The “MetaCopepod” project: Designing an integrated DNA metabarcoding and image analysis approach to study and monitor biodiversity of zooplanktonic copepods.

Panagiotis Kasapidis

Hellenic Centre for Marine Research (HCMR),
Institute of Marine Biology, Biotechnology and Aquaculture
P.O.Box 2214, 71003 Heraklion Crete, Greece

Ευρωπαϊκή Ένωση
Ευρωπαϊκό Κοινωνικό Ταμείο
ΕΠΙΧΕΙΡΗΣΙΑΚΟ ΠΡΟΓΡΑΜΜΑ
ΕΚΠΑΙΔΕΥΣΗ ΚΑΙ ΔΙΑ ΒΙΟΥ ΜΑΘΗΣΗ
επένδυση στην κοινωνία της γνώσης
ΥΠΟΥΡΓΕΙΟ ΠΑΙΔΕΙΑΣ ΚΑΙ ΘΡΗΣΚΕΥΜΑΤΩΝ
ΕΙΔΙΚΗ ΥΠΗΡΕΣΙΑ ΔΙΑΧΕΙΡΙΣΗΣ
ΕΣΠΑ 2007-2013
ΕΥΡΩΠΑΪΚΟ ΚΟΙΝΩΝΙΚΟ ΤΑΜΕΙΟ
Με τη συγχρηματοδότηση της Ελλάδας και της Ευρωπαϊκής Ένωσης
NSRF 2007-2013
programme for development

The “MetaCopepod” project

Aim: to develop a novel methodology, based on the combination of **DNA metabarcoding** and **image analysis**, to assess and monitor the diversity of marine zooplanktonic copepods (and cladocera), in the Mediterranean and the Black Sea, in a high-throughput, cost-effective, accurate and quantitative way.

Coordinator: Dr. Panagiotis Kasapidis, Hellenic Centre for Marine Research (HCMR), GREECE

Study area: Mediterranean and the Black Sea

Duration: Feb. 2014 – Oct. 2015

Budget: 180,000 euros

Funding: European Social Fund (ESF) and National Funds through the National Strategic Reference Framework (NSRF) 2007-2013, Operational Programme "Education and Life-Long Learning", Action "ARISTEIA II", Greek Ministry of Education and Religious Affairs, General Secretary of Research and Technology.

Studying zooplankton diversity: limitations of traditional approaches

- Quite laborious (sorting, identification under stereoscope) → bottleneck in sample processing.
- Requires local taxonomic expertise
- Difficult to identify immature stages
- Misidentifications
- Cryptic species

Image analysis

+ Pros

- High throughput
- Quantitative results (abundance, size spectra, biomass)

- Cons

- Low resolution → can recognize taxa similar to the ability of a trained taxonomists to identify a taxon under stereoscope at a glance
- “train” image analysis software separately for different types of zooplankton communities (not once for all).

DNA metabarcoding (NGS analysis) for studying marine biodiversity

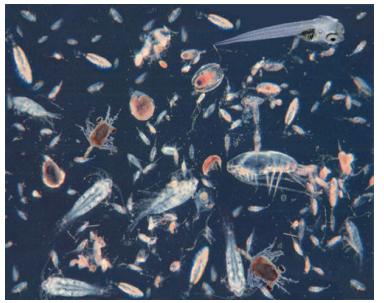
+ Pros:

- Faster processing of the samples
- Potentially higher accuracy in species' identification (even for difficult taxa, immature stages, cryptic species)
- No need of taxonomy expert but requires a well curated and complete reference genetic database

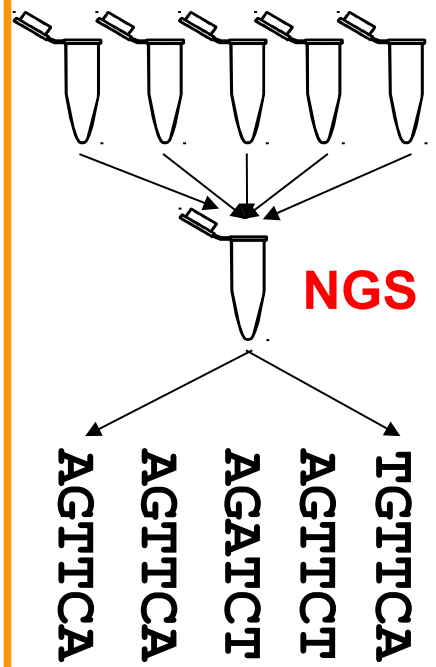
- Cons:

- Results are semiquantitative due to the PCR
- Biases in species' relative abundance mostly due to PCR amplification bias

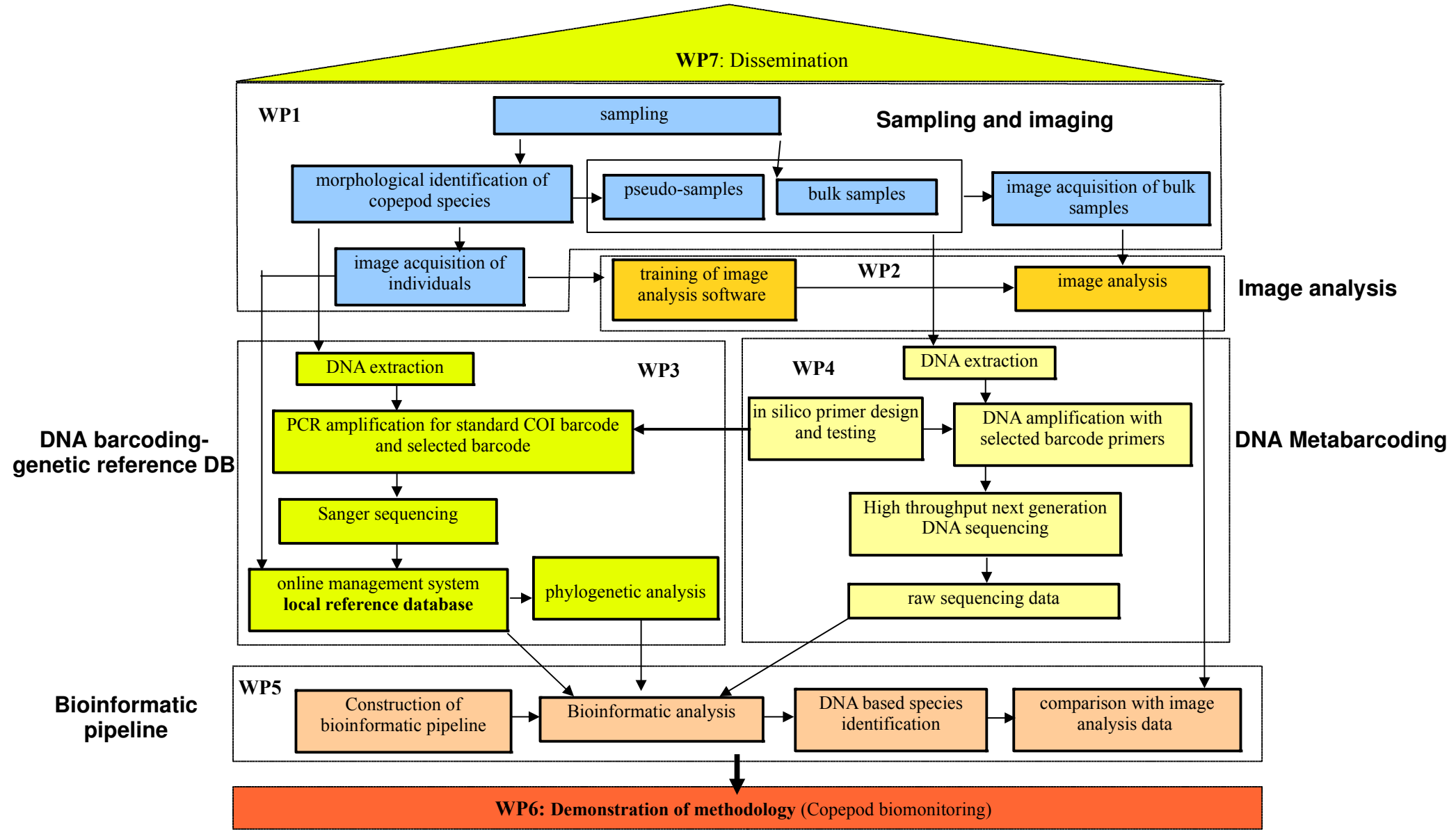
The "MetaCopepod" project aims to combine the Pros of both methods in order to increase accuracy in assessing zooplankton diversity.



