Analyzing Arthropod Biodiversity in Jinnah Garden, Lahore, Pakistan, Using DNA Barcoding (#94591)

First submission

Guidance from your Editor

Please submit by **29 Jan 2024** for the benefit of the authors (and your token reward) .

✓ **Structure and Criteria**
Please read the 'Structure and Criteria' page for general guidance.

▦ **Raw data check**
Review the raw data.

▨ **Image check**
Check that figures and images have not been inappropriately manipulated.

If this article is published your review will be made public. You can choose whether to sign your review. If uploading a PDF please remove any identifiable information (if you want to remain anonymous).

Files
Download and review all files from the materials page.

5 Figure file(s)
3 Table file(s)
5 Other file(s)

Structure and Criteria

2

Structure your review

The review form is divided into 5 sections. Please consider these when composing your review:

1. **BASIC REPORTING**
2. **EXPERIMENTAL DESIGN**
3. **VALIDITY OF THE FINDINGS**
4. General comments
5. Confidential notes to the editor

🗋 You can also annotate this PDF and upload it as part of your review

When ready submit online.

Editorial Criteria

Use these criteria points to structure your review. The full detailed editorial criteria is on your guidance page.

BASIC REPORTING

- ✓ Clear, unambiguous, professional English language used throughout.
- ✓ Intro & background to show context. Literature well referenced & relevant.
- ✓ Structure conforms to PeerJ standards, discipline norm, or improved for clarity.
- ✓ Figures are relevant, high quality, well labelled & described.
- ✓ Raw data supplied (see PeerJ policy).

EXPERIMENTAL DESIGN

- ✓ Original primary research within Scope of the journal.
- ✓ Research question well defined, relevant & meaningful. It is stated how the research fills an identified knowledge gap.
- ✓ Rigorous investigation performed to a high technical & ethical standard.
- ✓ Methods described with sufficient detail & information to replicate.

VALIDITY OF THE FINDINGS

- ℹ Impact and novelty not assessed. *Meaningful* replication encouraged where rationale & benefit to literature is clearly stated.
- ✓ All underlying data have been provided; they are robust, statistically sound, & controlled.
- ✓ Conclusions are well stated, linked to original research question & limited to supporting results.

Standout reviewing tips

The best reviewers use these techniques

Tip	***Example***
Support criticisms with evidence from the text or from other sources	*Smith et al (J of Methodology, 2005, V3, pp 123) have shown that the analysis you use in Lines 241-250 is not the most appropriate for this situation. Please explain why you used this method.*
Give specific suggestions on how to improve the manuscript	*Your introduction needs more detail. I suggest that you improve the description at lines 57- 86 to provide more justification for your study (specifically, you should expand upon the knowledge gap being filled).*
Comment on language and grammar issues	*The English language should be improved to ensure that an international audience can clearly understand your text. Some examples where the language could be improved include lines 23, 77, 121, 128 – the current phrasing makes comprehension difficult. I suggest you have a colleague who is proficient in English and familiar with the subject matter review your manuscript, or contact a professional editing service.*
Organize by importance of the issues, and number your points	*1. Your most important issue* *2. The next most important item* *3. …* *4. The least important points*
Please provide constructive criticism, and avoid personal opinions	*I thank you for providing the raw data, however your supplemental files need more descriptive metadata identifiers to be useful to future readers. Although your results are compelling, the data analysis should be improved in the following ways: AA, BB, CC*
Comment on strengths (as well as weaknesses) of the manuscript	*I commend the authors for their extensive data set, compiled over many years of detailed fieldwork. In addition, the manuscript is clearly written in professional, unambiguous language. If there is a weakness, it is in the statistical analysis (as I have noted above) which should be improved upon before Acceptance.*

Analyzing Arthropod Biodiversity in Jinnah Garden, Lahore, Pakistan, Using DNA Barcoding

Khush Bakhat Samreen Corresp., [1] , **Dr Farkhanda Manzoor Manzoor** [2]

[1] Department of Zoology, Lahore College for Women University, Lahore, Lahore, Pakistan

[2] Department of zoology, Lahore College for Women University, Lahore, Lahore, punjab, Pakistan

Corresponding Author: Khush Bakhat Samreen
Email address: cool.khushi19@gmail.com

Previous difficulties in arthropod taxonomy (such as limitations in conventional morphological approaches, the possibility of cryptic species and a shortage of knowledgeable taxonomists) has been overcome by the powerful tool of DNA barcoding. This study presents a thorough analysis of DNA barcoding in regards to Pakistani arthropods, which were collected from Lahore's Jinnah Garden. The 88 percent (9,451) of the 10,792 specimens that were examined were able to generate DNA barcodes and 83% (8,974) of specimens were assigned Barcode Index Numbers (BINs). However, the success rate differed significantly between the orders of arthropods, from 77% for Thysanoptera to an astounding 93% for Diptera. Through morphological exams, DNA barcoding, and cross-referencing with the Barcodei of Life Data system (BOLD), the Barcode iIndex Numbers (BINs) were assigned with a high degree of accuracy, both at the order (100%) and family (98%) levels. Though, identifications at the genus (37%) and species (15%) levels showed room for improvement. This underscores the ongoing need for enhancing and expanding the DNA barcode reference library. This study identified 324 genera and 191 species, underscoring the advantages of DNA barcoding over traditional morphological identification methods. Among the 17 arthropod orders identified, Coleoptera, Diptera, Hemiptera, Hymenoptera, and Lepidoptera from the class Insecta dominated, collectively constituting 94% of BINs. These results demonstrate that in Pakistani arthropods, DNA barcoding and BOLD are an invaluable tool for improving taxonomic understanding and biodiversity assessment, opening the door for further eDNA and metabarcoding research.

Analyzing Arthropod Biodiversity in Jinnah Garden, Lahore, Pakistan, Using DNA Barcoding

Khush Bakhat Samreen[1], Farkhanda Manzoor[1]

1- Department of Zoology, Lahore College for Women University, Lahore, Pakistan

Correspondence: cool.khushi19@gmail.com
Phone no: +92 336 4471371
Mailing address: house no 136, Street no 45, HassanPark, Ghorey Shah, Lahore, Pakistan

Author 2: Dr Farkhanda Manzoor
Email: drfarkhanda786@gmail.com
Phone no: +92 3334583936
Mailing Address: 281-B, Johar Town, Lahore, Pakistan

Abstract:

Previous difficulties in arthropod taxonomy (such as limitations in conventional morphological approaches, the possibility of cryptic species and a shortage of knowledgeable taxonomists) has been overcome by the powerful tool of DNA barcoding. This study presents a thorough analysis of DNA barcoding in regards to Pakistani arthropods, which were collected from Lahore's Jinnah Garden. The 88 percent (9,451) of the 10,792 specimens that were examined were able to generate DNA barcodes and 83% (8,974) of specimens were assigned Barcode Index Numbers (BINs). However, the success rate differed significantly between the orders of arthropods, from 77% for Thysanoptera to an astounding 93% for Diptera. Through morphological exams, DNA barcoding, and cross-referencing with the Barcode of Life Data system (BOLD), the Barcode Index Numbers (BINs) were assigned with a high degree of accuracy, both at the order (100%) and family (98%) levels. Though, identifications at the genus (37%) and species (15%) levels showed room for improvement. This underscores the ongoing need for enhancing and expanding the DNA barcode reference library. This study identified 324 genera and 191 species, underscoring the advantages of DNA barcoding over traditional morphological identification methods. Among the 17 arthropod orders identified, Coleoptera, Diptera, Hemiptera, Hymenoptera, and Lepidoptera from the class Insecta dominated, collectively constituting 94% of BINs. These results demonstrate that in Pakistani arthropods, DNA barcoding and BOLD are an invaluable tool for improving taxonomic understanding and biodiversity assessment, opening the door for further eDNA and metabarcoding research.

