

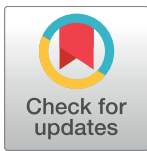
RESEARCH ARTICLE

Insect diversity in the Saharo-Arabian region: Revealing a little-studied fauna by DNA barcoding

Muhammad Ashfaq^{1*}, Jamal S. M. Sabir², Hosam O. El-Ansary^{3,4}, Kate Perez¹, Valerie Levesque-Beaudin¹, Arif M. Khan⁵, Akhtar Rasool⁶, Carlene Gallant¹, Joseph Addesi¹, Paul D. N. Hebert¹

1 Centre for Biodiversity Genomics, University of Guelph, Guelph, ON, Canada, **2** Biotechnology Research Group, Department of Biological Sciences, King Abdulaziz University, Jeddah, Saudi Arabia, **3** Floriculture, Ornamental Horticulture and Garden Design Department, Faculty of Agriculture (El-Shatby), Alexandria University, Alexandria, Egypt, **4** Department of Plant Production, College of Food and Agriculture Sciences, King Saud University, Riyadh, Saudi Arabia, **5** Department of Biotechnology, University of Sargodha, Sargodha, Pakistan, **6** Centre for Animal Science and Fisheries, University of Swat, Mingora, Pakistan

* mashfaq@uoguelph.ca



OPEN ACCESS

Citation: Ashfaq M, Sabir JSM, El-Ansary HO, Perez K, Levesque-Beaudin V, Khan AM, et al. (2018) Insect diversity in the Saharo-Arabian region: Revealing a little-studied fauna by DNA barcoding. PLoS ONE 13(7): e0199965. <https://doi.org/10.1371/journal.pone.0199965>

Editor: Bernd Schierwater, Tierärztliche Hochschule Hannover, GERMANY

Received: March 13, 2018

Accepted: June 18, 2018

Published: July 9, 2018

Copyright: © 2018 Ashfaq 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: Specimen data and sequences generated in this study are accessible on the Barcode of Life Data Systems (www.boldsystems.org) under the dataset DS-MAREG. (<https://doi.org/10.5883/DS-MAREG>).

Funding: This work was funded by the Deanship of Scientific Research (DSR) at King Abdulaziz University (grant number 99-130-35-HiCi); Ontario Ministry of Research, Innovation and Science; and 'Food From Thought' program supported by the Canada First Research Excellence Fund. The

Abstract

Although insects dominate the terrestrial fauna, sampling constraints and the poor taxonomic knowledge of many groups have limited assessments of their diversity. Passive sampling techniques and DNA-based species assignments now make it possible to overcome these barriers. For example, Malaise traps collect specimens with minimal intervention while the Barcode Index Number (BIN) system automates taxonomic assignments. The present study employs Malaise traps and DNA barcoding to extend understanding of insect diversity in one of the least known zoogeographic regions, the Saharo-Arabian. Insects were collected at four sites in three countries (Egypt, Pakistan, Saudi Arabia) by deploying Malaise traps. The collected specimens were analyzed by sequencing 658 bp of cytochrome oxidase I (DNA barcode) and assigning BINs on the Barcode of Life Data Systems. The year-long deployment of a Malaise trap in Pakistan and briefer placements at two Egyptian sites and at one in Saudi Arabia collected 53,092 specimens. They belonged to 17 insect orders with Diptera and Hymenoptera dominating the catch. Barcode sequences were recovered from 44,432 (84%) of the specimens, revealing the occurrence of 3,682 BINs belonging to 254 families. Many of these taxa were uncommon as 25% of the families and 50% of the BINs from Pakistan were only present in one sample. Family and BIN counts varied significantly through the year, but diversity indices did not. Although more than 10,000 specimens were analyzed from each nation, just 2% of BINs were shared by Pakistan and Saudi Arabia, 4% by Egypt and Pakistan, and 7% by Egypt and Saudi Arabia. The present study demonstrates how the BIN system can circumvent the barriers imposed by limited access to taxonomic specialists and by the fact that many insect species in the Saharo-Arabian region are undescribed.

funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

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

Introduction

Because insects are the major component of terrestrial metazoan biodiversity and important indicators of environmental conditions [1,2], their biomonitoring can aid efforts to conserve and restore biodiversity [3], to evaluate the impacts of climate change [4], and to protect ecological services [5,6]. However, comprehensive assessments of their diversity have been impossible because of the lack of taxonomic specialists, and because many insect species are undescribed [7]. These barriers have contributed to the current uncertainty in the global species count [8] which can only be resolved by new approaches as morphological studies could require a millennium to inventory all species [9].

Recent advances in sequencing technology have stimulated the adoption of DNA-based methods for documenting biodiversity [10,11]. The capacity of DNA barcoding to advance understanding of biodiversity at local [12–14] and continental scales [15] is now well-established. The Barcode Index Number (BIN) System [16] is a particularly important development as BINs are a strong proxy for morphological species [17,18]. Consequently, the BIN system is enabling both large-scale assessments of species diversity [19] and detailed studies of entire taxonomic assemblages [20]. The capacity of BINs to delimit species is particularly valuable in settings where prior taxonomic work has been constrained [21]. As a consequence, BINs are being used to advance biodiversity inventories [22], to reveal unknown faunas [23], and to explore biodiversity links among nations [24,25].

Among the 12 zoogeographic regions [26], data on insect diversity is least available for the Saharo-Arabian (Fig 1). For example, nearly half of the 371 arthropod families recently collected from the United Arab Emirates represented new records for that nation [27]. Egypt, Pakistan, and Saudi Arabia jointly comprise a land area of four million square-kilometers in the Saharo-Arabian region. Although each nation is thought to host more than 10,000 insect species, documentation of this diversity is very incomplete [28,29]. For instance, only 5,000 insect species have been reported from Pakistan [30], 3,000 from Saudi Arabia [31], and less than 2,500 from Egypt [32] in comparison with over 100,000 from Canada and the USA. The limited availability of biodiversity data for the Saharo-Arabian region restricts the recognition of invasive taxa and threatened species while also constraining deeper assessments of patterns of community structure and faunal evolution.

Aside from a lack of taxonomic resources, inaccessible terrain and political instability have contributed to the poor documentation of Saharo-Arabian biodiversity. This knowledge gap can be addressed by coupling efficient sampling methods, such as Malaise trap [33] with subsequent DNA barcode analysis to speed determinations of alpha and beta diversity and overlap [19]. By using this approach to examine patterns of insect diversity in Egypt, Pakistan, and Saudi Arabia, the present study provides a model for biodiversity surveillance in regions with limited previous biodiversity research.

Materials and methods

Ethics statement

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

Insect collection

No single collection method allows a comprehensive assessment of insect diversity [34,35]. As a result, diverse sampling methods are employed. These methods include, but are not limited to, light traps, Malaise traps, pitfall traps, pyrethrum knockdown, and Tullgren extractors.

