DNA barcoding makes an impact on the other side of the globe

Muhammad Ashfaq

Morphological identifications of organisms have always been a challenge (Hey 2001). Use of DNA sequences for identification and determination of phylogenetic relationships of species was embraced soon after publication of the first DNA sequences and continues today with a much broader scope (Bellis et al. 2003). DNA identification is based on the assumption that individuals from the same species carry specific DNA sequences that are different from those found in individuals from other species (Pereira et al. 2008). A wide range of molecular techniques have been used for the recognition of taxonomic units including (but not limited to) DNA hybridization, Random Fragment Length Polymorphisms (RFLP), Amplified Fragment Length Polymorphism (AFLP), Random Amplified Polymorphic DNA (RAPD), conventional Polymerase Chain Reaction (PCR), real-time PCR, sequencing and microarrays. Sequences from different loci from coding or non-coding DNA have been used for diverse animal groups with different preferences, and in most cases use of these loci across taxa has not been standardized (Avise and Liu 2011). Based on a comparison of sequences from divergent loci it is difficult to reach a conclusion on species identities and their relationships (Knowles and Carstens 2007). If a new genetic marker is used for the first time, the genetic composition of all species of that taxon should be resolved. This has made the application of DNA in resolving species conflicts not only difficult and non-productive, but in some cases counterproductive (Shaw 2002). The existence of some 5±3 million species of organisms on earth, of which only 1.5 million are named (Costello et al. 2012), with many more that are either morphological synonyms or cryptic complexes, warrants standardization of loci to be used for species identities.

In 2003 it was proposed that a single gene fragment of approximately 650 base pairs from the 5′ end of the mitochondrial cytochrome oxidase I (COI) could be used as a marker to identify an organism to its species (Hebert 2003). This gene region was termed a "DNA barcode" providing the foundation for DNA barcoding and predicting its utility in taxonomy (Hebert and Gregory 2005). Later, sections of two chloroplast genes, matK and rbcL, were selected as the DNA barcodes for plants (CBOL 2009). Soon DNA barcoding turned into an international initiative with an objective to barcode all species on earth with the International Barcode of Life Project (iBOL) centred at the University of Guelph. According to information on the iBOL website (ibol.org), "iBOL's main mission is extending the geographic and taxonomic coverage of the barcode reference library -- Barcode of Life Data Systems (BOLD) -- storing the resulting barcode records, providing community access to the knowledge they represent and creating new devices to ensure global access to this information". With these objectives in mind, iBOL has engaged nations across the globe to barcode all biodiversity. Nations have responded to the challenge and since the launch of iBOL in 2010, barcoding has progressed at a steady speed and currently barcode records are available for more than 2.8 million specimens representing roughly 350,000 species. Previous researchers have often assigned specimens to operational taxonomic units (OTUs) in cases where morphological identifications are difficult (Stackebrandt and Goebel 1994; Kauserud et al. 2008). Ratnasingham and Hebert (2013) recently developed the Barcode Index

Muhammad Ashfaq (mashfaq@uoguelph.ca) is a Research Associate in the Biodiversity Institute of Ontario, University of Guelph. He is the focal person for iBOL's Pakistan National Node and for the last 4 years has been engaged in DNA barcoding of biodiversity of Pakistan. His main focus has been pest and beneficial insects.

Number (BIN) system which provides a species-level taxonomic registry for animal groups (Hausmann et al. 2013) and has aided the discovery of new species (Landry and Hebert 2013). Consequently animal barcode data on BOLD has been organized by BINs.

