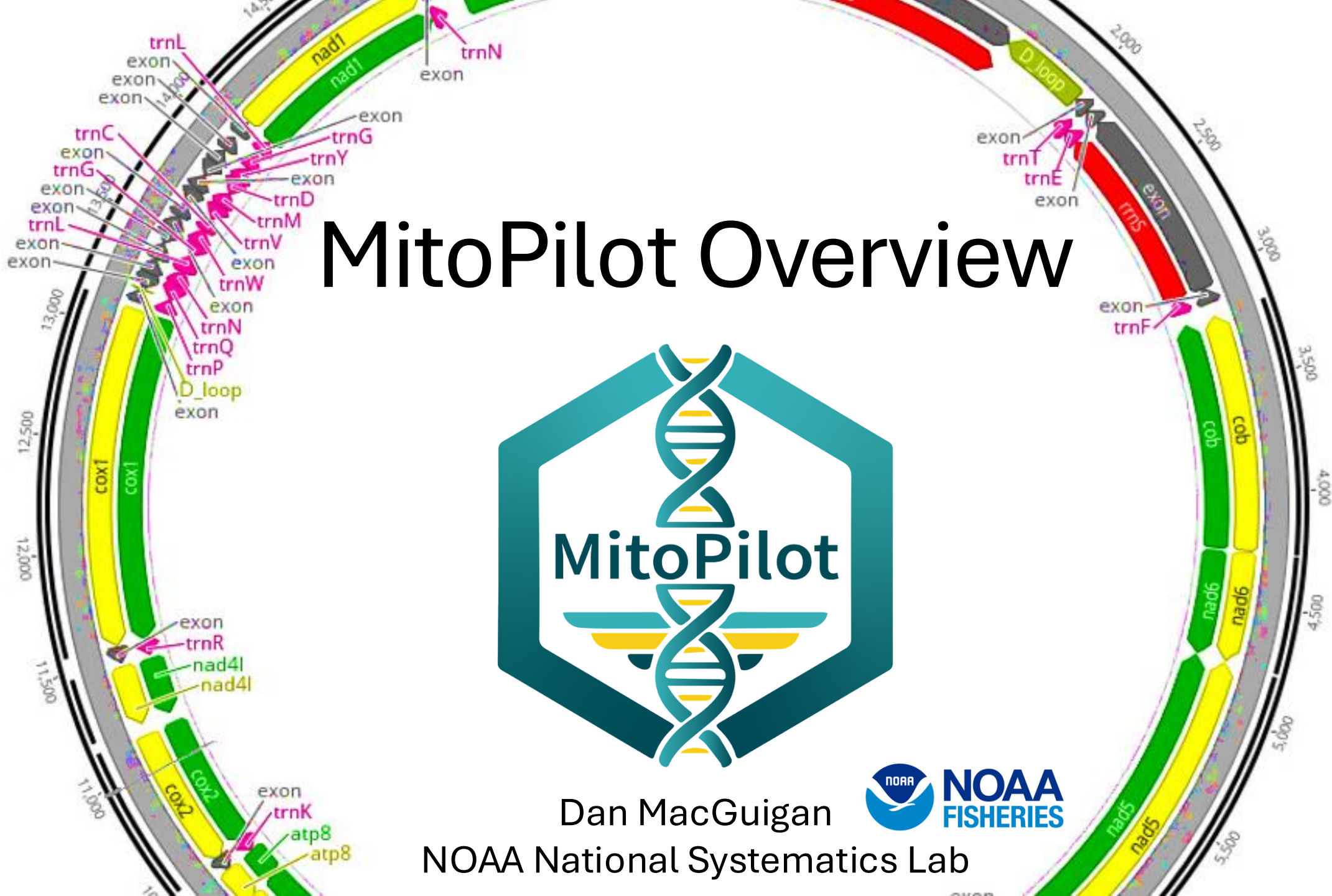


MitoPilot Overview



Dan MacGuigan

NOAA National Systematics Lab





What is MitoPilot?

