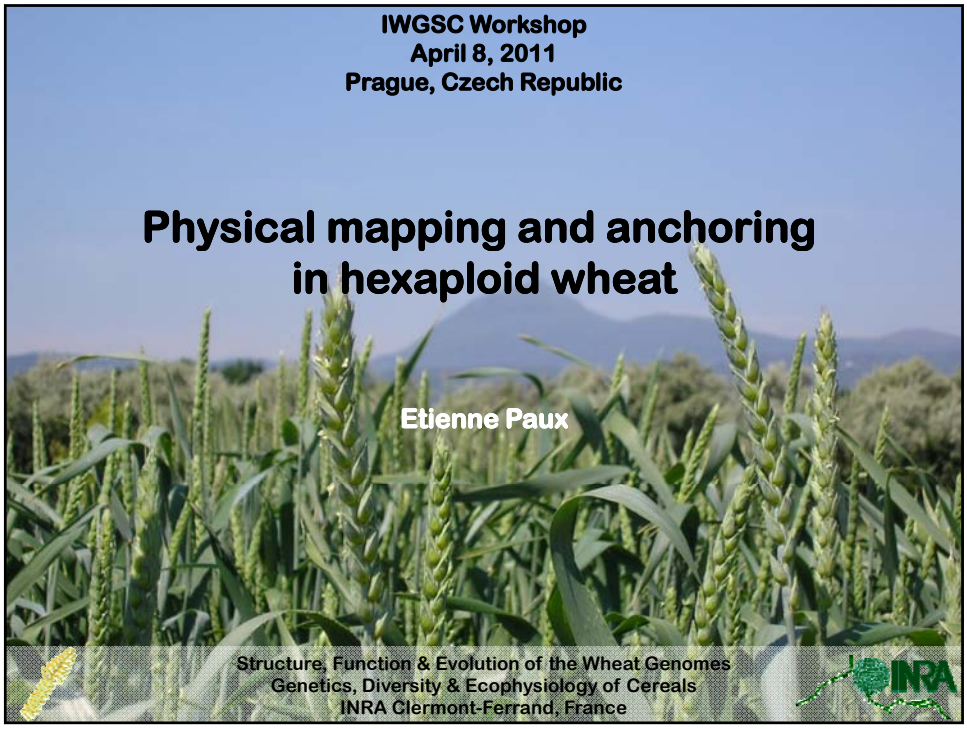
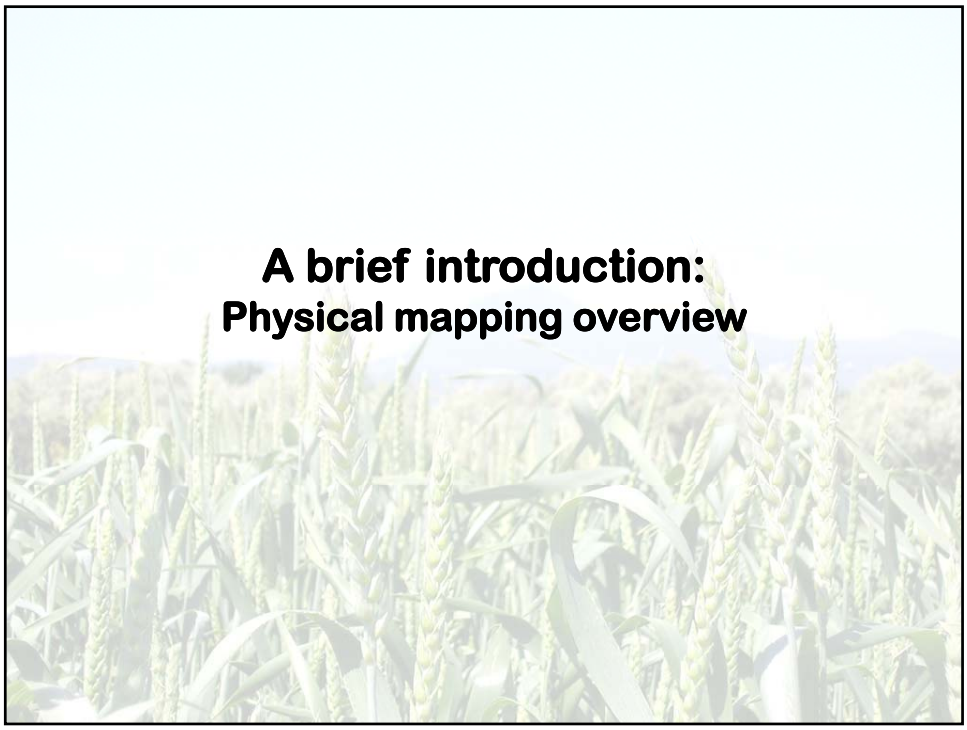
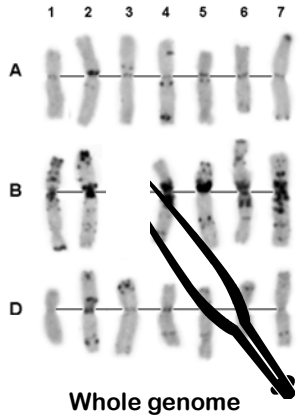
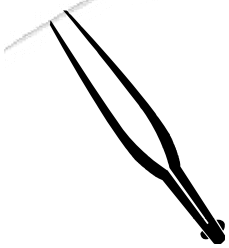

IWGSC Workshop
April 8, 2011
Prague, Czech Republic
Physical mapping and anchoring in hexaploid wheat
Etienne Paux
Structure, Function & Evolution of the Wheat Genomes
Genetics, Diversity & Ecophysiology of Cereals
INRA Clermont-Ferrand, France
INRA


A brief introduction:
Physical mapping overview

Construction of chromosome-specific integrated maps

Chromosome or chromosome arm sorting





Whole genome
- ✓ 21 chromosomes
- ✓ 17 000 Mb
- ✓ homoeologous loci

Sorted chromosome
- ✓ 1 chromosome (arm)
- ✓ 1000 Mb max.
- ✓ no homoeologous loci

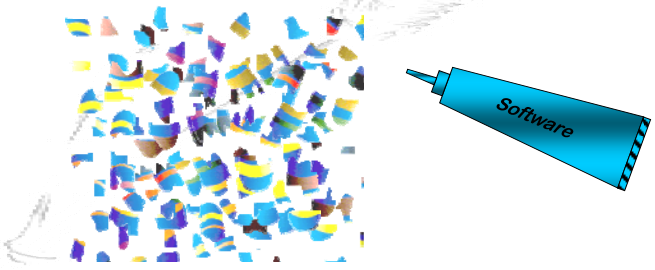
Construction of chromosome-specific integrated maps

BAC library construction

Ordered chromosome or chromosome arm
60,000-80,000 BACs per chromosome arm

Construction of chromosome-specific integrated maps

Physical contig assembly



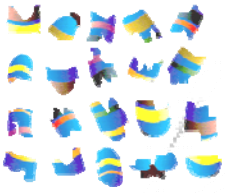
60,000 – 80,000 BAC contigs

500 BAC contigs per chromosome arm

Construction of chromosome-specific integrated maps

Physical contig anchoring

500 BAC contigs

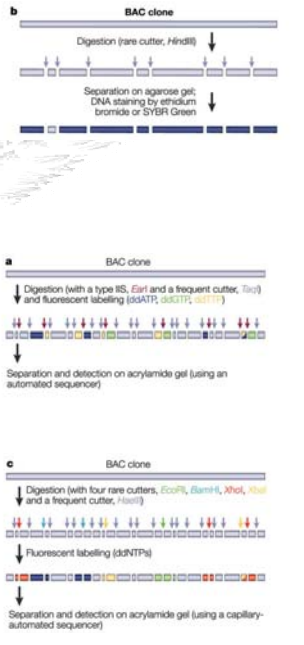


Molecular markers

BAC fingerprinting

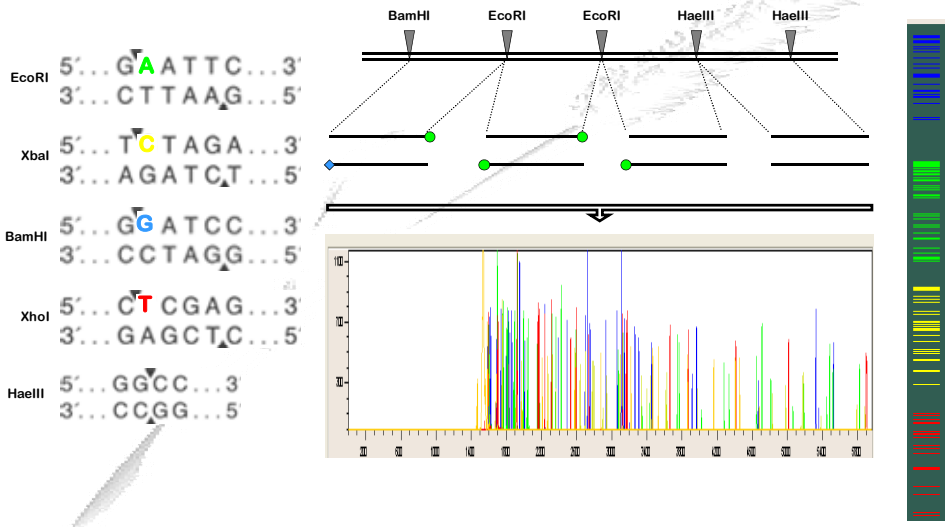
Different fingerprinting methods

- Coulson *et al.* (1986)
  - ✓ *Hin*dIII
  - ✓ BEt staining and agarose gel
  - ✓ 30-40 bands per BAC
- Zhang *et al.* (1997)
  - ✓ *Hin*dIII + *Hae*III
  - ✓ BEt staining and agarose gel
  - ✓ 45-50 bands per BAC
- Faller *et al.* (2000)
  - ✓ *Ear*I+ *Taq*I
  - ✓ fluorescent labelling (ddATP, ddGTP, ddTTP) and capillary sequencer
  - ✓ 100-110 bands per BAC
- Luo *et al.* (2003)
  - ✓ *Bam*HI+ *Hin*dIII + *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



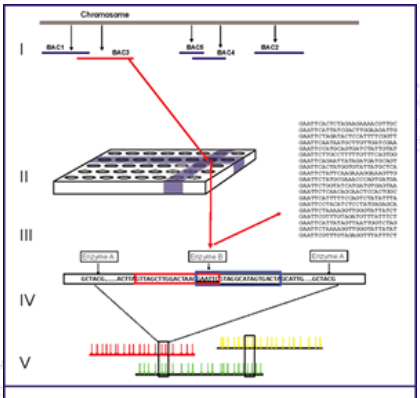
(adapted from Meyers *et al*, Nature Rev Genet, 2004)

SNaPshot fingerprinting



(adapted from Luo *et al*, Genomics, 2003)

Whole-genome profiling (KeyGene)



I BAC library construction (125 to 175 kb inserts)
II BAC clones are cleverly pooled
III Pooled BAC DNA is sequenced with Illumina GA II
IV BAC specific sequence tags, spaced every 2-6 Kb are generated (knowledge of entire BAC, not just end sequence)
V BAC clones are ordered into contigs, using sequence-based anchor points

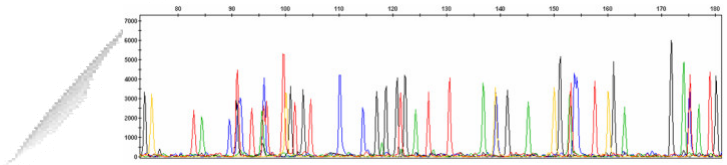
Editing fingerprints

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:

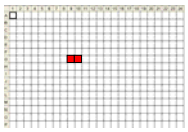
- ✓ "true peak" derived from a DNA insert digested band; → BAC fingerprint
- Background removal:
  - ✓ low signal peak produced by the machine;
  - ✓ partial digestion related peak;
  - ✓ star activity by-product;
  - ✓ *E. coli* genomic DNA band;
- Pre-processing:
  - ✓ vector band;
  - ✓ out of size standard range band (with unreliable sizing);
  - ✓ wide area peak (unreliable, resulting from co-migrating fragments).

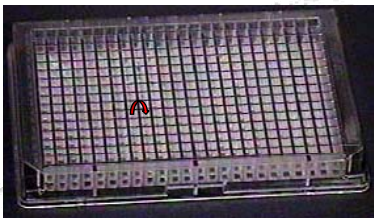


(adapted from Scalabrin *et al.*, *BMC Bioinformatics*, 2009)

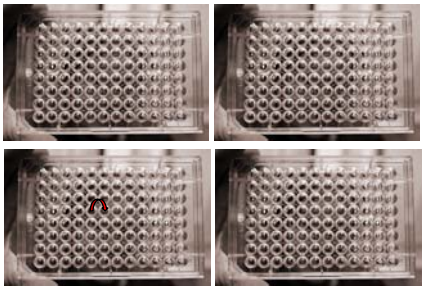
Well-to-well contamination removal using Genoprofiler

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

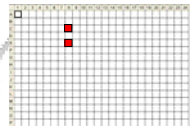




Splitting of 384-well plate into four 96-well plate during DNA extraction process.

