

Centre for Biodiversity Genomics: State of the Archives



Angela C. Telfer, Suresh Naik, Jaclyn McKeown, Jayme E. Sones, Allison Brown, Meredith E. Miller, Evgeny V. Zakharov, Paul D.N. Hebert & Jeremy R. deWaard Centre for Biodiversity Genomics, University of Guelph, Canada

Introduction

The Centre for Biodiversity Genomics (CBG) at the University of Guelph is home to a globally unique natural history collection supporting innovative biodiversity research within Canada and internationally. Comprised of three interrelated resources – specimen archive, imaging facility, DNA archive – the CBG stores, creates and shares vital reference material for diagnosing and discovering species, determining their relationships, and tracking change over time. All specimen data, images, and sequences from these resources are linked through the Barcode of Life Data Systems (BOLD)¹.

Specimen Archive

CBG's specimen archive is comprised of two physically separated components – a dry and a fluid archive – with a total capacity of approximately **10 million** specimens. The dry collection contains mostly pinned specimens stored in cabinet drawers while the fluid specimens are stored in ethanol in herbarium cabinets. Every specimen has been digitized and assigned a unique Sample ID. The Sample ID and the corresponding specimen data are recorded in an internal collection management information system for quick reference and retrieval of voucher specimens. At present, the specimen archive contains **3,341,263** specimens, and of those, **81%** have an Barcode Index Number (BIN)². The specimens represent a diverse mix of **25** phyla, **62** classes, **286** orders and **1,859** families (Fig. 1). Using the BIN as a proxy for species, there are **199,701** species with **148,592** representing the only records for that BIN on BOLD. **61%** of the specimens were collected from Canada, including from **159** national/provincial parks and other protected areas.

