



Vivaldi Project

Data management plan

18S sequencing in environmental samples

Water

Plankton

Sediment

Amplicon sequencing

29/05/2019

DATA MANAGEMENT PLAN

Template sheet for each dataset

Partner name

CSIC

Data category	<i>Microbiome (bivalve)</i>
Concerned WP	<i>WP1</i>
Name of the VIVALDI referent(s)	<i>Microbiome: Alberto Pallavicini</i>
Reference of the dataset <i>Please refer to the DMP table to find the appropriate reference.</i> <i>Ex: Genome-Patho/SubTaskN*/Pathogen/PartnerN*</i>	<i>Microbiome18S-environment/1.1/water_plankton_sediment/CSIC</i>
Description of the data	<i>Raw data without trimming in fastq format Amplicon sequencing of 18S</i>
Type	<i>Sequences</i>
Period and frequency of data collection	<i>Water filtered at > 200 µm (plankton), > 65 µm, > 0,22 µm and sediment were sampled in september 2016, october 2016, february 2017, June 2017, september 2017, february 2018, and june 2018 Samples were conveniently stored until DNA extraction.</i>
Geographical site of data collection (if applicable)	<i>The samples were obtained in the Meira beach, in the Ría de Vigo.</i>
Description of the material from which the dataset is generated <i>Information will be obtained from individuals, which can come from natural/hatchery population and/or from family produced in hatchery. Animals can be infected (naturally or experimentally). DNA extraction can be done from the whole animal, tissue.</i>	<i>Sequencing information has been obtained from DNA extracted from the environmental samples. DNA was extracted following the DNeasy PowerSoil DNA Kit QIAGEN protocol indications. Each sample belongs to water filtered at > 200 µm (plankton), > 65 µm, > 0,22 µm and sediment sampled in september 2016, october 2016, february 2017, June 2017, september 2017, february 2018, and june 2018.</i>
Protocols <i>Example: 16S ribosomal RNA gene sequencing by NGS Please refer to the DMP table* for more examples</i>	<i>Quality control of samples and preparation of the libraries was performed in Macrogen Korea amplicon sequencing by Illumina paired-end, Herculanase II fusion DNA polymerase Nextera XT index kit v2, Miseq 2x300bp</i>
Nature of the collected/generated data <i>Example: Raw dataset in .blc/.fastqc/.fasta formats for genomic information, and processed data set will be .vcf/.bed formats. Please refer to the DMP table* for more examples</i>	<i>Raw data without trimming in fastq format Amplicon sequencing of 18S</i>
Coverage (if applicable) <i>Example: random genomic regions covered at 50 X Please refer to the DMP table* for</i>	<i>NA</i>

more examples	
What are the prerequisites allowing to use the data as such? <i>Example: Any person able to use .fastqc file and .fasta file</i> Please refer to the DMP table* for more examples	<i>Any person who can deal with fasta quality files</i>
Sharing of main data	<i>Saved and shared after publication</i> <i>Accession number to download the raw reads will be available at NCBI SRA after publication</i>
Archiving and preservation <i>Example: data will be stored on a hard drive + online back up and then will be released on public database (Sinoe, Dryad) after publication.</i> Please refer to the DMP table* for more examples	<i>Data will be stored in our hard disks and in NCBI SRA repository.</i> <i>CSIC has another copy in an internal server.</i> <i>Data will be of public accession after publication</i>
List, description and storage of associated data (metadata) <i>Examples: environmental data, mortality monitoring, genotyping...</i>	<i>NA</i>
Sharing of metadata (if relevant)	<i>No relevant</i> <i>NA</i>

*To access the [DMP table](#), please login on the VIVALDI online platform

Once completed, this sheet has to:

1. Be sent to the referent(s) identified above for a final check
2. Be uploaded on the [VIVALDI online platform](#)