BRILL

Terrestrial Arthropod Reviews 4 (2011) 221–236

TAR
brill.nl/tar

Taxonomy and DNA sequence databases: A perfect match?

John James Wilson

Department of Integrative Biology, University of Guelph, Guelph, Ontario, N1G 2W1, Canada
e-mail: wilso04@gmail.com
Received: 3 May 2011; accepted: 15 June 2011

Summary
Despite the declining number of traditional taxonomists, our knowledge of Earth's biodiversity continues to grow in the form of DNA sequence data. Freely available through online databases, analyses of sequence datasets are increasingly used as an alternative for the traditional taxonomic process. Species identifications have become "DNA barcoding," new species discoveries are characterised by genetic divergences, and traditional classification has been supplanted by molecular phylogenetics. These developments are illustrated through a case study investigating the identities of *Taygetis* butterflies of Costa Rica. Here I review prospects and problems with the molecularization of taxonomy and the key role of publicly available nucleotide sequence databases in efforts to catalogue diversity of life.

Keywords
Taxonomy; systematics; DNA barcoding; species identification; species description; phylogenetics; classification; Lepidoptera; *Taygetis*

Introduction

The wealth of biological data freely available online through the International Nucleotide Sequence Database Collaboration (INSDC) comprising GenBank, EMBL (European Molecular Biology Laboratory) and DDBJ (DNA Data Bank of Japan), represents an unparalleled resource to the field of taxonomy and efforts to catalogue biological diversity. In this article I discuss the advantages and perils associated with three main applications of publicly available DNA sequences in taxonomy: (i) species identifications, (ii) facilitating species discoveries, (iii) a source of characters and/or taxa for phylogenetic reconstructions, and discuss the crucial role they could play in a DNA taxonomy of life. These developments are illustrated through a case study investigating the identities of the *Taygetis* butterflies of Costa Rica.

DOI 10.1163/187498311X591111

Species identifications

The "taxonomic impediment" and general decline of the traditional taxonomic workforce (Gaston and O'Neill, 2004) has resulted in biologists looking to new and inventive methods to identify specimens, whether, for example, for medical, ecological or agricultural purposes (e.g., Weeks et al., 1999; Hebert et al., 2003a). Advances in high throughput DNA sequencing (Shendure et al., 2004) and reductions in costs (Hajibabaei et al., 2005), means BLAST sequence similarity searching (BLAST is an algorithm used to search DNA sequence databases for a "best hit"; Wheeler et al., 2003) and the GenBank database are frequently employed as an identification tool, to determine the species or higher taxonomic grouping of an unknown specimen or sample from which a DNA sequence can be obtained. The use of short DNA sequences for species identification, DNA barcoding (Hebert et al., 2003a), with careful consideration of the gene region used, has been shown to be very effective in a wide variety of taxonomic groups (Hebert et al., 2003b; Waugh, 2007; Floyd et al., 2009). Sequence-based identification could enable species diagnosis from any life history stage and parts of whole organisms (Savolainen et al., 2005). This includes larval stages difficult to identify with traditional methods (Janzen et al., 2005), social insects in which several casts have "unrelated" morphologies (Smith et al., 2005) and historical fragments lacking species-specific morphological features (Noonan et al., 2005). Novel applications have included parasite forensics (De Bruyne et al., 2005), the fight against trade in endangered species (Hsieh et al., 2003), and classification of snake venoms (Pook and McEwing, 2005). It is not always necessary to sequence a DNA fragment from the unknown sample in order to use sequence databases as a tool in species identification. Sequences from GenBank have been used to develop cheaper diagnostic tools such as PCR assays (e.g. Chapman et al., 2003).

Exploitation of species barcodes, analogous with retail barcodes, first requires the assembly of a comprehensive database that links organisms and their sequences (Savolainen et al., 2005). Presently databases have very uneven distributions of sequences among taxa. Intensely studied groups and model organisms (e.g., *Arabidopsis thaliana*, *Drosophila melanogaster*) have many sequences and even entire genomes available. A few genes have been sequenced for many taxa but the vast majority of species diversity is unrepresented (Sanderson et al., 2003). Even amongst mammals diversity is insufficiently covered; ancient rabbit fragments could not be confidently identified due to lack of reliable conspecific reference sequences (Yang et al., 2005). This raises the possibility that inappropriate reference sequences could be applied resulting in spurious species diagnoses (Baker et al., 1996). Best BLAST hit, the simplest method of taxonomic assignment, is "essentially useless" when no relatives have appropriate sequences in GenBank (Tringe and Rubin, 2005). Tautz et al. (2003) suggest a DNA sequence is provided alongside all future taxonomic samples and species descriptions, and the current barcoding initiatives (www.ibol.org) go a long way to bridge the gap (for some major eukaryote groups at least). DNA barcoding currently operates through a system of "reciprocal illumination": improvement of the DNA sequence database through classical taxonomy, but also improvement of classical taxonomy through

flagging potential cryptic diversity and other taxonomic issues via DNA sequence analysis (e.g. Wilson et al., 2010).

Some authors argue GenBank is unsuitable for taxonomic purposes due to lack of provision to include morphological, biogeographical, and ecological information associated with the sequence entry (Tautz et al., 2003). However, the concept of "type sequences" with voucher specimens authenticated by experts on the taxa and with associated taxonomic data is becoming reality. In 2004 NCBI (National Centre for Biotechnology Information), GenBank's operators, sealed a partnership with the Consortium for the Barcode Of Life whereby "barcode standard" DNA sequences with relevant supporting data can now be archived with the INSDC (Figure 1; Hanner, 2005; Savolainen et al., 2005) with the keyword "BARCODE" attached. The BOLD (Barcode of Life Data Systems; www.barcodinglife.org; Ratnasingham and Hebert, 2007) is fast approaching 100,000 formally described species with barcodes (cytochrome *c* oxidase subunit 1 - *COI* - sequences), most associated with images, specimen and collection data, presenting an unparalleled opportunity for rapid, accurate species identification.

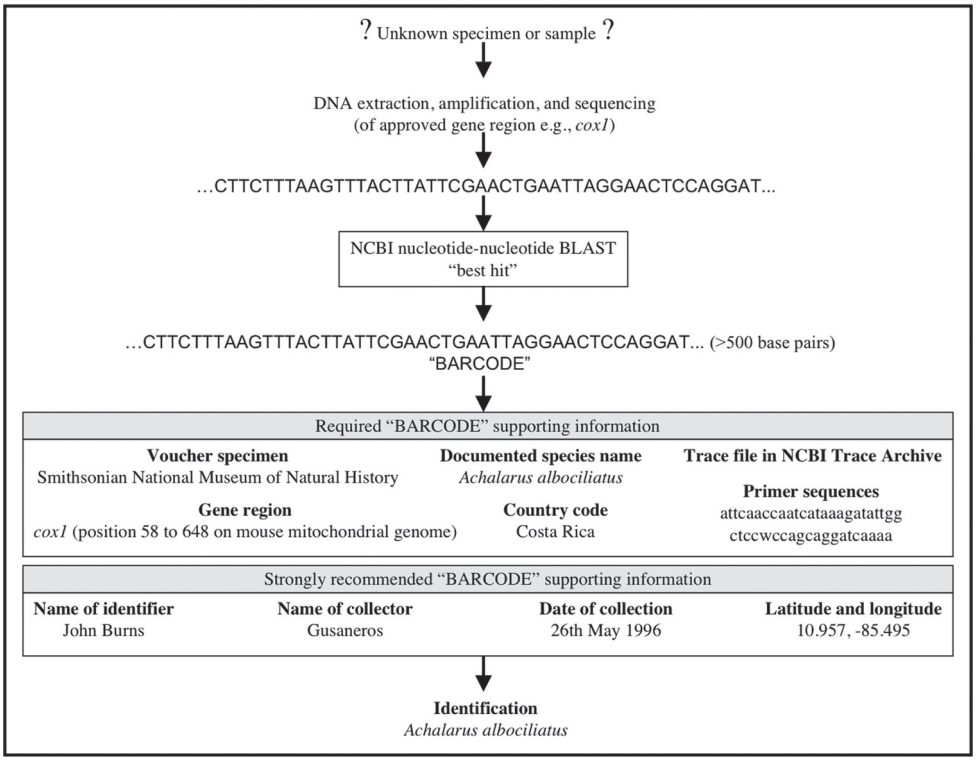

Figure 1. Possible route of species identification using the GenBank database and the criteria for assigning "BARCODE" status in INSDC (*Achalarus albociliatus* data from Hajibabaei et al., 2006).

