

A circular map of the human mitochondrial genome, approximately 16,500 base pairs in size. The map is color-coded: green for protein-coding genes, yellow for rRNA genes, and red for tRNA genes. The D-loop region is highlighted in grey. Genes shown include nad1, nad2, nad3, nad4, nad5, nad6, nad7, nad8, nad9, nad10, nad11, nad12, nad13, nad14, nad15, nad16, nad17, nad18, nad19, nad20, nad21, nad22, nad23, nad24, nad25, nad26, nad27, nad28, nad29, nad30, nad31, nad32, nad33, nad34, nad35, nad36, nad37, nad38, nad39, nad40, nad41, nad42, nad43, nad44, nad45, nad46, nad47, nad48, nad49, nad50, nad51, nad52, nad53, nad54, nad55, nad56, nad57, nad58, nad59, nad60, nad61, nad62, nad63, nad64, nad65, nad66, nad67, nad68, nad69, nad70, nad71, nad72, nad73, nad74, nad75, nad76, nad77, nad78, nad79, nad80, nad81, nad82, nad83, nad84, nad85, nad86, nad87, nad88, nad89, nad90, nad91, nad92, nad93, nad94, nad95, nad96, nad97, nad98, nad99, nad100. The map also shows the locations of exons and introns for several genes, including nad1, nad2, nad3, nad4, nad5, nad6, nad7, nad8, nad9, nad10, nad11, nad12, nad13, nad14, nad15, nad16, nad17, nad18, nad19, nad20, nad21, nad22, nad23, nad24, nad25, nad26, nad27, nad28, nad29, nad30, nad31, nad32, nad33, nad34, nad35, nad36, nad37, nad38, nad39, nad40, nad41, nad42, nad43, nad44, nad45, nad46, nad47, nad48, nad49, nad50, nad51, nad52, nad53, nad54, nad55, nad56, nad57, nad58, nad59, nad60, nad61, nad62, nad63, nad64, nad65, nad66, nad67, nad68, nad69, nad70, nad71, nad72, nad73, nad74, nad75, nad76, nad77, nad78, nad79, nad80, nad81, nad82, nad83, nad84, nad85, nad86, nad87, nad88, nad89, nad90, nad91, nad92, nad93, nad94, nad95, nad96, nad97, nad98, nad99, nad100. The map is labeled with base pair positions around the circle, from 0 to 16,500.

