

Research

Open Access

Does the DNA barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae)

Martin Wiemers* and Konrad Fiedler

Address: Department of Population Ecology, Faculty of Life Sciences, University of Vienna, Althanstrasse 14, 1090 Vienna, Austria

Email: Martin Wiemers* - martin.wiemers@univie.ac.at; Konrad Fiedler - konrad.fiedler@univie.ac.at

* Corresponding author

Published: 7 March 2007

Received: 1 December 2006

Frontiers in Zoology 2007, 4:8 doi:10.1186/1742-9994-4-8

Accepted: 7 March 2007

This article is available from: <http://www.frontiersinzoology.com/content/4/1/8>

© 2007 Wiemers and Fiedler; licensee BioMed Central Ltd.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/2.0>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

Background: DNA barcoding, i.e. the use of a 648 bp section of the mitochondrial gene cytochrome *c* oxidase I, has recently been promoted as useful for the rapid identification and discovery of species. Its success is dependent either on the strength of the claim that interspecific variation exceeds intraspecific variation by one order of magnitude, thus establishing a "barcoding gap", or on the reciprocal monophyly of species.

Results: We present an analysis of intra- and interspecific variation in the butterfly family Lycaenidae which includes a well-sampled clade (genus *Agrodiaetus*) with a peculiar characteristic: most of its members are karyologically differentiated from each other which facilitates the recognition of species as reproductively isolated units even in allopatric populations. The analysis shows that there is an 18% overlap in the range of intra- and interspecific *COI* sequence divergence due to low interspecific divergence between many closely related species. In a Neighbour-Joining tree profile approach which does not depend on a barcoding gap, but on comprehensive sampling of taxa and the reciprocal monophyly of species, at least 16% of specimens with conspecific sequences in the profile were misidentified. This is due to paraphyly or polyphyly of conspecific DNA sequences probably caused by incomplete lineage sorting.

Conclusion: Our results indicate that the "barcoding gap" is an artifact of insufficient sampling across taxa. Although DNA barcodes can help to identify and distinguish species, we advocate using them in combination with other data, since otherwise there would be a high probability that sequences are misidentified. Although high differences in DNA sequences can help to identify cryptic species, a high percentage of well-differentiated species has similar or even identical *COI* sequences and would be overlooked in an isolated DNA barcoding approach.

Background

Molecular tools have provided a plethora of new opportunities to study questions in evolutionary biology (e.g. speciation processes) and in phylogenetic systematics. Only recently, however, have claims been made that the sequencing of a small (648 bp) fragment at the 5' end of the gene cytochrome *c* oxidase subunit 1 (*COI* or *cox1*)

from the mitochondrial genome would be sufficient in most Metazoa to identify them to the species level [1,2]. This approach called "DNA barcoding" has gained momentum and the "Consortium for the Bar Code of Life (CBOL)" founded in September 2004 intends to create a global biodiversity barcode database in order to facilitate automated species identifications. Right from the start,

however, this approach received opposition, especially from the taxonomists' community [3-8]. Some arguments in this debate are political in nature, others have a scientific basis. Concerning the latter, one of the most essential arguments focuses on the so-called "barcoding gap". Advocates of barcoding claim that interspecific genetic variation exceeds intraspecific variation to such an extent that a clear gap exists which enables the assignment of unidentified individuals to their species with a negligible error rate [1,9,10]. The errors are attributed to a small number of incipient species pairs with incomplete lineage sorting (e.g. [11]). As a consequence, establishing the degree of sequence divergence between two samples above a given threshold (proposed to be at least 10 times greater than within species [10]) would indicate specific distinctness, whereas divergence below such a threshold would indicate taxonomic identity among the samples. Furthermore, the existence of a barcoding gap would even enable the identification of previously undescribed species ([11-13] but see [14]). Possible errors of this approach include false positives and false negatives. False positives occur if populations within one species are genetically quite distinct, e.g. in distant populations with limited gene flow or in allopatric populations with interrupted gene flow. In the latter case it must be noted that, depending on the amount of morphological differentiation and the species concept to be applied, such populations may also qualify as 'cryptic species' in the view of some scientists. False negatives, in contrast, occur when little or no sequence variation in the barcoding fragment is found between different biospecies (= reproductively isolated population groups *sensu* Mayr [15]). Hence, false negatives are more critical for the barcoding approach, because the existence of such cases would reveal examples where the barcoding approach is less powerful than the use of other and more holistic approaches to delimit species boundaries.

Initial studies on birds [10] and arthropods [9,16] appeared to corroborate the existence of a distinct barcoding gap, but two recent studies on gastropods [17] and flies [18] challenge its existence. The reasons for these discrepancies are not entirely clear. Although levels of *COI* sequence divergence differ between higher taxa (e.g. an exceptionally low mean *COI* sequence divergence of only 1.0% was found in congeneric species pairs of Cnidaria compared to 9.6–15.7% in other animal phyla [2]), Mollusca (with 11.1% mean sequence divergence between species) and Diptera (9.3%) are not peculiar in this respect. Meyer & Paulay [17] assume that insufficient sampling on both the interspecific and intraspecific level create the artifact of a barcode gap. Proponents of barcoding might argue, however, that the main reason for this overlap is the poor taxonomy of these groups, e.g. cryptic species may have been overlooked which are differentiated

genetically but very similar or even identical in morphology.

If the barcode gap does not exist, then the threshold approach in barcoding becomes inapplicable. Although more sophisticated techniques (e.g. using coalescence theory and statistical population genetic methods [19-21]) can sometimes help to delimit species with overlapping genetic divergences, these approaches require additional assumptions (e.g. about the choice of population genetic models or clustering algorithms) and are only feasible in well-sampled clades.

Barcoding holds promise nonetheless especially in the identification of arthropods, the most species-rich animal phylum in terrestrial ecosystems. Identification of arthropods is often extremely time-consuming and generally requires taxonomic specialists for any given group. Moreover, the fraction of undescribed species is particularly high, as opposed to vertebrates. Hence, there is substantial demand for improved (and rapid) identification tools by scientists who seek identification of large arthropod samples from complex faunas. Therefore arthropods deserve to be considered the yard-stick for the usefulness of barcoding approaches among Metazoa and it is not surprising that several recent studies have tried to apply DNA barcoding in arthropods [9,11-13,16,18,19,22-27]. Diversity is concentrated in tropical ecosystems, but measuring intra- and interspecific sequence divergence in tropical insects is hampered by the fragmentary knowledge of most taxa. In contrast, insects of temperate zones, and most notably the butterflies of the Holarctic region, are well known taxonomically compared to other insects. The species-rich Palaearctic genus (or subgenus) *Agrodiaetus* provides an excellent example to test the existence of the barcode gap in arthropods. This genus is exceptional because of its extraordinary interspecific variation in chromosome numbers which have been investigated for most of its ca 120 species ([28-30] and references therein). As a result several cryptic species which hardly or not at all differ in phenotype have been discovered (e.g. [31-39]). Available evidence suggests that apart from a few exceptions (e.g. due to supernumerary chromosomes) differences in chromosome numbers between butterfly species are linked to infertility in interspecific hybrids [40]. This is due to problems in the pairing of homologous chromosomes during meiosis. Since major differences in chromosome numbers are indicative of clear species boundaries, they are helpful also to infer species-level differentiation for allopatric populations. *Agrodiaetus* butterflies therefore are an ideal case for testing the validity of the barcoding approach. If valid, then it must be possible to safely recognize all species that can be distinguished by phenotype, karyotype or both character sets with reference to sequence divergences alone. On the contrary, failure of

DNA barcodes to differentiate between species that are distinguished by clear independent evidence would undermine the superiority of the barcoding approach, which has especially been attributed to taxa with "difficult" classical taxonomy, such as *Agrodiaetus*.

Results

Intraspecific divergence

The average divergence in 1189 intraspecific comparisons is 1.02% (SE = 1.13%). 95% of intraspecific comparisons have divergences of 0–3.2%. The few values higher than 3.2% are conspicuous and probably due to misidentifications (*Lampides boeticus*, *Neozephyrus japonicus*, *Arhopala atosia*, *Agrodiaetus kendeveni*, see below), unrecognized cryptic species (*Agrodiaetus altivagans* [41], *Agrodiaetus demavendi* [30]), hybridization events (*Meleageria marcida* [30,42]) or any of those (*Agrodiaetus mithridates*, *Agrodiaetus merhaba*).

The evidence for the possible misidentifications is the following:

- *Lampides boeticus* is the most widespread species of Lycaenidae and a well-known migrant which occurs throughout the Old World tropics and subtropics from Africa and Eurasia to Australia and Hawaii. Apart from a single unpublished sequence ([AB192475](#)), all other COI GenBank sequences of this species (from Morocco, Spain and Turkey) are identical with each other or only differ in a single nucleotide (= 0.15% divergence). They are also nearly identical to two specimens of *Lampides boeticus* in the CBOL database (BOLD) [43] from Tanzania and another sequence of this species from Papua New Guinea (Wiemers, unpubl. data). The GenBank sequence [AB192475](#) (of unknown origin, but possibly from Japan), however, differs strongly (8.2–8.7%) from all other *Lampides boeticus* sequences and therefore we assume this to represent a distinct species. Its identity however remains a mystery because it is not particularly close to any other GenBank sequence and a request for a check of the voucher specimen has remained unanswered for more than a year.
- The questionable unpublished sequence of *Neozephyrus quercus* ([AB192476](#)) is identical to a sequence of *Favonius orientalis* and therefore probably represents this latter species which is very similar in phenotype but well differentiated genetically (4.8% divergence).
- A similar situation constitutes the questionable unpublished sequence of *Arhopala atosia* ([AY236002](#)) which is very similar (0.4%) to a sequence of *Arhopala epimuta*.
- *Agrodiaetus kendeveni* is a local endemic of the Elburs Mts. in Iran. The two sequences of this species in the NCBI

database which exhibit a divergence of 5.4% have been published in two different papers by the same work group [29,44]. While one of them is identical to a sequence of *Agrodiaetus pseudoxerxes*, the other one is nearly identical to *Agrodiaetus elbursicus* (0.2% divergence). These latter two species however belong to separate species groups [30] and thus conspecificity of the two sequences of *A. kendeveni* is very improbable as there is no evidence of hybridization between members of different species groups in *Agrodiaetus* [30].

Higher intraspecific divergence values are also found between North African and Eurasian populations of *Polyommatus amandus* (3.8%) and *Polyommatus icarus* (5.7–6.8%). In the former species the North African population is also well differentiated in phenotype (ssp. *abdelaziz*), while in the latter species phenotypic differences have never been noted. Cases with substantial, but lower genetic divergence between North African and European populations which do not correspond to differentiation in phenotype also occur in the butterflies *Iphiclides podalirius feisthamelii* (2.1%; [30]) and *Pararge aegeria* (1.9%; [45]). In all cases these allopatric populations may actually represent distinct species, although we do not currently have additional evidence in support of this hypothesis.

