**Palacký University Olomouc**

**Faculty of Science**

**Department of Botany**


and


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

**Institute of Experimental Botany**

**Olomouc**


# Physical map of the wheat chromosome 4AL and positional cloning of a gene for yield


Ph.D. Thesis


**Mgr. Barbora Balcárková**


Olomouc 2017                    Supervisor: Mgr. Miroslav Valárik, Ph.D.

## Declaration:

I hereby declare that I elaborated this Ph.D. thesis independently under the supervision of Mgr. Miroslav Valárik Ph.D., using only information sources listed in the References.

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

# Bibliographic identification

| | |
|---|---|
| **Authors name:** | Mgr. Barbora Balcárková (Klocová) |
| **Title:** | Physical map of the wheat chromosome 4AL and positional cloning of a gene for yield |
| **Type of Thesis:** | Ph.D. thesis |
| **Department:** | Botany |
| **Supervisor:** | Mgr. Miroslav Valárik Ph.D. |
| **The Year of Presentation:** | 2017 |

**Abstract:**

Bread wheat (*Triticum aestivum* L.) is one of the most important crops worldwide. Wheat provides basic food for 35% of the world population and is grown in a wide range of environments, mainly in temperate zone. Its importance can be compared only with rice or maize. Fast growing human population dramatically increases importance on food production including wheat production.

Breeding of highly yielding wheat cultivars resistant to biotic and abiotic stresses is the main challenge for current breeding programs. Wheat genome is composed of three related diploid genomes (A, B and D) and contains more than 80 % of repetitive sequences. This great genome complexity makes breeding, molecular markers development, marker assisted selection (MAS) and wheat genome sequencing very complicated.

Reduction of the complexity by the flow sorting of chromosomes or chromosome arms greatly facilitates marker development and mapping. The sorting also provides sufficient quantity of high molecular weight DNA for survey sequencing and BAC library construction. Development of large numbers of chromosome specific markers allows construction of high density maps creating an ideal base for map based cloning of agronomically important genes.

The aim of the present Ph.D. thesis was to construct a sequencing-ready 4A chromosome-specific physical map and high density radiation hybrid map to use them for map saturation and cloning of the powdery mildew-resistance gene *QPm-tut-4A*, rust resistance gene (*Yr51*) and yield related gene *Qyi-4A-bga*.

| | |
|---|---|
| **Keywords:** | wheat, *Triticum aestivum*, physical map, radiation hybrid map, BAC library, positional cloning |
| **Number of Pages/Appendix:** | 61/6 |
| **Language:** | English |

# Bibliografická identifikace

**Jméno:** Mgr. Barbora Balcárková (Klocová)

**Téma:** Fyzická mapa chromozomu 4AL pšenice a poziční klonování genu ovlivňujícího výnos

**Typ práce:** Dizertační práce

**Obor:** Botanika

**Školitel:** Mgr. Miroslav Valárik Ph.D.

**Rok obhajoby:** 2017

**Abstrakt:**

Pšenice setá (*Triticum aestivum* L.) je jednou z nejdůležitějších plodin světa. Poskytuje základní potravu pro 35 % obyvatel Země a je pěstována v širokém rozmezí prostředí, hlavně v oblastech mírného pásu. Její důležitost se může srovnávat pouze s rýží nebo kukuřicí. Rychle rostoucí lidská populace klade stále větší nároky na produkci potravin a tedy i na pšenici.

Šlechtění výnosnějších a odolnějších odrůd pšenice je hlavní výzvou současných šlechtitelských programů. Genom pšenice se skládá ze tří homeologních subgenomů (A, B a D) a obsahuje více než 80 % repetitivních sekvencí. Velká komplexita genomu pšenice je hlavní překážkou šlechtění, efektivního vývoje molekulárních markerů, selekci za pomoci markeru (MAS, *Marker Assisted Selection*) a sekvenaci pšeničného genomu.

Jednou z možností, jak se vypořádat se složitostí pšeničného genomu, je použití jednotlivých chromozomů nebo jejich ramen tříděných průtokovou cytometrií, které jsou ideální základnou pro vývoj markerů a mapování s velkou hustotou. Třídění chromozomů anebo jejich ramen také poskytuje dostatečné množství vysokomolekulární DNA, která je využívána pro tzv. „survey" sekvenování a konstrukci knihoven DNA dlouhých inzertů (tzv. BAC knihoven).

Cílem dizertační práce byla konstrukce fyzické mapy dlouhého ramene chromozomu 4A pšenice a vysoko-hustotní radiační hybridní mapy tohoto chromosomu a jejich využití pro zahuštění genetické mapy klonování genů *QPm-tut-4A* zodpovědného za resistenci k padlí travnímu, *Yr51* odpovědného za rezistenci ke rzi a *Qyi-4A-bga* gene ovlivňujícího výnos.

**Klíčová slova:** pšenice, *Triticum aestivum*, fyzická mapa, radiační hybridní mapa, BAC knihovna, poziční klonování

**Počet stran/přílohy:** 61/6

**Jazyk:** Anglický

# Content

# 1    Introduction

Bread wheat (*Triticum aestivum* L.) is one of the most important crops worldwide. Wheat provides basic food for 35% of the population and is grown in a wide range of environments, mainly in temperate zone. Its importance can be compared only with rice or maize. Fast growing human population dramatically increases importance of food production including the production of wheat.

Breeding of highly yielding wheat cultivars resistant to biotic and abiotic stresses is the main challenge for current breeding programs. Achieving goals to generate more yielding cultivars is hampered by polyploid nature of wheat genome. The hexaploid genome of bread wheat (AABBDD) is result of two successive spontaneous hybridization events between three different diploid progenitors' species (*T. urartum*, AA; the most likely one of *Aegilops speltoides* from the *Sitopsis* group BB; goat grass *Ae. tauschii*, DD) from *Poaceae* family. The hybridisation resulted in large 17 Gb allohexaploid genome (Shewry, 2009; Matsuoka, 2011; Peng *et al*., 2011). All of the sub-genomes are closely related and similar to each other. Moreover, the majority of the genome (over 80 %) comprises repetitive elements (Moore 1995; Choulet *et al*., 2014a).

One of the options, how to overcome the wheat genome complexity, is utilization of its wild diploid ancestors like *T. urartum* and *T. monococcum* related to wheat genome A or *Aegilops tauschii* related to wheat genome D. Using this approach, a one third complexity reduction can be achieved. Another option represents usage of individual wheat flow sorted chromosomes or chromosome arms. This approach offers genome reduction up to fifty times. The flow sorting of chromosomes or chromosome arms provides sufficient quantity of high molecular weight DNA for survey sequencing and BAC library construction (Šafář *et al*., 2010). This approach also allowed distributing the workload for the wheat genome sequencing to different laboratories organized in the International Wheat Genome Sequencing Consortium (IWGSC, www.wheatgenome.org). In this approach, the chromosome specific physical maps and their Minimal Tilling Paths (MTPs) are crucial and serve as a basis for efficient gene mapping and cloning.

Only, MTPs of oriented and anchored physical maps can provide full utilization of their potential (Meyers *et al*., 2004; Paux *et al*., 2008a; Ariyadasa *et* Stein, 2012). However, genome regions with limited recombination lack information about marker order

and their utilization in physical map anchoring is limited. Even the GenomeZipper (GZ) uses for prediction of gene order synteny with sequenced grass genomes of rice, sorghum and *Brachypodium*, as a back bone was used recombination map (Mayer *et al*., 2011). The GZ for all wheat chromosomes are currently available (IWGSC, 2014) but their use is limited because at least one third of the wheat genome, especially in centromere and peri-centromeric regions, is located in recombination poor regions. Recombination mapping in such regions does not offer sufficient resolution for marker ordering (Akhunov *et al*., 2003). Ideal solution of this limitation offers Radiation hybrid (RH) mapping which is based on radiation-induced deletions and markers ordering is independent from recombination (Michalak de Jimenez *et al*., 2013). RH mapping was originally used to map human genome (Goss *et* Harris, 1975) and adapted to wheat (Hossain *et al*., 2004). At present the RH mapping together with high-throughput genotyping techniques (e.g. SNP chips, DArTs, POPSEQ, …) becomes a powerful tool for high-resolution mapping in wheat. RH maps were already used for anchoring of 3B and 6B physical maps (Paux *et al*., 2008b; Kobayashi *et al*., 2015).

The physical maps even without sequencing can significantly facilitate positional cloning of agronomically important genes because the final step in the cloning procedure is construction of physical map of cloned gene locus (Keller *et al*., 2005). The contigs of physical maps can also facilitate identification of markers which are tightly flanking the cloned genes (Feuillet *et al*., 2003). Such markers could be directly used in breeding process for MAS of desirable genes in breeding materials and their progenies.

# 2    Literature review

## 2.1    *Wheat (in general – genus Triticum)*

Wheat (in general) is the third most important cereal worldwide after rice and maize. It is intensively grown on 17 % of the total crop area from 67°N in Scandinavia and Russia to 45°S in Argentina, including tropics and sub-tropics, thanks to its enormous genetic diversity. Over 25 000 different cultivars and varieties of wheat are known which allows wheat to be grown over all climates (Shewry, 2009; Peng *et al.*, 2011).

In these days, 95 % of the world production is represented mostly by hexaploid bread wheat. Remaining 5 % of production is generated by tetraploid wheat, which is used for pasta production. Small amount of other varieties of wheat (spelta and einkorn) is grown in certain regions of Spain, Turkey, Balkans and India. This production is mainly used for cattle feeding (Shewry, 2009).

Wheat popularity and large expansion is caused by one very important characteristic. The wheat flour can be simply transformed into wide range of foods like various kinds of bread, cakes, pasta, noodles, and biscuits etc. This wheat property stems from storage proteins known as gluten proteins (Gustafson *et al.*, 2009; Shewry, 2009).

### 2.1.1    Systematic classification

Cultivated wheat and its wild relatives belong to the genus *Triticum* L., a member of the tribe *Triticeae*, which contains 300 species including genus *Hordeum* (barley), *Aegilops* (goat grass) and *Secale* (rye). Genus *Triticum* is divided in five sections based on polyploidy level, genome compositions, comparative genetic and mapping (Table 1). The section *Compositum* contains wheat species that were made by man. The remaining four sections vary in ploidy level and genome content (Goncharov, 2011; Matsuoka, 2011).

Interestingly, *T. urartu* exists only as a wild form, while *T. aestivum* and *T. zhukovkyi* are known exclusively as cultivated species. On the other hand, *T. monococcum*, *T. turgidum* and *T. timopheevii* are known in both forms, wild as well as cultivated. All wheat accessions originate in Fertile Crescent (Matsuoka, 2011).

**Table 1:** *Triticum* classification

Individual sub-genomes of wheat are differentiated according their origin. Genome $A^b$ comes from *T. boeticum*; genome $A^u$ originates from *T. urartum*; genome B from wild goat grass related to section *Sitopsis*; genome D *A. tauschii*; genome G from wild goat grass related to section *Sitopsis* (according Goncharov, 2011)

| Section | Group of specie | Species | 2n | Genomes |
|---|---|---|---|---|
| 1. *Monococcon* Dum. | Hulled | | | |
| | | *T. urartu* Thum. ex Gandil | 14 | $A^u A^u$ |
| diploid | | *T. boeticum* Boiss. | 14 | $A^b A^b$ |
| | | *T. monococcum* L. | 14 | $A^b A^b$ |
| | Naked | *T. sinskajae* A. Filat. Et Kurk. | 14 | $A^b A^b$ |
| 2. *Dicoccoides* Flaksb. | Hulled | *T. dicoccoides* (Körn. Ex Aschers. Et Graebn.) Schweinf. | 28 | $A^u A^u BB$ |
| | | *T. dicoccum* (Schrank) Schuebl. | 28 | $A^u A^u BB$ |
| tetraploid | Naked | *T. turgidum* L. | 28 | $A^u A^u BB$ |
| | | *T. durum* Desf. | 28 | $A^u A^u BB$ |
| | | *T. polonicum* L. | 28 | $A^u A^u BB$ |
| 3. *Triticum* | Hulled | *T. spelta* L. | 42 | $A^u A^u BBDD$ |
| hexaploid | Naked | *T. compactum* Host | 42 | $A^u A^u BBDD$ |
| | | *T. aestivum* L. | 42 | $A^u A^u BBDD$ |
| 4. *Timopheevii* A.Filat et Dorof. | Hulled | | | |
| | | *T. araraticum* Jakubz. | 28 | $A^u A^u GG$ |
| | | *T. timopheevii* (Zhuk.) Zhuk. | 28 | $A^u A^u GG$ |
| | | *T. zhukovskyi* Menabde et Erizjan | 42 | $A^b A^b A^u A^u GG$ |
| 5. *Compositum* N.P.Gontsch. | Hulled | | | |
| | | *T. soveticum* Zhebrak | 56 | $A^u A^u A^u A^u BBGG$ |
| manmade | | *T. borisii* Zhebrak | 70 | $A^u A^u A^u A^u BBDDGG$ |
| | Naked | *T. flaksbergeri* Navr. | 56 | $A^u A^u A^u A^u BBGG$ |

## 2.1.2 Evolution and domestication of wheat

The evolution of bread wheat started 300 000 – 500 000 years ago when wild diploid wheat *T. urartu* (genome $A^u A^u$, 2n = 2x = 14) hybridized with the B-genome ancestor that was closely related to section *Sitopsis* and the goat grass *Aegilops speltoides* Tausch. (genome SS, 2n = 2x = 14; Haider, 2013). This hybridization event gave rise to wild emmer wheat (*T. dicoccoides*, genome $A^u A^u BB$, 2n = 4x = 28). About 10 000 years ago hunters and gatherers in Fertile Crescent started with cultivation of wild emmer wheat (Figure 1). This was a big event of Neolithic revolution, when hunters and gatherers changed eating habits and transformed to settled agriculture.
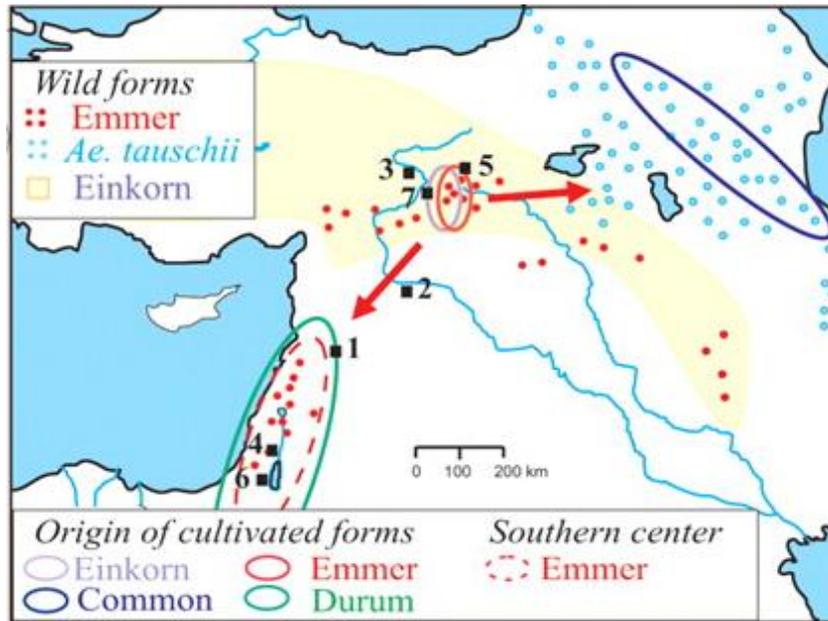
**Figure 1:** The origin of wheat

The solid line ovals represent the putative geographic regions of origin. Dots represent possible distribution of wild emmer and *Ae. tauschii*, whereas yellow area indicate distribution of wild *T. boeoticum*. Red arrows suggest spreading of cultivated emmer wheat. Numbers indicate archaeological sites of domestication (according to Dubcovsky *et* Dvorak, 2007).

The first cultivation of wheat occurred in south-eastern part of Turkey. The earliest cultivated forms of diploid wild einkorn and tetraploid emmer wheat were selected by farmers from wild populations. This was the first non-scientific breeding. Farmers selected wheat with greater yield. Already cultivated emmer wheat has spread to place of *Aegilops tauschii* (genome DD, $2n = 2x = 14$) origin in Near East (Figure 1). In about 9 000 years ago the second hybridization occurred between cultivated emmer wheat and wild goat grass *Ae. tauschaii* and early spelt (*T. spelta*, hexaploid genome $A^uA^uBBDD$, $2n = 6x = 42$) appeared. Most likely, this hybridization occurred independently and more than once in different places (Shewry, 2009; Matsuoka, 2011; Peng *et al*., 2011). About 8 500 years ago, natural mutation changed the ears of both, emmer and spelt to a more easily threshed type that later evolved into the free-threshing ears of durum wheat (*T. durum*) and bread wheat (*T. aestivum*; Figure 2; Peng *et al*., 2011). However, Dvořak *et al*. (2006) suggest that *T. spelta* is not the ancestor form of the free-threshing common wheat.
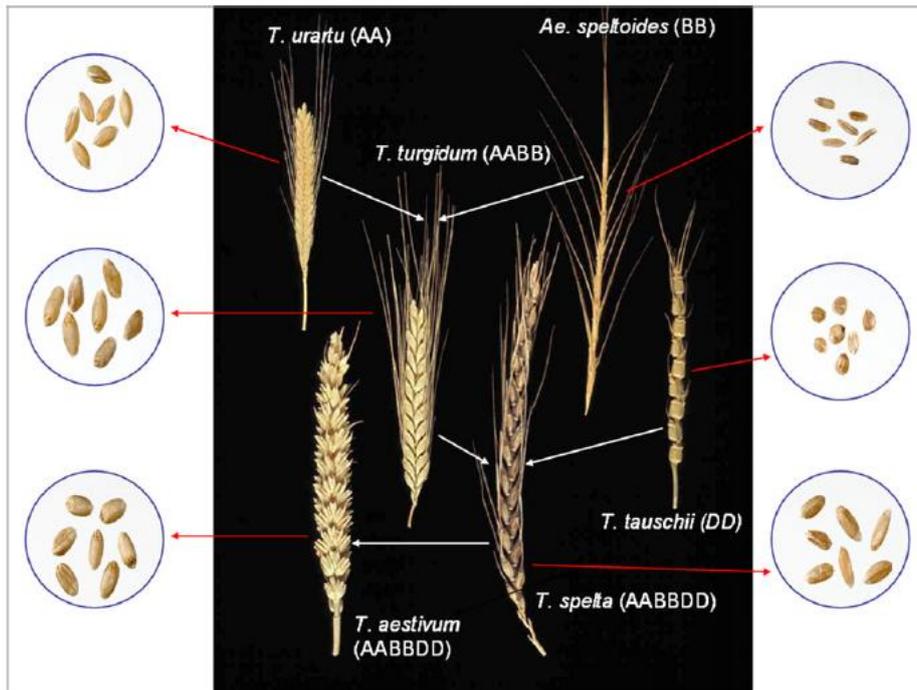
**Figure 2:** Scheme of bread wheat evolution (adapted from Shewry, 2009)

With the ongoing transformation of wild wheat to cultivated form, wheat lost several undesirable characteristics like brittle rachis which led to easy shattering of spikes at maturity and tightly hulled grains which hampered trashing. Unconsciously, farmers harvested only wheat which had non-shattering spikes. Farmers also preferred easily threshed grains (Figure 3). All these aspects caused that wild wheat (wild emmer and spelt) converted into cultivated wheat and since then it is no more able to survive in nature without help of man (Shewry, 2009; Matsuoka, 2011).
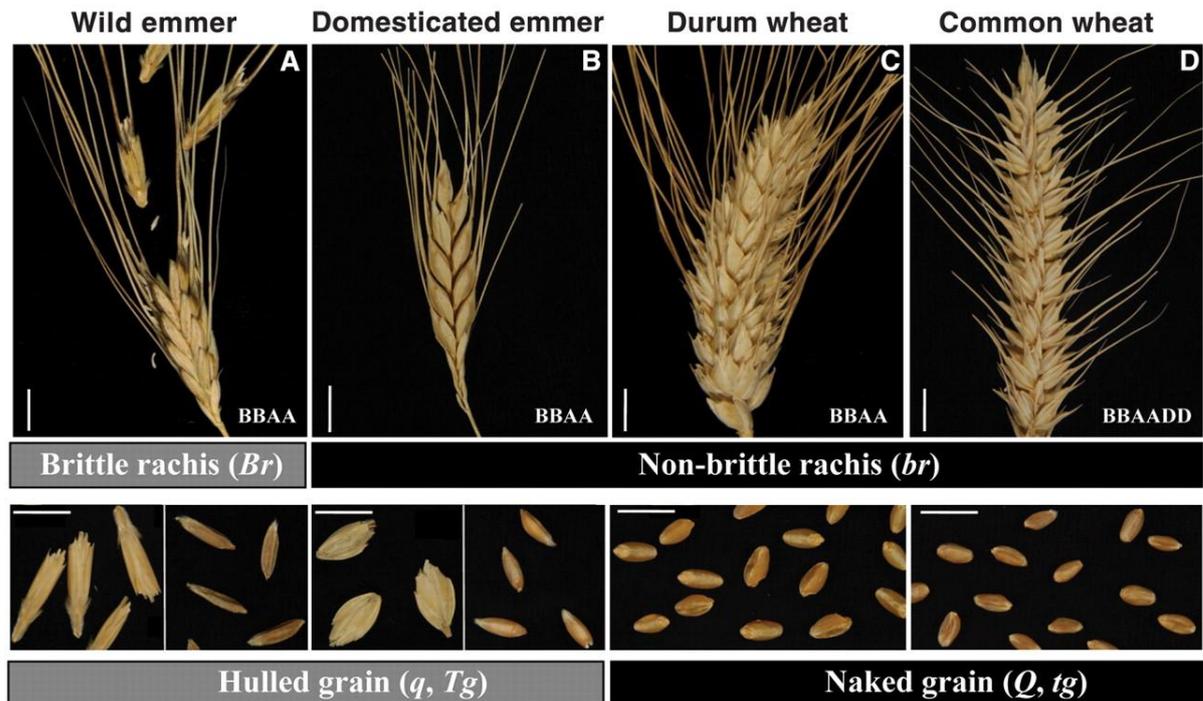
**Figure 3:** Wheat spikes

Wheat spikes showing (A) brittle rachis, (B to D) non-brittle rachis, (A and B) hulled grain, and (C and D) naked grain. (A) Wild emmer wheat (*T. dicoccoides*), (B) domesticated emmer (*T. dicoccum*), (C) durum (*T. durum*), and (D) common wheat (*T. aestivum*). White scale bars represent 1 cm. Letters at the lower right corner indicate the genome formula of each type of wheat. Gene symbols: *Br* brittle rachis, *Tg* tenacious glumes, and *Q* square head (reproduced from Dubcovsky *et* Dvorak 2007)

## 2.1.3 Wheat genome

Bread wheat, as most of plants (70% of plant species; reviewed in Wendel, 2000) is a polyploid species. Its allopolyploid genome (as mentioned above) originated from hybridization followed by polyploidization (Feldman *et* Levy, 2009). Polyploidy means, that two or more genomes are present in one nucleus. Based on the origin of each sub-genome, a distinction between autopolyploidy and allopolyploidy can be made. Wheat is allopolyploid species and its individual sub-genomes have arisen from different, but very closely related genomes (Feldman *et* Levy, 2009; Bento *et al*., 2011).

Bread wheat genome contains 42 chromosomes (2n = 6x = 42), seven pairs of chromosomes from each of the individual sub-genomes. Therefore, wheat genome is very complex and large (1C ~ 17 Gb; Bennett *et* Smith, 1976). Compared to rice, the wheat genome is about 40 times larger and has very high repetitive sequences content (> 80 %, Smith

*et* Flavell, 1975). The wheat genome is organized into two main parts with different evolutionary dynamics (Choulet *et al.*, 2014a).

- The first part is relatively small conserved corresponding mostly to gene space which is subject to a preservative selection pressure.
- The second part is much larger and more variable comprising repetitive and transposable elements (TE) with more dynamic evolution.

**The gene space**

The first molecular studies of wheat genome were carried out with the help of EST (expressed sequence tag) and mRNA (messenger RNA) sequences which were hybridized on a series of aneuploid lines (Qi *et al.*, 2004). Based on this study it was hypothesized that genes are located in a so called "gene-rich" regions. These "gene-rich" regions were found in recombination hot-spots in distal parts of the wheat chromosomes. It was also observed that gene density is increasing towards to the chromosome ends (Sandhu *et* Gill, 2002; Qi *et al.*, 2004). Almost 30 % of the wheat genome represents these "gene-rich" regions. The rest of the genome, considered as "gene-poor" would be composed mostly of repetitive sequences, TEs and a few isolated genes.

Over the past decade, large increase in the availability of BAC libraries and physical maps for individual wheat chromosomes (or chromosome arms) allowed sequencing of a substantial portion of the wheat BAC clones and BAC contigs. Based on the partial sequence of chromosome 3B (Choulet *et al.*, 2010) and *Ae. tauschii* genome (Massa *et al.*, 2011) sequence, it was assumed that genes are present along the whole chromosome and are clustered mainly into numerous small islands of 3-4 genes. These islands are detached by large blocks of TEs that are less than 1 Mb long (Choulet *et al.*, 2014a).

Nowadays, analysis of Chromosome Survey Sequences (CSSs) of individual chromosome arms allows a new and complex perspective on wheat genome and genes content. CSSs were obtained by flow-cytometric purification of individual chromosome arms and by the sequencing with Illumina with coverage 30 x to 241 x (IWGSC, 2014). CSSs enable to compare sub-genomes of bread wheat to each other and also to related species that were already sequenced, such as *T. urartu* (Ling *et al.*, 2013), *T. monococcum* (IWGSC, 2014), *Ae. tauschii* (Jia *et al.*, 2013), *T. turgidum* and *Ae. Speltoides* (IWGSC, 2014). It was estimated that the wheat genome contains 106 000 functional protein-coding genes.

This supports gene number estimates ranging between 32 000 and 38 000 for each diploid subgenome (IWGSC, 2014).

**The transposable elements (TE) space**

Wheat genome is composed of more than 83 % of repeated sequences. This was observed using reassociation kinetics of single-stranded DNA fragments (Flavell *et al*., 1977). Choulet *et al*. (2010) sequenced several BACs and BAC contigs and confirmed that the TE fraction represents more than 80 % of the wheat genome. Major part of TE fraction is represented by class I LTR retrotransposons. Approximately 50 % are representing by less than 10 families (Fatima, Jorge, Angela, Laura, Sabrina, WIS, Wilma and Nusif) and different families are found in centromeric and telomeric regions (Choulet *et al*., 2010).

Analysis of CSSs showed that 76,6 % of assembled sequences contained repeats. Moreover, analysis of the distribution of transposons across the three subgenomes revealed that class I retrotransposons were more abundant in the A genome chromosomes relative to B or D, whereas the opposite was true for the class II DNA transposons (IWGSC, 2014).

The TE content was also confirmed by analysis and annotation of chromosome 3B sequence. Choulet *et al*. (2014b) found 53 288 complete and 181 058 truncated copies of TEs representing 85 % of the 3B pseudomolecule.

## *2.2 Construction of physical map*

Construction of a whole-genome physical map has been an essential component of numerous genome sequencing projects initiated since the beginning of The Human Genome Project (Collins *et* Galas, 1993). A physical map is an ordered set of DNA fragments, among which the distances are expressed in physical distance units (base pairs). A physical map usually comprises a set of ordered large-insert clones such as BACs as the preferred building blocks of a physical map. A physical map is a model of a genome of interest (Meyers *et al*., 2004; Ariyadasa *et* Stein, 2012). Principally, physical maps simplify the task of positional cloning, genome sequencing and the targeted development of genetic markers.

BAC library of wheat whole-genome would contain about 1.2 million of clones. Doležel *et al*. (2007) pioneered a method how to slice the large wheat genome into smaller and easy-to-handle parts represented by individual chromosomes or chromosome arms. Wheat

chromosome arms are flow sorted, high molecular weight DNA (HMW DNA) is subsequently extracted and chromosome arm specific BAC libraries are prepared. Besides the rapid decrease in the complexity of the wheat genome, the BAC libraries of individual chromosome arms also facilitate dividing the huge task of sequencing the whole wheat genome between different laboratories in the world, hence splitting the workload and enabling faster and easier achievement of the goal (Feuillet *et* Eversole, 2008). The laboratories are organized under the IWGSC (www.wheatgenome.org) and contributions of independent countries to the wheat chromosome specific physical map construction and genome sequencing effort are marked in Figure 4.
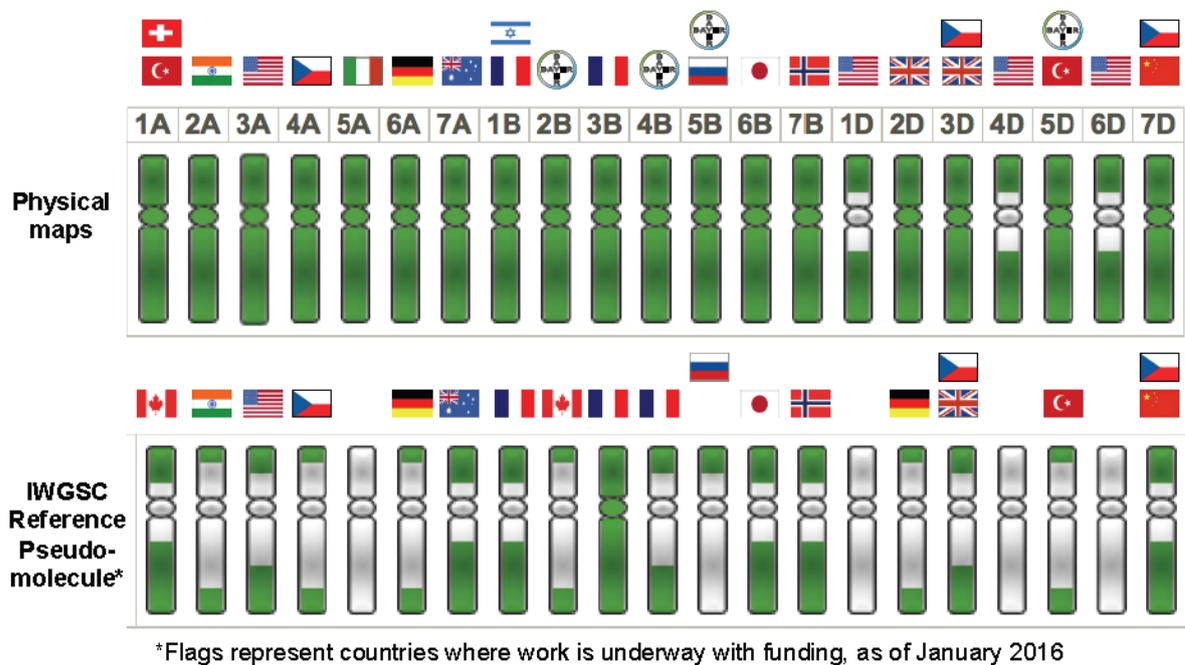


**Figure 4:** Scheme of distribution of the wheat genome mapping and sequencing efforts between worldwide laboratories.

Upper – distribution of projects on construction of physical maps of individual wheat chromosomes. Bottom - distribution of individual wheat chromosomes sequencing efforts. Flags represent countries of interest. Green colour of chromosome represents the percentage of what has already been achieved (adapted from www.wheatgenome.org).

## 2.2.1 Chromosome specific physical maps

The first step of chromosome specific physical map preparation is chromosome arm sorting. This step is necessary for genome size and complexity reduction. Wheat genome

can be reduced down to 1,3 % of the total genome size depending on chromosome arm size (Table 2; Šafář *et al*., 2010). DNA of sorted chromosome is extracted as intact and high quality (HMW) DNA so it can be used for next step BAC library construction (Šimková *et al*., 2011).

**Table 2:** Relative and molecular sizes of the hexaploid wheat chromosomes and their arms.

Calculated based on relative chromosome lengths (Gill 1987), 4C nuclear DNA amount 69,27 pg (Bennett *et* Smith, 1976), and considering 1 pg DNA = 0,978 Mbp (Doležel *et al*., 2003).

| Chromosome | Whole chromosome | | Long arm | | Short arm | |
| | Genome fraction (%) | Size (Mbp/1C) | Genome fraction (%) | Size (Mbp/1C) | Genome fraction (%) | Size (Mbp/1C) |
|---|---|---|---|---|---|---|
| 1A | 4.715 | 799 | 3.089 | 523 | 1.626 | 275 |
| 2A | 5.310 | 899 | 3.001 | 508 | 2.309 | 391 |
| 3A | 4.885 | 827 | 2.761 | 468 | 2.124 | 360 |
| 4A | 5.055 | 856 | 3.183 | 539 | 1.872 | 317 |
| 5A | 4.885 | 827 | 3.141 | 532 | 1.745 | 295 |
| 6A | 4.163 | 705 | 2.181 | 369 | 1.983 | 336 |
| 7A | 4.800 | 813 | 2.400 | 407 | 2.400 | 407 |
| 1B | 5.013 | 849 | 3.156 | 535 | 1.856 | 314 |
| 2B | 5.480 | 928 | 2.989 | 506 | 2.491 | 422 |
| 3B | 5.862 | 993 | 3.314 | 561 | 2.549 | 432 |
| 4B | 4.843 | 820 | 2.537 | 430 | 2.306 | 391 |
| 5B | 5.140 | 871 | 3.427 | 580 | 1.713 | 290 |
| 6B | 5.395 | 914 | 2.943 | 498 | 2.452 | 415 |
| 7B | 5.310 | 899 | 3.186 | 540 | 2.124 | 360 |
| 1D | 3.568 | 604 | 2.247 | 381 | 1.322 | 224 |
| 2D | 4.291 | 727 | 2.425 | 411 | 1.865 | 316 |
| 3D | 4.545 | 770 | 2.652 | 449 | 1.894 | 321 |
| 4D | 3.823 | 648 | 2.458 | 416 | 1.365 | 231 |
| 5D | 4.418 | 748 | 2.895 | 490 | 1.523 | 258 |
| 6D | 4.206 | 712 | 2.294 | 389 | 1.912 | 324 |
| 7D | 4.291 | 727 | 2.043 | 346 | 2.247 | 381 |
| Sum | 100.000 | 16937 | 58.322 | 9878 | 41.678 | 7060 |

The process of BAC library construction starts with digestion of flow-sorted chromosome DNA with restriction endonuclease *HindIII*, followed by size selection of DNA fragments. DNA fragments with the average size of 120 kbp are then ligated into pIndigoBAC-5 (*HindIII*) cloning vector. The resulting constructs are transformed into *E. coli* competent cells.

Last step of BAC library construction is ordering BAC clones into 386-well plates and storage at -80 °C (Šimková *et al*., 2011).

Before the physical map assembly, the individual clones of the BAC library are fingerprinted. Fingerprinting is based on the assumption, that clones digested with endonuclases produce unique set of fragments. Clones that are overlapping share part of their unique set of fragments (Soderlund *et al*., 1997). The most advanced fingerprinting technique used for BAC clones utilizes SNaPshot labelling kit (Luo *et al*., 2003). DNA from BAC clones are digested with five carefully selected endonucleases, four rarely and one often digesting nuclease. The often digesting nuclease was combined with each rarely digesting nuclease to produce short DNA fragments. The rarely digesting nucleases produce nucleotide overhangs which are labelled by the SNaPshot kit using single base extension of fluorescent labelled nucleotide. The labels are specific for each rarely digesting nuclease. Fragments are then separated and read by ABI DNA analyser. The obtained data are cleaned from background and contamination and only High Information Content Fingerprints (HICF) are kept for physical maps assembly (Luo *et al*., 2003).

Alternatively, the Whole Genome Profiling (WGP™) can be used to generate HICFs (Poursarebani *et al*., 2014; Kobayashi *et al*., 2015). BAC clones are pooled and then digested be restriction enzyme. Resulting fragments are ligated to adaptors and sequenced by Illumina. Pair-wise comparisons of tags allows identifying BAC overlaps and assembling contigs using physical map assembling softwares (Philippe *et al*., 2013).

## 2.2.2  Physical map assembly

There are two main approaches to the construction of physical map. The first is the FPC (FingerPrinted Contigs, Soderlund *et al*., 1997) and was developed for assembling, viewing and editing contigs of human genome (Gregory *et al*., 1996). FPC was used also for assembly of *Arabidopsis thaliana* (Marra *et al*., 1999) and rice (Chen *et al.*, 2002) physical maps. The second approach is Linear Topological Contig (LTC) and was created recently especially for complex genomes with high amount of repeat sequences, such as bread wheat and barley (Frenkel *et al*., 2010).

