Congruence between morphology-based species and Barcode Index Numbers (BINs) in Neotropical Eumaeini (Lycaenidae)

Carlos Prieto[1,2], Christophe Faynel[3], Robert K. Robbins[4] & Axel Hausmann[5]

[1] Departamento de Biología, Universidad del Atlántico, Barranquilla, Colombia
[2] Corporación Universitaria Autónoma del Cauca, Popayán, Colombia.
[3] 16 rue des Aspres, F-34160 Montaud, France.
[4] Department of Entomology, Smithsonian Institution, Washington, USA.
[5] SNSB-Bavarian State Collection of Zoology, Munich, Germany.

Corresponding Author:
Carlos Prieto[1,2]
Email address: cprieto50@gmail.com

Abstract

Background: With about 1000 species in the Neotropics, the Eumaeini butterflies (Lycaenidae, Theclinae) are one of the most diverse tribes among the Lycaenidae. Correct morphology-based identifications are challenging in many genera due to relatively little interspecific differences in wing patterns. Geographic infraspecific variation is sometimes more substantial than variation between species. In this paper we present a large DNA barcode dataset of South American Lycaenidae. We analyze how well DNA barcode BINs match morphologically delimited species.

Methods: We compare morphology-based species identifications with the clustering of molecular operational taxonomic units (MOTUs) delimitated by the RESL algorithm in BOLD, which assigns Barcode Index Numbers (BINs). We examine intra- and interspecific divergences for genera represented by at least four morphospecies. We discuss the existence of local barcode gaps in a genus by genus analysis. We also note differences in the percentage of species with barcode gaps in groups of lowland and high mountain genera.

Results: We identified 2213 specimens and obtained 1839 sequences of 512 species in 90 genera. Overall, the mean intraspecific divergence value of CO1 sequences was 1.20%, while the mean interspecific divergence between nearest congeneric neighbors was 4.89%, demonstrating the typical presence of a barcode gap. However, the gap seemed to disappear from the entire set when comparing the maximum intraspecific distance (8.40%) with the minimum interspecific distance (0.40%). Clear barcode gaps are present in many genera but absent in others. From the set of specimens that yielded COI fragment lengths of at least 650 bp, 75 % of the *a priori* morphology-based identifications were unambiguously assigned to a single Barcode Index Number (BIN). However after a taxonomic *a posteriori* review, the percentage of matched identifications rose to 85 %. BIN splitting was observed for 17 % of the species and BIN sharing for 9 %. We found that genera that contain primarily lowland species show higher percentages of

local barcode gaps and congruence between BINs and morphology than genera that contain exclusively high montane species. The divergence values to the nearest neighbors were significantly lower in high Andean species while the intra-specific divergence values were significantly lower in the lowland species. These results raise questions regarding the causes of observed low inter- and high intraspecific genetic variation. We discuss incomplete lineage sorting and hybridization as most likely causes of this phenomenon, as the montane species concerned are phylogenetically young and hybridization is probable. The release of our data set represents an essential baseline for a reference library for biological assessment studies of butterflies in mega diverse countries using modern high-throughput technologies and also highlights the necessity of taxonomic revisions for various genera combining both molecular and morphological data.

Introduction

The ability to delimit and identify species is the foundation for addressing taxonomic diversity issues in evolution, ecology, conservation, and biogeography. DNA barcodes potentially offer the opportunity for the rapid determination of species in large faunas, but reference libraries are needed to take advantage of this technique (Wirta et al. 2016; Hajibabaei et al. 2006). As of mid-2020, the Barcode of Life Database global repository (BOLD, http://www.boldsystems.org; Ratnasingham & Hebert 2007) includes more than 9 million DNA barcode sequences for over 224,000 metazoan (700,000 BINs, including many not yet identified taxa) and 69,000 plant species. There are DNA barcodes from species in every country worldwide, with many supporting national barcoding initiatives. ~~For e~~Each specimen in BOLD with a sequence longer than 500bp is automatically assigned a global unique identifier (BIN, Barcode Index Number) based on the Refined Single Linkage (RESL) algorithm (Ratnasingham & Hebert 2013). BIN assignments can be updated when new records reveal clear sequence divergence structure. DNA barcodes accurately delimit species in a number of large-scale studies (e.g., birds, Hebert et al. 2004b; Kerr et al. 2007; moths, Hebert et al. 2010; Hausmann et al 2011; Huemer et al. 2014; beetles, Hendrich et al. 2014; bees, Schmidt et al. 2015; dipterans, Morinière et al. 2019). They are often useful for discovering cryptic species, as has been shown with butterflies and flies (Hebert et al. 2004a; Smith et al. 2006; van Velzen et al. 2007; Riedel et al. 2013; Janzen et al. 2017; Espinoza et al., 2017; Dias et al., 2019; Tujuba et al. 2020). In many cases, BINs correspond with traditional taxonomy. However, perfect congruence is rare (e.g., Pyrcz et al 2018; Hawlitschek et al. 2017). While studies of the genetic diversity within a given species requires sampling from many localities (Bergsten et al. 2012), simple identification often requires only a single reference sequence (Hebert et al. 2003; Hausmann et al. 2013; Hawlitschek et al. 2017).

The utility of barcodes for describing several aspects of biodiversity depends on a strong correspondence between morphologically and genetically delimited entities. Although >20% of species pairs exhibit some level of incongruence in analyses at a continental scale (cf. Hausmann et al. 2013), the correlation increases significantly if the analyses are geographically restricted,

such as a single country (Hausmann 2011; Hausmann et al. 2013; Hendrich et al. 2014). For example, DNA barcodes accurately ~~identified~~ characterized more than 95% of Argentine butterfly species (Lavinia et al 2017). The success rate of DNA barcoding also varies among taxa, as can be seen among lepidopteran groups. DNA barcode species identifications were of more limited usefulness in neotropical Ithomiini~~ae~~ butterflies (Elias et al. 2007) and Palearctic Elachistidae moths (Kaila & Stahls 2006), but were more ~~accurate~~ useful in the lepidopteran families Hesperiidae, Sphingidae, Saturniidae, Geometridae and Erebidae (Hajibabaei et al. 2006; Hausmann et al. 2011; Rougerie et al. 2014; Ortiz et al. 2017).

