

DNA barcoding reveals hidden species diversity in *Cymothoe* (Nymphalidae)

Robin van Velzen^{1,2}, Freek T. Bakker¹ & Joop J.A. van Loon²

¹Nationaal Herbarium Nederland, Wageningen University branch, and Biosystematics Group, Wageningen University, Generaal Foulkesweg 37, 6703 BL Wageningen, The Netherlands, E-mail: robin.vanvelzen@wur.nl; ²Laboratory of Entomology, Wageningen University, Binnenhaven 7, 6709 PD Wageningen, The Netherlands

DNA barcoding has recently become a prominent tool for species identification and discovery. In this paper we assess whether DNA barcoding can be used to identify and match different life stages of *Cymothoe* butterflies. Our results showed that DNA barcode sequences cluster according to species and that interspecific divergence is higher than intraspecific variation. In three cases, high levels of intraspecific DNA sequence variation revealed sibling species in *Cymothoe*, which are supported by morphology and host-plant data. We conclude that DNA barcoding is a powerful tool for the identification of eggs and caterpillars. In addition, combined with morphology, ecology and biogeography, DNA barcoding can be valuable in revealing hidden species.

Keywords: *Cymothoe*, DNA barcoding, herbivory, host-plants, tropical Africa

Recently, the use of standardised short DNA sequences as a tool for species identification, *i.e.* DNA barcoding (Hebert *et al.* 2004, Hebert & Gregory 2005) is receiving increasing attention in the scientific world (Matz & Nielsen 2005; www.barcoding.si.edu; www.barcodeoflife.org; www.ecbol.org). The main scientific benefits of the DNA barcoding approach lie in the fast and digital species identification of any life stage or fragment of a specimen, and the facilitation of species discovery (Janzen *et al.* 2005, Savolainen *et al.* 2005). Its accuracy depends mainly on the separation between within-species (intraspecific) variation and interspecific divergence of the selected DNA sequence (Matz & Nielsen 2005, Meyer & Paulay 2005). In animals the mitochondrial DNA region *cox1* has successfully been applied as a DNA barcode in vertebrates (*e.g.* birds, fishes), insects and nematodes (see www.barcoding.si.edu). Using DNA barcodes, Hajibabaei and co-workers (2006) could discriminate Lepidoptera species in Costa Rica and were also able to associate sexes and reinforce the identification of butterfly species that are difficult to distinguish morphologically (Janzen *et al.*

2005, Hajibabaei *et al.* 2006). In addition, DNA sequence variation revealed a substantial number of morphologically similar species (sibling species; Mayr 1942), differing mainly in their choice of food plants and caterpillar morphology (Hebert *et al.* 2004, Janzen *et al.* 2005).

Our subject of study is the evolution of host plant use in the Afrotropical genus *Cymothoe* (Nymphalidae) and its monospecific sistergenus *Harma* (Nymphalidae) comprising species feeding on Achariaceae and Violaceae (species of *Rinorea*). Amiet & Achoundong (1996) were the first to note the high specificity of the relationships between red and yellow species of *Cymothoe* and their *Rinorea* host plants. They discovered that 18 *Cymothoe* species were strictly monophagous (each feeding on another *Rinorea* host species), and another six colonising only two or three (closely related) *Rinorea* species. Such high level of host specificity, together with the number of butterfly and plant species involved, makes this system an excellent subject for a species-level phylogenetic approach to unravelling the evolution of host plant associations.

An assessment of host plant use depends on reliable identifications of herbivores and plants (Novotny & Basset 2005). For *Cymothoe*, a recent taxonomic treatment is lacking and although some species are easy to recognise, there are a number of problems for others, mainly related to the identification of eggs and caterpillars. Although the final caterpillar stages of many species have been described by Amiet (Amiet & Achoundong 1996, Amiet 1997, 2000), early stages have not as they lack discriminative characters. Because the discovery of eggs or caterpillars on plants is one of the main observations underlying host plant records, a fast alternative to the time consuming practice of rearing eggs or larvae to adulthood for identification would offer an important time gain (Moritz & Cicero 2004, Janzen *et al.* 2005).

Here, we present the results of a pilot study aiming at evaluating the possibilities for DNA barcoding as a tool for the identification of *Cymothoe* and *Harma* specimens. We assess the intra- and interspecific level of divergence of *cox1* sequences, as well as the performance of DNA barcoding in associating eggs or caterpillars with identified adults.

