**Palacký University Olomouc**

**Faculty of Science**

**Department of Botany**


**and**


**Institute of Experimental Botany**

**Centre of the Region Haná for Biotechnological and Agricultural Research**


**Olomouc**

**Helena Staňková**


# Construction of physical map of 7DS wheat chromosome arm and its use for positional cloning


**Ph.D. Thesis**


**Supervisor: Ing. Hana Šimková, CSc.**


**Olomouc**

**2015**

**Acknowledgements**

I would like to express great thanks to my supervisor, Ing. Hana Šimková, CSc., for a professional guidance, inspirational ideas, patience and help. Further, I would like to thank to the head of the laboratory, prof. Ing. Jaroslav Doležel, DrSc., for the opportunity to work in the Centre of the Plant Structural and Functional Genomics and also to all the members of the team for a friendly and motivating atmosphere. My final thank goes to my family for their patience and support during the years of my studies.

**Declaration**

I hereby declare that I elaborated this Ph.D. thesis independently under the supervision of Ing. Hana Šimková, CSc. using only information sources referred in the chapter References.

...................................................

# BIBLIOGRAPHIC IDENTIFICATION

**Author's name:** Helena Staňková

**Title:** Construction of physical map of 7DS wheat chromosome arm and its use for positional cloning

**Type of Thesis:** Ph.D. thesis

**Department:** Botany

**Supervisor:** Ing. Hana Šimková, CSc.

**The Year of Presentation:** 2015

**Abstract:**

Bread wheat (*Triticum aestivum* L.) is one of the most important crop species providing the staple food for 30 % of the world's population. Despite of that, wheat genome research, namely genetic and physical mapping, sequencing and positional cloning, are significantly hampered by the huge genome size (~17 Gb), presence of three homoeologous subgenomes and prevalence of repetitive DNA sequences (>80 %). Those obstacles can be overcome by dissecting the genome into particular chromosomes or chromosome arms by flow-cytometric sorting. Dividing the genome into small defined parts makes the production of the reference genome sequence a feasible task.

This work focuses on the study of the short arm of the chromosome 7D (7DS) of wheat. Within the framework of the thesis, the physical contig map of the 7DS arm was constructed and sequencing of a minimal tilling path (MTP) was performed. This represents a crucial step towards the completion of the reference sequence of the 7DS. To validate the physical map assembly, sequences of MTP BAC clones were compared with the BioNano genome map constructed for the 7DS arm using Irys platform. The comparison proved that this novel genome mapping technology provides a useful tool for validation and improvement of the physical map assembly. It also holds a potential for studying tandem repeats and supporting a pseudomolecule construction.

The short arm of the chromosome 7D carries several agronomically important genes, including a Russian wheat aphid resistance gene *Dn2401*. The second part of the theses focuses on the positional cloning of this resistance gene. With the aim to construct a high density genetic map covering the gene region, which is a

prerequisite for the positional cloning, a new method for targeted marker development in a hexaploid wheat genome was introduced. Employing this approach, several new markers closely linked to the gene have been developed. This enabled to identify a BAC contig in the 7DS-specific physical map spanning the *Dn2401* region. Sequencing and annotation of the constituting BAC clones revealed several candidate genes. The availability of the physical contig map as well as MTP BAC clone sequences facilitated significantly the positional cloning process.

**Keywords:** bread wheat (*Triticum aestivum*), physical contig map, sequencing, genetic mapping, positional cloning

**Number of Pages/Appendices:** 89/VI

**Language:** English

# BIBLIOGRAFICKÁ IDENTIFIKACE

**Jméno:** Helena Staňková

**Téma:** Konstrukce fyzické mapy krátkého ramene chromozómu 7D pšenice a její využití pro poziční klonování

**Typ práce:** Disertační práce

**Obor:** Botanika

**Školitel:** Ing. Hana Šimková, CSc.

**Rok obhajoby:** 2015

**Abstrakt:**

Pšenice setá (*Triticum aestivum* L.) představuje jednu z nejvýznamnějších zemědělských plodin, jež poskytuje potravu pro 30% světové populace. Studium genomu pšenice, zejména pak genetické a fyzické mapování, sekvenování a poziční klonování, je ztíženo značnou velikostí genomu (~17 Gb), přítomností tří homeologních subgenomů a převahou repetitivních sekvencí (>80 %). Řešení těchto problémů nabízí třídění jednotlivých chromozómů a jejich ramen pomocí průtokové cytometrie. Toto rozdělení genomu na malé přesně definované části je nezbytným krokem k získání referenční sekvence genomu pšenice.

Předkládaná práce se zabývá studiem krátkého ramene chromozómu 7D (7DS). V rámci práce byla zkonstruována fyzická kontigová mapa ramene 7DS. Ta posloužila k výběru tzv. *minimal tilling path* (MTP), tedy minimální sestavy klonů z knihovny dlouhých inzertů, která reprezentuje celé chromozómové rameno. Klony z MTP byly následně sekvenovány a získaná data se stala základem pro sestavení referenční sekvence ramene 7DS. Za účelem ověření správnosti fyzické mapy a tedy i výsledné chromozómové sekvence bylo provedeno srovnání s genomovou mapou vytvořenou technologií BioNano na platformě Irys. Toto srovnání prokázalo, že genomová mapa může posloužit jako užitečný nástroj pro ověřování a zdokonalení fyzické mapy a má také potenciál uplatnit se při analýze tandemových repetitivních sekvencí. Zmíněná technologie může dále poskytnout jeden z podpůrných nástrojů pro sestavení pseudomolekuly, tedy kontinuální sekvence celého chromozómového ramene.

Krátké rameno chromozómu 7D nese několik genů pro agronomicky významné znaky, mezi nimi i gen *Dn2401*, jenž podmiňuje rezistenci k mšici

zhoubné. Druhá část předkládané práce je věnována pozičnímu klonování uvedeného genu. Jedním ze základních kroků v procesu pozičního klonování je konstrukce vysokohustotní genetické mapy pro oblast klonovaného genu. S cílem zahustit genetickou mapu byl navržen postup pro cílený vývoj markerů v prostředí hexaploidního pšeničného genomu. Za použití této metody bylo vyvinuto několik nových markerů v těsné vazbě na gen. Pomocí nově vyvinutých markerů byl identifikován kontig fyzické mapy, jenž překlenuje oblast genu. Sekvenování a následná anotace sekvencí klonů ze získaného kontigu odhalila přítomnost několika kandidátních genů. Dostupnost fyzické kontigové mapy a sekvencí jednotlivých klonů z MTP značně zjednodušila proces pozičního klonování.

**Klíčová slova:** pšenice setá (*Triticum aestivum*), fyzická kontigová mapa, sekvenování, genetické mapování, poziční klonování

**Počet stran/příloh:** 89/VI

**Jazyk:** anglický

# CONTENT

# 1 INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is one of the most important crop species worldwide, serving as a staple food for 30 % of the human population. It is cultivated on more land area than any other crop and its production goes over 700 million tons per year (http://faostat.fao.org/). As the human population is rising, expected to exceed 9 billion by 2050, it is presumed that the wheat production should increase by about 70 % to meet the future demand (IWGSC 2014). To approach this challenging target, efficient breeding for yield, quality and resistance to biotic and abiotic stresses will be essential.

A high quality reference sequence of the wheat genome will offer an insight into the genome composition, provide a complete gene catalogue with genes ordered along the chromosome and facilitate the exploration of regulatory network of developmental processes resulting in complex traits. This knowledge will support and simplify the process of selecting new improved wheat varieties and it will also significantly accelerate isolation of agronomically important genes by positional cloning. For a long time, reaching a high quality reference sequence of wheat was considered an impossible goal due to characteristic genome features such as huge size (~17 Gb), presence of three homoeologous subgenomes and a high portion of repetitive DNA (>80 %). However, the possibility of dissecting the wheat genome into individual chromosomes and chromosome arms by flow cytometry sorting in combination with BAC-by-BAC sequencing strategy made this goal feasible.

In this thesis, I focused on the construction of the physical map of the short arm of chromosome 7D (7DS) of wheat. Construction of the physical map represents a key step towards obtaining a quality reference sequence of this chromosome arm. The 7DS-specific physical map provided an important tool to facilitate positional cloning of a Russian wheat aphid resistance gene located on the 7DS chromosome arm.

# 2  LITERATURE REVIEW

## 2.1  Bread wheat (*Triticum aestivum* L.)

Bread wheat (*Triticum aestivum* L.) belongs to the *Poaceae* (grass) family, *Pooideae* subfamily and the tribe *Triticeae*. It is an allohexaploid species (2n = 6x = 42) whose genome comprises three individual subgenomes: A, B and D. The hybridization of three different progenitor species resulted in a large (1C ~16979 Mb; Bennett and Smith 1976) and highly redundant genome with more than 80 % of the genome consisting of repetitive sequences (Smith and Flavell 1975). Although wheat is a hexaploid, it behaves as a diploid in meiosis because the homoeologous pairing is prevented through the action of *Ph* genes (Martinez-Perez *et al.* 2001).

Wheat has been one of the first domesticated food crops and nowadays is cultivated on more land than any other crop species. Along with rice and maize, wheat represents one of the most important sources of plant calories consumed by human. Currently, the allohexaploid bread wheat dominates global wheat production. About 95 % of the wheat crop is bread wheat (used for making bread, cookies, and pastries) whereas the remaining 5 % are represented by durum wheat (*T. turgidum* ssp. *durum*) used for making pasta and other semolina products (Dubcovsky and Dvorak 2007). Since its origin, wheat has spread worldwide to be grown over a wide range of climates and soil conditions from Norway and Russia at 60°N to Argentina at 45°S (Dubcovsky and Dvorak 2007). The key factor in the success of wheat all over the world is its adaptability to a wide range of climatic conditions which can be partly attributed to the allohexaploid genome structure consisting of progenitor genomes previously adapted to different environments, thus creating the potential for the adaptation to a wider range of environmental conditions (IWGSC 2014).

### 2.1.1  Origin of bread wheat

Bread wheat cultivation and domestication has been directly associated with the spread of agriculture and settled societies ~10,000 years ago in the area of Fertile Crescent (present-day Israel, Jordan, Turkey, Syria, Iran, and Iraq; Feuillet *et al.* 2008).

Hexaploid genome of currently cultivated bread wheat arose from multiple successive spontaneous hybridization events between three different diploid progenitor species from *Poaceae* family (Figure 1). Each hybridization was followed by a polyploidization. Wheat diploid progenitors radiated from a common *Triticeae* ancestor 2.5 - 4.5 million years ago (MYA; Huang *et al.* 2002). The initial hybridization occurred less than 0.5 MYA involving the A genome donor *Triticum urartu* (2n = 2x = 14; $A^u A^u$) and the B genome donor - an unknown species from *Sitopsis* section closely related to *Aegilops speltoides* (2n = 2x = 14; SS). This hybridization resulted in tetraploid *Triticum turgidum* (2n = 4x = 28; AABB), an ancestor of a wild emmer wheat cultivated in the Middle East and *T. turgidum* ssp. *durum* grown for pasta production today (Huang *et al.* 2002; IWGSC 2014). Subsequent hybridization between the tetraploid *T. turgidum* and the donor of the D genome, *Aegilops tauschii*, resulted ~10,000 years ago in the hexaploid wheat *Triticum aestivum* (2n = 6x = 42; AABBDD; Feldman *et al.* 1995; Salamini *et al.* 2002; Petersen *et al.* 2006). Hybridization and polyploidization events were followed by a number of structural and functional rearrangements resulting in genome stabilization (Feldman and Levy 2009). Each of the three ancestral wheat diploid progenitor genomes is about 5.5 Gb in size. The hybridizations of the diploid genomes gave rise to a highly redundant ~17-Gb hexaploid genome with three homoeologous subgenomes, each carrying highly similar gene copies (Choulet *et al.* 2014a).
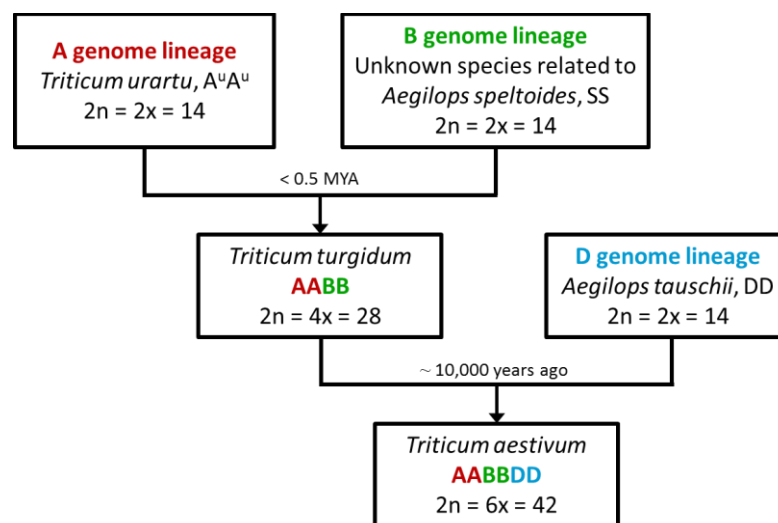


Figure 1: Scheme of hybridization events leading to development of the hexaploid genome of bread wheat (*Triticum aestivum*).

Recently, Marcussen *et al.* (2014) suggested that the D genome originates from the hybridization between the A and B genome lineages which occurred ~ 5.5 MYA. According to this study, the D genome was formed by the process of homoploid hybrid speciation – a hybridisation that does not result in a genome duplication event. However this hypothesis was recently re-evaluated by Li *et al.* (2015) taking into account chloroplast genome sequence information. The latter study revealed that the history of *Ae. tauschii* may be more complex than suggested by Marcussen *et al.* (2014), involving multiple rounds of both recent and ancients hybridisations between the species from *Aegilops-Triticum* complex.

Comparison of the hexaploid bread wheat gene sequences with gene repertoires from its closest extant relatives, which donated the A, B and D genomes showed limited gene loss during the evolution of the hexaploid wheat genome and frequent gene duplications after these genomes came together (Eversole *et al.* 2014). Gene expression patterns revealed that none of the subgenomes dominated gene expression and there is no bias in the gene content, structure, or composition between the different wheat subgenomes. Moreover, individual subgenomes exhibit a high degree of regulatory and transcriptional autonomy (Eversole *et al.* 2014; Pfeifer *et al.* 2014). Interestingly, this fact does not correspond with the findings in other paleopolyploid species (e.g. maize, soybean, cotton) where one of the subgenomes is more transcriptionally active than the others (Schnable *et al.* 2011; Rapp *et al.* 2009; Chaudhary *et al.* 2009). The apparent autonomy of the three wheat subgenomes may be explained by the relatively recent polyploidization and can also be related to regulatory mechanisms that control the transcriptional interplay of homoeologous genomes to balance expression of individual and groups of genes (IWGSC 2014).

## 2.1.2 Wheat genome information content

Until recently, relatively little was known about the composition of the wheat genome in terms of the number, position and distribution of genes and repetitive elements along the chromosomes. Sequencing of ESTs and cDNA clones (Erayaman *et al.* 2004; Gill *et al.* 1996a; Gill *et al.* 1996b; Sandhu and Gill 2002), selected BAC clones (Choulet *et al.* 2010; Devos *et al.* 2005) and BAC-end sequencing (Paux *et al.* 2006) provided a first insight into a composition of the wheat genome. However,

those data represented only a small fraction of the genome, thus could not be reliably extrapolated on the genome as a whole.

Recent BAC-by-BAC sequencing of the chromosome 3B resulting in a reference chromosome sequence (Choulet *et al.* 2014a), whole-chromosome shotgun sequencing of all individual chromosome (IWGSC 2014) as well as whole-genome shotgun sequencing of bread wheat (Brenchley *et al.* 2012; Chapman *et al.* 2015) shed a light on the wheat genome composition. Analysis of the chromosomal survey sequences produced by the whole-chromosome shotgun sequencing revealed that wheat genome is highly dynamic with a high level of elasticity and a changing gene repertoire in all wheat subgenomes (IWGSC 2014). New variation is rapidly generated through gene deletions and insertions of repetitive elements into coding and regulatory gene regions. Those disruptions are buffered by the polyploid nature of wheat. The buffering effect is corroborated by the fact that most of the 21 *T. aestivum* chromosomes can be removed to produce nullisomic plants exhibiting only minor phenotypic effects (Dubcovsky and Dvorak 2007).

In general, wheat genome is made up of two parts: (1) a small conservative part that is subjected to selection pressure and mostly corresponds to the "gene space", and (2) a much larger and more variable component which is under more dynamic evolution and comprises the "transposable elements (TE) space" as well as duplicated genes and gene fragments (Choulet *et al.* 2014b).

Study of the 3B chromosome pseudomolecule revealed that genes are not evenly distributed, and gene density is increasing on both arms along the centromere-telomere axis. This pattern is commonly observed in grass genomes. Distal segments of the 3B chromosome are exhibiting the highest gene density of 19 genes per Mb. The lowest gene density was observed in a segment encompassing the centromeric-pericentromeric region (five genes per Mb). As previously suggested by Choulet *et al.* (2010, 2014b) and Rustenholz *et al.* (2011), the genes are found everywhere along the chromosomes. There is no large region completely devoid of coding sequence, nevertheless the intergenic distances are extremely variable and a majority of genes are organized in small islands (Choulet *et al.* 2014a). Survey sequencing of all wheat chromosomes revealed that on average 23.6 % of genes are duplicated within a chromosome. The analysis of the 3B pseudomolecule sequence revealed even more duplicated genes (37 %; Choulet *et al.* 2014a). Estimated number of protein coding genes varies among studies. Choulet *et al.* (2014a) predict 93,000 genes for the bread

wheat genome as an extrapolation of the 3B data, Brenchley *et al.* (2012) estimate 94,000 – 96,000 genes in a whole-genome shotgun sequence, whilst 106,000 genes were identified in survey sequences of all individually sequenced chromosomes (IWGSC 2014).

According to studies based on renaturation kinetics (Smith and Flavell 1975) wheat genome harbours more than 80 % of the repetitive elements. Similar portion of DNA repeats (81 and 85 %) was also observed in chromosomal survey sequences and in the pseudomolecule of chromosome 3B, respectively (IWGSC 2014; Choulet *et al.* 2014a). The high amount of repetitive sequences results from a massive amplification in the ancestral genome before its divergence from related species of the *Triticeae* tribe around 15 million years ago (Choulet *et al.* 2014b). Estimated insertion dates for the most abundant LTR retrotransposon families showed a major peak at 1.5 million years. 3B sequencing supports the hypotheses that most of the transposable elements that shaped the B genome were inserted before polyploidization (0.5 million years ago) and have been less active since then. Detailed analysis of the 3B sequence revealed that most of the repetitive elements (67 %) are class I repetitive elements, represented mostly by LTR retrotransposons *Gypsy* (47 %) and *Copia* (16 %). 18 % of the sequence is composed of class II DNA transposons that mostly correspond to *CACTA* elements (16 %). Transposable elements are distributed along the whole chromosome with the highest density in centromeric-pericentromeric region (93 %). DNA transposons *Mutator*, *Harbinger* and MITEs are found close to genes, whereas LTR retrotransposons and *CACTA* elements tend to be located at much larger distances from the genes (Choulet *et al.* 2014a).

## 2.1.3 Reduction of the genome complexity

In a case of a large and complex genome of allohexaploid bread wheat, every reduction of the complexity helps to facilitate the genome study. First option is the indirect analysis of the wheat through the use of wheat progenitors and close relatives belonging to *Poaceae* family. The related species keep high degree of similarity with the wheat genome while having significantly lower genome size and reduced ploidy level. However, this approach is not suitable for particular types of wheat genome studies. Another option is a reduction of the genome complexity by

$C_0t$ filtration (Lamoureux *et al.* 2005; Šimková *et al.* 2007), methylfiltration (Rabinowicz *et al.* 2003) or most recently exome capture (Winfield *et al.* 2012). Nevertheless, these strategies are featured by eliminating a part of the genome. In order to get the high quality reference sequence of wheat, an approach reducing the wheat genome without any loss has to be applied. Solution for this offers flow cytometry sorting which enables to dissect the wheat genome into individual chromosomes and chromosome arms (Doležel *et al.* 2014).

### 2.1.3.1   Related species

All grasses show a high level of collinearity, i.e. a conservation of gene order along segments of the chromosome (Gale and Devos 1998a). The collinearity observed among the grass genomes enabled to partly characterise hexaploid wheat genome with the absence of its complete genome sequences by means of comparative genomics approaches. Comparative genomics in grasses has been used to define syntenic relationships between different species (Gale and Devos 1998b; Devos and Gale 2000) and provided insight into their evolution since the divergence from a common ancestor (Murat *et al.* 2010).

Having no reference genome sequence of wheat in hands, annotated genome sequences of wheat progenitors and closely related species, as well as other tools like genetic and physical maps, provided an invaluable source of information for wheat genome research. So far, genome sequences were generated for several wheat related crops, namely rice, *Brachypodium distachyon* and sorghum (IRGSP 2005; IBI 2010; Paterson *et al.* 2009) and diploid wild progenitors of wheat - *Aegilops tauschii* and *Triticum urartu* (Jia *et al.* 2013; Ling *et al.* 2013). Based on the sequence homology between related species and hexaploid wheat, genes in wheat chromosomal survey sequences could be identified and annotated (IWGSC 2014). The collinearity between wheat and its relatives played also an important role in many genetic mapping and map-based cloning projects as a clue for development of new markers and for identification of candidate genes (Kota *et al.* 2006; Griffiths *et al.* 2006; Yan *et al.* 2003, 2006; Terracciano *et al.* 2013). Finally, collinearity enabled to create so-called genome zippers. Genome zipper represents a collinearity based order of *Brachypodium distachyon*, rice and sorghum genes anchored to gene-based genetic map and supported by partial sequence information.

Despite the genomes of wheat ancestors and relatives were successfully used in several projects, they still represent at the best only an approximate model of the hexaploid genome. Even the use of diploid wheat progenitors (donors of A, B and D subgenomes) as a surrogate for the hexaploid is problematic. Regardless of a good conservation between their genomes and the genome of hexaploid wheat, specific rearrangements occurred during the evolution (Wicker *et al.* 2003; Chantret *et al.* 2005; Feldman and Levy 2005). As a result, the genome of hexaploid wheat does not correspond exactly at either the structural or the functional level to the sum of its three ancestral diploid genomes (IWGSC 2014). It is unequivocal that obtaining a reference genome sequence of hexaploid wheat is needed to enable proper characterisation of the genome.

### 2.1.3.2  Chromosome sorting

Dissecting the genome into individual chromosomes or chromosome arms enables to avoid the problem of genome redundancy which is characteristic for polyploids. One approach to isolating individual chromosomes is the microdissection (Matsunaga *et al.* 1999; Stein *et al.* 1998). However, this technique is quite labour-intensive and does not provide sufficient amount of the DNA. By contrast to this, sorting of individual chromosomes/arms using flow cytometry represents a high-throughput and automated method providing sufficient amount of the intact chromosomes which are suitable for most of the downstream applications (Doležel *et al.* 2014).

Flow cytometry sorting is based on the real-time analysis of the light scatter and fluorescence of sorted particles – in this case fluorescently labelled mitotic chromosomes (Doležel *et al.* 1992; Vrána *et al.* 2012). The output of the analysis is a histogram of relative fluorescence intensity, so-called flow karyotype. All particles with the same fluorescence intensity and thus of the same size, are located within the same peak. In ideal case, each chromosome is represented by one individual peak. Unfortunately, in a standard flow karyotype of wheat (Figure 2a), only one chromosome (3B) forms a distinct peak and thus can be discriminated and sorted. Other chromosomes are located in three large composite peaks, one of them containing chromosomes 1D, 4D and 6D, remaining two containing a mixture of chromosomes of similar size (Vrána *et al.* 2000).
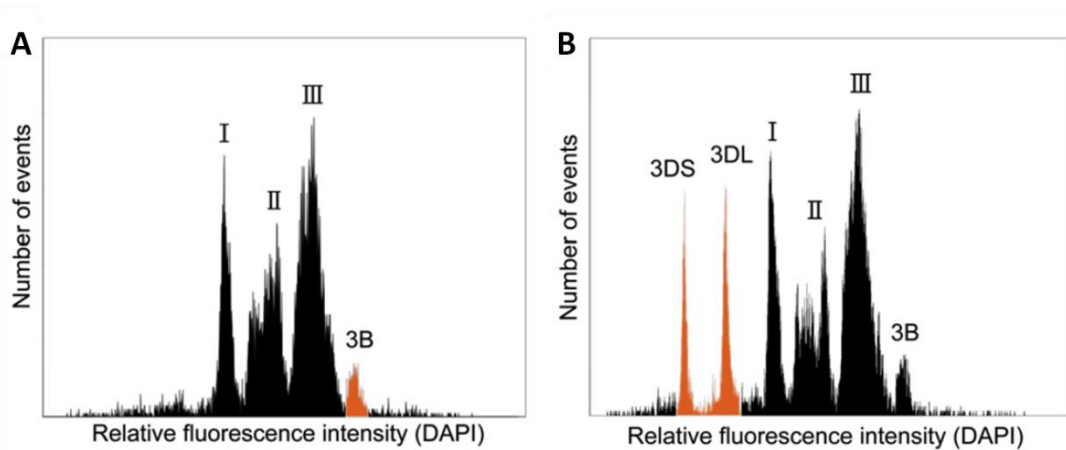
Figure 2: Histograms of relative fluorescence intensity (flow karyotypes) obtained from DAPI-stained suspensions of mitotic chromosomes. (**A**) A flow karyotype of hexaploid bread wheat cv. 'Chinese Spring' comprising three composite peaks (labelled I–III), representing groups of various chromosomes, and a peak of chromosome 3B. (**B**) Flow karyotype of a double-ditelosomic line dDt3D of cv. 'Chinese Spring' in which the chromosome 3D is represented by two telocentric chromosomes (telosomes) 3DS and 3DL (Doležel *et al.* 2014).

The need for sorting each of wheat chromosomes can be resolved by the use of special cytogenetic stocks, namely ditelosomic or double-ditelosomic lines (Sears and Sears 1978). Such wheat lines carry one or both arms (short and long) of a specific chromosome in the form of telosomes. All telosomes are considerably smaller than remaining chromosomes, thus creating discrete peaks which are not present in a standard flow karyotype (Figure 2b). This feature enables to sort them easily (Doležel *et al.* 2014). All wheat chromosomes, with the exception of chromosome 5BL can be sorted from the ditelosomic lines. Only the 5BL chromosome has to be sorted from a line carrying this arm as an isochromosome (Vrána *et al.* 2012). The level of contamination by other chromosomes is checked after sorting by fluorescence *in situ* hybridization (Doležel *et al.* 2004).

The disadvantage of the cytogenetic lines lies in the fact that all of them were developed from the wheat cultivar 'Chinese Spring' and the arms could not be sorted from a material of interest. This obstacle has been recently surmounted by Giorgi *et al.* (2013) who developed a method called FISHIS (fluorescence *in situ* hybridisation in suspension), which differentially labels chromosomes by hybridizing with oligonucleotide probes targeting specific microsatellite sequences. Hereby specifically labelled chromosomes can be identified in the flow karyotype and sorted with a high purity.

The DNA of flow-sorted chromosomes is of superior quality as their morphology after sorting is well preserved. Such a DNA is suitable for various applications including construction of BAC libraries, physical and genetic mapping, positional cloning, cytogenetic analysis, sequencing or genome mapping in nanochannel arrays (Doležel *et al.* 2014; Staňková *et al.* accepted). If higher amount of the DNA is needed without a special demand on DNA size, sorted chromosomes can be amplified using multiple displacement amplification (MDA) protocol based on Phi29 DNA polymerase (Šimková *et al.* 2008). A possibility of sorting individual chromosome arms enables to reduce the huge wheat genome into separate fractions of 224–580 Mb, each representing only 1.3–3.4 % of the genome size (Šafář *et al.* 2010). This highly simplifies the analysis of the wheat genome and paves the way towards the reference sequence of bread wheat.

## 2.2 Sequencing of the wheat genome

Obtaining a high quality reference sequence of the wheat genome is essential for the thorough characterisation of the genome and understanding biological processes in this world's most widely grown cereal crop. Complete reference sequence will provide access to a gene catalogue and regulatory elements and enable to determine their function and role in the wheat development. It will also allow to investigate the genome structure, organization, and evolution and to exploit the natural and induced genetic diversity of wheat. Finally, accessibility to a high quality reference genome sequence will be valuable for the accelerated development of improved wheat varieties (Feuillet *et al.* 2011; Choulet *et al.* 2014b).

### 2.2.1 Sequencing technologies

DNA sequencing is a process that determines the order of nucleotides in a DNA molecule. The original DNA sequencing was invented by Frederik Sanger (Sanger *et al.* 1977). Sanger sequencing is a chain-termination based method providing reads of the length up to 1,000 bp, with the significantly low error rate, thus generating reliable sequencing data (Choulet *et al.* 2014b). Nevertheless, the method is relatively expensive and have lower throughput in comparison with other techniques currently available. For those reasons it would not be feasible to use this method for the sequencing of huge plant genomes.

The advent of next-generation sequencing (NGS) in 2005 revolutionized genomics as it enabled to produce a huge amount of data in a short time with relatively low costs. Thus it made the sequencing of large genomes feasible. This new sequencing era started with the invention of highly parallelized pyrosequencing (Margulies *et al.* 2005) followed by many other technologies. In general, NGS techniques, whose common feature is a massively parallel sequencing producing as much as trillions base pairs in one run, can be divided into two groups. The first one, termed "second-generation sequencing" (SGS), includes a step of amplification of DNA fragments before the sequencing to ensure a sufficiently strong signal for correct reading of individual sequenced bases (Metzker *et al.* 2010; Mardis *et al.* 2008). Although several sequencing platforms have been developed and commercialised so far, including Roche/454 based on pyrosequencing, Sequencing

by Oligonucleotide Ligation and Detection (SOLiD; Valouev *et al.* 2008) or IonTorrent employing semiconductor sequencing (Rothberg *et al.* 2011), the sequencing platform of Illumina dominates in the majority of projects. The technology is based on bridge DNA amplification and "sequencing by synthesis" (Benett *et al.* 2005). Compared to other platforms, Illumina sequencing provides relatively long pair-end reads (up to 300 bp; http://www.illumina.com/) for a reasonable cost. Besides the pair-end sequencing, SGS technologies can be used for the mate-pair sequencing which enables to obtain short paired reads located tens of kilobase apart in the genome (http://www.illumina.com/; http://www.lucigen.com/).

In contrast to the above mentioned platforms, the "third-generation sequencing technologies" rely on direct single-molecule sequencing and provide reads of several kilobases in length. Compared to SGS technologies, the amplification step is avoided in order to avert potential artefacts that can arise during the amplification. Currently, the most advanced are the Single Molecule Real Time (SMRT) technology developed by Pacific Biosciences (PacBio platform; Eid *et al.* 2009), and a nanopore based technology (Mikheyev and Tin 2014) commercialized by Oxford Nanopore (https://www.nanoporetech.com/). While these technologies excel in the read length, which currently achieves 10-15 kb for PacBio and tens of kb for Oxford Nanopore, they are still not widely used for the genome sequencing because of a relatively high error rate and, in case of the PacBio, also a high cost (Metzker *et al.* 2010).

## 2.2.2  Plant genome sequencing strategies

The choice of the sequencing strategy depends on the size of the target genome, complexity, repeat content and on the ultimate goal of post sequencing research. Genomes of plants highly vary in the size, ploidy level and amount of repetitive sequences. The smallest known genome among angiosperms was found in the genus *Genlisea,* specifically in *Genlisea tuberosa*, a carnivorous species endemic to Brazil, with the genome size of 1C ~ 61 Mb (Fleischmann *et al.* 2014). In contrast, the largest genome is that of *Paris japonica*, a rhizomatous geophyte endemic to Japan, with the 1C genome size of 150 Gb (Pellicer *et al.* 2010). The amount of repetitive and transposable elements differs from 10 % in *Arabidopsis* to more than 80 % in wheat and the level of ploidy varies between diploid to octaploid and higher

(Feuillet *et al.* 2011). In general, model species with small, diploid genomes and a low portion of repeats were considered more tractable for sequencing and were the primary targets in whole-genome sequencing projects. Thus, the first plant genome for which reference sequence has been produced was that of *Arabidopsis thaliana* with genome size of 125 Mb (AGI 2000). *A. thaliana* was followed by the sequencing of nearly 400-Mb genome of rice (IRGSP 2005), which serves as a model for grass species. Nevertheless, many important crop species have genomes much larger and more complex than model organisms, which makes the sequencing and sequence assembly difficult. Considering this, appropriate strategy has to be selected in order to deliver the sequence of a certain required quality. The advent of next-generation sequencing technologies contributed to a possibility of sequencing such genomes as it brought a significant reduction in cost and increase of throughput.

Generally, two main strategies are utilised in genome sequencing projects – (1) the whole-genome shotgun sequencing and (2) a clone-by-clone sequencing (Figure 3). These strategies can be applied independently or can be combined in order to enhance the final sequence and to increase the probability of capturing the maximum information (Green 2001).

## 2.2.2.1 Whole-genome shotgun sequencing

Whole-genome shotgun (WGS) sequencing (Figure 3a) is an approach subsuming fragmentation of the whole genome into small pieces of a defined length which are then sequenced, usually from both ends producing pair-end reads. The whole set of pair-end data is then used for the sequence assembly. Originally, the genome fragments were cloned into a plasmid vector and inserts were sequenced by Sanger sequencing providing reads up to 1,000 bp. Nowadays, using the second-generation sequencing techniques, fragmented DNA is sequenced directly and the shorter length of sequencing reads (up to 300 bp) is compensated by high genome coverage. WGS sequencing represents straightforward, fast and relatively cheap procedure in comparison with clone-by-clone approach but at the same time brings some disadvantages.
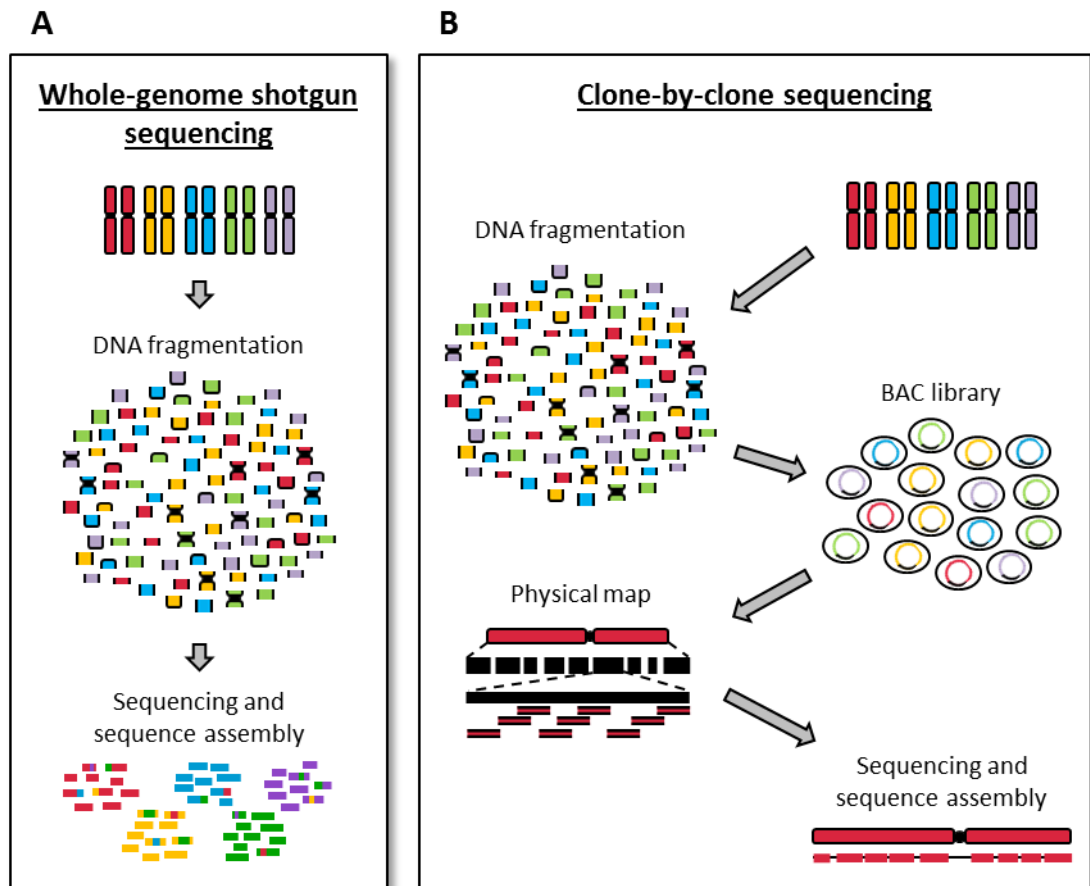
Figure 3: Scheme of the whole-genome shotgun (**A**) and clone-by-clone (**B**) sequencing strategy

Shotgun sequencing enables to establish a relatively comprehensive catalogue of genes and to define the content of repeats in the genome. However, distinguishing functional genes and pseudogenes and discerning recently duplicated gene copies may be sometimes problematic (Choulet *et al.* 2014b). The most difficult task poses the assembly of repetitive fraction of the genome. Sequence reads corresponding to repetitive sequences are prone to collapse in the assembly which results in an underrepresentation and misassembly of repetitive regions (Alkan *et al.* 2011).

Most of the whole-genome sequences published to date were produced using the WGS approach, which was for example applied in poplar (Tuskan *et al.* 2006), grapevine (Jaillon *et al.* 2007) or sorghum (Paterson *et al.* 2009). To date, the largest WGS assemblies were that of Norway spruce (Nystedt *et al.* 2013), white spruce (Birol *et al.* 2013) and loblolly pine (Zimin *et al.* 2014; each approximately 20 Gb). WGS sequencing of such large and highly repetitive plant genomes revealed that this approach is able to deliver a rough draft of the non-repetitive portion of a genome but assemblies remain highly fragmented with gaps of unknown sizes and sometimes

even not organized into chromosomes (Feuillet *et al.* 2011; Chapman *et al.* 2015). Potentially, the use of third-generation sequencing technologies providing long reads could increase the quality of WGS assemblies. However, these techniques are not widely used as mentioned previously. It is obvious that a high quality reference sequences with high contiguity and near-complete genome representation can be delivered only using a clone-by-clone sequencing strategy.

### 2.2.2.2 Clone-by-clone sequencing

Clone-by-clone (CBC) sequencing strategy (Figure 3b) is also known as hierarchical or BAC-by-BAC sequencing, the latter reflecting the fact that BAC vectors are currently most widely used for the clone-by-clone sequencing. The first and essential step, which precedes the very sequencing, is the construction of a physical contig map. Physical contig map is represented by an ordered set of overlapping BAC clones assembled into contigs and covering the genome. So-called minimal tilling path (MTP), the minimal set of overlapping clones ordered along the chromosomes and representing whole genome, is selected out of the whole physical map and sequenced. Individual MTP clones are fragmented before sequencing into small pieces and subsequently either cloned into a plasmid vector and sequenced using Sanger chemistry or sequenced directly by next-generation sequencing technologies, similarly as in the WGS approach. Pair-end reads are separately assembled for each clone. The physical contig map then provides a clue for merging of sequence contigs into scaffolds and finally for a construction of pseudomolecules of all individual chromosomes.

The apparent advantage of a clone-by-clone strategy compared to WGS approach is that it facilitates the correct assembly of regions harbouring sequences that occur in the genome in multiple identical or highly similar copies like duplicated and homoeologous genes, genes originating from multigene families and repetitive elements (Miller *et al.* 2010). The availability of a robust anchored physical map highly facilitates the production of pseudomolecules of individual chromosomes and prevents the assignment of sequences to incorrect chromosome. Since the anchoring is performed with BAC rather than sequence contigs, this approach, compared to the WGS, poses a significantly lower demand on the number of markers needed for positioning the sequences on chromosomes. The disadvantage lies in the necessity of

construction of a physical contig map and sequencing and assembling of clones individually. This is significantly more labour-intensive, time-consuming and expensive than direct sequencing of the genome as a whole.

The best quality plant genome sequences remain still those generated by the CBC approach (Feuillet *et al.* 2011), including sequences of *Arabidopsis thaliana* (AGI 2000) and rice (IRGSP 2005). In case of rice, other groups produced simultaneously draft sequences by the WGS strategy. A comparison of these drafts with the sequence produced by the CBC approach indicated that the clone-by-clone sequencing represents the only way to create a high quality reference genome sequence with a high contiguity and a near-complete genome representation. This approach remains a gold standard for sequencing of large and complex genomes containing a high amount of repetitive elements such as the genome of bread wheat (Philippe *et al.* 2012).

### 2.2.3  Wheat genome sequencing strategy

Typical features of the wheat genome, such as its huge size, presence of three nearly identical homoeologous subgenomes and high proportion of repeats, make production of a reference genome sequence a technical challenge. All these features have to be considered when selecting the appropriate sequencing strategy. The overview of wheat genome sequencing strategies is given in the Figure 4.

Some attempts were made using a WGS approach with the aim to deliver whole-genome data in a short time and for a reasonable cost (Brenchley *et al.* 2012; Chapman *et al.* 2015). Both these studies provided a rough insight into a wheat genome composition, however they also showed that a reference sequence cannot be obtain through the WGS approach and that the genome have to be dissected into smaller manageable pieces. This is enabled by the possibility of sorting individual chromosomes and chromosome arms using flow cytometry (Doležel *et al.* 2014). Individual fractions of the genome (chromosomes, arms) can be then sequenced separately applying either whole-chromosome shotgun or BAC-by-BAC sequencing approach (Figure 4). The strategy to sequence individual chromosomes or chromosome arms (chromosomal approach) was selected by the International Wheat Genome Sequencing Consortium (IWGSC) as a key strategy for obtaining a reference sequence of a wheat genome (Choulet *et al.* 2014b).
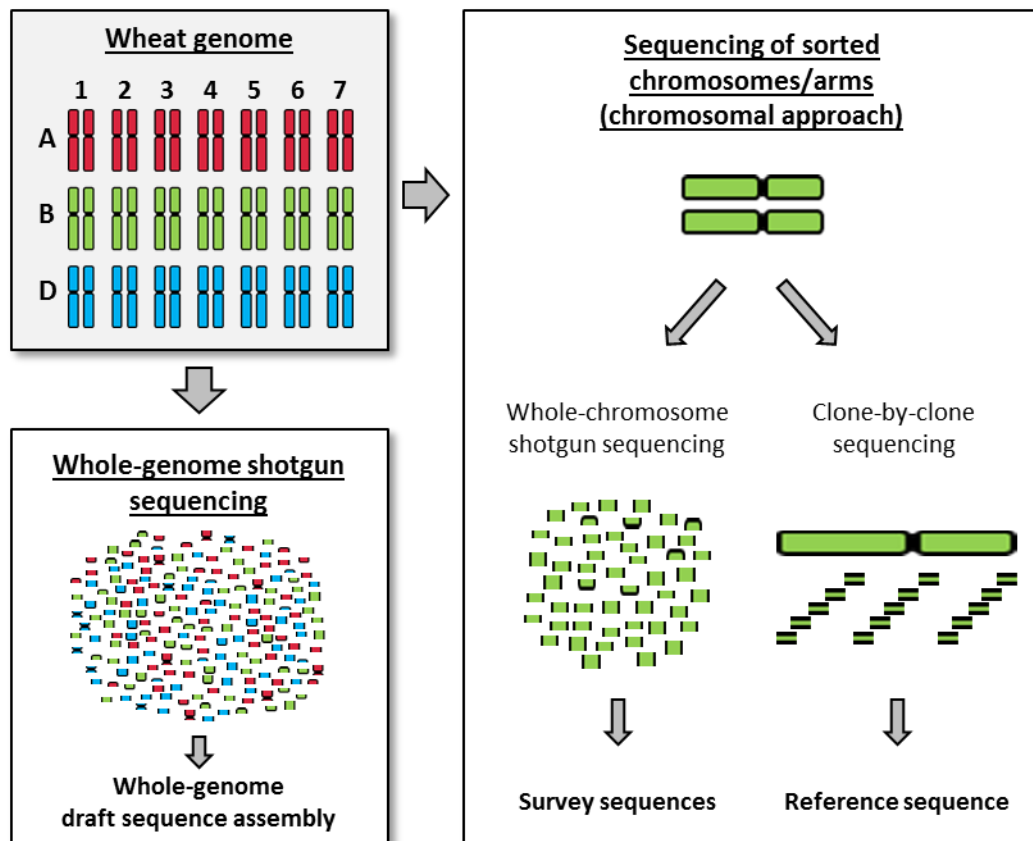
Figure 4: Overview of wheat genome sequencing strategies

### 2.2.3.1 Whole-genome shotgun sequencing

The first attempt to perform the whole-genome shotgun sequencing of wheat was made by Brenchley *et al.* in 2012. The genome of a wheat cultivar 'Chinese Spring' was sequenced with 5x coverage, using Roche/454 pyrosequencing technology. After filtering out reads containing repeats, a 3.8-Gb assembly containing low-copy sequences was obtained. The assembly resulted in identification of ~95,000 genes. A significant number of them were assigned to the A, B and D genome, respectively, by comparison with Illumina sequence assemblies of *Triticum monococcum*, cDNA assemblies of *Aegilops speltoides* and 454 sequences from *Ae. tauschii* representing wheat A-, B- and D-genome diploid progenitors. Searching for repetitive elements in sequencing reads revealed that repeats form 79 % of the sequencing data. Sequence analysis showed a highly dynamic structure of the wheat genome with significant loss of genes (between 10,000 and 16,000) compared to the three diploid progenitors, and an abundance of gene fragments.

