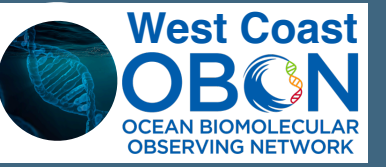


Operationalizing eDNA Metabarcoding Taxonomic Assignments for Routine Marine Biomonitoring



Zachary Gold^{1**}, Katherine Silliman^{2,3}, Emily E. Curd⁴, Nastassia Patin^{5,6}, Susanna Theroux⁶, Sean M. McAllister^{1,7}, Ramon Gallego⁸, Luna Gal⁹, Shaun Nielsen¹⁰, Erin V. Satterthwaite^{5,11}, N. Dean Pentcheff¹², Adam R. Wall¹², Regina Wetzer¹², Brice X. Semmens⁵, Rachel Storo^{2,3}, Luke R. Thompson^{2,3}

**zachary.gold@noaa.gov ¹NOAA PMEL Ocean Molecular Ecology Program, ²MSU Northern Gulf Institute, ³NOAA AOML, ⁴UV Vermont Biomedical Research Network, ⁵CalCOFI - UCSD SIO, ⁶SCCWRP, ⁷UW CICOES, ⁸Universidad Autónoma de Madrid, ⁹Landmark College, ¹⁰Independent Researcher, ¹¹CA Sea Grant, ¹²NHMLAC



INTRODUCTION

- 1) Environmental DNA (eDNA) is a powerful, cost-effective, and scalable tool for marine biomonitoring.
- 2) Improving taxonomic assignment of DNA sequences is needed for operationalizing species identification.
- 3) We outline a suite of tools for conducting accurate, repeatable, and defensible taxonomic assignments.

REFERENCE DATABASES

Curated, comprehensive reference barcode databases are key to making accurate taxonomic assignments.

EXAMPLE: CA CURRENT FISHES - MIFISH 12S MARKER

Barcoding Efforts Enhance Accuracy

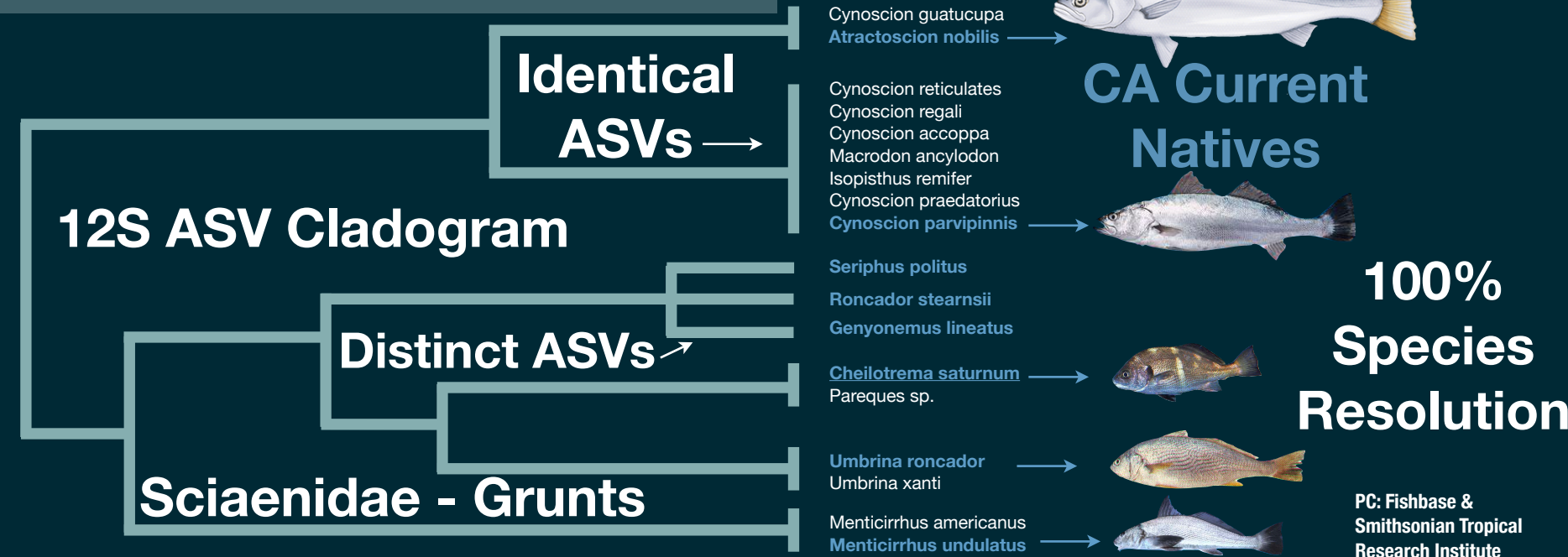
Extensive MiFish 12S barcoding effort added 252 species. Database now covers 92.7% of monitored species.