Although some of the other higher divergence values >2% are possibly due to cryptic species (e.g. in *Agrodiaetus demavendi*) or hybridization between closely related species (e.g. in the species pair *Lysandra corydonius* and *L. ossmar*, as evidenced by the comparative analysis of the nuclear rDNA internal transcribed spacer region ITS-2 [30]), most of those values represent cases in which there is hardly any doubt regarding the conspecificity of samples. The highest such value is 2.9% between distant populations of the widespread *Agrodiaetus damon* (from Spain and Russia). Outside the genus *Agrodiaetus* high values are also found between North African and Iranian populations of *Lycaena alciphron* (2.7%), Spanish and Anatolian populations of *Polyommatus dorylas* (2.3%) and even between Polish and Slovakian populations of *Maculinea nausithous* (2.3%). Table 1 lists mean intraspecific divergences in those species that are represented by more than one individual in the data set.

Interspecific divergence

The average divergence in 236348 interspecific comparisons is 9.38% (SE = 3.65%) ranging from 0.0% to 23.2% (between *Baliochila minima* and *Agrodiaetus poseidon*). Of these, 57562 are congeneric comparisons with an average divergence of 5.07% (SE = 1.73%) ranging from 0.0% (between 23 *Agrodiaetus* as well as 3 *Maculinea* species pairs) to 12.4% (between *Arhopala abseus* and *Arhopala ace*). 94% of those comparisons are within *Agrodiaetus*.

Table 1: Intraspecific nucleotide divergences

Species	No. of individuals	Mean percent divergence	Standard error (%)	Range (%)	Monophyly	corrected
<i>Acrodipsas aurata</i>	3	0.2	0.1	0.2 – 0.3	Mono	Mono
<i>Acrodipsas brisbanensis</i>	8	1.0	0.5	0.2 – 1.6	Mono	Mono
<i>Acrodipsas cuprea</i>	6	0.5	0.3	0.2 – 0.9	Mono	Mono
<i>Acrodipsas hirtipes</i>	2	1.0		---	Mono	Mono
<i>Acrodipsas mortoni</i>	2	0.2		---	Mono	Mono
<i>Agrodiaetus admetus</i>	4	1.7	0.7	0.5 – 2.5	Poly	Poly
<i>Agrodiaetus ainsae</i>	4	0.3	0.2	0 – 0.6	Poly	
<i>Agrodiaetus alcestis</i>	6	0.8	0.4	0 – 1.5	Poly	Poly
<i>Agrodiaetus altivagans</i>	9	1.8	1.5	0 – 5.5	Poly	
<i>Agrodiaetus antidolus</i>	4	0.3	0.3	0 – 0.7	Poly	Poly
<i>Agrodiaetus arasbarani</i>	2	1.0		---	Poly	Poly
<i>Agrodiaetus baytopi</i>	4	1.9	1.2	0.5 – 3.1	Poly	Poly
<i>Agrodiaetus birunii</i>	10	0.2	0.2	0 – 0.7	Para	Para
<i>Agrodiaetus caeruleus</i>	3	0.5	0.5	0 – 1	Mono	Mono
<i>Agrodiaetus carmon</i>	4	1.3	0.6	0.6 – 2	Poly	Poly
<i>Agrodiaetus cyaneus</i>	6	0.2	0.2	0 – 0.7	Poly	Poly
<i>Agrodiaetus damocles</i>	4	1.1	0.8	0.1 – 1.8	Poly	Poly
<i>Agrodiaetus damon</i>	5	1.6	0.8	0 – 2.9	Mono	Mono
<i>Agrodiaetus damone</i>	3	0.6	0.0	0.6 – 0.6	Para	Para
<i>Agrodiaetus dantchenkoi</i>	6	0.0	0.0	0 – 0	Poly	Poly
<i>Agrodiaetus darius</i>	3	0.0	0.0	0 – 0	Mono	Mono
<i>Agrodiaetus demavendi</i>	17	2.1	1.3	0 – 3.6	Poly	Poly
<i>Agrodiaetus elbursicus</i>	9	0.5	0.8	0 – 2.1	Poly	Poly
<i>Agrodiaetus erschoffii</i>	3	0.2	0.2	0 – 0.3	Mono	Mono
<i>Agrodiaetus fabressei</i>	3	0.1	0.1	0 – 0.2	Poly	Poly
<i>Agrodiaetus femininoides</i>	2	1.8		---	Poly	Poly
<i>Agrodiaetus firdussii</i>	9	0.5	0.4	0 – 1.3	Poly	Mono
<i>Agrodiaetus fulgens</i>	2	0.2		---	Poly	Poly
<i>Agrodiaetus glaucias</i>	2	0.2		---	Mono	Mono
<i>Agrodiaetus gorbunovi</i>	5	0.1	0.1	0 – 0.2	Para	
<i>Agrodiaetus haigi</i>	3	0.0	0.0	0 – 0	Poly	
<i>Agrodiaetus hamadanensis</i>	4	0.4	0.3	0 – 0.7	Mono	Mono
<i>Agrodiaetus hopfferi</i>	3	1.5	1.3	0.2 – 2.8	Para	Para
<i>Agrodiaetus huberti</i>	7	0.5	0.4	0 – 1.3	Poly	
<i>Agrodiaetus humedasae</i>	2	0.2		---	Mono	Mono
<i>Agrodiaetus iphidamon</i>	4	0.0	0.0	0 – 0	Mono	Mono
<i>Agrodiaetus iphigenia</i>	8	0.7	0.6	0 – 2	Mono	Mono
<i>Agrodiaetus iphigenides</i>	3	1.6	0.8	0.7 – 2.2	Poly	Poly
<i>Agrodiaetus kanduli</i>	2	2.7		---	Poly	
<i>Agrodiaetus kendeveni</i>	2	5.4		---	Poly	
<i>Agrodiaetus khorasanensis</i>	2	0.5		---	Mono	
<i>Agrodiaetus klausschuriani</i>	3	0.0	0.0	0 – 0	Mono	Mono
<i>Agrodiaetus kurdistanicus</i>	3	0.0	0.0	0 – 0	Poly	Poly
<i>Agrodiaetus lorestanus</i>	2	0.0		---	Mono	
<i>Agrodiaetus lycius</i>	2	0.8		---	Mono	Mono
<i>Agrodiaetus menalcas</i>	5	0.6	0.4	0 – 1.3	Mono	Mono
<i>Agrodiaetus merhaba</i>	3	2.4	1.2	1.1 – 3.5	Poly	Poly
<i>Agrodiaetus mithridates</i>	2	4.6		---	Poly	Poly
<i>Agrodiaetus mofidii</i>	2	1.0		---	Poly	Poly
<i>Agrodiaetus nephohiptamemos</i>	2	0.0		---	Mono	
<i>Agrodiaetus ninae</i>	5	0.7	0.3	0.2 – 1.3	Poly	Poly
<i>Agrodiaetus paulae</i>	2	0.0		---	Para	Para
<i>Agrodiaetus phyllides</i>	4	0.7	0.2	0.4 – 0.9	Poly	Poly
<i>Agrodiaetus phyllis</i>	4	1.7	0.7	0.5 – 2.5	Para	Mono
<i>Agrodiaetus pierceae</i>	3	0.3	0.2	0.2 – 0.5	Mono	Mono
<i>Agrodiaetus poseidon</i>	5	0.5	0.4	0 – 1	Poly	Mono
<i>Agrodiaetus posthumus</i>	3	0.1	0.1	0 – 0.2	Mono	Mono
<i>Agrodiaetus pseudactis</i>	2	1.0		---	Poly	
<i>Agrodiaetus pseudoxerxes</i>	2	1.8		---	Poly	Poly
<i>Agrodiaetus putnami</i>	3	0.0	0.0	0 – 0	Poly	
<i>Agrodiaetus ripartii</i>	17	1.4	0.8	0 – 3.3	Poly	Poly
<i>Agrodiaetus rjabovi</i>	2	1.1		---	Mono	Mono
<i>Agrodiaetus rovshani</i>	4	0.2	0.2	0 – 0.4	Mono	Mono
<i>Agrodiaetus sekercioglu</i>	2	0.5		---	Poly	
<i>Agrodiaetus shahrami</i>	2	0.3		---	Poly	Poly
<i>Agrodiaetus sigberti</i>	2	0.9		---	Poly	
<i>Agrodiaetus surakovi</i>	2	0.2		---	Para	Poly

Table 1: Intraspecific nucleotide divergences (Continued)

