



Genomes of skipper butterflies reveal extensive convergence of wing patterns

Wenlin Li^{a,b}, Qian Cong^{a,b,1}, Jinhui Shen^{a,b}, Jing Zhang^{a,b}, Winnie Hallwachs^c, Daniel H. Janzen^{c,2}, and Nick V. Grishin^{a,b,d,2}

^aDepartment of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75390-9050; ^bDepartment of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390-9050; ^cDepartment of Biology, University of Pennsylvania, Philadelphia, PA 19104; and ^dHoward Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390-9050

Contributed by Daniel H. Janzen, January 4, 2019 (sent for review December 17, 2018; reviewed by John Shuey and James Walters)

For centuries, biologists have used phenotypes to infer evolution. For decades, a handful of gene markers have given us a glimpse of the genotype to combine with phenotypic traits. Today, we can sequence entire genomes from hundreds of species and gain yet closer scrutiny. To illustrate the power of genomics, we have chosen skipper butterflies (Hesperiidae). The genomes of 250 representative species of skippers reveal rampant inconsistencies between their current classification and a genome-based phylogeny. We use a dated genomic tree to define tribes (six new) and subtribes (six new), to overhaul genera (nine new) and subgenera (three new), and to display convergence in wing patterns that fooled researchers for decades. We find that many skippers with similar appearance are distantly related, and several skippers with distinct morphology are close relatives. These conclusions are strongly supported by different genomic regions and are consistent with some morphological traits. Our work is a forerunner to genomic biology shaping biodiversity research.

genomics | taxonomy | higher classification | mimicry | Lepidoptera

It is the phenotype that is relevant when an animal interacts with its environment. Phenotypic traits have defined the place of an organism within its ecosystem and within a human-made classification system. Due to phenotypes being riddled by adaptive convergence, it is challenging to decipher phylogeny, and thus deduce phylogenetic classification from morphology. The phenotype is encoded by the genotype, which equally bears the footprint of evolution. DNA sequences are prime for evolutionary studies. Sanger-sequencing of select gene markers revolutionized phylogenetic research and refined morphology-based classification. However, frequent homoplasies are a serious obstacle to phylogeny reconstruction from small datasets. Since genomes are composed of millions of base pairs, the study of complete genotypes is expected to resolve many outstanding questions. While large-scale genomic studies are still scarce, they are most enlightening (1).

To exemplify a large-scale genomics project possible within a small laboratory, we use skipper butterflies (Hesperiidae). This family of about 4,000 species worldwide received less attention than other butterflies. Only a handful of studies have focused on skipper phylogeny (2–6). Pioneering work by Warren et al. (2, 3) revealed many surprising phylogenetic relationships compared with the morphological treatment (7–12). However, many questions remained unanswered. In particular, the subfamily Eudaminae was not analyzed in detail.

We obtained and analyzed genomic shotgun reads of 250 species of skippers covering all Eudaminae genera. We constructed a genome-level phylogeny of Hesperiidae and found that, due to convergence in wing patterns, many skippers have been placed in presumed evolutionary groups in which they do not belong. We also discovered a group of close relatives with disparate morphology. Furthermore, sequence data suggest possible genomic determinants of morphological traits.

Results and Discussion

Genomes of 250 Skippers. We selected 250 species of Hesperiidae from all subfamilies and tribes (*SI Appendix, Table S1*). Only 36

specimens were collected after 2012 and preserved in a DNA-friendly manner. Others were pinned in collections listed in *SI Appendix, Table S1* (also see *Acknowledgments*) and included specimens gathered over a century ago. Nevertheless, sequence reads targeting 10× coverage of the genome allowed us to assemble genomic regions by emphasizing protein-coding genes. We used 12,618 nuclear genes from the two reference genomes of Hesperiidae we have sequenced (13, 14). Aligned nuclear genomic regions covered 8,100,834 ± 2,387,641 positions in each specimen. Mitogenomes covered 10,663 ± 674 positions.

Genomic Tree of Skippers. We constructed genomic trees from concatenated alignments of coding regions (Fig. 1) and from coding regions with introns, and both trees had identical topologies. Furthermore, we constructed gene trees and combined them using the program ASTRAL (15), which produced the tree with the same topology. The trees were dated, i.e., transformed to predict absolute ages for its nodes (*SI Appendix, Methods*). The major phylogenetic groups in these trees agree with those defined by Warren et al. (2, 3) and follow-up publications that used larger sets of species and genes (4–6). Subfamily Coeliadinae (the awls) is sister to all other Hesperiidae (Fig. 1). Australian endemic *Euschemon*, as suggested by several studies (2, 3, 16), is unique and forms a

Significance

With decreasing cost, biodiversity genomics can be done by small laboratories, not large centers. Here, we overview the genomic landscape of an entire family of animals. By sequencing 250 representatives of skipper butterflies, we infer their phylogeny and search for genotypic determinants of phenotypic traits. We find that wing patterns are frequently convergent. This likely mimetic convergence is diversified, resulting in five distinct parallel patterns. Each of the five patterns occurs within at least two genera as well as in more distant relatives diverged more than 20 Mya. This work offers an example of how each of the nearly 6,000 families of animals will be studied in the future.

Author contributions: D.H.J. and N.V.G. designed research; W.L., Q.C., J.S., J.Z., W.H., and N.V.G. performed research; W.L., Q.C., J.Z., D.H.J., and N.V.G. analyzed data; and W.L., Q.C., J.S., J.Z., W.H., D.H.J., and N.V.G. wrote the paper.

Reviewers: J.S., The Nature Conservancy; and J.W., Kansas University.

The authors declare no conflict of interest.

