Molecular Biology Reports
https://doi.org/10.1007/s11033-020-05984-5

ORIGINAL ARTICLE

DNA barcoding and species delimitation of butterflies (Lepidoptera) from Nigeria

Lotanna Micah Nneji[1,2] · Adeniyi Charles Adeola[1,2] · Adeola Oluwakemi Ayoola[1] · Segun Olayinka Oladipo[3] · Yun-Yu Wang[1] · Yoila D. Malann[4] · Okorie Anyaele[5] · Ifeanyi Christopher Nneji[4] · Md Mizanur Rahman[1] · Caroline Samuel Olory[6]

Received: 22 August 2020 / Accepted: 5 November 2020
© Springer Nature B.V. 2020

Abstract

Accurate identification of species is a prerequisite for successful biodiversity management and further genetic studies. Species identification techniques often require both morphological diagnostics and molecular tools, such as DNA barcoding, for correct identification. In particular, the use of the subunit I of the mitochondrial *cytochrome c oxidase* (COI) gene for DNA barcoding has proven useful in species identification for insects. However, to date, no studies have been carried out on the DNA barcoding of Nigerian butterflies. We evaluated the utility of DNA barcoding applied for the first time to 735 butterfly specimens from southern Nigeria. In total, 699 DNA barcodes, resulting in a record of 116 species belonging to 57 genera, were generated. Our study sample comprised 807 DNA barcodes based on sequences generated from our current study and 108 others retrieved from BOLD. Different molecular analyses, including genetic distance-based evaluation (Neighbor-Joining, Maximum Likelihood and Bayesian trees) and species delimitation tests (TaxonDNA, Automated Barcode Gap Discovery, General Mixed Yule-Coalescent, and Bayesian Poisson Tree Processes) were performed to accurately identify and delineate species. The genetic distance-based analyses resulted in 163 well-separated clusters consisting of 147 described and 16 unidentified species. Our findings indicate that about 90.20% of the butterfly species were explicitly discriminated using DNA barcodes. Also, our field collections reported the first country records of ten butterfly species—*Acraea serena*, *Amauris* cf. *dannfelti*, *Aterica galena extensa*, *Axione tjoane rubescens*, *Charaxes galleyanus*, *Papilio lormieri lormeri*, *Pentila alba*, *Precis actia*, *Precis tugela,* and *Tagiades flesus*. Further, DNA barcodes revealed a high mitochondrial intraspecific divergence of more than 3% in *Bicyclus vulgaris vulgaris* and *Colotis evagore*. Furthermore, our result revealed an overall high haplotype (gene) diversity (0.9764), suggesting that DNA barcoding can provide information at a population level for Nigerian butterflies. The present study confirms the efficiency of DNA barcoding for identifying butterflies from Nigeria. To gain a better understanding of regional variation in DNA barcodes of this biogeographically complex area, future work should expand the DNA barcode reference library to include all butterfly species from Nigeria as well as surrounding countries. Also, further studies, involving relevant genetic and eco-morphological datasets, are required to understand processes governing mitochondrial intraspecific divergences reported in some species complexes.

Keywords DNA barcoding · Lepidoptera · Endemism · Mitochondrial DNA · Nigeria

Introduction

Butterflies, belonging to the order Lepidoptera, are distributed worldwide except in Antarctica [1]. An approximation of 18,000 species has been recorded globally, with more than 80% of the species found in the tropics [2–4]. Africa hosts over 4000 butterfly species, constituting 20% of the known butterfly species in the world [5, 6]. This makes Africa the second richest world region for butterfly diversity, after the Neotropics [7]. In the Afro tropics, West Africa harbors over

Lotanna Micah Nneji and Adeniyi Charles Adeola have contributed equally to this work.

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11033-020-05984-5) contains supplementary material, which is available to authorized users.

✉ Lotanna Micah Nneji
lotannanneji@gmail.com

✉ Adeniyi Charles Adeola
chadeola@mail.kiz.ac.cn

Extended author information available on the last page of the article

Published online: 16 November 2020

50% of its butterfly species, with Nigeria harboring more than 60% of the West African butterflies [8, 9]. The butterfly species richness in Nigeria is impressive when compared to other African countries [9]. Furthermore, new species and country first records are continually reported from Nigeria [10–12]. These studies suggest that butterfly diversity in Nigeria is vastly underestimated.

Accurate species identification is a pivotal component of biodiversity documentation and population genetic studies [10, 11]. In Nigeria, butterfly taxonomy is often an overlooked field of study. Most studies in this region addressed ecological questions that focused on diversity, abundance, and distribution patterns of butterflies [13–17], with only a few studies focusing on the taxonomy [18]. The routine identification of Nigerian butterflies is based on morphological characters. This approach faces enormous challenges due to the complex life cycle, morphological plasticity, and multiple morphs within butterfly species. For instance, in Nigeria, there are over 60 species of *Charaxes* with varying morphological traits and convergence. In such a scenario, morphology-based identification could be labor-intensive. Thus, it is essential to use alternative and rapid methods for solving taxonomic ambiguities and helping to identify species quickly.

In the last three decades, mitochondrial DNA, based on the *cytochrome c oxidase subunit I* (COI) gene, gained prominence as a molecular tool for identifying organisms and documenting biodiversity [19]. This process commonly referred to as DNA barcoding, is a quick and valid method for identifying species and delimiting complex cryptic species [20–28]. DNA barcoding has been applied in the biodiversity studies and species identification of diverse fauna, including fishes, amphibians, reptiles, mammals, birds, and invertebrates [20, 29–36]. Also, past studies have confirmed the relevance of this approach for lepidopterans with a variety of applications in taxonomy, biodiversity documentation, and ecology [35–45].

More than 1300 species of butterfly have been reported in Nigeria [9], even though only 3.6% of COI barcodes have been generated from the described species. To date, no studies have reported the DNA barcode analyses of Nigerian butterflies. Herein, our objectives are as follows: (i) to evaluate the effectiveness of DNA barcoding for identifying butterflies from southern Nigeria; and (ii) to establish the first curated DNA barcode reference library for butterfly species in Nigeria using the newly collected and previously published data. Our study substantially improved understanding of butterfly diversity in southern Nigeria, marked first country records of some butterfly species and enabled the creation of DNA barcodes of Nigerian butterflies. Besides, the information produced through this study will provide a credible knowledge base for entomologists, researchers, biodiversity scientists, and decision-makers for the development of sustainable conservation strategies for Nigerian butterflies.

Materials and methods

Site selection and sample collection

To create a first curated DNA reference library for Nigerian butterflies, we focused our field survey in Cross River National Park (CRNP) in southern Nigeria. A previous study by Terborgh [46] reported over 950 species of butterflies from the CRNP, with some species endemic to this protected area. Also, CRNP comprises a humid lowland forest [36], and according to Larsen [18], lowland forest is by far the most crucial ecological zone supporting abundant and diverse species of butterflies. To date, CRNP hosts 70% of the butterfly species in Nigeria [46]; thus, this highlights the uniqueness of CRNP as a hotspot for butterfly studies in Nigeria.

Ethical permission was acquired from the management of the Nigerian National Park Service. Between January 2017 and November 2019, we collected 735 specimens of butterflies from CRNP (see Table S1, supporting information). Field surveys were conducted in the morning (6:00–09:00am), afternoon (12:00–2:00 pm) and evening (4:00–6:00 pm). The specimens captured were euthanized with ethyl acetate. These were labeled, assigned reference codes, and kept in the insect collection of Dr. O. Anyaele of the Department of Zoology, University of Ibadan, Nigeria.

Morphology-based identification and voucher preservation

Specimens were identified based on the wing and genitalia patterns following the checklist and keys of identification guide for the butterflies of West Africa by Larsen [47]. We consulted the African butterfly database for further verification. Also, a qualified entomologist was assigned to verify the authenticity of the species name assigned to each specimen. For the DNA-based species identification, a hind leg was taken from each individual using sterile forceps. It was then preserved in a tube containing 95% ethanol. The ethanol-preserved tissue samples were stored at − 80 °C at the State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences (KIZ-CAS), China.

Molecular laboratory protocols

The extraction of the total genomic DNA, PCR amplification, and sequencing were performed following standard protocols [48–52]. For all individuals, we attempted to

generate approximately 500–640 base pairs (bp) of the partial sequence of the COI-5′ barcode region. The fragment was amplified with a primer pair designed by Hebert et al. [20]: LepF1 5′-ATTCAACCAATCATAAAGATATTGG-3′, and LepR1 5′-TAAACTTCTGGATGTCCAAAAAATCA-3′. Amplification was performed in a volume reaction of 25 µl comprising 1.5 µl of genomic working DNA, 18.5 µl of PCR water, 2.5 µl of PCR buffer, 1.2 µl of dNTP, 0.5 µl of each of the forward and reverse primers (10 pm/µl) and 0.30 µl of rTaq polymerase. The PCR cycle profiles were as follows: initial denaturation at 94° C for 5 min, accompanied by 35 cycles at 94 °C for 1 min, annealing at 46 °C for 45 s, an extension at 72 °C for 1 min, and a final extension at 72 °C for 10 min. The amplified products were imaged on 1.2% agarose gel containing gel red. Bidirectional sequencing of the purified PCR products was performed using an automated DNA sequencer (ABI 3730 xl).

