MitoPilot Overview

Dan MacGuigan

June 13, 2025

Introduction to MitoPilot

- A software package for mitogenome assembly and annotation from Illumina genome skimming data

- Developed by Devin Leopold (Jonah Ventures)

- Dan MacGuigan is leading the NMNH implementation and continued support

Goals of MitoPilot

- Parallelization of mitogenome assembly and annotation

- Curate and format annotation results to streamline GenBank submission

- Simplify execution across multiple computing environments

- Provide project management tools

- Improve reproducibility of analyses

- Single analysis platform for a broad range of taxa

MitoPilot architecture
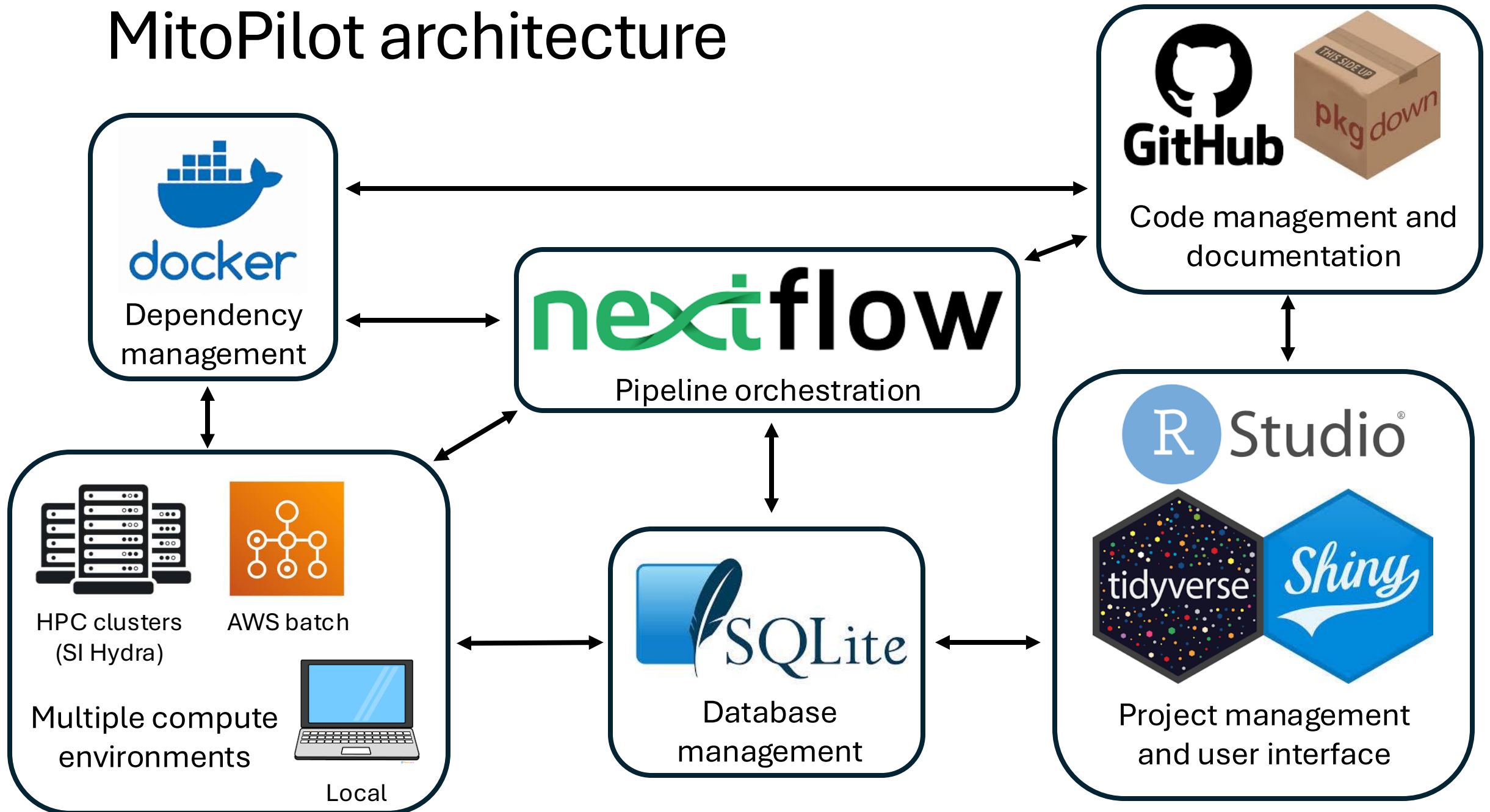
MitoPilot documentation website

https://smithsonian.github.io/MitoPilot/

Test project tutorial

Documentation
for all R functions

smithsonian.github.io/MitoPilot/

Instructions for using
MitoPilot on Hydra

MITOPILOT 0.0.0.9000   GET STARTED   REFERENCE   Articles ▾   Search for

Frequently Asked Questions
NMNH Hydra Setup
NOAA SEDNA Setup
Fish Mitogenome Curation
Building Custom Databases

OVERVIEW

DEVELOPERS

Devin Leopold
Author, maintainer

Dan MacGuigan
Contributor

Please see the underlined documentation site for more details.