This open access article is distributed under [Creative Commons Attribution-NonCommercial-NoDerivatives License 4.0 \(CC BY-NC-ND\)](https://creativecommons.org/licenses/by-nc-nd/4.0/).

Data deposition: The sequences reported in this paper have been deposited in the NCBI Sequence Read Archive (accession no. [SRP147939](https://www.ncbi.nlm.nih.gov/sra/SRP147939)) and the GenBank database (accession nos. [MH357724-MH357835](https://www.ncbi.nlm.nih.gov/nuclseq/MH357724-MH357835)). This work was registered with ZooBank (www.zoobank.org) as [8ABE82C4-AC8A-4B7B-8DF4-6BFF4BD48F3D](https://zoobank.org/8ABE82C4-AC8A-4B7B-8DF4-6BFF4BD48F3D), and registration numbers for 24 nomenclatural acts are given in Tables 1 and 2.

¹ Present address: Institute for Protein Design and Department of Biochemistry, University of Washington, Seattle, WA 98195.

² To whom correspondence may be addressed. Email: grishin@chop.swmed.edu or djanzen@sas.upenn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1821304116/-DCSupplemental.

Published online March 15, 2019.

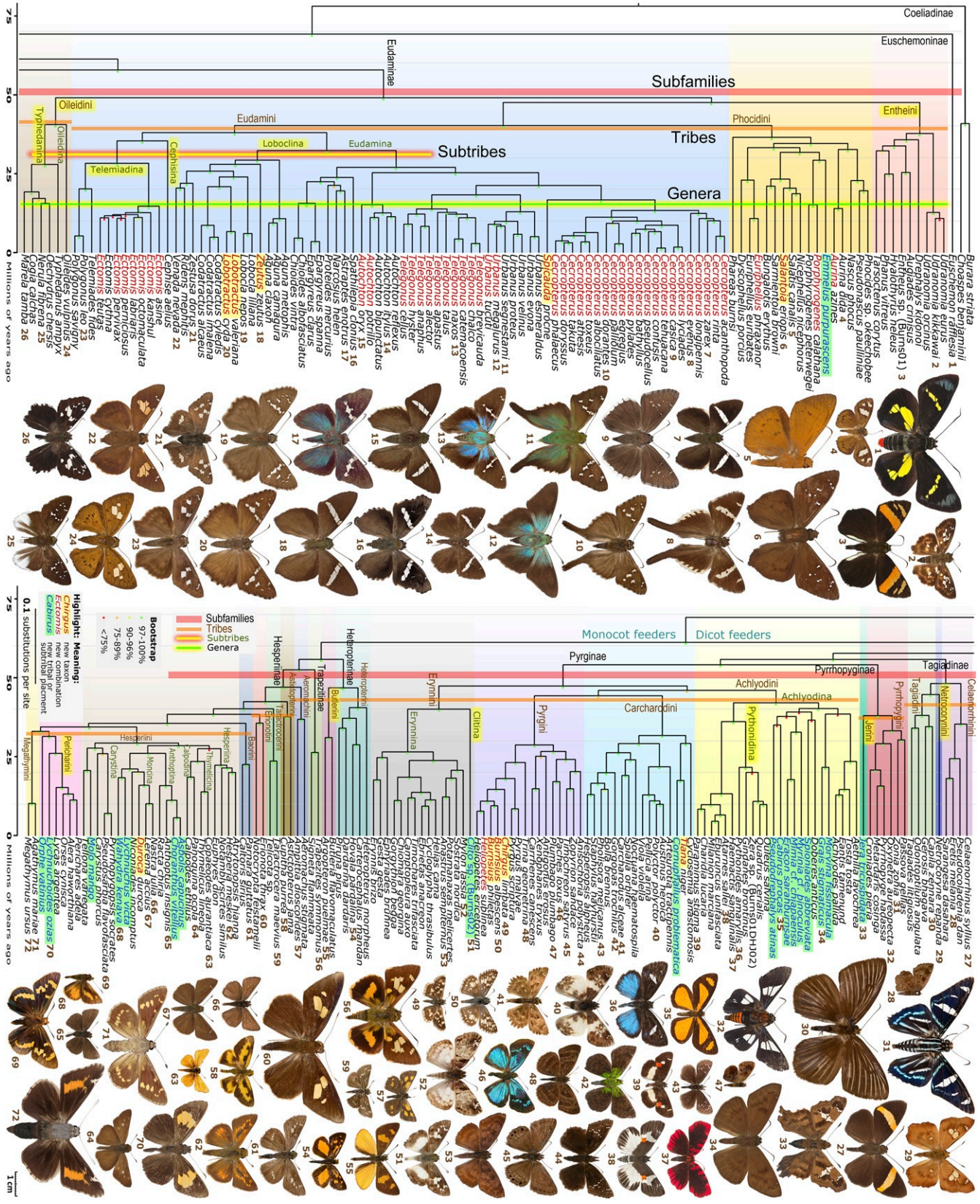


Fig. 1. Time-calibrated genomic tree of HesperIIDae constructed from nuclear gene coding regions of 250 species. The tree is cut in half: First and second segments are shown above and below. Colored horizontal lines delineate major taxonomic groups: red, subfamilies; orange (broken in several places), tribes; yellow shaded red (Eudaminae only), subtribes; and green shaded yellow (Eudaminae only), genera. New taxa are highlighted in yellow. New genera are in red font highlighted in yellow, and those used in different genus-species combinations than previously are in blue font highlighted in green. Segments of the tree corresponding to different tribes are highlighted in different colors. Illustrated HesperIIDae species are associated with names in the tree by numbers in brown font. See *SI Appendix* for details.

