Insect multilocus metabarcoding: *in silico* evaluation of old and new primers.

7th International Barcode of Life Conference
Kruger National Park, South Africa 23rd Nov 2017

Daniel Marquina[1,2], Anders F. Andersson[3] & Fredrik Ronquist[1]

[1] Department of Bioinformatics and Genetics, Swedish Museum of Natural History
[2] Department of Zoology, Stockholm University
[3] Science for Life Laboratory School of Biotechnology, KTH Royal Institute of Technology

Presence of highly conserved regions delimiting highly variable regions in COI?

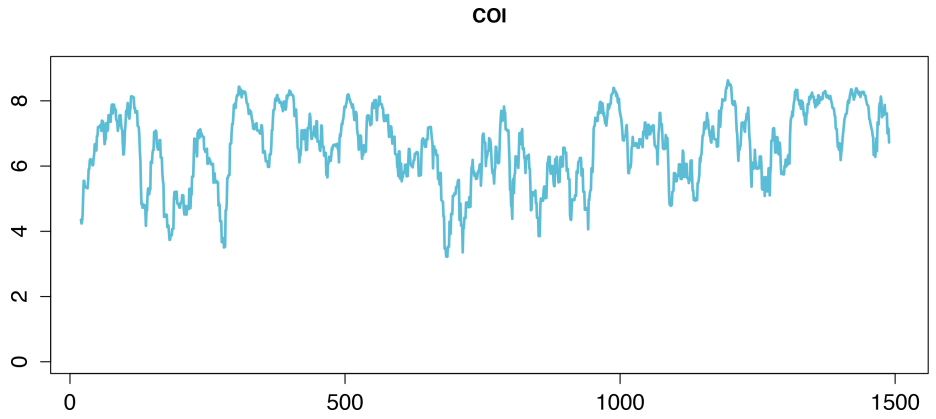
Universal primers with low degeneracy

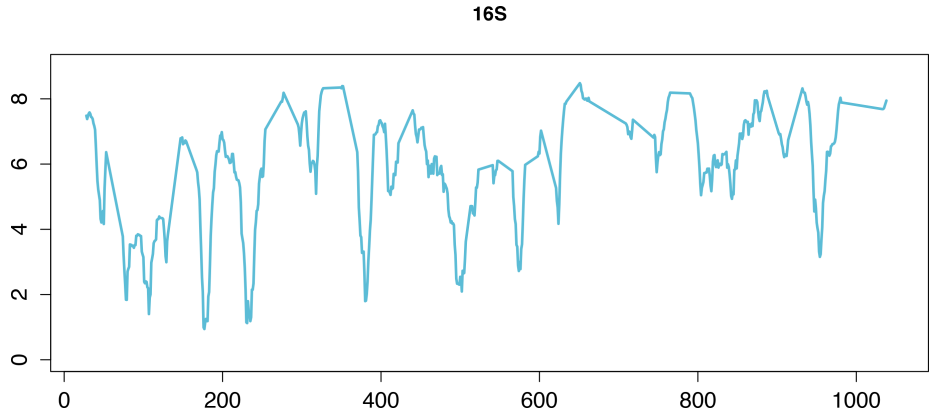
↓

Amplification bias

↓

Lower or null representation of some sequences





WORKFLOW

PRIMER DESIGN

1138 mitochondrial genomes

≈ 815 sp.

DegePrime:
– d=12 (low risk of non-target / primer dimers)
– d=48 (higher risk)

ecoPrimers:
Mismatches = 0 (70 %)
Mismatches = 3 (90 %)

IN SILICO PCR

1600 mitochondrial genomes

≈ 1115 sp.

ecoPCR:
NO MISMATCHES ALLOWED
PRIMER–TEMPLATE

Primers:
DegePrime 12/48
ecoPrimers
Previosuly Published

PRIMER EVALUATION

Taxonomic Coverage (B_C)

Taxonomic Resolution (B_S)

Exclusive Taxonomic Resolution (B_E)

Effective Taxonomic Resolution (ETR)

