



GENIALG
**GENetic diversity exploitation for Innovative macro-
 ALGal biorefinery**

Deliverable D2.4

**Establishment of genomic data platforms for *Laminaria*,
Macrocystis, *Undaria* and *Ascophylum* for future breeding
 initiatives**

Planned delivery date (as in DoA): June 2019

Actual submission date: January 2020

Workpackage: WP2

Workpackage leader: CNRS

Deliverable leader: CNRS

Version: 1.0

Project co-funded by the European Commission within the Horizon 2020 Programme (2014 - 2020)	
Dissemination Level	
PU Public	X
CI Classified, as referred to Commission Decision 2001/844/EC	
CO Confidential, only for members of the consortium (including the Commission Services)	

Research and Innovation action: GA no. 727892

Deliverable report

Project and work package:

Genialg, work package 2

Deliverable 2.4: Establishment of genomic data platforms for *Laminaria*, *Macrocystis*, *Undaria* and *Ascophylum* for future breeding initiatives

Participants:

CNRS (lead), SAMS

Context and objective

This deliverable is associated with Task 3.2. "Exploitation of genomic resources to develop breeding tools." The objective of the sub-task is to exploit genome data generated by the Phaeoexplorer project to prepare the establishment of breeding resources (such as collections of molecular markers) for additional brown algae of economic interest (i.e. in addition to *Saccharina latissima*) including the kelps *Laminaria*, *Macrocystis*, *Undaria* and the wrack *Ascophylum*. The specific aim is to carry out genome sampling to provide information about genetic diversity within these species and to mine genomes for genes of potential interest with respect for example to disease resistance or biomolecule production.

Phaeoexplorer is a large-scale sequencing project funded by the French organisation "France Genomique" (France Genomics) through the "Agence Nationale de la Recherche" (National Research Agency). The project involves the Station Biologique de Roscoff, the Genoscope sequencing centre in Paris and an international consortium of experts involved in genome analysis. The objective of Phaeoexplorer is to provide annotated genome assemblies for 67 strains corresponding to 44 diverse brown algal species plus four unicellular and multicellular sister species. The Phaeoexplorer builds on an earlier project with Genoscope which led to the publication of the first brown algal genome sequence (Cock et al., 2010).

Most of the production of the biological material and the nucleic acid extractions have been carried out in Roscoff, with input from some other members of the consortium, and the sequence data has been generated at Genoscope. This part of the project is virtually complete with only a few samples for recalcitrant strains still to be sequenced. Current efforts are focused on genome assembly and annotation and on making annotated genomes available to the consortium for analysis. The analyses planned within Genialg are integrated into this general strategy.

Status of the activity

Table 1 provides a summary of the status of sequencing for the species targeted within Genialg. Current efforts are focused on *Undaria* and *Macrocystis*, for which we have finished and draft annotated genome assemblies available, respectively.

Table 1. Status of genome sequencing and assembly by Phaeoexplorer for Genialg target species.

	Illumina DNA-seq data	Long read data	RNA-seq data	Genome assembly	Genome annotation
<i>Saccharina</i>	Illumina data available, additional sequencing planned	MinION data available, additional sequencing planned	RNA-seq data available, additional sequencing planned	Awaiting sequence data	Pending
<i>Laminaria</i>	Data obtained	MinION data (13.9 Gb)	Data obtained	In progress	Pending
<i>Macrocystis</i>	Data obtained	MinION data (4.5 Gb)	Data obtained	Draft in-house assembly available	Draft in-house annotation available
<i>Undaria</i>	Data obtained	n/a	Data obtained	594 Mbp	37383 genes
<i>Ascophylum</i>	Data obtained	n/a	Data obtained	750 Mbp	In progress

A large part of the work carried out in Roscoff so far has focused on setting up the bioinformatic infrastructure necessary for the planned analyses. This has included recovery and organisation of raw and processed sequence data from Genoscope and the setting up of an analysis framework, in particular genome browsers. For the raw and processed data files, secure, user-

friendly server access has been organised to protect the original data files. The data is housed on the SBR's in-house server which consists of 55 computing nodes (1,809 CPU Hyper threads), 10,098 Go RAM (to be updated in 2020) MySQL and PostGreSQL databases servers, docker and VM servers and one GPU server.

