

# **GGI-CBG 'Barcoding NMNH Genera' Project – Final Report**

**September 6, 2019**

**PI/Co-PIs and Project Team:** Jonathan Coddington, Niamh Redmond, Bernardo Santos, Ashton Smith, Jessica Bird, Scott Miller, Meredith Miller, Margarita Miklasevskaja, Allison Brown, Crystal Sobel and Jeremy deWaard

**NMNH Curators involved:** Brady, S., Buffington, M., Chamorro, L., Dikow, T., Erwin, T., Flint, O., Gates, M., Henry, T., Konstantinov, A., Kula, R., McKamey, S., Micheli, C., Norrbom, A., Santos, B., Schultz, T. and Solis, A.

## **Methods:**

### **1) Specimen Selection and Loan Organization**

In 2018, staff from the Centre for Biodiversity Genomics (CBG) completed three visits to the Smithsonian Institution National Museum of Natural History, Department of Entomology (NMNH): Phase 1 (May 29<sup>th</sup> to June 8<sup>th</sup>), Phase 2 (September 17<sup>th</sup> to 27<sup>th</sup>) and Phase 3 (Dec 3<sup>rd</sup> to 14<sup>th</sup>). Forty-five arrays of 95 specimens each (4274 specimens total) were selected and loaned to CBG for processing. Additionally, 116 specimens from various orders were contributed by curator Scott Miller. Two representatives from each genus (wherever possible) that were new to both GGBN and GenBank were selected, following the GGI-CBG project guidelines. Taxonomy, country of collection, sample ID, and specimen cabinet/drawer locations were carefully recorded by CBG staff at the time of loan organization. A report of the species names, sample IDs, and country of collection of all specimens was given to the GGI project manager. The loan was approved by museum curators prior to the transport of the specimens to CBG.

### **2) Imaging, Digitization, Subsampling and Sequencing**

Once transferred to CBG, specimens were accessioned and labelled with Barcode of Life Database (BOLD) and USNM ENT labels prior to digitization. All necessary precautions were taken to prevent cross-contamination of and/or damage to the specimens during imaging and subsampling. Digitization, imaging, and tissue sampling for all 45 arrays were completed following pre-determined specifications by museum curators and uploaded to BOLD. DNA samples were extracted using the silica-based protocol outlined in Ivanova, deWaard & Hebert (2006; DOI: 10.1111/j.1471-8286.2006.01428.x). DNA samples were PCR amplified and sequenced following protocols detailed in Hebert et al. (2013; DOI: 10.1371/journal.pone.0068535) that target overlapping fragments of the cytochrome c oxidase I (COI) gene. The BOLD projects "Coleoptera - USNM 2018" (Phase 1 - SICOA), "Small Orders and Hymenoptera - USNM 2018" (Phase 2 - SICOB) and "Hemiptera, Lepidoptera, Coleoptera and Diptera - USNM 2018" (Phase 3 - SICOC) were created to store all specimen data, images, sequence data and associated files. Jonathan Coddington, Niamh Redmond, Michael Trizna, Bernardo Santos and Scott Miller were added to the BOLD projects as full-access users. Using specifications outlined by GGI staff, 475 specimens were selected to be re-analyzed using Next Generation Sequencing (NGS)-based failure-tracking (Prosser et al. 2016; DOI: 10.1111/1755-0998.12474).

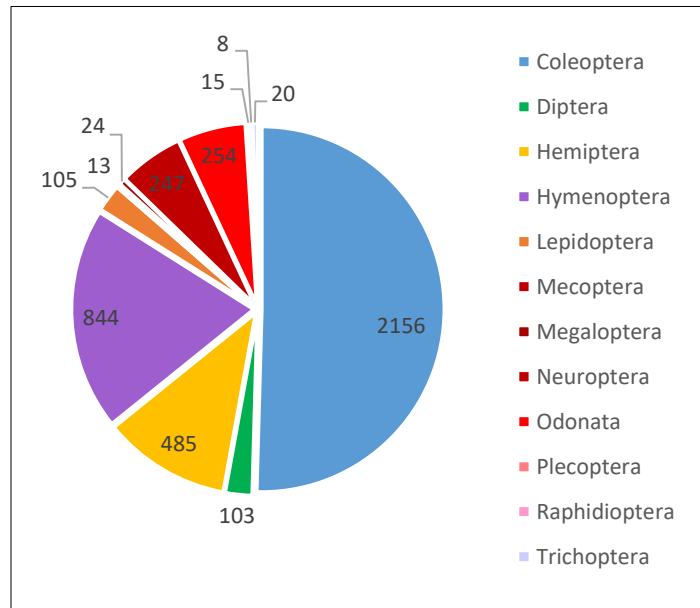
### **3) Data and Other Resources**

All specimen data and images were provided to the GGI project manager on a hard drive in April 2019. DNA bank data (following the GGBN Data Standard; Droege et al. 2016; DOI:

10.1093/database/baw125) were provided to the NMNH data manager for input into NMNH's EMu collection management system. Authorship of the specimen records was completed by CBG and GGI staff, and was sent to the NMNH data manager. DNA extracts were split (20 ul each) between the DNA archives of CBG and NMNH. Extracts for the NMNH were given to the GGI project manager in April 2019. All voucher specimens from Phase 1 (SICOA), Phase 2 (SICOB) and Phase 3 (SICOC) were returned to their original locations within the collection. All successfully sequenced records from the BOLD projects (> 200 bp) were submitted to GenBank (see Appendix 1) and will be moved to BioProject PRJNA81359 and made public by the GGI Project Manager. USNM voucher information will be listed in the "specimen voucher" field of all GenBank records, ensuring the correct linkage with records in the NMNH collections database (EMu).

