

DESIGN AND IN SILICO EVALUATION OF DEGENERATE PCR PRIMERS FOR INSECT METABARCODING

Daniel Marquina^{1,2}, Anders F. Andersson³, Fredrik Ronquist¹

¹Department of Bioinformatics and Genetics, Swedish Museum of Natural History, Stockholm, Sweden

²Department of Zoology, Stockholm University, Stockholm, Sweden

³Science for Life Laboratory, School of Biotechnology, Division of Gene Technology, KTH Royal Institute of Technology, Stockholm, Sweden

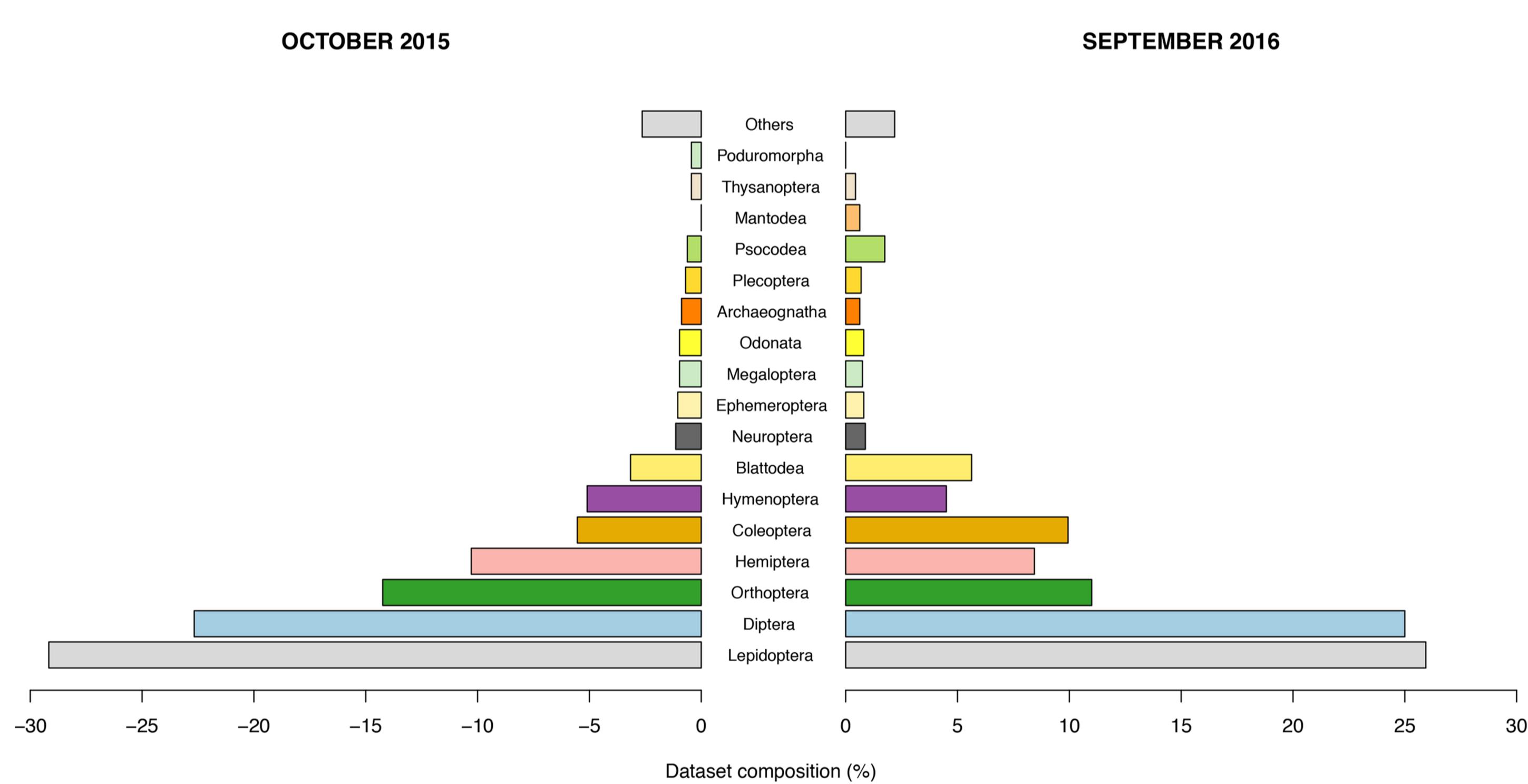
INTRODUCTION

Metabarcoding emerged as a high-throughput solution for DNA barcoding of samples containing large number of individuals (e.g., Malaise trap or pit-fall trap samples) or environmental DNA samples in which the organisms are not physically present. For animals, metabarcoding has been based mainly on sequencing short fragments within the ‘barcode region’ of the mitochondrial gene cytochrome oxidase I (COI). Nevertheless, several problems with PCR primers, resulting from the hypervariability of the third position of each codon of the gene, have been documented; these problems are potentially confounding and may introduce bias during the amplification step. These errors can cause less reliable results to be generated or impede the detection of species that are indeed present in the sample and the habitat.

For insects, several attempts have been made to find a universal and non-biased primer that can amplify a barcode fragment. In general, these primers present a very low success amplifying all the species present in a previously-known composition artificial mix. That is caused mainly because, until very recently, they didn't include any degeneracy and they used relatively few species.

MATERIALS AND METHODS

Two datasets (D1 and D2) of all available mitochondrial genomes were downloaded from GenBank (D1, October 2015; D2, September 2016). Primer design was conducted over D1 using two different software: DEGEPRIME and ECOPRIMERS.



After a *in silico* PCR amplification using the softwares ECOPCR and OBITOOLS, primers and resulting barcodes were evaluated by calculating the following indexes:

Taxonomic coverage

$$Bc = \frac{\text{no. species amplified}}{\text{total no. species}}$$

Taxonomic resolution

$$Bs = \frac{\text{no. species identified (incl. repeated clusters)}}{\text{total no. species}}$$

Exclusive taxonomic resolution

$$Bs-e = \frac{\text{no. species identified (excl. repeated clusters)}}{\text{total no. species}}$$

