Mining biodiversity databases establishes a global baseline of cosmopolitan Insecta mOTUs: a case study on Platygastroidea (Hymenoptera) with consequences for biological control programs

Matthew R. Moore, Elijah Talamas, Jonathan S Bremer, Natalie McGathey, James C Fulton, Zachary Lahey, Jessica Awad, Cheryl G Roberts, Lynn A Combee

Mining biodiversity databases establishes a global baseline of cosmopolitan Insecta mOTUs: a case study on Platygastroidea (Hymenoptera) with consequences for biological control programs

Authors: Matthew R. Moore[1]*, Elijah J. Talamas[1]*, Jonathan S. Bremer[1], Natalie McGathey[1], James C. Fulton[1], Zach Lahey[2], Jessica Awad[3], Cheryl G. Roberts[1], Lynn A. Combee[1]

*Co-first authors

1 Florida Department of Agriculture and Consumer Services, Division of Plant Industry, Gainesville, Florida, USA.

2 United States Department of Agriculture, Agricultural Research Service, US Vegetable Laboratory, Charleston, South Carolina, USA.

3 State Museum of Natural History Stuttgart, Stuttgart, Germany.

Matthew R. Moore   https://orcid.org/0000-0002-6313-3690.
Elijah J. Talamas   https://orcid.org/0000-0002-1048-6345.
Jonathan S. Bremer   https://orcid.org/0000-0002-4726-6995.
Natalie McGathey   https://orcid.org/0000-0002-8698-3567.
James C. Fulton   https://orcid.org/0000-0003-4093-7781.
Zach Lahey   https://orcid.org/0000-0002-9402-9570.
Jessica Awad   https://orcid.org/0000-0001-6441-4016.
Cheryl G. Roberts   https://orcid.org/0000-0002-3526-5864.
Lynn A. Combee   https://orcid.org/0000-0003-0014-5691.

Abstract

In the past decade, several platygastroid biological control agents were found to be adventive in North America and Europe while under evaluation in quarantine. The scope and relative risk of this phenomenon is not fully known, but it is clearly a trend with implications for classical biological control. As a means of assessing the issue, and providing a global baseline, we implemented a data-mining approach with DNA sequences in the Barcode of Life Database, which yielded 201 platygastroid BINs with intercontinental and island distributions. At least fifty-five BINs displayed exact COI barcode matches across continents, with many more BINs being inconclusive due to sequence length variation. These intercontinental and island BINs include biocontrol agents known to be adventive and many species identified only to genus. We provide 2,500 identifications for platygastroid BOLD BINs, 88% to genus, to encourage additional research on this distributional phenomenon. The intercontinental BOLD BINs were compared to literature records and GBIF occurrences of cosmopolitan species to identify gaps and discordance across data sources. A small COI barcode dataset from localities in Florida and Germany, including topotypical specimens, revealed more intercontinental matches. To assess the scale of intercontinental distributions for host taxa, and to examine the scale for a broader range of taxa, we analyzed COI sequences in BOLD for the entirety of Insecta. The discovery

that adventive parasitoids are following their invasive hosts has important implications for biosecurity and biological control and may lead to an increased emphasis on early detection.

Introduction

Human-mediated movement of insect pests is a well-known phenomenon and mitigation attempts often include classical biological control. Relatively recent invasive stink bugs (Hemiptera: Pentatomoidea) in the United States include the kudzu bug (*Megacopta cribraria* Fab.), the brown marmorated stink bug (*Halyomorpha halys* (Stål)) and bagrada bug (*Bagrada hilaris* Burmeister); management efforts for all of them included classical biological control using egg parasitoids in the family Scelionidae (Hymenoptera: Platygastroidea). In each case, the biological control agent in quarantine was found to be adventive prior to approved release, circumventing regulatory processes and establishing expanded geographic ranges without oversight (Gardner et al. 2013, Hogg et al. 2021, Talamas et al. 2021, Talamas et al. 2015, Milnes et al. 2016, Servick 2018, Ganjisaffar et al. 2018). In the case of *Trissolcus japonicus* attacking *H. halys*, this resulted in an immediate shift from determining if the parasitoids could be released, to tracking their movement and ecological effect (Buffington et al. 2018). This phenomenon, in which parasitoids are found in association with recently adventive pests, has been documented in Mexico, Chile, and throughout Europe (Sabbatini-Peverieri et al. 2018, Lomeli-Flores et al. 2019, Rojas-Gálvez et al. 2021, Stahl et al. 2019, Dieckhoff et al. 2021), and is not limited to biological control agents that are waiting to be deliberately introduced. In the United States, two species that parasitize stink bug eggs, *Psix tunetanus* and *Ps. striaticeps*, are known to be adventive (Johnson and Masner 1985, Birkmire et al. 2021), but were not imported into quarantine as part of a biological control program.

The prevalence of unintended introductions remains largely unevaluated and the numerous detections of scelionid parasitoids of stink bugs is probably a reflection of the attention given to these particular organisms. DeBach and Rosen (1991) estimated that forty-three percent of exotic beneficial arthropods in the United States arrived by accident. Sporadic appearance of biological control agents targeting invasive pests in new environments may be considered "fortuitous" (DeBach 1971). However, by definition, the consequences for native, non-target organisms remain unknown pending post hoc analysis. Tabulating the number of unintentionally introduced platygastroid wasps is even more problematic. An accurate estimate may be impossible to generate at present because taxonomic preparedness (Wheeler and Cracraft 1997) is generally lacking in hyper-diverse economic insect groups. Popovici et al. (2018) documented 'tramp' species in three platygastroid genera, all of which were identified morphologically. This approach requires intimate knowledge of the world fauna and access to a cosmopolitan representation of specimens. Even when unintended introductions are detected relatively quickly (e.g., *Trissolcus japonicus* (Ashmead)), verification of the species identity may require significant research and most cases are probably not extensively investigated. As of 2021, twenty-nine platygastroid wasp species are considered introduced or invasive in the United States (Simpson et al. 2021) and we consider this to be an underestimate based on recent research (Talamas et al. 2021, Birkmire et al. 2021). Only eight introduced platygastroid species are documented in Europe and three of these were deliberate releases (Katsanevakis et al. 2015).

We suspect that unintended introductions of platygastroids will be increasingly detected as 'dark taxa' (Page 2016) in sequencing databases. The aptly named *Gryon aetherium* Talamas (Talamas et al. 2021) perfectly illustrates an instance where the range expansion of a 'dark taxon' was first detected by COI barcode matches, and its classification and taxonomy were established subsequently. *Gryon aetherium* is a parasitoid of the pestiferous Bagrada bug (*B. hilaris*), which is found from South Africa to India. Given the size of this geographic range, hundreds of holotype specimens and descriptions had to be examined in detail before the species could be confidently described as new. The urgency increased as adventive populations were found in both North and South America, and it was only with years of preceding work that the adventive populations in the United States and Chile could be reported with a species name - concurrent with the description of *G. aetherium*. (Lomeli-Flores et al. 2019, Hogg et al. 2021, Rojas-Gálvez et al. 2021, Talamas et al. 2021). The potential for sequence databases to reveal other adventive populations is significant but limited by the level of identification associated with the sequences. The Barcode of Life Database (BOLD; Ratnasingham and Hebert 2007) contains nearly 140,000 scelionid and platygastrid sequence records as of October 2022. Most of these platygastroid BOLD records were bulk sequenced as part of national initiatives and did not receive thorough taxonomic treatments. A total of 128,000 of these BOLD records are unidentified at the species level, whereas 108,000 are unidentified at the genus level. Accordingly, the number of platygastroid dark taxa potentially present in BOLD is staggering, and association with existing names by comparison to type material, made by expert taxonomists, is a gigantic scientific effort.

In this contribution, we follow two research tracks to address the question of how many species may be moving over intercontinental distances while eluding detection. First, we used the existing BOLD infrastructure to identify platygastroid BINs (Barcode Index Numbers; database calculated clusters of highly similar sequences used to approximate species) that are potentially adventive across large geographic spans to serve as a global baseline. We then provided genus level identifications for these BINS whenever possible to encourage additional research. Furthermore, we generated a list of genus level identifications for about one fifth of all platygastroid BOLD BINs to ameliorate taxonomic impediments in the database more broadly. Lastly, we began DNA barcoding of topotype specimens as a decentralized method to provide the most reliable association possible between sequences and species names. Specifically, we provide a dataset from collections made in the Jacksonville, Florida, USA area, a large seaport and a historically important type locality for platygastroids, to establish whether this method was successful and to highlight interesting finds from the region. As the project progressed, we developed a simple method for analyzing BOLD BINs for all of Insecta to assess intercontinental distributions for putative hosts of platygastroids. Perhaps unexpectedly, this method revealed that the intercontinental distribution phenomenon is widespread with implications for assessing biological diversity, taxonomic methodology and regulatory frameworks.

Materials and Methods

BOLD Datamining

Beginning in October 2022, BOLD v4 (Ratnasingham and Hebert 2007) was manually datamined for Platygastroidea BINs with intercontinental distributions. Our search for

Platygastroidea followed the revised classification of Chen et al. (2021), which recognized Geoscelionidae, Janzenellidae, Neuroscelionidae, Nixoniidae, Platygastridae, Scelionidae, and Sparasionidae as the extant families in the superfamily. The BOLD BIN (Barcode Index Number) (Ratnasingham and Hebert 2013) database portal was queried with the following family- and genus level terms: "Geoscelionidae", "Janzenellidae", "Neuroscelionidae", "Nixoniidae", "Platygastridae", "Scelionidae", "Sparasionidae", "*Huddlestonium*", "*Plaumannion*", "*Janzenella*", "*Neuroscelio*", "*Nixonia*", "*Archaeoteleia*", "*Mexon*", "*Listron*", "*Sceliomorpha*" and "*Sparasion*". The resulting BINs were then assessed for the number of countries in which they occurred. BINs with multiple countries were further scrutinized. If the countries were on different continents or otherwise spanned large distances (e.g., Pacific islands and the mainland), then the BIN was included in the dataset. A set of information was recorded and downloaded for each intercontinental BIN: BOLD BIN, BOLD taxonomy lowest level, maximum intra-BIN p-distance, distance to nearest neighbor BIN, the BIN fasta file of COI sequences, and the Darwin Core (Wieczorek et al. 2012) BIN metadata text file (10.5281/zenodo.7930011; 10.5281/zenodo.7930274).

In February 2023, the dataset was expanded using an automated scripting approach (10.5281/zenodo.7930407). The Darwin Core data files for "Insecta" and "Araneae" were downloaded from BOLD. The Insecta files were downloaded on 3 January 2023. The Araneae files were downloaded on 16 March 2023. The Insecta file was examined for all of the categories listed under the field "country". The country fields were assigned to continents or island categories (Sup. File 1). Continent assignments largely followed the United Nations Statistics Division (2023) designations except in a few edge cases where changes were made to more closely follow the borders of zoogeographic regions. Edge cases were assigned to groups based upon the closest alignment of political boundaries and biogeographical barriers. To separate the Asian and European continents, Russia was divided along provincial borders that most closely continue the line of the Ural Mountains southward. Turkey, Georgia, Armenia, and Azerbaijan were included in Europe in order to avoid dividing the Caucasus region. Island nations sufficiently distant from continents were placed into four categories based on oceanic region. The Indonesian provinces of Papua and West Papua were included in the Australian region to avoid dividing the island of New Guinea. Solomon Islands, New Caledonia, and New Zealand were grouped with Australia.

An R (R Core Team 2022) script was developed that created two outputs: 1) a list of BINs with intercontinental or mainland/island distributions and their associated metadata (e.g., taxonomic information) and 2) a list of BINs with intercontinental or mainland/island distributions and an individualized tally of long-distance occurrences. The platygastroid data was compared to our manually extracted list. The taxonomic spread in the Insecta file was summarized at the family and genus level, with special attention given to known host groups of platygastroids. Summary statistics were completed with R 4.2.2 (R Core Team, 2022), using the dplyr (v1.0.10; Wickham et al., 2022), data.table (v1.14.8; Dowle & Srinivasan, 2023), and stringr (v1.4.1; Wickham, 2022) data manipulation packages. Data was visualized using the package ggplot2 (v3.4.1; Wickham, 2016). The full reproducible code is available in Supplementary Materials.

A randomized 5% of the recovered Insecta and Araneae BINs were validated by manual examination in the BOLD BIN database (Sup. File 2; Sup. File 19). Only the first page of returned search results (which contain order, species, and country level distribution) were

examined in the validation process. BINs with many different morphospecies identifications were scored as having minor taxonomic conflicts. BINs with multiple species or genus identifications were scored as having major taxonomic conflicts, except when it appeared that specific epithet discrepancies were due to varying genus classifications. The list of recovered Insecta and Araneae BINs were summarized at the family and genus level. BINs containing known host groups of Platygastroidea were similarly summarized. The Insecta and platygastroid occurrence datasets were transformed into pairwise matrices capturing their geographic distributions. Matrices (Sup. Files 10, 12, 13, 14, 15, 16) were imported to the online version of Circos (Kryzwinski et al. 2009) for visualization. Default settings were used in the online Circos viewer with the exception of the following: 1) labels segment set to large, 2) data filters intra-segment cells hidden, 3) row and column segments order set to col/row ratio, 4) row and column segments with normalized segment size set to remap segments size to 1000, 5) contribution tracks set to hide and no stroke, 6) ribbon caps completely disabled, 7) ratio layout enabled, 8) image format with no strokes and all tick labels hidden.

