Global Malaise Trap Program

Progress Report 2017

Prepared by the Collections Unit,
Centre of Biodiversity Genomics (previously Biodiversity Institute of Ontario)
February 2017

INTRODUCTION

The Global Malaise Program (GMP) is an international collaboration between the Centre of Biodiversity Genomics (CBG, previously Biodiversity Institute of Ontario) and a growing number of international contributors. The program represents a first step toward the acquisition of detailed temporal and spatial information on terrestrial arthropod communities across the globe. The program addresses the current lack of a systematic approach for tracking shifts in the species composition of terrestrial communities in response to environmental disturbance or global climate change. In comparison to water quality assessments, which are routinely based on surveys of the species composition of freshwater invertebrates; terrestrial environmental assessments lack a standard protocol to derive a biotic index, and instead generally rely on surveys of a few indicator taxa (e.g., birds, vascular plants) supplemented by qualitative habitat assessments. The use of indicator taxa disregards an important reality – most species in terrestrial ecosystems are arthropods.

Past efforts to include arthropods in terrestrial assessments have faced two serious barriers: ineffective sampling due to habitat complexities, and unreliable tools for species identification. The latter barrier has now been circumvented by DNA barcoding, a method that utilizes sequence variation in a standardized gene fragment to rapidly sort and objectively differentiate species (Hebert et al., 2003). This approach also makes it possible to carry out large-scale sampling programs and enables a time- and cost-efficient approach for biodiversity assessments. The present study represents a pilot phase of a longer-term program that will involve regular assessments of arthropod diversity with the intention of creating a globally-connected network of arthropod community monitoring sites.

To date, GMP has reached out to over forty countries and sampling has occurred at 63 sites. From 2012 to 2016, Malaise traps were deployed in ecosystems as diverse as Arctic tundra to tropical dry forest, running anywhere from 4-62 weeks with an average of 28 samples analyzed per location. Weekly samples were preserved in 95% ethanol and stored at -4°C to -20°C. All collection bottles were shipped for subsequent processing at CBG. Samples were accessioned, specimens were identified to order, arrayed, labeled, databased, and tissue-sampled for genetic analysis (Figure 1). All arthropods from samples selected for processing were barcoded, with the exception of a few very common species of Collembola, where only a few individuals from each trap sample were analyzed. Standard barcoding protocols (http://ccdb.ca/resources.php) were followed to recover the barcode region of cytochrome *c* oxidase subunit I (COI) gene. The barcode sequences, specimen images and collateral data are stored in the Barcode of Life Data Systems (BOLD; www.boldsystems.org). The project is available in the 'Global Malaise Program' campaign on BOLD. Barcoded specimens were assigned to an existing or new Barcode Index Number (BIN), a proxy for a formal Linnean species name, as outlined by Ratnasingham & Hebert (2013). Identifications were assigned by the BOLD-ID Engine where possible, allowing preliminary species inventories to be completed for each location and facilitating comparisons among them.