Combined Effective Taxonomic Resolution (ETR_C)
ETR_{SC} ETR_{RC}

Taxonomic resolution versus Exclusive Taxonomic Resolution

$$B_S/B_E = \frac{no.\ species\ unambiguously\ identified(*)}{no.\ species\ amplified} = [0\text{-}1]$$

MOTU 1

✓

Bombyx mori

Bombyx mori

Bombyx mori

$B_S = 3/4 = 0.75$

MOTU 2

✓

Drosophila simulans

Drosophila simulans

Drosophila simulans

MOTU 3

✓

Locusta migratoria

Locusta migratoria

Locusta migratoria

MOTU 4

✓

Drosophila simulans

Drosophila simulans

Drosophila simulans

MOTU 5

✗

Bombyx mori

Papilio machaon

Papilio machaon

Taxonomic resolution versus Exclusive Taxonomic Resolution

$$B_S/B_E = \frac{no.\ species\ unambiguously\ identified(*)}{no.\ species\ amplified} = [0\text{-}1]$$

MOTU 1

✗

Bombyx mori

Bombyx mori

Bombyx mori

$B_E = 1/4 = 0.25$

MOTU 2

✗

Drosophila simulans

Drosophila simulans

Drosophila simulans

MOTU 3

✓

Locusta migratoria

Locusta migratoria

Locusta migratoria

MOTU 4

✗

Drosophila simulans

Drosophila simulans

Drosophila simulans

MOTU 5

✗

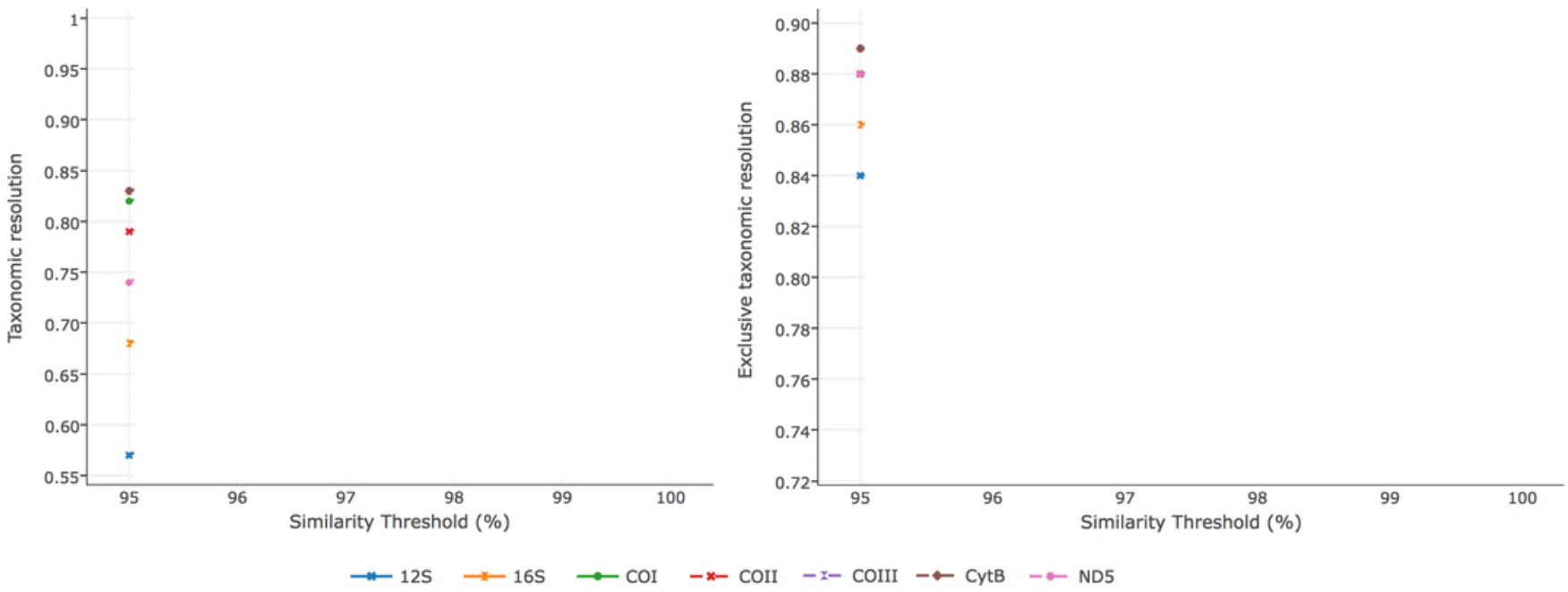
Bombyx mori

Papilio machaon

Papilio machaon

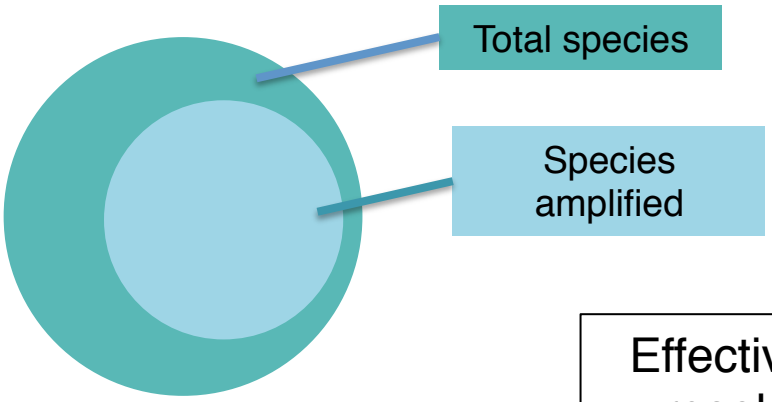
Taxonomic resolution versus Exclusive Taxonomic Resolution

B_E provides a more biologically accurate measure of the taxonomic resolution of a metabarcoding marker than B_S.

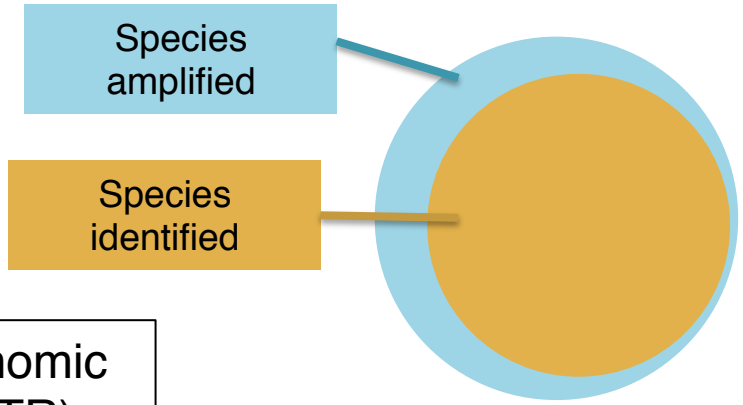


Effective Taxonomic Resolution

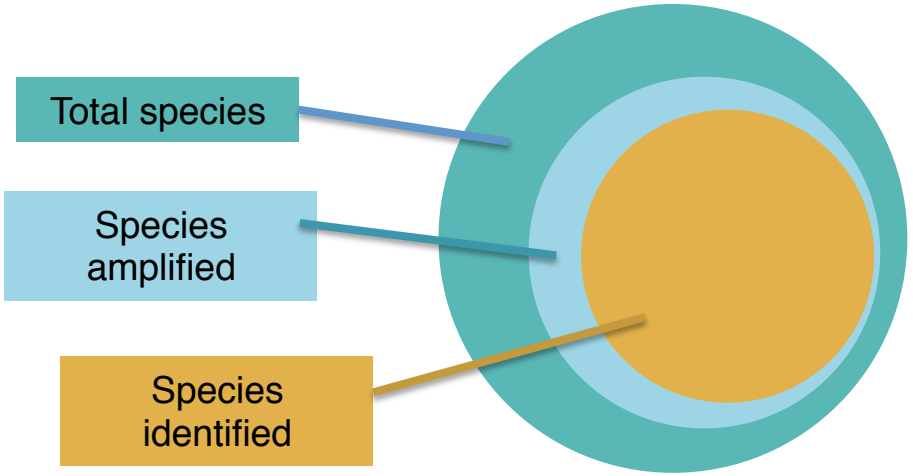
Taxonomic coverage (B_C)