subfamily of its own, sister to all other Hesperidae except Coeliadinae. The next split in the genomic tree is different from that appearing in a number of DNA studies, but in agreement with the morphological view. The Spreadwing skippers (mostly dicot feeders) and the Grass skippers (mostly monocot feeders) are sisters. Both Spreadwing and Grass skippers have been previously divided into several subfamilies. Genomic data very strongly support (bootstrap above 99%) monophyly of these subfamilies. The two latest diverging subfamilies (Hesperinae and Trapezitinae) have split off about 50 Mya (Fig. 1). Pyrginae, as defined by Warren et al. (3), diversified from their common ancestor before that time (>55 Mya). Moreover, while other subfamilies except Pyrginae (i.e., Eudaminae, Heteropterinae, and Hesperinae) are well separated from each other, Pyrginae sensu Warren have split into three compact groups, within a short time and before divergence of Grass skippers into subfamilies. Thus, these three groups within Warren's Pyrginae are no less distinct than Hesperidae subfamilies, and we treat them as such. Accordingly, we unify the sister tribes Tagiadini and Celaenorrhini in a subfamily Tagiadininae, and reinstate the subfamily Pyrrhopyginae. The latter are strikingly distinct in appearance from other skippers, and have been traditionally considered a subfamily. They diverged from their common ancestor with Pyrginae about 55 Mya, before the divergence of Grass skippers. Thus, genomic data suggest that the family Hesperidae consists of nine subfamilies (Fig. 1 and *SI Appendix*), and diversification into subfamilies occurred about 50 Mya.

Mitogenomes and Cytochrome Oxidase I Barcodes. In addition to a nuclear genome tree, we constructed a tree from mitogenomes. The resulting tree recapitulates major phylogenetic groups of the nuclear genome tree, but with weaker support. All of the subfamilies and tribes are composed of the same species in either mitogenome or nuclear genome phylogenies. Next, using mitogenomes of 250 species as a backbone, we increased taxonomic coverage by adding 290 species with cytochrome oxidase I (COI) barcodes only (*SI Appendix*, Fig. S1). These species grouped with mitogenomes of their expected relatives. Most of these barcode-only species were placed according to

their current classification, with several exceptions discussed in *SI Appendix*. We used this mitogenome + barcode tree together with the nuclear genome tree (Fig. 1) as the basis for our proposed classification of Hesperidae (*SI Appendix*, *Taxonomic Appendix*). Differences in barcodes and morphology suggested that 30 subspecies as defined by Evans (10–12) are more likely to be species, and such cases were analyzed in detail for wing pattern and genital differences (*SI Appendix*, *Taxonomic Appendix*).

Eudaminae Tribes and Subtribes. Previous studies refrained from defining tribes in the subfamily Eudaminae. We focused on this subfamily and attempted to delineate tribes consistently with the age of the tribes in Pyrginae (sensu stricto). The four Pyrginae tribes (Carcharodini, Achlyodini, Erynnini, and Pyrgini) outlined by Warren et al. (3) diverged around 42 Mya. Our genomic tree agrees with Warren's, except that it places *Grais*, *Tosta*, *Morvina*, *Myrnia*, *Xispia*, *Pseudodrephalys*, *Mimia*, *Eracon*, and *Spioniades* in Achlyodini; *Cornuphagus* in Carcharodini; *Clito* in Erynnini; and *Jera* in Pyrrhopyginae. Unexpectedly, *Emmelus* belongs to Eudaminae. Conversely, *Cabirus* is not Eudaminae (4), but Pyrginae: Achlyodini. All Pyrginae tribes received 100% bootstrap support and are conspicuous groups in the subfamily.

To achieve consistent classification, we cut the dated genomic tree around the time of Pyrginae divergence into tribes and find four Eudaminae phylogenetic groups supported by 100% bootstrap that we define as tribes. Two of the tribes that form best-separated groups are described in Table 1 as Oileidini and Entheini. The two others are closely related sisters Eudamini and Phocidini that diverged about 40 Mya, and are given a tribal rank due to their morphological distinction.

These four Eudaminae tribes correspond to groups with similar morphology (10, 11). For example, Entheini is largely the “B. Augiades group” of Evans. He defined it by the “divergent” third segment of palpi that stems from the outer edge of the second segment. Inconsistently, Evans included in this group three genera that have the central third segment: *Phocides* and *Hypocryptothrix*, that do not belong to Entheini, and *Cabirus*, that is not even in Eudaminae (Fig. 1). Interestingly,

