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

Progress Report for Jasper National Park

Report prepared by the Bio-Inventory and Collections Unit,

Biodiversity Institute of Ontario, University of Guelph

December 2013

INTRODUCTION

The Canadian National Parks 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 longer-term program that will involve regular assessments of arthropod diversity at sites across Canada.

A single Malaise trap was deployed by staff of the BIObus (www.biobus.ca) in a representative ecosystem at each park in the spring of 2012, and it was subsequently serviced by Parks Canada staff. Fourteen parks, from Pacific Rim National Park to St. Lawrence Islands National Park, participated in the program (Figure 1). The traps were deployed in a range of habitats from thick coastal rainforest to open prairie grasslands. Traps were deployed for roughly 20 weeks, with weekly samples preserved in 95% ethanol and then held at -20°C. All collection bottles were picked up by BIObus staff at the end of the season 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 few 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 that awaits an answer is how much of the arthropod fauna at a locality can be sampled with Malaise traps? BIO is exploring this with our 'Standardized Sampling' investigation in a subset of parks. Three sites in each park were selected and five standard collecting techniques were employed at each locality: Malaise, pan, pitfall, and flight-intercept traps, and 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.

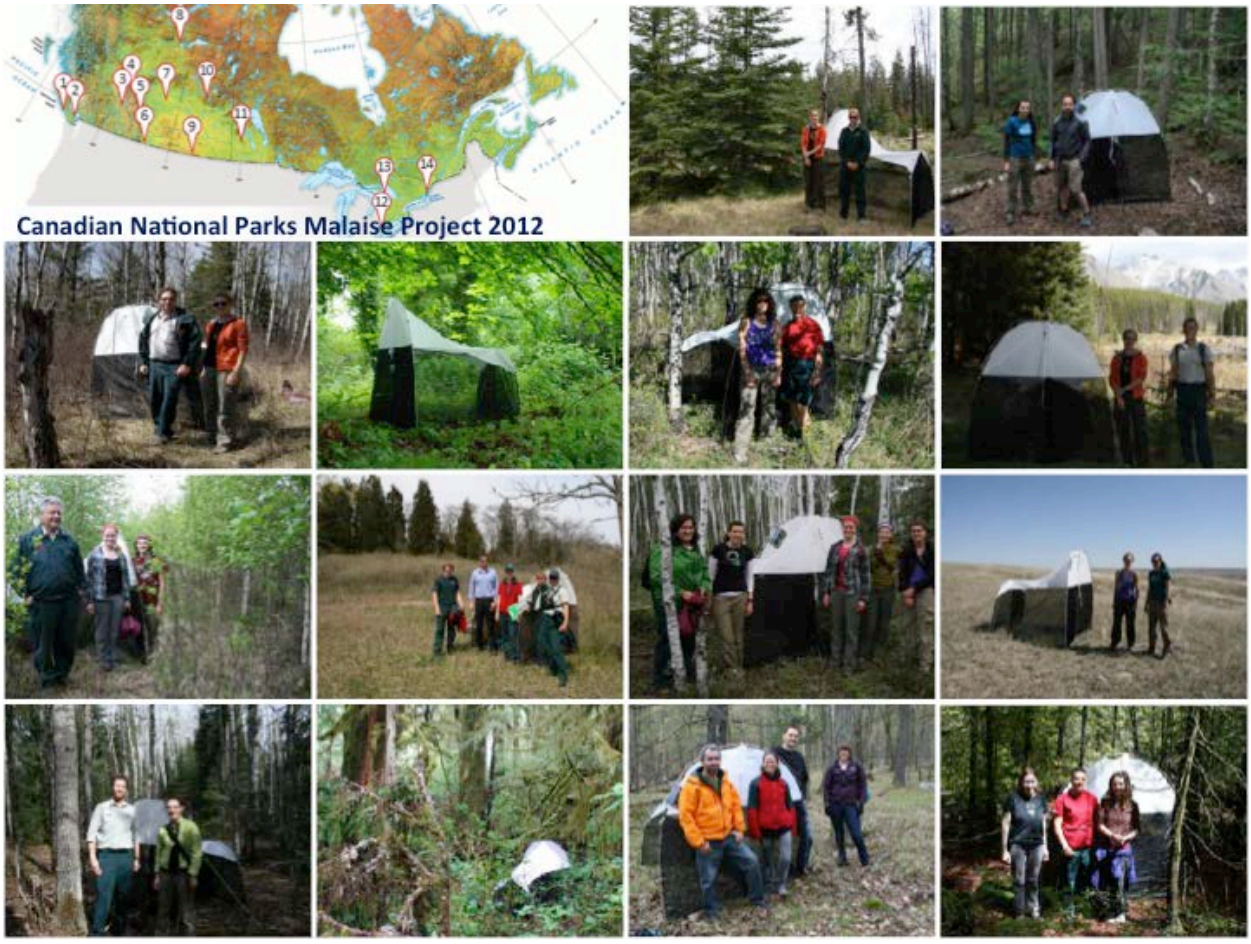

Figure 1. Sampling locations at the 14 Canadian National Parks surveyed in 2012

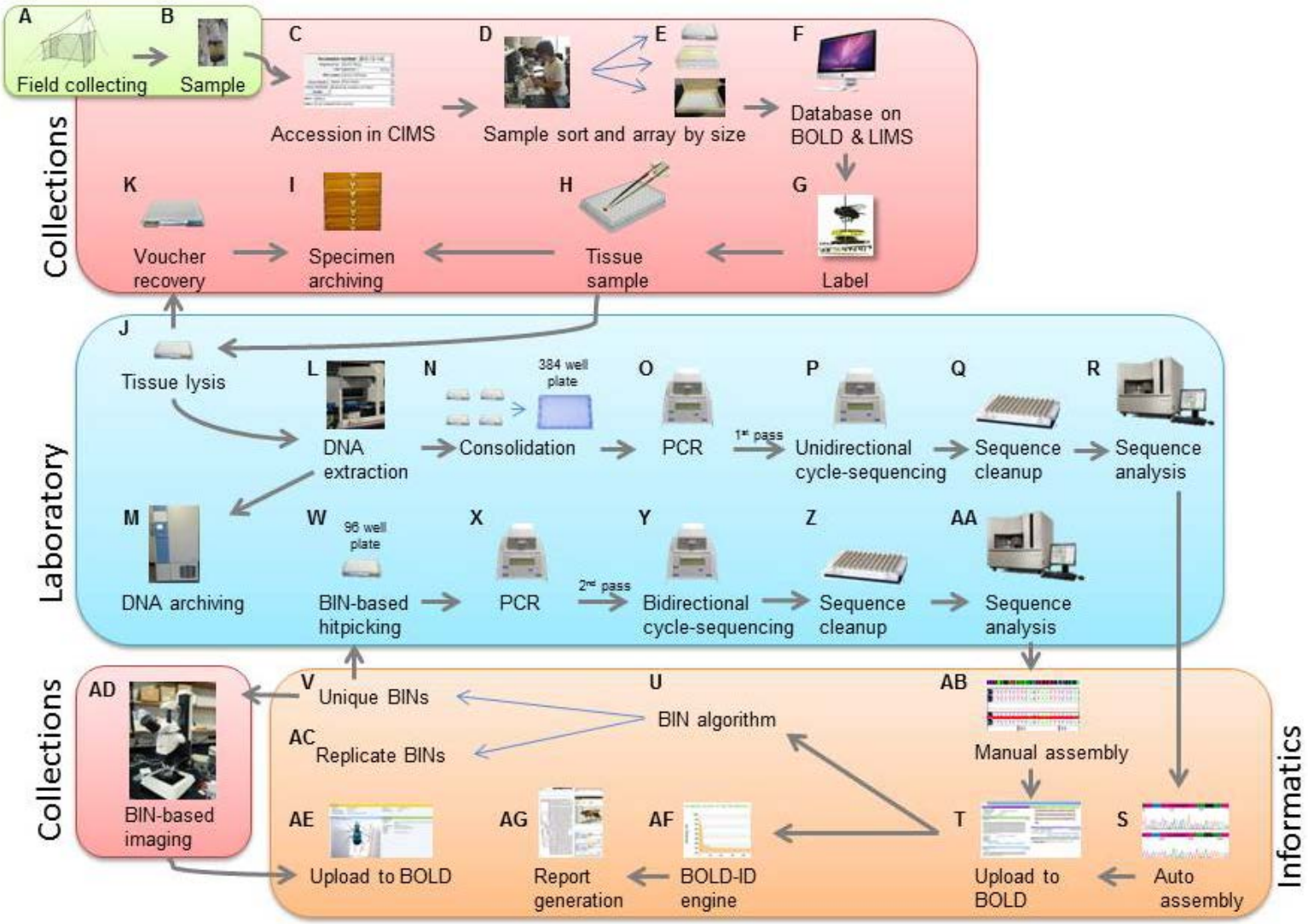

Figure 2. Schematic diagram showing the specimen workflow. Front end processing begins with field collecting (A) and proceeds through to archiving of specimens (I). Laboratory analysis begins with tissue lysis (J) through to sequence analysis (AA). The informatics workflow includes both manual (AB) and auto sequence assembly, and finishes with BIN assignments and subsequent imaging of each BIN (AD).

RESULTS FROM 2012 PROGRAM

The barcode analysis of all Malaise trap samples from 2012 was completed by fall 2013. In total, 189 weekly samples and nearly 150K specimens were analyzed. A total of 129,690 specimens generated barcode sequences that were long enough to allow a BIN assignment. Their analysis revealed a total of 15,814 BINS, while the Chao 1 (Magurran, 2003) species estimate for the total number of BINs that would be encountered with comprehensive sampling was 25,174 (Figure 3).

