DNA Barcode-based Assessment of Arthropod Diversity in Canada's National Parks:

Progress Report for Glacier National Park




Report prepared by the Bio-Inventory and Collections Unit,

Biodiversity Institute of Ontario, University of Guelph

December 2015

The Biodiversity Institute of Ontario at the University of Guelph is an institute dedicated to the study of biodiversity at multiple levels of biological organization, with particular emphasis placed upon the study of biodiversity at the species level. Founded in 2007, BIO is the birthplace of the field of DNA barcoding, whereby short, standardized gene sequences are used to accelerate species discovery and identification.

www.biodiversity.uoguelph.ca
Twitter handle @BIO_Outreach

International Barcode of Life Project	www.ibol.org
Canadian Centre for DNA Barcoding	www.ccdb.ca
Barcode of Life Datasystems	www.boldsystems.org
BIObus	www.biobus.ca Twitter handle @BIObus_Canada
School Malaise Trap Program	www.malaiseprogram.ca
DNA Barcoding blog	www.dna-barcoding.blogspot.ca

INTRODUCTION

The Canadian National Parks (CNP) Malaise Program, a collaboration between Parks Canada and the Biodiversity Institute of Ontario (BIO), represents a first step toward the acquisition of detailed temporal and spatial information on terrestrial arthropod communities across Canada. 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. By contrast, water quality assessments are routinely based on surveys of the species composition of freshwater invertebrates. Historically, assessments of terrestrial environments have lacked a standard protocol to derive a biotic index, and instead have generally relied 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 provides a time- and cost-efficient approach for biodiversity assessments. The present study represents a pilot phase of a long term program that will involve regular assessments of arthropod diversity at sites across Canada. To date, the CNP Malaise Program has sampled in all 43 accessible Canadian National Parks (Figure 1). The program was initiated in 2012 with the participation of 14 national parks in Central and Western Canada. In 2013, an additional 14 parks in Central and Eastern Canada were involved. Having collected from Western, Central, and Eastern Canadian National Parks, the program targeted Northern and remote National Parks in its third year (Figure 2). While only one Malaise trap was deployed in each park in 2012, two Malaise traps were deployed in 2013 and 2014 to increase overall specimen catch.

Due to the isolation and inaccessibility of some regions, Parks Canada staff and researcher volunteers facilitated the program by deploying and servicing the traps during their short field seasons. In addition, BIO conducted the CNP Malaise Program for some Southern Parks that were not previously sampled and also revisited several Parks to augment past collections. Weekly samples were preserved in 95% ethanol and then held at -20°C. At the end of the season, samples were shipped to BIO for analysis.

The trap samples were accessioned, specimens were identified to order, arrayed, labeled, databased, and tissue-sampled for genetic analysis (Figure 3). All arthropods were barcoded, with the exception of a few very common species (e.g., honeybee) where only a limited number of individuals from each trap sample were analyzed. Standard barcoding protocols (http://ccdb.ca/resources.php) were followed to recover the barcode region of the cytochrome *c* oxidase I (COI) gene. The barcode sequences, specimen images and collateral data

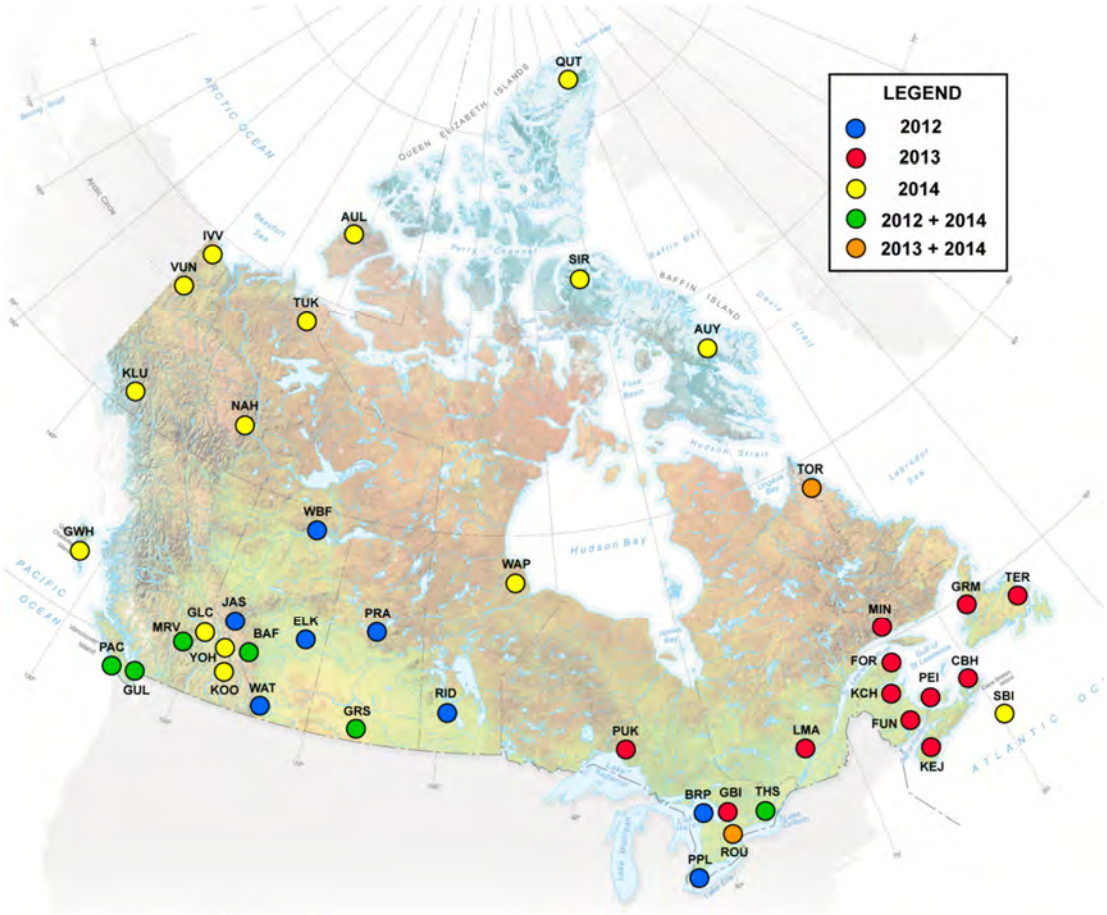

Figure 1. CNP Malaise Program sampling sites from 2012-2014.

2012	
BRP	Bruce Peninsula National Park
ELK	Elk Island National Park
JAS	Jasper National Park
PPL	Point Pelee National Park
PRA	Prince Albert National Park
RID	Riding Mountain National Park
WAT	Waterton Lakes National Park
WBF	Wood Buffalo National Park

2014	
AUL	Aulavik National Park
AUY	Auyuittuq National Park
GLC	Glacier National Park
GWH	Gwaii Haanas National Park
IVV	Ivvavik National Park
KLU	Kluane National Park
KOO	Kootenay National Park
NAH	Nahanni National Park
QUT	Quttinirpaaq National Park
SBI	Sable Island National Park
SIR	Sirmilik National Park
TUK	Tuktuk Nogait National Park
VUN	Vuntut National Park
WAP	Wapusk National Park
YOH	Yoho National Park

2013	
CBH	Cape Breton Highlands National Park
FUN	Fundy National Park
FOR	Forillon National Park
GBI	Georgian Bay Islands National Park
GRM	Gros Morne National Park
KCH	Kouchibouguac National Park
KEJ	Kejimkujik National Park
LMA	La Mauricie National Park
MIN	Mingan Archipelago National Park
PEI	Prince Edward Island National Park
PUK	Pukaskwa National Park
TER	Terra Nova National Park

2012 + 2014	
BAF	Banff National Park
GRS	Grasslands National Park
GUL	Gulf Islands National Park
MRV	Mount Revelstoke National Park
PAC	Pacific Rim National Park
THS	Thousand Islands National Park

2013 + 2014	
ROU	Rouge National Park
TOR	Torngat Mountains National Park

are stored in the Barcode of Life Data Systems (BOLD; www.boldsystems.org). The project is publicly available in the 'Canadian National Parks 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 park and facilitating comparisons among them.

A key question concerning this program relates to whether Malaise traps are the most effective method of capturing local arthropods. BIO is exploring this issue through a Standardized Sampling investigation in a subset of parks. In the selected parks, three sites were chosen and five standard collecting techniques were employed at each locality: Malaise, pan, pitfall, Berlese and flight-intercept traps, as well as sweep-netting. Each park was sampled by the BIObus staff for a one-week interval before the team proceeded to the next park with this weekly rotation continuing throughout the summer. All specimens collected with the different sampling methods were barcoded to permit a comparison among methods.