Table 1. Description of new tribes and subtribes of Hesperidae

New tribe or subtribe	Type genus	Diagnostic characters*	ZooBank ID†
Entheini Grishin, new tribe	<i>Entheus</i> Hübner, [1819]	276665.26:A192G; 85.28:C176T; 378.19:G1099C; 374.14:G1169T; keys to B.3a in ref. 10, but exclude B.9	303C1FD0-07CB-4919-900E-EA3D6347E5DD
Loboclina Grishin, new subtribe	<i>Lobocla</i> Moore, 1884	208.2:G145T, T146G; 18312.8:A619C, G620A; keys to B.4 in ref. 8 or C.5, C.10a, C.15.2 or C.18 in ref. 11	C606FC35-323D-4E55-AF5A-A86C6366BAFA
Cephisina Grishin, new subtribe	<i>Cephis</i> Evans, 1952	COI.bc:A44T, C84G, T479A; genitalia and palpi as described in ref. 17 (pp. 182–183) for <i>Cephis</i> are diagnostic	22B59811-F174-4FDF-A9D2-799897F4D44E
Telemiadina Grishin, new subtribe	<i>Telemiades</i> Hübner, [1819]	536.149:G1488C; 997.8:G514T; 860.7:A748G; 3001.3:C1773T; keys to B.2, C.3, C.7a (exclude C.7.6b), E.6a or E.9 in refs. 10 and 11	4AE0E59C-8B92-4C84-8651-E7A1C45C93C1
Oileidini Grishin, new tribe	<i>Oileides</i> Hübner, [1825]	1139.19:T562A; 851.8:C423A, G443A; 11945.11:G391A; 65.4:C330A; tuft of scales by anal fold from the base of hindwing, either above or below	CF9C3D29-523A-4D17-B140-9A69CFA98731
Typhedanina Grishin, new subtribe	<i>Typhedan</i> Butler, 1870	1341.12:T25841C; 489.5:G307T; 3446.8:T2308A, C2309G, A2500C; tuft of scales by anal fold from the base of hindwing above, but not below	B4D56F93-67F9-476F-B69C-133D988F8D58
Netrocorynini Grishin, new tribe	<i>Netrocoryne</i> C. & R. Felder, [1867]	2284.30:399A; 904.14:T439G; 275215.7:C925G; 998.8:G308A; 214.24:3520C; keys to B.1, C.1 or C.15 in ref. 8	DE61F048-02CF-4F8E-9392-D18A4618BABD
Jerini Grishin, new tribe	<i>Jera</i> Lindsey, 1925	103.23:796A; 420.27:G308A; 671.27:935C; 425.5:G1558T; keys to E.3 in ref. 11; forewing cell > 3/5 of costa	AF3B5CEA-880A-4CB2-AF40-E6D87C39C040
Pythonidina Grishin, new subtribe	<i>Pythonides</i> Hübner, [1819]	274.29:G397A; 3478.6:T116G; 7985.5:G916A; 925.10:G199C; in ref. 11, keys to E.44a, E.49.1 or, if uncut undivided, then E.37a or 40d	CB890271-5483-4B5A-A7BC-27DBC5E23DE5
Clitina Grishin, new subtribe	<i>Clito</i> Evans, 1953	COI.bc:G29T, 81A, 169A, 266A, 302T, A353T, A521T; keys to E.52 or E.13.8 in ref. 11	971884E2-E5F7-46A3-B182-657729B6A778
Butleriini Grishin, new tribe	<i>Butleria</i> Kirby, 1871	2627.8:A1459T; 141.4:C104A; 37338.38:G133T, G134C; keys to H.4 & 5 in ref. 12	D621EF81-FA65-4858-9450-E0C041598D7A
Pericharini Grishin, new tribe	<i>Perichares</i> Scudder, 1872	596.8:C1601G; 144.41:G201C; 83.15:G8658A, T8657G; keys to K.27a in ref. 12	94B68BD2-7F83-4E58-80E1-7F5AC8C56511

See *SI Appendix* for the lists of genera included in each taxon and sequences of protein-coding regions with diagnostic characters.

*Notation 272.1:A192G means position 192 in the gene 1 on the scaffold 272 is G, changed from A in the ancestor; 169A, means position 169 is A, but the ancestral state is unclear; COI.bc is the COI barcode region.

† ZooBank registration URL given for each taxon should be preceded by <http://www.zoobank.org>.

Phareas that has a divergent third segment (and was included in the Augiades group) is not in Entheini. Its males possess, absent in Entheini, tufts of hair-like scales in the groove along the hindwing vein 1A+2A. Our tree places this unusual skipper in Phocidini, implying that its peculiar palpi are convergent.

Oileidini is sister to other Eudaminae (Fig. 1). Genera in this tribe were grouped with some Pyrginae genera by Evans (11), and the tribe may be intermediate in morphology between Pyrginae and Eudaminae. This is the smallest tribe (six genera) and is characterized by tufts of hair-like scales in the groove along the hindwing vein 1A+2A in males, either below (*Oileides*) or above (the others). Similar structures in *Phareas* (Phocidini) are found on both sides of the hindwing.

The sister tribes Eudamini and Phocidini are separated from each other by a short branch and could be one tribe. However, each is strongly monophyletic, and Phocidini skippers stand out morphologically and ecologically: Their forewing veins R4 and R5 originate near each other, hindwing tornus is usually expanded (not lobed), skippers hold wings spread flat when resting, many are crepuscular, and many species are sexually dimorphic. Our Phocidini is the “D. Caelanorrhinus group” of Evans (10) after adding *Phocides*, *Phareas*, and *Emmelus* and removing *Cephise*, which has lobed or tailed hindwing tornus, and *Caelanorrhinus*, whose males have a tuft of long scales on the hind tibiae that fits into a thoracic pouch (not found in Eudaminae). The genomic tree suggests that *Oileides* is polyphyletic: one species together with *Aurina* belong to Phocidini.

Eudamini is the largest and most diverse tribe. It encompasses more than half of the subfamily. The genomic tree reveals groupings within the tribe that are described here as subtribes (Table 1). One of them, subtribe Cephisina, is monotypic for the genus *Cephise*, which diverged from its sister tribe Telemiadina 35 Mya and is unique in its morphological features (17). Telemiadina includes three genera: *Telemiades* with its close sister *Polygonus* and *Ectomis*, into which we sink *Hypocryptothrix*, *Heronia*, *Polythrix*, *Chrysoplectrum*, and *Speculum* (see *Uncanny Diver-*

gence Within a Genus). Along with *Lobocla*, *Loboclina* unifies genera with the arcuate antennal club from the “C. Urbanus group” of Evans plus *Venada*, *Aguna*, and a new genus, *Zeutus* (Table 2). Others belong to the “crown” group of Eudaminae. It includes an array of skippers that have been largely misclassified due to widespread and possibly mimetic convergence as detailed in *Widespread Convergence in Wing Patterns and Shapes*.

