UNIVERSITY of GUELPH

Centre for Biodiversity Genomics

The Spiders of Pakistan: Commencing the Assembly of a National DNA Barcode Reference Library

Muhammad Ashfaq[1], Gergin Blagoev[1], Arif Khan[2], Hafiz Tahir[3], Khalid Mukhtar[2], Paul Hebert[1]

[1] Centre for Biodiversity Genomics, University of Guelph, Canada; [2] Department of Biotechnology/Zoology, University of Sargodha, Pakistan; [3] Department of Zoology, GC University Lahore, Pakistan

Abstract: While DNA barcode coverage for spiders is well advanced in some regions, they are poorly studied in Pakistan. The current study focused on the creation of a comprehensive DNA barcode reference library of spider fauna of Pakistan. Spiders from 225 sites across the country were identified morphologically before being sequenced for the barcode region of COI. Barcodes of a total 1795 individuals, representing 28 families, 111 genera, and 199 species (including 86 species with interim names), were assigned to 218 BINs. The 199 named species were assigned to 215 BINs and 13 of them showed BIN splits. Maximum conspecific divergence ranged from 0% to 5.3% while congeneric distances ranged from 2.8% to 23%. 17 genera and 47 species are new taxonomic records to the spider fauna of Pakistan, while eight undescribed species are potentially new to science. Barcodes assisted the confirmation of opposite genders of the two endemic species, the undescribed male of *Cyclosa chichawatniensis* (Araneidae), and undescribed female of *Menemerus nigli* (Salticidae). Only one quarter of the taxa detected in this study were known from other countries which underscores the importance of regional barcode reference libraries.

Background:

The broad adoption of DNA barcoding has accelerated the documentation of animal biodiversity (Janzen et al. 2009) by constructing DNA barcode reference libraries (Gwiazdowski et al. 2015). These libraries have facilitated the identification of unknown organisms and are being developed at regional and global scales (Astrin et al. 2016). Use of the Barcode Index Number (BIN) system as species proxy (Ratnasingham and Hebert 2013) has accelerated the documentation and enumeration of biodiversity. Current progress for DNA barcoding spiders has resulted in 106,000 barcode records representing over 12% of the 46,000 species documented on the Barcode of Life Data Systems (BOLD). However, no spiders of Pakistan have been documented through barcoding. We take the first step to fill this gap by initiating the construction of a barcode reference library for spider fauna of Pakistan.

Methods:

1795 spiders collected from 225 diverse sites in the country (Fig. 1) were identified by morphology and subsequently sequenced for the DNA barcode region of COI (Hebert et al. 2003). Specimen data and images were submitted to BOLD (DS-MASPD) and barcoding was accomplished at the Canadian Centre for DNA Barcoding following standard methods (http://ccdb.ca/resources/). DNA sequences were submitted to BOLD where they were assigned to BINs after rigorous analysis. Accumulation curve analysis, BIN discordance, distance (K2P) analysis, and barcode gap analysis (BGA) were performed using analytical tools on BOLD. Neighbor-joining (NJ) analysis was performed by MEGA5 (Tamura et al. 2011) and Bayesian inference by MrBayes (Huelsenbeck and Ronquist 2001).
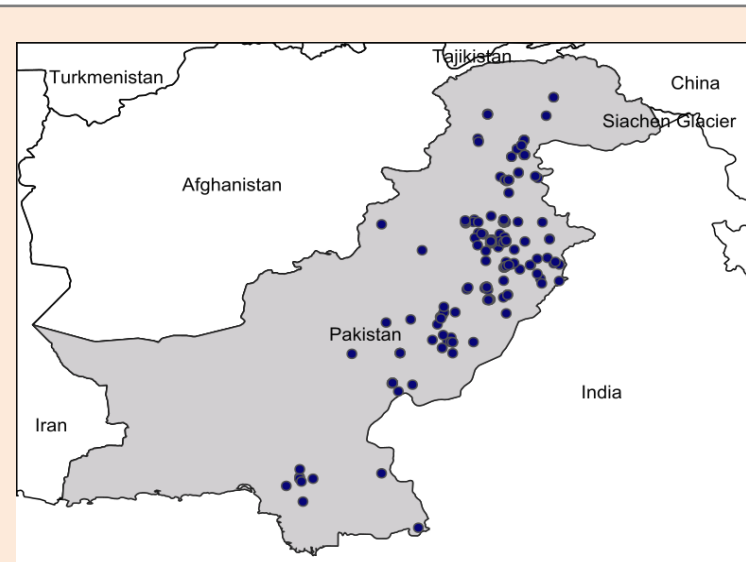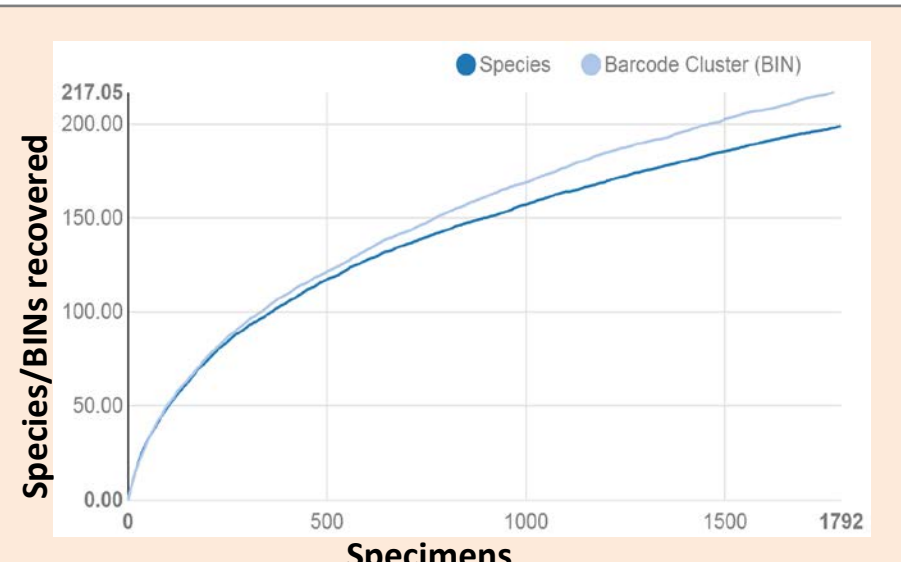

Fig 1: Spider collection sites in Pakistan

Fig 2: Accumulation curve of species/BINs

Results:

1579 of the 1795 specimens were identified to 113 Linnaean species. The other 213 were assigned to 86 interim species, while the remaining three were named only to family. Collectively, the specimens belonged to 28 families, 111 genera and 199 species. Araneidae, *Oxyopes*, and *Neoscona theisi* were the most common family, genus, and species in the collection. The 199 names species were assigned to 215 BINs, as twelve of them showed BIN splits. The NJ and Bayesian analysis supported the monophyly of BINs. Species recovery increased with the samples analyzed, and there was no sign of asymptote (Fig. 2). K2P distances varied at various taxonomic ranks (Table 1). Except for 13 species, conspecific divergence in all the Linnean species, with three of more records, was under 2% (Table 2).

Table 1: K2P sequence divergence in the COI barcode region for spider species from Pakistan.