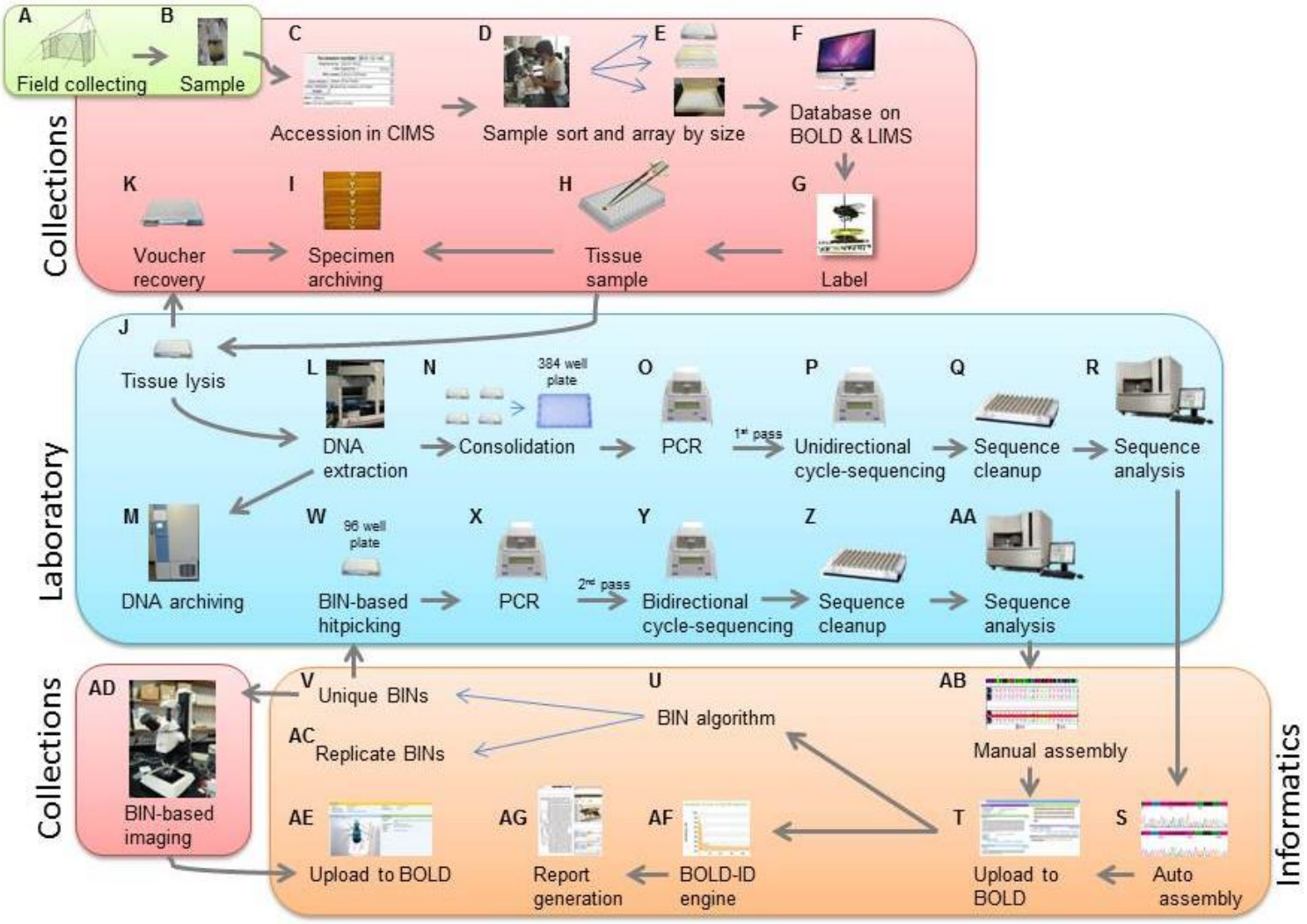

Figure 1. Schematic diagram showing specimen workflow. Front end processing begins with field collecting (A) and proceeds through to archiving of specimens (I). Laboratory analysis begins with tissue lysis (J) through to sequence analysis (AA). The informatics workflow includes both manual (AB) and auto sequence assembly (S), and finishes with BIN assignments and subsequent imaging of each BIN (AD).

PRELIMINARY RESULTS

Samples from 44 locations from 27 countries have been sequenced to date (Figure 2). A total of 1,262 Malaise samples from 44 sites which have completed processing or are near completion are included in this report (Figure 3). In total, over 1.03M specimens were analyzed and a total of 859,692 specimens generated barcode sequences that were long enough to allow a BIN assignment.

Their analysis revealed a total of 106,770 BINS (Figure 4). The usual 'hollow curve' species abundance pattern was observed, with 50,505 proxy species represented by just a single individual (singletons). By comparison, just 1206 BINs were represented by 100 or more individuals (Figure 5).

Figure 2. Sampling locations and their sample sizes (total specimens sorted) at the 44 GMP sites processed to date.

Figure 3. Temporal distribution of samples processed from each GMP site by year.

