Development of a new tool (4TREE) for adapted genome selection in European tree species

Romane Guilbaud, Chiara Biselli, Joukje Buiteveld, Luigi Cattivelli, Paul Copini, Arnaud Dowkiw, Danny Esselink, Agostino Fricano, Vanina Guerin, Véronique Jorge, et al.

To cite this version:

Romane Guilbaud, Chiara Biselli, Joukje Buiteveld, Luigi Cattivelli, Paul Copini, et al.. Development of a new tool (4TREE) for adapted genome selection in European tree species. GenTree, Jan 2020, Avignon, France. . hal-02928391

HAL Id: hal-02928391
https://hal.inrae.fr/hal-02928391v1

Submitted on 2 Sep 2020

Poster online

Development of a new tool (4TREE) for adapted genome selection in European tree species

Authors

R. Guilbaud[1], C. Biselli[2], J. Buiteveld[3], L. Cattivelli[4], P. Copini[3], A. Dowkiw[5], D. Esselink[3], A. Fricano[4], V. Guerin[5], V. Jorge[5], L.J. Kelly[6,7], L. Kodde[3], C.L. Metheringham[6,7], S. Pinosio[8], O. Rogier[5], V. Segura[5], I. Spanu[8], R.J.A. Buggs[6,7], S.C. González-Martínez[9], G. Nervo[2], M.J.M. Smulders[3], L. Sanchez Rodriguez[5], G.G. Vendramin[8], P. Faivre Rampant[1]*

* Corresponding author: Patricia.Faivre-Rampant@inrae.fr

Objective of the 4TREE array: To create a useful, affordable, high-throughput, robust (cross platform utilization), user-friendly and largely reusable genotyping tool for *Populus sp*, *Fraxinus sp*, *Pinus pinaster* and *Pinus pinea,* fostering collaboration between breeding programs across Europe.

Uses: Genome wide association studies, genomic predictions, monitoring and characterizing novel sources of variation.

This array will be used in B4EST for genotyping trees for multi-trait genotype-phenotype and genotype-environment association analyses to pinpoint genome regions related to the performance and adaptation in breeding.

The development of this **Axiom array** is a two step process: one step is the design of a **screening array** (450K SNPs, 480 individuals) to identify robust SNPs for the construction of one **final array** (12.5K SNPs/species) in a second step. This poster presents the screening array.

Pinus pinea

1

- RNA sequencing of 5 individuals from 5 countries
- Reference : Pinus taeda V2.0 genome, Zimin et al. 2017, doi: 10.1093/gigascience/gix072

2

- SNPs calling with GATK V4.0.3.0
- Min coverage: 8
- Min % of reads supporting the alternative allele: 20%
- No interspecific SNPs
- Positions informative in at least 2 samples

3

SNPs submitted	128,279
Recommended	3,495
Neutral	3,212
Not recommended	121,572

4

Priorities: Recommended and neutral markers

Filters:

- Probability of conversion
- Unique probes
- No interfering polymorphisms in the flanking sequences

36,784 SNPs

Populus sp.

1/2

Available data *

Populus nigra

- Infinium array, Faivre-Rampant et al. 2016, doi: 10.1111/1755-0998.12513
- RNAseq, Rogier et al. 2018, doi: 10.1186/s12864-018-5239-z
- WGS, Pégard et al. 2019, doi: 10.1186/s12864-019-5660-y
- GWAS candidates, V. Segura unpublished data

Populus deltoides

- WGS of 5 individuals, Pinosio et al. 2016, doi: 10.1093/molbev/msw161, unpublished data
- Exome capture, Fahrenkrog et al. 2017, doi: 10.1002/ece3.3466

3

SNPs submitted	2,213,994
Recommended	1,738,409
Neutral	196,596
Not recommended	278,989

4

Priorities and filters:

- Intraspecific SNPs to one or both species with a MAF>5%
- Biallelic interspecific SNPs (substitutions)
- SNPs from the existing Infinium array
- Candidate SNPs from GWAS for wood quality
- Homogeneous coverage over the genome
- 60% of SNPs in genes

137,545 SNPs

* No SNPs detection inside the project

Axiom screening array

1 Genomic resources

2 SNPs detection

3

Affymetrix scoring

- **"Recommended":**
  - Probability of conversion > 0.6
  - No interfering polymorphisms within 24 bases
  - 24 bases flanking sequences not shared with another SNP
- **"Not recommended":**
  - SNP and its flanking sequences duplicated in the list
  - Flanking sequences shared with another SNP
  - Probability of conversion < 0.4
  - Interfering polymorphisms in the flanking sequences
  - Homology problems
- **"Neutral":** All the other markers

4 SNPs selection

449,688 SNPs

Evaluation of SNPs selection

Pinus pinea	108 individuals
Pinus pinaster	119 individuals
Populus sp	118 individuals
Fraxinus sp	135 individuals

Final selection

Final array 4TREE
12.5K SNPs/species

Commercially available on demand in summer 2020

Pinus pinaster

1

Available data *

Infinium array, Plomion et al. 2016, doi: 10.1111/1755-0998.12464

Unpublished GenTree data (S. Scalabrin)

- Exome capture on 163 individuals
- Reference: exome *de novo* assembly

2

- SNPs calling with GATK V4.0.3.0
- GATK: QD < 2 , MQ < 40, MQRankSum < -12.5
- Min coverage: 8
- Min % of reads supporting the alternative allele: 20%
- Positions informative in at least 30 samples

3

SNPs submitted	191,785
Recommended	6,562
Neutral	7,898
Not recommended	177,325

4

Priorities: Markers from the Infinium array

Filters:

- Recommended or neutral
- Probability of conversion
- Unique probes
- No interfering polymorphisms in the flanking sequences

137,321 SNPs

Fraxinus sp.

1/2

Available data *

- RNAseq, Harper et al. 2016, doi: 10.1038/srep19335
- WGS, Sollars et al. 2017, doi: 10.1038/nature20786
- WGS, dieback disease, Stocks et al. 2019, doi: 10.1101/626234
- WGS, ash borer, Kelly et al. 2019, doi: 10.1101/772913

3

SNPs submitted	327,064
Recommended	279,337
Neutral	26,059
Not recommended	21,668

4

Priorities: Candidates associated with diseases and incompatibility (up to 8 SNPs per gene or transcript)

Filters:

- SNPs from associative transcriptomics
- Up to 8 SNPs per gene
- MAF> 5%
- MAF<5%
- Recommended
- Coverage

138,038 SNPs

B4EST website

H2020 n° 773383 B4EST

1 Université Paris-Saclay, INRAE, Etude du Polymorphisme des Génomes Végétaux (EPGV), 91000, Evry, France ; 2 Council for Agricultural Research and Economics (CREA), Research Unit for Intensive Wood Production, Casale Monferrato (AL), Italy; 3 Wageningen University and Research, Wageningen, The Netherlands; 4 Council for Agricultural Research and Economics (CREA), Research Centre for Genomics and Bioinformatics, Fiorenzuola d'Arda (PC), Italy; 5 BioForA, INRAE, ONF, 45075, Orléans, France; 6 Royal Botanic Gardens Kew, Richmond, Surrey TW9 3AB, UK; 7 School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NS, UK; 8 Institute of Biosciences and Bioresources, National Research Council, I-50019 Sesto Fiorentino, Florence, Italy ; 9 UMR BIOGECO, INRAE, Université Bordeaux, Cestas, France

Thanks to Marie Laure Schneider, ThermoFisher and Fikret Isik, Conifer SNP Consortium.