DNA Barcode-based Assessment of Canadian Arthropod Diversity:

British Columbia 2014

Report prepared by the Collections Unit,

Centre of Biodiversity Genomics (formerly Biodiversity Institute of Ontario),

University of Guelph

December 2016

INTRODUCTION

In 2014, the Centre of Biodiversity Genomics (CBG), formerly known as the Biodiversity Institute of Ontario (BIO), worked in collaboration with BC Parks to conduct a large-scale arthropod survey to determine Canadian species diversity using DNA barcoding. Large-scale trapping has not been previously attempted, most likely due to the limitations of taxonomists identifying such a large number of specimens. This problem will be overcome with the use of DNA barcoding techniques which differentiates species by variations in a short gene sequence (Hebert et al., 2003). Results from this program will provide an initial assessment of the arthropod diversity within parks in British Columbia (BC). Over the long term, this project will contribute to the creation of a complete DNA barcode library for all eukaryote species that occur in Canada.

Two types of specimen collecting programs were implemented: a Malaise Trap Program, which used a standard method of weekly Malaise sample collection, and a Standardizing Sampling (SS) Program, which involved various trapping methods to compare against Malaise traps. Eleven different BC parks were sampled in 2014, along with the Pacific Forestry Centre in Victoria for a total of 12 sampling locations (Figure 1).

Malaise traps are tent-like structures that are effective at capturing insects from various groups and are easily deployed and cost-effective. For the BC Malaise Program, traps were deployed in ten sites and sample collection was facilitated by BC Parks staff during the summer months. Weekly samples were preserved in 95% ethanol and then held at -20°C before being shipped to CBG at the end of the season.

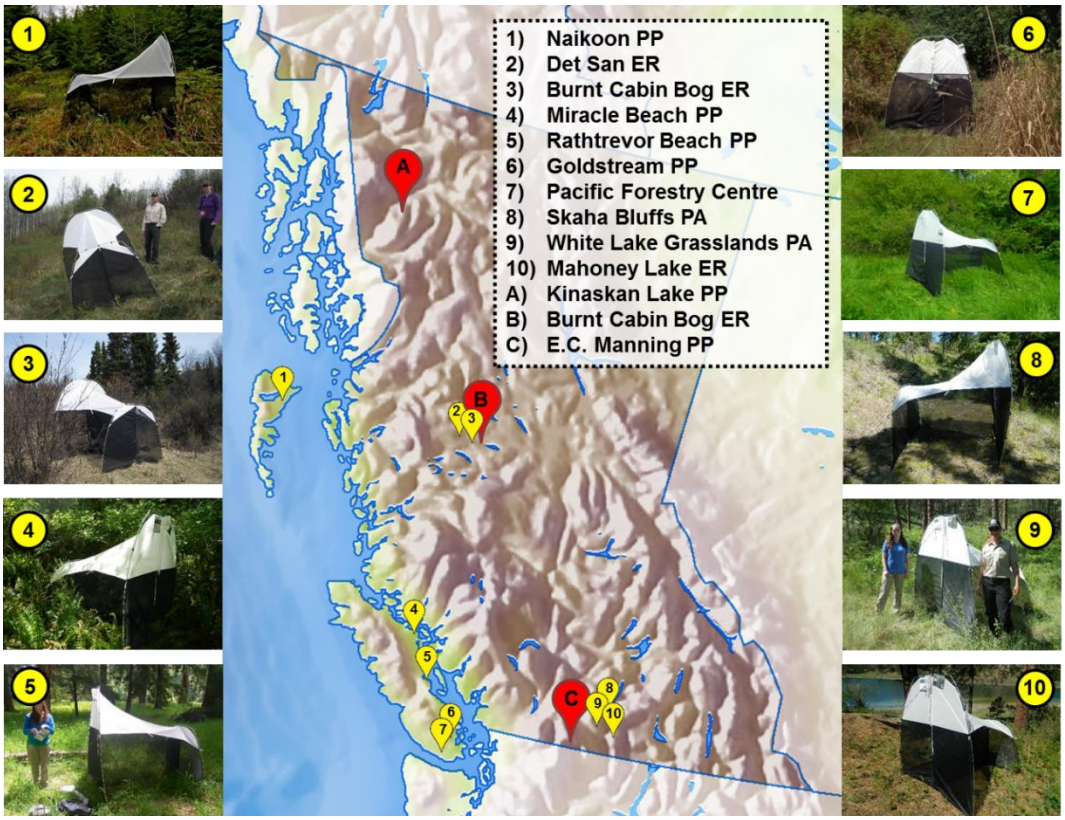


Figure 1. Map and list of sampling sites involved in the BC Malaise Program (yellow; sites 1-10), including images of the traps in each locality, and the Standardized Sampling Program (red; sites A-C).

The SS program was conducted in three parks in BC: Kinaskan Lake, Burnt Cabin Bog, and E.C. Manning. Three sites were chosen within each park and five standard collecting techniques were employed at each locality: 1 Malaise, 10 pan, 20 pitfalls, 3 Berlese, 1 flight-intercept trap, and standardized sweep-netting. Each park was sampled by the BIObus staff for one week and all samples were brought back to CBG for analysis.

All samples chosen for processing were sorted and 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 commons species where only a limited number of individuals from each trap sample was analyzed. Standard barcoding protocols 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). 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 taxonomy reports to be completed for each sampling site and facilitating comparisons among them.

