Sequencing and analysis of chr. 1B

Frédéric CHOULET
Genetics Diversity Ecophysiology of Cereals, INRA-UBP
Clermont-Ferrand, France

Partners, funders

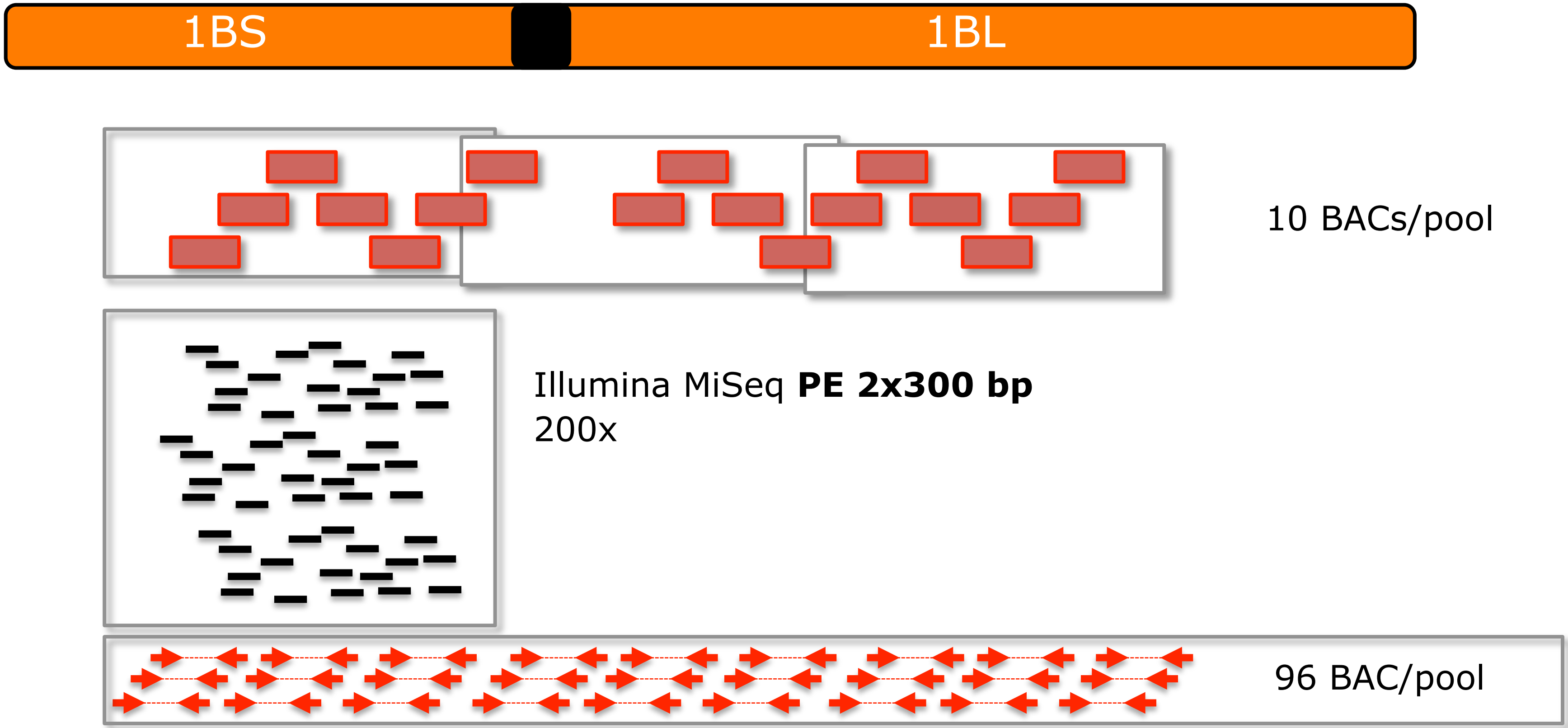
- *INRA GDEC*
- *Génoscope*
- *INRA CNRGV*
- *INRA URGI*
- *University of Haifa*

