Evolution of *Hypolimnas* butterflies (Nymphalidae): Out-of-Africa origin and Wolbachia-mediated introgression

Ranjit Kumar Sahoo\textsuperscript{a,⁎}, David J. Lohman\textsuperscript{b,c,d}, Niklas Wahlberg\textsuperscript{e}, Chris J. Müller\textsuperscript{g}, Oskar Brattström\textsuperscript{a,h}, Steve C. Collins\textsuperscript{h}, Djunjanti Peggie\textsuperscript{i}, Kwaku Aduse-Poku\textsuperscript{j,l}, Ullasa Kodandaramaiah\textsuperscript{a}

\textsuperscript{a} IISER-TVM Centre for Research and Education in Ecology and Evolution (ICREEE), School of Biology, Indian Institute of Science Education and Research Thiruvananthapuram, 695 551 Kerala, India
\textsuperscript{b} Biology Department, City College of New York, City University of New York, 160 Convent Avenue, NY 10031, United States
\textsuperscript{c} Biology Ph.D. Program, Graduate Center, City University of New York, 365 Fifth Ave., New York, NY 10016, United States
\textsuperscript{d} Entomology Section, National Museum of the Philippines, Manila 1000, Philippines
\textsuperscript{e} Department of Biology, Lund University, Sölvegatan 35, SE-223 62 Lund, Sweden
\textsuperscript{f} Australian Museum, 6 College Street, Sydney, NSW 2010, Australia
\textsuperscript{g} Department of Zoology, University of Cambridge, Downing Street, Cambridge CB2 3EJ, UK
\textsuperscript{h} African Butterfly Research Institute (ABRI), PO Box 14308, 0800 Nairobi, Kenya
\textsuperscript{i} Division of Zoology, Research Centre for Biology-LIPI, Jl. Raya Jakarta-Bogor Km. 46, Cibinong-Bogor 16911, Indonesia
\textsuperscript{j} School of Natural Resources, University of Energy and Natural Resources, Sunyani, Ghana

**Keywords:**
Arabia-Asia land bridge
Biogeography
Horizontal gene transfer
Host plant
Mid-Miocene climatic optimum
Oscillation hypothesis

**ABSTRACT**

*Hypolimnas* butterflies (Nymphalidae), commonly known as eggflies, are a popular model system for studying a wide range of ecological questions including mimicry, polymorphism, wing pattern evolution, and Wolbachia-host interactions. The lack of a time-calibrated phylogeny for this group has precluded understanding its evolutionary history. We reconstruct a species-level phylogeny using a nine gene dataset and estimate species divergence times. Based on the resulting tree, we investigate the taxon’s historical biogeography, examine the evolution of host plant preferences, and test the hypothesis that the endosymbiotic bacterium Wolbachia mediates gene transfer between species. Our analyses indicate that the species are grouped within three strongly supported, deeply divergent clades. However, relationships among these three clades are uncertain. In addition, many *Hypolimnas* species are not monophyletic or monophyletic with weak support, suggesting widespread incomplete lineage sorting and/or introgression. Biogeographic analysis strongly indicates that the genus diverged from its ancestor in Africa and subsequently dispersed to Asia; the strength of this result is not affected by topological uncertainties. While the larvae of African species feed almost exclusively on Urticaceae, larvae of species found further east often feed on several additional families. Interestingly, we found an identical mitochondrial haplotype in two *Hypolimnas* species, *H. bolina* and *H. alimena*, and a strong association between this mitotype and the Wolbachia strain wBo1a. Future investigations should explore the plausibility of Wolbachia-mediated introgression between species.

**1. Introduction**

The Old World butterfly genus *Hypolimnas* is remarkable for the highly varied wing patterns of the taxon’s 29 described species (Lamas, 2015; Table A1), which are commonly known as eggflies. Interspecific variability is so marked because different *Hypolimnas* species are Batesian and/or Müllerian mimics of different, unpalatable model species (Marsh et al., 1977; Vane-Wright et al., 1977). For this reason, the genus has been a model for ecological and evolutionary studies, beginning with Poulton’s (1897) writings on mimicry in *H. bolina* and *H. misippus*, and Marshall’s (1902) observations on mimicry and polymorphism of *H. dubius* (now known as *H. anthedon*). Since then, field and laboratory studies on members of this genus have provided insights into the selective forces that change frequencies of polymorphic forms in populations (Edmunds, 1966, 1969; Smith, 1973, 1976; Turner, 1978; Gordon, 1987; Gordon and Smith, 1998; but see Clarke et al., 1989). Classical genetics research (Clarke and Sheppard, 1975; Smith and Gordon, 1987) and pharmacological investigations (Marsh et al.,
1977) on Hypolimnas species have furthered understanding of wing pattern formation in butterflies. Although investigations on the relevance of wing patterns are mostly limited to two species, H. bolina and H. misippus, field observations suggest geographic patterns for model-mimic relationships among different lineages. While African species usually mimic Amauris and/or Acrea, Asian species generally mimic Danaus and/or Euploea (Table A1 and references therein). Females of the exceptionally widespread, pantropical species H. misippus mimic Danaus chrysippus throughout the Old World - a species that is itself unusually widespread (Vane-Wright et al., 1977; Braby et al., 2015).

