



The Importance of Taxonomic Quality Control in Paleontological Digitization: Strategies for Increasing Fitness for Use and Trust in Aggregated Data

**NATURAL
HISTORY
MUSEUM**
LOS ANGELES COUNTY

Katy Estes Smargiassi, Austin Hendy, Erica Krimmel, Jann Vendetti,
and Lindsay Walker

Invertebrate Paleontology and Malacology

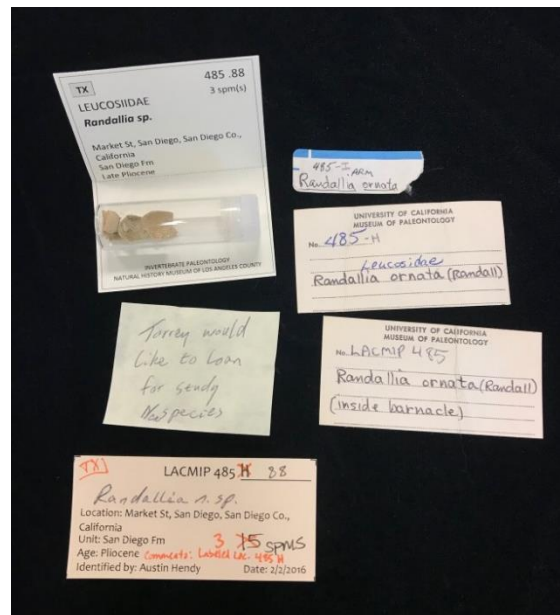
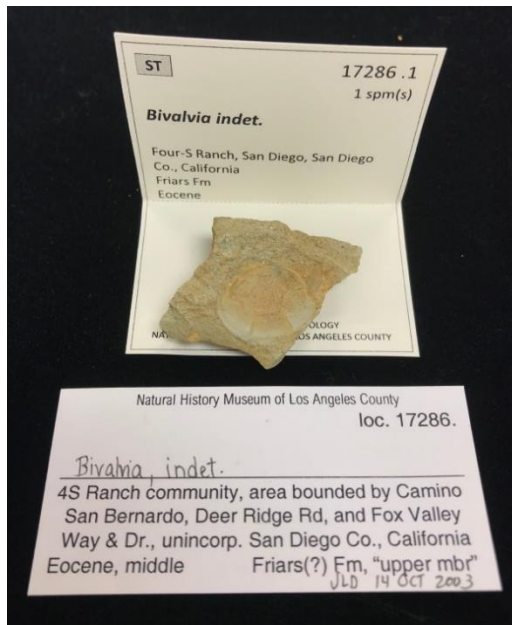
Natural History Museum of Los Angeles County

The Problem

- Responsibility for the quality of taxonomic data can be argued to belong to the
 - Data providers (who are responsible for identifications)
 - Data aggregators (who develop a unifying taxonomic backbone)
 - Downstream users (who analyze the data)

Taxonomic Quality over Quantity?

- Taxonomy of fossil specimens is fundamental to paleobiology research.
- Therefore, it is important that identifications of these specimens are as **accurate and precise as possible.**



Filling gaps in the LACMIP collection

| Class | % indetermined* |
|----------------|-----------------|
| Bivalvia | 3.7 |
| Echinoidea | 33.6 |
| Gastropoda | 4.0 |
| Malacostraca | 91.3† |
| Polyplacophora | 28.7 |
| Scaphopoda | 2.7 |

*Limited to data generated through the EPICC-TCN (Cenozoic only).

†Many reidentified by taxonomic expert, but awaiting updating of taxonomic dictionary

Filling gaps in the LACMIP collection

| Age | % indetermined |
|-------------|----------------|
| Pleistocene | 5 |
| Pliocene | 8.2 |
| Miocene | 9.2 |
| Oligocene | 8.8 |
| Eocene | 12.9 |
| Paleocene | 10.5 |

Where to Start?

- Fossil invertebrate taxonomic groups lack the species-level compendia that aid classification of many neontological plant and animal groups.
- Existing databases provide an easy solution;
 - WoRMS: World Registry of Marine Species (taxon matching tools)
 - PBDB: Paleobiology Database
- These are being used in both the digitization process **AND** as the primary taxonomic backbones for data aggregators (e.g., GBIF, iDigBio).
- How well do these tools perform as a service to a major digitization effort (Eastern Pacific Invertebrate Communities of the Cenozoic-TCN)?

Expert identification



Analysis of historic labels

Olivella biplicata Sewerby

LOCALITY NO. 35
Capistrano Beach, Orange
County, Calif. Quaternary

IDENTIFIED BY: USM

LOS ANGELES MUSEUM
EXPOSITION PARK LOS ANGELES

Museum of History, Science and Art
Los Angeles, Calif.

Ac. No. Mus. No.
Lyonsia californica (Carr)
IACMIP 2.