Species discoveries

Some taxonomists propose a wider use of DNA sequences and see them as facilitating species discoveries based on cluster or phylogenetic analyses of homologous sequences (Monaghan et al., 2005; Savolainen et al., 2005; Pons et al., 2006). Despite a history of cryptic species discoveries within morphological indiscriminate groups via DNA evidence, this application appears much more controversial than species identification (Will et al., 2005; DeSalle, 2006; Nielsen and Matz, 2006). However, many published empirical "barcoding" study have highlighted potential new species, with species delineation at least partially relying on DNA data (e.g., Hebert et al., 2004; Monaghan et al., 2005; Hajibabaei et al., 2006).

For routine specimen identification the BLAST algorithm is currently the only way to search GenBank for "matching" sequences, however, no one would seriously present lack of similar GenBank sequences as conclusive proof of a new undescribed species (Nielsen and Matz, 2006). The exception would be uncultured microbes where sequence similarity within rRNA genes is the accepted standard method of defining species (Venter et al., 2004). Neighbour-joining trees have frequently been used in barcoding analysis of *COI* sequence datasets; justified by barcoders as providing necessary speed of analysis for large datasets typical of barcoding studies, and because "the goal of barcoding is to provide species identification based on sequence similarity rather than to reconstruct deeper phylogenetic relationships accurately" (Ball et al., 2005). The observation of large inter- and low intra-species sequence variation, the "barcoding gap," promises easy identification of pre-existing species (Hebert et al., 2003a; Hebert et al. 2010; but see Wiemers and Fiedler, 2007).

However, there are both theoretical and practical concerns regarding phenetic threshold values (DeSalle et al., 2005; Meyer and Paulay, 2005; Monaghan et al., 2005), and this becomes an even more controversial practice when they are used to delimit new species. When analyzing *COI* sequence datasets, 3% divergence (Kimura 2-parameter) has been cited as sufficient genetic distance to characterize different species (Hebert et al., 2003b). Alternatively, divergences greater than a minimum threshold of one tenth of the average *p*-distance found between well established species in a lineage have been interpreted as indication of the possibility of cryptic species or misdiagnosis of specimens (Hebert et al., 2004; Monaghan et al., 2005; Smith et al., 2006). Meyer and Paulay (2005) determined that for three mollusc datasets a 3% threshold minimizes over-splitting but results in significant over-lumping of species (termed evolutionary significant units). It seems inevitable that taxonomists using phenetic thresholds will have to constantly revise similarity cut-offs from group to group making the delineation of species using distances fairly subjective (DeSalle et al., 2005; Meyer and Paulay, 2005).

Arguably, taxonomists are more willing to accept an evolutionary approach to new species recognition with DNA sequence data. This would be based on the principles of phylogenetic systematics, where evolutionary entities are inferred with shared derived characters (synapomorphies). Suitable methods of quantitative species delimitation might be to search for diagnostic sequence variation, then determine species limits

based on quantitative methods (see Sites and Marshall, 2003; Monaghan et al., 2005). One possibility is statistical parsimony haplotype network analysis (Templeton et al., 1992; Hart and Sunday, 2007). This method subdivides variation in a sequence dataset based on the level of homoplasy within the data themselves providing a relative measure of divergence rather than a pre-determined phenetic cut-off value (Monaghan et al., 2006). Other alternatives being investigated include elegant methodologies incorporating principles of population genetics (Nielsen and Matz, 2006) and an interpretation of branch length as the species level boundary (Monaghan et al., 2005; Pons et al., 2006; Vuataz et al., 2011). The "barcoder problem" (i.e. which decisionary system to employ to delimit species on the basis of DNA barcode sequences; DeSalle et al., 2005) remains an active area of research and discussion (e.g. Lohse, 2009; O'Meara, 2010).

The problem of species delineation is directly related to the "problem" of species concepts, themselves of perennial interest to biologists due to the unique position of species in the taxonomic hierarchy as the only "real" grouping of natural populations. If identification and discovery of "biological" species is the goal of taxonomy, using phenetic DNA distances to determine species boundaries could suffer the same failings inherent with the partial reality that morphological species equate with biological species, implicit in the traditional methods. Selection of appropriate markers and analytical tools to delimit clusters of interbreeding individuals will remain a challenging problem (Savolainen et al., 2005). Modern evolutionary species concepts with emphasis on monophyly or diagnosability (Baum and Donoghue, 1995) may be more compatible with DNA-based species delineation.

Characters and/or taxa for phylogenetic reconstructions

Though many systematists promote an "integrated" taxonomy and a "total evidence" (i.e., a combination of molecular and morphological characters) approach in phylogenetics, molecular data has distinct advantages over morphological characters in phylogenetic reconstructions (see Scotland et al., 2003) and researchers often obtain sequences from databases to supplement their own sequencing work or increase taxonomic coverage. The large numbers of characters available with sequence data increases robustness and support values of the phylogenetic hypothesis and the nature of DNA evolution allows explicit models of evolution to be incorporated into tree building algorithms (Holder and Lewis, 2003). Huge data matrices comprising unambiguous character states and homology assessments are possible with sequence data (Scotland et al., 2003) and phylogenetic hypotheses derived from molecular data can be used with calibrated molecular clocks to estimate divergence timings of taxa (Savolainen and Chase, 2003). However, there are still long recognised problems such as multiple sequence alignment in length variable regions and the phylogenetic treatment of gaps (Wheeler, 1996).

Short DNA sequences available from databases and their ever-increasing taxonomic coverage could become an unprecedented resource for phylogenetics in addition to being a diagnostic taxonomic tool (Savolainen et al., 2005). It would seem

short-sighted to disregard the potential value of short DNA sequences in phylogenetic analysis especially if they are standardized and used in combination to provide multiple independent sources of evidence, for example, both mitochondrial and nuclear loci. Despite concern over the use of DNA barcodes (specifically *COI*) in phylogenetics (Wilson, 2010), recent studies have demonstrated the value of *COI* in resolving deep divergences and report limited incongruence between mitochondrial and nuclear regions (Sihvonen et al. 2011). Dense taxon sampling of databases can limit the problem of missing taxa, a serious concern to accurate recovery of phylogenetic patterns (Wheeler, 2004), as long branches would be broken up in many cases and globally convergent characters become local homologies. With near complete taxon sampling, even short DNA sequences could theoretically resolve even deep phylogenies (see Pollock et al., 2002).

The most informative phylogenetic dataset will include as many taxa and nucleotides as possible while simultaneously limiting the amount of missing data (Yan et al., 2005). An investigation by Sanderson et al. (2003) found datasets with as much as 92% missing data can still provide accurate phylogenetic reconstructions. When different genes have been used extensively in different sections of the tree (i.e., different genes for different taxonomic groups) and only a minimally overlapping set of species is available, a synthetic approach known as "supertree construction" must then be used (Keeling et al. 2005). The technological challenges and processing power required when dealing with datasets from large sequence databases is an area of ongoing research (e.g. Sanderson and Driskell, 2003; Beiko et al., 2005).

