Using DNA metabarcoding to assess insect diversity in citrus orchards

Chenxi Liu [Corresp., 1], **Muhammad Ashfaq** [2], **Yanfang Yin** [1], **Yanjuan Zhu** [1], **Zhen Wang** [1], **Hongmei Cheng** [1], **Paul Hebert** [2]

[1] Sino-American Biological Control Laboratory, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China

[2] Centre for Biodiversity Genomics and Department of Integrative Biology, University of Guelph, Guelph, Ontario, Canada

Corresponding Author: Chenxi Liu
Email address: liuchenxi2004@126.com

Background. DNA metabarcoding is rapidly emerging as a cost-effective approach for large-scale biodiversity assessment and pest monitoring. The current study employed metabarcoding to assess insect diversity in citrus orchards in Ganzhou city, Jiangxi, China in 2018-2019. Insects were sampled using Malaise traps deployed in three citrus orchards producing a total of 43 pooled monthly samples.

Methods. The Malaise trap samples were sequenced following DNA metabarcoding workflow. Generated sequences were curated and analyzed using two cloud-based data storage and analytical platforms, the Barcode of Life Data System (BOLD) and Multiplex Barcode Research And Visualization Environment (mBRAVE).

Results. These platforms assigned the sequences to 2,141 Barcode Index Numbers (BINs), a species proxy . Only 63% of the BINs were shared among the three sampling sites (J=0.63) while BIN sharing between any two sites did not exceed 75% (J=0.73 to 0.75). Shannon diversity index (H') showed a similar pattern of BIN assortment at the 3 sampling sites. Comparison of BIN records against all those on BOLD made it possible to identify 40% of the BINs to a species, 57% to a genus, 97% to a family and 99% to an order. BINs which received a species match on BOLD were placed in one of four categories based on this assignment: pest, parasitoid, predator, or pollinator. As this study provides the first baseline data on insect biodiversity in Chinese citrus plantations, it is a valuable resource for research in a broad range of areas such as pest management and monitoring beneficial insects in citrus gardens.

Using DNA metabarcoding to assess insect diversity in citrus orchards

Chenxi Liu [1], Muhammad Ashfaq [2], Yanfang Yin [1], Yanjuan Zhu [1], Zhen Wang [1], Hongmei Cheng [1], Paul D. N. Hebert [2]

1 Sino-American Biological Control Laboratory, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing, China, 100193

2 Centre for Biodiversity Genomics and Department of Integrative Biology, University of Guelph, Guelph, ON, Canada, N1G 2W1

Corresponding author:

Chenxi Liu [1]

No. 2 Yuanmingyuan West Road, Haidian District, Beijing, 100193, China

Email address: liuchenxi@caas.cn

Abstract

Background. DNA metabarcoding is rapidly emerging as a cost-effective approach for large-scale biodiversity assessment and pest monitoring. The current study employed metabarcoding to assess insect diversity in citrus orchards in Ganzhou city, Jiangxi, China in 2018-2019. Insects were sampled using Malaise traps deployed in three citrus orchards producing a total of 43 pooled monthly samples.

Methods. The Malaise trap samples were sequenced following DNA metabarcoding workflow. Generated sequences were curated and analyzed using two cloud-based data storage and analytical platforms, the Barcode of Life Data System (BOLD) and Multiplex Barcode Research And Visualization Environment (mBRAVE).

Results. These platforms assigned the sequences to 2,141 Barcode Index Numbers (BINs), a species proxy. Only 63% of the BINs were shared among the three sampling sites (J=0.63) while BIN sharing between any two sites did not exceed 75% (J=0.73 to 0.75). Shannon diversity index (H') showed a similar pattern of BIN assortment at the 3 sampling sites. Comparison of BIN records against all those on BOLD made it possible to identify 40% of the BINs to a species, 57% to a genus, 97% to a family and 99% to an order. BINs which received a species match on BOLD were placed in one of four categories based on this assignment: pest, parasitoid, predator, or pollinator. As this study provides the first baseline data on insect biodiversity in Chinese citrus plantations, it is a valuable resource for research in a broad range of areas such as pest management and monitoring beneficial insects in citrus gardens.

Introduction

China ranks among the top three countries globally for citrus cultivation and production (Caserta et al. 2020). In fact, it ranked first in the world in 2017-2018, producing 32 million tons of fresh citrus (CGA 2020). Citrus fruit is considered a superior agricultural product and its production represents an important industry in the rural areas of southern China. However, crop yields are severely impacted by an array of pests. A large number of insect species, both pest and beneficial taxa (Niu et al. 2014) occur on citrus but their identification is difficult. Cryptic morphology and lack of taxonomic expertise are the major challenges in large-scale insect diversity assessments.

DNA barcoding (Hebert et al. 2003) the characterization of sequence variation in a standard 658 bp fragment of cytochrome *c* oxidase I gene (COI), has gained global acceptance for specimen identification and species discovery (Ashfaq et al. 2017; Gwiazdowski et al. 2015; Hebert et al. 2004; Taberlet et al. 2012). This method circumvents the limitations of morphology by identifying unknown organisms by matching their barcode sequences to reference sequences. The simplicity and reliability of this method has motivated expansion of the DNA barcode reference library in the Barcode of Life Data System (BOLD) (www.boldsystems.org). The BOLD system assigns eligible barcode sequences (>507 bp, <1% ambiguous bases, no stop codons, no contamination) to Barcode Index Numbers (BINs) which serve as proxy for species (Ratnasingham & Hebert 2013). BOLD currently holds seven million insect barcodes which have been assigned to more than 700,000 BINs. Implementation of the BIN system has enhanced the ability of DNA barcoding to discern and count species (Hebert et al. 2016), asses biodiversity composition (Ashfaq et al. 2018; Telfer et al. 2015), to map species distributions (Ashfaq et al. 2017), and to track species movements across borders (Ren et al. 2017). This success has led to the use of BINs in the analysis of bulk samples and biodiversity studies by high-throughput sequencing (HTS) (Cristescu 2014).

DNA metabarcoding is a developing approach that identifies the species present in a mixed sample (bulk DNA or environmental DNA) based on HTS of a specific DNA marker (Comtet et al. 2015; Hajibabaei et al. 2011; Moriniere et al. 2016; Yu et al. 2012). It differs from conventional DNA barcoding (usually based on Sanger sequencing of individual specimens)

because HTS allows taxonomy to be assigned to hundreds or even thousands of species in a bulk sample. It achieves this goal by generating amplicons of the barcode region from bulk DNA extracts which are then sequenced and assigned to operational taxonomic units (OTUs) that are queried against reference sequences to determine their source species (Cristescu 2014). Studies have now employed this approach to assess species composition in biological communities such as aquatic and terrestrial arthropods (Beng et al. 2016; Braukmann et al. 2019; Elbrecht & Steinke 2018; Ji et al. 2013) making metabarcoding an increasingly cost-effective approach for large-scale biodiversity studies.

Widespread interest in metabarcoding has resulted in data proliferation and the development of computational tools to aid data analysis. As the BIN system has offered a novel approach to circumvent morphological bottlenecks to discriminate species, pairing of BINs with HTS has accelerated biodiversity assessments. Prior studies have used metabarcoding to survey insect diversity in different ecological settings in China (Huang et al. 2022), but reports on the use of this technology to explore insect diversity in fruit gardens in this region are lacking. The current study aimed to fill this gap by coupling metabarcoding with the BIN system to explore insect diversity in citrus plantations. The composition of insect communities in citrus orchards was analyzed using DNA metabarcoding followed by data analysis on BOLD and mBRAVE. The species revealed by BIN matches on BOLD were then searched in citrus pest database – Citrus pest information system (CPIS, cpis.hzau.edu.cn) and the literature to allow their classification into pest, parasitoid, predator, or pollinator (Niu et al. 2014; Smaili et al. 2020; Urbaneja et al. 2020). The results provide a valuable resource for research on citrus pest management and beneficial insects. Exploration of pest and beneficial insect species in Chinese citrus orchards would allow scientists to screen the appropriate agents for pest management programs and promote pest biological control.

