Vivaldi Project

Data management plan

Genetic sequence data, 2bRAD and PoolSeq

27/05/2019

DATA MANAGEMENT PLAN

Template sheet for each dataset

Partner name	*UNIPD*
Data category	*Genome (host)*
Concerned WP	*WP3* *Choisissez un élément.* *Choisissez un élément.*
Name of the VIVALDI referent(s)	*Genome (Clam): Lucas Bargelloni*
Reference of the dataset *Please refer to the DMP table to find the appropriate reference.* *Ex: Genome-Patho/SubTaskN°/Pathogen/PartnerN°*	*Genome-host/3.1.2/Manila_clam/UNIPD*
Description of the data	*2bRAD and PoolSeq sequences of Manila clam used to identify genetic parkers for parentage assignment*
Type	*Sequences*
Period and frequency of data collection	- *2bRAD sequences (short sequences) from 50 individuals on a previous project* - *PoolSeq sequences (longer sequences, but pooled data) from the same 50 individuals* - *PoolSeq sequences from 109 clams*
Geographical site of data collection (if applicable)	*50 clams from Venice lagoon; 109 clams from SATMAR hatchery*
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***	- *DNA from 50 clams (previously sequences individually using 2bRAD) was sequenced as a single pool in 2017, then compared to individual 2bRAD sequences to validate the pooled sequencing method for SNP identification.* - *109 clams from the SATMAR hatchery were sampled after breeding, in 2016, and their DNA was extracted and sequenced (PoolSeq) in 2018 to select SNPs for parentage assignment of their descendants.* *Whole-genome 150 PE sequencing by Illumina*
Nature of the collected/generated data	*Raw sequence data in fasta format*

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***	
Coverage (if applicable) *Example: random genomic regions covered at 50 X* ***Please refer to the DMP table* for more examples***	*Poolseq per pool: 50x*
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***	*No prerequisites*
Sharing of main data	*Saved and shared after publication* *Publication expected for 2020*
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 is stored on an internal (private) server of the laboratory BCA at the University of Padova*
List, description and storage of associated data (metadata) *Examples: environmental data, mortality monitoring, genotyping…*	*No association with other data*
Sharing of metadata (if relevant)	*Saved and shared after publication* *Publication expected 2020*

*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