Phylogenetic inferences and species membership hypotheses generated from sequence database entries are only as good as the data on which they are based. Poor coverage of diversity in databases has already been discussed above, but what about reliability of the sequences already there? Most journals will require submission of sequences used in research articles to GenBank, although quality control of raw data is often solely dependent on the original scientists (Harris, 2003). Researchers are not obliged to correct existing database entries or use approved nomenclature (taxonomic and gene region), unless as a requirement of journal publication. Consequently, scientists including sequences from databases need to be aware that the quality is not always optimal, and should check unusual sequences in both a phylogenetic and functional context (Harris, 2003). Providing electropherograms to databases alongside the edited sequence read and not relying on computer base calls, opens up the actual sequence trace for scrutiny. Some apparent misidentifications may simply reflect disparities in taxonomy between specialists or knowledge of the systematics of the organism at the time (Bridge et al., 2004).

Case study: identities of the *Taygetis* butterflies of Costa Rica

An extensive ongoing inventory for all lepidopteran species in Area de Conservacion Guanacaste (ACG), Costa Rica was initiated in 2004 and incorporated DNA barcoding of selected specimens (Janzen et al. 2005, 2009; Janzen and Hallwachs, 2009).

When a high intra-specific sequence divergence was observed between two clusters of *Taygetis andromeda* (Cramer, 1776) (3% K2P vs. < 0.5%, the average conspecific K2P distance reported by Hajibabaei et al., 2006), a common satyrine recorded throughout much of South and Central America (DeVries, 1987), this prompted investigations into the identity of the two clusters and initiated a broader review of the taxonomic status of congeneric species known from Costa Rica. Of the roughly 27 species in the genus (Murray, 2001) DeVries (1987) lists 11 in Costa Rica, five of which are known from ACG (Janzen and Hallwachs, 2009).

After literature surveys, examination of images of types specimens, screening publicly available DNA sequences (GenBank) collected for other inventories, and study of diagnostic wing characters (Wilson, 2009), it was determined that of the 11 scientific names for *Taygetis* butterflies used by DeVries (1987), only three are still in valid use: *T. virgilia*, *T. kerea,* and *T. mermeria* (Figure 2). *T. andromeda* was determined to be a complex of two species: *T. laches* and *T. thamyra* (Janzen et al., 2009; Figure 2). Incorporation of DNA barcodes into the species inventory proved effective at flagging cryptic diversity and assisting the resolution of issues in species-level taxonomy (see also Hebert et al., 2004; Vaglia et al., 2007; Floyd et al., 2009; Hausmann et al., 2009; Wilson et al., 2010). Barcodes were also crucial for simply checking identities of sequences used for phylogenetic studies and submitted to GenBank (Regier et al., 2009; Wilson, 2010). Figures 3 and 4 shows the Linnaean names attached to the *Taygetis* sequences on GenBank and the names currently applied by the sequence authors (e.g. Peña, 2011).

Given the rampant instability of the names encountered in this study, it is hard to imagine a situation where combinations will not change, or where previously overlooked synonymies will not become apparent. For example, phylogenetic analysis of DNA barcodes suggests Miller (2004) may have unwittingly created two paraphyletic genera when he followed the suggestion of Freitas (2003) and moved *Taygetis celia* to *Taygetomorpha* (Figure 3). Admittedly the study of *Taygetis* is incomplete, with a narrow taxonomic and geographical focus and many species still have an incomplete chain connecting the "name" to a barcode (Figure 2). However, this may not be cause for alarm. Instead, the barcodes offer new hope as independent taxonomic anchors, "taxonomy objects" *sensu* Vogler and Monaghan (2007) around which to investigate, connect and partition biodiversity.

The future – DNA taxonomy?

Advocates of DNA taxonomy have great ambitions for DNA sequence databases. Their vision is of a universal DNA-based taxonomy across all organismal groups with DNA sequences providing precise, digital descriptions and the "scaffold" for a classification system (Tautz et al., 2003). DNA taxonomy is founded on the reality that evolutionary entities i.e. taxonomic (phylogenetic) groups including species, are equally as distinguishable with DNA sequences as with morphological characters. If this future of taxonomy is to be realized, build-up of sequence databases is essential for the identification and classification of organisms (Savolainen et al., 2005). However, DNA taxonomy is

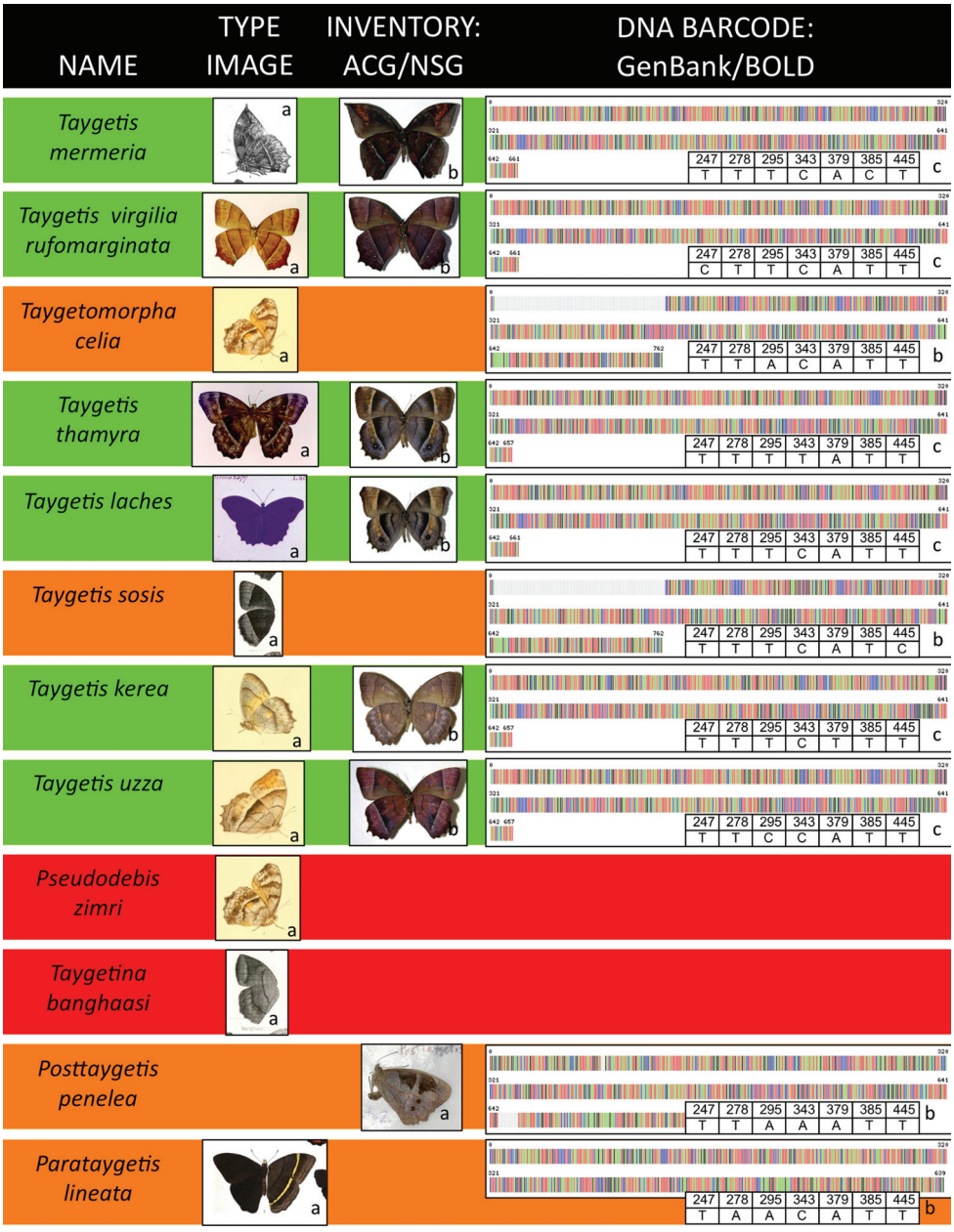

Figure 2. Diagnostic characters for ten *Taygetis* species from Costa Rica. Species of *Taygetis* found in Costa Rica presented in order of DeVries (1987) showing the connection between scientific names locked to type specimens and DNA barcodes through intermediary recently collected specimens. See Wilson (2009) for full details. *Taygetis mermeria*: (a) the type of *Taygetis excavate*; (b) ACG: 04-SRNP-11561; (c) BOLD: MHAAA177-05. *Taygetis virgilia*: (a) *Taygetis virgila rufomarginata* Lectoype 890329-18 (ZMHU); (b) ACG: 03-SRNP-17575; (c) BOLD: MHAAA737-05. *Taygetomorpha celia*: (a) the type of *Taygetis*

