Received: 21 January 2021 | Revised: 27 May 2021 | Accepted: 1 June 2021

DOI: 10.1111/1755-0998.13441

MOLECULAR ECOLOGY RESOURCES WILEY

RESOURCE ARTICLE

Genetic variation in neotropical butterflies is associated with sampling scale, species distributions, and historical forest dynamics

Natalí Attiná[1] | Ezequiel O. Núñez Bustos[1] | Darío A. Lijtmaer[1] | Paul D. N. Hebert[2] | Pablo L. Tubaro[1] | Pablo D. Lavinia[1,3]

[1]Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" (MACN–CONICET), Buenos Aires, Argentina

[2]Centre for Biodiversity Genomics, University of Guelph, Guelph, ON, Canada

[3]Universidad Nacional de Río Negro. CIT Río Negro (UNRN–CONICET). Sede Atlántica, Viedma, Río Negro, Viedma, Argentina

Correspondence
Natalí Attiná and Pablo D. Lavinia, Avenida Ángel Gallardo 470, C1405DJR, Buenos Aires, Argentina.
Emails: natali@attina.ar (NA); pablo.lavinia@conicet.gov.ar (PDL)

Funding information
Fundación Bosques Nativos Argentinos; Fundación Temaikèn; Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET); Richard Lounsbery Foundation; Natural Sciences and Engineering Resarch Council of Canada; Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación (Agencia I+D+i); Fundación Williams

Abstract

Previous studies of butterfly diversification in the Neotropics have focused on Amazonia and the tropical Andes, while southern regions of the continent have received little attention. To address the gap in knowledge about the Lepidoptera of temperate South America, we analysed over 3000 specimens representing nearly 500 species from Argentina for a segment of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene. Representing 42% of the country's butterfly fauna, collections targeted species from the Atlantic and Andean forests, and biodiversity hotspots that were previously connected but are now isolated. We assessed COI effectiveness for species discrimination and identification and how its performance was affected by geographic distances and taxon coverage. COI data also allowed to study patterns of genetic variation across Argentina, particularly between populations in the Atlantic and Andean forests. Our results show that COI discriminates species well, but that identification success is reduced on average by ~20% as spatial and taxonomic coverage rises. We also found that levels of genetic variation are associated with species' spatial distribution type, a pattern which might reflect differences in their dispersal and colonization abilities. In particular, intraspecific distance between populations in the Atlantic and Andean forests was significantly higher in species with disjunct distributions than in those with a continuous range. All splits between lineages in these forests dated to the Pleistocene, but divergence dates varied considerably, suggesting that historical connections between the Atlantic and Andean forests have differentially affected their shared butterfly fauna. Our study supports the fact that large-scale assessments of mitochondrial DNA variation are a powerful tool for evolutionary studies.

KEYWORDS
butterflies, diversification, DNA barcoding, Neotropics, South American forests, species traits

1 | INTRODUCTION

The Neotropics is arguably the most biodiverse region in the world, and unveiling how its remarkable species richness was generated and sustained through time has attracted researchers for centuries (Rull, 2020). During the Neogene and Quaternary, this region experienced dramatic landscape changes due to geotectonic events and cyclic climatic oscillations which shaped the evolutionary history of its biota (Cheng et al., 2013; Haffer, 1969; Hoorn et al., 2010; Ledo & Colli, 2017; Lundberg et al., 1998). Butterflies are one of the most conspicuous, widespread groups of animals in the Neotropics, with a species richness that accounts for ~40% of the global diversity of the group (~20,000 species; Lamas, 2004). This, in combination with their close association to vegetation and high sensitivity to environmental change, has made them a model group for the study of diversification patterns in the region (Bonebrake et al., 2010; Brower & Garzón-Orduña, 2020; Ebel et al., 2015; Lamas, 2004). However, most previous studies on Neotropical butterflies have examined Amazonia or the tropical Andes (Blandin & Purser, 2013; Chazot et al., 2016; Elias et al., 2009; Garzón-Orduña et al., 2014; Penz et al., 2015). By comparison, the butterfly fauna of temperate southern South America has received little attention (New & Samways, 2014).

More than 1200 species of butterflies are known from Argentina (Klimaitis et al., 2018), most in the Atlantic and Andean Forests (Figure 1), two biodiversity hotspots and priority areas for conservation (Olson & Dinerstein, 2002). Extending along the Brazilian coast, the Atlantic Forest reaches its southerly limit in Misiones province in northeastern Argentina, while the Central Andean forests extend from Peru south into northwestern Argentina (Godoy-Bürki et al., 2014; Ribeiro et al., 2009). The Andean forests (and adjacent Amazonia) are currently isolated from the Atlantic Forest by more open and drier environments (Caatinga, Cerrado, Chaco) that extend from northeastern Brazil to northern Argentina and are collectively known as the open vegetation corridor (OVC). In Argentina in particular (Figure 1), the Atlantic and Andean forests are separated by the xerophytic open forests and steppes of the Dry Chaco and by the Humid Chaco, a complex mosaic of savannas, grasslands, wetlands and gallery forests (Cabrera, 1976). Despite their current isolation, the Atlantic and Andean forests have been cyclically and transiently connected in the past (Ledo & Colli, 2017), promoting the interchange of their biota and creating a very interesting biogeographic scenario for evolutionary studies. While previous studies have investigated how these connections affected the diversification history of diverse vertebrate lineages (Costa, 2003; Lavinia et al., 2019; Prates et al., 2017; Trujillo-Arias et al., 2020), most invertebrate groups, including butterflies, have been neglected.

Nearly two decades ago, Hebert et al., (2003) proposed that the analysis of short, standardized segments of DNA, such as the 5′ region of the mitochondrial cytochrome *c* oxidase subunit I (COI) gene, would be effective in species discrimination. Termed DNA barcoding, many studies have now demonstrated the efficacy of this approach across most metazoan lineages and its collateral value for ecological and evolutionary analyses (Barreira et al., 2016; Dapporto et al., 2019; Kress et al., 2015). Among insects, the order Lepidoptera has received particular intensive study, work which has established that COI is very effective for species discrimination and specimen identification. As a result, DNA barcodes have been used for large-scale assessments of cryptic diversity and geographic patterns of genetic variation in Lepidoptera (Dincă et al., 2015; Gaytán et al., 2020; Hausmann et al., 2013; Huemer et al., 2014; Lavinia et al., 2017a), as well as to enable rapid diversity inventories and to delineate putative species in poorly-known groups (Kekkonen & Hebert, 2014; Zenker et al., 2016).

To address the gap in knowledge about the butterflies of temperate South America, we generated over 1000 COI sequences for more than 200 butterfly species from western Argentina with a focus on the Andean forests (Figure 1). We first used these records to test the effectiveness of DNA barcodes for species discrimination and identification within this region. We then combined these sequences with those from the butterflies of eastern Argentina (Lavinia et al., 2017a), mainly from the Atlantic Forest (Figure 1). This merged data set (>3000 COI sequences from almost 500 species) was used to assess the impact of increased spatial and taxonomic coverage on the performance of DNA barcoding, and to evaluate the spatial patterning of mitochondrial DNA variation across Argentina with emphasis on the Atlantic and Andean forests.

2 | MATERIALS AND METHODS

2.1 | Sampling, laboratory protocols and data sets

A total of 1150 specimens representing 241 species and 155 genera (Table 1; Tables S1 and S2) were collected using insect nets and fruit bait traps between 2014 and 2016 in nine provinces of Argentina (Figure 1). Since more than 99% of these specimens were collected in northwestern Argentina (NWA: Catamarca, Jujuy, La Rioja, Salta, Santiago del Estero and Tucumán provinces), including over 96% from the Andean forests, we refer to them as the NWA data set. All specimens were identified by ENB based on external morphology and following Klimaitis et al., (2018), and are deposited in the Museo Argentino de Ciencias Naturales "Bernardino Rivadavia" (MACN). Nine specimens (0.78%) lacked a species assignment but were identified to a generic level. DNA extraction and amplification were performed at either the MACN or the Centre for Biodiversity Genomics (CBG) following standard protocols (Ivanova et al., 2006; Kress & Erickson, 2012). A 658 bp region near the 5′ end of COI was amplified using either the primers LepF1 and LepR1 (Hebert et al., 2004), or the primer cocktails C_LepFolF and C_LepFolR (Folmer et al., 1994; Hebert et al., 2004). Amplicons were bidirectionally sequenced at the CCDB and sequences were edited and aligned using CodonCode Aligner (CodonCode Corporation) and MEGA 5.0 (Tamura et al., 2011). A sequence was not recovered from 120 specimens (10.4%) and eight sequences were discarded because of low quality (6) or contamination (2). As a result, the final NWA data set included 1022 sequences from 146 genera and 213 species (Table 1;