Hypolimnas butterflies have been the subject of research on reproductive diapause (Kemp, 2001a; Kemp and Jones, 2001), territorial productivity diapause (Kemp, 2001a; Pieloor and Seymour, 2001), and personal polyphenism (Kemp, 2000; Kemp and Jones, 2001), and parental care (Nafus and Schreiner, 1988; Schreiner and Nafus, 1991). Recent studies on Hypolimnas have focused on the microscopic ornamentation of wing scale arrangements (Saito, 2002; Kemp and Macedonia, 2006; Siddique et al., 2016). Moreover, decade-long investigations on Hypolimnas populations (Charlat et al., 2009; Duplouy et al., 2010) and breeding experiments (Dyson et al., 2002) have made Hypolimnas a model for studying population-level and evolutionary interactions between insects and alpha-proteobacterial lineage Wolbachia. This intracellular, bacterial parasite is found in more than half of all insect species. It is generally inherited maternally through its presence in the cytoplasm of the ovum, and the bacterium can manipulate reproduction of its host to favor production of females (Werren et al., 2008). Interestingly, studies on Hypolimnas-Wolbachia interactions have provided clear evidence for interference among selfish genetic elements (Charlat et al., 2006), and W.D. Hamilton’s theory on rapid sex-ratio flux due to an arms race between sex-ratio distorters and counteracting suppressor genes (Hornett et al., 2006; Charlat et al., 2007).

However, the absence of information on phylogenetic relationships among Hypolimnas species precludes broader understanding of evolutionary patterns in this group (Vane-Wright et al., 1977; Kemp et al., 2014). For instance, studies on butterfly-hostplant associations, historical biogeography, and wing-pattern evolution within the genus are hindered by the lack of a robust phylogeny.

Hypolimnas species are found in all tropical regions. Only one species, H. misippus, is found in Central and South America (where it is quite likely introduced) and it is distributed on every continent but Europe and Arctic. Hypolimnas bolina is also unusually widespread from Oceania west to tropical Asia and Madagascar (Vane-Wright et al., 1977; Tsuchada 1985; Parsons, 1998; Larsen, 2005). Although members of the genus are morphologically similar to each other, there is no known synapomorphic morphological character that validates the monophyly of the genus (Vane-Wright et al., 1977). Previous phylogenies of Nymphalidae (Wahlberg et al., 2005, 2009; Su et al., 2017) included few representative species of Hypolimnas, and recovered the genus as monophyletic and sister to the African genus Precis. The most comprehensive phylogeny of the group (Kodandaramaiah, 2009) included data from a maximum of three genes from 13 species, and was unable to resolve intrageneric relationships with strong support. Furthermore, the taxonomy of many species (e.g., H. bolina) has been challenging owing to their extensive wing pattern polymorphism (Vane-Wright et al., 1977).

In this study, we reconstruct phylogenetic relationships among 113 Hypolimnas specimens collected throughout the genus’ entire range using up to 7013 base pairs (bp) comprising sequences from one mitochondrial and eight nuclear loci. We also estimate divergence times of the species. Based on our results, we investigate the historical biogeographic patterns, evolution of larval host plant preferences, and examine the veracity of current species-level taxonomy. We also investigate patterns of Wolbachia infection in the group and evaluate the plausibility of Wolbachia-mediated gene transfer between species.

2. Methods

2.1. Taxon and marker sampling

We sampled 26 out of 29 recognized Hypolimnas species (Table A1). We were unable to sample H. aubergeri, H. chapmani, and H. euploeoides. The latter species is found on the Admiralty Islands north of New Guinea, but the others are African. Note that Lamas (2015) regards H. sumbawana as a subspecies of H. anomala, but our results demonstrate that sumbawana is clearly distinct from anomala (they are not sister taxa). Wherever possible, we included multiple specimens of each species from different geographic areas, for a total of 113 specimens. Preliminary species-level identifications were based on traditional morphological characters. We attempted to amplify eight nuclear loci (ArgKin, CAD, EF1a, GAPDH, MDH, RpS2 and RpS5, Wgl) totaling 5538 bp and one 1475 bp mitochondrial locus (COI) from each specimen using the primers and protocols described in Wahlberg and Wheat (2008). A three loci dataset (COI, EF1a, and Wgl) from 13 species (a subset of the species used here) that was previously used for phylogenetic analysis (Kodandaramaiah, 2009) was compiled from GenBank. Sequences of the sister genus Precis (Wahlberg et al., 2005, 2009) were also obtained from GenBank and served as the outgroup.

