HHS Public Access

Author manuscript

Trop Lepid Res. Author manuscript; available in PMC 2022 December 01.

Published in final edited form as:

Trop Lepid Res. 2022 December ; 32(2): 73–78. doi:10.5281/zenodo.7246139.

Genomic analysis reveals a new genus of Firetip skippers (Lepidoptera: Hesperiidae: Pyrrhopyginae)

Jing Zhang[1], **Jinhui Shen**[1], **Qian Cong**[2], **Geoff Martin**[3], **Nick V. Grishin**[1,4]

1. Departments of Biophysics, Biochemistry

2. Eugene McDermott Center For Human Growth & Development, University of Texas Southwestern Medical Center

3. Department of Life Sciences, Natural History Museum, London, SW7 5BD, UK.

4. Howard Hughes Medical Institute and 5323 Harry Hines Blvd, Dallas, TX, USA 75390-9050

Abstract

We obtained whole genome shotgun sequence reads for a number of Firetip skippers (subfamily Pyrrhopyginae), including all known species from the genera *Yanguna* Watson, 1893 and *Gunayan* Mielke, 2002 and representative species of *Pyrrhopyge* Hübner, [1819]. Phylogenetic analysis of their protein-coding regions unexpectedly revealed that *Yanguna tetricus* Bell, 1931 was not monophyletic with the other species of *Yanguna* (type species *Pyrrhopyga spatiosa* Hewitson, 1870). Instead, *Y. tetricus* formed a phylogenetic lineage as ancient as other three genera in its clade (*Pyrrhopyge*, *Yanguna* and *Gunayan*) that rapidly diversified from their ancestor. Therefore a new genus, *Guyanna* Grishin, **gen. n**. (type species *Yanguna tetricus*), is proposed for this lineage. The specimen that we sequenced was the *Y. tetricus* holotype in the Natural History Museum, London, leaving no doubt that we are dealing with this species. Genomic sequencing and comparison of specimens from museum collections offers a powerful strategy to reveal unforeseen phylogenetic relationships, and sequencing of primary types ensures that the conclusions are accurate in terms of nomenclature.

Keywords

biodiversity; butterflies; genomic sequencing; museomics; phylogeny

INTRODUCTION

A genus is a taxonomic category that applies to a group of species that are more closely related to each other than to other such groups. While the precise criteria to outline a genus are not defined, most taxonomists agree that genera should be monophyletic, include closely related species, and be similar to other genera in term of evolutionary divergence of species within a genus. Species within a genus share some common morphological or ecological

grishin@chop.swmed.edu .

features that are not present in other genera. These features, or “characters”, are used to delineate genera in animals. For instance, a long process on the sacculus of male genitalic valvae is apparently a derived character (synapomorphy) that was used by Mielke (2002) to distinguish a genus *Gunayan* he proposed for a group of Firetip skippers (subfamily Pyrrhopyginae).

With the advent of DNA-based phylogenies, it has become more straightforward to outline reliable monophyletic groups and to estimate the age of taxa. Nodes in the phylogenetic tree can be dated using fossil records, tree branch lengths rescaled, and this information can be used as an objective and consistent criterion to help delineate genera. For instance, an age of about 5 million years agreed well with genera as they were defined from morphology in a group of blue butterflies and was proposed as a guiding principle for the definition of genera (Talavera *et al.*, 2012). Genetic differentiation over a period of a few million years is likely to lead to the phenotypic divergence that has been used to outline genera by morphology. Understanding the correlation between genomic differences and phenotypic divergence is an emerging field of research.

To study the distinction between closely related genera, we have chosen the Firetip skipper butterflies (Hesperiidae: Pyrrhopyginae). Named for the prominent tuft of red or orange scales at the end of the abdomen present in many species of this subfamily, the Firetips are Neotropical in distribution. Only one species, *Apyrrothrix araxes* (Hewitson, 1867), reaches southern United States. The type genus, *Pyrrhopyge* Hübner, [1819], is marked by the reddish-orange tuft, but its closest relatives, *Yanguna* Watson, 1893 and *Gunayan* Mielke, 2002 have a dark or grayish abdomen end. To understand their relationships better, we obtained and analyzed genomic sequences of all known species of *Yanguna* and *Gunayan*, and representative species of *Pyrrhopyge*, along with several outgroup taxa. Phylogenetic trees obtained from these sequences revealed a surprise. While, as expected, the three genera indeed formed distinct phylogenetic clusters, *Yanguna tetricus* Bell, 1931 was not monophyletic with *Yanguna*, and stood out as its own ancient phylogenetic group on par with the others.