Recently, Chapman *et al.* (2015) produced another whole-genome shotgun assembly of a hexaploid bread wheat line 'Synthetic W7984', represented by 9.1 Gb of the sequence scaffolds of which 7.1 Gb were assigned to chromosomal locations. The genome was sequenced by Illumina sequencing platform with 30x coverage. Computational integration of the WGS data with the sequence-based genetic map produced linked assemblies spanning entire chromosomes. However, these assemblies include predominantly the accessible low-copy portion of the genome capturing approximately three-quarters of known genes in a reasonably complete form.

In both cases, the WGS strategy could not deliver a contiguous reference sequence of the whole genome. Even if the WGS sequencing strategy enables to assemble low-copy regions of the genome, it falls short of capturing and correctly assembling the repetitive fraction. Moreover, sequence information of recently duplicated gene families can be lost and ordering and assigning sequence scaffolds to individual chromosomes may be a challenging task. Thus, these fragmented assemblies cannot be considered as a reference sequence of the whole genome (Feuillet *et al.* 2011; Choulet *et al.* 2014b).

### 2.2.3.2   Chromosomal approach

Chromosomal approach employs the flow cytometry sorting for dividing the wheat genome into individual chromosomes and chromosome arms. This approach enables to significantly reduce the complexity of the genome and to avoid the problem of the presence of three highly similar homoeologous subgenomes. Possibility of isolation of individual chromosomes/arms also enables to divide the work among members of the IWGSC all over the world (Doležel *et al.* 2007; Choulet *et al.* 2014b).

#### 2.2.3.2.1  Whole-chromosome shotgun sequencing

Recently, the ordered and structured draft sequences (survey sequences) of all 21 chromosomes of wheat were generated by shotgun sequencing of individual flow-sorted chromosomes (IWGSC 2014). Each chromosome arm and the whole chromosome 3B of cv. 'Chinese Spring' ('CS') were sequenced to a depth of 30 -

241x with Illumina sequencing platform and assembled *de novo*. The final assembly is represented by 10.2 Gb, comprising 60 % of the whole genome.

Whole-chromosome shotgun sequencing provided a quick view into a genome composition. Annotation of genes performed by the comparison with annotated gene sequences of closely related species (*Brachypodium distachyon*, rice, sorghum, and barley), publically available wheat full length cDNAs and RNA-seq data revealed presence of 106,000 functional protein-coding genes. Comparative gene analysis with extant diploid and tetraploid wheat relatives showed a high sequence similarity and structural conservation. No significant gene loss was observed after polyploidization. However, dynamic gene gain, loss, and duplication took place across the genome since the divergence of the wheat lineages. Individual subgenomes exhibit high degree of transcriptional autonomy with no global dominance of a particular subgenome. Investigation of raw read data revealed that 81 % of them contain repetitive elements (IWGSC 2014).

Despite the fact that the draft assembly does not cover the whole genome, and suffers from multiple imperfections, it has a potential for accelerating the gene isolation and marker development. In combination with other complementary resources, draft assembly supports the assembly of the reference wheat genome sequence generated by BAC-by-BAC approach (IWGSC 2014).

### 2.2.3.2.2  BAC-by-BAC sequencing

BAC-by-BAC sequencing is a several-step procedure which allows producing a high quality reference genome sequence as proved by the recently published reference sequence of the largest wheat chromosome - 3B (Choulet *et al.* 2014a). The approach adopted by the IWGSC begins with the construction of a BAC library from flow-sorted chromosomes/arms. All individual clones in the library are fingerprinted and a physical contig map is constructed based on overlaps between the fingerprints. Then, the minimal tilling path (MTP) is established and clones representing the MTP are shotgun sequenced using NGS technologies. Final steps of the sequencing are the sequence assembly, construction of a pseudomolecule and its validation (Choulet *et al.* 2014b; Doležel *et al.* 2014).

### 2.2.3.2.2.1 BAC libraries

In general, DNA library consists of a set of DNA fragments isolated from the organism, ligated into a vector and introduced into a host cell. Several types of vectors have been used for the construction of the large insert DNA libraries, among them the BAC (bacterial artificial chromosome; Shizuya *et al.* 1992) vector. BAC became the predominantly used vector for its versatility, stable maintenance of insert sized between 100–200 kb, relatively easy library construction and easy manipulation of clones (Luo *et al.* 2003; Šafář *et al.* 2010).

First BAC libraries in wheat were constructed for the whole genome (Allouis *et al.* 2003; Nilmalgoda *et al.* 2003; Ling and Chen 2005; Ratnayaka *et al.* 2005; Shen *et al.* 2005). However, construction, analysis and maintenance of those libraries are expensive and laborious. Even if the fingerprinting was feasible, it would not be possible to construct a reliable physical map. In order to facilitate the wheat genome analysis, libraries from diploid and tetraploid progenitors and relatives of wheat were constructed (Akhunov *et al.* 2005; Lijavetzky *et al.* 1999; Chen *et al.* 2002; Cenci *et al.* 2003). Nevertheless, those libraries could not fully substitute libraries constructed directly from hexaploid wheat. The possibility of disecting the wheat genome into individual chromosomes/arms using flow cytometry enabled construction of wheat chromosome-specific BAC libraries.

The first chromosome-specific BAC library was constructed for the chromosome 3B (Šafář *et al.* 2004) as this chromosome was the only one that could be sorted from the standard karyotype of bread wheat. This library was followed by the library specific for the group of chromosomes 1D, 4D and 6D (Janda *et al.* 2004) and  libraries specific for chromosome arms 1BS (Janda *et al.* 2006) and 3AS (Šafář *et al.* 2009) respectively. All above mentioned libraries belong to a group of 'first generation' libraries, characterized by a small insert size (less than 100 kb) due to application of a single size-selection step. Improvement in cloning efficiency allowed introducing the second size-selection step, which laid the foundations for the 'second-generation libraries', whose average insert size exceeds 110 kb and reaches as much as 182 kb (Šafář *et al.* 2010). Hitherto, chromosome- or chromosome arm-specific BAC libraries have been constructed for all chromosomes of bread wheat (IEB genomic resource database, http://olomouc.ueb.cas.cz/dna-libraries/cereals). All these libraries have been constructed from cv. 'Chinese Spring' and besides being a

source of clones for BAC-by-BAC sequencing, they represent a useful resource for targeted marker development and positional cloning.

### 2.2.3.2.2.2 Physical mapping

Physical contig map is a contiguous ordered set of overlapping DNA fragments (BAC clones), reflecting the sequence of nucleotides in a chromosome. Distances in the physical map are expressed in physical distance units (base pairs) (Luo *et al.* 2003; Meyers *et al.* 2004). Sequencing projects aiming to produce a reference genome sequence require a robust physical map anchored to genetic, cytogenetic or radiation hybrid (RH) maps. The quality of the physical map relies on the technology and computational resources used for the map construction and also level of anchoring of the map to the chromosome. A sequencing-ready physical map should consist of a minimal number of contigs which are anchored, ordered and oriented and cover maximal length of the chromosome. The number of chimerical and misassembled contigs should be minimized (Philippe *et al.* 2012).

So far, physical contig maps in wheat (Table 1) were published for four whole chromosomes - 5A, 6A, 3B, 6B (Barabaschi *et al.* accepted; Poursarebani *et al.* 2014; Paux *et al.* 2008; Kobayashi *et al.* 2015) and five individual chromosome arms - 1AS, 1AL, 1BS, 1BL, 5DS (Breen *et al.* 2013; Lucas *et al.* 2013; Raats *et al.* 2013; Philippe *et al.* 2013; Akpinar *et al.* 2015).

Table 1: Currently available physical maps in wheat

| | Wheat chromosome/arm | Fingerprinting strategy | Assembly software | Reference |
|---|---|---|---|---|
| **Chromosome** | **5A** | HICF | FPC[1], LTC[1] | Barabaschi *et al.* (accepted) |
| | **6A** | WGP | FPC, LTC[1] | Poursarebani *et al.* (2014) |
| | **3B** | HICF | FPC | Paux *et al.* (2008) |
| | **6B** | WGP | FPC | Kobayashi *et al.* (2015) |
| **Chromosome arm** | **1AS** | HICF | FPC[1], LTC | Breen *et al.* (2013) |
| | **1AL** | HICF | FPC[1], LTC | Lucas *et al.* (2013) |
| | **1BS** | HICF | LTC | Raats *et al.* (2013) |
| | **1BL** | HICF | FPC[1], LTC | Philippe *et al.* (2013) |
| | **5DS** | HICF | LTC | Akpinar *et al.* (2015) |

[1] assembly used for the MTP selection

### 2.2.3.2.2.2.1 Basic strategy of physical map construction

Construction of the physical map starts with the generation of a specific fingerprint which characterises each BAC clone. Based on overlaps of these fingerprints, BACs are assembled into contigs. Once the assembly is finished, a minimal set of overlapping clones, the minimal tilling path (MTP), covering the whole chromosome or genome is selected. MTP clones then provide a template for sequencing (Meyers *et al.* 2004). Currently, two approaches are employed to obtain the fingerprints. The traditional one is based on generation of restriction patterns for individual clones (Luo *et al.* 2003), while the newer one, called whole genome profiling (WGP; van Oeveren *et al.* 2011), on generating short sequence tags.

The former method is based on the digestion of individual BAC clones by specific restriction enzymes yielding a set of restriction fragments typical for each clone. The most widely used approach is the high-information-content fingerprinting (HICF), utilizing the SNaPshot labelling kit to fluorescently label restriction fragments obtained with a set of restriction endonucleases (Figure 5). The restriction fragments are separated by capillary electrophoresis and fragment sizes are determined (Luo *et al.* 2003). The HICF method was applied in the construction of the majority of wheat physical maps published so far, namely the map of chromosome 5A, 3B and maps of chromosome arms 1AS, 1AL, 1BS, 1BL and 5DS.



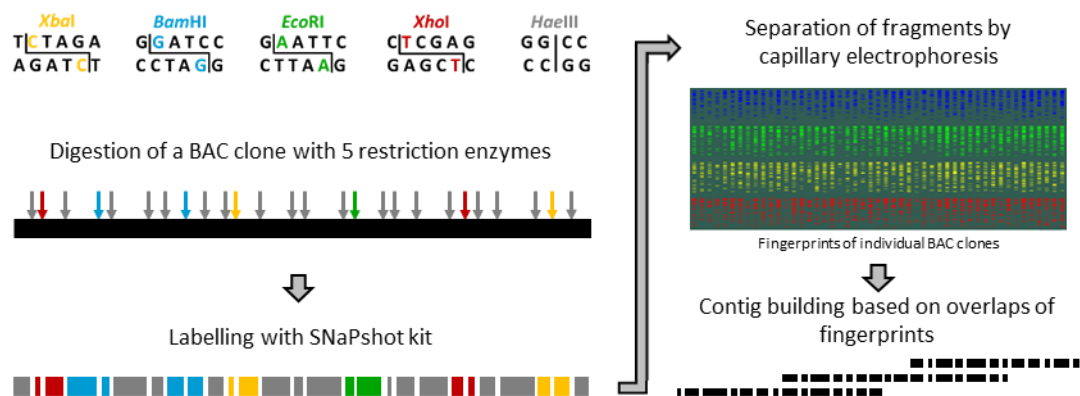Figure 5: Principle of HICF method. DNA of BAC clones is digested with 5 restriction enzymes. Fragments with recessive ends are labelled with SNaPshot kit. Fragments are separated using capillary electrophoresis. Those which are labelled are detected and sized. Physical contigs are assembled based on overlaps between BAC clones fingerprints (restriction spectra). (Adapted from Meyers *et al.* 2004.)

The whole genome profiling is a high-resolution sequence-based physical mapping technique employing NGS technologies in order to produce sequence tags along the BAC clones (Figure 6). The possibility of pooling BAC clones prior to the WGP, scaling the pooling complexity and adjusting of tag length allows to fully exploit the potential of the method while reducing costs for BAC DNA preparation. The first experiment in wheat conducted by Philippe *et al.* (2012) on a subset of BAC contigs from the 3B chromosome revealed that this method shows suitable even for the complex genome of bread wheat. Recently, a physical map of the wheat chromosome 6A followed by the map of the chromosome 6B were constructed by the WGP method. It was proved that the WGP is a more robust, less laborious and more efficient approach for building physical maps compared to the HICF (Philippe *et al.* 2012; Poursarebani *et al.* 2014), but its wider use is prevented by a higher cost.
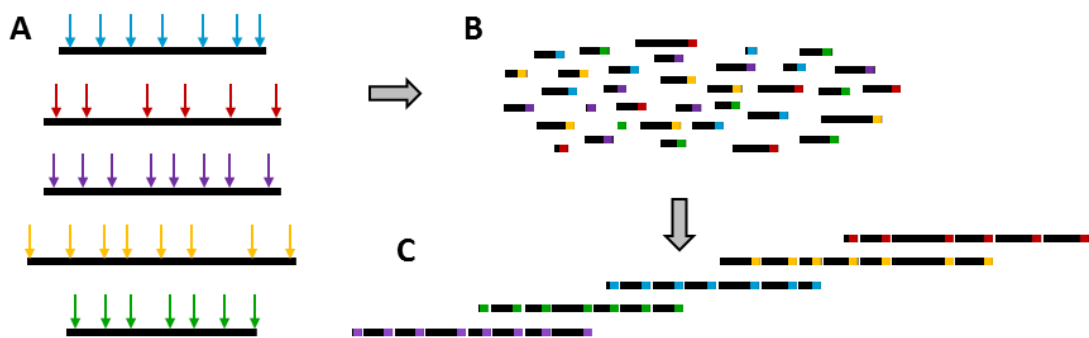


Figure 6: Scheme of the whole genome profiling. (**A**) BAC clones are cut by a couple of restriction enzymes. Barcoded adapters specific for each BAC clone are ligated to the ends of the fragments (marked in colours). (**B**) Short sequence (tag) from each fragment is obtained. (**C**) BAC contigs are assembled based on overlapping sets of sequence tags. (Adapted from van Oeveren *et al.* 2011.)

### 2.2.3.2.2.2.2 *Physical map assembly*

The assembly of the physical map is based on the identification of significant overlaps between BAC clones which are represented by shared sets of restriction fragments or sequence tags. Currently, two software packages employing different algorithms are widely used for the assembly of the physical map – FingerPrinted Contig (FPC; Soderlund *et al.* 1997) and Linear Topological Contig (LTC; Frenkel *et al.* 2010). Even though both these softwares were originally developed for the input

data generated by former fingerprinting techniques, they are able to process also WGP data (Poursarebani *et al.* 2014).

### *FingerPrinted Contig (FPC)*

FingerPrinted Contig (FPC) was the first software widely used for the construction of physical maps in wheat. This software uses a specific algorithm to compute the probability of overlap between clones, termed Sulston score (Sulston *et al.* 1988). A Sulston score cut-off value is set by the user and determines the threshold for considering two clones as overlapping. A low-score value corresponds to a larger overlap and thus higher stringency of the assembly (Luo *et al.* 2003).

FPC generates the physical map by building a consensus map (CB map) of each contig, which is a kind of a low-resolution restriction map. The initial CB map is built applying stringent criteria (low Sulston cut-off value) resulting in relatively large number of short reliable contigs. In subsequent steps, the cut-off value is increased which enables to add single clones to the end of contigs or merge contigs and thus increase the contig length (Soderlund *et al.* 2000). The cut-off value has to be adjusted for each individual genome based on its size and amount of repeats to assure the construction of a reliable physical map with a low number of false positive and non-identified overlaps (Luo *et al.* 2003). Once the assembly is complete, FPC selects automatically the minimal tilling path according to a user-defined size of the overlap between MTP clones.

### *Linear Topological Contig (LTC)*

Aiming to solve problems of FPC assemblies, such as short contigs built under stringent criteria and unreliable assembly of problematic regions, Frenkel *et al.* (2010) developed the Linear Topological Contig (LTC) software. This software employs systematic exploring of the topological contig (3D) structure and performs iterative clone clustering and ordering which leads to building of more reliable and longer contigs than those produced by FPC (Frenkel *et al.* 2010). Instead of Sulston score, LTC uses different more accurate metrics for estimation of the probability of clone overlaps.

In contrast to FPC, LTC starts building initial clusters with a liberal cut-off and continues with a gradual increase in stringency (Jain and Dubes 1988). Problematic clones and connections are identifiable by visual representation of

clusters as nets of significant clone overlaps. Visual representation also enables to identify branching clusters which contradict the one-dimensional linear organization of the chromosome (Figure 7). Hence, such clusters can be split into two or more linear structures. An end-to-end merging strategy can be finally applied to connect linear contigs into long scaffolds (Raats *et al.* 2013). Similarly like in FPC, the last step of the assembly process is the selection of MTP clones applying user-defined level of overlap between clones.

Apart from the *de novo* construction of the physical map, LTC software enables verification and repair of contigs built by FPC. LTC recalculates probabilities of clone overlaps and represents each contig as a net of significant overlaps. The connections within this net and topological linearity are tested, which allows to find branching nodes and split contigs containing the problematic sites. (Frenkel *et al.* 2010).



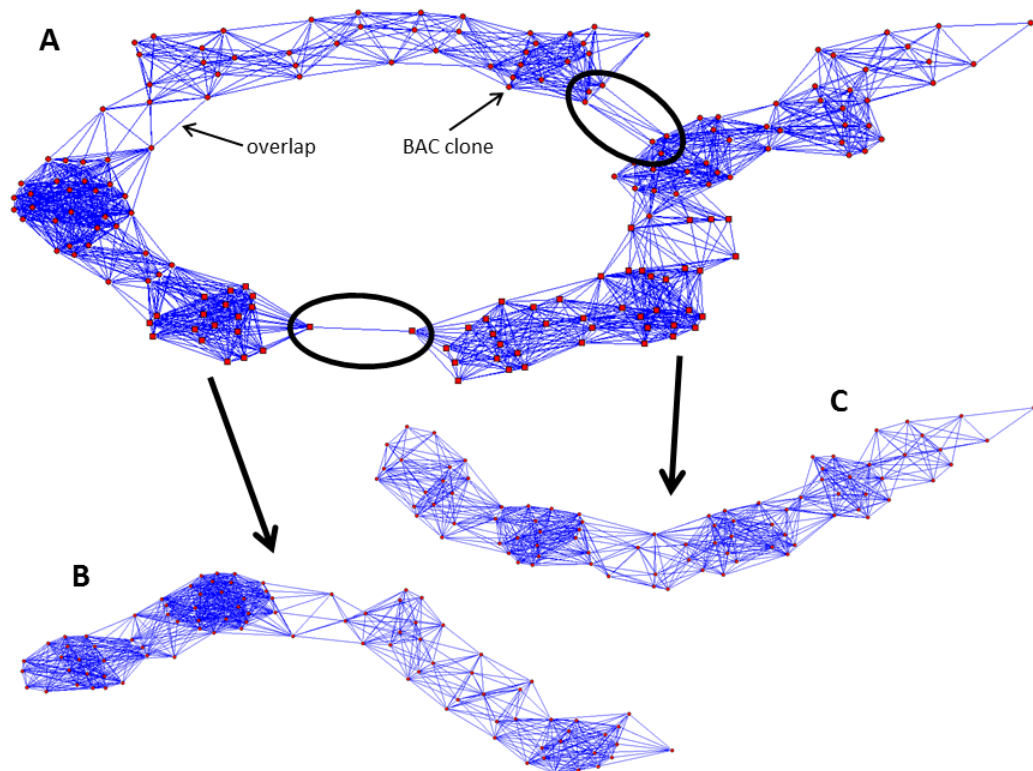Figure 7: Identification and correction of a branching cluster visualized by LTC. Physical contigs are visualised as nets of significant overlaps. Red dots represent clones; blue lines represent significant clone overlaps. Branching regions are marked by black ovals. Contig A exhibiting non-linear topological structure was split into contigs B and C by removing questionable clones causing branching.

*2.2.3.2.2.2.3 Anchoring of the physical map*

To fully draw the benefits that the physical maps provide for positional cloning, marker development and sequencing, BAC contigs have to be anchored to the chromosomes and ordered along them by molecular markers. In general, two opposite approaches are applicable for anchoring – the forward and the reverse approach.

Forward anchoring is based on localisation of previously mapped molecular markers in contigs of the physical map. Originally, genetically mapped markers stored in databases were used for screening the BAC library by PCR or by hybridization using high-density filters (Paux *et al.* 2008; Gardiner *et al.* 2004). Nowadays, high-throughput array-based platforms developed originally for highly parallel SNP genotyping (e.g. Infinium SNP array) and microarray platforms designated for the study of the gene expression (e.g. NimbleGen UniGene microarrays) are commonly used for forward anchoring of the physical maps (Luo *et al.* 2013; Breen *et al.* 2013; Lucas *et al.* 2013; Philippe *et al.* 2013; Raats *et al.* 2013). In order to maximize the effectiveness of anchoring procedures and minimize the costs, multidimensional pooling of BAC clones has been developed. BAC pools consist of a set of BAC clones representing a clearly defined fraction of the BAC library, combined into one sample. Different pooling strategies have been proposed, including 3D, 5D and 6D strategies (Barillot *et al.* 1991; Luo *et al.* 2009; You *et al.* 2010; Bruno *et al.* 1995). Out of them, the most widely used is the three-dimensional (3D) pooling for which pools are generated from plates, rows and columns of the BAC library plates.

Another approach to anchoring makes advantage of available partial (BAC ends) or complete sequence information of BAC clones, which can be used as a template for *in silico* anchoring. This procedure relies on the identification of homology between the sequence of a previously mapped marker and a sequence of a BAC clone. Besides genetic maps of bread wheat, the sequences of BAC clones can be anchored to genome zippers (Raats *et al.* 2013; Breen *et al.* 2013) or to a genetic map of a wheat related species. Recently, Akpinar *et al.* (2015) employed a SNP-based genetic map of *Aegilops tauschii*, the D-genome progenitor, for anchoring of BAC contigs of the chromosome arm 5DS.

The reverse approach to contig anchoring exploits sequences of the BAC clones or BAC-end sequences for the development of molecular markers, which are

subsequently positioned on the chromosome by linkage mapping or using special cytogenetic stocks - deletion lines (Raats *et al.* 2013; Philippe *et al.* 2013), or radiation hybrid (RH) panels (Riera-Lizarazu *et al.* 2010; Kumar *et al.* 2012; Tiwari *et al.* 2012). Radiation hybrid mapping, which relies on assaying radiation-induced chromosomal fragments for the presence of particular markers, is especially useful for anchoring BACs in regions with a low recombination frequency, such as centromeric and pericentromeric regions (Feuillet *et al.* 2012).

### 2.2.3.2.2.3 Sequencing and sequence assembly

The process of BAC-by-BAC sequencing of the chromosome is accomplished by the sequencing of MTP clones. Clones are sequenced individually or more often in pools of several BACs to reduce the sequencing costs (Doležel *et al.* 2014). Nowadays, mostly the NGS technologies are used for the shotgun sequencing of BAC clones of which the Illumina sequencing platform prevails over other technologies. Several approaches can be combined to achieve a higher sequence quality. In general, short paired-end reads with the standard distance of read pairs up to 300 bp are generated with high coverage. These data are assembled into sequence contigs. Subsequently, contigs are merged into scaffolds employing the information gained from mate-pair sequencing which provides short paired sequences situated in distances of several to tens of kilobases. Such information enables to span long regions which cannot be assembled from short pair-end read data (Choulet *et al.* 2014b). Additional information derived from sequencing of the ends of BAC clones (BES) is also useful for the identification of BAC ends and mainly for the deconvolution of pools containing several BACs. The above described pipeline can be modified or supplemented by single-molecule sequencing technologies (PacBio, Oxford Nanopore) providing several-kilobases reads (Metzker *et al.* 2010).

The generation of sequence reads is followed by sequence assembly. Basically, the sequence assembly represents a process in which bioinformatics software is used to align overlapping reads and thus assemble the continuous sequence of the genome or chromosome. The type of the assembly algorithm has to be considered with respect to input data (Schadt *et al.* 2010). Various assemblers based on different algorithms are used to assemble sequence reads produced by NGS technologies including Newbler (http://www.454.com/), Edna (Hernandez *et al.*

2008), Velvet (Zerbino and Birney 2008), ABySS (Simpson *et al.* 2009), SOAPdenovo (Li *et al.* 2009) and many others. For a more comprehensive assembly, pipelines combining several assembly tools are designed to join all the sequencing data generated by multiple sequencing technologies (Schatz *et al.* 2010). The final step of the assembly is the building of a pseudomolecule by ordering and orienting sequence scaffolds along the chromosome through the integration with the physical and genetic maps.

Till now, the reference sequence was generated only for the largest chromosome of wheat – the chromosome 3B (Choulet *et al.* 2014a). Sequencing strategy was based on mate-pair sequencing of pooled MTP BAC clones using 454/Roche NGS technology. Sequence assembly was performed using Newbler assembler and the final assembly was generated by incorporating the Illumina chromosome 3B shotgun and BAC-end sequences. Resulting sequence scaffolds were anchored to the physical map of 3B (Paux *et al.* 2008). Finally, the pseudomolecule was built by ordering 1,358 scaffolds along the chromosome using an ordered set of 2,594 anchored SNP markers. The length of the pseudomolecule is 774.4 Mb which represents 93 % of the initial sequence assembly obtained for the 3B chromosome (Choulet *et al.* 2014a).


### 2.2.3.2.2.4  Validation of the sequence assembly

Despite the fact that NGS technologies produce huge amount of data in a short time with relatively low costs and thus can produce easily a sufficient amount of data for whole-genome projects, the assembly can be problematic in particular regions of the genome. Plant genomes are known for the high portion of repetitive elements that often form nested structures which are prone to collapse in the sequence assembly. This results in the underrepresentation and misassembly of repetitive regions in the final assembly or merging unrelated regions of the genome (Alkan *et al.* 2011). In wheat, the size of transposable elements ranges between ~50 and ~31,000 bp and these elements form highly nested structures spread over distances as large as 200 kb (Choulet *et al.* 2010). Mate-pair reads are essential for traversing repeats in a genome (Choulet *et al.* 2014b) but with current technologies, they cannot span regions larger than approximately 50 kb in length (http://www.lucigen.com/). Another challenge to sequence assembling poses the

appearance of duplicated sequences (Alkan *et al.* 2011). Incorrect assembly of these sequences can lead to merging of regions which are in fact outlying in the genome. For these reasons, alternative technologies have to be used for the sequence validation and potential correction of the assembly.

One of the options for the sequence validation and identification of misassemblies is to perform the cytogenetic mapping (Febrer *et al.* 2010; Islam-Faridi *et al.* 2009). An example of such method is the fluorescence *in situ* hybridization (FISH) with BAC clones as probes (BAC –FISH). Using this method, Shearer *et al.* (2014) were able to find incorrectly arranged sequence scaffolds which affected in total one-third of the tomato genome sequence. Even if this approach is feasible, it is time consuming and laborious which hampers its applicability in large genomes. Moreover, the presence of dispersed repetitive DNA typical for the wheat genome makes preparation of single-copy probes a tedious task (Janda *et al.* 2006).

Alternatives to cytogenetic mapping provide optical mapping (Zhou and Schwartz 2004; Zhou *et al.* 2007a) and genome mapping in nanochannel arrays (Lam *et al.* 2012). Both these techniques produce a set of sequence motif consensus maps covering individual chromosome(s) or the whole genome. Each consensus map is assembled based on the overlaps between sequence motif patterns of individual DNA molecules which are up to several megabases long.

Optical mapping analyses long DNA molecules stretched in a microfluidic device (chip). DNA molecules are attached to the solid surface by electrostatic interactions and subsequently digested by a specific restriction endonuclease and stained. Molecules are automatically imaged and the length of fragments is estimated from their integrated fluorescence intensity. Consensus optical maps are constructed based on overlaps between restriction patterns of individual molecules (Zhou and Schwartz 2004). Despite the fact that this technique was used for mapping of many genomes, including microbial genomes (Zhou and Schwartz 2004; Zhou *et al.* 2004; Riley *et al.* 2011 etc.), genome of rice (Zhou *et al.* 2007b), maize (Zhou *et al.* 2009), human (Teague *et al.* 2010), budgerigar (Ganapathy *et al.* 2014) or ostrich (Zhang *et al.* 2015), it suffers from several disadvantages. The major ones are uneven and irreproducible stretching of DNA molecules, leading to inaccurate sizing of fragment length, and a low throughput of the technique. Those handicaps were overcome by genome mapping in nanochannel arrays (Irys platform developed by BioNano company; http://www.bionanogenomics.com/).

The genome mapping is based on labelling of the long DNA molecules in the enzyme-specific nicking sites and subsequent automatic massively parallel imaging of the DNA molecules in a chip-based nanochannel array. The diameter of nanochannels (45 nm) is specifically designed to harbour only one DNA molecule and to assure uniform linearization and stretching of the molecules. The analysis is performed in cycles. In each cycle, the labelled molecules flow from the reservoir into the nanochannels where they are imaged and subsequently washed out. As the chip can undergo tens of cycles, the technique enables high-throughput data collection (Lam *et al.* 2012; Cao *et al.* 2014). Hitherto, the genome mapping has been applied in several studies. The first one focused on mapping a 4.7-Mb highly variable region of the human genome harbouring the major histocompatibility complex (MHC) (Lam *et al.* 2012). Subsequently, a genome map of the human genome was constructed with the aim to detect structural variation (Cao *et al.* 2014). Recently, the data resulting from genome mapping in nanochannel arrays were used in combination with single-molecule sequencing data for the hybrid sequence assembly of the human genome (Pendleton *et al.* 2015). The technique was also applied in plants for *de novo* assembly of a 2.1-Mb region in the wheat D-genome progenitor *Aegilops tauschii* (Hastie *et al.* 2013). The study conducted by Staňková *et al.* (accepted) on the 7DS chromosome arm of wheat confirmed that the DNA of wheat chromosomes obtained by flow sorting is suitable for genome mapping. A *de novo* constructed genome map of the 7DS arm provided a tool for validation and improvement of the physical contig map and enabled to identify a chromosome region harbouring a long array of tandem repeats.

## 2.3 Positional cloning

Positional (or map-based) cloning is a process of the isolation of a gene underlying a trait of interest without a prior knowledge of the gene sequence and product. Isolation of the gene is essential for understanding the trait at molecular level and studying its mode of action in the plant (Krattinger *et al.* 2009a). Once isolated, the gene can be directly transferred to the plant lacking the trait of interest by means of genetic engineering. The positional cloning approach can be applied for isolation of major genes as well as quantitative trait loci (QTLs). Compared to major genes, positional cloning of QTLs is a complex procedure, which requires more sophisticated gene mapping and, optimally, trait mendelization. Since this work involved mapping and positional cloning of a major gene, general strategy of major gene positional cloning will be described further.

The positional cloning process (Figure 8) starts with determining the approximate position of the gene of interest on the chromosome by genetic mapping. A mapping population derived from the cross between parents differing in a trait of interest constitutes the basis for the reliable gene mapping (Krattinger *et al.* 2009a). A significant difference between parent phenotypes and the stability of the phenotype are important for the accurate phenotype scoring. In order to map the gene of interest, phenotypic data and molecular marker data are collected for all progenies in the population. Initially, the mapping population is screened with a spectrum of genetic markers which were previously mapped in various mapping populations. Screening the population with those markers mapped all over the genome enables to locate the gene on a particular chromosome and to determine its approximate position. Nowadays, high throughput arrays/chip-based genotyping platforms containing thousands markers (SNP, DArT, ISBP) are available enabling to speed up the first step of positional cloning (Gupta *et al.* 2013a). For a rough mapping of a gene, a population of 100-200 plants is sufficient (Krattinger *et al.* 2009a; Schneider 2005). Rough mapping is followed by a construction of a high-density genetic map. New markers have to be developed and mapped in a high-resolution mapping population comprising thousands of individuals (Krattinger *et al.* 2009a; Schneider 2005) in order to precisely allocate the gene. Only the individuals that are recombinant between available flanking markers are selected and used for further

genetic mapping and phenotyping. Once the interval delimited by flanking markers is reduced below 1 cM, the physical mapping stage can be approached.



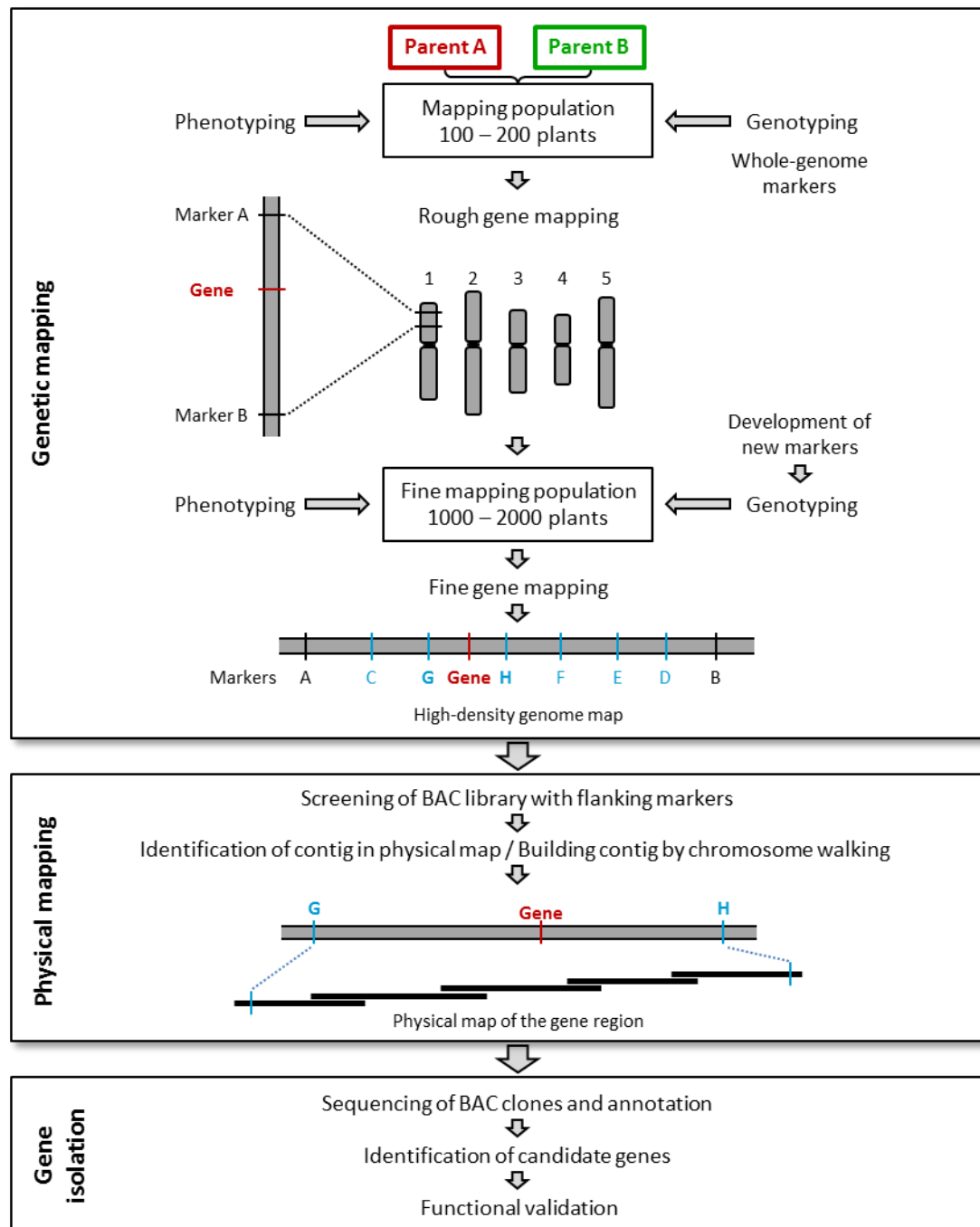Figure 8: Scheme of the positional cloning process. (Adapted from Krattinger *et al.* 2009a.)

Physical map spanning the region between flanking markers by BAC clones was traditionally constructed by a laborious process of chromosome walking (Krattinger *et al.* 2009a). The availability of the physical map of the whole genome or particular chromosome can highly facilitate positional cloning of the gene by

substituting for the chromosome walking step (Feuillet *et al.* 2012). In ideal scenario, screening of the BAC library with flanking markers reveals BAC clones belonging to one physical map contig and thus spanning continuously the gene region. Otherwise, new markers are developed from distal ends of contigs identified by the flanking markers and subsequently used to screen the library with the aim to find neighbouring contig(s). Once a BAC contig spanning the region between flanking markers is identified, the constituting BAC clones are sequenced. Proper annotation, based on the comparison with gene and protein sequences deposited in databases, leads to identification of candidate genes, i.e. genes that might be responsible for the trait of interest (Krattinger *et al.* 2009a; Salvi and Tuberosa 2005).

The final step of the positional cloning process is the validation of the gene function, which can be performed by over-expression or down-regulation of the gene through a stable transformation (*Agrobacterium* transformation, particle bombardment), RNAi silencing (Waterhouse and Helliwell 2003), virus induced gene silencing (VIGS; Gupta *et al.* 2013b) or by genetic complementation of a known mutant (Doebley *et al.* 1997). Other methods of gene knockout like TILLING (Slade *et al.* 2005) or gene editing by TALEN and CRISPR-Cas9 are also applicable (Wang *et al.* 2015).

## 2.3.1 Positional cloning in wheat

Positional cloning in wheat is hampered by the previously mentioned characteristic features of the genome but also by an uneven spatial distribution of the recombination (Saintenac *et al.* 2009) and a low level of polymorphism (Suenaga *et al.* 2005). Until recently, the tools facilitating the positional cloning including BAC libraries, physical maps and sequence information were underdeveloped in wheat. Recently, the chromosome based strategy enabled advancement of those tools. Combination of wheat genomic resources with those from diploid and tetraploid wheat relatives enabled successful isolation of several genes from wheat via positional cloning. As each positional cloning project is specific, it is not possible to outline a general approach for positional cloning in wheat. Several examples of strategies applied for the gene isolation in wheat are described below. Those approaches can be combined or supplemented by other techniques.

Like in other species, construction of a high-density genetic map is equally critical for positional cloning in wheat. To saturate the genetic map in the region of interest, many markers have to be developed in a targeted manner. This task is quite challenging in wheat where three highly similar subgenomes are present. With the limited sequence information in the past, represented mainly by ESTs or cDNA sequences (Choulet *et al.* 2014b), it was possible to design new markers predominantly in genic regions, which are highly similar among homoeologous subgenomes. This leads quite often to amplification of all three homoeologous products, which could not be distinguished from each other (Krattinger *et al.* 2009a). Chromosome-specific survey sequences produced for several individual chromosomes (Wicker *et al.* 2011; Vitulo *et al.* 2011; Hernandez *et al.* 2012; Berkman *et al.* 2013) and later for all chromosomes of wheat by IWGSC (2014) provide sufficient substrate for the marker design, as complete genes and also gene surrounding sequences are available. Sequences of all three homoeologs can be compared in order to find polymorphisms and design genome-specific markers. Moreover, amplified DNA of flow–sorted chromosomes can be used for the verification of the chromosomal specificity of each marker by PCR (Staňková *et al.* 2015).

In earlier times, the lack of the sequence information could be partially substituted also by sequences of wheat progenitors and close relatives from the grass family. Those species exhibit a high level of collinearity and a core set of coding genes conserved among their genomes was observed (Salse *et al.* 2009). A combination of the sequence information of model grasses like rice or *Brachypodium distachyon* with wheat EST sequences was used several times to design genetic markers (Schnurbusch *et al.* 2007; Terracciano *et al.* 2013; Qin *et al.* 2011; Quraishi *et al.* 2009). However, the efficiency of designing locus-specific primers was rather low due to insufficient subgenome specificity and lack of the information about the genomic context of ESTs. Recently, the collinearity between grass genomes was exploited for the construction of genome zippers. The first genome zippers were developed for individual barley chromosomes (Mayer *et al.* 2009, 2011). Barley genome zipper comprises an ordered set of *Brachypodium distachyon*, rice and sorghum genes anchored to a backbone of genetically mapped markers represented by barley ESTs (Close *et al.* 2009) and combined with barley 454 sequence reads, non-mapped barley ESTs and barley full-length cDNAs. A set of genome zippers

was developed in the same manner also for all wheat chromosomes (IWGSC 2014). The genome zipper showed to be a useful tool in targeted marker development. Based on the sequence homology between sequences from the region of interest and sequences of genes constituting the genome zipper, it is possible to delimit the interval in the genome zipper corresponding to the region around the gene to be cloned. Sequences found in this interval can be used together with chromosome specific survey sequences to design locus specific markers (Staňková *et al.* 2015).

Currently, construction of physical contigs maps for all wheat chromosomes is close to completion. Availability of the physical map enables to skip the laborious chromosome walking procedure. In early positional cloning projects, a so-called "shuttle mapping strategy" taking advantage of the availability of physical maps constructed in diploid and tetraploid relatives of wheat was applied in order to facilitate positional cloning (Huang *et al.* 2003; Feuillet *et al.* 2003; Yahiaoiu *et al.* 2004). In case of the shuttle mapping, genetic mapping and gene isolation were performed in hexaploid wheat while the physical contig spanning the gene region was identified in physical contig map originating from diploid or tetraploid species.

Up to now, 16 genes have been isolated from wheat since 2003. An overview of all those genes is given in Table 2. Half of the genes hitherto cloned are those underlying a resistance to different fungal pathogens including leaf rust (*Lr* genes), stem rust (*Sr* genes), stripe rust (*Yr36*) and powdery mildew (*Pm3b*). Another disease resistance-like gene *Tsn1* conferring sensitivity to *ToxA* effector produced by fungal pathogens was isolated by Faris *et al*. (2010). Genes conferring a resistance are of a big agronomical importance as the use of resistant cultivars increases the yield while minimizing the use of pesticides. Second group of genes is represented by genes controlling a flowering time (*VRN* genes). Remaining genes include a *Q* gene conferring the free-threshing character and a square-headed phenotype of spikes, *Ph1* gene controlling correct pairing of homologous chromosomes and *Gpc-B1* controlling senescence and grain protein, zinc and iron content. Although the number of genes cloned so far is not high, there is an assumption that the reference sequence available in a near future will significantly facilitate the positional-cloning process (Feuillet *et al.* 2011).