Figure 2. (Continued)

keneza; (b) GenBank: AY508572 (*Taygetis celia*). *Taygetis thamyra*: (a) *Taygetis thamyra* Lectotype 091677A-19 (van Lennep collection; BMNH); (b) ACG: 04-SRNP-14279; (c) BOLD: MHAAB021-05. *Taygetis laches*: (a) *Taygetis laches*, painting of the type (Jones Icones); (b) ACG: 03-SRNP-18609; (c) BOLD: MHAAA565-05. *Taygetis sosis*: (a) *Taygetis sosis* from Weymer (1910); (b) GenBank: AY508579. *Taygetis kerea*: (a) the type of *Taygetis kerea*; (b) ACG: 04-SRNP-13986; (c) MHAAB238-05. *Taygetis uzza:* (a) the type of *Taygetis uzza*; (b) ACG: 06-SRNP-44842; (c) BOLD: MHAAC383-07. *Pseudodebis zimri:* (a) the type of *Pseudodebis zimri*. *Taygetina banghaasi:* (a) the type of *Taygetina banghaasi*; *Posttaygetis penelea:* (a) NSG: NW126-13; (b) GenBank: DQ338813. *Parataygetis lineate:* (a) the type of *Parataygetis lineata*; (b) GenBank: AY508569. Information on the specimens pictured can be found by searching the specimen ID number at http://janzen.sas.upenn.edu (ACG) or http://nymphalidae.utu.fi/db.php (NSG). Information on the DNA barcodes can be found by searching the sequence ID at http://www.boldsystems.org (BOLD) or http://www.ncbi.nlm.nih.gov/nucleotide (GenBank). Seven nucleotide positions which in combination can distinguish the ten species, with barcodes, from each other are presented in boxes next to the illustrative barcode. This is very sensitive to the number of specimens analyzed – and the fewer specimens incorporated, the greater the likelihood that a rare haplotype is not reflected in the data. I present these diagnostics not to suggest that the coverage of each species is sufficient to reflect the variation within a species, but rather to demonstrate that such an analysis is possible within a region, when there is good representation of the variability within a species. The nucleotide position given refers to the barcode region, and can be compared to their full mitochondrial position by adding 48 (as aligned to the *Bos taurus* complete mitochondrial genome sequence Genbank ref AY676873). The colours surrounding each species refers to 1) an unbroken chain connecting the scientific name to a DNA barcode (GREEN), 2) a broken chain connecting the scientific name to a barcode (AMBER) or 3) no chain yet available to connect the scientific name to a DNA barcode (RED).

firmly rooted in the long-standing principles of traditional taxonomic work and reliant on names originating from previous classifications, otherwise new phylogenetic groupings would be incomprehensible to us (Franz, 2005). Vouchering specimens, including types and with the addition of DNA extractions, remains just as important a component as in the traditional approach.

Sequence information is easy to obtain, unambiguous, and makes species identification possible by non-specialists unfamiliar with the intricacies of morphology (Hebert et al., 2003a). Molecular operational taxonomic units (Blaxter, 2004), good species or not depending on the species concept applied, are nevertheless a good surrogate for identifying units of diversity in biodiversity studies. This enables users to obtain the information much faster than with the traditional morphological taxonomic process making surveys scalable across much larger taxonomic groupings and wider geographical regions (Smith et al., 2005) including correlation of "morphospecies" across diverse locations. It has taken two centuries to describe 1.7 million species using the traditional taxonomic approach. DNA-based species discovery has the potential to rapidly accelerate this process, an advantage which cannot be ignored in the light of the current biodiversity crisis affecting our planet. DNA sequences are already the *de facto* universal communication tool, providing data points for information about taxonomic

