Vivaldi Project

Data management plan

Crassostrea gigas miRNA sequencing

Oyster miRNA stimulation immunity

28/01/2020

DATA MANAGEMENT PLAN

Template sheet for each dataset

Partner name	*CNRS*
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/Crassostrea gigas/2*
Description of the data	*NGS; miRNA sequencing by* Illumina HiSeq
Type	*Sequences*
Period and frequency of data collection	*Data generated January 2017* An *in vivo* stimulation of *Crassostrea gigas* oysters was done by injecting two immunostimulants, namely poly(I:C) and inactivated gram negative bacteria (LGP32). Hemolymph was collected 8h after injection and RNA extracted
Geographical site of data collection (if applicable)	*Montpellier, France (oysters originating from Ifremer, Brittany)*
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.* **Protocols** *Example: 16S ribosomal RNA gene sequencing by NGS* ***Please refer to the DMP table* for more examples***	*Libraries were built using miRNA extracted from LPS-induced or poly(I:C) induced oysters (whole tissues, 5 months-old hatchery animals, 10 oysters / condition)* *10 oysters were sampled and pooled, snap frozen and reduce to powder. Two RNA extraction were performed per conditions and sequenced* *RNA Libraries were prepared after size selection of whole RNA (<60bp)using the TruSeq cDNA kit (Illumina, USA) and sequencing performed in Illumina HiSeq SR 50 bp by the Genome Quebec platform*
Nature of the collected/generated data *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***	*Raw data set (fastq format)* *Processed data will be .vcf format*

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

Coverage (if applicable) *Example: random genomic regions covered at 50 X* ***Please refer to the DMP table* for more examples***	*average of 20 million reads were obtained 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 able to use .fastqc file and .fasta file*
Sharing of main data	*Saved and shared after publication*
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 by CNRS-IHPE and then will be released on public database (NCBI) after publication*
List, description and storage of associated data (metadata) *Examples: environmental data, mortality monitoring, genotyping…*	*Metadata files contain all available information related to the study (oysters, treatment, etc...).* All these metadata will be available in both the SRA bio-project (to be released after publication) and the associated publication (in preparation).
Sharing of metadata (if relevant)	*Saved and shared after publication*

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