

Global Malaise Trap Program

Progress Report
2015



Global
Malaise
Trap
Program

INTRODUCTION

The Global Malaise Program (GMP) is an international collaboration between the Biodiversity Institute of Ontario (BIO) 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 thirty countries and sampling has occurred at 63 sites. In the 2012-2014 sampling season, 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 22 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 BIO. Samples were accessioned, specimens were identified to order, arrayed, labeled, databased, and tissue-sampled for genetic analysis (Figure 1). All arthropods 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 publicly 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.

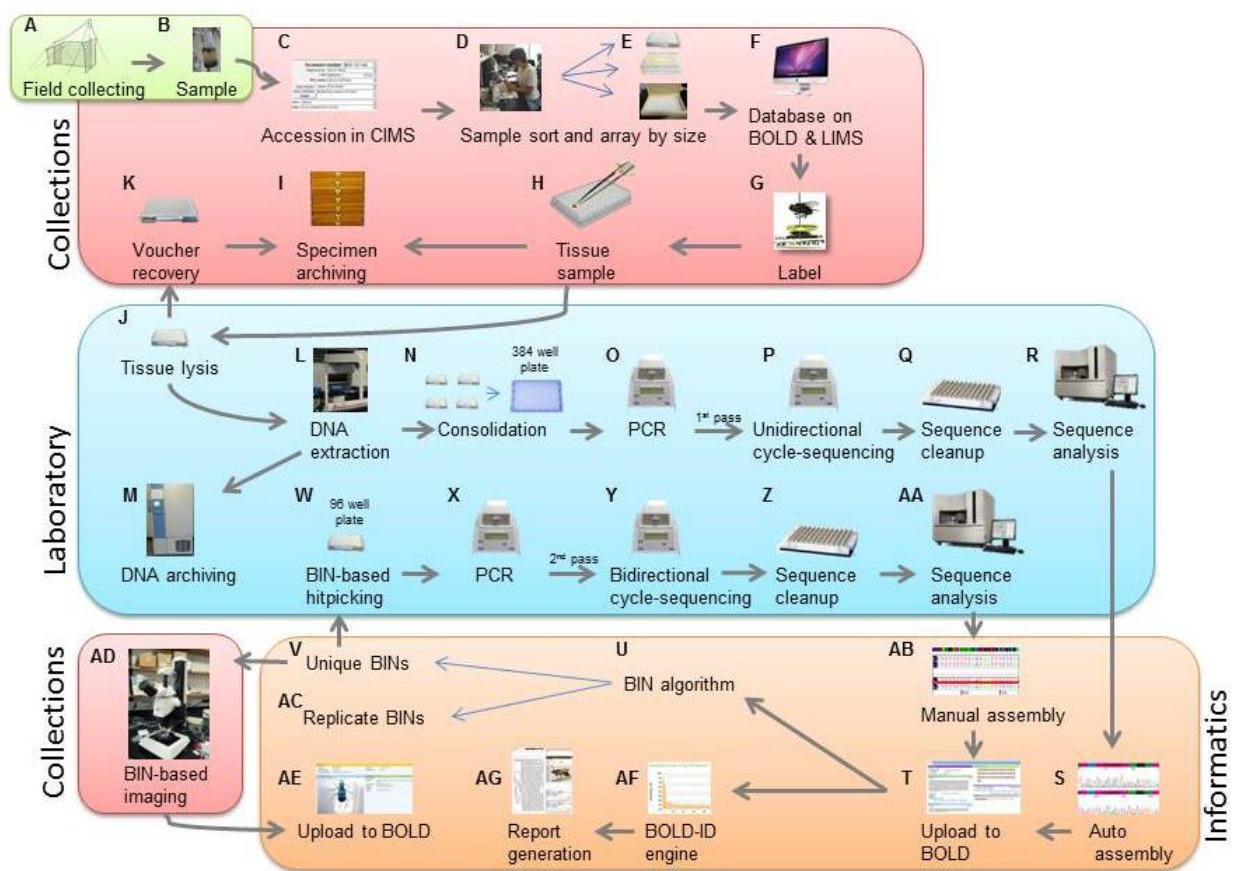


Figure 1. Schematic diagram showing specimen workflow. Front end processing begins with field collecting (A) and proceeds through to archiving of specimens (I). Laboratory analysis begins with tissue lysis (J) through to sequence analysis (AA). The informatics workflow includes both manual (AB) and auto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

PRELIMINARY RESULTS

Samples from thirty-seven locations from 27 countries have been sorted to date (Figure 2) with thirty-three having sequences on BOLD. A total of 703 Malaise samples from 30 sites which have completed processing or are near completion are included in this report (Table 1). In total, over 675K specimens were analyzed and a total of 502,364 specimens generated barcode sequences that were long enough to allow a BIN assignment.

Their analysis revealed a total of 67,202 BINS (Figure 3). The usual 'hollow curve' species abundance pattern was observed, with 33,162 proxy species represented by just a single individual (singletons). By comparison, just 712 BINs were represented by 100 or more individuals (Figure 4).