The primarily neotropical Eumaeini (Lycaenidae, Theclinae) contains more than a thousand species (Robbins 2004) and represents one of the most rapid radiations among the butterflies. Taxonomic difficulties, external similarity, small size, rarity, high species richness, and restricted geographical distributions (at least of high montane species) are the most likely causes of the relatively scarce knowledge of this butterfly tribe. In contrast with other, better known families, lycaenids lack sufficiently illustrated identification keys, monographs, field guides, or checklists covering regions or countries in a comprehensive and updated manner. The use of DNA barcode sequences and BINs in this group has been limited, but congruence between morphology and barcode sequences is variable (Prieto et al. 2011; Faynel et al. 2011, 2012; Prieto et al. 2016; Cong et al. 2016, 2017; Prieto & Lorenc-Brudecka, 2017; Busby et al. 2017, Prieto et al. 2018; Faynel 2019). In particular, in previous studies it appeared that strictly high Andean genera were more likely, on average, to show incongruence.

Incongruence between morphology and barcodes occurs when more than one BIN is detected in a traditionally recognized species or when a BIN number comprises members of more than one recognized species (Hebert et al. 2004 a, b). BIN discordance can be caused by unrecognized cryptic diversity, whereas BIN sharing may indicate recently separated lineages that are still undergoing genetic differentiation. In both cases, an evidence-based taxonomic choice must be made, either to describe a new species (BIN split) or to synonymize two names (BIN sharing). These taxonomic decisions can increase the percentage of congruence between DNA and morphology-based analyses.

In this paper we present a large DNA barcode dataset of South American Lycaenidae. We analyze genus by genus how well DNA barcode BINs match morphologically delimited species. The general goal is to quantify the potential usefulness of reference libraries of DNA barcodes ~~BINs~~ for identification and for resolution of taxonomic problems in this group. In previous studies (e.g. Faynel, 2019; Prieto *et al*. 2016; Prieto *et al*, 2018) we found that congruence between DNA and morphology varies among genera. We hypothesized that the incidence of congruence among strictly high Andean genera was lower than among lowland genera. A specific goal of this paper is to evaluate the hypothesis that the ability of DNA barcodes ~~BINs~~ to discriminate morphologically delimited species decreases in high elevation lineages.

Materials & Methods

Morphology-based species identifications

Commented [WR1]: This comparison may say less about differences between Lepidoptera lineages and more about the areas where the studies were conducted. E.g., the Ithomiini study comes from one of the most diverse communities for these butterflies, and therefore one most likely to contain multiple closely related species. I am not sure about the other studies you cite, but the last one, for example, comes from the Iberian peninsula, which presumably contains a Lepidoptera fauna that is much less diverse than tropical erebid faunas, where barcodes may be less useful. You could start this sentence with a caveat such as "Although some apparent differences among taxa may be due to biogeographic factors,…"

The basis for identifying the species analyzed in this study is the checklist of Robbins (2004), which includes 1058 species of Eumaeini in 83 genera. The checklist was updated using subsequent publications (e.g. Busby et al, 2017, Prieto & Vargas 2016, Prieto et al. 2016, Robbins et al 2015, Faynel et al 2012, Faynel et al 2011, Prieto 2011, Duarte & Robbins 2010, Prieto et al. 2008, Bálint & Faynel 2008). When necessary, identifications were verified through genitalic examination.

Sampling and sequencing

Collecting permits in Colombia were obtained from ANLA Agencia Nacional de Licencias Ambientales (00594 April 26th 2018). Tissue samples were taken from pinned Eumaeini (Theclinae) in the research collections of Carlos Prieto (RCCP) and Christophe Faynel (RCCF). We selected specimens collected in the past 10 years because older material is more likely to have degraded DNA. Samples came from Costa Rica, French Guiana, Colombia, Ecuador, Peru, and Brazil (Figure 1). One to three legs were removed from each sampled specimen. The sample included 2,214 specimens of 541 species identified a priori based on the existing classification. The number of specimens per species ranged from two to 23.

DNA extraction, amplification, and sequencing of the COI barcode region were carried out by the Canadian Centre for DNA Barcoding (CCDB), Ontario, Canada, using standard high throughput protocols (Ivanova et al. 2006; deWaard et al. 2008). PCR amplification with a single pair of standard primers targeted a 658 bp region near the 5' terminus of the mitochondrial cytochrome c oxidase I (COI) gene that included the standard 648 bp barcode region for the animal kingdom (Hebert et al. 2004a). Complete specimen data including images, voucher deposition, accession numbers, GPS coordinates, sequence and trace files are accessible in the Barcode of Life Data System (BOLD dataset: DS-CPCF Faynel-Prieto Neotropical Theclinae; doi: https://dx.doi.org/10.5883/ DS-CPCF). Distance-based Neighbor joining (NJ), available on the BOLD website, was used to construct DNA barcode gene trees and to quantify sequence divergence. We analyzed the entire dataset and each genus with the NJ algorithm. In some cases, nearest neighbor genera with few species were combined in a single tree.

Congruence between morphology and BINs

BOLD currently contains close to 9,000,000 barcodes and over 700,000 BINs generated with the Refined Single Linkage (RESL) algorithm. RESL employs a three-phased analysis to reach decisions on the number and circumscription of BINs (= MOTUs) in the sequence data set on BOLD (Ratnasingham & Hebert 2013). It is much faster than other approaches, ~~as for instance~~such as the generalized mixed Yule-coalescent model (Pons et al. 2006; Fujisawa & Barraclough 2013), a critical requirement for the analysis of large data sets.