Key Words: DNA Barcoding, Arthropods taxonomy, Biodiversity assessment, Barcode of Life Data System (BOLD), Barcode Index Numbers (BINs)

INTRODUCTION:

The problems with arthropod taxonomy are not localized; they exist worldwide. When employed extensively, traditional morphological techniques—which have been utilized for many years to identify species—present a number of challenges. First off, it can be challenging to develop consistent and trustworthy identification criteria due to the significant morphological variation that some species can show within their populations (Hebert et al. 2003). Furthermore, it can be particularly difficult to detect cryptic species—that is, animals that share physical features but have unique genetic characteristics—using conventional methods (William et al. 2006). Moreover, phenotypic plasticity—the ability of physical traits to alter in response to environmental factors—is exhibited by many species. This might result in inconsistent morphological identification, since members of the same species may display distinct characteristics depending on their physical characteristics can change in response to environmental conditions. This can lead to inconsistencies in morphological identification, as individuals within the same species may exhibit different traits based on their environmental context (Moczek 2010). The lack of qualified taxonomists exacerbates these problems by potentially impeding prompt and precise species identification. Examining preserved specimens is often necessary for morphological identification; however, poor preservation practices, specimen damage, or insufficient preservation can mask or change key morphological characteristics, making identification more challenging (Cognato et al. 2020). Finally, it may become more difficult to differentiate between closely related species based alone on morphology due to convergent evolution, which is fueled by comparable environmental forces and can result in the formation of similar features in various species (Montealegre et al. 2012).

This emphasizes the need for different techniques to improve the efficiency and accuracy of species identification. Researchers frequently use molecular techniques, such as DNA barcoding, in conjunction with morphological identification to overcome these challenges and improve accuracy and reliability (Seberg et al. 2003). DNA barcoding is a widely used method in many different fields, such as: phylogenetic studies (Hajibabaei et al. 2007), taxonomic analysis (Dewalt 2011), looking at biodiversity of insect communities (Hlebec et al. 2022), examining genetic patterns (Zhou et al. 2010), phylogenetic analysis (Huang et al. 2016), and food authentication and safety (Dawan and Ahn 2022). Through the examination of particular genetic markers, such as the cytochrome oxidase I (COI) gene, which is highly conserved among species and resistant to homoplasy, DNA barcoding enables precise species identification (Hajibabaei et al. 2007). Moreover, this 658-base pair sequence, often referred to as the "DNA barcode," acts as a unique marker for species identification due to its significant sequence variation, which helps differentiate species that are closely related (Jinbo et al. 2011).

DNA barcoding's rapid adoption in modern biodiversity research (Hebert et al. 2003) has been powered by its use in specimen identification. This impressive efficacy of DNA barcoding in enabling thorough assessments of biodiversity is also demonstrated by the studies of Wilson et al. 2017, Shashank et al. 2022, and D'Souza et al. 2021. DNA barcoding has significantly advanced our comprehension of biological diversity by focusing on specific, consistent DNA sequences like the internal transcribed spacer (ITS) region. This method delivers high accuracy and dependability, even among species that are closely related, as evidenced by the research of Tyagi et al. 2019. The study of Wilkinson et al. 2017 shows that development of next-generation sequencing technologies has expedited the identification and discovery of previously unknown species, significantly increasing the speed and efficiency of DNA barcoding.

DNA barcodes are unique sequences that act like biological identification tags for species. These barcodes are central to the Barcode of Life Data System (BOLD), which is an open-access platform that simplifies the tasks of classifying species, identifying unknown specimens, and discovering new species. BOLD is an abundant resource that provides a multitude of DNA barcode records from many taxonomic categories. By making a large database of barcode sequences easily accessible, this technology expedites the process of identifying species and enables scientists and researchers to compare and evaluate genetic data from different organisms (Ratnasingham and Hebert 2007). BOLD is an essential informatics platform for biodiversity and evolutionary research, offering a user-friendly interface that simplifies the management and analysis of genetic data. BOLD integrates molecular, morphological, and distributional data, bridging gaps in bioinformatics and supporting global research collaborations. By adhering to stringent data standards, BOLD ensures the quality and reliability of genetic information, making it an invaluable resource for the scientific community (Ratnasingham and Herbert 2007).

A Barcode Index Number (BIN) is a unique identification number assigned to each species. Its species' unique DNA barcode sequence serves as the basis for its BIN. The cytochrome c oxidase subunit I (COI) gene, a standardized region of the genome with notable species variation, serves as the foundation for everything. This is where what makes starts, because that special BIN can be created with just a few hundred base pairs. As to the findings of Ratnasingham and Hebert (2013) each BIN is assigned to a particular species or genus, making the complex subject of taxonomy a little easier to navigate and comprehend.

BINs have vast scope beyond their significance in the discovery of new species. Ren et al. 2018 demonstrated that they are useful tools for tracking the locations of current species. Moreover, BINs are also useful for estimating species abundances within large samples, as Andújar et al.

2018 and Braukmann et al. 2019 have shown. It's similar to using a microscope to enlarge on the populations of various species concealed in large samples, allowing us to learn more about the complex web of life. In addition, the use of BINs for DNA barcoding has made it easier for researchers to examine museum collections and learn more about past biological assemblages (Pentinsaari et al. 2020). Furthermore, scientists can assess the degree of similarity or dissimilarity across those populations by comparing the BIN profiles of fauna from other locations and the world at large (Ashfaq et al. 2017), which advances our knowledge of trends in global biodiversity.

Keeping in view the importance of DNA barcoding, the current study significantly expands the scope of DNA barcoding for Pakistani insects, thereby advancing our understanding of the country's taxonomic biodiversity and laying the foundation for future eDNA and metabarcoding investigations.

MATERIALS AND METHODS

Sample collection and preparation

1- Collections

Specimens were collected at various sites across Jinnah garden, Lahore, Pakistan. Jinnah garden is a public park that covers an area of approximately 16 hectares (39 acres). In terms of topography, Jinnah Garden features a mix of flat and hilly areas, with several walking paths and walkways that crisscross the park.

Between 2019 and 2021, collections were conducted using various methods such as light traps, Malaise traps, sweep nets, and hand collections for larvae (Fig 1). The collected specimens were euthanized in cyanide jars, placed in paper envelopes, and then relaxed, pinned, labeled, and stored at the Entomological Laboratory within Lahore College for Women University.

2- Specimen identification

Morphological identifications were carried out at the Entomological Laboratory at Lahore College for Women University, with valuable insights provided by various taxonomic experts. To enhance accuracy, morphological identifications were cross-referenced whenever feasible by comparing the Pakistani specimens' barcode records with pre-existing records on BOLD. The images, Collected data and specimen details were submitted to BOLD and can be accessed through the dataset DS-GMPJA Malaise trap Jinnah Garden Lahore.