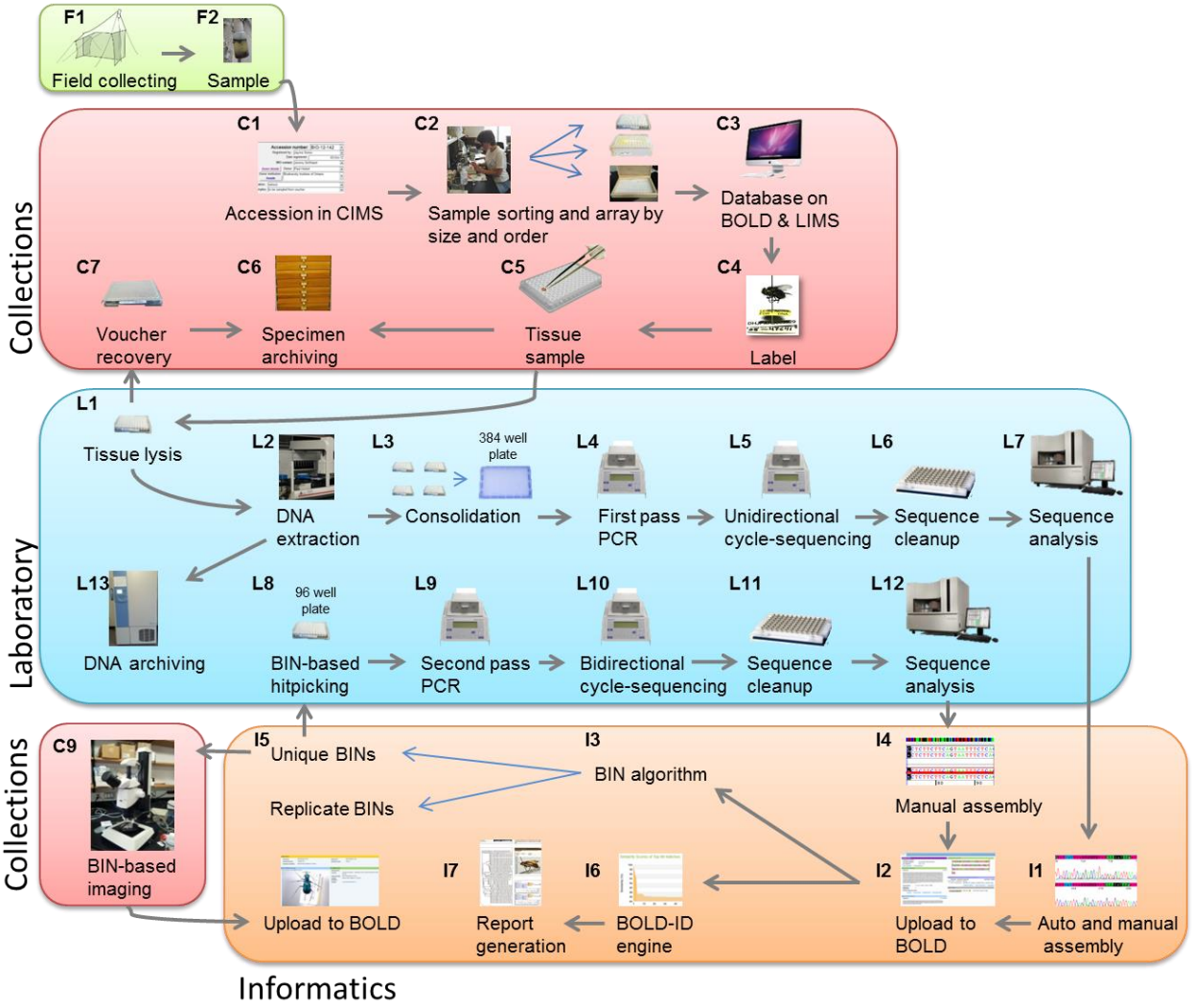


Figure 2. Schematic diagram showing the specimen workflow. Front end processing begins with field collecting (F1) and proceeds through to archiving of specimens (C6). Laboratory analysis begins with tissue lysis (L1) through to sequence analysis (L12). The informatics workflow includes both manual (I4) and auto sequence assembly, and finishes with BIN assignments and subsequent imaging of each unique BIN (C9).

RESULTS

Malaise Trap Program

One sample from each site was chosen for processing, collected approximately from the beginning of August to mid-August. A total of 11,348 specimens were captured in these samples with over half the individuals being flies (Diptera), followed in abundance by bees, ants and wasps (Hymenoptera), moths and butterflies (Lepidoptera), and true bugs (Hemiptera; Figure 3). Additionally, an excess amount of certain Collembola and mite morphospecies were observed; approximately 1000 and 270 individuals respectively. These specimens, along with an extra 400 marsh beetles (family: Scirtidae) from a single sample collected from Naikoon, were excluded from processing. A total of 2271 BINs were documented from all 10 sites (Appendix 1 and 2) and 390 are new to BOLD as of November 2016. Taxonomic breakdowns of individual sampling sites are provided in Tables 1 and 2.

