# Cytochrome b gene as a potential DNA barcoding marker in ecoraces of tropical Tasar silkworm *Antheraea mylitta* Drury

Priti Pragyan Ray [a,*], Barsha Barala [a], Purushottam Dash [b]

[a] *Sericulture Laboratory, Post Graduate Department of Zoology, Shailabala Women's Autonomous College, Cuttack, Odisha, India*
[b] *Directorate of Sericulture, Government of Odisha, India*

## ARTICLE INFO

## ABSTRACT

*Background:* Tropical Tasar silkworm *Antheraea mylitta* Drury plays a very crucial role in the economic development of the country employing around 9.76 million people in rural and semi-urban areas. Since the population of this species has spread over many states with varying environmental factors, the population has been isolated into many pockets, specifically adapted to that niche, and are called as "ecoraces". Even though phenotypically these ecoraces may look similar to some level but the intra-generic genotypic variability between them is correlated to the economically important yield traits of their cocoons. The purpose of our study was to evaluate the effectiveness of Cytochrome b (Cyt b) gene of mitochondrial genome as a potential DNA barcoding marker for identification of these ecoraces as to are they really ecoraces, and attempts to establish the first country report of DNA barcode reference library of this silkworm species.
*Methods and results:* DNA was extracted from pupae of pure ecoraces collected from core ecozones. PCR amplicon was bead purified and further subjected to Sanger Sequencing. The parsimony tree was also created using the MEGA 11 software. The genetic distance, gene flow and relation among the ecoraces was evaluated. The conserved regions of Cyt b gene were identified along with the documentation of the mode and tempo of nucleotide variation and evolution.
*Conclusion:* The present study confirms the efficiency of Cyt b gene as a potential DNA marker for identification of ecoraces of Tasar silkworm and is the first country report on barcoding of *A. mylitta* Drury. Apart from being used as an early-warning system for ecological degradation, this knowledge will help in assessing the evolutionary relationship of these economically important ecoraces and identify their position in the process of evolution aiding in the assessment of conservation priorities.

## 1. Introduction

The Tasar silk is obtained from the cocoons of Tasar silkworm, semi-domesticated wild type of silk producing variety, belonging to the genus *Antheraea* (Family: Saturniidae). Till date, more than 100 species of *Antheraea* have been reported on earth and out of those species (Kitching et al., 2018), only three species are exploited for silk production in India, namely *Antheraea mylitta* Drury (commonly known as Tropical Tasar silkworm), *Antheraea paphia* Linn. (commonly known as South India small Tasar silkworm), and *Antheraea assamensis* Helfer (Muga Silkworm). Since the species *Antheraea mylitta* Drury is subjected to a wide range of climatic and environmental conditions, the whole population has differentiated itself into many pockets of population termed as

"Ecoraces". These ecoraces exhibit phenotypic variation among themselves, owing to their polyphagous nature, the location of ecozones, etc. Till date, a total 44 ecoraces have been found in country India (Singh and Srivastava, 1997; Srivastava and Sinha, 2002) and in the state Odisha seven ecoraces have been reported, i.e. Modal, Nalia, Sukinda, Jata-Daba, Boudh, Umerkote and Adaba. The ecozones of these seven ecoraces have been mentioned in Table 1. These ecoraces are named in accordance to the name of the places where they were first identified in the state and studies till date are made only at the phenotypic level and there have been no reports of genetic characterisation of these ecologically and economically important ecoraces found in the state of Odisha so far. The variation in their genetic nature is an important factor for conservation of the biodiversity and gene pool of this animal. Many of

---

**Table 1**
Ecozones of the *A. mylitta* drury ecoraces of odisha.

| Sl no | Ecorace name | Ecozone | District of the ecozone |
|---|---|---|---|
| 1 | Modal | Similpal Biosphere Reserve and nearby forest areas | Mayurbhanj |
| 2 | Nalia | Some regions of Similpal Biosphere Reserve and Ghatagaon Block | Mayurbhanj and Keonjhar |
| 3 | Jata-Daba | Thakurmunda, Kendujuiani, Mahuldiha | Mayurbhanj |
| 4 | Sukinda | Sukindagarh | Jajpur |
| 5 | Boudh | Sunabeda forest range | Koraput |
| 6 | Adaba | Adaba forest | Gajapati |
| 7 | Umerkote | Umerkote | Nabarangpur |

the researchers involved with *A. mylitta* have previously worked on behaviour of oviposition pattern (Soundappan et al., 2021), host selection (Thirupam et al., 2023), avian pests (Reddy et al., 2020), while some of biochemical works are focused on biochemical characterisation of midgut bacterial symbionts through metagenomics study (Baig et al., 2023), immunohistochemical localisation of histamine neuropeptide (Barsagade et al., 2022), structural and functional characterisation of cocoonase enzyme (Sneha and Pandey, 2022; Rani et al., 2024), sericin protein (Jena et al., 2021) of various commercial ecoraces such as Daba, Railey, Andhra Local etc. Some molecular genetics works recently have been focused on molecular cloning and development of RAPD-SCAR (Random Amplified Polymorphic DNA-Sequence Characterized Amplified Region) markers (Prabhu et al., 2023), Arylphorin protein (Dutta et al., 2020) and genomic isolation and quantification of DNA of ecoraces such as Daba, Railey, Andhra local etc. (Manda et al., 2019). As far as characterisation and molecular studies is concerned, NGS (Next Generation Sequencing) has been implemented for some ecoraces (Andhra local and Daba) (Renuka and Shamitha, 2024), yet for characterisation and molecular studies, Odisha's endemic ecoraces (Modal, Nalia, Jata-Daba, Sukinda, and Boudh) are still under veil from global researchers, even when the potentiality of these ecoraces is higher for tasar silk production.

