Whole Genome Profiling (WGP) for physical

mapping of wheat chromosome 6A

Naser Poursarebani

Poursarebani@ipk-gatersleben.de

PBP Group, IPK-Gatersleben, Germany

IWGSC Workshop - 2012

Outline:

- **Introduction**

  - Importance of wheat for German agriculture,
  - Why wheat chromosome 6A?
  - And the important genes localized on this wheat chromosome

- **Methodologies of making a BAC based physical map**

  - High Information Content Fingerprinting (HICF) based physical map

  - Whole Genome Profiling (WGP^{TM}) based physical map (KeyGene Company)

- **Preliminary results _ WGP based 6A assembly**

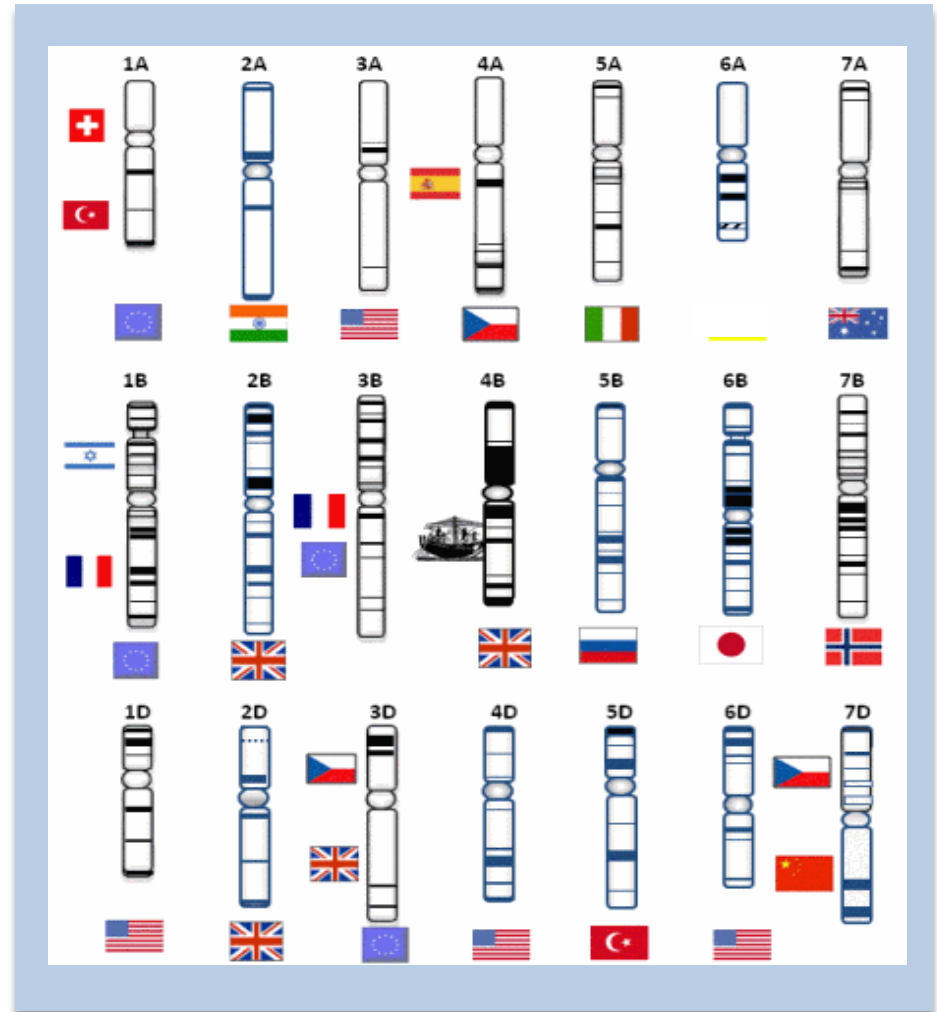- **Anchoring resources**

- **Summary and Outlook**

Wheat in Germany!