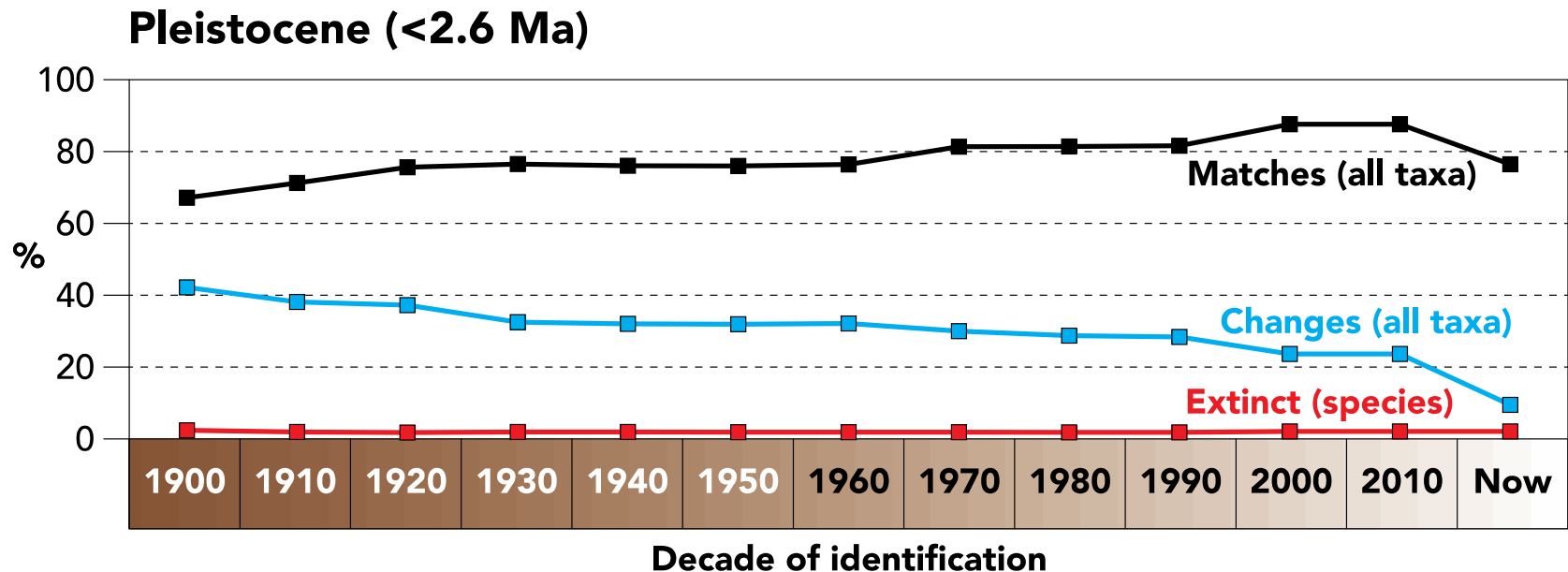
769P11
LOS ANGELES COUNTY MUSEUM OF NATURAL HISTORY
NO. 2
Nassarius delosi (Woodring, 1946)
Deadman Island, San Pedro, Los Angeles
County, California
Pleistocene
P.S.O. 7/4/76

103
Conus californicus Hds
Dead Man's Isl
Santa Catalina Isl
Pleistocene

Ophiodermella ophioderma
(Dall)
(*Moniliopsis incisa* var.)
Age ophioderma (Dall)
Form
Loc.

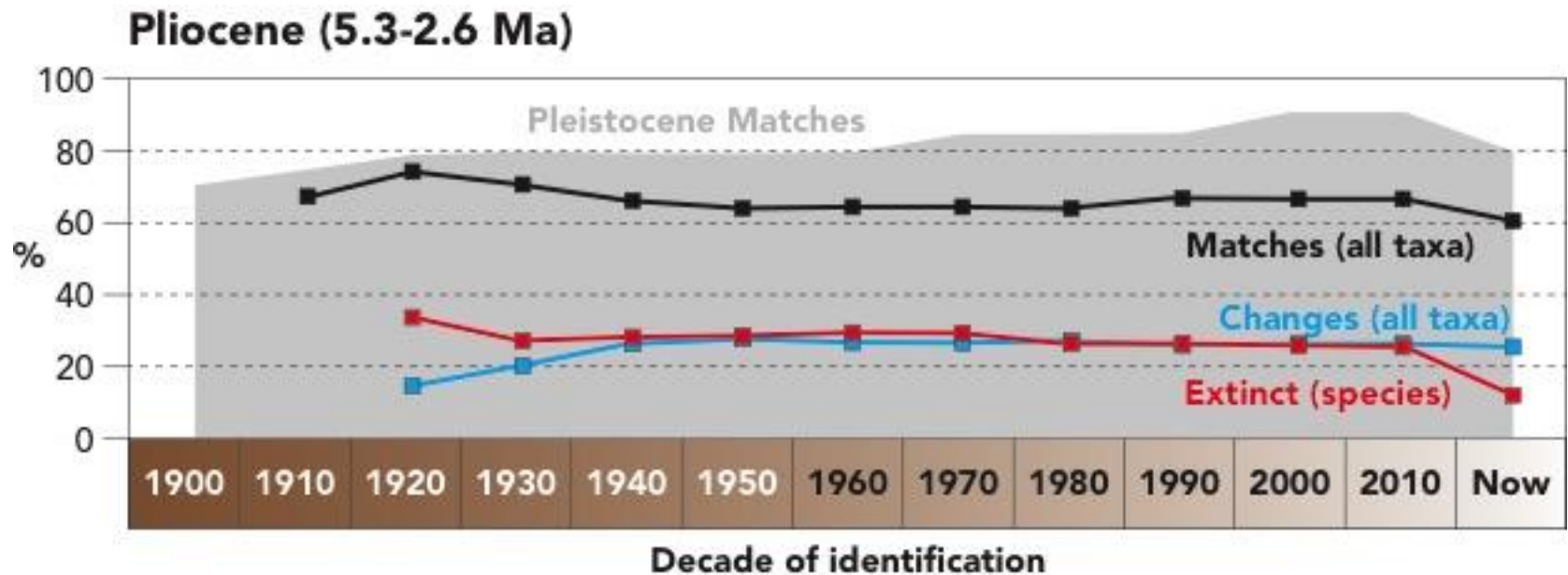
Acc. No. IACMIP
Los Angeles County Museum