MitoPilot Overview

Dan MacGuigan
June 13, 2025

June 13, 2025

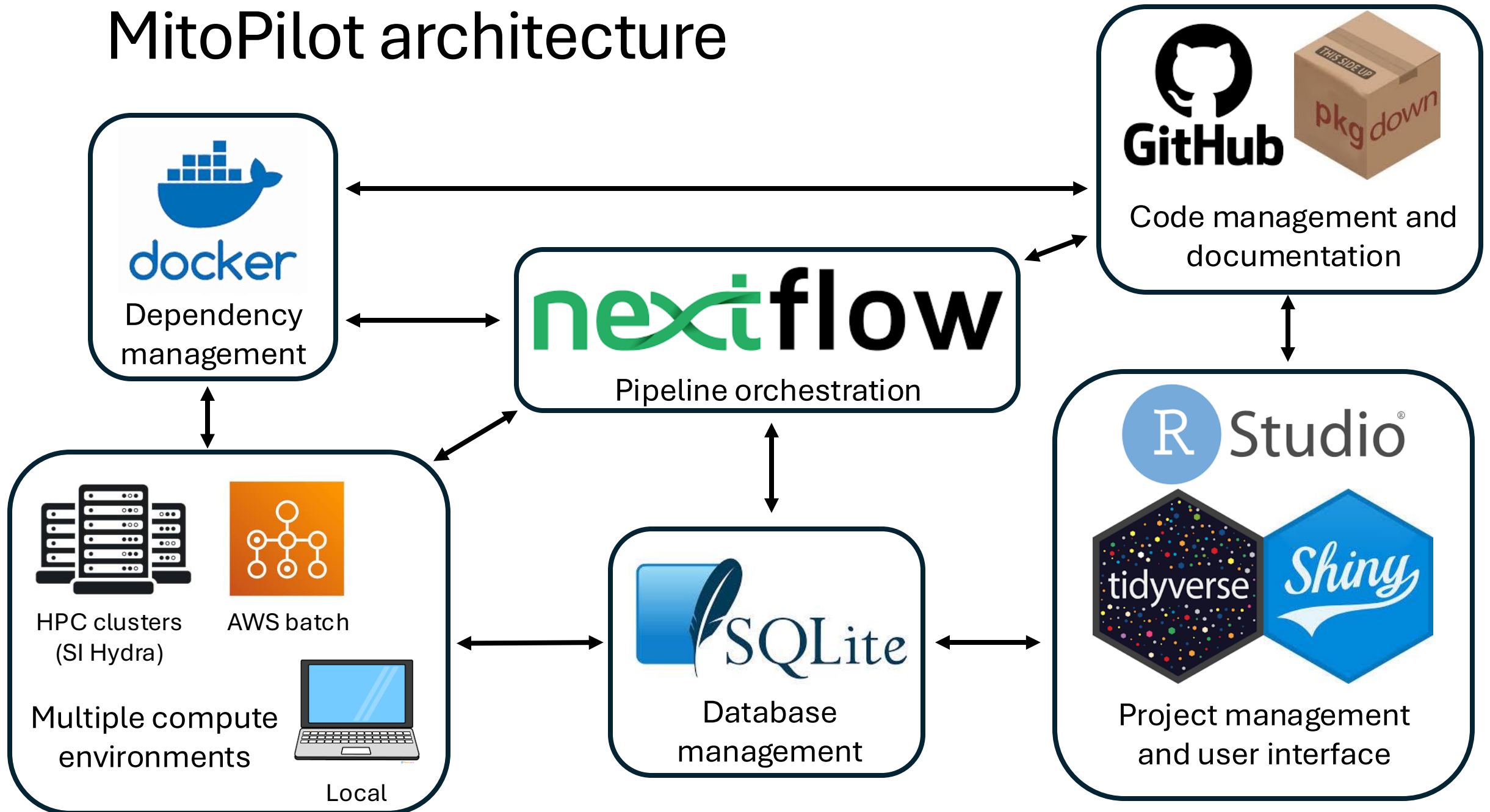
Introduction to [MitoPilot](#)

- 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/>

The screenshot shows the MitoPilot documentation website. At the top, a light blue header bar contains the URL 'smithsonian.github.io/MitoPilot/'. Below this, the main navigation area includes 'MITOPILOT 0.0.0.9000', two buttons 'GET STARTED' and 'REFERENCE' (both highlighted with black boxes), an 'Articles' dropdown menu (also highlighted with a black box), and a search bar labeled 'Search for'. The 'Articles' dropdown is open, showing a list of links: 'Frequently Asked Questions', 'NMNH Hydra Setup' (highlighted with a black box), 'NOAA SEDNA Setup', 'Fish Mitogenome Curation', and 'Building Custom Databases'. Annotations with arrows point to these elements: 'Test project tutorial' points to 'GET STARTED'; 'Documentation for all R functions' points to 'REFERENCE'; and 'Instructions for using MitoPilot on Hydra' points to 'NMNH Hydra Setup'. On the left, the 'OVERVIEW' section begins with the text 'Please see the [documentation site](#) for more details.' followed by a paragraph describing MitoPilot as a package for mitochondrial genome assembly and annotation, mentioning a 'Nextflow' pipeline and an 'R-Shiny' interface. On the right, the 'DEVELOPERS' section lists 'Devin Leopold' as the 'Author, maintainer' and 'Dan MacGuigan' as a 'Contributor'. Below this, the 'DEV STATUS' section shows two status bars: 'lifecycle experimental' (with 'experimental' in an orange box) and 'R-CMD-check.yaml passing' (with 'passing' in a green box).