Wheat is the number 1 crops regarding harvested acreage

cultivated are 3.2 mio ha, 7.2 t/h

Why Chromosome 6A
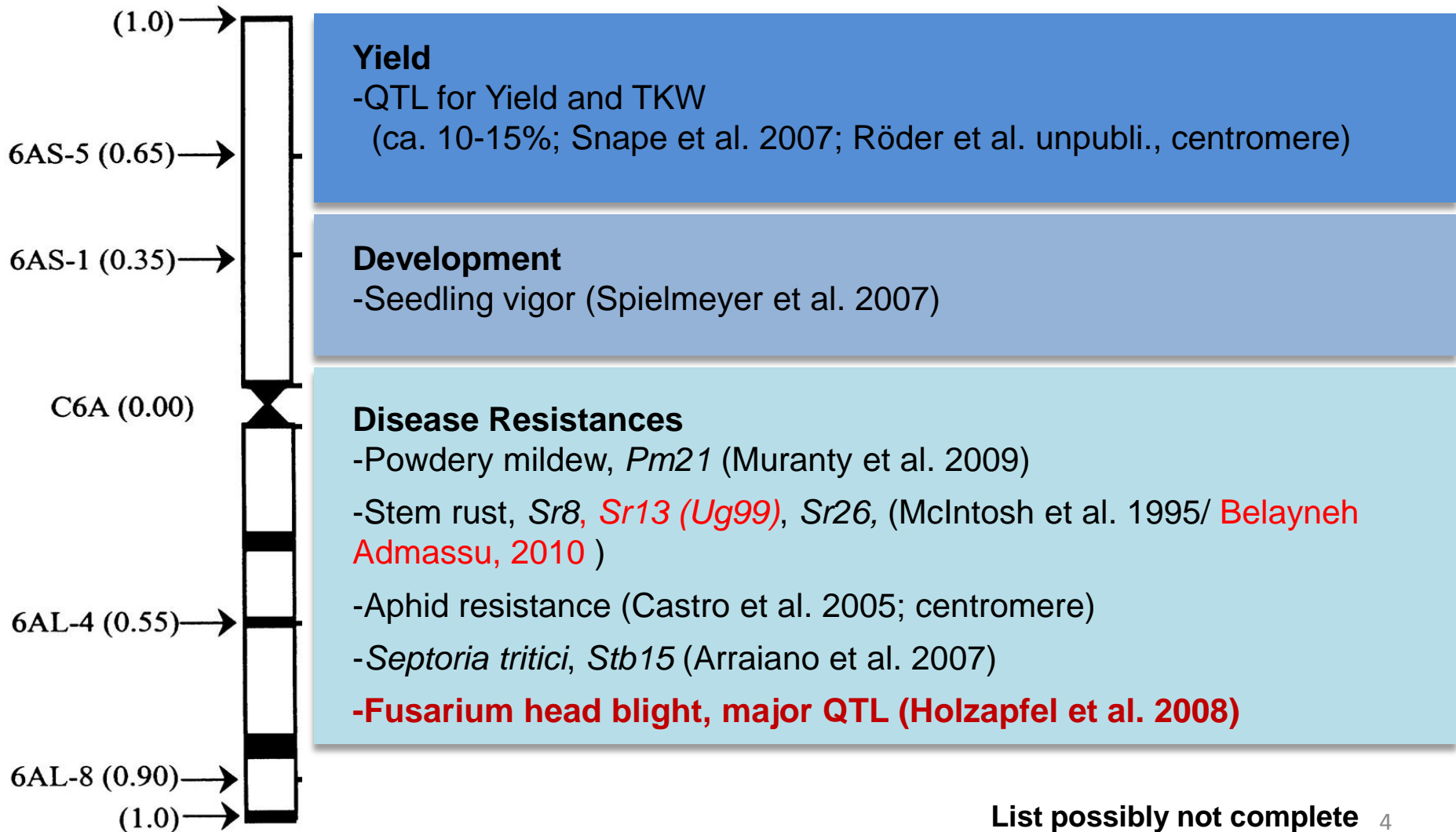
- last-remained chromsome

3

Selected Traits associated with Chromosome 6A

Chr 6A

(1.0)
6AS-5 (0.65)
6AS-1 (0.35)
C6A (0.00)
6AL-4 (0.55)
6AL-8 (0.90)
(1.0)

Yield
-QTL for Yield and TKW
 (ca. 10-15%; Snape et al. 2007; Röder et al. unpubli., centromere)

Development
-Seedling vigor (Spielmeyer et al. 2007)