- A software package for mitogenome assembly and annotation from Illumina genome skimming data
- Initial development by [Devin Leopold](#) at [Jonah Ventures](#)
- Dan MacGuigan (NOAA National Systematics Lab) is leading further development and NMNH implementation



Why do we need 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 pipeline overview



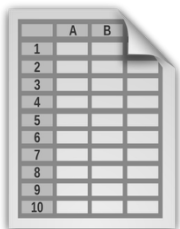
DNA sequence data



User-generated mitogenomes (optional)

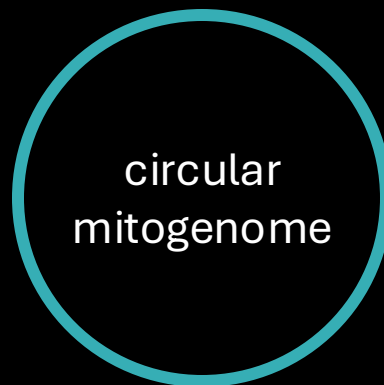


Sample metadata



ASSEMBLE

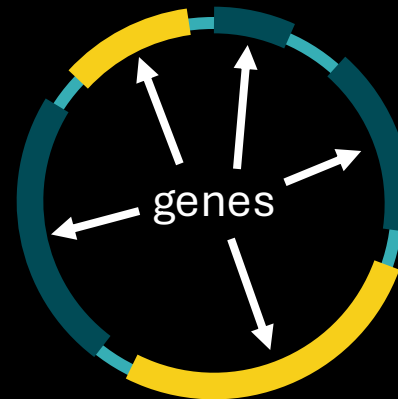
- QC sequence data
- Assemble mitogenome
- Map reads to assembly



fastp, GetOrganelle, MitoFinder, bowtie2

ANNOTATE

- Identify gene regions
- Automatic and manual curation
- Flag problematic samples/annotations



tRNAscan-SE, MITOS2

EXPORT

- Group samples
- Create sequence FASTA files
- Generate GenBank submission files

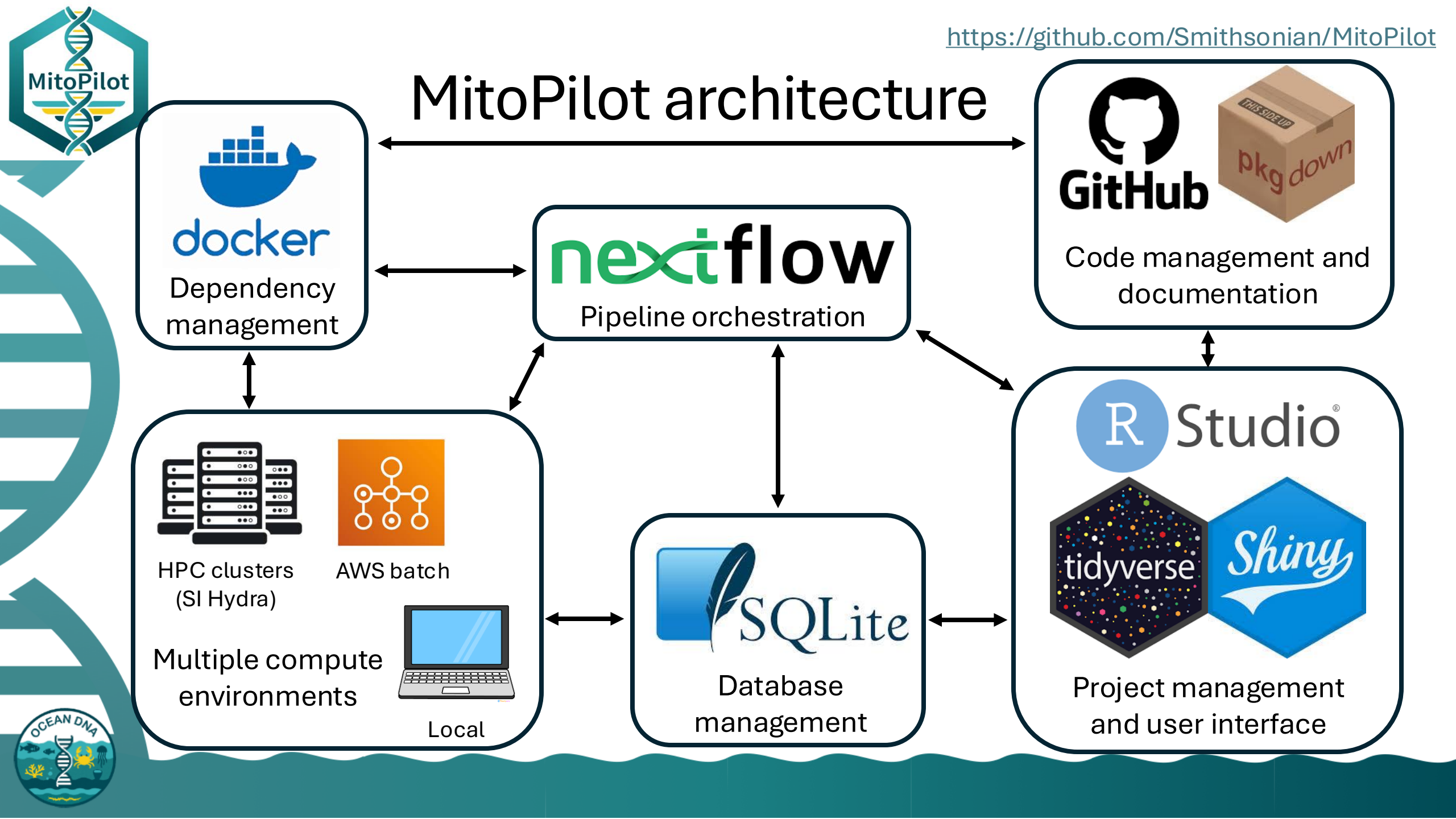
```
>Sample1
ATCCTTTACTGACT
GACCCCCTTATCG
```

```
> Sample1 NAD1
ATGACTTCTAATCT
CCTTACTCACCTA
```

```
>Feature Sample1
2729 3709 gene
gene nad1
```