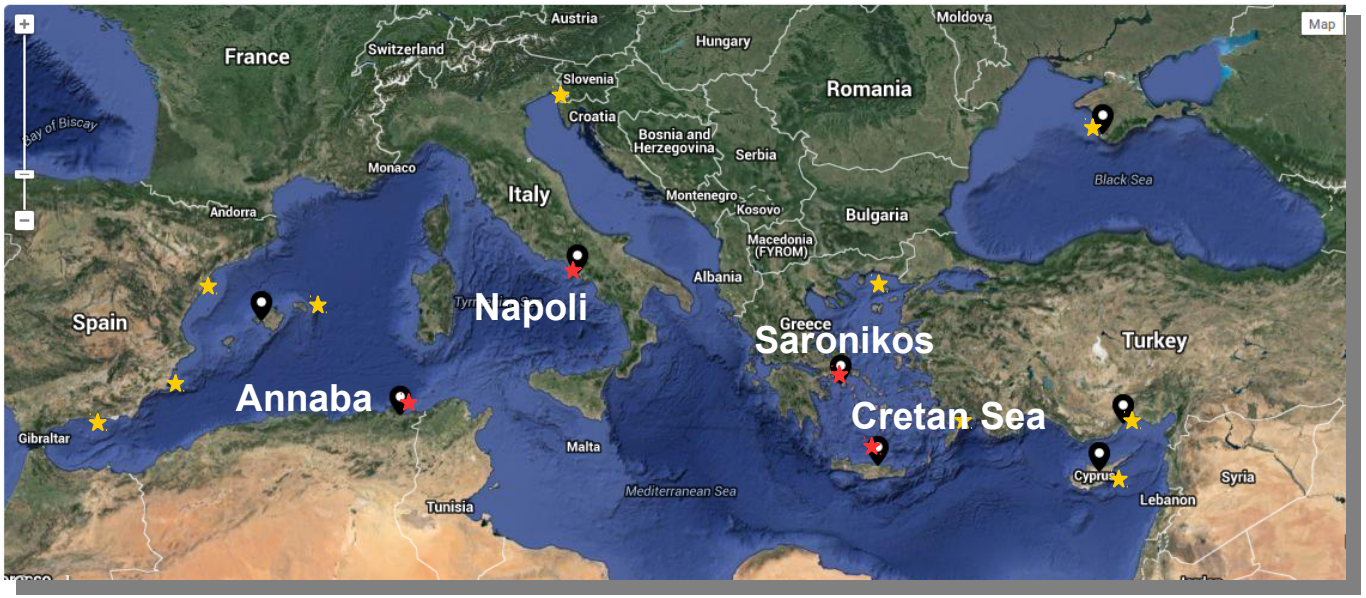
bulk sample

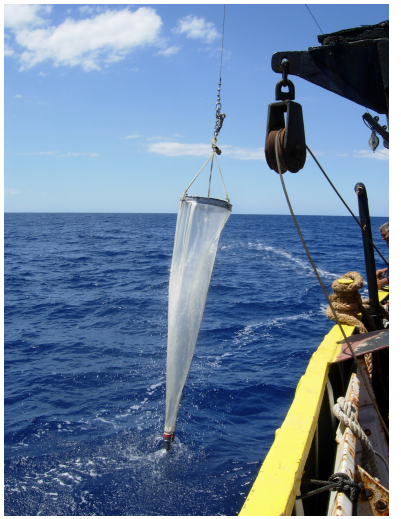


Workflow of the “MetaCopepod” project



Sampling





Partner Network and sampling stations.
Monthly sampling (★), non-regular sampling (★)

Samples

- 108 zooplankton samples collected and preserved in 95% EtOH
- 80 zooplankton samples were scanned for image analysis and then used for DNA metabarcoding
- 97 copepod species were morphologically identified (>350 specimen collections from different zooplankton samples)

For standardizing the methodology:

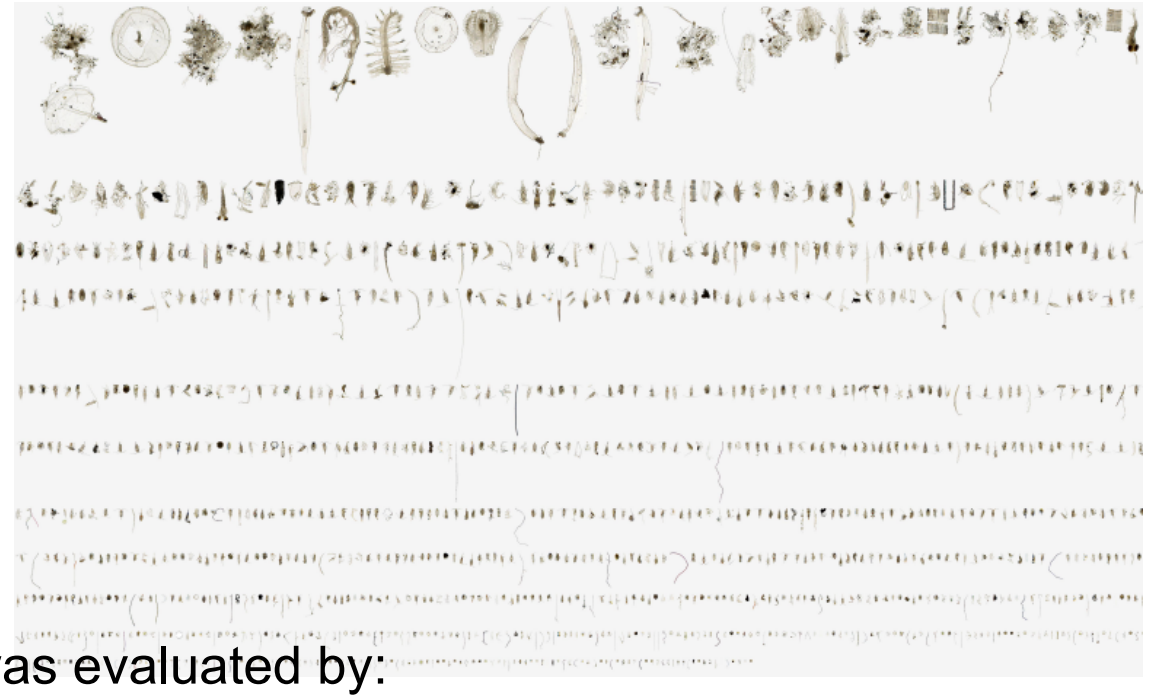
- Five of the zooplankton samples were taxonomically identified by taxonomist
- Six pseudosamples (mock samples) created by mixing taxonomically identified specimens at known numbers (6-16 species, 1-3 individuals per species)

Image analysis

(Epson scanner similar to Zooscan and ZooImage software)

For “training” the image analysis software we used:

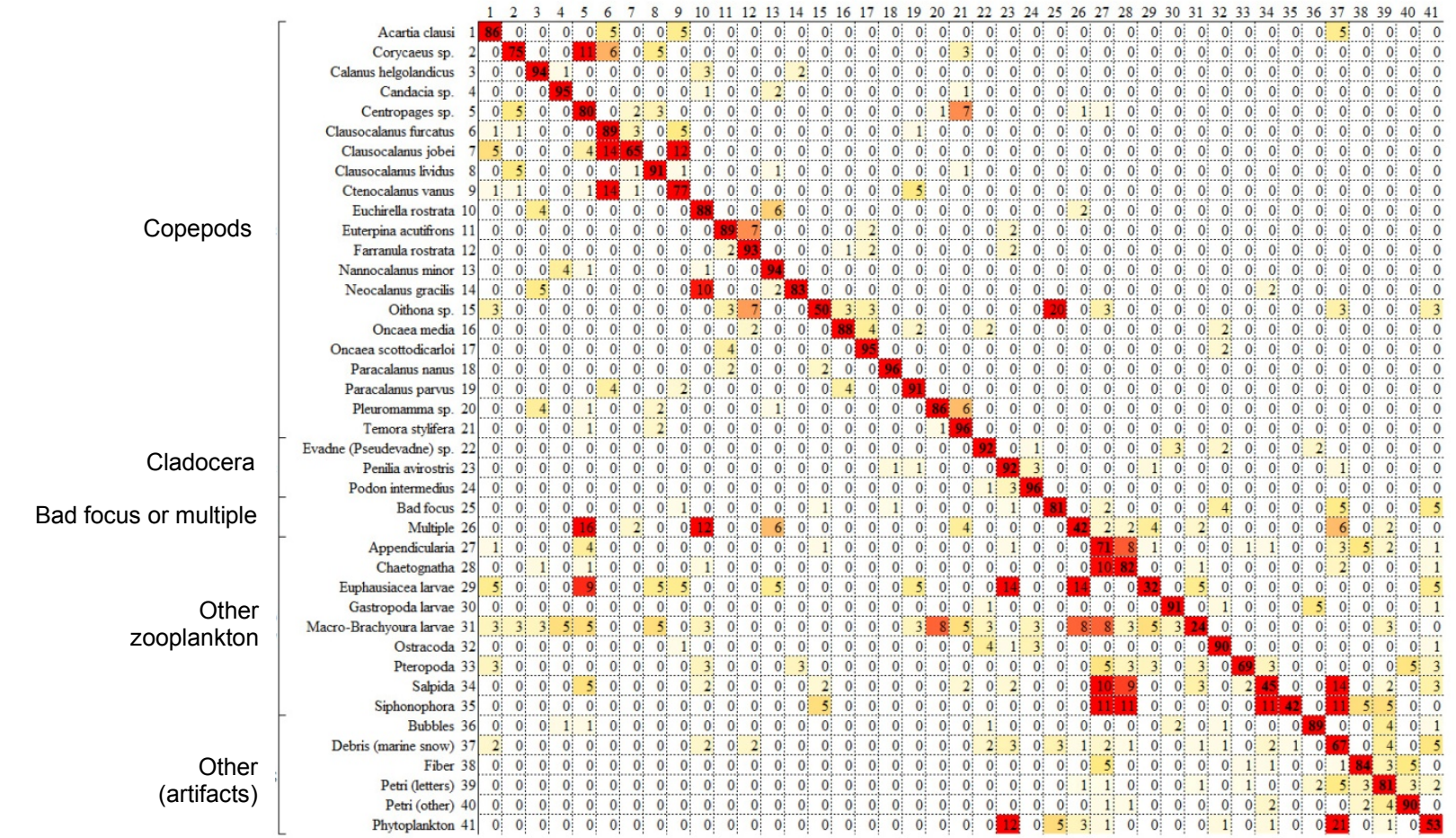
- scanned images of taxonomically identified taxa (“gold” standards)
- taxonomically identified taxa from scanned zooplankton images (“silver” standards)



The image analysis software was evaluated by:

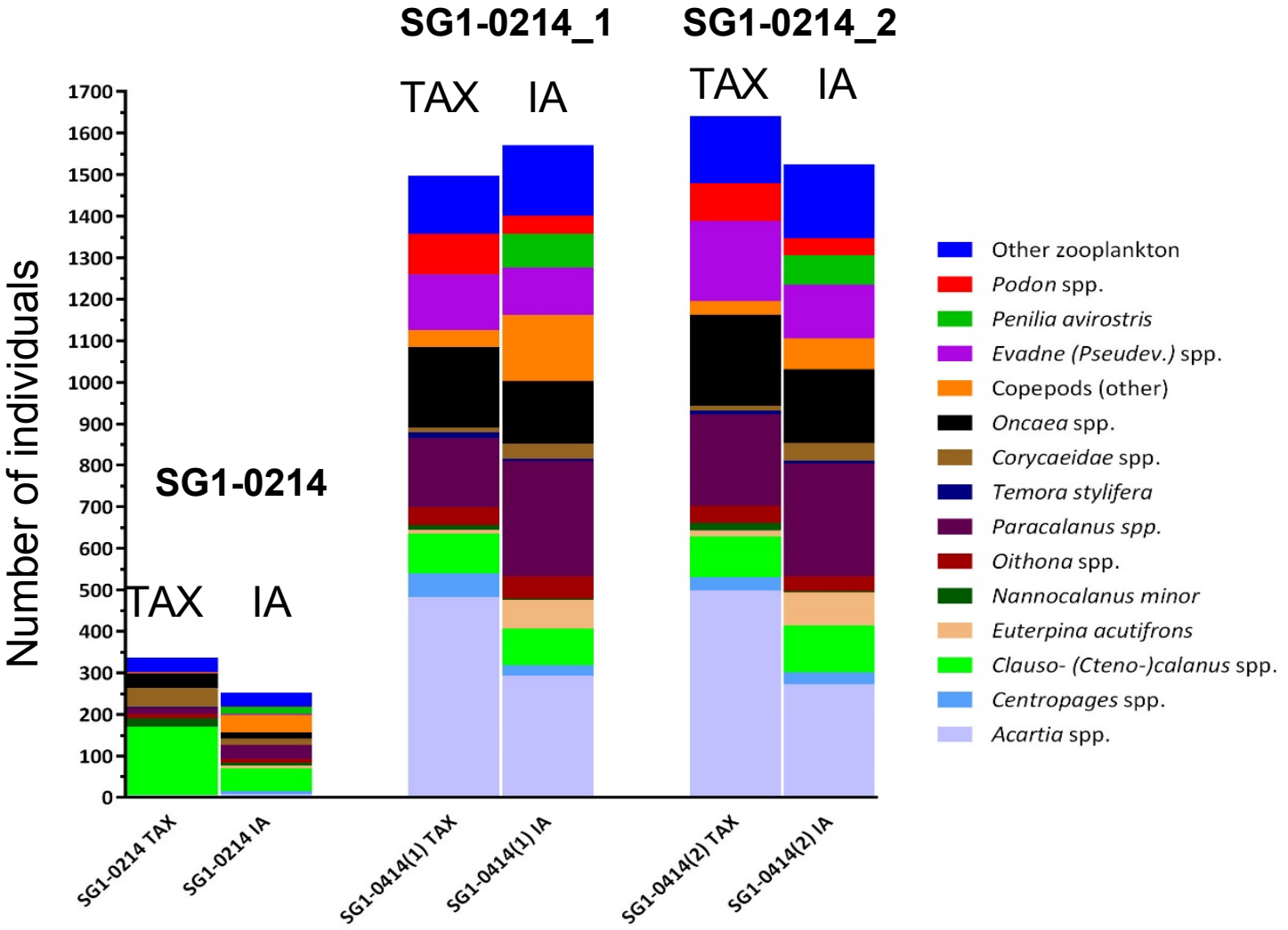
- self-evaluation
- using pseudosamples and taxonomically identified samples

Self evaluation of the Image Analysis software



Mostly correct identification, but confusion in certain taxa (e.g. *Clausocalanus* and *Ctenocalanus*)

Evaluation of the Image Analysis software with taxonomically identified samples



- Three samples from Saronikos Gulf analyzed both by a taxonomist and by IA
- IA software performs well both for taxa recognition and abundance estimation

Categories identified with accuracy by Image analysis (standardized for Saronikos Gulf)