Materials and Methods

A single Malaise trap (deWaard et al. 2019; Geiger et al. 2016) was deployed in three citrus orchards (GAN, QIU, SHI) about one kilometer apart in Ganzhou city, Jiangxi province, China (Figure 1). Samples were collected from April 2018 to July 2019 by replacing the collection bottles every month – so 43 samples were obtained from the three traps (Table S1). The collection bottles were stored at -20 ℃ until used. Specimens from each bottle were sorted into two size categories: small (e.g., parasitic wasps, *Drosophila*) and large (e.g., butterflies, locust). Large specimens were subsampled to obtain a tissue block (legs or partial abdomen) of a size similar to the small ones while small specimens were used in entirety. This was done to achieve comparable tissue-mass representation for all specimens in the DNA extracts. The specimens/tissue samples from each bottle were mixed, frozen in liquid nitrogen and then ground to a fine powder using a disposable mortar and pestle. DNA was extracted with the TIANamp Genomic DNA kit (DP304, TIANGEN Biotech, Beijing, China) following manufacturer's protocols. Briefly, 500 mg of the powder was lyzed in 2 ml Buffer GA overnight at 56 °C in the presence of 20 µl proteinase K. The lysate was centrifuged at 5000 g for 10 min and the supernatant was then aliquoted 40 µl into ten equal volumes for DNA extraction. The quality and purity of the DNA was assessed using a Nanodrop 2000 and 1% agarose gel electrophoresis. Three of the ten DNA extracts from each sample were randomly selected for PCR, creating a total of 129 DNA extracts for analysis.

PCR amplification

PCR employed a first round to amplify the target region of COI while the second added adapters to allow discrimination of the sequences derived from each DNA extract (Prosser et al. 2016). PCR1 employed the following primer pair - AncientLepF3 (TTATAATTGGDGGWTTTGGWAATTG) (Prosser et al. 2016) and LepR1 (TAAACTTCTGGATGTCCAAAAAATCA) (Hebert et al. 2004) and the following thermocycling regime (initial denaturation for 5 min at 95 °C, then 25 cycles of denaturation for 30 s at 95 °C, followed by annealing for 30 s at 55 °C and extension for 60 s at 72 °C, and a final extension for 10 min at 72 °C). For PCR2, adapters were added to primers employed in PCR1. All PCR reactions had a total volume of 25 µl and included 12.5 µl 2×PCR Master Mix, 2 µl each of sequencing primers described above (10 µM), and 1 µl of template purified from the

PCR1 product or water (negative template control). PCR2 employed the following thermocycling regime (Initial denaturation for 45 s at 98 °C, then 6 cycles of denaturation for 15 s at 98 °C, followed by annealing for 30 s at 60 °C and extension for 30 s at 72 °C followed by final extension for 1 min at 72 °C). Amplicons were purified using QIAquick Gel Extraction Kit and KAPA Library Quant (Illumina) DNA Standards & Primer Premix was used for amplicon quantification. Satisfactory amplification was achieved from 126 of the 129 DNA extracts (Table S1).

Next generation sequencing

Purified PCR2 products (126 reactions) were sequenced separately on an Illumina Miseq PE300 platform following standard protocols. Each reaction had two sequencing replicates to generate 252 products for sequencing. The amplicon libraries were prepared using MiSeq Reagent Kit v3 (Illumina, Inc.). Briefly, the pooled library was thawed on ice along with HT1 (hybridization buffer), and then diluted to 2nM in EB buffer. 5 µl of the library were mixed with 5 µl of NaOH at 0.2 N in a microcentrifuge tube with brief vortexing and 1 min of centrifugation at 280 g. After 5 min incubation at room temperature, 990 µl pre-chilled HT1 was added to the tube containing denatured library providing 1 ml of a 10 pM denatured library. The denatured library was diluted to 8 pM (480 µl of the 10 pM denatured library, 120 µl of the pre-chilled HT1) by inverted mixing and then pulse centrifugation. Subsequently, the libraries were loaded onto the reagent cartridge to set up the sequencing run. After cleaning the flow cell, the reagents were loaded into the flow cell to initiate the sequencing.

Sequencing data analysis

There are a range of cloud-based platforms and tools that are freely available to analyze NGS data (Bani Baker et al. 2020). One such tool, mBRAVE, the Multiplex Barcode Research and Visualization Environment, is a data storage and analytics platform with standardized pipelines and a sophisticated web interface designed to transform raw HTS data into biological insights (www.mbrave.net). mBRAVE integrates common analytical methods and links to the BOLD System for access to reference datasets (Young et al. 2021; Zieritz et al. 2022) and assignment of

sequences to BINs, the features that are unique to this platform only. Results from the 252 sequence libraries were uploaded to mBRAVE (www.mbrave.net) under the project "MBR-MTCHN1" where they were analyzed using a standard pipeline (Young et al. 2021; Zieritz et al. 2022) involving sequence trimming (25 bp on each side), quality filtering (minimum QV 25), de-replication, identification, and OTU generation. mBRAVE has direct access to the DNA barcode reference libraries on BOLD which allows comparison of the sequence data with the selected libraries and interpretation of the outcome. This also allows the assignment of the generated OTUs to the Barcode Index Numbers (BINs) and taxonomy on BOLD. The sequences were run against the DNA barcode reference library for Insecta on BOLD that represents 0.5 million BINs and 213,000 named species (the system reference library for mBRAVE ID engine – Insecta). The library also includes about 60,000 insect barcodes from China (DS-CHINAINS).

Sequences on mBRAVE were organized by 'sequence runs' and assigned to BINs and Linnaean taxonomy. The resultant data was downloaded from mBRAVE to summarize the results by Malaise trap location. Pairwise comparisons of BINs among trap locations were visualized by VennDiagram, an R package in the R statistical environment. To reduce the likelihood of false positives, a cleaning step was employed which excluded read counts in the BIN table that represented less than 0.01% of the total read count for their respective sample. BINs represented by a single sequence (singletons) were also excluded from the final BIN count. Concordance or discordance between a BIN and the associated species was determined using "BIN Discordance" tool on BOLD.

Diversity analysis within and among three Malaise trap sites was performed using BINs (as species proxies) recovered from the NGS sequence data. Alpha diversity at individual sites was analyzed by Shannon-Wiener index (Shannon 1948) while the diversity comparisons among the Malaise sites were conducted by Jaccard similarity coefficient (Jaccard 1912). Spread of BINs over time was determined by calculating BIN incidences in the Malaise sampling events for each site.

