



# Marine Microbial Biodiversity, Bioinformatics & Biotechnology



Grant agreement n°287589

Acronym : Micro B3

Start date of project: 01/01/2012, funded for 48 month

## Deliverable 6.2

# A viral/giral taxonomic annotation for two metagenomic data sets

Version: 1 (12 Dec. 2013)

Circulated to: Frank-Oliver Glöckner (12 Dec. 13)

Approved by: Pascal Hingamp, Hiroyuki Ogata, Chris Bowler (12 Dec. 13)

Expected Submission Date: 31.12.2013

Actual submission Date: 12.12.13

Lead Party for Deliverable: Partner 6: CNRS-IGS, Marseille

Mail: [emilie.villar@igs.cnrs-mrs.fr](mailto:emilie.villar@igs.cnrs-mrs.fr)

Tel.: +33 4 91 82 54 36

### Dissemination level:

Public (PU)	X
Restricted to other programme participants (including the Commission Services) (PP)	
Restricted to a group specified by the consortium (including the Commission Services) (RE)	
Confidential, only for members of the consortium (including the Commission Services) (CO)	



The Micro B3 project is funded from the European Union's Seventh Framework Programme ( Joint Call OCEAN.2011-2: Marine microbial diversity – new insights into marine ecosystems functioning and its biotechnological potential) under the grant agreement no 287589. The Micro B3 project is solely responsible for this publication. It does not represent the opinion of the EU. The EU is not responsible for any use that might be made of data appearing herein.



## **Summary**

In the MicroB3-WP5 (Deliverable 5.8), we provided an annotation pipeline: i/ To identify viral sequences in metagenomic reads or contigs, and ii/ To perform phylogenetic mapping on NCLDV marker genes. These two annotation strategies have been used for two ecological studies presented here:

1. Nucleo-cytoplasmic large DNA viruses in Tara-Oceans pyrosequence data

This study aimed to characterize the diversity, abundance and biogeography of marine NCLDVs in 17 metagenomes derived from microbial samples (0.2–1.6  $\mu\text{m}$  size range) collected during the Tara Oceans expedition.

2. ASPIC project - NCLDV transport through the Agulhas leakage off South Africa

The pipelines have been used to understand the role of Agulhas rings in viruses and particularly NCLDVs transport from the Indian Ocean to the South Atlantic Ocean.

## Dataset 1: Nucleo-cytoplasmic large DNA viruses in Tara-Oceans pyrosequence data

**Context:** The Tara-Oceans metagenomic samples were initially sequenced by GENOSCOPE (partner 17) using the Roche 454 GS-FLX Titanium pyrosequencing technology. The first available Tara Oceans pyrosequences (TOP) dataset consisting of 17 microbial metagenomes - corresponding to the size fraction 0.2 to 1.6  $\mu\text{m}$  - was used by the IGS laboratory (CNRS partner 6) to characterize diversity of Nucleo-cytoplasmic large DNA viruses (NCLDVs) in the different marine basins explored during the first year of the Tara Oceans cruise: Atlantic Ocean, Mediterranean Sea, Red Sea, Arabian Sea and Indian Ocean. The results were submitted as an open access article to ISME Journal and were published in April 2013 (reference below).

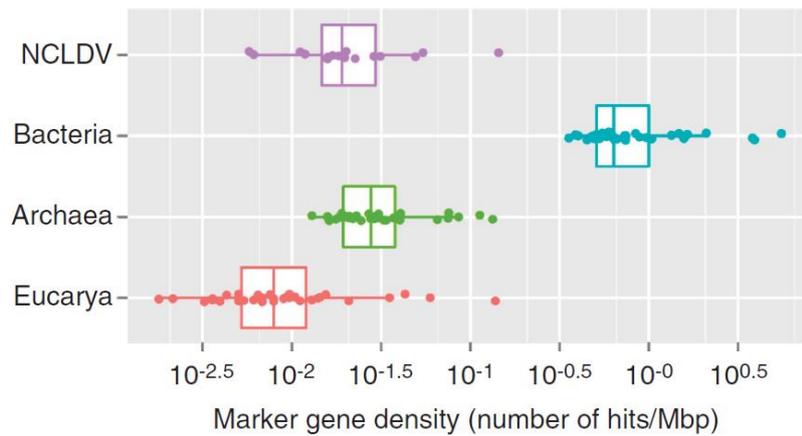
### Summary:

Nucleo-cytoplasmic large DNA viruses constitute a group of eukaryotic viruses that can have crucial ecological roles in marine environments by accelerating the turnover of their unicellular hosts or by causing diseases in animals. To better characterize the diversity, abundance and biogeography of marine NCLDVs, we analyzed 17 metagenomes derived from microbial samples (0.2–1.6  $\mu\text{m}$  size range) collected during the Tara Oceans expedition. The sample set includes ecosystems under-represented in previous studies, such as the Arabian Sea oxygen minimum zone (OMZ) and Indian Ocean lagoons.

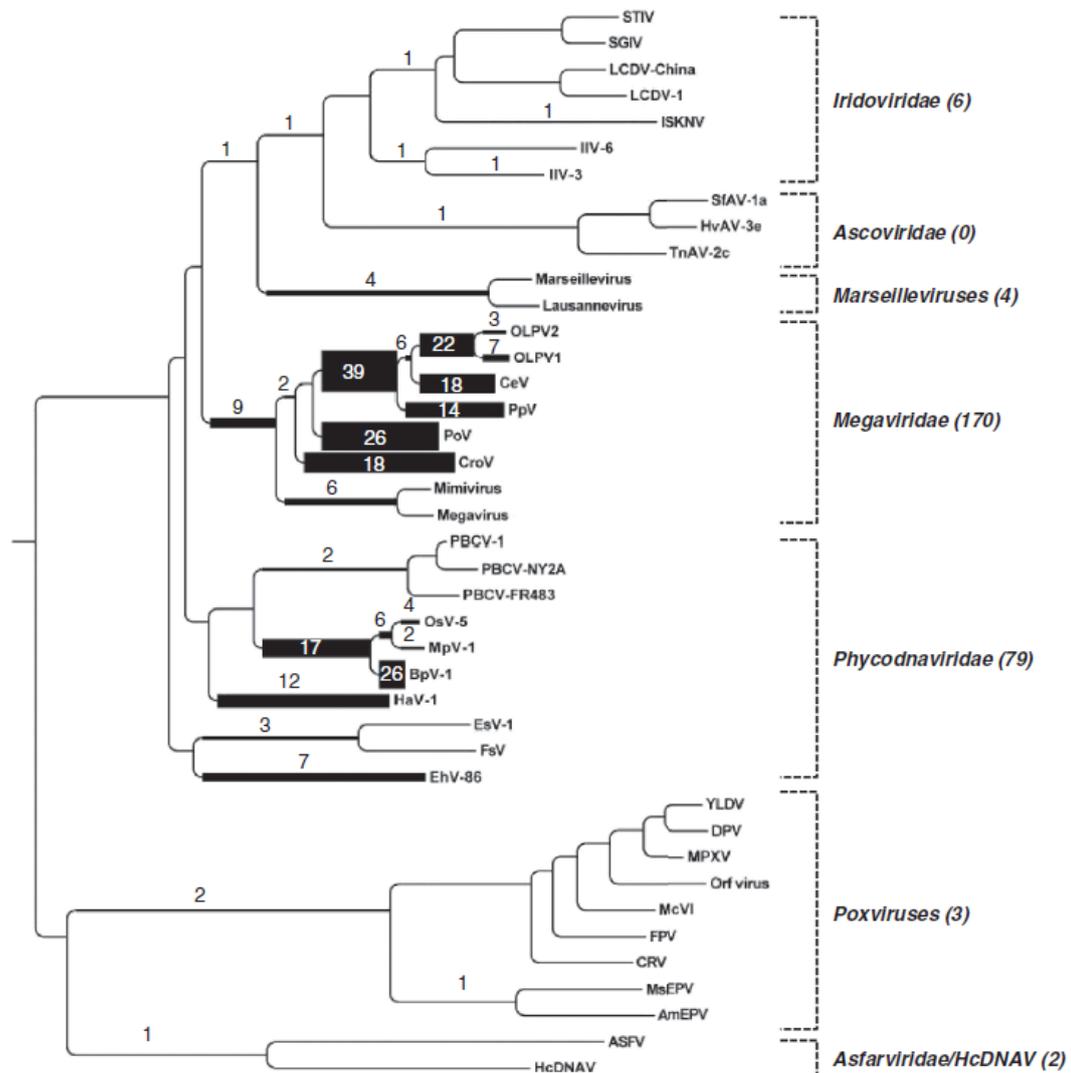
By combining computationally-derived relative abundance (Fig. 1) and direct prokaryote cell counts, the abundance of NCLDVs was found to be in the order of  $10^4$ – $10^5$  genomes  $\text{ml}^{-1}$  for the samples from the photic zone and  $10^2$ – $10^3$  genomes  $\text{ml}^{-1}$  for the OMZ. The Megaviridae and Phycodnaviridae dominated the NCLDV populations in the metagenomes (Fig. 2), although most of the reads classified in these families showed large divergence from known viral genomes.