DNA Sequence Alignment, Tree Building, and Match Scoring

Each BIN in the intercontinental and island playgastroid dataset was individually analyzed to determine whether they contained exact COI barcode matches across large distances. Each BIN's COI fasta file was aligned using the default settings of MUSCLE (Edgar 2004) as implemented in MEGA7 (Kumar et al. 2016). In MEGA7, these individual BIN alignments were used to build neighbor-joining trees (when there were more than two terminal taxa) and distance matrices. Analysis settings were identical for both approaches: the data were labeled as protein-coding nucleotide sequences with the invertebrate mitochondrial genetic code, p-distance set as the method/model, substitutions included transitions and transversions, rates among sites were uniform, and missing data treatment was set to partial deletion with a 95 percent site coverage cutoff. P-distance calculations involving ≤ 255 terminal taxa were exported as matrices. P-distance calculations involving > 255 terminal taxa were exported as pairwise columns. Neighbor-joining trees were exported as Newick tree files and uploaded into the online annotation portal of Interactive Tree of Life (iTOL) v5 (Letunic and Bork 2021).

Trees were viewed in iTOL to determine individual sequence membership in haplotype clusters. Apparent haplotype matches were then examined in the p-distance calculation files for confirmation. The geographic distribution of exact matches was evaluated by examining the specimen level metadata present in the BIN's Darwin Core text file. Putative exact intercontinental matches were then validated in the underlying DNA alignment. This was necessary due to slight variation in the length of COI barcode sequences. If longer sequences in the alignments displayed polymorphisms toward either the 5' or 3' ends and the putative intercontinental matches lacked these flanking data, then the matches were considered inconclusive. Ambiguous DNA base pairs were ignored for considering exact matches.

Individual BIN alignments were ultimately combined into one fasta file for tree building and visualization. A species of *Periclistus* Förster (Hymenoptera: Proctotrupomorpha: Cynipidae) with appropriate data coverage was selected to root subsequent analyses based on the sister relationship of Cynipoidea to Platygastroidea (Blaimer et al. 2023). Alignment of the combined intercontinental and island COI barcode dataset was performed in the MAFFT online service v7 (Katoh et al. 2019) with the FFT-NS-1 setting. This alignment (Sup. File 3) was used

for neighbor-joining analysis in MEGA11 (Tamura et al. 2021) using the same tree building parameters described above. This tree (Sup. File 4) topology was manipulated for viewing ease and annotated in iTOL v5 (Letunic and Bork 2021) and FigTree v1.4.4 (Rambaut 2018) (Sup. File 5).

Identification and BOLD Digital Morphology

During the above data mining activity, it was noticed that a large proportion of platygastroid BINs were unidentified below the family level. A list of Platygastroidea BINs were pulled from BOLD as candidates for identification using the provided digital morphology framework. Identifications were made, when possible, by comparison to images of primary type specimens provided by Talamas et al. (2017). Most Platygastroidea BINs were sight-identified using the BOLD BIN images to family, sub-family, tribe, or genus by the taxonomic authorities Elijah Talamas, Zachary Lahey, and Jessica Awad.

Taxonomic case study: Intercontinental Platygastrinae Taxa and BINs

The German Barcode of Life III: Dark Taxa project (Hausmann et al. 2020) provided 34 platygastrine specimens matching 14 BINs with intercontinental distributions. These were selected for closer morphological examination and taxonomic analysis. Methods for trapping, non-destructive DNA extraction, and COI sequencing follow Awad et al. (2021). Species partitioning was performed with ASAP (Puillandre et al. 2021) using the Jukes-Cantor model and default web server settings.

GBIF Data Mining and Literature Search

Genus level occurrence data (Tab. 1) were downloaded as Darwin Core Archives from The Global Biodiversity Information Facility (GBIF 2022) for all platygastroid genera identified in the intercontinental BIN dataset. These occurrence data were sorted by species, then country, to determine which species had records from more than one continent (Sup. File 6). Relevant taxonomic revisions were searched to corroborate intercontinental taxa reported in GBIF and BOLD. For a total evidence analysis of Platygastroidea, the GBIF, literature, and BOLD datasets were combined to assess their continental and island connections. Species with records from multiple data types were only counted once but their geographic occurrences were rectified for analysis.

Table 1. GBIF occurrence datasets used for comparison with the intercontinental BINs.

Family	Genus	GBIF Occurrence Dataset Citation
Platygastridae	*Allotropa*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.n3derh
Platygastridae	*Amitus*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.xjcn95
Platygastridae	*Amblyaspis*	GBIF.org (28 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.r8zptf

Platygastridae	*Aphanomerus*	GBIF.org (28 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.gw43u4
Platygastridae	*Euxestonotus*	GBIF.org (28 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.gu7djg
Platygastridae	*Fidiobia*	GBIF.org (14 February 2023) GBIF Occurrence Download https://doi.org/10.15468/dl.w478bm
Platygastridae	*Leptacis*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.efh6t8
Platygastridae	*Metaclisis*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.qjwpf5
Platygastridae	*Platygaster*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.d7n4wt
Platygastridae	*Synopeas*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.ev6apy
Scelionidae	*Anteris*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.sjsua4
Scelionidae	*Aradophagus*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.hbu7rt
Scelionidae	*Baeoneurella*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.dw6d4e
Scelionidae	*Baeus*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.9hxwkx
Scelionidae	*Baryconus*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.qaraf2
Scelionidae	*Calliscelio*	GBIF.org (15 February 2023) GBIF Occurrence Download https://doi.org/10.15468/dl.bsua2e
Scelionidae	*Dicroscelio*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.67g54v
Scelionidae	*Dyscritobaeus*	GBIF.org (20 February 2023) GBIF Occurrence Download https://doi.org/10.15468/dl.5mt6wa
Scelionidae	*Gryon*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.tn8kne
Scelionidae	*Hadronotus*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.9qz232

Scelionidae	*Idris*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.h2qj4j
Scelionidae	*Platyscelio*	GBIF.org (13 December 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.cprkz3
Scelionidae	*Psix*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.cc9rkg
Scelionidae	*Telenomus*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.32xjmc
Scelionidae	*Trimorus*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.ytcg37
Scelionidae	*Trissolcus*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.t4qbu5
Scelionidae	*Xenomerus*	GBIF.org (31 October 2022) GBIF Occurrence Download https://doi.org/10.15468/dl.5vzkr6

<u>Field Work</u>

Topotypes of platygastroid species described from Duval County, Florida, were collected between July 2018 and December 2021 in and around the Timucuan Ecological and Historic Preserve near the mouth of the St. Johns River. These collections were largely from Malaise traps that were placed at three different sites in the area. Collecting heads were provisioned with propylene glycol and wrapped with aluminum foil to prevent UV damage to the specimens. 3D printed yellow cylinder traps were experimentally deployed on the ground and suspended from overhanging branches during one sampling period, and additional specimens were collected in yellow pan traps and by sweep netting around the trap sites. Bulk samples were returned to the laboratory and sorted under a Zeiss Discovery V8 Stereo microscope. All platygastroid specimens were transferred to 95% ethanol and screened for matches to species described from that area.

<u>COI Barcoding of Florida Specimens</u>

DNA was non-destructively extracted from specimens using the Qiagen DNeasy Blood and Tissue Kit. Molecular voucher specimens were recovered and deposited at the Florida State Collection of Arthropods (Florida Department of Agriculture and Consumer Services - Division of Plant Industry; Gainesville, Florida). PCRs were conducted as 25 μL reactions using the KAPA HiFi HotStart Readymix Kit (Roche Diagnostics) per the manufacturer's recommended protocol. Oligonucleotide primers used for PCR and direct sequencing were the universal arthropod COI barcoding sets LCO1490/HCO2198 (Folmer et al. 1994) and LEP-F1/LEP-R1 (Hebert et al. 2004). PCR products were visually verified by gel electrophoresis and positive products were prepared for sequencing with the Qiagen QIAquick PCR purification kit. PCR products were bidirectionally sequenced utilizing BigDye Terminator v3.1 chemistry on the Applied Biosystems SeqStudio platform. Sequence traces were trimmed and assembled into

contigs in Sequencher 5.4.8 and Geneious Prime. New COI barcodes were uploaded to BOLD and GenBank (OQ561913–OQ561961) and assessed for their nearest matches.

<u>Photography</u>

Following non-destructive DNA extraction, voucher specimens were mounted and photographed with a Macropod imaging system consisting of a Canon EOS 6D Mark II camera body, EF 70–200mm lens, and 10× or 20× M Plan APO Mitutoyo objective lenses. Imaging software included Canon EOS Utility 3.14.30.4 and Helicon Focus Pro 7.7.5 for image stacking. Adobe Photoshop 23.2.2 was used for limited post-processing and addition of scale bars. Images were uploaded to BOLD along with other specimen metadata.

Results

<u>BOLD Data Mining</u>

The R script recovered 15,391 Insecta BOLD BINs with intercontinental and island distributions (Sup. File 7). Members of 23 Insecta orders were present in the dataset, with only Mecoptera, Notoptera (Mantophasmatodea and Grylloblattodea), and Zoraptera absent. Due to conflicting taxonomies present in BOLD, the family and genus representation numbers are close estimates of higher-level diversity. Orders with the most family- and genus-group diversity were Coleoptera (66 families; 560 genera), Lepidoptera (77 families; 1,984 genera), Diptera (82 families; 899 genera), Hymenoptera (62 families; 736 genera), and Hemiptera (64 families; 454 genera) (Tab. 2). Lepidoptera and Diptera contained the most BINs in the dataset (Tab. 2). The R script recovered 499 Araneae BOLD BINs with intercontinental and island distributions (Sup. File 18). Members of 42 Araneae families were present, comprising 224 genera (Sup. File 18).

Table 2. Taxonomic summary of Insecta BINs displaying geographic distributions spanning continents and islands.

Order	Families	Genera	BINs
Archaeognatha	1	1	1
Zygentoma	1	3	3
Odonata	8	51	91
Ephemeroptera	7	18	37
Dermaptera	4	7	11

Plecoptera	4	8	10
Orthoptera	6	44	62
Embioptera	2	3	5
Phasmatodea	2	4	4
Mantodea	1	5	7
Blattodea	7	31	60
Psocodea	21	30	107
Thysanoptera	3	29	86
Hemiptera	64	454	924
Hymenoptera	62	736	3,056
Raphidioptera	1	1	1
Neuroptera	5	15	43
Strepsiptera	2	2	2
Coleoptera	66	560	1,121
Trichoptera	17	47	126
Lepidoptera	77	1,984	5,336
Siphonaptera	3	6	12

Diptera	82	899	4,286
Total: 23	**446**	**4,938**	**15,391**

769 BINs from the Insecta dataset were randomly selected for cursory validation of the scripting process (Sup. File 2). Validated BINs were examined in Blattodea (1), Coleoptera (50), Diptera (166), Ephemeroptera (5), Hemiptera (38), Hymenoptera (96), Lepidoptera (391), Neuroptera (1), Orthoptera (6), Plecoptera (1), Psocodea (1), Thysanoptera (6), and Trichoptera (7). No validation BINs were discovered to be geographic false positives upon the initial pass. However, four BINs (BOLD:AAC6546, BOLD:AAP8198, BOLD:AAE7880, and BOLD:AAD4954) were recalculated in BOLD during the intervening time period of data gathering, analysis, and validation. These four BINs had the appearance of being geographic false positives, but they were confirmed to be accurate by comparison to the records present in the analyzed data files. Major taxonomic conflicts were present in Blattodea (1/1; 100%), Coleoptera (12/50; 24%), Diptera (44/166; 26%), Hemiptera (15/38; 39%), Hymenoptera (31/96; 32%), Lepidoptera (97/391; 25%), Neuroptera (1/1; 100%), Orthoptera (3/6; 50%), and Trichoptera (4/7; 57%) (Sup. File 2). There were far fewer minor taxonomic conflicts (Lepidoptera=19, Hymenoptera=3, Hemiptera=1, Diptera=4) (Sup. File 2). Twenty-five BINs from the Araneae dataset were randomly selected for validation of the scripting process (Sup. File 19). Validated BINs were examined in Araneidae (1), Clubionidae (1), Gnaphosidae (1), Linyphiidae (10), Lycosidae (4), Pholcidae (1), Tetragnathidae (1), Theridiidae (4), and Thomisidae (2) (Sup. File 19). No validation BINs were discovered to be geographic false positives upon the initial pass. Major taxonomic conflicts were present in Araneidae (1/1; 100%), Clubionidae (1/1; 100%), Linyphiidae (3/10; 30%), Lycosidae (1/4; 25%), and Thomisidae (1/2; 50%) (Sup. File 19).

