



WAGENINGEN
UNIVERSITY & RESEARCH

Allopatric origin of cryptic butterfly species that were discovered feeding on distinct host plants in sympatry.

Molecular Ecology

McBride, L.C.; Velzen, R.; Larsen, T.B.

<https://doi.org/10.1111/j.1365-294X.2009.04309.x>

This article is made publicly available in the institutional repository of Wageningen University and Research, under the terms of article 25fa of the Dutch Copyright Act, also known as the Amendment Taverne. This has been done with explicit consent by the author.

Article 25fa states that the author of a short scientific work funded either wholly or partially by Dutch public funds is entitled to make that work publicly available for no consideration following a reasonable period of time after the work was first published, provided that clear reference is made to the source of the first publication of the work.

This publication is distributed under The Association of Universities in the Netherlands (VSNU) 'Article 25fa implementation' project. In this project research outputs of researchers employed by Dutch Universities that comply with the legal requirements of Article 25fa of the Dutch Copyright Act are distributed online and free of cost or other barriers in institutional repositories. Research outputs are distributed six months after their first online publication in the original published version and with proper attribution to the source of the original publication.

You are permitted to download and use the publication for personal purposes. All rights remain with the author(s) and / or copyright owner(s) of this work. Any use of the publication or parts of it other than authorised under article 25fa of the Dutch Copyright act is prohibited. Wageningen University & Research and the author(s) of this publication shall not be held responsible or liable for any damages resulting from your (re)use of this publication.

For questions regarding the public availability of this article please contact openscience.library@wur.nl

Allopatric origin of cryptic butterfly species that were discovered feeding on distinct host plants in sympatry

CAROLYN S. McBRIDE,*§ ROBIN VAN VELZEN† and TORBEN B. LARSEN‡

*Center for Population Biology, University of California, Davis, CA 95616, USA, †National Herbarium of the Netherlands – Wageningen Branch, Biosystematics Group, Wageningen UR, Generaal Foulkesweg 37, 6703 BL Wageningen, The Netherlands, ‡Jacobys Alle 2, 1806 Frederiksberg, Denmark

Abstract

Surveys of tropical insects are increasingly uncovering cryptic species – morphologically similar yet reproductively isolated taxa once thought to comprise a single interbreeding entity. The vast majority of such species are described from a single location. This leaves us with little information on geographic range and intraspecific variation and limits our ability to infer the forces responsible for generating such diversity. For example, in herbivorous and parasitic insects, multiple specialists are often discovered within what were thought to be single more generalized species. Host shifts are likely to have contributed to speciation in these cases. But when and where did those shifts occur, and were they facilitated by geographic isolation? We attempted to answer these questions for two cryptic species within the butterfly *Cymothoe egesta* that were recently discovered on different host plants in central Cameroon. We first used mtDNA markers to separate individuals collected on the two hosts within Cameroon and then extended our analysis to incorporate individuals collected across the entire pan-Afrotropical range of the original taxon. To our surprise, we found that the species are almost entirely allopatric, dividing the original range and overlapping only in the narrow zone of West-Central Africa where they were first discovered in sympatry. This finding, combined with analyses of genetic variation within each butterfly species, strongly suggests that speciation occurred in allopatry, probably during the Pleistocene. We discuss the implications of our results for understanding speciation among other cryptic species recently discovered in the tropics and argue that more work is needed on geographic patterns and host usage in such taxa.

Keywords: Africa, allopatric speciation, cryptic species, *Cymothoe*, *Rinorea*, sympatric speciation

Received 10 November 2008; revision received 15 June 2009; accepted 18 June 2009

Introduction

The increasingly frequent discovery of cryptic species within what were previously thought to be single interbreeding entities calls previous estimates of biodiversity into question, challenges conservation efforts, and poses important questions about the speciation process (Bickford *et al.* 2007). This is nowhere more true than among

tropical herbivorous and parasitic insects. Studies of host use, immature stages, and/or DNA in these organisms often uncover multiple specialized species within purportedly more generalized taxa (Amiet 1997; Morehead *et al.* 2001; Berkov 2002; Hebert *et al.* 2004; Burns & Janzen 2005; Smith *et al.* 2006, 2007; Burns *et al.* 2008).

Despite the impressive efforts behind the discovery of many of these tropical host-specific cryptic species, almost all are described from just one location. This is due to the logistical difficulty of gathering the kinds of samples necessary for their identification (e.g. immature stages, host plants/animals) over large tropical areas. For example, *Palame mimetica* was once considered to be

Correspondence: Carolyn Sarah McBride, Fax: 212-327-7238; E-mail: lmcbride@rockefeller.edu

§Present address: Laboratory of Neurogenetics and Behavior, The Rockefeller University, 1230 York Avenue, Box 63, New York, NY 10065, USA

a single beetle species known from French Guiana, Brazil and Venezuela. In 2002, Amy Berkov published the results of a rearing experiment and phylogenetic analysis that revealed at least two cryptic species within the taxon, attacking distinct host trees at slightly different times of year in a forest just north of Saül, French Guiana (Berkov 2002). Likewise, *Astrartes fulgerator* was once thought to be a single variable butterfly species distributed across the Americas from the southern United States to northern Argentina. The collection and rearing of thousands of *A. fulgerator* caterpillars in one region of Costa Rica, however, revealed this taxon to comprise as many as 7, or possibly even 10, species, mostly limited to distinct host plants (Hebert *et al.* 2004; and see Brower 2006). In both of these cases, we know that the recently discovered cryptic species co-occur in one small area, but we can only guess how they share or partition the rest of the geographic range attributed to the taxon within which they were hidden. The same is true of many other sets of host-specific cryptic species recently discovered within tropical herbivorous and parasitic insects (Amiet 1997; Morehead *et al.* 2001; Burns & Janzen 2005; Smith *et al.* 2006, 2007; Burns *et al.* 2008).

Our lack of information on the distributions of these taxa makes it difficult to answer some of the most interesting questions posed by their discovery. For example, what factors were responsible for initial divergence? The fact that the species specialize on distinct hosts suggests that ecological adaptation to novel resources contributed (Stireman *et al.* 2005; Funk *et al.* 2006). But did such adaptation occur in sympatry or was it facilitated by geographic isolation? Allopatric speciation is generally accepted to be the dominant mode of speciation in animals and plants (Coyne & Orr 2004), and there is no reason why divergent host adaptation should not contribute to allopatric speciation events (Funk 1998). Nevertheless, substantially more attention has been devoted to the potential role of host adaptation and host shifts to speciation in sympatry than in allopatry. Indeed, host-specialized insects have several characteristics that should facilitate sympatric speciation: (i) their lives are often so intimately intertwined with those of their hosts that new hosts exert strong multifarious selection, (ii) alternative hosts provide discrete resources without intermediates and (iii) for insects that mate on or near their hosts, reproductive isolation can evolve as a direct pleiotropic effect of selection on host preference (Bush 1975; Rice 1984; Kondrashov & Mina 1986; Diehl & Bush 1989; Berlocher & Feder 2002; Dres & Mallet 2002).

It would be interesting to know what role, if any, sympatric speciation has played in generating cryptic tropical insect diversity. Gathering data on current and ancestral geographic ranges can help to answer this question (Diegisser *et al.* 2006; Fitzpatrick & Turelli

2006; Jordal *et al.* 2006) and may eventually lead to a general understanding of patterns of diversification among such taxa. Wide geographic sampling of cryptic tropical taxa is also important for basic taxonomy and the design of effective conservation strategies.