Sequence alignment and sequence-based specimen identification

Using SeqManTMII (DNASTAR Lasergene 7), the verification of the nucleotide sequences was done manually. Alignment used ClustalW with default parameters in MEGA 6.0 [53]. The aligned sequences were translated into amino acids to search for premature stop codons and confirm that the open reading frame was maintained in the protein-coding locus. After that, a DNA-based similarity check was performed using 'BLAST' (https:/blast.ncbi.nlm.nih.gov/Blast.cgi) in the GenBank and the 'Identification Request' function (https:/www.boldsystems.org/index.php/IDS IdentificationRequest) on the Barcode of Life Database (BOLD). Given that previous studies [20, 54] have reported species delimitation for most lepidopterans based on COI sequence divergence of > 2%, we used a threshold of 2% divergence in our study. The neighbor-joining (NJ) trees were constructed using our newly amplified COI sequences and other sequences of related species in the BOLD. The Barcode Index Number (BIN) analysis was performed in the BOLD. We also estimated the Kimura-2-Parameter (K2P) [55] sequence divergence within various taxonomic levels (family, genus, and species) using MEGA 6.0. For each taxonomic level, mean, maximum, and minimum K2P genetic distances were computed.

Sequence data assembly and estimation of COI genetic similarities and diversity

We searched for the previously published partial sequences of the COI-5′ barcode region of butterfly species from Nigeria. In this study, we used sequences retrieved from BOLD because it provides specifics on the voucher specimens. In sum, 108 COI sequences of 58 butterfly species were retrieved and used in this study (see Table S2, supporting information).

The NJ tree approach was used to evaluate the COI genetic distance among species. The tree was constructed using MEGA 7.0. Initial tree construction comprised 807 COI sequences of Nigerian butterflies as ingroup taxa. We downloaded the COI sequence of Asian peach fruit moth—*Carposina sasakii* (KM547875) as the primary outgroup taxon based on a previous study [56]. The K2P distance model, using 10,000 bootstrap pseudoreplicates [57], was employed to assess monophyly.

For a more comprehensive inference of the COI genetic distance, the COI sequence dataset was collapsed to encompass the haplotypes of Nigerian butterflies. Haplotypes were generated using DnaSP 5.10. Using the Akaike Information Criterion (AIC) implemented in MrModelTest version 3.7 [58], the best-fit nucleotide substitution model was selected. The maximum likelihood (ML) and Bayesian Inference (BI) approaches were used to estimate COI sequence distances. The ML tree was constructed in RAxML version 7.2.8 [59], using a gamma distribution evolutionary model with 100 random addition replicates and per partition branch lengths. Using 1000 bootstrap replications, the branch support for the optimal likelihood tree was calculated. The branch support for the optimal likelihood tree was estimated using 1000 bootstrap replications. The BI was estimated in MrBayes version 3.2.1 [60]. The analysis was run for 10 million generations with sampling every 1000th generation. Two independent runs with four Markov Chain Monte Carlo chains were performed. We excluded the first 25% of trees as burn-in before the log-likelihood scores stabilized. The trees were constructed in the CIPRES Science Gateway version 3.3 web service [61]. The resulting trees was displayed using FigTree v1.4.2 [62].

Species delimitation analyses

The species delimitation tests, based on the COI sequence data, were conducted using four different approaches. The species delimitation analyses involved a subset of the COI haplotypes of Nigerian butterflies. The clusters/matches that resulted were regarded as the Operational Taxonomic Units (OTUs).

Approach 1: TaxonDNA

The species cluster analysis was performed with the 'Cluster' algorithm in TaxonDNA version 1.8 [63]. The species unit, herein referred to as clusters, was defined for sequences within each cluster based on pairwise (uncorrected) distances. The pairwise distances were increased from 1.0% to a maximum of 3.0% with an increment of 0.5% in each step.

Approach 2: automated barcode gap discovery (ABGD)

The automated barcode gap discovery (ABGD; [64]) was conducted on the online server (https://www.abi.snv.jussieu.fr/public/abgd/abgdweb.html). Analyses used the Jukes-Cantor (JC69), K2P, and Simple model distances. Parameters were set as follows: prior intraspecific divergence of P-min = 0.001 and P-max = 0.1, steps = 10, relative gap width (X) = 1.0, and Nb bins (for distance distribution) = 20, with the other parameters in default settings.

Approach 3: Bayesian posterior tree Poisson (bPTP)

The Bayesian Posterior Tree Poisson (bPTP; [65]) analysis was performed using the webserver (https://species.h-its.org). This web interface only permits the input of a phylogenetic tree at a time. The bPTP analyses used the rooted RAxML tree for the evaluation. Parameters were specified as follows: Number of MCMC generations = 200,000, thinning = 100, burn-in 25%, seed = default, and the outgroup taxon specified.

Approach 4: general mixed yule-coalescent (GMYC)

Firstly, a linearized Bayesian phylogenetic tree was generated as implemented in BEAST v1.8.4. The analysis used a Yule pure birth model tree prior [66, 67]. Setting in BEAUti version 1.8 [68, 69] was as follows: the best substitution model, tree prior (Yule process), number of generations = 10,000,000 generations with sampling at every 1000 step, and other parameters in default settings. The evaluation of the Effective Sample Size (ESS > 200) and trace files of runs were performed in Tracer v1.5 [70]. Tree files generated from the BEAST analysis were combined using the LogCombiner. The maximum clade credibility tree was produced in TreeAnnotator v1.6.1 [71]. The single (sGMYC) and multiple (mGMYC) thresholds were conducted in the R platform using the APE [72] and SPLITS [73] packages. The support values were classified into three categories: weak (≤ 0.49), moderate (0.5–0.79), and strong (> 0.8).

Results

Collection of specimens and morphological identification

Of the 735 specimens collected, 116 butterfly species belonging to 57 genera from six families were morphologically described as follows: Hesperiidae (n = 10 species), Lycaenidae (n = 12 species), Nymphalidae (n = 74 species), Papilionidae (n = 10 species), Pieridae (n = 9 species) and Riodinidae (n = 1 species) (Table 1). Our field collections included ten butterfly species, *Acraea serena*, *Amauris* cf. *dannfelti*, *Aterica galena extensa*, *Axione tjoane rubescens*, *Charaxes galleyanus*, *Papilio lormieri lormeri*, *Pentila alba*, *Precis actia*, *Precis tugela,* and *Tagiades flesus* recorded for the first time from Nigeria. Also, our field collections consist of 16 butterfly species, namely *Acraea* sp., *Caenides* sp., *Catuna* sp. 1 and 2, *Eurema* sp., *Euryphura* sp., *Lepidochrysops* sp 1 and 2, *Neptis* sp., *Pseudophilotes* sp. 1 and 2, *Tarucus* sp., *Telipna* sp., *Papilio* sp. 1 and 2, *Semomesia* sp., that could not be classified to species-level and were assigned to genera-level.

Amplification success, tree-based specimen identification, and COI sequence information

The mitochondrial COI barcode sequences of more than 500 bp of 699 specimens were generated from the 735 (95.10% amplification success rate), providing at least one COI sequence for each of the 116 morphospecies. All attempts to generate quality sequences for some individuals (n = 36) failed. Thus, these individuals were excluded. No insertion, deletion, or stop codons were observed in the newly amplified sequences. This indicates that the amplified sequences represent functional mitochondrial COI sequences.

Overall, 85–100% sequence similarity was recovered during the BLAST search in the NCBI database. The NJ tree-based identification via the BOLD showed the clustering of COI sequences of the newly collected Nigerian butterfly species with other similar species in the database (see Figures S1, supporting information). This allowed for the unambiguous species-level identification of these species. However, some individuals collected in the present study could not be unambiguously placed on a species-level using morphology and DNA barcodes. These individuals were tentatively conferred (cf.) to species or referred to genus, pending further taxonomic identification. Table S1 shows the information of the newly amplified COI sequences, including the GenBank accession numbers.