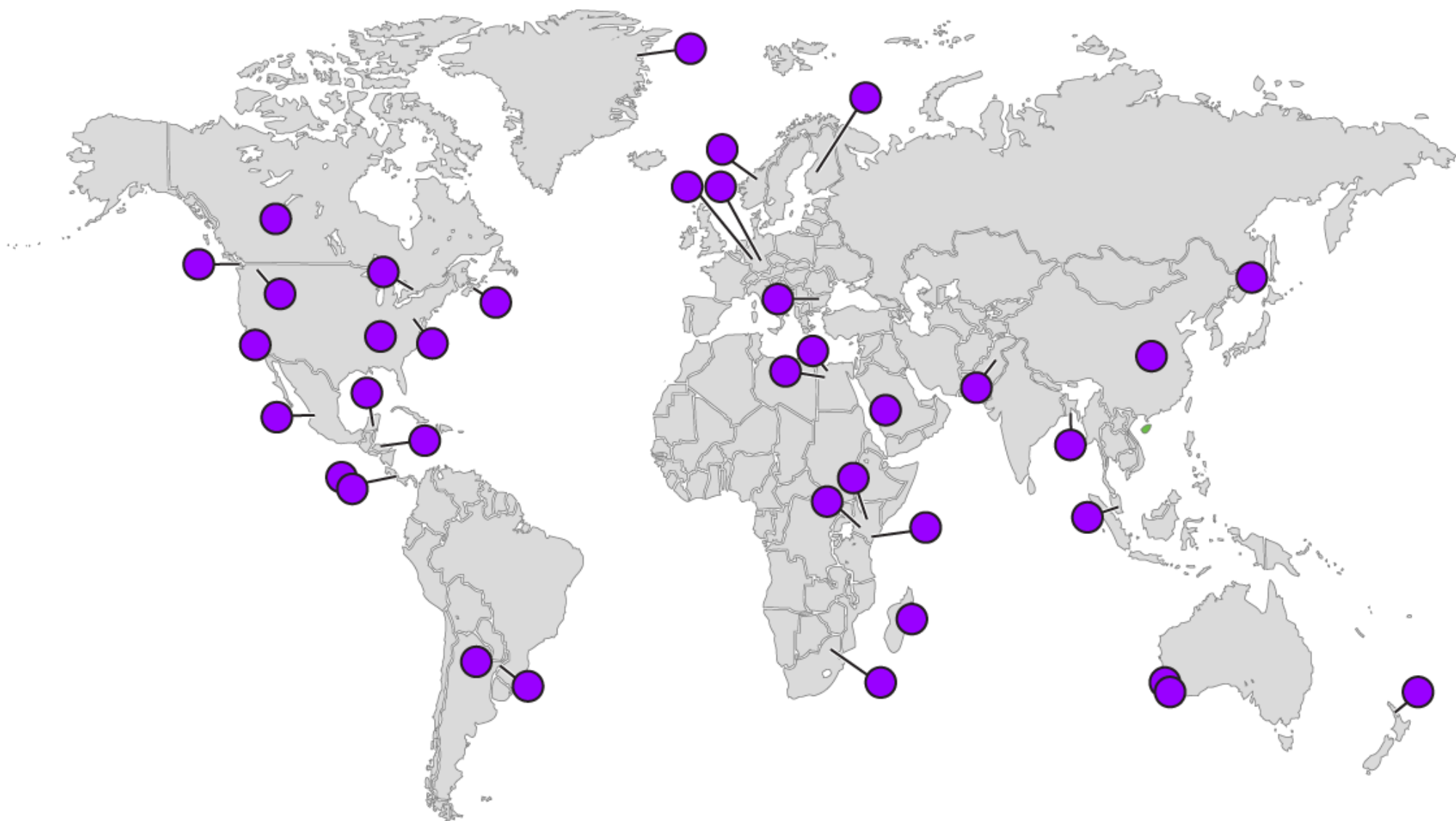


Figure 2. Sampling locations at the 37 GMP sites sorted to date.

Table 1. GMP sampling sites and number of Malaise samples processed to date.

GMP Year	Country	Site	Samples Processed
2012	Bulgaria	Sofiya, Godech	7
2012	Canada	Puslinch	24
2012	China	Haopingsi Nature Reserve Station	20
2012	Costa Rica	Santa Rosa (Dry Forest)	51
2012	Finland	Oulu, NE of Kiiminki	20
2012	Germany	Bayerischer National Park	9
2012	Greenland	Zackenberg Research Station	10
2012	Honduras	Cusuco National Park	16
2012	Malaysia	Gombak Field Studies Centre	20
2012	New Zealand	Waikato, Hamilton	25
2012	Pakistan	Pakistan Museum of Natural History	25
2012	USA	Great Smoky Mountain National Park	32
2012	USA	Northern Cascades National Park	16
2013	Argentina	Misiones	32
2013	Canada	Halifax, Point Pleasant Park	20
2013	Canada	New Gold New Afton Mine	32
2013	Egypt	Mariot	20
2013	Egypt	Smouha	17
2013	Germany	Landskrone	14
2013	Greenland	Zackenberg Research Station	10
2013	Honduras	Cusuco National Park	12
2013	Mexico	ECOSUR	5
2013	Mexico	Jalisco, Chamela	29
2013	South Africa	Magaliesburg	25
2014	Bangladesh	Chittagong	13
2014	Greenland	Zackenberg Research Station	14
2014	Honduras	Cusuco National Park	40
2014	Madagascar	Andasibe	14
2014	Norway	Trondheim, Sommerlystvegen	34
2014	Russia	Kiparisovo	4
2014	Saudi Arabia	Jeddah- Hada Al-Sham	15
2014	USA	ResMed	52
2014	USA	Front Royal	26