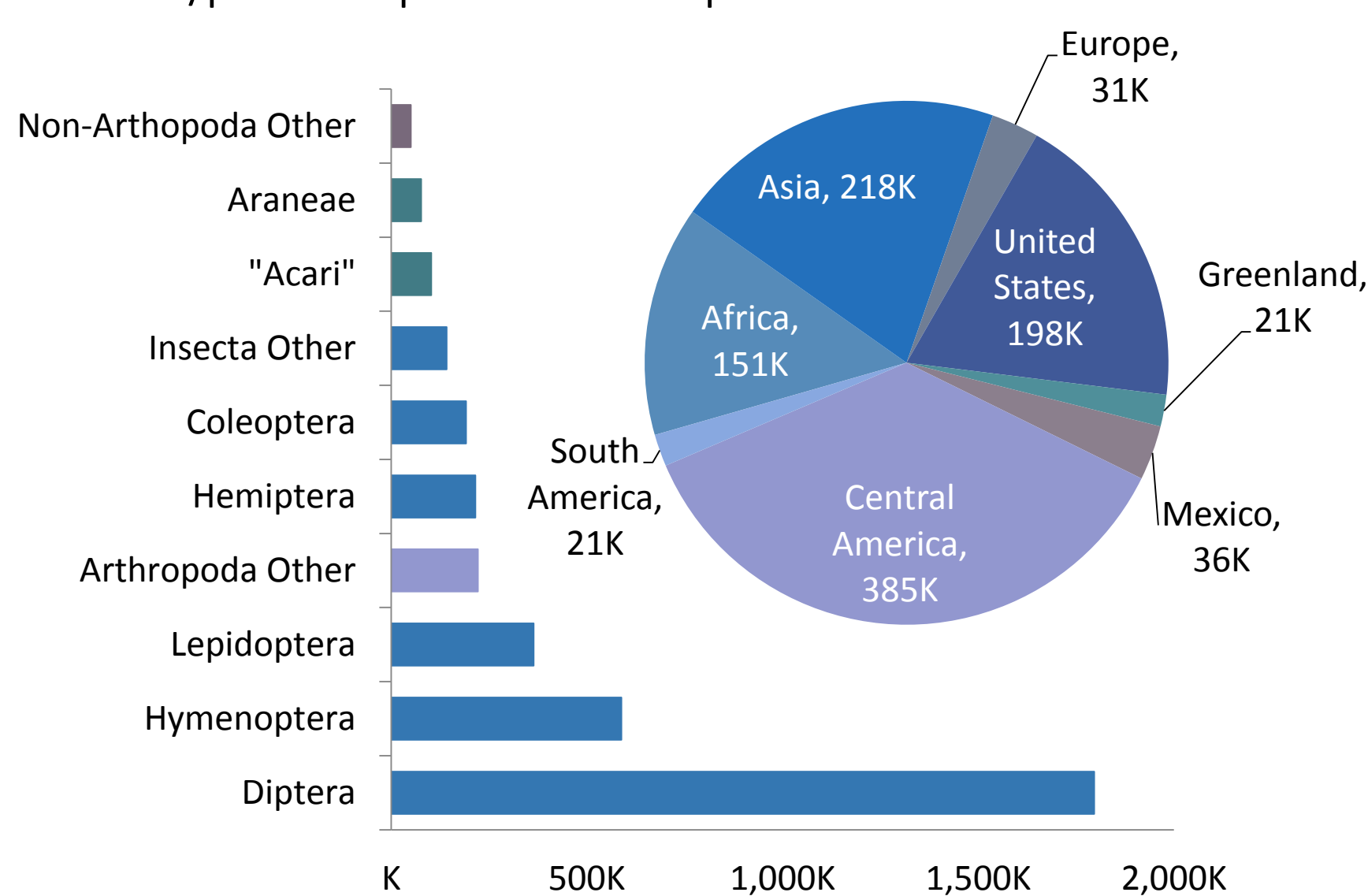


Figure 1: Orders and geographical regions, excluding Canada, represented in the specimen archive

Imaging Archive

The CBG imaging facility provides high-quality digital images for BOLD to support the identification and validation of barcoded specimens. At present, the imaging facility has **814,027** images on BOLD and **508,308** of these images are associated with specimens in the CBG specimen archive (Fig. 2). Approximately **60,000** images are taken each year utilizing a mix of full-time employees and undergraduate students. To ensure broad taxonomic coverage of the imaging, a pipeline is in place to prioritize the imaging of specimens representing newly established BINs. Since most BINs are represented by one or a few image(s), it is important that each specimen selected is an adult in good condition. Each image is z-stacked and consistent in orientation to maximize their quality and utility. All images are stored in BOLD and backed-up off-site. All CBG images have the copyright designation of Creative Commons - Attribution Non-Commercial Share-Alike (CC A-NC-SA).