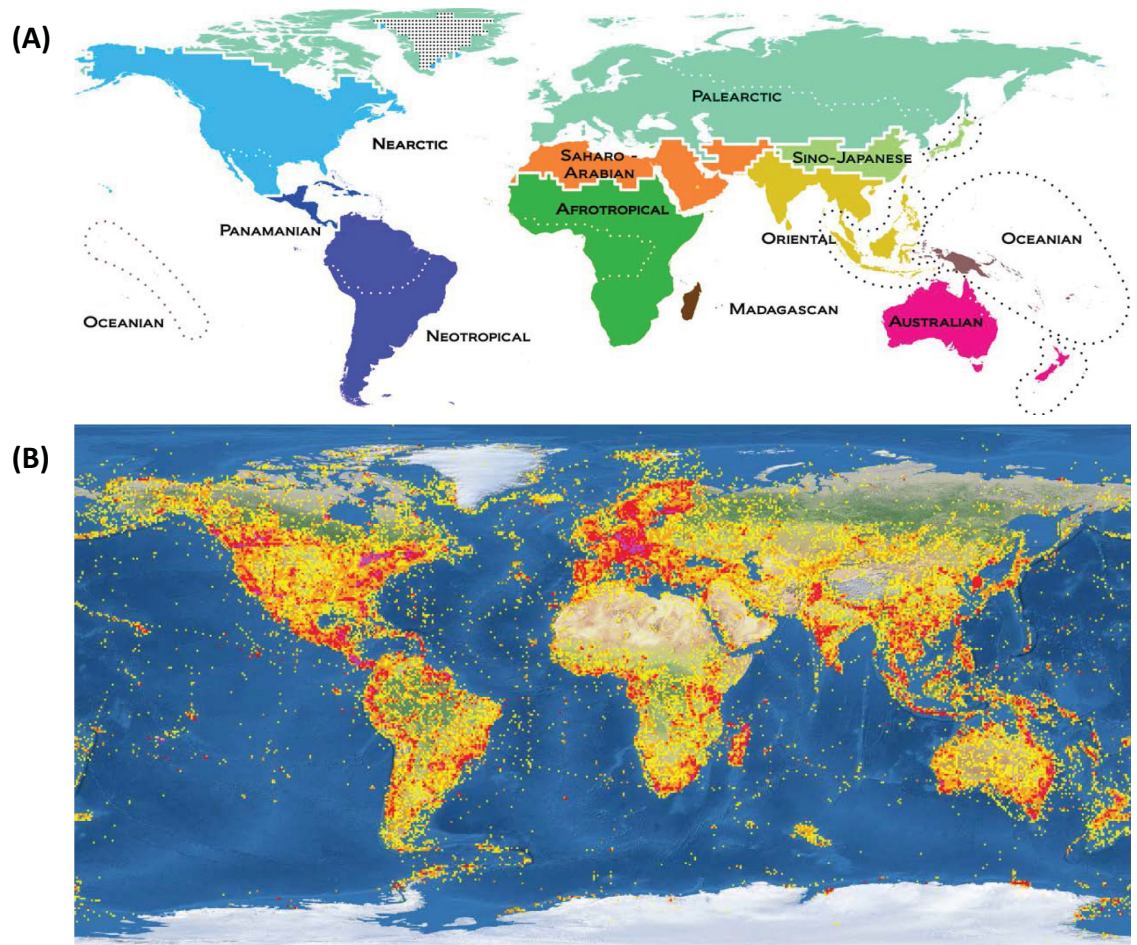


Fig 1. Map of the terrestrial zoogeographic regions (A; adopted from Holt et al. Science 2013; 339: 74–78) and biodiversity documentation by DNA barcoding (B; taken from the Barcode of Life Data Systems, www.boldsystems.org).

<https://doi.org/10.1371/journal.pone.0199965.g001>

Although Malaise traps are not an all-purpose collection method [36], they are generally accepted as the most cost- and time-effective [37], leading to their frequent use in biodiversity assessments [38]. Reflecting this fact, the Global Malaise Trap Program (<http://globalmalaise.org/>) has adopted this sampling method, and coupled it with DNA barcoding to advance understanding of insect diversity. As one element of this program, the present study deployed Malaise traps at four locations in three countries. One trap was installed at the Pakistan Museum of Natural History, Islamabad (33.686° N, 73.076° E) in the Shakarparian forest. Insects were collected from 7 February–13 December 2012 excepting 10 days in August/September and one week in July when samples were lost due to storm damage producing 39 collection events (weeks). A second trap was installed at Hada AL-Sham, a green valley, near Makkah, Saudi Arabia (21.795° N, 39.711° E) from 9 April–3 July 2014 (13 weeks) and from November 2014–February 2015 (5 weeks). The other traps were installed at Mostafa Kamel Village (30.92° N, 29.76° E) and Antoniodes Gardens (31.204° N, 29.95° E), Egypt. Both the sampling sites are situated on the delta of river Nile. Weekly samples were collected at site 1 from 27 May–7 October 2013 (20 weeks) and at site 2 from 29 May–18 September 2013 (17 weeks). Insects were collected into 95% ethanol, and samples were stored at -20°C until shipment to the Centre for Biodiversity Genomics at Guelph for DNA barcode analysis.

Molecular analysis

Specimens were sorted to order, arrayed, labeled, databased, and tissue sampled following standard workflows. Small specimens were transferred directly to 96-well microplates, and vouchers were recovered after DNA extraction for imaging and curation [39]. DNA extraction, PCR amplification, and sequencing were performed at the Canadian Centre for DNA Barcoding following standard protocols (<http://ccdb.ca/resources>). Except for hemipterans, PCR amplification of COI-5' was performed with primers C_LepFolF and C_LepFolR (http://ccdb.ca/site/wp-content/uploads/2016/09/CCDB_PrimerSets.pdf) following PCR conditions; 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). These primers are mixtures of LepF1 [40]/LCO1490 [41] and LepR1 [40]/HCO2198 [41], respectively. PCR amplification of hemipterans was conducted with LepF2_t1 [42] and LepR1 using the same thermocycling regime as above. All amplification reactions included 10.5 μ L (or 5.25 μ L) of standard PCR ingredients [40] and 2 μ L (or 1 μ L) of DNA template. PCR products were analyzed using the E-gel 96 system (Invitrogen Inc.) and amplicons were sequenced using BigDye v3.1 (Applied Biosystems) on an ABI 3730XL. Sequences were assembled, aligned, and edited using CodonCode Aligner (CodonCode Corporation, USA) and submitted to Barcode of Life Data Systems (BOLD) (www.boldsystems.org) [43]. All sequences generated in this study are accessible on BOLD under the dataset DS-MAREG (dx.doi.org/10.5883/DS-MAREG).

Data analysis

Sequences meeting quality criteria (>507 bp, <1% Ns, no stop codon or contamination flag) were assigned to a BIN by the Refined Single Linkage (RESL) algorithm on BOLD [16] which runs monthly on all eligible sequences. Shorter sequences (<507 bp), meeting all the other criteria, were assigned to the matching pre-existing BINs containing longer sequences. Results of this analysis are accessible through individual BIN pages. With few exceptions, each specimen was assigned to a family by sequence matches or by morphological analysis to existing records on BOLD (http://www.boldsystems.org/index.php/IDS_OpenIdEngine). The family level assignment of specimens was based on 90% or higher match of the unknown sequence with the known sequence. Sequences and their associated taxonomic data were subsequently downloaded from BOLD for analysis. Diversity indices, BIN/family incidences over time, and weather-diversity relationships were only examined for Pakistan since this was the sole location with collections for >75% of the year. Monthly mean temperature and relative humidity (RH) values for this site were obtained from the Pakistan Meteorological Department, Islamabad. Diversity indices, Simpson's D [44] and Shannon's H [45] were calculated for each collection event. The significance of diversity variation among the collection events was determined with a χ^2 test. BIN overlap among sites was calculated in Excel while BIN accumulation curves were generated on BOLD.

Results

The four Malaise traps collected 53,092 insects including 22,624 from Pakistan, 18,391 from Egypt (6,854 at site 1, 11,537 at site 2), and 12,077 from Saudi Arabia. Although these specimens were assigned to 17 orders, most were Diptera (42%), Hymenoptera (29%), Hemiptera (13%), Lepidoptera (7%) or Coleoptera (4%). A χ^2 test showed that the proportion of each order varied among the collection sites ($P < 0.05$) (Table 1). Although 84% (44,432) of the specimens generated barcodes, sequence recovery varied among sites with the lowest recovery from Saudi Arabia (79%). Because temperatures in Makkah often exceeded 40°C during the collection period, preservation was likely compromised [46]. Variation in barcode recovery

Table 1. Insect diversity analysis by Malaise trap collections at four sites in three countries.