Morphological species were partitioned into three categories following the comparative methodology of Hausmann et al. (2013): (I) those in which there was a perfect match between morphological species and BINs; (II) splits: those where morphological species placed in more than one BIN and (III) merges: those where different species shared the same BIN assignment or

mixtures where some individuals of a species shared a BIN with another morphological species. We re-examined each sample in the latter two cases by checking both the morphological identification and the alignment and trace files.

Barcode gaps

We analyzed "bar-code gaps" to evaluate the hypothesis that incongruence between morphological species and BINs increases in high Andean lycaenid genera (e. g. Prieto et al. 2018; Faynel 2019). The "barcode gap" is a comparison of intraspecific versus interspecific divergence in among the barcode CO1 sequences. A barcode gap exists if the intraspecific divergence (of a particular species) is smaller than its lowest interspecific divergence. For example, a small intraspecific divergence combined with a large interspecific divergence is a large gap. We compared these divergences in the entire dataset and in groups of genera partitioned by the elevational distribution of its species. The criterion for assigning elevational groups was that at least the 90% of the species in a genus occur in 1) high mountain habitats (+2200m), 2) middle mountain habitats (1220m - 2200m), 3) lowland habitats (0m – 1200m), 4) middle mountain + high mountain and, 5) or lowland + middle mountain.
As a quick visualization of barcode gaps, we made scatterplots showing maximum intraspecific variation plotted against the minimum distance to the nearest non-conspecific individual. A 1:1 relationship is the point at which the difference between the two is zero (Collins & Cruickshank 2013). To determine sampling size bias, we also made scatterplots with the number of individuals in each species plotted against their maximum intraspecific variation. These analyses were performed for genera with at least two species and five sequences and for the groups of genera according to their elevational category. To evaluate if the divergence patterns for intraspecific variation and distances to nearest neighbor differ between the sets of species occurring in high mountain and lowland habitats, a Shapiro-Wilk normality test (Shapiro & Wilk, 1965), and a Wilcoxon signed-rank tests with continuity correction were performed. The analyzes were carried out in R software, using the dplyr and car packages, in addition to the pre-installed packages of the program.

Results

DNA barcode sequences at least 500 base pairs (bp) in length were successfully recovered from 1839 specimens. These sequences were assigned to 556 BIN numbers that belong to 512 morphology-based species in 90 genera. From the congruence analysis (1597 sequences, 558 BINs, 398 species, 52 genera) mean intraspecific variation ranged from a low of 0.1% in *Paraspiculatus* to a maximum value of 3.85% in *Cyanophrys*. Mean distances to nearest neighbor species ranged from 2.3% in *Contrafacia* to 10.4% in *Aubergina*. Altogether, 299 (75%) morphology-based species perfectly matched a unique BIN, while 36 species (9%) shared a BIN with up to six species, and 60 species (17.33%) were placed in two or more BINs. After reevaluating the morphology-based species based on the molecular results, congruence between morphology and BINs rose to 85%. However, BIN sharing and BIN splitting were particularly

frequent in typically high montane genera such as *Johnsonita*, *Rhamma*, *Podanotum*, and *Penaincisalia* (Table 1).

Barcode gaps

The percentage of species with a barcode gap in the complete data set was 87.2%. However, the proportion of species from habitats at different elevations in the datasets affected barcode gap frequency. Gaps were observed in 95.7% of the species from lowland ecosystems (0–1200 m) while only 61.7% of species from exclusively high montane ecosystems (>2200 m) had clear barcode gaps (Figures 2, 3). The trend towards a higher percentages of barcode gaps in lowland species was also found when including genera with species distributed in both lowlands and mid-montane habitats (89.5%), exclusively mid montane habitats (82.8%) and genera with exclusively mid- and high-montane species (73.8%) (Table 2.). The divergence values to the nearest neighbor were significantly lower in the high Andean species (W = 3456, p-value = 0.000002463), while the intra-specific divergence values were significantly lower in the lowland species (W = 364, p-value = 0.01103).

Discussion

BIN sharing

In this study, we obtained 1839 sequences of 512 species distributed in 90 genera for 557 BIN numbers, representing 78% of the available data on BOLD for the Eumaeini. From the set of specimens that yielded COI fragment lengths of at least 650 bp, 75 % of the *a priori* morphology-based identifications were unambiguously assigned to a single Barcode Index Number (BIN). After a taxonomic *a posteriori* review, the percentage of perfect matching rose to 85 %. Very low levels of interspecific barcode variation can reflect overlooked synonymy if misidentifications are ruled out (e.g Puillandre et al. 2011), but low genetic divergence, particularly based on just one genetic locus, does not automatically invalidate established taxonomy. In cases of recent phylogenetic divergence, phenotypic differentiation can occur more rapidly than the complete sorting of mtDNA into the new, separated lineages. The decision to consider two species names as synonyms must be made by a taxonomist. That is why in our study we strived for accurate identification before and after delimiting the species ~~molecularly~~using molecular data. When species pairs with low barcode divergences are ~~presented~~ recovered as monophyletic groups in the cladograms or identification trees, and their morphology is highly divergent, they can be validated as two different species, particularly if both occur in sympatry. There is no fixed threshold level of divergence that indicates species status because the percentage of divergence that would indicate whether two entities belong to the same species depends on the taxonomic group being studied and its evolutionary history. ~~However~~Nevertheless, most studies have found that COI divergences rarely exceed 2 % within ~~a~~ named and morphologically validated species, while members of different species typically show higher divergences, and it has been shown repeatedly that this ‘threshold’ can be used in many or

most metazoans to determine species status (Ratnasingham & Hebert 2007, 2013). However, distance and geographic isolation are two aspects that must be taken into account when delimiting biological entities based on established thresholds. Two entities living in sympatry can be considered different species even when there are small genetic divergences of 2% or less. But if these same entities are geographically distant, ~~the~~ a 2% divergence may be considered irrelevant to define them as separate species.