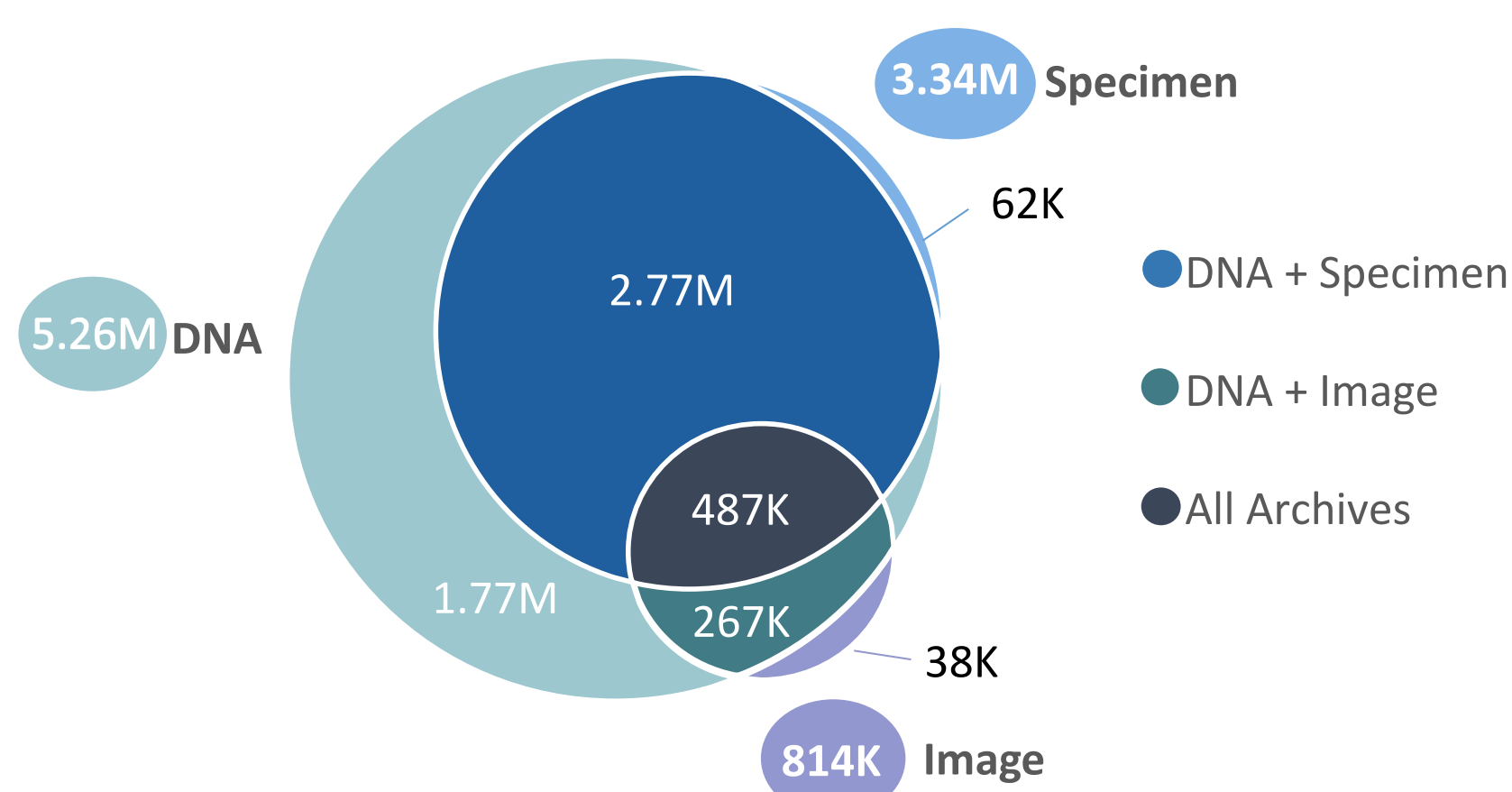


Figure 2: Overlap of records in each of CBG's archives

DNA Archive

CBG's DNA archive contains DNA extracts stored either as liquid in an ultra-cold -80°C freezer bank or dried in a trehalose-based preservative in -20°C freezers. Due to a finite storage capacity, a workflow is in place to prioritize which plates go into each archive. With a total capacity of **6.5 million** samples, the DNA archive is at **84%** capacity with **5,295,713** extracts. There is a heavy demand for these DNA extracts by the research community due to its unsurpassed taxonomic diversity and geographical scope (Table 1). These requests can be satisfied readily since all samples and plates are tracked in an internal database linked to CBG's laboratory information management system.

Table 1: Breakdown of the two freezer banks in the DNA Archive

	-80°C Freezer Bank	-20°C Freezer Bank
Storage Units	18	24
Capacity (plates)	27,476	41,500
Samples	2,394,303	3,081,012
Sequences	1,941,946	2,697,676
BINs	309,143	189,474
Phyla	52	16
Classes	144	31
Orders	702	111
Families	4,353	1,157
Genera	41,818	10,913
Countries/Oceans/Territories	246	59
With Images	1,885,433	240,914

Resource Accessibility

CBG's archives are highly connected resources. For example, **97%** of the vouchers in the specimen archive have high quality extracts in the DNA archive (Fig. 2). Similarly, **64%** of the digital photographs taken in the imaging facility are of vouchers located in the specimen archive. Together the resources in these archives represent **69%** of all the specimens on BOLD and contain at least one representative for **79%** of the BINs on BOLD. Notably, the archives contain the sole representative for **65%** of individual BINs on BOLD. CBG's specimens and DNA have proven to be valuable resources for external users from academic institutions, government agencies and private firms. Overall, **380,611** specimens and **207,335** DNA extracts have been loaned or donated to **263** institutions in **54** countries since 2008.

All CBG specimen vouchers, and their associated images and DNA sequence information are publicly available on BOLD. In 2015, to further increase the accessibility of their resources, CBG began releasing data to the Global Biodiversity Information Facility (GBIF) through its Canadian node, Canadensys (www.canadensys.net). To date, **1,005,634** Canadian specimen records have been published to the Canadensys repository. More recently, CBG has committed to making its DNA archive openly accessible, and will be releasing **1,500,515** Canadian specimen and DNA records to the Global Genome Biodiversity Network (GGBN; www.ggbn.org) in December 2017. These data will represent **183** orders, **1,161** families, **6,255** genera and **64,156** BINs. For both Canadensys and GGBN, the specimen data and sequences can easily be linked back to BOLD and GenBank via their Process IDs and GenBank accession numbers, respectively.

Acknowledgements

We would like to thank our colleagues at CBG – particularly Megan Milton, Kate Perez, Monica Young, Andrea Dobrescu, Crystal Sobel, Alex Borisenko, Sean Prosser, and Kirill Bessonov – for their roles in maintaining and supporting CBG's three archives.

References

- (1) Ratnasingham, S. & Hebert, P.D.N. (2013) doi:10.1371/journal.pone.0066213.
- (2) Ratnasingham, S. & Hebert, P.D.N. (2007). doi:10.1111/j.1471- 8286.2006.01678.x.