Vivaldi Project

Data management plan

Mussel hemocytes small RNA sequencing

Mytilus galloprovincialis
Hemocytes
miRNA
In vitro stimulation

31/07/2018

DATA MANAGEMENT PLAN

Template sheet for each dataset

Partner name	*CSIC*
Data category	*Transcriptome (host)*
Concerned WP	*WP2*
Name of the VIVALDI referent(s)	*Transcriptome: Beatriz Novoa & Paola Venier*
Reference of the dataset *Please refer to the DMP table to find the appropriate reference.* *Ex: Genome-Patho/SubTaskN°/Pathogen/PartnerN°*	*Transcriptome-host/2.4.2/mussel_smallRNA/CSIC*
Description of the data	*Raw data without trimming in fastq format* *Small RNA information* *Processed data will be a list of identified and novel smallRNAs*
Type	*Sequences*
Period and frequency of data collection	*Hemolymph from raft mussels was withdrawn, pooled, and stimulated with different molecules. Samples were collected 8h after the stimulation.*
Geographical site of data collection (if applicable)	*The mussels origin was the Ría de Vigo (raft mussels). The experiment was carried out in lab controlled conditions.*
Description of the material from which the dataset is generated *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tissue.*	*SmallRNA sequencing has been performed from hemocytes treated in vitro.* *Hemolymph from raft mussels were withdrawn and pooled, set in 6 well plates and stimulated. Samples were collected 8h after the stimulation RNA was extracted.* *Each sample (below, correspondence with the sampling table) belongs to a pool of 50 mussels and a specific stimulation (4 different stimulus + a control):* *CSIC-2.1.1_2017 CSIC-2.2.1_2017 CSIC-2.2.2_2017 CSIC-2.2.3_2017 CSIC-2.2.4_2017*
Protocols *Example: 16S ribosomal RNA gene sequencing by NGS* ***Please refer to the DMP table* for more examples***	*Quality control of samples and preparation of the libraries was performed in Macrogen Korea* *mRNA sequencing by Illumina single-end, TruSeq smallRNA Library prep kit, HiSeq 2500*
Nature of the collected/generated data	*Raw data without trimming in fastq format* *Small RNA information*

This project has received funding from the European Union's Horizon 2020 Research and innovation programme under grant agreement N° 678589

Example: Raw dataset in .blc/.fastqc/.fasta formats for genomic information, and processed datas set will be .vcf/.bed formats. ***Please refer to the DMP table* for more examples***	*Processed data will be a list of identified and novel smallRNAs*
Coverage (if applicable) *Example: random genomic regions covered at 50 X* ***Please refer to the DMP table* for more examples***	*An average of 27 million reads per sample*
What are the prerequisites allowing to use the data as such? *Example: Any person able to use .fastqc file and .fasta file* ***Please refer to the DMP table* for more examples***	*Any person who can deal with small RNA sequences*
Sharing of main data	*Saved and shared after publication* *Accession number to download the raw reads will be available at NCBI SRA after publication (BioProject PRJNA470756)*
Archiving and preservation *Example: data will be stored on a hard drive + online back up and then will be released on public database (Sinoe, Dryad) after publication.* ***Please refer to the DMP table* for more examples***	*Data will be stored in our hard disks and in NCBI SRA repository.* *It is foreseen to have another copy in an internal server of CSIC.* *Data will be of public accession after publication (BioProject PRJNA470756)*
List, description and storage of associated data (metadata) *Examples: environmental data, mortality monitoring, genotyping…*	*NA*
Sharing of metadata (if relevant)	*No relevant* *NA*

*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**

This project has received funding from the European Union's Horizon 2020 Research and innovation programme under grant agreement N° 678589