Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology

Daniel H. Janzen,1,2,3, John M. Burns,4,5, Qian Cong,6,7, Winnie Hallwachs,6,7, Tanya Dapkey,8, Ramya Manjunath,9 Mehrdad Hajibabaei,9, Paul D. N. Hebert,9, and Nick V. Grishin7,10

*Department of Biology, University of Pennsylvania, Philadelphia, PA 19104-6018; †Department of Entomology, National Museum of Natural History, Smithsonian Institution, Washington, DC 20013-7012; ‡Department of Biophysics, University of Texas Southwestern Medical Center, Dallas, TX 75390-9050; §Department of Biochemistry, University of Texas Southwestern Medical Center, Dallas, TX 75390-9050; ¶Department of Integrative Biology, University of Guelph, Guelph, ON, Canada N1G 2W1; and ¶Howard Hughes Medical Institute, University of Texas Southwestern Medical Center, Dallas, TX 75390-9050

Contributed by Daniel H. Janzen, January 16, 2017 (sent for review December 12, 2016; reviewed by May R. Berenbaum and John Shuey)

DNA sequencing brings another dimension to exploration of biodiversity, and large-scale mitochondrial DNA cytochrome oxidase I barcoding has exposed many potential new cryptic species. Here, we add complete nuclear genome sequencing to DNA barcoding, ecological distribution, natural history, and subtleties of adult color pattern and size to show that a widespread neotropical skipper butterfly known as Udranomia kikkawai (Weeks) comprises three different species in Costa Rica. Full-length barcodes obtained from all three century-old Venezuelan syntypes of U. kikkawai show that it is a rainforest species occurring from Costa Rica to Brazil. The two new species are Udranomia sallydaleyae Burns, a dry forest denizen occurring from Costa Rica to Mexico, and Udranomia tomdaleyi Burns, which occupies the junction between the rainforest and dry forest and currently is known only from Costa Rica. Whereas the three species are cryptic, differing but slightly in appearance, their complete nuclear genomes totaling 15 million aligned positions reveal significant differences consistent with their 0.00065-Mbp (million base pair) mitochondrial barcodes and their ecological diversification. DNA barcoding of tropical insects reared by a massive inventory suggests that the presence of cryptic species is a widespread phenomenon and that further studies will substantially increase current estimates of insect species richness.

cryptic species | ACG | butterflies | DNA barcoding | genomics

The 35-y ongoing biodiversity inventory of the estimated 15,000 species of moths and butterflies of Area de Conservación Guanacaste (ACG)—a large complex tropical conserved wildland in northwestern Costa Rica (1–4)—poses many questions of core interest to evolution, ecology, and conservation. The intense collecting (3) followed by mitochondrial DNA barcoding (4, 5) often suggests that what was known as a single species may comprise several species. But is there really more than one species? Subsequent study often finds characters from morphology, ecology, or natural history that covary with the barcodes and thereby confirm a complex of several species that may not differ from each other in their external appearance (3–10). Udranomia kikkawai, a small and long-known Costa Rican skipper butterfly (Hesperiidae), is one exemplar among many. Here, we unravel its biology and taxonomy, in part by augmenting barcoding with complete genomic analysis. Every species, even cryptic, carries a unique set of biological traits and should be recognized as a unit of biodiversity.

Act One
Udranomia kikkawai was first encountered in the ACG dry forest caterpillar inventory when it began in 1978. An adult was reared from a small caterpillar feeding only on the youngest leaves of Ourtarea lucens (Ochnaceae), a common dry forest understory evergreen shrub. Today, 1,303 reared wild-caught ACG caterpillars later (among 650,000+ reared wild-caught caterpillars of 7,000+ species), U. kikkawai still has the same limited diet. In 1999, the ACG inventory discovered its caterpillar in immediately adjacent ACG rainforest, feeding only on the newest leaves of Crescentia spathulata (Ochnaceae), a common dry forest understory evergreen shrub. Today, almost 888 wild-caught rainforest caterpillars later, they are still just as restricted in their diet. U. kikkawai occurs from Mexico to Brazil (11). It is one of the 500+ species of Hesperiidae now known to live in ACG (10), and it is visibly distinct from the two other species in ACG that belong to the same genus: Udranomia octonius and Udranomia eurus. A large, common, and showy hesperiid named Astreptes fulgerator in 1775, spans the neotropics and frequents ACG. In 2004, DNA barcoding a large sample of this skipper and correlating the barcodes with food plants, caterpillar color patterns, and subtle differences in adult patterns and size revealed 10 species under that one name, just in ACG (6). In 2004, stimulated by this stunning discovery, the inventory began barcoding many thousands of specimens amassed over 26 y (4). The barcodes of dry forest U. kikkawai and rain forest U. kikkawai formed two distinct clusters, shallowly separated in a neighbor-joining (NJ) phenogram (SI Appendix, Fig. S1 and Table S1) by just 1.5% (10 bp). They were dubbed U. kikkawaiDHJ01 and U. kikkawaiDHJ02, respectively, as with hundreds of other “barcode splits” encountered in the inventory (3–11). Examination of adults (Fig. 1) revealed no reliable interspecific differences in their color patterns. Genitalia