Table 2: Summary of genes isolated by positional cloning in wheat

| Gene | Function | Chromosome | Reference |
|------|----------|------------|-----------|
| *Lr1* | Leaf rust resistance gene | 5DL | Cloutier *et al.* (2007) |
| *Lr10* | Leaf rust resistance gene | 1AS | Feuillet *et al.* (2003) |
| *Lr21* | Leaf rust resistance gene | 1DS | Huang *et al.* (2003) |
| *Lr34* | Leaf rust resistance gene | 7DS | Krattinger *et al.* (2009b) |
| *Sr33* | Stem rust resistance gene | 1DS | Periyannan *et al.* (2013) |
| *Sr35* | Stem rust resistance gene | 3AL | Saintenac *et al.* (2013) |
| *Yr36* | Stripe rust resistance gene | 6BS | Fu *et al.* (2009) |
| *Pm3b* | Powdery mildew resistance gene | 1AS | Yahiaoui *et al.* (2004) |
| *Tsn1* | Disease resistance-like gene conferring sensitivity to *ToxA* effector produced by fungal pathogens | 5BL | Faris *et al.* (2010) |
| *VRN1* | MADS-box transcription factor controlling flowering | 5AL | Yan *et al.* (2003) |
| *VRN2* | Flowering repressor down-regulated by vernalization | 5AL | Yan *et al.* (2004) |
| *VRN3* | Vernalization gene, orthologue of *Arabidopsis FT* gene | 7BS | Yan *et al.* (2006) |
| *VRN4* | Vernalization gene | 5DS | Kippes *et al.* (2015) |
| *Q* | Gene conferring the free-threshing character and a square-headed phenotype of spikes | 5AL | Faris *et al.* (2003) |
| *Ph1* | Gene controlling correct pairing of homologous chromosomes | 5BL | Griffiths *et al.* (2006) |
| *Gpc-B1* | NAC transcription factor controlling senescence and grain protein, zinc and iron content | 6BS | Uauy *et al.* (2006) |

## 2.3.2  Resistance to Russian wheat aphid (*Diuraphis noxia*)

Wheat chromosome arm 7DS, which represents the object of this thesis, carries several agronomically important genes, including those conferring resistance to Russian wheat aphid.

Russian wheat aphid (RWA; *Diuraphis noxia*, Kurdjumov) is taxonomically classified in *Homoptera* order, family *Aphididae*. It is a serious invasive pest of

wheat and barley indigenous to Afghanistan, Iran, southern Russia, and countries bordering the Mediterranean Sea. Nowadays this pest is dispersed globally in all important wheat and barley growing areas except Australia (Ennahli *et al.* 2009) causing annually considerable economical losses (Valdez *et al.* 2012; Fazel-Najafabadi *et al.* 2015). The use of resistant varieties seems to be the most effective, economical and environmental safe means of controlling this pest (Liu *et al.* 2002; Fazel-Najafabadi *et al.* 2015).

Aphids belong to the major arthropod pests of agricultural crops (Smith *et al.* 2010). This probing-sucking insect inserts its stylets into the phloem sieve elements and feeds on the phloem sap while injecting salivary enzymes which causes plant damage. Aphids weaken the plants by removing photoassimilates and they can also be vectors of numerous devastating plant viruses (Botha *et al.* 2006; Smith *et al.* 2010). The infestation of the plant by Russian wheat aphid causes reduction of plant height, shoot weight, number of spikes, and yield (Girma *et al.* 1993). Symptoms like chlorosis, streaking along the entire leaf blade, leaf rolling, head trapping, and in severe cases, plant death are observed in susceptible plants. RWA feeding on susceptible plants results in lower levels of chlorophyll *a*, *b*, and carotenoids in damaged regions which subsequently leads to less efficient use of light energy and thus reduction in photosynthetic rates (Franzen *et al.* 2008). Rolled leafs sheltering aphid colonies provide the protection against predators and significantly hinder the effectiveness of applied insecticides.

Three types of resistance to aphids were described in plants – antixenosis, antibiosis and tolerance. Antixenosis (nonpreferance) comprises morphological or chemical plant factors that discourage the aphid of ingestion of plant fluids or oviposition. As a result, aphid selects an alternate host plant. Same like antixenosis, the antibiosis includes mechanical and chemical defensive factors. However, those factors have a negative effect on the aphid biology, causing moderate to lethal damage. Surviving individuals suffer from developmental defects and reduced fecundity. Antibiosis and antixenosis categories of resistance are commonly present together in plants and it is difficult to delineate between them. The last type of resistance is tolerance, which is described as a capability of plant to outgrow aphid infestation or to recover after the destruction or removal of damaged tissues. Tolerance often occurs in combination with antibiosis and antixenosis (Smith *et al.* 2005).

### 2.3.2.1 Mapping RWA resistance genes

Several RWA resistance genes have been identified in various cultivars and lines of wheat, rye and *Aegilops tauschii*, as described in the Table 3. All those genes with the exception of recessive *dn3* are major dominant genes. Besides these major genes, several QTLs were identified (Peng *et al.* 2009; Castro *et al.* 2004). The resistance mediated by various *Dn* genes may exhibit a combination of all three categories of resistance - antixenosis, antibiosis, and tolerance. Despite the fact that the effects of several RWA resistance genes are known, none have been isolated yet (Botha *et al.* 2014).

Seven out of the 12 genes given in the Table 3 were localized on the chromosome arm 7DS in different wheat cultivars or lines. With the exception of *Dn8*, all of them (*Dn1*, *Dn2*, *Dn5*, *Dn6*, *Dnx, Dn2401 and Dn626580*) were mapped to the central part of the chromosome arm 7DS. Marker-linkage analyses showed that those genes cluster in one region and are linked to genetic marker *Xgwm111* (Liu *et al.* 2005; Fazel-Najfabadi *et al.* 2015). Deletion bin mapping assigned marker *Xgwm111* to a central bin which is delimited as interval between 36 – 61 % of the chromosome arm length. Allelism tests revealed that genes clustering on the 7DS are either alleles of a single locus or are closely related members of a *Dn* gene family (Liu *et al.* 2005). Gene *Dn8* was genetically mapped to the distal end of the 7DS (Liu *et al.* 2001).

Table 3: Overview of genes underlying resistance to RWA

| Gene | Resistant line/cultivar | Chromosome location | Plant species | Reference |
|---|---|---|---|---|
| *Dn1* | PI 137739 | 7DS | wheat | Du Toit (1987) |
| *Dn2* | PI 262660 | 7DS | wheat | Du Toit (1987) |
| *dn3* | CO 03810 | unknown | *Aegilops tauschii* | Nkongolo *et al.* (1991) |
| *Dn4* | PI 372129 | 1DS | wheat | Nkongolo *et al.* (1991) |
| *Dn5* | PI 294994 | 7DS | wheat | Du Toit (1987) |
| *Dn6* | PI 243781 CI 6501 | 7DS | wheat | Nkongolo *et al.* (1991) |
| *Dn7* | 94M370 | 1RS.1BL translocation | rye/wheat | Marais *et al.* (1994) |
| *Dn8* | PI 294994 | 7DS | wheat | Liu *et al.* (2001) |
| *Dn9* | PI 294994 | 1DL | wheat | Liu *et al.* (2001) |
| *Dnx* | PI 220127 | 7DS | wheat | Harvey and Martin (1990) |
| *Dny* | PI 220350 | unknown | wheat | Martin *et al.* (2001) |
| *Dn2401* | CI2401 | 7DS | wheat | Voothuluru *et al.* (2006) |
| *Dn2414* | line 2414-11 | 1RS.1BL translocation | rye/wheat | Peng *et al.* (2007) |
| *Dn626580* | PI626580 | 7DS | wheat | Valdez *et al.* (2012) |

## 2.4 References

AGI - Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408:796–815

Akhunov ED, Akhunova AR, Dvorak J (2005) BAC libraries of *Triticum urartu*, *Aegilops speltoides* and *Ae. tauschii*, the diploid ancestors of polyploid wheat. Theor Appl Genet 111:1617–1622

Akpinar BA, Magni F, Yuce M, Stuart J, Lucas SJ, Šimková H, Šafář J, Vautrin S, Bergès H, Cattonaro F, Doležel J, Budak H (2015) The physical map of wheat chromosome 5DS revealed gene duplications and small rearrangements. BMC Genomics 16:453

Alkan C, Sajjadian S, Eichler EE (2011) Limitations of next-generation genome sequence assembly. Nat Methods 8:61–65

Allouis S, Moore G, Bellec A, Sharp R, Faivre Rampant P, Mortimer K, Pateyron S, Foote TN, Griffiths S, Caboche M, Chalhoub B (2003) Construction and characterisation of a hexaploid wheat (*Triticum aestivum* L) BAC library from the reference germplasm 'Chinese Spring'. Cereal Res Commun 31: 331-338

Barabaschi D, Magni F, Volante A, Gadaleta A, Šimková H, Scalabrin S, Maria Prazzoli MC, Bagnaresi P, Lacrima K, Michelotti V, Desiderio F, Orrù L, Mazzamurro V, Fricano A, MastrangeloxAM, Tononi P, Vitulo V, Jurman I, Frenkel Z, Cattonaro F, Morgante M, Blanco A, Doležel J, Delledonne M, Stanca AM, Cattivelli L, Valè G (accepted paper) Physical mapping of bread wheat chromosome 5A: An integrated approach. Plant Genome doi:103835/plantgenome2015030011

Barillot E, Lacroix B, Cohen D (1991) Theoretical analysis of library screening using a N-dimensional pooling strategy. Nucleic Acids Res 19:6241–6247

Bennett MD, Smith J (1976) Nuclear DNA amounts in angiosperms. Phil Trans R Soc B 274:227-274

Bennett ST, Barnes C, Cox A, Davies L, Brown C (2005) Toward the $1000 human genome. Pharmacogenomics 6:373–382

Berkman PJ, Visendi P, Lee HC, Stiller J, Manoli S, Lorenc MT, Lai K, Batley J, Fleury D, Šimková H, Kubaláková M, Weining S, Doležel J, Edwards D (2013) Dispersion and domestication shaped the genome of bread wheat. Plant Biotech J 11:564–571

Birol I, Raymond A, Jackman SD, Pleasance S, Coope R, Taylor GA, Yuen MM, Keeling CI, Brand D, Vandervalk BP, Kirk H, Pandoh P, Moore RA, Zhao Y, Mungall AJ, Jaquish B, Yanchuk A, Ritland C, Boyle B, Bousquet J, Ritland K, Mackay J, Bohlmann J, Jones SJ (2013) Assembling the 20 Gb white spruce

(*Picea glauca*) genome from whole-genome shotgun sequencing data. Bioinformatics 29:1492–1497

Botha AM, Lacock L, van Niekerk C, Matsioloko MT, du Preez FB, Loots S, Venter E, Kunert KJ, Cullis CA (2006) Is photosynthetic transcriptional regulation in *Triticum aestivum* L cv 'TugelaDN' a contributing factor for tolerance to *Diuraphis noxia* (Homoptera: Aphididae)? Plant Cell Rep 25:41-54

Botha AM, van Eck L, Burger NF, Swanevelder ZH (2014) Near-isogenic lines of *Triticum aestivum* with distinct modes of resistance exhibit dissimilar transcriptional regulation during *Diuraphis noxia* feeding. Biol Open 3:1116-1126

Breen J, Wicker T, Shatalina M, Frenkel Z, Bertin I, Philippe R, Spielmeyer W, Šimková H,ŠSafář J, Cattonaro F, Scalabrin S, Magni F, Vautrin S, Bergès H, International Wheat Genome Sequencing Consortium, Paux E, Fahima T, Doležel J, Korol A, Feuillet C, Keller B (2013) A physical map of the short arm of wheat chromosome 1A. PLoS ONE 8: e80272

Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705-710

Bruno WJ, Knill E, Balding DJ, Bruce DC, Doggett NA, Sawhill WW, Stallings RL, Whittaker CC, Torney DC (1995) Efficient pooling designs for library screening. Genomics 26:21–30

Cao H, Hastie AR, Cao D, Lam ET, Sun Y, Huang H, Liu X, Lin L, Andrews W, Chan S, Huang S, Tong X, Requa M, Anantharaman T, Krogh A, Yang H, Cao H, Xu X (2014) Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. Gigascience 3:34

Castro AM, Vasicek A, Ellerbrook C, Giménez DO,Tocho E, Tacaliti MS, Clúa A, Snape JW (2004) Mapping quantitative trait loci in wheat for resistance against greenbug and Russian wheat aphid. Plant Breeding 123:361-365

Cenci A, Chantret N, Kong X, Gu Y, Anderson OD, Fahima T, Distelfeld A, Dubcovsky J (2003) Construction and characterization of a half million clone BAC library of durum wheat (*Triticum turgidum* ssp. *durum*). Theor Appl Genet 107:931–939

Chantret N, Salse J, Sabot F, Rahman S, Bellec A, Laubin B, Dubois I, Dossat C, Sourdille P, Joudrier P, Gautier MF, Cattolico L, Beckert M, Aubourg S, Weissenbach J, Caboche M, Bernard M, Leroy P, Chalhoub B (2005) Molecular

basis of evolutionary events that shaped the hardness locus in diploid and polyploid wheat species (*Triticum* and *Aegilops*). Plant Cell 17:1033–1045

Chapman JA, Mascher M, Buluç A, Barry K, Georganas E, Session A, Strnadova V, Jenkins J, Sehgal S, Oliker L, Schmutz J, Yelick KA, Scholz U, Waugh R, Poland JA, Muehlbauer GJ, Stein N, Rokhsar DS (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. Genome Biol 16:26

Chaudhary B, Flagel L, Stupar RM, Udall JA, Verma N, Springer NM, Wendel JF (2009) Reciprocal silencing, transcriptional bias and functional divergence of homeologs in polyploid cotton (*Gossypium*). Genetics 182:503–517

Chen FG, Zhang XY, Xia GM, Jia JZ (2002) Construction and characterization of a bacterial artificial chromosome library for *Triticum boeoticum*. Acta Bot Sin 44:451–456

Choulet F,Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier MC, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphrey M, Liu S, Kong X, Jia J, Gut M, Brunel D, Anderson JA, Gill BS, Appels R, Keller B, Feuillet C (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell 22:1686–1701

Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Le Gouis J, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Šafář J, Doležel J, Rogers J, Vandepoele K, Aury JM, Mayer K, Bergès H, Quesneville H, Wincker P, Feuillet C (2014a) Structural and functional partitioning of bread wheat chromosome 3B. Science 345:1249721

Choulet F, Caccamo M, Wright J, Alaux M, Šimková H, Šafář J, Leroy P, Doležel J, Rogers J, Eversole K, Feuillet C (2014b) The Wheat black jack: Advances towards sequencing the 21 chromosomes of bread wheat. In: Tuberosa R, Graner A, Frison E (eds) Genomics of plant genetic resources. Volume 1. Managing, sequencing and mining genetic resources. pp 405-438 Springer Science+Business Media Dordrecht

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barely. BMC Genomics 10:582

Cloutier S, McCallum BD, Loutre C, Banks TW, Wicker T, Feuillet C, Keller B, Jordan MC (2007) Leaf rust resistance gene *Lr1*, isolated from bread wheat

(*Triticum aestivum* L) is a member of the large *psr567* gene family. Plant Mol Biol 65:93-106

Devos KM, Gale MD (2000) Genome relationships: The grass model in current research. Plant Cell 12:637–646

Devos KM, Ma J, Pontaroli AC, Pratt LH, Bennetzen JL (2005) Analysis and mapping of randomly chosen bacterial artificial chromosome clones from hexaploid bread wheat. Proc Natl Acad Sci USA 102:19243–19248

Doebley J, Stec A, Hubbard L (1997) The evolution of apical dominance in maize. Nature 386:485–488

Doležel J, Číhalíková J, Lucretti S (1992) A high-yield procedure for isolation of metaphase chromosomes from root tips of *Vicia faba* L. Planta 188:93–98

Doležel J, Kubaláková M, Bartoš J, Macas J (2004) Flow cytogenetics and plant genome mapping. Chromosome Res 12:77–91

Doležel J, Kubaláková M, Paux E, Bartoš J, Feuillet C (2007) Chromosome-based genomics in the cereals. Chromosome Res 15:51-66

Doležel J, Vrána J, Cápal P, Kubaláková M, Burešová V, Šimková H (2014) Advances in plant chromosome genomics. Biotechnol Adv 32:122-136

Du Toit F (1987) Resistance in wheat (*Triticum aestivum*) to *Diuraphis noxia* (Homoptera: Aphididae). Cereal Res Commun 15:175–179

Dubcovsky J, Dvorak J (2007) Genome plasticity a key factor in the success of polyploid wheat under domestication. Science 316:1862–1866

Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, deWinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, Zhong F, Korlach J, Turner S (2009) Real-time DNA sequencing from single polymerase molecules. Science 323:133–138

Ennahli S, El Bouhssini M, Grando S, Anathakrishnan R, Niide T, Starkus L, Starkey S, Smith CM (2009): Comparison of categories of resistance in wheat and barley genotypes against biotype 2 of the Russian wheat aphid, *Diuraphis noxia* (Kurdjumov). Arthropod-plant Inte 3:45-53

Erayman M, Sandhu D, Sidhu D, Dilbirligi M, Baenziger PS, Gill KS (2004) Demarcating the gene-rich regions of the wheat genome. Nucleic Acids Res 32:3546–3565

Eversole K, Feuillet C, Mayer KF, Rogers J (2014) Slicing the wheat genome. Introduction. Science 345:285-287

Faris JD, Zhang Z, Fellers JP, Gill BS (2003) Micro-colinearity between rice, *Brachypodium*, and *Triticum monococcum* at the wheat domestication locus *Q*. Funct Integr Genomics 8:149-164

Faris JD, Zhang Z, Lu H, Lu S, Reddy L, Cloutier S, Fellers JP, Meinhardt SW, Rasmussen JB, Xu SS, Oliver RP, Simons KJ, Friesen TL (2010) A unique wheat disease resistance-like gene governs effector-triggered susceptibility to necrotrophic pathogens. Proc Natl Acad Sci USA 107:13544–13549

Fazel-Najafabadi M, Peng J, Peairs FB, Šimková H, Kilian A, Lapitan NLV (2015) Genetic mapping of resistance to *Diuraphis noxia* (Kurdjumov) biotype 2 in wheat (*Triticum aestivum* L) accession CI2401. Euphytica 203:607–614

Febrer M, Goicoechea JL, Wright J, McKenzie N, Song X, Lin J, Collura K, Wissotski M, Yu Y, Ammiraju JS, Wolny E, Idziak D, Betekhtin A, Kudrna D, Hasterok R, Wing RA, Bevan MW (2010) An integrated physical, genetic and cytogenetic map of *Brachypodium distachyon*, a model system for grass research. PLoS One 5:e13461

Feldman M, Lupton FGH, Miller TE (1995) Wheat. In: Smartt J, Simmonds NW (eds) Evolution of crop plants. 2nd ed. pp 184–192 Longman Scientific and Technical New York

Feldman M, Levy AA (2005) Allopolyploidy – a shaping force in the evolution of wheat genomes. Cytogenet Genome Res 109: 250–258

Feldman M, LevyAA (2009) Genome evolution in allopolyploid wheat – a revolutionary reprogramming followed by gradual changes. J Genet Genomics 36:511–518

Feuillet C, Travella S, Stein N, Albar L, Nublat A, Keller B (2003) Map-based isolation of the leaf rust disease resistance gene *Lr10* from the hexaploid wheat (*Triticum aestivum* L) genome. Proc Natl Acad Sci USA 100:15253–15258

Feuillet C, Langridge P, Waugh R (2008) Cereal breeding takes a walk on the wild side. Trends Genet 24:24-32

Feuillet C, Leach JE, Rogers J, Schnable PS, Eversole K (2011) Crop genome sequencing: lessons and rationales. Trends Plant Sci 16:77–88

Feuillet C, Stein N, Rossini L, Praud S, Mayer K, Schulman A, Eversole K, Appels R (2012) Integrating cereal genomics to support innovation in the *Triticeae*. Funct Integr Genomics 12:573-583

Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, Temsch EM, Greilhuber J, Müller KF, Heubl G (2014) Evolution of genome size and

chromosome number in the carnivorous plant genus *Genlisea* (Lentibulariaceae), with a new estimate of the minimum genome size in angiosperms. Ann Bot 114:1651-1663

Franzen LD, Gutsche AR, Heng-Moss TM, Higley LG, Macedo TB (2008): Physiological responses of wheat and barley to Russian wheat aphid, *Diuraphis noxia* (Mordvilko) and bird cherry-oat aphid, *Rhopalosiphum padi* (L) (Hemiptera: Aphididae). Arthropod-plant Inte 2:227-235

Frenkel Z, Paux E, Mester D, Feuillet C, Korol A (2010) LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. BMC Genomics 11:584

Fu D, Uauy C, Distelfeld A, Blechl A, Epstein L, Chen X, Sela H, Fahima T, Dubcovsky J (2009) A kinase-START gene confers temperature-dependent resistance to wheat stripe rust. Science 323:1357–1360

Gale MD, Devos KM (1998a) Plant comparative genetics after 10 years. Science 282:656-659

Gale MD, Devos KM (1998b) Comparative genetics in the grasses. Proc Natl Acad Sci USA 95:1971–1974

Ganapathy G, Howard JT, Ward JM, Li J, Li B, Li Y, Xiong Y, Zhang Y, Zhou S, Schwartz DC, Schatz M, Aboukhalil R, Fedrigo O, Bukovnik L, Wang T, Wray G, Rasolonjatovo I, Winer R, Knight JR, Koren S, Warren WC, Zhang G, Phillippy AM, Jarvis ED (2014) High-coverage sequencing and annotated assemblies of the budgerigar genome. Gigasciensce 3:11

Gardiner J, Schroeder S, Polacco ML et al, (2004) Anchoring 9,371 maize expressed sequence tagged unigenes to the bacterial artificial chromosome contig map by two-dimensional overgo hybridization. Plant Physiol 134:1317–1326

Gill KS, Gill BS, Endo TR, Boyko EV (1996a) Identification and high-density mapping of gene-rich regions in chromosome group 5 of wheat. Genetics 143:1001–1012

Gill KS, Gill BS, Endo TR, Taylor T (1996b) Identification and high-density mapping of gene-rich regions in chromosome group 1 of wheat. Genetics 144:1883–1891

Giorgi D, Farina A, Grosso V, Gennaro A, Ceoloni C, Lucretti S (2013) FISHIS: fluorescence *in situ* hybridization in suspension and chromosome flow sorting made easy. PLoS One 8:e57994

Girma M, Wilde GE, Harvey TL (1993) Russian wheat aphid (Homoptera: Aphididae) affects yield and quality of wheat. J Econ Entomol 86:594–601

Green ED (2001) Strategies for the systematic sequencing of complex genomes. Nat Rev Genet 2:573-583

Griffiths S, Sharp R, Foote TN, Bertin I, Wanous M, Reader S, Colas I, Moore G (2006) Molecular characterization of *Ph1* as a major chromosome pairing locus in polyploid wheat. Nature 439:749-752

Gupta PH, Rustgi S, Mir RR (2013a) Array-based high-throughput DNA markers and genotyping platforms for cereal genetics and genomics. In: Gupta PK, Varshney RK (eds) Cereal Genomics II. pp 11-55 Springer Science+Business Media Dordrecht

Gupta B, Saha J, Sengupta A, Gupta K (2013b) Recent advances on Virus induced gene silencing (VIGS): Plant functional genomics. J Plant Biochem Physiol 1:e116

Harvey TL, Martin TJ (1990) Resistance to Russian wheat aphid, *Diuraphis noxia*, in wheat (*Triticum aestivum*). Cereal Res Commun 18:127-129

Hastie AR, Dong L, Smith A, Finklestein J, Lam ET, Huo N, Cao H, Kwok PY, Deal KR, Dvorak J, Luo MC, Gu Y, Xiao M (2013) Rapid genome mapping in nanochannel arrays for highly complete and accurate *de novo* sequence assembly of the complex *Aegilops tauschii* genome. PLoS One 8:e55864

Hernandez D, Francois P, Farinelli L, Osteras M, Schrenzel J (2008) *De novo* bacterial genome sequencing: Millions of very short reads assembled on a desktop computer. Genome Res 18:802–809

Hernandez P, Martis M, Dorado G, Pfeifer M, Gálvez S, Schaaf S, Jouve N, Šimková H, Valárik M, Doležel J, Mayer KF (2012) Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. Plant J 69:377-386

Huang S, Sirikhachornkit A, Su X, Faris J, Gill B, Haselkorn R, Gornicki P (2002) Genes encoding plastid acetyl-CoA carboxylase and 3-phosphoglycerate kinase of the *Triticum*/*Aegilops* complex and the evolutionary history of polyploid wheat. Proc Natl Acad Sci USA 99:8133–8138

Huang L, Brooks SA, Li WL, Fellers JP, Trick HN, Gill BS (2003) Map-based cloning of leaf rust resistance gene *Lr21* from the large and polyploid genome of bread wheat. Genetics 164:655–664

IBI - International Brachypodium Initiative (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768

IRGSP - International Rice Genome Sequencing Project (2005) The map-based sequence of the rice genome. Nature 436:793–800

Islam-Faridi MN, Nelson CD, DiFazio SP, Gunter LE, Tuskan GA (2009) Cytogenetic analysis of *Populus trichocarpa* – ribosomal DNA, telomere repeat sequence, and marker-selected BACs. Cytogenet Genome Res 125:74–80

IWGSC - International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788

Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, Choisne N, Aubourg S, Vitulo N, Jubin C, Vezzi A, Legeai F, Hugueney P, Dasilva C, Horner D, Mica E, Jublot D, Poulain J, Bruyère C, Billault A, Segurens B, Gouyvenoux M, Ugarte E, Cattonaro F, Anthouard V, Vico V, Del Fabbro C, Alaux M, Di Gaspero G, Dumas V, Felice N, Paillard S, Juman I, Moroldo M, Scalabrin S, Canaguier A, Le Clainche I, Malacrida G, Durand E, Pesole G, Laucou V, Chatelet P, Merdinoglu D, Delledonne M, Pezzotti M, Lecharny A, Scarpelli C, Artiguenave F, Pè ME, Valle G, Morgante M, Caboche M, Adam-Blondon AF, Weissenbach J, Quétier F, Wincker P, French-Italian Public Consortium for Grapevine Genome Characterization (2007) The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 449:463–467

Jain AK, Dubes RC (1988) Algorithms for clustering data. Prentice Hall Englewood Cliffs New Jersey

Janda J, Bartoš J, Šafář J, Kubaláková M, Valárik M, Číhalíková J, Šimková H, Caboche M, Sourdille P, Bernard M, Chalhoub B, Doležel J (2004) Construction of a subgenomic BAC library specific for chromosomes 1D, 4D and 6D of hexaploid wheat. Theor Appl Genet 109:1337-1345

Janda J, Šafář J, Kubaláková M, Bartoš J, Kovářová P, Suchánková P, Pateyron S, Číhalíková J, Sourdille P, Šimková H, Faivre-Rampant P, Hřibová E, Bernard M, Lukaszewski A, Doležel J, Chalhoub B (2006) Novel resources for wheat genomics: BAC library specific for the short arm of chromosome 1B. Plant J 47:977-986

Jia J, Zhao S, Kong X, Li Y, Zhao G, He W, Appels R, Pfeifer M, Tao Y, Zhang X, Jing R, Zhang C, Ma Y, Gao L, Gao C, Spannagl M, Mayer KF, Li D, Pan S, Zheng F, Hu Q, Xia X, Li J, Liang Q, Chen J, Wicker T, Gou C, Kuang H, He G, Luo Y, Keller B, Xia Q, Lu P, Wang J, Zou H, Zhang R, Xu J, Gao J, Middleton C, Quan Z, Liu G, Wang J, International Wheat Genome Sequencing Consortium, Yang H, Liu X, He Z, Mao L, Wang J (2013) *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496:91-95

Kippes N, Debernardi JM, Vasquez-Gross HA, Akpinar BA, Budak H, Kato K, Chao S, Akhunov E, Dubcovsky J (2015) Identification of the *VERNALIZATION 4*

gene reveals the origin of spring growth habit in ancient wheats from South Asia. Proc Natl Acad Sci USA E5401-E5410

Kobayashi F, Wu J, Kanamori H, Tanaka T, Katagiri S, Karasawa W, Kaneko S, Watanabe S, Sakaguchi T, Hanawa Y, Fujisawa H, Kurita K, Abe C, Iehisa JC, Ohno R, Šafář J, Šimková H, Mukai Y, Hamada M, Saito M, Ishikawa G, Katayose Y, Endo TR, Takumi S, Nakamura T, Sato K, Ogihara Y, Hayakawa K, Doležel J, Nasuda S, Matsumoto T, Handa H (2015) A high-resolution physical map integrating an anchored chromosome with the BAC physical maps of wheat chromosome 6B. BMC Genomics 16:595

Kota R, Spielmeyer W, McIntosh RA, Lagudah ES (2006) Fine genetic mapping fails to dissociate durable stem rust resistance gene *Sr2* from pseudo-black chaff in common wheat (*Triticum aestivum* L). Theor Appl Genet. 112:492-499

Krattinger S, Wicker T, Keller B (2009a) Map-based cloning of genes in *Triticeae* (wheat and barley). In: Feuillet C, Muehlbauer GJ (eds) Genetics and genomics of the Triticeae. pp 337–358 Springer New York

Krattinger SG, Lagudah ES, Spielmeyer W, Singh RP, Huerta-Espino J, McFadden H, Bossolini E, Selter LL, Keller B (2009b) A putative ABC transporter confers durable resistance to multiple fungal pathogens in wheat. Science 323:1360–1363

Kumar A, Simons K, Iqbal MJ, Michalak de Jimenéz M, Bassi FM, Ghavami F, Al-Azzam O, Drader T, Wanf Y, Luo MC, Gu YQ, Denton A, Lazo GR, Xu SS, Dvorak J, Kianian PM, Kianian SF (2012) Physical mapping resources for large plant genomes: radiation hybrids for wheat D-genome progenitor *Aegilops tauschii*. BMC Genomics 13:597

Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, Deshpande P, Cao H, Nagarajan N, Xiao M, Kwok PY (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 30:771–776

Lamoureux D, Peterson DG, Li W, Fellers JP, Gill BS (2005) The efficacy of Cot-based gene enrichment in wheat (*Triticum aestivum* L). Genome 48:1120–1126

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J (2009) SOAP2: an improved ultrafast tool for short read alignment. Bioinformatics 25:1966–1967

Li LF, Liu B, Olsen KM, Wendel JF (2015) A re-evaluation of the homoploid hybrid origin of *Aegilops tauschii*, the donor of the wheat D-subgenome. New Phytol 208:4-8

Lijavetzky D, Muzzi G, Wicker T, Keller B, Wing R, Dubcovsky J (1999) Construction and characterization of a bacterial artificial chromosome (BAC) library for the A genome of wheat. Genome 42:1176–1182

Ling P, Chen XM (2005) Construction of a hexaploid wheat (*Triticum aestivum* L) bacterial artificial chromosome library for cloning genes for stripe rust resistance. Genome 48:1028–1036

Ling HQ, Zhao S, Liu D, Wang J, Sun H, Zhang C, Fan H, Li D, Dong L, Tao Y, Gao C, Wu H, Li Y, Cui Y, Guo X, Zheng S, Wang B, Yu K, Liang Q, Yang W, Lou X, Chen J, Feng M, Jian J, Zhang X, Luo G, Jiang Y, Liu J, Wang Z, Sha Y, Zhang B, Wu H, Tang D, Shen Q, Xue P, Zou S, Wang X, Liu X, Wang F, Yang Y, An X, Dong Z, Zhang K, Zhang X, Luo MC, Dvorak J, Tong Y, Wang J, Yang H, Li Z, Wang D, Zhang A, Wang J (2013) Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496:87-90

Liu XM, Smith CM, Gill BS, Tolmay V (2001) Microsatellite markers linked to six Russian wheat aphid resistance genes in wheat. Theor Appl Genet 102:504–510

Liu XM, Smith CM, Gill BS (2002) Identification of microsatellite markers linked to Russian wheat aphid resistance genes *Dn4* and *Dn6*. Theor Appl Genet 104:1042-1048

Liu XM, C M Smith CM, Friebeb BR, Gill BS (2005) Molecular mapping and allelic relationships of Russian wheat aphid–resistance genes. Crop Sci 45:2273–2280

Lucas SJ, Akpınar BA, Kantar M, Weinstein Z, Aydınoğlu F, Šafář J, Šimková H, Frenkel Z, Korol A, Magni F, Cattonaro F, Vautrin S, Bellec A, Bergès H, Doležel J, Budak H (2013) Physical mapping integrated with syntenic analysis to characterize the gene space of the long arm of wheat chromosome 1A. PLoS ONE 8:e59542

Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378-389

Luo MC, Xu K, Ma Y, Deal KR, Nicolet CM, and Dvorak J (2009) A high-throughput strategy for screening of bacterial artificial chromosome libraries and anchoring of clones on a genetic map constructed with single nucleotide polymorphisms. BMC Genomics 10:28

Luo MC, Gu YQ, You FM, Deal KR, Ma Y, Hu Y, Huo N, Wang Y, Wang J, Chen S, Jorgensen CM, Zhang Y, McGuire PE, Pasternak S, Stein JC, Ware D, Kramer M, McCombie WR, Kianian SF, Martis MM, Mayer KF, Sehgal SK, Li W, Gill BS, Bevan MW, Šimková H, Doležel J, Weining S, Lazo GR, Anderson OD, Dvorak J (2013) A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. Proc Natl Acad Sci USA 110:7940-7945

Marais GF, Horn M, Du Toit F (1994) Intergeneric transfer (rye to wheat) of a gene(s) for Russian wheat aphid resistance. Plant Breeding 113:65–271

Marcussen T, Sandve SR, Heier L, Spannagl M, Pfeifer M, International Wheat Genome Sequencing Consortium, Jakobsen KS, Wulff BB, Steuernagel B, Mayer KF, Olsen OA (2014) Ancient hybridizations among the ancestral genomes of bread wheat. Science 345:1250092

Mardis ER (2008) Next-generation DNA sequencing methods. Annu Rev Genomics Hum Genet 9:387–402

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z, Dewell SB, Du L, Fierro JM, Gomes XV, Godwin BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer ML, Jarvie TP, Jirage KB, Kim JB, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, Rothberg JM (2005) Genome sequencing in microfabricated highdensity picolitre reactors. Nature 437:376–380

Martin TJ, Sears RG, Seifers DL, Harvey TL, Witt MD, Schlegel AJ, McCluskey PJ, Hatchett JH (2001) Registration of Trego wheat. Crop Sci 41:929-930

Martinez-Perez E, Shaw P, Moore G (2001) The *Ph1* locus is needed to ensure specific somatic and meiotic centromere association. Nature 411:204–207

Matsunaga S, Kawano S, Michimoto T, Higashiyama T, Nakao S, Sakai A, et al (1999) Semi-automatic laser beam microdissection of the Y chromosome and analysis of Y chromosome DNA in a dioecious plant, *Silene latifolia*. Plant Cell Physiol 40:60–68

Mayer KF, Taudien S,MartisM, Šimková H, Suchánková P, Gundlach H, Wicker T, Petzold A, FelderM, Steuernagel B, Scholz U, Graner A, Platzer M, Doležel J, Stein N (2009) Gene content and virtual gene order of barley chromosome 1H. Plant Physiol 151:496–505

Mayer KF, Martis M, Hedley PE, Šimková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Doležel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

Metzker ML (2010) Sequencing technologies – the next generation. Nat Rev Genet 11:31–46

Meyers BC, Scalabrin S, Morgante M (2004) Mapping and sequencing complex genomes: Let's get physical! Genetics 5:578-589

Mikheyev AS, Tin MM (2014) A first look at the Oxford Nanopore MinION sequencer. Mol Ecol Resour 14:1097–1102

Miller JR, Koren S, Sutton G (2010) Assembly algorithms for next-generation sequencing data. Genomics 95:315-327

Murat F, Xu JH, Tannier E, Abrouk M, Guilhot N, Pont C, Messing J, Salse J (2010) Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. Genome Res 20:1545–1557

Nilmalgoda SD, Cloutier S, Walichnowski AZ (2003) Construction and characterization of a bacterial artificial chromosome (BAC) library of hexaploid wheat (*Triticum aestivum* L) and validation of genome coverage using locus-specific primers. Genome 46:870–878

Nkongolo KK, Quick JS,Limin E, Fowler DB (1991) Sources and inheritance of resistance to Russian wheat aphid in *Triticum* species amphiploids and *Triticum tauschii*. Can J Plant Sci 71:703-708

Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, Vezzi F, Delhomme N, Giacomello S, Alexeyenko A, Vicedomini R, Sahlin K, Sherwood E, Elfstrand M, Gramzow L, Holmberg K, Hällman J, Keech O, Klasson L, Koriabine M, Kucukoglu M, Käller M, Luthman J, Lysholm F, Niittylä T, Olson A, Rilakovic N, Ritland C, Rosselló JA, Sena J, Svensson T, Talavera-López C, Theißen G, Tuominen H, Vanneste K, Wu ZQ, Zhang B, Zerbe P, Arvestad L, Bhalerao R, Bohlmann J, Bousquet J, Garcia Gil R, Hvidsten TR, de Jong P, MacKay J, Morgante M, Ritland K, Sundberg B, Thompson SL, Van de Peer Y, Andersson B, Nilsson O, Ingvarsson PK, Lundeberg J, Jansson S (2013) The Norway spruce genome sequence and conifer genome evolution. Nature 497:579–584

Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, Haberer G, Hellsten U, Mitros T, Poliakov A, Schmutz J, Spannagl M, Tang H, Wang X, Wicker T, Bharti AK, Chapman J, Feltus FA, Gowik U, Grigoriev IV, Lyons E, Maher CA, Martis M, Narechania A, Otillar RP, Penning BW, Salamov AA, Wang Y, Zhang L, Carpita NC, Freeling M, Gingle AR, Hash CT, Keller B, Klein P, Kresovich S, McCann MC, Ming R, Peterson DG, Mehboob-ur-Rahman, Ware D, Westhoff P, Mayer KF, Messing J, Rokhsar DS (2009) The *Sorghum bicolor* genome and the diversification of grasses. Nature 457:551–556

Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, Feuillet C (2006) Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. Plant J 48:463–474

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, Lagudah E, Somers D, Kilian A, Alaux M, Vautrin S,

Bergès H, Eversole K, Appels R, Šafář J, Šimková H, Doležel J, Bernard M, Feuillet C (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322:101-104

Pellicer J, Fay MF, Leitch IJ (2010) The largest eukaryotic genome of them all? Botanical Journal of the Linnean Society 164:10-15

Pendleton M, Sebra R, Pang AW, Ummat A, Franzen O, Rausch T, Stütz AM, Stedman W, Anantharaman T, Hastie A, Dai H, Fritz MH, Cao H, Cohain A, Deikus G, Durrett RE, Blanchard SC, Altman R, Chin CS, Guo Y, Paxinos EE, Korbel JO, Darnell RB, McCombie WR, Kwok PY, Mason CE, Schadt EE, Bashir A (2015) Assembly and diploid architecture of an individual human genome via single-molecule technologies. Nat Methods 12:780-786

Peng JH, Wang H, Haley SD, Peairs FB, Lapitan NLV (2007) Molecular mapping of the Russian wheat aphid resistance gene *Dn2414* in wheat. Crop Sci 47:2418–2429

Peng JH ,Bai Y, Haley SD, Lapitan NLV (2009) Microsatellite-based molecular diversity of bread wheat germplasm and association mapping of wheat resistance to the Russian wheat aphid. Genetica 135:95-122

Periyannan S, Moore J, Ayliffe M, Bansal U, Wang X, Huang L, Deal K, Luo M, Kong X, Bariana H, Mago R, McIntosh R, Dodds P, Dvorak J, Lagudah E (2013) The gene *Sr33*, an ortholog of barley *Mla* genes, encodes resistance to wheat stem rust race Ug99. Science 341:786–788

Petersen G, Seberg O, Yde M, Berthelsen K (2006) Phylogenetic relationships of *Triticum* and *Aegilops* and evidence for the origin of the A, B, and D genomes of common wheat (*Triticum aestivum*). Mol Phylogenet Evol 39:70–82

Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, International Wheat Genome Sequencing Consortium, Mayer KF, Olsen OA (2014) Genome interplay in the grain transcriptome of hexaploid bread wheat. Science 345:1250091

Philippe R, Choulet F, Paux E, van Oeveren J, Tang J, Wittenberg AH, Janssen A, van Eijk MJ, Stormo K, Alberti A,Wincker P, Akhunov E, van der Vossen E, Feuillet C (2012) Whole Genome Profiling provides a robust framework for physical mapping and sequencing in the highly complex and repetitive wheat genome. BMC Genom 13:47

Philippe R, Paux E, Bertin I, Sourdille P, Choulet F, Laugier C, Šimková H, Šafář J, Bellec A, Vautrin S, Frenkel Z, Cattonaro F, Magni F, Scalabrin S, Martis MM, Mayer KF, Korol A, Bergès H, Doležel J, Feuillet C (2013) A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. Genome Biol 14:R64

Poursarebani N, Nussbaumer T, Šimková H, Šafář J, Witsenboer H, van Oeveren J, Doležel J, Mayer KF, Stein N, Schnurbusch T (2014) Whole-genome profiling and shotgun sequencing delivers an anchored, gene-decorated, physical map assembly of bread wheat chromosome 6A. Plant J 79:334-347

Qin B, Cao A, Wang H, Chen T, You FM, Liu Y, Ji J, Liu D, Chen P, Wang XE (2011) Collinearity-based marker mining for the fine mapping of *Pm6*, a powdery mildew resistance gene in wheat. Theor Appl Genet 123:207-218

Quraishi UM, Abrouk M, Bolot S, Pont C, Throude M, Guilhot N, Confolent C, Bortolini F, Praud S, Murigneux A, Charmet G, Salse J (2009) Genomics in cereals: from genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection. Funct Integr Genom 9:473–484

Raats D, Frenkel Z, Krugman T, Dodek I, Sela H, Šimková H, Magni F, Cattonaro F, Vautrin S, Bergès H, Wicker T, Keller B, Leroy P, Philippe R, Paux E, Doležel J, Feuillet C, Korol A, Fahima T (2013) The physical map of wheat chromosome 1BS provides insights into its gene space organization and evolution. Genome Biol 14:R138

Rabinowicz PD,McCombieWR,Martienssen RA (2003) Gene enrichment in plant genomic shotgun libraries. Curr Opin Plant Biol 6:150–156

Rapp RA, Udall JA, Wendel JF (2009) Genomic expression dominance in allopolyploids. BMC Biol 7:18

Ratnayaka I, Baga M, Fowler DB, Chibbar RN (2005) Construction and characterization of a BAC library of a cold-tolerant hexaploid wheat cultivar. Crop Sci 45:1574–1577

Riera-Lizarazu O, Leonard JM, Tiwari VK, Kianian SF (2010) A method to produce radiation hybrids for the D-genome chromosomes of wheat (*Triticum aestivum* L). Cytogenet Genome Res 129:234–240

Riley MC, Lee JE, Lesho E, Kirkup BC Jr (2011) Optically mapping multiple bacterial genomes simultaneously in a single run. PLoS One 6:e27085