Submit to GenBank







MitoPilot documentation website

Test project
tutorial

Documentation for
all R functions

Instructions for using
MitoPilot on Hydra

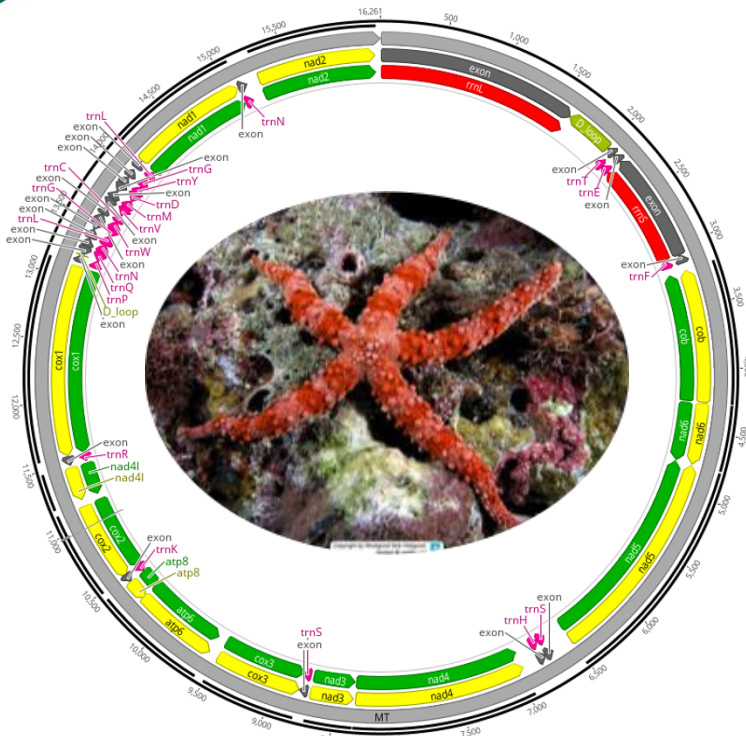
The screenshot shows the MitoPilot documentation website at smithsonian.github.io/MitoPilot/. The browser address bar is at the top. Below it, the page header includes 'MITOPILOT 1.2.8', a 'GET STARTED' button, a 'REFERENCE' button, and an 'Articles' dropdown menu. The 'Articles' menu is open, showing options: 'Frequently Asked Questions', 'NMNH Hydra Setup' (highlighted with a box), 'NOAA SEDNA Setup', 'Fish Mitogenome Curation', and 'Building Custom Databases'. A search bar is on the right. The main content area has an 'OVERVIEW' section with the text: '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:'. On the right, there are sections for 'DEVELOPERS' (listing Devin Leopold and Dan MacGuigan) and 'DEV STATUS' (showing 'lifecycle' as 'experimental' and 'R-CMD-check.yaml' as 'passing').

