

DNA Barcoding of 15 Species of African Saturniidae (Lepidoptera) across 12 Genera Reveals High Incidence of Non-Monophyly †

Zwannda Nethavhani *, Kayleigh Hiscock and Barbara van Asch

Department of Genetics, Stellenbosch University, Private Bag X1, Matieland 7602, South Africa; 20998880@sun.ac.za, kayleighanske@gmail.com, bva@sun.ac.za

* Correspondence: 20998880@sun.ac.za

† The 1st International Electronic Conference on Entomology, Sciforum, online, 1–15 July 2021

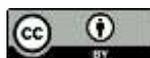
Abstract: African Saturniidae (Lepidoptera) been poorly described despite their nutritional and economic significance in Southern Africa. DNA barcodes for 15 species across 12 genera were generated as a starting point to create a reference sequence library for African Saturniidae. Phylogenetic clustering and genetic divergence estimates were used to assess patterns of genetic diversity in each genus, including all sequences available on BOLD Systems. Cases of polyphyletic and paraphyletic species affected 81% of the dataset. Taxonomic misidentifications, clerical and laboratory errors, introgression, hybridization and incomplete lineage sorting may contribute to explain these results. We propose that the dataset available for African Saturniidae does not correctly represent the genetic diversity of a high proportion of the taxa in the family.

Keywords: *COI*; genetic divergence; phylogenetics; species identification; Taxonomy

Citation: Nethavhani, Z.; Hiscock, K.; van Asch, B. DNA Barcoding of 15 Species of African Saturniidae (Lepidoptera) across 12 Genera Reveals High Incidence of Non-Monophyly, in Proceedings of the 1st International Electronic Conference on Entomology, 1–15 July 2021, MDPI: Basel, Switzerland, doi:10.3390/IECE-10573

Published: 2 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Submitted for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The family Saturniidae (Lepidoptera) contains 3,454 species distributed in 169 genera worldwide across diverse terrestrial ecosystems [1]. Saturniidae include a large number of Emperor moth species that play roles in human societies [2]. Some species are economically relevant as silk producers, e.g., *Samia cynthia* (Tribe: Attacini) and *Antheraea assamensis* (Tribe: Saturniini), distributed and farmed in India, Myanmar, Indonesia, and Sri Lanka [3].

In Africa, Saturniidae include 185 species in 46 genera and over 80 of these are edible at their larval stage [4–6]. Caterpillars are mostly consumed in southern, central, and western Africa where they provide an important source of protein and vitamins, as well as seasonal income to many rural communities [7]. *Bunaea alcinoe* and *Cirina forda* are widely distributed and consumed in sub-Saharan Africa, whereas *Gonimbrasia belina* and *Gynanisa maja* are more prevalent in southern and central Africa [7]. The increasing demand for favored species has become a threat to their sustainability due to overharvesting and habitat erosion.

Despite their importance, African Saturniidae have been poorly described, and genetic data that could assist in understanding patterns of biodiversity in this group is limited, except for DNA barcodes available on BOLD Systems (<https://v4.boldsystems.org/>) for a small proportion of species. Additionally, a recent study reported the complete mitochondrial genomes of two economically relevant species (*Gonimbrasia belina* and *Gynanisa maja*) and identified the most informative regions for use in future population studies [8]. Genetic information is a prerequisite for understanding the distribution of the diversity of populations and for managing the sustainable use of wild resources.

Baseline genetic data on African Saturniidae is currently scarce, and the taxonomic coverage across the rich diversity of species on the continent is incomplete.

DNA barcoding is a convenient tool for cataloguing biodiversity by sequencing a short standardized region of a well-defined gene [9]. The use of a fragment of the mitochondrial *cytochrome oxidase I (COI)* gene is widely applied as a species identification marker in insects due to its highly conserved nature, and low intraspecific and high interspecific variation. DNA barcoding also has the capability of linking the different life stages of an organism, thus allowing for species identification in the absence of well-described taxonomic characters, or incomplete specimens. DNA barcoding relies on the use of reference barcodes acquired from expertly identified specimens and deposited on publicly available online platforms (e.g. BOLD Systems, <http://www.boldsystems.org/>).