Disease Resistances
-Powdery mildew, *Pm21* (Muranty et al. 2009)
-Stem rust, *Sr8*, *Sr13 (Ug99)*, *Sr26,* (McIntosh et al. 1995/ Belayneh Admassu, 2010)
-Aphid resistance (Castro et al. 2005; centromere)
-*Septoria tritici*, *Stb15* (Arraiano et al. 2007)
-Fusarium head blight, major QTL (Holzapfel et al. 2008)
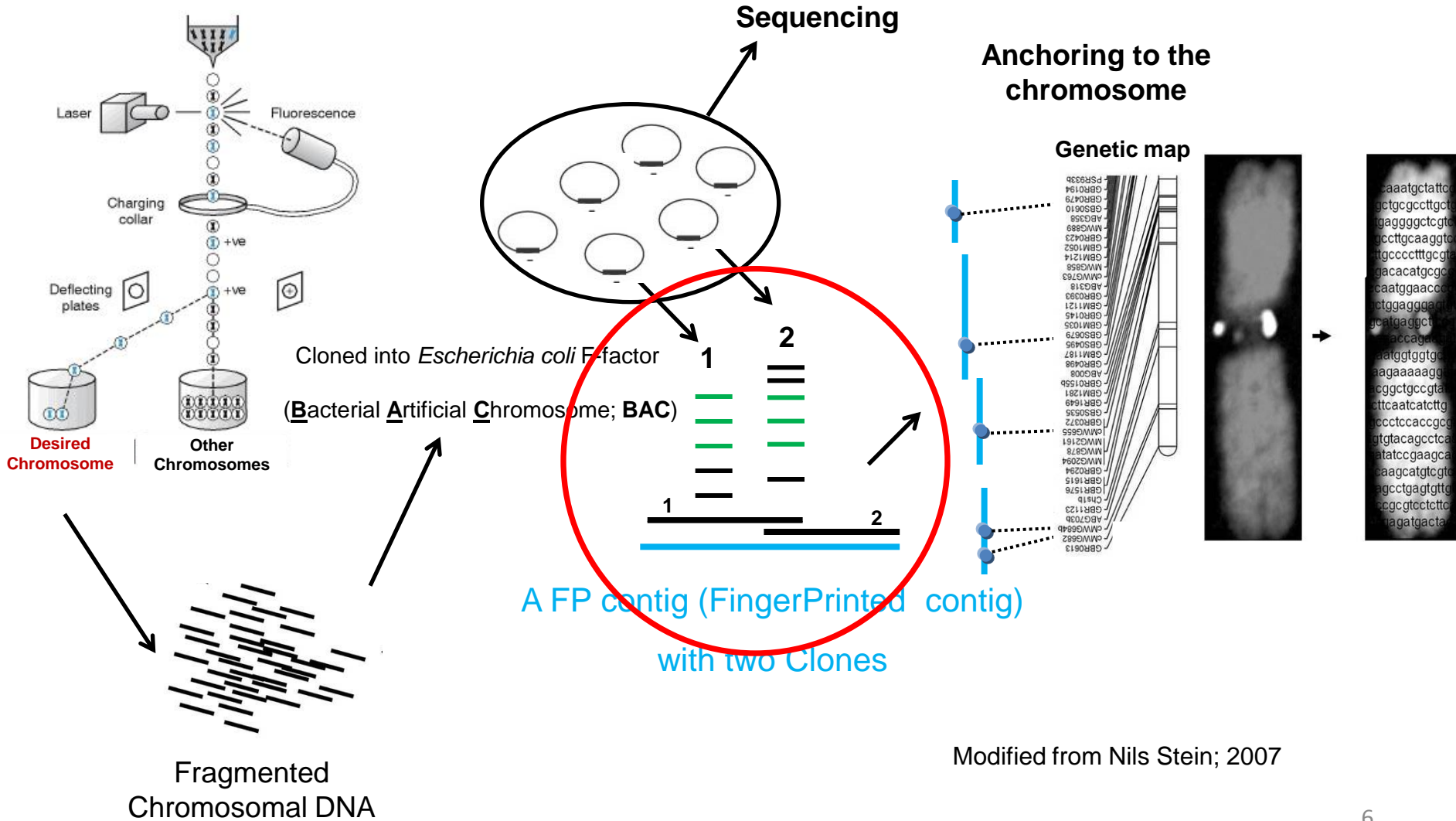
List possibly not complete 4

BAC based Physical map

Basic principles and the new technology

Get the chromosome sequenced;
From BAC library to Physical mapping and Sequencing
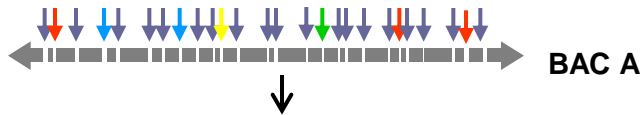
Chromosome Sorting by Flow Cytometry

Sequencing

Anchoring to the chromosome

Genetic map

Laser

Fluorescence

Charging collar

+ve

Deflecting plates

+ve

Desired Chromosome

Other Chromosomes

Cloned into *Escherichia coli* F-factor

(**B**acterial **A**rtificial **C**hromosome; **BAC**)

1   2

1   2

A FP contig (FingerPrinted contig)

with two Clones

Fragmented Chromosomal DNA

Modified from Nils Stein; 2007
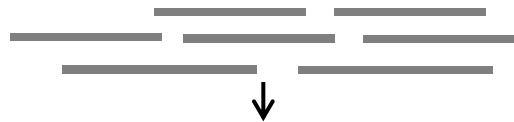
Fingerprinting strategies

High Information content fingewrprinting (HICF)

I: Multiple RE digestion
XbaI, *BamH1*, *XhoI*, *EcoR1* and *HaeIII*
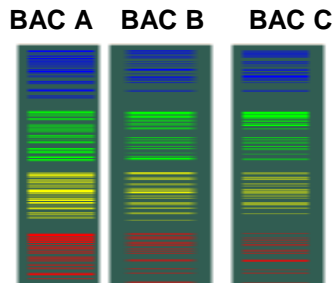
BAC A

II: Enzymatic digested fragments: BAC fingerprints

III: Fragment labeling using fluorescent SNaPshot chemistry: required for size calling

