Vivaldi Project

Data management plan

BS-Seq, RNA-Seq – DNA methylation and RNA-Seq analysis on whole body oyster samples

30/03/2020

DATA MANAGEMENT PLAN

Template sheet for each dataset

Partner name	*CNRS*
Data category	*Phenotypic markers (host)*
Concerned WP	*WP2* *WP2* *WP2*
Name of the VIVALDI referent(s)	*Phenotypic markers (host): Florian Enez & Christine Paillard*
Reference of the dataset *Please refer to the DMP table to find the appropriate reference.* *Ex: Genome-Patho/SubTaskN°/Pathogen/PartnerN°*	*Phenotypic-markers-host/2.1.4/C.Gigas/P2*
Description of the data	The occurrence of epigenetic changes in *C. gigas* resisting to mortalities was studied in 2 full-sib families obtained from the DECIPHER project. The DNA methylation pattern (DNA methylation), gene expression (RNAseq) and phenotypic changes (increased resistance) in 2 generations (F1 and F2) of oysters submitted (or not) to immune stimulus (e.g. rich microbial environment) during ontogenesis or before the challenge was assessed.
Type	*Sequences*
Period and frequency of data collection	*4 times: march 2017, July 2017, March 2018, July 2018*
Geographical site of data collection (if applicable)	*n/a*
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.*	*A biparental oyster family (Decipher #11 and #32)* *DNA was extracted from whole body oysters sampled during larval development and juveniles.*
Protocols *Example: 16S ribosomal RNA gene sequencing by NGS* ***Please refer to the DMP table* for more examples***	*DNA methylation data was generated from bisulfite converted gDNA sequenced by HiSeqX PE150bp (illumina)* *RNA data was generated by HiSeq4000 PE100bp (illumina)fastq*

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

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***	*fastq*
Coverage (if applicable) *Example: random genomic regions covered at 50 X* ***Please refer to the DMP table* for more examples***	*BS-Seq: 100-120 millions reads per sample* *RNA-Seq: 30 millions reads per samples*
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*
Sharing of main data	*Saved and shared after publication* *Please specify*
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***	*SRA database.*

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