BOLD Systems can assist in assigning newly encountered specimens to their species [10–11], as the platform provides a species identification tool that retrieves an identification when the query sequence has less than 3.00% divergence from a reference sequence. However, the efficacy of BOLD Systems has recently been examined for having high incidences of non-monophyly presumed to result from taxonomic misidentifications across many animal phyla [12, 13]. Therefore, it is indispensable to combine expertise in classical taxonomy and DNA barcoding for building useful databases [10–13].

This study aimed to contribute to cataloging the biodiversity of African Saturniidae using DNA barcoding. The main objective was to sequence the standard *COI* barcoding region of 15 species found in South Africa, as a starting point for creating a highly curated library of DNA barcodes for the family.

2. Materials and Methods

2.1. Specimen collection and DNA extraction, PCR and Sequencing

Adult and caterpillars were collected from various locations in Namibia and South Africa from March 2018 to January 2021. All specimens were identified based on currently available taxonomic keys. Total DNA was extracted from adult and caterpillar legs using a standard phenol-chloroform protocol [15]. A total of 198 specimens was sequenced for the standard *COI* barcoding region (702 bp) using a variety of PCR primers (Table 1). PCR amplifications were performed using the QIAGEN Multiplex PCR Kit (QIAGEN), according to manufacturer’s protocol. Sanger sequencing was performed at the Central Analytical Facilities (CAF) of Stellenbosch University, South Africa.

Table 1. Primers used for the amplification of the standard *COI* barcoding region of 15 African Saturniidae species.

Primer	Sequence (5' - 3')
LCO1490	5' GGTCAACAAATCATAAAGATATTGG 3'
HCO2198*	5' TAAACTTCAGGGTGACCAAAAAATC 3'
Lep-F	5' ATTCAACCAATCATAAAGATATTGG 3'
Lep-R*	5' TAAACTTCTGGATGTCCAAAAAATCA 3'
Bar-Bel-F	5' ATTCTACTAATCATAAAGATATTGG 3'
Bar-Bel-R*	5' ACTTCTGGGTGTCCAAAAAATC 3'
Veg-F**	5' ATTCAACTAACCATATAAAGAAATTGG 3'

*Primer used for Sanger sequencing, **Primers designed in this study.

2.3. DNA analyses

Genetic clustering and genetic divergences were calculated using the new sequences generated in this study and DNA barcodes taxonomically assigned to the genera under study and available on BOLD Systems as of September 2020. The final dataset included a total of 1,868 sequences representing 191 species across 12 genera. Multiple sequence alignments were performed using the MAFFT algorithm [16] in Geneious Prime

v2021.1 (<https://www.geneious.com>). Genetic clustering of sequences was assessed using the Neighbour-joining (NJ) method in MEGA X [17], under the Kimura 2-parameter model (K2P) [18].

Genetic divergences were estimated as maximum pairwise distances (max p-distance, %) in MEGA X, under the K2P model. Intraspecific max p-distances were calculated for species (i.e., sequences were grouped according to species names), and intragroup max p-distances were calculated for groups of sequences according to the clusters recovered on the NJ tree, disregarding the names of the sequences.

3. Results and Discussion

This study investigated patterns of genetic diversity amongst 15 African Saturniidae species in 12 genera to assess the congruence between morphological and molecular identification, in the context of all data available on BOLD Systems for these genera. The new sequences represent only a small fraction (8%) of the Saturniidae species found on the continent. DNA barcodes were generated for 198 specimens for assessment of genetic clustering and estimates of genetic diversity. Intraspecific p-max distances and genetic clustering of sequences on the NJ trees showed that 83% (10/12) of the genera had cases of non-monophyletic species, representing 81% of the total sequence dataset (Table 2).

Table 2. Incidence of species-level non-monophyly (NM) across 12 genera of African Saturniidae. NM was calculated as the proportion of non-monophyletic species in each genus.