MATERIALS AND METHODS

In April and May 2006, van Velzen has conducted an expedition to Cameroon with the goal to study *Cymothoe* butterflies and their associated host plants in the field. Three regions were visited (Fig. 1): Yaoundé (Central Province), Edéa (Littoral Province) and Kribi (South Province). Specimens of *Cymothoe* and *Harma* were collected using hand nets and fruit traps, pinch-killed and dried on silica gel for conservation.

For the DNA barcoding trial we have sampled 47 collections of *Cymothoe* and *Harma* butterflies, comprising 9 and 1 species respectively (taxon list with collection data is available upon request). In addition, 19 unidentified eggs and

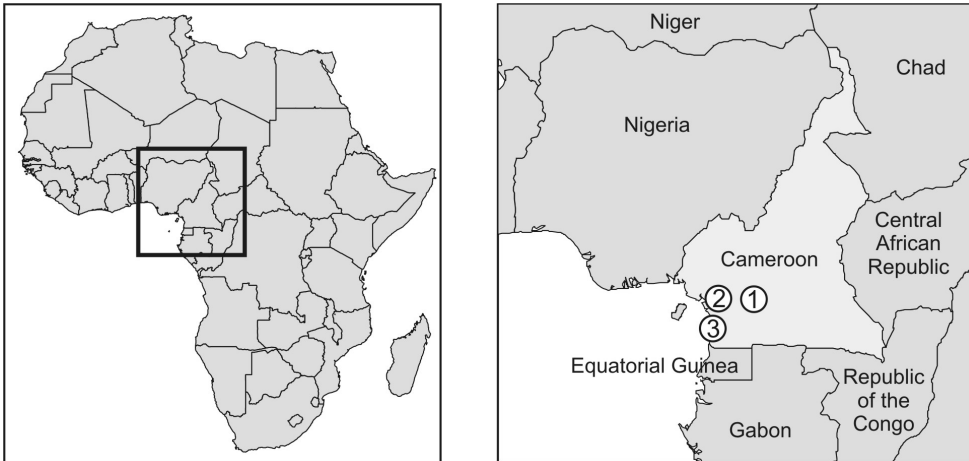


Figure 1. Map of the three visited regions in Cameroon: 1. Yaoundé (Central Province); 2. Edéa (Littoral Province); 3. Kribi (South Province).

caterpillars were sampled to assess the applicability of DNA barcoding for their identification. All sequences were processed by the Canadian Centre for DNA Barcoding (CCDB, Guelph) where DNA was extracted from single legs, eggs or caterpillar prolegs using an automated Glass Fiber plate extraction method on a Beckman-Coulter Biomek NX Span-8 liquid handling system (Ivanova *et al.* 2006). The target 658-bp fragment of *cox1* was amplified using primer pair LepF (5'-ATTCAACCAATCATAAAGATATTGG-3') and LepR (5'-TAAACTTCTGGATGTCCAAAAAATCA-3'). In the cases where these primers did not produce a PCR product, two internal mini-primers were added, using primer pairs LepF with mLr (5'-CCTGTTCCAGCTCCATTTTC-3') and mLr (5'-GCTTCCCACGAATAAATAATA-3') with LepR. All generated sequences, together with photographs and collection details, are deposited at the Barcoding of Life Database (BoLD; www.boldsystems.org) under the project code CYMO.

DNA barcoding analyses were done through the online interface of the BoLD website. The taxon identification tree was based on the Kimura 2-parameter distance model (Kimura 1980), with the filter set to sequences with length >100 basepairs, and all codon positions included. The resulting tree was transformed into a radial tree in TreeIllustrator 0.52 beta (Trooskens *et al.* 2005). A sequence distance summary was provided through the online interface of BoLD, with only sequences longer than 425 base pairs included in order to ensure reliable sequence distance estimates (Hanner *et al.* in press).

RESULTS

The radial taxon identification tree based on DNA barcodes shows that all sequences clustered nicely according to the species (Fig. 2). It appears that 9

A number of possible flaws in this study need to be discussed, however. First, the number of samples was very low, both in number of species and in the number of sampled individuals per species. An assessment on the reliability of DNA barcoding for species identification depends greatly on the number of samples (Moritz & Cicero 2004, Matz & Nielsen 2005), and there need to be at least 10 samples per species (Janzen *et al.* 2005). In this study the only adequately sampled species is *Harma theobene*, for the other species the number of samples ranged from only 1 to 9. Moreover, ideally all species within a genus should be included to ensure that the between-species distance always exceeds the intraspecific variation (Moritz & Cicero 2004). Finally, taxa should be included from more than one geographic region in order to adequately reveal the extent of intraspecific variation (Moritz & Cicero 2004, Janzen *et al.* 2005). Although we included samples from three different regions in Cameroon, this geographical range is limited compared with the distribution of some of the widespread species occurring throughout tropical Africa.