Rothberg JM, Hinz W, Rearick TM, Schultz J, Mileski W, Davey M, Leamon JH, Johnson K, Milgrew MJ, Edwards M, Hoon J, Simons JF, Marran D, Myers JW, Davidson JF, Branting A, Nobile JR, Puc BP, Light D, Clark TA, Huber M, Branciforte JT, Stoner IB, Cawley SE, Lyons M, Fu Y, Homer N, Sedova M, Miao X, Reed B, Sabina J, Feierstein E, Schorn M, Alanjary M, Dimalanta E, Dressman D, Kasinskas R, Sokolsky T, Fidanza JA, Namsaraev E, McKernan KJ, Williams A, Roth GT, Bustillo J (2011) An integrated semiconductor device enabling non-optical genome sequencing. Nature. 475:348-352

Rustenholz C, Choulet F, Laugier C, Šafář J, Šimková H, Doležel J, Magni F, Scalabrin S, Cattonaro F, Vautrin S, Bellec A, Bergès H, Feuillet C, Paux E

(2011) A 3,000-loci transcription map of chromosome 3B unravels the structural and functional features of gene islands in hexaploid wheat. Plant Physiol 157:1596-1608

Šafář J, Bartoš J, Janda J, Bellec A, Kubaláková M, Valárik M, Pateyron S, Weiserová J, Tušková R, Číhalíková J, Vrána J, Šimková H, Faivre-Rampant P, Sourdille P, Caboche M, Bernard M, Doležel J, Chalhoub B (2004) Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. Plant J 39:960-968

Šafář J, Šimková H, Kubaláková M, Suchánková P, Číhálíková J (2009) Generating resources for genomics of wheat homoeologous chromosome group 3: 3AS- and 3DS-specific BAC libraries. J Genet Breeding 61:151–160

Šafář J, Šimková H, Kubaláková M, Číhalíková J, Suchánková P, Bartoš J, Doležel J (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. Cytogenet Genome Res 129:211-23

Saintenac C, Falque M, Martin OC, Paux E, Feuillet C, Sourdille P (2009) Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (*Triticum aestivum* L). Genetics 181:393–403

Saintenac C, Zhang W, Salcedo A, Rouse MN, Trick HN, Akhunov E, Dubcovsky J (2013) Identification of wheat gene *Sr35* that confers resistance to Ug99 stem rust race group. Science 341:783–786

Salamini F, Özkan H, Brandolini A, Schäfer-Pregl R, Martin W (2002) Genetics and geography of wild cereal domestication in the near east. Nat Rev Genet 3:429–441

Salse J, Abrouk M, Bolot S, Guilhot N, Courcelle E, Faraut T,Waugh R, Close TJ, Messing J, Feuillet C (2009) Reconstruction of monocotelydoneous proto-chromosomes reveals faster evolution in plants than in animals. Proc Natl Acad Sci USA 106:14908–14913

Salvi S, Tuberosa R (2005) To clone or not to clone plant QTLs: present and future challenges. Trends Plant Sci 10:297-304

Sandhu D, Gill KS (2002) Gene-containing regions of wheat and the other grass genomes. Plant Physiol 128:803–811

Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain-terminating inhibitors. Proc Natl Acad Sci USA 74:5463–5467

Schadt EE, Linderman MD, Sorenson J, Lee L, Nolan GP (2010) Computational solutions to large-scale data management and analysis. Nat Rev Genet 11:647–657

Schatz MC, Delcher AL, Salzberg SL (2010) Assembly of large genomes using second-generation sequencing. Genome Res 20:1165–1173

Schnable JC, Springer NM, Freeling M (2011) Differentiation of the maize subgenomes by genome dominance and both ancient and ongoing gene loss. Proc Natl Acad Sci USA 108:4069–4074

Schneider K (2005) Mapping populations and principles of genetic mapping. In: Meksem K, Kahl G (eds) The handbook of plant genome mapping. Genetic and physical mapping. pp 3 - 22 Wiley-WCH Weinheim

Schnurbusch T, Collins NC, Eastwood RF, Sutton T, Jefferies SP, Langridge P (2007) Fine mapping and targeted SNP survey using rice-wheat gene colinearity in the region of the *Bo1* boron toxicity tolerance locus of bread wheat. Theor Appl Genet 115:451–461

Sears ER, Sears LMS (1978) The telocentric chromosomes of common wheat. In: Ramanujams S (ed) Proc 5th Int Wheat Genetics Symp. pp 389–407 Indian Agricultural Research Institute New Delhi

Shearer LA, Anderson LK, de Jong H, Smit S, Goicoechea JL, Roe BA, Hua A, Giovannoni JJ, Stack SM (2014) Fluorescence *in situ* hybridization and optical mapping to correct scaffold arrangement in the tomato genome. G3 (Bethesda) 4:1395-405

Shen B, Wang DM, McIntyre CL, Liu CJ (2005) A 'Chinese Spring' wheat (*Triticum aestivum* L) bacterial artificial chromosome library and its use in the isolation of SSR markers for targeted genome regions. Theor Appl Genet 111:1489–1494

Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, Simon M (1992) Cloning and stable maintenance of 300-kilobasepair fragment of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci USA 89:8794–8797

Šimková H, Janda J, Hřibová E, Šafář J, Doležel J (2007) Cot-based cloning and sequencing of the short arm of wheat chromosome 1B. Plant Soil Environ 53:437-441

Šimková H, Svensson JT, Condamine P, Hřibová E, Suchánková P, Bhat PR, Bartoš J, Šafář J, Close TJ, Doležel J (2008) Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. BMC Genomics 9:294

Simpson JT, Wong K, Jackman SD et al (2009) ABySS: a parallel assembler for short read sequence data. Genome Res 19:1117–1123

Slade AJ, Fuerstenberg SI, Loeffler D, Steine MN, Facciotti D (2005) A reverse genetic, nontransgenic approach to wheat crop improvement by TILLING. Nat Biotechnol 23:75–81

Smith DB, Flavell RB (1975) Characterization of wheat genome by renaturation kinetics. Chromosoma 50:223–242

Smith CM (2005) Plant resistance to arthropods. Molecular and conventional aproaches. Springer Dordrecht

Smith CM, Liu X, Wang LJ, Liu X, Chen MS, Starkey S, Bai J (2010) Aphid feeding activates expression of a transcriptome of oxylipin-based defense signals in wheat involved in resistance to herbivory. J Chem Ecol 36:260–276

Soderlund C, Longden I, Mott R (1997) FPC: a system for building contigs from restriction fingerprinted clones. Comput Appl Biosci 13:523-535

Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. Genome Res 10:1772-1787

Staňková H, Valárik M, Lapitan NL, Berkman PJ, Batley J, Edwards D, Luo MC, Tulpová Z, Kubaláková M, Stein N, Doležel J, Šimková H (2015) Chromosomal genomics facilitates fine mapping of a Russian wheat aphid resistance gene. Theor Appl Genet 128:1373-1383

Staňková H, Hastie AR, Chan S, Vrána J, Tulpová Z, Kubaláková M, Visendi P, Hayashi S, Luo MC, Batley J, Edwards D, Doležel J, Šimková H (accepted paper) BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. Plant Biotechnol J

Stein N, Ponelies N, Musket T, McMullen M, Weber G (1998) Chromosome micro-dissection and region-specific libraries from pachytene chromosomes of maize (*Zea mays* L). Plant J 13:281–289

Suenaga K, Khairallah M, William HM, Hoisington DA (2005) A new intervarietal linkage map and its application for quantitative trait locus analysis of 'gigas' features in bread wheat. Genome 48: 65–75

Sulston J, Mallett F, Staden R, Durbin R, Horsnell T, Coulson A (1988) Software for genome mapping by fingerprinting techniques. Comput Appl Biosci 4:125-132

Teague B, Waterman MS, Goldstein S, Potamousis K, Zhou S, Reslewic S, Sarkar D, Valouev A, Churas C, Kidd JM, Kohn S, Runnheim R, Lamers C, Forrest D, Newton MA, Eichler EE, Kent-First M, Surti U, Livny M, Schwartz DC (2010) High-resolution human genome structure by single-molecule analysis. Proc Natl Acad Sci USA 107:10848-10853

Terracciano I, Maccaferri M, Bassi F, Mantovani P, Sanguineti MC, Salvi S, Šimková H, Doležel J, Massi A, Ammar K, Kolmer J, Tuberosa R (2013) Development of COS-SNP and HRM markers for haplotyping and marker-assisted selection of *Lr14* in durum wheat (*Triticum durum* Desf). Theor Appl Genet 26:1077-1101

Tiwari VK, Riera-Lizarazu O, Gunn HL, Lopez K, Iqbal MJ, Kianian SF, Leonard JM (2012) Endosperm tolerance of paternal aneuploidy allows radiation hybrid mapping of the wheat D-genome and measure of γ ray-induces chromosome breaks. PLoS ONE 7: e48815

Tuskan GA, Difazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, Putnam N, Ralph S, Rombauts S, Salamov A, Schein J, Sterck L, Aerts A, Bhalerao RR, Bhalerao RP, Blaudez D, Boerjan W, Brun A, Brunner A, Busov V, Campbell M, Carlson J, Chalot M, Chapman J, Chen GL, Cooper D, Coutinho PM, Couturier J, Covert S, Cronk Q, Cunningham R, Davis J, Degroeve S, Déjardin A, Depamphilis C, Detter J, Dirks B, Dubchak I, Duplessis S, Ehlting J, Ellis B, Gendler K, Goodstein D, Gribskov M, Grimwood J, Groover A, Gunter L, Hamberger B, Heinze B, Helariutta Y, Henrissat B, Holligan D, Holt R, Huang W, Islam-Faridi N, Jones S, Jones-Rhoades M, Jorgensen R, Joshi C, Kangasjärvi J, Karlsson J, Kelleher C, Kirkpatrick R, Kirst M, Kohler A, Kalluri U, Larimer F, Leebens-Mack J, Leplé JC, Locascio P, Lou Y, Lucas S, Martin F, Montanini B, Napoli C, Nelson DR, Nelson C, Nieminen K, Nilsson O, Pereda V, Peter G, Philippe R, Pilate G, Poliakov A, Razumovskaya J, Richardson P, Rinaldi C, Ritland K, Rouzé P, Ryaboy D, Schmutz J, Schrader J, Segerman B, Shin H, Siddiqui A, Sterky F, Terry A, Tsai CJ, Uberbacher E, Unneberg P, Vahala J, Wall K, Wessler S, Yang G, Yin T, Douglas C, Marra M, Sandberg G, Van de Peer Y, Rokhsar D (2006) The genome of black cottonwood, *Populus trichocarpa* (Torr & Gray). Science 15:1596–1604

Uauy C, Distelfeld A, Fahima T, Blechl A, Dubcovsky J (2006) A NAC Gene regulating senescence improves grain protein, zinc, and iron content in wheat. Science 314:1298-1301

Valdez VA, Byrne PF, Lapitan NLV, Peairs FB, Bernardo A, Bai G, Haley SD (2012) Inheritance and genetic mapping of Russian wheat aphid resistance in iranian wheat landrace accession PI626580. Crop Sci 52:676-682

Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, Zeng K, Malek JA, Costa G, McKernan K, Sidow A, Fire A, Johnson SM (2008) A high-resolution, nucleosome position map of *C. elegans* reveals a lack of universal sequence-dictated positioning. Genome Res 18:1051–1063

van Oeveren J, de Ruiter M, Jesse T, van der Poel H, Tang J, Yalcin F, Janssen A, Volpin H, Stormo KE, Bogden R, van Eijk MJ, Prins M (2011) Sequence-based physical mapping of complex genomes by whole genome profiling. Genome Res 21:618–625

Vitulo N, Albiero A, Forcato C, Campagna D, Dal Pero F, Bagnaresi P, Colaiacovo M, Faccioli P, Lamontanara A, Šimková H, Kubaláková M, Perrotta G, Facella P, Lopez L, Pietrella M, Gianese G, Doležel J, Giuliano G, Cattivelli L, Valle G,

Stanca AM (2011) First survey of the wheat chromosome 5A composition through a next generation sequencing approach. PLoS One 6:e26421

Voothuluru P, Meng J, Khajuria C, Louis J, Zhu L, Starkey S, Wilde GE, Baker CA, Smith CM (2006) Categories and inheritance of resistance to Russian wheat aphid (Homoptera: Aphididae) biotype 2 in a selection from wheat cereal introduction 2401. J Econ Entomol 99:1854-1861

Vrána J, Kubaláková M, Šimková H, Číhalíková J, Lysák MA, Doležel J (2000) Flow-sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L). Genetics 156: 2033–2041

Vrána J, Šimková H, Kubaláková M, Číhalíková J, Doležel J (2012) Flow cytometric chromosome sorting in plants: The next generation. Methods 57:331-337

Wang M, Wang S, Xia G (2015) From genome to gene: a new epoch for wheat research? Trends Plant Sci 20:380-387

Waterhouse PM, Helliwell CA (2003) Exploring plant genomes by RNA-induced gene silencing. Nat Rev Genet 4:29–38

Wicker T, Yahiaoui N, Guyot R, Schlagenhauf E, Liu ZD, et al (2003) Rapid genome divergence at orthologous low molecular weight glutenin loci of the A and Am genomes of wheat. Plant Cell 15:1186–1197

Wicker T, Mayer KF, Gundlach H, Martis M, Steuernagel B, Scholz U, Šimková H, Kubaláková M, Choulet F, Taudien S, Platzer M, Feuillet C, Fahima T, Budak H, Doležel J, Keller B, Stein N (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. Plant Cell 23:1706–1718

Winfield MO, Wilkinson PA, Allen AM, Barker GL, Coghill JA, Burridge A, Hall A, Brenchley RC, D'Amore R, Hall N, Bevan MW, Richmond T, Gerhardt DJ, Jeddeloh JA, Edwards KJ (2012) Targeted re-sequencing of the allohexaploid wheat exome. Plant Biotechnol J 10:733-742

Yahiaoui N, Srichumpa P, Dudler R, Keller B (2004) Genome analysis at different ploidy levels allows cloning of the powdery mildew resistance gene *Pm3b* from hexaploid wheat. Plant J 37:528-538

Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J (2003) Positional cloning of the wheat vernalization gene *VRN1*. Proc Natl Acad Sci USA 100:6263-6268

Yan L, Loukoianov A, Blechl A, Tranquilli G, RamakrishnaW, SanMiguel P, Bennetzen JL, Echenique V, Dubcovsky J (2004) The wheat *VRN2* gene is a flowering repressor down-regulated by vernalization. Science 303:1640–1644

Yan L, Fu D, Li C, Blechl A, Tranquilli G, Bonafede M, Sanchez A, Valárik M, Yasuda S, Dubcovsky J (2006) The wheat and barley vernalization gene *VRN3* is an orthologue of *FT*. Proc Natl Acad Sci USA 103:19581–19586

You FM, Luo MC, Xu K, Deal KR, Anderson OD, and Dvorak J (2010) A new implementation of high-throughput five-dimensional clone pooling strategy for BAC library screening. BMC Genomics 11:692

Zerbino DR, Birney E (2008) Velvet: algorithms for *de novo* short read assembly using de Bruijn graphs. Genome Res 18:821–829

Zhang J, Li C, Zhou Q, Zhang G (2015) Improving the ostrich genome assembly using optical mapping data. Gigascience 4:24

Zhou S, Schwartz DC (2004) The Optical Mapping of microbial genomes. ASM News 70:323-330

Zhou S, Kile A, Bechner M, Place M, Kvikstad E, Deng W, Wei J, Severin J, Runnheim R, Churas C, Forrest D, Dimalanta ET, Lamers C, Burland V, Blattner FR, Schwartz DC (2004) Single-molecule approach to bacterial genomic comparisons via optical mapping. J Bacteriol 186:7773-7782

Zhou S, Herschleb J, Schwartz DC (2007a) A single molecule system for whole genome analysis. In: Perspectives in bioanalysis. Volume 2. pp 269-304 Elsevier BV

Zhou S, Bechner MC, Place M, Churas CP, Pape L, Leong SA, Runnheim R, Forrest DK, Goldstein S, Livny M, Schwartz DC (2007b) Validation of rice genome sequence by optical mapping. BMC Genomics 8:278

Zhou S, Wei F, Nguyen J, Bechner M, Potamousis K, Goldstein S, Pape L, Mehan MR, Churas C, Pasternak S, Forrest DK, Wise R, Ware D, Wing RA, Waterman MS, Livny M, Schwartz DCA (2009) Single molecule scaffold for the maize genome. PLoS Genet 5:e1000711

Zimin A, Stevens KA, Crepeau MW, Holtz-Morris A, Koriabine M, Marçais G, Puiu D, Roberts M, Wegrzyn JL, de Jong PJ, Neale DB, Salzberg SL, Yorke JA, Langley CH (2014) Sequencing and assembly of the 22-Gb loblolly pine genome. Genetics196:875–90

# 3 AIMS OF THE THESIS

**I**   **Construction and validation of the physical contig map of the wheat chromosome arm 7DS**

First aim of the thesis was to construct a physical contig map for the short arm of the wheat chromosome 7D, select and sequence minimal tilling path (MTP), and finally validate the physical map by an alternative mapping technology.

**II**   **Construction of a high-density genetic map in *Dn2401* gene region**

The second aim of the thesis was to construct a high-density genetic map in the region of an aphid resistance gene - *Dn2401*. This key step in the positional cloning process should enable to span the region by a physical map contig and potentially identify candidate gene(s).

# 4 RESULTS

## 4.1 Original papers

4.1.1    BioNano genome mapping of individual chromosomes supports physical
         mapping and sequence assembly in complex plant genomes
         (Appendix I)


4.1.2    Chromosomal genomics facilitates fine mapping of a Russian wheat aphid
         resistance gene
         (Appendix II)

### 4.1.1 BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes

Staňková H, Hastie AR, Chan S, Vrána J, Tulpová Z, Kubaláková M, Visendi P, Hayashi S, Luo MC, Batley J, Edwards D, Doležel J, Šimková H

**Abstract:**

The assembly of a reference genome sequence of bread wheat is challenging due to its specific features such as the genome size of 17 Gb, polyploid nature and prevalence of repetitive sequences. BAC-by-BAC sequencing based on chromosomal physical maps, adopted by the International Wheat Genome Sequencing Consortium as the key strategy, reduces problems caused by the genome complexity and polyploidy, but the repeat content still hampers the sequence assembly. Availability of a high-resolution genomic map to guide sequence scaffolding and validate physical map and sequence assemblies would be highly beneficial to obtaining an accurate and complete genome sequence. Here we chose the short arm of chromosome 7D (7DS) as a model to demonstrate for the first time that it is possible to couple chromosome flow sorting with genome mapping in nanochannel arrays and create a *de novo* genome map of a wheat chromosome. We constructed a high-resolution chromosome map composed of 371 contigs with an N50 of 1.3 Mb. Long DNA molecules achieved by our approach facilitated chromosome-scale analysis of repetitive sequences and revealed a ~800-kb array of tandem repeats intractable to current DNA sequencing technologies. Anchoring 7DS sequence assemblies obtained by clone-by-clone sequencing to the 7DS genome map provided a valuable tool to improve the BAC-contig physical map and validate sequence assembly on a chromosome-arm scale. Our results indicate that creating genome maps for the whole wheat genome in a chromosome-by-chromosome manner is feasible and that they will be an affordable tool to support the production of improved pseudomolecules.

## 4.1.2 Chromosomal genomics facilitates fine mapping of a Russian wheat aphid resistance gene

Staňková H, Valárik M, Lapitan NL, Berkman PJ, Batley J, Edwards D, Luo MC, Tulpová Z, Kubaláková M, Stein N, Doležel J, Šimková H

**Abstract:**

Positional gene cloning and targeted marker development in bread wheat are hampered by high complexity and polyploidy of its nuclear genome. Aiming to clone a Russian wheat aphid resistance gene *Dn2401* located on wheat chromosome arm 7DS, we have developed a strategy overcoming problems due to polyploidy and enabling efficient development of gene-associated markers from the region of interest. We employed information gathered by Genome Zipper, a synteny-based tool combining sequence data of rice, *Brachypodium*, sorghum and barley, and took advantage of a high-density linkage map of *Aegilops tauschii*. To ensure genome- and locus-specificity of markers, we made use of survey sequence assemblies of isolated wheat chromosomes 7A, 7B and 7D. Despite the low level of polymorphism of the wheat D subgenome, our approach allowed us to add in an efficient and cost-effective manner 11 new gene-associated markers in the *Dn2401* region and narrow down the target interval to 0.83 cM. Screening 7DS-specific BAC library with the flanking markers revealed a contig of four BAC clones that span the *Dn2401* region in wheat cultivar 'Chinese Spring'. With the availability of sequence assemblies and GenomeZippers for each of the wheat chromosome arms, the proposed strategy can be applied for focused marker development in any region of the wheat genome.

## 4.2  Published abstracts – poster presentations

4.2.1    Sequence-based tools facilitate high-density mapping of a Russian wheat
         aphid resistance gene in wheat
         (Appendix III)


4.2.2    Poziční klonování genu pro rezistenci k mšici zhoubné (*Diuraphis noxia*):
         Konstrukce vysokohustotní genetické mapy [In Czech]
         (Appendix IV)


4.2.3    Physical map and sequencing of wheat chromosome arm 7DS
         (Appendix V)


4.2.4    Improving the physical map of the wheat 7DS chromosome arm with a
         BioNano genome map
         (Appendix VI)

## 4.2.1 Sequence-based tools facilitate high-density mapping of a Russian wheat aphid resistance gene in wheat

Staňková H, Valárik M, Šafář J, Lapitan N, Stein N, Berkman P, Edwards D, Doležel J, Šimková H

**Abstract:**

Bread wheat (*Triticum aestivum* L.) is one of the most important crop species providing the staple food for 35 % of the world's population. Genome mapping as well as positional cloning in bread wheat are hampered by its huge genome size (~17 Gb), polyploidy and large amount of repetitive sequences (>80 %).

Russian wheat aphid (*Diuraphis noxia*) is an important world invasive pest of bread wheat. Host resistance is the most efficient, economical, and environmentally safe approach to protect wheat from pathogens while minimizing the use of pesticides. Several genes contributing to RWA resistance were identified in various wheat lines. CI2401 line carries *Dn2401* gene that underlies resistance to this aphid species on the short arm of chromosome 7D (7DS).

Construction of a   high-density genetic map covering the *Dn2401* gene region is essential for positional cloning of this resistance gene as well as marker-assisted selection. Here, we present an approach to marker development employing new sequence-based tools and resources. Using the GenomeZipper created by combining 454 reads of particular barley chromosomes, barley ESTs and *Brachypodium*, rice and sorghum sequences, we succeeded in delimiting syntenic regions in genomes of barley, rice, *Brachypodium* and sorghum comprising our region of interest. Data from GenomeZipper was combined with Illumina paired-read sequence data generated recently from the flow-sorted 7AS, 7BS and 7DS wheat chromosome arm, respectively. High coverage of the reads enabled assembly and scaffolding of unique and low copy regions of the chromosome arms. Sufficiently long sequence contigs enabled development of SNP markers from regions surrounding EST sequences, which are characterised by a higher level of polymorphism than genic regions.

Knowledge of sequences from all three homoeologous chromosome arms enabled us to cope with polyploidy in wheat and to develop markers specific for 7DS.

## 4.2.2 Poziční klonování genu pro rezistenci k mšici zhoubné (*Diuraphis noxia*): Konstrukce vysokohustotní genetické mapy

Staňková H, Valárik M, Lapitan N, Berkman P, Edwards D, Luo MC, Tulpová Z, Kubaláková M, Stein N, Doležel J, Šimková H

[In Czech]

**Abstrakt:**

Pšenice setá (*Triticum aestivum* L.) je jednou z ekonomicky nejvýznamnějších kulturních plodin, poskytující zdroj potravy pro 35 % obyvatel světa. Jedná se o allohexaploidním druh (2n = 6x = 42) s celkovou velikostí genomu téměř 17 x $10^9$ bp. Genom je tvořen třemi homeologními subgenomy (A, B a D) a jeho podstatnou část (přes 80 %) tvoří repetitivní sekvence. Všechny výše zmíněné vlastnosti pšeničného genomu znesnadňují jeho analýzu, genetické i fyzické mapování, sekvenování či pozičního klonování. Třídění jednotlivých chromozómů a jejich ramen pomocí průtokové cytometrie umožňuje rozložit tento obrovský genom na malé a snadno analyzovatelné části.

Na krátkém rameni chromozómu 7D (7DS) pšenice se nachází řada agronomicky významných genů, včetně genu *Dn2401* pro rezistenci k mšici zhoubné (*Diuraphis noxia*). Mšice zhoubná je jedním z nejvýznamnějších škůdců pšenice a ječmene. Chemické i biologické postupy hubení nejsou v případě mšice zhoubné dostatečně účinné. Z tohoto důvodu se jeví jako nejvýhodnější způsob ochrany pěstování odrůd nesoucích geny pro rezistenci vůči tomuto škůdci.

Konstrukce vysokohustotní genetické mapy pokrývající oblast zkoumaného genu je nezbytná pro jeho následné poziční klonování, tedy izolaci genu na základě jeho pozice na genetické či fyzické mapě. Za účelem konstrukce této mapy byla vyvinuta metoda pro cílené odvozování markerů z úzké oblasti genomu, a to v podmínkách polyploidního genomu pšenice. Tato metoda využívá syntenie mezi pšenicí a jejími příbuznými druhy (ječmen, *Brachypodium*, rýže, čirok, *Aegilops tauschii*) v kombinaci se sekvencemi jednotlivých chromozómů skupiny 7,

získanými celochromozómovým neuspořádaným (shotgun) sekvenováním. Za pomocí genových markerů vymezujících oblast genu na rameni 7DS je možno identifikovat orthologní oblasti v genomech příbuzných druhů. Geny z těchto oblastí poté slouží k nalezení odpovídajících sekvenčních kontigů pocházejících z chromozómů 7A, 7B a 7D pšenice. Na základě polymorfizmů mezi jednotlivými homeologními chromozómy je možno designovat jednolokusové, genomově specifické markery. Popsaným postupem bylo odvozeno 11 nových vysoce specifických markerů, které posloužily k zahuštění mapy v okolí genu *Dn2401*. Skríning BAC knihovny z ramene 7DS umožnil identifikaci kontigu v 7DS-specifické fyzické mapě, který kompletně překlenuje oblast genu. BAC klony z této oblasti byly osekvenovány a v současné době probíhá jejich anotace. Sekvence BAC klonů mohou rovněž posloužit k odvození dalších markerů v blízkosti genu *Dn2401*. Prezentovaná metodika vývoje markerů je obecně aplikovatelná pro jakýkoliv chromozóm pšenice.

## 4.2.3 Physical map and sequencing of wheat chromosome arm 7DS

Staňková H, Luo MC, Visendi P, Tulpová Z, Batley J, Bartoš J, Doležel J,

Edwards D, Šimková H

**Abstract:**

Bread wheat (*Triticum aestivum*) is one of the most important crop species in the world. It provides the staple food for 35 % of world's population and its annual production can be challenged only by rice. Wheat huge genome size (1C = 17 Gb), presence of three homoeologous subgenomes (A, B and D) and prevalence of repetitive sequences (>80 %) hamper any genomic research in wheat including physical map construction and sequencing. The possibility to divide the wheat genome into individual chromosome arms by flow cytometry sorting enables coping with polyploidy and studying the wheat genome in a stepwise manner. In our project, we focused on sequencing the 7DS chromosome arm based on physical map of the 7DS. First, a 7DS-specific BAC library comprising 49,152 BAC clones representing 12.2 equivalents of the chromosome arm was constructed and all BAC clones were fingerprinted using SNaPshot-based HICF technology. Clones were automatically assembled into contigs based on fingerprint overlaps using FPC software. Integration of the 7DS physical map with that of *Aegilops tauschii* (D genome ancestor) provided a clue for further merging of contigs. Reliability of the assembly was verified through LTC software. Final BAC contig assembly resulted in 931 BAC contigs and a minimal tiling path (MTP) of 4,608 BAC clones. The physical map has been anchored to 7DS genetic maps by 583 markers, which located 307 contigs on the chromosome arm.  MTP BAC clones were pooled in fours and the pools were sequenced by Illumina MiSeq platform. Using Sassy software developed to assemble short complex sequences from Illumina paired read data, we succeeded in assembling the reads into 1 – 7 sequence contigs per BAC. Optimisation of the procedure aiming to further reduce the number of contigs per BAC is in progress. Sequences of BAC clones from the 7DS MTP will be used for anchoring of the 7DS

physical map. The anchored 7DS physical map as well as the sequence of the 7DS arm will become valuable tool for genetic mapping and positional cloning of genes located in this part of the wheat genome.

## 4.2.4 Improving the physical map of the wheat 7DS chromosome arm with a BioNano genome map

Staňková H, Hastie A, Cao H, Vrána J, Kubaláková M, Visendi P, Hayashi S, Luo MC, Batley J, Edwards D, Doležel J, Šimková H

**Abstract:**

Specific features of the wheat genome such as a huge genome size (17 Gb), presence of three homoeologous subgenomes (A, B and D) and prevalence of repetitive sequences hamper mapping and sequencing of this complex genome. To overcome these obstacles, a BAC-by-BAC sequencing strategy based on chromosomal physical maps has been adopted by the International Wheat Genome Sequencing Consortium.

In order to sequence a short arm of chromosome 7D (7DS), SNaPshot-based physical map was constructed using FPC and verified by LTC software. BAC clones representing minimal tilling path (MTP) were sequenced using Illumina HiSeq platform and assembled into contigs using Sassy software. With the aim to validate and improve the physical map as well as sequence assembly of 7DS arm, BioNano genome was constructed using Irys platform. The technology is based on visualisation of enzyme-specific nicking sites distributed along the DNA molecules hundreds to thousands of kilobases in length. For the 7DS, genome map was constructed using Nt.BspQI nickase and the final assembly consisted of 371 fragments with average length of 0.9 Mb and N50 of 1.3 Mb. Alignment of the sequences of 7DS MTP BAC clones enabled to verify the physical map assembly, size gaps and scaffold contigs. Comparison of the physical map assembly with the BioNano map also allowed anchoring, orientating, and ordering contigs along genome maps. Our results suggest that BioNano genome map provides a valuable tool to support, validate and improve the physical map and sequence assembly in the complex wheat genome.

## 4.3 Published abstract – oral presentation

4.3.1    BioNano genome mapping of flow-sorted chromosomes supports physical mapping and sequence assembly in complex plant genomes

### 4.3.1 BioNano genome mapping of flow-sorted chromosomes supports physical mapping and sequence assembly in complex plant genomes

Staňková H, Hastie A, Vrána J, Kubaláková M, Visendi P, Hayashi S, Luo MC, Batley J, Edwards D, Doležel J, Šimková H

**Abstract:**

Currently, majority of sequencing projects rely on the use of Next-Generation Sequencing technologies, which produce a huge amount of data in a relatively short time and at a low cost. However, assemblies based on NGS short reads are generally incomplete and highly fragmented due to large duplications and high proportion of repetitive DNA. Moreover, numerous finished genome assemblies were reported to suffer from serious errors.

BioNano genome mapping in nanochannel arrays using Irys platform offers a complementary tool for verifying and completion of sequence assemblies. The technology is based on single-molecule mapping of stretched DNA molecules hundreds to thousands kilobases in length. These molecules are nicked and labelled at sequence specific sites, linearized and stretched in nanochannels and subsequently imaged by a high-resolution camera. Consensus genome maps of megabase size showing distribution of the sequence motif along the chromosome are produced based on the single molecule data.

Here we report on using BioNano genome mapping for analysing bread wheat genome. High complexity (17 Gb) and polyploidy (2n = 6x = 42) of the wheat genome prevent analysis on a whole-genome level. Flow sorting showed to be an efficient approach to dissect the genome into particular chromosomes or even chromosome arms and thus significantly reduce the complexity. Moreover, high molecular weight DNA prepared from flow-sorted chromosomes showed to be of superior quality. Our pilot experiment on wheat 7DS chromosome arm resulted in constructing quality *de novo* genome map from 1.6 million flow-sorted 7DS arms. The obtained genome map covering 92 % of the arm was used for verification of the 7DS physical map, anchoring contigs, sizing gaps and identification of

misassemblies in sequence contigs. Moreover, several regions of hundreds kilobases composed of one type of tandem repeat were identified. Comparison of genome maps of two cultivars of bread wheat proved to be useful for detection of structural variation. Hitherto obtained results suggest that the BioNano genome maps derived from individual flow-sorted wheat chromosomes/arms represent an affordable tool to support physical map and sequence assembly and lead to the production of improved pseudomolecules.

# 5  CONCLUSIONS

In this thesis I focused on the study of the short arm of the wheat chromosome 7D (7DS). The first goal of this work was the construction and validation of the 7DS physical contig map. The other part of the thesis is dedicated to the positional cloning of a Russian wheat aphid resistance gene *Dn2401* located on the 7DS chromosome arm, particularly to the construction of a high-density genetic map covering the gene region.

## 5.1  Construction and validation of the physical contig map of the wheat chromosome arm 7DS

Construction of a quality physical contig map and a sequence assembly of MTP BAC clones represent indispensable steps towards obtaining the reference sequence of the genome through the BAC-by-BAC sequencing strategy. The physical contig map of the 7DS chromosome arm was constructed based on fingerprints of clones from a 7DS-specific BAC library, the minimal tilling path was selected and MTP BAC clones were sequenced by the Illumina next-generation sequencing technology. With the aim to validate and improve the physical map assembly, a 7DS-specific BioNano genome map was constructed using Irys platform and a comparison with the sequences of BAC clones representing the 7DS MTP was performed. The comparison showed that the BioNano mapping technology represents a useful tool for the validation and improvement of the physical map. It was demonstrated that integration of the physical and BioNano genome mapping can facilitate scaffolding, anchoring and ordering of contigs constituting the physical map. Based on the presented results, it is obvious that the BioNano mapping technology offers a missing tool needed to complement the extant genomics tools to deliver high quality reference genome sequence.

## 5.2  Construction of a high-density genetic map in *Dn2401* gene region

*Dn2401*, a gene underlying resistance to Russian wheat aphid, is one of the agronomically important genes located on the 7DS chromosome arm. Aiming to perform positional cloning of this gene, a high-density genetic map around the gene was constructed using a newly developed method for targeted development of highly

specific markers in the hexaploid wheat genome. The method benefits from the synteny with close relatives of wheat and chromosome shotgun sequences of a complete wheat homoeologous group. Described strategy is applicable in any region of the wheat genome and thus can be used in any positional cloning project conducted in wheat. Applying this strategy, 11 new gene-associated markers were developed and used for the saturation of the genetic map in the region of interest. Markers flanking the gene from both sites enabled to identify one contig from the 7DS physical map spanning the *Dn2401* region. Annotation of sequenced MTP BAC clones from this contig revealed presence of several candidate genes. The study proved that availability of the physical contig map can significantly facilitate and speed up the positional cloning process and that the physical map is an indispensable supplementary tool for positional cloning.

# 6 LIST OF ABBREVIATIONS

BAC            bacterial artificial chromosome

BES            BAC-end sequence/sequencing

bp            base pairs

Cas9            CRISPR associated protein 9

CB            consensus band

CBC            clone-by-clone

cDNA            complementary DNA

CRISPR            clustered regularly interspersed palindromic repeats

CS            Chinese Spring

DAPI            4',6-diamidino-2-phenylindole

DArT            diversity arrays technology

dDt            double-ditelosomic

DNA            deoxyribonucleic acid

EST            expressed sequence tag

FISH            fluorescence *in situ* hybridization

FISHIS            fluorescence *in situ* hybridization in suspension

Gb            gigabase pairs

HICF            high-information-content fingerprinting

ISBP            insertion site-based polymorphism

kb            kilobase pairs

LTR            long terminal repeat

Mb            megabase pairs

MDA            multiple displacement amplification

MHC            major histocompatibility complex

MITE            miniature inverted–repeat transposable element

| | |
|---|---|
| MTP | minimal tilling path |
| MYA | million years ago |
| NGS | next-generation sequencing |
| PCR | polymerase chain reaction |
| QTL | quantitative trait locus |
| RH | radiation hybrid |
| RNA | ribonucleic acid |
| RNAi | RNA interference |
| RNA-seq | RNA sequencing |
| SGS | second-generation sequencing |
| SNP | single nucleotide polymorphism |
| ssp. | subspecies |
| TALEN | transcription activator-like effector nuclease |
| TE | transposable element |
| TILLING | targeting induced local lesions in genomes |
| VIGS | virus-induced gene silencing |
| WGS | whole-genome shotgun |

# 7 LIST OF APPENDICIES

## Original papers

Appendix I:     BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes

Appendix II:    Chromosomal genomics facilitates fine mapping of a Russian wheat aphid resistance gene

## Published abstracts – poster presentations

Appendix III:   Sequence-based tools facilitate high-density mapping of a Russian wheat aphid resistance gene in wheat

Appendix IV:    Poziční klonování genu pro rezistenci k mšici zhoubné (*Diuraphis noxia*): Konstrukce vysokohustotní genetické mapy [In Czech]

Appendix V:     Physical map and sequencing of wheat chromosome arm 7DS

Appendix VI:    Improving the physical map of the wheat 7DS chromosome arm with a BioNano genome map

# APPENDIX I

**BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes**

Staňková H, Hastie AR, Chan S, Vrána J, Tulpová Z, Kubaláková M, Visendi P, Hayashi S, Luo M-C, Batley J, Edwards D, Doležel J, Šimková H

# BioNano Genome Mapping of Individual Chromosomes Supports Physical Mapping and Sequence Assembly in Complex Plant Genomes

**Full title:**

**BioNano Genome Mapping of Individual Chromosomes Supports Physical Mapping and Sequence Assembly in Complex Plant Genomes**

**Authors:**

Helena Staňková[1], Alex R. Hastie[2], Saki Chan[2], Jan Vrána[1], Zuzana Tulpová[1], Marie Kubaláková[1], Paul Visendi[3], Satomi Hayashi[4], Mingcheng Luo[5], Jacqueline Batley[4,6], David Edwards[6], Jaroslav Doležel[1], Hana Šimková[1]

**Names and addresses of institutions:**

[1]Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, CZ-78371, Olomouc, Czech Republic

[2]BioNano Genomics, 9640 Towne Centre Drive, San Diego, CA 92121, USA

[3]Australian Centre for Plant Functional Genomics, University of Queensland, Brisbane, QLD 4072, Australia

[4]School of Agriculture and Food Sciences, University of Queensland, Brisbane, QLD 4072, Australia

[5]Department of Plant Sciences, University of California, Davis, CA 95616, USA

[6]School of Plant Biology, University of Western Australia, Crawley, WA 6009, Australia

**Address for correspondence:**

Dr. Hana Šimková

Institute of Experimental Botany

Šlechtitelů 31

CZ-78371 Olomouc

Czech Republic

Phone: +420 585 238 715

Fax: +420 585 238 704

E-mail: simkovah@ueb.cas.cz


**Authors´s email adresses:**

Helena Staňková[1]        stankovah@ueb.cas.cz

Alex R. Hastie[2]         ahastie@bionanogenomics.com

Saki Chan[2]             schan@bionanogenomics.com

Jan Vrána[1]             vrana@ueb.cas.cz

Zuzana Tulpová[1]        tulpova@ueb.cas.cz

Marie Kubaláková[1]       kubalakova@ueb.cas.cz

Paul Visendi[3]          paul.muhindira@uqconnect.edu.au

Satomi Hayashi[4]        s.hayashi@uq.edu.au

Mingcheng Luo[5]         mcluo@ucdavis.edu

Jacqueline Batley[4,6]     jacqueline.batley@uwa.edu.au

David Edwards[6]         Dave.Edwards@uwa.edu.au

Jaroslav Doležel[1]       dolezel@ueb.cas.cz

Hana Šimková[1]          simkovah@ueb.cas.cz

## Summary

The assembly of a reference genome sequence of bread wheat is challenging due to its specific features such as the genome size of 17 Gbp, polyploid nature and prevalence of repetitive sequences. BAC-by-BAC sequencing based on chromosomal physical maps, adopted by the International Wheat Genome Sequencing Consortium as the key strategy, reduces problems caused by the genome complexity and polyploidy, but the repeat content still hampers the sequence assembly. Availability of a high-resolution genomic map to guide sequence scaffolding and validate physical map and sequence assemblies would be highly beneficial to obtaining an accurate and complete genome sequence. Here we chose the short arm of chromosome 7D (7DS) as a model to demonstrate for the first time that it is possible to couple chromosome flow sorting with genome mapping in nanochannel arrays and create a *de novo* genome map of a wheat chromosome. We constructed a high-resolution chromosome map composed of 371 contigs with an N50 of 1.3 Mb. Long DNA molecules achieved by our approach facilitated chromosome-scale analysis of repetitive sequences and revealed a ~800-kb array of tandem repeats intractable to current DNA sequencing technologies. Anchoring 7DS sequence assemblies obtained by clone-by-clone sequencing to the 7DS genome map provided a valuable tool to improve the BAC-contig physical map and validate sequence assembly on a chromosome-arm scale. Our results indicate that creating genome maps for the whole wheat genome in a chromosome-by-chromosome manner is feasible and that they will be an affordable tool to support the production of improved pseudomolecules.

## Introduction

Recent progress in understanding eukaryotic genome structure and function lead to the realization that a majority of genome sequences is transcribed and that, in addition to protein coding sequences, the so-called non-coding DNA may also be functionally significant (ENCODE Project Consortium, 2012). In addition, unexpected plasticity of eukaryotic genomes, and functional significance of copy number and structural variation has been revealed (Zarrei *et al.*, 2015). These observations underline the need for high quality reference genome sequences, which are a prerequisite to study these phenomena and discover genome features other than genes underlying traits of agronomic importance. Whilst next generation sequencing (NGS) technologies excel in huge throughput, reaching as much as trillions base pairs within a few days, the prevalent technologies provide short reads of only several hundred base pairs, making the assembly of large and complex genomes a daunting task.

As discussed recently, the published reference genome sequences obtained using whole genome shotgun strategies may suffer from extensive mis-assemblies and comprise gaps (Ruperao *et al.*, 2014; Ganapathy *et al.*, 2014; Pendelton *et al.*, 2015). This is also true to some extent for genome assemblies obtained even using the robust BAC-by-BAC approach (Shearer *et al.*, 2014; Callaway, 2014), indicating problems with the assembly of BAC contig physical maps. Thus, not a single plant or animal genome is truly complete (Kelley and Salzberg, 2010), and even the golden standard sequence of the human genome is known to be missing some genomic regions (Callaway 2014). Wider application of technologies providing reads in the kilobase range, such as Single Molecule Real-Time sequencing adopted by Pacific Biosciences (Chaisson *et al.*, 2015), and nanopore technologies (Mikheyev and Tin, 2014) promise to improve the quality of whole genome shotgun assemblies. Yet, even reads of this length are not enough, as it has been estimated that reads exceeding 200 kb would be needed to resolve repeats and other problematic regions (Marx, 2013). Current NGS technologies fall short of this, and thus the introduction of errors in whole genome assemblies cannot be avoided, which may negatively influence various downstream applications.

Several approaches have been applied for validation and correction of genome assemblies. For example, errors in clone-based physical maps and pseudomolecule mis-

assemblies can be identified using fluorescence *in situ* hybridization (FISH) with BAC clones as probes. Using this method, Shearer *et al.* (2014) showed that scaffolds representing one-third of the tomato genome were arranged incorrectly. Unfortunately, BAC-FISH is time consuming and laborious, and its applicability in large genomes is hampered by the presence of dispersed repetitive DNA, which make preparation of single-copy probes a tedious task (Janda *et al.*, 2006). Using a recently developed approach, incorrect assignment of sequences to particular chromosomes can be revealed by sequencing DNA of flow-sorted chromosomes (Ruperao *et al.*, 2014). Although this approach does not detect errors in ordering clone or sequence contigs within a chromosome, its relative simplicity makes it applicable in all species, from which chromosomes can be sorted. The principles of optical mapping were developed some time ago (Zhou and Schwartz, 2004), but only recently the technology and its modifications, such as genome mapping in nanochannel arrays (Lam *et al.*, 2012), became suitable for mapping large genomes (Zhou *et al.*, 2009; Young *et al.*, 2011; Dong *et al.*, 2013; Shearer *et al.*, 2014; Hastie *et al.*, 2013; Zhang *et al.*, 2015; Ganapathy *et al.*, 2014; Pendelton *et al.*, 2015). The method produces physical maps of short sequence motifs (i.e., recognition sites of nicking/restriction enzymes) along hundreds to thousands of kilobases-long stretches of DNA, and provides a high-throughput tool for ordering and orienting contigs of physical maps and validation of genome assemblies.