Significance
Thirteen years of mitochondrial DNA barcoding of 15,000+ species of Lepidoptera and their parasitoids living in Area de Conservación Guanacaste, northwestern Costa Rica, indicate several thousand cases where barcodes combined with ecology suggest unrecognized cryptic species, substantially increasing species counts. Here, we show that the slightly different barcodes of three extremely similar parapatric-sympatric species of butterflies covary not only with ecology and subtle morphological traits but also with nuclear genomes—a finding that we predict will be commonplace and a method that we predict will be widely used. The barcodes of the century-old type specimens of Udranomia kikkawai from Venezuela reveal that this name applies to one of the three Costa Rican cryptic species; the others we describe as new.


Reviewers: M.R.B., University of Illinois at Urbana-Champaign; and J.S., The Nature Conservancy.

The authors declare no conflict of interest.

Data deposition: The barcode sequences reported in this paper have been deposited in the GenBank database (accession nos. KY421070, KY421071, and KY421072).

1To whom correspondence should be addressed. Email: djanzen@sas.upenn.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1621504114/-/DCSupplemental.
(SI Appendix, Figs. S2 and S3) showed none of the various differences that often characterize hesperiid species. Whereas some taxonomists are reluctant to recognize species distinguished only by “invisible characters” such as DNA barcodes and ecology, even when those characters covary, we felt that two species were involved. Moreover, each one feeds on just one of two species of Ochnaceae, limited to dry forest and rainforest. There was one anomalous specimen among 43 (voucher code 01-SRNP-17596), whose barcode differed by about 2.5% from the barcodes of both U. kikkawaiDHJ01 and -DHJ02. It was reared from a caterpillar eating O. lucens found side by side with numerous U. kikkawaiDHJ01, on the boundary between dry forest and rainforest. This odd specimen was baptized U. kikkawaiDHJ03.

Today, 342 DHJ01, 164 DHJ02, and 171 DHJ03 have been reared and DNA barcoded. DHJ03 (Fig. 1) occupies only the intergrade between dry forest and rainforest (Fig. 2). DHJ03 is a low-density dry-forest differentiate mixed in with DHJ01 along its eastern edge adjoining the rainforest. Their caterpillars eat the same species of food plant and even cooccur on the same
Fig. 2. Map of the three species of the U. kikkawai complex. Blue solid circles, U. kikkawai in rainforest; yellow solid circles, U. sallydaleyae in dry forest and rainforest–dry forest intergrade; and red solid circles, U. tomdaleyi in the rainforest–dry forest intergrade. Map is by Waldy Medina.

individual plants. However, there are a few places where DHJ03 occurs in the dry forest, 1–10 km west of the rainforest edge. These are well inside the dry forest (five such sites are evident as red solid circles in Fig. 2). However, each of these places is a local small moist site. Throughout this complex mingling of ecosystems and their occupants, there is even a single case of DHJ02 invading the edge of the dry forest, in one very local evergreen forest site. It is the blue solid circle to the west of the volcanic mountain range signaled by the cloud forest in Fig. 2. There, where all three species occur in the same hectare, the DHJ02 caterpillar was eating O. lucens rather than C. spathulata, its usual rainforest food plant, which is missing from that site.

Act Two

Further reference to the players in this taxonomic drama calls for scientific names. Now that U. kikkawai is apparently more than one species, which, if any, is the “real” U. kikkawai? Three specimens, one male and two females, collected in 1899–1900 in Suapure, deep in the Venezuelan rainforest, were described by Andrew Gray Weeks (14) in 1906 as U. kikkawai and deposited in the Museum of Comparative Zoology (MCZ) at Harvard University (MCZ type 00016685). To fix a single specimen as the name bearer, we designate the male apparently illustrated in Weeks (plate VI in ref. 15), specimen no. MCZ-ENT 00016685, collected on January 18, 1900, as the lectotype of U. kikkawai. Because the Udranomia species in ACG looked the same superficially and genitalically, we needed to DNA barcode all three of the ancient specimens of the type series. Full-length DNA barcode sequences were assembled from nine overlapping PCR products amplified from DNA extracted from abdomens of the three U. kikkawai type specimens (GenBank accessions KY421070, KY421071, and KY421072). All three barcodes turned out to be identical to each other and to that of DHJ02, the ACG rainforest species, which we therefore consider to be the true U. kikkawai.