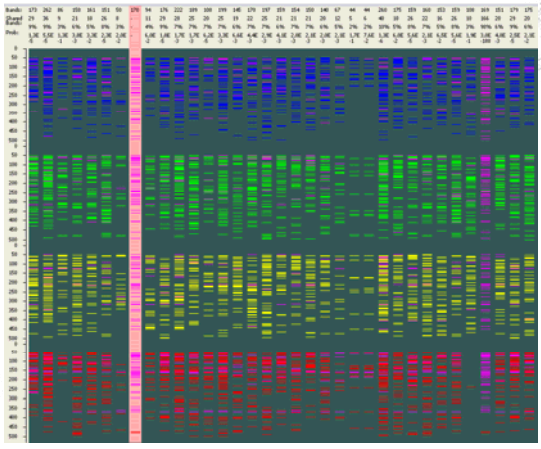


✓ Well-to-well contamination in 96-well plate format
➔ Non-adjacent wells showing similar profiles



'One-to-one' contamination

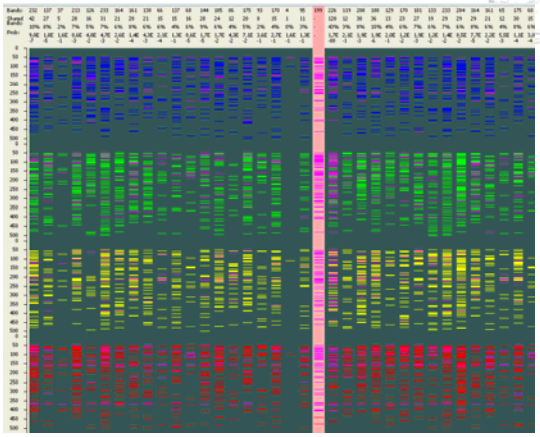
Two adjacent wells contain the same clone B1



80-100% identity of fingerprints

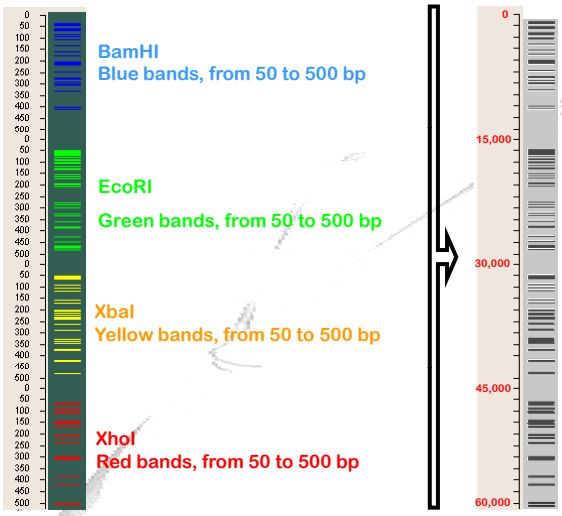
‘One-to-two’ contamination

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



BamHI
Blue bands, from 50 to 500 bp

EcoRI
Green bands, from 50 to 500 bp

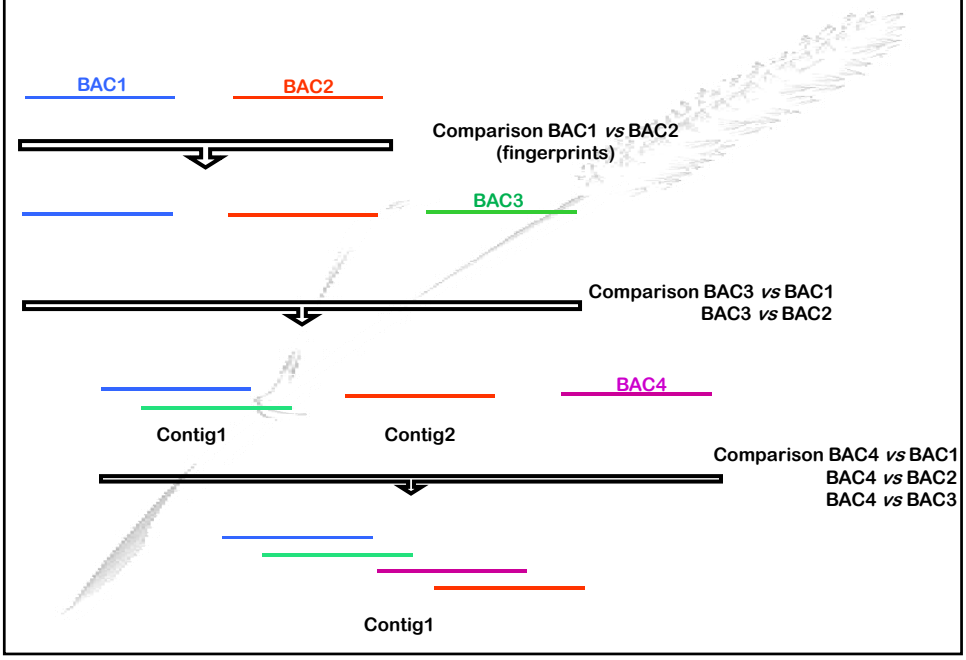
XbaI
Yellow bands, from 50 to 500 bp

XhoI
Red bands, from 50 to 500 bp

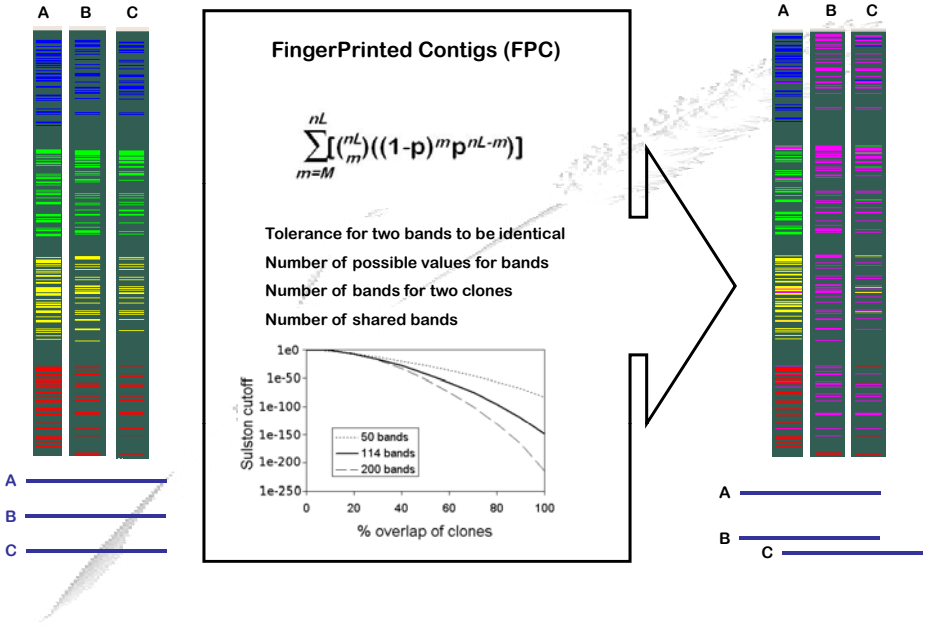
Complete fingerprint
‘Black’ bands, from 0 to 60,000
(with 4 gaps)