MATERIALS AND METHODS

A leg was removed from the *Y. tetricus* holotype (specimen number NHMUK012824232 in the Natural History Museum, London, United Kingdom, NHMUK) using fine tweezers and placed in a plastic tube, being assigned the molecular code NHMUK0247272661, and later processing code NVG-18083F05 (Fig. 1). DNA was extracted from the leg non-destructively using Macherey-Nagel (MN) reagents; 70 µl buffer T1 and 10 µl protK were simply added to the tube without crushing the leg, and the mixture was incubated at 57°C for 24 hours. Then, 80 µl buffer B3 was added and incubation continued for 2 hours, after which 85 µl of absolute EtOH was added and thoroughly mixed. The resulting liquid was transferred to a different tube and DNA extraction continued according to MN protocol (Li *et al.*, 2019), leaving the leg intact. This procedure resulted in about 14 ng of genomic DNA of *Y. tetricus* dissolved in 35 µl. About 70% of the DNA extract was used to construct a mate-pair library according to our published protocols (Cong *et al.*, 2017). The library was sequenced for 150 bp from both ends targeting 6 Gbp of data on Illumina HiSeq x10

at GENEWIZ. The resulting reads were matched using Diamond (Buchfink *et al.*, 2015) to the exons of the reference genome of *Cecropterus lyciades* (Shen *et al.*, 2017) that we obtained previously, and exons assembled and aligned to other Hesperiidae genomes we obtained using the same methods. Coding regions of mitochondrial genome (including the COI barcode) were assembled similarly. For the COI barcode, due to its frequent use, a specialized procedure was developed (Li *et al.*, 2019), where all sequence reads matching this region were filtered for contamination, aligned, and this lineup inspected manually to check for possible errors. Exons expected to be from the Z chromosome were predicted assuming similar syntenic arrangement with *Heliconius* Kluk, 1780 (Heliconius Genome Consortium, 2012). Phylogenetic trees were generated from 4 sets of exons: whole nuclear genome, whole mitochondrial genome and Z-chromosome using RAxML-NG (Kozlov *et al.*, 2019) with default parameters (-m GTRGAMMA). PhyML (Guindon *et al.*, 2010) was used to construct the COI barcode tree. For additional technical details of experimental and computational protocols we refer to the SI Appendix of our recently published study (Li *et al.*, 2019). The data used in this project were deposited in the NCBI database <https://www.ncbi.nlm.nih.gov/> as BioProject PRJNA889040, and the COI barcode sequence of the *Y. tetricus* holotype in GenBank with accession ON480167. The supplementary file is deposited at <https://osf.io/n934q/>.

We assembled protein-coding regions from the whole genome shotgun reads of 36 Pyrrhopyginae species that we have sequenced. See Table S1 in supplementary file <https://osf.io/n934q/> for detailed specimen data in addition to brief information about the specimens shown in trees in Fig. 1. These species included all 6 known *Yanguna*, 3 *Gunayan* species and 21 representative *Pyrrhopyge* (out of 42 known species) (Mielke, 2005). In addition, 5 more distant Pyrrhopyginae genera were selected from close relatives of *Pyrrhopyge* as outgroups (Fig. 1b), and the entire tree was further rooted with *Mysoria* (*Sarbia*) *xanthippe spixii* (Plötz, 1879). The study included two holotypes from the NHMUK: *Yanguna tetricus* Bell, 1931 (see Methods above) and *Yanguna timaeus* Bell, 1931 (currently in *Gunayan*, specimen number NHMUK012824233, molecular code 0247278334, processing code NVG-18083F06). The lengths of resulting genomic regions were: nuclear total 7,481,792+/−3,155,310, Z-chromosome 305,487+/−130,366; mitogenomes 10,403+/−296. We considered Z-chromosome separately. Males of butterflies have two copies of Z, while females have Z and W. In Z, recombination is reduced to half of that in autosomes, and sexual selection acts differently on genes encoded by it. Thus a separate analysis of genes encoded by Z chromosome may offer additional insights about evolution of these species. Phylogenetic trees were constructed from coding regions of nuclear genome, Z chromosome and mitogenome. In addition, a COI barcode region dendrogram was also obtained.

