Data Article

Stool microbiome dataset of the critically endangered Puerto Rican parrot (*Amazona vittata*)

Miguel G. Rodriguez-Reyes\(^a\), Carlos Rios-Velazquez\(^b, *\)

\(^a\) Industrial Biotechnology Program, University of Puerto Rico at Mayagüez, Puerto Rico
\(^b\) Biology Department, University of Puerto Rico at Mayagüez, Puerto Rico

**Abstract**

The Puerto Rican parrot (*Amazona vittata*), endemic to Puerto Rico, is the only native parrot in the United States and is classified as a critically endangered species. There are two captive populations of *A. vittata* in Puerto Rico located in the Iguaca Aviary in El Yunque National Rainforest and the José L. Vivaldi Aviary in the Río Abajo Forest. To characterize the microbial communities of *A. vittata*’s stool, 21 stool samples from captive birds were collected, DNA extracted and sequenced using Illumina MiSeq. Sequences were processed by removing host sequences (*A. vittata* genome) and low-quality reads. Taxonomic and functional profiles were generated using MG-RAST. The most abundant domain was Bacteria (96%), followed by Virus (3%), and Eukaryota (0.6%). Among the functions in the microbiome, the most abundant was related to carbohydrates (14%), followed by clustering-based subsystems (12%), protein metabolism (8%), and amino acids and derivatives (7%). This dataset describes the stool microbiome of *A. vittata* using a metagenomics approach. Data can be used to develop holistic conservation strategies for *A. vit-
tata and other endangered birds, as well as to search for bioprospects with potential biomedical and biotechnological applications.

Published by Elsevier Inc.
This is an open access article under the CC BY-NC-ND license (http://creativecommons.org/licenses/by-nc-nd/4.0/)

Specifications Table

| Subject | Microbiology |
|---------|--------------|
| Specific subject area | Metagenomics |
| Type of data | FASTQ files, Figures |
| How data were acquired | Illumina MiSeq, MG-RAST |
| Data format | Raw, Processed |
| Parameters for data collection | Stool samples were collected from cages of captive *A. vittata* |
| Description of data collection | Metagenomic DNA was extracted from *A. vittata* stool samples and shotgun-sequenced using Illumina MiSeq. Sequences were annotated using MG-RAST. |
| Data source location | U.S. Fish and Wildlife Service Iguaca Aviary located in El Yunque National Forest in Rio Grande, Puerto Rico (18.3338, −65.7743). |
| Data accessibility | Raw data and annotations of this metagenome are available in the MG-RAST server under Study ID mgp94593 (https://www.mg-rast.org/linkin.cgi?project=mgp94593) and in NCBI database under BioProject PRJNA700670 (https://www.ncbi.nlm.nih.gov/bioproject/PRJNA700670). |

Value of the Data

- The dataset generated describes the stool microbiome of *A. vittata* using a metagenomics approach.
- Conservation scientists could benefit from this data to develop holistic conservation strategies that integrates the microbiome.
- This microbiome dataset of captive *A. vittata* could be compared to released specimens to detect potential microbiome shifts that could be associated to survival fitness after reintroduction into the wilderness.

1. Data Description

The Puerto Rican parrot (*Amazona vittata*) (Fig. 1), endemic to Puerto Rico, is the only native parrot in the United States [1]. It is categorized as Critically Endangered in the IUCN Red List and considered one of the ten most endangered birds in the world [1,2]. There are two *A. vittata* conservation facilities in eastern and west-central Puerto Rico: the Iguaca Aviary in El Yunque National Rainforest and the José L. Vivaldi Aviary in the Río Abajo Forest, respectively. As of December 2020, there were 215 captive and 8–12 released birds in El Yunque, and 204 captive and 130–140 released birds in Río Abajo, approximately (P. Torres-Báez in litt. 2021). For this study, stool samples from *A. vittata* specimens at Iguaca Aviary were collected. The stool metagenome was sequenced and described in terms of the diversity (Fig. 2) and functional (Fig. 3) profiles of the microbial communities, which represents the averaged stool microbiome with individual deviations from the composition treated as typical. Raw FASTQ files and figures data can be found under MG-RAST Study mgp94593 and NCBI BioProject PRJNA700670.
Fig. 1. *Amazona vittata* is small in size with average weight and length of 270 gs and 29 cm, respectively. Its main color is green, with traces of sky blue on the tip of the wings, a white ring around the eyes, and a red band above the beak. Credit: U.S. Fish and Wildlife Service, Tom MacKenzie.

Fig. 2. Taxonomic diversity of the *Amazona vittata* stool microbiome (931,628 hits against RefSeq database at 85% identity). The metagenome shows Bacteria as the most abundant domain (96%), followed by Virus (3%), Eukaryota (0.6%), other sequences (0.09%), and Archaea (0.002%). From the 26 bacterial phyla detected, the most abundant was Proteobacteria (58%), followed by Firmicutes (41%), Bacteroidetes (0.7%), Actinobacteria (0.3%), and other 22 phyla that represent the remaining 0.03%. Among the Bacteria, 99 orders were detected, from which the most abundant were Enterobacteriales (46%), followed by Lactobacillales (38%), Pseudomonadales (11%), Bacillales (2%), and other 95 orders that represent the remaining 3%.
Fig. 3. Functional profile of the Amazona vittata stool microbiome (404,686 hits against Subsystems-level 1 database at 85% identity). The most abundant function among A. vittata stool microbiome corresponded to carbohydrates (14%), followed by clustering-based subsystems (12%), protein metabolism (8%), amino acids and derivatives (7%), phages, prophages, transposable elements, plasmids (6%), miscellaneous (5%), membrane transport (5%), cell wall and capsule (5%), DNA metabolism (4%), among other categories which represent the remaining 34%.