Results

General patterns of citrus insects

The 252 NGS libraries included 14 Malaise trap collections from GAN, 15 from QIU and 14 from SHI (Table S1). In total, these runs yielded 9.5 million (M) DNA sequences which dereplicated to 2.7M barcodes averaging 250 bp in length. The 1,515,627 sequences remaining after filtration were assigned to BINs. Most sequences belonged to Diptera (64%), Hymenoptera (16.1%), Lepidoptera (15.3%), Hemiptera (2.9%) and Coleoptera (1.3%) (Figure 2). The BIN system (Ratnasingham and Hebert 2013) linked the cumulative NGS sequences from the three Malaise sites to a total of 2,141 BINs, and their counts were similar for the three orchards (GAN = 1,795; QIU = 1,792; SHI = 1,712) (Table 1). These BINs were used as a proxy for species to analyze insect biodiversity assemblages at the collection sites. Alpha diversity analysis showed a similar Shannon index (*H'*) value (GAN = 6.9; QIU = 6.9; SHI = 6.8) for the three Malaise sites (Figure 3). Biodiversity overlap determined by Jaccard similarity coefficient (*J*) suggested a high level of BIN sharing among the three sites in pairwise comparisons (*J* = 0.73-0.75) or in total (*J* = 0.63) (Figure 3). At all three sampling sites, most BINs (GAN=769; QIU=739; SHI=751) were encountered only once in the 14/15 collection events and just a few (GAN=12; QIU=15; SHI=17) were detected in all samples. However, the variance (R^2) of BIN occurrence in the collection events at the three sampling sites remained low (Figure 4).

When the sequences were compared against all insect records on BOLD, 40% (875) of the BINs showed a match to a known species while 59% (1229) were placed to a genus, 97% (2076) to a family, and 100% (2141) to an order. However, a check of the correspondence between the BINs and their associated taxonomy revealed 235 discordances at the species level, 98 at the genus level and 20 at the family level. In GAN, 220 discordant BINs were found at the species, 85 at the genus and 15 at the family. In QIU, 218 discordant BINs were found at the species, 87 at the genus and 18 in the family. In SHI, 216 discordant BINs were found at the species, 89 at the genus and 19 at the family (Table 1).

Screening pest and beneficial insect species

Most (96%) of the 2,141 BINs were linked to 14 insect orders, predominantly Diptera (49%), Hymenoptera (22%), Lepidoptera (14%), Coleoptera (5%), and Hemiptera (4.5%), and of them 875 were linked to a known species. Among the 875 BINs that were identified to a known insect

species, 443 were pests, 223 pest/pollinators, 140 parasitoids, 52 predators, 2 predator/pollinators and 15 were pollinators (Figure 5, Table S2).

Discussion

Challenges of morphology and lack of taxonomic expertise have limited the understanding of insect pest and natural enemy complexes in citrus orchards, compromising the efficacy of pest management tactics (Niu et al. 2014). The present study circumvented these limitations by employing DNA metabarcoding to assess insect diversity in three citrus plantations. The coupling of Malaise traps with metabarcoding has been successfully used to develop inventory for insect faunas in Sweden (Karlsson et al. 2020) and Germany (Moriniere et al. 2016). Malaise traps are useful for capturing flying insects and have been frequently used for barcode-based insect diversity assessments (Hardulak et al. 2020; Karlsson et al. 2020). DNA metabarcoding offers simultaneous identification of multiple species in bulk samples (Yu et al. 2012). Prior studies have used metabarcoding to examine insect biodiversity, pest prevalence (Piper et al. 2022), and to evaluate predator-prey relationships (Yang et al. 2022). For example, (Huang et al. 2022) used metabarcoding to reveal the composition of Diptera communities in a subtropical system, while (Kirse et al. 2021) used this method to analyze seasonal shifts in arthropod diversity in Malaise trap catches. The effective implementation of metabarcoding for simultaneous, multi-species identification of complex mixed communities have helped scaling up pest surveillance efforts (Piper et al. 2019).

BINs were used as a proxy to count species and to link the generated sequences to species represented on BOLD. This approach not only revealed the presence of potentially more than 2,000 insect species, but it also linked 866 (40%) of the BINs to known insects. Interestingly, the number of BINs revealed at each of the three sites were similar, but 25% of the BINs were unique to each site. This result supports the utility of BINs as an effective approach for counting species (Hebert et al. 2016) and assessing insect diversity (Telfer et al. 2015). While most BINs were assigned to single taxon, 12% were linked to more than one species or genus. The lower number of BINs with species matches on BOLD indicates the incompleteness of the current barcode reference library, emphasizing the need to develop regional DNA barcode reference

libraries. BIN discordance has been reported in many barcode studies (De Leon et al. 2020; Gibbs 2018) and the issue has generally been linked to misidentifications (Hebert et al. 2004), heteroplasmy, or incomplete lineage sorting (Kang et al. 2016; Weber et al. 2019). These issues can only be resolved by detailed taxonomic analysis. Nonetheless, use of BINs to identify or count species in pooled samples where sequenced specimens may not be validated by morphology has limitations (Elbrecht et al. 2017). To investigate the spread of BINs over time we calculated BIN incidences in the collection events. This analysis revealed that almost half of the BINs at each site were encountered only once in the total collection events at each site and just a fraction of them were detected in all samples. The points to the seasonality of the activity of various insect species (Wolda 1988).

Insects were classified as either pest or beneficial taxa based on the barcode sequence/BIN association with known species. The number of pest species detected was far higher than documented in the literature on citrus pests in China (Niu et al. 2014; Urbaneja et al. 2020). The present data provides a valuable resource for research in citrus pest management and monitoring. For example, the Asian citrus psyllid, *Diaphorina citri,* a devastating exotic pest of citrus in the US, was detected in samples collected from January to April, 2019. Its detection suggests the need for further monitoring targeted at this species to assist the timely implementation of control measures to reduce the risk of damage to the citrus production.

Numerous studies have been conducted to reveal the trophic relationships between predator and prey by detecting host DNA from the gut contents or feces of the predators using DNA barcoding (Furlong 2015; Galan et al. 2018; McClenaghan et al. 2019; Rytkonen et al. 2019; Verdasca et al. 2022). In our current study, we deployed the Malaise traps in the center of each citrus orchard, that reduced the possibility of adventitious predators. However, possibility of migration of insect predators from the neighboring crops to the sampling sites who ended up in the traps in search of their prey cannot be ruled out. Likewise, it is possible that some insects revealed in our study are not truly citrus pests instead they had been consumed by a predator or had moved to the gardens in search of nectar or to benefit from honeydew secreted by other arthropods. To minimize this deviation at most, we manually verified the identified insect species in the CPIS database and some published references described in materials and methods section, but contamination with non-target DNA originating from predator-prey interactions

(Eitzinger et al. 2019) or airborne eDNA from off-site insects cannot be ruled out (Roger et al. 2022).

Biological control is the most environmentally safe and cost-effective pest management strategy. The present study revealed the presence of known species of Diptera and Hymenoptera which may have potential as biological control agents. For example, BINs linked to 32 species of Braconidae were encountered which belong to three genera (9 *Apanteles* , 14 *Cotesia*, 9 *Microplitis*) that are potential biological control agents for Lepidoptera (Wharton 1993). In addition, syrphid flies (48 species) revealed by BIN-linkages also can be useful natural enemies of aphids (Nelson et al. 2012).

In summary, metabarcoding of bulk insect collections provides a cost- and time-effective way to assess insect communities in citrus orchards. However, the present study only examined one region. Constructing a comprehensive insect inventory for citrus orchards across China will require surveys at the representative sites in other citrus areas.

Conclusions

Our study analyzed the composition of insect communities in citrus orchards using DNA metabarcoding followed by data analysis on BOLD and mBRAVE. The species revealed by BIN matches on BOLD were then searched in citrus pest database and the literature to allow their classification into pest, parasitoid, predator, or pollinator. Our results provide a valuable resource for research on citrus pest management and beneficial insect exploration.