<i>Agrodiaetus tankeri</i>	3	1.7	0.7	1 – 2.3	Poly	Poly
<i>Agrodiaetus tenhageni</i>	2	0.0		---	Mono	Mono
<i>Agrodiaetus turcicolus</i>	5	0.8	0.4	0 – 1.3	Poly	
<i>Agrodiaetus turcicus</i>	4	0.8	0.3	0.5 – 1.1	Mono	Mono
<i>Agrodiaetus valiabadi</i>	2	0.0		---	Mono	Mono
<i>Agrodiaetus vanensis</i>	5	0.5	0.3	0 – 0.8	Mono	
<i>Agrodiaetus wagneri</i>	2	0.1		---	Para	
<i>Agrodiaetus zapvadi</i>	4	0.0	0.1	0 – 0.1	Poly	
<i>Agrodiaetus zarathustra</i>	2	0.0		---	Mono	Mono
<i>Arhopala achelous</i>	6	1.3	0.9	0.2 – 3.1	Poly	Poly
<i>Arhopala antimuta</i>	2	1.1		---	Mono	Mono
<i>Arhopala atosia</i>	3	2.9	1.7	1 – 4.3	Poly	Poly
<i>Arhopala barami</i>	2	0.3		---	Mono	Mono
<i>Arhopala democritus</i>	2	1.0		---	Mono	Mono
<i>Arhopala epimuta</i>	6	0.2	0.3	0 – 0.8	Para	Para
<i>Arhopala labuana</i>	2	0.5		---	Mono	Mono
<i>Arhopala major</i>	2	0.3		---	Mono	Mono
<i>Arhopala moolaiana</i>	2	2.3		---	Mono	Mono
<i>Aricia agestis</i>	6	0.7	1.0	0 – 2.4	Poly	Poly
<i>Aricia artaxerxes</i>	3	1.8	0.5	1.2 – 2.1	Poly	
<i>Celastrina argiolus</i>	2	1.3		---	Mono	Mono
<i>Chrysoritis nigricans</i>	2	1.1		---	Mono	Mono
<i>Chrysoritis pyrois</i>	2	0.0		---	Mono	Mono
<i>Cyaniris semiargus</i>	4	1.0	0.4	0.3 – 1.5	Mono	Mono
<i>Favonius cognatus</i>	3	0.3	0.2	0.1 – 0.5	Poly	Poly
<i>Favonius jezoensis</i>	2	0.6		---	Mono	Mono
<i>Favonius korshunovi</i>	2	0.4		---	Mono	Mono
<i>Favonius orientalis</i>	3	0.7	0.5	0.1 – 1	Poly	Mono
<i>Favonius saphirinus</i>	8	1.1	0.7	0 – 2	Mono	Mono
<i>Favonius taxila</i>	3	0.1	0.1	0 – 0.1	Mono	Mono
<i>Favonius ultramarinus</i>	5	1.0	0.4	0.4 – 1.4	Poly	Poly
<i>Favonius yuasai</i>	3	0.5	0.4	0.1 – 0.8	Mono	Mono
<i>Flos anniella</i>	2	2.6		---	Mono	Mono
<i>Jalmenus evagoras</i>	12	0.6	0.3	0.2 – 1.1	Mono	Mono
<i>Lampides boeticus</i>	4	4.3	4.5	0.2 – 8.7	Poly	Mono
<i>Lucia limbaria</i>	2	1.7		---	Mono	Mono
<i>Lycaeides melissa</i>	5	0.7	0.5	0 – 1.2	Poly	Poly
<i>Lycaena alciphron</i>	2	2.7		---	Mono	Mono
<i>Lysandra albicans</i>	3	0.9	0.1	0.8 – 0.9	Para	
<i>Lysandra bellargus</i>	6	0.2	0.3	0 – 0.8	Mono	Mono
<i>Lysandra coridon</i>	5	1.6	0.5	0.7 – 2.1	Poly	Poly
<i>Lysandra corydonius</i>	4	1.7	1.2	0 – 2.7	Poly	Poly
<i>Lysandra ossmar</i>	2	2.1		---	Poly	
<i>Maculineaalcon</i>	7	0.0	0.1	0 – 0.2	Poly	Mono
<i>Maculinea arion</i>	10	0.2	0.2	0 – 0.6	Para	Para
<i>Maculinea arionides</i>	4	0.5	0.4	0 – 0.9	Poly	Poly
<i>Maculinea nausithous</i>	3	2.2	0.3	1.9 – 2.4	Mono	Mono
<i>Maculinea rebeli</i>	3	0.1	0.2	0 – 0.3	Poly	
<i>Maculinea teleius</i>	5	0.9	0.5	0.2 – 1.6	Mono	Mono
<i>Meleageria daphnis</i>	4	2.1	0.4	1.5 – 2.6	Poly	Mono
<i>Meleageria marcida</i>	2	4.4		---	Poly	
<i>Neolysandra fatima</i>	2	0.0		---	Mono	Mono
<i>Neozephyrus japonicus</i>	2	4.8		---	Poly	
<i>Plebejus argus</i>	5	1.0	0.8	0 – 1.9	Mono	Mono
<i>Polyommatus amandus</i>	3	2.6	2.0	0.3 – 3.8	Para	Para
<i>Polyommatus cornelia</i>	3	1.1	0.5	0.6 – 1.5	Para	Para
<i>Polyommatus dorylas</i>	4	1.6	0.4	1.2 – 2.3	Mono	Mono
<i>Polyommatus eroides</i>	2	1.4		---	Poly	Poly
<i>Polyommatus escheri</i>	2	2.0		---	Mono	Mono
<i>Polyommatus icarus</i>	8	2.2	2.3	0 – 6.8	Poly	Poly
<i>Polyommatus menelaos</i>	2	0.0		---	Mono	Mono
<i>Polyommatus myrrhinus</i>	3	0.1	0.1	0 – 0.1	Mono	Mono
<i>Polyommatus thersites</i>	5	0.9	0.6	0 – 1.6	Mono	Mono
<i>Pseudophilotes vicrama</i>	2	0.0		---	Mono	Mono
<i>Quercusia quercus</i>	2	0.6		---	Mono	Mono
<i>Vacciniina alcedo</i>	2	0.0		---	Mono	Mono

Mean and range of intraspecific nucleotide divergences for 133 Lycaenidae species, using Kimura's two parameter model. The column "Monophyly" states if conspecific sequences form a monophylum ("Mono"), a paraphylum ("Para") or a polyphylum ("Poly") and the subsequent column gives the corrected status (if presumable errors are excluded and critical taxa are lumped together).

Only congeneric comparisons were included in subsequent analyses in order to make comparisons feasible across taxonomic levels. Table 2 lists mean interspecific divergences in genera of which at least two species are represented in the data set. Sequence divergence in 95% of interspecific (congeneric) comparisons is above 1.9%, and 87.6% of such comparisons reveal distances above 3%.

The barcode gap

As apparent in Figure 1 (and Figure 2 for comparisons within *Agrodiaetus* only) no gap exists between intraspecific and interspecific divergences. Since some (0.14%) interspecific divergences are as low as 0% no safe threshold can be set to strictly avoid false negatives. Although species pairs with such low divergences include some whose taxonomic status as distinct species is debatable, they also include many pairs which are well differentiated in phenotype, have a very different karyotype (in *Agrodiaetus*), and occur sympatrically without any evidence for interbreeding. Examples include *Agrodiaetus peilei* – *A. morgani* (0.0%), *Agrodiaetus fabressei* – *A. ainsae* (0.2%), *Agrodiaetus peilei* – *A. karindus* (0.2%), *Polyommatus myrrhinus* – *P. cornelia* (0.4%), or *Agrodiaetus poseidon* – *A. hopfferi* (0.6%).

The minimum cumulative error based on false positives plus false negatives is 18% at a threshold level of 2.8%

(Figure 3). Minimum errors are very similar for *Agrodiaetus* (18.6% at 3.0% threshold, not shown) and other Lycaenidae (18.6% at 2.0% threshold, not shown), but much lower in *Arhopala* (5.3% at 3.4% threshold, Figure 4).

For safe identification, minimum distances between species (Figure 5) are critical and not average distances. In *Agrodiaetus*, all but two species (= 98.3%) have close relatives with interspecific distances below 3%. In the other genera combined, "only" 74% of taxa are affected but this lower rate is probably due to undersampling and would rise, if more sequences of more closely related species become available for the analysis.

Identification with NJ tree profile

The approach of species identification with a Neighbour-Joining (NJ) tree profile as proposed by [9] does not necessarily depend on the barcoding gap but on the coalescence of conspecific populations and the monophyly of species (details see Data analysis).

The success rate in the identification of our Lycaenidae data set with this method was 58%. Five out of 158 misidentifications or ambiguous identifications (3.2%) can be attributed to incorrectly identified specimens (*Lampides boeticus*, *Neozephyrus japonicus*, *Agrodiaetus kendevani*, see above). Further 90 cases (57%) were among closely

Table 2: Interspecific nucleotide divergences

Genus	No. of species	Mean percent divergence	Standard error (%)	Range (%)
<i>Acrodipsas</i>	9	3.1	1.0	0.5 – 5.7
<i>Agriades</i>	2	4.7		---
<i>Agrodiaetus</i>	117	5.1	1.7	0 – 10.1
<i>Arhopala</i>	30	6.8	1.7	0.4 – 12.4
<i>Aricia</i>	7	3.4	1.9	0.2 – 7.5
<i>Chrysothrix</i>	19	7.0	2.5	0.8 – 10.9
<i>Euphilotes</i>	2	10.3		---
<i>Favonius</i>	9	4.0	0.9	0.1 – 5.4
<i>Glaucopteryx</i>	2	1.3		---
<i>Lycaeides</i>	3	1.7	0.9	0.5 – 3.0
<i>Lycaena</i>	9	4.5	1.1	1.2 – 6.8
<i>Lysandra</i>	9	2.2	0.7	0.7 – 4.0
<i>Maculinea</i>	7	2.8	1.4	0 – 6.0
<i>Meleageria</i>	2	2.6	1.6	0.1 – 4.4
<i>Neolysandra</i>	5	4.6	1.6	1 – 6.3
<i>Phengaris</i>	3	3.8	2.1	1.3 – 5.1
<i>Plebejus</i>	5	5.6	1.6	2.4 – 7.4
<i>Polyommatus</i>	12	5.9	2.5	0.1 – 10.5
<i>Pseudophilotes</i>	4	2.7	1.6	0.6 – 4.5
<i>Satyrium</i>	3	4.5	0.5	4 – 4.9
<i>Trimenia</i>	2	6.1		---
<i>Turanana</i>	2	4.8		---
<i>Vacciniina</i>	3	7.2	0.3	6.8 – 7.5

Mean and range of interspecific nucleotide divergences for species in 22 Lycaenidae genera, using Kimura's two parameter model

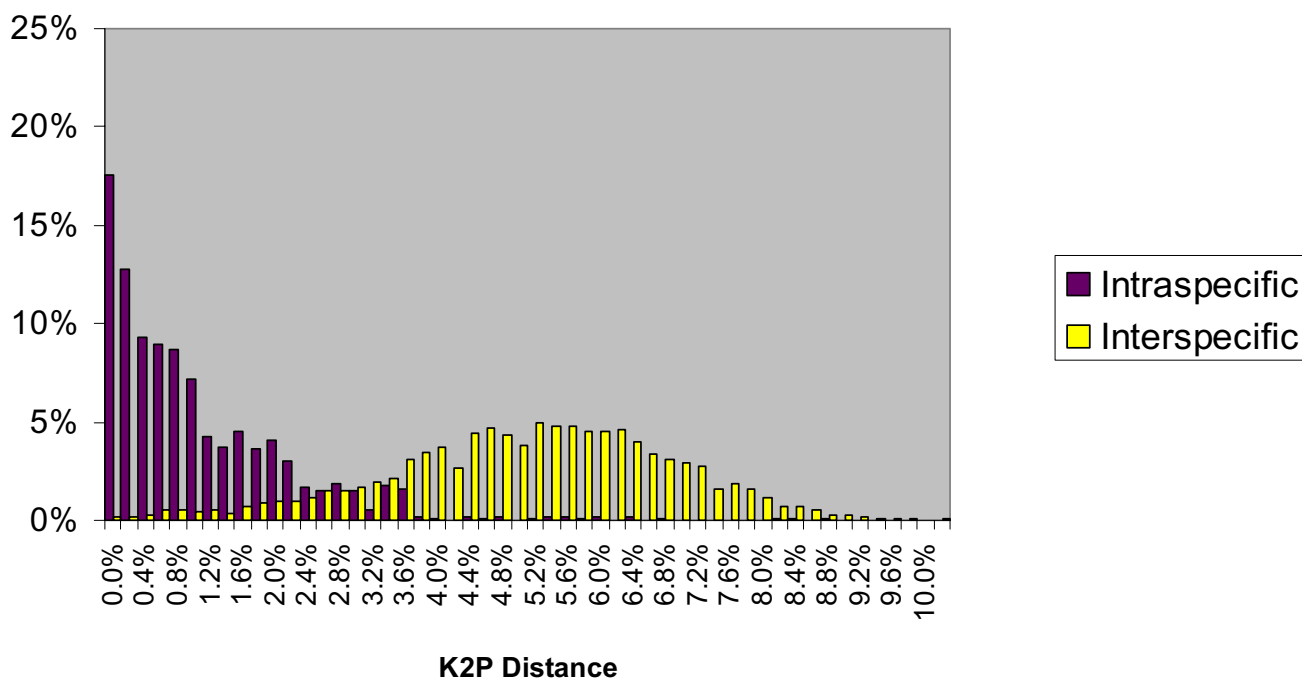


Figure 1
Frequency distribution of intraspecific and interspecific (congeneric) genetic divergence in Lycaenidae. Total number of comparisons: 1189 intraspecific and 57562 interspecific pairs across 315 Lycaenidae species. Divergences were calculated using Kimura's two parameter (K2P) model.