Historical trends in taxon matching



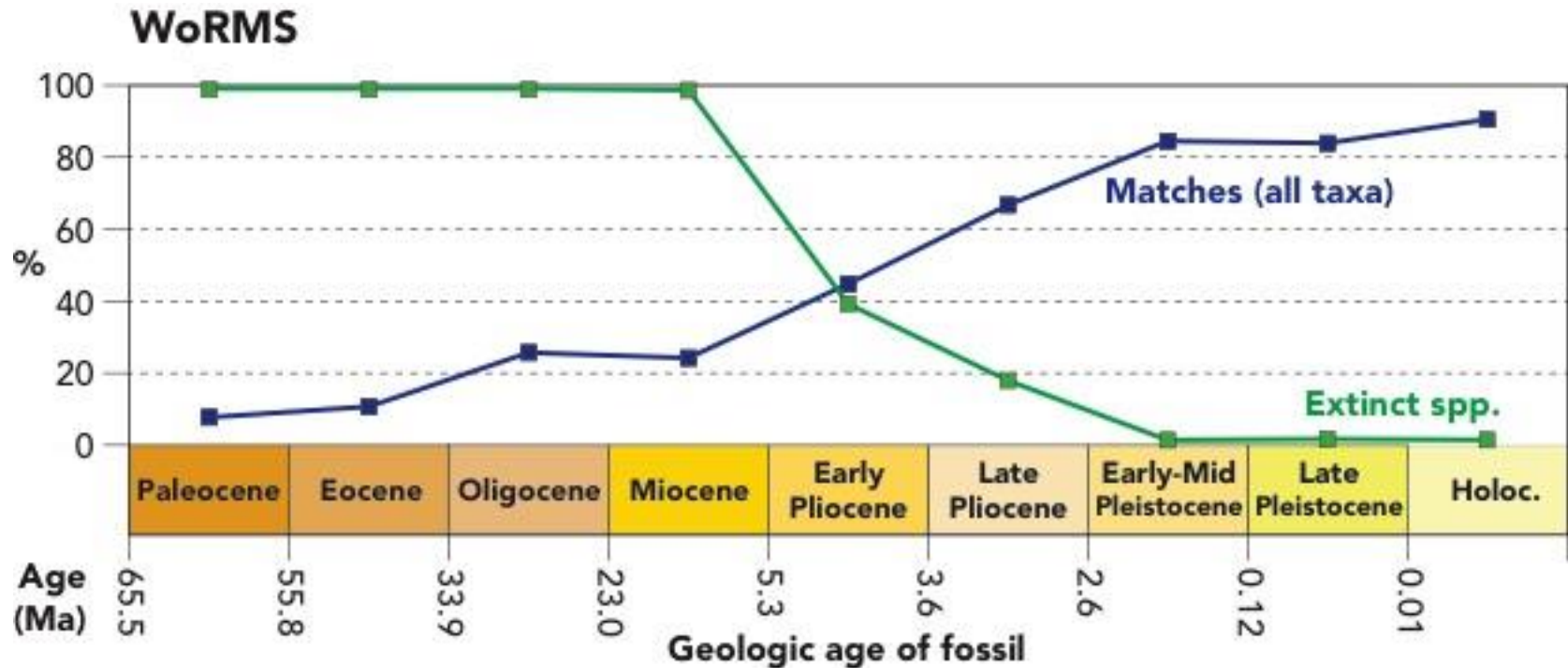
- For Pleistocene age fossils, a taxon-match with WoRMS will capture about 65-85% of specimen records
- This decreases with age of original identification
- Very few species are extinct!