The percentage of clearly different morphological species grouped within the same BIN was predominantly high for the montane~~untain~~ genus *Rhamma*, where the BIN BOLD:ABX0547 is shared by five well-differentiated morphospecies. We suggest that most of the cases of BIN sharing between morphologically divergent high mountain species represent recently separated lineages that are still undergoing genetic differentiation. As most of these cases were recovered as monophyletic clades in the trees, a lower basic threshold setting in the algorithm would separate these species into different BIN numbers. However, it should be noted that with such modified settings and parameters the number of cases of BIN discordance in the same group of species may increase. In the case of the genus *Rhamma* the assignment of a single BIN number for clearly different morphological species is a result of the basic settings chosen for ~~the~~ delimitation of sequences into BINs ~~system~~ and not an intrinsic error of the barcode methodology.

Incomplete lineage sorting is relatively common in recently and rapidly radiating species groups as these species often have not yet had the necessary time to fix alternative haplotypes or alleles (Galtier & Daubin 2008). As a result, the relationships of incipient species typically progress from initial polyphyly through paraphyly and reach monophyly once lineage sorting is complete in the two sister species. Thus, in mtDNA analyses, relatively young species may appear polyphyletic or paraphyletic owing to incomplete lineage sorting (Tang et al. 2012). This phenomenon seems to be particularly common in high Andean genera such as *Rhamma* and has important effects on species identification and delimitation based on genetic sequence analysis. Further studies comparing genetic distances of sympatric and allopatric populations of several pairs of species can help to detect evidence of incomplete lineage sorting, and its prevalence in high mountain species of Theclinae.

BIN splits

High levels of intraspecific barcode variation often reflect cryptic species (e.g , Puillandre et al. 2011, 2012). However, deep barcode splits can also be the result of the recovery of pseudogenes, as a consequence of hybridization, or *Wolbachia* infection (Huemer et al. 2018, Mally et al. 2018, Werren et al. 2008). High percentages of BIN splits were found in some genera with typical mid- and high mountain representatives such as *Podanotum*, *Johnsonita*, *Thaeides* and *Rhamma*.

As noted by Prieto et al. (2018), the genus *Rhamma* includes several species presenting both types of discordance, BIN sharing and BIN splitting (e.g., *R. arria* and *R. bilix*). Species with ~~a~~ wider geographical distributions in high Andean ecosystems~~,~~ seem to show a greater number of

Commented [WR2]: Would be nice to add a sentence to clarify whether any of these are sympatric, and for those that are allopatric, what kinds of morphological differences are observed.

Commented [WR3]: 'Chosen' by whom? Presumably this threshold changes depending on the analysis, and is therefore 'chosen' by the program, not by the user a priori. Could clarify this.

Commented [WR4]: Not clear what this means – do you mean a flaw in the theory of applying barcodes to delimit species?

incongruences. Morphologically identified specimens were placed in three well-differentiated BINs for *R. arria* (BOLD:ABX0547, BOLD:ADD3784, BOLD:ADD3785) and four BINs for *R. bilix* (BOLD:ACF3699, BOLD:ABX0491, BOLD:ADD1839, BOLD:ABX0493). These clades might correspond to either divergent conspecific lineages, or unconfirmed putative species separated, in some cases, by deep genetic divergences. Cases of mitochondrial introgression can hinder the delimitation of some Eumaeini species in the genus *Calycopis* (Cong et al 2017), and we suppose that such processes occur more frequently in high Andean genera. Introgressive hybridization may have been common throughout the evolutionary history of these genera which are, therefore, of particular interest to taxonomists and evolutionary biologists because partial and unequal gene exchange can have important effects on the dynamics of speciation~~,~~ and phylogenetic patterns (Grant 1998; Grant et al. 2005; Funk and Omland 2003), and affect species identification and delimitation based on DNA sequences.

Barcode gap analysis

A ~~good~~ useful display of distance data for species delimitation is a scatterplot showing for each species the maximum intraspecific variation against the minimum distance to the nearest non-conspecific species ('nearest neighbor'), with a 1:1 slope representing the point at which the difference between the two is zero (Collins & Cruickshank 2013). This type of representation shows the barcode gap for each species in the dataset and can be an accurate display of the percentage of species in the study group that have a barcode gap (Figures 2, 3). Since the identification of species and the delimitation of taxonomic entities using barcodes~~,~~ depend on the existence of a clear DNA barcode gap, a quick visualization of the existence of these gaps in each species can ~~be an indication~~indicate ~~of~~ the usefulness of the DNA barcoding approach in a given genus.

~~The m~~Major topographic and climatic variations are important factors that determine the nature of South American biodiversity. The geography of South America, together with its climate and biodiversity, ha~~s~~ve evolved over a very long period, initiated about 100 MY ago on the ancient Gondwanan continent. The Andes rose much later (about 25 MY ago) and were populated by butterflies mainly originating in the eastern parts of the continent (Purser, 2015). Many species of butterflies are found near, or within, the complex valley systems of the Andes, a result of the combination of at least two important factors: altitudinal gradients and geographic barriers created by the intricate system of valleys and ridges (e.g. Willmott 2003, Holzinger & Holzinger, 1994, Ebel et al. 2015). The tectonic rise of the Andes has created new environments and modified others, and the uplift of the cordilleras has separated butterfly communities favoring the evolution of allopatric vicariants. Dramatic changes in global climate during glaciations, accompanied by major adjustments in vegetation, created new biomes which again may have stimulated the evolution of new species and subspecies, especially ~~mainly~~ at high altitudes (Pyrcz et al. 2017, Purser 2015). These changes are quite rapid on a geological scale, and certain subspecies, notably among the high altitude pronophilines (Satyrinae), seem to have evolved since the last glacial maximum 20,000 years ago (Adams 1985, Pyrcz et al 2009, Casner & Pyrcz

Commented [WR5]: This still seems a really inappropriate generalization to make here, the reference is not a phylogenetic study and therefore inferences about areas of origin could well be wrong. There are plenty of biogeographic studies of Neotropical butterfly lineages that could be used to make some generalizations, however. Also, it's not really clear what you mean by 'originating' – an Andean lineage that has its inferred area of origin in eastern South America? Presumably you don't mean species that originated in the east and subsequently moved to colonize the Andes, since there are very few species occurring in both areas.