Contig assembly
1- Overview

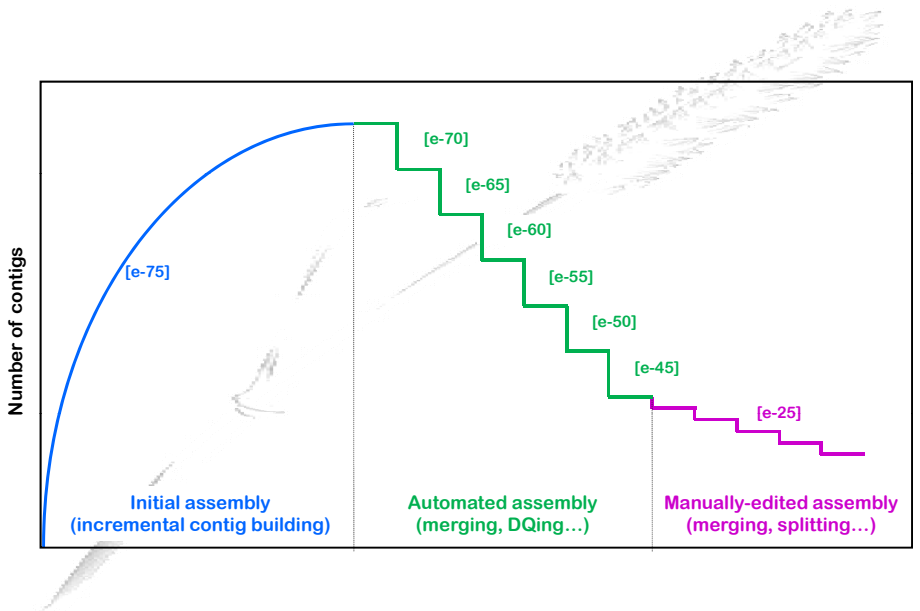
Pairwise comparison and contig assembly



Overlap calculation: the Sulston score

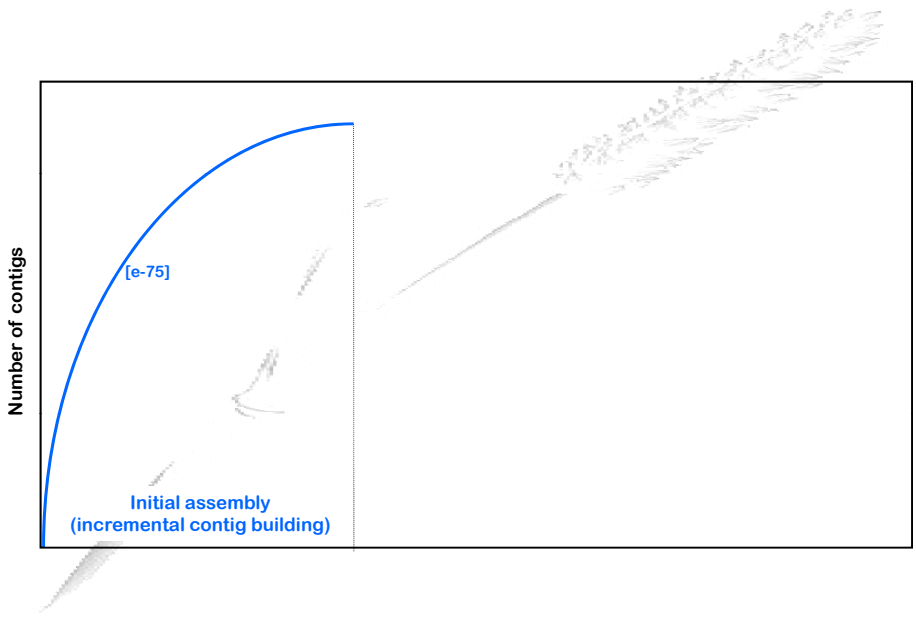


Assembly of the physical map

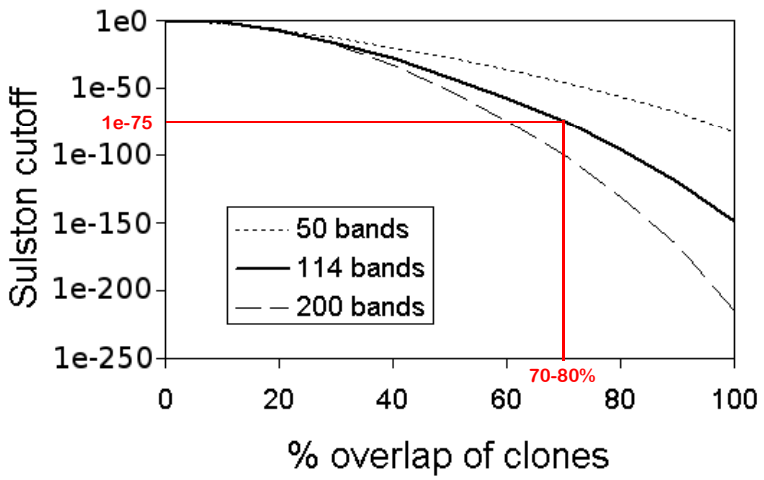


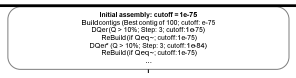
Contig assembly
2- Initial assembly

Assembly of the physical map



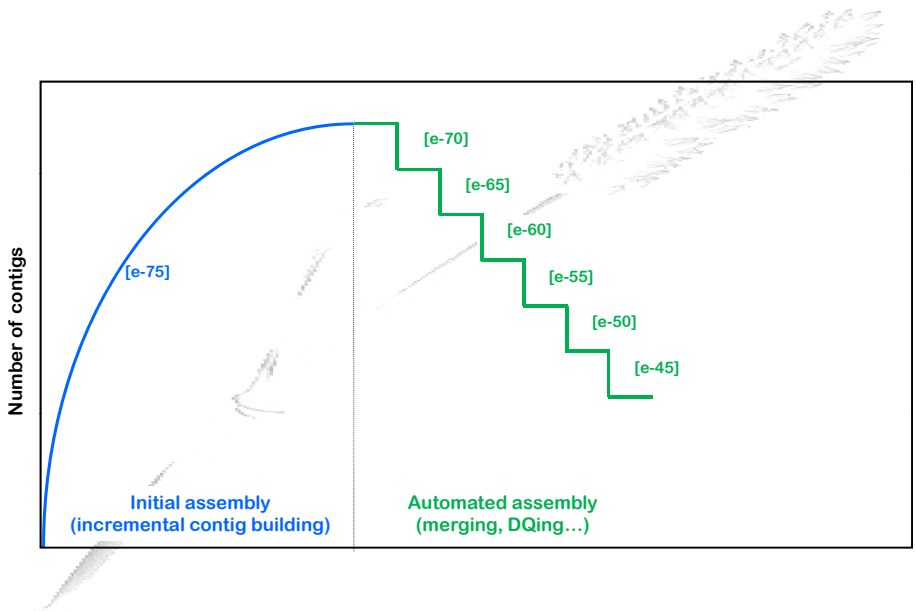
Sulston score overlap



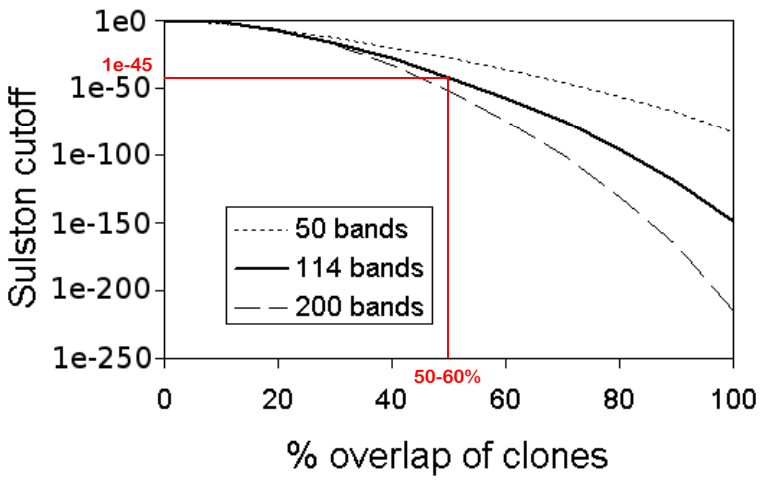


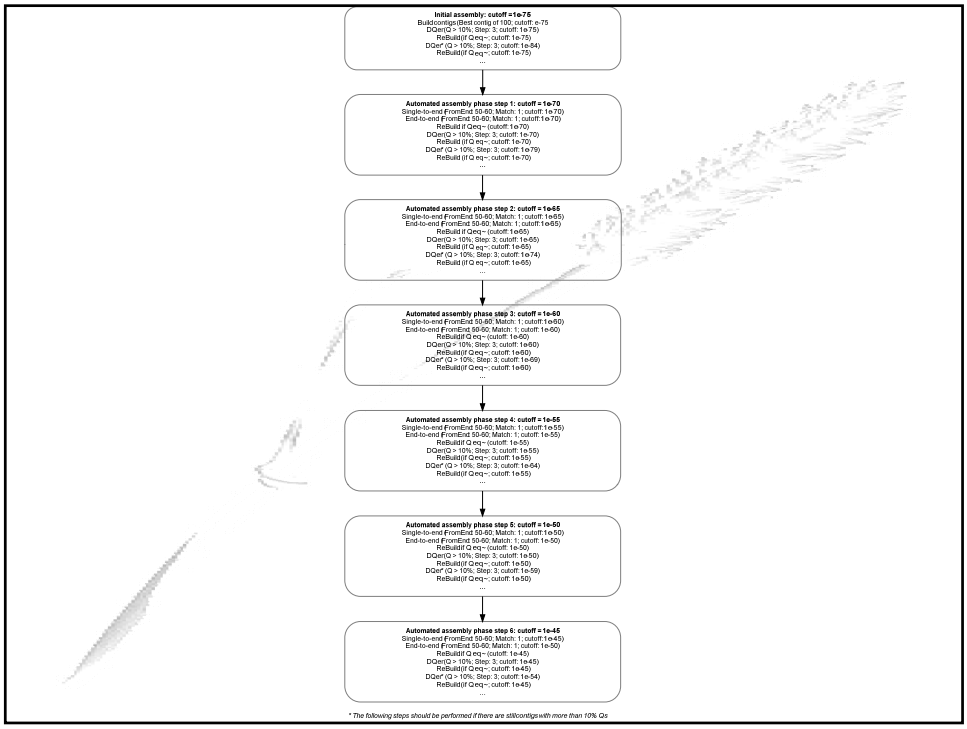
Contig assembly
3- Automated assembly

Assembly of the physical map



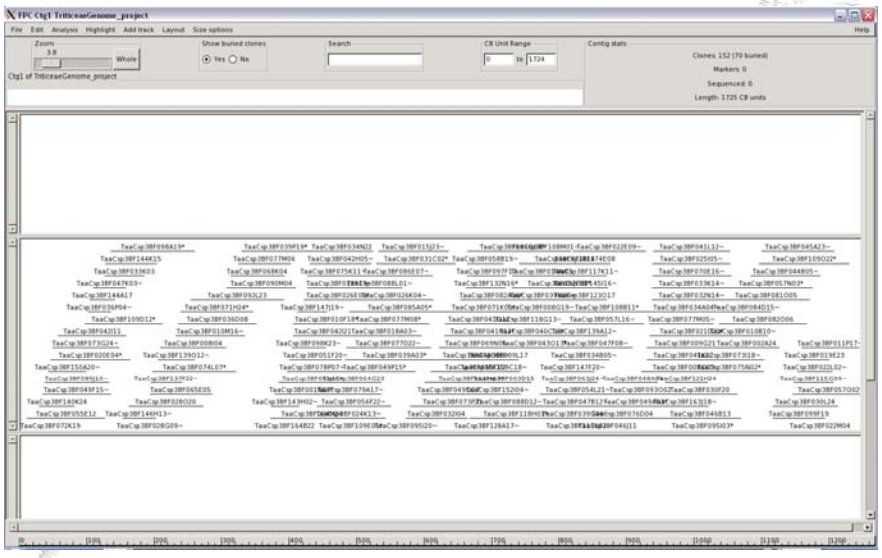
Sulston score overlap



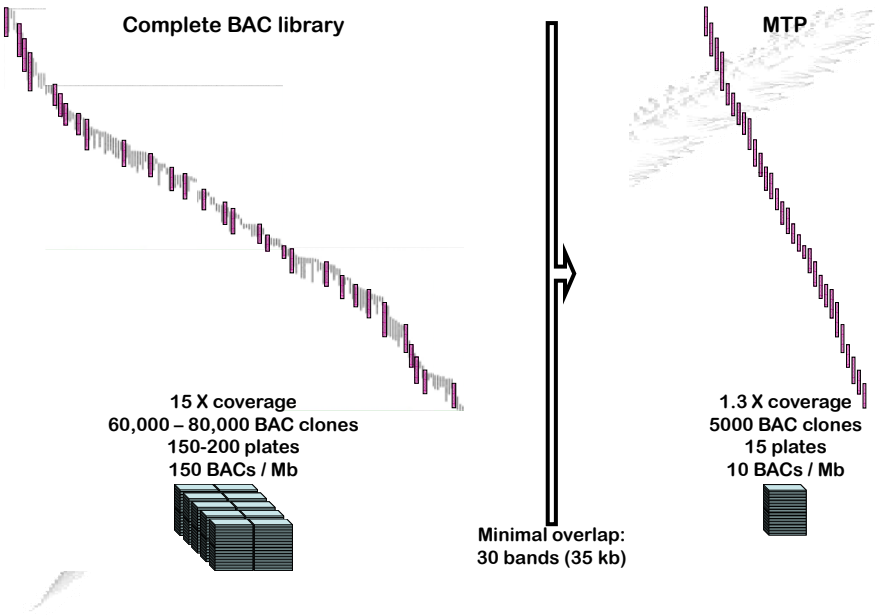


Contig assembly
4- Establishing a Minimal Tiling Path (MTP)

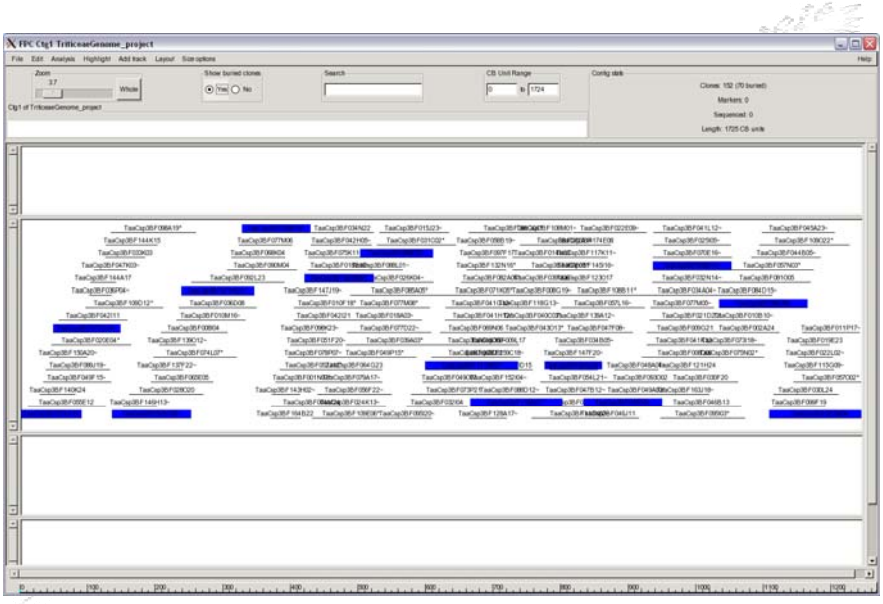
Contig display



Establishment of the Minimal Tiling Path

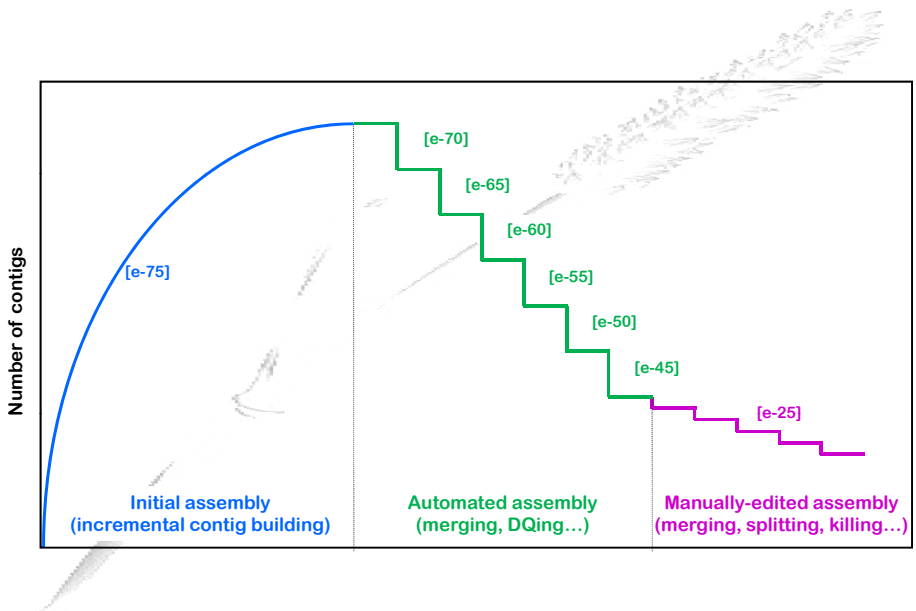


Selecting an MTP



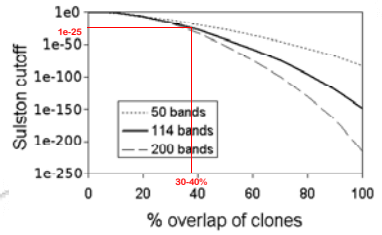
Contig assembly
5- Manually-edited assembly

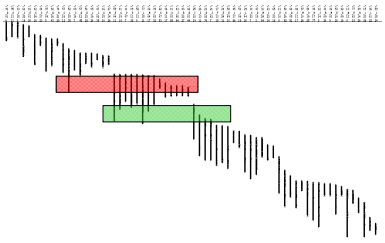
Assembly of the physical map



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)



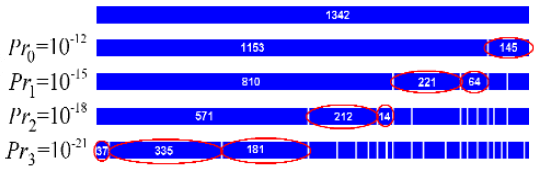


Contig assembly
6- LTC

LTC program

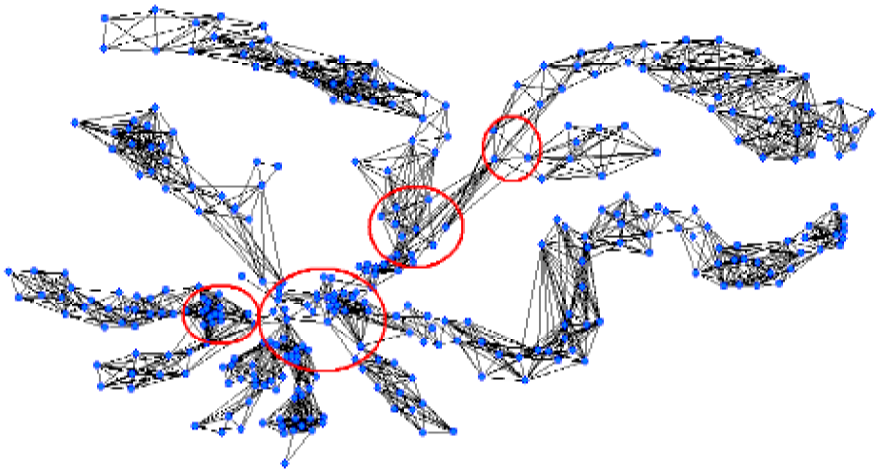
Frenkel *et al.*, (2010) LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. *BMC Bioinformatics*, 11, 584.

- ✓ 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.



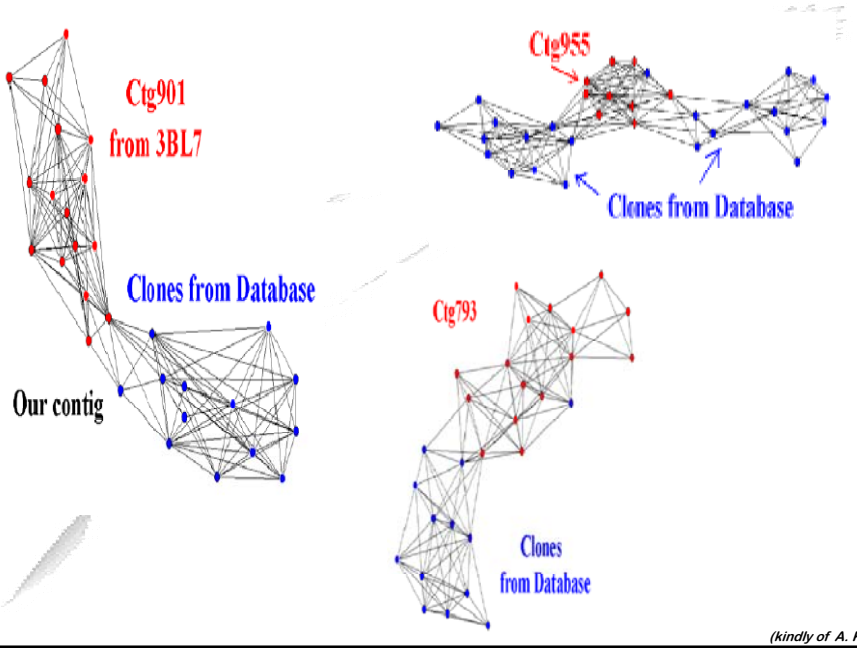
(adapted from Frenkel *et al.*, *BMC Bioinformatics*, 2010)

"Linearization" by removing clones in cluster branching



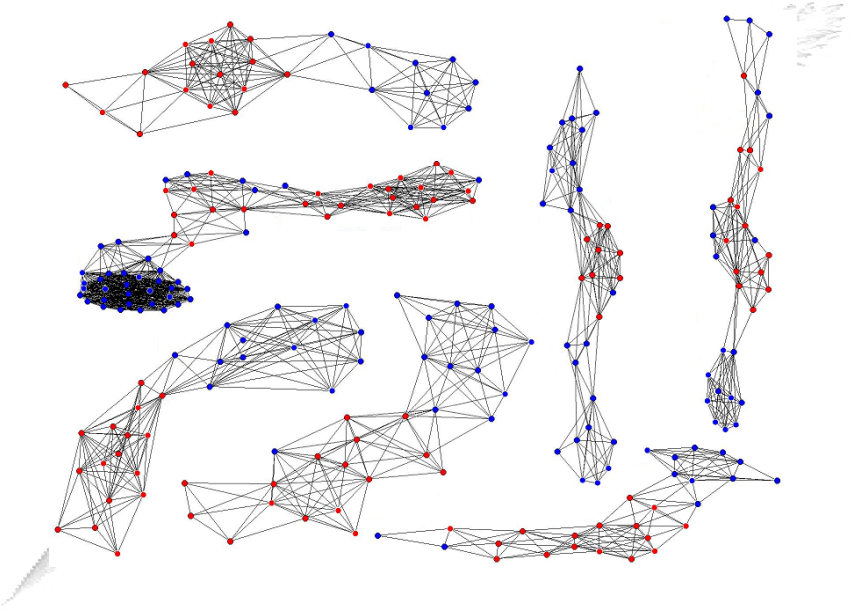
(kindly of A. Korol)

Examples of contig elongation



(kindly of A. Korol)

Examples of *de novo* assembled contigs

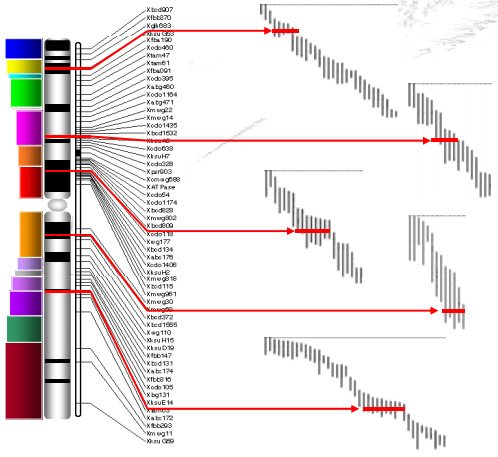


(kindly of A. Korol)