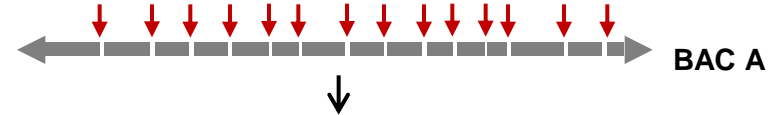IV: BAC fragment based profile: Between 40 to 250 bans per BAC

BAC A   BAC B   BAC C

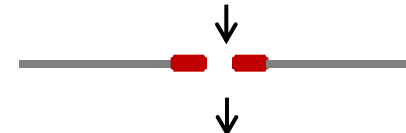IV: BAC overlap identification and contig assembly
tolerance of 0.4 bp

BAC A
BAC B
BAC C

Whole Genome Profiling (**WGP**TM)
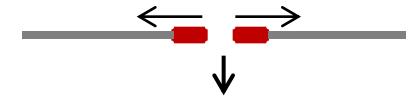
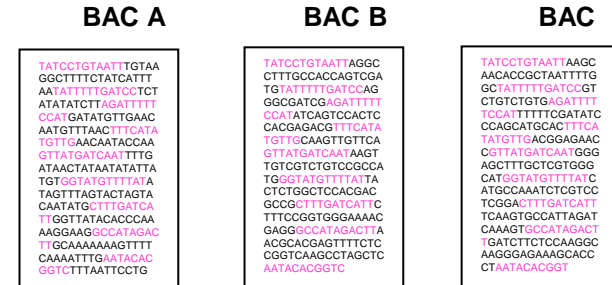### I: Single RE digestion
HindIII

BAC A

II: Sequencing adaptor ligation
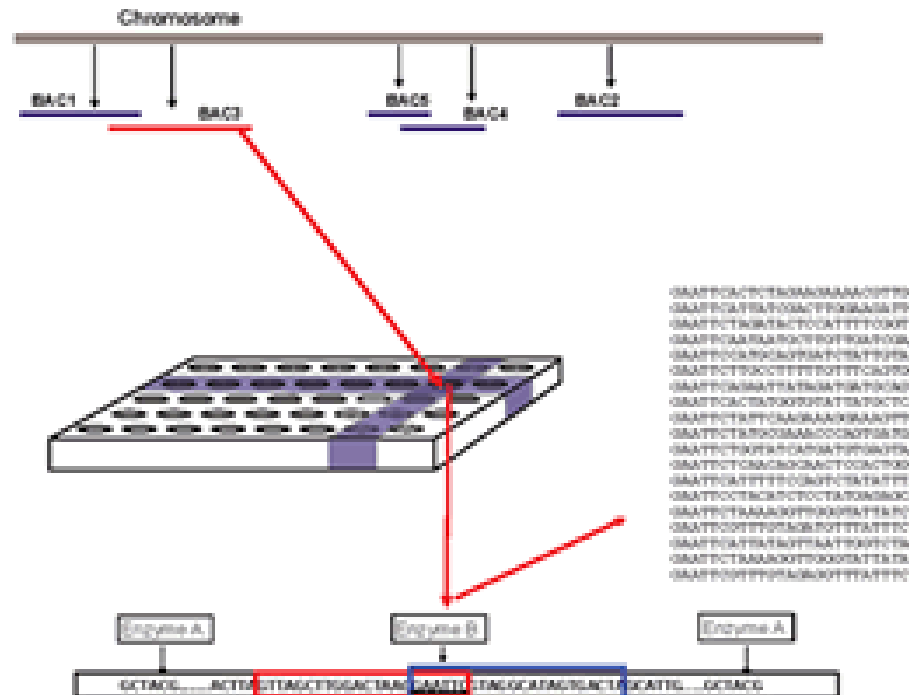
III: End sequencing of enzymatic digested fragments

IV: BAC sequence based profile: on average 50 sequence tags of ~100bp per BAC (126kb, 1HindIII site/2.5Kb)