**FPC (FingerPrinted Contigs)**

The FPC algorithm generates an approximate CB (consensus band) map which can be considered as a low-resolution restriction map, it does not try to resolve all bands (Figure 5). CB maps are built using a fast approximation algorithm (hybrid greedy/stochastic algorithm) and inexact data. To produce this map program goes through four basic steps (Soderlund *et al.*, 1997; Engler *et al.*, 2003).
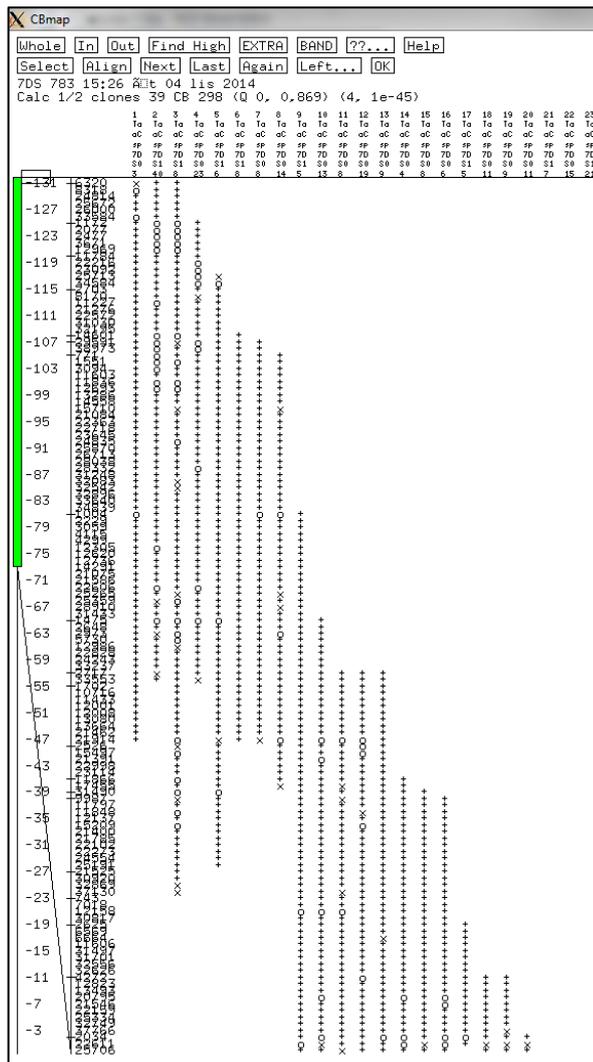


**Figure 5:** Consensus Bands (CB) map displayed in FPC

The consensus bands are shown along the left. The tick marks represent partially ordered groups. The columns represent the clones. A '+' indicates a match with the band to the left within the tolerance, a 'x' indicates a match within twice the tolerance and a 'o' indicates no match (adapted from Engler *et* Soderlund, 2006).

i.    Compare two clones

In the first step, FPC is searching for overlapping clones. In contig assembly systems, contigs of overlapping clones are represented by an artificial coordinate system based on the clones' band data. The bands of clones can be very close to each other. The tolerance determines how close two bands in one clone must be to consider them as the same band. Each clone in a contig has a left and right coordinate and the overlap between two clones corresponds to the number of bands they share, which is calculated with the Sulston formula (Sulston *et al*., 1988). The resulting score is compared to a given threshold called "cut-off". Cut-off is a probability score expressed in scientific notation ($1e^{-5}$). The higher the exponent is the lower the score and *vice versa*, the lower the score, the higher is the stringency. Cut-off discriminates between true overlaps (below the cut-off) and false overlaps resulting from randomly shared bands (above the cut-off; Soderlund *et al*., 1997; Engler *et* Soderlund, 2002).

For the wheat physical map assembly, the cut-off was set to a very stringent value ($1e^{-75}$; Paux *et al*., 2008b).due to the very high content of repeats in wheat genome. Higher stringency prevents false positive overlaps and generation of chimerical contigs (Apples *et al*., 2010).

ii.    Bury clones

In this step, FPC buried shorter clone in the longer clone given that all the bands of the shorter clone are a subset of the longer clone. The shorter clones are labelled as buried clones and they identify good clones, meaning that the bands are probably correct and not just coincidental. Clones without an existing overlap with any other clone are not placed in any contig and are called singletons (Soderlund *et al*., 1997; Soderlund *et al*., 2000).

iii.    Build a consensus band map (CB map)

BAC clones are ordered to contigs using CB maps (Figure 5). FPC is using hybrid greedy/stochastic algorithm to build the initial map, and then greedily extends the map. Two clones are declared as adjacent if they have a coincidence score that is below the user-defined cut-off. If there is a severe problem aligning a clone to the CB map, it is marked as a Q (questionable) clone. The Q-clones are exactly defined as clones for which the CB algorithm cannot order at least 50 % of the bands in CB map (Appels *et al*, 2010). If there are many Q-clones in the contig, the ordering is most likely incorrect. The contig must

be recalculated with more stringent cut-off and rebuild or split in to smaller contigs (DQer tool; Sonderlund *et al*., 1997; Sonderlund *et al*., 2000; Engler *et* Soderlung, 2002).

iv. MTP (Minimal Tiling Path) clones selection

The last step of physical map assembly is selection of MTP clones. These clones represent a minimal subset that covers the whole genome with defined overlaps. It is used for anchoring the physical map and as a template for sequencing.

MTP picking was automated and firstly included in the version V7 of the FPC programme. The selection was based on draft sequence and BES (BAC-end sequences). In the later version of FPC (V8), selection of MTP clones was based on HICF. MTP clones picking is based on the assumption that clones that are close to each other in the physical map have a greater likelihood of overlapping than those that are distant. Candidate MTP pairs (overlapping BACs) are preselected based on user defined number of shared bands. For each MTP pair a "spanner" and two flanking clones ("flankers") are selected (Figure 6). The pair of overlapping MTP clones must be supported by a "spanner" clone which supports presence of the clones overlap and two "flankers" which determine if the contig continues or not. (Nelson *et* Soderlund, 2009).
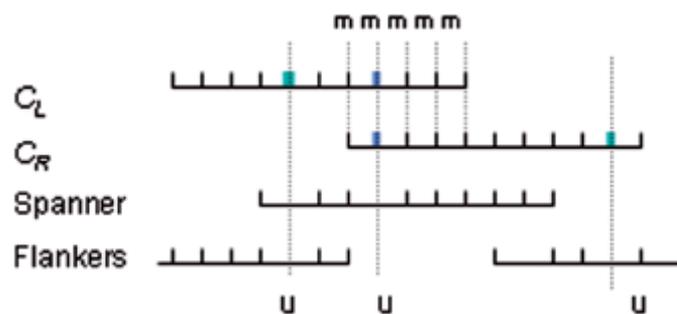


**Figure 6:** MTP clone pair, with confirming spanner and flankers.

Horizontal lines represent clones, vertical lines represent shared bands. $C_L$ and $C_R$ represent the MTP clone pair. Light blue highlight bands which are not in "flankers", dark blue mark bands which are not in "spanner" (adapted from Nelson *et* Soderlund, 2009).

**LTC (Linear Topological Contig)**

Second assembly software, the LTC (Linear Topology Contigs) developed by Frenkel *et al*. (2010) provides a novel approach how to order and build contigs of physical maps. The approach is based on presumption that the chromosome structure is linear, so the BAC

contigs have to be also linear. In contrast to FPC, the LTC starts clustering with a relatively relaxed cut-off value ($e^{-15}$) to build a net of significant clone overlaps and uses increasing cut-off and the topology testing (linearity) of significant clones overlap to get longer contigs with a realistic structure. Algorithms of LTC take into account the structure and complexity of cereals genomes.

The contig construction has eight steps, which are described in detail as follows.

### i.    Net of significant clone overlaps

In the first step, LTC creates 'nets of significant clone overlaps'. Because LTC algorithm was developed especially for genomes with high amount of repetitions, the "nets of significant clone overlaps" is created and assembly initiated on a very relaxed cut-off (e.g. $1e^{-15}$, low stringency; Frenkel *et al.*, 2010). The program calculates p-values for all the clone overlaps. Then the algorithm declares for a given cut-off (same criterion as in FPC) two clones that are significantly overlapped. Unlike FPC that used the Sulston formula (Soderlund *et al.*, 1997) for the p-value calculation, the LTC algorithm applies the $Pr^{(Siid)}$ and $Pr^{(Sind)}$ metrics. These two models enable p-values to be calculated more accurately and can take into account available markers, if there are any. $Pr^{(Siid)}$ model assumes that appearances of bands of different lengths are independent and identically distributed. Whereas the $Pr^{(Sind)}$ is based on the same hypothesis but bands may not be identically distributed (Frenkel *et al.*, 2010).

### ii.    Temporary exclusion of the clones and clone overlaps unproven by parallel paths

In the subsequent step, the LTC excludes putatively false significant overlaps and putatively problematic clones before clustering (questionable clones and questionable overlaps). This step is based on the assumption that each part of the chromosome is most probably covered by more than two significantly overlapped clones. This means that there will be parallel paths for each clone and overlaps, but for chimerical clones (Q-clones) and false overlaps (Q-overlap) there will be no such overlapping clone or parallel path, respectively (Figure 7; Arratia *et al.*, 1991; Wang *et al.*, 1994).

The program searches for supporting parallel significant overlaps for each overlap. If there is no such supporting overlap (Figure 7, red edge), program labels this overlap as questionable. The same method is applied to each clone (Figure 7, red node). Significant overlaps and clones that do not have parallel paths are temporary excluded from clustering. However this does not imply that such clones and overlaps are false necessarily. Likely,

an excluded clone is not necessarily chimerical. They can be used later to merge or elongate contigs (Frenkel *et al*., 2010).
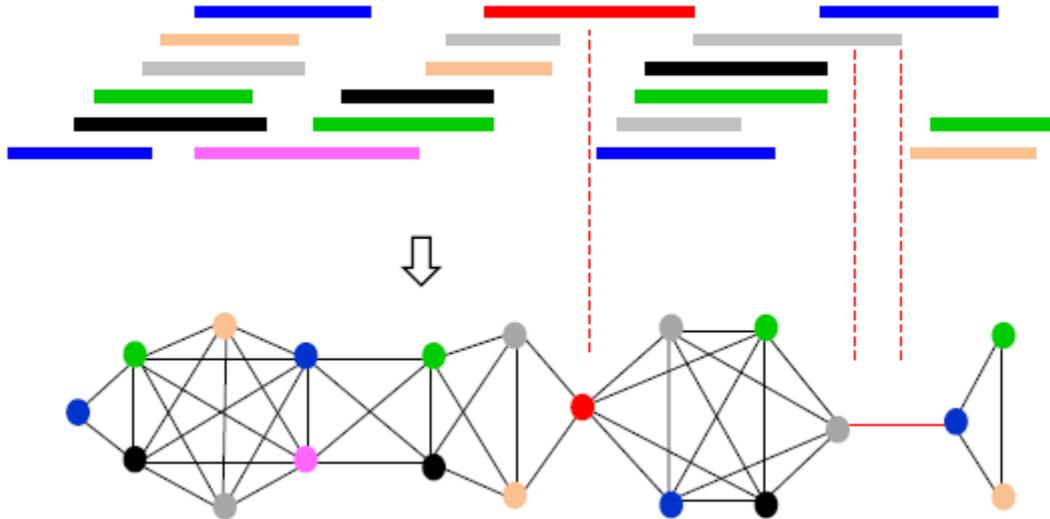


**Figure 7:** Network representation of clone overlaps

Top: Scheme of physical clone overlaps. Bottom: Network representation of the overlapping clones (nodes) and clone overlaps (edges). Colors are used to clarify the correspondence between physical and network representations of clone overlaps. Weak connections (Q-clones and Q-overlaps) caused by low coverage are in red. Q-clone = questionable clone; Q-overlap = questionable overlap (reproduced from Raats *et al*., 2013).

iii.     Adaptively changing cut-off clustering

In this step LTC adaptively raises cut-off and repeatedly passes through four steps to get reasonably sized clusters.

Circle of adaptive clustering starts with excluding putatively false clone overlaps and putative chimerical clones (described in previous step). Then clustering is initiated by the single-linkage algorithm (Jain *et* Dubes, 1988), clusters are created at a given cut-off. Subsequently, LTC selects reasonably sized clusters, which are optimally defined as clusters with 6 - 500 clones. The last step of the circle is increasing the cut-off stringency and next round of adaptive clustering starts. The process of adaptively changing cut-off clustering is illustrated on Figure 8 (Frenkel *et al*., 2010).
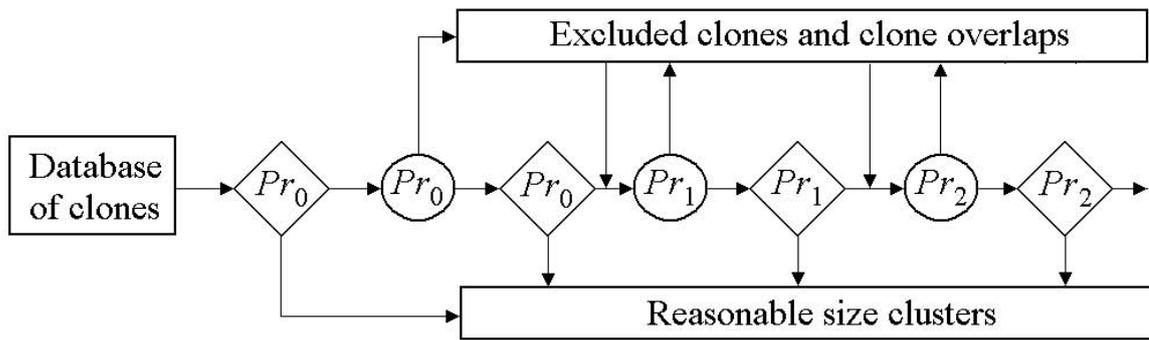
**Figure 8:** Scheme of clone clustering with adapting cut-off

Diamonds denote single-linkage clustering with corresponding cut-off. Circles denotes procedure of excluding clones and clones overlaps not approved by parallel paths in the net of significant (relative to the corresponding cut-off) clone overlaps from the analysis (reproduced from Frenkel *et al.*, 2010)

iv.     Looking for linear topological structure

DNA in chromosomes is a linear structure, so the BAC clusters should also be linear. In this step LTC is searching for non-linear clusters and splits them into smaller and linear clusters by excluding clones responsible for branching. LTC employs a representation of clusters as nets of significant overlaps. The longest path of the net is scored as rank = 0 (as is shown in Figure 9). The presence of clones with rank 2 and higher indicates non-linear structure (Frenkel *et al.*, 2010).
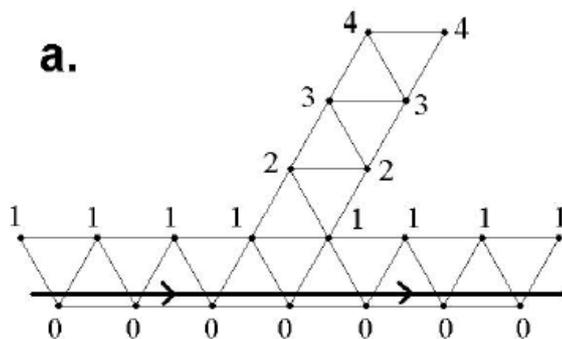


**Figure 9:** Net of significant overlaps of branching cluster.

Nodes denote clones, vertex denotes overlaps. Red nodes represent longest path and rank 0. Blue and grey notes represent rank 2 and 3, respectively and indicate branching of cluster (adapted from Frenkel *et al.*,2010).

v.       Global optimization ordering

LTC orders clusters with linear topological structures without constructing their band map, in contrast with FPC which is using CB maps for ordering. Here, the ordering problems are formulated in terms of global optimization of some criterion. The maximization of such criterion can be formulated as the well-known and intensively studied Travelling Salesman Problem (TSP) without the need to return to the starting point. TSP is mathematical problem, which seeks the most effective route among a particular number of cities (clones). Using such global optimization approach may result in a reduced number of Q-clones and number of places in the contig where two neighbouring clones have an unexpected non-significant overlap (Frenkel *et al*., 2010).

vi.      Verification of the orders by re-sampling

In this step, LTC is testing the contig stability. To evaluate this, LTC uses jack-knife re-sampling iterations (Mester *et al*., 2003). Firstly, LTC scores clones overlaps over all bands and then it orders clone overlaps using only a subsets of bands. Excluding parallel clones allows the construction of a stable skeleton map. This approach is similar to the one used in the MultiPoint software used for genetic map building (Mester *et al*., 2003; Frenkel *et al*., 2010).

vii.     Sub-contig merging into contigs

After the previous step, LTC tries to elongate contigs by merging those that has significant end-to-end overlaps or by adding singletons. LTC reanalyses all clones that were excluded in previous steps. Subsequently, LTC is searching for all clones connected with the clones from both contig ends. Such elongation is accepted only if adding singletons does not lead to contig branching (Frenkel *et al*., 2010).

viii.    Selection of MTP clones

Selection of MTP clones is ruled by the topology of clone overlaps and additional three requirements (Frenkel *et al*., 2010):
1)   Terminal clones of the contigs should be included in the list of MTP clones.
2)   Neighbouring clones in the MTP should significantly overlap at the chosen threshold.
3)   List of MTP clones should include the minimal number of clones.

**Comparison of assembling softwares**

FPC was used for building a physical map for 3B chromosome of wheat (Paux *et al*., 2008b). It was a pioneering work on the first physical map of wheat. That time, FPC was the only software for physical map assembly. Following physical maps of other chromosome arms of wheat (Lucas *et al*., 2013; Breen *et al*., 2013; Raats *et al*., 2013; Philippe *et al*., 2013; Poursarebani *et al*., 2014; Kobayashi *et al*., 2015; Akpinar *et al*, 2015; Barabaschi *et al*., 2015) were assembled by both softwares. They compared these two programs and observed that clusters form LTC assemblies are longer and cover greater length of chromosome arm.

LTC differs from FPC in these features:
- Calculation of metrics of clones overlaps (p-value)
- Consideration of the band abundances
- Algorithm for calculation of clone clustering
- Usage of topological cluster structure for contig construction
- Global optimization methods (TSP) are used for clone ordering in LTC
- Re-sampling is used for rating the credibility of the map

The above listed comparison applies for wheat chromosome arm BAC libraries, however, FPC may prove to be more suitable in physical map assembly for other species (Frenkel *et al*., 2010).

### 2.2.3 Anchoring of physical map to genetic map

The full potential of physical maps, which will be used for positional cloning, marker development and sequencing, can only be achieved by contig linkage to high resolution genetic map. Thus, once physical maps have been assembled, it is essential that the contigs are anchored to high resolution genetic map by molecular markers. Contigs of physical maps are anchored to genetic map by screening BAC libraries by molecular markers (Feuillet *et al*., 2012). Wheat BAC libraries contain tens of thousands of clones. To perform PCR whit every single clone would be extremely time consuming and expensive. The more effective way is using multidimensional BAC pools which allow identifying of BAC clone/clones positive for tested marker in one step and limited number of PCR reactions. The three dimensional pool (3D pools) are, the most common for wheat physical maps .The 3D pools comprise DNAs representing individual plates (dimension 1), DNA from clones of each particular col-

umn of all plates (dimension 2) and DNA from clones of each particular row of all plates (dimension 3) (Bruno *et al.*, 1995; Ariyadasa *et* Stein, 2012); e.g. the 3D pool of eight 384-well microtiter plates would contain 48 DNA pools (Figure 10; Paux *et al.*, 2008b).
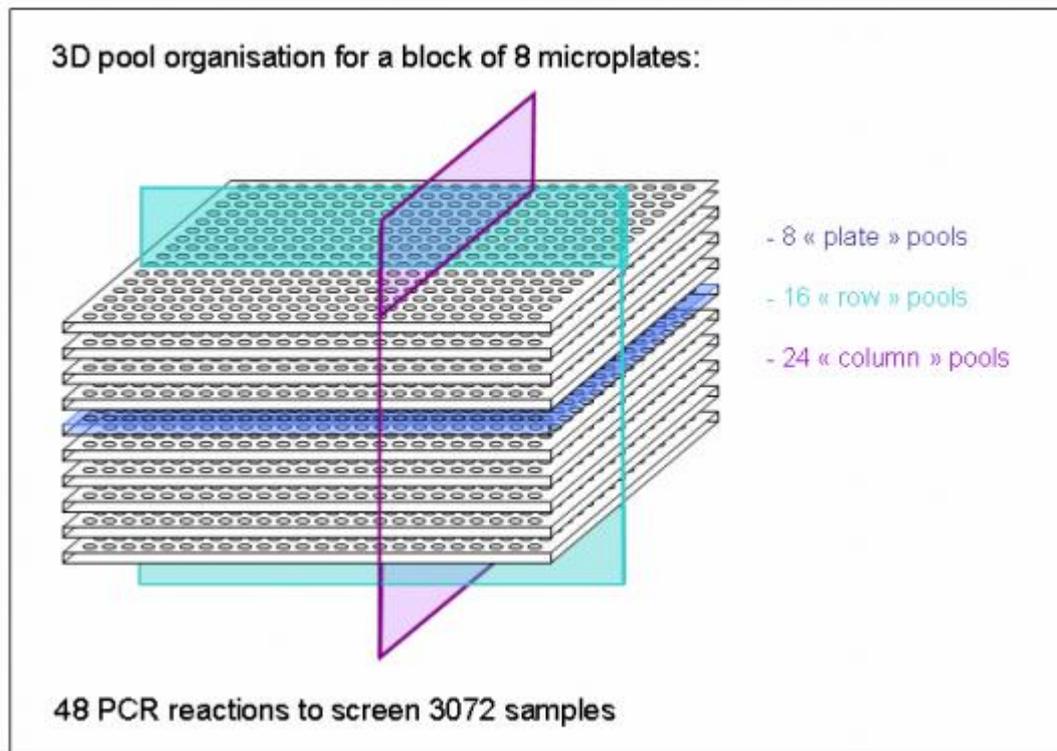


**Figure 10:** Scheme of 3D pooling of BAC library.
Adapted from http://cnrgv.toulouse.inra.fr/Services/DNA-Pool-production

Strategies of BAC pool screening and anchoring are summarized in Figure 11. Approaches of wheat physical map integrating can be divided into experimental and *in silico* procedures. Screening of BAC pools by molecular markers ranks among the experimental approaches. However, this approach has one big disadvantage. It is necessary to screen BAC pools with thousands of markers to orientate and anchor every single contig, which is laborious and time consuming (Ariyadasa *et* Stein, 2012; Feuillet *et al.,* 2012).

Recently, the PCR screening was replaced by cheaper high-throughput arrays, with the wheat NimbleGen 40K Unigene microarray (Rustenholz *et al.*, 2011) being the most widely used for wheat physical map anchoring. This array was used for anchoring of 1BL (Philippe *et al.*, 2013), 1BS (Raats *et al.*, 2013), 1AS (Breen *et al.*, 2013) and 1AL (Lucas *et al.*, 2013) physical map. Information gathered from the utilization of this array is also used

for *in silico* anchoring through so-called GenomeZipper (Mayer *et al*., 2009; Mayer *et al*., 2011), which was originally developed for barley chromosome survey sequences to determine a virtual order of genes along the chromosome using inference of synteny information with sequenced related grass genomes (Mayer *et al*., 2009).
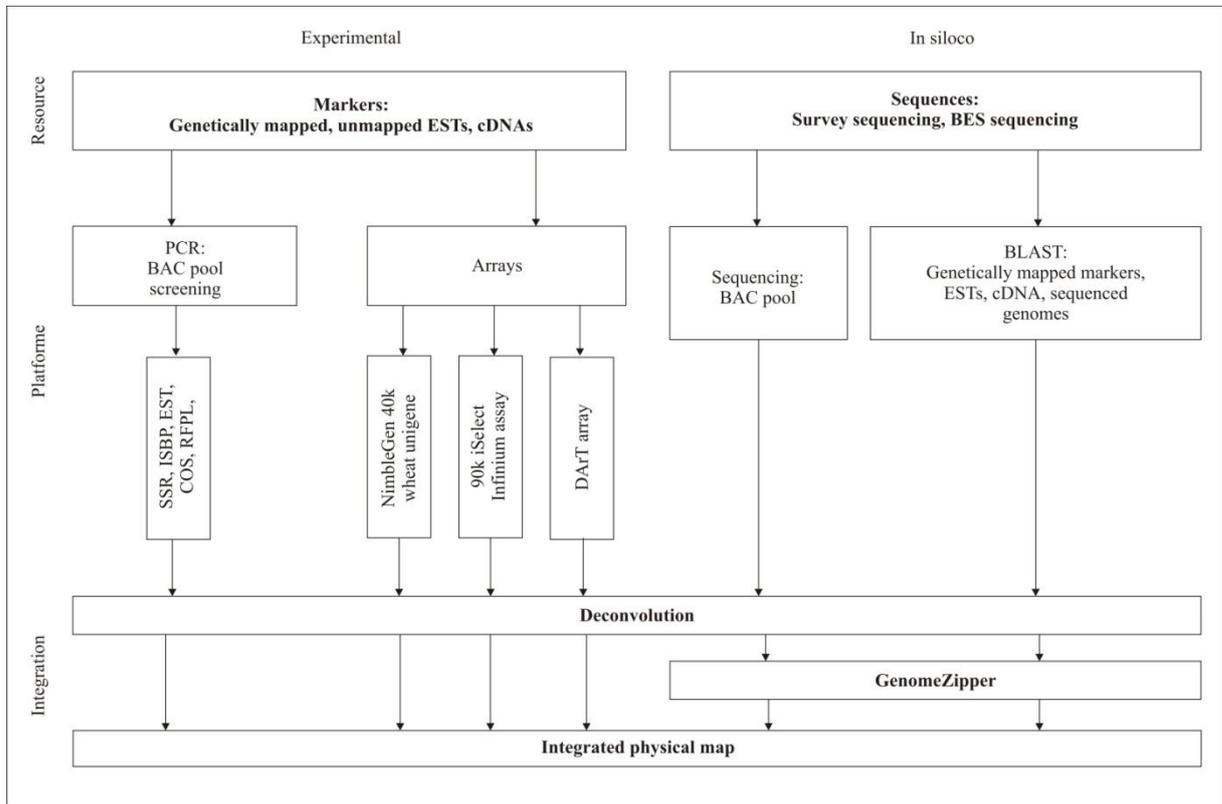


**Figure 11:** Approaches of wheat physical map anchoring

Genetically mapped markers and unmapped sequence information such as cDNA and ESTs can be used to screen multidimensional BAC pools via PCR-based methods or highly parallel array platforms. The resulted data are deconvoluted to obtain marker BAC relationships prior to integrating to the contig map. Alternatively, the BAC sequence information obtained through survey sequencing (NGS/Sanger sequencing) or BES could be searched against genetically mapped markers, ESTs, cDNA, etc. of the same organism or different species (adjusted according to Ariyadasa *et* Stein, 2012).

## 2.2.4  Anchoring of physical map to radiation hybrid map

All the previous approaches used information from a genetic map as a backbone, to some extent. Using such resources, anchoring in regions near centromeres or within large blocks of heterochromatin where little or no recombination occurs may not be possible (Akhunov *et al*., 2003; Kumar *et al*., 2012). An alternative strategy for getting enough mapping

resolution in peri-centromeric regions and other low recombining regions offers the radiation hybrid mapping (RH mapping) approach, which was originally developed in human genetics (Goss *et* Harris, 1975). In wheat the radiation hybrid mapping was successfully used for 3B (Paux *et al.*, 2008b) and 6B (Kobayashi *et al.*, 2015) physical map anchoring, so far.

Radiation hybrid mapping have several advantages over the genetic mapping. The biggest advantage is that polymorphism is not a requirement. Molecular markers are mapped based on their presence/absence (Kumar *et al.*, 2014).

Another advantage is the resolution. The resolution in RH mapping is determined by the number of breaks along the chromosome. By the enhancement of radiation dosage the number of breaks along chromosome can be increased and also resolutions of RH map increases. Thus, RH panel with higher resolution than mapping population (used for genetic mapping) are produced without the need to extend population size. However, the tolerance for radiation dosage in plants is limited, especially when viable plant is needed (Kumar *et al.*, 2014). This can be overcome by using endosperm radiation hybrid (ERH) panel. ERH panel allows applying higher radiation dosage and to create and detect more breaks along chromosome. DNA of ERH panel is extracted from endosperms and plant (if needed) can be produced by embryo rescue technique (Tiwari *et al.*, 2012).

The third advantage is uniformity of resolution across the chromosome. Radiation in-duced breakages are assumed to be random along the whole chromosome. Kumar *et al.*, 2012 observed that resolution of RH map in centromeric region was about 100-fold higher than the genetic map of the same segment. RH maps represent the physical genome better than do the genetic maps (Kumar *et al.*, 2014).

**Wheat RH maps**

Wheat RH panels can be divided into two groups, namely the whole genome (WG) panels and single chromosome panels. WG panels were developed only for D-genome by crossing irradiated pollen (endosperm RH panel) or seeds (seed RH panel) of hexaploid wheat (A*A*B*B*D*D*) with tetraploid wheat (AABB). The resulting RH lines were quasi-pentaploid (AA*BB*D*) with single copy of an irradiated D*-genome (Kumar et al., 2012; Tiwari *et al.*, 2016). So far, three additional WG panels were developed. First WG RH panel was only characterized by 10 SSR markers. Only chromosomes 5D and 6D were assayed by 20 and 15 additional markers, respectively (Riera-Lizarazu *et al.*, 2010). Second WG panel

was developed by Kumar et al., 2012 and characterized by 35 SSR and 60 repeat DNA junction markers (RJM). RH map was constructed only for 2D chromosome which was saturated with additional 8 SSR, 8 RJM and 14 EST markers (Kumar *et al*., 2012). Whole RH map for this panel was constructed using 609 markers (580 DArT and 29 SSR) and total length of the map was 14 706,7 cR (Kumar *et al*., 2015). Tiwari *et al*. (2012) developed WG endosperm radiation hybrid panel and constructed RH map of D-genome with 737 DArT markers and total length of 5 054 cR.

Single chromosome RH panels are used mainly for physical map anchoring. Wide range of aneuploid stocks like substitution lines (Joppa *et* Williams, 1977) and nullisomic-tetrasomic lines (Sears, 1954) allowed development of single chromosome RH panels. First single chromosome RH map was constructed for 1D chromosome of wheat (Hossain *et al*., 2004). Following RH maps of chromosome 3B, 6B were mainly used for physical map anchoring. Chromosome additional line of bread wheat cultivar Chinese Spring, disomic for barley chromosome 3H, was used for development of single chromosome RH panel for barley chromosome 3H (Mazaheri *et al*., 2015).

## 2.2.5 Role of physical map for positional cloning of important genes

Anchored physical map is indispensable assistant for sequencing of individual chromosome arms of wheat but also for map-based positional cloning of agronomically and economically important genes.

Positional cloning is defined as a method with which the gene's location can be pinpointed with sufficient precision to narrow down its location to a DNA segment that is small enough to be sequenced. For this method it is not necessary to know the gene product. A physical map serves as a tool to circumvent large distances that can comprise of repetitive sequences (especially complex and large genomes like barley or wheat; Krattinger *et al*., 2009, Weikuan *et* Goldowitz, 2011).

Positional cloning has several steps which are schematically illustrated in Figure 12. The whole process starts with mapping population and identification of chromosome arm and location of the gene of interest. When molecular markers closely linked (flanking markers) to the gene of interest are identified, a physical map can be screened with these markers. Two BAC contigs can be identified in this step usually, which is followed by a pro-

cedure known as Chromosome walking. BAC clones from the contig ends are sequenced and new markers are designed from these sequences, which are mapped to the mapping population and used for next round of screening the physical map. These steps are repeated until complete physical map of the gene locus is constructed. The goal is to identify the BAC clone containing the gene of interest. The target BAC clone can be sequenced, annotated and candidate genes can be identified (Krattinger *et al*., 2009, Weikuan *et* Goldowitz, 2011). Table 3 summarized examples of positional cloning with the use of physical map or BAC library.
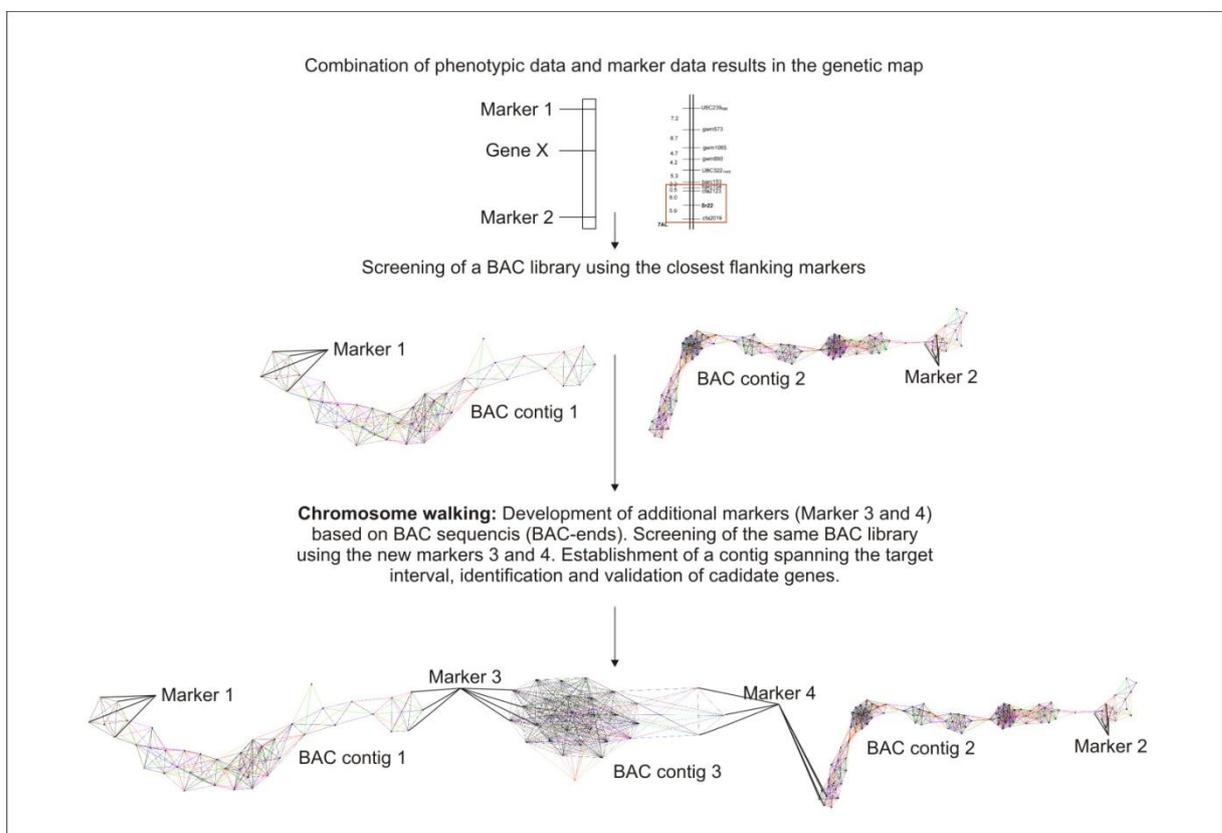


**Figure 12:** Strategy to isolate a gene of interest (gene X) via map-based cloning (adjusted according to Krattinger *et al*., 2009).

Nowadays, new method for positional cloning of major genes is being used. The method named MutChromSeq is not depending on recombination and fine-mapping. Sánchez-Martín *et al*. (2016) combined mutagenesis, genome complexity reduction and high-throughput sequencing and provides fast and inexpensive method to cloned genes in wheat or barley (Sánchez-Martín *et al*., 2016).