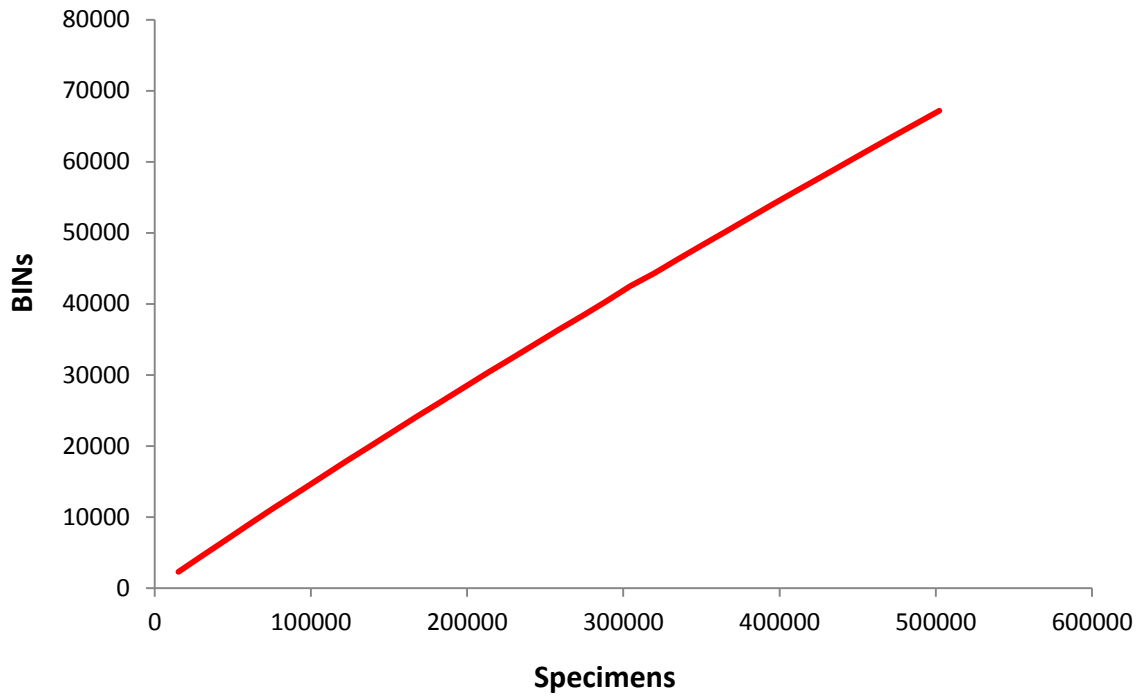


Figure 3. BIN accumulation curve for the 703 Malaise trap samples collected in 33 GMP sites analyzed to date.

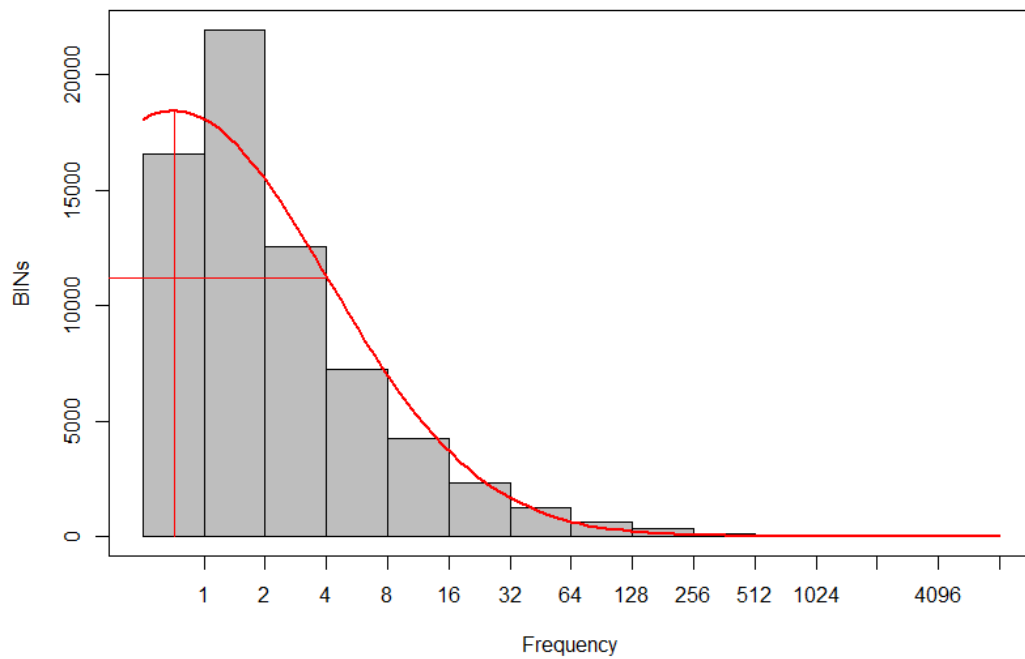


Figure 4. Lognormal species abundance curve, showing the total BINs within each \log_2 abundance frequency interval (Preston, 1962).

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 63,514 specimens from 51 samples collected from the Santa Rosa Sector of the Area de Conservacion Guanacaste (ACG) in Costa Rica. The number of BINS

detected ranged from a low of 114 from ECOSUR Chetumal, Mexico, to a high of 6344 at the Gombak Field Studies Centre in Malaysia (Figure 5). There was no evidence for a strong correlation between sample size and the number of BINS detected (Figure 6, $r^2 = 0.1118$).

