Towards a reference sequence of the 1 Gb wheat chromosome 3B

Frédéric Choulet, Sébastien Theil, Adriana A. Alberti, Valérie Barbe, Sophie Mangenot, Arnaud Couloux, Nicolas N. Guilhot, Philippe Leroy, Josquin Daron, Michael M. Alaux, et al.

To cite this version:

Frédéric Choulet, Sébastien Theil, Adriana A. Alberti, Valérie Barbe, Sophie Mangenot, et al.. Towards a reference sequence of the 1 Gb wheat chromosome 3B. 16. Annual Advances in Genome Biology and Technology (AGBT), Feb 2012, Marco Island, United States. ⟨hal-02807639⟩

HAL Id: hal-02807639
https://hal.inrae.fr/hal-02807639v1

Submitted on 6 Jun 2020

Advances In Genome Biology and Technology – Marco Island, FL, USA – Feb. 15-18, 2012

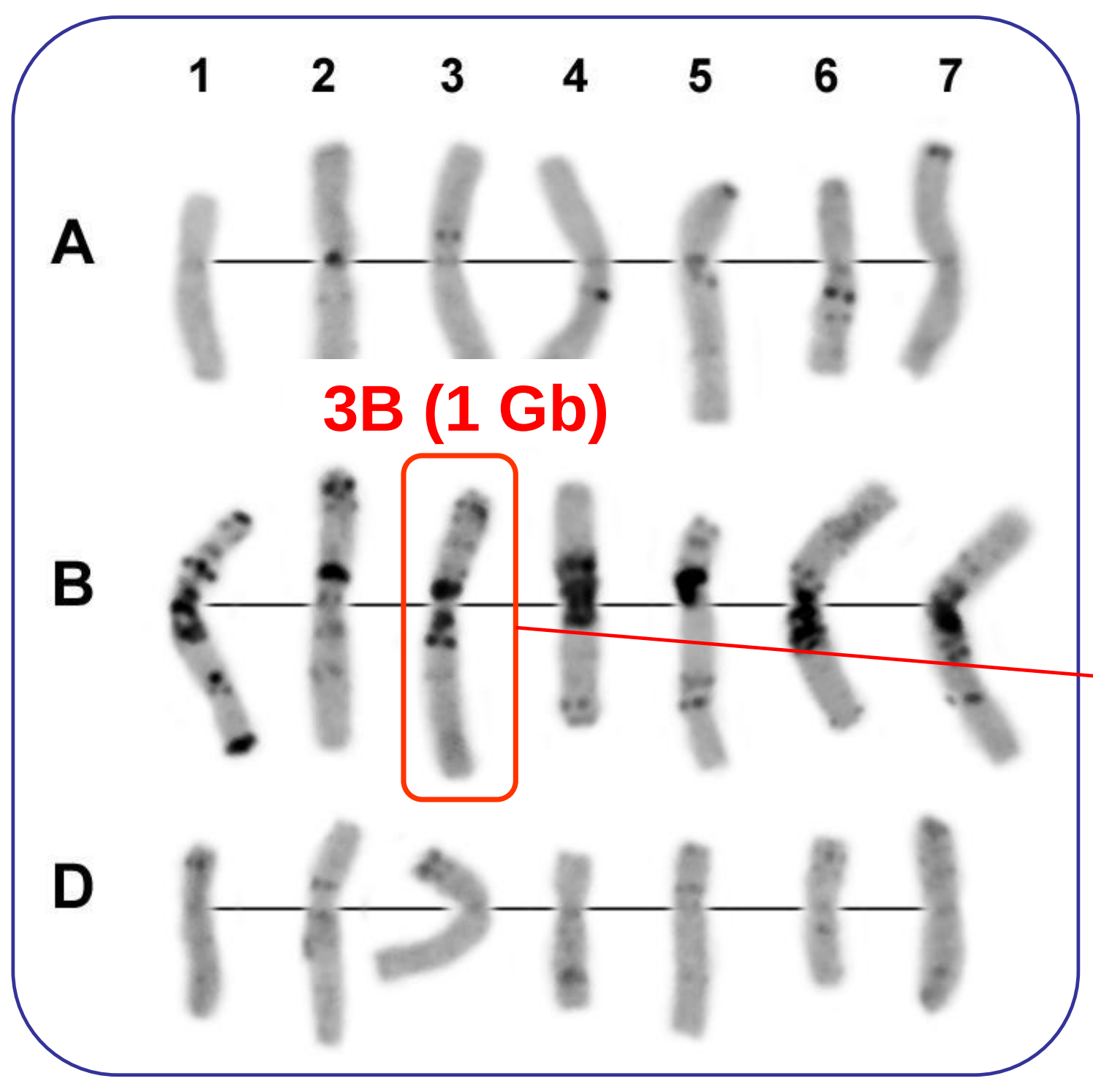
Towards a reference sequence of the 1 Gb wheat chromosome 3B