2010). Several phylogenetic studies of~~n~~ butterflies indicate that the most recent diversification events tend~~s~~ to occur ~~near the summits of the Andes~~at high elevations and that the highest altitude species and subspecies are the youngest (e.g., Casner & Pyrcz 2010, Pyrcz et al. 2017, Jiggins et al. 2006,).

Due to incomplete lineage sorting in very young species, it is not easy to accurately define taxonomic boundaries, and~~,~~ additional difficulties may be caused by hybridization. Although the ~~Refined Single Linkage (~~RESL~~)~~ algorithm as the basis for the BIN system (Ratnasingham & Hebert 2013) provides a powerful tool to propose primary, tentative species hypotheses for large datasets (Ratnasingham and Hebert, 2007), such an mtDNA-based approach cannot prove the absence of gene flow and still depends on arbitrary, a priori settings and assumptions. The efficiency of these methods largely depends on the accumulation of mtDNA mutations since species separation, and thus can only delimit lineages with sufficiently long isolation (Rannala 2015). Moreover, nothing is known about how the kind of speciation process (vicariance of an existing species versus a small founder colonization) might affect the ability of barcodes to identify species correctly. We assume that incomplete lineage sorting and occasional hybridization are usual phenomena in the very young species of South American Eumaeini, and that these are the two most likely causes of the low percentages of DNA barcode gaps found in high Andean species in comparison with the older species from the lowlands.

However, other very species-rich groups of Andean butterflies with recent speciation processes, such as the subtribe Pronophilina, have shown very high percentages of barcode gaps and perfect congruence between morphology and DNA barcodes (e.g. Marín et al. 2017; Pyrcz et al. 2018). This shows that, in certain groups, other biological factors allow young high Andean species to present more complete DNA lineage sorting in short periods of time. In the case of Pronophilina, the low vagility of its species could be a determining factor that ~~avoids~~ limits gene flow between~~the meeting of~~ separate populations ~~and therefore the incidence of hybridization, allowing populations to remain strictly separated for longer periods of time~~promotes lineage sorting.

Conclusions

In mtDNA analyses, relatively young species may appear polyphyletic or paraphyletic owing to incomplete lineage sorting, and other aspects such as introgressive hybridization may have been common throughout the evolutionary history of Eumaeini genera. Partial gene exchange can have important effects on the dynamics of speciation, and affect species delimitation based on DNA sequences. These phenomena seem to be particularly common in high Andean genera such as *Rhamma* and to have important effects on species identification based on genetic sequence analysis.

Since we found evidence, at least in the tribe Eumaeini, that relatively young species in young ecosystems tend to have more incongruences between morphology and DNA delimitation, and thus lower percentages of DNA barcode gaps, it would be interesting to find out if there are

similar patterns when comparing groups of species belonging to related genera in young and old ecosystems at the same altitude. This could be done by comparing a group of high mountain species from the northern part of the Andes in Venezuela, with their relatives in the central part of the Andes in Peru. These two regions exhibit a different geological age of around 40 MY, with the Venezuelan part being the youngest.

Acknowledgements

For help and support in many ways we acknowledge Daniel Augusto Mantilla, (Popayán, Colombia), Evgeny V. Zakharov (CCBD, Guelph, Canada), Pierre Boyer (France), Gregory Nielsen, (Colombia), Myriam Nicolle Pertuz (Barranquilla, Colombia), Tomasz Pyrcz (Kracow, Poland), Michael Balke (ZSM, Munich, Germany), Soranggy Cruzco (Barranquilla, Colombia) and Zsolt Balint (HNHM, Budapest, Hungary). This research was supported with funds from the Georg Forster Research Fellowship Program of the Alexander von Humboldt Foundation (Bonn), the Research Group Linkage Programme: Evolution of the high Andean insect fauna project, the Federal Ministry for Education and Research (Germany), the Corporación Universitaria Autónoma del Cauca (Popayán, Colombia) and the Vice-Rectorate for Research of the Universidad del Atlántico, (Barranquilla, Colombia) under resolution number 3247 12th June of 2015.

References

Adams MJ 1985. Speciation in the Pronophiline Butterflies (Satyridae) of the Northern Andes. *Journal of Research on the Lepidoptera*, Supplement 1: 33–49.

Balint Zs, Faynel C. 2008. Review of the genus *Brangas* Hübner, 1819 (Lepidoptera: Lycaenidae) with description of a new genus. *Annales Historico-Naturales Musei Nationalis Hungarici*, 100: 271–306.

Bergsten J, Bilton DT, Fujisawa T et al. 2012. The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, 61, 851–869.

Busby R, Faynel C, Moser A, Robbins RK. 2017. Sympatric diversification in the upper Amazon: A revision of the Eumaeine genus *Paraspiculatus* (Lepidoptera: Lycaenidae). *Smithsonian Contributions to zoology*, 649: 1–63.

Casner KL, Pyrcz TW. 2010. Patterns of timing of diversification in a tropical montane butterfly genus, *Lymanopoda* (Nymphalidae, Satyrinae). *Ecography*, 35: 1–9.

Collins RA, Cruickshank RH, 2013. The seven deadly sins of DNA barcoding. *Molecular Ecology Resources*, 6: 969–975.

Cong Q, Shen J, Borek D, Robbins RK, Otwinowski Z, Grishin NV. 2016. Complete genomes of Hairstreak butterflies, their speciation, and nucleomitochondrial incongruence. *Scientific Reports*, 6: 24863. (doi:10.1038/srep24863)

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 of London B*, 284: 20161735.

deWaard JR, Ivanova NV, Hajibabaei M, Hebert PDN. 2008. Assembling DNA barcodes: analytical protocols. In: C. Martin (Ed.), Methods in molecular biology: environmental genetics. (pp. 275–293) Totowa: Humana Press Inc.