Historical trends in taxon matching



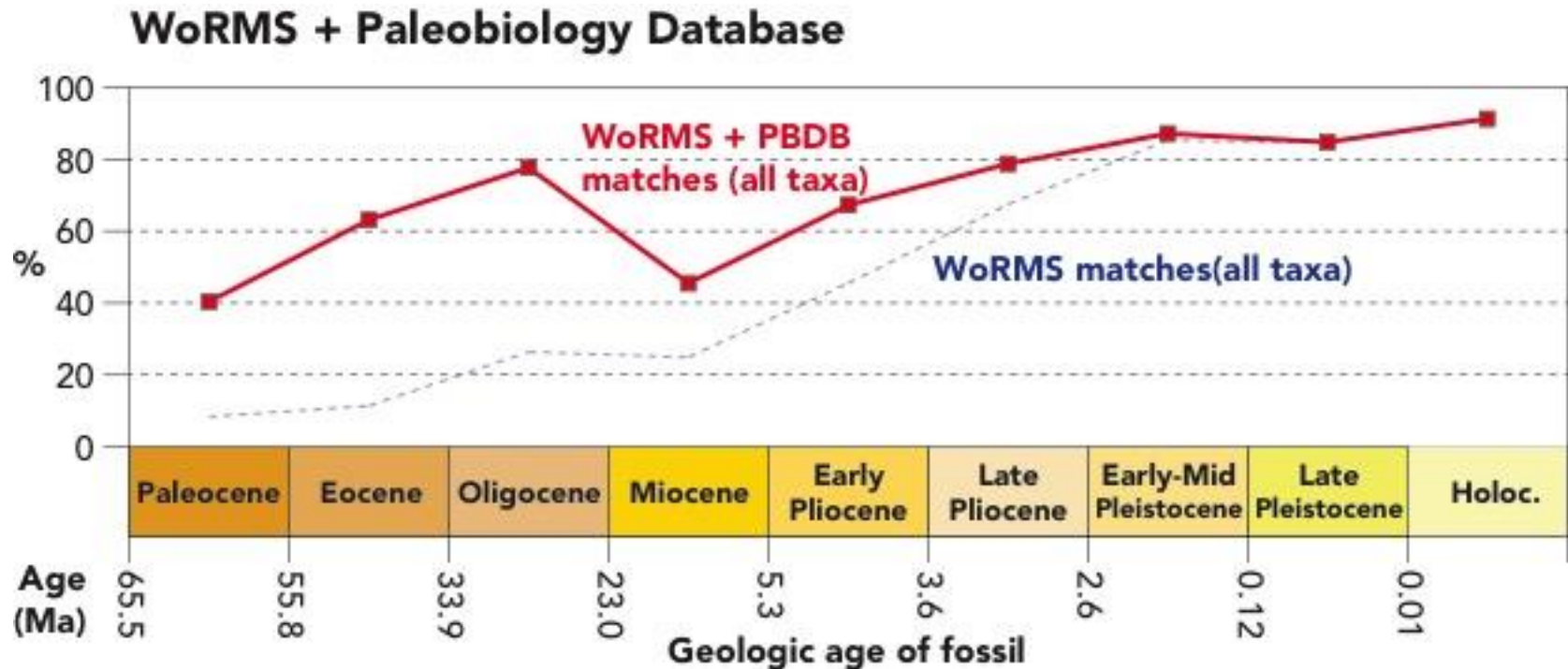
- For Pliocene-age fossils, a taxon-match with WoRMS will capture no more than 75% of v
- More species are extinct!

Using WoRMS for fossil invertebrates



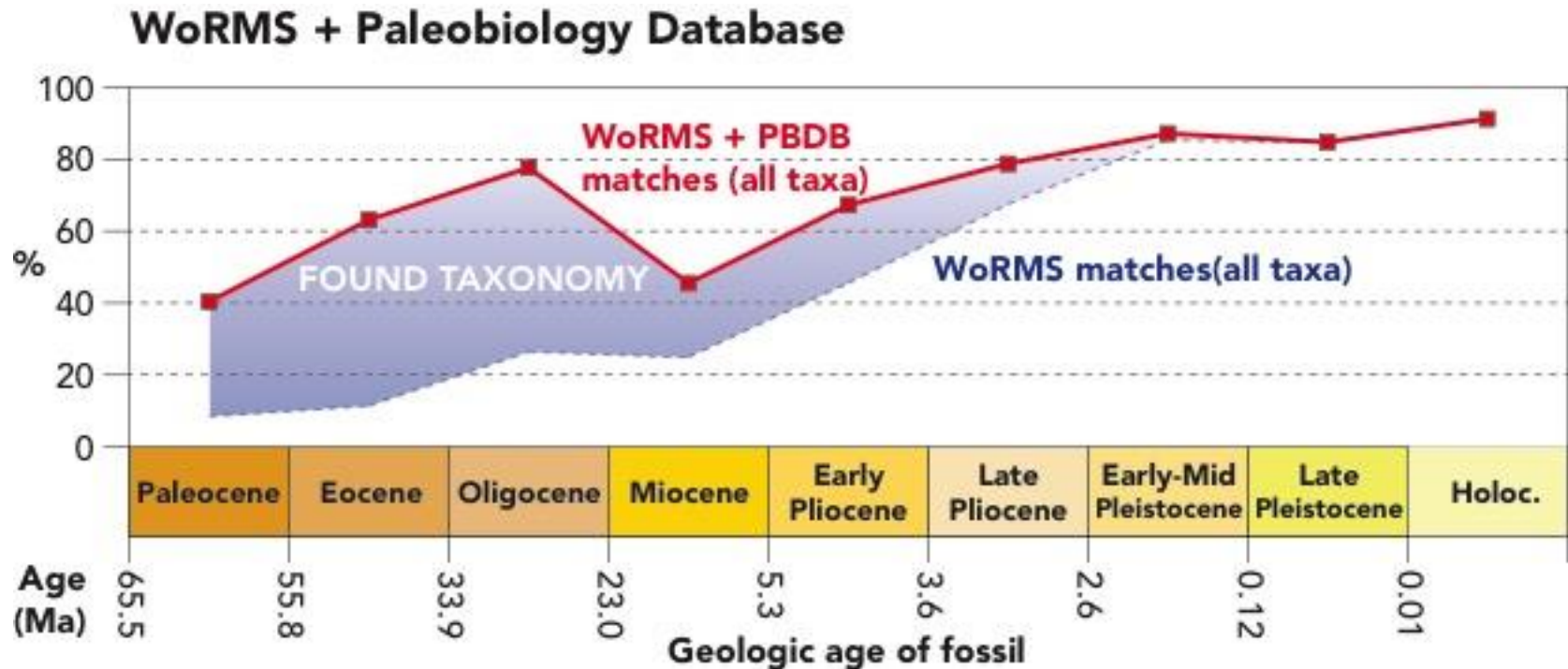
- Through geologic time the matches of specimen records increases as the number of extant species increase.
- Only really useful for Pliocene-age fossils onwards