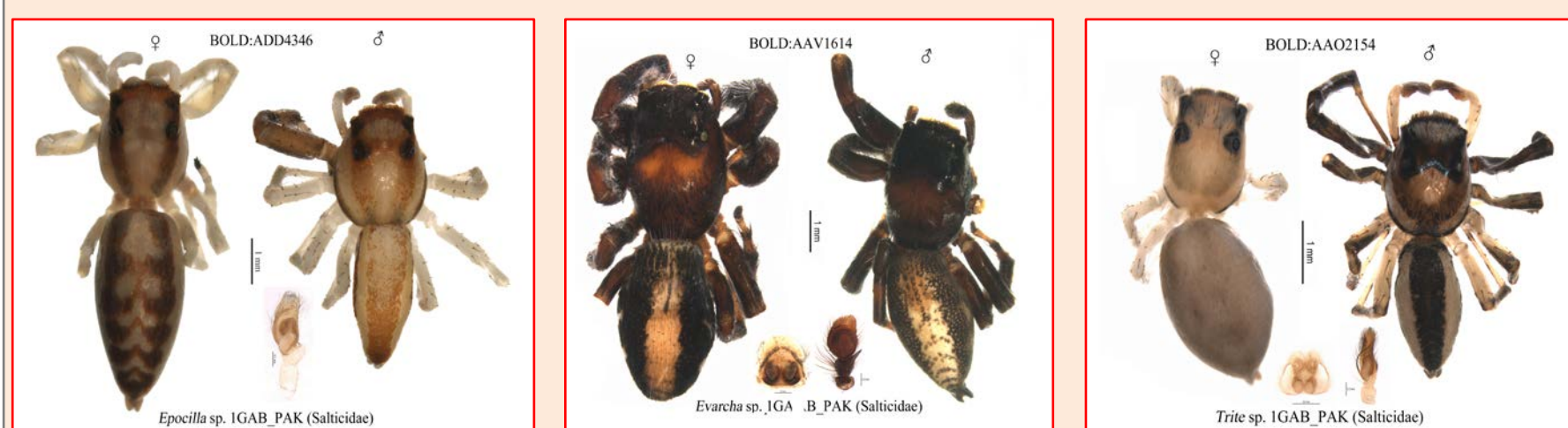| Distance class	n	Taxa	Comparisons	Min (%)	Mean (%)	Max (%)
Intraspecific	1545	90	43805	0	0.8	5.3
Congeners	961	18	52668	2.8	8.7	23.2
Confamilial	1464	12	113301	5.3	15.4	26.7

Fig. 3: Specimen images (both sexes), with BINs, of three of the eight spider species potentially new to science.
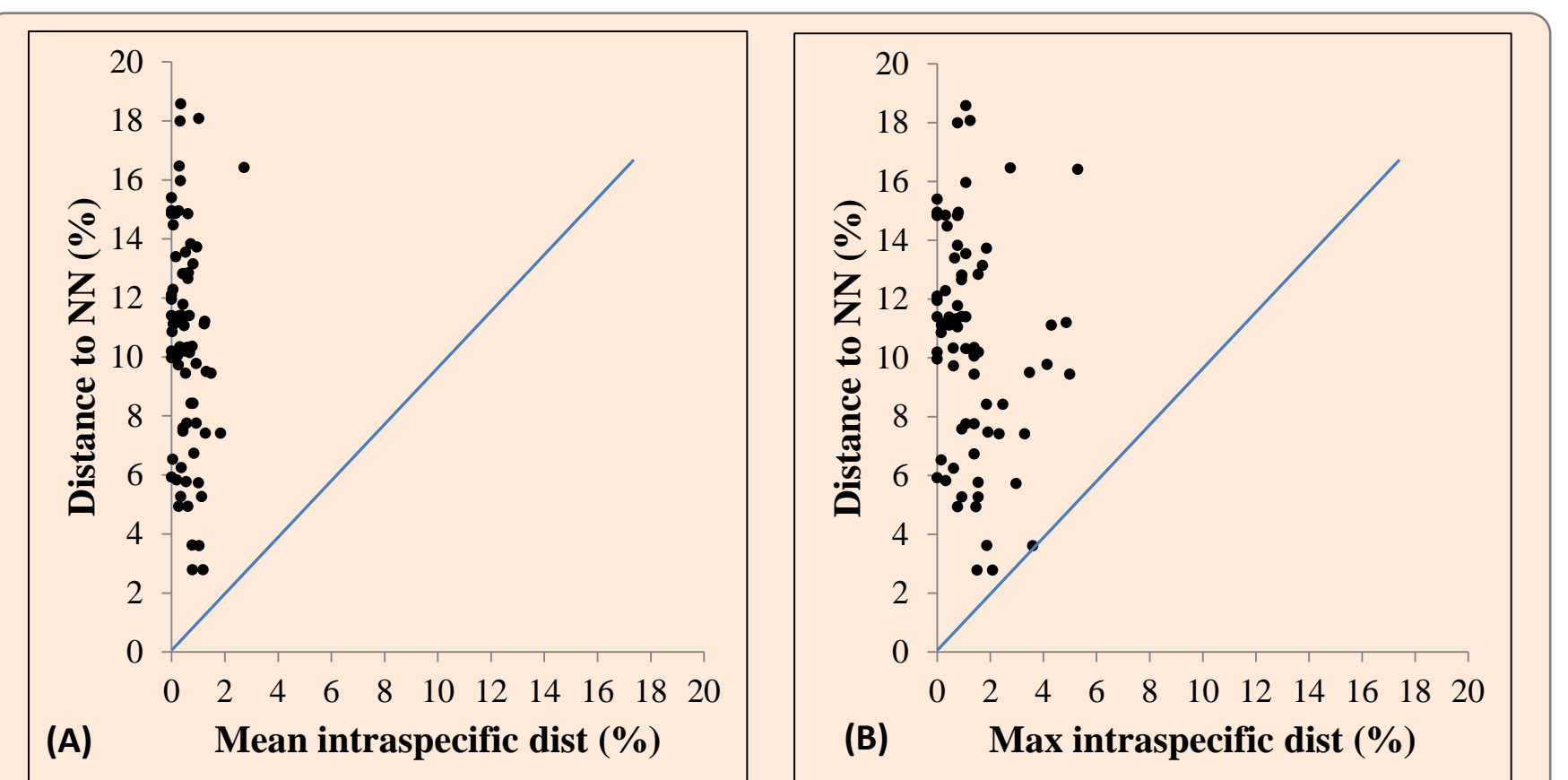