related sister species whose taxonomic status is in dispute (Table 3). If these cases are not taken into account (i.e. counted as successful identifications, an unrealistic best case scenario for barcoding success), the success rate would rise to 84%. In *Agrodiaetus* the success rate would remain lower (79%) while in the remaining genera it would reach 91%. But even with these corrections, 61 cases of misidentifications (16%) remain, 46 of these in *Agrodiaetus* (affected taxa in Table 4). The complete Neighbour-joining tree (available for download as additional file 1: NJ-tree) shows the reason for this failure: Only 46% of conspecific sequences form a monophyletic group on this tree while the others are either paraphyletic (10%) or even polyphyletic (44%). In *Agrodiaetus*, only 34% of species are monophyletic (Table 1), while the others are paraphyletic (11%) or polyphyletic (55%). If incorrectly identified specimens are excluded and critical taxa (Table 3) are lumped together, still only 59% of species are monophyletic (43% in *Agrodiaetus*) while 7% are paraphyletic and 34% polyphyletic (49% in *Agrodiaetus*).

Conclusion

We found an upper limit for intraspecific sequence divergences in a wide range of species of the diverse butterfly family Lycaenidae, but no lower limit for interspecific

divergences and thus no barcoding gap. This result is especially well documented in the comprehensively sampled genus *Agrodiaetus* (114 of ca 130 recognized species sequenced) while the smaller overlap in *Arhopala* can be attributed to the lower percentage of species sampled (33 of more than 200 species). The choice of species by [46] was to maximize coverage of divergent clades while minimizing the total number of species which is a common and sensible approach for phylogenetic studies, but undermines the power of such sequence data as critical tests for the barcoding approach. The general level of sequence divergence is not exceptionally low in Lycaenidae compared to other Lepidoptera. The mean congeneric interspecific sequence divergence of 5.1% in Lycaenidae (5.1% in *Agrodiaetus* and 5.0% in the other genera) was only slightly lower than the mean value of 6.6% found by [2] in various families of Lepidoptera.

We thus confirm the results of Meyer & Paulay [17] and Meier et al. [18]. Our results also agree with those from a recent study in the Neotropical butterfly subfamily Ithomiinae (Nymphalidae) [47] which records highly variable levels of divergence in mtDNA (*COI* & *COII*) between taxa of the same rank. Our results however fail to agree with those of Barrett & Hebert [9] on arachnids. In that study

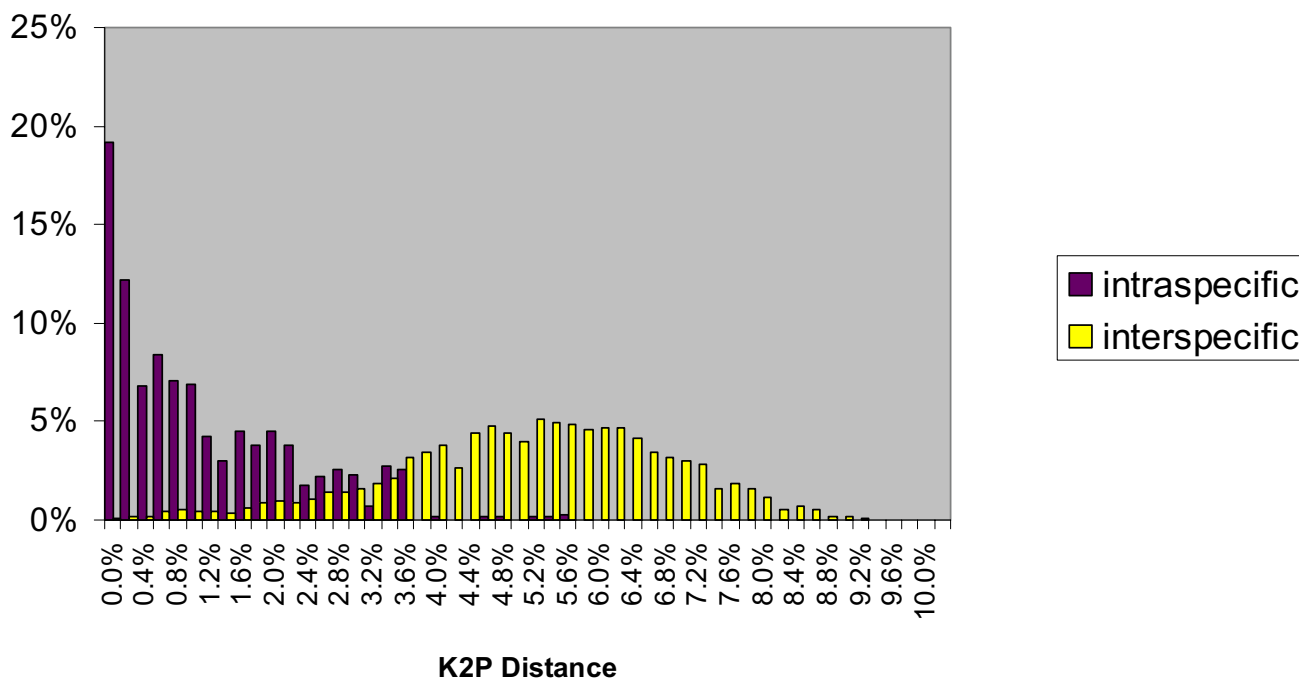


Figure 2

Frequency distribution of intraspecific and interspecific (congeneric) genetic divergences in *Agrodiaetus*. Total number of comparisons: 737 intraspecific and 54209 interspecific pairs across 114 *Agrodiaetus* species. Divergences were calculated using Kimura's two parameter (K2P) model.

the mean percent sequence divergence between congeneric species was 16.4% (SE = 0.13) and thus three times higher than in our study while the divergence among conspecific individuals was only slightly higher with 1.4% (SE = 0.16). The contradiction between our study and theirs can be explained by the very incomplete and sparse taxon sampling in their data set amounting to just 1% of the species contained within the families. We conclude that the reported existence of a barcode gap in arachnids appears to be an artifact based on insufficient sampling across taxa.

Despite these difficulties, species identification of unidentified samples with the help of barcodes is entirely possible. The NJ tree profile approach which does not rely on a barcode gap enabled the correct assignment of many sequences, and other methods (e.g. applying population genetic approaches) might further increase the success rate. However, 17% of test sequences could still not be identified correctly, even in some sympatric species pairs which clearly differ in phenotype and chromosome number (e.g. *Agrodiaetus ainsae* [n = 108–110]/*fabressei* [n = 90], *Agrodiaetus hopfferi* [n = 15]/*poseidon* [n = 19–22]). The main reason for this failure is that a large proportion of species are not reciprocally monophyletic, e.g. due to

incomplete lineage sorting, which is in accordance with a previous study [48]. Moreover, the success with this method is again completely dependent on comprehensive sampling. If the correct species is not included in the profile, the assignment must by necessity be incorrect and misleading. Because of the non-existence of a barcoding gap, this error will often be impossible to detect. This limits possible applications of the barcoding approach. For example, cryptic species can only be detected with the help of a barcoding approach at high genetic divergence from all phenotypically similar species. An example is *Agrodiaetus paulae* which was discovered in this way [41]. In contrast, and on the one hand, the sympatric species pairs *Agrodiaetus ainsae-fabressei*, *A. hopfferi-poseidon* and *A. morgani-peilei* would have gone unnoticed by barcoding approaches even though their strong phenotypical and karyological differentiation (n = 108 vs. n = 90, n = 15 vs. n = 19–22 and n = 27 vs. n = 39, respectively) clearly indicates their specific distinctness. On the other hand, sequence divergence in what is currently believed to represent one species does not *per se* prove the specific distinctness of the entities in question. In *Polyommatus icarus* or *P. amandus*, for example, the high divergences between North African and Eurasiatic samples is a strong hint for the presence of unrecognized cryptic species, but this

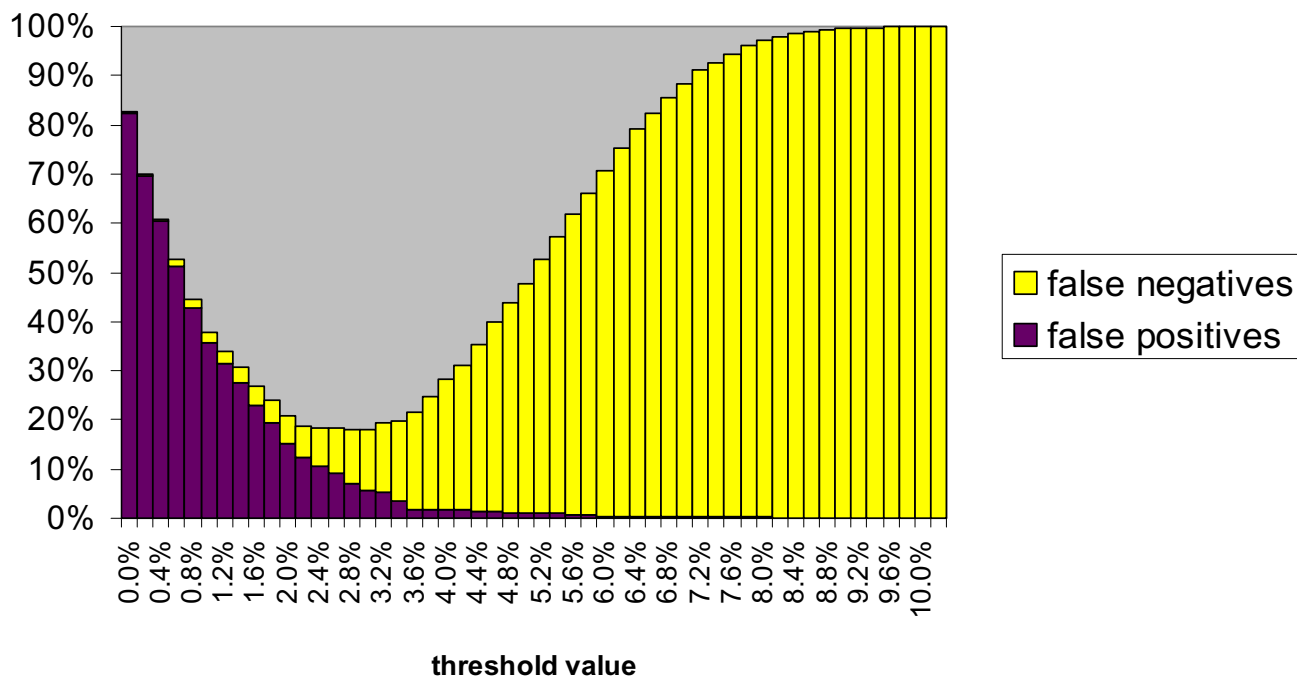


Figure 3
Cumulative error based on false positives plus false negatives for each threshold value in 315 Lycaenidae species including only congeneric comparisons. The optimum threshold value is 2.8%, where error is minimized at 18.0%.