Frédéric Choulet[1], **Sébastien Theil**[1], **Adriana Alberti**[2], **Valérie Barbe**[2], **Sophie Mangenot**[2], **Arnaud Couloux**[2], **Nicolas Guilhot**[1], **Philippe Leroy**[1], **Josquin Daron**[1], **Michael Alaux**[3], **Hana Šimková**[4], **Jaroslav Doležel**[4], **Arnaud Bellec**[5], **Hélène Bergès**[5], **Pierre Sourdille**[1], **Etienne Paux**[1], **Hadi Quesneville**[3], **Patrick Wincker**[2], **Catherine Feuillet**[1]

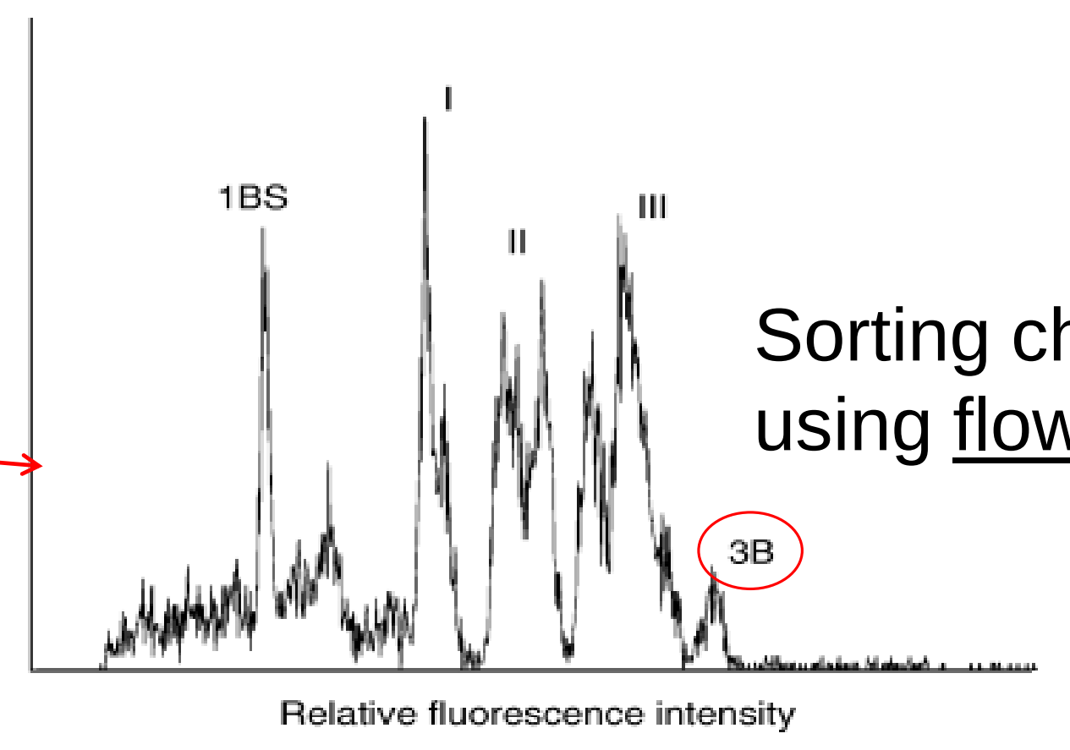
[1]INRA-University Blaise Pascal Joint Research Unit 1095 GDEC, Clermont-Ferrand 63100, France
[2]Genoscope, Institute of Genomics, CEA, Evry 91057, France
[3]INRA Research Unit 1164 Genomics-Informatics, Versailles 78026, France
[4]Institute of Experimental Botany, Olomouc 77200, Czech Republic
[5]INRA French Plant Genomic Resource Centre, Castanet Tolosan 31326, France