BAC A	BAC B	BAC
TATCCTGTAATTGTAA	TATCCTGTAATTAGGC	TATCCTGTAATTAAGC
GGCTTTTCTATCATTT	CTTTGCCACCAGTCGA	AACACCGCTAATTTTG
AATATTTTTGATCCTCT	TGTATTTTTGATCCAG	GCTATTTTTGATCCGT
ATATATCTTAGATTTTT	GGCGATCGAGATTTTT	CTGTCTGTGAGATTTT
CCATGATATGTTGAAC	CCATATCAGTCCACTC	TCCATTTTTTCGATATC
AATGTTTAACTTTCATA	CACGAGACGTTTCATA	CCAGCATGCACTTTCA
TGTTGAACAATACCAA	TGTTGCAAGTTGTTCA	TATGTTGACGGAGAAC
GTTATGATCAATTTTG	GTTATGATCAATAAGT	CGTTTATGATCAATGGG
ATAACTATAATATATTA	TGTCGTCTGTCCGCCA	AGCTTTGCTCGTGGG
TGTGGTATGTTTTATA	TGGGTATGTTTTATTA	CATGGTATGTTTTATC
TAGTTTAGTACTAGTA	CTCTGGCTCCACGAC	ATGCCAAATCTCGTCC
CAATATGCTTTGATCA	GCCGCTTTGATCATTC	TCGGACTTTGATCATT
TTGGTTATACACCCAA	TTTCCGGTGGGAAAAC	TCAAGTGCCATTAGAT
AAGGAAGGCCATAGAC	GAGGGCCATAGACTTA	CAAAGTGCCATAGACT
TTGCAAAAAAGTTTT	ACGCACGAGTTTTCTC	TGATCTTCTCCAAGGC
CAAAATTTGAATACAC	CGGTCAAGCCTAGCTC	AAGGGAGAAAGCACC
GGTCTTTAATTCCTG	AATACACGGTC	CTAATACACGGT

IV: BAC overlap identification and contig assembly
tolerance of 0.bp

BAC A
BAC B
BAC C

7

Modified from Catherine Feuillet ; PAG, 2012

A pooling scheme was utilized for WGP of the 6A BACs

BAC Clones are cleverly pooled

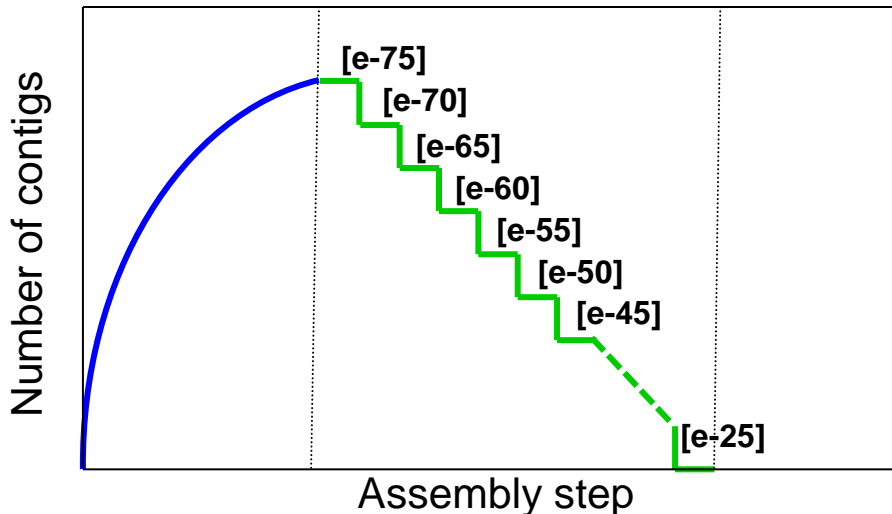Pooled BAC DNA are digested and sequenced with Illumina GAII

Sequence tags are de- convoluted and assigned to individual BAC **(2-6 kb apart)**

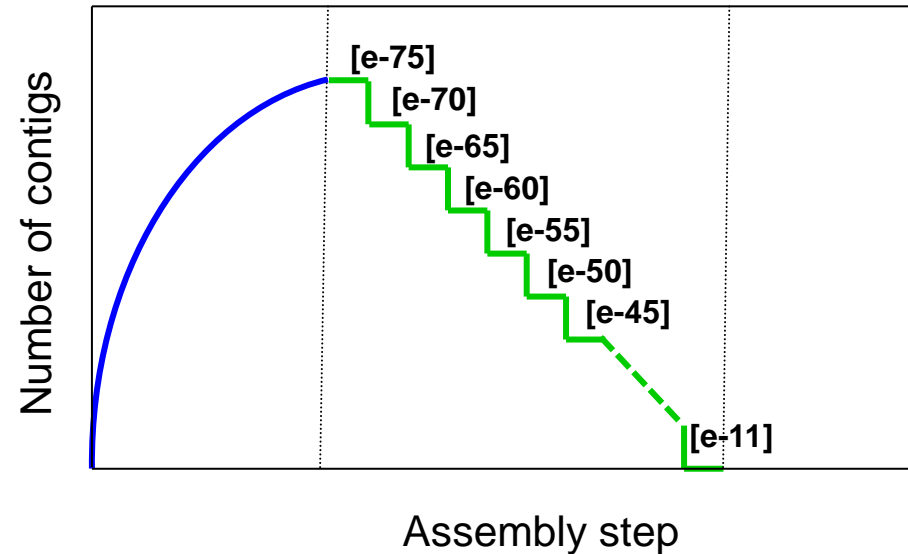**Jan von Overen et al. Genome.Res 2011; Arabidopsis**

http://www.keygene.com/services/technologies_whole_genome_profiling.php

Comparison of HICF and WGP using wheat 3B BAC clones (Philippe et al., 2012)