Metric	Before Barcoding	After Barcoding
Amplicon Sequence Variants (ASVs) Assigned to Species	145	156
Reads Assigned to Species	192,808	248,677
ASVs Assigned to Native Species	25	37

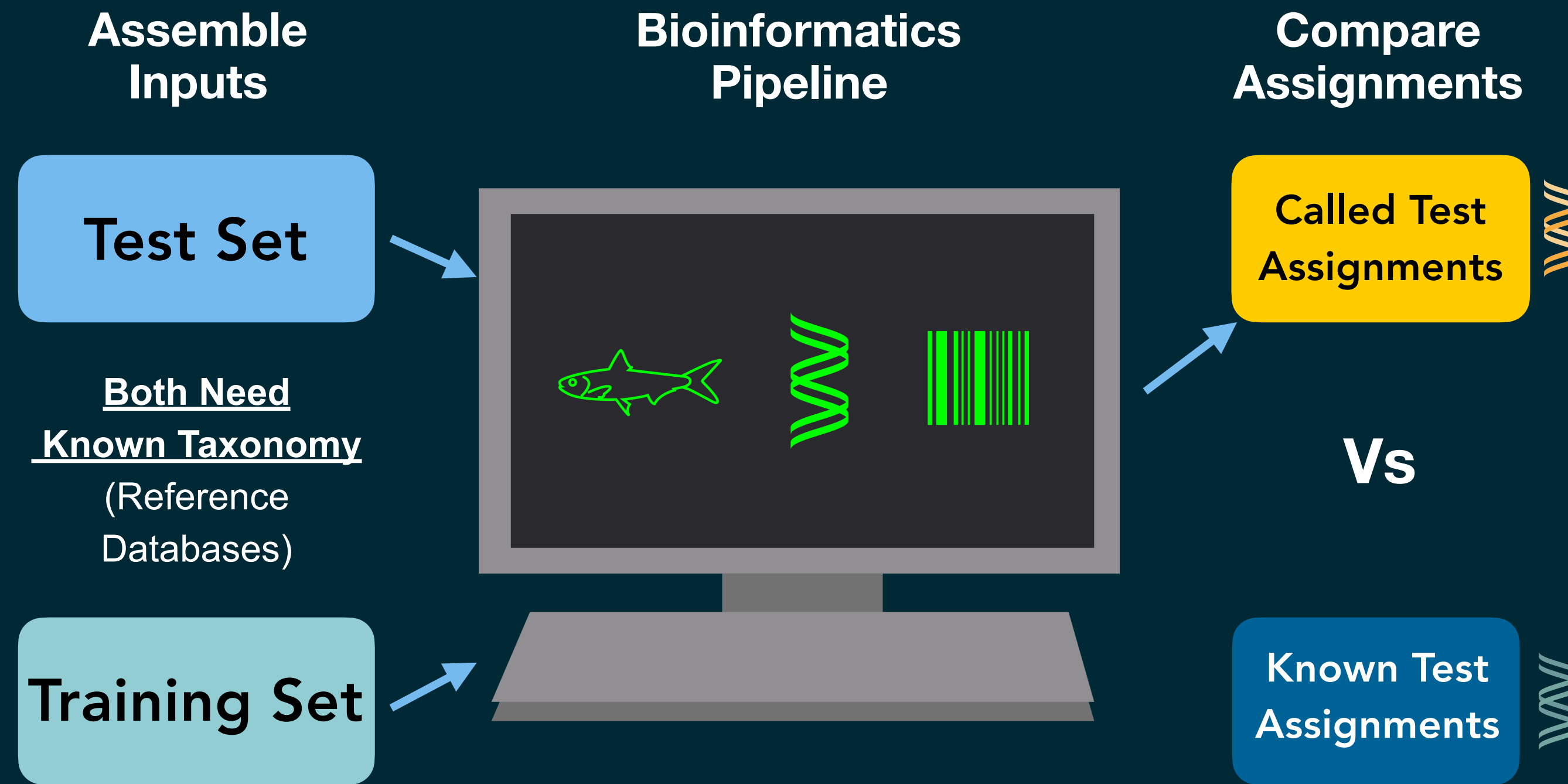
Regional Databases Outperform Global Databases