Dias FMS, Janzen D, Hallwachs W, Chacón I, Willmott K, Ortiz-Acevedo E, Mielke OHH, Casagrande MM, 2019. DNA Barcodes uncover hidden taxonomic diversity behind the variable wing patterns in the Neotropical butterfly genus *Zaretis* (Lepidoptera: Nymphalidae: Charaxinae). *Zoological Journal of the Linnean Society*, 185 (1): 132–192.

Duarte M, Robbins RK. 2010. Description and phylogenetic analysis of the Calycopidina (Lepidoptera, Lycaenidae, Theclinae, Eumaeini): a subtribe of detritivores, *Revista Brasileira de Entomologia*, 54 (1): 45–65.

Ebel E, Da Costa JM, Sorenson M, Hill RI, Briscoe A, Willmott K, Mullen SP. 2015. Rapid diversification associated with ecological specialization in Neotropical *Adelpha* butterflies. *Molecular Ecology*, 24(10): 2392–2405.

Elias M, Hill R I, Willmott, KR, Mallet J, Jiggins CD. 2007. Limited performance of DNA barcoding in a diverse community of tropical butterflies. *Proceedings of the Royal Society of London B*, 274: 2881–2889.

Espinoza B, Janzen DH, Hallwachs W. 2017. 17 new species hiding in 10 long-named gaudy tropical moths (Lepidoptera: Erebidae, Arctiinae). *Tropical Lepidoptera Research*, 27 (1): 1–29.

Faynel C. 2019. Le genre *Chalybs* Hübner, [1819] avec la description d'une nouvelle espéce du plateau des Guyanes (Lepidoptera, Lycaenidae, Theclinae). *Lambillionea*, CXIX, 3, (Supplément): 3–34.

Faynel C, Busby R, Moser A, Robbins RK. 2011. Species level taxonomy of the Neotropical hairstreak genus *Porthecla* (Lepidoptera: Lycaenidae: Theclinae: Eumaeini). *Annales de la Société entomologique de France*, 47 (1–2): 241–259.

Faynel C, Busby R, Robbins RK. 2012. Review of the species level taxonomy of the neotropical butterfly genus *Oenomaus* (Lycaenidae, Theclinae, Eumaeini). *ZooKeys*, 222: 11–45.

Fujisawa T, Barraclough TG. 2013. Delimiting species using single-locus data and the Generalized Mixed Yule Coalescent (GMYC) approach: a revised method and evaluation on simulated datasets. *Systematic Biology*, 62: 707– 724.

Funk DJ, Omland KD. 2003. Species-level paraphyly and polyphyly: frequency, causes, and consequences, with insights from animal mitochondrial DNA. *Annual Review of Ecology, Evolution, and Systematics*, 34: 397–423.

Galtier N, Daubin V. 2008. Dealing with incongruence in phylogenomic analyses. *Philosophical Transactions of the Royal Society B-Biological Sciences*, 363: 4023–4029.

Grant PR. 1998. Evolution on islands. Oxford (UK): Oxford University Press; p. 334.

Grant PR, Grant BR, Petren K. 2005. Hybridization in the recent past. *American Naturalist*, 166: 56–67.

Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proceedings of the National Academy of Sciences of the United States of America*, 103: 968–971.

Hausmann A. 2011. An integrative taxonomic approach to resolving some difficult questions in the Larentiinae of the Mediterranean region (Lepidoptera, Geometridae). *Mitteilungen der Münchner Entomologischen Gesellschaft*, 101: 73–97.

Hausmann A, Haszprunar G, Segerer AH, Speidel W, Behounek G, Hebert PDN. 2011. Now DNA-barcoded: The Butterflies and Larger Moths of Germany (Lepidoptera: Rhopalocera, Macroheterocera). *Spixiana*, 34 (1): 47–58.

Hausmann A, Godfray HCJ, Huemer P, Mutanen M, Rougerie R, van Nieukerken EJ, et al. 2013. Genetic Patterns in European Geometrid Moths Revealed by the Barcode Index Number (BIN) System. *PLoS ONE*, 8 (12). doi: 10.1371/journal.pone.0084518.

Hawlitschek O, Morinière J, Lehmann GUC, Lehmann AW, Kropf M, Dunz A, Glaw F, Detcharoen M, Schmidt S, Hausmann A, Szucsich NU, Caetano-Wyler SA, Haszprunar G. 2017. DNA barcoding of crickets, katydids, and grasshoppers (Orthoptera) from Central Europe with focus on Austria, Germany, and Switzerland. *Molecular Ecology Resources*, 17 (5). doi: 10.1111/1755-0998.12638

Hebert PDN, Cywinska A, Ball SL, DeWaard JR 2003. Biological identifications through DNA barcodes. *Proceedings of the Royal Society B: Biological Sciences*, 270: 313–321.

Hebert PDN, Penton, EH, Burns JM, Janzen DH, Hallwachs W. 2004a. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly Astraptesfulgerator. *Proceedings of the National Academy of Sciences of USA*, 101(41): 14812–14817.

Hebert PDN, Stoeckle MY, Zemlak TS, Francis CM. 2004b. Identification of birds through DNA barcodes. *PLoS Biology*, 2:1657–1663.

Hebert PDN, deWaard JR, Landry JF. 2010. DNA barcodes for 1/1000 of the animal kingdom. *Biological Letters*, 6: 359–362. doi:10.1098/rsbl.2009.0848. PubMed: 20015856.

Holzinger H, Holzinger R. 1994. *Heliconius* and related genera. In: Heliconius and related genera: Lepidoptera Nymphalidae; the genera *Eueides*, *Neruda* and *Heliconius*. Venette, France, Sciences Nat. 328 pp., 51 pls.

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. doi:10.1371/journal.pone.0115774.