In the description of species of the U. kikkawai group (Table 1), visible characters too complex to convey both concisely and completely in a few words are clearer in the text below and in Fig. 1. Subtle differences in facies may appear to distinguish one species from another, but individuals vary and overlap so much in these traits that even many newly eclosed and pinned fresh specimens cannot be identified by eye to any one of the three species with certainty (and especially not when their wings are worn). Imperfect detection of interspecific visual differences has been possible only with repeated perusal of hundreds of reared specimens grouped by barcodes and sex.

The intricately patterned ventral hindwing (Fig. 1) includes two relatively conspicuous bands of pale spots whose scales range
from beige to cream to white. The basal band consists primarily of two spots that are especially prominent when they are semihyaline. The irregular medial band consists primarily of seven spots, of which a small triangular one in cell M3–Cu1 and a larger rectangular one in cell Cu1–Cu2 are almost always semihyaline so that they also show dorsally in the middle of the hindwing (Fig. 1). These two contiguous spots generally overlap little if at all in Udranomia sallydaleyae and U. tomdaleyi but noticeably in U. kikkawai (Fig. 1).

The pale bands of the ventral hindwing (Fig. 1) connect via a weak to strong stripe of more or less white scales (some hair-like) near the inner margin of the wing. Between the pale bands is a band of brown ground color that may exhibit grayish (to almost semihyaline) overscaling (more extensive in females than in males) and dark.

By mutual transplants of eggs and larvae, we determined that U. kikkawai, the rainforest species, survives to adulthood if fed only leaves of O. lucens in laboratory confinement, but the field collections show no sign of population-level invasion of dry forest and its large stands of O. lucens. Equally, both U. sallydaleyae and U. tomdaleyi survive to adulthood when fed only new leaves of C. spathulata in the laboratory, but show no sign of having invaded the adjacent rainforest ecosystem in which this plant is common. The caterpillars and pupae of all three are similar and colored as the light pinkish yellows of the new expanding leaves commonly. The caterpillars and pupae of all three are similar and colored as the light pinkish yellows of the new expanding leaves of both species of food plants (2, 11).

Whereas DNA barcoding is cheap and fast for large numbers of specimens, a deep dive into the nuclear genome, although slow and expensive, was worthwhile in this exemplary case to exclude the possibility of irregularities with barcodes (16). There-
both nuclear and mitochondrial genes. The mitogenome tree based on 6,891 positions recapitulated the barcode phenogram (Fig. 3, Right). The nuclear genome tree (Fig. 3, Left) recovered all three species as monophyletic with 100% bootstrap using both RAxML (12) and PhyloBayes (13), supporting the conclusions based on barcodes and ecology. We found protein-coding genes that are conserved within, but divergent among, the three species as measured by a fixation index (17) larger than 0.8. In all three species, GO term enrichment analysis (18) detected proteins that participate in circadian temperature homeostasis, transcription regulation, circadian behavior, copulation, and eclosion rhythm as the most statistically significant (SI Appendix, Table S2). Previously, we found circadian clock proteins to be speciation hotspots in several pairs of closely related species (18, 19), and the divergence in proteins involved in copulation is expected to correlate with speciation. Thus, the nuclear genomic regions that diverged the most between the three Udranomia species may be responsible for the mechanics of their speciation in terms of both reproductive isolation and adaptation to different ecological conditions. This further supports the distinctiveness of the three Udranomia species.

The tree built from concatenated nuclear genes differs in topology from the mitogenome tree. Using the F4 statistic (20, 21), we detected statistically significant (P < 0.001) introgression between U. sallydaleya and U. tomdaleyi that may bias the nuclear tree and explain why they are grouped together. Thus, we selected genes that give trees in which all three species are monophyletic (i.e., have phylogenetic signal and show no introgression), resulting in an alignment of 397 nuclear genes. The tree built from them matches the topology of the mitogenome tree, placing U. tomdaleyi at the root.

Act Three

Even though U. kikkawai comprises three species, without the barcoding data that suggested splits it would still be considered just one, occurring in both dry forest and rainforest through much of the Neotropics. When only a few specimens were barcoded, the apparent barcode split was so shallow (less than 2%) that it could have been ignored. Indeed, the rainforest species and the dry forest species are placed in the same barcode index number (BIN) (BOLD:AAAS981) by one of the standard ways to group DNA barcodes at the species level (22, 23). When the single specimen of U. tomdaleyi turned up among the mass of U. kikkawai (rainforest) and U. sallydaleya (dry forest) specimens, it was viewed as a peculiar outlier, pseudogene, sequencing error, etc., and would have remained that way if haphazardly increasing the sample size had detected no more of them. However, as it went on to demonstrate, the ACG inventory heavily sampled other sites that appeared to have the same ecological circumstances as the site where the first U. tomdaleyi was encountered. Sampling went backward through specimens previously accumulated by the inventory and forward through new ones explicitly collected and reared for barcoding. The result thoroughly demonstrated that U. tomdaleyi is a real species that is ecologically specialized to live on the fragile join of dry forest with rainforest (24).