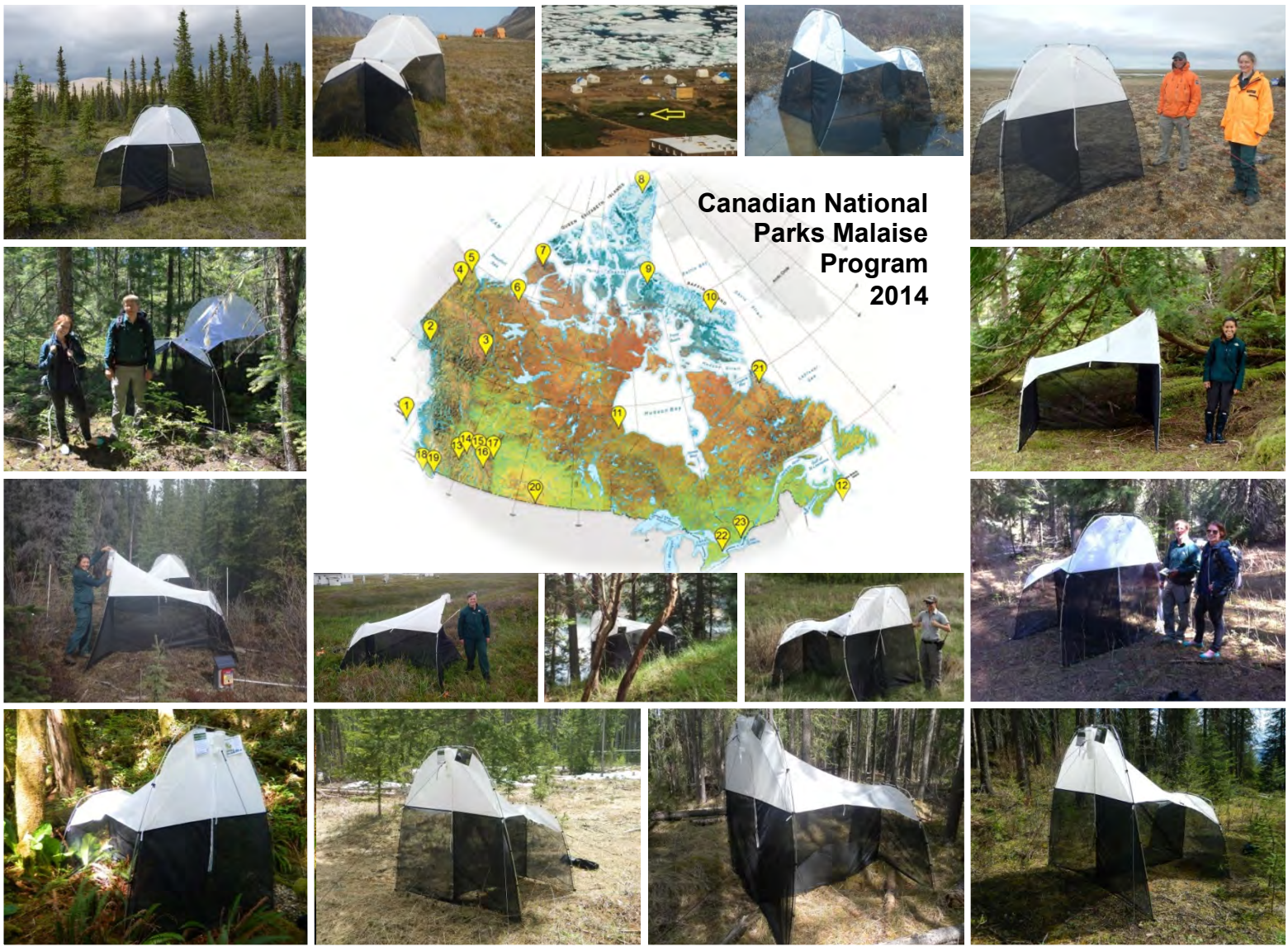

Figure 2. Sampling locations at 16 of the 23 Canadian National Parks surveyed in 2014.

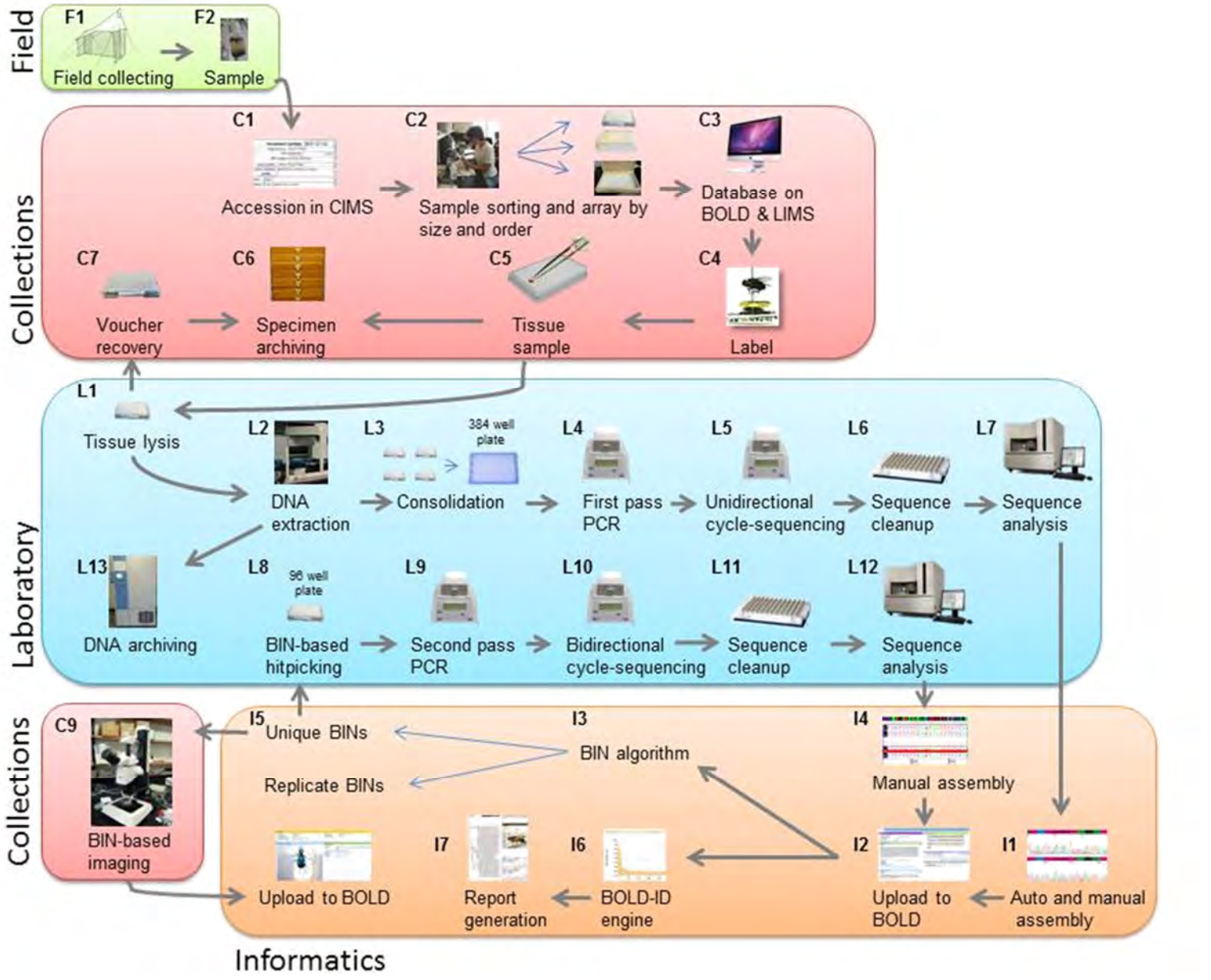

Figure 3. Schematic diagram showing the specimen workflow. Front end processing begins with field collecting (F1) and proceeds through to archiving of specimens (C6). Laboratory analysis begins with tissue lysis (L1) through to sequence analysis (L12). The informatics workflow includes both manual (I4) and auto sequence assembly, and finishes with BIN assignments and subsequent imaging of each BIN (C9).

2012-2014: RESULTS FOR 43 NATIONAL PARKS

The barcode analysis of all Malaise trap samples from 2014 was completed by fall 2015. In total, 339 weekly samples and nearly 295K specimens were analyzed. A total of 254,323 specimens generated barcode sequences that were long enough to allow a BIN assignment. Their analysis revealed a total of 30,335 BINs from the 2014 collection.

In combination with the 2012 and 2013 samples, the CNP Malaise program has collected over 725K specimens to date. The average sequence success rate was 90% which led to 619,995 records with sequences long enough for a BIN assignment. A total of 36,423 BINs were revealed while the Chao 1 (Magurran, 2003) species estimate for the total number of BINs that would be encountered with complete sampling using this method would be 51,722 (Figure 4).

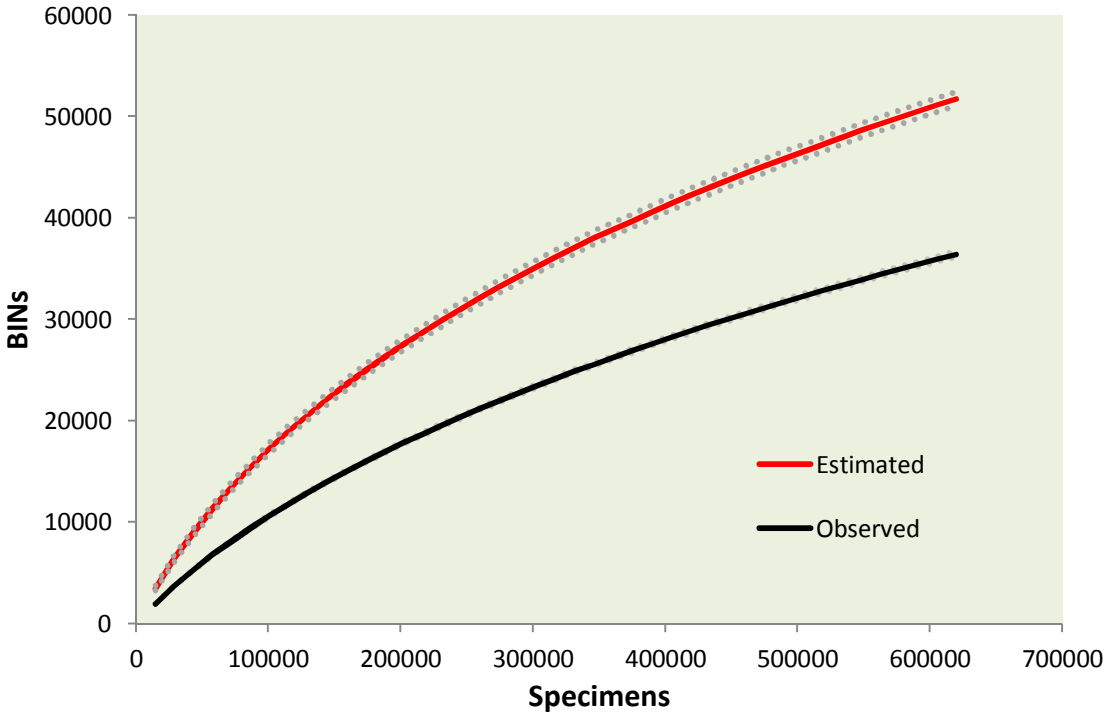

Figure 4. BIN accumulation curve for the 746 Malaise trap samples collected in 43 Canadian National Parks during 2012-2014.