## Results:

In total, 4274 specimens (45 arrays) were borrowed from NMNH between June and December 2018. Appendix 1 has a complete list of specimens and their associated data, downloaded from BOLD in July 2019 (Ratnasingham & Hebert, 2007). This represents 12 orders (Figure 1), 127 families, 2198 genera and 2556 identified species collected from 118 countries (Figure 2). As of July 2019, 2070 of the 2198 selected genera were new to GGBN, 1948 were new to GenBank and 1188 were new to BOLD. This constitutes 939 BINs with 58% (541 BINs) being unique to this project. Specimen collection dates (by decade) are summarized in Table 1 and Figure 3. Overall sequencing success was 60% (2562/4274) and is summarized in Tables 2 and 3 and Figures 4 and 5.



**Figure 1:** Taxonomic breakdown of specimens sampled

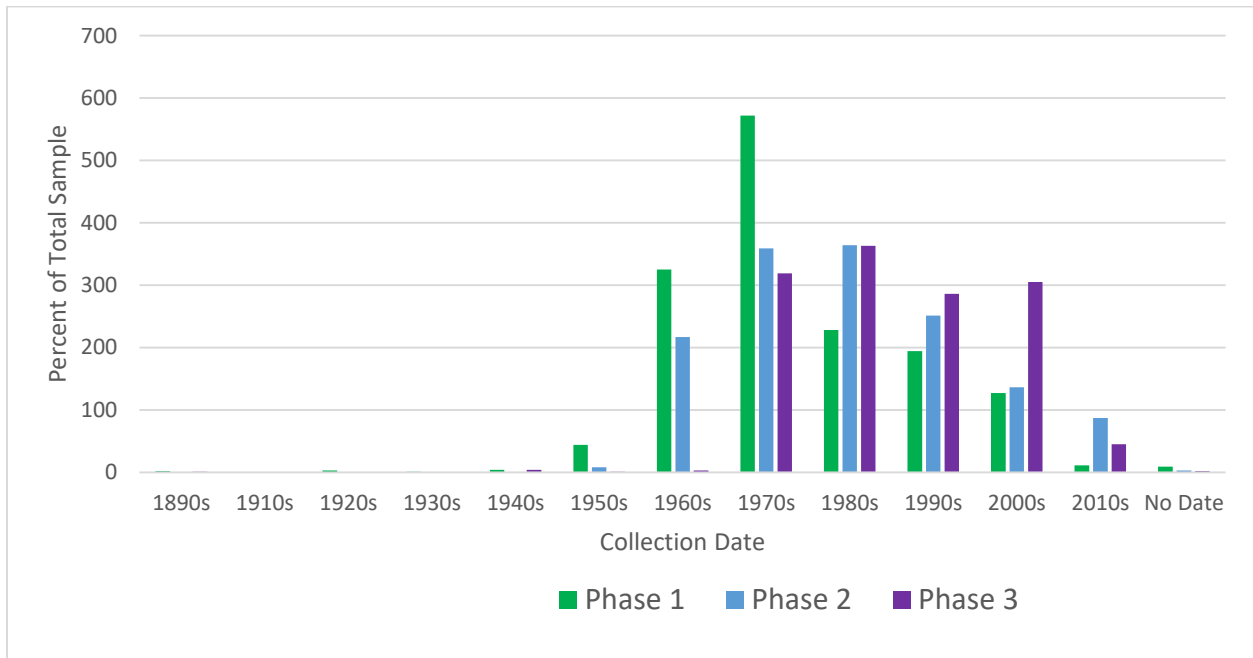


**Figure 2:** Map of collection locations for specimens borrowed from NMNH (BOLD, 2019)

Sequencing success by taxonomic group is summarized in Tables 4 to 15 and Figures 6 to 17. Since Coleoptera represent over half of the total specimens selected, results by family of Coleoptera are summarized in Tables 16 to 25 and Figures 18 to 27. Sequencing success of additional specimens (116 total, provided by Scott Miller) are summarized in Tables 26 to 27, and Figures 28 to 29. NGS-based failure-tracking of 475 specimens resulted in 370 recovered sequences (77.9%). Of the 370 specimens that gained a sequence, 291 (61.3%) were 300 bp or greater and 131 (27.6%) were 500 bp or greater. After validation, the 2489 successfully sequenced records were added to the private dataset DS-NMNH2019 titled 'Barcoding NMNH Insect Genera 2018-19' (which will receive the following DOI when made public: <http://dx.doi.org/10.5883/DS-NMNH2019>). In total, 4274 label images and 7709 specimen images were completed by CBG imaging technicians. Specimen images were largely habitus and/or dorsal view, with additional frontal images completed for all 726 Curculionoidea specimens. Specimen and label images (in TIF format and labelled with USNM ENT numbers) were provided to the GGI project manager in April 2019. Specimen images are also viewable on BOLD and in Appendix 2.

**Table 1:** Specimen Collection Dates

Collection Date Decade	Specimen Count	Percent of Total Samples
1890 – 1899	3	0.07%
1910 - 1919	0	0.00%
1920 - 1929	3	0.07%
1930 - 1939	1	0.02%
1940 - 1949	8	0.19%
1950 - 1959	53	1.24%
1960 - 1969	545	12.75%
1970 - 1979	1250	29.25%
1980 - 1989	955	22.34%
1990 - 1999	731	17.10%
2000 - 2009	568	13.29%
2010 - 2018	143	3.35%
No Year	14	0.33%
<b>Total</b>	<b>4274</b>	



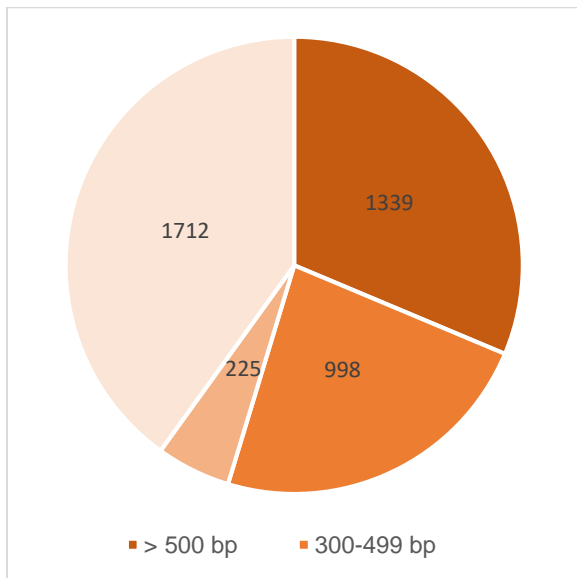
**Figure 3:** Collection Dates – Phase 1 (SICOA), Phase 2 (SICOB) and Phase 3 (SICOC)