Genus	Number of sequences (number of species)	NM (%)
<i>Bunaea</i>	87 (3)	67%
<i>Eochroa</i>	12 (1)	0%
<i>Epiphora</i>	306 (39)	82%
<i>Gonimbrasia</i>	552 (57)	79%
<i>Gynanisa</i>	179 (21)	90%
<i>Heniocha</i>	104 (9)	56%
<i>Holocerina</i>	237 (17)	76%
<i>Ludia</i>	171 (21)	86%
<i>Nudaurelia</i>	26 (2)	0%
<i>Pselaphelia</i>	86 (10)	70%
<i>Usta</i>	76 (7)	100%
<i>Vegetia</i>	32 (4)	25%

3.1. Phylogenetic clustering and genetic divergence estimates

3.1.1. *Bunaea*

The NJ tree for the genus *Bunaea* recovered three main clusters, of which cluster 2 (2.58%) and cluster 3 (1.38%) were monophyletic. Cluster 1 had a high level of incongruity between species names and genetic groups (12 novel *B. alcinoe*, all *Bunaea alcinoe caffraria* and five *Bunaea alcinoe* sequences publicly available) with a max intraspecific p-distance of 1.70% despite different taxonomic designations (Figure 1a). However, the max intraspecific p-distance for *B. alcinoe* was 5.68%, which is above the range commonly accepted for conspecific individuals.

3.1.2. *Vegetia*

The clusters on the NJ tree for the genus *Vegetia* were largely congruent with species, except for the sequence of *Vegetia ducalis* 4. The intraspecific variation of *V. ducalis* as a species was 4.08%, suggestive of incorrect identification of *Vegetia ducalis* 4 (Figure

2b). Based on sequence similarity, *Vegetia ducalis* 4 could represent a different species; however, there were no other sequences with high similarity to confirm this hypothesis.

