



Maritime pine (*Pinus pinaster*) de novo transcriptome assembly

Philippe Label, Céline Noirot, Grégoire G. Le Provost, María-Teresa Cervera, Christophe Plomion, Christophe C. Klopp

► To cite this version:

Philippe Label, Céline Noirot, Grégoire G. Le Provost, María-Teresa Cervera, Christophe Plomion, et al.. Maritime pine (*Pinus pinaster*) de novo transcriptome assembly. 3. workshop on "Statistical challenges on the 1000€ genome sequences in plants";StatSeq'2011, Apr 2011, Toulouse, France. hal-02750183

HAL Id: hal-02750183

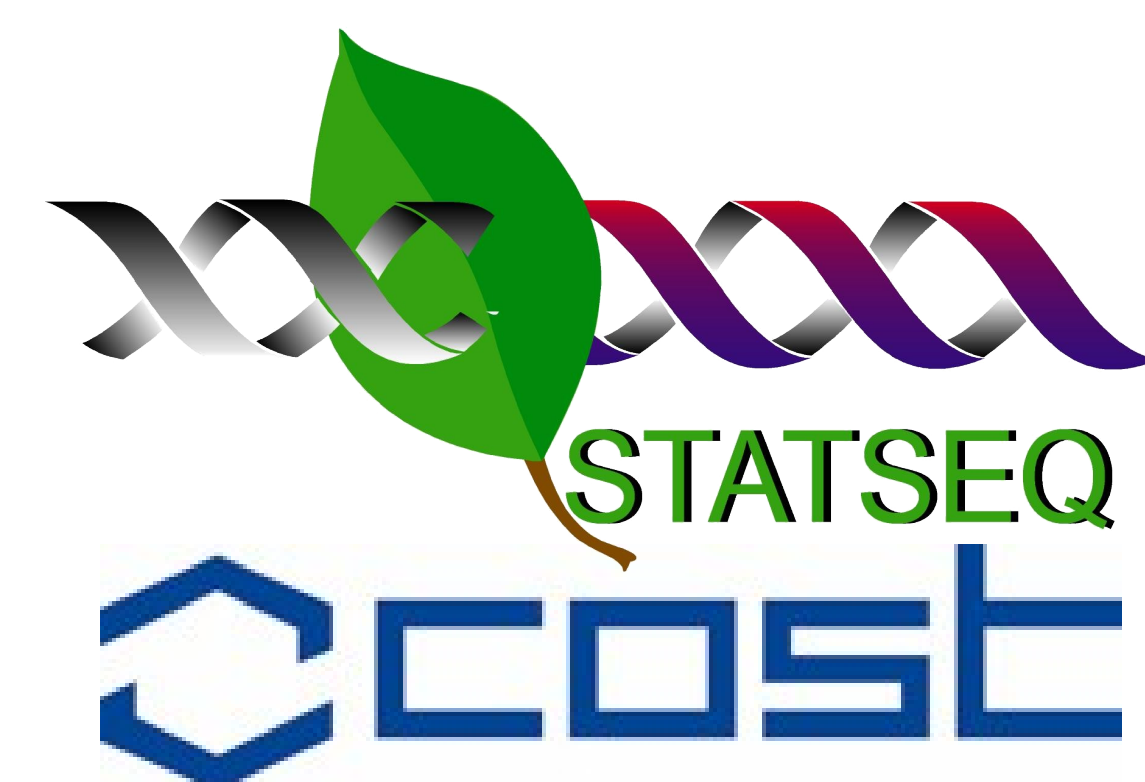
<https://hal.inrae.fr/hal-02750183>

Submitted on 3 Jun 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Maritime pine (*Pinus pinaster*) de novo transcriptome assembly



Philippe Label^{1,2}, Céline Noirot¹, Grégoire Le Provost³, Maria-Teresa Cervera⁴, Christophe Plomion³ and Christophe Klopp¹

¹, INRA-Toulouse, Plateforme Bioinformatique GenoToul, INRA, Chemin de Borde Rouge, BP52 627, 31326 Castanet Tolosan cedex, France
², INRA-Orléans, UAGPF, 2163 Avenue de la Pomme de Pin, CS 40001 Ardon, 45075 Orléans cedex 2, France ; Université Paris Diderot-Paris 7, France
³, INRA-Pierroton, BIOGECO, Domaine de l'Hermitage, 69 route d'Arcachon, 33612 Cestas cedex, France
⁴, INIA-Madrid, CIFOR-INIA, Genómica y Ecología Forestal, Dept. Ecología y Genética Forestal, Carretera de La Coruña km 7,5, 28040 Madrid, Espagne

Rationale

1 Fast increase of **world population** implies increased **wood demand** (heating, pulp & paper, construction).
 Developing the wood resource calls for **cultivated forest** species such as **maritime pine**.

2 Major steps of a typical **sequence processing pipeline** include sequence cleaning, fragment **assembly** and clustering.