Larger categories	Categories for Image Analysis	Taxa in each category
Big Calanoida	Calanus helgolandicus	Calanus_helgolandicus
	Candacia-Paracandacia	Candacia Paracandacia simplex
	Euchirella	Euchirella
	Nannocalanus minor	Nannocalanus minor
	Neocalanus	Neocalanus
	Pleuromamma	Pleuromamma
Medium Calanoida	Acartiidae	Acartia Paracartia grani
	Centropagidae	Centropages
	Temora	Temora Aetidus Lucicutia
Small Calanoida	Small Calanoida	Calocalanus Clausocalanus Paracalanus Ctenocalanus vanus Phaenna spinifera Scolecithricella & Scolecithrix Isias clavipes Mecynocera clausi
Cyclopoida	Oithona	Oithona
Harpacticoida	Euterpina acutifrons Clytemnestra	Euterpina acutifrons Clytemnestra other Harpacticoida n.d.
	Micro & Macrosettela	Micro & Macrosettela
Poecilostomatoida	Coryceaidae	Coryceaus Farranula rostrata
	Oncaeidae	Oncaea Triconia
Ctenopoda	Penilia_avirostris	Penilia avirostris
Onychopoda	Onychopoda	Evadne Pseudevadne Podon Pleopis polyphemoides

Critical factors for DNA metabarcoding

- **Primers**
  - designed on conserved regions across target taxa→ reduce amplification bias
  - amplify a variable region → high resolution, ideally at species' level
  - amplify a short region→ easier PCR amplification even for degraded samples

- **Genetic reference database**
  - well curated (no errors or misidentifications)
  - as complete as possible

Primer design and reference database

Primers were designed (ecoPrimer software), which amplify a region of ~150 bp of the mitochondrial 16S rRNA gene.

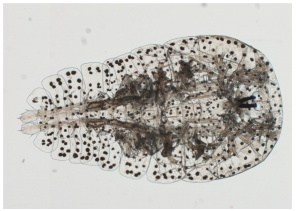
Reference database (GenBank largely incomplete for 16S rRNA)

Out of the 97 taxa taxonomically identified

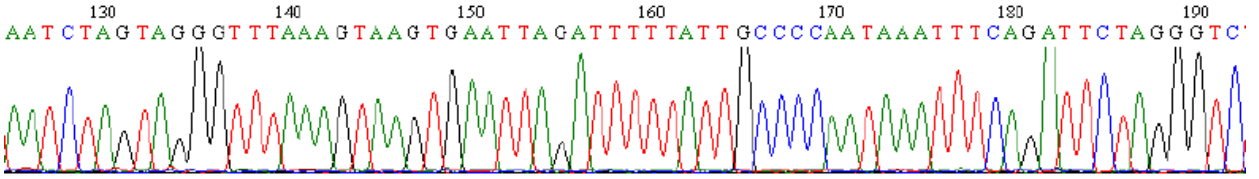
- 93 lacked 16S sequences and 36 lacked COI sequences in GenBank

We performed DNA barcoding for both COI and 16S genes

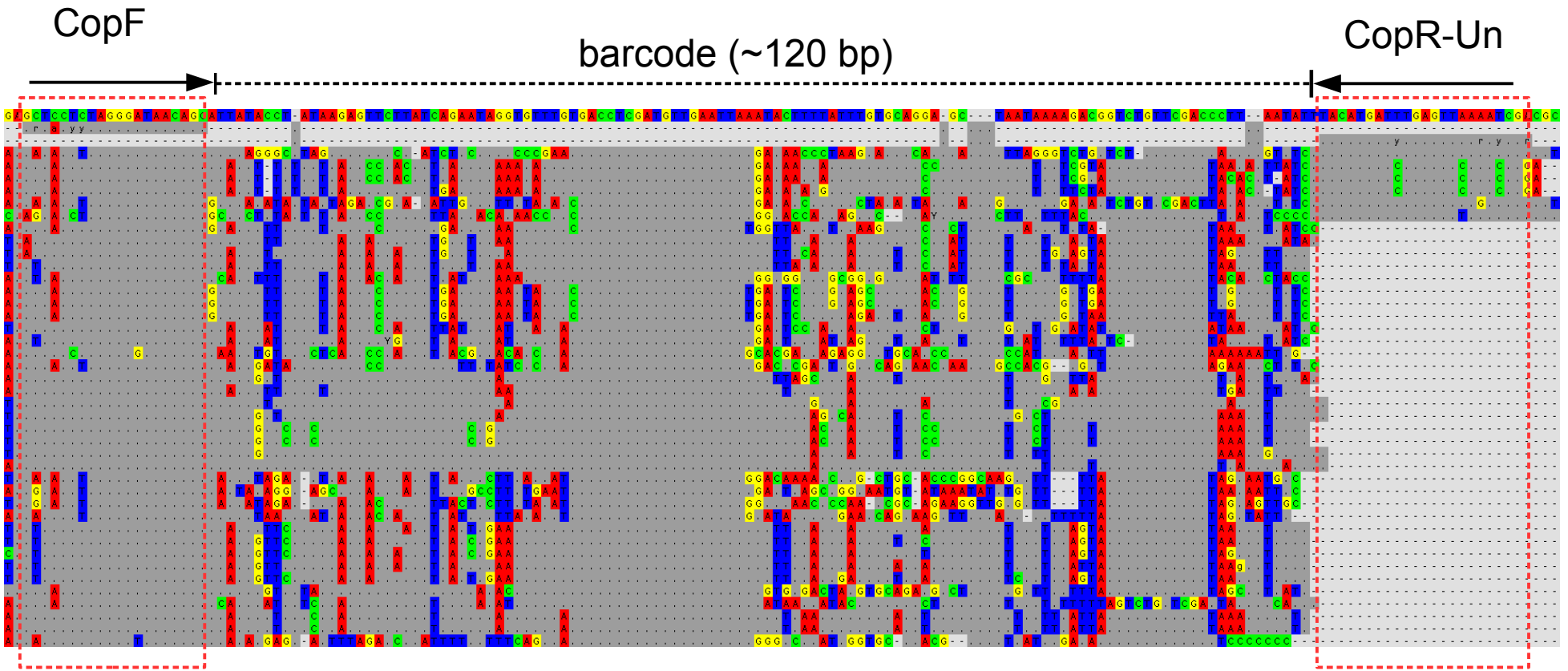
- 50 new additions of 16S barcode (55 species sequenced in total)
- 8 new additions for standard COI barcode (48 species sequenced in total)