In this study, we address two cryptic species formerly hidden within the Afrotropical butterfly *Cymothoe egesta* (Nymphalidae). For most of the 20th century, *C. egesta* was thought to be a single, variable butterfly species distributed in forested areas across sub-Saharan Africa from Sierra Leone in the west to Tanzania and Uganda in the east. The species comprised two subspecies; West African populations were considered to belong to the nominate *C. e. egesta*, while Central/East African populations were considered to belong to *C. e. confusa*. This situation changed in 1997, when Jean-Louis Amiet described two forms of *C. egesta* feeding on different host plants in sympatry in the vicinity of Yaoundé, central Cameroon (within the range of *C. e. confusa*) (Amiet 1997). Adults of the two forms were nearly indistinguishable – the most consistent difference being that the background wing colour of freshly eclosed males was the typical *C. egesta*-like dull yellow in one form (referred to by Amiet as *C. e. confusa*), and slightly more orange in the other form (referred to by Amiet as *C. 'egesta orange'*). In contrast, the immature stages of the two forms showed obvious colour differences and, as mentioned above, were found feeding on distinct host plants; Amiet collected *C. e. confusa* on plants in the *Rinorea ilicifolia* species group (Violaceae) and found *C. 'egesta orange'* on *Rinorea lepidobotrys*, a species representing a distinct lineage within the genus *Rinorea* (Amiet 1997).

We used molecular methods to describe the full geographic distributions of these two cryptic species, and then to infer the geographic circumstances under which they initially diverged. We found neither that the two species were fully sympatric nor that one was endemic to a portion of the broader range of the other. Instead, they were largely allopatric, one in the west and the other in the east, overlapping only in the relatively narrow zone of central Cameroon surveyed by Amiet. This finding, combined with analyses of population structure and intraspecific genetic diversity, strongly suggests that speciation occurred in allopatry. We discuss the implications of these results for understanding speciation among other cryptic tropical taxa and emphasize the importance of continued attention to such species.

Materials and methods

Insect collection

We searched known and potential *Rinorea* host plants for immature stages of *C. egesta* in six localities across

central and south Cameroon in 2003 and 2006. We assembled additional specimens, mostly captured as adults by other collectors, from over 30 localities across the entire range formerly attributed to *C. egesta*, including, from west to east, Sierra Leone, Guinea (Conakry), Ghana, Nigeria, Cameroon, Gabon, Democratic Republic of Congo, Zambia and Tanzania. Figure 1 shows a map of all collection localities, and Table S1 (Supporting Information) provides detailed information on each of the 86 assembled specimens, including collection locality, collector, year collected, life stage collected and host plant (if known).

PCR amplification and sequencing

We generated two DNA sequence data sets from the assembled specimens. The 'phylogenetic' set was designed to maximize phylogenetic signal and thus comprised relatively long 1345 bp segments of the mtDNA gene *COI* from 59 geographically representative individuals. The 'barcode' set was designed to maximize taxon sampling and comprised shorter, but overlapping, 658-bp

fragments of *COI* (DNA barcodes) from a larger group of 86 individuals (the same 59 plus an additional 27). The 'phylogenetic' data set was generated in the following way: We extracted DNA from 1 to 2 legs per specimen using a QIAGEN DNeasy kit. We then amplified the 1345-bp segment of *COI* for the first several specimens using the generic Nymphalid primer pairs HCO/LCO (Folmer *et al.* 1994) and Jerry/Pat (Simon *et al.* 1994) and the PCR recipe and cycling protocol of Zimmermann *et al.* (2000). For all subsequent specimens we used the custom designed primers found in Table S2 and the PCR recipes and cycling protocols found in Tables S3 and S4 (Supporting Information) respectively. The region was amplified in two segments from well-preserved specimens and in four shorter segments from poorly preserved specimens. The PCR amplicons were purified using either a Qiaquick (QIAGEN) or MSB Spin PCRapace (Invitex) kit, sequenced on an ABI 3730 Capillary Electrophoresis Genetic Analyzer (both the forward and reverse strands using the same primers as in the PCR reactions), and edited/assembled using CodonCode Aligner (CodonCode Corporation) or the

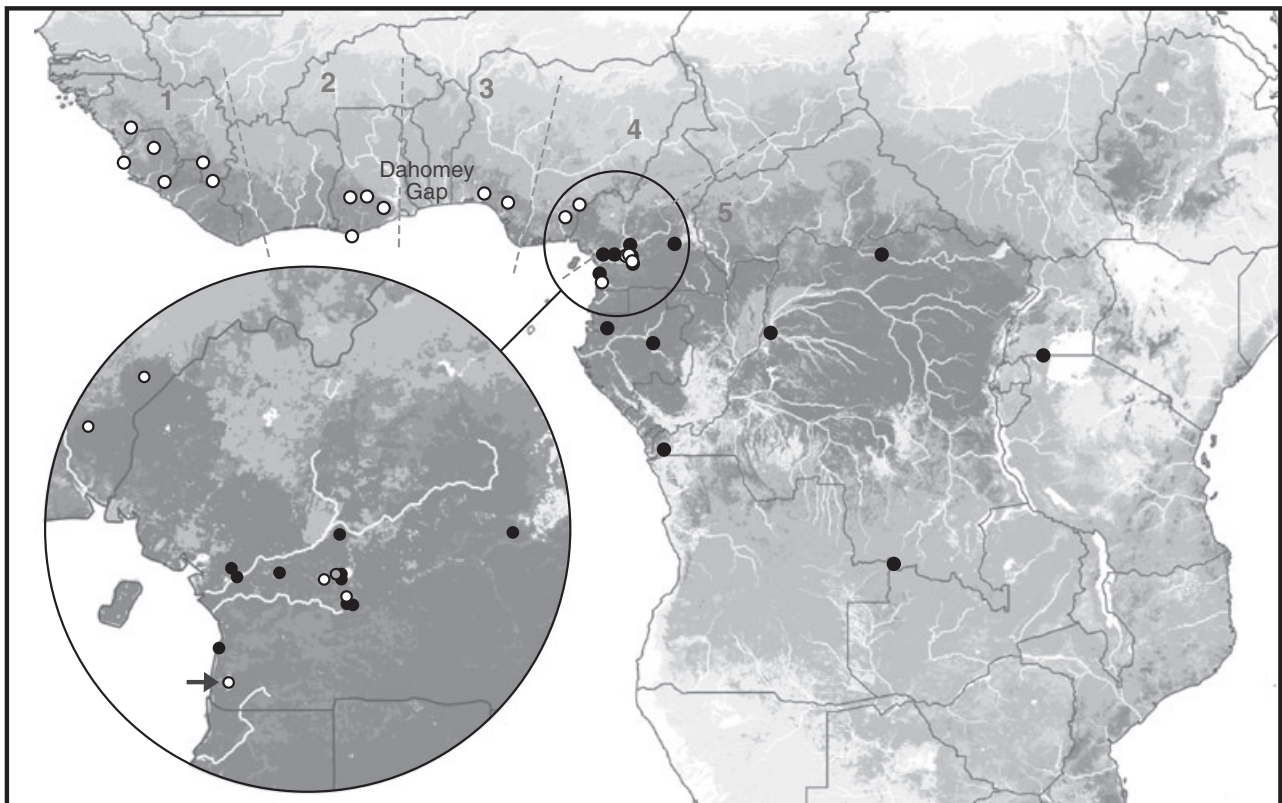


Fig. 1 Map of sub-Saharan Africa showing the collection sites for the specimens examined in this study. Sites where *C. egesta* s. s., *C. confusa*, or both were collected are represented by white, black and grey circles respectively. The inset highlights the zone of overlap within Cameroon. The small black arrow in the inset points to the locality where six larvae were collected on *R. brevibracmosa*. Dashed grey lines demarcate the five biogeographical sub-regions in which *C. egesta* s. s. occurs. These subregions are well defined by butterfly endemism and distribution patterns, but boundaries are not absolute (Larsen 2005).