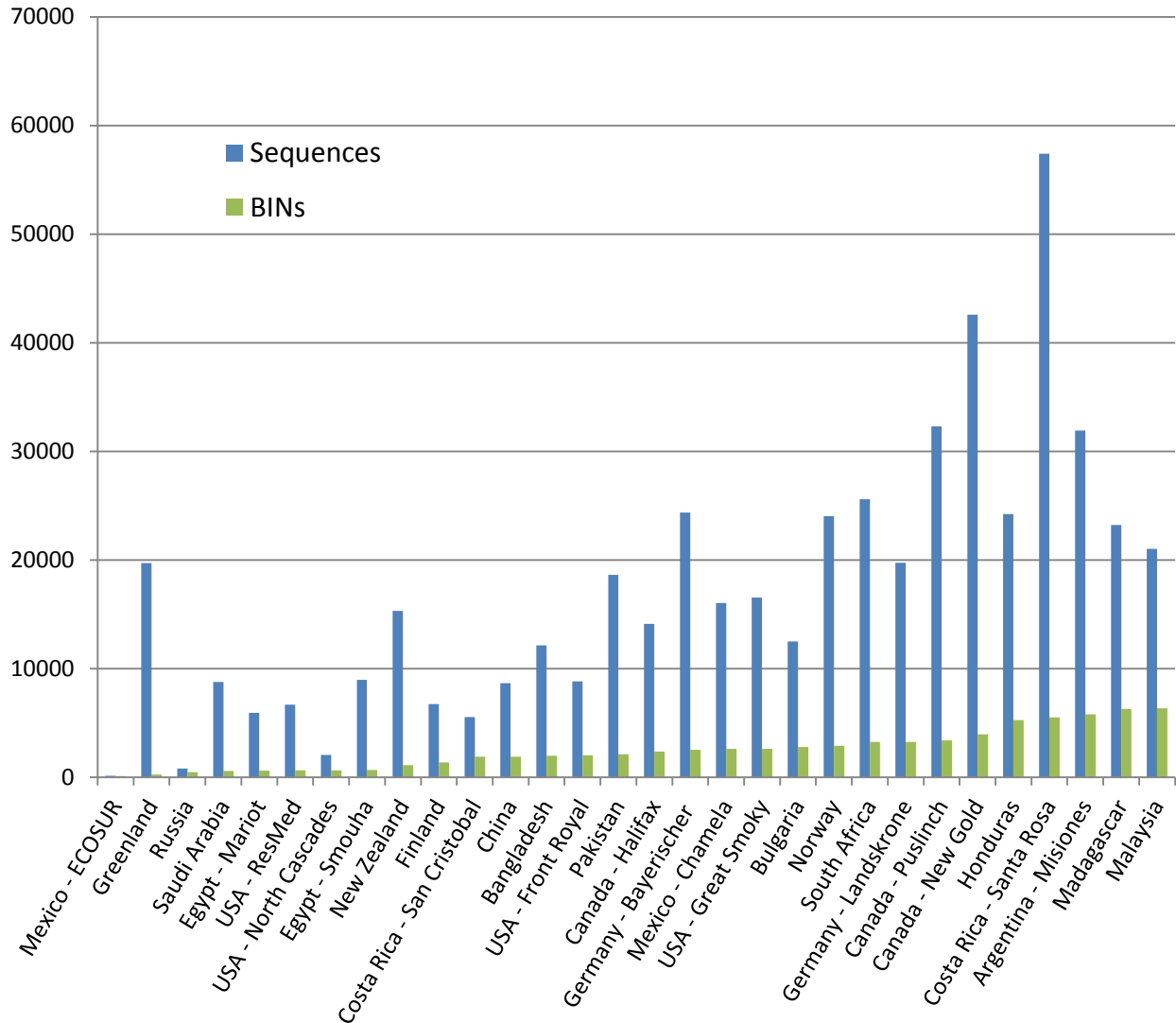


Figure 5. Total sequences and number of BINs generated from 30 GMP locations.

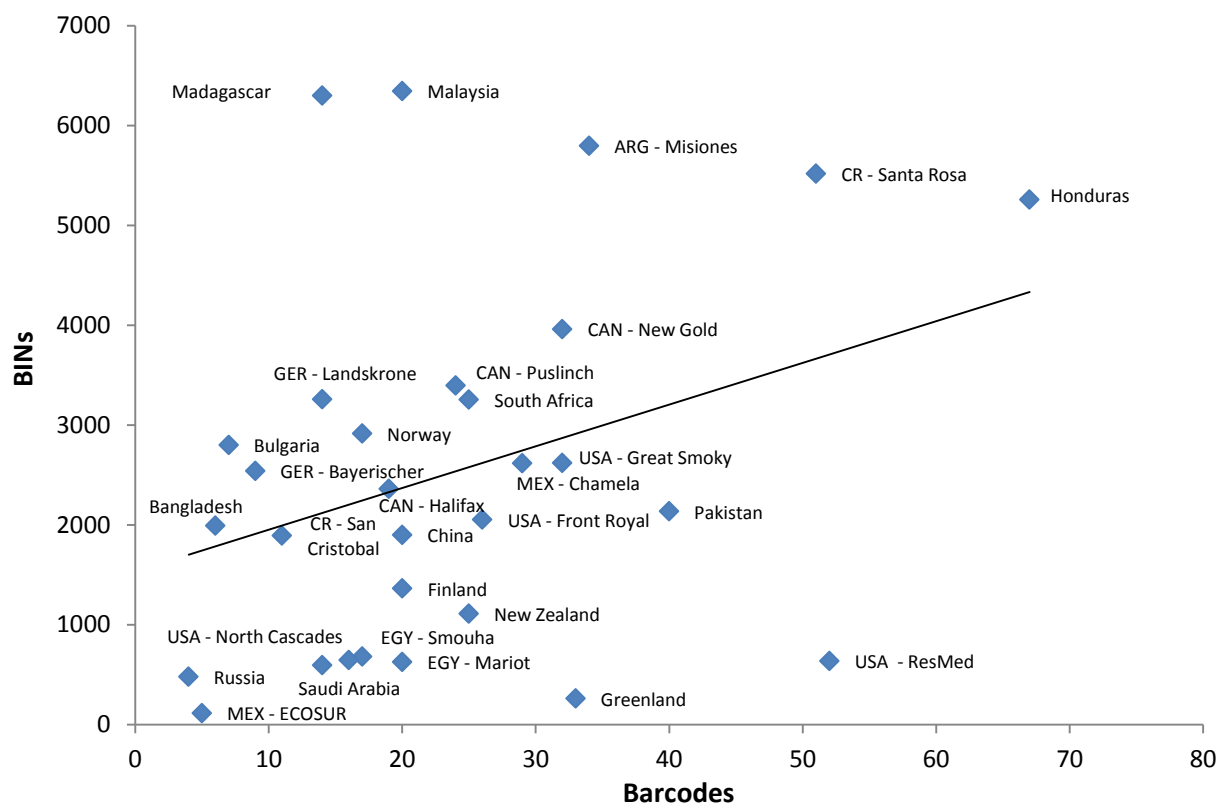


Figure 6. Regression analysis examining the relationship between the number of barcoded specimens and number of BINs ($r^2 = 0.1118$, $p > 0.05$).