Acknowledgements

We thank Changxiu Xia, Qinqin Wei and Yiwen Fang for the help with field sampling.

References

Ashfaq M, Akhtar S, Rafi MA, Mansoor S, and Hebert PDN. 2017. Mapping global biodiversity connections with DNA barcodes: Lepidoptera of Pakistan. *PLoS One* 12:e0174749. 10.1371/journal.pone.0174749

Ashfaq M, Sabir JSM, El-Ansary HO, Perez K, Levesque-Beaudin V, Khan AM, Rasool A, Gallant C, Addesi J, and Hebert PDN. 2018. Insect diversity in the Saharo-Arabian region: Revealing a little-studied fauna by DNA barcoding. *PLoS One* 13:e0199965. 10.1371/journal.pone.0199965

Bani Baker Q, Hammad M, Al-Rashdan W, Jararweh Y, Al-Smadi M, and Al-Zinati M. 2020. Comprehensive comparison of cloud-based NGS data analysis and alignment tools. *Informatics in Medicine Unlocked* 18. 10.1016/j.imu.2020.100296

Beng KC, Tomlinson KW, Shen XH, Surget-Groba Y, Hughes AC, Corlett RT, and Slik JW. 2016. The utility of DNA metabarcoding for studying the response of arthropod diversity and composition to land-use change in the tropics. *Sci Rep* 6:24965. 10.1038/srep24965

Braukmann TWA, Ivanova NV, Prosser SWJ, Elbrecht V, Steinke D, Ratnasingham S, de Waard JR, Sones JE, Zakharov EV, and Hebert PDN. 2019. Metabarcoding a diverse arthropod mock community. *Mol Ecol Resour* 19:711-727. 10.1111/1755-0998.13008

Caserta R, Teixeira-Silva NS, Granato LM, Dorta SO, Rodrigues CM, Mitre LK, Yochikawa JTH, Fischer ER, Nascimento CA, Souza-Neto RR, Takita MA, Boscariol-Camargo RL, Machado MA, and De Souza AA. 2020. Citrus biotechnology: What has been done to improve disease resistance in such an important crop? *Biotechnology Research and Innovation*. 10.1016/j.biori.2019.12.004

CGA. 2020. Citrus Growers' Association of Southern Africa. Key industry statistics for citrus growers (Available from: www.cga.co.za).

Comtet T, Sandionigi A, Viard F, and Casiraghi M. 2015. DNA (meta)barcoding of biological invasions: a powerful tool to elucidate invasion processes and help managing aliens. *Biological Invasions* 17:905-922. 10.1007/s10530-015-0854-y

Cristescu ME. 2014. From barcoding single individuals to metabarcoding biological communities: towards an integrative approach to the study of global biodiversity. *Trends Ecol Evol* 29:566-571. 10.1016/j.tree.2014.08.001

De Leon LF, Cornejo A, Gavilan RG, and Aguilar C. 2020. Hidden biodiversity in Neotropical streams: DNA barcoding uncovers high endemicity of freshwater macroinvertebrates at small spatial scales. *PLoS One* 15:e0231683. 10.1371/journal.pone.0231683

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, and Hebert PDN. 2019. Expedited assessment of terrestrial arthropod diversity by coupling Malaise traps with DNA barcoding (1). *Genome* 62:85-95. 10.1139/gen-2018-0093

Eitzinger B, Abrego N, Gravel D, Huotari T, Vesterinen EJ, and Roslin T. 2019. Assessing changes in arthropod predator-prey interactions through DNA-based gut content analysis-variable environment, stable diet. *Mol Ecol* 28:266-280. 10.1111/mec.14872

Elbrecht V, and Steinke D. 2018. Scaling upDNAmetabarcoding for freshwater macrozoobenthos monitoring. *Freshwater Biology*. 10.1111/fwb.13220

Elbrecht V, Vamos EE, Meissner K, Aroviita J, Leese F, and Yu D. 2017. Assessing strengths and weaknesses of DNA metabarcoding-based macroinvertebrate identification for routine stream monitoring. *Methods in Ecology and Evolution* 8:1265-1275. 10.1111/2041-210x.12789

Furlong MJ. 2015. Knowing your enemies: Integrating molecular and ecological methods to assess the impact of arthropod predators on crop pests. *Insect Sci* 22:6-19. 10.1111/1744-7917.12157

Galan M, Pons JB, Tournayre O, Pierre E, Leuchtmann M, Pontier D, and Charbonnel N. 2018. Metabarcoding for the parallel identification of several hundred predators and their prey: Application to bat species diet analysis. *Mol Ecol Resour* 18:474-489. 10.1111/1755-0998.12749

Geiger MF, Moriniere J, Hausmann A, Haszprunar G, Wagele W, Hebert PD, and Rulik B. 2016. Testing the Global Malaise Trap Program - How well does the current barcode reference library identify flying insects in Germany? *Biodivers Data J*:e10671. 10.3897/BDJ.4.e10671

Gibbs J. 2018. DNA barcoding a nightmare taxon: assessing barcode index numbers and barcode gaps for sweat bees. *Genome* 61:21-31. 10.1139/gen-2017-0096