**Table 2:** Overall Success

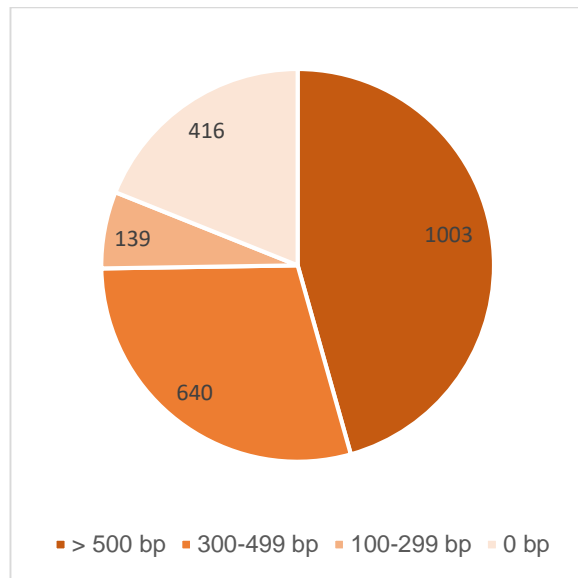
OVERALL SUCCESS – All Specimens		
<b>Total Records</b>	<b>4274</b>	
<b>&gt; 500 bp</b>	1339	31%
<b>300-499 bp</b>	998	23%
<b>100-299 bp</b>	225	5%
<b>0 bp</b>	1712	40%

**Table 3:** Overall Genera Success

OVERALL SUCCESS – Genera		
<b>Total Genera</b>	<b>2198</b>	
<b>&gt; 500 bp</b>	1003	46%
<b>300-499 bp</b>	640	29%
<b>100-299 bp</b>	139	6%
<b>0 bp</b>	416	19%



**Figure 4:** Overall Success



**Figure 5:** Overall Genera Success

## **RESULTS: Success by Taxonomic Group**

**Table 4: Diptera – Overall Success**

<b>OVERALL SUCCESS – All Specimens</b>		
<b>Total Records</b>	103	
<b>&gt; 500 bp</b>	40	39%
<b>300-499 bp</b>	21	20%
<b>100-299 bp</b>	1	1%
<b>0 bp</b>	41	40%

**Table 5: Diptera – Genera Success**

<b>OVERALL SUCCESS – Genera</b>		
<b>Total Genera</b>	53	
<b>&gt; 500 bp</b>	28	53%
<b>300-499 bp</b>	16	30%
<b>100-299 bp</b>	1	2%
<b>0 bp</b>	8	15%

**Table 6: Hemiptera – Overall Success**

<b>OVERALL SUCCESS – All Specimens</b>		
<b>Total Records</b>	465	
<b>&gt; 500 bp</b>	130	28%
<b>300-499 bp</b>	161	35%
<b>100-299 bp</b>	17	4%
<b>0 bp</b>	157	34%

**Table 7: Hemiptera – Genera Success**

<b>OVERALL SUCCESS – Genera</b>		
<b>Total Genera</b>	252	
<b>&gt; 500 bp</b>	94	37%
<b>300-499 bp</b>	103	41%
<b>100-299 bp</b>	14	6%
<b>0 bp</b>	41	16%

**Table 8: Lepidoptera – Overall Success**

<b>OVERALL SUCCESS – All Specimens</b>		
<b>Total Records</b>	105	
<b>&gt; 500 bp</b>	73	70%
<b>300-499 bp</b>	16	15%
<b>100-299 bp</b>	4	4%
<b>0 bp</b>	12	11%

**Table 9: Lepidoptera – Genera Success**

<b>OVERALL SUCCESS – Genera</b>		
<b>Total Genera</b>	56	
<b>&gt; 500 bp</b>	47	84%
<b>300-499 bp</b>	6	11%
<b>100-299 bp</b>	0	0%
<b>0 bp</b>	3	5%

**Table 10: Neuropteroids, Mecoptera, Odonata, Plecoptera & Trichoptera – Overall Success**

<b>OVERALL SUCCESS – All Specimens</b>		
<b>Total Records</b>	266	47%
<b>&gt; 500 bp</b>	152	27%
<b>300-499 bp</b>	19	3%
<b>100-299 bp</b>	131	23%
<b>0 bp</b>	266	47%

**Table 11: Neuropteroids, Mecoptera, Odonata, Plecoptera & Trichoptera – Genera Success**

<b>OVERALL SUCCESS – Genera</b>		
<b>Total Genera</b>	297	
<b>&gt; 500 bp</b>	181	61%
<b>300-499 bp</b>	83	28%
<b>100-299 bp</b>	13	4%
<b>0 bp</b>	20	7%

**Table 12: Hymenoptera – Overall Success**

<b>OVERALL SUCCESS – All Specimens</b>		
<b>Total Records</b>	827	
<b>&gt; 500 bp</b>	183	22%
<b>300-499 bp</b>	154	19%
<b>100-299 bp</b>	96	12%
<b>0 bp</b>	394	48%

**Table 13: Hymenoptera – Genera Success**

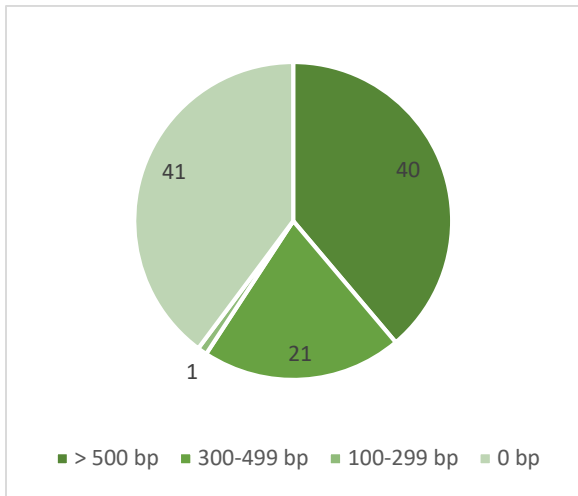
<b>OVERALL SUCCESS – Genera</b>		
<b>Total Genera</b>	439	
<b>&gt; 500 bp</b>	159	36%
<b>300-499 bp</b>	118	27%
<b>100-299 bp</b>	59	13%
<b>0 bp</b>	103	23%