Total taxonomic resolution

$$TTR = Bc \cdot Bs-e$$

Combined total taxonomic resolution

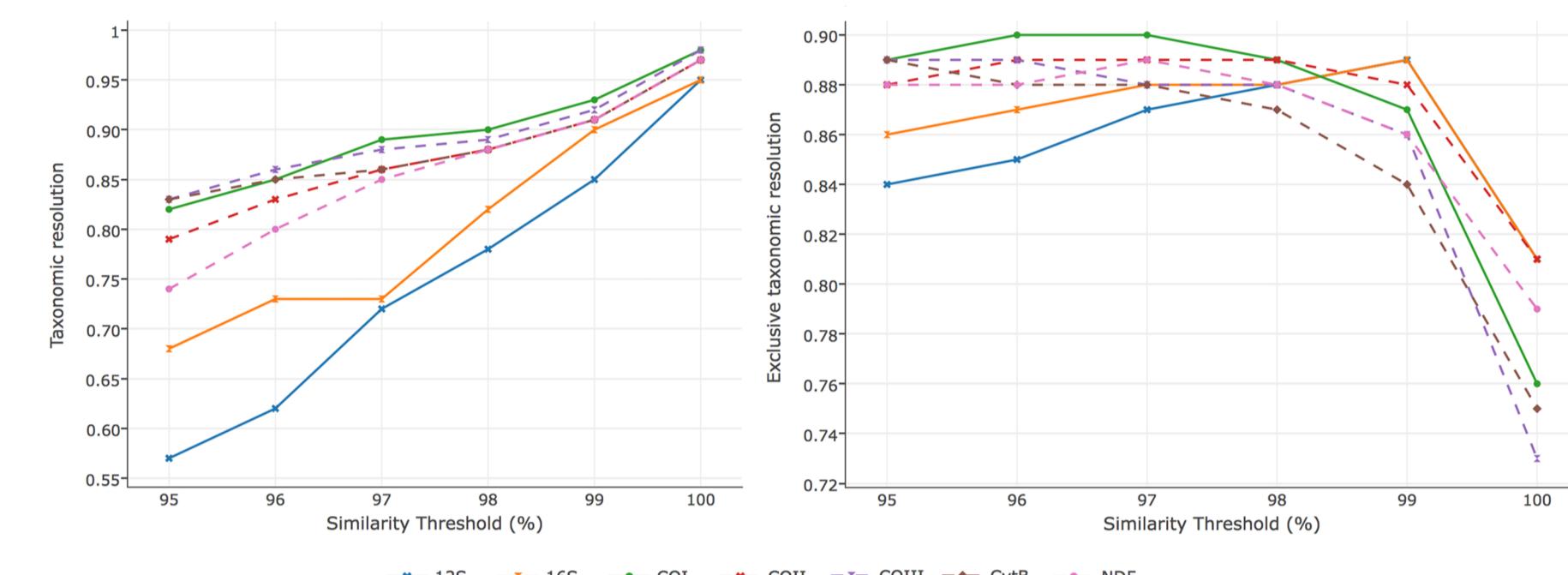
$$TTRc = TTRr + TTRu$$