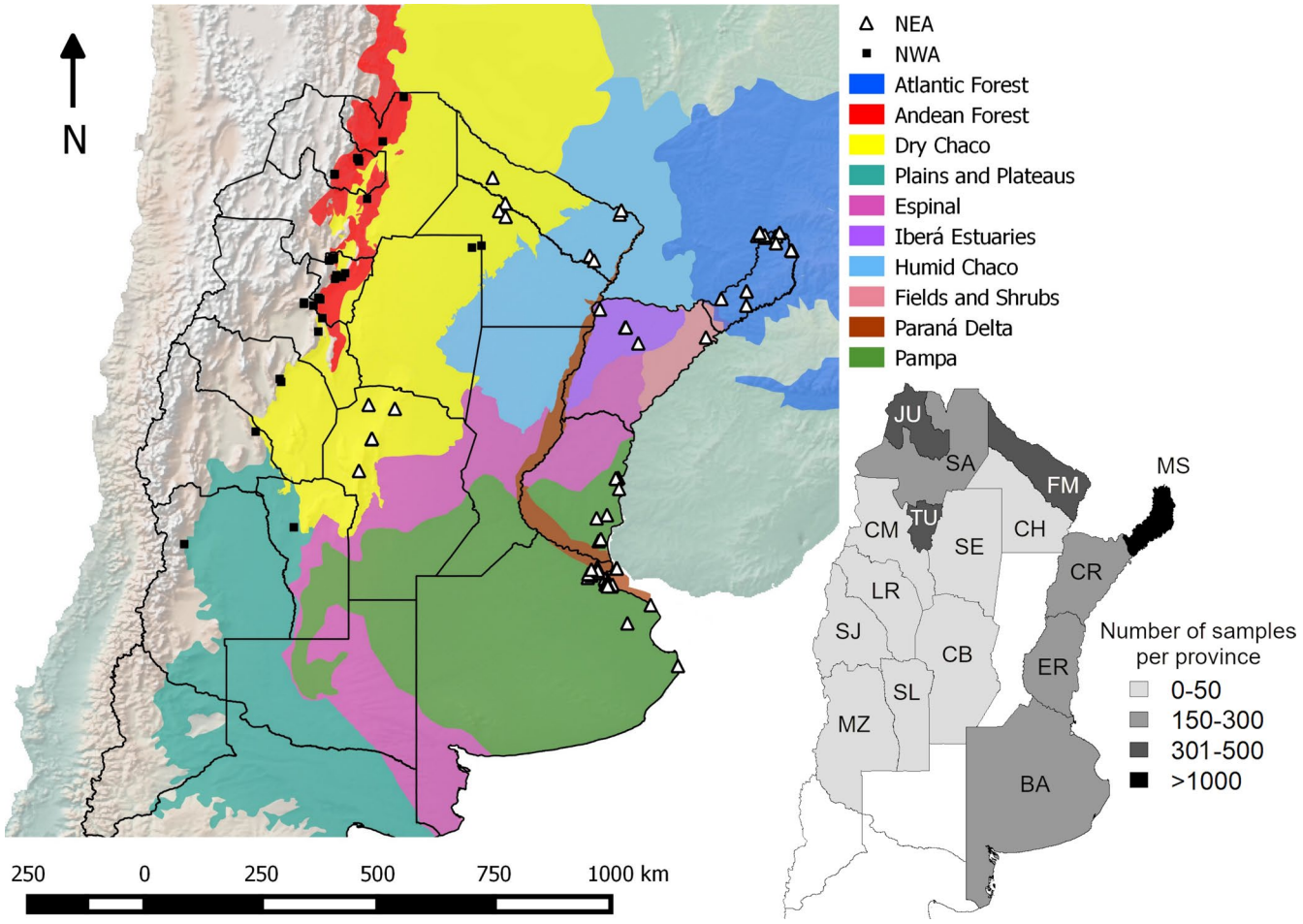

FIGURE 1 Sampling localities for the 1150 butterflies that compose the northwestern Argentina data set (NWA; black squares), and for the 2161 specimens from the northeastern Argentina data set (NEA; white triangles). Ecoregions relevant for the analyses are indicated by different colours. The map in the lower right depicts the total number of specimens collected in each of 16 provinces: BA, Buenos Aires; CB, Córdoba; CH, Chaco; CM, Catamarca; CR, Corrientes; ER, Entre Ríos; FM, Formosa; JU, Jujuy; LR, La Rioja; MZ, Mendoza; MS, Misiones; SA, Salta; SJ, San Juan; SL, San Luis; SE, Santiago del Estero; TC, Tucumán

TABLE 1 Numbers of individuals and species analyzed for six butterfly families from the northwestern Argentina (NWA) data set

Family	Specimens/species sampled	Specimens/species sequenced	Sequencing success (%) for specimens/species
Hesperiidae	359/88	322/79	89.7/89.8
Lycaenidae	70/35	50/27	71.4/77.1
Nymphalidae	475/75	433/71	91.1/94.6
Papilionidae	28/8	26/7	92.9/87.5
Pieridae	185/29	158/23	85.4/79.3
Riodinidae	33/6	33/6	100/100
Total	1150/241	1022/213	88.9/88.4

Note: Success in recovery of a COI sequence reflects the final data set used for the analyses.

Table S1). On average, 4.8 sequences were analysed per species (range 1–21), with 156 species represented by more than one specimen. The 57 species represented by a single sequence (singletons) accounted for 5.6% of the records and 27% of the species.

We previously assembled a database with 2020 sequences derived from 417 butterfly species of Argentina (Lavinia et al., 2017a). Since nearly 90% of these sequences derived from specimens collected in northeastern Argentina (NEA: Chaco, Corrientes, Formosa and Misiones provinces), including almost 60% from the Atlantic Forest in Misiones province (Figure 1), we refer to them as the NEA data set. In addition to separately analysing the NWA data set, we compared it to and combined it with the NEA library to allow the

analyses described below. Unless otherwise indicated, analyses included all specimens from each database as long as they were successfully sequenced and passed our quality filters.

2.2 | Genetic distances and gene trees

Only sequences longer than 500 bp and with less than 1% ambiguous sites were included in the analyses. Uncorrected genetic distances (p-distances) and Kimura 2-parameter (K2P) divergences (Kimura, 1980) were computed within and between species using SPIDER (Brown et al., 2012) in R 3.5.2 (R Core Team, 2018). Because results were almost identical for the two distance metrics and K2P is the most common substitution model in DNA barcoding studies, we only report the latter values. We then ascertained the distance of each individual to its furthest conspecific and its closest nonconspecific (i.e., nearest neighbour) to assess the presence/absence of a barcode gap: a separation between maximum intraspecific and minimum interspecific divergences that is key for species discrimination and identification. Singletons were considered distinguishable from other species when they possessed an unique COI sequence that was separated from the nearest neighbour taxon in gene trees ("Tree-Based Identification" approach; Wilson et al., 2013).

A Neighbour-Joining (NJ) gene tree was generated on BOLD (http://www.boldsystems.org; Ratnasingham & Hebert, 2007) using the K2P distance model and the pairwise deletion option for missing data. Node support values were computed in MEGA through 1000 bootstrap pseudoreplicates and printed on the NJ tree (Appendix S1). We also built a maximum likelihood (ML) gene tree with RAxML 8.1.22 (Stamatakis, 2014) based on 100 independent ML tree searches under the GTRGAMMA model of evolution, and 1000 rapid bootstrap pseudoreplicates for node support values. The latter were printed on the best-scoring ML tree (Appendix S2). While inferring the phylogenetic relationships among species was beyond the scope of our study, these trees were used to assess the distinctiveness of singletons and to measure the support of terminal nodes and intraspecific clades.

2.3 | Specimen identifications

We simulated a sequence-based identification process by treating each sequence as an unknown specimen and querying it against the COI library for NWA. A species name was assigned to each query based on three criteria: Best Match (BM) and Best Close Match (BCM) following Meier et al., (2006), and the BOLD Identification Criterion (BIC) from BOLD's ID engine (Ratnasingham & Hebert, 2007). BM assigns a species name to the query according to its closest match in the library irrespective of genetic divergence (except when two or more species are equally distant to the query; in that case the identification is considered as ambiguous). On the other hand, BCM and BIC only assign a species name to the query if its genetic divergence from the closest match is below a set threshold. BCM considers only the closest match below the threshold, while all sequences below the threshold are considered under BIC. As a result, an identification made by BCM is correct when the closest match below the threshold is from the same species as the query, but incorrect when it is from another species, and ambiguous when two or more species are equally distant from the query. With BIC, the identification is correct when all sequences below the threshold derive from the same species as the query, but incorrect when they all correspond to another species, and ambiguous when sequences from multiple species appear below the threshold. Lastly, for both BCM and BIC, queries remain unidentified when there is no match below the threshold. All simulations were carried out in SPIDER.

Four different thresholds were employed for BCM and BIC: (i) the 95th percentile of all intraspecific distances (2.02%), (ii) BOLD's ID engine threshold of 1%, (iii) the genetic distance (1.50%) that minimized the sum of false positive and false negative identifications, and (iv) the lowest value (1.64%) in a density plot of all genetic distances which should correspond to the transition between intra- and interspecific distances. We used SPIDER to obtain the latter two values with the functions "threshVal" and "localMinima" respectively (Figures S1 and S2). Singletons remained in the library as potential matches for other sequences but were not used as queries. We performed all analyses using both K2P and uncorrected p-distances but only report the former since results were almost identical.

2.4 | DNA barcode performance over large geographic distances

To assess the influence of geographic distance on the performance of DNA barcoding, we repeated the sequence-based identification process but used the NEA data set as reference sequences to identify individuals from NWA. Sampling localities from NWA and NEA databases are separated by 956 km on average (range = 136–1649 km), and the mean geographic distance between conspecifics from different data sets is 933 km (range = 195–1646 km). A total of 1022 sequences representing 213 species from the NWA database were queried against the NEA library using BLAST 2.10 (Altschul et al., 1990) on command line. A species name was assigned to each query based on the same three criteria previously explained. Regarding the thresholds for BCM and BIC, we used the BOLD's ID engine threshold of 1% and those reported by Lavinia et al., (2017) for the NEA library: 0.85% ("threshVal"), 1.39% (95th percentile of intraspecific distances) and 2.06% ("localMinima").

2.5 | An expanded COI library for the butterflies of Argentina

We merged data from the NEA and NWA databases into a single COI library (NEA + NWA) composed of 3042 sequences from

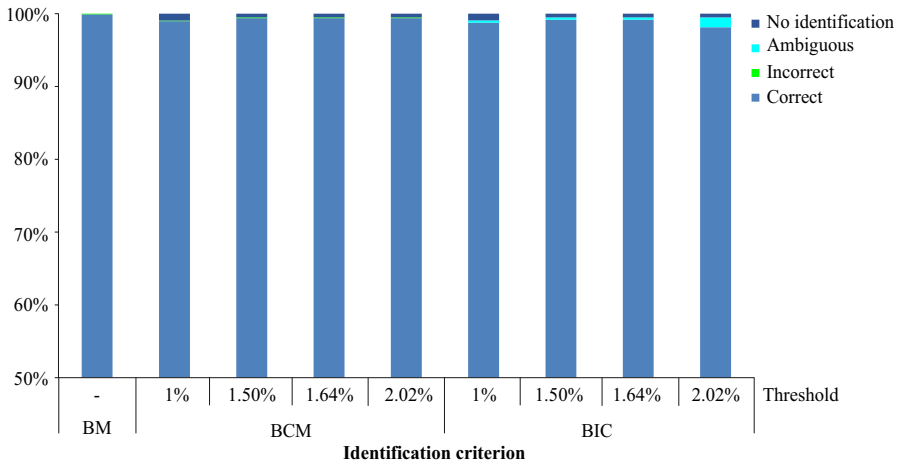

FIGURE 2 Results from the sequence-based identification simulations based on 965 queried individuals from 156 species from the NWA database. The total number of specimens (% in parenthesis) assigned to each of the four categories are shown for three identification criteria (BM, Best Match; BCM, Best Close Match; BIC, BOLD Identification Criterion) and four threshold values: 1.00% (BOLD's identification threshold), 1.50% ("threshVal"), 1.64% ("localMinima") and 2.02% (95th percentile of all intraspecific distances). Singletons were not used as queries but remained as potential matches for other sequences. Note that the y-axis starts at 50%

495 species and 283 genera. On average, 6.15 sequences were analysed per species (range 1–46) with 385 species represented by two or more individuals. Singletons (110) represented 22% of the species and 3.62% of the sequences. We recalculated all summary statistics and simulated a sequence-based identification process for this merged data set. Simulations employed the same three identification criteria and four sequence thresholds introduced above. All threshold values other than BOLD's 1% were re-estimated for the NEA + NWA database: 0.85% ("threshVal"), 1.12% ("localMinima"), and 2.02% (95th percentile of intraspecific distances). Lastly, we estimated new NJ and ML gene trees (Appendix S3 and S4 respectively) for the NEA + NWA data set following the same procedures described above.

2.6 | Spatial patterns of intra- and interspecific variation across Argentina

We first compared levels of within and between species variation for the three COI libraries: NEA, NWA, and NEA + NWA. We repeated these analyses for a subset of 135 species shared by NWA and NEA databases, and also compared the degree of intraspecific differentiation between versus within regions. Next, we focused on 85 of these 135 species that occur in both the Atlantic (NEA) and Andean forests (NWA). Of these, 27 species have a disjunct distribution between forests, while the other 58 possess a more continuous range (Klimaitis et al., 2018). We compared the level of intraspecific divergence between forest populations of species with disjunct distributions versus those with continuous distributions. We then examined in more detail a subset of species with relatively high genetic differentiation between forests. A species was included in this subset if it met one or both of the following criteria: (a) maximum intraspecific divergence between forests >0.98% (the average distance to the furthest conspecific for NEA + NWA); (b) individuals from the Atlantic Forest and the Andean forests were clustered into two or more distinct, well-supported (bootstrap support values ≥80%) mitochondrial lineages in the NEA + NWA gene trees, regardless of sequence divergence between them and irrespective of their relationships with conspecifics from other ecoregions. Lastly, we dated divergence events between populations from the Atlantic and Andean forests based on COI sequence variation. We focused only on species with a clear split between forests and applied both a slow (2.3% sequence divergence per million years; Brower, 1994) and a fast (3.36% sequence divergence per million years; Papadopoulou et al., 2010) COI molecular rate.