2.2. Phylogenetic analyses

We used PartitionFinder v1.1.1 (Lanfear et al., 2012) to determine optimal data partitioning schemes and evolutionary models before performing ML analyses using RAxML v8 (Stamatakis, 2014) on the CIPRES web portal (Miller et al., 2010). Nodal support values were calculated from 1000 bootstrap trees. We performed separate ML analyses for (a) the mitochondrial dataset, (b) the nuclear dataset, and (c) the combined nuclear and mitochondrial dataset of all sequenced loci. We used MrBayes v3.2 (Ronquist et al., 2012) on the CIPRES web portal (Miller et al., 2010) for Bayesian Inference (BI) of phylogeny of the combined nuclear and mitochondrial dataset and evolutionary models from PartitionFinder v1.1.1 (Lanfear et al., 2012). The program MrBayes was set to estimate the base frequencies and shape parameters from the data. We performed two independent runs with two chains per run for 10 million generations, sampling trees every 1000 generations. The convergence of independent runs was analyzed from the values of Potential Scale Reduction Factors (value close to one indicates convergence) (Gelman and Rubin, 1992). The consensus tree was reconstructed after discarding the first 25% of trees as burn-in. Posterior probability (PP) values provided a measure of nodal support.

2.3. Molecular dating and diversification analyses

For the molecular dating analysis, we included sequences of all other genera in Junonini (the tribe to which Hypolimnas belongs) (Wahlberg et al., 2005; Kodandaramaiah and Wahlberg, 2007) and used four members of the sister tribe Victoriniini (Wahlberg et al., 2005; Su et al., 2017) as outgroup. We estimated divergence times using BEAST v2.4 (Bouckaert et al., 2014) taking into account the secondary calibrations from Wahlberg et al. (2009) which is a family-level study of Nymphalidae that used seven fossils and host plant ages for node calibrations, and is based on a dataset of 10 gene regions. A recent paper by Su et al. (2017) on the subfamily Nymphalinae suggests that all clades are older than in Wahlberg et al. (2009), using a dataset of 3 gene regions calibrated with only two of the above mentioned seven fossils (Vanessa amerindica and Prodryas persephone: late Eocene) and host plant ages. The placement of these two fossils in the butterfly tree has recently been questioned in a morphological revision of all butterfly fossils (de Jong, 2017), with a recommendation to place the former fossil at the base of the subfamily Nymphalinae (rather than within the tribe Nymphalini) and not to use the later fossil at all, due to ambiguous characters. In a very recent study taking these recommendations into
account (Chazot et al., 2018), we find ages for Junoniini similar to Wahlberg et al. (2009). We thus base our analysis on times of divergence calibrated according to Wahlberg et al. (2009), where the age of Junoniini was estimated 29.07 million years ago (Mya) compared to 42.3 Mya by Su et al. (2017). However, we examined the extent to which variation in time estimates of higher clades could potentially influence age estimation of *Hypolimnas* (details in Table A2).

In our dataset, we calibrated the higher-level nodes (Junoniini + Victoriniini and Junoniini) with ages from the previous estimates and assigned normal distribution priors to the calibrated nodes. Following Wahlberg et al. (2009), we assigned an age of 29.07 Mya (95% highest posterior density [HPD]: 24.27–34.0) to the crown age of Junoniini and 41.08 Mya (95% HPD: 35.61–45.75) to the root (Junoniini + Victoriniini). Using the partitioning scheme from PartitionFinder v1.1.1 (Lanfear et al., 2012), we performed two independent runs of 25 million generations each, sampling every 2500 generations. We also examined the effect of tree priors (birth-death and Yule model) on age estimation. After confirming the convergence of runs with Tracer v1.6 (Rambaut et al., 2014), we discarded the initial 20% of each run as burn-in and summarized the tree using TreeAnnotater v2.3 (Bouckaert et al., 2014).