	Egypt			Pakistan	Saudi Arabia	Total
	Site 1	Site 2	Egypt total			
Collection period	27 May– 7 Oct, 2013	29 May– 18 Sep, 2013		15 Feb– 6 Dec 2012	9 Apr– 3 July, 2014 + one week each in Nov, 2014 –Feb, 2015	
Total catch (specimens)	6854	11537	18391	22624	12077	53092
DNA barcodes recovered (%)	5611 (82)	10169 (88)	15780 (86)	19068 (84)	9584 (79)	44432 (84)
BINs	571	636	991	2248	728	3682
Singleton BINs	256	239	367	1052	323	1566
Orders	14	12	14	17	12	17
Families	129	137	164	214	132	254
Most common orders: n (% of total catch):						
i) Coleoptera	324 (5)	153 (1)	477 (3)	1083 (5)	702 (6)	2262 (4)
ii) Diptera	4717 (69)	3280 (28)	7997 (43)	11799 (52)	2559 (21)	22355 (42)
iii) Hemiptera	197 (3)	1145 (10)	1342 (7)	1592 (7)	4097 (34)	7031 (13)
iv) Hymenoptera	981 (14)	5008 (43)	5989 (33)	6684 (30)	2847 (24)	15520 (29)
v) Lepidoptera	389 (6)	1583 (14)	1972 (11)	1028 (5)	622 (5)	3622 (7)
Chi-square = 124.8; <i>P</i> <0.05						
DNA barcodes recovered: n (%)						
i) Coleoptera	254 (78)	137 (90)	391 (82)	877 (81)	326 (46)	1594 (70)
ii) Diptera	3898 (83)	3161 (96)	7059 (89)	11126 (94)	2318 (91)	20503 (92)
iii) Hemiptera	125 (63)	833 (73)	958 (71)	1026 (64)	3298 (80)	5282 (75)
iv) Hymenoptera	780 (80)	4138 (83)	4918 (82)	4728 (71)	2139 (75)	11785 (76)
v) Lepidoptera	366 (94)	1568 (99)	1934 (98)	970 (94)	609 (98)	3513 (97)
Chi-square = 54.98; <i>P</i> <0.05						
Order diversity: BINs (BIN: specimen ratio)						
i) Coleoptera	58 (0.23)	37 (0.27)	81 (0.21)	207 (0.24)	76 (0.23)	348 (0.22)
ii) Diptera	230 (0.06)	261 (0.08)	381 (0.05)	818 (0.07)	214 (0.09)	1285 (0.06)
iii) Hemiptera	30 (0.24)	52 (0.06)	71 (0.07)	161 (0.16)	81 (0.02)	277 (0.05)
iv) Hymenoptera	191 (0.24)	195 (0.05)	327 (0.07)	829 (0.18)	244 (0.11)	1328 (0.11)
v) Lepidoptera	41 (0.11)	56 (0.04)	83 (0.04)	155 (0.16)	65 (0.11)	281 (0.08)
Chi-square = 0.35; <i>P</i> >0.05						

<https://doi.org/10.1371/journal.pone.0199965.t001>

was observed among the major orders at each site, although these differences varied among sites ($\chi^2 = 54.98, P < 0.05$) (Table 1). Considering all sites, success in barcode recovery was higher for Lepidoptera (97%) and Diptera (92%) than for Coleoptera (70%) and Hemiptera/Hymenoptera (75%/76%).

By considering sequence matches to records on BOLD, the specimens with barcodes were assigned to 254 families. Most of these families (216/254) belonged to five orders: Diptera (61),

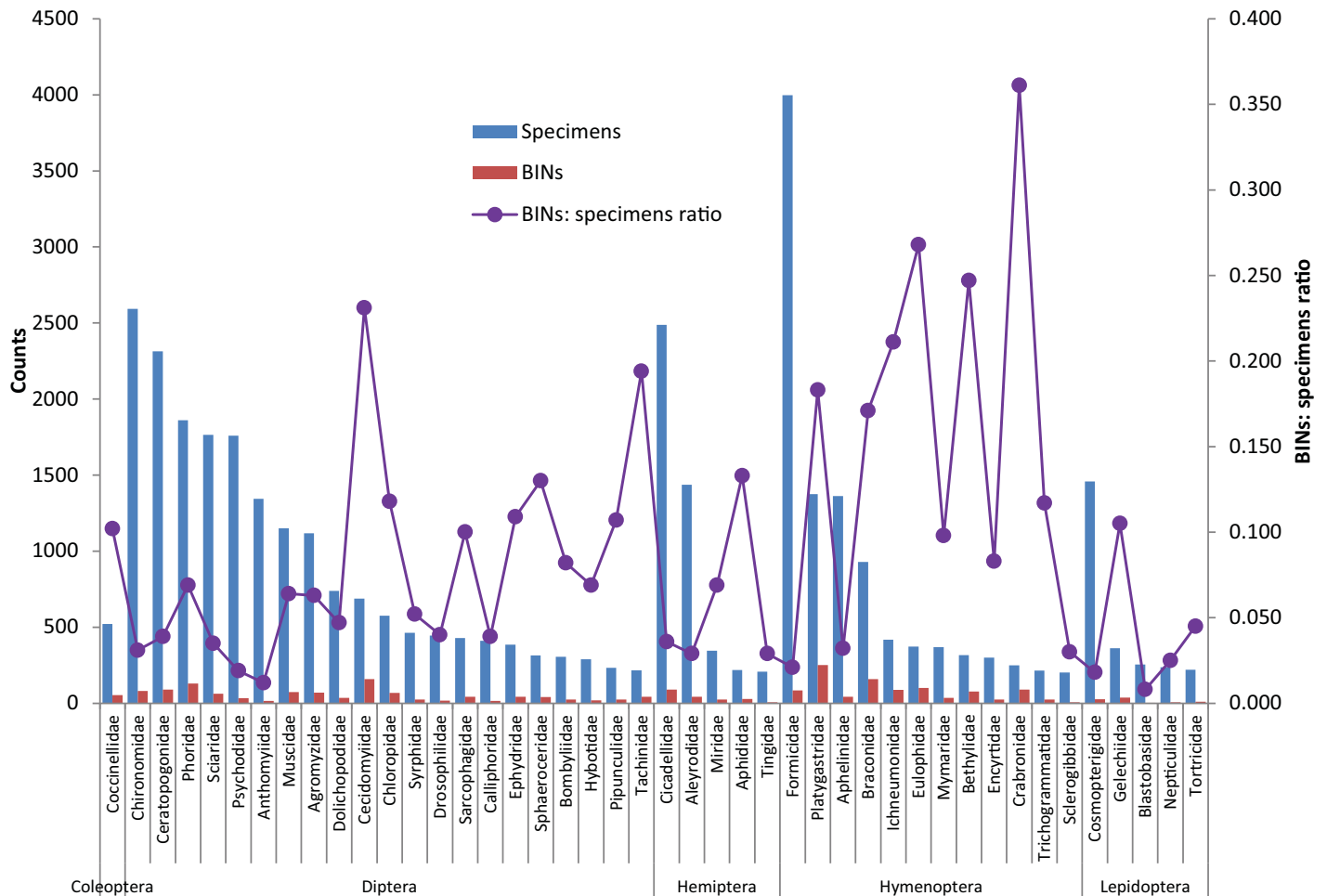


Fig 2. Abundance of families with their corresponding BINs (bars) and the BIN/specimen ratio (line). Only families with >100 specimens were analyzed.

<https://doi.org/10.1371/journal.pone.0199965.g002>

Hymenoptera (44), Coleoptera (42), Lepidoptera (38), and Hemiptera (31). The samples from Pakistan included 214 families while 132 were collected in Saudi Arabia and 164 in Egypt (129 at site 1 and 137 at site 2).

The 42,510 qualifying sequences were assigned to 3,682 BINs with 2,248 (61%) derived from Pakistan, 728 (20%) from Saudi Arabia, and 991 (27%) from Egypt (571 = 16% at site 1; 636 = 17% at site 2). 1,601 sequences missed the BIN assignment due to short reads (<507 b) while 101 due to high number of ambiguous bases (>1%). Another 220 sequences could not be assigned to a BIN, though they met the quality criteria. Moreover, more than half (1192/1,922) of the sequences that failed a BIN assignment could be associated with a known BIN through NJ clustering. Nearly half (43%) of all BINs (1,566/3,682) were represented by a single specimen.