**Table 3:** List of genes that have been / are positionally cloned in wheat using BAC library

| Gene | Trait | BAC librery | Chromosome | Physical map | Reference |
|------|-------|-------------|------------|--------------|-----------|
| *Lr10* | Leaf rust resistance | *T. monococcum* DV92 | 1A$^m$S | - | Stein et al., 2000 |
| *Q* | The main domestication locus infuencing threshability and spike morphology | *T. monococcum* DV92 | 5AL | - | Faris et al., 2003 |
| *VRN1* | Flowering control | *T. monococcum* DV92 | 5A$^m$ | - | Yan et al., 2003 |
| *Pm3* | Powdery mildew resistance | *T. turgidum* ssp *durum*, *T. monococcum* DV92 | 1AS | Cenci et al., 2003 | Yahiaoui et al., 2004 |
| *VRN2* | Flowering control | *T. monococcum* DV92 | 5A$^m$ | - | Yan et al., 2004 |
| *Gpc-B1* | Zn and Fe grain uptake and senescence | *T. turgidum* ssp *durum* | 6BS | Cenci et al., 2003 | Distelfeld et al., 2006 |
| *Ph1* | Major chromosome pairing locus in polyploid wheat | ? | 5B | - | Griffiths et al., 2006 |
| *Tsn1* | Necrotrophic pathogenes resistance | *T. turgidum* ssp *durum* | 5BL | Cenci et al., 2003 | Lu et Faris, 2006 |
| *Lr1* | Leaf rust resistance | *Ae. tauschii* accession AL8/78, *T.aestivum* Xiaoyan 54 | 5DL | - | Qiu et al., 2007 |
| *Skr* | Crossability gene | *T. aestivum*, Tae-B-Chinese spring | 5BS | - | Alfares et al., 2009 |
| *Yr36* | Stripe rust resistance | *T. turgidum* ssp *durum* | 6BS | Cenci et al., 2003 | Fu et al., 2009 |
| *PSH* | Pre-harvest sprouting | *T. aestivum* TaaCsp3AShA | 3AS | not yet published | Liu et al., 2013 |
| *MlIW172* | Powdery mildew resistance | *T. aestivum* TaaCsp7ALhA | 7AL | not yet published | Quyang et al., 2014 |
| *Sr2* | Stem rust resistance | *T. aestivum* cv Hope Taa-Hop3BFh | 3B | - | Mago et al., 2014 |
| *Dn2401* | Russian wheat aphid resistance | *T. aestivum* TaaCsp7DShA | 7DS | not yet published | Staňková et al., 2015 |

## 2.3   References

Akhunov ED, Goodyear AW, Geng S, Qi L, Echalier B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao S, Anderson OD, Linkiewicz AM, Dubcovsky J, La Rota M, Sorrells ME, Zhang D, Nguyen HT, Kalavacharla V, Hossain K, Kianian, SF, Peng J, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Choi D, Close TJ, Dilbirligi M, Gill KS, Walker-Simmons MK, Steber C, McGuire PE, Qualset, CO, Dvorak J (2003): The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome research 13:753-763

Akpinar BA, Magni F, Yuce M, Lucas SJ, Šimková H, Šafář J, Vautrin S, Berges 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

Alfares W, Bouguennec A, Balfourier F, Gray G, Berges H, Vautrin S, Sourdille P, Bernard M, Feuillet C (2009): Fina mapping and marker development for the crossability gene SKr on chromosome 5BS of hexaploid wheat (Triticum aestivum L). Genetics 183(2):469-81

Appels R, Akhunov E, Alaux M, Caccamo M, Cattonaro F, Dolezel J, Edwards D, Luo MC, Matthews D, Guilhot N, Paul E, Wicker T, Eversole K and Feuillet C (2010) IWGSC: physical mapping standard protocols workshop report. In: W.G.R.C. Kansas State University (ed.) Plant and animal genome meeting vol. 56. San Diego, CA, USA, pp. 5–9

Ariyadasa R, Stein N (2012): Advances in BAC-Based Physical Mapping and Map Integration Strategies in Plants. Journal of Biomedicine and Biotechnology 2012:184854

Arratia R, Lander ES, Tavare S, Waterman MS (1991): Genomic mapping by anchoring random clones: a mathematical analysis. Genomics, 11(4):806-827.

Barabaschi D, Magni F, Volante A, Gadaleta A, Šimková H, Scalabrin S, Prazzoli ML, Bagnaresi P, Lacrima K, Michelotti V, Desiderio F, Orru L, Mazzamurro V, Fricano A, Mastrangelo A, Tononi P, Vitulo N, Jurman I, Frenekl Z, Cattonaro F, Morgante M, Blanco A, Doležel J, Delledonne M, Stanca AM, Cattivelli L, Vale G (2015): Physical

Mapping of Bread Wheat Chromosome 5A: An Integrated Approach. The Plant Genome 8(3)

Bennett MD, Smith JB (1976): Nuclear DNA amounts in angiosperms. Philos. T. R. Soc. B., 274:227-274

Bento M, Gustafson JP, Viegas W, Silva M (2011): Size matters in *Triticeae* polyploids: larger genomes have higher remodelling. Genome 54:175-183

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, Berges 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(11): e80272

Bruno WJ, Knill E, Balding DJ, Bruce DC, Doggett NA, Sawhill WW, Stallings RL, Whitaker CC, Torney DC (1995): Efficient pooling designs for library screening. Genomics 26(1):21-30

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(5):931-9

Collins F, Galas D (1993): A new five-year plan for the U.S. Human Genome Project. Science, 262(5130):43-6

Distelfeld A, Uauy C, Fahima T, Dubcovsky J (2006): Physical map of the wheat high-grain protein content gene GPC-B1 and development of a high-throughput molecular marker. New Phytol 169(4):753-63

Doležel J, Bartoš J, Voglmayr H, Greilhuber J (2003): Nuclear DNA content and genome size of trout and human. Cytometry, 51:127-128

Doležel J, Kubaláková M, Paux E, Bartoš J, Feuillet C (2007): Chromosome-based genomics in the cereals, Chromosome Res, 15(1):51-66

Dubcovsky J, Dvorak J (2007): Genome Plasticity a Key Factor in the Success of Polyploid Wheat under Domestication. Science 316:1862-1866

Dvorak J, Akhunov ED, Akhunov AR, Deal KR, Luo MC (2006): Molecular Characterization of a Diagnostic DNA Marker for Domesticated Tetraploid Wheat Provides Evidence for Gene Flow from Wild Tetraploid Wheat to Hexaploid Wheat. Mol Biol Evol, 23(7):1386-1396

Engler F, Soderlund C (2002): Software for Physical Maps. In: Ian Dunham (ed) *Genomic Mapping and Sequencing*, Horizon Press, Genome Technology series. Norfolk, UK, pp. 201-236.

Engler FW, Hatfield J, Nelson W, Soderlund CA (2003): Locating Sequence on FPC Maps and Selecting a Minimal Tilling Path. Genome research 13:152-2163

Engler F, Soderlung C (2006): FPC V8: A tutorial;
http://www.agcol.arizona.edu/software/fpc/

Faris JD, Fellers JP, Brooks SA, Gill BS (2003): A bacterial artificial chromosome contig spanning the major domestication locus Q in wheat and identification of a candidate gene. Genetics 164(1):311-21

Feldman M, Levy AA (2009): Genome evolution in allopolyploid wheat – a revolutionary reprogramming followed by gradual changes. Journal of Genetics and 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 U S A., 100(25):15253-8.

Feuillet C, Eversole K (2008). Physical mapping of the wheat genome: A coordinated effort to lay the foundation for genome sequencing and develop tools for breeders. Israel Journal of Plant Sciences 55: 307-313.

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

Flavell RB, Rimpau J, Smith DB (1977): Repeated sequence DNA relationship in four cereals genomes. Chromosoma, 63:205-222

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 Bioinformatics 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(5919):1357-60

Gill BS (1987): Chromosome banding methods, standard, chromosome nomenclature and applications in cytogenetic analysis. In: E. G. Heyne (Ed.): Wheat and wheat improvements. American Society of Agronomy, Madison, pp. 243-254

Goncharov NP (2011): Genus *Triticum* L. taxonomy: the present and the future. Plant Syst Evol 295:1-11

Goss SJ, Harris H (1975): New method for mapping genes in human chromosome. Nature 255:680-684

Gregory S, Soderlund C, Coulson A (1996): Contig assembly by fingerprinting. In: Dear PH (ed.), Genome Mapping: A Practical Approach. pp 227-254 Oxford University Press

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 polyploidy wheat. Nature 439(7077):749-52

Gustafson P, Raskina O, Ma XF, Nevo E (2009): Wheat Evolution, Domestication, and Improvement. In: Carver BF (eds.): Wheat Science and Trade, pp. 3-30, Wiley-Blackwell, Oxford, UK.

Haider N (2013): The Origin of the B-Genome of Bread Wheat (*Triticum aestivum* L.). Russian Journal of Genetics 49(3):263-274

Hossain KG, Riera-Lizarazu O, Kalavacharla V, Vales MI, Maan SS, Kianian SF (2004): Radiation hybrid mapping of the species cytoplasm-specific (*scsae*) gene in wheat. Genetics 168(1):415-23

Chen M, Presting G, Barbazuk WB, Goicoechea JL, Blackmon B, Fang G, Kim H, Frisch D, Yu Y, Sun S, Higingbottom S, Phimphilai J, Phimphilai D, Thurmond S, Gaudette B, Li P, Liu J, Hatfield J, Main D, Farrar K, Henderson C, Barnett L, Costa R, Williams B, Walser S, Atkins M, Hall C, Budiman MA, Tomkins JP, Luo M, Bancroft I, Salse J, Regad F, Mohapatra T, Singh NK, Tyagi AK, Soderlund C, Dean RA, Wing RA (2002): An integrated physical and genetic map of the rice genome. Plant Cell 14(3):537-545

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlu S, LePaslier MC, Magdelenat G, Gonthier C, Couloux A, Budak H, Breen J, Pumphery M, Liu S, Kong X, Jia J, Gut M, Brunel D, Anderson JA, Gill BS, 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, Caccamo M, Wright J, Alaux M, Šimková H, Šafář J, Leroy P, Doležel J, Rogers J, Eversole K, Feuillet C (2014a): 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. Pp. 405-438. Springer

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, Berges H, Quesneville H, Wincker P, Feuillet C (2014b): Structural and functional partitioning of bread wheat chromosome 3B. Science 345(6194): 1249721.

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

Jain AK, Dubes RC (1988): Algorithms for clustering data N.J.: Prentice Hall: Englewood Cliffs

Jia JZ, Zhao SC, Kong XY, Li YR, Zhao GY, He WM, Appels R, Pfeifer M, Tao Y, Zhang XY, Jing RL, Zhang C, Ma YZ, Gao LF, Gao C, Spannagl M, Mayer KFX, Li D, Pan SK, Zheng FY, Hu Q, Xia XC, Li JW, Liang QS, Chen J, Wicker T, Gou CY, Kuang HH, He GY, Luo YD, Keller B, Xia QJ, Lu P, Wang JY, Zou HF, Zhang RZ, Xu JY, Gao JL, Middleton C, Quan ZW, Liu GM, Wang J, Yang HM, Liu X, He ZH, Mao L, Wang J (2013): *Aegilops tauschii* draft genome sequence reveals a gene repertoire for wheat adaptation. Nature 496, 91–95

Joppa LR, Williams ND (1977): D-genome substitution monosomics of durum wheat. Crop Sci 17:772-776

Keller B, Feuillet C, Yahiaoui N (2005): Map-based isolation of disease resistance genes from bread wheat: cloning in a supersize genome. Genet Res., 85(2):93-100

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 JCM, 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

Krattinger S, Wicker T, Keller B (2009): Map-Based Cloning of Genes in Triticeae (Wheat and Barley). In: Feuillet C, Muehlbauer GJ (eds.): Genetics and Genommics of the Triticeae, Plant Genetics and Genomics: Crops and Models 7. pp 337-357, Springer Science US

Kumar A, Simons K, Iqbal MJ, Michalak de Jiménez M, Bassi FM, Ghavami F, Al-Azzam O, Drader T, Wang Y, Luo MCh, Gu YQ, Denton A, Lazo GR, Xu SS, Dvorak J, Kianian

PMA, Kianian SF (2012): Physical mapping resources for large palnt genomes: radiation hybrids for whet D-genome progenitor Aegilops tauschii. BMC Genomics 13:597

Kumar A, Bassi FM, Michalak de Jimenez MK, Ghavami F, Mazaheri M, Simons K, Iqbal MJ, Mergoum M, Kianian SF, Kianian PMA (2014). Radiation hybrids: a valuable tool for genetic, genomic and functional analysis of plant genomes, Genomics of Plant Genetic Resources, Vol.1, eds R.Tuberosa, A .Graner, and E. Frison (Dordrecht: Springer), 285–318.

Kumar A, Seetan R, Mergoum M, Tiwari VK, Iqbal MJ, Wang Y, Al-Azzam O, Šimková H, Luo MC, Dvorak J, Gu YQ, Denton A, Kilian A, Lazo GR, Kianian SF (2015): Radiation hybrid maps of the D-genome of Aegilops tauschii and their application in sequence assembly of large and complex plant genomes. BMC Genomics, 16:800

Ling HQ, Zhao SC, Liu DC, Wang JY, Sun H, Zhang C, Fan HJ, Li D, Dong LL, Tao Y, Gao C, Wu HL, Li YW, Cui Y, Guo XS, Zheng SS, Wang B, Yu K, Liang QS, Yang WL, Lou XY, Chen J, Feng MJ, Jian JB, Zhang XF, Luo GB, Jiang Y, Liu JJ, Wang ZB, Sha YH, Zhang BR, Wu HJ, Tang DZ, Shen QH, Xue PY, Zou SH, Wang XJ, Liu X, Wang FM, Yang YP, An XL, Dong ZY, Zhang KP, Zhang XQ, Luo MC, Dvorak J, Tong YP, Wang J, Yang HM, Li ZS, Wang DW, Zhang AM, Wang J (2013): Draft genome of the wheat A-genome progenitor *Triticum urartu*. Nature 496, 87-90

Liu S, Sehgal SK, Li J, Lin M, Trick HN, Yu J, Gill BS, Bai G (2013): Cloning and characterization of a critical regulator for preharvest sprouting in wheat. Genetics 195(1):263-73

Lu H, Faris JD (2006): Macro- and microcolinearity between the genomic region of wheat chromosome 5B containing the Tsn1 gene and rice genome. Funct Integr Genomics 6(2):90-103

Lucas SJ, Akpınar BA, Kantar M, Weinstein Z, Aydınoglu F, Šafář J, Šimková H, Frenkel Z, Korol A, Magni F, Cattonaro F, Vautrin S, Bellec A, Berges 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(4): e59542

Luo MC, Thoma C, You F M, 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 labelling kit and sizing of restriction fragments by capillary electrophoresis. Genomics 82:378 – 389

Mago R, Tabe L, Vautrin S, Šimková H, Kubaláková M, Upadhyaya N, Berges H, Kong X, Breen J, Doležel J, Appels R, Ellis JG, Spielmeyer W (2014): Major hyplotype divergence including multiple germin-like protein genes, at the wheat Sr2 adult plant stem rust resistance locus. BMC Plant Biol 14:379

Marra M, Kucaba T, Sekhon M, Hillier L, Martienssen R, Chinwalla A, Crockett J, Fedele J, Grover H, Gund Ch, McCombie WR, McDonald K, McPherson J, Mudd N, Parnell L, Schein J, Seim R, Shelby P, Waterston R, Wilson R (1999): A map for sequence analysis of the *Arabidopsis thaliana* genome. Nature genetics 22, 265 - 270

Massa AN, Wanjugi H, Deal KR, O'Brien K, You FM, Maiti R, Chan AP, Gu YQ, Luo MC, Anderson OD, Rabinowicz PD, Dvorak J, Devos KM (2011): Gene Space Dynamics During the Evolution of *Aegilops tauschii*, *Brachypodium distachyon*, *Oryza sativa*, and *Sorghum bicolor* Genomes. Mol Biol Evol 28:2537-2547

Matsuoka Y (2011): Evolution of Polyploid *Triticum* Wheats under Cultivation: The Role of Domestication, Natural Hybridization and Allopolyploid Speciation in their Diversification. Plant Cell Physiol. 52(5):750-764

Mayer KF, Taudien S, Martis M, Simková H, Suchánková P, Gundleach H, Wicker T, Petzold A, Felder M, Steuenagel B, Sholz U, Graner A, Platzer M, Dolezel 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, Simková 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, Dolezel J, Waugh R, Stein N (2011): Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1246-1263

Mazaheri M, Kianian PMA, Kumar A, Mergoum M, Seetan R, Soltani A, Lund LI, Pirseyedi SM, Denton, AM, Kianian SF (2015): Radiation Hybrid Map of Barley Chromosome 3H. The Plant Genome 8(2)

Mester D, Ronin Y, Minkov D, Nevo E, Korol A (2003): Constructing large-scale genetic maps using an evolutionary strategy algorithm. Genetics, 165(4):2269-2282.

Meyer BC, Scalabrin S, Morgane M (2004): Mappimg and sequencing complex genomes: Let's get physical! Nature reviews 5:578-588

Michalak de Jimmenez MK, Bassi FM, Ghavami F, Simons K, Dizon R, Seetan RI, Alnemer LM, Denton AM, Doğramaci M, Šimková H, Doležel J, Seth K, Luo M, 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

Moore G (1995): Cereal genome evolution: pastoral pursuits with 'Lego' genomes. Curr Opin Genet Dev, 5(6):717-24.

Nelson W, Soderlund C (2009): Integrating semence with FPC fingerpint maps. Nucleic Acids Research 1-11

Paux E, Legeai F, Guilhot N, Adam-Blomdon A, Alaux M, Salse J, Sourdille P, Leroy P, Feuillet C (2008a): Physical mapping in large genomes: accelerating anchoring of BAC contigs to genetic maps through in silico analysis. Funct Integr Genomics 8:29-32

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 (2008b): A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322: 101-104

Peng JH, Sun D, Nevo E (2011): Domestication evolution, genetics and genomics in wheat. Mol Breeding 28:281-301

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 KFX, 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 KFX, Stein N, Schnurbusch T (2014): Whole-genome profiling and shotgun sequencing delivers an anchored, gene-decorated, physical map assembly of bread wheat chromosome 6A. The Plant Journal 79:334-347

Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, Akhunov ED, Dvořák J, Linkiewicz AM, Ratnasiri A, Dubcovsky J, Bermudez-Kandianis CE, Greene RA, Kantety R, La Rota CM, Munkvold JD, Sorrells SF, Sorrells ME, Dilbirligi M, Sidhu D, Erayman M, Randhawa HS, Sandhu D, Bondareva SN, Gill KS, Mahmoud AA, Ma XF, Miftahudin, Gustafson JP, Conley EJ, Nduati V, Gonzalez-Hernandez JL, Anderson JA, Peng JH, Lapitan NLV, Hossain KG, Kalavacharla V, Kianian SF, Pathan MS, Zhang DS, Nguyen HT, Choi DW, Fenton RD, Close TJ, McGuire PE, Qualset CO, Gill BS (2004): A Chromosome Bin Map of 16,000 Expressed Sequence Tag Loci and Distribution of Genes Among the Three Genomes of Polyploid Wheat. Genetics 168:701-712

Qiu JW, Schürch AC, Yahiaoui N, Dong LL, Fan HJ, Zhang ZJ, Keller B, Ling HQ (2007): Physical mapping and identification of a candidate for the lef rust resistance gene Lr1 of wheat. Theor Appl Genet 115(2):159-68

Quyang S, Zhang D, Han J, Zhao X, Cui Y, Song W, Huo N, Liang Y, Xie J, WangZ, Wu Q, Chen YX, Lu P, Zhang DY, Wang L, Sun H, Yang T, Keeble-Gagnere G, Appels R, Doležel J, Ling HQ, Luo M, Gu Y, Sun Q, Liu Z (2014): Fine physical and genetic mapping of powdery mildew resistance gene MIIW172 originating from wild emmer (Triticum dicoccoides). PLoS One 9(6):e100160

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 Biology 14: R138

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

Rustenholz C, Choulet F, Laugier C, Safar J, Simkova H, Dolezel J, Magni F, Scalabrin S, Cattonaro F, Vautrin S, Bellec A, Berges H, Feuillet C, Paux E (2011): A 3000-loci transcription map of chromosome 3B unravels the structural and functional features of gene islands in hexaploid wheat. Plant Physiol 157:1596-1608

Sánchez-Martín J, Steuernagel B, Ghosh S, Herren G, Hurni S, Adamski N, Vrána J, Kubaláková M, Krattinger SG, Wicker T, Doležel J, Keller B, Wulff BB (2016): Rapid gene isolation in barley and wheat by mutant chromosome sequencing. Genome Biol. 17(1):221

Sandhu D, Gill KS (2002): Gene-Containing Regions of Wheat and the Other Grass Genomes. Plant Physiol 128:803-811

Sears ER (1954): The aneuploid of common wheat. Mo Agr Exp Sta Res Bull 572: 1-58

Shewry PR (2009): Wheat. Journal of Experimental Botany 60 (6):1537-1553

Smith DB, Flavell RB (1975): Characterization of Wheat Genome by Renaturation Kinetics. Chromosoma 50:223-242

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 Research 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 facili-

tates fine mapping of a Russian wheat aphid resistance gene. Theor Appl Genet 128(7):1373-83

Stein N, Feuillet C, Wicker T, Schlagenhauf E, Keller B (2000): Subgenome chromosome walking in wheat: a 450-kb physical contig in Triticum monococcum L. spans the Lr10 resistance locus in hexaploid wheat (Triticum aestivum L.). Proc Natl Acad Sci USA 97(24):13436-41

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

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

Šimková H, Šafář J, Kubaláková M, Suchánková P, Číhalíková J, Robert-Quatre H, Azhaguvel P, Weng Y, Peng J, Lapitan NLV, Ma Y, You FM, Luo MCh, 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.

Tiwari VK, Heesacker A, Riera-Lizarazu O, Gunn H, Wang S, Wang Y, Gu YQ, Paux E, Koo DH, Kumar A, Luo MC, Lazo G, Zemetra R, Akhunov E, Friebe B, Poland J, Gill BS, Kianian S, Leonard JM (2016): A whole-genome, radiation hybrid mapping resource of hexaploid wheat. Plant J, 86(2):195-207

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 a measure of g ray-induced chromosome breaks. PLoS ONE 7:e48815

Wang Y, Prade RA, Griffith J, Timberlake WE, Arnold J (1994): A fast random cost algorithm for physical mapping. Proc Natl Acad Sci, 91(23):11094-11098.

Weikuant G, Goldowitz D (2011): Gene Discovery: From Positional Cloning to Genomic Cloning. In: Weikuant G, Wang Y (ed.): Gene Discovery for Disease Models. pp 1-9 John Wiley & Sons

Wendel JF (2000): Genome evolution in polyploids. Plant Molecular Biology 42:225-249

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(4):528-38

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(10):6263-8

Yan L, Loukoianov A, Blechl A, Tranquilli G, Ramakrishna W, SanMiguel P, Benetzen JL, Echenique V, Dubcovsky J (2004): The wheat VRN2 gene is a flowering repressor down-regulated by vernalization. Science 303(5664):1640-4

# 3  Aims of the Thesis

**I.  Construction of physical map of the wheat chromosome 4A**

The main aims of the thesis were construction of physical map of wheat chromosome 4A, the minimal tilling path (MTP) selection, preparation and sequencing of MTP 3D pools, and anchoring of physical map contigs to chromosome through available mapping and genomic resources.

**II.  Construction of high density RH map of the chromosome 4A**

The second aim was construction of RH map which resolution is independent from natural recombination and allows marker ordering and the physical map anchoring in low recombining regions.

**III.  Application of the physical map for positional cloning of important wheat genes**

The additional aim of the study was application of the created genomic resources for high density mapping and positional cloning of agronomically important genes.

# 4 Results

## *4.1 Original Papers*

4.1.1 A High Resolution Radiation Hybrid Map of Wheat Chromosome 4A (Appendix I)

4.1.2 The *in silico* identification and characterization of a bread wheat/*Triricum militinae* introgression line (Appendix II)

4.1.3 The wheat *Phs-A1* pre-harvest sprouting resistance locus delays dormancy loss during seed after-ripening and maps 0.3 cM distal to the *PM19* genes in UK germplasm (Appendix III)

4.1.4 Molecular mapping of stripe rust resistance gene *Yr51* in chromosome 4AL of wheat (Appendix IV)

### 4.1.1  A High Resolution Radiation Hybrid Map of Wheat Chromosome 4A

Balcárková B, Frenkel Z, Škopová M, Abrouk M, Kumar A, Chao S, Kianian SF, Akhunov E, Korol AB, Doležel J, Valárik M

**ABSTRACT:** Bread wheat has a large and complex allohexaploid genome with low recombination level at chromosome centromeric and peri-centromeric regions. This significantly hampers ordering of markers, contigs of physical maps and sequence scaffolds and impedes obtaining of high-quality reference genome sequence. Here we report on the construction of high-density and high-resolution radiation hybrid (RH) map of chromosome 4A supported by high-density chromosome deletion map. A total of 119 endosperm-based RH lines of two RH panels and 15 chromosome deletion bin lines were genotyped with 90K iSelect single nucleotide polymorphism (SNP) array. A total of 2316 and 2695 markers were successfully mapped to the 4A RH and deletion maps, respectively. The chromosome deletion map was ordered in 19 bins and allowed precise identification of centromeric region and verification of the RH panel reliability. The 4A-specific RH map comprises 1080 mapping bins and spans 6550.9 cR with a resolution of 0.13 Mb/cR. Significantly higher mapping resolution in the centromeric region was observed as compared to recombination maps. Relatively even distribution of deletion frequency along the chromosome in the RH panel was observed and putative functional centromere was delimited within a region characterized by two SNP markers.

### 4.1.2 The *in silico* identification and characterization of a bread wheat/*Triricum militinae* introgression line

Abrouk M, <u>Balcárková B</u>, Šimková H, Komínkova E, Martis MM, Jakobson I, Timofejeva L, Rey E, Vrána J, Kilian A, Järve K, Doležel J and Valárik M

IF: 6.09

**ABSTRACT:** The capacity of the bread wheat (Triticum aestivum) genome to tolerate introgression from related genomes can be exploited for wheat improvement. A resistance to powdery mildew expressed by a derivative of the cross-bread wheat cv. Tähti × *T. militinae* (*Tm*) is known to be due to the incorporation of a Tm segment into the long arm of chromosome 4A. Here, a newly developed in silico method termed rearrangement identification and characterization (RICh) has been applied to characterize the introgression. A virtual gene order, assembled using the GenomeZipper approach, was obtained for the native copy of chromosome 4A; it incorporated 570 4A DArTseq markers to produce a zipper comprising 2132 loci. A comparison between the native and introgressed forms of the 4AL chromosome arm showed that the introgressed region is located at the distal part of the arm. The Tm segment, derived from chromosome 7G, harbours 131 homoeologs of the 357 genes present on the corresponding region of Chinese Spring 4AL. The estimated number of *Tm* genes transferred along with the disease resistance gene was 169. Characterizing the introgression's position, gene content and internal gene order should not only facilitate gene isolation, but may also be informative with respect to chromatin structure and behaviour studies.

### 4.1.3 The wheat *Phs-A1* pre-harvest sprouting resistance locus delays dormancy loss during seed after-ripening and maps 0.3 cM distal to the *PM19* genes in UK germplasm

Shorinola O, Bird N, Simmonds J, Berry S, Henriksson T, Jack P, Werner P, Gergets T, Scholefield D, Balcárková B, Valárik M, Holdsworth M, Flintham J, Uauy C

IF: 5.677

**ABSTRACT:** The precocious germination of cereal grains before harvest, also known as pre-harvest sprouting, is an important source of yield and quality loss in cereal production. Pre-harvest sprouting is a complex grain defect and is becoming an increasing challenge due to changing climate patterns. Resistance to sprouting is multi-genic, although a significant proportion of the sprouting variation in modern wheat cultivars is controlled by a few major quantitative trait loci, including *Phs-A1* in chromosome arm 4AL. Despite its importance, little is known about the physiological basis and the gene(s) underlying this important locus. In this study, we characterized *Phs-A1* and show that it confers resistance to sprouting damage by affecting the rate of dormancy loss during dry seed after-ripening. We show *Phs-A1* to be effective even when seeds develop at low temperature (13 °C). Comparative analysis of syntenic *Phs-A1* intervals in wheat and *Brachypodium* uncovered ten orthologous genes, including the *Plasma Membrane 19* genes (*PM19-A1* and *PM19-A2*) previously proposed as the main candidates for this locus. However, high-resolution fine-mapping in two bi-parental UK mapping populations delimited *Phs-A1* to an interval 0.3 cM distal to the *PM19* genes. This study suggests the possibility that more than one causal gene underlies this major pre-harvest sprouting locus. The information and resources reported in this study will help test this hypothesis across a wider set of germplasm and will be of importance for breeding more sprouting resilient wheat varieties.

### 4.1.4 Molecular mapping of stripe rust resistance gene *Yr51* in chromosome 4AL of wheat

Randhawa M, Bansal U, Valárik M, Klocová B, Doležel J, Bariana H

**ABSTRACT:** A wheat landrace, AUS27858, from the Watkins collection showed high levels of resistance against Australian pathotypes of *Puccinia striiformis* f. sp. *tritici*. It was reported to carry two genes for stripe rust resistance, tentatively named *YrAW1* and *YrAW2*. One hundred seeds of an F3 line (HSB#5515; *YrAW1yrAW1*) that showed monogenic segregation for stripe rust response were sown and harvested individually to generate monogenically segregating population (MSP) #5515. Stripe rust response variation in MSP#5515 conformed to segregation at a single locus. Bulked segregant analysis using high-throughput DArT markers placed *YrAW1* in chromosome 4AL. MSP#5515 was advanced to F6 and phenotyped for detailed mapping. Novel wheat genomic resources including chromosome specific sequence and genome zipper were employed to develop markers specific for the long arm of chromosome 4A. These markers were used for further saturation of the *YrAW1* carrying region. *YrAW1* was delimited by 3.7 cM between markers *owm45F3R3* and *sun104*. Since there was no other stripe rust resistance gene located in chromosome 4AL, *YrAW1* was formally named *Yr51*. Reference stock for Yr51 was lodged at the Australian Winter Cereal Collection, Tamworth, Australia and it was accessioned as AUS91456. Marker sun104 was genotyped on a set of Australian and Indian wheat cultivars and was shown to lack the resistance-linked sun104-225 bp allele. Marker sun104 is currently being used for marker-assisted back-crossing of Yr51 in Australian and Indian wheat backgrounds.

## 4.2 Published abstracts

4.2.1 Sequence ready physical map of bread wheat chromosome 4A (Appendix V)

4.2.2 Construction and characterization of wheat 4A chromosome specific physical map as a base step for the chromosome sequencing (Appendix VI)

### 4.2.1  Sequence ready physical map of bread wheat chromosome 4A

Klocová B, Abrouk M, Frenkel Z, Kumar A, Kianian SF, Šimková H, Šafář J, Hu Y, Luo M, Carling J, Kilian A, Korol AB, Wang S, Akhunov E, Doležel J, Valárik M

**ABSTRACT:** Despite progress in next generation sequencing technologies sequencing of large and complex plant genomes like that of bread wheat is still a challenge. Size, complexity, and repetitive element content of the allohexaploid bread wheat (*Triticum aestivum* L., 17 Gb) genome are the main obstructions from acquiring a reference genome sequence. Dividing the wheat genomes to chromosomes or chromosomal arms is powerful means to overcome the difficulties. As part of the coordinated effort by the International Wheat Genome Sequencing Consortium, we have constructed physical map of chromosome 4A using BAC libraries from flow-sorted short (4AS) and long (4AL) chromosome arms. The fingerprinted BAC clones were assembled to physical maps using LTC (Linear Topology Contig) software and additional merging and elongating was facilitated by super-scaffolding tool of the LTC software. The 4AS library was assembled into 250 super-contigs and Minimum Tiling Path (MTP) consist of 4 422 clones. The 4AL physical map was assembled into 924 super-contigs and MTP contains of 8 369 clones. The physical maps represent 86% and 89% of 4AS and 4AL, respectively. 67 and 74 super-contigs with more than 100 clones were assembled for 4AS and 4AL respectively, with the longest super- contig comprising 998 clones. Three dimensional pools of the MTP were sequenced and used for in silico anchoring of the contigs. A total of 1780 DArT markers were used to construct GenomeZipper from the 4AS and 4AL survey sequences. Additionally, 54125 and 62656 low-copy markers were identified from the survey sequences of the 4AS and 4AL chromosome arms, respectively. All these resources facilitated anchoring 100% of 4AS contigs and 99% of 4AL contigs. In order to facilitate ordering the physical map contigs in pericentromeric region and regions with low recombination rate, a Radiation Hybrid panel was constructed for chromosome 4A and genotyped with 90k SNP Infinium chip.This work has been supported by the grant LO1204 from the National Program

### 4.2.2 Construction and characterization of wheat 4A chromosome specific physical map as a base step for the chromosome sequencing

Klocová B, Kladivová M, Frenkel Z, Kumar A, Kianian S F, Šimková H, Šafář J, Hu Y, Zhang Y, You F M, Luo M, Korol A, Doležel J, Valárik M

**ABSTRACT:** *Triticum astivum* L. (bread wheat) has large and complex genome. The genome size is approximately 17 Gb with 80% of repetition and is composed of three homeologous genomes A, B and D. All of this makes sequencing, mapping, and marker assisted breeding a difficult task. To facilitate sequencing of wheat 4A chromosome we are developing sequence-ready chromosome arm specific physical maps. Physical maps for both arms were assembled from fingerprinted arm-specific BAC libraries. The physical maps were assembled using both, LTC (Linear Topology Contig) and FPC (FingerPrint Contigs) softwares. The 4AS library was assembled into 415 contigs and MTP consist of 4 422 clones. After superclustering, the advanced tools of LTC, there were 67 super-contigs with more than 100 clones in one contig. The longest supercontig comprised 582 clones. Also, 4AL physical map was assembled into 1127 contigs and MTP contains of 8 369 clones. After additional superclustering there were 74 supercontigs with more than 100 clones in one contig. Moreover, the supercontigs of both physical maps contains more than half of informative clones. 3D pools of MTP clones were prepared to simplify marker anchoring to the maps. To the physical map of 4AL were anchored 20 markers, four of this markers were found in one supercontig. After sequencing of the MTP BAC clones high-capacity physical map anchoring will be performed *in silico*. Endosperm radiation hybrid panel comprising 1 100 lines will be used for contig ordering and orientation. This work has been supported by the Czech Science Foundation (P501/10/1740), MSMT CR and EU (Operational Programme Research and Development for Innovations No. ED0007/01/01) and Internal Grant Agency PrF-2012-003.

# 5    Conclusions

In this thesis I focused on the study of wheat chromosome 4A. The first goal of this work was the construction of sequence ready physical map of long arm of chromosome 4A and anchoring of physical map to genetic recourses. The second aim was to constructed radiation hybrid map specific for chromosome 4A as genome recourse for physical map anchoring. The last part of the thesis was focused on positional cloning with the help of physical map. The 4A physical map was used for saturation in regions of genes responsible for resistance to powdery mildew (*QPm-tut-4A*), rust resistance (*Yr51*), pre-harvest sprouting gene (*Phs-A1*) and yield related locus (*Qyi-4A-bga).*

## *5.1    Physical map construction and anchoring*

Sequence-ready physical map of chromosome 4A is a backbone for positional cloning, markers for MAS development and especially for acquisition of reference sequence. Fingerprinted BAC library of 4A chromosome were assembled using LTC assembling software. MTP BACs were selected and pooled to 3D pools. Physical map had been further improved by manual contig merging. 3D MTP BAC pools were Illumina sequenced to facilitate physical map anchoring.

Physical map was anchored in three levels. CSS were used as first level. All scaffolds and contigs of 4AS physical map were anchored to CSS; on the contrary 98.6 % of 4AL physical map was anchored to CSS. 4A specific RH map was used as second level of physical map anchoring. Second level of anchoring enabled to anchored 79.6 % and 54.4 % of 4AS and 4AL physical map, respectively. As third level of physical map anchoring was used 4A-specific GZ. GZ allowed anchored 84.8 % and 67.5 % of 4AS and 4AL physical map, respectively. Three levels of different genomic resources allowed anchored, ordered and oriented of 100 % and 98.6 % of 4AS and 4AL physical map, respectively.

Web application was developed to open the anchored physical map to public (https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_4AL_v1).

## 5.2    Radiation hybrid map

4A-specific ERH panel was developed to construct recombination independent 4A-specific RH map. 119 lines were genotyped by 90K wheat SNP chip. Final RH map comprise 2 317 markers and spans 6 550.9 cR. RH map allowed anchoring of 79.6 % and 54.4 % of 4AS and 4AL physical maps, respectively.

## 5.3    Positional cloning

The anchored 4A physical map facilitated significantly narrowing down all regions for which was used. The sequences of BAC clones of the physical map allowed development of locus specific markers and anchoring to CSSs scaffolds.

The physical map was used for markers development, region saturation with markers and narrowing down the *QPm-tut-4A* locus to 0.02 cM. Region of *Phs-A1* gene was narrowed down to 0.3 cM. Physical map was used for identification and sequencing of candidate BAC clones. Sequences were annotated and causal gene identified. Region of gene resistance *Yr51* was delimited by 3.7 cM between markers *owm45F3R3* and *sun104* employing the 4A physical map and the 4A GZ. The marker sun104 was developed from the resources and is successfully used for marker-assisted selection. The *Qyi-4A-bga* gene locus was reduced to 0.5 cM using the 4A chromosome specific maps and the flanking markers are successfully used in pre-breeding and breeding process.