Due to poor convergence of two tree parameters (prior and posterior) in the BEAST analyses, we compared the age estimates from the BEAST analyses with those from a path-length based molecular dating analysis (Bouckaert et al., 2014). We performed a ML analysis of the Junoniini dataset following the procedure described in the previous section. The resulting ML tree was used to estimate relative ages of the nodes using the PATHd8 method (Britton et al., 2007), which uses the mean path length algorithm (Britton et al., 2002) with correction for a molecular clock. The absolute ages were estimated using two secondary calibrations as mentioned above, except that the root of the tree was fixed with the mean estimated age without any mention of age ranges. After comparing the divergence times estimated across analyses, we considered the time tree from BEAST analysis under a birth-death prior for further diversification analyses, retaining only one randomly chosen specimen per species.

We estimated the gamma statistic (Pybus and Harvey, 2000) using the R (v3.4.1; R Core team, 2017) package Laser v2.4.1 (Rabosky, 2006) to test whether the diversification rate varies across the tree. A negative gamma statistic indicates that the internal nodes are closer to the root than expected under a constant rate model, suggesting high diversification rate towards the base of the tree. A positive gamma statistic suggests that the diversification rate increased recently. We used the Monte Carlo Constant Rates (MCCR) test (Pybus and Harvey, 2000) to account for incomplete taxon sampling by adjusting the critical value of gamma.

We generated a Lineage Through Time (LTT) plot for the time tree and evaluated the pattern of lineage accumulation in the empirical LTT plot by comparing it with simulated LTT plots. We simulated 1000 trees under a constant diversification model (parameters from pure birth model) with 26 taxa using the R package TreeSim v2.3 (Stadler, 2011) to construct the simulated LTT curves. We compared the lineage accumulation pattern on the time tree to different models of diversification (Rabosky and Lovette, 2008) using log-likelihood scores and Akaiki Information Criterion (AIC) in the R package Laser v2.4.1 (Rabosky, 2006).

We used a Bayesian framework, employed in BAMM v2.5 (Bayesian Analysis of Macreovolutionary Mixtures; Rabosky, 2014), for modelling speciation and extinction rates over the phylogeny. We ran the analysis with four independent chains of 20 million generations each sampling every 20,000 generations. After removing first 10% of generations as burn-in, we analysed the sampled generations for rate shift configurations using the R package BAMMtools v2.1.0 (Rabosky et al., 2014).

### 2.4. Character state mapping

Information about geographic ranges of each species was compiled from various sources (Table A1) and each species was classified as present or absent in six biogeographic regions – Neotropics (South and Central America), Nearctic (Canada and the continental United States), Palearctic (East Asia, West Asia, Central Asia, North Asia and Europe), Afrotropics (Africa and Malagasy region), Oriental (South Asia and Southeast Asia) and Australasia (Australia, New Zealand, New Guinea and South Pacific Islands). The distributions were mapped onto the phylogeny using the online tool Interactive Tree Of Life (iTOl) v3 (Letunic and Bork, 2016). We also compiled larval host plant data from various sources (Table A1) and mapped them onto the phylogeny in the same manner as above.

### 2.5. Wolbachia infection

We tested all samples for *Wolbachia* infection using a PCR-based assay by amplifying the *Wolbachia* surface protein gene *wsp* using the primer pair wsp 81F – wsp 691R (Zhou et al., 1998). The infection status of samples that tested positive based on amplification of *wsp* was confirmed by amplifying the faster evolving cell-cycle gene *fts*Z (using the primer pair *fts*Z1F – *fts*Z1R; Baldo et al., 2006). All samples that were positive for *wsp* were also positive for *fts*Z, and were thus scored as positive for presence of *Wolbachia*. We further amplified four additional loci (*cox*1, *fbpA*, *gatB*, *hcpA*) from infected samples for multi-locus sequence typing (MLST) (Baldo et al., 2006) to identify and explore *Wolbachia* strain diversity.

### 3. Results

#### 3.1. Phylogenetic relationships

The monophyly of *Hypolimnas* was strongly supported (Bootstrap [BS] = 100; Posterior Probability [PP] = 1) in all analyses. Irrespective of the dataset used or type of analysis, the trees had three deeply divergent, strongly supported (BS > 90; PP = 1) branches, which we refer to as clades I, II and III. For simplicity, we describe results from ML and BI analyses of the combined nuclear and mitochondrial dataset (Fig. 1), except where otherwise stated. The results of the ML analyses of mitochondrial and nuclear loci are shown in Figs. A1 and A2.

Clade I comprised three strongly supported (BS > 90; PP = 1) lineages; however, the inter- and intralineage relationships are not strongly supported (BS < 60; PP < 0.6) for some nodes. The mitochondrial and nuclear ML trees supported the monophyly of one or more of these three lineages poorly (BS < 60) or did not recover them as monophyletic at all (Figs. A1 and A2).