Bread wheat (*Triticum aestivum* L.), together with rice (*Oryza sativa*, L.) and maize (*Zea mays* L.) are the three most important crops and significant sources of calories and proteins for humankind. However, their genomes differ considerably in size and complexity, with bread wheat having by far the largest (~17 Gb) and polyploid genome consisting of three homoeologous subgenomes, A, B and D, with inter- and intrachromosomal duplications, and a high proportion of repetitive DNA (e.g. 85% for chromosome 3B; Choulet *et al.*, 2014a). The availability of a wheat reference genome sequence is needed urgently to employ molecular and genomic tools more extensively to speed up breeding improved varieties (Feuillet *et al.*, 2012; Choulet *et al.*, 2014b). The availability of a genome sequence would also make wheat an attractive model to study genome changes accompanying evolution of polyploid crop genomes and their

domestication. While various strategies have been employed to tackle the huge and complex bread wheat genome, including shotgun sequencing of the whole genome (Brenchley *et al.*, 2012; Chapman *et al.*, 2015) and shotgun sequencing of chromosomes isolated by flow sorting (IWGSC, 2014), it became obvious that a high quality genome sequence cannot be obtained from short-read shotgun data.

Considering the peculiarities of the wheat genome, The International Wheat Genome Sequencing Consortium (IWGSC) selected a clone-by-clone sequencing strategy based on physical maps constructed from chromosome (arm)-specific BAC libraries as a key approach towards obtaining the reference genome sequence (http://www.wheatgenome.org/). This approach offers a lossless reduction in complexity, in which the genome is sequenced *per partes*; avoids problems due to genome size, polyploidy, and large duplications; and greatly simplifies the genome assembly. In order to completely reconstruct the genomic sequence from BAC sequence data, contigs of the physical map are anchored and oriented. This relies on markers that are present in the contigs, and whose position on chromosomes is known. To satisfy this demand, large numbers of markers evenly distributed along chromosomes are needed. While high density linkage maps of wheat were recently constructed using high-throughput approaches, their resolution is limited due to relatively small sizes of the mapping populations. A particular challenge is posed by low-recombining regions, which may represent more than one third of the chromosome, and in which the resolution of genetic maps is poor (Erayman *et al.*, 2004; Paux *et al.*, 2008; Luo *et al.*, 2013). Radiation hybrid maps (Kumar *et al.*, 2012; Tiwari *et al.*, 2012) are largely independent of recombination and may aid in resolving this problem, but are not yet available for each of the wheat chromosomes. Alternative recombination-independent approaches are thus needed, and the BioNano genome mapping appears highly promising in this respect.

High accuracy of the genome mapping in nanochannel arrays enables *de novo* assembly of genome maps even without prior knowledge of genome sequence (Lam *et al.*, 2012). However, the huge and polyploid bread wheat genome appears too complex to be analyzed as a whole. Moreover, as the reference genome sequence is being produced by sequencing physical maps of individual chromosomes, or chromosome

arms, it seems practical to follow the chromosome-based strategy of IWGSC and produce BioNano maps from individual chromosomes. Here we chose the short arm of chromosome 7D (7DS) with the size of 381 Mb (Gill *et al.*, 1991; Šafář *et al.*, 2010) as a model to demonstrate for the first time that it is possible to couple chromosome flow sorting with genome mapping in nanochannel arrays to create a *de novo* genome map. DNA prepared from flow-sorted chromosomes was of superior quality and enabled construction of a high-resolution chromosome map. Moreover, long molecules achieved by our approach facilitated chromosome-scale analysis of repetitive sequences. Anchoring the 7DS genome map to the 7DS sequence assemblies obtained by clone-by-clone sequencing provided a valuable tool to improve the physical map and validate sequence assembly of the chromosome arm.

## Results

### *De novo assembly of a 7DS genome map*

The genome map of the 7DS chromosome arm of wheat was built from molecules treated by the nicking enzyme *Nt.BspQI* (labelled motif GCTCTTC). Statistics for data collection and genome map assembly are given in Table 1. In total 68.8 Gb data of DNA molecules over 150 kb were collected from one Irys chip, which corresponds to 180 equivalents of the 7DS chromosome arm. This coverage was compiled of 209,788 molecules, the largest of which exceeded 2 Mb in size (see Suppl. Figure 1). The N50 of the size-filtered molecules (>150 kb) was 354 kb. The 7DS genome map was assembled *de novo* and consisted of 371 constituent genome maps with average length of 0.9 Mb and N50 of 1.3 Mb. The size of the largest genome map was 4.6 Mb. The 7DS genome map has a total length of 350 Mb and covers 92% of the estimated arm length.

### *Tandem repeat detection and analysis*

Long DNA molecules obtained in our study enabled chromosome-scale analysis of repetitive sequences. During image acquisition on Irys, striking DNA molecules can

be seen that have evenly spaced labels (i.e. fluorescently labelled *Nt.BspQI*-nicked sites) that span over hundreds of kilobases. The regular labelling pattern indicates the presence of tandem repeats. In the wheat 7DS data, a particular labelled segment was seen with 9.3 kb spacing that spreads over a region of ~1 Mb in the genome map No. 350 (see Figure 1). Among single molecules underlying this map, we detected several comprising arrays of the evenly spaced labels of a minimum of 800 kb in length. Anchoring the available 7DS sequence scaffolds to the genome map No. 350 did not provide any significant match. This indicates the genome mapping revealed a hitherto unknown genome region composed of a long tandemly organized repeat, which is in its entirety intractable by traditional sequencing methods. Quantitative analysis of labelled tandem repeats within the whole 7DS dataset revealed that the majority of these repeats fall into the size category of 9.25-9.75 kb (Suppl. Figure 2), to which significantly contributes the repeat constituting the map No. 350. Potentially, the peak in repeat size can represent one type of repeat only and the size span is given by mutations or by variability in stretching among single molecules.

### *Optimization of sequence anchoring*

The genome map can serve as a guide for sequence assembling, provided available sequence contigs/scaffolds are long enough to be reliably anchored to the genome map. To determine the minimal sequence length needed for reliable anchoring within a wheat chromosome arm, we randomly selected ten BAC clones with inserts over 120 kb and typical labelling frequency (~12 sites/100 kb) assembled as one contiguous sequence (Suppl. Table 1). Comparison of these clones with the complete set of genome maps revealed their locations, which were determined as the best hits, reaching confidence value ranging from 15.85 to 24.89. The allocations were confirmed through anchoring of overlapping or neighbouring clones, which in all cases hit the selected genome map. Identical position for each of the clones was also obtained after truncating them to the size of 120 kb. Using a sliding window approach, 210 sequence fragments of three size categories (30 kb, 60 kb, and 90 kb) were generated. These comprised 100, 70 and 40 sequences of 30, 60 and 90 kb respectively. Applying this approach, a variety of nicking site patterns were obtained for each size category.

Comparison of the 210 sequences with genome maps provided multiple hits for all of the sequences. The best hit (highest confidence value) in the correct position was observed for 109 of them (52%). Data for particular size categories are given in Table 2. From the total number of one hundred 30-kb sequences, only 12% were assigned to the correct position, though with generally low confidence values (5.86 - 7.60). In the 60-kb size category, 57 out of 70 (81%) sequences were assigned correctly with confidence ranging from 6.07 to 13.87. The most reliable anchoring results were obtained with 90-kb sequences. In this category, all 40 sequences gained the highest-confidence value for the correct genome-map position. The variation in confidence level within a size category was mainly due to differing number of recognition sites of the nicking enzyme (Table 2): higher density of recognition sites generally increases reliability of the assignment.

The study indicated that without additional information, 30-kb sequences could not be reliably assigned to a genome map. In the 60-kb category, 70% sequences could be anchored with confidence value above 7. Knowledge of the sequence context, e.g. other sequence contigs belonging to the same or a neighbouring BAC clone known from the physical map, can aid reliable assignment of the short contigs through co-anchoring of the short sequences to genome maps. With preceding determination of a corresponding genome map, nearly double of the 30-kb sequences (21 %) could be assigned to the right position. For 60-kb sequences, the percentage of correctly positioned sequences rose from 81 to 87%. This approach can be used to order and orient shorter sequence contigs within a BAC clone or a pool of overlapping BAC clones.

Comparison of genome maps with complete sequences of the above BAC clones (in total 1376 kb sequence) enabled investigating error in size measurement introduced by mapping in nanochannel arrays. The size estimates showed to be highly precise, underestimating the sequence length by 1.4 kb (±0.56 kb) per 100 kb sequence.

### *BAC contig scaffolding and validation*
Long genome maps spanning several BAC contigs serve as a guide for building contig scaffolds in the length of megabases with precisely estimated gap sizes. They can

also point to potential contig overlaps and mis-assemblies, as demonstrated in Figure 2 for genome map No. 19 (GM19). This map with a length of 3.69 Mb is one of the largest in the assembly, thus having a potential to span several contigs of the physical map. Available sequence contigs of the minimum tilling path (MTP) BAC clones larger than 20 kb were aligned to GM19. Out of the complete set of 5847 contigs, 21 mapped to GM19 with a significant level of confidence. These sequence contigs anchored in total nine 7DS BAC contigs covering as a whole 89% of GM19 and oriented six of them (Fig. 2a). A list of MTP BAC clones from the anchored contigs is given in Suppl. Table 2. Contigs 713, 763 and 1857, which were anchored through one BAC clone each only, were oriented by allocating BAC-end sequences of overlapping clones within the sequence of the anchored one. We also identified three potential overlaps between BAC contigs ctg454 and ctg962, ctg546 and ctg1080, and ctg1080 and ctg3912, respectively. The overlaps between ctg546 and ctg1080, and ctg1080 and ctg3912, respectively, were confirmed by BLAST alignment of sequences of overlapping BAC clones, while the overlap between ctg454 and ctg962 could not be validated due to the lack of sequence data. We revealed five gaps between the contigs of the physical map, covering 410 kb total. Potentially, some of these gaps can be closed in the future once complete sequence information of all MTP clones is available. In other genome maps we confirmed that the map resolution was sufficient to resolve BAC contigs as short as three BAC clones represented by two MTP clones only. An example of a three-clone contig successfully anchored to a genome map is ctg1974 in Figure 4.

Alignment of BAC clone sequences to the GM19 pointed to a BAC clone that was incorrectly assigned to ctg3865 (Figure 2b). In contrast to another two BAC clones of this contig, which matched GM 19 with a high confidence, the end clone TaaCsp067A19 had no match with this genome map. Sequences of the mis-assigned clone showed no homology with sequences of overlapping clones, neither from ctg3865, nor from the potentially overlapping ctg1857. At the same time, the TaaCsp067A19 clone matched genome map No. 36, which proposed its positioning in ctg40. The proposed position has been confirmed by sequence overlaps with neighbouring clones.

*Anchoring of BAC contigs to the 7DS arm*

Traditionally, BAC contigs have been ordered along chromosomes through harboured markers with known positions in genetic or radiation hybrid (RH) maps, which created a big demand on the number of markers applied. A combination of various types of genomic resources (genetic and RH maps or synteny-based tools), which need to be integrated, is typical for the majority of physical mapping projects. The BioNano genome map provides an alternative means for positioning BAC contigs and at the same time enables a straightforward integration of various maps.

GM19 positioned four BAC contigs, (ctg454, ctg713, ctg763 and ctg3912) without any marker into the context of five contigs (ctg962, ctg546, ctg1080, ctg3865 and ctg1857) that had been anchored by a total of ten markers to the *Ae. tauschii* genetic map (Figure 2a). The marker order proposed by the genome map (Suppl. Table 2) was in agreement with the order of the markers in the genetic map (Luo *et al.*, 2013), which provides a support for the correctness of the genome map.

An example of the integration of various genetic maps through a genome map is given in Figure 3. Available sequence contigs of the MTP BAC clones >20 kb were aligned to the genome map 15, which resulted in anchoring seven sequences coming from four different BAC contigs. These contigs were previously allocated by markers to a total of three genetic maps: contig ctg192 was anchored to the genetic map of *Ae. tauschii* (Luo *et al.*, 2013), contigs ctg2449 and ctg864 were anchored to the consensus DArTseq map of bread wheat (Kilian, unpublished) and contig ctg738 was anchored to the wheat composite microsatellite map (http://wheat.pw.usda.gov/GG2/index.shtml). While the mutual positions of ctg2449 and ctg864 could be deduced from the DArTseq map and could also be confirmed by a sequence overlap between clones constituting the two contigs, the positions of ctg192 and ctg738 were only revealed from the genome map, which enabled estimating mutual positions of all the contigs as well as positions and approximate physical distances of *Ae. tauschii* SNP marker AT7D6156, DArTseq markers 7D_1191028 and 7D_1233886, and a microsatellite marker *Xbarc092*, respectively.

*Merging and scaffolding of genome maps*

Limitations for BioNano genome map assembly are posed by regions with low density of nicking sites and also 'fragile sites', caused by the occurrence of proximally located nicking sites on opposite DNA strands, which induce a biased fragmentation of the DNA (Lam *et al.*, 2012). Both limitations can be overcome by aligning the genome maps with BAC contigs that may span the problematic region and reliably scaffold the genome maps or serve as a guide for merging particular genome maps as demonstrated in Figure 4.

Alignment of the available set of 7DS MTP sequences to genome map No. 243 yielded four reliably anchored sequence contigs, which belonged to three BAC contigs - ctg3770, ctg1974 and ctg547. Since the outer contigs ctg3770 and ctg547 extended far beyond GM243, we aligned available sequences of other BAC clones from these contigs to the complete set of genome maps, which anchored genome map No. 158 proximal and genome map No. 68 distal of GM243, respectively. The alignment also revealed a small overlap between GM243 and GM158 and between GM243 and GM68, respectively. Thus three genome maps could be joined based on the information from the physical contig map. This approach provides a potential for a significant improvement of genome map assembly parameters.

## Discussion

The principles of optical mapping were developed some time ago (Zhou and Schwartz, 2004), but only recently the technology and its modifications such as genome mapping in nanochannel arrays (Lam *et al.*, 2012) became suitable for mapping large genomes. The largest optical/genome map assembled so far is that of human (3.2 Gb; Teague *et al.*, 2010; Pendelton *et al.*, 2015). It is obvious that such a tool is also extremely necessary for crops with larger genomes, including that of bread wheat.

Since the huge size (~17 Gb) and polyploid nature of the bread wheat genome may pose a serious obstacle to assembling a map on a whole-genome level, we proposed coupling the genome mapping with flow sorting of particular chromosomes/arms, which dissects the wheat genome into manageable portions of 224 to 993 Mb (Šafář *et al.*,

2010). The present work on the 7DS chromosome arm demonstrates that DNA prepared from flow-sorted chromosomes is of superior quality and allows *de novo* assembly of a quality genome map. This result confirms that the protocol of HMW DNA preparation from flow-sorted chromosomes (Šimková *et al.*, 2003), developed and applied previously for construction of chromosomal BAC libraries (Šafář *et al.*, 2010; http://olomouc.ueb.cas.cz/genomic-resources), is highly compatible with the BioNano Genomics Irys platform. Since the procedure of chromosome sorting has been elaborated for use in more than twenty plant species (Doležel *et al.*, 2014), this approach can find a wider application. Flow sorting of particular chromosome types, which substantially reduces sample complexity and helps to deal with polyploidy and segmental duplications, is limited to chromosomes with distinct size or to the availability of special cytogenetic stocks (addition, translocation, ditelosomic lines), which enable discrimination and flow-sorting of desired chromosomes (Doležel *et al.*, 2014). If discrimination of individual chromosomes is not possible, fractions enriched for particular chromosomes can be sorted to reduce sample complexity (Vrána *et al.*, 2015). In species, in which the chromosome sorting is not feasible, our flow-cytometry-based protocol can be used for purification of cell nuclei. In contrast to traditional methods for HMW DNA preparation, DNA prepared from flow-sorted nuclei is not contaminated by plastid and mitochondrial DNA and the protocol greatly reduces negative effects of secondary metabolites (Šimková *et al.*, 2003; Šafář *et al.* 2004).

The high quality of the HMW DNA prepared from the flow-sorted 7DS arm, which was reflected by the large size of single molecules (Suppl. Figure 1), enabled revealing an array of tandem repeats exceeding ~800 kb in length under the support of single molecules carrying a regular labelling pattern along DNA stretches of this size. Such regions are intractable to current short-read sequencing technologies, whose output assemblies are heavily biased against repeats and duplications because of short-read mapping ambiguity and assembly collapse (Alkan *et al.*, 2011). Even long-read technologies such as Single Molecule Real-Time sequencing using the PacBio platform or long-insert mate-pair sequencing are not able to reliably span a repeat region of this size. In the light of this, the missing match in the available 7DS sequence assemblies for the genome map carrying the array is not surprising. Alternatively, the region may have

been absent in the 7DS BAC library, which was the source of the sequence data, since tandem repeat regions inserted in a BAC vector induce recombination within the clone and thus are refractory to cloning. Genome mapping in nanochannel arrays relying on the analysis of single molecules of hundreds to thousands of kilobases in length proved useful for identifying regions of tandem repeats also in other organisms. Hastie *et al.* (2013) found two blocks of tandem repeats in a partial genome map of *Aegilops tauschii*, which were missing in the sequence assembly of the 2.1-Mb prolamin gene family region. Cao *et al.* (2014) analysed structural variation in a human genome and found an intact molecule of 633 kb harbouring two tracts of 2.5-kb tandem repeats, one of at least 53 copies, the other of at least 21 copies. In their study, the 2.5 kb showed the most abundant size category among labelled repeats in YH cell line (male), in contrast to line NA12878 (female), in which the frequency of the 2.5 kb repeat was 19 times lower. Based on additional genome mapping in other males and females, the 2.5 kb repeat appeared male-specific, indicating a potential biological role of tandem repeats in the genome and predicting them a source of structural variability.

Besides being an invaluable tool to study structural variation, the optical/genome maps were also highly beneficial in assembling genome sequences by aiding ordering, orienting, and joining contigs and scaffolds; sizing and closing gaps; anchoring the scaffolds; and identifying and correcting mis-assemblies (Zhou *et al.*, 2007, 2009; Young *et al.*, 2011; Dong *et al.*, 2013; Shearer *et al.*, 2014; Hastie *et al.*, 2013; Zhang *et al.*, 2015; Ganapathy *et al.*, 2014; Pendelton *et al.*, 2015). The majority of current sequencing projects rely on assembling short-read data obtained by shotgun sequencing, which results in heavily fragmented and frequently incorrect assemblies. While these can be improved through the genome maps, the length of assembled contigs/scaffolds must be sufficient to ensure reliable anchoring. Our study carried out on a wheat chromosome arm of relatively low complexity (381 Mb), but with a high proportion of repeats (over 80%), showed that sequence contigs of 90 kb could be unambiguously anchored to the 7DS genome map. We can extrapolate that the required sequence length will increase with the genome complexity and in genomes with a lower frequency of nicking sites. In our experiment, we were partially successful even with shorter sequence contigs, observing 81 and 12% of correct assignments for 60- and 30-kb sequences,

respectively. Whilst the confidence value of ~6, obtained for some of the shorter sequences (Table 2), would be prohibitively low if anchoring sequences from shotgun assemblies, it can still be applied in case of the BAC-by-BAC approach, thanks to the support of other clones from the same BAC contig, which reliably preselect the corresponding genome map. BAC-by-BAC sequencing of complex crops genomes, including wheat, is frequently carried out on pools of several overlapping BAC clones (Choulet *et al.*, 2014a; http://www.wheatgenome.org/). If validating, scaffolding and correcting sequence assemblies within a narrow genome region determined as a BAC contig, the alignment can be performed within one or two genome maps only, which allows decreasing the confidence value and mapping even the short sequences. This indicates that coupling the genome mapping with the BAC-by-BAC sequencing strategy is a powerful approach to resolving complex genomes.

With 90-kb anchorable sequence length, the genome mapping in nanochannel arrays outperforms the optical mapping technology of OpGen, Inc., which generally requires scaffolds ≥200 kb with less than 5% Ns for reliable map assignment (Zhang *et al.*, 2015). The reason for the difference may lie in inherent features of the two technologies. While genome mapping on the Irys platform is based on labelling DNA molecules in enzyme-specific nicking sites and subsequent automatic massively parallel imaging of DNA molecules in nanochannel arrays (Lam *et al.*, 2012), optical mapping is based on stretching long DNA molecules in a microfluidic device, attachment of molecules to the surface of the device by electrostatic interactions, subsequent digestion by specific restriction endonuclease, and staining (Zhou and Schwartz, 2004). The latter technique suffers from lower uniformity in DNA stretching and has a higher error rate (Cao *et al.*, 2014), which has to be compensated by a higher coverage of input data (Zhou *et al.*, 2007, 2009; Zhang *et al.*, 2015). In our study we observed high concordance between size estimates based on sequencing and the genome mapping; the genome maps in all cases underestimated the sequence length, on average by 1.4%. This systematic underestimate was probably due to non-discriminated recognition sites whose distance was under the resolution limit of the technology (1.5 kb). Despite higher error rate in the single-molecule data, genome maps generated by the optical mapping approach appear less fragmented and have larger contigs than those generated on the

Irys platform (e.g. Zhou *et al.*, 2007; 2009; Ganapathy *et al.*, 2014). This is mainly due to 'fragile sites' associated with nick sites adjacent to each other on the opposite DNA strands, which are specific to BioNano Genomics technology. Stitching these sites through contigs of the physical map, as demonstrated in our study, or through long sequence contigs, as shown by Cao *et al.* (2014) or Pendelton *et al.* (2015), can improve the assembly metrics significantly (Cao *et al.*, 2014).

In our study, we collected from one Irys chip 68.8 Gb size-filtered data, which corresponds to 180 equivalents of the 7DS chromosome arm. This exceeds the coverage of 70-80x, required for the BioNano technology (Cao *et al.*, 2014; Pendelton *et al.*, 2015), and suggests that one chip may provide sufficient data for two wheat chromosome arms. This implies that the whole wheat genome could be analysed chromosome-by-chromosome using only 21 Irys chips, which makes the analysis of a polyploid ~17 Gb genome a realistic goal.

We demonstrate that the BioNano genome map is a useful tool for ordering and orienting BAC contigs along a chromosome. This is extremely beneficial in non- or low-recombining regions of the genome, in which genetic mapping fails. Studies in wheat and its relatives revealed that low recombination rate may affect more than one third of chromosomal length (Erayman *et al.*, 2004; Paux *et al.*, 2008; Luo *et al.*, 2013), which suggests that the role of genome mapping may be invaluable in a significant part of the genome. Another challenge in projects based on BAC-by-BAC sequencing pose short contigs of a few BAC clones that are not easy to anchor and nearly impossible to orient. For this reason, contigs shorter than five (Paux *et al.*, 2008; Philippe *et al.*, 2013) or even six BAC clones (Breen *et al.*, 2013; Poursarebani *et al.*, 2014) were excluded from wheat physical map assemblies and are not subjected to sequencing. This approach can introduce gaps in sequence assemblies. The present study demonstrates that genome maps have a sufficient resolution to position and orient contigs consisting of as little as three BAC clones, which can contribute to higher completeness of generated genome sequences.

To conclude, using wheat chromosome arm 7DS as a model, we demonstrate the suitability of flow-sorted chromosomes for BioNano mapping technology. This approach facilitates physical map scaffolding, validation, correction, and anchoring. As

such it provides a missing tool needed to complement the extant genomics tools to deliver high quality reference genome sequences and analyse structural genome variation.

## Experimental procedures

### *Building BioNano genome map*

High molecular weight (HMW) DNA was prepared from wheat 7DS chromosome arm as described in Šimková *et al.* (2003). The 7DS arm was flow-sorted from a double ditelosomic line of wheat *Triticum aestivum* L. cv. Chinese Spring carrying both arms of chromosome 7D as telosomes. The seeds were kindly provided by Prof. B.S. Gill (KSU, Manhattan, USA) and Prof. A. Lukaszewski (UC, Riverside, USA). Liquid suspensions of intact chromosomes were prepared according to Kubaláková *et al.* (2002) by mechanical homogenization of 20-25 formaldehyde-fixed root-tip meristems enriched for metaphase cells in 1 ml ice-cold isolation buffer (IB; Šimková *et al.*, 2003). Chromosomes in suspension were stained with 2 µg/ml DAPI (4',6-diamidino-2-phenylindole) and analyzed using a FACSAria SORP flow cytometer (Becton Dickinson, San Jose, USA). Purity of the flow-sorted 7DS arm as estimated by FISH was 84%. The major contaminant in the sorted fraction was the 7DL telosome, which formed 1.1% of the sorted fraction; the remaining ~15% were made up of a mixture of other chromosomes. In total, 1.6 x $10^6$ 7DS arms corresponding to 1.2 µg DNA were flow-sorted and embedded in three agarose miniplugs of total volume 60 µl. DNA embedded in plugs was purified by proteinase K (Roche) treatment as described in Šimková *et al.* (2003). The miniplugs were washed four times in wash buffer (10 mM Tris, 50 mM EDTA, pH 8.0) and four times in TE buffer (10 mM Tris, 1 mM EDTA, pH 8.0), melted for 2 min at 70°C and solubilized with GELase (Epicentre, Madison, USA) for 45 min. The purified DNA underwent 30 minutes of drop dialysis (Merck Millipore, Billerica, USA) against TE buffer and was quantified using Quant-iT™ PicoGreen® dsDNA assay (Thermo Fisher Scientific, Waltham, USA).

Survey sequences of the 7DS chromosome arm (Berkman *et al.*, 2011; IWGSC, 2014) were inspected for frequency of recognition sites of particular nicking enzymes,

and nicking endonuclease *Nt.BspQI* with an estimated frequency of 12 sites/100 kb (labelling frequency) was selected for the labelling. DNA was labelled using the IrysPrep® Reagent Kit (BioNano Genomics, San Diego, USA) following manufacturer's instructions with modifications suitable for samples with lower DNA concentration. Specifically, 200 ng of purified chromosomal DNA were nicked using 2U of *Nt.BspQI* (New England BioLabs, Beverly, USA) at 37°C for two hours in NEBuffer 3. The nicked DNA was labelled with a fluorescent-dUTP nucleotide analogue using Taq polymerase (New England BioLabs) for one hour at 72°C. After labelling, the nicks were ligated with Taq ligase (New England BioLabs) in the presence of dNTPs. The backbone of the labelled DNA was stained with IrysPrep DNA Stain (BioNano Genomics).

Labelled and stained DNA was loaded on the Irys chip and run for two runs for a total of 41 cycles. A total of 82.5 Gb data was generated, of which 68.8 Gb exceeded 150 kb. After single molecules were detected to find the label positions on the DNA backbone, *de novo* assembly was performed by a pairwise comparison of all single molecules and graph building (Cao *et al.*, 2014). A p-value threshold of $10e^{-9}$ was used during the pairwise assembly, $10e^{-10}$ for extension and refinement steps, and $10e^{-11}$ for a final refinement.

***Repeat detection and analysis***

An algorithm included in the IrysView 2.0 software package (BioNano Genomics) was used to identify tandem repeats with one nick site per repeat motif (labelled tandem repeats), in both the assembly and the raw data. Detected repeats were quantified and their unit size and frequency in the dataset were plotted in a histogram for visual analysis. Arrays of five or more repeat units were considered in our analysis.

***Physical map construction, anchoring and sequencing***

The 7DS physical contig map (https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_7DS_v1/) was constructed from a 7DS-specific BAC library using FPC software (Soderlund *et al.*, 2000) as described in Šimková *et al.* (2011). A minimal tiling path (MTP) of 4,608 BAC clones selected from

the physical map were sequenced using a pooling strategy in which 96 pools, each consisting of four BACs, were indexed and sequenced on a single lane using the Illumina HiSeq 2000 platform. Sequences were de-multiplexed and assembled using the SASSY assembler (Kazakoff *et al.*, 2012). Deconvolution was supported by BAC-end sequences generated from the MTP BAC clones by Sanger sequencing. The resulting assembly has a mean N50 of 65 kb and currently covers about 75% of the 7DS arm. The sequence contigs were used for *in silico* anchoring of the 7DS physical map to the *Ae. tauschii* SNP genetic map (Luo *et al.*, 2013) and a consensus bread wheat DArTseq genetic map (Kilian, unpublished). 7DS-specific microsatellite markers from GrainGenes database (http://wheat.pw.usda.gov/GG2/index.shtml) were anchored manually by PCR screening of three-dimensional pools of the 7DS BAC library (Šimková *et al.*, 2011).

### *Anchoring 7DS sequence assemblies to the genome map*

Comparison of sequence assembly with the genome map was performed using the IrysView 2.0 software package. Based on the type of analysis, individual sequences representing 7DS MTP clones or complete pools of MTP BAC clones were compared with the complete set of genome maps or with individual genome maps, respectively. Prior to comparison, *cmap* files were generated from *fasta* files of individual sequences or BAC pools, respectively. Query-to-anchor comparison was performed with default parameters and variable p-value threshold ranging from $1e^{-6}$ to $1e^{-10}$, based on the type of analysis.

To estimate the minimum length of sequence needed for identifying the corresponding genome map, we randomly selected ten MTP BAC clones with inserts exceeding 120 kb assembled as a contiguous sequence (Suppl. Table 1). BAC clone sequences, checked for the typical frequency of nicking sites (~12 sites/100 kb), were compared with the whole set of genome maps as described above. Clone assignments to particular genome maps were validated by anchoring overlapping or neighboring clones identified previously by FPC. Subsequently, sequences of the analyzed BAC clones were truncated to the length of 120 kb and analyzed using a sliding window approach, applying three window sizes – 30 kb, 60 kb and 90 kb – and a window shift of 10 kb. All

generated sequence fragments were compared with the complete set of genome maps in IrysView software, which calculated a confidence value for each of the aligned sequences as -log10(p-value) where the p-value calculation is described in Anantharaman and Mishra (2001). To maximize the number of aligned sequences, the p-value threshold was set to $10e^{-4}$.

## Conflict of Interest

## Acknowledgements

# References

Alkan, C., Sajjadian, S., Eichler, E.E. (2011) Limitations of next-generation genome sequence assembly. *Nat. Methods*, 8, 61–65.

Anantharaman, T., Mishra, B. (2001) A probabilistic analysis of false positives in optical map alignment and validation. *Algorithms in Bioinformatics*, First International Workshop, WABI 2001 Proceedings, LNCS 2149: 27-40, Springer-Verlag

Berkman, P.J., Skarshewski, A., Lorenc, M.T., Lai, K., Duran, C., Ling, E.Y., Stiller, J., Smits, L., Imelfort, M., Manoli, S., McKenzie, M., Kubaláková, M., Šimková, H., Batley, J., Fleury, D., Doležel, J., Edwards, D. (2011) Sequencing and assembly of low copy and genic regions of isolated Triticum aestivum chromosome arm 7DS. *Plant Biotechnol. J.* 9, 768-775.

Breen, J., Wicker, T., Shatalina, M., Frenkel, Z., Bertin, I., Philippe, R., Spielmeyer, W., Šimková, H., Šafář, J., Cattonaro, F., Scalabrin, S., Magni, F., Vautrin, S., Bergès, H., International Wheat Genome Sequencing Consortium, Paux, E., Fahima, T., Doležel, J., Korol, A., Feuillet, C., Keller, B. (2013) A Physical Map of the Short Arm of Wheat Chromosome 1A. *PLoS ONE*, 8, e80272.

Brenchley, R., Spannagl, M., Pfeifer, M., Barker, G.L., D'Amore, R., Allen, A.M., McKenzie, N., Kramer, M., Kerhornou, A., Bolser, D., Kay, S., Waite, D., Trick, M., Bancroft, I., Gu, Y., Huo, N., Luo, M.C., Sehgal, S., Gill, B., Kianian, S., Anderson, O., Kersey, P., Dvorak, J., McCombie, W.R., Hall, A., Mayer, K.F., Edwards, K.J., Bevan, M.W., Hall, N. (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature*, 491, 705-710.

Callaway, E. (2014) 'Platinum' genome shapes up. *Nature*, 515, 323-323.

Cao, H., Hastie, A.R., Cao, D., Lam, E.T., Sun, Y., Huang, H., Liu, X., Lin, L., Andrews, W., Chan, S., Huang, S., Tong, X., Requa, M., Anantharaman, T., Krogh, A., Yang, H., Cao, H., Xu, X. (2014) Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. *GigaScience*, 3, 34.

Chaisson, M.J.P., Huddleston, J., Dennis, M.Y., Sudmant, P.H., Malig, M., Hormozdiari, F., Antonacci, F., Surti, U., Sandstrom, R., Boitano, M., Landolin, J.M., Stamatoyannopoulos, J.A., Hunkapiller, M.W., Korlach, J., Eichler, E.E. (2015)

Resolving the complexity of the human genome using single-molecule sequencing. *Nature*, 517, 608-611.

Chapman, J.A., Mascher, M., Buluç, A., Barry, K., Georganas, E., Session, A., Strnadova, V., Jenkins, J., Sehgal, S., Oliker, L., Schmutz, J., Yelick, K.A., Scholz, U., Waugh, R., Poland, J.A., Muehlbauer, G.J., Stein, N., Rokhsar, D.S. (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol*. 16, 26.

Choulet, F., Alberti, A., Theil, S., Glover, N., Barbe, V., Daron, J., Pingault, L., Sourdille, P., Couloux, A., Paux, E., Leroy, P., Mangenot, S., Guilhot, N., Le Gouis, J., Balfourier, F., Alaux, M., Jamilloux, V., Poulain, J., Durand, C., Bellec, A., Gaspin, C., Šafář, J., Doležel, J., Rogers, J., Vandepoele, K., Aury, J.M., Mayer, K., Bergès, H., Quesneville, H., Wincker, P., Feuillet, C. (2014a) Structural and functional partitioning of bread wheat chromosome 3B. *Science*, 345, 1249721.

Choulet, F., Caccamo, M., Wright, J., Alaux, M., Šimková, H., Šafář, J., Leroy, P., Doležel, J., Rogers, J., Eversole, K., Feuillet, C. (2014b) The Wheat Black Jack: Advances towards sequencing the 21 chromosomes of bread wheat. In: *Genomics of plant genetic resources* (Tuberosa, R., Graner, A., Frison, E., eds). pp. 405-438. Dordrecht: Springer Science+Business Media.

Doležel, J., Vrána, J., Cápal, P., Kubaláková, M., Burešová, V., Šimková, H. (2014) Advances in plant chromosome genomics. *Biotechnol. Adv*. 32, 122-136.

Dong, Y., Xie, M., Jiang, Y., Xiao, N., Du, X., Zhang, W., Tosser-Klopp, G., Wang, J., Yang, S., Liang, J., Chen, W., Chen, J., Zeng, P., Hou, Y., Bian, C., Pan, S., Li, Y., Liu, X., Wang, W., Servin, B., Sayre, B., Zhu, B., Sweeney, D., Moore, R., Nie, W., Shen, Y., Zhao, R., Zhang, G., Li, J., Faraut, T., Womack, J., Zhang, Y., Kijas, J., Cockett, N., Xu, X., Zhao, S., Wang, J., Wang, W. (2013) Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (Capra hircus). *Nat. Biotechnol*. 31, 135-141.

ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature*, 489, 57-74.

Erayman, M., Sandhu, D., Sidhu, D., Dilbirligi, M., Baenziger, P.S., Gill, K.S. (2004) Demarcating the gene-rich regions of the wheat genome. *Nucleic Acids Res*. 32, 3546–3565.

Feuillet, C., Stein, N., Rossini, L., Praud, S., Mayer, K., Schulman, A., Eversole, K., Appels, R., (2012) Integrating cereal genomics to support innovation in the Triticeae. *Funct. Integr. Genomics*, 12, 573-583.

Ganapathy, G., Howard, J.T., Ward, J.M., Li, J., Li, B., Li, Y., Xiong, Y., Zhang, Y., Zhou, S., Schwartz, D.C., Schatz, M., Aboukhalil, R., Fedrigo, O., Bukovnik, L., Wang, T., Wray, G., Rasolonjatovo, I., Winer, R., Knight, J.R., Koren, S., Warren, W.C., Zhang, G., Phillippy, A.M., Jarvis, E.D. (2014) High-coverage sequencing and annotated assemblies of the budgerigar genome. *GigaScience*, 3, 11.

Gill, B.S., Friebe, B., Endo, T.R. (1991) Standard karyotype and nomenclature system for description of chromosome bands and structural aberrations in wheat (Triticum aestivum). *Genome*, 34, 830-839.

Hastie, A.R., Dong, L., Smith, A., Finklestein, J., Lam, E.T., Huo, N., Cao, H., Kwok, P.Y., Deal, K.R., Dvorak, J., Luo, M.C., Gu, Y., Xiao, M. (2013) Rapid genome mapping in nanochannel arrays for highly complete and accurate de novo sequence assembly of the complex Aegilops tauschii genome. *PLoS One*, 8, e55864.

International Wheat Genome Sequencing Consortium (IWGSC) (2014) A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science*, 345, 1251788.

Janda, J., Šafář, J., Kubaláková, M., Bartoš, J., Kovářová, P., Suchánková, P., Pateyron, S., Číhalíková, J., Sourdille, P., Šimková, H., Faivre-Rampant, P., Hřibová, E., Bernard, M., Lukaszewski, A., Doležel, J., Chalhoub, B. (2006) Novel resources for wheat genomics: BAC library specific for the short arm of chromosome 1B. *Plant J*. 47, 977-986.

Kazakoff, S.H., Imelfort, M., Edwards, D., Koehorst, J., Biswas, B., Batley, J., Scott, P.T., Gresshoff, P.M. (2012) Capturing the biofuel wellhead and powerhouse: the chloroplast and mitochondrial genomes of the leguminous feed-stock tree Pongamia pinnata. *PLoS One*, 7, e51687.

Kelley, D.R., Salzberg, S.L. (2010) Detection and correction of false segmental duplications caused by genome mis-assembly. *Genome Biol*. 11, R28.

Kubaláková, M., Vrána, J., Číhalíková, J., Šimková, H., Doležel, J. (2002) Flow karyotyping and chromosome sorting in bread wheat (Triticum aestivum L.). *Theor. Appl. Genet*. 104, 1362–1372.

Kumar, A., Simons, K., Iqbal, M.J., Michalak de Jimenéz, M., Bassi, F.M., Ghavami, F., Al-Azzam, O., Drader, T., Wang, Y., Luo, M.C., Gu, Y.Q., Denton, A., Lazo, G.R., Xu, S.S., Dvorak, J., Kianian, P.M., Kianian, S.F. (2012) Physical mapping resources for large plant genomes: radiation hybrids for wheat D-genome progenitor Aegilops tauschii. *BMC Genomics*, 13, 597.

Lam, E.T., Hastie, A., Lin, C., Ehrlich, D., Das, S.K., Austin, M.D., Deshpande, P., Cao, H., Nagarajan, N., Xiao, M., Kwok, P.Y. (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat. Biotechnol*. 30, 771–776.

Luo, M.C., Gu, Y.Q., You, F.M., Deal, K.R., Ma, Y., Hu, Y., Huo, N., Wang, Y., Wang, J., Chen, S., Jorgensen, C.M., Zhang, Y., McGuire, P.E., Pasternak, S., Stein, J.C., Ware, D., Kramer, M., McCombie, W.R., Kianian, S.F., Martis, M.M., Mayer, K.F., Sehgal, S.K., Li, W., Gill, B.S., Bevan, M.W., Šimková, H., Doležel, J., Weining, S., Lazo, G.R., Anderson, O.D., Dvořák, J. (2013) A 4-gigabase physical map unlocks the structure and evolution of the complex genome of Aegilops tauschii the wheat D-genome progenitor. *Proc. Natl. Acad. Sci. USA*, 110, 7940-7945.

Marx, V. (2013) Next-generation sequencing: The genome jigsaw. *Nature*, 501, 263–268.

Mikheyev, A.S., Tin, M.M. (2014) A first look at the Oxford Nanopore MinION sequencer. *Mol. Ecol. Resour*. 14, 1097–1102.

Paux, E., Sourdille, P., Salse, J., Saintenac, C., Choulet, F., Leroy, P., Korol, A., Michalak, M., Kianian, S., Spielmeyer, W., Lagudah, E., Somers, D., Kilian, A., Alaux, M., Vautrin, S., Bergès, H., Eversole, K., Appels, R., Šafář, J., Šimková, H., Doležel, J., Bernard, M., Feuillet, C. (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. *Science*. 322, 101–104.

Pendleton, M., Sebra, R., Pang, A.W., Ummat, A., Franzen, O., Rausch, T., Stütz, A.M., Stedman, W., Anantharaman, T., Hastie, A., Dai, H., Fritz, M.H., Cao, H., Cohain, A., Deikus, G., Durrett, R.E., Blanchard, S.C., Altman, R., Chin, C.S., Guo, Y., Paxinos, E.E., Korbel, J.O., Darnell, R.B., McCombie, W.R., Kwok, P.Y., Mason, C.E., Schadt, E.E., Bashir, A. (2015) Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat. Methods*, 12, 780-786.

Philippe, R., Paux, E., Bertin, I., Sourdille, P., Choulet, F., Laugier, C., Šimková, H., Šafář, J., Bellec, A., Vautrin, S., Frenkel, Z., Cattonaro, F., Magni, F., Scalabrin, S., Martis, M.M., Mayer, K.F., Korol, A., Berges, H., Doležel, J., and Feuillet, C. (2013) A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. *Genome Biol*, 14, R64.

Poursarebani, N., Nussbaumer, T., Šimková, H., Šafář, J., Witsenboer, H., van Oeveren, J., Doležel, J., Mayer, K.F., Stein, N., Schnurbusch, T. (2014) Whole-genome profiling and shotgun sequencing delivers an anchored, gene-decorated, physical map assembly of bread wheat chromosome 6A. *Plant J*. 79, 334-347.

Ruperao, P., Chan, C.K., Azam, S., Karafiátová, M., Hayashi, S., Čížková, J., Saxena, R.K., Šimková, H., Song, C., Vrána, J., Chitikineni, A., Visendi, P., Gaur, P.M., Millán, T., Singh, K.B., Taran, B., Wang, J., Batley, J., Doležel, J., Varshney, R.K., Edwards, D. (2014) A chromosomal genomics approach to assess and validate the desi and kabuli draft chickpea genome assemblies. *Plant Biotechnol J*. 12, 778-786.

Šafář, J., Noa-Carrazana, J.C., Vrána, J., Bartoš, J., Alkhimova, O., Lheureux, F., Šimková, H., Caruana, M.L., Doležel, J., Piffanelli, P. (2004) Creation of a BAC resource to study the structure and evolution of the banana (*Musa balbisiana*) genome. *Genome*, 47, 1182–1191.

Šafář, J., Šimková, H., Kubaláková, M., Číhalíková, J., Suchánková, P., Bartoš, J., Doležel, J. (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet. Genome Res*. 129, 211-223.

Shearer, L.A., Anderson, L.K., de Jong, H., Smit, S., Goicoechea, J.L., Roe, B.A., Hua, A., Giovannoni, J.J., Stack, S.M. (2014) Fluorescence in situ hybridization and optical mapping to correct scaffold arrangement in the tomato genome. *G3 (Bethesda)*, 4, 1395-1405.