needs to be rigorously tested with sequence data from samples that cover the geographic range more comprehensively. Also in practical application the problem of misidentified specimens and sequences in GenBank remains a real threat to the accuracy of barcode-based identifications. An example is the GenBank sequence [AB192475](#) of *Lampides boeticus* which is also used in the CBOL database (see above). This underscores the importance of voucher specimens and documentation of locality data, an issue raised by barcoding supporters but unfortunately still much neglected by GenBank. Another case of misidentification (GenBank sequence [AF170864](#) of *Plebejus acmon* which was originally submitted as *Euphilotes bernardino*) [30] has already been corrected with the help of the voucher specimen.

In conclusion, the barcoding approach can be very helpful, e.g. in identifying early stages of insects or when only fragments of individuals are available for analysis. However, correct identification requires that all eligible species can be included in the profile and that sufficient information is available on the amount of intraspecific genetic variation and genetic distance to closely related species.

The barcoding procedure is not very well suited for identifying species boundaries but it may help to give mini-

imum estimates of species numbers in very diverse and inadequately known taxonomic groups at single localities. Our case study on *Agrodiaetus* shows that a substantial number of species would have gone unnoticed by the barcoding approach as 'false negatives'. Thus, especially in clades where many species have evolved rapidly as a result of massive radiations with minimum sequence divergence, the barcoding approach holds little promise of meeting the challenge of rapid and reliable identification of large samples. Yet, it is exactly these situations which pose the most problematic tasks in the morphological identification of insects.

Although molecular data can be helpful in discovering new species, a large genetic divergence is not sufficient proof since it must be corroborated by other data. Furthermore, most closely related species which are difficult to identify with traditional means, are also similar genetically and would go unnoticed by an isolated barcoding approach. Mathematical simulations have shown that populations have to be isolated for more than 4 million generations (i.e. 4 million years in the mostly univoltine *Agrodiaetus* species) for two thresholds proposed by the barcoding initiative (reciprocal monophyly, and a genetic divergence between species which is 10 times greater than within species) to achieve error rates less than 10% [49].

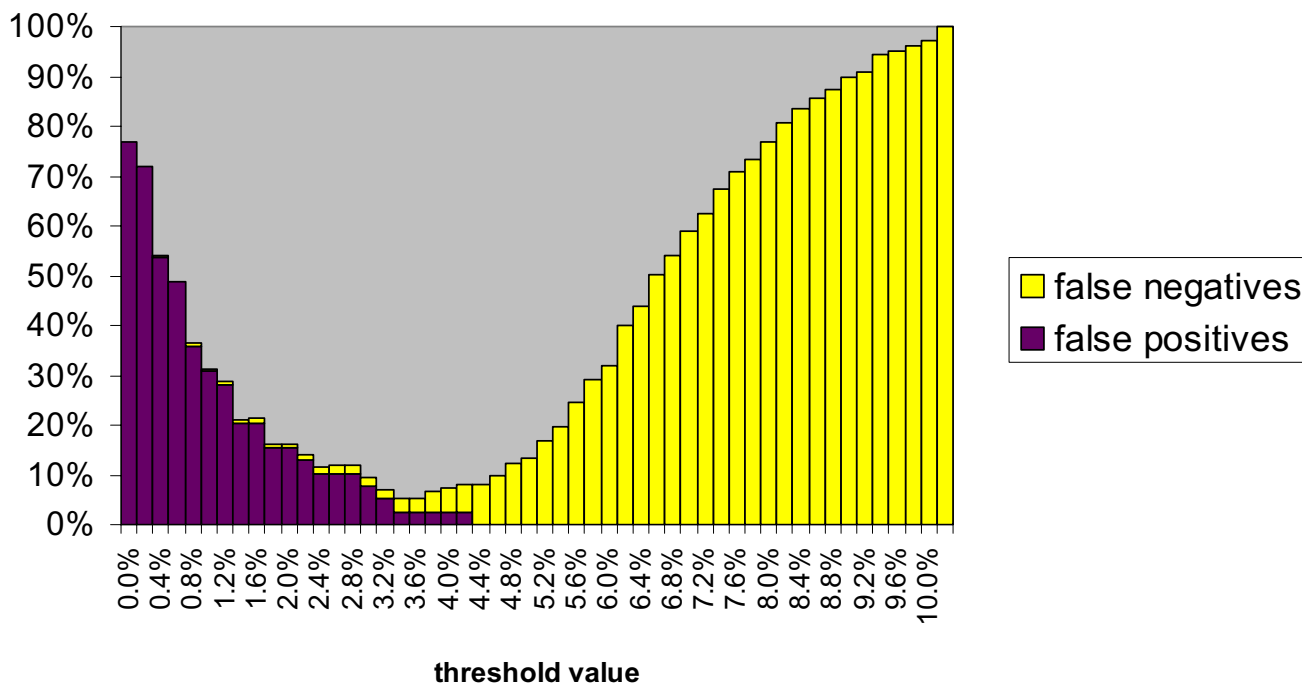


Figure 4
Cumulative error based on false positives plus false negatives for each threshold value in 30 *Arhopala* species.
 The optimum threshold value is 3.4%, where error is minimized at 5.3%.

This might help to explain why the barcoding approach appears to be more successful in the Oriental genus *Arhopala* which is thought to represent a phylogenetically older lineage of Lycaenidae estimated to be about 7–11 Million years old [50], while the origin of the Palearctic genus *Agrodiaetus* is dated at only 2.5–3.8 Million years [44].

Our data show that the lack of a barcoding gap and reciprocal monophyly in Lycaenidae is not confined to the genus *Agrodiaetus* with its extraordinary interspecific variation in chromosome numbers, but also to other genera of Lycaenidae with stable chromosome numbers. It should also be noted that in *Agrodiaetus* there is neither evidence for exceptional rapid radiation as in cichlids of the East African lakes [51] nor for unusual (i.e. sympatric) speciation patterns caused by karyotype evolution. Rather, karyotype diversification seems to have been a mere by-product of the usual mode of allopatric speciation [29,30,44].

Methods

Data sources

A total of 694 barcode sequences were used for our analysis. We used a 690 bp fragment at the 5' end of cytochrome c oxidase subunit I (*COI*) of 309 Lycaenidae sequences from a molecular phylogenetic study by Wiem-

ers [30]. Most sequences belong to *Agrodiaetus* (198), the others (111) mostly to closely related Polyommatae. All sequences have been deposited in GenBank [52] ([AY556844-AY556867](#), [AY556869-AY556963](#), [AY556965-AY557155](#)) with LinkOuts provided to images of the voucher specimens deposited with MorphBank [53]. These sequences were supplemented by 385 further sequences of Lycaenidae deposited in GenBank as of March, 2006 (Table 5). They include sequences from further studies on *Agrodiaetus* [29,44], the Palearctic genus *Maculinea* [54], Nearctic *Lycaeides melissa* [55], the Oriental genus *Arhopala* [46,50], the Australian genera *Acrodipsas* [56] and *Jalmenus* [57], and the South African *Chrysoritis* [58] as well as a few sequences which have only been used as outgroups in non-Lycaenidae studies (e.g. [59,60]). Sequence length in the 5' region as defined by CBOL ranged between 240 bp and the maximum of 987 bp. (18 *COI* sequences from a study on *Japonica* only contained a 3' end fragment and therefore were not included.) Of these, 89% are at least 648 bp long as recommended by CBOL and 98% at least 500 bp long which is deemed sufficient for barcode sequences [13]. However, sequence overlap for sequences from different studies was sometimes lower because of slightly different sequence locations within the barcode region (Figure 6). It should be noted that these inconsistencies in barcode comparisons

Minimum interspecific distances

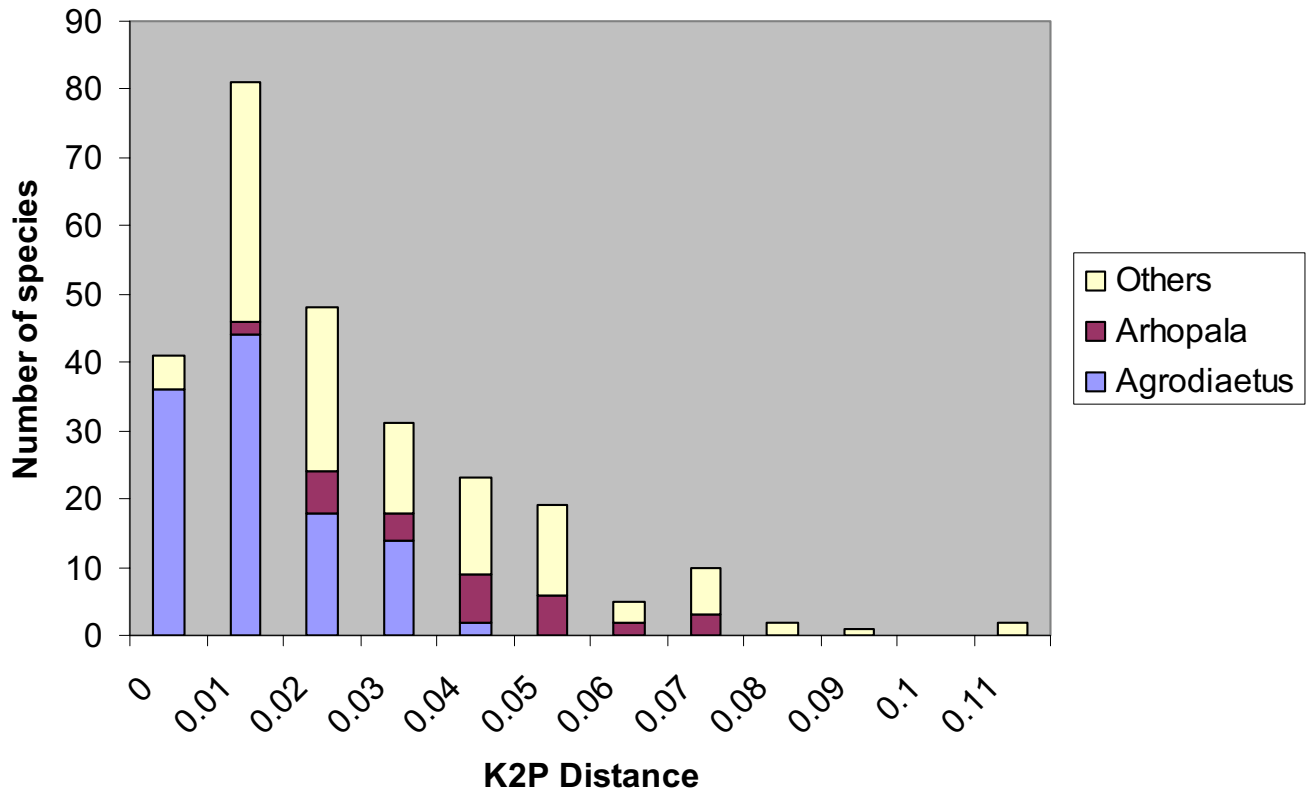


Figure 5
Frequency distribution of minimum interspecific (congeneric) genetic distances across 263 Lycaenidae species.

are a common situation in barcode sequences due to differences in primer use (e.g. [2]).