In clade II, *H. deceptor* was sister to *H. anthedon* with strong support (BS = 100; PP = 1). The taxon known as *dubius* was previously regarded as a full species, and our results confirm that it is a junior synonym of *H. anthedon*. At every locus, *H. anthedon* was polyphyletic with regard to *H. “dubius*”, but the clade with both taxa was monophyletic with strong support (BS = 98; PP = 1; Figs. A1 and A2).

The relationship between clade I and clade II is unclear: their sister group, *H. bo- morrhous* + *H. alimena*, was not strongly supported (BS < 60; PP < 0.6) for some nodes. The mitochondrial and nuclear ML trees supported the monophyly of one or more of these three lineages poorly (BS < 60) or did not recover them as monophyletic at all (Figs. A1 and A2).

Clade III included the remaining 16 species (Fig. 1). Within this group, *H. misippus* diverged first followed by the divergences of *H. bolina*, *H. inopinata*, *H. almena*, and the clade *H. anomala + H. antilope*. This was followed by the divergences of all other species.

Within the strongly supported clade of *H. anomala + H. antilope* (BS = 100; PP = 1), the monophyletic *H. antilope* (BS = 74; PP = 0.7) was nested within *H. anomala*, rendering the latter paraphyletic in all analyses. Four species (*deceptor, monteironis, salmacis* and *errabunda*)
4.3. Molecular dating and diversification rates

The divergence analyses with BEAST using secondary calibrations from Wahlberg et al., (2009) estimated the mean crown age of *Hypolimnas* to be 13.4 Mya (under a birth-death prior) and 16.8 Mya (under a Yule-model), while the PATHd8 analysis suggested an origin 15.1 Mya.

Using the time tree from BEAST analysis under a birth-death prior, the MCCR test (using total species = 29, sampled species = 26, gamma = −1.163) resulted in a critical gamma of −1.83 (corrected p-value = 0.17). As the critical gamma is more than −1.645, there is weak signal for diversification rate variation across the tree; a lower value would indicate an early burst of diversification (Fordyce, 2010). Moreover, model comparison indicated that a constant rate model (pure birth model; AIC = 10.852; logLik = −4.426) best fits the lineage accumulation pattern in comparison to the models that allow variation in speciation or extinction rates or both (Table A3). The placement of the empirical LTT plot within the simulated curves generated under a constant diversification rate (Fig. A3) further corroborates the finding that the diversification rate was constant. This was corroborated by the BAMM analysis which supported no change in diversification rate, with constant speciation and extinction rates (Fig. A4).

4.4. Host plant and biogeography

Mapping the phylogeny with the geographic distributions of the species (Fig. 2) showed that two deeply divergent lineages, clade I and clade II, are distributed across the Afrotopics. While most species in clade III are distributed in the Oriental and Australasian regions, two species in this group are widespread globally – *bolina* is found in Australasia, Oriental, Paleartctic and Afrotopics, while *misippus* is found in Australasia, the Oriental region, the Afrotropics and the New World (Nearctic + Neotropic; possibly introduced (Larsen, 1996; Parsons, 1998).

Most species of *Hypolimnas* feed on Urticaceae (Fig. 2), while the African clades are confined to Urticaceae, the Oriental and Australasian species feed on Euphorbiaceae, Malvaceae, Moraceae, Portulacaceae and Acanthaceae in addition to Urticaceae. The two widespread species, *H. misippus* and *H. bolina*, feed on a large number of species across many different (> 10) families.
4. Discussion

4.1. Phylogenetic relationships and taxonomic implications

The relationships in the tree from the combined nuclear and mitochondrial dataset are largely concordant with those inferred independently from the nuclear-only dataset and from the mitochondrial dataset. Deeper nodes were generally strongly supported; however, recent divergences among species were not well resolved (BS < 60). Across all analyses, the species were recovered in three strongly supported deeply divergent clades (I, II and III) (BS > 90; PP = 1).

The relationships among species within two of the deeply divergent clades – clade I and II – are not well resolved at the internal nodes; only four of the eight nodes were resolved (BS > 80) in the mitochondrial tree and only three were resolved in the combined nuclear tree (BS > 80) and five in the concatenated tree (BS > 80). In clade III, most of the nodes were poorly resolved (BS < 60) across all analyses, except the divergence of H. misippus from the rest of the species in the clade. Overall, the branch lengths for recent species divergences were relatively smaller in the combined nuclear tree compared to those in the mitochondrial tree (Figs. A1 and A2). The nodes corresponding to these small branch lengths are often associated with low node support (BS < 60). This pattern of short branch length being associated with low node support indicates weak phylogenetic signal even after including a relatively large dataset of eight loci. Weak phylogenetic signal in the dataset may result from incomplete lineage sorting, introgression and hybridization events.