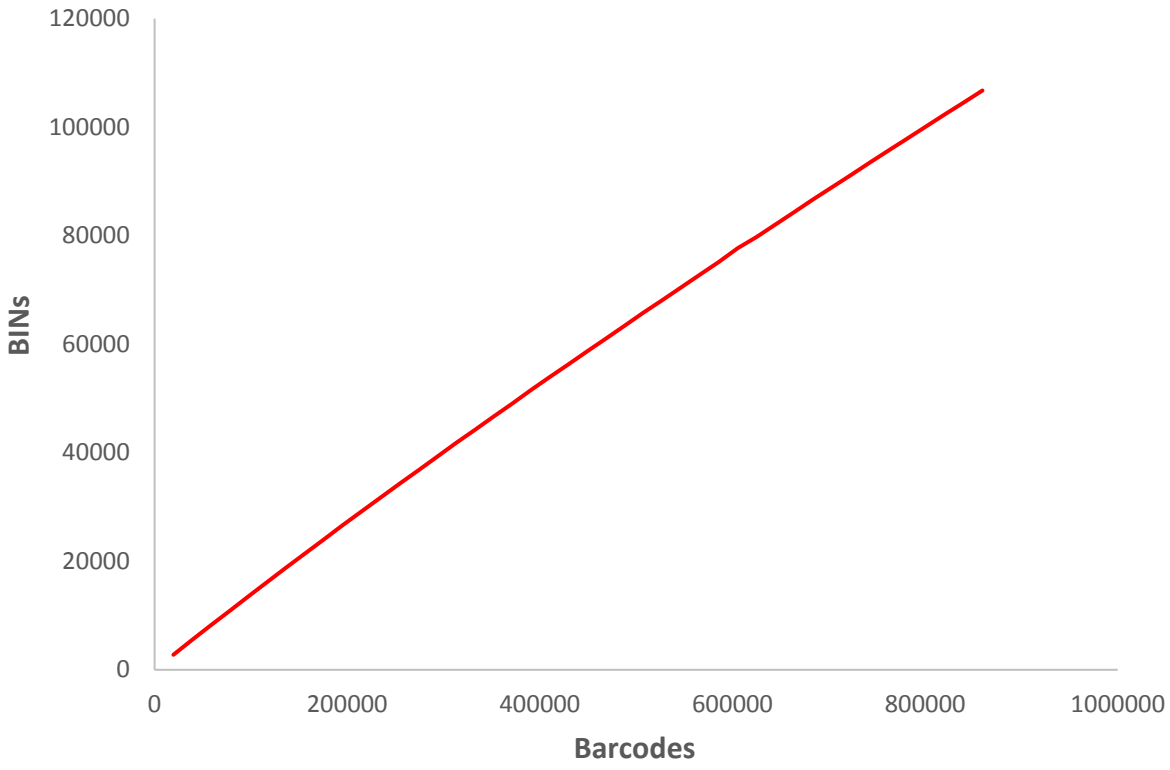

Figure 4. BIN accumulation curves for the 1262 Malaise trap samples collected in 44 GMP sites analyzed to date.

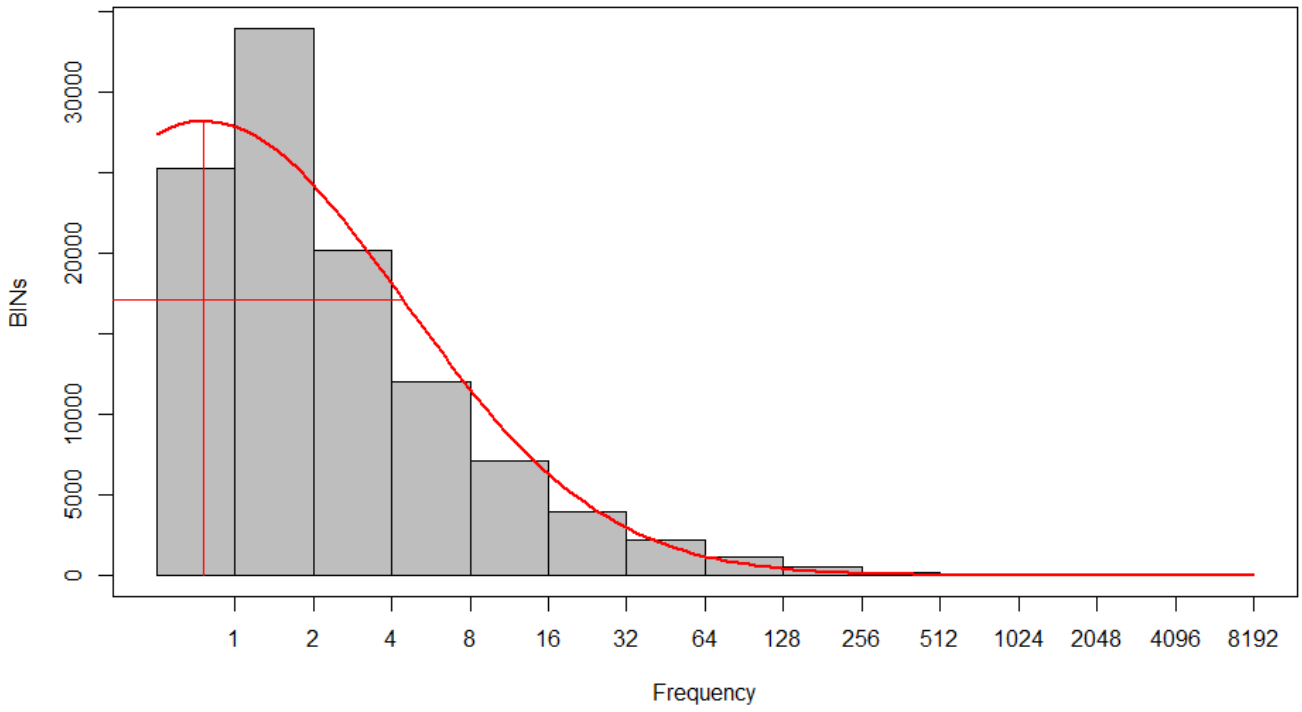

Figure 5. Lognormal species abundance curve, showing the total BINs within each log2 abundance frequency interval (Preston, 1962).