The usual 'hollow curve' species abundance pattern was observed, with 12,942 species represented by just a single individual (singletons) (Figure 5). By comparison, just 1062 BINs were represented by 100 or more individuals. The most commonly encountered species was *Smittia sp.* – a non-biting midge belonging to the family Chironomidae – with 7243 individuals sequenced. Species richness extrapolation using the lognormal species abundance distribution (Preston, 1962) suggests that nearly twice as many BINs exist in these 43 National Parks (61,760 BINs) as were collected. Despite the discrepancy between the two methods of estimating species richness (Chao and Preston), both results suggest that a considerable fraction of the species still awaits collection.

The average number of specimens collected per weekly sample was 1713. Only 14 of the 24 parks sampling in 2014 collected at least 10 weekly samples. Of these, the total number of individuals captured varied from a low of 6227 from 20 samples at Pacific Rim National Park to a high of 36,274 specimens from 10 analyzed samples in Grasslands National Park. Sequencing success also varied among parks, from a low of 83.7% at Wapusk National Park (20,224 barcode records from 20,941 specimens), versus 98.5% for Auyuittuq National Park. The number of BINs detected ranged from a low of 78 at Auyuittuq to a high of 3795 at Thousand Islands (Figure 7).

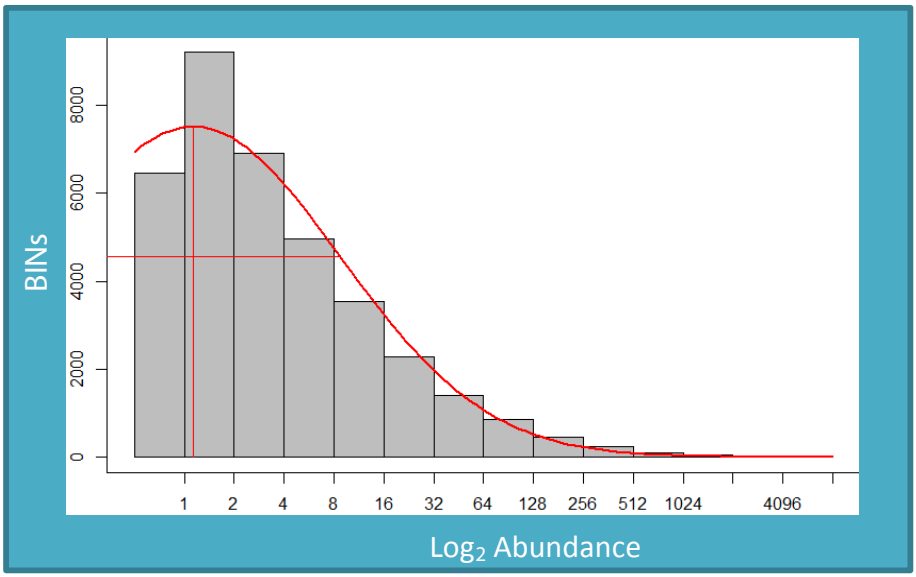

Figure 5. Lognormal species abundance curve, showing the total BINs within each log $_2$ abundance frequency interval (Preston, 1962).

These results are comparable to the BIN accumulation curves observed at Malaise collecting sites that are part of BIO's Global Malaise Program (Figure 6). The total BIN richness in each National Park (mean = 1319 BINs) is generally less than those of highly diverse global sites (e.g. Argentina, Costa Rica). However, the slopes of the accumulation curves suggest that Malaise traps enable us to survey biodiversity at comparable rates across a range of biomes.

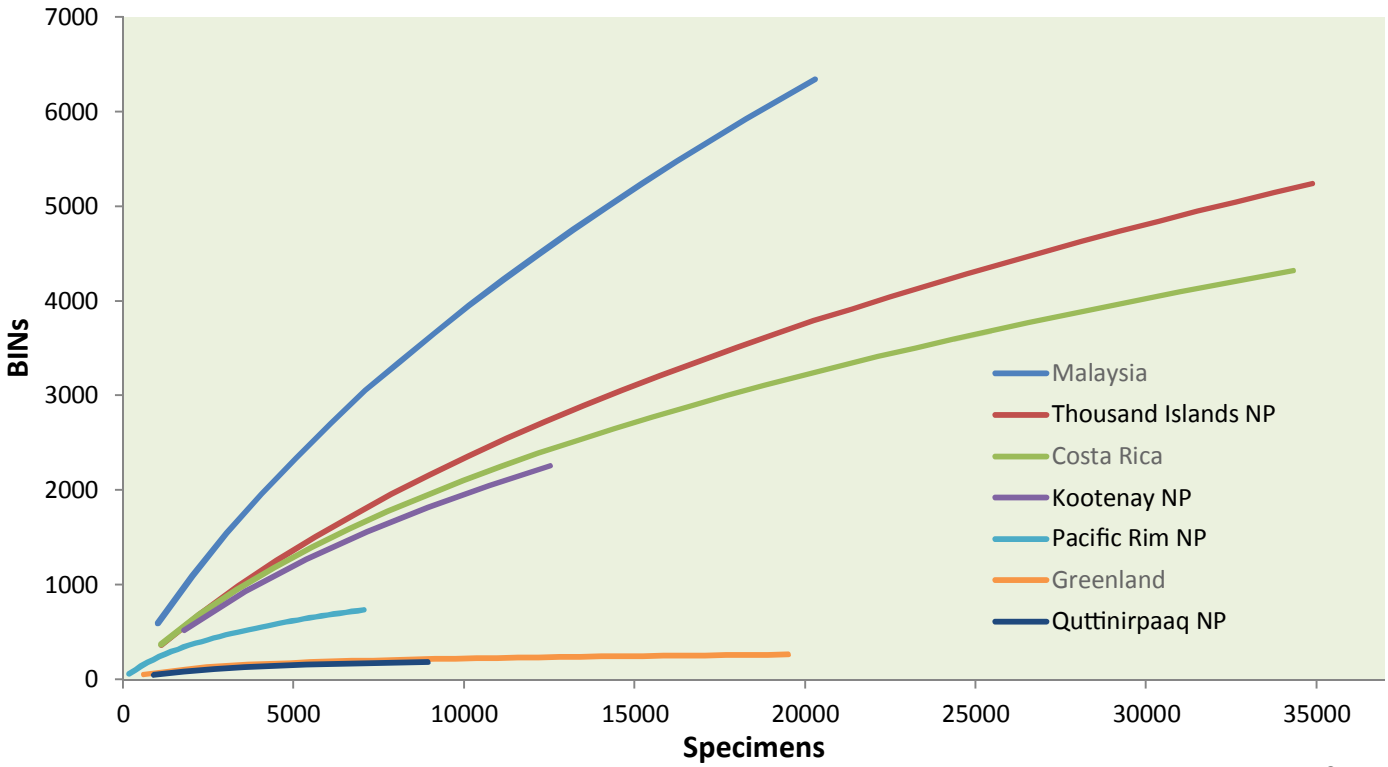

Figure 6. Comparison of BIN accumulation curves for 127 Malaise samples collected from 7 different sampling sites. Grey text indicates Global Malaise locations.

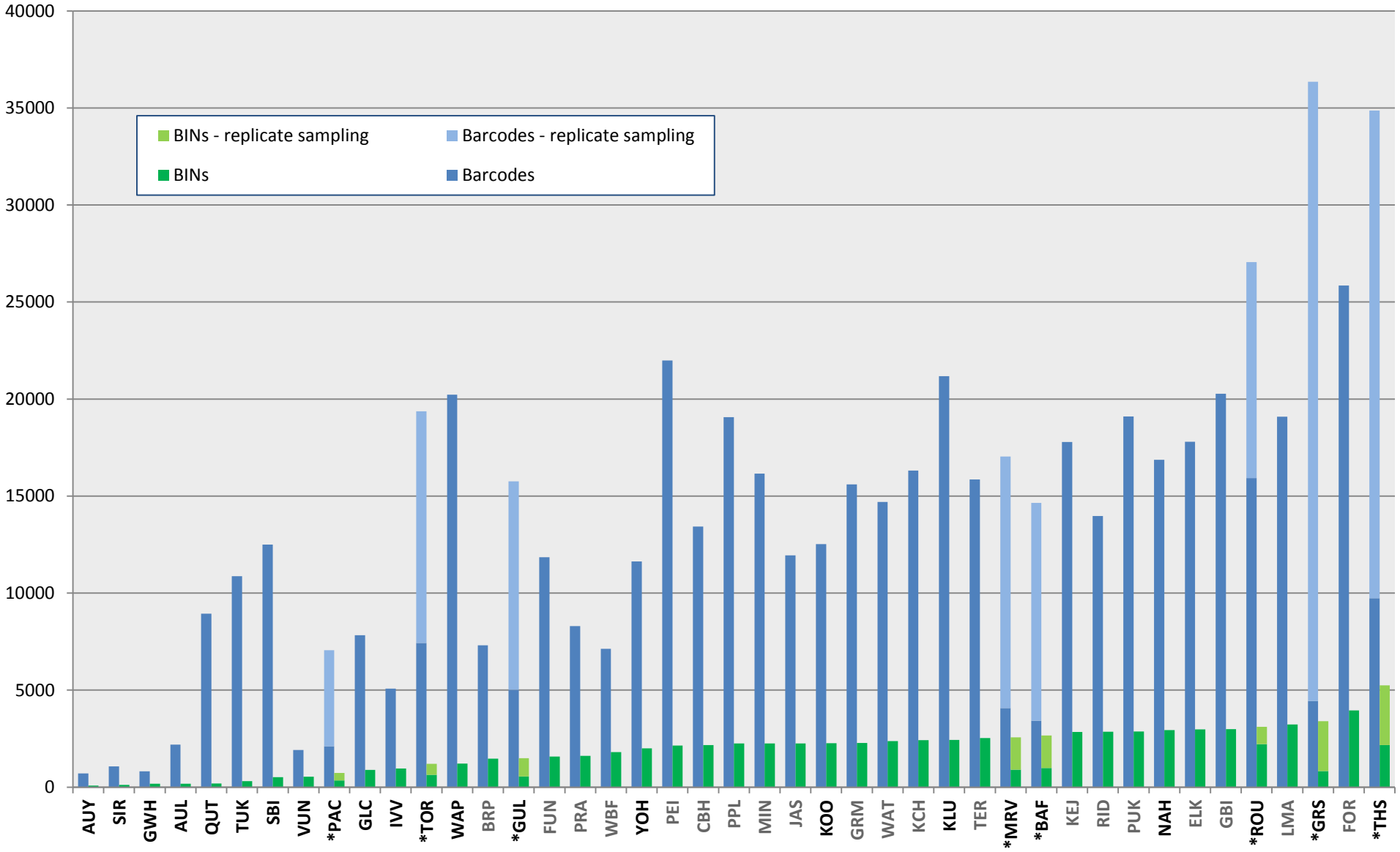

Figure 7. Total sequences and number of BINs generated from each of the 43 parks; grey text indicates 2012/2013 sampling (*repeated parks).