### Table 1

Percent sequence divergence in mitochondrial haplotypes (mitotypes; COI: 580 bp) among three *Hypolimnas* species with respect to their *Wolbachia* infection status. One *H. bolina* specimen (DL95-0001) was excluded from this analysis because its infection status was uncertain; another *Wolbachia*-infected *H. bolina* (NW98-4) specimen had a mitotype that differed from other infected specimens (see Section 4.4) and was therefore excluded.

<table>
<thead>
<tr>
<th></th>
<th>bolina (8)</th>
<th>alimena (4)</th>
<th>anomal (2)</th>
<th>bolina (3)</th>
<th>alimena (4)</th>
<th>anomal (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Infected</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bolina</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>alimena</td>
<td>0.07</td>
<td>0.07</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>anomal</td>
<td>0.07</td>
<td>0.07</td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Uninfected</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>bolina</td>
<td>0.07</td>
<td>0.07</td>
<td>0.08</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>alimena</td>
<td>0.05–0.06</td>
<td>0.05–0.06</td>
<td>0.08</td>
<td>0.05–0.06</td>
<td>0.01</td>
<td></td>
</tr>
<tr>
<td>anomal</td>
<td>0.06</td>
<td>0.06</td>
<td>0.02</td>
<td>0.08</td>
<td>0.07</td>
<td>0</td>
</tr>
</tbody>
</table>

The number in the parentheses represents the number of specimens.
among recent lineages, or slow rates of molecular evolution. Discordance among gene trees is another plausible reason for poor phylogenetic resolution (Smith et al., 2015). However, low nodal support within individual gene trees suggests that strong phylogenetic discordance among gene trees is unlikely to result in the patterns found here. Moreover, it is possible that discordances in different regions of the tree result from different evolutionary phenomena.

Only seven species were recovered as monophyletic (BS > 80) and monophyly of five additional species could not be tested objectively because they were rare and represented in our analyses by single specimens. The monophyly of some other species were uncertain because of: (i) paraphyly (H. anomala), (ii) discrepancies between clades (e.g., H. pandarus), and (iii) poor nodal support (BS < 60) for monophyly (e.g., H. deois). Incomplete lineage sorting or gene flow between recently diverged species are plausible explanations for these patterns. Additionally, it is possible that morphologically delimited species in this genus of phenotypically divergent and highly evolvable Batesian mimics may not reflect true species boundaries (Wei et al., 2017).

We note that some inconsistencies between the nuclear and mitochondrial trees and non-monophyly of some species may be due to incomplete sequences in our datasets. For example, the two H. deceptor specimens we include have just one or two nuclear loci plus COI, but without a single nuclear locus shared between both specimens. As a result, in the COI tree they group together as sister to H. anthedon whereas they are within the H. anthedon clade in the nuclear tree. The phylogenetic position of H. inopinata is tentative because we only had 758 bp data from three loci for this specimen and note that the nuclear and mitochondrial trees differ in the placement of this taxon. Although the two specimens of H. errabunda formed a distinct clade in the mitochondrial tree, they were paraphyletic with regards to H. octocula and H. pihoeka in the nuclear tree. This may result from poor phylogenetic signal from only two nuclear loci that amplified in the H. errabunda specimens.

Our molecular phylogenetic results confirm that H. dubius Palisot de Beauvois (1820) is conspecific with H. anthedon Doubleday (1845). The name H. dubius is invalid because it is a junior homonym, and the taxon should be called H. anthedon (Lamas 2015). In addition, our results suggest that H. pihoeka Kirsch (1877) is a junior synonym of H. octocula Butler (1869) and that H. antilope Cramer (1877) is a junior synonym of H. anomala Wallace (1869). Though we do not formally synonymize these latter two pairs in this publication, it likely that future molecular and/or morphological study will support these tentative taxonomic conclusions.

4.2. Historical biogeography

Although relationships among the three deeply divergent clades were not strongly supported, we observed a common historical biogeographic pattern across trees generated with all datasets. Irrespective of the conflicting topologies, the earliest diverging clade of Hypolimnas (depending on the tree, either only clade I or the clade containing both clade I and II; Figs. 1 and A5) and the sister genus Precis are both restricted to Africa. Other genera in the tribe Junoniini also likely originated in Africa, except H. alimena and H. anthedon, which are widespread in various regions of the world (e.g., North America, Europe, Asia, Australia, and South America).