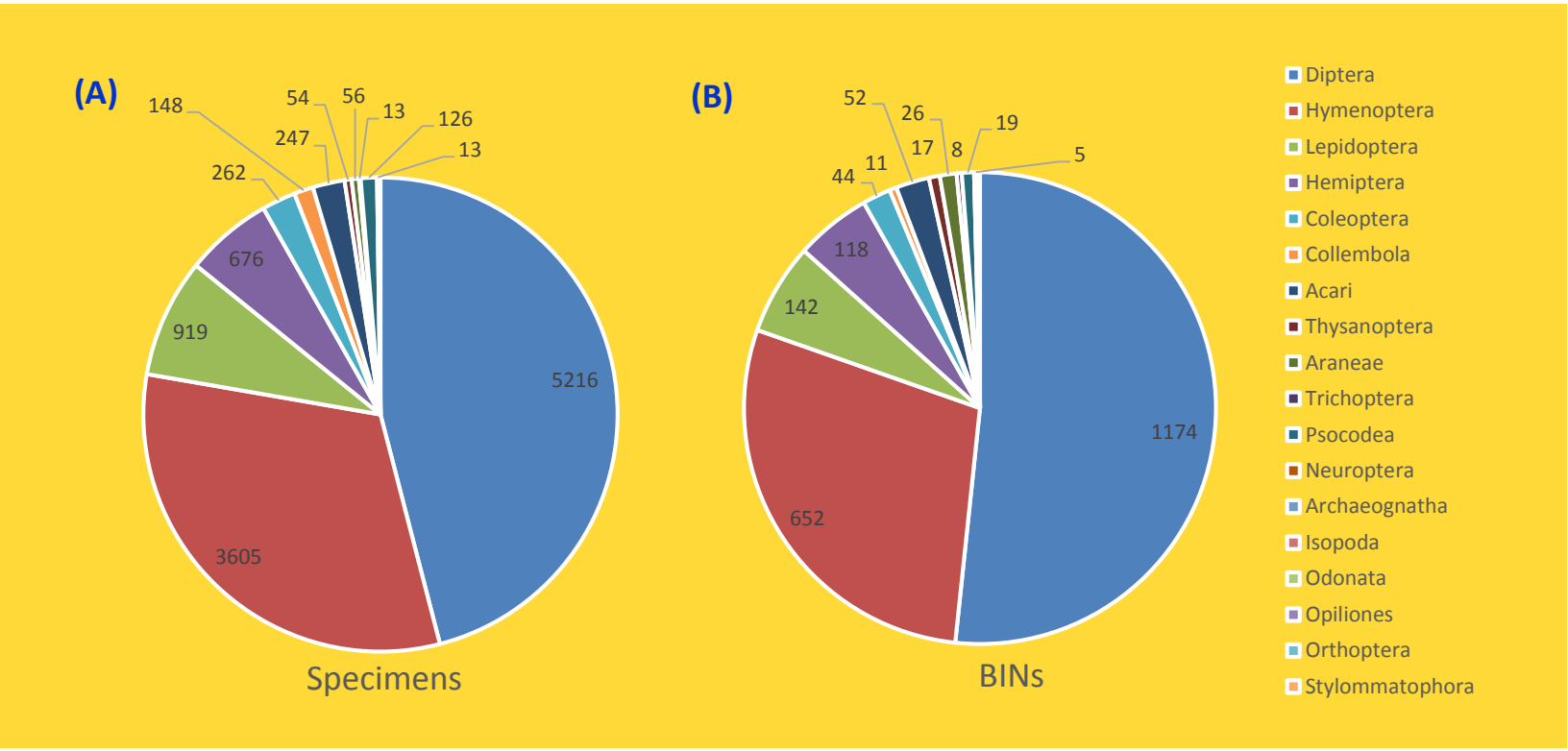


Figure 3. Taxonomic breakdown of (A) 11,348 total specimens processed and (B) 2271 total BINs collected in the BC Malaise Trap Program 2014.

The Chao species estimate suggest that approximately 4583 BINs are present in these sites and could be collected with this method if sampling effort was extended (Figure 4). The pattern of relative species abundance is quite typical, with a few species represented by many individuals (7 species with >100 individuals) – including 205 individuals of *Pigritia sp*., a species of micromoth – and a large number of species with few individuals (1281 singletons). Species richness extrapolation using the lognormal species abundance distribution suggests that 4596 BINs exist in the park (Figure 5).

Table 1. Taxonomy breakdown of specimens captured and processed from each site.

Taxon	Burnt Cabin	Det San	PFC	Naikoon	Goldstream	Skaha	White Lake	Mahoney	Miracle	Rathtrevor
Diptera	1924	667	669	772	555	188	214	134	69	24
Hymenoptera	967	738	457	342	510	286	151	76	54	24
Lepidoptera	205	123	26	3	22	297	166	59	10	8
Hemiptera	149	110	67	44	98	74	74	50	5	5
Coleoptera	74	7	39	106	9	10	5	3	4	5
Collembola	68	0	32	15	0	0	16	0	11	6
Acari	37	4	40	58	0	6	75	4	16	7
Thysanoptera	11	29	3	3	0	7	1	0	0	0
Araneae	9	5	6	8	0	5	1	3	18	1
Trichoptera	7	0	0	4	1	0	0	1	0	0
Psocodea	4	10	41	14	0	1	2	0	18	36
Neuroptera	2	2	0	0	1	1	5	0	1	1
Archaeognatha	0	0	0	0	0	0	0	0	1	0
Isopoda	0	0	0	0	0	0	0	0	0	3
Odonata	0	0	0	0	1	0	0	1	0	0
Opiliones	0	0	0	1	0	0	0	0	0	0
Orthoptera	0	0	0	0	0	3	0	1	0	0
Plecoptera	0	0	0	0	0	0	0	0	0	0
Stylommatophora	0	0	0	2	0	0	0	0	0	0
TOTAL	3457	1695	1380	1372	1197	878	710	332	207	120

Table 2. Taxonomy breakdown of BINs observed from each site.