Improving on a good taxonomic backbone



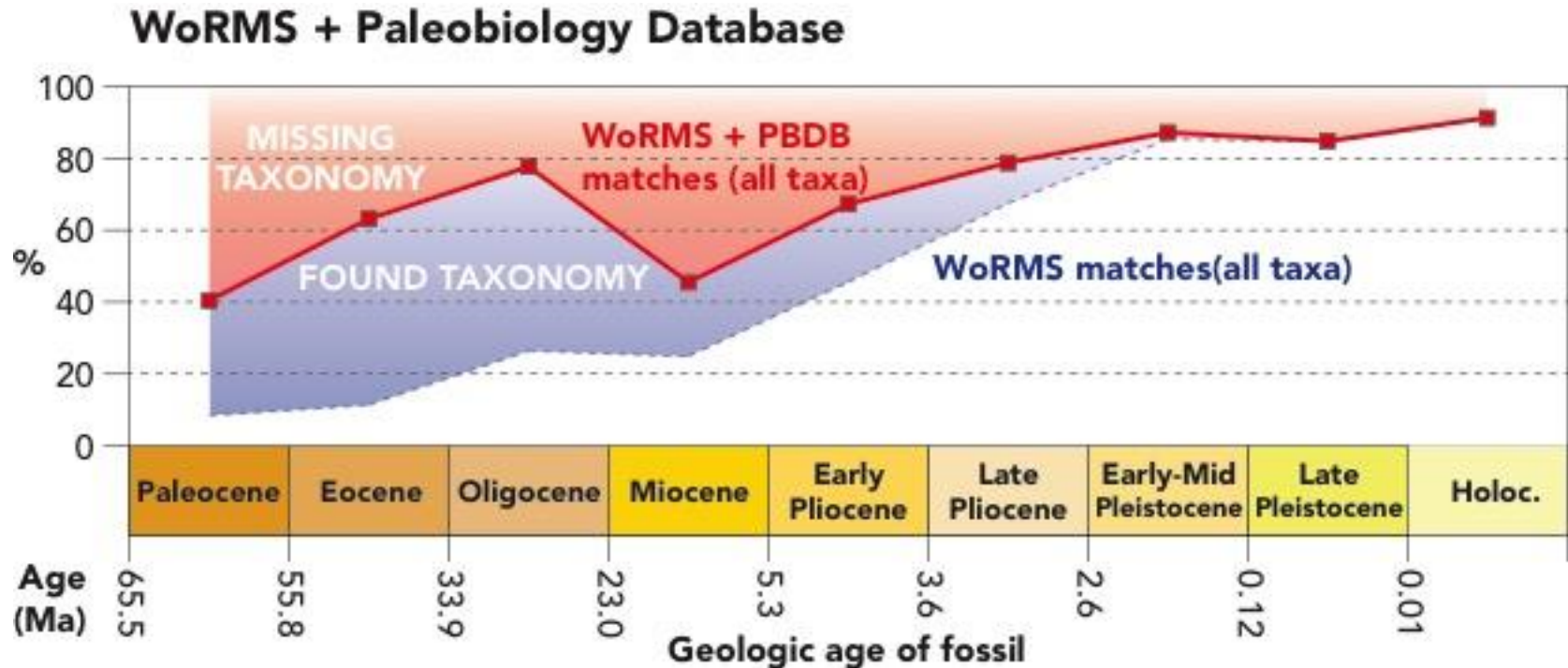
- Adding in the taxonomic opinions of the Paleobiology Database improves the % of matches with specimen records

Improving on a good taxonomic backbone



- Now about 40-80% of pre-Pliocene specimen records have matches with our two taxonomic resources

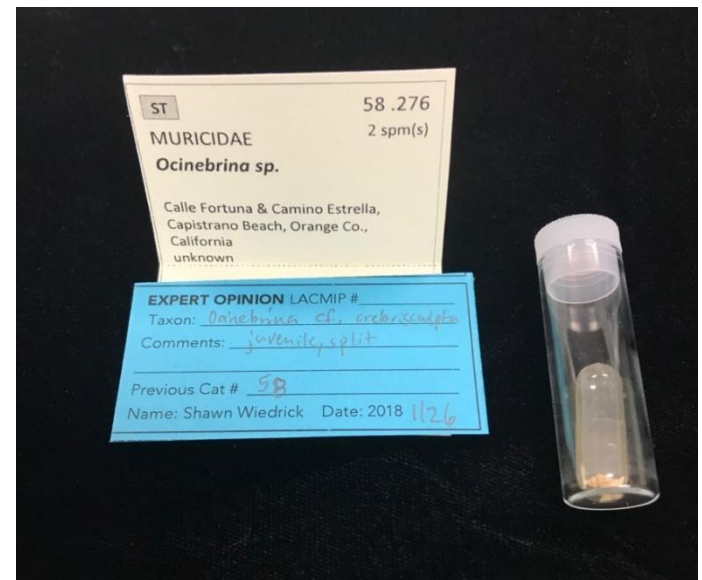
Improving on a good taxonomic backbone



- But, what about the gap?
- We (paleontology community) need to resolve this

Strategies for success

- Involvement of experts
 - identification of specimens
 - building taxonomic dictionaries
 - project design
- Develop taxonomic dictionaries with internal consistency
- Implement internally consistent taxonomic dictionaries when migrating to a new database or when starting fresh
- Work together to identify and ENHANCE taxonomic resources