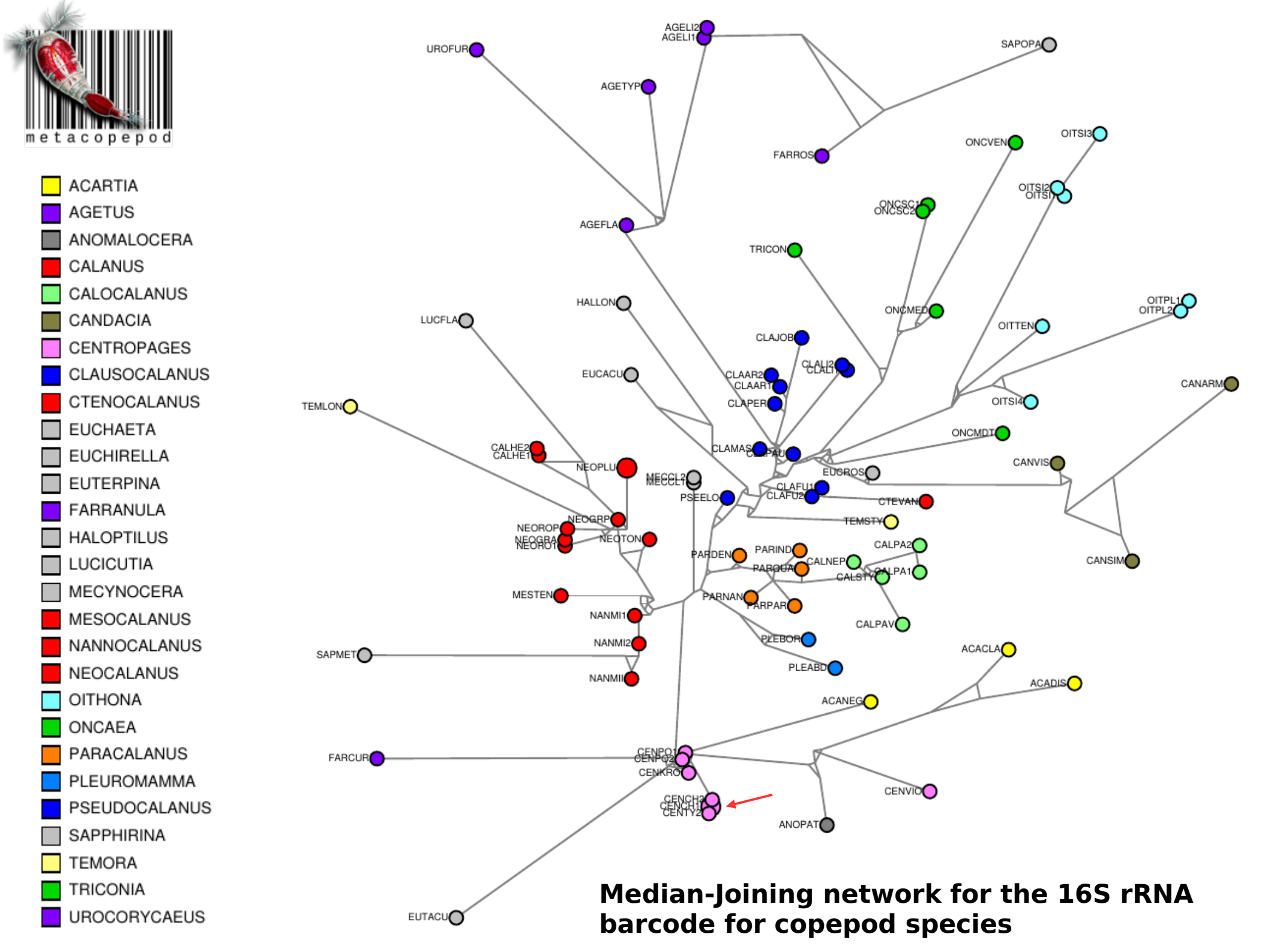
Sapphirina opalina



Evaluation of the 16S barcode to discriminate species



Alignment of 16S copepod sequences from NCBI and from MetaCopepod (produced with universal primers)


metacopepod
ACARTIA
AGETUS
ANOMALOCERA
CALANUS
CALOCALANUS
CANDACIA
CENTROPAGES
CLAUSOCALANUS
CTENOCALANUS
EUCHAETA
EUCHIRELLA
EUTERPINA
FARRANULA
HALOPTILUS
LUCICUTIA
MECYNOCERA
MESOCALANUS
NANNOCALANUS
NEOCALANUS
OITHONA
ONCAEA
PARACALANUS
PLEUROMAMMA
PSEUDOCALANUS
SAPPHIRINA
TEMORA
TRICONIA
UROCORYCAEUS
AGELI2
AGELI1
UROFUR
AGETYP
SAPOPA
FARROS
ONCVEN
OITSI3
OITSI2
OITSI1
ONCSC1
ONCSC2
AGEFLA
TRICON
HALLON
ONCMED
OITPL1
OITPL2
LUCFLA
OITTEN
CLAJOB
CLALI2
CLALI1
EUCACU
CLAAR2
CLAART1
CANARM
CLAPER
TEMLON
OITSI4
ONCMDT
CALHE2
CALHE1
CLAMAS
CLAPAU
NEOPLU
MECCL2
MECCL1
EUCROS
CANVIS
CLAFU1
PSEELO
CLAFU2
CTEVAN
NEOGRP
TEMSTY
NEOROP
NEOGBA
NEOROT
NEOTON
PARDEN
PARIND
CALPA2
PAROUA
CALNEP
CANSIM
CALSTY
CALPA1
MESTEN
PARNAN
PARPAR
NANMI1
CALPAV
PLEBOR
NANMI2
SAPMET
ACACLA
PLEABD
NANMI3
ACADIS
ACANEG
FARCUR
CENPO1
CENPG2
CENKRO
CENVIO
CENCH2
CENCH1
CENTY2
ANOPAT
EUTACU
Median-Joining network for the 16S rRNA barcode for copepod species

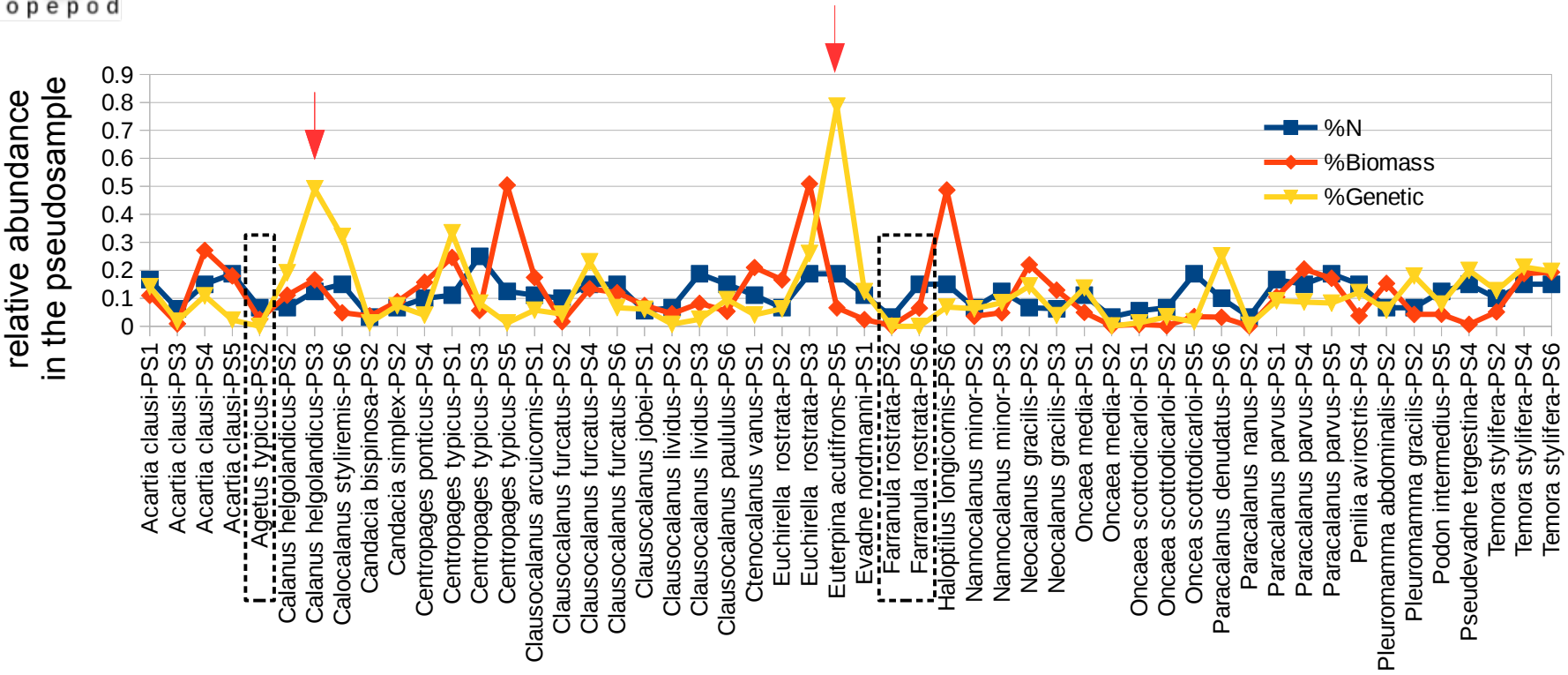
DNA metabarcoding of zooplankton samples

- Unsorted zooplankton samples, taxonomically identified samples and pseudosamples → DNA extraction, PCR and sequencing on a MiSeq Illumina platform.

- Raw sequencing data analyzed using a bioinformatic pipeline constructed for the project.