When analyzing all CNP Malaise Programs from 2012-2014, the number of BINs detected in each park was strongly influenced by sample size (Figure 8, R^2 = 0.6873, $p<0.05$). With 29K specimens analyzed, Thousand Islands National Park displayed the highest BIN count (N = 3795), while Auyuittuq National Park captured less than 100 BINs from 727 specimens.

Of the 36K BINs captured, more than half were unique to a single collection site; i.e. 21,072 BINs occurred in only one of the 43 parks. The number of BINs unique to each park varied (Figure 9). Thousand Islands National Park exhibited the highest count of unique BINs with 35% of its BINs being unique to the site (1851 BINs of 5244). Despite having lower BIN counts, northern parks exhibited fairly average ratios of unique BINs to BINs captured (for example, Auyuittuq 23%, Sirmilik 23%, Kluane 27%, and Nahanni 28%). Grasslands had the highest proportion of unique BINs with 51% of its BINs being unique. In contrast, Fundy National Park, with 203 unique BINs, had the lowest ratio of unique BINs to BINs captured, only 13%. Insular parks, such as Sable Island and Gwaii Haanas, also displayed high percentages of unique BINs. This indicates considerably high diversity despite the perception of low diversity given current sampling efforts.

The similarity in species composition between parks showed marked variation (Figure 10). For example, Gros Morne and Terra Nova National Parks – 301km apart – shared the highest proportion of BINs (938 shared species), with a Chao's Sorenson Similarity index (Chao et al., 2005) of 0.39.

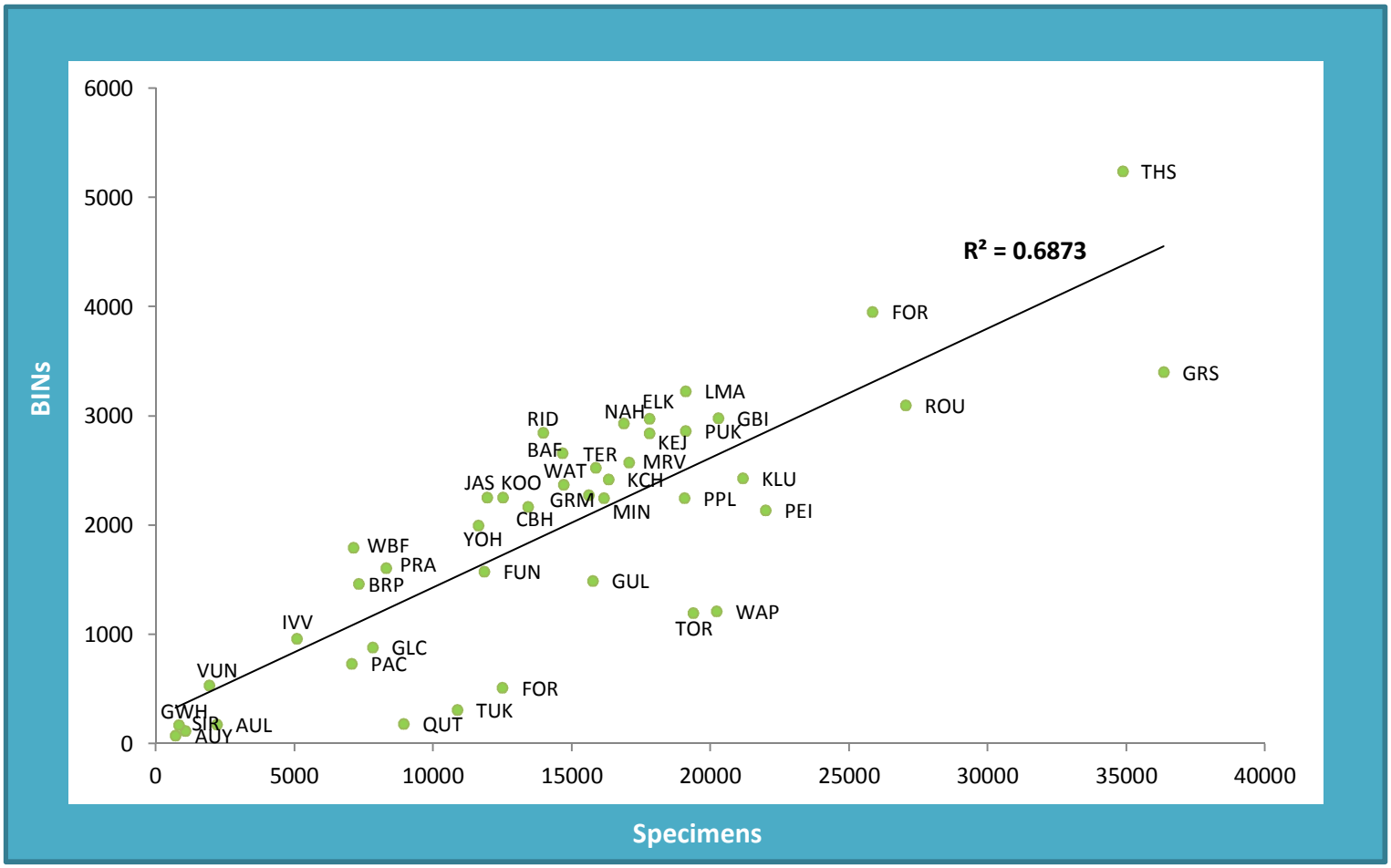

Figure 8. Regression analysis examining the relationship between the number of barcoded specimens and number of BINs (BINs = 0.1187(Specimens) + 243.18).

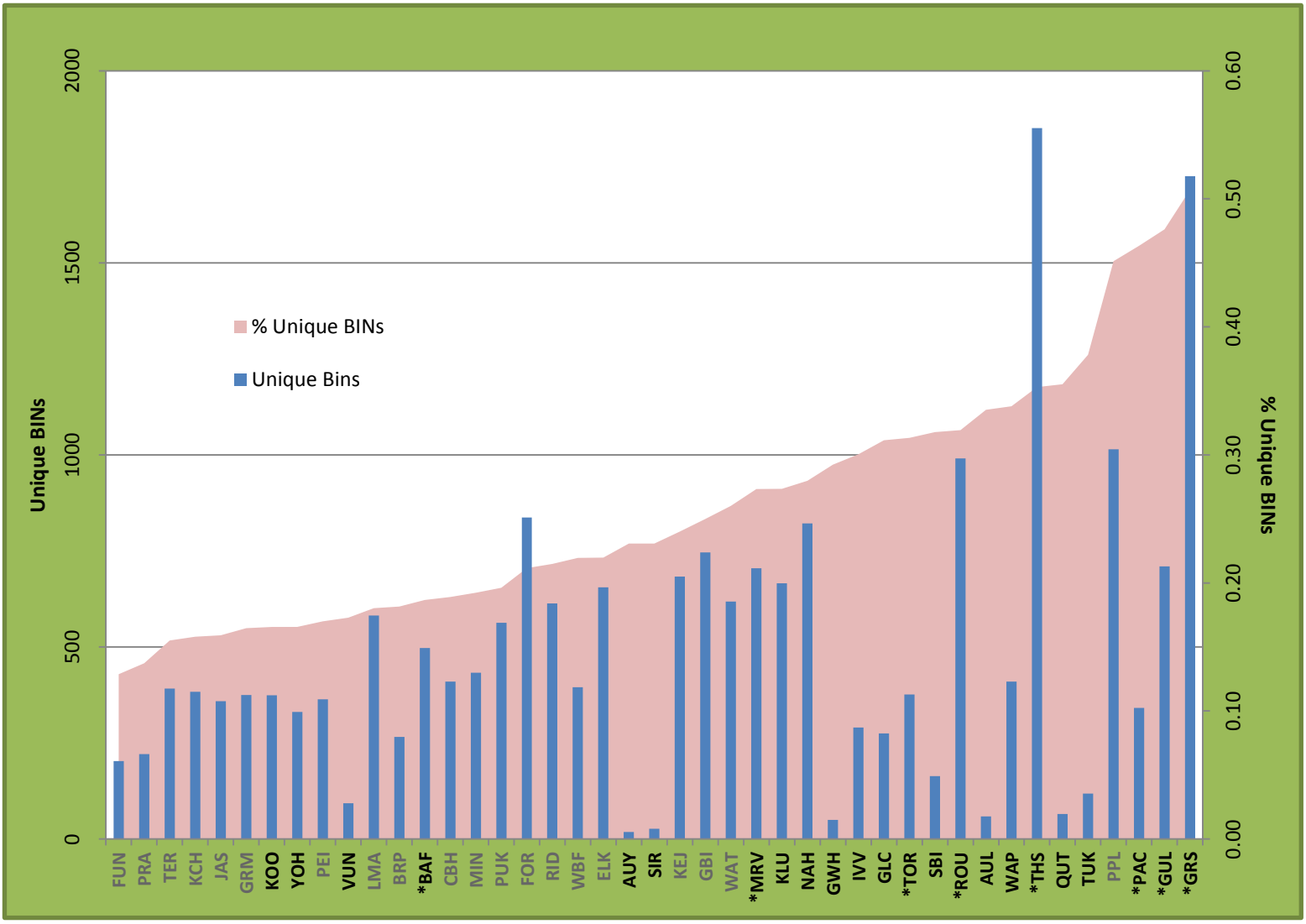

Figure 9. Total number of BINs unique to each park (bars) and the percentage of unique BINs collected in each park (Unique BINs/Total BINs); grey text indicates 2012/2013 sampling (*repeated parks).