Šimková, H., Číhalíková, J., Vrána, J., Lysák, M.A., Doležel, J. (2003) Preparation of high molecular weight DNA from plant nuclei and chromosomes isolated from root tips. *Biol. Plantarum*, 46, 369-373.

Šimková, H., Šafář, J., Kubaláková, M., Suchánková, P., Číhalíková, J., Robert-Quatre, H., Azhaguvel, P., Weng, Y., Peng, J., Lapitan, N.L., Ma, Y., You, F.M., Luo, M.C., Bartoš, J., Doležel, J. (2011) BAC libraries from wheat chromosome 7D: efficient tool for positional cloning of aphid resistance genes. *J. Biomed. Biotechnol*. 2011, 302543.

Soderlund, C., Humphray, S., Dunham, A., French, L. (2000) Contigs built with fingerprints, markers, and FPC V4.7. *Genome Research*, 10, 1772-1787.

Teague, B., Waterman, M.S., Goldstein, S., Potamousis, K., Zhou, S., Reslewic, S., Sarkar, D., Valouev, A., Churas, C., Kidd, J.M., Kohn, S., Runnheim, R., Lamers, C., Forrest, D., Newton, M.A., Eichler, E.E., Kent-First, M., Surti, U., Livny, M., Schwartz, D.C. (2010) High-resolution human genome structure by single-molecule analysis. *Proc. Natl. Acad. Sci. USA*, 107, 10848-10853.

Tiwari, V.K., Riera-Lizarazu, O., Gunn, H.L., Lopez, K., Iqbal, M.J., Kianian, S.F., Leonard, J.M. (2012) Endosperm tolerance of paternal aneuploidy allows radiation hybrid mapping of the wheat D-genome and a measure of γ ray-induces chromosome breaks. *PLoS One*, 7, e48815.

Vrána, J., Kubaláková, M., Číhalíková, J., Valárik, M., Doležel, J. (2015) Preparation of sub-genomic fractions enriched for particular chromosomes in polyploid wheat. *Biol. Plantarum*, 59, 445-455.

Young, N.D., Debellé, F., Oldroyd, G.E., Geurts, R., Cannon, S.B., Udvardi, M.K., Benedito, V.A., Mayer, K.F., Gouzy, J., Schoof, H., Van de Peer, Y., Proost, S., Cook, D.R., Meyers, B.C., Spannagl, M., Cheung, F., De Mita, S., Krishnakumar, V., Gundlach, H., Zhou, S., Mudge, J., Bharti, A.K., Murray, J.D., Naoumkina, M.A., Rosen, B., Silverstein, K.A., Tang, H., Rombauts, S., Zhao, P.X., Zhou, P., Barbe, V., Bardou, P., Bechner, M., Bellec, A., Berger, A., Bergès, H., Bidwell, S., Bisseling, T., Choisne, N., Couloux, A., Denny, R., Deshpande, S., Dai, X., Doyle, J.J., Dudez, A.M., Farmer, A.D., Fouteau, S., Franken, C., Gibelin, C., Gish, J., Goldstein, S., González, A.J., Green, P.J., Hallab, A., Hartog, M., Hua, A., Humphray, S.J., Jeong, D.H., Jing, Y., Jöcker, A., Kenton, S.M., Kim, D.J., Klee, K., Lai, H., Lang, C., Lin, S., Macmil,

S.L., Magdelenat, G., Matthews, L., McCorrison, J., Monaghan, E.L., Mun, J.H., Najar, F.Z., Nicholson, C., Noirot, C., O'Bleness, M., Paule, C.R., Poulain, J., Prion, F., Qin, B., Qu, C., Retzel, E.F., Riddle, C., Sallet, E., Samain, S., Samson, N., Sanders, I., Saurat, O., Scarpelli, C., Schiex, T., Segurens, B., Severin, A.J., Sherrier, D.J., Shi, R., Sims, S., Singer, S.R., Sinharoy, S., Sterck, L., Viollet, A., Wang, B.B., Wang, K., Wang, M., Wang, X., Warfsmann, J., Weissenbach, J., White, D.D., White, J.D., Wiley, G.B., Wincker, P., Xing, Y., Yang, L., Yao, Z., Ying, F., Zhai, J., Zhou, L., Zuber, A., Dénarié, J., Dixon, R.A., May, G.D., Schwartz, D.C., Rogers, J., Quétier, F., Town, C.D., Roe, B.A. (2011) The Medicago genome provides insight into the evolution of rhizobial symbioses. *Nature*, 480, 520-524.

Zarrei, M., MacDonald, J.R., Merico, D., Scherer, S.W. (2015) A copy number variation map of the human genome. *Nat. Rev. Genet*. 16, 172–183.

Zhang, J., Li, C., Zhou, Q., Zhang, G. (2015) Improving the ostrich genome assembly using optical mapping data. *GigaScience*, 4, 24.

Zhou, S., Schwartz, D.C. (2004) The optical mapping of microbial genomes. *ASM News*, 70, 323-330.

Zhou, S., Bechner, M.C., Place, M., Churas, C.P., Pape, L., Leong, S.A., Runnheim, R., Forrest, D.K., Goldstein, S., Livny, M., Schwartz, D.C. (2007) Validation of rice genome sequence by optical mapping. *BMC Genomics*, 8, 278.

Zhou, S., Wei, F., Nguyen, J., Bechner, M., Potamousis, K., Goldstein, S., Pape, L., Mehan, M.R., Churas, C., Pasternak, S., Forrest, D.K., Wise, R., Ware, D., Wing, R.A., Waterman, M.S., Livny, M., Schwartz, D.C. (2009) Single Molecule Scaffold for the Maize Genome. *PLoS Genet*. 5, e1000711.

# Tables

**Table 1. Data collection and assembly statistics**

|  | No. molecules/ genome maps | Total length | 7DS arm coverage | Molecule/ mapN50 | Longest molecule/map |
|---|---|---|---|---|---|
| **Single molecules (>150 kb)** | 209,788 | 68.8 Gb | 180x | 354 kb | 2.1 Mb |
| **Map assembly** | 371 | 350 Mb | 0.92x | 1.3 Mb | 4.6 Mb |

**Table 2: Assignment of 30-, 60-, and 90-kb sequences to 7DS genome maps**

| Sequence length | No. sequences | No. correctly assigned | Percentage correctly assigned | Lowest confidence Highest confidence | No. labels[1] |
|---|---|---|---|---|---|
| **30 kb** | 100 | 12 | 12% | 5.86 | 5 |
|  |  |  |  | 7.60 | 6 |
| **60 kb** | 70 | 57 | 81 % | 6.07 | 5 |
|  |  |  |  | 13.87 | 10 |
| **90 kb** | 40 | 40 | 100 % | 7.91 | 6 |
|  |  |  |  | 19.99 | 14 |

[1] No. labels corresponds to number of distinguishable *Nt.BspQI* recognition sites in the sequence

# Figure legends

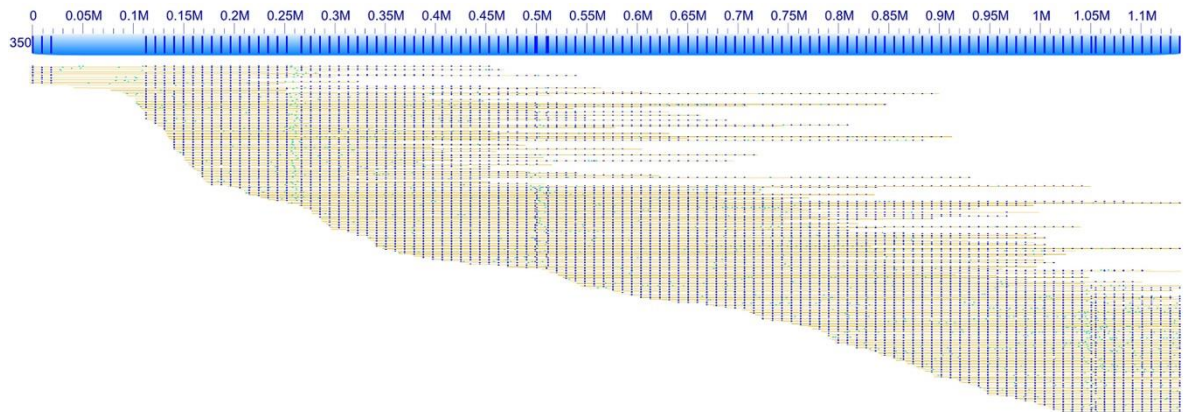**Figure 1. Genome map No. 350 comprising a long array of tandem repeats**

The pile of single molecules, depicted as yellow lines with blue and green dots corresponding to mapped and unmapped labels, respectively, was a source for building the consensus genome map (blue bar). The regular labelling pattern indicates presence of tandem repeats.

**Figure 2. Scaffolding and correcting physical map contigs based on the genome map No. 19.**

(a) In total, nine contigs of the physical map (black bars) could be anchored through sequences of constituting BAC clones (blue lines) to the genome map No. 19 (green bar). Short red bars indicate approximate positions of *Ae. tauschii* SNP markers anchored to particular clones. The purple line and bar in ctg3865 represent clone TaaCsp067A19, which was incorrectly assigned to this contig. Detail is shown in (b). A *cmap* of the clone TaaCsp067A19 does not match the corresponding region in GM19. The green bar corresponds to GM19, while the blue bars represent *in silico* digested BAC sequences (*cmaps*).

**Figure 3. Local integration of three genetic maps through the genome map No. 15**

Contigs of the physical map (black bars) were aligned to the genome map No. 15 (green bar) through sequences of constituting BAC clones (blue bars). The BAC contigs carry genetic markers (red) originating from three genetic maps, which could be integrated through the genome map.

**Figure 4. Merging genome maps**

Three genome maps (green bars) could be merged together after aligning sequences of BAC clones (blue bars) from three contigs of the physical map (black bars).

# Figures



**Figure 1. Genome map No. 350 comprising a long array of tandem repeats**
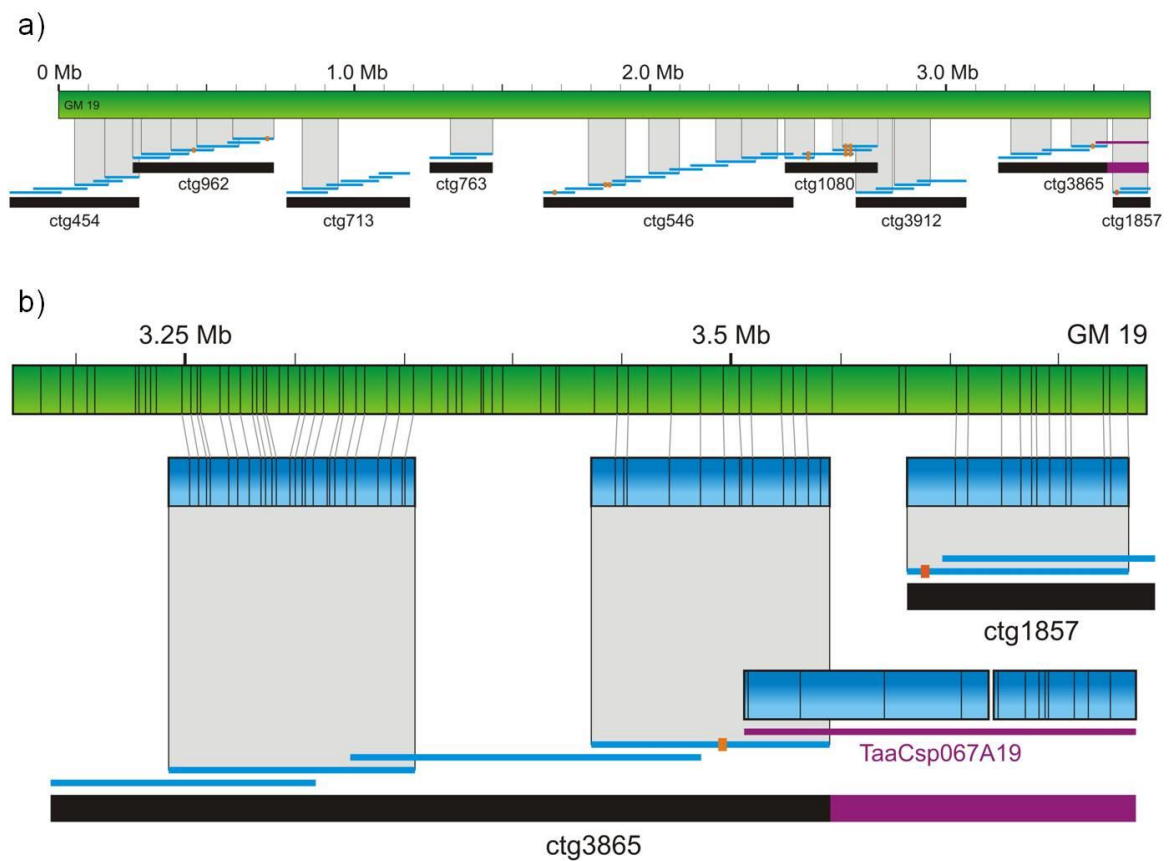


**Figure 2. Scaffolding and correcting physical map contigs based on the genome map No. 19.**
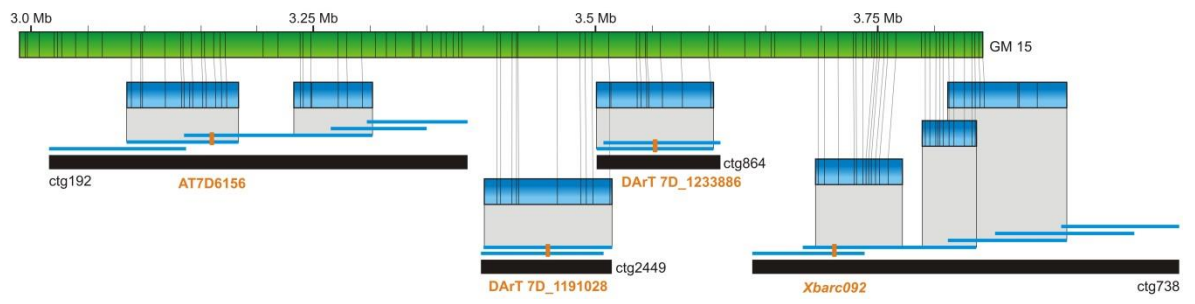
**Figure 3. Local integration of three genetic maps through the genome map No. 15**
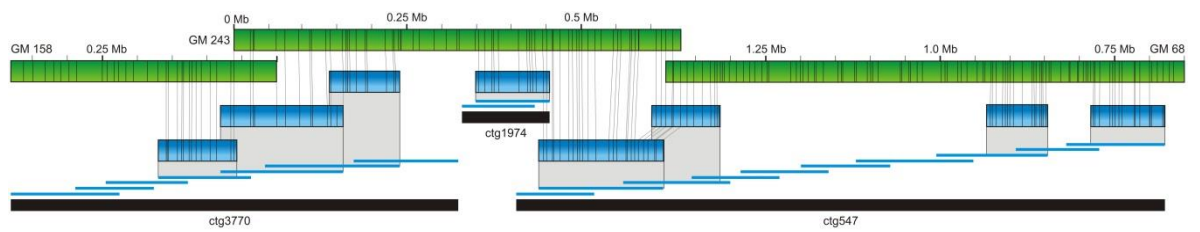


**Figure 4. Merging genome maps**

# APPENDIX II

**Chromosomal genomics facilitates fine mapping of a Russian wheat aphid resistance gene**

Staňková H, Valárik M, Lapitan NL, Berkman PJ, Batley J, Edwards D, Luo MC, Tulpová Z, Kubaláková M, Stein N, Doležel J, Šimková H

CrossMark

ORIGINAL PAPER

# Chromosomal genomics facilitates fine mapping of a Russian wheat aphid resistance gene

Helena Staňková[1] · Miroslav Valárik[1] · Nora L. V. Lapitan[2,3] · Paul J. Berkman[4] ·
Jacqueline Batley[5] · David Edwards[5,6] · Ming-Cheng Luo[7] · Zuzana Tulpová[1] ·
Marie Kubaláková[1] · Nils Stein[8] · Jaroslav Doležel[1] · Hana Šimková[1]

## Abstract

*Key message* **Making use of wheat chromosomal resources, we developed 11 gene-associated markers for the region of interest, which allowed reducing gene interval and spanning it by four BAC clones.**

*Abstract* Positional gene cloning and targeted marker development in bread wheat are hampered by high complexity and polyploidy of its nuclear genome. Aiming to clone a Russian wheat aphid resistance gene *Dn2401*

located on wheat chromosome arm 7DS, we have developed a strategy overcoming problems due to polyploidy and enabling efficient development of gene-associated markers from the region of interest. We employed information gathered by GenomeZipper, a synteny-based tool combining sequence data of rice, *Brachypodium*, sorghum and barley, and took advantage of a high-density linkage map of *Aegilops tauschii*. To ensure genome- and locus-specificity of markers, we made use of survey sequence assemblies of isolated wheat chromosomes 7A, 7B and 7D. Despite the low level of polymorphism of the wheat D subgenome, our approach allowed us to add in an efficient and cost-effective manner 11 new gene-associated markers in the *Dn2401* region and narrow down the target interval to 0.83 cM. Screening 7DS-specific BAC library with the flanking markers revealed a contig of four BAC clones that span the *Dn2401* region in wheat cultivar 'Chinese Spring'. With the availability of sequence assemblies and GenomeZippers for each of the wheat chromosome arms, the proposed strategy can be applied for focused marker development in any region of the wheat genome.

Communicated by H.-Q. Ling.

✉ Hana Šimková
simkovah@ueb.cas.cz

1 Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, 783 71 Olomouc, Czech Republic

2 Department of Soil and Crop Sciences, Colorado State University, Fort Collins, CO 80523, USA

3 USAID/Bureau for Food Security, 1300 Pennsylvania Ave. NW, Washington, DC 20523, USA

4 CSIRO Plant Industry, Brisbane, QLD 4072, Australia

5 School of Plant Biology, University of Western Australia, Perth, WA 6009, Australia

6 Institute of Agriculture, University of Western Australia, Perth, WA 6009, Australia

7 Department of Plant Sciences, University of California, Davis, CA 95616, USA

8 Department Genebank, AG Genome Diversity, Leibniz Institute of Plant Genetics and Crop Plant Research, Corrensstr. 3, Seeland (OT), 06466 Gatersleben, Germany

## Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most important crops providing staple food for 35 % of the world's population. Its annual production of 713 million tons (FAOSTAT 2013) is topped only by rice and maize. Wheat yields are markedly reduced due to attack of a large variety of pests and diseases causing significant economic losses.

Russian wheat aphid, RWA (*Diuraphis noxia* Kurdjumov) is globally a major invasive pest affecting wheat, as well as barley (Morrison and Peairs 1998). Numerous *D.*

*noxia* strains (biotypes) varying in virulence have spread in all wheat and barley growing areas with the exception of Australia (Lapitan et al. 2007). Besides causing chlorosis and reducing plant biomass and consequently grain yield (Burd and Burton 1992; Girma et al. 1993), aphid infestation introduces leaf rolling, which reduces the effectiveness of contact insecticides by sheltering the insect colonies. Thus, development of new varieties carrying genes controlling resistance to the pest is the most efficient, economical and environmentally safe approach to reducing the damage. Supplementation of conventional wheat breeding approaches for pest resistance with marker-assisted selection and, potentially, direct gene transfer by molecular methods, promises to enhance the efficiency of the breeding process. Prerequisites of this approach are saturation of genetic maps in the region of interest and, ideally, identification of the gene underlying the resistance.

Numerous RWA resistance genes have been identified in various cultivars and lines of wheat, rye and *Aegilops tauschii* (Nkongolo et al. 1991; Marais et al. 1998; Martin et al. 2001; Liu et al. 2001, 2002; Peng et al. 2009), but only a few of them, namely *Dn7*, *Dn626580* and *Dn2401* (Lapitan et al. 2007; Valdez et al. 2012; Qureshi et al. 2006; Fazel-Najafabadi et al. 2014), confer resistance to the highly virulent US RWA biotype 2, the prevailing biotype in the United States. In the present work, we made efforts towards cloning the *Dn2401* gene located in the central part of the short arm of chromosome 7D (7DS). A previous study (Fazel-Najafabadi et al. 2014) identified several 7DS-located microsatellite markers linked with the trait, including *Xbarc214* and *Xgwm473*, delimiting an interval of 2.9 cM. In our work, we aimed to narrow down the *Dn2401* interval and potentially span it with a contig of BAC clones.

Marker development and mapping in hexaploid bread wheat ($2n = 6x = 42$) is highly challenging due to the complex nature of its genome, combining large size (~17 Gbp/1 C), high repeat content (>80 % repetitive sequences), and the presence of three homoeologous subgenomes A, B and D. Moreover, the level of polymorphism among elite wheat cultivars, especially within the D subgenome, the youngest of the subgenomes, is limited (Chao et al. 2009; Berkman et al. 2013).

Previously, positional cloning projects took advantage of mapping in diploid wheat ancestors and close relatives to avoid problems due to polyploidy. Nevertheless, mapping in diploid species is often complicated by inconsistencies between genetic maps of the hexaploid and the diploid and/or absence of the gene of interest from the diploid genome. This is also the case with *Dn2401*. In such situation, mapping in the hexaploid becomes a necessity. In the past decade, the major source of information for targeted marker development aiming to saturate a narrow region was sequence information of wheat expressed sequence

tags (ESTs) and colinearity between wheat and sequenced genomes of model grass species like rice and *Brachypodium* (Keller et al. 2005; Qin et al. 2011; Terracciano et al. 2013). However, the efficiency of designing locus-specific primers based on ESTs was rather low due to insufficient specificity, accentuated in polyploid genomes, and lack of wheat sequence information that would reveal genomic context of the ESTs (Schnurbusch et al. 2007; Terracciano et al. 2013).

In the past two decades, chromosomal genomics has been developed as a strategy to tackle problems associated with genome complexity and polyploidy in wheat (Doležel et al. 2012). The approach is based on dissecting large genomes into particular chromosomes or chromosome arms by flow-cytometric sorting and has also been applied to other Triticeae species like barley (Lysák et al. 1999), rye (Kubaláková et al. 2003) and *Aegilops* sp. (Molnár et al. 2011). DNA of flow-sorted chromosomes is of high quality, which enabled the construction of a set of wheat chromosome (arm)-specific BAC libraries (Šafář et al. 2010; http://olomouc.ueb.cas.cz/). The International Wheat Genome Sequencing Consortium (IWGSC, http://www.wheat-genome.org/) adapted the chromosomal approach employing chromosome-specific physical maps as the major strategy towards obtaining a complete reference sequence of the bread wheat genome. Moreover, combining next-generation sequencing technologies with the flow sorting enabled to obtain survey sequences of individual chromosomes (arms) of barley (Mayer et al. 2009, 2011), wheat (IWGSC 2014) as well as rye (Martis et al. 2013). The generated barley chromosomal sequences were exemplarily used to establish so-called 'GenomeZipper', a colinearity-based tool comprising of an ordered set of *Brachypodium*, rice and sorghum genes anchored to a backbone of linkage-mapped barley ESTs (Close et al. 2009), the corresponding barley 454 sequence reads, non-mapped barley ESTs and barley full-length cDNAs (Mayer et al. 2011). Implying virtual gene order in barley, GenomeZipper promises to be a useful tool for targeted development of gene-derived markers from desired regions of the genome.

Application of these chromosomal resources, including a 7DS-specific BAC library and physical map (Šimková et al. 2011), survey sequence assemblies of the 7AS, 7BS and 7DS chromosome arms (Berkman et al. 2013; IWGSC 2014), and barley GenomeZipper (Mayer et al. 2011), together with a new high-density linkage map of *Ae. tauschii* (Luo et al. 2013), the progenitor of the D subgenome, enabled us to develop an efficient approach for targeted identification of gene-associated markers from the region of interest and to place 11 new markers around the *Dn2401* gene. This aided reducing interval delimited by flanking markers and identifying BAC contig spanning the region of interest.

## Materials and methods

### Plant material

Genetic mapping was performed in 184 F2 plants derived from a cross between RWA resistant winter wheat line CI2401 and susceptible cultivar 'Glupro'. To identify BAC clones and contigs bearing markers linked to the *Dn2401* gene, a BAC library specific for the 7DS chromosome arm of cv. 'Chinese Spring' (Šimková et al. 2011) and a 7DS-specific physical contig map (https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_7DS_v1/) were utilized.

7AS, 7BS and 7DS chromosome arms were isolated from double-ditelosomic lines of wheat cv. 'Chinese Spring' carrying arms of chromosomes 7A, 7B and 7D as telocentric chromosomes. Seeds for these lines were kindly provided by Prof. Bikram Gill (Kansas State University, Manhattan, USA) and Prof. Adam Lukaszewski (University of California, Riverside, USA).

### Preparation of chromosome-specific DNA

7AS, 7BS and 7DS chromosome arms were isolated from the double-ditelosomic lines by flow-cytometric sorting as described in Kubaláková et al. (2002). Chromosomal DNA was amplified by isothermal multiple displacement amplification (MDA) as described in Šimková et al. (2008).

### Scoring RWA response

RWA screening was done at the Colorado State University Insectary greenhouse under ambient conditions (14 h, ~25.5 °C days with light intensities between 1100 and 1400 μM m$^2$ s$^{-1}$ and 10 h, ~20 °C nights). A randomized complete block design with three replications was used. Twelve to fifteen seeds of each F2 derived F3 family (F$_{2:3}$) were planted. Each tray contained the parents as well as 'Gamtoos- R' (containing the *Dn7* resistance gene) and 'Yuma' which served as resistant and susceptible controls, respectively. 7-day-old seedlings were infested with RWA biotype 2 by scattering aphids evenly across the trays. Each plant was scored for leaf rolling and chlorosis at 7 and 10 days post infestation, as previously described (Collins et al. 2005). Chlorosis scores ranged from 1 (=healthy plants with small hypersensitive lesions) to 9 (=dead or unrecoverable), while leaf rolling scores were on a scale of 1 (=completely flat leaves) to 4 (=tightly rolled leaves with leaf trapping) (Collins et al. 2005). Plants with chlorosis scores of ≤4 and leaf rolling scores ≤2 were considered resistant and those exhibiting chlorosis ≥5 and leaf rolling ≥3 were considered susceptible. The F2 individuals were assigned a phenotypic class (homozygous resistant, heterozygous, or homozygous susceptible) based on the number of resistant and susceptible individuals observed in the F$_{2:3}$ families, under RWA infestation. Homozygous resistant and susceptible designations were only given to F2 individuals when all (or all but one) of the plants in the F$_{2:3}$ family were scored with the same corresponding designation.

### Mapping of SSR markers

All 184 plants of the mapping population were scored for SSR markers *Xcfd14*, *Xcfd68* and *Xgwm473* using primer sequences and PCR conditions given in GrainGenes 2.0 (http://wheat.pw.usda.gov/). Marker *Xcfd14* was separated in 6 % non-denaturing polyacrylamide gel (19:1 acrylamide:bisacrylamide) and visualized by ethidium bromide staining. Genotyping of markers *Xcfd68* and *Xgwm473* was performed by PCR amplification using fluorescently labeled primers and subsequent separation of the PCR products by ABI 3730xl DNA analyzer (Applied Biosystems). Data were analyzed by GeneMarker® software (SoftGenetics).

### Marker development strategy

Markers specific for the *Dn2401* gene region were developed by two approaches, both employing colinearity with related grass genomes (scheme of the procedure shown in Supplementary File 1).

Since the available flanking markers were microsatellites, the procedure started with identification of genes delimiting the *Dn2401* region. BAC pools of the 7DS-specific BAC library were PCR screened with flanking microsatellite markers *Xcfd68* and *Xgwm473* and corresponding BAC clones as well as BAC contigs were identified (Šimková et al. 2011). The *Xcfd68* marker was used in our study instead of the closest marker *Xbarc214*, located 0.3 cM proximal to *Xcfd68* (Fazel-Najafabadi et al. 2014), because of easier and more reliable scoring. Two BAC clones from each of the contigs identified by the above markers were sequenced using Roche/454 technology and the sequence reads were assembled using Newbler software package. Genes adjacent to both *Xcfd68* and *Xgwm473* were identified and used to initially delimit syntenic regions in sequenced grass genomes.

The first approach to marker development was based on information provided by the barley GenomeZipper (Mayer et al. 2011). After identifying the syntenic region, sequences of relevant barley ESTs were retrieved from the 'HarvEST: Barley' database (Close et al. 2007), version 1.77. These were compared with assembled 7AS, 7BS and 7DS sequences of cv. 'Chinese Spring' (Berkman et al. 2013; http://www.wheatgenome.info/) using BLASTn algorithm. Genetically mapped barley ESTs located at

syntenic positions in *Brachypodium*, rice and sorghum were selected initially, followed by ESTs only inferred by synteny. Wheat genomic sequences corresponding to 7AS, 7BS and 7DS contigs were identified based on sequence identity with barley ESTs. Contigs from each subgenome were aligned and annotated using GeniousPro software, version 5.4.4 (Drummond et al. 2010) using the Genious or Muscle alignment software with default parameters. Sequences annotated with marked exons and introns were the basis for designing a first set of 7DS-specific primers.

An alternative approach employing synteny between wheat and *Ae. tauschii* was applied to design a second set of markers using a gene-based linkage map of *Ae. tauschii* that included information of colinear rice, *Brachypodium* and sorghum genes (Luo et al. 2013). Sequences of markers from orthologous region of *Ae. tauschii* were compared with the 7AS, 7BS and 7DS assemblies using BLASTn, and the sequence contigs thus identified were aligned as described above. Exons and introns were determined by comparison with rice genes using BLASTn.

### Primer design and identification of polymorphisms

Aligned and annotated sequence contigs from 7AS, 7BS and 7DS were used to design subgenome-specific primers for the region of interest. To avoid amplification from conserved gene domains, primers were preferentially designed in non-coding regions. Primer3 v.0.4.0 software was used for the primer design with default parameters (Rozen and Skaletsky 2000). To verify the specificity of the PCR primers, DNA of flow-sorted chromosome arms 7AS, 7BS and 7DS was used as PCR template. PCR reactions were carried out in 20-µl volume comprising 10 ng of template DNA, 1× PCR buffer (10 mM Tris-HCl, pH 8.8 at 25 °C, 1.5 mM MgCl$_2$, 50 mM KCl, 0.1 % Triton X-100), 1 µM primers, 200 µM dNTPs and 1 U of DyNAzyme DNA Polymerase (Finnzymes, Espoo, Finland). The amplification was performed under the following conditions: initial denaturation at 94 °C for 5 min followed by 35 cycles of 94 °C for 30 s, 58 °C for 30 s, 72 °C for 1 min and final extension at 72 °C for 10 min. PCR products were separated in 4 % non-denaturing polyacrylamide gel and visualised by ethidium bromide staining. PCR conditions (annealing temperature and cycle number) were optimized to obtain a single 7DS-specific product. Primer pairs that proved to be 7DS-specific were applied to amplify DNA of the mapping population parents, CI2401 and 'Glupro'. PCR was performed under optimized conditions with 20 ng of template DNA. The amplification products were cleaned up by Exonuclease I/FastAP™ Thermosensitive Alkaline Phosphatase (Fermentas, Burlington, Canada) treatment and cycle sequencing was performed using a BigDye® Terminator v3.1 Cycle Sequencing kit (Applied Biosystems,

Foster City, USA) following manufacturer's instructions. Sequencing products were purified using an Agencourt Clean SEQ kit (Beckman Coulter, Beverly, MA) and analyzed on an ABI 3730xl DNA analyzer (Applied Biosystems). Sequences were aligned to identify polymorphisms between the parents using MEGA 4 software and ClustalW with default parameters. To verify the locus specificity of the primers, the parental sequences were aligned to the 'Chinese Spring' 7DS survey sequence (Berkman et al. 2011).

### Linkage mapping and integration with the physical map

Mapping of the identified SNP markers was done by amplicon sequencing on 184 plants of the CI2401/'Glupro' F2 mapping population. PCR products from individual F2 plants were purified, sequenced and analyzed as described above. Marker *Xowm711* showing length polymorphism in the amplicon was scored after separation in 4 % polyacrylamide gel. Genotyping data were analyzed using Join-Map 4.0 (Van Ooijen and Voorrips 2001). Genetic distances between markers were calculated applying the Kosambi function (Kosambi 1944). The markers were then integrated into the existing microsatellite genetic map covering the *Dn2401* gene region (Fazel-Najafabadi et al. 2014). Aiming to integrate genetic and physical maps of the region, the mapped markers were used to screen the 7DS-specific BAC library as described in Šimková et al. (2011) and contigs from the 7DS physical map were identified through positive BAC clones.

## Results

### Delimiting syntenic regions

In order to apply a synteny-based approach to marker development, we had to identify genic sequences adjacent to the available microsatellite markers and through them delimit syntenic regions in genomes of relative species.

Screening of the 7DS-specific BAC library with *Dn2401*-flanking SSR markers *Xcfd68* and *Xgwm473* and sequencing of BAC clones from marker-comprising contigs revealed genes corresponding to barley ESTs that enabled identification of the syntenic region in the barley GenomeZipper. This region was delimited by barley EST 40502 and the genes Bradi1g43560.1, Os06g0284200 and Sb10g010460.1 on the distal side, and barley EST 7580 and Bradi3g39810.1, Os08g0503800 and Sb07g028020.1 on the proximal side, and contained 255 predicted genes. In addition, a 2-cM region proximal of *Xgwm473* was included in the study. The *Dn2401* region comprised a

previously described breakpoint of colinearity between the barley, as well as wheat genome and genomes of *Brachypodium*, rice and sorghum, respectively (Sorrells et al. 2003; Mayer et al. 2011), combining parts of *Brachypodium* chromosomes 1 and 3, rice chromosomes 6 and 8 and sorghum chromosomes 7 and 10.

Based on genetic positions of the first batch of synteny-derived SNP markers we could more precisely identify the syntenic region in *Ae. tauschii*. In subsequent experiments we focused on a region distal of the colinearity breakpoint delimited by markers AT7D6468 and AT7D6420 (Luo et al. 2013), which is colinear to a part of rice chromosome 6, *Brachypodium* chromosome 1 and sorghum chromosome 10.

### Development of locus-specific primers

Barley ESTs and *Ae. tauschii* marker sequences from the selected syntenic regions were used to identify homologous sequences within wheat 7AS, 7BS and 7DS sequence assemblies. In total, 11 ESTs from the GenomeZipper and 17 marker sequences from the *Ae. tauschii* linkage map were selected for marker development. The respective genic sequences for the three homoeologs, if available, were aligned, annotated and used as a template for designing 7DS-specific primers based on homoeolog sequence variants (Fig. 1a). From the total of 161 primer pairs, 114 primer pairs had at least one primer designed as 7DS-specific based on known sequence information from all three homoeologs (Table 1). For 86 out of these 114 primer pairs, both forward and reverse primers were designed to be 7DS-specific. In the remaining 28 cases, only one primer could be designed as a 7DS homoeolog specific variant (HSV), either due to missing sequence information from 7AS and 7BS (25 primer pairs) or due to a lack of polymorphism between the subgenomes on one side of the locus (3 primer pairs). Out of the 114 7DS-specific primer pairs, 99 produced a single 7DS-specific PCR amplification product, a success rate of 87 %. Further 47 primer pairs were designed with incomplete knowledge of homoeolog sequence. In these cases, 39 (83 %) produced a 7DS-specific PCR amplification product (Table 1). The relatively high success rate in cases of missing homoeolog sequence information reflects the fact that most of the primers were designed in gene-flanking regions which show relatively low level of conservation between A, B and D subgenomes.

In total, out of 161 primer pairs, 138 (listed in Suppl. File 2) provided 7DS-specific product using DNA from flow-sorted 7AS, 7BS and 7DS arms as a template, but only 15 primer pairs provided the specific product when using the annealing temperature specified by the Primer3 software. In the other 123 cases, PCR conditions had to be optimized by increasing the annealing temperature (Fig. 1b) and/or decreasing number of PCR cycles (data not shown). Eight out of 161 primer pairs amplified the locus from all three
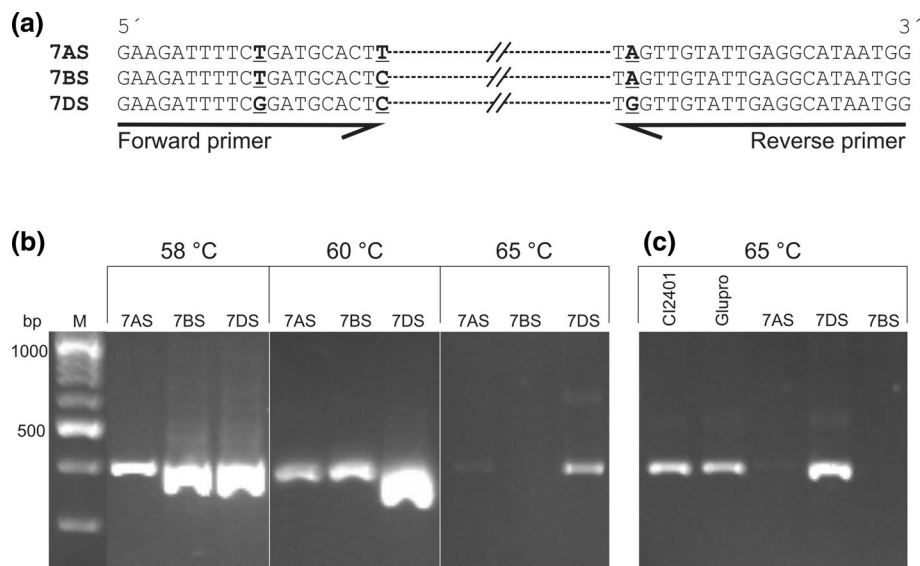


**Fig. 1** Development of genome-specific markers. **a** 7DS-specific primers were designed based on HSVs between 7AS, 7BS and 7DS chromosome arms (HSVs are *underlined*). **b** PCR was optimized on Phi29-amplified DNA of isolated chromosome arms 7AS, 7BS and 7DS, respectively, aiming to obtain 7DS-specific product only. The increase in annealing temperature (indicated above the *gel images*) was essential for most loci. **c** PCR was run on CI2401 and cv. 'Glupro' at the optimal annealing temperature. Resulting PCR products were sequenced and aligned to reveal polymorphisms between the parents of the mapping population. The figure shows the procedure for locus STS94

**Table 1** The efficiency of development of locus-specific primers

| | Complete knowledge of homoeolog sequences[a] | Incomplete knowledge of homoeolog sequences[b] | Overall |
|---|---|---|---|
| | Number of primer pairs | Number of primer pairs | Number of primer pairs |
| Total number of primer pairs | 114 | 47 | 161 |
| Single 7DS-specific product | 99 | 39 | 138 |
| Amplification from more subgenomes | 4 | 4 | 8 |
| Amplification from more loci | 6 | 2 | 8 |
| No product | 5 | 2 | 7 |

[a] At least one primer designed as 7DS-specific based on the knowledge of 7AS, 7BS and 7DS sequences

[b] Primer designing without knowing 7AS and/or both 7AS and 7BS sequences

subgenomes, while eight primer pairs produced more than one PCR product for each of the 7S arms. In seven cases, no product was obtained. Overall, we obtained a single 7DS-specific PCR product with 86 % (138) of all designed primers. Out of 138 primer pairs successfully tested on the flow-sorted chromosome arms originating from 'Chinese Spring', only one failed to provide a product in the parents of the mapping population (CI2401, 'Glupro').

**Analysis of polymorphisms**

All 7DS-specific PCR products from both parental DNAs (CI2401, 'Glupro') were sequenced (Suppl. File 3). The sequencing confirmed locus-specificity of the PCR products. A total of 73,070 bp were sequenced for both CI2401 and 'Glupro'. Comparison of parental sequences revealed 11 single-nucleotide polymorphisms between the parents, of which two were present in one PCR product. One of these two SNPs, representing an indel in a [T]$^n$ microsatellite, was excluded from the mapping. The remaining ten SNP markers were designated *Xowm701–Xowm710* (Table 2). The last marker (*Xowm711*) was a 43-bp indel between CI2401 and 'Glupro'. Based on the sequencing data, we calculated the frequency of polymorphisms between the CI2401 line and 'Glupro', as one polymorphism (SNP/indel) per 6,089 bp. Out of the 11 SNPs, four were found in introns and six in gene-flanking regions, just as the long indel. Only one SNP was identified in an exon, reflecting the fact that exon regions show a higher level of sequence conservation compared to introns and intergenic regions.

**Mapping new markers**

Mapping the *Xowm701–Xowm711* markers resulted in six clusters comprising one to three co-segregating markers each (Fig. 2). All markers mapped in the *Dn2401* gene region. Genotypes of all markers across the mapping population are stated in Suppl. File 4, segregation

ratios in Suppl. File 5. The first two markers, *Xowm701* and *Xowm702*, mapped outside the region delimited by SSR markers *Xcfd68* and *Xgwm473*, 0.28 cM proximal to *Xgwm473*. In case of the *Xowm701*, this fitted well with the position of the barley ortholog in the GenomeZipper, whereas a more distal position was expected for marker *Xowm702*. The remaining nine markers (*Xowm703–Xowm711*) were located within the region delimited by *Xcfd68* and *Xgwm473*. The order of *Ae. tauschii*-derived markers *Xowm704–Xowm711* in wheat fully corresponded with the order of their orthologs in *Ae. tauschii*. The data indicate that both the barley GenomeZipper and the *Ae. tauschii* genetic map are valuable resources for targeted marker development in the region of interest in wheat. Mapping the newly developed markers resulted in the reduction of the *Dn2401* interval from 2.47 cM, delimited by *Xcfd68* and *Xgwm473*, to 0.83 cM, delimited by *Xowm711* on the distal side of *Dn2401*, and co-segregating markers *Xowm704*, *Xowm705* and *Xowm706* on the proximal side. This also aided a more precise determination of syntenic regions in wheat relatives, positioning the *Dn2401* interval distal of the colinearity breakpoint in rice chromosome 6, *Brachypodium* chromosome 1 and sorghum chromosome 10. The distance from the breakpoint is around 12.5 Mb in *Ae. tauschii*. In the barley GenomeZipper, the *Dn2401* interval comprises 31 predicted genes.