The BIN analysis included 82 BIN clusters that were congruent with other barcode data in the BOLD (Table 1). At the species level, the BINs were concordant with morphology-based identification. Among these BINs, 55 (67.10%) were indicated as singleton which implies that the BIN only refers to one species (Table 1). Also, we observed that 27 (32.90%) BIN clusters were assigned to two or more species (Table 1). Even in this case, we found the species name assigned to our specimens to be included among the names assigned to the BIN. Howbeit, the discrepancies in the species name assigned to BIN evidenced the need for taxonomic reassessment of some butterfly species. The BIN analyses revealed that the average intra-specific distance of the BIN

Table 1 Summary information of newly collected butterflies from Cross River National Park, southern Nigeria, including the Barcode Index Number details of obtained from the BOLD database (N = Number of Individuals barcoded; Avd = Average Distance; MxD = Maximum Distance; DNN = Distance to the Nearest Neighbor; AvD_NN = Average Distance to the nearest neighbor; MxD_NN = Maximum Distance to the nearest neighbor; N/A = Information not available in BOLD)

Family	Genus	Species Name	N	BIN	AvD (%)	MxD (%)	DNN (%)	BIN of the nearest neighbor	Nearest Member Taxonomy	AvD_NN (%)	MxD_NN (%)
Hesperiidae	*Caenides*	*Caenides dacela*	2	BOLD:AAI3824	0.25	0.46	3.85	BOLD:ADK1559	*Caenides dacela*	N/A	N/A
		Caenides sp.	2	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Celaenorrhinus	*Celaenorrhinus plagiatus*	1	BOLD:AAI3775 *	0.47	1.23	1.28	BOLD:ACF4534	*Celaenorrhinus zanqua uzungwae*	0.17	0.49
	Ceratrichia	*Ceratrichia phocion phocion*	3	BOLD:AAK0407	0	0	3.18	BOLD:AAF0912	*Ceratrichia semlikensis*	0.03	0.15
	Monza	*Monza alberti*	1	BOLD:ABZ8105	N/A	N/A	1.93	BOLD:AAK6587		0	0
		Monza cretacea	1	BOLD:AAI3294	0.57	0.84	6.6	BOLD:ADK0805	*Semalea* sp.	N/A	N/A
	Pardaleodes	*Pardaleodes incerta murcia*	2	BOLD:ACG5871	0.7	1.13	2.73	BOLD:ACR1685	*lepidoptera*	0	0
	Pteroteinon	*Pteroteinon ceucaenira*	1	BOLD:ADK7700	N/A	N/A	3.61	BOLD:AAL2299	*Pteroteinon pruna*	N/A	N/A
		Pteroteinon laufella	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Tagiades	*Tagiades flesus**	5	BOLD:AAH9519	0	0	3.88	BOLD:AAE9650	*Tagiades samborana samborana*	0.21	0.32
Lycaenidae											
	Anthene	*Anthene larydas*	5	BOLD:AAK2334*	0.27	0.67	3.1	BOLD:AAJ2410	*Anthene princeps*	1.01	1.01
	Axiocerses	*Axiocerses tjoane rubescens**	1	BOLD:AAI8079*	0.31	1.69	1.15	BOLD:ACE7107	*Axiocerses tbo* sp5	0.04	0.31
	Azanus	*Azanus mirza*	1	BOLD:ABV3861	0.28	0.96	2.25	BOLD:AAE1107	*Azanus sitalces*	0.05	0.15
	Lepidochrysops	*Lepidochrysops* sp1	3	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Lepidochrysops sp2	13	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Mimeresia	*Mimeresia cellularis*	1	BOLD:AAK6216	N/A	N/A	9.79	BOLD:AAK6217	*Mimeresia debora*	N/A	N/A
	Pentila	*Pentila alba**	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Pseudophilotes	*Pseudophilotes* sp1	9	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Pseudophilotes sp2	9	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Tarucus	*Tarucus* sp	10	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Telipna	*Telipna* sp	3	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Tuxentius	*Tuxentius cf. margaritaceus*	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
Nymphalidae											
	Acraea	*Acraea acerata*	8	BOLD:AAE7511	0.13	0.32	7.15	BOLD:ABZ7234	*Actinote conspicua*	0.35	0.35
		Acraea alcinoe	2	BOLD:AAE7545	0.28	0.96	9.95	N/A		N/A	N/A
		Acraea alciope	16	BOLD:AAH9115*	0.2	0.53	2.09	BOLD:ABY4558	*Acraea alciope*	0.37	0.92
		Acraea bonasia	5	BOLD:AAF5433	0.35	0.64	2.17	BOLD:AAN3288		1.03	1.77

Table 1 (continued)

Family	Genus	Species Name	N	BIN	AvD (%)	MxD (%)	DNN (%)	BIN of the nearest neighbor	Nearest Member Taxonomy	AvD_NN (%)	MxD_NN (%)
		Acraea cf. eponina	7	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Acraea cf. parrhasia parrhasia	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Acraea consanguinea consanguinea	1	BOLD:AAF5510	0.59	1.27	5.94	BOLD:AAF5266	*Acraea vestalis*	0.68	1.28
		Acraea epaea	6	BOLD:AAC6677	0.39	0.96	5.46	BOLD:AAC6676	*Acraea epaea*	0.68	1.53
		Acraea lycoa	10	BOLD:AAF5811	1.09	2.41	7.19	BOLD:AAD6718	*Acraea jodutta*	0.36	0.64
		Acraea polis	1	BOLD:AAE8460	0.11	0.16	1.74	BOLD:AAO0510	*Acraea pentapolis*	0.12	0.19
		Acraea pseudegina	6	BOLD:AAF5468*	0.5	1.12	2.89	BOLD:AAJ2899	*Acraea oncaea*	0.75	1.93
		*Acraea serena**	27	BOLD:ADQ5678	0.5	0.5	2	BOLD:ADQ5676	*Acraea eponina*	0.32	1.91
		Acraea sp	13	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Acraea zetes zetes	2	BOLD:AAC9577	0.4	0.96	1.77	BOLD:ABY8103	*Acraea acara*	0.31	0.46
	Amauris	*Amauris cf. dannfelti**	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Amauris democles democles	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Antanartia	*Antanartia cf. delius*	8	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Aterica	*Aterica galene extensa**	1	BOLD:AAN2209	N/A	N/A	6.26	BOLD:AAD7879	*Aterica rabena*	0.49	0.99
	Bebearia	*Bebearia carshena*	1	BOLD:AAF1153	0.32	0.48	4.01	BOLD:AAD2586	*Bebearia subtentyris*	N/A	N/A
		Bebearia cocalia katera	2	BOLD:AAC9553	0.08	0.35	3.53	BOLD:AAC9554	*Bebearia cocalia*	N/A	N/A
		Bebearia mardania	1	BOLD:AAK0052	0	0	3.53	BOLD:AAC9555	*Bebearia cocalia*	N/A	N/A
		Bebearia oxione oxione	1	BOLD:AAF0774	0.24	0.48	5.14	BOLD:ACF5559	*Bebearia seeldrayersi*	N/A	N/A
	Bicyclus	*Bicyclus dorothea dorothea*	36	BOLD:AAW5137*	0.83	1.5	7.45	BOLD:AAW5138	*Bicyclus vulgaris*	0.71	2.28
		Bicyclus ignobilis ignobilis	1	BOLD:AAI8585	1.53	2.25	6.13	BOLD:ADK2482	*Bicyclus maesseni*	N/A	N/A
		Bicyclus italus	1	BOLD:AAE3469*	0.27	0.65	2.66	BOLD:AAD6325	*Bicyclus graueri*	0.41	N/A
		Bicyclus martius sanaos	9	BOLD:ADK6543	1.47	1.89	3.12	BOLD:AAI8586	*Bicyclus martius*	1.33	1.33
		Bicyclus safitza	5	BOLD:AAI8583*	1.14	2.82	6.39	BOLD:ADK6749	*Bicyclus cottrelli*	N/A	N/A
		Bicyclus sandace	2	BOLD:ADK4207	1.17	1.17	6.85	BOLD:ADK6237	*Bicyclus smithi*	1.27	2.95
		Bicyclus technatis	1	BOLD:AAI8590	0.55	0.67	6.4	BOLD:ADK2894	*Bicyclus mollitia*	0.97	1.66

Table 1 (continued)

Family	Genus	Species Name	N	BIN	AvD (%)	MxD (%)	DNN (%)	BIN of the nearest neighbor	Nearest Member Taxonomy	AvD_NN (%)	MxD_NN (%)
	Catuna	*Catuna oberthueri*	8	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Catuna sp1	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Catuna sp2	8	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Charaxes	*Charaxes candiope*	1	BOLD:ABZ2098*	0.24	1	1.8	BOLD:AAC4995	*Charaxes antamboulou*	0	0
	Charaxes	*Charaxes cedreatis cedreatis*	3	BOLD:AAA6009*	1.73	4.81	1.61	BOLD:ACF4061	*Charaxes marieps*	0	0
		Charaxes etesipe etesipe	1	BOLD:AAB3242*	1.22	2.88	1.93	BOLD:ABY7585	*Charaxes cacuthis*	0.19	0.46
		Charaxes eupale eupale	3	BOLD:AAA9696*	0.45	1.56	2.25	BOLD:AAD7822	*Charaxes dilutus dilutus*	0.34	0.93
		*Charaxes galleyanus**	1	BOLD:AAA6009*	1.73	4.81	1.61	BOLD:ACF4061	*Charaxes marieps*	0	0
		Charaxes hadrianus	1	BOLD:AAC8826	0.47	1.09	5.7	BOLD:AAC4995	*Charaxes antamboulou*	0	0
		Charaxes imperialis albipuncta	1	BOLD:AAE2527	0.78	1.68	5.79	BOLD:AAC2113	*Charaxes violetta*	0.44	0.92
		Charaxes lucretius intermedius	15	BOLD:AAD7321	0.17	0.33	1.05	BOLD:ACE2933	*Charaxes lemosi*	0.15	0.15
		Charaxes numenes aequatorialis	1	BOLD:AAF4312*	0.14	0.32	2.62	BOLD:AAC2113	*Charaxes violetta*	0.44	0.92
		Charaxes porthos porthos	1	BOLD:ACE7389	0.47	1.4	1.61	BOLD:AAE8750	*Charaxes porthos gallayi*	0.82	1.07
	Charaxes	*Charaxes subornatus subornatus*	2	BOLD:AAC8841	0.15	0.63	2.73	BOLD:AAD7213	*Charaxes minor minor*	0.05	0.15
	Cymothoe	*Cymothoe beckeri beckeri*	4	BOLD:AAB5754	0.06	0.61	3.21	BOLD:AAD1357	*Cymothoe distincta*	1.35	3.31
		Cymothoe fumana balluca	3	BOLD:AAB7140*	0.87	1.82	2.68	BOLD:ABZ4416	*Cymothoe ogova*	0.19	0.54
	Cynandra	*Cynandra opis opis*	2	BOLD:ADR8965	N/A	N/A	5.78	BOLD:AAL5339	*Euriphene melanops*	N/A	N/A
	Danaus	*Danaus chrysippus*	2	BOLD:ABX5122*	0.54	2.02	1.58	BOLD:AAB3216	*Danaus chrysippus*	0.02	0.2
	Elymniopsis	*Elymniopsis bammakoo bammakoo*	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Euphaedra	*Euphaedra ceres lutescens*	4	BOLD:AAB0368*	1.09	2.92	2.91	BOLD:ADR5728	*Euphaedra medon*	1.17	1.17
		Euphaedra hewitsoni sumptuosa	1	BOLD:ADR1993	N/A	N/A	1.93	BOLD:AAE4896	*Euphaedra abri*	0.08	0.15