The BIN/specimen ratio (0.06 vs 0.23) was lower for the most frequent order, Diptera ($n = 22,355$) than for the least frequent, Coleoptera ($n = 2,262$) (Table 1). This pattern was also evident at a family level as the three commonest families had lower BIN/specimen ratios than many of the infrequently collected families. For example, the Formicidae ($n = 3,997$, ratio = 0.02), Chironomidae ($n = 2,593$, ratio = 0.03) and Cicadellidae ($n = 2,488$, ratio = 0.04)

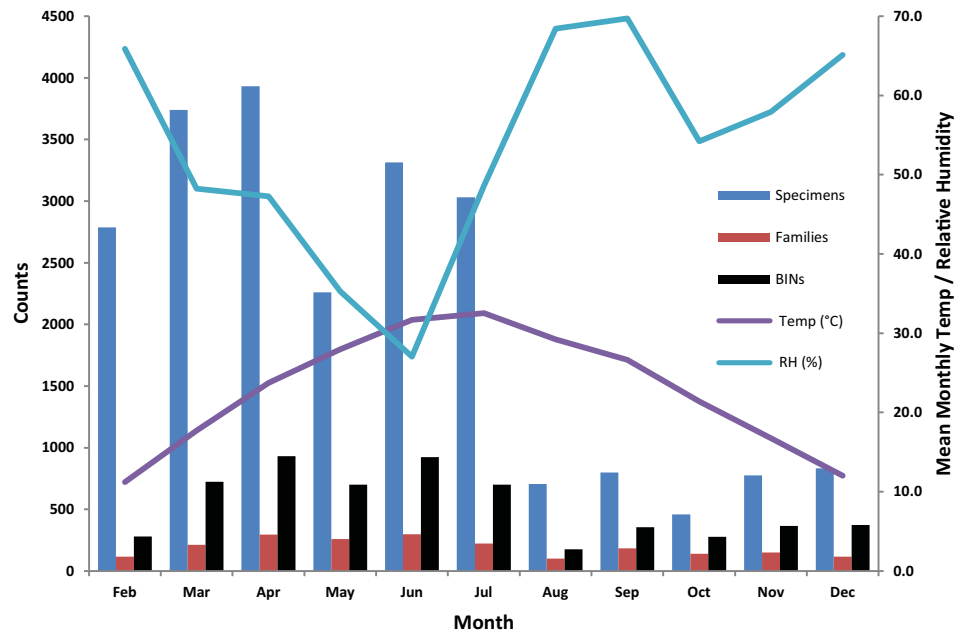


Fig 3. Number of insects, families, and BINs (bars) collected by the Malaise trap deployed in Islamabad, Pakistan versus temperature and relative humidity data (lines).

<https://doi.org/10.1371/journal.pone.0199965.g003>

had much lower ratios than the uncommon Crabronidae ($n = 249$, ratio = 0.36), Eulophidae ($n = 373$, ratio = 0.27) and Bethylidae ($n = 316$, ratio = 0.25) (Fig 2).

Among the 254 insect families, Formicidae was the most abundant ($n = 3997$) while the Crabronidae was the most diverse (BIN: specimen ratio = 0.36). While Formicidae dominated both ($n = 1896$) Egyptian sites, Chironomidae ($n = 2422$) was most common in Pakistan, and Cicadellidae ($n = 2146$) in Saudi Arabia. Eulophidae had the highest BIN: specimen ratio (0.27) in Egypt, while Cecidomyiidae had the highest (0.35) in Pakistan, and Platygastriidae (0.27) in Saudi Arabia.

Diversity analysis at the Pakistan site

The 19,068 specimens from Pakistan with barcodes included representatives of 15 orders, 214 families, and 2,248 BINs. Specimens of Diptera (11,126), Hymenoptera (4,728), Hemiptera (1,026), Lepidoptera (970), and Coleoptera (877) dominated the collections with only two other orders contributing more than 100 records (Orthoptera and Psocodea were each represented by six families and by 29/16 BINs respectively). More specimens (16,031 versus 6,593) and BINs (1,592 versus 1,155) were collected from February–June than from July–December, a difference associated with the higher temperatures and lower humidity in the first half of the year (Fig 3).

Most families (54) and BINs (1,090) were encountered only once in the 39 collection events, but two families (Agromyzidae, Formicidae) and one BIN (ACF1938 –Formicidae) were detected in most weeks (38 and 28, respectively) (Fig 4A and 4B). Just 27 BINs were represented by >100 records and only 26 occurred in more than a third of the collection events (S1A and S1B Fig). The number of families and BINs varied significantly ($P < 0.01$) among collection events (Table 2). Diversity indices for family and BIN richness in each collection event were determined by Simpson's D and Shannon's H and compared among collection events (Table 2). Simpson's D for families ranged from 0.20–0.97 and for BINs from 0.50–0.99, while

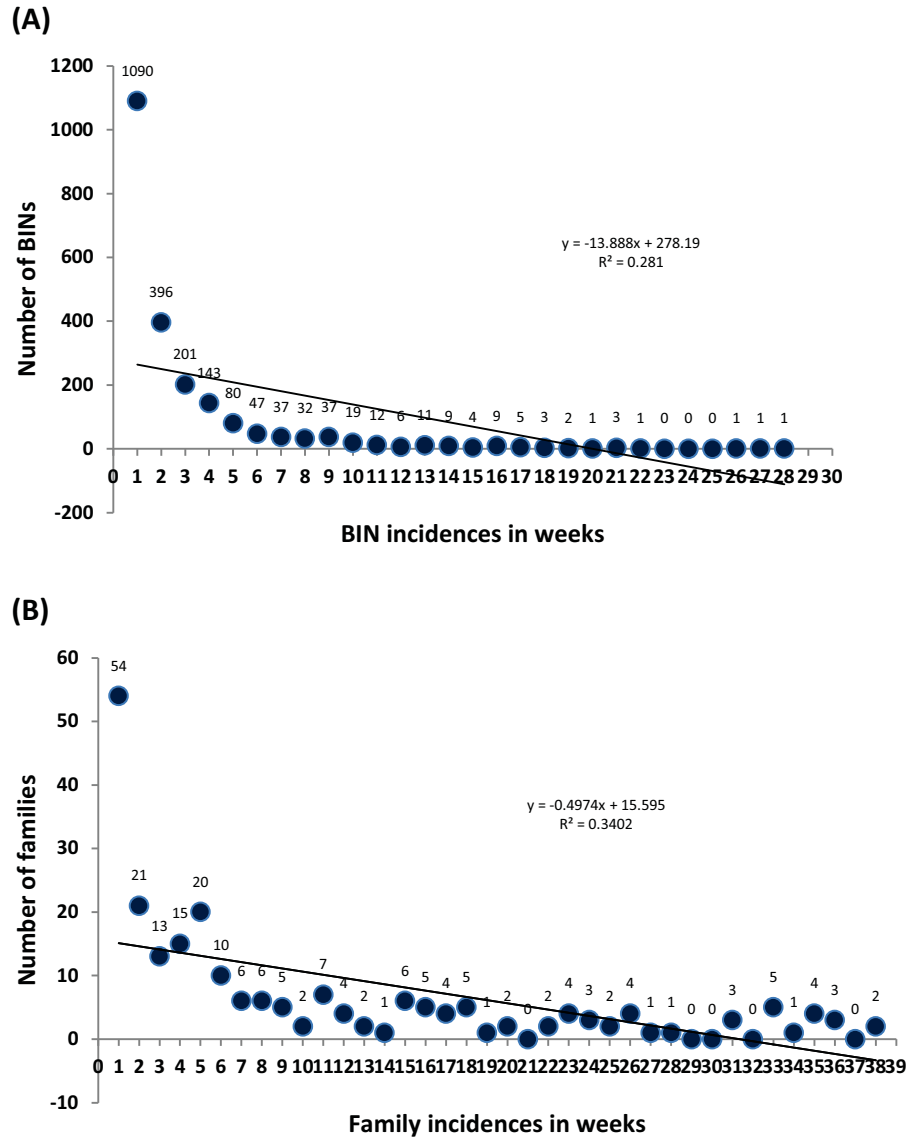


Fig 4. Incidence of BINs (A) and families (B) in the 39 collections from Islamabad, from February to December 2012.