The two parks closest together in geographic distance (only 45.7km apart), Banff and Kootenay, had a similarity index of 0.33. Surprisingly, the two parks furthest apart with a distance of 5228km, Kluane and Sable Island, still had a similarity index value of 0.02. Meanwhile, 19 other park-pairings shared no similar species (similarity index = 0.00). Although similarity indices were still shown to be negatively correlated to geographic distance between parks (Figure 11, $p<0.05$), individual parks did not necessarily share the highest species with their nearest geographic neighbour (Figure 12). It is interesting to note that Torngat Mountains shared the most species with Wapusk rather than Mingan Archipelago, which is 844km closer. This likely reflects the similar habitats and elevations between the two locations despite being further apart. It is tempting to make a similar hypothesis for the similarity index between Sable Island and Wood Buffalo. However, the more likely scenario is that the small sample size from Sable Island is the cause for this relationship. Collecting more of the diversity through repeated sampling in Sable Island may lead to the discovery of new BINs that are shared with parks closer in geographic distance than Wood Buffalo. Further sampling must be done to test either hypothesis.

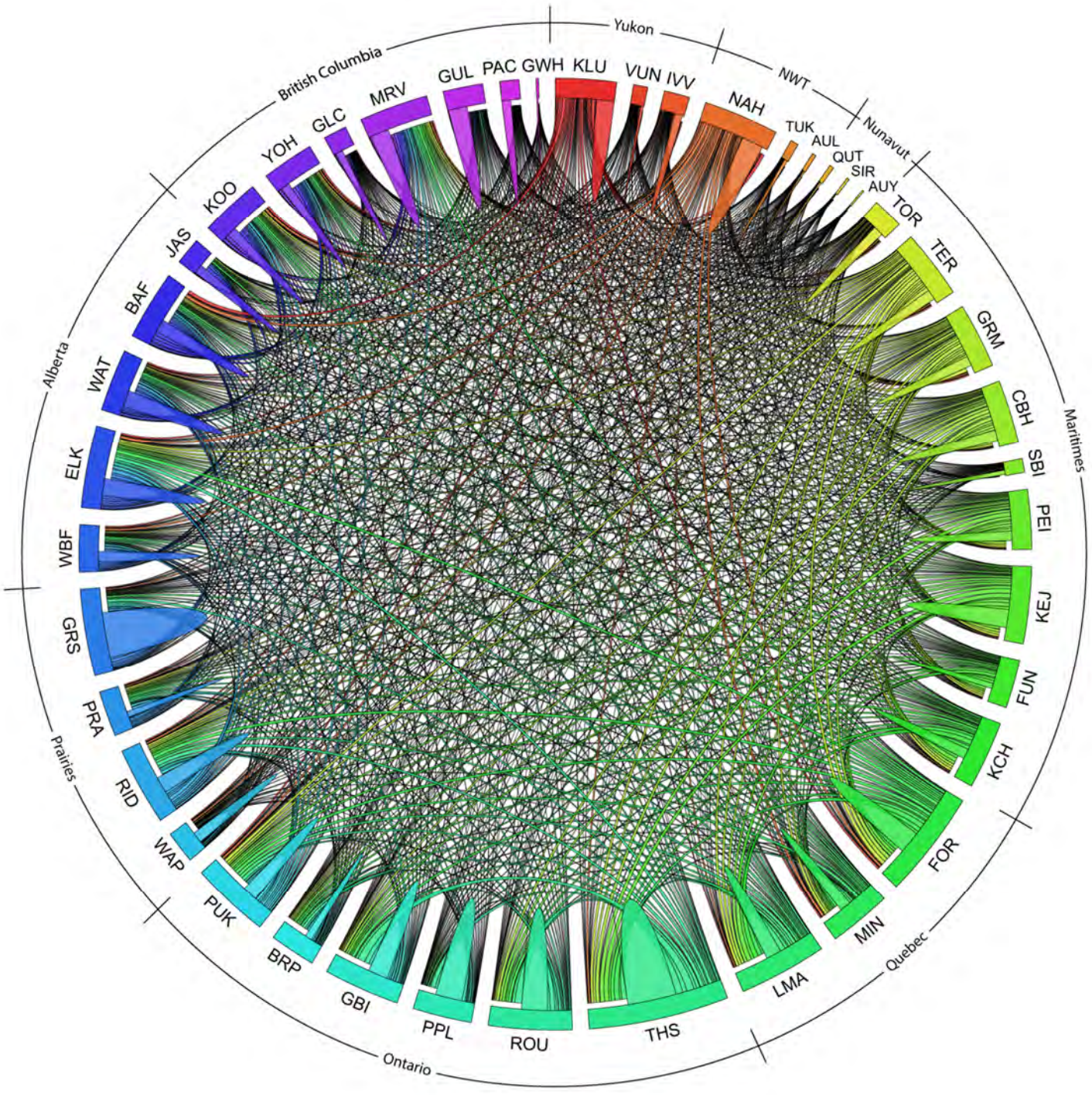

Figure 10. Chord diagram of species overlap between all 43 National Parks, organized by provinces and territories. The width of each wedge reflects the number of BINs captured in each park relative to the others. The widths of internal humps are proportional to the unique BINs within each park. Arcs connecting the parks reflect the proportion of shared species between any two parks, but have been scaled to account for BINs which are found in more than just two parks such that their widths are not directly proportional to the number of shared.

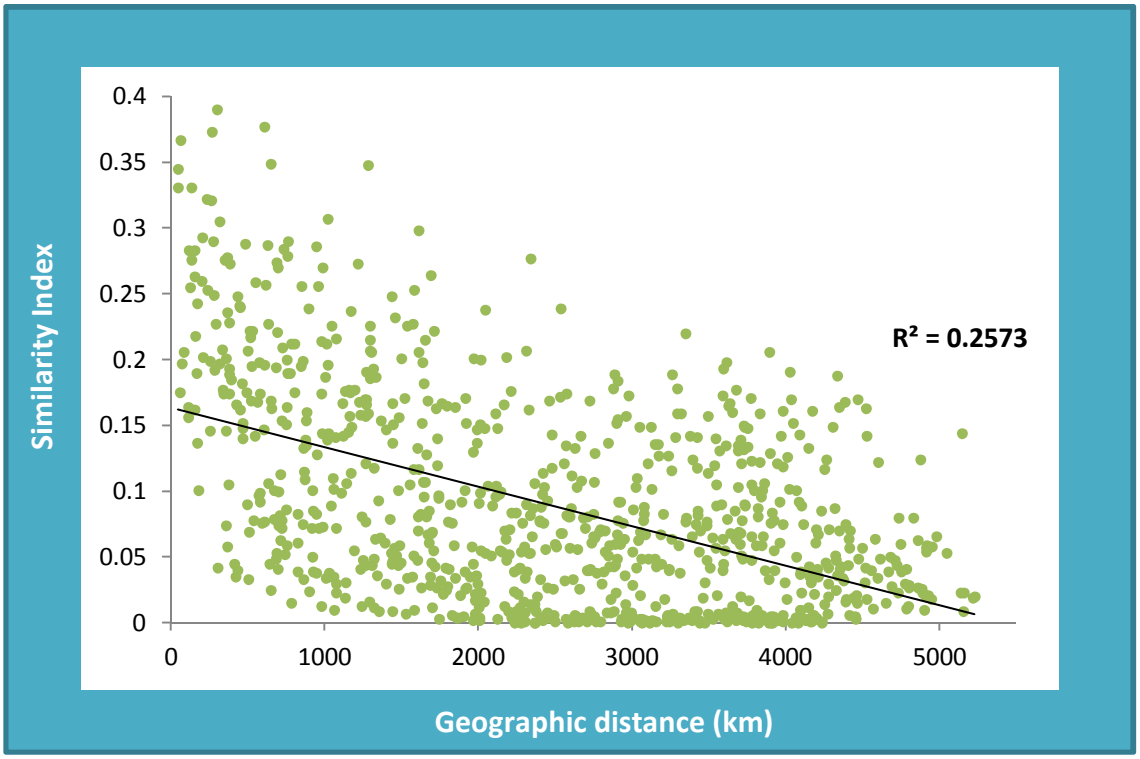

Figure 11. The relationship between geographic distance and species similarity. Similarity is based on Chao-Sorensen Raw Abundance data; each point represents a pair of locations. There is a significant negative correlation between the two variables ($p<0.05$).