$$TTRr = (Bc_1 + Bc_2) \cdot Bs-e_{\text{Higest 1-2}}$$

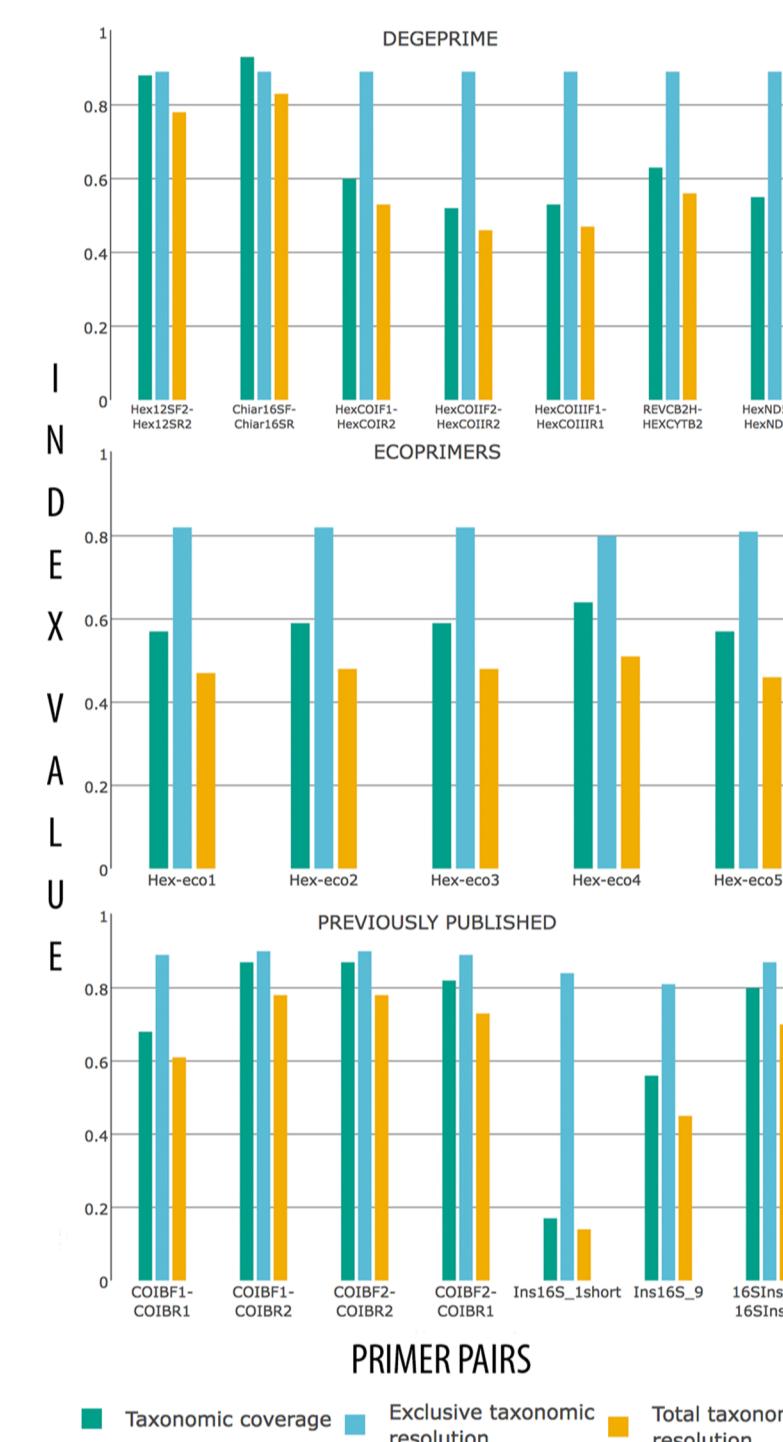
$$TTRu = (Bc_{u1} \cdot Bs_{u1}) + (Bc_{u2} \cdot Bs_{u2})$$