Greater than 80% of recovered Insecta BINs were present on two continents or islands (Fig. 1; Sup. File 8), totalling 254 unique combinations of continent/island points (Sup. File 9). Europe–North America (4,207 points), Asia–Europe (2,324 points), North America–South America (2,248 points), Asia–Australia (827 points), and Africa–Asia (665 points) were the five most common geographic combinations captured within Insecta BINs (Sup. File 9). In total, North America (12,612 connections), Europe (11,402 connections), and Asia (8,728 connections) were the most connected geographic categories (Fig. 2; Sup. File 10). Europe–North America (5,563 connections), Europe–Asia (3,549 connections), North America–South America (2,718 connections), North America–Asia (1,520 connections), Africa–Asia (1,458 connections), and Australia–Asia (1,345 connections) were the most common connections in Insecta BINs (Fig. 2; Sup. File 10).
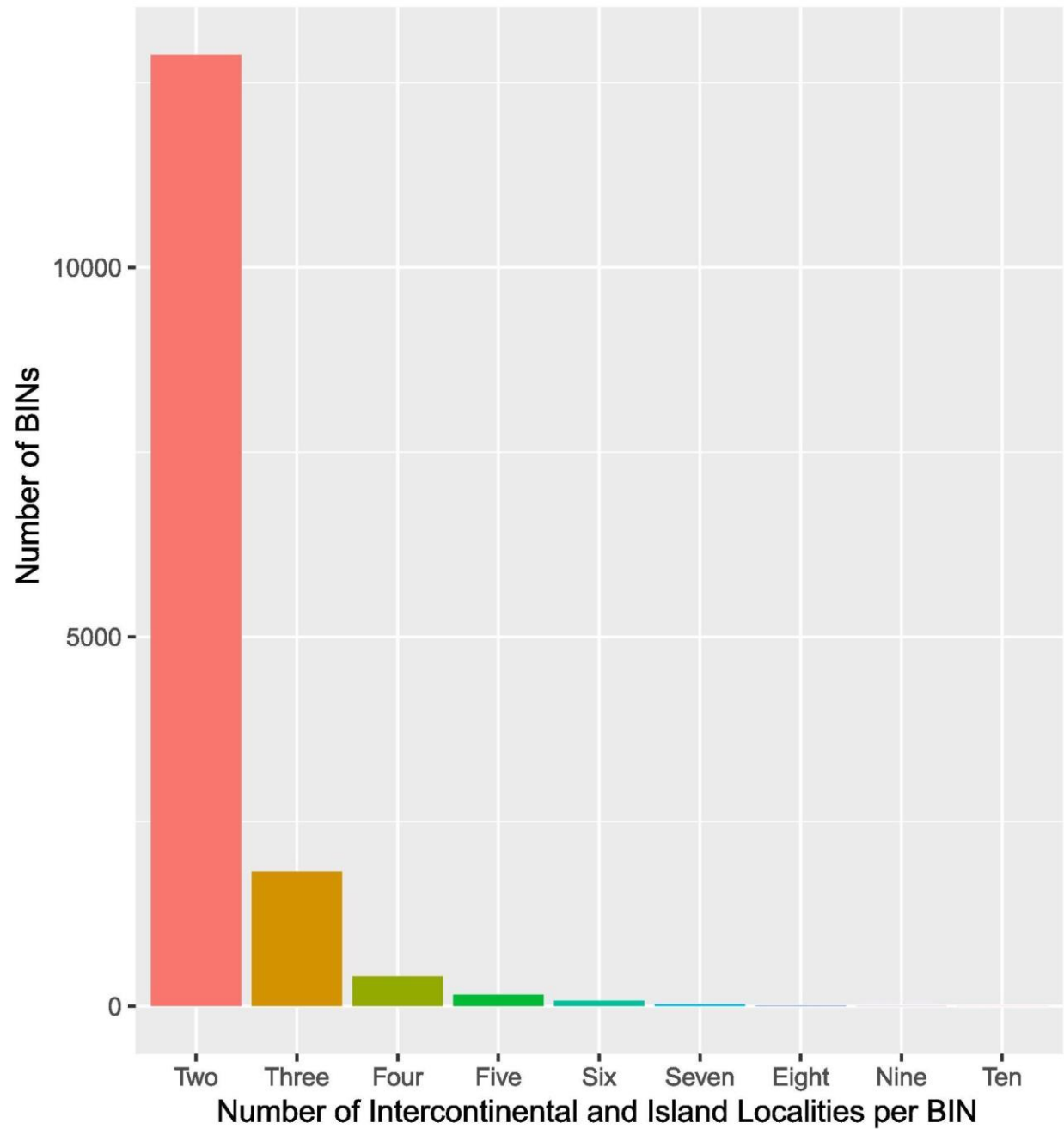
Figure 1. Histogram displaying the percentage of Insecta BINs with a given number of continent and island data points.
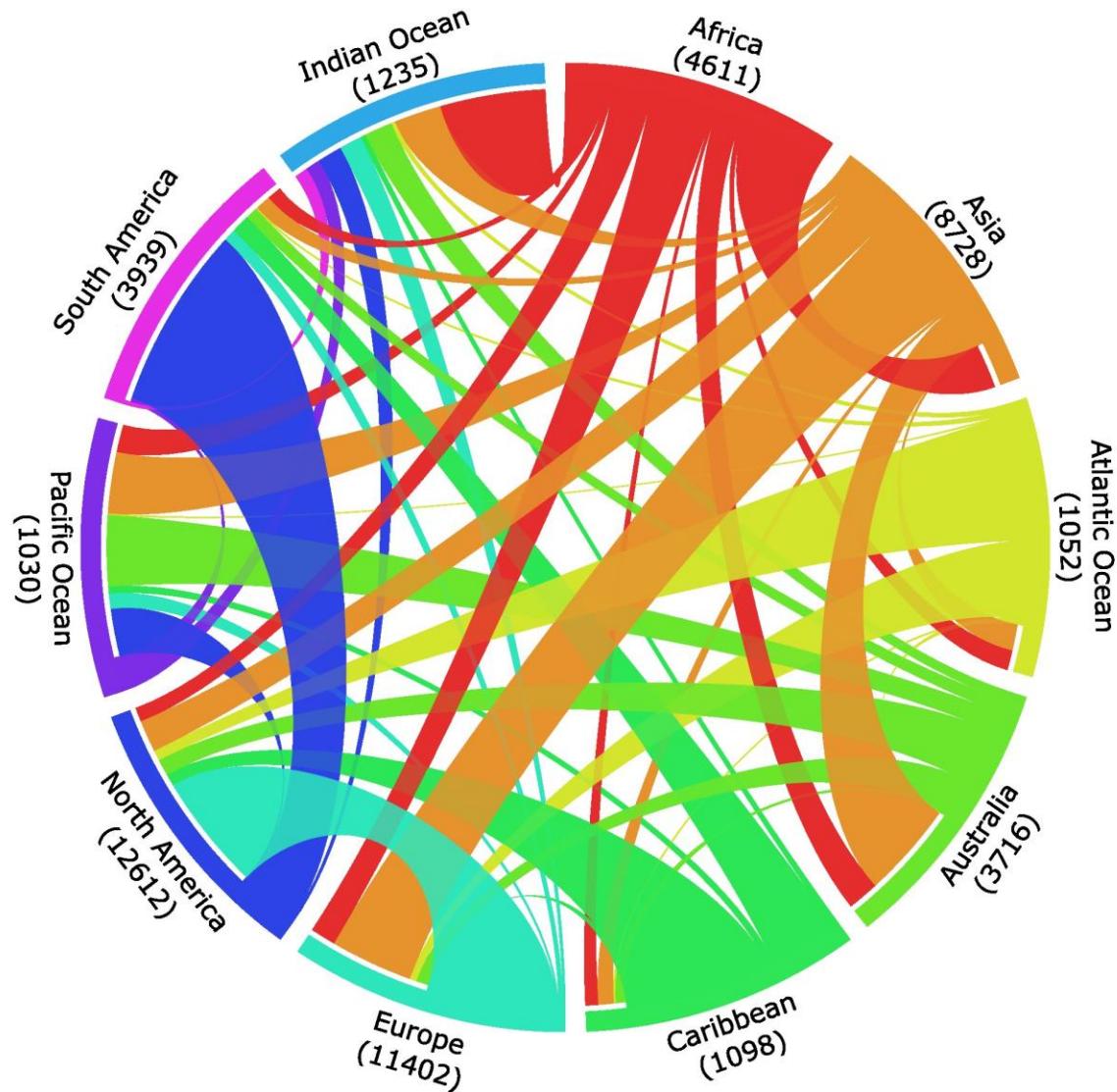
Figure 2. Circos plot displaying the relative contribution of continents and oceanic island groups to geographic connections in the Insecta BOLD BIN dataset. The single data point for Antarctica was eliminated from this visualization. Numbers in parentheses indicate the total number of connections to that geographic grouping.

Two hundred and one platygastrid and scelionid BINs were identified with intercontinental and island distributions (Tab. 3; Sup. File 11). No cases were detected in the other platygastroid families. One hundred-forty of these BINs were identified only to the family or subfamily level in BOLD. We identified 152 of these family level BINs to genus using the specimen images provided by BOLD. The remaining eight BINs were identified to subfamily. 27 genera are represented in the dataset (10 Platygastridae genera; 17 Scelionidae genera) (Tab. 3; Sup. File 11). *Telenomus* (69 BINs) and *Platygaster* (35 BINs) are the most represented genera in their respective families. Canada (101 points) and Germany (90 points) are the most

represented country level occurrence points in the platygastroid BIN dataset (Tab. 3; Sup. File 11), likely biased by the existence of nationally directed barcoding initiatives.

Table 3. Summary of platygastroid and host BINs displaying intercontinental and island distributions in BOLD. Data on host associations were derived from summaries in Masner and Huggert (1989) and AustIn and Field (1997).

Genus (parasitoid)	Intercontinental BINS (parasitoid)	Host taxon	Intercontinental genera (host)	Intercontinental BINs (host)
Platygastridae				
Amblyaspis	5	Cecidomyiidae	43	476
Euxestonotus	2			
Leptacis	7			
Metaclisis	1			
Platygaster	35			
Synopeas	13			
Platygastrinae	4			
Amitus	1	Aleyrodidae	13	35
Aphanomerus	1	Flatidae	2	4
		Ricaniidae	3	4
Allotropa	1	Pseudococcidae	13	34
Fidiobia	2	Chrysomelidae	45	94

		Curculionidae	92	210
Scelionidae				
Anteris	1			
Aradophagus	2	Theridiidae	20	52
Baeoneurella	1	Lygaeidae	7	20
Baeus	4	Araneae		
Baryconus	1	Tettigoniidae	8	11
Calliscelio	1	Gryllidae	6	8
Dicroscelio	2			
Dyscritobaeus	1			
Gryon	6	Heteroptera Lepidoptera		
Hadronotus	5			
Idris	9	Araneae		
Psix	1	Heteroptera		
Scelioninae	2			
Scelionidae	1			

Telenomus	69	Heteroptera Lepidoptera Diptera Neuroptera		
Trissolcus	11	Pentatomidae	21	27
Telenominae	3			
Trimorus	7	Carabidae	48	105
Xenomerus	1			
Teleasinae	1			