The genes of mitochondrial DNA (mt-DNA) are considered one of the greatest genetic markers to study ecology, evolutionary biology, and phylo-genetics of a species because of their maternal inheritance, near-neutrality (Ballard and Kreitman, 1995), technical ease of use, lack of recombination, and rapid evolutionary rate (almost 4times) compared to nuclear DNA (Ballard and Rand, 2005). The insect mitogenome is circular and double stranded, consisting of 15-18 kb in length (Cameron, 2014) and 37 genes including 13 mitochondrial protein-coding genes (PCGs), 2 ribosomal RNA genes, and 22 t-RNA genes, and non- coding A + T-rich region which participate in control of replication and transcription (Cameron, 2014; Boore, 1999). The 13 PCGs (i.e. ND1, ND2, ND3, ND4, ND4L, ND5, ND6, COI, COII, COIII, ATP6, ATP8 and Cytochrome b) of mt-DNA are better targets as compared to nuclear genes for the barcoding purpose since the indels (insertion/deletion polymorphism) are rare and most of it led to a shift in the reading frame (Parson et al., 2000). Even though COI gene has proven itself very useful for the barcoding purpose, in many cases Cytochrome b (Cyt-b) gene has proven to be very useful as well (Parson et al., 2000; De Pancorbo et al., 2004; Caine et al., 2006; Farag et al., 2020). In class Insecta, two supra-generic studies have been done on the parasitic wasp family (Gimeno et al., 1997) and among basal families of Beetles (Ballard et al., 1998). Simmons and Weller (2001a, 2001b) had conducted research to understand the evolution and utility of the Cyt-b in insects, where they had included 4 lepidopteran species; Wasp moth or *Hyalaceria gigantean* (Arctiidae), Ghost moth or *Hepialus dongyuensis* (Hepialidae), and two butterfly species *Aricia artaxerxes* (Lycaenidae), and *Euphydryas aurinia* (Nymphalidae) (Simmons and Weller, 2001a). In our present work, we have used this gene as our marker, to evaluate the potentiality of it as barcoding marker in *A. mylitta* Drury, the Tropical Tasar Silkworm

species of India.

*A. mylitta* mainly resides in forests which are filled with its host plants such as *Shorea robusta* (Sal), *Terminalia arjuna* (Arjun), *Terminalia tomentosa* (Asan) etc. Many ecologists believe that the continuous anthropogenic activity and habitat fragmentation has led to the isolation of the main population into many small population (now known as ecoraces) which originally may have had a single long stretch of population. The most interesting fact about these ecoraces is there is no specific boundary between these ecorace population and they don't obey the concept of static boundary, for which the identification of the ecoraces have always been a difficult task among the ecologists. To overcome this problem, in our current work we have used Cyt-b gene as a DNA barcoding marker for five ecoraces namely Modal, Nalia, Jata-Daba, Sukinda and Boudh of Odisha, to characterize them and identify them as to are they really ecoraces?

## 2. Materials and methods

### 2.1. Collection of specimen

Around 50 healthy cocoons from each of the 5 ecoraces were collected from their respective ecozones (Ref: Table 1). The cocoons were carefully cut and pupae were taken out to and were subjected to DNA extraction.

### 2.2. DNA extraction

In our current study, we have taken Cytochrome b or Cyt-b gene as our DNA molecular maker for analyzing the intra-specific and inter-specific relations between these ecoraces of *A. mylitta* and other lepidopterans. For genetic analysis of the individual ecoraces of *A. mylitta* Drury, mitochondrial DNA from individual ecoraces was isolated. The DNA was eluted with nuclease free water. The extracted mtDNA was then subjected to electrophoresis system and was photographed. The concentration of each DNA sample was measured by Nanodrop (Biotech instruments, USA). The purified DNA was then stored at −80 °C for further use.

### 2.3. PCR amplification and sequencing

For the amplification of mitochondrial cytochrome *b* gene, universal Cyt-b amplification primer named Cytb_mcb 398 (TACCATGAGGA-CAAATATCATTCTG) and Cytb_mcb 869 (CCTCCTAGTTTGTTAGG-GATTGATCG) were used (Verma et al., 2006). The amplification was carried out in Veriti® 96 well thermal cycler, with 25 µl containing 10 pmol of each both forward and reverse primers, 2.5 mM of MgCl2, 200 mM of each of the four dNTPs (deoxy riboNucleotide Tri-Phosphates), 0.5 U of Taq polymerase enzyme, 1× of PCR Buffer (Invitrogen, Life Technologies, Brazil) and 50-100 ng of isolated genomic DNA. The template was denatured by heating at pre-denaturation of 95 °C for 5 min. This was followed by 39 cycles of denaturation 30 s at 95 °C, 45 s annealing at 55 °C and 1 min elongation at 72 °C, with a final extension of 7 min at 72 °C and a single discrete PCR amplicon band of ~450 bp was observed. The PCR amplicon was bead purified and further subjected to Sanger Sequencing. Bi-directional DNA sequencing reaction of PCR amplicon was carried out with cytb_mcb 398& cytb_mcb 869 primers using BigDye™ Terminator v3.1Cycle sequencing kit on Biosystem Analyzer 3500Dx Genetic Analyzer.

### 2.4. Sequence editing and sequence data analysis and phylogenetic studies

The obtained Cyt-b sequences were edited and validated using Mega 11 software (Tamura et al., 2021). The assembly of the sequences were done by Clustal X2 (Larkin et al., 2007). To detect the similarity between the sequences, and the inter-generic similarity between ecoraces and other species, BLAST search was performed (Zhang et al., 2000). After

the proper alignment, the ends of the sequences were trimmed and submitted to GenBank. Analysis of the obtained sequences like nucleotide frequencies, nucleotide pair sequences, pairwise distance, Transition/Transversion ratio, Overall Transition/Transversion Bias (R), and Nucleotide substitution per site was calculated using Maximum Composite Likelihood parameter in MEGA 11 software. The intraspecific nucleotide distance between the five ecoraces was also calculated by the same MEGA 11 software. To understand the intra-specific phylogenetic relationship between the ecoraces of *A. mylitta* Drury, and the interspecific phylogenetic relationship between the ecoraces of *A. mylitta* Drury and other lepidopteran species, we constructed 6 phylogenetic trees based on the mitochondrial Cyt-b sequences obtained from the sequencing. These trees were generated using Neighbour-Joining Method (Saitou and Nei, 1987a), Maximum Likelihood criterion in MEGA 11 and Bayesian method (Berli, 2006; Noscimento et al., 2017). The Neighbour-Joining tree was generated using MEGA11 with Kimura 2 Parameter molecular evolutionary model. Statistical support for the nodes and the internodes were calculated using bootstrap analysis with 1000 replicates. Maximum likelihood phylogeny analysis was carried out in MEGA 11 adjusted to 1000 bootstrap 4 replicates under Tamura-3-Parameter. The strongness of the clade was assessed using bootstrap analysis. The parsimony tree was also created using the same MEGA 11 software (Tamura et al., 2021).