INRA

*PIA **BreedWheat** project*

*France Génomique **WheatSeq** project*

Strategy



10 BACs/pool

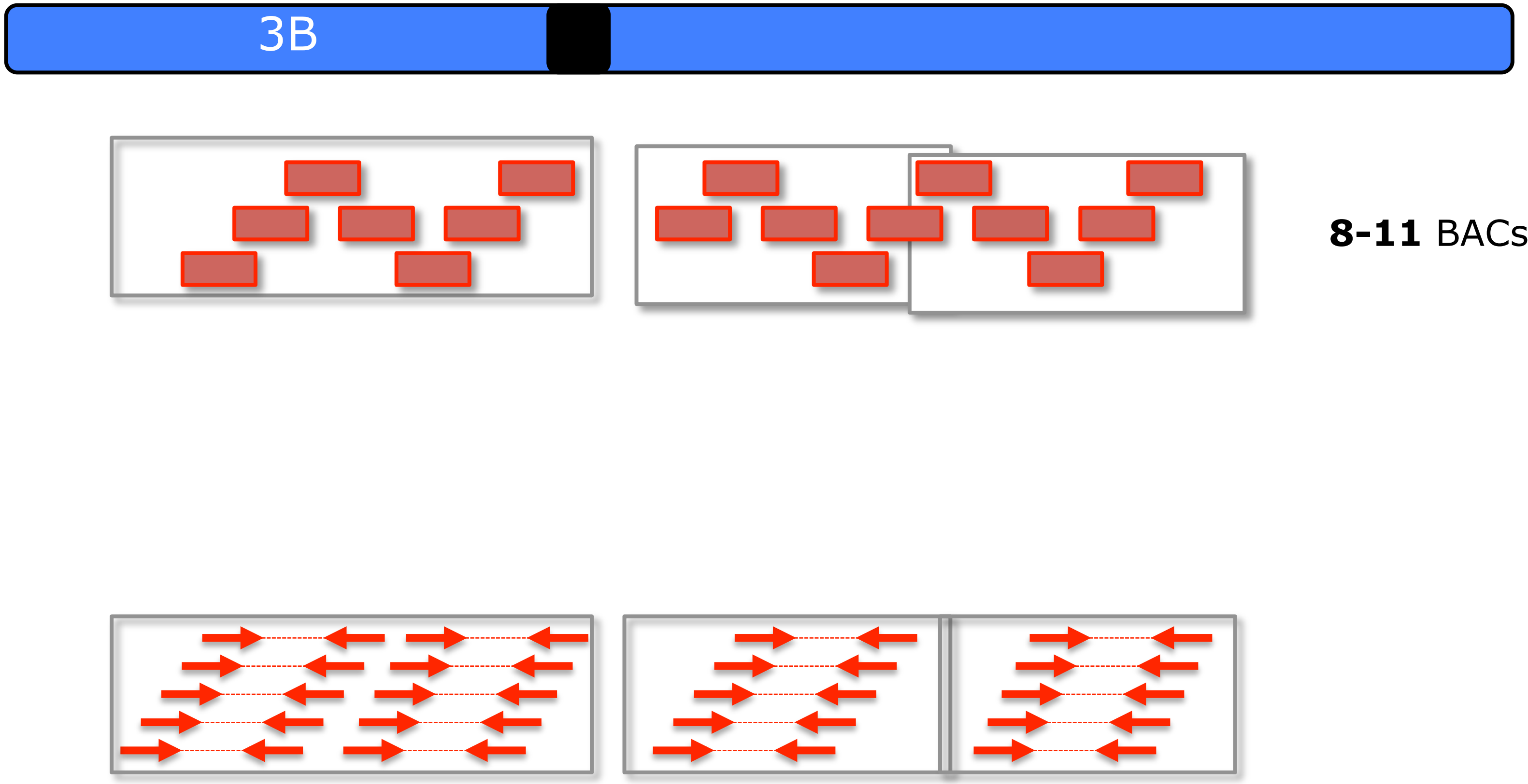
Illumina MiSeq **PE 2x300 bp**
200x

96 BAC/pool

Illumina HiSeq **5kb MP 2x150 bp**
100x

CEA-IG-Génoscope
- ○ *Valérie Barbe*
- ○ *Adriana Alberti*
- ○ *Karine Labadie*
- ○ *Sophie Mangenot*
- ○ *Arnaud Couloux*
- ○ *Jean-Marc Aury*
- ○ *Patrick Wincker*