The number of individuals collected in each park varied nearly 100-fold ranging from a low of 679 specimens from 5 weekly samples at ECOSUR Chetumal, Mexico to over 75K specimens from 52 samples collected from the Misiones Province in Argentina. The number of BINS detected ranged from a low of 114 from ECOSUR Chetumal, Mexico, to a high of 8664 at the Misiones, Argentina (Figure 6). There was evidence for a correlation between sample size and the number of BINs detected (r^2 = 0.6785, $p<0.001$).

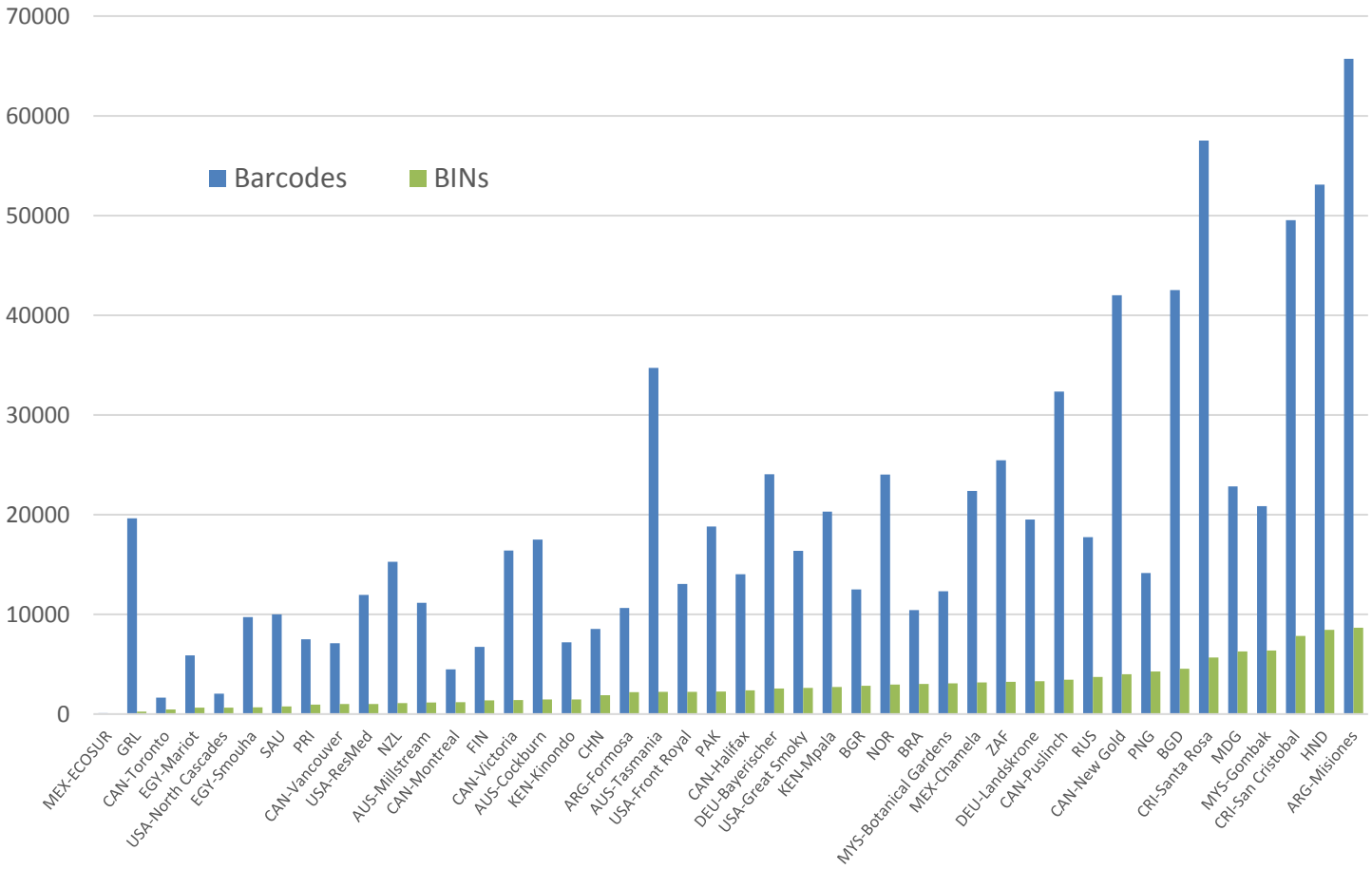

Figure 6. Total sequences and number of BINs generated from 44 GMP locations.