Contact: frederic.choulet@clermont.inra.fr

• A chromosome-by-chromosome approach to face the highly complex genome of bread wheat



- ❑ Allohexaploid (3 subgenomes: A, B & D)
- ❑ 17 Gb (1C)
- ❑ 85% of repeated sequences (TEs)

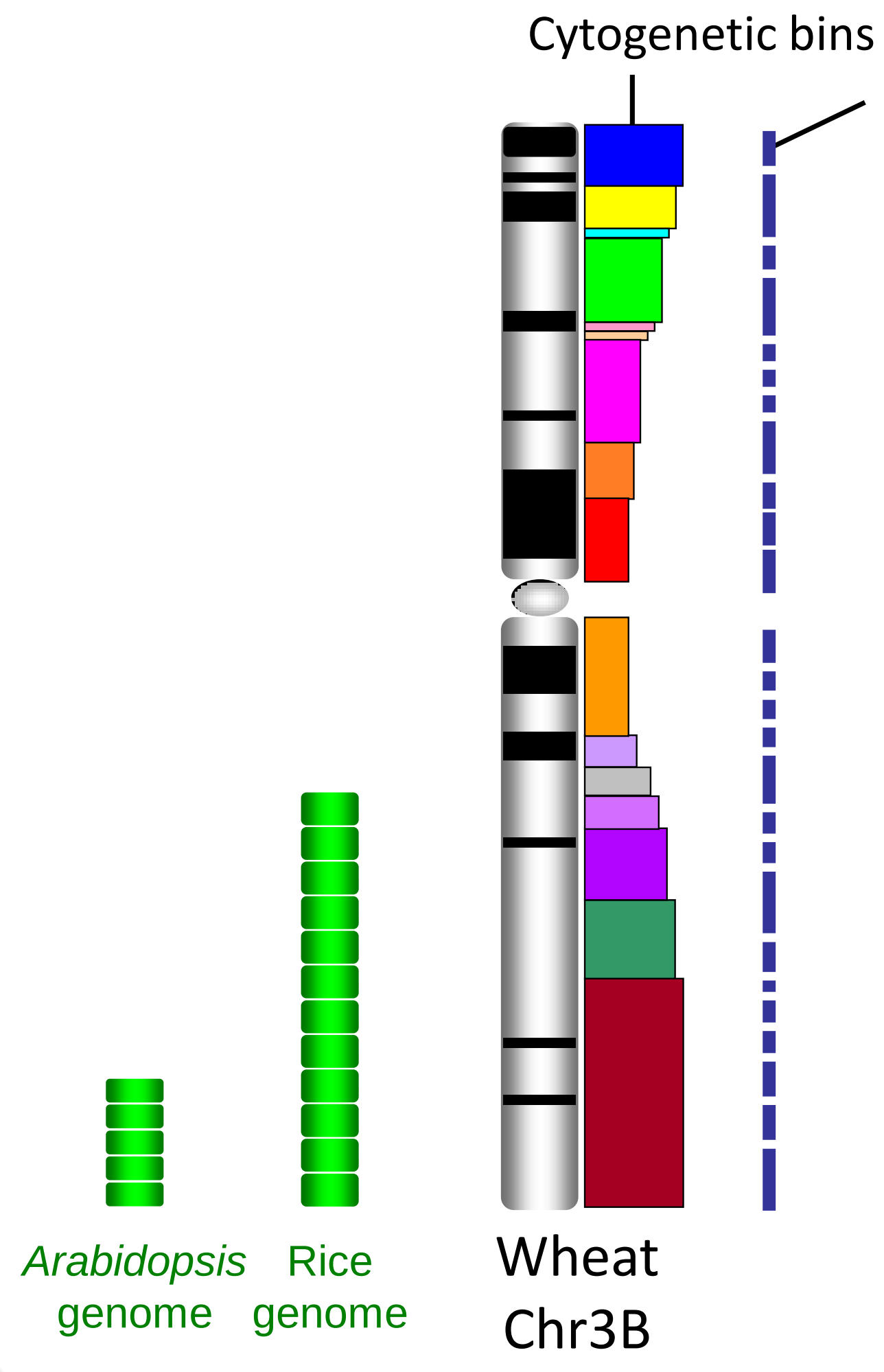


Sorting chromosomes using flow cytometry

The chromosome-based approach is the strategy established by the IWGSC (International Wheat Genome Sequencing Consortium)