Staden Package (Staden 1996; version 1.6.0). All sequences were deposited in GenBank (accession nos FJ422147–FJ422211). The 'barcode' data set was generated by trimming the 'phylogenetic' data set down to the appropriate segment and then adding the barcodes of 27 additional individuals. The new barcodes were generated by the Canadian Centre for DNA barcoding in Guelph, Canada in the following way: DNA was extracted using a routine silica-based 96-well extraction automation protocol (Ivanova *et al.* 2006), and the target 658-bp fragment was amplified and sequenced as described by Hajibabaei *et al.* (2006). These were deposited both in GenBank (accession nos GQ200316–GQ200343) and the Barcode of Life Database (BOLD; project code EGES).

We also attempted to amplify a 350-bp segment of the nuclear gene *wingless* from a subset of our samples. Unfortunately, however, the locus would not amplify from the majority, which had been kept as papered/pinned specimens at room temperature for 5 or more years. The data that were obtained (from four individuals) showed little variation (mean pairwise divergence <0.3%) and were not analysed further.

Statistical analyses

To infer the evolutionary relationships among the 59 DNA sequences in the 'phylogenetic' data set, we first identified the most appropriate models of sequence evolution using ModelTest version 3.7 (Posada & Crandall 1998), and then, for each model in the 95% credible set, conducted Bayesian phylogenetic analyses using MrBayes version 3.1.2 (Huelsenbeck & Ronquist 2001) with 3 outgroup sequences (one each from *C. orphnina*, *C. ogova* and *C. fontainei*). All models yielded similar results, and we report here those inferred under the best model only – a general-time-reversible model with a proportion of invariant sites and gamma distributed rate variation among the remaining sites (nst = 6, rates = invgamma). The two Markov Chain Monte Carlo runs were allowed to run for 4 million generations with the temperature differential reduced from 0.2 to 0.1 to optimize mixing among hot and cold chains within each run (ngen = 4 000 000, Temp = 0.1). The two runs converged (with the standard deviation of split frequencies dropping below 0.01 and the log-likelihood scores becoming stable) after approximately 2 million generations and so only trees sampled after this point were used to generate the consensus tree. We conducted a second analysis for the 86 sequences in the 'barcode' data set; we used the program MEGA4 (Tamura *et al.* 2007) to compute pairwise distances under the Kimura-2-parameter model (K2P; Kimura 1980) and infer a haplotype tree by Neighbour Joining.

We examined the structure of genetic variation within each cryptic species using the 'phylogenetic' data and the program Arlequin 2.000 (Schneider *et al.* 2000). First we pruned from the data set all but a single representative from groups of potential siblings – individuals that shared identical COI haplotypes and were collected as larvae on plants in close proximity on the same day. Removal of these sequences did not significantly alter the results. We then implemented one AMOVA for each of the cryptic species to quantify the effects of major biogeographic breaks on intraspecific gene flow. The model used for *C. egesta sensu stricto* partitioned variation (i) between the regions on either side of the Dahomey Gap, (ii) among biogeographic zones nested within each of those regions and (iii) within biogeographic zones. The model used for *C. confusa* partitioned variation (i) among biogeographic zones (Cameroon vs. the combination of all regions further east) and (ii) within biogeographic zones. Third, we estimated two measures of nucleotide diversity within zones – π (Tajima 1983) and θ_w (Waterson 1975). Finally, we examined the hypothesis that low overall nucleotide diversity in certain zones was the result of recent range/population expansion by computing Tajima's *D* (Tajima 1989).

We estimated the time of divergence between the two cryptic species in two different ways based on the 'phylogenetic' data set. First, we used the program MDIV (Nielsen & Wakeley 2001) with an HKY model that excludes migration ($M = 0$; reciprocally monophyletic data provide essentially no evidence of migration). Four independent runs of 5 million cycles each gave similar results, indicating convergence. Scaled divergence times were converted into absolute divergence times using two different mutation rate estimates for COI in insects: a relatively high rate of 1.15% per million years (Brower 1994) and a relatively low rate of 0.78% per million years (Zakharov *et al.* 2004). Second, we calculated the 'net divergence' between the two species. This estimator of divergence assumes that the ancestral population size was the average of the descendent population sizes and is calculated by subtracting the average of the within species π values from the mean pairwise divergence between species (Nei 1987). Net nucleotide divergence was converted into absolute divergence time assuming a molecular clock and using the same two mutation rate estimates mentioned above.

Results

Genetic separation of the cryptic species

The 'phylogenetic' data set (1345 bp of COI from 59 specimens) included 45 polymorphisms defining 23 unique haplotypes. The Bayesian tree in Fig. 2

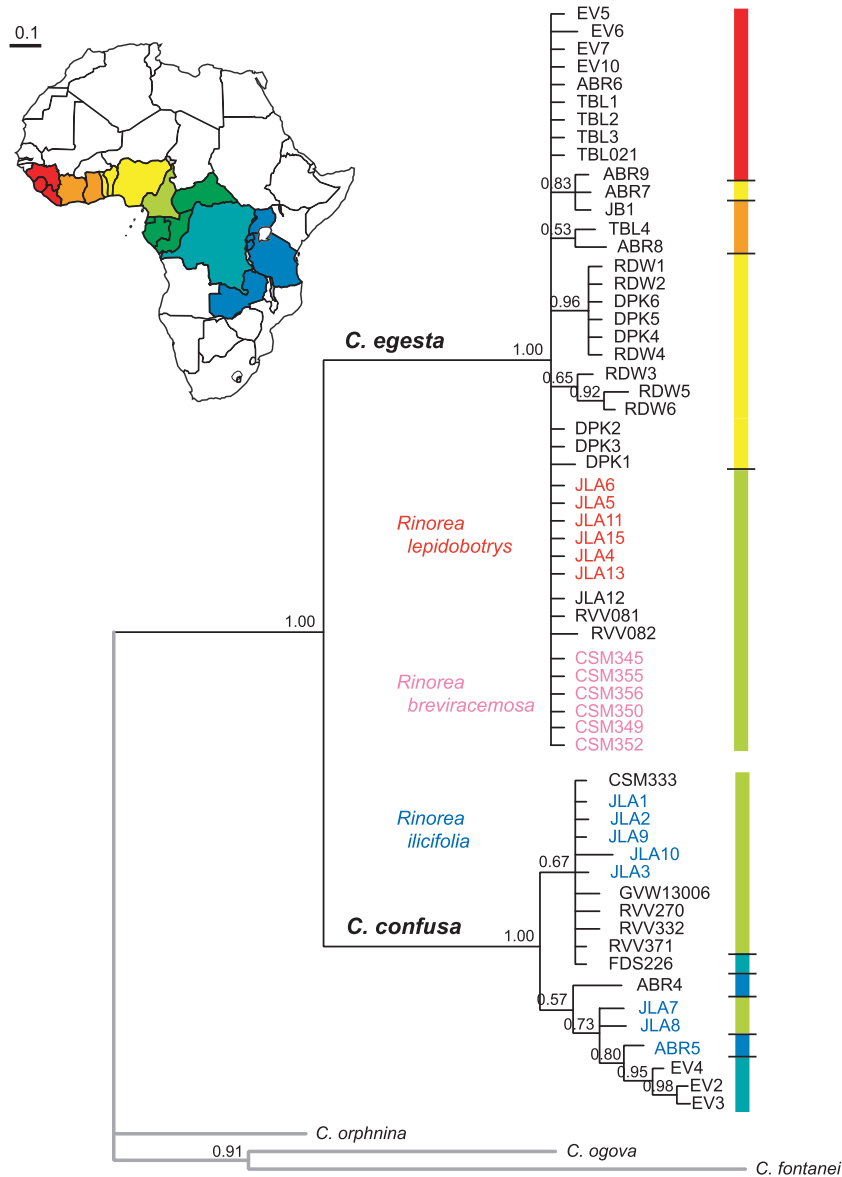


Fig. 2 Bayesian tree showing the phylogenetic relationships among *C. egesta sensu lato* specimens collected across the forested areas of sub-Saharan Africa. Coloured bars to the right of the tree indicate the region of origin of each specimen. Numbers at nodes are Bayesian posterior probabilities. Specimens printed in coloured lettering were collected as larvae on the following host plants: *R. lepidobotrys* (red), *R. breviracemosa* (pink), *R. ilicifolia* group species (blue). Only those specimens with full COI sequences are pictured. See Figure S1 (Supporting Information) for a tree based on shorter COI barcodes from the larger set of 86 specimens.