Of the 106K BINs captured, 64.8% were unique to a single collection site; i.e. 69,161 BINs occurred in only one of the 44 sites analyzed so far. The number of BINs unique to each location varied (Figure 7). Malaysia exhibited the highest count of unique BINs with 6967 of 8664 being unique (80%) while the Zackenberg Research Station in Greenland had the fewest unique BINs (N = 16) and also the lowest ratio of unique BINs to BINs captured. The site with the highest proportion of unique BINs to total BINs is Madagascar with 93% of its BINs only captured at that site. A significant negative correlation was observed between the proportion of unique BINs at a site and its distance from the equator (Figure 8).

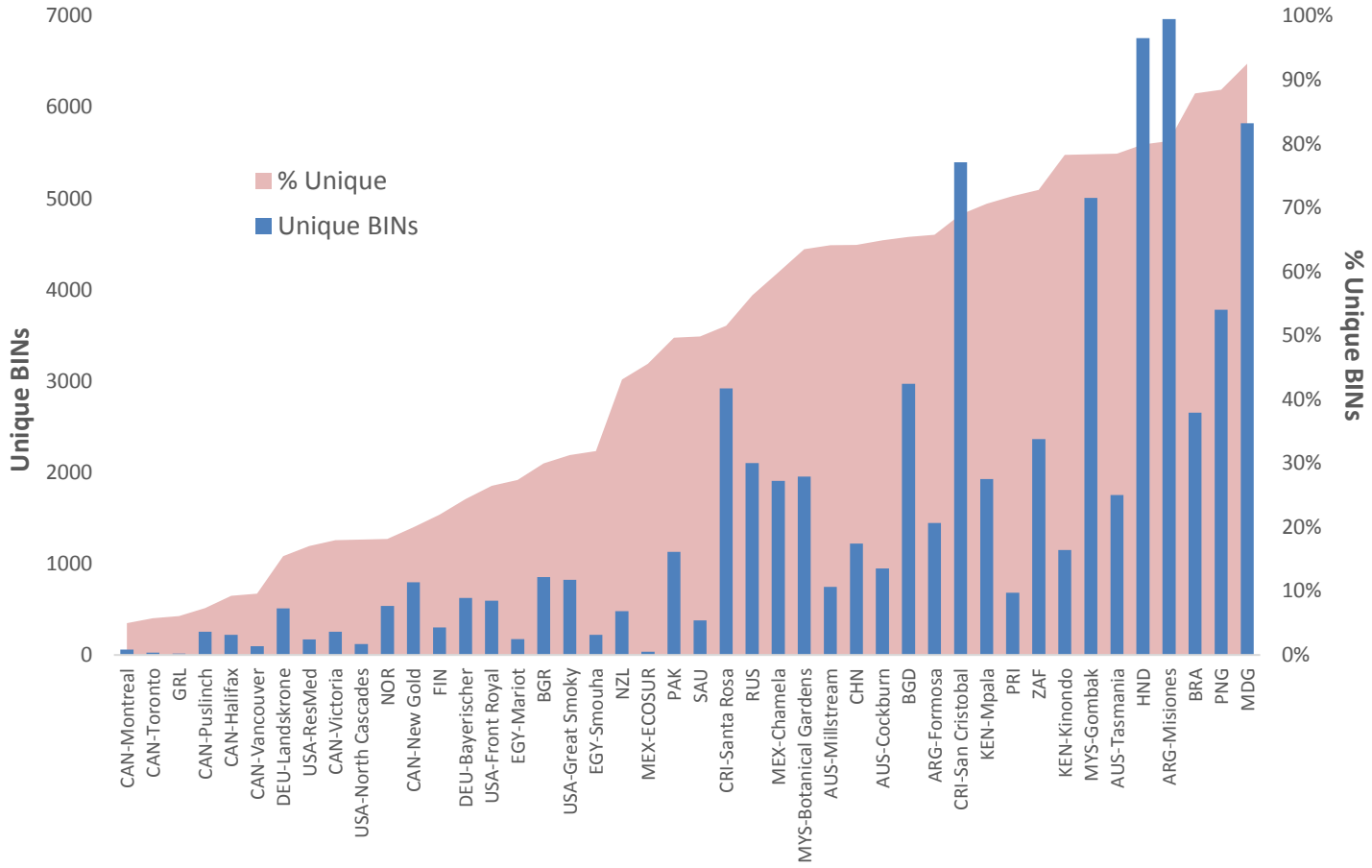

Figure 7. Total number of BINs unique to each GMP site (bars) and the percentage of unique BINs collected in each site (Unique BINs/Total BINs).

