

RESEARCH ARTICLE

BIN overlap confirms transcontinental distribution of pest aphids (Hemiptera: Aphididae)

Muhammad Tayyib Naseem^{1,2}, Muhammad Ashfaq^{3*}, Arif Muhammad Khan^{1,4}, Akhtar Rasool^{1,5}, Muhammad Asif¹, Paul D. N. Hebert³

1 National institute for Biotechnology and Genetic Engineering, Faisalabad, Pakistan, **2** Pakistan Institute of Engineering and Applied Sciences, Islamabad, Pakistan, **3** Centre for Biodiversity Genomics & Department of Integrative Biology, University of Guelph, Guelph, ON, Canada, **4** Department of Biotechnology, University of Sargodha, Sargodha, Pakistan, **5** Department of Zoology, University of Swat, Swat, Pakistan

* mashfaq@uoguelph.ca



OPEN ACCESS

Citation: Naseem MT, Ashfaq M, Khan AM, Rasool A, Asif M, Hebert PDN (2019) BIN overlap confirms transcontinental distribution of pest aphids (Hemiptera: Aphididae). PLoS ONE 14(12): e0220426. <https://doi.org/10.1371/journal.pone.0220426>

Editor: Feng ZHANG, Nanjing Agricultural University, CHINA

Received: July 11, 2019

Accepted: November 24, 2019

Published: December 10, 2019

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: <https://doi.org/10.1371/journal.pone.0220426>

Copyright: © 2019 Naseem et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: The specimen metadata and sequences generated in this study are available on the Barcode of Life Data System

Abstract

DNA barcoding is highly effective for identifying specimens once a reference sequence library is available for the species assemblage targeted for analysis. Despite the great need for an improved capacity to identify the insect pests of crops, the use of DNA barcoding is constrained by the lack of a well-parameterized reference library. The current study begins to address this limitation by developing a DNA barcode reference library for the pest aphids of Pakistan. It also examines the affinities of these species with conspecific populations from other geographic regions based on both conventional taxonomy and Barcode Index Numbers (BINs). A total of 809 aphids were collected from a range of plant species at sites across Pakistan. Morphological study and DNA barcoding allowed 774 specimens to be identified to one of 42 species while the others were placed to a genus or subfamily. Sequences obtained from these specimens were assigned to 52 BINs whose monophyly were supported by neighbor-joining (NJ) clustering and Bayesian inference. The 42 species were assigned to 41 BINs with 38 showing BIN concordance. These species were represented on BOLD by 7,870 records from 69 countries. Combining these records with those from Pakistan produced 60 BINs with 12 species showing a BIN split and three a BIN merger. Geo-distance correlations showed that intraspecific divergence values for 49% of the species were not affected by the distance between populations. Forty four of the 52 BINs from Pakistan had counterparts in 73 countries across six continents, documenting the broad distributions of pest aphids.

Introduction

Although aphids (Hemiptera: Aphididae) are important plant pests, their life stage diversity and phenotypic plasticity have constrained the development of effective taxonomic keys in the past [1,2]. With over 5,000 described species, the Aphididae represents the largest family within the Aphidoidea [3]. Most pest aphids belong to the subtribe Aphidina which includes

(www.boldsystems.org (BOLD) in the dataset DS-MAAPH.

Funding: This research was supported by a grant from the Higher Education Commission (HEC), Pakistan and by grant 106106-001 (Engaging Developing Nations in iBOL) from IDRC. MT Naseem was supported by the Higher Education Commission (HEC), Pakistan under the Indigenous PhD Fellowship. This is a contribution to the Food from Thought project supported by the Canada First Research Excellence Fund.

Competing interests: The authors have declared that no competing interests exist.

670 described species so far [3,4]. Nearly 100 aphid species have been listed as serious agricultural pests; they damage more than 300 plant species [5,6], and lower crop yield by direct feeding and by transmitting viral diseases [7].

Sibling species complexes are common in many pest aphids [8]. Very often, these species are morphologically identical but genetically distinct [9]. They often include anholocyclic biotypes (= clones) with differing host preferences and varying competency for disease transmission [10,11]. Species identification is so challenging that taxonomic keys are either ineffective or only useful for a particular geographic area or taxonomic group [12]. These deficits have prompted the search for alternative approaches for identification such as protein profiling [13] and the use of DNA sequence data [14,15]. However, the later approach has gained stronger uptake due to its universal applicability, low cost, and strong performance [16].

Past studies have demonstrated that DNA-based approaches can enable both specimen identification and the clarification of putative cryptic species complexes [17,18]. Diverse mitochondrial and nuclear genes have been used individually and in combination to discriminate insect species [19–21]. Although multigene analyses are valuable in resolving complex taxonomic situations and essential for phylogenetic reconstructions [22,23], it has seen little application in routine identifications [18]. By contrast, DNA barcoding [24] employs a 658 base pair (bp) segment of a single mitochondrial gene, cytochrome *c* oxidase I, to discriminate animal species. Because of its ease of application, DNA barcoding has become the most popular approach for the identification of specimens in diverse insect groups including aphids [25–32]. Its effortless integration with high-throughput sequencing workflows has made DNA barcoding an effective tool for large-scale pest diagnosis, biosurveillance, and biodiversity assessments [33–35].

The application of DNA barcoding requires bioinformatics support and a comprehensive reference library [36]. The Barcode of Life Data System (BOLD— www.boldsystems.org) [37] meets the former need and currently includes more than six million barcode records from animals. Most of these records are from insects (5.2 million) and 49,000 of them derive from aphids (accessed 3 July 2019). All barcode sequences meeting quality criteria receive a Barcode Index Number (BIN) [38]. BINs are an effective species proxy because they correspond closely with species designated through morphological study [39,40]. As a result, BINs are now routinely employed for biodiversity assessments, counting species, analyzing cryptic species complexes, and assessing species ranges [41–43]. All these developments have generated considerable interest in DNA barcoding, leading to the development of well-parameterized reference barcode libraries for some groups at continental and global scales [32,44–48].

Although the DNA barcode library for insects is still incomplete, it is already highly valuable for identifying various pest species and assessing their distributions [29,45,49–52]. However, the lack of reference sequences constrains the utility of DNA barcoding in many situations. Although barcode coverage for the aphid fauna of some countries is extensive [46,53,54], DNA barcoding studies in other nations, including Pakistan, for these pest species are limited. The current study addresses this gap by generating a barcode reference library for the pest aphids of Pakistan, and by using BINs to reveal their links to aphid assemblages in other regions.

Materials and methods

Ethics statement

No specific permissions were required for this study. The study did not involve endangered or protected species.

Aphids were sampled from 123 plant species representing 33 families at 87 sites in Pakistan (Fig 1, S1 Table) during 2010–2013. These sites included agricultural settings, nurseries, national parks, botanical gardens, natural forests, and disturbed habitats. Based on GPS

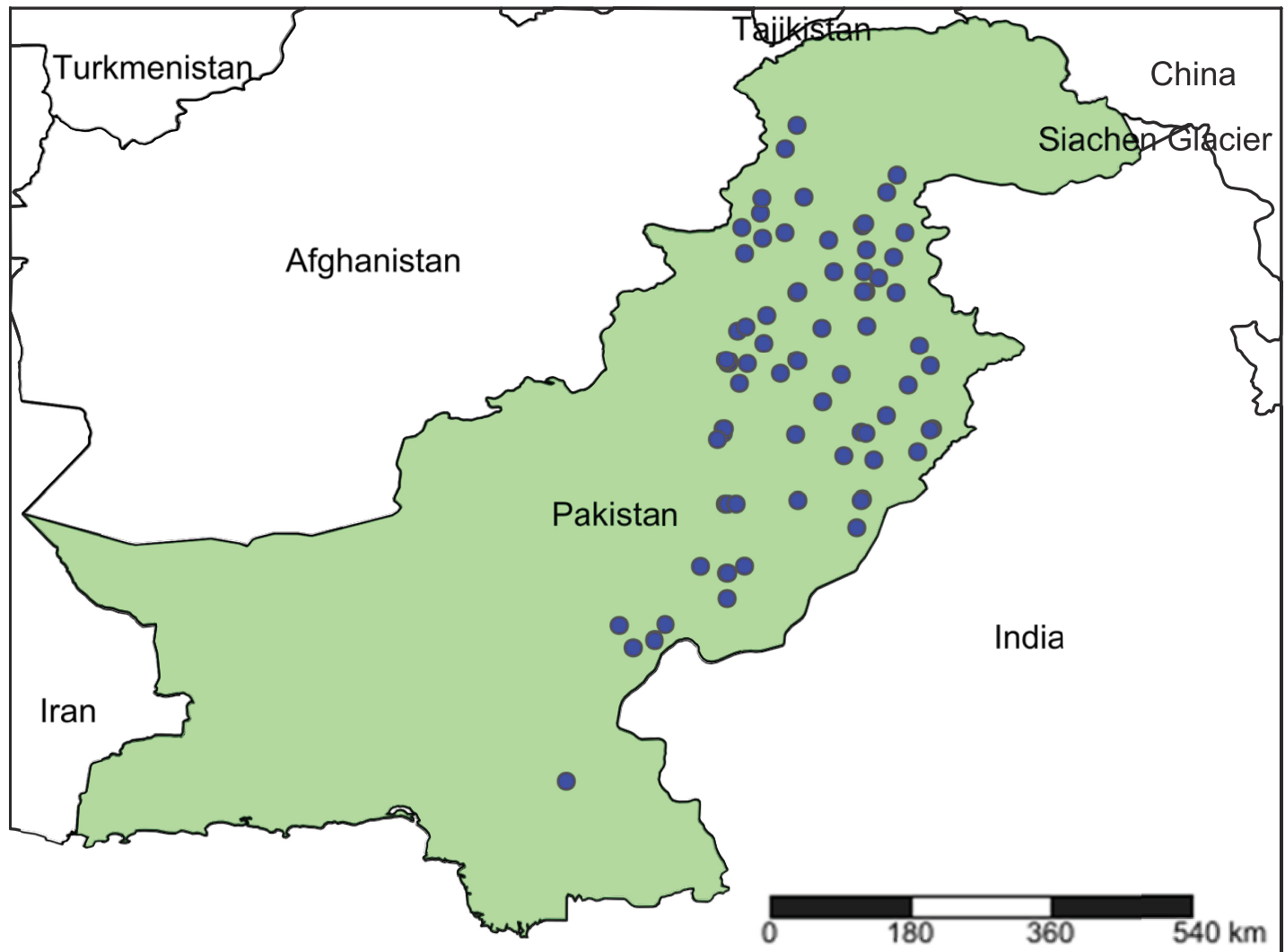


Fig 1. Collection sites for aphids in Pakistan. The map was generated by www.simplemapp.net using GPS coordinates.