Taxon	Burnt Cabin	Det San	PFC	Naikoon	Goldstream	Skaha	White Lake	Mahoney	Miracle	Rathtrevor
Diptera	520	185	125	186	180	92	80	51	36	9
Hymenoptera	243	148	68	76	72	83	57	34	16	5
Lepidoptera	46	35	10	3	9	22	34	20	4	2
Hemiptera	37	27	19	8	23	13	12	7	2	3
Acari	13	2	3	18	0	2	4	3	6	3
Coleoptera	10	1	10	9	6	4	1	2	3	4
Thysanoptera	7	7	2	1	0	4	1	0	0	0
Araneae	6	3	3	4	0	5	1	2	6	1
Trichoptera	5	0	0	2	1	0	0	1	0	0
Psocodea	3	4	7	2	0	0	1	0	6	4
Collembola	2	0	3	4	0	0	1	0	2	2
Neuroptera	2	2	0	0	0	0	1	0	0	0
Isopoda	0	0	0	0	0	0	0	0	0	1
Odonata	0	0	0	0	1	0	0	1	0	0
Orthoptera	0	0	0	0	0	1	0	1	0	0
Plecoptera	0	0	0	0	0	0	0	0	0	0
Stylommatophora	0	0	0	1	0	0	0	0	0	0
TOTAL	894	414	250	314	292	226	193	122	81	34

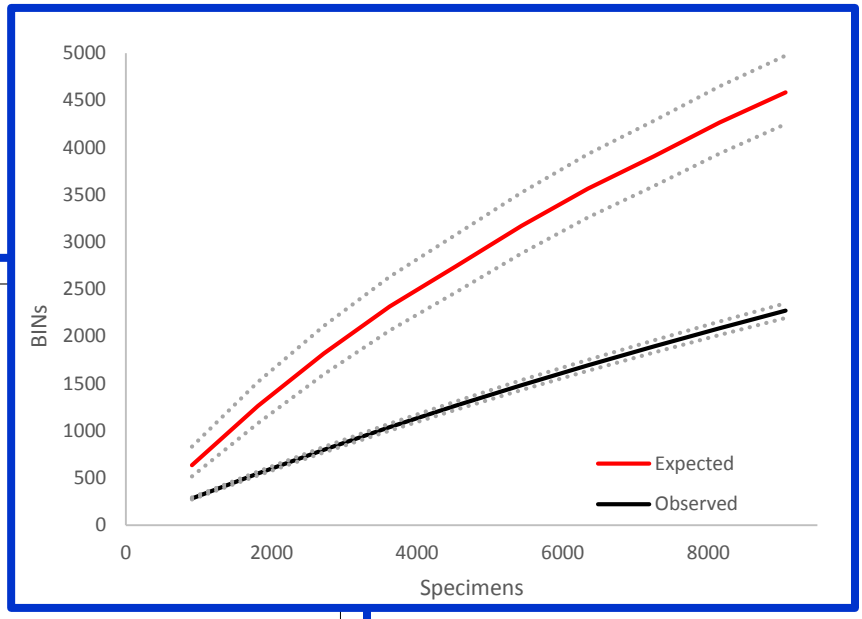


Figure 4. BIN accumulation curve for all specimens collected in the BC Malaise Trap Program.

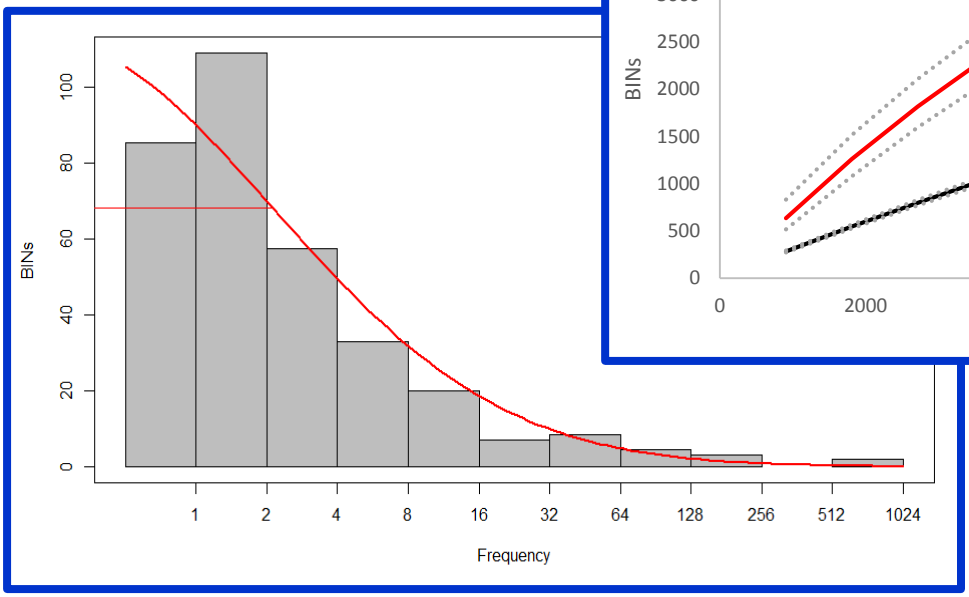


Figure 5. Lognormal species abundance curve showing the total BINs within each $\log_2$ abundance frequency interval (Preston 1962).