Test project tutorial

Documentation for all R functions

Instructions for using MitoPilot on Hydra

smithsonian.github.io/MitoPilot/

MITOPILOT 0.0.0.9000

GET STARTED

REFERENCE

Articles ▾

Frequently Asked Questions

NMNH Hydra Setup

NOAA SEDNA Setup

Fish Mitogenome Curation

Building Custom Databases

Search for

OVERVIEW

Please see the [documentation site](#) for more details.

MitoPilot is a package for the assembly and annotation of mitochondrial genomes 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

DEVELOPERS

Devin Leopold
Author, maintainer

Dan MacGuigan
Contributor

DEV STATUS

lifecycle experimental

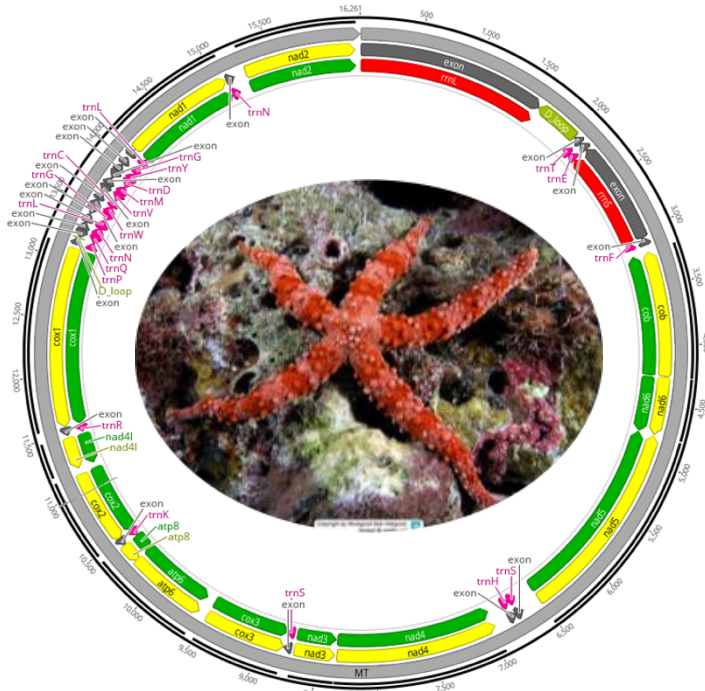
R-CMD-check.yaml passing

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

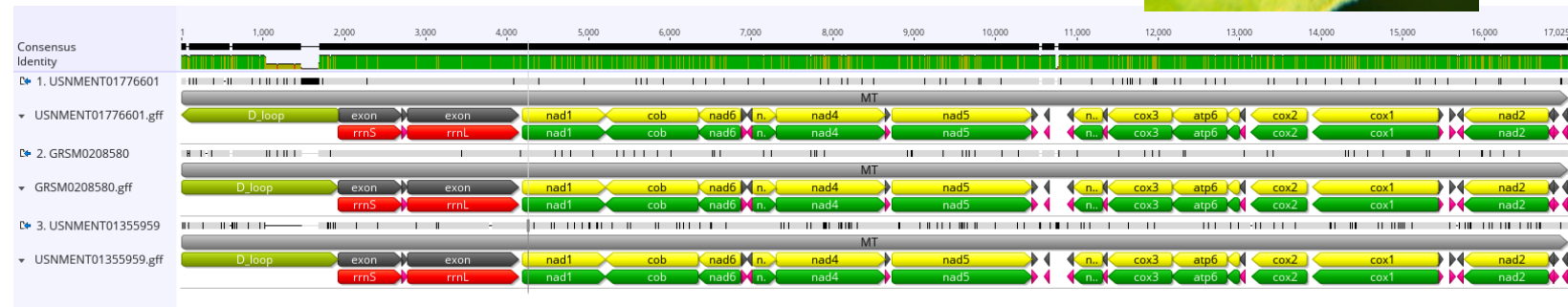
MitoPilot usage at the NMNH...so far

- 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