Nevertheless, we feel that there are no serious problems to be expected, for several reasons. First, the *cox1* sequences can discriminate between hidden sibling species that had not been discovered until recently, due to their high morphological similarities (see below). Therefore, it is likely that other, less closely related species can also be discriminated based on DNA barcodes (Moritz & Cicero 2004). Second, an analysis including sequences of *C. egesta* from West and East Africa shows that these sequences group nicely within the variation already found in the Cameroonian specimens (McBride, pers. comm.), indicating that intraspecific divergence is low even at a wide geographical scale.

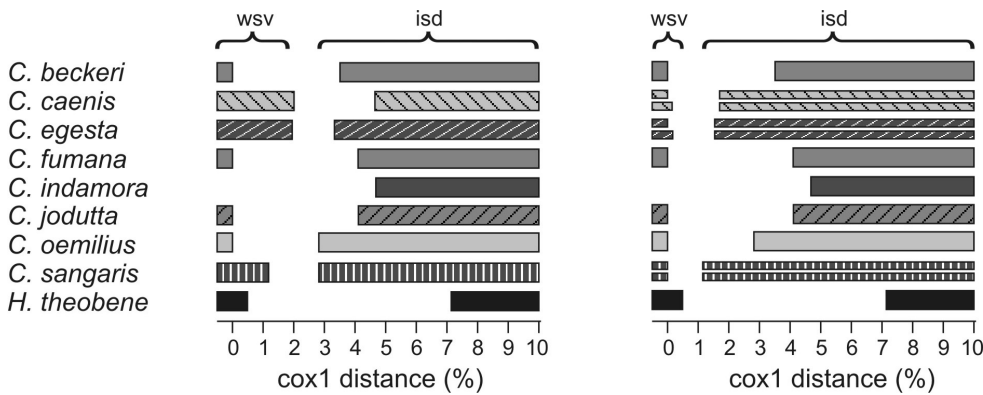


Figure 3. Summaries of within-species variation (wsv) and interspecific divergence (isd) of mtDNA *cox1* sequences >425 base pairs long, for species of *Cymothoe* and *Harma*. Left: without sibling species taken into account; Right: with sibling species taken into account.

Hidden species diversity in *Cymothoe*

As mentioned above, intraspecific DNA sequence variation indicates that at least three species may contain two or more sibling species: *C. egesta*, *C. sangaris* and *C. caenis*. This is concordant with the findings of Amiet (1997) who claims that *C. egesta* and *C. sangaris* each consist of several hidden sibling species which differ in subtle morphological differences and host plant use.

The DNA barcode sequence diversities within *C. egesta* samples show high levels of divergence between two clusters (Table 1), that coincide with the two sibling species described by Amiet (1997); one feeding on the closely related *Rinorea lepidobotrys* and *R. breviracemosa*, the other colonising *R. ilicifolia* and its relatives. McBride (University of California, Davis, USA) found that these same two *cox1* sequence clusters are associated with *C. egesta* from West Africa and *C. confusa* from Central and East Africa (McBride, pers. comm.).

The complex around *C. sangaris* has posed a problem for taxonomists for a long time already (Larsen 2005). The males are well characterised by a blood-red ground colour and the lack of a white spot on the hind wing costa (Larsen 2005), but in contrast to the morphological uniformity of the males, the females exhibit much variation in wing patterns and colour (Amiet 1997, Larsen 2005). Based on larval and pupal characters in combination with female morphology, Amiet (1997) concluded that there are at least four distinct species hiding in the *C. sangaris* complex, each having their own *Rinorea* host plant(s). We have found two populations of *C. sangaris* in the Yaoundé area: one feeding on *R. preussi* and another feeding on *R. batesii*, which is in compliance with two of the four species mentioned by Amiet (1997). Although the localities are only 20 km apart, the sequence divergence is 1.15% (Table 2), which is unusually high for intraspecific variation (Hebert *et al.* 2004). Apparently, here the *cox1* sequence divergences also confirm previous hypotheses based on morphological characters and host plant data.