Pakistan barcode project

DNA barcoding in Pakistan was initiated in 2010 under the project "DNA barcoding economically important insect species of Pakistan". This collaborative project between the National Institute for Biotechnology and Genetic Engineering (NIBGE), Faisalabad, and the Biodiversity Institute of Ontario (BIO), University of Guelph, aimed at sequencing important pest and beneficial insect species and constructing a regional barcode reference library for species identification. The project barcoded more than 10,000 insect specimens during the first 18 months. In 2012, the project was expanded with a grant from the International Development Research Centre (IDRC) through the University of Guelph/iBOL with an added objective of barcode applications in studying special interest pest insects and disease vectors impacting with socio-economic impacts on the country. The IDRC grant was additionally supported by sequencing and data management support provided at BIO through the Canadian Centre for DNA Barcoding (CCDB) and BOLD. This support enabled collection of fresh specimens from remote geographical areas in Pakistan and enhanced the capacity for front end specimen processing and data basing. Arthropod collections for barcoding included agricultural and forest pests, human/animal disease vectors and beneficial predators and parasitoids. Collected specimens were identified at least to order and labeled appropriately, and the data were submitted to BOLD. Barcodes were sequenced following standard barcoding protocols and the sequences obtained were assigned BINs.

Revealing biodiversity

The Pakistan project generated barcode data for 31,333 arthropod and 1,259 plant specimens representing roughly 4,510 insect and 365 plant species. The arthropod specimens represented 12 orders of Insecta and 2 of Arachnida. Thirty seven percent of the BINs were represented by single records (singletons) (Table 1). A significant number of Hymenoptera were barcoded, but almost half of their BINs were singletons. Only 22% of the Pakistan BINs were shared with other countries. A relatively large number of BINs were Coleoptera (668) but just 9% were shared with other countries. Similarly, out of 937 hymenopteran BINs, only 15% found a match with other countries. The highest number of BINs on BOLD came from Lepidoptera (>96,000), but just 29% of Pakistan Lepidoptera found a match on BOLD. A significant number of invasive pest species are Hemiptera, which makes it a particularly important order. Out of 418 BINs from this order, only 20% were common with other countries. Mantodea, which represent important insect predators (beneficial insects), shared only 8% of the BINs with other geographic regions. Spiders (Araneae) are also important predators, and sequences from 1,228 specimens were assigned to 183 BINs, though only 20 (11%) BINs were found in other nations (6,500). The Pakistan data shows the localization of biodiversity, but to reach to a final conclusion on biodiversity overlap between Pakistan and other countries, much more coverage of the regional arthropod fauna is required. A recent publication on the butterflies of Pakistan (Ashfaq et al. 2013) points to the presence of regional endemism in this group of insects. Barcode libraries of butterflies of Pakistan, and cotton species are public and are accessible through the web at dx.doi.org/10.5883/DS-MABUTPUB and dx.doi.org/10.5883/DS-MAPLTPUB, respectively.

Table 1. Number of specimens with a barcode sequence and number of BINs represented in 12 insect orders and spiders from Pakistan

Order	Number of barcodes	BINs recovered	Singleton BINs (%)	BINs shared with other countries (%)
Lepidoptera	4589	1030	284 (28)	397 (29)
Coleoptera	2912	668	305 (46)	59 (9)
Diptera	11677	876	340 (39)	186 (21)
Hemiptera	2900	418	150 (36)	85 (20)
Orthoptera	1256	158	38 (24)	33 (21)
Thysanoptera	471	52	15 (29)	18 (35)
Isoptera	98	9	3 (33)	1 (11)
Phthiraptera	706	7	0	3 (43)
Hymenoptera	4554	937	418 (45)	136 (15)
Neuroptera	499	84	29 (35)	5 (6)
Mantodea	103	37	19 (51)	3 (8)
Odonata	340	51	8 (16)	24 (47)
Araneae	1228	183	78 (43)	20 (11)
Total	31333	4510	1683 (37)	970 (22)

Examples of barcode applications: Revealing cryptic species complexes and discovering new species

Whiteflies:

Whiteflies (*Bemisia tabaci*) are important pests not only due to their direct damage to host plants but also due to their role as virus vectors. This taxon is known to be a complex of at least 35 cryptic species (De Barro and Boykin 2013). Barcode data for 593 *B. tabaci* specimens from Pakistan revealed the presence of 9 BINs in this complex from 2 cotton-growing provinces (Punjab and Sindh). Previously, three *B. tabaci* lineages were documented from the country (Ahmed et al. 2011). The barcode-based species distribution analysis showed that Asia II 1, vector of leaf curl virus on cotton, has expanded its range in other areas. Barcoding has helped our understanding of the expansion of cotton leaf curl disease southward.