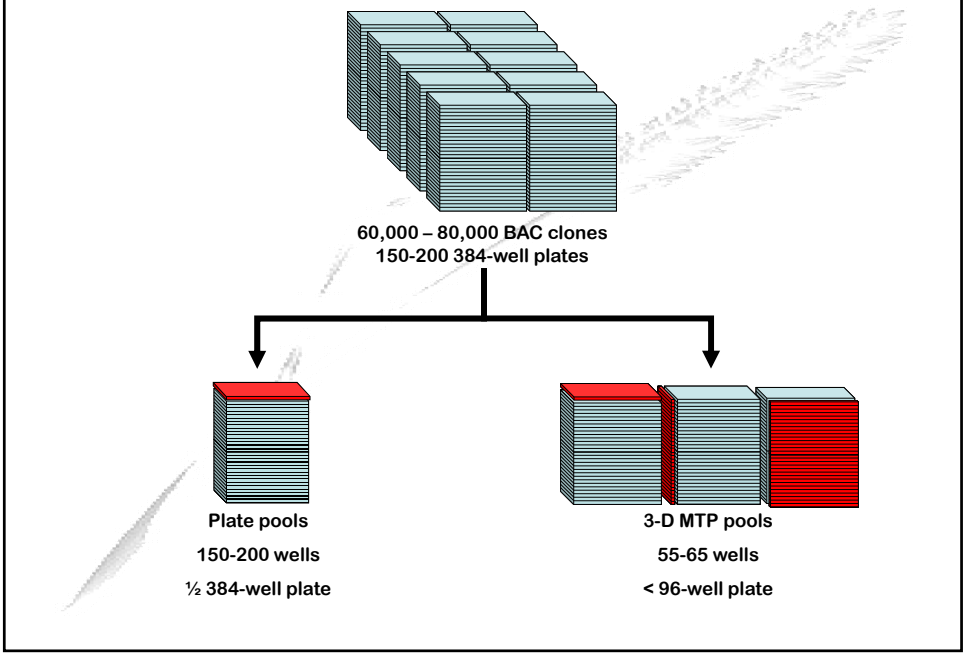
Contig anchoring
1- Forward anchoring

Forward contig anchoring

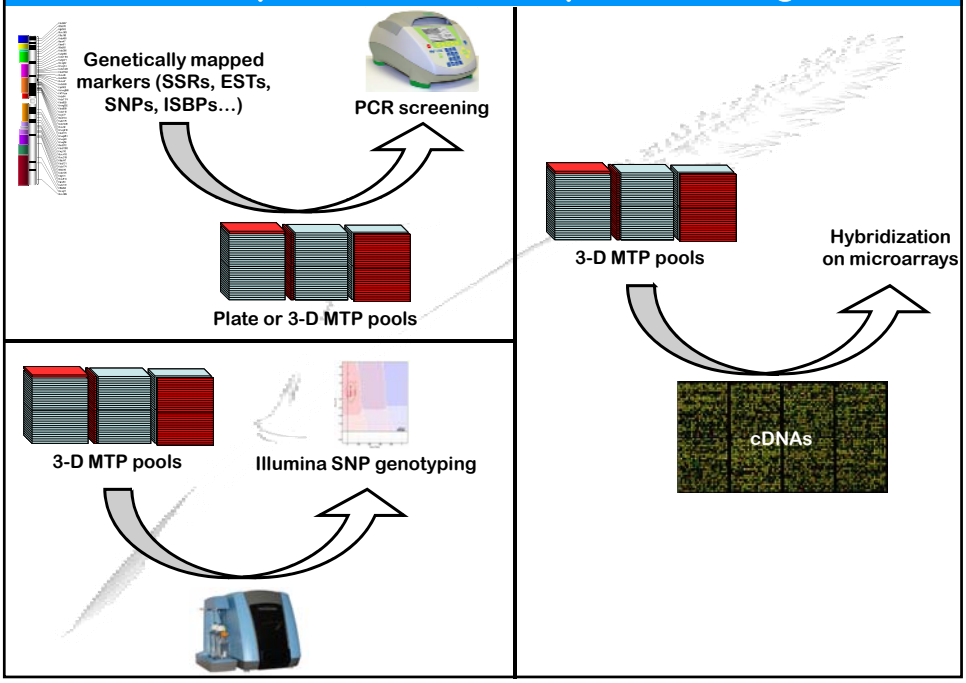
Forward anchoring: from genetic maps to contigs using genetically-mapped markers (SSRs, ESTs, RFLPs, DArTs, SNPs…)



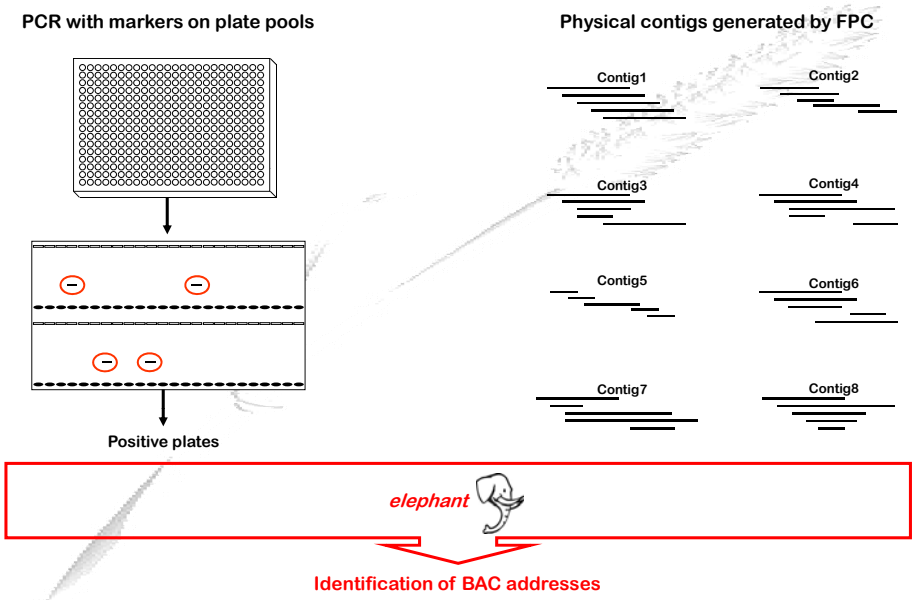
Resources for forward anchoring



Examples of methods for pool screening

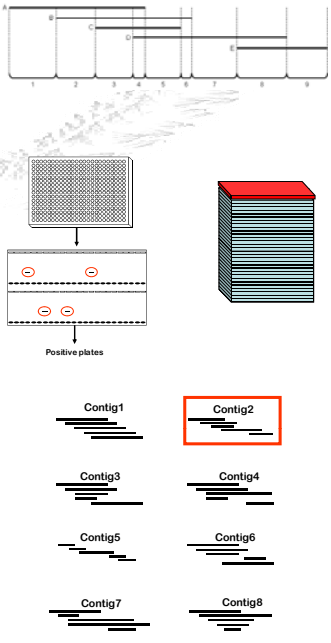


in silico physical map anchoring through pool screening

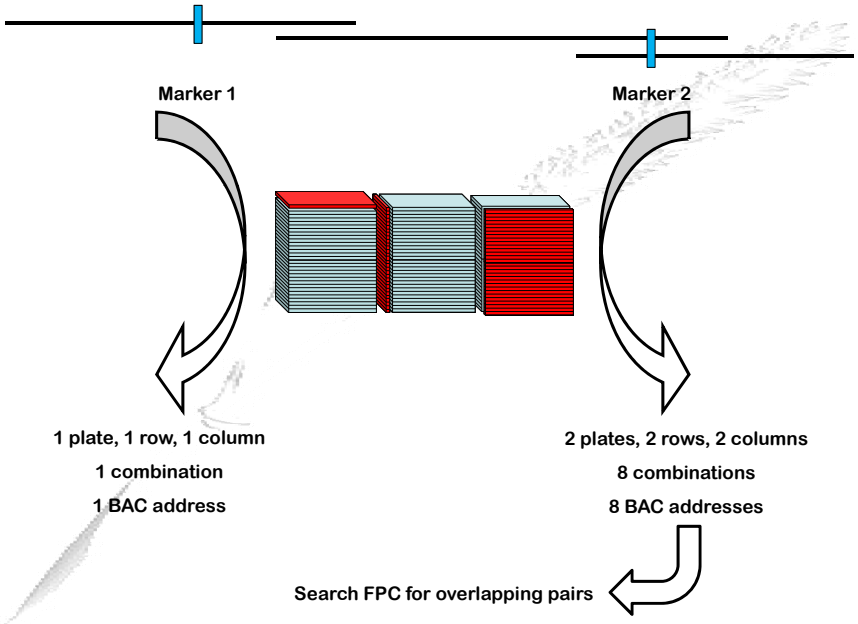


elephant: electronic physical map anchoring tool

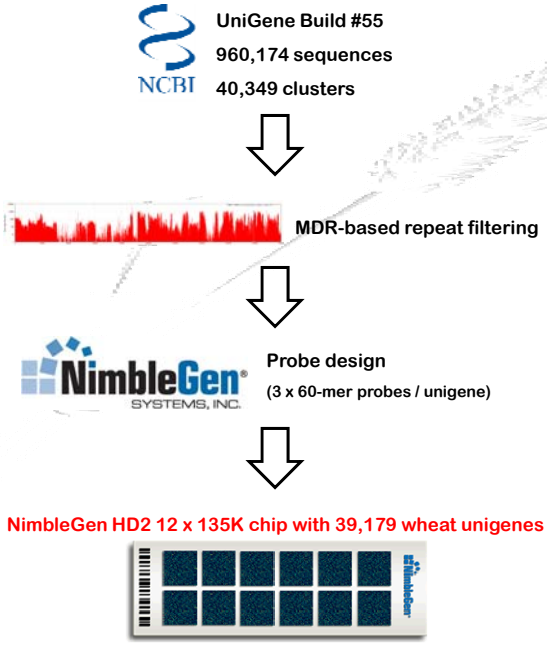
- ✓ Using the FPC file, *elephant* partitions the contigs into short segments, by splitting the contig at each branching point (*e.g.* a clone finishing or joining the assembly) and establishes a list of clones for each segment. Each segment contains a list of overlapping BAC clones that are different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



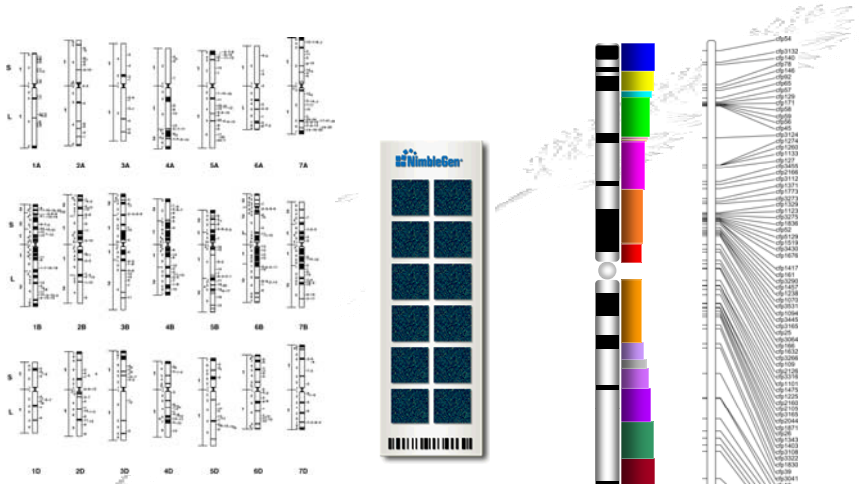
3-D MTP pool screening



The wheat 12x135 K NimbleGen array



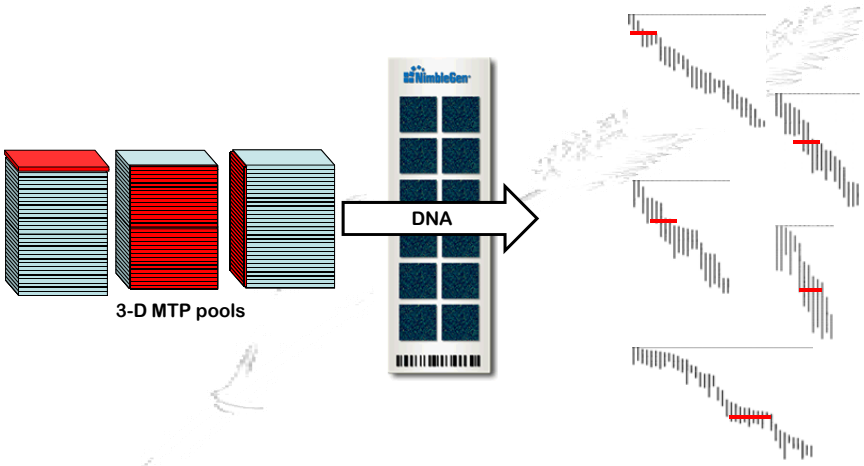
Assigning UniGenes to deletion bins



98.6% of NSF mapped loci
on the NimbleGen chip

Genetic mapping through COS or SNP markers?

Chip-based anchoring



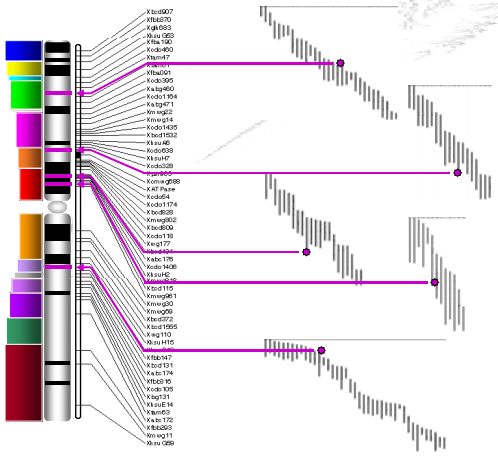
3-D MTP pools

2836 unigenes mapped to chromosome 3B
Chromosome 1BL underway
Chromosomes 1AS and 1BS to come

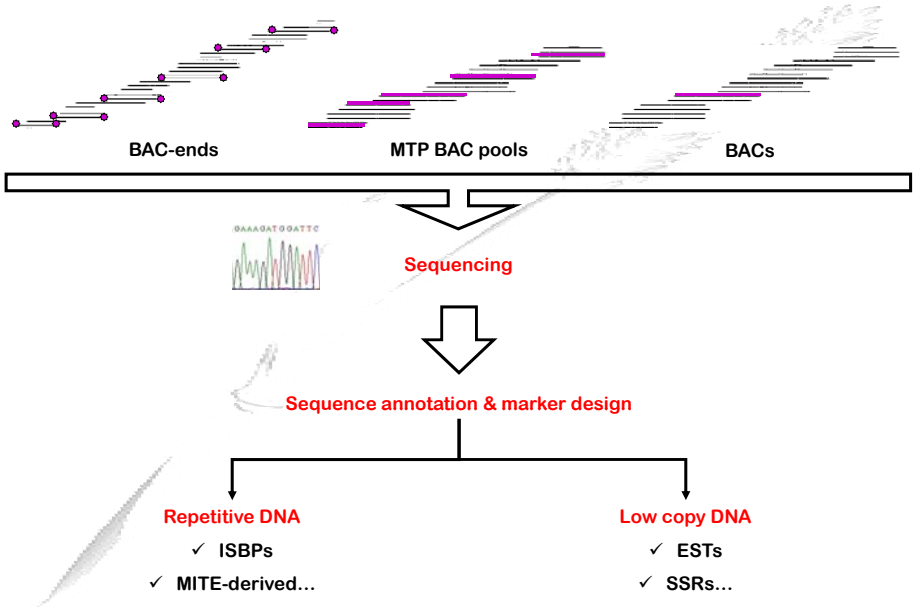
Contig anchoring
2- Reverse anchoring

Reverse contig anchoring

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

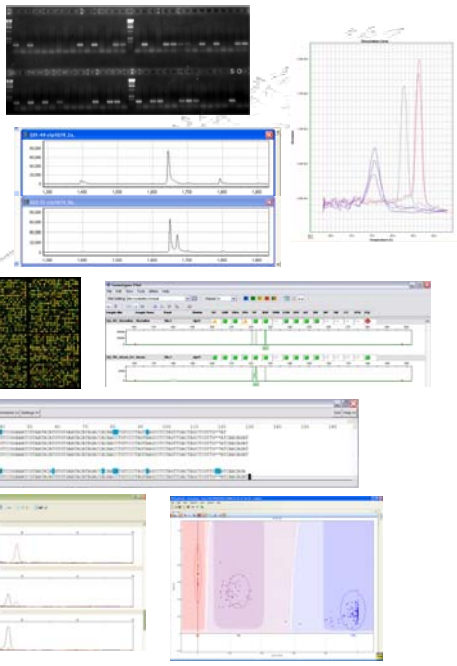


Developing new markers from sequences

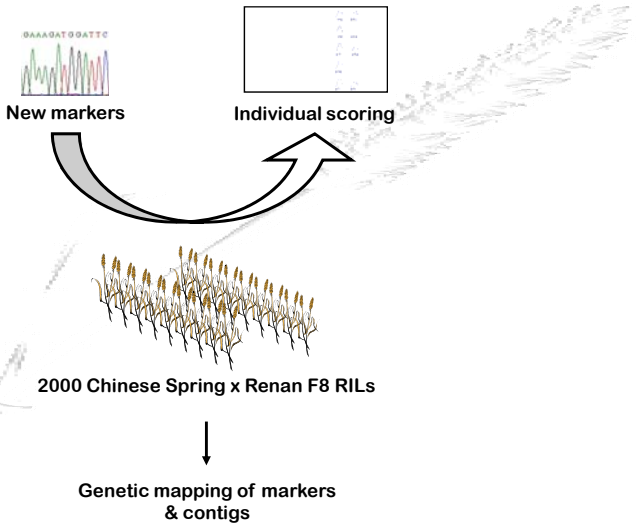


Examples of methods for genetic mapping

- ✓ 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

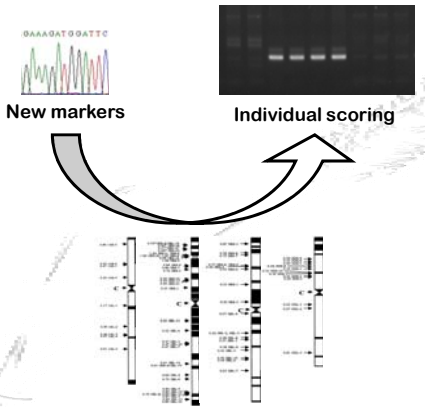
Advantages:

- ✓ Relative order of markers
- ✓ Links to QTLs

Drawbacks:

- ✓ Dependent on recombination
- ✓ Polymorphic markers

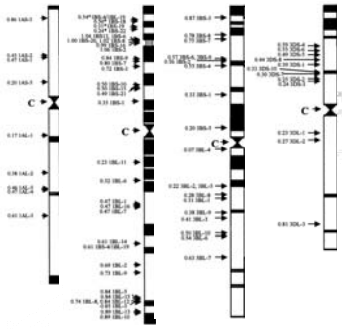
Deletion bin mapping



Deletion bin mapping of markers & contigs

(adapted from Dilbirligi *et al.*, *Genetics*, 2004)

Deletion bin mapping



Advantages:

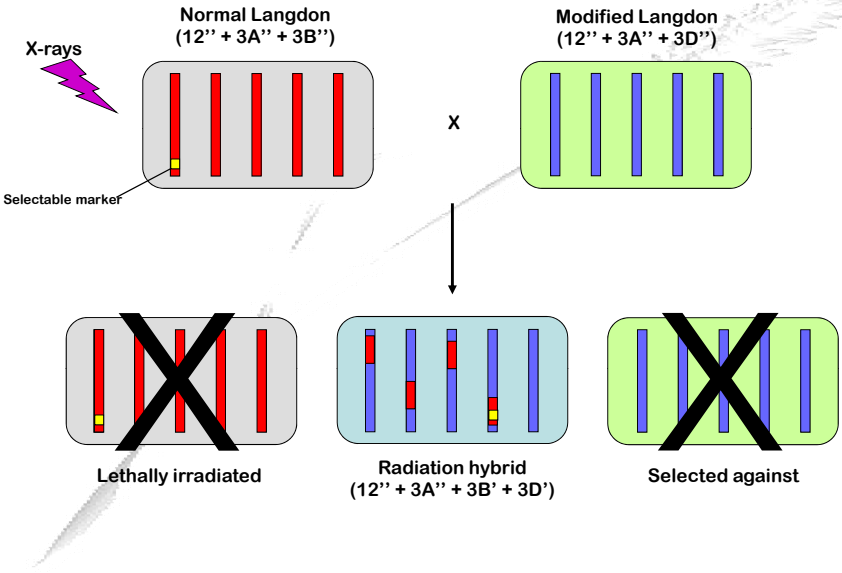
- ✓ Independent on recombination
- ✓ No need for polymorphic markers

Drawbacks:

- ✓ No relative order in bins
- ✓ Large genomic segments

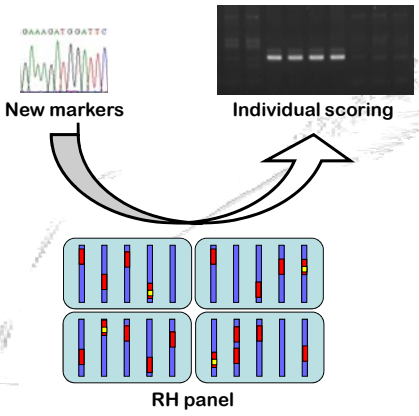
(adapted from Dilbirligi *et al.*, *Genetics*, 2004)

Radiation hybrid mapping

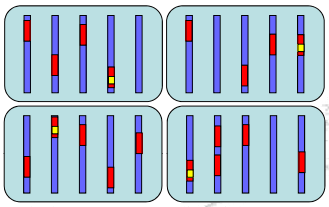


(kindly of S. Kianian)

Radiation hybrid 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

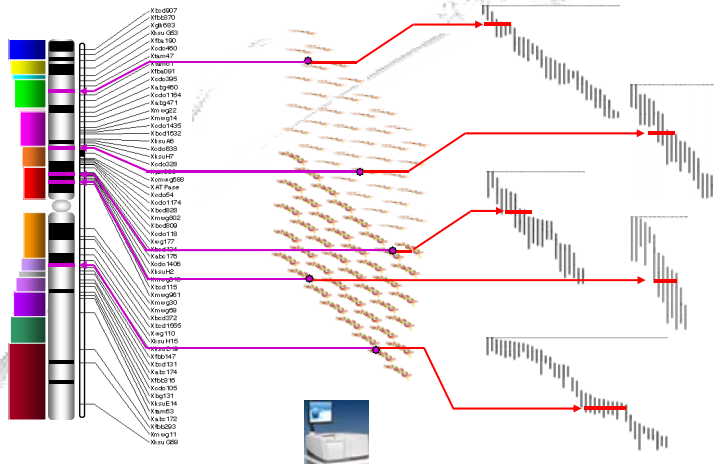
Drawbacks:

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

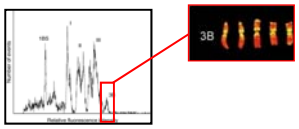
Contig anchoring
3- Hybrid anchoring

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



Flow-sorted chromosome

Next generation sequencing

- ✓ Roche rather than Illumina because of read length
- ✓ low coverage

Annotation and marker design

(SSRs, ESTs, ISBPs…)

BAC screening

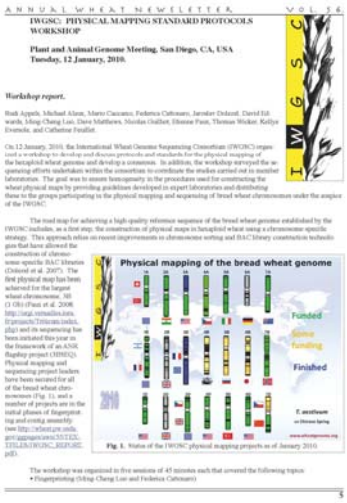
- ✓ Plate pools
- ✓ MTP pools

Genetic mapping

- ✓ Recombination-based mapping
- ✓ Deletion mapping

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

Standard protocols & guidelines


ANNUAL WHEAT NEWSLETTER VOL 56

IWGSC: PHYSICAL MAPPING STANDARD PROTOCOLS WORKSHOP

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

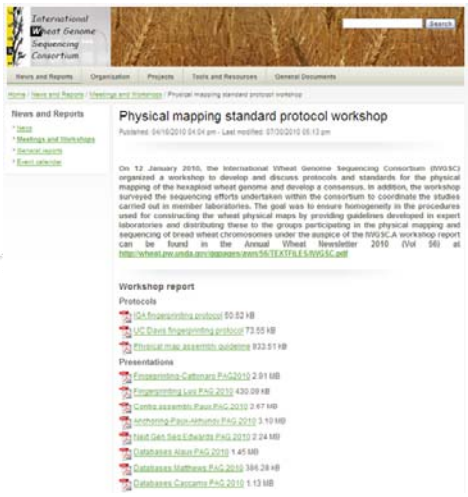
Workshop report.

Physical mapping of the bread wheat genome

Funded
Some funding
Finished

Annual Wheat Newsletter 2010

http://wheat.pw.usda.gov/ggpages/awn/56/TEXTFILES/IWGSC.pdf


International Wheat Genome Sequencing Consortium

News and Reports

Physical mapping standard protocol workshop

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 auspice of the IWGSC.A workshop report can be found in the Annual Wheat Newsletter 2010 (Vol 56) at http://wheat.pw.usda.gov/ggpages/awn/56/TEXTFILES/IWGSC.pdf

Workshop report

Protocols

Presentations

IWGSC

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

Standard protocols & guidelines


Triticeae Genome

Guidelines for physical map assembly

1. BAC naming convention:

TriticeaeGenome

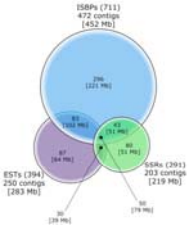
http://www.triticeaegenome.eu/page.php?optim=Deliverables

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

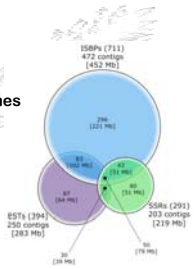
To be discussed…

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…

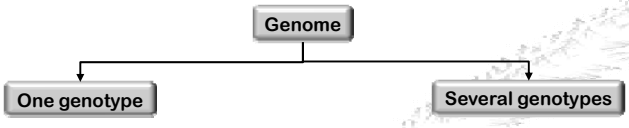
IWGSC Workshop
April 8, 2011
Prague, Czech Republic

Sequence-enabled marker development in hexaploid wheat

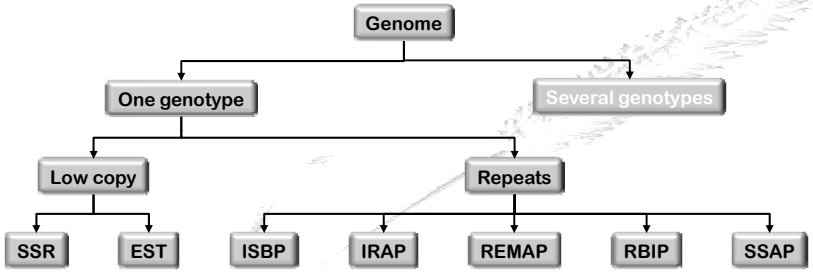
Etienne Paux

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

A decision tree for marker development in wheat



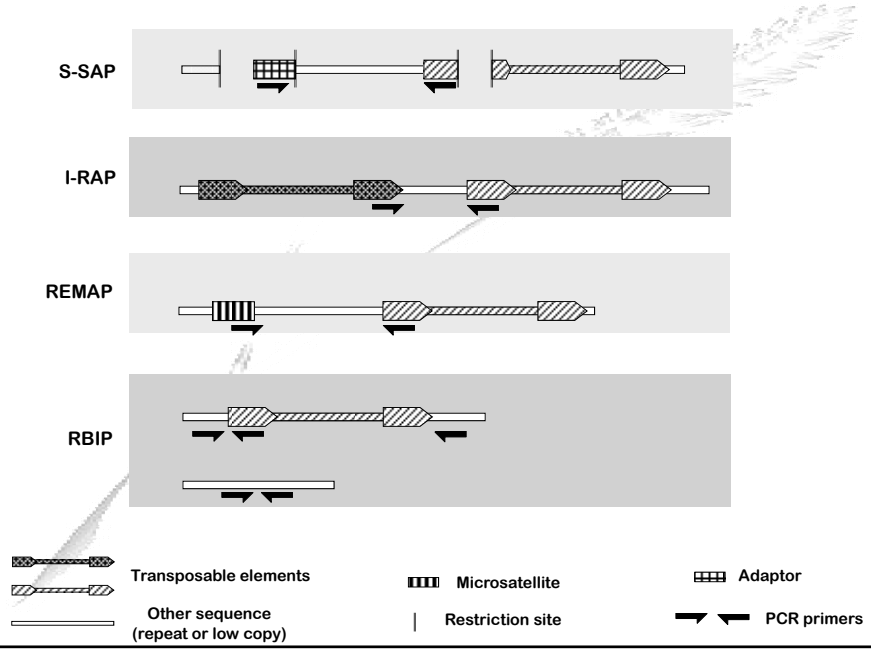
A decision tree for marker development in wheat



SSR and EST markers

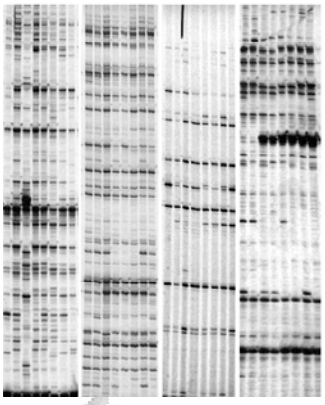
- SSRs
  - ✓ Two categories
    - ❖ Genomic DNA → genomic SSRs (gSSRs)
    - ❖ ESTs → EST-SSRs
  - ✓ Origin: Transcriptome, enriched libraries, BES or whole genome shotgun
  - ✓ Detection tools: SciRoKo, MISA, SSRFinder, SSRIT, TRF, TROLL, Sputnik, SSRsearch, SSR Locator, SSRPrimer…
  - ✓ Detection technique: capillary electrophoresis
- ESTs
  - ✓ More than one million wheat ESTs in databases
  - ✓ Chips (Affymetrix, Agilent, NimbleGen)
  - ✓ Detection techniques: HRM, SSCP, SFP
  - ✓ Conserved Ortholgous Set markers (presence of an intron to increase the polymorphism level)

Repeat-based markers



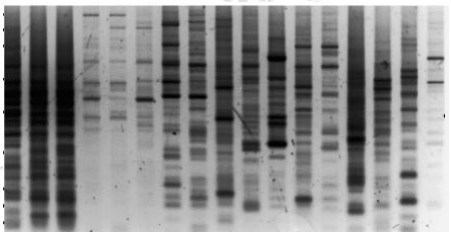
Sequence-Specific Amplified Polymorphism, Inter-Retrotransposon Amplified Polymorphism and Retrotransposon Microsatellite Amplified Polymorphism

S-SAP



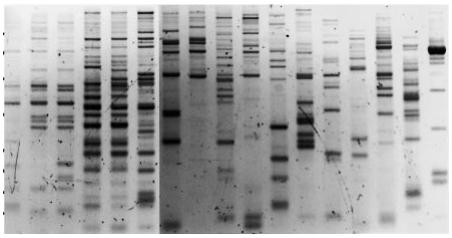
(adapted from Queen *et al.*, *Mol Gen Genomics*, 2004)

I-RAP



(adapted from Kalendar *et al*, Theor Appl Genet 1999)