MitoPilot is a package for the assembly and annotation of mitochondrial genomess
from genome skimming data. The core application consists of a Nextflow pipeline that
is wrapped in an R package, which includes an R-Shiny graphical interface to monitor
and interact with processing parameters and outputs. Currently the pipeline expects
paired-end Illumina reads as the raw input and performs the following steps:

1. Mitogenome assembly
   ○ fastp for quality control and adapter trimming

DEV STATUS

lifecycle experimental
R-CMD-check.yaml passing

MitoPilot pipeline overview

- Three workflow modules
  - Assemble
    - Read trimming with `fastp`
    - Assembly with `GetOrganelle`
    - Read mapping for coverage statistics with `bowtie2`
  - Annotate
    - Annotation of protein-coding genes with `Mitos2`
    - Annotation of tRNAs with `tRNAscan-SE`
    - Automatic curation with taxon-specific rulesets
    - Manual review and editing of annotations
  - Export
    - Custom scripts to parse results
    - Generates gene-by-gene alignment report, whole mitogenome FASTA, individual gene FASTAs, GenBank feature tables, and GFF formatted annotations
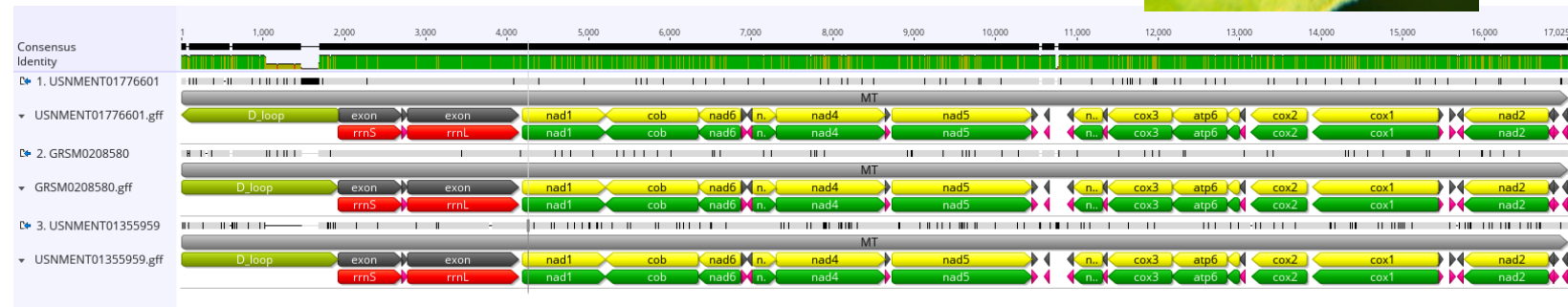
MitoPilot usage at the NMNH...so far

- Matt Girard has processed 1,328 fish samples
    - 269 published on GenBank

- Dan MacGuigan has processed 65 starfish (from Allen Collins & Chris Mah) and 4 dipteran (from Julia Steier) samples
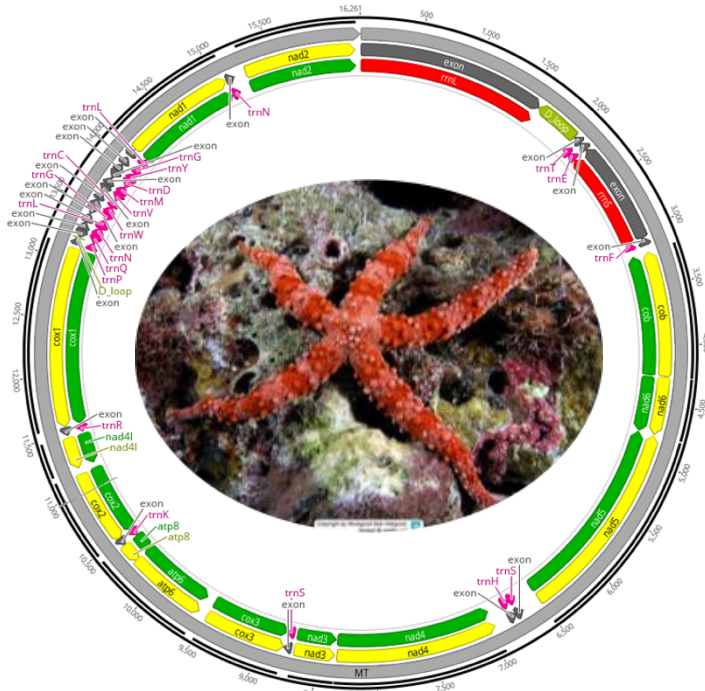
Nippled Sea Star (*Gomophia nardoa*)

tephritid fruit fly (*Strauzia*) mitogenomes from MitoPilot

Challenges of adapting MitoPilot for new taxa

- Need existing reference mitogenomes for assembly with GetOrganelle

- Taxa with "unusual" mitogenome features
  - Fragmented
  - Mitochondrial telomeres
  - Gene duplication
  - Pseudogenes
  - Bi-parental inheritance

- NCBI GenBank has an opaque, taxon-specific review processes