

NORDIC BARCODING SYMPOSIUM 2007

Molecular Operational Taxonomic Units:
Treasure Trove or Trivial Pursuit?



"... endless forms
most beautiful
and most wonderful
have been,
and are being,
evolved"

(Darwin 1859)

Molecular Operational Taxonomic Units: Treasure Trove or Trivial Pursuit?

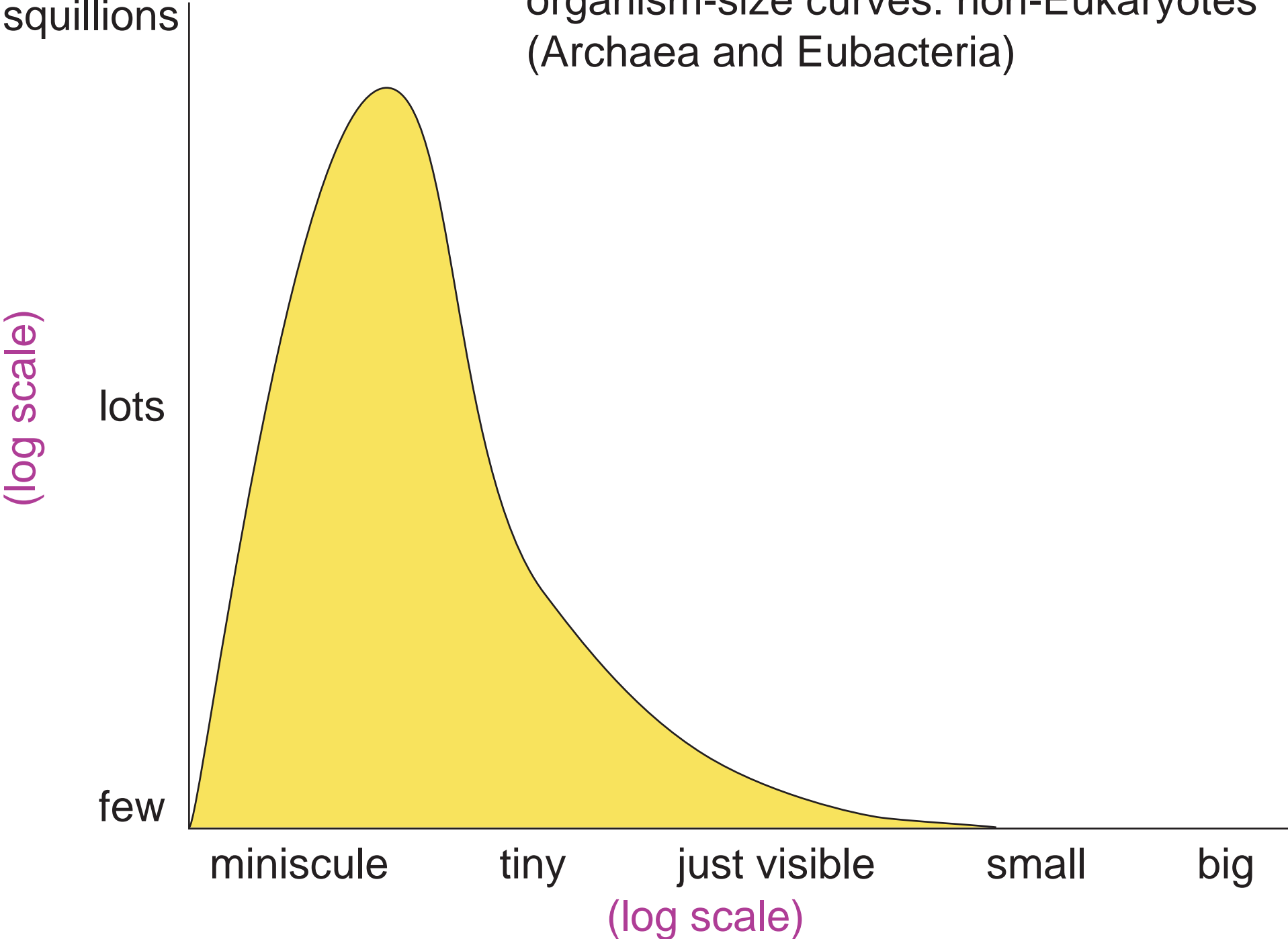
Jenna Mann, Robin Floyd

Ben Elsworth, Tom Chapman, Felicity Royds,
