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

Naseem, Muhammad Tayyib; Ashfaq, Muhammad; Khan, Arif Muhammad; Rasool, Akhtar; Asif, Muhammad; Hebert, Paul D N

Published in:
PLoS ONE

DOI:
10.1371/journal.pone.0220426

Publication date:
2019

Document version
Publisher's PDF, also known as Version of record

Document license:
CC BY

Citation for published version (APA):
https://doi.org/10.1371/journal.pone.0220426
BIN overlap confirms transcontinental distribution of pest aphids (Hemiptera: Aphididae)

Muhammad Tayyib Naseem1,2, Muhammad Ashfaq3*, Arif Muhammad Khan1,4, Akhtar Rasool1,5, 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

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
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
coordinates, the collection sites were rendered using SimpleMappr.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 μl 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-50 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 μL reactions that included standard PCR ingredients [59] and 2 μL 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 (TGTAAAACGACGCTAATCATAARGATATYGG) [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 v3.2.6 [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 FindModel (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>
<thead>
<tr>
<th>Distance class</th>
<th>n</th>
<th>Taxa</th>
<th>Comparisons</th>
<th>Min (%)</th>
<th>Mean (%)</th>
<th>Max (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within Species</td>
<td>764</td>
<td>36</td>
<td>30756</td>
<td>0</td>
<td>0.1</td>
<td>3.6</td>
</tr>
<tr>
<td>Within Genus</td>
<td>434</td>
<td>6</td>
<td>32305</td>
<td>0.8</td>
<td>7.4</td>
<td>10.3</td>
</tr>
<tr>
<td>Within Family</td>
<td>770</td>
<td>1</td>
<td>233004</td>
<td>3.7</td>
<td>9.6</td>
<td>17.3</td>
</tr>
</tbody>
</table>

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.

https://doi.org/10.1371/journal.pone.0220426.t001
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 (Acyrthosiphon, 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

![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)
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 *Acyrthosiphon 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:AAAT865) was employed as outgroup.

https://doi.org/10.1371/journal.pone.0220426.g004
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).