Strategy



8-11 BACs/pool

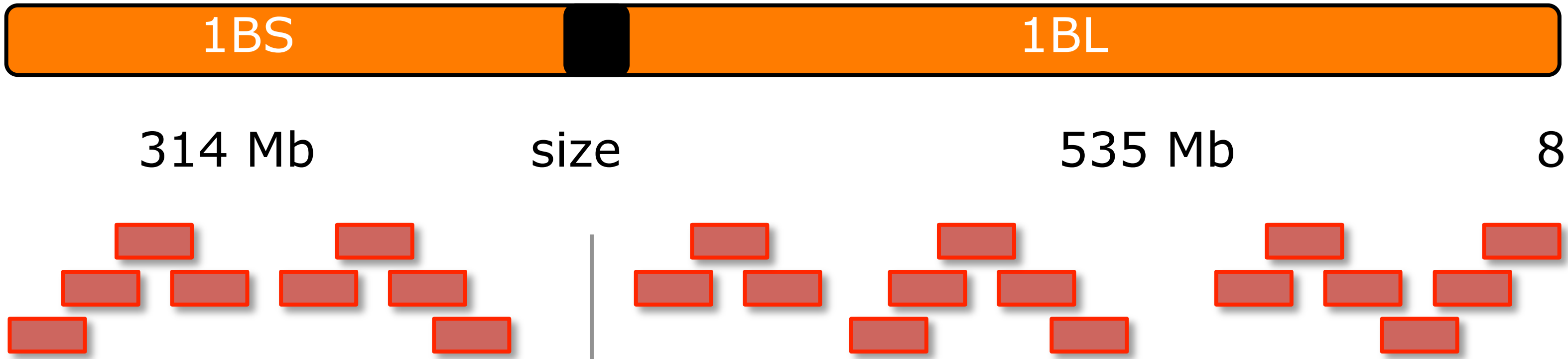
Roche454 **8kb-MP 450 bp**

40x

CEA-IG-Génoscope

- *Valérie Barbe*
- *Adriana Alberti*
- *Karine Labadie*
- *Sophie Mangenot*
- *Arnaud Couloux*
- *Jean-Marc Aury*
- *Patrick Wincker*