<https://doi.org/10.1371/journal.pone.0199965.g004>

Shannon's H for families ranged from 0.63–3.70 and for BINs from 1.50–5.10. A contingency χ^2 test indicated there was no significant variation in family/BIN diversity among collection events.

BIN overlap among collection sites

BIN assignments allowed quantification of the overlap in species assemblages among the four sites. Among the 2,248 BINs collected in Pakistan and the 728 in Saudi Arabia, just 2% were shared (Fig 5). The overlap between Egypt (991 BINs from the two trap sites) and Saudi Arabia was 7% while that between Egypt and Pakistan was only 4%. The BIN overlap between the two Egyptian sites was higher (19%), but the sites were just 45 km apart. The BIN accumulation curves showed no sign of reaching an asymptote at any of the sites (Fig 6).

Table 2. Number of insect specimens, barcodes, families and BINs recovered from weekly Malaise trap collections.

Month	Week	Specimens	Barcodes	Number of Families (D / H)	Number of BINs (D / H)
February	1	1495	1454	35 (0.2 / 0.63)	91 (0.51 / 1.48)
	2	1045	1027	35 (0.44 / 1.2)	105 (0.72 / 2.23)
	3	248	228	41 (0.89 / 2.89)	86 (0.94 / 3.69)
March	4	534	508	34 (0.86 / 2.4)	105 (0.94 / 3.47)
	5	851	819	34 (0.72 / 1.87)	113 (0.84 / 2.74)
	6	1441	1363	70 (0.91 / 2.97)	280 (0.94 / 4.13)
April	7	914	843	71 (0.94 / 3.24)	237(0.97 / 4.48)
	8	757	698	61 (0.93 / 3.18)	206 (0.97 / 4.38)
	9	1541	1323	87 (0.94 / 3.34)	306 (0.97 / 4.5)
May	10	1035	863	87 (0.95 / 3.47)	258 (0.98 / 4.72)
	11	598	546	61 (0.92 / 3.05)	184 (0.97 / 4.35)
	12	681	615	66 (0.92 / 3.0)	193 (0.96 / 4.18)
June	13	718	549	78 (0.95 / 3.5)	232 (0.99 / 4.91)
	14	488	401	54 (0.88 / 2.87)	136 (0.91 / 3.71)
	15	372	322	62 (0.95 / 3.44)	161 (0.99 / 4.67)
July	16	1027	799	80 (0.93 / 3.44)	285 (0.91 / 4.95)
	17	267	200	50 (0.96 / 3.50)	116 (0.99 / 4.46)
	18	627	501	71 (0.95 / 3.54)	224 (0.99 / 5.0)
August	19	1392	981	94 (0.95 / 3.53)	356 (0.99 / 5.11)
	20	505	324	58 (0.95 / 3.46)	168 (0.99 / 4.71)
	21	2063	1543	105 (0.89 / 3.2)	409 (0.96 / 4.8)
September	22	454	349	59 (0.93 / 3.28)	158 (0.98 / 4.53)
	23	415	200	57 (0.97 / 3.66)	112 (0.99 / 4.41)
	24	138	61	23 (0.83 / 2.41)	37 (0.96 / 3.24)
October	25	153	93	25 (0.73 / 1.91)	39(0.93 / 3.03)
	26	128	103	24 (0.78 / 2.22)	41 (0.9 / 2.95)
	27	281	178	57 (0.93 / 3.43)	117 (0.99 / 4.47)
November	28	190	163	51 (0.96 / 3.49)	94 (0.97 / 4.11)
	29	199	154	48 (0.97 / 3.56)	112 (0.99 / 4.51)
	30	192	166	47 (0.97 / 3.06)	91 (0.99 / 4.29)
December	31	62	57	30 (0.94 / 3.05)	45 (0.99 / 3.72)
	32	52	44	21 (0.95 / 2.85)	36 (0.99 / 3.5)
	33	153	135	41 (0.96 / 3.35)	110 (0.99 / 4.38)
Chi-square	34	83	78	30 (0.91 / 2.95)	53 (0.98 / 3.75)
	35	471	422	60 (0.95 / 3.42)	203 (0.99 / 4.79)
	36	14	13	9 (0.86 / 1.84)	9 (0.94 / 2.09)
Chi-square	37	208	184	42 (0.96 / 3.35)	115 (0.99 / 4.46)
	38	431	395	54 (0.95 / 3.32)	194 (0.99 / 4.79)
	39	401	366	59 (0.96 / 3.43)	190 (0.99 / 4.79)
Chi-square		$P < 0.01$	$P < 0.01$	$P < 0.01 (P > 1 / P > 1)$	$P < 0.01 (P > 1 / P > 1)$

D = Simpson's diversity index (Simpson 1949); H = Shannon's diversity index (Shannon and Weaver 1948).

<https://doi.org/10.1371/journal.pone.0199965.t002>

Discussion

Comprehensive assessments of the insect species present at even a single locale have, until now, been impossible. For example, the analysis of 129,494 insects from a single 0.5 hectare site in Panama required contributions from 102 taxonomists over an 8-year interval to place

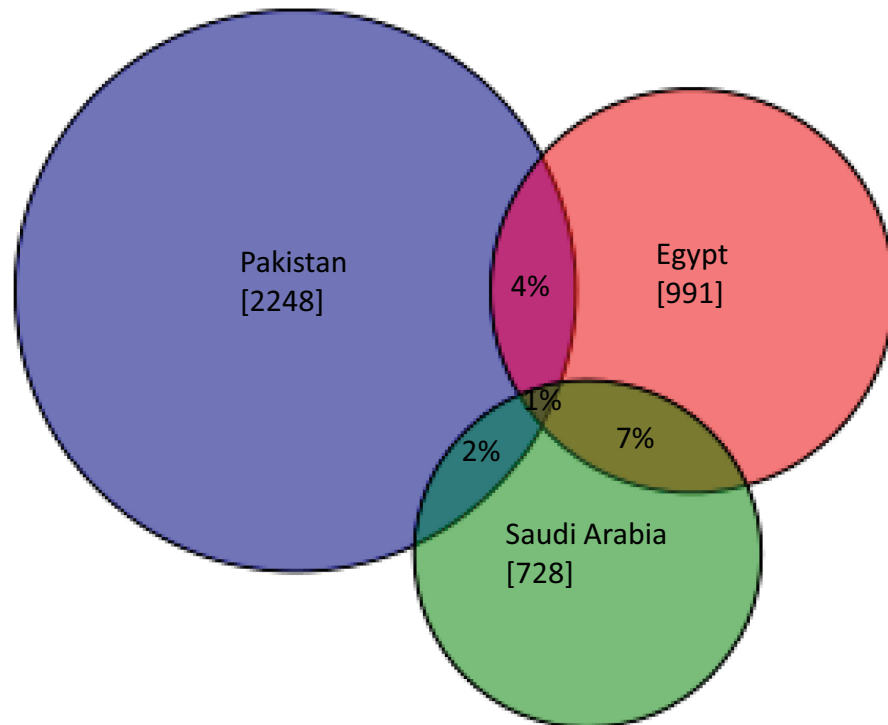


Fig 5. BIN overlap among the three countries—Egypt, Pakistan, Saudi Arabia.