Our taxon co-occurrence analysis revealed a potential association between viruses of the Megaviridae family and eukaryotes related to oomycetes. In support of this predicted association, we identified six cases of lateral gene transfer between Megaviridae and oomycetes. Our results suggest that marine NCLDVs probably outnumber eukaryotic organisms in the photic layer (per given water mass) and that metagenomic sequence analyses promise to shed new light on the biodiversity of marine viruses and their interactions with potential hosts.

**Publication:** Hingamp, P., Grimsley, N., Acinas, S.G., Clerissi, C., Subirana, L., Poulain, J., Ferrera, I., Sarmiento, H., Villar, E., Lima-Mendez, G., et al. (2013). Exploring nucleo-cytoplasmic large DNA viruses in Tara Oceans microbial metagenomes. ISME J 7, 1678–1695. doi:10.1038/ismej.2013.59



**Figure 1:** Metagenome-based relative abundance of NCLDV and cellular genomes in the TOP data set.



**Figure 2:** Phylogenetic positions of metagenomic reads closely related to NCLDV DNA polymerase sequences.



## **Dataset 2: ASPIC project - NCLDV transport through Agulhas leakage off South Africa**

**Context:** The ASPIC project aims at understanding the role of Agulhas rings in plankton transport from the Indian Ocean to the South Atlantic Ocean. The underlying dataset is provided by the second year of the Tara Oceans expedition, during which 9 stations were sampled from the three relevant Indian, Atlantic and Southern Oceans, together with two sampling stations specifically targeting Agulhas rings. This integrative interdisciplinary study is a collaboration between IGS, SBR and ENS (CNRS, partner 6) and Stazione Zoologica Anton Dohrn (partner 8). The manuscript currently in preparation, coordinated by Emilie Villar (IGS), is planned to be submitted together with a limited set of other Tara Oceans manuscripts in a high profile journal in 2014.

### **Manuscript provisional abstract:**

As the main choke point of the global ocean circulation (Bjostoch et al., 2008), the Agulhas leakage is intensively studied by oceanographers. Despite the fact that Agulhas leakage and the invisible plankton life are two significant climate regulators (Beal et al., 2011; Falkowski, 2011), precious little is known about plankton fate when transported from Indian to Atlantic oceans. The major part of the Agulhas leakage occurs as anticyclonic rings, which are formed at the Agulhas retroflection and move across the South Atlantic gyre for up to several years (Lutjeharms and Gordon, 1987). Agulhas rings are very often associated with an initial severe cooling and deeper mixed layers than the surrounding subtropical waters (Arhan et al., 2011). Nonetheless, according to Cermeno and Falkowski (2009) the Agulhas choke point is not a barrier to plankton dispersal. We tested this hypothesis during the Tara Oceans expedition across the Indian, South Atlantic and Southern Oceans. We also specifically sampled two Agulhas rings (9 months and 3 years-old, respectively) that were identified by satellite altimetry. Combined with the Tara Oceans holistic analytical approach (remote sensing, physics, chemistry, modeling, imagery, and genetics; Karsenti *et al.*, 2011), this station collection provides an unprecedented dataset to explore the biological connections amongst oceanic basins.