Andrew Langland

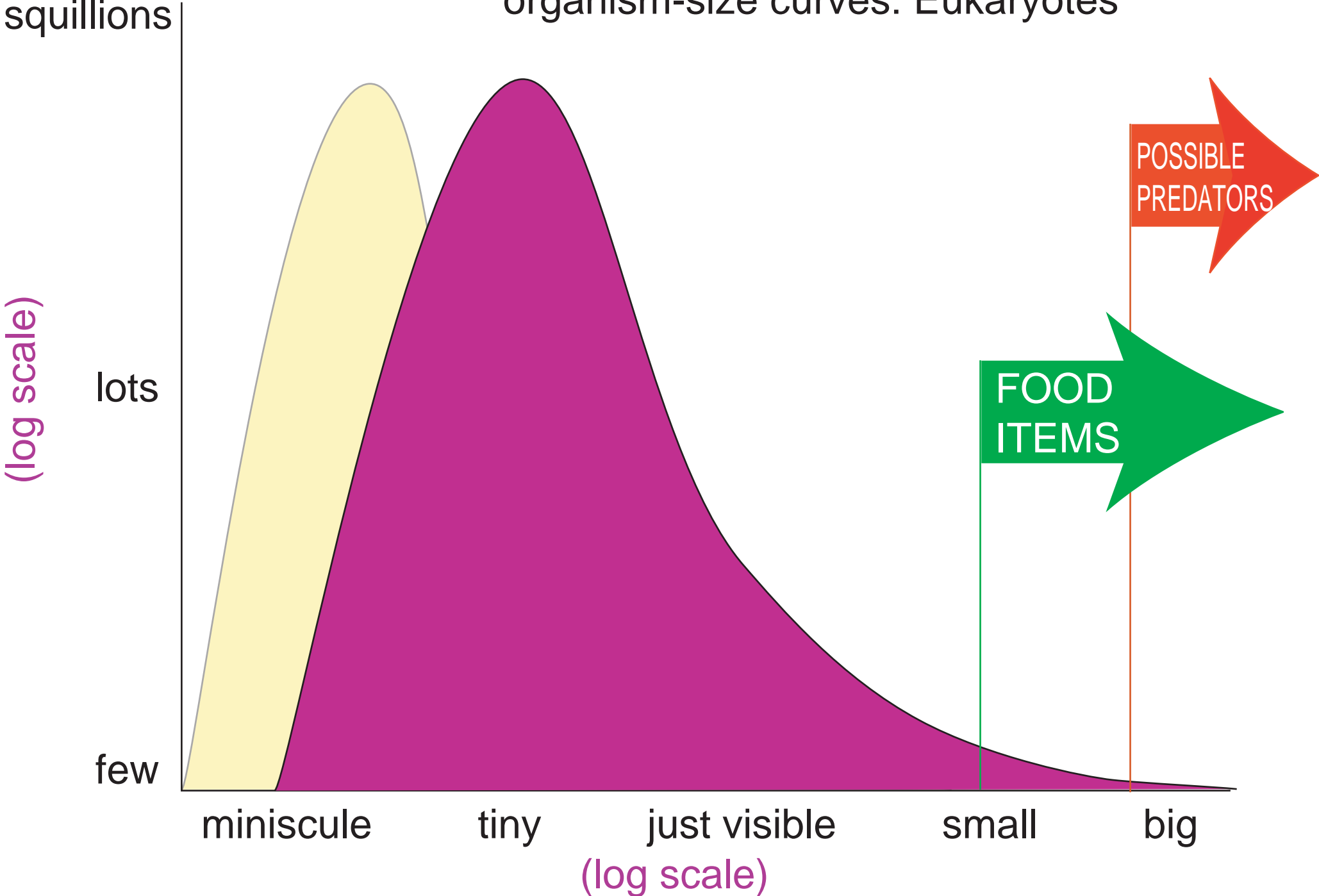
Jennifer Daub, Fran Thomas, Paul DeLey

Outline

organism-size curves: non-Eukaryotes
(Archaea and Eubacteria)



organism-size curves: Eukaryotes



- 1: Barcoding assumptions
- 2: When the assumptions are not met
- 3: Turning barcode data into taxon lists
[Trees & Clustering]
- 4: Neglected meiofauna
- 5: Charismatic megafauna

