



Figure S1. Output of the multivariate analyses: a) Canonical Variates Analysis (CVA) b) Canonical Correspondence Analysis (CCA)

a) Canonical Variates Analysis (CVA) also referred as Linear Discrimination Analysis (LDA) was performed using the function `lda` found in the R library MASS [48]. LDA is used for the analysis of group structure in multivariate data. Canonical variate axes are directions in multivariate space that maximally discriminate the pre-defined groups of interest specified in the data.

CVA has been used here to study the ability of the Ward's-Manhattan clustering (based on sample depth and salinity only) in predefining groups to discriminate among the four systems. In particular 14 parameters were considered: chemical-physical parameters (Depth, Temperature, Salinity, density, Oxygen - in ml/L and in % - Fluorescence), four enzymes (AP, LAP, a-Glu, b-Glu), PPT, Secondary Production and Total Phytoplankton Abundance.

Circles represent confidence regions ($p > 95\%$) in the CVA space where we expect to find respectively the group means (continuous line) and sample population belonging to a given group (dashed line).

b) The Canonical Correspondence Analysis (CCA) was applied to the dataset using the PAST (version 3.11) application [53]. The outputs of this multivariate analysis can be visualised through a graph ("biplot") that allows depicting in a reduced dimensional space the relationships between microbial variables (microbial activities, prokaryotic abundances) and linear constraints (such as depth, temperature, salinity, fluorescence, oxygen, and nutrients). A CCA ordination diagram consists of ordination axes, points for qualitative variables and arrows for quantitative variables.

The implementation in PAST follows the eigenanalysis algorithm given in [52]

CVA confirms that CS is well distinct with respect to IS and OS, whose populations conversely exhibit a certain degree of overlap. DS is completely separated.