## 3. Result

After the PCR amplification, a ~ 450 bp amplicon of Cyt b gene was amplified. This Cyt-b was tested for an appropriate DNA Barcoding marker for *A. mylitta* Drury ecoraces identification. The obtained sequences of all the five ecoraces have been submitted to the GenBank (Table 2). The partial Cyt-b gene obtained from the PCR has high A + T content with an average T percentage ranging up-to 42.6 %. The average A, T, C and G frequencies in the Cyt-b gene sequences have been estimated to be 31.6, 42.6, 16.2 and 9.6 respectively. The details of nucleotide frequency (in percentage) of ecoraces have been mentioned in Table 3. The alignment of the obtained sequences was done by MEGA 11 software. The result reported that the consensus sequence between the ecoraces have been found to be more as compared to reserved sequences. Out of average ~ 459 bp, 413 bp are conserved, 46 bp are variable, and 6are singleton sites (Refer Fig. 1). Likewise, parsimony informative sites have been found to be 40. The codon usage analysis showed that the five ecoraces sequences have total average of 151 codon, and highest 18.4 % of it codes for Phenylalanine. Disparity Index per site for all sequence pairs is shown in Table 4. Values greater than 0 indicate the larger differences in base composition biases than expected based on evolutionary divergence between sequences and by chance alone. This analysis involved 5 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. There was a total of 460 positions in the final dataset. The difference in base composition bias per site between sequence pairs (Kumar and Gadagkar, 2001) is shown in Table 5. This analysis involved 5 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. The maximum composite likelihood for Estimation of the Pattern of Nucleotide Substitution was calculated with complete deletion option using TamuraNei

**Table 2**
GenBank accession number of the submitted ecoraces.

| Sl no. | Ecorace | GenBank accession number |
|---|---|---|
| 1 | Modal | BankIt2693962 A_mylitta_modal OQ818307 |
| 2 | Nalia | BankIt2695088 Antheraea_mylitta_nalia_ecorace OQ829357 |
| 3 | Jata-Daba | BankIt2695716 Antheraea_mylitta_JataDaba_ecorace OQ850265 |
| 4 | Sukinda | BankIt2695774 Antheraea_mylitta_Sukinda_ecorace OQ850266 |
| 5 | Boudh | BankIt2695897 Antheraea_mylitta_Boudh_ecorace OQ850267 |

**Table 3**
Nucleotide frequency of Cyt B gene in ecoraces of *A. mylitta* Drury.

| Ecorace | T | C | A | G | bp Number |
|---|---|---|---|---|---|
| Modal | 40.8 % | 16.9 % | 32.5 % | 9.9 % | 456 |
| Nalia | 42.0 % | 16.9 % | 31.2 % | 9.9 % | 455 |
| Jata-Daba | 43.4 % | 15.7 % | 31.6 % | 9.4 % | 459 |
| Sukinda | 43.4 % | 15.7 % | 31.6 % | 9.4 % | 459 |
| Boudh | 43.6 % | 15.7 % | 31.4 % | 9.4 % | 459 |
| Average | 42.6 % | 16.2 % | 31.6 % | 9.6 % | 457.6 |

model in MEGA 11 software (Tamura et al., 2004). For simplicity in calculation, the sum of r values is made equal to 100. The rate of interchange between purine and pyrimidines or the translational substitute rate is more when compared with rate of interchange between purine and pyrimidine i.e. overall transversional substitutes in Cyt b gene. The overall transition/transversion bias has been found to be $R = 1.6$ %. The transition/transversion rate ratios have been found to be for purines $k1 = 0.025$ and for pyrimidines $k2 = 0.169$. Maximum composite likelihood estimate of the pattern of nucleotide substitution was done by same MEGA11 software and the results obtained from that is given in Table 6. The evolutionary history was inferred using the Neighbour-Joining method, both Inter and Intra specifically (Saitou and Nei, 1987b). The bootstrap consensus tree inferred from 1000 replicates is taken to represent the evolutionary history of the taxa were analysed (Felsenstein, 1985). Branches corresponding to partitions reproduced in less than 50 % bootstrap replicates are collapsed. The percentage of replicate trees in which the associated taxa clustered together in the bootstrap test with 1000 replicates are shown next to the branches. The evolutionary distances were computed using the Kimura 2-parameter method (Kimura, 1980) and are in the units of the number of base substitutions per site. This analysis involved 5 nucleotide sequences. All positions containing gaps and missing data were eliminated with complete deletion option. The intraspecific tree constructed on Cyt-b gene revealed that Modal and Nalia ecoraces of *A. mylitta* Drury have highest similarities between them which means the genetic flow between these two ecoraces is highest. Similarly, another cluster of the tree revealed that Jata-Daba ecorace and Sukinda ecorace share high genetic similarities between them as compared to Boudh ecorace. The intra-generic phylogenetic tree obtained from the Neighbour-Joining Method revealed that, the *A. mylitta* ecoraces share evolutionary phylogeny with its sister species from USA, the *Antheraea polyphempous* moth and from China, the Chinese Oak Tasar the *Antheraea pernyi* (Figs. 2–4).