definitively separated these specimens into two sister clades with 100% posterior probabilities. Average pairwise nucleotide divergence between the clades was 2.39%, exceeding the average pairwise divergence within each clade (0.12% and 0.27%) by an order of magnitude. To confirm the correspondence of these two clades to the cryptic species described by Amiet (1997), we examined 13 individuals collected by Amiet himself within Cameroon. All six individuals found on *R. lepidobotrys* and referred to by Amiet as *C. 'egesta orange'* fell into one clade, while all seven individuals found on

R. ilicifolia group host species and referred to by Amiet as *C. egesta ssp. confusa* fell into the other clade (specimens are highlighted in Fig. 2 with red and blue lettering respectively). Although mtDNA is known to introgress across species boundaries more easily than nuclear DNA (Hudson & Turelli 2003; Ballard & Whitlock 2004) and is not necessarily a reliable tool for *de novo* species identification (Moritz & Cicero 2004), the fact that taxa described by other means (morphology and host plants) correspond perfectly to two strongly supported sister clades provides excellent

evidence that mtDNA can be used to separate species in this case. The NJ tree based on the larger 'barcode' data set (658 bp from 86 specimens) corroborated these results. The 86 sequences were split into two clusters, and two additional individuals collected on *R. ilicifolia* group host plants within Cameroon grouped with those collected on the same hosts by Amiet (Fig. S1, Supporting Information; see RVVE007 and RVVC025).

The two species are largely allopatric despite their co-occurrence within Cameroon

The two cryptic species within *C. egesta s. l.* were originally discovered in central Cameroon. To determine how they share/partition the rest of the pan-Afrotropical range previously attributed to *C. egesta*, we examined the position in the phylogeny of 32 specimens collected mostly as adults in seven additional countries across sub-Saharan Africa. All 26 specimens collected in countries west of Cameroon grouped with the individuals found on *R. lepidobotrys*, while all six specimens collected in countries east of Cameroon grouped with the individuals found on *R. ilicifolia* group hosts (Fig. 2). The additional specimens from the 'barcode' data set also assorted perfectly among the two clades/clusters according to where they were collected – 10 additional individuals from West African countries on the one hand and 18 additional individuals from Central/East African countries on the other (Fig. S1, Supporting Information). The two species thus appear to have largely allopatric distributions with a narrow zone of overlap (Fig. 1).

Species names, taxonomy and host plants

Information on the geographic ranges of the two cryptic species allowed us to match them with previously described type specimens and assign each a name. The type specimen of *Cymothoe egesta* (originally *Papilio egesta*) was described by Cramer (1775) and came from 'Surinamen' (patria falsa). It is generally accepted that the true type locality was Sierra Leone, where the species is common and from where much of Cramer's material in 1775–1779 came. As Sierra Leone is well outside the area of overlap of the two cryptic species (Fig. 1), we can confidently retain this name for the western taxon, including specimens referred to by Amiet within Cameroon as *C. 'egesta orange'*. *Cymothoe confusa* was described by Aurivillius from Congo, Brazzaville in 1887 (Aurivillius 1887) and has subsequently been treated as an eastern subspecies of *C. egesta sensu lato*. We raise this name to its original status as a distinct species, *C. confusa* Aurivillius 1887 **stat. rev.** and assign it to the eastern taxon. In summary, *C. egesta sensu stricto* is

essentially a western butterfly found from Sierra Leone to Cameroon, while *C. confusa* is an eastern species found from Cameroon to Uganda, northeastern Tanzania and Zambia. An east–west cline in the extent of melanization in adult males transcends the species boundary, creating a situation where adults of the two taxa are readily distinguishable from each other in allopatry (Fig. S2). Adults from the area of overlap in Cameroon, however, are nearly identical (Fig. 3), and appear intermediate in melanization to those found further west or further east. Immature stages and host plants, of course, are clearly and consistently different within the zone of overlap (Amiet 1997). Appendix S1 (Supporting Information) provides a more detailed taxonomic review and description of adult morphology.

While searching for larvae on host plants in the field, we found six individuals in the vicinity of Campo, Cameroon (see small black arrow in the inset of Fig. 1) feeding on *R. breviracemosa*, a previously unrecognized host plant closely related to *R. lepidobotrys* (G. Achoung personal communication and unpublished molecular data). These larvae were morphologically intermediate to those of the two species described by Amiet (1997), with reddish-mauve sides resembling *R. ilicifolia*-feeding *C. confusa* (Fig. 4, arrow 1), but a bright white latero-dorsal line separating the sides from the back (arrow 2), a yellowish fringe separating the sides from the belly (arrow 3), and a less extensive black mask (arrow 4) all resembling *R. lepidobotrys*-feeding *C. egesta s. s.* In contrast to the larvae, the pupae and mtDNA sequences were not intermediate; they exactly matched those associated with *C. egesta s. s.* (for description of pupae see Amiet 1997; for mtDNA sequences see pink lettering in Fig. 2). It is possible that these caterpillars represent a third species, so recently diverged from that feeding on *R. lepidobotrys* that there has been little time for morphological divergence or the development of fixed mtDNA differences. The fact that they were collected over 100 miles from the nearest *R. lepidobotrys*-feeding specimens, however, leaves open the possibility that they are instead simply geographic variants. Until further evidence comes to light, we consider *C. egesta s. s.* to be a single species with somewhat variable larval coloration, feeding on two closely related host plants, *R. lepidobotrys* and *R. breviracemosa*, within Cameroon. We do not know what this species feeds on in the western portion of its range, but *R. lepidobotrys* and *R. breviracemosa* appear to have distributions that closely resemble that of the butterfly, from Sierra Leone in the west to Cameroon/Gabon in the east (Hawthorne & Jongkind 2006). It is therefore probable that *C. egesta s. s.* larvae feed on these hosts throughout.

Cymothoe confusa uses host plants in the *Rinorea ilicifolia* group within Cameroon. This was true both for the



Fig. 3 Adult males of *C. egesta* s. s. (columns 1–3) and *C. confusa* (columns 4–6) from the area of sympatry in Cameroon. Note that there is no consistent difference in the extent and intensity of black markings, and that the difference in background colour is slight – *C. egesta* being more orange than *C. confusa*. The background colour of western, allopatric populations of *C. egesta* is more similar to that of *C. confusa* than to the orange tone of the conspecific population shown here. Plate courtesy of Jean-Louis Amiet.

larvae described by Amiet in 1997 and for two individuals we collected in 2006. *C. confusa* has also been found on *R. ilicifolia* at the opposite end of its range in northwestern Tanzania (African Butterfly Research Institute, specimen ABR5 in blue lettering in Fig. 2). The morphology of this Tanzanian larva exactly matched that of Cameroonian specimens (Fig. 4, compare bottom two images). We therefore suspect that *C. confusa* has relatively stable larval morphology and uses *R. ilicifolia* group species as hosts throughout its Central/East African range. More host records and larval collections, particularly from the central portion of the range in Congo, are needed to confirm this suspicion. Note that while *R. ilicifolia* is reported from throughout the Afrotropics – including from West Africa where *C. confusa* does not occur, preliminary data suggest that West African populations of this plant ‘species’ are genetically distinct from Central African populations (unpublished data).