<https://doi.org/10.1371/journal.pone.0220426.g001>

coordinates, the collection sites were rendered using SimpleMapp.net (Fig 1). Aphids were collected by either beating foliage above a white paper sheet or by removing them from their host plant with a camel hair brush [55]. Collections were transferred into Eppendorf tubes pre-filled with 95% ethanol and stored at -20°C until analysis.

Identification

Aphids were identified using standard taxonomic keys [55,56]. Morphological characters were examined with a Labomed CZM6 stereomicroscope (Labo America) equipped with an ocular micrometer. Each specimen was identified to species-level based on morphology. This identification was later validated by DNA barcode sequence matches on BOLD.

DNA barcoding

Front-end processing, including specimen sorting, arraying, databasing, and imaging was performed at the Insect Molecular Biology Laboratory, National Institute for Biotechnology and

Genetic Engineering (NIBGE), Faisalabad. Individual specimens were placed into 96-well format in a microplate pre-filled with 30 μ l of 95% ethanol in each well. Each specimen was photographed dorsally using a 12 megapixel Olympus μ -9000 camera (Olympus America Inc., USA) mounted on a stereomicroscope. Specimen metadata (collection information and taxonomy) and images were submitted to BOLD under the project MAAPH, “Barcoding Aphid Species of Pakistan”. DNA extraction, PCR amplification, and sequencing were carried out at Centre for Biodiversity Genomics at Guelph. DNA extraction followed Ivanova et al. [57] with voucher recovery protocol [58]. PCR amplification of the COI-5' (barcode region) [24] was performed using primer pair C_LepFolF (forward) and C_LepFolR (reverse) (http://ccdb.ca/site/wp-content/uploads/2016/09/CCDB_PrimerSets.pdf) in 12.5 μ L reactions that included standard PCR ingredients [59] and 2 μ L of DNA template. The thermocycling regime was: 94°C (1 min), 5 cycles at 94°C (40 s), 45°C (40 s), 72°C (1 min); 35 cycles at 94°C (40 s), 51°C (40 s), 72°C (1 min); and a final extension at 72°C (5 min). PCR success was verified by analyzing the amplicons on 2% agarose E-gel® 96 system (Invitrogen Inc.). Specimens which failed to amplify in the first round of PCR were re-run with primers LepF2_t1 (TGTAACGACG GCCAGTAATCATAARGATATYGG) [60] and LepR1 using the same PCR conditions. PCR products were sequenced bidirectionally on an Applied Biosystems 3730XL DNA Analyzer using the BigDye Terminator Cycle Sequencing Kit (v3.1) (Applied Biosystems). Sequences were edited using CodonCode Aligner (CodonCode Corporation, USA), and translated on MEGA v6 [61] to confirm they were free of stop codons, and submitted to BOLD. The specimen metadata and sequences generated in this study are available on BOLD in the dataset DS-MAAPH. Vouchers were deposited at the Insect Museum, NIBGE, Faisalabad, Pakistan (with sample ID prefix NIBGE) and at the Centre for Biodiversity Genomics, Guelph, ON, Canada (with ID prefix BIOUG).

Data analysis

All barcode sequences were compared with those on BOLD and GenBank to ascertain sequence similarities. Sequence matches on BOLD were obtained using the “Identification Engine” tool whereas nBLAST (<http://www.ncbi.nlm.nih.gov/blast/>) was used on GenBank. All sequences meeting quality standards (>500 bp, <1% ambiguous bases, no stop codon or contamination flag) were assigned to a BIN [38] (as of 18 January 2019). BIN discordance and BIN overlap reports were generated using analytical tools on the BOLD workbench. As a test of the reliability of species discrimination, the presence or absence of a ‘barcode gap’ [62] among species was determined on BOLD. A species was considered distinct when its maximum intraspecific distance was less than the distance to its nearest neighbor (NN).

ClustalW nucleotide alignments [63] and neighbor-joining (NJ) analysis [63] were conducted in MEGA6. The NJ analysis employed the Kimura-2-Parameter (K2P) [64] distance model, with pairwise deletion of missing sites, and 1000 non-parametric bootstrap [65] replicates for the nodal support. Bayesian inference was performed in MrBayes v3.2.0 [66] employing the Markov Chain Monte Carlo (MCMC) technique. This analysis was based on representative sequences from 67 aphid haplotypes in the dataset extracted using DNAsp v5.10 [67] with *Diaphorina citri* (Hemiptera: Psyllidae) as outgroup. The data were partitioned in two ways; i) a single partition with parameters estimated across all codon positions, ii) a codon-partition in which each codon position was allowed different parameter estimates. The evolution of sequences was modelled by the GTR+ Γ model independently for the two partitions using the “unlink” command in MrBayes, and the best model was selected using Find-Model (www.hiv.lanl.gov/cgi-bin/findmodel/findmodel.cgi). The analyses were run for 10 million generations using four chains with sampling every 1000 generations. Bayesian

posterior probabilities were calculated from the sample points once the MCMC algorithm converged. Convergence was defined as the point where the standard deviation of split frequencies was less than 0.002 and the PSRF (potential scale reduction factor) approached 1, and both runs converged to a stationary distribution after the burn-in (by default, the first 25% of samples were discarded). Each run produced 100001 samples of which 75001 samples were included. The trees generated through this process were visualized using FigTree v1.4.0.

BOLD was searched for barcode records for the 42 species encountered in this study. The resultant records were examined for BIN assignment [38] and used in a geo-distance correlation analysis to examine the relationship between geographic and genetic distance in each species. Two methods were employed in the latter analysis; the Mantel Test [68] was used to examine the relationship between geographic (km) and genetic (K2P) distance matrices, while the second approach compared the spread of the minimum spanning tree of collection sites and maximum intra-specific divergence [69]. The relationship between geographic and intra-specific distances was analyzed for each species with at least one individual from three or more sites. BINs recovered from Pakistan were also used in BIN-overlap analysis on BOLD to ascertain the incidence of unique BINs in a region, and to estimate overlap in BIN composition.

Results

Morphological analysis facilitated by the barcode data enabled identification of most specimens (774/809) and revealed 42 species, each representing an important crop pest (S1 Table). Another 32 specimens could be placed to a genus while the remaining three could only be assigned to a subfamily (Aphidinae). Overall, the 809 specimens included representatives of 30 genera and five subfamilies (Aphidinae, Calaphidinae, Chaitophorinae, Eriosomatinae, Lachninae) of the Aphididae (S2 Table). Members of the Aphidinae were dominant (n = 780) as the other four subfamilies were represented by just 29 specimens with Chaitophorinae and Lachninae each contributing one specimen (S2 Table). Among the determined genera, *Aphis* was most common one (n = 306), and was represented by eight identified and three undetermined species. Furthermore, *Myzus* was the second most frequent genus (n = 170), but it was only represented by one species, *Myzus persicae*. *Rhopalosiphum*, the third most abundant (83) genus, was represented by three major pest species (*R. maidis*, *R. padi*, *R. rufiabdominale*). Two species (*Aphis astragalina*, *Periphyllus lyropictus*) represented first records for Pakistan whereas two others (*Lipaphis pseudobrassicae*, *Sarucallis kahawaluokalani*) were known, but were recorded as *Lipaphis erysimi* and *Tinocallis kahawaluokalani*.

The 809 barcode sequences provided two or more records for 36 of the 42 species and single records for the rest (Tables 1 and S1). Maximum K2P divergence values within species ranged from 0–3.6% (mean = 0.1%), whereas distance values within genera were between 0.8–10.3% (mean = 7.4%), and within family (Aphididae) 3.7–17.3% (mean = 9.6%) (Table 1). Barcode gap analysis examined the ability of barcodes to discriminate the 42 named species. With the exception of one species (*Aphis gossypii*), where the maximum intraspecific distance (3.6%) overlapped with *A. affinis*, the maximum intraspecific distance for each species was less than

Table 1. Sequence divergence (K2P) in the COI barcode region for aphid species from Pakistan with more than three specimens, genera with three or more species, and families with three or more genera. This analysis only considers specimens that were assigned to a Linnaean species.