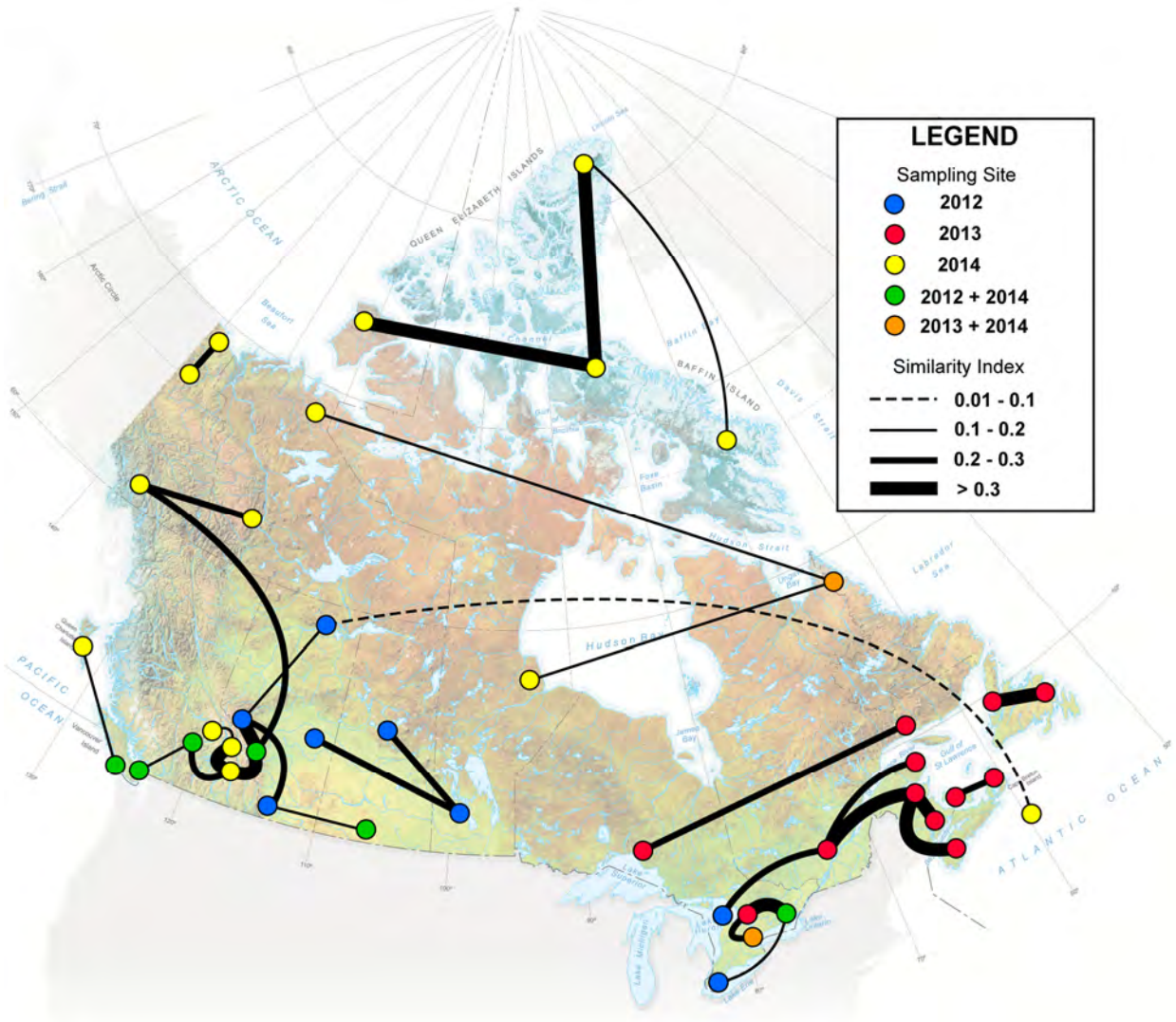

Figure 12. Map indicating the highest similarity index values for each park.

The Standardized Sampling Program was executed in a total of 23 national parks from 2012-2014 (Figure 13a). The processing of collected samples is still ongoing; however, preliminary results from a total of 17 standardized sampling weeks from 12 national parks are available. To date, the program has captured a total of 310,327 specimens from 42 sites after 994 individual collection events. The number of barcodes collected so far is 231,942 leading to the generation of 19,887 BINs. When comparing this to the CNP dataset, 12,207 of these BINs are shared between the two programs and the total number of BINs combined is 44,106.

Malaise traps captured more specimens ($p<0.001$) than the other trap types (Figure 13b), revealing a significantly higher proportion of the local fauna (29% of total BINs, and 39% of unique BINs). Moreover, collector effort varied drastically between methods, with Malaise traps capturing the most specimens, BINs, and unique BINs per unit of time ($p<0.05$). On the other hand, even though sweep netting appears to capture a high volume of specimens, it requires 15 times more effort than Malaise traps to be comparable.

The taxonomic diversity captured with each method also varied (Figure 14). As expected, Malaise, intercept, and pan traps captured more flying insects (flies, wasps, bees) while pitfalls and Berlese funnels captured more soil arthropods such as beetles and mites.

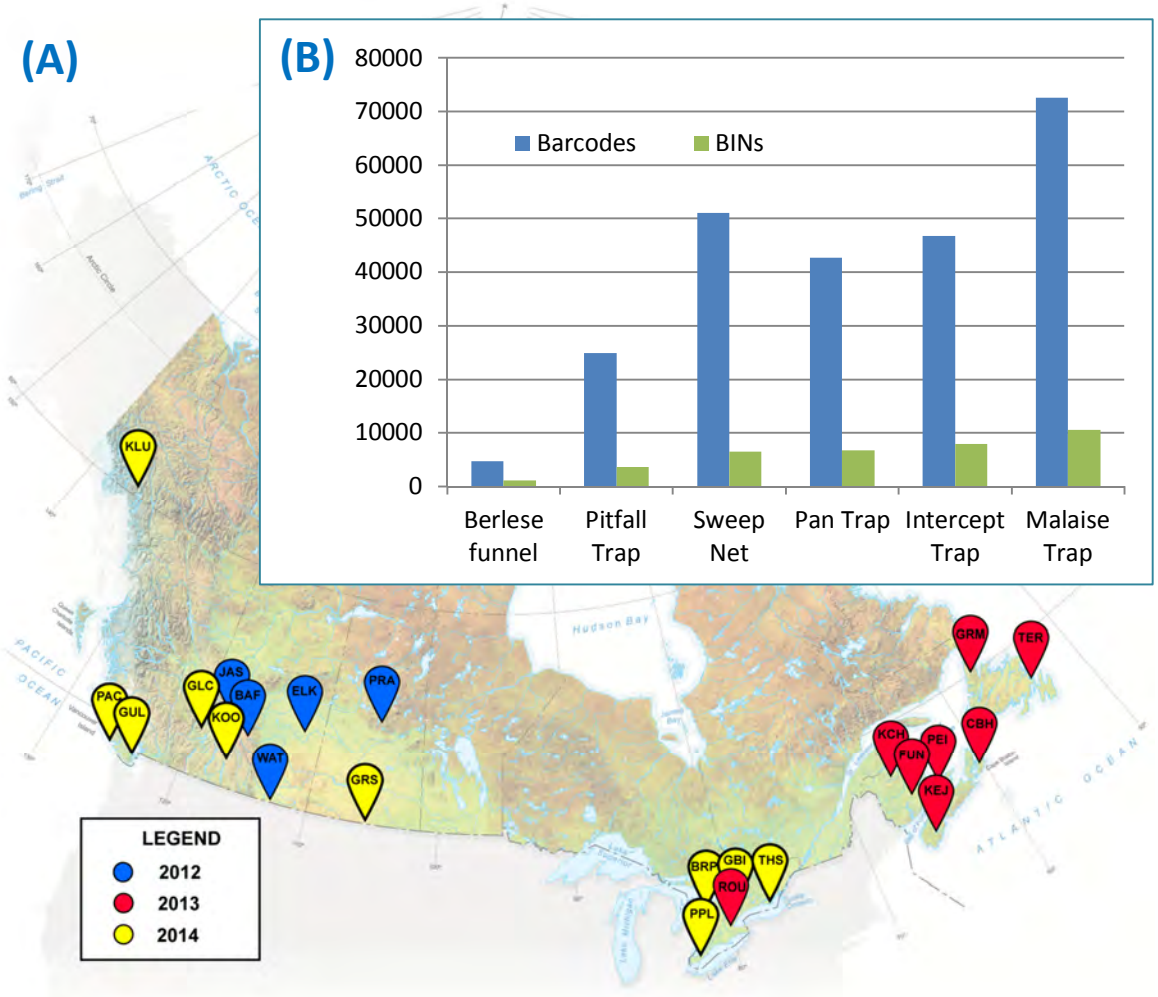

Figure 13. (A) Map indicating the parks where standardized sampling was conducted over 3 years. **(B)** Total number of sequences and BINs captured using different collecting methods.

Of the 19K BINs captured, more than half were unique to a single sampling method (N= 11,638). The number of BINs unique to each method varied and the majority of unique BINs were captured in Malaise traps (Figure 14). It is important to note that although Berlese funnels collected the fewest unique BINs (N = 592), this method had the highest ratio of unique BINs to BINs captured (52%). The majority of these BINs belong to specimens from the subclass Acari (mites) and a large number of Collembola (springtail) BINs as well.