*Discussion*

How abundant are barcode splits among some 15,000 species of Lepidoptera, and fly and wasp parasitoids of their caterpillars, now DNA barcoded by the ACG inventory? As many as 4,500 species have barcode splits. At least 1,000 of these splits are as shallow as the 1–2% difference between the skipper butterflies U. kikkawai and U. sallydaleya (NJ tree in SI Appendix, Fig. S1 and refs. 4, 6, 10). Barring further analysis, we cannot confidently say how many of the barcode splits reflect biological species. But from long experience, we hazard that 10–20% of the traditional, morphologically based “single” species will turn out to be two or more (3, 4, 8–10). Sound evidence for species status of problematic barcode clusters separated by shallow splits comes from covariance between those clusters and independent characters. We demonstrate here that molecular data from nuclear DNA support those from mitochondrial DNA.

Altogether, covariance between traits as disparate as DNA barcodes, nuclear genomes, ecology, and wing facies shows that there are indeed three distinct biological species comprising a long unsuspected U. kikkawai complex in ACG and beyond (i.e., Mexico to Brazil). Due to new tools and their widespread use, it is increasingly clear that diverse geographic regions and
taxonomic groups contain many cryptic species. Their prior obscurity stems largely from our limited, and therefore biased, visual perception of the living world.

Forty-five protein-coding genes associated with specific functions differ among the species of the *U. kikkawai* complex (SI Appendix, Table S2). None of these genes directly affects an external visually obvious trait. Some of them relate to various aspects of behavior, including copulation, that can support differentiation that leads to speciation. Under such circumstances, which are doubtless common to many organisms, offspring species may, by chance, not noticeably differ outwardly from one another or from their parent species; and these cryptic species may endure in sympathy. A similar result is even more likely when, for any of various reasons (including mimicry and camouflage), selection favors retention of the parental appearance. Although the evolution and persistence of cryptic species frustrates some taxonomists, it stimulates others.

**Materials and Methods**

Procedures for collecting and barcoding specimens were described previously (6). Barcodes of *U. kikkawai* type specimens were amplified in nine overlapping segments. Primers for them are given in GenBank entries KY421070, KY421071, and KY421072. Methods for DNA extraction, library preparation, next-generation sequencing, and computational analysis of complete nuclear and mitochondrial genomes have been reported previously (19). See SI Appendix for details.

**ACKNOWLEDGMENTS.** We gratefully acknowledge the unflagging support of the team of ACG parataxonomists (3, 5) who found and reared the specimens used in this study and the team of biodiversity managers who protect and manage the ACG forests that host these insects and their food plants. We thank Donald J. Harvey for dissecting genitalia and Karie Darrow for photographing some of them, as well as adult specimens. We are indebted to Phil Perkins, Rachel Hawkins, and Naomi Pierce (Museum of Comparative Zoology, Harvard University) for the loan of *U. kikkawai* type series specimens for DNA analysis essential for this study. The two new species are named in honor of Mr. and Mrs. Tom and Sally Daley of Moraga, CA, in recognition of their many years of diligent management of the Guanacaste Dry Forest Conservation Fund financial bureaucracy. This study has been supported by US National Science Foundation Grants BSR 9024770 and DEB 9306296, 9400829, 9705072, 0072730, and 0515699; and grants from the Wege Foundation, International Conservation Fund of Canada, Jessie B. Cox Charitable Trust, Blue Moon Fund, Guanacaste Dry Forest Conservation Fund, Area de Conservación Guanacaste, Peruvian Global, Smithsonian Institution (National Museum of Natural History), US Department of Agriculture Systematic Entomology Laboratory, individual private donors, the University of Pennsylvania (D.H.J. and W.H.), National Institutes of Health Grant GM094575, and the Welsh Foundation Grant I-1505 (to N.V.G.). This study has been supported by the Government of Canada through its ongoing support of the Canadian National Collection, Genome Canada, the Biodiversity Institute of Ontario, Centre for Biodiversity Genomics, the Ontario Genomics Institute, and the Natural Sciences and Engineering Research Council of Canada.

25. Evans WH (1952) *Evans WH (1952)*
31. Evans WH (1952) *Evans WH (1952)*