1: Barcoding assumptions

- we can PCR and sequence our target from ALL specimens

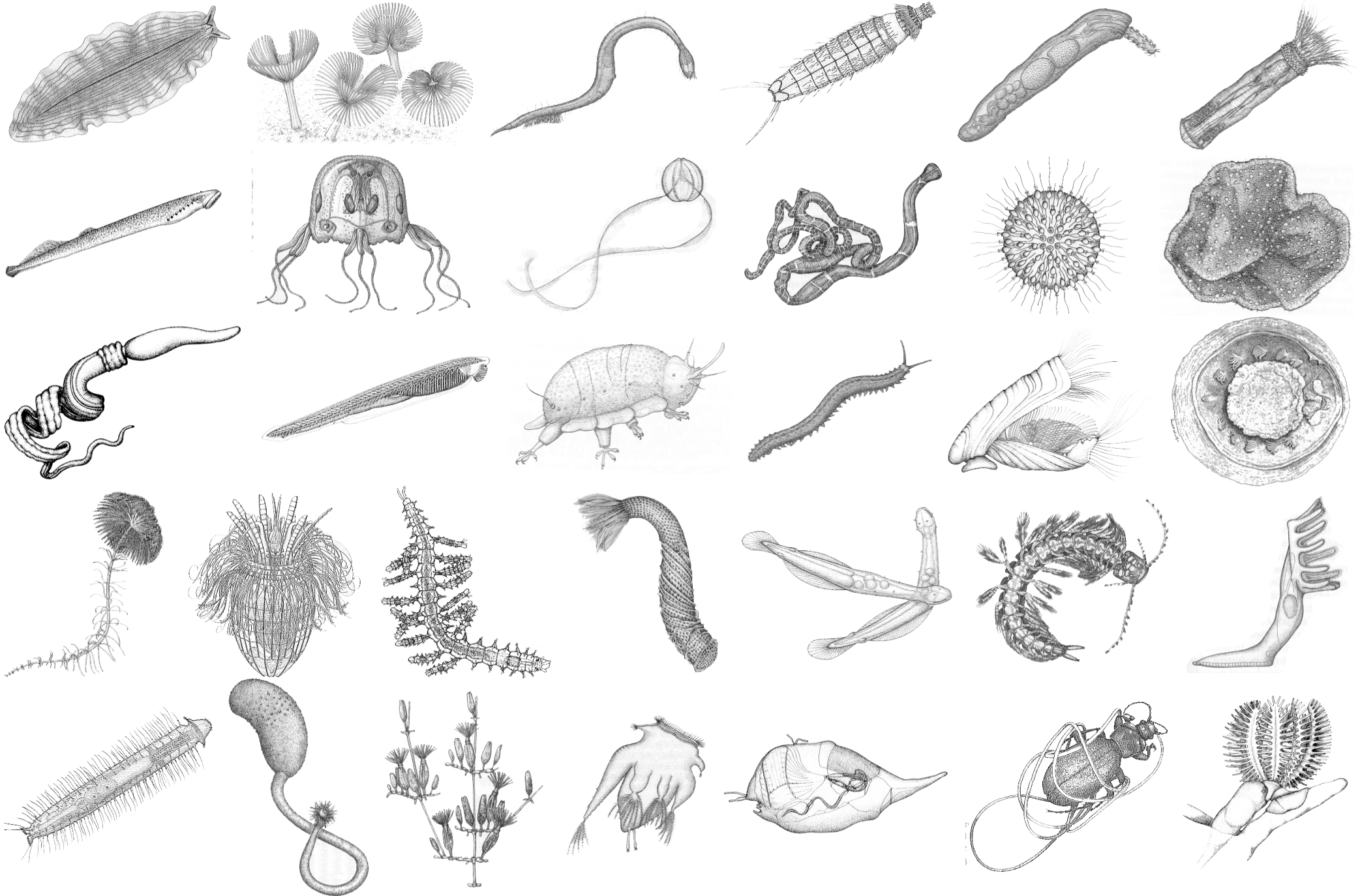
1: Barcoding assumptions

- we can PCR and sequence our target from ALL specimens

but many taxa are proving problematic for CO1 (eg Nematoda, protists)

multiple barcode targets (MLST) may be required

35 (or so) metazoan phyla



1: Barcoding assumptions

- we can PCR and sequence our target from ALL specimens
- we can generate a database of holotype/paratype sequences for all taxa

1: Barcoding assumptions

- we can PCR and sequence our target from ALL specimens
- we can generate a database of holotype/paratype sequences for all taxa

this may be problematic...

... assumes we KNOW what a species is.

1: Barcoding assumptions

- we can PCR and sequence our target from ALL specimens
- we can generate a database of holotype/paratype sequences for all taxa
- **divergence in the barcode target sequence faithfully reports on ‘species’ boundaries**

1: Barcoding assumptions

- we can PCR and sequence our target from ALL specimens
- we can generate a database of holotype/paratype sequences for all taxa
- **divergence in the barcode target sequence faithfully reports on ‘species’ boundaries**

we know this assumption may be violated

we know this assumption may be violated

- e.g. for plants, mitochondrial evolutionary rates are ‘too slow’ [another gene is needed]
- e.g. specimens may share mitochondrial haplotypes due to incomplete lineage sorting since speciation
- e.g. specimens may share mitochondrial haplotypes due to introgression driven by parasitic or mutualist cytoplasmic symbionts

1: Barcoding assumptions

- we can PCR and sequence our target from ALL specimens
 - we can generate a database of holotype/paratype sequences for all taxa
- divergence in the barcode target sequence faithfully reports on 'species' boundaries