Of the 67K BINs captured, 37.6% were unique to a single collection site; i.e. 25,267 BINs occurred in only one of the 33 sites analyzed so far. The number of BINs unique to each location varied (Figure 7). Malaysia exhibited the highest count of unique BINs as over half of its BINs were unique (3707 of 6344) while the Zackenberg Research Station in Greenland had the fewest unique BINs ($N = 8$) and also the lowest ratio of unique BINs to BINs captured.

The similarity in species composition between sites showed marked variation (Table 2). For example, the two sites in Egypt, which are separated by less than 50km, shared the highest

proportion of BINs, with a Chao's Sorensen Similarity Index (Chao et al., 2005) of 0.338. This high species similarity was followed by two sites in the USA (Front Royal, Virginia and Great Smoky National Park), separated by over 3,500km, with a Similarity Index of 0.232. The two sites farthest apart were Hamilton, New Zealand and Landskrone, Germany, and had a Similarity Index of 0.019. While the two closest sites from different countries, New Gold, Canada and North Cascades, USA, had a Similarity Index of 0.044. In addition, a significant negative correlation was observed between geographic distance and Chao's Sorensen Similarity values (Figure 8).

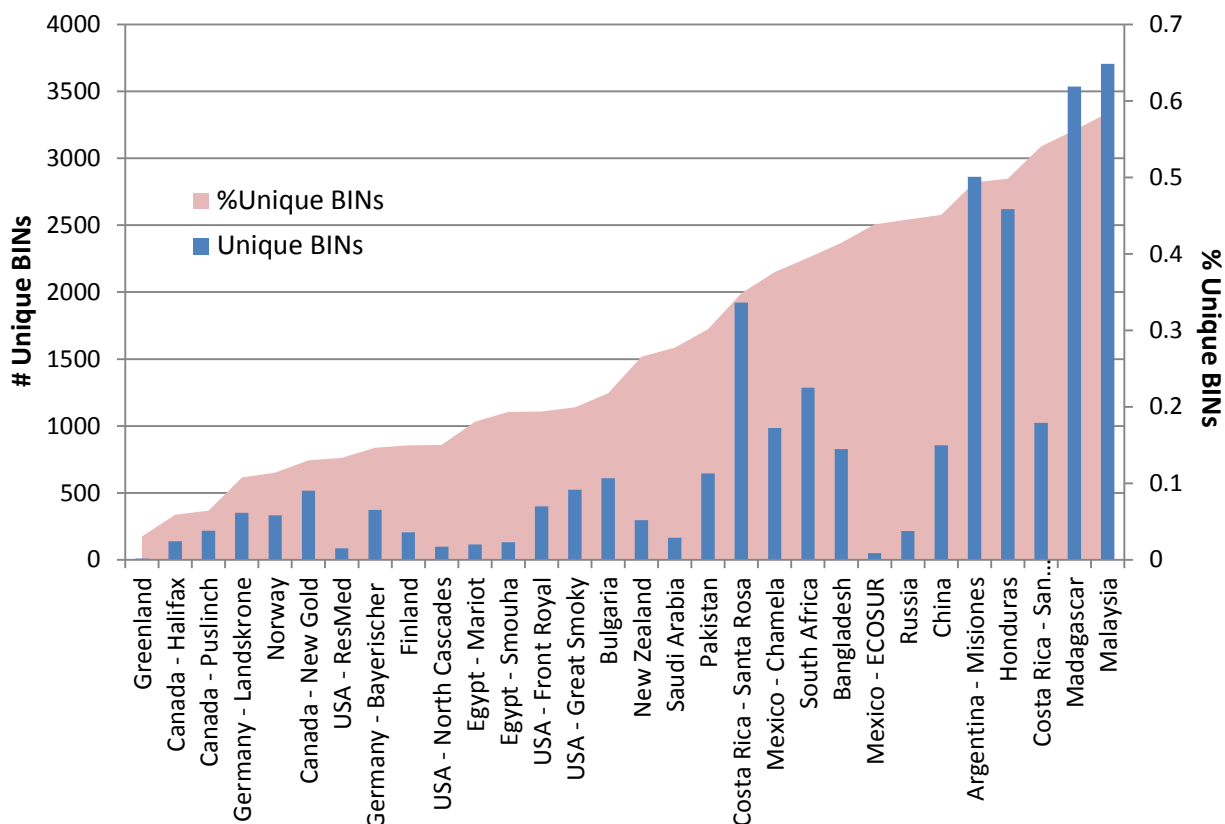


Figure 7. Total number of BINs unique to each GMP site (bars) and the percentage of unique BINs collected in each site (Unique BINs/Total BINs).