Statistical differences were assessed through Welch's *F* tests followed by Games-Howell post hoc tests carried out in SPSS 25.0.

3 | RESULTS

3.1 | Genetic distances and gene trees

Mean intraspecific divergence among the 156 species (110 genera) represented by two or more specimens (6.19 sequences per species on average) was 0.29% (95% confidence interval [CI]: 0.20%–0.37%). By contrast, the average interspecific distance among 123 pairs of congeneric species (106 species from 39 genera) was 7.24% (95% CI: 6.38%–8.10%). More importantly, mean divergence (7.56%, 95% CI: 7.41%–7.70%) to the nearest neighbour species was nearly 13× the average distance to the furthest conspecific (0.60%, 95% CI: 0.54%–0.66%). As a result, a barcode gap was present for all species with multiple individuals except for *Calycopis* sp. 1 and *Urbanus pronta* (Figure S3). The lowest interspecific distance registered was 0.15% between *Calycopis* sp. 1 and the single

specimen of *Calycopis caulonia*, followed by 1.70% between *Tegosa* sp. and *Tegosa claudina*. Except for *C. caulonia*, all singletons had a unique COI sequence that distinguished them from their nearest neighbour in the gene trees (Appendix S1 and S2). Distance to the nearest neighbour averaged 7.50% (95% CI: 6.89%–8.12%) among singletons, with 96.5% (55 species) showing a minimum interspecific divergence that was greater than the lower 5% of all congeneric distances (3.36%).

3.2 | Specimen identification simulations

BM generated 99.9% correct and 0.1% incorrect identifications while identification success with BCM and BIC varied only slightly depending on the threshold value employed (Figure 2; Table S3). BCM with a threshold value between 1.5% and 2.02% produced 99.38% correct identifications, 0.10% incorrect identifications, and 0.52% unidentified queries (Figure 2; Table S3). With BOLD's ID engine threshold of 1%, correct identifications decreased marginally (98.96%) due to a higher incidence (0.93%) of unidentified queries. Results with BIC were almost identical, the main differences being a lack of incorrect identifications and the appearance of some ambiguous calls (Figure 2; Table S3). Use of the "localMinima" (1.64%) and "ThreshVal" (1.5%) thresholds to guide decisions returned 99.17% correct identifications, 0.31% ambiguous identifications, and 0.52% unidentified queries. Adoption of BOLD's threshold (1%) led to 98.76% correct identifications, reflecting the higher incidence (0.93%) of unidentified queries while ambiguous calls remained the same. Finally, when employing the 95th percentile of all intraspecific distances (2.02%) as a threshold, correct identifications decreased slightly (98.13%) because ambiguous assignments increased to 1.35%.

3.3 | DNA barcode performance over large geographic distances

Based on the use of the NEA database as reference and the BM criterion, most (98.11%) specimens from the NWA data set were correctly identified while 1.73% were incorrect and 0.16% were ambiguous (Figure 3). BCM and BIC generated identical results when threshold values ≤1% were employed. With a threshold of 0.85%, both methods produced 77.04% correct, 0.47% incorrect, and 0.16% ambiguous identifications with 22.33% unidentified queries. Results were very similar with a threshold of 1%, with only a slightly higher percentage (77.99%) of correct identifications and slightly fewer (21.38%) unidentified queries (Figure 3; Table S4). Identification success with the BCM criterion increased at the highest thresholds (1.39%, 2.06%) because the number of unidentified queries decreased while the incidence of incorrect and ambiguous calls remained the same (Figure 3; Table S4). When a threshold of 1.39% was employed, BCM produced 84.28% correct identifications with 15.09% unidentified queries, while the use of a 2.06% threshold raised the correct calls to 87.74% and decreased unidentified queries to 11.64% (Figure 3; Tables S4 and S5). BIC showed a similar pattern, but the increase in identification success was lower due to a rise in the number of ambiguous identifications as thresholds increased. Ambiguous calls reached 2.83% and 4.87% with the 1.39% and 2.06% thresholds respectively, while the proportion of unidentified queries was the same as with BCM (Tables S4 and S5). As a result, BIC produced between 81.60% and 83.02% correct identifications with the 1.39% and 2.06% thresholds respectively (Figure 3; Table S4).

We then queried 386 individuals from 78 species from the NWA database that lacked a conspecific match in the NEA data set and that were excluded from the previous sequence-based identification

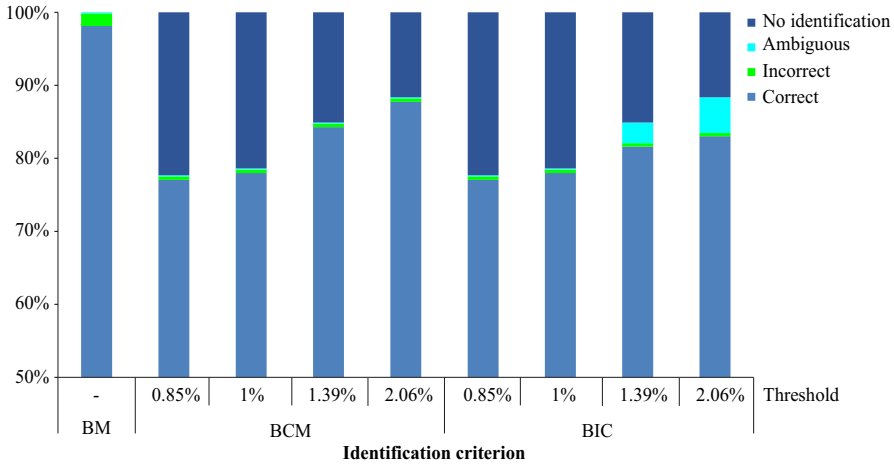

FIGURE 3 Results from the sequence-based identification of specimens from the NWA database using the NEA data set as reference library. Percentages are based on 636 queries for 135 species shared between the two databases (BM, Best Match; BCM, Best Close Match; BIC, BOLD Identification Criterion). The four thresholds implemented correspond to BOLD's 1% and the three values reported by Lavinia et al., (2017) for the NEA library: 0.85% ("threshVal"), 1.39% (95th percentile of all intraspecific distances) and 2.06% ("localMinima"). Note that the *y*-axis starts at 50%

procedure. Only a small number of specimens (3–13) were incorrectly identified with both BCM and BIC (Table S4). Involving just four species (*Hectarides lamarchei*, *Codatractus alcaeus*, *Tegosa* sp. 1, *Hylephila ancora*), these specimens were always assigned to a congeneric species.

When the identification process was repeated in the reverse direction (i.e., querying sequences from the NEA library against that of NWA), very similar results were obtained (Table S6).

3.4 | A comprehensive COI database for the butterflies of Argentina

Based on the merged NWA+NEA library (Table 2), mean intraspecific divergence was 0.43% (95% CI: 0.35%–0.50%) among the 385 species (222 genera) represented by two or more COI sequences (7.62 individuals per species on average). By comparison, average distance among 518 congeneric species pairs (310 species in 99 genera) was 7.08% (95% CI: 6.61%–7.55%). Mean divergence to the nearest neighbour (6.55%, 95% CI: 6.46%–6.64%) was nearly 7× the average distance to the furthest conspecific (0.98%, 95% CI: 0.94%–1.03%). As a result, a barcode gap was present in 96.88% of the species represented by multiple individuals with only 12 species showing maximum intraspecific distances higher than the distance to their nearest neighbour (Figure S4). Only 14 species were recovered as paraphyletic in the NJ (Appendix S3) and/or ML (Appendix S4) gene trees, and 10 of them lacked a barcode gap. Singletons showed a similar pattern as the mean divergence to their nearest neighbour was 6.90% (95% CI: 6.53%–7.26%), with 98% of the 108 species showing a minimum interspecific divergence higher than the lower 5% of all congeneric distances (3.13%). Consistently, all singletons were clearly distinct from their nearest neighbour in the gene trees. Sequence-based identification simulations delivered high identification success with varied identification criteria and threshold values (Table S7). Correct identifications ranged from 93.62% with BIC at the highest threshold value (2.02%) to 99.45% with BM. BCM and BIC produced always the same percentage of no identifications, which decreased from 2.01% at the lowest threshold (0.85%) to 0.95% at the highest (2.02%). Correct identifications increased together with the threshold value when BCM was employed, reaching 98.67% at the highest threshold. By comparison, increasing the threshold value raised the incidence of ambiguous identifications from 0.65% to 5.39% with BIC. As a result, correct identifications with BIC decreased from 97.34% to 93.62% as the threshold increased (Table S7).

3.5 | Geographic patterns of intra- and interspecific variation across northern Argentina

Genetic variation within and between species differed significantly among the NEA, NWA, and NEA + NWA data sets (Table 3; Table S8). Mean intraspecific divergence was higher in NEA + NWA than in either NWA or NEA, but differences were only significant between the first two; no significant differences were found between NEA and NWA. The pattern was clearer for the maximum intraspecific distance which was significantly higher in NEA + NWA than in either NEA or NWA, and also in NEA than in NWA (Figure 4a). No differences were found among data sets in the mean divergence among congeneric species. However, minimum interspecific distance was significantly lower in NEA + NWA than in either NEA or NWA, and also in NEA than in NWA (Figure 4b).

The 135 species shared by the NEA and NWA databases showed the same trend, but with sharper differences (Table 3; Tables S8 and S9). Distance to the nearest neighbour was significantly lower in NEA + NWA than in the individual data sets, as well as in NEA than in NWA. Mean intraspecific distance for the species in NEA + NWA was significantly higher than within NEA or NWA with no significant differences between the latter. Maximum intraspecific distance was significantly higher in NEA + NWA than in either individual library, and also in NEA than in NWA. At the same time, intraspecific variation was significantly higher between than within regions. Both mean and maximum distances between conspecifics from NEA and NWA were significantly higher than those for NEA and NWA (Figure 4c).

Among the 135 species shared by the NEA and NWA data sets (Table S9), 85 occur in the currently isolated Atlantic Forest (NEA) and Andean forests (NWA). Mean and maximum intraspecific divergence values between forest populations were both significantly higher for the 27 species with disjunct distributions than for the 58 species with a continuous range (Figure 4d, Table 3; Table S8). Nearly 60% of species with disjunct distributions showed high genetic differentiation between the Atlantic and Andean forests versus 47% for species with continuous ranges (Table 4). Both mean and maximum intraspecific divergence between forests for this subset of divergent species were also significantly higher among disjunctly distributed taxa than among those with a continuous range (Table 3).

Finally, all splits between populations from the Atlantic Forest and the Andean forests were dated to the Pleistocene (Table 4). Populations of species with a disjunct distribution were estimated to have diverged from 0.83 Ma (95% CI: 0.53–1.13) to 1.21 Ma (95% CI: 0.78–1.65) depending on whether a fast or slow rate was adopted, while those from species with continuous ranges appear to have diverged from 0.46 Ma (95% CI: 0.25–0.67) to 0.67 Ma (95% CI: 0.37–0.97).