# 6    List of abbreviations

BAC          Bacterial artificial clone

BES          BAC-end sequence

bp           base pair

CB           consensus band

CS           Chinese Spring

cDNA         complementary DNA

cM           centy Morgan

cR           centy Ray

CSS          Chromosome Survey Sequence

DArT         Diversity Array Technology

DNA          deoxyribonucleic acid

ERH          endosperm radiation hybrid

EST          expressed sequence tag

FPC          FingerPrinted Contigs

Gb           giga base pair

GZ           GenomeZipper

HICF         High informative contend fingerprint

HMW          high molecular weight

IWGSC        International Wheat Genome Sequencing Consortium

kbp          kilo base pairs

LTC          Linear Topological Contig

LTR          long terminal repeat

MAS          Marker assisted selection

Mbp          mega base pair

mRNA         messenger ribonucleic acid

| | |
|---|---|
| MTP | Minimal Tilling Path |
| NGS | next generation sequencing |
| PCR | polymerase chain reaction |
| POPSEQ | population sequencing |
| RH | Radiation Hybrid |
| RJM | repeat junction marker |
| SNP | single nucleotide polymorphism |
| ssp. | subspecies |
| SSR | Simple Sequence Repeat |
| TE | transposable element |
| TSP | Travelling Salesman Problem |
| V | version |
| WG | Whole Genome |
| WGP | Whole Genome Profiling |
| Q | questionable |
| 3D | three dimensional |

# 7    List of appendices

## *Original papers*

Appendix I:    A High Resolution Radiation Hybrid Map of Wheat Chromosome 4A


Appendix II:    The *in silico* identification and characterization of a bread wheat/*Triricum militinae* introgression line


Appendix III:    The wheat *Phs-A1* pre-harvest sprouting resistance locus delays dormancy loss during seed after-ripening and maps 0.3 cM distal to the *PM19* genes in UK germplasm


Appendix IV:    Molecular mapping of stripe rust resistance gene *Yr51* in chromosome 4AL of wheat




## *Published abstracts – poster presentation*

Appendix V:    Sequence ready physical map of bread wheat chromosome 4A


Appendix VI:    Construction and characterization of wheat 4A chromosome specific physical map as a base step for the chromosome sequencing

# APPENDIX I

**A High Resolution Radiation Hybrid Map of Wheat Chromosome 4A**

Balcárková B, Frenkel Z, Škopová M, Abrouk M, Kumar A, Chao S, Kianian SF, Akhunov E, Korol AB, Doležel J, Valárik M

IF: 4.495

# A High Resolution Radiation Hybrid Map of Wheat Chromosome 4A

Barbora Balcárková[1], Zeev Frenkel[2], Monika Škopová[1†], Michael Abrouk[1], Ajay Kumar[3], Shiaoman Chao[4], Shahryar F. Kianian[5], Eduard Akhunov[6], Abraham B. Korol[2], Jaroslav Doležel[1] and Miroslav Valárik[1]*

[1] Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czechia, [2] Institute of Evolution, University of Haifa, Haifa, Israel, [3] Department of Plant Sciences, North Dakota State University, Fargo, ND, USA, [4] Biosciences Research Laboratory, United States Department of Agriculture-Agricultural Research Service, Fargo, ND, USA, [5] Cereal Disease Laboratory, United States Department of Agriculture-Agricultural Research Service, University of Minnesota, St. Paul, MN, USA, [6] Department of Plant Pathology, Kansas State University, Manhattan, KS, USA

Bread wheat has a large and complex allohexaploid genome with low recombination level at chromosome centromeric and peri-centromeric regions. This significantly hampers ordering of markers, contigs of physical maps and sequence scaffolds and impedes obtaining of high-quality reference genome sequence. Here we report on the construction of high-density and high-resolution radiation hybrid (RH) map of chromosome 4A supported by high-density chromosome deletion map. A total of 119 endosperm-based RH lines of two RH panels and 15 chromosome deletion bin lines were genotyped with 90K iSelect single nucleotide polymorphism (SNP) array. A total of 2316 and 2695 markers were successfully mapped to the 4A RH and deletion maps, respectively. The chromosome deletion map was ordered in 19 bins and allowed precise identification of centromeric region and verification of the RH panel reliability. The 4A-specific RH map comprises 1080 mapping bins and spans 6550.9 cR with a resolution of 0.13 Mb/cR. Significantly higher mapping resolution in the centromeric region was observed as compared to recombination maps. Relatively even distribution of deletion frequency along the chromosome in the RH panel was observed and putative functional centromere was delimited within a region characterized by two SNP markers.

Keywords: endosperm radiation hybrid panel, radiation hybrid map, wheat chromosome 4A, chromosome deletion bin map, Triticum aestivum, SNP iSelect array

## INTRODUCTION

Bread wheat (*Triticum aestivum* L.) is one the four most important crops grown world-wide. The availability of its genome sequence may significantly facilitate breeding for improved yield and resistance to biotic and abiotic stresses to withstand the changing environmental conditions. However, wheat genome sequencing is hampered by its large size (1C ∼ 17 Gb) and allohexaploid nature (AABBDD genome). To facilitate the wheat genome sequencing, the International Wheat Sequencing Consortium (IWGSC[1]) was established in 2005. The main challenge in obtaining a reference sequence of bread wheat is the ability to contiguously order BAC contigs or sequence scaffolds along the chromosomes (Choulet et al., 2014b). To achieve this goal, IWGSC follows

---

[1]www.wheatgenome.org

a strategy of physical mapping and sequencing of the individual chromosomes and chromosome arms (Eversole et al., 2014). An ideal physical map is fully oriented and anchored to high-resolution genetic map with high marker density (Meyers et al., 2004; Paux et al., 2008a; Ariyadasa and Stein, 2012).

Genetic maps are based on recombination between polymorphic molecular markers. A shortcoming of genetic maps is the variation in recombination rate along chromosomes and, in particular, strong suppression of recombination in peri-centromeric regions (Korol, 2013). About one-third of the wheat genome is located in recombination-poor regions (Akhunov et al., 2003). Thus, the order of loci within centromeric regions cannot be determined merely through recombination mapping. This problem can be solved by using the radiation hybrid (RH) mapping approach (Kumar et al., 2012a).

Radiation hybrid mapping is a physical mapping approach based on radiation-induced deletions for mapping markers (Michalak de Jimenez et al., 2013). A panel of independently derived RHs is assayed for the presence or absence of marker loci, and the patterns and frequencies of marker co-retention are used to calculate their physical proximity and to develop a RH map. RH mapping method has several advantages over genetic mapping: (1) recombination independence, (2) higher resolution in peri-centromeric and centromeric region, and (3) polymorphic marker independence (Kumar et al., 2014). Because of these advantages, RH mapping was used to anchor BAC (Bacterial Artificial Chromosome) contigs on chromosomes 3B and 6B of wheat (Paux et al., 2008b; Kobayashi et al., 2015).

Molecular markers are a critical component needed to anchor and orient physical maps. To anchor a given contig, one marker locus placed on a map is the necessary minimum, although two are required for proper contig orientation. Despite the fact that physical maps have improved over the past few years and the number of contigs covering a chromosome or whole genome has been decreasing (Frenkel et al., 2010), their number may still exceed several hundred per chromosome. Thus, hundreds of evenly distributed markers are needed and, in order to achieve a proper orientation of the contigs, the number of markers must be doubled (Kumar et al., 2012a). A similar situation concerns anchoring, orienting, and eventually merging of sequence scaffolds in genome sequencing projects (e.g., Choulet et al., 2014a). To satisfy this need, high-throughput mapping systems such as array based platforms have been employed (Ariyadasa and Stein, 2012). Currently, the 90K Illumina iSelect chip (Wang et al., 2014) is one of arrays with highest number of available SNP (single nucleotide polymorphism) markers in wheat. The chip allows mapping of approximately 2000 marker loci per chromosome (Wang et al., 2014). In ideal cases, it could allow anchoring and orienting of one thousand physical contigs per chromosome. Additionally, the high density maps with high resolution can facilitate precise mapping and cloning of agronomically important genes. For example, chromosome 4A of wheat has been reported to harbor over 50 genes involved in regulation of grain yield and quality, reaction to biotic and abiotic stresses, and genes regulation of physiological traits such as height, maturity and dormancy (Araki et al., 1999; Börner et al.,

2002; McCartney et al., 2005; Jakobson et al., 2006; Shorinola et al., 2016).

In this work, two endosperm radiation hybrid (ERH) panels consisting of 1069 lines were developed for wheat chromosome 4A as well as chromosome 4A deletion map. Selected lines of the ERH panels and wheat chromosome deletion bin lines were genotyped by the 90K iSelect chip (Wang et al., 2014) and corresponding maps were constructed. This work is a part of 4A project to aid development of a complete reference sequence of this chromosome.

## MATERIALS AND METHODS

### Plant Material

Endosperm radiation hybrid panels were developed from two independent crosses between hexaploid wheat (*T. aestivum*, L.) cultivar 'Chinese Spring' (CS) whose pollen was irradiated and nulli-tetrasomic (NT) lines N4AT4B and N4AT4D, which served as female parents (Supplementary Figure S1). Fifteen chromosome deletion lines of CS chromosome 4A (four for short arm and 11 for long arm; Endo and Gill, 1996; **Table 1**) provided by the NBRP-WHEAT Centre (Japan) were used for chromosome deletion bin map construction and validation of RH maps. The 4AS and 4AL chromosome arms were flow sorted from the 4A double ditelosomic line of CS in which chromosome 4A is represented by a pair of telosomes representing the short (4AS) and the long arm (4AL) of 4A (Sears and Sears, 1978). Grains of this stock and CS were kindly provided by Dr. Bikram S. Gill (Department of Plant Pathology, Kansas State University, Manhattan, NY, USA).

### ERH Panels Development

Endosperm radiation hybrid panels were prepared as described by Kumar et al. (2012a) and Tiwari et al. (2012). Briefly, dehiscent spikes of CS were irradiated using gamma rays (10 Gy, 15

**TABLE 1 | 4A deletion stock with their fragment length and accession number.**

| Deletion stock | Breakpoint | Accession number |
|---|---|---|
| 4AS-01 | 0.20 | LPGKU1156 |
| 4AS-02 | 0.71 | LPGKU1157 |
| 4AS-03 | 0.76 | LPGKU1158 |
| 4AS-04 | 0.63 | LPGKU1159 |
| 4AL-01 | 0.85 | LPGKU1143 |
| 4AL-02 | 0.75 | LPGKU1144 |
| 4AL-04 | 0.80 | LPGKU1146 |
| 4AL-05 | 0.66 | LPGKU1147 |
| 4AL-06 | 0.84 | LPGKU1148 |
| 4AL-07 | 0.66 | LPGKU1149 |
| 4AL-09 | 0.73 | LPGKU1150 |
| 4AL-10 | 0.82 | LPGKU1151 |
| 4AL-11 | 0.66 | LPGKU1152 |
| 4AL-13 | 0.59 | LPGKU1154 |
| 4AL-14 | 0.79 | LPGKU1155 |

Gy, or 20 Gy) and the pollen from the irradiated spikes was immediately used to pollinate previously emasculated spikes of N4AT4B and N4AT4D lines (Supplementary Figure S1). The irradiation was done using an Acel Gamma Cell 220 irradiator (Gamma Irradiation Facility, North Dakota State University, Fargo, ND, USA). Seeds were harvested 20 days after pollination and endosperm was dissected from each seed as described by Tiwari et al. (2012). Endosperms were individually placed in microtubes and stored at −80°C until DNA extraction.

## DNA Extraction and Characterization of ERH Panels

DNA extraction from plant tissues and endosperm of the 4A RH panels was done using Invisorb® Spin Plant Mini Kit (Stratec Biomedical, Berlin, Germany) following the manufacturer's protocol. Leaf tissues from chromosome deletion lines and NT stocks were harvested from 3 weeks-old seedlings and lyophilized. Dry tissues were homogenized using two 5 mm glass beads and MM3 mill (Retsch, Haan, Germany). The homogenization was done 4 min and 27 Hz. Similarly, the endosperm tissues of individual 4A RH lines were desiccated and homogenized two times for 4 min at 27 Hz. The amount and quality of extracted DNA samples was checked using a NanoDrop Spectrophotometer (Thermo Scientific, Waltham, MA, USA) and electrophoresis in 0.8% agarose gel, respectively.

Telocentric chromosomes 4AS and 4AL were flow-sorted as described by Kubaláková et al. (2002) and their DNA amplified by multiple displacement amplification as described by Šimková et al. (2008).

Equivalent of 4A monosomic lines in N4AT4B and N4AT4D genetic background were used as controls (Table 2). The 4A monosomic lines were reconstructed by equimolar pooling of 10 ERH lines which showed no deletions (100% of marker retention) for each genetic background.

## Panel of 4AS- and 4AL-Specific Markers Design and ERH Lines Verification and Characterization

To select ERH lines with maximal informative content, a set of reliable 4A-specific sequence-tagged sites (STS) markers was developed. Chromosomal distribution of the marker candidate loci was deduced from the 4A GenomeZipper (Hernandez et al., 2012). mRNA sequences of rice genes in the selected

**TABLE 2 | Controls included in single nucleotide polymorphism (SNP) genotyping.**

| Control | Number of 4A copies |
| --- | --- |
| CS | 2 |
| Mono4ATetra4B | 1 |
| Mono4ATetra4D | 1 |
| N4AT4B | 0 |
| N4AT4D | 0 |
| 4AS | Short arm |
| 4AL | Long arm |

loci (Table 3) were used to annotate homologous wheat genes in the linked Chinese Spring chromosome survey (CSS) sequences scaffolds (IWGSC, 2014) of all three wheat sub-genomes. For each gene, at least two pairs of primers were designed in the most variable regions. Specificity of the primers for 4A was tested on DNA of NT lines and chromosome deletion lines by multiplex PCR with an internal standard *owm37* (F: CAGACACGAGATTTGATAAGGCTA, R:TGCTGAAAACACTCTTTCAACAC). PCR reactions were carried out in 20 µl of reaction volume using 15 ng of genomic DNA, 100 mM Tris-Hcl, 500 mM KCl, 10 mM $MgCl_2$, 1% Triton X-100), 1 µM primers, 200 µM dNTPs (Invitrogen, Waltham, MA, USA), 0.01% Cresol red (Sigma–Aldrich, Dorset, UK), 1.5% sucrose (Lachner, Neratovice, Czech Republic) and 0.5 units of Taq polymerase (Finzyme, Vantaa, Finland). PCR conditions were used as follow: initial denaturation 95°C for 5 min, 40 cycles of: 30 s at 95°C, 30 s at 60°C, 60 s at 72°C, followed by a final extension at 72°C for 10 min. The PCR products were separated on a 4% non-denaturing polyacrylamide gel and visualized using ethidium bromide staining.

Confirmed chromosome 4A-specific markers and the internal standard were used for characterization of 414 ERH lines in a multiplex setup. The multiplex PCR reactions were carried out in 10 µl volumes and analyzed as described above. ERH lines with at least one deletion and not missing the whole chromosomal arm were considered for further analysis.

## iSELECT SNP Genotyping and Data Analysis

DNA samples of 119 selected ERH lines and DNA of control lines (Table 2) and lines of the chromosome deletion stocks (Table 1) were sent for genotyping at USDA-ARS Small Grains Genotyping Center, Fargo[2]. The samples were genotyped with the Illumina iSelect 90K SNP array (Wang et al., 2014), following Akhunov et al. (2009). The control DNA (Table 2) and DNA of 15 chromosome deletion lines (Table 1), were genotyped in three replicates.

The genotyping results were manually analyzed using Illumina's GenomeStudio (GS) v2011.1 software. In this case, identification of signal differentiated to four categories in the workspace was expected. The categories were: (1) no signal (deleted marker locus: NT, RH, and cytological deletion lines with absence of the marker locus), (2) signal representing monosomic 4A chromosome (reconstructed 4A monosomic lines, arm-specific DNA and RH lines with presence of marker locus), (3) disomic 4A chromosome (chromosome deletion lines with presence of marker locus and CS), and (4) questionable signals. The module of GS for mapping in tetraploid species which allows discrimination of five different signal clusters was used. Each genotype call was manually evaluated and clustered in one of the above mentioned categories manually selected based on signal intensity. Clustering decisions were made according to the signal level from the controls including the 4A nullisomic, monosomic, and disomic lines (Table 2; Supplementary Figure S2). The resulting genotypes were exported to MS Excel sheet and the

[2]http://wheat.pw.usda.gov/GenotypingLabs/

**TABLE 3 | Panel of chromosome 4A-specific STS markers.**

| Rice gene | SCC scaffold | Marker | Primer sequence | 4A deletion bin |
|---|---|---|---|---|
| Os03g0180400 | 4AS_5944160 | **owm160** | AAGGGCCCATATCATATCACAC | 4AS3 – 0.76−1 |
| LOC_Os03g08280 | | | AACAGTGGAGGGCTTTGCTA | |
| Os03g0161800 | 4AS_5934783 | **owm161** | TTTTCAAGCAGGTTTTGTGC | 4AS3 – 0.76−1 |
| LOC_Os03g06620 | | | TCACTTCTCTTCTTTGCGTTCA | |
| Os03g0203700 | 4AS_5934571 | **owm129** | TGATGATACCAGAGCGTACAAT | 4AS3 – 0.76−1 |
| LOC_Os03g10640 | | | CCTTTGATAAGAGGCCCTCAG | |
| Os03g0297400 | 4AS_6015094 | **owm126** | CCAGTCAGAAATTATTATGAACCTATC | 4AS1 – 0.20–0.63 |
| LOC_Os03g18590 | | | CGCTGTCTCGAGATTGGAGT | |
| | | owm127 | CAGCAAATGCATGATTTCACTAAT | |
| | | | TTCAGATACAGTTCCTGATCTTGC | |
| Os03g0736300 | 4AL_7166434 | **owm121** | ATTGCCGTCGCGAACTAGA | 4ALC – 0–0.59 |
| LOC_Os03g52630 | | | CGGGACGAGCTTGACGAT | |
| | | owm162 | TGTTCAAGGACAGCAAGACG | |
| | | | CATTTAGATGCTGTCATATTGCTTG | |
| Os03g0684700 | 4AL_7176697 | owm165 | TGAGTTACAGCCACTCTTGTGC | 4AL13 – 0.59–0.66 – 1 |
| LOC_Os03g48030 | | | ACCACCTGCCAAGGTTCCTA | |
| | | **owm166** | TGCTACCATGGTTCAGAATGA | |
| | | | AGTTGACGAAGCGGCCTTT | |
| Os03g0854800 | 4AL_7142517 | **owm119** | ACTTGGGAACATTCAGCTCTT | 4AL13 – 0.59–0.66 – 1 |
| LOC_Os03g63770 | | | TTTCTCCTCTGTTGGAACATCA | |
| Os03g0145800 | 4AL_7091911 | **owm167** | TTTTCTTGGTCAGTATAACCTGTTTTT | 4AL2 – 0.75–0.79 |
| LOC_Os03g05260 | | | TGAGCAGAGAAAAATTTCCAAG | |

*Set of 4A specific STS markers designed for characterisation of ERH panels with distribution supported by 4A GenomeZipper (Hernandez et al., 2012). Markers in bold were used for ERH lines characterization.*

genotypes were converted to binary codes where questionable signals were labeled as missing data.

## iSELECT 90K SNP Cytological Deletion and RH Maps

### Verification of Chromosome Deletion Map

The CS chromosome deletion lines (**Figure 1**; Endo and Gill, 1996) were genotyped in three replicates. Deletion lines which exhibit disomic-like genotype were included in the CS-like cluster and deletion lines with 4A nullisomic-like signals were designated as members of NT-like cluster (Supplementary Figure S2). Clusters were considered reliable when at least two control replicates were in the cluster. Signals of SNP markers which cannot be clearly assigned to either cluster were marked as questionable and were not included in the map construction.

Assignment of markers to cytological deletion bins was based on an MS Excel ordering tool. Markers specific to a particular chromosome arm were verified by comparing its cluster position with the marker signals on the 4AS and 4AL chromosome specific DNAs. Markers specific for the arm, but missing in all deletion lines, were assigned to the most distal deletion bin of the arm. Additionally, markers common for both arm DNAs and present in all deletion lines were assigned to the centromeric bin defined as an overlap between the 4AS and 4AL telosomic chromosomes. The arm specific peri-centromeric bins were defined as group of markers specific for the chromosomal arm and common for all deletion lines, but not present in the centromeric bin.

## RH Map

Direct ordering of all markers based on minimizing the map length (calculated as sum of RH distances between sequential markers; Newell et al., 1998) failed due to artificially small distances caused by missing data (pmissing is 0.08, STDV is 0.06) and highly varying marker presence (pA) frequencies (analogous to allele segregation in genetic mapping: average pA is 0.75, STDV is 0.09). Data filtering based on the assumption that co-segregating markers ("twins") are more reliable for mapping than singleton markers (Ronin et al., 2015) also failed because only a small number of "exact twins" of markers were found in this dataset.

For the RH map construction combined data of B and D panels (2711 markers) were used. Markers with missing genotypes in more than 50 lines and markers having pA higher than 0.8 or smaller than 0.7 were temporally excluded from the analysis. The remaining 1405 markers were clustered by single-linkage algorithm with a cut-off for the proportion of difference in genotypes (at lines where genotypes were not missed in both markers) equals to 0.06.

As chromosome structure is linear, marker clusters with non-linear structure were considered as problematic and subdivided into parts with linear structure. This linearization editing was conducted using LTC software (Frenkel et al., 2010). Markers corresponding to nodes from the diametric paths of the resulted linear clusters were used as candidate skeleton markers. Using MultiPoint program (Mester et al., 2003; Ronin et al., 2010), the set of skeleton markers was globally ordered
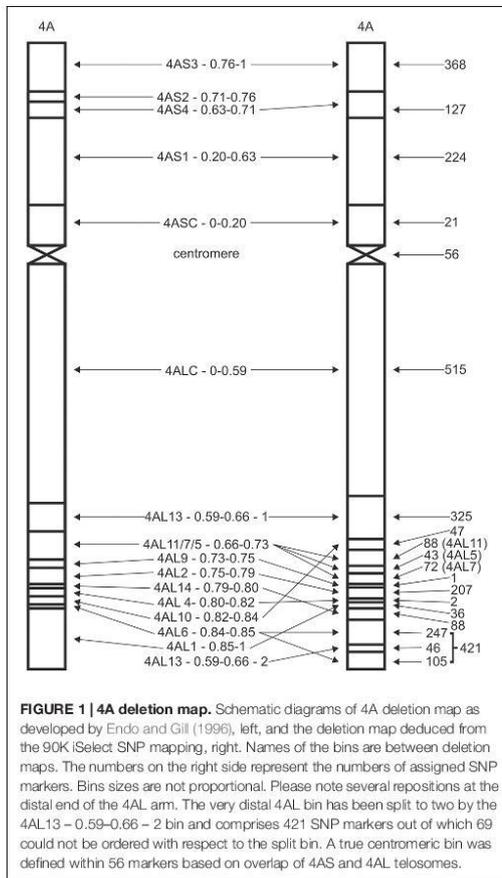
**FIGURE 1 | 4A deletion map.** Schematic diagrams of 4A deletion map as developed by Endo and Gill (1996), left, and the deletion map deduced from the 90K iSelect SNP mapping, right. Names of the bins are between deletion maps. The numbers on the right side represent the numbers of assigned SNP markers. Bins sizes are not proportional. Please note several repositions at the distal end of the 4AL arm. The very distal 4AL bin has been split to two by the 4AL13 – 0.59–0.66 – 2 bin and comprises 421 SNP markers out of which 69 could not be ordered with respect to the split bin. A true centromeric bin was defined within 56 markers based on overlap of 4AS and 4AL telosomes.

and further edited to exclude markers causing (i) local map instability upon jackknife re-sampling and (ii) violation of local monotonic increase of distance along the map, and include additional markers that fit well into large or moderate gaps. Coordinates of the obtained 144 skeleton markers were calculated as cumulative sum of distances between adjacent markers. The resulted skeletal map was verified using the bin map.

For the remaining non-skeleton markers, their distances for all skeleton markers were calculated using RH formula (Cox et al., 1990) and Haldane formula for simple recombination event, because at small distances the corresponding two mapping metrics result in highly correlated distances. A marker was attached to the RH map if the Haldane distance to the closest skeleton map marker was less than "10 cM." Markers, which did not fit this condition, were added to larger map regions using their bin map positions.

The resulting RH map was manually checked taking into the account the position in chromosomal deletion bin map and RH genotype. The STS markers used for ERH line characterization were added. Markers with higher level of missing data and problematic lines were excluded before final map length calculation.

## Mapping of SNP Markers to Chromosome Survey Sequences

To verify marker specificity for chromosome 4A and its regions, comparative analysis was performed with survey sequences from chromosome arms 4AS, 4AL, 4BS, 4BL, 4DS, 4DL, 5AS, 5AL, 5BS, 5BL, 5DS, 5DL, 7AS, 7AL, 7BS, 7BL, 7DS, and 7DL (IWGSC, 2014). The arm sequences were repeat masked using the mips-REdat database (Nussbaumer et al., 2013) and Vmatch software (Abouelhoda et al., 2002). The alignments between marker sequences and the wheat chromosome arm sequences were performed using the BLAST algorithm (Altschul et al., 1990). The BLAST outputs were post-processed by an in-house perl script, which filtered the results based on the following criteria: a minimum identity 90% and a minimum alignment length of 50 base pairs. Finally, the best hit for each alignment was selected.

## RESULTS

### 4AS and 4AL Specific Marker Panel for Characterization of ERH Lines

Selection of RH lines with the highest information content (high number of small deletions, but not missing whole chromosome 4A, or its arm) was ensured by a set of 23 primer pairs that were designed for five and six STS specific for 4AS and 4AL chromosomal arms, respectively. Even distribution of the loci along chromosome 4A was facilitated by 4A GenomeZipper (Hernandez et al., 2012). Five and six primer pairs were found specific for four loci on each 4AS and 4AL chromosome arms, respectively, and denominated as *owm119 – owm167* markers (**Table 3**). Eight most reliable markers, one for each locus were used for characterization of the ERH panels (**Table 3**).

### ERH Panels Development and Characterization

A total of 1069 ERH lines specific for chromosome 4A were developed using three dosages of radiation (10, 15, and 20 Gy; **Table 4**). According to the female parent used (N4AT4B or N4AT4D), the lines were divided into two ERH panels (Supplementary Figure S1). As expected, the recovery of ERH lines negatively correlated with radiation dosage (**Table 4**). DNA was extracted from endosperm of 414 ERH lines representing all three irradiation levels (**Table 4**). A total of 140, 137, 138 lines belonging to 10, 15, and 20 Gy panels, respectively, were characterized with eight 4A-specific STS markers (**Table 3**) using multiplex PCR with internal standard. Estimation of the retention frequency (Kumar et al., 2012a), based on the

| Dose (Gy) | ERH lines prepared | | DNA extracted | | ERH lines genotyped | |
|---|---|---|---|---|---|---|
| | N4AT4B | N4AT4D | N4AT4B | N4AT4D | N4AT4B | N4AT4D |
| 10 | 250 | 339 | 72 | 68 | 12 | 15 |
| 15 | 69 | 163 | 68 | 68 | 26 | 24 |
| 20 | 105 | 143 | 70 | 68 | 22 | 20 |
| Total | 424 | 645 | 210 | 204 | 60 | 59 |

eight STS markers, showed that 7.1% lines lost the entire chromosome (retention frequency = 0%), whereas 27.1% of the lines did not show any deletion on chromosome 4A (retention frequency = 100%). Additionally, 17.7 and 19.2% of the lines lost complete short and long arm of chromosome 4A, respectively (Supplementary Table S1). A total of 113 lines, 57 and 56 for N4AT4B and N4AT4D ERH panel, respectively, were selected for genotyping by 90K iSelect chip (Wang et al., 2014). The main criterions for selection were the amount of available DNA (at least 2 μg), retention frequency (10–90%) and homogeneous distribution of lines across the three irradiation doses and both female parents. In addition, six lines showing 100% retention frequency were randomly selected from 20 Gy panels (**Table 4**).

## Marker Calling and RH Map Construction

The selected 119 ERH lines were genotyped along with DNA of three replicates of 15 chromosome deletion lines, five control lines and DNA amplified from flow-sorted 4AS and 4AL arms (**Tables 1** and **2**). The samples were genotyped with the iSelect chip (Wang et al., 2014) and genotypes of 81587 marker loci were retained. Manual analysis was found the most effective, as it allowed resolving clusters with non-standard shape and individually judge questionable signals. Additionally, average distance between clusters, especially for clusters with nullisomic and monosomic signals, changed significantly from sample to sample (Supplementary Figure S2). The analysis yielded 2711 SNP markers specific for CS chromosome 4A and polymorphic for the ERH panels. Six and eight markers were found polymorphic only for the N4AT4B and N4AT4D based ERH lines, respectively. A total of 2697 SNP markers (>99%) were common for both ERH panels. Based on the results of SNP genotyping, a total of 23 ERH lines were excluded from RH map construction. These included ten N4AT4B and eight N4AT4D lines, either with 100% retention frequency, or with questionable calls across all markers.

## Verification Chromosomal Deletion Map

Construction of the 4A deletion map was included in this study to verify marker ordering along the chromosome arms in RH map. First, chromosome 4A-specific SNPs were assigned to corresponding chromosome deletion bins based on their presence in chromosome deletion lines (**Table 1**). Out of 2711 4A-specific markers, 24 were excluded because of unreliable signal calling. In total 2687 SNP markers were assigned to 16 different

deletion bins (**Figure 1**; Supplementary Table S2). Surprisingly, no SNP was assigned to 4AS2 – 0.71–0.76 bin and only one SNP was assigned to 4AL9 – 0.73–0.75 bin. The highest number of SNPs (515) was assigned to 4ALC – 0–0.59 bin (**Figure 1**; Supplementary Table S2). The large number of markers offered high resolution and, in combination with DNA obtained from telocentric chromosomes 4AS and 4AL, allowed identification of a new deletion bin in 4AL telomeric region and precise characterization of centromeric bin of the CS 4A chromosome. Besides the new bin identification, the higher resolution enabled better characterization of bins and their reordering (**Figure 1**; Supplementary Table S3). Linear order of 4AS bins was compared against deletion maps published at Graingenes[3], and confirmed previous findings. On the other hand, the order of bins in the distal part of the 4AL chromosome arm changed significantly. Bins, 4AL1 —0.85–0.86, 4AL14 – 0.79–0.80, and 4AL10 – 0.82–0.84 were rearranged (**Figure 1**) and the bin 4AL – 0.66–0.73 previously delimited by deletion lines 4AL-11/7/5 (LPGKU1152, LPGKU1149, LPGKU1147; **Table 1**) was split to three (**Figure 1**; Supplementary Table S3).

### The New 4AL Distal Chromosomal Deletion Bin

Data analysis showed that the deletion line 4AL-13 (LPGKU1154; **Table 1**) comprises two fragments of the 4AL chromosome arm. The sub-centromeric region represents original observation of Endo and Gill (1996), but its distal end is combined with a short fragment of 4AL sub-telomeric region; this was not reported previously. This finding was confirmed by the fact that 46 markers identified in this new bin were present only in 4AL-13 deletion line, telocentric chromosome 4AL and control lines with complete chromosome 4A. Additionally, sequences of these markers were found homoeologous to chromosome specific CSS of chromosome arms 4AL, 7AS and 7DS (IWGSC, 2014; Supplementary Table S3).

### 4AL Centromeric Bin

The centromeric bin was defined as overlap of flow-sorted telocentric chromosomes 4AS and 4AL and determined with 56 SNP marker loci (2% of all mapped markers). The bin was named as 4A-centromere. The original centromeric bin of 4A chromosome delimited by the 4AL-13 and 4AS-01 lines (**Table 1**) was partitioned to three bins and increased the number of detected bins to 18 (**Figure 1**). This partition divided marker loci of the original centromeric bin to 21 for 4ASC – 0–0.20 bin, 56

---

[3]http://wheat.pw.usda.gov

for 4A-centromere and 515 for 4ALC – 0–0.59 bin (**Figure 1**; Supplementary Table S3).

## Radiation Hybrid Map of Chromosome 4A

The RH map was constructed using all 2687 polymorphic SNPs following Newell et al. (1998) and Ronin et al. (2015). In the first step, markers with missing data genotypes in more than 50 lines and markers with highly varying pA frequencies higher than 0.8 or smaller than 0.7 were temporally excluded from the analysis. The remaining 1405 loci were clustered by single-linkage algorithm. Two clusters were obtained, one with 931 and the second with 434 linked marker loci. The remaining 40 loci were singletons (presumably due to higher level of errors). Both clusters were checked for linear structure (Frenkel et al., 2010). Markers corresponding to nodes from diametric paths of the resulted parts with linear clusters were used as candidate skeleton markers. Using MultiPoint program (Mester et al., 2003; Ronin et al., 2010), the skeleton markers were globally ordered and further edited to exclude problematic markers. A total of 144 skeleton markers were obtained and the resulting map was verified using chromosome deletion bin map (Supplementary Figure S3). Remaining non-skeleton markers were added to RH map based on their chromosome deletion bin position. The resulting RH map comprised 2467 markers, which were then manually checked. The final RH map comprises 2316 markers (2308 SNP, 8 STS; Supplementary Table S3) spanning the length of 6550.9 centi Ray (cR). Considering the molecular size of chromosome 4A of 856 Mb (Šafář et al., 2010), the average resolution of the final RH map is 0.13 Mb/cR.

In the 96 informative ERH lines (lines with deletions), the number of deletion events ranged from 1 to 72 per line, with an average of 11.25 deletions per line. For the 4AS arm, 369 deletions were identified with a maximum 31 and an average 3.84 deletions per line. For the 4AL arm, 711 deletions were retained with a maximum 65 and an average 7.41 deletions per line. Average of 1.16 and 1.32 deletions per Mb for 4AS and 4AL chromosome arms, respectively, were observed considering their respective sizes to be 316 and 540 Mb, (Šafář et al., 2010).

The deletion frequency per marker locus in the ERH panels varied from 0 to 39 with an average 27.7 deletions per locus (**Figure 2**). The frequency of deletions along the chromosomal arms was slightly higher for the long arm (average 28.75) compared to short arm (average 25.43) (**Figure 2**). This corresponds with higher deletion rate per line as described above. If we consider regions on the 4A RH map with continuous marker deletion frequency below average and along the centromere as peri-centromeric regions (**Figure 2**), then the 4AS and 4AL peri-centromeric regions comprise 175 and 387 markers, respectively. If we consider, even distribution of the markers along the chromosome, then the 4AL peri-centromeric region would be 2.21-fold longer compared to the 4AS peri-centromeric region (**Figure 2**). This assumption is corroborated by the presence of relevant markers in the centromeric bins of the chromosome deletion map. This would also mean that only portion of 4AS peri-centromeric region flanked by markers *tplb0056l21_172* and *Kukri_c8543_3646* is covered by 4ASC – 0–0.20 deletion bin. The 4AS peri-centromeric region is also included in 4AS1 – 0.20–0.63 and 4AS4 – 0.63–0.71 bins whereas the 4AL peri-centromeric region flanked by markers *Kukri_c42920_338* and *Kukri_rep_c104642_308* is fully covered with the 4ALC – 0–0.59 deletion bin.

## Evaluation of 4A RH and Chromosome Deletion Maps

To evaluate the mapping accuracy, resolution, and specificity, we analyzed SNP markers and mapping results using two approaches. First, 4A-specific markers for chromosome deletion and RH maps identified in this study were compared with 4A recombination based consensus genetic maps developed using eight hexaploid (1928 lines) and 13 tetraploid (1332 lines) wheat mapping populations genotyped with the same Illumina platform (**Figure 3**; Wang et al., 2014; Maccaferri et al., 2015). Consensus genetic maps of 4A for hexaploid and tetraploid wheat (Wang et al., 2014; Maccaferri et al., 2015) shared 909 and 677 SNP markers, respectively, with our 4A RH map. In the hexaploid consensus recombination map, markers from deletion bins 4AS4 – 0.63–0.71 to 4ALC – 0–0.59 representing 63.4% of chromosome 4A are mapping to a region of 30 cM representing only 4.6% of the map. On the other hand, in the 4A RH map, the same region represents 3000.3 cR, or 45.8% of the map (**Figure 3A**). In the tetraploid wheat based map the region corresponds to 18.9 cM of the map and represents 10.8% of the map (**Figure 3B**).

Additionally, both recombination maps being in concordance with the RH map suggests that the newly identified 4AL distal bin (based on line 4AL-13; **Table 1**) belongs to the middle part of the most distal 4AL bin denominated by line 4AL-06 (**Figure 3A**; Supplementary Table S3) increasing of number of detected deletion bins to 19 (**Figure 1**). According to the recombination map, the orientation for a portion of markers from the bins 4AS3 – 0.76-1, 4AL2 – 0.75–0.79, and 4AL6 – 0.84–0.85 - 1 are inverted in the 4A RH map (**Figure 3**). On the other hand, besides the inverted block of markers, there are only few markers (1 and 2 markers in hexaploid and tetraploid consensus map, respectively) misplaced and caused the discrepancy between the recombination consensus maps and the deletion and RH map (**Figure 3**).