Taxonomic Cross Validation Metric	Assigned by Global	Assigned by Regional
Accuracy	86.5%	90.9%
Sensitivity	88.0%	92.1%
Specificity	98.2%	98.7%

Curation Enhances Resolution



TAXONOMIC CROSS VALIDATION



VERSATILE TOOL FOR GENERATING COMPREHENSIVE REFERENCE LIBRARIES



- 1) eDNA monitoring requires a greater diversity of specialized gene regions than are currently curated by professional staff.
- 2) We present rCRUX: an easy-to-implement tool to generate curated, comprehensive reference libraries for any bespoke locus.

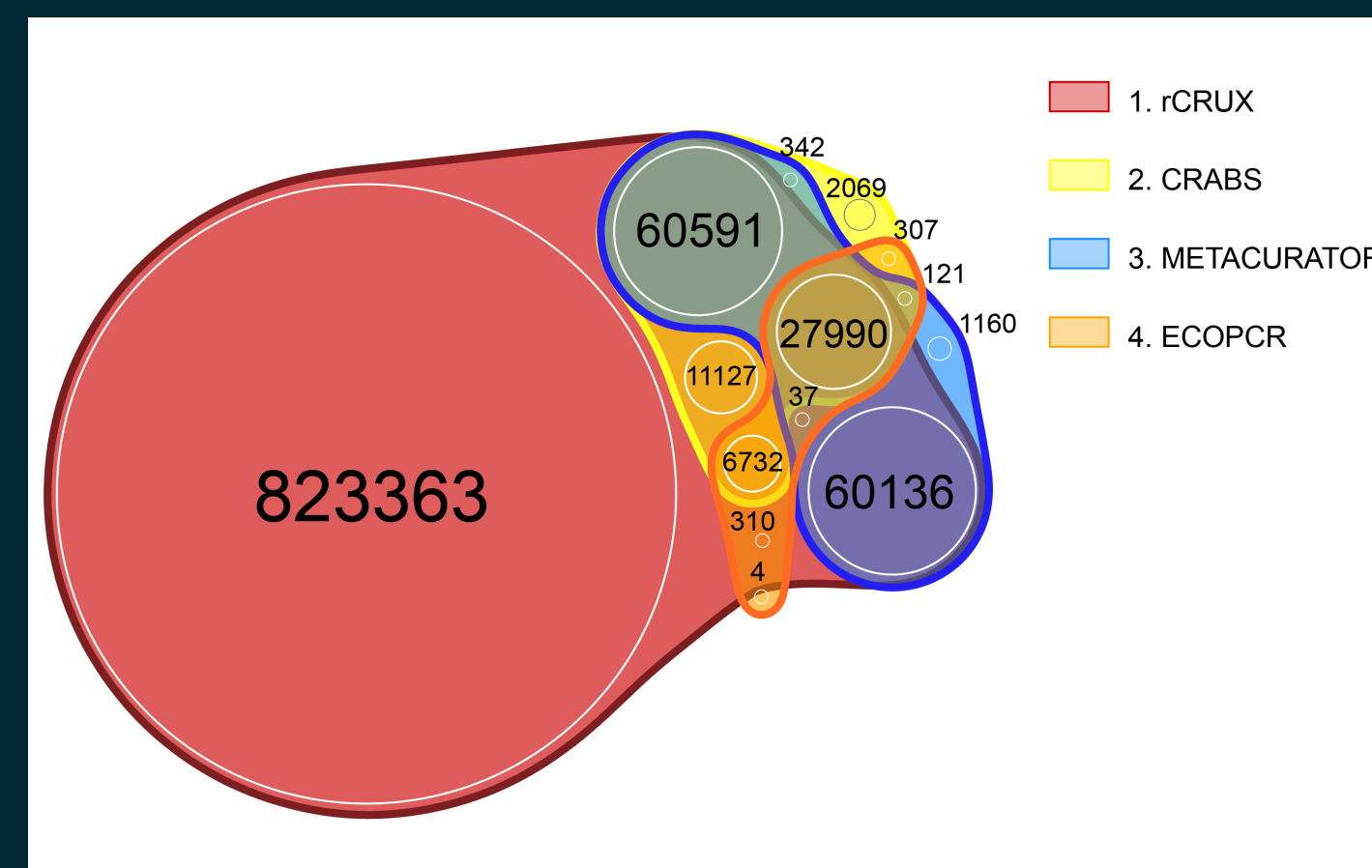
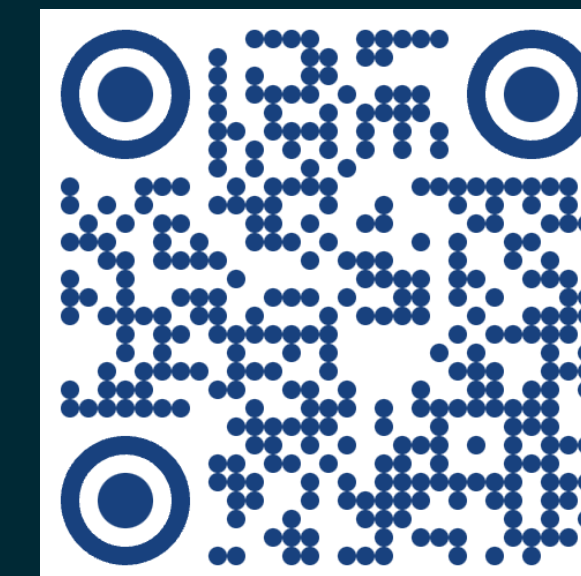