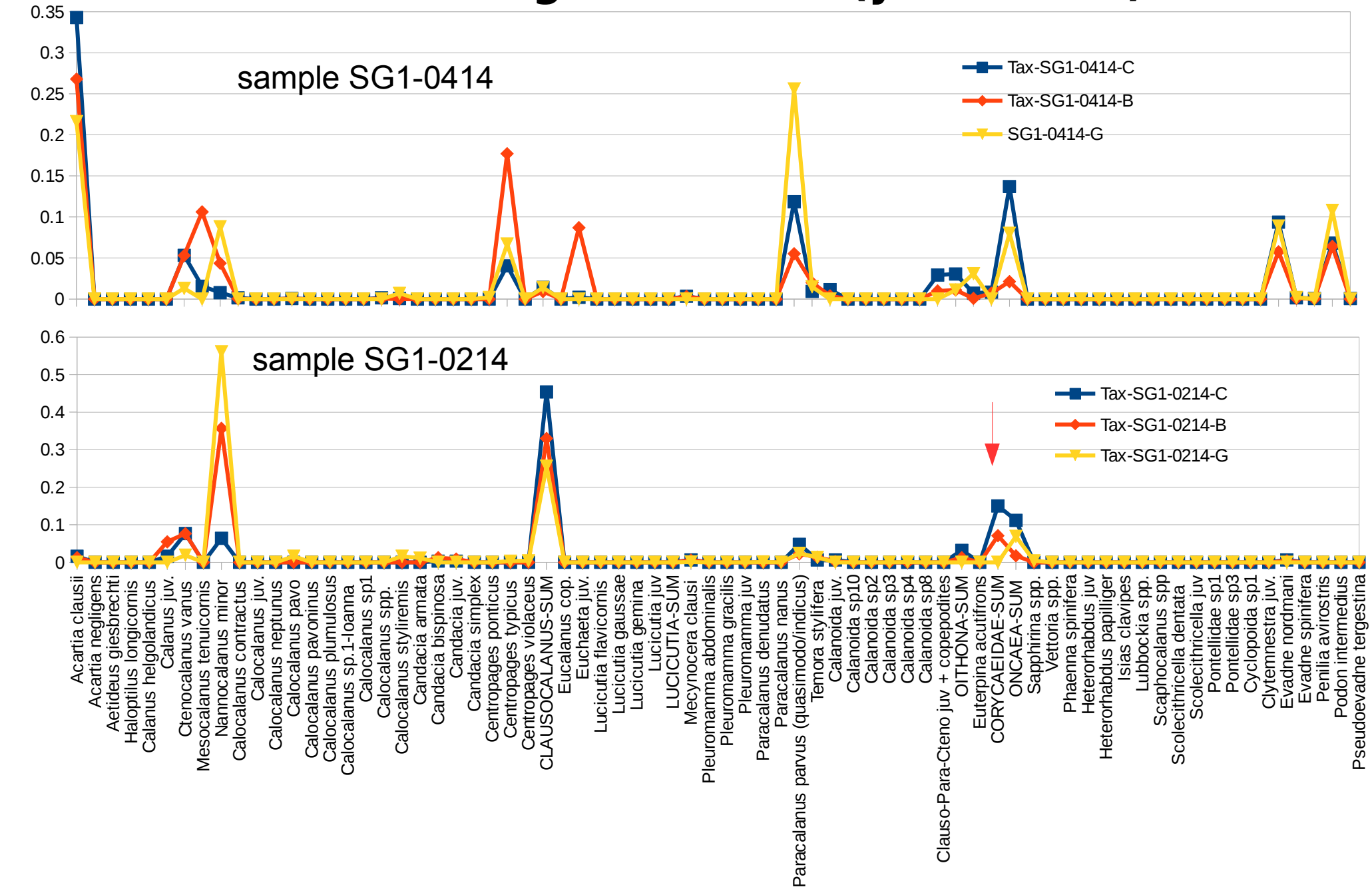
- Sequences assigned to taxa using the reference database

Pseudosamples: comparison between genetic data, % abundance (%N) and % biomass

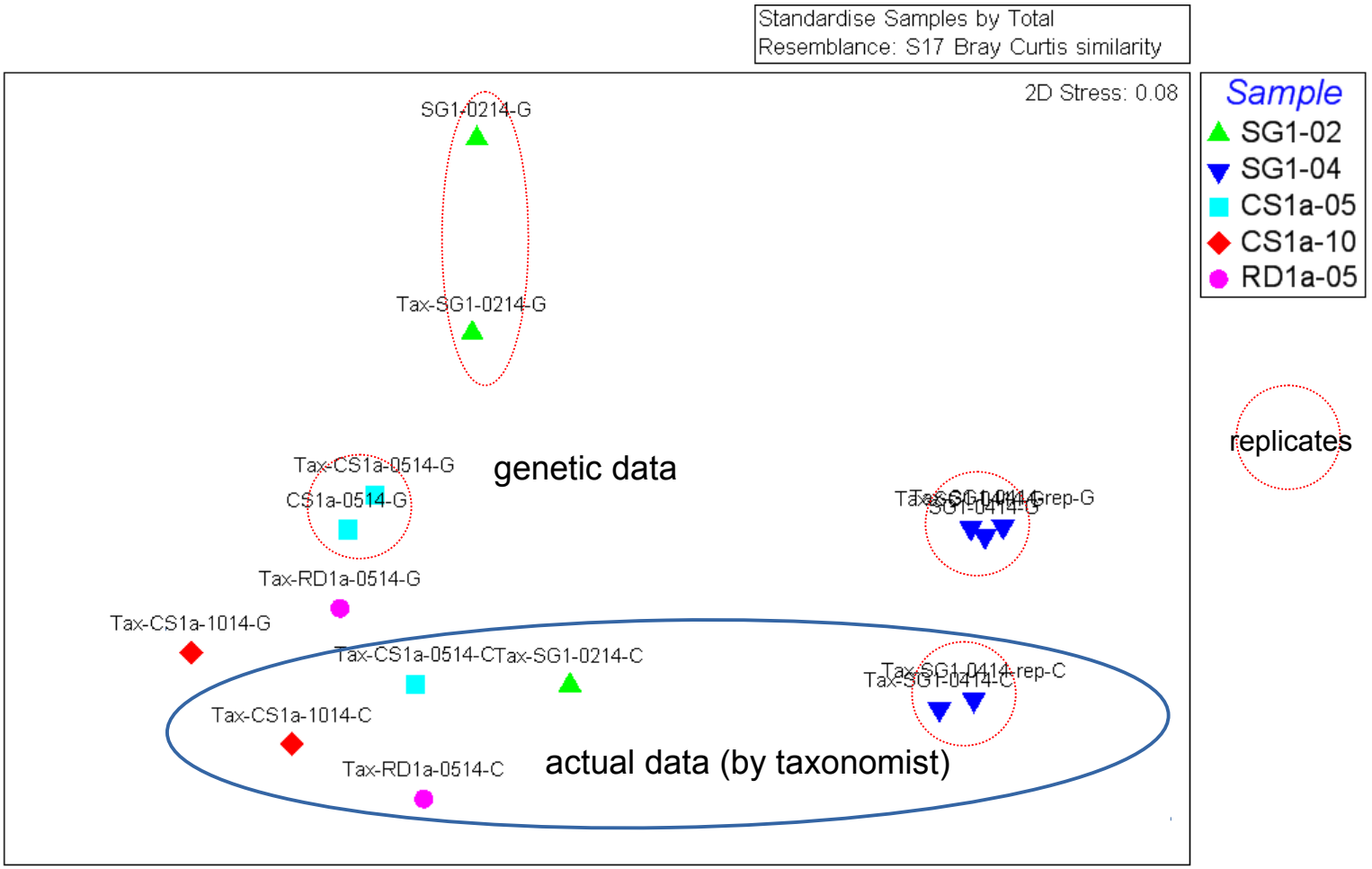


- Pseudosamples contained 6-16 species of 1-3 individuals each.
- All species detected by genetic analysis (even as low as 0.12% of the total sample biomass), except for the Corycaeidae [*Farranula rostrata* (max 6.5% biomass) and *Agetus typicus* (2.5% biomass)].
- For few species (e.g. *Calanus helgolandicus* and *Euterpina acutifrons*), large discrepancies between genetic and actual data. Systematic? Due to technical handling?→ more checks necessary.