Within Hypolimnas, the Early Miocene lineage diversification is critical to understanding the biogeographic history of the genus. This diversification occurred during the late Early Miocene to early Middle Miocene (17–15 Ma), which is consistent with the widespread distribution of the genus in various regions of the world at that time. The Early Miocene was a period of increased aridification and fragmentation of the landmass, which likely contributed to the diversification of Hypolimnas across different regions.

The biogeographic history of Hypolimnas is characterized by two major events: the radiation of the African lineage and the dispersal of the Oriental lineage into Asia and eventually the New World. The African lineage diversified in Africa and then expanded into Asia, with the Oriental lineage diverging later and dispersing into the New World. This biogeographic pattern indicates that the African lineage was the ancestral lineage of Hypolimnas, and the dispersal into Asia and the New World occurred through long-distance dispersal events.

The dispersal of Hypolimnas into Asia likely occurred during the Late Miocene, with the radiation of the Oriental lineage in the Late Miocene. This radiation occurred in the region of present-day India, where the climate was more suitable for the survival of Hypolimnas. The dispersal into the New World likely occurred in the Late Pliocene, with the radiation of the North American lineage. This radiation occurred in the region of present-day North America, where the climate was more suitable for the survival of Hypolimnas.

The biogeographic history of Hypolimnas is characterized by the radiation of the African lineage in Africa and then its dispersal into Asia, with the Oriental lineage diverging later and dispersing into the New World. This biogeographic pattern indicates that the African lineage was the ancestral lineage of Hypolimnas, and the dispersal into Asia and the New World occurred through long-distance dispersal events.
GenBank) from Africa, India, China, Myanmar, Thailand, Malaysia, and islands of the South Pacific (Charlat et al., 2009; Duplouy et al., 2010). The divergent mitotypes specific to *H. alimena* occurred in the specimens from Indonesia and Australia. However, the divergent mitotypes specific to *H. bolina* occurred only in Indonesia, French Polynesia, Fiji, Papua New Guinea, and Australia (compiled from our samples from Southeast Asia and 13 specimens from GenBank; Charlat et al., 2009; Duplouy et al., 2010) (Fig. 3). Phylogenetic placement of these mitochondrial sequences of these additional *H. bolina* specimens and conspecifics in our dataset (Fig. A6) further corroborate the geographic distribution of mitotypes.

To explain this distribution pattern of mitotypes, we predicted that the common mitotype occurs in specimens infected with *Wolbachia*. This prediction was based on two previous findings. First, many populations of *H. bolina* from Southeast Asia and islands of the South Pacific are infected by the intracellular bacterium *Wolbachia* (Charlat et al., 2009; Duplouy et al., 2010). Second, *Wolbachia* can mediate transfer and spread of mitochondrial genes across recently diverged species (Hurst and Jiggins, 2005). Thus, we hypothesized that *H. bolina* and *H. alimena* specimens that share the common mitotype have the same *Wolbachia* strain, and that the rest of the *H. bolina* and *H. alimena* specimens either have different *Wolbachia* strains or are uninfected. Characterization of the sequence types of *Wolbachia* infecting specimens in our analysis confirmed that *H. bolina* and *H. alimena* specimens that share the common mitotype also have the male-killing *Wolbachia* sequence type wBol (MLST 125). All other specimens of these species were not infected. We were unable to characterize the particular variant of the wBol strain genetically. However, Charlat et al., (2009) and Duplouy et al., (2010) previously reported that the common mitotype is in strong linkage disequilibrium with wBol1a, and we therefore surmise that the same *Wolbachia* strain infects our specimens. Moreover, one specimen of *H. bolina* from Sulawesi, Indonesia, tested positive for *Wolbachia* but lacked the common mitotype. It is plausible that this specimen may be infected with a non-male-killing *Wolbachia* strain like wBol1b or wBol2 (following Charlat et al., 2009).