Figure 1. Leray CO1 rCRUX database comparison. rCRUX captured 99.6% ($n=990,286$) of all species observed across other publicly available databases.

PARAMETER BENCHMARKING

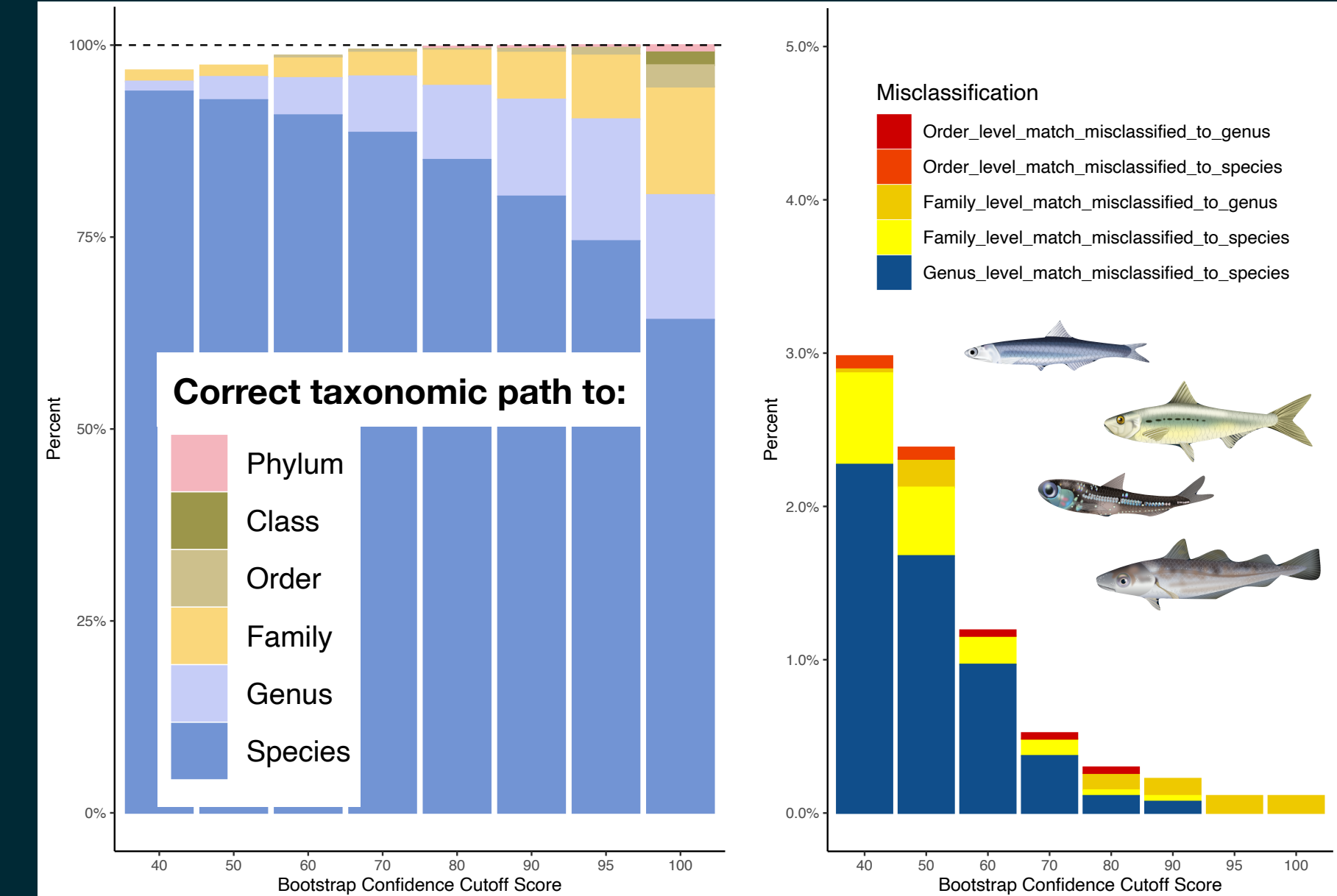


Figure 3. Parameter testing reveals confidence vs. resolution tradeoff. Relaxed confidence results in increased accuracy and misclassification rates. Optimization for invasive & protected species requires more stringency.

DEFENSIBLE STRINGENT TAXONOMIC BEST PRACTICES:

- 1) Anacapa Classifier - 95% identity & query coverage
- 2) Unambiguous BLAST top hit
- 3) Reference sequences for target & sister natives exist
- 4) Documented in region (e.g. Invasive Species Profiles List)

FUTURE DIRECTIONS

Formal comparison of metabarcoding pipelines

Tourmaline: Amplicon Sequence Processing Workflow

ANACAPA

REVAMP: Metabarcoding Pipeline

MMARINeDNA

github.com/lpipes/tronko

REFERENCES

1. Curd, E. E. et al. (2023). rCRUX: A rapid and versatile tool for generating metabarcoding reference libraries in R. *Environmental DNA*.