3- DNA barcoding / Molecular analysis

A total of 10792 insects were subjected to barcoding in Jinnah Garden, Lahore, following the established protocols (deWaard et al. 2019a, 2019b). Briefly, for larger specimens, a leg was carefully removed using sterile forceps and transferred to a well containing 30 ml of 95% ethanol. Smaller specimens, already on plates, were prepared for analysis, with vouchers retrieved after DNA extraction (Porco et al. 2010). At the Canadian Centre for DNA Barcoding (CCDB), we followed well-established procedures for DNA extraction, PCR amplification, and sequencing. These methods were described in previous publications (Ivanova et al. 2006, Hebert et al. 2018, deWaard et al.,2019b). Depending on the specific experiment, we used either six or twelve milliliters of material for the PCR reactions, as outlined in Hebert et al. 2013.

Using an Applied Biosystems 3730XL DNA Analyzer and the BigDye Terminator Cycle Sequencing Kit (v3.1), specimens were subjected to Sanger sequencing. Afterward, CodonCode Aligner was used to assemble, align, and modify the sequences before being submitted to BOLD. All DNA extracts are stored within the DNA archive facility at Centre for Biodiversity Genomics (CBG), Guelph, Canada.

4- **Data analysis**

The final dataset consisted of 8974 records, which received BINs and taxonomy assignments according to the workflow proposed by deWaard et al. 2019b. This involved a two-step process, where first, the barcode data was uploaded onto BOLD, and then each record underwent taxonomic assignment and verification. Morphological study by taxonomic specialists was also conducted alongside the molecular analysis to enhance species delimitation. Prior studies have shown the benefit of integrating both approaches, such as those by Silva-Brandão et al. 2009 on Lepidoptera, Blagoev et al. 2013 on Araneae (Theridioidea), and Naseem et al. 2019 on thrip specimens collected from cotton fields in Pakistan. These studies demonstrate the advantages of combining molecular and morphological techniques for accurate species identification. Only sequences that met the criteria of quality were either assigned to already existing BINs or used to create new ones, following the approach outlined by Ratnasingham and Hebert 2013.

To delineate new Barcode Index Numbers (BINs), the protocol necessitated adherence to stringent quality criteria. Eligibility for BIN classification required sequences to span at least 500 base pairs of the barcode region, specifically between positions 70 and 700 on the alignment of BOLD contain less than 1% ambiguous bases, and be devoid of stop codons or contamination indicators (Ratnasingham and Hebert 2013). Additionally, sequences of shorter length (300-495 base pairs) that met the quality standards—lacking ambiguous bases and stop codons—and demonstrated high similarity to an established BIN were consolidated under the corresponding BIN (deWaard et al. 2019a). Comprehensive BIN data, inclusive of specimen records and their images where available, are accessible through the BOLD interface at DS-GMPJA Malaise trap Jinnah Garden Lahore.

Complete BIN data, including sample records and representative

A "BIN discordance" analysis was employed to ascertain the proper BIN assignments within BOLD. Unassigned specimens underwent scrutiny via the BOLD Identification Engine (http://www.boldsystems.org/index.php/IDS_OpenIdEngine). Ensuing assignments underwent corroboration through the taxon ID tree to ensure accuracy. Sequences identified as contamination were consequently flagged, cataloged as such on BOLD, and excised from both the analysis and their associated BIN entries.

The "BIN discordance" report leveraged the comprehensive suite of functions within MS Excel to compute summary statistics. Furthermore, indices of species richness and evenness were evaluated using the "Diversity measure" function provided by BOLD.

RESULTS

DNA barcoding analysis of 10,792 specimens yielded successful results for 9,451 (88%), the remaining 12% were excluded from additional investigation because they were either unable to amplify or produced undesirable sequences (such as contamination, NUMTs, stop codons, or endosymbionts) (Supp. File 1 & 2). Sequence recovery varied widely among orders with 100 or more specimens, ranging from 77% for Thysanoptera to 93% for Diptera. There was significant difference in the sequence recovery for the other four major orders of insects, which are the Coleoptera (83%), Hymenoptera (81%), Hemiptera (80%), and Lepidoptera (72%).(Supp. Table S1)

Among the 9,541 successfully barcoded records, 8,974 (95%) were assigned to BINs (Barcode Index Numbers), leaving 477 records that did not qualify for BINs. These 477 sequences, not meeting BIN criteria, underwent analysis using the stand-alone version of the RESL algorithm (via the 'Cluster Sequences' function on BOLD), revealing 386 Operational Taxonomic Units (OTUs), potentially representing distinct species. Of these, only 18 OTUs (encompassing 24 records) were free from contamination or stop codons (Supp. File 3). Each of these 18 OTUs was cross-referenced with the BOLD ID Engine, revealing no matches to known BINs and suggesting they are novel to BOLD, as further supported by “taxon ID tree” analysis on BOLD. (Supp. File 4)

The 8,974 barcodes successfully assigned were distributed across 1,361 BINs. Notably, 191 unique BINs (14%) were exclusively identified at the Jinnah Garden site, while the remaining 1,170 BINs (86%) were shared with other locations, both within and outside Pakistan. In terms of taxonomic classification, 98% of the barcodes (9,255) belonged to the Class Insecta, followed by Class Arachnida (99 barcodes, 1%), Class Collembola (91 barcodes, 0.96%), and Class Malacostraca (5 barcodes, 0.05%). The Class Arachnida specimens were further categorized into four orders

(Araneae, Mesostigmata, Sarcoptiformes, and Trombidiformes), encompassing 17 families, 14 genera, and 9 species. Collembola included two orders (Entomobryomorpha and Symphypleona), with the former yielding 3 families, 3 genera, and 2 species. Malacostraca featured only the order Isopoda, with four barcodes across one species.

In the class Insecta, specimens were assigned across 10 orders, with 98% falling into 149 families (as detailed in Tables 1). The majority (92%) belonged to three orders: Diptera (66%), Hymenoptera (16%), and Hemiptera (10%), as shown in Figure 2. Other orders like Coleoptera, Lepidoptera, and Thysanoptera each had over 50 specimens, while Neuroptera, Odonata, Orthoptera, and Psocodea had just fewer.

Among 10792 specimens, 97% of specimens (N = 10,448) were accompanied by images. Most sequences (95%) received a BIN assignment, cumulating in 1,361 BINs. Over half (51%) of the 1,361 Barcode Index Numbers (BINs) were represented by a minimum of two or more sequences, while the remaining 49% were represented by only a single specimen. Notably, the proportion of these single-specimen BINs exceeded 40% in the orders of Coleoptera, Diptera, Hemiptera, and Hymenoptera, with the highest occurrence in Hymenoptera (58%, N=283). Additionally, a significant majority of specimens (97%, N=10,448) were documented with images. The assignment of BINs varied across different orders, with the order Araneae (Class Arachnida) showing an 82% success rate, and the order Entomobryomorpha (Class Collembola) achieving 85%. Within Class Insecta, the distribution of BIN assignments was as follows: Diptera and Hymenoptera both at 79%, Coleoptera at 75%, Lepidoptera at 64%, and Hemiptera at 52% (refer to Table 1). Together these five orders contributed to 94% of the BINs and 81% of the families identified (as shown in Table 1, Figures 2 and 3).

Regarding taxonomic resolution, 37% of BINs were identified to the genus level, and 15% to the species level. This led to the identification of 324 genera and 191 species (Table 1). In Class Insecta, a higher proportion of BINs were identified to the genus (32%, N=172) and species (12%, N=65) levels, particularly in the orders Diptera and Hymenoptera (Genus: 31%, N=153; Species: 7%, N=37). The representation of the 170 families was highly variable. Twenty-one families were represented by more than 100 specimens each, while 33 families had only a single specimen (Table 1). This variation was also observed in the number of BINs: 16 of the 21 families with more than 100 specimens had over 20 BINs, whereas 59 families had only one BIN. The families Ceratopogonidae (N=1,489) and Chloropidae (N=680) had the highest specimen counts in the order Diptera, Class Insecta. In terms of diversity, Cecidomyiidae (92 BINs) in order Diptera and Scelionidae (70 BINs) in order Hemiptera were the most diverse.