## 4. Discussion

The intraspecific phylogenetic analysis of these five ecoraces showed that Nalia, Modal and Jata-Daba have more gene flow between them as compared to Boudh and Sukinda ecorace. Boudh and Sukinda ecorace Cyt b sequence showed less variance in between their sequence. Molecular phylogenetic trees made by N-J method and Maximum-Likelihood method also showed similar kind of clustering among the ecorace sequences. In the N-J method tree Nalia and Modal ecorace, and Jata-Daba and Sukinda ecorace showed highest similarity between them, whereas the origin of Boudh is believed to be similar and of the same origin to the other two cluster. This origin and the gene flow can be explained by positive correlation between the genetic distance of the ecoraces and the geographical distance between them. The ecoraces that have ecozones nearer to each other, seems to have higher genetic flow between them and hence more genetic similarity, whereas the ecozones distant from each other tend to have less genetic outflow between them. If we look at the geographic availability of the ecoraces Modal and Nalia are available in the different regions of Similpal Biosphere Reserve, hence the chances of gene flow are more as compared to Boudh ecorace ecozones which is in Sunabeda Forest Range of Koraput, a place farther away from Similpal. Jata-Daba ecorace being native to Similapal nearby

```
#A_mylitta_modal              GGA CAA ATA TCA TTC TGA GGG GCT ACT GTA ATT ACA AAT CTT CTT TCA GCA ATT CCA TAT TTA GGT ACT ATA TTA GTA
#Antheraea_mylitta_Boudh_ecorace     GGA CAA ATA TCA TTT TGA GGA GCT ACT GTT ATT ACT AAT CTT CTT TCA GCA ATC CCT TAT TTA GGA ACT ATA TTA GTA
#Antheraea_mylitta_JataDaba_ecorace  GGA CAA ATA TCA TTT TGA GGA GCT ACT GTT ATT ACT AAT CTT CTT TCA GCA ATC CCT TAT TTA GGA ACT ATA TTA GTA
#Antheraea_mylitta_nalia_ecorace     GGA CAA ATA TCA TTC TGA GGG GCT ACT GTA ATT ACT AAT CTT CTT TCA GCA ATT CCA TAT TTA GGT ACT ATA TTA GTT
#Antheraea_mylitta_Sukinda_ecorace   GGA CAA ATA TCA TTT TGA GGA GCT ACT GTT ATT ACT AAT CTT CTT TCA GCA ATC CCT TAT TTA GGA ACT ATA TTA GTA

#A_mylitta_modal              AAC TGA ATT TGA GGG GGA TTT GCT GTA GAT TAA TGC TAC TTA AAC TCG ATT TTA TAC ATT TCA CTT -CT ATT ACC TTT
#Antheraea_mylitta_Boudh_ecorace     AAC TGA ATT TGA GGG GGA TTT GCT GTA GAT -AA TGC TAC CTT AAC TCG ATT TTA TAC ATT TCA TTT TTT ACT TCC TTT
#Antheraea_mylitta_JataDaba_ecorace  AAC TGA ATT TGA GGG GGA TTT GCT GTA GAT -AA TGC TAC CTT AAC TCG ATT TTA TAC ATT TCA TTT TTT ACT TCC TTT
#Antheraea_mylitta_nalia_ecorace     AAC TGA ATT TGA GGG GGA TTT GCT GTA GAT -AA TGC TAC TCT AAC TCG ATT TTA TAC ATT TCA CTT TCT ATT ACC TTT
#Antheraea_mylitta_Sukinda_ecorace   AAC TGA ATT TGA GGG GGA TTT GCT GTA GAT -AA TGC TAC CTT AAC TCG ATT TTA TAC ATT TCA TTT TTT ACT TCC TTT

#A_mylitta_modal              -AT TAT TTT AAT ATT AAC TAT AAT TCA TCT TCT TTT TCT TCA TCA AAC TGG ATC AAA TAA CCC TTT AGG ATT AAA TAG
#Antheraea_mylitta_Boudh_ecorace     TAT TAT TTT AAT ATT AAC TAT AAT TCA CCT TCT TTT TTT ACA TCA AAC TGG TTC AAA TAA CCC TTT AGG TTT AAA TAG
#Antheraea_mylitta_JataDaba_ecorace  TAT TAT TTT AAT ATT AAC TAT AAT TCA CCT TCT TTT TTT ACA TCA AAC AGG TTC AAA TAA CCC TTT AGG TTT AAA TAG
#Antheraea_mylitta_nalia_ecorace     TAT TAT TTT AAT ATT AAC TAT AAT TCA TCT TCT TTT TCT TCA TCA AAC TGG ATC AAA TAA CCC TTT AGG ATT AAA TAG
#Antheraea_mylitta_Sukinda_ecorace   TAT TAT TTT AAT ATT AAC TAT AAT TCA CCT TCT TTT TTT ACA TCA AAC AGG TTC AAA TAA CCC TTT AGG TTT AAA TAG

#A_mylitta_modal              AAA TAT TGA TAA AAT TCC TTT CCA TCC TTA TTT TAC ATA TAA GGA CTT AAT TGG ATT TAT TAT TTT AAT TAC ACT TTT
#Antheraea_mylitta_Boudh_ecorace     AAA TAT AGA TAA AAT TCC CTT TCA CCC TTA TTT TAC ATA TAA AGA TTT AAT TGG ATT TAT TAT TTT AAT TAT ATT TTT
#Antheraea_mylitta_JataDaba_ecorace  AAA TAT AGA TAA AAT TCC CTT TCA CCC TTA TTT TAC ATA TAA AGA TTT AAT TGG ATT TAT TAT TTT AAT TAT ATT TTT
#Antheraea_mylitta_nalia_ecorace     AAA TAT TGA TAA AAT TCC TTT CCA TCC TTA TTT TAC ATA TAA GGA CTT AAT TGG ATT TAT TAT TTT AAT TAC ACT TTT
#Antheraea_mylitta_Sukinda_ecorace   AAA TAT AGA TAA AAT TCC CTT TCA CCC TTA TTT TAC ATA TAA AGA TTT AAT TGG ATT TAT TAT TTT AAT TAT ATT TTT

#A_mylitta_modal              ATT ATT ATT AAC ATT AAT TAA CCC CTA TTT ATT AGG AGA CC- TGA TAA C-T TAT TCC TGC AAA TCC TTT AGT TAC ACC
#Antheraea_mylitta_Boudh_ecorace     ATC TTT ATT AAC TTT AAT TAA CCC TTA TTT ATT AGG AGA CCC AGA TAA TTT TAT TCC TGC TAA TCC TTT AGT AAC TCC
#Antheraea_mylitta_JataDaba_ecorace  ATC TTT ATT AAC TTT AAT TAA TCC TTA TTT ATT AGG AGA CCC AGA TAA TTT TAT TCC TGC TAA TCC TTT AGT AAC TCC
#Antheraea_mylitta_nalia_ecorace     ATT ATT ATT AAC ATT AAT TAA CCC CTA TTT ATT AGG AGA CCC TGA TAA CMT TAT TCC TGC TAA TCC TTT AGT TAC ACC
#Antheraea_mylitta_Sukinda_ecorace   ATC TTT ATT AAC TTT AAT TAA TCC TTA TTT ATT AGG AGA CCC AGA TAA TTT TAT TCC TGC TAA TCC TTT AGT AAC TCC

#A_mylitta_modal              TAT TCA TAT TCA ACC AGA ATG ATA TTT TCT ATT TGC TTA TGC TAT TTT ACG ATC AAT CCC AAA CAA ACT A
#Antheraea_mylitta_Boudh_ecorace     CAT TCA TAT TCA ACC AGA ATG ATA TTT TTT ATT TGC ATA TGC TAT TTT ACG ATC AAT CCC TAA TAA ATT A
#Antheraea_mylitta_JataDaba_ecorace  CAT TCA TAT TCA ACC AGA ATG ATA TTT TTT ATT TGC ATA TGC TAT TTT ACG ATC AAT CCC TAA TAA ATT A
#Antheraea_mylitta_nalia_ecorace     TAT TCA TAT TCA ACC AGA ATG ATA TTT TCT ATT TGC TTA TGC TAT TTT ARG ATR AWT CCC TAA CAA ACT A
#Antheraea_mylitta_Sukinda_ecorace   CAT TCA TAT TCA ACC AGA ATG ATA TTT TTT ATT TGC ATA TGC TAT TTT ACG ATC AAT CCC TAA TAA ATT A
```