Laboratory protocols

DNA was extracted from thorax tissue recently collected and preserved in 100% ethanol using Qiagen® DNeasy Tissue Kit according to the manufacturer's protocol for mouse tail tissue. In a few cases only dried material was available and either thorax or legs were used for DNA extraction.

Amplification of DNA was conducted using the polymerase chain reaction (PCR). The reaction mixture (for a total reaction volume of 25 µl) included: 1 µl DNA, 16.8 µl

ddH₂O, 2.5 µl 10 × PCR II buffer, 3.2 µl 25 mM MgCl₂, 0.5 µl 2 mM dNTP-Mix, 0.25 µl Taq Polymerase and 0.375 µl 20 pm of each primer. The two primers used were:

Primer 1: k698 TY-J-1460 TAC AAT TTA TCG CCT AAA CTT CAG CC [61]

Primer 2: Nancy C1-N-2192 (CCC) GGT AAA ATT AAA ATA TAA ACT TC [61]

PCR was conducted on thermal cyclers from Biometra® (models Uno II or T-Gradient) or ABI Biosystems® (model GeneAmp® PCR-System 2700) using the following profiles:

Table 3: Sister species or species complexes with disputable species borders

Agrodiaetus altivagans/**damocles**/*ectabanensis*/*gorbunovi*/*kanduli*/*maraschi*/*wagneri* [30, 41]
Agrodiaetus artvinensis/*bilgini*/**firdussii**/*pseudactis*/*sigberti* [30]
Agrodiaetus aserbeidschanus/*huberti*/**ninae**/*turcicolus*/*zapvadi* [30]
Agrodiaetus baytopi/*iphicarmon* [30]
Agrodiaetus carmon/*schuriani* [30]
Agrodiaetus cyaneus/*kermansis*/*paracyaneus*
Agrodiaetus demavendi/*lorestanus* [30]
Agrodiaetus khorasanensis/*nephohiptamenos*/**ripartii** [30]
Agrodiaetus phyllis/*vanensis* [30]
Agrodiaetus poseidon/*putnami* [30]
Agrodiaetus sekercioglu/**surakovi** [30]
Aricia agestis/*artaxerxes* [30]
Lysandra albicans/*caelestissimus*/**coridon**/*gennargenti* [30]
Lysandra caucasicus/**corydonius**/*ossmar* [30]
Maculineaalcon/*rebeli* [30, 54]
Meleageria daphnis/*marcida* [30, 42, 54]
Polyommatus andronicus/**icarus** [30]
Polyommatus eros/*eroides* [30]

List of disputable species complexes due to e.g. incomplete speciation and gene flow or, in *Agrodiaetus*, very similar phenotype and only slight differences in karyotype. The taxonomically oldest name is marked in bold.

Table 4: Taxa misidentified with the NJ tree profile approach

Agrodiaetus admetus (78–80)/**demavendi** (≈67)/*nephohiptamenos* (≈90)
Agrodiaetus ainsae (108–110)/**fabressei** (90)
Agrodiaetus alcestis (19–21)/**dantchenkoi** (40–42)/*eriwanensis* (28–32)/*interjectus* (29–32)
Agrodiaetus altivagans (18–22)/*ciscaucasicus* (16)
Agrodiaetus antidolus (42)/**femininoides** (27)/*kurdistanicus* (62)
Agrodiaetus arasbarani (25)/*elbursicus* (16)/*lukhtanovi* (22)/**paulae** (17)/*zarathustra* (≈22)
Agrodiaetus baytopi (27–28)/**tanker** (20–21)
Agrodiaetus birunii (10–11)/*brandti* (19)
Agrodiaetus carmon (81–82)/*surakovi* (50)
Agrodiaetus ciscaucasicus (16)/**mofidii** (35)
Agrodiaetus cyaneus (19)/**pseudoxerxes** (15–16)
Agrodiaetus damone (66–68)/**iphigenides** (67)/*juldusus* (67)/*karatavicus* (67)/**phyllides** (67)
Agrodiaetus elbursicus (17)/*turcicolus* (20)
Agrodiaetus hopfferi (15)/**poseidon** (19–22)
Agrodiaetus lorestanus (68)/**ripartii** (90)
Arhopala achelous/*muta*
Favonius cognatus/**ultramarinus**
Maculinea arion/**arionides**
Polyommatus amandus **abdelaziz** /*Meleageria daphnis*
Polyommatus cornelia/*myrrhinus*

List of taxa which were misidentified with the NJ tree profile approach (excluding possible errors and critical taxa listed in Tab.3). Misidentified test taxa (in bold font) and their identifications are placed jointly in a single line. Haploid chromosome numbers of *Agrodiaetus* species (taken from [29, 30, 44]) are given in parenthesis.

Table 5: Material

GenBank accession no.	Number of sequences	Reference	Taxa in focus
<u>AY556844 – AY556867</u> <u>AY556869 – AY556963</u> <u>AY556965 – AY557155</u>	309	[30, 41]	<i>Agrodiaetus</i>
<u>AY496709 – AY496821</u> <u>AY502111 – AY502112</u> <u>AY953984 – AY954025</u>	157	[29, 44]	<i>Agrodiaetus</i>
<u>AY235861 – AY235903</u> <u>AY235955 – AY236006</u>	52	[46, 50]	<i>Arhopala</i>
<u>AY675402 – AY675448</u>	47	[54]	<i>Maculinea</i>
<u>DQ234691 – DQ234695</u>	5	[55]	<i>Lycaeides</i>
<u>AY091712 – AY091741</u>	30	[56]	<i>Acrodipsas</i>
<u>DQ249942 – DQ249953</u>	12	[57]	<i>Jalmenus</i>
<u>AF279217 – AF279244</u>	28	[58]	<i>Chrysoritis</i>
<u>AF170864</u>	1	[59]	Papilionidae
<u>AY350456 – AY350459</u>	4	[60]	Lepidoptera
<u>DQ018938 – DQ018948</u>	11	[67]	Papilionoidea & Hesperioidea
<u>AB195510 – AB195545</u>	36	Odagiri et al. (unpubl.)	<i>Favonius</i>
<u>AB192475 – AB192476</u>	2	Tanikawa et al. (unpubl.)	Hesperiidae

List of GenBank accession numbers used for analysis including references and taxa which were the focus of these studies

Initial 4 minutes denaturation at 94°C and 35 cycles of 30 seconds denaturation at 94°C, 30 seconds annealing at 55°C and 1 minute extension at 72°C.

PCR products were purified using purification kits from Promega® or Sigma® and checked with agarose gel electrophoresis before and after purification.

Cycle sequencing was carried out on Biometra®T-Gradient or ABI Biosystems® GeneAmp® PCR-System 2700 thermal cyclers using sequencing kits of MWG Biotech® (for Li-cor® automated sequencer) or ABI Biosystems® (for ABI® 377 automated sequencer) according to the manufacturers' protocols and with the following cycling times: initial 2 minutes denaturation at 95°C and 35 cycles of 15 seconds denaturation at 95°C, 15 seconds annealing at 49°C and 15 seconds extension at 70°C. Primers used were the same as for the PCR reactions for the ABI (primer 1 was used for forward and primer 2 for independent reverse sequencing), but for Li-cor truncated and labelled primers were used with 3 bases cut off at the 5' end and labelled with IRD-800. For ABI sequencing the products were cleaned using an ethanol precipitation protocol. Electrophoresis of sequencing reaction products was carried out on Li-cor® or ABI® 377 automated sequencers using the manufacturer's protocols.

Data analysis

Sequences were aligned with BioEdit 7.0.4.1 [62] and pruned to a maximum of 987 bp, the section proposed by CBOL for barcoding. Pairwise sequence divergences were calculated separately for intraspecific as well as for interspecific, but intrageneric comparisons with Mega 3.1 [63] using Kimura's two parameter (K2P) distance model. This

is not necessarily the best model to analyze the data (see [64]), but it was chosen to facilitate comparisons with other barcode studies of Hebert and co-workers [1,9-12,16] who have been using this model. Distance tables were processed to calculate divergence means (incl. standard errors and ranges) within and between species.

The taxonomy was taken from GenBank in most cases but two minor spelling inconsistencies were corrected. In four cases where a taxon within *Agrodiaetus* was treated as a species taxon by one author but only as a subspecies by another, we matched them by treating those taxa as distinct species. The generic subdivision of Lycaenidae is very much in flux. Some genera are only treated as subgenera by some authors and many genera (like *Polyommatus* or *Plebejus*) are probably paraphyletic or polyphyletic, however we undertook no revision of the GenBank taxonomy since it appeared consistent enough for our analysis. The remaining inconsistencies only affect few taxa in our analysis and include the treatment of *Sublysandra* (distinct genus or subgenus of *Polyommatus*), *Eumedonia* (distinct genus or subgenus of *Aricia*), *Otrnjukovia* (synonym to *Turanana*), *Maculinea* (synonym to *Phengaris*) and *Callipsyche* (synonym to *Satyrrium*). (A complete list of sequences with corresponding taxa names and voucher numbers is found in the additional file 1: NJ tree.)