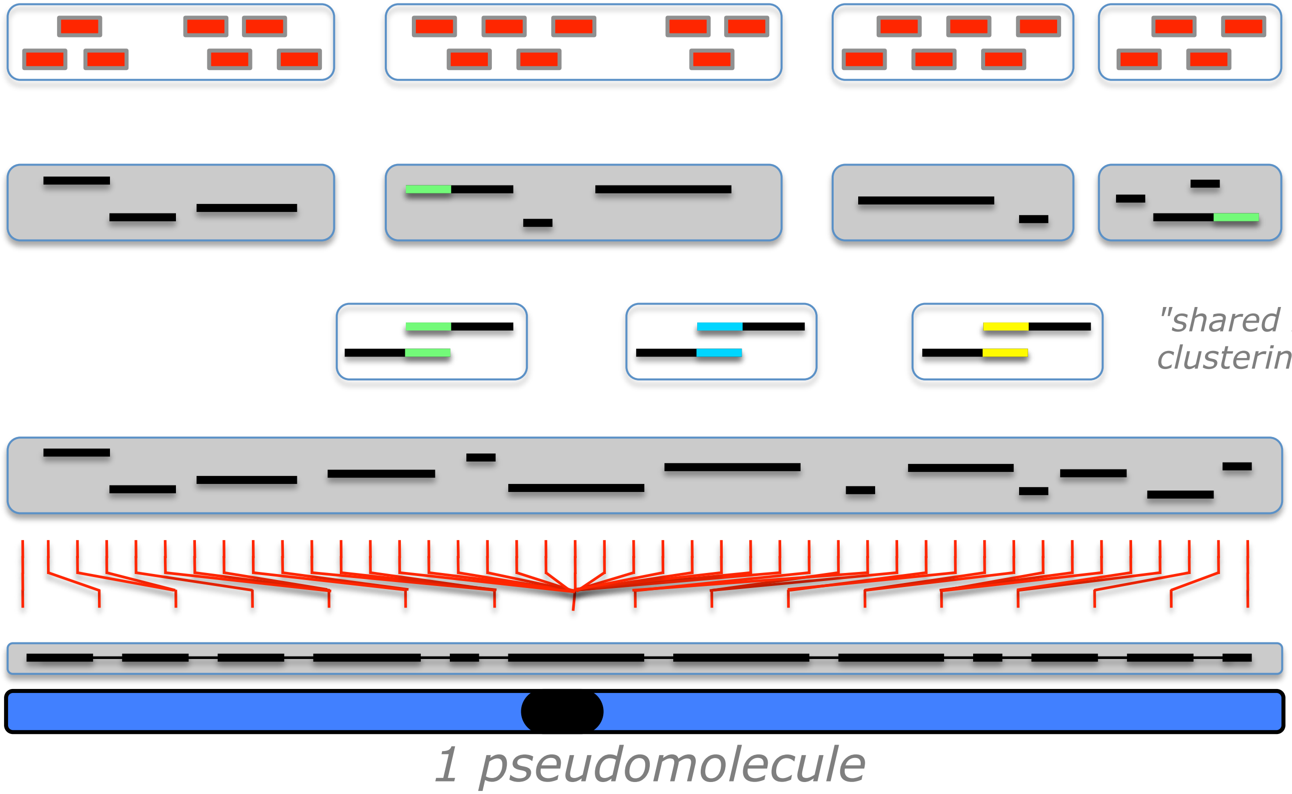
Physical maps



314 Mb	size	535 Mb	849 Mb
4372	#MTP-BACs	6023	10395
-	#PE-pools	-	1083
-	#MP-pools	-	109

INRA CNRGV
- ○ *Arnaud Bellec*
- ○ *Hélène Bergès*

Pseudomolecule construction



BAC-pools

Assembly
Scaffolding
Gap filling

"shared ISBP"-based clustering

Merging

map

Anchoring

1 pseudomolecule

Assembly – scaffolding – merging

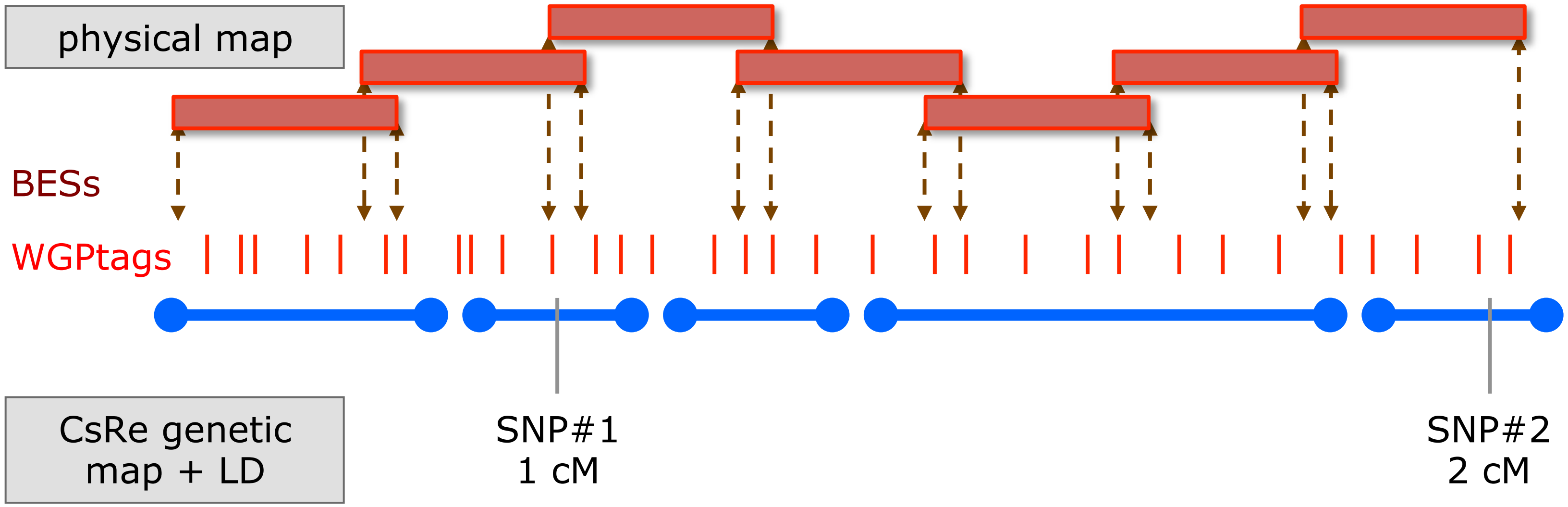
	1B	3B
expected size	849 Mb	995 Mb
pool assembly	923 Mb **13386** scaff, N50=**350 kb** **1% Ns**	1040 Mb **16136** scaff, N50=**275 kb**
Gap filling, Manual Scaffolding Improvement (BACends)	-	992 Mb **4999** scaff, N50=**463 kb** **7% Ns**
Merging	807 Mb (95%) redundancy removed=13% **10979** scaff, N50=**656 kb**	833 Mb (84%) redundancy removed=16% **2808** scaff, N50=**892 kb**
with cutoff >10kb	**2746** scaff, N50=**694 kb**	**1913** scaff, N50=**895 kb**