The average number of specimens collected per weekly sample was 1135. The number of individuals captured varied from a low of 120 at Rathtrevor Beach to a high of 3457 from Burnt Cabin Bog (Figure 6). The number of BINs observed in each sample was strongly influenced by sample size ($R^2 = 0.9749$, $p<0.05$). As expected, Burnt Cabin Bog also displayed the highest BIN count (N = 894) while Rathtrevor Beach captured the lowest BIN count (N = 34).

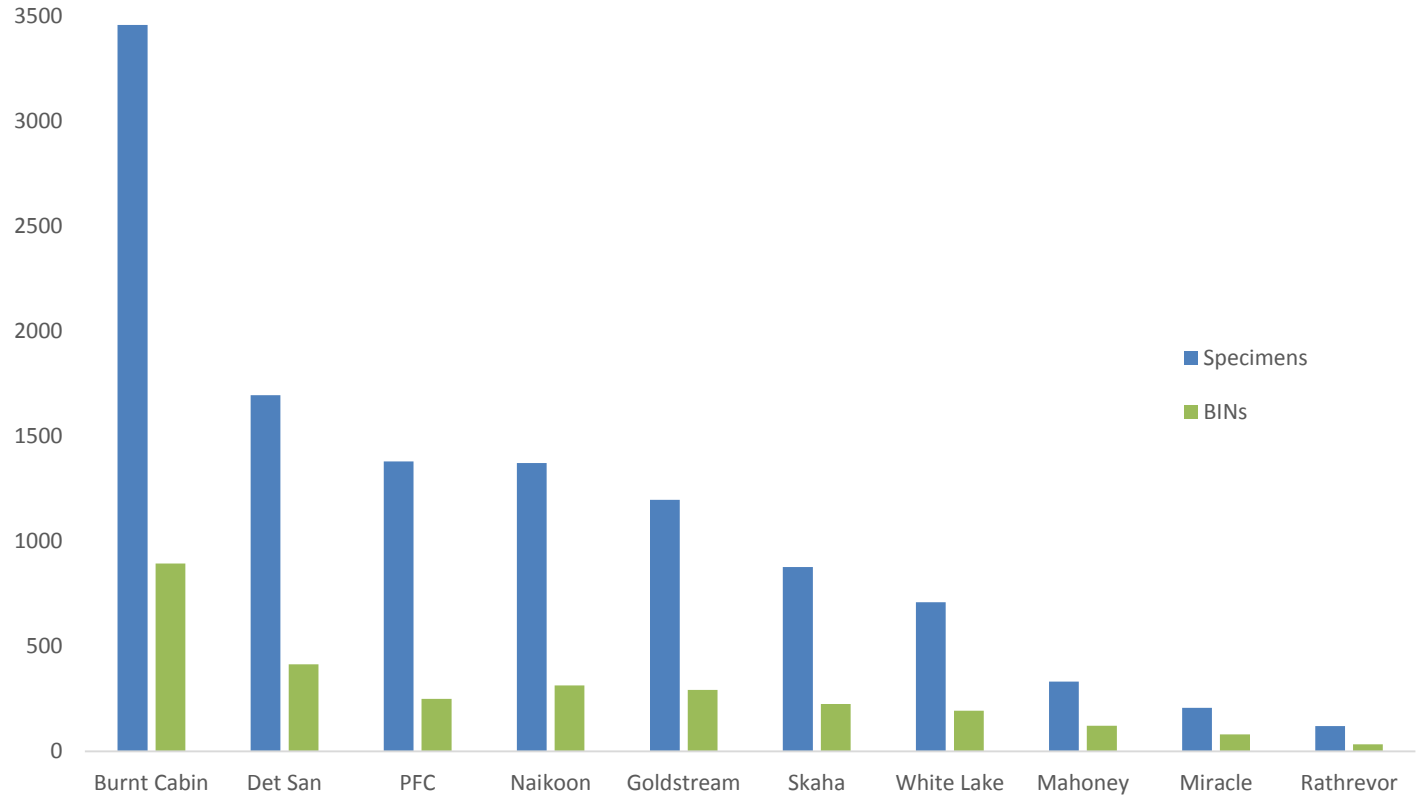


Figure 6. Total specimens and number of BINs generated from each of the 10 sampling sites.

The similarity in species composition between parks showed some variation (Figure 7). As expected, parks within a short geographical distance of each other typically shared a higher proportion of BINs than parks that were further apart (Figure 8). For example, PFC and Goldstream, only 11.4km apart, had the greatest species overlap (83 BINs) and a Chao's Sorensen Similarity Index of 0.306. Similarly, the 3 southern Okanagan parks had a Chao's Sorensen Similarity index of approximately 0.22 and are within 20km of each other. Det San and Burnt Cabin Bog, also approximately 11km apart, had a species overlap of 111 BINs but a smaller similarity index (0.17) due to the greater number of BINs collected at both parks. The insular park Naikoon on Haida Gwaii had the highest similarity index with Goldstream, which is also an insular park located on the southern point of Vancouver Island.

In total, 515 arthropod species were named, representing 23% of the BINs; 99.7% were assigned at least to family, and 54% of the BINs were assigned to a genus. Specimens collected from the BC Malaise Program represent 201 different families and 611 genera. Taxonomy reports for each sampling site are provided in Appendix 5. 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.