A Lycaenidae species profile was created according to [9]. Of the 694 barcode sequences, we excluded 9 short *Arhopala* sequences with a barcode length of only 240 bp. (To check the position of those sequences, a separate analysis was run containing only the *Arhopala* sequences.) Of the remaining 685 sequences, we randomly selected 1 sequence from each of the 308 Lycaenidae species for

Sequence overlap for pairwise barcode comparisons

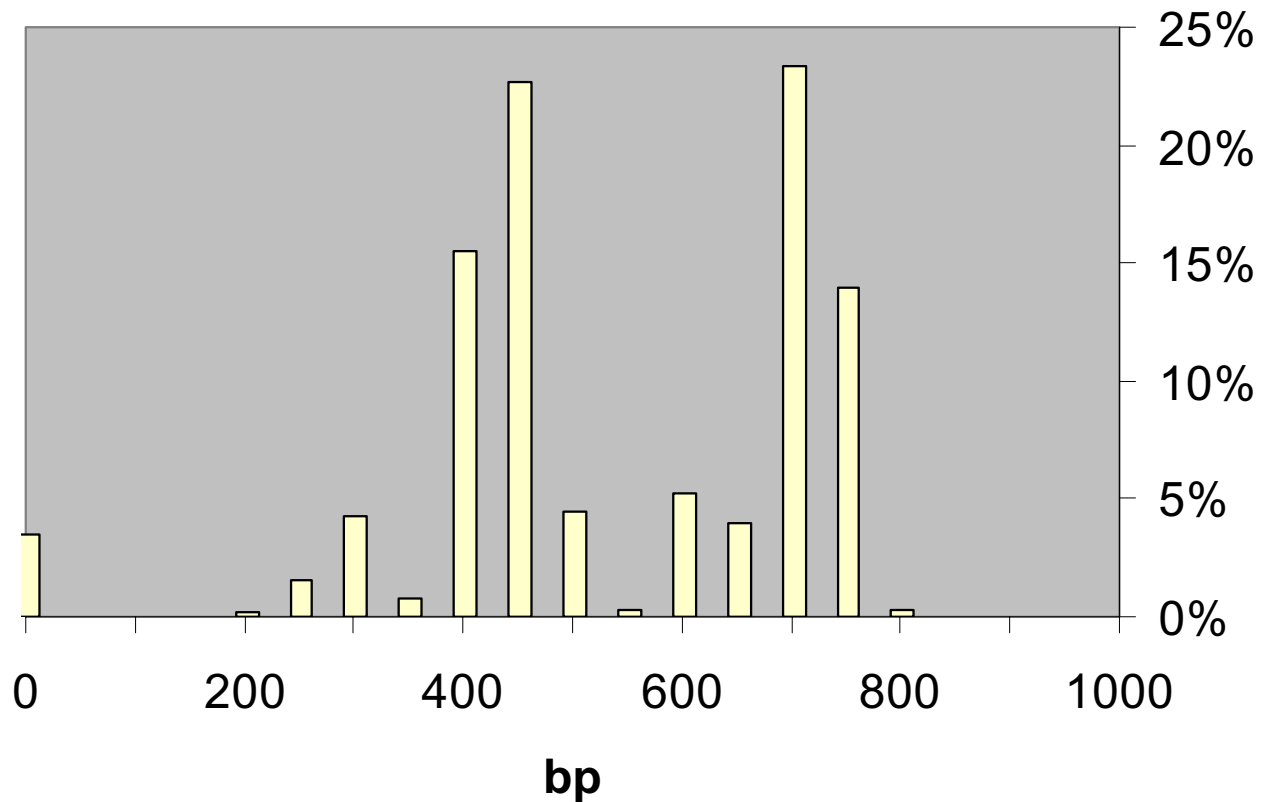


Figure 6
Sequence overlap for pairwise barcode comparisons. Length of sequence overlap in 246229 cross-comparisons of 694 aligned sequences

inclusion into a *COI* species profile. We chose a sequence of *Apodemia mormo* (GenBank accession number [AF170863](#)) from the family Riodinidae as outgroup because this family appears to represent the sister group to Lycaenidae [65-67]. The other 377 sequences which had not been included in the profile were used as "test" sequences: They were singly added to the test profile in repeated Neighbour-joining analyses and their "classification success" was recorded. A test was recorded as successful if the test sequence grouped most closely with the conspecific profile sequence and not with another species. Results of three GenBank sequences which were not identified to species level (all belonging to the genus *Agrodiaetus*) were not counted. After the classification test, another NJ analysis was run including all sequences in order to understand possible failures in classification. The main

reason for using the Neighbour-joining as a tree-building method is its computational efficiency. Although this method is well suited for grouping closely related sequences, it should be noted that other methods (such as Maximum Parsimony, Maximum Likelihood or Bayesian inference of phylogeny) are usually superior in constructing phylogenetic trees.

Competing interests

The author(s) declare that they have no competing interests.

Authors' contributions

MW carried out the molecular genetic studies, sequence alignment, statistical analysis and drafted the manuscript. KF participated in the design of the study and the statisti-

cal analysis and helped to draft the manuscript. All authors read and approved the final manuscript.

Additional material

Additional file 1

Neighbour-joining tree (Distance model: Kimura-2-Parameter) of profile and test taxa; includes a list of GenBank sequences with taxa names and corresponding voucher codes.

Click here for file

[<http://www.biomedcentral.com/content/supplementary/1742-9994-4-8-S1.xls>]

Acknowledgements

Most of the sequencing work was carried out by the first author at the molecular lab of the Alexander Koenig Research Institute and Museum of Zoology in Bonn. We thank the late Clas Naumann for supervision and assistance in many ways; Bernhard Misof for supervision of the molecular work; Esther Meyer, Ruth Rottscheidt, Meike Thomas, Manuela Brenk and Claudia Huber for assistance in DNA sequencing; Axel Hille, Claudia Etzbauer, Rainer Sonnenberg, Anja Schunke and Oliver Niehuis for general assistance in the lab; Karen Meusemann, Jurate De Prins and Vladimir Lukhtanov for karyological preparations; Wolfgang Eckweiler, Klaus G. Schurian, Alexandre Dantchenko, John Coutsis, José Munguira and Otakar Kudrna for specimen samples; Sabine Fischer for assistance with computerized analyses; James Mallet and an anonymous reviewer for corrections and helpful comments to the first draft of the manuscript. This study was supported by the Deutsche Forschungsgemeinschaft (DFG grant Na 90/14).

References

1. Hebert PD, Cywinska A, Ball SL, deWaard JR: **Biological identifications through DNA barcodes.** *Proc Biol Sci* 2003, **270(1512)**:313-321.
2. Hebert PD, Ratnasingham S, deWaard JR: **Barcoding animal life: cytochrome c oxidase subunit I divergences among closely related species.** *Proc Biol Sci* 2003, **270 Suppl 1**:S96-9 [<http://www.journals.royalsoc.ac.uk/openurl.asp?genre=article&id=doi:10.1098/rsbl.2003.0025>].
3. Ebach MC, Holdrege C: **DNA barcoding is no substitute for taxonomy.** *Nature* 2005, **434(7034)**:697.
4. Moritz C, Cicero C: **DNA barcoding: promise and pitfalls.** *PLoS Biol* 2004, **2(10)**:e354.
5. Smith VS: **DNA barcoding: perspectives from a "Partnerships for Enhancing Expertise in Taxonomy" (PEET) debate.** *Syst Biol* 2005, **54(5)**:841-844.
6. Sperling FA: **DNA Barcoding: Deus ex Machina.** *Newsl Biol Surv Canada (Terr Arthropods)* 2003, **22(2)**:50-53.
7. Will KW, Mishler BD, Wheeler QD: **The perils of DNA barcoding and the need for integrative taxonomy.** *Syst Biol* 2005, **54(5)**:844-851.
8. Will KW, Rubinoff D: **Myth of the molecule: DNA barcodes for species cannot replace morphology for identification and classification.** *Cladistics* 2004, **20**:47-55.
9. Barrett RDH, Hebert PD: **Identifying spiders through DNA barcodes.** *Can J Zool* 2005, **83**:481-491.
10. Hebert PD, Stoeckle MY, Zemlak TS, Francis CM: **Identification of Birds through DNA Barcodes.** *PLoS Biol* 2004, **2(10)**:e312.
11. Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W: **Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fulgerator*.** *Proc Natl Acad Sci U S A* 2004, **101(41)**:14812-14817.
12. Smith MA, Fisher BL, Hebert PD: **DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar.** *Philos Trans R Soc Lond B Biol Sci* 2005, **360(1462)**:1825-1834.
13. Smith MA, Woodley NE, Janzen DH, Hallwachs W, Hebert PD: **DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae).** *Proc Natl Acad Sci U S A* 2006, **103(10)**:3657-3662.
14. Brower AVZ: **Problems with DNA barcodes for species delimitation: 'ten species' of *Astrartes fulgerator* reassessed (Lepidoptera: Hesperidae).** *Syst Biodiv* 2006, **4(2)**:127-132.
15. Mayr E: **Principles of systematic zoology.** New York, McGraw-Hill; 1969.
16. Hajibabaei M, Janzen DH, Burns JM, Hallwachs W, Hebert PD: **DNA barcodes distinguish species of tropical Lepidoptera.** *Proc Natl Acad Sci U S A* 2006, **103(4)**:968-971.
17. Meyer CP, Paulay G: **DNA barcoding: error rates based on comprehensive sampling.** *PLoS Biol* 2005, **3(12)**:e422.
18. Meier R, Shiyang K, Vaidya G, Ng PKL: **DNA barcoding and taxonomy in Diptera: a tale of high intraspecific variability and low identification success.** *Syst Biol* 2006, **55(5)**:715-728.
19. Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, Kamoun S, Sumlin WD, Vogler A: **Sequence-based species delimitation for the DNA taxonomy of undescribed insects.** *Syst Biol* 2006, **55(4)**:595-609.
20. Matz MV, Nielsen R: **A likelihood ratio test for species membership based on DNA sequence data.** *Philos Trans R Soc Lond B Biol Sci* 2005, **360(1462)**:1969-1974.
21. Nielsen R, Matz M: **Statistical approaches for DNA barcoding.** *Syst Biol* 2006, **55(1)**:162-169.
22. Hogg ID, Hebert PDN: **Biological identification of springtails (Collembola: Hexapoda) from the Canadian Arctic, using mitochondrial DNA barcodes.** *Can J Zool* 2004, **82**:749-754.
23. Janzen DH, Hajibabaei M, Burns JM, Hallwachs W, Remigio E, Hebert PD: **Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding.** *Philos Trans R Soc Lond B Biol Sci* 2005, **360(1462)**:1835-1845.
24. Monaghan MT, Balke M, Gregory TR, Vogler AP: **DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers.** *Philos Trans R Soc Lond B Biol Sci* 2005, **360(1462)**:1925-1933.
25. Monaghan MT, Balke M, Pons J, Vogler AP: **Beyond barcodes: complex DNA taxonomy of a South Pacific Island radiation.** *Proc Biol Sci* 2006, **273(1588)**:887-893.
26. Paquin P, Hedin M: **The power and perils of 'molecular taxonomy': a case study of eyeless and endangered *Cicurina* (Araneae: Dictynidae) from Texas caves.** *Mol Ecol* 2004, **13**:3239-3255.
27. Scheffer SJ, Lewis ML, Joshi RC: **DNA barcoding applied to invasive leafminers (Diptera: Agromyzidae) in the Philippines.** *Ann Entomol Soc Am* 2006, **99(2)**:204-210.
28. Lesse H: **Spéciation et variation chromosomiques chez les Lépidoptères Rhopalocères.** *Annls Sci nat, Zool (sér 12)* 1960, **2(1-14)**:1-223.
29. Lukhtanov VA, Kandul NP, Plotkin JB, Dantchenko AV, Haig D, Pierce NE: **Reinforcement of pre-zygotic isolation and karyotype evolution in *Agrodiaetus* butterflies.** *Nature* 2005, **436(7049)**:385-389.
30. Wiemers M: **Chromosome differentiation and the radiation of the butterfly subgenus *Agrodiaetus* (Lepidoptera: Lycaenidae: *Polyommatus*) – a molecular phylogenetic approach.** *phD thesis* 2003:1-198 [http://hss.ulb.uni-bonn.de/diss_online/math_nat_fak/2003/wiemers_martin]. Bonn, University of Bonn
31. Lesse H: **Description de deux nouvelles espèces d'*Agrodiaetus* (Lep. Lycaenidae) séparées à la suite de la découverte de leurs formules chromosomiques.** *Lambillionea* 1957, **57(9/10)**:65-71.
32. Lesse H: **Note sur deux espèces d'*Agrodiaetus* (Lep. Lycaenidae) récemment séparées d'après leurs formules chromosomiques.** *Lambillionea* 1959, **59(1-2)**:5-10.
33. Lesse H: **Les nombres de chromosomes dans la classification du groupe d'*Agrodiaetus ripartii* FREYER (Lepidoptera, Lycaenidae).** *Revue fr Ent* 1960, **27(3)**:240-263.
34. Lesse H: ***Agrodiaetus iphigenia* H.S. et son espèce jumelle *A. tankeri* n. sp. séparées d'après sa formule chromosomique (Lepid. Lycaenidae).** *Bull Soc ent Mulhouse* 1960, **1960**:75-78.
35. Lesse H: **Variation chromosomique chez *Agrodiaetus dolus* HB. (Lep. Lycaenidae).** *Alexanon* 1962, **2**:283-286.
36. Lukhtanov VA, Dantchenko A: **Descriptions of new taxa of the genus *Agrodiaetus* Hübner, [1822] based on karyotype inves-**