To facilitate the analysis of the genomes, JBrowse genome browsers are being set up within the context of project website that provides easy access to visualisations of the data (Figure 1).

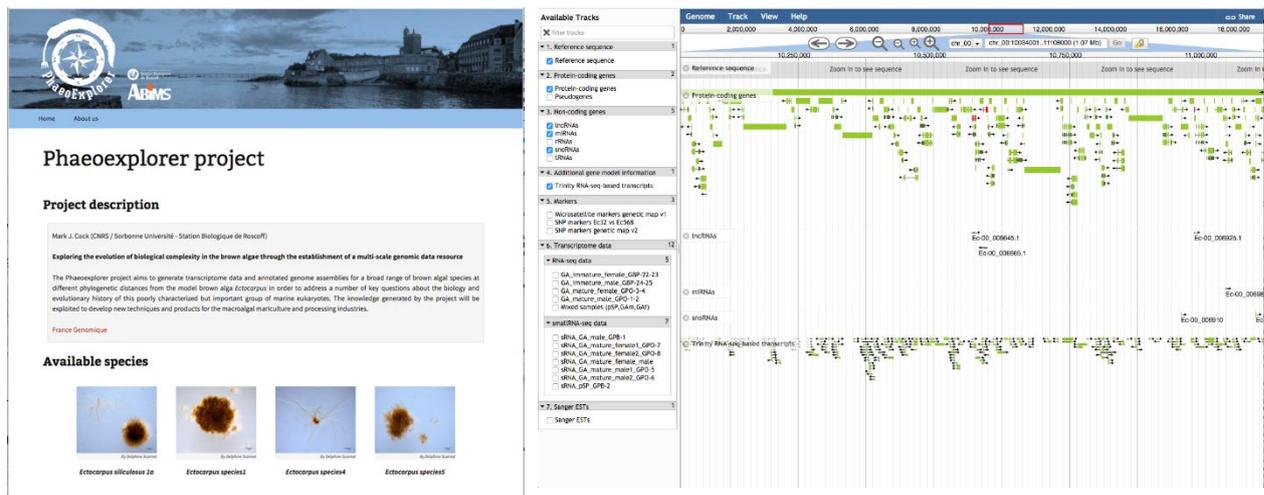


Figure 1. Screen views of the Phaeoexplorer website that has been set up at Roscoff (left) and an example of a genome browser view for the associated JBrowse analysis interfaces (right).

In preparation for the gene mining phase, lists of potentially interesting candidate genes have been established based on an earlier analysis of the *Ectocarpus* genome (Cock et al., 2010; Cormier et al., 2017). These lists include genes putatively involved in key metabolic processes such as cell wall biosynthesis, storage carbohydrate biosynthesis and halide metabolism, putative defence genes, key signalling genes including genes involved in life cycle regulation and sex determination, master developmental regulators and genes at QTL loci associated with temperature and salinity stress resistance (Avia et al., 2017). Table 2 indicates the planning of tasks for 2020 to complete this work.

Table 2. Future planning for the gene mining task

	Month											
	J	F	M	A	M	J	J	A	S	O	N	D
Complete genome assembly and annotation	■	■	■	■								
Upload genomes onto server system		■	■	■	■							
Gene mining				■	■	■	■	■	■			

References

- Avia K, Coelho SM, Montecinos AE, Cormier A, Lerck F, Mauger S, Faugeron S, Valero M, Cock JM, Boudry P (2017) High-density genetic map and identification of QTLs for responses to high temperature and low salinity stresses in the model alga *Ectocarpus* sp. **Sci Rep** 7:43241.
- Cock JM et al. (2010) The *Ectocarpus* genome and the independent evolution of multicellularity in brown algae. **Nature** 465:617–621.
- Cormier A, Avia K, Sterck L, Derrien T, Wucher V, Andres G, Monsoor M, Godfroy O, Lipinska A, Perrineau M-M, Van De Peer Y, Hitte C, Corre E, Coelho SM, Cock JM (2017) Re-annotation, improved large-scale assembly and establishment of a catalogue of noncoding loci for the genome of the model brown alga *Ectocarpus*. **New Phytol** 214:219–232.