RESULTS

All the trees lead to the same conclusions (Fig. 1b). First, the branch separating the ingroups from outgroups is the longest internal branch in the trees, suggesting phylogenetic closeness of the three ingroup genera (*Pyrrhopyge*, *Yanguna* and *Gunayan*). Second, each of the three genera forms a compact cluster well-separated from the others, with one exception. Third, the exception, *Yanguna tetricus*, is not monophyletic with other *Yanguna* species. Instead,

while being a member of this group of relatives, it was well-removed from all ingroup Pyrrhopyginae taxa. This is an unexpected and interesting result. We sequenced the type species of all three genera: *Pyrrhopyge phidias bixae* (Linnaeus, 1758), *Yanguna spatiosa* (Hewitson, 1870), and *Gunayan rhacia* (Hewitson, 1875), and *Y. tetricus* didn't group with any of them. Therefore, it does not belong to these genera as currently defined, and to restore monophyly of *Yanguna*, a new genus is proposed for *Y. tetricus*.

Guyanna Grishin, **gen. n**.

Zoobank registered: https://zoobank.org/80EBCB78-D65A-4317-AD1A-0994C6DEB640

Type species: *Yanguna tetricus* Bell, 1931

Diagnosis.

Morphologically similar to *Yanguna* (see diagnosis in Mielke (2002)) but differs from it by the shortened tip of the aedeagus after the lobule (Mielke, 2002), an apparent synapomorphy of the new genus. Additionally, the costa-ampulla of the right valva is distally squared (as in many *Pyrrhopyge* species) rather than triangular (as in *Yanguna*) and the central tooth of the right harpe is broad and directed dorsad rather than dorsocaudad (Fig. 2, which is a reproduction of fig. 14 from Bell (1934)). The genus keys to A.1.49 in Evans (1951), and is distinguished from other genera of Pyrrhopyginae by the combination of the following characters: head with white lines and dots; abdomen black, striped gray; end of abdomen brown or grayish (not red or orange); legs black, only forecoxae orange, without white scales; fringes not checkered; male genitalic valvae asymmetric, sacculus without a long process that is present in *Gunayan* (see figs. 11–12 in (Bell 1934)). In nuclear DNA, a combination of the following base pairs is diagnostic (see supplementary file <https://osf.io/n934q/> for the sequences with these characters). This character unifies *Guyanna* with *Pyrrhopyge* and excludes others: aly728.44.1:G672C. The meaning is: position 672 in exon 1 of the gene 44 from the scaffold 728 of *Cecropterus lyciades* [formerly in *Achalarus*] (aly) reference genome (Shen *et al.*, 2017) is C, changed from G in the ancestor. These characters separate *Pyrrhopyge* from others (character state in *Pyrrhopyge* is given after "not"): aly5196.6.2:T61T (not A), aly536.33.3:T477T (not C), aly770.15.10:T48T (not C). These characters separate *Gunayan* from others (character state in *Gunayan* is given after "not"): aly9588.6.1:G777G (not A), aly1244.1.2:T276T (not C), aly393.2.2:C91C (not T). These characters separate *Yanguna* from others (character state in *Gunayan* is given after "not"): aly27.7.1:A504A (not G), aly5729.9.1:C189C (not T), aly594.11.4:C288C (not T). These characters unify *Guyanna*, *Pyrrhopyge*, *Gunayan*, and *Yanguna* and exclude others: aly1539.14.4:A159T, aly1018.6.2:A267G, aly2178.13.1:T351C. This complex combination of characters was chosen to mitigate possible negative consequences of diagnosing a taxon without close relatives (*G. tetricus*) and to distinguish between synapomorphy and error or unique base pair in this particular specimen. Thus, we looked for possible synapomorphic characters in all other, more speciose clades and define the new genus by the lack of these synapomorphies combined with the synapomorphies for the clades that include it. In the COI barcode region, the following characters are diagnostic in combination: T56T (not A) (the meaning is: position 56 in the barcode is T, and is T in the ancestor, but not A as

in other taxa), C81T, 82A (not C or T) (ancestral state could not be confidently deduced), G86G (not A), A214T, C287C (not T), T289T (not A), T319T (not A), G474A, T607T (not A).