RESULTS

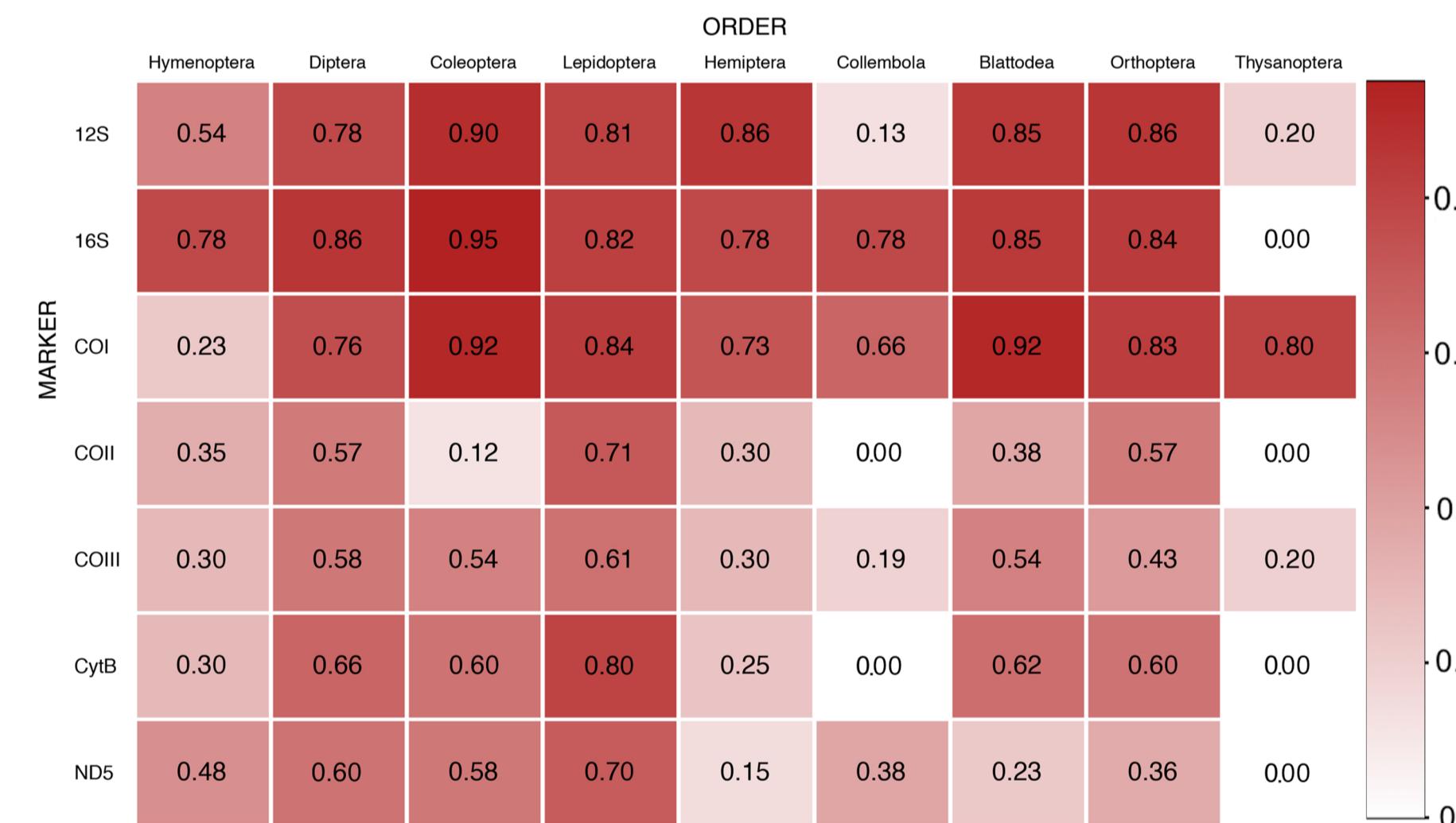
1. Exclusive taxonomic resolution is a better indicator of the barcoding capability of a marker.



