversity to resolve species boundaries.

COI gene of mitochondria does have two important advantages: First, the universal primers for this gene are very robust, enabling recovery of its 5'end from respective of most, if not all, animal phyla. Second, COI appears to possess greater range of phylogenetic signal than any other mitochondrial gene. Along with other protein coding genes, its third-position nucleotides show a high incidence of base substitutions, leading to a rate of molecular evolution that is about three times greater than that of 12S and 16S.The evolution of this gene is rapid enough to discriminate not only closely related species, but also phylogeographic groups within a single species. Below species level, COI and other mitochondrial DNA are chosen as markers insome species because of its high mutation rate (Vandewoestijne *et al*., 2004). At the same time COI is the most slowly evolving gene of the mitochondrial protein coding genes (Simon *et al.,* 1994).

In some animal groups, CO1 has failed to deliver reliable DNA barcodes. In cnidarians and sponges, CO1 divergences are extraordinarily low compared with bilaterian animals (Shearer *et al.,* 2002; Park *et al.,* 2007). On the other hand, in aves, gastropods and amphibians, interspecific variation and intraspecific variation in CO1 are very high (Remigio and Hebert, 2003; Hebert *et al.,* 2004b). In 449 dipteran species, the identification success

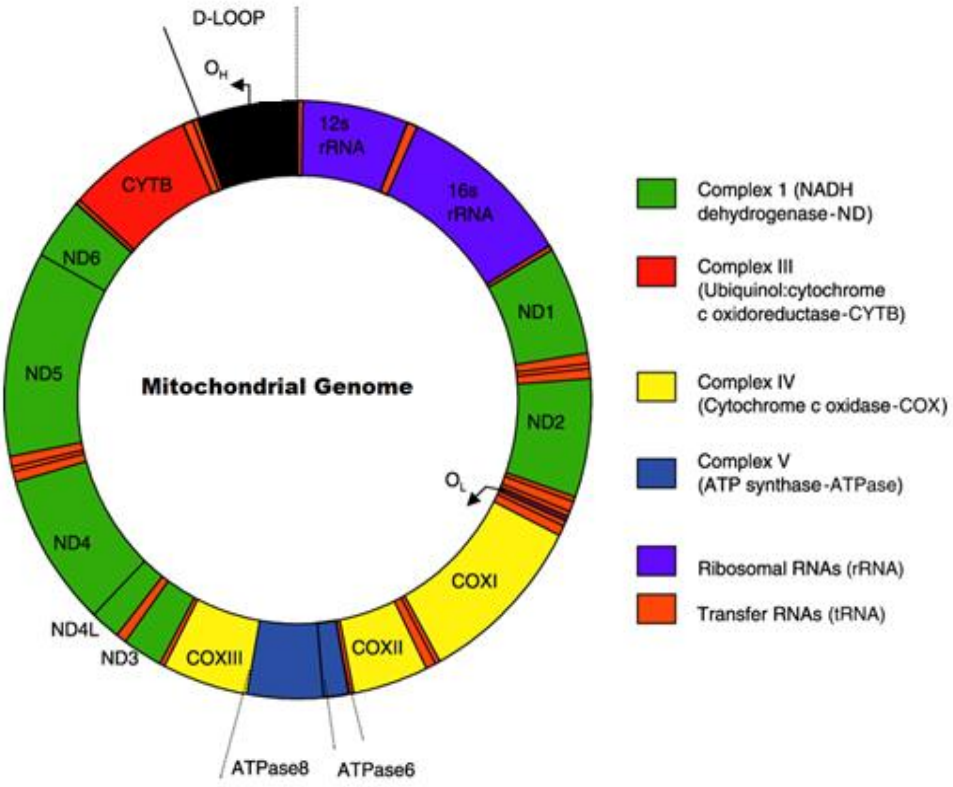


Figure 1: Insect Mitochondrial genome

through CO1 ‘barcodes’ was low due to substantial overlaps in inter- and intraspecific divergences. Moreover, it was shown that the vast majority of nucleotide substitutions within the CO1 fragment occur at the third codon position, which might lead to rapid saturation (Vences *et al.,* 2005).

However, COI gene has proved to be suitable for species identification in a large range of animal taxa, including butterflies and moths (Hebert *et al.,* 2004a; Janzen *et al.,* 2005; Hajibabaei *et al.,* 2006; Burns *et al.,*2008; De-Mandal *et al.,* 2014), birds (Hebert *et al.,* 2004b; Kerr *et al.,* 2007), mayflies (Ball *et al.,* 2005), spiders (Greenstone *et al.,* 2005), fishes (Ward *et al.,* 2005), ants (Schlick-Steiner *et al.,* 2006), Crustacea (Costa *et al.,* 2007), gastropods (Remigio and Hebert, 2003), mosquitoes (Kumar *etal.,* 2007), and wasps (Smith *et al.,* 2008). The efficacy of COI based barcoding is also documented for few other kingdoms like fungi (Seifert *et al.,* 2007), macroalgae (McDevit and Saunders, 2009) and ciliates (Chantangsi *et al.,* 2007).

Mitochondrial Subunit ND1 (mtND1) gene is involved in the first step of the electron transport chain of oxidative phosphorylation. In general, NADH dehydrogenase subunits are rapidly evolving at the amino acid level than the cytochrome oxidase subunits (Simon *et al*., 1994). NDI+tRNA are suitable for resolving relationship among closely related taxa in coleopteran: Chrysomelidae (Stapel *et al.,* 2008).

In insects, the mitochondrial NADH dehydrogenase subunit 1 (ND1) gene region has proved to be another suitable marker especially for the identification of lower level taxonomic entities such as populations and sister species. The mitochondrial ND1 (NADH dehydrogenase 1) gene region, for example, showed better performance than CO1 in

resolving phylogenetic relationships especially in insects such as in aphids (Lin and Danforth, 2004), in Hawaiian drosophilids (Baker and DeSalle, 1997) and odonates (Hadrys *et al*., 2006; Dijkstra *et al*., 2007 and Rach *et al.*, 2008). In mammals, the estimated variability in ND1 is slightly higher than in CO1 (Saccone *et al.*, 1999).But study by Weller *et al*.(1996) using characters from a portion of the ND1 gene does not recover the monophyly of, or any resolution within, the nymphalids. These findings suggest that the inability to resolve all families at once may not be due to the shortcomings of a particular gene itself, but might instead be due to biological differences in the radiations of these lineages, or a different rate of molecular evolution in the Nymphalidae. At the same time, ND1 has been successfully applied to phylogenetic and population genetic studies in Odonates and seems to be well suited as an alternative or complement to CO1 (Rach *et al.*, 2008). It has also been used for studying the evolutionary relationship in Butterflies (Aubert *et al.*, 1999; Martin *et al.*, 2000; Zimmermann *et al*., 2000), fruit fly (Wu *et al.*, 2014).

Both ND1 and CO1 are suitable DNA barcoding markers and deliver reliable character-based DNA barcodes for the vast majority of species. However, neither one alone could resolve all species. Combining both markers is highly beneficial for discriminating species, in particular sister species as well as geographical entities. It cannot be predicted which marker delivers the higher degree of information in which species. It shows that using more than one marker is the best for studying the phylogeny or relationship between and among organisms (Bergmann *et al*., 2013).

Mitochondrial protein-coding genes seem to meet the best criteria as markers for studying phylogeny because of several reasons: (i) high copy numbers per cell (Hoy, 2003; Avise, 2004) generally enhance PCR amplification (Lin and Danforth, 2004); (ii) the haploid

character allows the direct sequencing of PCR products (Saccone *et al.*, 1999; Hurst and Jiggins, 2005); (iii) the lack of introns, rare occurrence of indels (Hebert *et al.*, 2003a) and low recombination rate ease the alignment; and (iv) the lack of proofreading mechanisms leads to higher evolutionary rates than in nuclear genes (Hoy, 2003).

The full-length mitogenome sequences have phylogenetic utility within several insect lineages including Lepidoptera (Kim *et al*., 2014). Mitogenomic data are effective for revealing higher-level relationships of diverse animal groups. Although Sanger sequencing has been used to obtain high-quality mitogenomic sequences, the time required for primer design, and the cost invested to recover large numbers of mitogenomic sequences remains challenging. Recently, next-generation sequencing (NGS) methods have been shown to overcome such shortcomings. This makes practical the task of re-examining and re-evaluating phylogenies with much larger datasets (Wu *et al*., 2014).So far, about 500 species of mitogenomic sequenced have been recorded for insect species (Cameron, 2014).In the second largest insect order Lepidoptera, there were about 172 mitogenome sequences in NCBI.

Nymphalinae (Lepidoptera: Nymphalidae) comprises about 500 species and is distributed nearly all around the world (Harvey, 1991). However, the taxonomy and systematics of Nymphalinae are still standing as a controversial issue, and waiting for further investigations (Shi *et al*., 2013). Currently, only five complete metagenomes of Nymphalinae have been reported, which were *Kallima inachus*, *Melitaea cinxia*, *Junonia orithya*, *Junonia almana* and *Yoma sabina*.

Estimation of synonymous and nonsynonymoussubstitution rates is important in understanding the dynamics of molecular sequence evolution (Kimura, 1983; Gillespie, 1991; Ohta, 1995). As synonymous (silent) mutations are mostly invisible to natural selection, while nonsynonymous (amino- acid-replacing) mutations may be under strong selective pressure, comparison of the rates of fixation of those two types of mutations provides a powerful tool for understanding the mechanisms of DNA sequence evolution. Models of variable non-synonymous/ synonymous rate ratios among sites may provide important insights into functional constraints at different amino acid sites and may be used to detect sites under positive selection (Nielsen and Yang, 1998).Study in codon usage bias is important for understanding the evolution of gene or genome among closely related individuals. Major causes that change the codon bias in a gene or genome are expression level, gene length, composition bias, recombination rate (Powell *et al.*, 1997; Sharp, 1994). Insects have an exceptional diversity within animal kingdom and the exponential increase of sequence data makes it an ideal choice for studying codon usage bias.

Relative Synonymous Codon Usage (RSCU) is a simple measure of non-uniform usage of synonymous codonsin a coding sequence. An RSCU value for a codon is simply the observed frequency of that codon divided by the frequency expected under the assumption of equal usage of the synonymous codons for an amino acid (sharp and Li, 1986; 1987). In the absence of any codon usage bias, the RSCU values would be 1.00. A codon that is used less frequently than expected will have an RSCU value of less than 1.00 and vice versa for a codon that is used more frequently than expected.

Restriction Fragment Length Polymorphism (RFLP) is a method of indirect sequence detection. By using enzymes which cuts the DNA at specific sites, we can detect whether

individual DNA sequences (usually PCR products) possess the restriction site or not. Fragments can be separated by electrophoresis on an agarose or a polyacrylamide gel. By quantifying fragment length, we can also estimate the relative position of restrictions sites to each other. RFLP is an easy and quick method of surveying genetic variation; though much of the sequence information is lost in the procedure.RFLP can be applied to a range of problems from population genetics to species identification to phylogenetic. Because of the dropping costs of sequencing, it is these days mainly applied if a larger number of samples have to be surveyed. Most RFLP markers are co-dominant (both alleles in heterozygous sample will be detected) and highly locus-specific.

PCR-restriction fragment length polymorphism (RFLP)-based analysis, also known as cleaved amplified polymorphic sequence (CAPS), is a popular technique for genetic analysis. It has been applied for the detection of intra-species as well as interspecies variation (Rasmussen, 2012). It is a technique that exploits variations in homologous DNA sequences.The DNA sample is digested by restriction enzymes and the resulting restriction fragments are separated according to their lengths by gel electrophoresis (Botstein *et al.*, 1980).

Nakamura *et al*. (1998) reported that PCR-RFLP analysis is more useful in detecting DNA polymorphism in objective sequences than through sequencing analysis. The analysis of restriction fragment length polymorphism (RFLP) of PCR fragments was already successfully applied for species differentiation (Meyer *et al.,* 1995). PCR-RFLP of different genes was demonstrated to detect inter- and intraspecific variations in several animals such as Atlantic snapper (Chow, 1993), tuna (Chow and Inogue, 1993), lobsters (Silberman and Walsh, 1992), turtles (Karl *et al*., 1992).

The Hoxgenes are an ancient family of developmentalregulatory genes that, in arthropods, are differentially expressed along the anterior/posterior axis of the body to define tagmosis and many finer details of segment organization (Akam, 2001). *Hox* genes have been conserved since theearly divergence of the bilaterian animals, and comparisons of the sequences and functions of Hoxgenes can reveal evolutionary relationships (deRosa *et al.*, 1999).

Hox genes development are amenable to detailed characterization ranging from the genetic pathways involved in establishing the pattern, to the molecular and cellular interactions underlying pattern specification or about the extent to which they contribute to phenotypic variation in wing colorization or eyespot morphology or fore-leg formation (Beldade *et al*., 2002). The regulatory activities of the Hox proteins play a key role in determining segmental morphologies along the insect body, including the presence or absence of limbs (Lewis, 1978; Kaufman *et al*., 1980). The development of distal limb structures in arthropods is controlled by a Hox regulatory target, the *Distal-less* (Dll) gene (Cohen *et al.*, 1989; Panganiban *et al*., 1995; Schoppmeier and Damen, 2001). The Dllgene is expressed in the primordia, and later in the distal regions of the developing limbs of all arthropods (Panganiban*et al.*, 1997).

In *Drosophila melanogaster*, limbnumber and type are controlled by homeotic and limb patterning genes *Distal-less* (Dll*)* and it is expressed specifically within the primordia of cephalic and thoracic limbs and is excluded from the abdomen by the action of the homeotic genes of the bithorax complex (William and Carroll, 1993; Cohen, 1990). The gene *distal-less* (dll) shows expression at the tip of most protruding structures, such as limbs, antennae,

and setae, in a variety of Lepidoptera, as well as in other insects and invertebrates (Carroll *et a1.*,1994; Panganiban *et a1.,* 1995).

In order to reconstruct the biotic history of this region, it is imperative that we construct the evolutionary history (phylogeny) of species distributed therein. Unfortunately, there is very little phylogenetic work done on the Oriental biota as a whole but quite a few studies have been done on smaller areas such as in the Western Ghats and the Sunda shelf. In this regard, butterflies are an ideal system to address questions pertaining to biogeography of the Oriental region. Butterflies are therefore a good model for studying the distribution of biodiversity and the factors influencing population divergence and speciation across the region.So far, to our knowledge, there has been no published work related to Butterfly diversity from Northeast India.

In the present study, Nuclear markers (RAPD) and Mitochondrial genes (COI and NDI) have been used to determine whether the river Brahmaputra acts as a barrier for gene flow among the selected butterflies species or if hybridization occur among them. Also, in our studies, we sequence the first complete mitogenome sequence of *Junonia iphita*, and compared its sequence to other Nymphalidae metagenomes, in order to provide more useful information for the taxonomic and phylogenetic studies of Nymphalidae butterflies in the Northeast Indian region.

OBJECTIVES OF THE STUDY

1. Collection, morphological analysis and comparative phylogeny of selected species of Nymphalidae butterflies across North East India.

2. To compare the Polymorphism within and between populations of butterfly in north and south Brahmaputra river using molecular markers (RAPD, PCR-RFLP, Cytochrome c Oxidase I and NADH Dehydrogenase 1).

3. To characterize the *hox* gene (distal-less (*dll*) in relation to their physicochemical properties and evolutionary relationships in the selected species of Nymphalids.

III. MATERIALS AND METHODS

3.1. Survey and sample collection across Northeast India

Butterflies were collected from different parts of Northeast India by direct searching, visual observation and sweep netting method with opportunistic sample collection. In this methods target species or group of species were observed visually and it was a very effective method for day flying butterflies, moths, dragonflies and damselflies (Sutherland, 1996).Modification of transect method described by Pollard (1977) in sampling design was used. The adult butterflies were collected using butterfly net and kept in a small and dry envelopes and were later stored in a fridge. The legs were taken and stored in 70% alcohol for further use in the molecular experiments. Identification of the butterflies wasdone based on the informationgiven by Evans (1932), Wynter-Blyth (1957), Wahlberg (NSG, 2009) and Isaac Kehimkar (2008).

Butterflies are collected from different parts of Northeast India viz., Assam, Sikkim, Arunachal Pradesh in the North Brahmaputra and Mizoram, Meghalaya, Tripura in the South of Brahmaputra from April 2010 to May 2013.Samples were collected from different parts of Northeast India based on four criteria:a) Species should be distributed only in the Oriental region. b) They must be easy to identify. c) They should not exhibit local migration d) must be locally common for ease of collection. Based on the survey and preliminary studies, five species were selected: *Junonia almana, Junonia iphita, Junonia atlites*, *Elymnias hypernmestra* and *Ariande merione* across Northeast India.

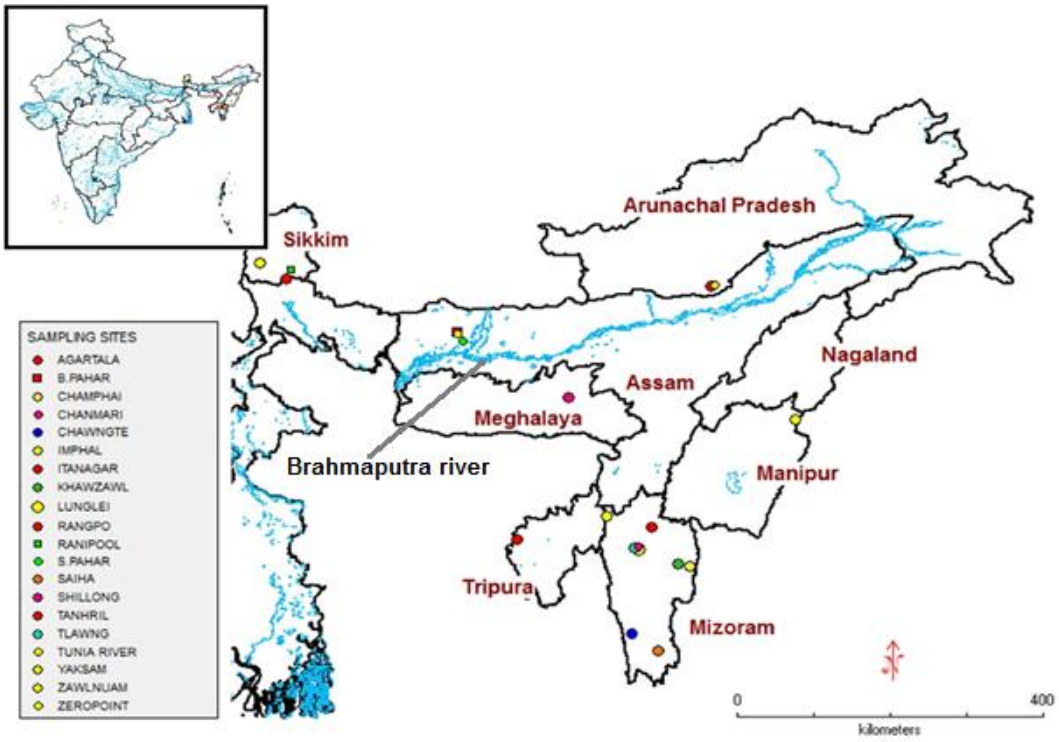


Figure 2: Map showing different collection sites across Northeast India

Table 1: Different sampling sites with coordinates

Different States	Collection sites	Coordinates
Mizoram	Aizawl	23˚44′33″N 92˚43′25″E
	Saiha	22˚29′40″N 92˚58′50″E
	Zawlnuam	23˚37′37″N 92˚17′35″E
	Lunglei	22°53'51"N 92°45'15"E
Meghalaya	Barapani	25˚39′09.9″N 91˚52′45″E
	Nehu Campus	25°36'42"N 91°53'57"E
Tripura	Agartala	23˚50′20″N 91˚16′0″E
Manipur	Imphal	24°47'24"N 93°56'24"E
Assam	Bongaigaon	26°28'48"N 90°32'24"E
Sikkim	Ranipool	27°17'36"N 88°35'25"E
	Rangpo	27°10'48"N 88°31'48"E
	Yuksam	27°22'10"N 88°13'14"E
Arunachal Pradesh	Itanagar	27˚5′13″N 93˚37′19″E

3.2. Diversity studyof Nymphalidae in Mizoram

The diversity studies of Nymphalidae family on Mizoram were done with respect to three forest type of Mizoram which were Tropical Wet Evergreen Forest, Tropical Semi-Evergreen Forestand Mountain Sub-Tropical Forest. The numbers of species sighted during the survey between 2010 and 2013 were counted in the sampling areas and the Butterflies (Family:Nymphalidae) were categorized on the basis of their abundance in the different forest types. During the surveyed period between 2010 and 2013, monthly average rainfall, temperature and humidity in each forest types of Mizoram were recorded (Table 2).

3.2.1Statistical analysis

Shannon and Simpson diversity indices were calculated as a measure of diversity in eachforest types. Berger-Parker dominance has also been evaluated for each forest types. Mean and standard deviation of butterfly's abundance from 2010 to 2013 was recorded. All the calculations were performed using commercially available GraphPad InStat version 3.06 software (GraphPad Software Inc., San Diego, CA), PAST 1.86b (Hammer *et al*., 2001) and Biodiversity Pro software (Mcaleece *et al*., 1997).A p-value of <0.05 was noted to have statistically significant value.

Table 2: Monthly average temperature, humidity and rainfall in three forest types of Mizoram during 2010-2013.

Foresttype1	**Jan**	**Feb**	**Mar**	**Apr**	**May**	**June**	**July**	**Aug**	**Sept**	**Oct**	**Nov**	**Dec**
Temperature(°C)	19	19.15	23.7	24.4	24.2	24.02	25	30.38	27.1	26.3	23.2	17.1
Humidity (%)	58.3	51.1	52.3	52.3	85.5	95.7	97	92.4	90.1	85.7	79.9	73
Rainfall (mm)	13.2	11.22	92.5	92.5	398.33	479.17	288.33	454.64	562.7	315.3	15.87	20
Forest type2												
Temperature(°C)	18	20.16	22.92	24.13	23.6	23.75	24.56	24.8	23.99	23.99	21.25	18.52
Humidity (%)	61.53	54.52	54.33	73.24	82.29	89.09	89.37	90.05	81.7	81.7	75.03	64.73
Rainfall (mm)	11.53	6.23	94.66	253	379.54	493.33	372.37	447.83	1168.6	688.3	71.266	19.33
Forest type3												
Temperature(°C)	14.5	16.95	21.7	23.25	23.75	22	22.75	23	22	21	18.75	18
Humidity (%)	40.4	49.1	48.1	69.1	81.4	90.6	90.8	90.1	86.8	80.7	67.2	47.1
Rainfall (mm)	10.63	11.3	75	198.4	228.67	286.57	339.5	330.58	801.67	389	39.63	17.47

Forest type 1: Tropical Wet Evergreen Forest

Forest type 2: Tropical Semi-Evergreen Forest

Forest type 3: Mountain Sub-Tropical Forest

Sources: Directorate of Economics & Statistics and Directorate of Science & Technology, Aizawl, Mizoram.

Shannon index

$$H' = \sum pi \ln pi$$

Where, H' = Shannon index of diversity, *p*i = the proportion of important value of the ith species (*p*i = *n*i / N, *n*i is the important value index of ith species and N is the important value index of all the species).

Shannon's equitability (J) can be calculated by dividing H by Hmax (here Hmax = lnS).

Simpson index

$$D = \sum \left(\frac{ni(ni-1)}{N(N-1)} \right)$$

Where $n_i$is the number of individual in the *ith* species

And *N* = total number of individuals. As *D* increases, diversity decreases. Simpson's index is therefore usually expressed as 1-*D* or 1/*D*.

Berger-Parker index

The Berger-Parker index, *d*, is a simple dominance measure. It expresses the proportional abundance of the most abundant species.

$$d = Nmax/N$$

Where *Nmax* = the number of individuals in the most abundant species.

3.3. Extraction of Genomic DNA

DNA was extracted from the leg which was stored in 4°C prior to extraction. The extraction method followed modified protocol of Zimmermann *et al.*(2000). The leg was washed with distilled water and dried. The leg was macerated with the help of scissor in 1.5ml eppendorf tube, the leg tissues were homogenized with pestle and 250µl of extraction buffer (50mM Tris HCl, 25mM NaCl, 25mM EDTA,1% SDS) was added and mixed gently. 2µl of proteinase *k* (20mg/ml) was then added and incubated in an oven at 56°C for half hour. To this, 250µl of phenol/chloroform (1:1) was added and mixed gently and centrifuged at 13,000 rpm for 5 minutes. Supernatant was carefully taken out and collected in a new eppendorf tube. 450µl of absolute ice cold ethanol was added to the supernatant and then mixed gently by inverting the tube several times and then kept in 20°C for 30 mins. The tube was then centrifuged at 13,000 rpm for 5 mins at 4°C. Ethanolwas poured off without dislodging the pellet, and200µl of 70% ethanol wasadded, flash spin at 6000 rpm for 1 mins. The ethanol was poured off and the pellet was dried in an oven for 5 mins. 30µl of distilled water was added to the tube; the pellet was re-suspended by gently flicking the tube and was stored at 20°C for further use.

3.4. Genomic DNA profiling through RAPD-PCR

The Genomic DNA was amplified using RAPD procedure of Gallusser *et al.*(2004) with modification. Five random primers were used(Table 6). 25µl of reaction mixture contained: 1X amplification buffer, 3mM MgCl, 2mM dNTPs, 1U of Taq polymerase, 0.8µl of BSA and 2µl of template DNA. The amplification was carried out in thermal-cycler Gradient (Eppendorf, Germany) using the following condition. Initial Denaturation at 94°C

for 5min followed by 35 cycles, 1min at 94°C for Denaturation, 1min at 37°C for annealing and 1min at 72°C for extension and final extension at 72°C for 5min. The amplified products were stored at 4°C.

3.5.PCR amplificationof CO1 Gene

PCR was performed with universal primers described in Hebert *et al.*(2003) forward primer LCO 5' - TAA TAC GAC TCA CTA TAG GGG GTC AAC AAA TCA TAA AGA TAT TGG -3' and reverse primer HCO 5'- ATT AAC CCT CAC TAA AGT AAA CTT CAG GGT GAC CAA AAA ATC A -3'. The 25µl reaction mixed contained: 1X amplification buffer, 2.5mM $MgCl_2$, 0.25mM dNTPs, 0.2pM each primer, 0.8µl BSA, 2µl Genomic DNA and 1U Taq DNA polymerase. The PCR thermal regime for amplification was 5 min at 95°C for initial denaturation, followed by 30 cycles of 30sec at 95°C for denaturation, 40sec for annealing at 51°C – 54°C, elongation for 30sec at 72°C and a final elongation for 6min at 72°C.

3.6.PCR amplification of ND1 Gene

PCR was performed with specific primers, ND1 (fwd) 5'- TTC AAA CCGGTG TAA GCC AGG -3' and ND1 (rev) 5'- TAG AAT TAG AAG ATC AAC CAG-3' described by Rach *et al*. (2008). The 25µl reaction mix contained: 1X amplification buffer, 2.5mM $MgCl_2$, 0.25mM dNTPs, 0.2pM each primer, 0.8µl BSA, 2µl Genomic DNA and1U Taq DNA polymerase. The PCR thermal regime for amplification was: 5 min at 95°C for initial denaturation, followed by 30 cycles of 30sec at 95°C for denaturation, 40sec for annealing at 51°C -54°C, elongation for 30sec at 72°C and a final elongation for 6min at 72°C.

3.7. PCR-restriction fragment length polymorphism of COI gene

Amplified PCR (COI) product was subjected to restriction enzymes digestion with*TaqI* (T/CGA), *AluI* (AG/CT) and *RsaI* (GT/AC) (Fermentas, Thermo Scientific) enzymes as per the protocol of Sambrook *et al.* (1989). 10µl reaction volumes contained: 0.2/0.3µl of restriction endonuclease, 1µl endonuclease buffer, 3µl of PCR product and 5.8/5.7 µl of sterile water. Then incubated or digested for 2 hour with the enzymes *Alu1*, at 37˚C, *Taq1* at 65˚C and 6 hour at 37˚Cfor *RsaI.*After the incubation it is then deactivated by adding 1µl of loading dye which contain 1% EDTA which will stop the enzymes reaction. Restriction fragments were separated on 12% acrylamide gel, visualized and photograph using UV trans-illuminator.

3.8. PCR amplification of Dll (Distal-less) Gene

We amplified short fragment of the Dll gene from genomic DNA of *Elymnias hypernmestra* using degenerate primers design to match the conserved region of Dll gene in insects from database.Dll(fwd)5'-CSTTCRTVGAGYTRCAGCA-3'and(rev)5'-GCVGCCTTCATCATCTTYTTG-3'. The 25µl reaction mixes contained: 1.2X amplification buffer, 2mM $MgCl_2$, 0.25mM dNTPs, 0.35pM each primer, 0.4µl gelatin, 2µl genomic DNA and 1.5U Taq DNA polymerase. The PCR thermal regime for amplification was 5 min at 95°C for initial denaturation, followed by 30 cycles of 40sec at 95°C for denaturation, 40sec for annealing at 51°C- 54°C, elongation for 50sec at 72°C and a final elongation for 5min at 72°C.Dll genes were unambiguously identified by alignment with previously published sequences from NCBI.

3.9. Agarose gel electrophoresis and sequencing of the PCR products

The agarose gel electrophoresis followed the protocol described by Sambrook *et al*.(1989).To visualize the different PCR products (2µl of COI, NDI and Dll and 10µl of RAPD PCR products) was loaded in 1.5% agarose gel, 1X TAE pH 8 (Tris-acetate-EDTA electrophoresis buffer)and ethidium bromide (10mg/ml final concentration). The gel was run at 100V for 1-2 hours and visualized with UV trans-illuminator and photographed using Syngene G: BOX, Germany. PCR amplicons were estimated usinglow range ruler plus or 100 base pair DNA ruler(GeNei).The PCR products were sequenced using Sanger's di-deoxy method and sequencing reactions were carried out in one direction on a sequencer (SciGenom Labs Pvt Ltd., Cochin, India). All the sequences were checked using BLAST (NCBI). Sequences were aligned and checked using Pairwise Sequence Alignment (EMBOSS-water, EBI) and FinchTV version 1.4.0 (Patterson *et al*., 2006) followed by manual adjustments. All the protein coding sequences were translated into amino acids and their ORFs checked (ORF Finder (NCBI)).

3.10. Polyacrylamide gel electrophoresis for PCR-RFLP products

The polyacrylamide gel electrophoresis was done following Sambrook *et al*. (1989). 12% gel was made from 30% acrylamide,10% ammonium persulfate and 10X TBE buffer and run in 1X TBE buffer (Tris-base, Boric acid, EDTA). The gel was run at 100V for 1-2 hours and visualized with UV trans-illuminator and photograph using Syngene G: BOX, Germany. PCR amplicons were estimated using low range ruler plus or 100 base pair DNA ruler (GeNei).

3.11. Data Analysis

3.11.1 Analysis of data produced by RAPD-PCR and PCR-RFLP profiling

For the analysis and comparison of RAPD patterns and PCR-RFLP, a set of distinct, well separated bands were selected. The molecular weight of the bands on gels was documented using the Gel documentation system (UVItech Reader, Cambridge). Only distinct and polymorphic bands were used in the analysis. Bands were scored as 1 for the presence or 0 for the absence of band for all of the individual samples. The binary matrix was used to estimate genetic similarities using Jaccard's coefficients.For RAPD data the similarity matrix was subjected to unweighted pair group method of arithmetic averages clustering in order to construct the phenetic dendrograms using UPGMA employing the SAHN algorithm with 1000 bootstrap value. All parameters were computed using NTSYS-pc 2.01i (Rohlf, 1998) and DARWIn 5 (Perrier and Jacquemoud-Collet 2006).

Genotyping data from RAPD-PCR and PCR-RFLP were used to evaluate three parameters, namely, (1) polymorphism information content (PIC), (2) marker index (MI), and (3) resolving power (RP). The polymorphic information content of each RAPD-PCR marker was computed as $PIC_i=2_{fi}$ (1-*fi*), where f is the frequency of the amplified allele (band present) and (1-fi) is the frequency for null allele. Only polymorphic bands were used for calculating PIC values. (Roldan-Ruiz *et al*., 2000).The MI was calculated using the formula, MI=PIC × EMR, where effective multiplex ratio (EMR) is the total number of polymorphic fragment per primer. Resolving power of each primer was calculated using the formula, $RP=\Sigma I_{b,}$ where I_b represent band informativeness expressed as $I_b=1-(2\times|0.5-P|)$ where p is the fraction of the total accession in which the band is present. (Prevost and Wilkinson,

1999).By treating each specific RAPD multiband as a distinct haplotype, each species were defined in terms of haplotype identity. Analysis of variance was done to partition RAPD variation into within and among population components using software PAST (Hammer, 2001).

3.11.2 Analysis of COI, NDI sequences using MEGA 5.1

Phylogenetic and molecular evolutionary analyses were conducted using MEGA version 5.1 (Tamura *et al*., 2011). The sequences of CO1 and ND1 from each species were aligned separately using ClustalW implemented in the program MEGA5.1. Once the alignment was done,nucleotide composition, analysis of various sites such as conserved sites, parsimonious informative sites, variable sites, singleton, zero, two and four fold degeneratives sites are performed. The evolutionary distances were computed using the p-distance method and are in the units of the number of base differences per site. The analysis involved 32, 30nucleotide sequences for COI, NDI genes. All positions containing gaps and missing data were eliminated. There were a total of 653 nucleotides in the final dataset of the CO1, 389nucleotides forNDI and 1042 nucleotides for combined data of COI and NDI genes for analysis. RSCU value and maximum composite likelihood estimates of transition transversion bias of COI and NDI genes were also calculated using Mega version 5.

3.11.3 Codon usage analysis

Codon usage analysis like codon bias index (Morton, 1993), the effective number of codons (Wright, 1990), and G + C content at second and third positions as well as overall were calculated for each gene using DnaSPv5 (Librado and Rozas, 2009). The effective number of

codons (ENC) was estimated to quantify the codon usage bias (CUB). Comparison of synonymous codon bias and GC difference was carried out by calculating SCUO (synonymous codon usage order) (Wan *et al.,* 2004) among COI and NDI sequences using CodonO (http://sysbio.cvm.msstate.edu/ CodonO/). The ratio of the number of no synonymous nucleotide substitutions per site (dN) to that of synonymous nucleotide substitutions (dS) based on a set of aligned COI and NDI sequences had been performed to investigate the selection pressure among the two genes by using online tool SNAP v1.1.1 (Korber, 2000).