2: When the assumptions are not met

most organisms are small

small organisms may lack light-microscope-visible morphological characters

the 'taxonomic deficit' is most extreme for small taxa

e.g **Nematoda:**

23000 species described \Leftrightarrow 1-10 million predicted

3: Turning barcode data into taxon lists [Trees & Clustering]

Tree Based Definition of Taxa

using...

tree shape

branching pattern/topology

distance measures

relative branch lengths

statistical support for clades

(bootstrap, posterior probability, decay indices)

presence of 'holotype' sequences within clade

Tree Based Definition of Taxa

sensitive to alignment

sensitive to error in sequences

can be subjective

(one person's 'well defined clade' is another's 'nested subpopulation')

has to be repeated for each new sequence added

what if 'holotype' is not available?

Tree Based Definition of Taxa

(Tim Barraclough)

looking for the transition between the **phylogenetic**
(BETWEEN-taxon processes)

and the **coalescent**

(WITHIN-taxon processes)

assessed by likelihood

has to be repeated for each new sequence added

Clustering Sequences

single-linkage clustering

MOTU_define.pl

[can be updated incrementally]

[is amenable to multiple resampling]

Clustering Sequences

multiple linkage clustering

Tribe-MCL

[has to be repeated for each new sequence added]

[is amenable to multiple resampling]

What if 'holotype' is not available?

global metazoan species list
~10 million

(of which ~1.5 million have been described, we hope)

The Taxonomic Delimitation Uncertainty Principle

‘the more closely a species is delimited
the less able we are to assign any
individual specimen to that species’

Adams, B. J. 1998. Species concepts and the evolutionary paradigm in modern nematology. *Journal of Nematology* 30:1–21.