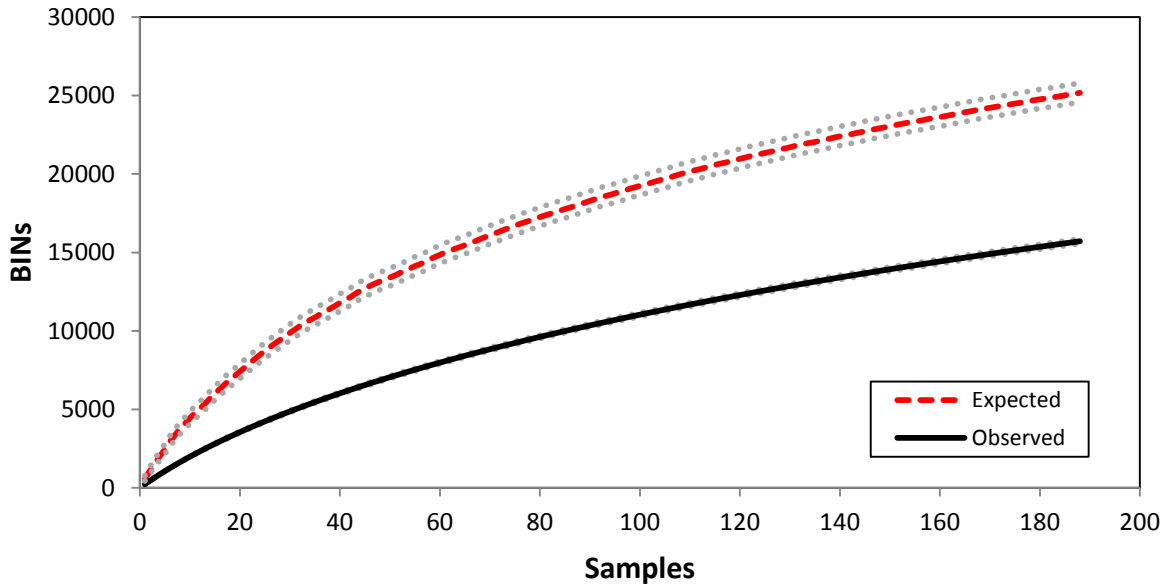

Figure 3. BIN accumulation curve for the 189 Malaise trap samples collected in 14 Canadian National Parks during 2012.

The usual 'hollow curve' species abundance pattern was observed, with 6985 species represented by just a single individual (singletons). By comparison, just 203 BINs were represented by 100 or more individuals. Species richness extrapolation using the lognormal species abundance distribution suggests that nearly twice as many BINs exist in these 14 National Parks (29,844 BINs; Figure 4) as were collected. The most commonly encountered species was the onion fly (*Delia antiqua*) – a cosmopolitan agricultural pest – with exactly 2500 individuals trapped.

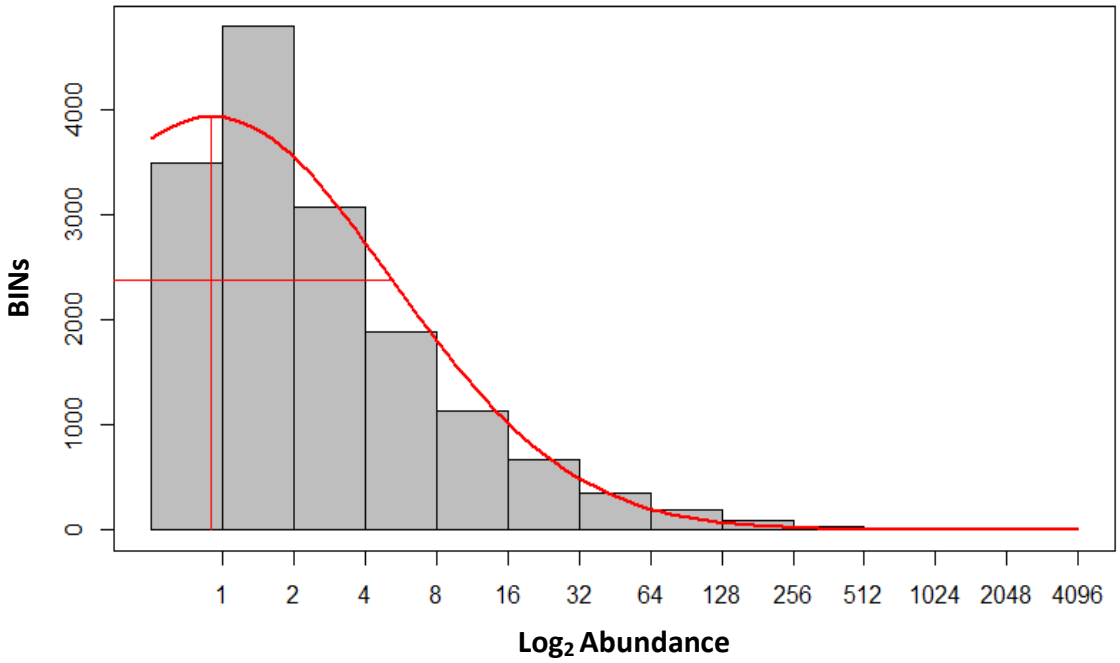

Figure 4. Lognormal species abundance curve, showing the total BINs within each log $_2$ abundance frequency interval (Preston, 1962).

