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

Progress Report for Cape Breton Highlands National Park

Report prepared by the Bio-Inventory and Collections Unit,

Biodiversity Institute of Ontario, University of Guelph

December 2014

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. There are four units with complementary mandates that are housed within BIO and interact to further knowledge of biodiversity.

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
International Barcode of Life Conference 2015	www.dnabarcodes2015.org

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.

The CNP Malaise Program was initiated in 2012 with the participation of 14 national parks in Central and Western Canada. In 2013, an additional 14 parks were involved, from Rouge National Urban Park to Terra Nova National Park (Figure 1). While only one Malaise trap was deployed in each park in 2012, two Malaise traps were deployed (within ten metres of each other) in 2013 to increase overall specimen catch.

The two Malaise traps were deployed by BIO staff in a representative ecosystem at the parks in the spring of 2013, and were subsequently serviced by Parks Canada staff. The traps were deployed in a range of habitats including coniferous forests, mixed forests, marshes, and bogs. The Malaise traps were deployed for roughly 20 weeks, with the exception of Torngat Mountains National Park which only collected for 3 weeks due to a short field season for Park Staff. Weekly samples were preserved in 95% ethanol and then held at -20°C. All trap samples were then assembled for subsequent processing at BIO.

The trap samples were accessioned, specimens were identified to order, arrayed, labeled, databased, and tissue-sampled for genetic analysis (Figure 2). 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 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 1. Sampling locations at the 14 Canadian National Parks surveyed in 2013.

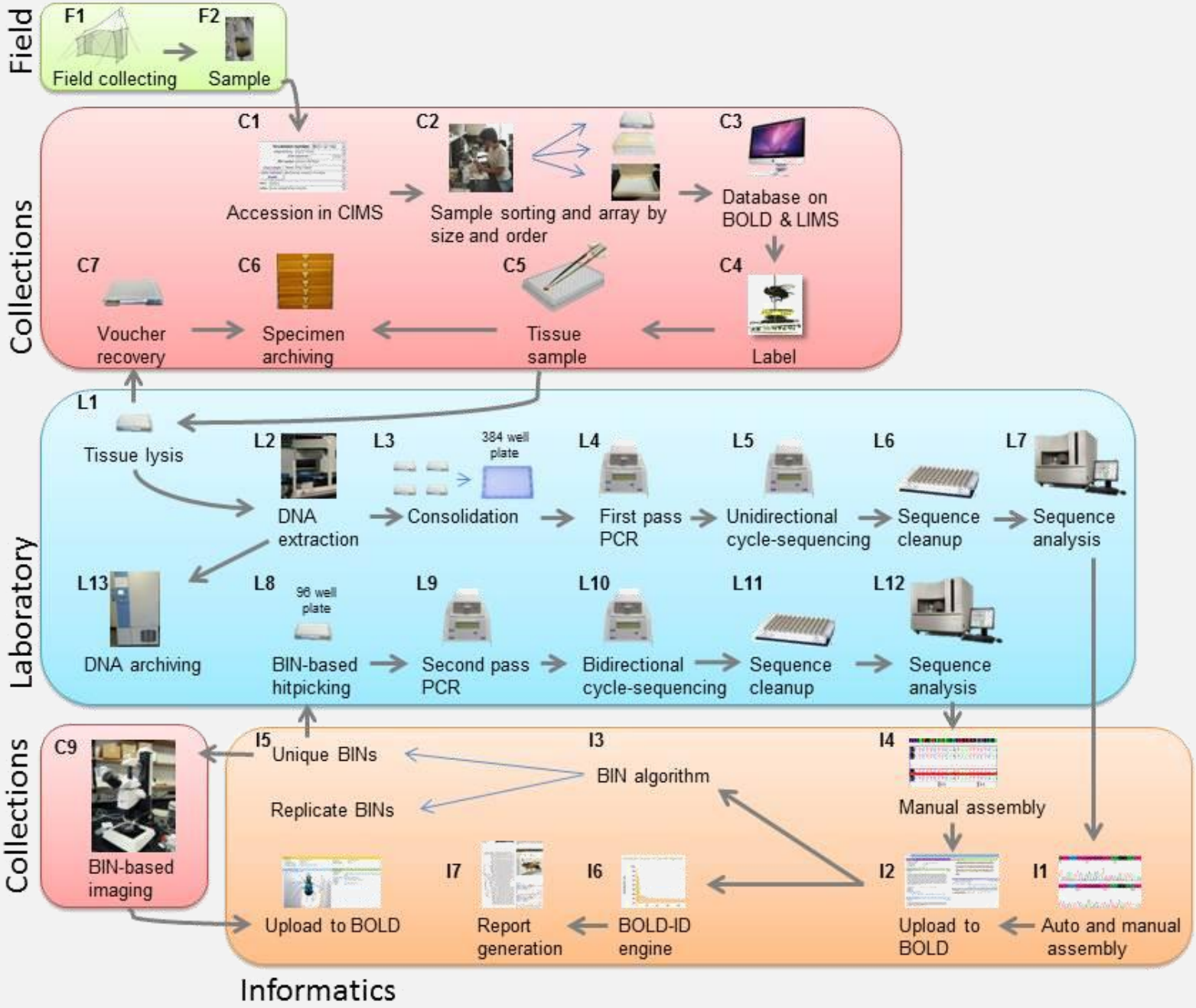


Figure 2. Schematic diagram showing the specimen workflow. Front end processing begins with field collecting (F1) and proceeds through to archiving of specimen Laboratory analysis begins with tissue lysis (L1) through to sequence analysis (L12). The informatics workflow includes both manual (I4) and auto sequence assembly finishes with BIN assignments and subsequent imaging of each BIN (C9).

2012-2013: RESULTS FOR 28 NATIONAL PARKS

The barcode analysis of all Malaise trap samples from 2013 was completed by fall 2014. In total, 227 weekly samples and nearly 280K specimens were analyzed. A total of 240,373 specimens generated barcode sequences that were long enough to allow a BIN assignment. Their analysis revealed a total of 17,427 BINs from the 2013 collection.

In combination with the 2012 samples, the CNP Malaise program has collected over 430K specimens to date. The average sequence success rate was 90% which led to 371,387 records with sequences long enough for a BIN assignment. A total of 26,989 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 39,457 (Figure 3).

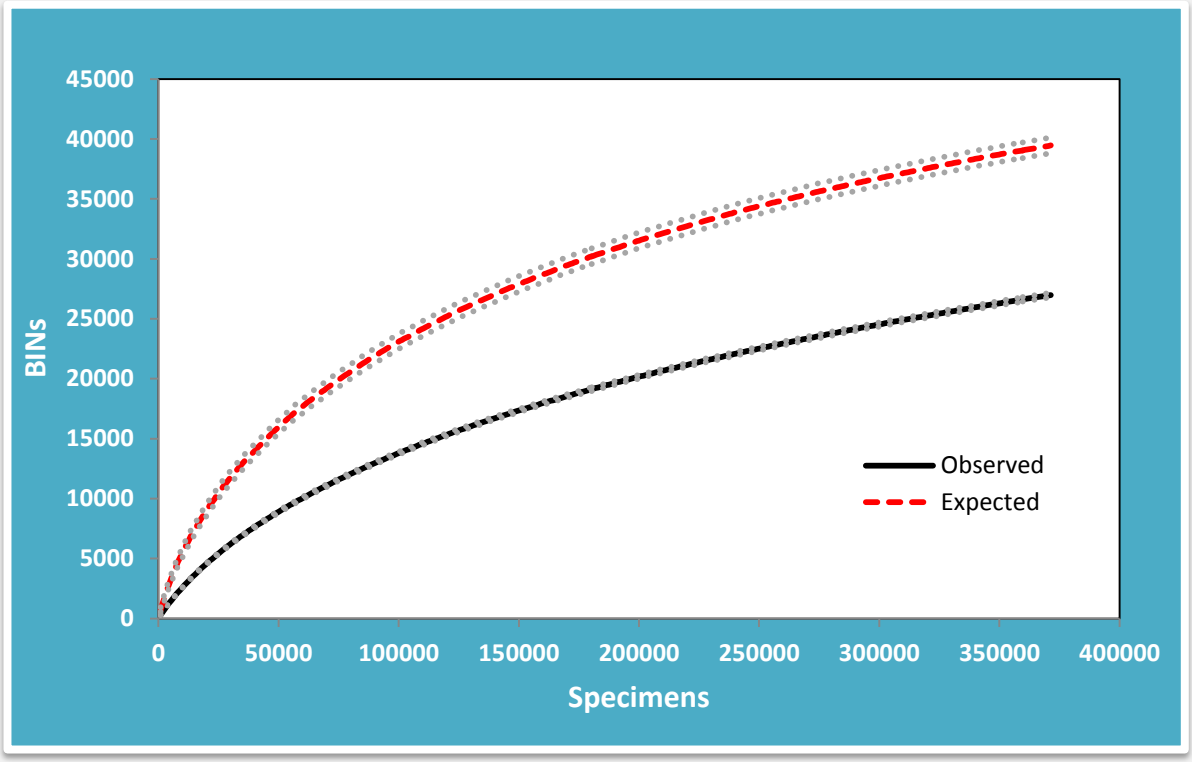


Figure 3. BIN accumulation curve for the 414 Malaise trap samples collected in 28 Canadian National Parks during 2012-2013.