The minimum BIN size was two (necessary for a geographic match), and the largest, a *Platygaster* species, contained 677 COI sequences. Most BINs (180 of 201) contained fewer than 100 COI sequences (Fig. 3; Sup. File 11). Intra-BIN maximum p-distances reported by BOLD ranged from 0.00% to 5.42% (Sup. File 11). Most BINs (157 of 201) in the platygastroid dataset had an intra-BIN maximum p-distance of less than 2.50% (Sup. File 11). P-distance to the nearest-neighbor BIN was reported from all of the platygastroid BINs. These p-distances ranged from 0.99% to 17.86% (Sup. File 11). Twenty-eight BINs had larger intra-BIN p-distances than the p-distances to their nearest-neighbor BIN (Sup. File 11). 55 of the platygastroid BINs displayed exact COI matches across distant localities. These exact matches were discovered in *Amblyaspis* (2), *Amitus* (1), *Anteris* (1), *Aphanomerus* (1), *Dicroscelio* (1), *Euxestonotus* (2), *Fidiobia* (1), *Gryon* (2), *Idris* (3), *Leptacis* (2), *Platygaster* (9), *Psix* (1), *Synopeas* (5), *Telenomus* (18), *Trimorus* (1), *Trissolcus* (1), unidentified Platygastrinae (2), unidentified Scelioninae (1), and an unidentified Telenominae (1) (Sup. File 11). Six BINs also included two exact intercontinental haplotype matches: *Euxestonotus* (1), *Gryon* (1), *Platygaster* (1), *Synopeas* (2), and *Telenomus* (1) (Sup. File 11). An additional 28 BINs contained COI sequences that were exact matches which were scored as inconclusive due to sequence length variation, precluding meaningful comparisons (Sup. File 11).
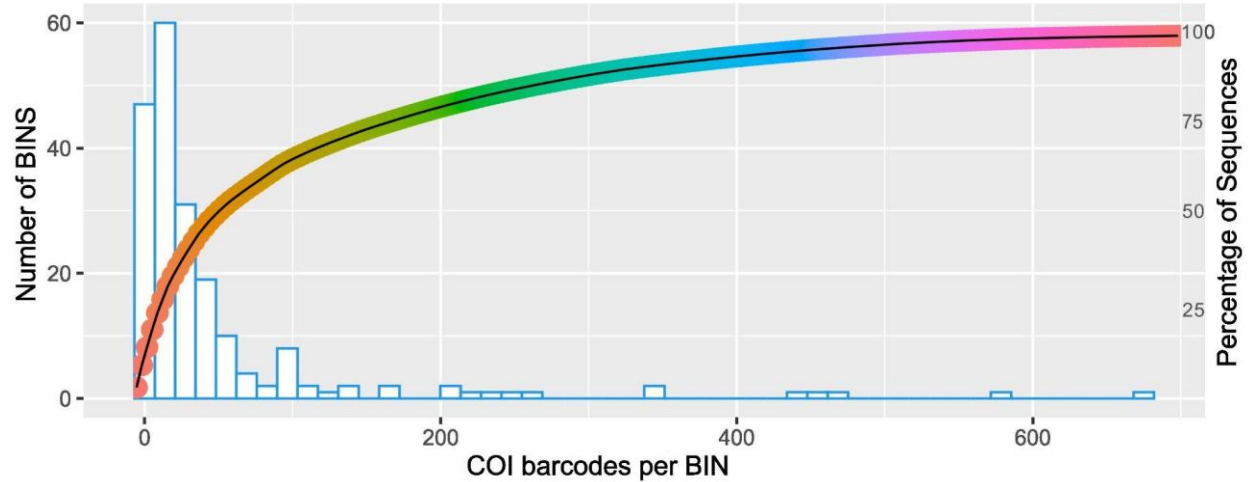
Figure 3. Pareto chart displaying the number of COI barcodes contained per BIN in the intercontinental and island platygastroid dataset.

In the total evidence Platygastroidea dataset, North America (248 connections), Europe (211 connections), and Asia (198 connections) were the most common connections in BINs (Fig. 4; Sup. File 12). *Telenomus* and *Trissolcus* were the most represented scelionid genera across continents and islands. In Platygastridae, *Platygaster* and *Synopeas* were the most represented genera. The webs of geographic connection found in *Telenomus* and *Trissolcus* are more complex than those of *Platygaster* and *Synopeas* (Fig. 5; Sup. Files 13, 14). For example, *Synopeas* and *Platygaster* lack any conclusive geographic data points in the Caribbean and Pacific Ocean and only one each in Africa. Europe and North America are the most connected continents in these genera, except for *Trissolcus* in which Asia and Europe are the most connected (Fig. 5; Sup. Files 15, 16).
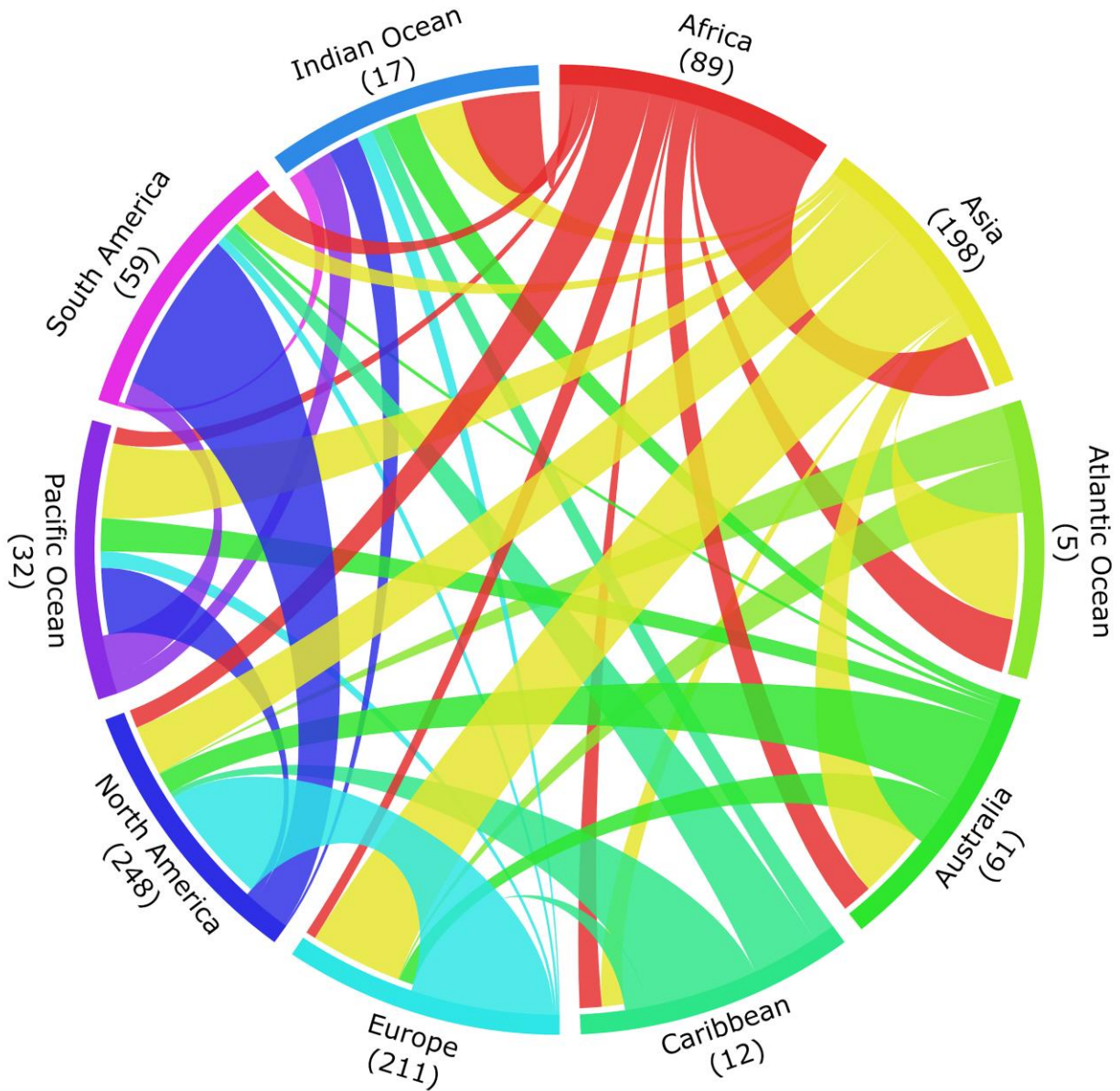
Figure 4. Circos plot displaying the relative contribution of continents and oceanic island groups to geographic connections in the combined platygastroid BOLD BIN, GBIF, and literature dataset. Numbers in parentheses indicate the total number of connections to that geographic grouping.
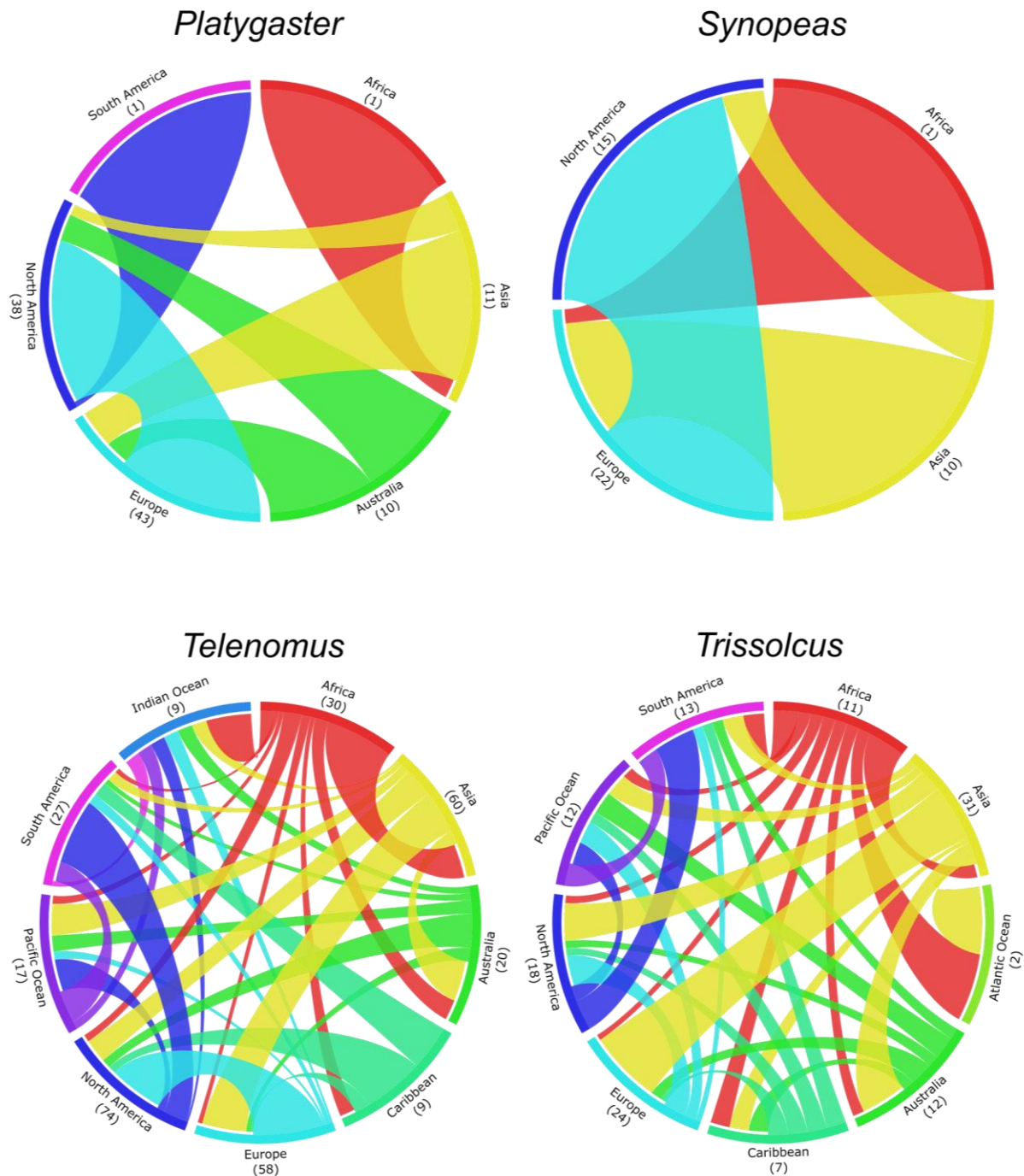
Figure 5. Circos plot displaying the relative contribution of continents and oceanic island groups to geographic connections in the combined *Platygaster*, *Synopeas*, *Telenomus*, and *Trissolcus* BOLD BIN, GBIF, and literature datasets. Numbers in parentheses indicate the total number of connections to that geographic grouping.

DNA Sequence Alignment and Tree Building

The intercontinental and island BIN dataset contains 9,874 platygastroid COI barcode sequences. The MAFFT alignment was 669 base pairs wide after trimming excessive data from the 3' end of some barcodes. The alignment contains some gappy regions due to varying COI amino acid phenotypes present among Platygastroidea (Talamas et al. 2021). All positions with less than 95% site coverage were eliminated in the combined BIN p-distance neighbor-joining analysis; 520 positions were included in the final analysis. Terminal sequence clusters were collapsed and annotated to reflect their BIN assignment, revised or confirmed at the genus level of identification, and the number of sequences in that cluster (Fig. 6; Sup. File 5). These terminal clusters corresponded unambiguously to their BIN assignment (i.e., they formed "monophyletic groups"), except in the case of *Telenomus* BOLD:AAU4881.

Gryon aetherium

Platygaster sp.