The number of individuals collected in each park varied more than 10-fold ranging from a low of 3010 specimens from 20 weekly samples at Pacific Rim National Park, to 21,191 specimens from 10 bi-weekly samples at Point Pelee National Park. Sequencing success also varied among parks, from a low of 70.5% at Pacific Rim National Park (2125 records from 3010 specimens), versus 91% for Point Pelee National Park (Figure 5), likely reflecting differences in specimen preservation. The number of BINS detected ranged from a low of 504 at Gulf Islands National Park to a high of 3038 at Elk Island National Park (Figure 5). The number of BINs detected in each park was strongly influenced by sample size (Figure 6, $r^2 = 0.75$), but there was evidence of other effects. For example, four parks (Gulf Islands, Pacific Rim, Prince Albert, Point Pelee) showed lower BIN counts than the rest. The number of BINs unique to each park also varied, with Point Pelee National Park exhibiting the highest count (N = 1603; 69.8% of the BINs at this park were only collected in this park; Figure 7). Prince Albert National Park displayed the most unique BINs when standardized for collecting effort, as 84.1% of its 629 BINs were only collected in that park (Figure 7).

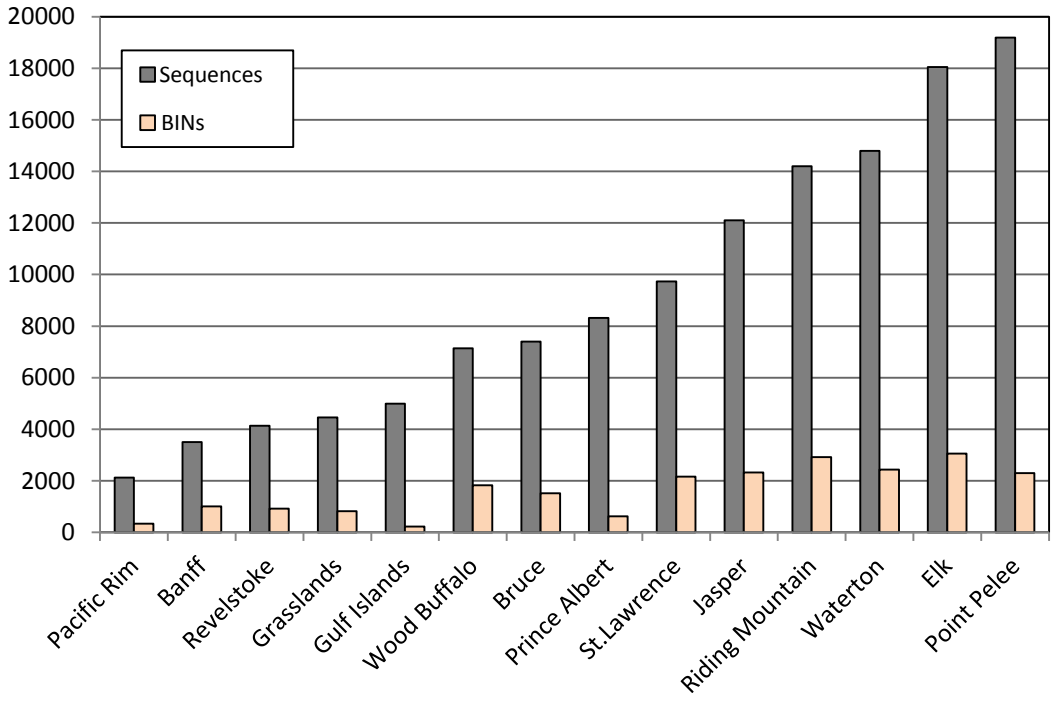

Figure 5. Total sequences and number of BINs generated from each of the 14 parks.