The second approach to evaluate the specificity of the mapped SNPs was by BLASTn comparison with survey sequences of individual chromosomes (4B, 4D, 5B, 5D, 7A, and 7D; IWGSC, 2014). A total of 2381 SNP loci could be reliably mapped to chromosome 4A survey sequences. 306 of the SNPs were not homologous with chromosome 4A survey sequences (Supplementary Table S3), but they produced reliable genotyping signals on the 4AS/4AL arm-specific DNA. Also, homology test with homoeologous chromosomes (4B, 4D, 5B, 5D, 7A, and 7D) confirmed the specificity of 132 of the markers to the expected genomic regions (Supplementary Table S3).

**FIGURE 2 | Deletion rate along the 4A chromosome in 96 RH lines.** The chart shows number of lines with deletion per SNP locus (y-axis) along the 4A chromosome based on marker order in the 4A RH map (x-axis). The deletion rate is stable in the most of the chromosome arms with significant drop in peri-centromeric regions. The peri-centromeric regions could be delimited as continual decrease below an average line. However, the decrease in the number of identified deletions in the peri-centromeric regions may be also accounted to lower marker density because used markers are associated with transcribed genome regions, only. The zero deletion rate is presumably associated with functional centromere.

## DISCUSSION

Main advantages of RH maps are higher resolution in chromosomal regions with suppressed recombination and a possibility to map markers without a need for polymorphism [reviewed by Kumar et al. (2014)]. In this work, two endosperm-based RH panels were developed as sets of lines in which the irradiated 4A chromosome was in monosomic composition to avoid heterozygosity issues (Supplementary Figure S1). Due to a limited resolution of the 90K SNP iSelect array (Wang et al., 2014) in distinguishing signal levels from chromosomes in nullisomic and monosomic status, a wide range of controls including lines without chromosome 4A (nullisomic), with one copy of 4A (monosomic) and two copies of this chromosome (disomics) were used. Without the controls reliable identification of signal clusters would not be possible in many cases (**Table 2**; Supplementary Figure S2).

### 4A SNP Chromosome Deletion Bin Map

To facilitate and verify marker ordering during RH map construction, 15 deletion lines with defined deletions in chromosome 4A were selected (**Table 1**; Supplementary Table S2) and genotyped with the iSelect chip (Wang et al., 2014). In previous studies 786 EST loci were mapped in 10 4A-specific deletion bins (Miftahudin et al., 2004; Qi et al., 2004). Nine bins were delimited using the deletion lines while the centromeric bin was delimited by markers presence in all deletion lines but absent in NT lines for specific chromosome (Qi et al., 2004).

In the present study, out of the 81587 genotyped SNP markers, 2687 4A-specific SNP markers could be mapped on the 15 deletion lines and 4AL and 4AS telosomes. The mapped markers divided the chromosome 4A into 19 bins (4 and 14 deletion bins for 4AS and 4AL, respectively, and centromere). Reliability of the marker ordering was tested by *in silico* mapping of the markers to homoeologous chromosomal loci (Supplementary Tables S2 and S3). Sixteen bins were identified based on the deletion lines and two new bins were identified because of the chromosomal rearrangement. The centromeric bin was delimited as an overlap between the 4AL and 4AS telosomic chromosomes within the 56 markers (**Figure 1**). In contrast to Qi et al. (2004), this approach

allows more precise separation of the centromeric bin from the sub-centromeric bins.

Ordering the deletion map based on chromosomal fraction length (FL) (Endo and Gill, 1996) did not match the SNP marker genotypes. Although this was unexpected observation, such inconsistencies in measurements of chromosomal deletions were reported previously (Qi et al., 2003; Dilbirligi et al., 2004; Sourdille et al., 2004; Philippe et al., 2013; Belova et al., 2014; Akpinar et al., 2015). In this study, a single bin 4AL – 0.66–0.73 previously defined by three lines 4AL-05 (LPGKU1147), 4AL-07 (LPGKU1149), and 4AL-11 (LPGKU1152) was split to three bins containing 43, 72, and 88 SNP loci, respectively (**Figure 1**; Supplementary Table S2). On the other hand, in some lines, even the deleted chromosomal fraction could be distinguished, little or no markers could be mapped in bins defined by them. For example the 4A bins originally defined as 4AS2 – 0.71–0.76, 4AL9 – 0.73–0.75, and 4AL4 – 0.80–0.82 by Endo and Gill (1996) comprised only 0, 1, and 2 SNP markers, respectively (**Figure 1**; **Table 4**). Furthermore, these bins were not supported by any previous cytological deletion mapping studies (Qi et al., 2003, 2004; Miftahudin et al., 2004). These discrepancies could be explained by uneven chromosome condensation resulting in low resolution of cytogenetic measurements during the FL calculations and possibility that the small bin size or DNA content do not allow for suitable SNP discovery according to Wang et al. (2014) approach.

### The New 4AL Distal Bin

A new 4AL distal deletion bin was delimited through identification of telomeric chromosomal fragment attached to the chromosome 4A deletion line 4AL-13 (LPGKU1154; **Figure 1**; Supplementary Table S2). However, construction of the RH map and comparison with the consensus recombination maps indicated that the situation is more complicated and that the region likely originated from the central part of the 4AL6 – 0.84–0.85 bin splitting it into two. This increases the number of 4A deletion bins to 19. Additionally, 45 out of 46 SNPs from this new bin were mapped to survey sequence homoeologs (Supplementary Table S3) in 7AS and 7DS arms, confirming the distal localization of the fragment in the translocated

FIGURE 3 | Comparing of 4A RH map with recombination consensus maps of 4A chromosome of bread and pasta wheats. (A) Comparison of the deletion map supported 4A RH map with hexaploid bread wheat recombination consensus 4A map (Wang et al., 2014). From the left to right – deletion map, RH map and recombination map. The overall marker order agrees with a few exceptions. The major differences are the significant increase in the resolution and marker density at centromeric region of the RH map, and significant differences at distal ends of the chromosomal arms (e.g., the distal end of the 4AL arm shows inversions and marker translocation). The regions (similarly to the centromere region) show significant decrease in marker synteny exhibited as a lack of lines connecting the common markers. (B) Comparison of the deletion map supported 4A RH map with tetraploid durum wheat recombination consensus 4A map (Maccaferri et al., 2015). From the left to right – deletion map, RH map and recombination map. As expected, this map shares fewer markers with the RH map. Most of the features are similar to the comparison shown in (A) except for more conserved marker synteny at the end of the 4AL arm. The resolution of this map at the centromeric region seems to be better as compared to the hexaploid recombination map. Gray lines between the maps link positions of the same SNP markers. Centromeres are denominate by the deletion bin "centromere."

**FIGURE 4 | Comparison of chromosome 4A maps developed in this work with the published 4A SNP maps of bread wheat.** Chromosome 4A RH and deletion maps are compared with the recombination consensus map of 4A chromosome (Wang et al., 2014) and whole genome RH map of bread wheat in which 4A is represented by linkage groups 4AS, 4AL1, and 4AL2 (Tiwari et al., 2016). Note the absence of SNP markers in centromeric region in the whole genome RH map and disagreement in marker order and the number between the whole genome RH map and the remaining maps. All maps were obtained using 90k iSelect SNP array (Wang et al., 2014).

7BS segment. Endo and Gill (1996) also observed fusion of chromosomal fragments from different chromosomes and such lines were excluded from deletion stocks. However, this seems to be a small fragment that was likely below the detection limit of cytological banding analysis. This finding may explain the observation of Miftahudin et al. (2004) who identified 7AS/7DS homologous segment before 5AL translocated segment of 4AL chromosome arm (Supplementary Figure S4). Identification of the 7AS/7DS segment may be a result of the attachment of the sub-telomeric region in the deletion line 4AL13 (**Figure 1**). This is corroborated by the absence of such 7AS/7DS segment in the RH map. Additionally, one of the ESTs delimiting the region (BE426203) maps to the CSS sequence scaffold 4AL_7170232 which contain SNP marker Tdurum_contig82236_117 mapped to the very distal 4AL bin 4AL6 – 0.84–0.85. Unfortunately, none of other ESTs from the region is homologous with the SNPs or SNP linked CSS sequence scaffolds. These findings will require further verification.

Additional discrepancy with previously published deletion maps represents segment of ancestral 4AS chromosome on the very distal region of present time 4AS chromosome (Supplementary Figure S4, Miftahudin et al., 2004). This segment was not supported by the 4A RH map. Comparison of the ESTs used to identify the region (BE518074 and BE494743)

with the CSS sequences (IWGSC, 2014) showed that BE518074 has 100% identity with sequences of all three chromosomes of group 3 and 96% and only 84% identity with the 4AS and 4DS chromosome, respectively. Also the BE494743 EST has only 93 and 89% identity with the 4AS and 4DS chromosomes, respectively. Both ESTs were mapped to 4AS and 4DS arms, only by hybridization[4]. This suggests that the region was most likely assigned incorrectly to the ancestral 4AS because of paralogous sequences. Moreover, the BE518074 EST contains conserved domain PTZ00067 belonging to 40S ribosomal protein S23 subfamily, corroborating a possibility of misleading hybridization signal. However, more evidences will be needed to resolve these inconsistencies.

## Centromeric Bin

Genetic characterization of centromere is difficult since (peri)centromeric regions exhibit low levels of recombination, if any at all (Chen et al., 2008). Several approaches have been used to define centromeric regions in wheat with the deletion bin mapping being the most popular. Most proximal bins of short and long arms are marked as peri-centromeric and centromeric (Qi et al., 2004; Belova et al., 2014; Wang et al.,

---

[4]http://wheat.pw.usda.gov

2015). More precision was acquired by combining deletion bin maps and recombination maps (Marone et al., 2012). In a previous study (Qi et al., 2004), presence of markers in all deletion lines and absence in NT and DT lines led to their assignment to a centromeric bin of the specific arm. In this work, simultaneous presence of markers in the 4AS and 4AL telocentric chromosomes was used to define centromeric region. The centromeric region was delimited by 56 SNP marker loci (Supplementary Table S3). Further narrowing of the region may be possible by using additional ERH lines. As the chromosome requires an intact centromere to survive cell division, markers from the centromere should not be deleted in any of the RH lines. Out of the 56 markers from the centromeric region, only two markers (Excalibur_rep_c94194_201, and Kukri_c51716_802; Supplementary Table S3) were not deleted in the ERH lines (**Figure 2**). The region characterized by these markers could be considered as the region representing functional centromere of chromosome 4A.

## 4A SNP ERH Map

The 4A ERH panel developed in this work comprises 1069 lines obtained after pollinating N4AT4B and N4AT4D lines with irradiated pollen to minimize the risk of bias, if any, due to the presence of either chromosome in tetrasomic condition. The pollen was irradiated at three different dosages to facilitate map construction by implementing lines with different density of breaks. Out of the 414 lines identified by initial screening, 119 lines with the lowest retention frequency were genotyped using 90K iSelect array (Wang et al., 2014). In these lines, the retention frequency of 76% was observed. It was significantly lower when compared to RH panels developed by seed irradiation. In these cases, the average retention frequency was 80–90% (Kumar et al., 2012a,b, 2015; Michalak de Jimenez et al., 2013; Mazaheri et al., 2015). The main reason for higher retention frequencies in seed panels is the fact, that seeds with higher chromosome breakage level would not produce viable plants (Kumar et al., 2014).

The final 4A ERH map comprises 2316 markers with a length of 6550.9 cR. In the ERH panel, a total of 1080 mapping bins were identified. In comparison, the most marker populated wheat genetic maps, at present, developed by the Population Sequencing (POPSEQ) approach and the Axiom Wheat SNP Genotyping Arrays mapped large amount of markers but with limited resolution. The POPSEQ map based on 90 DH lines ordered its 112,687 markers in total of 1335 mapping bins across the whole wheat genome (21 chromosomes), with 71 mapping bins for chromosome 4A (Chapman et al., 2015). The Axiom consensus map ordered 2358 SNPs in only 45 mapping bins of the 4A chromosome (Winfield et al., 2015). This suggests 15 and 24-fold higher resolution power of the 4A ERH map when compared to the POPSEQ and Axiom wheat maps, respectively.

To further validate the 4A ERH map, comparisons with consensus recombination maps of tetraploid and hexaploid wheat developed using the same 90K iSelect SNP array, were made. The consensus recombination maps were constructed using 12 RIL and one DH populations with 1928 lines for tetraploid, and eight DH mapping populations with 1332 lines for

hexaploid wheat (Wang et al., 2014; Maccaferri et al., 2015). The 4A consensus map of the hexaploid wheat (Wang et al., 2014) has a length of 652 cM (271 recombination mapping bins) and comprises of 1928 SNP markers from which only 973 mapped to the 4A ERH map. The 4A consensus map of the tetraploid wheat (Maccaferri et al., 2015) has a length of 176.5 cM (320 recombination bins) and contains 1346 SNP markers out of which only 670 mapped to the 4A ERH map. However, none of the recombination mapping populations was derived from cv. Chinese Spring and different levels of marker polymorphism could explain the discrepancy in the number of markers mapped on chromosome 4A.

Because of different mapping populations and map length calculations used in the recombination and RH mapping, a meaningful comparison of the sensitivity and resolution can be done only by comparing the number of mapped markers and number of mapping bins, respectively. The 4A RH map comprises of 1.7- and 1.2-fold more SNP markers as compared to the consensus tetraploid and hexaploid maps, respectively (Wang et al., 2014; Maccaferri et al., 2015). The higher number of mapped markers is due to the advantage of RH mapping, which does not require polymorphism of marker alleles. Another important reason contributing to the higher map resolution was the single chromosome-focused RH mapping approach. The 4A RH map contains 3.4- and 4-fold higher number of mapping bins as compared to the tetraploid and hexaploid recombination maps, respectively. This is despite the fact that our RH map was constructed using 20- and 14-fold fewer lines. Marker order in the recombination maps and the 4A RH map was overall well preserved with only a few exceptions (**Figure 3**). However, the most striking resolution difference was observed in the centromeric and peri-centromeric regions of chromosome 4A (**Figure 3**). This phenomenon was most pronounced in the hexaploid map where the peri-centromeric region in the RH map represents about half of the map, but in the recombination map it is less than 5% of the map (**Figure 3A**). On the other hand, despite the fact that higher and more even frequency of radiation induced deletions along chromosome 4A (**Figure 2**), some interstitial regions of chromosome arms (e.g., region delimited by bin 4AL2 – 0.75–0.79) showed higher resolution in the recombinant maps (**Figure 3A**).

Uneven distribution of recombination frequency along the length of the wheat chromosomes, which usually increases significantly with the growing distance from the centromere were described before (Lukaszewski and Curtis, 1993; Akhunov et al., 2003; Saintenac et al., 2009; Choulet et al., 2014a). The low recombination rate prevents high resolution marker ordering in centromeric regions. The RHs can order markers along the chromosome independently of recombination even in the centromeric regions (Kumar et al., 2014), and our results confirm these findings. The 4A RH map has an even deletion frequency (**Figure 2**) along most of the chromosome 4A, except peri-centromeric regions, where the rate dropped significantly to zero deletions, most likely at the centromere, as described above. Kumar et al. (2015) hypothesized that lower deletion rate in centromeric region may be due to presence of important genes needed for survival. Taking into the account that SNP

markers from the 90K iSelect array are largely gene-based (Wang et al., 2014), we can hypothesize that the drop of the deletion frequency in the peri-centromeric regions (**Figure 2**) may be the result of lower gene/marker density and hence detection of fewer deletion events. The low gene density at the centromeric and peri-centromeric regions of the chromosomes has been described previously (Sandhu and Gill, 2002; Qi et al., 2004; Choulet et al., 2010, 2014b). The use of high-throughput markers without sequence preference, such as DArTSeq (Cruz et al., 2013), POPSEQ (Mascher et al., 2013), or ISBP/RJM (Paux et al., 2006), could help clarify these hypotheses.

Additional, discrepancy in marker order between RH and recombination maps was observed in the very distal parts of both chromosome arms (marker translocations and inversions, **Figure 3**). This is in concordance with the higher frequency of recombination in these regions. Different haplotypes of the distal end of the 4AL chromosome end arm associated with a loss of alleles and size differentiation of this chromosome in 200 wheat accessions were recently described by Tsõmbalova et al. (2016).

To date, four RH maps have been constructed from single plant chromosomes. However, none of them matches the map constructed in this study in the density and the resolution. There are two RH maps developed for wheat chromosome 3B based on seed panels and either combination of ISBP, SSR and DArT markers on 70 RH lines (Kumar et al., 2012a), or DArT markers on 463 RH lines (Bassi et al., 2013). The first map, with the length of 1871 cR comprises 541 markers. The second map with final length of 2,852 cR was constructed using 696 RH lines and only 140 markers. For comparison, RH map of barley chromosome 3H constructed using 113 EST and RJM markers on 202 RH lines resulted in a map of 3066 cR (Mazaheri et al., 2015). The fourth map was developed for wheat chromosome 2D using 92 RH lines and 25 markers with the final length of 453 cR (Kumar et al., 2012b).

While we have developed RH map of chromosome 4A, Tiwari et al. (2016) developed RH map for the whole wheat genome. Their map was constructed using iSelect 90K SNP array and fertile plants of RH lines derived from hybridization of tetraploid *T. turgidum* cv. Altar 84 and irradiated pollen of cv. Chinese Spring. A total of 833 markers in three linkage groups (one for 4AS and two for 4AL) were identified for chromosome 4A (**Figure 4**) and average mapping resolution was estimated to 2.4 Mb per mapping bin (Tiwari et al., 2016). Out of the 833 markers, 733 mapped in our RH map and 431 in consensus genetic map of 4A developed by Wang et al. (2014). Comparison of the three maps revealed inconsistencies in marker order especially in 4AL2 linkage group (**Figure 4**). Similar to genetic maps, the Tiwari's RH map comprises small number of markers in centromeric region

(**Figure 4**). Markers from our chromosome deletion map mapped in the 4ASC – 0–0.20 bin are not present at all and only 82 markers from the bin 4ALC – 0–0.59 are present and represent the 4AL1 linkage group (**Figure 4**). This shortage limits the utility of the whole genome RH map as a tool for mapping low recombining centromeric regions, at least for chromosome 4A. Additionally, the whole genome approach allowed identification of 35.5% of 4A markers, only. These findings suggest that the whole genome RH mapping in wheat remains challenging and utilization of heterogeneous RH lines may limit reliability and efficiency of marker development and ordering. Our results suggests that the use of reliable high-throughput genotyping platforms with carefully preselected ERH panel and sufficient controls can provide reliable high density maps with significantly higher resolution especially in the centromeric regions.

## AUTHOR CONTRIBUTIONS

## FUNDING

## ACKNOWLEDGMENTS

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fpls.2016.02063/full#supplementary-material

## REFERENCES

Abouelhoda, M. I., Kurtz, S., and Ohlebusch, E. (2002). "The enhanced suffix array and its applications to genome analysis," in *Proceedings of the Second Workshop on Algorithms in Bioinformatics: Lecture Notes in Computer Science*, Vol. 2452, (Berlin: Springer-Verlag), 449–463.

Akhunov, E., Nicolet, C., and Dvorak, J. (2009). Single nucleotide polymorphism genotyping in polyploidy wheat with the Illumina GoldenGate assay. *Theor. Appl. Genet.* 119, 507–517. doi: 10.1007/s00122-009-1059-5

Akhunov, E. D., Goodyear, A. W., Geng, S., Qi, L., Echalier, B., Gill, B. S., et al. (2003). The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. *Genome Res.* 13, 753–763. doi: 10.1101/gr.808603

Akpinar, B. A., Magni, F., Yuce, M., Lucas, S. J., Šimková, H., Šafář, J., et al. (2015). The physical map of wheat chromosome 5DS revealed gene duplications and small rearrangements. *BMC Genomics* 16:453. doi: 10.1186/s12864-015-1641-y

Altschul, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2

Araki, E., Miura, H., and Sawada, S. (1999). Identification of genetic loci Affecting amylose content and agronomic traits on chromosome 4A of wheat. *Theor. Appl. Genet.* 98, 977–984. doi: 10.1007/s001220051158

Ariyadasa, R., and Stein, N. (2012). Advances in BAC-based physical mapping and map integration strategies in plants. *J. Biomed. Biotechnol.* 2012:184854. doi: 10.1155/2012/184854

Bassi, F. M., Kumar, A., Zhang, Q., Paux, E., Huttner, E., Kilian, A., et al. (2013). Radiation hybrid QTL mapping of Tdes2 involved in the first meiotic division of wheat. *Theor. Appl. Genet.* 126, 1977–1990. doi: 10.1007/s00122-013-2111-z

Belova, T., Grønvold, L., Kumar, A., Kianian, S., He, X., Lillemo, M., et al. (2014). Utilization of deletion bins to anchor and order sequences along the wheat 7B chromosome. *Theor. Appl. Genet.* 127, 2029–2040. doi: 10.1007/s00122-014-2358-z

Börner, A., Schumann, E., Fürste, A., Cöster, H., Leithold, B., Röder, S., et al. (2002). Mapping of quantitative trait loci determining agronomic important characters in hexaploid wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 105, 921–936. doi: 10.1007/s00122-002-0994-1

Chapman, J. A., Mascher, M., Buluc, A., Barry, K., Georganas, E., Session, A., et al. (2015). A whole-genome shotgun approach for assembling and anchoring the hexaploid bread wheat genome. *Genome Biol.* 16:26. doi: 10.1186/s13059-015-0582-8

Chen, S. Y., Tsubouchi, T., Rockmill, B., Sandler, J. S., Richards, D. R., Vader, G., et al. (2008). Global analysis of the meiotic crossover landscape. *Dev. Cell* 15, 401–415. doi: 10.1016/j.devcel.2008.07.006

Choulet, F., Alberti, A., Theil, S., Glover, N., Barbe, V., Daron, J., et al. (2014a). Structural ad functional partitioning of bread wheat chromosome 3B. *Science* 345:1249721. doi: 10.1126/science.1249721

Choulet, F., Caccamo, M., Wright, J., Alaux, M., Šimková, H., Šafář, J., et al. (2014b). "The wheat black jack: advances towards sequencing the 21 chromosomes of bread wheat," in *Genomics of Plant Genetic Resources*, eds R. Tuberosa, A. Graner, and E. Frison (Dordrecht: Springer), 405–438.

Choulet, F., Wicker, T., Rustenholz, C., Paux, E., Salse, J., Leroy, P., et al. (2010). Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable elements space. *Plant Cell* 22, 1686–1701. doi: 10.1105/tpc.110.074187

Cox, D. R., Burmeister, M., Price, E. R., Kim, S., and Myers, R. M. (1990). Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes. *Science* 250, 245–250. doi: 10.1126/science.2218528

Cruz, V. M. V., Kilian, A., and Dierig, D. A. (2013). Development of DArT marker platforms and genetic diversity assessment of the U.S. collection of the new oilseed crop lesquerella and related species. *PLoS ONE* 8:e64062. doi: 10.1371/journal.pone.0064062

Dilbirligi, M., Erayman, M., Sandhu, D., Sidhu, D., and Gill, K. S. (2004). Identification of wheat chromosomal regions containing expressed resistance genes. *Genetics* 166, 461–481. doi: 10.1534/genetics.166.1.461

Endo, T. R., and Gill, B. S. (1996). The deletion stocks of common wheat. *J. Hered.* 87, 295–307. doi: 10.1093/oxfordjournals.jhered.a023003

Eversole, K., Feuillet, C., Mayer, K. F. X., and Rogers, J. (2014). Slicing the wheat genome. *Science* 345, 285–287. doi: 10.1126/science.1257983

Frenkel, Z., Paux, E., Mester, D., Feuillet, C., and Korol, A. (2010). LTC: a novel algorithm to improve the efficiency of contig assembly for physical mapping in complex genomes. *BMC Bioinformatics* 11:584. doi: 10.1186/1471-2105-11-584

Hernandez, P., Martis, M., Dorado, G., Pfeifer, M., Gálvez, S., Schaaf, S., et al. (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. doi: 10.1111/j.1365-313X.2011.04808.x

IWGSC (2014). A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. *Science* 345:1251788. doi: 10.1126/science.1251788

Jakobson, I., Peusha, H., Timofejeva, L., and Järve, K. (2006). Adult plant and seedling resistance to powdery mildew in a *Triticum aestivum* x *Triticum militinae* hybrid line. *Theor. Appl. Genet.* 112, 760–766. doi: 10.1007/s00122-005-0181-2

Kobayashi, F., Wu, J., Kanamori, H., Tanaka, T., Katagiri, S., Karasawa, W., et al. (2015). A high-resolution physical map integrating an anchored chromosome with the BAC physical maps of wheat chromosome 6B. *BMC Genomics* 16:595. doi: 10.1186/s12864-015-1803-y

Korol, A. B. (2013). "Recombination," in *Encyclopedia of Biodiversity*, 2nd edn, Vol. 6, ed A. S. Levin (Waltham, MA: Academic Press), 353–369.

Kubaláková, M., Vrána, J., Číhaliková, J., Šimková, H., and Doležel, J. (2002). Flow karyotyping and chromosome sorting in bread wheat (*Triticum aestivum* L.). *Theor. Appl. Genet.* 104, 1362–1372. doi: 10.1007/s00122-002-0888-2

Kumar, A., Bassi, F. M., Michalak de Jimenez, M. K., Ghavami, F., Mazaheri, M., Simons, K., et al. (2014). "Radiation hybrids: a valuable toll for genetic, genomic and functional analysis of plant genomes," in *Genomics of Plant Genetic Resources*, Vol. 1, eds R. Tuberosa, A. Graner, and E. Frison (Dordrecht: Springer), 285–318.

Kumar, A., Bassi, F. M., Paux, E., Al-Azzam, O., Michalak de Jimenez, M., Denton, A. M., et al. (2012a). DNA repair and crossing over favour similar chromosome regions as discovered in radiation hybrid of *Triticum*. *BMC Genomics* 13:339. doi: 10.1186/1471-2164-13-339

Kumar, A., Seetan, R., Mergoum, M., Tiwari, V. K., Iqbal, M. J., Wang, Y., et al. (2015). Radiation hybrid maps of the D-genome of *Aegilops tauschii* and their application in sequence assembly od large and complex plant genomes. *BMC Genomics* 16:800. doi: 10.1186/s12864-015-2030-2

Kumar, A., Simons, K., Iqbal, M. J., Michalak de Jiménez, M., Bassi, F. M., Ghavami, F., et al. (2012b). Physical mapping resources for large plant genomes: radiation hybrids for wheat D-genome progenitor *Aegilops tauschii*. *BMC Genomics* 13:597. doi: 10.1186/1471-2164-13-597

Lukaszewski, A., and Curtis, C. A. (1993). Physical distribution of recombination in B-genome chromosome of tetraploid wheat. *Theor. Appl. Genet.* 86, 121–127. doi: 10.1007/BF00223816

Maccaferri, M., Ricci, A., Salvi, S., Milner, S. G., Noli, E., Martelli, P. L., et al. (2015). A high-density, SNP-based consensus map of tetraploid wheat as a bridge to integrate durum and bread wheat genomics and breeding. *Plant Biotechnol. J.* 13, 648–663. doi: 10.1111/pbi.12288

Marone, D., Laidò, G., Gadaleta, A., Colasuonno, P., Ficco, D. B. M., Giancaspro, A., et al. (2012). A high-density consensus map of A and B wheat genomes. *Theor. Appl. Genet.* 125, 1619–1638. doi: 10.1007/s00122-012-1939-y

Mascher, M., Muehlbauer, G. J., Rokhsar, D. S., Chapman, J., Schmutz, J., Barry, K., et al. (2013). Anchoring and ordering NGS contig assemblies by population sequencing (POPSEQ). *Plant J.* 76, 718–727. doi: 10.1111/tpj.12319

Mazaheri, M., Kianian, P. M. A., Kumar, A., Mergoum, M., Seetan, R., Soltani, A., et al. (2015). Radiation hybrid map of barley chromosome 3H. *Plant Genome* 8, 1–11. doi: 10.3835/plantgenome2015.02.0005

McCartney, C. A., Somers, D. J., Humphreys, D. G., Lukow, O., Ames, N., Noll, J., et al. (2005). Mapping quantitative trait loci controlling agronomic traits in the spring wheat cross RL4452x'AC Domain'. *Genome* 48, 870–883. doi: 10.1139/g05-055

Mester, D., Ronin, Y., Minkov, D., Nevo, E., and Korol, A. (2003). Constructing large-scale genetic maps using an evolutionary strategy algorithm. *Genetics* 165, 2269–2282.

Meyers, B. C., Scalabrin, S., and Morgane, M. (2004). Mappimg and sequencing complex genomes: let's get physical! *Nat. Rev.* 5, 578–588. doi: 10.1038/nrg1404

Michalak de Jimenez, M. K., Bassi, F. M., Ghavami, F., Simons, K., Dizon, R., Seetan, R. I., et al. (2013). A radiation hybrid map of chromosome 1D reveals synteny conservation at a wheat speciation locus. *Funct. Integr. Genomics* 13, 19–32. doi: 10.1007/s10142-013-0318-3

Miftahudin, K. R., Ma, X. F., Mahmoud, A. A., Layton, J., Rodriguez Milla, M. A., Chikmawati, T., et al. (2004). Analysis of expressed sequence tag loci on wheat chromosome group 4. *Genetics* 168, 651–663. doi: 10.1534/genetics.104.034827

Newell, W., Beck, S., Lehrach, H., and Lyall, A. (1998). Estimation of distances and map construction using radiation hybrids. *Genome Res.* 8, 493–508.

Nussbaumer, T., Martis, M. M., Roessner, S. K., Pfeifer, M., Bader, K. C., Sharma, S., et al. (2013). MIPS PlantsDB: a database framework for comparative plant genome research. *Nucleic Acids Res.* 41, D1144–D1151. doi: 10.1093/nar/gks1153

Paux, E., Legeai, F., Guilhot, N., Adam-Blomdon, A., Alaux, M., Salse, J., et al. (2008a). Physical mapping in large genomes: accelerating anchoring of BAC contigs to genetic maps through in silico analysis. *Funct. Integr. Genomics* 8, 29–32. doi: 10.1007/s10142-007-0068-1

Paux, E., Roger, D., Badaeva, E., Gay, G., Bernard, M., Sourdille, P., et al. (2006). Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. *Plant J.* 48, 463–474. doi: 10.1111/j.1365-313X.2006.02891.x

Paux, E., Sourdille, P., Salse, J., Saintenac, C., Choulet, F., Leroy, P., et al. (2008b). A physical map of the 1-gigabase bread wheat chromosome 3B. *Science* 322, 101–104. doi: 10.1126/science.1161847

Philippe, R., Paux, E., Bertin, I., Sourdille, P., Choulet, F., Laugier, C., et al. (2013). A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. *Genome Biol.* 14:R64. doi: 10.1186/gb-2013-14-6-r64

Qi, L., Echalier, B., Friebe, B., and Gill, B. S. (2003). Molecular characterization of a set of wheat deletion stocks for use in chromosome bin mapping of ESTs. *Funct. Integr. Genomics* 3, 39–55.

Qi, L. L., Echalier, B., Chao, S., Lazo, G. R., Butler, G. E., Anderson, O. D., et al. (2004). A chromosome bin map of 16,000 expressed sequence tag loci and distribution of gene among the three genomes of polyploid wheat. *Genetics* 168, 701–712. doi: 10.1534/genetics.104.034868

Ronin, Y., Mester, D., Minkov, D., and Korol, A. (2010). Building reliable genetic maps: different mapping strategies may result in different maps. *Nat. Sci.* 2, 576–589. doi: 10.4236/ns.2010.26073

Ronin, Y., Minkov, D., Mester, D., Akhunov, A., and Korol, A. (2015). "Building ultra-dense genetic maps in the presence of genotyping errors and missing data," in *Advances in Wheat Genetics: From Genome to Field*, eds Y. Ogihara, S. Takumi, and H. Handa (Tokyo: Springer), 127–133.

Šafář, J., Šimková, H., Kubaláková, M., Číhalíková, J., Suchánková, P., Bartoš, J., et al. (2010). Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet. Genome Res.* 129, 211–223. doi: 10.1159/000313072

Saintenac, C., Falque, M., Martin, O. C., Paux, E., Feuillet, C., and Sourdille, P. (2009). Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (*Triticum aestivum* L.). *Genetics* 181, 393–403. doi: 10.1534/genetics.108.097469

Sandhu, D., and Gill, K. S. (2002). Gene-containing regions of wheat and the other grass genomes. *Plant Physiol.* 128, 803–811. doi: 10.1104/pp.010745

Sears, E. R., and Sears, L. (1978). "Tetelocentric chromosomes of common wheat," in *Proceedings of the 5th International Wheat Genetics symposium*, ed. S. Ramanujams (New Delhi: Indian Agricultural Reseach Institute), 389–407.

Shorinola, O., Bird, N., Simmonds, J., Berry, S., Henriksson, T., Jack, P., et al. (2016). The wheat Phs-A1 pre-harvest sprouting resistance locus delays the rate of seed dormancy loss and maps 0.3 cM distal to the PM19 genes in UK germplasm. *J. Exp. Bot.* 67, 4169–4178. doi: 10.1093/jxb/erw194

Šimková, H., Svensson, J. T., Condamine, P., Hříbová, E., Suchánková, P., Bhat, P. R., et al. (2008). Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. *BMC Genomics* 9:294. doi: 10.1186/1471-2164-9-294

Sourdille, P., Singh, S., Cadalen, T., Brown-Guedira, G. L., Gay, G., Qi, L., et al. (2004). Microsatellite-based deletion bin system for establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). *Funct. Integr. Genomics* 4, 12–25. doi: 10.1007/s10142-004-0106-1

Tiwari, V. K., Heesacker, A., Riera-Lizarazu, O., Gunn, H., Wang, S., Wang, Y., et al. (2016). A whole-genome, radiation hybrid mapping resource of hexaploid wheat. *Plant J.* 86, 195–207. doi: 10.1111/tpj.13153

Tiwari, V. K., Riera-Lizarazu, O., Gunn, H. L., Lopez, K., Iqbal, M. J., Kianian, S. F., et al. (2012). Endosperm tolerance of paternal aneuploidy allows radiation hybrid mapping of the wheat D-genome and a measure of γ ray-induced chromosome breaks. *PLoS ONE* 7:e48815. doi: 10.1371/journal.pone.0048815

Tsömbalova, J., Karafiátová, M., Vrána, J., Kubaláková, M., Peuša, H., Jakobson, I., et al. (2016). A haplotype specific to North European wheat (*Triticum Q14 aestivum* L.). *Genet. Resour. Crop Evol.* 1–12. doi: 10.1007/s10722-016-0389-9

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B. E., et al. (2014). Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* 12, 787–796. doi: 10.1111/pbi.12183

Wang, Y., Drader, T., Tiwari, V. K., Dong, L., Kumar, A., Huo, N., et al. (2015). Development of a D genome specific marker resource for diploid and hexaploid wheat. *BMC Genomics* 16:646. doi: 10.1186/s12864-015-1852-2

Winfield, M. O., Allen, A. M., Burridge, A. J., Barker, G. L. A., Benbow, H. R., Wilkinson, P. A., et al. (2015). High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* 13, 733–742. doi: 10.1111/pbi.12485

# APPENDIX II

**The *in silico* identification and characterization of a bread wheat/*Triricum militinae* introgression line**

Abrouk M, Balcárková B, Šimková H, Komínkova E, Martis MM, Jakobson I, Timofejeva L, Rey E, Vrána J, Kilian A, Järve K, Doležel J and Valárik M

IF: 6.09

# The *in silico* identification and characterization of a bread wheat/*Triticum militinae* introgression line

Michael Abrouk[1], Barbora Balcárková[1], Hana Šimková[1], Eva Komínkova[1], Mihaela M. Martis[2,3], Irena Jakobson[4], Ljudmilla Timofejeva[4], Elodie Rey[1], Jan Vrána[1], Andrzej Kilian[5], Kadri Järve[4], Jaroslav Doležel[1] and Miroslav Valárik[1,*]

[1]*Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Olomouc, Czech Republic*
[2]*Munich Information Center for Protein Sequences/Institute of Bioinformatics and Systems Biology, Institute for Bioinformatics and Systems Biology, Helmholtz Center Munich, Neuherberg, Germany*
[3]*Division of Cell Biology, Department of Clinical and Experimental Medicine, Bioinformatics Infrastructure for Life Sciences, Linköping University, Linköping, Sweden*
[4]*Department of Gene Technology, Tallinn University of Technology, Tallinn, Estonia*
[5]*Diversity Arrays Technology Pty Ltd, Canberra, ACT, Australia*

## Summary

The capacity of the bread wheat (*Triticum aestivum*) genome to tolerate introgression from related genomes can be exploited for wheat improvement. A resistance to powdery mildew expressed by a derivative of the cross-bread wheat cv. Tähti × *T. militinae* (*Tm*) is known to be due to the incorporation of a *Tm* segment into the long arm of chromosome 4A. Here, a newly developed *in silico* method termed rearrangement identification and characterization (RICh) has been applied to characterize the introgression. A virtual gene order, assembled using the GenomeZipper approach, was obtained for the native copy of chromosome 4A; it incorporated 570 4A DArTseq markers to produce a zipper comprising 2132 loci. A comparison between the native and introgressed forms of the 4AL chromosome arm showed that the introgressed region is located at the distal part of the arm. The *Tm* segment, derived from chromosome 7G, harbours 131 homoeologs of the 357 genes present on the corresponding region of Chinese Spring 4AL. The estimated number of *Tm* genes transferred along with the disease resistance gene was 169. Characterizing the introgression's position, gene content and internal gene order should not only facilitate gene isolation, but may also be informative with respect to chromatin structure and behaviour studies.