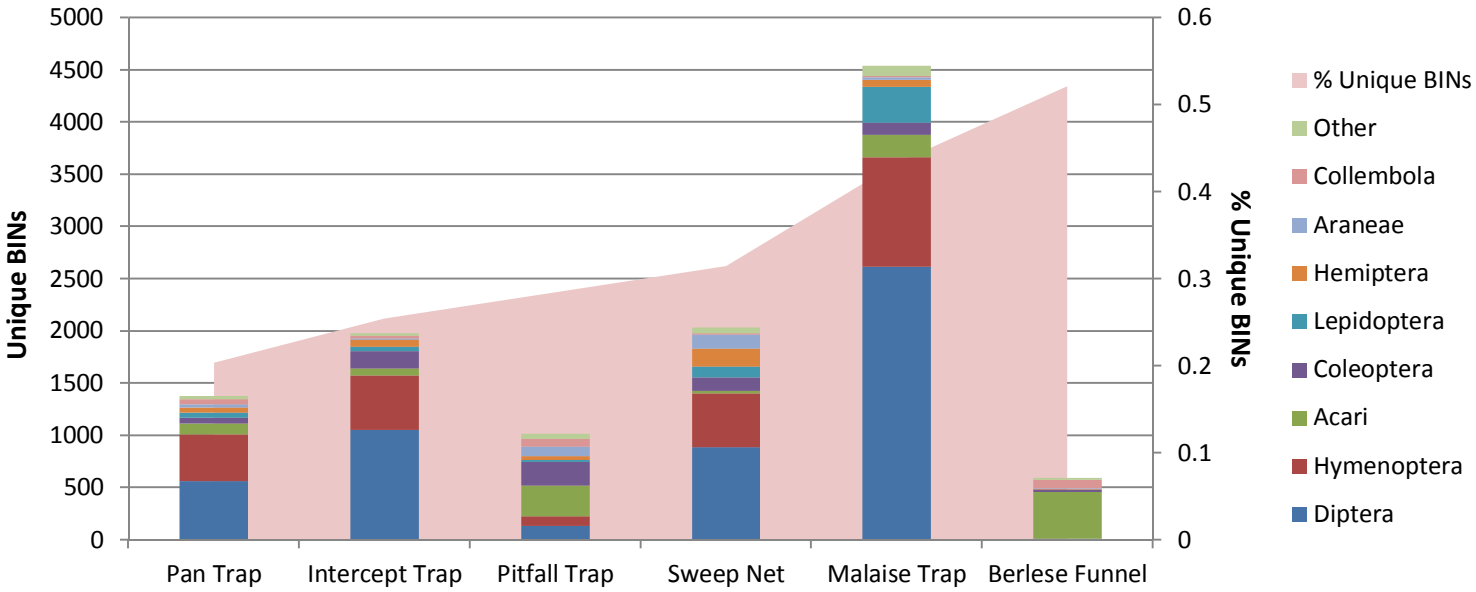

Figure 14. Total number of BINs unique to each collecting method and their taxonomic breakdown (bars) and the percentage of unique BINs collected with each method (Unique BINs/Total BINs).

The diversity of species collected by Malaise traps is impressive. The combined results from CNP Malaise Programs 2012-2014 included representatives for 36,423 BINs from 726,606 total specimens collected from Malaise traps in 43 Canadian National Parks. This BIN count represents 113% of the total number (N = 32,045) of terrestrial arthropod species recorded in all prior taxonomic studies, and 57.2% of the estimated total number of terrestrial arthropod species (N = 63,643) found in Canada (Mosquin et al. 1995).

BIO is edging closer to a comprehensive dataset to estimate alpha and beta diversity of the terrestrial arthropod fauna in our National Parks. Simultaneously, it is constructing the barcode reference library to rapidly and accurately re-identify those species – a critical first step towards a terrestrial biotic index for Canada. The next step involves sampling diverse environments and disturbance regimes, as well as to examine replicate samples. We expect to then be able to link the condition of the environment with attributes of the community composition (for instance, the diversity of rare, indicator, pest, pioneer, and/or exotic species). As our reference barcode library for Canadian arthropods matures, the ability to conduct comprehensive terrestrial diversity assessments will strengthen. Ultimately, this will allow the calculation of a standardized terrestrial biotic index that can assist with determining how to balance ecological benefits with economic benefits associated with land management practices.

2014 RESULTS – GLACIER NATIONAL PARK

Two Malaise traps were deployed approximately 10m apart in a coniferous forest near a stream in the Rogers Pass Compound at Glacier National Park (51.300933N 117.5219W, 1341m ASL, Figure 15). These traps collected arthropods weekly from June 6th to September 30th 2014. Ten Malaise trap samples were analyzed which contained a range of 477 to 1581 specimens each. A total of 9224 specimens were captured, and a barcode recovery rate of 89.3% was observed (Appendix 1-3).

Figure 15. Parks Canada staff deploying a Malaise trap at Glacier National Park in 2014.

Most of the specimens collected were flies (Diptera), followed in abundance by mites (Acari), bees, ants and wasps (Hymenoptera), book lice (Psocoptera), and springtails (Collembola; Figure 16). A total of 883 BINs were observed and the Chao species estimate suggests that approximately 1831 BINs are present in the park and could be collected with this method if sampling effort was extended (Chao et al., 2005; Figure 17).

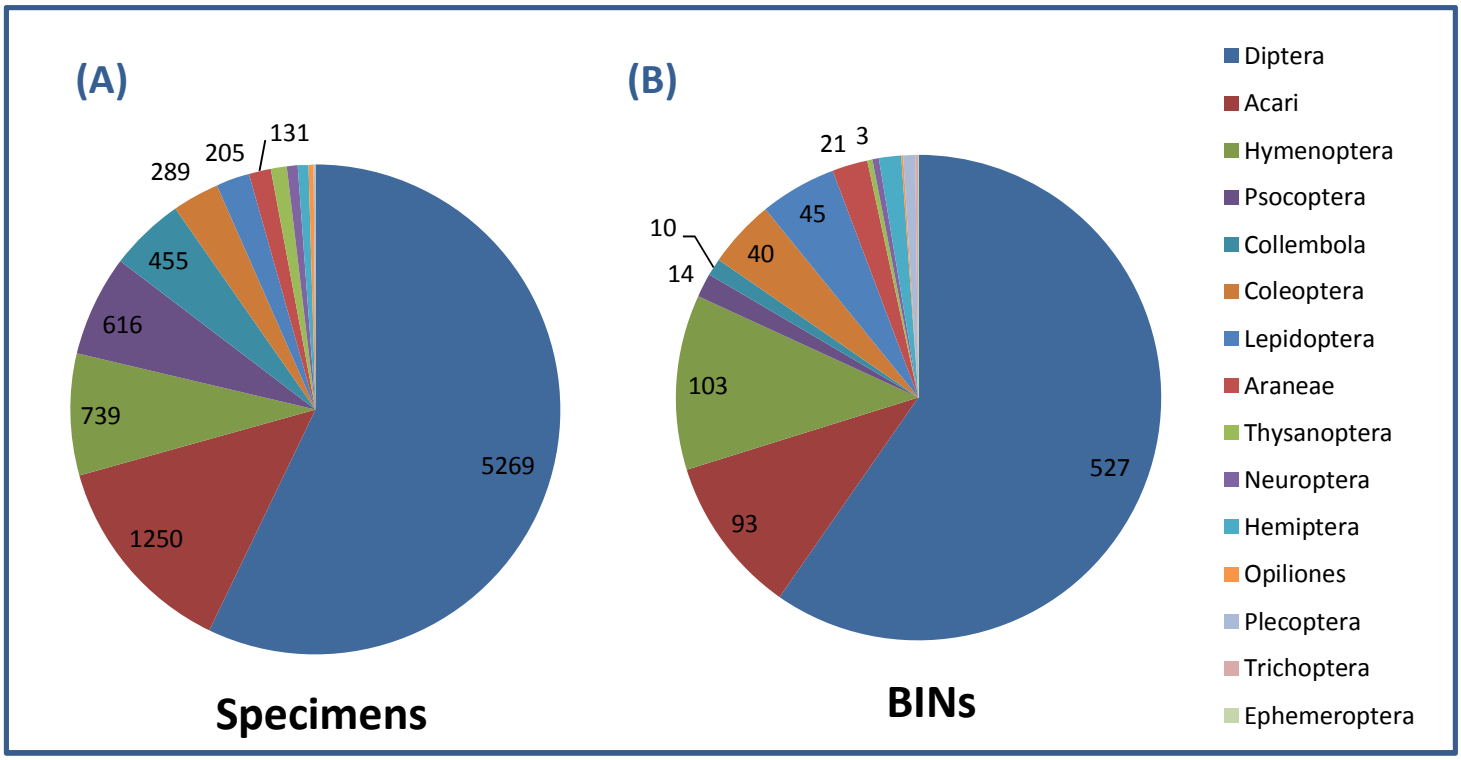

Figure 16. Taxonomic breakdown of (A) 9224 total specimens and (B) 883 total BINs collected by Malaise traps at Glacier National Park in 2014.

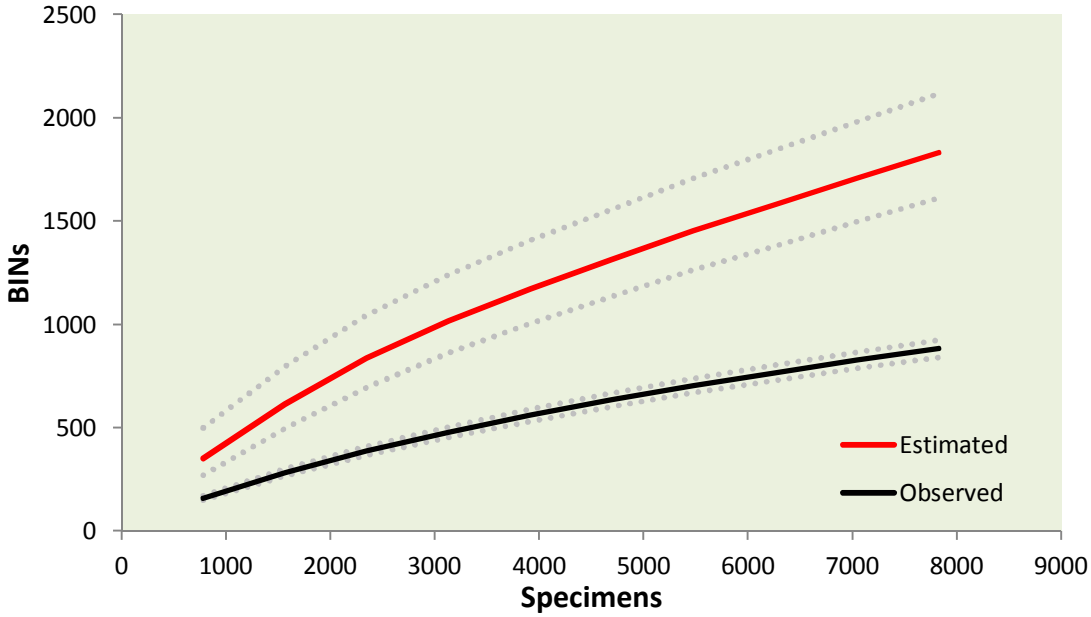

Figure 17. BIN accumulation curve for all specimens collected by the Malaise traps at Glacier National Park in 2014.