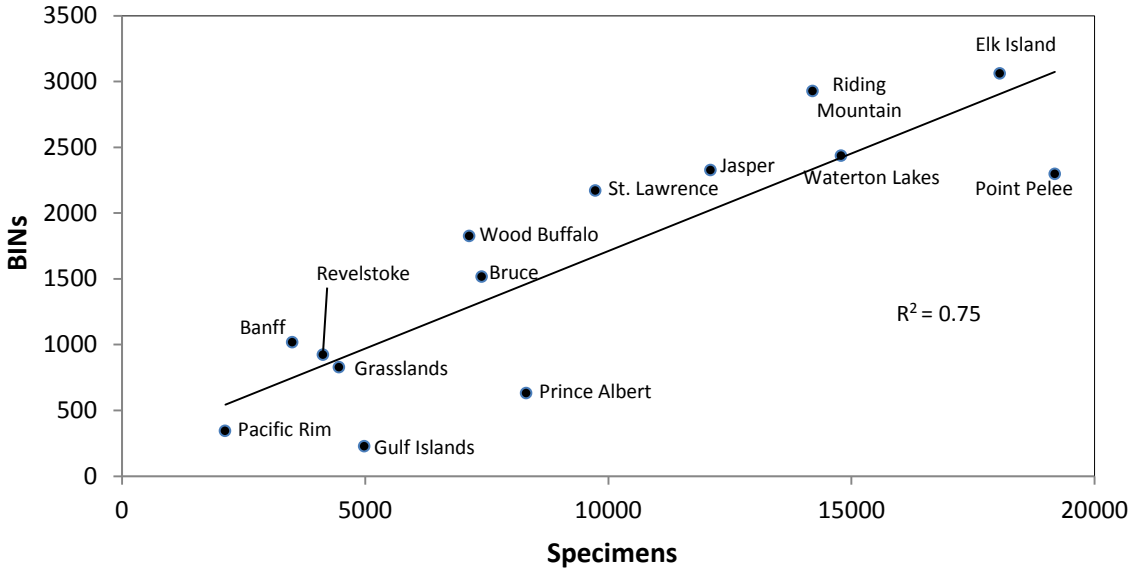

Figure 6. Regression analysis examining the relationship between the number of barcoded specimens and number of BINs (p>0.000). The formula for the linear regression is BINs = 0.15(Specimens) + 229.12.

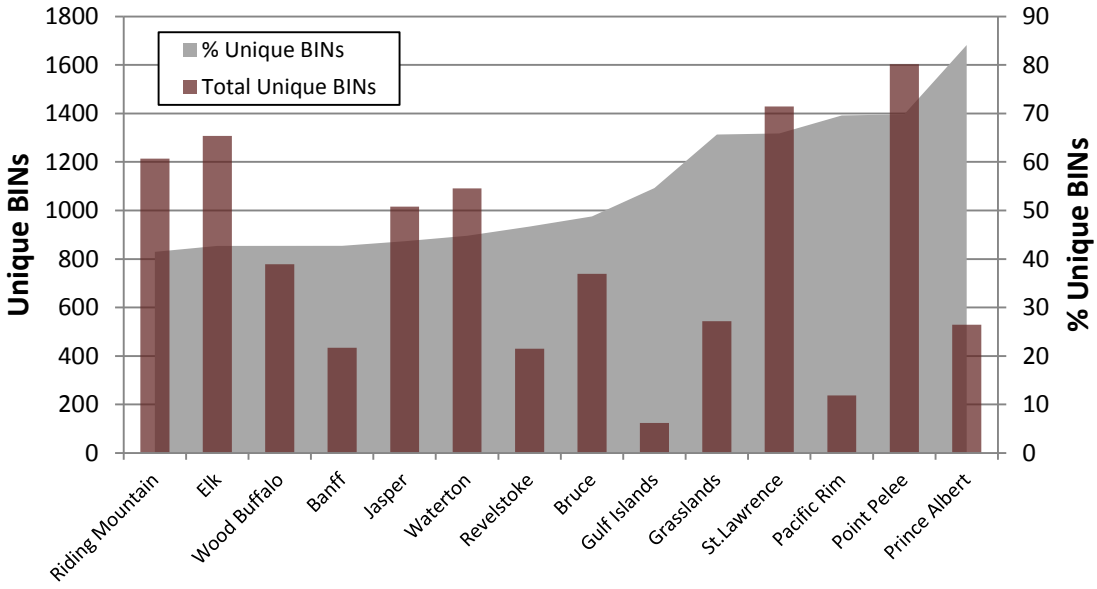

Figure 7. Total number of BINs unique to each park (bars) and the percentage of unique BINs collected in each park (Unique BINs/Total BINs).

The similarity in species composition between parks showed marked variation. For example, Elk Island (7) and Riding Mountain National Parks (11) shared the highest proportion of BINs, with a Chao's Sorenson Similarity index (Chao et al., 2005) of 0.71, which equates to 854 overlapping BINS (Figure 8). By contrast, Gulf Islands (2) and Point Pelee National Parks (12) – over 3000km apart – shared only 22 BINs (Chao's Sorenson Similarity index = 0.014; Figure 8). 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.