3.11.4 Phylogenetic Inference

Phylogenetic relationships were inferred using Bayesian inference (BI) and Maximum Parsimony (MP) for COI, NDI and Combined COI+ NDI datasets. jModeltest2.1.3 (Darriba *et al*., 2012)was used to determine substitution models according to Akaike Information criterion, being GTR+I+G for COI, HKY+I+G for NDI and combined COI+NDI datasets. For the whole mitogenomes analysis Maximum parsimony was perform using PAUP*, Maximum likelihood (ML) analysis was performed with RaxML (raxmlGUI, Silvestro and Michalac, 2012) and Bayesian Inference using MrBayes. Phylip file was generated for RaxML analysis using ALTER Alignment Transformation Environment (Glez-Peña *et al*., 2010). For ML analysis the bootstrap was set at 500 and the model was set at GTRGAMMAI**.**Maximum parsimony (MP) trees were obtained using PAUP* (Swofford, 2002) byheuristic search option with tree-bisection-reconnection (TBR) branch-swapping. The number of bootstrap replicates was set at 1000. Starting tree was obtained via stepwise addition and the number of trees held at each step during stepwise addition equals 1. All characters were equally weighted, and zero length branches were collapsed to polytomies.

Multistate taxa were interpreted as uncertainty, topological constraint was not enforced and the generated 50% consensus trees were saved as .tree file.The Maximum likelihood (ML) analyses for COI and NDI were conducted using PHYML (Guindon *et al.,* 2010) under the following conditions: the proportion of invariable sites as "estimated", number of substitution rate categories as four, gamma distribution parameter as "estimated", the substitution model used was GTR for both the gene COI and NDI and the starting tree as a BIONJ distance-based tree. The confidence values of the ML tree were evaluated via the bootstrap test with 500 iterations.

The Bayesian Inference (BI) analyses was conducted using MrBayes ver. 3.1(Ronquist and Huelsenbeck, 2003). Two independent runs of four incrementallyheated MCMC chains (one cold chain and three hotchains) were simultaneously run for eight million generationsdepending on the dataset, with sampling conducted every 100 generations. The convergence of MCMC, which was monitored by determining the average standard deviation of split frequencies, was achieved (< 0.01) within eight million generations depending on the dataset, and the first 20% of the sampled trees were discarded as burn-in. The confidence values of the BI tree are presented as the Bayesian posterior probabilities in percentages (BPP).with the partitioned strategy. Starting trees were random.Convergence, ESS values, and burn-in were assessed with Tracer v1.5 (Rambaut and Drummond, 2007).

3.12.Sequencing, genome characterization and Phylogenetic Analysis of whole mitochondrial genome of *Junonia iphita.*

Sequencing was performed by using paired end Illumina Miseq platform. All the raw sequences were analysed based on base quality score (> 30), average base content per reads and GC distribution followed by filtered Illumina adapters using Cutadapt (Martin, 2011). Assembly was done by using *Junonia almana* (NCBI accession: NC_024407) as a reference in MITObim (Hahn*et al*., 2013).The assembly resulted in a single contig. The locations of the 13 protein-coding genes and two rRNA were initially identified by DOGMA and mitos webserver (Wyman, 2004; Bernt *et al.*, 2013)with default settings, and refined by alignment with published nymphylinae mitochondrial genome. Further the map of the predicted genes is generated using GenomeVx (Conant and Wolfe, 2008). The tRNA genes were identified with tRNAscan-SE Search Server (Schattner *et al*.,2005) with a COVE score cutoff of 1.0. Composition skew analysis was carried out to describe the base composition of nucleotide sequences, which measures the relative number of As to Ts (AT skew=[A–T]/[A+T]) and Gs to Cs (GC skew=[G–C]/[G+C]) (Perna and Kocher, 1995). In addition, the codon usage and nucleotide compositions were analyzed with MEGA 5 (Tamura, 2011).

Sixteen Nymphalidae complete mitochondrial genome sequence were downloaded from NCBI including two newly determined mt genomes, one Lycaenidae (*Protantigius supernan*) were also downloaded and used as out group to illustrate the phylogenetic placement of the newly determined mt genomes. Nucleic acid sequences of 13 concatenated protein-coding genes (PCGs) were used in phylogenetic analysis. The alignment of the nucleic acid sequences of each 13 mitochondrial PCGs was aligned with Clustal W using MEGA 5.

3.13. Characterization of Distal-less (Dll) gene in relation to their physicochemical properties

For characterization of dll gene, two species of *Elymnias hypernmestra* from north and south Brahmaputra River was sequenced and additional sequences of different butterflies species was retrieved from Swiss-Prot database in FASTA format for analysis. The physico-chemical characterization of the above proteins was computed using Expasy's Protparam tools and server for the following parameters viz., amino acid composition, theoretical isoelectric point (pI), molecular weight, negative and positive residues, extinction coefficient. The 3D structure was generated by using Esypred server, a homology based tool and was evaluated using servers Rampage, ProQ (Protein Quality server) and CE (Combinatorial Extension). Rasmol (http://openrasmol.org) was used to visualize the modeled 3D structure (Sivakumar *et al.*, 2007). Thephylogenetic tree was constructed using neighbouring joining method with 100 bootstrap support using MEGA 5.

IV. RESULTS

4.1. Samples selected for studying the intraspecific variation across Northeast India

The following eleven samples of butterflies were selected to study the intraspecific variation within the nymphalidae species across Northeast India and to determine whether Brahmaputra acts as a barrier for gene flow: *Junonia lemonias, Junonia iphita, Junonia almana, Elymnias hypernmestra, Euploea mulciber, Euthalia aconthea, Euthalia phemius, Vindula erota, Parantica aglea, Ariadne merione and Junonia atlites*. Initial studies related to the field work taking into account the four criteria for sample collection, availability of the species and preliminary results with COI gene led to the selected of final five individuals for the study viz; *Junonia almana, Junonia iphita, Junonia atlites, Elymnias hypernmestra and Ariadne merione*. These species were used for studying the molecular phylogenetic analysis and to discriminate whether it will be useful for detecting intraspecific variation between the nymphalids of north and south Brahmaputra. The sequences generated for the present study were submitted and published in GenBank, the species name, location, the voucher name and accession numbers are given (Table 3).

All the five species selected were collected from different parts of Northeast India: Sikkim, Arunachal Pradesh, and Bongaigaon in Assam which are on the north of Brahmaputra River. Meghalaya, Tripura, Manipur and different parts of Mizoram which are on the south Brahmaputra sites. From each location at least two or more individual of the selectedspecies were collected.

4.2. Morphological Characteristics of selected Nymphalidae species

1. *Junonia atlites* (Grey Pansy)

UP wings creamy grey with dark brown lines. Complete row of discal eyespots on both wings.

2.*Junonia almana* (Peacock Pansy)

UP tawny orange, UPF apex square-cut.Very large unmistakable eyespots on UPH and two smaller eyespots on UPF.

3. *Junonia iphita* (Chocolate Pansy)

UP pale to dark brown with darker brown bands. FW apex slightly produced and square-cut. UPH with row of small eyespots.UPF with or without small eyespots.

4. *Ariadne merione* (Common castor)

UP of both wings rusty brown.Discal line beyond cell double and wavy on UP.UN greyish brown with dark brown narrow bands.

5. *Elymnias hypernmestra* (Common Plamfly)

UP blackish brown with purple gloss.Broad diffuse chestnut border on UPH and marginal series of blue spots on UPF.

Figure 3: Five nymphalidae selected for studies.**UPF-Upper Forewing, UPH-Upper hind wing, FW-Fore wing, HW-Hind wing, UP-Upper, UN-Under.**

4.3. Distribution of Nymphalidae in Mizoram

During the systematic survey, a total of 2200 individuals of 115 species belonging to family Nymphalidae have been recorded from different parts of Mizoram. Mizoram was divided into three forest types: Tropical Wet Evergreen Forest (type1), Tropical Semi-Evergreen Forest (type2) and Mountain Sub-Tropical Forest (type3). Out of the 115 species, a maximum of 929 species were recorded in forest type 1 followed by forest type 2 with 837 individuals and the lowest with 434 individuals in forest type 3.The species were ranked according to the abundance, the K dominance curve for type 3 forests showed the lowest diversity with high dominance while type 1 and type 2 forests indicates that the species are evenly distributed and the diversity is also high compared to type 3 forests (Figure4).

The diversity was calculated using different indices and type 1 forest had the highest diversity (1.927), type 2 forest follows (1.914) and the lowest was type 3 forests (1.729) from Shannon H′s result. Simpson's diversity also shows the same trends with type 1 (74.66) as the highest, followed by type 2 (72.78) and the lowest is type 3 (44.36). The difference between type 1 and type 2 forests is very low in both the cases. The Berger-parker dominance is the highest in type 3 forest (7.37), while for type 1 (3.23) and type 2 (3.22) the dominance value is very similar (Table 3).Mountain Sub-Tropical forest was found to have the lowest species diversity (1.729 in Shannon and 44.36 in Simpson), lowest average rainfall (227.36), lowest average temperature (20.63) and lowest average humidity (70.11). For Tropical Wet Evergreen Forest and Tropical Semi-Evergreen Forest the average rainfall, temperature and humidity as shown in Table 4.Species richness, temperature and humidity were positively and highly correlated, whereas in rainfall and species richness the correlation was less. The Berger-parker dominance with species richness, temperature, humidity and rainfall were also negatively correlated (Table 4-6).

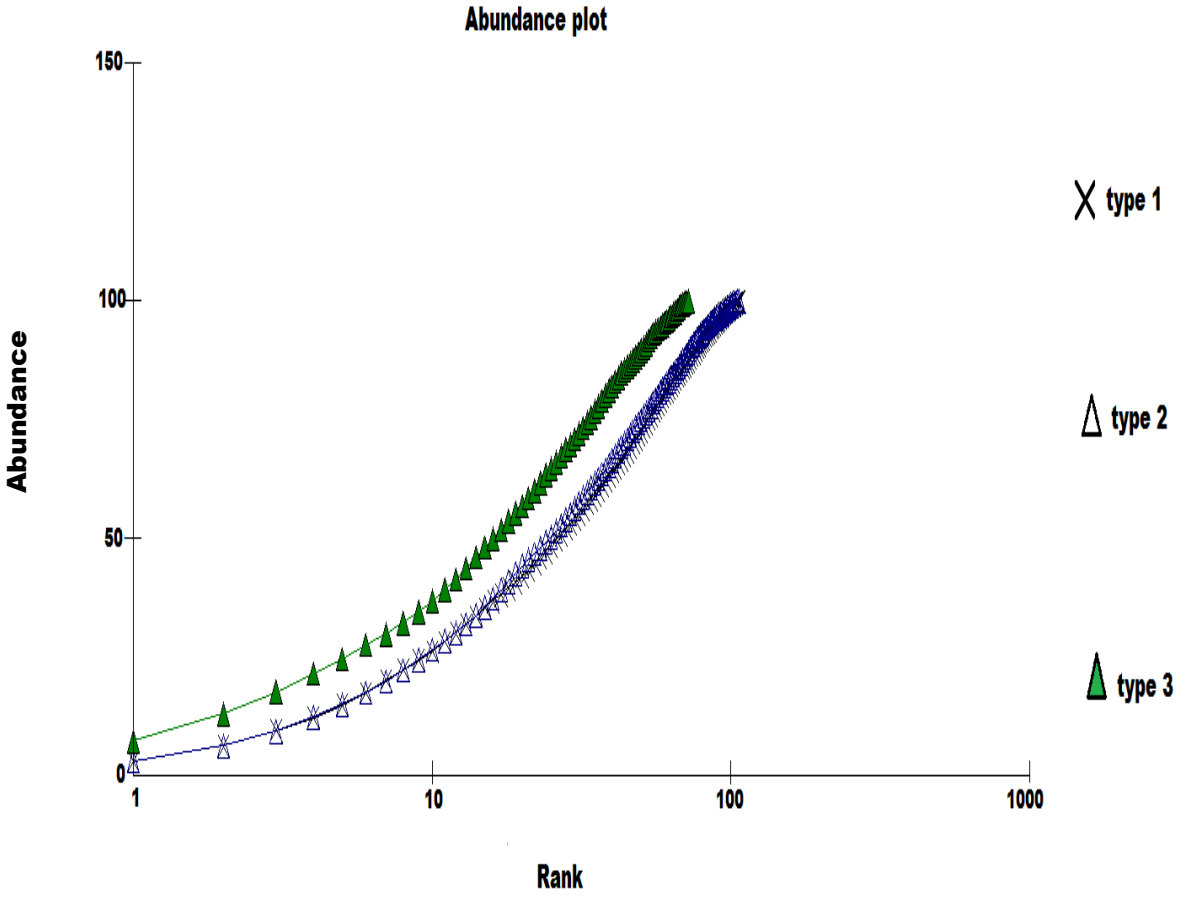


Figure 4: A ranked species K-dominance curve of Nymphalidae species collected from three forest type of Mizoram during 2010 – 2013

Table 3: Species names, Locations, Voucher name and Accession no. of selected Nymphalidae species

Name of the species	Locations	COI voucher Name	COI Accession No.	NDI Voucher Name	NDI Accession No.
1.*Junonia almana*	Bongaigaon(Assam)	JA_N1	KM115643	JAN_N1	KP258676
2.*Junonia almana*	Ranipool (Sikkim)	JA_N2	KM115626	JAN_N2	KP258678
3.*Junonia almana*	Itanagar(Arunachal Pradesh)	JA_N3	KM115619	JAN_N3	KP258679
4.*Junonia almana*	Buhchang(Mizoram)	JA_S1	KM115633	JAN_S1	KP258677
5.*Junonia almana*	Zawlnuam(Mizoram)	JA_S2	KM115637	JAN_S2	KP258680
6.*Junonia almana*	Chawngte (Mizoram)	JA_S3	KM115630	JAN_S3	KP258681
7.*Junonia iphita*	Bongaigaon(Assam)	JI_N1	KM115644	JIN_N1	KP258688
8.*Junonia iphita*	Rangpo (Sikkim)	JI_N2	KM115622	JIN_N2	KP258690
9.*Junonia iphita*	Itanagar(Arunachal Pradesh)	JI_N3	KM115617	JIN_N3	KP258691
10.*Junonia iphita*	Aizawl (Mizoram)	JI_S1	KM115634	JIN_S1	KP258689
11.*Junonia iphita*	Barapani(Meghalaya)	JI_S2	KM115639	JIN_S2	KP258692
12.*Junonia iphita*	Saiha (Mizoram)	JI_S3	KM115628	JIN_S3	KP258693
13.*Junonia atlites*	Bongaigaon(Assam)	JAT_N1	KM115646	JATN_N1	KP258682
14.*Junonia atlites*	Yuksam (Sikkim)	JAT_N2	KM115625	JATN_N2	KP258684
15.*Junonia atlites*	Itanagar(Arunachal Pradesh)	JAT_N3	KM115621	JATN_N3	KP258685

Name of the species	Locations	COI voucher Name	Coi Accession No.	NDI Voucher Name	NDI Accession No.
16.*Junonia atlites*	Aizawl (Mizoram)	JAT_S1	KM115631	JATN_S1	KP258683
17.*Junonia atlites*	Imphal(Manipur)	JAT_S2	KM115638	JATN_S2	KP258686
18.*Junonia atlites*	Chawngte (Mizoram)	JAT_S3	KM115645	JATN_S3	KP258687
19.*Ariadne merione*	Bongaigaon(Assam)	AM_N1	KM115641	AMN_N1	KP258664
20.*Ariadne merione*	Yuksam (Sikkim)	AM_N2	KM115623	AMN_N2	KP258666
21.*Ariadne merione*	Itanagar(Arunachal Pradesh)	AM_N3	KM115618	AMN_N3	KP258667
22.*Ariadne merione*	Aizawl (Mizoram)	AM_S1	KM115635	AMN_S1	KP258665
23.*Ariadne merione*	Lunglei (Mizoram)	AM_S2	KM115640	AMN_S2	KP258668
24.*Ariadne merione*	Zawlnuam(Mizoram)	AM_S3	KM115629	AMN_S3	KP258669
25.*Elymnias hypernmestra*	Bongaigaon(Assam)	EH_N1	KM115642	EHN_N1	KP258670
26.*Elymnias hypernmestra*	Ranipool (Sikkim)	EH_N2	KM115624	EHN_N2	KP258672
27.*Elymnias hypernmestra*	Itanagar(Arunachal Pradesh)	EH_N3	KM115620	EHN_N3	KP258673
28.*Elymnias hypernmestra*	Aizawl (Mizoram)	EH_S1	KM115632	EHN_S1	KP258671
29.*Elymnias hypernmestra*	Agartala (Tripura)	EH_S2	KM115636	EHN_S2	KP258674
30.*Elymnias hypernmestra*	Chawngte (Mizoram)	EH_S3	KM115627	EHN_S3	KP258675

Table 4: Diversity indices among the different Forest types in Mizoram

Forest type	Shannon H' log base 10	Shannon Hmax Log Base 10.	Shannon J'	Simpsons Diversity (D)	Simpsons Diversity (1/D)	Berger-Parker Dominance (d)	Berger-Parker Dominance (1/d)	Berger-Parker Dominance (d%)
Type 1	1.927	2.02	0.95	0.013	74.66	0.032	30.96	3.23
Type 2	1.914	2.03	0.94	0.014	72.78	0.032	31	3.22
Type 3	1.729	1.85	0.93	0.023	44.36	0.074	13.56	7.37

Forest type 1: Tropical Wet Evergreen Forest

Forest type 2: Tropical Semi-Evergreen Forest

Forest type 3: Mountain Sub-Tropical Forest

Table 5: Nymphalid Diversity indices, temperature, humidity and rainfall in the three forest types of Mizoram

Forest type	Shannon H'Log Base 10.	Simpsons Diversity (1/D)	Berger-Parker Dominance (d %)	Average Rainfall (mm)	Average Temp (°C)	Average Humidity (%)
Type 1	1.927	74.668	3.229	245.66	23.62	78.17
Type2	1.914	72.783	3.226	336.32	22.53	75.4
Type 3	1.729	44.363	7.373	227.36	20.63	70.11

Table 6: Correlation between Diversity, temperature, humidity and rainfall in The three forest types

Diversity Indices	Average rainfall (mm)	Average temperature (˚C)	Average humidity (%)	Shannon H'Log Base 10.	Simpsons Diversity (1/D)	Berger-Parker Dominance (d %)
Shannon H'Log Base 10.	0.583	0.952	0.959	1		
Simpsons Diversity (1/D)	0.586	0.951	0.958	0.999	1	
Berger-Parker Dominance(d%)	-0.63	-0.932	-0.94	-0.998	-0.998	1

4.4. Analysis of Data produced by RAPD-PCR

Five random primers (Bangalore Genei) were tested on thirty individual comprising five species of nymphalidae butterflies in three locations from north and south of Brahmaputra River in Northeast India. The PCR product size ranged from 100 to 1500bp. Only clear and distinct bands were selected for analysis (Figure 5). Gel electrophoresis of different sets of RAPD reaction products resulted in a total of 465 bands from North and 460 from South Brahmaputra River. The number of polymorphic bands from North is 232 (49.91%) and south is 243 (53%). The total no of bands (TNB), number of polymorphic bands (NPB), percentage of polymorphic bands (P %), Polymorphic information content (PIC), Resolving power (RP) and Marker index (MI) for individual primers are shown in Table 7. Primer OPT1 gives the highest total number of bands in both sides with 104 for north and 106 for south and number of polymorphic bands 55 and 56, respectively. The resolving power of marker is also highest with a value of 13.9 and 14 for both sides. PIC is used for estimating the degree of polymorphism of marker; it is a good measure for heterozygosity. High PIC value indicates rich heterozygosity which indicates high degree ofpolymorphism in the individual. The PIC value ranged from 0.19 to 0.24.Anova was done for binary data and PIC value and the result were summarized in Table 8.

The UPGMA dendrogram constructed from Jaccard's similarity coefficient based on the RAPD binary data implemented in NTSYS-pc 2.01i software separates the 30 individuals into five main groups with bootstraps support value at the nodescalculated using Darwin 5, all the *Elymnias hypernmestra* forms group I, group II constitute *Junonia almana* and it forms a sister clade to *Adriane merione* which makes up group III, group IV is *Junonia atlites* and form sister clades to *Junonia iphita* as the V group (Figure6).

Table 7: RAPD primers and the polymorphism detected in the present study

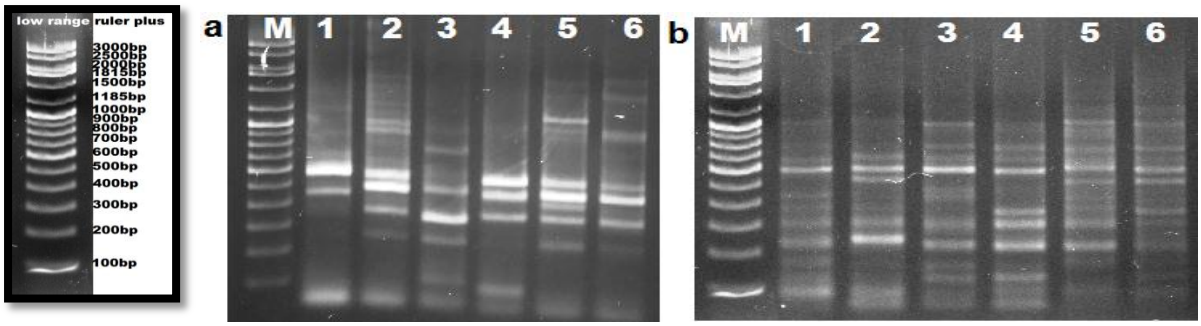
Primers	TNB		NPB		P %		PIC		RP		MI	
5'→3'	N	S	N	S	N	S	N	S	N	S	N	S
m18TGGTCAGTGA	98	91	41	42	41.8	46.2	0.24	0.23	13.1	12.1	0.1	0.11
m26GACGTGGTGA	89	84	46	47	51.7	56	0.21	0.20	11.9	11.2	0.11	0.11
opt1GGGCCACTCA	104	106	55	56	52.9	52.8	0.21	0.20	13.9	14	0.11	0.11
opt4GTGTCTCAGG	94	85	50	52	53.2	61.2	0.21	0.19	12.5	11.3	0.11	0.12
opt5GGGTTTGGCA	80	94	40	46	50	48.9	0.21	0.22	10.3	12.1	0.11	0.11
Total	465	460	232	243	49.9	53	0.22	0.21	12.3	12.2	0.11	0.11

Total number of bands (TNB), Number of polymorphic bands (NPB), Polymorphism Percentage (P %), Polymorphism information content (PIC), Resolving Power (RP) and Marker index (MI) of butterfly population. North-N; South-S.

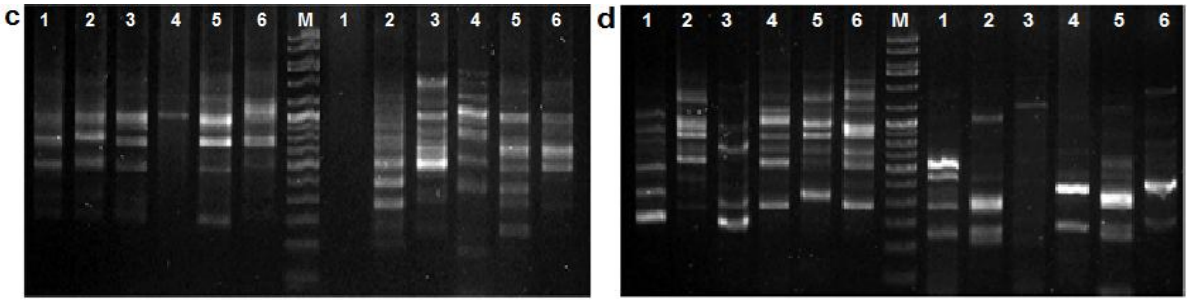
Table 8: One-way ANOVA of RAPD binary data and PIC value using PAST software

One way ANOVA	df	sum of sqrs	mean sqrs	F	p-value
RAPD binary data					
Between groups	29	4.183	0.144	1.697	0.01
Within groups	9810	833.86	0.085		
Total	9839	838.046			
PIC value					
Between groups	9	0.009	0.001	14.64	1.35E-09
Within groups	40	0.003	7.60E-05		
Total	49	0.012			

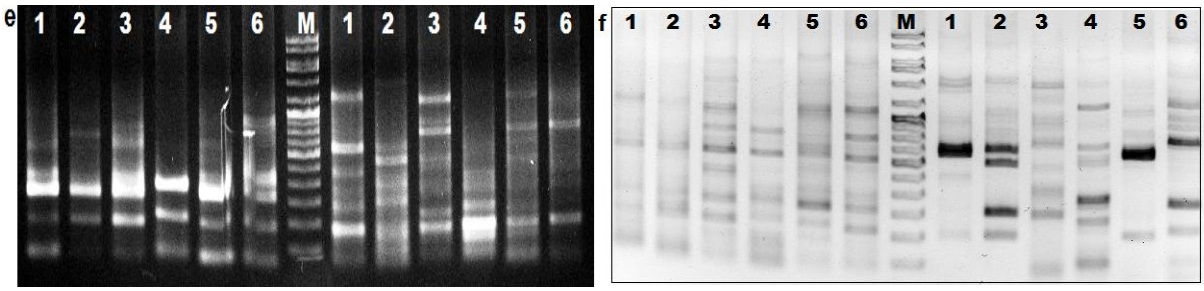
The total dataset contains 30 individuals from six location (five nymphalidae species), using 5 RAPD markers. Two analyses were conducted: the first between and within groups using RAPD binary data, and the second using PIC value. *P*-value and fixation indices (F) are given. df= degree of freedom; p value < 0.05 is statistically considered significant.



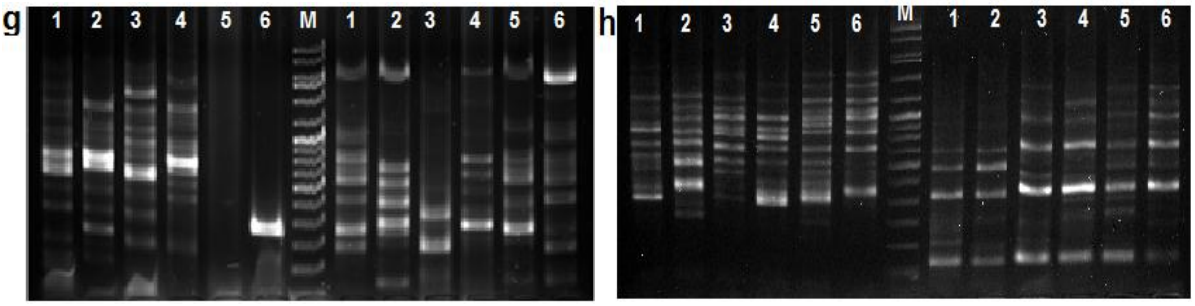
DNA ladder *J. atlites* P-M18*E. hypernmestra* P-M18



J.iphita and *A. merione*, P-OPT1 *J.atlites*, P- OPT1 and OPT4



J. iphita and A. merione, P-OPT5 *E. hypernmestra* and *J. almana*, P-M26



J.iphita and *J. almana*, P-OPT5 *E.hypernmestra*, P- OPT1 and OPT4

Figure 5: RAPD profile of five nymphalidae species using different primers.

M-Marker; (Lane 1, 2, 3 are from north Brahmaputra
and Lane 4, 5, 6 from south Brahmaputra). P-Primer.

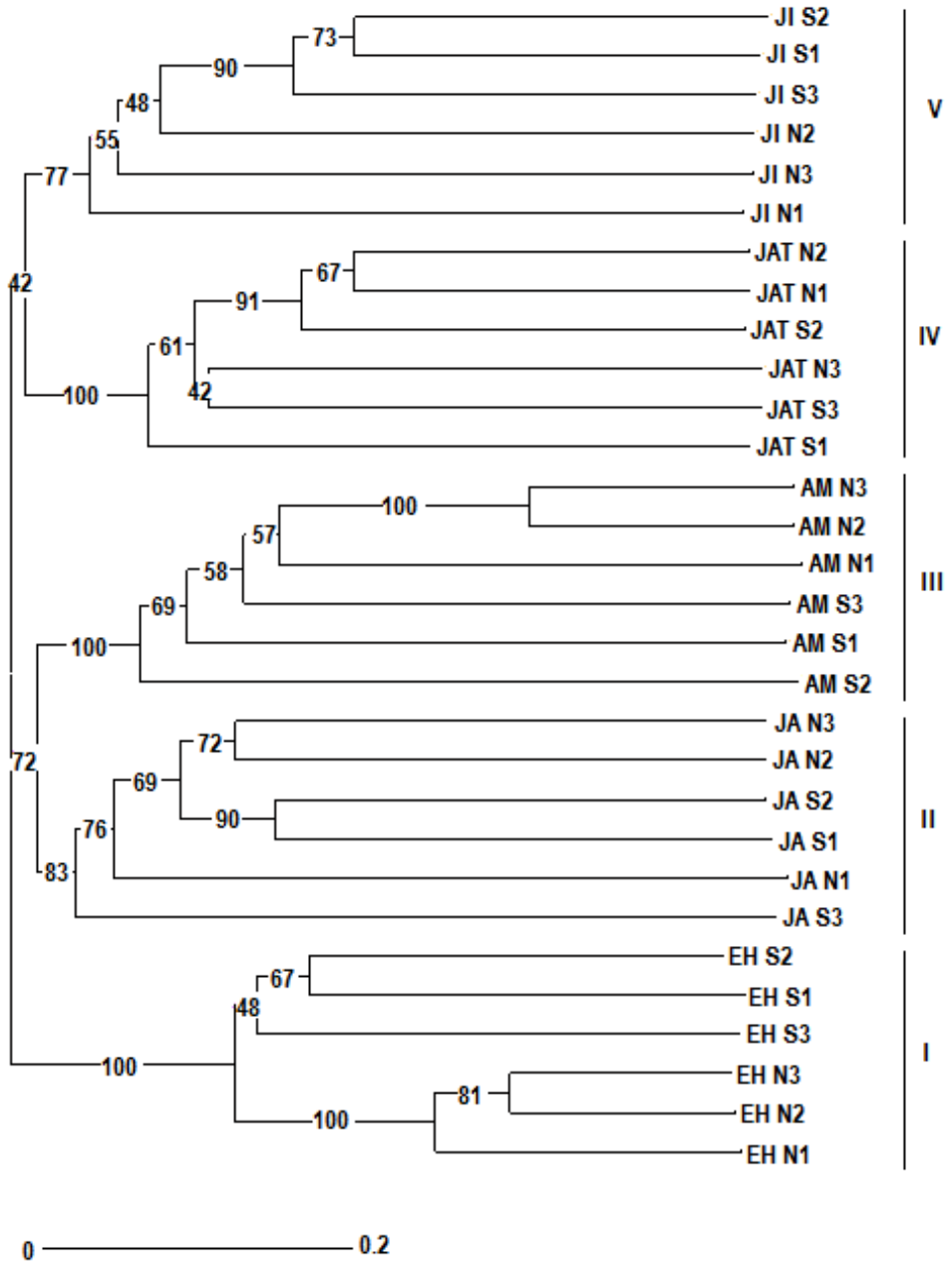


Figure 6: Dendrogram of the butterfly species generated using RAPD data.

EH-*Elymnias hypernmestra* (I); JAT-*Junonia atlites* (IV); JI-*Junonia iphita* (V); AM-*Ariadne merione* (III); JA-*Junonia almana* (II); S-South; N-North.

4.5. Amplification of COI, NDI and Dll genes

CO1 primers amplified approximately 750bp long fragment of the COI gene from the mitochondrial genome (Figure 7). Thirty individuals representing 5 species of Nymphalidae from different locations, selected for the studies were sequenced. ND1 primer amplified approximately 580bp long fragment of the mitochondrial genome, which includes fragments of 16S rRNA, the intervening tRNA leu region and the ND1 gene region (Figure 8). Like COI for ND1 gene all the thirty individuals from 5 species were sequenced. For dll gene two individual of *E.hypernmestra* around 380 bp from both sides of Brahmaputra river was amplified (Figure 9).

4.6. General properties of mitochondrial genes (COI and NDI)

The COI and NDI genes of the five species were separately analyzed in MEGA5. The total sequences length for COI is 653 bp and for NDI 389 bp. The nucleotide frequencies of all the species for COI and NDI genes were given in Tables 9 and 10. All the selected nymphalidae species analyzed using both the gene shows that the nucleotide composition is mostly A+T rich. AT content is the highest in the third codon followed by second codon and lowest in the first codon. The GC content is low for all the butterfly species and it is lowest in the third codon for both the genes.

Tables 11 and 12 show the maximum composite likelihood estimate of the pattern of nucleotide substitution COI and NDI gene sequences of selected nymphalidae species. Each entry shows the probability of substitution from one base (row) to another base (column) instantaneously. Only entries within a row have been compared. Substitution pattern and rates

were estimated under the Tamura-Nei model. The pattern of substitution matrix for COI (Table 11) and the pattern of substitution matrix for NDI (Table 12) shows that in both the cases transitional substitution is higher compared to transversional substitution. And between the two genes the substitutional pattern was also different COI shows higher transitional substitution than ND1 gene.

Comparison of evolutionary relationship among the nymphalidae species was done and a detailed summary of base characteristics, degenerative site's information and synonymous and non-synonymous substitution patterns within COI and NDI genes of butterfly species is represented in Table 13. Codon usage in the butterfly gene finds that UUA and AUU are the most frequently used codon in the butterfly species. Relative synonymous codon usage values (RSCU) of butterfly speciesare summarized in Table 14 and 15.

The pairwise genetic distance was calculated for both COI and NDI sequences using kimura-2-parameters (Tables 16 and 17). The evolutionary distance within the species was low in most of the cases for both the gene. In COI sequences the p-distance was mainly between 0.00 to 0.01 except for *Junonia iphita* where it ranged between 0.00 to 0.07. In NDI sequences, the p-distance range between 0.00 to 0.02. Codon Bias Index (CBI) values in our study,ranged from 0.657 to 0.702 in COI gene and 0.741 to 0.837 in NDI gene (Tables 18 and 19). CBI is a measure of deviation from the uniform use of synonymous codons thatachieves values between 0 and 1 for random use and maximum biasamong synonymous codons, respectively. Effective number of codon (ENC)value range from 33.26 to 36.53 in COI among the 32 sequences while for NDI gene the ENC value range from 25.64 to 31.12.ENC value would be 20 if individual codon is used for single amino acid where as if all codons were equally used value would be 61; and if ENC value is more than 40, the CUB was

regarded as a low bias.GC3S values range from 0.06 to 0.11 for COI gene and it range from 0.05 to 0.09 in NDI gene.The G + C3index, measures the fraction of thirdpositions of the codons that is G or C. Based upon the codon O tool the minimum codon usage bias is in (0.402) and maximum is (0.505) for COI gene and for NDI the minimum codon bias is (0.612) and maximum is (0.712).