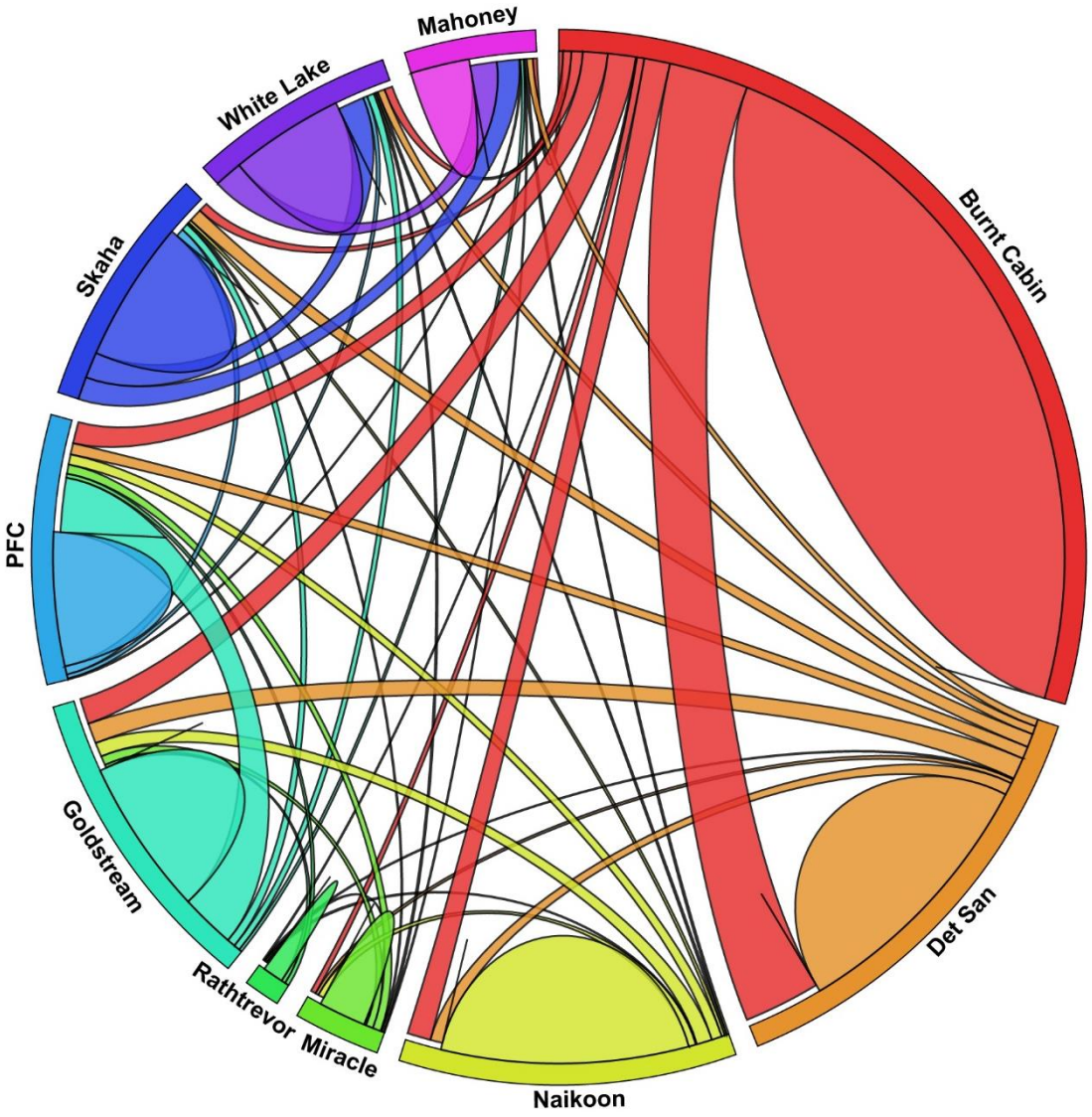
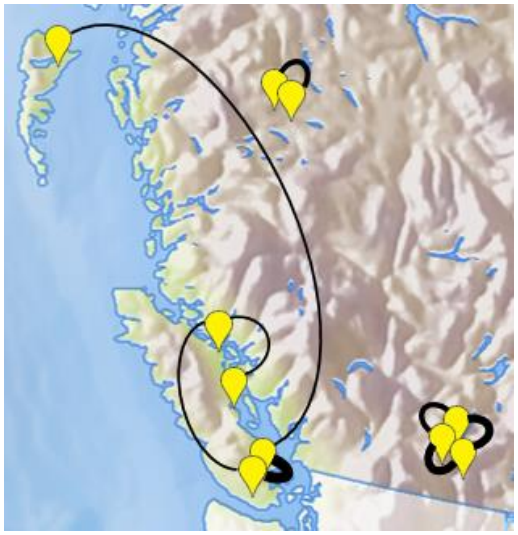


Figure 7. Chord diagram of species overlap between 10 sampling sites. The width of each wedge reflects the number of BINs captured in each site relative to the others. The widths of internal humps are proportional to the unique BINs within each site. Arcs connecting the sites reflect the proportion of shared species between any two sites, 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.