The usual 'hollow curve' species abundance pattern was observed, with 10,245 species represented by just a single individual (singletons) (Figure 4). By comparison, just 623 BINs were represented by 100 or more individuals. The most commonly encountered species was *Entomobrya nivalis* – a common 'slender springtail' – with 4813 individuals sequenced. Species richness extrapolation using the lognormal species abundance distribution (Preston, 1962) suggests that nearly twice as many BINs exist in these 28 National Parks (47,303 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.

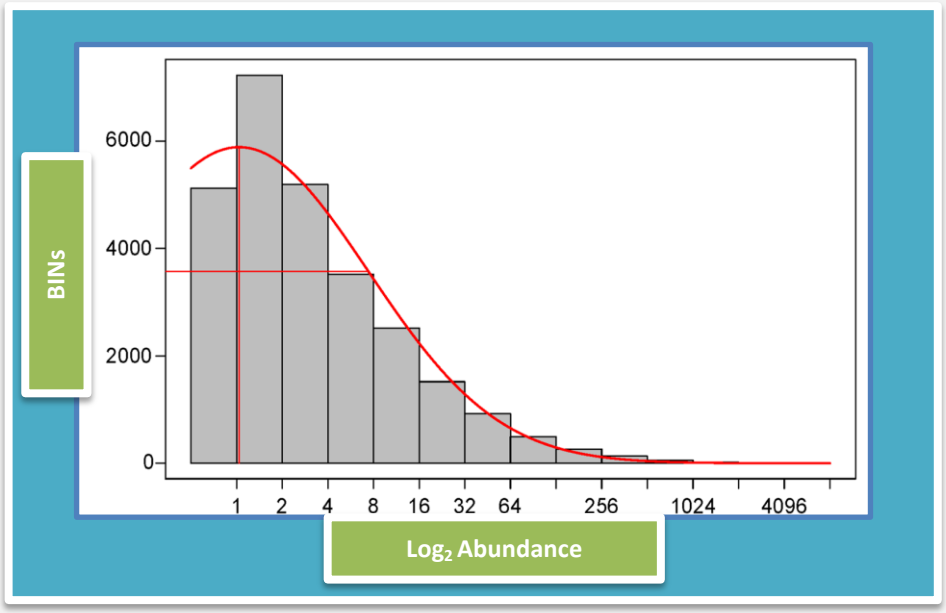


Figure 4. Lognormal species abundance curve, showing the total BINs within each $\log_2$ abundance frequency interval (Preston, 1962).

Among the 2013 parks with full sampling seasons, the number of individuals collected varied from a low of 15,280 specimens from 21 samples at Cape Breton Highlands National Park to a high of 30,188 specimens from 21 samples at Forillon National Park. Sequencing success also varied among parks, from a low of 84.9% at Fundy National Park (13,111 barcode records from 15,435 specimens), versus 94% for Prince Edward Island National Park (Figure 5). The number of BINs detected ranged from a low of 1592 at Fundy National Park to a high of 4017 at Forillon National Park (Figure 5).

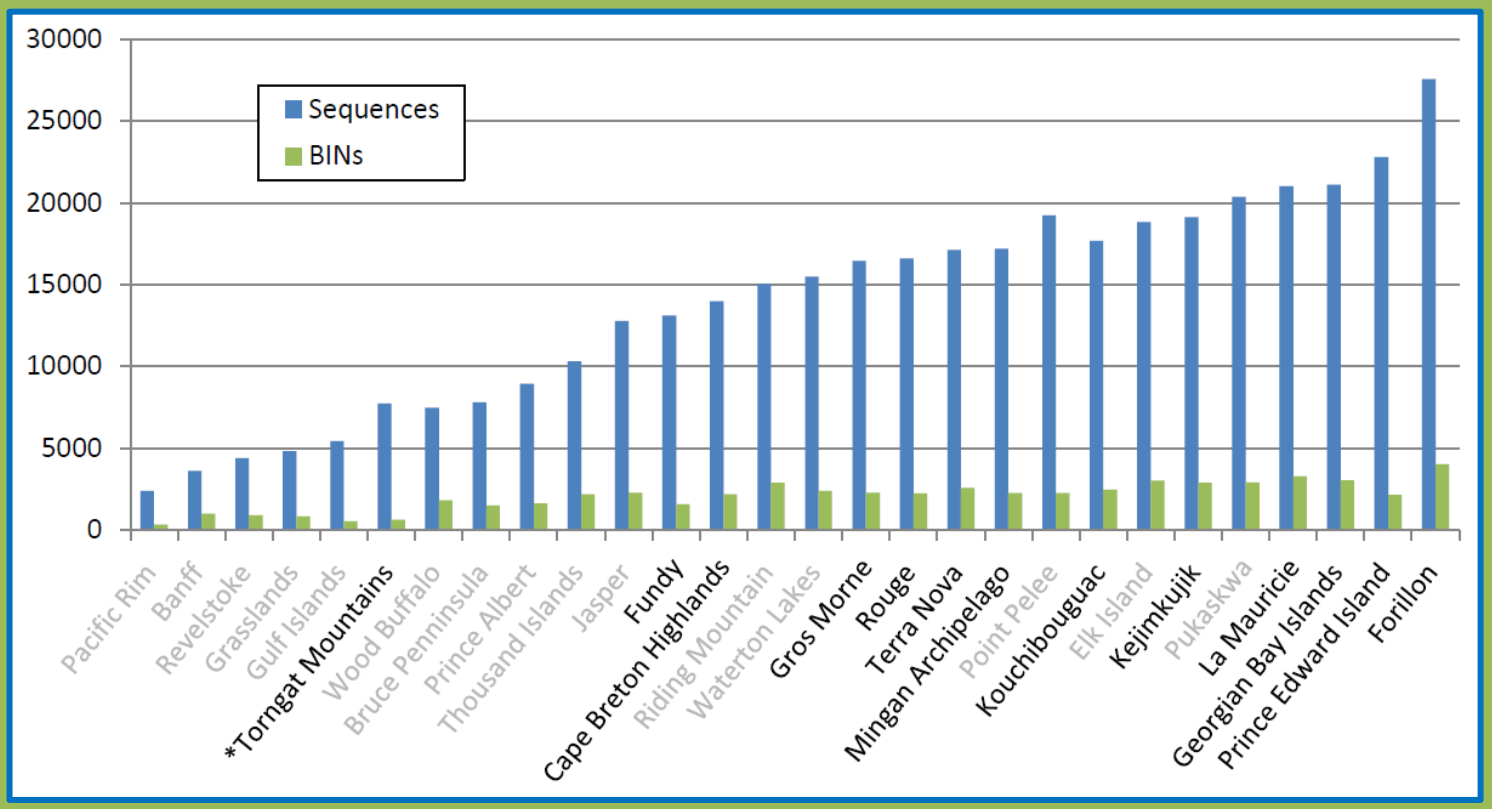


Figure 5. Total sequences and number of BINs generated from each of the 28 parks; grey text indicates 2012 sampling (*only 3 sampling weeks).

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 = 2086 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.