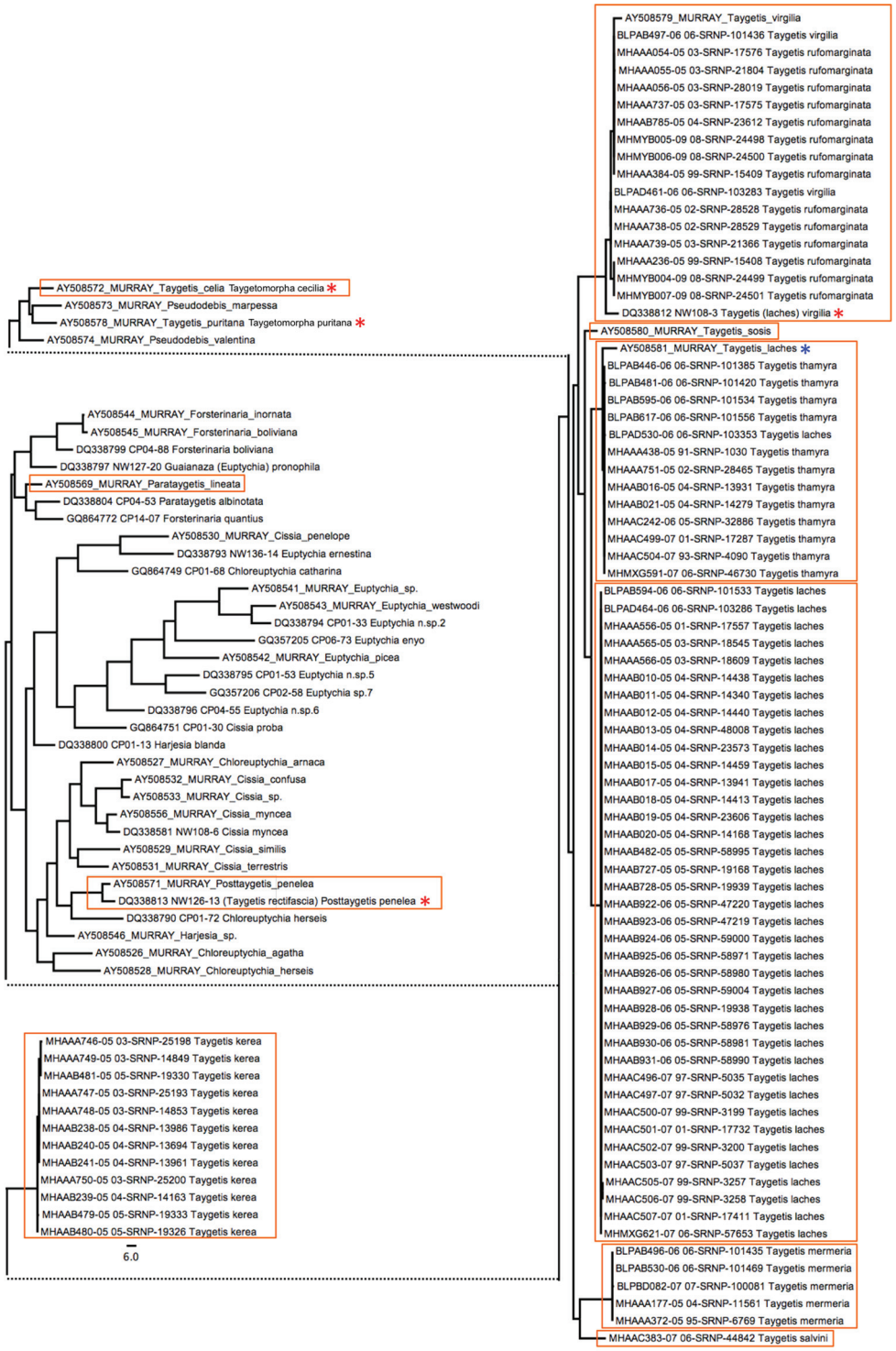
AY508572_MURRAY_Taygetis_celia Taygetomorpha cecilia
AY508573_MURRAY_Pseudodebis_marpessa
AY508578_MURRAY_Taygetis_puritana Taygetomorpha puritana
AY508574_MURRAY_Pseudodebis_valentina
AY508544_MURRAY_Forsterinaria_inornata
AY508545_MURRAY_Forsterinaria_boliviana
DQ338799 CP04-88 Forsterinaria boliviana
DQ338797 NW127-20 Guaianaza (Euptychia) pronophila
AY508569_MURRAY_Parataygetis_lineata
DQ338804 CP04-53 Parataygetis albinotata
GQ864772 CP14-07 Forsterinaria quantius
AY508530_MURRAY_Cissia_penelope
DQ338793 NW136-14 Euptychia ernestina
GQ864749 CP01-68 Chloreuptychia catharina
AY508541_MURRAY_Euptychia_sp.
AY508543_MURRAY_Euptychia_westwoodi
DQ338794 CP01-33 Euptychia n.sp.2
GQ357205 CP06-73 Euptychia enyo
AY508542_MURRAY_Euptychia_picea
DQ338795 CP01-53 Euptychia n.sp.5
GQ357206 CP02-58 Euptychia sp.7
DQ338796 CP04-55 Euptychia n.sp.6
GQ864751 CP01-30 Cissia proba
DQ338800 CP01-13 Harjesia blanda
AY508527_MURRAY_Chloreuptychia_arnaca
AY508532_MURRAY_Cissia_confusa
AY508533_MURRAY_Cissia_sp.
AY508556_MURRAY_Cissia_myncea
DQ338581 NW108-6 Cissia myncea
AY508529_MURRAY_Cissia_similis
AY508531_MURRAY_Cissia_terrestris
AY508571_MURRAY_Posttaygetis_penelea
DQ338813 NW126-13 (Taygetis rectifascia) Posttaygetis penelea
DQ338790 CP01-72 Chloreuptychia herseis
AY508546_MURRAY_Harjesia_sp.
AY508526_MURRAY_Chloreuptychia_agatha
AY508528_MURRAY_Chloreuptychia_herseis
MHAAA746-05 03-SRNP-25198 Taygetis kerea
MHAAA749-05 03-SRNP-14849 Taygetis kerea
MHAAB481-05 05-SRNP-19330 Taygetis kerea
MHAAA747-05 03-SRNP-25193 Taygetis kerea
MHAAA748-05 03-SRNP-14853 Taygetis kerea
MHAAB238-05 04-SRNP-13986 Taygetis kerea
MHAAB240-05 04-SRNP-13694 Taygetis kerea
MHAAB241-05 04-SRNP-13961 Taygetis kerea
MHAAA750-05 03-SRNP-25200 Taygetis kerea
MHAAB239-05 04-SRNP-14163 Taygetis kerea
MHAAB479-05 05-SRNP-19333 Taygetis kerea
MHAAB480-05 05-SRNP-19326 Taygetis kerea
6.0
AY508579_MURRAY_Taygetis_virgilia
BLPAB497-06 06-SRNP-101436 Taygetis virgilia
MHAAA054-05 03-SRNP-17576 Taygetis rufomarginata
MHAAA055-05 03-SRNP-21804 Taygetis rufomarginata
MHAAA056-05 03-SRNP-28019 Taygetis rufomarginata
MHAAA737-05 03-SRNP-17575 Taygetis rufomarginata
MHAAB785-05 04-SRNP-23612 Taygetis rufomarginata
MHMYB005-09 08-SRNP-24498 Taygetis rufomarginata
MHMYB006-09 08-SRNP-24500 Taygetis rufomarginata
MHAAA384-05 99-SRNP-15409 Taygetis rufomarginata
BLPAD461-06 06-SRNP-103283 Taygetis virgilia
MHAAA736-05 02-SRNP-28528 Taygetis rufomarginata
MHAAA738-05 02-SRNP-28529 Taygetis rufomarginata
MHAAA739-05 03-SRNP-21366 Taygetis rufomarginata
MHAAA236-05 99-SRNP-15408 Taygetis rufomarginata
MHMYB004-09 08-SRNP-24499 Taygetis rufomarginata
MHMYB007-09 08-SRNP-24501 Taygetis rufomarginata
DQ338812 NW108-3 Taygetis (laches) virgilia
AY508580_MURRAY_Taygetis_sosis
AY508581_MURRAY_Taygetis_laches
BLPAB446-06 06-SRNP-101385 Taygetis thamyra
BLPAB481-06 06-SRNP-101420 Taygetis thamyra
BLPAB595-06 06-SRNP-101534 Taygetis thamyra
BLPAB617-06 06-SRNP-101556 Taygetis thamyra
BLPAD530-06 06-SRNP-103353 Taygetis laches
MHAAA438-05 91-SRNP-1030 Taygetis thamyra
MHAAA751-05 02-SRNP-28465 Taygetis thamyra
MHAAB016-05 04-SRNP-13931 Taygetis thamyra
MHAAB021-05 04-SRNP-14279 Taygetis thamyra
MHAAC242-06 05-SRNP-32886 Taygetis thamyra
MHAAC499-07 01-SRNP-17287 Taygetis thamyra
MHAAC504-07 93-SRNP-4090 Taygetis thamyra
MHMXG591-07 06-SRNP-46730 Taygetis thamyra
BLPAB594-06 06-SRNP-101533 Taygetis laches
BLPAD464-06 06-SRNP-103286 Taygetis laches
MHAAA556-05 01-SRNP-17557 Taygetis laches
MHAAA565-05 03-SRNP-18545 Taygetis laches
MHAAA566-05 03-SRNP-18609 Taygetis laches
MHAAB010-05 04-SRNP-14438 Taygetis laches
MHAAB011-05 04-SRNP-14340 Taygetis laches
MHAAB012-05 04-SRNP-14440 Taygetis laches
MHAAB013-05 04-SRNP-48008 Taygetis laches
MHAAB014-05 04-SRNP-23573 Taygetis laches
MHAAB015-05 04-SRNP-14459 Taygetis laches
MHAAB017-05 04-SRNP-13941 Taygetis laches
MHAAB018-05 04-SRNP-14413 Taygetis laches
MHAAB019-05 04-SRNP-23606 Taygetis laches
MHAAB020-05 04-SRNP-14168 Taygetis laches
MHAAB482-05 05-SRNP-58995 Taygetis laches
MHAAB727-05 05-SRNP-19168 Taygetis laches
MHAAB728-05 05-SRNP-19939 Taygetis laches
MHAAB922-06 05-SRNP-47220 Taygetis laches
MHAAB923-06 05-SRNP-47219 Taygetis laches
MHAAB924-06 05-SRNP-59000 Taygetis laches
MHAAB925-06 05-SRNP-58971 Taygetis laches
MHAAB926-06 05-SRNP-58980 Taygetis laches
MHAAB927-06 05-SRNP-59004 Taygetis laches
MHAAB928-06 05-SRNP-19938 Taygetis laches
MHAAB929-06 05-SRNP-58976 Taygetis laches
MHAAB930-06 05-SRNP-58981 Taygetis laches
MHAAB931-06 05-SRNP-58990 Taygetis laches
MHAAC496-07 97-SRNP-5035 Taygetis laches
MHAAC497-07 97-SRNP-5032 Taygetis laches
MHAAC500-07 99-SRNP-3199 Taygetis laches
MHAAC501-07 01-SRNP-17732 Taygetis laches
MHAAC502-07 99-SRNP-3200 Taygetis laches
MHAAC503-07 97-SRNP-5037 Taygetis laches
MHAAC505-07 99-SRNP-3257 Taygetis laches
MHAAC506-07 99-SRNP-3258 Taygetis laches
MHAAC507-07 01-SRNP-17411 Taygetis laches
MHMXG621-07 06-SRNP-57653 Taygetis laches
BLPAB496-06 06-SRNP-101435 Taygetis mermeria
BLPAB530-06 06-SRNP-101469 Taygetis mermeria
BLPBD082-07 07-SRNP-100081 Taygetis mermeria
MHAAA177-05 04-SRNP-11561 Taygetis mermeria
MHAAA372-05 95-SRNP-6769 Taygetis mermeria
MHAAC383-07 06-SRNP-44842 Taygetis salvini

Figure 3. *COI* MP tree for *Taygetis* and allies. A single most parsimonious cladogram (Tree length 1230, CI 0.321, RI 0.712) was found by a heuristic search of a datamatrix containing 650bp *COI* "DNA barcodes" from *Taygetis* species and allies available on GenBank or BOLD. Species of *Taygetis* listed in DeVries (1987) and present on the tree are highlighted in orange boxes. All species represented by more than one sequence are "monophyletic" on the cladogram. Where sequences are labelled with two names (marked with a red star) these refer to the names attached to the sequences on GenBank and the names currently applied by the sequence authors. In one case (blue star) it was impossible to confirm the identity of the specimen due to images of the specimen not being made available by the sequence author.