Figure 4 illustrates the diversity of BINs and the BIN-to-specimen ratio for the 21 families with more than 20 BINs. The ratio was highest for Scelionidae in Hemiptera (0.42) and lowest for Anthomyiidae and Ceratopogonidae in Diptera (0.03). Based on the Preston log-normal species distribution, species richness extrapolation indicates that a thorough examination of Jinnah Garden's Malaise-trappable arthropod fauna might yield roughly 2,785 BINs, or almost twice as many as have been found. However, the Chao-1 Index estimates the species richness at 2,389.74 BINs. The Simpson Index of Diversity (1-D) is 0.989, indicating very high species diversity, and the Shannon Index is 5.77, suggesting high species richness and evenness (Table 2, Fig 5). (Supp. File 5)

DISCUSSION

Pakistan's arthropod biodiversity has been estimated to contain anywhere from approximately 5,000 to 20,000 species, according to various sources (Ministry of Climate Change, Pakistan 2019, Hasnain 1998). Nevertheless, these previous estimates have been deemed insufficient by recent studies (Baig and Al-Subaiee 2009, Rana et al 2019). This research aimed to provide a more accurate assessment by utilizing DNA barcoding and the BIN system. A thorough examination was conducted on more than 10,792 specimens gathered from Jinnah Garden in Lahore, Pakistan. This exhaustive endeavor aimed to establish a comprehensive DNA barcode library of the region's arthropod fauna.

Although the success rate (88%) for recovering DNA barcodes was good, it differed significantly between orders from 77% for thysanoptera to 93% for Diptera (with specimens more than 100). In other research, similar variance in barcode recovery across several arthropod taxa has been documented.Park et al. 2011 reported sequence recovery rates ranging from 75% to 80% in Hemiptera. Ashfaq et al. (2012) reported similar trends of variation in sequence recovery rates in various insect orders and in a broader scope, across Canada, DNA barcoding was performed on 1,500,003 animal specimens across diverse taxonomic levels, resulting in 38% of specimens being assigned taxonomically at the species level, with higher proportions at the genus and family levels (deWaard et al. 2019a).

However, low sequence recovery rates in DNA barcoding can be attributed to various factors. One significant factor is the failures in primer binding, which can result from genetic variation within the primer binding sites or mismatches between the primers and the targeted DNA, leading to reduced sequence recovery (Elbrecht et al. 2018, Wilson et al. 2017). Such primer binding failures can especially affect the amplification of DNA from diverse or taxonomically complex groups, contributing to underestimations of species richness (Wilson et al. 2017). Furthermore, co-amplification of pseudogenes, which are non-functional DNA sequences that resemble the target genes, can introduce errors and reduce the accuracy of species identification (Leite 2012). The presence of endosymbionts, like Wolbachia, in the host species can interfere with DNA extraction and amplification, impacting the success of barcoding (Jones et al., 2011). Recent speciation events and incomplete lineage sorting can lead to genetic similarity among closely related species, making it challenging to distinguish them using a single barcode marker (Soria-Carrasco et al. 2014, Yasuda et al. 2015).

The combination of morphological examination and barcode matching on BOLD (deWaard et al. 2019a, 2019b) proved highly effective in assigning BINs to an order level with 100% efficacy and 98% to the family level. However, just 37% of BINs were identified to the genus level, and 15% to the species level. This resulted in the identification of 324 genera and 191 species, demonstrating improved parameterization of the barcode reference library. This was especially true for the two most diverse orders (Hymenoptera: 8% and Diptera: 12%), where species assignments were less than 15%. Notably, similar studies conducted in other regions have achieved considerably higher assignment success rates; Canada (38%) and Germany (34%) (e.g., Geiger et al. 2016, deWaard et al. 2019a), suggesting that further optimization of the DNA barcode reference library may be necessary to improve identification accuracy.

Although the reference database used in the current analysis was limited (Virgilio et al. 2010), the study still managed to identify representatives from a significant number of genera and species. Specifically, 324 genera and 191 species successfully identified through the use of the global reference library known as BOLD. These findings highlight the advantages of using DNA barcoding over traditional morphological identification methods (Marshall et al. 2009).

Furthermore, the analysis yielded 1361 Barcode Index Numbers (BINs), indicating a high level of species richness in the fauna of Jinnah Garden in Lahore, Pakistan. However, It is crucial to emphasize that these estimates are derived from a restricted sample collection and limited geographic coverage, so more comprehensive efforts and broader sampling would likely result in even higher estimates of species richness.

Of the 17 arthropod orders identified in the study, five orders (Coleoptera, Diptera, Hemiptera, Hymenoptera and Lepidoptera), from the class Insecta were the most abundant, collectively making up a substantial portion of 94% BINs. This finding corroborates earlier investigations employing both morphological (Stork 2018) and molecular techniques (Ashfaq et al. 2022; Pentinsaari et al. 2020). The dominance of Diptera and Hymenoptera can be attributed to the Malaise trap collection method, which preferentially captures low-flying insects, including these orders (Cooksey and Barton, 1981, deWaard et al. 2019b). Comparable patterns have been documented in other studies, such as those conducted in Canada where Diptera constituted approximately 57% of collections (deWaard et al. 2019b).

With 100 or more specimens, 16 of the 362 families dominated, and the BIN diversity mirrored this trend. The fact that 59 families were represented by a single BIN and 33 families by a single specimen lends credence to the survey's inconsistent detection of families. It's interesting to note that of the 21 families with the highest BINs, eight were dipterans, and the greatest BIN: specimen ratio was found in the family Cecidomyiidae (Fig 4).

The analysis of species richness extrapolation, utilizing the Preston log-normal species distribution model, suggest that a comprehensive sampling effort of the fauna at the Jinnah Garden as it could uncover a significantly higher number of Barcode Index Numbers (BINs) than what has been observed to date. A similar results were documented by (deWaard et al. 2019).

The Barcode of Life Data System (BOLD) houses an extensive database, and is a dependable platform for evaluating faunal overlap through BINs, with over nine million DNA barcode records for over 760,000 animal species. The results emphasize the importance of compiling local biodiversity inventories. This study represents a significant advancement in establishing an inventory of the arthropod fauna in Lahore, Pakistan.

REFERENCES:

Andújar C, García-de la Torre G, and Rodríguez JM. 2018. Using DNA barcodes to identify and quantify fish species in market products. Int. Food Res. J, 109:443-451. https://doi.org/10.1371/journal.pone.0202976

Ashfaq M, Akhtar S, Rafi MA, Mansoor S. and Hebert PD. 2017. Mapping global biodiversity connections with DNA barcodes: lepidoptera of Pakistan. PloS one. 12:e0174749. https://doi.org/10.1371/journal.pone.0174749

Ashfaq, M., Khan, A. M., Rasool, A., Akhtar, S., Nazir, N., Ahmed, N., and Hebert, P. D. 2022. A DNA barcode survey of insect biodiversity in Pakistan. *PeerJ*, **10**: e13267. https://doi.org/10.7717/peerj.13267

Baig MB, Al-Subaiee FS. 2009. Biodiversity in Pakistan: key issues. Biodiversity 10:20–29. https://doi.org/10.1080/14888386.2009.9712858.