Genetic structure and demographic history within species

We used the ‘phylogenetic’ data set to examine intra-specific variation within both species, starting with *C. egesta* s. s. An AMOVA indicated that 44% of the variation within *C. egesta* s. s. can be accounted for by differences among the major biogeographic zones of West Africa, defined by such barriers as the Volta River/Dahomey Gap, and the Niger, Cross and Sanaga Rivers ($P < 0.0001$; zones described in Booth 1958; Larsen 2005 and illustrated in Fig. 1). This result may reflect the effect of these barriers on historical gene flow in the species or it may simply reflect isolation by distance. Two additional results support the latter alternative. First, the majority of variation fell within zones (66%, $P < 0.0001$). Second, there was no genetic signal specifically associated with the most prominent biogeographic break in West Africa – the Dahomey Gap ($P = 0.9$). This

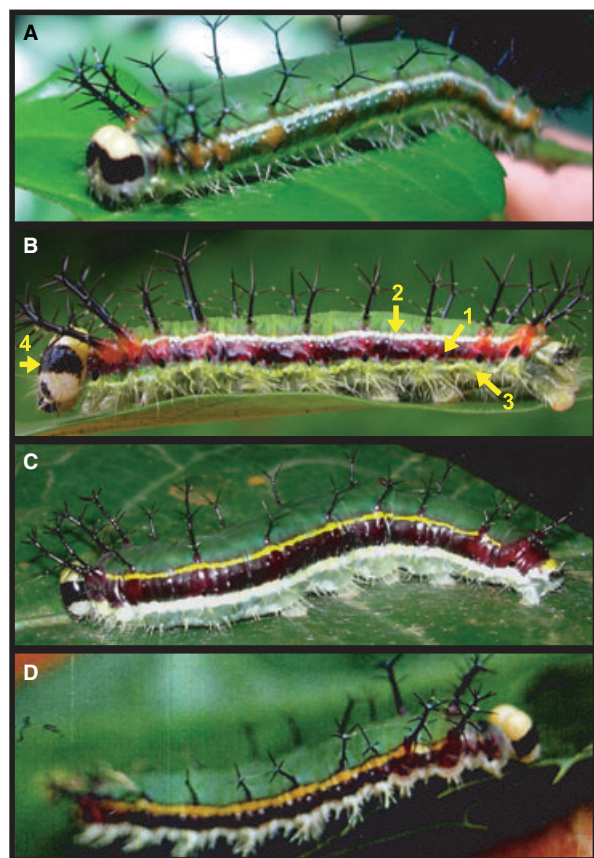


Fig. 4 Photographs illustrating variation in larval coloration within and between *C. egesta* s. s. and *C. confusa*. The first two photographs show *C. egesta* s. s. collected within Cameroon on *R. lepidobotrys* (A) and *R. breviracemosa* (B). The second two photographs show *C. confusa* collected on *R. ilicifolia* in Cameroon (C) and Tanzania (D). The numbered arrows in B point to morphological characters that make this larva intermediate to those pictured in A and C/D (see text for more detail). The photograph in D was taken by Colin Congdon of the African Butterfly Research Institute.

gap is a strip of open Savannah habitat in eastern Ghana, Togo and Benin (Fig. 1) that has separated the forests of West Africa from those of Nigeria and Camer-

oon repeatedly during arid periods over the course of the past million years – the most recent separation beginning approximately 8000 years ago (Maley 1996). The Gap has facilitated recent speciation among many forest animals, including some butterflies (see Larsen 2005). The fact that we found no lasting mark of the Dahomey Gap on population genetic variation within the butterfly suggests that the species is highly mobile.

To gain insight into the geographic origin and demographic history of *C. egesta* s. s., we estimated two measures of sequence diversity (π and θ_w) for the same five zones mentioned above. Table 1 shows that genetic variation was low at the western end of the species range (zone 1), increased towards the centre (zones 2 and 3), and then decreased again at the eastern end of the range (zones 4 and 5; zones illustrated in Fig. 1). If we consider only those zones where at least eight individuals were sampled, it becomes clear that western Nigeria (zone 3) in the centre of the species range harbours more diversity than either the Liberian subregion to the west (zone 1) or the area of overlap with *C. confusa* in Cameroon to the east (zone 5). None of the zones gave Tajima's *D* values that were significantly different from zero, but the value for zone 1 trended negative (Table 1). Taken together, these data suggest that populations in the centre of the species range have been the most stable over evolutionary time and that populations currently found further to the west and east are the result of range/population expansion out of western Nigeria and possibly Ghana.

Limited sampling from politically unstable regions of Central and East Africa (particularly the Democratic Republic of Congo) prevented us from conducting detailed analyses of population structure within *C. confusa* and from identifying its geographic centre of diversity. Nevertheless, we did observe significant variation between the Cameroonian specimens and those collected further to the south and east in Congo, Tanzania and Zambia (AMOVA $P = 0.002$; 43% of variation) – again likely a reflection of isolation by distance. We also

Table 1 Nucleotide diversity and neutrality estimates at mtDNA COI for *C. egesta* and *C. confusa* within biogeographic zones. *S* is the number of segregating sites. Tajima's *D* was only estimated for zones where >5 individuals were sampled

Species	Zone	Sub-region	<i>N</i>	<i>S</i>	Nucleotide diversity $\times 10^3$		Tajima's <i>D</i>	<i>P</i>
					$\pi \pm SE$	$\theta_w \pm SE$		
<i>C. egesta</i>	1	Liberian	10	2	0.31 \pm 0.11	0.54 \pm 0.13	-1.40	0.08
	2	Ghana	3	3	1.56 \pm 0.85	1.56 \pm 0.68		
	3	Western Nigeria	10	5	1.98 \pm 0.43	1.67 \pm 0.30	0.73	0.8
	4	Eastern Nigeria/Cross River Loop	3	1	0.50 \pm 0.36	0.50 \pm 0.29		
	5	Cameroon	8	2	0.53 \pm 0.18	0.60 \pm 0.16	-0.45	0.3
<i>C. confusa</i>		Cameroon	12	9	1.69 \pm 0.34	2.63 \pm 0.37	-1.46	0.07
		Congo/Tanzania/Zambia	6	6	2.78 \pm 0.81	2.73 \pm 0.66	0.08	0.6

observed a nearly significant negative Tajima's *D* within Cameroon (-1.46 , $P = 0.07$, Table 1), again suggestive of recent range/population expansion there.

The two methods we used to estimate the time of species divergence gave slightly different results, but both placed the split within the Pleistocene. The coalescent-based approach implemented in MDIV gave a 90% credible interval of 0.1–0.6 Ma (high mutation rate) or 0.2–0.9 Ma (low mutation rate) (Fig. S3, Supporting Information). Net divergence between the two species was 2.2%, corresponding to a divergence time of 0.95 Ma (high mutation rate) or 1.41 Ma (low mutation rate).

Discussion

Many cryptic species of specialized insect herbivores and parasitoids have been discovered in the tropics during the past 10 years (e.g. Amiet 1997; Morehead *et al.* 2001; Berkov 2002; Burns & Janzen 2005; Smith *et al.* 2006, 2007; Burns *et al.* 2008). Collecting and studying insects across large tropical areas are logistically difficult, however, and despite the impressive and coordinated efforts behind many of the recent discoveries (e.g. the biodiversity inventory in Area de Conservación Guanacaste, Costa Rica, <http://janzen.sas.upenn.edu>), the vast majority of new species are described from just a single location or region (but see Condon *et al.* 2008 for broad geographic study of flies that specialize on different plant parts rather than different plant species). We are left with little knowledge of the geographic distributions of these new species or of the extent of intraspecific variation within them.

This knowledge gap has several important consequences. For one, it complicates basic taxonomy. The morphological similarity of cryptic species makes it difficult for taxonomists to match new species with existing type specimens and thus to determine which should retain the original specific name. Amiet (1997) refrained from naming the cryptic species of *C. egesta* for this reason. The type specimen of *C. egesta* came from Sierra Leone, and he had no way to tell which, if any, of the Cameroonian species it represented. Our investigation of geographic distributions allowed us to resolve this dilemma since we found that only one of the species occurs in Sierra Leone. Information on geographic distributions will clearly be less useful in cases where multiple cryptic species occur in the type locality.

