

# GEM3-sponsored training opportunities: Insights from the Smithsonian Conservation Genomics Bioinformatics Workshop

Morgan Calahan and Mosope Abanikannda



Smithsonian



Smithsonian-Mason School of Conservation

# Bioinformatics Analysis for Conservation Genomics Workshop

- March 20 - 29, 2023
- SMSC Campus in Front Royal, VA
- 5 GEM3 Attendees!
  - Boise State University
  - College of Western Idaho
  - Idaho State University
  - University of Idaho



Smithsonian



Smithsonian-Mason School of Conservation

# Purpose of the Course

1. Design conservation genomics projects
2. Handle genomic data
3. Use common population genomics software packages or tools
4. Interpret, visualize, and publish results
5. Practice collaborative research



# Instructors

- **Klaus Koepfli**, Ph.D., Senior Research Scientist, Smithsonian-Mason School of Conservation, George Mason University
- **Rebecca Dikow**, Ph.D., Data Scientist, Smithsonian Office of the Chief Information Officer Data Science Lab
- **Carlos Arias**, Ph.D., Data Scientist and Research Manager, Ecological and Evolutionary Genomics Lab (EEG), Smithsonian Tropical Research Institute (STRI)
- **Henrique Figueiro**, Ph.D., Postdoctoral Researcher, Smithsonian-Mason School of Conservation, George Mason University
- **Jennifer Spillane**, Ph.D., Postdoctoral Fellow, Smithsonian Office of the Chief Information Officer Data Science Lab
- **Michael Trizna**, Data Scientist, Smithsonian Office of the Chief Information Officer Data Science Lab
- **Ekaterina Noskova**, Researcher, Computer Technologies Laboratory, ITMO University, Russia



# SMSC Conservation Genomics Module

- Introduction to shell and hydra
- Clouded Leopard (*Neofelis nebulosa*)
- Genome assembly
- Assembly QC & Repeat annotation
- Variant Calling
- Popgenome
- Genetic Diversity
- Deleterious variants



# The Clouded Leopard (*Neofelis nebulosa*)



- Native to Southeast Asia; India, Nepal
- Adapted to life in the trees
- Hunt small mammals, birds, gibbons
- Solitary and secretive nature
- Conservation status: Vulnerable
- Habitat loss
- Poaching
- Lack of Data

# The Clouded Leopard (*Neofelis nebulosa*)



- Resequenced 6 mainland clouded leopards
- Whole genome and mitogenome resequencing
- Read alignment and variant calling
- Pedigree reconstruction
- Genetic structure
- Genetic diversity + Runs of homozygosity
- Historic  $N_e$

# GEM3 Genomics Workflow Module

[https://gem3genomics.github.io/Genomics\\_Workflow.github.io/index.html](https://gem3genomics.github.io/Genomics_Workflow.github.io/index.html)

- Learn how to process raw genomic data
- Produce summary statistics
- Initial annotation
- Call and identify sequence variants
- Common Unix / Linux commands
- Glossary of common genomics terms
- Repository with downstream analysis options





# GEM3 Genomics Workflow Module

Genomics Workflow

- Home
- Scan Raw Data
- Trim Adaptors
- Create Histogram
- Assemble with Various Data Types
- Assembly Statistics
- Assembly Completeness and Quality
- Mask Repetitive Elements
- Gene Prediction and Begin Annotation
- Call Variants
- Glossary
- Common Unix / Linux Commands
- Citations

- 1 Vision Statement**
- 2 What Will Be Covered?
- 3 What is Required?
- 4 Acknowledgements
- 5 Licensing

## Genomics Workflow



# GEM3

Genes by Environment  
Modeling · Mechanisms · Mapping

### 1 Vision Statement

This genomics workflow was created by Idaho EPSCoR [GEM3](#) participants and attendees of the 2023 Smithsonian School of Conservation Bioinformatics Analysis of Conservation Genomics Workshop with George Mason University through support of GEM3. The purpose of this workflow module is to share what was learned in this workshop and build capacity in conservation genomics and bioinformatics.

### 2 What Will Be Covered?

Please follow this workflow to learn about how to process raw genomic resequencing data, as well as downstream analysis actions. Further downstream analyses are captured in the following repository: [GEM3Genomics/Genomics\\_Downstream\\_Analyses](#).

# GEM3 Genomics Workflow Module: Downstream Analyses

The screenshot shows the GitHub interface for the repository 'Gem3Genomics / Genomics\_Downstream\_Analyses'. The repository is public and has 0 stars and 0 forks. The main branch is 'main'. The repository contains several files, including README.md, Admixture.Rmd, Admixture.html, Citations.Rmd, Citations.html, Citations.md, Demographic\_History\_via\_PSMC\_..., and Demographic\_Info\_PSMC.html. The 'About' section is empty, and there is one release labeled 'Original' (Latest) from last week.

**Repository Information:**

- Repository: Gem3Genomics / Genomics\_Downstream\_Analyses (Public)
- Stars: 0
- Forks: 0
- Unwatched: 1

**Branches and Tags:**

- main (selected)
- 1 branch
- 1 tag

**Files:**

File Name	Action	Last Commit
Gem3Genomics Update README.md		#190a7c last week (31 commits)
Admixture.Rmd	Add files via upload	last month
Admixture.html	Add files via upload	last month
Admixture.md	Update Admixture.md	3 weeks ago
Citations.Rmd	Add files via upload	3 weeks ago
Citations.html	Add files via upload	3 weeks ago
Citations.md	Update Citations.md	3 weeks ago
Demographic_History_via_PSMC_...	Add files via upload	last month
Demographic_History_via_PSMC_...	Add files via upload	last month
Demographic_Info_PSMC.html	Add files via upload	last month

**About:** No description, website, or topics provided.

**Releases:** 1 release

- Original (Latest) - last week

# GEM3 Genomics Workflow Module

[https://gem3genomics.github.io/Genomics\\_Workflow.github.io/index.html](https://gem3genomics.github.io/Genomics_Workflow.github.io/index.html)



# Outcomes

1. Collaboration with instructors and attendees from around the world
2. Open-source GEM3 Genomics Workflow Module Github
3. Attendees co-authoring clouded leopard genome / mitogenome annotation manuscript



# Acknowledgements



# GEM3

Genes by Environment  
Modeling · Mechanisms · Mapping



Smithsonian



Smithsonian-Mason School of Conservation

# GEM3-sponsored training opportunities: Insights from the Smithsonian Conservation Genomics Bioinformatics Workshop

Morgan Calahan and Mosope Abanikannda

Contact: [morgancalahan@boisestate.edu](mailto:morgancalahan@boisestate.edu), [mosopea@uidaho.edu](mailto:mosopea@uidaho.edu)