**Physical map of the *Dn2401* region**

7DS-specific BAC library of a cv. 'Chinese Spring' was screened with all genetically mapped markers aiming to estimate local physical/genetic distance ratio and, potentially, identify a BAC contig spanning the *Dn2401* region. Most of the markers located in separate contigs, which prevents the exact calculation of the ratio, but an estimate based on the size of the BAC contigs and the position of the individual markers suggested that on average, the genetic distance of 1 cM exceeded 3 Mbp in the studied region. In contrast to this observation, markers *Xowm705* and

**Table 2** Gene-associated markers mapped in the *Dn2401* gene region

| Marker | Polymor-phism | | Forward primer Reverse primer | PCR product size (bp) | Corresponding sequence in barley/*Aegilops tauschii*[a] | |
|--------|---------------|---|-------------------------------|------------------------|------------------------------------------------------------|---|
| *Xowm701* | SNP | T/C | ACGTAACCCTTCCACAATCG TGATTTCCCTTGACCTGCTT | 1305 | Barley EST (unigene name) | 20049 |
| *Xowm702* | SNP | C/T | AGAGGGGAACTTTCTTTAGAAGATG GAAGAAGTGGCAGTTTTCTCG | 617 | | 21318 |
| *Xowm703* | SNP | A/G | AGAGAATGCTCCGGATGAC GTTTCTGAAGCTAACAATCTATTACG | 630 | | 7349 |
| *Xowm704* | SNP | C/T | AAGCGGATCCCAGTAACGAT CGATGTAATCTCGGCGGAAC | 587 | *Aegilops tauschii* marker | AT7D6458 |
| *Xowm705* | SNP | A/T | TGGCAGGAAATGGAAGTTTT GAATGTTCAGAATTTCCTCCGTA | 394 | | AT7D6434 |
| *Xowm706* | SNP | A/- | TAGCACAGGATAAATACAACCACT AAGACGAATGCAGTAAATAACACA | 353 | | AT7D6440 |
| *Xowm707* | SNP | G/A | ATGGACATCTTCAATAGGAATGG GCTAAGTTCTTGAACATTTTCTCAT | 494 | | AT7D6421 |
| *Xowm708* | SNP | C/G | CTGTCTCTCCACAGCGTGTC CGCCCCTGCATGGTAGTA | 581 | | AT7D6420 |
| *Xowm709* | SNP | T/C | AGGTTGTCAACTTCCACTGC AAACTAATATAGGAGTGACTGAGCAAT | 553 | | AT7D6420 |
| *Xowm710* | SNP | A/G | TGACAATTTTGACAAGCCACA CGAAAGCATTTTGGAGCATT | 666 | | BE443936 |
| *Xowm711* | indel | | GCCGTGGTATGTGATTCTGTT ACAACGTTCATGGTCTGTGG | 584/541 | | AT7D6430 |

[a] Barley ESTs as in HarvEST: Barley database version 1.77, and *Ae. tauschii* markers as in Luo et al. (2013)

*Xowm711,* flanking the gene locus in the genetic distance of 0.83 cM, were found to allocate both to the same BAC contig (ctg783, https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_7DS_v1/) with an estimated distance of ~300 kb from each other represented by four BAC clones of minimum tilling path (Fig. 2). Thus the physical/genetic distance between these markers equaled ~360 kb per 1 cM. The high recombination rate around the gene made possible positioning of the trait in ~300 kb interval with a population of only 184 F2 individuals.

## Discussion

Gene-associated markers are important for plant molecular breeding. Their development in hexaploid wheat is hindered by the presence of gene homoeologs and paralogs. Thus, designing genome- and locus-specific primers is an essential step in the development of molecular markers. In our study, we applied a procedure exploiting shotgun-sequence information from isolated chromosome arms, which allowed developing locus-specific primer pairs with an efficiency of 86 %.

In previous studies employing synteny-based approaches for targeted marker development, markers were developed based on genome sequence and gene order in rice or *Brachypodium*, in combination with sequence information from wheat ESTs (Bossolini et al. 2006; Schnurbusch et al. 2007; Qin et al. 2011; Quraishi et al. 2009). As only conserved transcribed sequences were available, this approach suffered from a lack of specificity in the initial steps of locus-specific marker development. Additional steps had to be incorporated to increase marker specificity, including cloning and sequencing of primary PCR products to identify homoeolog specific sequences and to design locus-specific primers based on homoeolog sequence variants (HSVs). Such approach was used by Schnurbusch et al. (2007) who developed markers linked to the *Bo1* gene located on the wheat 7BL chromosome arm. In their work, design of a 7B-specific primer harboring at least one HSV was possible in 24 out of 41 tested primer pairs and a 7B-specific PCR product was obtained for 16 primer pairs. Overall, 7B-specific products could be obtained only for 39 % of the investigated loci.

The success rate achieved in our study was also significantly higher as compared to other studies on development of locus- or genome- specific markers in wheat. Terracciano et al. (2013) derived genome-specific markers to map *Lr14* gene on the 7BL chromosome arm of tetraploid durum wheat. Lacking sequence information from 7A and 7B chromosomes, the authors began by sequencing PCR products obtained from exon-derived primers using

**Fig. 2** Linkage and physical map of the *Dn2401* region and its synteny with *Ae. tauschii* and barley orthologous regions. Linkage map of the central part of wheat chromosome arm 7DS comprising the *Dn2401* locus and newly developed markers (*Xowm*) in the context of previously mapped SSR markers. Colinearity with orthologous loci in *Ae. tauschii* linkage map (7D$^t$S) and barley GenomeZipper (7HS) is indicated by *solid lines*. *Dashed lines* show colinearity for genes adjacent to flanking SSR markers. The *numbers* on the *left* indicate positions of the loci in the *Ae. tauschii* map, the GenomeZipper, and the wheat 7DS linkage map in cM units. *Asterisks* point out virtual positioning of the respective barley ESTs on the barley genetic map (HarvEST, version 1.77). Minimum tilling path of BAC contig 783 assigned to the *Dn2401* region with approximate positions of flanking markers is shown on the *right*
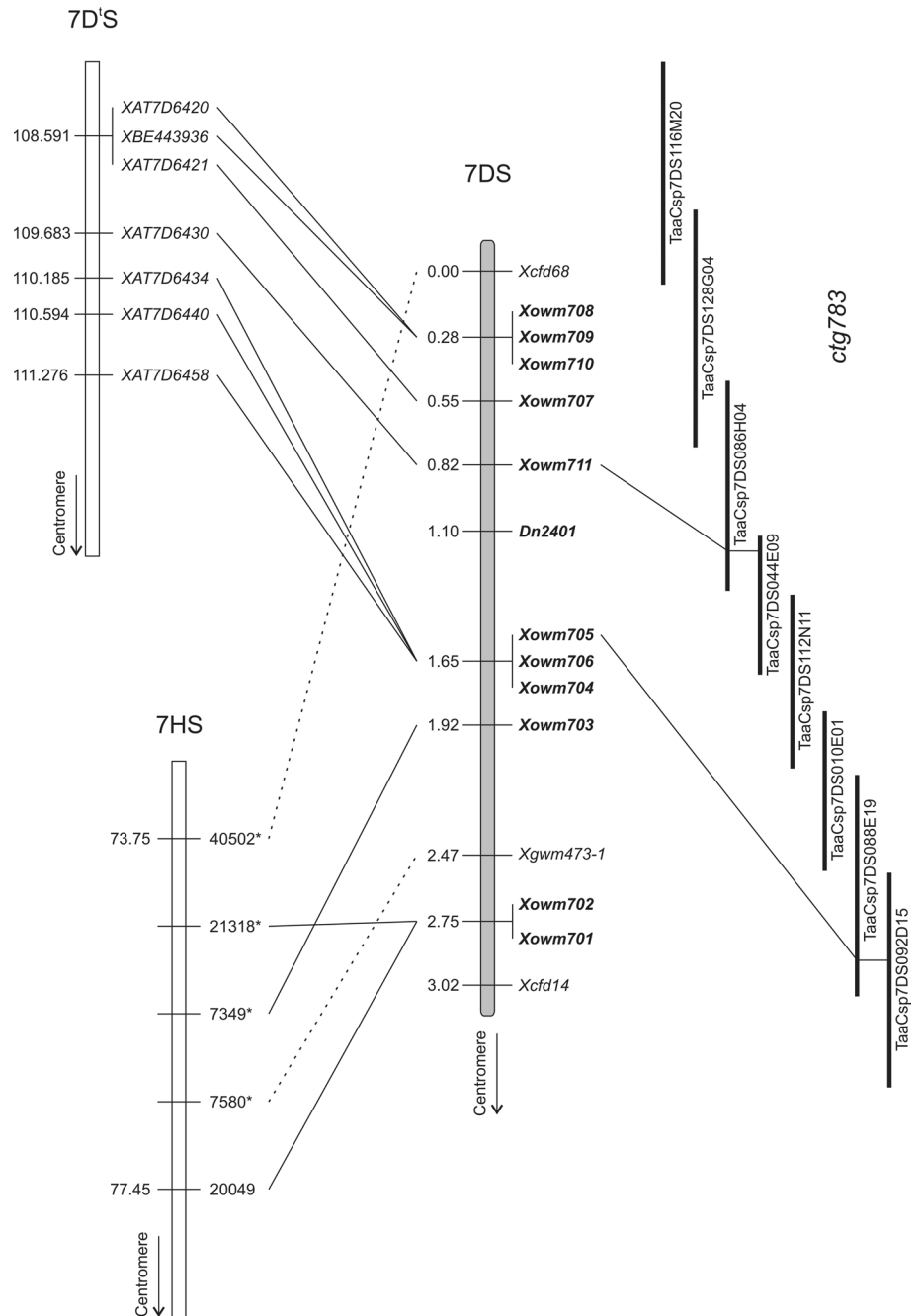


flow-sorted chromosome arms as a template. The chromosome-specific sequences obtained were used to identify HSVs for the design of genome-specific primers. This approach led to obtaining 7B-specific products for only 30 % of primer pairs, perhaps due to an insufficient level of polymorphism used to secure the primer specificity. The same strategy, including sequencing of PCR products from flow-sorted chromosome arms 1AS, 1AL, 1BS, 1BL, 1DS and 1DL was applied by Michalak de Jimenez et al. (2013) to develop chromosome-specific markers for radiation hybrid mapping on chromosome 1D. They succeeded

in developing 1D-specific markers for 39 % of investigated loci harboring at least one HSV in the primer sequence.

These results suggest that the availability of a quality sequence assembly for the desired chromosome arm and its homoeologs can significantly increase the efficiency of marker development. The knowledge of genomic context of genes used for marker development enabled us to efficiently design locus-specific primers, placing them preferentially into introns and gene-flanking regions, which exhibit lower levels of conservation. This markedly reduced non-specific amplification from homoeologous

and paralogous loci as well as conserved domains of other genes, which often occurs if the primers are designed on exon sequences (Šimková, unpublished). Comparison of the studies mentioned above indicates that a careful optimization of PCR conditions is essential to obtain locus-specific sequence (also reported by Schnurbusch et al. 2007). Although the initial phase, consisting of precise designing 7DS-specific primers and adjusting PCR conditions, is relatively laborious, it is highly compensated by the reduction of labor and costs that would be expended on sequencing and mapping of non-specific products.

The main challenge affecting the efficiency of our approach was a low level of polymorphism in our mapping population. In the present study, a total of 73,070 bp were sequenced from both 'Glupro' and CI2401, yielding 12 polymorphisms (11 SNPs and one indel), which implies a polymorphism frequency of 1 per 6,089 bp. This is ten times lower than the frequency observed by Schnurbusch et al. (2007) who found 1 SNP/613 bp when comparing close to 20 kb of sequence information from two cultivars of hexaploid wheat in the *Bo1* region. The low level of polymorphism observed in our region is probably due to the location of the *Dn2401* region on the D subgenome of wheat, known to have lower level of polymorphism as compared to the A and B subgenomes (Chao et al. 2009). Low level of polymorphism in the D subgenome was also observed by Berkman et al. (2013) who analyzed the frequency of intervarietal SNPs on 7A, 7B and 7D chromosomes.

The positions of the newly developed markers on the genetic map were compared with those of barley genic sequences in the barley GenomeZipper and markers in the *Ae. tauschii* linkage map (Fig. 2). Comparison between the mapped *Xowm* markers developed in the present work and virtually ordered barley ESTs revealed a small discrepancy in case of marker *Xowm702*, which mapped in our population more proximally as compared to the position predicted by the barley GenomeZipper. As the GenomeZipper represents a hypothetic order of genes, this discrepancy could be due to incorrect positioning of the gene on the virtual barley map. Alternatively, since the GenomeZipper has been constructed based on barley linkage map, the discrepancy may indicate disruption of micro-colinearity between barley and wheat due to a translocation or inversion. Disruptions of colinearity between wheat and barley were reported before (Mayer et al. 2011). When comparing our high-density map of the *Dn2401* region with that of *Ae. tauschii*, no mismatch was detected. The region showed complete colinearity, which is in line with the fact that *Ae. tauschii* is the ancestor of the wheat D subgenome and became a part of the bread wheat genome only 10,000 years ago. Thus, the linkage map of *Ae. tauschii* proved to be the most reliable source of markers in our study. On the other hand, the GenomeZipper provides a denser map of genes, including

links to annotations in the related genomes (http://mips.helmholtz-muenchen.de/plant/barley/gz/index.jsp), which gives an opportunity to search for candidate genes within the interval of interest. GenomeZippers were recently constructed based on wheat chromosome shotgun sequences and genetic maps (IWGSC 2014). They combine a higher information value of wheat-derived genetic maps with the high density of ordered genes introduced by GenomeZipper, resulting in the most dense virtual gene map for wheat. We expect the wheat chromosome Zippers will serve as an efficient template for wheat marker development.

Integration of genetic and physical maps for the *Dn2401* region permitted the analysis of recombination rates around the gene, and the results suggested at least eightfold increase in recombination rate in the proximity of the gene compared to the neighboring regions. Since the flanking markers *Xowm705* and *Xowm711* have been derived from *Ae. tauschii* markers AT7D6434 and AT7D6430, respectively, which have been anchored to *Ae. tauschii* physical map (Luo et al. 2013), we could precisely examine recombination rate in the orthologous region of *Ae. tauschii* genome. BAC contigs comprising orthologs of the wheat flanking markers showed about threefold increase in recombination rate compared to the average of the *Ae. tauschii* genome (0.32 cM/Mb, corresponding to 3.13 Mb/cM) and more than tenfold increase when comparing to neighboring BAC contigs. Highly recombining regions of *Ae. tauschii* were shown to be associated with disease resistance genes and signal transduction genes (GO class "receptor activity") (Luo et al. 2013), which play important roles in the innate immunity in plants (Vakhrusheva and Nedospasov 2011). In a study of Azhaguvel et al. (2012), another aphid resistance gene—*Gb3*, which confers resistance to greenbug and locates in the long arm of the 7D (7DL), was also observed to be situated in a highly recombining region (558 kb/cM and 360 kb/cM from the left and the right site, respectively) in *Ae. tauschii* 7DL. Alternative explanation for the decreased physical/genetic distance ratio around *Dn2401* is that the region in cv. 'Chinese Spring' has been affected by a deletion, which reduces the physical distance. Comparative analysis methods such as genome mapping in nanochannel arrays (Lam et al. 2012) could shed light on the structure of the region in the concerned cultivars.

## Conclusions

We have developed a powerful strategy for targeted marker development in hexaploid wheat, which relies on synteny with close relatives of wheat and on shotgun wheat chromosome sequence information from a complete homoeologous group. The approach was efficient and cost-effective, despite the low level of polymorphism between parents of

the available mapping population. It allowed us to increase marker density around the RWA resistance gene *Dn2401* by adding 11 new gene-associated markers and reduce interval delimited by flanking markers to 0.83 cM. Screening a 7DS-specific BAC library with the flanking markers identified four BAC clones that span the *Dn2401* region in cv. 'Chinese Spring'. Sequencing of these clones will provide data for development of additional markers and potentially suggest candidates for the resistance gene. With the availability of the shotgun sequence assemblies and GenomeZippers for each of the wheat chromosome arms, the strategy described here can be applied for targeted marker development in any region of the wheat genome.

# References

Azhaguvel P, Rudd JC, Ma Y, Luo MC, Weng Y (2012) Fine genetic mapping of greenbug aphid-resistance gene *Gb3* in *Aegilops tauschii*. Theor Appl Genet 124:555–564

Berkman PJ, Skarshewski A, Lorenc M, Lai K, Duran C, Ling EYS, Stiller J, Smits L, Imelfort M, Manoli S, McKenzie M,

Kubaláková M, Šimková H, Batley J, Fleury D, Doležel J, Edwards D (2011) Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. Plant Biotech J 9:768–775

Berkman PJ, Visendi P, Lee HC, Stiller J, Manoli S, Lorenc MT, Lai K, Batley J, Fleury D, Šimková H, Kubaláková M, Weining S, Doležel J, Edwards D (2013) Dispersion and domestication shaped the genome of bread wheat. Plant Biotech J 11:564–571

Bossolini E, Krattinger SG, Keller B (2006) Development of SSR markers specific for the *Lr34* resistance region of wheat using sequence information from rice and *Aegilops tauschii*. Theor Appl Genet 113:1049–1062

Burd JD, Burton RL (1992) Characterization of plant-damage caused by the Russian wheat aphid (Homoptera, Aphididae). J Econ Entomol 85:2017–2022

Chao S, Zhang W, Akhunov E, Sherman J, Ma Y, Luo MC, Dubcovsky J (2009) Analysis of gene-derived SNP marker polymorphism in US wheat (*Triticum aestivum* L.) cultivars. Mol Breed 23:23–33

Close TJ, Wanamaker S, Roose ML, Lyon M (2007) HarvEST: an EST database and viewing software. Methods Mol Biol 406:161–178

Close TJ, Bhat PR, Lonardi S, Wu Y, Rostoks N, Ramsay L, Druka A, Stein N, Svensson JT, Wanamaker S, Bozdag S, Roose ML, Moscou MJ, Chao S, Varshney RK, Szucs P, Sato K, Hayes PM, Matthews DE, Kleinhofs A, Muehlbauer GJ, DeYoung J, Marshall DF, Madishetty K, Fenton RD, Condamine P, Graner A, Waugh R (2009) Development and implementation of high-throughput SNP genotyping in barely. BMC Genom 10:582

Collins MB, Haley SD, Peairs FB, Rudolph JB (2005) Biotype 2 Russian Wheat Aphid resistance among wheat germplasm accessions. Crop Sci 45:1877–1880

Doležel J, Vrána J, Šafář J, Bartoš J, Kubaláková M, Šimková H (2012) Chromosomes in the flow to simplify genome analysis. Funct Integr Genomics 12:397–416

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2010) Geneious v5.3. http://www.geneious.com

FAOSTAT (2013) FAOSTAT, Food and Agriculture Organization of the United Nations. http://faostat.fao.org/

Fazel-Najafabadi M, Peng J, Peairs FB, Simkova H, Kilian A, Lapitan NLV (2014) Genetic mapping of resistance to *Diuraphis noxia* (Kurdjumov) Biotype 2 in wheat (*Triticum aestivum* L.) accession CI2401. Euphytica. doi:10.1007/s10681-014-1284-0

Girma M, Wilde GE, Harvey TL (1993) Russian wheat aphid (*Homoptera*: *Aphididae*) affects yield and quality of wheat. J Econ Entomol 86:594–601

Keller B, Feuillet C, Yahiaoui N (2005) Map-based isolation of disease resistance genes from bread wheat: cloning in a supersize geonome. Genet Res 85:93–100

Kosambi D (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Kubaláková M, Vrána J, Číhalíková J, Šimková H, Doležel J (2002) Flow karyotyping and chromosome sorting in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 104:1362–1372

Kubaláková M, Valárik M, Bartoš J, Vrána J, Číhalíková J, Molnár-Láng M, Doležel J (2003) Analysis and sorting of rye (*Secale cereale* L.) chromosomes using flow cytometry. Genome 46:893–905

Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, Deshpande P, Cao H, Nagarajan N, Xiao M, Kwok PY (2012) Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 30:771–776

Lapitan NLV, Peng JH, Sharma V (2007) A high-density map and PCR markers for Russian wheat aphid resistance gene *Dn7* on chromosome 1RS/1BL. Crop Sci 47:809–818

Liu XM, Smith CM, Gill BS, Tolmay V (2001) Microsatellite markers linked to six Russian wheat aphid resistance genes in wheat. Theor Appl Genet 102:504–510

Liu XM, Smith CM, Gill BS (2002) Identification of microsatellite markers linked to Russian wheat aphid resistance genes *Dn4* and *Dn6*. Theor Appl Genet 104:1042–1048

Luo MC, Gu YQ, You FM, Deal KR, Ma Y, Hu Y, Huo N, Wang Y, Wang J, Chen S, Jorgensen CM, Zhang Y, McGuire PE, Pasternak S, Stein JC, Ware D, Kramer M, McCombie WR, Kianian SF, Martis MM, Mayer KF, Sehgal SK, Li W, Gill BS, Bevan MW, Šimková H, Doležel J, Weining S, Lazo GR, Anderson OD, Dvorak J (2013) A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. Proc Natl Acad Sci USA 110:7940–7945

Lysák MA, Číhalíková J, Kubaláková M, Šimková H, Künzel G, Doležel J (1999) Flow karyotyping and sorting of mitotic chromosomes of barley (*Hordeum vulgare* L.). Chrom Res 7:431–444

Marais GF, Wessels WG, Horn M (1998) Association of a stem rust resistance gene (*Sr45*) and two Russian wheat aphid resistance genes (*Dn5* and *Dn7*) with mapped structural loci in common wheat. S Afr J Plant Soil 15:67–71

Martin TJ, Sears RG, Seifers DL, Harvey TL, Witt MD, Schlegel AJ, McCluskey PJ, Hatchett JH (2001) Registration of Trego wheat. Crop Sci 41:929–930

Martis MM, Zhou R, Haseneyer G, Schmutzer T, Vrána J, Kubaláková M, König S, Kugler KG, Scholz U, Hackauf B, Korzun V, Schön CC, Doležel J, Bauer E, Mayer KFX, Stein N (2013) Reticulate evolution of the rye genome. Plant Cell 25:3685–3698

Mayer KFX, Taudien S, Martis M, Šimková H, Suchánková P, Gundlach H, Wicker T, Petzold A, Felder M, Steuernagel B, Scholz U, Graner A, Platzer M, Doležel J, Stein N (2009) Gene content and virtual gene order of barley chromosome 1H. Plant Physiol 151:496–505

Mayer KFX, Martis M, Hedley PE, Šimková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Doležel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

Michalak de Jimenez MK, Bassi FM, Ghavami F, Simons K, Dizon R, Seetan RI, Alnemer LM, Denton AM, Doğramaci M, Šimková H, Doležel J, Seth K, Luo MC, Dvorak J, Gu YQ, Kianian SF (2013) A radiation hybrid map of chromosome 1D reveals synteny conservation at a wheat speciation locus. Funct Integr Genomics 13:19–32

Molnár I, Kubaláková M, Šimková H, Cseh A, Molnár-Láng M, Doležel J (2011) Chromosome isolation by flow sorting in *Aegilops umbellulata* and *Ae. comosa* and their allotetraploid hybrids *Ae. biuncialis* and *Ae. geniculata*. PLoS One 6:e27708

Morrison WP, Peairs FB (1998) Response model concept and economic impact. In: Thomas Say Publications in Entomology, Entomological Society of America, Lanham, MD

Nkongolo KK, Quick JS, Limin AE, Fowler DB (1991) Sources and inheritance of resistance to Russian wheat aphid in *Triticum* species amphiploids and *Triticum tauschii*. Can J Plant Sci 71:703–708

Peng JH, Bai Y, Haley SD, Lapitan NLV (2009) Microsatellite-based molecular diversity of bread wheat germplasm and association mapping of wheat resistance to the Russian wheat aphid. Genetica 135:95–122

Qin B, Cao A, Wang H, Chen T, You FM, Liu Y, Ji J, Liu D, Chen P, Wang XE (2011) Colinearity-based marker mining for the fine mapping of *Pm6*, a powdery mildew resistance gene in wheat. Theor Appl Genet 123:207–218

Quraishi UM, Abrouk M, Bolot S, Pont C, Throude M, Guilhot N, Confolent C, Bortolini F, Praud S, Murigneux A, Charmet G, Salse J (2009) Genomics in cereals: from genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection. Funct Integr Genomics 9:473–484

Qureshi JA, Michaud JP, Martin TJ (2006) Resistance to biotype 2 Russian wheat aphid (Homoptera: Aphididae) in two wheat lines. J Econ Entomol 99:544–550

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. Humana Press, Totowa, pp 365–386

Šafář J, Šimková H, Kubaláková M, Číhalíková J, Suchánková P, Bartoš J, Doležel J (2010) Development of chromosome-specific BAC resources for genomics of bread wheat. Cytogenet Genome Res 129:211–223

Schnurbusch T, Collins NC, Eastwood RF, Sutton T, Jefferies SP, Langridge P (2007) Fine mapping and targeted SNP survey using rice–wheat gene colinearity in the region of the *Bo1* boron toxicity tolerance locus of bread wheat. Theor Appl Genet 115:451–461

Šimková H, Svensson JT, Condamine P, Hřibová E, Suchánková P, Bhat PR, Bartoš J, Šafář J, Close TJ, Doležel J (2008) Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. BMC Genom 9:294

Šimková H, Šafář J, Kubaláková M, Suchánková P, Číhalíková J, Robert-Quatre H, Azhaguvel P, Weng Y, Peng J, Lapitan NL, Ma Y, You FM, Luo MC, Bartoš J, Doležel J (2011) BAC libraries from wheat chromosome 7D: efficient tool for positional cloning of aphid resistance genes. J Biomed Biotechnol 2011:302543

Sorrells ME, La Rota M, Bermudez-Kandianis CE, Greene RA, Kantety R, Munkvold JD, Miftahudin, Mahmoud A, Ma X, Gustafson PJ, Qi LL, Echalier B, Gill BS, Matthews DE, Lazo GR, Chao S, Anderson OD, Edwards H, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorak J, Zhang D, Nguyen HT, Peng J, Lapitan NL, Gonzalez-Hernandez JL, Anderson JA, Hossain K, Kalavacharla V, Kianian SF, Choi DW, Close TJ, Dilbirligi M, Gill KS, Steber C, Walker-Simmons MK, McGuire PE, Qualset CO (2003) Comparative DNA sequence analysis of wheat and rice genomes. Genome Res 13:1818–1827

Terracciano I, Maccaferri M, Bassi F, Mantovani P, Sanguineti MC, Salvi S, Šimková H, Doležel J, Massi A, Ammar K, Kolmer J, Tuberosa R (2013) Development of COS-SNP and HRM markers for haplotyping and marker-assisted selection of *Lr14* in durum wheat (*Triticum durum* Desf.). Theor Appl Genet 26:1077–1101

The International Wheat Genome Sequencing Consortium (IWGSC) (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788

Vakhrusheva OA, Nedospasov SA (2011) System of innate immunity in plants. Mol Biol 45:16–23

Valdez VA, Byrne PF, Lapitan NLV, Peairs FB, Bernardo A, Bai G, Haley SD (2012) Inheritance and genetic mapping of Russian wheat aphid resistance in iranian wheat landrace accession PI 626580. Crop Sci 52:676–682

Van Ooijen JW, Voorrips RE (2001) JoinMap 3.0, software for the calculation of genetic linkage maps. Plant J 3:739–744

# APPENDIX III

**Sequence-based tools facilitate high-density mapping of a Russian wheat aphid resistance gene in wheat**

Staňková H, Valárik M, Šafář J, Lapitan N, Stein N, Berkman P, Edwards D, Doležel J, Šimková H

# Sequence–based tools facilitate high-density mapping of a Russian wheat aphid resistance gene in wheat

Helena Staňková[1], Miroslav Valárik[1], Jan Šafář[1], Nora Lapitan[2], Nils Stein[3], Paul Berkman[4], David Edwards[4], Jaroslav Doležel[1], Hana Šimková[1]

[1] Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Sokolovská 6, CZ-77200 Olomouc, Czech Republic
[2] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado, 80524, USA
[3] Leibnitz Institute of Plant Genetics and Crop Plant Research, Department Genebank, AG Genome Diversity, D-06466 Gatersleben, Germany
[4] Australian Centre for Plant Functional Genomics, University of Queensland, Brisbane, QLD 4072, Australia

## Background

Bread wheat (*Triticum aestivum* L.) is one of the most important crop species providing staple food for 35% of the world's population. It is an allohexaploid species (2n = 6x = 42, AABBDD genome) originating from two interspecific hybridizations. Genome mapping as well as positional cloning in bread wheat are hampered by its huge genome size (~ 17 Gbp), polyploidy and large amount of repetitive sequences (> 80%).

Russian wheat aphid (*Diuraphis noxia*), native in Afghanistan, is an important world invasive pest of wheat and barley crops. Several *D. noxia* biotypes with various virulency have spread in all wheat and barley growing areas with the exception of Australia. Host resistance is the most efficient, economical, and environmentally safe approach to protect wheat from pathogens while minimizing the use of pesticides. Several genes contributing to RWA resistance were identified in various wheat lines but only a few confer resistance to highly virulent US RWA biotype 2. CI2401 line carries *DnCI2401* gene that underlies resistance to RWA biotype 2 on the short arm of chromosome 7D (7DS).

Construction of a high-density genetic map covering the Dn*CI2401* gene region is essential for positional cloning of this resistance gene as well as marker-assisted selection. In this work, we present an approach to marker development employing new sequence-based tools and resources.

## Genetic map
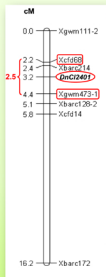


- Flanking SSR markers **cfd68** and **gwm473** delimit an interval of 2.5 cM around the resistance gene *DnCI2401*

- Saturation of genetic map in this region is essential for positional cloning of the *DnCI2401* gene

## Mapping populations

- **CI2401 x Glupro**

  CI2401 – resistant – carrying *DnCI2401*
  cv. Glupro – susceptible
  152 F2 plants

- **Synthetic W7984 x Opata 85 (ITMI)**

  Subsidiary population
  with a higher level of polymorphism
  115 F8 RILs

**Bread wheat**
(*Triticum aestivum*)

**Russian wheat aphid**
(*Diuraphis noxia*)

http://www.forestryimages.org/

### 7AS, 7BS and 7DS Illumina paired-read sequence data

- Flow-sorted 7AS, 7BS and 7DS arms amplified by multiple displacement amplification have been sequenced by Illumina GAIIx technology
- Unique and low copy regions of the chromosome arms were assembled and scaffolded

| Chromosome arm | Length (Mbp) | Sequence data (Mbp) | Coverage |
|---|---|---|---|
| 7AS | 407 | 8,700 | 21,3x |
| 7BS | 360 | 10,800 | 30x |
| 7DS | 381 | 13,050 | 34,2x |

Draft assemblies are publicaly available at www.wheatgenome.info

## Designing 7DS specific SNP markers

- Syntenic regions in genomes of barley, rice, *Brachypodium* and sorghum aligned in GenomeZipper corresponding to our region of interest were delimited
- Barley ESTs from HarvEST database with a known position on barley genetic map were blasted against the wheat 7AS, 7BS and 7DS assemblies
- Contigs containing sequences homologous to the barley ESTs were identified

## GenomeZipper



GenomeZipper was created by combining *Brachypodium*, rice and sorghum gene sequences, 454 reads of particular barley chromosomes and barley ESTs.
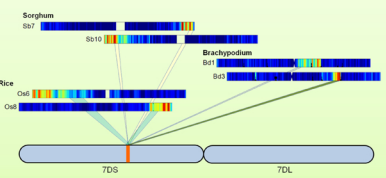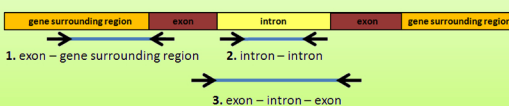
Syntenic regions in rice, *Brachypodium* and sorghum identified by GenomeZipper

- 7AS, 7BS and 7DS sequences were aligned and SNPs and/or INDELs were identified
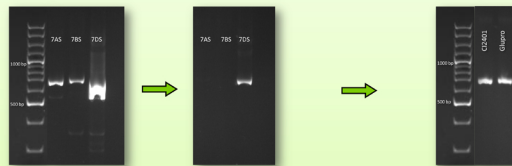


**Forward primer**   **Reverse primer**

- 7DS specific primers were designed

**3 types of primer pairs:**



1. exon – gene surrounding region   2. intron – intron

3. exon – intron – exon

- After optimizing PCR conditions, 7DS specific product was obtained

- 7DS specific sequences were amplified from DNA of parents of mapping populations


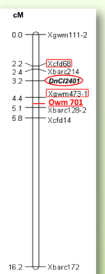
- PCR products were sequenced and SNPs between parents of mapping populations were identified
- Mapping of the SNPs is in progress





**New SNP marker Owm701 was placed on genetic map in the *DnCI2401* region**

## Results

- In total, 9 SNPs in both mapping populations were identified
  - 5 in gene surrounding regions
  - 3 in introns
  - 1 in exon

| Mapping population | Sequences compared | Number of SNPs | Frequency of SNPs |
|---|---|---|---|
| CI2401 x Glupro | 11,600 bp | 2 | 1 SNP / 5,800 bp |
| ITMI | 5,300 bp | 7 | 1 SNP / 750 bp |

- New SNP marker **Owm701** was placed on genetic map close to the *DnCI2401* gene, confirming colinearity between wheat and barley in the *DnCI2401* region.

## Conclusions

- The presented approach employing sequence information from individual wheat chromosome arms helps reduce problems associated with polyploidy and enables targeted development of markers from the region of interest
- Most of the SNPs were found in gene surrounding regions, which will be a target for the future SNP identification
- The present mapping population (CI2401 x Glupro) shows very low level of polymorphism ➡ development of a new mapping population including synthetic wheat will be essential

# APPENDIX IV

**Poziční klonování genu pro rezistenci k mšici zhoubné (*Diuraphis noxia*):
Konstrukce vysokohustotní genetické mapy**

Staňková H, Valárik M, Lapitan N, Berkman P, Edwards D, Luo MC, Tulpová Z, Kubaláková M, Stein N, Doležel J, Šimková H

[In Czech]

# Poziční klonování genu pro rezistenci k mšici zhoubné (*Diuraphis noxia*): konstrukce vysokohustotní genetické mapy

Helena Staňková[1], Miroslav Valárik[1], Nora Lapitan[2], Paul Berkman[3], David Edwards[3], Ming-Cheng Luo[4], Zuzana Tulpová[1], Marie Kubaláková[1], Nils Stein[5], Jaroslav Doležel[1], Hana Šimková[1]

[1] Centrum regionu Haná pro biotechnologický a zemědělský výzkum, Ústav experimentální botaniky, Šlechtitelů 31, 783 71, Olomouc - Holice, Czech Republic
[2] Department of Soil and Crop Sciences, Colorado State University, Fort Collins, Colorado, 80524, USA
[3] Australian Centre for Plant Functional Genomics, University of Queensland, Brisbane, QLD 4072, Australia
[4] Department of Plant Sciences, University of California, Davis, CA 95616, USA
[5] Leibniz Institute of Plant Genetics and Crop Plant Research, Department Genebank, AG Genome Diversity, D-06466 Gatersleben, Germany

## Úvod

Pšenice setá (*Triticum aestivum* L.) představuje jednu z ekonomicky nejvýznamnějších kulturních plodin, jež poskytuje zdroj potravy pro 35% obyvatel světa. Jedná se o allohexaploidním druh (2n = 6x = 42) s celkovou velikostí genomu téměř $17 \times 10^9$ bp. Genom je tvořen třemi homeologními subgenomy (A, B a D) a jeho podstatnou část (přes 80%) tvoří repetitivní sekvence. Všechny výše zmíněné vlastnosti pšeničného genomu znesnadňují jeho analýzu, genetické i fyzické mapování, sekvenování či poziční klonování. Třídění jednotlivých chromozómů a jejich ramen pomocí průtokové cytometrie umožňuje rozložit tento obrovský genom na malé a snadno analyzovatelné části.

Na krátkém rameni chromozómu 7D (7DS) pšenice byl identifikován gen *Dn2401* pro rezistenci k mšici zhoubné (*Diuraphis noxia*). Mšice zhoubná je jedním z nejvýznamnějších škůdců pšenice a ječmene. Chemické i biologické postupy hubení nejsou v případě mšice zhoubné dostatečně účinné. Z tohoto důvodu se jeví jako nejvýhodnější způsob ochrany pěstování odrůd nesoucích geny pro rezistenci vůči tomuto škůdci. Konstrukce vysokohustotní genetické mapy pokrývající oblast zkoumaného genu je nezbytná pro jeho následné poziční klonování, tedy izolaci genu na základě jeho pozice na genetické či fyzické mapě. Za účelem konstrukce mapy byla vyvinuta metoda pro cílené odvozování vysoce specifických markerů ze zkoumané oblasti v podmínkách polyploidního genomu. Naše metoda využívá syntenie mezi pšenicí a jejími příbuznými druhy (ječmen, *Brachypodium*, rýže, čirok, *Aegilops tauschii*) v kombinaci se sekvencemi jednotlivých chromozómů skupiny 7, získanými celochromozómovým neuspořádaným (shotgun) sekvenováním. Tímto způsobem jsme získali nové markery specifické pro oblast studovaného genu. Následně byl za pomoci těchto markerů identifikován kontig ve fyzické mapě ramene 7DS, jenž překlenuje oblast genu *Dn2401*. Klony z tohoto kontigu byly osekvenovány a následně anotovány. Anotace sekvencí odhalila přítomnost několika kandidátních genů.

## Gen *Dn2401*

- Podmiňuje rezistenci k mšici zhoubné (*Diuraphis noxia*)
- Identifikován v linii CI2401
- Lokalizován na krátkém rameni chromozómu 7D pšenice (7DS)
- Zamapován mezi markery *cfd68* a *gwm473* (2,39 cM)
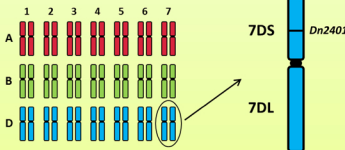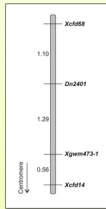
### Mapovací populace F2
- Křížení **CI2401** (rezistentní) x **Glupro** (citlivá)
- 158 jedinců

## Markery *owm*

- Celková spěšnost navrhování primerů specifických pro rameno 7DS – 86% (138 ze 161 párů)
- Pouze 11 párů primerů poskytovalo produkt polymorfní mezi rodiči mapovací populace

- Odvozeno 11 nových markerů
  - *owm701* – *owm710* – SNP markery
  - *owm711* – délkový polymorfismus
- Interval mezi hraničními markery zkrácen z 2,39 cM na 0,81 cM
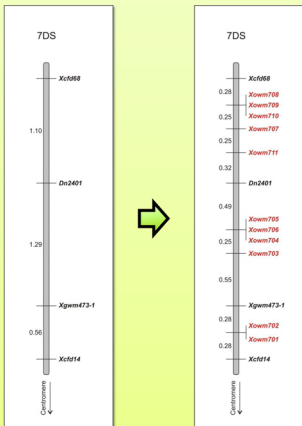
## Konstrukce vysokohustotní genetické mapy

- využití syntenie mezi pšenicí a příbuznými druhy - ječmen, *Brachypodium*, rýže, čirok, *Aegilops tauschii*
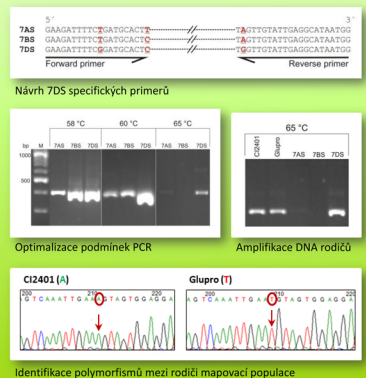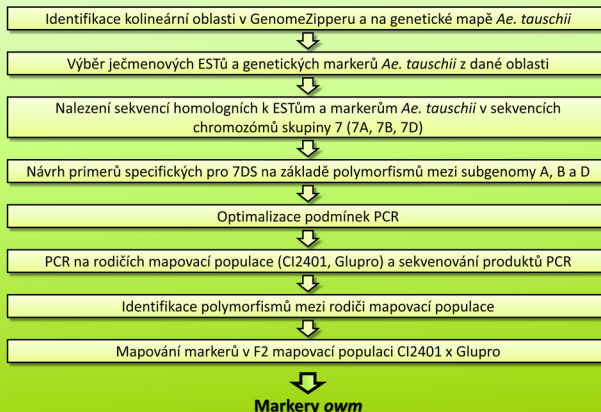
### Genomické zdroje pro vývoj genetických markerů
- **GenomeZipper**
  - Soubor genů ječmene seřazených na základě kolinearity s osekvenovanými genomy rýže, *Brachypodia* a čiroku a ukotvený na genetickou mapu ječmene tvořenou EST markery
- **Genetická mapa *Aegilops tauschii***
  - Tvořena geneticky zamapovanými SNP markery
  - Většina markerů odvozena z genů
- **Sekvence chromozómů skupiny 7**
  - Získány celochromozómovým neuspořádaným (shotgun) sekvenováním

**GenomeZipper**

| Marker | cM | Brachypodium | Rýže | Čirok | Ječmenový EST |
|---|---|---|---|---|---|
| - | - | Brad1g44100.1 | Os06g0257200 | - | 270 |
| - | - | Brad1g44110.1 | Os06g0256900 | Sb10g031100.1 | 8010;30324 |
| - | - | - | Os06g0256500 | - | - |
| 3_0496 | 73.75 | Brad1g44690.1 | Os06g0127900 | - | 1896 |
| - | - | Brad1g44150.1 | Os06g0256200 | Sb10g009240.1 | 18538;13626 |
| - | - | Brad1g44160.1 | Os06g0256100 | Sb10g009230.1 | 45494;25645 |
| - | - | - | Os06g0255700 | - | - |
| - | - | Brad1g44180.1 | Os06g0255400 | Sb10g009200.1 | 43211;39112 |
| - | - | - | Os06g0255200 | - | - |
| 1_1257 | 73.75 | Brad1g44190.1 | Os06g0256100 | Sb10g009180.1 | 37017;17759 |
| - | - | Brad1g44200.1 | Os06g0090170.1 | - | 3189;3290 |
| - | - | Brad1g44220.1 | Os06g0253600 | Sb10g009120.1 | - |
| - | - | Brad1g44230.1 | Os06g0253100 | Sb10g009090.1 | - |
| - | - | - | - | Sb10g009080.1 | - |
| 1_0153 | 73.75 | Brad1g44460.1 | Os06g0247500 | Sb10g008850.1 | 15071 |
| - | - | Brad1g44480.1 | Os06g0246700 | - | 50561;9698;35077 |
| - | - | Brad1g44490.1 | Os06g0246500 | - | 16068 |
| - | - | Brad1g49730.1 | Os06g0129200 | Sb10g001800.1 | - |
| - | - | Brad1g49740.1 | Os06g0129100 | Sb10g001810.1 | 8409;31629 |
| - | - | Brad1g49740.1 | Os06g0129000 | Sb10g001820.1 | 8647 |
| - | - | Brad1g49750.1 | - | Sb10g001830.1 | 23662 |
| 3_0832 | 73.75 | Brad1g48760.1 | - | - | 55684 |

### Postup vývoje lokusově specifických genetických markerů

- Identifikace kolineární oblasti v GenomeZipperu a na genetické mapě *Ae. tauschii*
- Výběr ječmenových ESTů a genetických markerů *Ae. tauschii* z dané oblasti
- Nalezení sekvencí homologních k ESTům a markerům *Ae. tauschii* v sekvencích chromozómů skupiny 7 (7A, 7B, 7D)
- Návrh primerů specifických pro 7DS na základě polymorfismů mezi subgenomy A, B a D
- Optimalizace podmínek PCR
- PCR na rodičích mapovací populace (CI2401, Glupro) a sekvenování produktů PCR
- Identifikace polymorfismů mezi rodiči mapovací populace
- Mapování markerů v F2 mapovací populaci CI2401 x Glupro
- **Markery *owm***

Návrh 7DS specifických primerů

Optimalizace podmínek PCR

Amplifikace DNA rodičů

Identifikace polymorfismů mezi rodiči mapovací populace

## Fyzické mapování a sekvenování

### BAC knihovna DNA z ramen 7DS
- Skríning 3D poolů hraničními markery *owm705* a *owm711*
- Identifikace pozitivních BAC klonů
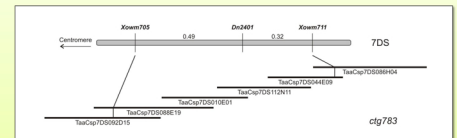
### Fyzická mapa 7DS
- Nalezení BAC kontigu, jenž obsahuje pozitivní BAC klony → *ctg783*

### Sekvenování
- MTP klony mezi hraničními markery
- Sekvenování technologií Illumina
- Sekvence skládány programem Sassy

## Anotace sekvencí BAC klonů
- Porovnání sekvencí s databází anotovaných genů ječmene
- Nalezeny dva kandidání geny
  - salycylát o-methyltransferáza (součást metabolické dráhy kyseliny salicylové)
  - epoxid hydroláza (pozorována zvýšená exprese při napadení rostliny mšicí)

- Kandidátní geny budou sekvenování za použití DNA obou rodičů mapovací populace → hledání polymorfismů

## Souhrn

- Bylo odvozeno 11 nových markerů
  → Zamapovány do oblasti genu *Dn2401*
- Interval vymezující oblast genu byl zkrácen z 2,39 cM na 0,81 cM
- Nové hraniční markery umožnily identifikovat kontig *ctg783* překlenující studovanou oblast
- Sekvenování BAC klonů z *ctg 783* odhalila přítomnost dvou kandidátních genů

# APPENDIX V

**Physical map and sequencing of wheat chromosome arm 7DS**

Staňková H, Luo MC, Visendi P, Tulpová Z, Batley J, Bartoš J, Doležel J, Edwards D, Šimková H

# Physical map and sequencing of wheat chromosome arm 7DS

Helena Staňková[1], Ming-Cheng Luo[2], Paul Visendi[3], Zuzana Tulpová[1], Jacqueline Batley[4], Jan Bartoš[1], Jaroslav Doležel[1], David Edwards[3], Hana Šimková[1]

[1] Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, CZ- 783 71, Olomouc - Holice, Czech Republic
[2] Department of Plant Sciences, University of California, Davis, CA 95616
[3] School of Agriculture and Food Sciences and Australian Centre for Plant Functional Genomics, University of Queensland, Brisbane, QLD 4072, Australia
[4] Centre for Integrative Legume Research, University of Queensland, School of Agriculture and Food Sciences, Brisbane, QLD 4072, Australia

## Background

Bread wheat (*Triticum aestivum*) is one of the most important crop species in the world providing staple food for 35% of world's population. Physical map construction and sequencing of wheat genome are hampered by huge genome size (1C = 17 Gb), presence of three homeologous subgenomes (A, B and D) and prevalence of repetitive sequences (>80%). The possibility to divide the wheat genome into individual chromosome arms by flow cytometric sorting enables to cope with polyploidy and to focus on the desired part of the genome.