4 | DISCUSSION

We assembled and analysed over 1000 COI sequences representing 213 butterfly species mainly from northwestern Argentina. This library (NWA) was then combined with the one for the butterflies of northeastern Argentina (NEA), providing coverage for 42% of the butterfly fauna of the country (Klimaitis et al., 2018; Lavinia et al., 2017a). We first examined the effectiveness of DNA barcodes for discrimination and identification of the species found in northwestern Argentina. We then analysed the insights

TABLE 2 Summary of the individuals and species sampled and sequenced for the seven butterfly families present in Argentina

Family	Species in Argentina	Specimens/species sampled	Specimens/species sequenced	Sequencing success (%) for specimens/species	Species covered (%)
Hesperiidae	505	958/202	901/193	94.1/95.6	40.0
Lycaenidae	195	194/59	171/52	88.1/88.1	30.3
Nymphalidae	332	1522/182	1392/173	91.5/95.1	54.8
Papilionidae	34	95/16	89/14	93.7/87.5	47.1
Pieridae	72	383/36	333/29	86.9/80.6	50.0
Riodinidae	115	159/33	156/33	98.1/100	28.7
Hedylidae	1	-	-	-	-
Total	1254	3311/528	3042/495	91.9/93.8	42.1

Note: Numbers are based on the NEA + NWA data set.

provided by the use of both libraries into both the effect of spatial and taxonomic coverage on DNA barcoding performance, as well as on the geographic patterns of genetic variation in butterflies of subtropical South America, with an emphasis on the historical relationship between the currently isolated Atlantic and Andean forests.

4.1 | Species discrimination and identification within NWA

Maximum intraspecific distance was lower than the minimum distance to the nearest neighbour for most species represented by multiple individuals, while all singletons possessed a COI sequence distinct from those of any other species (Wilson et al., 2013). Hence, COI was extremely effective in discriminating the butterfly species found in northwestern Argentina. There were three exceptions: *Urbanus pronta*, *Calycopis* sp. 1 and *Calycopis caulonia*. The first one was paraphyletic in the NJ gene tree (monophyletic with node support <50% in the ML topology), with two specimens being more closely related with *Urbanus* sp. 1 than to their other conspecifics. The second case involved three specimens of *Calycopis* sp. 1 which formed a cluster with low (56%) bootstrap support and possessed a minimum interspecific distance of 0.15% to the singleton *Calycopis caulonia*. It is worth noting that these cases involve taxa that could only be identified to a generic level based on external morphology and wing coloration patterns, suggesting that the incapacity of DNA barcodes to discriminate these species could be actually a consequence of taxonomic inaccuracy or limitations (Mutanen et al., 2016). At the same time, COI might not be able to distinguish these taxa. For instance, species assignments are certainly difficult in the genus *Calycopis*, where taxa are not only hard to differentiate based on morphology and male genitalia, but also mitochondrial variation among species is minimal and inconsistent with morphology or sampling localities (Cong et al., 2016; Duarte & Robbins, 2005; Lavinia et al., 2017a).

Simulation tests showed that the current COI library is extremely effective in specimen identification with success exceeding 98% irrespective of the identification criterion and divergence threshold employed. There were almost no incorrect assignments and a low frequency (<1%) of unidentified queries or ambiguous calls (<1.5%). Best Match (BM) generated the highest percentage of correct identifications (99.9%) with only one specimen being incorrectly identified. Because this criterion assigns a species name to each query regardless of its level of sequence divergence to its closest match in the reference library, this reflects the near absence of problematic taxa in the NWA database. Best Close Match (BCM) and BOLD Identification Criterion (BIC) approaches, which incorporate sequence divergence thresholds, returned similarly high percentages of correct identifications. Because higher thresholds help to identify species with deeper intraspecific divergence that otherwise will remain unidentified, correct identifications increased as the threshold value raised. However, when the highest threshold was coupled with BIC this effect was counteracted by more ambiguous calls. Since higher thresholds may fall within the overlap between intra- and interspecific distances, the probability increases that closely related species will fall under the threshold when larger values are implemented. This raises the frequency of ambiguous identifications with BIC but not BCM, since only the former considers all sequences under the threshold to assign a species name to the query.

The identification success with the NWA database was only slightly higher than that of the NEA library, which ranged from 95.6% to 99.42% (Lavinia et al., 2017a). Interestingly, Lavinia et al., (2017a) suggested that lower thresholds were more effective for the butterflies of NEA, mainly because the incidence of ambiguous identifications with BIC always increased with the threshold value, decreasing identification success. This pattern was absent for NWA as identification success with BIC only declined at the highest threshold (2.02%). This difference reflects the fact that minimum interspecific distances are lower and maximum intraspecific distances higher among NEA than NWA butterflies, therefore reducing the barcode gap and increasing identification ambiguity in the former.

Our results emphasize that optimal sequence thresholds vary depending on the properties of each library that will, in turn, be impacted by geographic and taxonomic coverage. Additionally, patterns of mitochondrial DNA variation within a set of species will

TABLE 3 Intra- and interspecific genetic distances (K2P, %) for different data sets

Genetic distance	Data set	Mean	95% Confidence interval of the mean	
			Lower limit	Upper limit
Mean intraspecific	NWA	0.29	0.20	0.37
	NEA	0.31	0.24	0.39
	NEA + NWA	0.43	0.35	0.5
	NWA (shared)	0.33	0.21	0.44
	NEA (shared)	0.35	0.25	0.45
	NEA + NWA (shared)	0.71	0.55	0.86
	NEA vs NWA (shared)	1.02	0.78	1.26
	Atlantic vs Andean forests (disjunct)	1.65	1.01	2.30
	Atlantic vs Andean forests (continuous)	0.78	0.51	1.04
	Atlantic vs Andean forests (disjunct, divergent)	2.64	1.88	3.40
	Atlantic vs Andean forests (continuous, divergent)	1.40	0.92	1.88
Maximum intraspecific	NWA	0.60	0.54	0.66
	NEA	0.69	0.65	0.74
	NEA + NWA	0.98	0.94	1.03
	NWA (shared)	0.71	0.63	0.78
	NEA (shared)	0.84	0.77	0.91
	NEA + NWA (shared)	1.36	1.29	1.43
	NEA vs NWA (shared)	1.16	1.10	1.23
	Atlantic vs Andean forests (disjunct)	1.73	1.51	1.95
	Atlantic vs Andean forests (continuous)	1.05	0.96	1.13
	Atlantic vs Andean forests (disjunct, divergent)	2.62	2.36	2.88
	Atlantic vs Andean forests (continuous, divergent)	1.60	1.48	1.72
Mean interspecific	NWA	7.24	6.38	8.10
	NEA	7.18	6.64	7.71
	NEA + NWA	7.08	6.61	7.55
Minimum interspecific	NWA	7.56	7.41	7.70
	NEA	6.91	6.81	7.01
	NEA + NWA	6.55	6.46	6.64
	NWA (shared)	7.35	7.16	7.55
	NEA (shared)	6.89	6.74	7.05
	NEA + NWA (shared)	6.41	6.28	6.54

Note: The mean distance together with its corresponding 95% confidence interval is reported in each case. Genetic distances within NEA, NWA and NEA + NWA databases are shown for the complete data sets and for the 135 species shared by the NEA and NWA libraries. In the latter case, genetic distances between conspecifics from different data sets (NEA vs NWA) are also shown. Intraspecific divergence values between populations from the Atlantic and Andean forests are also calculated for 27 species with disjunct distributions and for 58 continuously distributed species, and for a subset of species with high genetic differentiation between forests.

be shaped by the evolutionary forces that have shaped the genetic diversity of that particular regional fauna (Bergsten et al., 2012; Dapporto et al., 2019; Gaytán et al., 2020; Lavinia et al., 2017a). As a result, optimal thresholds should be estimated independently for each library with the recognition that these might change as the database grows.

4.2 | Spatial patterns of intra- and interspecific mitochondrial DNA variation

Expanding the geographic and taxonomic sampling of COI libraries can increase intraspecific variation due to phylogeographic structure and isolation by distance, and reduce interspecific divergence

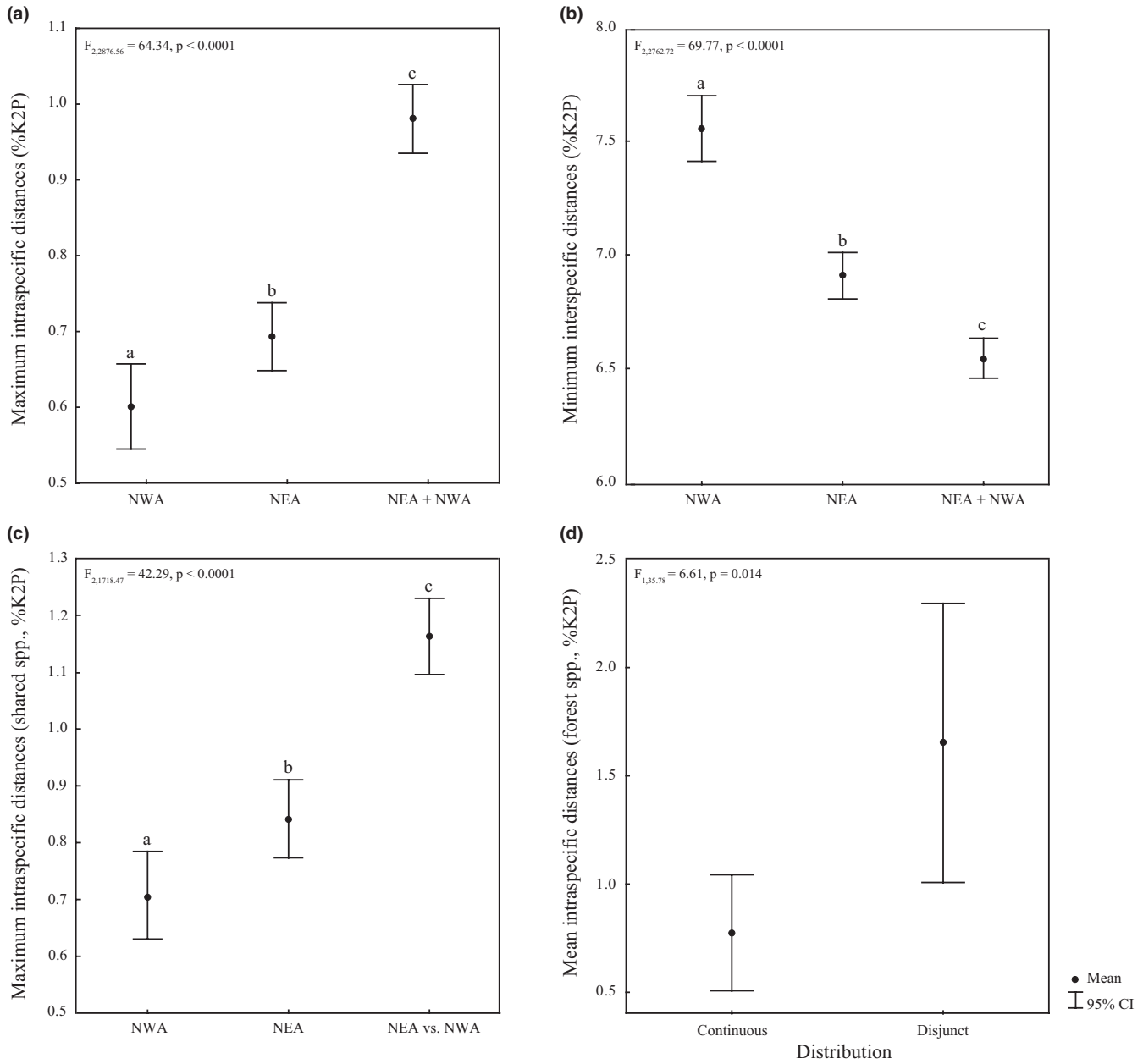

FIGURE 4 Maximum intraspecific distance (a) and minimum interspecific distance (b) for the three data sets. (c) Maximum intraspecific distance for the 135 species shared by the NEA and NWA within each data set and between them (NEA vs. NWA). (d) Mean intraspecific distance for 85 forest species, 27 of which have a disjunct distribution between the Atlantic and Andean forests, while the other 58 possess a continuous range. Welch's *F* statistic and the significance of the tests (*p*) are shown within each panel. Letters in panels (a) to (c) indicate pairs of data sets that differed significantly in post hoc pairwise comparisons

as more closely related and disjunctly distributed taxa are encountered (Barco et al., 2016; Bergsten et al., 2012; Gaytán et al., 2020; Marín et al., 2017; Virgilio et al., 2010). Consistently, we found that the separation between intra- and interspecific variation narrows as the spatial and taxonomic coverage rises across the three butterfly assemblages analysed here.