2. Curd, E. E. et al. (2019). Anacapa Toolkit: An environmental DNA toolkit for processing multilocus metabarcoding datasets. *MEE* 10.9: 1469-1475.
3. Gold, Z. et al. (2021). Improving metabarcoding taxonomic assignment: A case study of fishes in a large marine ecosystem. *MER* 21.7: 2546-2564.
4. Gold, Z. et al. (2022). A manager's guide to using eDNA metabarcoding in marine ecosystems. *PeerJ* 10: e14071.
5. Thompson, L. R. et al. (2022). Tourmaline: A containerized workflow for rapid & iterative amplicon sequence analysis using QIIME 2 & Snakemake. *GigaSci* 11.
6. Pipes, L. & Rasmus N. (2022). A rapid phylogeny-based method for accurate community profiling of large-scale metabarcoding datasets. *BioRxiv*: 2022.12.
7. McAllister, S. M. et al. (2023). REVAMP Rapid Exploration and Visualization through an Automated Metabarcoding Pipeline. *Oceanography* 36.2/3: 114-119.
8. Hakimzadeh, Ali et al. (2023). A pile of pipelines: An overview of the bioinformatics software for metabarcoding data analyses. *Molecular Ecology Resources*.
9. Bokulich, N. A. et al. (2018). Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. *Microbiome*.