The aim of this work was to construct physical contig map of 7DS chromosome arm and to sequence minimum tilling path (MTP). BAC clones from the 7DS-specific BAC library were fingerprinted using SNaPshot-based HICF technology and than automatically assembled into contigs using FPC software. Integration of the 7DS physical map with that of *Aegilops tauschii* (D genome ancestor) provided a clue for further merging of contigs. Reliability of the assembly was verified through LTC software. The physical map has been anchored to the genetic map applying both forward and reverse anchoring strategy. BAC clones representing 7DS MTP are being sequenced by Illumina platform and assembled into sequence contigs using Sassy software which was designed to assemble short complex sequences from Illumina paired read data. Anchored 7DS physical map and the sequence of the 7DS MTP will become a valuable tool for genetic mapping and positional cloning of genes located on 7DS chromosome arm of wheat.

## Physical map construction

### 1) Fingerprinting

- All 49,157 BAC clones from 7DS-specific BAC library were fingerprinted with the SNaPshot high-information-content-fingerprinting (HICF) technology
- 39,765 fingerprints were useful for contig assembly

#### 7DS-specific BAC library

Size of the 7DS chromosome arm is 381 Mb (2,2% of the wheat genome)

7DS arms were flow-sorted from 7DS/7DL double ditelosomic line of cv. Chinese Spring
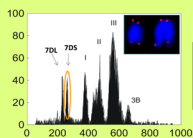
| Library characteristics | |
|---|---|
| Number of clones | 49,152 |
| Average insert size | 113 kb |
| Coverage | 12.2x |

### 2) Contig assembly

- Clones were assembled into contigs based on fingerprint overlaps

**FPC software (FingerPrinted Contigs)**
- Automatic assembly up to cut-off value 1e[-45]
- Manual end-merging of automatically assembled contigs
  → Based on integration of *Aegilops tauschii* physical contig map with physical contig map of 7DS chromosome arm

**LTC software (Linear Topology Contigs)**
- Three-dimensional view of contigs assembled by FPC software → verification of the assembly made by FPC
- Contigs were manually edited in FPC based on the 3-D view from LTC
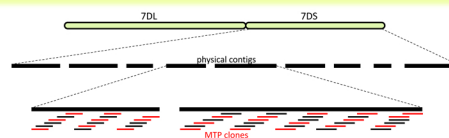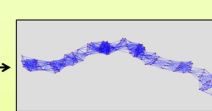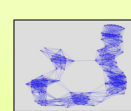
**Selection of MTP BAC clones**
- Selected from final assembly
- FPC software

#### Parameters of the final assembly

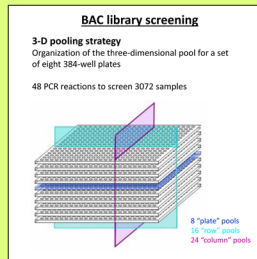| | |
|---|---|
| Length of the assembly | 362 Mbp (95% of the 7DS) |
| Number of clones | 39,765 |
| Number of contigs | 931 |
| Number of singletons | 11,426 |
| **Number of MTP BAC clones** | **4,608** |

## Physical map anchoring

### Landing on the chromosome
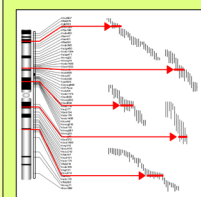
**Forward anchoring**
- **PCR screening of 3-D pools** from 7DS-specific BAC library with genetically mapped wheat SSR and STS markers
- *In silico* anchoring – integration of 7DS physical contig map with anchored physical contig map of *Aegilops tauschii*

- 31 markers anchored by PCR screening (20 SSR markers, 11 STS markers)
- 583 markers anchored *in silico*

**BAC library screening**
**3-D pooling strategy**
Organization of the three-dimensional pool for a set of eight 384-well plates

48 PCR reactions to screen 3072 samples

8 "plate" pools
16 "row" pools
24 "column" pools

**Reverse anchoring**
- STS markers were developed based on sequences of MTP BAC clones
  - 9216 BAC-end sequences
  - Sequences of 7 completely sequenced and assembled BAC clones
- Markers were mapped on D-genome specific radiation hybrid (RH) panel (Kumar *et al*., 2012)

- 2 STS markers were mapped on D-genome specific RH panel
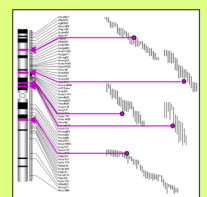
## Sequencing of 7DS MTP

### Sequencing of MTP BAC pools
- 4 BAC clones were pooled into one pool
  → **1152 BAC pools** out of 4608 MTP BAC clones

- DNA sequencing libraries were constructed using TruSeq DNA PCR-Free Sample Preparation Kit
  - Size selection – fragments size ~550 bp
- Pools were sequenced on Illumina MiSeq sequencer
  → 250 bp pair-end reads

### BAC-end sequencing
- All MTP BAC clones were sequenced by Sanger sequencing from both ends
  → **9216 BAC-end sequences**

### Assembly of sequencing data
- Illumina pair-end reads were assembled into sequence contigs using **Sassy** software
- 250 bp paired reads from each BAC pool were cut into 61 bp offset paired reads and assembled

~700 bp fragment with 250 bp paired reads

61 bp paired reads
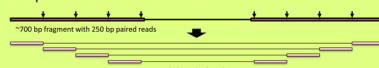
- BAC-end sequences were mapped to the sequence contigs
  → The ends of BAC assemblies and individual BACs in the pool were identified
- The assembly is supported by whole genome mate-pair data
- Sequence reads of individual BAC clones were assembled into 1 - 7 sequence contigs per BAC

## Results and conclusions

- Physical map was anchored by 616 markers located in 309 contigs → **51%** of the assembly
- Wheat genome sequencing by proposed strategy is feasible
- Sequencing of all 7DS MTP BAC clones is in progress
- Sequences of BAC clones from the 7DS MTP will be used for further anchoring of the 7DS physical map

**Reference**
Kumar, A., Simons, K., Iqbal, M.J., Michalak de Jimenéz, M., Bassi, F.M., Ghavami, F., Al-Azzam, O., Drader, T., Wang, Y., Luo, M.-Ch., Gu, Y.Q., Denton, A., Lazo, G.R., Xu, S.S., Dvorak, J., Kianian, P.M., Kianian, S.F. (2012): Physical mapping resources for large plant genomes: radiation hybrids for wheat D-genome progenitor *Aegilops tauschii*. BMC Genomics 13: 597.

# APPENDIX VI

**Improving the physical map of the wheat 7DS chromosome arm with a BioNano genome map**

Staňková H, Hastie A, Cao H, Vrána J, Kubaláková M, Visendi P, Hayashi S, Luo MC, Batley J, Edwards D, Doležel J, Šimková H

# Improving the physical map of the wheat 7DS chromosome arm with a BioNano genome map

Helena Staňková[1], Alex Hastie[2], Han Cao[2], Jan Vrána[1], Marie Kubaláková[1], Paul Visendi[3], Satomi Hayashi[4], Mingcheng Luo[5], Jacqueline Batley[4], David Edwards[3], Jaroslav Doležel[1], Hana Šimková[1]

[1] Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, CZ-78371, Olomouc, Czech Republic
[2] BioNano Genomics, 9640 Towne Centre Drive, San Diego, CA 92121, USA
[3] Australian Centre for Plant Functional Genomics, University of Queensland, Brisbane, QLD 4072, Australia
[4] School of Agriculture and Food Sciences, University of Queensland, Brisbane, QLD 4072, Australia
[5] Department of Plant Sciences, University of California, Davis, CA 95616, USA

## Background

Specific features of the wheat genome such as a huge genome size (17 Gb), presence of three homoeologous subgenomes (A, B and D) and prevalence of repetitive sequences hamper mapping and sequencing of this complex genome. To overcome these obstacles, a BAC-by-BAC sequencing strategy based on chromosomal physical maps has been adopted by the International Wheat Genome Sequencing Consortium.

In order to sequence a short arm of chromosome 7D (7DS), a BAC contig physical map was constructed. BAC clones representing minimum tilling path (MTP) were sequenced using Illumina HiSeq platform and assembled using Sassy and Sspace software. With the aim to validate and improve the physical map as well as sequence assembly of the 7DS arm, 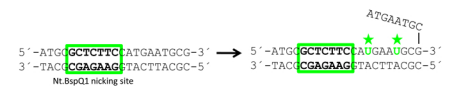a BioNano genome map was constructed using Irys platform. The technology is based on visualisation of enzyme-specific nicking sites distributed along the DNA molecules hundreds to thousands of kilobases in length. For the 7DS, a genome map was constructed using Nt.BspQI nickase and the final assembly consisted of 371 fragments with average length of 0.9 Mb and N50 of 1.3 Mb. Alignment of the sequences of 7DS MTP BAC clones to the genome map enabled to verify the physical map assembly, size gaps and scaffold contigs. Comparison of the physical map assembly with the BioNano map also allowed anchoring, orientating, and ordering contigs along genome maps. Our results suggest that BioNano genome map provides a valuable tool to support, validate and improve the physical map and sequence assembly in the complex wheat genome.

## BioNano genome mapping on Irys platform

- BioNano mapping is based on single-molecule mapping of stretched DNA molecules hundreds to thousands of kilobases in length
- Sequence-motif maps providing long-range template for ordering genomic sequences are created

### Sequence-specific labelling

- DNA is nicked and fluorescently labelled at specific sequence motifs

5´-ATGG**GCTCTTC**CATGAATGCG-3´ → 5´-ATGG**GCTCTTC**CATGAAUGCG-3´
3´-TACC**CGAGAAG**GTACTTACGC-5´ → 3´-TACC**CGAGAAG**GTACTTACGC-5´
Nt.BspQ1 nicking site

ATGAATGC

- Whole DNA molecule is labelled by Yoyo dye (background colour)

labells
dsDNA

### DNA linearization and fluorescence imaging

- Labelled DNA is loaded on the Irys chip containing nanochannel array
- DNA molecules are uncoiled in a pillar region and driven into the nanochannels
- Nanochannels assure equal linearization and stretching of the DNA molecules
- Fluorescently labelled DNA is imaged by Irys instrument

pillar region   nanochannels
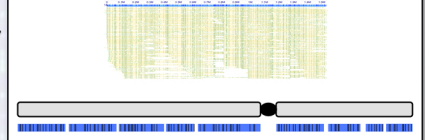
Irys instrument
Irys chip
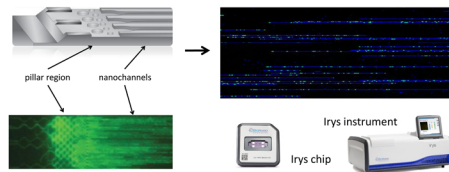
### Building consensus genome map

- Nicking pattern is created for each DNA molecule

- Consensus maps are built based on overlaps between nicking patterns of individual DNA molecules
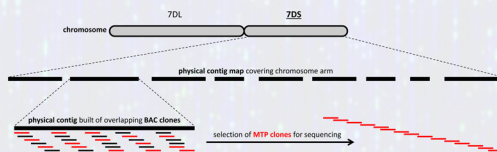
## 7DS physical map

### Assembly

- Automatic assembly by FPC software
- Validation of the assembly by LTC software
- Final assembly
  → 931 contigs, 4608 MTP BAC clones

### MTP sequencing

- Sequencing of pools containing 4 BACs
- Illumina pair-end + mate-pair sequencing
- Sequences assembled by Sassy and Sspace software

### Anchoring

- 73 % of the map (527 contigs) were anchored by 1885 markers

chromosome   7DL   7DS
physical contig map covering chromosome arm
physical contig built of overlapping BAC clones
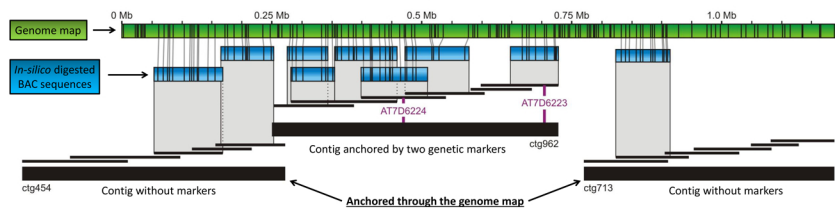selection of **MTP clones** for sequencing

## 7DS BioNano genome map

- Map was constructed as described by Hastie *et al.* (2013) from flow-sorted 7DS arm after digestion with Nt.BspQI (~12 sites per 100kb)
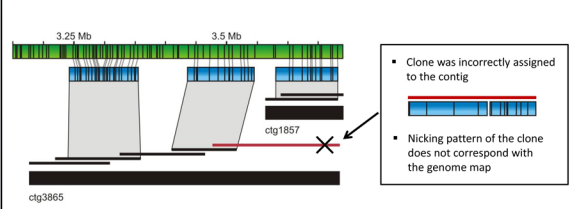
| Parameters of the map | | | |
|---|---|---|---|
| Map size | 350 Mb | Coverage | 94% |
| No. of genome maps | 371 | Map N50 | 1.3 Mb |

- Alignment of sequences of 7DS MTP BAC clones to genome map enabled to
  → anchor and order contigs of the physical map
  → identify potential overlaps and size gaps between contigs
  → identify miss-assemblies in the physical map
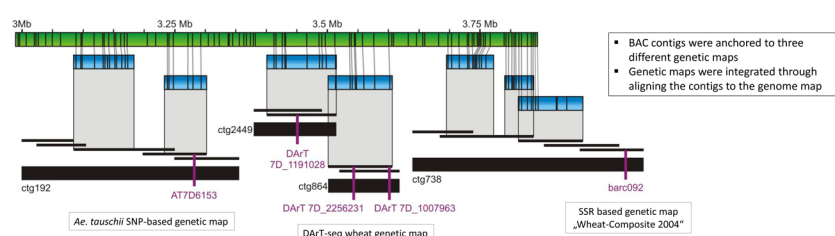  → integrate different genetic maps through the physical and BioNano map

### Anchoring of contigs without markers

0 Mb   0.25 Mb   0.5 Mb   0.75 Mb   1.0 Mb
Genome map
*In-silico* digested BAC sequences
AT7D6224   AT7D6223
Contig anchored by two genetic markers   ctg962
ctg454
Contig without markers   ctg713   Contig without markers
**Anchored through the genome map**

### Identification of a miss-assembly

3.25 Mb   3.5 Mb
- Clone was incorrectly assigned to the contig
ctg1857
- Nicking pattern of the clone does not correspond with the genome map
ctg3865

### Integration of different genetic maps

3Mb   3.25 Mb   3.5 Mb   3.75 Mb
- BAC contigs were anchored to three different genetic maps
- Genetic maps were integrated through aligning the contigs to the genome map
ctg2449
DArT 7D_1191028
ctg192   ctg864   ctg738   barc092
AT7D6153   DArT 7D_2256231   DArT 7D_1007963
*Ae. tauschii* SNP-based genetic map   SSR based genetic map „Wheat-Composite 2004"
DArT-seq wheat genetic map

## Conclusions

- Genome map proved useful for the improvement of the physical contig map of the 7DS chromosome arm by
  → ordering contigs without marker
  → ordering contigs in non-recombining regions
  → integrating various maps
  → identifying miss-assemblies
- Genome map will help in validating sequence assembly, sizing gaps between sequence contigs and finalizing and validation of the 7DS pseudomolecule

**Reference:** Hastie AR, Dong L, Smith A et al. (2013) Rapid genome mapping in nanochannel arrays for highly complete and accurate de novo sequence assembly of the complex *Aegilops tauschii* genome. PLoS One. 8:e55864.

**Palacký University Olomouc**

**Faculty of Science**

**Department of Botany**

**and**

**Institute of Experimental Botany**

**Centre of the Region Haná for Biotechnological and Agricultural Research**

**Olomouc**

**Helena Staňková**

# Construction of physical map of 7DS wheat chromosome arm and its use for positional cloning

**P1527 Biology - Botany**

**Summary of Ph.D. Thesis**

**Olomouc 2015**

Ph.D. thesis was carried out at the Department of Botany, Faculty of Science, Palacký University Olomouc, between the years 2010-2015.

Ph.D. candidate:       **Mgr. Helena Staňková**


Supervisor:         **Ing. Hana Šimková, CSc.**
Institute of Experimental Botany, Olomouc

Reviewers:          **Ing. Petr Smýkal, Ph.D.**
Palacký University, Olomouc

**RNDr. Aleš Kovařík, CSc.**
Institute of Biophysics, Brno

**RNDr. Helena Štorchová, CSc.**
Institute of Experimental Botany, Praha

The evaluation of this Ph.D. thesis was written by Prof. Ing. Aleš Lebeda, DrSc., Department of Botany, Faculty of Science, Palacký University Olomouc.

The summary of the Ph.D. thesis was sent for distribution on …………………

The oral defense will take place on ………………… before the Commission for the Ph.D. thesis of the Study Program Botany in a conference room of the Institute of Experimental Botany, Šlechtitelů 31, Olomouc-Holice.

The Ph.D. thesis is available in the Library of the Biological Departments of the Faculty of Science at Palacký University Olomouc, Šlechtitelů 11, Olomouc-Holice.

**Prof. Ing. Aleš Lebeda, DrSc.**
Chairman of the Commission for the Ph.D.
Thesis of the Study Program Botany
Department of  Botany, Faculty of Science
Palacký University Olomouc

# CONTENT

# 1    INTRODUCTION

Bread wheat (*Triticum aestivum* L.) belongs among the most important agricultural crops worldwide, along with rice and maize. It has been one of the first domesticated crop species which is nowadays cultivated on more land than any other crop. Hexaploid genome of bread wheat arose from multiple successive spontaneous hybridization events between three different diploid progenitor species from *Poaceae* family. Each hybridisation was followed by a polyploidization. This process resulted in a huge ~17-Gb genome (Bennett and Smith 1976) formed of three homoeologous subgenomes. Moreover, the majority of the genome (>80 %) consists of repetitive elements (IWGSC 2014). All those above mentioned characteristics greatly hamper wheat genome analysis. Dissection of the genome into individual chromosomes and chromosome arms using flow-cytometry sorting provided an elegant solution to these problems  and made the wheat genome research including genetic and physical mapping, positional cloning and sequencing, a feasible task (Doležel *et al.* 2007).

Obtaining a high quality reference sequence of the wheat genome poses an essential step for the thorough characterisation of the genome.  Complete reference sequence will provide access to a gene catalogue and regulatory elements and enable to determine their function and role in the wheat development. It will also allow to investigate the genome structure, organization and evolution, and to exploit the natural and induced genetic diversity of wheat. This knowledge will be highly valuable for the accelerated development of improved wheat varieties needed in the future (Feuillet *et al.* 2011; Choulet *et al.* 2014).

All the features of the wheat genome have to be considered when selecting the appropriate sequencing strategy. In general, two main approaches are followed in genome sequencing projects – the whole genome shotgun (WGS) and the clone-by-clone (CBC) sequencing. These strategies can be applied independently or can be combined in order to enhance the sequence quality and to increase the probability of capturing the maximum information (Green 2001). Several wheat sequencing projects were based on the WGS approach (Brenchley *et al.* 2012; Chapman *et al.* 2015). Although those studies provided a rough insight into the genome composition, they also showed that a reference sequence cannot be obtain through the WGS approach and that the genome have to be dissected into smaller manageable pieces. Individual fractions of the genome (chromosomes, arms) obtained by flow-cytometry sorting can be sequenced separately applying either whole chromosome shotgun or clone-by-clone (BAC-by-BAC) sequencing approach. The chromosomal approach was adopted by the International Wheat

Genome Sequencing Consortium (IWGSC) as a key strategy towards obtaining a reference sequence of wheat (Choulet *et al.* 2014).

Recently, the chromosome-based draft sequence of the wheat genome produced by the whole chromosome shotgun sequencing was released (IWGSC 2014). Despite the fact that the whole chromosome shotgun sequencing offered a quick view into a genome composition it is unequivocal that chromosome-based BAC-by-BAC sequencing strategy represents a favourable approach towards production of a reference wheat genome sequence. BAC-by-BAC sequencing is a several-step procedure which begins with the construction of a BAC library from genomic or, in this case, chromosomal high-molecular weight (HMW) DNA. All clones in the library are fingerprinted and physical contig map is constructed based on overlaps between the fingerprints. Then, the minimal set of clones covering the chromosome (minimal tilling path, MTP) is selected and the MTP clones are shotgun sequenced using next-generation sequencing (NGS) technologies. Final steps of the sequencing are the sequence assembly and construction of a pseudomolecule (Choulet *et al.* 2014). A validation of the physical contig map as well as the sequence assembly is essential for the production of a high quality reference sequence.

Several agronomically important genes have been identified on the short arm of chromosome 7D (7DS) of wheat, including the Russian wheat aphid resistance gene *Dn2401* (Fazel-Najafabadi *et al.* 2015). Russian wheat aphid (RWA; *Diuraphis noxia*, Kurdjumov) is a serious invasive pest of wheat and barley. It spread in all important wheat and barley growing areas with the only exception of Australia and it causes considerable economical losses every year. The use of resistant varieties seems to be the most effective, economical and environmentally safe mean of controlling this serious pest (Liu *et al.* 2002). Positional (or map-based) cloning is a process of the isolation of the gene underlying trait of interest without prior knowledge of the gene sequence and product. Once isolated, the gene can be transferred to the plant lacking the trait of interest which would results, in this case, in a production of RWA resistant wheat variety. The availability of a physical contig map of the 7DS arm as well as the reference sequence provides an invaluable tool to facilitate positional cloning of the *Dn2401* resistance gene.

# 2 AIMS OF THE THESIS

**I Construction and validation of the physical contig map of the wheat chromosome arm 7DS**

First aim of the thesis was to construct a physical contig map for the short arm of the wheat chromosome 7D, select and sequence minimal tilling path (MTP), and finally validate the physical map by an alternative mapping technology.

**II Construction of a high-density genetic map in *Dn2401* gene region**

The second aim of the thesis was to construct a high-density genetic map in the region of an aphid resistance gene - *Dn2401*. This key step in the positional cloning process should enable to span the region by a physical map contig and potentially identify candidate gene(s).

# 3 MATERIALS AND METHODS

## 3.1 Construction and validation of the physical contig map of the wheat chromosome arm 7DS

**Plant material**

Chromosome arm 7DS was flow-sorted from double-ditelosomic line of bread wheat (*Triticum aestivum* L.) cv. 'Chinese Spring' carrying both arms of chromosome 7D as telosomes.

**Construction of a 7DS-specific physical contig map**

7DS-specific BAC library (Šimková *et al.* 2011) was used as a genomic resource for the construction of a 7DS-specific physical contig map. BAC library was fingerprinted using the SNaPshot-based high-information-content fingerprinting technology (Luo *et al.* 2003). A fingerprint-based assembly of the physical map was performed using FPC (FingerPrinted Contig) software (Soderlund *et al.* 2000). Resulting assembly was verified by LTC (Linear Topological Contig) software (Frenkel *et al.* 2010). A minimal tiling path was selected from the physical map and clones representing the MTP were sequenced using a pooling strategy in which 96 pools, each consisting of four BACs, were indexed and sequenced on a single lane using the Illumina Hiseq 2000 platform. Sequences were de-multiplexed and assembled using the SASSY assembler (Kazakoff *et al.* 2012). Deconvolution was supported by BAC-end sequences generated from the MTP BAC clones by Sanger sequencing.

**Construction of a BioNano genome map of 7DS chromosome arm**

HMW DNA of the flow-sorted 7DS arm embedded in agarose miniplugs was used for the construction of a BioNano genome map. The DNA was released from the plugs by GELase treatment and labelled at Nt.BspQI nicking sites using the IrysPrep Reagent Kit (BioNano Genomics). The backbone of the DNA molecules was stained with IrysPrep DNA Stain (BioNano Genomics). The labelled and stained DNA was loaded on the Irys chip v2 and analysed by Irys instrument (BioNano Genomics).

After single molecules were detected to find the label positions on the DNA backbone, de novo assembly was performed by a pairwise comparison of all single molecules and graph building (Cao *et al.* 2014).

**Validation of a physical contig map**

Validation of the physical contig map was performed by comparison of sequences of MTP BAC clones with the 7DS-specific BioNano genome map. Comparison was done using the IrysView 2.0 software package (BioNano Genomics). Prior to comparison, *cmap* files were generated from *fasta* files of individual sequences or BAC pools. Query-to-anchor comparison was performed with default parameters and variable p-value threshold ranging from $1e^{-6}$ to $1e^{-10}$, based on the type of analysis.

## 3.2 Construction of a high-density genetic map in *Dn2401* gene region

**Plant material**

Genetic mapping was carried out in a mapping population of 184 F2 plants derived from a cross between RWA resistant wheat line CI2401 and susceptible cultivar 'Glupro'.

**Development of gene-associated markers in hexaploid genome**

Markers specific for the *Dn2401* gene region were developed by two approaches, both employing collinearity with related grass genomes. The first approach was based on the information provided by the barley GenomeZipper (Mayer *et al.* 2011). A syntenic region delimited by markers flanking the *Dn2401* gene was identified in the Zipper. Sequences of barley ESTs belonging to this interval were compared with assembled 7AS, 7BS and 7DS sequences of cv. 'Chinese Spring' (Berkman *et al.* 2013) using BLASTn algorithm. EST-harbouring sequence contigs were identified in 7AS, 7BS and 7DS sequence assemblies. These contigs were aligned using GeniousPro software (Drummond *et al.* 2010) in order to identify polymorphisms between 7AS, 7BS and 7DS sequences. To assure a subgenome-specific amplification, at least one primer from the primer pair was designed in the polymorphic region. Primer3 v.0.4.0 software was used for the primer design with default parameters (Rozen and Skaletsky 2000).

An alternative approach employing synteny between wheat and *Aegilops tauschii* was applied for the second set of markers. An interval in gene-based linkage map of *Ae. tauschii* (Luo *et al.* 2013) collinear to the *Dn2401* gene region was identified. Sequences of markers from this region were compared with the 7AS, 7BS and 7DS assemblies using BLASTn, and corresponding sequence

contigs were thus identified. The contigs were aligned and primers were designed as described above.

To verify specificity of the primers, amplified DNA of flow-sorted chromosome arms 7AS, 7BS and 7DS was used as PCR template. 7DS-specific primers were subsequently used for the amplification of the DNA of parents of the mapping population (CI2401 and 'Glupro'). PCR products were sequenced and polymorphisms were identified.

**Mapping of genetic markers and screening of the BAC library**

Mapping of the newly developed SNP markers was performed by amplicon sequencing on 184 plants of the mapping population. Genotyping data were analysed using JoinMap 4.0 (van Ooijen and Voorrips 2001) applying the Kosambi function (Kosambi 1944).

Mapped markers were used to screen the 7DS-specific BAC library as described in Šimková *et al.* (2011) and contigs from the 7DS physical map were identified through positive BAC clones.

# 4    SUMMARY OF RESULTS

This thesis focuses on the study of the short arm of the wheat chromosome 7D (7DS). The first goal of this work was the construction and validation of the 7DS physical contig map. The second part of the thesis is dedicated to the positional cloning of the Russian wheat aphid resistance gene *Dn2401* located on the 7DS chromosome arm, particularly construction of a high-density genetic map covering the gene region.

## 4.1    Construction and validation of the physical contig map of the wheat chromosome arm 7DS

Construction of the physical contig map represents one of the essential steps towards obtaining the reference sequence of the genome through the BAC-by-BAC sequencing strategy. Aiming to deliver a reference sequence of the 7DS chromosome arm, the physical contig map of this chromosome arm was constructed. The final assembly consisted of 931 contigs. In total, 4,608 BAC clones representing minimal tilling path were selected and sequenced in pools of four clones, yielding 17 contigs per pool on average. Part of the contigs were assigned to particular clones through available BAC-end sequences. In order to prove the quality and correctness of the physical map assembly, a 7DS-specific BioNano genome map was constructed using Irys platform and a comparison with the sequences of BAC clones representing the 7DS MTP was performed. The comparison showed that the BioNano mapping technology represents a useful tool for the validation and improvement of the physical map. It was demonstrated that integration of the physical and BioNano genome mapping can facilitate scaffolding, anchoring and ordering of contigs constituting the physical map. Based on the presented results, it is obvious that the BioNano mapping technology offers a missing tool needed to complement the extant genomics tools to deliver high quality reference genome sequence.

## 4.2    Construction of a high-density genetic map in *Dn2401* gene region

*Dn2401*, a gene underlying resistance to Russian wheat aphid, is one of the agronomically important genes located on the 7DS chromosome arm. Aiming to perform positional cloning of this gene, a high-density genetic map around the gene was constructed using a newly developed method for targeted development of highly specific markers in the hexaploid wheat genome.  The method benefits from the

synteny with close relatives of wheat and chromosome shotgun sequences of a complete wheat homoeologous group. Applying the presented approach, 11 new gene-associated markers were developed and used for the saturation of the genetic map in the region of interest. The interval between markers flanking the gene was reduced to 0.83 cM. Screening of a 7DS-specific BAC library with the flanking markers allowed identification of four MTP BAC clones spanning the *Dn2401* region in cv. 'Chinese Spring' representing one physical contig. Annotation of sequences of MTP BAC clones from this contig revealed presence of several candidate genes. The study proved that availability of the physical contig map can significantly facilitate and speed up the positional cloning process and that the physical map is an indispensable supplementary tool for positional cloning.

# 5    CONCLUSIONS

Within the framework of this thesis, I focused on the 7DS chromosome arm of hexaploid bread wheat aiming to construct a physical contig map of this chromosome arm and to make a step forward in positional cloning of the *Dn2401* resistance gene.

In the first part of this thesis, a 7DS-specific BAC-based physical contig map was constructed, MTP clones were sequenced and the physical map assembly was validated. This study proved that the novel technology, a BioNano genome mapping, provides a useful tool for the verification and improvement of the physical map and has a potential to guide a construction of a pseudomolecule. A quality physical map and a sequence assembly of MTP BAC clones represent indispensable steps towards obtaining a reference sequence of the 7DS chromosome arm of wheat.

In the second part of the thesis, a high-density genetic map of *Dn2401* gene region was constructed with the aim to isolate this gene by positional cloning. A new approach, enabling development of highly specific markers in hexaploid wheat genome, was presented. The described strategy is applicable for targeted marker development in any region of the wheat genome and thus can be useful in any positional cloning project conducted in wheat. Availability of a 7DS-specific physical contig map and sequences of MTP BAC clones greatly facilitated identification of candidate genes and thus accelerated the positional cloning process.

# 6    REFERENCES

Bennett MD, Smith J (1976) Nuclear DNA amounts in angiosperms. Phil Trans R Soc B 274:227-274

Berkman PJ, Visendi P, Lee HC, Stiller J, Manoli S, Lorenc MT, Lai K, Batley J, Fleury D, Šimková H, Kubaláková M, Weining S, Doležel J, Edwards D (2013) Dispersion and domestication shaped the genome of bread wheat. Plant Biotech J 11:564–571

Brenchley R, Spannagl M, Pfeifer M, Barker GL, D'Amore R, Allen AM, McKenzie N, Kramer M, Kerhornou A, Bolser D, Kay S, Waite D, Trick M, Bancroft I, Gu Y, Huo N, Luo MC, Sehgal S, Gill B, Kianian S, Anderson O, Kersey P, Dvorak J, McCombie WR, Hall A, Mayer KF, Edwards KJ, Bevan MW, Hall N (2012) Analysis of the bread wheat genome using whole-genome shotgun sequencing. Nature 491:705-710

Cao H, Hastie AR, Cao D, Lam ET, Sun Y, Huang H, Liu X, Lin L, Andrews W, Chan S, Huang S, Tong X, Requa M, Anantharaman T, Krogh A, Yang H, Cao H, Xu X (2014) Rapid detection of structural variation in a human genome using nanochannel-based genome mapping technology. Gigascience 3:34

Chapman JA, Mascher M, Buluç A, Barry K, Georganas E, Session A, Strnadova V, Jenkins J, Sehgal S, Oliker L, Schmutz J, Yelick KA, Scholz U, Waugh R, Poland JA, Muehlbauer GJ, Stein N, Rokhsar DS (2015) A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. Genome Biol 16:26

Choulet F, Caccamo M, Wright J, Alaux M, Šimková H, Šafář J, Leroy P, Doležel J, Rogers J, Eversole K, Feuillet C (2014) The Wheat black jack: Advances towards sequencing the 21 chromosomes of bread wheat. In: Tuberosa R, Graner A, Frison E (eds) Genomics of plant genetic resources. Volume 1. Managing, sequencing and mining genetic resources. pp 405-438 Springer Science+Business Media Dordrecht

Doležel J, Kubaláková M, Paux E, Bartoš J, Feuillet C (2007) Chromosome-based genomics in the cereals. Chromosome Res 15:51-66

Drummond AJ, Ashton B, Buxton S, Cheung M, Cooper A, Heled J, Kearse M, Moir R, Stones-Havas S, Sturrock S, Thierer T, Wilson A (2010) Geneious v5.3. http://www.geneious.com

Fazel-Najafabadi M, Peng J, Peairs FB, Šimková H, Kilian A, Lapitan NLV (2015) Genetic mapping of resistance to *Diuraphis noxia* (Kurdjumov) biotype 2 in wheat (*Triticum aestivum* L) accession CI2401. Euphytica 203:607–614

Feuillet C, Leach JE, Rogers J, Schnable PS, Eversole K (2011) Crop genome sequencing: lessons and rationales. Trends Plant Sci 16:77–88

Frenkel Z, Paux E, Mester D, Feuillet C, Korol A (2010) LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. BMC Genomics 11:584

Green ED (2001) Strategies for the systematic sequencing of complex genomes. Nat Rev Genet 2:573-583

IWGSC - International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345:1251788

Kazakoff SH, Imelfort M, Edwards D, Koehorst J, Biswas B, Batley J, Scott PT, Gresshoff PM (2012) Capturing the biofuel wellhead and powerhouse: the chloroplast and mitochondrial genomes of the leguminous feed-stock tree *Pongamia pinnata*. PLoS One 7:e51687

Kosambi D (1944) The estimation of map distances from recombination values. Ann Eugen 12:172–175

Liu XM, Smith CM, Gill BS (2002) Identification of microsatellite markers linked to Russian wheat aphid resistance genes *Dn4* and *Dn6*. Theor Appl Genet 104:1042-1048

Luo MC, Thomas C, You FM, Hsiao J, Ouyang S, Buell CR, Malandro M, McGuire PE, Anderson OD, Dvorak J (2003) High-throughput fingerprinting of bacterial artificial chromosomes using

the SNaPshot labeling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378-389

Luo MC, Gu YQ, You FM, Deal KR, Ma Y, Hu Y, Huo N, Wang Y, Wang J, Chen S, Jorgensen CM, Zhang Y, McGuire PE, Pasternak S, Stein JC, Ware D, Kramer M, McCombie WR, Kianian SF, Martis MM, Mayer KF, Sehgal SK, Li W, Gill BS, Bevan MW, Šimková H, Doležel J, Weining S, Lazo GR, Anderson OD, Dvorak J (2013) A 4-gigabase physical map unlocks the structure and evolution of the complex genome of *Aegilops tauschii*, the wheat D-genome progenitor. Proc Natl Acad Sci USA 110:7940-7945

Mayer KF, Martis M, Hedley PE, Šimková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, Kubaláková M, Suchánková P, Murat F, Felder M, Nussbaumer T, Graner A, Salse J, Endo T, Sakai H, Tanaka T, Itoh T, Sato K, Platzer M, Matsumoto T, Scholz U, Doležel J, Waugh R, Stein N (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263

Rozen S, Skaletsky HJ (2000) Primer3 on the WWW for general users and for biologist programmers. In: Krawetz S, Misener S (eds) Bioinformatics methods and protocols: methods in molecular biology. pp 365–386 Humana Press Totowa

Šimková H, Šafář J, Kubaláková M, Suchánková P, Číhalíková J, Robert-Quatre H, Azhaguvel P, Weng Y, Peng J, Lapitan NL, Ma Y, You FM, Luo MC, Bartoš J, Doležel J (2011) BAC libraries from wheat chromosome 7D: Efficient tool for positional cloning of aphid resistance genes. J Biomed Biotechnol 2011:302543

Soderlund C, Humphray S, Dunham A, French L (2000) Contigs built with fingerprints, markers, and FPC V4.7. Genome Res 10:1772-1787

van Ooijen JW, Voorrips RE (2001) JoinMap 3.0, software for the calculation of genetic linkage maps. Plant J 3:739–744

# 7 LIST OF AUTHOR´S PUBLICATIONS

## 7.1 Original papers

Staňková H, Valárik M, Lapitan NL, Berkman PJ, Batley J, Edwards D, Luo MC, Tulpová Z, Kubaláková M, Stein N, Doležel J, Šimková H (2015) Chromosomal genomics facilitates fine mapping of a Russian wheat aphid resistance gene. Theor Appl Genet 128:1373-1383

Staňková H, Hastie AR, Chan S, Vrána J, Tulpová Z, Kubaláková M, Visendi P, Hayashi S, Luo MC, Batley J, Edwards D, Doležel J, Šimková H (accepted paper) BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. Plant Biotechnol J

## 7.2 Published abstracts – poster presentations

Staňková H, Valárik M, Šafář J, Lapitan N, Stein N, Berkman P, Edwards D, Doležel J, Šimková H (2011) Sequence-based tools facilitate high-density mapping of a Russian wheat aphid resistance gene in wheat. In: Abstracts of the „Olomouc Biotech 2011. Plant Biotechnology: Green for Good". Olomouc, Czech Republic.

Staňková H, Luo MC, Visendi P, Tulpová Z, Batley J, Bartoš J, Doležel J, Edwards D, Šimková H (2013) Physical map and sequencing of wheat chromosome arm 7DS. In: Abstracts of the „Olomouc Biotech 2013. Plant Biotechnology: Green for Good II". Olomouc, Czech Republic.

Staňková H, Valárik M, Lapitan N, Berkman P, Edwards D, Luo MC, Tulpová Z, Kubaláková M, Stein N, Doležel J, Šimková H (2014) Poziční klonování genu pro rezistenci k mšici zhoubné (*Diuraphis noxia*): Konstrukce vysokohustotní genetické mapy. In: Sborník abstrakt, Bulletin České společnosti experimentální biologie rostlin, 6. Metodické dny. Seč, Česká republika.

Staňková H, Hastie A, Cao H, Vrána J, Kubaláková M, Visendi P, Hayashi S, Luo MC, Batley J, Edwards D, Doležel J, Šimková H (2015) Improving the physical map of the wheat 7DS chromosome arm with a BioNano genome map. In: Abstracts of the „Olomouc Biotech 2015. Plant Biotechnology: Green for Good III". Olomouc, Czech Republic.

## 7.3 Published abstract – oral presentation

Staňková H, Hastie A, Vrána J, Kubaláková M, Visendi P, Hayashi S, Luo MC, Batley J, Edwards D, Doležel J, Šimková H (2015) BioNano genome mapping of flow-sorted chromosomes supports physical mapping and sequence assembly in complex plant genomes. In: Book of abstracts of the "V4 International Conferencce, Analytical Cytometry VIII". Olomouc, Czech Republic.

# 8    SUMMARY (in Czech)

**Konstrukce fyzické mapy krátkého ramene chromozómu 7D pšenice a její využití pro poziční klonování**

Pšenice setá (*Triticum aestivum* L.) představuje jednu z nejvýznamnějších zemědělských plodin, jež poskytuje potravu pro 30% světové populace. Studium genomu pšenice, zejména pak genetické a fyzické mapování, sekvenování a poziční klonování, je ztíženo značnou velikostí genomu (~17 Gb), přítomností tří homeologních subgenomů a převahou repetitivních sekvencí (>80 %). Řešení těchto problémů nabízí třídění jednotlivých chromozómů a jejich ramen pomocí průtokové cytometrie. Toto rozdělení genomu na malé přesně definované části je nezbytným krokem k získání referenční sekvence genomu pšenice.

Předkládaná práce se zabývá studiem krátkého ramene chromozómu 7D (7DS). V rámci práce byla zkonstruována fyzická kontigová mapa ramene 7DS. Ta posloužila k výběru tzv. *minimal tilling path* (MTP), tedy minimální sestavy klonů z knihovny dlouhých inzertů, která reprezentuje celé chromozómové rameno. Klony z MTP byly následně sekvenovány a získaná data se stala základem pro sestavení referenční sekvence ramene 7DS. Za účelem ověření správnosti fyzické mapy a tedy i výsledné chromozómové sekvence bylo provedeno srovnání s genomovou mapou vytvořenou technologií BioNano na platformě Irys. Toto srovnání prokázalo, že genomová mapa může posloužit jako užitečný nástroj pro ověřování a zdokonalení fyzické mapy a má také potenciál uplatnit se při analýze tandemových repetitivních sekvencí. Zmíněná technologie může dále poskytnout jeden z podpůrných nástrojů pro sestavení pseudomolekuly, tedy kontinuální sekvence celého chromozómového ramene.

Krátké rameno chromozómu 7D nese několik genů pro agronomicky významné znaky, mezi nimi i gen *Dn2401*, jež podmiňuje rezistenci k mšici zhoubné. Druhá část předkládané práce je věnována pozičnímu klonování uvedeného genu. Jedním ze základních kroků v procesu pozičního klonování je konstrukce vysoko-hustotní genetické mapy pro oblast klonovaného genu. S cílem zahustit genetickou mapu byl navržen postup pro cílený vývoj markerů v prostředí hexaploidního pšeničného genomu. Za použití této metody bylo vyvinuto několik nových markerů v těsné vazbě na gen. Pomocí nově vyvinutých markerů byl identifikován kontig fyzické mapy, jež překlenuje oblast genu. Sekvenování a následná anotace klonů ze získaného kontigu odhalila přítomnost několika kandidátních genů. Dostupnost fyzické kontigové mapy a sekvencí jednotlivých klonů z MTP značně zjednodušila proces pozičního klonování.