➢ 3B is the biggest wheat chromosome
➔ used as a model for physical mapping and sequencing

• Establishment of a physical map of chromosome 3B



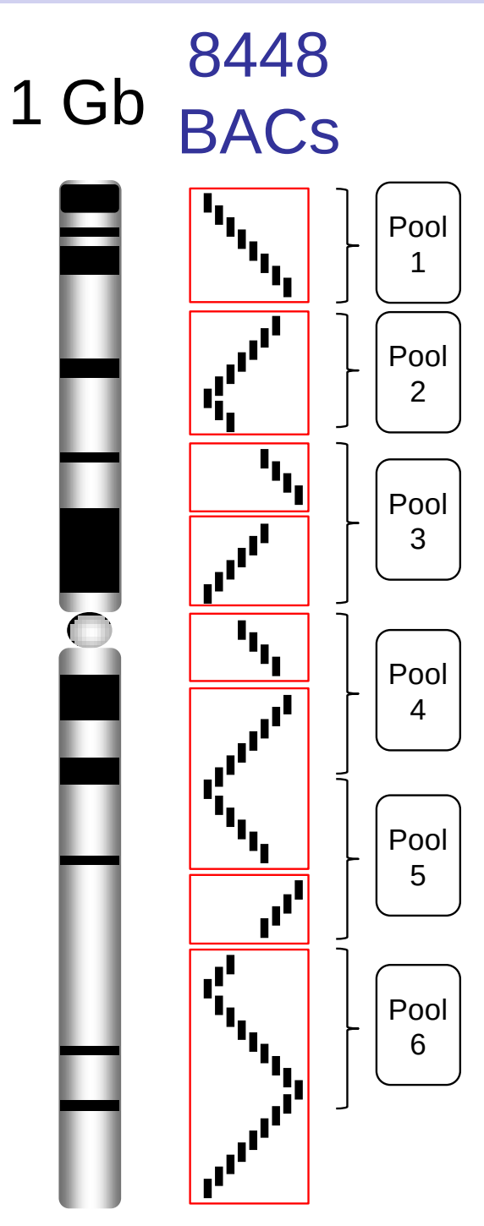
- ❑ 132,000 fingerprinted BACs (19x coverage)
- ❑ 1,282 BAC-contigs (99% of the chromosome)
- ❑ 8448 BAC clones in the Minimal Tilling Path
- ❑ ~4,300 markers assigned to BAC-contigs

Rustenholz et al. Plant Physiol. 2011
Paux et al. Science 2008

• 3BSEQ project: Getting a reference sequence of chromosome 3B

http://urgi.versailles.inra.fr/Projects/3BSeq

❑ *Roche/454 Sequencing of the Minimal Tilling Path*



928 pools of 10 BACs

928 libraries LPE 8kb

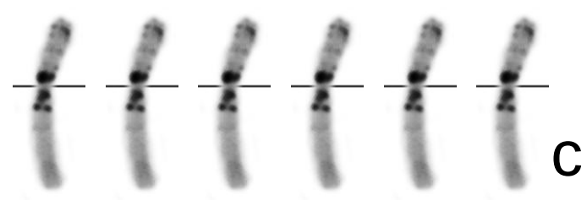
Pool 1 | Pool 2 | Pool 3 | Pool 4 | Pool 5 | Pool 6

150 runs
Roche/454 GSFLX Titanium

➢ **40 Gb** (40x cov.)

Assembly of each pool individually (**NEWBLER**)

❑ *Illumina Whole 3B shotgun*



Sorted 3B chromosomes

Amplified DNA

PE library 0.5 kb

Illumina HiSeq2000

➢ 845 M reads (2x108 bp)
➢ **82 Gb** (82x cov.)

Assembly (**ABYSS**)