Ignorance of the geographic ranges of tropical cryptic species also makes it difficult for conservationists to determine how best to conserve them. Is one or more of the new species a narrowly distributed endemic that must be conserved through preservation of habitat in

one specific region? Or are the taxa broadly sympatric such that preservation of any number of regions within the purported range of the original taxon will help to conserve all of them? These questions remain unanswered for the majority of cryptic species so far discovered in the tropics.

Finally, our ignorance of phylogeographic variation across the ranges of these cryptic species makes it difficult for us to answer some of the most interesting questions posed by their discovery. For example, there is widespread agreement that host shifts, and subsequent host adaptation, can contribute to speciation in herbivorous and parasitic insects (Funk *et al.* 2002). The geographic context under which this is likely to take place, however, has been hotly debated (see Via 2001). And although many authors have noted that cryptic taxa found feeding on distinct host plants in the same location make good candidates for sympatric speciation (e.g. Morehead *et al.* 2001; Al-Barrak *et al.* 2004; Blair *et al.* 2005; Xue & Yang 2007; Barat *et al.* 2008), few data are available to test these hypotheses.

Here we examined the geographic distribution and mode of speciation of two cryptic Afrotropical butterfly species. We show that although *C. egesta* and *C. confusa* were first discovered through their sympatry, they are largely allopatric today, and are likely to have been completely allopatric during speciation.

Current geographic distributions and incongruence between old and new species

All 36 specimens we examined from countries west of Cameroon clustered genetically with one cryptic species (*C. egesta* s. s.) and all 24 specimens from countries south/east of Cameroon clustered genetically with the other (*C. confusa*). The probability of obtaining such perfect geographic assortment of 60 samples by chance alone (assuming the two species are equally likely to be collected in any given place) is less than 10^{-17} . We are thus confident that the two species are largely allopatric.

This finding may seem unsurprising since until approximately 100 years ago the taxon was classified as two distinct, allopatric species by the same names as those we apply here. But the 'new' species do not correspond exactly to those previously recognized on the basis of adult morphology. The previous species were differentiated by the extent of black markings on the wings of males, with western males being lighter than eastern males. This trait varies clineally, however, and the cline transcends the 'new' species boundary (Fig. S2 and Appendix S1, Supporting Information). The result is that the 'new' species have indistinguishable black

markings in the area of overlap – both resembling what was considered *C. confusa* (until early 1900s) or *C. egesta confusa* (early 1900s to present). Moreover, the subtle difference in wing ground colour that does differentiate adult males of the ‘new’ species where they occur in sympatry (Fig. 3), does not do so in allopatry and was not recognized as important in the previous classification (Appendix S1, Supporting Information). Performing this study, we therefore expected that one of the cryptic species would remain *C. egesta confusa*, the eastern subspecies of a broadly distributed Afrotropical butterfly, while the other would be a novel taxon with slightly orange wings and a narrower, but fully sympatric, range. Instead, we found two mostly allopatric species exhibiting an unusual pattern where adults are morphologically distinct outside the area of overlap, but nearly indistinguishable within. Explained from a different perspective, we found two mostly allopatric species whose clear differences in host plants, immature morphology and mtDNA were incongruent with previously described variation in adult morphology.

Several phenomena may explain this incongruence. First, male-mediated hybridization within the area of overlap may have resulted in gene flow at nuclear, but not mitochondrial loci. It is not clear, however, why such hybridization would obscure the differences in adult morphology without also affecting those in immature morphology and host plant choice. The latter traits may be sex-linked like the mtDNA (note that species differences in Lepidoptera are often sex-linked; Prowell 1998), but sex-linkage should create mismatches between phenotypic traits and mtDNA rather than preserve correlations among them. This is because females are the heterogametic sex in Lepidoptera, carrying a single Z-chromosome inherited from their fathers. The daughters of between species crosses would thus carry the mtDNA of one species and the Z-chromosome (and sex-linked phenotypes) of the other. A second explanation for adult morphological similarity within, but not outside, the area of overlap is convergent selection – either for Mullerian mimicry or for melanization patterns that are adaptive in the climatic conditions of west/central Cameroon. We know of no *Cymothoe* species in which males are mimics, but several other butterflies do become increasingly melanic from Cameroon eastwards (e.g. *Acraea pseudogina*), perhaps due to increasingly cool/humid environments.

More work addressing these issues, including analyses of variable nuclear markers, would clearly be interesting. Unfortunately, this will require fresh collections. We were only able to achieve the broad sampling we present here by assembling specimens of variable age and state of preservation. The majority of our material (particularly that with known host plants and larval

morphology) proved too old and poorly preserved for nuclear DNA sequencing.

Historical geographic distributions

To infer the extent of historical range overlap between *C. egesta* and *C. confusa*, we examined patterns of population genetic variation within species. The centre of genetic diversity for *C. egesta* was in the middle of its range in western Nigeria, well to the west of the range limit of *C. confusa*. Although we could not identify the centre of diversity for *C. confusa*, the frequency spectrum of polymorphic sites segregating among *C. confusa* individuals within Cameroon (summarized by Tajima's *D*) suggests that the species has not been stable in this area of sympatry. Instead, it likely invaded or increased in population size in recent evolutionary time. For example, the estimated Pleistocene species divergence time is consistent with the idea that *C. confusa* evolved in the stable forests of eastern Congo during arid periods when the remainder of its contemporary range in central/western Congo and adjacent eastern Cameroon were largely savannah (see discussion of African forest refugia in Douglass & Miller 2003). It would have then only colonized Cameroon relatively recently. In support of this hypothesis, other pairs of forest butterfly taxa share similar distributions. For example, *Papilio menestheus* and *P. lormieri* are also largely allopatric, but overlap in the narrow zone of Cameroon east of the Sanaga River. Whatever the exact history may be, the pattern of intraspecific genetic variation indicates that current sympatry between *C. egesta* and *C. confusa* is likely the result of secondary contact between taxa that originally diverged in allopatry.

Given the current distributions, this result should come as no surprise. Even insects with more sympatric, or fully sympatric, contemporary ranges are sometimes inferred to have originated in allopatry or parapatry. Diegisser *et al.* (2006) used phylogeographic data to infer that broadly sympatric host races of the fruit fly *Tephritis conura* diverged in allo- or peripatry before the postglacial range expansion of one of the two host plants to current areas of sympatry. Jordal *et al.* (2006) inferred that two completely sympatric taxa from the island of La Palma, which appear to be each other's closest relatives based on combined analysis of mitochondrial and nuclear loci, are likely the result of independent colonization events. Although host-specialized insects do provide some of the most plausible candidates for sympatric speciation, it is important to emphasize that the very same characteristics that make them good candidates for speciation via host shift in sympatry should also facilitate the secondary co-existence of

young taxa that originated in allopatry. Discrete alternative resources should help to prevent competitive exclusion, and mating on or near host plants should help to prevent hybridization and species fusion.

Implications for understanding the origins of other cryptic species in the tropics

It is difficult to extrapolate our findings to the many other sets of cryptic taxa that continue to be discovered in tropical areas. It may be that many of those discovered in known biogeographical contact zones such as West/Central Cameroon or Central America will also end up being largely allopatric and having allopatric origins. But we do not expect this to be true for all taxa, and each case will surely prove unique in some way. True understanding of broad patterns can only come through the accumulation of many individual examples. Here we present one of the first such examples for cryptic tropical insects. We hope that future work will add additional examples and in so doing contribute to species description and taxonomy, promote the design of efficient conservation priorities, and help to answer questions about speciation.