Differences due to:

- Physical map 1B > 3B
- Sequencing tech 1B > 3B
- BAC pooling 3B > 1B
- Manual finishing 3B >> 1B

	1B	3B
expected size	849 Mb	995 Mb
all scaffolds	807 Mb (95%)	833 Mb (83%)
completeness	96%	94%

Anchoring



	1B	3B
expected size	849 Mb	995 Mb
all scaffolds	807 Mb (95%)	833 Mb (83%)
completeness	96%	94%

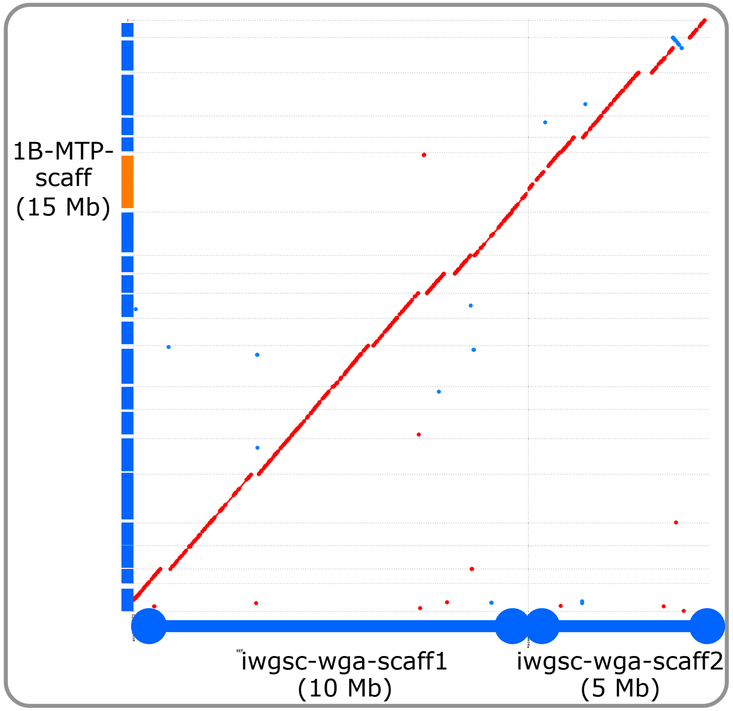
Anchoring

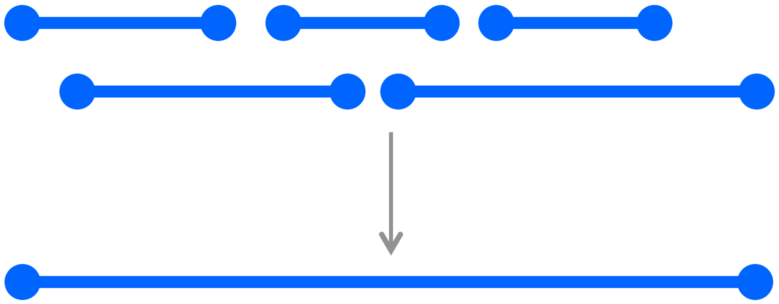
physical map		
BESs	**9706** BACs with BESs	**20953** BACs with BESs
WGPtags	WGPtags to come !!!! ☺	**14199** BACs with WGPtags
CsRe genetic map + LD	**5636** SNPs	~**2000** SNPs
Pseudomolecule	`v0.2.0` **1356 scaff / 552 Mb / 71%**	*Choulet et al. 2014* 1358 scaff / 774 Mb / **93%**