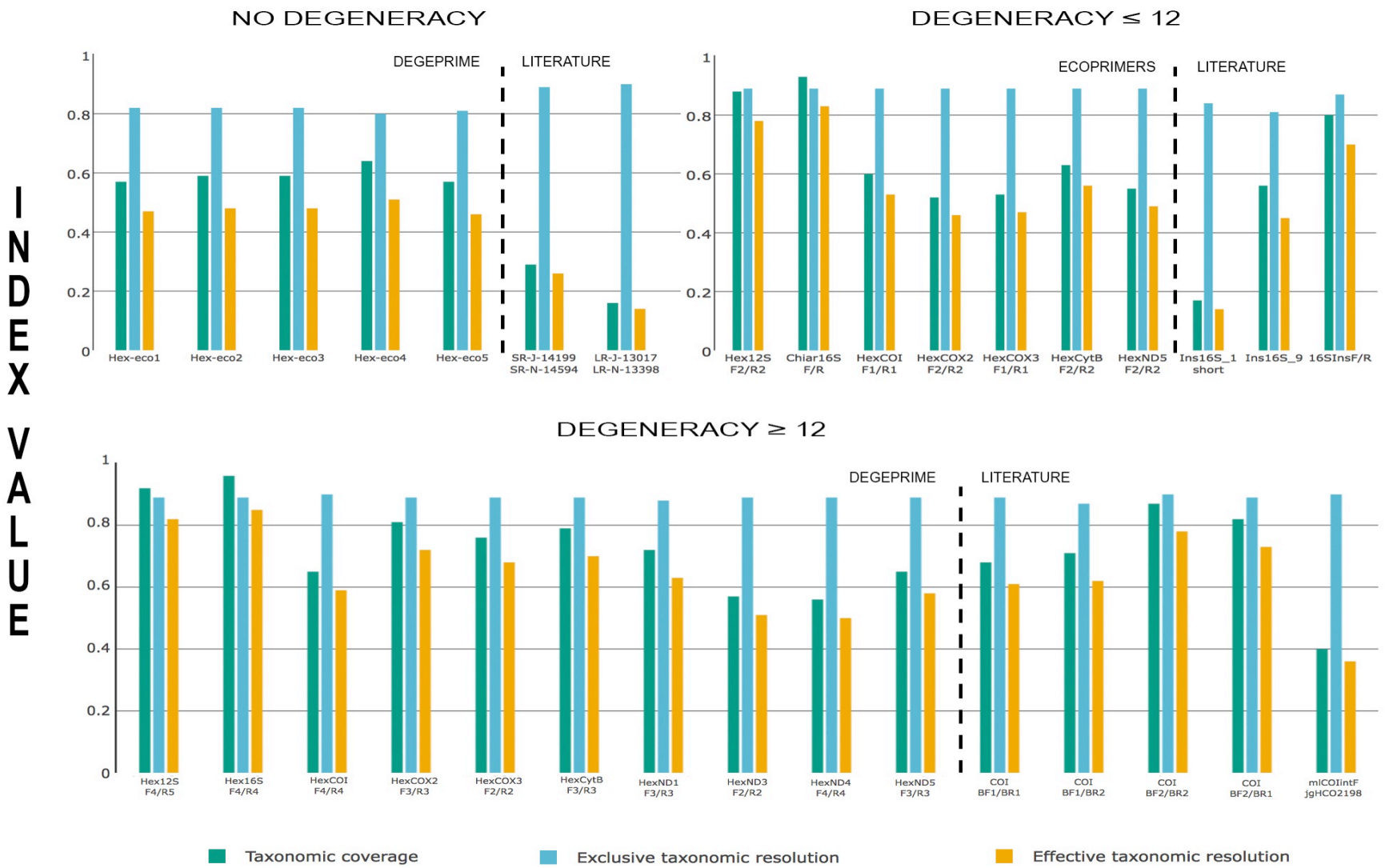
Exclusive taxonomic resolution (B_E)



Effective taxonomic resolution (ETR)



Which is the best primer pair for metabarcoding of Hexapoda?