Thus, we show that when a particular invariant mitotype is found in populations of *H. bolina* throughout its geographic range, it is mostly associated with the *Wolbachia* strain wBol (Fig. 3). We found that all populations of *H. bolina* in Africa, East Asia, and South Asia have this identical, common mitotype, but individuals in Southeast Asian and South Pacific populations have different mitotypes co-occurring with the common mitotype. From this pattern, we infer that (i) the strong association between *Wolbachia* strain and the common mitotype indicates that their co-transmission is perfect (e.g., strong linkage disequilibrium), and (ii) the absence of the particular *Wolbachia* strain in the island populations of Southeast Asia and South Pacific indicates that dispersal of that strain to these regions is recent or that competitive interactions among multiple strains occur in those populations (Charlat et al., 2009). We also detect the presence of the common mitotype and associated *Wolbachia* strain in *H. alimena*, which indicates plausible occurrence of *Wolbachia*-mediated mitochondrial gene flow between *H. bolina* and *H. alimena*. The most likely explanation is that the *Wolbachia* strain and associated mitochondrial genes were transferred between species via hybridization. Following a selective sweep of *Wolbachia*, the associated mitotype spread into the new species. Further investigation is necessary to address the extent of gene transfer between the species and the role of *Wolbachia* in mediating interspecific transfer. Such strong concordance between *Wolbachia* and mitochondrial genes, and their co-transmission across species boundary have been recorded in *Drosophila* (between *D. simulans* and *D. mauritiana* (Rousset and Solignac, 1995; Ballard, 2000), *Acraea* butterflies (A. ecedon and A. ecedana) (Jiggins, 2003) and *Eupristina* wasps (cryptic species *E. verticillate*-1 and *E. verticillate*-2) (Sun et al., 2011). Therefore, *Wolbachia*-mediated introgression is probably more prevalent than previously thought. Interestingly, a divergent *Wolbachia* strain was found in *H. anomala*, a close relative of *H. bolina* and *H. alimena*. We noticed that out of three *H. anomala* specimens in our study, two specimens are infected and share an identical mitotype while one uninfected specimen has a distinct mitotype.

5. Summary and conclusions

We present the most comprehensive phylogeny of *Hypolimnas* butterflies based on a large dataset of nine genetic markers. Our analyses suggest that widespread topological uncertainty and non-monophyly of species results from the recency of species divergences, possibly
through incomplete lineage sorting and/or introgression. We find support for the ‘Out-of-Africa’ hypothesis in the group, and hypothesize that the Mid-Miocene Climatic Optimum facilitated dispersal from Africa to Asia. In Hypolimnas, the non-African Hypolimnas lineages have broader host plant ranges compared to the African lineages; however, we find no support for change in diversification rate associated with the host range extension. We surmise that the wide geographic range and broad host plant use by *H. bolina* and *H. misippus* could potentially result in population fragmentation; future evaluations to investigate these possibilities will be interesting. Intriguingly, some individuals from the species *H. bolina* and *H. almeta* share both a mitochondrial haplotype and the same *Wolbachia* strain, and our analyses suggest *Wolbachia*-mediated inter-specific introgression. Further investigations might explore the plausibility of *Wolbachia*-mediated gene transfer between *Hypolimnas* species and associated geographical patterns.

Acknowledgements

We thank W. John Tennent, Brian Patrick, Szabolcs Sáfián, and James Stewart for providing additional specimens for the study. We thank Teresa Di Micco de Santo who helped out with sampling and photography of specimens from ABRI. We are grateful to Perry Buenavente and Ven Samarita of the National Museum of the Philippines for coordinating relevant permits and authorizations needed under the National Cultural Heritage Act. Field collection in Indonesia was conducted under permits to DJL and DP from Ristek and other relevant Indonesian authorities.

Funding sources

The majority of the sequencing costs was covered by grants from the Department of Science and Technology to UK (INSPIRE Faculty Award DST/INSPIRE/04/2013/000476 and DST-RFBR-P-155). RKS is supported by a research fellowship from Council of Scientific and Industrial Research, India. DJL and KAP are supported by grant DEB-1541557 from the US National Science Foundation. Specimen collection in the Philippines was funded NSF DEB-110380 to DJL and fieldwork in Indonesia was funded in part by Grant 9285-13 from the National Geographic Society Committee for Exploration and Research to DJL. NW acknowledges support from the Swedish Research Council.

Data deposition

The sequences generated in the study are deposited into GenBank (Accession # MG886873 - MG887746, MG920062 - MG920140, and MG934344 - MG934359) (detail in Appendix A).

Appendix A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at https://doi.org/10.1016/j.ympev.2018.02.001.

References


Saito, A., 2002. Mimicry in butter
Rousset, F., Solignac, M., 1995. The evolution of single and double
Smith, D.A.S., Gordon, J., 1987. The genetics of the butterfly Hypolimnas misippus (L.):
Wangberg, N., Leneveu, J., Kodandaramaiah, U., Peña, C., Nylin, S., Freitas, A.V.L.,
Wahlberg, N., Wheat, C.W., 2008. Genomic outposts serve the phylogenomic pioneers:
Saito, A., 2002. Mimicry in butter