Acknowledgements

We are grateful to Jean-Louis Amiet for advice, encouragement, and the photograph of his bred specimens from Cameroon, and to Gaston Achoundong for crucial help finding and identifying *Rinorea* host plants within Cameroon. We would also like to thank J.-L. Amiet, S. C. Collins of the African Butterfly Research Institute (ABRI), J. H. Boersma, J. Bossart, F. de Smet, D. P. Knoop, G. van de Weghe, E. Vingerhoedt, and R. D. Warren for generously contributing specimens for genetic analysis. W. Savage helped to develop a PCR protocol for degraded specimens, and Michael Turelli and J.-L. Amiet commented on earlier versions of this manuscript. CSM was supported by an NSF Predoctoral Fellowship, a UC Regents Dissertation-Year Fellowship, and NSF grant DEB 0815145 to Michael Turelli. The initial sequencing was supported by NIH grant RO161773 to Sergey Nuzhdin. RVV acknowledges the Alberta Mennega Foundation and the Hugo de Vries Foundation for supporting field work in Cameroon and the European Commission's Research Infrastructure Action for supporting research visits to the RMCA (Tervuren) and NHM (London) via the SYNTHESYS Project (BE-TAF-3810 and GB-TAF-4003). TBL is as always grateful for continued financial support to studies on African butterflies from the Carlsberg Foundation in Denmark. DNA barcode sequences were kindly generated by the Biodiversity Institute of Ontario with sequence analysis funded by grants to Paul Hebert from the Natural Sciences and Engineering Research Council of Canada (NSERC) and Genome Canada through the Ontario Genomics Institute. We thank the Cameroon Ministry of Environment and Forests for facilitating our research within Cameroon.

References

- Al-Barrak M, Loxdale HD, Brookes CP *et al.* (2004) Molecular evidence using enzyme and RAPD markers for sympatric evolution in British species of Tetramesa (Hymenoptera : Eurytomidae). *Biological Journal of the Linnean Society*, **83**, 509–525.
- Amiet J-L (1997) Spécialisation trophique et premiers états chez les Cymothoe: implications taxonomiques (Lepidoptera, Nymphalidae). *Bulletin de la Société entomologique de France*, **102**, 15–29.
- Aurivillius C (1887) Förteckning över en samling Coleoptera och Lepidoptera från Kongoflodens område, skänkt till Riksmuseet af Löjtnant M. Juhlin-Dannfelt. *Öfversigt af Kongliga Vetenskaps-Akademiens Förhandlingar Stockholm*, **5**, 305–314.
- Ballard JWO, Whitlock MC (2004) The incomplete natural history of mitochondria. *Molecular Ecology*, **13**, 729–744.
- Barat M, Tarayre M, Atlan A (2008) Genetic divergence and ecological specialisation of seed weevils (*Exapion* spp.) on gorses (*Ulex* spp.). *Ecological Entomology*, **33**, 328–336.
- Berkov A (2002) The impact of redefined species limits in Palame (Coleoptera : Cerambycidae : Lamiinae : Acanthocinini) on assessments of host, seasonal, and stratum specificity. *Biological Journal of the Linnean Society*, **76**, 195–209.
- Berlacher SH, Feder JL (2002) Sympatric speciation in phytophagous insects: Moving beyond controversy? *Annual Review of Entomology*, **47**, 773–815.
- Bickford D, Lohman DJ, Sodhi NS *et al.* (2007) Cryptic species as a window on diversity and conservation. *TREE*, **22**, 148–155.
- Blair CP, Abrahamson WG, Jackman JA, Tyrrell L (2005) Cryptic speciation and host-race formation in a purportedly generalist tumbling flower beetle. *Evolution*, **59**, 304–316.
- Booth AH (1958) The Niger, the Volta, and the Dahomey Gap as geographic barriers. *Evolution*, **12**, 48–62.
- Brower AVZ (1994) Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial-DNA evolution. *PNAS*, **91**, 6491–6495.
- Brower AVZ (2006) Problems with DNA barcodes for species delimitation: 'ten species' of *Astraptus fulgurator* reassessed (Lepidoptera: Hesperidae). *Systematics and Biodiversity*, **4**, 127–132.
- Burns JM, Janzen DH (2005) Pan-neotropical genus *Venada* (Hesperidae: Pyrginae) is not monotypic: four new species occur on one volcano in the Area de Conservacion Guanacaste, Costa Rica. *Journal of the Lepidopterists' Society*, **59**, 19–34.
- Burns JM, Janzen DH, Hajibabaei M, Hallwachs W, Hebert PDN (2008) DNA barcodes and cryptic species of skipper butterflies in the genus *Perichares* in Area de Conservacion Guanacaste, Costa Rica. *PNAS*, **105**, 6350–6355.
- Bush GL (1975) Sympatric speciation in phytophagous parasitic insects. In: *Evolutionary Strategies of Parasitic Insects and Mites* (ed. Price PW), pp. 187–207. Plenum, New York.
- Condon MA, Scheffer SJ, Lewis ML, Swensen SM (2008) Hidden neotropical diversity: Greater than the sum of its parts. *Science*, **320**, 928–931.
- Coyne JA, Orr HA (2004) *Speciation*. Sunauer & Associates, Sunderland, Massachusetts.