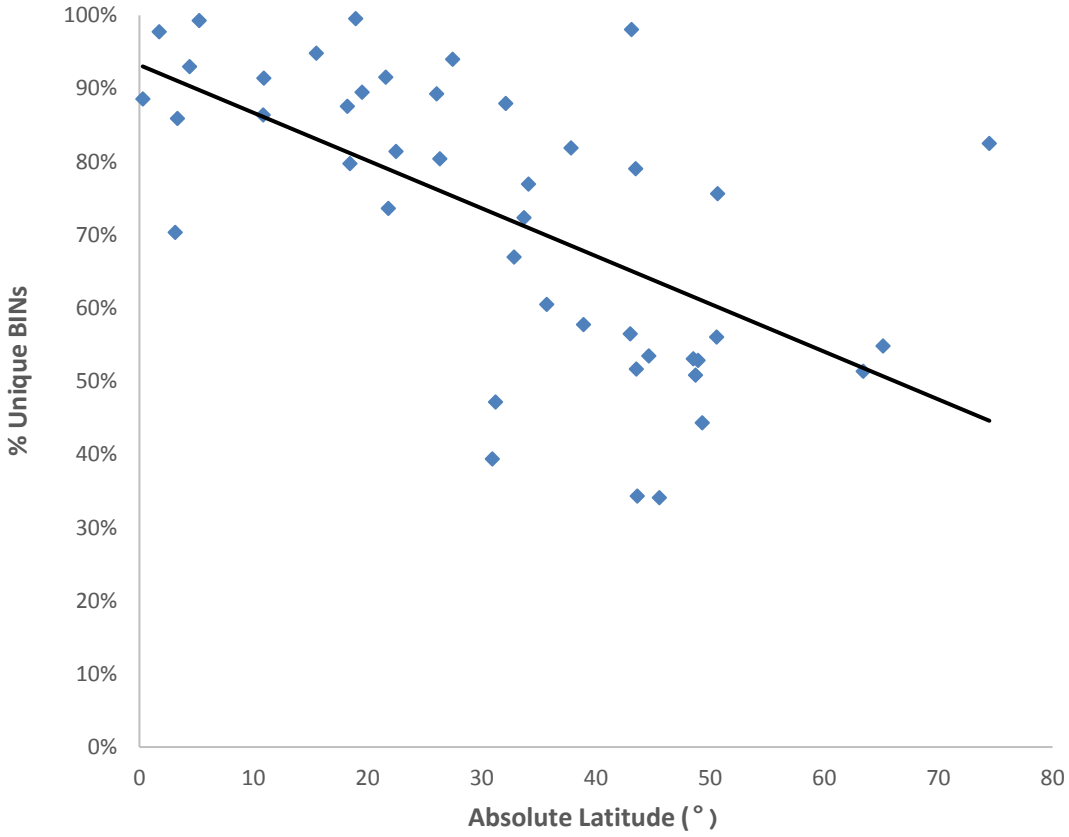

Figure 8. Regression analysis examining the relationship between the proportion of unique BINs in each site and its distance from the equator or absolute latitude (r^2= 0.3728, p<0.001).

The similarity in species composition between sites showed marked variation (Appendix 1). For example, the two sites in Egypt, which are separated by less than 50km, shared the highest proportion of BINs, with a Chao's Sorenseon Similarity Index (Chao et al., 2005) of 0.342. This high species similarity was followed by two sites in the USA (Front Royal, Virginia and Great Smoky National Park), separated by nearly 600km, with a Similarity Index of 0.245. The two sites farthest apart were Millstream National Park, Australia and Puerto Rico, and had a Similarity Index of 0.002. While the two closest sites from different countries, Vancouver, Canada and North Cascades, USA, had a Similarity Index of 0.188. In addition, a moderate negative correlation was observed between geographic distance and Chao's Sorenson Similarity values (Figure 9).

