

**IWGSC Workshop
April 8, 2011
Prague, Czech Republic**

Physical mapping and anchoring in hexaploid wheat

Etienne Paux

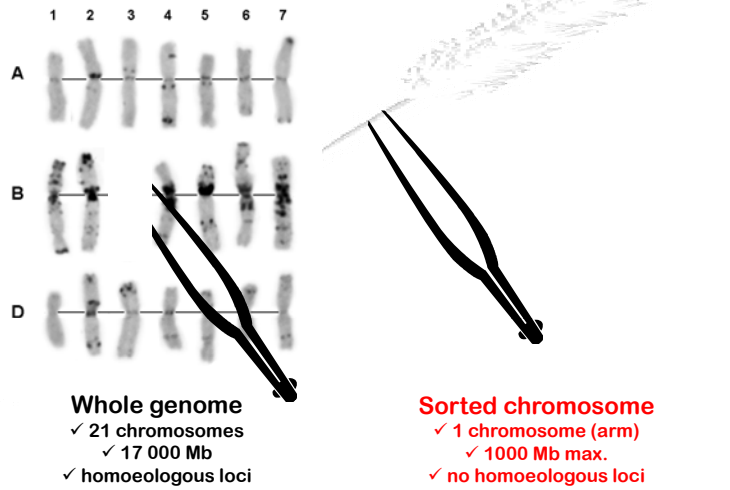
**Structure, Function & Evolution of the Wheat Genomes
Genetics, Diversity & Ecophysiology of Cereals
INRA Clermont-Ferrand, France**



A brief introduction: Physical mapping overview

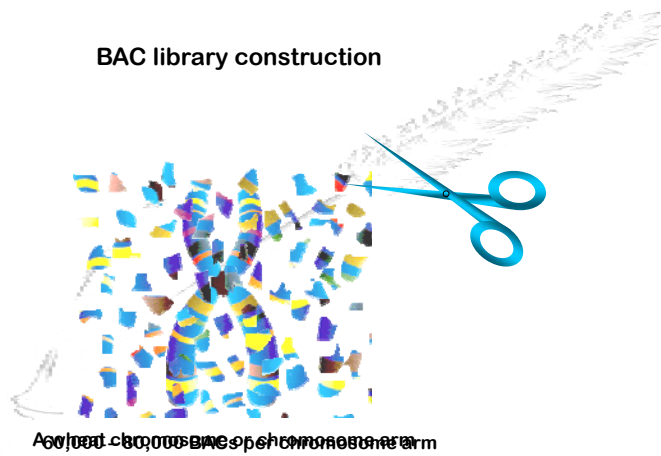
Construction of chromosome-specific integrated maps

Chromosome or chromosome arm sorting



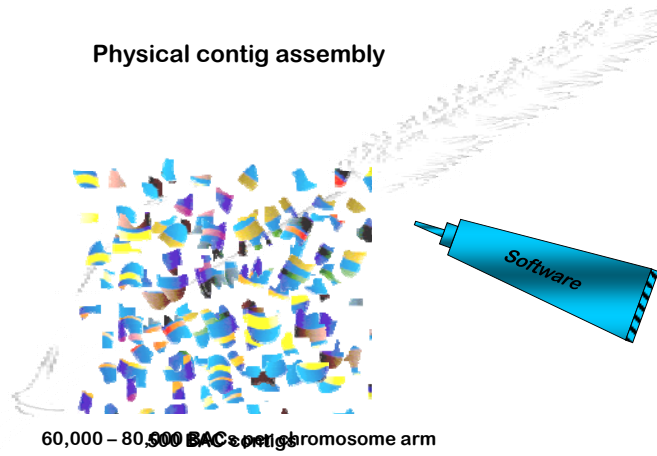
Construction of chromosome-specific integrated maps

BAC library construction



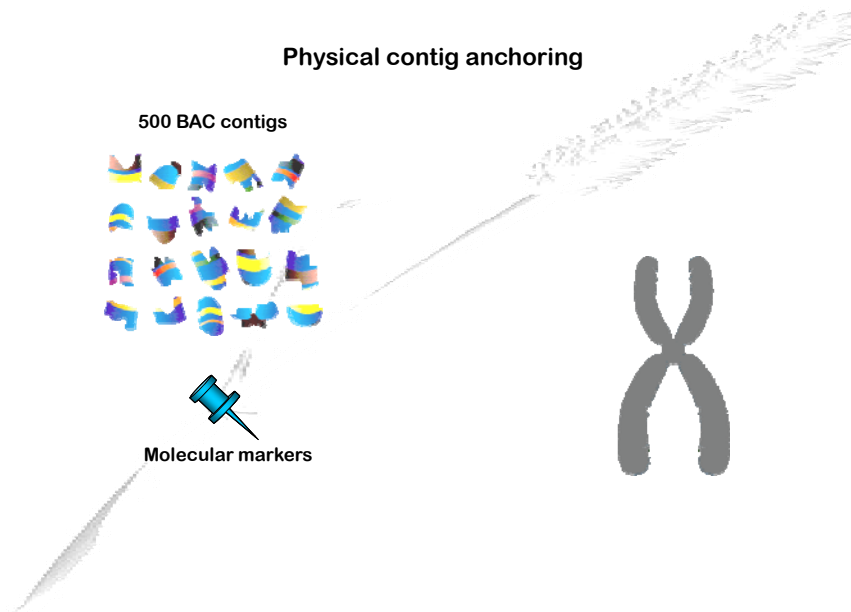
Construction of chromosome-specific integrated maps

Physical contig assembly



Construction of chromosome-specific integrated maps

Physical contig anchoring





BAC fingerprinting

Different fingerprinting methods

- **Coulson *et al.* (1986)**
 - ✓ *Hind*III
 - ✓ BEt staining and agarose gel
 - ✓ 30-40 bands per BAC
- **Zhang *et al.* (1997)**
 - ✓ *Hind*III + *Hae*III
 - ✓ BEt staining and agarose gel
 - ✓ 45-50 bands per BAC
- **Faller *et al.* (2000)**
 - ✓ *Eat*I + *Taq*I
 - ✓ fluorescent labelling (ddATP, ddGTP, ddTTP) and capillary sequencer
 - ✓ 100-110 bands per BAC
- **Luo *et al.* (2003)**
 - ✓ *Bam*HI + *Hind*III + *Xba*I + *Xho*I + *Hae*III
 - ✓ fluorescent labelling (ddNTPs) and capillary sequencer
 - ✓ 120-150 bands per BAC
- **KeyGene (2008)**
 - ✓ *Eco*RI + *Mse*I
 - ✓ Illumina sequencing
 - ✓ 15-60 tags per BAC

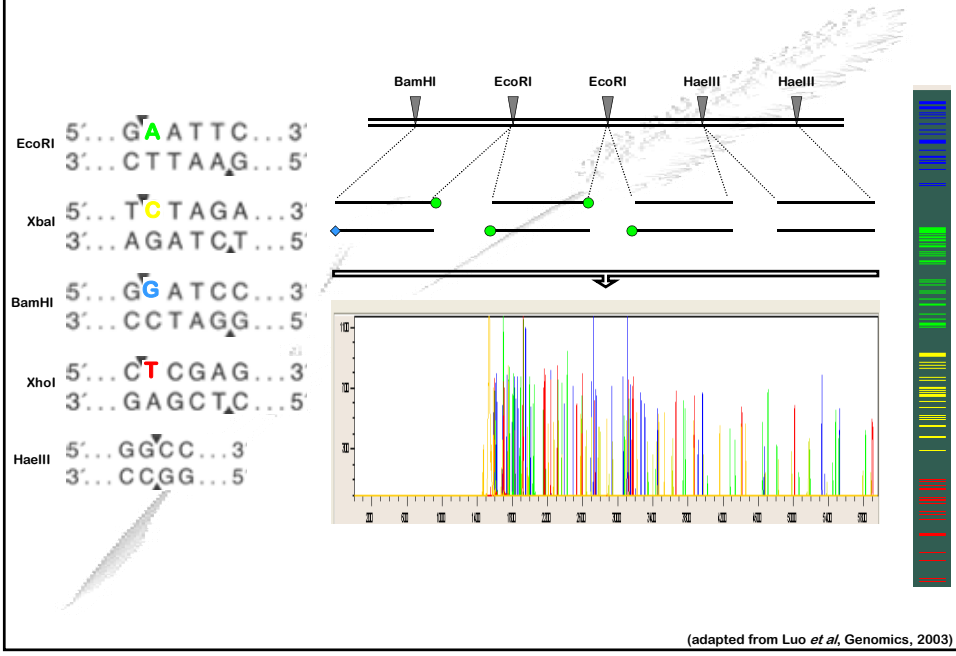
b BAC clone
 Digestion (rare cutter, *Hind*III)
 Separation on agarose gel; DNA staining by ethidium bromide or SYBR Green

a BAC clone
 Digestion (with a type IIS, *Eat*I and a frequent cutter, *Tsp*I) and fluorescent labelling (ddATP, ddGTP, ddTTP)
 Separation and detection on acrylamide gel (using an automated sequencer)

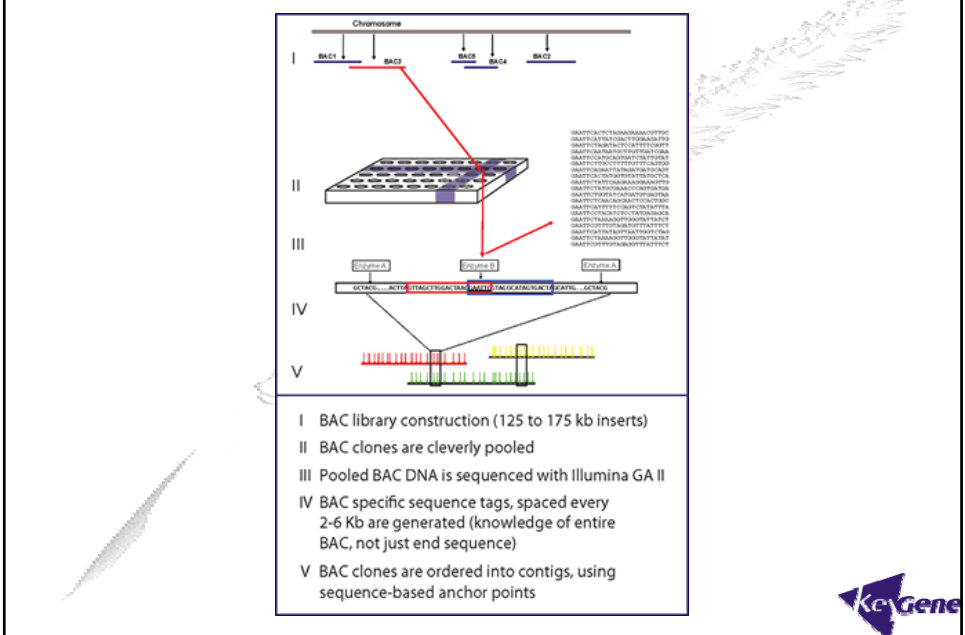
c BAC clone
 Digestion (with four rare cutters, *Eco*RI, *Bam*HI, *Xho*I, *Xba*I and a frequent cutter, *Hae*III)
 Fluorescent labelling (ddNTPs)
 Separation and detection on acrylamide gel (using a capillary-automated sequencer)

(adapted from Meyers *et al.*, Nature Rev Genet, 2004)

SNAPshot fingerprinting



Whole-genome profiling (KeyGene)





Editing fingerprints

Cleaning fingerprints using FPB

Scalabrin *et al.* (2009) Automated FingerPrint Background removal: FPB. *BMC Bioinformatics*, 10:127

Each peak represents a fragment with a certain size and intensity and it can derive from different sources:

Background removal

Pre-processing

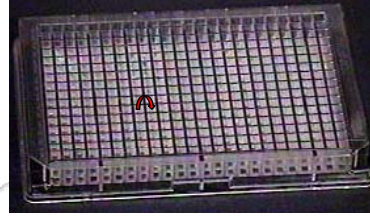
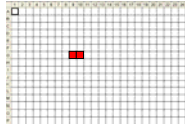
- ✓ "true peak" derived from a DNA insert digested band; BAC fingerprint
- ✓ low signal peak produced by the machine;
- ✓ partial digestion related peak;
- ✓ star activity by-product;
- ✓ *E. coli* genomic DNA band;

- ✓ vector band;
- ✓ out of size standard range band (with unreliable sizing);
- ✓ wide area peak (unreliable, resulting from co-migrating fragments).

(adapted from Scalabrin *et al.*, *BMC Bioinformatics*, 2009)

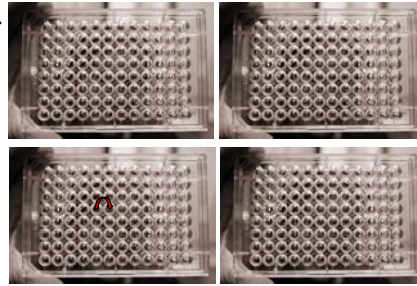
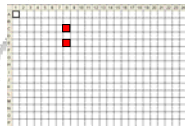
Well-to-well contamination removal using Genoprofiler

- ✓ Well-to-well contamination in **384-well plate** format
→ Adjacent wells showing similar profiles



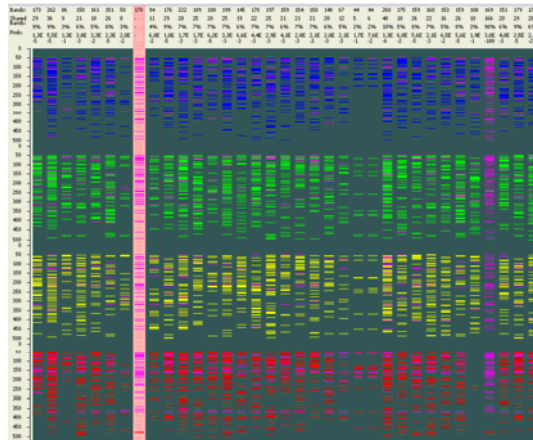
Splitting of 384-well plate into four 96-well plate during DNA extraction process.

- ✓ Well-to-well contamination in **96-well plate** format
→ Non-adjacent wells showing similar profiles



'One-to-one' contamination

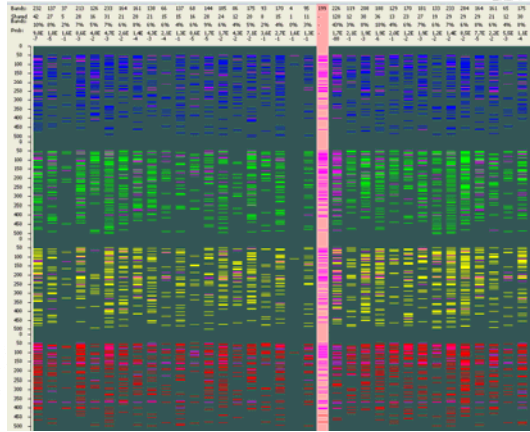
Two adjacent wells contain the same clone B1



80-100% identity of fingerprints

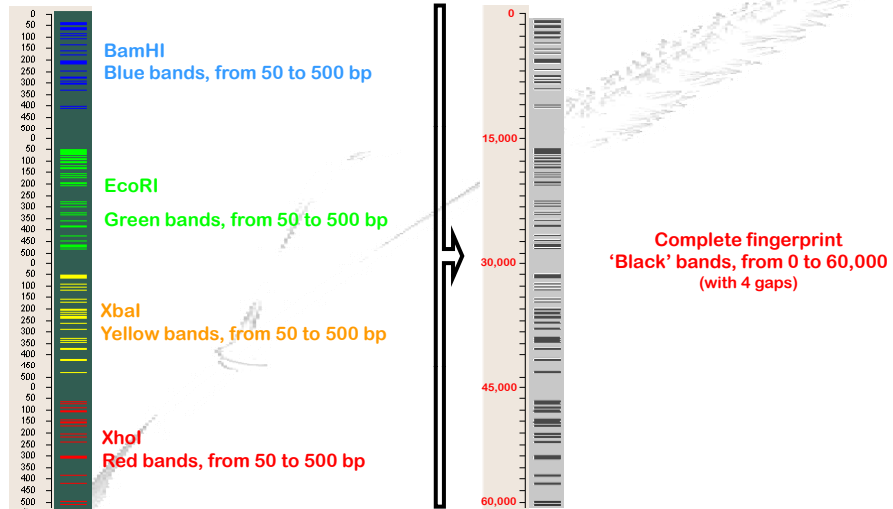
'One-to-two' contamination

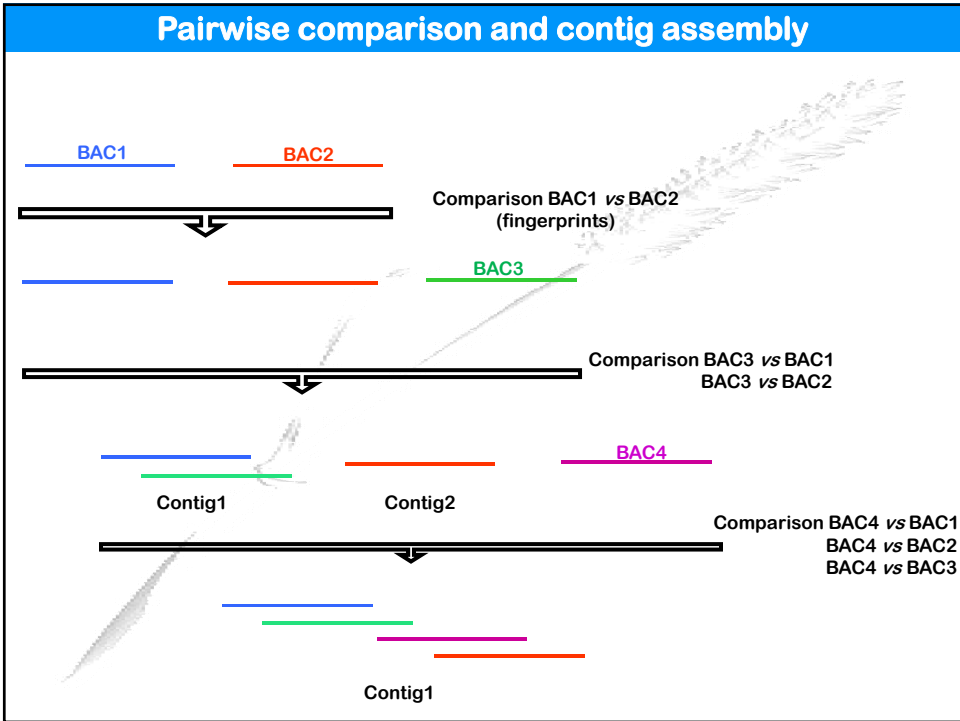
One well contains one clone B1 and the adjacent one contains the same clone B1 and another one B2

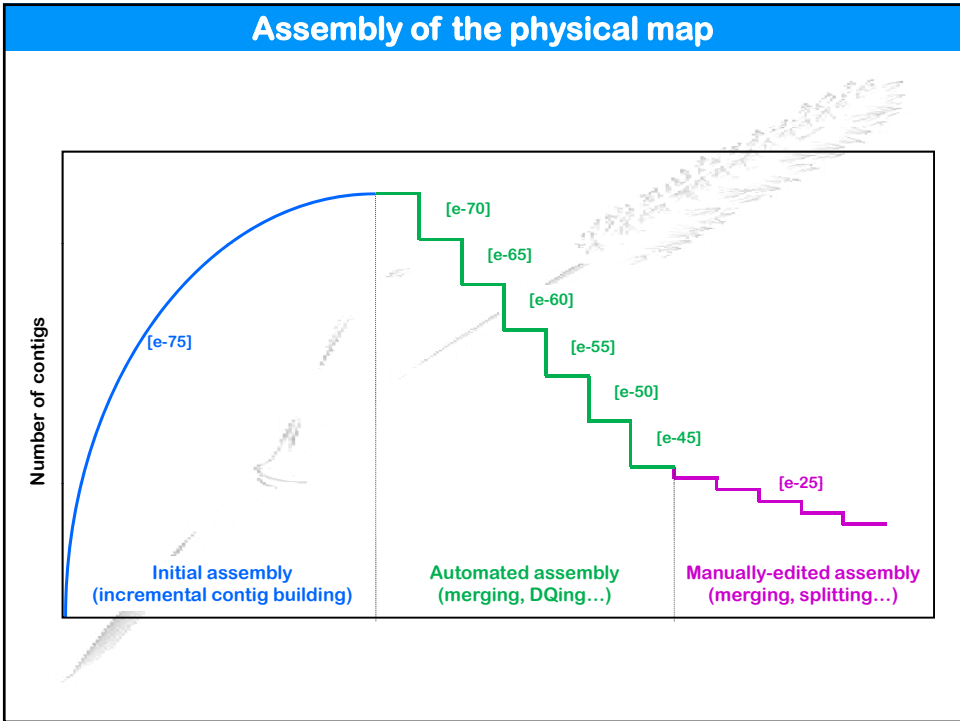
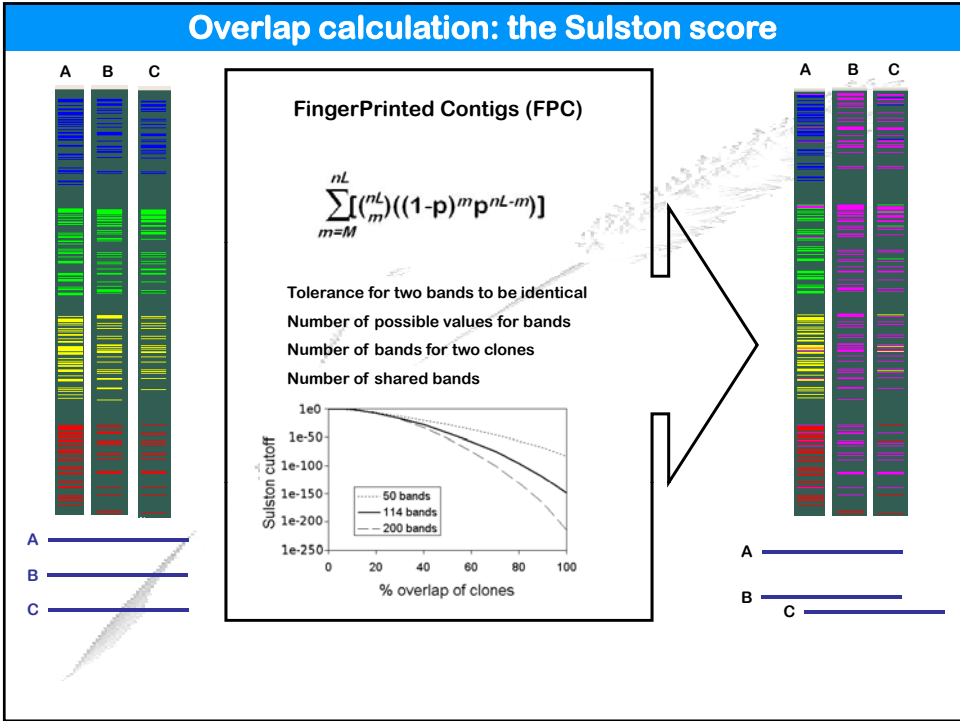