Blagoev GA, Nikolova NI, Sobel CN, Hebert PD and Adamowicz SJ. 2013. Spiders (Araneae) of Churchill, Manitoba: DNA barcodes and morphology reveal high species diversity and new Canadian records. BMC Ecol. 13(1):1-7. https://doi.org/10.1186/1472-6785-13-44

Braukmann TWA, Ivanova NV, Prosser SWJ, et al. 2019. Metabarcoding a diverse arthropod mock community. Mol. Ecol. Resour. 19:711–727. https://doi.org/1010.1111/1755-0998.13008.

Cognato AI, Sari G, Smith SM, Beaver RA,et al. 2020. The essential role of taxonomic expertise in the creation of DNA databases for the identification and delimitation of Southeast Asian ambrosia beetle species (Curculionidae: Scolytinae: Xyleborini). Front. Ecol. Evol. 8:27 https://doi.org/10.3389/fevo.2020.00027

Cooksey LM, Barton HE. 1981. Flying insect populations as sampled by Malaise trap on Crowley’s Ridge in northeast Arkansas. JAAS. 35:29–32. https://scholarworks.uark.edu/jaas/vol35/iss1/9

Dawan J, Ahn J. 2022. Application of DNA barcoding for ensuring food safety and quality. Food Sci. Biotechnol. 11:1355-64. https://doi.org/10.1007/s10068-022-01143-7 deWaard JR, Levesque-Beaudin V, deWaard SL, Ivanova NV, McKeown JTA, Miskie R, Naik S, Perez KHJ, Ratnasingham S, Sobel CN, Sones JE, Steinke C, Telfer AC, Young AD, Young MR, Zakharov EV, Hebert PDN, Wilson JJ. 2019b. Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding. Genome 62(3):85–95. https://doi.org/10.1139/gen-2018-0093.

deWaard JR, Ratnasingham S, Zakharov EV, Borisenko AV, Steinke D, Telfer AC, Perez KHJ, Sones JE, Young MR, Levesque-Beaudin V, Sobel CN, Abrahamyan A, Bessonov K, Blagoev G, deWaard SL, Ho C, Ivanova NV, Layton KKS, Lu L, Manjunath R, McKeown JTA, Milton MA, Miskie R, Monkhouse N, Naik S, Nikolova N, Pentinsaari M, Prosser SWJ, Radulovici AE, Steinke C, Warne CP, Hebert PDN. 2019a. A reference library for the identification of Canadian invertebrates: 1.5 million DNA barcodes, voucher specimens, and genomic samples. Scientific Data 6(1):308. https://doi.org/10.1038/s41597-019-0320-2.

DeWalt RE. 2011. DNA barcoding: a taxonomic point of view. Journal of the North American Benthological Society. 30(1):174-81. https://doi.org/10.1899/10-021.1

D'Souza ML, Van der Bank M, Shongwe Z, Rattray RD, Stewart R, Van Rooyen J, Govender D, Hebert PD. 2021. Biodiversity baselines: tracking insects in Kruger National Park with DNA barcodes. Biol. Conserv. 256:109034. https://doi.org/10.1016/j.biocon.2021.109034

Elbrecht V, Hebert PD, Steinke D. Slippage of degenerate primers can cause variation in amplicon length. Sci Rep. 8(1):10999. https://doi.org/10.1038/s41598-018-29364-z

Geiger M, Moriniere J, Hausmann A, Haszprunar G, Wägele W, Hebert PDN, Rulik B. 2016. Testing the global malaise trap program–how well does the current barcode reference library identify flying insects in Germany? Biodiversity Data Journal 4:e10671 https://doi.org/10.3897/BDJ.4.e10671.

Government of Pakistan. 2000. Biodiversity action plan for Pakistan. Available at https://portals. iucn.org/library/efiles/documents/2000-081.pdf.

Hajibabaei M, Singer GA, Hebert PD, Hickey DA. 2007. DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. TiG. 23(4):167-

172. https://doi.org/10.1016/j.tig.2007.02.001

Hajibabaei, M., Singer, G. A., Hebert, P. D., and Hickey, D. A. (2007). DNA barcoding: how it complements taxonomy, molecular phylogenetics and population genetics. *TRENDS in Genetics*, *23*(4), 167-172.

Hasnain T. 1998. Implementation of convention on biological diversity in Pakistan: policy brief series # 2. Islamabad, Pakistan: Sustainable Development Policy Institute (SDPI).

Hausmann A, Godfray HCJ, Huemer P, Mutanen M, Rougerie R, van Nieukerken EJ, Ratnasingham S, Hebert PDN. 2013. Genetic patterns in European geometrid moths revealed by the barcode index number (BIN) system. PloS one. 8:e84518 https://doi.org/10.1371/journal.pone.0084518.

Hebert PDN, Braukmann TWA, Prosser SWJ, Ratnasingham S, deWaard JR, Ivanova NV, Janzen DH,Hallwachs W,NaikS,Sones JE, Zakharov EV. 2018. ASequel to Sanger: amplicon sequencing that scales. BMC Genom, 19:219. https://doi.org/10.1186/s12864-018-4611-3.

Hebert PDN, Cywinska A, Ball SL, deWaard JR. 2003. Biological identifications through DNA barcodes. Proc. R. Soc. B: Biol. Sci. 270:313–321. https://doi.org/10.1098/rspb.2002.2218.

Hlebec D, Sivec I, Podnar M, Kučinić M. 2022. DNA barcoding for biodiversity assessment: Croatian stoneflies (Insecta: Plecoptera). PeerJ. 2022 Apr 20;10:e13213. https://doi.org/10.7717/peerj.13213

Huang Z, Yang C, Ke D. 2016. DNA barcoding and phylogenetic relationships in Anatidae. Mitochondrial DNA Part A. 27(2):1042-4. https://doi.org/10.3109/19401736.2014.926545

Ivanova NV, deWaard JR, Hebert PDN. 2006. An inexpensive, automation-friendly protocol for recovering high quality DNA. Mol. Ecol. Notes. 6(4):998–1002. https://doi.org/10.1111/j.1471-8286.2006.01428.x.

Jinbo U, Kato T, Ito M. 2011. Current progress in DNA barcoding and future implications for entomology. Entomol. Sci. 14(2):107-24. https://doi.org/10.1111/j.1479-8298.2011.00449.x

Jinbo, U., Kato, T. and Ito, M. (2011). Current progress in DNA barcoding and future implications for entomology. *Entomological Science*, *14*(2), 107-124.

1Jones M, Ghoorah A, Blaxter M. 2011. MOTU and taxonerator: turning DNA barcode sequences into annotated operational taxonomic units. PLoS one. 6(4):e19259. https://doi.org/10.1371/journal.pone.0019259

Leite LA. 2012. Mitochondrial pseudogenes in insect DNA barcoding: differing points of view on the same issue. Biota Neotrop. 12:301-8. https://doi.org/10.1590/S1676-06032012000300029

Marshall, S., Paiero, S. and Buck, M. 2009. Point Pelee National Park species list. Available at http:// www.uoguelph.ca/debu/pelee_specieslist.htm (accessed 20 June 2020).

Ministry of Climate Change, Pakistan. 2019. Pakistan's sixth national report to the United Nations Convention on Biological Diversity. Convention on Biological Diversity, Montreal, Canada. Available at https://www.cbd.int/doc/nr/nr-06/pk-nr-06-en.pdf.