359 Gwiazdowski RA, Foottit RG, Maw HE, and Hebert PDN. 2015. The hemiptera (insecta) of Canada:
360 constructing a reference library of DNA barcodes. *PLoS One* 10:e0125635.
361 10.1371/journal.pone.0125635
362 Hajibabaei M, Shokralla S, Zhou X, Singer GA, and Baird DJ. 2011. Environmental barcoding: a next-
363 generation sequencing approach for biomonitoring applications using river benthos. *PLoS One*
364 6:e17497. 10.1371/journal.pone.0017497
365 Hardulak LA, Moriniere J, Hausmann A, Hendrich L, Schmidt S, Doczkal D, Muller J, Hebert PDN, and
366 Haszprunar G. 2020. DNA metabarcoding for biodiversity monitoring in a national park:
367 Screening for invasive and pest species. *Mol Ecol Resour* 20:1542-1557. 10.1111/1755-
368 0998.13212
369 Hebert PDN, Cywinska A, Ball SL, and deWaard JR. 2003. Biological identifications through DNA
370 barcodes. *Proc Biol Sci* 270:313-321. 10.1098/rspb.2002.2218
371 Hebert PDN, Penton EH, Burns JM, Janzen DH, and Hallwachs W. 2004. Ten species in one: DNA
372 barcoding reveals cryptic species in the neotropical skipper butterfly Astraptes fulgerator.
373 *Proceedings of the National Academy of Sciences of the United States of America* 101:14812-
374 14817. 10.1073/pnas.0406166101
375 Hebert PDN, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA, Pedersen S,
376 Jannetta P, and deWaard JR. 2016. Counting animal species with DNA barcodes: Canadian
377 insects. *Philos Trans R Soc Lond B Biol Sci* 371. 10.1098/rstb.2015.0333
378 Huang J, Miao X, Wang Q, Menzel F, Tang P, Yang D, Wu H, and Vogler AP. 2022. Metabarcoding reveals
379 massive species diversity of Diptera in a subtropical ecosystem. *Ecol Evol* 12:e8535.
380 10.1002/ece3.8535
381 Jaccard P. 1912. The Distribution of the Flora in the Alpine Zone.1. *New Phytologist* 11:37-50.
382 10.1111/j.1469-8137.1912.tb05611.x
383 Ji Y, Ashton L, Pedley SM, Edwards DP, Tang Y, Nakamura A, Kitching R, Dolman PM, Woodcock P,
384 Edwards FA, Larsen TH, Hsu WW, Benedick S, Hamer KC, Wilcove DS, Bruce C, Wang X, Levi T,
385 Lott M, Emerson BC, and Yu DW. 2013. Reliable, verifiable and efficient monitoring of
386 biodiversity via metabarcoding. *Ecol Lett* 16:1245-1257. 10.1111/ele.12162
387 Kang AR, Kim MJ, Park IA, Kim KY, and Kim I. 2016. Extent and divergence of heteroplasmy of the DNA
388 barcoding region in Anapodisma miramae (Orthoptera: Acrididae). *Mitochondrial DNA A DNA*
389 *Mapp Seq Anal* 27:3405-3414. 10.3109/19401736.2015.1022730
390 Karlsson D, Forshage M, Holston K, and Ronquist F. 2020. The data of the Swedish Malaise Trap Project,
391 a countrywide inventory of Sweden's insect fauna. *Biodivers Data J* 8:e56286.
392 10.3897/BDJ.8.e56286
393 Kirse A, Bourlat SJ, Langen K, and Fonseca VG. 2021. Metabarcoding Malaise traps and soil eDNA reveals
394 seasonal and local arthropod diversity shifts. *Sci Rep* 11:10498. 10.1038/s41598-021-89950-6
395 McClenaghan B, Nol E, and Kerr KCR. 2019. DNA metabarcoding reveals the broad and flexible diet of a
396 declining aerial insectivore. *The Auk* 136. 10.1093/auk/uky003
397 Moriniere J, Cancian de Araujo B, Lam AW, Hausmann A, Balke M, Schmidt S, Hendrich L, Doczkal D,
398 Fartmann B, Arvidsson S, and Haszprunar G. 2016. Species Identification in Malaise Trap
399 Samples by DNA Barcoding Based on NGS Technologies and a Scoring Matrix. *PLoS One*
400 11:e0155497. 10.1371/journal.pone.0155497
401 Nelson EH, Hogg BN, Mills NJ, and Daane KM. 2012. Syrphid flies suppress lettuce aphids. *BioControl*
402 57:819-826. 10.1007/s10526-012-9457-z
403 Niu J-Z, Hull-Sanders H, Zhang Y-X, Lin J-Z, Dou W, and Wang J-J. 2014. Biological control of arthropod
404 pests in citrus orchards in China. *Biological Control* 68:15-22. 10.1016/j.biocontrol.2013.06.005

405 Piper AM, Batovska J, Cogan NOI, Weiss J, Cunningham JP, Rodoni BC, and Blacket MJ. 2019. Prospects
406 and challenges of implementing DNA metabarcoding for high-throughput insect surveillance.
407 *Gigascience* 8. 10.1093/gigascience/giz092
408 Piper AM, Cunningham JP, Cogan NOI, and Blacket MJ. 2022. DNA Metabarcoding Enables High-
409 Throughput Detection of Spotted Wing Drosophila (Drosophila suzukii) Within Unsorted Trap
410 Catches. *Frontiers in Ecology and Evolution* 10. 10.3389/fevo.2022.822648
411 Prosser SW, deWaard JR, Miller SE, and Hebert PD. 2016. DNA barcodes from century-old type
412 specimens using next-generation sequencing. *Mol Ecol Resour* 16:487-497. 10.1111/1755-
413 0998.12474
414 Ratnasingham S, and Hebert PDN. 2013. A DNA-based registry for all animal species: the barcode index
415 number (BIN) system. *PLoS One* 8:e66213. 10.1371/journal.pone.0066213
416 Ren J-M, Ashfaq M, Hu X-N, Ma J, Liang F, Hebert PDN, Lin L, Germain JF, and Ahmed MZ. 2017. Barcode
417 index numbers expedite quarantine inspections and aid the interception of nonindigenous
418 mealybugs (Pseudococcidae). *Biological Invasions* 20:449-460. 10.1007/s10530-017-1546-6
419 Roger F, Ghanavi HR, Danielsson N, Wahlberg N, Löndahl J, Pettersson LB, Andersson GKS, Boke Olén N,
420 and Clough Y. 2022. Airborne environmental DNA metabarcoding for the monitoring of
421 terrestrial insects—A proof of concept from the field. *Environmental DNA* 4:790-807.
422 10.1002/edn3.290
423 Rytkonen S, Vesterinen EJ, Westerduin C, Leviakangas T, Vatka E, Mutanen M, Valimaki P, Hukkanen M,
424 Suokas M, and Orell M. 2019. From feces to data: A metabarcoding method for analyzing
425 consumed and available prey in a bird-insect food web. *Ecol Evol* 9:631-639. 10.1002/ece3.4787
426 Shannon C. 1948. A mathematical theory of communication. *The Bell System Technical Journal*:379–423
427 and 623–656.
428 Smaili MC, Boutaleb-Joutei A, and Blenzar A. 2020. Beneficial insect community of Moroccan citrus
429 groves: assessment of their potential to enhance biocontrol services. *Egyptian Journal of*
430 *Biological Pest Control* 30. 10.1186/s41938-020-00241-0
431 Taberlet P, Coissac E, Hajibabaei M, and Rieseberg LH. 2012. Environmental DNA. *Mol Ecol* 21:1789-
432 1793. 10.1111/j.1365-294X.2012.05542.x
433 Telfer AC, Young MR, Quinn J, Perez K, Sobel CN, Sones JE, Levesque-Beaudin V, Derbyshire R,
434 Fernandez-Triana J, Rougerie R, Thevanayagam A, Boskovic A, Borisenko AV, Cadel A, Brown A,
435 Pages A, Castillo AH, Nicolai A, Glenn Mockford BM, Bukowski B, Wilson B, Trojahn B, Lacroix CA,
436 Brimblecombe C, Hay C, Ho C, Steinke C, Warne CP, Garrido Cortes C, Engelking D, Wright D,
437 Lijtmaer DA, Gascoigne D, Hernandez Martich D, Morningstar D, Neumann D, Steinke D, Marco
438 DeBruin DD, Dobias D, Sears E, Richard E, Damstra E, Zakharov EV, Laberge F, Collins GE, Blagoev
439 GA, Grainge G, Ansell G, Meredith G, Hogg I, McKeown J, Topan J, Bracey J, Guenther J, Sills-
440 Gilligan J, Addesi J, Persi J, Layton KK, D'Souza K, Dorji K, Grundy K, Nghidinwa K, Ronnenberg K,
441 Lee KM, Xie L, Lu L, Penev L, Gonzalez M, Rosati ME, Kekkonen M, Kuzmina M, Iskandar M,
442 Mutanen M, Fatahi M, Pentinsaari M, Bauman M, Nikolova N, Ivanova NV, Jones N, Weerasuriya
443 N, Monkhouse N, Lavinia PD, Jannetta P, Hanisch PE, McMullin RT, Ojeda Flores R, Mouttet R,
444 Vender R, Labbee RN, Forsyth R, Lauder R, Dickson R, Kroft R, Miller SE, MacDonald S, Panthi S,
445 Pedersen S, Sobek-Swant S, Naik S, Lipinskaya T, Eagalle T, Decaens T, Kosuth T, Braukmann T,
446 Woodcock T, Roslin T, Zammit T, Campbell V, Dinca V, Peneva V, Hebert PD, and deWaard JR.
447 2015. Biodiversity inventories in high gear: DNA barcoding facilitates a rapid biotic survey of a
448 temperate nature reserve. *Biodivers Data J*:e6313. 10.3897/BDJ.3.e6313
449 Urbaneja A, Grout TG, Gravena S, Wu F, Cen Y, and Stansly PA. 2020. Chapter 16 - Citrus pests in a global
450 world. In: Talon M, Caruso M, and Gmitter FG, eds. *The Genus Citrus*: Woodhead Publishing,
451 333-348.