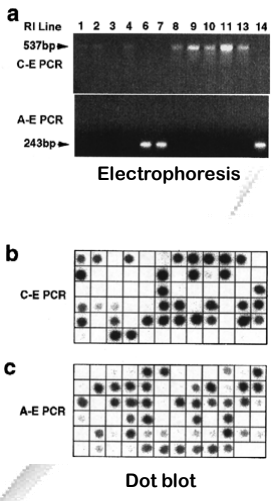
REMAP



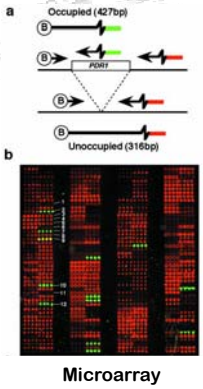
(adapted from Kalendar *et al*, Theor Appl Genet 1999)

Retrotransposon-Based Insertional Polymorphism and Tagged Microarray Markers

RBIP



TAM

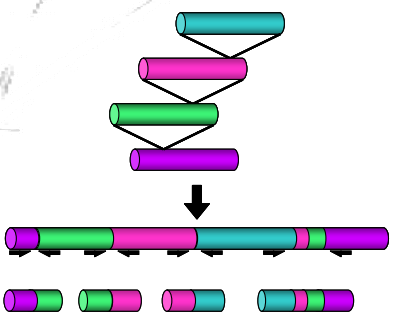


(adapted from Flavell *et al*, Plant J 1998)

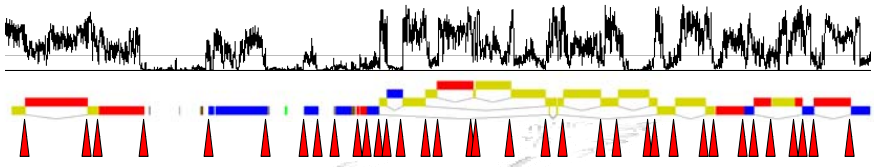
(adapted from Flavell *et al*, Nucleic Acids Res 2003)

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

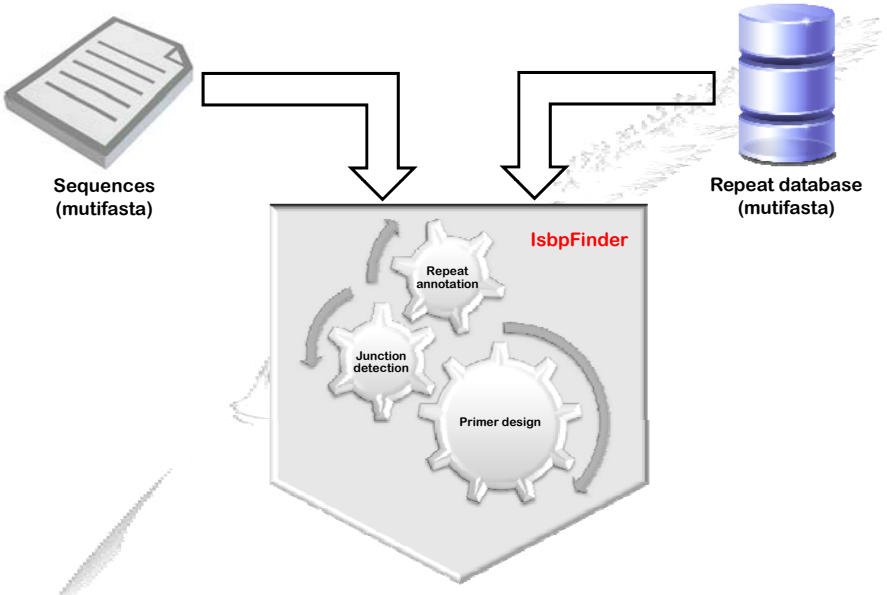
70% genome-specific markers

One ISBP every 5.4 kb

17-Gb genome

More than 3 million ISBPs in the whole hexaploid wheat genome

IsbpFinder

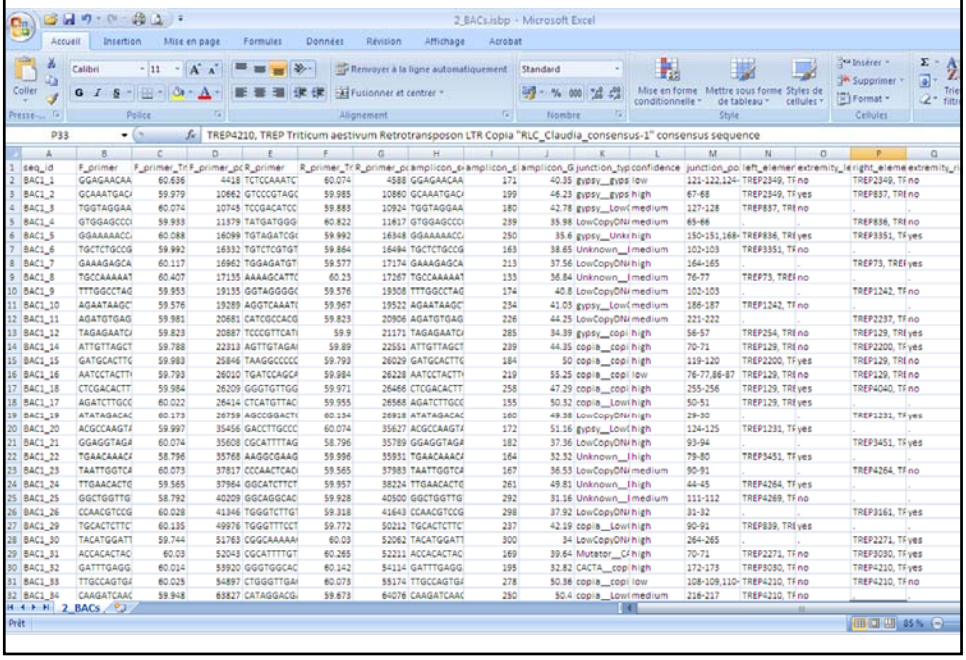


Putative ISBP markers
(table and embl files)

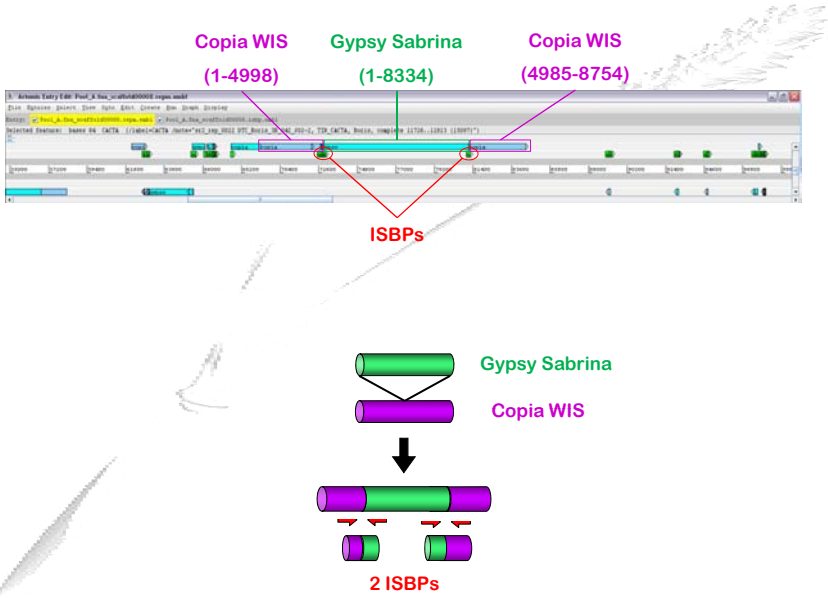
IsbpFinder command line

```
### IsbpFinder ###
#
# AUTHOR:     Frederic CHOULET
# 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.xm  -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 extrimity 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
```

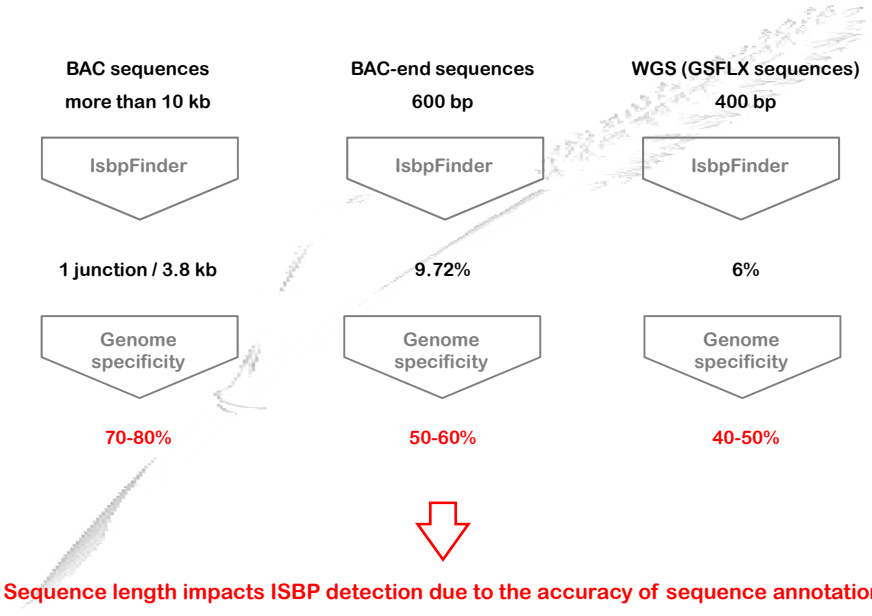
IsbpFinder output table



Visualization of IsbpFinder results under Artemis

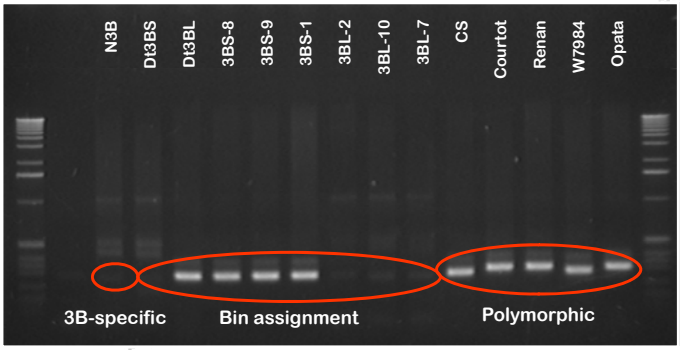


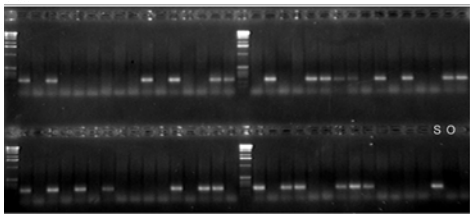
ISBP design and genome-specificity



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

Agarose gel electrophoresis

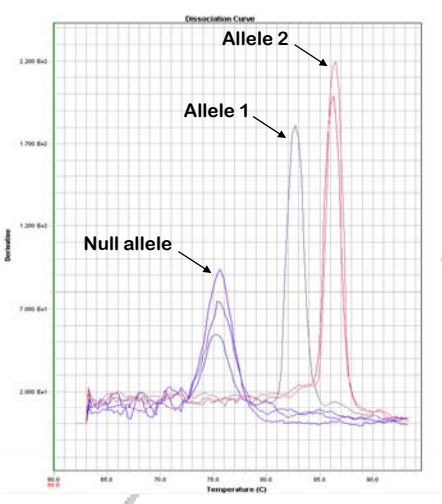


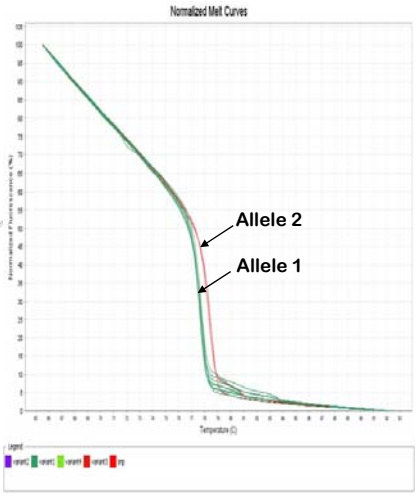


Fluorescent capillary electrophoresis

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

Melting curve-based genotyping





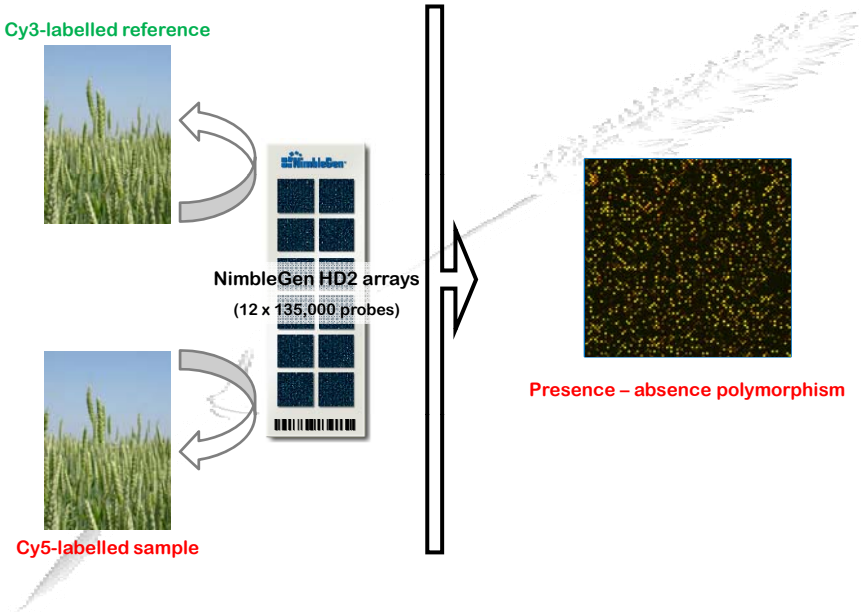
Melting curve analysis (MCA)

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

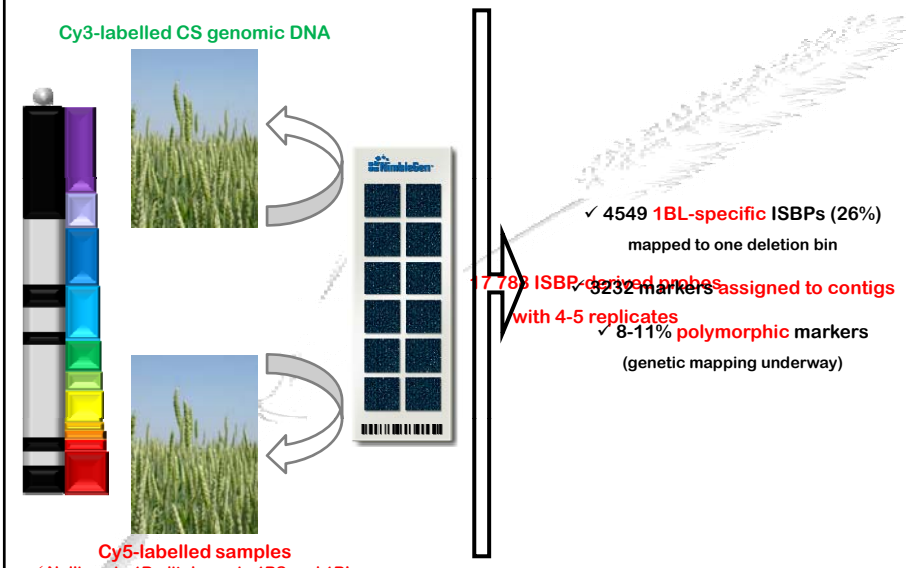
High resolution melting (HRM)