- tigation (Lepidoptera, Lycaenidae). *Atalanta* 2002, **33**(1/2):81-107, col. pl. I.
37. Lukhtanov VA, Dantchenko AV: **Principles of the highly ordered arrangement of metaphase I bivalents in spermatocytes of *Agrodiaetus* (Insecta, Lepidoptera).** *Chromosome Research* 2002, **10**(1):5-20.
 38. Lukhtanov VA, Wiemers M, Meusemann K: **Description of a new species of the "brown" *Agrodiaetus* complex from South-East Turkey.** *Nota lepid* 2003, **26**(1/265-71) [http://www.socourlep.com/downloads/pdf_nota_26_065_071.pdf].
 39. Olivier A, Puplesiene J, van der Poorten D, De Prins W, Wiemers M: **Revision of some taxa of the *Polyommatus* (*Agrodiaetus*) *transcaasicus* group with description of a new species from Central Anatolia (Lepidoptera: Lycaenidae).** *Phegea* 1999, **27**(1):1-24.
 40. Lorkovic Z: **The butterfly chromosomes and their application in systematics and phylogeny.** In *Butterflies of Europe Volume 2: Introduction to Lepidopterology*. Edited by: Kudrna O. Wiesbaden , Aula; 1990:332-396.
 41. Wiemers M, De Prins J: ***Polyommatus* (*Agrodiaetus*) *paulae* sp. nov. (Lepidoptera: Lycaenidae) from Northwest Iran, discovered by means of molecular, karyological and morphological methods.** *Entomol Z* 2004, **114**(4):155-162.
 42. Schurian KG: **Zur Biologie, Ökologie und Taxonomie von *Polyommatus* (*Meleageria*) *daphnis brandti* (Pfeiffer, 1938) und *Polyommatus* (*Meleageria*) *daphnis marcida* (Lederer, 1870) aus Nordiran (Lepidoptera: Lycaenidae).** *Entomol Z* 2006, **116**(5):219-225.
 43. **Barcode of Life Data Systems (BOLD)** [<http://www.boldsystems.org/>]
 44. Kandul NP, Lukhtanov VA, Dantchenko AV, Coleman JW, Sekercioglu CH, Haig D, Pierce NE: **Phylogeny of *Agrodiaetus* Hübner 1822 (Lepidoptera: Lycaenidae) inferred from mtDNA sequences of COI and COII and nuclear sequences of EF1-alpha: karyotype diversification and species radiation.** *Syst Biol* 2004, **53**(2):278-298.
 45. Weingartner E, Wahlberg N, Nylin S: **Speciation in *Pararge* (*Satyrinae*: *Nymphalidae*) butterflies – North Africa is the source of ancestral populations of all *Pararge* species.** *Syst Ent* 2006, **31**(4):621-632.
 46. Megens HJ, van Nes WJ, van Moorsel CHM, Pierce NE: **Molecular phylogeny of the Oriental butterfly genus *Arhopala* (Lycaenidae, Theclinae) inferred from mitochondrial and nuclear genes.** *Syst Entomol* 2003, **29**:115-131.
 47. Whinnett A, Zimmermann M, Willmott KR, Herrera N, Mallarino R, Simpson F, Joron M, Lamas G, Mallet J: **Strikingly variable divergence times inferred across an Amazonian butterfly 'suture zone'.** *Proceedings of the Royal Society B* 2005, **272**:2525-2533.
 48. Funk DJ, Omland KE: **Species-level paralogy and polyphyly: Frequency, causes, and consequences, with insights from animal mitochondrial DNA.** *Annu Rev Ecol Syst* 2003, **34**:397-423.
 49. Hickerson MJ, Meyer CP, Moritz C: **DNA barcoding will often fail to discover new animal species over broad parameter space.** *Syst Biol* 2006, **55**(5):729-739.
 50. Megens HJ, van Moorsel CH, Piel WH, Pierce NE, de Jong R: **Tempo of speciation in a butterfly genus from the Southeast Asian tropics, inferred from mitochondrial and nuclear DNA sequence data.** *Mol Phylogenet Evol* 2004, **31**(3):1181-1196.
 51. Sturmbauer C, Meyer A: **Genetic divergence, speciation and morphological stasis in a lineage of African cichlid fishes.** *Nature* 1992, **358**:578-581.
 52. **National Center for Biotechnology Information** [<http://www.ncbi.nlm.nih.gov/>]
 53. **MorphBank** [<http://www.morphbank.net/>]
 54. Als TD, Vila R, Kandul NP, Nash DR, Yen SH, Hsu YF, Mignault AA, Boomsma JJ, Pierce NE: **The evolution of alternative parasitic life histories in large blue butterflies.** *Nature* 2004, **432**(7015):386-390.
 55. Gompert Z, Nice CC, Fordyce JA, Forister ML, Shapiro AM: **Identifying units for conservation using molecular systematics: the cautionary tale of the Karner blue butterfly.** *Mol Ecol* 2006, **15**(7):1759-1768.
 56. Eastwood R, Hughes JM: **Molecular phylogeny and evolutionary biology of *Acrodipsas* (Lepidoptera: Lycaenidae).** *Mol Phylogenet Evol* 2003, **27**(1):93-102.
 57. Eastwood R, Pierce NE, Kitching RL, Hughes JM: **Do ants enhance diversification in Lycaenid butterflies? Phylogeographic evidence from a model myrmecophile, *Jalmenus evagoras*.** *Evolution* 2006, **60**(2):315-327.
 58. Rand DB, Heath A, Suderman T, Pierce NE: **Phylogeny and life history evolution of the genus *Chrysoritis* within the *Aphnaeini* (Lepidoptera: Lycaenidae), inferred from mitochondrial cytochrome oxidase I sequences.** *Mol Phylogenet Evol* 2000, **17**(1):85-96.
 59. Caterino MS, Reed RD, Kuo MM, Sperling FA: **A partitioned likelihood analysis of swallowtail butterfly phylogeny (Lepidoptera: Papilionidae).** *Syst Biol* 2001, **50**(1):106-127.
 60. Vila M, Bjorklund M: **The utility of the neglected mitochondrial control region for evolutionary studies in Lepidoptera (Insecta).** *J Mol Evol* 2004, **58**(3):280-290.
 61. Caterino MS, Sperling FA: ***Papilio* phylogeny based on mitochondrial cytochrome oxidase I and II genes.** *Mol Phylogenet Evol* 1999, **11**(1):122-137.
 62. Hall TA: **BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT.** *Nucl Acids Symp Ser* 1999, **41**:95-98.
 63. Kumar S, Tamura K, Nei M: **MEGA3: Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment.** *Briefings in Bioinformatics* 2004, **5**:150-163.
 64. Nei M, Kumar S: **Molecular Evolution and Phylogenetics.** Oxford , Oxford Univ Press; 2000.
 65. Campbell DL, Brower AV, Pierce NE: **Molecular evolution of the wingless gene and its implications for the phylogenetic placement of the butterfly family Riodinidae (Lepidoptera: Papilionoidea).** *Mol Biol Evol* 2000, **17**(5):684-696.
 66. Eliot JN: **The higher classification of the Lycaenidae (Lepidoptera): a tentative arrangement.** *Bulletin of the British Museum (Natural History) Entomology* 1973, **28**(6):371-505.
 67. Wahlberg N, Braby MF, Brower AV, de Jong R, Lee MM, Nylin S, Pierce NE, Sperling FA, Vila R, Warren AD, Zakharov E: **Synergistic effects of combining morphological and molecular data in resolving the phylogeny of butterflies and skippers.** *Proc Biol Sci* 2005, **272**(1572):1577-1586.
 68. Lukhtanov VA, Vila R: **Rearrangement of the *Agrodiaetus dolus* species group (Lepidoptera, Lycaenidae) using a new cytological approach and molecular data.** *Insect Syst Evol* 2006, **37**:325-334.

Publish with **BioMed Central** and every scientist can read your work free of charge

"BioMed Central will be the most significant development for disseminating the results of biomedical research in our lifetime."

Sir Paul Nurse, Cancer Research UK

Your research papers will be:

- available free of charge to the entire biomedical community
- peer reviewed and published immediately upon acceptance
- cited in PubMed and archived on PubMed Central
- yours — you keep the copyright

Submit your manuscript here:
http://www.biomedcentral.com/info/publishing_adv.asp