**Fig. 1.** Multiple alignment of the amplified partial Cyt-b gene of ecoraces of *A. mylitta* of Odisha.

**Table 4**
Estimation of net base composition bias disparity between sequences.

|            | Nalia  | Modal  | Jata-Daba | Boudh  | Sukinda |
|------------|--------|--------|-----------|--------|---------|
| Nalia      |        |        |           |        |         |
| Modal      | 0.0160 |        |           |        |         |
| Jata-Daba  | 0.3720 | 0.1440 |           |        |         |
| Boudh      | 0.2180 | 0.0700 | 1.0000    |        |         |
| Sukinda    | 0.2920 | 0.1300 | 1.0000    | 1.0000 |         |

**Table 5**
Estimation of base composition bias difference between sequences.

|            | Nalia   | Modal   | Jata-Daba | Boudh   | Sukinda |
|------------|---------|---------|-----------|---------|---------|
| Nalia      |         |         |           |         |         |
| Modal      | 0.04555 |         |           |         |         |
| Jata-Daba  | 0.09698 | 0.17559 |           |         |         |
| Boudh      | 0.12967 | 0.21739 | 0.00216   |         |         |
| Sukinda    | 0.11422 | 0.18124 | 0.00212   | 0.00215 |         |

**Table 6**
Maximum composite likelihood estimate of the pattern of nucleotide substitution.

|   | A     | T     | C    | G    |
|---|-------|-------|------|------|
| A | –     | 20.18 | 7.64 | 0.11 |
| T | 15.04 | –     | 1.29 | 4.56 |
| C | 15.04 | 3.4   | –    | 4.56 |
| G | 0.37  | 20.18 | 7.64 | –    |



**Fig. 2.** Phylogenetic tree made with Boot strap consensus.

regions like Thakurmunda, but showing highest gene-flow with Sukinda which is native to Sukindagarh of Jajpur district, the distance between the ecozones being nearly 100 km. The possible explanations for this dilemma are that the two ecoraces aren't true ecoraces, rather just man-made hybrids. The inter-specific analysis based on the phylogenetic tree
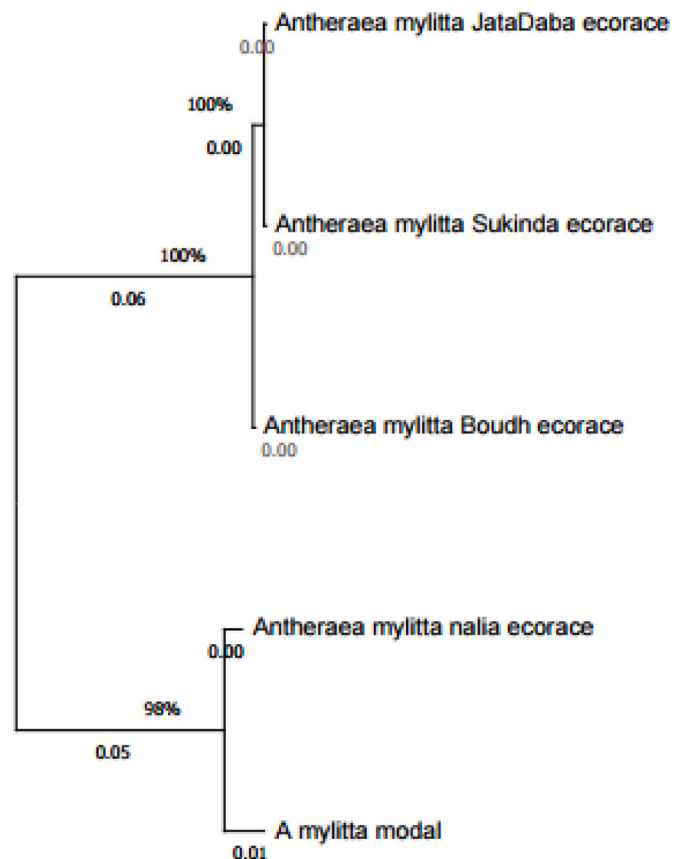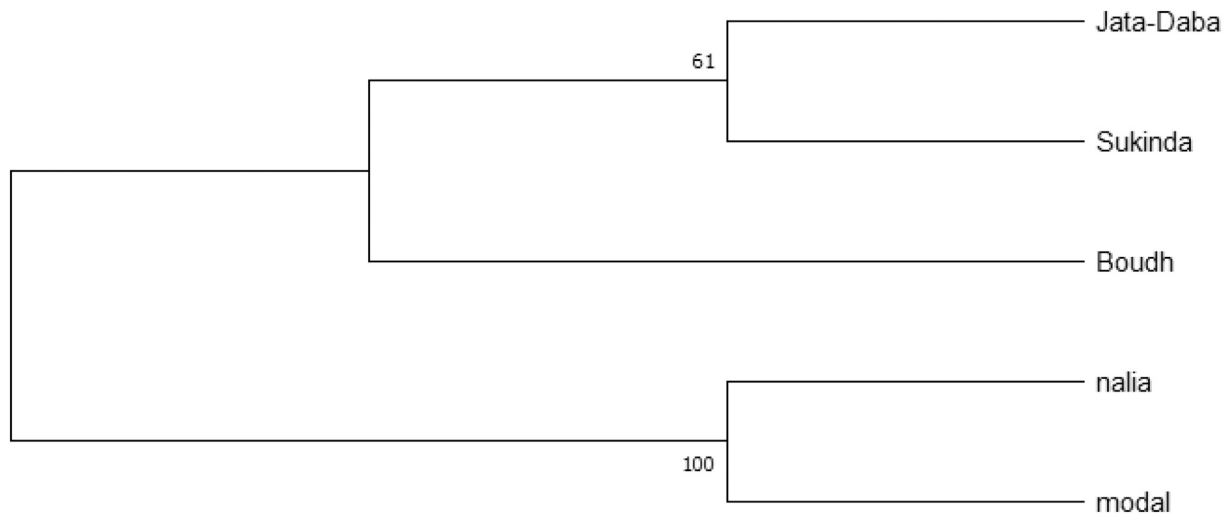
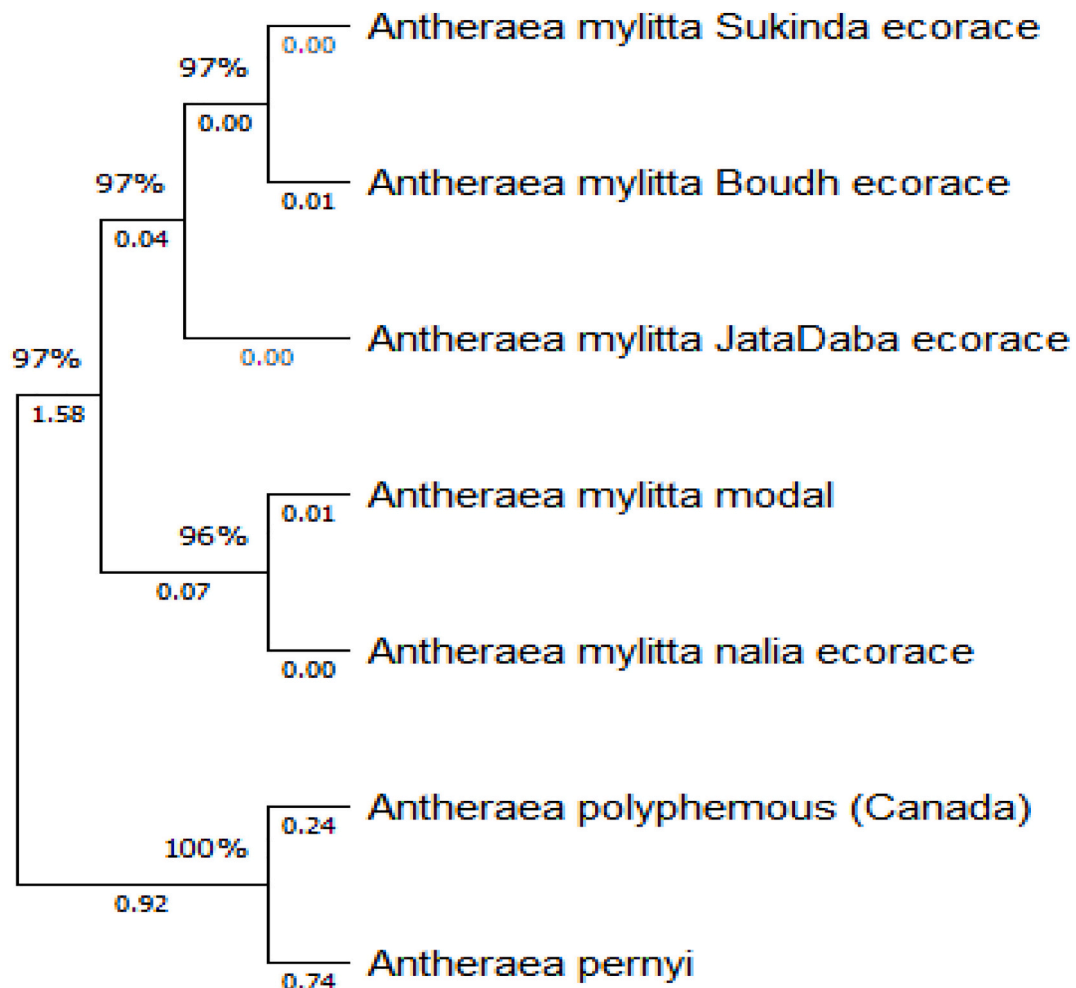**Fig. 3.** Intraspecific Neighbour-Joining tree.