Eudaminae Genera and Subgenera. We define a major phylogenetic cluster of species with a common ancestor existing within a certain timeframe as being a genus (18). We cut a dated phylogenetic tree at a time point to maximize agreement with the current classification. This neither splits nor merges most genera that are well defined by morphological features, and we treat as genera the groups of species supported by the cut branches. The time 15 Mya corresponds to such point. It keeps well-known genera *Aguna*, *Udranomina*, and *Urbanus proteus* group unsplit. However, it separates traditional and morphologically distinct pairs of genera such as *Epargyreus* and *Chioides*. We attempt to reduce the number of monotypic genera, unless the genus is strongly distinct, because we wish to indicate relationships to other species by the name of a genus. As a result (Fig. 1 and *SI Appendix*), we delineated 50 Eudaminae genera, 4 of which are described as new in Table 2. The number of monotypic genera decreased from 10 to 4: *Nicephellus*, *Spathilepia*, and *Zeutus* (19), plus *Emmelus* transferred from Pyrginae. These four genera diverged from their sister taxa at least 18 Mya and are morphologically distinct.

Within some genera, we see groups of species that may be defined as subgenera (*SI Appendix*), and three new subgenera are described (Table 2). Some of the subgenera, such as *Thorybes*, have been used as genera for decades, but their genomic and morphological distinctness is smaller compared with most genera.

Widespread Convergence in Wing Patterns and Shapes. The most unexpected result of this study is the astounding number of misplacements of species into genera they do not belong to. The

Table 2. Description of new genera and subgenera of Hesperidae

New genus or subgenus	Type species	Diagnostic characters*	Derivation of the name†	ZooBank ID‡
<i>Tekliades</i> Grishin, new genus	<i>Thymele ramanatek</i> Boisduval, 1833	68A, C90T, T145C, 412T, 553A, 583T; keys to I.1.9 in ref. 7	Masculine, a blend of [ramana]Tek and [Coe]liades	0815648A-DA0C-4C46-AEAB-6C00131AC8BD
<i>Salantoia</i> Grishin, new genus	<i>Eudamus eriopis</i> Hewitson, 1867	59C, A79T, T163A, 530T, 598A, T637A; D.3.2 or 3 in ref. 10; harpe flat, not hook-shaped	Feminine, a blend of Sala[tis] and [Sarnie]ntoia	3F82E9DE-A5A2-44B3-A13D-53CF8A673FAE
<i>Spicauda</i> Grishin, new genus	<i>Goniurus procne</i> Plötz, 1881	307T, T349A, 424(not T), G506A, T562A; keys to C.13.13c in ref. 10; harpe dorsally spiked	Feminine, a blend of spica and cauda	14D26B57-940C-407B-8E70-4E25203044B8
<i>Urbanoides</i> Grishin, new subgenus	<i>Goniurus esmeraldus</i> Butler, 1877	T49A, A85T or C, T212A, C542T, T544A, A607C or T, T619A or G; keys to C.13.6a in ref. 10	Masculine, means similar to <i>Urbanus</i>	20FAC3B6-F038-40A0-B182-3C7F32A40702
<i>Zeutus</i> Grishin, new genus	<i>Cecropterus zeutus</i> Möschler, 1879	A22T, C271A, T278A, A526T, T548C, A607T; genitalia as for <i>zeutus</i> in ref. 19 (p. 27)	Masculine, echoes the type species name	75715B9C-46AB-40F5-B738-420DABD56863
<i>Lobotractus</i> Grishin, new genus	<i>Eudamus valeriana</i> Plötz, 1881	49A, T400A, 401T, A477G, 517T, C542T, T619A; as given for the “cyda group” in ref. 17 (p. 196)	Masculine, a blend of Lobo[cla] and [Coda]tractus	C6E5B5DF-1C74-4DBD-85C3-7285209F6F03
<i>Caudatractus</i> Grishin, new subgenus	<i>Eudamus alcaeus</i> Hewitson, 1867	355A, T556A, A592T; <i>Codatractus</i> (C.11 in ref. 10) with tailed hindwing	Masculine, includes tailed species of <i>Codatractus</i>	DF0F3C91-F56E-4B65-B86C-385A36F9D7FD
<i>Asina</i> Grishin, new subgenus	<i>Eudamus asine</i> Hewitson, 1867	T70A, T127A, T197C, 206T, 208A, A256T, T346A, 373A; keys to C.7.2a in ref. 10	Feminine, derived from the type species name	B3B7A6F6-A95C-4A2E-B9FB-80A7A8F86761
<i>Tiana</i> Grishin, new genus	<i>Ebrietas niger</i> Williams & Bell, 1940	T16C, A43T, G86A, T142C, A196G, T278A, 283C; keys to F.7.3 or 4 in ref. 11	Feminine, a blend of T[osta] and [I]ljiana	89382699-24FB-4466-B39B-94E6B544C425
<i>Chirgus</i> Grishin, new genus	<i>Hesperia limbata</i> Erschoff, 1876	85A, 205A, 223A, 241A, 263(not C), T277A, A415T, 479T, T574A; keys to G.1.2e or 9 in ref. 11	Masculine, a blend of Chi[lean] and [Py]rgus	7B1905F1-9471-48BF-90BF-32360783AB1E
<i>Burnsius</i> Grishin, new genus	<i>Syrictthus communis</i> Grote, 1872	205T, 223A, 241T, 263(not C), T277A, 479T; keys to G.1.5, 8, or 10a in ref. 11	Masculine, honors skipper taxonomist John M. Burns	48996B74-3AB1-4DEA-9A64-B8F112E62343
<i>Duroca</i> Grishin, new genus	<i>Hesperia duroca</i> Plötz, 1882	127T, 163C, 349C, T424C; keys to J.39.5a in ref. 12, harpe broad, hook-shaped	Feminine, echoes the type species name	476FE13C-5895-4139-BB11-44F835E21565

See *SI Appendix* for the lists of species included in each taxon.