Psix striaticeps

Trissolcus basalis

Trissolcus japonicus

Telenomus sp.
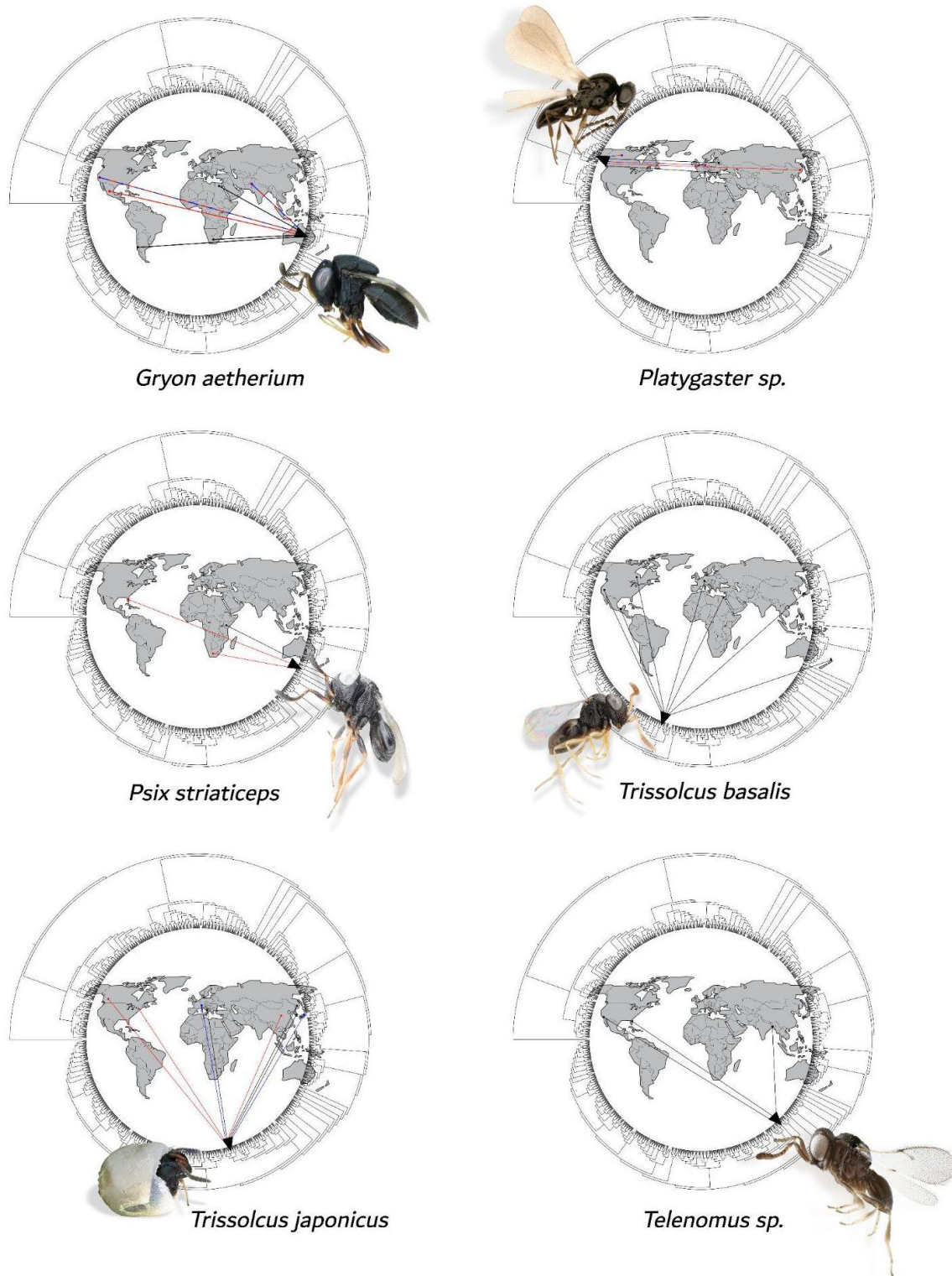
Figure 6. Circularized p-distance neighbor-joining tree of the intercontinental and island BOLD BIN dataset. The BOLD BINs containing *Gryon aetherium*, *Platygaster* sp. (BOLD:ACI8542), *Psix striaticeps, Trissolcus basalis*, *Tr. japonicus*, and *Telenomus* sp. (BOLD:ACY0393) are

highlighted by an enlarged terminal cluster. Lines emanating from these clusters show generalized geographic localities where the BIN was detected. Solid black lines indicate different COI haplotypes. Solid red or blue lines indicate exact COI haplotype matches across continents. Striped red and blue lines indicate multiple exact COI haplotype matches found at that generalized geographic locality.

BOLD Digital Morphology of Platygastroidea

A total of 2,565 Platygastroidea BINs were evaluated for their identification accuracy using the specimen images provided by BOLD (Sup. File 17). Nearly all of these BINs were correctly identified to superfamily (12 were misidentified to order or family). Updated sub-family classification was provided for 2,551 platygastroid BINs (Sup. File 17). A total of 2,209 BINs had images of sufficient quality to add genus level identifications. Sixty-four platygastroid genera were present among these BINs (Sup. File 17), of which eight represent the first DNA sequence data for the genus in GenBank or BOLD: *Euxestonotus*, *Gastrotrypes*, *Isocybus*, *Metaclisis*, *Parabaeus*, *Embioctonus*, *Styloteleia*, and *Xenomerus*.

Taxonomic Case Study: Intercontinental Platygastrinae Taxa and BINs

Comparison of BOLD data to platygastrine specimens in the State Museum of Natural History Stuttgart yielded 14 intercontinental BINs, representing 11 species in five genera (*Amblyaspis, Euxestonotus*, *Leptacis*, *Platygaster*, and *Synopeas*). BOLD identified six species (seven BINs) only to family, two species (four BINs) to subfamily, and three species (three BINs) with binomials. Of the three species identifications provided by BOLD, we verified two (*Pl. demades* and *Pl. sagana*) by comparison of voucher specimens to primary types, while one (*Pl. tuberosula*) was unverifiable. One more species (*E. error*) was unidentified in BOLD but identifiable by our own examination. The remaining seven species were unidentifiable due to the superficial species impediment in platygastrine taxonomy.

The distributions of two species (*Pl. demades* and putative *Pl. tuberosula*) are explained by deliberate introductions for pest control on apple/pear and wheat, respectively. One species (*E. error*) is probably an unintentional introduction, moving with its host, the wheat midge *Sitodiplosis mosellana* (Echegaray et al. 2016). The biogeographic history of the remaining eight species could not be determined. It is possible that one of the unidentified *Platygaster* species (either BOLD:ACP1536 or BOLD:AAZ3286) is *Platygaster hiemalis* Forbes, deliberately introduced to New Zealand for control of the Hessian fly, *Mayetiola destructor* (Say) (Ferguson et al. 2007). However, the type material is unknown and the historical literature provides conflicting diagnoses.

GBIF Data Mining and Literature Search

In the process of analyzing the GBIF data, we encountered a number of errors that were likely to be detected only by those with intimate knowledge of the group: *Trissolcus japonicus* (Ashmead) was listed as *Gryon japonicum* (Ashmead); records for *Trissolcus cultratus* (Mayr) were incomplete; *Calliscelio elegans* (Perkins) was listed as *Caenoteleia elegans* (Perkins); and *Telenomus dalmanni* (Ratzeburg) was misspelled as *Telenomus dalmani*. Notable name changes are that *Trissolcus davatchii* (Javahery) is now treated as a junior synonym of *Tr. elasmuchae* (Watanabe) and *Tr. grandis* (Thomson) is a junior synonym of *Tr. belenus*

(Walker). GBIF and literature searches returned 130 intercontinental taxa (37 Platygastridae; 93 Scelionidae), with an overall discrepancy of plus 67 BINs (Tab. 4). The genera *Amblyaspis*, *Leptacis*, *Baeoneurella*, *Baryconus*, and *Xenomerus* were present only in the BOLD BIN dataset (Tab. 4; Sup. File 11). Conversely, the genera *Inostemma*, *Tetrabaeus*, *Duta*, *Platyscelio*, *Probaryconus*, and *Scelio* were present only in the GBIF or literature datasets (Tab. 4; Sup. File 6).

Table 4. Comparison of total platygastroid intercontinental and island distributions present in different data sources.

Taxon	GBIF/Literature	BOLD BINs	Discrepancy
Platygastridae Total	37	72	+35 DNA
Allotropa	3	1	-2 DNA
Amblyaspis	0	5	+5 DNA
Amitus	4	1	-3 DNA
Aphanomerus	2	1	-1 DNA
Euxestonotus	1	2	+1 DNA
Fidiobia	2	2	-/-
Inostemma	1	0	-1 DNA
Leptacis	0	7	+7 DNA
Metaclisis	1	1	-/-
Platygaster	15	35	+20 DNA
Synopeas	7	13	+6 DNA
Tetrabaeus	1	0	-1 DNA
Scelionidae	93	129	+36 DNA
Anteris	1	1	-/-
Aradophagus	4	2	-2 DNA
Baeoneurella	0	1	+1 DNA
Baeus	2	4	+2 DNA
Baryconus	0	1	+1 DNA

Calliscelio	14	1	-13 DNA
Dicroscelio	1	2	+1 DNA
Duta	1	0	-1 DNA
Dyscritobaeus	3	1	-2 DNA
Gryon	4	6	+2 DNA
Hadronotus	4	5	+1 DNA
Idris	3	9	+6 DNA
Platyscelio	2	0	-2 DNA
Probaryconus	1	0	-1 DNA
Psix	3	1	-2 DNA
Scelio	1	0	-1 DNA
Telenomus	24	69	+45 DNA
Trimorus	1	7	+6 DNA
Trissolcus	24	11	-13 DNA
Xenomerus	0	1	+1 DNA

Specimen Collection, Determinations, and Image Plates

Talamas et al. (2017) provided names and images for species described from Jacksonville with primary types in the National Museum of Natural History. This enabled identifications for taxa that have yet to be revised with modern standards, at least for species with distinctive morphology. For example, *Synopeas cynipsiphilum* (BOLD record SUPER036-23) has a conspicuous divide between the mesoscutum and mesoscutellum that make it easy to recognize. For many others, diagnostic characters were either unknown or too subtle for us to confidently make a determination of species without studying the taxon in detail. In some cases, we did not make a determination because the specimens we sequenced were not the same sex as the primary type. As these taxa are revised, we are certain that more matches will be made between our vouchers and primary types from the region. Of the specimens that yielded COI barcode data, we matched seven specimens to primary types, totalling six species (Tab. 5).

COI Barcoding Platygastroids from the Jacksonville Area

Forty-nine specimens were COI barcoded from Timucuan Ecological and Historic Preserve and Buck Island (Tab. 5). Twenty-eight specimens had BOLD identification hits greater than 97% matches (Tab. 5), representing 25 BINs. Specimen vouchers FSCA 00094179 and FSCA 00094185, both identified as *Telenomus* sp., were 100% matches to a specimen from San Diego, California USA in the BIN BOLD:ACY0393 (Tab. 5). This is an intercontinental

Telenomus BIN identified from the Insecta scripting procedure (Fig. 6), now with geographic data points in California, Florida, and Bangladesh.

Table 5. BOLD identification engine identifications for COI barcoded platygastroid specimens collected at Timucuan Ecological and Historic Preserve and Buck Island (Jacksonville Port Authority).

Process id	Sampleid	Morphological Identification	Topotype	% Match	Match Lowest Taxonomy	Match BIN
SUPER 001-23	FSCA 00091067	*Metaclisis*	No	92.26	Platygastridae	BOLD:AEC 9177
SUPER 002-23	FSCA 00090995	*Hadronotus carinatifrons*	No	99.5	*Hadronotus carinatifrons*	BOLD:AET1 244
SUPER 003-23	FSCA 00095771	*Synopeas*	No	93.94	Platygastridae	BOLD:ADH 8704
SUPER 004-23	FSCA 00094172	*Phanuromyia*	No	89.62	Scelionidae	BOLD:AEN 8490
SUPER 005-23	FSCA 00094176	*Phanuromyia*	No	86.37	Scelionidae	BOLD:ADH 6867
SUPER 006-23	FSCA 00094152	*Phanuromyia*	No	86.9	Scelionidae	BOLD:ADH 6867
SUPER 007-23	FSCA 00094154	*Phanuromyia*	No	98.72	*Telenomus*	BOLD:AAN 8100
SUPER 008-23	FSCA 00094151	*Baryconus floridanus*	Yes	92.4	Scelionidae	BOLD:ABA 5967
SUPER 009-23	FSCA 00094175	*Scelio incertus*	No	99.53	Scelionidae	BOLD:ACA 7140
SUPER 010-23	FSCA 00094167	*Scelio floridanus*	Yes	98.74	Scelionidae	BOLD:ACA 7140

SUPER 011-23	FSCA 00094179	*Telenomus*	No	100	Scelionidae	BOLD:ACY 0393
SUPER 012-23	FSCA 00094185	*Telenomus*	No	100	Scelionidae	BOLD:ACY 0393
SUPER 013-23	FSCA 00094149	*Telenomus*	No	100	Scelionidae	BOLD:ACV 4748
SUPER 014-23	FSCA 00094165	*Telenomus*	No	100	Scelionidae	BOLD:AEO 7335
SUPER 015-23	FSCA 00094156	*Telenomus*	No	100	*Telenomus*	BOLD:AAN 8031
SUPER 016-23	FSCA 00094137	*Telenomus*	No	100	*Telenomus*	BOLD:ABW 3189
SUPER 017-23	FSCA 00094197	*Synopeas*	No	93.38	Platygastridae	BOLD:ACM 5719
SUPER 018-23	FSCA 00095010	*Synopeas*	No	94.16	Platygastridae	BOLD:ADH 9879
SUPER 019-23	FSCA 00095011	*Synopeas*	No	93.62	Platygastridae	BOLD:ADH 9879
SUPER 020-23	FSCA 00095851	*Synopeas*	No	99.84	*Synopeas*	BOLD:AEP 1939
SUPER 021-23	FSCA 00097239	*Synopeas*	No	99.68	Platygastridae	BOLD:AEP 1939
SUPER 022-23	FSCA 00095852	*Platygaster*	No	96.54	Platygastridae	BOLD:AAN 8090
SUPER 023-23	FSCA 00094140	*Telenomus*	No	97.11	Scelionidae	BOLD:ACX 8754