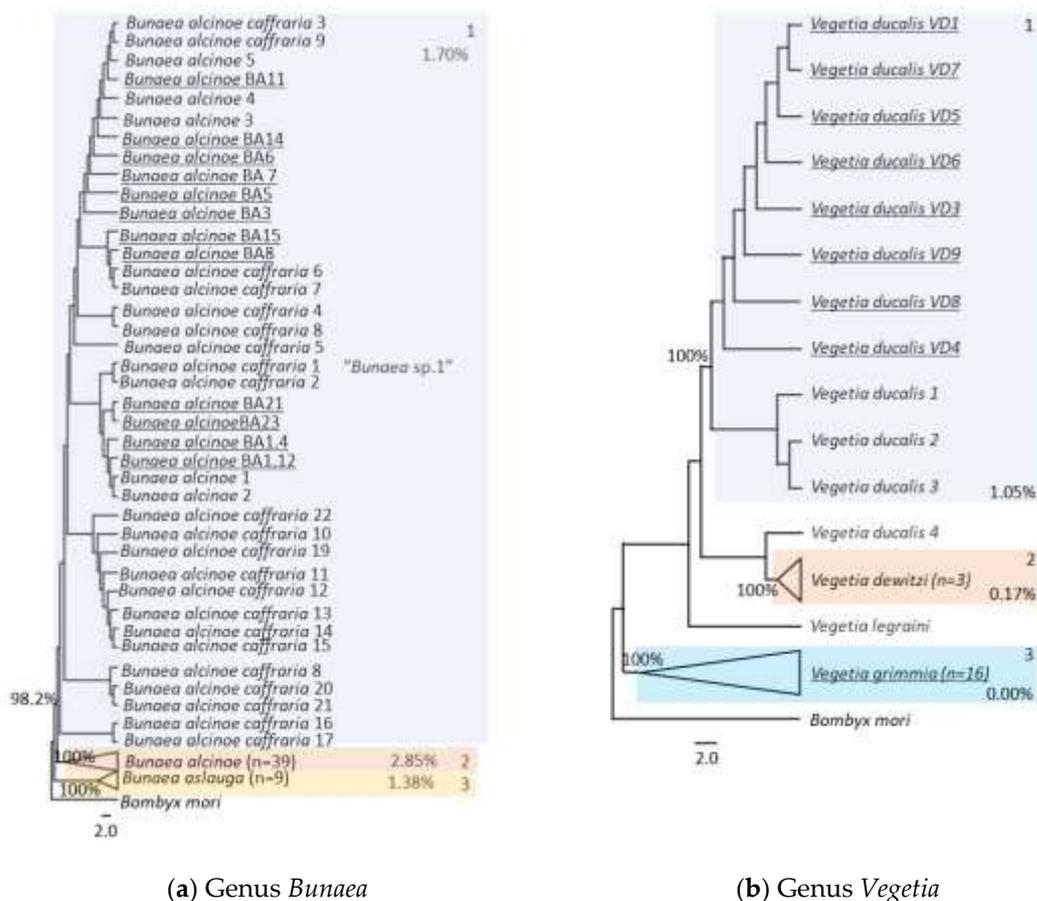


Figure 1. Neighbor-joining (K2P) trees of (a) 87 COI sequences of *Bunaea*, (b) 32 COI sequences (658 bp) of *Vegetia*. The numbers on the NJ trees represent genetic clusters (intraspecific maximum p-distances < 3%), and novel sequences are underlined.

3.1.3. *Eochroa*

Prior to this study, only two sequences were available on BOLD Systems for the genus *Eochroa*. A total of 10 new sequences were generated for *Eochroa trimenii*. Both novel and public sequences formed a cluster with intraspecific p-max = 2.58%, supporting the morphological identification of the new specimens.

3.1.4. *Epiphora*

The NJ tree of the genus *Epiphora* included 303 sequences in 39 species and recovered 21 clusters of which 12 were monophyletic. The new *E. bauhiniae* sequences clustered with mixed sequences from other seven species with an intraspecific distance of 2.50%, suggesting conspecificity. Groups of sequences with identical names had p-distances ranging between 3.42% - 7.97%, suggesting that some specimens may have been misidentified.

3.1.5. *Gonimbrasia*

The genetic clustering analysis of the genus *Gonimbrasia* included 539 sequences available on BOLD System distributed amongst 57 species and 13 new sequences for *Gonimbrasia tyrrhea*. A complex tree topology recovered with 17 monophyletic species, including *G. tyrrhea* (max p-distance = 0.15%) The remaining 40 species were distributed across non-monophyletic clusters.

3.1.6. *Gynanisa*

The genus *Gynanisa* had nine clusters and a high level of incongruity between the names of the sequences and the clusters on the NJ tree. Intraspecific distances of sequences from BOLD Systems designated as conspecific were high, and ranged between 3.20% and 5.90%. The new sequences of *G. maja* had an intraspecific p-distance of 3.04%, formed two distinct groups, clustering with *G. maja*, *G. nigra*, and *G. ata*.

3.1.7. *Holocerina*

The NJ tree for the genus *Holocerina* showed 15 clusters of which 10 were monophyletic with conspecific individuals and intraspecific distances below 3.00%. The new *H. smilax* sequence clustered with some of the other *H. smilax* with an intraspecific distance of 1.51% supporting conspecificity of specimens.

3.1.8. *Ludia*

The genus *Ludia* showed a total of 12 clusters with a high level of incongruity between species and genetic clusters. The new sequence for *L. delegorguei* clustered with sequences of *L. delegorguei* and *L. goniata* with intraspecific distance of 2.01%.

3.1.9. *Nudaurelia*

The NJ tree for *Nudaurelia* consisted entirely of novel sequences for *N. cytherea* and *N. wahlbergi*, and both species had low intraspecific max p-distances (0.00% and 0.47%, respectively) in congruency with the morphological identification of the specimens.

3.1.10. *Heniocha*

The NJ tree for the genus *Heniocha* showed seven clusters of which three were monophyletic. The intraspecific divergence among *H. dyops* from BOLD Systems and all novel sequences of *H. dyops* was 3.71%. When *H. dyops* was divided into two groups, the intragroup divergences were low: novel sequences of *H. dyops*– 1.76%, and BOLD Systems *H. dyops*– 2.87%, suggesting that the *H. dyops* dataset may represent two species or two diverged populations of the same species. The new *H. apollonia* sequence clustered with other sequences of *H. apollonia* with an intraspecific distance of 2.17% supporting conspecificity.