35-50% identity of fingerprints:
one of the well displays two merged fingerprints

Multiplication factor & color shift



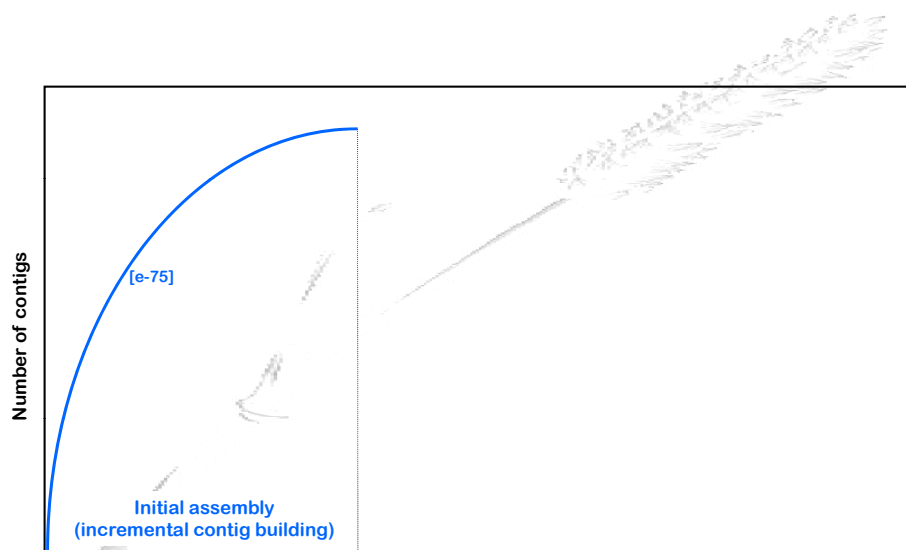


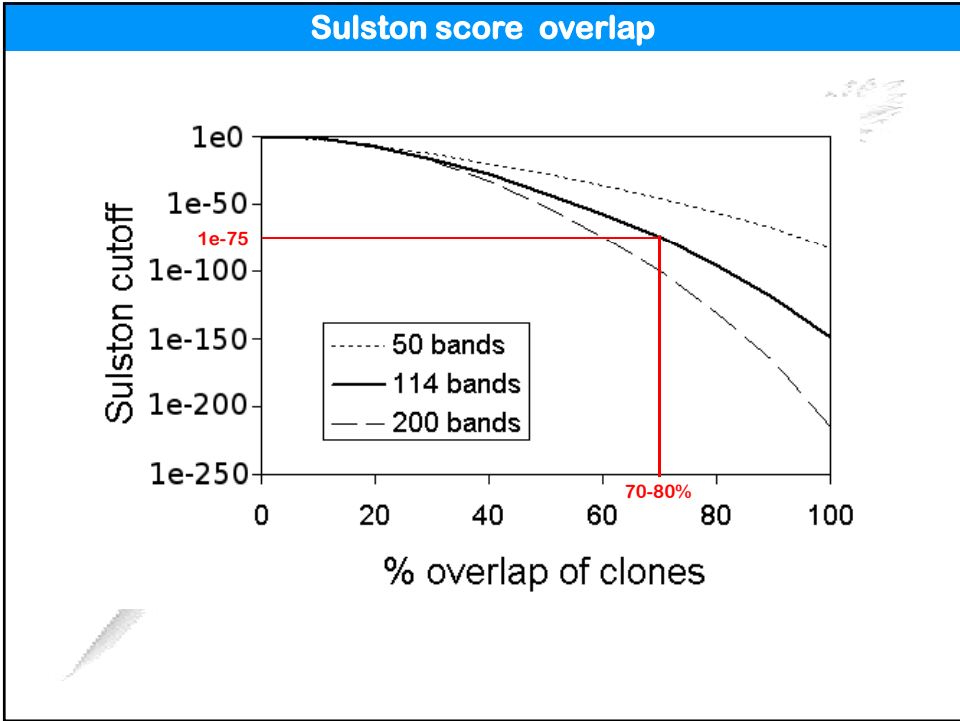


Contig assembly 2- Initial assembly



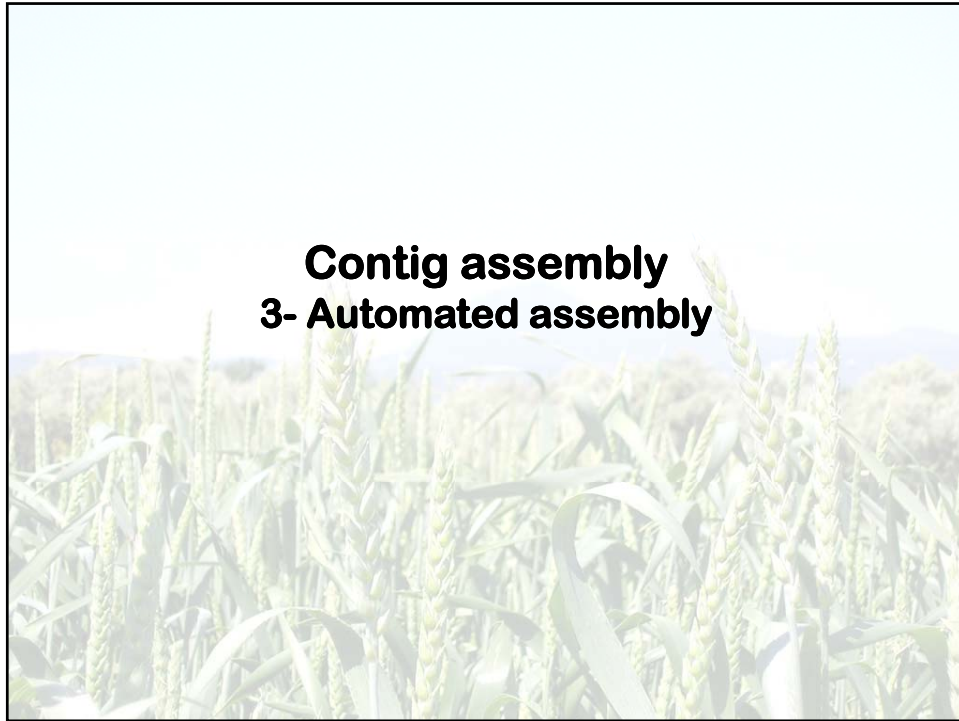
Assembly of the physical map



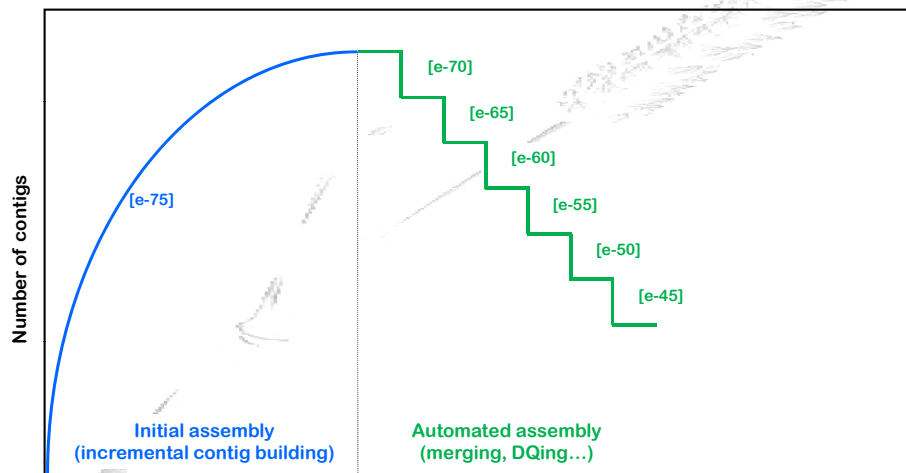


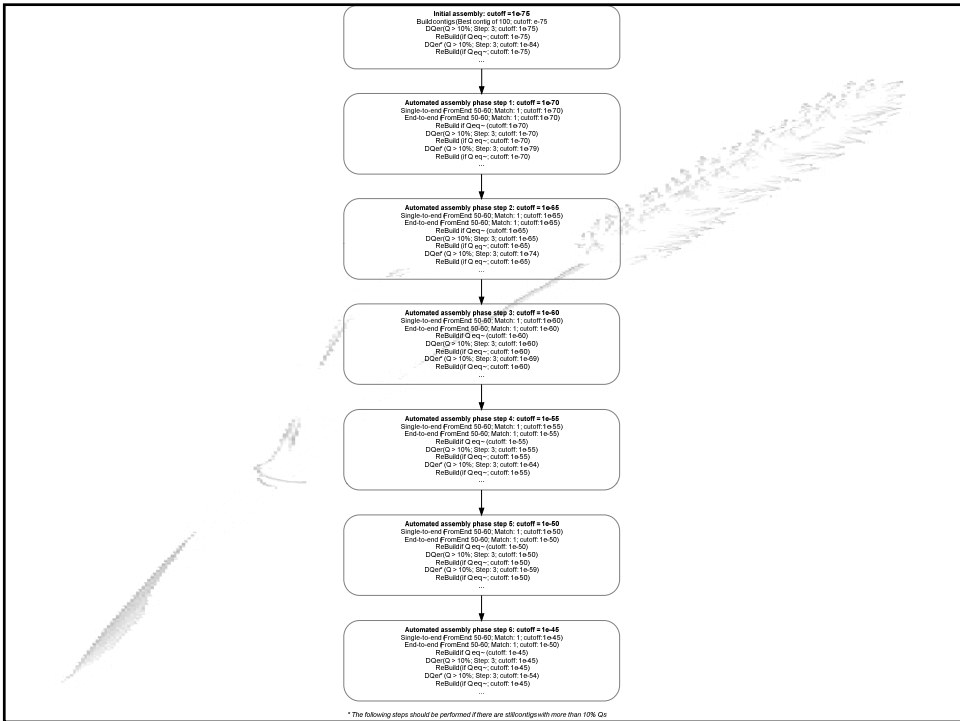
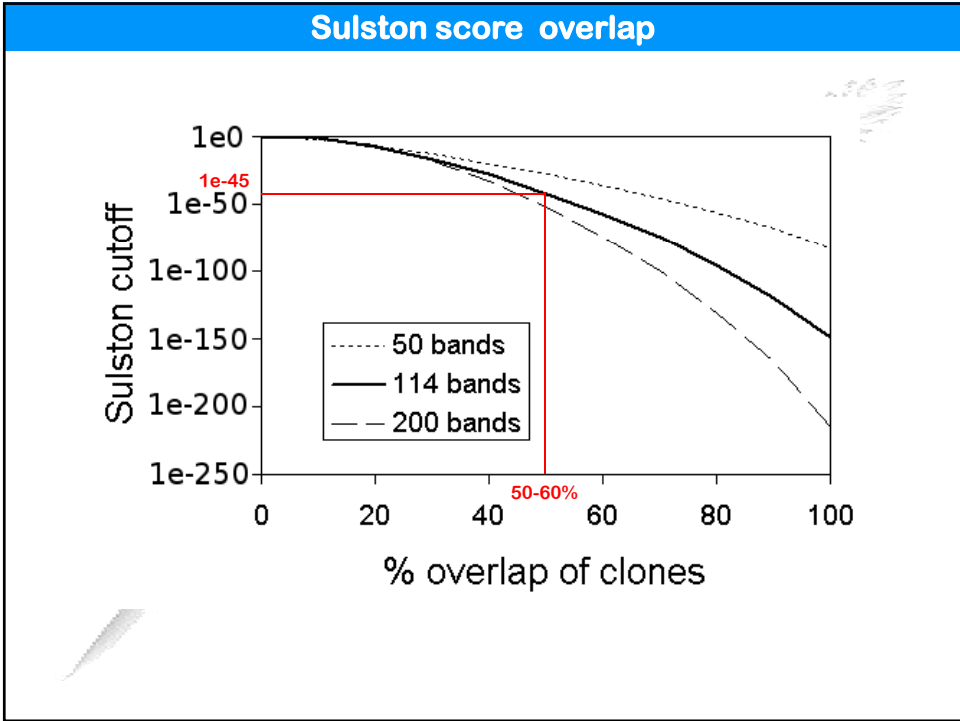
Contig assembly

3- Automated assembly



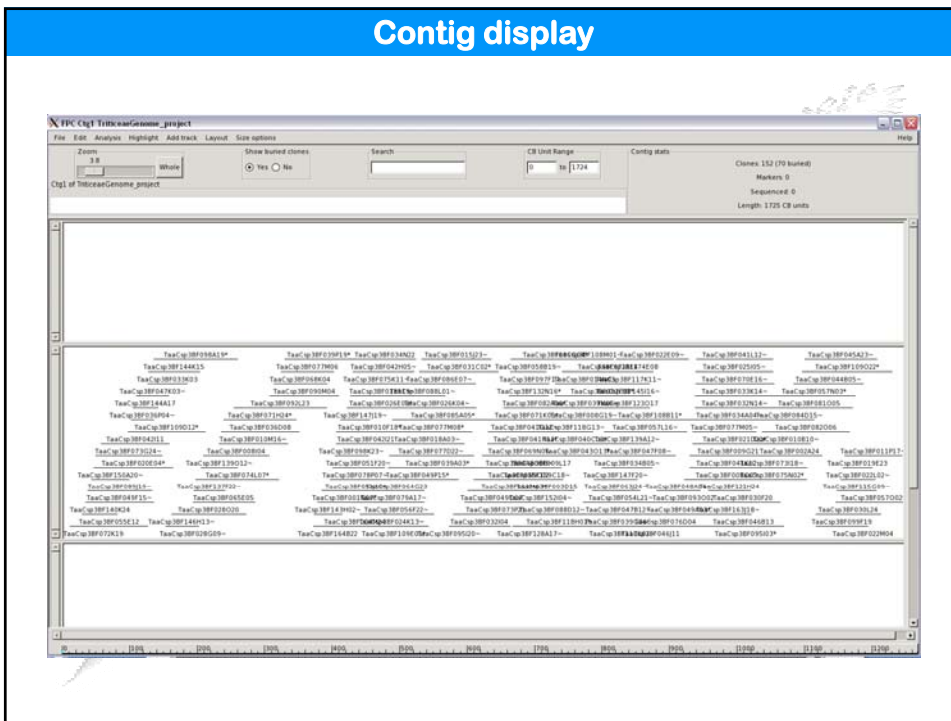
Assembly of the physical map



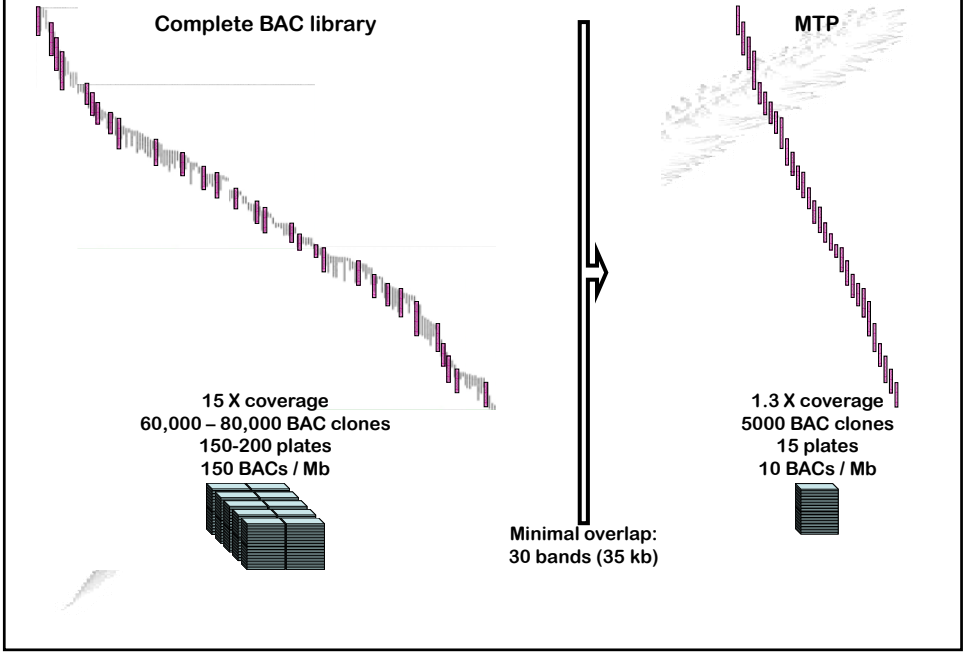


Contig assembly

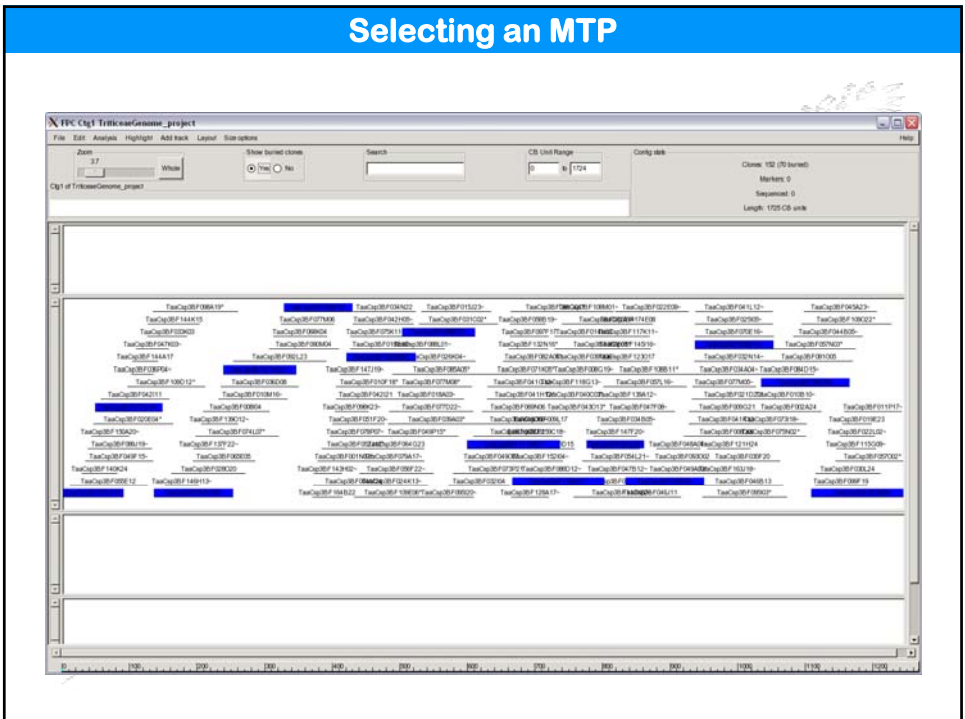
4- Establishing a Minimal Tiling Path (MTP)