Of the species *C. caenis*, specimens have been found at three localities in the littoral plain of Cameroon. Although the first two localities are both in the same region (Kribi), we found a sequence divergence of 1.7% (Table 3). In contrast, the sequences coming from one of these localities had only 0.3% difference with the sequence coming from Edéa, more than 80 km north, indicating that the sequence divergences are not related to geographical distance. The Achariaceae-feeding *C. caenis* is less well documented by Amiet, but different forms and sub-species have been described in the past (Birket-Smith 1960, Fontaine 1982) and we are currently assessing whether the *cox1* sequence clusters correlate with morphological differences.

When taking these hidden species into account, the minimum distance between species decreases to 1.15%. The maximum within-species divergence also decreases to a mere 0.49% (*Harma theobene*; see Fig. 3). These values are similar to those found in Costa Rican Hesperiidæ, where the average within- and between-

Table 1. Sequence diversity within the *C. egesta* complex, based on Kimura 2 parameter distances of mtDNA *cox1* sequences >425 base pairs long.

Sample ID	Region	Locality	Distance				
			RVV B270	RVV B300	RVV B372	RVV B081	RVV B082
RVV B270	Edéa	Eding	-				
RVV B300	Edéa	Ducam-Duclair	0	-			
RVV B372	Kribi	Londji 2	0	0	-		
RVV B081	Yaoundé	Kala	1.54	1.73	1.54	-	
RVV B082	Yaoundé	Kala	1.95	1.91	1.86	0.17	-

Table 2. Sequence diversity within the *C. sangaris* complex, based on Kimura 2 parameter distances of mtDNA *cox1* sequences >425 base pairs long.

Sample ID	Region	Locality	Distance			
			RVV B346	RVV B354	RVV E011	RVV B199
RVV B346	Yaoundé	Elounden	-			
RVV B354	Yaoundé	Elounden	0	-		
RVV E011	Yaoundé	Elounden	0	0	-	
RVV B199	Yaoundé	Kala	1.15	1.15	1.18	-

Table 3. Sequence diversity within *C. caenis*, based on Kimura 2 parameter distances of mtDNA *cox1* sequences >425 base pairs long.

Sample ID	Region	Locality	Distance				
			RVV B249	RVV B384	RVV B385	RVV B364	RVV B383
RVV B249	Edéa	Malimba	-				
RVV B384	Kribi	Djabilobe	0.15	-			
RVV B385	Kribi	Djabilobe	0.3	0.15	-		
RVV B364	Kribi	Londji 2	2.01	1.86	1.7	-	
RVV B383	Kribi	Londji 2	2.01	1.86	1.7	0	-

species diversity was 0.17% and 4.58, respectively (Hajibabaei *et al.* 2006). Although DNA divergences differ between groups (DeSalle *et al.* 2005) and should not be a primary criterion for recognising species boundaries (Moritz & Cicero 2004, Will & Rubinoff 2004, Hebert & Gregory 2005, Hajibabaei *et al.* 2006), and the morphology and ecology of an organism should always be taken into account (Will & Rubinoff 2004, DeSalle 2006, Rubinoff 2006), it appears that in our particular case of *Cymothoe* a sequence divergence of 1% correlates with sibling species.

The three examples above show that analysis of morphology, biogeography and DNA sequence data, considered in concert, can be valuable in revealing hidden species.

Acknowledgements We thank Jean-Louis Amiet and Gaston Achoundong; without their previous work, this study would not have been possible. We are indebted to the people of the staff and laboratory of the Canadian Centre for DNA Barcoding in Guelph; particularly Paul Hebert, who generously sequenced our butterfly material. Olivier Sene Belinga has been an invaluable collaborator in the field. Torben Larsen and Michel Beaurain have been of great help in identifying butterfly specimens; Rien de Jong and Yde Jongema have provided much practical advice. This project has received financial support from the Hugo de Vries Foundation, Stichting der Korinthiërs, Wageningen University Travel Fund, Alberta Mennega Foundation and the Biosystematics Fund.