	Burnt Cabin	Det San	Goldstream	Mahoney	Miracle	Naikoon	PFC	Rathtrevor	Skaha
Det San	0.17								
Goldstream	0.066	0.088							
Mahoney	0.012	0.037	0.019						
Miracle	0.021	0.024	0.07	0.02					
Naikoon	0.06	0.047	0.076	0.014	0.056				
PFC	0.059	0.051	0.306	0.011	0.091	0.053			
Rathtrevor	0.004	0.009	0.037	0	0.07	0.011	0.049		
Skaha	0.027	0.063	0.046	0.201	0.02	0.019	0.055	0	
White Lake	0.022	0.04	0.041	0.267	0.022	0.012	0.027	0	0.196

Figure 8. Map indicating the highest similarity index values for each site. Similarity Indices displayed on table to the right, greater thickness of lines in map correspond to higher similarity index values.

Standardized Sampling Program

One site from each park was chosen for processing: the Rhododendron Flats in E.C. Manning, the Moor Bog in Burnt Cabin Bog, and the site on Kinaskan Lake Trail in Kinaskan Lake. An excess amount of Collembola was observed in the samples collected from the Rhododendron Flats and the Kinaskan Lake Trail (approximately 3830 individuals). There was also an excess of Cherry Leaf Beetles (*Tricholochmaea cavicollis*) observed in multiple Moor Bog samples (nearly 4K individuals), as well as ~300 excess sawfly larvae, ~300 excess ants, and nearly 1000 excess mites. All excess specimens were excluded from processing and are stored in a freezer archive. In total, 18,944 specimens were processed from 31 individual collecting events leading to the generation of 2626 BINs (Appendix 3 and 4).

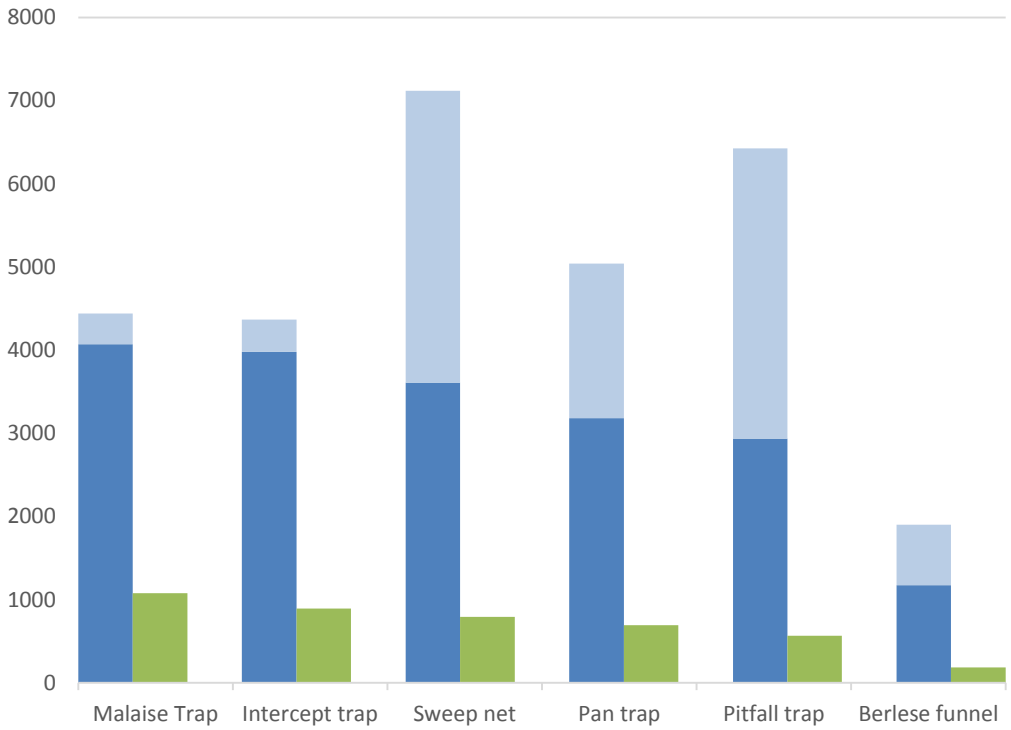


Figure 9. Total number of specimens (processed and excess) and BINs captured using different collecting methods in the BC SS Program 2014.

While Malaise traps did not capture more specimens than other trap types (Figure 9), they revealed a significantly higher proportion of the local fauna (41% of total BINs, and 31% 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 captures the highest volume of specimens, it requires 15 times more effort than Malaise traps to be comparable. Pitfall traps and pan traps also collected a high volume of specimens but these two traps require more regular servicing (2-3 times per week) while Malaise traps can be left for 1 to 2 weeks without servicing.

The taxonomic diversity captured with each method also varied (Figure 10). 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. A large amount of beetles and spiders were also captured using sweep nets.