❑ *BAC-ends sequencing*

Bac-End sequencing of 3 versions of the MTP

40175 BacEnds ⇔ 25 Mb

• 3BSEQ assembly results

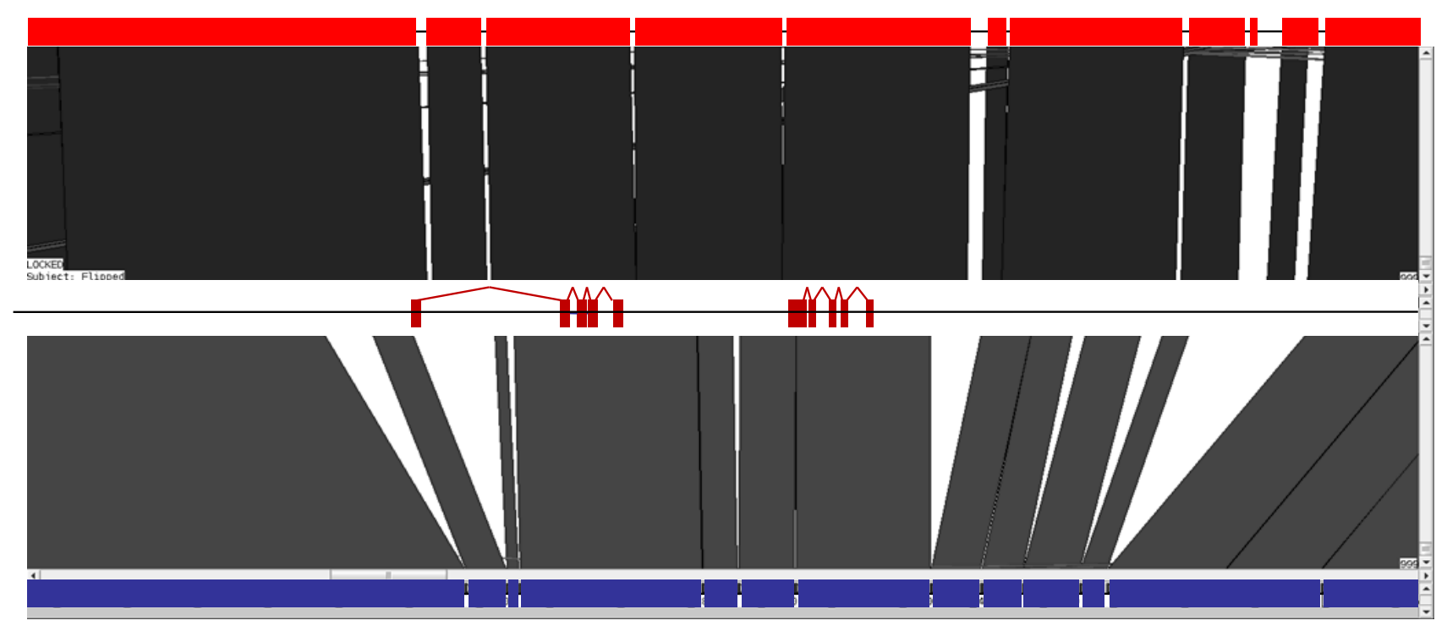
❑ *Roche/454 Sequencing of the Minimal Tilling Path*

	Scaffolds
# Scaffolds	`16,167 scaff`
# bps	`1028 Mb`
Average	`64 kb`
L50	`1028 scaff`
N50	`277 kb`
Max size	`1.3 Mb`

❑ *Illumina Whole 3B shotgun*

	Contigs >200bp
# Contigs	`546,922 ctg`
# bp	`639 Mb`
Average size	`1.2 kb`
N50	`2.7 kb`
Max size	`48 kb`

Example of assembly accuracy according to a 50kb reference sequence:



1 scaffold (Roche/454)