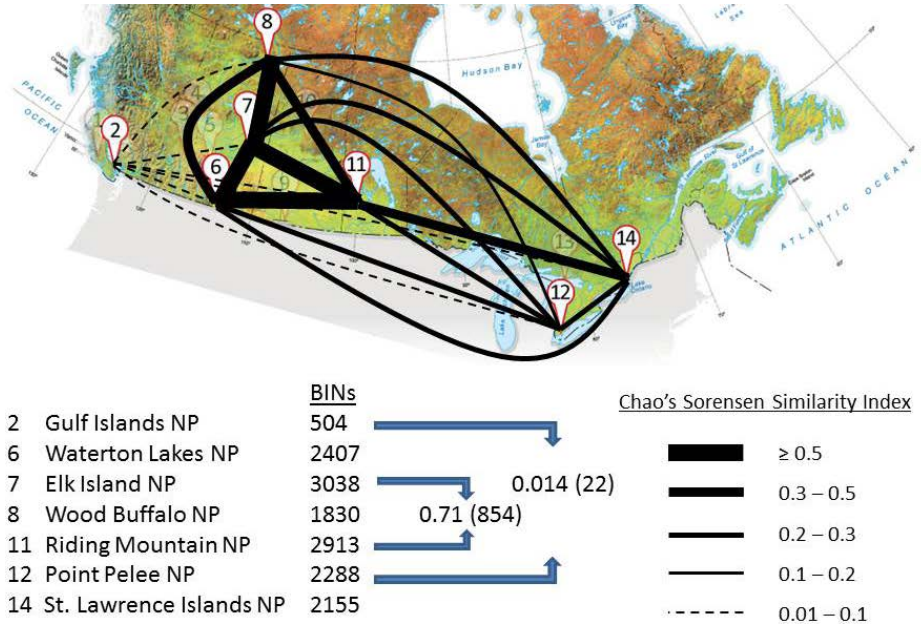

Figure 8. Variation in the proportions of shared species between selected parks as measured by Chao's Sorenson species similarity. Values range from 0.014 to 0.71.

Preliminary results for the Standardized Sampling project from three sites in four national parks (N = 44K specimens) indicated that Malaise traps capture a significant proportion of the local arthropod fauna (Figure 9). Malaise traps captured 27% to 45% of total collected arthropod BINs, and roughly 30% of unique BINs.

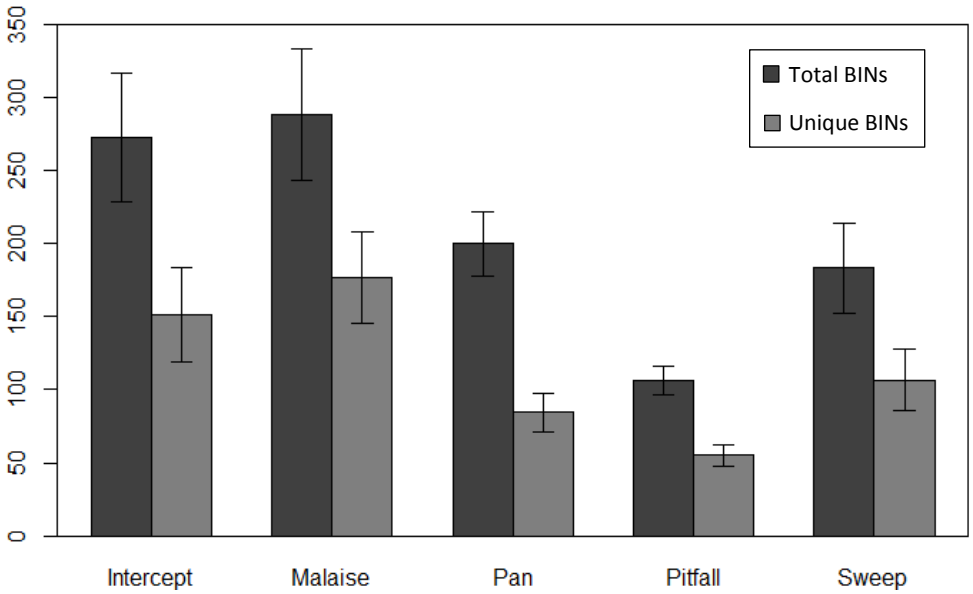

Figure 9. Comparison of four sampling methods, showing the total number of BINs along with the number of unique BINs collected by each method across all 14 parks.

The diversity of species collected by Malaise traps is impressive. The 1068 Lepidoptera BINs (N = 7882) captured in 189 samples from 14 Parks in 2012 represents over 20% of Canada's Lepidoptera diversity (Pohl G., NRCan, pers. comm.). When the 2012 results are combined with those BIO obtained in 2008-2011 in Canada's national parks, 243,137 specimens have now been barcoded from 26 national parks, providing representation for 26,485 BINs. This total represents 83.4% of the total number of terrestrial arthropod species recorded in all prior work (N = 31,598), and 41.6% 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 calculate 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 will be to sample a wide spectrum of environments and disturbance regimes, as well as to examine replicate samples. We expect at that point to 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.