Mosquitoes:

Recent outbreaks of dengue fever caused by mosquito-vectored dengue virus have underscored the need to identify and analyze mosquito species from the region. Barcode analysis of 1684 mosquitoes from dengue-affected areas of Pakistan revealed 32 species. The genus *Aedes* was represented by six taxa with the two dengue vector species, *Aedes albopictus* and *Aedes aegypti*, dominant and broadly distributed. The BIN system revealed the presence of four cryptic lineages of *Aedes* which have not been previously reported from the country. The barcode-based distribution analysis when compared with prior studies showed a shift in *Ae. albopictus* habitat from rural (forested areas) to urban areas. Barcodes also discovered the presence of *Anopheles culicifacies-A* and *Anopheles annularis-B*, which are the first reports from Pakistan.

Noctuid species in/ around cotton fields:

Cotton contributes significantly to Pakistan's economy. Although the crop is attacked by a number of insect pests, larvae of Noctuidae are the major concern. Barcode data revealed the

presence of 25 noctuid species in and around cotton fields in Punjab. Barcodes identified four species of *Spodoptera*, of which three, *S. littoralis*, *S. cilium* and *S. exigua*, were the first reports from agricultural areas of Pakistan.

Mango mealybug complex:

The species status of the mango mealybug, *Drosicha mangiferae*/*Drosicha stebbingi*, in Pakistan has been debated for decades (Latif 1949). This species attacks a wide range of host plants and gains its name from its host. The mealybug from mango is called *D. mangiferae* while similar-looking specimens from other trees are called *D. stebbingi*. Barcode data was used to analyze the genetic differences among mealybug species from different host plants from three regions of the country. As nucleotide sequences failed to differentiate specimens from different hosts, it was concluded that the same mealybug species is found on both mango and citrus trees.

In summary, results from the Pakistan barcode project suggest that DNA barcoding provides an efficient tool for assessing and documenting regional biodiversity and comparing and connecting them with global databases. Results support the value of developing regional biodiversity reference libraries through the involvement of all nations across the globe.

Acknowledgements

I thank colleagues at the CCDB for aid with sequence analysis, and staff employed with the DNA barcoding project at NIBGE, Faisalabad, for their diligence in collecting specimens. Financial support from the Higher Education Commission Pakistan and IDRC Canada is acknowledged.

References

Ahmed, M.Z., De Barro, P.J., Greeff, J.M., Ren, S.-X., Naveed, M., and Qiu, B.-L. 2011. Genetic identity of the *Bemisia tabaci* species complex and association with high cotton leaf curl disease (CLCuD) incidence in Pakistan. Pest Management Science, **67**: 307-317.

Ashfaq, M., Akhtar, S., Khan, A.M., Adamowicz, S.J., and Hebert, P.D.N. 2013. DNA barcode analysis of butterfly species from Pakistan points towards regional endemism. Molecular Ecology Resources, **13**: 832-843.

Avise, J.C., and Liu, J.-X. 2011. On the temporal inconsistencies of Linnean taxonomic ranks. Biological Journal of the Linnean Society, **102**: 707-714.

Bellis, C., Ashton, K.J., Freney, L., Blair, B., and Griffiths, L.R. 2003. A molecular genetic approach for forensic animal species identification. Forensic Science International, **134**: 99-108.

CBOL Plant Working Group et al. 2009. A DNA barcode for land plants. Proceedings of the National Academy of Sciences USA, **106**: 12794–12797.

Costello, M.J., Wilson, S., and Houlding, B. 2012. Predicting total global species richness using rates of species description and estimates of taxonomic effort. Systematic Biology, **61**: 871-883.

De Barro, P., and Boykin, L.M. 2013. Global *Bemisia* dataset release version 31 December 2012. v1. CSIRO. Data Collection. 10.4225/08/50EB54B6F1042.