SUPER 024-23	FSCA 00094150	*Telenomus*	No	93.81	Scelionidae	BOLD:ADY 7126
SUPER 025-23	FSCA 00094139	*Telenomus*	No	99.68	*Telenomus*	BOLD:ABY 2759
SUPER 026-23	FSCA 00094173	*Telenomus*	No	93.84	*Telenomus podisi*	BOLD:ADK 2938
SUPER 027-23	FSCA 00094159	*Baeus*	No	97.19	Scelionidae	BOLD:ACZ 5774
SUPER 028-23	FSCA 00094161	*Phanuromyia*	No	91.2	Scelionidae	BOLD:ACJ7 306
SUPER 029-23	FSCA 00094138	*Scelio floridanus*	Yes	98.89	Scelionidae	BOLD:AEY 5850
SUPER 030-23	FSCA 00094249	*Trimorus*	No	94.94	Scelionidae	BOLD:AEJ7 657
SUPER 031-23	FSCA 00094136	*Trimorus*	No	99.52	*Trimorus*	BOLD:ABV 9390
SUPER 032-23	FSCA 00094153	*Telenomus*	No	100	Scelionidae	BOLD:ACI3 554
SUPER 033-23	FSCA 00094146	*Phanuromyia*	No	99.84	*Telenomus autumnalis* (unavailable name)	BOLD:ACM 1917
SUPER 034-23	FSCA 00094155	*Calotelea*	No	100	Scelionidae	BOLD:AAN 8024
SUPER 035-23	FSCA 00093874	*Scelio pumilis*	No	99.81	*Scelio*	BOLD:ACA 7141
SUPER 036-23	FSCA 00097487	*Synopeas cynipsiphilum*	Yes	99.04	*Synopeas*	BOLD:ADX 3415

SUPER 037-23	FSCA 00097245	*Metanopedias brunneipes*	Yes	98.77	Platygastridae	BOLD:ABY 3815
SUPER 038-23	FSCA 00095854	*Synopeas*	No	99.53	Platygastridae	BOLD:AAY 6787
SUPER 039-23	FSCA 00097243	*Platygaster*	No	87.64	Platygastridae	BOLD:ADI9 080
SUPER 040-23	FSCA 00097242	*Inostemma*	No	98.79	Platygastridae	BOLD:ACT 8287
SUPER 041-23	FSCA 00093881	*Scelio opacus*	No	95.27	Scelionidae	BOLD:ACT 6721
SUPER 042-23	FSCA 00093865	*Hadronotus chelinideae*	No	98.87	*Gryon chelinideae*	BOLD:ACN 3082
SUPER 043-23	FSCA 00093888	*Calotelea bicolor*	No	93.78	Scelionidae	BOLD:ADY 8276
SUPER 044-23	FSCA 00093928	*Leptacis puncticeps*	Yes	97.47	Platygastridae	BOLD:AEE 0956
SUPER 045-23	FSCA 00093946	*Leptacis longipes*	Yes	93.63	*Leptacis*	BOLD:ABV 2678
SUPER 046-23	FSCA 00093915	*Telenomus*	No	99.35	*Telenomus* sp. SL017	BOLD:AEI6 588
SUPER 047-23	FSCA 00091193	*Hadronotus bicolor*	No	99.36	*Gryon bicolor*	BOLD:AAN 8046
SUPER 048-23	FSCA 00090993	*Hadronotus bicolor*	No	98.84	*Gryon bicolor*	BOLD:AAN 8046
SUPER 049-23	FSCA 00091003	*Hadronotus bicolor*	No	99.36	*Gryon bicolor*	BOLD:AAN 8046

Discussion

<u>The Approach: Is it Viable?</u>

Our BOLD data mining approach confirmed several well-characterized cases of parasitoid range expansion in *Tr. basalis*, *Tr. japonicus*, *Tr. hyalinipennis*, *Te. remus*, *G. aetherium*, *Ps. striaticeps*, and *Pl. demades*. Therefore, the geographic distribution patterns found in the platygastroid dataset are generally considered credible even when lacking any directionality. The effort of COI barcoding of topotypical specimens can be helpful for determining directionality, particularly if the specimens match primary types described a long time ago. Antiquity of specimens does not preclude the possibility that they were adventive at the time of collection, but it gives, at minimum, a historical perspective.

Exact COI barcode matches from geographically disparate populations might be the most conclusive evidence of a new adventive population within our framework. However, even among the validated cases mentioned above, only *Te. remus*, *G. aetherium*, and *Ps. striaticeps* were scored as exact long-distance matches. *Trissolcus japonicus* and *Tr. basalis* both have the appearance of exact intercontinental matches but were scored inconclusive. *Trissolcus hyalinipennis* and *Pl. demades* have no evidence for exact intercontinental matches in our dataset. Specimens identified as *Pl. demades* are divided into five BOLD BINS, two of which contain populations introduced into Canada and the USA (Mason et al. 2017). These Nearctic *Pl. demades* BINs were not captured by our method. These cases show that the expectation of exact matches may be unnecessary when inferring range expansion, although it does lend additional confidence. The BIN calculations still captured relevant information in these cases even though additional geographic data points were present in neighboring BINs.

A majority of the intercontinental platygastroid BINs were unidentified below the family level at the beginning of this study, affording our group a "clean" taxonomic slate on which to analyze and interpret these results. Alternatively, the identifications for many of the validated intercontinental Insecta BINs were conflicted, which confounds interpretation of the data. Conflicting taxonomies present in BOLD BINs complicated our ability to extract and summarize higher-level taxonomic data from the Insecta dataset. Even if these taxonomic conflicts can be rectified or understood on a case-by-case basis by expert systematists, that particular database problem is likely to persist in future analyses. We advocate that experts use our baseline dataset to closely examine the intercontinental BINs in their group of interest and make informed judgments about their veracity. We closely examined only about 200 BINs, a paltry portion of the total number of intercontinental insects detected.

Many of the BOLD BIN criticisms leveled by Meier et al. (2022) are applicable to our approach. We found that BOLD BIN recalculations affected our study, once again demonstrating their instability. A few of our validation BINs were missing when they were queried back to BOLD for checking, having been recalculated between the time of database mining and data analysis. This feature of BOLD BINs will certainly complicate repetition of our analyses in the future. For better or worse, we must rely on BINs calculated database-wide for our proposed method, as BINs are tied to the specimen metadata in an efficient way for mining. To further verify our relatively small (nearly 10,000 COI barcodes) platygastroid dataset the

sequences could be aligned, phylogenetically analyzed, and those results used to provide *de novo* mOTUs by varying methods with more public documentation than RESL (Ratnasingham and Hebert 2013). However, this would be too biased and would be better accomplished by analyzing all available Platygastroidea data in BOLD, comprising well over 140,000 COI barcodes and growing. Such large analyses would firmly move our method into the realm of bioinformaticians and supercomputing, rather than being repeatable by scientists with less capability for big data. We consider it a major strength that our approach can be performed on an internet-connected, standard desktop computer with freely available statistical software.

Our case study of Platygastrinae using GBOL specimens indicated that BINs overestimated "species richness" and we reiterate that the BINs in the Insecta and Platygastroidea dataset do not necessarily equate to species. Results for poorly known or hyperdiverse insect groups must be interpreted with caution. However, our results were intuitive for insect BOLD BINs with the most widespread geographic data; these included economic species long associated with human activity. For example, *Ctenocephalides felis* (cat flea; BOLD:AAY6332), *Aedes aegypti* (yellow fever mosquito; BOLD:AEI9358), *Culex quinquefasciatus* (southern house mosquito; BOLD:AAA4751), *Aphis gossypii* (cotton aphid; BOLD:AAA3070) were recovered in the analysis with 9 or 10 intercontinental or island data points. Cecidomyiidae (Diptera), focused on as a host group of platygastrines, had 476 intercontinental BINs. Most Cecidomyiidae BINs are unidentified below the family level in a situation analogous to the platygastroid dataset, emphasizing how a lack of data resolution in one taxon can affect the ability to interpret patterns in other ecologically linked groups.

Natural or Adventive Distributions?

Platygastroid wasps are one of the most dominant flying insect groups worldwide, with a high rate of geographically structured community turnover and high taxonomic neglect, further complicating faunistic studies (Srivathsan et al. 2022). Srivathsan et al. (2022) discovered 44 platygastroids mOTUs in multiple sampling sites. They also reported that usually less than 3% (1–9% within a given insect family) of their mOTUs in the top 20 most dominant insect families were found at multiple sites. About five members of our platygastroid BIN dataset display geographic ranges which may be "false positives" based on our intentions. For example, BINs shared between the Republic of Georgia and Iran, or Egypt and Saudi Arabia were scored as intercontinental even though these countries are in close geographic proximity. Conservatively, we detected 195 platygastroid BINs comprising specimens with vast geographic ranges. As of February 2023, BOLD analytics has calculated 11,468 Platygastroidea BINs. From our study, approximately 1.7% of available platygastroid BINs display this striking distribution pattern.

Determining whether these distributions are natural or adventive for most of the platygastroid BINs is difficult pending taxonomic revisions and follow-up research. Gilligan et al. (2020) suggested that a species association with Beringia supports hypothetically natural Holarctic distributions in Tortricidae (Lepidoptera), while also providing a good list of expectations for Holarctic species versus recent introductions. For truly Holarctic organisms, expectations include 1) no direct evidence of introduction, 2) association with native hosts, 3) initial reports from inland areas, 4) lack of recent range expansion, and 5) presence in the Arctic biogeographic region (Gilligan et al. 2020). These expectations are a useful framework for

assessing many of the intercontinental platygastroid BINs present in Canada, Europe, and northeast Asia.

Many Platygastroidea distributions in BOLD dataset were from the southern hemisphere, between the Northern and Southern Tropics, or island localities, precluding the need to consider a naturally holarctic distribution as an explanation for the pattern. We think those cases are best considered introductions most likely mediated by human activity. However, in some *Platygaster* and *Telenomus* BINs the locality data imply enormous geographic ranges across the entirety of Canada, northern and central Europe, northeastern Asia, and other spurious localities. For example, *Telenomus* BIN (BOLD:AAV1142; 461 public barcodes) has occurrences across Canada including Nunavut in the north, south to desert regions of eastern California, and eastern Europe. One unidentified platygastrine (BOLD:ABW3192; 37 public barcodes with several exact matches) has occurrences in eastern and western Canada, Germany, the Russian Far East, South Africa, and California. These geographic ranges encompass several climates and biomes, showing the apparent ability of the wasps to tolerate dramatically different environmental conditions for at least a short period, considering that species that fail to establish would have to persist long enough to be collected.

Historically, the Palearctic and Nearctic regions have been regarded as separate by platygastroid taxonomists. Early European authors rarely made comparisons to the fauna of neighboring countries, let alone distant continents. Likewise, early American hymenopterists often treated the Nearctic fauna as unique. This approach contrasts with that of early American lepidopterists, who tended to misidentify Nearctic species as similar-looking European species (Gilligan et al. 2020). The discrepancy may be explained by the availability of detailed keys and illustrations for Lepidoptera in the 19th century, while descriptions of Platygastroidea were generally short and vague.

Contrary to the assumptions of the past, the results of our study suggest that some platygastroid genera, such as *Platygaster*, include many naturally holarctic species as well as human-mediated introductions. Landry et al. (2013) came to a similar conclusion for Lepidoptera, with the added benefit of host plant data to help distinguish natural from anthropogenic distributions. It is difficult to make such determinations for Platygastrinae, as little is known of their host repertoire, and Cecidomyiidae itself is a dark taxon. However, our results also indicate a high number of intercontinental cecidomyiid BINs, which match well with the parasitoid distributions.

Constraints due to a lack of data are a consistent theme throughout platygastroid taxonomy, especially when compared with better-studied groups of insects. Our method offers a path to gather and interpret the available data, albeit with limitations. For example, the Nearctic and western Palearctic were remarkably well sampled, allowing for more detailed examination of distribution patterns. On the other hand, the Pacific Islands and Caribbean yielded no *Platygaster* or *Synopeas* records. This likely reflects reality in the Pacific Islands, but is a result of undersampling in the Caribbean, a distinction which cannot be made by our method alone. Ashmead (1900) recorded 29 species of platygastrine wasps from St. Vincent and Grenada, while the only platygastrine in Hawaii is considered an accidental introduction (Drake 1969).