Table 1 (continued)

Family	Genus	Species Name	N	BIN	AvD (%)	MxD (%)	DNN (%)	BIN of the nearest neighbor	Nearest Member Taxonomy	AvD_NN (%)	MxD_NN (%)
	Euriphene	*Euriphene barombina*	3	BOLD:AAD1170*	0.25	0.64	1.44	BOLD:ACF4121	*Euriphene conjugens*	0.69	1.38
		Euriphene cf. barombina	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Euriphene gambiae gabonica	1	BOLD:AAJ0448	0	0	4.17	BOLD:AAF6518	*Euriphene iris*	0.61	0.76
	Euryphura	*Euryphura cf. plautilla*	5	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Euryphura sp.	2	BOLD:AAA7865*	1.71	3.67	5.48	N/A		N/A	N/A
	Euxanthe	*Euxanthe crossleyi crossleyi*	1	BOLD:ACE7270	0.15	0.46	1.61	BOLD:ACE7271	*Euxanthe wakefieldi*	0.25	0.37
	Hypolimnas	*Hypolimnas anthedon anthedon*	5	BOLD:ADR7650	0.17	0.63	1.29	BOLD:AAX2107	*Hypolimnas anthedon*	0.22	0.51
		Hypolimnas salmacis salmacis	2	BOLD:ADR2909	0.63	0.63	1.73	BOLD:ADR2484	*Hypolimnas salmacis*	0.26	0.63
	Junonia	*Junonia cf. stygia*	4	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Junonia oenone oenone	23	BOLD:ABZ7488*	0.81	2.19	1.83	BOLD:ABZ6191	*Junonia orithya*	0.61	3.34
		Junonia sophia sophia	19	BOLD:AAI8028	0.21	0.66	4.84	BOLD:AAI8026	*Junonia* sp.	0	0
		Junonia terea	86	BOLD:AAD8116	0.03	0.16	1.11	BOLD:ACE4154	*Junonia terea*	0.88	0.88
	Kallima	*Kallima rumia*	9	BOLD:AAX0519*	1.75	1.75	8.28	N/A		N/A	N/A
	Neptis	*Neptis melicerta*	1	BOLD:AAI3132*	0.98	1.93	5.94	BOLD:AAI3144	*Neptis species79_ANEPT*	0.23	0.31
		Neptis nemetes nemetes	1	BOLD:ACU2956	0.37	0.76	6.74	BOLD:ACU3896	*Neptis frobenia*	0.04	0.16
		Neptis nicoteles	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Neptis sp.	2	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Palla	*Palla ussheri ussheri*	3	BOLD:AAD5991*	0.9	1.9	2.89	BOLD:AAX1840	*Palla publius centralis*	1.43	1.43
	Precis	*Precis actia**	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		*Precis tugela**	2	BOLD:ABA9386	0.08	0.15	3.37	BOLD:ACG5635	*Precis rauana*	N/A	N/A
	Protogoniomorpha	*Protogoniomorpha parhussus parhussus*	3	BOLD:AAF1736*	0.74	2.01	4.36	BOLD:AAE7781	*Protogoniomorpha anacardii*	0.19	0.32
	Pseudacraea	*Pseudacraea eurytus eurytus*	1	BOLD:AAE1021*	0.31	1.12	9.08	N/A		N/A	N/A
	Tirumala	*Tirumala petiverana*	7	N/A	N/A	N/A	N/A	N/A		N/A	N/A

Table 1 (continued)

Family	Genus	Species Name	N	BIN	AvD (%)	MxD (%)	DNN (%)	BIN of the nearest neighbor	Nearest Member Taxonomy	AvD_NN (%)	MxD_NN (%)
	Ypthima	*Ypthima doleta*	26	BOLD:AAG5140	0.44	1.6	4.98	BOLD:AAG5138	*Ypthima*	N/A	N/A
Papilionidae											
	Graphium										
		Graphium cf. policenes	37	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Graphium latreillianus	3	BOLD:AAF7123	0.41	0.61	3.53	BOLD:AAI6671	*Graphium tynderaeus*	0	0
		Graphium leonidas leonidas	4	BOLD:AAI6670	0.16	0.33	2.94	BOLD:AAE3691	*Graphium cyrnus*	0	0
		Graphium policenes policenes	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Papilio	*Papilio dardanus dardanus*	2	BOLD:ADR7455	0.67	1.06	4.17	BOLD:ADS3161		N/A	N/A
		Papilio demodocus demodocus	6	BOLD:AAE2913	0.44	1.28	2.84	BOLD:AAD3215	*Papilio erithonioides*	0.05	0.17
		*Papilio lormieri lormieri**	9	BOLD:AAF7776	1.21	2.09	6.54	BOLD:AAI5285	*Papilio syfanius*	0.14	1.29
		Papilio nireus nireus	1	BOLD:AAI2538	0.33	0.67	3.05	BOLD:ADS3162	*Papilio bromius*	N/A	N/A
		Papilio sp1	2	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Papilio sp2	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A
Pieridae											
	Appias	*Appias epaphia*	4	BOLD:AAJ1893*	0.62	1.2	5.33	BOLD:AAE7114	*Appias sabina*	0.38	0.78
		Appias sabina sabina	4	BOLD:AAE7114	0.42	0.81	5.4	BOLD:AAJ1893	*Appias sabina*	0.62	1.2
	Catopsilia	*Catopsilia florella*	5	BOLD:AAD2080	0.22	1.28	3.26	BOLD:AAK2431	*Catopsilia pyranthe*	0.2	1.74
	Colotis	*Colotis euippe euippe*	2	BOLD:ABY7106	N/A	N/A	1.95	BOLD:AAA9767	*Colotis euippe*	0.63	1.93
	Eurema	*Eurema brigitta brigitta*	1	BOLD:AAD1193*	0.04	0.76	2.36	BOLD:AAF6473	*Eurema brigitta pulchella*	0	0
		Eurema hecabe solifera	21	N/A	N/A	N/A	N/A	N/A		N/A	N/A
		Eurema senegalensis	18	BOLD:ABX5427	0.56	0.76	1.93	BOLD:ABZ7036	*Eurema senegalensis*	N/A	N/A
		Eurema sp.	12	N/A	N/A	N/A	N/A	N/A		N/A	N/A
	Leptosia	*Leptosia alcesta alcesta*	40	BOLD:AAF1830	N/A	N/A	4	BOLD:AAF1829	*Leptosia alcesta*	0.16	0.31

Table 1 (continued)

Family	Genus	Species Name	N	BIN	AvD (%)	MxD (%)	DNN (%)	BIN of the nearest neighbor	Nearest Member Taxonomy	AvD_NN (%)	MxD_NN (%)
Riodinidae											
	Semomesia	*Semomesia* sp.	1	N/A	N/A	N/A	N/A	N/A		N/A	N/A

Species with asterisk (*) represent individuals recorded for the first time in Nigeria; BIN number with asterisk (*) represent BIN assigned to two or more species, and BIN with asterisk (*) refers to BIN with only one species

for each species ranged between 0.00 and 1.75%, with the maximum intra-specific distance ranging between 0.00 and 4.81%, and the mean genetic distance to the nearest neighbor ranging between 1.05 and 9.95% (Table 1). Analyses showed that the average genetic distance of nearest neighbor for each species ranged between 0.00 and 1.43%, with a maximum genetic distance ranging between 0.00 and 3.34%.