Distance class	n	Taxa	Comparisons	Min (%)	Mean (%)	Max (%)
Within Species	764	36	30756	0	0.1	3.6
Within Genus	434	6	32305	0.8	7.4	10.3
Within Family	770	1	233004	3.7	9.6	17.3

<https://doi.org/10.1371/journal.pone.0220426.t001>

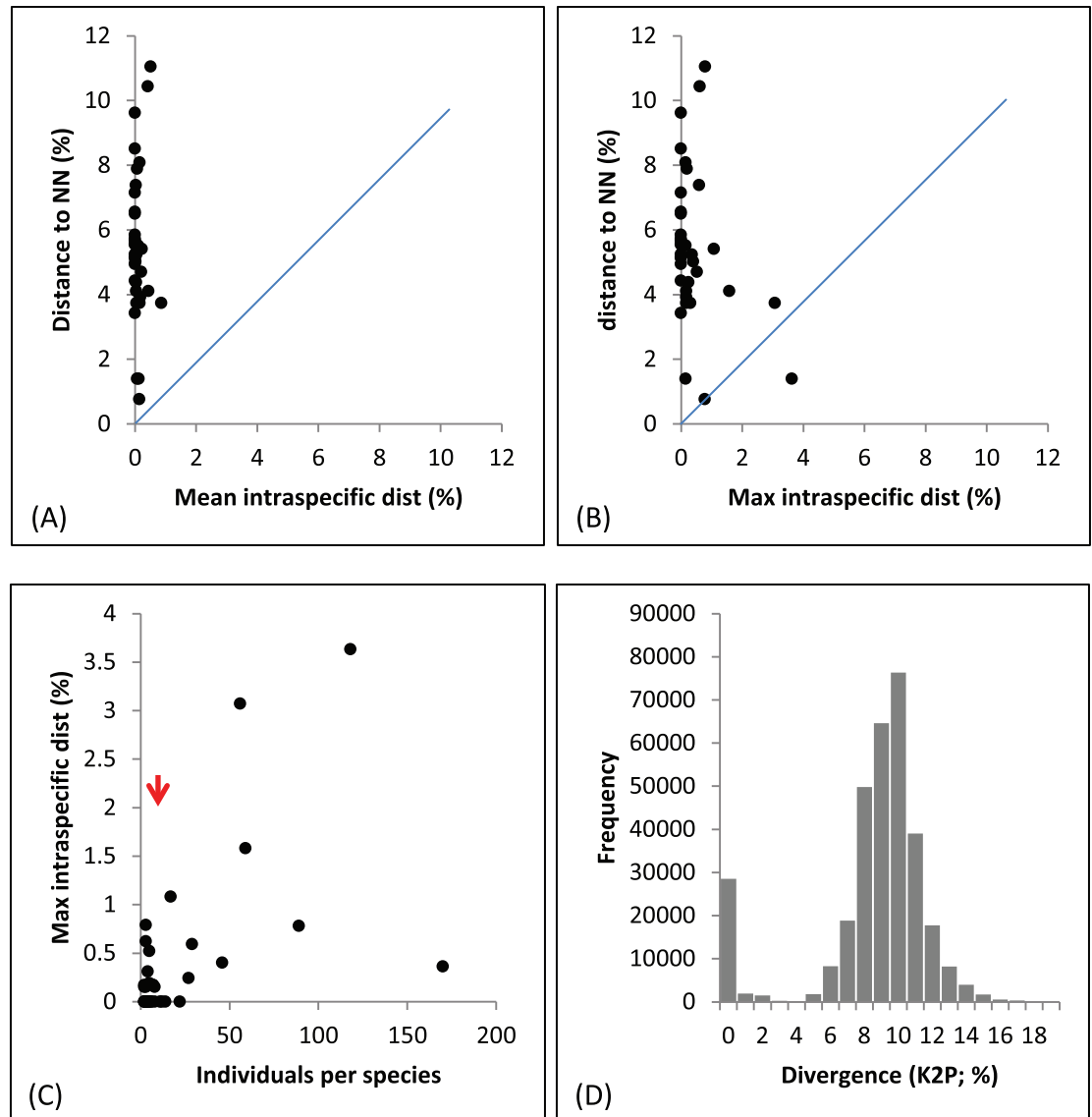


Fig 2. Barcode gap analysis for species of aphids with three or more specimens collected in Pakistan. NN = nearest neighbor.

<https://doi.org/10.1371/journal.pone.0220426.g002>

its NN distance (Fig 2A, Fig 2B). This pattern did not change with increased sample size (Fig 2C).

Nearly all sequences (801/809) fulfilled the criteria for a BIN assignment, and they were placed in 52 BINs. The 774 specimens of the 42 species were assigned to 41 BINs; 38 showed BIN concordance (species members = BIN members), one species (*Rhopalosiphum padi*) was split (AAA9899, ACF2924), and two species (*Aphis affinis*, *A. gossypii*) were merged (AAA3070), while another, *Aphis astragalina* lacked a BIN assignment due to its low quality sequence (410 bp, 9 Ns). The 32 specimens lacking a species assignment were placed in 9 BINs—three for *Aphis* and one for each of the other six genera (*Acyrtosiphon*, *Capitophorus*, *Forda*, *Hyalopterus*, *Macrosiphoniella*, *Schizaphis*). The three specimens only identified to a subfamily were assigned to two BINs. NJ analysis (Fig 3) and Bayesian inference (BI) (Fig 4) supported the monophyly of each of the 52 BINs. The NJ and BI also discriminated the species

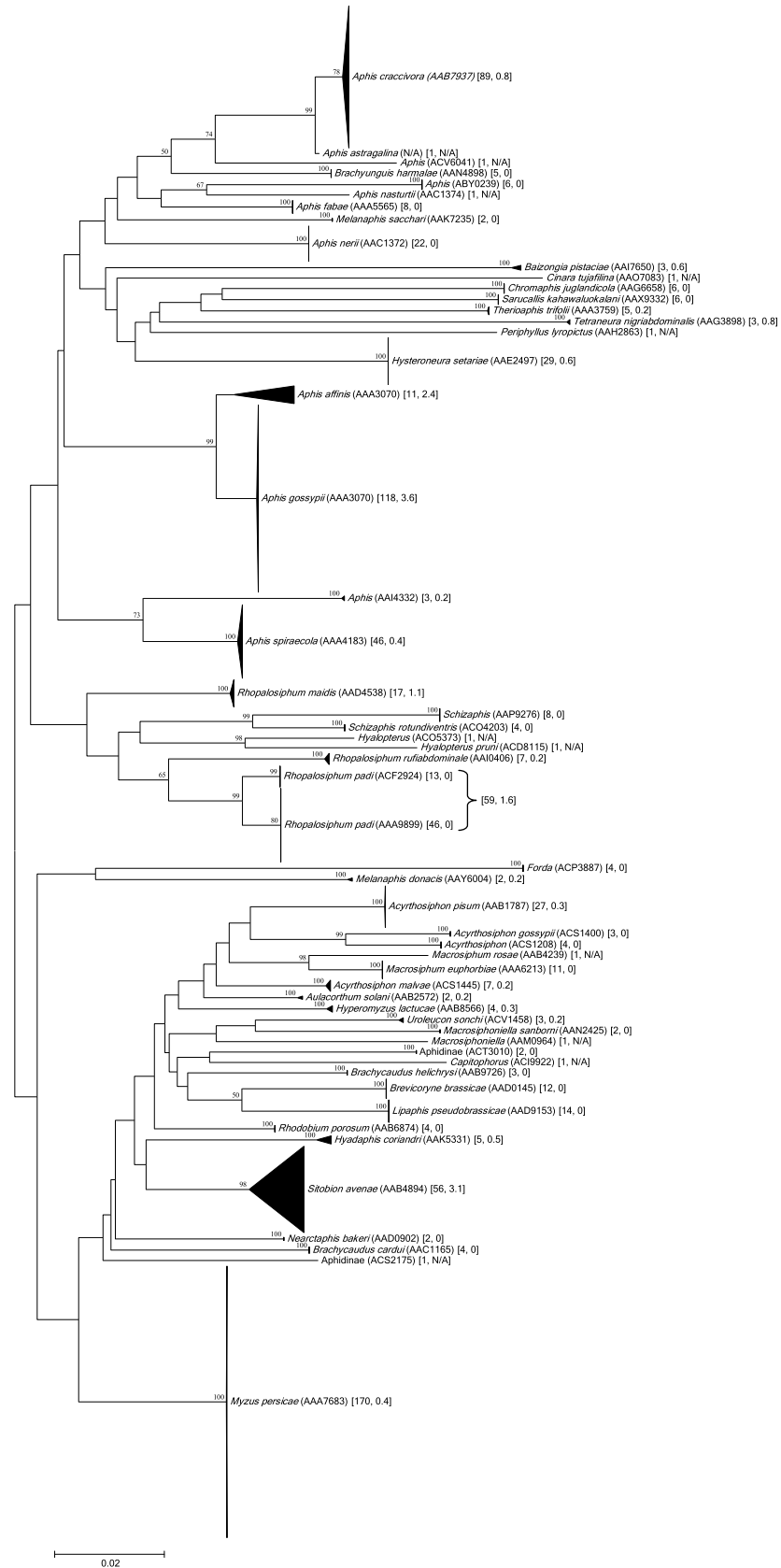


Fig 3. NJ analysis of COI-5' sequences from species/BINs of aphids from Pakistan. Bootstrap values (%) (1,000 replicates) are shown above the branches (values <50% are not shown) while the scale bar shows K2P distances. The node for each species/BIN with multiple specimens was collapsed to a vertical line or triangle, with the horizontal depth indicating the level of intraspecific divergence. BIN numbers are shown for species with only family- or genus-level identification or those split into two BINs.

<https://doi.org/10.1371/journal.pone.0220426.g003>

or genera that either lacked (*Aphis astragalina*) or shared BINs (*Aphis gossypii*, *A. affinis*), as they formed distinct branches on the NJ and BI trees (Fig 3, Fig 4).

Geo-distance correlation analysis for 37 species was conducted by including an additional 5,067 sequences from conspecific individuals deposited on BOLD. This analysis showed that intraspecific divergences in 49% of the species were not affected by expanding analysis to consider their entire ranges (Mantel test, $P > 0.01$) (Table 2). The other 51%, that were affected by geographic range, included six species with BIN splits and eight with intraspecific divergence higher than >2%. The distributional patterns of aphids detected in Pakistan were further analyzed by examining BIN overlap between Pakistan and other countries, a comparison that involved 9,905 barcode records assigned to the 52 BINs. This analysis showed that 27 of the 52 BINs were recorded from four or more continents while eight were unique to Pakistan (Table 3). Except for *Acyrtosiphon malvae* and *Uroleucon sonchi*, all named species (40) analyzed in this study already had barcode records from multiple countries and continents (Table 3).