Moczek AP. 2010. Phenotypic plasticity and diversity in insects. Philos. Trans. R. Soc. Lond., B, Biol. Sci. 365(1540):593-603. https://doi.org/10.1098/rstb.2009.0263

Montealegre-Z F, Jonsson T, Robson-Brown KA, Postles M, Robert D. 2012 Convergent evolution between insect and mammalian audition. Sci. 338(6109):968-71. https://doi.org/10.1126/science.1225271

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

Park DS, Suh SJ, Hebert PD, Oh HW, Hong KJ. 2011. DNA barcodes for two scale insect families, mealybugs (Hemiptera: Pseudococcidae) and armored scales (Hemiptera: Diaspididae). Bull. Entom. Res., Lond. 101(4):429-34. https://doi.org/10.1017/S0007485310000714

Pentinsaari M, Blagoev GA, HoggID, Levesque-Beaudin V, Perez K, Sobel CN, Vandenbrink B, Borisenko A. 2020. DNA barcoding survey of an Arctic arthropod community: implications for future monitoring. Insects. 11(1):46. https://doi.org/10.3390/insects11010046

Porco D, Rougerie R, Deharveng L, Hebert P. 2010. 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. 10(6):942–945. https://doi.org/10.1111/j.1755-0998.2010.2839.x.

Rana, N, Saleem M, Majeed W, Jalal F, Ehsan N, Nargis S. 2019. Diversity of arthropods regarding habitat specialty in agro-ecosystem of Faisalabad, Pakistan. GSC biol. pharm. sci. 6(2):01-08. https://doi.org/10.30574/gscbps.2019.6.2.0008

Ratnasingham S, Hebert PD. 2013 A DNA-based registry for all animal species: the Barcode Index Number (BIN) system. PloS one.8(7):e66213. https://doi.org/10.1371/journal.pone.0066213

Ratnasingham S, Hebert PD. 2017. BOLD: The Barcode of Life Data System (http://www. barcodinglife. org). Mol. Ecol. Notes. 7(3):355-64. https://doi.org/10.1111/j.1471-8286.2007.01678.x

Ren J, Ashfaq M, Hu X, Ma J, Liang F, Hebert PDN, Lin L, Germain JF, Ahmed MZ. 2018. Barcode index numbers expedite quarantine inspections and aid the interception of nonindigenous mealybugs (Pseudococcidae). Biological Invasions 20:449–460. https://doi.org/10.1007/s10530-017-1546-6.

Seberg O, Humphries CJ, Knapp S, Stevenson DW, Petersen G, Scharff N, Andersen NM. 2003. Shortcuts in systematics? A commentary on DNA-based taxonomy. Trends Ecol. Evol. 18(2):63-65. https://doi.org/10.1016/S0169-5347(02)00059-9

Seberg, O., Humphries, C. J., Knapp, S., Stevenson, D. W., Petersen, G., Scharff, N., and Andersen, N. M. (2003). Shortcuts in systematics? A commentary on DNA-based taxonomy. *Trends in Ecology & Evolution*, *18*(2), 63-65.

Shashank PR, Naveena NL, Rajgopal NN, Elliott TA, Sreedevi K, Sunil S, Meshram NM. DNA barcoding of insects from India: Current status and future perspectives. Mol. Biol. Rep. 49(11):10617-10626. https://doi.org/10.1007/s11033-022-07628-2

Shashank, P. R., Naveena, N. L., Rajgopal, N. N., Elliott, T. A., Sreedevi, K., Sunil, S., & Meshram, N. M. (2022). DNA barcoding of insects from India: Current status and future perspectives. *Molecular Biology Reports*, *49*(11), 10617-10626.

Silva-Brandão KL, Lyra ML, Freitas AV. 2009. Barcoding Lepidoptera: current situation and perspectives on the usefulness of a contentious technique. Neotrop. entomol. 38:441-51. https://doi.org/10.1590/S1519-566X2009000400001

Smith MA, Bertrand C, Crosby K, Eveleigh ES, Fernandez-Triana J, Fisher BL, Gibbs J, Hajibabaei M, Hallwachs W, Hind K, Hrcek J, Huang D-W, Janda M, Janzen DH, Li Y, Miller SE, Packer L, Quicke D, Ratnasingham S, Rodriguez J, Rougerie R, Shaw MR, Sheffield C, Stahlhut JK, Steinke D, Whitfield J, Wood M, Zhou X. 2012. Wolbachia and DNA barcoding insects: patterns, potential, and problems. Plos One 7:e36514 https://doi.org/10.1371/journal.pone.0036514.

Soria-Carrasco V, Gompert Z, Comeault AA,et al. 2014. Stick insect genomes reveal natural selection's role in parallel speciation. Science. 344(6185):738-42. https://doi.org/10.1126/science.1252136

Stork NE. 2018. How many species of insects and other terrestrial arthropods are there on Earth?. Annu. Rev. Entomol. 63:31-45. https://doi.org/10.1146/annurev-ento-020117-043348

Virgilio M, Backeljau T, Nevado B, De Meyer M. 2010. Comparative performances of DNA barcoding across insect orders. BMC bioinform. 11(1):1-10. https://doi.org/10.1186/1471-2105-11-206

Wilkinson MJ, Szabo C, Ford CS, Yarom Y, Croxford AE, Camp A, Gooding P. 2017. Replacing Sanger with Next Generation Sequencing to improve coverage and quality of reference DNA barcodes for plants. Sci Rep. 7(1):46040. https://doi.org/10.1038/srep46040

Williams HC, Ormerod SJ, Bruford MW. 2006. Molecular systematics and phylogeography of the cryptic species complex Baetis rhodani (Ephemeroptera, Baetidae). Mol. Phylogenet. Evol. 40(2):370-382. https://doi.org/10.1016/j.ympev.2006.03.004

Wilson JJ, Sing KW, Floyd RM, Hebert PD. 2017. DNA barcodes and insect biodiversity. J. Insect Biodivers. Syst. 575-592. https://doi.org/10.1002/9781118945568.ch17

Yasuda N, Taquet C, Nagai S. Fortes M, Fan TY, Harii S, Yoshida T, Sito Y, Nadaoka K.

2015. Genetic diversity, paraphyly and incomplete lineage sorting of mtDNA, ITS2 and microsatellite flanking region in closely related Heliopora species (Octocorallia). Mol. Phylogenet. Evol. 93: 161-171. https://doi.org/10.1016/j.ympev.2015.07.009

Zhou X, Jacobus LM, DeWalt RE, Adamowicz SJ, Hebert PD. 2010. Ephemeroptera, Plecoptera, and Trichoptera fauna of Churchill (Manitoba, Canada): insights into biodiversity patterns from DNA barcoding. Journal of the North American Benthological Society. 29(3):814-837. https://doi.org/10.1899/09-121.1

Table 1(on next page)

Table 1: Count of specimens having DNA barcode records from Jinnah Garden in Lahore, Pakistan, that belong to 17 orders. For every order, the number of families, genera, species, and BINs is given.