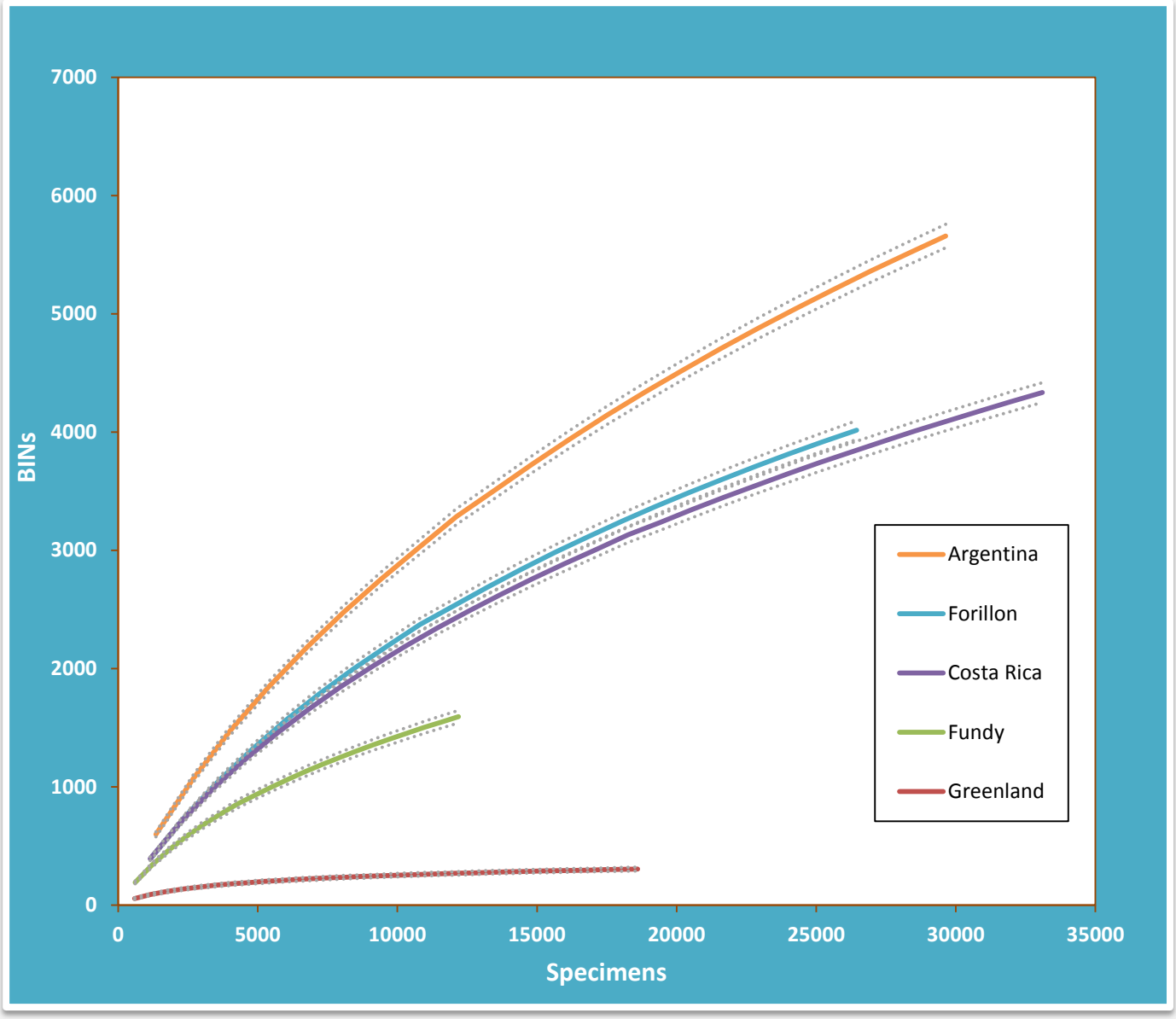


Figure 6. Comparison of BIN accumulation curves for 125 Malaise samples collected from 5 different sampling sites.

Figure 7. Regression analysis examining the relationship between the number of barcoded specimens and number of BINs

(BINs = 6.62(Specimens) + 46.89).

When analyzing both CNP Malaise Programs from 2012 and 2013, the number of BINs detected in each park was strongly influenced by sample size (Figure 7, R^2 = 0.82, p>0.0001). With over 30K specimens analyzed, Forillon National Park displayed the highest BIN count, while Pacific Rim National Park captured less than 400 BINs from 3010 specimens.

Of the 27K BINs captured, more than half were unique to a single collection site; i.e. 16,094 BINs occurred in only one of the 28 parks. The number of BINs unique to each park varied (Figure 8). Point Pelee National Park exhibited the highest count of unique BINs as nearly half of its BINs were unique (1207 BINs of 2270).

Pacific Rim National Park had the fewest unique BINs (N = 208), but the highest ratio of unique BINs to BINs captured (60%). In contrast, Fundy National Park, with 227 unique BINs, had the lowest ratio of unique BINs to BINs captured. This indicates considerably high diversity at Pacific Rim National Park (and others) despite the perception of low diversity given current sampling efforts (ie. Gulf Islands, Grasslands, Torngat Mountains). It is evident that the Malaise trapping method is less effective when sampling in dense rain forests such as Pacific Rim, despite being deployed for the full 20 week collection period. Since flight paths are reduced, fewer specimens are captured, but the specimens that are collected are extremely diverse.

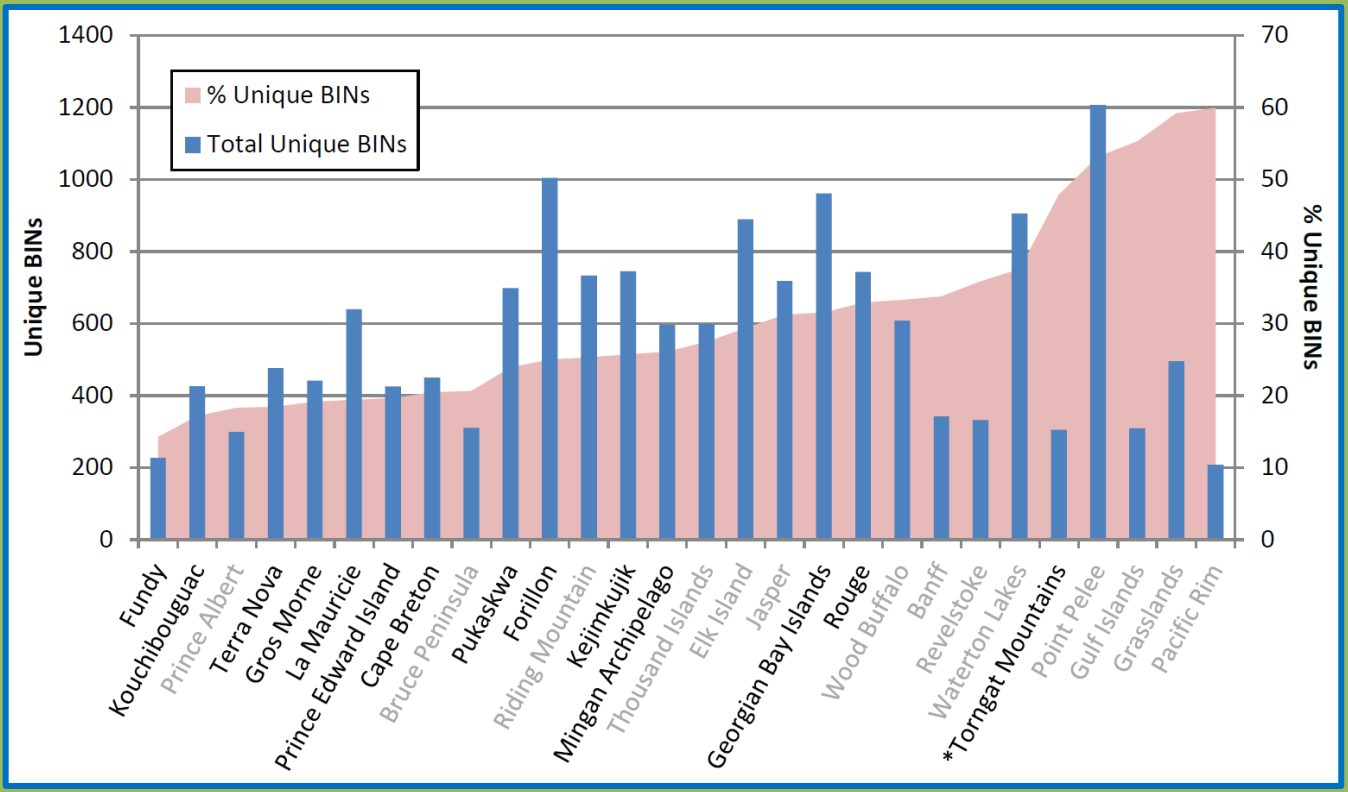


Figure 8. Total number of BINs unique to each park (bars) and the percentage of unique BINs collected in each park (Unique BINs/Total BINs); grey text indicates 2012 sampling (*only 3 sampling weeks).