Verdasca MJ, Godinho R, Rocha RG, Portocarrero M, Carvalheiro LG, Rebelo R, and Rebelo H. 2022. A metabarcoding tool to detect predation of the honeybee Apis mellifera and other wild insects by the invasive Vespa velutina. *Journal of Pest Science* 95:997-1007. 10.1007/s10340-021-01401-3

Weber AA, Stohr S, and Chenuil A. 2019. Species delimitation in the presence of strong incomplete lineage sorting and hybridization: Lessons from Ophioderma (Ophiuroidea: Echinodermata). *Mol Phylogenet Evol* 131:138-148. 10.1016/j.ympev.2018.11.014

Wharton RA. 1993. Bionomics of the Braconidae. *Annual Review of Entomology* 38:121-143. 10.1146/annurev.en.38.010193.001005

Wolda H. 1988. Insect Seasonality: Why? *Annual Review of Ecology and Systematics* 19:1-18. 10.1146/annurev.es.19.110188.000245

Yang T, Song X, Zhong Y, Wang B, and Zhou C. 2022. Field investigation- and dietary metabarcoding-based screening of arthropods that prey on primary tea pests. *Ecol Evol* 12:e9060. 10.1002/ece3.9060

Young RG, Milian-Garcia Y, Yu J, Bullas-Appleton E, and Hanner RH. 2021. Biosurveillance for invasive insect pest species using an environmental DNA metabarcoding approach and a high salt trap collection fluid. *Ecol Evol* 11:1558-1569. 10.1002/ece3.7113

Yu DW, Ji Y, Emerson BC, Wang X, Ye C, Yang C, and Ding Z. 2012. Biodiversity soup: metabarcoding of arthropods for rapid biodiversity assessment and biomonitoring. *Methods in Ecology and Evolution* 3:613-623. 10.1111/j.2041-210X.2012.00198.x

Zieritz A, Lee PS, Eng WWH, Lim SY, Sing KW, Chan WN, Loo JS, Mahadzir FN, Ng TH, Yeo DCJ, Gan LX, Gan JY, Gibbins C, Zoqratt MZHM, and Wilson JJ. 2022. DNA metabarcoding unravels unknown diversity and distribution patterns of tropical freshwater invertebrates. *Freshwater Biology* 67:1411-1427. 10.1111/fwb.13926

Table 1(on next page)

Insect BINs recovered from three Malaise traps in citrus orchards and their assignment and discordance at different taxonomic ranks.

1

2 **Table 1: Insect BINs recovered from three Malaise traps in citrus orchards and their**
3 **assignment and discordance at different taxonomic ranks.**

	Location			
	GAN	QIU	SHI	All locations
Total BINs	1,795	1,782	1,712	2,141
BINs assigned to order	1,784	1,776	1,707	2,128
BINs assigned to family	1,744	1,739	1,664	2,076
BINs assigned to genus	1,053	1,034	1,019	1,229
BINs assigned to species	766	755	733	875
BINs with family-level discordance	15	18	19	20
BINs with genus-level discordance	85	87	89	98
BINs with species-level discordance	220	218	216	235

4

5

Figure 1

Three sites in Ganzhou (Jiangxi province, China) where Malaise traps were deployed.

The map was created with GPSVISUALIZER (https://www.gpsvisualizer.com).

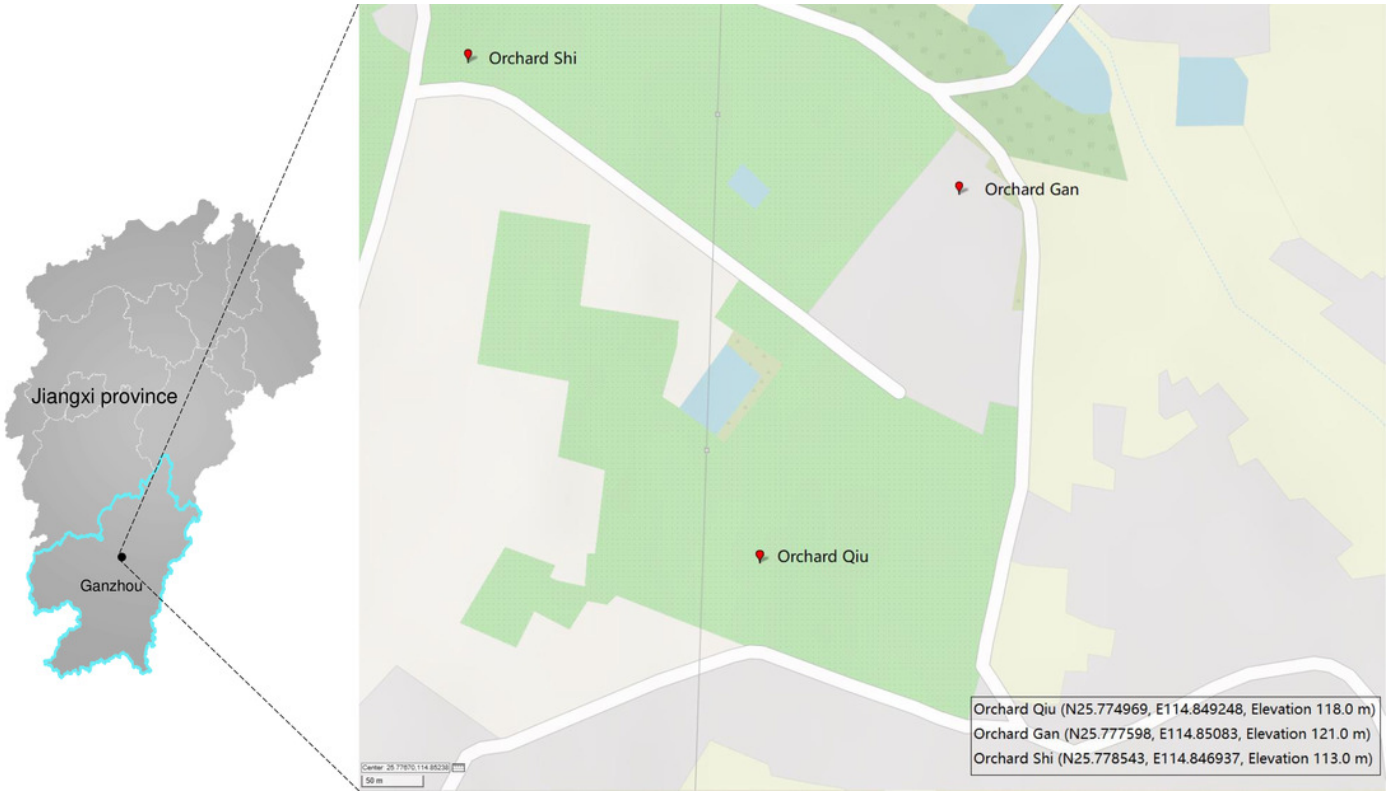


Figure 2

Number of sequences assigned to 14 insect orders based on Malaise trap collections from three Chinese citrus orchards