4.7. Phylogenetic analysis of COI and NDI gene

Phylogenetic tree was constructed for COI and NDI gene using different tree building methods like maximum parsimony (PAUP), maximum likelihood (PHYML) and using bayesian analysis (MrBayes). Using Maximum parsimony (MP) method a phylogenetic tree was constructed with the COI and NDI gene sequences in PAUP* 4.01 version beta (Swofford, 1998). *Libythea celtis* was used as out-group. All the characters were are of type 'unord' and assign equal weight. Heuristic MP Searches were performed with 50 repetitions using random addition of taxa. Gaps were treated as a missing character. Optimal tree was then made by the tree bisection reconnection (TBR) branch swapping algorithm. All the analysis involved 32 nucleotide sequences for COI and 30 nucleotide sequences for NDI gene. MP method calculates the tree based on least number of evolutionary changes. One thousand bootstrap replications were performed to check the internal stability of the tree. Parsimony analysis for CO1 included a total of 653 characters. Out of the 653 characters, 487 characters were conserved, 150 variables characters were parsimony informative and 26 variable characters were parsimony uninformative characters.Parsimony analysis for ND1 included a total of 389 characters, in which 283 characters were conserved and 12 variable characters were parsimony uninformative, 94 were parsimony informative sites for all the

samples used for the studies. All position containing gaps and missing data were eliminated from the sequences for both COI and NDI genes.

The jModeltest result showed that GTR+I+G model scores the lowest AIC value for CO1 and the negative log likelihood value of 1938.5223 and for ND1 it shows that GTR+I+G model with the lowest AIC score and the negative log likelihood value of 1033.005. Bayesian analysis for CO1 was set at 8 million generations,the effective sample size was 3672.0185 and LnL score was -2191.8762. The first 20% trees were considered as the burn-in phase and discarded. Bayesian analysis for ND1 was also set at 8 million generations the effective sample size was 1280.985 and the LnL score was -1246.909. The first 20% trees were considered as the burn-in phase and discarded. Maximum likelihood (ML) tree was constructed using PhyML 3.0. (Guindon, 2010). The bootstrap support was set at 1000 and the model used was GTR for the both genes. Optimal tree was made by the Nearest Neighbor Interchange (NNI) branch swapping algorithm.All the analysis involved 33 nucleotide sequences for COI and 30 nucleotide sequences for NDI genes.

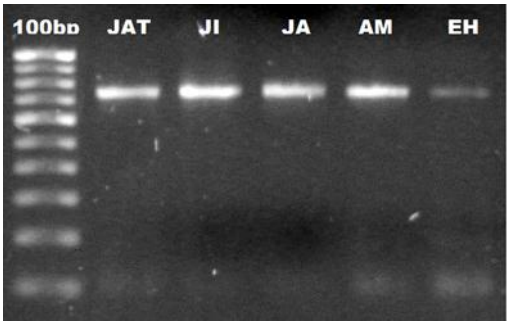


Figure 7: PCR Amplification product of COI gene for the five samples

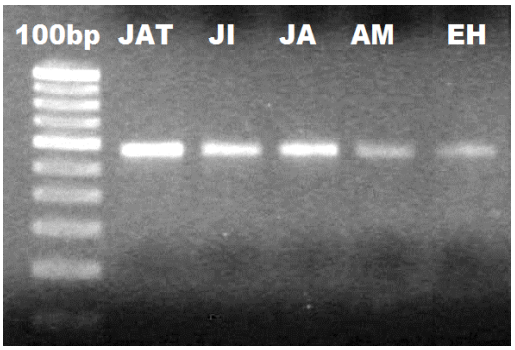


Figure 8: PCR Amplification product of ND1 gene for the five samples

Lane 1: 100bp Ladder; JAT-*J.atlites*; JI- *J. iphita*; JA- *J. almana*;

AM- *A.merione*; EH- *E. hypernmestra*

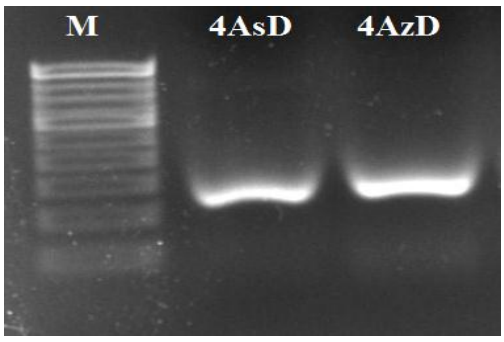


Figure 9: PCR Amplification product of dll gene for the two species

M-100bp ladder; 4AsD- *Elymnias hypernmestra* north;

4AzD- *Elymnias hypernmestra* south.

Table 9: Nucleotide frequency in COI gene of butterfly species

Species	All codon						First codon						Second codon						Third codon					
name	T	C	A	G	A+T	G+C	T	C	A	G	A+T	G+C	T	C	A	G	A+T	G+C	T	C	A	G	A+T	G+C
AM N1	40.6	15.5	29.9	14.1	70.4	29.6	29	14.2	30.7	25.7	60.1	39.9	43	25.2	16.5	15.6	59.2	40.8	50	6.9	42.4	.9	92.2	8
AM N2	40.6	15.5	29.9	14.1	70.4	29.6	29	14.2	30.7	25.7	60.1	39.9	43	25.2	16.5	15.6	59.2	40.8	50	6.9	42.4	.9	92.2	8
AM N3	40.4	15.5	29.7	14.4	70.1	29.9	29	14.2	30.7	25.7	60.1	39.9	43	25.2	16.5	15.6	59.2	40.8	49	6.9	41.9	1.8	91.2	9
AM S1	40.4	15.5	29.9	14.2	70.3	29.7	29	14.2	30.7	25.7	60.1	39.9	43	25.2	16.5	15.6	59.2	40.8	49	6.9	42.4	1.4	91.7	8
AM S2	40.5	15.5	29.9	14.1	70.4	29.6	29	14.2	30.7	25.7	60.1	39.9	42	25.3	16.6	15.7	59.0	41.0	50	6.9	42.4	.9	92.2	8
AM S3	40.6	15.5	29.9	14.1	70.4	29.6	29	14.2	30.7	25.7	60.1	39.9	43	25.2	16.5	15.6	59.2	40.8	50	6.9	42.4	.9	92.2	8
EH N1	40.3	16.4	29.1	14.2	69.4	30.6	28	15.6	31.7	25.2	59.2	40.8	43	24.3	17.0	15.6	60.1	39.9	50	9.2	38.7	1.8	88.9	11
EH N2	40.6	16.4	28.8	14.2	69.4	30.6	28	15.6	31.2	25.2	59.2	40.8	43	24.3	17.0	15.6	60.1	39.9	51	9.2	38.2	1.8	88.9	11
EH N3	40.6	16.4	28.8	14.2	69.4	30.6	28	15.6	31.2	25.2	59.2	40.8	43	24.3	17.0	15.6	60.1	39.9	51	9.2	38.2	1.8	88.9	11
EH S1	40.1	16.5	29.2	14.1	69.4	30.6	28	16.1	30.7	25.7	58.3	41.7	43	24.3	17.0	15.6	60.1	39.9	50	9.2	40.1	.9	89.9	10
EH S2	40.1	16.5	29.2	14.1	69.4	30.6	28	16.1	30.7	25.7	58.3	41.7	43	24.3	17.0	15.6	60.1	39.9	50	9.2	40.1	.9	89.9	10
EH S3	40.1	16.5	29.1	14.2	69.2	30.8	28	16.1	30.7	25.7	58.3	41.7	43	24.3	17.0	15.6	60.1	39.9	50	9.2	39.6	1.4	89.4	11
JA N1	38.3	15.6	31.5	14.5	69.8	30.2	28	15.1	28.9	27.5	57.3	42.7	43	24.8	16.1	16.1	59.2	40.8	43	6.9	49.8	.0	93.1	7
JA N2	38.6	15.3	31.5	14.5	70.1	29.9	28	15.1	28.9	27.5	57.3	42.7	43	24.8	16.1	16.1	59.2	40.8	44	6.0	49.8	.0	94.0	6
JA N3	38.6	15.3	31.5	14.5	70.1	29.9	28	15.1	28.9	27.5	57.3	42.7	43	24.8	16.1	16.1	59.2	40.8	44	6.0	49.8	.0	94.0	6
JA S1	38.3	15.5	31.9	14.4	70.1	29.9	28	15.1	29.4	27.1	57.8	42.2	43	24.8	16.1	16.1	59.2	40.8	43	6.5	50.2	.0	93.5	6
JA S2	38.6	15.3	31.5	14.5	70.1	29.9	28	15.1	28.9	27.5	57.3	42.7	43	24.8	16.1	16.1	59.2	40.8	44	6.0	49.8	.0	94.0	6
JA S3	38.4	15.5	31.5	14.5	70.0	30.0	28	15.1	28.9	27.5	57.3	42.7	43	24.8	16.1	16.1	59.2	40.8	44	6.5	49.8	.0	93.5	6
JAT N1	38.4	16.1	30.8	14.7	69.2	30.8	28	14.7	29.8	27.1	58.3	41.7	43	24.8	16.1	16.1	59.2	40.8	44	8.8	46.5	.9	90.3	10
JAT N2	38.4	16.2	30.8	14.5	69.2	30.8	28	14.7	29.8	27.1	58.3	41.7	43	24.8	16.1	16.1	59.2	40.8	44	9.2	46.5	.5	90.3	10
JAT N3	38.3	16.2	30.9	14.5	69.2	30.8	28	14.7	30.3	27.1	58.3	41.7	43	24.8	16.1	16.1	59.2	40.8	44	9.2	46.5	.5	90.3	10
JAT S1	38.4	16.2	30.8	14.5	69.2	30.8	28	14.7	29.8	27.1	58.3	41.7	43	24.8	16.1	16.1	59.2	40.8	44	9.2	46.5	.5	90.3	10
JAT S2	38.3	16.2	30.9	14.5	69.2	30.8	28	14.7	29.8	27.1	58.3	41.7	43	24.8	16.1	16.1	59.2	40.8	43	9.2	47.0	.5	90.3	10
JAT S3	38.3	16.4	30.9	14.4	69.2	30.8	28	14.7	30.3	26.6	58.7	41.3	43	24.8	16.1	16.1	59.2	40.8	43	9.7	46.5	.5	89.9	10
JI N1	38.6	15.8	30.9	14.7	69.5	30.5	29	14.2	29.8	27.1	58.7	41.3	43	24.8	16.1	16.1	59.2	40.8	44	8.3	47.0	.9	90.8	9
JI N2	38.6	15.8	30.9	14.7	69.5	30.5	29	14.2	29.8	27.1	58.7	41.3	43	24.8	16.1	16.1	59.2	40.8	44	8.3	47.0	.9	90.8	9
JI N2b	38.4	15.6	31.1	14.9	69.5	30.5	28	14.2	30.7	27.1	58.7	41.3	43	24.8	15.6	16.5	58.7	41.3	44	7.8	47.0	.9	91.2	9
JI N3	38.3	16.1	31.1	14.5	69.4	30.6	27	16.1	30.7	26.1	57.8	42.2	43	24.8	16.1	16.1	59.2	40.8	45	7.4	46.5	1.4	91.2	9
JI S1	38.6	15.8	31.1	14.5	69.7	30.3	29	14.2	29.8	27.1	58.7	41.3	43	24.8	16.1	16.1	59.2	40.8	44	8.3	47.5	.5	91.2	9
JI S2	38.4	15.9	31.1	14.6	69.5	30.5	28	14.4	30.2	27.0	58.6	41.4	43	25.2	15.4	16.4	58.4	41.6	44	7.9	47.7	.5	91.6	8
JI S3	38.2	16.0	31.2	14.6	69.4	30.6	27	16.1	30.9	26.3	57.6	42.4	43	24.9	16.1	16.1	59.0	41.0	45	6.9	46.8	1.4	91.7	8
JI S3b	38.1	15.6	31.5	14.7	69.7	30.3	28	14.2	31.2	27.1	58.7	41.3	43	24.8	16.5	16.1	59.2	40.8	44	7.8	47.0	.9	91.2	9
Avg.	39.2	15.9	30.5	14.4	69.7	30.3	28	14.9	30.2	26.5	58.6	41.4	43	24.8	16.3	15.9	59.3	40.7	46	7.9	45.1	.8	91.3	9

JA-*J.almana*,JAT-*J.altites*,JI-*J.iphita*, EH-*E.hypernmestra*, AM-*A.merione.*

N-north and S-South.

For each species the nucleotide frequency for the overall codon and the nucleotide frequencies at each codon position are given. The result is given in percentage.

Note: A+T content is the highest in the third codon followed by second codon and lowest in the first codon.

Table 10: Nucleotide frequency in NDI gene of butterfly species

Species	All codon						First codon						Second codon						Third codon					
name	T	C	A	G	A+T	G+C	T	C	A	G	A+T	G+C	T	C	A	G	A+T	G+C	T	C	A	G	A+T	G+C
AM.N1	47.6	7.5	31.6	13.4	79.2	20.8	40	6.9	33.1	20.0	73.1	26.9	51	13.8	20.8	14.6	71.5	28.5	52	1.6	41.1	5.4	93.0	7
AM.N2	47.6	7.5	31.6	13.4	79.2	20.8	40	6.9	33.1	20.0	73.1	26.9	51	13.8	20.8	14.6	71.5	28.5	52	1.6	41.1	5.4	93.0	7
AM.N3	47.6	7.5	31.6	13.4	79.2	20.8	40	6.9	33.1	20.0	73.1	26.9	51	13.8	20.8	14.6	71.5	28.5	52	1.6	41.1	5.4	93.0	7
AM.S1	47.6	7.5	31.6	13.4	79.2	20.8	40	6.9	33.1	20.0	73.1	26.9	51	13.8	20.8	14.6	71.5	28.5	52	1.6	41.1	5.4	93.0	7
AM.S2	47.6	7.5	31.6	13.4	79.2	20.8	40	6.9	33.1	20.0	73.1	26.9	51	13.8	20.8	14.6	71.5	28.5	52	1.6	41.1	5.4	93.0	7
AM.S3	47.6	7.5	31.6	13.4	79.2	20.8	40	6.9	33.1	20.0	73.1	26.9	51	13.8	20.8	14.6	71.5	28.5	52	1.6	41.1	5.4	93.0	7
EH.N1	48.3	6.0	32.1	13.6	80.4	19.6	38	7.0	35.9	18.8	74.2	25.8	52	10.2	21.1	17.2	72.7	27.3	55	.8	39.4	4.7	94.5	6
EH.N2	48.0	6.3	31.9	13.8	79.9	20.1	38	7.0	35.9	18.8	74.2	25.8	52	10.2	21.1	17.2	72.7	27.3	54	1.6	38.6	5.5	92.9	7
EH.N3	48.3	6.0	31.9	13.8	80.2	19.8	38	7.0	35.9	18.8	74.2	25.8	52	10.2	21.1	17.2	72.7	27.3	55	.8	38.6	5.5	93.7	6
EH.S1	48.8	5.7	31.1	14.4	79.9	20.1	38	7.0	35.9	18.8	74.2	25.8	52	10.2	21.1	17.2	72.7	27.3	57	.0	36.2	7.1	92.9	7
EH.S2	48.8	5.7	31.1	14.4	79.9	20.1	38	7.0	35.9	18.8	74.2	25.8	52	10.2	21.1	17.2	72.7	27.3	57	.0	36.2	7.1	92.9	7
EH.S3	48.8	5.7	31.1	14.4	79.9	20.1	38	7.0	35.9	18.8	74.2	25.8	52	10.2	21.1	17.2	72.7	27.3	57	.0	36.2	7.1	92.9	7
JA.N1	48.8	6.7	30.1	14.4	78.9	21.1	39	7.7	33.1	20.0	72.3	27.7	52	12.3	20.8	14.6	73.1	26.9	55	.0	36.4	8.5	91.5	9
JA.N2	47.8	6.7	30.8	14.7	78.7	21.3	39	7.7	33.1	20.0	72.3	27.7	52	12.3	20.8	14.6	73.1	26.9	52	.0	38.8	9.3	90.7	9
JA.N3	48.8	6.7	30.1	14.4	78.9	21.1	39	7.7	33.1	20.0	72.3	27.7	52	12.3	20.8	14.6	73.1	26.9	55	.0	36.4	8.5	91.5	9
JA.S1	48.8	6.7	30.1	14.4	78.9	21.1	39	7.7	33.1	20.0	72.3	27.7	52	12.3	20.8	14.6	73.1	26.9	55	.0	36.4	8.5	91.5	9
JA.S2	48.8	6.7	30.1	14.4	78.9	21.1	39	7.7	33.1	20.0	72.3	27.7	52	12.3	20.8	14.6	73.1	26.9	55	.0	36.4	8.5	91.5	9
JA.S3	48.8	6.7	30.1	14.4	78.9	21.1	39	7.7	33.1	20.0	72.3	27.7	52	12.3	20.8	14.6	73.1	26.9	55	.0	36.4	8.5	91.5	9
JAT.N1	48.8	6.4	30.1	14.7	78.9	21.1	40	6.9	30.8	22.3	70.8	29.2	52	12.3	20.8	14.6	73.1	26.9	54	.0	38.8	7.0	93.0	7
JAT.N2	48.8	6.4	30.1	14.7	78.9	21.1	40	6.9	30.8	22.3	70.8	29.2	52	12.3	20.8	14.6	73.1	26.9	54	.0	38.8	7.0	93.0	7
JAT.N3	48.8	6.4	30.1	14.7	78.9	21.1	40	6.9	30.8	22.3	70.8	29.2	52	12.3	20.8	14.6	73.1	26.9	54	.0	38.8	7.0	93.0	7
JAT.S1	48.8	6.4	30.1	14.7	78.9	21.1	40	6.9	30.8	22.3	70.8	29.2	52	12.3	20.8	14.6	73.1	26.9	54	.0	38.8	7.0	93.0	7
JAT.S2	48.8	6.4	30.1	14.7	78.9	21.1	40	6.9	30.8	22.3	70.8	29.2	52	12.3	20.8	14.6	73.1	26.9	54	.0	38.8	7.0	93.0	7
JAT.S3	48.8	6.4	30.1	14.7	78.9	21.1	40	6.9	30.8	22.3	70.8	29.2	52	12.3	20.8	14.6	73.1	26.9	54	.0	38.8	7.0	93.0	7
JI.N1	50.1	5.5	30.5	13.9	80.6	19.4	40	6.6	32.2	21.5	71.9	28.1	54	10.0	21.7	14.2	75.8	24.2	57	.0	37.5	5.8	94.2	6
JI.N2	49.6	6.4	30.3	13.6	79.9	20.1	40	6.9	32.3	20.8	72.3	27.7	52	12.3	20.8	14.6	73.1	26.9	57	.0	38.0	5.4	94.6	5
JI.N3	49.6	6.4	30.1	13.9	79.7	20.3	40	6.9	31.5	21.5	71.5	28.5	52	12.3	20.8	14.6	73.1	26.9	57	.0	38.0	5.4	94.6	5
JI.S1	49.4	6.7	30.1	13.9	79.4	20.6	40	6.9	31.5	21.5	71.5	28.5	52	12.3	20.8	14.6	73.1	26.9	56	.8	38.0	5.4	93.8	6
JI.S2	49.4	6.7	30.1	13.9	79.4	20.6	40	6.9	31.5	21.5	71.5	28.5	52	12.3	20.8	14.6	73.1	26.9	56	.8	38.0	5.4	93.8	6
JI.S3	49.4	6.7	30.1	13.9	79.4	20.6	40	6.9	31.5	21.5	71.5	28.5	52	12.3	20.8	14.6	73.1	26.9	56	.8	38.0	5.4	93.8	6
Avg.	48.6	6.6	30.7	14.1	79.3	20.7	39	7.1	32.9	20.5	72.4	27.6	52	12.1	20.9	15.1	72.8	27.2	55	.5	38.4	6.5	93.0	7

JA-*J.almana,*JAT-*J.altites,*JI-*J.iphita,* EH-*E.hypernmestra,* AM-*A.merione.*

N-north and S-South.

For each species the nucleotide frequency for the overall codon and the nucleotide frequencies at each codon position are given. The result is given in percentage.

Table 11: Maximum composite likelihood estimate of the pattern of nucleotide substitution of coi gene

	A	T	C	G
A	-	*6.75*	*2.74*	**9.16**
T	*5.26*	-	**10.69**	*2.49*
C	*5.26*	**26.34**	-	*2.49*
G	**19.36**	*6.75*	*2.74*	-

NOTE: Each entry shows the probability of substitution (r) from one base (row) to another base (column). Rates of different transitional substitutions are shown in **bold** and those of transversionsal substitutions are shown in *italics*.The analysis involved 32 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 653 positions in the final dataset.

Table 12: Maximum composite likelihood estimate of the pattern of nucleotide substitution of ndi gene

	A	T	C	G
A	-	*6.74*	*0.91*	**8.36**
T	*4.26*	-	**5.42**	*1.95*
C	*4.26*	**40.23**	-	*1.95*
G	**18.26**	*6.74*	*0.91*	-

NOTE: Each entry shows the probability of substitution (r) from one base (row) to another base (column). Rates of different transitional substitutions are shown in **bold** and those of transversionsal substitutions are shown in *italics.* The analysis involved 30 nucleotide sequences. All ambiguous positions were removed for each sequence pair. There were a total of 389 positions in the final dataset

Table 13: Comparison of the evolutionary features in COI and NDI gene of butterflies

Species		Base Characteristics					Degenerative sites			S	N	Lc	dS	dN	t	w	Ps/pn	ds/dn	K1	K2	P1/s	P_S^*	P1/N	P_N*	R
		T	C	V	Si	Pi	0	2	4																
JA	COI	653	649	4	3	1	410	136	104	161.9	492.1	217.9	0.01	0.002	0.02	5.5	5.5	5.6	5.4	11.8	0.24	0.64	0.75	0.35	3.82
JAT			647	6	6	0	410	139	102	160.5	493.4	217.9	0.0	0.002	0.01	4	3.4	3.4	1.8	2.7	0.24	0.56	0.75	0.43	1.00
JI			606	47	5	42	409	134	100	160.1	489.7	216.5	0.18	0.01	0.21	23	21	23.8	16.8	28.1	0.24	0.88	0.75	0.11	10.0
EH			641	12	3	9	412	137	101	158.4	495.6	217.9	0.04	0.003	0.05	14.3	16	16.6	12.4	4.6	0.24	0.82	0.75	0.17	3.36
AM			649	4	4	1	408	144	99	147.7	461.7	203.1	0.01	0	0.01	0	0	0	5.8	0	0.24	1	0.75	0	1.01
JA	NDI	389	385	4	4	0	236	112	39	84.3	305.7	129.9	0.02	0	0.02	0	0	0	0	0	0.21	1	0.78	0	0
JAT			389	0	0	0	235	111	41	85.3	304.7	129.9	0	0	0	0	0	0	1	1	0.21	0	0.78	0	0.31
JI			386	3	2	1	235	112	39	82.9	297.1	126.6	0.01	0.003	0.02	4	3.6	3.6	6	15.2	0.21	0.52	0.78	0.47	2.82
EH			376	7	1	6	235	108	38	79.9	304.1	127.9	0.04	0.01	0.07	4.8	4.8	4.9	23.3	13.9	0.20	0.56	0.79	0.43	5.57
AM			389	0	0	0	237	110	40	84.7	305.3	129.9	0	0	0	0	0	0	1	1	0.21	0	0.78	0	0.31

T= Total no of sites; C= Conserved sites, V= Variable sites; Si= Singleton sites ; Pi= Parsimony informative sites; 0= 0-fold degenerate sites; 2= 2-fold degenerate sites ;4= 4-fold degenerate sites; S= Number of synonymous sites; N= Number of nonsynonymous sites; Lc= Number of codons; dS= Number of synonymous substitutions per synonymous site; dN= Number of nonsynonymous substitutions per nonsynonymous site; t= Time (distance) measured by the expected number of nucleotide substitutions per codon; w= Nonsynonymous/synonymous rate ratio; K1(purine) & K2(pyrimidine)= Transition/transversion (mutation) rate ratio; P1/s= Proportion of nonsynonymous sites; P^{s*}= Proportion of synonymous substitutions; P1/N= Proportion of nonsynonymous sites; PN*= Proportion of nonsynonymous substitutions; R= Transition/Transversion bias. JA-*J.almana*,JAT-*J.altites*,JI-*J.iphita*, EH-*E.hypernmestra*, AM-*A.merione.*

Table 14: Relative synonymous codon usage (RSCU) among the COI gene sequences

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	13.5	1.8	UCU(S)	5.8	2.51	UAU(Y)	2.1	1.05	UGU(C)	0	0
UUC(F)	1.5	0.2	UCC(S)	1.6	0.71	UAC(Y)	1.9	0.95	UGC(C)	0	0
UUA(L)	22.7	4.65	UCA(S)	6.6	2.89	UAA(*)	0	0	UGA(W)	5	2
UUG(L)	0.2	0.03	UCG(S)	0	0	UAG(*)	0	0	UGG(W)	0	0
CUU(L)	4.3	0.89	CCU(P)	4.6	1.31	CAU(H)	3.8	1.52	CGU(R)	0.6	0.56
CUC(L)	0.9	0.19	CCC(P)	2.9	0.84	CAC(H)	1.2	0.47	CGC(R)	0	0
CUA(L)	1.2	0.25	CCA(P)	6.5	1.85	CAA(Q)	3	2	CGA(R)	3.4	3.44
CUG(L)	0	0	CCG(P)	0	0	CAG(Q)	0	0	CGG(R)	0	0
AUU(I)	22.1	1.91	ACU(T)	4.9	1.6	AAU(N)	11.3	1.81	AGU(S)	1.4	0.6
AUC(I)	1	0.09	ACC(T)	1.4	0.45	AAC(N)	1.2	0.19	AGC(S)	0	0
AUA(M)	13.6	1.99	ACA(T)	5.8	1.88	AAA(K)	0	2	AGA(S)	3	1.29
AUG(M)	0.1	0.01	ACG(T)	0.2	0.07	AAG(K)	0	0	AGG(S)	0	0
GUU(V)	5.7	1.94	GCU(A)	9	2.61	GAU(D)	6.2	1.55	GGU(G)	5	0.94
GUC(V)	0.3	0.11	GCC(A)	1.4	0.4	GAC(D)	1.8	0.45	GGC(G)	0	0
GUA(V)	4.9	1.68	GCA(A)	3.4	0.98	GAA(E)	3	2	GGA(G)	15.8	2.97
GUG(V)	0.8	0.28	GCG(A)	0	0.01	GAG(E)	0	0	GGG(G)	0.5	0.09

The analysis involved 32 COI sequences.The codon with highest RSCU was UUA (L)

Table 15: Relative synonymous codon usage (RSCU) among the NDI gene sequences

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	9.6	1.99	UCU(S)	4.3	4.35	UAU(Y)	9	1.96	UGU(C)	2.7	1.93
UUC(F)	0	0.01	UCC(S)	0	0	UAC(Y)	0.2	0.04	UGC(C)	0.1	0.07
UUA(L)	20.2	5.56	UCA(S)	0.6	0.64	UAA(*)	0	0	UGA(W)	2	2
UUG(L)	1.4	0.39	UCG(S)	0	0	UAG(*)	0	0	UGG(W)	0	0
CUU(L)	0.2	0.06	CCU(P)	3.3	3.33	CAU(H)	0	0	CGU(R)	1.6	3.2
CUC(L)	0	0	CCC(P)	0	0	CAC(H)	0	0	CGC(R)	0	0
CUA(L)	0	0	CCA(P)	0.4	0.44	CAA(Q)	2.4	1.6	CGA(R)	0.4	0.8
CUG(L)	0	0	CCG(P)	0.2	0.24	CAG(Q)	0.6	0.4	CGG(R)	0	0
AUU(I)	13.7	1.99	ACU(T)	1.8	2.6	AAU(N)	5.8	2	AGU(S)	0.7	0.7
AUC(I)	0.1	0.01	ACC(T)	0	0	AAC(N)	0	0	AGC(S)	0	0
AUA(M)	10.6	1.74	ACA(T)	0.8	1.11	AAA(K)	4	1.6	AGA(S)	2.3	2.31
AUG(M)	1.6	0.26	ACG(T)	0.2	0.29	AAG(K)	1	0.4	AGG(S)	0	0
GUU(V)	7.4	3.03	GCU(A)	3	3.96	GAU(D)	1.8	2	GGU(G)	5.2	2.12
GUC(V)	0	0	GCC(A)	0	0	GAC(D)	0	0	GGC(G)	0.2	0.08
GUA(V)	2	0.81	GCA(A)	0	0.04	GAA(E)	1.8	1.64	GGA(G)	1.8	0.74
GUG(V)	0.4	0.16	GCG(A)	0	0	GAG(E)	0.4	0.36	GGG(G)	2.6	1.06

The analysis involved 30 NDI sequences. The codon with highest RSCU was UUA (L)

Table 16: The genetic distance (p-distance) of the COI gene sequences of the 5 Nymphalidae species

AM_N1																															
AM_N2	0.00																														
AM_N3	0.00	0.00																													
AM_S1	0.00	0.00	0.01																												
AM_S2	0.00	0.00	0.00	0.00																											
AM_S3	0.00	0.00	0.00	0.00	0.00																										
EH_N1	0.13	0.13	0.13	0.13	0.13	0.13																									
EH_N2	0.12	0.12	0.13	0.13	0.13	0.12	0.01																								
EH_N3	0.12	0.12	0.13	0.13	0.13	0.12	0.01	0.00																							
EH_S1	0.12	0.12	0.13	0.12	0.12	0.12	0.01	0.01	0.01																						
EH_S2	0.12	0.12	0.13	0.12	0.12	0.12	0.01	0.01	0.01	0.00																					
EH_S3	0.12	0.12	0.13	0.13	0.13	0.12	0.02	0.01	0.01	0.00	0.00																				
JA_N1	0.13	0.13	0.13	0.13	0.13	0.13	0.15	0.15	0.15	0.14	0.14	0.14																			
JA_N2	0.13	0.13	0.13	0.13	0.13	0.13	0.15	0.15	0.15	0.14	0.14	0.14	0.00																		
JA_N3	0.13	0.13	0.13	0.13	0.13	0.13	0.15	0.15	0.15	0.14	0.14	0.14	0.00	0.00																	
JA_S1	0.13	0.13	0.13	0.13	0.13	0.13	0.15	0.15	0.15	0.14	0.14	0.15	0.00	0.00	0.00																
JA_S2	0.13	0.13	0.13	0.13	0.13	0.13	0.15	0.15	0.15	0.14	0.14	0.14	0.00	0.00	0.00	0.00															
JA_S3	0.13	0.13	0.13	0.13	0.13	0.13	0.15	0.15	0.15	0.14	0.14	0.14	0.00	0.00	0.00	0.00	0.00														
JAT_N1	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.13	0.13	0.13	0.08	0.07	0.07	0.08	0.07	0.07													
JAT_N2	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.13	0.13	0.13	0.08	0.07	0.07	0.08	0.07	0.07	0.00												
JAT_N3	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.13	0.13	0.13	0.08	0.07	0.07	0.08	0.07	0.08	0.00	0.00											
JAT_S1	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.13	0.13	0.13	0.08	0.07	0.07	0.08	0.07	0.07	0.00	0.00	0.00										
JAT_S2	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.13	0.13	0.13	0.07	0.07	0.07	0.08	0.07	0.07	0.00	0.00	0.00	0.00									
JAT_S3	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.14	0.14	0.14	0.08	0.08	0.08	0.08	0.08	0.08	0.01	0.00	0.00	0.00	0.00								
JI_N1	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.13	0.13	0.13	0.07	0.07	0.07	0.07	0.07	0.07	0.02	0.02	0.02	0.02	0.02	0.02							
JI_N2	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.13	0.13	0.13	0.07	0.07	0.07	0.07	0.07	0.07	0.02	0.02	0.02	0.02	0.02	0.02	0.00						
JI_N2b	0.14	0.14	0.14	0.14	0.13	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.08	0.08	0.08	0.08	0.08	0.08	0.03	0.03	0.02	0.03	0.03	0.03	0.01	0.01					
JI_N3	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.08	0.08	0.08	0.08	0.08	0.08	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07				
JI_S1	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.13	0.13	0.13	0.13	0.13	0.07	0.07	0.07	0.07	0.07	0.07	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.01	0.06			
JI_S2	0.13	0.13	0.13	0.13	0.13	0.13	0.14	0.14	0.14	0.13	0.13	0.13	0.07	0.07	0.07	0.07	0.07	0.07	0.02	0.02	0.02	0.02	0.02	0.02	0.00	0.00	0.00	0.07	0.00		
JI_S3	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.08	0.08	0.08	0.08	0.08	0.08	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.06	0.07	0.00	0.06	0.07	
JI_S3b	0.14	0.14	0.14	0.14	0.14	0.14	0.14	0.15	0.15	0.14	0.14	0.14	0.08	0.08	0.08	0.08	0.08	0.08	0.03	0.03	0.03	0.03	0.03	0.03	0.01	0.01	0.01	0.07	0.01	0.00	0.07

JA-*J.almana*,JAT-*J.altites*,JI-*J.iphita*, EH-*E.hypernmestra*, AM-*A.merione*.N-north and S-South

Table 17: The genetic distance (p-distance) of the NDIgene sequences of the 5 Nymphalidae species