The similarity in species composition between parks showed marked variation (Figure 9). For example, Kouchibougac and La Mauricie National Parks – 607km apart – shared the highest proportion of BINs (1093 BINs), with a Chao's Sorenson Similarity index (Chao et al., 2005) of 0.70. By contrast, Grasslands and Pacific Rim National Parks – 1313km apart – shared only one BIN; a species of fly (*Helina* sp.) from the Muscidae family (Chao's Sorenson Similarity index = 0.001). Parks in two of the east coast provinces shared relatively higher proportions of BINs (Chao's Sorenson Similarity index for both = 0.66); specifically Fundy and Kouchibouguac in New Brunswick and Terra Nova and Gros Morne in Newfoundland. Surprisingly, Torngat Mountains shared the highest proportion of BINs with Jasper National Park, despite being over 3000km apart. This likely reflects the similar habitats and elevations between the two locations. In addition, an interesting, although not unexpected pattern was apparent – the Rocky Mountains act as a major barrier to species as evidenced by the low connectivity between sites on opposite sides of the range (See 2 Gulf Islands and 7 Elk Island in Figure 10).

Within Point Pelee National Park (Figure 11a) the number of shared BINs between weekly samples ranged from 56 (between weeks 1 and 7) to 266 (between weeks 4 and 5). Similarly at Rouge National Park (Figure 11b), the number of shared BINs ranged from 9 to 135 BINs between weekly samples. Species overlap trends (Figure 11) suggest that BINs tend to become more common later in the season (increased likelihood of being detected in more than one sample). Despite this trend, most BINs were only detected in a single sample, suggesting substantial species turnover across seasons.

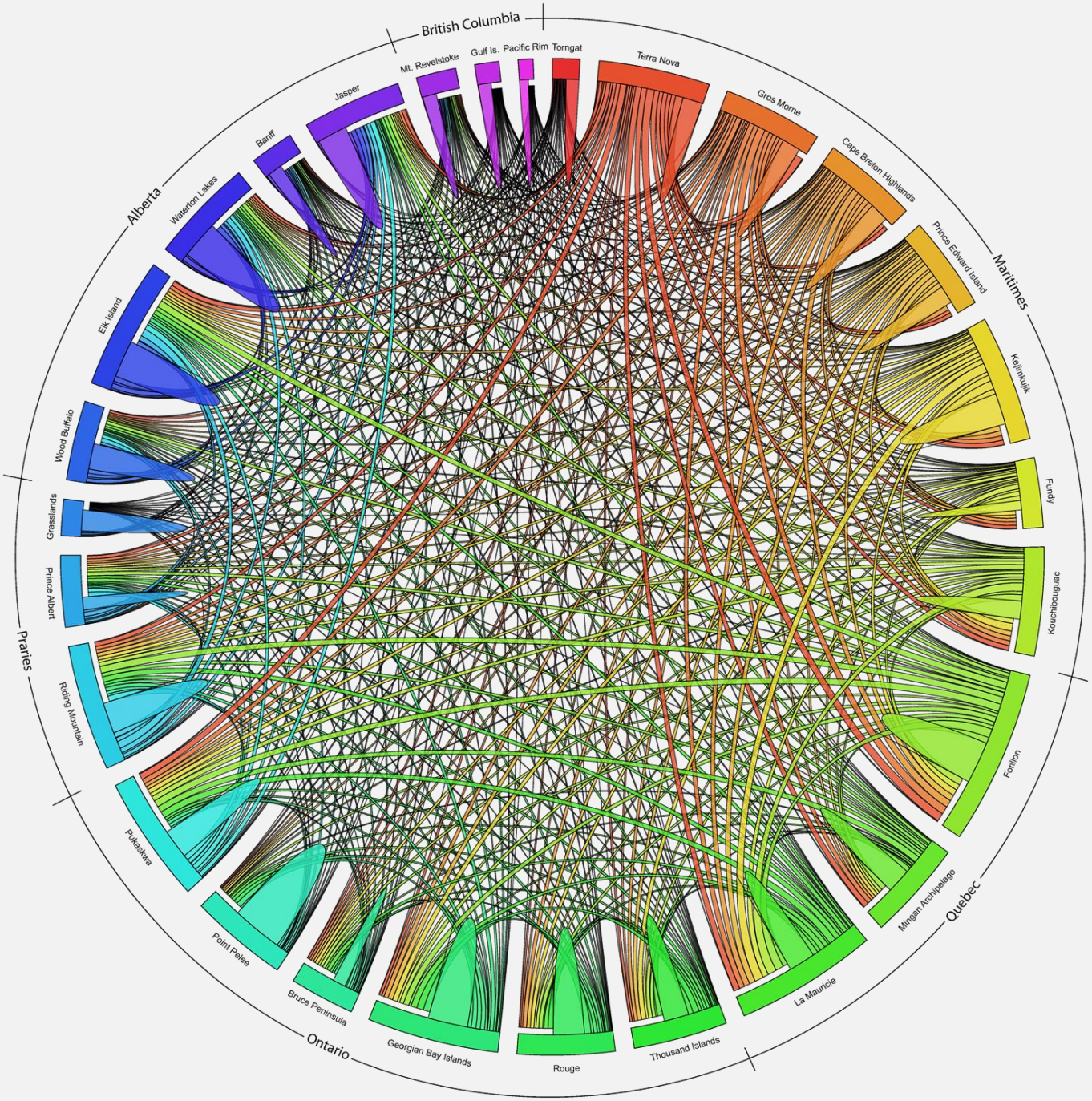


Figure 9. Chord diagram of species overlap between all 28 National Parks, arranged East to West in a clockwise fashion. The width of each wedge reflects the number of BINs captured in each park relative to the others. The widths of internal humps are proportional to the unique BINs within each park. 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.

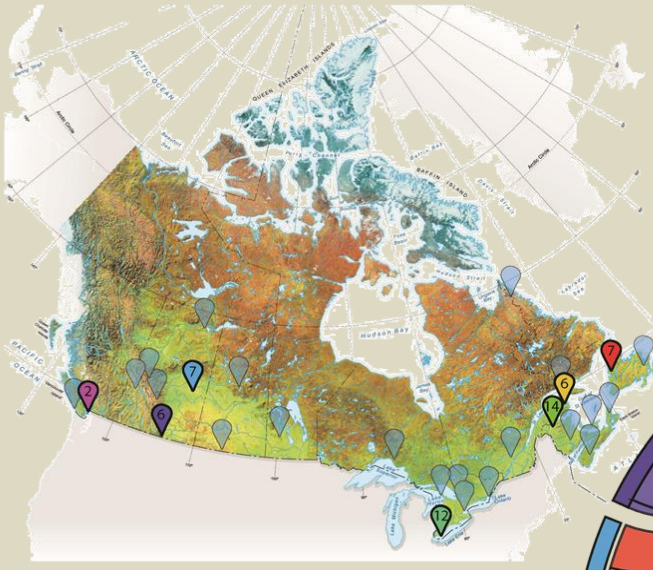
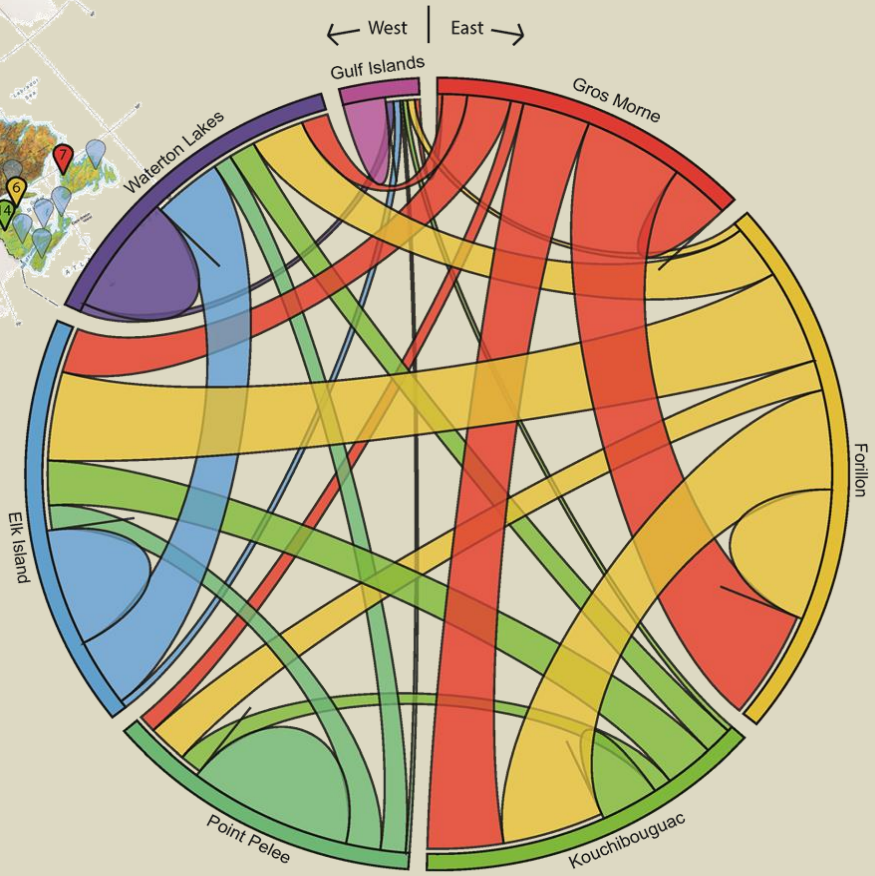


Figure 10. Locations and chord diagram of species overlap between seven of the sampled parks. See Figure 9 for a description of the chord diagram.