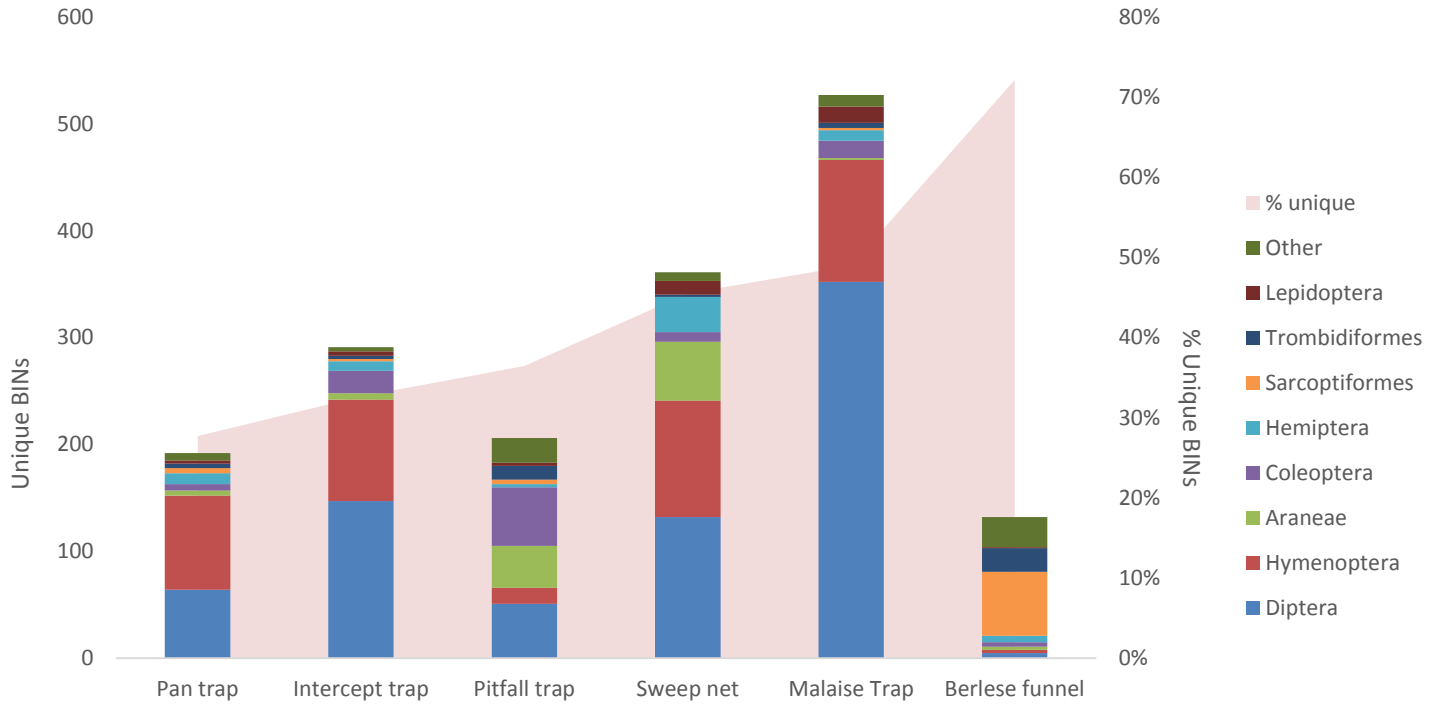


Figure 10. Total number of BINs unique to each collecting method and their taxonomic breakdown (bars) and the percentage of unique BINs collected with each method (unique BINs/total BINs; shaded area).

Of the 2626 BINs captured, more than half were unique to a single sampling method (N = 1709). The number of BINs unique to each method varied and the majority of unique BINs were captured in Malaise traps (Figure 10). It is important to note that although Berlese funnels collected the fewest BINs and unique BINs (132 out of 183 total), this method had the highest ratio of unique BINs to BINs captured (72%). The majority of these BINs belong to specimens from the subclass Acari (mites).

When comparing the preliminary SS results to the Malaise Program dataset, 413 out of the 2626 BINs are shared between the two and the total number of BINs combined is 4484. Only approximately 15-18% of

species overlap between programs indicating the importance of employing a variety of collecting methods to develop a more complete picture of arthropod fauna within an environment. While Malaise traps are the simplest and most cost-effective trapping method, other techniques greatly complement these samples. In particular, Berlese funnels capture BINs that otherwise would not be collected with other methods, including Malaise traps which targets flying arthropods rather than soil and ground specimens.

In total, 704 arthropod species were named, representing 28% of the BINs; 99% were assigned at least to family, and 59% of the BINs were assigned to a genus. Specimens collected from the BC SS Program represent 245 different families and 721 genera. A taxonomy report is provided in the Appendix 5. 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.

As this reference library for Canadian arthropods matures, the ability to conduct comprehensive terrestrial diversity assessments will strengthen. The next step involves sampling diverse environments and disturbance regimes, as well as to examine replicate samples. 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.

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.

APPENDICES

Appendices can also be accessed from http://biobus.ca/reports/.

Appendix 1. Neighbour-joining tree of representative specimens from each BIN collected by Malaise traps in the BC Malaise Program 2014 (colourized based on taxonomic order).

Appendix 2. Image library of 2263 (out of 2271) BIN representatives collected in the BC Malaise Program (in alignment with Appendix 1).

Appendix 3. Neighbour-joining tree of representative specimens from each BIN collected in the BC Standardized Sampling Program 2014 (colourized based on taxonomic order).

Appendix 4. Image library of 2600 (out of 2626) BIN representatives collected in the BC Standardized Sampling Program (in alignment with Appendix 3).

Appendix 5. Taxonomy reports for individual parks sampled in the BC Malaise Program and for the entire Standardized Program.

Appendix 6. Complete data spreadsheet of all specimens processed for the BC Malaise Program and Standardized Sampling 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 Malaise Trap Program and Standardized Sampling Program in BC. This project would not have been possible without the indispensable staff at the Centre of Biodiversity Genomics (formerly Biodiversity Institute of Ontario), and the many students who embarked on the cross country BIObus journeys. Special thanks to BC Parks for their partnership in the program, and parks staff and volunteers in particular for servicing the Malaise traps week after week.

Contacts:

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

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

Links:

BIObus	www.biobus.ca
Centre of Biodiversity Genomics	www.biodiversitygenomics.net
Barcode of Life Datasystems	www.boldsystems.org