## Introduction

Using interspecific hybridization to widen a crop's gene pool is an attractive strategy for reversing the genetic bottleneck imposed by domestication and for compensating the genetic erosion, which has resulted from intensive selection (Feuillet *et al.*, 2008). Much of the pioneering research in this area has focused on bread wheat (*Triticum aestivum*), in which over 50 related species have been exploited as donors thanks to the plasticity of the recipient's genome (Jiang *et al.*, 1993; Wulff and Moscou, 2014). Typically, introgression events have involved the transfer of a substantially sized donor chromosome segment, which, along with the target, probably bears gene(s), which impact negatively on the host's fitness (a phenomenon also called 'linkage drag') (Gill *et al.*, 2011; Qi *et al.*, 2007; Zamir, 2001). For this reason, very few introgression lines are represented in commercial cultivars (Rey *et al.*, 2015). The prime means of reducing the length of an introgressed segment is to induce recombination with its homoeologous region (Niu *et al.*, 2011). The success of this strategy is highly dependent on the conservation of gene content and order between the donor segment and its wheat equivalent.

The level of resolution with which introgression segments can be characterized has developed over the years along with advances in DNA technology. Large numbers of genetic markers have been identified in many crop species, including wheat (Bellucci *et al.*, 2015; Chapman *et al.*, 2015; Sorrells *et al.*, 2011; Wang *et al.*, 2014). In a recent example, a wheat mapping population has been genotyped with respect to >100 000 markers, but the mapping resolution achieved has only enabled the definition of around 90 mapping bins per chromosome (Chapman *et al.*, 2015). Given that the genomes of most donor species are poorly characterized, marker data at best allow only the position of an introgressed segment to be defined on the basis of the loss of wheat markers; they cannot determine either the size of the introduced segment or analyse its genetic content. The recently developed 'Introgression Browser' (Aflitos *et al.*, 2015) combines genotypic data with phylogenetic inferences to identify the origin of an introgressed segment, but to do so, a high-quality reference sequence of the host genome is needed, along with a large set of donor sequence data. The first of these requirements is being addressed by a concerted effort to acquire a reference sequence for bread wheat (www.wheatgenome.org). So far, only chromosome (3B) has been fully sequenced, and the gene content of each wheat chromosome has been obtained (Choulet *et al.*, 2014; IWGSC, 2014). The so-called GenomeZipper method (Mayer *et al.*, 2011), based on a variety of resources, has been used to predict gene order along each of the 21 bread wheat chromosomes (IWGSC, 2014).

The improved resistance to powdery mildew of an introgressive line 8.1 derived from the cross of bread wheat cv. Tähti (genome formula ABD) and tetraploid *T. militinae* (*Tm*; genome formula A[t]G) is known to be mainly due to the incorporation of a segment of *Tm* chromatin containing the resistance gene *QPm-tut-4A* into the long arm of chromosome 4A (Jakobson *et al.*, 2006, 2012). Here, a

novel *in silico*-based method, termed rearrangement identification and characterization (RICh), has been developed to identify the sequences suitable for generating markers targeting an introgression segment such as the one from *Tm*. The method integrates the GenomeZipper approach with shotgun sequences of chromosome with the introgression. The RICh method was also effective in confirming the identity of the chromosomal rearrangements, which occurred during the evolution of modern wheat.

## Results

### Chromosome sorting, sequencing and assembly

The flow karyotype derived from the DAPI-stained chromosomes of the DT4AL-TM line included a distinct peak (Figure S1) corresponding to the 4AL telosome (4AL-TM), which enabled it to be sorted to an average purity of 86.2%. The contaminants in the sorted peak comprised a mixture of fragments of various chromosomes and chromatids. DNA of all 45 000 sorted 4AL-TM telosomes was amplified by DNA multiple displacement amplification (MDA). To minimize the risk of representation bias, the products from three independent amplification reactions were pooled. From the resulting 4.5 µg DNA, a total of ~6.2 Gb of sequence was obtained, which was subsequently assembled into 279 077 contigs of individual length >200 bp, with an N50 of 2068 bp (Table 1). When the assembly was aligned with the reference genome sequences of *Brachypodium distachyon* (Vogel *et al.*, 2010), rice (IRGSP, 2005) and sorghum (Paterson *et al.*, 2009), it was apparent that the 4AL-TM telosome shares synteny with segments of *B. distachyon* chromosomes Bd1 and Bd4, rice chromosomes Os3, Os6 and Os11 and sorghum chromosomes Sb1, Sb5 and Sb10 (Figure S2).

### Origin of the introgression segment

The chromosomal origin of the *Tm* introgression segment was established by initially flow sorting the *Tm* chromosome complement. This was achieved by pretreating the chromosomes with fluorescence *in situ* hybridization in suspension (FISHIS) (Giorgi *et al.*, 2013) in which GAA microsatellites were fluorescently labelled by FITC. The resulting DAPI *vs* GAA bivariate flow karyotype succeeded in defining 13 distinct clusters (Figure 1). As the haploid chromosome number of *Tm* is 14, one of the clusters was therefore

deemed likely to harbour a mixture of two distinct chromosomes. Two of the clusters (#4 and #8) contained sequences that were amplified by the *Xgwm160* (Roder *et al.*, 1998) and *owm82* primers (these two markers are linked to the *QPm-tut-4A* gene from *Tm* introgression). The dispersed profile of cluster #4 (Figure 1) suggested that it was composed of two different A$^t$ genome chromosomes, because all G chromosomes were identified due to a higher GAA content (Badaeva *et al.*, 2010). The *owm72* marker, also linked to the *QPm-tut-4A* gene, amplified two fragments in *Tm*, one of size 205 bp and the other of size 250 bp; only the former was amplified from 4AL-TM telosome or of cluster #8. The fluorescence *in situ* hybridization (FISH) profile of the chromosomes present in cluster #8 unambiguously identified the introgressed segment as deriving from chromosome 7G.

### GenomeZipper improvement

A chromosome 4A zipper was constructed based on Chinese Spring (CS) chromosome specific survey sequences (CSSs) using 1780 specific DArTseq markers ordered in consensus genetic map (Table S1). As DArTseq marker sequences are short (69 nt) and generally nongenic, they were initially anchored to the CSS assembly; this step reduced the number of useful markers to 632 (CSS-DArTseq markers), of which 102 mapped to the short arm and 530 to the long arm. The first version of the zipper comprised a total of 2398 loci. The resulting model for 4AS was collinear with Bd1, Os3 and Sb01, as reported previously (Hernandez *et al.*, 2012). However, the one for 4AL was a mosaic of 15 orthologous blocks (based on the rice genome as the reference), derived from Os11/Bd4/Sb5, Os3/Bd1/Sb1 and Os6/Bd1/Sb10 (Figure S3a). Validation for this complex structure was sought from analysis of the subset of 2638 SNP loci (Wang *et al.*, 2014), which had been assigned a bin locations based on an analysis of a panel of established 4A deletion lines (Endo and Gill, 1996): of these, 750 mapped to five deletion bins on 4AS

**Table 1** Assembly statistics of chromosome arms 4AL-TM, 4AS-CS and 4AL-CS

|  | 4AS-CS | 4AL-CS | 4AL-TM |
|---|---|---|---|
| Sequencing read depth | 241x | 116x | 23x |
| Total contigs | 301 954 | 362 851 | 279 077 |
| Total bases (bp) | 282 335 959 | 361 971 522 | 266 737 930 |
| Assembly coverage* | 0.89x | 0.67x | 0.49x |
| Min contig length (bp) | 200 | 200 | 200 |
| Max contig length (bp) | 70 057 | 129 043 | 28 604 |
| Average contig length (bp) | 935 | 998 | 956 |
| N50 length (bp) | 2782 | 3053 | 2068 |

The data for 4AS-CS and 4AL-CS arms are taken from IWGSC (2014) and data for 4AL-TM were acquired in this study.

*The size of chromosome arms 4AS-CS (318 Mbp) and 4AL-CS (540 Mbp) were taken from Šafář *et al.* (2010). To estimate the assembly coverage of the 4AL-TM arm, the 4AL-CS size was used.



**Figure 1** The bivariate flow karyotype of *T. militinae*. Mitotic chromosomes at metaphase were stained with DAPI and GAA microsatellites were labelled with FITC. A set of 13 distinct clusters were obtained (shown boxed). Cluster #8 harbours the *Tm* chromosome (7G) which was the origin of the introgression segment present in line 8.1. Cluster #4 harbours a putative homoeolog of 7G and based on its width and shape most likely comprises a mixture of two distinct chromosomes.

and 1888 to 13 deletion bins on 4AL (Figure S3, Balcárková *et al.*, unpublished). The analysis allowed 329 SNP loci (113 on 4AS, 216 on 4AL) to be integrated into the new 4A zipper. Of the 113 4AS SNP loci, just four mapped to an inconsistent locations, demonstrating the model's accuracy; however, six (#3, #6, #8, #12, #14 and #16) of the 15 4AL blocks were inconsistent with respect to the multiple SNP loci allocations. For example, block #12—positioned in the subtelomeric region according to the zipper—included 18 SNP loci assigned to the pericentromeric region. The GenomeZipper was therefore rerun after first removing the 62 CSS-DArTseq markers associated with the misassignment of the blocks (Table S2); of the 570 CSS-DArTseq markers retained (Table S3), 79 were anchored to at least one of the *B. distachyon*, rice or sorghum scaffolds. The set of 2132 loci (745 on 4AS and 1387 on 4AL) revealed just six (rather than 16) blocks (Figure S3b, Table S2). The final structure resembles that described by Hernandez *et al.* (2012). When the model was retested with SNP markers, no further discrepancies were flagged along distal part of chromosome arm 4AL (Figure S3b).

### The *in silico* characterization of the evolutionary chromosome rearrangements on 4AL

The RICh method is based on a stringent identification and density estimation of homoeologs and is validated using a segmentation analysis. To test the approach, the CSS-based scaffolds of chromosome arms 4BS, 4BL, 4DS, 4DL, 5BL, 5DL, 7AS and 7DS (IWGSC, 2014) were compared with that of chromosome 4A, applying as the criteria a 90% level of identity and a minimum alignment length of 100 bp. The numbers of homoeologous loci obtained were, respectively, 719, 762, 636, 877, 850, 673, 602 and 627 (mean 718), but no common distinct blocks allowing for the definition of evolutionary translocations could be identified. A window size of eleven genes was then selected from the 4A zipper for the subsequent segmentation analysis. The ancestral 4AS and 4AL arms had an average density of 0.83, while the remainder of 4AL had a density of only 0.41 (Figure 2a). 4BL and 4DL sequences were homologous to 4AS, and 4BS and 4DS ones to 4AL, confirming the pericentromeric inversion event uncovered before (Devos *et al.*, 1995; Hernandez *et al.*, 2012; Ma *et al.*, 2013; Miftahudin *et al.*, 2004). Immediately following the ancestral 4AL region, the density of homoeologs associated with chromosome group 5 increased (5BL and 5DL: 147 genes, density 0.73), identifying the presence of ancestral 5AL chromatin on this arm (Figure 1b). Finally, the most distal segment of 4AL was associated with an increased density of chromosome group 7 (7AS and 7DS: 557 genes, density 0.45), confirming the ancestral translocation event involving 7BS (Figure 1C).

### Characterization of the *Tm* introgression segment

The RICh approach was then used to characterize the 4AL introgressed *Tm* segment. A direct comparison between the 4AL-TM sequence assembly and the 4A-CS zipper (95% identity, 100 bp minimum alignment length) was then made. For the long arm, the segmentation analysis revealed two distinct regions (Figure 3): the more proximal one had a high density of homologous genes (~0.84, 863 loci), so likely corresponds to a region of the 4AL telosome inherited from bread wheat (Figure 3). However, in the distal part of the arm, the homologous gene density fell to ~0.37, suggesting this as the site of the translocation event (Figure 3). Considering the same number of

genes in the homologous regions of CS DT4AL chromosome arm (4AL-CS) and 4AL-TM, the comparison between these proximal segments revealed that 16% of homologous genes (167 of 1030) in the 4AL-TM assembly were not identified and may be accounted to the sequencing and assembly imperfection. If this rate of imperfection is applied to the regions including the introgressed segment (357 CS genes *vs* 131 *Tm* homologous genes), the presence of 169 CS nonhomologous genes in the introgression segment could be estimated. The number of such genes represents the size of linkage drag (neglecting allelic variation of the homologous genes).

## Discussion

Introgression from related species provides many opportunities to broaden the genetic base of wheat, but its impact on wheat improvement has been limited by a combination of imperfect homology between donor and recipient chromatin, the loss of key recipient genes, the suppression of recombination and linkage drag effect. Thus, obtaining an accurate understanding of the size, homology, orientation and position of an introgressed segment could help to determine which introgression events are more likely to avoid incurring a performance penalty. Such knowledge would also be informative in the context of isolating a valuable gene introduced via an introgression event. Gaining this information requires saturating the target region with molecular markers. In an effort to clone of *QPm-tut-4A* gene introgressed to the wheat 4A chromosome from *T. militinae,* we developed new method for chromosome rearrangements and introgressions identification and characterization.

The presence of ancient intra- and interchromosomal rearrangements is a known complicating issue in the polyploid wheat genome, and the 4AL chromosome arm, which is one of the site of the introgression event selected in line 8.1, has a particularly complex structure. The composition of the proximal segment of the 4AL telosomes carried by DT4AL-TM and the standard CS DT4AL stock was largely identical, as expected. However, distal part of the telosomes differs in presence of *Tm* introgressive segment (Jakobson *et al.*, 2012), but no difference by synteny blocks could be detected. In hybrids between the tetraploid forms *T. turgidum* and *T. timopheevi,* Gill and Chen (1987) noted that while the latter's G genome chromosomes paired most frequently with those from the B genome, chromosome 4A was occasionally involved in pairing with chromosome 7G, presumably as a result of the presence of the 7BS segment on the *T. turgidum* 4AL arm. The likelihood is therefore that the *Tm* chromosome 7G segment, which has contributed the 4A-based powdery mildew resistance of line 8.1, was introduced via homologous recombination with the segment of 4AL carrying 7BS chromatin.

To increase resolution of the analysis, the GenomeZipper method (Mayer *et al.*, 2011), combining genetic maps, data from chromosome shotgun sequencing, and synteny information with sequenced model genomes has been adopted. The method has been useful for developing virtual gene orders in both wheat and barley chromosomes (IWGSC, 2014; Mayer *et al.*, 2011). The most crucial data set is a reliable genetic map, which serves as backbone to integrate and orient the identified syntenic blocks. Two zippers for chromosome 4A have been published to date. The first was based on relatively low coverage sequencing of the chromosome, employing as its backbone a barley linkage map formed from expressed sequence tags distributed over the
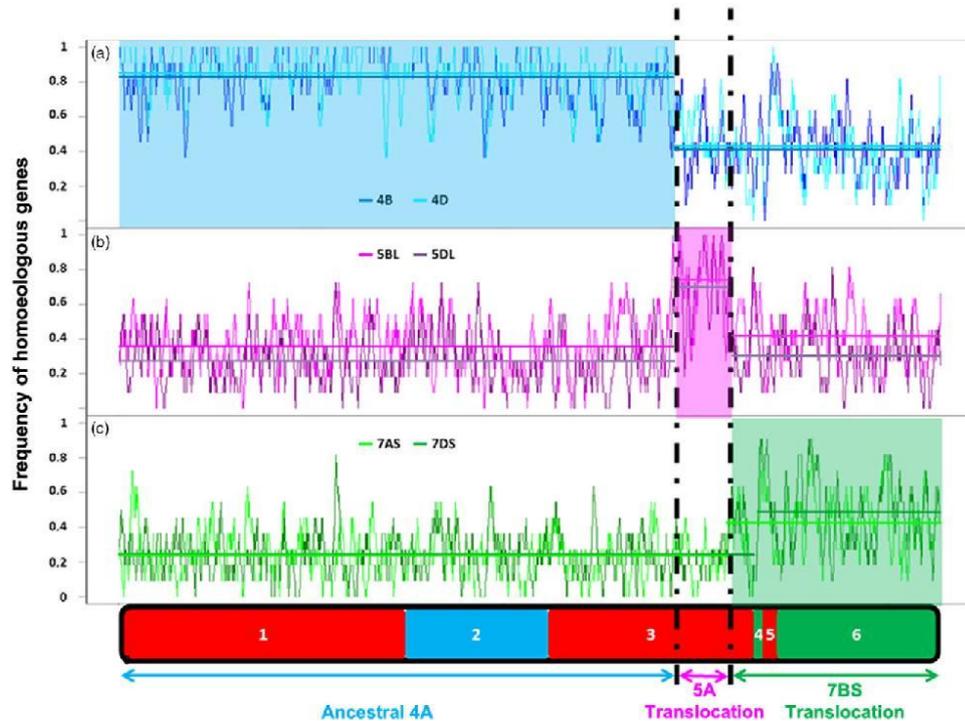
**Figure 2** Variation in homoeologous gene density along the various 4A-CS chromosome segments compared to their homoeologous chromosomes. The structure of native 4A-CS chromosome is represented at the bottom with syntenic blocks with rice genome shown in different colours (red = Os3; blue = OS11; green = Os6). (a) The 4A homoeologous gene density compared to 4B and 4D chromosomes, (b) comparison with the 5BL and 5DL chromosome arms and (c) comparison with chromosome arms 7AS and 7DS is shown as homoeologous genes frequency histogram. Homoeologous regions are characterized by a high average frequency (denoted by the horizontal lines). The lower average frequency shown by the group 7 chromosomes reflects a significantly lower sequencing coverage.

chromosome arms 4HS (117 loci), 4HL (16 loci), 5HL (36 loci) and 7HS (36 loci) (Hernandez *et al.*, 2012). The second was based on the 4A CSS and wheat SNP map and consisted of 167 markers on 4AS ordered into 56 mapping bins and 200 (92 mapping bins) markers on 4AL; these were combined with a linkage map developed from a mapping population bred from a cross between bread wheat cv. Opata and a synthetic wheat (Sorrells *et al.*, 2011). Neither of these two zippers was able to provide a sufficient level of resolution to identify the *Tm* introgression into 4AL chromosome arm. The present new zipper was based on consensus DArT map derived from crosses with CS and comprised 55% more markers and 25% more mapping bins than the latter one, which approximately doubled the number of ordered genes/loci (2132 *vs* 1004), and was informative with respect to the *Tm* introgression. When this improved zipper was used in conjunction with the RICh method, it was also possible to recognize the three evolutionary rearrangements, which have long been known to have generated the structure of the modern chromosome arm 4AL (Figure 2) (Devos *et al.*, 1995; Hernandez *et al.*, 2012; Ma *et al.*, 2013; Miftahudin *et al.*, 2004). Similarly, it was able to

identify that a lower density of homologous genes obtained at the distal end of the 4AL-TM telosome (Figure 3) is representing the region harbouring the segment introgressed from *Tm*. The *Tm* introgression overlaps with almost the entire chromosome 7BS segment now present on 4AL (Figure 3, Table S2), while the proximal region of the 4AL-CS and 4AL-TM telosomes is essentially of bread wheat origin. The number of wheat loci retained in this latter region did, however, differ by 16% in gene content (4AL-CS—1030 and 4AL-TM—863 genes). This difference may be result of lower sequencing coverage of the 4AL-TM (30x compared to 116x of the 4AL-CS (IWGSC, 2014)) and thus lower representation of the 4AL-TM sequence assembly. If we assume the similar gene density in homologous chromosomes of relative species, as reported before by Tiwari *et al.* (2015), and if the same rate of missing genes as above due to sequencing and assembly imperfections is assumed, estimated 169 CS nonhomologous genes were carried by the introgression in linkage drag. Knowledgeable selection of parental lines that have relatively high frequency of homologous genes in the region of interest (e.g. QTL for resistance in the *Tm* introgression, Figure 3) may
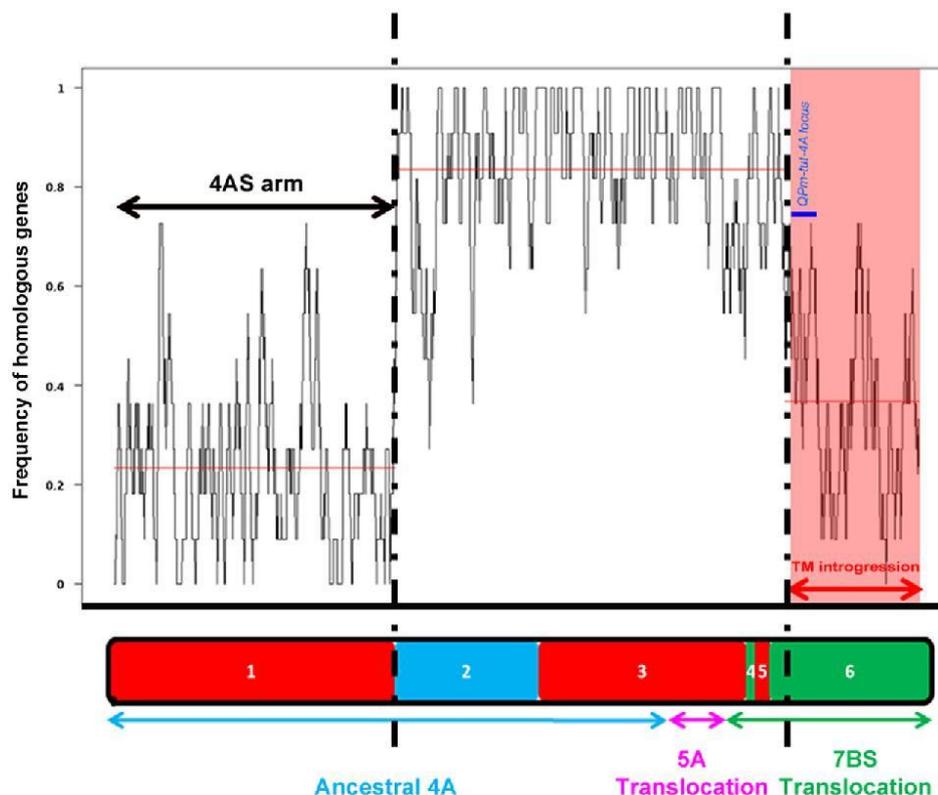
**Figure 3** Variation in homologous gene density between 4A-CS chromosome and 4AL-TM telosome. The structure of native 4A-CS chromosome is represented at the bottom with syntenic blocks with rice genome shown in different colours (red = Os3; blue = OS11; green = Os6). The homologous gene density along the 4A-CS zipper compare to 4AL-TM assembly is shown with the black line. The segment of the *Tm* introgression overlaps the 7BS translocation in 4AL (red highlight). The equivalent region on the 4AL-CS telosome harbours 357 genes, only 131 have homologous genes on the *Tm* segment. The dark blue bar represents approximate localization of the *QPm-tut-4A* locus.

increase chances of unobstructed recombination as was observed in the *QPm-tut-4A* locus (Jakobson *et al.*, 2012). So, reducing the length of the introgression segment by inducing further rounds of recombination can lessen (or even eliminate) any negative effects of linkage drag. Application of the RICh approach should prove informative regarding the order or frequency of homologous genes of any such selections. Overall, the RICh method offers a robust means of both characterizing chromosome rearrangements and of predicting the gene content of a specific chromosomal region. Recent advances in high-throughput genotyping permits the elaboration of ever higher density linkage maps (Bellucci *et al.*, 2015; Chapman *et al.*, 2015; Sorrells *et al.*, 2011; Wang *et al.*, 2014). The status of chromosome flow sorting is such that almost any wheat chromosome (Tsõmbalova *et al.*, 2016) and also chromosomes in many crops (Doležel *et al.*, 2014) can now be isolated to a reasonable purity, while the advances in NGS sequencing make RICh widely affordable. These developments should facilitate the preparation of materials needed for

applying the RICh approach, thereby offering novel opportunities for a wide range of prebreeding activities, positional cloning, chromatin hybridization and structural studies.

## Experimental procedures

### Plant materials

Grains of the bread wheat ditelosomic CS DT4AL line were provided by Dr. Bikram Gill (KSU, Manhattan, KS), those of the two nullisomic–tetrasomic lines N4AT4B and N4AT4D (Sears and Sears, 1978) by the National BioResource Centre (Kyoto, Japan), those of *Tm* ($2n = 4x = 28$, genome formula $A^tA^tGG$) accession K-46007 by the N.I. Vavilov Institute of Plant Industry (St. Petersburg, Russia). The line denoted DT4AL-TM was generated from the cross CS DT4AL × 8.1: the line carries 40 bread wheat chromosomes and a pair of 4AL telosomes with the *Tm* introgression (4AL-TM) and is resistant to powdery mildew (Jakobson *et al.*, 2012).

## Flow sorting and amplification of the 4AL telosome carried by 4AL-TM

Liquid suspensions of mitotic chromosomes were prepared from root tips of 4AL-TM seedlings as described by (Vrána *et al.*, 2000). The telosomes were separated from the rest of the genome by flow sorting, using a FACSAria II SORP flow cytometer and sorter (BD Biosciences, San Jose, CA). The level of contamination within a sorted peak was determined using FISH, based on probes detecting telomeric repeats, the Afa repeat and $(GAA)_n$, following the methods described by Kubaláková *et al.* (2003). The flow-sorted 4AL-TM telosomes were treated with proteinase, after which DNA was extracted using a Millipore Microcon YM-100 column (www.millipore.com). Chromosomal DNA was MDA amplified using the Illustra GenomiPhi V2 DNA amplification kit (GE Healthcare) as described by Šimková *et al.* (2008).

## Identifying the origin of the introgression segment on the 4AL-TM telosome

Chromosomes of *T. militinae* were flow sorted as described above. However, prior to flow cytometry, GAA microsatellites on chromosomes were labelled by FITC using FISHIS protocol (Giorgi *et al.*, 2013). Bivariate analysis (DNA content/DAPI *vs* GAA/FITC) enabled discrimination of 13 of 14 chromosomes of *T. militinae*. Individual chromosome fractions were sorted into tubes for PCR amplification and onto microscopic slides for identification of sorted chromosomes by FISH. Three markers linked to the *Tm* powdery mildew resistance gene *QPm-tut-4A* were used for the selection of the critical cluster: these were the microsatellite *Xgwm160* (Roder *et al.*, 1998) and two unpublished, one (*owm72*) amplified by the primer pair 5′-TGCTTGCTTGTA GATTGTGCA/5′-CCAGTAAGCTTTGCCGTGTG) and the other (*owm82*) by 5′-GGGAGAGACGAAAGCAGGTA/5′-CTTGCATG CACGCCAGAATA. Each 20 μL PCR contained 0.01% (w/v) o-cresol sulphonephtalein, 1.5% (w/v) sucrose, 0.2 mM dNTP, 0.6 U Taq DNA polymerase and 1 μM of each primer in 10 mM Tris-HCl/50 mM KCl/1.5 mM MgCl$_2$/0.1% (v/v) Triton X-100. The template comprised about 500 sorted chromosomes. Test reactions were seeded with either 20 ng genomic DNA extracted from CS, *Tm*, N4AT4B or N4AT4D, or with 50 pg of MDA amplified DNA from 4AL-CS and 4AL-TM telosomes. The reactions were subjected to an initial denaturation (95 °C/5 min), followed by 40 cycles of 95 °C/30 s, 55 °C/30 s and 72 °C/30 s, and completed with an elongation of 72 °C/10 min. The products were electrophoretically separated through 4% nondenaturing polyacrylamide gels and visualized by EtBr staining. The markers were mapped using a F$_2$ population bred from the cross CS × 8.1 (Jakobson *et al.*, 2012).

## Sequencing of the 4AL telosome

A CSS assembly of CS chromosome arms 4AS (4AS-CS) and 4AL (4AL-CS) were acquired from Internation Wheat Genome Sequencing Consortium (IWGSC, 2014). Two sequencing libraries of DNA amplified from the 4AL-TM telosome were constructed using a Nextera kit (Illumina, San Diego, CA) with the insert size adjusted to 500 and 1000 bp. The resulting clones were sequenced as paired-end reads by IGA (Udine, Italy) using a HiSeq 2000 device (Illumina). The 4AL-TM reads were assembled with SOAPdenovo2 software, applying a range of k-mers (54–99, with a step size of 3) to select the assembly with the highest coverage and the largest N50. Assembled scaffolds (k-mer of 69,

minimum length 200 bp) were chosen for further analysis (Table 1).

## DArTseq and SNP maps for GenomeZipper construction and validation

A DArTseq consensus map, based on four crosses involving cv. Chinese Spring as a parent has been provided by DArT PL (www.diversityarrays.com). Individual maps were created using DArT PL's OCD MAPPING program (Petroli *et al.*, 2012) to order DArTseq and array-based DArTs. DArT PL's consensus mapping software (Raman *et al.*, 2014) was applied to create a consensus map using similar strategy as described in Li *et al.* (2015). Version 3.0 of consensus map with approximately 70 000 markers was used in this study.

A SNP deletion map (Balcárková *et al.*, unpublished) was used for validation. Genomic DNAs of a set of 15 chromosome 4A deletion lines (Endo and Gill, 1996) and DNAs amplified from 4AL-CS and 4AS-CS chromosome arms as controls were genotyped at USDA-ARS (Fargo, ND) using a iSelect 90k SNP array (Wang *et al.*, 2014) on Infinium platform (Illumina). The raw genotypic data were manually analysed using GenomeStudio V2011.1 software (Illumina).

## Comparative analysis and GenomeZipper analysis

Synteny between related genomic segments was assessed using ChromoWIZ software (Nussbaumer *et al.*, 2014). The number of conserved genes present within a series of 0.5-Mbp genomic windows (window shift 0.1 Mbp) was determined. The consensus chromosome 4A linkage map used as the backbone for the GenomeZipper analysis comprised 1780 DArTseq markers (Table S1). As these sequences are mostly short (69 nt) and few identify coding sequences, they were first aligned to the set of 4A CSS contigs, preserving only those contigs that matched the entire DArTseq marker sequence at a level of at least 98% identity. The retained CSS contigs ('CSS-DArTseq markers') were used for the construction of the zipper, which was subsequently validated against the SNP deletion map (2706 SNPs). Similarly as above, only those 4A CSS contigs that aligned with SNP loci along their entire length (98% identity threshold) were retained. Ordering of the CSS-DArTseq markers was compared with that ordered by SNPs from the deletion bin map and CSS-DArTseq markers which do not follow the SNP order were eliminated, and a second version of the zipper was generated using the remaining markers (Table S3). This version was revalidated against the SNP deletion map.

## The RICh approach

To identify introgressed/translocated regions, the final 4A zipper was compared to the complete set of CSS sequences obtained from chromosome arms 4BS, 4BL, 4DS, 4DL, 5BL, 5DL, 7AS and 7DS (IWGSC, 2014). Alignments were performed using the BLAST algorithm (Altschul *et al.*, 1990). The BLAST outputs were filtered by applying the following criteria: a minimum identity of either 90% (translocation analysis) or 95% (introgression analysis) and a minimum alignment length of 100 bp. For each comparison, the density of homologous genes was evaluated using a sliding window of eleven genes (five upstream and five downstream), and a segmentation analysis was performed using the R package changepoint v1.1 (Killick and Eckley, 2014), applying the parameter segment neighbourhoods method with a BIC penalty on the mean change. The method allows a statistical detection of gene density changes along the chromosome, corresponding to an

increase or decrease in the level of synteny. For translocation events, an increase in synteny level with one group of homoeologs is required, while for an introgression, a loss of orthology is anticipated.

## Acknowledgements

## Conflict of interests

Dr. A Kilian is head of Diversity Arrays Technology Pty Ltd.

## References

Aflitos, S.A., Sanchez-Perez, G., de Ridder, D., Fransz, P., Schranz, M.E., de Jong, H. and Peters, S.A. (2015) Introgression browser: high-throughput whole-genome SNP visualization. *Plant J.* **82**, 174–182.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W. and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410.

Badaeva, E.D., Budashkina, E.B., Bilinskaya, E.N. and Pukhalskiy, V.A. (2010) Intergenomic chromosome substitutions in wheat interspecific hybrids and their use in the development of a genetic nomenclature of Triticum timopheevii chromosomes. *Russian Journal of Genetics*, **46**, 769–785.

Bellucci, A., Torp, A.M., Bruun, S., Magid, J., Andersen, S.B. and Rasmussen, S.K. (2015) Association mapping in Scandinavian winter wheat for yield, plant height, and traits important for second-generation bioethanol production. *Front. Plant Sci.* **6**, 1046.

Chapman, J.A., Mascher, M., Buluc, A., Barry, K., Georganas, E., Session, A., Strnadova, V. *et al.* (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. *et al.* (2014) Structural and functional partitioning of bread wheat chromosome 3B. *Science*, **345**, 1249721.

Devos, K.M., Dubcovsky, J., Dvořák, J., Chinoy, C.N. and Gale, M.D. (1995) Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. *Theor. Appl. Genet.* **91**, 282–288.

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

Endo, T. and Gill, B. (1996) The deletion stocks of common wheat. *J. Hered.* **87**, 295–307.

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

Gill, B.S. and Chen, P.D. (1987) Role of cytoplasm-specific introgression in the evolution of the polyploid wheats. *Proc. Natl Acad. Sci. USA*, **84**, 6800–6804.

Gill, B.S., Friebe, B.R. and White, F.F. (2011) Alien introgressions represent a rich source of genes for crop improvement. *Proc. Natl Acad. Sci. USA*, **108**, 7657–7658.

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

Hernandez, P., Martis, M., Dorado, G., Pfeifer, M., Gálvez, S., Schaaf, S., Jouve, N. *et al.* (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.

IRGSP. (2005) The map-based sequence of the rice genome. *Nature*, **436**, 793–800.

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

Jakobson, I., Peusha, H., Timofejeva, L. and Järve, K. (2006) Adult plant and seedling resistance to powdery mildew in a *Triticum aestivum* x *Triticum militinae* hybrid line. *Theor. Appl. Genet.* **112**, 760–769.

Jakobson, I., Reis, D., Tiidema, A., Peusha, H., Timofejeva, L., Valárik, M., Kladivová, M. *et al.* (2012) Fine mapping, phenotypic characterization and validation of non-race-specific resistance to powdery mildew in a wheat-*Triticum militinae* introgression line. *Theor. Appl. Genet.* **125**, 609–623.

Jiang, J., Friebe, B. and Gill, B. (1993) Recent advances in alien gene transfer in wheat. *Euphytica*, **73**, 199–212.

Killick, R. and Eckley, I.A. (2014) changepoint: an R package for changepoint analysis. *J. Stat. Softw.* **58**, 1–19.

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

Li, H., Vikram, P., Singh, R., Kilian, A., Carling, J., Song, J., Burgueno-Ferreira, J. *et al.* (2015) A high density GBS map of bread wheat and its application for dissecting complex disease resistance traits. *BMC Genom.* **16**, 216.

Ma, J., Stiller, J., Berkman, P.J., Wei, Y., Rogers, J., Feuillet, C., Doležel, J. *et al.* (2013) Sequence-based analysis of translocations and inversions in bread wheat (*Triticum aestivum* L.). *PLoS ONE*, **8**, e79329.

Mayer, K., Martis, M., Hedley, P., Simkova, H., Liu, H. and Morris, J. (2011) Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell*, **23**, 1249–1263.

Miftahudin, Ross, K., Ma, X.-F., Mahmoud, A.A., Layton, J., Milla, M.A.R., Chikmawati, T. *et al.* (2004) Analysis of expressed sequence tag loci on wheat chromosome group 4. *Genetics*, **168**, 651–663.

Niu, Z., Klindworth, D.L., Friesen, T.L., Chao, S., Jin, Y., Cai, X. and Xu, S.S. (2011) Targeted introgression of a wheat stem rust resistance gene by DNA marker-assisted chromosome engineering. *Genetics*, **187**, 1011–1021.

Nussbaumer, T., Kugler, K.G., Schweiger, W., Bader, K.C., Gundlach, H., Spannagl, M., Poursarebani, N. *et al.* (2014) chromoWIZ: a web tool to query and visualize chromosome-anchored genes from cereal and model genomes. *BMC Plant Biol.* **14**, 348.

Paterson, A.H., Bowers, J.E., Bruggmann, R., Dubchak, I., Grimwood, J., Gundlach, H., Haberer, G. *et al.* (2009) The Sorghum bicolor genome and the diversification of grasses. *Nature*, **457**, 551–556.