Establishment of the Minimal Tiling Path

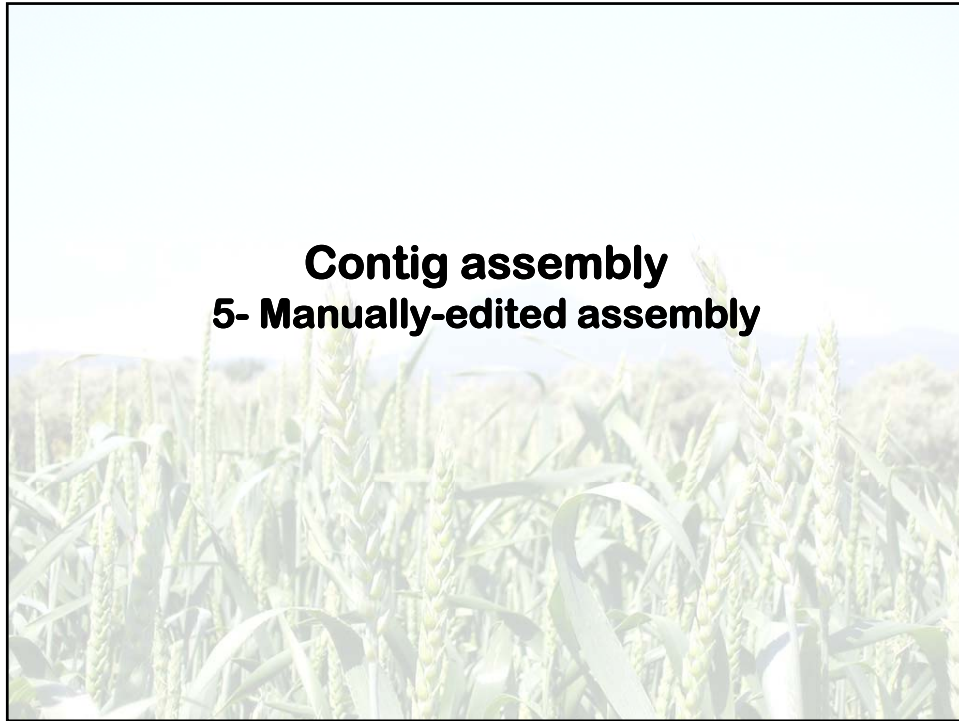


Selecting an MTP

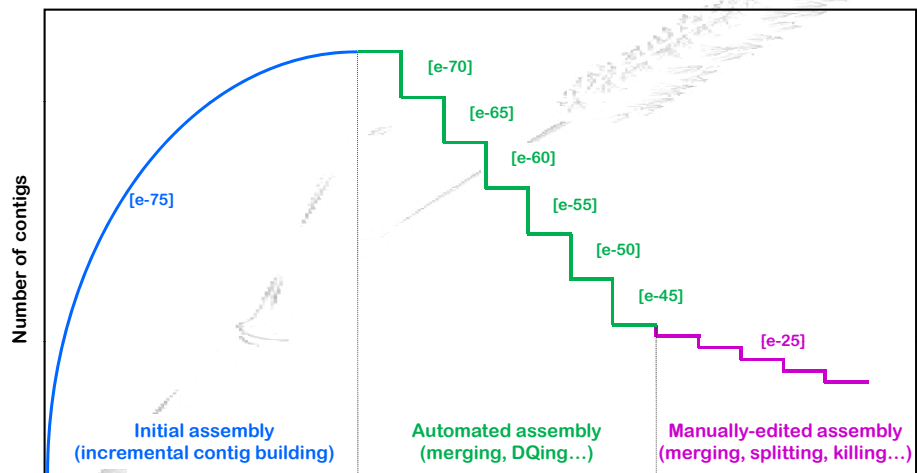


Contig assembly

5- Manually-edited assembly

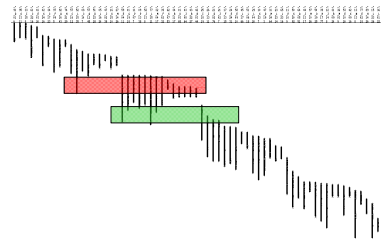
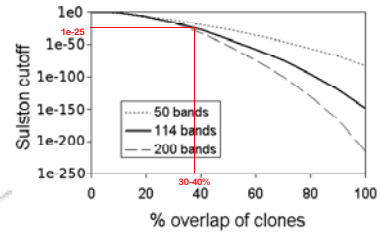


Assembly of the physical map



Manually-edited assembly

- ✓ Looking for **small reliable overlaps** (with or without marker data)
- ✓ Merging non-overlapping contigs based on **marker data**
- ✓ Checking **questionable** contigs (automatically or manually)
- ✓ Removing **problematic** clones
- ✓ Splitting **chimeric** contigs
- ✓ Killing **small** contigs (less than 6 clones or 300 kb)

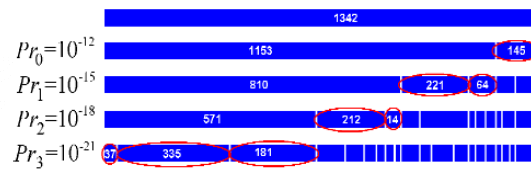


Contig assembly 6- LTC

LTC program

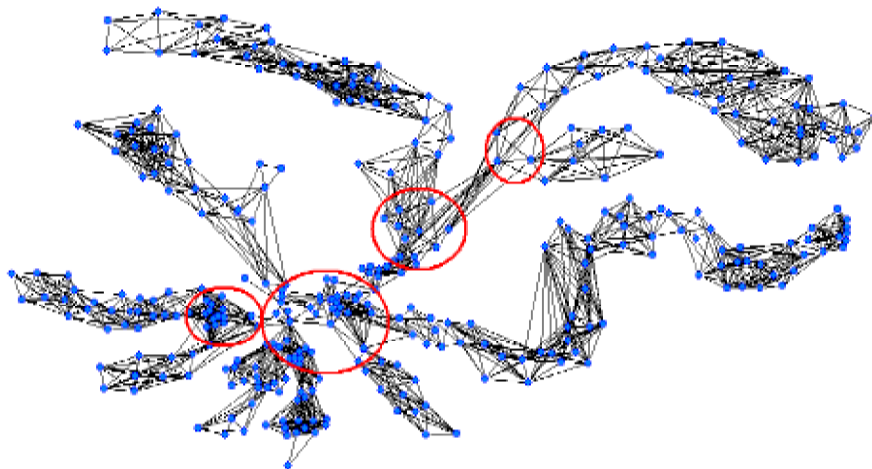
Frenkel *et al.*, (2010) LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. *BMC Bioinformatics*, 11, 584.

- ✓ LTC program starts clustering with a relatively relaxed cutoff and uses the topology of significant clone overlapping to obtain longer contigs with realistic (linear) structure.
- ✓ In each cluster, clones are ordered based on a global optimization procedure and clones that disturb the order stability (assessed by re-sampling analysis) are excluded from the contig.
- ✓ Ordered contigs are then merged upon a relaxed cutoff into longer contigs using for control of the contig topology the network representation of the significant clone overlaps.

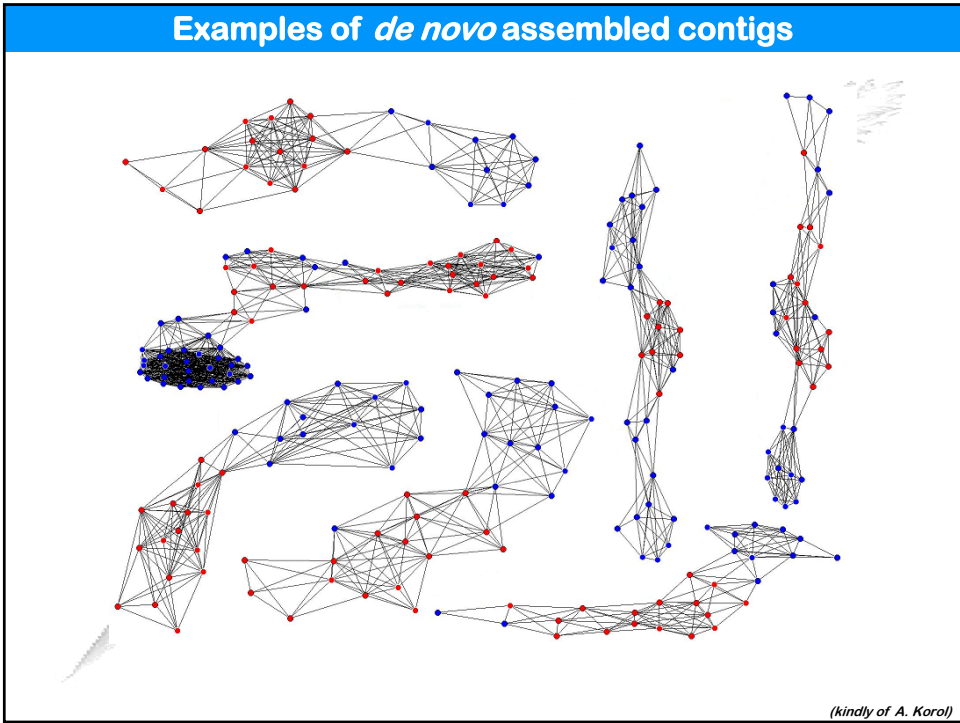
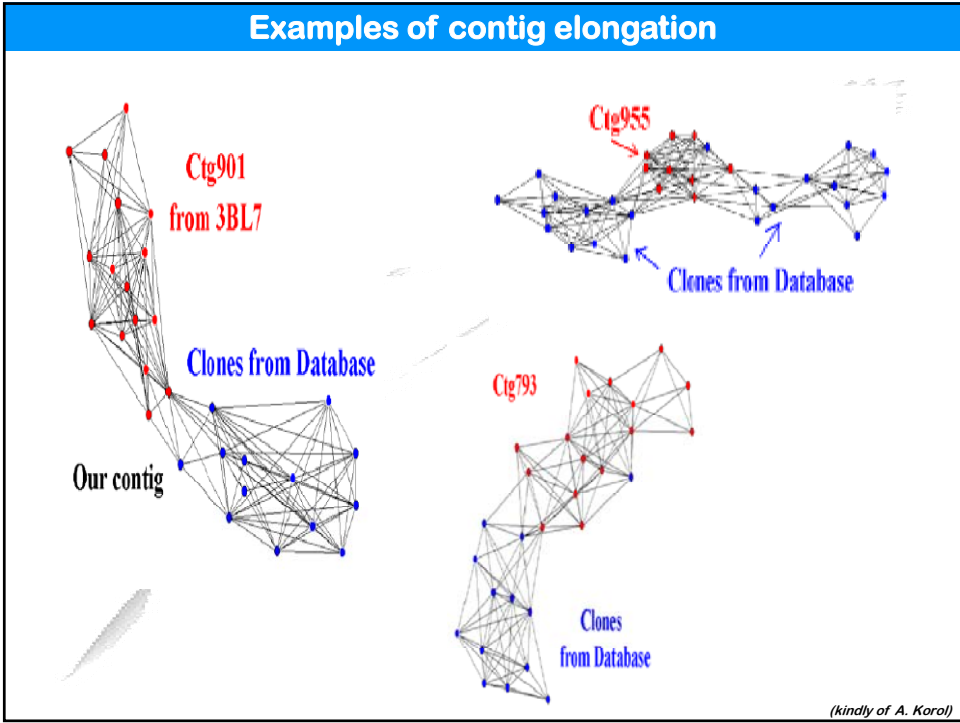


(adapted from Frenkel *et al.*, *BMC Bioinformatics*, 2010)

“Linearization” by removing clones in cluster branching



(kindly of A. Korol)



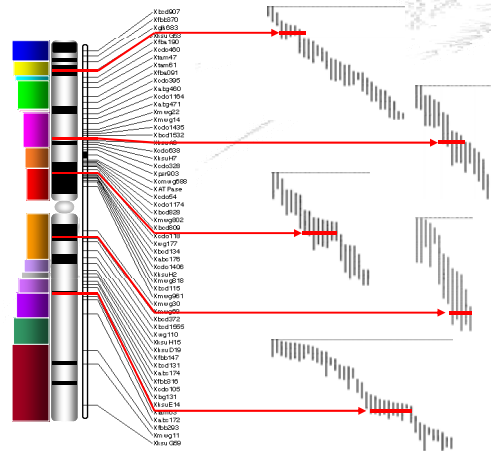
Contig anchoring

1- Forward anchoring

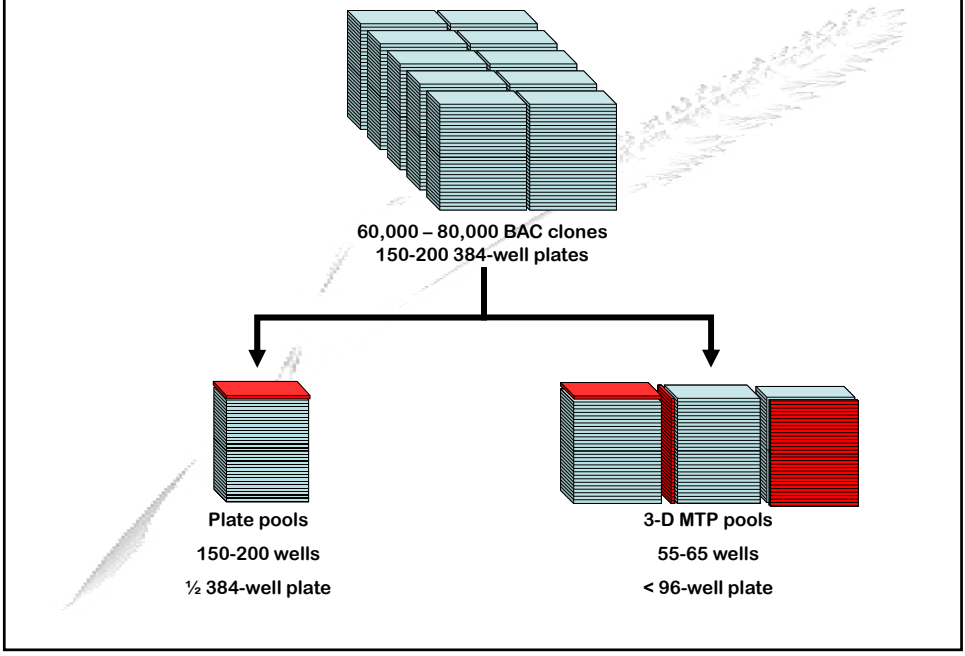


Forward contig anchoring

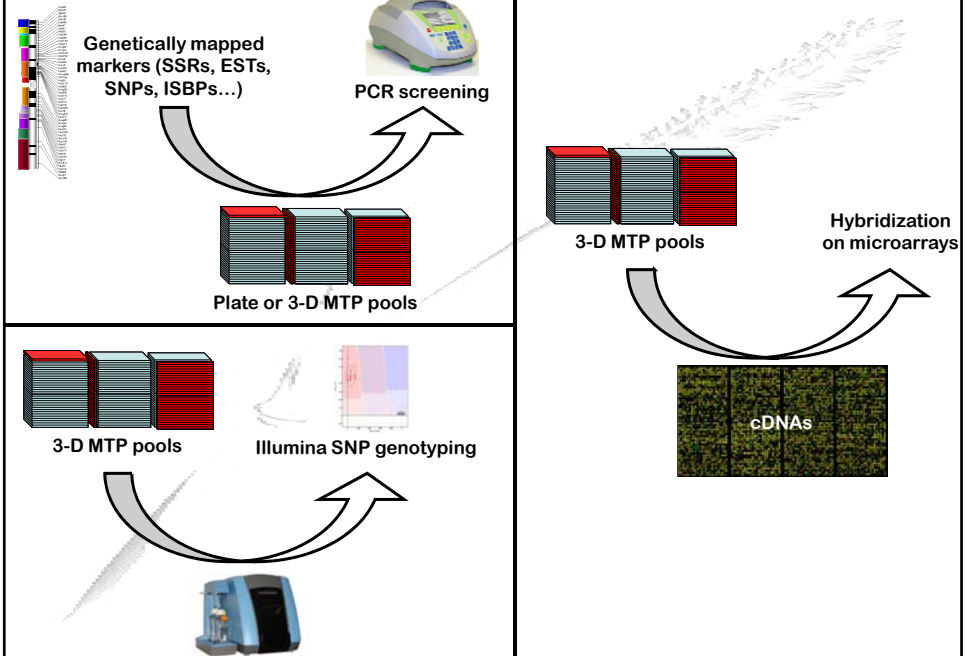
Forward anchoring: from genetic maps to contigs using genetically-mapped markers (SSRs, ESTs, RFLPs, DArTs, SNPs...)



Resources for forward anchoring

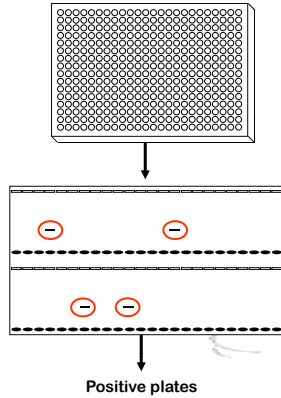


Examples of methods for pool screening

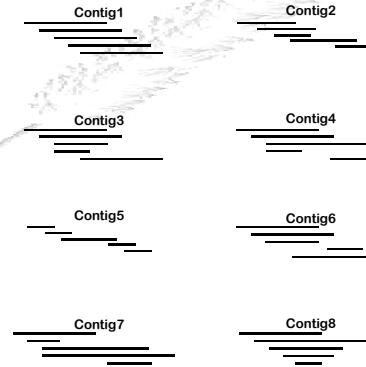


in silico physical map anchoring through pool screening

PCR with markers on plate pools



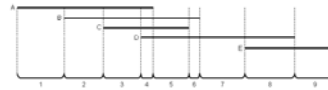
Physical contigs generated by FPC



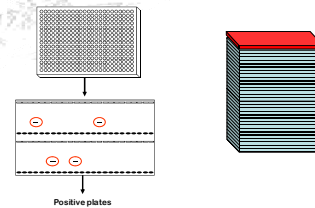
Identification of BAC addresses

elephant: electronic physical map anchoring tool

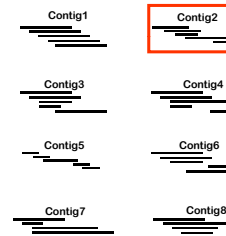
✓ Using the FPC file, *elephant* partitions the contigs into short segments, by splitting the contig at each branching point (e.g. a clone finishing or joining the assembly) and establishes a list of clones for each segment. Each segment contains a list of overlapping BAC clones that are different from the contiguous ones.



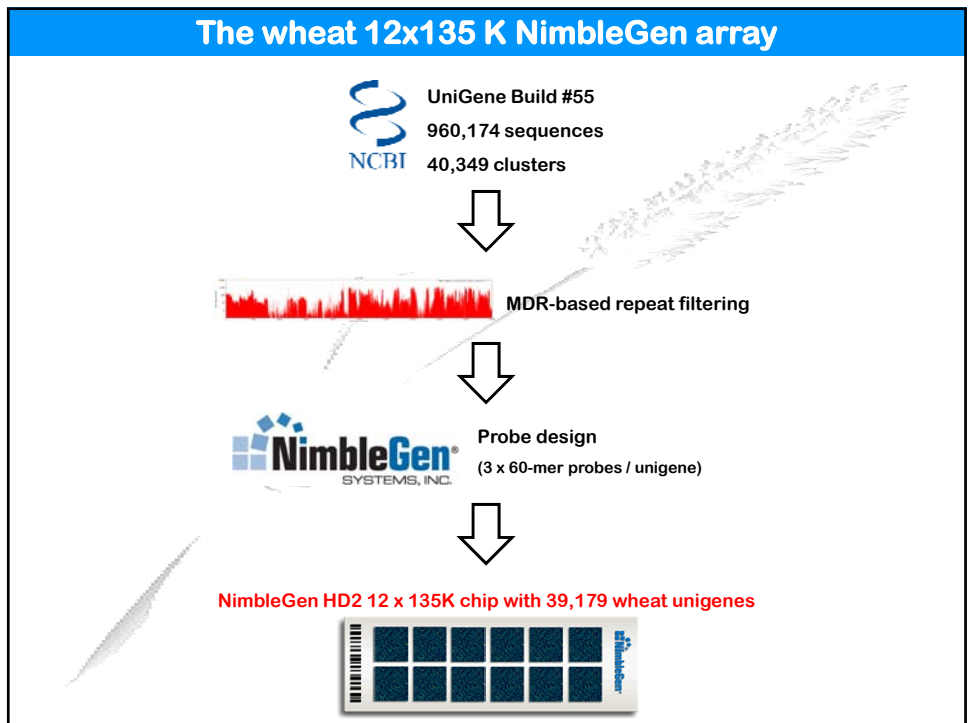
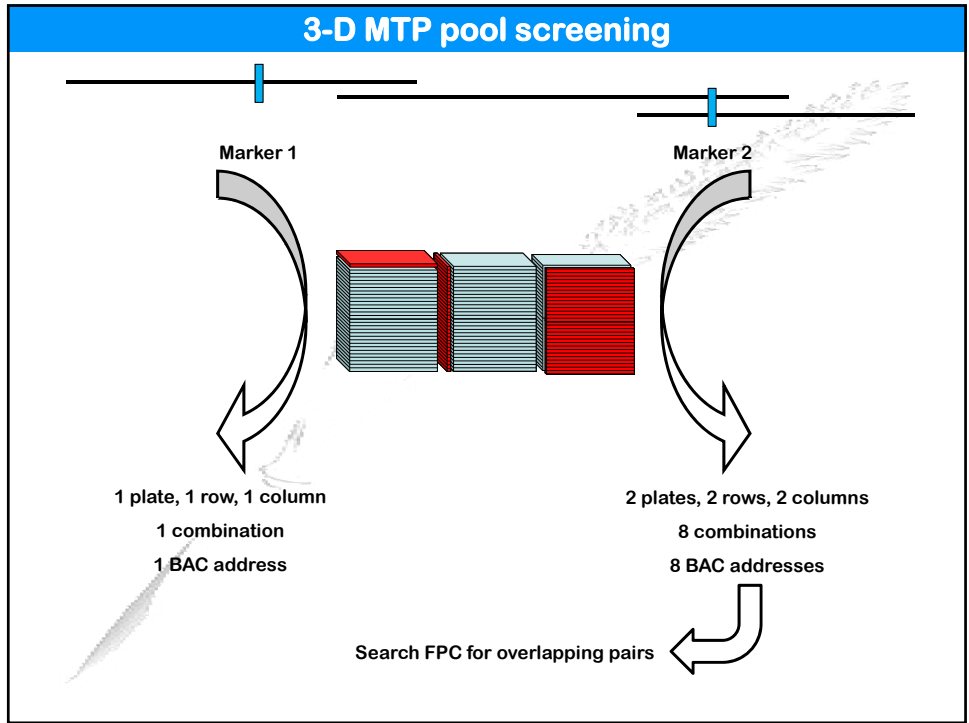
✓ For each marker, *elephant* combines the results from pool screening with the pool composition to establish a list of candidate clones harbouring the markers.