Discussion

Prior morphological surveys on the aphids of Pakistan have reported the presence of nearly 300 species [70–72]. Most of this work focused on specific geographic regions [73] or species attacking crops [74,75]. The current study surveyed aphids across major agricultural areas of Pakistan from a wider range of host plants, but primarily aimed to develop a barcode reference library for the fauna for the first time. Prior studies have begun to construct barcode reference libraries for some pest insect groups, such as aphids in Canada [29], leafminers in USA [76], fruit flies in Africa [45], food pests in Korea [49], thrips in Pakistan [28], looper moths in British Columbia [77], and mealybugs in China [52]. These libraries have stimulated the use of DNA barcoding in biosecurity and plant protection programs [78], but their use revealed the need for expanded parameterization of the libraries in order to improve their utility in diagnosing newly encountered species. Barcode libraries for two major pest insect groups in Pakistan, thrips and whiteflies, have progressed well [28,79], but other groups have seen little attention in this country so far. The current study not only expands on the prior efforts by barcoding another group of insect pests but also maps the global presence of pest aphids by using BINs.

Most aphids analyzed in this study were assigned to a species, but 35 specimens could only be determined to genus or subfamily level. In part, this difficulty reflected the fact that many important pest aphids are cryptic species complexes whose members are almost impossible to discriminate using morphological traits only [42] or their identification was beyond our expertise. For example, *Aphis gossypii* is a particularly challenging species complex [5,13]; it includes at least 18 morphologically indistinguishable species [80] likely explaining its wide range of primary and secondary host plants [81]. In the present study, DNA barcoding separated all eight species of the genus *Aphis* that were encountered. Although K2P distances between two species pairs; i) *A. affinis* and *A. gossypii* (1.4%), ii) *A. astragalina* and *A. craccivora* (0.8%) were low, both NJ analysis and Bayesian inference supported the monophyly of each species. The COI divergences in this study are similar to those reported in prior investigations

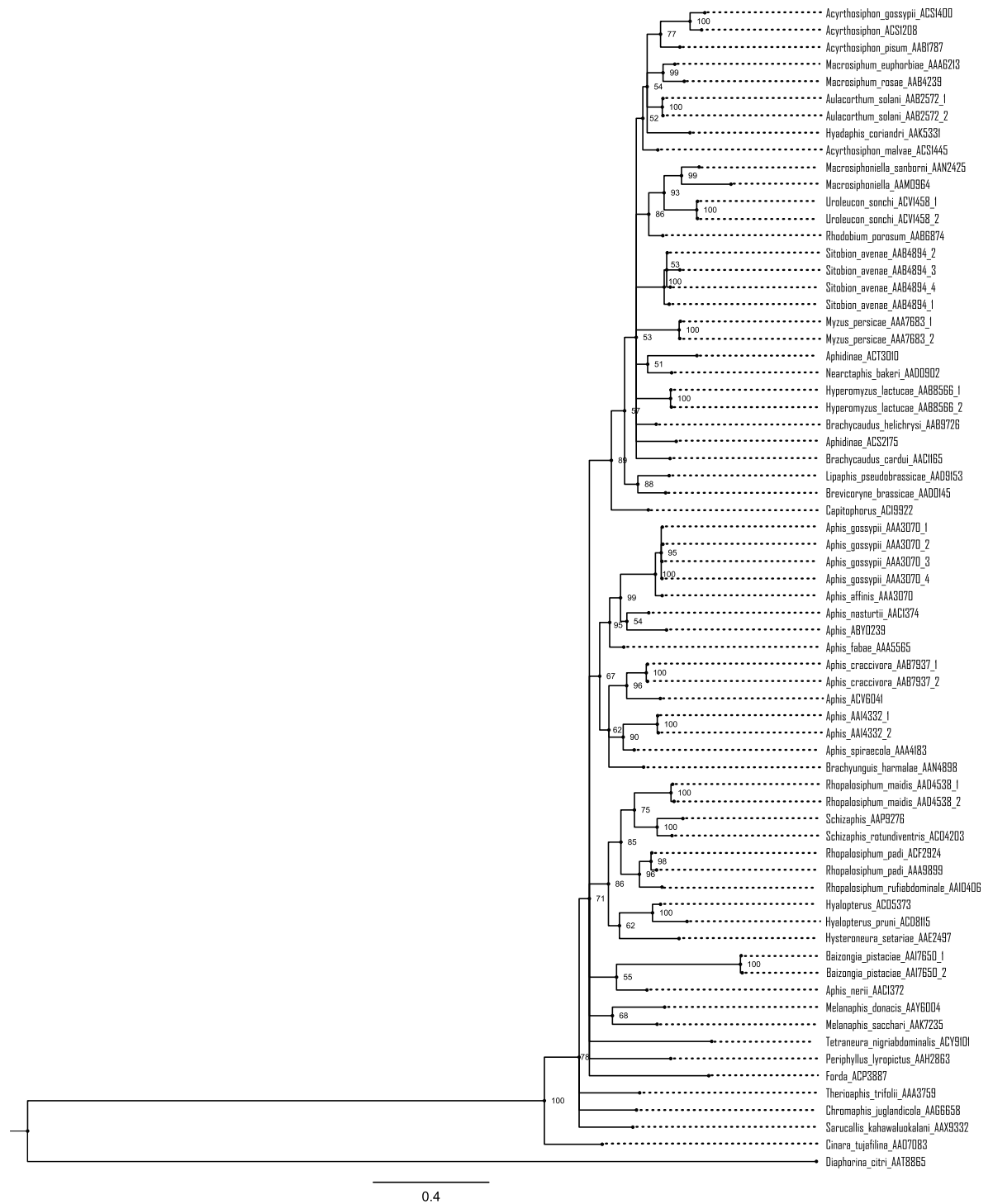


Fig 4. Phylogenetic analysis of aphid species/BINs from Pakistan based on COI-5' sequences. The tree was estimated using Bayesian inference. Posterior probabilities are indicated at nodes. The analysis was based on representative sequences from 67 aphid haplotypes in the dataset that were extracted using DnaSP v5.10 (Librado and Rozas 2009). Taxa are followed by the BINs and haplotype numbers. *Diaphorina citri* (BOLD:AAT8865) was employed as outgroup.

<https://doi.org/10.1371/journal.pone.0220426.g004>

Table 2. Geographic (km) and genetic (K2P) distance correlation analysis for 42 aphid species from Pakistan combined with conspecifics from 69 other countries.

Species	Record Count	BINs	Linear Regression R ²	Gen Dist Max	Geo Dist Max	Mantel R ²	Mantel P value
<i>Acyrtosiphon malvae</i>	54	2	0.13	4.8	18675	0.13	0.01
<i>Acyrtosiphon pisum</i>	205	1	0.01	1.7	19312	0.01	0.05
<i>Aphis affinis</i>	8	1	0.18	0.2	356	0.18	0.04
<i>Aphis astragalina</i>	37	1	0.71	0.8	10789	0.71	0.01
<i>Aphis craccivora</i>	420	1	0.09	1.4	19417	0.00	0.01
<i>Aphis fabae</i>	426	1	0.01	2.5	19456	0.01	0.01
<i>Aphis gossypii</i>	362	1	0.00	3.9	19369	0.00	0.8
<i>Aphis nasturtii</i>	38	1	0.48	1.6	11656	0.48	0.01
<i>Aphis nerii</i>	99	1	0.02	1.2	19110	0.02	0.01
<i>Aphis spiraeola</i>	277	2	0.00	3.1	19355	0.00	0.2
<i>Aulacorthum solani</i>	118	1	0.13	2.0	19291	0.13	0.01
<i>Baizongia pistaciae</i>	5	2	0.48	4.4	5967	0.48	0.23
<i>Brachycaudus cardui</i>	54	1	0.17	0.9	11929	0.17	0.01
<i>Brachycaudus helichrysi</i>	108	4	0.01	3.1	19426	0.01	0.01
<i>Brevicoryne brassicae</i>	166	2	0.02	6.3	19178	0.02	0.04
<i>Chromaphis juglandicola</i>	11	1	0.52	0.8	10908	0.52	0.01
<i>Hyadaphis coriandri</i>	13	1	0.40	1.4	685	0.40	0.01
<i>Hyalopterus pruni</i>	151	3	0.06	6.6	16867	0.06	0.01
<i>Hyperomyzus lactucae</i>	87	1	0.05	0.3	19474	0.05	0.02
<i>Hysteroneura setariae</i>	53	1	0.04	1.9	15843	0.04	0.01
<i>Lipaphis pseudobrassicae</i>	91	2	0.12	3.6	18212	0.13	0.01
<i>Macrosiphoniella sanborni</i>	5	1	0.01	0.2	15935	0.01	0.43
<i>Macrosiphum euphorbiae</i>	198	1	0.14	1.6	19482	0.14	0.01
<i>Macrosiphum rosae</i>	80	1	0.00	1.2	19379	0.00	0.14
<i>Melanaphis donacis</i>	7	1	0.16	0.2	6092	0.16	0.35
<i>Melanaphis sacchari</i>	225	1	0.02	1.9	18400	0.02	0.01
<i>Myzus persicae</i>	322	1	0.00	2.2	19234	0.00	0.68
<i>Nearctaphis bakeri</i>	70	2	0.04	6.8	11961	0.04	0.09
<i>Periphyllus lyropictus</i>	33	1	0.00	0.2	11002	0.00	0.92
<i>Rhodobium porosum</i>	7	1	0.08	0.2	12350	0.08	0.1
<i>Rhopalosiphum maidis</i>	63	1	0.00	2.0	18405	0.00	0.40
<i>Rhopalosiphum padi</i>	1189	2	0.30	5.0	19841	0.29	0.01
<i>Rhopalosiphum rufiabdominale</i>	18	1	0.03	0.2	15787	0.03	0.99
<i>Sitobion avenae</i>	314	1	0.02	4.6	18405	0.00	0.01
<i>Tetraneura nigriabdominalis</i>	41	2	0.16	8.6	16534	0.16	0.01
<i>Therioaphis trifolii</i>	470	2	0.00	13.0	18823	0.00	0.33
<i>Uroleucon sonchi</i>	45	2	0.11	2.3	19003	0.11	0.03
<i>Acyrtosiphon gossypii</i>	5	1	N/A	N/A	N/A	N/A	N/A
<i>Brachyunguis harmalae</i>	8	1	N/A	N/A	N/A	N/A	N/A
<i>Cinara tujafilina</i>	8	1	N/A	N/A	N/A	N/A	N/A
<i>Sarucallis kahawaluokalani</i>	10	1	N/A	N/A	N/A	N/A	N/A
<i>Schizaphis rotundiventris</i>	5	1	N/A	N/A	N/A	N/A	N/A

N/A: Data for the correlation analysis was missing.