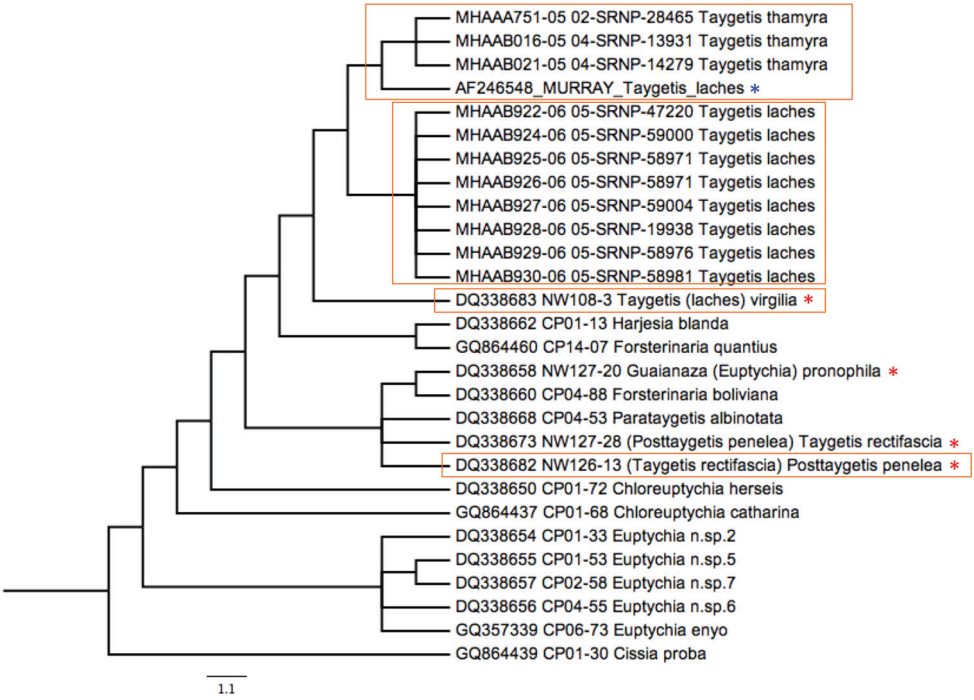

Figure 4. *Wingless* MP consensus tree for *Taygetis* and allies. The tree is a strict consensus tree of 29 most parsimonious cladograms (Tree length 386, CI 0.630, RI 0.701) found by a heuristic search of a datamatrix containing *wingless* gene fragments (400bp) from *Taygetis* species and allies available on GenBank or newly generated for this study. Species of *Taygetis* listed in DeVries (1987) and present on the tree are highlighted in orange boxes. The two species represented by more than one sequence are "monophyletic" on the cladogram. Where sequences are labelled with two names (marked with a red star) these refer to the names attached to the sequences on GenBank and the names currently applied by the sequence authors. In one case (blue start) it was impossible to confirm the identity of the specimen due to images of the specimen not made available by the sequence author.

specimens and species, in a way that complicated, mostly incomprehensible and unobtainable morphological descriptions can never be.

To some extent DNA taxonomy is already reality above the species level with systematists predominantly employing sequence data (and sequence databases) in phylogenetic reconstructions. The failure of researchers to translate molecular phylogenetic analyses into traditional classifications, and whether this is in fact necessary and useful, is discussed by Franz (2005). Sequence data collected for lots of different kinds of studies and purposes and submitted to sequence databases can be drafted into attempts to reconstruct the Tree of Life (Driskell et al., 2004).

In my view, taxonomy and DNA sequence databases are a good match for the following compelling reasons; DNA sequences are unambiguous digital data not influenced by subjective assessment and open to repeated analysis and testing of the species and phylogenetic hypotheses generated. Once submitted to online databases nucleotide sequences represent a freely available taxonomic resource that democratizes the way species are recognized and opens up biodiversity to all users.

Acknowledgements

Bob Hanner (University of Guelph) helped improve an earlier version of this manuscript. Dan Janzen, Winnie Hallwachs (University of Pennsylvania), Lee Miller, Jackie Miller (Florida Museum of Natural History), Mehrdad Hajibabaei and Paul Hebert (University of Guelph) are co-authors of the case study - We thank parataxonomists for finding and rearing the specimens from the ACG, the BOLD team (Megan Milton and Kara Layton) and CCDB laboratory staff (Suresh Naik) for technical support, Heather Braid (University of Guelph) for help while learning how to create a scratchpad, and Niels Kristensen (Denmark), John Chainey (BMNH), Andre Freitas (Brazil) and Niklas Wahlberg (University of Turku) for help with taxonomic investigations.

References

Baker, C. S, F. Cipriano, and S. R. Palumbi. 1996. Molecular genetic identification of whales and dolphin products from commercial markets in Korea and Japan. Molecular Ecology 5:671-685.

Ball, S. L., P. D. N. Hebert, S. K. Burian, and J. M. Webb. 2005. Biological identifications of mayflies (Ephemeroptera) using DNA barcodes. Journal of the North American Benthological Society 24:508-524.

Baum, D. A. and M. J. Donoghue. 1995. Choosing among alternative “phylogenetic” species concepts. Systematic Botany 20:560-573.

Beiko, R. G., T. J. Harlow, and M. A. Ragan. 2005. Highways of gene sharing in prokaryotes. Proceedings of the National Academy of Science USA 102:14332-14337.

Blaxter, M. L. 2004. The promise of a DNA taxonomy. Philosophical Transactions of the Royal Society of London B 359:669-679.

Bridge, P. D., B. M. Spooner, and P. J. Roberts. 2004. Reliability and use of published sequence data. The New Phytologist 161:15-17.

Chapman, D. D., D. L. Ambercrombie, C. J. Dovady, E. K. Pikitch, M. J. Stanhope, and M. S. Shivji. 2003. A streamlined, bi-organelle, multiplex approach to species identification: application to global

conservation and trade monitoring of the great white shark, *Carcharodon carcharias*. Conservation Genetics 4:415-425.

Cramer, P. 1775-1776. De Uitlandsche Kapellen voorkomende in de drie waereld-deelen Asia, Africa en America. Papillons exotiques des trois parties du monde l'Asie, l'Afrique et l'Amerique. Baalde, S. J., Amsteldam & Wild, B., J. Van Schoonhoven, and Comp, Utrecht. Volume 1. [vi], xxx, 16 pp., 155 pp.

De Bruyne, A., H. U. Year, F. Guerhier, P. Boireau, and J. Dupouy-Camet. 2005. Simple species identification of *Trichinella* isolates by amplification and sequencing of the 5S ribosomal DNA intergenic spacer region. Veterinary Parasitology 132:57-61.

DeSalle, R., M. G. Egan, and M. Sidall. 2005. The unholy trinity: taxonomy, species delimitation and DNA barcoding. Philosophical Transactions of the Royal Society of London B 360:1905-1916.