For a complete inference of the COI genetic similarities, we combined our newly acquired dataset with other COI sequences of Nigerian butterflies from the BOLD. Taken together, our dataset for the phylogenetic analyses included a total of 807 individuals consisting of 163 species to 57 genera of six families, viz., Hesperiidae (n = 16 species), Lycaenidae (n = 12 species), Nymphalidae (n = 110 species), Papilionidae (n = 10 species), Pieridae (n = 14 species) and Riodinidae (n = 1 species) (see Table S1 and 2, supporting information).

Approximately 625 bp of COI sequence was aligned after trimming the ambiguous bases. The nucleotide diversity of the entire dataset was 0.13015. A total of 274 polymorphic sites and 511 conserved sites were recorded. Parsimony informative sites with two, three, and four variants were 116, 73, and 74. The complete data set was into 304 haplotypes with a diversity of 0.9764. Table 2 shows detailed information on the sequence compositions. Overall, the highest GC% (31.60%) was recorded in the family Pieridae (Table 2).

Mean genetic divergence analysis

The mean K2P genetic divergences increased with the taxonomic level (Table 3). Intraspecific divergences ranged from 0.0 to 5.92% with a mean of 0.45%, while conspecific divergences ranged from 1.70 to 15.96% with a mean of 5.97%. The distances within families ranged from 9.58 to 15.68%, with a mean of 12.39%. The minimum mean K2P congeneric divergence (1.70%, SE ± 0.44) was almost four-fold higher than the average conspecific (0.45%, SE ± 0.15) divergence. Two species had intraspecific divergences of > 3.00%; *Bicyclus vulgaris vulgaris* (3.10%, SE ± 0.77) and *Colotis evagore* (3.72%, SE ± 0.77). Other species had low COI intraspecific genetic distances of < 3.00%.

Estimation of COI genetic similarities

The NJ tree, representing both new and previously published COI gene datasets, clustered individuals of the same species with high bootstrap values (Fig. 1 and Figure S2). In our current study, most morphologically described species were well-differentiated and formed separate COI clusters supported by high bootstrap values (Fig. 1). The ML (Figure S3) and BI (Figure S4) analyses of the haplotypes of the

Table 2 Nucleotide composition, overall and family-wise GC content and GC at codon position 1, 2 and 3

S/no	Nucleotide	All Species (%)	Family-wise Nucleotide composition (%)					
			Hesperiidae	Lycaenidae	Nymphalidae	Papilionidae	Pieridae	Riodinidae
1	G	14.7	14.4	14.6	14.8	14.4	14.7	14.1
2	C	16.9	16.9	15.2	17.2	15.7	16.9	14.6
3	A	30.1	29.4	32.3	30.0	30.3	29.7	33
4	T	38.2	39.3	37.8	38.0	39.6	38.7	38.4
5	GC	31.6	31.3	29.8	32	30.1	31.6	28.7
6	GC1	41.6	42.2	38.2	42.1	40.5	41.9	38.8
7	GC2	42.4	42.5	42.3	42.3	42.5	43	43.2
8	GC3	10.9	9.1	9.2	11.7	7.38	9.88	3.9

G guanine, *C* cytosine, *A* adenine, *T* thymine

Table 3 Percentage K2P sequence divergence at the *COI* barcode region among the 97 butterfly species with > 2 specimens, among the 26 genera with two or more species and among the five families with two or more genera [Min = Minimum K2P distance (%), Max = Maximum K2P distance (%), SE = Standard Error of the Mean K2P distance(%)]

Taxonomic level	Number of taxa	Min (%)	Mean (%)	Max (%)	SE (%)
Within species	97	0.00	0.45	5.92	0.15
Within genus	26	1.70	5.97	15.96	0.94
Within family	5	9.58	12.39	15.68	1.2

complete COI sequence dataset created monophyletic groups consisting of 163 species.

Species delimitation

For the species delineation using TaxonDNA, we only included species with two or more COI sequences. When a 2% threshold was applied, all the morphospecies were delineated. With a few exceptions, the OTUs formed by the ABGD, bPTP, and GMYC agreed with the species' clusters recovered in the tree-based identification (Fig. 2).

For the ABGD, the resulting OTUs were largely similar to the JC69 and K2P methods (Figures S5-7 of the supporting information). In contrast, the simple distance method differs in its result (Figure S5 of the supporting information). The JC69 and K2P produced six partitions with multiple prior maximal distances of three each for the recursive and initial partitions, which delimited the total dataset into 150 (P = 0.001000, 0.001668, and 0.002783) and 149 (P = 0.004642, 0.007743, and 0.012915) putative species, respectively (see Table S3, supporting information). The simple distance method produced lesser partitions (n = 5) with one and multiple prior maximal distances of four, which delimited the entire dataset into 148 (Initial partition; P = 0.004642 and 0.007743) and 149 (recursive partition; P = 0.001000, 0.001668, and 0.002783) known species, respectively (see Table S3, supporting information). The ABGD tool with prior maximal distance (P = 0.001000, 0.001668, 0.002783, 0.004642, and 0.007743) delimited the species according to the morphological identification and largely in accordance with the other tree-based analyses (see Table S3, supporting information).

Our tree-based species delimitation method (bPTP and GMYC) produced congruent OTUs. The bPTP analysis was performed with 200,000 MCMC sampling recommended when the number of species is more than 50. The analysis showed an acceptance rate of 0.15, with merges of 50,016, splits of 49,984, and 161–182 estimated number of species with a mean value of 170.51 (see Table S4 and Figure S8, supporting information). The GMYC model generated 62 ML clusters, 159 ML entities with a confidence interval between 60 and 62 (see Table S5, supporting information). The number of identified species were higher for the bPTP and GMYC; however, only marginal differences were observed among these species' delimitation methods. The numbers of species delimited by bPTP (161–182 putative species) and GMYC (156–164 putative species) analyses were similar to those recovered by the ABGD (i.e., 148–150 putative species with > 1.00% inter-specific divergence).

Discussion

Our study presents the first DNA barcode analysis of the butterflies from Nigeria, incorporating both the newly generated and previously published mitochondrial COI gene sequences. Compared to the rates of some other insect groups, the overall success rate of DNA sequencing and barcode generation of 95.10% in our current study is high. Specifically, the overall success rate of our current study was high compared to 63% in beetles [74], 83% in neuropterans [75], and 93.50% in butterflies from southern South America [76]. Our DNA barcoding success rate, however, was

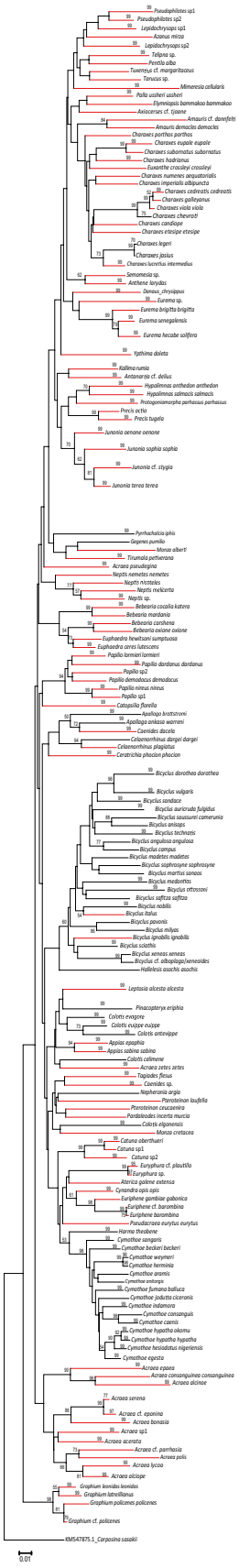


Fig. 1 The collapsed Neighbor-joining tree of the 807 individuals of butterflies from Nigeria. The tree has been collapsed to show the 163 taxa presented in this work and their names. The NJ tree was estimated on the basis of Kimura 2-parameter. Red branches indicate taxa for which sequences were generated for the first time from Nigeria. Black branches indicate taxa for which sequences have been generated by previous studies. Numbers above branches correspond to bootstrap support based on 1000 pseudoreplicates. The primary outgroup taxon was specified as *Carposina sasakii* (KM547875)