	BINs
Gulf Islands NP	559
Waterton Lakes NP	2412
Elk Island NP	3029
Point Pelee NP	2270
Kouchibouguac NP	4016
Forillon NP	2308
Gros Morne NP	2482



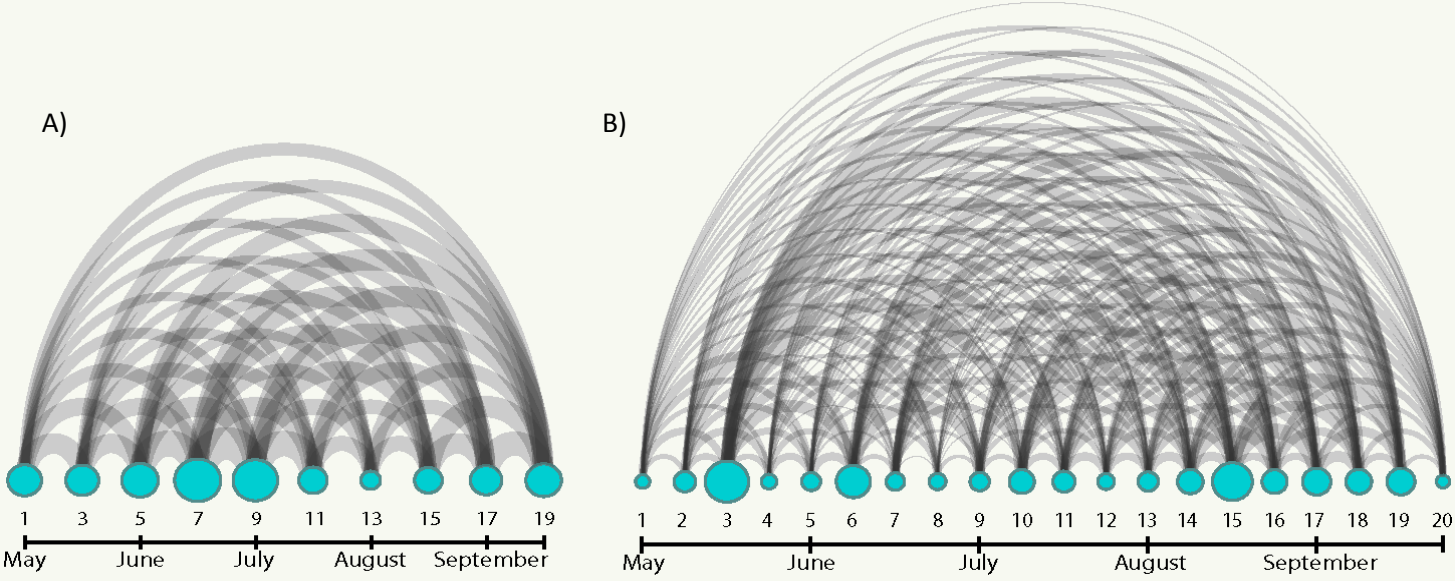


Figure 11. Species overlap between A) 10 bi-weekly samples collected in Point Pelee National Park in 2012, and B) 20 weekly samples collected in Rouge National Park in 2013. Size of nodes indicate the number of BINs in each sample, and the width of the arcs reflect the number of species shared between each sample.

The Standardized Sampling Program was executed in five national parks in 2012: Banff, Elk Island, Jasper, Prince Albert, and Waterton Lakes National Parks. In three parks (Banff, Jasper, Waterton Lakes), standardized sampling was performed for a second week, resulting in 24 comparisons of six trapping methods (N = 144, 144K specimens). The number of BINs captured in each sample was significantly associated with the number of specimens in that sample [BINs = 80.5 + 0.20(Specimens)] (Figure 12). While the slope of this relationship for Malaise traps alone is steeper than all other trapping methods (Figure 12), this difference is not significant. Malaise traps captured more specimens ($p<0.0001$) than the other trap types (Figure 13), revealing a significantly higher proportion of the local fauna (33% of total BINs, and 40% 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.

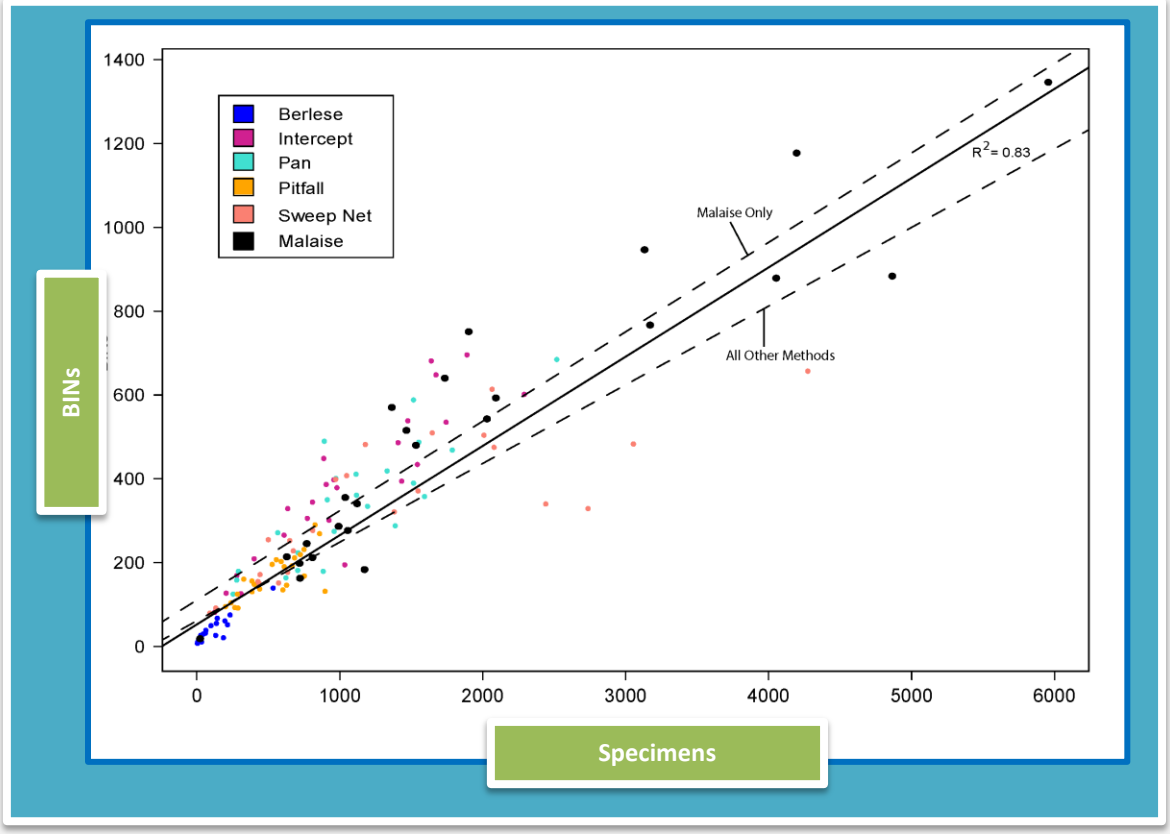


Figure 12. Regression analysis examining the relationship between specimens collected and BINs detected using different sampling strategies.