Taxonomically identified samples: Comparison between actual counts (blue line) and genetic data (yellow line)



MDS for taxonomically identified samples: Comparison between genetic data (-G) and actual counts (-C)



The observed shift between the two data sets is mainly due to the inability to match exactly some taxa categories (e.g. Calocalanus juv. and Calocalanus sp. identified by the taxonomist to which of the 7 Calocalanus species identified by DNA metabarcoding correspond to ?)

Conclusions from standardization of DNA metabarcoding

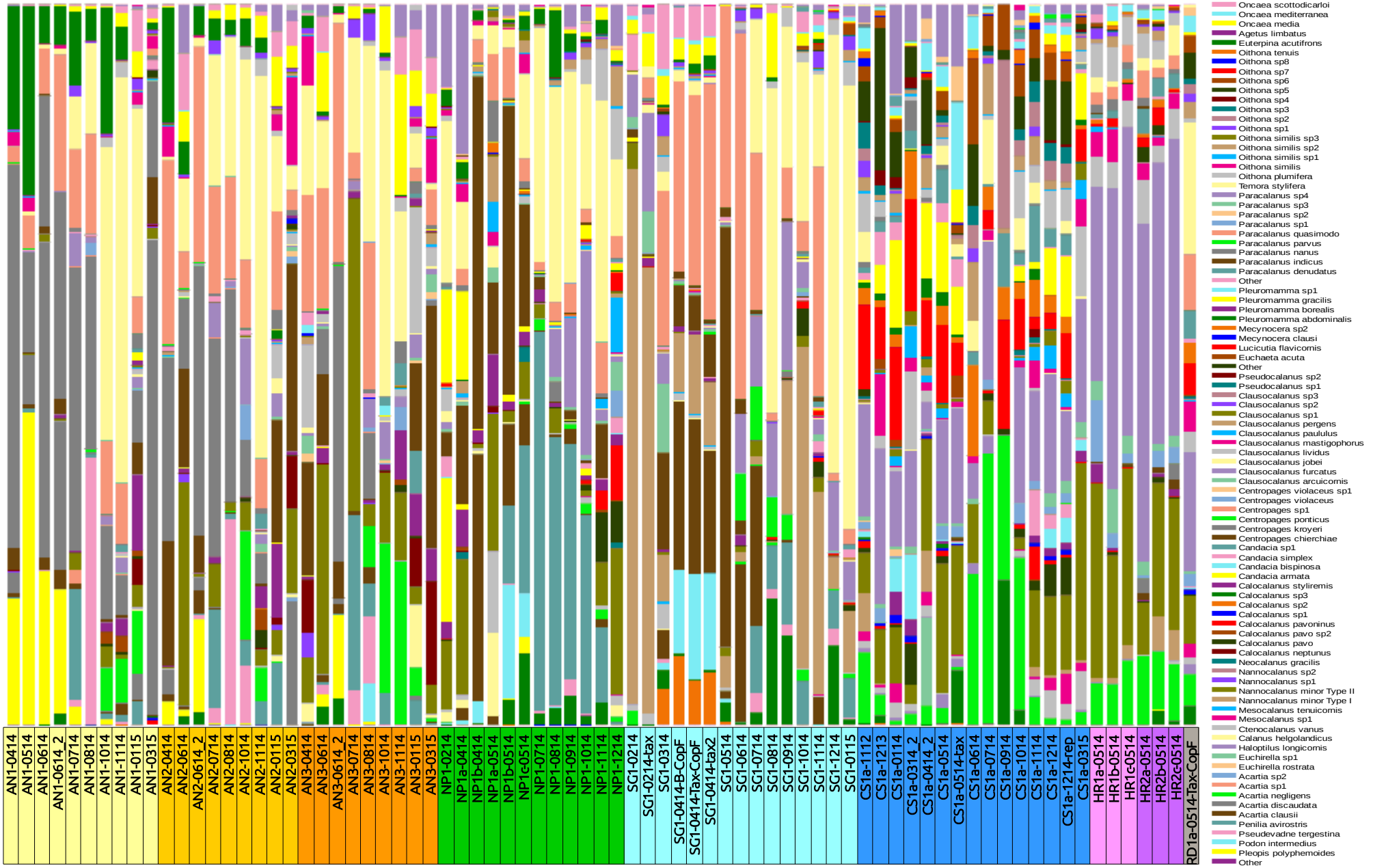
High repeatability (sub-samples of the same sample give very similar results)

Generally good concordance between morphological identification and NGS analysis

- but, *Corycaeidae* almost absent in DNA metabarcoding analysis (both in pseudo and taxonomically identified samples), although some species present in the reference DB and are individually amplified in PCR with the 16S primers.
- redesign primers? problems in DNA extraction?

Some taxa are systematically over or under-represented in DNA metabarcoding analysis → image analysis can greatly assist reducing the bias

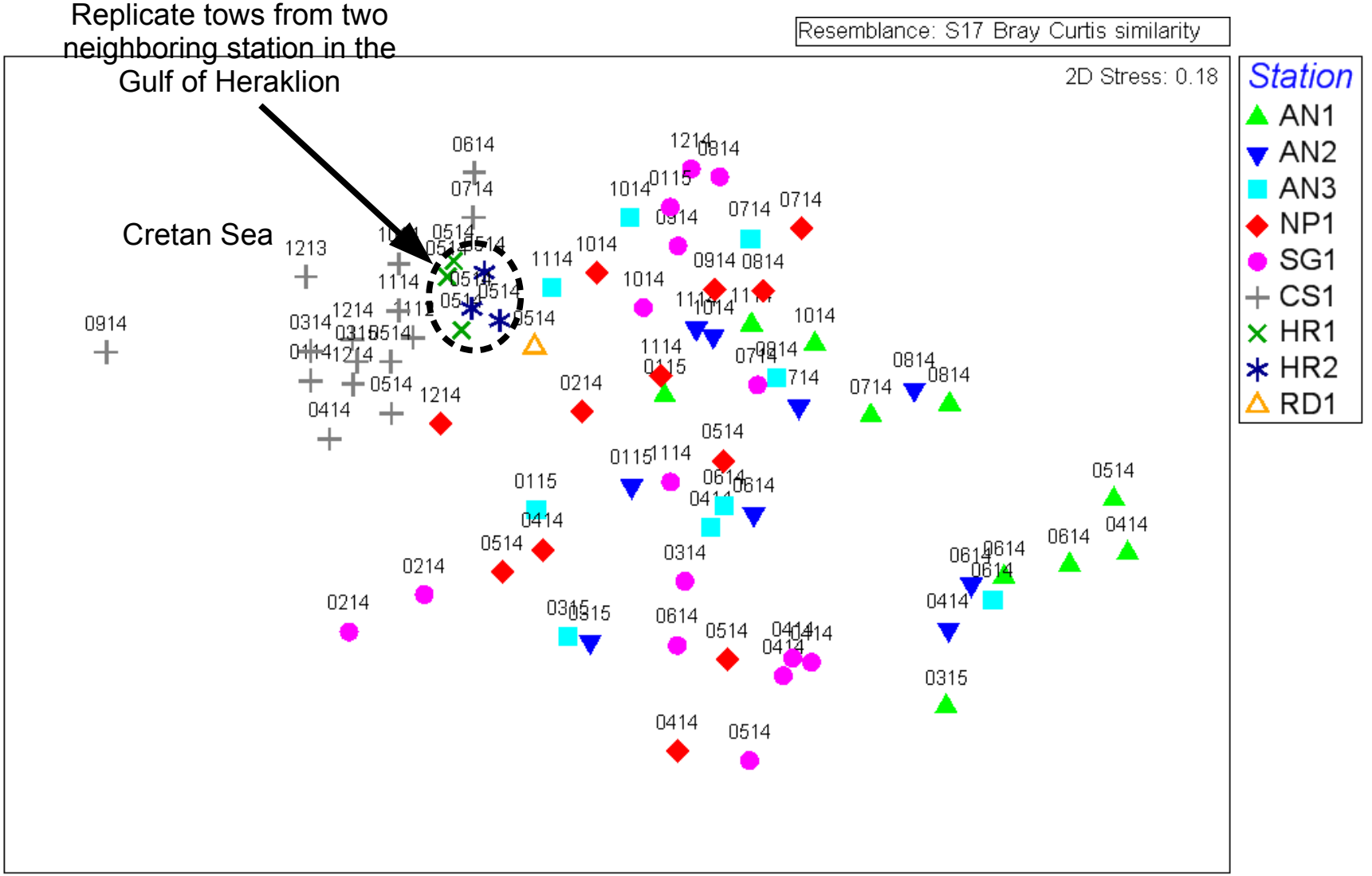
Relative abundance (species level) of copepoda & cladocera in monitoring stations of the Mediterranean (genetic data)