<https://doi.org/10.1371/journal.pone.0220426.t002>

[29,82,83] which reported low sequence divergence between sibling species such as *A. gossypii* and *A. affinis* [29].

Table 3. Occurrence of 52 pest aphid BINs across six continents and their association with Linnaean species on the Barcode of Life Data System (BOLD).

BIN	Countries	Continents	(Number) and names of the associated species
AAA3070	44	6	(35) <i>Aphis affinis</i> , <i>A. aliena</i> , <i>A. argrimoniae</i> , <i>A. cf. frangulae</i> , <i>A. chloris</i> , <i>A. cisticola</i> , <i>A. clerodendri</i> , <i>A. confusa</i> , <i>A. crepidis</i> , <i>A. egomae</i> , <i>A. frangulae</i> , <i>A. gossypii</i> , <i>A. hieracii</i> , <i>A. hypericiphaga</i> , <i>A. hypochoeridis</i> , <i>A. idaei</i> , <i>A. leontodontis</i> , <i>A. lichtensteini</i> , <i>A. longirostrata</i> , <i>A. madderiae</i> , <i>A. mamonthovae</i> , <i>A. monardae</i> , <i>A. nivalis</i> , <i>A. oestlundii</i> , <i>A. origani</i> , <i>A. parietariae</i> , <i>A. punicae</i> , <i>A. ruborum</i> , <i>A. sedi</i> , <i>A. serpylli</i> , <i>A. sumire</i> , <i>A. taraxacicola</i> , <i>A. teucritii</i> , <i>A. viticis</i>
AAA3759	15	6	(1) <i>Therioaphis trifolii</i>
AAA4183	29	6	(1) <i>Aphis spiraeicola</i>
AAA5565	31	6	(9) <i>Aphis fabae</i> , <i>A. solanella</i> , <i>A. hederiae</i> , <i>A. ilicis</i> , <i>A. viburni</i> , <i>A. newtoni</i> , <i>A. fukii</i> , <i>A. lambersi</i> , <i>A. seselii</i>
AAA6213	19	6	(13) <i>Macrosiphum albifrons</i> , <i>M. cerinthiacum</i> , <i>M. cholodkovskiyi</i> , <i>M. corydalis</i> , <i>M. daphnidis</i> , <i>M. euphorbiae</i> , <i>M. gaurae</i> , <i>M. gei</i> , <i>M. hellebori</i> , <i>M. impatientis</i> , <i>M. sileneum</i> , <i>M. valerianae</i> , <i>M. zionense</i>
AAA7683	22	6	(1) <i>Myzus persicae</i>
AAA9899	16	4	(1) <i>Rhopalosiphum padi</i>
AAB1787	19	6	(1) <i>Acyrtosiphon pisum</i>
AAB2572	16	5	(2) <i>Aulacorthum solani</i> , <i>Macrosiphum gei</i>
AAB4239	20	4	(3) <i>Macrosiphum rosae</i> , <i>M. funestum</i> , <i>Sitobion rosivorum</i>
AAB4894	17	5	(1) <i>Sitobion avenae</i>
AAB6874	10	5	(6) <i>Ericaphis scammelli</i> , <i>E. fimbriata</i> , <i>Rhodobium porosum</i> , <i>Wahlgreniella nervata</i> , <i>W. vaccinii</i> , <i>W. arbuti</i>
AAB7937	30	6	(8) <i>Aphis craccivora</i> , <i>A. masoni</i> , <i>A. intybi</i> , <i>A. rumicis</i> , <i>A. spiraeicola</i> , <i>A. tirucallis</i> , <i>A. coronillae</i> , <i>A. fabae</i>
AAB8566	20	5	(2) <i>Hyperomyzus lactucae</i> , <i>H. carduellinus</i>
AAB9726	14	6	(1) <i>Brachycaudus helichrysi</i>
AAC1165	13	5	(2) <i>Brachycaudus cardui</i> , <i>B. lateralis</i>
AAC1372	22	6	(1) <i>Aphis nerii</i>
AAC1374	8	3	(5) <i>Aphis nasturtii</i> , <i>A. davletshinae</i> , <i>A. umbrella</i> , <i>A. althaeae</i> , <i>A. cf. rostella</i>
AAD0145	18	6	(1) <i>Brevicoryne brassicae</i>
AAD0902	4	3	(1) <i>Nearctaphis bakeri</i>
AAD4538	12	6	(1) <i>Rhopalosiphum maidis</i>
AAD9153	11	6	(2) <i>Lipaphis pseudobrassicae</i> , <i>L. erysimi</i>
AAE2497	13	5	(1) <i>Hysteroneura setariae</i>
AAG3896	14	5	(1) <i>Tetraneura nigriabdominalis</i>
AAG6658	6	3	(1) <i>Chromaphis juglandicola</i>
AAH2863	4	3	(1) <i>Periphyllus lyropictus</i>
AAI0406	13	5	(1) <i>Rhopalosiphum rufiabdominale</i>
AAI4332	3	2	(NA) Identified to genus— <i>Aphis</i>
AAI7650	2	1	(1) <i>Baizongia pistaciae</i>
AAK5331	3	2	(1) <i>Hyadaphis coriandri</i>
AAK7235	22	5	(2) <i>Melanaphis sacchari</i> , <i>M. japonica</i>
AAM0964	7	4	(2) <i>Macrosiphoniella yomogifoliae</i> , <i>M. abrotani</i>
AAN2425	5	4	(1) <i>Macrosiphoniella sanborni</i>
AAN4898	2	2	(1) <i>Brachyunguis harmalae</i>
AAO7083	6	3	(1) <i>Cinara tujafilina</i>
AAP9276	5	3	(NA) Identified to genus— <i>Schizaphis</i>
AAX9332	4	3	(1) <i>Sarucallis kahawaluokalani</i>
AAZ6004	4	2	(1) <i>Melanaphis donacis</i>
ACD8115	2	1	(1) <i>Hyalopterus pruni</i>
ACF2924	11	6	(1) <i>Rhopalosiphum padi</i>
ACI9922	5	3	(NA) Identified to genus— <i>Capitophorus</i>
ACO4203	4	2	(1) <i>Schizaphis rotundiventris</i>
ACO5373	2	1	(1) <i>Hyalopterus pruni</i>
ACS1400	2	1	(1) <i>Acyrtosiphon gossypii</i>

(Continued)

Table 3. (Continued)

BIN	Countries	Continents	(Number) and names of the associated species
ABY0239	1	1	(NA) Identified to genus– <i>Aphis</i>
ACP3887	1	1	(NA) Identified to genus– <i>Forda</i>
ACS1208	1	1	(NA) Identified to genus– <i>Acyrtosiphon</i>
ACS1445	1	1	(1) <i>Acyrtosiphon malvae</i>
ACS2175	1	1	(NA) Identified to subfamily– <i>Aphidinae</i>
ACT3010	1	1	(NA) Identified to subfamily– <i>Aphidinae</i>
ACV1458	1	1	(1) <i>Uroleucon sonchi</i>
ACV6041	1	1	(NA) Identified to genus– <i>Aphis</i>

<https://doi.org/10.1371/journal.pone.0220426.t003>

Prior studies revealed a strong correspondence between BINs and known species [39], in particular when reference specimens are identified by experts [84]. The same pattern was apparent in this study as 38 of 41 species were assigned to a single BIN. There were only two exceptions; *R. padi* was assigned to two BINs and *A. gossypii*–*A. affinis* were assigned to the same BIN. By comparison, when barcode sequences from conspecific specimens from other countries were considered, 12 of the 42 species showed a BIN split, an outcome which likely indicates incorrectly identified specimens [39]. Interestingly, the BIN (AAA3070) shared by specimens of *A. gossypii* and *A. affinis* from Pakistan included 31 additional species names when all records for it on BOLD were considered. Misidentifications and overlooked cryptic species may often cause conflicts between BIN and species morphology [85], but this can only be resolved by detailed taxonomic studies [86]. As well, heteroplasmy, hybridization, and incomplete lineage sorting can also cause BIN-morphology conflicts [87,88]. Furthermore, host affinities of sympatric populations, which have been observed in aphids, also expand intraspecific divergence [89], possibly resulting in BIN splits as we observed in *R. padi*.

Geo-distance correlations showed that the genetic divergence increased with geographic distance in almost half of the aphid species. Interestingly, the inclusion of conspecific sequences from other regions also increased the incidence of BIN splits. Since these analyses included all the conspecific sequences on BOLD, this outcome may reflect taxonomic errors [90]. Although spatial variation in conspecific sequences sometimes leads to increased intra-specific divergence values [91], it is usually too low to reduce the capacity of DNA barcodes to deliver reliable species identifications [47,92].