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



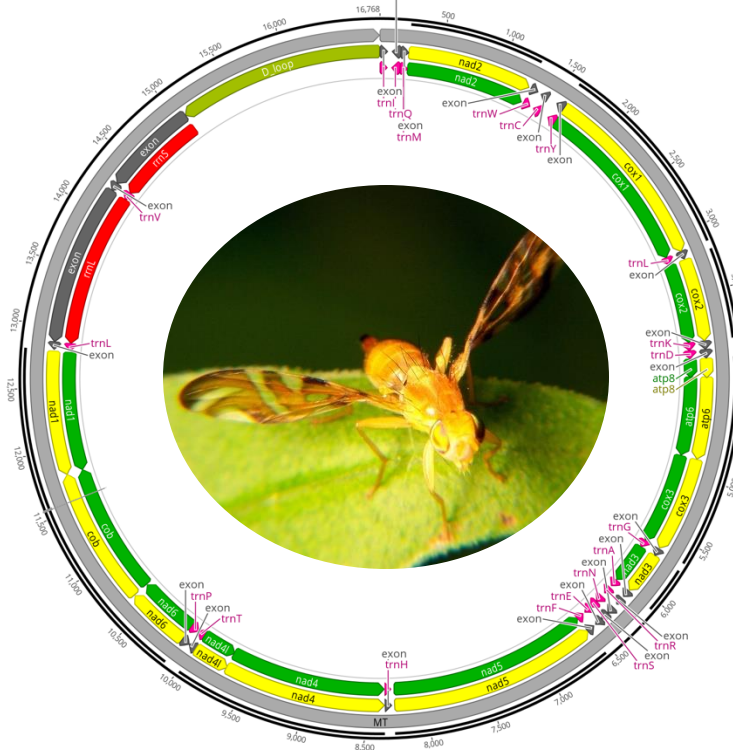


MitoPilot usage at the NMNH so far...

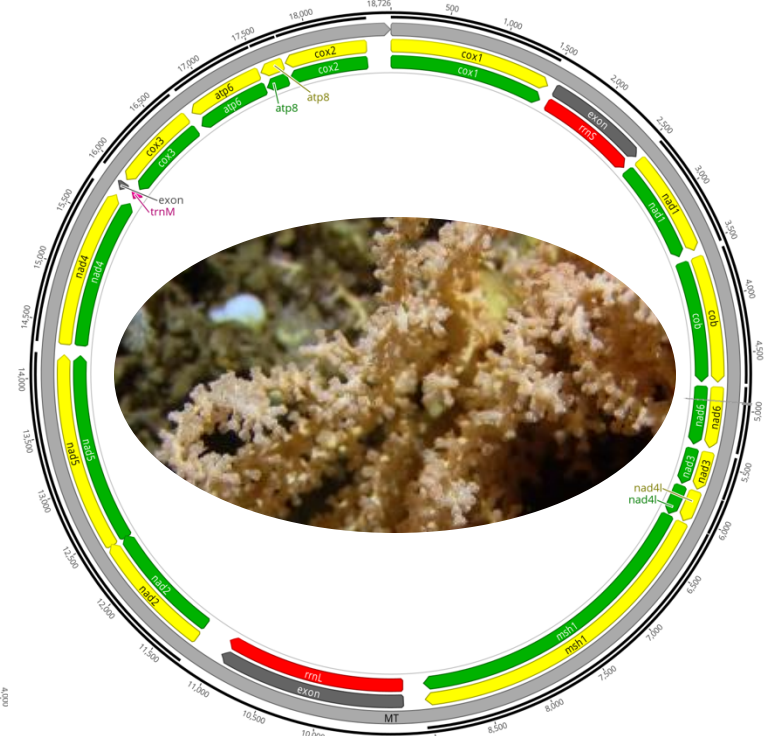


Starfish
(*Gomophia nardoa*)

Tephritid fruit flies
(*Strauzia rugosum*)



Octocorals
(*Lateothela grandiflora*)





Challenges ahead

- Some taxa lack reference mitochondrial genomes, which makes assembly difficult
- Taxa with unusual mitogenome features
 - Fragmented
 - Mitochondrial telomeres
 - Gene duplication
 - Pseudogenes
 - Bi-parental inheritance
- NCBI GenBank has an opaque, taxon-specific review processes