- Cramer P (1775) *De Uitlandsche Kapellen voorkomende in de drie Waereld-deelen Asia, Afrika en America*. Volume 1, S.J. Baalde & B. Wild, Amsteldam & Utrecht.
- Diegisser T, Seitz A, Johannesen J (2006) Phylogeographic patterns of host-race evolution in *Tephritis conura* (Diptera : Tephritidae). *Molecular Ecology*, **15**, 681–694.
- Diehl SR, Bush GL (1989) The role of habitat preference in adaptation and speciation. In: *Speciation and its Consequence* (eds Otte D, Endler JA), pp. 345–365, Sinauer & Associates, Sunderland, Massachusetts.
- Douglas JF, Miller LD (2003) Afrotropical skippers (Lepidoptera; Hesperioidea) and the emergence of the combined refugium theory. *Bulletin Allyn Museum*, **143**, 1–18.
- Dres M, Mallet J (2002) Host races in plant-feeding insects and their importance in sympatric speciation. *Philosophical Transactions of the Royal Society of London. Series B*, **357**, 471–492.
- Fitzpatrick BM, Turelli M (2006) The geography of mammalian speciation: Mixed signals from phylogenies and range maps. *Evolution*, **60**, 601–615.
- 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**, 294–299.
- Funk DJ (1998) Isolating a role for natural selection in speciation: Host adaptation and sexual isolation in *Neochlamisus bebbianae* leaf beetles. *Evolution*, **52**, 1744–1759.
- Funk DJ, Filchak KE, Feder JL (2002) Herbivorous insects: model systems for the comparative study of speciation ecology. *Genetica*, **116**, 251–267.
- Funk DJ, Nosil P, Etges WJ (2006) Ecological divergence exhibits consistently positive associations with reproductive isolation across disparate taxa. *PNAS*, **103**, 3209–3213.
- Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PDN (2006) DNA barcodes distinguish species of tropical Lepidoptera. *PNAS*, **103**, 968–971.
- Hawthorne WD, Jongkind CCH (2006) *Woody Plants of Western African Forests*. Royal Botanic Gardens, Kew, Richmond, Surrey.
- Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *PNAS*, **101**, 14812–14817.
- Hudson RR, Turelli M (2003) Stochasticity overrules the “three-times rule”: Genetic drift, genetic draft, and coalescence times for nuclear loci versus mitochondrial DNA. *Evolution*, **57**, 182–190.
- Huelsensbeck JP, Ronquist F (2001) MR BAYES: Bayesian inference of phylogeny. *Bioinformatics*, **17**, 754–755.
- Ivanova NV, deWaard JR, Hebert PDN (2006) An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes*, **6**, 998–1002.
- Jordal BH, Emerson BC, Hewitt GM (2006) Apparent ‘sympatric’ speciation in ecologically similar herbivorous beetles facilitated by multiple colonizations of an island. *Molecular Ecology*, **15**, 2935–2947.
- Kimura M (1980) A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *Journal of Molecular Evolution*, **16**, 111–120.
- Kondrashov AS, Mina MV (1986) Sympatric speciation – when is it possible. *Biological Journal of the Linnean Society*, **27**, 201–223.
- Larsen TB (2005) *The Butterflies of West Africa*. Apollo Books, Stenstrup, Denmark.
- Maley J (1996) The African rain forest – Main characteristics of changes in vegetation and climate from the Upper Cretaceous to the Quaternary. In: *Essays on the Ecology of the Guinea-Congo Rain Forest* (eds Alexander IJ, Swaine MD, Watling R). *Proceedings of the Royal Society of Edinburgh Series B* **104**, 31–73.
- Morehead SA, Seger J, Feener DH, Brown BV (2001) Evidence for a cryptic species complex in the ant parasitoid *Apocephalus paraponerae* (Diptera: Phoridae). *Evolutionary Ecology Research*, **3**, 273–284.
- Moritz C, Cicero C (2004) DNA barcoding: Promise and pitfalls. *PLoS Biology*, **2**, 1529–1531.
- Nei M (1987) *Molecular Evolutionary Genetics*. Columbia University Press, New York, New York.
- Nielsen R, Wakeley J (2001) Distinguishing migration from isolation: a Markov chain Monte Carlo approach. *Genetics*, **158**, 885–896.
- Posada D, Crandall KA (1998) MODELTEST: testing the model of DNA substitution. *Bioinformatics*, **14**, 817–818.
- Prowell DP (1998) Sex linkage and speciation in Lepidoptera. In: *Endless Forms: Species and Speciation* (eds Howard DJ, Berlocher SH), pp. 309–319, Oxford University Press, New York.
- Rice WR (1984) Disruptive selection on habitat preference and the evolution of reproductive isolation – a simulation study. *Evolution*, **38**, 1251–1260.
- Schneider S, Roessli D, Excoffier L (2000) ARLEQUIN 2.0. A Software for Population Genetic Data Analysis. Genetics and Biometry Laboratory, Department of Anthropology, University of Geneva, Switzerland.
- Simon C, Frati F, Beckenbach A *et al.* (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene-sequences and a compilation of conserved polymerase chain-reaction primers. *Annals of the Entomological Society of America*, **87**, 651–701.
- Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PDN (2006) DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera : Tachinidae). *PNAS*, **103**, 3657–3662.
- Smith MA, Wood DM, Janzen DH, Hallwachs W, Hebert PDN (2007) DNA barcodes affirm that 16 species of apparently generalist tropical parasitoid flies (Diptera, Tachinidae) are not all generalists. *PNAS*, **104**, 4967–4972.
- Staden R (1996) The Staden sequence analysis package. *Molecular Biotechnology*, **5**, 233–241.
- Stireman JO, Nason JD, Heard SB (2005) Host-associated genetic differentiation in phytophagous insects: General phenomenon of isolated exceptions? Evidence from a goldenrod-insect community. *Evolution*, **59**, 2573–2587.
- Tajima F (1983) Evolutionary relationship of DNA-sequences in finite populations. *Genetics*, **105**, 437–460.
- Tajima F (1989) Statistical-method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics*, **123**, 585–595.
- Tamura K, Dudley J, Nei M, Kumar S (2007) MEGA4: Molecular evolutionary genetics analysis (MEGA) software version 4.0. *Molecular Biology and Evolution*, **24**, 1596–1599.

- Via S (2001) Sympatric speciation in animals: the ugly duckling grows up. *TREE*, **16**, 381–390.
- Waterson GA (1975) On the number of segregating sites in genetical models without recombination. *Theoretical Population Biology*, **7**, 256–276.
- Xue HJ, Yang XK (2007) Host plant use in sympatric closely related flea beetles. *Environmental Entomology*, **36**, 468–474.
- Zakharov EV, Caterino MS, Sperling FAH (2004) Molecular phylogeny, historical biogeography, and divergence time estimates for swallowtail butterflies of the genus *Papilio* (Lepidoptera: Papilionidae). *Systematic Biology*, **53**, 193–215.
- Zimmermann M, Wahlberg N, Descimon H (2000) Phylogeny of *Euphydryas* checkerspot butterflies (Lepidoptera: Nymphalidae) based on mitochondrial DNA sequence data. *Annals of the Entomological Society of America*, **93**, 347–355.

The authors met and began this collaboration through a mutual interest in Afrotropical butterflies. Carolyn McBride uses a combination of field, molecular genetic, and computational approaches to study host adaptation in insects, particularly as it relates to speciation. She recently finished her PhD at the University of California Davis and is now a postdoc at The Rockefeller University in New York. Robin van Velzen is a PhD student at the Wageningen University, supervised by Freek Bakker, working on ecology, evolution and systematics of species of *Cymothoe* and their *Rimorea* host plants. He is especially interested in host plant associations and underlying mechanisms, as well as their influence on patterns of diversification. Torben Larsen is the author of the standard works on butterflies of Arabia, Kenya and West Africa. He is currently working on a monograph of the African Hesperidae as well as coordinating in-depth butterfly surveys of selected African forests for long-term monitoring.

Supporting information

Additional supporting information may be found in the online version of this article:

Appendix S1 Taxonomic summary of *C. egesta* and *C. confusa*.

Fig. S1 Neighbour-Joining tree of short *COI* barcodes from 85 specimens of *C. egesta sensu lato*. The specimens fall into two major clades corresponding to the cryptic species, *C. egesta s. s.* and *C. confusa*. Specimen labels include the country and localities of origin, with the country names coloured to illustrate their position in the map of Africa (top left). The two species are mostly allopatric, overlapping only in Cameroon. The

specimen ABR5 from northwest Tanzania was excluded due to poor sequence quality in the region of the barcode.

Fig. S2 Variation in the extent of melanization on wings of adult male *C. egesta* (A–C) and *C. confusa* (D–F). The specimens of each species are arranged from west to east – with the eastern most *C. egesta* (C) and the western most *C. confusa* (D) coming from the area of sympatry in Cameroon. The cross-species longitudinal cline in melanization can be seen by disregarding the species assignments and examining the specimens in order from west to east (A–F). A) Light *C. egesta* f. *degesta* from Vane, Ghana. B) Typical western *C. egesta* from Bobiri, Ghana. C) Darker and slightly orange *C. egesta* from Cameroon. D) Light *C. confusa* from Cameroon. E) Typical *C. confusa* from Shaba, Congo (DRC) that matches the holotype closely. F) Very dark *C. confusa* male from Tanganyika Province, eastern Congo (DRC). Background yellow colour tone of different specimens is not directly comparable due to differences in photographic methods and postprocessing. Males average 6.5 cm in wing-span. Females are not shown.

Fig. S3 Estimated divergence time between *C. egesta* and *C. confusa*. The curves represent the posterior probability densities of divergence time estimated using the program MDIV assuming two different values for the mutation rates (u). The set of curves on the left and right represent the same four independent runs of the program with u assumed to be either 1.15% (Brower 1994) or 0.78% (Zakharov *et al.* 2004) respectively.

Table S1 Collection information associated with the butterfly specimens examined this study

Table S2 Primers used for amplifying and sequencing the mtDNA *COI* gene from low and high quality (poorly and well preserved) *C. egesta* specimens

Table S3 PCR recipes used in combination with the *C. egesta* specific primers for specimens yielding high and low quality DNA

Table S4 PCR cycling protocols used in combination with the *C. egesta* specific primers. The standard protocol was used for high quality specimens, and the ‘touchdown’ protocol was used specifically for older specimens that yielded low quality DNA

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.