AM.N1																													
AM.N2	0.00																												
AM.N3	0.00	0.00																											
AM.S1	0.00	0.00	0.00																										
AM.S2	0.00	0.00	0.00	0.00																									
AM.S3	0.00	0.00	0.00	0.00	0.00																								
EH.N1	0.18	0.18	0.18	0.18	0.18	0.18																							
EH.N2	0.19	0.19	0.19	0.19	0.19	0.19	0.01																						
EH.N3	0.18	0.18	0.18	0.18	0.18	0.18	0.00	0.00																					
EH.S1	0.18	0.18	0.18	0.18	0.18	0.18	0.01	0.02	0.02																				
EH.S2	0.18	0.18	0.18	0.18	0.18	0.18	0.01	0.02	0.02	0.00																			
EH.S3	0.18	0.18	0.18	0.18	0.18	0.18	0.01	0.02	0.02	0.00	0.00																		
JA.N1	0.14	0.14	0.14	0.14	0.14	0.14	0.18	0.19	0.18	0.18	0.18	0.18																	
JA.N2	0.15	0.15	0.15	0.15	0.15	0.15	0.19	0.19	0.19	0.20	0.20	0.20	0.01																
JA.N3	0.14	0.14	0.14	0.14	0.14	0.14	0.18	0.19	0.18	0.18	0.18	0.18	0.00	0.01															
JA.S1	0.14	0.14	0.14	0.14	0.14	0.14	0.18	0.19	0.18	0.18	0.18	0.18	0.00	0.01	0.00														
JA.S2	0.14	0.14	0.14	0.14	0.14	0.14	0.18	0.19	0.18	0.18	0.18	0.18	0.00	0.01	0.00	0.00													
JA.S3	0.14	0.14	0.14	0.14	0.14	0.14	0.18	0.19	0.18	0.18	0.18	0.18	0.00	0.01	0.00	0.00	0.00												
JAT.N1	0.13	0.13	0.13	0.13	0.13	0.13	0.19	0.19	0.19	0.19	0.19	0.19	0.06	0.07	0.06	0.06	0.06	0.06											
JAT.N2	0.13	0.13	0.13	0.13	0.13	0.13	0.19	0.19	0.19	0.19	0.19	0.19	0.06	0.07	0.06	0.06	0.06	0.06	0.00										
JAT.N3	0.13	0.13	0.13	0.13	0.13	0.13	0.19	0.19	0.19	0.19	0.19	0.19	0.06	0.07	0.06	0.06	0.06	0.06	0.00	0.00									
JAT.S1	0.13	0.13	0.13	0.13	0.13	0.13	0.19	0.19	0.19	0.19	0.19	0.19	0.06	0.07	0.06	0.06	0.06	0.06	0.00	0.00	0.00								
JAT.S2	0.13	0.13	0.13	0.13	0.13	0.13	0.19	0.19	0.19	0.19	0.19	0.19	0.06	0.07	0.06	0.06	0.06	0.06	0.00	0.00	0.00	0.00							
JAT.S3	0.13	0.13	0.13	0.13	0.13	0.13	0.19	0.19	0.19	0.19	0.19	0.19	0.06	0.07	0.06	0.06	0.06	0.06	0.00	0.00	0.00	0.00	0.00						
JI.N1	0.12	0.12	0.12	0.12	0.12	0.12	0.19	0.19	0.19	0.19	0.19	0.19	0.07	0.08	0.07	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03					
JI.N2	0.12	0.12	0.12	0.12	0.12	0.12	0.18	0.18	0.18	0.18	0.18	0.18	0.07	0.08	0.07	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.01				
JI.N3	0.12	0.12	0.12	0.12	0.12	0.12	0.18	0.18	0.18	0.18	0.18	0.18	0.07	0.08	0.07	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.00	0.00			
JI.S1	0.12	0.12	0.12	0.12	0.12	0.12	0.18	0.19	0.18	0.19	0.19	0.19	0.07	0.08	0.07	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.01	0.01	0.00		
JI.S2	0.12	0.12	0.12	0.12	0.12	0.12	0.18	0.19	0.18	0.19	0.19	0.19	0.07	0.08	0.07	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.01	0.01	0.00	0.00	
JI.S3	0.12	0.12	0.12	0.12	0.12	0.12	0.18	0.19	0.18	0.19	0.19	0.19	0.07	0.08	0.07	0.07	0.07	0.07	0.03	0.03	0.03	0.03	0.03	0.03	0.01	0.01	0.00	0.00	0.00

Table 18: Summary of codon usage index of Cytochrome oxidase I (COI) gene sequences from 32 butterfly species

Species	ENC	CBI	GC	G+C2	G+C3s	G+Cc	SCUO
AM_N1	34.96	0.7	0.30	0.40	0.08	0.30	0.479
AM_N2	34.96	0.7	0.30	0.40	0.08	0.30	0.479
AM_N3	35.69	0.69	0.30	0.40	0.09	0.30	0.457
AM_S1	35.03	0.7	0.30	0.40	0.09	0.30	0.468
AM_S2	34.96	0.7	0.30	0.40	0.8	0.30	0.472
AM_S3	34.96	0.7	0.30	0.40	0.8	0.30	0.472
EH_N1	36.04	0.68	0.31	0.39	0.11	0.31	0.410
EH_N2	36.53	0.67	0.31	0.39	0.11	0.31	0.408
EH_N3	36.53	0.67	0.31	0.39	0.11	0.31	0.408
EH_S1	35.39	0.68	0.31	0.39	0.10	0.31	0.432
EH_S2	35.39	0.68	0.31	0.39	0.10	0.31	0.432
EH_S3	35.81	0.68	0.31	0.39	0.11	0.31	0.424
JA_N1	33.71	0.69	0.30	0.40	0.07	0.30	0.474
JA_N2	33.28	0.70	0.30	0.40	0.06	0.30	0.505
JA_N3	33.28	0.70	0.30	0.40	0.06	0.30	0.505
JA_S1	33.26	0.7	0.30	0.40	0.07	0.30	0.497
JA_S2	33.26	0.70	0.30	0.40	0.06	0.30	0.505
JA_S3	33.32	0.70	0.30	0.40	0.07	0.30	0.499
JAT_N1	34.47	0.67	0.31	0.40	0.10	0.31	0.423
JAT_N2	34.72	0.66	0.31	0.40	0.10	0.31	0.408
JAT_N3	34.85	0.66	0.31	0.40	0.10	0.31	0.405
JAT_S1	34.72	0.66	0.31	0.40	0.10	0.31	0.408
JAT_S2	34.89	0.66	0.31	0.40	0.10	0.31	0.407
JAT_S3	34.82	0.66	0.31	0.40	0.10	0.31	0.402
JI_N1	33.72	0.68	0.31	0.40	0.09	0.31	0.451
JI_N2	33.72	0.68	0.31	0.40	0.09	0.31	0.451
JI_N2b	33.69	0.69	0.31	0.40	0.09	0.30	0.454
JI_N3	34.57	0.68	0.31	0.40	0.09	0.31	0.422
JI_S1	33.48	0.69	0.30	0.40	0.09	0.30	0.455
JI_S2	33.56	0.69	0.31	0.40	0.09	0.30	0.456
JI_S3	34.57	0.69	0.31	0.40	0.09	0.31	0.425
JI_S3b	33.73	0.69	0.30	0.40	0.09	0.33	0.455

JA-*J.almana*,JAT-*J.altites*,JI-*J.iphita*, EH-*E.hypernmestra*, AM-*A.merione.* N-north and S-South.ENC= Effective number of codon; CBI= Codon bias index;SCUO= synonymous codon usage order

Table 19:Summary of codon usage index of NDI genes sequences from 30 butterfly species

Species	ENC	CBI	GC	G+C2	G+C3s	G+Cc	SCUO
AMN_N1	26.29	0.78	0.21	0.27	0.06	0.21	0.678
AMN_N2	26.29	0.78	0.21	0.27	0.06	0.21	0.678
AMN_N3	26.29	0.78	0.21	0.27	0.06	0.21	0.678
AMN_S1	26.29	0.78	0.21	0.27	0.06	0.21	0.678
AMN_S2	26.29	0.78	0.21	0.27	0.06	0.21	0.678
AMN_S3	26.29	0.78	0.21	0.27	0.06	0.21	0.678
EHN_N1	27.11	0.78	0.20	0.26	0.05	0.19	0.634
EHN_N2	27.48	0.78	0.20	0.26	0.06	0.20	0.672
EHN_N3	27.08	0.79	0.20	0.26	0.05	0.20	0.669
EHN_S1	27.70	0.79	0.20	0.26	0.06	0.20	0.669
EHN_S2	27.70	0.79	0.20	0.26	0.06	0.20	0.669
EHN_S3	27.70	0.79	0.20	0.26	0.06	0.20	0.668
JAN_N1	27.45	0.76	0.21	0.25	0.07	0.21	0.638
JAN_N2	31.12	0.74	0.21	0.25	0.08	0.21	0.612
JAN_N3	27.45	0.76	0.21	0.25	0.07	0.21	0.638
JAN_S1	27.45	0.76	0.21	0.25	0.07	0.21	0.638
JAN_S2	27.45	0.76	0.21	0.25	0.07	0.21	0.638
JAN_S3	27.45	0.76	0.21	0.25	0.07	0.21	0.638
JATN_N1	29.05	0.80	0.21	0.25	0.06	0.21	0.652
JATN_N2	29.05	0.80	0.21	0.25	0.06	0.21	0.652
JATN_N3	29.05	0.80	0.21	0.25	0.06	0.21	0.652
JATN_S1	29.05	0.80	0.21	0.25	0.06	0.21	0.652
JATN_S2	29.05	0.80	0.21	0.25	0.06	0.21	0.652
JATN_S3	29.05	0.80	0.21	0.25	0.06	0.21	0.652
JIN_N1	25.64	0.84	0.19	0.25	0.05	0.19	0.703
JIN_N2	26.05	0.84	0.20	0.25	0.05	0.20	0.712
JIN_N3	26.32	0.84	0.20	0.25	0.05	0.20	0.700
JIN_S1	27.23	0.82	0.21	0.25	0.06	0.20	0.677
JIN_S2	27.23	0.82	0.21	0.25	0.06	0.20	0.677
JIN_S3	27.23	0.82	0.21	0.25	0.06	0.20	0.677

JA-*J.almana*,JAT-*J.altites*,JI-*J.iphita*, EH-*E.hypernmestra*, AM-*A.merione.*
N-north and S-South.ENC= Effective number of codon; CBI= Codon bias index;SCUO= synonymous codon usage order

4.7.1 Phylogenetic analysis of COI gene using different tree building methods

For COI gene, 32 individuals of nymphalidae species comprising five species were analyzed using *Libythea celtis* (Accession no - NC_016724.1) as out-group. Different tree building methods like Maximum Parsimony, Maximum likelihood and Bayesian Inference were used (Figures 10 - 12).

From the three tree building method, two major clade was obtained. One clade containing the *Junonia* species and the other clade constitute *Ariadne merione* and *Elymnias hypernmestra*. All the tree building methods used gives congruent result. The 20 species belonging to the genus *Junonia* formed a monophyletic clade with a bootstrap support value of MP (97.2), ML (92) and BI (1.00). While the *A.merione* species and *E.hypernmestra* species forms separate clade with bootstrap support value of MP (72.9), ML (79), BI (0.80).Except for two species of *J. iphita* (S3 and N3) all the species forms a monophyletic clade. These two species from two different location forms a basal clade for *J. atlites* and the rest of the *J.iphita* species with a bootstrap support of MP (43.9), ML (60), BI (0.63).The rest of the *J.iphita* and *J.atlites* forms sister clade with a high bootstrap value of MP (98.7), ML (84), BI (0.98). Within the species *J.iphita*ofS1 and S2 forms the basal clade for JI- N1, N2, N2b, and S3b. The intra specific variation of *J.atlites* (JAT) and *A.merione* (AM) was low; it resulted in polytomies in the phylogenetic tree. For *J.almana* (JA) it forms two clade with 100% bootstrap value, one clade consist of JA-N1, S3, S1 (N-north, S-South) and the other clade comprises JA-N2, S2, N3. But the *E.hypernmestra* species forms two separate clusters. One clade consisted of North Brahmaputra samples (EH- N1, N2 and N3) and the other clade consisted of the South Brahmaputra species (EH- S1, S2 and S3) in all the three tree building methods.

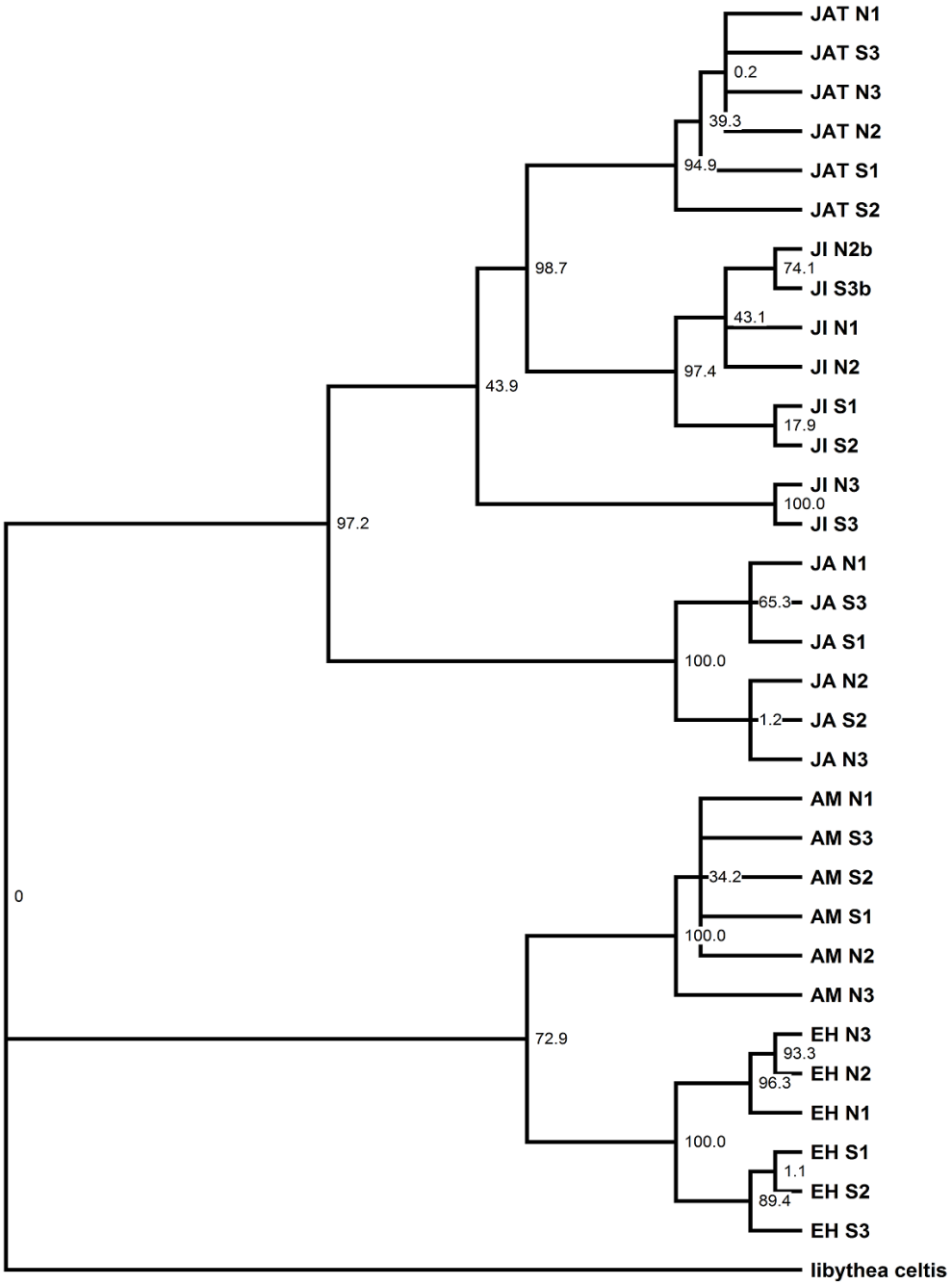


Figure 10: Maximum parsimony tree constructed for COI gene using PAUP software

JA-*J.almana*,JAT-*J.altites*,JI-*J.iphita*, EH-*E.hypernmestra*, AM-*A.merione*.

. N-North Brahmaputra and S-South Brahmaputra. Outgroup- *L.celtis*.

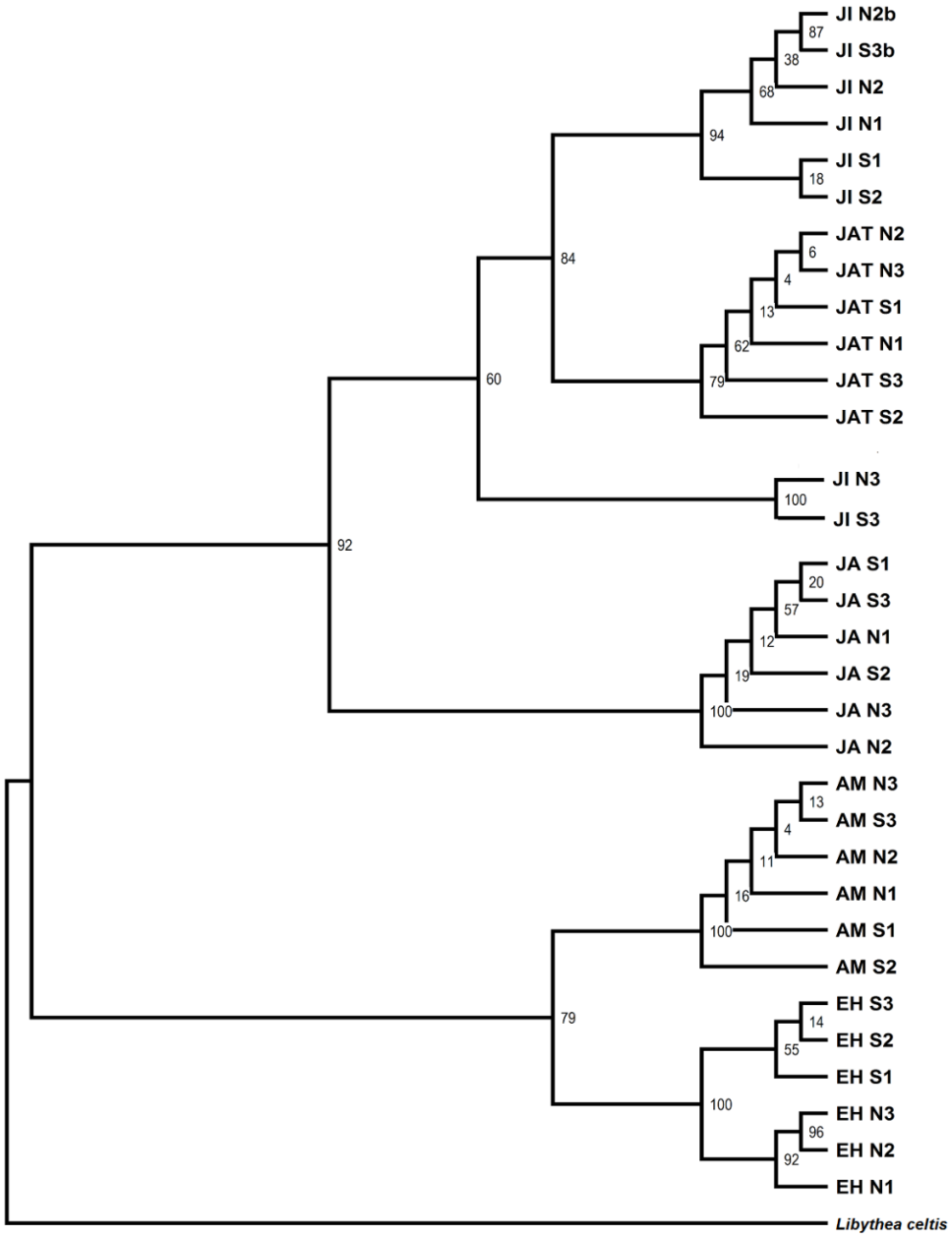


Figure 11: Maximum Likelihood tree constructed for COI gene using PhyML Software

JA-*J.almana*,JAT-*J.altites*,JI-*J.iphita*, EH-*E.hypernmestra*, AM-*A.merione*.

. N-North Brahmaputra and S-South Brahmaputra. Outgroup- *L.celtis.*

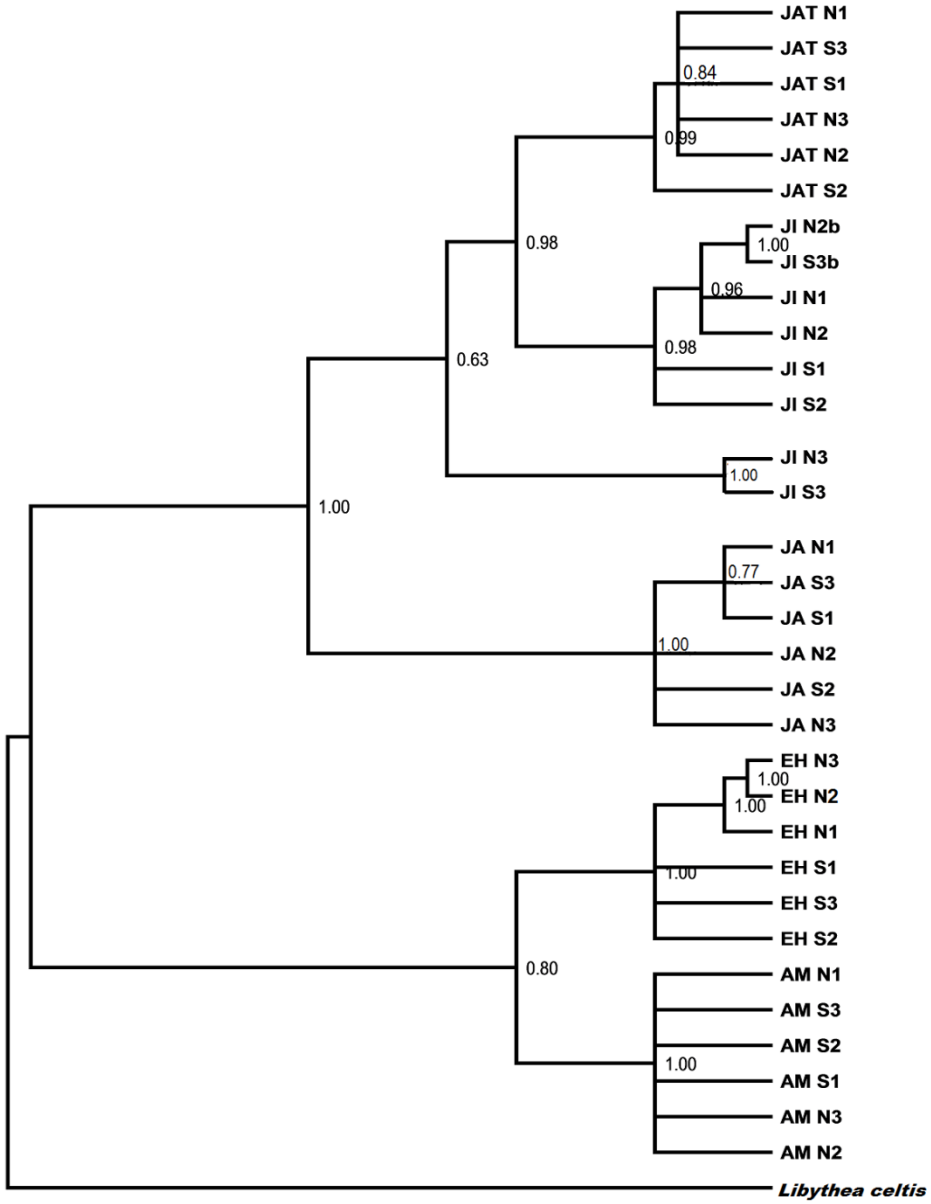


Figure 12: Bayesian tree constructed for COI geneusing MrBayes

JA-*J.almana*,JAT-*J.altites*,JI-*J.iphita*, EH-*E.hypernmestra*, AM-*A.merione*.

. N-North Brahmaputra and S-South Brahmaputra. Outgroup- *L.celtis.*

4.7.2 Phylogenetic analysis of NDI gene using different tree building methods

Using NDI gene a phylogenetic tree was constructed for the selected Nymphalidae species, with three tree building methods - Maximum Parsimony, Maximum likelihood and Bayesian Inference (Figures 13 - 15). *L.celtis* (NC_016724.1) was used as out-group.

In all the tree building methods, *E.hypernmestra* formed the basal clade. It was further sub divided into two sub-clades. One clade constitutes all the *Junonia* species and the other clades constitute the *A.merione* species with the bootstrap support of 66.1 (MP), 69 (ML), 0.89 (BI). Within the Junonia species, *J. almana* forms the basal clade and further sub divided to *J.atlites* and *J.iphita* with a high bootstrap support of 96.2(MP), 92(ML), 1.00(BI). The species *J. atlites, J. almana* and *A.merione* with 50% consensus tree result in polytomies because of high sequences similarity within the species level. For *J.iphita* inMP tree JI-NI form the basal clade with a high bootstrap value of 82.2, it diverge into two more sub clade; one clade comprising of JI- N2, N3 and the other clade the south species but with low bootstrap value of 22. For BI tree JI species separate into two clusters with bootstrap support of 0.77 for the North and South. Within the North, the species were similar with low bootstrap support resulting in polytomies. For ML tree JI is divided into two clusters with bootstrap value of 86, one contained N2, N3 and the other clade comprising of S1, S2, S3 and N1. The tree topology as a whole was similar for all the three tree building methods, but within the species level there was difference in the clustering pattern. Finally, for the *E.hypernmestra* species there was a separate clade for the North and South Brahmaputra species with 100% (1.00) bootstrap support value indicating intraspecific variation within the *E.hypernmestra* species.

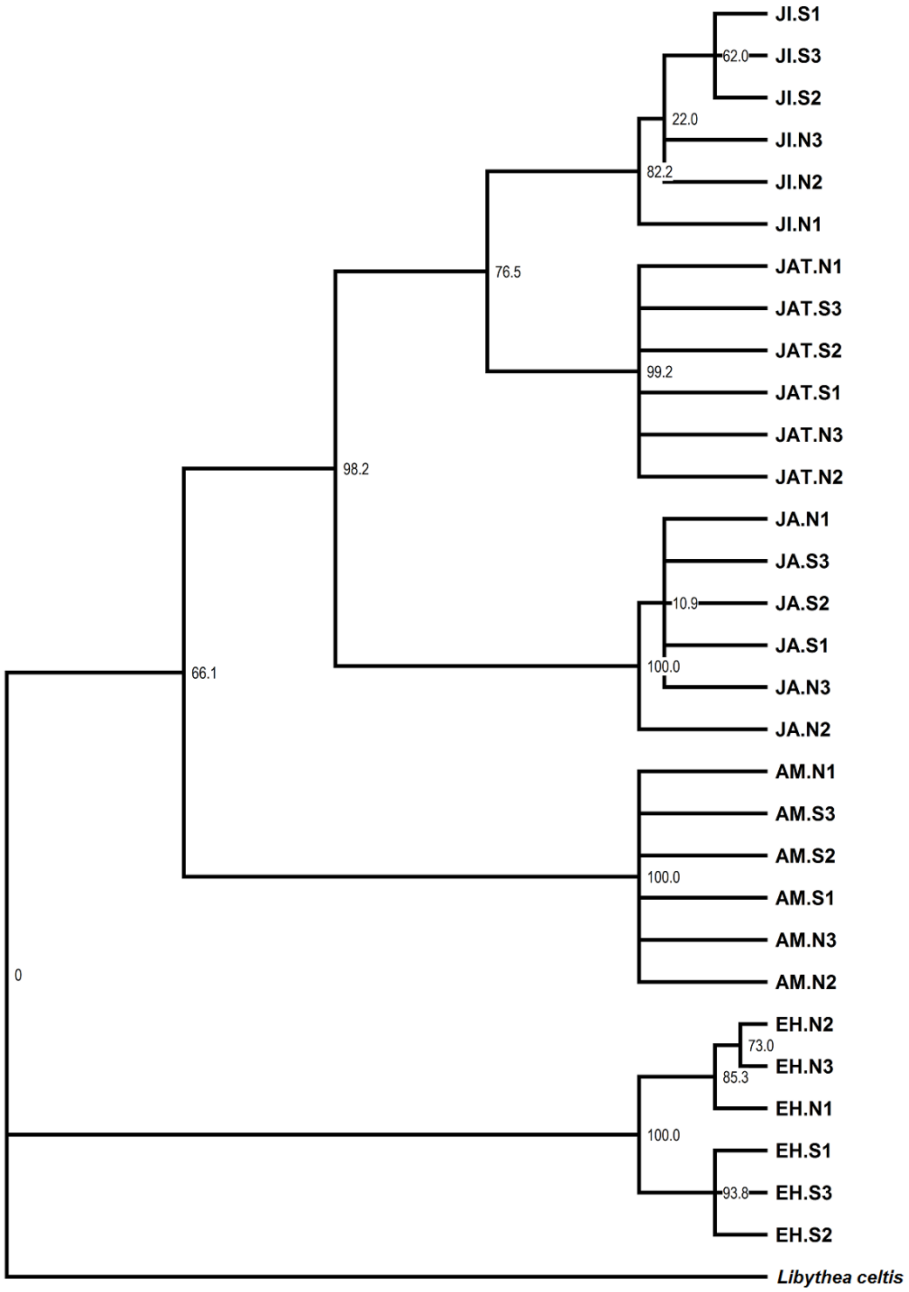


Figure 13: Maximum Parsimony tree constructed for NDI gene using PAUP Software

JI- *J. iphita,* JAT- *J. atlites,* JA- *J. almana,* AM- *A. merione,* EH-*E.hypernmestra.* N-North Brahmaputra and S-South Brahmaputra. Outgroup- *Libythea celtis*

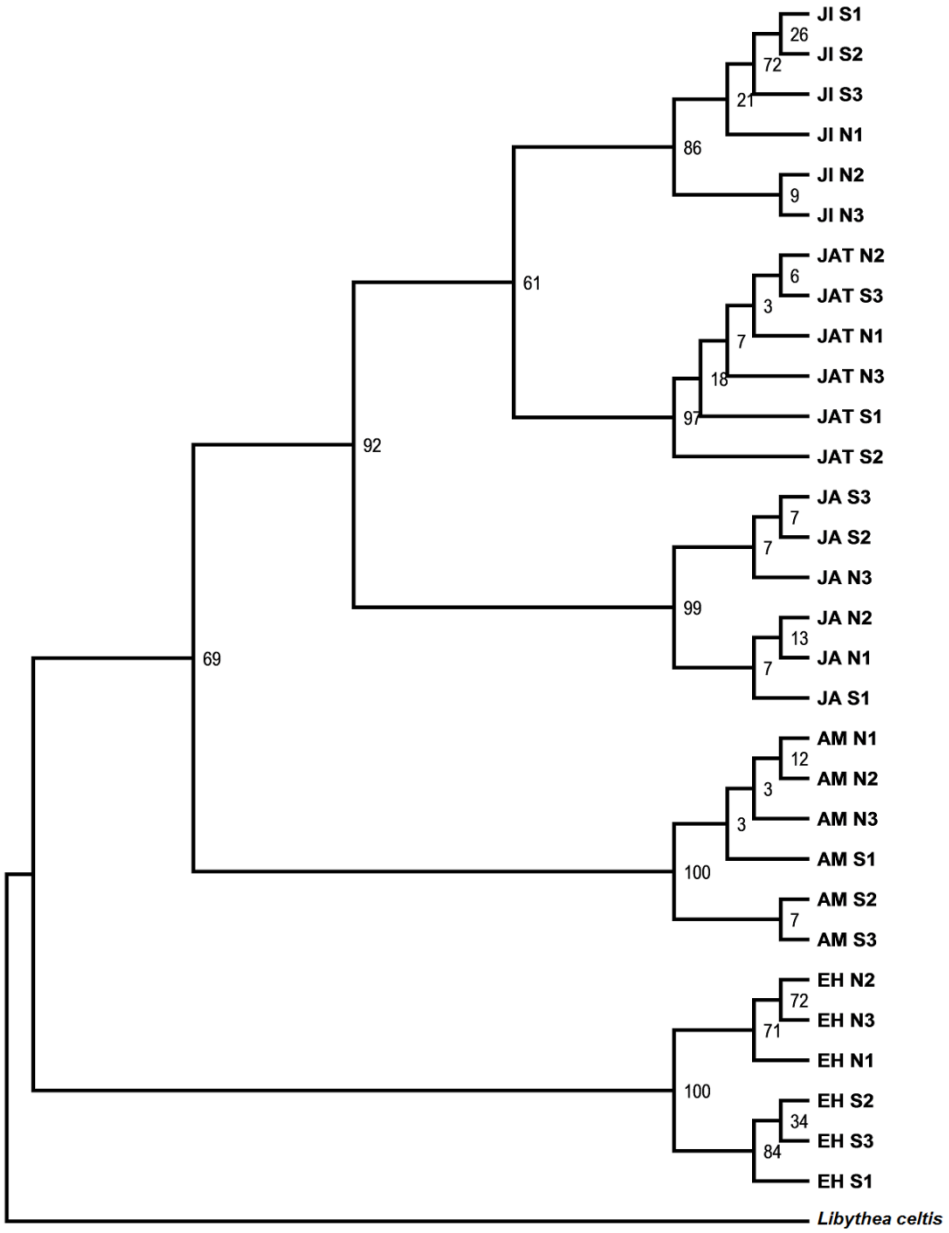


Figure 14: Maximum Likelihood tree constructed for NDI gene using PhyML Software

JI- *J. iphita*, JAT- *J. atlites*, JA- *J. almana*, AM- *A. merione*, EH-*E.hypernmestra.* N-North Brahmaputra and S-South Brahmaputra.Outgroup- *Libythea celtis*

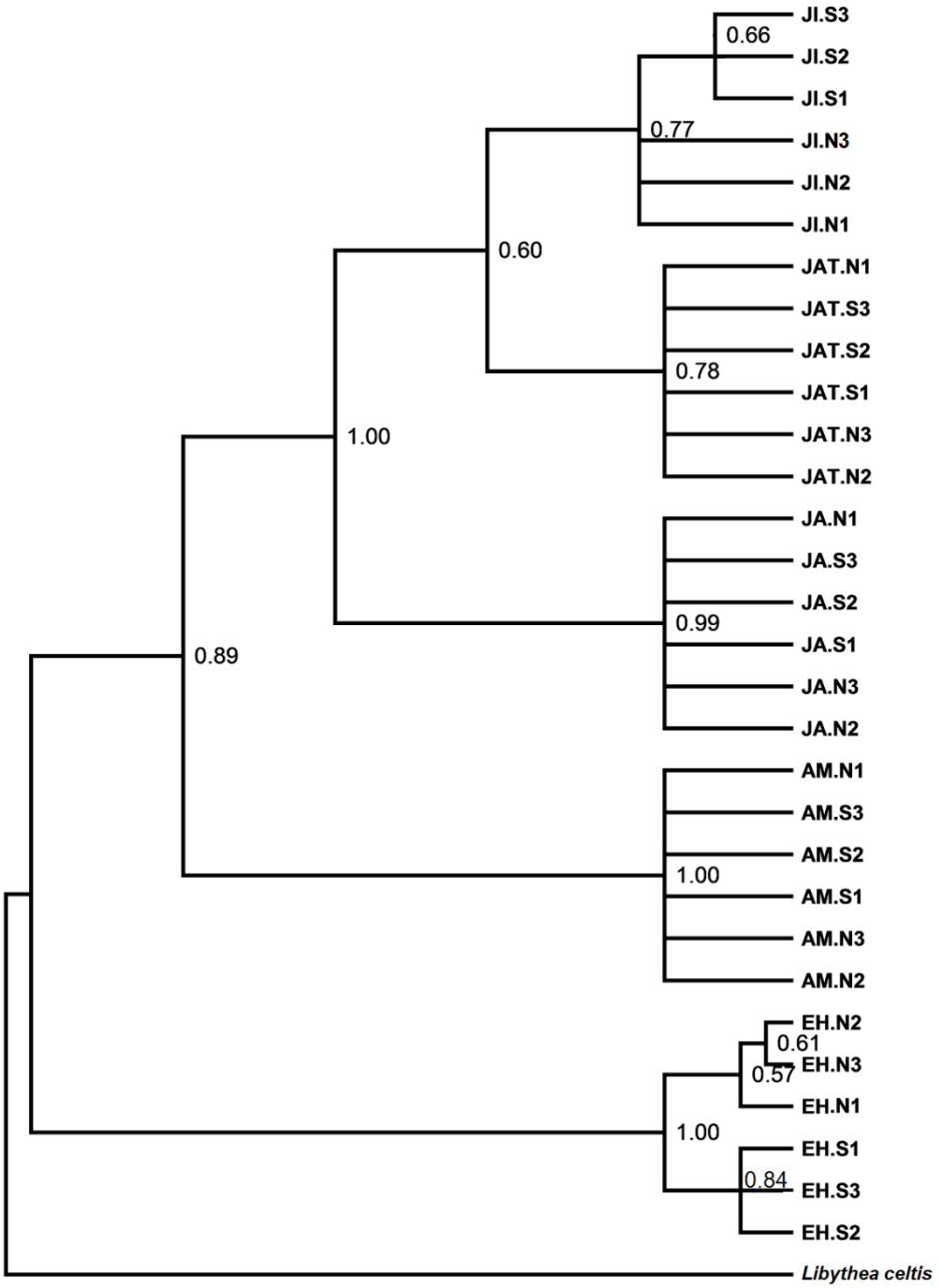


Figure 15: Bayesian tree constructed for NDI gene using MrBayes

JI- *J. iphita,* JAT- *J. atlites,* JA- *J. almana,* AM- *A. merione,* EH-*E.hypernmestra.* N-North Brahmaputra and S-South Brahmaputra.Outgroup- *Libythea celtis*

.