Huemer P, Karsholt O, Aarvik L, Berggren K, Bidzilya O, Junnilainen J, Landry JFL, Mutanen M, Nupponen K, Segerer A, Šumpich J, Wieser Ch, Wiesmair B, Hebert PDN. 2018. DNA barcode library for European Gelechiidae (Lepidoptera) suggests greatly underestimated species diversity. *Zookeys*, 921: 141–157. doi: 10.3897/zookeys.921.49199

Hendrich L, Morinière J, Haszprunar G, Hebert PDN, Hausmann A, Köhler F, Balke M. 2014. A comprehensive DNA barcode database for Central European beetles with a focus on Germany: adding more than 3500 identified species to BOLD. *Molecular Ecology Resources*, 15(4):795–818 doi: 10.1111/1755-0998.12354.

Ivanova NV, deWaard JR, Hebert PDN. 2006. An inexpensive, automation-friendly protocol for recovering highquality DNA. *Molecular Ecology Notes*, 6: 998–1002.

Janzen DH, Burns JM, Cong Q, Hallwachs W, Dapkey T, Manjunath R, Hajibabaei M, Hebert PDN, Grishin NV. 2017. Nuclear genomes distinguish cryptic species suggested by their DNA

barcodes and ecology. *Proceedings of the National Academy of Sciences of the United States of America*, 114 (31): 8313–8318. doi:10.1073/pnas.1621504114.

Jiggins CD, Mallarino R, Willmott KR, Bermingham E. 2006. The phylogenetic pattern of speciation and wing pattern change in Neotropical *Ithomia* butterflies. *Evolution*, 60 (7): 1454 – 1466.

Kaila L, Stahls G. 2006. DNA barcodes: Evaluating the potential of COI to diffentiate closely related species of Elachista (Lepidoptera: Gelechioidea: Elachistidae) from Australia. *Zootaxa*, 1170: 1–26.

Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM, Hebert PDN. 2007. Comprehensive DNA barcode coverage of North American birds. *Molecular Ecology Notes*, 7: 535–543.

Lavinia PD, Nuñez Bustos EO, Kopuchian C, Lijtmaer DA, García 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(10): e0186845. https://doi.org/ 10.1371/journal.pone.0186845

Mally R, Huemer P, Nuss M. 2018. Deep intraspecific DNA barcode splits and hybridisation in the Udea alpinalis group (Insecta, Lepidoptera, Crambidae)–an integrative revision. *ZooKeys*, 746: 51–90. 10.3897/zookeys.746.22020

Marín MA, Cadavid IC, Valdés L, Álvarez CF, Uribe SI, Vila R, Pyrcz TW. 2017. DNA barcoding of an assembly of montane Andean butterflies (Satyrinae): Geographical scale and identification performance. *Neotropical Entomology*, 46: 514–523.

Morinière J, Balke M, Doczkal D, Geiger MF, Hardulak LA, Haszprunar G, Hausmann A, Hendrich L, Regalado L, Rulik B, Schmidt S, Wägele JW, Hebert PDN. 2019. A DNA barcode library for 5,200 German flies and midges (Insecta: Diptera) and its implications for metabarcoding-based biomonitoring. *Molecular Ecology Resources*, 19: 900–928.

Ortiz A, Rubio R, Guerrero JJ, Garre MJ, Serrano J, Hebert PDN, Hausmann, A. 2017. Close congruence between Barcode Index Numbers (BINs) and species boundaries in the Erebidae (Lepidoptera: Noctuoidea) of the Iberian Peninsula. *Biodiversity Data Journal*, 5.e19840. DOI: 10.3897/BDJ.5.e19840.

Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin W, Vogler AP. 2006. Sequence based species delimitation for the DNA taxonomy of undescribed insects. *Systematic Biology*, 55: 595–609.

Prieto C. 2011 The genus *Micandra* Staudinger (Lepidoptera: Lycaenidae: Theclinae) in Colombia, with the description of a new species from the Sierra Nevada de Santa Marta, *Zootaxa*, 3040: 55–68.

Prieto C, Bálint Zs, Boyer P, Micó E. 2008. A review of the "*browni* group" of *Penaincisalia* with notes on their distribution and variability (Lepidoptera: Lycaenidae). *Zootaxa*, 1941: 1–24.

Prieto C, Micó E, Galante E. 2011. Molecules, Wing Pattern and distribution: an approach to species delimitation in the "loxurina group" (Lepidoptera: Lycaenidae: Penaincisalia). *Neotropical Entomology*, 40 (5): 553– 559.

Prieto C, Grishin N, Hausmann A, Lorenc-Brudecka J. 2016. The *Penaincisalia amatista* species-group (Lepidoptera: Lycaenidae, Eumaeini) in Colombia, insights from mtDNA barcodes and the description of a new species. *Systematics and Biodiversity*, 14 (2): 171–183.

Prieto C, Vargas MA. 2016. Elfin butterflies of the genus *Rhamma* Johnson (Lepidoptera: Lycaenidae: Theclinae): A review of the Colombian species. *Zootaxa*, 4093 (3): 323–342. https://doi.org/10.11646/zootaxa.4093.3.2

Prieto C, Lorenc-Brudecka J. 2017. Description of *Rhamma dawkinsi* (Lepidoptera: Lycaenidae) a new mountain butterfly from Colombia. *Zootaxa*, 4338 (3): 587–594.

Prieto C, Nuñez R, Hausmann A. 2018. Molecular species delimitation in the genus *Rhamma* Johnson, 1992 (Lepidoptera: Lycaenidae, Theclinae). *Mitochondrial DNA*, 30 (1): 1 –17.

Puillandre N, Macpherson E, Lambourdiere J, Cruaud C, Boisselier MC, Samadi S. 2011. Barcoding type specimens helps to identify synonyms and an unnamed new species in *Eumunida* Smith, 1883(Decapoda: Eumunididae). *Invertebrate Systematics*, 25: 322–333.