3.1.11. *Pselaphelia*

The NJ tree of the genus *Pselaphelia* recovered 10 clusters, of which seven clusters were monophyletic and comprised of conspecific individuals. The new *P. flavivitta* sequence clustered with other *P. flavivitta* with intraspecific distance of 1.51% supporting conspecificity.

3.1.12. *Usta*

The NJ tree for the genus *Usta* recovered 10 clusters, of which seven were monophyletic. The remaining three had a high level of incongruity between species names and genetic clusters; however, the intragroup distances within each of the three clusters were below 3.00% suggesting that each cluster is composed of conspecific individuals. The new sequences of *U. terpsichore* clustered with sequences of *U. terpsichore*, *U. subangulata*, and *U. angulata* with an intraspecific distance of 2.10% suggesting that this cluster represents a single species.

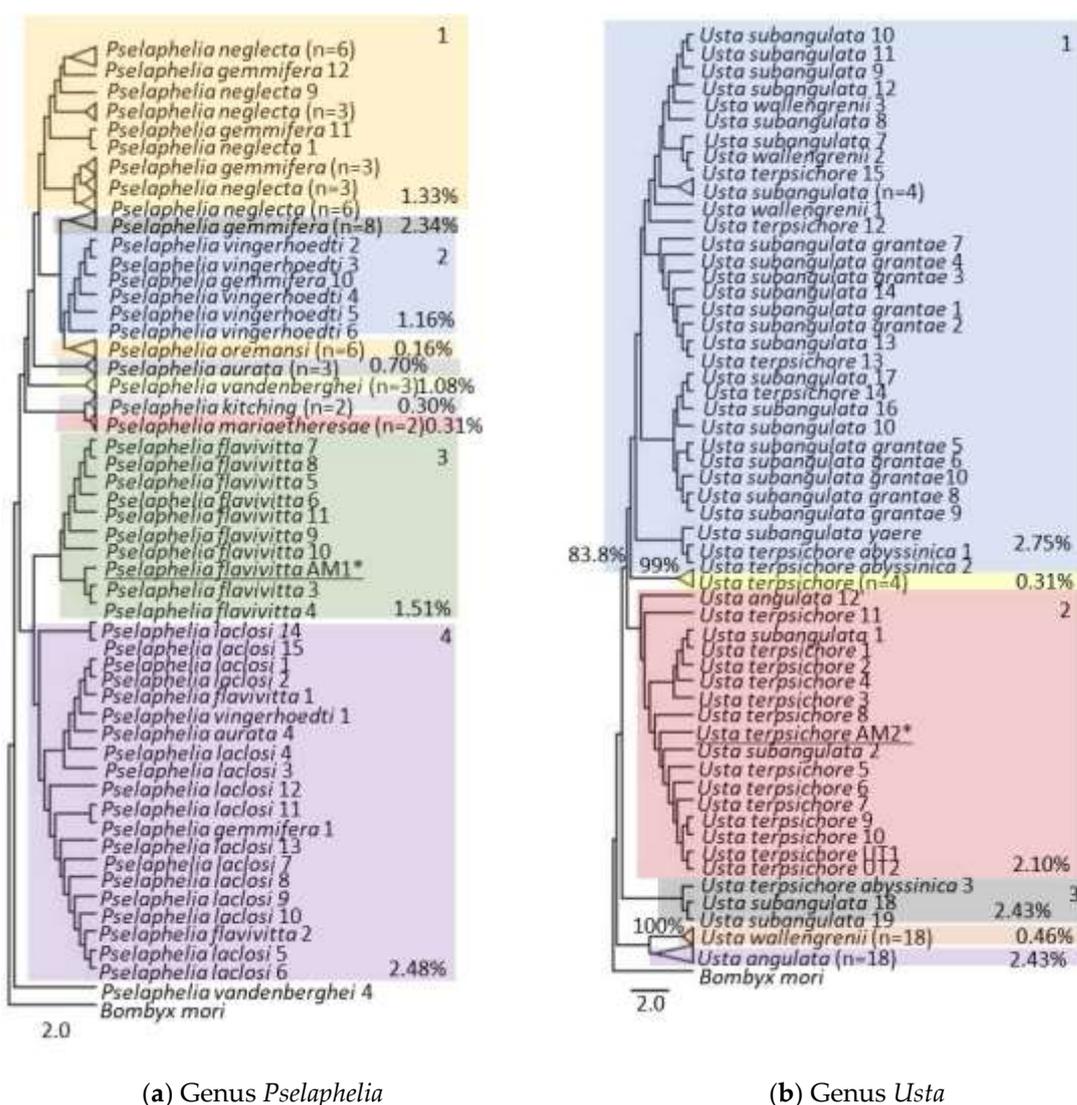


Figure 2. Neighbor-joining (K2P) trees of (a) 86 COI sequences of genus *Pselaphelia* (b) 76 COI sequences (658 bp) of genus *Usta*. The numbers on the NJ trees represent genetic clusters (intraspecific maximum p-distances <3%), and novel sequences are underlined.

A large proportion of the dataset showed the similar patterns of paraphyly. In these cases, more than one species clustered together on the NJ tree with intraspecific p-max distances less than 3.00% (e.g., Figure 2a and 2b). Most of the sequences on the NJ trees belonged to clusters that were not congruent with their species designations, and genetic divergence calculated based on groups of sequences with the same name were higher (>3.00%) than the range expected among conspecific individuals.