**Table 14: Coleoptera – Overall Success**

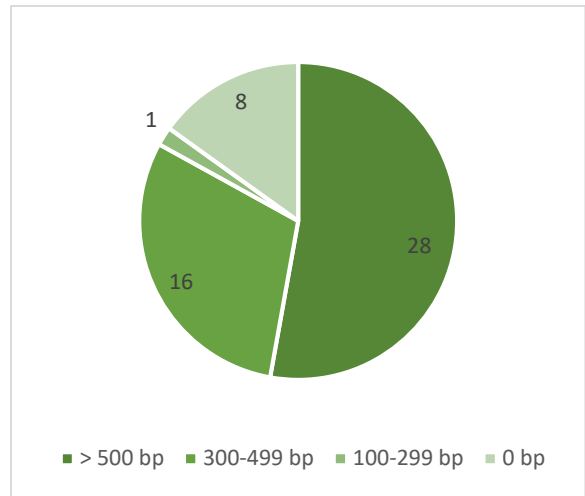
<b>OVERALL SUCCESS – All Specimens</b>		
<b>Total Records</b>	2090	
<b>&gt; 500 bp</b>	607	29%
<b>300-499 bp</b>	459	22%
<b>100-299 bp</b>	84	4%
<b>0 bp</b>	940	45%

**Table 15: Coleoptera – Genera Success**

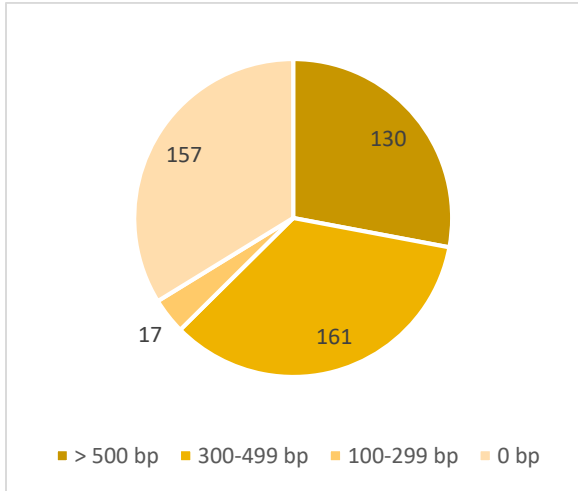
<b>OVERALL SUCCESS – Genera</b>		
<b>Total Genera</b>	1054	
<b>&gt; 500 bp</b>	472	23%
<b>300-499 bp</b>	297	14%
<b>100-299 bp</b>	51	2%
<b>0 bp</b>	234	11%



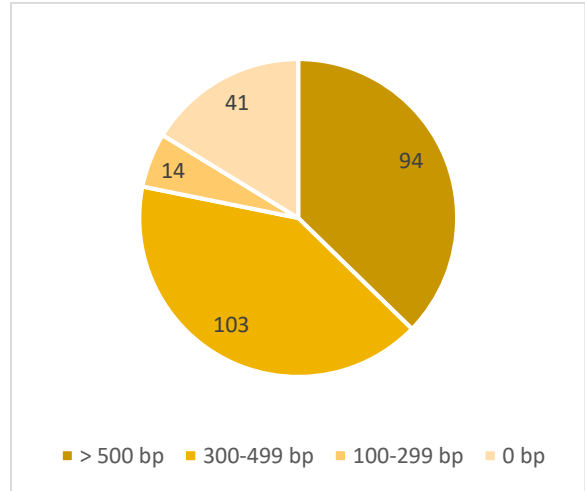
**Figure 6:** Diptera – Overall Success



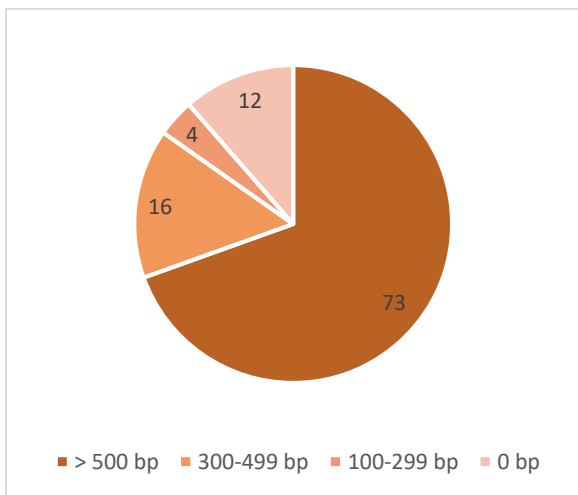
**Figure 7:** Diptera – Genera Success



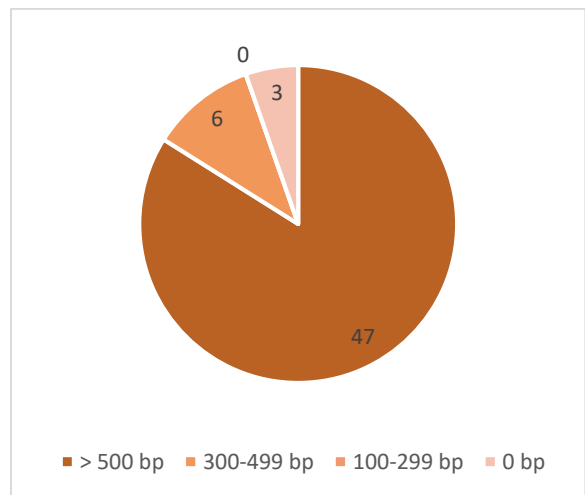
**Figure 8:** Hemiptera – Overall Success