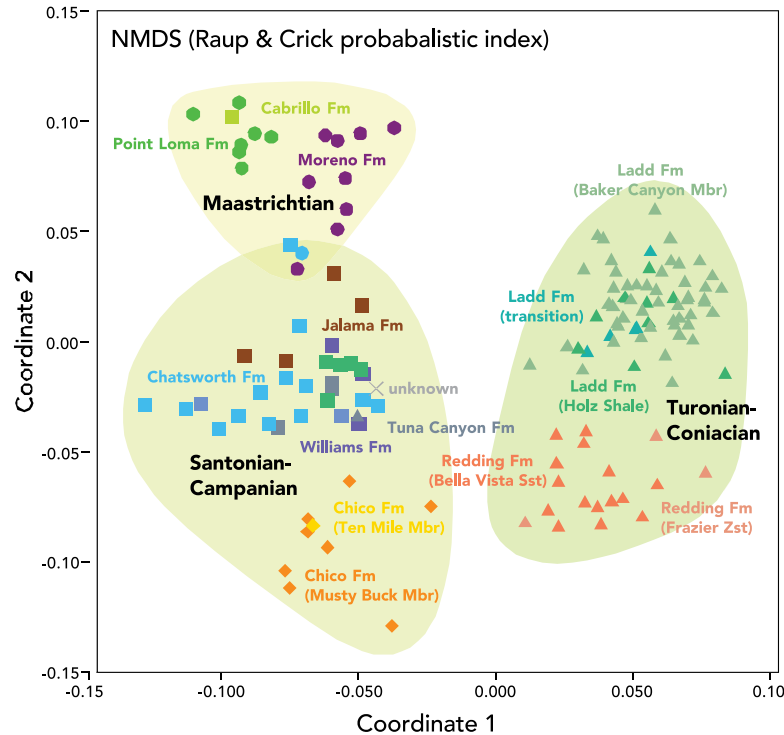


Putting the dead to work: Late Cretaceous biogeography

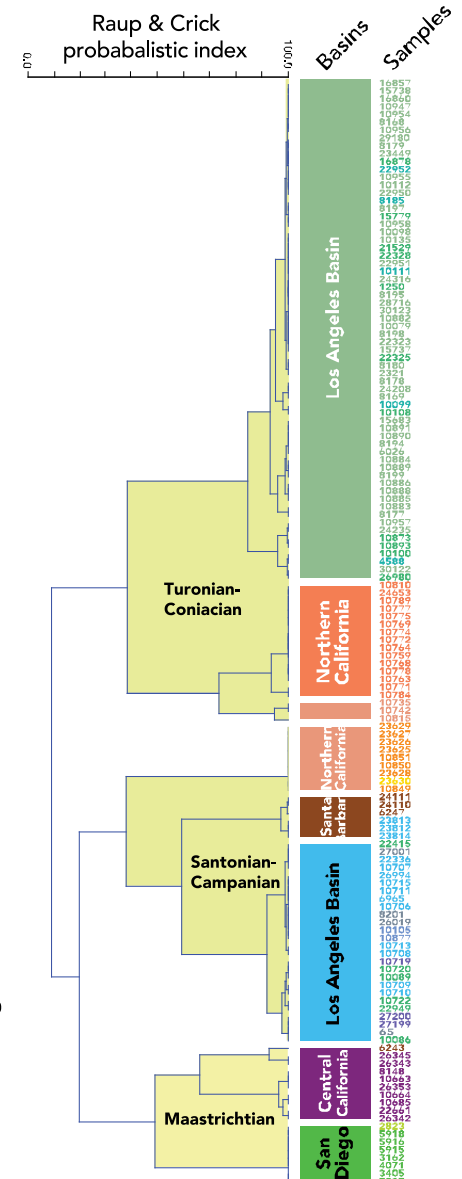
Taxonomic checklists

Mollusca
 Gastropoda
 Neogastropoda
 Volutidae
Drillula
Drillula jacksonensis (Anderson, 1958)
Konista
Konista bi conica (Anderson, 1958)
Longoconcha
Longoconcha eumeka Saul & Squires, 2008
Retipirula
Retipirula calidula Saul & Squires, 2008
Retipirula crassitesta (Gabb, 1869)
Retipirula pinguis Saul & Squires, 2008
Varens
Varens anae Saul & Popenoe, 1993
Varens formosus Saul & Popenoe, 1993
Volutoderma
Volutoderma angelica Saul & Squires, 2008
Volutoderma averillii (Gabb, 1864)
Volutoderma eldери Saul & Squires, 2008
Volutoderma magna Packard, 1922
Volutoderma perissa Saul & Squires, 2008
Volutoderma blakei Saul & Squires, 2008
Volutoderma californica Dall, 1903
Volutoderma gabbi White, 1889
Volutoderma jalama Saul & Squires, 2008
Volutoderma querna Saul & Squires, 2008
Volutoderma santana Packard, 1922
Volutoderma suciana Dall, 1907
Volutoderma ynezae Saul & Squires, 2008
Volutoderma? antherena Saul & Squires, 2008