2012 RESULTS – JASPER NATIONAL PARK

The 10 Malaise trap samples analyzed from Jasper National Park in 2012 contained 739 to 2622 individuals, and the barcode recovery rate was 85%. Over half of the specimens were flies (Diptera), followed in abundance by bees, ants and wasps (Hymenoptera), springtails (Collembola), moths and butterflies (Lepidoptera), and mites (Acari) (Figure 10). A total of 2327 BINs were observed, and rarefaction analysis suggests that approximately 4370 BINs are present in the park and could be collected with this method if sampling effort was extended (Figure 11).

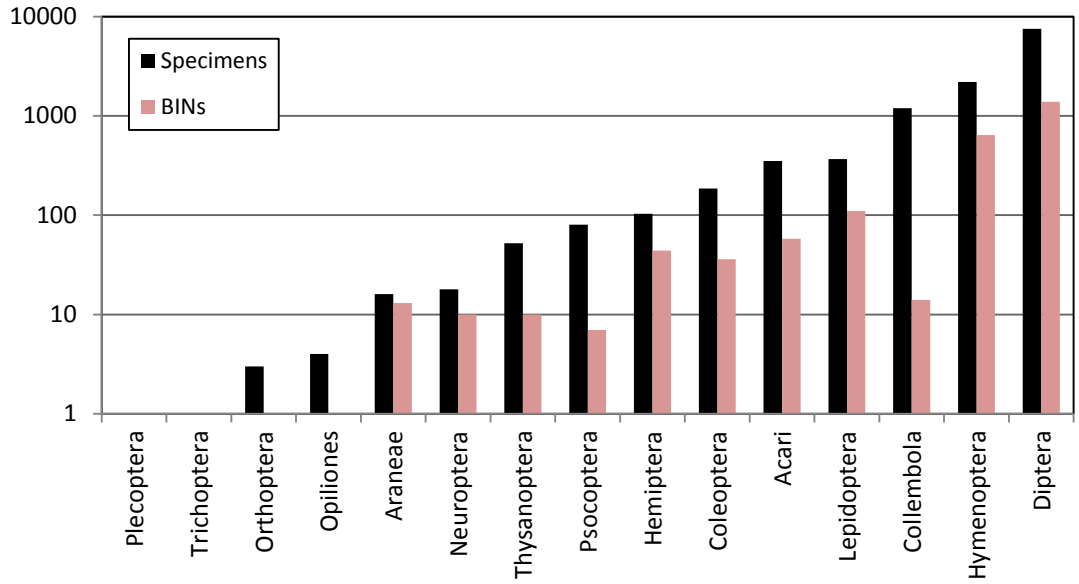

Figure 10. Taxonomic breakdown of specimens collected by Malaise trap at Jasper National Park in 2012.

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. In total, 345 species were named, representing 14.1% (1703) of the specimens barcoded from the park (Appendix 1). These specimens represent 88 families and 247 genera (Appendix 2). Appendix 3 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 (9 species with >100 individuals) – including 1058 individuals of one springtail (Collembola: Entomobrya) species – and a large number of species with few individuals (1228 singletons) (Figure 12).

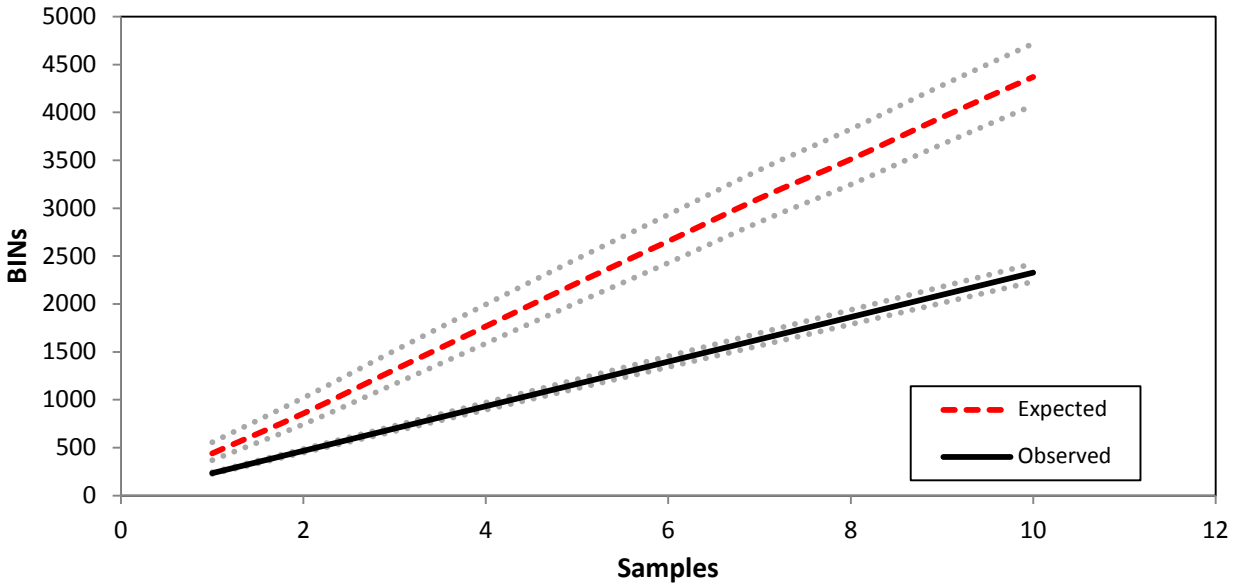

Figure 11. BIN accumulation curve for all specimens collected by Malaise trap at Jasper National Park in 2012.