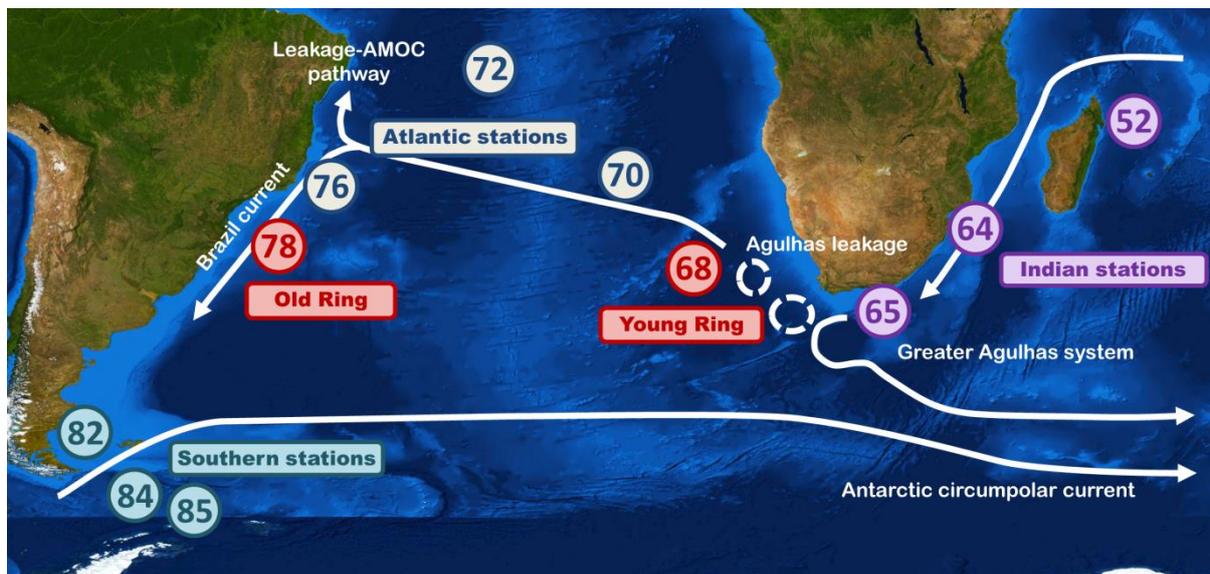
We found water in the young ring to be 5°C lower in temperature than Indian Ocean source waters, with a particularly deep mixed layer of 250m depth. A nitrogen cycle disturbance, predicted by Agulhas ring modeling, was confirmed by elevated nitrite concentration in the young ring as well as related functional and taxonomic signatures. The shift in biodiversity observed in the young ring appeared to oppose no significant barrier to plankton flow, since Atlantic and Indian Ocean plankton communities were essentially indistinguishable in our study. This biodiversity shift was likely transient, since the old ring looked like a typical Atlantic plankton community. In contrast and as expected, the Southern Ocean plankton community appeared isolated in our results. We also demonstrated that taxonomic groups were not equally affected by environmental changes occurring during the ring formation,

e.g. smaller organisms (< 5  $\mu\text{m}$ ) seemed to be more resilient than larger plankton (>5 $\mu\text{m}$ ).

Given that Indian and Atlantic Ocean plankton communities are mostly similar even at the finest grained genetic tracers level, we confirm that Agulhas leakage is efficient in plankton transport across the choke point. However, as the young ring is shown here to be physically and biologically distinct from both source and destination oceans, this transport must occur through successions of heterogeneous rings, outside of rings, or as a rare biosphere flow inside the rings.