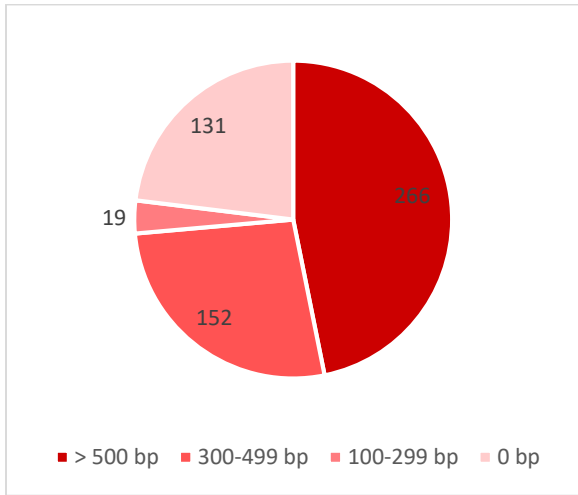
**Figure 9:** Hemiptera – Genera Success



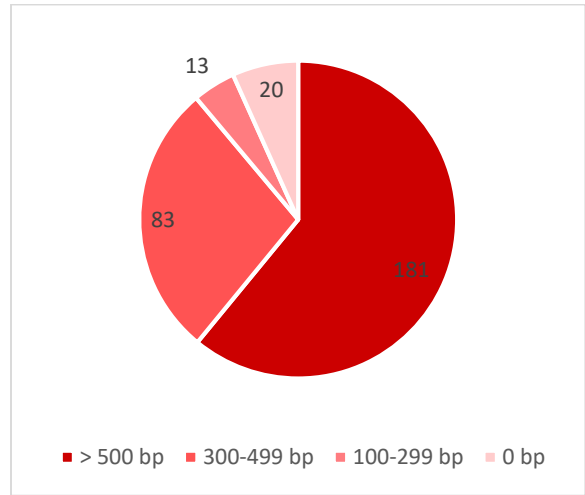
**Figure 10:** Lepidoptera – Overall Success



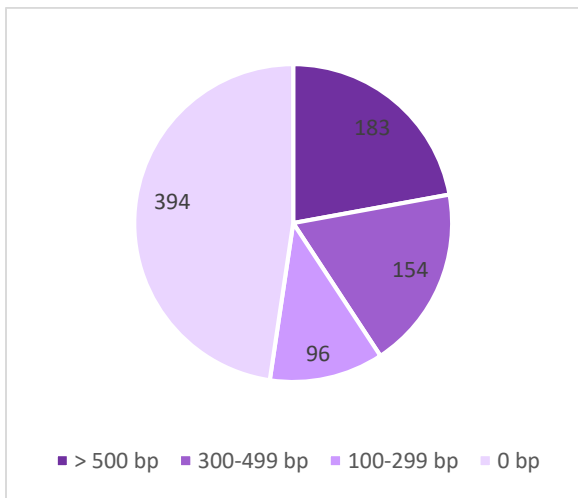
**Figure 11:** Lepidoptera – Genera Success



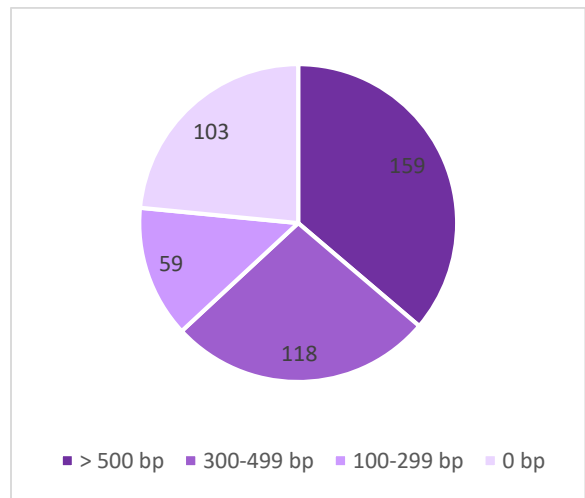
**Figure 12:** Neuropteroids, Mecoptera, Odonata, Plecoptera & Trichoptera – Overall Success



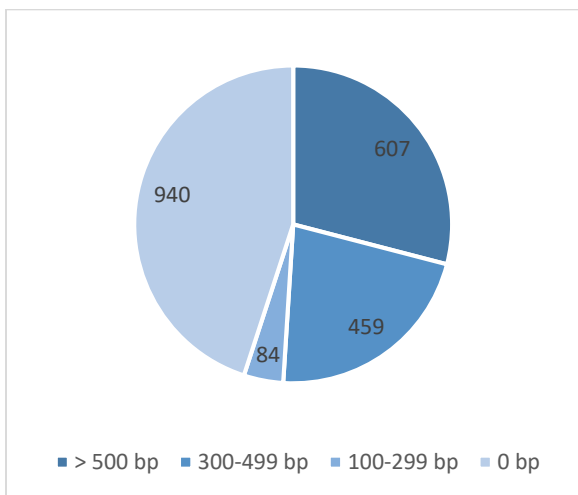
**Figure 13:** Neuropteroids, Mecoptera, Odonata, Plecoptera & Trichoptera – Genera Success



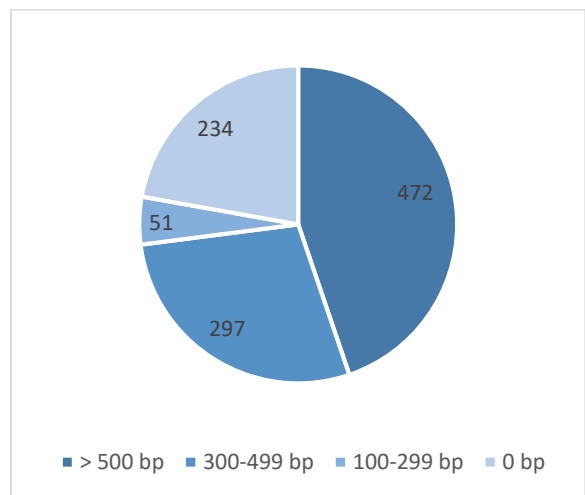
**Figure 14:** Hymenoptera – Overall Success