MOTU

Molecular Operational Taxonomic Unit

assumption-explicit

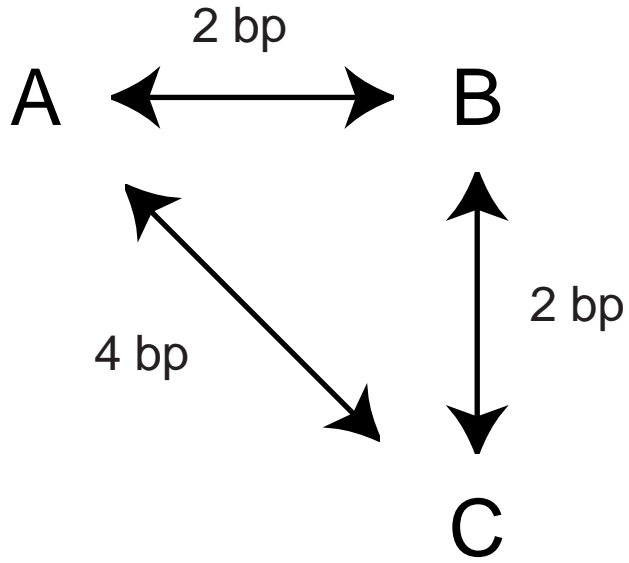
objective

universal

MOTU

- can be derived using any **explicit** criteria
 - a MOTU is defined by its members
(and not the other way round)
- a MOTU may correspond to a taxonomist's idea of a 'species', but could just as well be a 'population' or 'genus', etc.

Sequence differences

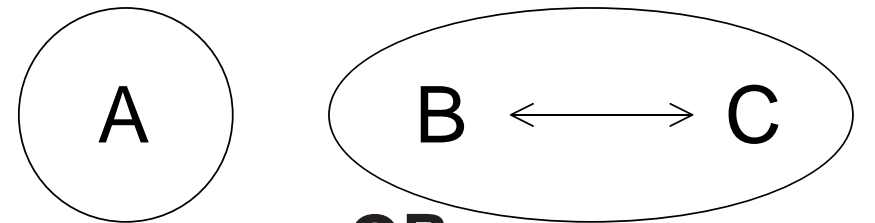
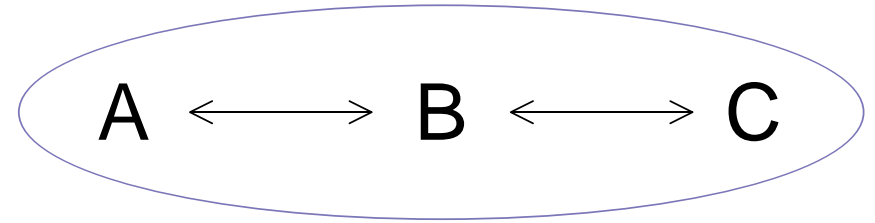