Derivation of the name.

The name is a feminine noun in the nominative singular. Similarly to *Gunayan*, it is an anagram of *Yanguna*. These three genera plus *Pyrrhopyge* are close relatives and form a prominent clade in the phylogenetic tree. The anagram is also a hint to the type locality of the type species (Roraima) suggested by Evans (1951: 34) as being in Guyana, the assessment followed by Mielke (2005). It should be noted that just “Roraima” present on the specimen label is not sufficient by itself to deduce Guyana (British Guiana at that time) as the type locality country, since prior to 1900 (Fig.1a, Crowley died in 1900) “Roraima” could also have referred to a locality in modern day Brazil or Venezuela.

Species included: Only the type species.

DISCUSSION

In the absence of DNA sequences it is not readily apparent that *Yanguna tetricus* is significantly more distinct from its former congeners than they are from one another. Indeed, its general genitalic morphology, body and wing patterns, and wing venation and shape, place it among *Yanguna* as originally described (Mielke, 2002). Nevertheless, Mielke (2002) noticed the uniqueness of this species in the shape of the aedeagus tip, the only exceptional character in *Yanguna* that he mentioned. However, it is hardly possible to put a time-scale on such a genitalic difference and predict when the *Y. tetricus* lineage split from others. DNA information is indispensible to reveal the magnitude of the distinction between *Y. tetricus* and other members of the *Pyrrhopyge* clade (*Pyrrhopyge*, *Yanguna*, and *Gunayan*). In fact, the genetic differentiation of *Y. tetricus* is sufficiently profound that the phylogenetic relationships among the four genera (*Guyanna* gen. n., *Yanguna*, *Gunayan*, and *Pyrrhopyge*) remain unresolved and the relative order of their divergence from their ancestors is not clear (Fig. 1b). This lack of resolution is a result of these four genera being nearly equidistant from one another in genetic differentiation, with all four genera apparently diverging at about the same time, nearly 10 Mya (Li *et al.*, 2019; Sahoo *et al.*, 2017; Zhang *et al.*, 2019). Generic splits around that time have been suggested for other groups of animals, for instance, humans versus chimpanzees (Kumar *et al.*, 2005; Moorjani *et al.*, 2016). Thus the generic status for these four groups seems reasonable and is consistent with how genera are defined in other animals.

Even if dating divergence events has significant errors, unscaled and not-dated trees with branch lengths proportional to the number of base pair substitutions along them (Fig. 1b) suggest the same division into four groups (blue, red, green and magenta). These four clades together form a monophyletic group strongly supported in all trees (100% of partitions have this group). Each non-singular clade individually is a strongly supported monophyletic group: 100% of partitions support blue (*Pyrrhopyge*), green (*Gunayan*), and magenta (*Yanguna*) clades in all trees. However, the mutual arrangement of these clades differs among the trees and is more weakly supported. Therefore, we see two confident

evolutionary levels in the trees. The first level unifies these four groups, corresponding to rapid diversification of the clade into these four groups, possibly complicated by incomplete lineage sorting and gene exchange through hybridization. These events and rapid diversification obscure phylogenetic relationships between the four groups, revealing a conflict between the gene trees that results in low statistical support for any topology. The second level is the diversification of each of these four groups into species. We do not observe a most confidently supported level in between these two. Therefore, we either treat the assemblage of the four groups as a single genus (*Pyrrhopyge s. l.*), or divide it into four genera, the solution adopted here, following recent treatments (Mielke, 2002; Zhang *et al.*, 2019). A solution that divides *Pyrrhopyge* sensu lato into two or three groups seems less meaningful, because such groups may not be monophyletic (due to lower statistical support) and they are supported by relatively shorter branches (fewer changes in genomic DNA). For example, all trees suggest that *Y. tetricus* is monophyletic with *Pyrrhopyge*, and while this relationships may be correct, it is not obvious from morphology: the common branch of the two groups is short and may not have genomic changes that encode significant morphological changes.