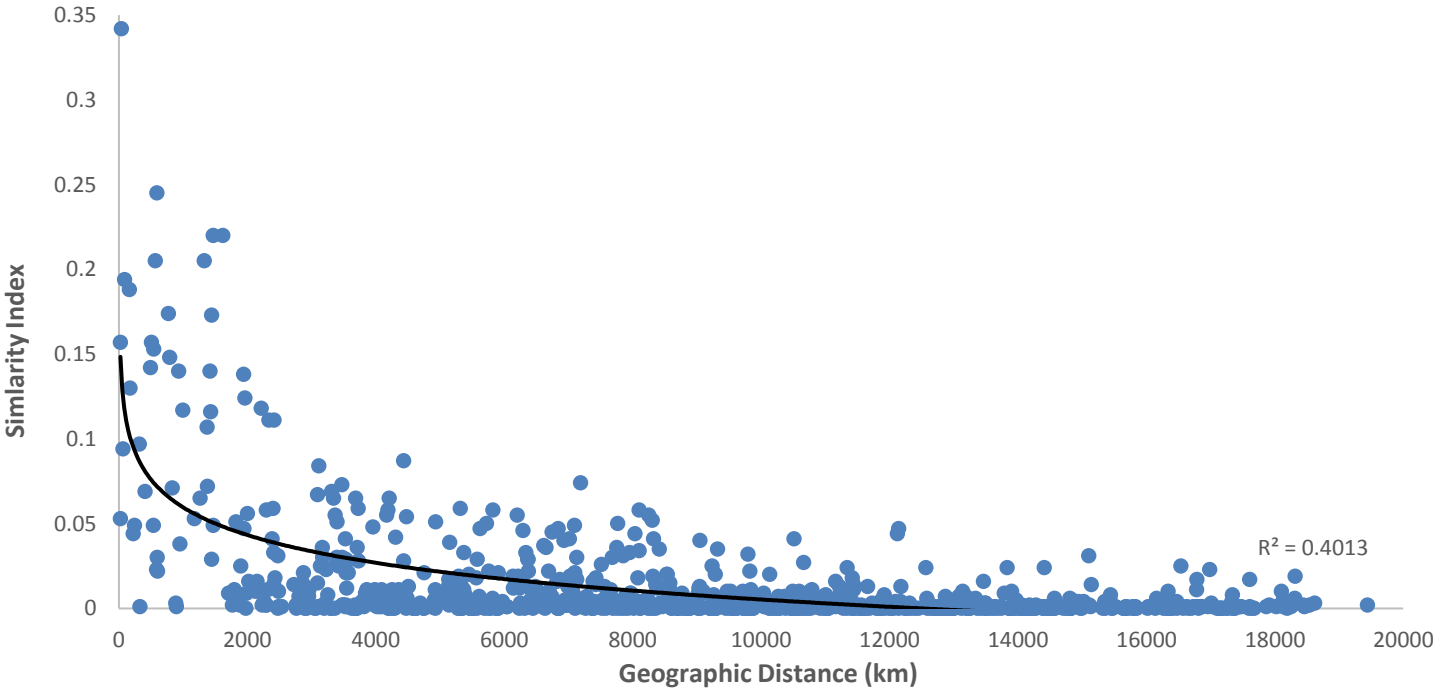

Figure 9. The relationship between geographic distance and species similarity. Similarity is based on Chao-Sorensen Raw Abundance data; each point represents a pair of locations.

The 106K BINs detected so far from 860K records were classified under 57 different orders and 785 different families. As expected, the most abundant order collected in GMP was Diptera, comprising 55% of the collected taxa. This was followed by Hymenoptera which comprised 17% then Hemiptera, Lepidoptera, and Coleoptera comprising 7%, 5%, and 4% respectively. While the major insect orders were encountered the most, the traps have also captured a considerable amount of diversity with another 51 orders from 9 taxonomic classes comprising 10.8% of all collected taxa (Figure 10).

As of February 2017, 93.8% of GMP BINs were identified to at least family using the BOLD ID Engine and morphological analysis of certain groups by taxonomic experts. In total 9547 arthropod species were named, representing 10% of the total BINs detected. It is important to emphasize that it will be possible to identify taxa lacking a species name as the barcode reference library becomes more complete.