The NWA library consists of over 1000 specimens separated on average by 307 km and representing 213 species, while the NEA database covers 417 species represented by more than 2000 sequences and a mean distance between sampling localities of 581 km. The NEA + NWA library includes nearly 500 species represented by more than 3000 specimens from localities separated by 719 km on average, and with a mean distance of 956 km and a maximum of 1649 km between localities in eastern and western Argentina. As expected, the ratio between maximum intraspecific and minimum interspecific distances nearly doubled from 0.08 for the library with the smallest scale (NWA) to 0.15 for that with the largest (NEA+NWA), with NEA showing an intermediate value (0.10).

Species with relatively high intraspecific variation were comparatively more common in the NEA library than in that of NWA.

TABLE 4 Species with high genetic differentiation between the Atlantic (NEA) and the Andean (NWA) forests

Species	Distribution between forests	Mean divergence (%K2P)			Max divergence (%K2P)			Distinct mitochondrial lineages
		Within Atlantic Forest	Within Andean forests	Between forests	Within Atlantic Forest	Within Andean forests	Between forests	
Actinote pellenea	Continuous	0.12	0.99	1.66	0.30	2.33	2.17	NO
Battus polydamas	Continuous	0.00	0.96	0.53	0.00	2.71	2.71	NO
Biblis hyperia	Continuous	0.41	0.20	0.54	0.77	0.61	1.08	NO
Cymaenes gisca	Continuous	0.08	0.24	0.21	0.15	1.23	1.08	NO
*Dircenna dero**	Continuous	0.07	0.15	0.42	0.30	0.15	0.61	YES
*Eantis thraso**	Continuous	0.00	0.10	1.89	0.00	0.31	2.01	YES
Emesis ocypore	Continuous	0.23	0.18	0.81	0.47	0.30	1.10	NO
Emesis russula	Continuous	0.00	0.00	1.08	0.00	0.00	1.10	NO
Eurema albula	Continuous	0.00	1.15	0.83	0.00	2.33	2.01	NO
Eurema elathea	Continuous	0.15	3.64	5.21	0.15	7.08	6.89	YES
*Hamadryas epinome**	Continuous	0.08	0.31	1.70	0.30	0.47	1.90	YES
Heliconius erato	Continuous	1.53	0.83	1.20	3.45	1.89	3.28	NO
*Heliopetes libra**	Continuous	0.00	0.04	0.46	0.00	0.16	0.62	YES
Heraclides astyalus	Continuous	0.61	0.42	0.64	1.23	0.62	1.23	NO
*Hermeuptychia gisella**	Continuous	0.00	0.00	0.31	0.00	0.00	0.32	YES
*Hypanartia lethe**	Continuous	0.00	0.08	1.43	0.00	0.31	1.73	YES
Junonia genoveva	Continuous	0.61	2.50	1.46	1.23	3.74	3.91	NO
Marpesia petreus	Continuous	0.61	0.61	0.58	0.92	0.61	1.38	NO
Memphis moruus	Continuous	0.00	0.00	4.25	0.00	0.00	4.25	YES
*Ministrymon azia**	Continuous	0.10	NA	1.75	0.15	NA	1.85	NO
*Morpho helenor**	Continuous	0.22	0.10	2.86	0.61	0.16	3.03	YES
Ortilia ithra	Continuous	0.78	0.23	1.16	1.70	0.47	1.70	NO
Phoebis neocypris	Continuous	0.11	1.07	0.70	0.46	2.87	3.04	NO
Pyrgus orcus	Continuous	0.92	0.59	1.22	1.54	1.10	1.90	NO
Smyrna blomfildia	Continuous	0.38	0.30	0.35	1.26	0.30	1.08	NO
*Synapte silius**	Continuous	0.00	0.06	1.42	0.00	0.15	1.54	YES
*Trina geometrina**	Continuous	1.20	NA	3.15	2.51	NA	3.62	YES
Epargyreus socus	Disjunct	4.63	NA	3.52	7.44	NA	6.41	NO
*Epiphile orea**	Disjunct	NA	NA	4.10	NA	NA	4.10	YES
*Haematera pyrame**	Disjunct	0.14	NA	3.70	0.63	NA	4.04	YES
*Hamadryas fornax**	Disjunct	0.33	0.00	0.74	0.46	0.00	0.79	NO
Heliopetes alana	Disjunct	0.07	0.54	0.76	0.15	0.92	1.08	NO
*Leptophobia aripa**	Disjunct	0.09	0.00	5.33	0.15	0.00	5.56	YES
*Lychnuchoides ozias**	Disjunct	NA	0.15	1.56	NA	0.15	1.70	YES
Mechanitis lysimnia	Disjunct	0.51	0.69	0.91	1.54	1.54	1.54	YES
Memphis acidalia	Disjunct	NA	0.00	3.14	NA	0.00	3.18	YES
*Pseudopieris nehemia**	Disjunct	0.08	0.03	3.32	0.30	0.16	3.61	YES

(Continues)

TABLE 4 (Continued)

Species	Distribution between forests	Mean divergence (%K2P)			Max divergence (%K2P)			Distinct mitochondrial lineages
		Within Atlantic Forest	Within Andean forests	Between forests	Within Atlantic Forest	Within Andean forests	Between forests	
*Siproeta epaphus**	Disjunct	0.06	0.08	1.85	0.15	0.31	2.05	YES
*Staphylus incisus**	Disjunct	0.00	0.41	3.82	0.00	0.46	4.09	YES
*Strymon bubastus**	Disjunct	NA	NA	2.69	NA	NA	2.69	YES
*Taygetis ypthima**	Disjunct	0.61	NA	3.17	0.61	NA	3.33	YES
*Thyridia psidii**	Disjunct	0.00	NA	0.46	0.00	NA	0.46	YES
Urbanus pronta	Disjunct	NA	3.14	3.14	NA	4.71	4.71	NO

Note: Distribution between forests and mean and absolute maximum intraspecific divergences (K2P, %) within and between forests are indicated for each species. The last column indicates if individuals from the Atlantic Forest and the Andean forests were clustered into two or more distinct, well-supported (bootstrap support values ≥80%) mitochondrial lineages in the gene trees based on the NEA + NWA data set. Asterisks indicate species used to estimate divergence times between forests.

Even though this can be at least partially attributed to the larger geographic distances covered in the NEA database, it could also be associated with differences in landscape sampling within each region. Butterflies in the NWA database were collected almost exclusively in montane forests located on the east slope of the Southern Central Andes in western Argentina. On the other hand, sampling covered a more heterogeneous landscape for the NEA library (Figure 1), with collection taking place across different ecoregions and physical barriers from eastern Argentina such as the Paraná-Paraguay River axis, the largest subtropical fluvial system in South America (Kopuchian et al., 2020). This, together with the fact that intraspecific variation was only weakly associated with geographic distances in NEA (Lavinia et al., 2017a), suggests that not only larger geographic distances but also the sampling of ecologically and climatically distinct habitats, alone or in combination with other evolutionary drivers, could explain the higher intraspecific variation found in the NEA data set (Gaytán et al., 2020; Kopuchian et al., 2020; Lavinia et al., 2017a). Further studies are needed to elucidate whether differences between NEA and NWA libraries are the consequence of different sampling strategies, evidence of a genuine difference between regions in the level of intraspecific variation in their butterfly fauna, or a combination of both effects.

4.3 | DNA barcode performance over large geographic distances

In order to assess the impact of large geographic distance on identification success, individuals from the NWA library were identified using that of NEA as reference. BM produced the highest percentage of correct identifications, with only 12 out of 636 queries failing to receive a correct species assignment. As a result, all NWA specimens from 130 out of the 135 species shared between databases were correctly identified using the NEA library. Identification success decreased markedly with BCM and BIC. Correct identifications dropped to between 77.04% and 87.74%, meaning that unequivocal identifications could not be established for 24 to 40 species. Identification success raised when larger threshold values were used, as these allowed the identification of species with high intraspecific variation between NWA and NEA. However, with BIC this positive effect was counterbalanced by a raise in the proportion of ambiguous identifications as thresholds increased. On top of that, 74 NWA queries representing 22 species remained unidentified even when the highest threshold (2.06%) was implemented. Genetic distance between these unidentified individuals and their closest conspecifics in NEA averaged 3.62% (95% CI: 3.34%–3.91%), being considerably larger than the highest threshold and significantly higher (Welch's $F = 469.90$, $p < .001$) than the minimum distance to their conspecifics in NWA (0.17%, 95% CI: 0.03%–0.31%).

The great number of ambiguous calls and unidentified NWA queries is a direct consequence of the effect of increased geographic and taxonomic sampling on intra- and interspecific variation. Maximum genetic distance was significantly higher (Welch's $F = 63.48$, $p < .001$) between NWA queries and their conspecifics in the NEA library (mean 1.24%, 95% CI: 1.13%–1.34%), than to those within the NWA data set (mean 0.71%, 95% CI: 0.63%–0.78%). At the same time, NWA individuals had a significantly smaller (Welch's $F = 87.26$, $p < .001$) minimum interspecific distance when queried against the NEA library (mean 6.06%, 95% CI: 5.87%–6.25%) than within the NWA database (mean 7.35%, 95% CI: 7.16%–7.55%). This resulted in the reduction of the barcode gap which, in turn, raised identification ambiguity and diminished the effectiveness of COI for species discrimination and identification.

A total of 386 individuals from 78 species from the NWA database did not have a potential conspecific match in the NEA library. When these specimens were included in the simulations, correct identifications with BM plummeted to 61% as all of them were incorrectly identified. This error was minimized with BCM and BIC, which require query sequences to meet a set sequence divergence threshold before they gain an assignment. As expected, between 96.63% and 99.22% of these NWA individuals remained unidentified with these criteria. Interestingly, specimens in the NWA database

of *Heraclides lamarchei*, which is absent in eastern Argentina, were always erroneously identified as *Heraclides hectorides* (NEA library), which is not found in western Argentina. Mean distance between these allopatric taxa is 0.91% (range 0.77%–1.23%), being remarkably lower than the ~6% sequence divergence between each of them and their corresponding nearest neighbour in NWA and NEA, illustrating how more closely related species can be encountered as the sampling scale increases (Barco et al., 2016; Bergsten et al., 2012; Marín et al., 2017; Virgilio et al., 2010).