When sequencing a unique and old specimen, a question about data quality and its possible negative influence on the conclusions of the study comes up. For instance, DNA degradation and contamination may affect sequence quality and lead to faulty phylogenetic analyses. Indeed, DNA degradation results in miscalling of base pairs in individual sequence reads. However, due to the random nature of these events, higher coverage of genomic regions by sequence reads removes these individual and random errors in each read. For low coverage regions, these random errors would not correlate with phylogeny, and will mostly contribute to the length of the terminal branch of this specimen, because these incorrect base pairs will be mostly unique to this specimen. While we observed this effect for old specimens in a number of projects, the terminal branch of *Guyanna tetricus* comb. n. is not longer than most others (Fig. 1b), suggesting that its sequence is of a reasonable quality. Moreover, trees constructed from different genomic regions, nuclear and mitochondrial, result in the same conclusion. We also analyzed the COI barcodes, and they constitute one of the regions best-covered by sequence reads. The reads mapping to the barcode were inspected manually for consistency and overlap. The dendrogram from the barcodes placed the four taxa in the same topology, further supporting our conclusions. Furthermore, even if these sequencing results are flawed for some unknown reason, they pointed to the morphological uniqueness of *G. tetricus* comb. n. that was already noticed in the original description of the genus *Yanguna* by Mielke (2002). The structure of the aedeagus tip is likely synapomorphic and can be used to differentiate between the four taxa (*Guyanna* gen. n., *Yanguna*, *Gunayan*, and *Pyrrhopyge*), as was suggested by Mielke (2002); he proposed names for the two genera (*Yanguna*, *Gunayan*), and mentioned the third (*Guyanna* gen. n.), including it in *Yanguna* as the only exception to the state of the character of the genus in which he placed it.

Finally, we note that it is optimal to sequence primary type specimens, which ensures that we are analyzing the correct species and there is no misidentification involved. In sum, DNA analysis has been instrumental to discover that *Yanguna tetricus* does not belong in *Yanguna* and, as hinted by morphological evidence, should be placed in the new genus that we named here.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

We are grateful to Robert K. Robbins and Brian Harris (National Museum of Natural History, Smithsonian Institution, Washington DC), Wolfram Mey and Viola Richter (Museum für Naturkunde, Berlin, Germany), David A. Grimaldi and Courtney Richenbacher (American Museum of Natural History, New York, NY, USA), John Rawlins (Carnegie Museum of Natural History), John R. MacDonald and Richard L. Brown (Mississippi Entomological Museum), for facilitating access to the collections under their care and stimulating discussions; to Bernard Hermier for many enlightening discussions and numerous suggestions; and anonymous reviewers for helpful comments. Special thanks to Olaf H. H. Mielke and Carlos Mielke for discussions, comments, and sampling specimens for DNA analysis. We acknowledge the Texas Advanced Computing Center (TACC) at The University of Texas at Austin (http://www.tacc.utexas.edu) for providing invaluable HPC resources that were essential to carry out this study, which has been supported by grants (to NVG) from the National Institutes of Health GM127390 and the Welch Foundation I-1505.

LITERATURE CITED

Bell EL 1934. Studies in the Pyrrhopyginae, with descriptions of new species (Lepidoptera, Rhopalocera, Hesperiidae). Journal of the New York Entomological Society 41: 481–529.

Buchfink B, Xie C, Huson DH 2015. Fast and sensitive protein alignment using DIAMOND. Nature Methods 12: 59–60. [PubMed: 25402007]

Cong Q, Shen J, Borek D, Robbins RK, Opler PA, Otwinowski Z, Grishin NV (2017) When COI barcodes deceive: complete genomes reveal introgression in hairstreaks. *Proceedings of the Royal Society B*: Biological Sciences 284: 20161735.

Evans WH 1951. A catalogue of the American Hesperiidae indicating the classification and nomenclature adopted in the British Museum (Natural History). Part I. Introduction and Group A Pyrrhopyginae. British Museum (Natural History), London, x + 92 pp., pls. 91–99. pp.

Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O.2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic Biology 59: 307–321. [PubMed: 20525638]

Heliconius Genome Consortium. 2012. Butterfly genome reveals promiscuous exchange of mimicry adaptations among species. Nature 487: 94–98 [PubMed: 22722851]

Kozlov A, Darriba D, Flouri T, Morel B, Stamatakis A.2019. RAxML-NG: A fast, scalable, and user-friendly tool for maximum likelihood phylogenetic inference. Bioinformatics 35: 4453–4455. [PubMed: 31070718]

Kumar S, Filipski A, Swarna V, Walker A, Hedges SB 2005. Placing confidence limits on the molecular age of the human-chimpanzee divergence. Proceedings of the National Academy of Sciences USA 102: 18842–18847.