*Sequence characters are for the COI barcode region and only their combination is diagnostic to distinguish from former genera of these species. Notation A79T means position 79 is T, changed from A in the ancestor; 59C means position 59 is C, but the ancestral state is unclear.

†All names are treated as nouns in the nominative singular.

‡ZooBank registration URL given for each taxon should be preceded by <http://www.zoobank.org/>.

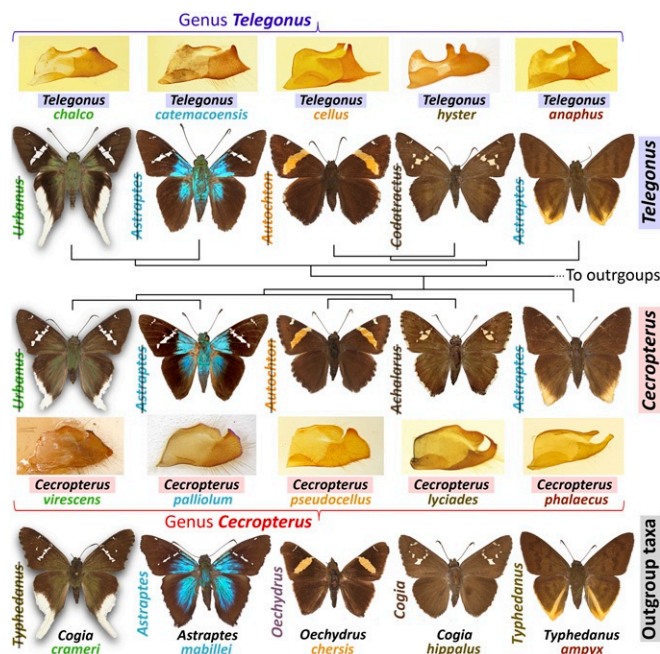


Fig. 2. Convergent wing patterns in *Telegonus*, *Cecropterus*, and others. Previously, they were placed in crossed out genera. Genitalia valvae are shown for the first two genera.

genera themselves, proposed over the years of classic entomological studies, mostly stood the test of genomic data: 55 genera were recognized before our work, and we revise them to be 50. We eliminated several monotypic genera for which visual morphological differences hindered close relationships with other species, and merged several phenotypically diverse but genotypically close genera (*SI Appendix*). Apparently, phenotypic distinction (e.g., wing shape, such as a tailed hindwing, or wing pattern, such as a pale stripe across the forewing) may be indicative of genetic differentiation. However, placement of a species into a genus by its dominant to human eye phenotypic feature is more problematic. For example, before our work, many tailed skippers were placed in the genus *Urbanus* based on the tail. Genomic data imply that half of them do not belong there, and we transferred them to three other genera. One genus is named here (*Spicauda*, Table 2), while the others have not previously included tailed skippers. Moreover, we transferred some skippers without long tails from *Astrartes* to *Urbanus*. They were formerly misclassified due to similar wing patterns consisting of shiny metallic-cyan wing bases and white forewing spots.

We found misclassifications to be widespread across Eudaminae and attribute them to convergence driven by the selection for large mimicry rings (20–22). This convergence is not confined to one or two basic patterns, but is more diverse. Some patterns are common in both the Old and New World, and, in addition to skippers, include butterflies from other families and even moths, flies, and beetles (20, 23). We find five different phenotypes (Fig. 2) that parallel each other in two genera (*Telegonus* and *Cecropterus*) and their more distant relatives (outgroups): (i) greenish bases of brown wings, white stripe on the forewing, and hindwing with a white tail and margins; (ii) metallic-blue wing bases, and forewing with white stripe (20); (iii) brown forewing with a yellow stripe across and apical white spots; (iv) cream-white, semitranslucent spots on the brown forewing; and (v) brown wings, and hindwing with yellow tornus.

At least four of these phenotypes are not ancestral and thus are convergent. Curiously, every species (10 out of 10) was placed by its appearance in a wrong genus before our study (crossed

out genus in Fig. 2). In retrospect, assignment to a correct genus could have been possible through detailed comparison of male genitalia. In *Telegonus*, the dorsal side of the valva is concave in the middle and forms a mouth-like structure with two “kissing lips.” In *Cecropterus*, the valva is dorsally and terminally convex, without a “kiss.” These genitalic features agree with genomic phylogeny and reinforce our conclusions.

Uncanny Divergence Within a Genus. Before our work, each of the three genera *Ectomis*, *Hypocrypthrix*, and *Heronia* consisted of single species of unique appearance. No taxonomist had ever thought of them being a group. To our surprise, all phylogenetic trees we obtained (even from COI barcodes only) revealed only a slight divergence among these three and two other genera, *Polythrix* and *Chrysoplectrum*, suggesting that it is best to place all these skippers in a single genus, *Ectomis*. Moreover, their most divergent phylogenetic lineage was part of *Polythrix*, and is described here as a new subgenus *Asina* (Table 2). Barcode divergence among the subgenus *Ectomis* is within 10% and is less than within the genus of swallowtail butterflies *Pterourus* (24), which some researchers consider a subgenus of *Papilio*.

Despite limited genetic differentiation, the expanded *Ectomis* contains species of profound phenotypic divergence. All skippers in the former genus *Polythrix* (subsumed by *Ectomis*) are tailed. Others are not, although their hindwing is usually lobed at the tornus. While most *Ectomis* are brown skippers with a pale forewing band frequently divided into spots, some vary from solid dark brown to dark metallic green with a forewing central spot, or tawny with many white spots. Some are even part of the mimetic complex with brilliant blue thorax and wing bases above (20). Males of some species possess tufts of hairlike scales on the forewing below, while others have a double row of yellow spines on the hind tibiae. Male genitalia are as diverse as wing shapes and patterns, and the phylogenetic closeness is not apparent from genitalia. The valva varies from a simple curved plate without elaborations (in *Ectomis cythna*) to very complex with several processes (in “*Heronia*” *labriaris*). It would be of interest to investigate genetic mechanisms for such a rapid phenotypic divergence within *Ectomis*.