**Figure 15:** Hymenoptera – Genera Success



**Figure 16:** Coleoptera – Overall Success



**Figure 17:** Coleoptera – Genera Success

## RESULTS: Coleoptera Success by Family

Table 16: Curculionoidea – Overall Success

OVERALL SUCCESS – All Specimens		
Total Records	759	
> 500 bp	112	15%
300-499 bp	144	19%
100-299 bp	34	4%
0 bp	469	62%

Table 18: Carabidae & Cleridae – Overall Success

OVERALL SUCCESS – All Specimens		
Total Records	709	
> 500 bp	196	28%
300-499 bp	173	24%
100-299 bp	28	4%
0 bp	312	44%

Table 20: Chrysomeloidea – Overall Success

OVERALL SUCCESS – All Specimens		
Total Records	471	
> 500 bp	209	44%
300-499 bp	114	24%
100-299 bp	17	4%
0 bp	131	28%

Table 22: Cerambycidae – Overall Success

OVERALL SUCCESS – All Specimens		
Total Records	97	
> 500 bp	61	63%
300-499 bp	24	25%
100-299 bp	2	2%
0 bp	10	10%

Table 24: Scarabaeidae – Overall Success

OVERALL SUCCESS – All Specimens		
Total Records	51	
> 500 bp	27	53%
300-499 bp	4	8%
100-299 bp	3	6%
0 bp	17	33%

Table 17: Curculionoidea – Genera Success

OVERALL SUCCESS – Genera		
Total Genera	393	
> 500 bp	104	26%
300-499 bp	120	31%
100-299 bp	25	6%
0 bp	144	37%

Table 19: Carabidae & Cleridae – Genera Success

OVERALL SUCCESS – Genera		
Total Genera	332	
> 500 bp	154	46%
300-499 bp	103	31%
100-299 bp	16	5%
0 bp	59	18%

Table 21: Chrysomeloidea – Genera Success

OVERALL SUCCESS – Genera		
Total Genera	249	
> 500 bp	154	62%
300-499 bp	60	24%
100-299 bp	7	3%
0 bp	28	11%

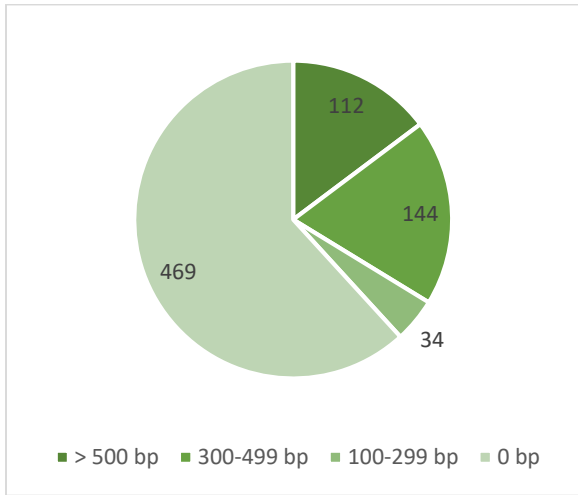
Table 23: Cerambycidae – Genera Success

OVERALL SUCCESS – Genera		
Total Genera	54	
> 500 bp	41	76%
300-499 bp	12	22%
100-299 bp	0	0%
0 bp	1	2%

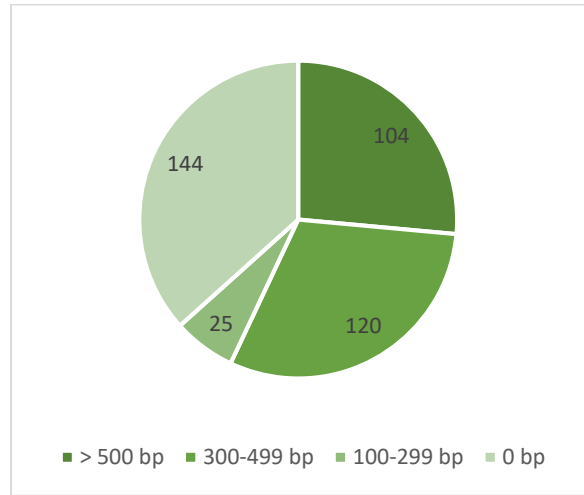
Table 25: Scarabaeidae – Genera Success

OVERALL SUCCESS – Genera		
Total Genera	25	
> 500 bp	18	72%
300-499 bp	2	8%
100-299 bp	3	12%
0 bp	2	8%

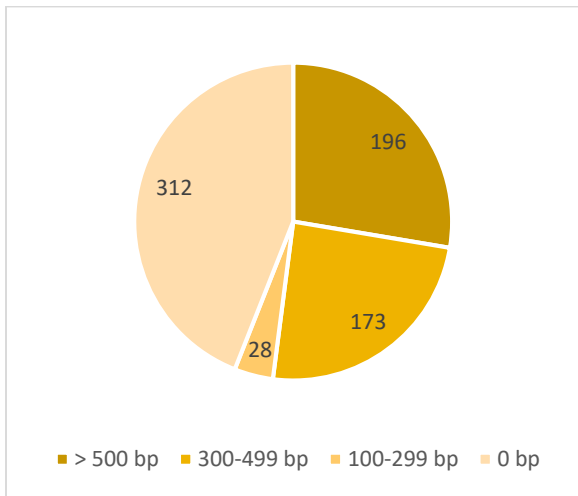




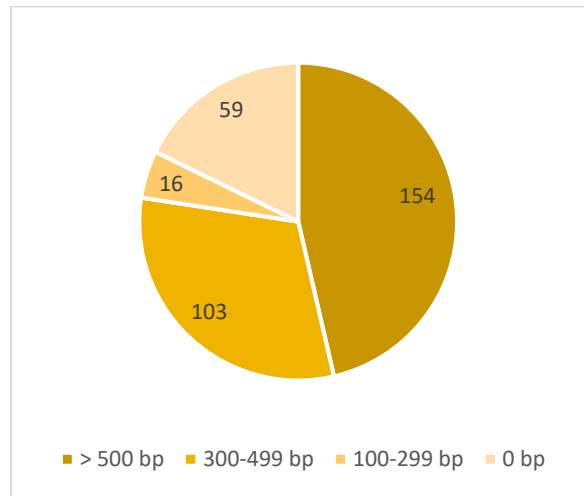
**Figure 18:** Curculionoidea – Overall Success