Li W, Cong Q, Shen J, Zhang J, Hallwachs W, Janzen DH, Grishin NV 2019. Genomes of skipper butterflies reveal extensive convergence of wing patterns. Proceedings of the National Academy of Sciences USA 116: 6232–6237.

Mielke OHH 2002. Pyrrhopyginae: gêneros novos e revalidados (Lepidoptera: Hesperiidae). Revista brasileira de Zoologia 19: 217–228.

Mielke OHH 2005. Catalogue of the American Hesperioidea: Hesperiidae (Lepidoptera). Curitiba, Sociedade Brasileira de Zoologia. xiii + 1536 pp.

Moorjani P, Amorim CE, Arndt PF, Przeworski M.2016. Variation in the molecular clock of primates. Proceedings of the National Academy of Sciences USA 113: 10607–10612.

Sahoo RK, Warren AD, Collins SC, Kodandaramaiah U.2017. Hostplant change and paleoclimatic events explain diversification shifts in skipper butterflies (Family: Hesperiidae). BMC Evolutionary Biology 17: 174. [PubMed: 28768477]

Shen J, Cong Q, Borek D, Otwinowski Z, Grishin NV 2017. Complete genome of Achalarus lyciades, the first representative of the Eudaminae subfamily of skippers. Current Genomics 18: 366–374. [PubMed: 29081692]

Talavera G, Lukhtanov VA, Pierce NE, Vila R.2012. Establishing criteria for higher-level classification using molecular data: the systematics of Polyommatus blue butterflies (Lepidoptera, Lycaenidae). Cladistics 29: 166–192. [PubMed: 34818826]

Zhang J, Cong Q, Shen J, Brockmann E, Grishin NV 2019. Genomes reveal drastic and recurrent phenotypic divergence in firetip skipper butterflies (Hesperiidae: Pyrrhopyginae). Proceedings of the Royal Society B: Biological Sciences 286: 20190609.