Connecting Genotype to Phenotype. *Telegonus chalco* (Fig. 2) and *Telegonus brevicauda* are sister species that do not look alike. *T. brevicauda* belongs to the “blue-wing-base” mimicry complex (Figs. 1–3) and is nearly indistinguishable from its more distant relative *Urbanus tucuti* (Fig. 1). We compared genomes of these three species to find rapidly diverging proteins between the two sisters (*SI Appendix, Methods*). Genes involved in the circadian clock system, transcription regulation, wing morphogenesis, fatty acid, and vitamin metabolism (*SI Appendix, Tables S4 and S5*) stood out. The differences in morphogenesis genes may be related to the hindwing tail development in *T. chalco*. Next, we looked for fixed mutations in the uniquely patterned *T. chalco* compared with the two similar species. Xanthine dehydrogenase

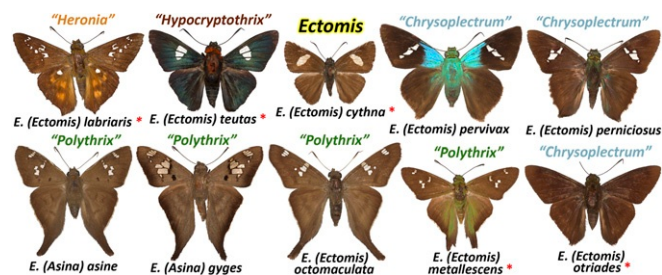


Fig. 3. Uncanny divergence within *Ectomis*. These skippers belonged to five genera listed above each image. Type species for these genera are marked with red asterisks.

had the largest number of such mutations (*SI Appendix, Table S6*). This enzyme is involved in pterin metabolism, and pterin derivatives were suggested to be responsible for wing coloration in pierid butterflies (25). The *T. chalco*-specific mutations may alter interactions between the enzyme and the transport proteins that deliver the enzyme to different pigment granules (26), and thus affect the wing color in *T. chalco*. Some of these mutations are present in other skipper species with extensive white scaling on hindwings.

Discussion: A Broad Picture. Today, and more so tomorrow, genome sequencing provides an efficient and cost-effective way to gain rapid insights about biodiversity. Given a taxonomic group, comparative analysis of genomic sequence along with morphology and ecology will be rich in discoveries. Even now, this can be done within a small laboratory. With the family Hesperidae, we have shown how such a project can be accomplished and have reported some of the results that can be expected. Not only is this family interesting as a diverse group of butterflies, it also emphasizes several generalities. First, while the reference genomes required freshly collected specimens, the bulk of the project was done with old museum samples. Even specimens collected a century ago yielded usable genomic data. Second, we provided a genome-based phylogeny of the family and reclassified it taxonomically. We found that many species were classified differently than previously thought. Other families may display the same problem when examined genomically. Third, we encountered many examples of phenotypic convergence and divergence, and mined genomic data for the links between genotype and phenotype. With the ever-decreasing cost of sequencing, we expect that, soon, any phylogenetic project will start from sequencing of genomes.

Materials and Methods

For freshly collected specimens, DNA was extracted from tissue of a specimen (minus wings and genitalia that were kept in an envelope) field-stored

in a vial with RNAlater. For pinned and dry specimens from museum collections, DNA was extracted from a whole abdomen or a leg. We have previously reported details of methods for DNA extraction, genomic library preparation, next-generation sequencing, and computational analysis of nuclear and mitochondrial genomes (13, 27). Phylogenetic trees were constructed with the programs RAxML (28) and ASTRAL (15), and were dated using the procedure described in *SI Appendix*, where further details of the methods are given. Genomic data obtained in the project have been deposited in the SRA database (accession SRP147939), and COI barcodes have GenBank accessions MH357724 through MH357835. ZooBank registration for this work, published on March 15, 2019, is 8A8E82C4-AC8A-4B7B-8DF4-6BFF4BD48F3D.

ACKNOWLEDGMENTS. We thank the Area de Conservación Guanacaste, Costa Rica (ACG) parataxonomists who found and reared many specimens used in this study. We are grateful to Robert K. Robbins, John M. Burns, and Brian Harris (National Museum of Natural History, Smithsonian Institution); Blanca Huertas, David Lees, and Geoff Martin (Natural History Museum, London); David Grimaldi and Courtney Richenbacher (American Museum of Natural History); Andrew D. Warren and Andrei Sourakov (McGuire Center for Lepidoptera and Biodiversity); Edward G. Riley, Karen Wright, and John Oswald (Texas A&M University Insect Collection), John Rawlins (Carnegie Museum of Natural History); Weiping Xie (Los Angeles County Museum of Natural History); Crystal Maier and Rebekah Baquiran (The Field Museum of Natural History); Vince Lee (California Academy of Sciences); James R. Reddell (University of Texas Biodiversity Center); Wolfram Mey (Berlin Museum für Naturkunde) for granting access to the collections under their care and for stimulating discussions; Bernard Hermier, Ernst Brockmann, and Paul A. Opler for specimens and generously sharing COI barcode sequences through the Barcode of Life Data System (BOLD) database (www.boldsystems.org); Jim P. Brock, William R. Dempwolf, Bernard Hermier, John MacDonald, and James A. Scott for specimens; Gerardo Lamas, Bernard Hermier, and Jonathan Pelham for fruitful discussions; Bernard Hermier, Lisa N. Kinch, and R. Dustin Schaeffer for critical reading of the manuscript; Texas Parks and Wildlife Department (Natural Resources Program Director David H. Riskind) for Research Permit 08-02Rev; and US National Park Service for the research permits: Big Bend (Raymond Skiles) for BIBE-2004-SCI-0011 and Yellowstone (Erik Oberg and Annie Carlson) for YELL-2017-SCI-7076. We acknowledge the Texas Advanced Computing Center at The University of Texas at Austin for providing high-performance computing resources. The study was supported, in part, by National Institutes of Health Grants GM094575 and GM127390 (to N.V.G.) and the Welch Foundation Grant I-1505 (to N.V.G.).