Standardized sampling was also carried out at Jasper National Park in 2012. A total of 21,323 individuals comprising 3318 BINs were collected with the five sampling techniques (Appendices 4, 5). Malaise trap was the most effective collection technique, resulting in 1941 BINs, followed by pan trap (850 BINs), flight intercept trap (756 BINs), sweep-netting (647 BINs) , and pitfall trap (460 BINs). The Malaise trap was responsible for capturing 58.5% of total arthropod BINs, and 40.2% of unique BINs.

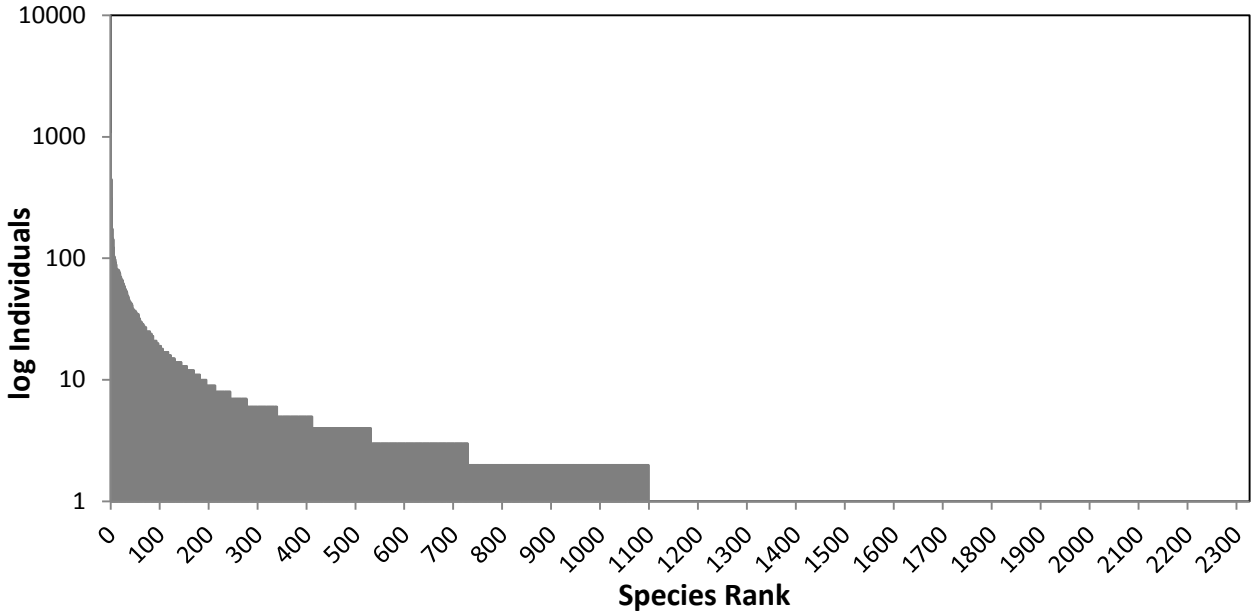

Figure 12. Rank abundance plot for all specimens collected by Malaise trap at Jasper National Park in 2012.

MALAISE TRAP PROGRAM IN 2013

The program continued in 2013 with traps deployed in 14 eastern national parks, ranging from Pukaskwa National Park in Ontario to Terra Nova National Park in Newfoundland (Figure 13). Due to the under-sampling of some parks in 2012, two traps were deployed at each park in 2013. Trap samples were retrieved by BIObus staff (or shipped directly to BIO) in September/October and barcode analysis is in progress.

Figure 13. Sampling locations at the 14 National Parks surveyed in 2013.

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 specimens successfully barcoded from the Malaise trap deployed at Jasper National Park in 2012.

Appendix 2. Images for 108 of the 110 Lepidopteran BINs collected in Jasper National Park; of these, 86 include a species name.

Appendix 3. Complete data spreadsheet of specimens collected by Malaise trap with available taxonomy and collection information.

Appendix 4. Neighbour-joining tree of specimens collected as part of the Standardized Sampling project at Jasper National Park in 2012.

Appendix 5. Complete data spreadsheet of specimens collected through the Standardized Sampling project with available taxonomy and collection information.