Order	Specimens with barcodes	Specimens assigned to BINs (%)	BINs recovered	OTUs without BIN	Singleton BINs (%)	BINs assigned to family (%)	Families recovered	BINs assigned to genus (%)	Genera recovered	BINs assigned to species (%)	Species recovered
Class Arachnida											
Araneae	36	59.89	15	1	64.29	85.71	11	71.43	10	28.57	4
Mesostigmata	24	100.00	11	0	36.36	81.82	4	54.55	3	27.27	3
sarcoptiformes	3	100.00	1	0	0.00	0.00	0	0.00	0	0.00	0
Trombidiformes	15	100.00	5	0	60.00	80.00	2	40.00	1	40.00	2
Class Collembola											
Entomobryomorpha	89	97.75	12	0	33.33	83.33	3	41.67	3	16.67	2
Symphypleona	2	50.00	1	0	100.00	0.00	0	0.00	0	0.00	0
Class Malacostraca											
Isopoda	5	80.00	1	0		100.00	1	100.00	1	100.00	1
Class Insecta											
Coleoptera	277	88.45	86	2	54.65	100.00	24	30.23	21	18.60	16
Diptera	6142	97.75	530	6	40.75	99.81	38	32.33	103	12.22	64
Hemiptera	996	89.06	122	0	42.62	99.18	23	59.02	51	31.15	33
Hymenoptera	1547	91.79	490	8	57.76	97.55	39	31.22	89	7.55	37
Lepidoptera	155	83.23	48	0	52.08	87.50	14	58.33	25	31.25	15
Neuroptera	1	100.00	1	0	100.00	100.00	1	100.00	1	0.00	0
Odonata	5	100.00	2	0	50.00	100.00	1	50.00	1	0.00	1
Orthoptera	6	83.33	3	1	66.67	100.00	3	66.67	2	66.67	2
Psocodea	34	88.24	9	0	44.44	77.78	4	33.33	2	22.22	2
Thysanoptera	92	89.01	24	0	41.67	100.00	2	70.83	11	41.67	10
Total	**9450**	**95%**	**1361**	**18**	**49%**	**98%**	**170**	**37%**	**324**	**15%**	**191**

Table 1: Count of specimens having DNA barcode records from Jinnah Garden in Lahore, Pakistan, that belong to 17 orders. For every order, the number of families, genera, species, and BINs is given. (Supp. Table S1).

Table 2(on next page)

Table 2: Species richness estimates based on the abundance of 1361 BINs encountered at various sites in Jinnah garden, Lahore, Pakistan

Specimens	BINs	Preston	Chao-1 index	Simpson Index of Diversity (1-D)	Shannon Index
8,974	1361	2785	2389	0.98	5.77

Table 2: Species richness estimates based on the abundance of 1361 BINs encountered at various sites in Jinnah garden, Lahore, Pakistan. (Supp File 5)

Figure 1

Jinnah garden map showing collection sites for arthropod specimens examines in this study. The color of each site point indicate the number of specimens sampled. Map was generated in GIS-map satellite imagery.

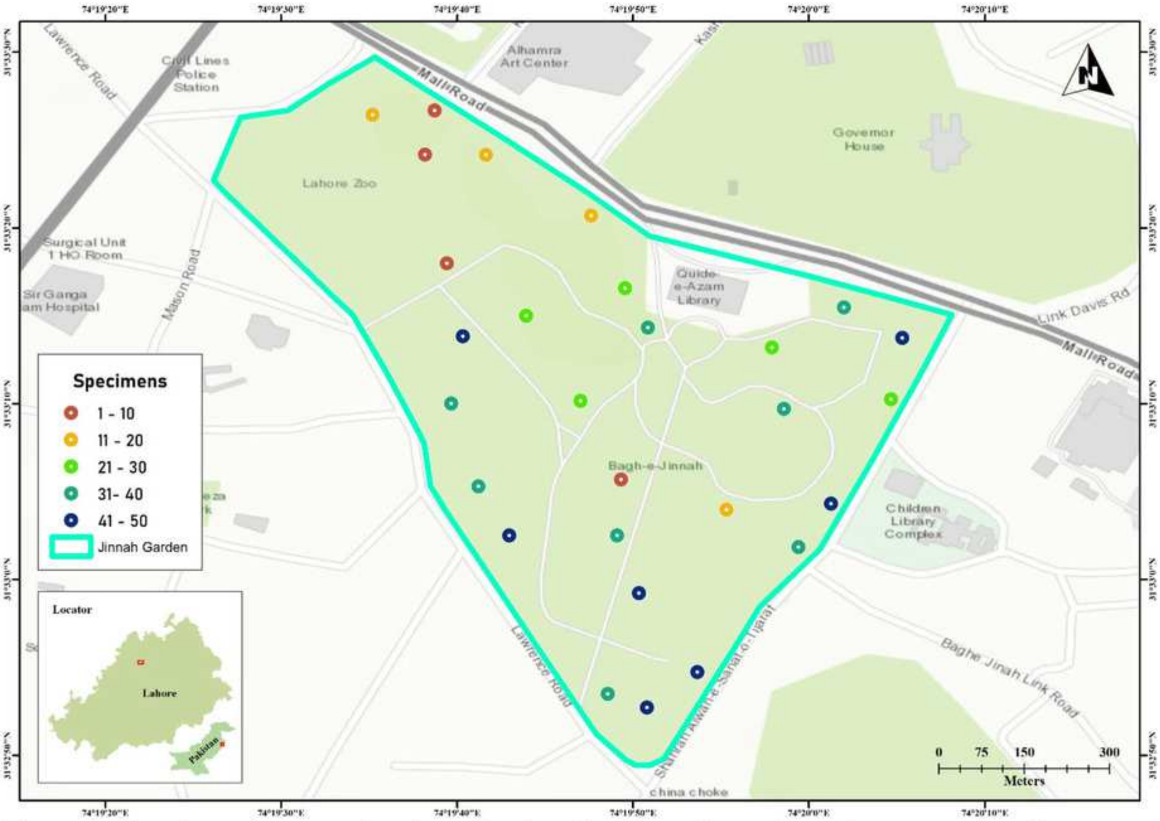

Figure 1: Jinnah garden map showing collection sites for arthropod specimens examines in this study. The color of each site point indicate the number of specimens sampled. Map was generated in GIS-map satellite imagery.

Figure 2(on next page)