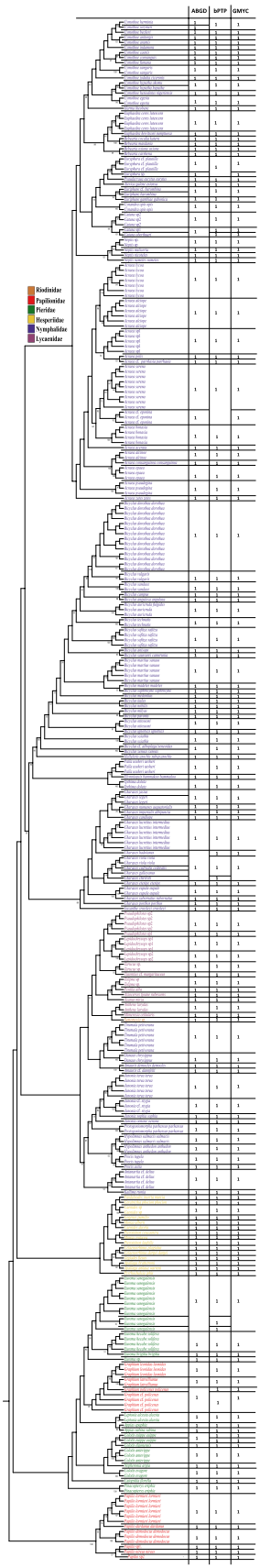


Fig. 2 The COI gene tree (Maximum likelihood (ML) tree based on the GTR + G model) showing relationships based on the 304 unique COI haplotypes of Nigerian butterflies. Numbers above or below branches correspond to bootstrap support based on 1000 pseudoreplicates. The numbers on the right represent the taxa predicted by each of the species delimitation methods—General Mixed Yule-coalescent (GMYC), mPTP, Automated Barcode Gap Discovery (ABGD; based on K2P method) and Bayesian Posterior Tree Poisson (bPTP, based on ML estimation)

lower when compared to the 100% success rates recorded by Yang et al. [77] and Marín et al. [78] in Satyrine butterflies. While specimens that failed to amplify could have been re-sequenced with an additional primer set, due to the cost and because at least one COI sequence was amplified for each morphospecies, we did not attempt to re-amplify failed sequences. Nevertheless, the sequences generated in our current study are generally essential. The COI sequences of butterflies from Nigeria are typically scarce in the global database. Until this study, less than 200 COI sequences of Nigerian butterflies were found on the BOLD and GenBank databases. With the 699 sequences presented in this study, we increased the data available of an additional ~12.00%. The newly generated sequences are relevant for the molecular characterization of Nigerian butterfly species and further molecular ecology studies.

One of the main objectives of this study is to explore the suitability of DNA barcodes to identify Nigerian butterflies. Applying the DNA barcoding approach, 90.20% of the newly collected morphospecies from CRNP were assigned to a species-level identification. Some of these newly recorded species in Nigeria constitute threats to some economically important trees. For instance, the larva of *A. tjoane*. *rubescens* feeds on *Acacia* (including *A. abyssinica*, *A. polyacantha campylacantha*, and *A. sieberiana* var. *woodii*), *Peltophorum africanum,* and *Brachystegia* species. More new records are likely to be present among the 16 taxa from CRNP that could not be identified to species-level. Previous studies [79–81] have shown that the identification of insects based on morphology alone could be challenging. This is particularly true for some morphologically cryptic insect pests. In these circumstances, our study, in accordance with previous studies [20, 82–86], indicated that DNA barcoding offers a fast and reliable means of identifying organisms. However, incorporating DNA barcodes with morphological and ecological characters will aid the rapid and authentic classification of Nigerian butterflies.

Another goal of the present study is to establish the first curated DNA barcode reference library for butterflies in Nigeria using the newly collected and previously published data. Combining recently assembled sequences with others in the global database, our research contributes to creating a DNA barcode reference library for Nigerian butterflies with more than 600 DNA barcodes. Our results broadly support those of previous studies by Hebert et al. [20], Yang et al. [77], Lavinia et al. [76], Marín et al. [78] and Ashfaq et al. [87, 88], suggesting that DNA barcoding should be universally applicable for rapid identification and delineation of butterfly species. According to the sequence threshold of 10X of Hebert et al. [20], the mean intraspecific mt COI variation is an approximate criterion for species identification. Based on the sequence threshold of 10X, the intraspecific nucleotide divergence percentage is expected to be less than 3%, and that of interspecies should be more than 3% for accurate species designation. Further, the previous study by Carew et al. [89] has shown the maximum intraspecific variation of COI sequences ranging from 3 to 3.9% enabled species identification. In our present study, the minimum mean K2P con-generic divergence was four-times higher than the mean con-specific divergence. The COI sequences analyzed in our study exhibited high interspecies variability based on nucleotide sequences. According to 10X rule, the observed intraspecific divergence was higher to discriminate between the individuals, suggesting distinctive species delineation.

The average COI intraspecific genetic divergence of 0.45% observed in our present study is higher than previous studies on butterflies (0.17% in [23]; 0.26% in [40]; 0.2% in [88]; 0.8% in [77]). This may be attributed to extensive sampling in our current study compared to other previous studies. The DNA barcodes generated in our study revealed high intraspecific divergence within some species, which deserve to be studied in more detail. Specifically, our study shows that *B. vulgaris vulgaris* and *C. evagore* had COI intraspecific divergences of > 3%, possibly indicating the presence of cryptic diversity. These species have a wide distribution range across several geographic barriers (e.g., a mountain, Cameroon Volcanic Active Line, etc.), with previously described ecological heterogeneity and morphological plasticity. For instance, *B. vulgaris* is widely distributed throughout the forest (including pre-forest and riparian forest), dense savanna, the edges of farmland, woodland, and marshy areas of tropical east and west Africa [90]. On the other hand, *C. evagore* is widely observed in the dry parts of tropical Africa, northern Africa, southern Spain, and southwest Arabia [90]. These results reinforce the need for detailed phylogeographic studies investigating processes governing the diversification of these species. The present data with only the barcode data from Nigeria did not enable us to test this hypothesis. More extensive genetic sampling throughout the full range may reveal genetic differences apparent in its range and examine the possible correlation of the genetic variation with geographic distance and barriers. It is obvious that the establishment of DNA barcode reference libraries will allow not only the accurate species-identification but also aid in answering molecular ecology questions regarding the patterns of diversification across the continent.

The tree-based identification using NJ, BI, and ML methods showed that all species were unambiguously distinguishable from all other species, by the formation of distinct, non-overlapping COI clusters. This is particularly true for the newly barcoded specimens. The results of different species delimitation methods (TaxonDNA, ABGD, GMYC, and bPTP) allowed comparison to observe any discrepancies. Cluster analysis using TaxonDNA revealed that most of the morphospecies with 2 or more barcodes

were correctly identified. Comparing the species delineation methods, our study showed that the concordance with the numbers of species delimited by bPTP, GMYC analyses, ABGD and the tree-based identification using NJ, BI, and ML methods. Nevertheless, discrepancies were observed in some cases (Fig. 2). For instance, individuals designated as *C. hesiodotus nigeriensis* by the bPTP (bootstrap support of 0.95) clustered with *Cymothoe hesiodotus nigeriensis, C.hypatha hypatha* and *C. hypatha okomu* in the ABGD analyses (Fig. 2). Our K2P genetic divergence revealed low genetic divergence of 2.5% between *C. hesiodotus nigeriensis* and *C. hypatha hypatha*, and 2.0% genetic divergence between *C. hesiodotus nigeriensis* and *C. hypatha okomu*. A shallow K2P genetic divergence of 1.3% was observed between *C. hypatha hypatha* and *C. hypatha okomu*. Such a case, as mentioned, was also observed between *Charaxes galleyanus*, *Charaxes cedreatis cedreatis* versus *Charaxes viola viola*; *Cymothoe weymeri* versus *Cymothoe herminia*; and *Charaxes jasius* versus *Charaxes legeri*. We conclude that these species groups might represent complexes of species or even a case of misidentification. Therefore, studies are required to confirm the taxonomic status of these species. Such studies will include extensive sample collection and analyses involving multiple genetic markers, morphological characters, and ecological traits.

It is worth noting that the public reference libraries, such as BOLD, are not yet adequately filled with COI sequences of African butterfly species. For the biodiversity and bioassessment studies of African butterflies to take advantage of the benefits that DNA barcoding offers, inventories and taxonomic studies of butterflies should incorporate its DNA barcodes. If such studies could use combined molecular and morphological approaches to identify all the butterfly species in the region, gaps in the reference library can be filled, genetic variations examined, and eco-morphological characteristics evaluated.

Conclusion

Our study confirmed the use of DNA barcoding as a successful and reliable tool to identify Nigerian butterflies. A high degree of mitochondrial intraspecific divergence within some species may indicate cryptic diversification that requires further study. Given the threats facing butterflies in the forest ecosystems in Africa, there is an urgent need for comprehensive biodiversity inventory and documentation using a rapid and reliable taxonomic tool. Our study concludes that establishing a reliable COI barcode reference library would assist entomologists, geneticists, researchers, students, biodiversity managers, and policymakers in the biodiversity studies and implementation of conservation plans for Nigerian butterflies. Finally, our study advocates the need for integration of molecular, morphological, and ecological datasets in the studies on the diversity and taxonomy of butterflies from Nigeria.

Acknowledgements We are grateful to field assistants at the Cross River National Park Nigeria, for their help during sampling. The authors wish to thank Deme Gideon Gyma and Daniel Bassey for their assistance during the field survey.

Author contributions LMN and ACA designed and supervised the project. LMN, AOA, OA, YM., SOO, ICN, and AOA collected the samples. LMN, ACA, and YYW performed the molecular laboratory works. LMN and MMR performed genetic analyses. CSO provided technical assistance. LMN wrote the initial draft of the manuscript. ACA revised the initial draft of the manuscript. All authors read and approved the final manuscript.