4.8. Phylogenetic analysis of COI and NDI combined sequence

The mitochondrial COI and NDI sequence were combined together for building phylogenetic tree using Bayesian Inference (Figure 16).

The Phylogenetic tree obtained for combined COI and NDI sequence is very similar with COI phylogenetic tree compared to NDI phylogenetic tree. It differ slightly only in the bootstrap value. And for the analysis of COI gene sequence,for the species *J.iphita* eight individuals were use, but since we don't have the NDI data for two species JI-N3 and JI-S3, in the combined data analysis we did not include that and in the final combined data we had 30 individuals. For each species the branching pattern is in concordant with the COI data. For the species *A.merione*, it results in polytomy even in the combined data. J.atlites S2 forms the base for all the *J.atlites* species with high bootstrap value. *J.almana* divided into two clades, one clade comprises species from N1, S3, S1 and the other clade consist N2, N3 and S3. *J.iphita* forms two clades, *J.iphita* from S1 and S2 forms one clade and the remaining species from S3, N1, N2 and N3 forms the other clade.*E.hypernmestra* separate into two clade, North Brahmaputra clade and South Brahmaputra clade.

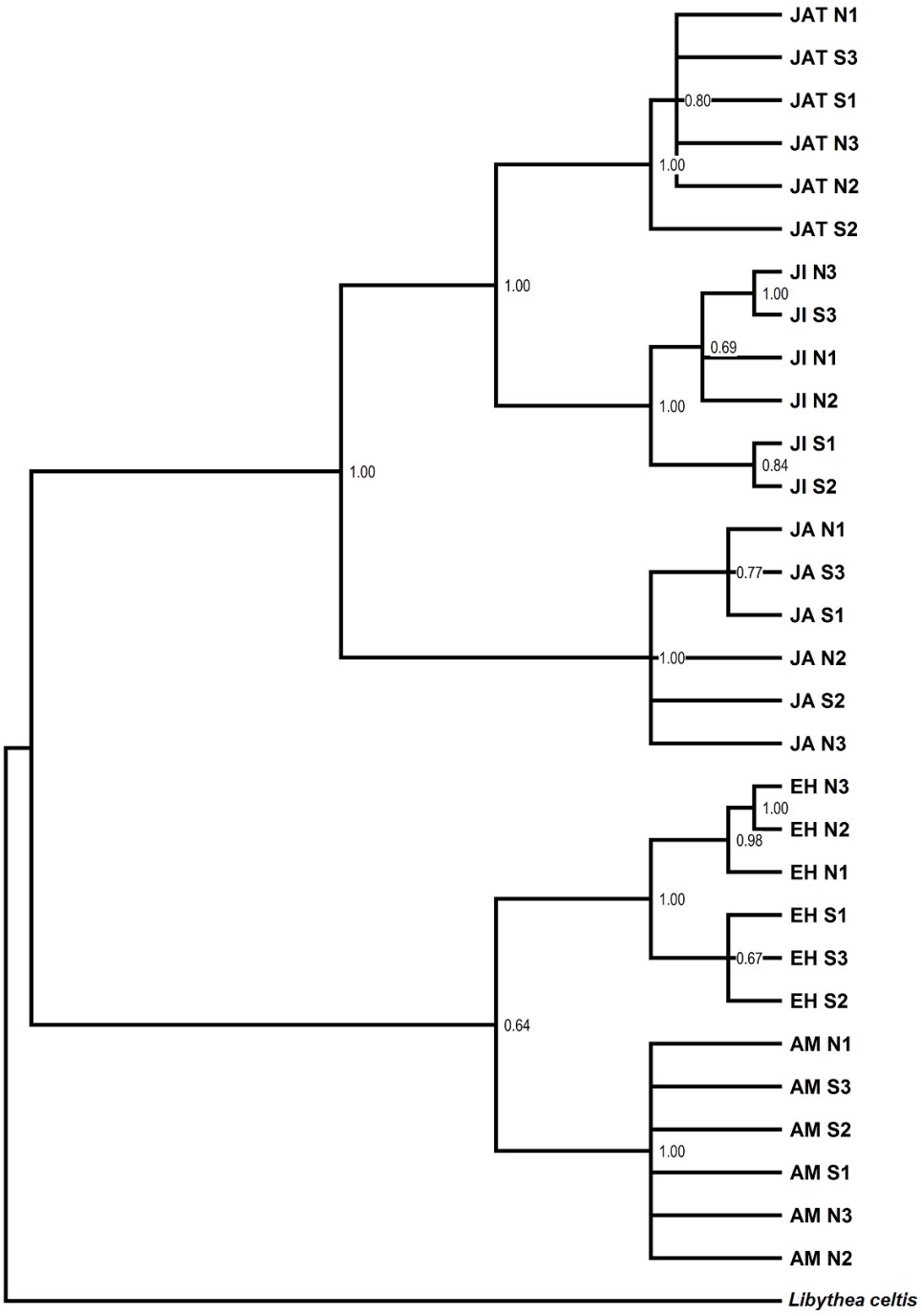


Figure 16: Bayesian tree constructed for combined COI and NDI gene

JI- *J. iphita*, JAT- *J. atlites*, JA- *J. almana*, AM- *A. merione*, EH-*E.hypernmestra.* N-North Brahmaputra and S-South Brahmaputra.Outgroup- *Libythea celtis*

4.9. Genome organization of the two *Junonia iphita*from both north and south Brahmaputra River

Junonia iphita from both south and north Brahmaputra River were selected for whole mitochondrial genome analysis. The complete mitogenome of *J. iphita* from south Brahmaputra River (JI_MZ) is a circular molecule of 15,433 bp in length and thenear complete mitogenome of *J. iphita* from north Brahmaputra River (JI_AZ) is 14,892 bp in length. Both the species consisted of typical 13 protein-coding genes (PCGs), 22 transfer RNA genes (tRNAs) and 2 ribosomal RNA genes (rrnL and rrnS) and one major non-coding AT-rich region except for JI_AZ where AT-rich region is not sequenced (Tables 20-21). Arrangement of the gene and their orientation is similar to other butterfly species. The overall nucleotide compositions are significantly biased toward AT (80.5 % for South and 80 % for North Brahamaputra species) which are similar to other work on nymphalids in *Eumenis autonoe* 79.1 % (Kim *et al*., 2010) and for *Parathyma sulpitia* 81.9 % (Tian *et al*., 2012).The mitogenome nucleotide skewness for JI_MZ and JI_AZ were AT-skew=-0.003; GC-skew=0.215 and AT-skew=-0.002; GC-skew=-0.213**,** respectively(Table 22).

Comparison of base composition (total basepair 14,894) excluding the Dloop region for both the *J.iphita*shows 14,833 conserved sites which is 99.59 %.An analysis of the Mutation/SNP sites revealed that there were altogether 51 sites along the entire length of the genome where SNPs were observed. This amounted to 0.4 % of the entire mitochondrial genome. These SNP sites were infered by correlating with the number of variable sites.

In both the *Junonia iphita* (JI_MZ and JI_AZ) intergenic spacers are dispersed throughout the whole mitochondrial genome which differs slightly between the two species (14 for South and 12 for North Brahamaputra species). The largest intergenic spacer is

located between tRNAG and *ND2* and consists of 52 bp in both the species. For JI_MZ the A+T rich region is 546 bp in length containing ATAGA motif followed by a 18 bp poly-T stretch, two microsatellite-like $(TA)_9$elements and 8 bp poly-A stretch immediately upstream of trnM gene and the A+T content in the sequence of the A+T-rich region is 96%, also within the range observed in the completely sequenced lepidopteran insects, with the value from 89.17% in *A. melete* to 98.25% in *P. atrilineata.*

The highest RSCU was UUA (L) for both the species the codon.All of the remaining codons with RSCU greater than 2 have U or, particularly, A in codon position 3 (Table 23 and 24).All PCGs are initiated by a typical ATN codon, except for the cox1 gene, which uses unusual CGA as start codon as observed in most of the other sequenced nymphalids (Shi *et al*., 2013; Wang *et al*., 2011).Seven protein coding genes start with ATG (*cox2, atp6, cox3, nad4, nad4 l, cytb* and *nad1*), five with ATT (*nad2, atp8, nad3, nad5* and *nad6*) in both the species.Eight genes use complete termination codon (TAA) for the species JI_MZ but for JI_AZ nine gene use complete termination codon (TAA), three genes (cox1, cox2 and nad5) end with single T; nad4 and nad1 use TA as stop codon for JI_MZ species. While for JI_AZ three genes have incomplete stop codon T (*cox1, cox2* and *nad1*) and one gene (*nad4*) ended with TA (Figure 17).

The mt DNA includes standard 22 tRNA genes. The 22 tRNAgenes sequence range from 61-71 bp. All the secondary structure of tRNA gene folds into typical cloverleaf structure except $tRNA^{Ser(AGN)}$ and $tRNA^{Arg}$. The*tRNAs* possess 7 bp aminoacyl stems but other portions of tRNAs are variable in length, 7 -9 bp anticodon loop, 4 - 7 bp anticodon stems, DHU stem 3 – 4 bp, DHU loop 4 – 11 bp, TΨC loops 4 – 10 bp and finally TΨC stem 3 – 5 bp. A total of 21 mismatch base pairs are found in the tRNA stem region. 18 of them

are G-U mismatch and 3 U-U mismatch at $tRNA^{Ala}$, $tRNA^{Leu(UUR)}$ in the Aminoacyl stem and at $tRNA^{Ser(UCN)}$ in anticodon stem. Like other Lepidoptera, the two rRNAs genes are found in *J.iphita* mitogenomes (1332 bp *lrRNA* and 786 bp srRNA for JI_AZ; 1330 bp *lrRNA* and 785 bp *srRNA* for JI_MZ) and are also significantly biased towards AT nucleotides (85 % and 82.7% for *lrRNA*; 84.9% and 85.3 % for *srRNA*) which are well within the range of other Lepidoptera species (Zhang *et al*., 2012).

4.9.1Intraspecies variation within *Junonia iphita* AZ (North) and *Junonia iphita* MZ (South Brahamaputra species)

The two *J. iphita* from the north and south Brahmaputra River were compared to study the presence of any intraspecific variations among them. And comparison between these two sequences reveals minor variation. We excluded the D loop portion in our comparisons since the D loop of *J.iphita* AZ was not sequenced. From Table 25 we observed that the nad1 region of JI_Az ended with TA while that of JI_Mz ends with only T. In the case of nad5 JI_Az ends with an incomplete stop codon T and JI_Mz ends with a complete stop codon TAA.

Analysis of the tRNA structure revealed similar structures except for the tRNA of Glycine. Here it was observed that JI_Mz has an A-A mismatch in its anti-codon stem. This mismatch was absent in the case of JI_Az Minor variations were observed in the number of nucleotide residues A, T, G and C; in the number of intergenic spacers and also in the number of base pairs of nad5, cytB, rrnaL and rrnaS. Variations in terms of base pair numbers were also observed in tRNAs.

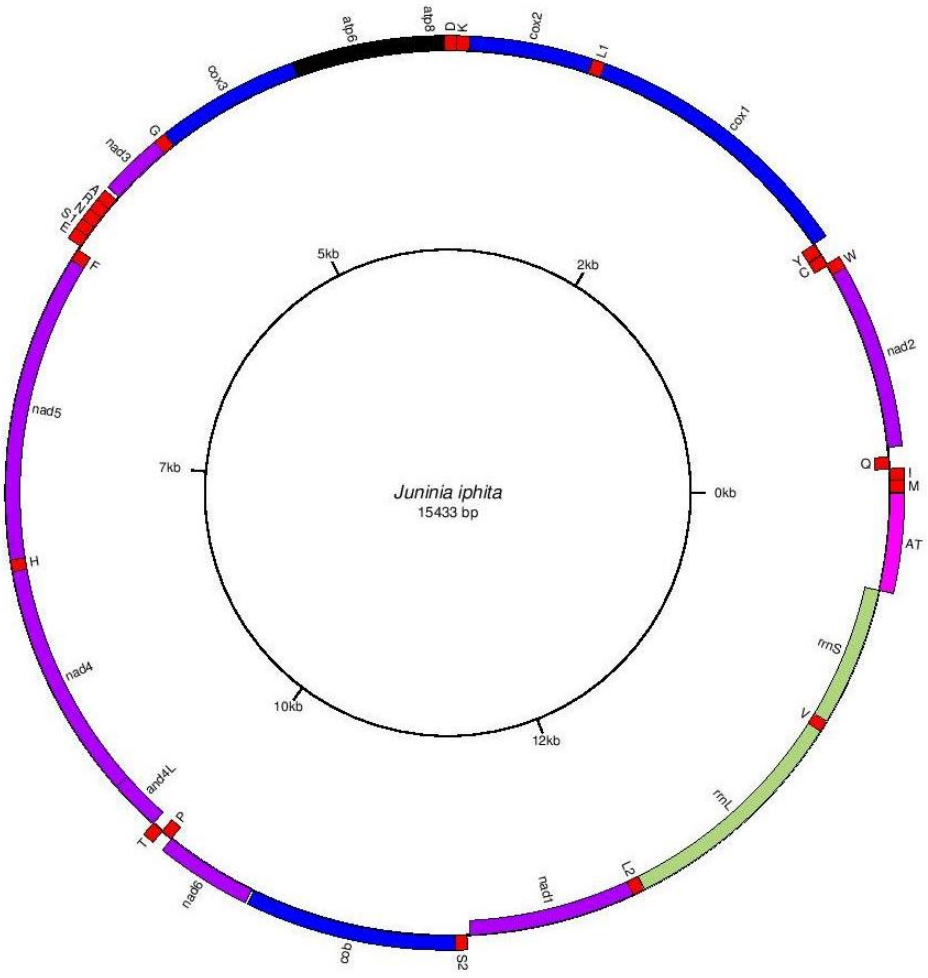


Fig 17: Complete mitochondrial genome of *Junonia iphita* (JI_MZ Accession no SRP053690)

Table 20: Organization of the complete mitochondrial genome in *Junonia iphita* of Mizoram (JI_MZ Accession no SRP053690)

Feature	Strand	Start	End	Size	Anticodon	Intergenic length*	Start	End
Trnm(Met)	+	1	68	68	CAT			
Trni(Ile)	+	70	134	65	GAT	1		
Trnq(Gln)	-	132	200	69	TTG	-3		
Nad2	+	253	1266	1014		52	ATT	TAA
Trnw(Trp)	+	1265	1331	67	TCA	-2		
Trnc(Cys)	-	1324	1388	65	GCA	-8		
Trny(Tyr)	-	1395	1462	68	GTA	6		
Cox1	+	1465	2995	1531		2	CGA	T
Trnl(Leu)(UUR)	+	2996	3063	68	TAA	0		
Cox2	+	3064	3739	676		0	ATG	T
Trnk(Lys)	+	3740	3810	71	CTT	0		
Trnd(Asp)	+	3811	3875	65	GTC	0		
Atp8	+	3876	4040	165		0	ATT	TAA
Atp6	+	4034	4711	678		-7	ATG	TAA
Cox3	+	4711	5499	789		-1	ATG	TAA
Trng(Gly)	+	5502	5568	67	TCC	2		
Nad3	+	5569	5918	350		0	ATT	TAA
Trna(Ala)	+	5942	6006	65	TGC	23		
Trnr(Arg)	+	6006	6069	64	TCG	-1		
Trnn(Asn)	+	6070	6136	67	GTT	0		
Trns1(Ser)(AGN)	+	6136	6196	61	GCT	-1		
Trne(Glu)	+	6197	6265	69	TTC	0		
Trnf(Phe)	-	6290	6356	65	GAA	28		
Nad5	-	6359	8090	1732		3	ATT	T
Trnh(His)	-	8091	8157	67	GTG	0		
Nad4	-	8157	9496	1340		-1	ATG	TA
Nad4l	-	9497	9784	288		0	ATG	TAA
Trnt(Thr)	+	9787	9851	65	TGT	2		
Trnp(Pro)	-	9852	9916	65	TGG	0		
Nad6	+	9919	10446	528		2	ATT	TAA
Cob	+	10461	11616	1154		14	ATG	TAA
Trns2(Ser)(UCN)	+	11615	11681	67	TGA	-2		
Nad1	-	11699	12636	938		17	ATG	TA
Trnl1(Leu)(CUN)	-	12638	12706	69	TAG	1		
Rrnl	-	12708	14037	1330		1		
Trnv(Val)	-	14038	14102	65	TAC	0		
Rrns	-	14103	14887	785		0		
A+T		14888	15433	546		0		

Table 21: Organization of the complete mitochondrial genome in *Junonia iphita* of Sikkim (JI_AZAccession no SRP053322)

Feature	Strand	Start	End	Size	Anticodon	Intergenic length*	Start	End
Trnm(Met)	+	1	68	68	CAT			
Trni(Ile)	+	69	133	65	GAT	0		
Trnq(Gln)	-	131	199	69	TTG	-3		
Nad2	+	252	1265	1014		52	ATT	TAA
Trnw(Trp)	+	1264	1330	67	TCA	-2		
Trnc(Cys)	-	1323	1389	67	GCA	-8		
Trny(Tyr)	-	1394	1461	68	GTA	4		
Cox1	+	1464	2994	1531		2	CGA	T
Trnl(Leu)(UUR)	+	2995	3062	68	TAA	0		
Cox2	+	3063	3738	676		0	ATG	T
Trnk(Lys)	+	3739	3809	71	CTT	0		
Trnd(Asp)	+	3810	3874	65	GTC	0		
Atp8	+	3875	4039	165		0	ATT	TAA
Atp6	+	4033	4710	678		-7	ATG	TAA
Cox3	+	4710	5498	789		-1	ATG	TAA
Trng(Gly)	+	5501	5567	67	TCC	2		
Nad3	+	5568	5917	350		0	ATT	TAA
Trna(Ala)	+	5941	6005	65	TGC	23		
Trnr(Arg)	+	6005	6068	64	TCG	-1		
Trnn(Asn)	+	6069	6135	67	GTT	0		
Trns1(Ser)(AGN)	+	6135	6195	61	GCT	-1		
Trne(Glu)	+	6196	6264	69	TTC	0		
Trnf(Phe)	-	6293	6359	67	GAA	28		
Nad5	-	6360	8093	1734		0	ATT	TAA
Trnh(His)	-	8094	8160	67	GTG	0		
Nad4	-	8160	9499	1340		-1	ATG	TA
Nad4l	-	9499	9786	288		-1	ATG	TAA
Trnt(Thr)	+	9789	9853	65	TGT	2		
Trnp(Pro)	-	9854	9918	65	TGG	0		
Nad6	+	9921	10448	528		2	ATT	TAA
Cob	+	10463	11618	1156		14	ATG	TAA
Trns2(Ser)(UCN)	+	11617	11683	67	TGA	-2		
Nad1	-	11701	12638	938		17	ATG	T
Trnl1(Leu)(CUN)	-	12640	12708	69	TAG	1		
Rrnl	-	12710	14041	1332		1		
Trnv(Val)	-	14042	14106	65	TAC	0		
Rrns	-	14107	14892	786		0		

Table 22: Nucleotide composition for thirteen Protein Coding Genes region in the two *Junonia iphita* species

Species name	All codon						First codon					Second codon						Third codon					
	T	C	A	G	A+T	G+C	T	C	A	G	A+T	T	C	A	G	A+T	G+C	T	C	A	G	A+T	G+C
JI_AZ(north Brahmaputra	40.4	11.8	40.1	7.6	80.5	19.4	39	11.6	39.5	10.2	78.5	42	13.4	36.7	7.6	78.7	21	40	10.5	44.1	5.1	84.1	15.6
JI_MZ(south Brahmaputra	40.1	12.2	39.9	7.9	80	20.1	40	14.1	36.6	9.4	76.6	42	11	41.7	4.9	83.7	15.9	38	11.4	41.3	9.3	79.3	20.7

Table 23: Relative synonymous codon usage of *Junonia iphita*AZ

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	332	1.57	UCU(S)	46	1.33	UAU(Y)	213	1.61	UGU(C)	29	1.38
UUC(F)	91	0.43	UCC(S)	31	0.9	UAC(Y)	52	0.39	UGC(C)	13	0.62
UUA(L)	429	4.02	UCA(S)	100	2.9	UAA(*)	199	1.57	UGA(W)	71	1.8
UUG(L)	24	0.23	UCG(S)	14	0.41	UAG(*)	54	0.43	UGG(W)	8	0.2
CUU(L)	53	0.5	CCU(P)	32	0.99	CAU(H)	82	1.59	CGU(R)	5	0.61
CUC(L)	23	0.22	CCC(P)	38	1.18	CAC(H)	21	0.41	CGC(R)	3	0.36
CUA(L)	99	0.93	CCA(P)	55	1.71	CAA(Q)	86	1.43	CGA(R)	22	2.67
CUG(L)	12	0.11	CCG(P)	4	0.12	CAG(Q)	34	0.57	CGG(R)	3	0.36
AUU(I)	364	1.62	ACU(T)	50	1.27	AAU(N)	312	1.56	AGU(S)	25	0.72
AUC(I)	85	0.38	ACC(T)	31	0.79	AAC(N)	88	0.44	AGC(S)	9	0.26
AUA(M)	261	1.84	ACA(T)	72	1.83	AAA(K)	275	1.6	AGA(S)	37	1.07
AUG(M)	22	0.16	ACG(T)	4	0.1	AAG(K)	69	0.4	AGG(S)	14	0.41
GUU(V)	30	1.22	GCU(A)	33	1.83	GAU(D)	42	1.47	GGU(G)	23	0.84
GUC(V)	8	0.33	GCC(A)	13	0.72	GAC(D)	15	0.53	GGC(G)	4	0.15
GUA(V)	49	2	GCA(A)	26	1.44	GAA(E)	51	1.48	GGA(G)	74	2.72
GUG(V)	11	0.45	GCG(A)	0	0	GAG(E)	18	0.52	GGG(G)	8	0.29

Table 24: Relative synonymous codon usage of *Junonia iphita* MZ

Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU	Codon	Count	RSCU
UUU(F)	334	1.58	UCU(S)	45	1.31	UAU(Y)	209	1.59	UGU(C)	29	1.38
UUC(F)	90	0.42	UCC(S)	32	0.93	UAC(Y)	54	0.41	UGC(C)	13	0.62
UUA(L)	428	4.01	UCA(S)	100	2.91	UAA(*)	198	1.57	UGA(W)	73	1.78
UUG(L)	24	0.23	UCG(S)	14	0.41	UAG(*)	55	0.43	UGG(W)	9	0.22
CUU(L)	53	0.5	CCU(P)	32	1	CAU(H)	80	1.54	CGU(R)	5	0.63
CUC(L)	23	0.22	CCC(P)	36	1.13	CAC(H)	24	0.46	CGC(R)	2	0.25
CUA(L)	100	0.94	CCA(P)	55	1.72	CAA(Q)	86	1.46	CGA(R)	22	2.75
CUG(L)	12	0.11	CCG(P)	5	0.16	CAG(Q)	32	0.54	CGG(R)	3	0.38
AUU(I)	366	1.63	ACU(T)	50	1.27	AAU(N)	315	1.57	AGU(S)	25	0.73
AUC(I)	82	0.37	ACC(T)	31	0.78	AAC(N)	87	0.43	AGC(S)	10	0.29
AUA(M)	262	1.85	ACA(T)	73	1.85	AAA(K)	276	1.6	AGA(S)	36	1.05
AUG(M)	22	0.15	ACG(T)	4	0.1	AAG(K)	70	0.4	AGG(S)	13	0.38
GUU(V)	29	1.2	GCU(A)	33	1.81	GAU(D)	42	1.47	GGU(G)	24	0.89
GUC(V)	8	0.33	GCC(A)	13	0.71	GAC(D)	15	0.53	GGC(G)	3	0.11
GUA(V)	49	2.02	GCA(A)	27	1.48	GAA(E)	51	1.48	GGA(G)	75	2.78
GUG(V)	11	0.45	GCG(A)	0	0	GAG(E)	18	0.52	GGG(G)	6	0.22

Table 25: Variation detected in the two Mitogenome of *J. iphita* species

Composition of mt genome	*J. iphita* AZ (north)	*J. iphita* MZ (south)
Number of IS* and its size	12 (148 bp)	14 (152 bp)
Total overlapping region	10	9
IL* in tRNA(Ile)	1	0
Size of tRNA(Cys)	67 bp	65 bp
Size of tRNA (Phe)	67 bp	65 bp
Size of nad5	1734	1732
IL in nad5	3	0
nad5 end (stop codon)	T	TAA
Size of cytb	1156 bp	1154 bp
Size of rrnaL	1332 bp	1330 bp
rrnaS	786 bp	785 bp
nad1 end (stop codon)	TA	T
IL in tRNA(Tyr)	6	4
tRNA(Gly) mismatch	Absent	A-A in AC* stem
Percentage of A	40.1	39.9
Percentage of T	40.4	40.1
Percentage of C	11.8	12.2
Percentage of G	7.6	7.9

IL*= Intergenic length; IS*= Intergenic spacer; AC*= Anticodon

Table 26: Species and voucher number of whole genome sequences downloaded from NCBI for the present study

Species name	length	Accession no
Junonia almana	15,256 bp	NC_024407.1 GI:658608477
Junonia orithya	15,214 bp	NC_022697.1 GI:556505845
Yoma sabina	15,330 bp	NC_024403.1 GI:658608408
Melitaea cinxia	15,171 bp	CM002851.1 GI:662986786
Kallima inachus	15,183 bp	NC_016196.1 GI:357018090
Apatura metis	15,236 bp	NC_015537.1 GI:333236263
Argynnis children	15,131 bp	NC_024415.1GI:658608625
Acraea issoria	15,245 bp	NC_013604.1GI:280978083
Parthenos Sylvia	15,249 bp	NC_024417.1GI:658608662
Dichorragia nesimachus	15,355 bp	NC_024409.1GI:658608515
Libythea celtis	15,164 bp	NC_016724.1GI:372292061
Melanitis leda	15,122 bp	NC_021370.1GI:511347533
Calinaga davidis	15,267 bp	NC_015480.1GI:331746893
Polyura arja	15,363 bp	NC_024408.1GI:658608494
Parantica sita	15,211 bp	NC_024412.1GI:658608570
Euploea mulciber	15,166 bp	NC_016720.1GI:372292005
Protantigius superans (Outgroup)	15,248 bp	NC_016016.1 GI:347600337

4.9.2 Phylogenetic tree constructed for whole mitochondrial genome of Nymphalidae

A phylogenetic tree was constructed for the whole mitochondrial genome using three tree building methods MP and ML (Figures 18 and 19).

In order to place the mt DNA sequence of *J.iphita* AZ in relation to other butterfly mitogenomes and to its phylogenetic relationship with other Nymphalidae a dataset containing concatenated 13 protein coding genes was generated. The final alignment contains 13,269 sites in the matrix with 18 in-group and one out-group taxa. Of these sites 7547 were conserved, 5713 were variables and 3963 were parsimony informative sites. The MP and ML analysis gives overall similar topology except in the branching pattern within the subfamily Satyrinae, Calinaginae and Charaxinae. In MP tree *M.leda* and *C.davidis* forms a sister clade with *P.arja* at the basal clade, but in ML tree it is vice versa with *M.leda* and *P.arja* forms sister clade with *C.davidis* at the base but the bootstrap support is higher in the MP tree. In MP analysis *L.celtis* and *D.nesimachus* clade together with low bootstrap value (35.7) whereas in ML L celtis is forms a basal clade for *D.nesimachus*. Two heliconiinae species crustered together and limentidinae form the basal clade in both MP and ML analysis. Apaturinae form the basal clade for the Nymphalinae subfamily. All the nymphylinae shows similar topology in both the tree. Within the nymphalinae the four *Junonia* species cluster together, which further sub divided into two clade, one clade consist of the *J.almana* and *J.orithya*; the other clade comprises the two *J.iphita* with high bootstrap support.

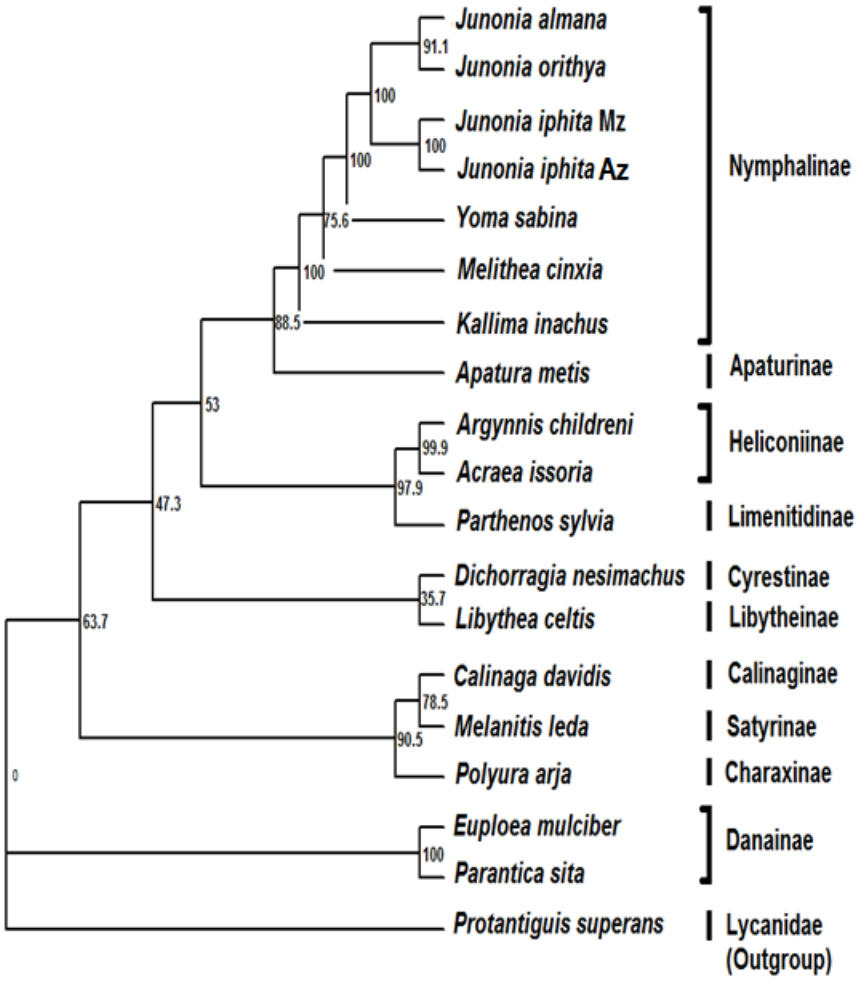


Figure 18: Maximum Parsimony tree of whole mitochondrial genome using PAUP Software

Number of bootstrap replicates = 1000 Starting tree(s) obtained via stepwise addition random, Number of trees held at each step during stepwise addition, Branch-swapping algorithm: tree-bisection-reconnection (TBR)

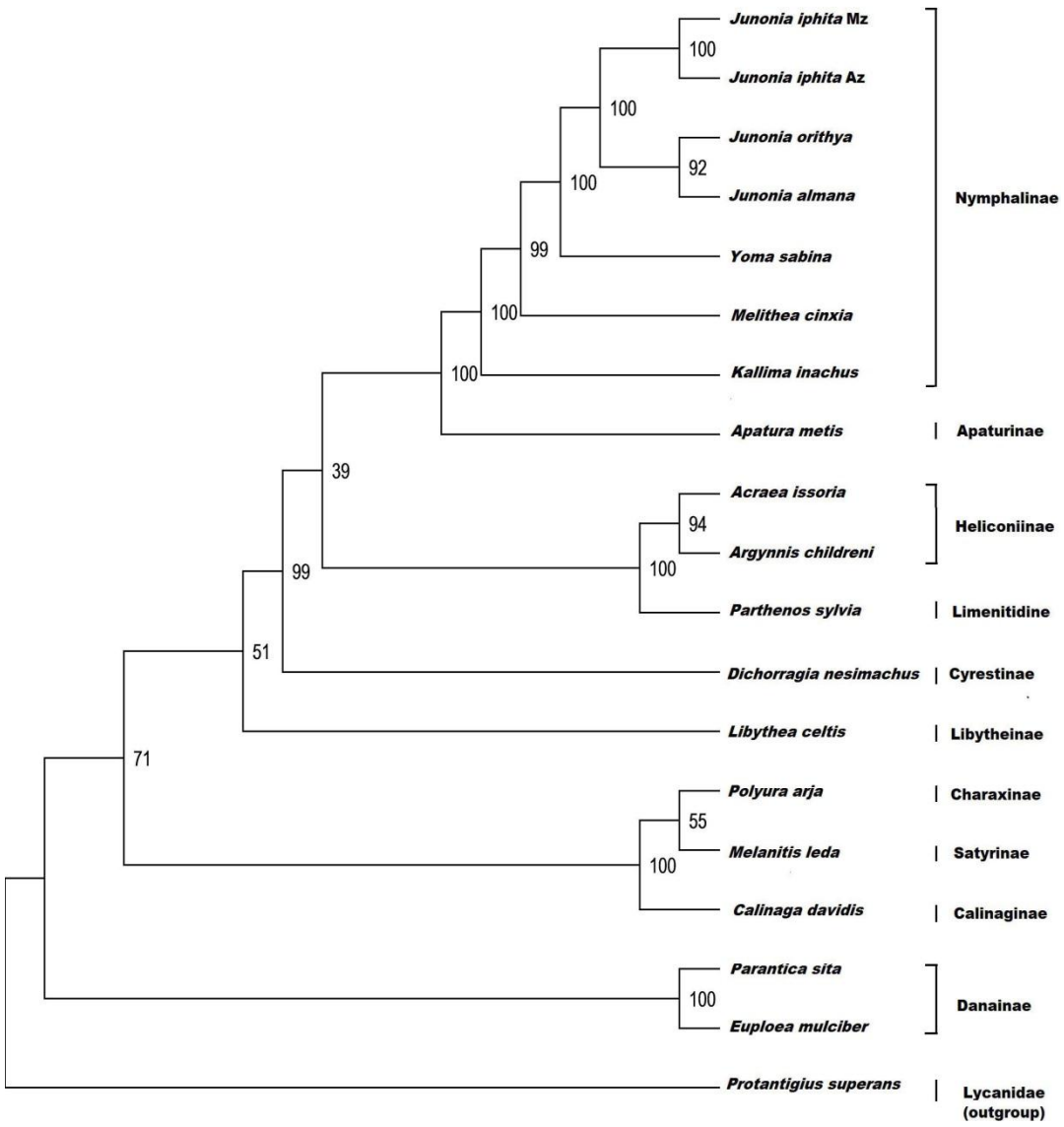


Figure 19: Maximum Likelihood tree of whole mitochondrial genome using RaxML Software

ML analysis inferred from all codon positions of 13 PCGs with model GTRGAMMAI.Bootstrap support values are indicated at each node.