<table>
<thead>
<tr>
<th>BIN</th>
<th>Countries</th>
<th>Continents</th>
<th>Number</th>
<th>Name of the associated species</th>
</tr>
</thead>
<tbody>
<tr>
<td>AAA3759</td>
<td>15</td>
<td>6</td>
<td>(1)</td>
<td>Therioaphis trifolii</td>
</tr>
<tr>
<td>AAA4183</td>
<td>29</td>
<td>6</td>
<td>(1)</td>
<td>Aphis spiracola</td>
</tr>
<tr>
<td>AAA7683</td>
<td>22</td>
<td>6</td>
<td>(1)</td>
<td>Myzus persicae</td>
</tr>
<tr>
<td>AAA9899</td>
<td>16</td>
<td>4</td>
<td>(1)</td>
<td>Rhopalosiphum padi</td>
</tr>
<tr>
<td>AAB1787</td>
<td>19</td>
<td>6</td>
<td>(1)</td>
<td>Acrysaphion pisum</td>
</tr>
<tr>
<td>AAB2572</td>
<td>16</td>
<td>5</td>
<td>(2)</td>
<td>Aulacorthum solani, Macrosiphum gei</td>
</tr>
<tr>
<td>AAB4239</td>
<td>20</td>
<td>4</td>
<td>(3)</td>
<td>Macrosiphum rosae, M. funestum, Sitobion rosivorum</td>
</tr>
<tr>
<td>AAB8949</td>
<td>17</td>
<td>5</td>
<td>(1)</td>
<td>Sitobion avenae</td>
</tr>
<tr>
<td>AAB8972</td>
<td>14</td>
<td>6</td>
<td>(1)</td>
<td>Hyperomyzus lactucae, H. carduallinus</td>
</tr>
<tr>
<td>AAC1165</td>
<td>13</td>
<td>5</td>
<td>(2)</td>
<td>Brachycyclus cardui, B. lateralis</td>
</tr>
<tr>
<td>AAC1372</td>
<td>22</td>
<td>6</td>
<td>(1)</td>
<td>Aphis nerii</td>
</tr>
<tr>
<td>AAC1374</td>
<td>8</td>
<td>3</td>
<td>(5)</td>
<td>Aphis nasturtii, A. davlethsinae, A. umbrella, A. althaeae, A. cf. rostella</td>
</tr>
<tr>
<td>AAD0145</td>
<td>18</td>
<td>6</td>
<td>(1)</td>
<td>Brevicoryne brussicae</td>
</tr>
<tr>
<td>AAD0902</td>
<td>4</td>
<td>3</td>
<td>(1)</td>
<td>Nearchaphis bakeri</td>
</tr>
<tr>
<td>AAD4538</td>
<td>12</td>
<td>6</td>
<td>(1)</td>
<td>Rhopalosiphum maidis</td>
</tr>
<tr>
<td>AAD9153</td>
<td>11</td>
<td>6</td>
<td>(2)</td>
<td>Lipaphis pseudobrussicae, L. erysimi</td>
</tr>
<tr>
<td>AAE2497</td>
<td>13</td>
<td>5</td>
<td>(1)</td>
<td>Hysteroneura setariae</td>
</tr>
<tr>
<td>AAG3896</td>
<td>14</td>
<td>5</td>
<td>(1)</td>
<td>Tetraneura nigrabdominalis</td>
</tr>
<tr>
<td>AAG6658</td>
<td>6</td>
<td>3</td>
<td>(1)</td>
<td>Chromaphis juglandicola</td>
</tr>
<tr>
<td>AAH2863</td>
<td>4</td>
<td>3</td>
<td>(1)</td>
<td>Periphyllus lyopictus</td>
</tr>
<tr>
<td>AAI4046</td>
<td>13</td>
<td>5</td>
<td>(1)</td>
<td>Rhopalosiphum rufiabdominale</td>
</tr>
<tr>
<td>AAIH332</td>
<td>3</td>
<td>2</td>
<td>(NA)</td>
<td>Identified to genus-Aphis</td>
</tr>
<tr>
<td>AAI7650</td>
<td>2</td>
<td>1</td>
<td>(1)</td>
<td>Baizongia pistaciae</td>
</tr>
<tr>
<td>AAK3331</td>
<td>3</td>
<td>2</td>
<td>(1)</td>
<td>Hyadaphis coriandri</td>
</tr>
<tr>
<td>AAK7235</td>
<td>22</td>
<td>5</td>
<td>(2)</td>
<td>Melanaphis sacchari, M. japonica</td>
</tr>
<tr>
<td>AAM0964</td>
<td>7</td>
<td>4</td>
<td>(2)</td>
<td>Macrosiphoniella yomogifoliae, M. abrotani</td>
</tr>
<tr>
<td>AAN2425</td>
<td>5</td>
<td>4</td>
<td>(1)</td>
<td>Macrosiphoniella sanborni</td>
</tr>
<tr>
<td>AAN4898</td>
<td>2</td>
<td>2</td>
<td>(1)</td>
<td>Brachyuguis harmalae</td>
</tr>
<tr>
<td>AAO7083</td>
<td>6</td>
<td>3</td>
<td>(1)</td>
<td>Cinara tujaflina</td>
</tr>
<tr>
<td>AAP9276</td>
<td>5</td>
<td>3</td>
<td>(NA)</td>
<td>Identified to genus-Schizaphis</td>
</tr>
<tr>
<td>AAX5332</td>
<td>4</td>
<td>3</td>
<td>(1)</td>
<td>Saracallis kahawaluokalani</td>
</tr>
<tr>
<td>AAY6004</td>
<td>4</td>
<td>2</td>
<td>(1)</td>
<td>Melanaphis donacis</td>
</tr>
<tr>
<td>ACD8115</td>
<td>2</td>
<td>1</td>
<td>(1)</td>
<td>Hyalocephalus pruni</td>
</tr>
<tr>
<td>ACF2924</td>
<td>11</td>
<td>6</td>
<td>(1)</td>
<td>Rhopalosiphum padi</td>
</tr>
<tr>
<td>ACI9222</td>
<td>5</td>
<td>3</td>
<td>(NA)</td>
<td>Identified to genus-Capitophorus</td>
</tr>
<tr>
<td>ACO4203</td>
<td>4</td>
<td>2</td>
<td>(1)</td>
<td>Schizaphis rotundiventris</td>
</tr>
<tr>
<td>ACO5373</td>
<td>2</td>
<td>1</td>
<td>(1)</td>
<td>Hyalocephalus pruni</td>
</tr>
<tr>
<td>ACS1400</td>
<td>2</td>
<td>1</td>
<td>(1)</td>
<td>Acrystaphion gossypii</td>
</tr>
</tbody>
</table>

(Continued)
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 *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 intraspecific 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), *Acythosiphon pisum* (pea aphid), *Bachycaudus helichrysi* (plum aphid), *Breviscoryne brassicae* (cabbage aphid), *L. pseudobrassicae* (turnip aphid), *R. padi* (oat aphid), *R. maidis* (corn aphid), *Macrosiphum euphorbiae* (potato aphid), *M. persicae* (peach aphid), and *Theroaphis 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.

Table 3. (Continued)

<table>
<thead>
<tr>
<th>BIN</th>
<th>Countries</th>
<th>Continents</th>
<th>(Number) and names of the associated species</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABY0239</td>
<td>1</td>
<td>1</td>
<td>(NA) Identified to genus–<em>Aphis</em></td>
</tr>
<tr>
<td>ACP3887</td>
<td>1</td>
<td>1</td>
<td>(NA) Identified to genus–<em>Forda</em></td>
</tr>
<tr>
<td>ACS1208</td>
<td>1</td>
<td>1</td>
<td>(NA) Identified to genus–<em>Acythosiphon</em></td>
</tr>
<tr>
<td>ACS1445</td>
<td>1</td>
<td>1</td>
<td>(1) <em>Acythosiphon malvae</em></td>
</tr>
<tr>
<td>ACS2175</td>
<td>1</td>
<td>1</td>
<td>(NA) Identified to subfamily–<em>Aphidinae</em></td>
</tr>
<tr>
<td>ACT3010</td>
<td>1</td>
<td>1</td>
<td>(NA) Identified to subfamily–<em>Aphidinae</em></td>
</tr>
<tr>
<td>ACV1458</td>
<td>1</td>
<td>1</td>
<td>(1) <em>Uroleucon sonchi</em></td>
</tr>
<tr>
<td>ACV6041</td>
<td>1</td>
<td>1</td>
<td>(NA) Identified to genus–<em>Aphis</em></td>
</tr>
</tbody>
</table>

https://doi.org/10.1371/journal.pone.0220426.1003

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


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


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