- Jarvis ED, et al. (2014) Whole-genome analyses resolve early branches in the tree of life of modern birds. *Science* 346:1320–1331.
- Warren AD, Ogawa JR, Brower AVZ (2008) Phylogenetic relationships of subfamilies and circumscription of tribes in the family Hesperidae (Lepidoptera: Hesperidae). *Cladistics* 24:642–676.
- Warren AD, Ogawa JR, Brower AVZ (2009) Revised classification of the family Hesperidae (Lepidoptera: Hesperidae) based on combined molecular and morphological data. *Syst Entomol* 34:467–523.
- Sahoo RK, et al. (2016) Ten genes and two topologies: An exploration of higher relationships in skipper butterflies (Hesperidae). *PeerJ* 4:e2653.
- Sahoo RK, Warren AD, Collins SC, Kodandaramiah U (2017) Hostplant change and paleoclimatic events explain diversification shifts in skipper butterflies (Family: Hesperidae). *BMC Evol Biol* 17:174.
- Toussaint EFA, et al. (2018) Anchored phylogenomics illuminates the skipper butterfly tree of life. *BMC Evol Biol* 18:101.
- Evans WH (1937) *A Catalogue of the African Hesperidae Indicating the Classification and Nomenclature Adopted in the British Museum* (British Museum, London).
- Evans WH (1949) *A Catalogue of the Hesperidae from Europe, Asia, and Australia in the British Museum (Natural History)* (British Museum, London).
- Evans WH (1951) *A Catalogue of the American Hesperidae Indicating the Classification and Nomenclature Adopted in the British Museum (Natural History). Part I. Introduction and Group A Pyrrhopyginae* (British Museum, London).
- Evans WH (1952) *A Catalogue of the American Hesperidae Indicating the Classification and Nomenclature Adopted in the British Museum (Natural History). Part II (Groups B, C, D) Pyrginae. Section 1* (British Museum, London).
- Evans WH (1953) *A Catalogue of the American Hesperidae Indicating the Classification and Nomenclature Adopted in the British Museum (Natural History). Part III (Groups E, F, G) Pyrginae. Section 2* (British Museum, London).
- Evans WH (1955) *A Catalogue of the American Hesperidae Indicating the Classification and Nomenclature Adopted in the British Museum (Natural History). Part IV (Groups H to P) Hesperinae and Megathyminae* (British Museum, London).
- Cong Q, Borek D, Otwinowski Z, Grishin NV (2015) Skipper genome sheds light on unique phenotypic traits and phylogeny. *BMC Genomics* 16:639.
- 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. *Curr Genomics* 18:366–374.
- Mirarab S, et al. (2014) ASTRAL: Genome-scale coalescent-based species tree estimation. *Bioinformatics* 30:i541–i548.
- Zhang J, et al. (2017) The complete mitogenome of *Euschemon rafflesia* (Lepidoptera: Hesperidae). *Mitochondrial DNA B* 2:136–138.
- Burns JM (1996) Genitalia and the proper genus: *Codatractus* gets *mysie* and *uydixa*—in a compact *cyda* group—as well as a *hysterectomy*, while *Cephise* gets part of *Polythrix* (Hesperidae: Pyrginae). *J Lepidopterists Soc* 50:173–216.
- 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.
- Williams RC, Bell EL (1934) Studies in the American Hesperioidea. Paper II (Lepidoptera). *Trans Am Entomol Soc* 60:17–30.
- Janzen DH, et al. (2009) Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol Ecol Resour* 9(Suppl 1):1–26.
- Janzen DH, Hallwachs W (2016) DNA barcoding the Lepidoptera inventory of a large complex tropical conserved wildland, Area de Conservación Guanacaste, northwestern Costa Rica. *Genome* 59:641–660.
- Janzen DH, Hallwachs W, Burns JM (2010) A tropical horde of counterfeit predator eyes. *Proc Natl Acad Sci USA* 107:11659–11665.
- Eisner T, et al. (2008) Defensive chemistry of lycid beetles and of mimetic cerambycid beetles that feed on them. *Chemoecology* 18:109–119.
- Shiraiwa K, Cong Q, Grishin NV (2014) A new *Heraclides* swallowtail (Lepidoptera, Papilionidae) from North America is recognized by the pattern on its neck. *Zookeys* 468:85–135.
- Morehouse NI, Vukusic P, Rutowski R (2007) Pterin pigment granules are responsible for both broadband light scattering and wavelength selective absorption in the wing scales of pierid butterflies. *Proc Biol Sci* 274:359–366.
- Reaume AG, Knecht DA, Chovnick A (1991) The rosy locus in *Drosophila melanogaster*: Xanthine dehydrogenase and eye pigments. *Genetics* 129:1099–1109.
- Cong Q, Grishin NV (2016) The complete mitochondrial genome of *Lerema accius* and its phylogenetic implications. *PeerJ* 4:e1546.
- Stamatakis A (2014) RAxML Version 8: A tool for phylogenetic analysis and postanalysis of large phylogenies. *Bioinformatics* 30:1312–1313.