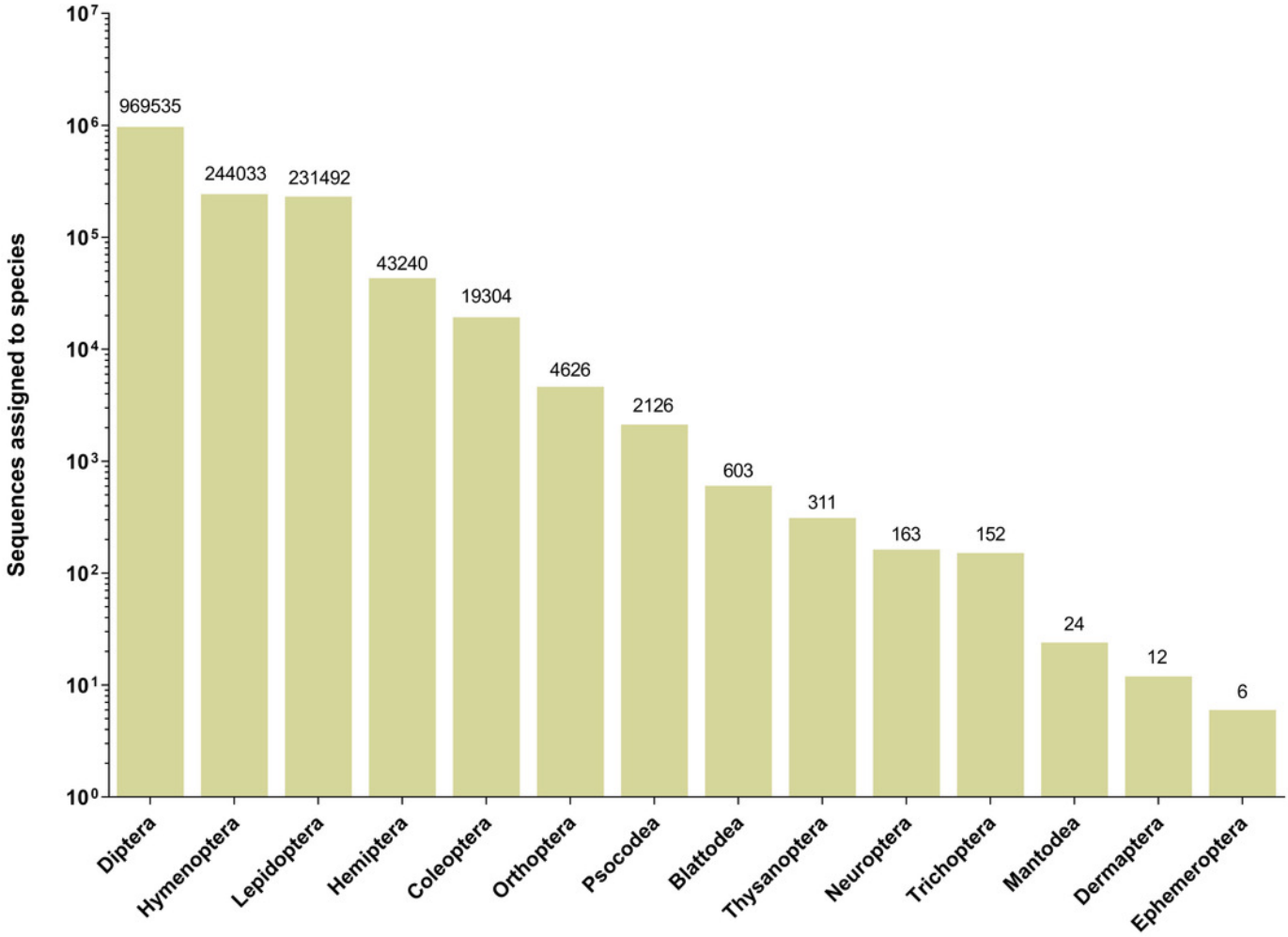


Figure 3

Venn diagram depicting the alpha diversity (Shannon index, *H′*) and the overlap (Jaccard index, *J*) in BIN composition among insect collections from three citrus orchards.