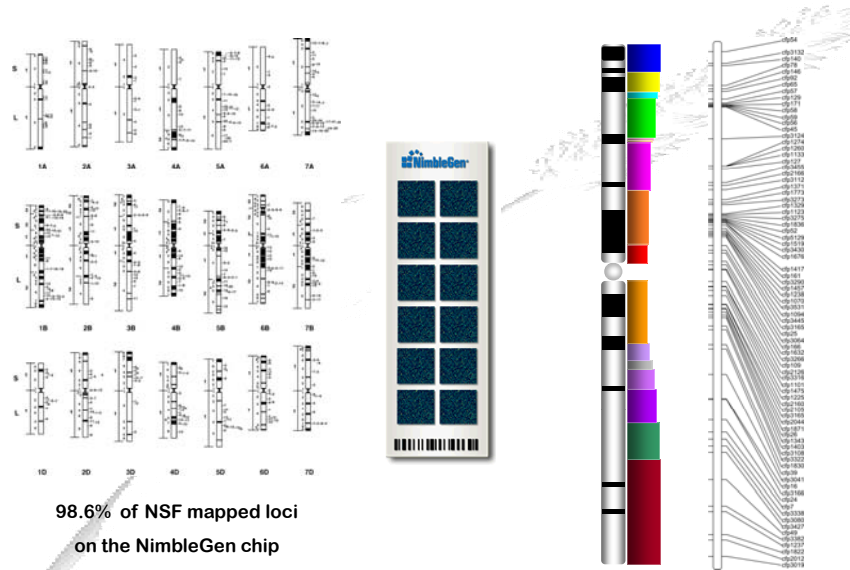
✓ Then, *elephant* compares the two lists and scores each segment using positive clones, missing clones and negative clones. All segments with a score above a given threshold are selected as candidates.



✓ The output text file reports the candidate contig(s) for each marker.

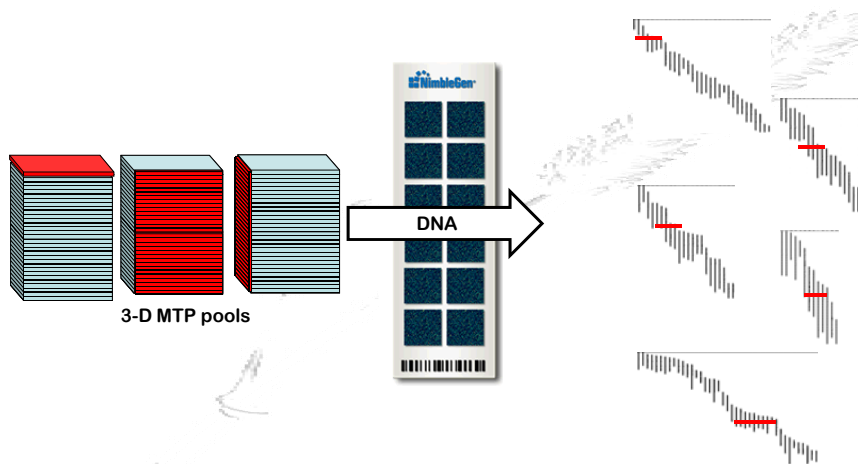


Assigning UniGenes to deletion bins



Genetic mapping through COS or SNP markers?

Chip-based anchoring



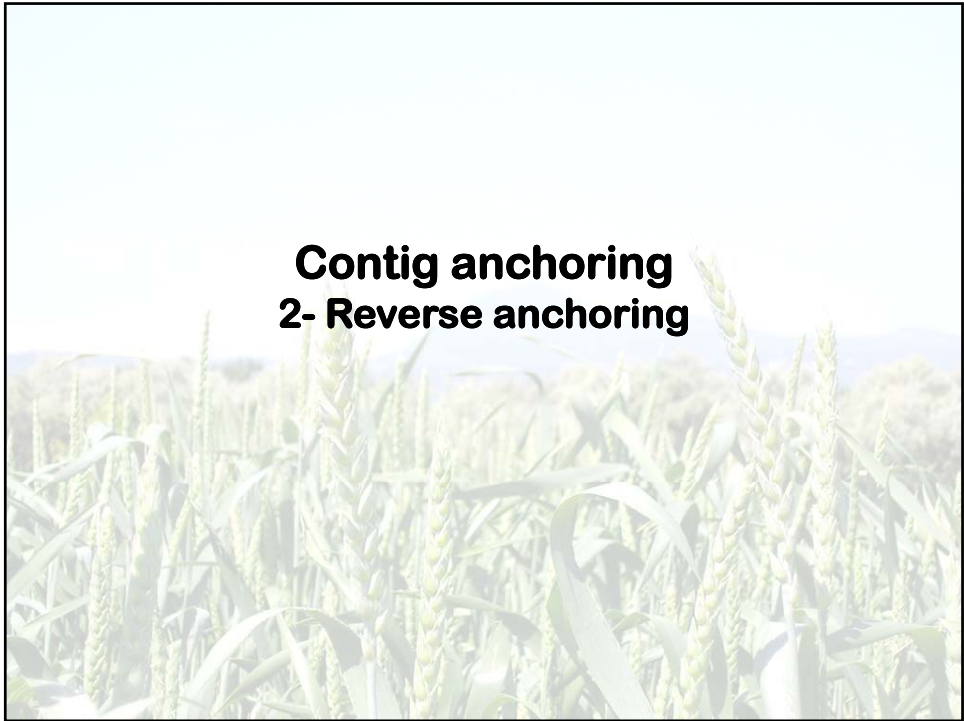
2836 unigenes mapped to chromosome 3B

Chromosome 1BL underway

Chromosomes 1AS and 1BS to come

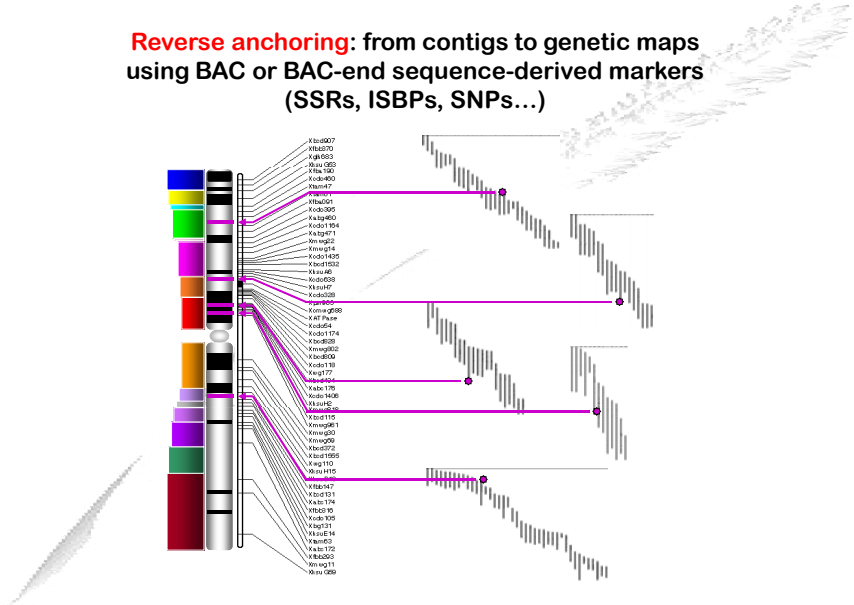
Contig anchoring

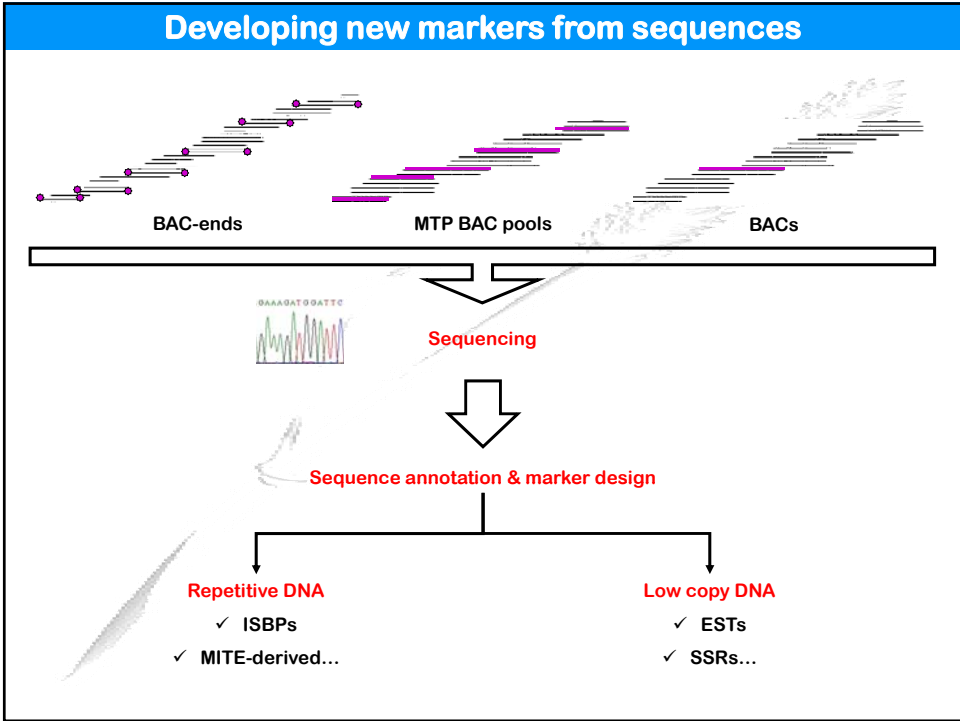
2- Reverse anchoring



Reverse contig anchoring

Reverse anchoring: from contigs to genetic maps using BAC or BAC-end sequence-derived markers (SSRs, ISBPs, SNPs...)

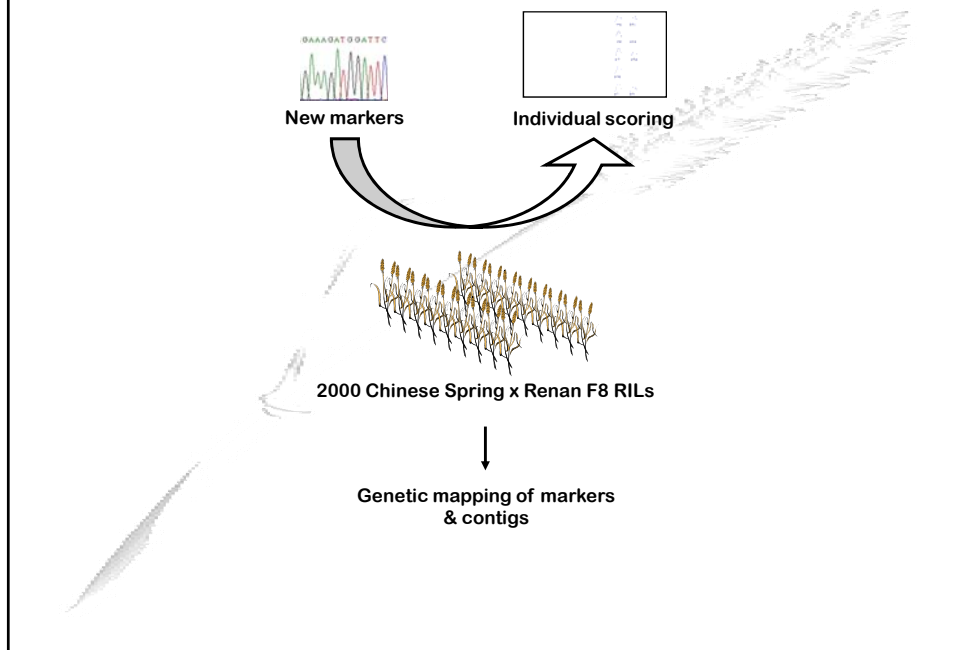




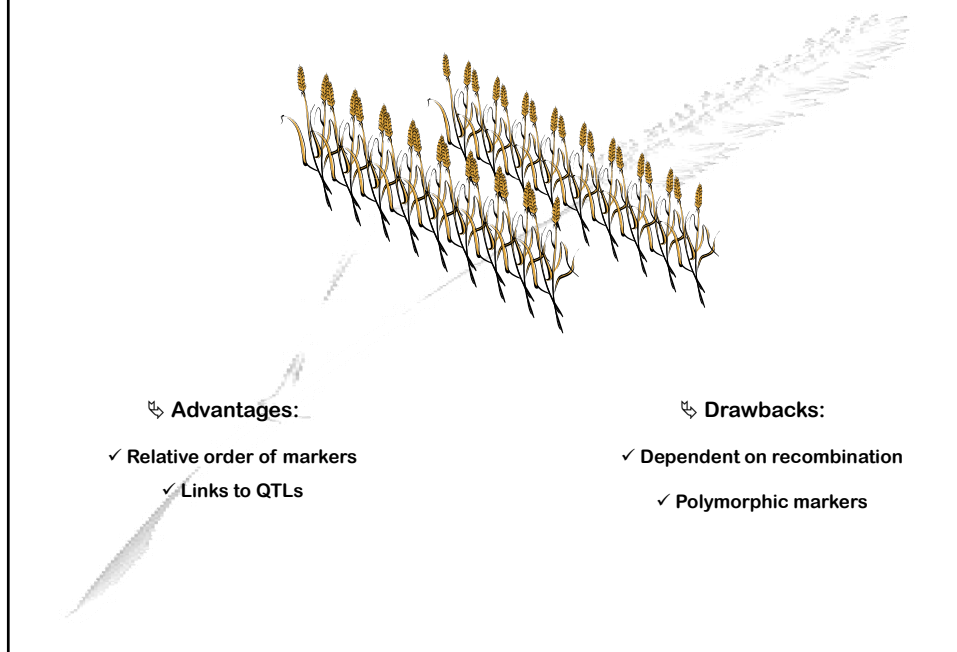
Examples of methods for genetic mapping

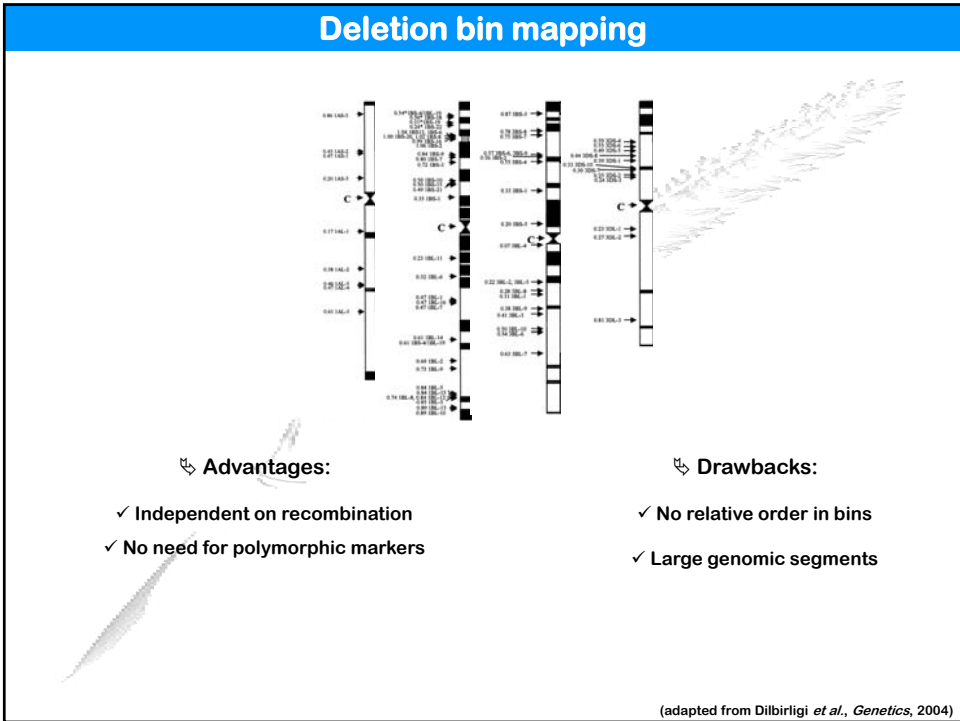
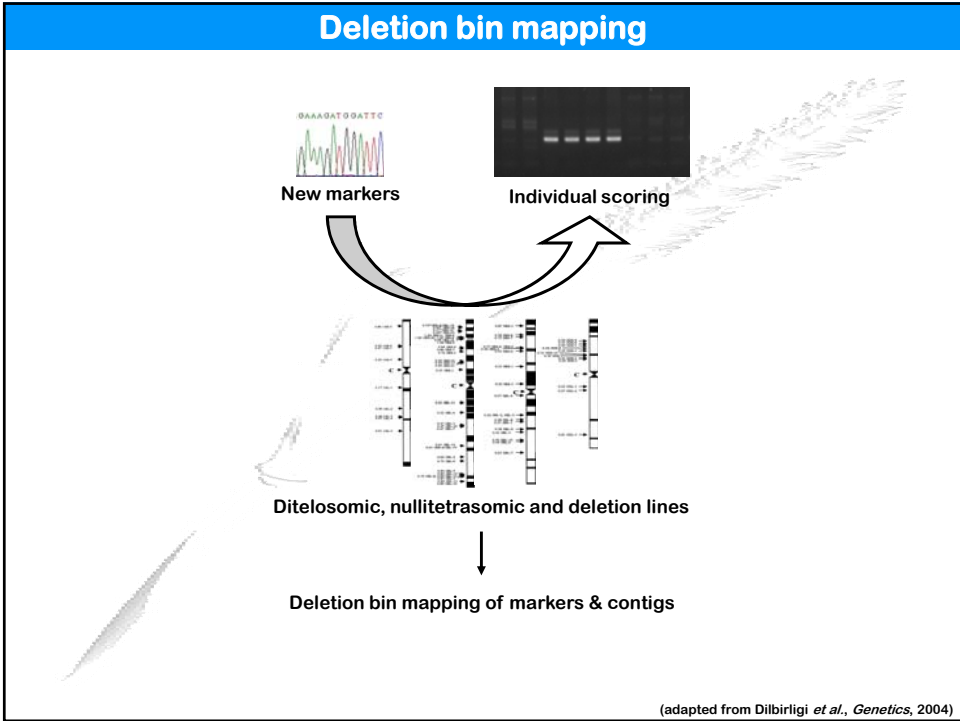
- ✓ Agarose gel electrophoresis
- ✓ Melting curve analysis
- ✓ Temperature gradient capillary electrophoresis
- ✓ Capillary electrophoresis
- ✓ Direct allelic sequencing
- ✓ Allele-specific PCR
- ✓ SNaPshot
- ✓ SNPLex
- ✓ Illumina
- ✓ KASPar
- ✓ Hybridization on microarray ...