<u>Type Material, Vouchers, and Historical Biological Control Releases</u>

Where type material is available for comparison, BIN distributions are helpful in matching DNA barcode data to nomenclature, as it was in the case of *E. error*. *Euxestonotus error* was originally described from New York state and it primarily attacks midges infesting wheat and rye, both of Palearctic origin (Zohary et al. 2012; Haas et al. 2019). The particular case of *E. error* has long caused debate in the literature over whether it is a Holarctic, Nearctic, or Palearctic species (Masner 1989; Doane and Masner 1991). Unfortunately, type material is not always available. Historical type loss is especially common in Europe due to war, neglect, or documentation problems, but it also occurs elsewhere. One *Xenomerus* BIN (BOLD:ACP7487) was identified from British Columbia, Canada, Germany, and Norway. Should this ultimately prove accurate with additional research, it would be the first *Xenomerus* record for the western Hemisphere, and the first country level records for the genus in Germany and Norway (Mikó et al. 2010). In this case, a species identification by comparison with verified type material should be possible due to the recent revision of *Xenomerus* (Mikó et al. 2010).

These case studies highlight the importance of specimen vouchering for biological control. Modern introductions typically have a high standard of documentation, but historically introduced biological control agents often lack molecular and morphological data. Both *Pl. tuberosula* and *Pl. hiemalis* have no known type material and past identification of these species for the purpose of introduction was likely based on host identity. However, numerous species of *Platygaster* are associated with Hessian fly and other gall midges on wheat (Gahan 1933). Additionally, many platygastrine species are associated with multiple hosts and host plants (Cossentine et al. 2020, Vlug 1995), raising the significant possibility of misidentified or multi-species biocontrol introductions from the 19th and early 20th century.

Digital Morphology in BOLD

Specimen images in BOLD allowed us to provide a list of genus level identifications for about one fifth of all platygastroid BOLD BINs. This has just begun the process of overcoming taxonomic impediments in the group and the database more broadly, as many thousands of BINs remain to be examined and identified. Platygastroid wasps are generally small insects (0.5 to 10 mm), making species level characters difficult to assess without proper microscopy and high-resolution images. Regardless, the image quality and habitus views in BOLD were generally sufficient to identify BINs to genus. We encourage systematists to examine BIN images in their group of expertise to see if this process can be repeated for other under-studied insect families. Given that we discovered several BINs which contained genera lacking any other DNA sequence data, is it likely that more such cases remain to be found.

Data presented here suggest that taxonomic revisions should be targeted at the global level as much as possible, even if there are feasibility concerns. A consideration of data from across the world will likely be necessary for the accurate and precise description of some sections of platygastrid and scelionid biodiversity. The BIN database should be preemptively searched to quickly quantify diversity present in an area and inform the taxonomic approach. Our genus level identification of BINS has already facilitated one such study. Melotto et al. (2023) described a new species, *Synopeas maximum* Awad & Talamas, which parasitizes an emerging pest of soybean. Given that soybean is a cosmopolitan crop, it was a possibility that this parasitoid was adventive. A phylogenetic analysis of *Synopeas* BINS enabled them to determine that the closest relatives of *S. maximum* are from North America, suggesting that it

was not adventive. Furthermore, specimen images in BOLD allowed them to assess if certain morphological characters corresponded to a monophyletic species group.

On another research track, the molecular evolution of COI barcodes across Animalia was evaluated by Pentisaari et al. (2016), demonstrating that parasitic groups had many convergent amino acid variations. Several COI barcode amino acid phenotypes have been noted in Platygastroidea and some appear informative at the genus level (Talamas et al. 2021). Increased BOLD resolution for Platygastroidea genera could allow for new analyses of these patterns for the superfamily (Chen et al. 2021).

Regulatory and Biosecurity Applications

Biosecurity practices entail anticipatory risk assessment and preparedness, methods of surveillance, emergency management, and policy enforcement (Barker and Francis 2021). This dataset has potential applications for risk assessment and surveillance. Our approach sought to use large geographic spans in BINs with the goal of identifying putatively adventive platygastroids. Small adjustments to the data mining and analysis methods could be made to focus on administrative borders instead. BOLD data have already been used to inventory fauna at the country level (Hebert et al. 2016; Geiger et al. 2016), making new BIN additions to Canada and Germany straightforward to detect with subsequent assessment.

Emerging infectious disease surveillance programs have implemented several automated, internet-based data mining approaches for intelligence gathering (Stevens 2021). Similarly, automation and scheduled repetition of a modified version of our analysis could provide information supporting risk assessment and invasion preparedness as BINs accumulate in new geographic areas. Detection at the earliest stages of invasion, followed by rapid response, provides the best economic and environmental outcomes. This proposed method of biosecurity surveillance will probably not be timely enough to meet early detection criteria; once new data points are detected in an area the proverbial cat is already out of the bag. Agencies also rely on in-hand specimens for diagnosis of regulated species and are unlikely to take a decentralized BIN record as definitive evidence of new introductions. However, this may be a situation analogous to several cases of verified arthropod introductions that were first noticed or tracked as records in iNaturalist and BugGuide (Halbert et al. 2020, Hayden et al. 2022, iNaturalist 2023, Iwane 2018, USDA 2021, Chuang et al. 2023, VanDyk 2023).

Further scrutinizing the Insecta BINs with consideration for specimen spatiotemporal data might also inform pathway analyses for species moving via global trade, weather phenomena, or animal migration events. Data from the Jacksonville area uncovered an intercontinental *Telenomus* species, with matching specimens collected in 2014 and 2015 from coastal areas of Bangladesh and California. Relevant agencies could have been alerted to the risk of this new introduction, possibly centered around international seaports-of-entry, eight years prior if the available data were scanned systematically. Thus, nearly a decade of potential research progress on this unidentified, globally mobile *Telenomus* went unrealized. Cases like these might be valuable input for pathway models or agent-based models of long-distance insect dispersal. Detailed examination of the Insecta BIN dataset should reveal similar cases for groups with a widely variable range of life history traits, dispersal potentials, and introduction histories or modalities.

Implications for Biological Control Programs

These analyses highlight the urgent need for more detailed and comprehensive approaches (Talamas et al. 2021) when identifying biological control agents, whether classical or augmentative. This is especially the case for platygastroid egg parasitoids; they are small, cryptic, and their biodiversity dwarfs what has been adequately described. They also attack small, cryptic, and poorly described insect life stages. COI barcoding of proposed classical biological control agents must occur immediately at program outset. New data can then be quickly assessed as distinct or not present in BINs calculated from other non-native areas. Putatively "native" communities of parasitoids being evaluated for use as augmentative biological control agents should be similarly treated. Our results and those of Srivathsan et al. (2022) demonstrate the significant likelihood of unintended and undocumented platygastroid introductions.

Acknowledgments

Literature Cited

Ashmead WH (1900) Report upon the aculeate Hymenoptera of the islands of St. Vincent and Grenada, with additions to the parasitic Hymenoptera and a list of the described Hymenoptera of the West Indies. Transactions of the Entomological Society of London 48: 207–367.

Austin AD, Field SA (1997) The ovipositor system of scelionid and platygastrid wasps (Hymenoptera: Platygastroidea): comparative morphology and phylogenetic implications. Invertebrate Taxonomy 11: 1-87.

Awad J, Vasilița C, Wenz S, Alkarrat H, Zimmermann O, Zebitz C, Krogmann L (2021) New records of German Scelionidae (Hymenoptera: Platygastroidea) from the collection of the State

Museum of Natural History Stuttgart. Biodiversity Data Journal 9: e69856. https://doi.org/10.3897/BDJ.9.e69856

Barker K, Francis RA (2021) Spatial tensions and divides in grappling with 'invasive life'. In: Barker K, Frances RA (Eds) Routledge Handbook of Biosecurity and Invasive Species. Routledge. pp. 1–11

Birkmire S, Penca C, Talamas EJ, Moore MR, Hodges AC (2021) *Psix striaticeps* (Dodd) (Hymenoptera, Scelionidae): an Old World parasitoid of stink bug eggs arrives in Florida, USA. In: Lahey Z, Talamas E (Eds) Advances in the Systematics of Platygastroidea III. Journal of Hymenoptera Research 87: 503–521. https://doi.org/10.3897/ jhr.87.76191.

Blaimer BB, Santos BF Cruaud A, Gates MW, Kula RR, Mikó I, Rasplus J-Y, Smith DR, Talamas EJ, Brady SQ, Buffington ML (2023) Key innovations and the diversification of Hymenoptera. Nature Communications 14: 1212.

Buffington ML, Talamas EJ, Hoelmer KA (2018) Team *Trissolcus*: Integrating Taxonomy and Biological Control to Combat the Brown Marmorated Stink Bug. American Entomologist 64: 224–232.

Chen H, Lahey Z, Talamas EJ, Valerio AA, Popovici OA, Musetti L, Klompen H, Polaszek A, Masner L, Austin AD, Johnson NF (2021) An integrated phylogenetic reassessment of the parasitoid superfamily Platygastroidea (Hymenoptera: Proctotrupomorpha) results in a revised familial classification. Systematic Entomology 46: 1088–1113.

Chuang A, Deitsch JF, Nelson DR, Sitvarin MI, Coyle DR (2023) The Jorō spider (*Trichonephila clavata*) in the southeastern US: an opportunity for research and a call for reasonable journalism. Biological Invasions 25: 17-26.

Cossentine JE, Brauner AM, Franklin JL, Robertson MC, Buhl PN, Blatt S, Gariepy T, Fraser H, Appleby M, Grigg-McGuffin K, Mason PG (2020) Parasitism and phenology of *Dasineura mali* (Diptera: Cecidomyiidae) in Canadian apple (Rosaceae) orchards. Canadian Entomologist 152(3): 355–373. https://doi.org/10.4039/tce.2020.15

DeBach P (1971) Fortuitous biological control from ecesis on natural enemies. Pages 293–307 In: Asahinas S, Gressitt JL, Hidaka Z, Nishida T, Nomura K (Eds) Entomological essays to commemorate the retirement of Professor K. Yasumatsu. Hokuryucan Publishing Company; Tokyo.

DeBach P, Rosen D (1991) Biological control by natural enemies. Second edition. Cambridge University Press; New York/Port Chester/Melbourne/Sydney. 440 p.

Doane JF, Masner L (1991) Discovery of the wheat midge parasitoid, *Euxestonotus error* (Fitch) (Hymenoptera: Platygastridae), in British Columbia. Biocontrol News 5: 50-51.

Dowle M, Srinivasan A (2023) Data.table: Extension of `data.frame`. R package version 1.14.8. [accessed 3 May 2023]. https://CRAN.R-project.org/package=data.table.

Drake EF (1969) A new species of Platygasteridae from Hawaii (Hym.: Proctotrupoidea). Proceedings of the Hawaiian Entomological Society 20(2): 323–326.

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research 32: 1792–1797.

Dieckhoff C, Wenz S, Renninger M, Reißig A, Rauleder H, Zebitz CPW, Reetz J, Zimmermann O (2021) Add Germany to the List—Adventive Population of *Trissolcus japonicus* (Ashmead) (Hymenoptera: Scelionidae) Emerges in Germany. *Insects* 12:414. https://doi.org/10.3390/insects12050414

Echegaray ER, Stougaard RN, Bohannon B (2016) First record of *Euxestonotus error* (Hymenoptera: Platygastridae) in the Pacific Northwest, United States of America. Canadian Entomologist 148: 616–618.

Ferguson CM, Moeed A, Barratt BIP, Hill RL, Kean JM (2007) BCANZ - Biological Control Agents introduced to New Zealand. http://www.b3nz.org/bcanz (accessed January 2023).

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R (1994) DNA primers for amplification of mitochondrial Cytochrome C oxidase subunit I from diverse metazoan invertebrates. Molecular Marine Biology and Biotechnology 3: 294–299.

Gahan AB (1933) The serphoid and chalcidoid parasites of the Hessian fly. United States Department of Agriculture Miscellaneous Publications 174: 1–147.

Gardner W, Blount J, Golec J, Jones W, Hu X-P, Talamas E, Evans R, Dong X, Ray C, Buntin G, Gerardo N, Couret J (2013) Discovery of *Paratelenomus saccharalis* (Dodd) (Hymenoptera: Platygastridae), an Egg Parasitoid of Megacopta cribraria F. (Hemiptera: Plataspidae) in its Expanded North American Range. Journal of Entomological Science 48: 355–359.

Ganjisaffar F, Talamas EJ, Bon M-C, Gonzalez L, Brown BV, Perring TM (2018) *Trissolcus hyalinipennis* Rajmohana & Narendran (Hymenoptera, Scelionidae), a parasitoid of *Bagrada hilaris* (Burmeister) (Hemiptera: Pentatomidae), emerges in North America. Journal of Hymenoptera Research 65: 111–130.

GBIF: The Global Biodiversity Information Facility (2022) What is GBIF? Available from https://www.gbif.org/what-is-gbif [31 Oct 2022].

Geiger MF, Astrin JJ, Borsch T, Burkhardt U, Grobe P, Hand R, Hausmann A, Hohberg K, Krogmann L, Lutz M, Monje C, Misof B, Morinière J, Müller K, Pietsch S, Quandt D, Rulik B, Scholar M, Traunspurger W, Haszprunar G, Wägele W (2016) How to tackle the molecular species inventory for an industrialized nation - lessons from the first phase of the German Barcode of Life initiative GBOL (2012-2015). Genome 59: 661-670.