**Fig. 4.** Interspecific Phylogenetic tree using Neighbour joining method.

obtained from the MEGA 11 shows that the Modal and Nalia ecorace cluster having out-group of *Antheraea pernyi*, the Chinese oak tasar, and for Sukinda ecorace, *Antheraea polyphemous*, for Boudh ecorace the *Saturnia pavonia*. All the out-groups in their natural habitat prefer tropical and sub-tropical environment, which has almost the same climatic range as the ecozones of *A. mylitta*. *A. polyphemous* is mostly found in Mexico and Florida of southern parts of USA, where the primary habitation is tropical or Sub-tropical humid, similarly *Saturnia pavonia*, even though is mainly distributed in Palaearctic region, but the main population is more prevalent in the warm temperate regions. The type of

phylogeny we have obtained from the analysis where the many outgroups are embedded in the ecorace clusters, and showing similar ancestry with the ecoraces might can be explained better by climatic and host plant adaptations. Most of the *Antheraea* species are polyphagous in nature and they thrive on a wide range of host plants depending on the availability of the vegetation, hence the characters and genotypic variability between the same species might can vary depending on the host plant nutrition available. Since the ecoraces of *A. mylitta* can thrive on different host plants such as Sal (*Shorea robusta*), Arjun (*Terminalia arjuna*), Asan (*Terminalia tomentosa*), Oak (*Quercus*), etc. hence the ecorace characteristics are different from each other. Similarly, the outgroup species we have found *A. polyphemous*, and *A. pernyi* feed on wide range of host plants such as Oak (*Quercus*), Maple (*Acer*), Birch (*Betula*), Hickory (*Caryu*), Willow (*Salix*). The tropical and humid subtropical climate of these regions along with very little Snow blizzard makes the geographical location similar to that of Tropical moist deciduous forest ecozones of our *A. mylitta*. Safran and Nosil in 2012 have proven that the populations pairs (might be of same species or not) feeding on the same host plant, in different geographic locations, are ecologically similar and assumed to not be subject to divergent selection, whereas in contrast, pairs of populations (might be of same species or not) feeding on different host plant species are ecologically divergent and subject to divergent selection (Safran and Nosil, 2012). Hence the similarities in the climatic conditions and similar host plants might be the reason why Modal, Nalia ecoraces showing high genetic similarities with *A. pernyi*, Sukinda with *A. polyphemous* and Boudh with *S. pavonia,* forming three different clusters rather than clustering all the ecoraces together in one cluster, and outgroups in different clusters. In the past, assessing polymorphism among ecoraces required using SSR (Simple Sequence Repeats) (Renuka and Shamitha, 2016), RAPD-SCAR (Random Amplification of Polymorphic DNA-Sequence Characterized Amplified Region) (Pranhee et al., 2023), ISSR (Inter Simple Sequence Repeat) (Chatterjeee et al., 2004), RAPD (Random Amplification of Polymorphic DNA) Markers (Saha et al., 2008), Repetitive Taq1 genomic marker (Mahendran et al., 2006a), RFLP (Restriction Fragment Length Polymorphism) (Mahendran et al., 2006b) and similar techniques, necessitating over a dozen custom markers to amplify genes for phylogenetic analysis. However, in our work, only two readily accessible universal primers were needed to amplify the Cyt b gene. This streamlined approach not only addressed manpower and financial constraints but also expedited the workflow. This technique not only achieves a higher success rate in identifying and analyzing ecoraces compared to previous methods but also offers a clearer understanding of the phylogeny of *Antheraea mylitta* and sibling species, shedding light on evolutionary and ecological perspectives.