### Contextual data:

We studied 11 stations from the TARA-Oceans cruise (ASPIC dataset, Fig. 3), with 9 stations representing the 3 oceans studied (Indian Ocean, IO: stations 52, 64, 65; Atlantic Ocean, AO: stations 70, 72, 76; and Southern Ocean: stations 82, 84, 85), and 2 stations representing rings (stations 68 & 78). The station 68 represented a “young” 6 months-old ring of around 110 km in diameter, and the station 78 the tail of an “old” 36 months-old ring. These two stations were characterized by a deep pycnocline. After six months (station 68), the ring was characterized by a heat loss of  $-5^{\circ}\text{C}$  compared to Indian stations 64 and 65, and a particularly deep mixed layer of 250m depth, (vs. 80m depth in the stations 64 and 65).



**Figure 3:** Agulhas leakage and location of Tara-Oceans stations.

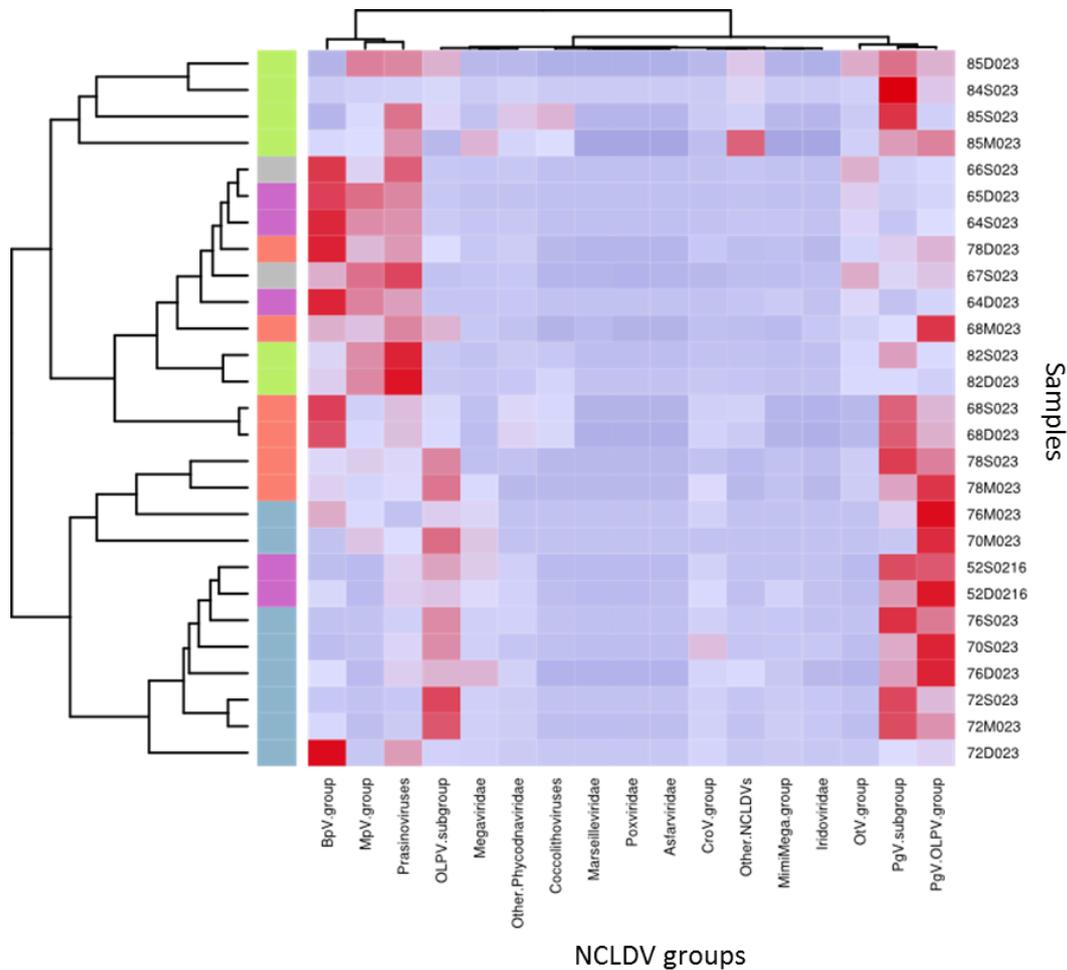
### Main results:

Here are presented the results concerning viruses and NCLDVs obtained using the pipelines developed in the framework of the deliverable 5.8.

*NCLDV distribution across the choke point:*

The Nucleocytoplasmic large DNA viruses (NCLDV) diversity seemed to cluster ASPIC samples according to the station latitudes (Fig. 4):

- The most tropical stations (52, 70, 72, 76 and 78) were characterized by high proportions of Phaeocystis globosa virus (PgV) and Organic Lake phycodnaviruses (OLPV).
- Choke point stations (64, 65, 66, 67 and 68), were characterized by high proportions of Bathycoccus prasinus virus (BpV), Micromonas pusilla virus (MpV) and Prasinoviruses.
- Southern stations are characterized by a high proportion of PgV subgroup, but few OLPV.



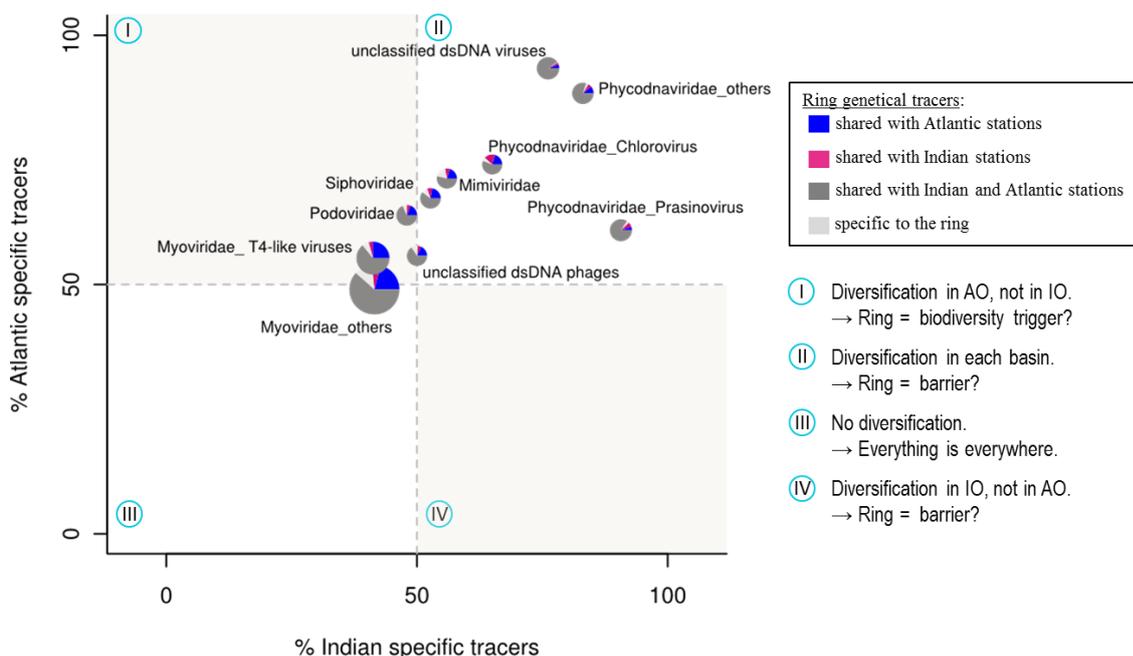
**Figure 4:** Heatmap of NCLDV group abundances.