These patterns of paraphyly could be explained by misidentification of specimens due to lack of expertise and databasing mistakes. However, it may also indicate that species names fail to represent the genetic boundaries that delimit species into separate evolutionary entities, i.e., the evolutionary nature of species leads to taxonomic uncertainties which are reflected on the high level of paraphyly observed in this study. The observed phenotypic variation used for taxonomic names, despite gene flow, may be related to genetic or ecological selection, which makes it difficult to distinguish and classify specimens.

4. Conclusion

The reliability of DNA barcodes for identifying unknown specimens by comparison with publicly available sequences depends on the quality of the data that constitutes the reference library. Our detailed analyses of patterns of genetic diversity in 12 genera of African Saturniidae suggest that a large proportion of the data presently available on BOLD Systems is not in agreement with the notion that conspecific individuals have maximum p-distance < 3%. Therefore, we recommend caution when utilizing the available sequences to infer species identification of unknown specimens based on the current publicly available dataset.

Supplementary Materials: This work does not include supplementary material.

Author Contributions: BvA conceptualized the study, ZN and KH performed laboratory work. All authors performed data analyses, contributed to the preparation of this manuscript and approved the final version for publication.

Funding: The work was funded by the National Research Foundation of South Africa (grant number 121293). ZN was supported by the Foundational Biodiversity Information Programme (PhD grant number 128333).

Informed Consent Statement: Not applicable.

Data Availability Statement: The sequences generated in this study will be deposited on GenBank.

Acknowledgments: The authors are grateful to Andrew Morton, Bronwyn Egan, Chrissie Fourie, Gail Morland, Margriet Brink, and William Versveld for assistance with collection and identification of specimens.

Conflicts of Interest: The authors declare no conflict of interest.