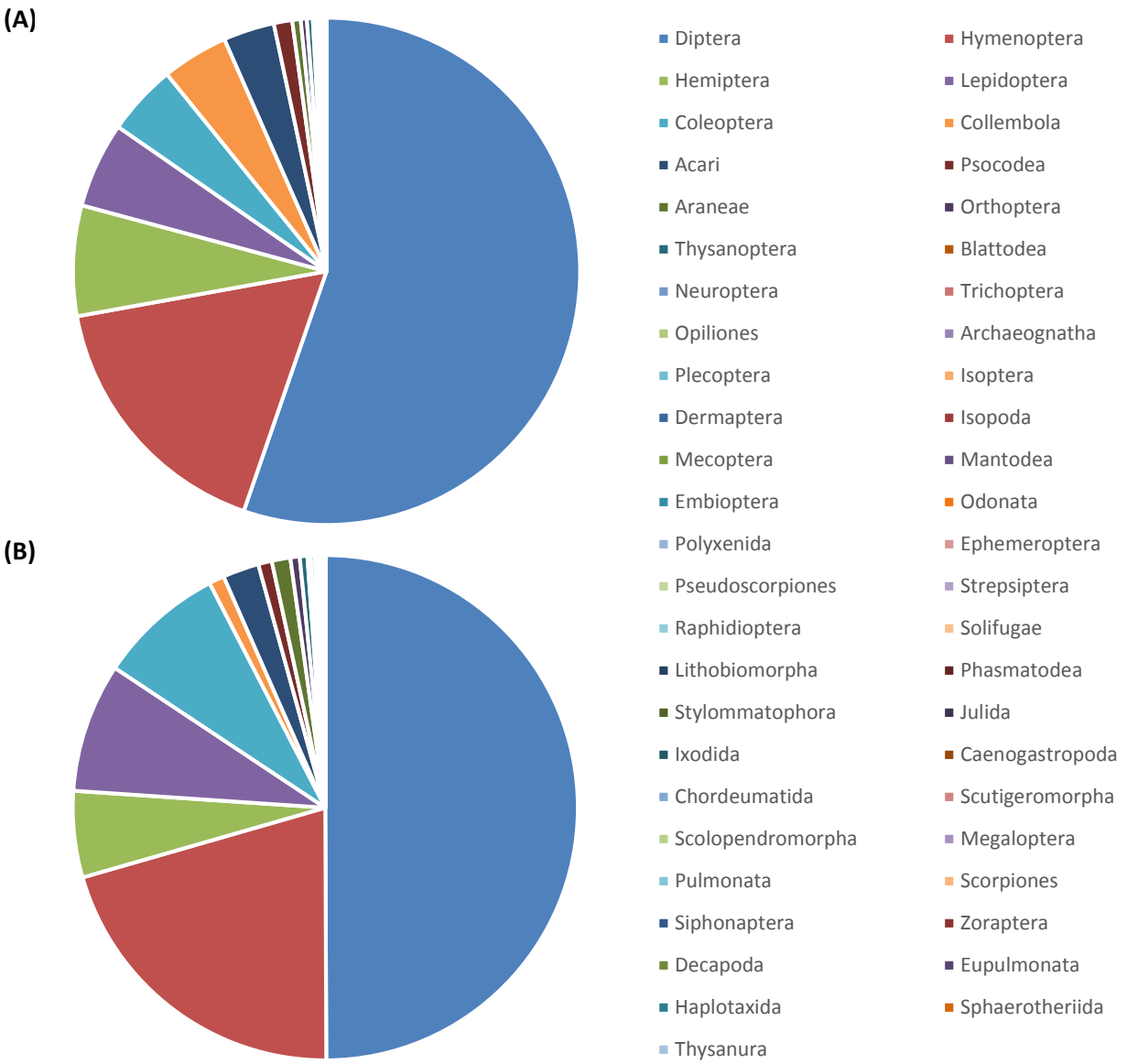

Figure 10. Taxonomic breakdown of (A) N total specimens and (B) N total BINs collected and analyzed from 44 GMP sites.

REFERENCES

Chao, A., R.L. Chazdon, R.K. Colwell, and T.-J. Shen (2005). A new statistical approach for assessing compositional similarity based on incidence and abundance data. Ecology Letters 8: 148-159.

Hebert, P.D.N., A. Cywinska, S.L. Ball, and J.R. deWaard (2003). Biological identifications through DNA barcodes. Proceedings of the Royal Society of London B 270: 313-321.

Preston, F.W. (1962). The Canonical Distribution of Commonness and Rarity: Part I. Ecology 43: 185-215.

Ratnasingham, S. and P.D.N. Hebert (2013). A DNA-based registry for all animal species: the Barcode Index Number (BIN) System. Public Library of Science ONE 8: e66213.

GLOBAL MALAISE TRAP PROGRAM DISTRIBUTION MAP