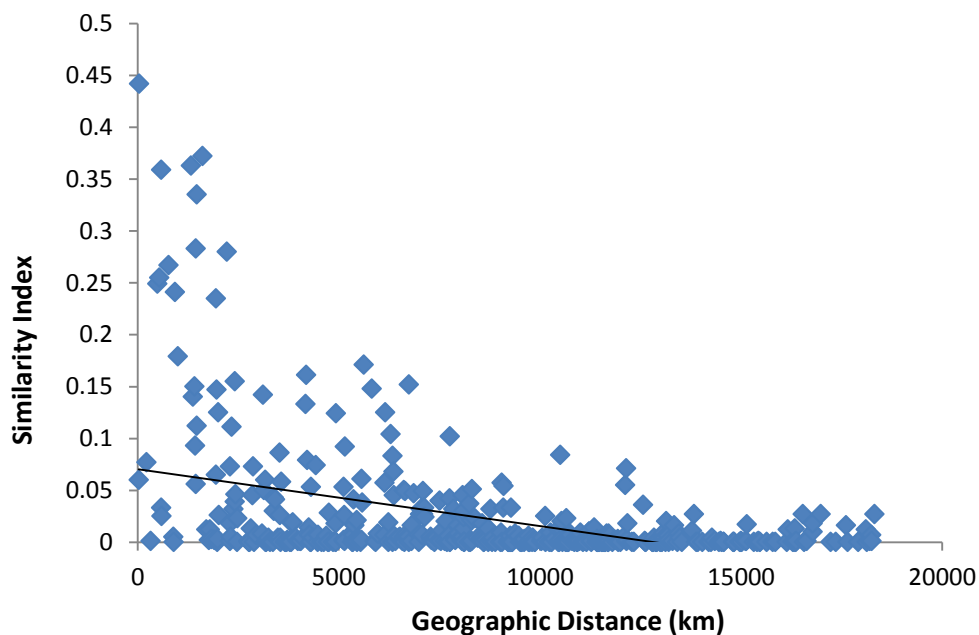


Figure 8. The relationship between geographic distance and species similarity. Similarity is based on Chao-Sorensen Raw Abundance data; each point represents a pair of locations. There is a significant negative correlation between the two variables ($r^2 = 0.18$, $p < 0.001$).

Table 2. Species overlap between GMP locations based on Chao-Sorensen Raw Abundance.

	Argentina - Misiones	Bangladesh	Bulgaria	Canada - Halifax	Canada - New Gold	Canada - Puslinch	China	Costa Rica - San Cristobal	Costa Rica - Santa Rosa	Egypt - Mariot	Egypt - Smouha	Finland	Germany - Bayerischer	Germany - Landskrone
Bangladesh	0.001													
Bulgaria	0.001	0.006												
Canada - Halifax	0.002	0.001	0.038											
Canada - New Gold	0.001	0.002	0.039	0.065										
Canada - Puslinch	0.001	0.002	0.049	0.205	0.083									
China	0.001	0.015	0.03	0.008	0.008	0.009								
Costa Rica - San Cristobal	0.004	0.001	0	0	0	0	0							
Costa Rica - Santa Rosa	0.008	0.002	0	0.001	0.001	0.002	0.001	0.03						
Egypt - Mariot	0.002	0.023	0.049	0.006	0.01	0.009	0.017	0	0.002					
Egypt - Smouha	0.002	0.028	0.028	0.008	0.007	0.008	0.014	0.001	0.002	0.338				
Finland	0	0.001	0.031	0.03	0.015	0.029	0.012	0	0	0.001	0.003			
Germany - Bayerischer	0	0.001	0.117	0.048	0.032	0.046	0.025	0	0	0.018	0.014	0.136		
Germany - Landskrone	0.001	0.004	0.22	0.038	0.036	0.047	0.031	0	0.001	0.021	0.015	0.048	0.141	
Greenland	0	0	0.005	0.007	0.011	0.006	0.003	0	0	0.004	0.002	0.01	0.008	0.007
Honduras	0.002	0.001	0	0.001	0.001	0.001	0	0.013	0.019	0.002	0.002	0	0	0
Madagascar	0	0.001	0	0	0	0	0	0	0.001	0.001	0.001	0	0	0
Malaysia	0	0.026	0.002	0.001	0.001	0	0.002	0.001	0.003	0.002	0.002	0	0	0.001
Mexico - Chamela	0.003	0.001	0	0.002	0.002	0.003	0	0.005	0.046	0.001	0.002	0	0	0
Mexico - ECOSUR	0	0	0	0	0	0	0	0	0.004	0	0	0	0	0
New Zealand	0.001	0.003	0.017	0.014	0.013	0.023	0.005	0	0.001	0.018	0.012	0.01	0.01	0.019
Norway	0	0.002	0.112	0.052	0.036	0.059	0.018	0	0	0.01	0.009	0.173	0.221	0.172
Pakistan	0.002	0.114	0.028	0.008	0.009	0.008	0.027	0.001	0.002	0.06	0.057	0.003	0.013	0.018
Russia	0.001	0.003	0.012	0.006	0.005	0.004	0.036	0	0.001	0.007	0.007	0.008	0.011	0.013
Saudi Arabia	0.001	0.011	0.009	0.002	0.003	0.002	0.007	0	0	0.123	0.105	0	0.003	0.004
South Africa	0.002	0.01	0.013	0.006	0.009	0.012	0.006	0	0.002	0.033	0.029	0	0.004	0.008
USA - Front Royal	0.001	0.001	0.006	0.103	0.016	0.132	0.004	0	0.001	0.004	0.003	0.005	0.008	0.006
USA - Great Smoky	0.001	0	0.007	0.121	0.029	0.137	0.004	0	0.002	0.002	0.003	0.006	0.014	0.01
USA - North Cascades	0	0.001	0.007	0.04	0.044	0.029	0.002	0	0	0.003	0.005	0.018	0.018	0.008
USA - ResMed	0.003	0.01	0.02	0.018	0.041	0.017	0.012	0.002	0.003	0.046	0.041	0.004	0.009	0.018

Table 2 continued. Species overlap between GMP locations based on Chao-Sorensen Raw Abundance.