The DNA barcode reference library on BOLD has recently gained increased species coverage for spiders (order: Araneae), allowing the discovery of many new species for Canada, for a province, and in science. Many of these new species were collected in National Parks and these new records based on park and province are detailed in Table 1. In each of the National Parks sampled, at least one new species occurrence record was noted (meaning that species was never before recorded in the province). As well, three species of spiders never recorded in Canada were found within the Ontario National Parks. In

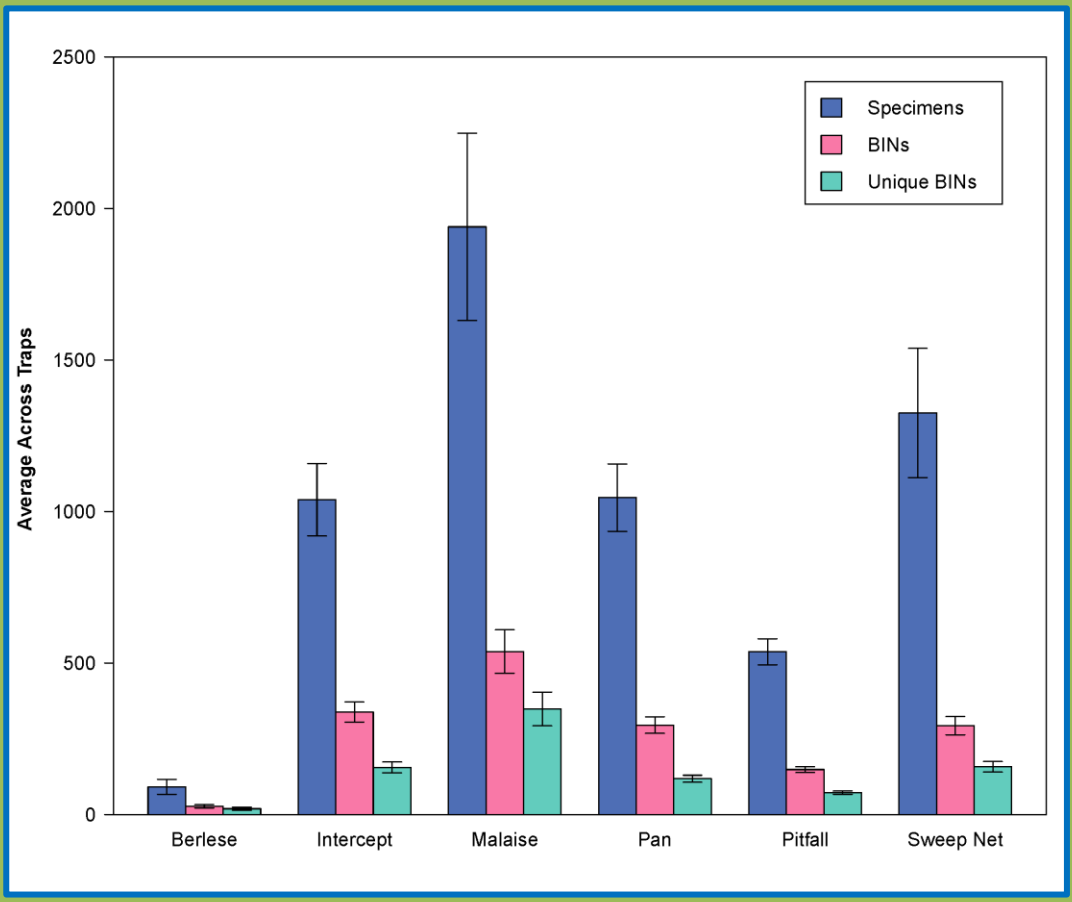


Figure 13. Comparison of six sampling methods, showing the total number of BINs along with the number of unique BINs collected by each method across all 14 parks.

total, ten species new to science were also recorded. Despite being a newly discovered species, *Mughiphantes* sp. has an apparently broad geographic range as it was collected in three different provinces – Alberta, Quebec, and Newfoundland.

The diversity of species collected by Malaise traps is impressive. The combined results from CNP Malaise Programs 2012 and 2013 included representatives for 26,989 BINs from 431,825 total specimens collected from Malaise traps in 28 Canadian National Parks. This BIN count represents 85.4% of the total number (N = 31,598) of terrestrial arthropod species recorded in all prior taxonomic studies, and 42.4% 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.

Table 1. New species records for spiders in Canada by province and park (some species found in multiple parks).

Province (Park)	New Species to Province	New Species to Canada	New Species to Science
Alberta	**42**	**0**	**6**
Banff National Park	14		2
Elk Island National Park	10		
Jasper National Park	21		3
Waterton Lakes National Park	20		3
Wood Buffalo National Park (AB section)	6		
British Columbia	**15**	**0**	**4**
Glacier National Park	3		1
Gulf Islands National Park	1		
Kootenay National Park	7		
Mount Revelstoke National Park	4		1
Pacific Rim National Park	4		2
Yoho National Park	5		1
Manitoba	**11**	**0**	**0**
Riding Mountain National Park	11		
New Brunswick	**25**	**0**	**0**
Fundy National Park	17		
Kouchibouguac National Park	12		
Newfoundland and Labrador	**8**	**0**	**1**
Gros Morne National Park	7		
Terra Nova National Park	3		1
Northwest Territories	**4**	**0**	**0**
Wood Buffalo National Park (NWT section)	5		
Nova Scotia	**16**	**0**	**0**
Cape Breton Highlands National Park	7		
Kejimkujik National Park	10		
Ontario	**17**	**3**	**0**
Bruce Peninsula National Park	1		
Georgian Bay Islands National Park	2	1	
Point Pelee National Park	3	1	
Pukaskwa National Park	5		
Rouge National Urban Park	10	1	
Thousand Islands National Park	1		
Prince Edward Island	**45**	**0**	**0**
Prince Edward Island National Park	45		
Quebec	**3**	**0**	**1**
Forillon National Park	1		
La Mauricie National Park	1		
Mingan Archipelago National Park Reserve	1		1
Saskatchewan	**24**	**0**	**0**
Grasslands National Park	4		
Prince Albert National Park	21		

Two Malaise traps were deployed in the forest canyon at Clyburn Valley Road in Cape Breton Highlands National Park (46.65529 N, 60.42849 W, 21m elevation, Figure 14). They collected arthropods weekly from May 10th to September 27th 2013. Nineteen Malaise trap samples were analyzed which contained a range of 248 to 2098 individuals. A total of 15,280 specimens were captured and a barcode recovery rate of 91.5% was observed (Appendix 1). Over half of the specimens were flies (Diptera), followed in abundance by bees, ants and wasps (Hymenoptera), beetles (Coleoptera), true bugs (Hemiptera), and moths and butterflies (Lepidoptera) (Figure 15). A total of 2198 BINs were observed and the Chao species estimate suggests that approximately 4351 BINs are present in the park and could be collected with this method if sampling effort was extended (Chao; Figure 16).

Figure 14. Map of Eastern Canada indicating location of Cape Breton Highlands National Park.