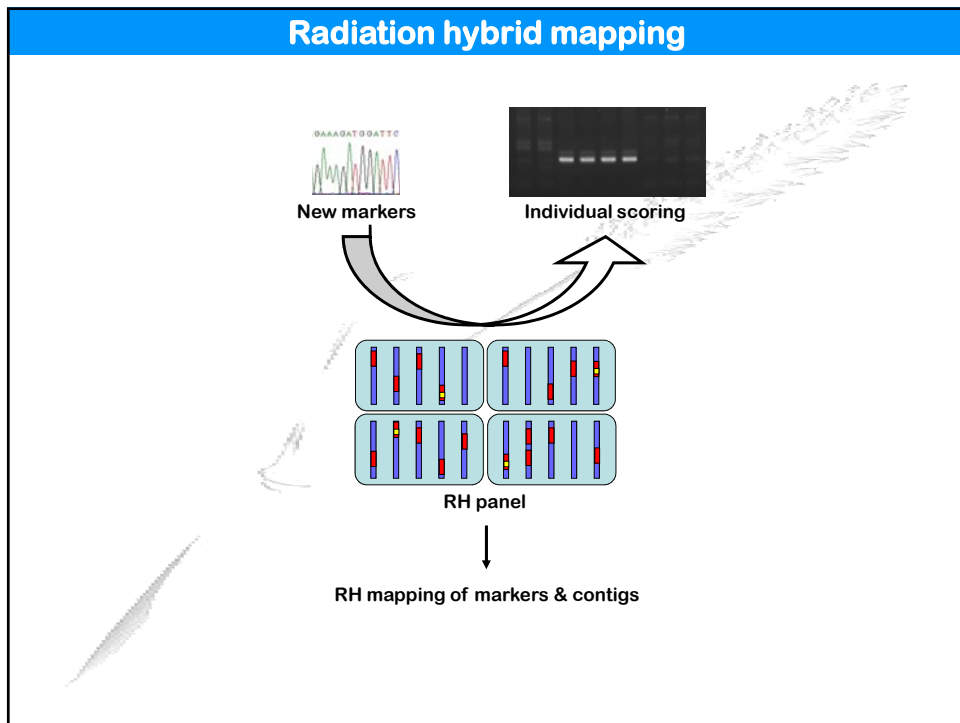
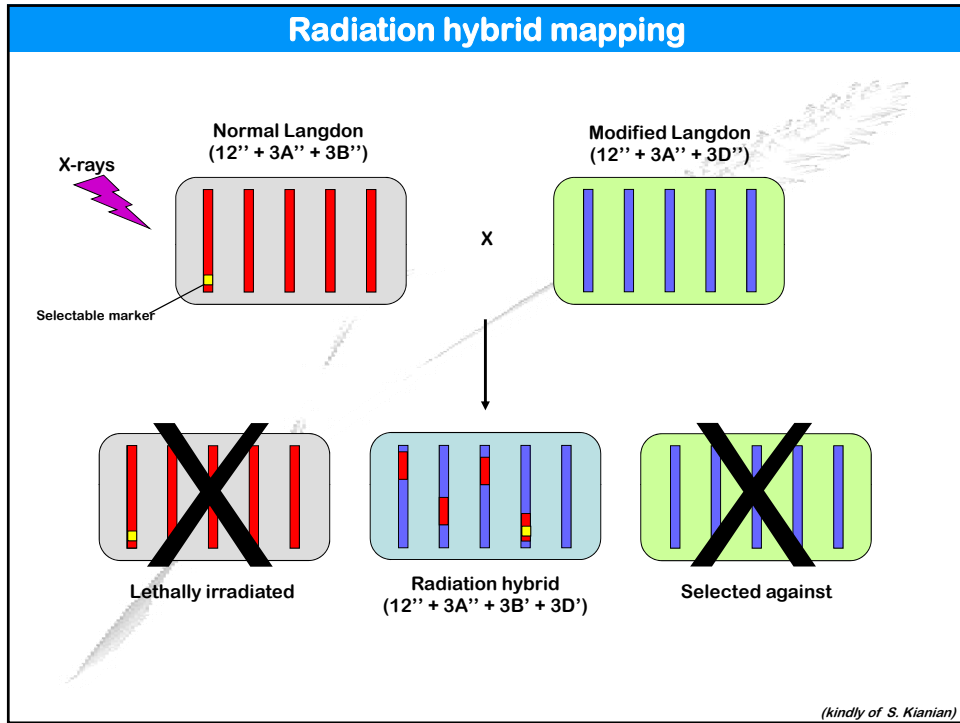
Recombination-based mapping



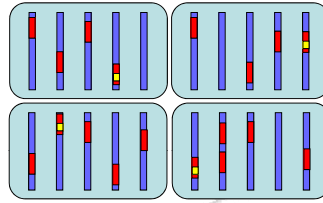
Recombination-based mapping







Radiation hybrid mapping



Advantages:

- ✓ Independent on recombination
- ✓ No need for polymorphic markers
- ✓ Resolution compatible with marker ordering (300 kb)
- ✓ No need for large population

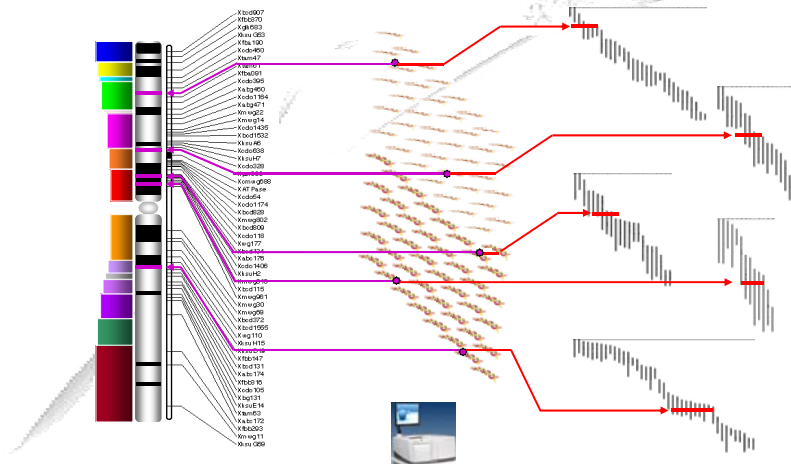
Drawbacks:

- ✓ Few results on wheat
- ✓ Tricky to develop RH panel

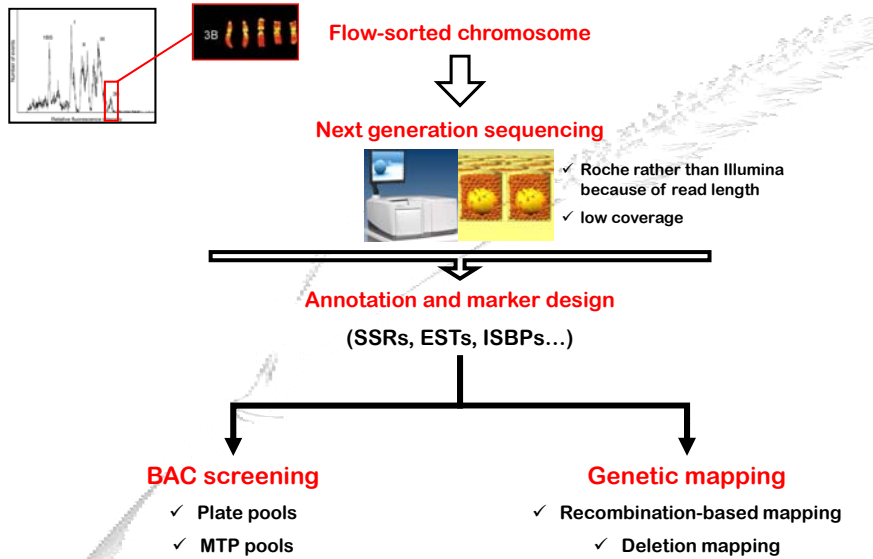
Contig anchoring 3- Hybrid anchoring

Third approach for physical contig anchoring

Hybrid anchoring: from random sorted chromosome shotgun sequences to genetic maps and contigs using sequence-derived markers (ISBPs, SSRs, ESTs, SNPs...)



Hybrid anchoring workflow



(adapted from Janda *et al.*, *Genetics*, 2006)

Standard protocols & guidelines

ANNUAL WHEAT NEWSLETTER VOICE

IWGSC: PHYSICAL MAPPING STANDARD PROTOCOLS WORKSHOP

Plant and Animal Genome Meeting, San Diego, CA, USA
Tuesday, 12 January, 2010.

Workshop report.

Rudi Appels, Michael Akmal, Maria Casanova, Patricia Cattivani, Jennifer Deland, David Eder, Wang Chang-Lin, Claire Matthews, Nicholas Oubadi, Shihong Peng, Thomas Weber, Kjetil Isenhardt, and Catherine Feuillet.

On 12 January 2010, the International Wheat Genome Sequencing Consortium (IWGSC) organized a workshop to develop and discuss protocols and standards for the physical mapping of the hexaploid wheat genome and develop a consensus. In addition, the workshop surveyed the sequencing efforts undertaken within the consortium to coordinate the studies carried out in member laboratories. The goal was to ensure homogeneity in the procedures used for constructing the physical maps by providing guidelines developed in expert laboratories and distributing these to the groups participating in the physical mapping and sequencing of wheat wheat chromosomes under the auspices of the IWGSC.

The first step for selecting a high quality reference sequence of the bread wheat genome established by the IWGSC includes, as a first step, the construction of physical maps to be used as a chromosome specific strategy. This approach relies on repeat arrangements in chromosome walking and BAC library construction technology that have been discussed.

Physical mapping of the bread wheat genome

Figure 1 shows the status of the IWGSC physical mapping projects as of January 2010. The figure is divided into three columns: 'Proposed', 'Active', and 'Finished'. Each column contains a list of chromosome arms (e.g., 1A, 1B, 1D, 2A, 2B, 2D, 3A, 3B, 3D, 4A, 4B, 4D, 5A, 5B, 5D, 6A, 6B, 6D, 7A, 7B, 7D, 8A, 8B, 8D, 9A, 9B, 9D, 10A, 10B, 10D, 11A, 11B, 11D) with corresponding project names and dates. A legend indicates the status of each project: 'Proposed' (green), 'Active' (yellow), and 'Finished' (red).

The workshop was organized in five sessions of 45 minutes each that covered the following topics:

- Fluorescence (Chang-Lin Wang and Patricia Cattivani)

International Wheat Genome Sequencing Consortium

Home and Reports | Organization | Projects | Tools and Resources | General Documents

Home | News and Reports | Meetings and Workshops | Physical mapping standard protocol workshop

Physical mapping standard protocol workshop

Published: 04/16/2010 04:04 pm. Last modified: 07/05/2010 09:13 pm

On 12 January 2010, the International Wheat Genome Sequencing Consortium (IWGSC) organized a workshop to develop and discuss protocols and standards for the physical mapping of the hexaploid wheat genome and develop a consensus. In addition, the workshop surveyed the sequencing efforts undertaken within the consortium to coordinate the studies carried out in member laboratories. The goal was to ensure homogeneity in the procedures used for constructing the wheat physical maps by providing guidelines developed in expert laboratories and distributing these to the groups participating in the physical mapping and sequencing of bread wheat chromosomes under the auspices of the IWGSC. A workshop report can be found in the Annual Wheat Newsletter 2010 (Vol. 56) at <http://www.wheatgenome.org/newsletters/aw56/TEXTFILES/IWGSC.pdf>

Workshop report

Protocols

- [Chromosome walking protocol](#) 50.02 KB
- [BAC Cloning and Sequencing protocol](#) 73.05 KB
- [Physical map assembly guidelines](#) 932.91 KB

Presentations

- [Cattivani Patricia Ph.D. 2010](#) 2.91 MB
- [Cattivani Patricia Ph.D. 2010](#) 430.09 KB
- [Cattivani Patricia Ph.D. 2010](#) 3.47 MB
- [Workshop Plan Abstracts Ph.D. 2010](#) 3.10 MB
- [Wang Chang-Lin Ph.D. 2010](#) 2.24 MB
- [Cattivani Patricia Ph.D. 2010](#) 1.45 MB
- [Cattivani Patricia Ph.D. 2010](#) 384.28 KB
- [Cattivani Patricia Ph.D. 2010](#) 1.13 MB

Annual Wheat Newsletter 2010
<http://wheat.pw.usda.gov/ggpages/awn/56/TEXTFILES/IWGSC.pdf>

IWGSC
<http://www.wheatgenome.org/News-and-Reports/Meetings-and-Workshops/Physical-mapping-standard-protocol-workshop>

Standard protocols & guidelines

Triticaceae Genome

Guidelines for physical map assembly

1. BAC naming convention:

In this section we remind what was established at the beginning of the project http://www.triticaceagenome.eu/pdf/Triticaceae_Genome_Report_2007.pdf

- IGA: Chromosome (Genus) Function: provides the BAC library
- IGA: CNRSDV in Truncates (Function) Barcode: the plate, makes copies and sends them to
- IGA as Update: ready for BAC fingerprinting
- IGA sends (or updates) the fingerprinting data to the Triticaceae Genome (TG) partners

BAC naming conventions are as described at <http://www.triticaceagenome.eu/faq/faqBAC>

Clicking on one of the "Triticaceae Genome" libraries, e.g. <http://www.triticaceagenome.eu/faq/faqBAC/TriticaceaeGenome/TriticaceaeGenome> provides you with examples of the naming convention that is used internationally and internally at CNRSDV (and some more information)

Example:

The international name is TricCapS02A for a single library (here, for a clone from the 3D5 BAC library).

Note: CNRSDV has used accidentally Taa as a prefix instead of Taa, as defined for the international convention. Unfortunately, that "typo" was propagated also at IGA, therefore names of barcodes, plates, and fingerprints contain Taa instead of Taa as a prefix for all the BAC libraries from the Triticaceae Genome project. Though, it is not a big problem, as explained below.

Thus, in the example, TricCapS02A, the nomenclature means:

- Tric: Triticum aestivum subsp. aestivum
- Cap: Chromosome Spring
- S02: S0 is the chromosome, L is the arm. There are two possible arms, L and S, standing for Long and Short. Note that chromosome 2B was fingerprinted entirely and is not split in the two arms.
- 2: it stands for the HindIII enzyme used for the library construction (it can be also a for EcoRI or a for BamHI, however these two enzymes are not used for library construction)
- A: is the library code. Normally it is A, standing for first library, it is possible to have B, for a second library. The only chromosome with two libraries, at the moment, is 2B.

Each clone is identified by the library name (with mislabeled Taa to Taa for the TG project) and by the plate number and well position.

Each library is sent from CNRSDV to IGA in a set of plates that are barcoded each with the library name followed by the number "0" (used as separator) and four digits identifying the plate number. E.g. plate number 23 of library TricCapS02A is TricCapS02A_0023 (the four digits are padded with zeros if the plate number contains less than 4 digits). If more than a library is needed for a chromosome then plate numbers are progressive in the libraries, therefore, the plate number is already sufficient to uniquely identify a plate inside a library (e.g. There was only one plate 16 for library A and B of chromosome 3B).

Inside each plate a clone is identified by its well position. E.g. clone A01 of plate 23 of library TricCapS02A_0023 is labeled as TricCapS02A_0023A01.

Triticaceae Genome
<http://www.triticaceagenome.eu/page.php?optim=Deliverables>

The Physical MAP (Minimum Anchor Points) standard

✓ Type of markers to be used (avoid RFLPs, DArTs...)

- ❖ SSRs
- ❖ ISBPs
- ❖ ESTs

→ Complementarity of markers

✓ Number of markers per chromosome arm

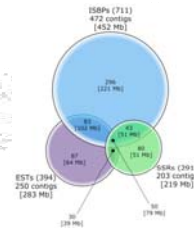
- ❖ in contigs: 1000 (SSRs, ISBPs, and ESTs)
- ❖ in deletion bins: 500 (SSRs, ISBPs, and some ESTs)
- ❖ on genetic maps: 150 (SSRs, ISBPs, and some ESTs)

→ 75-100% of contigs should be anchored in deletion bins

✓ Strategies

- ❖ PCR screening for SSRs, ISBPs or ESTs: 300-500 each
- ❖ Microarray screening for ESTs: 500-1500
- ❖ Microarray screening for ISBPs (to be validated): 2000-5000

To be discussed...



The Physical MAP (Minimum Anchor Points) standard

✓ Types of markers to be used (avoid RFLPs, DArTs...)

- ❖ SSRs: for genetic mapping and contig / QTL links
- ❖ ISBPs: for genetic and deletion bin mapping
- ❖ ESTs: for comparative genomics including between homoeologous chromosomes

→ Complementarity of markers

✓ PCR-based strategy for 'backbone' anchoring

- ❖ SSRs: all publicly available markers
- ❖ ISBPs: ~500 from BES or sorted chromosome shotgun sequences
- ❖ ESTs: ~500 (from the 10,000 whole genome COS markers)

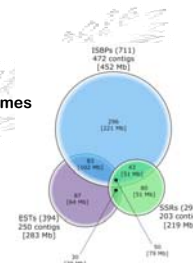
→ 75% of the contigs mapped to deletion bins

✓ Additional high-throughput strategies for physical map saturation

- ❖ Microarray screening for ESTs: 500-1500
- ❖ Microarray screening for ISBPs (to be validated): 2000-5000

→ 100% of the contigs mapped to deletion bins with several markers per contig

To be discussed...



IWGSC Workshop
April 8, 2011
Prague, Czech Republic

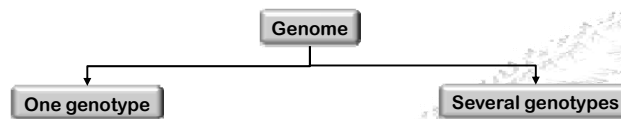
Sequence-enabled marker development in hexaploid wheat

Etienne Paux

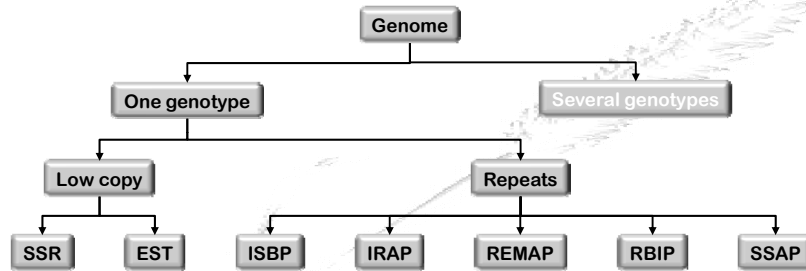
Structure, Function & Evolution of the Wheat Genomes
Genetics, Diversity & Ecophysiology of Cereals
INRA Clermont-Ferrand, France



A decision tree for marker development in wheat



A decision tree for marker development in wheat



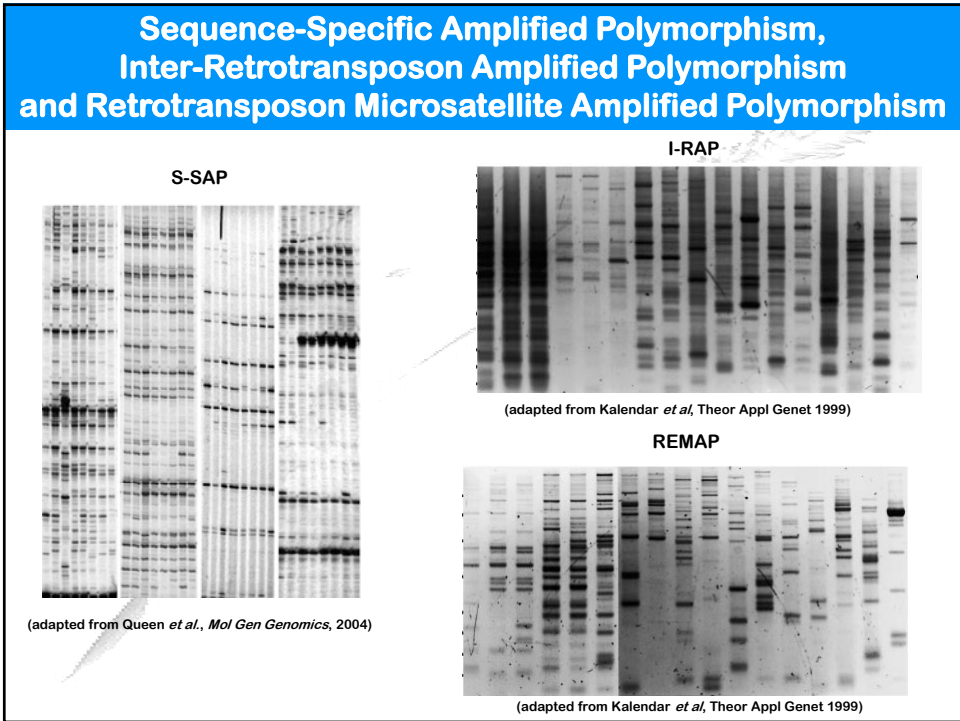
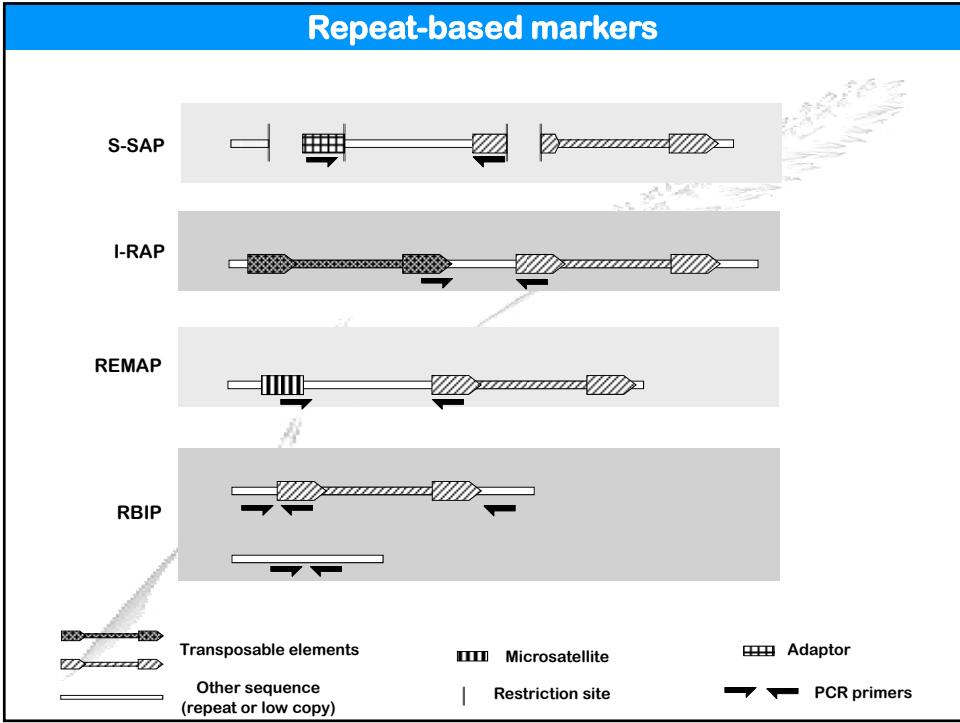
SSR and EST markers

➤ SSRs

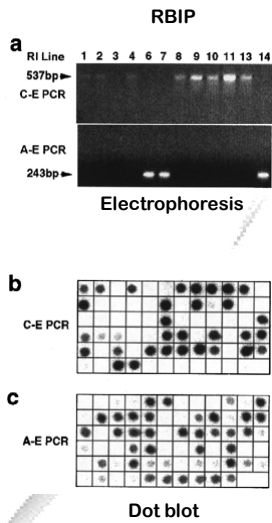
- ✓ Two categories
 - ❖ Genomic DNA → genomic SSRs (gSSRs)
 - ❖ ESTs → EST-SSRs
- ✓ Origin: Transcriptome, enriched libraries, BES or whole genome shotgun
- ✓ Detection tools: SciRoKo, MISA, SSRFinder, SSRIT, TRF, TROLL, Sputnik, SSRsearch, SSR Locator, SSRPrimer...
- ✓ Detection technique: capillary electrophoresis

➤ ESTs

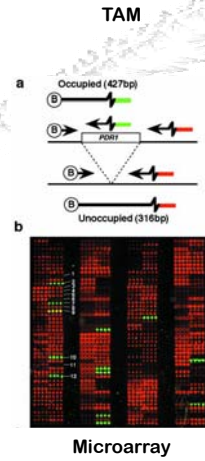
- ✓ More than one million wheat ESTs in databases
- ✓ Chips (Affymetrix, Agilent, NimbleGen)
- ✓ Detection techniques: HRM, SSCP, SFP
- ✓ Conserved Orthologous Set markers (presence of an intron to increase the polymorphism level)



Retrotransposon-Based Insertional Polymorphism and Tagged Microarray Markers



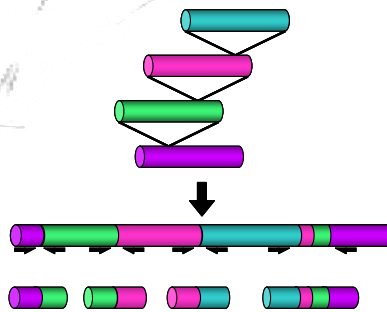
(adapted from Flavell *et al.*, Plant J 1998)



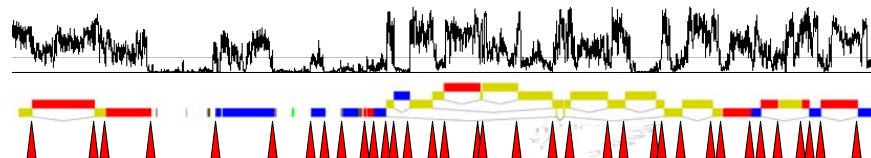
(adapted from Flavell *et al.*, Nucleic Acids Res 2003)

Insertion Site-Based Polymorphism markers

- ✓ based on polymorphism of **TE insertion site**
- ✓ **PCR** amplification of TE junctions using primers in TE and flanking DNA
- ✓ **genome-specific** amplicon
- ✓ **polymorphic** between and within species
- ✓ **easy to derive** from genomic sequences (BES, WGS, BAC...)