Funding This work was supported by the Whitley Wildlife Conservation Trust, Sino-Africa Joint Research Centre, Chinese Academy of Sciences (SAJC201611), National Natural Science Foundation of China (31750110480), and the Animal Branch of the Germplasm Bank of Wild Species, Chinese Academy of Sciences (the Large Research Infrastructure Funding).

Compliance with ethical standards

Conflict of interest All authors have declare that no conflicts of interest exist.

Ethical approval Permission to collect animal specimens (NPH/GEN/121/XXV/461) from Cross River National Park was obtained from the National Park Service Headquarter, Abuja, Nigeria.

References

1. Gullan PJ, Cranston PS (2004) The insects: an outline of entomology, 3rd edn. Wiley-Blackwell, Hoboken, pp 198–199
2. Boriani L, Burgio G, Marini M, Genghini M (2005) Faunistic study on butterflies collected in Northern Italy rural landscape. Bull Insectol 58(1):49–56
3. Bonebrake TC, Ponisio LC, Boggs CL, Ehrlich PR (2010) More than just indicators: a review of tropical butterfly ecology and conservation. Biol Conserv 143:1831–1841
4. Aiswarya VN, Pradarsika M, Soma AB (2014) Studies on the diversity and abundance of butterfly (Lepidoptera: Rhopalocera) fauna in and around Sarojini Naidu college campus, Kolkata, West Bengal, India. J Entomol Zool Stud 2(4):129–133
5. Ackery PR, Smith CP, Vane-Wright R (eds) (1995) Carcasson's African butterflies. an annotated catalogue of the Papilionoidea and Hesperoidea of the Afrotropical region. CSIRO Publishing, Melbourne, p 816
6. Kristensen NP, Scoble MJ, Karsholt O (2007) Lepidoptera phylogeny and systematics: the state of inventorying moth and butterfly diversity. Zootaxa 1668:699–747
7. Emmel TC, Larsen TB (1997) Butterfly diversity in Ghana West Africa. Trop Lepidoptera 8(supplement 3):1–13
8. Viejo P, Paul L, Curie N (2000) Conservation biology, 1st ed. p 300
9. African Butterfly Database (2020) Species list of butterflies in Nigeria. https://abdb-africa.org/locality/Nigeria?oid=13. Accessed 5 May 2020
10. Tropek R, Jansta P, Lestina D (2013) Acraea wigginsi occidentalis (Bethune-Baker, 1926), a new butterfly for Nigeria, with

remarks on its habitat and known distribution (Lepidoptera: Nymphalidae). Shilap 41:163–165
11. Sáfián S, Pyrcz TW, Brattström O (2016) Two new species of *Bebearia* Hemming, 1960, as further evidence of centre of endemism of butterflies in Western Nigeria (Lepidoptera: Nymphalidae: Limenitinae). Zootaxa 4175(5):449
12. Pyrcz TW, Sáfián S (2018) A new subspecies of *Cymothoe fumana* (Westwood, 1850) from Western Nigeria (Lepidoptera: Nymphalidae: Limenitidinae). Metamorphosis 29:5–8
13. Amusan B, Ojianwuna C, Kehinde T, Akanbi A (2014) Butterfly diversity in Obafemi Awolowo University, Ile-Ife, Southwest Nigeria. Zoologist 12:1–7
14. Alarape AA, Omifolaji JK, Mwansat GS (2015) Butterfly species diversity and abundance in university of Ibadan Botanical Garden, Nigeria. Open J Ecol 5:352–360
15. Kemabontal KA, Ebiyon AS, Olaleru F (2015) The Butterfly fauna of three varying habitats in South Western Nigeria. FUTA J Res Sci 1:1–6
16. Yager GO, Agbidye FS, Okoh AO (2016) Diversity and abundance of butterfly species (lepidoptera) fauna in Federal University of Agriculture, Makurdi Forestry Nursery, Benue State Nigeria. J Res For Wildl Environ 8(3):83–89
17. Orimaye JO, Ogunyemi OO, Okosodo EF, Ojo VA, Agbelusi TO (2016) Butterfly species diversity in protected and unprotected habitat of Ise Forest Reserve, Ise Ekiti, Ekiti State. Adv Ecol 2016:1–5
18. Larsen TB (1995) Butterfly research in the Oban Hills, cross River National Park. Calabar: Oban Hills Programme, Second Interim Report
19. Hebert PDN, Cywinska A, Ball SL, Waard JR (2003) Biological identifications through DNA barcodes. Philos Trans R Soc B 270:313–321
20. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. Philos Trans R Soc B 101:14812–14817
21. Ward RD, Zemlak ST, Innes BH, Last PR, Hebert PDN (2005) DNA barcoding Australia's fish species. Philos Trans R Soc B 360:1847–1857
22. Cywinska A, Hunter FF, Hebert PDN (2006) Identifying Canadian mosquito species through DNA barcodes. Med Vet Entomol 20:413–424
23. Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. Philos Trans R Soc B 103:968–971
24. Hajibabaei M, Smith MA, Janzen DH, Rodriguez JJ, Whitefiels JB, Hebert PDN (2006) A minimalist barcode can identify a specimen whose DNA is degraded. Mol Ecol Notes 6:959–964
25. Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PDN (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. Philos Trans R Soc B 104:4967–4972
26. Borisenko AV, Burton KL, Ivanova NV, Hanner RH, Hebert PDN (2008) DNA barcoding in surveys of small mammal communities: a field study in Suriname. Mol Ecol Res 8:471–479
27. Janzen DH, Hallwachs W, Blandin P, Burns JM et al (2009) Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. Mol Ecol Res 9(1):1–26
28. Kerr KCR, Lijtmear DA, Barriera AS, Hebert PDN, Tubaro PL (2009) Probing evolutionary patterns in neotropical birds through DNA barcodes. PLoS ONE 4:e4379
29. Smith MA, Fernandez-Triana J, Roughley R, Hebert PDN (2009) DNA barcode accumulation curves for understudied taxa and areas. Mol Ecol Res 9:08–216. https://doi.org/10.1111/j.1755-0998.2009.02646
30. Foottit RG, Maw HEL, von Dohlen CD, Hebert PDN (2008) Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. Mol Ecol Res 8:1189–1201
31. Hastings JM, Schultheis PJ, Whitson M et al (2008) DNA barcoding of new world cicada killers (Hymenoptera: Crabronidae). Zootaxa 1713:27–38
32. Hubert N, Hanner R, Holm E et al (2008) Identifying Canadian freshwater fishes through DNA barcodes. PLoS ONE 3:e2490
33. Hou ZE, Li Z, Li SQ (2009) Identifying Chinese species of *Gammarus* (Crustacea: Amphipoda) using DNA barcoding. Curr Zool 55:158–164
34. Wong E, Shivji MS, Hanner RH (2009) Identifying sharks with DNA barcodes: assessing the utility of a nucleotide diagnostic approach. Mol Ecol Res 9(1):243–256
35. Clare EL, Lim BK, Fenton MB, Hebert PDN (2011) Neotropical bats: estimating species diversity with DNA barcodes. PLoS ONE 6:e22648
36. Nneji LM, Adeola AC, Wang Y-Y et al (2020) Testing the effectiveness of DNA barcoding for biodiversity assessment of moths from Nigeria. Diversity 12:85
37. Nneji LM, Adeola AC, Okeyoyin A et al (2019) Assessing the effectiveness of molecular data for species identification and diversity studies of Nigerian herpetofauna. Mitochondrial DNA B. https://doi.org/10.1080/23802359.2018.1561219
38. Lukhtanov VA, Sourakov A, Zakharov EV, Hebert PDN (2009) DNA barcoding Central Asian butterflies: increasing geographical dimension does not significantly reduce the success of species identification. Mol Ecol Res 9:1302–1310
39. Dincă V, Zakharov EV, Hebert PDN, Vila R (2011) Complete barcode library for a country's butterfly fauna reveals high performance for temperate Europe. Philos Trans R Soc B 278:347–355
40. Gaikwad SS, Ghate HV, Ghaskadbi SS, Patole MS, Shouche YS (2012) DNA barcoding of nymphalid butterflies (Nymphalidae: Lepidoptera) from Western Ghats of India. Mol Biol Rep 39:2375–2383
41. Huemer P, Mutanen M, Sefc KM, Hebert PDN (2014) Testing DNA barcode performance in 1000 species of European Lepidoptera: large geographic distances have small genetic impacts. PLoS ONE 9(12):e115774
42. Hausmann A, Haszprunar G, Hebert PDN (2011) DNA barcoding the geometrid fauna of Bavaria (Lepidoptera): successes, surprises, and questions. PLoS ONE 6:e17134
43. Mutanen M, Hausmann A, Hebert PDN, Landry J-F, de Waard JR, Huemer P (2012) Allopatry as a Gordian knot for taxonomists: patterns of DNA barcode divergence in Arctic-Alpine Lepidoptera. PLoS ONE 7(10):e47214
44. Dincă V, Montagud S, Talavera G et al (2015) DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. Sci Rep 5:12395
45. Mutanen M, Kivelä SM, Vos RA, Doorenweerd C, Ratnasingham S, Hausmann A et al (2016) Species-level para- and polyphyly in DNA barcode gene trees: strong operational bias in European Lepidoptera. Syst Biol 65(6):1024–1040
46. Terborgh J (2002) Making parks work: strategies for preserving tropical nature. Island Press, Washington DC, p 65
47. Larsen TB (2005) Butterflies of West Africa. Apollo Books, Svendborg
48. Ivanova NV, Grainger CM (2007a) CCDB protocols, COI amplification. https://www.dnabarcoding.ca/CCDB-DOCS/CCDB_amplification.pdf. Accessed 15 July 2020
49. Ivanova NV, Grainger CM (2007b) CCDB protocols, sequencing. https://www.dnabarcoding.ca/CCDB_DOCS/CCDB_Sequencing.pdf. Accessed 15 July 2020
50. Ivanova NV, Grainger CM (2007c) CCDB protocols, primer sets. https://www.dnabarcoding.ca/CCDB_DOCS/CCDB_Amplification.pdf. Accessed 15 July 2020