DeSalle, R. 2006. Species discovery versus species identification in DNA barcoding efforts: response to Rubinoff. Conservervation Biology 20(5):1545-1547.

DeVries, P. J. 1987. The Butterflies of Costa Rica and their Natural History. Volume I. Papilionidae, Pieridae and Nymphalidae. Princeton University Press, New Jersey, USA. 288 pp.

Driskell, A. C., C. Ane, J. G. Burleigh, M. M. McMahon, B. C. O'Meara, and M. L. Sanderson. 2004. Prospects for building the tree of life from large sequence databases. Science 306:1172-1174.

Floyd, R., J. J. Wilson, and P. D. N. Hebert. 2009. DNA barcodes and insect biodiversity. pp. 417-431. *In*, R. G. Footit, and P. H. Adler (Editors). Insect Biodiversity: Science and Society. Blackwell Publishing, Oxford, United Kingdom. 632 pp.

Franz, N. M. 2005. On the lack of good scientific reasons for the growing phylogeny/classification gap. Cladistics 21:495-500.

Freitas, A. V. L. 2003. Description of a new genus for "*Euptychia*" *peculiaris* (Nymphalidae: Satyrinae): Immature stages and systematic position. Journal of the Lepidopterists' Society 57:100-106.

Gaston, K.J. and M. A. O'Neill. 2004. Automated species identification: why not? Philosophical Transactions of the Royal Society of London B 359:655-667.

Hajibabaei, M., J. R. deWaard, N. V. Ivanova, S. Ratnasingham, R. T. Dooh, S. L. Kirk, P. M. Mackie, and P. D. N. Hebert. 2005. Critical factors for assembling a high volume of DNA barcodes. Philosophical Transactions of the Royal Society of London B 360:1959-1967.

Hajibabaei, M, D. H. Janzen, J. M. Burns, W. Hallwachs, and P. D. N. Hebert. 2006 DNA barcodes distinguish species of tropical Lepidoptera. Proceedings of the National Academy of Science USA 103:968-971.

Hanner, R. 2005. Proposed standards for BARCODE records in INSDC (BRIs). Available from http://barcoding.si.edu/PDF?DWG_data_standards-Final.pdf (Accessed 6 April 2006).

Harris, D. J. 2003. Can you bank on GenBank? Trends in Ecology and Evolution 18:317-319.

Hart, M. W. and J. Sunday. 2007 Things fall apart: biological species form unconnected parsimony networks. Biology Letters 3:509-512.

Hausmann, A., P. D. N. Hebert, A. Mitchell, R. Rougerie, M. Sommerer, and C. J. Young. 2009. Revision of the Australian *Oenochroma vinaria* Guenée, 1858 species-complex (Lepidoptera, Geometridae, Oenochrominae): DNA barcoding reveals cryptic diversity and assesses status of type specimen without dissection. Zootaxa 2239:1-21.

Hebert, P. D. N., A. Cywinska, and S. L. Ball. 2003a. Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B 270:313-321.

Hebert, P. D. N., S. Ratnasingham, and J. R. deWaard. 2003b. Barcoding animal life: cytochrome c oxidase subunit 1 divergence among closely related species. Proceedings of the Royal Society of London B 270:S96–S99.

Hebert, P. D. N., E. H. Penton, J. M. Burns, D. H. Janzen, and W. Hallwachs. 2004. Ten species in one: DNA barcoding reveals cryptic species in the Neotropical skipper butterfly *Astraptes fulgerator*. Proceedings of the National Academy of Science USA 101:14812-14817.

Hebert, P. D. N., J. R. deWaard and J.-F. Landry. 2010. DNA barcodes for 1/1000 of the animal kingdom. Biology Letters 6:359-362.

Holder, M. and P. O. Lewis. 2003. Phylogeny estimation: traditional and Bayesian approaches. Nature Reviews Genetics 4:275-284.

Hsieh, H. M., L. H. Huang, L. C. Tsai, Y. C. Kuo, H. H. Meng, A. Linacre, and J. C. I. Lee. 2003. Species identification of rhinoceros horns using the cytochrome b gene. Forensic Science International 136:1-11.

Janzen, D. H. and W. Hallwachs. 2009. Dynamic database for an inventory of the macrocaterpillar fauna, and its food plants and parasitoids, of Area de Conservacion Guanacaste (ACG), northwestern Costa Rica (nn-SRNP-nnnnn voucher codes). Available from: http://janzen.sas.upenn.edu/ (Accessed 23 March 2011).

Janzen, D. H., M. Hajibabaei, J. M. Burns, W. Hallwachs, E. Remigio, and P. D. N. Hebert. 2005. Wedding biodiversity inventory of a large and complex Lepidoptera fauna with DNA barcoding. Philosophical Transactions of the Royal Society of London B 360:1835-1845.

Janzen, D. H., W. Hallwachs, P. Blandin, J. M. Burns, J.-M. Cadiou, I. Chacon, T. Dapkey, A. Deans, M. Epstein, B. Espinoza, J. Franclemont, W. Haber, M. Hajibabaei, J. Hallwachs, P. D. N. Hebert, I. D. Gauld, D. Harvey, A. Hausmann, I. Kitching, D. Lafontaine, J.-F. Landry, C. Lemaire, J. Miller, J. Miller, L. Miller, S. E. Miller, J. Montero, E. Munroe, S. Green, J. Rawlins, R. Robbins, J. Rodriguez, R. Rougerie, M. Sharkey, M. A. Smith, M. A. Solis, B. Sullivan, P. Thiaucourt, D. Wahl, S. Weller, J. Whitfield, K. Willmott, D. M. Wood, N. Woodley, and J. J. Wilson. 2009. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. Molecular Ecology Resources 9:1-25.

Keeling, P. J., G. Burger, D. G. Durnford, B. F. Lang, R. W. Lee, R. E. Pearlman, A. J. Roger, and M. W. Gray. 2005. The tree of eukaryotes. Trends in Ecology and Evolution 20:670-676.

Lohse, K. 2009. Can mtDNA barcodes be used to delimit species? a response to Pons et al. (2006). Systematic Biology 58:439-442.

Meyer, C. P. and G. Paulay. 2005. DNA barcoding: error rates based on comprehensive sampling. PLoS Biology 3:2229-2238.

Miller, L. 1978. Notes and descriptions of Euptychiini (Lepidoptera, Satyridae) from the Mexican Region. Journal of the Lepidopterists' Society 32(2):75-85.

Monaghan, M. T., M. Balke, T. R. Gregory, and A. P. Vogler. 2005. DNA-based species delineation in tropical beetles using mitochondrial and nuclear markers. Philosophical Transactions of the Royal Society of London B 360:1925-1933.

Monaghan, M. T., M. Balke, J. Pons, and A. P. Vogler. 2006. Beyond barcodes: complex DNA taxonomy of a South Pacific island radiation. Proceedings of the Royal Society of London B. 273:887-893.

Murray, D. and D. P. Prowell. 2005. Molecular phylogenetics and evolutionary history of the neotropical satyrine subtribe Euptychiina (Nymphalidae: Satyrinae). Molecular Phylogenetics and Evolution 34:67-80.

Nielsen, R and M. Matz. 2006. Statistical approaches for DNA barcoding. Systematic Biology 55:162-169.

Noonan, J. P., M. Hofreiter, D. Smith, J. R. Priest, N. Rohland, G. Rabeder, J. Krause, J. C. Detter, S. Paabo, and E. M. Rubin. 2005. Genomic sequencing of Pleistocene cave bears. Science 309:597-600.

O'Meara, B. C. 2010. New heuristic methods for joint species delimitation and species tree inference. Systematic Biology 59:59-73.

Peña, C. 2011. The NSG Voucher Specimen Database. Available from http://nymphalidae.utu.fi/Vouchers.htm (Accessed 11 April 2011).

Pollock, D. D., D. J. Zwickl, J. A. McGuire, and D. M. Hillis. 2002. Increased taxon sampling is advantageous for phylogenetic inference. Systematic Biology 51:664-671.