## 5. Conclusion

Our study is the first study to compare the relative genetic distance and relation among the ecoraces of *A. mylitta* Drury in Odisha. In many of the studies carried out by researchers have showed that Cyt b and COI gene shows similar kind of results in the identification process (Tobe et al., 2010). Currently in our study we have used Cyt b gene to test it as a potential barcoding marker for *A. mylitta* Drury and use it ecoraces identification for this species. Previously many Marker such as ISSR Markers, Repetitive Taq1 genomic marker, RAPD Markers, Microsatellite markers (Chakraborty et al., 2015) and SSR Markers had been used for ecorace identification, but using mitochondrial gene Cyt b as ecorace identification is the first attempt by us. RAPD, RFLP, and ISSR markers are not only expensive to employ, but their development is also time-consuming and labour-intensive, with lower success rates, especially in genera where polymorphism is scarce. The primers employed for Cyt b and COI DNA Barcoding are not only universal but have also been endorsed by iBOL (International Barcode of Life) with established rates of genetic divergence at both intraspecific and interspecific levels across different classes of organisms. This enables swift identification and

phylogenetic analysis, reducing the overall cost of the procedure compared to alternative molecular methods. Furthermore, standardising a particular marker for identification of the ecoraces will make the sericulturists and researchers easy to access the true breed ecoraces faster.

What we have found in our study is similar to that of Simmons and Weller (2001a, 2001b) where they had found that A/T bias (A = 41 %, *T* = 45 %) is more in Insect Cyt b gene (Simmons and Weller, 2001b), and in our sample A and T on average have been found to be 31.86 % and 42.61 %. The phylogenetic relationship between the ecoraces have shown that Modal and Nalia show highest similarity which means the gene flow between the ecoarces are more since they belong to nearby regions in the Similapal Biosphere reserve. Whereas Boudh and Sukinda shared less genetic similarity between them as compared to Modal and Nalia, since the ecozones of these two ecoraces are farther away from each other. The interspecific phylogenies also have shed some light on the environmental adaptations by the out group sister species similar to that of ecoraces. By understanding these ecoraces and their ecozones we will be able to save the ecoraces which are on the verge of extinction, because of the overexploitation of certain hybrids or good yielding ecoraces like Sukinda and Jata-Daba. Not only that, we can also strengthen the ecorace line by interbreeding or inbreeding and save a lot of financial loss by the rearer community. This study also strengthens the belief of using Cyt b gene as a barcoding marker, which can be useful in identification of silk rearing lepidopterans and other related species.

## Ethics approval

Not Applicable.

## Consent to participate

All authors have given consent to participate for publication of their work.

## Consent for publication

All authors have given consent to participate for publication of their work.

## Funding

Not applicable.

## Availability of data and materials

Not applicable.

## Code availability

Not applicable.

## Animal ethics compliance

The experimental animal used in our experiment i.e. *A. mylitta* Drury, belongs to commercial invertebrate and does not come under the purview of Wildlife Protection Act -1972, India.

## CRediT authorship contribution statement

**Priti Pragyan Ray:** Writing – review & editing, Validation, Supervision, Project administration, Formal analysis, Conceptualization. **Barsha Barala:** Writing – original draft, Methodology, Investigation. **Purushottam Dash:** Resources.

## Declaration of competing interest

All the authors have no conflict of interest.

## Data availability

No data was used for the research described in the article.

## References

Baig MM, Singh G, Prabhu DIG, Manyappa, Sahay A, SAthyanarayan K (2023) Characterisation of tasar silkworm *Antheraea mylitta* Drury (Saturniidae: Lepidoptera) midgut bacterial symbionts through metagenomics analysis. International Journal of Tropical Insect Science 43:999–1011.

Ballard, J.W.O., Kreitman, M., 1995. Is mitochondrial DNA a strictly neutral marker? Trends in Ecology and Evolution, Cell Press 10 (12), 485–488.

Ballard, J.W.O., Rand, D.M., 2005. The population biology of mitochondrial DNA and its phylogenetic implications. Annual review Ecology and Evolution Systematics 36, 621–642.

Ballard, J.W.O., Thayer, M.K., Newton, A.F., Grismer, E.R., 1998. Data sets, partitions and characters: philosophies and procedures for analysing multiple data sets. Syst. Biol. 47, 427–436.

Barsagade, D.D., Kirsan, J., Chankapure, P.N., Gharde, S.A., Barsagade, V.G., Thakre, M.P., 2022. Immunohistochemical localisation of histamine neuropeptide in the brain neurosecretory cells of tasar silkworm *Antheraea mylitta* Drury. Journal of Entomology and zoology studies 10 (6), 229–233.

Berli, P., 2006. Comparison of Bayesian and Maximum Likelihood inference of population genetic parameters 22 (3), 341–345.

Boore, J.L., 1999. Animal Mitochondrial Genome. Nucleic Acid Research 27 (8), 1767–1780.

Caine, L., Lima, G., Pontes, L., Abrantes, D., Pereira, M.F., 2006. Species identification by cytochrome b gene: casework samples. International Congress Series 1228, 145–147.

Cameron, S.L., 2014. Insect mitochondrial genomics: implications for evolution and phylogeny. Annu. Rev. Entomol. 59, 95–117.

Chakraborty, S., Muthulakshmi, M., Vardhini, D., Jayaprakash, P., Nagaraju, J., Arunkumar, K.P., 2015. Genetic analysis of Indian Tasar Silkmoth (Antheraea mylitta) populations. Sci. Rep. 5, 15728.

Chatterjee, S.N., Vijayan, K., Roy, G.C., Nair, C.V., 2004. ISSR profiling of genetic variability in the ecotypes of Antheraea mylitta Drury, the tropical Tasar silkworm. Russ. J. Genet. 40 (2), 152–159.

De Pancorbo, M.M., Casbro, A., Fernandez-Fernandez, I., Cuevac, N., 2004. Cytochrome b for identification pf animal species in processed food. Int. Congr. Ser. 1261, 592–594.

Dutta, S., Mohapatra, J., Ghosh, A.K., 2020. Molecular characterisation of *Antheraea mylitta* Arylphorin gene and its encoded protein. Arch. Biochem. Biophys. 692, 108540.

Farag, M.R., El Bohi, K.M., Khalil, S.R., Alagawany, M., Arain, M.A., Sharun, K., Tiwari, R., Dhama, K., 2020. Forensic application of mitochondrial cytochrome b gene in the identification of domestic and wild animal species. Journal of Experimental Biology and Agricultural Sciences 8 (1), 1–8.

Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.

Gimeno, C., Belshaw, R., Quicke, D.L.J., 1997. Phylogenetic relationships of the Alysiinae/Opiinae (Hymenoptera:Braconidae) and utility of cytochrome b. Insect Mol. Biol. 6, 273–284.