<https://doi.org/10.1371/journal.pone.0199965.g005>

them to 6,144 putative species [47]. Despite this massive effort, just 23% of the specimens gained a formal species assignment, and nearly half of all specimens had to be excluded from the identification effort because they belonged to groups lacking taxonomic specialists. Viewed from this perspective, the capacity of a single Malaise trap to gather 30,000–50,000 specimens per year represents an overwhelming challenge for any identification effort employing morphological approaches. The present study overcame this barrier by using DNA barcoding and the BIN system to assign each specimen to a putative species. Although prior studies have demonstrated the power of this approach for work on well-known faunas [19,48,49], the present study breaks new ground by employing this method to explore insect diversity in a region with limited biodiversity knowledge.

This study examined 53,000 specimens derived from single traps deployed for 17–39 weeks at sites in Egypt, Pakistan, and Saudi Arabia. As these deployments totaled 93 weeks, trap catches averaged 570 specimens a week. As reported in prior studies [19,48,50], Diptera and Hymenoptera dominated the catch. BIN compliant sequences were recovered from 84% of the specimens, a value similar to those reported for specimens from Malaise traps deployed in Canada [19] and Europe [48]. Recovery did vary among insect orders, being higher for Diptera and Lepidoptera than for Coleoptera, Hemiptera and Hymenoptera, a pattern documented in prior studies that likely reflects primer mismatches [19]. Targeting shorter sequence fragments or employing HTS approaches [51] may alleviate the issues related to primer mismatches.

The abundance and diversity of collected insects varied over the collection period. The pattern of bigger and diverse catches in earlier than the later months of the year coincided with a rise in temperature from March to June and a fall from September to December. Weather is known to influence both spatial and temporal patterns of insect communities [52]. It is also

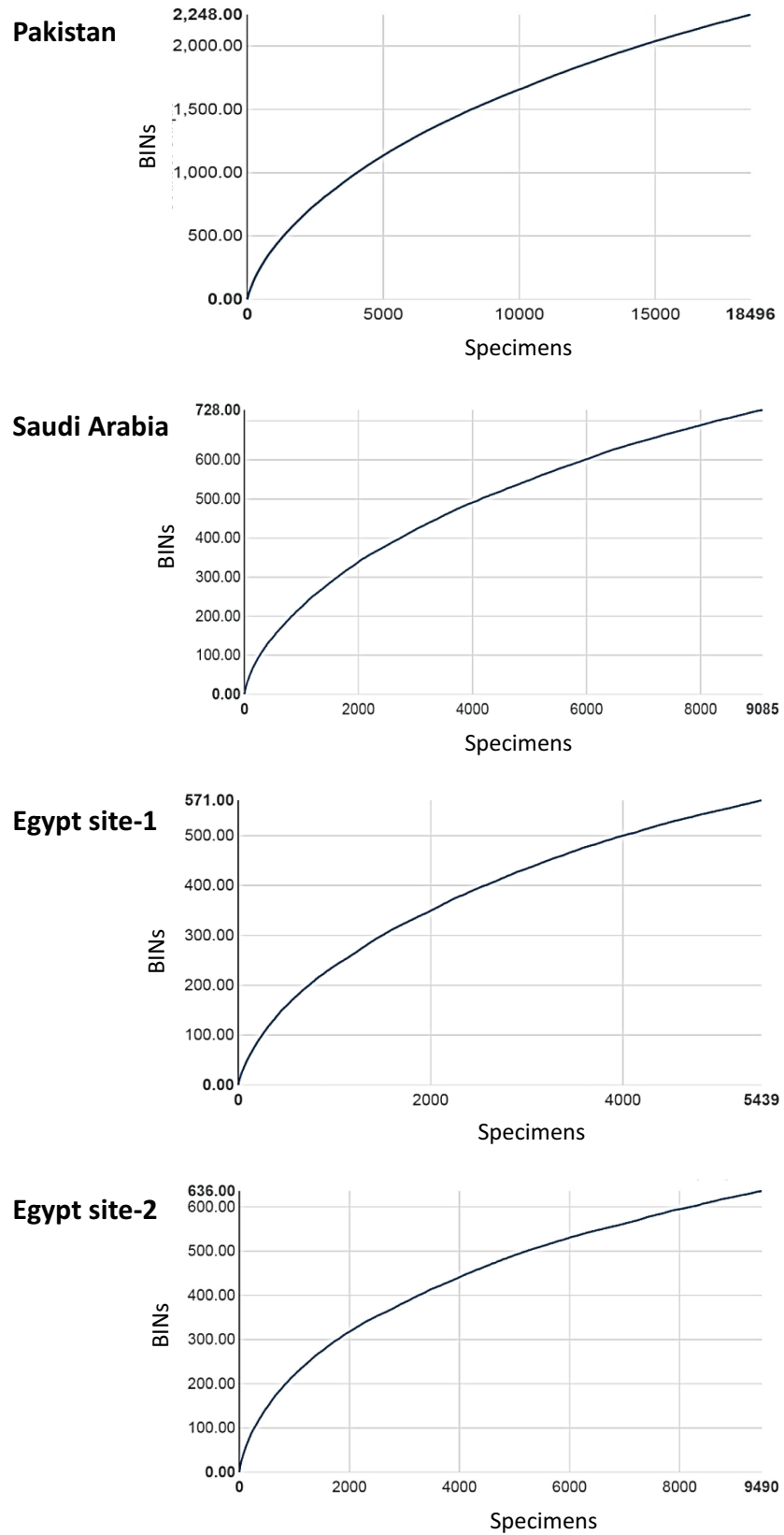


Fig 6. Accumulation curve for Barcode Index Numbers (BINs) recovered from insect samples collected at the four Malaise trap sites.

<https://doi.org/10.1371/journal.pone.0199965.g006>

known that insect emergence is driven by temperature that also affects their development, survival and abundance [53].

Analysis indicated that the 44,000 specimens which generated sequences belonged to 3,682 BINs with representatives of 254 families and 17 insect orders. These samples could not have been processed morphologically because such a large fraction of the Saharo-Arabian insect fauna is undescribed. For example, 20% of the species encountered in a recent survey of the United Arab Emirates were new, even though the groups analyzed were among those with the best taxonomy [27]. Because of this taxonomic barrier, the species present in Malaise trap samples from the Saharo-Arabian region have never been comprehensively assessed. Despite this lack of directly comparable data, the present study has shown that the barcode analysis of specimens from brief sampling of a few sites recovered half as many insect species as reported from all prior studies in these nations. The present results further indicated that more sampling is required to ascertain the number of species in each nation and the extent of overlap in their faunas. Certainly, the BIN overlap values (2–7%) reported in this study are underestimates of actual faunal overlap because of under-sampling. Because measures of endemism [54] play such an important role in conservation planning [55,56], further surveys of insect diversity in the Saharo-Arabian region are needed. These surveys can potentially employ several other methods of passive collection such as emergence traps, light traps, pitfall traps and coloured pan traps. The lower level of sequence recovery observed in samples from one of the study locations could be related to poor preservation or storage condition of specimens [46]. The issue of deteriorating DNA quality in preserved samples may be resolved by using better preservatives, such as 95% ethanol [57]. While the sampling effort required to properly estimate overall diversity and overlap values cannot be determined without further sampling, the analysis of a Malaise trap catch from each ecoregion within these nations would represent an important first step.

This present study has demonstrated how the BIN system can circumvent the barriers imposed by limited access to taxonomic specialists and by the fact that many insect species in the Saharo-Arabian region are undescribed. As such, it demonstrates how biodiversity analysis can be accelerated in regions that have seen little exploration [58]. Because it remains critical to extend the DNA barcode reference library through specimen-based analysis [59], it is important that new sequencing platforms are leading to substantial reductions in analytical costs [51]. Moreover, once libraries are well parameterized, metabarcoding studies [50,60] will permit detailed tracking of biodiversity trajectories at scales that would otherwise be impossible.

Supporting information

S1 Fig. Spatial (A) and temporal (B) abundance of BINs. (A) shows BINs with at least 100 specimens in the total collection while (B) shows BINs detected in at least 14 of 39 collection events.