ISBP markers open new perspectives for genome saturation in hexaploid wheat



IsbpFinder

One TE junction every 3.8 kb on average

70% genome-specific markers

One ISBP every 5.4 kb

17-Gb genome

More than 3 million ISBPs in the whole hexaploid wheat genome

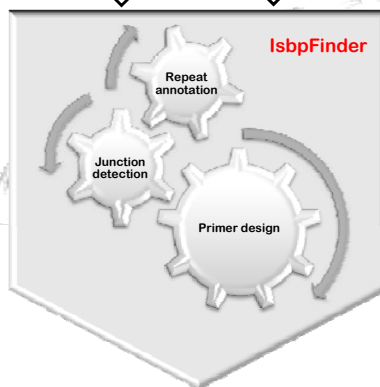
IsbpFinder



Sequences
(mutifasta)



Repeat database
(mutifasta)



Putative ISBP markers
(table and embl files)

IsbpFinder command line

```

### IsbpFinder ###
#
# AUTHOR:      Frederic CHOLET
# VERSION:    1.8
# CREATED:    2007-06-11
# LAST MODIF: 2010-9-9
# PURPOSE     This script is used to design primers for amplification of ISBP markers.
#             It can run and/or parse RepeatMasker results and design primers with "Primer3" program.

USAGE:
IsbpFinder [-dorm | -rm RepeatMasker_file] [-lib Repeat_library] [OPTIONS] fasta_file1 fasta_file2 ...

EXAMPLES:
IsbpFinder -dorm -lib /usr/local/db/TREP.fas -f EMBL MyFastaFile.fas
IsbpFinder -rm MyRmFile.out.xml -lib /usr/local/db/TREP.fas -f tab MyFastaFile.fas

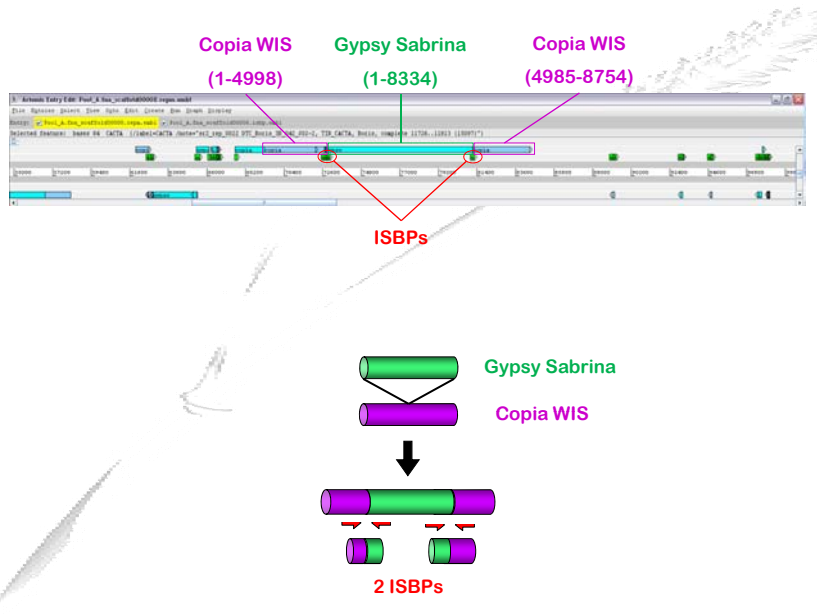
### OPTIONS ###
-dorm      run RepeatMasker for each FASTA sequence before primer design
-rm <file> repeatMasker result file
-lib <file> PATH to repeatmasker lib in FASTA format
-f string  output format: tab|csv|embl
           [tab] -> format with 1 line per couple of primers with field separator = "tab" [default]
           [csv] -> format with 1 line per couple of primers with field separator = ";"
           [embl] -> print results as EMBL features (to be read under ARTEMIS)
-o <file> output file (for tab|csv outputs)
           [default: STDOUT]
-e integer distance from the extremity of the element
           [default:50 nt]
-V:        print version
-h:        print this help

[epaux@pegase ISBP]$ IsbpFinder -dorm -lib /db/natif/TREPtotal.tfa -f embl -o Sequences.isbp Sequences.fas
  
```

IsbpFinder output table

seq_id	F_primer	F_primer_pos	R_primer	R_primer_pos	amplicon_pos	amplicon_size	junction_pos	junction_size	confidence	junction_pos	left_extremity	right_extremity
BAC1_1	GGAGAACAA	60.636	4428 TCTCCAATC	60.074	4588 GGAGAACAA	171	40.35	gypsy_low	121-122,124	TREP2349	Tf no	TREP2349, Tf no
BAC1_2	GCATATGAC	59.979	10862 GTCCCTATG	59.985	10866 GCATATGAC	199	46.23	gypsy_high	67-68	TREP2349	Tf yes	TREP837, TREP no
BAC1_3	TGGTAGGAA	60.074	10758 TCCGACATC	59.889	10924 TGGTAGGAA	180	42.78	gypsy_low/medium	127-128	TREP837	TREP no	TREP836, TREP no
BAC1_4	GTGGAGCCG	59.938	11879 TATGATGGG	60.822	11873 GTGGAGCCG	239	35.98	LowCopy/NI/medium	65-66			TREP836, TREP no
BAC1_5	GGAAAACCC	60.088	16099 TGTAGATCG	59.992	16948 GGAAAACCC	250	35.6	gypsy_unkn/high	150-151,168	TREP836	TREP yes	TREP836, TREP yes
BAC1_6	TGCTCTCCG	59.992	16332 TGCTCTCGT	59.864	16494 TGCTCTCCG	163	38.65	Unknown_/medium	102-103	TREP3551	Tf no	
BAC1_7	GAAGAGACA	60.117	16862 TGGAGATGT	59.577	17174 GAAGAGACA	215	37.56	LowCopy/NI/high	164-165			TREP73, TREP yes
BAC1_8	TGCCAAAAT	60.407	17135 AAAAGCATT	60.23	17267 TGCCAAAAT	133	36.84	Unknown_/medium	76-77	TREP73	TREP no	TREP73, TREP no
BAC1_9	TTTGCCCTAG	59.953	19135 GGTAGGGG	59.376	19308 TTTGCCCTAG	174	40.8	LowCopy/NI/medium	102-103			TREP1242, Tf no
BAC1_10	AGATAGAAC	59.576	19289 AGGTCAAAAT	59.967	19521 AGATAGAAC	234	41.03	gypsy_low/medium	186-187			TREP1242, Tf no
BAC1_11	AGATGTGAG	59.981	20681 TCCGCCACG	59.823	20906 AGATGTGAG	226	44.25	LowCopy/NI/medium	221-222			TREP2237, Tf no
BAC1_12	TAGAGATCA	59.823	20887 TCCCGTCAI	59.9	21171 TAGAGATCA	285	34.39	gypsy_copi/high	56-57	TREP234	TREP no	TREP129, TREP yes
BAC1_14	ATTGTAGAT	59.788	22313 AGTGTAGAI	59.89	22551 ATTGTAGAT	239	44.35	copi_copi/high	70-71	TREP129	TREP no	TREP2200, TREP yes
BAC1_15	GTAGCAGTC	59.983	25846 TAAGGCCCG	59.793	26029 GTAGCAGTC	184	50	copi_copi/high	119-120	TREP129	TREP yes	TREP129, TREP no
BAC1_16	AATCTCACT	59.793	26010 TGATCAGGA	59.984	26228 AATCTCACT	219	55.25	copi_copi/low	76-77,86-87	TREP129	TREP no	TREP129, TREP no
BAC1_18	CTGCACACT	59.984	26209 GGGTGTGG	59.971	26466 CTGCACACT	258	47.29	copi_copi/high	255-256	TREP129	TREP yes	TREP400, TREP no
BAC1_17	AGATCTTCC	60.022	26414 CTGATTTAG	59.955	26568 AGATCTTCC	155	50.32	copi_copi/low/high	50-51	TREP129	TREP yes	TREP129, TREP yes
BAC1_19	ATATAGACA	60.073	26719 AGCCGACTC	60.134	26928 ATATAGACA	160	49.38	LowCopy/NI/high				TREP1231, Tf yes
BAC1_20	AGCCCAAGTA	59.997	35456 GACCTTGCC	60.074	35627 AGCCCAAGTA	172	51.16	gypsy_low/high	124-125	TREP129	TREP yes	TREP1231, Tf yes
BAC1_21	GGAGGTAGA	60.074	35608 CGCATTTAG	58.796	35789 GGAGGTAGA	182	37.36	LowCopy/NI/high	93-94			TREP3451, Tf yes
BAC1_22	TGAACAAAC	58.796	35768 AAGCGGAAG	59.996	35931 TGAACAAAC	164	32.32	Unknown_/high	79-80	TREP3451	Tf yes	
BAC1_23	TAATGTGTC	60.073	37817 CCCAATCAI	59.565	37883 TAATGTGTC	167	36.53	LowCopy/NI/medium	90-91			TREP2464, Tf no
BAC1_24	TTGACACTG	59.565	37964 GGCATCTTC	59.937	38224 TTGACACTG	261	49.81	Unknown_/high	44-45	TREP2464	Tf yes	
BAC1_25	GGGTGTGTC	58.792	40209 GGCAGCGAC	59.928	40509 GGGTGTGTC	292	31.18	Unknown_/medium	111-112	TREP2469	Tf no	
BAC1_26	CGACGTCCG	60.028	41346 TGGGTTCTG	59.818	41643 CGACGTCCG	298	37.92	LowCopy/NI/high				TREP3161, Tf yes
BAC1_29	TGCATCTTC	60.135	49976 TGGGTTGCT	59.772	50212 TGCATCTTC	237	42.19	copi_copi/low/high	90-91	TREP839	TREP yes	
BAC1_30	TACATGATT	59.744	51763 GGGCAAAA	60.03	52062 TACATGATT	300	34	LowCopy/NI/high	264-265			TREP2271, Tf yes
BAC1_31	ACCACTAGC	60.03	52043 GGCATTTGAT	60.265	52211 ACCACTAGC	169	39.64	Mutator_c/high	70-71	TREP2271	Tf no	TREP3030, Tf yes
BAC1_32	GATTTGAGG	60.014	53920 GGGTGGGAC	60.142	54114 GATTTGAGG	195	32.82	CACTA_copi/high	172-173	TREP3030	Tf no	TREP210, Tf yes
BAC1_33	TTGCAGTGA	60.025	54897 CTGGTTGAI	60.075	55174 TTGCAGTGA	278	50.36	copi_copi/low	108-109,110	TREP210	Tf no	TREP210, Tf yes
BAC1_34	CGAGTCAAC	59.948	83827 CATAGGACG	59.873	84078 CGAGTCAAC	250	50.4	copi_copi/low/medium	216-217	TREP210	Tf no	

Visualization of IsbpFinder results under Artemis

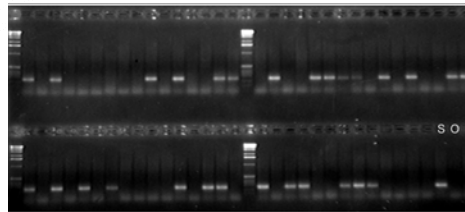
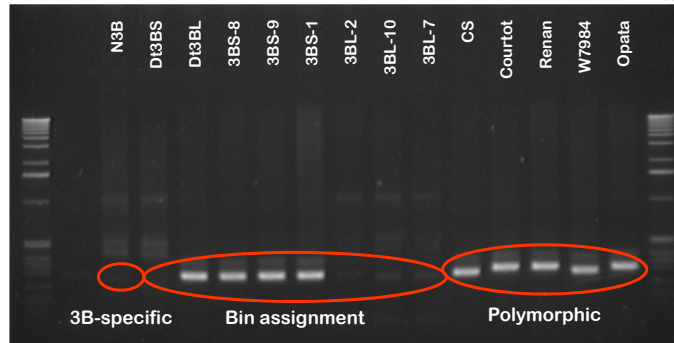


ISBP design and genome-specificity

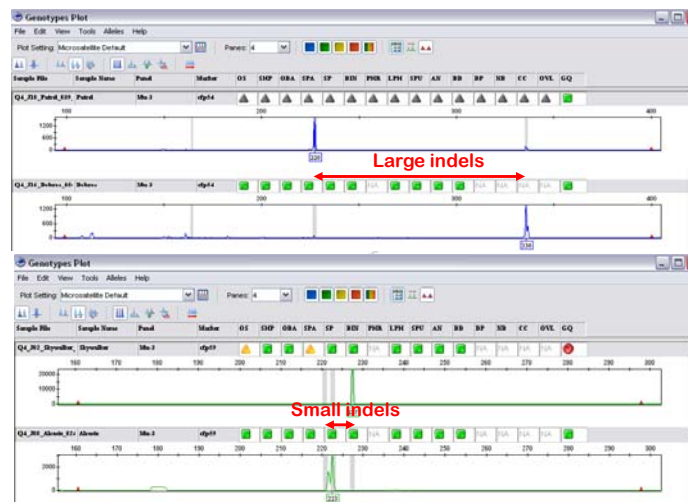
BAC sequences more than 10 kb	BAC-end sequences 600 bp	WGS (GSFLX sequences) 400 bp
IsbpFinder	IsbpFinder	IsbpFinder
1 junction / 3.8 kb	9.72%	6%
Genome specificity	Genome specificity	Genome specificity
70-80%	50-60%	40-50%

Sequence length impacts ISBP detection due to the accuracy of sequence annotation

Agarose gel electrophoresis

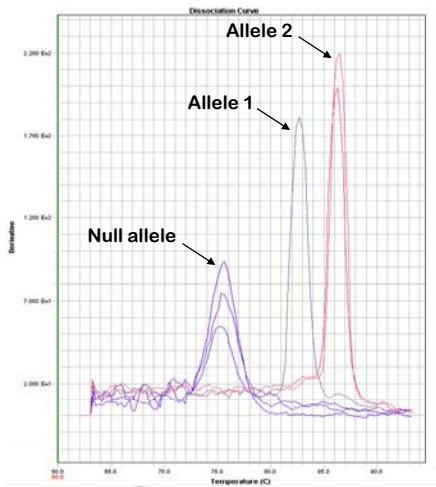


Fluorescent capillary electrophoresis



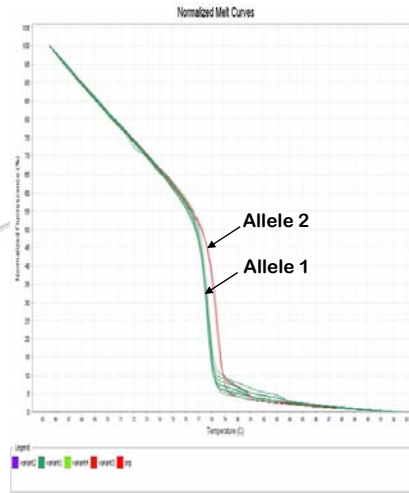
- ✓ **Small (1 – 4 nucleotide) deletions**
- ✓ **Microsatellite motifs (1.6% of the ISBP markers contain SSRs)**
- ✓ **Small TE insertions (MITE...)**

Melting curve-based genotyping



Melting curve analysis (MCA)

- ✓ Presence – absence polymorphism
- ✓ Size polymorphism
- ✓ Sequence polymorphism (several SNPs)



High resolution melting (HRM)

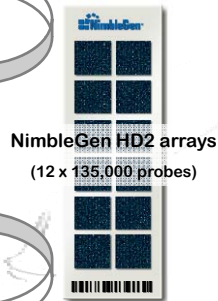
- ✓ SNP genotyping
- ✓ SNP discovery

IMaGe: ISBP Microarray-based Genotyping

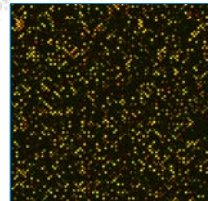
Cy3-labelled reference



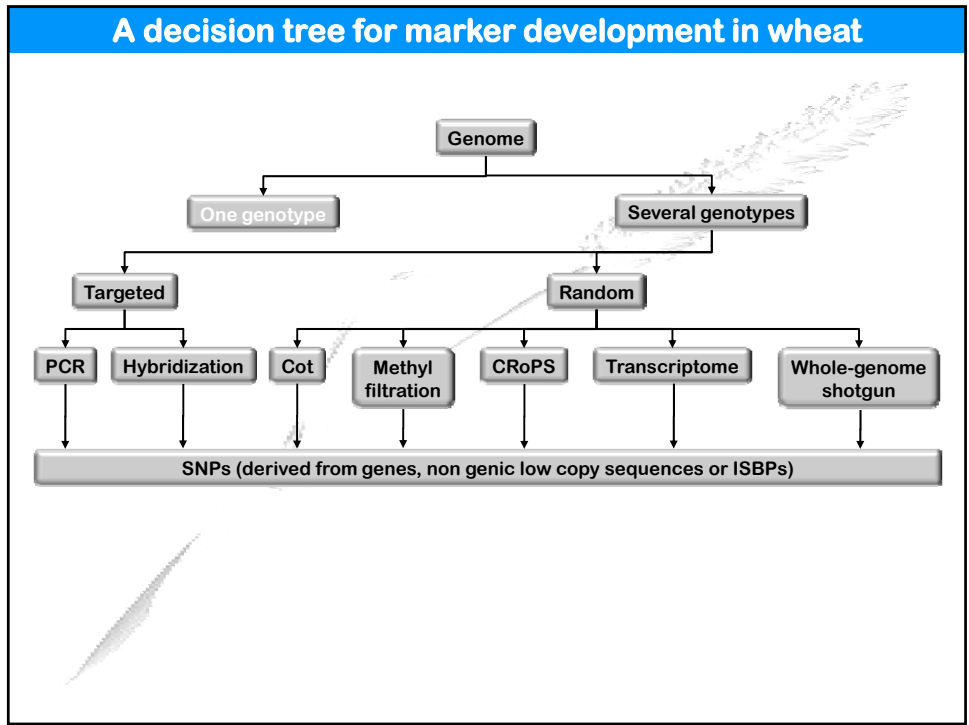
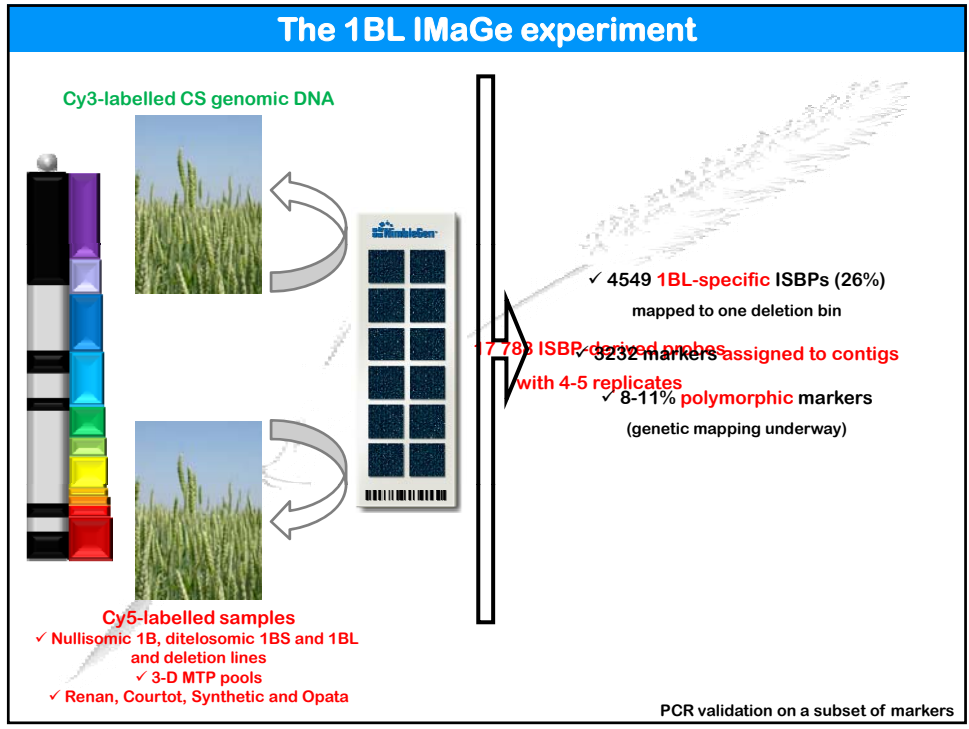
Cy5-labelled sample