➢ 11, 238 BAC clones (230 Mb) including 1380 fully sequenced BACs (12 Mb) were used for both HICF and WGP approaches

Physical map assembly using HICF

Physical map assembly using WGP

- Initial Assembly
- Automated Assembly (Single to end & end to end Merging , DQing)

9

Comparison of HICF and WGP

	HICF (1e-25)	WGP™ (1e-11)
Estimated coverage in length	236 Mb +- 65	199 Mb +- 42
Number of contigs	631	434
Average contig size (kb)	374	469
N50 (kb)	455	567

Comparison to 12 sequenced contigs

Coverage percentage in length	95.8%	94.9%
Number of chimerical contigs for 10 Mb	0.6%	0.6%
Percentage of mis-assembled BACs	9.5%	2.7%

✓ Equivalent coverage in lenth
✓ Equivalent number of chimerical contigs
✓ Less miss-assembled BACs in WGP
✓ physical map fully enriched by sequence info in WGP
✓ less time and effort (in case of keygene collabroation) in WGP
~ same cost

Wheat 6A physical map was decided to be performed in collaboration with KeyGene using Whole Genome Profiling (WGP) approach

6A BAC library resource utilized

No of BAC clone	Long Arm(369 Mb)	Short Arm (336Mb)	Chromosome coverage
Complete BAC library	55,296 (123 kb)	46.080 (130 kb)	~16x
BACs for WGP™	23,040	24,576	~8 - 9x
WGP™ Output (FPC ready _keygene)	18,660	19,183	6.2 - 7.4x

Library RE :HindIII

WGP RE: HindIII/MseI

Detail results of the WGP

&

Perliminary results on the **automated assembly**

(6AL and 6AS)

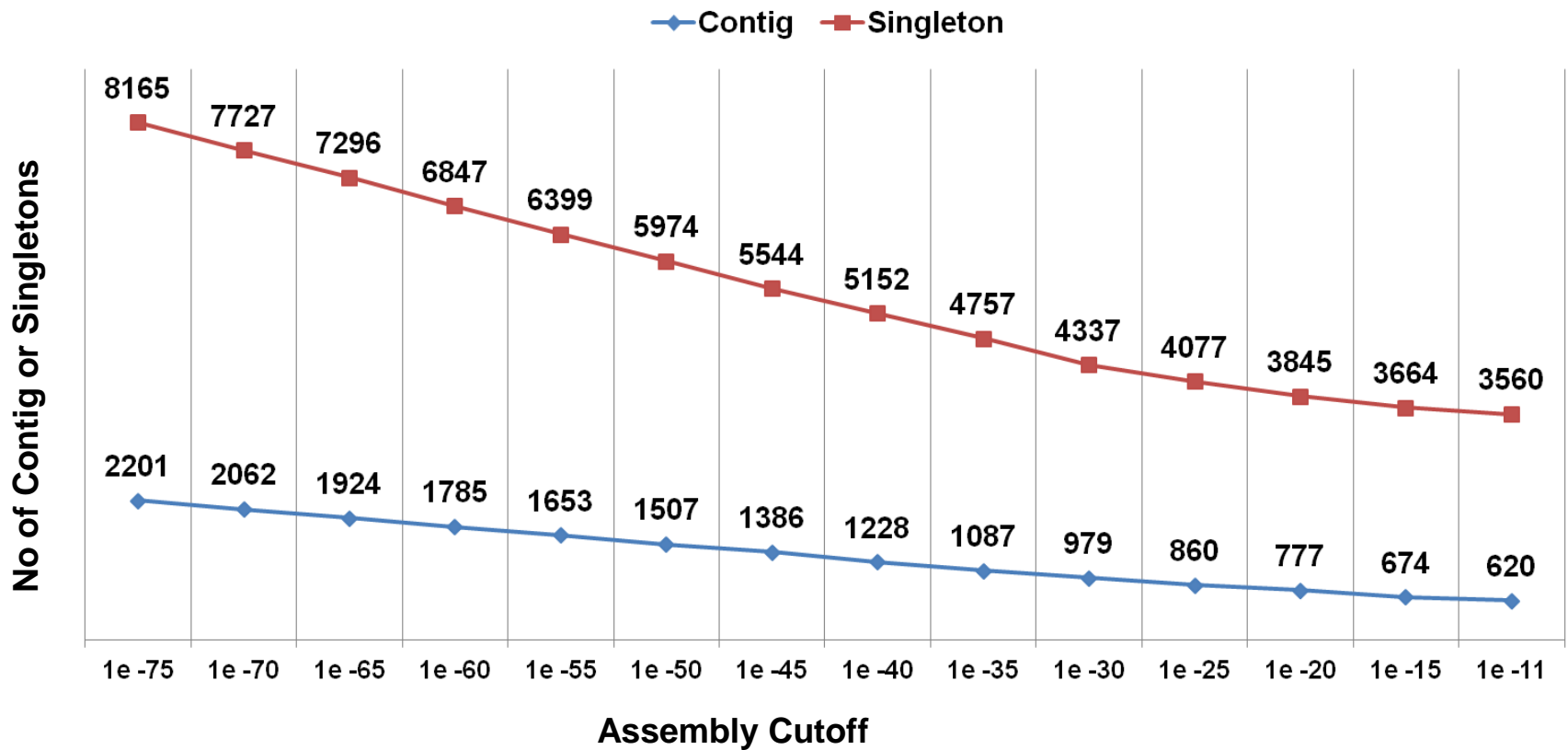# Overview of general WGP output parameters and sequence data processing (Illumina HiSeq2000 sequencing)