Taxonomic assignment for 17 orders of four classes of Phylum Arthropoda

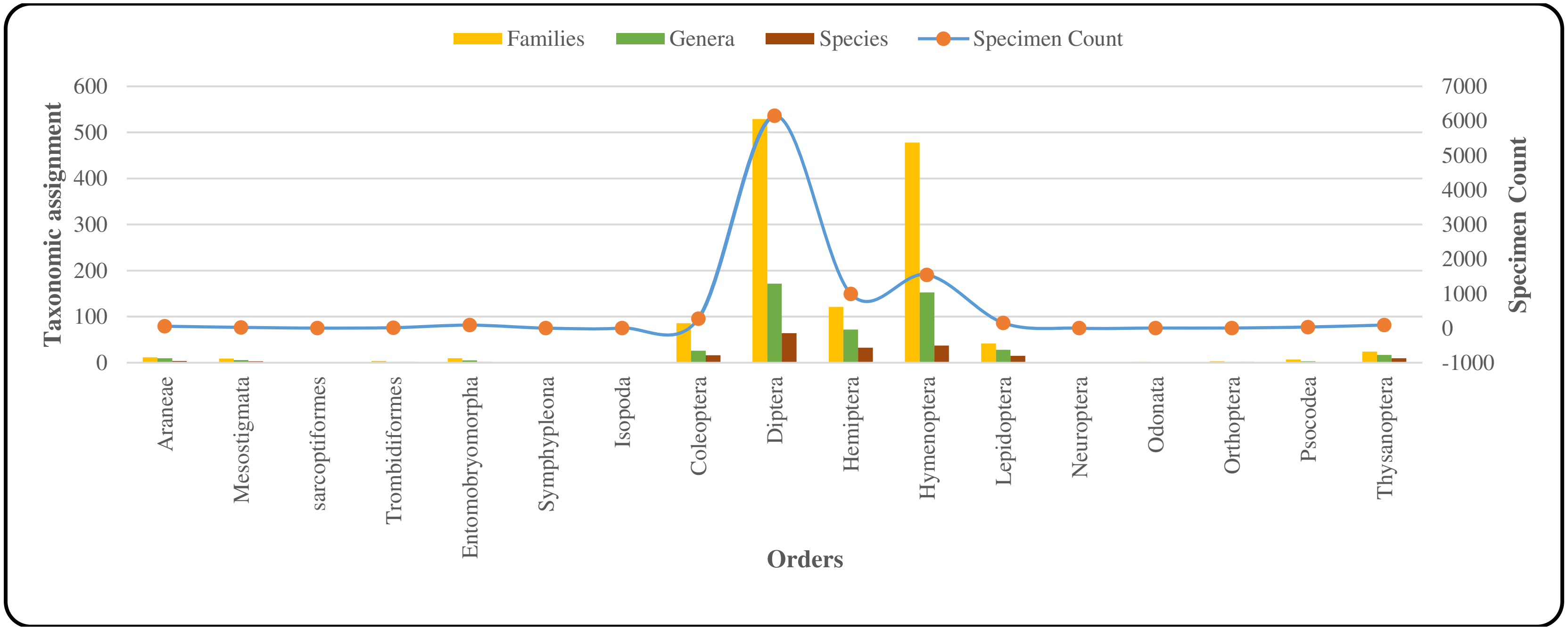

Figure 2: Taxonomic assignment for 17 orders of four classes of Phylum Arthropoda.

Figure 3

BINs assignment for 17 orders of four classes of Phylum Arthropoda.

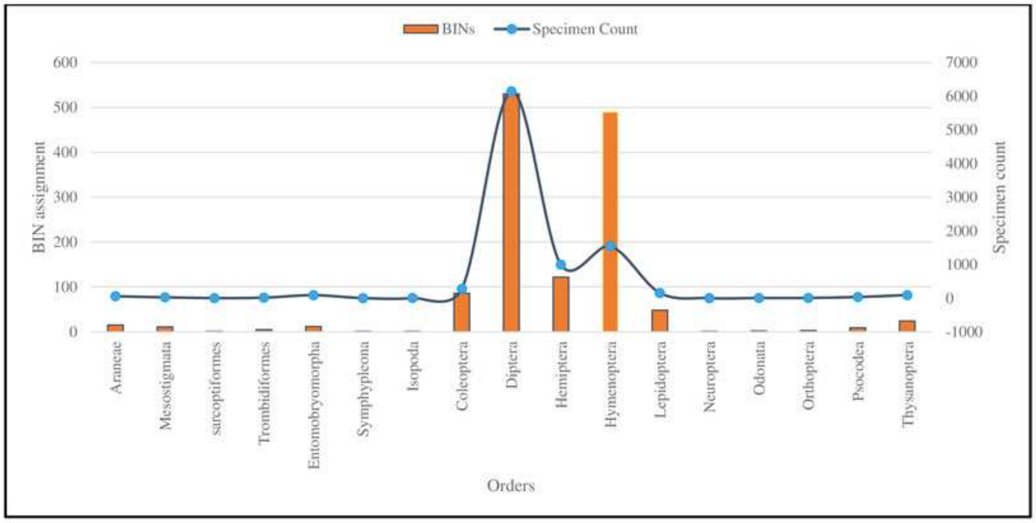

Figure 3: BINs assignment for 17 orders of four classes of Phylum Arthropoda.

Figure 4

BIN: Specimen ratio and BIN diversity for 21 insect families (of three orders) with more than 20 BINs each

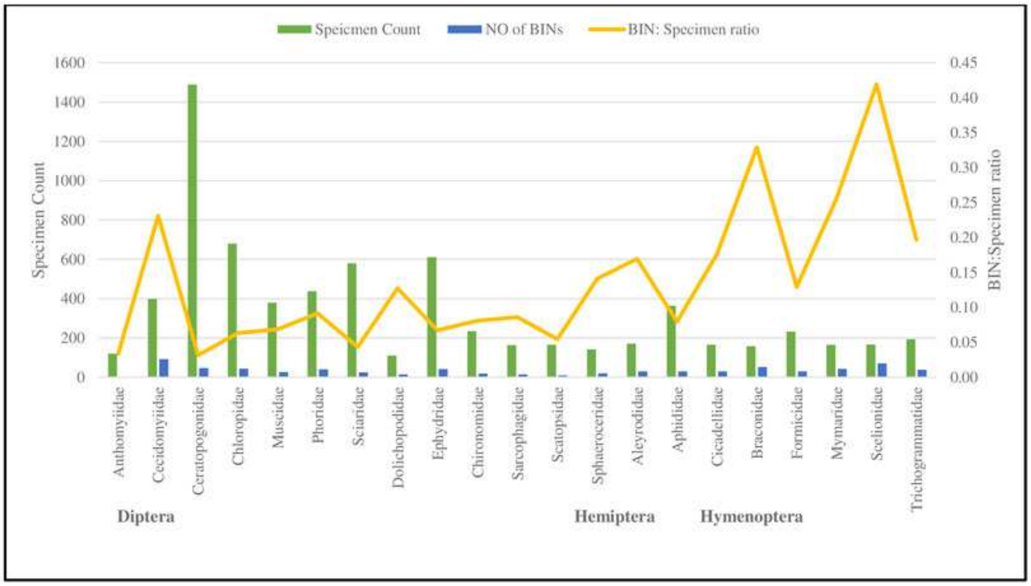

Figure 4: BIN: Specimen ratio and BIN diversity for 21 insect families (of three orders) with more than 20 BINs each.

Figure 5

Preston plot with veil line (and species richness extrapolation) based upon the abundance data of 8451 arthropods taxa which generated a sequence.

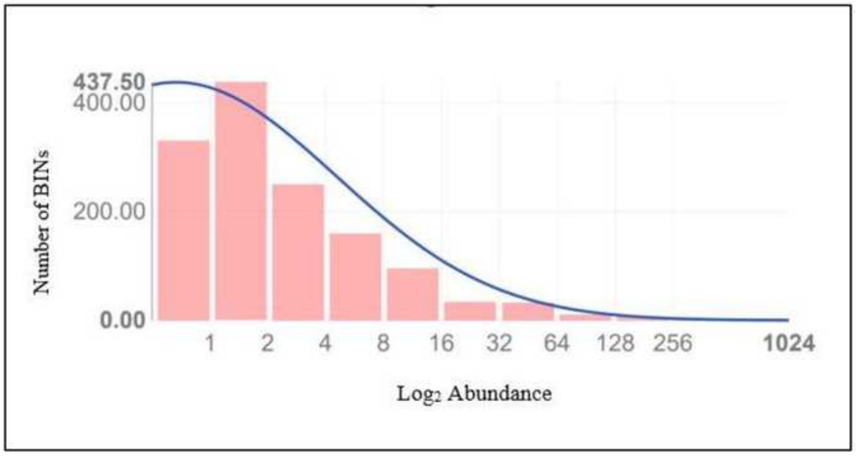

Figure 5: Preston plot with veil line (and species richness extrapolation) based upon the abundance data of 8451 arthropods taxa which generated a sequence.