Pons, J., T. G. Barraclough, J. Gomez-Zurita, A. Cardosa, D. P. Duran, S. Hazell, S. Kamoun, W. D. Sumlin and A. P. Vogler. 2006. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. Systematic Biology 55:595-609.

Pook, C. E. and R. McEwing. 2005. Mitochondrial DNA sequences from dried snake venom: a DNA barcoding approach to the identification of venom samples. Toxicon 46:711-715.

Ratnasingham, S. and P. D. N. Hebert. 2007. BOLD: The Barcode of Life Data System. Molecular Ecology Notes 7:355-364.

Regier, J. C., A. Zwick, M. P. Cummings, A. Y. Kawahara, S. Cho, S. Weller, A. Roe, J. Baixeras-Almela, J. Brown, C. Parr, D. Davis, M. Epstein, W. Hallwachs, A. Hausmann, D. Janzen, I. Kitching, A. Solis, S.-H. Yen, A. Bazinet, and C. Mitter. 2009. Toward reconstructing the evolution of advanced moths and butterflies (Lepidoptera: Ditrysia): an initial molecular study. BMC Evolutionary Biology 9:280.

Sanderson, M. J. and A. C. Driskell. 2003. The challenge of constructing large phylogenetic trees. Trends in Plant Science 8:374-379.

Sanderson, M. J., A. C. Driskell, R. H. Ree, O. Eulenstein, and S. Langley. 2003. Obtaining maximal concatenated phylogenetic data sets from large sequence databases. Molecular Biology and Evolution 20:1036-1042.

Savolainen, V. and M. W. Chase. 2003. A decade of progress in plant molecular phylogenetics. Trends in Genetics 19:717-724.

Savolainen, V., R. S. Cowan, A. P. Vogler, G. K. Roderick, and R. Lane. 2005. Towards writing the encyclopedia of life: an introduction to DNA barcoding. Philosophical Transactions of the Royal Society of London B 360:1805-1811.

Scotland, R. W., R. G. Olmstead, and J. R. Bennett. 2003. Phylogeny reconstruction: the role of morphology. Systematic Biology 52:539-548.

Shendure, J., R. D. Mitra, C. Varma, and G. M. Church. 2004. Advanced sequencing technologies: methods and goals. Nature Reviews Genetics 5:335-344.

Sihvonen, P., M. Mutanen, L. Kaila, G. Brehm, A. Hausmann and H. S. Staude. 2011. Comprehensive molecular sampling yields a robust phylogeny for geometrid moths (Lepidoptera: Geometridae). PLoS One 6(6):e20356.

Sites, J. W. and J. C. Marshall. 2003. Delimiting species: a renaissance issue in systematic biology. Trends in Ecology and Evolution 18:462-470.

Smith, M. A., B. L. Fisher, and P. D. N. Hebert. 2005. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. Philosophical Transactions of the Royal Society of London B 360:1825-1834.

Smith, M. A., N. E. Woodley, D. H. Janzen, W. Hallwachs, and P. D. N. Hebert. 2006. DNA barcodes reveal cryptic host-specificity within the presumed polyphagous members of a genus of parasitoid flies (Diptera: Tachinidae). Proceedings of the National Academy of Science USA 103:3657-3662.

Tautz, D., P. Arctander, A. Minelli, R. H. Thomas, and A. P. Vogler. 2003. A plea for DNA taxonomy. Trends in Ecology and Evolution 18:70-74.

Templeton, A. R., K. A. Crandall, and C. F. Sing. 1992. A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and sequencing data. III. Cladogram estimation. Genetics 132:619-633.

Tringe, S. G. and E. M. Rubin. 2005. Metagenomics: DNA sequencing of environmental samples. Nature Reviews Genetics 6:805-814.

Vaglia, T., J. Haxaire, I. J. Kitching, I. Meusnier, and R. Rougerie. 2008. Morphology and DNA barcoding reveal three cryptic species within the *Xylophanes neoptolemus* and *loelia* species-group (Lepidoptera: Sphingidae). Zootaxa 1923:18-36.

Venter, J. C., K. Remington, J. F. Heidelberg, A. L. Halpern, D. Rusch, J. A. Eisen, D. Wu, I. Paulsen, K. E. Nelson, W. Nelson, D. E. Fouts, S. Levy, A. H. Knap, M. W. Lomas, K. Nealson, O. White, J. Peterson, J. Hoffman, R. Parsons, H. Baden-Tillson, C. Pfannkoch, Y.-H. Rogers, and H. O. Smith. 2004. Environmental genome shotgun sequencing of the Sargasso Sea. Science 304:66-74.

Vogler, A. P., and M. T. Monaghan. (2007) Recent advances in DNA taxonomy. Journal of Zoological Systematics and Evolutionary Research 45:1-10.

Vuataz, L., M. Sartori, A. Wagner and M. T. Monaghan. 2011. Toward a DNA taxonomy of alpine *Rhithrogena* (Ephemeroptera: Heptageniidae) using a mixed yule-coalescent analysis of mitochondrial and nuclear DNA. PLoS ONE 6(5):e19728.

Waugh, J. 2007. DNA barcoding in animal species: progress, potential and pitfalls. BioEssays 29:188-197.

Weeks, P. J. D., M. A. O'Neill, K. J. Gaston, and I. D. Gauld. 1999. Automating insect identification: exploring the limitations of a prototype system. Journal of Applied Entomology 123:1-8.

Weymer, G. 1910. Satyridae. pp. 184-192, pl. 44-46. *In*, A. Seitz (Editor). Gross-Schmettlinge der Erde. A. Kernen, Stuttgart, Germany.

Wheeler, D. L., T. Barrett, D. A. Benson, S. H. Bryant, K. Canese, V. Chetvernin, D. M. Church, M. DiCuccio, R. Edgar, S. Federhen, L. Y. Geer, W. Helmberg, Y. Kapustin, D. L. Kenton, O. Khovayko, D. J. Lipman, T. L. Madden, D. R. Maglott, J. Ostell, K. D. Pruitt, G. D. Schuler, L. M. Schrimi, E. Sequeira, S. T. Sherry, K. Sirotkin A. Souvorov, G. Starchenko, T. O. Suzek, R. Tatusov, T. A. Tatusova, L. Wagner, and E. Yaschenko. 2003. Database resources of the National Center for Biotechnology. Nucleic Acid Research 31:28-33.

Wheeler, Q. D., P. H. Raven, and E. O. Wilson. 2004. Taxonomic triage and the poverty of phylogeny. Philosophical Transactions of the Royal Society of London B 359:571-583.

Wheeler, W. 1996. Optimization alignment: the end of multiple sequence alignment in phylogenetics? Cladistics 12:1-9.

Wiemers, M. and K. Fiedler (2007) Does the barcoding gap exist? – a case study in blue butterflies (Lepidoptera: Lycaenidae). Frontiers in Zoology 4:8.

Wilson, J. J. 2009. *Taygetis* – Identities of the "*Taygetis*" butterflies of Costa Rica. Available at http://taygetis.myspecies.info (Accessed 11 April 2011).

Wilson, J. J. 2010. Assessing the value of DNA barcodes and other priority gene regions for molecular phylogenetics of Lepidoptera. PLoS ONE 5(5):e10525.

Wilson, J. J., J.-F. Landry, D. H. Janzen, W. Hallwachs, V. Nazari, M. Hajibabaei, and P. D. N. Hebert. 2010. Identity of the ailanthus webworm moth, a complex of two species: evidence from DNA barcoding, morphology and ecology. Zookeys 46:41-60.

Yan, C. H., J. G. Burleigh, and O. Eulenstein. 2005. Identifying optimal incomplete phylogenetics data sets from sequence databases. Molecular Phylogenetics and Evolution 35:528-535.

Yang, D. Y., J. R. Woiderski, and J. C. Driver. 2005. DNA analysis of archaeological rabbit remains from the American southwest. Journal of Archaeological Science 32:567-578.