

Figure S1. Canonical correspondence analysis for the relationship between BGCs distribution in the genomes and the genus to which the genome belongs to (shown as the different shapes) and the ecosystem from which it was binned (shown as the different colors). The model of BGC distribution using the genus and ecosystem explained 87.11% of the variance and was significant ($P=0.001$). Both factors (genus and ecosystem) were also significant predictors of the BGC distribution ($P=0.001$). Analysis was carried out in R using the function `cca` in the `Vegan` package.