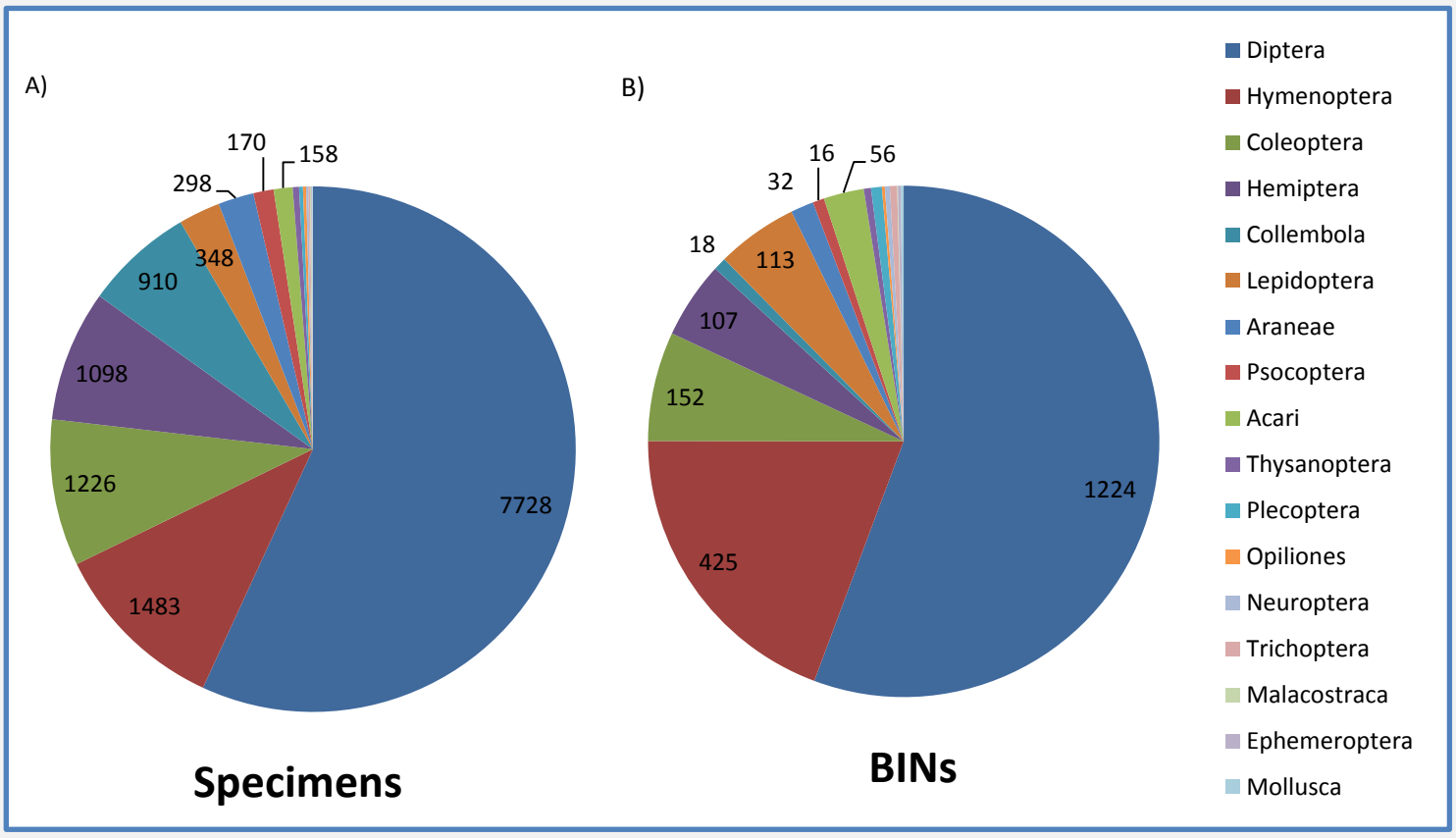


Figure 15. Taxonomic breakdown of (A) N total specimens and (B) N total BINs collected by Malaise traps at Cape Breton Highlands National Park in 2013.

Most specimens have a species-level identification in some taxonomic groups (e.g., Lepidoptera, Araneae), but the taxonomic framework required to provide names is lacking for many BINs in other groups. The order Lepidoptera has the largest barcode coverage and representative images of the 112 Lepidopteran BINs collected in Cape Breton are provided in Appendix 2.

In total, 367 arthropod species were named, representing 17% of the BINs from the park (Appendix 3). Over 96% of BINs were assigned at least to family, and 39% of the BINs were assigned to a genus. Specimens collected from Cape Breton represent 205 different families and 540 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.

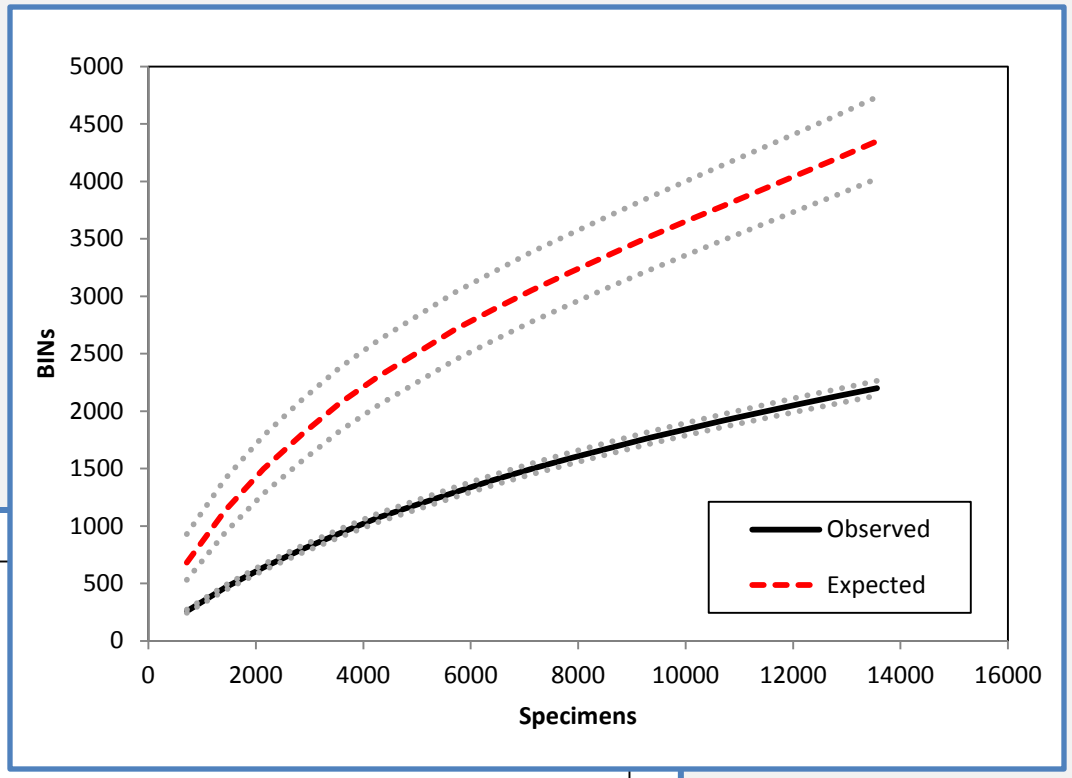


Figure 16. BIN accumulation curve for all specimens collected by the Malaise traps at Cape Breton Highlands National Park in 2013.

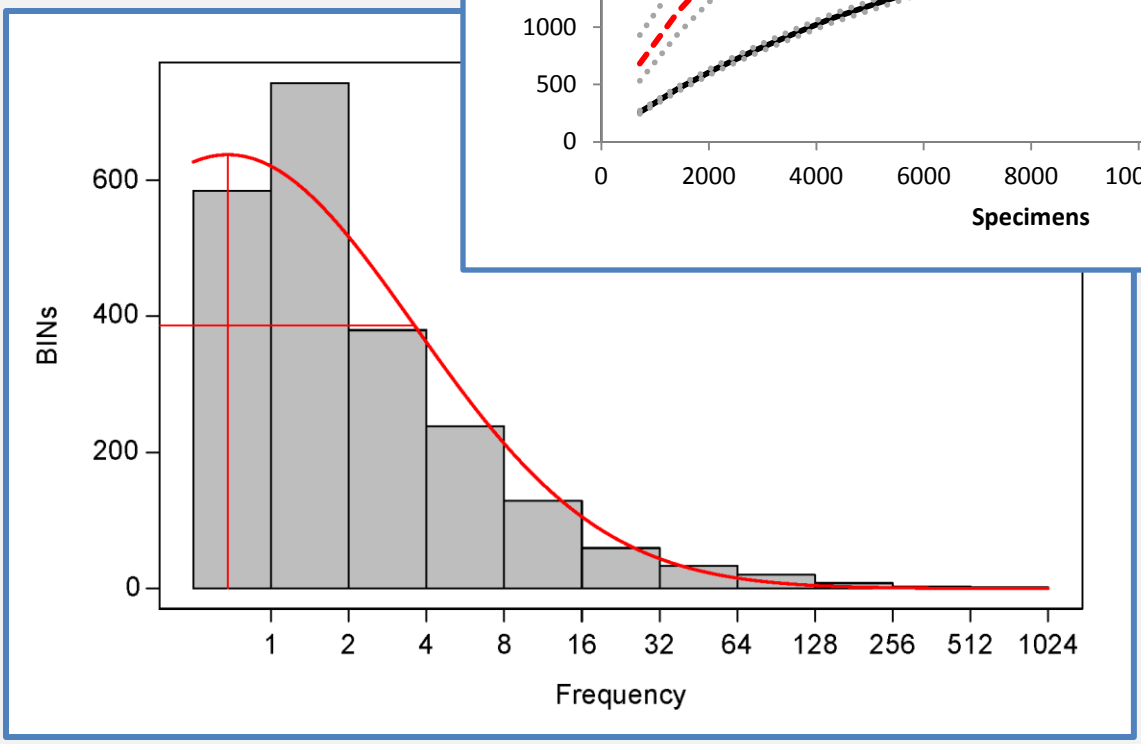


Figure 17. Lognormal species abundance curve, showing the total BINs within each log $_2$ abundance frequency interval.

The pattern of relative species abundance is quite typical, with a few species represented by many individuals (16 species with >100 individuals) – including 711 individuals of *Ctenosciara hyalipennis* (Diptera: Sciaridae) – and a large number of species with few individuals (1168 singletons) (Figure 17). Species richness extrapolation using the lognormal species abundance distribution suggests that 5348 BINs exist in the park (Preston).