In conclusion, COI sequences from one region in Argentina can be used to correctly identify a large proportion of the butterfly fauna from a distant locality in the country. That being said, identification success over large geographic distances is still considerably lower than that obtained from either the regional NWA and NEA (Lavinia et al., 2017a) libraries or the expanded NEA+NWA database (see below). Finally, it is important to note that even though increasing taxonomic and geographic sampling do affect the identification performance of DNA barcodes, their capacity to discriminate species should not be affected as long as minimum interspecific distances remain greater than maximum intraspecific ones (Hausmann et al., 2013; Huemer et al., 2014; Lukhtanov et al., 2009; Marín et al., 2017).

4.4 | Historical relationship between South American forests

The Atlantic Forest to the east and the Andean forests to the west concentrate the majority of the butterfly diversity of Argentina. Even though these forests are currently isolated by the open vegetation corridor (OVC, Figure 1), they have been connected in the past (Ledo & Colli, 2017; Lundberg et al., 1998). Among the 135 species of butterflies shared between NWA and NEA, 85 inhabit both the Atlantic and Andean forests and 27 of these have a disjunct distribution between them (Klimaitis et al., 2018). The remaining 58 species have a comparatively more continuous range across northern Argentina, occurring also in the relatively more open and drier environments of the Humid and Dry Chaco that separate the Atlantic and Andean forests (Figure 1). Results show that intraspecific divergence between forests is significantly higher among species with disjunct distributions than for those with continuous ranges. Even though the frequency of species with relatively high genetic differentiation between forests was only slightly higher among those with disjunct distributions, support for the mitochondrial lineages matching the Atlantic and Andean forests was notably different between the two groups. In 75% of the species with disjunct distributions, individuals from these forests were clustered into two or more distinct, well-supported mitochondrial lineages (bootstrap support values between 80%–100%) in the NEA + NWA gene trees. In contrast, less than 45% of the species with continuous ranges were clustered in distinct, well-supported lineages (Table 4; Appendix S3 and S4).

The species' spatial distribution type has an impact on the level of intraspecific genetic differentiation across northern Argentina in general, and between forest populations in particular. Our findings are in agreement with those of a similar study on moths from the southern European peninsulas (Gaytán et al., 2020), and contrast with the minimal to nonexistent association between distribution patterns and intraspecific divergence reported for the Lepidoptera of Asia and northcentral Europe (Huemer et al., 2014, 2018), where current distributions seem to be the result of recent range expansions after the last glacial maximum (i.e., within less than 15,000 years). On the contrary, the Iberian and Italian Peninsulas, which were refugia during Pleistocene glaciations, host intraspecific lineages that never spread north and eastwards after the retreat of the ice sheets. Therefore, the undersampling of these peninsulas results in an underestimation of intraspecific genetic diversity more marked than that expected by the sole reduction of the geographic scale (Gaytán et al., 2020). Our results and those of Gaytán et al., (2020) reflect that the knowledge of species' distribution ranges as well as their comprehensive sampling are key to better understand spatial patterns of intraspecific variation and to increase DNA barcoding performance, as we discussed above.

All splits between Atlantic Forest and Andean forest populations reported here were dated to the last 2.5 million years, a time period when these currently isolated forests experienced multiple transient connections promoted by the cyclical climate changes of the Quaternary (Lavinia et al., 2019; Trujillo-Arias et al., 2017, 2020; Turchetto-Zolet et al., 2016; Zachos et al., 2001). Divergence times estimated here support an older diversification history than that of most Eurasian Lepidoptera (Huemer et al., 2014, 2018; but see Gaytán et al., 2020), as most splits were estimated to have occurred before the last glacial cycle (~0.1 Ma). This is consistent with previous studies that suggest that around 70% of sister species pairs of Neotropical butterflies diverged along the Pleistocene (Brower & Garzón-Orduña, 2020; Garzón-Orduña et al., 2014; Matos-Maraví, 2016). Since the Andean forests (and the adjacent Amazonia) have been connected with the Atlantic Forest through different and not mutually exclusive spatiotemporal routes throughout the Neogene and Quaternary (Batalha-Filho et al., 2013; Lavinia et al., 2019; Prates et al., 2016; Prates, Xue, et al., 2016; Trujillo-Arias et al., 2020), our results do not reject the possibility of older diversification events for the butterfly fauna of these forests, especially considering the fact that here we focused on intraspecific splits instead of dating speciation events or crown group ages (Blandin & Purser, 2013; Elias et al., 2009).

The contrasting temporal patterns of divergence across the OVC suggest that species with disjunct and continuous distributions between the Atlantic and Andean forests have been differentially affected by the establishment of this environmental barrier and the historical cycles of connection and disconnection between these forests. At the same time, variation in the level of population divergence between forests was also observed within each group of species, indicating that current distribution patterns might be the result of multiple diversification events promoted by past connections between these habitats. This is congruent with recent evidence that shows that phylogeographical patterns shared among codistributed

species separated by a common barrier cannot be explained by a single vicariant or dispersal event. On the contrary, diversification histories are idiosyncratic as species' responses to a common barrier, such as the OVC, will depend greatly on the biological attributes of each taxon (Kopuchian et al., 2020; Lavinia et al., 2019; Penz et al., 2015; Prates, Rivera, et al., 2016; Prates, Xue, et al., 2016; Smith et al., 2014). In particular, it has been shown that species traits associated with dispersal and colonization abilities (e.g., habitat choice, feeding generalism and other ecophysiological traits) can explain variation in the levels of intraspecific divergence among butterfly (Dapporto et al., 2019; Huemer et al., 2014; Penz et al., 2015) and bird species (Burney & Brumfield, 2009; Harvey et al., 2017; Lavinia et al., 2015, 2019). In this context, one possibility is that species with disjunct distributions represent forest specialists for which the OVC constitutes a more critical barrier to dispersal and gene flow than for species with continuous ranges, which could in turn be classified as more generalist. Future studies should assess the existence of biological differences that could account for the distinct spatial distribution patterns of the species here analysed and, in turn, explain the contrasting levels of intraspecific differentiation found among them.

4.5 | A reference library for the butterflies of Argentina

A database with over 3000 COI sequences from nearly 500 butterfly species, ~40% of the fauna of Argentina (Klimaitis et al., 2018), is now publicly available (https://doi.org/10.5883/DS-NEANWA). A barcode gap was present in all but 12 of the species represented by multiple individuals, and all singletons were distinguishable from their nearest neighbour. Fourteen species were paraphyletic in the gene trees based on the combined data set, but nearly 60% of them were paraphyletic within the individual databases. As a consequence, expanding both taxonomic and geographic coverage did not greatly increase the incidence of paraphyletic or otherwise problematic taxa. This result contrasts with that reported for a group of aquatic beetles where the frequency of nonmonophyletic species increased significantly as the geographic scale expanded (Bergsten et al., 2012).

Identification simulations showed that the current COI library is very effective in identifying unknown specimens, despite the reduction of the barcode gap in the expanded database. However, this reduction did increase identification uncertainty and reduced identification success with BIC, as expected given its stringent nature (Bergsten et al., 2012; Virgilio et al., 2010). As a result, the expanded database performed on average slightly worse than the local NEA and NWA libraries when this criterion was applied. By contrast, the narrowed barcode gap had a null or negligible effect on the more liberal BM and BCM criteria, which returned between 97.68% and 99.45% correct species assignments. These values are similar to those derived from the NEA and NWA databases independently, and considerably better than the ones obtained when specimens from NWA were identified using reference sequences from the geographically distant NEA. In the case of the latter, identification success decreased between 11% and 22%, depending on the identification criterion and sequence threshold implemented, in comparison to that obtained with the NWA, NEA and NEA + NWA libraries. These results highlight the importance of increasing the spatial and taxonomic coverage of DNA barcode reference libraries to better capture intra- and interspecific diversity levels and to improve identification success, and of considering the use of more regional databases for the identification of local fauna when a large-scale library is not available (Bergsten et al., 2012; Gaytán et al., 2020).

5 | CONCLUSION

By extending the DNA barcode coverage for the butterflies of Argentina through the generation of over 1000 new COI sequences from more than 200 species and the survey of western populations, this study has provided new insights into butterfly diversification patterns in the southern Neotropics. We showed that expanding the geographic and taxonomic sampling increases maximum intraspecific divergences and reduces distances to the closest heterospecific, diminishing identification success, especially when strict identification criteria are employed. Furthermore, our results evidenced that patterns of mitochondrial variation are influenced by species' spatial distribution type, probably reflecting biological differences that impact their dispersal and colonization abilities. In particular, the present results suggest that past connections between the currently isolated Atlantic and Andean forests have differentially affected their shared butterfly fauna, and that all diversification events between these environments took place in the Pleistocene. To our knowledge, this constitutes the first multispecies assessment of the historical relationship between these forests using butterfly species as model organisms. Finally, our research supports the fact that, even in the era of genomic data, large-scale analyses of mitochondrial DNA variation are still extremely useful for evolutionary studies, as they unveil spatial diversification patterns and highlight cases that deserve further investigation (Barreira et al., 2016; Dapporto et al., 2019; Kress et al., 2015).

ACKNOWLEDGEMENTS

We thank colleagues from the MACN and the CBG for processing the tissue samples and generating the COI sequences. We also thank the two reviewers for their comments and suggestions which improved significantly our manuscript. This project was supported by the Richard Lounsbery Foundation, Natural Sciences and Engineering Resarch Council of Canada, the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), the Agencia Nacional de Promoción de la Investigación, el Desarrollo Tecnológico y la Innovación, Fundación Williams, Fundación Bosques Nativos Argentinos and Fundación Temaikèn. For granting the permits and transit guides, we thank the Offices of Fauna of the Argentinian provinces in which fieldwork was conducted, the National Parks

Administration, and the Ministerio de Ambiente y Desarrollo Sostenible from Argentina. We thank Leonardo Demartino for his aid with data analysis.

CONFLICT OF INTEREST

The authors have no conflicts of interest to declare.

AUTHOR CONTRIBUTIONS

NA and PDL conceived de project. NA and PLD designed research and performed all analyses with contribution from all the authors. ENB, NA and PDL performed specimen and genetic data curation. DAL, PDNH, PLT and PDL acquired funding and administrated the project. NA and PDL wrote the manuscript with contributions from all authors. PDL supervised the project.

DATA AVAILABILITY STATEMENT

All specimen and sequence data for the NWA, NEA and NEA + NWA databases are available in their respective public data sets "DS-NWAR" (https://doi.org/10.5883/DS-NWAR), "DS-BUNEACAR" (https://doi.org/10.5883/DS-BUNEACAR) and "DS-NEANWA" (https://doi.org/10.5883/DS-NEANWA) on BOLD (www.boldsystems.org), and in Dryad (https://doi.org/10.5061/dryad.rfj6q5790) (Lavinia et al., 2017b). COI sequences can also be found in GenBank under accession numbers MZ334986-MZ336013.

ORCID

Pablo D. Lavinia https://orcid.org/0000-0002-3583-9637

REFERENCES

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. J. (1990). Basic local alignment search tool. *Journal of Molecular Biology*, *215*(3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Barco, A., Raupach, M. J., Laakmann, S., Neumann, H., & Knebelsberger, T. (2016). Identification of North Sea molluscs with DNA barcoding. *Molecular Ecology Resources*, *16*(1), 288–297. https://doi.org/10.1111/1755-0998.12440.