- ✓ SNP genotyping
- ✓ SNP discovery

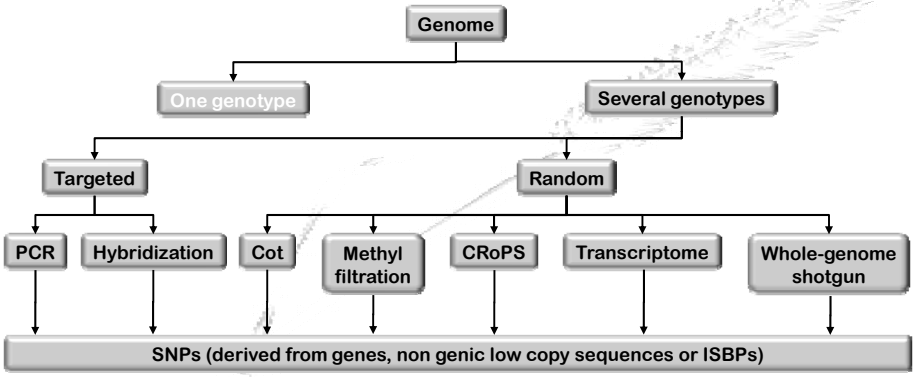
IMaGe: ISBP Microarray-based Genotyping



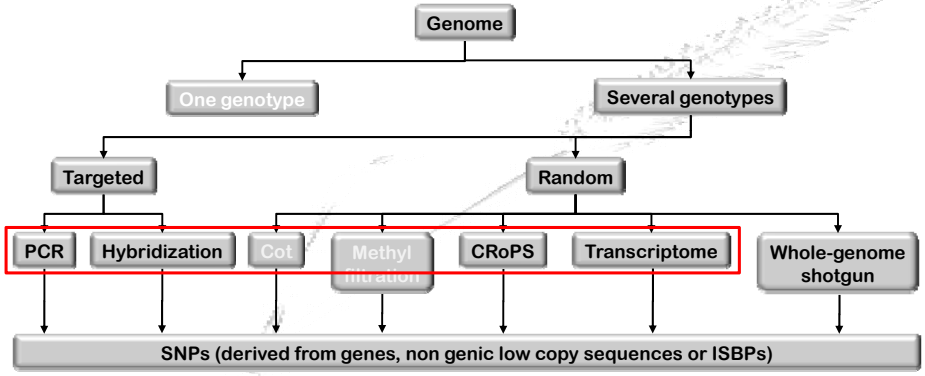
The 1BL IMaGe experiment



A decision tree for marker development in wheat

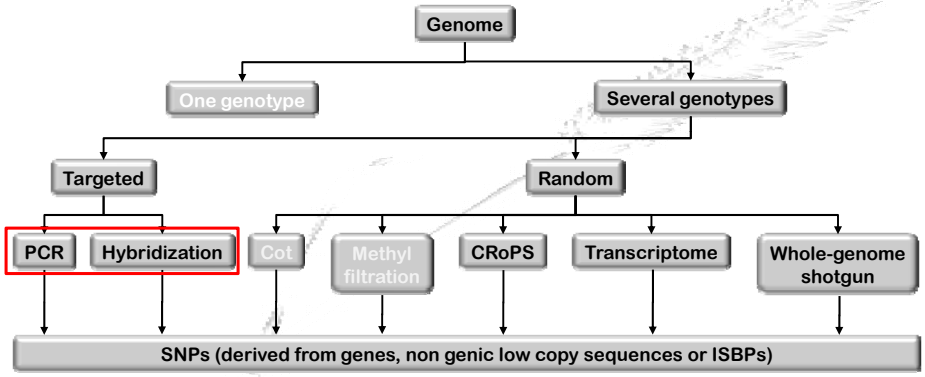


A decision tree for marker development in wheat



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

A decision tree for marker development in wheat



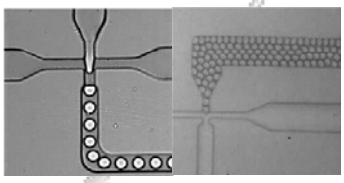
Prerequisite: contiguous or discrete target regions clearly identified

PCR-based target enrichment

Simplex or multiplex PCR

Microfluidic PCR

(Fluidigm Access Array system)



Microdroplet PCR

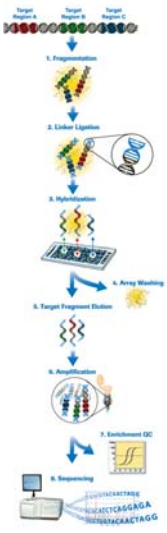
(RainDance Technologies)

Sequencing

(Roche, Illumina, Applied…)

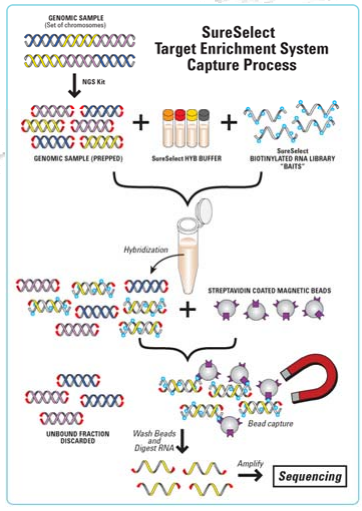
Hybridization-based target enrichment

Solid phase



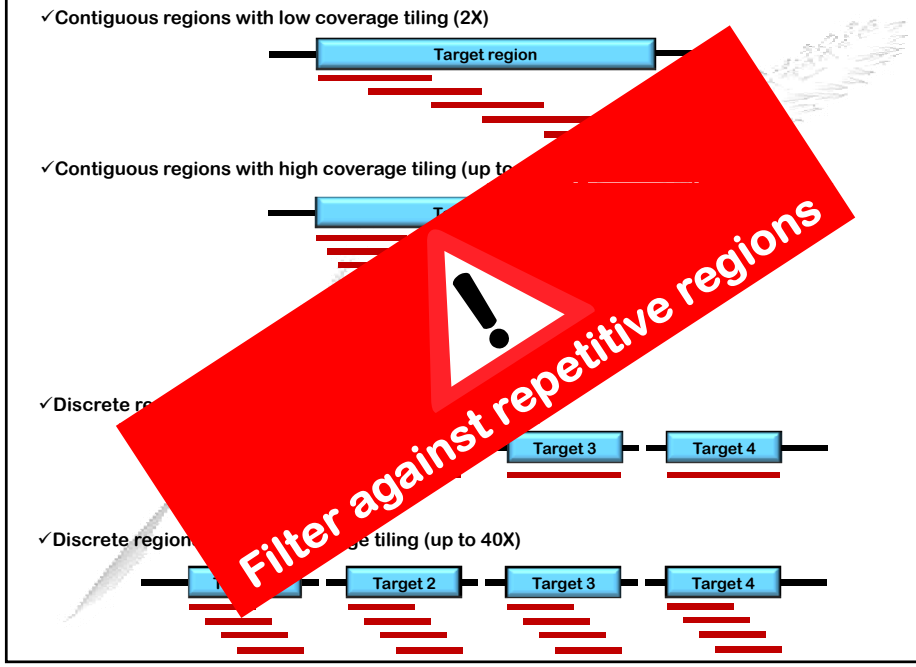
(NimbleGen, Agilent, Febit…)

Liquid phase

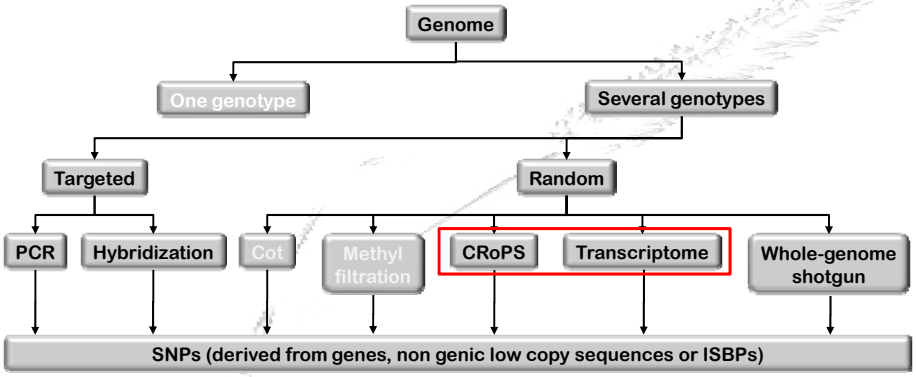


(Agilent, NimbleGen, Illumina…)

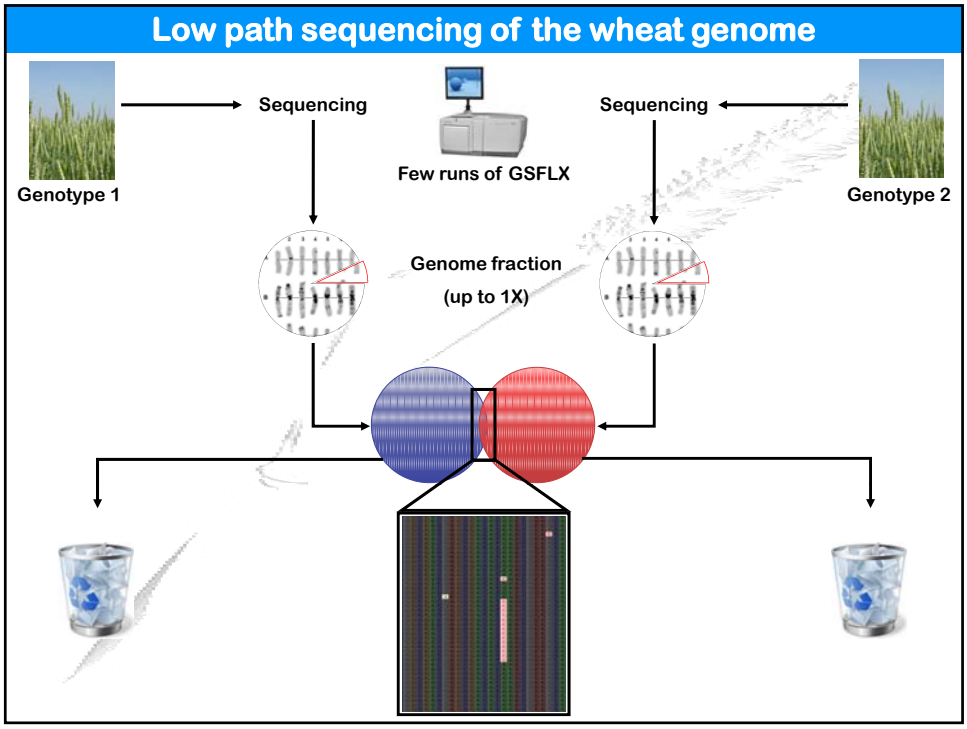
Sequence capture design

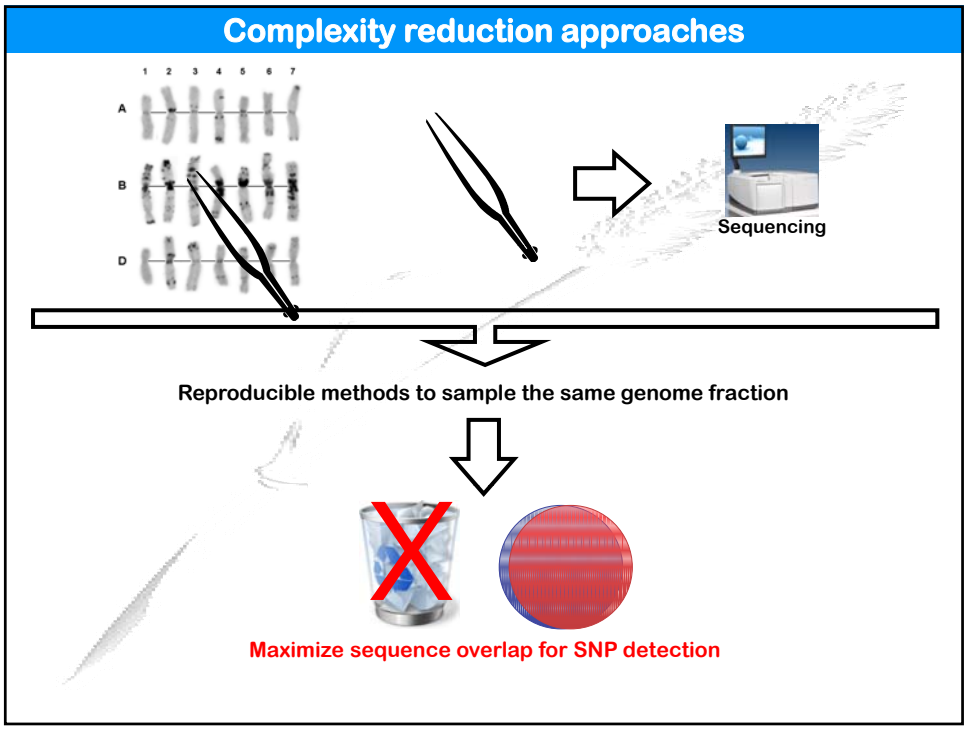


A decision tree for marker development in wheat

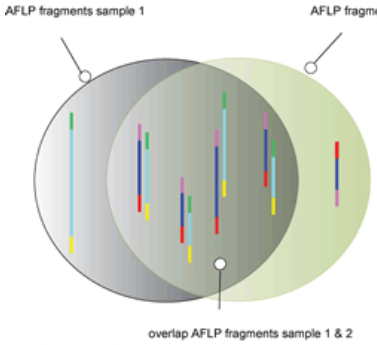


Prerequisite: reproducible protocols to reduce complexity
and sample the same fraction of the genome between different genotypes


Low path sequencing of the wheat genome
Genotype 1
Sequencing
Few runs of GSFLX
Sequencing
Genotype 2
Genome fraction
(up to 1X)

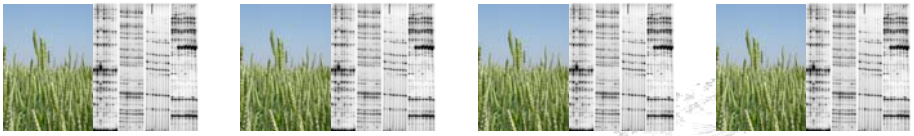

Complexity reduction approaches
1 2 3 4 5 6 7
A
B
D
Sequencing
Reproducible methods to sample the same genome fraction
Maximize sequence overlap for SNP detection

Complexity Reduction of Polymorphic Sequences (CRoPS)

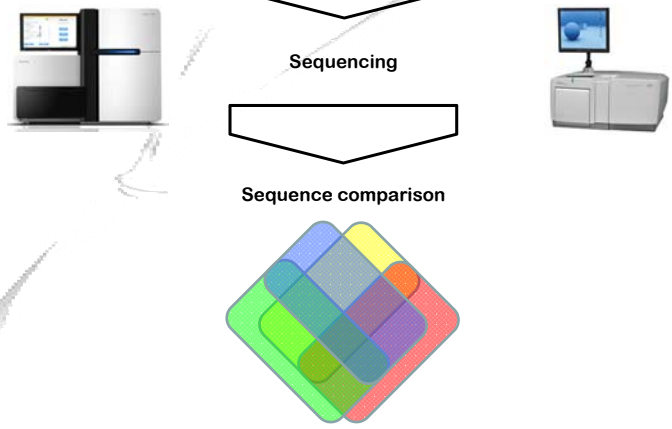


Sequencing
(Roche, Illumina, Applied…)