2. Experimental Design, Materials and Methods

2.1. Sampling

Stool samples were collected from captive A. vittata specimens at Iguaca Aviary in El Yunque National Rainforest, Rio Grande, P.R. (18.3338, −65.7743). Twenty-one cages containing a pair of birds were sampled using sterile spatulas and microtubes. Stool collected was considered fresh due to short-term exposure to the environment. Samples were transported on ice until safely stored at −20°C. Birds were not touched or disturbed during sampling.

2.2. DNA extraction

Metagenomic DNA of the 21 samples was extracted individually using the FastDNA™ Spin Kit for Soil (MP Biomedicals) following the manufacturer’s protocol, except that a second ethanol wash was performed to remove impurities and humic acids, followed by an ethanol-sodium acetate DNA precipitation (2 vol. 95% EtOH, 0.10 vol. NaOAc 3 M pH 5.2) and resuspension in 50 μL of water.

2.3. Metagenome sequencing

The extracted DNA was pooled using 55 ng of each sample to a final normalized sample concentration of 26.3 ng/μL. To sequence the DNA at the Molecular Research DNA Laboratory (MR
DNA, Shallowater, TX, USA, www.mrdnalab.com), a genomic library was prepared with 50 ng
of DNA using the Nextera DNA Flex Library Preparation kit (Illumina) following the manu-
facturer’s protocol. After fragmentation and addition of adapter sequences through a limited-cycle
PCR, the final concentration of the library was 8.78 ng/µL using the Qubit® dsDNA HS Assay Kit
(Life Technologies) with an average length of 551 bp using the Agilent 2100 Bioanalyzer (Agi-
lent Technologies). The library was diluted to 10.0 pM and pair-end-sequenced using the MiSeq
Reagent Kit v3 for 600 cycles (Illumina).

2.4. Preprocessing sequences

The sequencing process yielded 1,840,483 paired-end reads. Those were preprocessed us-
ing FASTQC [3] as quality check, Cutadapt [4] to remove adapter sequences (CTGTCCTPATA-
CACATCT), and Sickle [5] to trim low-quality reads of Phred score <20. Bowtie2 [6] was used to
remove host sequences (1,806 reads) using A. vittata genome (Accession ID: GCA_000332375.1)
[7] as reference.

2.5. Taxonomic and function profiling

After preprocessing, 1,826,825 sequences of 210 ± 87 bp average length, totaling
384,525,358 bp, were uploaded to the Metagenomics Rapid Annotation using Subsystems Tech-
nology server (MG-RAST, www.mg-rast.org) [8]. In silico profiles of A. vittata’s stool microbiome
were generated in MG-RAST using representative hit and 85% identity. The profiles include tax-
onomic diversity (Fig. 2) and genes’ functionality (Fig. 3) based on RefSeq and Subsystems-level
1 reference databases, respectively.

Ethics Statement

This article is an original work of the authors. This research didn’t involve experiments in
animal subjects. Amazona vittata birds were not used for experimenting purposes. Stool on bird’s
cage floor was collected for sampling purposes; birds were not touched, disturbed, or harmed,
since researchers were not in direct contact with the birds.

CRediT Author Statement

Miguel G. Rodriguez-Reyes: Investigation, Formal analysis, Writing - Original Draft; Carlos
Rios-Velazquez: Supervision, Writing - Review & Editing.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal rela-
tionships which have, or could be perceived to have, influenced the work reported in this article.

Acknowledgments

This work was supported by the Maximizing Access to Research Careers Program (MARC
U*STAR), NIH-T34GM008419. Special thanks to the United States Fish and Wildlife Services staff
at the Iguaca Aviary under the Puerto Rican Parrot Recovery Program, in particular to Pablo
Thanks to Cristian Sanlatte-Reyes for helping with the collection of samples.

References

[1] U.S. Fish and Wildlife Service, Recovery Plan For the Puerto Rican Parrot (*Amazona vittata*), US Fish and Wildlife Service, Atlanta, GA, 2009.

[2] BirdLife International, *Amazona vittata*. The IUCN Red List of Threatened Species 2020: e.T22686239A179276011, 2020 https://dx.doi.org/10.2305/IUCN.UK.2020-3.RLTS.T22686239A179276011.en.

[3] S. Andrews, FastQC: A quality Control Tool For High Throughput Sequence Data, 2010. http://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

[4] M. Martin, Cutadapt removes adapter sequences from high-throughput sequencing reads, EMBOJ. 17 (1) (2011) 10–12.

[5] N.A. Joshi, J.N. Fass, Sickle: A sliding-window, adaptive, Quality-Based Trimming Tool For FastQ Files, 2011. https: //github.com/najoshi/sickle.

[6] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods 9 (4) (2012) 357–359.

[7] T.K. Oleksyk, J.F. Pombert, D. Siu, A. Mazo-Vargas, B. Ramos, W. Guiblet, Y. Afanador, C.T. Ruiz-Rodriguez, M.L. Nicker- son, D.M. Logue, M. Dean, L. Figueroa, R. Valentín, J.C. Martinez-Cruzado, A locally funded Puerto Rican parrot (*Amazona vittata*) genome sequencing project increases avian data and advances young researcher education, GigaSci 1 (1) (2012) 1–7.

[8] F. Meyer, D. Paarmann, M. D’Souza, R. Olson, E.M. Glass, M. Kubal, J. Wilkening, The metagenomics RAST server –a public resource for the automatic phylogenetic and functional analysis of metagenomes, BMC Bioinf. 9 (1) (2008) 386.