BINs are valuable in evaluating the geographic range of aphid species because they circumvent taxonomic uncertainties. In addition, BINs are gaining increased use to estimate species numbers [41] and to understand their distributions [52]. This analysis revealed that 27 of the 44 BINs with prior records on BOLD occurred on four or more continents, highlighting the broad ranges of many pest aphids. For example, BINs for *Aphis fabae* (black bean aphid), *A. nerii* (oleander aphid), *A. craccivora* (groundnut aphid), *Acyrtosiphon pisum* (pea aphid), *Brahycaudus helichrysi* (plum aphid), *Brevicoryne brassicae* (cabbage aphid), *L. pseudobrassicae* (turnip aphid), *R. padi* (oat aphid), *R. maidis* (corn aphid), *Macrosiphum euphorbiae* (potato aphid), *M. persicae* (peach aphid), and *Therioaphis trifolii* (alfalfa aphid) were all recorded from six continents. Interestingly, BINs associated with some of these species were also linked with other species on BOLD. For instance, AAA3070 was linked to 33 other species of *Aphis* while AAA6213 was associated with 13 species of *Macrosiphum*, and AAA5565 with nine species of *Aphis*. Although some of these cases may involve BIN sharing by different species [29], most cases likely reflect misidentifications.

The level of BIN overlap between the aphid fauna of Pakistan is much higher (85%) than levels for moths (44%) [93] and spiders (24%) [94]. This difference, may, be due, in part, to the

fact that the winged alates of aphids can disperse long distances and their dispersal capacity with the broad availability of the crop plants that they attack [95]. Consequently, the number of aphid species known from Europe has increased by 20% in the last 30 years [96] reflecting their transport on produced fruits [52], coupled with shifting environmental regimes. Reports suggest that with every 1°C increase, some 15 additional aphid species were recorded in Europe [97]. In North America, about 18% of all aphid species are introduced, and nearly half are plant pests [98]. Rapid developments in DNA sequencing are enabling the documentation of pest species and their distribution across the globe, but conflicts between taxonomic assignments and sequences have limited the full utility of these data. Given this difficulty, the BIN system provides an alternative path to document and track the pest species on a planetary scale.

Supporting information

S1 Table. Plant-host family range for aphid species/BINs collected in Pakistan.
(XLSX)

S2 Table. Identification and BIN assignment of 809 specimens of Aphididae collected from 123 plant hosts in Pakistan.
(XLSX)

Acknowledgments

We are very grateful to C Favret, University of Montreal, Quebec, Canada, for his help in aphid identification. We thank staff at the CCDB for aiding the sequence analysis which was made possible by a grant from Genome Canada and Ontario Genomics in support of the International Barcode of Life (iBOL) project. This study was also enabled by support from the Canada First Research Excellence Fund provided to the Food From Thought research program.

Author Contributions

Conceptualization: Muhammad Tayyib Naseem, Muhammad Ashfaq, Paul D. N. Hebert.

Data curation: Muhammad Tayyib Naseem, Muhammad Ashfaq, Paul D. N. Hebert.

Formal analysis: Muhammad Tayyib Naseem, Muhammad Ashfaq.

Funding acquisition: Muhammad Ashfaq, Paul D. N. Hebert.

Investigation: Muhammad Tayyib Naseem, Muhammad Ashfaq, Akhtar Rasool, Paul D. N. Hebert.

Methodology: Muhammad Tayyib Naseem, Muhammad Ashfaq, Arif Muhammad Khan, Akhtar Rasool.

Project administration: Muhammad Ashfaq, Muhammad Asif, Paul D. N. Hebert.

Resources: Muhammad Ashfaq, Paul D. N. Hebert.

Software: Muhammad Tayyib Naseem, Muhammad Ashfaq, Arif Muhammad Khan.

Supervision: Muhammad Ashfaq, Muhammad Asif, Paul D. N. Hebert.

Validation: Muhammad Tayyib Naseem, Muhammad Ashfaq, Paul D. N. Hebert.

Visualization: Muhammad Tayyib Naseem.

Writing – original draft: Muhammad Tayyib Naseem, Arif Muhammad Khan, Akhtar Rasool, Muhammad Asif.

Writing – review & editing: Muhammad Ashfaq, Paul D. N. Hebert.

References

1. Braendle C, Davis GK, Brisson JA, Stern DL. Wing dimorphism in aphids. *Heredity*. 2006; 97: 192. <https://doi.org/10.1038/sj.hdy.6800863> PMID: 16823401
2. Ogawa K, Miura T. Aphid polyphenisms: trans-generational developmental regulation through viviparity. *Front Physiol*. 2014; 5: 1. <https://doi.org/10.3389/fphys.2014.00001> PMID: 24478714
3. Favret C. Aphid species file version 5.0/5.0 [2019.02]. <http://aphid.speciesfile.org/>.
4. Blackman R, Eastop V. Aphids on the world's plants: An online identification and information guide; 2014.
5. Blackman RL, Eastop VF. Taxonomic issues. In: Emden H F van, Harrington R, editors. *Aphids as crop pests*: CAB International; 2007. pp. 1–27.
6. Dixon AFG, Kindlmann P, Leps J, Holman J. Why there are so few species of aphids especially in the tropics. *Am Nat*. 1987; 129: 580–592.
7. Sewell GH, Storch RH, Manzer FE, Forsythe HY Jr. The relationship between coccinellids and aphids in the spread of potato leafroll virus in a greenhouse. *Am J Potato Res*. 1990; 67: 865–868.
8. Mosco MC, Arduino P, Bullini L, Barbagallo S. Genetic heterogeneity, reproductive isolation and host preferences in mealy aphids of the *Hyalopterus pruni* complex (Homoptera, Aphidoidea). *Mol. Ecol*. 1997; 6: 667–670.
9. Carletto J, Blin A, Vanlerberghe-Masutti F. DNA-based discrimination between the sibling species *Aphis gossypii* Glover and *Aphis frangulae* Kaltentbach. *Syst Entomol*. 2009; 34: 307–314.
10. Sorensen JT. Aphids In: Cardé RT, Resh VH, editors. *Encyclopedia of insects*, 2nd edition. San Diego, USA: Academic Press; 2009. p. 27–31.
11. Raymond B, Searle JB, Douglas AE. On the processes shaping reproductive isolation in aphids of the *Aphis fabae* (scop.) complex (Aphididae: Homoptera). *Biol J Linnean Soc*. 2001; 74: 205–215.
12. Brisson JA. Aphid wing dimorphisms: linking environmental and genetic control of trait variation. *Philosophical Transactions of the Royal Society of London Series B, Biol Sci*. 2010; 365(1540): 605–16.
13. Jayaseelan M, Roesler Uwe R. MALDI-TOF MS Profiling-Advances in species identification of pests, parasites, and vectors. *Front Cell Infect Microbiol*. 2017; 7: 184. <https://doi.org/10.3389/fcimb.2017.00184> PMID: 28555175
14. Naaum AM, Footitt RG, Maw HEL, Hanner R. Real-time PCR for identification of the soybean aphid, *Aphis glycines* Matsumura. *J Appl Entomol*. 2014; 138: 485–9.
15. Kinyanjui G, Khamis FM, Mohamed S, Ombura LO, Warigia M, Ekesi S. Identification of aphid (Hemiptera: Aphididae) species of economic importance in Kenya using DNA barcodes and PCR-RFLP-based approach. *Bull Entomol Res*. 2016; 106: 63–72. <https://doi.org/10.1017/S0007485315000796> PMID: 26490301
16. Nunez JCB, Oleksiak MF. A cost-effective approach to sequence hundreds of complete mitochondrial genomes. *PLOS ONE*. 2016; 11: e0160958. <https://doi.org/10.1371/journal.pone.0160958> PMID: 27505419
17. Caterino MS, Cho S, Sperling FA. The current state of insect molecular systematics: A thriving Tower of Babel. *Annu Rev Entomol*. 2000; 45: 1–54. <https://doi.org/10.1146/annurev.ento.45.1.1> PMID: 10761569
18. Pons J, Barraclough TG, Gomez-Zurita J, Cardoso A, Duran DP, Hazell S, et al. Sequence-based species delimitation for the DNA taxonomy of undescribed insects. *Syst Biol*. 2006; 55: 595–609. <https://doi.org/10.1080/10635150600852011> PMID: 16967577
19. Erlandson M, Braun L, Baldwin D, Soroka J, Ashfaq M, Hegedus D. Molecular markers for *Peristenus* spp. (Hymenoptera: Braconidae) parasitoids associated with *Lygus* spp. (Hemiptera: Miridae). *Can Entomol*. 2003; 135: 71–83.
20. Ashfaq M, Ara J, Noor AR, Hebert PDN, Mansoor S. Molecular phylogenetic analysis of a scale insect (*Drosicha mangiferae*; Hemiptera: Monophlebidae) infesting mango orchards in Pakistan. *Eur J Entomol*. 2011; 108: 553–559.
21. Chen R, Jiang L-Y, Qiao G-X. The effectiveness of three regions in mitochondrial genome for aphid DNA barcoding: A case in Lachninae. *PLOS ONE*. 2012; 7: e46190. <https://doi.org/10.1371/journal.pone.0046190> PMID: 23056258