Fig. 4: Barcode gap analysis (BGA) of spider species with >3 records. Dots above the diagonal line indicate existence of a barcode gap.

The BGA showed that intraspecific distances did not exceed the distance to NN species. Using the BGA tool on BOLD, a species is unambiguously identifiable if its max. intraspecific distance is less than its NN distance.

Spider biodiversity connections with other regions were calculated by BIN-overlap analysis. Just 51 of the 218 BINs in this study found a match against >10,000 spider BINs on BOLD representing 45 countries.

Table 2 and Fig. 5: Six of the 12 spider species with high intraspecific divergence and BIN splits (A) and their clustering pattern on NJ tree (B).

	Species	K2P divergence (n)	BINs
1	*Crossopriza maculipes*	5.3 (12)	3
2	*Hippasa pisaurina*	4.2 (17)	2
3	*Plexippus paykulli*	5.0 (41)	5
4	*Runcinia insecta*	4.9 (42)	2
5	*Spariolenus tigris*	4.1 (2)	2
6	*Thomisus zaheeri*	4.3 (31)	2

(A)   (B)

Conclusions and Significance:

- The study generates the largest dataset of DNA barcodes for spider fauna of Pakistan to initiate the construction of barcode reference library;
- A high level of congruence between the Linnaean species and BINs support the validity of BINs as proxy for morphological species;
- High intraspecific divergence, and BIN splits, in the known species suggest the presence of cryptic species complexes;
- 47 species and 17 genera are reported for the first time in Pakistan and eight species are potentially new to science;
- Barcodes confirmed the opposite genders of the two endemic species in the region;
- One jumping spider, *Sonoita lightfooti*, found in Pakistan through this study was previously reported only in Africa, which is a new record for Eurasia;
- The low level of BIN overlap with other regions underscores the need to construct regional DNA barcode reference libraries.

References:
Janzen DH. et al. (2009). Molecular Ecology Resources 9(s1): 1-26.
Gwiazdowski et al. (2015). PLOS ONE 10: e0125635.
Astrin JJ, et al. (2016). PLOS ONE11: e0162624.
Hebert PDNH (2003). Proceedings of the Royal Society B: Biological Sciences 270:313-321.
Tamura et al. (2011). Molecular Biology and Evolution 28: 2731-2739.
Huelsenbeck and Ronquist (2001). Bioinformatics 17:754-755.