Gradient analysis



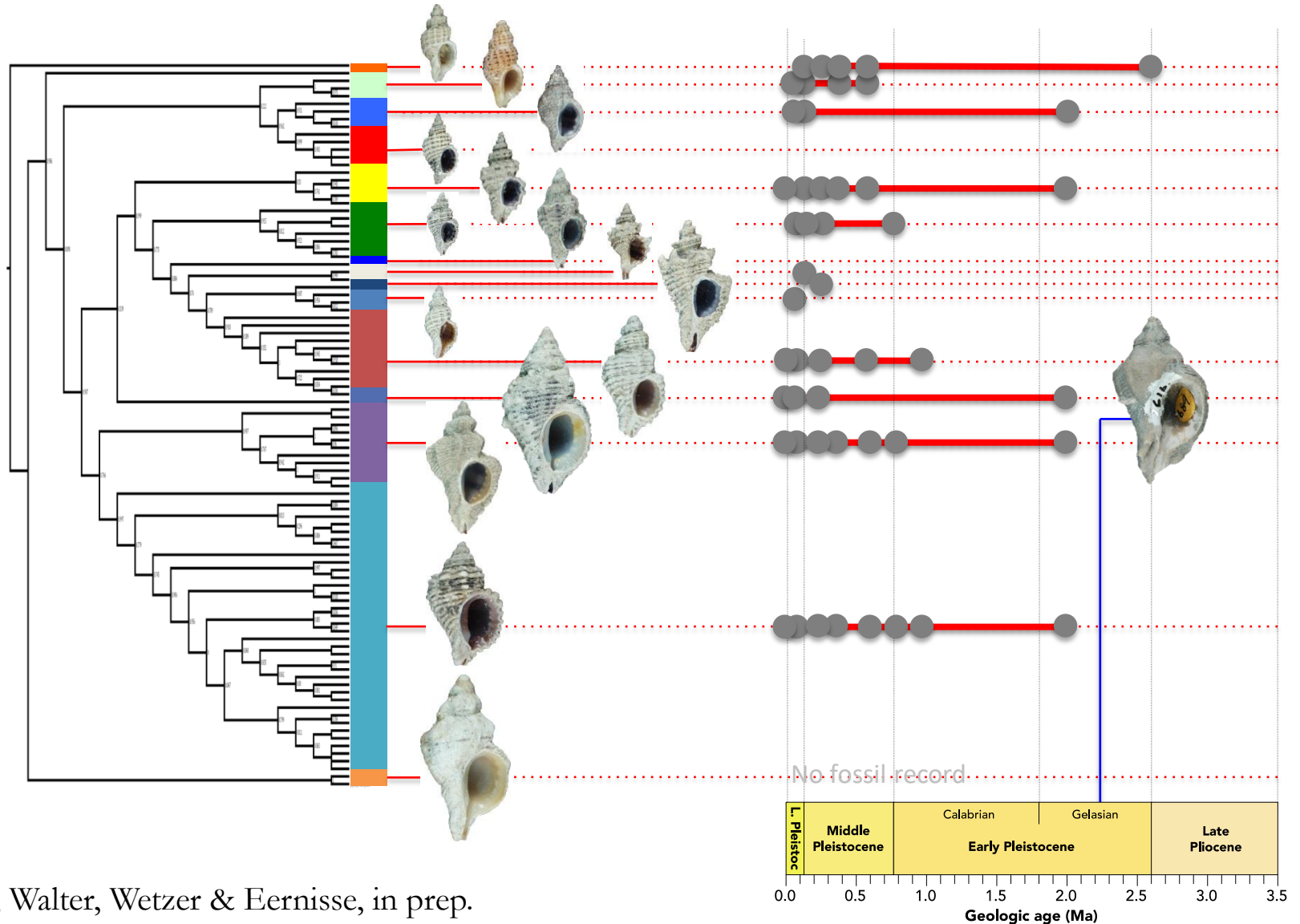
Cluster analysis



Walker et al. (this meeting) – Revitalizing the Cretaceous Seas of California (CSBR)