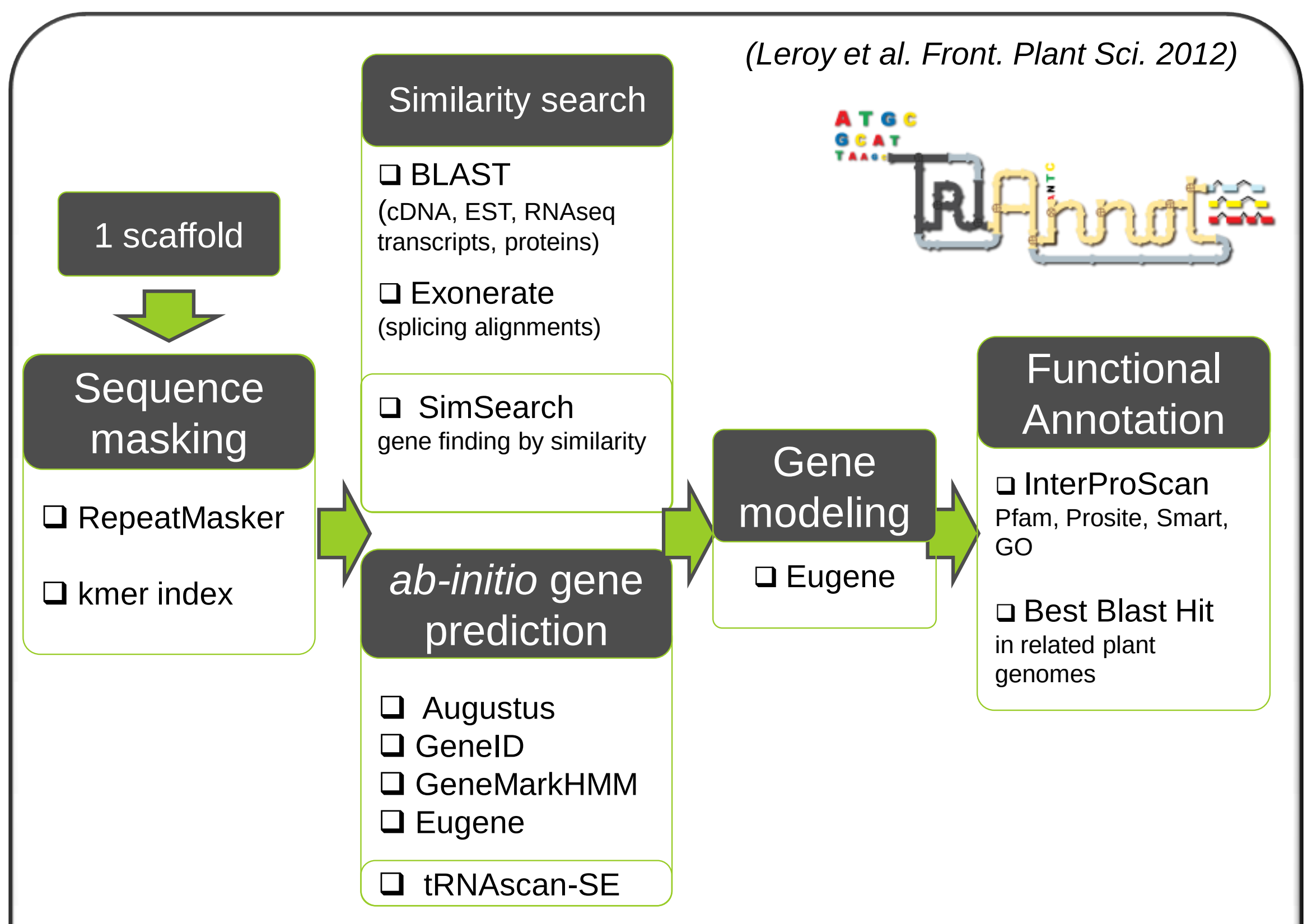
Reference seq. (with 2 genes)

13 unordered contigs (Illumina)

• The TriAnnot pipeline

(http://www.clermont.inra.fr/triannot)

(Leroy et al. Front. Plant Sci. 2012)



1 scaffold

Sequence masking
- ❑ RepeatMasker
- ❑ kmer index

Similarity search
- ❑ BLAST (cDNA, EST, RNAseq transcripts, proteins)
- ❑ Exonerate (splicing alignments)
- ❑ SimSearch gene finding by similarity