Jena, K., Kumari, R., Pandey, J.P., Kar, P.K., Akhtar, J., Singh, A.K., Gupta, V.P., Sinha, A.K., 2021. Biochemical characterisation of sericin isolated from cocoons of tropical Tasar silkworm *Antheraea mylitta* raised on three different host plants for perspective utilisation. J. Asia Pac. Entomol. 24 (3), 903–911.

Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16 (2), 11–120.

Kitching, I.J., Rougerie, R., Zwick, A., Hamilton, C.A., St Laurent, R.A., Naumann, S., Mejia, L.B., Kawahara, A.Y., 2018. A global checklist of the Bombycoidea (Insecta: Lepidoptera). Biodiversity Data Journal 6, e22236.

Kumar, S., Gadagkar, S.R., 2001. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics 158, 1321–1327.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., Higgins, D.G., 2007. Clustal W and Clustal X version 2.0. Bioinformatics 23 (21), 2947–2948.

Mahendran, B., Acharya, C., Dash, R., Ghosh, S.K., Kundu, S.C., 2006a. Repetitive DNA in tropical tasar silkworm *Antheraea mylitta*. Gene 370, 51–57.

Mahendran, B., Padhi, B.K., Ghosh, S.K., Kundu, S.C., 2006b. Genetic variation in ecoraces of tropical tasar silkworm, *Antheraea mylitta* D. Using RFLP technique. Curr. Sci. 90 (1), 100–103.

Manda, S., Gangupantula, S., Gattu, R., 2019. Genomic DNA isolation and quantification in various ecoraces of Tasar silkmoths *Antheraea mylitta* D. International Journal of Engineering Applied Sciences and Technology 4 (4), 295–300.

Noscimento, F.F., dos Reis, M., Yang, Z., 2017. A Biologist's guide to Bayesian Phylogentics analysis. Nature Ecology and Evolution 1, 1446–1454.

Parson, W., Pegoraro, K., Niederstatter, H., Foger, M., Steinlechner, M., 2000. Species identification by means of the cytochrome b gene. Int. J. Leg. Med. 114 (1–2), 8–24.

Prabhu, I.G., Manappa, Baig M.M., Kumar, N., Sinha, A.K., Kutala, S., 2023. Molecular cloning and development of RAPD-SCAR markers for the selection of thermos-tolerant line of tropical tasar silkworm. J. Environ. Biol. 44 (3).

Pranhee, I.G., Manjappa, Baig M.M., Kumar, N., Sinha, A.K., Kutala, S., 2023. Molecular cloning and development of RAPD-SCAR markers for the selection of thermos-tolerant line of tropical tasar silkworm. J. Environ. Biol. 44 (3).

Rani, A., Pandey, D.M., Pandey, J.P., 2024. Biomolecular characterisation of *Antheraea mylitta* cocoonase: a secreted protease. Anal. Biochem. 686.

Reddy, B.T., Baig, M.M., Bawaskar, D.M., Gadad, H., 2020. Avian pests of Tasar silkworm *Antheraea mylitta* (Drury) (Saturniidae:Lepidoptera) from India. Journal of Experimental Zoology India 23 (2), 1079–1084.

Renuka, G., Shamitha, G., 2016. Genetic variation in ecoraces of tropical tasar *silkworm Antheraea mylitta*, using SSR markers. J. Genet. 95 (4), 777–785.

Renuka, G., Shamitha, V., 2024. Next Generation Sequencing (NGS) in Populations of Tropical Tasar Silkworm *Antheraea Mylitta*. Authorea.

Safran, R.J., Nosil, P., 2012. Speciation: the origin of new species. Nature Education Knowledge 3 (10), 17.

Saha, M., Mahendran, B., Kindu, S.C., 2008. Development of random amplified polymorphic DNA markers for tropical Tasar silkworm Antheraea mylitta. J. Econ. Entomol. 101 (4), 1176–1182.

Saitou, N., Nei, M., 1987a. The neighbour joining method: a new method for Recinstructing phylogenetic trees. Mol. Biol. Evol. 4 (4), 406–425.

Saitou, N., Nei, M., 1987b. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.

Simmons, R.B., Weller, S.J., 2001a. Utility and evolution of cytochrome b in insects. Mol. Phylogenet. Evol. 20 (2), 196–210.

Simmons, R.B., Weller, S.J., 2001b. Utility and evolution of cytochrome b in insects. Mol. Phylogenet. Evol. 20 (2), 196–210.

Singh, B.M.K., Srivastava, A.K., 1997. Ecoraces of Antheraea mylitta Drury and exploitation strategy through hybridization. CTR&TI, current technology seminar in non-mulberry sericulture. Base Paper. 6, 1–39.

Sneha, S., Pandey, D.M., 2022. In Silico structural and functional characterisation of Antheraea mylitta Cocoonase. 20 (102).

Soundappan, M., Kumar, D., Bhagavanulu, M.V.K., Chandrashekharaiah, M., Rathore, M.S., Srinivas, C., 2021. Wing and leg amputation induced oviposition patteren of tropical tasar silk moth *Antheraea mylitta* Drury (Lepidoptera:Saturnidae) in oviposition devives- earthen cups and paper sheets. J. Asia Pac. Entomol. 24 (3).

Srivastava, A.K., Sinha, A.K., 2002. Present Status of Tropical Silkworm Germplasm Management. Workshop on Germplasm Management and Utilisation at CSGRC, Hosur, Base Paper, pp. 1–12.

Tamura, K., Nei, M., Kumar, S., 2004. Prospects for inferring very large phylogenies by using the neighbor-joining method. Proceedings of the National Academy of Sciences (USA) 101, 11030–11035.

Tamura, K., Stecher, G., Kumar, S., 2021. MEGA11: molecular evolutionary genetics analysis version 11. Mol. Biol. Evol. 38 (7), 3022–3027.

Thirupam, Reddy B., Chandrashekharaiah, M., Mozumdar, S.M., Sathyanaryana, K., Chowdary, N.B., 2023. Host selection behaviour of tasar silkworm *Antheraea mylitta* (Drury) for oviposition. J. Environ. Biol. 44, 452–457.

Tobe, S.S., Kitchner, A.C., Lincare, A.M.T., 2010. Reconstructing mammalian phylogenies: a detailed comparison of the cytochrome b and cytochrome oxidase subunit I mitochondrial genes. PlosOne 5 (11), e14156.

Sunil Kumar Verma, Lalji Singh; Council of Scientific and Industrial Research. Universal Primers for Wildlife Identification. United States Patent. Patent Number: US 7,141,364 B1. Isuued Date: Nov 28th 2006.

Zhang, Z., Schwartz, S., Wagner, L., Miller, W., 2000. A greedy algorithm for aligning DNA sequences. J. Comput. Biol. 7 (1–2), 203–214.