4.10. PCR-Restriction fragment length polymorphism of COI gene

Amplified PCR (COI) products were subjected to restriction digestion with *TaqI* (T/CGA), *AluI* (AG/CT) and *RsaI* (GT/AC) (Fermentas, Thermo Scientific) enzymes using the protocol of Sambrook *et al.*(1989).

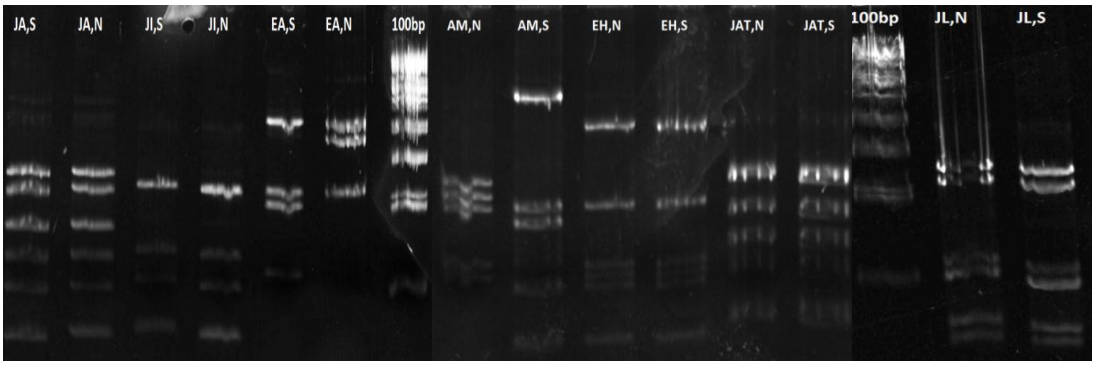
The different restriction enzymes resulted with total bandsof 59 from bothNorth and South Brahmaputra River. The number of polymorphic bands is 39fromNorth and 34 from south, respectively. The total no of bands (TNB), number of polymorphic bands (NPB), percentage of polymorphic bands (P %), Polymorphic information content (PIC), Resolving power (RP) and Marker index (MI) for individual primers are shown in Table 27. The average percentage of polymorphic bands is 65.67 and 57.93. The average PIC value is 0.3 for north and 0.32 for south. The restriction enzyme, *Alu I* is a frequent cutter, so it result in more number of bands with 28 RE cutting sites for both the sites and number of polymorphic bands 19 and 16, respectively. The resolving power of marker is also highest with a value of 7.42 for both the sites. The result in tabular form and the gel picture (Figure20 a-c) of the three restriction enzymes shows that the variation between the species from north and south Brahmaputra River is very low.

Table 27: Restriction enzymes and Polymorphism detected in nymphalidae

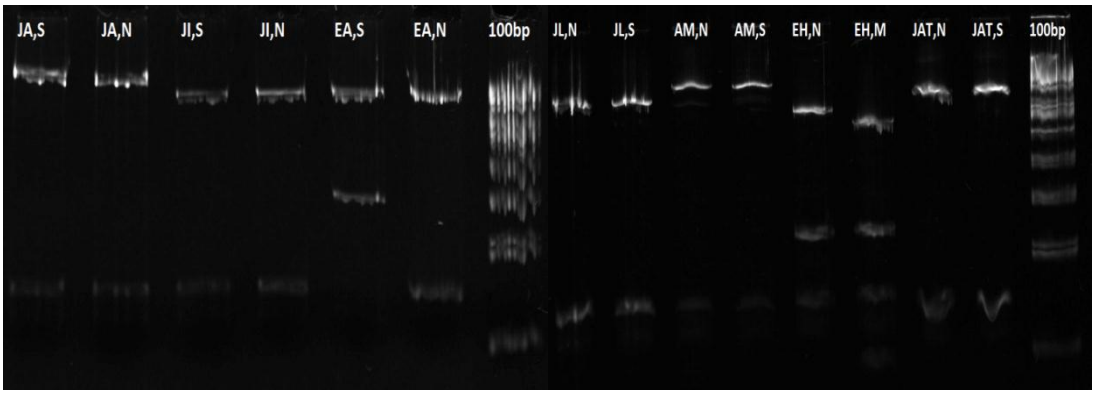
RE	**TNB**		**NPB**		**% POLY**		**PIC**		**RP**		**MI**	
sites	N	S	N	S	N	S	N	S	N	S	N	S
Alu 1	28	28	19	16	67.85	57.14	0.29	0.33	7.42	7.42	0.19	0.18
Taq I	15	15	10	10	66.66	66.66	0.3	0.31	4	4.3	0.2	0.2
Rsa I	16	16	10	8	62.5	50	0.31	0.33	4.57	4	0.19	0.16
Total	59	59	39	34	65.67	57.93	0.3	0.32	5.33	5.24	0.19	0.18

Total number of bands (TNB), Number of polymorphic bands (NPB), Polymorphism Percentage (P %), Polymorphism information content (PIC), Resolving Power (RP) and Marker index (MI) of butterfly population. N- North, S-South.

a) Alu I



b) *Taq I*



c) *RsaI*

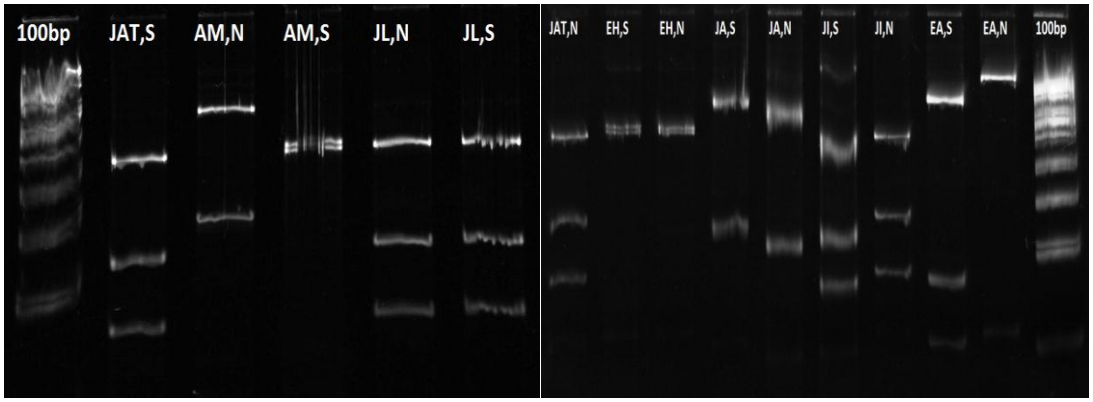


Figure 20:RFLP amplification products of three Restriction Enzymes (a) *AluI*(b) *TaqI* and (c) *RsaI*

4.11.Physico-chemical characterization ofDistal-less (dll) genes

Two distal-less genes were isolated from *Elymnias hypernmestra* (4AzD and 4AsD) species from north and south Brahmaputra respectively. Their nucleotides were sequenced and using EMBL-EBI's 'EMBOSS Transeq' software, the sequenced nucleotides were translated to their corresponding Protein sequences. Three distal-less proteinsequences i.e. *Bicyclus anynana*, *Heliconius cydno galanthus* and *Junonia coenia* were retrieved from the Swiss-Prot database. The results of their primary sequence analysis are given in (Tables28 - 30).

The computed isoelectric point (pI) indicated that the distal-less proteins of most of the samples required abasic buffer solution with a pH range of 8 to 11.*Heliconius cydno galanthus* requiredan acidic buffer solution with a pH range of 4 to 6 while *E.hypernmestra* (4AzD)required a buffer solution around the acidic and basic pH borderline i.e. 6 to 8.On the basis of instability index, the distal-less protein of only*E.hypernmestra*(4AzD),with a score of less than 40 indicated it was stable in a test tube environment. Extinction coefficient of the distal-less proteins at 280 nm ranged from 2045-60975 M–1 cm–1. *E. hypernmestra* (4AzD) scored the least and *Bicyclus anynana*scored the highest.

The computed aliphatic index inferred that, in comparison to the other samples, the distal-less proteins of*E.hypernmestra* (4AsD and 4AzD), both with an AI>70 may be the only samples among the selected five that are able to withstand a wide range of temperatures. With respect to amino acid composition, there seems to be an almost equal distribution of both non-polar and polar amino acids with a slight edge towards the polar amino acids.

Serine (Ser)consistently comprises the highest percentage of amino acid content. The average molecular weight of the distal-less protein calculated was 19643.8 Da. (Table 30).The calculated GRAVY index for all the distal-less samples, showed extremely low values. This inferred that the protein was hydrophilic.The server SOSUI classifiedall the distal-less proteins as soluble proteins and therefore no trans-membrane regions were identified. Cys_Rec software computation showed that both species of *E.hypernmestra*had a number of cysteine residues none showed any indication of forming a disulphide (SS) bond. However, *Bicyclus anynana* with three cysteine residues was shown to have the best potential for forming a disulphide (SS) bond. Cys_Rec calculated CYS113 as the most probable to form a disuphide (SS) bond with a score of 1.4.

The 3D structure was predicted for all the five distal-less proteins and then analyzed using Rasmol, a protein molecular viewer. It showed presence of cysteine residues in all the five samples (Fig. 21). However, disulphide bond (S-S) formation seemed unlikely except in the case of *Bicyclus anynana*. With respect to secondary structures, the distal-less protein of *Junonia coenia,Heliconius cydno galanthus* and *E.hypernmestra* (4AzD) are shown to contain helix (pink) and turns (blue). There are no beta sheets and with respect to cysteine molecules the latter two are shown to possess one cysteine residue each while that of *Junonia coenia*has two cysteine residues. However, they are spaced far apart from each other to disallow any disulphide (SS) bond formation.*E.hypernmestra*(4AsD) also has a helix (pink) and turns (blue). It also has beta sheets (yellow) and one cysteine residue. *Bicyclus anynana* has all three secondary structures i.e. helix (pink), turns (blue) and sheets (yellow). It also has three cysteine residues. One of which (CYS113) was computed to form a disulphide (SS) bond by Cys_Rec software.

Table 28. Distal-less protein sequences of butterflies retrieved from Swiss-Prot database

Accessionnumber	Organism name	Sequence description
AAW70235.1	*Heliconius cydno galanthus*	Distal less protein
AFM73657.1	*Bicyclus anynana*	Distal less protein
AAB32450.1	*Junonia coenia*	Distal-less protein

Table 29. Distal-less protein sequences isolated and sequenced from *Elymnias hypernmestra*

Organism I.D	Organism Name	Sequence description
4AsD	*Elymnias hypernmestra*(North Brahmaputra)	Distal less protein
4AzD	*Elymnias hypernmestra*(South Brahmaputra)	Distal less protein

Table 30. Parameters of Butterfly distal less genes computed using Expasy's ProtParam tool

Accession Number/ Organism I.D	Sequence length	M. wt	pI	– R	+ R	EC	II	AI	GRAVY	Cysteine Residues
AAW70235.1	75	7670.2	5.96	5	4	13980	83.50	33.73	-0.923	1
AFM73657.1	358	38584.8	9.43	19	29	60975	56.51	48.55	-0.872	3
AAB32450.1	197	21936.5	9.85	11	23	19035	47.92	49.09	-0.971	2
4AsD	131	15002.7	9.16	5	10	19160	48.57	120.53	0.574	4
4AzD	132	15026.6	7.94	5	6	2045	32.43	124.77	0.597	3

M. wt.,Molecular weight; pI, Isoelectric point; –R, Number of negative residues; +R, Number of positive residues; EC, Extinction coefficient at 280 nm; II, Instability index.; AI, Aliphatic index; GRAVY,Grand Average Hydropathy (Sequences selected from Swiss-prot and sequences isolated from *E.hypernmestra*of north and south of Brahmaputra).

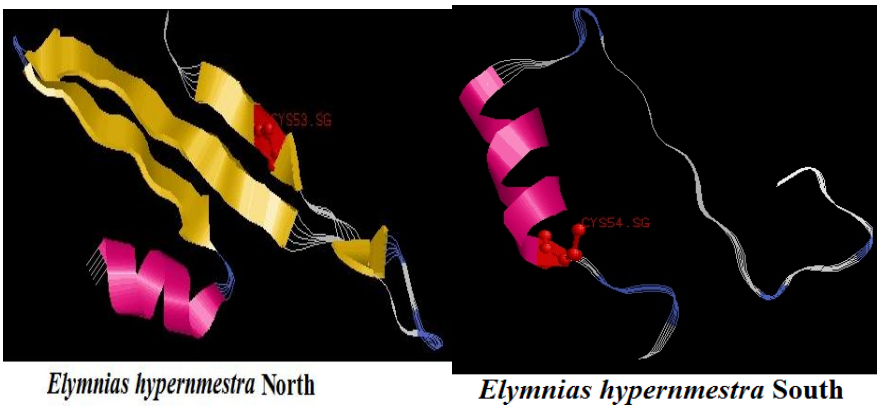


Figure 21: Homology based 3D structure of distal-less protein

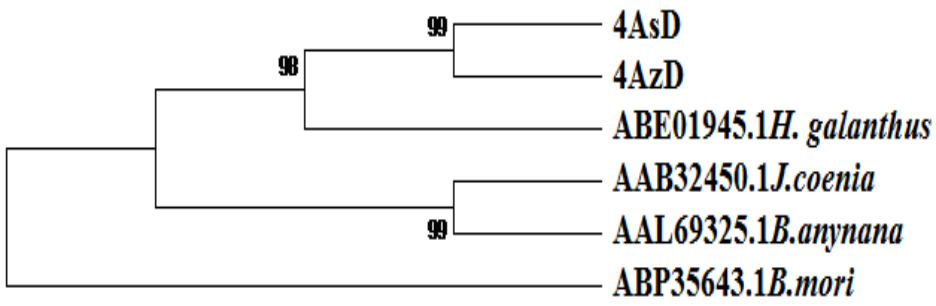


Figure 22: Maximum parsimony tree constructed for distal-less (dll) gene

4Asd- *Elymnias hypernmestra* (North Brahmaputra), 4AzD- *Elymnias hypernmestra*(South Brahmaputra). Outgroup- *Bombyx mori*

The phylogenetic tree was constructed for distal-less (dll) gene using Maximum parsimony methods (Fig. 22), four dll gene sequences were downloaded from NCBI, *Bombyx mori* was used as outgroup. The tree was divided into two cluster using *B.mori* as outgroup. One clade consists of *J.coenia* and *B.anynana* with high bootstrap value (99). The other clade consists of *H.cyndo galanthus* and the *E.hypernmestra* species. We can see that two species of *E.hypernmestra*forms a cluster with a high bootstrap support (99).*H.cyndo galanthus* forms the base for the *E.hypernmestra*with bootstrap support (98). Here the two species of *E.hypernmestra*from north and south Brahmaputra river does not forms any separate cluster.

IV. DISCUSSION

For studying the intraspecific variation across Northeast India, eleven species were selected. But after preliminary analysis based on field work, availability of species, species which fulfill the four criteria for samples collection which are; First the species should be distributed only in the Oriental region (India and Southeast Asia); Secondly they must be easy to identify (to prevent taxonomic errors); Thirdly the selected species should not exhibit local migration (low dispersal ability); Finally they must be locally common thereby facilitating ease of collection. Also preliminary work done was done using COI gene, neighbor joining tree was built using mega5, the five species which are *Junonia almana, Junonia iphita, Junonia atlites, Ariadne merione* and *Elymnias hypernmestra* shows intraspecific variability and was thus selected for further studies to see whether it will be useful for detecting the genetic difference across the north and south Brahmaputra River.

The *K*-dominance curve is apowerful tool for measuring abundance trends in communities over time (Lambshead *et al*., 1983). *K*-dominance curves are the cumulative ranked abundance against species rank. The logic behind the use of these curves as indicators is that only the subset of species that can tolerate perturbation will thrive and the rest will decline or disappear. Thus, the steepest and most elevated curve shows the lowest diversity and the most perturbed system state (Soki wiki, 2014; Voronkov *et al.,* 2013). So in our study it is observed that type 3 forest had the lowest diversity. While type 1 and type 2 forest shows high diversity with very little difference among the two forest types. The two diversity indices (Shannon and Simpson) for type 1 and type 2 are very similar with a slight difference, type 1 shows higher value than type 2 and type 3 is the lowest among them. So in Berger-

parker dominance type 3 has the highest dominance value among the three forest type while type 1 and 2 forest had almost similar dominance.

The low dominance value of type 1 and 2 forest indicate that butterfly species are more or less evenly distributed in terms of abundance compared to type 3 forest. The present results indicate an availability of host plants for each butterfly species and efficient resource sharing by each species in type 1 and 2 forests.Several studies revealed that habitat specificity is directly linked to the availability of host plants for larvae and adults (Thomas, 2005). High species dominance in type 3 forest can be link to availability of specific host plants for a particular species of butterfly.

The two diversity indices (Shannon and Simpson) for type 1 and type 2 are very similar with a slight difference, type 1 shows higher value than type 2 and type 3 is the lowest among them. So in Berger-parker dominance type 3 has the highest dominance value among the three forest type while type 1 and 2 forest had almost similar dominance. And members of the Nymphalidae were always dominant in the tropical region because most of the species are polyphagous in nature, consequently helping them to live in all the habitats. Additionally, many species of this family are strong, active fliers that might help them in searching for resources in large areas (Majumder *et al.*, 2012; Zothansangi *et al*., 2011).

Thus, the diversity of Nymphalidae in Mizoram is influence by temperature, humidity, rainfall and availability of host plant for the larval and adult stage. Butterflies and their caterpillars are dependent on specific host plants for foliage, nectar and pollen as their food. Butterflies are often considered opportunistic foragers, which visit a wide variety of available flowers. Thus butterfly diversity reflects overall plant diversity, especially that of

herbs and shrubs in the given area (Feltwell, 1986; Tiple *et al.,* 2007).Thus butterfly population dynamics are generally influenced by environmental factors like temperature, rainfall, humidity, photoperiod, and variations in the availability of food resources and vegetation cover for the larval and adult stage (Anu *et al.,* 2009; Tiple *et al.,* 2009; Rajagopal *et al.,* 2011).

Random amplified polymorphic DNA-PCR (RAPD)method has shown to function well with insects (Kambhampati*et al.,* 2002), as these organisms have a relatively large genome size, which increases the probability of finding polymorphism (Dezfouli *et al.,* 2002). Black *et al.* (1992) has suggested that RAPD can be used to measure clonal diversity even in insects such as aphids which reproduce asexually but show extensive polymorphism and Lou *et al.* (1998) has reported variation in RAPD within and between geographic populations of the wheat stem sawfly (*Cephus cinctus* Norton, 1872). And also because of its low cost, easy to get result and its ability to detect polymorphism we choose RAPD as a tool for studying the intraspecific variation.

The dendrogram of RAPD clearly shows that the rapd markers can be used as a marker to distinguish at the species level, each of the species were separately cluster and rapd is a useful marker for detecting polymorphism. *E. hypernmestra* (EH) species from both sites of river Brahmaputra forms separate cluster with a strong bootstrap support of 100%. Within the south, EH S3 species showed low bootstrap support which may be because in the south region intraspecific variation is less. For *J. iphita* there was separate cluster for the site A (North Brahmaputra) and Site B (South Brahmaputra), but the north sample forms the basal line for the south sample with high bootstrap support of 90%. In *A. merione*, the north and south were separated with a bootstrap support of 58% and the south forms the basal clade to

the north samples. In the case of *J. atlites* and *J. almana,* the species from both sites were not clustered separately, there might be horizontal gene transfer occurring with the sample from both sides of the Brahmaputra. In general the intraspecific variation is low within and between the samples from both sides.

One way analysis of ANOVA was done using binary data and PIC value of RAPD data. Both results give a p value less than 0.05, which is considered statistically significant. It shows that RAPD can be used to detect polymorphism among and within species. Among the species intraspecific variation was less compared to interspecific variation.

Using RAPD inter- and intraspecific variations in insects as well as differentiation of very closely related species or even geographical populations have been clarified: Lepidoptera – Indian meal moth, *Plodiainterpunctella* (Hübner) (Dowdy and McGaughey, 1996), European corn borer, *Ostrinia nubilalis* (Hübner) (Pornkulwat *et al.,* 1998), gypsy moth, *Lymantria dispar* (L.) (Garner and Slavicek, 1996).Al-Barrak *etal.* (2004)successfully discriminated the closely related sibling species of *T. calamagrostidis*, *T. longicornis* and *T. petiolata* using RAPD-PCR.

Two mitochondrial genes- COI and NDI were used to study the relationship within and between the selected five Nymphalidae species across Northeast India using different parameters. The nucleotide composition was calculated and result in a strong A+T bias in both the genes which is similar with other previous studies on Butterfly.A strong A/T bias in mitochondrial DNA has also been reported in many mitochondrial genome studies; including studies of the Lepidoptera (Kim *et al*., 2006; 2009). Also, several previous studies advocated the use of mtDNA sequences as aids in identifying closely related species (Sperling

andHickey, 1994; Kruse and Sperling, 2001), and many studies explicitly stated the appropriateness of mtDNA in resolving the relationships of closely related butterfly species (Brunton and Hurst, 1998; Caterino and Sperling, 1999; Rand *et al.*, 2000; Monteiro and Pierce, 2001). A broad sampling strategy is imperative for closely related species, as has been emphasized by Funk (1999).

Protein coding genes of mitochondrial DNA are relatively conserved within insects (Simon, 1991), coding sequences contains numerous sites where synonymous substitutions can occur. Sperling and Hickey, 1994 reasoned that 'approximately one-third of sites would potentially provide fine-grained information about evolution at the species level and below. Transition transversion ratio changes according to evolutionary distance.The transition transversion ratio are best estimated between very closely related species in order to observe the instantaneous substitution pattern, but may cause problems as the small number of nucleotide changes causes a large variation in the estimation (Yang and Yoder, 1999). Studies on *Drosophila melanogaster* species group have shown that there is no strong transition-transversion bias (Wolstenholme and Clary, 1985; Sharp and Li, 1989), the reason is that the estimates were obtained from comparisons of distantly related species. Sequence comparison within the species and among closely related species showed strong transition which is also observed in mammals (deBruijn, 1983; Satta *et al.*, 1987). So analysis with closely related species found more transitions than transversions; transversions is more in distantly related species due to multiple substitutions (Tamura, 1992; Wakeley, 1996; Simon *et al.*, 1994) which supports our studies. Our studies in transition transversion bias using maximum composite likelihood estimate also shows similar result.The species we selected are closely related comprising of three genera within the family nymphalidae. So within and

between species level transition was higher compared to transversion. COI shows higher transition ratio compare to NDI genes.

Relative synonymous codon usage(RSCU)was used to investigate variance in codon usage. Synonymous codons of one codon family were expected to have the same frequency. The RSCU represented the observed frequency of a codon divided by the frequency expected if all synonymous codons if a particular amino acid were used equally. Thus, the value of RSCU reflected the codon usage bias in a single codon family (Sharp *et al.*, 1986). In the present study, RSCU values give an estimate of the preference for alternative synonymous codons (Sharp *et al.*, 1986). RSCU>1 indicates codons which were used more frequently than expected whereas RSCU<1 and RSCU=1 indicates a codon is used as expected by random usage. The codon with highest RSCU was UUA (L). All of the remaining codons with RSCU greater than 2 have U or, particularly, A in codon position 3. The AT content of insect mtDNA is high, and codon usage tends toward A or U at the third position (Wolstenholme and Clary, 1985; Crozier and Crozier,1993; Brower and DeSalle, 1998).The selective pressure on the amino acid composition would be most notable on the hydrophobic amino acids and could cause substantial degrees of parallel evolution even in the polypeptide chain (Foster and Hickey, 1999). In our studies more frequently encoded hydrophobic amino acids were UUU (F), UUA (L), AUU (I), AUA (M) and GUU (V).

The p-distance or evolutionary distance was calculated for all the species using the kimura-2-parameters for both the genes, within the species the genetic distance were quite low and range from 0.0 to 0.02, but for one species of *Junonia iphita* the maximum genetic distance (p-distance) was 0.07 which is the highest intraspecific distance in the studied species. The maximum p-distance among the five selected species is 0.15 for COI gene and

for NDI gene it is 0.20. Our results on inter and intraspecific variation are comparable with others studies on Nymphalidae by Gaikwad (2012) and Ashfaq *et al.* (2013). In their studies for some species intraspecies nucleotide divergent is less than 0.1%. The relationship between geographical distance and the level of intraspecific divergence was not strong. Geographical distance is often associated with an increased genetic divergence, but the increase is too small to impede the identification of species

Effective number of codon (ENC) depends upon nucleotide composition of a gene it ranged from 20(strongest codon bias) to 61 (no codon bias) (Wright, 1990). Analysis reveals that the ENC mainly lays between 33.263 to 36.526 in the COI gene with the mean of 34.558, whereas for NDI gene the ENC value ranged between 25.369 to 31.121 with the mean of 27.495which is slightly higher than the result found in another lepidopteran,*Baronia*(26.65), but less than coleopteran *Neochlamisus*(45.78) (Herbeck and Novembre, 2003). In all the five species for both COI and NDI genes, the ENC range was nearer to 20, which means the codon bias was quite high. NDI had high ENC valuecompared to COI gene sequence. Synonymous codon usage orders (SCUO) did not vary much in the selected butterfly sequences for both the genes. It has been generally accepted that genome GC content is correlated with amino acid usage and codon usage (Ermolaeva, 2001). A very low or very high GC composition is associated with a large codon usage bias (Wan *et al.*, 2004). Studied conducted by Knight *et al.*(2001) showed that the GC composition was the main force that drives codon and amino-acid usage although both mutation and selection play important roles. Lynn *et al.*(2002) further showed that codon usage bias was affected by GC composition and environment (e.g., temperature). As we can see most of the study conducted on codon usage bias depended on the GC content especially GC_3region. But in our studies, the GC content is low for all the species whereas the AT content is quite high especially for

AT_3,for our data the bias is related to AT content which is in supported by other work on insect: hymenoptera (Behura and Severson, 2012).The codon usage bias is associated with the gene expression. Highly expressed genes have always high codon usage bias and tend to used that more frequently and found to be least evolving (Moura *et al.*, 2011; Deka and Chakraborty, 2014).

The ratio of non-synonymous (dN) to synonymous (dS) changes between taxa is used to estimate the direction of selection.Ratio of dN/dS = 1, dN/ dS < 1 and dN/dS > 1 represents neutral evolution, purifying selection and positive Darwinian selection, respectively. dN/dS ratio indicates the butterfly COI and NDI gene sequence evolved under purifying selection (Ophir *et al*., 1999).

The phylogenetic tree constructed for both COI and NDI at the subfamily was congruent with other studied on Nymphalidae butterfly.The relative branching order among intra-familial clades in Nymphalidae remains uncertain and has been presented by most authors (Ackery, 1984).The relationships among nymphalid tribesandsubfamilies remain poorly understood (Brower, 2000).The phylogenetic tree obtained for both the gene COI and NDI forms monophyletic clade within the species level, except in COI gene for *J.iphita* species (eight samples) from different locations across northeast India. Samples JIN3 and JIS3 formed basal clade for the rest of the *J.iphita* and *J.atlites* species from both sides of the Brahmaputra river. Interspecies hybridization is known to occur in the genus *Junonia*in the past or due to the fact that these are recently diverging lineages(Kodandaramaiah and Wahlberg, 2007). So in our case that might be the reason for two *J.iphita* species from both sides of the Brahmaputra river forming the base and from evolutionary point of view these

two species (JAT and JI) are more closely related within the *Junonia* species compared to other *Junonia* species.

COI and NDI sequences were combined together to build a phylogenetic tree. The Phylogenetic tree obtained for combined COI and NDI sequence is very similar with COI phylogenetic tree compared to NDI phylogenetic tree. It differ slightly only in the bootstrap value.

Hybridization and incomplete lineage sorting of ancestral polymorphic copies are usually the main cause of polyphyly within the intra-specific copies in the phylogenetic tree (Zheng *et al.,* 2014). In the analysis of the mtDNA COI and NDI sequence data in the present study, there is considerable congruence in topology, indicating that certain clades are well differentiated phylogenetically. But within the species level for *J.almana, J.atlites* and *A.merione* the difference or intraspecific variation was low. So in the phylogenetic tree it results in polytomy for both the genes with the different tree building methods (ML, MP and BI).With COI and NDI gene *J. iphita* result in different branching pattern. With COI gene the species from both the sites does not forms separate cluster and at the same time two species from different location forms a basal clade for the rest of the *J.iphita* and *J.atlites*. Using NDI gene, the *J.iphita* from north and south were separately cluster using two tree building methods (MP and BI) but with ML methods one of the species from north forms the basal clade for the south species.

In the present studies using COIand NDI for the *Elymnias hypernmestra* species we get the pattern of differentiation between north and south Brahmaputra river which is in concordant with Remington's (1968)work on *Papilio zelicaon* between two region of north

and south Alberta, where he found that some differentiation occurred although not to the degree that was implied by his description of a new species. In the present study, whereas for the other selected species from North and South Brahmaputra there was no specific barrier or hybridization might have occurred between the two regions as the maximum width of the Brahmaputra river extent up to 23 km and the minimum width is 1.20 km in the valley of Assam.

In our studies whole mitogenomes of two *J. iphita* representing north and south Brahmaputra river sites were sequenced to study the intraspecific variation. With our two species nucleotide composition are significantly biased toward AT and the mitogenome nucleotide skewness in our studies for both the species indicate a slight AT skew and moderate GC skew a common phenomenon of nymphalids (Kim *et al.,* 2010; Tian *et al.,* 2012). In both the *Junonia iphita* (JI_MZ and JI_AZ) intergenic spacers are dispersed throughout the whole mitochondrial genome and differ slightly between them. Intergenic spacer sequences (IGS) show significant differences even among closely related insect species, maybe because of their rapid evolutionary rates. The partial duplication of the ND2 gene is thought to be the reason for the origination of this spacer which underwent rapid sequence divergence for their noncoding nature among even closely related taxa (Kim *et al.,* 2009). Previous study reported on other nymphalidae species shows that $tRNA^{Ser(AGN)}$ has a reduced dihydrouridine, which was seen with our two Junonia species also (Hu *et al.,* 2010; Wang *et al.,* 2011; Zhao *et al.,* 2013). In case of $tRNA^{Arg}$ DHU loop is absent which is previously detected in other nymphalinae like *J. almana* isolate N877 (KF590539), *Yoma Sabina* (NC_024403), *J. orithya* (NC_022697). The phylogenetic tree constructed for the whole mitogenomes using the 13 PCGs region defines the genetic relatedness of the two *J.iphita* species.

A comparison analysis of the mitogenomes of both *J. iphita* (JI_AZ and JI_MZ) revealed minor variation. These variations were mainly in terms of differences in base pair numbers and composition of the end codon of nad1 and nad5. Another variation is in the structure of tRNA glycine where JI_Az showed an A-A mismatch at the anticodon stem. Analysis for SNPs by correlating with number of variable sites revealed that variation between the two genomes amounted to 0.4%. However inclusion of more samples from both sites may ultimately reveal enough reasons to establish intraspecific variation.

PCR-RFLP using three Restriction enzymes *Alu*I, *Taq*I and *Rsa*I on 7 samples from north and south of Brahmaputra was done to see whether PCR-RFLP will be able to detect intraspecific variation among the 7 species selected from both the sites. The use of only two or three Restriction enzyme (RE) will besufficient for identification of a single species dependingon the number and range of species among which todistinguish (Wolf *et al*., 2000).In our study we observed that most of the species we compared for both the sites had similar REs cutting sites using all the three enzymes. For *A. merione,* the species from both sites shows different cutting sites using enzyme *AluI.A.merione* from north had restriction site at 310 and 370bp which is not present in the south species, whereas south species had cutting sites at 670bp which is absent in the North species. Using Enzyme *RsaI* for the south, *A.merione* species there is only one cutting sites at 600bp which is totally different from north species which had cutting sites on 440 and 910bp. For the species, *euthalia aconthea* there was also difference in the cutting sites, but the species was not selected for studies because of sampling problems. Here, we tested only one species each from both the sites to see the difference within species and only *A. merione* and *E. aconthea* shows variation. But the PCR-RFLP result is not congruent with further works on mtDNA COI and NDI. In both the genes, *A.merione* does not show any intraspecific variation. *E.hypernmestra*showedintraspecific

variation with the species from north and south Brahmaputra river region using COI, NDI and RAPD-PCR while variation was not found in PCR-RFLP.

Hoxgene sequences from a range of arthropod taxa, including new data from a basal hexapod and a myriapod was used to estimate a phylogeny of the arthropods and the data shows thatinsects and crustaceans form a single clade within the arthropods to the exclusion of myriapods. They also suggest that myriapods are more closely allied to the chelicerates than to this insect/crustaceanclade (Cook *et al.,* 2001).

Dllis required for the elaboration of proximo-distal positionalinformation within developing adult limbs (Cohen and Jurgens, 1989). But, the larvae (caterpillars) of lepidopteranhave fully elaborated antennae, mouthparts, thoracic legs and abdominal prolegs, but the evolutionary relationshipsamong these structures, and the origin of the prolegs areunclear (Birket-Smith, 1984).Panganiban*et al*.(1994) isolated and characterizedthe expression of a lepidopteran Dllgene and identify common aspects of limb patterning in insects,and suggest that the regulation of Dllmight be modifiedto generate different types of limb.

An*in silico* characterization of the physico-chemical properties of five distal-less genes (three samples retrieved from the Swiss-Prot database and two sequenced of *E.hypernmestra* from north and south Brahmaputra riverrevealed that the distal-less proteins of the butterflies had a high pI while only that of *Heliconius cydno galanthus* requiredan acidic buffer solution. Since pI is actually the pH at which a protein carries no net charge, this information could be used to prepare the ideal pH of a buffer. An ideal pH would thus be at least one unit away from the calculated pI i.e. one unit above or below the calculated

pI.Therefore in this case the distal-less proteins of most of the samples required abasic buffer solution with a pH range of 8 to 11.*Heliconius cydno galanthus* requiredan acidic buffer solution with a pH range of 4 to 6 while *E.hypernmestra* (4AzD) required a buffer solution around the acidic and basic pH borderline i.e. 6 to 8.The instability index which provides a measure of how stable a protein may be in a test tube conditions is vital due to the fact that proteins in general are highly unstable in test tube environments. A protein with an instability index smaller than 40 is predicted as stable. In the present estimate, of the five distal-less protein samples only *E.hypernmestra* (4AzD) computed an instability index of less than 40 i.e. 32.43. Thus indicating only *E.hypernmestra* (4AzD) distal-less protein is stable.