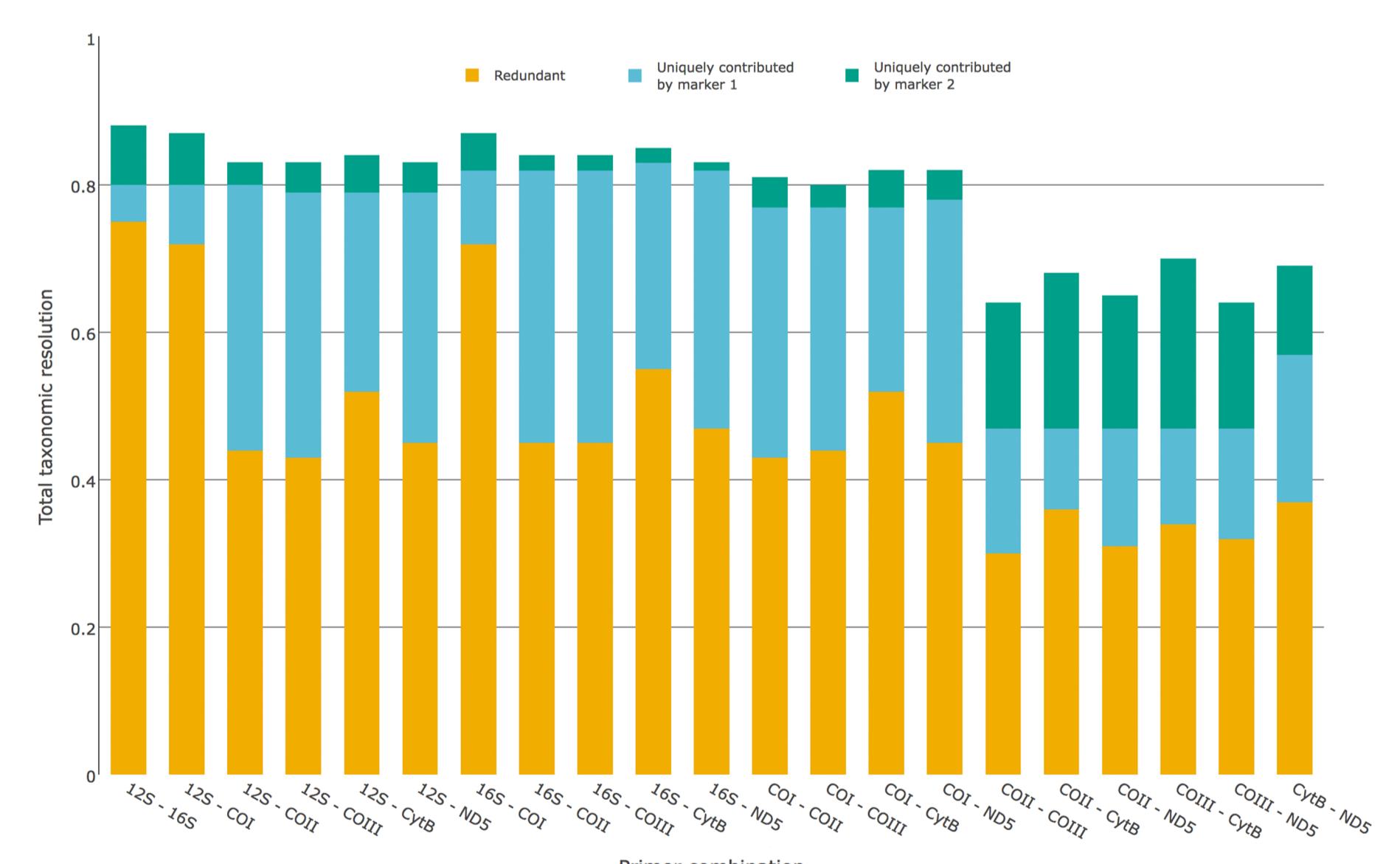
2. DEGEPRIME is a better tool for primer design when using large number of sequences than ECOPRIMERS. For all evaluated genes, except COI, the primers designed with DEGEPRIME outperformed the already published primers.



3. 16S is the best marker for 4 of the 8 most abundant orders in bulk samples, followed by COI (best for 3) and 12S (best for 1). Although 16S is more constant though orders, there is much more reference data for COI.



4. Combining two genes in the metabarcoding pipeline will increase total detection and identification of species up to 90 %.



CONCLUSION: The use of degenerate primers allows to amplify a larger proportion of the insects in a sample in contrast to those primers that doesn't include wobble bases. 12S, 16S and COI are the best metabarcoding markers for insects, and COI should always be present. The development of references of whole mitochondrial genomes will allow to use better markers (12S and 16S) and, locally or for smaller groups, other markers with good resolution capacity but today out of use.