(PDF)

Acknowledgments

We thank Dr. Muhammad Rafique, Director General of the Pakistan Museum of Natural History, for permission to deploy a Malaise trap and for facilitating sample collections. We thank Ahmad Atef, Ahmad Bahieldin, Sherif Edris, Nahid H. Hajrah and Haitham A. Yacoub for aid in collecting samples within Saudi Arabia.

Author Contributions

Conceptualization: Paul D. N. Hebert.

Data curation: Muhammad Ashfaq, Kate Perez, Valerie Levesque-Beaudin, Carlene Gallant, Joseph Addesi, Paul D. N. Hebert.

Formal analysis: Muhammad Ashfaq, Kate Perez, Valerie Levesque-Beaudin.

Funding acquisition: Muhammad Ashfaq, Jamal S. M. Sabir, Paul D. N. Hebert.

Investigation: Muhammad Ashfaq, Hosam O. El-Ansary, Kate Perez, Valerie Levesque-Beaudin, Arif M. Khan, Akhtar Rasool, Carlene Gallant, Joseph Addesi, Paul D. N. Hebert.

Methodology: Muhammad Ashfaq, Jamal S. M. Sabir, Kate Perez, Valerie Levesque-Beaudin, Arif M. Khan, Akhtar Rasool, Carlene Gallant, Joseph Addesi, Paul D. N. Hebert.

Project administration: Muhammad Ashfaq, Jamal S. M. Sabir, Hosam O. El-Ansary, Kate Perez, Paul D. N. Hebert.

Resources: Muhammad Ashfaq, Jamal S. M. Sabir, Paul D. N. Hebert.

Software: Muhammad Ashfaq, Valerie Levesque-Beaudin.

Supervision: Muhammad Ashfaq, Jamal S. M. Sabir, Hosam O. El-Ansary, Akhtar Rasool, Paul D. N. Hebert.

Validation: Paul D. N. Hebert.

Writing – original draft: Muhammad Ashfaq.

Writing – review & editing: Paul D. N. Hebert.

References

1. Azam I, Afsheen S, Zia A, Javed M, Saeed R, Sarwar MK, et al. Evaluating insects as bioindicators of heavy metal contamination and accumulation near industrial area of Gujrat, Pakistan. *BioMed Res Int*. 2015; <http://dx.doi.org/10.1155/2015/942751>.
2. Gerlach J, Samways M, Pryke J. Terrestrial invertebrates as bioindicators: An overview of available taxonomic groups. *J Insect Conserv*. 2013; 17: 831–850.
3. Samways MJ. Insects in biodiversity conservation: some perspectives and directives. *Biodiver Conserv*. 1993; 2: 258–282.
4. McMahon SM, Harrison SP, Armbruster WS, Bartlein PJ, Beale CM, Edwards ME, et al. Improving assessment and modelling of climate change impacts on global terrestrial biodiversity. *Trends Ecol Evol*. 2011; 26: 249–259. <https://doi.org/10.1016/j.tree.2011.02.012> PMID: 21474198
5. Miller SE, Rogo LM. Challenges and opportunities in understanding and utilization of African insect diversity. *Cimbebasia* 2001; 17: 197–218.
6. Adams WM, Aveling R, Brockington D, Dickson B, Elliott J, Hutton J, et al. Biodiversity conservation and the eradication of poverty. *Science* 2004; 306: 146–1149.
7. Ward DF, Lariviere M-C. Terrestrial invertebrate surveys and rapid biodiversity assessment in New Zealand: lessons from Australia. *New Zealand J Ecol*. 2004; 28: 151–159.
8. Scheffers BR, Joppa LN, Pimm SL, Laurance WF. What we know and don't know about earth's missing biodiversity. *Trends Ecol Evol*. 2012; 27: 501–510. <https://doi.org/10.1016/j.tree.2012.05.008> PMID: 22784409
9. Rafael JA, Aguiar AP, Amorim DS. Knowledge of insect diversity in Brazil: Challenges and advances. *Neotrop Entomol*. 2009; 38: 565–570. PMID: 19943001
10. Hebert PDN, Cywinska A, Ball SL, deWaard JR. Biological identifications through DNA barcodes. *Proc Biol Sci*. 2003; 270: 313–321. <https://doi.org/10.1098/rspb.2002.2218> PMID: 12614582
11. Wilson JJ, Sing KW, Floyd RM, Hebert PDN. DNA barcodes and insect biodiversity. <https://doi.org/10.1002/9781118945568.ch17>. In: Footitt RG & Adler PH, Editors. *Insect Biodiversity: Science and Society*. 2017. Vol. 1, 2nd Edition. Blackwell Publishing Ltd., Oxford.

12. Smith MA, Fisher BL, Hebert PDN. DNA barcoding for effective biodiversity assessment of a hyperdiverse arthropod group: the ants of Madagascar. *Phil Trans R Soc B*. 2005; 360: 1825–1834. <https://doi.org/10.1098/rstb.2005.1714> PMID: 16214741
13. Janzen DH, Hallwachs W, Blandin P, Burns JM, Cadiou JM, Chacon I, et al. Integration of DNA barcoding into an ongoing inventory of complex tropical biodiversity. *Mol Ecol Resour*. 2009; Suppl s1: 1–26. <https://doi.org/10.1111/j.1755-0998.2009.02628.x> PMID: 21564960
14. Janzen DH, Hallwachs W. DNA barcoding the Lepidoptera inventory of a large complex tropical conserved wildland, Area de Conservacion Guanacaste, northwestern Costa Rica. *Genome*. 2016; 59: 641–660. <https://doi.org/10.1139/gen-2016-0005> PMID: 27584861
15. Kerr KCR, Stoeckle MY, Dove CJ, Weigt LA, Francis CM, Hebert PDN. Comprehensive DNA barcode coverage of North American birds. *Mol Ecol Notes*. 2007; 7: 535–543. <https://doi.org/10.1111/j.1471-8286.2007.01670.x> PMID: 18784793
16. 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
17. Hausmann A, Godfray HCJ, Huemer P, Mutanen M, Rougerie R, van Nieukerken EJ, et al. Genetic patterns in European geometrid moths revealed by the Barcode Index Number (BIN) system. *PLOS ONE*. 2013; 8: e84518. <https://doi.org/10.1371/journal.pone.0084518> PMID: 24358363
18. Pentinsaari M, Hebert PDN, Mutanen M. Barcoding beetles: A regional survey of 1872 species reveals high identification success and unusually deep interspecific divergences. *PLOS ONE*. 2014; 9: e108651. <https://doi.org/10.1371/journal.pone.0108651> PMID: 25255319
19. Hebert PDN, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA, et al. Counting species with DNA barcodes: Canadian insects. *Phil Trans R Soc B*. 2016; 371: 20150333. <https://doi.org/10.1098/rstb.2015.0333> PMID: 27481785
20. Zahiri R, Lafontaine JD, Schmidt BC, deWaard JR, Zakharov EV, Hebert PDN. Probing planetary biodiversity with DNA barcodes: The Noctuoidea of North America. *PLOS ONE*. 2017; 12: e0178548. <https://doi.org/10.1371/journal.pone.0178548> PMID: 28570635
21. Cristescu ME. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends Ecol Evol*. 2014; 29: 566–571. <https://doi.org/10.1016/j.tree.2014.08.001> PMID: 25175416
22. Telfer AC, Young MR, Quinn J, Perez K, Sobel CN, Sones JE, et al. Biodiversity inventories in high gear: DNA barcoding facilitates a rapid biotic survey of a temperate nature reserve. *Biodiversity Data J*. 2015; 3: e6313.
23. Lees DC, Kawahara AY, Rougerie R, Ohshima I, Kawakita A, Bouteleux O, et al. DNA barcoding reveals a largely unknown fauna of Gracillariidae leaf-mining moths in the Neotropics. *Mol Ecol Resour*. 2013; 14: 286–296. <https://doi.org/10.1111/1755-0998.12178> PMID: 24119085
24. 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
25. Sing KW, Wang WZ, Wan T, Lee PS, Lee ZX, Chen X, et al. Diversity and human perceptions of bees in Southeast Asian megacities. *Genome*. 2016; 59: 851–865. <https://doi.org/10.1139/gen-2015-0166> PMID: 27608265
26. Holt BG, Lessard JP, Borregaard MK, Fritz SA, Araújo MB, Dimitrov D, et al. An update of Wallace's zoogeographic regions of the world. *Science*. 2013; 339: 74–78. <https://doi.org/10.1126/science.1228282> PMID: 23258408
27. von Harten A. Arthropod fauna of the UAE. 2008–2014; Volumes 1–5. Dar Al Ummah Publishers, Abu Dhabi, UAE.
28. El-Hawagry M, Khalil M, Sharaf M, Fadl H, Aldawood A. A preliminary study on the insect fauna of Al-Baha Province, Saudi Arabia, with descriptions of two new species. *ZooKeys*. 2013; 274: 1–88.
29. Malik IU, Faiz AH, Abbas F. Biodiversity assessment and its effect on the environment of Shakarparian forest. *J Biores Manag*. 2014; 1: 21–35.
30. Convention on Biological Diversity. Pakistan Fifth National Report to CBD. Climate Change Division, Government of Pakistan, Islamabad. 2014; 86pp.
31. Al-Ahmadi AZ, Salem MM. Entomofauna of Saudi Arabia: General survey of insects reported in the Kingdom of Saudi Arabia. Part 1: Checklist of Insects. Academic Publishing and Press, King Saud University. Riyadh, KSA. 1999; 240pp.
32. Salem MM. The insect fauna of Egypt: A comprehensive checklist of insects recorded in Egypt. Lambert Academic Publishing. 2017; 700pp.