Barreira, A. S., Lijtmaer, D. A., Tubaro, P. L., & Adamowicz, S. (2016). The multiple applications of DNA barcodes in avian evolutionary studies. *Genome*, *59*(11), 899–911. https://doi.org/10.1139/gen-2016-0086.

Batalha-Filho, H., Fjeldså, J., Fabre, P.-H., & Miyaki, C. Y. (2013). Connections between the Atlantic and the Amazonian forest avifaunas represent distinct historical events. *Journal of Ornithology*, *154*(1), 41–50. https://doi.org/10.1007/s10336-012-0866-7.

Bergsten, J., Bilton, D. T., Fujisawa, T., Elliott, M., Monaghan, M. T., Balke, M., Hendrich, L., Geijer, J., Herrmann, J., Foster, G. N., Ribera, I., Nilsson, A. N., Barraclough, T. G., & Vogler, A. P. (2012). The effect of geographical scale of sampling on DNA barcoding. *Systematic Biology*, *61*(5), 851–869. https://doi.org/10.1093/sysbio/sys037.

Blandin, P., & Purser, B. (2013). Evolution and diversification of Neotropical butterflies: Insights from the biogeography and phylogeny of the genus *Morpho fabricius*, 1807 (Nymphalidae: Morphinae), with a review of the geodynamics of South America. *Tropical Lepidoptera Research*, *23*(2), 62–85.

Bonebrake, T. C., Ponisio, L. C., Boggs, C. L., & Ehrlich, P. R. (2010). More than just indicators: A review of tropical butterfly ecology and conservation. *Biological Conservation*, *143*(8), 1831–1841. https://doi.org/10.1016/j.biocon.2010.04.044.

Brower, A. V. Z. (1994). Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. *Proceedings of the National Academy of Sciences of the United States of America*, *91*(14), 6491–6495.

Brower, A. V. Z., & Garzón-Orduña, I. J. (2020). Contrasting patterns of temporal diversification in neotropical butterflies: An overview. In V. Rull, & A. Carnaval (Eds.), *Neotropical diversification: Patterns and processes* (pp. 189–222). Springer. https://doi.org/10.1007/978-3-030-31167-4_9

Brown, S. D. J., Collins, R. A., Boyer, S., Lefort, M.-C., Malumbres-Olarte, J., Vink, C. J., & Cruickshank, R. H. (2012). Spider: An R package for the analysis of species identity and evolution, with particular reference to DNA barcoding. *Molecular Ecology Resources*, *12*(3), 562–565. https://doi.org/10.1111/j.1755-0998.2011.03108.x.

Burney, C. W., & Brumfield, R. T. (2009). Ecology predicts levels of genetic differentiation in Neotropical birds. *American Naturalist*, *174*(3), 358–368. https://doi.org/10.1086/603613.

Cabrera, A. L. (1976). *Regiones Fitogeográficas Argentinas*. Encicl. Arg. de Agric. y Jardineria, II, ACME.

Chazot, N., Willmott, K. R., Condamine, F. L., De-Silva, D. L., Freitas, A. V. L., Lamas, G., Morlon, H., Giraldo, C. E., Jiggins, C. D., Joron, M., Mallet, J., Uribe, S., & Elias, M. (2016). Into the Andes: Multiple independent colonizations drive montane diversity in the Neotropical clearwing butterflies *Godyridina*. *Molecular Ecology*, *25*(22), 5765–5784. https://doi.org/10.1111/mec.13773.

Cheng, H., Sinha, A., Cruz, F. W., Wang, X., Edwards, R. L., d'Horta, F. M., Ribas, C. C., Vuille, M., Stott, L. D., & Auler, A. S. (2013). Climate change patterns in Amazonia and biodiversity. *Nature Communications*, *4*, 1411. https://doi.org/10.1038/ncomms2415.

Cong, Q., Shen, J., Borek, D., Robbins, R. K., Otwinowski, Z., & Grishin, N. V. (2016). Complete genomes of Hairstreak butterflies, their speciation, and nucleo-mitochondrial incongruence. *Scientific Reports*, *6*, 24863. https://doi.org/10.1038/srep24863.

Costa, L. P. (2003). The historical bridge between the Amazon and the Atlantic Forest of Brazil: A study of molecular phylogeography with small mammals. *Journal of Biogeography*, *30*, 71–86. https://doi.org/10.1046/j.1365-2699.2003.00792.x.

Dapporto, L., Cini, A., Vodă, R., Dincă, V., Wiemers, M., Menchetti, M., Magini, G., Talavera, G., Shreeve, T., Bonelli, S., Casacci, L. P., Balletto, E., Scalercio, S., & Vila, R. (2019). Integrating three comprehensive data sets shows that mitochondrial DNA variation is linked to species traits and paleogeographic events in European butterflies. *Molecular Ecology Resources*, *19*(6), 1623–1636. https://doi.org/10.1111/1755-0998.13059.

Dincă, V., Montagud, S., Talavera, G., Hernández-Roldán, J., Munguira, M. L., García-Barros, E., Hebert, P. D. N., & Vila, R. (2015). DNA barcode reference library for Iberian butterflies enables a continental-scale preview of potential cryptic diversity. *Scientific Reports*, *5*, 12395. https://doi.org/10.1038/srep12395.

Duarte, M., & Robbins, R. K. (2005). Immature stages of *Calycopis caulonia* (Hewitson, 1877) (Lepidoptera, Lycaenidae, Theclinae, Eumaeini), with notes on rearing detritivorous hairstreaks on artificial diet. *Zootaxa*, *1063*, 1–31. https://doi.org/10.11646/zootaxa.1063.1.1.

Ebel, E. R., DaCosta, J. M., Sorenson, M. D., Hill, R. I., Briscoe, A. D., Willmott, K. R., & Mullen, S. P. (2015). Rapid diversification associated with ecological specialization in Neotropical *Adelpha* butterflies. *Molecular Ecology*, *24*(10), 2392–2405. https://doi.org/10.1111/mec.13168.

Elias, M., Joron, M., Willmott, K., Silva-brandão, K. L., Kaiser, V., Arias, C. F., Piñerez, L. M. G., Uribe, S., Brower, A. V. Z., Freitas, A. V. L., & Jiggins, C. D. (2009). Out of the Andes: Patterns of diversification in clearwing butterflies. *Molecular Ecology*, *18*(8), 1716–1729. https://doi.org/10.1111/j.1365-294X.2009.04149.x.

Folmer, O., Black, M., Hoeh, W., Lutz, R., & Vrijenhoek, R. (1994). DNA primers for amplification of mitochondrial cytochrome c oxidase

subunit I from diverse metazoan invertebrates. *Molecular Marine Biology and Biotechnology*, *3*(5), 294–299.

Garzón-Orduña, I. J., Benetti-Longhini, J. E., & Brower, A. V. Z. (2014). Timing the diversification of the Amazonian biota: Butterfly divergences are consistent with Pleistocene refugia. *Journal of Biogeography*, *41*(9), 1631–1638. https://doi.org/10.1111/jbi.12330.

Gaytán, A., Bergsten, J., Canelo, T., Pérez-Izquierdo, C., Santoro, M., & Bonal, R. (2020). DNA Barcoding and geographical scale effect: The problems of undersampling genetic diversity hotspots. *Ecology and Evolution*, *10*, 10754–10772. https://doi.org/10.1002/ece3.6733.

Godoy-Bürki, A. C., Ortega-Baes, P., Sajama, J. M., & Aagesen, L. (2014). Conservation priorities in the Southern Central Andes: Mismatch between endemism and diversity hotspots in the regional flora. *Biodiversity and Conservation*, *23*(1), 81–107. https://doi.org/10.1007/s10531-013-0586-1.

Haffer, J. (1969). Speciation in Amazonian forest birds. *Science*, *165*(3889), 131–137. https://doi.org/10.1126/science.165.3889.131.

Harvey, M. G., Aleixo, A., Ribas, C. C., & Brumfield, R. T. (2017). Habitat association predicts genetic diversity and population divergence in Amazonian birds. *The American Naturalist*, *190*(5), 631–648. https://doi.org/10.1086/693856.

Hausmann, A., Godfray, H. C. J., Huemer, P., Mutanen, M., Rougerie, R., van Nieukerken, E. J., Ratnasingham, S., & Hebert, P. D. N. (2013). Genetic patterns in European geometrid moths revealed by the Barcode Index Number (BIN) System. *PLoS One*, *8*(12), e84518. https://doi.org/10.1371/journal.pone.0084518.

Hebert, P. D. N., Cywinska, A., Ball, S. L., & DeWaard, J. R. (2003). Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *270*(1512), 313–321. https://doi.org/10.1098/rspb.2002.2218

Hebert, P. D. N., Penton, E. H., Burns, J. M., Janzen, D. H., & Hallwachs, W. (2004). Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proceedings of the National Academy of Sciences of the United States of America*, *101*(41), 14812–14817. https://doi.org/10.1073/pnas.0406166101.

Hoorn, C., Wesselingh, F. P., ter Steege, H., Bermudez, M. A., Mora, A., Sevink, J., Sanmartin, I., Sanchez-Meseguer, A., Anderson, C. L., Figueiredo, J. P., Jaramillo, C., Riff, D., Negri, F. R., Hooghiemstra, H., Lundberg, J., Stadler, T., Sarkinen, T., & Antonelli, A. (2010). Amazonia through time: Andean uplift, climate change, landscape evolution, and biodiversity. *Science*, *330*(6006), 927–931.

Huemer, P., Hebert, P. D. N., Mutanen, M., Wieser, C., Wiesmair, B., Hausmann, A., Yakovlev, R., Möst, M., Gottsberger, B., Strutzenberger, P., & Fiedler, K. (2018). Large geographic distance versus small DNA barcode divergence: Insights from a comparison of European to South Siberian Lepidoptera. *PLoS One*, *13*(11), e0206668. https://doi.org/10.1371/journal.pone.0206668.

Huemer, P., Mutanen, M., Sefc, K. M., & Hebert, P. D. N. (2014). Testing DNA Barcode performance in 1000 species of European Lepidoptera: Large geographic distances have small genetic impacts. *PLoS One*, *9*(12), e115774. https://doi.org/10.1371/journal.pone.0115774.

Ivanova, N. V., Dewaard, J. R., & Hebert, P. D. N. (2006). An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, *6*(4), 998–1002. https://doi.org/10.1111/j.1471-8286.2006.01428.x.

Kekkonen, M., & Hebert, P. D. N. (2014). DNA barcode-based delineation of putative species: Efficient start for taxonomic workflows. *Molecular Ecology Resources*, *14*(4), 706–715. https://doi.org/10.1111/1755-0998.12233.

Kimura, M. (1980). A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *Journal of Molecular Evolution*, *16*(2), 111–120. https://doi.org/10.1007/BF01731581.

Klimaitis, J., Núñez Bustos, E., Klimaitis, C., & Güller, R. (2018). *Mariposas-Butterflies-Argentina. Guía de Identificación-identification guide*. : Vazquez Mazzini Editores.

Kopuchian, C., Campagna, L., Lijtmaer, D. A., Cabanne, G. S., García, N. C., Lavinia, P. D., Tubaro, P. L., Lovette, I., & Di Giacomo, A. S. (2020). A test of the riverine barrier hypothesis in the largest subtropical river basin in the Neotropics. *Molecular Ecology*, *29*(12), 2137–2149. https://doi.org/10.1111/mec.15384.