	Greenland	Honduras	Madagascar	Malaysia	Mexico - Chamela	Mexico - ECOSUR	New Zealand	Norway	Pakistan	Russia	Saudi Arabia	South Africa	USA - Front Royal	USA - Great Smoky	USA - North Cascades
Honduras	0														
Madagascar	0	0.001													
Malaysia	0	0.001	0.001												
Mexico - Chamela	0	0.004	0	0											
Mexico - ECOSUR	0	0.001	0	0	0.003										
New Zealand	0	0.001	0	0.001	0	0									
Norway	0.008	0	0	0	0	0	0.023								
Pakistan	0.004	0.001	0.001	0.011	0.001	0	0.009	0.011							
Russia	0	0	0	0.001	0	0	0.003	0.013	0.008						
Saudi Arabia	0.005	0	0	0.001	0.001	0	0.004	0.002	0.032	0.002					
South Africa	0.001	0.001	0.002	0.001	0.001	0	0.014	0.004	0.019	0.003	0.012				
USA - Front Royal	0.001	0.001	0	0	0	0	0.008	0.008	0.003	0.003	0	0.005			
USA - Great Smoky	0.003	0.002	0	0	0.003	0	0.005	0.011	0.004	0.001	0.001	0.003	0.232		
USA - North Cascades	0.002	0	0	0	0.001	0	0	0.017	0.004	0.002	0.002	0	0.019	0.025	
USA - ResMed	0.007	0.002	0.001	0.002	0.006	0	0.038	0.01	0.022	0.009	0.011	0.021	0.006	0.01	0.006

The 67K BINs detected so far from 500K records were classified under 41 different orders and 594 different families. As expected, the most abundant order collected in GMP was Diptera, comprising over 60% of the collected taxa. This was followed by Hymenoptera which comprised 14% then Lepidoptera, Coleoptera, and

Hemiptera each comprised around 5-7%. While the major insect orders were encountered the most, the traps have also captured a considerable amount of diversity with another 36 orders comprising 7% of all collected taxa (Figure 9).

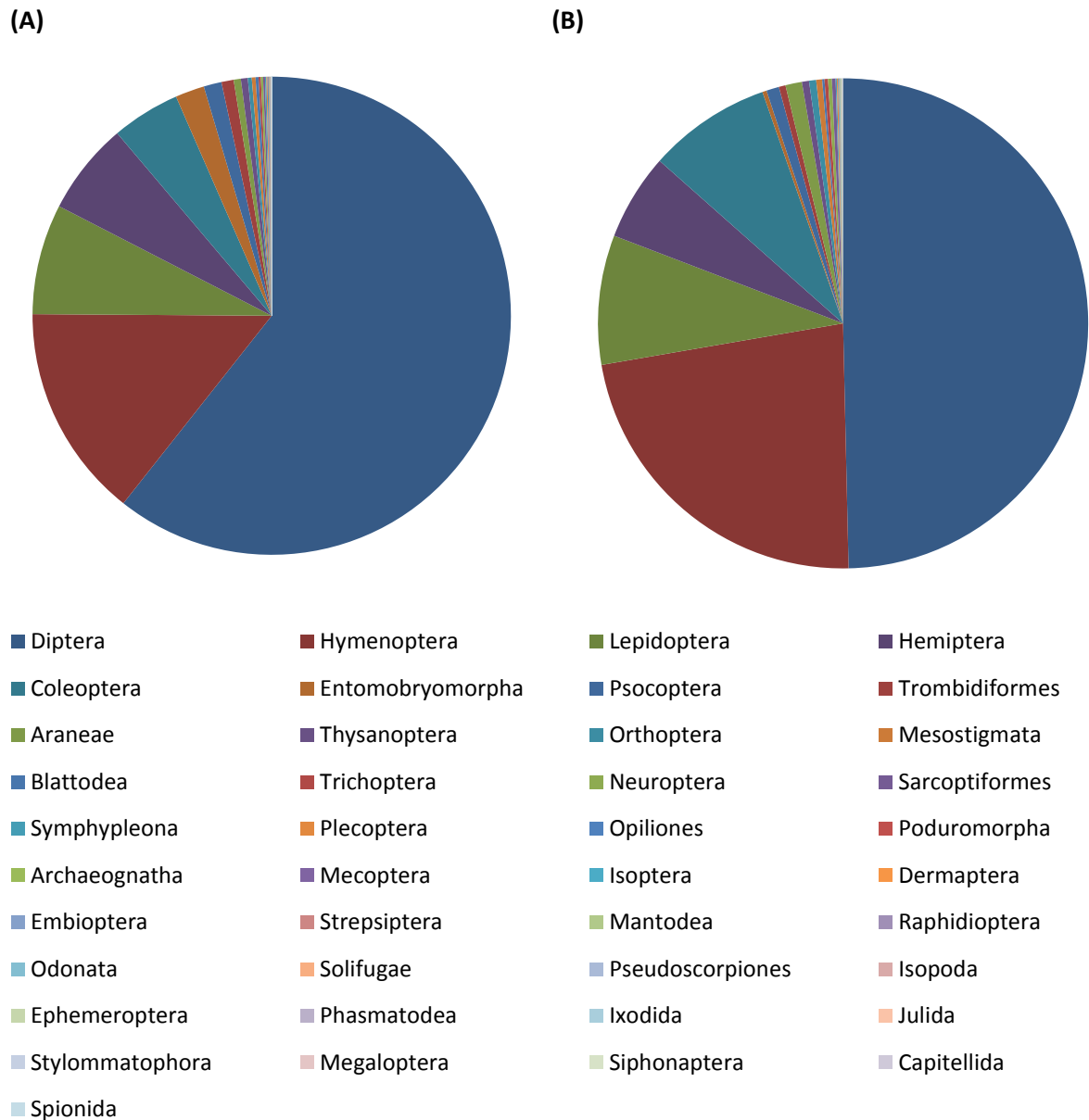


Figure 9. Taxonomic breakdown of (A) N total specimens and (B) N total BINs collected and analyzed from 33 GMP sites.