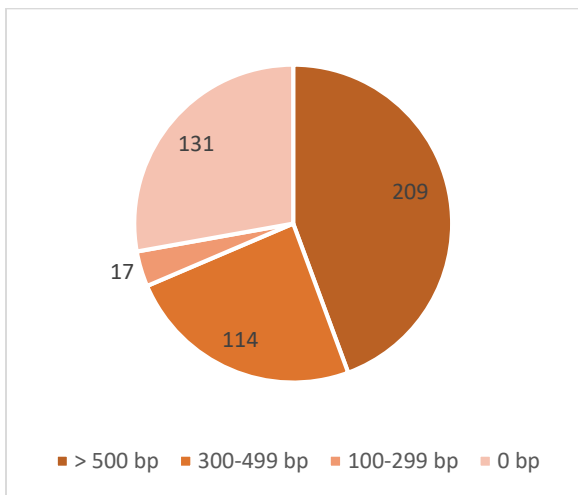
**Figure 19:** Curculionoidea – Genus Success



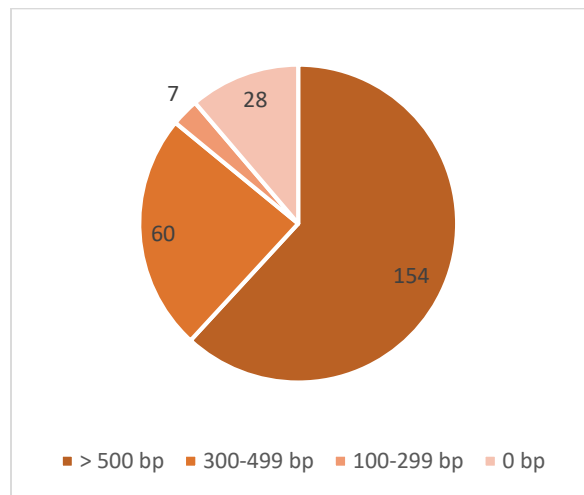
**Figure 20:** Carabidae/Cleridae – Overall Success



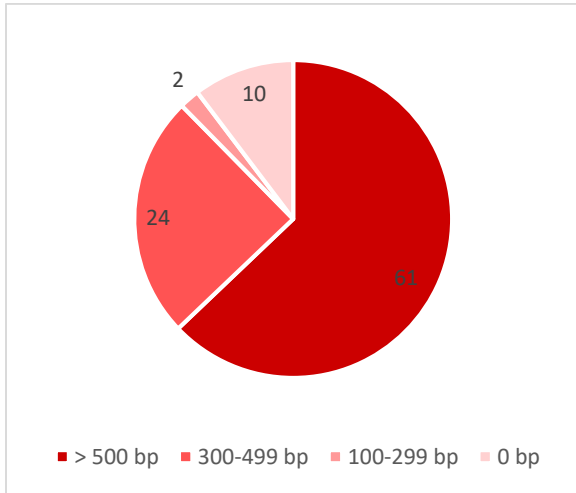
**Figure 21:** Carabidae/Cleridae – Genus Success



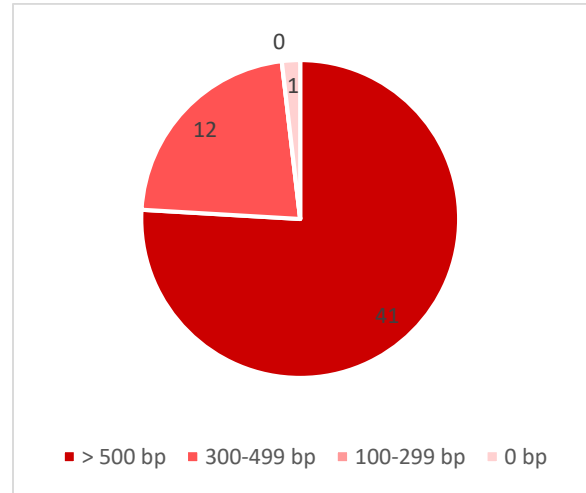
**Figure 22:** Chrysomeloidea – Overall Success



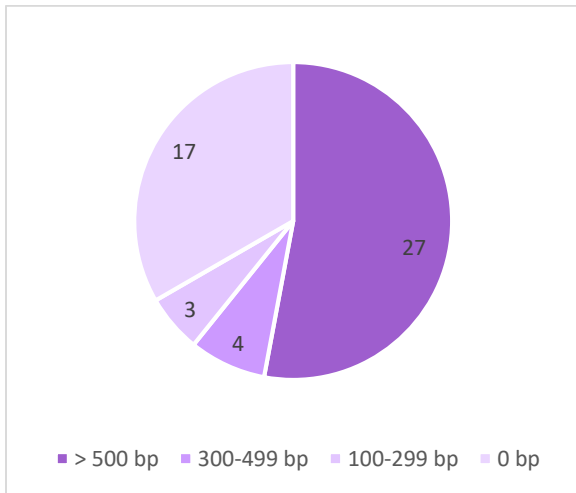
**Figure 23:** Chrysomeloidea – Genus Success