WGP parameter	6AS	6AL
Estimated size of chromosome	336 Mbp	369 Mbp
# BACs tested	23,040	24,576
tagged BACs (FPC ready)	19,183 (83,3%)	18,660 (75,9%)
Enzyme combination	HindIII/MseI	HindIII/MseI
% deconvolutable reads	51,6%	47.30%
# unique WGP tags (FPC ready)	85,013	108,811
average # WGP tags/ BAC	27.1	27
average # reads/ tag	122.4	111.1

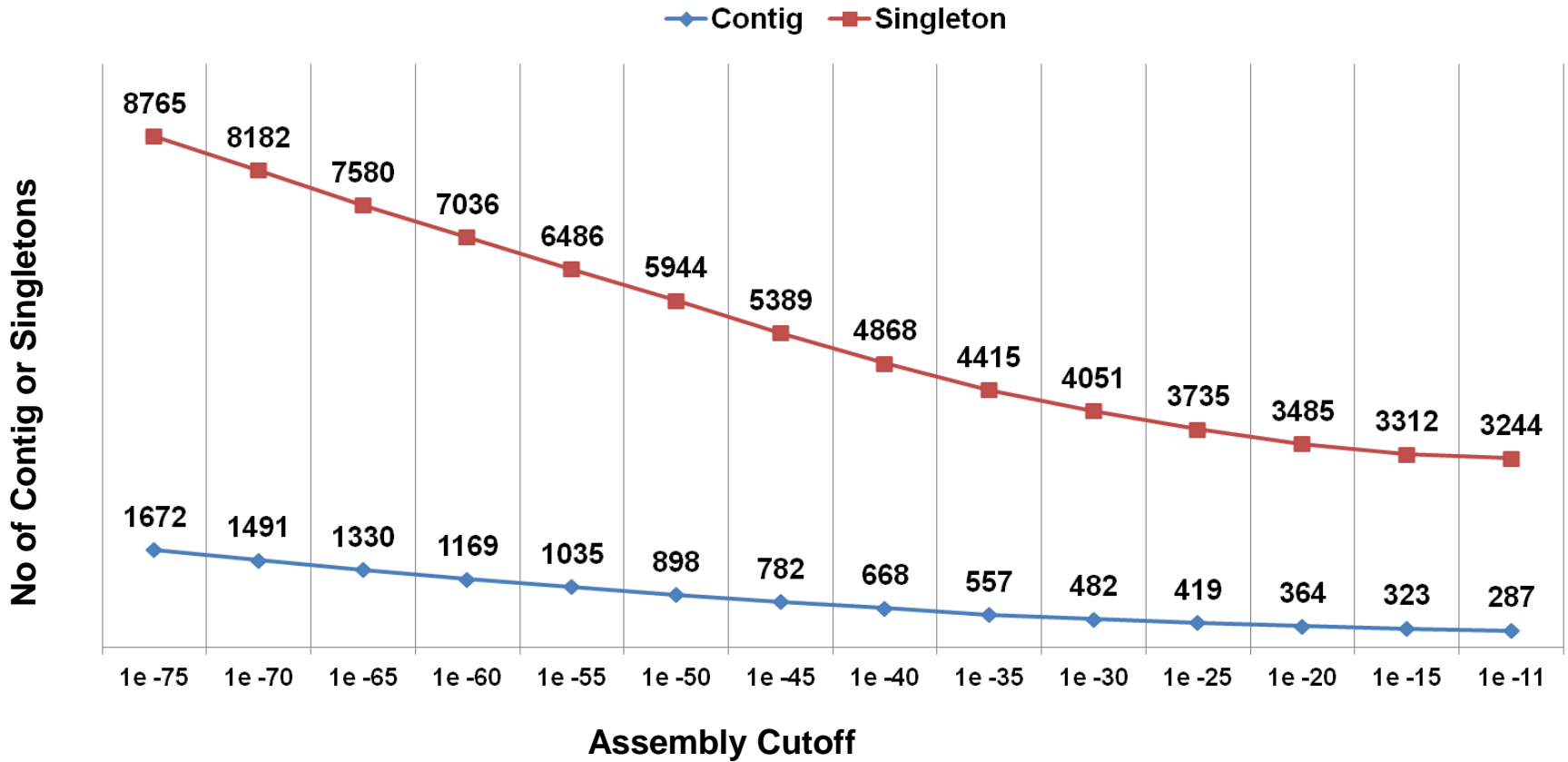# Automated physical map assembly using WGP fingerprints (performed by FPC tool)