***ab-initio* gene prediction**
- ❑ Augustus
- ❑ GeneID
- ❑ GeneMarkHMM
- ❑ Eugene
- ❑ tRNAscan-SE

Gene modeling
- ❑ Eugene

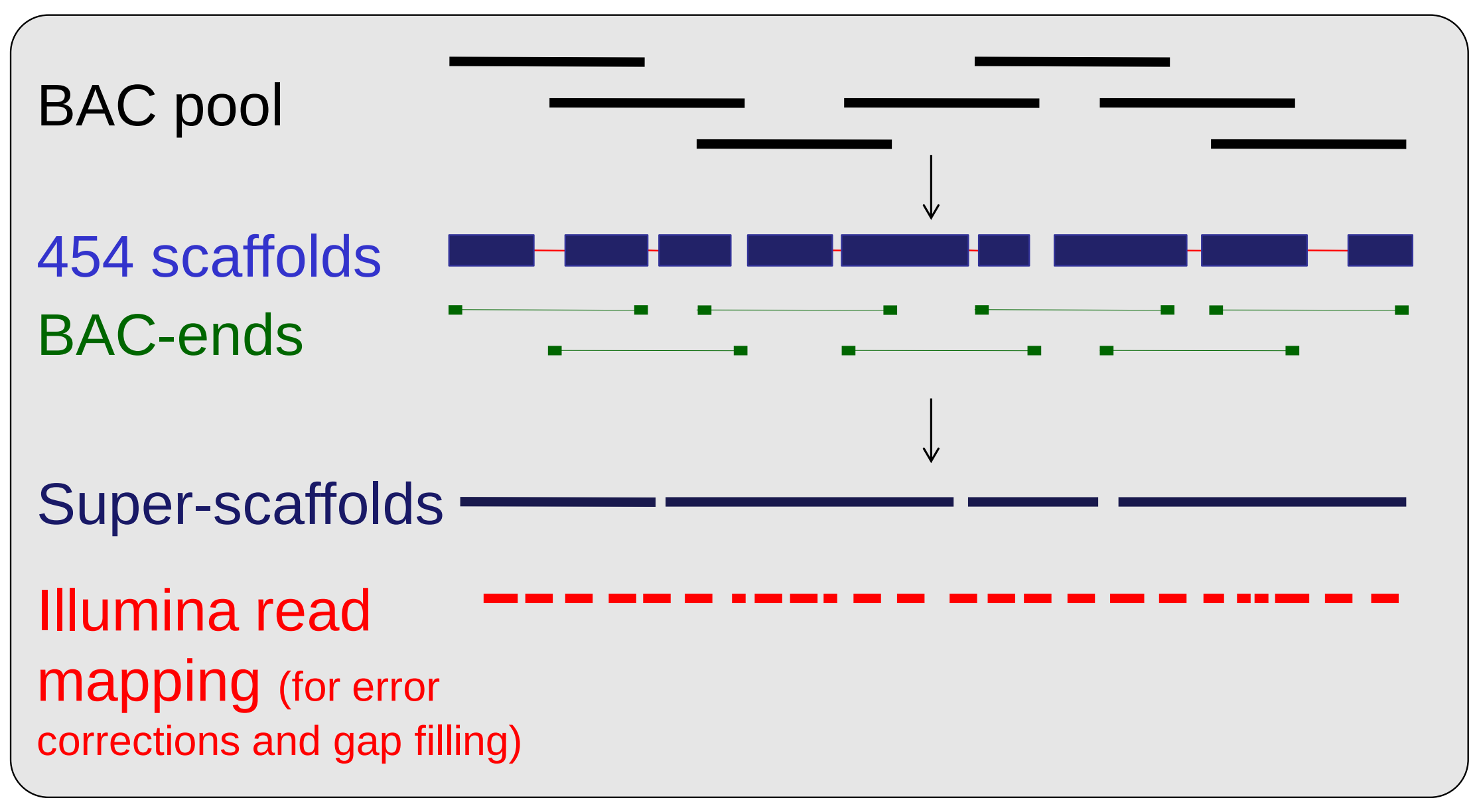
Functional Annotation
- ❑ InterProScan Pfam, Prosite, Smart, GO
- ❑ Best Blast Hit in related plant genomes

8734 genes predicted (1/114 kb)

• Work in progress to build a pseudomolecule

➢ **Merging scaffolds of different BAC pools based on sequence comparison**

➢ **Integrating 454Scaffolds/Illumina reads/BES**



BAC pool

454 scaffolds
BAC-ends

Super-scaffolds

Illumina read mapping (for error corrections and gap filling)

• Additional studies

➢ **Anchoring and orientation of scaffolds**

4000 SNPs to be developed
- ➢ **Cytogenetic bin mapping (aneuploid lines)**
- ➢ **Genetic mapping (381 F2 individuals + 2 parents)**
- ➢ **Radiation hybrid mapping**

➢ **RNASeq (15 different conditions)**

40x2(PE) M reads x 2 replicates x 15 conditions

➢ **Structural polymorphism (CNVs, PAVs)**

Illumina shotgun sequencing of chromosomes 3A and 3D
Illumina shotgun sequencing of chromosomes 3B from 18 wheat varieties

FRENCH NATIONAL INSTITUTE FOR AGRICULTURAL RESEARCH
Center of Clermont-Ferrand / Theix
Joint Research Unit INRA-University Blaise Pascal 1095, Genetics Diversity and Ecophysiology of Cereals
http://www.clermont.inra.fr/umr1095_eng/