51. Ivanova NV, DeWaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. Mol Ecol Notes 6:998–1002
52. Ivanova NV, Zemlak TS, Hanner RH, Hebert PDN (2007) Universal primer cocktails for fish DNA barcoding. Mol Ecol Notes 7:544–548
53. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S (2013) MEGA6: molecular evolutionary genetics analysis version 6.0. Mol Biol Evol 30(12):2725–2729
54. Strutzenberger P, Brehm G, Fiedler K (2011) DNA barcoding-based species delimitation increases species count of Eois (Geometridae) moths in a well-studied tropical mountain forest by up to 50%. Insect Sci 18:1744–7917
55. Kimura M (1980) A simple method for estimating evolutionary rate of base substitutions through comparative studies of nucleotide sequences. J Mol Evol 16:111–120
56. Kim M, Xinlong W, Kim JM et al (2010) Phylogenetic relationships of true butterflies (Lepidoptera: Papilionoidea) inferred from COI, 16S rRNA and EF-1α sequences. Mol Cells 30:409–425
57. Felsenstein J (1985) Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39:783–791
58. Posada D, Buckley TR (2004) Model selection and model averaging in phylogenetics: advantages of Akaike information criterion and Bayesian approaches over likelihood ratio tests. Syst Biol 53:793–808
59. Stamatakis A (2006) RAxML-VI-HPC: maximum likelihood-based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics 22:2688–2690
60. Ronquist F et al (2012) MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539–542
61. Miller MA, Pfeiffer W, Schwartz T (2012) The CIPRES science gateway: enabling high-impact science for phylogenetics researchers with limited resources. In: Proceedings of the 1st conference of the extreme science and engineering discovery environment: Bridging from the extreme to the campus and beyond. Association for Computing Machinery, USA, pp 1–8
62. Rambaut A (2012) Figtree v 1.4.0. https://tree.bio.ed.ac.uk/software/figtree/
63. Meier R, Kwong S, Vaidya G, Ng PKL (2006) DNA Barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success. Syst Biol 55:715–55728
64. Puillandre N, Lambert A, Brouillet S, Achaz G (2011) ABGD, Automatic barcode discovery for primary species delimitation. Mol Ecol 21:211864–211877
65. Saddhe AA, Jamdade RA, Kumar K (2017) Evaluation of multilocus marker efficacy for delineating mangrove species of West Coast India. PLoS ONE 12:e0183245
66. Nee S (2006) Birth-death models in macroevolution. Ann Rev Ecol Evol Syst 37:1–17
67. Gernhard T (2008) The conditioned reconstructed process. J Theor Biol 253(4):769–778
68. Drummond AJ, Rambaut A (2007) Beast: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol 7:214
69. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with beauti and the beast 1.7. Mol Biol Evol 29:1969–1973
70. Rambaut A, Drummond AJ (2009) Tracer version 1.5 [computer program]. https://beast.bio.ed.ac.uk
71. Rambaut A, Drummond AJ (2010) TreeAnnotator version 1.6.1 [computer program]. https://beast.bio.ed.ac.uk
72. Paradis E, Claude J, Strimmer K (2004) APE: analysis of phylogenetics and evolution in R language. Bioinformatics 20:289–290
73. Ezard T, Fujisawa Tand Barraclough TG (2009). splits: SPecies' LImits by Threshold Statistics. R package version. 1.0–14/r31. https://R-Forge.R-project.org/projects/splits/
74. Hendrich L, Morinière J, Haszprunar G, Hebert PDN, Hausmann A, Kohler F et al (2015) A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3,500 identified species to BOLD. Mol Ecol Res 15:795–818
75. Morinière J, Hendrich L, Hausmann A, Hebert P, Haszprunar G et al (2014) Barcoding Fauna Bavarica: 78% of the Neuropterida fauna barcoded! PLoS ONE 9(10):e109719
76. Lavinia PD, Nuñez Bustos E, Kopuchian C, Lijtmaer DA, Garcia NC, Hebert PDN, Tubaro PL (2017) Barcoding the butterflies of southern South America: species delimitation efficacy, cryptic diversity and geographic patterns of divergence. PLoS ONE 12:0186845
77. Yang M, Zhai Q, Yang Z, Zhang Y (2015) DNA barcoding Satyrine butterflies (Lepidoptera: Nymphalidae) in China. Mitochondrial DNA 27(4):2523–2528
78. Marín MA, Cadavid IC, Valdés L et al (2016) DNA barcoding of an Assembly of Montane Andean butterflies (Satyrinae): geographical scale and identification performance. Neotrop Entomol 46:514–523
79. Busvine JR (1980) Cryptic species of insect disease vectors and their importance. Endeavour 4:108–112
80. Della Torre A, Costantini C, Besansky NJ, Caccone A, Petrarca V, Powell JR, Coluzzi M (2002) Molecular and ecological aspects of incipient speciation within Anopheles gambiae: the glass is half full. Science 289:115–117
81. Clark AR, Armstrong KF, Carmichael AE, Milne JR, Raghu S, Roderick GK, Yeates DK (2005) Invasive phytophagous pests arising through a recent tropical evolutionary radiation: the Bactrocera dorsalis complex of fruit bies. Ann Rev Entomol 50:293–319
82. Ball SL, Armstrong KF (2006) DNA barcodes for insect pest identification: a test case with tussock moths (Lepidoptera: Lymantriidae). Can J For Res 36:337–350
83. Blagoev GA, Nikolova NI, Sobel CN, Hebert PDN, Adamowicz SJ (2013) Spiders (Araneae) of Churchill, Manitoba: DNA barcodes and morphology reveal high species diversity and new Canadian records. BMC Ecol 13:44–44
84. Raso L, Sint D, Rief A, Kaufman R, Traugott M (2014) Molecular identification of adult and juvenile Linyphiid and Theridiid spiders in Alpine Glacier Foreland communities. PLoS ONE 9:101755
85. Dona J, Diaz-Real J, Mironov S, Bazaga P, Serrano D, Jovani R (2015) DNA barcoding and mini barcoding as a powerful tool for feather mite studies. Mol Ecol Res 15:1216–1225
86. Xu X, Liu F, Chen J, Li D, Kuntner M (2015) Integrative taxonomy of the primitively segmented spider genus Ganthela (Araneae: Mesothelae: Liphistiidae): DNA barcoding gap agrees with morphology. Zool J Linn Soc 175:288–306
87. Ashfaq M, Ara J, Noor AR, Hebert PDN, Mansoor S (2011) Molecular phylogenetic analysis of a scale insect (Drosicha mangiferae; Hemiptera: Monophlebidae) infesting mango orchards in Pakistan. Eur J Entomol 108:553–559
88. Ashfaq M, Akhtar S, Khan AM, Adamowicz SJ, Hebert PDN (2013) DNA barcode analysis of butterfly species from Pakistan points towards regional endemism. Mol Ecol Resour 13(5):832–843
89. Carew ME, Pettigrove V, Cox RL, Hoffmann AA (2007) DNA identification of urban Tanytarsinichironomids (Diptera: Chironomidae). J N Am Benthol Soc 26:587–600
90. Williams MC (2020) Afrotropical butterflies. https://www.lepsocafrica.org/?p=publications&s=atb

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Lotanna Micah Nneji[1,2] · Adeniyi Charles Adeola[1,2] · Adeola Oluwakemi Ayoola[1] · Segun Olayinka Oladipo[3] · Yun-Yu Wang[1] · Yoila D. Malann[4] · Okorie Anyaele[5] · Ifeanyi Christopher Nneji[4] · Md Mizanur Rahman[1] · Caroline Samuel Olory[6]

1 State Key Laboratory of Genetic Resources and Evolution, Kunming Institute of Zoology, Chinese Academy of Sciences, Kunming 650223, Yunnan, China

2 Sino-Africa Joint Research Centre, Chinese Academy of Sciences, Kunming 650223, Yunnan, China

3 Department of Biosciences and Biotechnology, Kwara State University, Malete, Kwara, Nigeria

4 Department of Biological Science, University of Abuja, Abuja 902101, Federal Capital Territory, Nigeria

5 Department of Zoology, Faculty of Science, University of Ibadan, Ibadan 200, Oyo, Nigeria

6 Cross River National Park, Calabar 542, Cross River, Nigeria