Kress, W. J., & Erickson, D. L. (2012). *DNA barcodes: Methods and protocols*. *Methods in molecular biology*. Humana Press. https://doi.org/10.1007/978-1-61779-591-6_1.

Kress, W. J., García-Robledo, C., Uriarte, M., & Erickson, D. L. (2015). DNA barcodes for ecology, evolution, and conservation. *Trends in Ecology & Evolution*, *30*(1), 25–35. https://doi.org/10.1016/j.tree.2014.10.008.

Lamas, G. (2004). *Checklist: Part 4 A Hesperioidea — Papilionoidea*. Association for Tropical Lepidoptera/Scientific Publishers.

Lavinia, P. D., Barreira, A. S., Campagna, L., Tubaro, P. L., & Lijtmaer, D. A. (2019). Contrasting evolutionary histories in Neotropical birds: Divergence across an environmental barrier in South America. *Molecular Ecology*, *28*(7), 1730–1747. https://doi.org/10.1111/mec.15018.

Lavinia, P. D., Escalante, P., García, N. C., Barreira, A. S., Trujillo-Arias, N., Tubaro, P. L., Naoki, K., Miyaki, C. Y., Santos, F. R., & Lijtmaer, D. A. (2015). Continental-scale analysis reveals deep diversification within the polytypic Red-crowned Ant Tanager (Habia rubica, Cardinalidae). *Molecular Phylogenetics and Evolution*, *89*, 182–193. https://doi.org/10.1016/j.ympev.2015.04.018.

Lavinia, P. D., Núñez Bustos, E. O., Kopuchian, C., Lijtmaer, D. A., García, N. C., Hebert, P. D. N., & Tubaro, P. L. (2017a). 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.

Lavinia, P. D., Núñez Bustos, E. O., Kopuchian, C., Lijtmaer, D. A., García, N. C., Hebert, P. D. N., & Tubaro, P. L. (2017b). Butterflies of Northeastern and Central Argentina. *BOLD System*. https://doi.org/10.5883/DS-BUNEACAR.

Ledo, R. M. D., & Colli, G. R. (2017). The historical connections between the Amazon and the Atlantic Forest revisited. *Journal of Biogeography*, *44*(11), 2551–2563. https://doi.org/10.1111/jbi.13049.

Lukhtanov, V. A., Sourakov, A., Zakharov, E. V., & Hebert, P. D. N. (2009). DNA barcoding Central Asian butterflies: Increasing geographical dimension does not significantly reduce the success of species identification. *Molecular Ecology Resources*, *9*(5), 1302–1310. https://doi.org/10.1111/j.1755-0998.2009.02577.x.

Lundberg, J. G., Marshall, L. G., Guerrero, J., Horton, B., Malabarba, M. C. S. L., & Wesselingh, F. (1998). The stage for neotropical fish diversification: A history of tropical South American rivers. In L. Malabarba, R. Reis, R. Vari, L. Zms, & L. Cas (Eds.), *Phylogeny and classification of neotropical fishes* (pp. 13–48). Edipucrs.

Marín, M. A., Cadavid, I. C., Valdés, L., Álvarez, C. F., Uribe, S. I., Vila, R., & Pyrcz, T. W. (2017). DNA barcoding of an assembly of montane Andean butterflies (Satyrinae): Geographical scale and identification performance. *Neotropical Entomology*, *46*(5), 514–523. https://doi.org/10.1007/s13744-016-0481-z.

Matos-Maraví, P. (2016). Investigating the timing of origin and evolutionary processes shaping regional species diversity: Insights from simulated data and neotropical butterfly diversification rates. *Evolution*, *70*(7), 1638–1650. https://doi.org/10.1111/evo.12960.

Meier, R., Shiyang, K., Vaidya, G., & Ng, P. K. L. (2006). DNA barcoding and taxonomy in Diptera: A tale of high intraspecific variability and low identification success. *Systematic Biology*, *55*(5), 715–728. https://doi.org/10.1080/10635150600969864.

Mutanen, M., Kivelä, S. M., Vos, R. A., Doorenweerd, C., Ratnasingham, S., Hausmann, A., Huemer, P., Dincă, V., van Nieukerken, E. J., Lopez-Vaamonde, C., Vila, R., Aarvik, L., Decaëns, T., Efetov, K. A., Hebert, P. D. N., Johnsen, A., Karsholt, O., Pentinsaari, M., Rougerie, R., … Godfray, H. C. J. (2016). Species-level para- and polyphyly

in DNA barcode gene trees: Strong operational bias in European Lepidoptera. *Systematic Biology*, *65*(6), 1024–1040. https://doi.org/10.1093/sysbio/syw044.

New, T. R., & Samways, M. J. (2014). Insect conservation in the southern temperate zones: An overview. *Austral Entomology*, *53*(1), 26–31. https://doi.org/10.1111/aen.12071.

Olson, D. M., & Dinerstein, E. (2002). The global 200: Priority ecoregions for global conservation. *Annals of the Missouri Botanical Garden*, *89*(2), 199–224. https://doi.org/10.2307/3298564.

Papadopoulou, A., Anastasiou, I., & Vogler, A. P. (2010). Revisiting the insect mitochondrial molecular clock: The mid-Aegean trench calibration. *Molecular Biology and Evolution*, *27*(7), 1659–1672. https://doi.org/10.1093/molbev/msq051.

Penz, C., DeVries, P., Tufto, J., & Lande, R. (2015). Butterfly dispersal across Amazonia and its implication for biogeography. *Ecography*, *38*(4), 410–418. https://doi.org/10.1111/ecog.01172.

Prates, I., Melo-Sampaio, P. R., Drummond, L. D. O., Teixeira, M., Rodrigues, M. T., & Carnaval, A. C. (2017). Biogeographic links between southern Atlantic Forest and western South America: Rediscovery, redescription, and phylogenetic relationships of two rare montane anole lizards from Brazil. *Molecular Phylogenetics and Evolution*, *113*, 49–58. https://doi.org/10.1016/j.ympev.2017.05.009.

Prates, I., Rivera, D., Rodrigues, M. T., & Carnaval, A. C. (2016). A mid-Pleistocene rainforest corridor enabled synchronous invasions of the Atlantic Forest by Amazonian anole lizards. *Molecular Ecology*, *25*(20), 5174–5186. https://doi.org/10.1111/mec.13821.

Prates, I., Xue, A. T., Brown, J. L., Alvarado-Serrano, D. F., Rodrigues, M. T., Hickerson, M. J., & Carnaval, A. C. (2016). Inferring responses to climate dynamics from historical demography in neotropical forest lizards. *Proceedings of the National Academy of Sciences of the United States of America*, *113*(29), 7978–7985. https://doi.org/10.1073/pnas.1601063113.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing.

Ratnasingham, S., & Hebert, P. D. N. (2007). BOLD: The barcode of life data system. *Molecular Ecology Notes*, *7*(3), 355–364. https://doi.org/10.1111/j.1471-8286.2007.01678.x.

Ribeiro, M. C., Metzger, J. P., Martensen, A. C., Ponzoni, F. J., & Hirota, M. M. (2009). The Brazilian Atlantic Forest: How much is left, and how is the remaining forest distributed? Implications for conservation. *Biological Conservation*, *142*(6), 1141–1153. https://doi.org/10.1016/j.biocon.2009.02.021.

Rull, V. (2020). Neotropical diversification: Historical overview and conceptual insights. In V. Rull, & A. C. Carnaval (Eds.), *Neotropical diversification: Patterns and processes* (pp. 13–49). Springer, Cham.

Smith, B. T., McCormack, J. E., Cuervo, A. M., Hickerson, M. J., Aleixo, A., Cadena, C. D., Pérez-Emán, J., Burney, C. W., Xie, X., Harvey, M. G., Faircloth, B. C., Glenn, T. C., Derryberry, E. P., Prejean, J., Fields, S., & Brumfield, R. T. (2014). The drivers of tropical speciation. *Nature*, *515*(7527), 406–409. https://doi.org/10.1038/nature13687.

Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics*, *30*(9), 1312–1313. https://doi.org/10.1093/bioinformatics/btu033.

Tamura, K., Peterson, D., Peterson, N., Stecher, G., Nei, M., & Kumar, S. (2011). MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. *Molecular Biology and Evolution*, *28*(10), 2731–2739. https://doi.org/10.1093/molbev/msr121.

Trujillo-Arias, N., Dantas, G. P. M., Arbeláez-Cortés, E., Naoki, K., Gómez, M. I., Santos, F. R., Miyaki, C. Y., Aleixo, A., Tubaro, P. L., & Cabanne, G. S. (2017). The niche and phylogeography of a passerine reveal the history of biological diversification between the Andean and the Atlantic forests. *Molecular Phylogenetics and Evolution*, *112*, 107–121. https://doi.org/10.1016/j.ympev.2017.03.025.

Trujillo-Arias, N., Rodríguez-Cajarville, M. J., Sari, E., Miyaki, C. Y., Santos, F. R., Witt, C. C., Barreira, A. S., Gómez, I., Naoki, K., Tubaro, P. L., & Cabanne, G. S. (2020). Evolution between forest macrorefugia is linked to discordance between genetic and morphological variation in Neotropical passerines. *Molecular Phylogenetics and Evolution*, *149*, 106849. https://doi.org/10.1016/j.ympev.2020.106849.

Turchetto-Zolet, A. C., Salgueiro, F., Turchetto, C., Cruz, F., Veto, N. M., Barros, M. J. F., & Margis, R. (2016). Phylogeography and ecological niche modelling in Eugenia uniflora (Myrtaceae) suggest distinct vegetational responses to climate change between the southern and the northern Atlantic Forest. *Botanical Journal of the Linnean Society*, *182*(3), 670–688. https://doi.org/10.1111/boj.12473.

Virgilio, M., Backeljau, T., Nevado, B., & De Meyer, M. (2010). Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics*, *11*(206), 206. https://doi.org/10.1186/1471-2105-11-206.

Wilson, J. J., Sing, K. W., & Sofian-Azirun, M. (2013). Building a DNA barcode reference library for the true butterflies (Lepidoptera) of Peninsula Malaysia: What about the subspecies? *PLoS One*, *8*(11), e79969. https://doi.org/10.1371/journal.pone.0079969.

Zachos, J., Pagani, H., Sloan, L., Thomas, E., & Billups, K. (2001). Trends, rhythms, and aberrations in global climate 65 Ma to present. *Science*, *292*(5517), 686–693. https://doi.org/10.1126/science.1059412.

Zenker, M. M., Rougerie, R., Teston, J. A., Laguerre, M., Pie, M. R., Freitas, A. V. L., & Constantino, R. (2016). Fast census of moth diversity in the Neotropics: A comparison of field-assigned morphospecies and DNA barcoding in Tiger Moths. *PLoS One*, *11*(2), e0148423. https://doi.org/10.1371/journal.pone.0148423.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Attiná, N., Núñez Bustos, E. O., Lijtmaer, D. A., Hebert, P. D. N., Tubaro, P. L., & Lavinia, P. D. (2021). Genetic variation in neotropical butterflies is associated with sampling scale, species distributions, and historical forest dynamics. *Molecular Ecology Resources*, 00, 1–17. https://doi.org/10.1111/1755-0998.13441