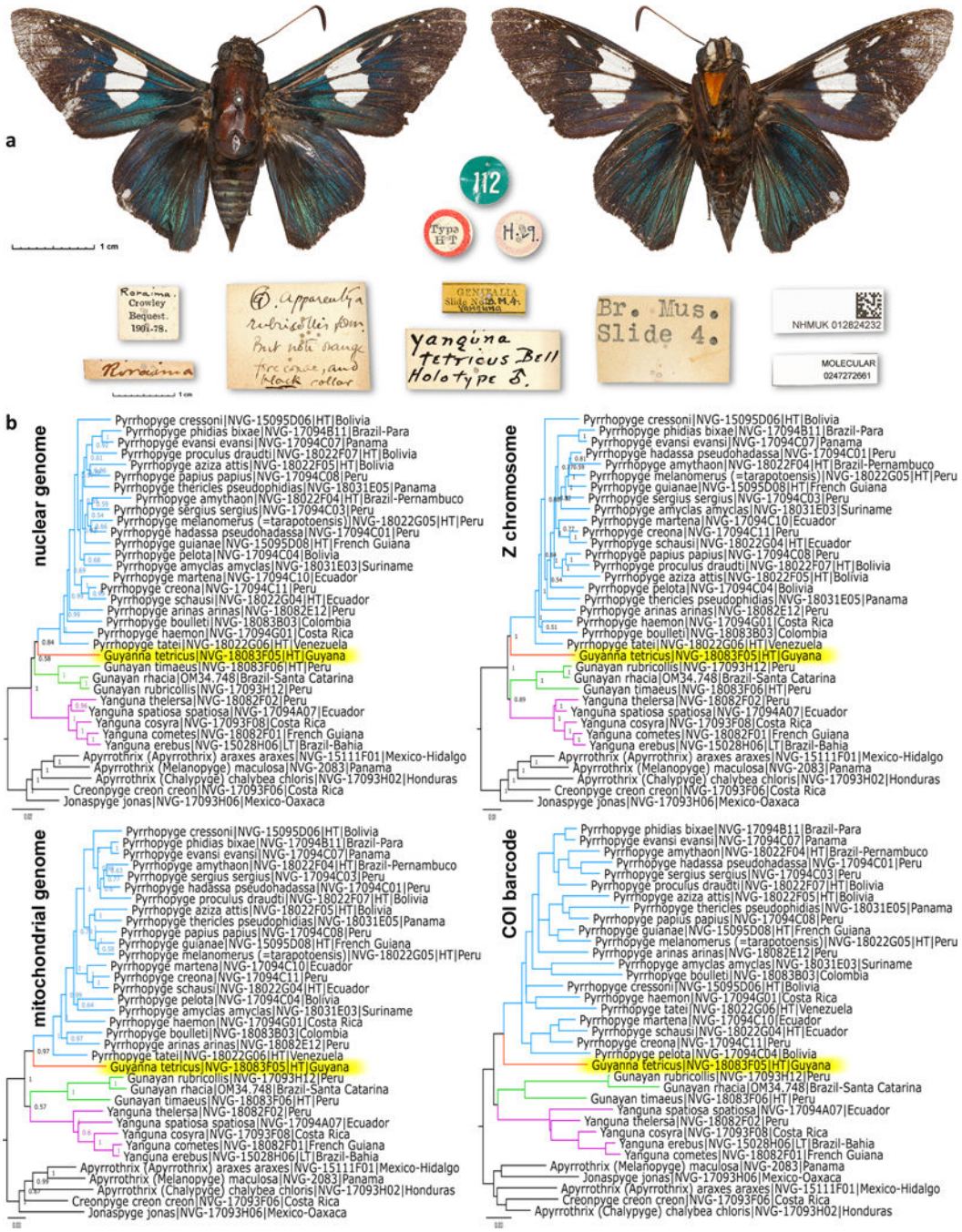

Figure 1.

Guyanna tetricus: (a) holotype in dorsal (left) and ventral (right) views with its labels (below), both sides of the round “Type” label are displayed, labels are shown at 3/4 of the specimen scale; (b) phylogenetic trees constructed from various portions of nuclear and mitochondrial genomes. The trees were rooted with *Mysoria (Sarbia) xanthippe spixii* (NVG-17094E01), not shown and would refer to the stub-like root in the trees.

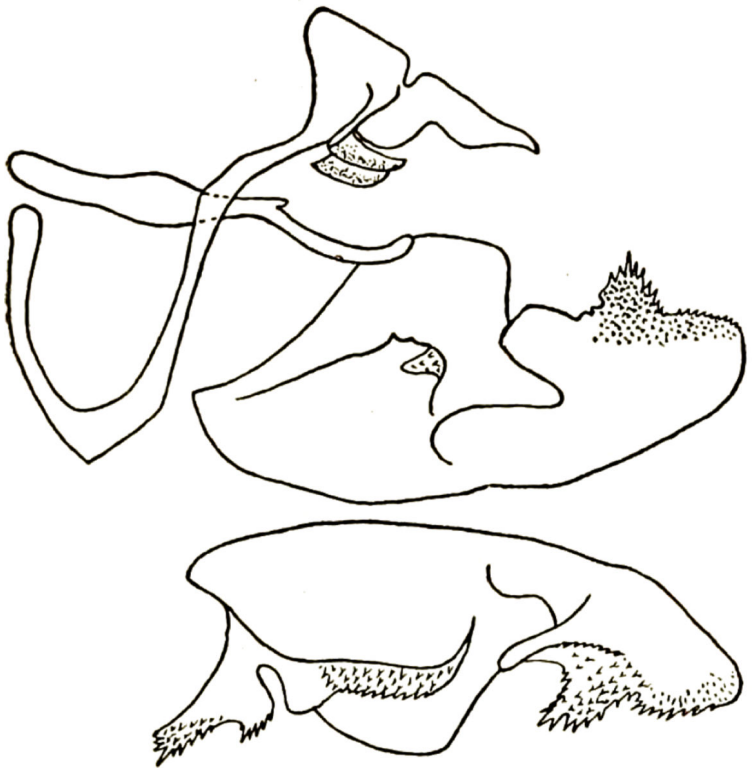

Figure 2.

Genitalia illustration of the holotype of *Guyanna tetricus* reproduced from Bell (1934), Fig. 14, downloaded from <https://www.biodiversitylibrary.org/item/205826#page/549/> and modified to remove colors. Interior view of left valva rotated 180° along horizontal axis (dorsal side down, “opened book” view of both valvae) from its position within the genital capsule shown above, is shown below.