Extinction coefficient of a protein is a measure of how strongly it absorbs light at a given wave length. In proteins it depends almost exclusively on the number of aromatic residues, particularly tryptophan and cystine. Cystine is the amino acid formed by the oxidation of two cysteine molecules that covalently link via a disulphide bond. *Bicyclus anynana* computed the highest extinction coefficient of 60975 at 280 nm while *E.hypernmestra* (4AzD)computed the lowest at 2045. It's worth mentioning that observation of the Cys_Rec results revealed disulphide (SS) bond formation in the distal-less protein of *Bicyclus anynana* which might be the reason for the highest extinction coefficient reading.

The aliphatic index of a protein is defined as the relative volume occupied by aliphatic side chain amino acids–alanine, valine, isoleucine and leucine. This index may be regarded as a positive factor for increased thermostability as it has been recorded that proteins of thermophilic bacteria contain high concentrations of aliphatic amino acids in comparison to other proteins.The distal-less proteins of*E.hypernmestra (*4AsDand 4AzD*)*, both with an AI>70 may be the only samples among the selected five that are able to withstand a wide

range of temperatures.*E.hypernmestra* (4AzD) with an aliphatic index of 124.77 and *E.hypernmestra* (4AsD) with a score of 120.53 may be the most thermostable protein among the five selected samples.

GRAVY value for a protein is the sum of its hydropathy values which is nothing but the hydrophobic and hydrophilic properties of its side chains. Larger the GRAVY values more hydrophobic the protein. The distal-less proteins all computed GRAVY values of less than 1. This indicates that these proteins are hydrophilic. An analysis of the amino acid composition reveals that polar amino acids i.e. Serine (Ser), Threonine (Thr), Glycine (Gly) and Glutamine (Gln) predominate. The polar amino acid although uncharged overall, they have an uneven charge distribution. Because of this, these amino acids can form hydrogen bonds with water. As a result, such amino acids are hydrophilic and are often found on the outer surface, in contact with the watery environment of the cell. Distal-less proteins sequence of three samples i.e. *Bicyclus anynana*, *Heliconius cydno galanthus* and *Junonia coenia* showed a high percentage of proline (Pro) residue. High number of Proline (Pro) residues on the other hand indicates presence of many conformationally rigid regions. This is because the distinctive cyclic structure of proline's side chain gives it an exceptional conformational rigidity compared to other amino acids. It therefore affects the rate of peptide bond formation between proline and the other amino acids. Also, it is this rigid nature that affects the secondary structure of proteins and disrupts alpha helices and beta sheets. SOSUI server predicts presence and location of secondary structures of a protein. Absence or rare occurrences of secondary structure could infer that the protein is soluble and that it is not a trans-membrane protein. This is because trans-membrane proteins or trans-membrane regions in a protein are rich in secondary structures. Distal-less proteins from the five butterfly samples was computed by SOSUI server as soluble and also that it is not a trans-membrane

protein. This is corroborated by the 3D structure which shows the position of their secondary structure as not transmembrane in nature.

Cys_Rec software was used to identify the number of cysteine residues in each of the five distal-less proteins. Along with identifying these residues, the software also identifies possible or probable disulphide (SS) bond formation in any of these protein sequences. *Junonia coenia and Bicyclus anynana* contained two and three cysteine residues respectively. However the software computed only *Bicyclus anynana* as having any chance of forming disulphide (SS) bond formation.

3D structure analysis of the distal-less proteins revealed useful information. The absence or the presence of only one or two cysteine residues indicated that there was no chance of disulphide (SS) bond formation. In the case of *Junonia coenia*, even though there were two cysteine residues, Cys_Rec software analysis failed to confirm any disulphide bond formation. This was further confirmed by its 3D structure analysis where the two cysteine residues were spaced too far apart from each other to allow for any type of interaction. However, the protein of *Bicyclus anynana* had three cysteine residues and observation of their 3D structure seems to confirm Cys_Rec software prediction of atleast one disulphide (SS) bond formation. Cys_Rec computed that CYS113 was the most probable candidate to form a disulphide (SS) bond. The two other cysteine residues in the viscintiy of CYS113 are CYS29 and CYS104. Analysis of the 3D structure using Rasmol computed that CYS104 could be the probable candidate for CYS113 to form a disulphide (SS) bond, with a bond distance of 14.20A (bond distance of CYS113 and CYS29 was 22.96). Further 3D structural analysis revealed that proline (Pro) residues formed constituents of turns and if at all they formed part of a secondary structure they were situated towards the end. This confirms the

fact that proline (pro) acts as disrupters of secondary structures and are thus avoided as constituents of secondary structures, especially helices. Transmembrane proteins are generally characterized by a majority of alpha-helix structures as they form a channel connecting the extra-cellular to the intra-cellular regions. SOSUI server findings that distal-less proteins are not transmembrane proteins is confirmed by analysis of the 3D structure which reveals very less alpha-helix structures (Figure 21).

The maximum parsimony tree constructed for dll gene revealed that both the *E.hypernmestra* sequence forms a sister clade with high bootstrap support instead of forming separate cluster (Figure 22). In our study Distal-less protein shows minor variation and is not a good marker for studying intra-specific variation within the selectedbutterfly species. This study provides base information for further studies as very few information are present in databases.

The present study showed RAPD can be used as a tool for rapid detection of genetic polymorphism at the species level and to some extent at within species level. Despite its drawbacks ofpoor levels of reproducibility and dominant inheritance of markers, RAPD is widely used as areliable, quick, easy, and cost-effective method in genetic relationship study also, particularlyat intraspecific level and among closely-related species (Ayana *et al*., 2000; Fahima *et al*., 1999; Millan *et al*., 1996; Hoey *et al*., 1996). In RAPD-PCR, all the species were separately clustered or forms monophyly with a good bootstrap support, but at the genus level, JA forms a sister clade with AM instead of forming a clade with the other *Junonia* species. So using other molecular markers like COI and NDI we try to clarify that problem. The two mtDNA gene COI and NDI result in a tree where the three Junonia species forms a sister clade, and for COI gene *A.merione* and *E.hypernmestra* forms a sister clade although

other studies on Nymphalidae shows that *A. merione* is more closely related to *Junonia* species than the *E.hypernmestra* where as for NDI gene *A. merione* forms the basal clade to the *Junonia* species and *E.hypernmestra* form paraphyletic clade. In the present study, using COI, NDI and RAPD-PCR for the *E. hypernmestra* species we get the pattern of differentiation between north and south Brahmaputra river.And for the other selected species from North and South Brahmaputra there was no specific barrier or hybridization might have occurred between the two regions which is a natural phenomenon as butterfly can fly even for a long distant. But with Dll gene there was no variation between the two species of *E.hypernmestra*, they form sister clade with a high bootstrap value (99).

This is the first scientific research done on comparative phylogeny on butterflies across NE India using different molecular markers. It can form the platform for further studies on molecular phylogeny of butterfly of NE India especially in Mizoram. Lots of work still need to be done using more molecular markers, more field sampling and increasing number of individuals to know the true evolutionary or phylogenetic relationship of Butterflies across NE India particularly Mizoram.

VI. SUMMARY

✓ Based on the survey and preliminary field and molecular studies, five species of butterfly were selected for studying the intraspecific variation across NE India. The five species selected were *J. almana, J. iphita, J. atlites, E. hypernmestra* and *A. merione.*

✓ 120 nymphalidae species were identified and documented from Mizoram.

✓ The diversity of butterfly in Mizoram was found to be highest in tropical wet evergreen forest (1.927 in Shannon and 74.6 in Simpsons) and lowest in Mountain Sub-Tropical Forest (1.729 in Shannon and 44.3 in Simpsons).

✓ In our studies using RAPD-PCR, *E.hypernmestra, J.iphita* and *A.merione* species were differentiated between the north and south Brahmaputra River. *E. hypernmestra* species from both sites of river Brahmaputra formed a separate cluster with a strong bootstrap support of 100%. *J. iphita* and *A. merione*species formed a separate cluster for both sides with bootstrap support of 90% and 58%, respectively. In the case of *Junonia atlites* and *J.almana,*the species from both sites did not cluster separately as there might be horizontal gene transfer occurring with the sample from both sides of the Brahmaputra.

✓ In general, the intraspecific variation is low within and between the samples from both sides. One way analysis of ANOVA was done using binary data and PIC value of RAPD data. Both results give a p value less than 0.05, which is statistically considered significant. It shows that RAPD can be used to detect polymorphism among and within species.

✓ Two mitochondrial genes- COI and NDI were used to study the relationship within and between the selected five Nymphalidae species across Northeast India using different parameters. The nucleotide composition was calculated and result in a strong A+T bias in both the gene. Analysis with closely related species found more transitions than transversions; transversions is more in distantly related species due to multiple substitutions.The species we selected are closely related comprising of three genera within the family nymphalidae. So, within and between species level transition was higher compared to transversion. COI gene shows higher transition ratio compared to NDI gene.

✓ Relative synonymous codon usage (RSCU) values gave an estimate of the preference for alternative synonymous codons. The codon with highest RSCU was UUA (L). All of the remaining codons with RSCU greater than 2 have U or, particularly, A in codon position 3. In our studies, the hydrophobic amino acids codons which were more frequently encoded were UUU (F), UUA (L), AUU (I), AUA (M) and GUU (V).

✓ The p-distance or evolutionary distance was calculated for all the species using the kimura-2-parameters for both the genes. It was observed that within the species, the genetic distance was quite low and ranged from 0.0 to 0.02, but for *J. iphita* the maximum genetic distance (p-distance) was 0.07 which is the highest intraspecific distance in the studied species. The maximum p-distance among the five selected species is 0.15 for COI gene and for NDI gene it is 0.20.

✓ Analysis revealed that the Effective Number of Codons (ENC)rangedbetween 33.26 to 36.53 in the COI gene with a mean value of 34.56, whereas for NDI gene, the ENC value ranged between 25.64 to 31.12 with a mean value of 27.50.

✓ Phylogenetic analysis using COI and NDI shows that the difference or intraspecific variation was low within the species level for *J.almana, J.atlites* and *A.merione*. So in the phylogenetic tree it results in polytomy for both the genes with the different tree building methods (ML, MP and BI). Using COI and NDI gene *J. iphita* result in different branching pattern.

✓ With COI gene the species from both the sites does not forms separate cluster and at the same time two species from different location forms a basal clade for the rest of the *J.iphita* and *J.atlites*.

✓ Using NDI gene, *J.iphita* from north and south were separately clustered using two tree building methods (MP and BI), but with ML one of the species from north forms the basal clade for the south species. But for *E.hypernmestra* species, the pattern of differentiation between north and south Brahmaputra River using different tree building methods which are in concordant with earlier works.

✓ The Phylogenetic tree obtained for combined COI and NDI sequence is very similar with COI phylogenetic tree compared to NDI phylogenetic tree. It differ slightly only in the bootstrap value.

✓ The average PIC value for PCR-RFLP for COI gene is 0.3 for north and 0.32 for south. The restriction enzyme *Alu I* is a frequent cutter, so it resulted in more number of bands with 28 RE cutting sites for both the sites and number of polymorphic bands were 19 and 16, respectively. The resolving power of marker is also highest with a value of 7.42 for both the sites. Only *A.merione* and *E. aconthea* shows variation in the banding pattern, but

the PCR-RFLP result is not congruent with further works on mtDNA COI and NDI. While *E. hypernmestra* result in intraspecific variation with the species from north and south Brahmaputra Rivers using COI, NDI and RAPD-PCR.

✓ The complete mitogenome of *J. iphita* from south Brahmaputra River (JI_MZ) is a circular molecule of 15,433 bp in length and the near complete mitogenome of *J. iphita* from north Brahmaputra River (JI_AZ) is 14892 bp in length.The overall nucleotide compositions are significantly biased toward AT (80.5 % and 80 % respectively). The mitogenome nucleotide skewness for JI_MZ and JI_AZ were AT-skew=-0.003; GC-skew=0.215 and AT-skew=-0.002; GC-skew=-0.213 respectively.

✓ SNPs were identified at 51 sites along the length of the genome. This amounted to 0.4 % of the entire mitochondrial genome. These SNP sites were infered by correlating with the number of variable sites.

✓ In both the *J. iphita* (JI_MZ and JI_AZ) intergenic spacers are dispersed throughout the whole mitochondrial genome which differs slightly between the two species (14 and 12). The largest intergenic spacer is located between tRNAG and *ND2* and consists of 52 bp in both the species.For JI_MZ the A+T rich region is 546 bp in length containing ATAGA motif followed by a 18 bp poly-T stretch, two microsatellite-like $(TA)_9$ elements and 8 bp poly-A stretch immediately upstream of trnM gene.

✓ All PCGs are initiated by a typical ATN codon, except for the cox1 gene, which uses unusual CGA as start codon. Seven protein coding genes start with ATG, five with ATT in both the species. Eight genes use complete termination codon (TAA) for the species JI_MZ

but for JI_AZ nine gene use complete termination codon (TAA), three genes (*cox1, cox2* and *nad5*) end with single T; nad4 and nad1 use TA as stop codon for JI_MZ species. While for JI_AZ three genes have incomplete stop codon T (*cox1, cox2* and *nad1*) and one gene (*nad4*) ended with TA.

✓ Comparison between these two sequences reveals minor variation. We excluded the D loop portion in our comparisons since the D loop of *J.iphita* AZ was not sequenced. It was observed that the nad1 region of JI_Az ended with TA while that of JI_Mz ends with only T. In the case of nad5 JI_Az ends with an incomplete stop codon T and JI_Mz ends with a complete stop codon TAA. Analysis of the tRNA structure revealed similar structures except for the tRNA of Glycine. Here it was observed that JI_Mz has an A-A mismatch in its anti-codon stem. This mismatch was absent in the case of JI_Az Minor variations were observed in the number of nucleotide residues A, T, G and C; in the number of intergenic spacers and also in the number of base pairs of nad5, cytB, rrnaL and rrnaS. Variations in terms of base pair numbers were also observed in tRNAs.

✓ An*In silico* characterization of the physico-chemical properties of five distal-less genes (three samples retrieved from the Swiss-Prot database and two sequenced of *E.hypernmestra* from north and south Brahmaputra riverrevealed that the distal-less proteins of the butterflies had a high pI.

✓ A protein with an instability index smaller than 40 is predicted as stable. In the present estimate, of the five distal-less protein samples only *E.hypernmestra* (4AzD) computed an instability index of less than 40 i.e. 32.43. Thus indicating only *E.hypernmestra* (4AzD) distal-less protein is stable. *B. anynana* computed the highest extinction coefficient of 60975 at 280 nm while *E.hypernmestra* (4AzD)computed the lowest at 2045.

✓ The distal-less proteins (dll) of*E.hypernmestra (*4AsD and 4AzD*)*, both with an AI>70 may be the only samples among the selected five that are able to withstand a wide range of temperatures. *E.hypernmestra* (4AzD) with an aliphatic index of 124.77 and *E.hypernmestra* (4AsD) with a score of 120.53 may be the most thermostable protein among the five selected samples.

✓ In our study the phylogenetic tree constructed for Dll gene did not give much information regarding the intraspecific variation between the two species of *E.hypernmestra* from north and south of Brahmaputra. Distal less protein shows minor variation. This study provides base information for further studies as very few information are present in databases.

VII. BIBILOGRAPHY

Al-Barrak M, Loxdale HD, Brookes CP, Dawah HA, Biron DG and Alsagair O (2004) Molecular evidence using enzyme and RAPD markers for sympatric evolution in British species of *Tetramesa* (Hymenoptera: Eurytomidae).Biol. J. Linnean. Soc. 83: 509-525.

Akam M (2000) Arthropods: developmental diversity with a (super) phylum. Proc. Natl. Acad. Sci. 97: 4438-4441.

Ackery PR (1984) Systematic and faunistic studies on butterflies.In The biology of butterflies (ed. R. I. Vane-Wright & P. R. Ackery). 9-24. London: Academic Press.

Anu A, Sabu TK and Vineet PJ (2009) Seasonality of litter insects and relationship with rainfall in a wet ever green forest in south Western Ghats. J. Ins. Sci. 9: 46.

Ashfaq M, Akhtar S, Khan AM, Adamowicz and Hebert PDN (2013) DNA barcode analysis of butterfly species from Pakistan points towards regional endemism. Mol. Ecol. Res. 13: 832-843.

Aubert J, Legal L, Descimon H and Michel F (1999) Molecular Phylogeny of Swallowtail Butterflies of the Tribe Papilionini (Papilionidae, Lepidoptera) Mol. Phylogenet. Evol. 12(2): 156-167.

Avise JC (2004) Molecular Markers, Natural History, and Evolution, 2nd edn. Sinauer Associates, Sunderland, Massachusetts.

Ayana A, Bekele E and Bryngelsson T (2000) Genetic variation in wild sorghum (Surghum biculur ssp.verticillijlorum (L.) Moench) germplasm from Ethiopia assessed by random amplified polymorphic DNA (RAPD). Hereditas 132: 249-254.

Ayres DR, Garcia-Rossi D, Davis HG and Strong DR (1999) Extent and degree of hybridization between exotic (*Spartinaalterniflora*) and native (*S. foliosa*) cordgrass (*Poaceae*) in California, USA determined by random amplified polymorphic DNA (RAPD). Mol. Ecol. 8: 1179-1186.

Baker RH and DeSalle R (1997) Multiple sources of character information and the phylogeny of Hawaiian drosophilids. Syst. Biol. 46: 654-673.

Ball SL, Hebert PDN, Burian SK and Webb JM (2005) Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. J. North. Am. Benthol. Soc. 24: 508-524.

Bansal R and Karanth KP (2010) Molecular phylogeny of *Hemidactylus* geckos (Squamata: Gekkonidae) of the Indian subcontinent reveals a unique Indian radiation and an Indian origin of Asian house geckos. Mol. Phylogenet. Evol. 57 (1): 459-465.

Bardakci F (2001) Random Amplified Polymorphic DNA (RAPD) Markers. Turk. J. Biol. 25: 185-196.

Barrett RDH and Hebert PDN (2005) Identifying spiders through DNA barcodes.Can. J. Zool.83:481-491.

Bartish IV, Jeppson N and Nybom H (1999) Population genetic structure in the dioecious pioneer plant species *Hippophaerhamnoides* investigated by random amplified polymorphic DNA (RAPD) markers. Mol. Ecol. 8: 791-802.

Behura SK and Severson DW (2012) Comparative analysis of Codon usage bias and codon context patterns between dipteran and hymenopteran sequenced genomes.PloS One. 7(8): e43111.

Beldade P and Brakefield PM (2002) The genetics and evo–devo of butterfly wing patterns. Nature Rev. Genet.3: 442-452.

Benton TG (1995) Biodiversity of Henderson Island insects.Bio. J. Linn. Soc. 56: 245-259.

Bergmann T, Jach R, Damm S, Desalle R, Schierwater B and Hadrys H (2013) The potential of distance-based thresholds and character-based DNA barcoding for defining problematic taxonomic entities by CO1 and ND1. Mol. Ecol. Resour. 13: 1069-1081.

Bernt M, Donath A, Jühling F, Externbrink F, Florentz C, Fritzsch G and Stadler PF (2013) MITOS: Improved de novo metazoan mitochondrial genome annotation. Mol. Phylogenet. Evol. 69(2): 313-319.

Birket-Smith SJR (1984) Prolegs, Legs and Wings of Insects.Copenhagen: Scandinavian Science Press Ltd.

Black WC, Duteau NM, Puterka GJ, Nechols JR and Pettorini JN (1992) Use of the random amplified polymorphic DNA polymerase chain reaction (RAPD-PCR) to detect DNA polymorphisms in aphids (Homoptera: Aphididae). Bull. Entmol. Res. (82):151-159.

Blaxter M (2003) Counting angels with DNA. Nature 421: 122-124.

Bolaric S, Barth S, Melchinger AE and Posselt UK (2005) Genetic diversity in European perennial ryegrass cultivars investigated with RAPD markers. Plant. Breed. 124: 161-166.

Boore JL (1999) Animal mitochondrial genomes.Nucl. Acids Res. 27: 1767-1780.

Botstein D, White RL, Skolnick M and Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. 32: 314-331.

Brandão KLS, Lyra ML and Freitas AVL (2009) Barcoding Lepidoptera: Current Situation and Perspectives on the Usefulness of a Contentious Technique.Neotrop.Entomol. 38(4): 441-451.

Brook BW, Sodhi NS and Ng PKL (2003) Catastrophic extinctions follow deforestation in Singapore. Nature 424: 420-424.

Brower AVZ and DeSalle R (1998) Mitochondrial vs. nuclear DNA sequence evolution among nymphalid butterflies: the utility of Wingless as a source of characters for phylogenetic inference. Insect Mol. Biol. 7: 1-10.

Brunton CFA and Hurst GDD (1998) Mitochondrial DNA phylogeny of brimstone butterflies (genus *Gonepteryx*) from the Canary Islands and Madeira. Biol. J. Linn. Soc. 63: 69-79.

Burns JM, Janzen DH, Hajibabaei M, Hallwachs W and Hebert PDN (2008) DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica. Proc. Natl. Acad. Sci. 105: 6350-6355.

Bush RM, Bender CA, Subbarao K, Cox NJ and Fitch WM (1999) Predicting the evolution of human influenza A. Science 286: 1921-1925.

Bussell JD (1999) The distribution of random amplified polymorphic DNA (RAPD) diversity amongst populations of *Isotomapetraea* (*Lobeliaceae*). Mol. Ecol. 8: 775-789.

Cameron SL, Yoshizawa K, Mizukoshi A, Whiting MF and Johnson KP (2011) Mitochondrial genome deletions and minicircles are common in lice (Insecta: Phthiraptera). BMC Genomics12: 394.

Cameron SL (2014) Insect mitochondrial genomics: implications for evolution and phylogeny.Annu. Rev. Entomol. 59: 95-117.

Carroll SB, Gates J, Keys D, Paddock SW, Panganiban GEF, Selegue J and Williams JA (1994) Pattern formation and eyespot determination in butterfly wings. Science 265: 109-114.

Castro LR, Austin AD and Dowton M (2002) Contrasting rates of mitochondrial molecular evolution in parasitic Diptera and Hymenoptera. Mol. Biol. Evol.19: 1100-13.

Caterino MS and Sperling FAH (1999) Papilio phylogeny based on mitochondrial cytochrome oxidase I and II genes. Mol. Phylogenet. Evol 11: 122-137.

Chantangsi C, Lynn DH, Brand MT, Cole JC, Hetrick N and Ikonomi P (2007) Barcoding ciliates: a comprehensive study of 75 isolates of the genus *Tetrahymena*. Int. J. Syst. Evol. Microbiol. 57: 2412-2425

Chow S (1993) PCR-RFLP analysis on thirteen western Atlantic snappers (subfamily Lutjaninae): a simple method for species and stock identificaton. Fish. Bull.91: 619-627.

Chow S and Inogue S (1993) Intra- and interspecific restriction fragment length polymorphism in mitochondrial genes of *Thunnustuna* species. Bull. Nat. Res. Inst. Far. Seas.Fish.30: 207-224.

Conant GC and Wolfe KH (2008) GenomeVx: simple web-based creation of editable circular chromosome maps. Bioinf. 24(6): 861-862.

Corbet GB and Hill JE (1992) The mammals of the Indomalayan region: a systematic review. Oxford University Press, New York.

Cohen SM (1990) Specification of limb development in the *Drosophila* embryo by positional cues from segmentation genes.Nature343: 173-177.

Cohen SM, Brönner G, Küttner F, Jürgens G, Jäckle H (1989) *Distal-less* encodes a homeodomain protein required for limb developmentin *Drosophila*. Nature 338: 432-434.

Cook CE, Smith ML, Telford MJ, Bastianello A and Akam M (2001) *Hox* genes and the phylogeny of the arthropods. Curr. Biol. 11(10): 759-763.

Costa FO, deWaard JR, Boutillier J, Ratnasingham S, Dooh RT, Hajibabaei M and Hebert PDN (2007) Biological identifications through DNA barcodes: the case of the Crustacea. Can. J. Fish.Aquat. Sci. 64: 272-295.

Crozier RH and Crozier YC (1993) The Mitochondrial Genome of the Honeybee Apis mellifera: Complete Sequence and Genome Organization. Genetics 133(1): 97-117.

CYS_REC.http://sun1.softberry.com/berry.phtml?topic=cys_rec &group=help &subgroup-propt. (27/10/2006).

Darriba D, Taboada GL, Doallo R and Posada D (2012) "jModelTest 2: more models, new heuristics and parallel computing". Nat. Methods 9(8): 772.

DeBruijn MHL (1983) Drosophila melanogaster mitochondrial DNA, a novel organization and genetic code. Nature 304: 234-241.

Deka H and Chakraborty S (2014) Compositional constraint is the key force in shaping codon usage bias in haemagglutinin gene in H1N1 subtype of influenza *A* virus. Int. J. Genomics 2014:7.

DeMandal SD, Liansangmwii C, Gurusubramanian G and Senthil Kumar N (2014) Mitochondrial markers for identification and phylogenetic studies in insects- A Review. DNA Barcode 2: 1-9.

deRosa R, Grenier JK, Andreeva T, Cook CE, Adoutte A and Akam M (1999) *Hox* genes in brachiopods and priapulids and protostome evolution. Nature 399: 772-776.

Dezfouli SRN, Oshaghi MA, Vatandoost H, Djavadian E, Telmadarei Z and Assmar M (2002) Use of Random Amplified Polymorphic DNA Polymerase ChainReaction (RAPD-PCR) and ITS2 PCR assays for differentiation of populations and putative sibling species of *Anopheles fluviatilis* (Diptera: Culicidae) in Iran. Iranian J. Publ. Health. 31(3-4): 133-137.

Dijkstra KDB, Groeneveld LF, Clausnitzer V and Hadrys H (2007) The Pseudagrion split: molecular phylogeny confirms the morphological and ecological dichotomy of Africa′s most diverse genus of Odonata (Coenagrionidae). Int. J. Odonatol. 10: 31-41.

Dowdy AK and Mcgaughey WH (1996) Using Random Amplified Polymorphic DNA to differentiate strains of the Indian meal moth.(Lepidoptera:Pyralidae). Environ. Entomol. 25: 396-400.

Dowton M, Cameron SL, Dowavic JI, Austin AD and Whiting MF (2009) Characterization of 67mitochondrial tRNA gene rearrangements in the Hymenoptera suggests that mitochondrial tRNA gene position is selectively neutral. Mol. Biol. Evol.26: 1607-17.

Ermolaeva MD (2001) Synonymous codon usage in bacteria.Curr.Issues. Mol .Biol. 3: 91-97.

Evans WH (1932) The identification of Indian butterflies. Bombay Natural Hist. Soc.

Fahima T, Sun GL, Beharav A, Krugman T, Beiles A and Nevo E (1999) RAPD polymorphism of wild emmer wheat populations, Triticum dicoccoides, in Israel. Theor. Appl. Genet. 98:434-447.

Felsenstein J (1983) Statistical Inference of Phylogenies.J. R. Stat. Soc. series A, 146: 246-272.

Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evol. 39: 783-791.

Feltwell J (1986) The Natural History of Butterflies. Groom Helem Ltd. Provident House, Bureel Row, Beckenham Kent BR3 IAT, 133PP.

Fernández ME, Figueiras AM and Benito C (2002) The use of ISSR and RAPD markers for detecting DNA polymorphism, genotype identification and genetic diversity among barley cultivars with known origin. Theor. Appl. Genet. 104: 845-851.

Foster PG and Hickey DA (1999) Compositional bias may affect both DNA-based and protein-based phylogenetic reconstructions. J. Mol. Evol. 48: 284-290.

Funk DJ (1999) Molecular systematics of cytochrome oxidase I and 16S from *Neochlamisus* leaf beetles and the importance of sampling. Mol. Biol. Evol. 16: 67-82.

Gadgil M (1996) Documenting diversity: An experiment. Curr. Sci. 70: 36-44.

Gaikwad SS, Ghate HV, Ghaskadbi SS, Patole MS and Shouche YS (2013) DNA barcoding of nymphalid butterflies (Nymphalidae:Lepidoptera) from Western Ghats of India. Mol. Biol. Rep. 39: 2375-2383.

Gallusser S, Guadagnuolo R and Rahier M (2004) Genetic (RAPD) diversity between Oleria onega agarista and Oleria onega ssp. in north-eastern Peru. Genetica 121: 65-74. Geospiza Inc.

Garner KJ and Slavicek JM (1996) Identification and characterization of a RAPD-PCR marker for distinguishing Asian and North American gypsy moths. Insect Mol. Biol. 5: 81-90.

Gillespie JH (1991) The causes of molecular evolution. Oxford University Press, Oxford, England.

Glez-Peña D, Gómez-Blanco D, Reboiro-Jato M, Fdez-Riverola Fand Posada D (2010) ALTER: program-oriented format conversion of DNA and protein alignments. Nucl.Acids. Res. Web Server issue. ISSN: 0305-1048.

Gómez DRS (2004) Intraspecific variation and population structure of the Velvetbean Caterpillar, Anticarsia gemmatalis Hübner, 1818 (Insecta: Lepidoptera: Noctuidae). Genet. Mol. Biol. 27(3): 378-384.

Goodsell D (2000) Cytochrome C Oxidase.RCSB Protein Data Bank(PDB).

Gray MW, Burger G and Lang BF (1999). Mitochondrial evolution. Science 283: 1476-1481.

Gray MW, Burger G and Lang BF (2001). The origin and early evolution of mitochondria. Genome Biol. 2: 1-5.

Greenstone MH, Rowley DL, Heimbach U, Lundgren JG, Pfannenstiel RS and Rehner SA (2005) Barcoding generalist predators by polymerase chain reaction: carabids and spiders. Mol. Ecol. 14: 3247-3266.

Guadagnuolo R, Savova-Bianchi D and Felber F (2001a) Gene flow from wheat (*Triticum aestivum* L.) to jointed goatgrass (*Aegilopscylindrica* Host.), as revealed by RAPD and microsatellite markers. Theor. Appl. Genet. 103: 1-8.

Guadagnuolo R, Savova-Bianchi D and Felber F (2001b) Search for evidence of introgression of wheat (*Triticum eastivum* L.) traits into sea barley (*Hordeum marinum* s.str. Huds) and bearded wheatgrass (*Elymus caninus* L.) in central and northern Europe, using isozymes, RAPD and microsatellite markers.Theor. Appl. Genet. 103: 191-196.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W and Gascuel O (2010) New Algorithms and Methods to Estimate Maximum-Likelihood Phylogenies: Assessing the Performance of PhyML 3.0. Syst. Biol. 59(3): 307-21.

Gupta S and Preet S (2013) Genetic differentiation of invasive *Aedes albopictus* by RAPD-PCR: implications for effective vector control. Parasitol. Res. 113: 2137-2142.

Hadrys H, Clausnitzer V and Groeneveld LV (2006) The present role and future promise of conservation genetics for forest Odonates. In: Forests and Dragonflies (ed. Rivera A). 279-299. Pensoft Publishers Sofia- Moscow, Pontevedra, Spain.

Hahn V, Blankenhorn K, Schwall M and Melchinger AE (1995) Relationships among early European maize inbreds. III. Genetic diversity revealed with RAPD markers and comparison with RFLP and pedigree data. Maydica. 40: 299-310.

Hahn C, Bachmann L and Chevreux B (2013) Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads - a baiting and iterative mapping approach. Nucl. Acids Res. 41(13): e129.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W and Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. Proc. Nat. Acad. Sci. USA.103: 968-971.

Hammer R, Harper DAT and Ryan PD (2001) PAST: Palaeontological Statistics Software package for education and data analysis. Palaeontol.Electron. 4(1): 1-9.

Haribal M (2003) Butterflies of Sikkim Himalaya and Their Natural History.Natraj Publishers.217p.

Harvey DJ (1991) Higher classification of the Nymphalidae, Appendix B [M] //Nijhout HF. (Ed.), The Development and Evolution of Butterfly Wing Patterns. Washington, DC: Smithsonian Institution Press. 255-273.

Hebert PD, Cywinska A, Ball SL and Dewaard JR (2003) Biological identifications through DNA barcodes.Proc. R. Soc. BBiol. Sci.270: 313-321.

Hebert PDN, Penton EH, Burns JM, Janzen DH and Hallwachs W (2004a) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Proc. Natl. Acad. Sci. USA. 101: 14812-14817.

Hebert PDN, Ratnasingham S and Dewaard JR (2003a) Barcoding animal life:Cytochrome c oxidase subunit 1 divergences among closely related species. Proc. R. Soc. Lond. B (Suppl.) 270: 96-99.

Hebert PDN, Stoeckle MY, Zemlak TS and Francis CM (2004b) Identification of birds through DNA barcodes.Plos. Biol. Public Library of Science Biology 2: 1657-1663.

Herbeck JT and Novembre J (2003) Codon usage patterns in cytochrome oxidase I across multiple insect orders**.**J. Mol. Evol. 56(6)**:** 691-701.

Hirokawa T, Boon-Chieng S and Mitaku S (1998) SOSUI: classification and secondary structure prediction system for membrane proteins. Bioinf. 14: 378-9.

Hoey BK, Crowe KR, Jones VM and Polans NO (1996) Aphylogenetic analysis of *Pisum* based on morphologicalcharacters, and allozyme and RAPD markers. Theor. Appl.Genet. 92: 92-100.

Hoole JC, Joyce DA and Pullin AS (1999) Estimates of gene flow between populations of the swallowtail butterfly *Papilio machaon* in Broadland, UK and implications for conservation. Biol. Conserv. 89: 293-299.

Hoy M (2003) Insect Molecular Genetics, 2nd edn. Academic Press, San Diego, California.

Hu J, Zhang DX, Hao JS, Huang DY, Cameron S and Zhu CD (2010) The complete mitochondrial genome of the yellow coaster, *Acraea issoria* (Lepidoptera: Nymphalidae: Heliconiinae: Acraeini): Sequence, gene organization and a unique tRNA translocation event. Mol. Biol. Rep. 37:3431-8.

Hunt GJ and Page RE.JR (1992) Patterns of inheritance with RAPD molecular markers reveal novel types of polymorphism in the honey bee. Theo. Appl. Genetics. 85: 15-20.

Hurst GD and Jiggins FM (2005) Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc. Biol. sci. R. Soc. 272: 1525-1534.

Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E and Hebert P (2005) Wedding biodiversity inventory of a large complex Lepidoptera fauna with DNA barcoding. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360: 1835-1845.

Kambhampati S, Black WC IV and Rai KS (1992) Random amplified polymorphic DNA of mosquito species and populations (Diptera: Culicidae): Techniques, statistical analysis, and applications. J. Med. Entomol. 29**:** 939-94.