Petroli, C., Sansaloni, C., Carling, J., Steane, D., Vaillancourt, R. and Myburg, A. (2012) Genomic characterization of DArT markers based on high-density linkage analysis and physical mapping to the eucalyptus genome. *PLoS ONE*, **7**, e44684.

Qi, L., Friebe, B., Zhang, P. and Gill, B.S. (2007) Homoeologous recombination, chromosome engineering and crop improvement. *Chromosome Res.* **15**, 3–19.

Raman, H., Raman, R., Kilian, A., Detering, F., Carling, J. and Coombes, N. (2014) Genome-wide delineation of natural variation for pod shatter resistance in *Brassica napus*. *PLoS ONE*, **9**, e101673.

Rey, E., Molnár, I. and Doležel, J. (2015) Genomics of wild relatives and alien introgressions. In *Alien Introgression in Wheat: Cytogenetics, Molecular Biology, and Genomics* (Molnár-Láng, M., Ceoloni, C. and Doležel, J., eds), pp. 347–381. Cham: Springer International Publishing.

Roder, M.S., Korzun, V., Wendehake, K., Plaschke, J., Tixier, M.H., Leroy, P. and Ganal, M.W. (1998) A microsatellite map of wheat. *Genetics*, **149**, 2007–2023.

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

Sears, E.R. and Sears, L.M.S. (1978) The telocentric chromosomes of common wheat. In *Proc 5th Int Wheat Genet Symp* (Ramanujam, S., ed), pp. 389–407. New Delhi: Indian Soc of Genet Plant Breed.

Šimková, H., Svensson, J.T., Condamine, P., Hřibová, E., Suchánková, P., Bhat, P.R., Bartoš, J. *et al.* (2008) Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. *BMC Genom.* **9**, 294.

Sorrells, M.E., Gustafson, J.P., Somers, D., Chao, S.M., Benscher, D., Guedira-Brown, G., Huttner, E. *et al.* (2011) Reconstruction of the synthetic W7984 x Opata M85 wheat reference population. *Genome*, **54**, 875–882.

Tiwari, V.K., Wang, S., Danilova, T., Koo, D.H., Vrana, J., Kubalakova, M., Hribova, E. *et al.* (2015) Exploring the tertiary gene pool of bread wheat: sequence assembly and analysis of chromosome 5M(g) of *Aegilops geniculata. Plant J.* **84**, 733–746.

Tsömbalova, J., Karafiátová, M., Vrána, J., Kubaláková, M., Peuša, H., Jakobson, I., Järve, M. *et al.* (2016) A haplotype specific to North European wheat (*Triticum aestivum* L.). *Genet. Resour. Crop Evol.* DOI 10.1007/s10722-016-0389-9

Vogel, J.P., Garvin, D.F., Mockler, T.C., Schmutz, J., Rokhsar, D., Bevan, M.W., Barry, K. *et al.* (2010) Genome sequencing and analysis of the model grass *Brachypodium distachyon. Nature,* **463**, 763–768.

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

Wang, S., Wong, D., Forrest, K., Allen, A., Chao, S., Huang, B.E., Maccaferri, M. *et al.* (2014) Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. *Plant Biotechnol. J.* **12**, 787–796.

Wulff, B.B.H. and Moscou, M.J. (2014) Strategies for transferring resistance into wheat: from wide crosses to GM cassettes. *Front. Plant Sci.* **5**, 692. doi: 10.3389/fpls.2014.00692

Zamir, D. (2001) Improving plant breeding with exotic genetic libraries. *Nat. Rev. Genet.* **2**, 983–989.

## Supporting information

Additional Supporting Information may be found online in the supporting information tab for this article:

**Figure S1** The flow karyotype of DT4AL-TM, a bread wheat line ditelosomic for 4AL, the distal portion of which includes a segment translocated from *T. militinae.*

**Figure S2** A comparative analysis of the telosomes 4AL-CS and 4AL-TM with the *B. distachyon*, rice and sorghum genomes.

**Figure S3** Refining the robustness of the 4A zipper.

**Table S1** Consensus 4A DArTseq map, based on four independent populations, each involving CS as one parent.

**Table S2** The new 4A zipper, composed of 2132 loci, constructed using the CS-based 4A specific DArTseq map and validated by reference to SNPs mapped using a panel of deletion lines.

**Table S3** The set of CSS-DArTseq markers used to construct the new zipper.

# APPENDIX III

**The wheat *Phs-A1* pre-harvest sprouting resistance locus delays dormancy loss during seed after-ripening and maps 0.3 cM distal to the *PM19* genes in UK germplasm**

Shorinola O, Bird N, Simmonds J, Berry S, Henriksson T, Jack P, Werner P, Gergets T, Scholefield D, Balcárková B, Valárik M, Holdsworth M, Flintham J, Uauy C

IF: 5.677

RESEARCH PAPER

# The wheat *Phs-A1* pre-harvest sprouting resistance locus delays the rate of seed dormancy loss and maps 0.3 cM distal to the *PM19* genes in UK germplasm

Oluwaseyi Shorinola[1], Nicholas Bird[1,2], James Simmonds[1], Simon Berry[3], Tina Henriksson[4], Peter Jack[5], Peter Werner[2], Tanja Gerjets[6], Duncan Scholefield[6], Barbara Balcárková[7], Miroslav Valárik[7], M. J. Holdsworth[6], John Flintham[1] and Cristobal Uauy[1,*]

[1] John Innes Centre, Norwich Research Park, NR4 7UH, UK
[2] KWS UK Ltd, Hertfordshire, SG8 7RE, UK
[3] Limagrain UK Ltd, Woolpit Business Park, IP30 9UP, UK
[4] Lantmannen, SE-268 81, Svalov, Sweden
[5] RAGT Seeds, Essex, CB10 1TA, UK
[6] Division of Plant and Crop Sciences, School of Biosciences, University of Nottingham, LE12 5RD, UK
[7] Centre of Plant Structural and Functional Genomics, Olomouc, 783 71, Czech Republic

* Correspondence: cristobal.uauy@jic.ac.uk

## Abstract

The precocious germination of cereal grains before harvest, also known as pre-harvest sprouting, is an important source of yield and quality loss in cereal production. Pre-harvest sprouting is a complex grain defect and is becoming an increasing challenge due to changing climate patterns. Resistance to sprouting is multi-genic, although a significant proportion of the sprouting variation in modern wheat cultivars is controlled by a few major quantitative trait loci, including *Phs-A1* in chromosome arm 4AL. Despite its importance, little is known about the physiological basis and the gene(s) underlying this important locus. In this study, we characterized *Phs-A1* and show that it confers resistance to sprouting damage by affecting the rate of dormancy loss during dry seed after-ripening. We show *Phs-A1* to be effective even when seeds develop at low temperature (13 °C). Comparative analysis of syntenic *Phs-A1* intervals in wheat and *Brachypodium* uncovered ten orthologous genes, including the *Plasma Membrane 19* genes *(PM19-A1* and *PM19-A2)* previously proposed as the main candidates for this locus. However, high-resolution fine-mapping in two bi-parental UK mapping populations delimited *Phs-A1* to an interval 0.3 cM distal to the *PM19* genes. This study suggests the possibility that more than one causal gene underlies this major pre-harvest sprouting locus. The information and resources reported in this study will help test this hypothesis across a wider set of germplasm and will be of importance for breeding more sprouting resilient wheat varieties.

Key words: After-ripening, dormancy, PM19, pre-harvest sprouting, seed, synteny, *Triticum aestivum*.

## Introduction

Pre- and post-harvest crop losses caused by biotic or abiotic stress factors are major drawbacks to attaining global food security. In addition to their detrimental effects on crop yield, their effects on quality are equally damaging. Pre-harvest

sprouting (PHS) represents one such source of both yield and quality loss in global wheat production. PHS is characterized by the precocious germination of grains before harvest with consequent reductions in seed viability and end-use value, particularly for bread-making purposes. PHS is strongly influenced by the environment and is especially prevalent in wheat-growing regions with high rainfall during the period of grain maturation and ripening (Mares and Mrva, 2014). In addition, adverse environmental conditions like heat stress or water deficit during grain development have been generally associated with higher levels of seed germination upon grain maturation and this predisposes plants to incidences of PHS (Rodriguez *et al*, 2011). Given the current climate change projections of increased temperature and precipitation in parts of the world (Walther *et al.*, 2002; Trenberth, 2011), the incidence of PHS is expected to increase and become a greater challenge in wheat production areas.

Breeding for PHS resistance is an effective and environmentally sustainable strategy to address this problem. This can be achieved by breeding for seed dormancy, which is the most dominant component of PHS resistance. In general, plants with higher seed dormancy and responsiveness to abscisic acid (ABA) – a key hormone regulating seed dormancy – display higher sprouting resistance (Walker-Simmons, 1987; Gerjets *et al.*, 2010). On the other hand, prolonged seed dormancy is not desirable as it delays growing cycles and results in non-uniform germination upon sowing. The desired breeding objective is therefore an intermediate level of seed dormancy that confers resistance to sprouting during seed development and maturation, but declines shortly thereafter. However, breeding for PHS resistance is made difficult by the highly quantitative nature of this trait: numerous quantitative trait loci (QTL) accounting for varying degrees of resistance have been identified on all 21 wheat chromosomes (reviewed in Flintham *et al.*, 2002; Gao *et al.*, 2013*b*; Mares and Mrva, 2014), highlighting the complexity of its genetic control.

Despite this, a few QTL have been consistently identified across diverse germplasm which account for a large proportion of the sprouting variation observed in modern wheat cultivars. These loci include the *Red grain* colour genes (*R-1*) on the long arm of group 3 chromosomes, which encode a Myb transcription factor (*TaMyb10*) that regulates flavonoid biosynthesis (Himi and Noda 2005; Himi *et al.* 2011). Although not well understood, *TaMyb10* also has pleiotropic effects on seed dormancy, with red-grained wheat being generally, but not always, more dormant than white-grained wheat (Mares *et al.* 2005; Kottearachchi *et al.*, 2006; Mares and Mrva 2014). Besides the *R-1* loci, two major QTL on chromosome arms 3AS and 4AL account for a considerable proportion of the variation in PHS resistance when independent of grain colour. The 3AS QTL was recently cloned and shown to encode for Mother of Flowering Time (TaMFT: Nakamura *et al.*, 2011; Lei *et al.*, 2013; Liu *et al.*, 2013*b*), which promotes seed dormancy in wheat embryos. *TaMFT* is upregulated when seeds develop at temperatures below 13 °C, suggesting its involvement in the integration of environmental inputs into the regulation of seed dormancy (Nakamura *et al.*, 2011).

The 4AL QTL for PHS resistance was first reported by Flintham (2000) and was originally named *Phs*, although here we refer to it as *Phs-A1* in agreement with the current wheat gene nomenclature. This locus was subsequently identified in diverse germplasm (Imtiaz *et al.*, 2008; Ogbonnaya *et al.*, 2008; Torada *et al.*, 2008). In these studies the QTL effect was rarely characterized, thus limiting the understanding of its physiological basis. Also, the gene(s) underlying *Phs-A1* have recently been investigated. A number of candidate genes including those encoding for Aquaporin (Lohwasser *et al.*, 2013) and GA20-Oxidase (Tyagi and Gupta, 2012; Cabral *et al.*, 2014) have been suggested based on homology to model species. Recently a transcriptomics study of wheat seed dormancy by Barrero *et al.* (2015) identified two tandem genes, *PM19-A1* and *PM19-A2*, encoding ABA-inducible Plasma Membrane 19 proteins, as the main candidates for *Phs-A1*. Furthermore, the *PM19* genes were shown to be positive regulators of wheat seed dormancy as RNA interference (RNAi) lines with reduced expression of both genes show reduced level of seed dormancy (Barrero *et al.*, 2015).

In the present study we identified, validated and characterized the effect of the 4AL *Phs-A1* QTL. We show that this QTL exerts its effect by regulating the duration of seed dormancy loss after seed maturation. Furthermore, through multiple fine-mapping experiments in independent bi-parental mapping populations, we narrowed down the effective resistance locus to an interval less than 0.5 cM which, surprisingly, excludes the *PM19* candidate genes reported by Barrero *et al.* (2015). We discuss the conflicting nature of these results and the implication of our finding on the effort to breed for more PHS resilient wheat varieties.

## Materials and methods

### Plant materials

The identification, characterization and high-resolution fine-mapping of *Phs-A1* were done in two experimental populations made from the Option × Claire and Alchemy × Robigus crosses. These four cultivars are UK winter bread-wheat varieties with Alchemy and Option being PHS resistant while Claire and Robigus are PHS susceptible. The 4AL *Phs-A1* QTL was originally detected in the doubled haploid (DH) populations made from these crosses using the wheat × maize technique from $F_1$ plants (Laurie and Bennett, 1988). Forty-eight and 122 DH individuals were analysed in the Alchemy × Robigus and the Option × Claire population, respectively. Selected DH lines from these populations were subsequently used to develop the mapping population used in this study as detailed below.

### Alchemy × Robigus fine-mapping population

For subsequent characterization and fine mapping of *Phs-A1* in the Alchemy × Robigus population, we developed near isogenic lines (NILs) and recombinant inbred lines (RILs). To accomplish this, five SSR markers including *barc170*, *wmc420*, *wmc707*, *wmc760* and *wmc313* were used to select DH lines homozygous for Alchemy in different, but overlapping, intervals across the 4AL chromosome arm. These were independently backcrossed to the recurrent parent Robigus and advanced to the $BC_3$ generation by crossing heterozygous plants selected at each generation. Following self-pollination of selected $BC_3F_1$ lines, NILs homozygous for the Alchemy introgression found in the original DH lines were selected using the SSR markers flanking the introgressions. For the development of RILs used for high resolution fine-mapping, $BC_3F_2$ lines heterozygous for

*Phs-A1* interval (*barc170-wmc420*) were self-pollinated and BC$_3$F$_3$ lines with recombination events between the critical *Phs-A1* interval were selected. These were advanced to the BC$_3$F$_4$ generation by self-pollination to obtain homozygous RILs.

### Option × Claire fine-mapping population

We also developed F$_4$ RILs from the Option × Claire cross. This was accomplished by crossing a DH line (OC69) homozygous for Option across the QTL interval with Claire. Following self-fertilization of F$_1$ progeny, 2400 F$_2$ plants were screened and 85 F$_2$ recombinant lines with recombination events between markers *barc170, wms894 and xhbe03* were recovered. Thirty of these lines were randomly selected, self-fertilized and lines with homozygous recombinant haplotype were extracted from the F$_3$ population. In addition, lines with Claire or Option non-recombinant haplotype were also selected as controls. However, only 27 of these were initially phenotyped and advanced to the F$_4$ generation for further phenotyping.

### Growth conditions

Three germination index (GI) experiments and five sprouting experiments were conducted in the Alchemy × Robigus and the Option × Claire populations. All the GI experiments were conducted in the Alchemy × Robigus population, whereas sprouting experiments were conducted in both the Alchemy × Robigus (sprouting experiment-1 and 5) and Option × Claire populations (sprouting experiment-2, -3 and -4). In GI experiment-1 and sprouting experiment-1, -3 and -5, plants were grown in the glasshouse under long day conditions with 16h light (300 mmol) at 18 °C, 8h darkness at 15 °C and at relative humidity of 70%. In GI experiment-2 and sprouting experiment-4, plants were grown in controlled environment room (CER) under long day conditions with 16h light (250–400 mmol) at 20 °C, 8h darkness at 15 °C and at 70% relative humidity. GI experiment-3 was designed to test if *Phs-A1* is still effective when grains are developed at low temperature. In this experiment, plants were transferred 1–7 d after anthesis into a CER and maintained at constant day and night temperature of 13 °C. Plant materials for sprouting experiment-2 were grown in the field at Thriplow, UK (52.1000° N, 0.1000° E) as single rows in 1 m$^2$ plots using in a randomized complete block design with two replications per line.

### SNP and SSR genotyping

Five SSR markers including *wmc420, barc170, wmc707, wmc760* and *wmc313* were used for the development of Alchemy × Robigus NILs. For the fine-mapping of *Phs-A1*, SSR *wms894* and *xhbe03* were used to genotype Option × Claire RILs while only *xhbe03* was used for Alchemy × Robigus RILs as *wms894* is not polymorphic in this cross. The primer sequences of SSR markers were obtained from the GrainGenes database (http://wheat.pw.usda.gov/GG3, last accessed 4 February 2016), except for *wms894* which was obtained from RAGT Seed, UK. These were labelled with the FAM, VIC, NED or PET fluorescent dye (Applied Biosystems) for the multiplexing of assays. PCR were performed with the Qiagen Hotstart Master Mix (Qiagen, Cat No: 203443) and in volume of 6.25 µl containing 3.125 µl of Hotstart mix, 0.625 µl of primer mix and 2.5 µl of DNA. Thermal cycling conditions was as follow: Hotstart at 95 °C for 15min, 35 cycles of 95 °C for 1min; 50–60 °C (depending on annealing temperature of primers) for 1min, 72 °C for 1min and a final extension step of 72 °C for 10min. PCR amplicon were afterwards run on an Applied Biosystems 3730 DNA Analyzer using GeneScan 500 LIZ (Thermo Fisher Scientific; Cat. No:4322682) as size standard. Genotype data were analysed on the GeneScan® Analysis Software (Applied Biosystems). SSR markers *barc170, wmc707, wmc760* and *wmc313* gave Alchemy/Robigus band sizes of 170/180bp, 165/190bp, 110/100bp and 185320/340325bp, respectively, while *wms894* and *xhbe03 gave* Option/Claire band sizes of 160/125bp and 140/138bp, respectively.

For the development of SNP markers, sequences of wheat genes orthologous to *Brachypodium* genes in the syntenic *Phs-A1* interval were amplified and sequenced to identify SNPs between parental lines. Kompetitive Allelic Specific PCR assays (KASP: Smith and Maughan, 2015) were developed for each SNP. Assays were performed in 384 well plate format in a 5.07 µl volume containing 2.5 µl of DNA, 2.5 µl of KASP master mix (LGC, UK) and 0.07 µl of primer mix. PCR was performed on an Eppendorf Mastercycler pro 384 using the following protocol: Hotstart at 95 °C for 5min, ten touchdown cycles (95 °C for 20 s; touchdown 65 °C, −1 °C per cycle, 25 s) followed by 30–40 cycles of amplification (95 °C for 10 s; 57 °C for 1 min). No extension step is necessary as KASP amplicons are smaller than 100 bp. Plates were read using the Tecan SAFIRE Fluorescent Scanner and genotype data was viewed graphically with the KlusterCaller™ software (LGC, UK).

### Germination index (GI) assays

At four stages of grain maturation and after ripening including physiological maturity, harvest maturity (7 d after physiological maturity), as well as 14 and 28 d after harvest maturity, ears were harvested and gently threshed to obtain grains from the central portion of the ear. For plants grown under constant 13 °C in GI experiment-3, harvest maturity was reached 12 d after physiological maturity. Twenty grains were placed with the crease facing down in 90mm petri dishes containing two layers of Sartorius filter paper and were incubated in 5ml of sterile water for 7 d. After each day of incubation, germinated seeds (with ruptured seed coat) were counted and removed from the plate. The number of germinated seeds per day was used to calculate a weighted GI score using the formula described by Walker-Simmons (1987) with a slight modification: GI=(7×n1)+(6×n2)+(5×n3)+(4×n4)+(3×n5)+(2×n6)+(1×n7)/7×(N−M). Where *n1, n2, … n7* are the number of germinated grains on the first, second, and nth days until the 7th day, respectively; N is the total number of grains per plate and M is the number of mouldy grains after the 7 d of incubation. GI test were conducted on ears from six independent plants per NIL group in GI experiment-1 and three to four independent plants per line in GI experiment-2 and -3. The average GI scores are presented.

### Sprouting test

Plants were synchronized by days to flowering and maturity and ears were harvested at similar times post-anthesis and allowed to after-ripen at room temperature (20 °C) until GI differences were observed between the parental varieties. After-ripened ears (two to three per plant) were arranged standing upright on wire racks loaded on a revolving wheel inside a sprouting chamber. The ears were then misted for 5–7 d under 100% humidity. Misted ears were dried and ears from the same plant were gently threshed together to collect grains, which were examined for the symptoms of sprouting damage (breakage of the seed coat near the embryo). This was used to calculate the percentage of sprouting in each plant. The number of independent plants phenotyped per line ranged from 12 plants per NIL in sprouting experiment-1; two plants per RIL in sprouting experiment-2, 2–6 plants per RIL in sprouting experiment-3; 4–24 independent plants per RIL in sprouting experiment-4 and three plants per RIL in sprouting experiment-5. The total number of seeds across the experiments ranged from 50 to 200 seeds per genotype. The NILs and RILs were grouped according to their haplotype and the number of lines in each group are indicated in the figure legends. The average sprouting percentages of haplotype groups are presented in the main figures while the average sprouting percentages of individual RILs in experiments-3 to -5 are presented in the supplementary data.

### Statistical analyses

Statistical significance was calculated using either one-way or two-way analyses of variance (ANOVA). Tukey's Honestly Significant Difference (HSD) tests and Dunnett's tests (using parental varieties as controls) were performed for multiple comparisons between NIL, RILs and parents. Data that did not meet the ANOVA assumption

of homogeneity of variance were arcsin transformed and confirmed to meet the assumptions before being used for the ANOVA analysis. Statistical analyses were performed in Genstat (version15.2.0.8821) and Minitab (Version, 17.2.1).

## Results

### Validation of Phs-A1 in UK bi-parental populations

The 4AL QTL effect is known to segregate in UK wheat varieties (Flintham *et al.*, 2011) and has been identified in DH populations derived from crosses between Alchemy × Robigus and Option × Claire (Alchemy and Option providing the resistance allele). The QTL is flanked by markers *barc170* and *wmc491* in the Alchemy × Robigus population (Supplementary Fig. S1 at *JXB* online) and as such collocated with the major *Phs-A1* QTL identified across multiple studies (Flintham *et al.*, 2002; Imtiaz *et al.*, 2008; Torada *et al.*, 2008; Barrero *et al.*, 2015). To independently validate the effect of *Phs-A1*, we developed NILs from the Alchemy × Robigus cross through marker-assisted backcrossing (details in Methods). Five markers distributed across the 4AL chromosome arm were used for NIL development, including *barc170* and *wmc420* (1 cM proximal to the *wmc491* flanking marker) as well as *wmc707*, *wmc760* and *wmc313* which are distal to *barc170* (Fig. 1A). Seven overlapping recombination haplotypes (designated as NIL Groups 1–7) were developed (Fig. 1B) with only Group 3 NILs containing the Alchemy resistant haplotype across the complete QTL interval (*barc170–wmc420*).

We assessed the seed dormancy and PHS resistance phenotype of these NILs through a GI test on threshed seeds (GI experiment-1) and an artificial sprouting test on whole spikes (sprouting experiment-1). In the GI test, highly significant differences were observed between the Robigus and Alchemy parental controls (*P*<0.001; Fig. 1C). NILs were classified as either resistant or susceptible based on a Dunnett's test to the parental controls. NILs with higher or non-significant GI differences than Robigus were classified as susceptible, while NILs with lower or non-significant GI differences than Alchemy were classified as resistant (Supplementary Table S1). NIL Groups 1, 5 and 7 with the Robigus haplotype across the QTL interval, and NIL Groups 2 and 4 with recombinant haplotypes within the QTL interval, all showed the susceptible GI phenotype (Fig. 1C). Group 3 NILs showed significantly lower GI than Robigus (*P*<0.001) but also significantly higher GI than Alchemy (*P*<0.001). Likewise, Group 6 NILs also showed significant differences from both parents but the GI was only slightly lower than the susceptible Robigus parent.

In the sprouting test, all the NIL groups (except Group 3) were significantly different than Alchemy but not Robigus and were therefore classified as being susceptible to sprouting (sprouting experiment-1; Fig. 1D). Group 3 NILs showed comparable sprouting levels to the resistant variety Alchemy, and were significantly different from Robigus (*P*<0.001), consistent with the GI results. Taken together, the GI and sprouting results validate the resistance effect of *Phs-A1* in NILs with the Alchemy haplotype across the complete *barc170-wmc420* interval. NILs from Group 2 and 4, which have the Alchemy allele at either one or other – but not both – flanking markers, were susceptible suggesting that the *Phs-A1* resistance locus is delimited by, but not linked to, these markers.

### Phs-A1 reduces the rate of dormancy loss during dry seed after-ripening

The validation experiments suggest that *Phs-A1* confers PHS resistance by affecting seed dormancy since the effect was not only observed in whole spikes (sprouting test), but was also evident in the reduced germination potential of threshed seeds (GI test). To further understand this resistance mechanism, we examined the rate of dormancy loss in one of the resistant and susceptible NIL groups used in the previous experiment through the GI test (GI experiment-2). This was done across different stages of seed maturation and after-ripening including physiological maturity (PM, ~40% seed moisture content), harvest maturity (HM, ~20% seed moisture content) and two post-harvest time points (14 and 28 d post-harvest, DPH). The NILs used for the experiment (NIL Group 2 and 3) were
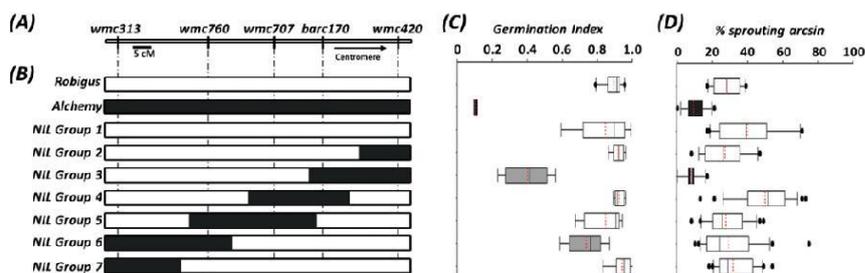


**Fig. 1.** Validation of *Phs-A1* in Alchemy × Robigus NILs. (A) Genetic map of SSR markers across the 4AL chromosome arm used to develop the NILs. (B) Graphical genotypes of Alchemy × Robigus NILs. The NILs are grouped based on their recombination haplotype across the marker intervals, with each group comprising two independent NILs. The black filled portion in the graphical genotype represents the Alchemy alleles, whereas the white sections represent the Robigus alleles. (C) Mean germination index of each NIL group in GI experiment-1. (D) Sprouting phenotype of each NIL group in sprouting experiment-1. The left and right boundaries of the boxplot indicate the 25th and 75th percentile, respectively, while the error bars (whiskers) on either side of the boxplot indicate the 10th and 90th percentiles. The solid line within the boxplot marks the median (50th percentile) while the red line within the box marks the mean.

genetically isogenic and only differed in the extent of the Alchemy introgression between *barc170* and *wmc420* (Fig. 1A). This allows for a precise characterization of this region without the confounding effect of other background loci.

At PM and HM, all lines showed similarly low GI with no difference between the contrasting alleles ($P>0.51$; Fig. 2). However, at 14 DPH there was a significant GI difference between the contrasting NILs and parents, with Robigus and the susceptible Group 2 NIL showing increased germination potential compared to Alchemy and the resistant Group 3 NIL ($P<0.001$). This difference in germination potential was maintained at 28 DPH, although the differences were reduced and less significant than at 14 DPH ($P<0.05$).

The previously cloned seed-coat-independent PHS QTL regulated by *TaMFT* is effective when seeds develop under low temperature (Nakamura *et al.*, 2011). We therefore examined if *Phs-A1* was effective at low temperature by measuring the rate of seed germination in NILs grown at 13 °C
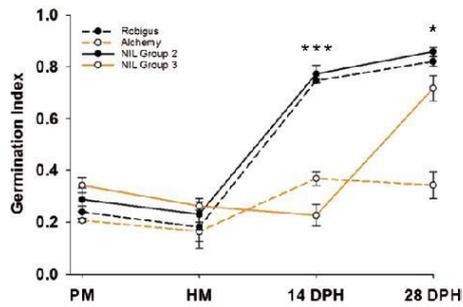


**Fig. 2.** *Phs-A1* delays the rate of seed dormancy loss during after-ripening. The germination index of seeds harvested from Robigus, Alchemy and NILs with either the recombinant haplotype (NIL Group 2) or Alchemy haplotype (NIL Group 3) between *barc170* and *wmc420* in GI experiment-2. Seeds were tested at physiological maturity (PM), harvest maturity (HM; 7 d after PM), 14 and 28 d post-harvest (DPH) and germinated at 16 °C. Error bars represent standard error of the means (SEM) of three biological replications for each time point. Significant differences between NILs at $P<0.05$ (*) and $P<0.001$ (***) are indicated.

post-anthesis (GI experiment-3). While the NILs displayed an overall higher depth of seed dormancy when plants were grown at 13 °C post anthesis, the same pattern of GI differences at the post-harvest time-points (but not at PM and HM) was observed (Supplementary Fig. S2). Together, these experiments suggest that *Phs-A1* delays the rate of seed dormancy loss during after-ripening.

### Phs-A1 *maps to a 0.5 cM interval between wms894 and xhbe03*

As previously stated, *Phs-A1* is flanked by *barc170* and *wmc420* in the bi-parental populations used in this study. Torada *et al.* (2008) mapped *Phs-A1* to a 2.6 cM interval between *barc170* and *xhbe03*. We therefore used *barc170*, *xhbe03*, and another marker – *wms894* – in the same physical bin (4AL_13-0.59–0.66), to characterize Option × Claire F$_4$ RILs (Fig. 3A, B). We selected 27 homozygous recombinants across the interval, grouped these according to their haplotypes (Fig. 3B), and assessed the sprouting phenotype using the artificial sprouting test (sprouting experiment-2). Two significantly different sets were identified in this experiment: one was made up of RIL Group 2 and Option Control RILs with between 3 and 5% sprouting, whereas the second set contained RIL groups 1, 3, 4 and the Claire Control RILs with average sprouting between 15 and 22% (Fig. 3C). RIL Group 2 and the Option Control RILs were similar to the resistant Option parent and had the Option haplotype between *wms894* and *xhbe03*. RIL Groups 1, 3, 4 and the Claire Control RILs were similar to the susceptible Claire parent and carried a homozygous Claire or recombinant haplotype across the *wms894–xhbe03* interval. This suggests that the *Phs-A1* resistance is only observed when RILs have the Option haplotype across the 0.5 cM *wms894-xhbe03* interval.

### Synteny reveals the putative gene content of the Phs-A1 *locus*

Given the small genetic interval to which *Phs-A1* mapped, we evaluated the gene content across this locus. We first
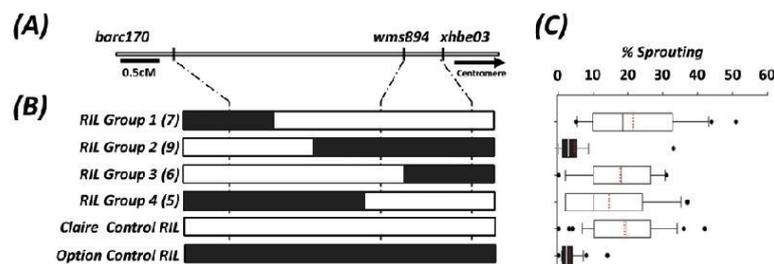


**Fig. 3.** Interval mapping of *Phs-A1* in the Option × Claire RIL population. (A) Genetic map of the SSR markers flanking *Phs-A1*. (B) Graphical genotypes of RILs and controls are presented with the Option and Claire alleles represented in black and white, respectively. The RILs are grouped according to their fixed genotype across the *Phs-A1* interval and the number of lines in each RIL group is indicated in parenthesis. (C) Sprouting phenotype of RIL groups and controls in sprouting experiment-3. The left and right boundaries of the boxplot indicate the 25th and 75th percentile, respectively, while the error bars (whiskers) on either side of the boxplot indicate the 10th and 90th percentiles. The solid line within the boxplot marks the median (50th percentile) while the red line within the box marks the mean. Boxes with the same colour are similar to each other based on pairwise comparisons.

identified genes containing the flanking markers (*xhbe03* and *wms894*) in wheat: *xhbe03* is designed from the 3′ UTR sequence of *PM19-A2* (*Traes_4AL _F99FCB25F*) while the sequence of *wms894* is located in the promoter region of an *OTU Cysteine Protease* gene (*Traes_4AL _F00707FAF*). We next examined the collinear region in *Brachypodium*: reciprocal BLASTs against the *Brachypodium* genome identified *Bradi1g00600* and *Bradi1g00720* as orthologues of *PM19-A2* and *OTU Cysteine Protease,* respectively. This defined the collinear *Phs-A1* interval in *Brachypodium* to a 75 kb region which contains 11 genes (*Bradi1g00607* to *Bradi1g00710*).

The *Brachypodium* genes were used to search the wheat chromosome arm assemblies of the hexaploid wheat cultivar, Chinese Spring (IWGSC, 2014). Orthologous contigs and gene models to *Bradi1g00600*–*Bradi1g00620* and *Bradi1g00670*–*Bradi1g00720* were identified on chromosome arm 4AL. These included *PM19-A2* and its paralogue *PM19-A1,* as well as genes encoding for Myosin J protein, Ubiquitin conjugating enzyme, Ethylene Responsive Factor-1B-Like (ERF-1B-Like), Activating Signal Co-integrator-1 (ASC1), Protein Phosphatase 1-Like (PP1-Like), a phosphate transporter, a hypothetical protein and the OTU Cysteine Protease (Fig. 4, Supplementary Table S2). No wheat orthologues were identified for *Bradi1g00630* to *Bradi1g00660*. Within the wheat IWGSC contigs, a non-collinear gene encoding for an Aminocyclopropane Carboxylate Oxidase 1-Like protein (ACC Oxidase-1; *Traes_4AL_65DF744B71*) was also identified. All these genes/contigs were mapped within or linked to the critical *wms894*–*xhbe03* interval using SNP-based KASP assays (Supplementary Table S3), except for *Traes_4AL_C56125840* which we did not map due to the lack of a genetic marker. This confirmed the collinear gene order between wheat and *Brachypodium* and suggests possible candidate genes for *Phs-A1*.

## Phs-A1 *maps distal to the PM19 genes in two UK fine-mapping populations*

Barrero *et al.* (2015) identified *PM19-A1* and *PM19-A2* as the main candidates for a seed dormancy QTL on wheat 4AL chromosome arm in a multi-parental mapping population. To determine if these genes determined the allelic variation observed in the UK populations, we further fine-mapped *Phs-A1* in the Option × Claire $F_4$ RILs with homozygous recombinant and non-recombinant haplotypes in the *Phs-A1* interval. We first defined the linkage between the gene-based KASP assays previously used to map the syntenic genes (Fig. 5A). The two *PM19* genes were completely linked and so too were the *PP1-like*, *ERF-1B-like* and *ASC1* genes. There were however recombination events between *PP1-Like/ ERF-1B-Like/ ASC1* and *OTU Cysteine Protease* and between *ACC Oxidase-1* and *PM19-A2/PM19-A1*. Given the genetic linkage between some of these markers, four SNP markers including *OTU Cysteine Protease, PP1-Like, ACC Oxidase-1* and *PM19-A2* (Supplementary Table S3) were used to define five distinct recombinant haplotypes (RIL Group 11–15; Fig. 5B).