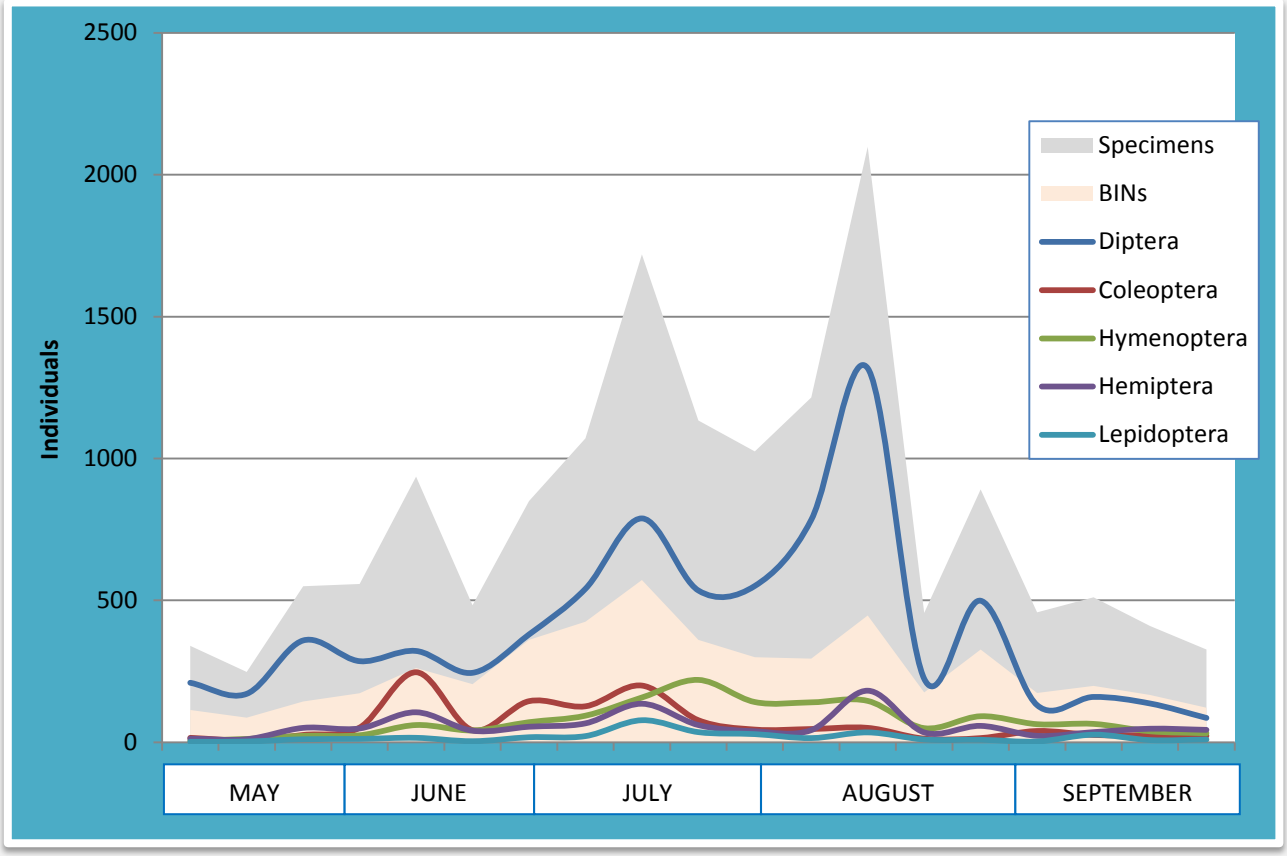


Figure 18. Seasonal variation of the major insect orders as well as specimen and BIN counts collected at Cape Breton Highlands National Park in 2013.

Several population trends of the major insect orders were observed over the 20-week sampling period (Figure 18). Diptera was consistently the most abundant order with several major peaks in mid-July and especially mid-August. Coleoptera peaked earlier in June, while Hemiptera peaked in mid-August. It is interesting to note that the most diverse week (mid-July where the highest number of BINs was recorded) was not the week with the highest specimen count, which is in mid-August. The great diversity in this sample can be accounted for by the relatively higher abundance of other major insect orders besides Diptera.

MALAISE TRAP PROGRAM 2014

Having collected from Western, Central, and Eastern Canadian National Parks, the CNP Malaise Program targeted the Northern and remote National Parks in its third year (Figure 19). Due to the isolation and inaccessibility of these regions, Parks Canada staff and researcher volunteers facilitated the program by deploying and servicing two Malaise traps during the short field season. At the end of the season, samples were shipped to BIO for analysis. 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. As of December 2014, BIO has completed processing 29% of the 2014 CNP Malaise samples collected.

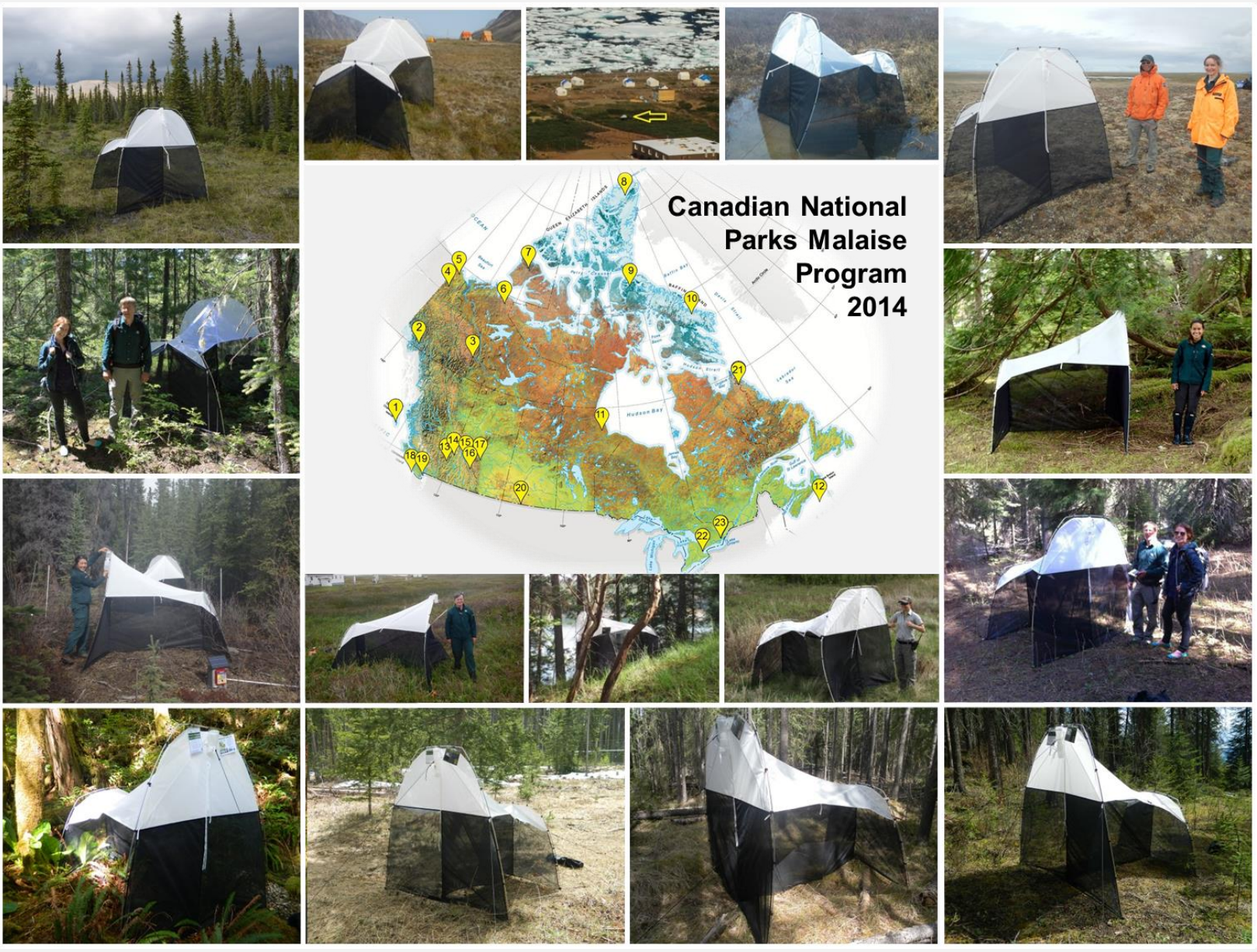


Figure 19. Sampling locations at 16 of the 23 National Parks surveyed in 2014.

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 Cape Breton Highlands National Park in 2013 (colourized based on Taxonomic Order).

Appendix 2. Images for 112 Lepidopteran BINs collected in Cape Breton Highlands National Park; of these, 58 include a species name.

Appendix 3. Taxonomy report for Cape Breton Highlands National Park.

Appendix 4. Complete data spreadsheet of all specimens collected from Cape Breton Highlands National Park 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. 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.

Special thanks to Parks Canada for their participation, and parks staff in particular for servicing the Malaise traps week after week.

Contacts:

Jeremy deWaard
Director of Collections
jdewaard@uoguelph.ca

Kate Perez, MSc
BIObus/Malaise Program Coordinator
kperez@uoguelph.ca

Castor canadensis DNA barcode sequence on the CN Tower

Biodiversity Institute of Ontario
http://www.biodiversity.uoguelph.ca/

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