1B will reach a gold standard…

- ○ Completeness +++
- ○ Contiguity ++
- ○ %Anchored + (->to be increased)
  - WGP-tags
  - TGAC WGS assembly
  - IWGSC WGA *(Illumina+NRGene)*

Combine all available resources ***-> 1B gold standard pseudomol***

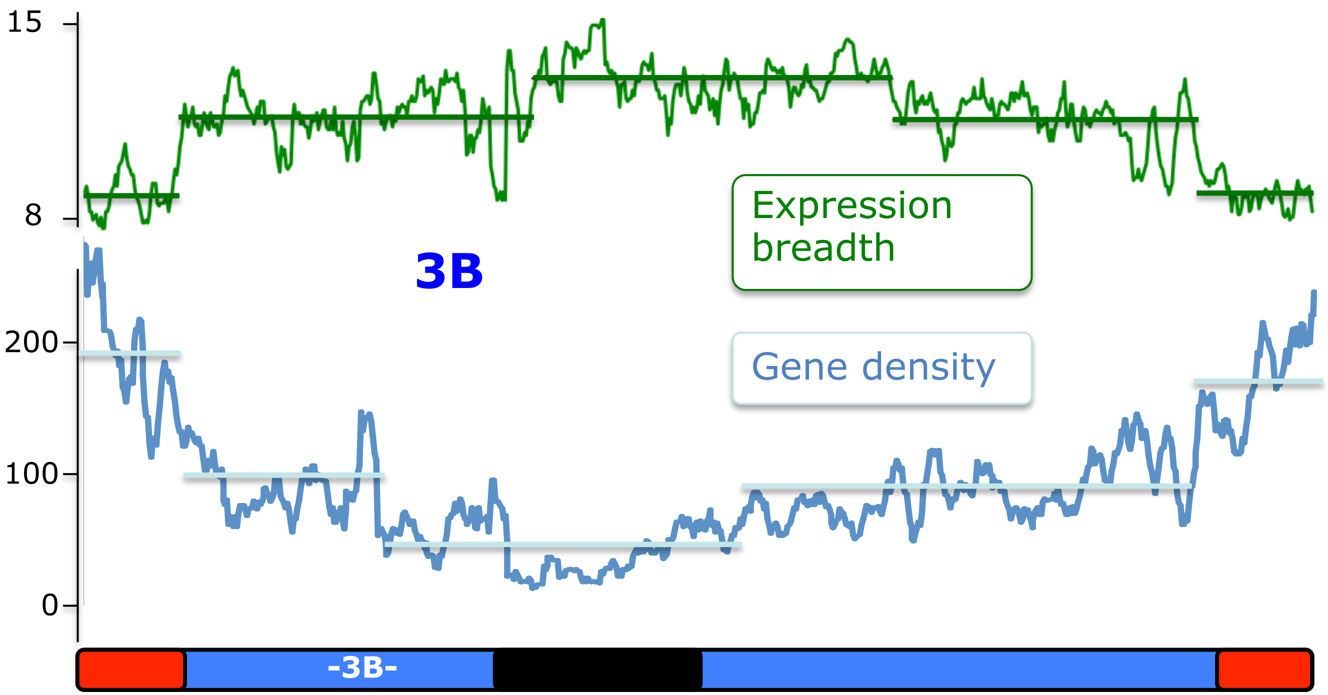




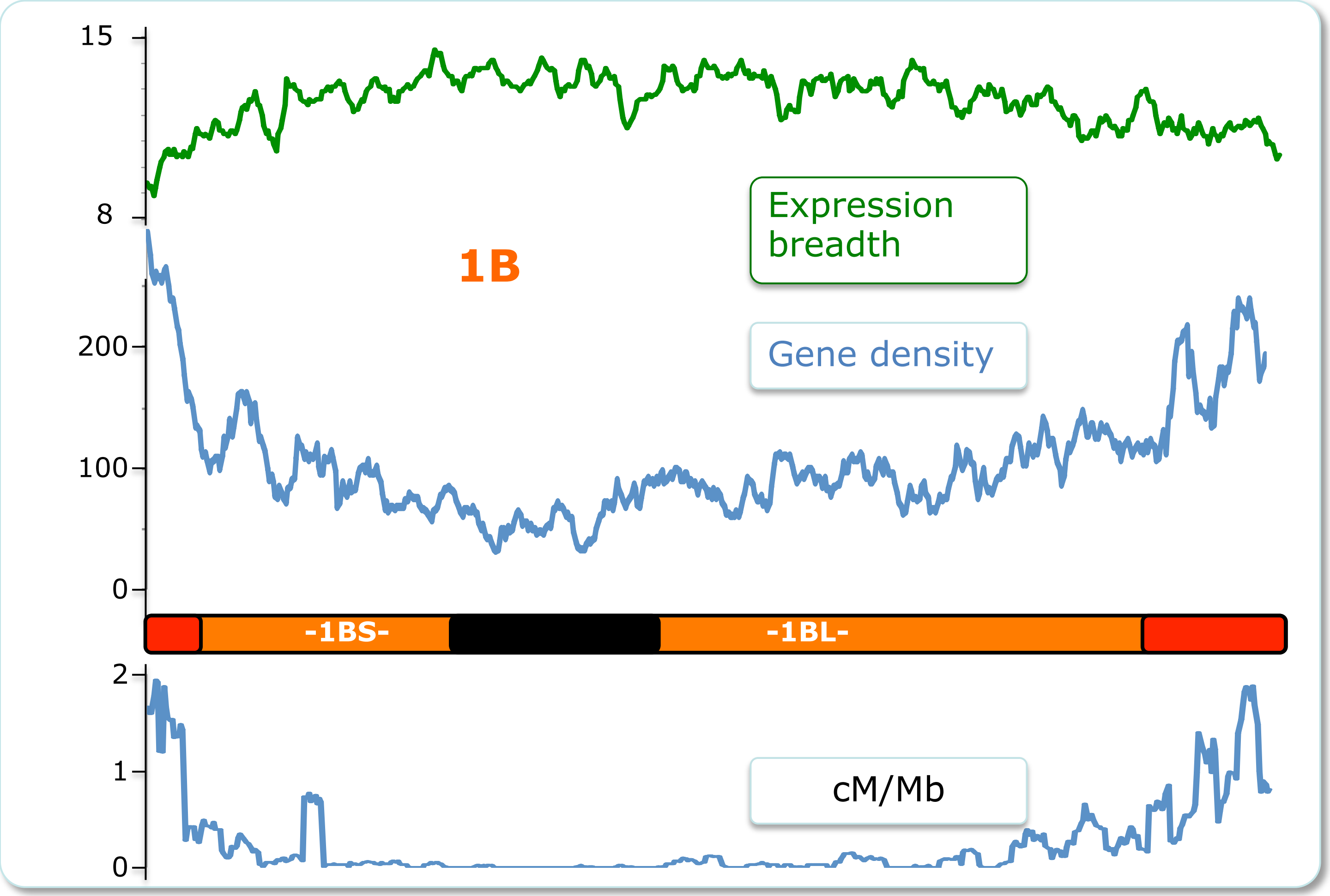
Sequence content

- Gene and TE annotation (TriAnnot):
  - **6369** gene models
  - **70%** expressed (RNASeq 15 samples *Pingault et al. 2015*)
  - **88%** TEs

Chromosome organization



Chromosome organization



❑ Acknowlegments

GDEC

AA. Josselin
H. Rimbert
B. Darrier
P. Sourdille
F. Balfourier
J. Kitt
R. Philippe
E. Paux

University of Haifa

T. Fahima
A. Korol
Z. Frenkel

INRA CNRGV

A. Bellec
H. Bergès

INRA URGI

T. Letellier
M. Alaux
H. Quesneville

CEA-IG-Génoscope

V. Barbe
A. Alberti
K. Labadie
S. Mangenot
A. Couloux
JM. Aury
P. Wincker

Gene name nomenclature

TRAES1BS**000000**CFD	*gene*
TRAES1BS**000000**CFD_t1	*transcript*
TRAES1BS**000000**CFD_c1	*CDS*
TRAES1BS**000000**CFD_p1	*protein*

6 digits

#scaff #gene

TRAES3BF**000100010**CFD