Vivaldi Project

Data management plan

Gene expression profile of Vibrio aestuarianus in presence and absence of marine snow

Vibrio aestuarianus
Marine snow
Virulence expression

21/01/2019

DATA MANAGEMENT PLAN

Template sheet for each dataset

Partner name	*UNIGE*
	UNITS
Data category	*Transcriptome (pathogen)*
Concerned WP	*WP4*
	WP4
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-pathogen/Task 4.4/V.aestuarianus/UNIGE*
Description of the data	*Transcriptomic information* *Vibrio aestuarianus genomic transcript sequences (fastq)*
Type	*Sequences*
Period and frequency of data collection	*The gene expression of Vibrio aestuarianus in presence and absence (control) of marine snow has been evaluated at a unique time point*
Geographical site of data collection (if applicable)	*NA*
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.*	*Transcriptomic information has been obtained by Vibrio aestuarianus in presence and absence of marine snow.* *Vibrio aestuarianus grew in a liquid medium. The broth culture has been centrifuged and pelleted, after washing, re-suspended in artificial sea water medium. The concentration of Vibrio aestuarianus has been standardized and both, three aliquots as they are, and other three aliquots mixed with marine snow, were processed for RNA extraction, libraries preparation and Next generation sequencing.*
Protocols *Example: 16S ribosomal RNA gene sequencing by NGS* **Please refer to the DMP table* for more examples**	*RNA sequencing by Illumina paired-end, TruSeq* *Illumina NextSeq 550 platform and a high output v2 kit*
Nature of the collected/generated data *Example: Raw dataset in .blc/.fastqc/.fasta formats for genomic*	*Transcriptomic information* *Processed data will be a fastq transcriptome*

information, and processed datas set will be .vcf/.bed formats. **Please refer to the DMP table* for more examples**	
Coverage (if applicable) *Example: random genomic regions covered at 50 X* **Please refer to the DMP table* for more examples**	*Approximately 15.000.000 total 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 fasta quality files*
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 in UNIGE hard disks and in NCBI repository (no uploaded yet).* *Data will be of public accession after publication*
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**