Addition order

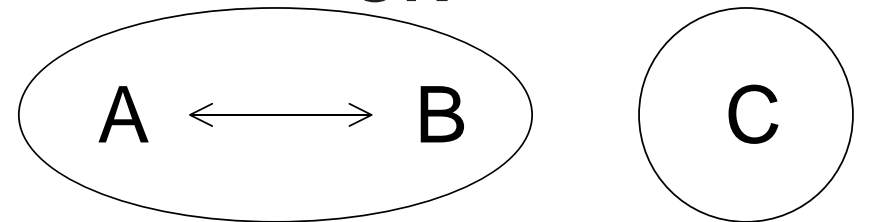
A, B, C
C, A, B
B, A, C
C, B, A

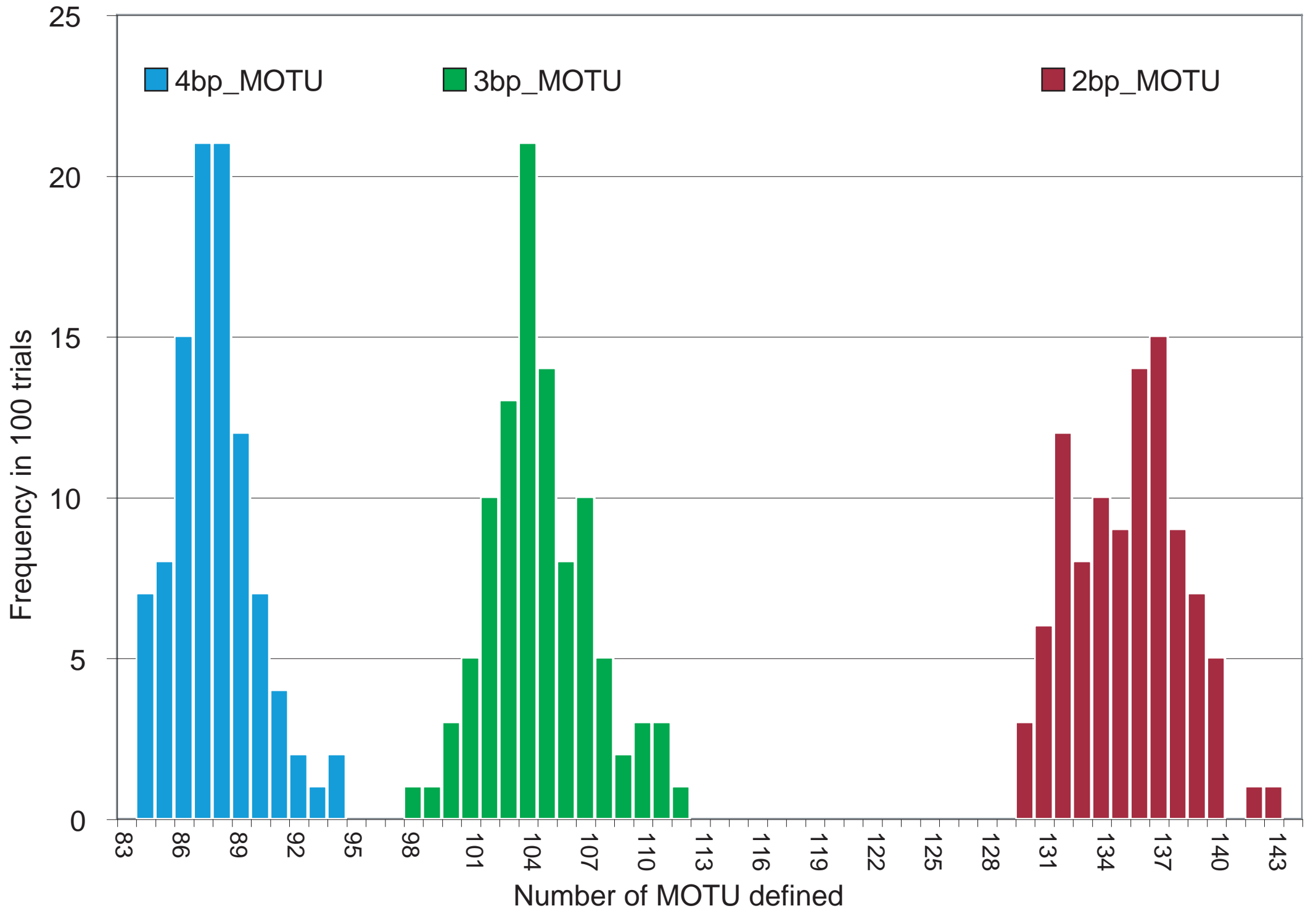
A, C, B
B, C, A

Resulting clusters



OR





4: Neglected meiofauna

ongoing surveys of tardigrades and nematodes in terrestrial and marine habitats

using specimen-centric and ‘ecosystem DNA’ approaches

5: Charismatic megafuna

a reanalysis of the *Astrartes* 'fulgerator'
dataset of CO1 barcodes

Hebert PD, Penton EH, Burns JM, Janzen DH, Hallwachs W.

Ten species in one: DNA barcoding reveals cryptic species in the neotropical skipper butterfly *Astrartes fulgerator*.

Proc Natl Acad Sci U S A. 2004;101:14812–7.

**MOTU defined using various cutoffs
in the *Astraptes 'fulgerator'* mtCOX1 barcode dataset**

