

# Nick Gabry

Data Scientist & Bioinformatician



linkedin.com/in/NGabry



github.com/NGabry



ngabry.github.io



N.T.Gabry@gmail.com



248.766.9365



Indianapolis, IN 46254

## SKILLS

### Primary Languages

*Proficient:* Python/Jupyter Notebooks

*Sufficient:* R/Rstudio, SQL, Bash/Zsh

### Technical Tools

*Data Management* - MySQL, sqlite, git, remote HPC, Unix/Linux shell tools

*Data Manipulation* - Pandas, NumPy, SciPy, dplyr, grep/awk/sed, SQLite, MS Excel

*Machine Learning* - scikit-learn, PyTorch, TensorFlow

*Data Visualization* - Matplotlib, Seaborn, ggplot

*Bioinformatics* - QIIME2, BedTools, SamTools, FASTX-Toolkit, FQC, Cutadapt, DADA2, RAxML, mafft, MEGAX, SeqIO

## EXPERIENCE

### Ph.D. Graduate Assistant - Biology • *Indiana State University*

Aug 2018 – Present

- Performed 3 next-generation sequencing runs from 150 nestling fecal samples and 70 bulk insect slurries generating close to 60,000,000 short read sequences for metabarcoding
- Developed a naive bayes taxonomic classifier using the BOLD API and scikit-learn implementation in QIIME2 to assign taxonomy to approximately 45,000,000 filtered short read sequences
- Reduced costs of experimentation by more than \$20,000 and eliminated common sequencing pitfall known as tag-jumps through design of novel dual indexed fusion-primers compatible with the Ion Torrent sequencing ecosystem
- Oversaw and assisted in the maintenance of a 30+ year long-term SQL database of genetic parentage for a breeding population of white-throated sparrows obtained through fragment analysis of micro-satellite alleles.

### M.S. Graduate Assistant - Computer Science • *Indiana State University*

May 2021 – Present

- Published MetaPlex, a bioinformatic toolkit, to BioConda package management channel
- Provided consultatory programming services to PhD and Masters level researchers on the analysis of single cell RNA-seq datasets
- Designed laboratory next-generation sequencing workflow and assisted in creating bioinformatic pipeline for *drosophila melanogaster* ChIP-seq experiments

## PROJECTS

### MetaPlex - NGS Read Processing Toolkit • [github.com/NGabry/MetaPlex](https://github.com/NGabry/MetaPlex)

- NGS metabarcoding workflow and read processing toolkit / software package published to BioConda

### K-Means Clustering of a Metacognitive Dataset • [github.com/NGabry/K-Means-Clustering-Metacognition](https://github.com/NGabry/K-Means-Clustering-Metacognition)

- Demonstrated K-means clustering algorithm from scratch for the Metacognition Assessment Scale (MAS-A)

### K-Nearest Neighbor Classifier for State of the Unions • [github.com/NGabry/SOTU-Classifer-KNN](https://github.com/NGabry/SOTU-Classifer-KNN)

- Built KNN classifier with sk-learn and NLP tools to assign presidential orator to State of Union speeches

## EDUCATION

**PH.D. BIOLOGY (Ecology)**  
Indiana State University  
Terre Haute, IN / **2023**

**M.S. COMP SCI (Bioinformatics)**  
Indiana State University  
Terre Haute, IN / **2023**

**B.S. PSYCHOLOGY (Biopsychology)**  
University of Michigan  
Ann Arbor, MI / **2017**