

Combination of classic biodiversity analyses with novel algorithms which extract environmental information from microbial sequences

Christina Pavlou^{1,2,3*}, Anastasios Oulas¹, Katerina Vasileiadou¹, Christos Arvanitidis¹

¹ Institute of Marine Biology, Biotechnology and Aquaculture, Hellenic Centre for Marine Research, 71500 Gournes, Heraklion, Crete, Greece

² Department of Biology, Faculty of Sciences, University of Ghent, 9000 Ghent, Belgium

³ Department of Microbial Ecophysiology, Faculty of Biology, University of Bremen, 28359, Bremen, Germany

* Corresponding author: cpavlou@hcmr.gr

Introduction

Lagoons are enriched habitats, with unstable environmental conditions caused by their constraint from the sea and their shallowness. The frequent fluctuations of the abiotic parameters cause severe changes in the abundance and distribution of organisms; while this relationship has been studied for macrofaunal organisms, only little is known about the lagoonal microbial diversity.

The continuous drop in the associated costs combined with the increased efficiency of the latest high-throughput sequencing technologies has resulted in an unprecedented growth in sequencing projects.

However, despite the vast availability of sequence data for microorganisms, little is known on the ecological features and processes prevailing on the ecosystems they have been found, especially since most of them can only be characterized as "uncultured".

The aim of the present study was to combine the novel pipeline, **SEQenv**, used to annotate sequences with environment descriptive terms occurring in the relevant literature, with classic biodiversity analyses.

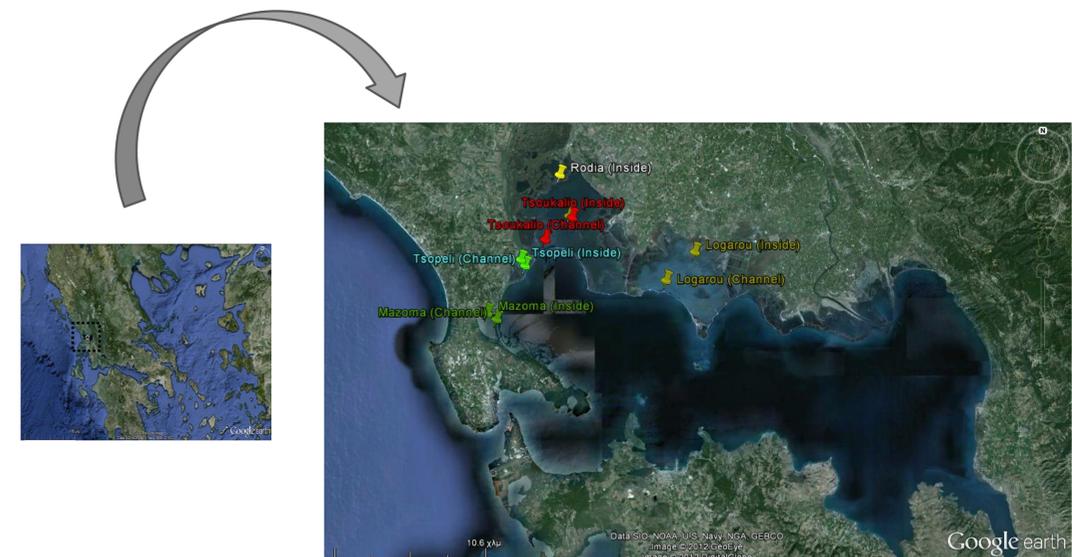


Figure 1: Map of the sampling stations in Amvrakikos Gulf.

Conclusions

- There is a clear spatial diversity pattern of the bacteria found in the lagoonal complex of the Amvrakikos Gulf.
- Each lagoonal community is annotated with different environmental descriptive terms, at least as their abundance is concerned.
- This may indicate that the communities have, to some extent, different sources of origin.

Materials and Methods

Collection of sediment samples



- DNA extraction
- PCR amplification (V5-V6 region of the 16S rRNA gene)
- Next generation sequencing

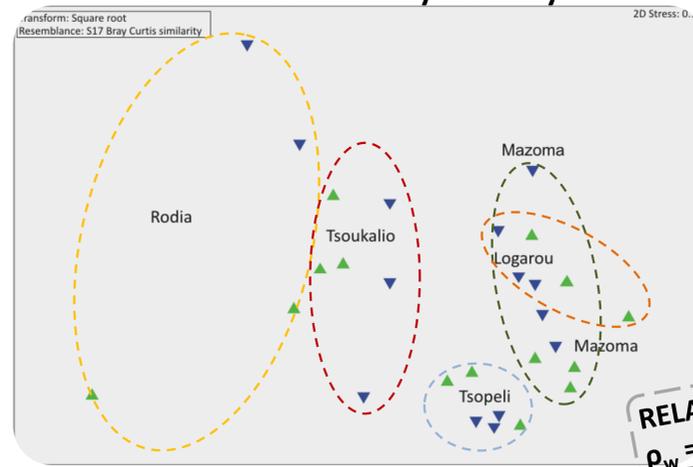


454 GS FLX Titanium Series

Extraction of abundant OTUs

ampliconnoise
Noise removal from pyrosequenced amplicons

Bacterial community diversity



▲ : stations located inside the lagoons
▼ : stations located in the channel connecting each lagoon to the Gulf
ANOSIM (Factor: Lagoon)
R = 0.796; p = 0.001

ENVIRONMENTS project



SEQenv pipeline



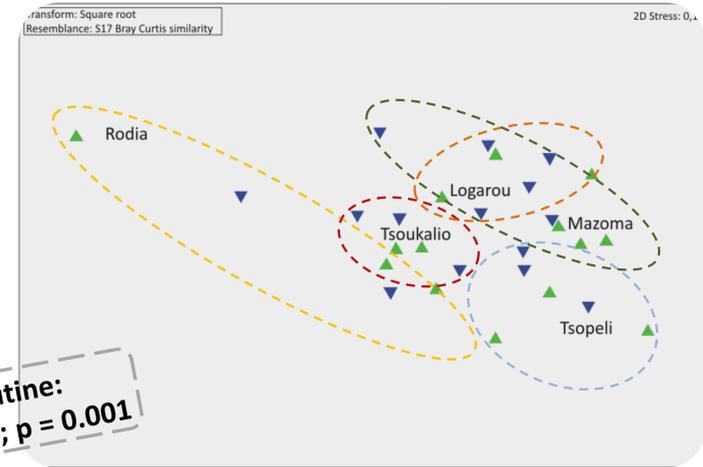
Environment Ontology



Extraction of Isolation Sources



Environmental descriptive terms



▲ : stations located inside the lagoons
▼ : stations located in the channel connecting each lagoon to the Gulf
ANOSIM (Factor: Lagoon)
R = 0.595; p = 0.001

RELATE routine:
 $p_w = 0.683$; $p = 0.001$

Results

As expected the most abundant environmental terms that characterize the samples are *sediment* (ENVO: 2007) and *marine sediment* (ENVO: 2113). Among the environmental descriptive terms found in all the samples were *sea* (ENVO: 16), *lake* (ENVO: 20), *mangrove* (ENVO: 1000181) and *soil* (ENVO: 1998).

However, only samples from Rodia and Tsoukalio lagoons were associated with *fresh water* (ENVO: 2011) and only in samples from Rodia was the term *lake sediment* (ENVO: 546) found.

Interestingly, the term *beef* (ENVO: 3067) was found in Tsopeli and Tsoukalio lagoons which may suggest that they are influenced from the adjacent meat producing factories.