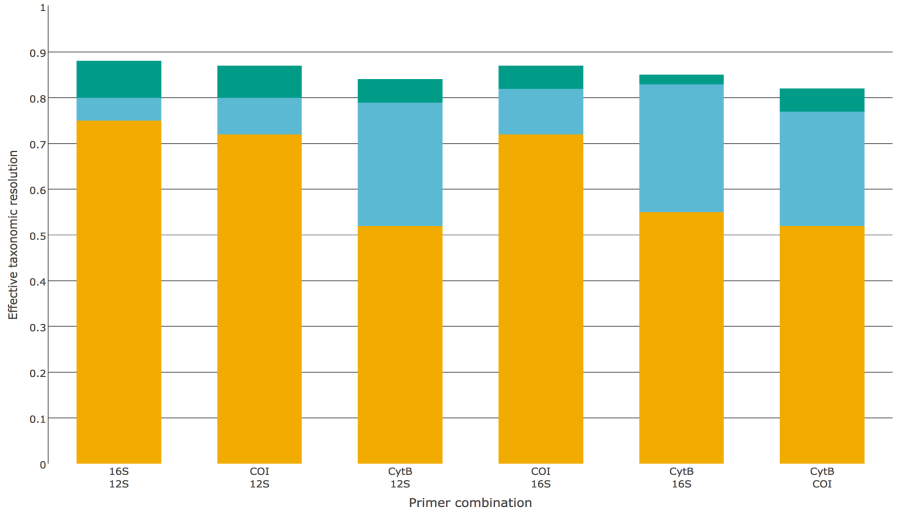


Are all markers equally good/bad for the different orders of Hexapoda?

Best combination of two markers

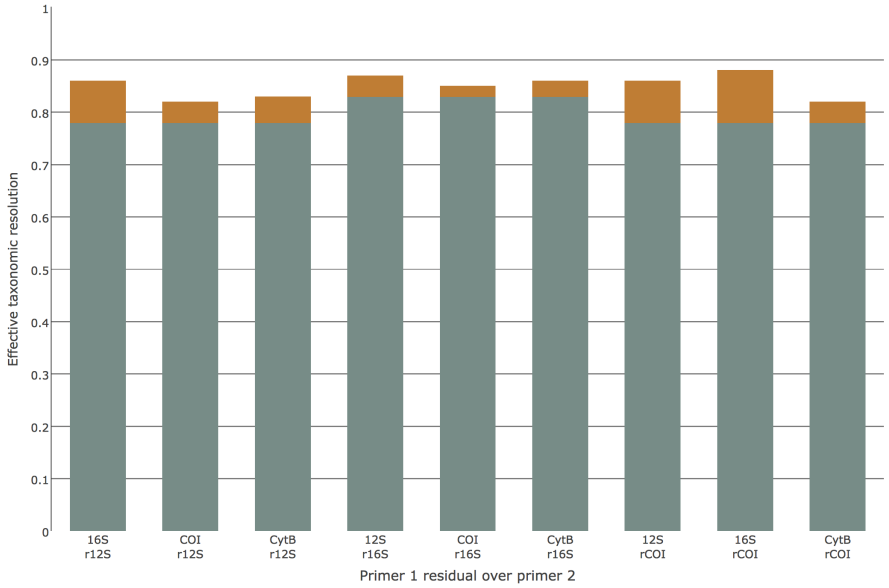
Simultaneously combined ETR:

- Primer pairs for two markers over the entire dataset.
- Redundant and uniquely contributed ETR calculated
- 12S + 16S / 16S + COI best combinations



Residually combined ETR:

- Primer pairs for one marker over the species not amplified by the other.
- Original and residual ETR calculated
- 16S + COI best combination
- For well-know biotas



In summary

B_E is a more biologically accurate indicator than B_S. Other indices useful for barcode and primer evaluation.

Degeneracy and strict PCR conditions allows for amplification with lower risk of bias.

All mitochondrial genes provide good resolution. Lack of reference data impedes their use.

Best strategy: simultaneously sequence COI and 16S (megadiverse biotas) or sequence COI and 16S residually (well-known biotas).

Thanks for listening!

Questions?

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 642241.

Banner and background: Erik Erskmark (NRM-BIO)