33. Steinke D, Breton V, Berzitis E, Hebert PDN. The School Malaise Trap Program: Coupling educational outreach with scientific discovery. *PLOS Biol.* 2017; 15: e2001829. <https://doi.org/10.1371/journal.pbio.2001829> PMID: 28437475
34. Oliver I, Beattie AJ. Designing a cost-effective invertebrate survey: A test of methods for rapid assessment of biodiversity. *Ecol Appl.* 1996; 6: 594–607.
35. Wilkie L, Cassis G, Gray M. A quality control protocol for terrestrial invertebrate biodiversity assessment. *Biodivers Conserv.* 2003; 12: 121–146.
36. Russo L, Stehouwer R, Heberling JM, Shea K. The composite insect trap: An innovative combination trap for biologically diverse sampling. *PLOS ONE.* 2011; 6: e21079. <https://doi.org/10.1371/journal.pone.0021079> PMID: 21698160
37. Hutcheson J, Jones D. Spatial variability of insect communities in a homogenous system: Measuring biodiversity using Malaise trapped beetles in a *Pinus radiata* plantation in New Zealand. *Forest Ecol Manag.* 1999; 118: 93–105.
38. Hill D, Fasham M, Tucker G, Shewry M, Shaw P. *Handbook of biodiversity methods: Survey, evaluation and monitoring.* Cambridge University Press. 2005; 573p.
39. Porco D, Rougerie R, Deharveng L, Hebert PDN. 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
40. Hebert PDN, deWaard JR, Zakharov EV, Prosser SWJ, Sones JE, McKeown JTA, 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
41. Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R. DNA primers for amplification of mitochondrial cytochrome c oxidase subunit I from diverse metazoan invertebrates. *Mol Marine Biol Biotechnol.* 1994; 3: 294–299. PMID: 7881515
42. Park DS, Suh SJ, Hebert PDN, Oh HW, Hong KJ. DNA barcodes for two scale insect families, mealybugs (Hemiptera: Pseudococcidae) and armored scales (Hemiptera: Diaspididae). *Bull Entomol Res.* 2011; 101: 429–434. <https://doi.org/10.1017/S0007485310000714> PMID: 21272395
43. Ratnasingham S, Hebert PDN. BOLD: The Barcode of Life Data System (www.barcodinglife.org). *Mol Ecol Notes.* 2007; 7: 355–364. <https://doi.org/10.1111/j.1471-8286.2007.01678.x> PMID: 18784790
44. Simpson EH. Measurement of diversity. *Nature.* 1949; 163: 688.
45. Shannon CE, Weaver W. A mathematical theory of communication. *Bell Syst Tech J.* 1948; 27: 379–423 and 623–656.
46. Stein ED, White BP, Mazor RD, Miller PE, Pilgrim EM. Evaluating ethanol-based sample preservation to facilitate use of DNA barcoding in routine freshwater biomonitoring programs using benthic macroinvertebrates. *PLOS ONE.* 2013; 8: e51273. <https://doi.org/10.1371/journal.pone.0051273> PMID: 23308097
47. Bassett Y, Cizek L, Cuénoud P, Didham RK, Guilhaumon F, Missa O, et al. Arthropod diversity in a tropical forest. *Science.* 2012; 338: 1481–1484. <https://doi.org/10.1126/science.1226727> PMID: 23239740
48. Geiger MF, Moriniere J, Hausmann A, Haszprunar G, Wägele W, Hebert PDN, et al. Testing the Global Malaise Trap Program—How well does the current barcode reference library identify flying insects in Germany? *Biodiv Data J.* 2016; 4: e10671.
49. Moriniere J, de Araujo BC, Lam AW, Hausmann A, Balke M, Schmidt S, et al. Species identification in Malaise trap samples by DNA barcoding based on NGS technologies and a scoring matrix. *PLOS ONE.* 2016; 11: e0155497. <https://doi.org/10.1371/journal.pone.0155497> PMID: 27191722
50. Brandon-Mong GJ, Littlefair JE, Sing KW, Lee YP, Gan HM, Clare EL, et al. Temporal changes in arthropod activity in tropical anthropogenic forests. *Bull Entomol Res.* 2018; 14: 1–8.
51. Hebert PDN, Braukmann TWA, Prosser SWJ, Ratnasingham S, deWaard JR, Ivanova NV, et al. A Sequel to Sanger: Amplicon sequencing that scales. *BMC Genomics.* 2018; 19: 219. <https://doi.org/10.1186/s12864-018-4611-3> PMID: 29580219
52. Gandhi KJK, Gilmore DW, Katovich SA, Mattson WJ, Spence JR, Seybold SJ. Physical effects of weather events on abundance and diversity of insects in North American forests. *Environ Rev.* 2007; 15: 113–152.
53. Bale JS, Masters GJ, Hodkinson ID, Awmack C, Bezemer TM, Brown VK, et al. Herbivory in global climate change research: direct effects of rising temperature on insect herbivores. *Global Change Biol.* 2002; 8: 1–16.
54. 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: 832–843. <https://doi.org/10.1111/1755-0998.12131> PMID: 23789612

55. Gaston KJ. Global patterns in biodiversity. *Nature*. 2000; 405: 220–227. <https://doi.org/10.1038/35012228> PMID: 10821282
56. Lamoreux JF, Morrison JC, Ricketts TH, Olson DM, Dinerstein E, McKnight MW, et al. Global tests of biodiversity concordance and the importance of endemism. *Nature*. 2006; 440: 212–214. <https://doi.org/10.1038/nature04291> PMID: 16382239
57. Moreau CS, Wray BD, Czekanski-Moir JE, Rubin BER. DNA preservation: a test of commonly used preservatives for insects. *Invertebr Syst*. 2013; 27: 81–86.
58. Lawton JH, Bignell DE, Bolton B, Bloemers GF, Eggleton P, Hammond PM, et al. Biodiversity inventories, indicator taxa and effects of habitat modification in tropical forest. *Nature*. 1998; 391: 72–75.
59. Joly S, Davies TJ, Archambault A, Bruneau A, Derry A, Kembel SW, et al. Ecology in the age of DNA barcoding: the resource, the promise and the challenges ahead. *Mol Ecol Resour*. 2013; 14: 221–232. <https://doi.org/10.1111/1755-0998.12173> PMID: 24118947
60. Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, et al. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods Ecol Evol*. 2012; 3: 613–623.