Puillandre N, Lambert A, Brouillet S, Achaz G. 2012. ABGD, Automatic barcode gap discovery for primary species delimitation. *Molecular Ecology*, 21(8): 1864–1877. https://doi.org/10.1111/j.1365-294X.2011.05239.x

Purser B. 2015. Butterflies of the Andes: their biodynamics and diversification. The International Biodiversity Foundation, Mariposa Press, Gainesville, Florida 251 pp.

Pyrcz T, Wojtusiak J, Garlacz R. 2009. Diversity and distribution patterns of pronophilina butterflies (Lepidoptera: Nymphalidae: Satyrinae) along an altitudinal transect in North-Western Ecuador. *Neotropical Entomology*, 38(6):716–726.

Pyrcz, T, Lorenc-Brudecka J, Zubek A, Boyer P, Gabaldon C, Mavarez J. 2017. Taxonomy, phylogeny and distribution of the genus *Steromapedaliodes* sensu novo in the Cordillera de Mérida, Venezuela (Lepidoptera: Nymphalidae: Satyrinae: Satyrini). *Arthropod Systematics and Phylogeny*, 75(2): 195–243.

Pyrcz T, Prieto C, Boyer P, Lorenc-Brudecka J. 2018. Discovery of a remarkable new species of *Lymanopoda* Westwood and considerations on its position in the generic phylogeny: an integrative taxonomic approach (Lepidoptera, Nymphalidae, Satyrinae). *European Journal of Entomology*, 115: 387–399.

612 Tang QY, Liu SQ, Yu D, Liu HZ, Danley PD. 2012 Mitochondrial captureand incomplete
613 lineage sorting in the diversification of balitorine loaches (Cypriniformes, Balitoridae) revealed
614 by mitochondrial and nuclear genes. *Zoologica Scripta*, 41: 233–247.
615
616 Rannala B. 2015. The art and science of species delimitation. *Current Zoology*, 61: 846–853.
617 https://doi.org/10.1093/czoolo/61.5.846
618
619 Ratnasingham S, Hebert PDN. 2007. BOLD: the barcode of life data system
620 (www.barcodinglife.org). *Molecular Ecology Notes*, 7: 355–364.
621
622 Ratnasingham S, Hebert PDN. 2013. A DNA-based registry for all animal species: The Barcode
623 Index Number (BIN) System. *PLOS ONE*, 8(8): e66213. doi:10.1371/journal.pone.0066213.
624
625 Riedel A, Sagata K, Surbakti S, Tänzler R, Balke M. 2013. One hundred and one new species of
626 *Trigonopterus* weevils from New Guinea. *ZooKeys*, 280.
627
628 Robbins RK. 2004. Lycaenidae. Theclinae. Tribe Eumaeini. In: Lamas G (Ed.) Checklist: Part
629 4A. Hesperioidea — Papilionoidea. In: Heppner JB (Ed.) Atlas of Neotropical Lepidoptera.
630 Volume 5A. Association for Tropical Lepidoptera and Scientific Publishers, Gainesville, FL,
631 118–137.
632
633 Robbins RK, Heredia AD, Busby RC. 2015. Male secondary sexual structures and the
634 systematics of the *Thereus oppia* species group (Lepidoptera, Lycaenidae, Eumaeini). *ZooKeys*,
635 520: 109–130.
636
637 Rougerie R, Kitching IJ, Haxaire J, Miller SE, Hausmann A, Hebert PDN. 2014. Australian
638 Sphingidae - DNA barcodes challenge current species boundaries and distributions. *PLoS ONE*,
639 9(7): e101108. doi:10.1371/journal.pone.0101108
640

Schmidt S, Schmid-Egger C, Moriniere J, Haszprunar G, Hebert P. 2015. DNA barcoding largely supports 250 years of classical taxonomy: Identifications for Central European bees (Hymenoptera, Apoidea partim). *Molecular Ecology Resources*, 15. 10.1111/1755-0998.12363.

Shapiro SS, & Wilk, MB. 1965. An analysis of variance test for normality (complete samples). *Biometrika*, 52(3–4), 591–611. https://doi.org/10.1093/biomet/52.3-4.591

Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PDN. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). *Proceedings of the National Academy of Sciences of USA*, 103: 3657–3662.

Tujuba TF, Hausmann A, Sciarretta A. 2020. Revision of the *Orbamia* Herbulot, 1966 group of genera with description of two new genera, ten new species, and two new subspecies (Lepidoptera, Geometridae, Ennominae, Cassymini). *ZooKeys*, 929: 53–77. https://doi.org/10.3897/zookeys.929.50391.

vanVelzen R, Bakker FT, vanLoon JJA. 2007. DNA barcoding reveals hidden species diversity in *Cymothoe* (Nymphalidae). *Proceedings of the Netherlands Entomological Society Meeting*, 18: 95–103.

Werren JH, Baldo L, Clark ME. 2008. *Wolbachia*: master manipulators of invertebrate biology. *Nature Reviews Microbiology*, 6: 741–751.

Willmott, K. 2003. The genus Adelpha: its systematics, biology and biogeography (Lepidoptera: Nymphalidae: Limenitidini). Gainesville, FL, Scientific Publishers, 322pp.

Wirta H, Várkonyi G, Rasmussen C, Kaartinen R, Schmidt NM, Hebert PDN, Barták M, Blagoev G, Disney H, Ertl S, Gjelstrup P, Gwiazdowicz DJ, Huldén L, Ilmonen J, Jakovlev J, Jaschhof M, Kahanpää J, Kankaanpää T, Krogh PH, Labbee R, Lettner C, Michelsen V, Nielsen SA, Nielsen TR, Paasivirta L, Pedersen S, Pohjoismäki J, Salmela J, Vilkamaa P, Väre H,

672 Tschirnhaus Mv, Roslin T. 2015. Establishing a community-wide DNA barcode library as a new
673 tool for arctic research. *Molecular Ecology Resources*, 16 (3): 809-822.
674 https://doi.org/10.1111/1755-0998.12489.