Rows represent ASPIC samples, columns represent NCLDV taxonomic groups. Cells are warm (red) when abundant in the samples, and cold (blue) when rare. The two dendrograms are obtained using hierarchical clustering (complete linkage) on Spearman distance matrices of samples (horizontally) and on ward distance matrices of groups (vertically). Sample codes contain the station number (52 to 85), the depth (S: Surface, D: Deep Chlorophyll Maximum, M: Mesopelagic) and the fraction size (023: 0.2-3 $\mu$ m; 0216: 0.2-1.6 $\mu$ m) sampled. Dendrogram leaves are colored in blue for Atlantic stations, purple for Indian stations, red for ring stations and green for Southern ocean stations.

As observed for eukaryotic barcodes from larger size fractions (from 5 to 2000  $\mu\text{m}$ ), samples were more discriminated by geographical position than by the sampled depth for surface and DCM samples. In contrast to previous studies where mesopelagic samplings showed a specific deep plankton community (Decelle *et al.*, 2013), the ASPIC dataset suggests that NCLDV communities were relatively homogeneous in the water column. Is it the case for their hosts?

### Ring transport of viruses from the Indian to Atlantic Oceans

The majority of the viral families displayed high levels of diversification in each oceanic basin (sector II in Fig. 5), suggesting that the ring capture may have been acting as a bottleneck for viral diversity diffusion. Moreover, the corresponding genetic tracers found in the young ring were mostly either shared with the two connected oceans, or shared only with the Atlantic Ocean, suggesting that the Indian Ocean viral diversity reduction was either occurring in the 9-month old ring (e.g. ds DNA viruses, Phycodnaviridae), or had just occurred (Myoviridae).



**Figure 5:** Ring viruses origin and fate through the Agulhas choke point.

Viral families are plotted on the chart with coordinates proportional to genetic tracers specific to Indian (x-axis, IO) and Atlantic Oceans (y-axis, AO). Each quarter sector corresponds to different scenarios I-IV described in the right legend. The circle size is proportional to the number of unique co-mapped reads considered. The pie charts describe similarity of the genetic tracers found in the young ring with the two surrounding oceanic basins.

## References

- Arhan, M., Speich, S., Messenger, C., Dencausse, G., Fine, R., and Boye, M. (2011). Anticyclonic and cyclonic eddies of subtropical origin in the subantarctic zone south of Africa. *Journal of Geophysical Research (Oceans)* 116, 11004.
- Barton, A.D., Dutkiewicz, S., Flierl, G., Bragg, J., and Follows, M.J. (2010). Patterns of Diversity in Marine Phytoplankton. *Science* 327, 1509–1511.
- Beal, L.M., Ruijter, W.P.M.D., Biastoch, A., Zahn, R., and 136, S.W.G. (2011). On the role of the Agulhas system in ocean circulation and climate. *Nature* 472, 429–436.
- Biastoch, A., Böning, C.W., and Lutjeharms, J.R.E. (2008). Agulhas leakage dynamics affects decadal variability in Atlantic overturning circulation. *Nature* 456, 489–492.
- Cermeño, P., and Falkowski, P.G. (2009). Controls on Diatom Biogeography in the Ocean. *Science* 325, 1539–1541.
- Decelle, J., Martin, P., Paborstava, K., Pond, D.W., Tarling, G., Mahé, F., de Vargas, C., Lampitt, R., and Not, F. (2013). Diversity, Ecology and Biogeochemistry of Cyst-Forming Acantharia (Radiolaria) in the Oceans. *PLoS ONE* 8, e53598.
- Falkowski, P. (2012). Ocean Science: The power of plankton. *Nature* 483, S17–S20.
- Karsenti, E., Acinas, S.G., Bork, P., Bowler, C., De Vargas, C., Raes, J., Sullivan, M., Arendt, D., Benzoni, F., Claverie, J.-M., et al. (2011). A Holistic Approach to Marine Eco-Systems Biology. *PLoS Biol* 9, e1001177.
- Lutjeharms, J.R.E., and Gordon, A.L. (1987). Shedding of an Agulhas ring observed at sea. *Nature* 325, 138–140.