References

- [1] I. J. Kitching *et al.*, "A global checklist of the Bombycoidea (Insecta: Lepidoptera)," *Biodivers. Data J.*, vol. 6, 2018, doi: 10.3897/BDJ.6.e22236.
- [2] J. C. REGIER, M. C. GRANT, C. MITTER, C. P. COOK, R. S. PEIGLER, and R. ROUGERIE, "Phylogenetic relationships of wild silkmoths (Lepidoptera: Saturniidae) inferred from four protein-coding nuclear genes," *Syst. Entomol.*, vol. 33, no. 2, pp. 219–228, Apr. 2008, doi: 10.1111/j.1365-3113.2007.00416.x.
- [3] T. Kalita and K. Dutta, "Characterisation of cocoon of different population of *Antheraea assamensis* (Lepidoptera: Saturniidae)," *Orient. Insects*, vol. 54, no. 4, pp. 574–590, 2020, doi: 10.1080/00305316.2020.1727376.
- [4] "WORLD LIST OF EDIBLE INSECTS 2015 (Yde Jongema) WORLD LIST OF EDIBLE INSECTS 2015 (Yde Jongema)," vol. 2015, pp. 1–75, 2015.
- [5] Z. T. Hlongwane, R. Slotow, and T. C. Munyai, "Nutritional composition of edible insects consumed in africa: A systematic review," *Nutrients*, vol. 12, no. 9, pp. 1–28, 2020, doi: 10.3390/nu12092786.
- [6] T. Lautenschläger *et al.*, "Edible insects of Northern Angola," *African Invertebr.*, vol. 58, no. 2, pp. 55–82, 2017, doi: 10.3897/afrinvertebr.58.21083.
- [7] S. Kelemu *et al.*, "African edible insects for food and feed : inventory , diversity , commonalities and contribution to food security," vol. 1, no. 2, pp. 103–119, 2015, doi: 10.3920/JIFF2014.0016.
- [8] J. Langley, S. Van der Westhuizen, G. Morland, and B. van Asch, "Mitochondrial genomes and polymorphic regions of *Gonimbrasia belina* and *Gynanisa maja* (Lepidoptera: Saturniidae), two important edible caterpillars of Southern Africa," *Int. J. Biol. Macromol.*, vol. 144, pp. 632–642, 2020, doi: 10.1016/j.ijbiomac.2019.12.055.
- [9] P. D. N. Hebert, A. Cywinska, S. L. Ball, and J. R. DeWaard, "Biological identifications through DNA barcodes," *Proc. R. Soc. B Biol. Sci.*, vol. 270, no. 1512, pp. 313–321, 2003, doi: 10.1098/rspb.2002.2218.
- [10] L. M. Nneji *et al.*, "Testing the effectiveness of DNA barcoding for biodiversity assessment of moths from Nigeria," *Diversity*, vol. 12, no. 2, pp. 1–12, 2020, doi: 10.3390/d12020085.
- [11] M. Pentinsaari, S. Ratnasingham, S. E. Miller, and P. D. N. Hebert, "BOLD and GenBank revisited - Do identification errors arise in the lab or in the sequence libraries?," *PLoS One*, vol. 15, no. 4, pp. 1–10, 2020, doi: 10.1371/journal.pone.0231814.
- [12] H. A. Ross, "The incidence of species-level paraphyly in animals: A re-assessment," *Mol. Phylogenet. Evol.*, vol. 76, no. 1, pp. 10–17, 2014, doi: 10.1016/j.ympev.2014.02.021.
- [13] M. Mutanen *et al.*, "Species-level para- and polyphyly in DNA barcode gene trees: Strong operational bias in European Lepidoptera," *Syst. Biol.*, vol. 65, no. 6, pp. 1024–1040, 2016, doi: 10.1093/sysbio/syw044.
- [14] R. A. Collins, N. Zealand, and N. Zealand, "The seven deadly sins of DNA barcoding," 2012, doi: 10.1111/1755-0998.12046.

-
- [15] J. Sambrook and D. W. Russell, "Molecular Cloning-Sambrook & Russel-Vol. 1, 2, 3," *Cold Spring Harb Lab Press*, 2001.
- [16] K. Katoh and D. M. Standley, "MAFFT multiple sequence alignment software version 7: improvements in performance and usability," *Mol. Biol. Evol.*, vol. 30, no. 4, pp. 772–780, 2013.
- [17] S. Kumar, G. Stecher, M. Li, C. Knyaz, and K. Tamura, "MEGA X: Molecular evolutionary genetics analysis across computing platforms," *Mol. Biol. Evol.*, vol. 35, no. 6, pp. 1547–1549, Jun. 2018, doi: 10.1093/molbev/msy096.
- [18] M. Kimura, "A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences," *J. Mol. Evol.*, vol. 16, no. 2, pp. 111–120, Jun. 1980, doi: 10.1007/BF01731581.