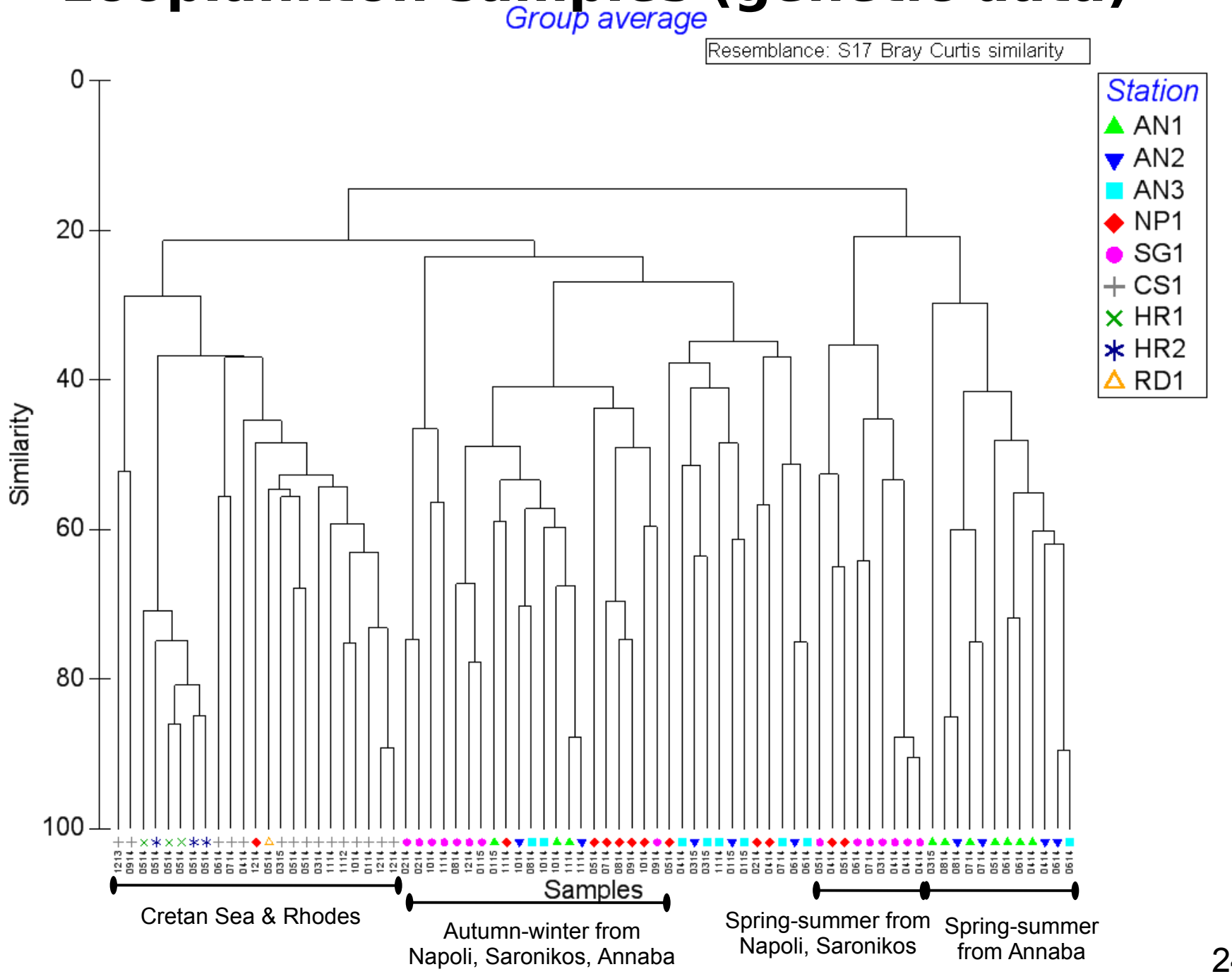
MDS plot based on genetic data for all samples

Replicate tows from two neighboring station in the Gulf of Heraklion

Cretan Sea



Similarity dendrogram (Bray Curtis) of zooplankton samples (genetic data)



Integration between image analysis and DNA metabarcoding analysis (under way)

A script is developed to combine automatically the output of image analysis and genetic analysis.

IA categories	IA %	DNA metabarcoding categories	Genetic data %	Final combined (%)
Calanus helgolandicus	20	Calanus helgolandicus	40	20
		Subtotal	**40**	**20**
Acartiidae	35	Acartia clausii	4	14
		Acartia negligens	6	21
		Subtotal	**10**	**35**
Small Calanoida	30	Clausocalanus arcuicornis	10	15
		Clausocalanus furcatus	2	3
		Clausocalanus jobei	5	7.5
		Ctenocalanus vanus	2	3
		Mecynocera clausi	1	1.5
		Subtotal	**20**	**30**
Oithona	15	Oithona similis	8	4
		Oithona nana	12	6
		Oithona sp.	10	5
		Subtotal	**30**	**15**
Total	**100**		**100**	**100**

→ Transformed to abundances and biomass.

What next

- Results are currently reanalyzed and the methodology is optimized for higher integration between image analysis and NGS analysis

- The 16S primers designed perform quite well and have high resolution (also amplify quite well the other components of zooplankton but not evaluated yet for taxa other than copepods and cladocera). Redisign to amplify Corycaeidae.

- Genetic reference DB needs further improvement (few important species still missing, taxonomic issues for certain taxa)

- Image analysis software should be “trained” for the oligotrophic regions of Eastern Mediterranean and other regions (e.g Algeria)

Partners of the "MetaCopepod" project

Genetics

Panagiotis Kasapidis

Jon Kristoffersen

Gianpaolo Zampicinini

Valentina Tsartsianidou

Vasso Terzoglou

Biodiversity informatics

Sarah Faulwetter

Bioinformatics

Jacques Lagnel

Tereza Manousaki

Image analysis

Costas Frangoulis

Stratos Batziakas

Taxonomy

Ioanna Siokou

Nontas Christou

external collaborators

Maria Grazia Mazzocchi (SZN, Naples)

Maria-Luz Fernandez de Puelles (IEO, Spain)

Meriem Khelifi Touhami (Annaba Univ., Algeria)

Yesim Ak Örek (METU, Turkey)

Rana Abu Alhaija (The Cyprus Institute, Cyprus)

Alexandra Gubanova (IBSS, Ukraine)

Acknowledgments

Many thanks to people who contributed with samples:

- Prof. Alenka Malej, National Institute of Biology, Ljubljana, Slovenia
- Dr. Soultana Zervoudaki, HCMR, Greece
- Dr. Maria Corsini-Foka, HCMR, Rhodes, Greece

The project's webpage: http://metacopepod.hcmr.gr/

For more info you may contact Dr. P. Kasapidis <kasapidi@hcmr.gr>

Thank you for your attention!