- Initial Assembly
- Automated Assembly (Single to end & end to end Merging , Dqing, Rebuilding)

6AL _Stepwise assembly _ reduction of Contigs and Singletons (17309 BACs with 6 to 68 tags, 5.7 x)

6AS _stepwise assembly _ Reduction of Contigs and Singletons (17853 BACs with 6 to 68 tags, 6.9x)

Anchoring strategies

Recombination based genetic maps

➢ Genotyping By Sequencing based (GBS) markers (Jesse Poland, PLoS One 2012);

1400 genetic markers from 6A with known sequences,
215 Double haploid individuals ('Opata / Syn W9784)

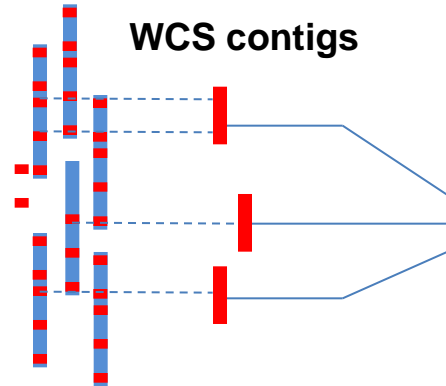Whole chromosome 6A shotgun sequence contigs

GBS genetic markers

Physical contig x

WCS contigs
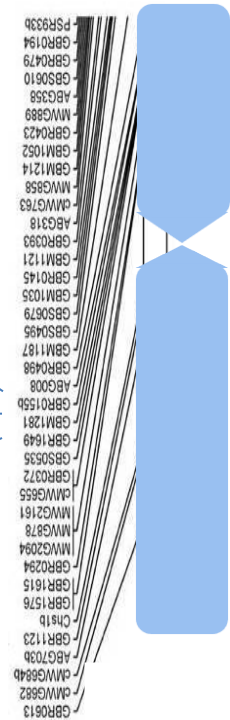
➢ SNP markers, Röder et al. unpubli.,

➢ Wheat KASPar SNP database

• currently 93 SNPs mapped on chr 6A
• ca. 4000 SNPs mapped by the end of this year (in total)
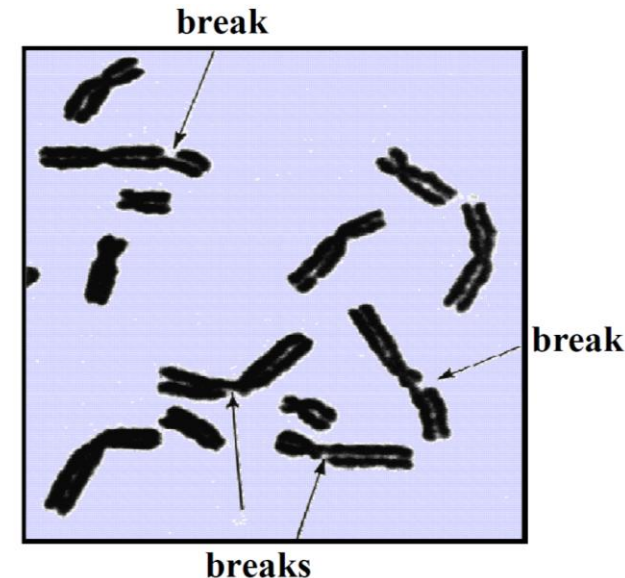• monthly update of the SNP database

Anchoring strategies

Recombination independent genetic maps
(Radiation Hybrid)

❖ Physical mapping based on radiation induced chromosome breakage and reconstruction of marker order based on their co-retention pattern

Wheat A genome RH map is being produced!

(Dr. Shahryar Kianian, North Dakota U, Fargo, USA)

http://www.ratical.org/radiation/CNR/PP/chp3.html

Summary

- The WGP approach is shown to be potentially applicable to make a robust physical map

- The HindIII enzyme utilized in the WGP panel increased the tag density

- The assembly length of the 6AS is corresponding well to the estimated size of the arm while this is not the case for 6AL

and the outlook

- Perform the genetic anchoring using marker info mentioned

- Start the manual editing and merging using the marker info from earlier step of the assembly (e.g. cutoff e-30)

Acknowledgment

IPK, Germany

Thorsten Schnurbusch

Nils Stein

Uwe Scholz / Group

Jenny Knibbiche

KeyGene, Wageningen

Hanneke Witsenboer

Jan van Oeveren

IEB, Czech Republic

Jaroslav Dolezel / Group

INRA, France

Helene Berges / Group

MIPS, Germany

Klaus Mayer / Group

21