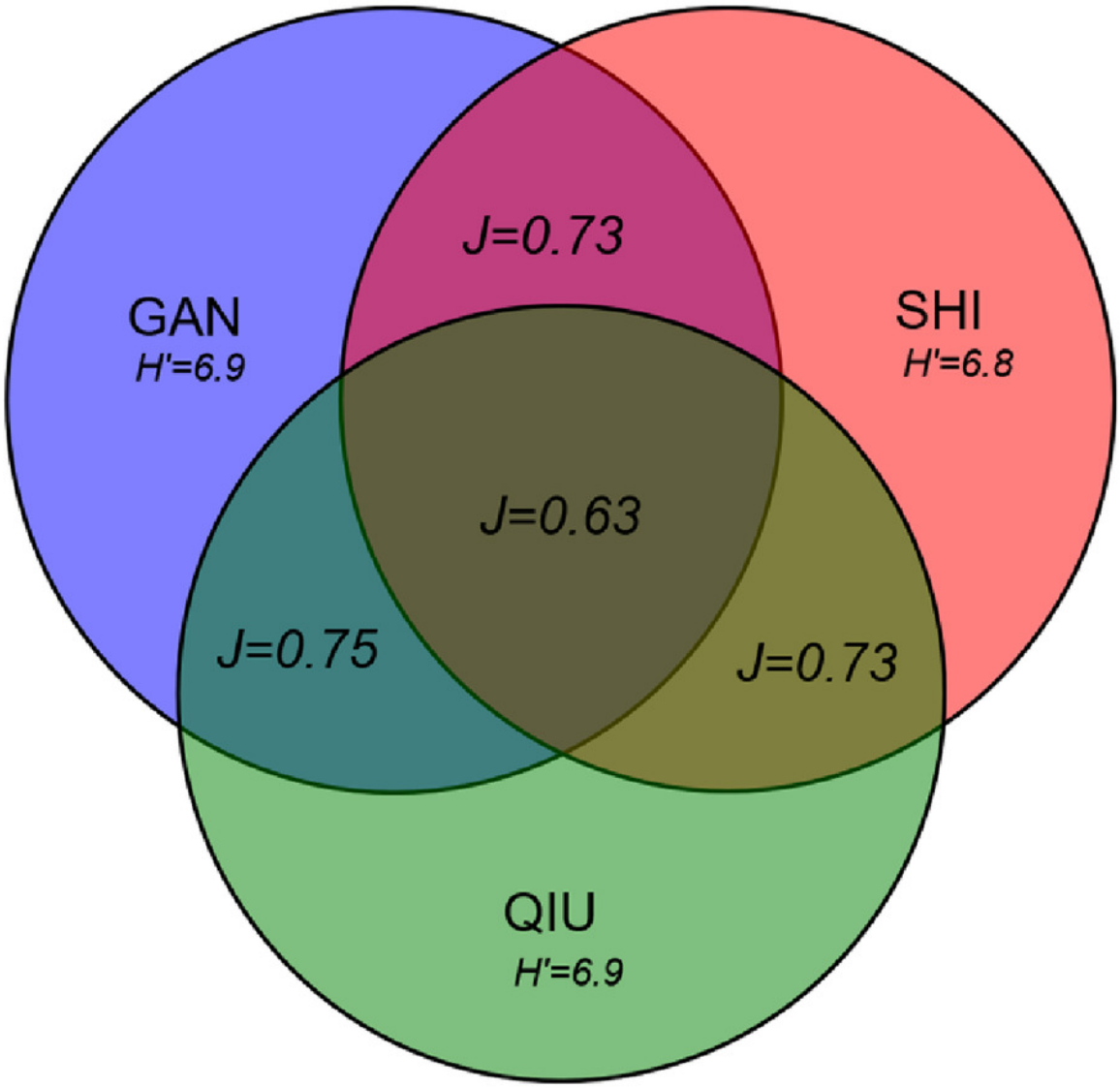
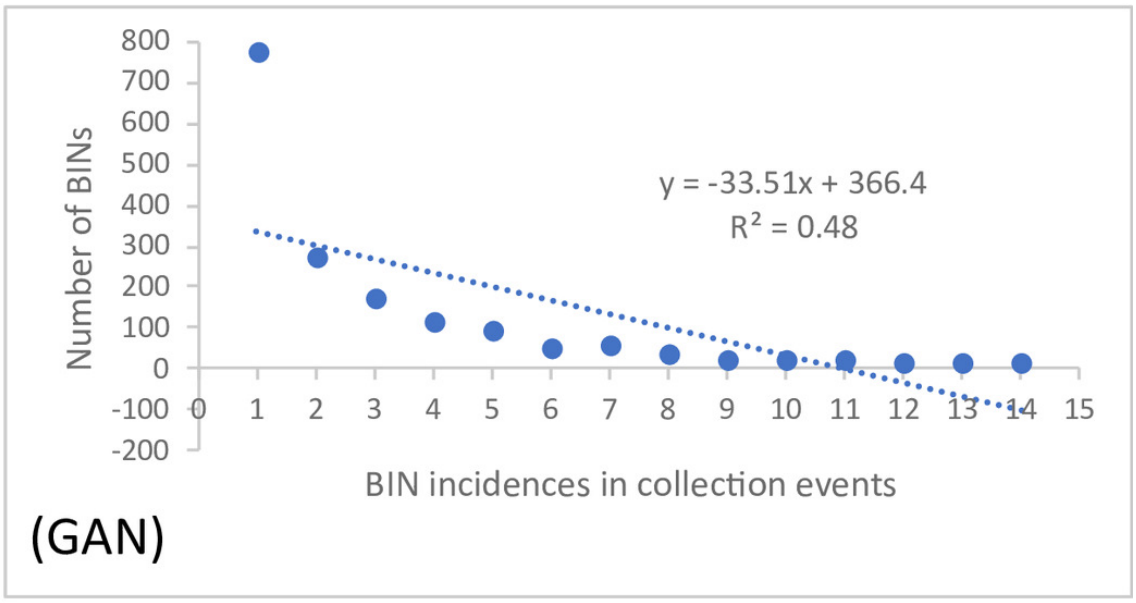
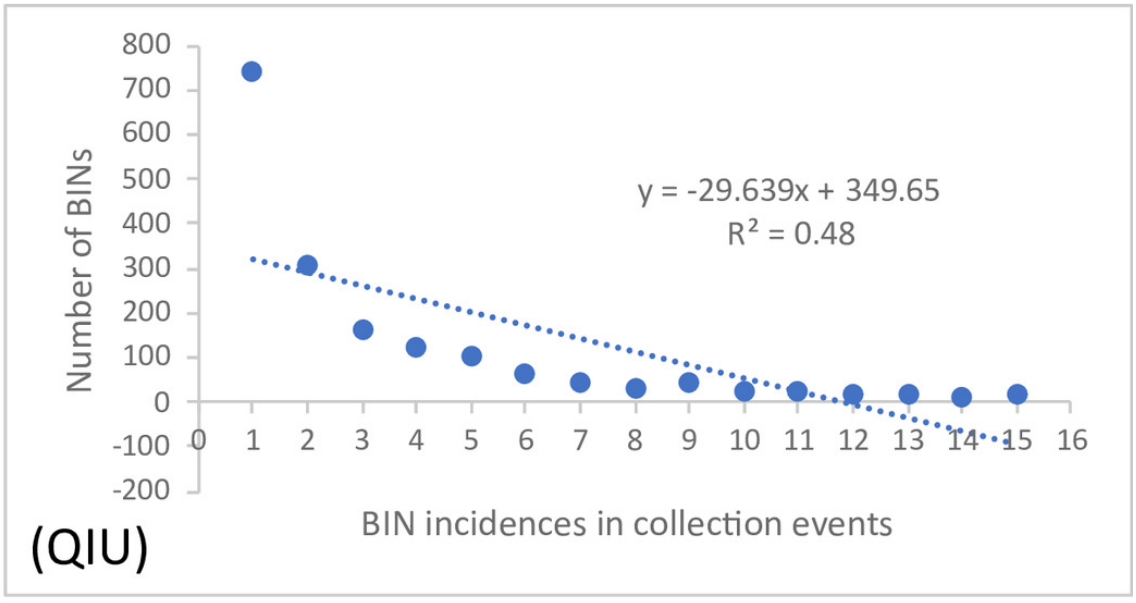


Figure 4

Incidence of BINs in the Malaise collections at three sampling cites, GAN, QIU and SHI.


Number of BINs
y = -33.51x + 366.4
R² = 0.48
BIN incidences in collection events
(GAN)


Number of BINs
y = -29.639x + 349.65
R² = 0.48
BIN incidences in collection events
(QIU)

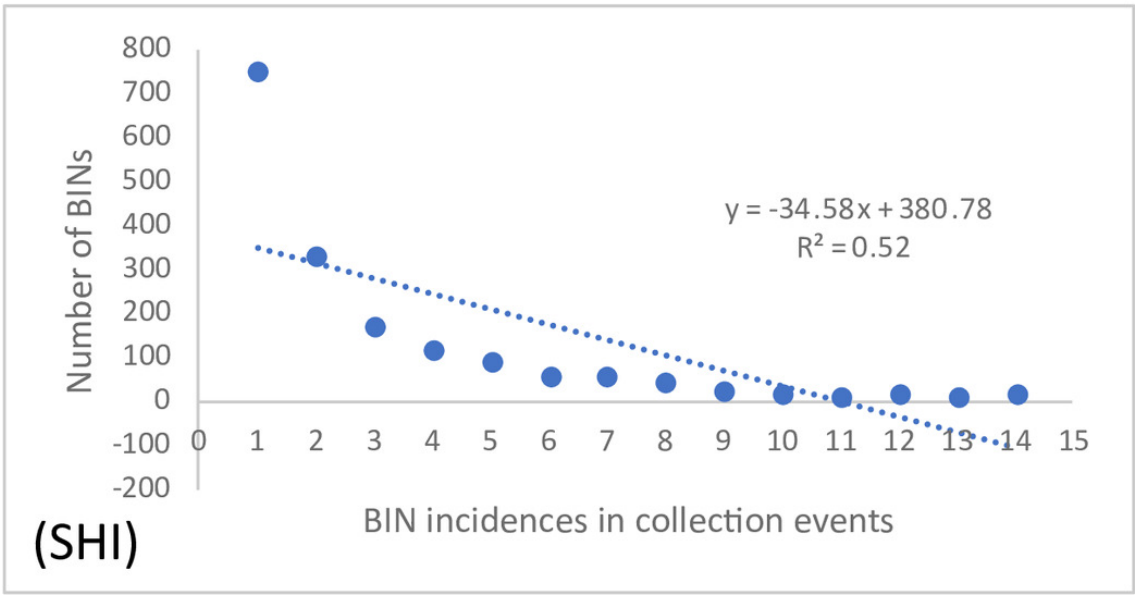

Number of BINs
y = -34.58x + 380.78
R² = 0.52
BIN incidences in collection events
(SHI)

Figure 5

Number of BINs associated with pest and beneficial insect species.