3 The sequence **cleaning** step usually includes filtering of **duplicated** reads, short reads, low **quality** reads, **contaminations**, and reads containing **ambiguous bases** (N) above a certain threshold.

Starting Point

Pinus pinaster contigs database: TGICL assembly mainly based on Roche 454 GS-FLX sequences



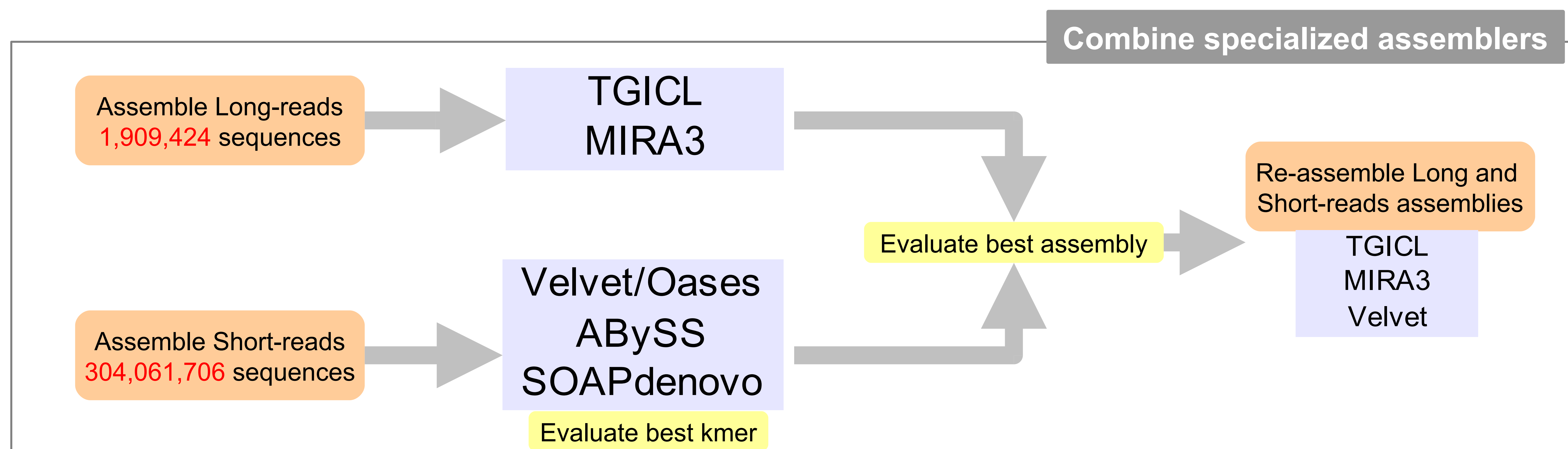
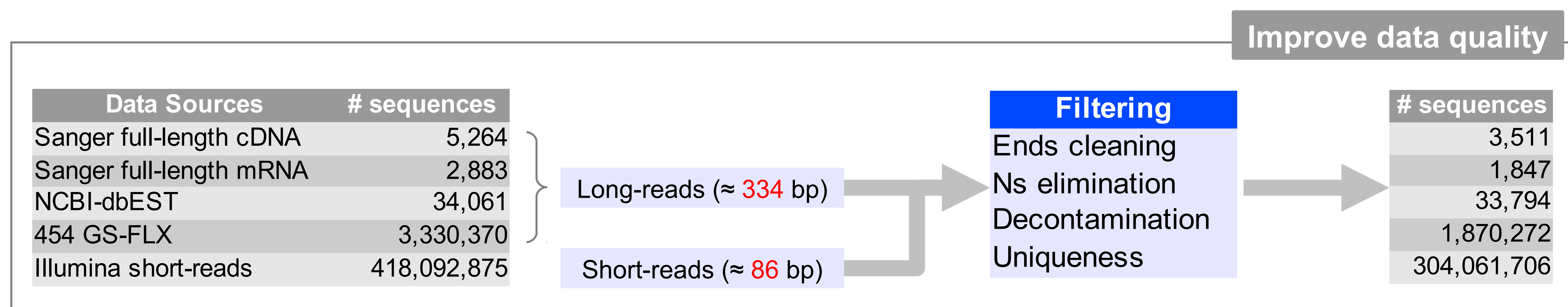
Landes forest in France, 465,389 ha of maritime pine

Assembly:	lpp2, July 2010
Number of base pairs:	102,376,712
Number of sequences:	2,016,588
Total number of contigs:	198,425
Number of singlets:	124,542
Polymorphism:	July 2010
Number of putative SNPs:	30,803
Number of contigs including SNP:	11,936

http://genotoul-contigbrowser.toulouse.inra.fr:9092/Pinus_pinaster2/index.html



Present Strategy



StatSeq'2011

14-15 April, Toulouse. The third StatSeq workshop on
 "Statistical challenges on the 1000€ genome sequences in plants"

ALIMENTATION
 AGRICULTURE
 ENVIRONNEMENT