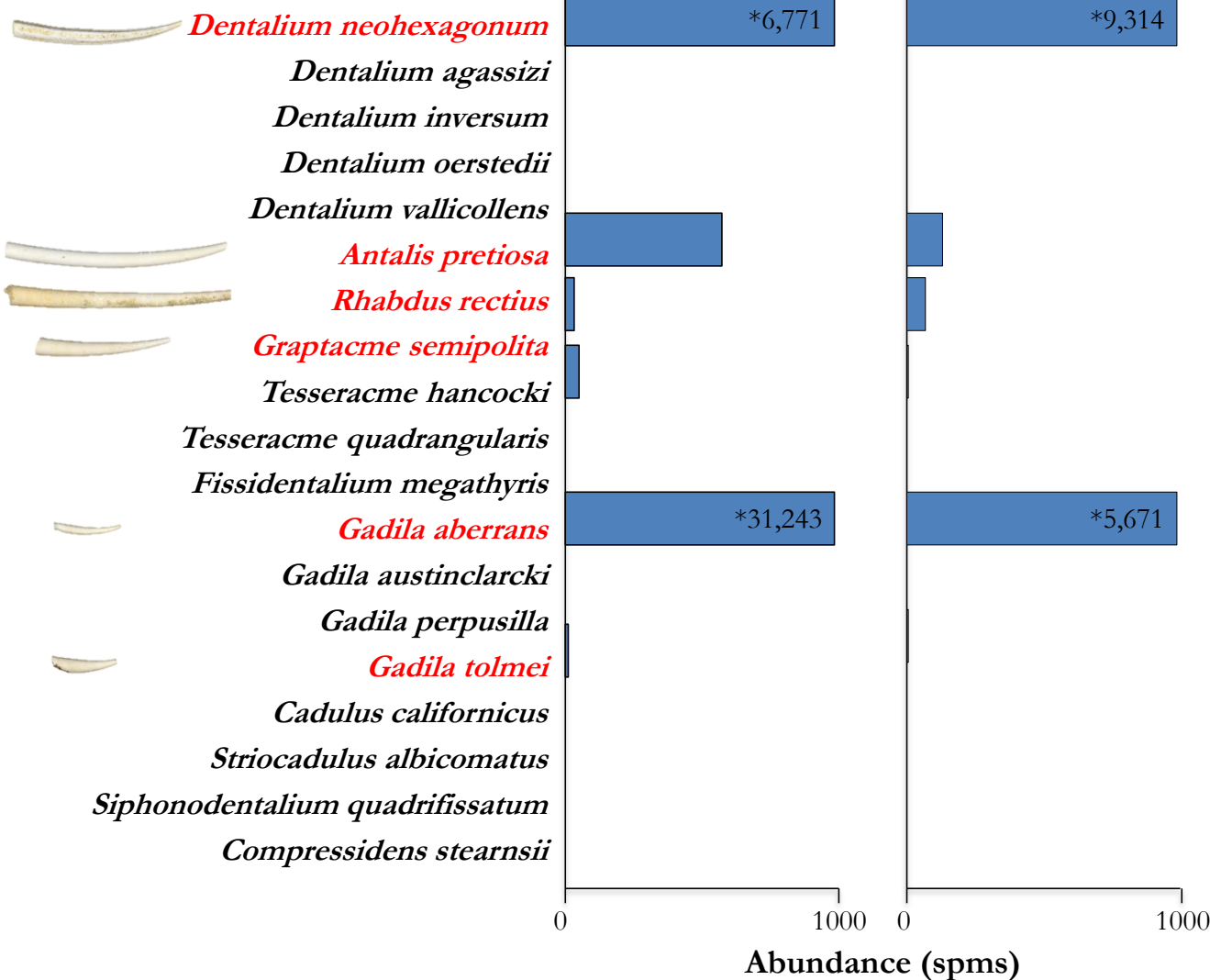
Putting the dead to work:

Plio-Pleistocene *Paciocinebrina* among the collections

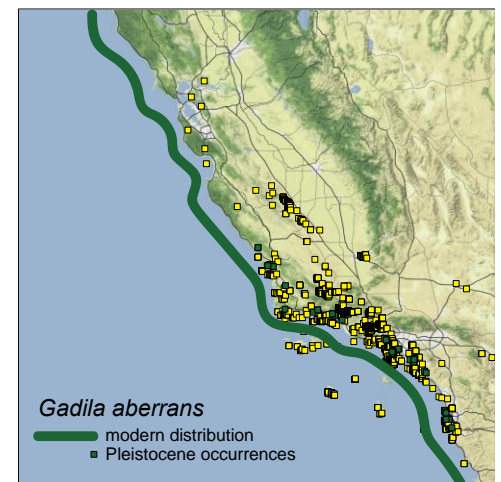
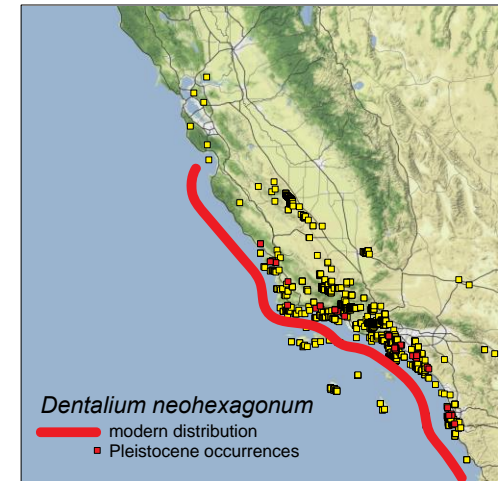


Putting the dead to work: Plio-Pleistocene scaphopods

Modern biodiversity



Species distribution (past & present)



THANKS!

- Co-authors Austin Hendy, Erica Krimmel, Lindsay Walker, and Jann Vendetti
- Shawn Weidrick, Scott Rugh, Torey Nyborg, and Chuck Powell, for their expert identifications and research contributions
- Thanks to the many LACMIP students who cataloged thousands of specimens to make this project possible!



iDigBio
Integrated Digitized Biocollections