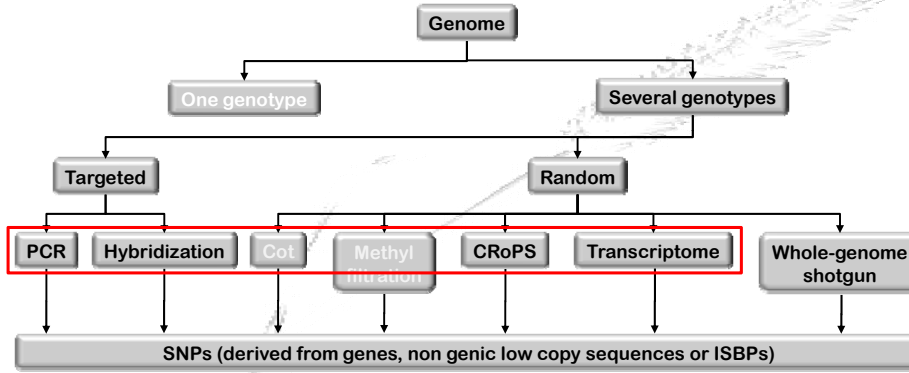
NimbleGen HD2 arrays
(12 x 135,000 probes)



Presence – absence polymorphism

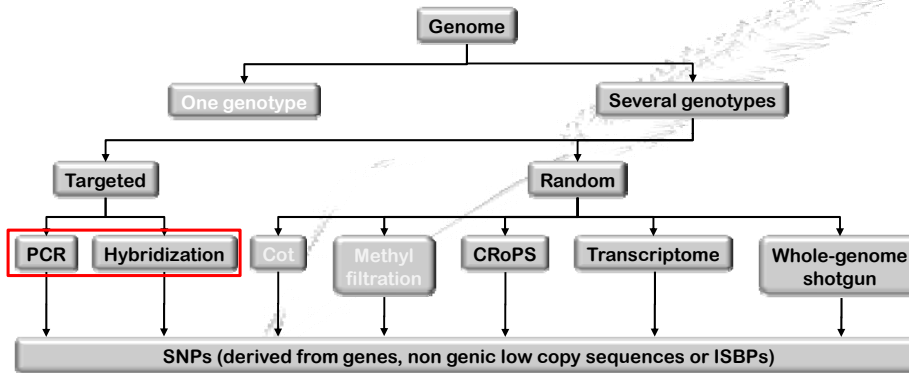


A decision tree for marker development in wheat



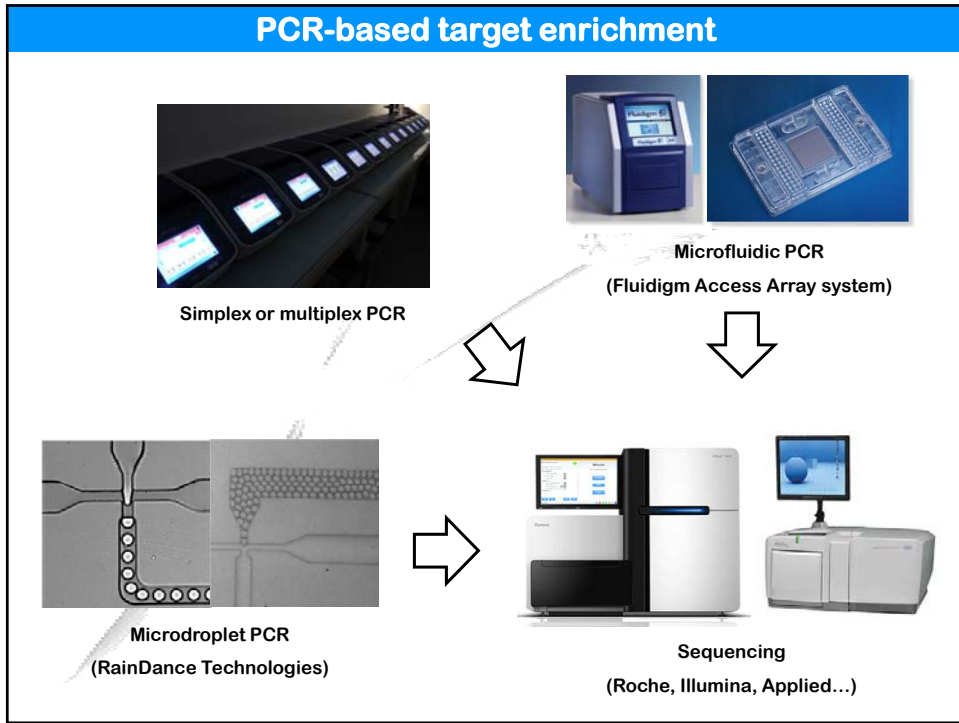
Reproducible sampling of the same fraction of the genome between different genotypes

A decision tree for marker development in wheat

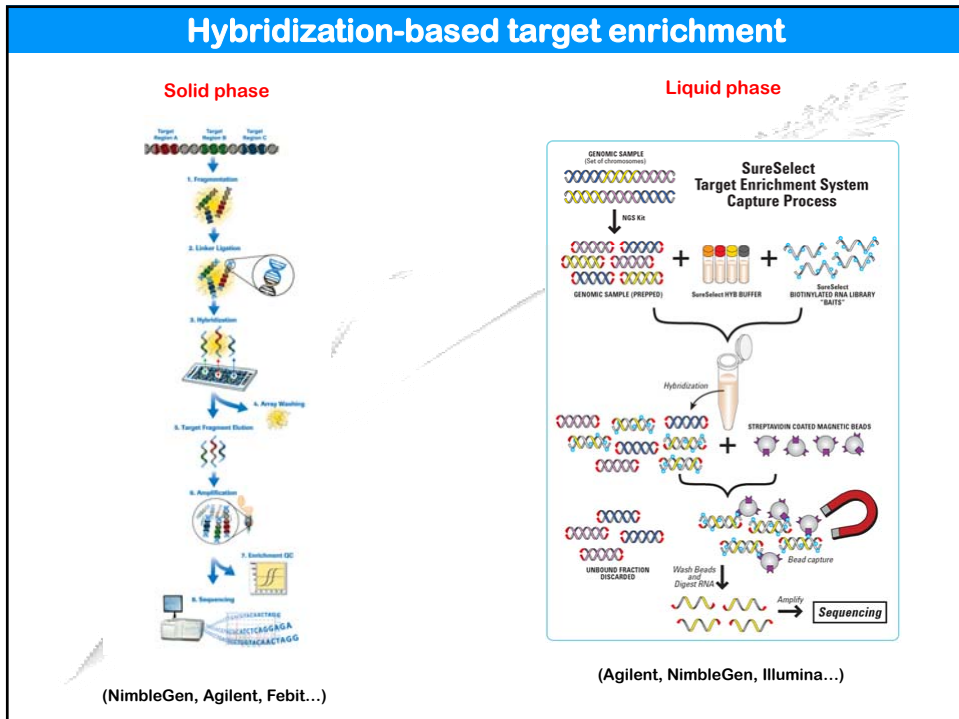


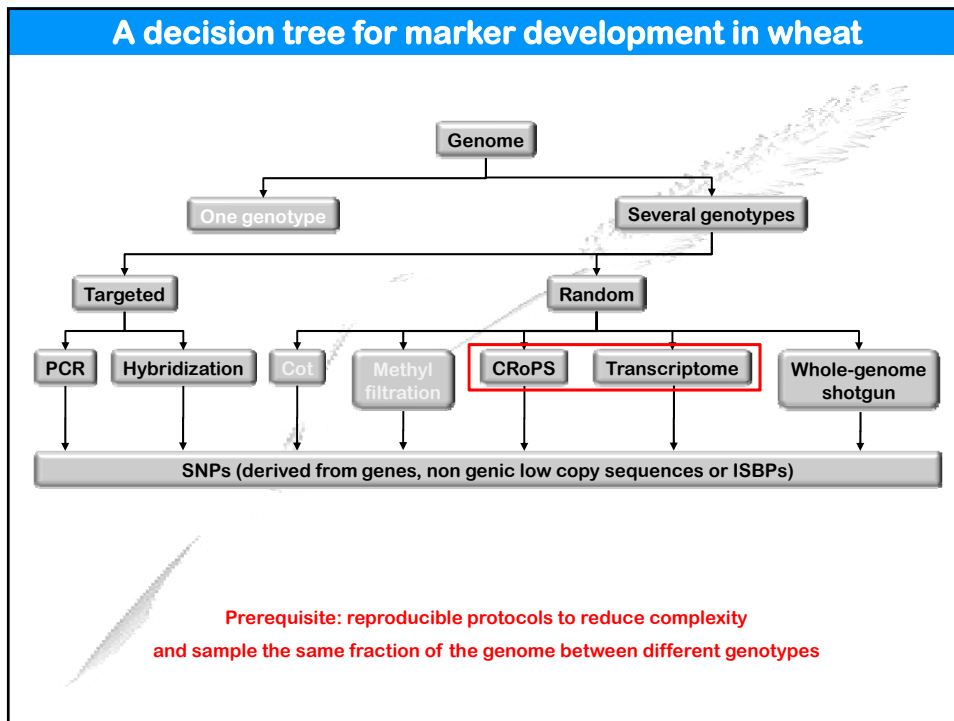
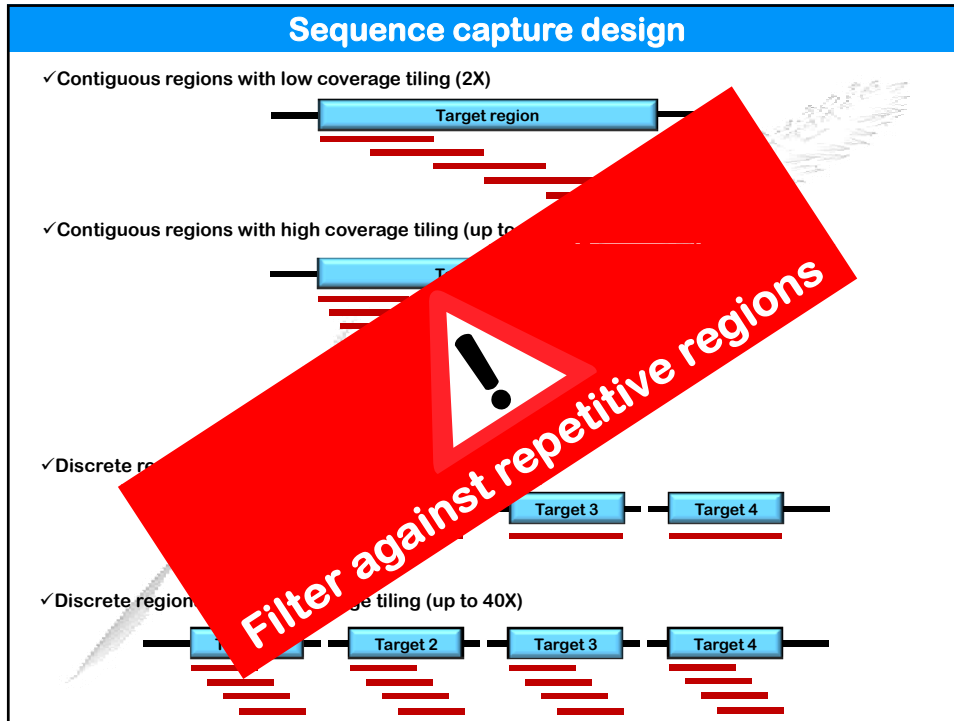
Prerequisite: contiguous or discrete target regions clearly identified

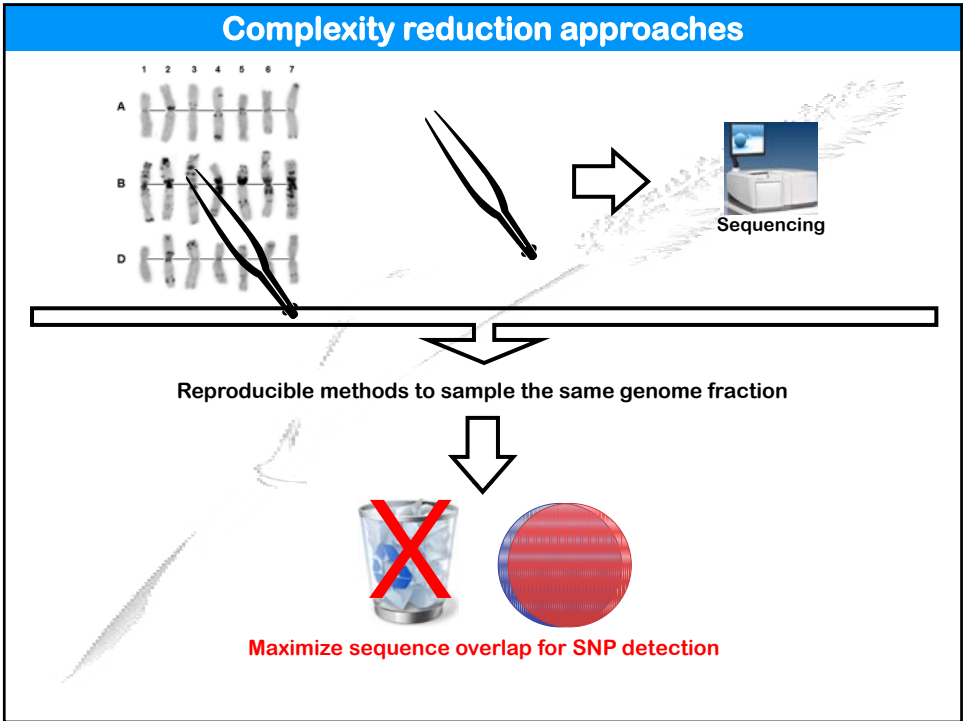
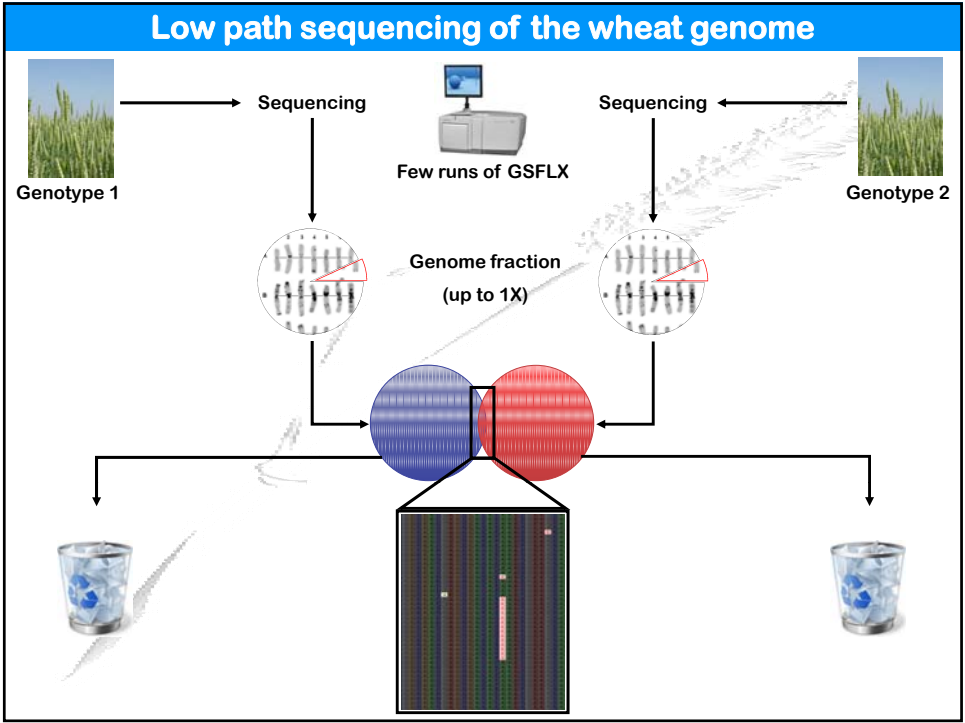
PCR-based target enrichment



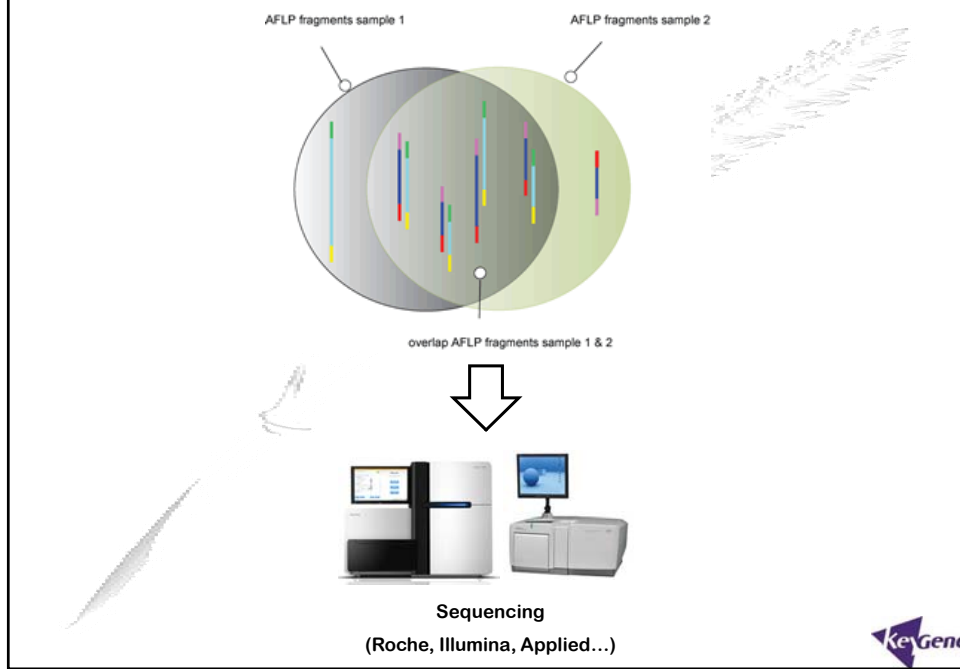
Hybridization-based target enrichment



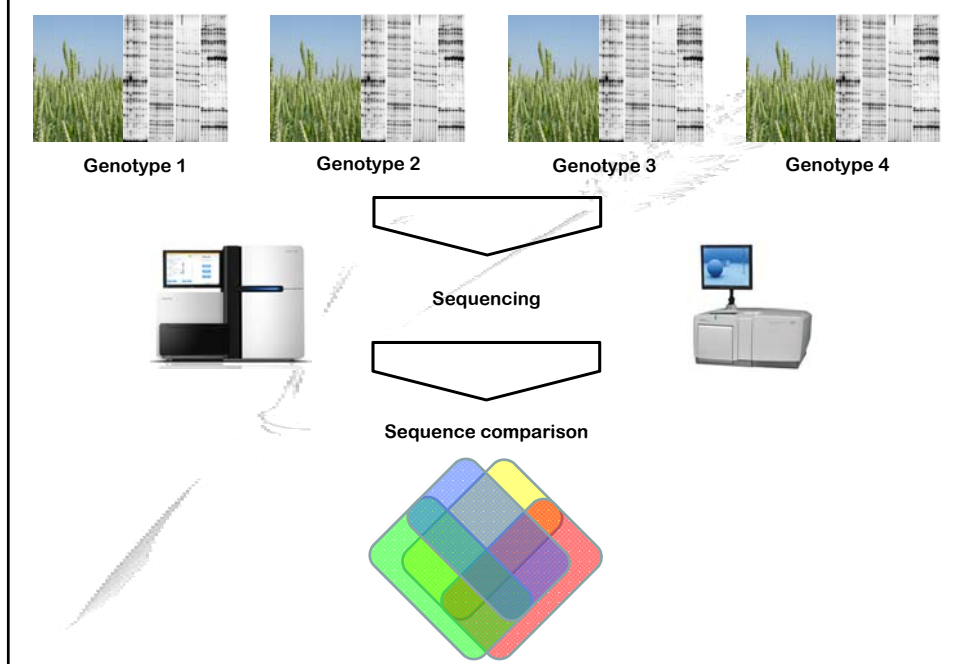


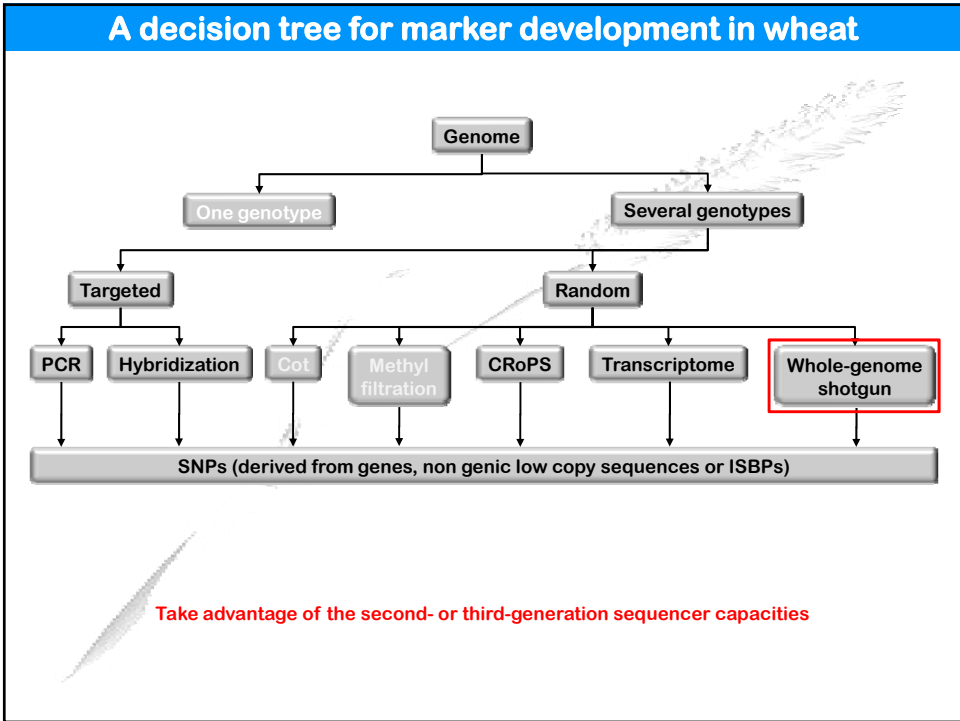
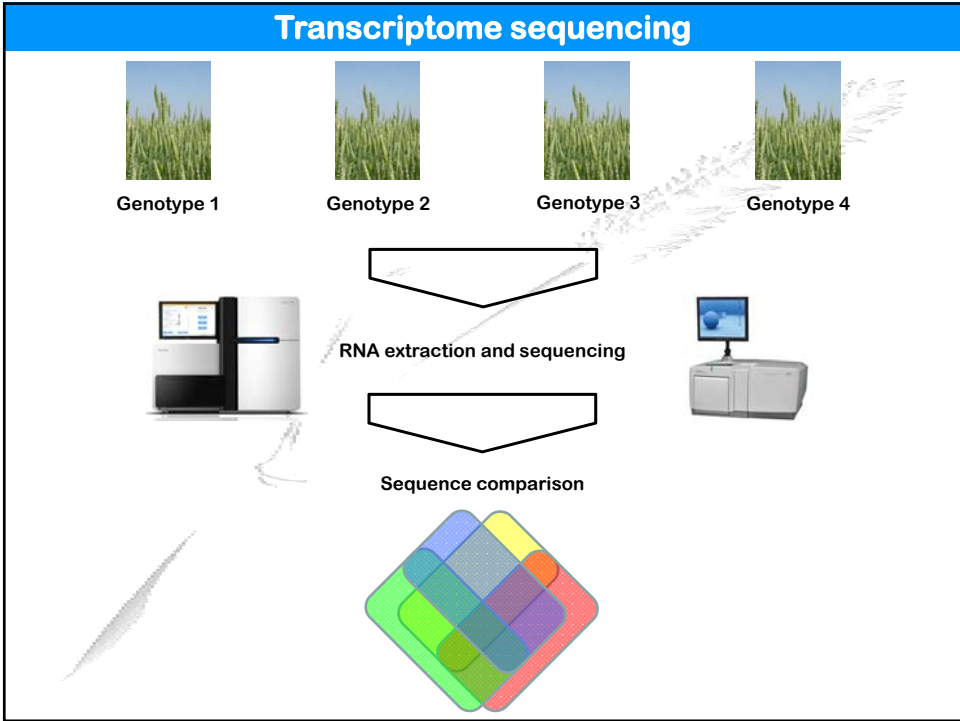