In total, 153 arthropod species were named, representing 18% of the BINs from the park (Appendix 4). 99% of BINs were assigned at least to family, and 35% of the BINs were assigned to a genus. Specimens collected from Glacier represent 141 different families and 209 genera. Appendix 4 provides a complete list of specimens with available taxonomy and collection information. It is important to emphasize that it will be possible to identify many of the taxa which currently lack a species name as the barcode reference library becomes more complete.

The pattern of relative species abundance is quite typical, with a few species represented by many individuals (12 species with >100 individuals) – including 1096 individuals of a species of non-biting midge (Diptera: Chironomidae) – and a large number of species with few individuals (467 singletons; Figure 18). Species richness extrapolation using the lognormal species abundance distribution suggests that 3879 BINs exist in the park (Preston, 1962).

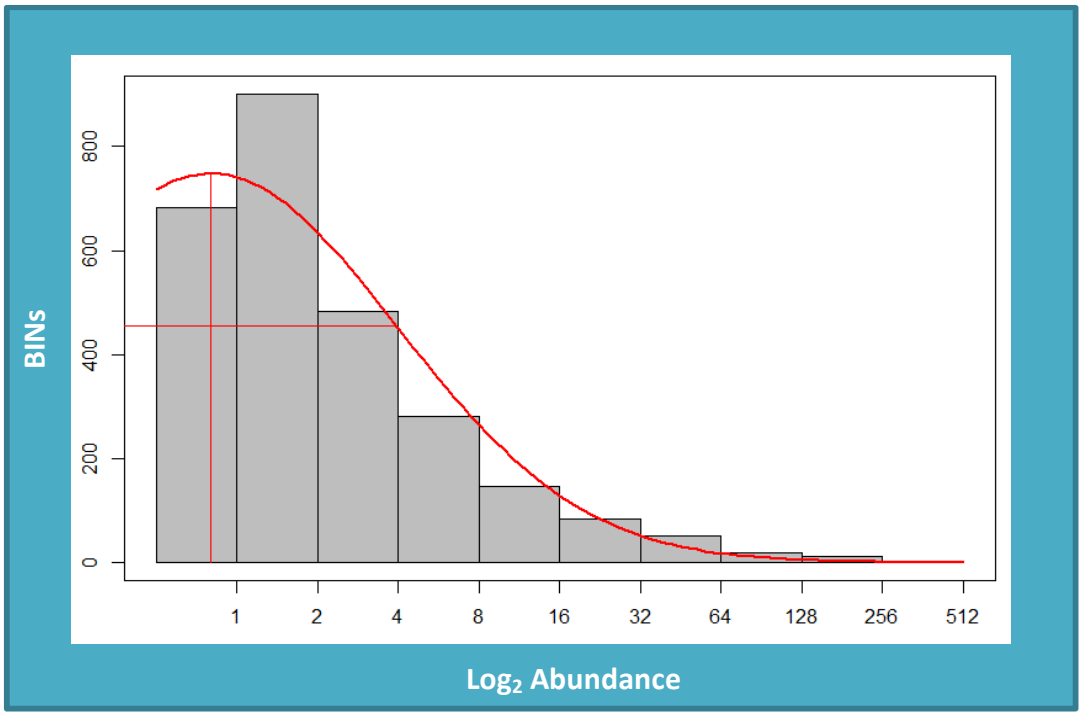

Figure 18. Lognormal species abundance curve, showing the total BINs within each log $_2$ abundance frequency interval (Preston, 1962).

Standardized sampling was also carried out at Glacier National Park in 2014. BIO is currently working to process the 41 collected lot samples. To date, 16,522 individuals have been sorted from 3 sampling sites. The taxonomic breakdown based on collection method is shown in Figure 19. Analysis of current data revealed 1415 BINs so far. It is interesting to note that in these particular sites, the Malaise trap does not appear to be the most effective collection technique. This gives more incentive to sample using other methods when collecting arthropods. These specimens will help supplement the species inventory obtained through the CNP Malaise Program for a more complete picture of the arthropod diversity within the park.

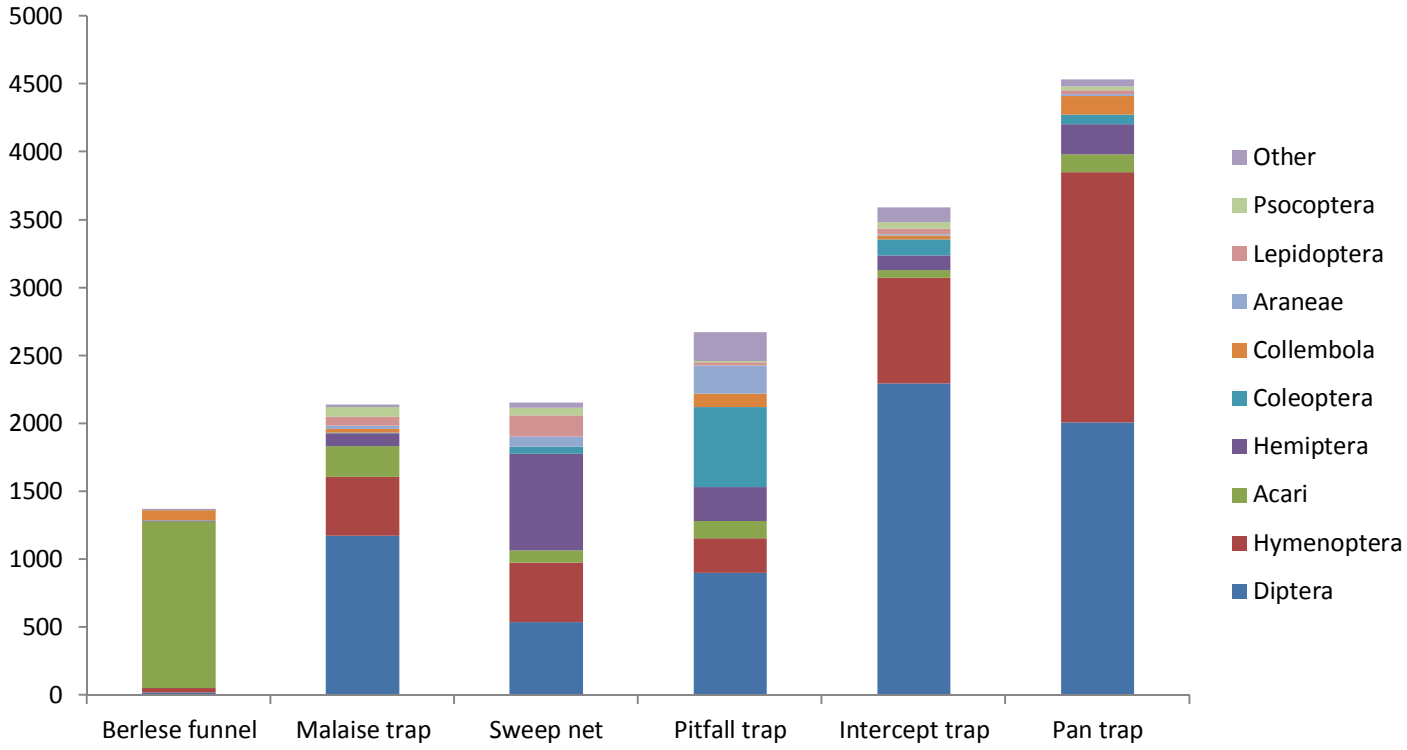

Figure 19. Total number of specimens sorted out so far from the standardized sampling collections at Glacier National Park in 2014. Colours indicate the taxonomic breakdown captured using different collecting methods.

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.

Magurran, A.E. (2003). Measuring Biological Diversity. Wiley-Blackwell, Malden, Massachusetts. 264 pp.

Mosquin, T., P.G. Whiting and D.E. McAllister (1995). Canada's Biodiversity: The Variety of Life, its Status, Economic Benefits, Conservation Costs and Unmet Needs. Canadian Centre for Biodiversity, Canadian Museum of Nature, Ottawa, Ontario. 293 pp.

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.

APPENDICES

Appendix 1. Neighbour-joining tree of representative specimens from each BIN collected by the Malaise traps deployed at Glacier National Park in 2014 (colourized based on Taxonomic Order).

Appendix 2. Image library of 868 (out of 878) BIN representatives collected in Glacier National Park (in alignment with Appendix 1).

Appendix 3. Taxonomy report for Glacier National Park.

Appendix 4. Complete data spreadsheet of all specimens collected from Glacier National Park through the CNP Malaise Program 2014 with available taxonomy and collection information.

ACKNOWLEDGEMENTS

We are grateful for the overwhelming support of many people who have enabled the successful execution of the Canadian National Malaise Program over the past three years. Special thanks to Parks Canada for their partnership in this program. Additionally, this project would not have been possible without the indispensable staff at the Biodiversity Institute of Ontario, and the many students who embarked on the cross country BIObus journeys.

Contacts:

Jeremy deWaard, PhD
Director of Collections
dewaardj@uoguelph.ca

Kate Perez, MSc
Field Operations Lead
kperez@uoguelph.ca

BIO Biodiversity Institute of Ontario

Biodiversity Institute of Ontario

University of Guelph,
50 Stone Road East,
Guelph, ON, Canada N1G 2W1
519-824-4120 x56393
Fax: 519-824-5703

Canada
OntarioGenomicsInstitute
GenomeCanada
NSERC CRSNG
Ontario Ministry of Research and Innovation
INNOVATION.CA
Canada Foundation for Innovation | Fondation canadienne pour l'innovation