REFERENCES

- Amiet, J.-L. & Achoundong, G. 1996. An example of trophic specialization within the Lepidoptera: The Cameroonian *Cymothoe* feeding on *Rinorea* (Violaceae) (Lepidoptera, Nymphalidae). *Bulletin de la Societe Entomologique de France* 101: 449-466.
- Amiet, J.-L. 1997. Trophic specialization and early stages in *Cymothoe*: Taxonomic implications (Lepidoptera, Nymphalidae). *Bulletin de la Societe Entomologique de France* 102: 15-29.
- Amiet, J.-L. 2000. Les premiers etats des *Cymothoe*: Morphologie et interet phylogenetique (Lepidoptera, Nymphalidae). *Bulletin de la Societe Entomologique de France* 106: 349-390.
- Birket-Smith, J. 1960. Results from the Danish expedition to French Cameroons (1948-1950). XXVII – Lepidoptera (part II). *Bulletin de l'institut Français d'Afrique Noire* 22: 924-983.
- Coley, P.D. 1999. Paleoecology – Hungry herbivores seek a warmer world. *Science* 284: 2098-2099.
- DeSalle, R., Egan, M.G. & Siddall, M. 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. *Philos. Trans. R. Soc. B-Biol. Sci.* 360: 1905-1916.
- DeSalle, R. 2006. Species discovery versus species identification in DNA barcoding efforts: Response to Rubinoff. *Conserv. Biol.* 20: 1545-1547.
- Fontaine, M. 1982. Genre *Cymothoe* Hübner. Lep. Nymphalidae – S. fam. Nymphalinae. Note sur les premiers états. *Lambillionea* 82: 63-64, 67-72, 95-98.
- Gaston, K.J. 2000. Global patterns in biodiversity. *Nature* 405: 220-227.
- Gentry, A.H. 1992. Tropical forest biodiversity – Distributional patterns and their conservation significance. *Oikos* 63: 19-28.
- Hajibabaei, M., Janzen, D.H., Burns, J.M., Hallwachs, W. & Hebert, P.D.N. 2006. DNA barcodes distinguish species of tropical Lepidoptera. *Proc. Natl. Acad. Sci. USA* 103: 968-971.
- Hebert, P.D.N., Penton, E.H., Burns, J.M., Janzen, D.H. & Hallwachs, W. 2004. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptus fulgerator*. *Proc. Natl. Acad. Sci. USA* 101: 14812-14817.
- Hebert, P.D.N. & Gregory, T.R. 2005. The promise of DNA barcoding for taxonomy. *Syst. Biol.* 54.
- Ivanova, N.V., deWaard, J.R. & Hebert, P.D.N. 2006. An inexpensive, automation-friendly protocol for recovering high-quality DNA. *Molecular Ecology Notes* 6: 998-1002.

- Janzen, D.H., Hajibabaei, M., Burns, J.M., Hallwachs, W., Remigio, E. & Hebert, P.D.N. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. *Philos. Trans. R. Soc. B-Biol. Sci.* 360: 1835-1845.
- Kimura, M. 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide-sequences. *J. Mol. Evol.* 16: 111-120.
- Larsen, T.B. 2005. *Butterflies of West Africa*. Apollo Books, Stenstrup.
- Matz, M.V. & Nielsen, R. 2005. A likelihood ratio test for species membership based on DNA sequence data. *Philos. Trans. R. Soc. B-Biol. Sci.* 360: 1969-1974.
- Mayr, E. 1942. *Systematics and the Origin of Species from the viewpoint of a zoologist*. Columbia University Press, New York.
- Meyer, C.P. & Paulay, G. 2005. DNA barcoding: Error rates based on comprehensive sampling. *PLoS Biol.* 3: 2229-2238.
- Moritz, C. & Cicero, C. 2004. DNA barcoding: Promise and pitfalls. *PLoS Biol.* 2: 1529-1531.
- Myers, N., Mittermeier, R.A., Mittermeier, C.G., da Fonseca, G.A.B. & Kent, J. 2000. Biodiversity hotspots for conservation priorities. *Nature* 403: 853-858.
- Novotny, V. & Basset, Y. 2005. Review - Host specificity of insect herbivores in tropical forests. *Proc. R. Soc. Lond. Ser. B-Biol. Sci.* 272: 1083-1090.
- Rubinoff, D. 2006. DNA barcoding evolves into the familiar. *Conserv. Biol.* 20: 1548-1549.
- Savolainen, V., Cowan, R.S., Vogler, A.P., Roderick, G.K. & Lane, R. 2005. Towards writing the encyclopaedia of life: an introduction to DNA barcoding. *Philos. Trans. R. Soc. B-Biol. Sci.* 360: 1805-1811.
- Trooskens, G., De Beule, D., Decouttere, F. & Van Criekinge, W. 2005. Phylogenetic trees: visualizing, customizing and detecting incongruence. *Bioinformatics* 21: 3801-3802.
- Will, K.W. & Rubinoff, D. 2004. Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification. *Cladistics* 20: 47-55.