Karanth KP (2006) Out-of-India Gondwanan origin of some tropical Asian biota.Curr. Sci. 90: 789-792.

Karanth KP (2008) Primate Numts and reticulate evolutions of capped and golden leaf monkeys (Primates: Colobinae). J. Biosci. 33(5): 761-770.

Karanth KP, Singh L, Collura RV and Stewart CB (2008) Molecular phylogeny and biogeography of langurs and leaf monkeys of South Asia (Primates: Colobinae). Mol. Phylogenet. Evol. 46: 683-694.

Karl SA, Bowen BW and Avise JC (1992) Global population genetic structure and male-mediated gene flow in the green turtle (*Chelonia mydas*): RFLP analyses of anonymous nuclear loci. Genetics131: 163-73.

Kaufman TC, Lewis R, Wakimoto B (1980) Cytogenetic analysis of chromosome 3 in *Drosophila melanogaster*: the homeotic complex in polytene chromosome interval 84 A-B. Genetics 94: 115-133.

Kehimkar I (2008) The Book of Indian Butterflies. Bombay Natural History Society.

Kelly DW, Macisaac HJ and HeathDD (2006) Vicariance and dispersal effects on phylogeographic structure and speciation in a widespread estuarine invertebrate. Evol. 60(2): 257-267.

Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM and Hebert PDN (2007) Comprehensive DNA barcode coverage of North American birds. Mol. Ecol. Notes 7: 535-543.

Kimura M (1983) The neutral theory of molecular evolution. Cambridge University Press, Cambridge, England.

Kim I, Lee EM, Seol KY, Yun EY, Lee YB, Hwang JS and Jin BR (2006). The mitochondrial genome of the Korean hairstreak, Coreana raphaelis (Lepidoptera: Lycaenidae). Insect Mol. Biol. 15: 217-225.

Kim MI, Baek JY, Kim MJ, Jeong HC, Kim KG, Bae CH, Han YS, Jin BR and Kim I (2009) Complete nucleotide sequence and organization of the mitogenome of the red-spotted apollo butterfly, Parnassius bremeri (Lepidoptera: Papilionidae) and comparison with other lepidopteran insects. Mol. Cells 28: 347-363.

Kim M J, Wan X, Kim K, Hwang JS and Kim I (2010) Complete nucleotide sequence and organization of the mitogenome of endangered Eumenis autonoe (Lepidoptera: Nymphalidae). Afr. J. Biotechnol. 9(5): 735-754.

Kim MJ, Wang AR, Park JS and Kim I (2014) Complete mitochondrial genomes of five skippers (Lepidoptera: Hesperiidae) and phylogenetic reconstruction of Lepidoptera. Gene 549: 97-112.

Kodandaramaiah U and Wahlberg N (2007) Out of Africa origin and dispersal mediated diversification of the butterfly genus Junonia (Nymphalidae: Nymphalinae). J. Evol. Biol. 20: 2182-2191.

Korber B (2000) HIV Signature and Sequence Variation Analysis. Computational Analysis of HIV Molecular Sequences, Chapter 4: 55-72. Allen G. Rodrigo and Gerald H. Learn, eds. Dordrecht, Netherlands: Kluwer Academic Publishers. HIV database website (www.hiv.lanl.gov).

Knight RD, Freeland SJ and Landweber LF (2001) A simple model based on mutation and selection explains trends in codon and amino-acid usage and GC composition within and across genomes. Genome Biol. 2: RESEARCH0010.

Kruse JJ and Sperling FAH (2001) Molecular phylogeny within and between species of the *Archips argyrospila* complex (Lepidoptera: Tortricidae). Ann. Entomol. Soc. Am. 94: 166-173.

Kumar NP, RajavelAR, Natarajan R and Jambulingam P (2007) DNA barcodes can distinguish species of Indian mosquitoes (Diptera: Culicidae). J. Med. Entomol. 44: 1-7.

Lambshead PJD, Platt HM and Shaw KM (1983) Detection of differences among assemblages of marine benthos species based on an assessment of dominance and diversity. J. Nat. Hist. 17: 859-874.

Lemey P, Rambaut A, Drummond AJ and Suchard MA (2009) Bayesian Phylogeography Finds Its Roots. PLoS.Comput. Biol. 5(9): e1000520. doi:10.1371/journal.pcbi.1000520.

Lery X, LaRue B, Cossette J, and Charpentier G (2003) Characterization and authentication of insect cell lines using RAPD markers.Insect.Biochem. Mol. Biol. 33: 1035-1041.

Lewis EB (1978) A gene complex controlling segmentation in *Drosophila*. Nature 276: 565-570.

Librado P and Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinf. 25: 1451-1452.

Lin CP and Danforth BN (2004) How do insect nuclear and mitochondrial gene substitution patterns differ? Insights from Bayesian analyses of combined datasets. Mol. Phylogenet. Evol. 30: 686-702.

Linares MC, Soto-Calderón ID, Lees DC and Anthony NM (2009). High mitochondrial diversity in geographically widespread butterflies of Madagascar: A test of the DNA barcoding approach Mol. Phylogenet. Evol. 50: 485-495.

Linnaeus C (1758). Systema naturæ per regna tria naturæ, secundum classes, ordines, genera, species, cum characteribus, differentiis, synonymis, locis 1 (10thed.). Stockholm: Laurentius Salvius. (1-4): 1-824.

Liu XZ, Peng ZB, Fu JH and Huang CL (1998) Maize inbred line grouping by using cluster analysis of RAPD molecular marker, phenotype and heterosis. Acta.Agriculturae.Boreali.Sinica. 13: 36-41.

Lou KF, Weiss MJ, Bruckner PL, Morrill WL, Talbert LE and Martin JM (1998) RAPD Variatin within and among Geographix populations of wheat stem Sawfly (*Cephus cintus Norton*). AGA.89**:** 329-335.

Lübberstedt T, Melchinger AE, Duble C, Vuylsteke M and Kuiper M (2000) Relationships among early European maize inbreds. IV. Genetic diversity revealed with AFLP markers and comparison with RFLP, RAPD and pedigree data. Crop. Sci. 40: 783-791.

Lynn DJ, Singer GA and Hickey DA (2002) Synonymous codon usage is subject to selection in thermophilic bacteria. Nucl. Acids Res. 30: 4272-277.

Ma C, Yang PC, Jiang F, Chapuis MP and Shall Y (2012) Mitochondrial genomes reveal the global phylogeography and dispersal routes of the migratory locust. Mol. Ecol.21: 4344-58.

Majumder J, Lodh R and Agarwala BK (2012) Variation in butterfly diversity and unique species richness along different habitats in Trishna Wildlife Sanctuary, Tripura, northeast India. Check List 8(3): 432-436.

Mani MS (1974) Ecology and Biogeography in India- Vol-1, Dr. W. Junk B.V. Publishers, The Hague.725pp.

Martin JF, Gilles A and Descimon H (2000) Molecular Phylogeny and Evolutionary Patterns of the European Satyrids (Lepidoptera: Satyridae) as Revealed by Mitochondrial Gene Sequences. Mol. Phylogenet. Evol. 15(1): 70-82.

Martin M (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.Journal. 17(1): 10.

McaleeceN, Gage JD, Lambshead J, and Patterson GLJ (1997) Biodiversity Professional. The Natural History Museum &The Scottish Association for Marine Science.

McDevit DC and Saunders GW (2009) On the utility of DNA barcoding for species differentiation among brown macroalgae (Phaeophyceae) including a novel extraction protocol. Phycol. Res. 57: 131-141.

Meyer CP and Paulay G (2005) DNA Barcoding: Error Rates Based on Comprehensive Sampling. PLoS Biol. 3(12): e422.

Meyer R, Hofelein C, Luthy J and Candrian U (1995) Polymerase chain reaction-restriction fragment length polymorphism analysis: a simple method for species identification in food. J. AOAC. Int. 78: 1542-51.

Milan T, Osuna F, Cobos S, Torres AM and Cubero JI (1996) Using RAPDs to study phylogenetic relationships in Rosa.Theor. Appl. Genet. 92: 237-277.

Mitaku S and Hirokawa T (1999) Physicochemical factors for discriminating between soluble and membrane proteins: hydrophobicity of helical segments and protein length. Protein Eng. 11

Monteiro A and Pierce NE (2001) Phylogeny of *Bicyclus* (Lepidoptera: Nymphalidae) inferred from COI, COII and EF-1alpha gene sequences. Mol. Phylogenet. Evol. 18: 264-281.

Moritz C, Patton JL, Schneider CJ and Smith TB (2000) Diversification of Rainforest Faunas: an integrated molecular approach. Annu. Rev. Ecol. Syst. 31: 533-563.

Morton BR (1993) Chloroplast DNA codon use: Evidence for selection at the psbA locus based on tRNA availability. J. Mol. Evol. 37: 273-280.

Moura GR, Pinheiro M, Freitas A, Oliveira JL, Frommlet JC, Carreto L, Soares AR, Bezerra AR and Santos MAS (2011) Species-Specific Codon Context Rules Unveil Non-Neutrality Effects of Synonymous Mutations. PLoS ONE. 6 (10): e26817.

Moya A, Goya P, Cifuentes F and Cenis JL (2001) Genetic diversity of Iberian populations of *Bemisia tabaci* (*Hemiptera: Aleyrodidae*) based on random amplified polymorphic DNA-polymerase chain reaction. Mol. Ecol. 10: 891-897.

Murthy MS, Sannaveerappanavar VT and Shankarappa KS (2014) Genatic Diversity of diamondback Moth, *Plutella xylostella* L. (Yponomeutidae: Lepidoptera) populations in India using RAPD markers. J. Entomol. 11(2): 95-101.

Myers N, Mittermeier RA, Mittermeier CG, da Fonseca GAB and Kent J (2000) Biodiversity hotspots for conservation priorities. Nature 403: 853-858.

Nakamura H, Kaneko S, Yamaoka Y, and Kakishima M (1998) Differentiation of Melampsora rust species on willows in Japan using PCR-RFLP analysis of ITS regions of ribosomal DNA. Mycoscience39: 105-113.

Nelson LA, Lambkin CL, Batterham P, Wallman JF and Dowton M (2012) Beyond barcoding: a mitochondrial genomics approach to molecular phylogenetics and diagnostics of blowflies (Diptera: Calliphoridae). Gene511: 131-42.

Nielsen R and Yang Z (1998) Likelihood models for detecting positively selected amino acid sites and applications to the HIV-1 envelope gene. Genetics 148: 929-936.

Norton E (1872) Notes on North American Tenthridinae with descriptions of new species. Trans. Am. Entomol. Soc. 4: 77-86.

NSG (Nymphalidae Systematics Group) (2009) The NSG's voucher specimen database of Nymphalidae butterflies. Version 1.0.15.

Ohta T (1995) Synonymous and nonsynonymous substitutions in mammalian genes and the nearly neutral theory. J. Mol. Evol. 40: 56-63.

Ophir R, Itoh T, Graur D, Gojobori T (1999) A simple method for estimating the intensity of purifying selection in protein-coding genes.Mol. Biol. Evol.16: 49-53.

Pachuau R **(**1994) Geography of Mizoram; *R.T. Enterprise,* Aizawl.

Pachuau L, Vanlalruati C and Senthil Kumar N (2012) Morphological versus molecular characterization of three similar pierid species of Butterflies. Int. J. Pharm. Bio. Sci. 3(4): 1091-1102.

Pagel M (1999) Inferring the historical patterns of biological evolution. Nature 401: 877-884.

Panganiban G, Nagy L and Carroll SB (1994) The role of the *Distal-less* gene in the development and evolution of insect limbs. Curr. Biol. 4: 671-675.

Panganiban G, Sebring A, Nagy L, Carroll SB (1995) The development of crustacean limbs, and the evolution of arthropods. Science 270: 1363-1366.

Panganiban G, Irvine SM, Lowe C, Roehl H, Corleyi LS, Sherbon B, Grenier JK, FallonJF, Kimble J, Walker M, Wray GA, Swallai BJ, Martindale MQ and Carroll SB (1997) The origin and evolution of animal appendages. Proc. Natl. Acad. Sci. USA. (94): 5162-5166.

Paradis E (1998) Detecting shifts in diversification rates without fossils. Am. Nat. 152(2): 176-187.

Park MH, Sim CJ, Baek J and Min GS (2007) Identification of genes suitable for DNA barcoding of morphologically indistinguishable Korean Halichondriidae sponges. Mol. Cell. 23: 220-227.

Patterson C, Williams DM and Christopher J (1993) Congruence between Molecular and Morphological Phylogenies. Humpries Annu. Rev. Ecol. Syst. 24: 153-188.

Patterson J, Chamberlain B and Thayer D (2006).FinchTV 1.4.0 (Geospiza, Inc.; Seattle, WA, USA; http://www.geospiza.com).

Perna NT and Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. 41: 353-358.

Perrier X and Jacquemoud-Collet JP (2006)DARwinsoftware.http://darwin.cirad.fr/darwin.

Pollard E (1977) A method for assessing changes in the abundance of butterflies. Biol. Conserv. 12: 115-124.

Pornkulwat S, Skoda SR, Thomas GD and Foster JE (1998) Random amplified polymorphic DNA used to identify genetic variation in ecotypes of the European corn borer (Lepidoptera: Pyralidae). Ann. Entomol. Soc. Amer. 91: 719-725.

Powell JR and Moriyama EN (1997) Evolution of codon usage bias in Drosophila. Proc. Natl. Acad. Sci. U S A. 94: 7784-7790.

Prevost A and Wilkinson MJ (1999) A new system of comparing PCR primers applied to ISSR fingerprinting of potato cultivars. Theor. Appl. Genet. 98: 107-112.

Rach J, DeSalle R, Sarkar IN, Schierwater Band Hadrys H (2008) Character-based DNA barcoding allows discrimination of genera, species and populations in Odonata. Pro. R. Soc. B. Biol. Sci. 275: 237-247.

Rajagopal T, Sekar M and Archunan G (2011) Diversity and community structure of butterfly of Arignar Anna Zoological Park, Chennai, Tamil Nadu. J. Env. Biol. 32: 201-207.

Rambaut A and Drummond AJ (2007) Tracer v1.4.Published by authors: http://beast.bio.ed.ac.uk/Tracer.

Rand DB, Heath A, Suderman T and Pierce NE (2000) Phylogeny and life history evolution of the genus *Chrysotis* within Aphnaeini (Lepidoptera: lycaenidae), inferred from mitochondrial cytochrome oxidase I sequences. Mol. Phylogenet. Evol. 17: 85-96.

Rasmussen HB (2012) Restriction Fragment Length Polymorphism Analysis of PCR-Amplified Fragments (PCR-RFLP) and Gel Electrophoresis - Valuable Tool for Genotyping and Genetic Fingerprinting, Gel Electrophoresis - Principles and Basics, Dr. Sameh Magdeldin (Ed.), ISBN: 978-953-51-0458-2, InTech.

Remigio EA and Hebert PDN (2003) Testing the utility of partial COI sequences for phylogenetic estimates of Gastropod relationships. Mol. Phylogenet. Evol. 29: 641-647.

Remington CL (1968) A new sibling *Papilio* from the Rocky Mountains, with genetic and biological notes (Insecta, Lepidoptera).Postilla, 119: 1-40.

Rohlf FJ (1998) NTSYS-pc: Numerical Taxonomy and Multivariate Analysis System, Version 2.201i. New York: Exeter Software, Applied Biostatics Inc.

Ronquist F and Huelsenbeck JP (2003) MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinf.19: 1572-1574.

Roldan-Ruiz I, Dendauw J, VanBockstaele E, Depicker A and Loose MD (2000). AFLP markers reveal high polymorphic rates in ryegrasses (Lolium spp.). Mol. Breed. 6: 125-134.

Roy, AD and Karanth KP (2009) The Out-of-India Hypothesis: What do Molecules suggest?. J. Biosci. 34(5): 687-697.

Saccone C, De Giorgi C, Gissi C, Pesole G and Reyes A (1999) Evolutionary genomics in Metazoa: the mitochondrial DNA as a model system. Gene 238: 195-209.

Salvato P, Simonato M, Battisti A and Negrisolo E (2008) The complete mitochondrial genome of the bag-shelter moth *Ochrogaster lunifer* (Lepidoptera: Notodontidae). BMC Genomics9: 331.

Sambrook J, Fritsch EF and Maniatis T (1989) Molecular Cloning: A Laboratory Manual, 2nd ed. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Santos VM, Maria Macoris, Maria Andrighetti, Avila PE and Kirchgatter K (2003) Analysis of genetic relatedness between populations of aedes aegypti from different geographic regions of são paulo state, brazil. Rev. Inst. Med. trop. S. Paulo. 45(2): 99-101.

Satta Y, Ishiwa H and Chigusa.SI (1987) Analysis of nucleotide substitutions of mitochondrial DNAs in Drosophila melanogaster and its sibling species. Mol. Biol. Evol. 4: 638- 650.

Schattner P, Brooks AN and Lowe TM (2005) The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucl. Acids Res. 33: 686-689.

Schlick-Steiner B C, Steiner F M, Moder K, Seifert B, Sanetra M, Dyreson E, Stauffer C and Christian E (2006) A multidisciplinary approach reveals cryptic diversity in Western Palearctic *Tetramorium* ants (Hymenoptera: Formicidae). Mol. Phylogenet. Evol. 40: 259-273.

Schoppmeier M, Damen WGM (2001) Double-stranded RNA interference in the spider *Cupiennius salei*: the role of *Distalless* is evolutionarily conserved in arthropod appendage formation. Dev. Genes. Evol. 211: 76-82.

Seifert KA, Samson RA, Dewaard JR, Houbraken J, Levesque CA, Moncalvo JM, Louis-Seize G and Hebert PDN (2007) Prospects for fungus identification using C01DNAbarcodes, with *Penicillium* as a test case. Proc. Natl. Acad. Sci. 104: 3901-3906.

Shao R, Dowton M, Murrell A and Barker SC (2003) Rates of gene rearrangements and nucleotide substitution are correlated in the mitochondrial genomes of insects. Mol. Biol. Evol.20: 1612-19.

Sharma VL, Bhatia S, GillnTK, Badran AA, Kumari M, Singh JJ and Sobti RC (2006) Molecular characterization of two species of butterflies (Lepidoptera: Insecta) through RAPD-PCR technique. Cytologia 71: 81-85.

Sharma VL, Kaur P, Gill TK, Kumari M and Sobti RC (2010) Genetic Characterisation in two species of Catopsilia (Pieridae: Lepidoptera) by RAPD-PCR technique. Caryologia 63(3): 250-256.

Sharma VL, Sobti RC, Gill TK, Suman K, Badran and Mamtesh K (2006) Molecular Studies of Five Species of Butterflies (Lepidoptera: Insecta) Through RAPD-PCR Technique. Caryologia 59(3): 226-234

Sharp MP and Li WH (1986) Codon usage in regulatory genes in *Escherichia coli* does not reflect selection for rare codons. Nucl. Acids Res. 14(19): 7737-7749.

Sharp MP and Li WH (1987) The codon adaptation index- a measure of directional synonymous codon usage bias, and its potential applications. Nucl. Acids Res. 15(3): 1281-1295.

Sharp MP and Li WH (1989) On the rate of DNA sequence evolution in Drosophila. J. Mol. Evol. 28: 398-402.

Sharp MP and Matassi G (1994) Codon usage and genome evolution.Curr.Opin.Genet. Dev. 4: 851-860.

Shearer TL, Van-Oppen MJ, Romano SL and Worheide G (2002) Slow mitochondrial DNA sequence evolution in the Anthozoa (Cnidaria). Mol. Ecol. 11: 2475-2487.

Shi QH, Zhao F, Hao JS and Yang Q (2013) Complete mitochondrial genome of the Common Evening Brown, Melanitis leda Linnaeus (Lepidoptera: Nymphalidae: Satyrinae). Mitochondr.DNA. 24: 492-494.

Silberman JD and Walsh PJ (1992) Species identification of spiny lobster phyllosome larvae via ribosomal DNA analysis.Mol.Mar. Biol. Biotechnol.1: 195-205.

Silva DL, Day JJ, Elias M, Willmott K, Whinnett A and Mallet J (2010). Molecular phylogenetics of the neotropical butterfly subtribe Oleriina. Mol. Phylogenet. Evol. 5: 1032-1041.

Silvestro D and Michalak I (2012) raxmlGUI: A graphical front-end for RAxML. Org. Divers.Evol.12: 335-337.

Simon C (1991) Molecular systematics at the species boundary: exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from amplified DNA. 3-71 in G. M. HEWITT, A. JOHNSTON, and J. YOUNG, eds. Molecular taxonomy. NATO Advanced Studies Institute, Springer, Berlin.

Simon C, Frati F, Beckenbach A,Crespi B,Liu H and Flook P (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87: 651-701.

Sivakumar K, Balaji S and Gangaradhashnan (2007) In silico characterization of antifreeze proteins using computational tools and server. J. Chem. Sci. 119: 571-579.

Smith MA, Fisher BL and Hebert PDN (2005) DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar.*Philos.* Trans. R. Soc. Lond. B. Biol. Sci.360(1462)**:** 1825-1834.

Smith MA, Rodriguez J, Whitfield J, Deans A, Janzen DH, Hallwachs W and Hebert PDN (2008) Extraordinary diversity of parasitoid wasps exposed by iterative integration of natural history, DNA barcoding, morphology and collections. Proc. Natl. Acad. Sci. USA. 105: 12359-12364.

Sodhi NS, Koh LP, Brook BW and Ng PKL (2004) Southeast Asian biodiversity: an impending disaster. Trends. Ecol. Evol. 19: 654-660.

SokiWiki (2014) *K dominance curves - Indicators - Confluence*, SOKI, Antarctic Climate and Ecosystems Co-operative Research Centre. http://www.soki.aq/x/e4Fm

Sperling FAH and Hickey DA (1994) Mitochondrial DNA sequence variation in the spruce budworm species complex (*Choristoneura:*Lepidoptera). Mol. Biol. Evol. 11(4): 656-665.

Stapel H, Misof B and Wagner T (2008) A Molecular and Morphological Phylogenetic analysis of Afrotropical monolepta species and related Galerucinae (Coleoptera: Chrysomelidae). Arthropod Syst. Phylo. 66(1): 3-17.

Sutherland WJ (1996) Ecological Census Techniques.University Press, Cambridge.

Swofford DL (2002) PAUP (Phylogenetic Analysis Using Parsimony and Other Methods), version 4.0b10 (Altivec).Massachussets, USA, Sinauer, Sundaland.

Taanman JW (1999) The mitochondrial genome: structure, transcription, translation and replication. Biochim.Biophys.Acta. 1410: 103-123.

Tamura K (1992) The Rate and Pattern of Nucleotide Substitution in Drosophila Mitochondrial DNA. Mol. Biol. Evol. 9(5): 814-825.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, and Kumar S (2011) MEGA5: Molecular Evolutionary Genetics Analysis Using Maximum Likelihood, Evolutionary Distance, and Maximum Parsimony Methods. Mol. Biol. Evol. 28(10): 2731-2739.

Thomas JA (2005) Monitoring change in the abundance and distribution of insects using butterflies and other indicator groups. Philo. Trans. R. Soc. B. Biol. Sci. 360: 339-357.

Thomas JA, Telfer MG, Roy DB, Preston CD, Greenwood JJD, Asher J, Fox R, Clarke RT and Lawton JH (2004) Comparative losses of British butterflies, birds, and plants and the global extinction crisis. Science 303: 1879-1881.

Tian LL, Sun XY, Chen M, Gai Y, Hao JS and Yang Q (2012) Complete mitochondrial genome of the Five-dot Sergeant *Parathyma sulpitia* (Nymphalidae: Limenitidinae) and its phylogenetic implications. Zool. Res. 33:133-43.

Tiple AD, Khurad AM and Dennis RLH (2007) Butterfly diversity in relation to human-impact gradient on an Indian University Campus, Nota. Lepidopterol. 30(1): 179-18.

Tiple A, Agasthe D, Khurad AM and Kunte K (2009) Population dynamics and seasonal polyphenism of Chilades pandava butterfly (Lycaenidae) in Central India. Curr. Sci. 97(12): 1774-1779.

Turner H, Hovenkamp P and van Welzen PC (2001) Biogeography of Southeast Asia and the West Pacific. J. Biogeogr. 28: 217- 230.

Vandewoestijne S and Baguette M (2002) The genetic structure of endangered populations in the Cranberry Fritillary, *Boloria aquilonaris* (Lepidoptera, Nymphalidae): RAPDs vs allozymes. Heredity 89: 439-445.

Vandewoestijne S, Baguette M, Brakefield PM, and Saccherib IJ (2004) Phylogeography of Aglais urticae (Lepidoptera) based on DNA sequences of the mitochondrial COI gene and control region. Mol. Phylogenet. Evol. 31: 630-646.

Vanlalruati C, Zothansangi, Gurusubramanian G and Senthil Kumar N (2011) Morphological and molecular studies of six Junonia species of Butterflies using RAPD-PCR technique. Sci. Vis. 11(3): 141-146.

van Swaay C, Warren M and Lois G (2006) Biotope use and trends of European butterflies. J. Insect. Conserv. 10: 189-209.

Varshney RK (2006) An estimate of the numbers of butterfly species in the Indian Region. Bionotes 8(3): 61-63.

Vences M, Thomas M, Bonett RM and Vieites DR (2005) Deciphering amphibian diversity through DNA barcoding: chances and challenges. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360: 1859-1868.

Vogel G (1997) Phylogenetic analysis: getting its day in court. Science275: 1559-1560.

Voronkov A, Hop H and Gulliksen (2013) Diversity of hard-bottom fauna relative to environmental gradients in Kongsfjorden, Svalbard.Polar. Res. 32: 11208.

Vucetich LM, Vucetich JA, Joshi CP, Waite TA and Peterson RO (2001) Genetic (RAPD) diversity in *Peromyscus maniculatus* populations in a naturally fragmented landscape. Mol. Ecol. 10: 35-40.

Wahlberg N, Weingartner E, Warren AD and Nylin S (2009) Timing major conflict between mitochondrial and nuclear genes in species relationships of *Polygonia* butterflies (Nymphalidae: Nymphalini). BMC Evol. Biol. 9: 92.

Wakeley J (1996) The variance of pairwise nucleotide differences in two populations with migration. Theor.Popul. biol. 49: 39-57.

Wan XF, Xu D, Kleinhofs A and Zhou J (2004) Quantitative relationship between synonymous codon usage bias and GC composition across unicellular genomes.BMC Evol. Biol. 4(1): 19.

Wang XC, Sun XY, Sun QQ, Zhang DX, Hu J, Yang Q and Hao JS (2011) The completemitochondrial genome of the laced fritillary *Argyreushyperbius*(Lepidoptera: Nymphalidae).Zool. Res. 32: 465-475.

Ward RD, Zemlak TS, Innes BH, Last PR and Hebert PDN (2005) DNA barcoding Australia's fish species. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 360: 1847-1857.

Warren RW, Nagy LM, Seleque J, Gates J and Carroll SB (1994) Evolution of homeotic gene regulation in flies and butterflies. Nature 372: 458–461.

Wiesing K, Nybom H, Wolff K and Meyer W (1995) DNA fingerprinting in plants and fungi. CRC Press, Boca Raton. 322pp.

Weller SJ, Pashley DP and Martin JA (1996) Reassessment of butterfly family relationships using independent genes and morphology.Ann. Entomol. Soc. Am. 89: 184–192.

Welsh J and McClelland M (1990) Fingerprinting genomes using PCR with arbitrary primers.Nucl. Acids Res. 18:7213-7218.

Williams CL, Goldson SL, Baird DB and Bullock DW (1994) Geographical origin of an introduced insect pest, *Listronotus bonariensis* (Kuschell), determined by RAPD analysis. Heredity 72: 412-419.

Williams J and Carroll SB (1993) The origin, patterning and evolution of insect appendages. BioEssays 15: 567-577.

Williams JGK, Kubelik AR, Livak KJ, Rafalski JA and Tingey SV (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucl. Acids Res. 18: 6531-6535.

Wilson EQ (1992) Fluctuations in abundance of tropical insects.Am. Nat. 112: 1017-1045.

Wolf *C*, Burgener M, HuKbner P and LuKthy J (2000) PCR-RFLP Analysis of Mitochondrial DNA:Differentiation of Fish Species. Lebensm.-Wiss. u.-Technol. 33: 144-150.

Wolstenholme DR and Clary DO (1985) Sequence evolution of Drosophila mitochondrial DNA. Genetics 109(4): 725-744.

Wolstenholme DR (1992) Animal mitochondrial DNA: structure and evolution. Int. Rev. Cyt. 141: 173-216.

Wright F (1990) The 'effective number of codons' used in a gene. Gen. 87: 23-29.

Wu LW, Lin LH, Lees DC and HsuYF (2014) Mitogenomic sequences effectively recover relationships within brush-footed butterflies (Lepidoptera: Nymphalidae). BMC Genomics 15: 468.

Wu M (2000) Genetic diversity and its relationship to hybrid performance and heterosis in maize as revealed by AFLPs and RAPDs.Maize Genet. Coop. Newsl. 74: 62-63.

Wu ZZ, Li HM, Bin SY, Ma J, He HL, Li XF, Gong FL and Lin JT (2014) Sequence analysis of mitochondrial ND1 gene can reveal the genetic structure and origin of Bactrocera dorsalis s.s. BMC Evol. Biol. 14: 55.

Wyman SK, Jansen RK and Boore JL (2004) Automatic annotation of organellar genomes with DOGMA. Bioinf. 20(17): 3252-3255.

Wynter-Blyth MA (1957) Butterflies of the Indian Region.Bombay Natural History Society.

Yang Z and Rannala B (2012) Molecular phylogenetics: principles and practice. Nat. rev. Genet. 13: 303-314.

Yang Z and Yoder AD (1999) Estimation of the Transition/Transversion rate bias and species sampling. J. Mol. Evol. 48(3): 274-283.

Zachariah G, Fordyce JA, Forister ML and Nice CC (2008) Recent colonization and radiation of North American Lycaeides (Plebejus) inferred from mtDNA. Mol. Phylogenet. Evol. 48: 481-490.

Zakharov EV (2001) Natural Hybridization between Two Swallowtail Species Parnassius nomion and Parnassius bremeri (Lepidoptera, Papilionidae) Shown by RAPD–PCR.Russ. J. Genet. 37 (4): 375-383. Translated from Genetika. 37(4): 475-484.

Zakharov EV, Caterino MS and Sperling FAH (2004) Molecular Phylogeny, Historical Biogeography, and Divergence Time Estimates for Swallowtail Butterflies of the Genus *Papilio* (Lepidoptera: Papilionidae). Syst. Biol.53(2): 193-215.

Zhang M, Nie X, Cao T, Wang J, Li T, Zhang X, Guo Y, Ma E and Zhong Y (2012) The complete mitochondrial genome of the butterfly Apatura metis (Lepidoptera: Nymphalidae). Mol. Biol. Rep. 39: 6529-6536.

Zhao F, Huang DY, Shi QH, Hao JS, Sun XY, Zhang LL and Yang Q (2013) The firstmitochondrial genome for the butterfly family Riodinidae and its systematic implications.Zool. Res. 34: 109-119.

Zheng X, Cai D, Potter D, Postman J, Liu J and Teng Y (2014) Phylogeny and evolutionary histories of Pyrus L revealed by phylogenetic trees and networks based on data from multiple DNA sequences. Mol. Phylogenet. Evol. 80: 54-65.

Zimmermann M, Walhberg N and Descimon H (2000) Phylogeny of Euphydryas Checkerspot Butterflies (Lepidoptera: Nymphalidae) Based on Mitochondrial sequence data. Ann. Entomol. Soc. Am. 93: 347-355.

Zothansangi, Vanlalruati C, Laldikbera I, Gurusubramanian G and Senthil Kumar N (2011) Butterfly diversity in Mizoram University Campus, Aizawl, India and their global distributions. AEC-070. ISBN: 978-93-81361-53-5. Excel India Publishers. New Delhi.

Papers published and submitted

Vanlalruati C, De Mandal S, Gurusubramanian G, Karanth P and Senthil Kumar N (2015) The Complete mitochondrial genome of Chocolate Pansy, *Junonia iphita* (Lepidoptera: Nymphalidae: Nymphalinae). Mitochondrial DNA (DOI: 10.3109/19401736.2015.1033701).

Vanlalruati C, Zothansangi, Gurusubramanian G and Senthil Kumar N (2011) Morphological and molecular studies of six *Junonia* species of butterflies using RAPD-PCR technique. Sci. Vis. 11(3): 141-146.

Pachuau L, Vanlalruati C and Senthil Kumar N (2012) Morphological versus molecular characterization of three similar pierid species of butterflies. International Journal of Pharmaceutical and BioSciences 3(4): 1091-1102.

Vanlalruati C, Zothansangi, Gurusubramanian G and Senthil Kumar N (2015) RAPD-PCR discriminates the intra-specific genetic polymorphism in a metapopulation of nymphalidae species. Journal of King Saud University-Science (In press).

De Mandal S, Vanlalruati C, Gurusubramanian G and Senthil Kumar N (2015) Genetic variation in Mitochondrial Cytochrome Oxidase I Gene within the Butterfly species. Computational Life Sciences: Interdiscip. Sci. (Communicated).

Vanlalruati C, De Mandal S, Gurusubramanian G, and Senthil Kumar N (2015) Near complete mitochondrial genome of the *Junonia iphita* and its intraspecific variation. Genomic data (Communicated).

Zothansangi, Vanlalruati C, Senthil Kumar N and Gurusubramanian G (2011) Genetic variation within two cryptic species of *Cirrochroa* (Heliconiinae: Lepidoptera) by RAPD-PCR technique. Sci. Vis. 11(3): 165-170.

Zothansangi, Vanlalruati C, Lalremsanga HT, Senthil Kumar N and Gurusubramanian G (2011) Butterflies of Mizoram, North-eastern India with distinction between two

cryptic species (*Cirrochroa sps.*) through RAPD-PCR analysis. International conference on conservation and management of pollinators.Kuching, Sarawak, Malaysia.

Zothansangi, Vanlalruati C, Laldikbera I, Gurusubramanian G and Senthil Kumar N (2011) Butterfly diversity in Mizoram University Campus, Aizawl, India and their global distributions. Advance in Environmental Chemistry-070. ISBN: 978-93-81361-53-5. Excel India Publishers. New Delhi.

Lalrotluanga, Laldikbera I, Vanlalruati C, Zothansanga, Brindha S, Senthil Kumar N and Gurusubramanian G (2008) Butterfly faunal diversity in Aizawl, Mizoram, India. Sci.Vis. 8(3): 65-75.

Vanlalruati C, Gurusubramanian G, Karanth P and Senthil Kumar N (2015) Intra-specific variations in *Elymnias hypernmestra* (Lepidoptera: Nymphalidae: Satyrinae) as revealed by the mitochondrial COI and ND1 genes (Manuscript under preparation to Molecular Phylogenetics and Evolution).