FUTURE DIRECTION

Going forward our plans will be two-fold. First, we will continue to sample many of the already established locations. Our hope is that these locations will act as nodes for more intense, localized sampling efforts, paving the way for a globally connected bio-monitoring program. Secondly, we will attempt to reach out to new locations (see Participation List). With new countries joining, the need for funds to cover analytical costs are rising. We are asking participants to sponsor the analysis of a single site for the insect flight season (anywhere from 10-52 weeks). We are approximating a range of

sampling regimes and costs, from Northern sites reaching over 10,000 specimens in 10 week seasons (~\$20,000) to temperate and tropical sites reaching 20,000+ specimens in 28-52 week seasons (~\$40,000+).

We anticipate that the researcher(s) overseeing malaise traps in a particular nation will generate a publication detailing the results obtained from the analysis of their collections. We also anticipate that all members of the project will join in a publication that will synthesize the overall results.

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.

PARTICIPANTS

2012-2013 participants

Argentina

Pablo Tubaro, Museo Argentino de Ciencias Naturales

Bulgaria

Gergin Blagoev, Biodiversity Institute of Ontario

Canada

Tyler Zemplak, Biodiversity Institute of Ontario

Paul Hebert, Biodiversity Institute of Ontario

Luke Holdstock, New Gold

China

Zhaofu Yang, Haopingsi Nature Reserve Station

Costa Rica

Daniel Janzen, Area de Conservacion Guanacaste

Egypt

Hosam El-Ansary, Alexandria University

Finland

Marko Mutanen, University of Oulu

Germany

Axel Hausmann & Stefan Schmidt, Zoological State Collection (Bavaria)

Greenland

Tomas Roslin, Zackenberg Research Station

Honduras

Michelle D'Souza, Biodiversity Institute of Ontario

Malaysia

John Wilson, University of Malaysia

Mexico

Manuel Elias Gutierrez, ECOSUR

Virginia Leon Regagnon, Chamela Field Station

Ian Hogg, University of Waikato

Pakistan

Muhammad Ashfaq, National Institute of Biotechnology

Russia

Evgeny Zakharov, Biodiversity Institute of Ontario

South Africa

Hermann Staudt, Busmark 2000

United States

Becky Nichols, Great Smoky Mountains National Park

Doug Gibson, San Elijo Lagoon Conservancy

Regina Rochefort, North Cascades National Park

2014 participants

Argentina

Pablo Tubaro, Museo Argentino de Ciencias Naturales

Australia

Denis Crosbie, Cockburn Wetland Education Centre

Mark Heath, private collector

Mike Liddell, James Cook University

Tim Wardlaw, Forestry Tasmania

Bangladesh

Badrul Amin Bhuiya, University of Chittagong

Brazil

Marlucia Martins, Emilio Goeldi Museum

Canada

Dave Fraser, British Columbia Ministry of Environment

Cameroon

Thibaud Decaens, Université de Rouen

Costa Rica

Daniel Janzen, Area de Conservacion Guanacaste

Ecuador

Yves Bassett, Smithsonian Tropical Research Institute

Gabon

Thibaud Decaens, Université de Rouen

Germany

Matthias Geiger, Zoological Research Museum Alexander Koenig

Kenya

Dino Martins, Turkana Basin Institute

Margaret Kinnaird, Mpala Research Centre

Robert Copeland, ICIPE

Scott Miller, Smithsonian National Museum of Natural History

Madagascar

David Lees, University of Cambridge

Malaysia

John Wilson, University of Malaya

Panama

Hector Barios & Yves Bassett, Smithsonian Tropical Research Institute

Papua New Guinea

Vojtech Novotny, New Guinea Binatang Research Center

New Zealand

Ian Hogg, University of Waikato

Norway

Torbjorn Ekrem, Museum of Natural History Trondheim

Saudi Arabia

Jamal Sabir & Ahmed Bahieldin, King Abdulaziz University

South Africa

Hermann Staudt, Busmark 2000

USA

Bradley Zlotnick, San Diego Natural History Museum

Joshua Kohn, University of California San Diego

Scott Miller, Smithsonian National Museum of Natural History

2015 participants**Australia**

Peter Cale, Australian Landscape Trust

Costa Rica

Daniel Janzen, Area de Conservacion Guanacaste

Indonesia

Stefan Schmidt, Zoological State Collection (Bavaria)

Kenya

Margaret Kinnaird, Mpala Research Centre

Pakistan

Muhammad Ashfaq, National Institute of Biotechnology

Philippines

Eddie Mondejar, Mindanao State University

Puerto Rico

Sean Locke, University of Puerto Rico

Russia

Oleg Ermakov, Penza State University

Evgeny Zakharov, Biodiversity Institute of Ontario

Saudi Arabia

Jamal Sabir & Ahmed Bahieldin, King Abdulaziz University

Thailand

Yves Bassett, Smithsonian Tropical Research Institute

USA

Bradley Zlotnick, San Diego Natural History Museum

Joshua Kohn, University of California San Diego

Scott Miller, Smithsonian National Museum of Natural History

GLOBAL MALAISE TRAP PROGRAM DISTRIBUTION MAP