Hausmann, A., Godfray, H.C.J., Huemer, P., Mutanen, M., Rougerie, R., van Nieukerken, E.J., Ratnasingham, S., and Hebert, P.D.N. 2013. Genetic patterns in European Geometrid moths revealed by the Barcode Index Number (BIN) system. PLoS ONE, **8**: e84518.

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

Hebert, P.D.N., and Gregory, T.R. 2005. The promise of DNA barcoding for taxonomy. Systematic Biology, **54**: 852-859.

Hey, J. 2001. The mind of the species problem. Trends in Ecology and Evolution, **16**: 326-329.

Kauserud, H., Mathiesen, C., and Ohlson, M. 2008. High diversity of fungi associated with living parts of boreal forest bryophytes. Botany, **86**: 1326-1333.

Knowles, L.L., and Carstens, B.C. 2007. Delimiting species without monophyletic gene trees. Systematic Biology, **56**: 887-895.

Landry, J.-F., and Hebert, P.D.N. 2013. *Plutella australiana* (Lepidoptera, Plutellidae), an overlooked diamondback moth revealed by DNA barcodes. ZooKeys, **327**: 43-63.

Latif, A. 1949. The taxonomic status of *Drosicha stebbingi* (Green) and *Droshicha mangiferae* (Green) (Hem., Coccid.). Bulletin of Entomological Research, **40**: 351-354.
Pereira, F., Carneiro, J., and Amorim, A. 2008. Identification of species with DNA based technology: current progress and challenges. Recent Patents on DNA & Gene Sequencing, **2**: 187-199.
Ratnasingham, S., and Hebert, P.D.N. 2013. A DNA-based registry for all animal species: the Barcode Index Number (BIN) System. PLoS ONE, **8**: e66213.
Shaw, K.L. 2002. Conflict between nuclear and mitochondrial DNA phylogenies of a recent species radiation what mitochondrial DNA reveals and conceals about modes of speciation in Hawaiian crickets. Proceedings of the National Academy of Sciences USA, **99**: 16122-7.
Stackebrandt, E., and Goebel, B. 1994. Taxonomic note: A place for DNA-DNA reassociation and 16S rDNA sequence analysis in the present species definition in bacteriology. International Journal of Systematic Bacteriology, **44**: 846-849.

Entomological insights from studying a bird

Sean McCann

When I started my doctoral research, I had never considered working with birds. I was an entomologist! I had finished a Master's degree studying reproductive ecology of mosquitoes, and wanted to branch out into insects with a bigger brain…maybe Hymenoptera. Over the last several years, I have indeed studied several hymenopterans, but mainly through the eyes of a specialist vertebrate predator.

Red-throated Caracaras (*Ibycter americanus*) are members of the Falconidae found in forested areas from Central America to southern Brazil. Until recently, the behaviour and life history of this strikingly-marked raptor (Fig. 1) was little known to scientists. For my doctoral research, I have been lucky enough to lead some fascinating fieldwork on these birds in the pristine rainforests of French Guiana.


Sean McCann

Figure 1. The Red-throated Caracara (*Ibycter americanus*) is an unusual member of the Falconidae that preys primarily on social wasps.

From observations and examination of gut-contents, Red-throated Caracaras had been reported to be specialist predators of social wasps (Thiollay 1991), but there was little in the way of quantitative data. Our 2010 nest camera study (McCann et al. 2010) was the first to provide a close glimpse at the nesting biology and diet of these birds. Confirming some previous reports, we found that the caracaras are highly cooperative, with up to seven adults providing care to a single chick. Surprisingly, we found the caracaras nested in large

Sean McCann (smmccann@sfu.ca), a PhD student at Simon Fraser University, has been a member of the ESC since 2003, when he was an undergraduate student of Richard Ring at the University of Victoria. He has studied Medical Entomology at the University of Florida and is currently engaged in conservation research in Honduras. You can follow his research and photography at www.ibycter.com.