22. Ashfaq M, Noor AR, Mansoor S. DNA-based characterization of an invasive mealybug (Hemiptera: Pseudococcidae) species damaging cotton in Pakistan. *Appl Entomol Zool.* 2010; 45: 395–404.
23. Liu J, Jiang J, Song S, Tornabene L, Chabarria R, Naylor GJP, et al. Multilocus DNA barcoding—species identification with multilocus data. *Sci Rep.* 2017; 7: 16601. <https://doi.org/10.1038/s41598-017-16920-2> PMID: 29192249
24. Hebert PDN, Cywinska A, Ball SL. Biological identifications through DNA barcodes. *Proceedings of the Royal Society of London B: Biol Sci.* 2003; 270: 313–321.
25. Hebert PDN, Ratnasingham S, de Waard JR. Barcoding animal life: Cytochrome *c* oxidase subunit 1 divergences among closely related species. *Proceedings of the Royal Society of London B: Biol Sci.* 2003; 270(Suppl 1): S96–S99.
26. Ander M, Troell K, Chirico J. Barcoding of biting midges in the genus *Culicoides*: a tool for species determination. *Med Vet Entomol.* 2013; 27: 323–31. <https://doi.org/10.1111/j.1365-2915.2012.01050.x> PMID: 23106166
27. Adams M, Raadik TA, Burrige CP, Georges A. Global biodiversity assessment and hyper-cryptic species complexes: more than one species of elephant in the room? *Syst Biol.* 2014; 63: 518–33. <https://doi.org/10.1093/sysbio/syu017> PMID: 24627185
28. Iftikhar R, Ashfaq M, Rasool A, Hebert PDN. DNA barcode analysis of thrips (Thysanoptera) diversity in Pakistan reveals cryptic species complexes. *PLOS ONE.* 2016; 11: e0146014. <https://doi.org/10.1371/journal.pone.0146014> PMID: 26741134
29. Footitt RG, Maw HEL, Von Dohlen CD, Hebert PDN. Species identification of aphids (Insecta: Hemiptera: Aphididae) through DNA barcodes. *Mol Ecol Resour.* 2008; 8: 1189–1201. <https://doi.org/10.1111/j.1755-0998.2008.02297.x> PMID: 21586006
30. Liu Q-H, Jiang L-Y, Qiao G-X. DNA barcoding of Greenideinae (Hemiptera: Aphididae) with resolving taxonomy problems. *Invertebr Syst.* 2013; 27: 428–438.
31. Cocuzza GE, Cavalieri V. Identification of aphids of *Aphis frangulae*-group living on Lamiaceae species through DNA barcode. *Mol Ecol Resour.* 2014; 14: 447–57. <https://doi.org/10.1111/1755-0998.12199> PMID: 24188728
32. Gwiazdowski RA, Footitt RG, Maw HEL, Hebert PDN. The Hemiptera (Insecta) of Canada: constructing a reference library of DNA barcodes. *PLOS ONE.* 2015; 10: e0125635. <https://doi.org/10.1371/journal.pone.0125635> PMID: 25923328
33. Sow A, Brévault T, Benoit L, Chapuis M-P, Galan M, Coeur d'acier A, et al. Deciphering host-parasitoid interactions and parasitism rates of crop pests using DNA metabarcoding. *Sci Rep.* 2019; 9: 3646. <https://doi.org/10.1038/s41598-019-40243-z> PMID: 30842584
34. Piper AM, Batovska J, Cogan NOI, Weiss J, Cunningham JP, Rodoni BC, et al. Prospects and challenges of implementing DNA metabarcoding for high-throughput insect surveillance. *GigaScience.* 2019; 8: 1–22.
35. Ritter CD, Häggqvist S, Karlsson D, Sääksjärvi IE, Muasya AM, Nilsson RH, et al. Biodiversity assessments in the 21st century: the potential of insect traps to complement environmental samples for estimating eukaryotic and prokaryotic diversity using high-throughput DNA metabarcoding. *Genome.* 2019; 62: 147–159. <https://doi.org/10.1139/gen-2018-0096> PMID: 30673361
36. Galimberti A, Ferri E, De Mattia F, Labra M, Casiraghi M. DNA barcoding: a six-question tour to improve users' awareness about the method. *Brief Bioinform.* 2010; 11: 440–53. <https://doi.org/10.1093/bib/bbq003> PMID: 20156987
37. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System (<http://www.barcodinglife.org>). *Mol Ecol Notes.* 2007; 7: 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x> PMID: 18784790
38. Ratnasingham S, Hebert PDN. A DNA-based registry for all animal species: The Barcode Index Number (BIN) system. *PLOS ONE.* 2013; 8: e66213. <https://doi.org/10.1371/journal.pone.0066213> PMID: 23861743
39. Mutanen M, Kekkonen M, Prosser SWJ, Hebert PDN, Kaila L. One species in eight: DNA barcodes from type specimens resolve a taxonomic quagmire. *Mol Ecol Resour.* 2015; 15: 967–984. <https://doi.org/10.1111/1755-0998.12361> PMID: 25524367
40. Ortiz AS, Rubio RM, Guerrero JJ, Garre MJ, Serrano J, Hebert PDN, et al. Close congruence between Barcode Index Numbers (BINs) and species boundaries in the Erebidae (Lepidoptera: Noctuoidea) of the Iberian Peninsula. *Biodivers data J.* 2017; 5: e19840.
41. Hebert PDN, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA, et al. Counting animal species with DNA barcodes: Canadian insects. *Phil Trans R Soc Lond B: Biol Sci.* 2016; 371: 1702.

42. Lee Y, Lee W, Kanturski M, Footitt RG, Akimoto SI, Lee S. Cryptic diversity of the subfamily Calaphidinae (Hemiptera: Aphididae) revealed by comprehensive DNA barcoding. PLOS ONE. 2017; 12: e0176582. <https://doi.org/10.1371/journal.pone.0176582> PMID: 28448639
43. Ashfaq M, Sabir JSM, El-Ansary HO, Perez K, Levesque-Beaudin V, Khan AM, et al. Insect diversity in the Saharo-Arabian region: Revealing a little-studied fauna by DNA barcoding. PLOS ONE. 2018; 13: e0199965. <https://doi.org/10.1371/journal.pone.0199965> PMID: 29985924
44. deWaard JR, Mitchell A, Keena MA, Gopurenko D, Boykin LM, Armstrong KF, et al. Towards a global barcode library for *Lymantria* (Lepidoptera: Lymantriinae) tussock moths of biosecurity concern. PLOS ONE. 2010; 5: e14280. <https://doi.org/10.1371/journal.pone.0014280> PMID: 21151562
45. Virgilio M, Jordaens K, Breman FC, Backeljau T, De Meyer M. Identifying insects with incomplete DNA barcode libraries, African fruit flies (Diptera: Tephritidae) as a test case. PLOS ONE. 2012; 7: e31581. <https://doi.org/10.1371/journal.pone.0031581> PMID: 22359600
46. Coeur d'Acier A, Cruaud A, Artige E, Genson G, Clamens AL, Pierre E, et al. DNA barcoding and the associated phylaphidb@se website for the identification of European aphids (Insecta: Hemiptera: Aphididae). PLOS ONE. 2014; 9: e97620. <https://doi.org/10.1371/journal.pone.0097620> PMID: 24896814
47. Ashfaq M, Akhtar S, Khan AM, Adamowicz SJ, Hebert PDN. DNA barcode analysis of butterfly species from Pakistan points towards regional endemism. Mol Ecol Resour. 2013; 13(5):832–843. <https://doi.org/10.1111/1755-0998.12131> PMID: 23789612
48. Raupach MJ, Hannig K, Moriniere J, Hendrich L. A DNA barcode library for ground beetles of Germany: the genus *Amara* Bonelli, 1810 (Insecta, Coleoptera, Carabidae). Zookeys. 2018; 759: 57–80.
49. Cho SY, Suh KI, Bae YJ. DNA barcode library and its efficacy for identifying food-associated insect pests in Korea. Entomol Res. 2013; 43: 253–261.
50. Ashfaq M, Hebert PDN. DNA barcodes for bio-surveillance: Regulated and economically important arthropod plant pests. Genome 2016; 59: 933–945. <https://doi.org/10.1139/gen-2016-0024> PMID: 27753511
51. Kirichenko NI, Petkoa VM, Magnoux CE, Lopez-Vaamonde C. Diversity and distribution of leaf mining insects on birches (*Betula* spp.) in Siberia. Entomol Rev. 2017; 97: 183–197.
52. Ren J-M, Ashfaq M, Hu X-N, Ma J, Liang F, Hebert PDN, et al. Barcode Index Numbers expedite quarantine inspections and aid the interception of nonindigenous mealybugs (Pseudococcidae). Biol Invasions 2018; 20: 449–460.
53. Footitt RG, Maw HEL, Pike KS. DNA barcodes to explore diversity in aphids (Hemiptera: Aphididae and Adelgidae). Redia 2009; 92: 87–91.
54. Lee W, Kim H, Lim J, Choi HR, Kim Y, Kim YS, et al. Barcoding aphids (Hemiptera: Aphididae) of the Korean Peninsula: Updating the global data set. Mol Ecol Resour. 2011; 11: 32–37. <https://doi.org/10.1111/j.1755-0998.2010.02877.x> PMID: 21429098
55. Blackman RL, Eastop VF. Aphids on the world's crop: An identification and information guide, 2nd Edition. Wiley; 2000.
56. Martin JH. The identification of common aphid pests of tropical agriculture. Trop Pest Manag. 1983; 29: 395–411.
57. Ivanova NV, deWaard JR, Hebert PDN. An inexpensive, automation-friendly protocol for recovering high quality DNA. Mol Ecol Notes 2006; 6: 998–1002.
58. Porco D, Rougerie R, Deharveng L, Hebert P. Coupling non-destructive DNA extraction and voucher retrieval for small soft-bodied arthropods in a high-throughput context: The example of Collembola. Mol Ecol Resour. 2010; 10: 942–945. <https://doi.org/10.1111/j.1755-0998.2010.2839.x> PMID: 21565103
59. Hebert PDN, Dewaard JR, Zakharov EV, Prosser SW, Sones JE, McKeown JT, et al. A DNA 'barcode blitz': rapid digitization and sequencing of a natural history collection. PLOS ONE. 2013; 8: e68535. <https://doi.org/10.1371/journal.pone.0068535> PMID: 23874660
60. Park DS, Footitt R, Maw E, Hebert PDN. Barcoding bugs: DNA-based identification of the true bugs (Insecta: Hemiptera: Heteroptera). PLOS ONE. 2011; 6: e18749. <https://doi.org/10.1371/journal.pone.0018749> PMID: 21526211
61. Tamura K, Stecher G, Peterson D, Filipski A, Kumar S. MEGA6: Molecular evolutionary genetics analysis version 6.0. Mol Biol Evol. 2013; 30: 2725–2729. <https://doi.org/10.1093/molbev/mst197> PMID: 24132122
62. Meyer CP, Paulay G. DNA barcoding: Error rates based on comprehensive sampling. PLOS BIOL. 2005; 3: e422. <https://doi.org/10.1371/journal.pbio.0030422> PMID: 16336051
63. Thompson JD, Higgins DG, Gibson TJ. Clustal W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. 1994; 22: 4673–4680. <https://doi.org/10.1093/nar/22.22.4673> PMID: 7984417

64. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol.* 1980; 16: 111–120. <https://doi.org/10.1007/bf01731581> PMID: 7463489
65. Felsenstein J. Confidence limits on phylogenies: An approach using the bootstrap. *Evolution.* 1985; 39: 783–791. <https://doi.org/10.1111/j.1558-5646.1985.tb00420.x> PMID: 28561359
66. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. Mrbayes 3.2: Efficient Bayesian phylogenetic inference and model choice across a large model space. *Syst Biol.* 2012; 61: 539–542. <https://doi.org/10.1093/sysbio/sys029> PMID: 22357727
67. Librado P, Rozas J. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* 2009; 25: 1451–1452. <https://doi.org/10.1093/bioinformatics/btp187> PMID: 19346325
68. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 1967; 27: 209–220. PMID: 6018555
69. Blagoev GA, deWaard JR, Ratnasingham S, deWaard SL, Lu L, Robertson J, et al. Untangling taxonomy: a DNA barcode reference library for Canadian spiders. *Mol Ecol Resour.* 2016; 16: 325–341. <https://doi.org/10.1111/1755-0998.12444> PMID: 26175299
70. Naumann-Etienne K, Remaudiere G. A commented preliminary checklist of the aphids (Homoptera: Aphididae) of Pakistan and their host plants. *Parasitica* 1995; 51: 61.
71. Irshad M. Aphids and their biological control in Pakistan. *Pak J Biol Sci.* 2001; 4: 537–541.
72. Kanturski M, Zia A, Rafi MA. The *Lachnus* of Pakistan with description of a new species (Hemiptera: Aphididae: Lachninae). *J Asia-Pac Entomol.* 2017; 20: 1219–1227.
73. Bodlah I, Naeem M, Mohsin A. Checklist distribution host range and ecology of Aphidoidea (Homoptera) from the rainfed region of Punjab province of Pakistan, *Sarhad J Agric.* 2011; 27: 93–101.
74. Hamid S. Natural balance of graminicolous aphids in Pakistan. *Survey of populations.* *Agronomie* 1983; 3: 665–673.
75. Amin M, Mahmood K, Bodlah I. Aphid species (Hemiptera: Aphididae) infesting medicinal and aromatic plants in the Poonch division of Azad Jammu and Kashmir, Pakistan. *J Animal Plant Sci.* 2017; 27: 1377–1385.
76. Scheffer SJ, Lewis ML, Gaimari SD, Reitz SR. Molecular survey for the invasive leafminer pest *Liriomyza huidobrensis* (Diptera: Agromyzidae) in California uncovers only the native pest *Liriomyza langei*. *J Econ Entomol.* 2014; 107: 1959–64. <https://doi.org/10.1603/EC13279> PMID: 26309286
77. deWaard JR, Hebert PDN, Humble LM. A comprehensive DNA barcode library for the looper moths (Lepidoptera: Geometridae) of British Columbia, Canada. *PLOS ONE.* 2011; 6: e18290. <https://doi.org/10.1371/journal.pone.0018290> PMID: 21464900
78. Hodgetts J, Ostojca-Starzewski JC, Prior T, Lawson R, Hall J, Boonham N. DNA barcoding for biosecurity: case studies from the UK plant protection program. *Genome.* 2016; 59: 1033–48. <https://doi.org/10.1139/gen-2016-0010> PMID: 27792411
79. Ashfaq M, Hebert PDN, Mirza MS, Khan AM, Mansoor S, Shah GS, et al. DNA barcoding of *Bemisia tabaci* complex (Hemiptera: Aleyrodidae) reveals southerly expansion of the dominant whitefly species on cotton in Pakistan. *PLOS ONE.* 2014; 9: e104485. <https://doi.org/10.1371/journal.pone.0104485> PMID: 25099936
80. Lagos-Kutz D, Favret C, Giordano R, Voegtlin DJ. Molecular and morphological differentiation between *Aphis gossypii* Glover (Hemiptera, Aphididae) and related species, with particular reference to the North American Midwest. *ZooKeys.* 2014; 459: 49–72.
81. Singh G, Singh NP, Singh R, Singh G, Singh NP. Food plants of a major agricultural pest *Aphis gossypii* Glover (Homoptera: Aphididae) from India: An updated checklist. *Int J Life Sci Biotechnol Pharma Res.* 2014; 3: 1–26.
82. Wang JF, Qiao GX. DNA barcoding of genus *Toxoptera* Koch (Hemiptera: Aphididae): Identification and molecular phylogeny inferred from mitochondrial COI sequences. *Insect Sci.* 2009; 16: 475–484.
83. Cocuzza GE, Cavalieri V, Barbagallo S. Preliminary results in the taxonomy of the cryptic group *Aphis frangulae/gossypii* obtained from mitochondrial DNA sequence. *Bull Insectology.* 2008; 61: 125–126.
84. Janzen DH, Burns JM, Cong Q, Hallwachs W, Dapkey T, Manjunath R, et al. Nuclear genomes distinguish cryptic species suggested by their DNA barcodes and ecology. *Proc Natl Acad Sci USA.* 2017. <https://doi.org/10.1073/pnas.1621504114>.
85. Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W. Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. *Proc Natl Acad Sci USA.* 2004; 101: 14812–7. <https://doi.org/10.1073/pnas.0406166101> PMID: 15465915

86. Bortolus A. Error cascades in the biological sciences: the unwanted consequences of using bad taxonomy in ecology. *Ambio*. 2008; 37: 114–8. [https://doi.org/10.1579/0044-7447\(2008\)37\[114:ecitbs\]2.0.co;2](https://doi.org/10.1579/0044-7447(2008)37[114:ecitbs]2.0.co;2) PMID: 18488554
87. Kang AR, Kim MJ, Park IA, Kim KY, Kim I. Extent and divergence of heteroplasmy of the DNA barcoding region in *Anapodisma miramae* (Orthoptera: Acrididae). *Mitochondrial DNA A*. 2016; 27: 3405–3414.
88. Weber AA.-T, Stöhr S, Chenuil A. Species delimitation in the presence of strong incomplete lineage sorting and hybridization: Lessons from *Ophioderma* (Ophiuroidea: Echinodermata). *Mol Phylogenet Evol*. 2019; 131: 138–148. <https://doi.org/10.1016/j.ympev.2018.11.014> PMID: 30468939
89. Peccoud J, Ollivier A, Plantegenest M, Simon J-C. A continuum of genetic divergence from sympatric host races to species in the pea aphid complex. *Proc Natl Acad Sci USA*. 2009; 106: 7495–7500. <https://doi.org/10.1073/pnas.0811117106> PMID: 19380742
90. Virgilio M, Backeljau T, Nevado B, De Meyer M. Comparative performances of DNA barcoding across insect orders. *BMC Bioinformatics*. 2010; 11: 206. <https://doi.org/10.1186/1471-2105-11-206> PMID: 20420717
91. Bergsten J, Bilton DT, Fujisawa T, Elliott M, Monaghan MT, Balke M, et al. The effect of geographical scale of sampling on DNA barcoding. *Syst Biol*. 2012; 61: 851–69. <https://doi.org/10.1093/sysbio/sys037> PMID: 22398121
92. Huemer P, Mutanen M, Sefc KM, Hebert PDN. Testing DNA barcode performance in 1000 species of European Lepidoptera: large geographic distances have small genetic impacts. *PLOS ONE*. 2014; 9: e115774. <https://doi.org/10.1371/journal.pone.0115774> PMID: 25541991
93. Ashfaq M, Akhtar S, Rafi MA, Mansoor S, Hebert PDN. Mapping global biodiversity connections with DNA barcodes: Lepidoptera of Pakistan. *PLOS ONE*. 2017; 12: e0174749. <https://doi.org/10.1371/journal.pone.0174749> PMID: 28339501
94. Ashfaq M, Blagoev G, Tahir HM, Khan AM, Mukhtar MK, Akhtar S, et al. Assembling a DNA barcode reference library for the spiders (Arachnida: Araneae) of Pakistan. *PLOS ONE*. 2019; 14: e0217086. <https://doi.org/10.1371/journal.pone.0217086> PMID: 31116764
95. Irwin ME, Kampmeier GE, Weisser WW. Aphid movement: process and consequences. In: Emden HF, van Harrington R, editors. *Aphids as crop pests*. 2007; p. 153–186.
96. Harrington R, Verrier P, Denholm C, Hulle M, Maurice D, Bell N, et al. EXAMINE (Exploitation of Aphid Monitoring in Europe): an European thematic network for the study of global change impacts on aphids. In: Simon JC, Dedryver CA, Rispe C, Hulle M, editors. *Aphids in a New Millennium*. Science Update 2004. p. 45–49.
97. Hull M, Renoust M, Turpeau E. New aphid species detected by permanent aerial sampling programmes in France. In: Nieto Nafria JM, Dixon AEG, editors. *Aphids in natural and managed ecosystems*. Universidad de Leon, Leon, Spain. 1998. pp. 365–369.
98. Miller GL, Halbert SE, Foottit RG. North American adventive aphids: appraising the faunal components. 7th International Symposium on Aphids, Freemantle Australia. 2005.