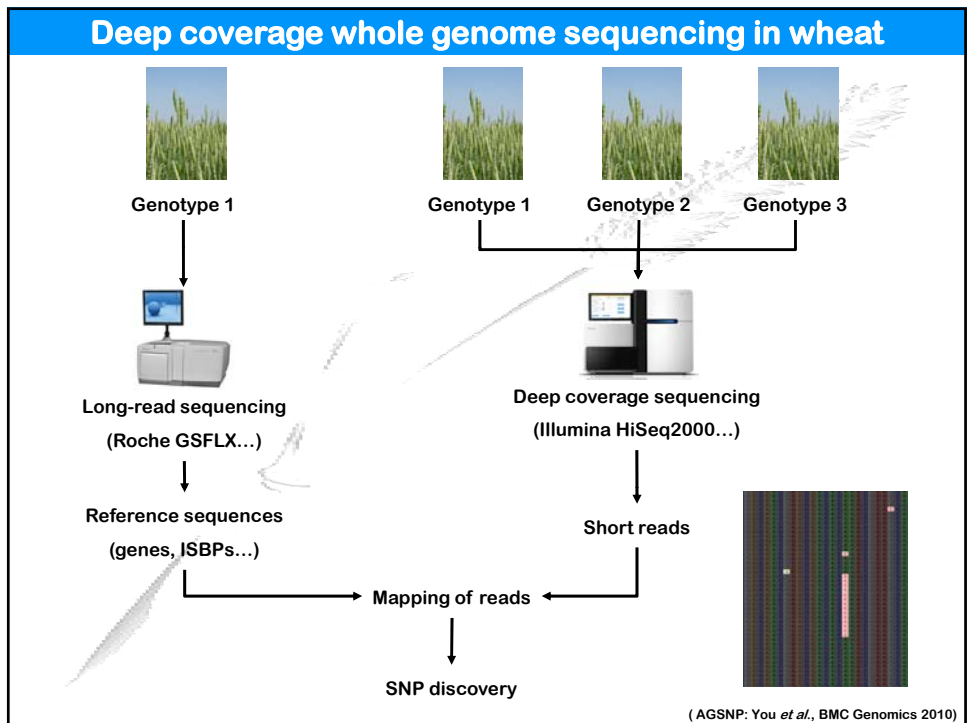
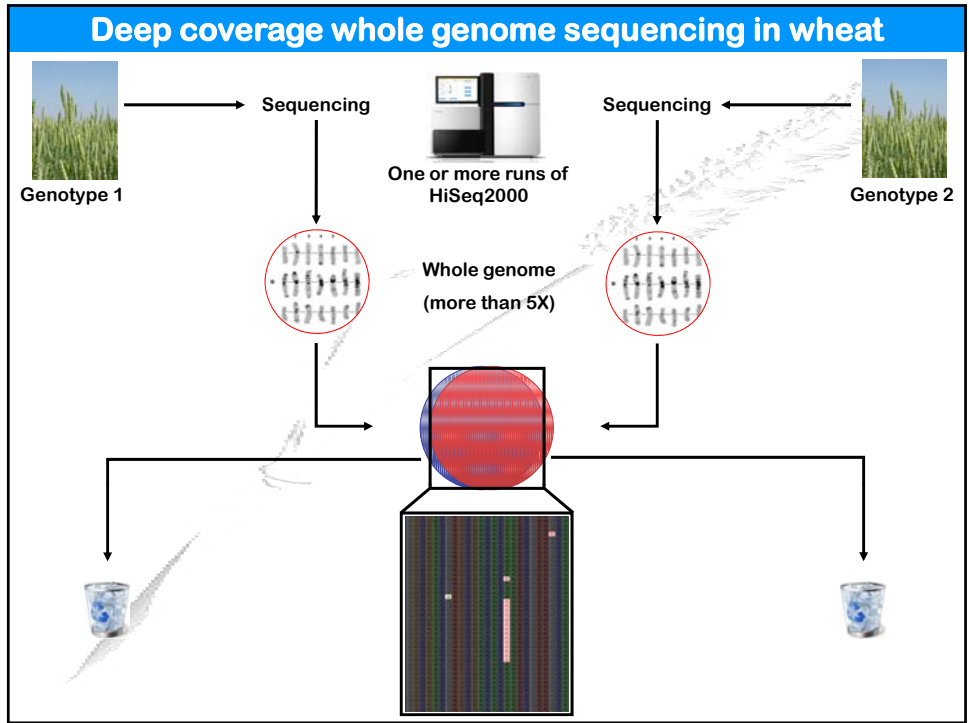
Complexity Reduction of Polymorphic Sequences (CRoPS)



Sequencing SSAP markers







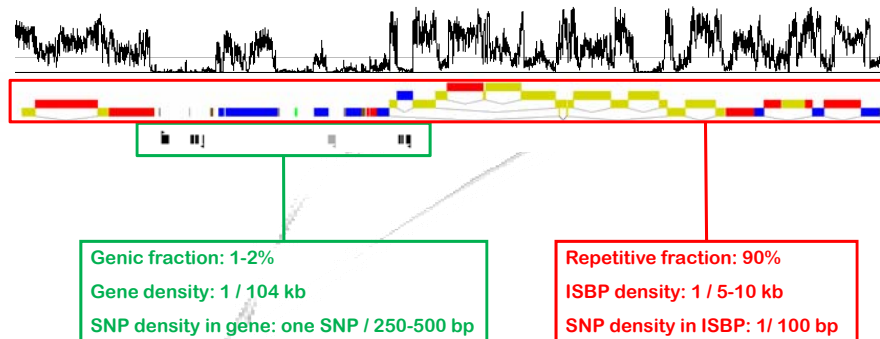
International wheat SNP initiatives

- ✓ Single Nucleotide Polymorphism (SNP) markers for high-throughput genotyping to advance genomic, genetic and breeding research in wheat (Eduard Akhunov, USA)
- ✓ Australian SNP discovery and genotyping project (Matt Hayden, Australia)
- ✓ Haplotype Polymorphism in polyploid wheats and their Diploid Ancestors (Jan Dvorak, USA)
- ✓ Investigating gene function in cereals (Keith Edwards, UK)
- ✓ SNP discovery in complexity reduced libraries and amplicon pools (Matt Hayden, Australia)
- ✓ Population genomics and association mapping in Israeli populations of wild relatives of wheat (Adina Breiman, Israel)
- ✓ Whole genome sequencing of Australian wheat cultivars (Dave Edwards, Australia)
- ✓ SNP discovery in durum wheat (Luigi Cattivelli, Italy)

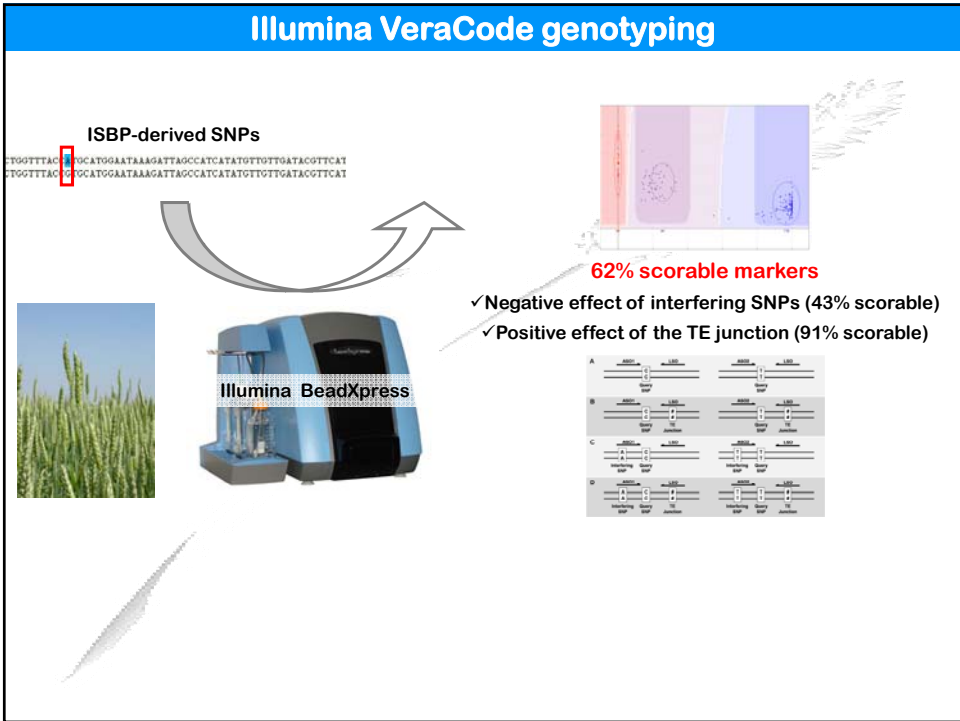
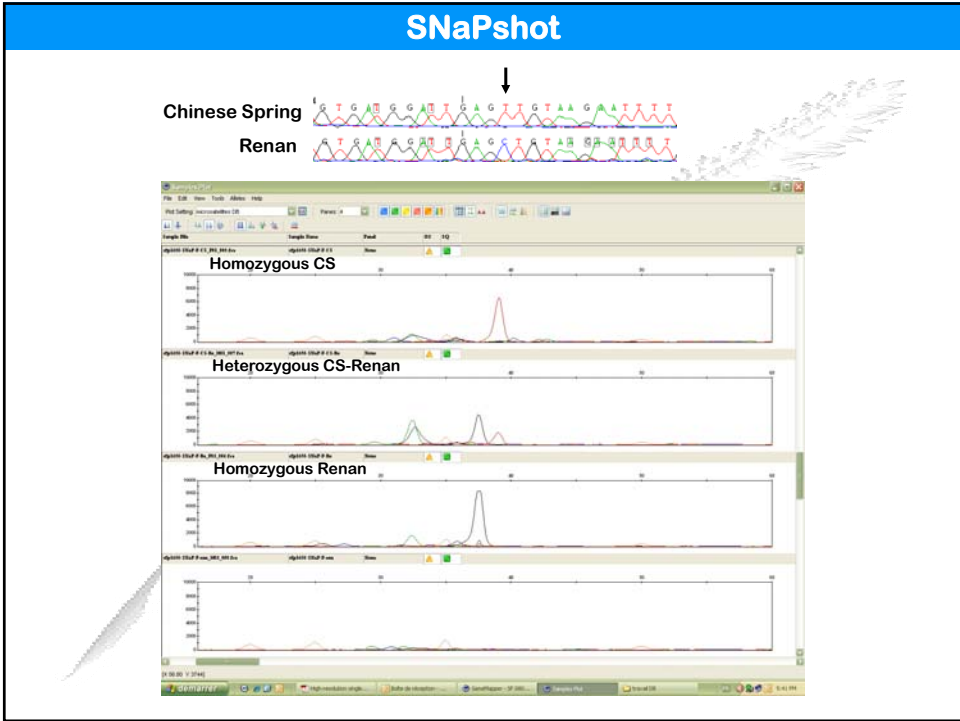
Several thousands of gene-derived SNPs will be available in the coming years
→ Development of Illumina Infinium 9K and 50K chips

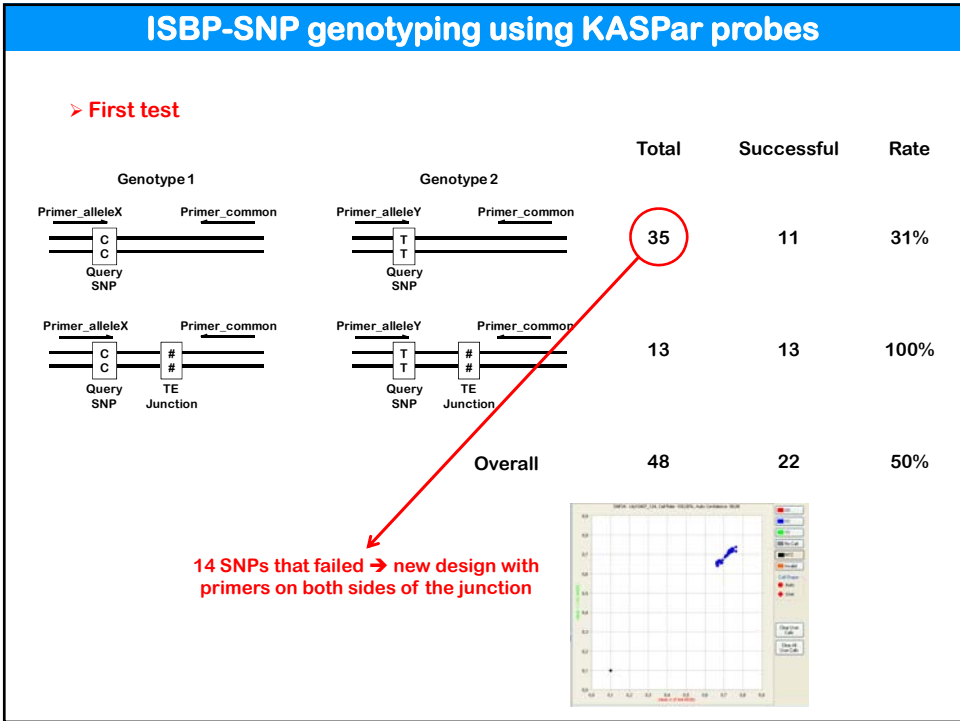
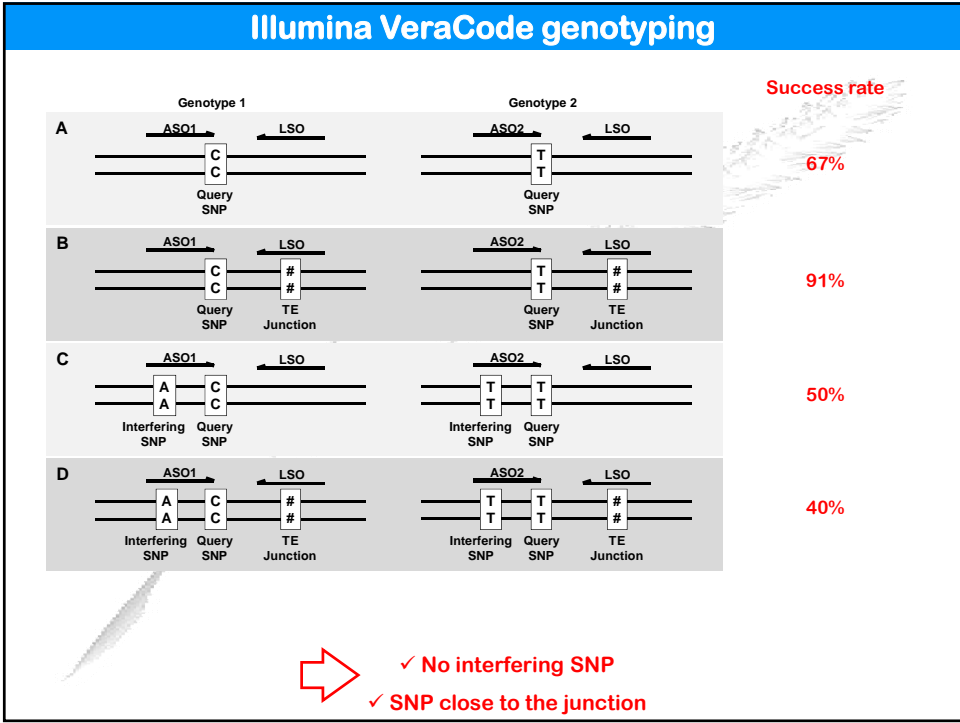
(http://wheatgenomics.plantpath.ksu.edu/IWSWG/snp_projects)

Mining the repetitive fraction



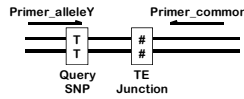
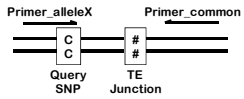
ISBP-SNPs not as a surrogate but in complement to genic SNPs



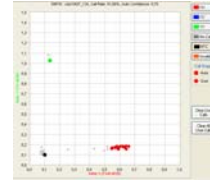
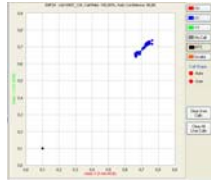


ISBP-SNP genotyping using KASPar probes

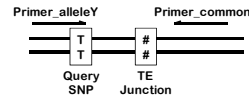
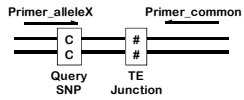
➤ **Second test**



	Total	Successful	Rate
Second test	14	13	93%
Overall	27	26	96%

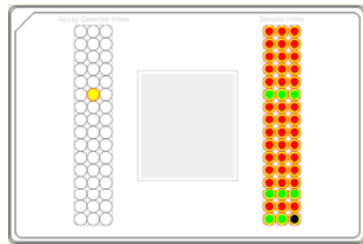
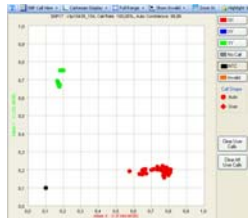


➤ **Overall**

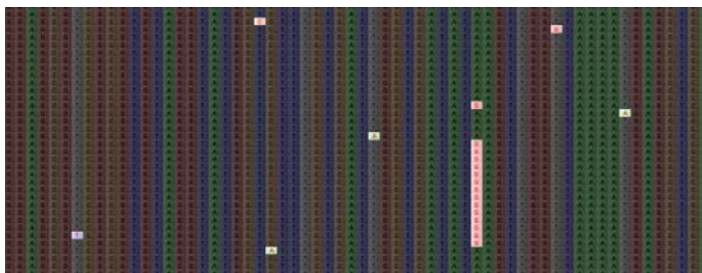


ISBP-SNP genotyping using KASPar probes

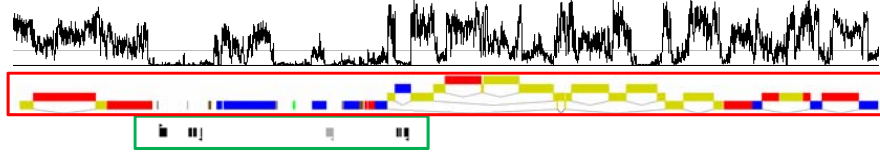
➤ **Reproducibility: 100%**



➤ **Consistency with sequencing data: 100%**



Advantages and drawbacks of ISBP-SNPs



- ⊕ Numerous
- ⊕ Easy to detect (no confusion with homoeologous variations)
- ⊕ Easy to score (pseudo-diploid clusters)

- ⊖ Not linked to genes (but causal polymorphisms are not always in genes)

- ⊖ Cannot be scored on Illumina Infinium
- ⊖ Not transferable to other species

A decision tree for marker development in wheat