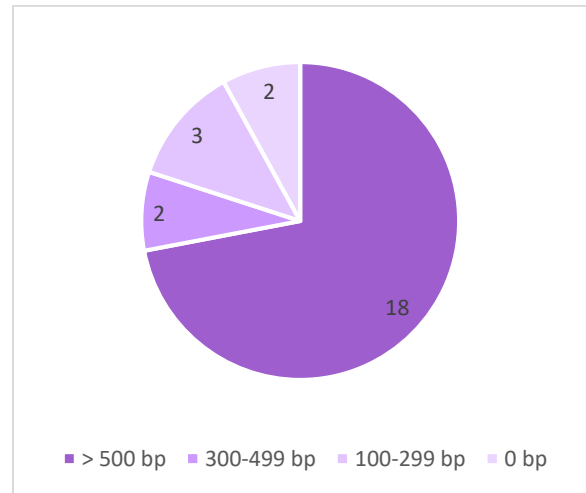
**Figure 24:** Cerambycidae – Overall Success



**Figure 25:** Cerambycidae – Genus Success



**Figure 26:** Scarabaeidae – Overall Success



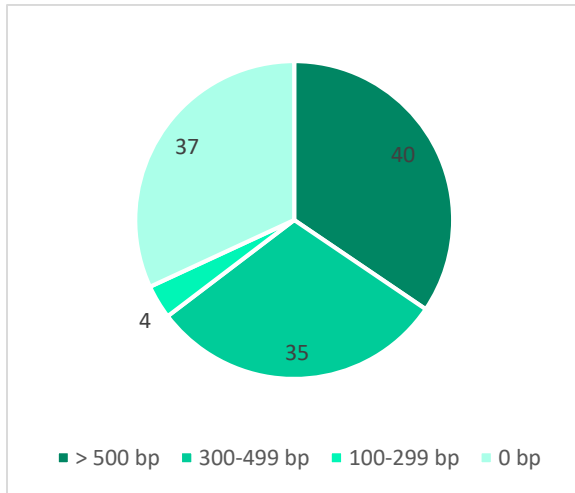
**Figure 27:** Scarabaeidae – Genus Success

**RESULTS: Additional Specimens (provided by Scott Miller)**

(Specimen Orders: Coleoptera, Hymenoptera, Hemiptera & Neuroptera)

**Table 26:** Additional Specimens – Overall Success

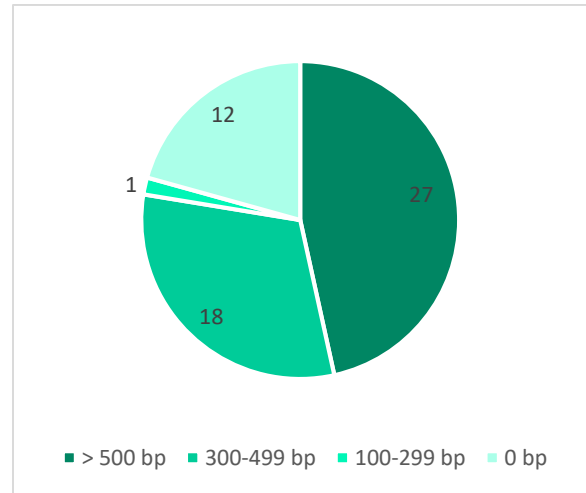
OVERALL SUCCESS – All Specimens		
Total Records	116	
> 500 bp	40	34%
300-499 bp	35	30%
100-299 bp	4	3%
0 bp	37	32%



**Figure 28:** Additional Specimens – Overall Success

**Table 27:** Additional Specimens – Genus Success

OVERALL SUCCESS – Genera		
Total Genera	58	
> 500 bp	27	47%
300-499 bp	18	31%
100-299 bp	1	2%
0 bp	12	21%



**Figure 29:** Additional Specimens – Genus Success

## **Year 2 Progress Update**

Staff from CBG completed the first of three visits to the NMNH from May 28<sup>th</sup> to June 6<sup>th</sup> 2019 (Year 2 Phase 1). Eighteen arrays of Hemiptera, Hymenoptera, Coleoptera specimens (1710 specimens total), were selected and loaned to CBG for processing (CCDB-34061 to CCDB-34074, CCDB-34077 to CCDB-34080). Fourteen of these arrays are being sequenced using the Sanger sequencing protocol, and four will be sequenced using Next-Generation Sequencing (NGS). Three of the Sanger plates (CCDB-34065, CCDB-34069 and CCDB-34072) and one NGS plate (CCDB-34079) will be processed as whole vouchers. This whole voucher protocol will involve removing specimens from points and placing them directly into sampling plates by CBG staff, performing voucher recovery, and re-pointing specimens by NMNH staff once returned to the museum. Two representatives from each genus (wherever possible) that are new to both GGBN and GenBank were selected, following the GGI-CBG project guidelines. Taxonomy, country of collection, sample ID, and specimen cabinet/drawer locations were carefully recorded by CBG staff. A report of the species names, sample ID's, and country of collection of all specimens was given to the GGI Project manager. The loan was approved by museum curators Tom Henry, Robert Kula, Sean Brady and Scott Miller prior to the transport of the specimens to CBG. The second visit (Phase 2) is planned for September 10<sup>th</sup> to 19<sup>th</sup> 2019, and the third visit (Phase 3) is planned for December 3<sup>rd</sup> to 12<sup>th</sup> 2019 to select the remaining 21 Sanger plates and 6 NGS plates (2565 specimens total). These trips will include selecting and borrowing specimens currently held at the NMNH Museum Support Centre (MSC). Pick lists for targeted taxa will be arranged with museum curators prior to future visits, and may include (but are not limited to) genera from the following orders: Hemiptera (remaining families within the Heteroptera), Lepidoptera (Saturniidae, Papilionoidea, etc.), Coleoptera (Scarabaeidae, Cerambycidae, Staphlinidae, Curculionoidea, etc.), Hymenoptera (Platygastroidea, etc.), Diptera (Tabanidae, Certopogonidae, Culicidae, Tipulomorpha, etc.) and Araneae.

## **Attached Appendices**

**Appendix 1:** BOLD data spreadsheet for the 4274 NMNH specimens analyzed in Year 1.

**Appendix 2:** Image library for the 4274 NMNH specimens analyzed in Year 1.