Gilligan TM, Brown JW, Baixeras J (2020) Immigrant Tortricidae: Holarctic versus introduced species in North America. Insects 11 (594): doi:10.3390/insects11090594.

Haas M, Schreiber M, Mascher M (2019) Domestication and crop evolution of wheat and barely: genes, genomics, and future directions. Journal of Integrative Plant Biology 61: 204-225.

Halbert SE, Eger JE, Allen JS (2020) *Brachyplatys subaeneus* (Westwood), black bean bug, detected in South Florida. FDACS Pest Alert. FDACS-P-01936.

Hausmann A, Krogmann L, Peters RS, Rduch V, Schmidt S (2020) GBOL III: Dark Taxa. iBOL Barcode Bulletin 10: 1. https://doi.org/10.21083/ibol.v10i1.6242

Hayden JE, Burnette KM, Moore MR (2022) *Papillio demoleus*, Lime Swallowtail, on Key West, Florida. FDACS Pest Alert. FDACS-P-01665.

Hebert PDN, Penton EH, Burns JM, Janzen DH, Hallwachs W (2004) Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astraptes fulgerator*. PNAS 101: 14812–14817.

Hebert PDN, Ratnasingham S, Zakharov EV, Telfer AC, Levesque-Beaudin V, Milton MA, Pedersen S, Jannetta P, deWaard JR (2016) Counting animal species with DNA barcodes: Canadian insects. Philosophical Transactions of the Royal Society B: Biological Sciences 371: 20150333.

Hogg BN, Hougardy E, Talamas E (2021) Adventive *Gryon aetherium* Talamas (Hymenoptera, Scelionidae) associated with eggs of *Bagrada hilaris* (Burmeister) (Hemiptera, Pentatomidae) in the USA. In: Lahey Z, Talamas E (Eds) Advances in the Systematics of Platygastroidea III. Journal of Hymenoptera Research 87: 481-492. https://doi.org/10.3897/jhr.87.73778.

iNaturalist (2023) Available from https://www.inaturalist.org. Accessed 23 March 2023.

Iwane T (2018) An invasive moth is recorded in Ontario, Canada for the first time - Observation of the week, 9/9/18. iNaturalist Blog. An Invasive Moth is Recorded in Ontario, Canada for the First Time - Observation of the Week, 9/9/18 · iNaturalist. Accessed 23 March 2023.

Johnson NF, Masner L (1985) Revision of the genus *Psix* Kozlov & Lê (Hymenoptera: Scelionidae). Systematic Entomology 10: 33–58.

Katsanevakis S, Deriu I, D'Amico F, Nunes AL, Pelaez Sanchez S, Crocetta F, Arianoutsou M, Bazos I, Chistopoulou A, Curto G, Delipetrou P, Kokkoris Y, Panov VE, Rabitsch W, Roques A, Scalera R, Shirley SM, Tricarico E, Vannini A, Zenetos A, Zervou S, Zikos A, Cardoso AC (2015) European Alien Species Information Network (EASIN): supporting European policies and scientific research. Management of Biological Invasions 6: 147–157.

Katoh K, Rozewicki J, Yamada KD (2019) MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Briefings in Bioinformatics 20: 1160–1166. https://doi.org/10.1093/bib/bbx108.

Kryzwinski MI, Schein JE, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Genome Research Published in Advance June 18, 2009: doi:10.1101/gr.092759.109

Kumar S, Stecher G, Tamura K (2016) MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets. Molecular Biology and Evolution 33: 1870–1874.

Landry J-F, Nazari V, DeWaard JR, Mutanen M, Lopez-Vaamonde C, Huemer P, Hebert PDN (2013) Shared but overlooked: 30 species of Holarctic Microlepidoptera revealed by DNA barcodes and morphology. Zootaxa 3749: 1–93.

Letunic I, Bork P (2021) Interactive Tree of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Research 49: W293–W296. https://doi.org/10.1093/nar/gkab301.

Lomeli-Flores JR, Rodríguez-Rodríguez SE, Rodríguez-Levya E, González-Hernández H, Gariepy TD, Talamas EJ (2019) Field studies and molecular forensics identify a new association: *Idris elba* Talamas, sp. nov. parasitizes the eggs of *Bagrada hilaris* (Burmeister). In: Talamas E (Eds) Advances in the Systematics of Platygastroidea II. Journal of Hymenoptera Research 73: 125–141.

Masner L (1989) *Euxestonotus error* (Fitch) - Palearctic or Neartic. Biocontrol News 2: 23-24.

Masner L, Huggert L (1989) World review and keys to the genera of the subfamily Inostemmatinae with reassignment of the taxa to the Platygastrinae and Sceliotrachelinae (Hymenoptera: Platygastridae). Memoirs of the Entomological Society of Canada 147: 1–214. https://doi.org/10.4039/entm121147fv.

Mason PG, Olfert OO, Haye T, Gariepy TD, Abram PK, Gillespie DR (2017) Risks and benefits of accidental introductions of biological control agents in Canada. In: Mason PG, Gillespie DR, Vincent C (Eds) Proceedings of the 5th International Symposium on Biological Control of Arthropods: 6-8. CABI.

Meier R, Blaimer BB, Buenaventura E, Hartop E, vonRintelen T, Srivathsan A, Yeo D (2022) A reanalysis of the data in Sharkey et al.'s (2021) minimalist revision reveals that BINs do not deserve names, but BOLD Systems needs a stronger commitment to open science. Cladistics 38: 264-275.

Mikó I, Masner L, Deans AR (2010) World revision of *Xenomerus* Walker (Hymenoptera: Platygastroidea, Platygastridae) Zootaxa 2708: 1–73.

Milnes JM, Wiman NG, Talamas EJ, Brunner JF, Hoelmer KA, Buffington ML, Beers EH (2016) Discovery of an exotic egg parasitoid of the Brown Marmorated Stink Bug, *Halyomorpha hays*

(Stål) in the Pacific northwest. Proceedings of the Entomological Society of Washington 118: 466–470.

Page RDM (2016) DNA barcoding and taxonomy: dark taxa and dark texts. Philosophical Transactions of the Royal Society B: Biological Sciences 371: 20150334.

Pentisaari M, Salmela H, Mutanen M, Roslin T (2016) Molecular evolution of a widely-adopted taxonomic marker (COI) across the animal tree of life. Scientific Reports 6: 35275.

Popovici OA, Masner L, Viciriuc M, Pintilioaie A, Notton DG, Talamas E (2018) New distribution data for some charismatic tramp species of Platygastroidea (Hymenoptera). Zootaxa 4370: 1–22.

Puillandre N, Brouillet S, Achaz G (2021) ASAP: assemble species by automatic partitioning. Molecular Ecology Resources 21(2): 609–620. https://doi.org/10.1111/1755-0998.13281

R Core Team (2022) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing. [accessed 3 May 2023]. https://www.R-project.org.

Rambaut A (2018) FigTree v.1.4.4. [accessed 27 October 2022]. http://tree.bio.ed.ac.uk/software/figtree.

Ratnasingham S, Hebert PDN (2007) BOLD: The Barcode of Life Data System (www.barcodinglife.org). Molecular Ecology Notes 7: 355–364.

Ratnasingham S, Hebert PDN (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.

Rojas-Gálvez NR, Talamas E, Albornoz MV, Flores MF, Barros-Parada W, Bout A (2021) *Gryon aetherium* Talamas (Hymenoptera, Scelionidae): Parasitoid of *Bagrada hilaris* (Burmeister) (Hemiptera, Pentatomidae) Adventive in Chile. In: Lahey Z, Talamas E (Eds) Advances in the Systematics of Platygastroidea III. Journal of Hymenoptera Research 87: 493–501. https://doi.org/10.3897/jhr.87.75363.

Sabbatini Peverieri G, Talamas E, Bon MC, Marianelli L, Bernardinelli I, Malossini G, Benvenuto L, Roversi PF, Hoelmer K (2018) Two Asian egg parasitoids of *Halyomorpha halys* (Stål) (Hemiptera, Pentatomidae) emerge in northern Italy: *Trissolcus mitsukurii* (Ashmead) and *Trissolcus japonicus* (Ashmead) (Hymenoptera, Scelionidae). Journal of Hymenoptera Research 67: 37–53.

Servick K (2018) Control freaks. Science 361(6402): 542–545. doi: 10.1126/science.361.6402.542.

Simpson A, Turner R, Blake R, Liebhold A, Dorado M (2021) United States Register of Introduced and Invasive Species: U.S. Geological Survey data release, https://doi.org/10.5066/P95XL09Q.

Srivathsan A, Ang Y, Heraty JM, Hwang WS, Jusoh WFA, Kutty SN, Puniamoorthy J, Yeo D, Roslin T, Meier R (2022) Global convergence of dominance and neglect in flying insect diversity. BioRxiv: https://doi.org/10.1101/2022.08.02.502512.

Stahl J, Tortorici F, Pontini M, Bon M-C, Hoelmer K, Marazzi C, Tavella L, Haye T (2019) First discovery of adventive populations of *Trissolcus japonicus* in Europe. Journal of Pest Science 92: 371–379.

Stevens KB (2021) Novel scientific approaches to understanding emerging infectious diseases. In: Barker K, Francis RA (Eds) Routledge Handbook of Biosecurity and Invasive Species. Routledge. pp. 90-101.

Talamas E, Thompson J, Cutler A, Fitzsimmons Schoenberger S, Cuminale A, Jung T, Johnson N, Valerio A, Smith A, Haltermann V, Alvarez E, Schwantes C, Blewer C, Bodenreider C, Salzberg A, Luo P, Meislin D, Buffington M (2017) A photographic catalog of primary types of Platygastroidea (Hymenoptera) in the National Museum of Natural History, Smithsonian Institution. In: Talamas EJ, Buffington ML (Eds) Advances in the Systematics of Platygastroidea. Journal of Hymenoptera Research 56: 187–224.

Talamas EJ, Bremer JS, Moore MR, Bon M-C, Lahey Z, Roberts CG, Combee LA, McGathey N, van Noort S, Timokhov AV, Hougardy E, Hogg B (2021) A maximalist approach to the systematics of a biological control agent: *Gryon aetherium* Talamas, sp. nov. (Hymenoptera, Scelionidae). In: Lahey Z, Talamas E (Eds) Advances in the Systematics of Platygastroidea III. Journal of Hymenoptera Research 87: 323–480. https://doi.org/10.3897/jhr.87.72842.

Talamas EJ, Herlihy MV, Dieckhoff C, Hoelmer KA, Buffington ML, Bon M-C, Weber DC (2015) *Trissolcus japonicus* (Ashmead) (Hymenoptera, Scelionidae) emerges in North America. Journal of Hymenoptera Research 43: 119–128. doi: 10.3897/JHR.43.4661.

Tamura K, Stecher G, Kumar S (2021) MEGA11: Molecular Evolutionary Genetics Analysis version 11. Molecular Biology and Evolution 38: 3022–3027.

United Nations Statistics Division (2023) https://unstats.un.org/unsd/methodology/m49/. Accessed 28 March 2023.

USDA (2021) USDA confirms Box Tree Moth and takes action to contain and eradicate the pest. USDA Animal and Plant Health Inspection Service. USDA APHIS | USDA Confirms Box Tree Moth and Takes Action to Contain and Eradicate the Pest. Accessed 23 March 2023.

VanDyk J (2023) BugGuide.Net: Identification, Images, & Information For Insects, Spiders & Their Kin For the United States & Canada. Iowa State University. https://bugguide.net/. Accessed 23 March 2023.

Vlug H (1995) Catalogue of the Platygastridae (Platygastroidea) of the world (Insecta: Hymenoptera). C. van Achterberg (Ed) Hymenopterum catalogus Pars 19. SPB Academic Publishing, Amsterdam.

Wheeler QD, Cracraft J (1997) Taxonomic preparedness: Are we ready to meet the biodiversity challenge? Pp. 435–446. In: Reaka-Kudla ML, Wilson DE, Wilson EO (Eds) Biodiversity II: Understanding and protecting our biological resources. Joseph Henry Press; Washington, D. C.

Wickham H (2016) Ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.

Wickham H (2022) Stringr: Simple, Consistent Wrappers for Common String Operations. R package version 1.4.1. [accessed 3 May 2023]. https://CRAN.R-project.org/package=stringr.

Wickham H, François R, Henry L, Müller K (2022) dplyr: A Grammar of Data Manipulation. R package version 1.0.10. [accessed 3 May 2023]. https://CRAN.R-project.org/package=dplyr.

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, Robertson T, Vieglais D (2012) Darwin Core: an evolving community-developed biodiversity data standard. PLoS ONE 7(1): e29715. https://doi.org/10.1371/journal.pone.0029715.

Zohary D, Hopf M, Weiss E (2012) Domestication of plants in the Old World. Fourth Edition. Oxford University Press.