Sequencing SSAP markers



Genotype 1 | Genotype 2 | Genotype 3 | Genotype 4



Sequencing

Sequence comparison

Transcriptome sequencing


Genotype 1

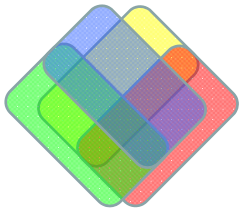

Genotype 2


Genotype 3

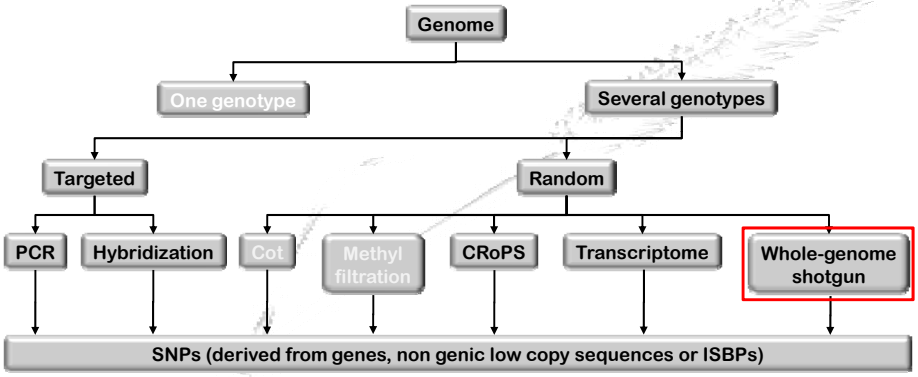

Genotype 4

RNA extraction and sequencing

Sequence comparison

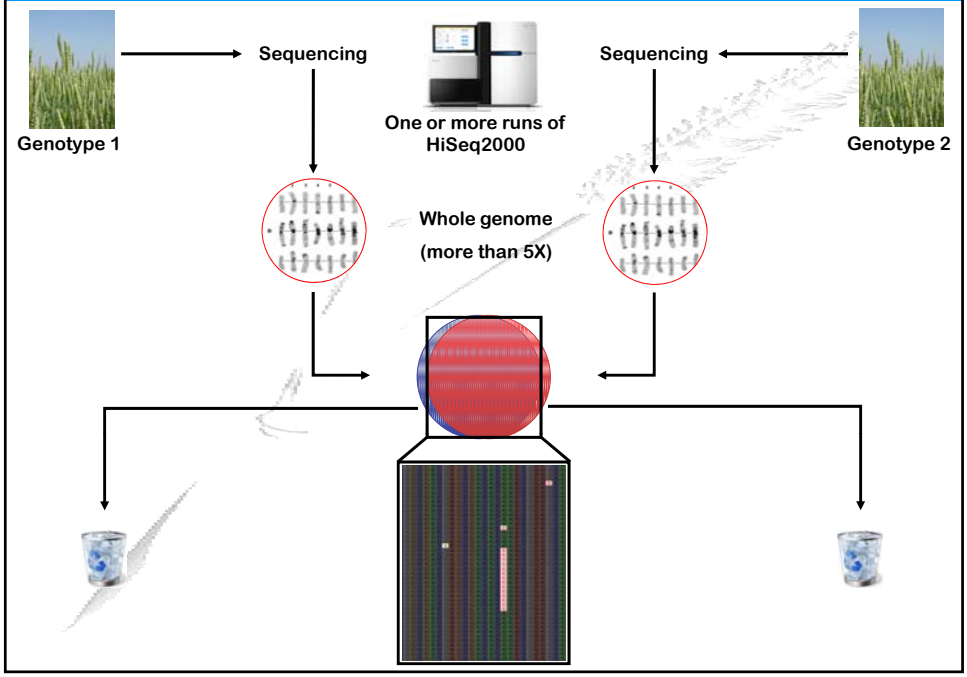


A decision tree for marker development in wheat

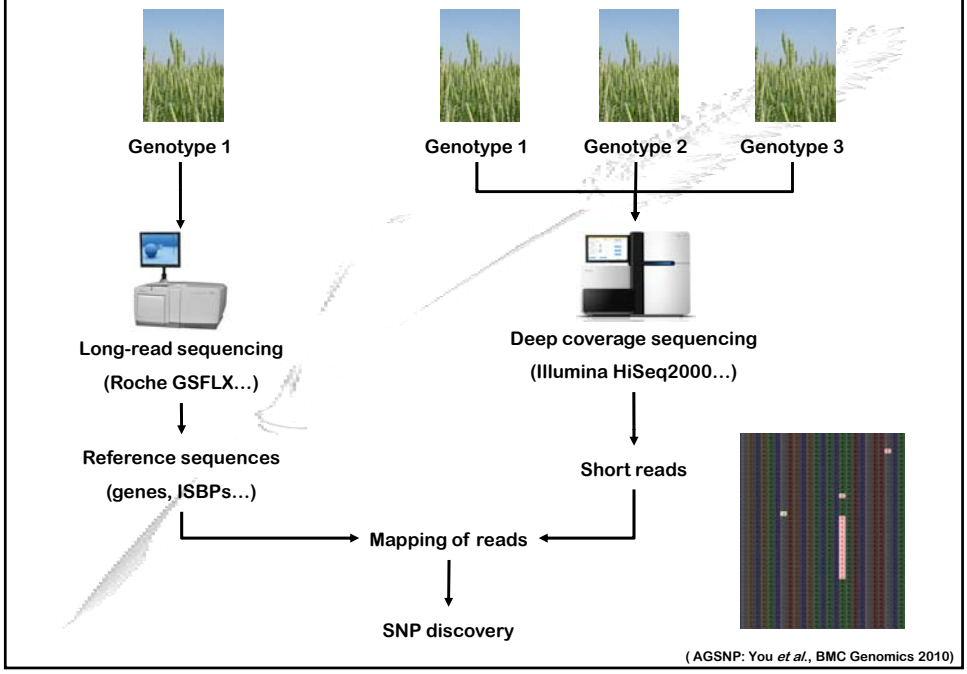


Take advantage of the second- or third-generation sequencer capacities

Deep coverage whole genome sequencing in wheat



Deep coverage whole genome sequencing in wheat



(AGSNP: You *et al.*, BMC Genomics 2010)

International wheat SNP initiatives

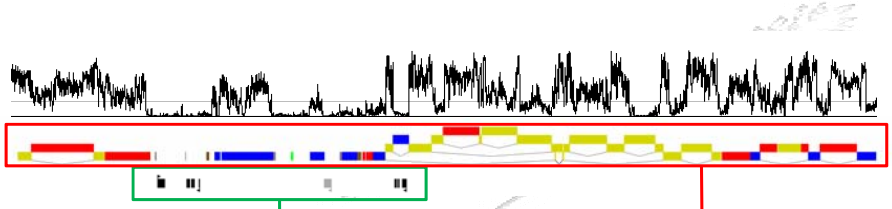
- ✓ Single Nucleotide Polymorphism (SNP) markers for high-throughput genotyping to advance genomic, genetic and breeding research in wheat (Eduard Akhunov, USA)
- ✓ Australian SNP discovery and genotyping project (Matt Hayden, Australia)
- ✓ Haplotype Polymorphism in polyploid wheats and their Diploid 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

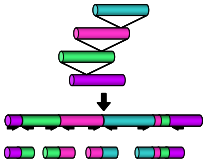


Genic fraction: 1-2%
Gene density: 1 / 104 kb
SNP density in gene: one SNP / 250-500 bp

Repetitive fraction: 90%
ISBP density: 1 / 5-10 kb
SNP density in ISBP: 1/ 100 bp

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

ISBP markers as a source of SNPs



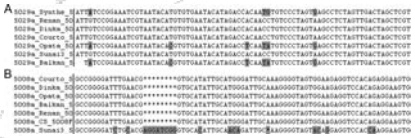
157 ISBPs x 8 wheat lines

Sequencing

105 polymorphic ISBPs

✓401 SNPs

✓19 IDPs

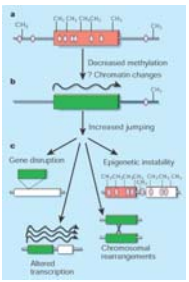


One SNP every 99 bp

(*vs.* 1 SNP every 250-500 bp in genes)

More than 6 million ISBP-derived SNPs in the whole hexaploid wheat genome

TE methylation and SNP frequency



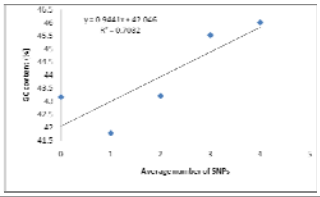
(adapted from Symer and Bender, 2001)

✓ Transposable elements are methylated to prevent uncontrolled movement

✓ Methylation leads to an increase in mutation frequency at deaminated sites (C/T transition)

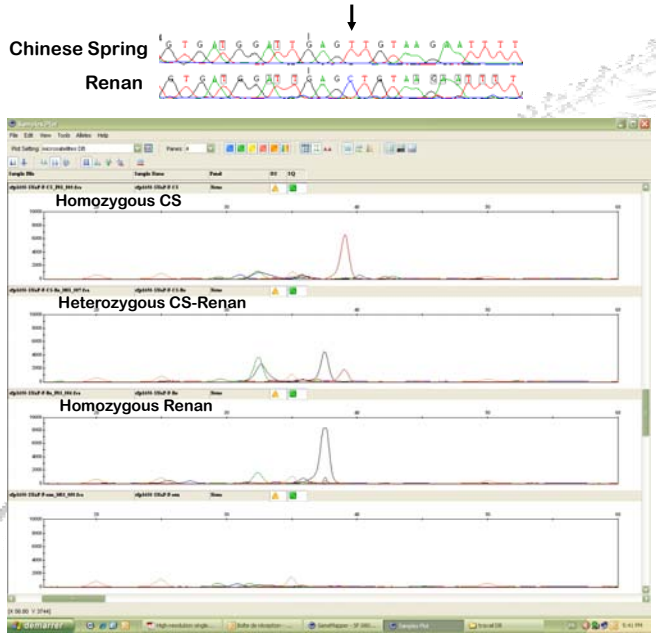
5-methylcytosine thymine

The SNP density is correlated with the GC content of the sequence



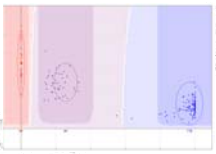
(adapted from Paux et al, Plant Biotech J 2010)

SNaPshot



Illumina VeraCode genotyping

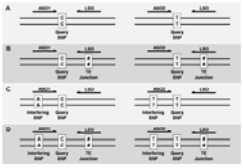
ISBP-derived SNPs



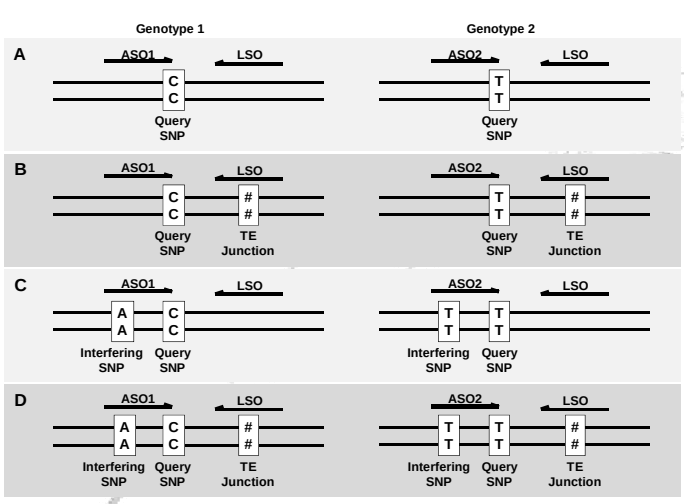
62% scorable markers

✓Negative effect of interfering SNPs (43% scorable)

✓Positive effect of the TE junction (91% scorable)



Illumina VeraCode genotyping

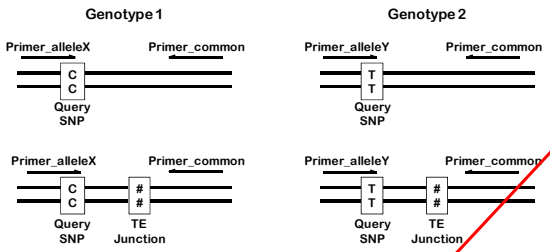


✓ No interfering SNP

✓ SNP close to the junction

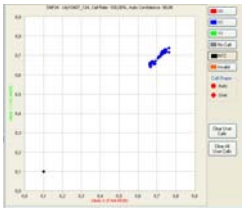
ISBP-SNP genotyping using KASPar probes

➢ First test



	Total	Successful	Rate
	35	11	31%
	13	13	100%
Overall	48	22	50%

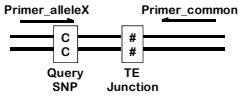
14 SNPs that failed ➔ new design with primers on both sides of the junction

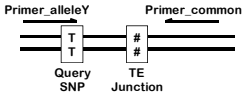


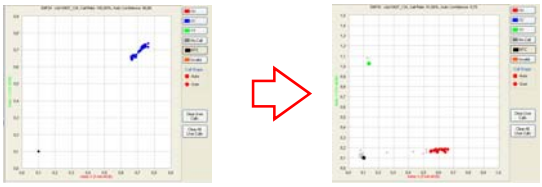
ISBP-SNP genotyping using KASPar probes

➢ Second test

	Total	Successful	Rate
	14	13	93%

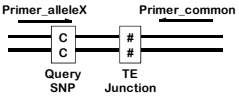


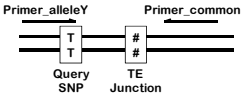




➢ Overall

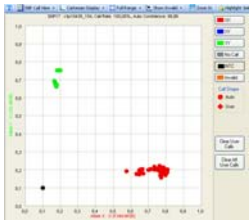
	27	26	96%

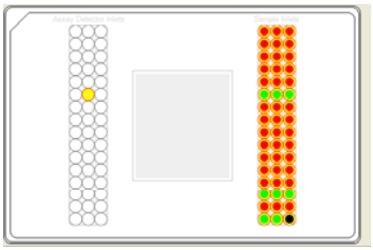




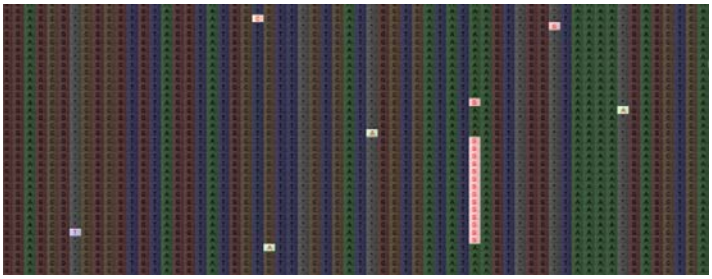
ISBP-SNP genotyping using KASPar probes

➢ Reproducibility: 100%

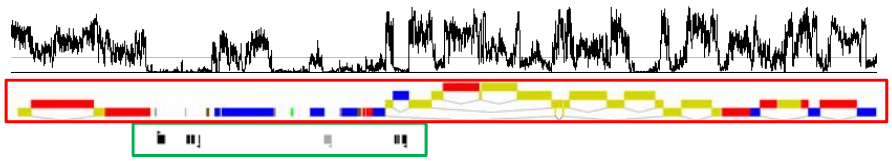




➢ Consistency with sequencing data: 100%



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