A subset of lines from each RIL group, in addition to the parental cultivars and non-recombinant Claire and Option
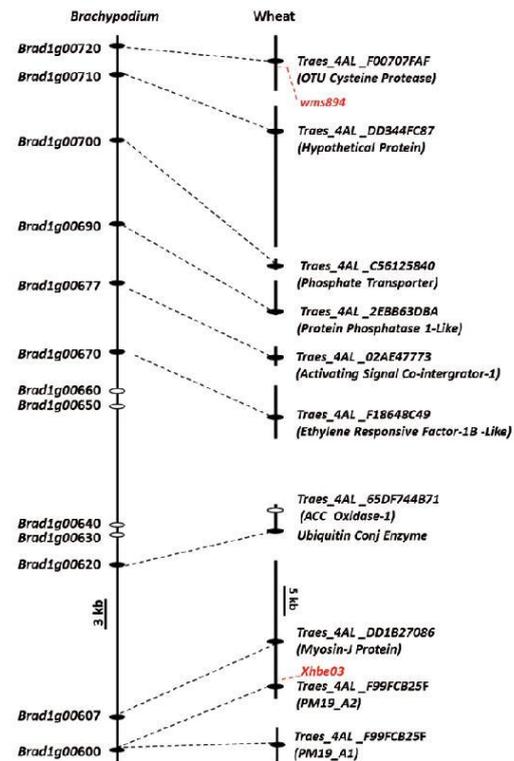


**Fig. 4.** Synteny reveals the putative gene content of the *Phs-A1* locus. Sequences of genes containing the *Phs-A1* flanking markers (*wms894* and *xhbe03*; in red) were used to obtain genes in the orthologous *Brachypodium* interval (*Brad1g00600–Brad1g00720*). Collinear genes are represented by black ovals while non-collinear genes are represented by the white ovals. Orthologous wheat contigs (black lines) and gene models are connected to their corresponding *Brachypodium* genes. All the wheat genes were genetically mapped within or linked to the *wms894–xhbe03* interval except for *Traes_4AL _C56125840*. Wheat orthologue could not be found for *Brad1g00630–Brad1g00660*.

control RILs were phenotyped using the artificial sprouting test (sprouting experiment-3; Fig. 5B). Variation in sprouting percentage was observed defining a bimodal distribution (Supplementary Fig. S3). To unequivocally assign sprouting phenotypes to these lines, the mean sprouting percentages of each RIL group (Fig. 5C) as well as the individual sprouting percentages of each RIL (Supplementary Table S4) were compared against those of Claire and Option using the Dunnett's test. This showed that the sprouting phenotype is completely associated to the *PP1-Like/ERF-1B-Like/ASC1* linkage in all the lines tested. Five independent recombination events (Group 12 and 13) map *Phs-A1* proximal to the *OTU Cysteine Protease* (*wms894*) marker. Similarly, the six lines from RIL Groups 11, 14 and 15 map *Phs-A1* distal to both the *ACC Oxidase-1* and the *PM19* genes. This was unexpected given the reported association of the *PM19* genes with sprouting
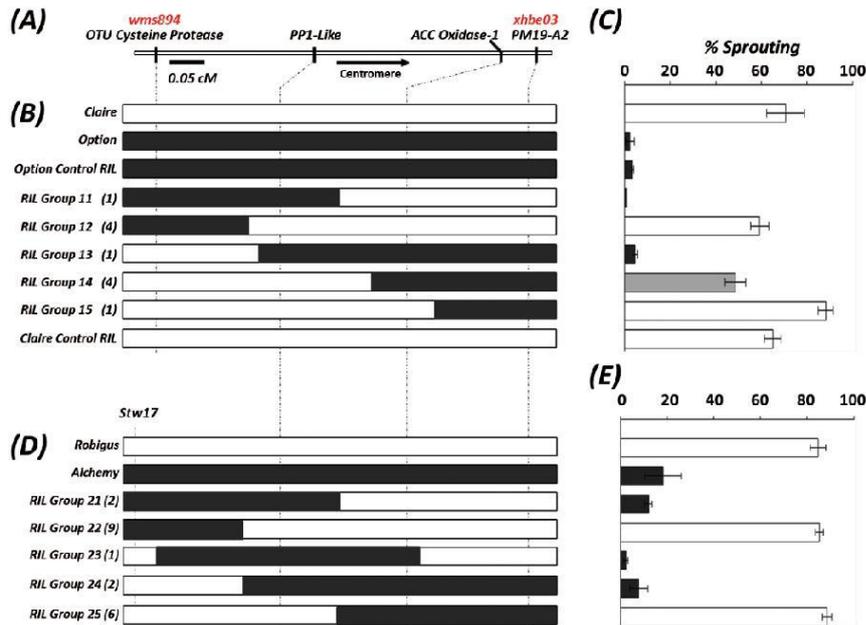
**Fig. 5.** High-resolution fine-mapping of *Phs-A1* in Option × Claire and Alchemy × Robigus RIL populations. (A) Linkage map of SNP (black) and SSR (red) markers across the *Phs-A1* interval. The graphical genotype of Option × Claire RILs (B) and Alchemy × Robigus RILs (D) are aligned against their sprouting phenotype (C and E, respectively). RILs are grouped based on their recombination haplotype across the marker interval and the number of lines in each group is indicated in parentheses. Resistant parent alleles (Option and Alchemy) are represented in black, whereas the susceptible parent alleles (Claire and Robigus) are shown in white. Marker *stw17* (2 cM distal to *wms894*) was used in the Robigus × Alchemy population as *wms894* and OTU Cysteine Protease are monomorphic. The sprouting phenotype of each RIL group is designated as susceptible (white), moderate (grey) or resistant (black) based on statistical comparison with the parental controls. Error bars represents SEM.

resistance. We confirmed this result in an independent sprouting experiment (sprouting experiment-4; Supplementary Table S5) using the critical Group 12–14 RILs. Although a higher level of sprouting was observed in this experiment, *Phs-A1* still conferred a moderate level of resistance, which was associated with the *PP1-Like/ERF-1B-Like/ASC1* linkage.

We also independently fined-mapped *Phs-A1* in an Alchemy × Robigus RIL population, which contained similar recombination haplotypes (RIL Groups 21–25; Fig. 5D) as in the Option × Claire population. However, marker *stw17* was used in place of the *OTU Cysteine Protease* marker as this was not polymorphic in the Alchemy × Robigus cross. We assessed the sprouting phenotype of these lines using the artificial sprouting test (sprouting experiment-5, Supplementary Table S6). Similar to the previous results, the mean sprouting percentages of each RIL group (Fig. 5E) confirmed the complete linkage of *Phs-A1* to *PP1-Like*, *ERF-1B-Like* and *ASC1* genes. Eleven independent recombination events in RIL Group 22 and 24 map *Phs-A1* proximal to *stw17*, whereas nine independent RILs map *Phs-A1* distal to the *ACC Oxidase-1* and *PM19* genes (RIL Groups 21, 23 and 25). This provides strong genetic evidence that in the two UK mapping populations *Phs-A1* maps distal to the *PM19* genes.

## Discussion

### Phs-A1 *confers resistance to sprouting in wheat*

In this study we characterized and fine-mapped *Phs-A1*, a major PHS resistance and seed dormancy QTL in wheat. This QTL has been previously identified across diverse germplasm and agro-ecological zones, including Australia, Canada, China, Japan and Europe (Kato *et al.*, 2001; Li *et al.*, 2004; Mares *et al.*, 2005; Torada *et al.*, 2005, 2008; Ogbonnaya *et al.*, 2007; Chen *et al.*, 2008; Cabral *et al.*, 2014; Albrecht *et al.*, 2015). Given its widespread identification and the magnitude of its effect, *Phs-A1* plays a crucial role in providing resilience to pre-harvest sprouting in wheat.

Despite its consistent identification, only Torada *et al.* (2008) had previously validated the effect of *Phs-A1* in independent isogenic material. In this study, we also validated the effect in UK germplasm using isogenic lines. In both the GI and the artificial sprouting test, we show the QTL to be effective only in lines carrying the Alchemy (resistant) haplotype across the entire *barc170–wmc420* QTL interval. However, the level of dormancy of these NILs in the GI test was only intermediate to that of Alchemy, unlike in the artificial sprouting test, where NILs showed similar sprouting resistance as Alchemy. This suggests the presence of additional loci

controlling seed dormancy in the Alchemy × Robigus cross, which are independent of the 4AL region.

### Phs-A1 *delays the rate of dormancy loss during seed after-ripening*

Understanding the mode and timing of expression of QTL is important to inform effective deployment strategies in breeding programmes and further define the underlying mechanisms. Our physiological characterization shows that *Phs-A1* confers sprouting resistance by delaying the rate of dormancy loss during after-ripening (2–4 weeks after harvest ripeness). After-ripening could be described as a period of dry seed storage during which physiological changes within seeds ultimately lead to the release from dormancy (Gao and Ayele, 2014). Some of these physiological changes include non-enzymatic oxidation of mRNA and protein by Reactive Oxygen Species (ROS) giving rise to changes in protein levels, properties and function upon imbibition (Bykova *et al.*, 2011; Gao *et al.*, 2013a; Gao and Ayele, 2014). Also, changes in the transcript level of genes involved in the biosynthesis or signalling of several hormones, including ABA, indole acetic acid, brassinosteroid, ethylene, cytokinin and salicylic acid have been reported (Liu *et al.*, 2013a; Chitnis *et al.*, 2014). When and how these physiological and transcriptional changes are initiated in dry or imbibed seeds are not well understood. However, the strong effect of *Phs-A1* on after-ripening provides an entry point to further understand the molecular pathway regulating seed after-ripening in wheat and possibly other cereals.

### Phs-A1 *maps 0.3 cM distal to the PM19 genes*

Fine-mapping of *Phs-A1* to an initial 0.5 cM interval revealed the presence of at least ten genes in the syntenic *Brachypodium* region with varied biological functions. However due to differential gene loss and duplication events between wheat and *Brachypodium* (Faris *et al.*, 2008; Glover *et al.*, 2015), it is possible that these do not represent the complete gene content in wheat. This is supported by the physical map surrounding the *PM19* genes by Barrero *et al.* (2015), which revealed the presence of additional genes besides the ones defined solely by syntenic relationships. Importantly, the high-resolution fine-mapping conclusively excludes at least six of these genes as being causal for *Phs-A1*, including the *OTU Cysteine Protease, ACC oxidase-1, Ubiquitin Conjugating Enzyme, Myosin J, PM19-A1* and *PM19-A2* genes. These results are particularly surprising for the *PM19* genes, but data from three independent experiments across two different mapping populations showed that *Phs-A1* maps 0.3 cM distal to the *PM19* locus in at least 16 recombinant lines (15 lines in Fig. 5 and one additional line in Supplementary Table S5). Based on these results, we argue that the *PM19* genes are not the main cause of the *Phs-A1* effect, at least in UK wheat varieties. In further support of this conclusion, Torada *et al.* (2008) mapped *Phs-A1* 0.5 cM distal to *xhbe03* (located in the 3′ UTR of *PM19-A2*) in two independent populations derived from Japanese and Canadian germplasm. The results of Torada *et al.* (2008) are

equivalent to those presented here and suggest that the two *PM19* loci are not the causal genes defining the 4AL sprouting resistance in the UK and possibly other germplasm pools.

The discrepancy, with the results obtained by Barrero *et al.* (2015), could be explained by a number of factors. Firstly, although the *PM19* genes affect seed dormancy in wheat, it is possible that they do not account for the natural variation in sprouting mapped to *Phs-A1* but are instead closely linked to the causal locus. The transgenic data reported by Barrero *et al.* (2015) convincingly support the role of the *PM19* genes in promoting primary seed dormancy in wheat. However, the genetic evidence presented by Barrero *et al.* (2015) is based on the phenotype of only one of five heterogeneous inbred families (F1038) developed from a four-parent MAGIC population. Given the diverse nature of this F$_7$ MAGIC population, it is possible that the phenotype of the F1038 family might have been conditioned by other loci that are independent of *Phs-A1*, but were not accounted for in the study. Indeed, there is evidence of this occurring in this population: the phenotype of another heterogeneous inbred family presented in the study (F1516) did not completely support the *PM19* genes as the causal genes.

Alternatively, it is possible that the 4AL PHS resistance originates from two independent natural variants in the closely linked *PM19* and *Phs-A1* loci. This scenario would be reminiscent of the *VRN1* locus controlling vernalization requirement and flowering in wheat (Yan *et al.*, 2003). The *VRN1* locus contains three closely linked genes: *APETALA1 (AP1)*, *AGAMOUS – LIKE GENE 1 (AGLG1)* and *Phytochrome C (PHYC)*, which all function in transitioning to flowering in Arabidopsis (Irish and Sussex, 1990; Flanagan and Ma, 1994; Balasubramanian *et al.*, 2006; Preston *et al.*, 2009; Heijmans *et al.*, 2012). Detailed genetic, expression and sequence analysis initially showed *AP1* to be the main gene underlying *VRN1*. However, recent evidence suggests that *PHYC* is also important in accelerating flowering in wheat under long day conditions (Chen *et al.*, 2014). Analogously, allelic variation at both *PM19* genes and the *Phs-A1* locus could determine resistance to sprouting, but distinct allelic variants could have been selected in Australian and UK germplasm. If true, this would account for the widespread identification of the wider *Phs-A1* locus in diverse germplasms around the world.

In support of this multiple causal gene hypothesis, we identified similar but distinct polymorphisms in the *PM19* genes in our experimental populations compared to those reported by Barrero *et al.* (2015). In the sequence of *PM19-A1* (Supplementary Fig. S4), we found only six of the seven SNPs reported between the dormant (Yitpi) and one of the non-dormant (Chara) parents used by Barrero *et al.* (2015), with only one of these leading to amino acid change between our contrasting parents (Alchemy/Option and Robigus/Clare). We also found a 12-bp deletion close to the translation stop codon in our dormant parent which resulted in the loss of four amino acid residues. This deletion was not reported in the Australian germplasm. Similar haplotype differences were observed in the *PM19-A2* sequence (Supplementary Fig. S5). A smaller deletion polymorphism (188 bp) was found in the promoter sequence of *PM19-A2* in UK germplasm compared

to the 216 bp deletion in Australian lines. Likewise, only one SNP was identified in the coding sequence of *PM19-A2* by Barrero *et al.* (2015), while we identified nine SNPs of which three were non-synonymous. This clearly suggests differences in haplotype structure in these genes between Australian and UK germplasm. This might have functional implications on the regulation and role of *PM19* in these different germplasm pools.

Based on the present study, it is difficult to determine which of these hypotheses is true. We envisage that similar high-resolution fine-mapping of the *Phs-A1* locus in different genetic backgrounds might help resolve this discrepancy. Importantly, the phenotypic characterization of the onset of PHS resistance, as well as the marker resources developed in this study, will facilitate comparative studies across other germplasm, helping clarify the association of the *PM19* gene with the *Phs-A1* natural allelic variation. If the multiple causal gene hypothesis is established, it would also be interesting to test the epistatic interaction between *PM19* and *Phs-A1* by examining lines with recombinant and non-recombinant Yipti and Alchemy/Option haplotypes between the two loci.

### *Towards the identification of* Phs-A1 *causal gene*

In three independent fine-mapping experiments (sprouting experiment-3 to -5), *Phs-A1* showed complete linkage to three genes: *ERF-1B-Like, ASC1* and *PP1-Like*. None of the genes has previously been implicated in the control of dormancy and the induction of germination. *ERF-1B-Like* encodes for an ethylene responsive transcription factor orthologous to Arabidopsis *AtERF1*, which belongs to the AP2/ERF Group-IX family involved in defence response against pathogens (Licausi *et al.*, 2013). *ERF-1B-Like* also shows good homology to *AtERF15*, another member of this AP2/ERF family, which was recently shown to be a positive regulator of ABA response (Lee *et al.*, 2015). ABA is a critical regulator of seed dormancy, with higher seed responsiveness to ABA associated with the PHS resistance. *ASC1* has not been characterized in plants, but sequence analysis of genes containing the ASC-homology domain suggest that they are involved in RNA metabolism as transcription co-activator, RNA processor and regulator of translation (Iyer *et al.*, 2006). *PP1-like* is a member of the serine threonine phosphoprotein phosphatase (PPP), which are involved in a wide range of cellular processes. Since there are likely to be other non-syntenic genes within the interval, it would be premature to suggest these as unique candidate genes for *Phs-A1*. We are currently working to establish a physical map of this interval to reveal other possible candidate genes for *Phs-A1*.

### Supplementary data

Supplementary data are available at *JXB* online.

Fig. S1. PHS resistance QTL on chromosome arm 4AL in the Alchemy × Robigus DH population.

Fig. S2. After-ripening effect of *Phs-A1* in NILs grown at 13 °C post-anthesis.

Fig. S3. Distribution of the sprouting percentage of Option × Claire $F_4$ RILs in sprouting experiment-3.

Fig. S4. Alignment of *PM19-A1* coding sequences of UK and Australian germplasm.

Fig. S5. Alignment of *PM19-A2* coding sequences of UK and Australian germplasm.

Table S1. Statistical comparison of the GI and sprouting phenotype of Alchemy × Robigus NILs.

Table S2. Information on genes found in the syntenic *Phs-A1* intervals in wheat and *Brachypodium*.

Table S3. KASP SNP assays used to fine map *Phs-A1*.

Table S4. Statistical comparison of the sprouting scores of Option × Claire RILs in sprouting experiment-3.

Table S5. Statistical comparison of the sprouting score of Option × Claire RILs in sprouting experiment-4.

Table S6. Statistical comparison of the sprouting score of Alchemy × Robigus RILs in sprouting experiment-5.

### Acknowledgements

### References

**Albrecht T, Oberforster M, Kempf H, Ramgraber L, Schacht J, Kazman E, Zechner E, Neumayer A, Hartl L, Mohler V.** 2015. Genome-wide association mapping of preharvest sprouting resistance in a diversity panel of European winter wheats. Journal of Applied Genetics 1–9.

**Balasubramanian S, Sureshkumar S, Agrawal M, Michael TP, Wessinger C, Maloof JN, Clark R, Warthmann N, Chory J, Weigel D.** 2006. The PHYTOCHROME C photoreceptor gene mediates natural variation in flowering and growth responses of *Arabidopsis thaliana*. Nature Genetics 38, 711–715.

**Barrero JM, Cavanagh C, Verbyla KL, et al.** 2015. Transcriptomic analysis of wheat near-isogenic lines identifies PM19-A1 and A2 as candidates for a major dormancy QTL. Genome Biology 16, 93.

**Bykova NV, Hoehn B, Rampitsch C, Hu J, Stebbing JA, Knox R.** 2011. Thiol redox-sensitive seed proteome in dormant and non-dormant hybrid genotypes of wheat. Phytochemistry 72, 1162–1172.

**Cabral AL, Jordan MC, McCartney CA, You FM, Humphreys DG, MacLachlan R, Pozniak CJ.** 2014. Identification of candidate genes, regions and markers for pre-harvest sprouting resistance in wheat (*Triticum aestivum* L.). BMC Plant Biology 14, 340.

**Chen A, Li C, Hu W, Lau MY, Lin H, Rockwell NC, Martin SS, Jernstedt JA, Lagarias JC, Dubcovsky J.** 2014. PHYTOCHROME C plays a major role in the acceleration of wheat flowering under long-day photoperiod. Proceedings of the National Academy of Sciences, USA 111, 10037–10044.

**Chen CX, Cai SB, Bai GH.** 2008. A major QTL controlling seed dormancy and pre-harvest sprouting resistance on chromosome 4A in a Chinese wheat landrace. Molecular Breeding 21, 351–358.

**Chitnis VR, Gao F, Yao Z, Jordan MC, Park, SE, Ayele BT.** 2014. After-ripening induced transcriptional changes of hormonal genes in wheat seeds: the cases of brassinosteroids, ethylene, cytokinin and salicylic acid. PLoS ONE 9, e87543.

**Faris JD, Zhang Z, Fellers JP, Gill BS.** 2008. Micro-colinearity between rice, *Brachypodium* and *T. monococcum* at the wheat domestication locus Q. Functional & Integrative Genomics 8, 149–164.

**Flanagan CA, Ma H.** 1994. Spatially and temporally regulated expression of the MADS-box gene AGL2 in wild-type and mutant arabidopsis flowers. Plant Molecular Biology **26,** 581–595.

**Flintham J, Adlam R, Bassoi M, Holdsworth M, Gale MD.** 2002. Mapping genes for resistance to sprouting damage in wheat. Euphytica **126,** 39–45.

**Flintham J, Holdsworth M, Jack P, Kettlewell PS, Phillips A.** 2011. An integrated approach to stabilising HFN in wheat: screens, genes and understanding. Project Report 480. HGCA, Agriculture and Horticulture Development Board.

**Flintham JE.** 2000. Different genetic components control coat-imposed and embryo-imposed dormancy in wheat. Seed Science Research **10,** 43–50.

**Gao F, Ayele BT.** 2014. Functional genomics of seed dormancy in wheat: advances and prospects. Frontiers in Plant Science **5,** 458.

**Gao F, Rampitsch C, Chitnis VR, Humphreys GD, Jordan MC, Ayele BT.** 2013a. Integrated analysis of seed proteome and mRNA oxidation reveals distinct post-transcriptional features regulating dormancy in wheat (*Triticum aestivum* L.). Plant Biotechnology Journal **11,** 921–932.

**Gao X, Hu CH, Li HZ, Yao YJ, Meng M, Dong L, Zhao WC, Chen QJ, Li XY.** 2013b. Factors affecting pre-harvest sprouting resistance in wheat (*Triticum aestivum* L.): a review. The Journal of Animal and Plant Sciences **23,** 556–565.

**Gerjets T, Scholefield D, Foulkes MJ, Lenton JR, Holdsworth MJ.** 2010. An analysis of dormancy, ABA responsiveness, after-ripening and pre-harvest sprouting in hexaploid wheat (*Triticum aestivum* L.) caryopses. Journal of Experimental Botany **61,** 597–607.

**Glover NM, Daron J, Pingault L, Vandepoele K, Paux E, Feuillet C, Choulet F.** 2015. Small-scale gene duplications played a major role in the recent evolution of wheat chromosome 3B. Genome Biology **16,** 188.

**Heijmans K, Morel P, Vandenbussche M.** 2012. MADS-box genes and floral development: the dark side. Journal of Experimental Botany **63,** 5397–5404.

**Himi E, Maekawa M, Miura H, Noda K.** 2011. Development of PCR markers for Tamyb10 related to R-1, red grain color gene in wheat. Theoretical and Applied Genetics **122,** 1561–1576.

**Himi E, Noda K.** 2005. Red grain colour gene (R) of wheat is a Myb-type transcription factor. Euphytica **143,** 239–242.

**Imtiaz M, Ogbonnaya FC, Oman J, van Ginkel M.** 2008. Characterization of quantitative trait loci controlling genetic variation for preharvest sprouting in synthetic backcross-derived wheat lines. Genetics **178,** 1725–1736.

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

**Irish VF, Sussex IM.** 1990. Function of the apetala-1 gene during Arabidopsis floral development. The Plant Cell **2,** 741–753.

**Iyer LM, Burroughs AM, Aravind L.** 2006. The ASCH superfamily: novel domains with a fold related to the PUA domain and a potential role in RNA metabolism. Bioinformatics **22,** 257–263.

**Kato K, Nakamura W, Tabiki T, Miura H, Sawada S.** 2001. Detection of loci controlling seed dormancy on group 4 chromosomes of wheat and comparative mapping with rice and barley genomes. Theoretical and Applied Genetics **102,** 980–985.

**Kottearachchi NS, Uchino N, Kato K, Miura H.** 2006. Increased grain dormancy in white-grained wheat by introgression of preharvest sprouting tolerance QTLs. Euphytica **152,** 421–428.

**Laurie DA, Bennett MD.** 1988. The production of haploid wheat plants from wheat × maize crosses. Theoretical and Applied Genetics **76,** 393–397.

**Lee SB, Lee SJ, Kim SY.** 2015. AtERF15 is a positive regulator of ABA response. Plant Cell Reports **34,** 71–81.

**Lei L, Zhu X, Wang S, Zhu M, Carver BF, Yan L.** 2013. TaMFT-A1 is associated with seed germination sensitive to temperature in winter wheat. PLoS ONE **8,** e73330.

**Li C, Ni P, Francki M, Hunter A, et al**. 2004. Genes controlling seed dormancy and pre-harvest sprouting in a rice-wheat-barley comparison. Functional & Integrated Genomics **4,** 84–93.

**Licausi F, Ohme-Takagi M, Perata P.** 2013. APETALA2/Ethylene Responsive Factor (AP2/ERF) transcription factors: mediators of stress responses and developmental programs. New Phytologist **199,** 639–649.

**Liu A, Gao F, Kanno Y, Jordan MC, Kamiya Y, Seo M, Ayele BT.** 2013a. Regulation of wheat seed dormancy by after-ripening is mediated by specific transcriptional switches that induce changes in seed hormone metabolism and signaling. PLoS ONE **8,** e56570.

**Liu S, Sehgal SK, Li J, Lin M, Trick HN, Yu J, Gill BS, Bai G.** 2013b. Cloning and characterization of a critical regulator for preharvest sprouting in wheat. Genetics **195,** 263–273.

**Lohwasser U, Rehman Arif MA, Börner A.** 2013. Discovery of loci determining pre-harvest sprouting and dormancy in wheat and barley applying segregation and association mapping. Biologia Plantarum **57,** 663–674.

**Mares D, Mrva K.** 2014. Wheat grain preharvest sprouting and late maturity alpha-amylase. Planta **240,** 1167–1178.

**Mares D, Mrva K, Cheong J, Williams K, Watson B, Storlie E, Sutherland M, Zou Y.** 2005. A QTL located on chromosome 4A associated with dormancy in white- and red-grained wheats of diverse origin. Theoretical and Applied Genetics **111,** 1357–1364.

**Nakamura S, Abe F, Kawahigashi H, et al**. 2011. A wheat homolog of MOTHER OF FT AND TFL1 acts in the regulation of germination. The Plant Cell **23,** 3215–3229.

**Ogbonnaya FC, Imtiaz M, DePauw RM.** 2007. Haplotype diversity of preharvest sprouting QTLs in wheat. Genome **50,** 107–118.

**Ogbonnaya FC, Imtiaz M, Ye G, Hearnden PR, Hernandez E, Eastwood RF, van Ginkel M, Shorter SC, Winchester JM.** 2008. Genetic and QTL analyses of seed dormancy and preharvest sprouting resistance in the wheat germplasm CN10955. Theoretical and Applied Genetics **116,** 891–902.

**Rodriguez MV, Toorop PE, Benech-Arnold RL.** 2011 Challenges facing seed banks in relations to seed quality. In: Kermode AR, ed. Seed dormancy: methods and protocols, method in molecular biology, Vol **773**. New York: Humana Press, 17–40.

**Preston JC, Christensen A, Malcomber ST, Kellogg EA.** 2009. MADS-box gene expression and implications for developmental origins of the grass spikelet. American Journal of Botany **96,** 1419–1429.

**Smith SM, Maughan PJ.** 2015. SNP genotyping using KASPar assays. Methods in Molecular Biology **1245,** 243–256.

**Torada A, Ikeguchi S, Koike M.** 2005. Mapping and validation of PCR-based markers associated with a major QTL for seed dormancy in wheat. Euphytica **143,** 251–255.

**Torada A, Koike M, Ikeguchi S, Tsutsui I.** 2008. Mapping of a major locus controlling seed dormancy using backcrossed progenies in wheat (*Triticum aestivum* L.). Genome **51,** 426–432.

**Trenberth KE.** 2011. Changes in precipitation with climate change. Climate Research **47,** 123–138.

**Tyagi S, Gupta PK.** 2012. Meta-analysis of QTLs involved in pre-harvest sprouting tolerance and dormancy in bread wheat. Triticeae Genomics and Genetic **3,** 9–24.

**Walker-Simmons M.** 1987. ABA levels and sensitivity in developing wheat embryos of sprouting resistant and susceptible cultivars. Plant Physiology **84,** 61–66.

**Walther GR, Post E, Convey P, Menzel A, Parmesan C, Beebee TJC, Fromentin JM, Hoegh-Guldberg O, Bairlein F.** 2002. Ecological responses to recent climate change. Nature **416,** 389–395.

**Yan L, Loukoianov A, Tranquilli G, Helguera M, Fahima T, Dubcovsky J.** 2003. Positional cloning of the wheat vernalization gene VRN1. Proceedings of the National Academy of Sciences, USA **100,** 6263–6268.

# APPENDIX IV

**Molecular mapping of stripe rust resistance gene *Yr51* in chromosome 4AL of wheat**

Randhawa M, Bansal U, Valárik M, <u>Klocová B</u>, Doležel J, Bariana H

ORIGINAL PAPER

# Molecular mapping of stripe rust resistance gene *Yr51* in chromosome 4AL of wheat

**Mandeep Randhawa · Urmil Bansal · Miroslav Valárik ·
Barbora Klocová · Jaroslav Doležel · Harbans Bariana**

## Abstract

***Key message*** **This manuscript describes the chromo-
somal location of a new source of stripe rust resistance
in wheat. DNA markers closely linked with the resist-
ance locus were identified and validated.**

*Abstract*   A wheat landrace, AUS27858, from the Watkins
collection showed high levels of resistance against Austral-
ian pathotypes of *Puccinia striiformis* f. sp. *tritici*. It was
reported to carry two genes for stripe rust resistance, ten-
tatively named *YrAW1* and *YrAW2*. One hundred seeds of
an F3 line (HSB#5515; *YrAW1yrAW1*) that showed mono-
genic segregation for stripe rust response were sown and
harvested individually to generate monogenically segregat-
ing population (MSP) #5515. Stripe rust response variation
in MSP#5515 conformed to segregation at a single locus.
Bulked segregant analysis using high-throughput DArT
markers placed *YrAW1* in chromosome 4AL. MSP#5515
was advanced to F6 and phenotyped for detailed mapping.
Novel wheat genomic resources including chromosome-
specific sequence and genome zipper were employed to
develop markers specific for the long arm of chromosome
4A. These markers were used for further saturation of the
*YrAW1* carrying region. *YrAW1* was delimited by 3.7 cM

M. Randhawa · U. Bansal · H. Bariana (✉)
Department of Plant and Food Sciences, Faculty of Agriculture
and Environment, The University of Sydney PBI-Cobbitty,
PMB 4011, Narellan, NSW 2567, Australia
e-mail: harbans.bariana@sydney.edu.au

M. Valárik · B. Klocová · J. Doležel
Centre of the Region Haná for Biotechnological and Agricultural
Research, Institute of Experimental Botany AS CR, Slechtitelu
31, 78371 Olomouc, Czech Republic

between markers *owm45F3R3* and *sun104*. Since there
was no other stripe rust resistance gene located in chromo-
some 4AL, *YrAW1* was formally named *Yr51*. Reference
stock for *Yr51* was lodged at the Australian Winter Cereal
Collection, Tamworth, Australia and it was accessioned as
AUS91456. Marker *sun104* was genotyped on a set of Aus-
tralian and Indian wheat cultivars and was shown to lack
the resistance-linked *sun104*-225 bp allele. Marker *sun104*
is currently being used for marker-assisted backcrossing of
*Yr51* in Australian and Indian wheat backgrounds.

## Introduction

Global wheat production is affected significantly by
stripe rust, caused by *Puccinia striiformis* f. sp. *tritici*
(Pst). The historic breakdown of stripe rust resistance
gene combination *Yr9* and *Yr27* in the 'Veery' derivatives
alarmed wheat-growing nations. Even though more than
50 stripe rust resistance genes have been identified in
wheat (McIntosh et al. 2011), virulent races of the patho-
gen continue to emerge rapidly to overcome resistance
genes. Stripe rust resistance can be classified as all stage
resistance (ASR) or adult plant resistance (APR) on the
basis of their expression at different growth stages. Vari-
ous terms have been used to describe these two types of
resistance (Bariana 2003).

Deployment of ASR genes singly does not often pro-
vide durable resistance due to the emergence of viru-
lence in pathogen populations. Pyramiding of two or
more genes in a single genotype can be difficult using
conventional selection system based on bioassays, espe-
cially in the event of resistance genes expressing simi-
lar infection types and absence of epistatic interactions.
Recent developments in molecular biology have provided

phenotype neutral selection technology based on marker-trait associations.

Identification of markers closely linked with disease resistance genes has progressed in the last decade through the development of high-throughput and cost-effective genotyping facilities. One of the first high-throughput platforms in wheat, diversity arrays technology (DArT), exploits independent chip hybridization of genome representation for diversity assessment of tested genomes and can test hundreds to thousands of genomic loci in parallel (Jaccoud et al. 2001; Akbari et al. 2006). This approach can be more efficient using high-throughput next-generation sequencing (NGS) platforms for genome representations sequencing referred to as genotyping-by-sequencing (GBS) and can identify several hundred thousand genome tags (Poland et al. 2012). Another approach includes the use of advances in wheat genome sequencing and NGS technologies to develop SNP chips for wheat with 9,000 sequences of wheat transcriptome and with 92,000 markers from the wheat genome sequence (E. Akhunov personal communication, http://wheat.pw.usda.gov/ggpages/9K_assay_available.html). All these technologies individually or in combination can be used to fine map the gene of interest.

Wheat landraces are valuable sources of genetic diversity for resistance to biotic and abiotic stresses. A common wheat landrace, AUS27858, was observed to be resistant against a range of Australian Pst pathotypes both under the greenhouse and field conditions. It was demonstrated to carry two genes for seedling resistance based on analysis of AUS27858/Westonia F3 population (Bariana and Bansal unpublished results). F3 lines segregating at a single locus were identified, and monogenically segregating populations (MSPs) were developed. Stripe rust resistance genes were temporarily named *YrAW1* and *YrAW2*. This investigation was planned to determine chromosomal location of *YrAW1*.

## Materials and methods

### Host materials

One hundred seeds from the F3 family HSB#5515 (*YrAW-1yrAW1*) were grown and harvested individually to develop monogenically segregating population MSP#5515. A recombinant inbred line (RIL) F6 population (89 lines) was subsequently developed from MSP#5515.

### Pathogen material

Pst pathotype, 134 E16A+Yr17+Yr27+ (culture number 617), was used for testing MSP#5515 and F6 RIL population. Two resistant and two susceptible RILs were also tested against six Australian Pst pathotypes 134 E16A+ (572), 134 E16A+Yr17+ (599), 134 E16A+Yr17+Yr27+ (617), 110 E143A+ (444), 108 E141A+ (420), and 104 E137+ (414). Avirulence/virulence formulae of Pst pathotypes used are presented in Table 1.

### Greenhouse screening

Twenty seeds of each F3 line were sown in 9-cm pots filled with a mixture of pine bark and river sand in the ratio of 2:1. In the case of RILs, six seeds of each line and four lines per pot were sown. Parents AUS27858 and Westonia were included as controls. Ten grams of water-soluble fertilizer Aquasol® was dissolved in 10 l of tap water and applied to 100 pots. A single application of nitrogenous fertilizer urea was applied at the same rate as Aquasol® to 7-day-old seedlings.

Twelve-day-old seedlings (two leaf stage) were inoculated by atomising Pst pathotype 134 E16A+Yr17+Yr27+ urediniospores suspended in light mineral oil (Isopar L) using a hydrocarbon propellant pressure pack. Inoculated

**Table 1** Virulence/avirulence formulae of Pst pathotypes used

| Pst pathotype | Culture no. | Virulnce/avirulence formulae |
|---|---|---|
| 104 E137A+ | 414 | *Yr2, Yr3, Yr4, Yr34/Yr1, Yr5, Yr6, Yr7, Yr8, Yr9, Yr10, Yr15, Yr17, Yr24, Yr27, Yr32, Yr33, Yr35, Yr36, Yr37, Yr47, YrA, YrSp* |
| 108 E141A+ | 420 | *Yr2, Yr3, Yr4, Yr6, YrSD, YrSu, YrND, YrA, Yr34/Yr1, Yr5, Yr7, Yr8, Yr9, Yr10, Yr15, Yr17, Yr24, Yr27, Yr32, Yr33, Yr35, Yr36, Yr37, Yr47, YrA, YrSp* |
| 110 E143A+ | 444 | *Yr2, Yr3, Yr4, Yr6, Yr7, YrSD, YrSu, YrND, YrA, Yr34/Yr1, Yr5, Yr8, Yr9, Yr10, Yr15, Yr17, Yr24, Yr27, Yr32, Yr33, Yr35, Yr36, Yr37, Yr47, YrA, YrSp* |
| 134 E16A+ | 572 | *Yr2, Yr6, Yr7, Yr8, Yr9, YrA/Yr1, Yr3, Yr4, Yr5, Yr10, Yr15, Yr17, Yr24, Yr27, Yr32, Yr33, Yr34, Yr35, Yr36, Yr37, Yr47, YrSp* |
| 134 E16A+Yr17+ | 599 | *Yr2, Yr6, Yr7, Yr8, Yr9, Yr17, YrA/Yr1, Yr3, Yr4, Yr5, Yr10, Yr15, Yr24, Yr27, Yr32, Yr33, Yr34, Yr35, Yr36, Yr37, Yr47, YrSp* |
| 134 E16A+Yr17+Yr27+ | 617 | *Yr2, Yr6, Yr7, Yr8, Yr9, Yr17, Yr27, YrA/Yr1, Yr3, Yr4, Yr5, Yr10, Yr15, Yr24, Yr27, Yr32, Yr33, Yr34, Yr47, YrSp* |

seedlings were incubated at 9–12 °C for 24 h on trolleys covered with polythene hoods to provide 100 % humidity in a temperature controlled cool room. Inoculated seedlings were then moved to a microclimate growth room maintained at 17 ± 2 °C. Seedling responses were scored on a 0–4 scale as described in Bariana and McIntosh (1993).

## Molecular mapping

### DNA isolation and quantification

Genomic DNA was isolated from seedlings of MSP#5515, MSP#5515-derived F6 RIL population and parents AUS27858 and Westonia following the procedure described on the Diversity Arrays Technology (DArT) Pty. Ltd. website (http://www.diversityarrays.com). DNA was quantified using a Nanodrop ND-1000 spectrophotometer (Nanodrop Technologies). DNA dilutions with final concentration of 50 ng/μl were prepared.

### Bulked segregant analysis

Bulked segregant analysis (BSA) was performed to establish the genomic location of *YrAW1* in MSP#5515. Equal amounts of DNA from 20 homozygous resistant and 20 homozygous susceptible lines were bulked together to constitute resistant and susceptible bulks, respectively. High-density DArT array Wheat *Pst*I (*Taq*I) 3 (http://www.diversityarrays.com) was used for BSA.

### Saturation of chromosome 4AL map

Fourteen simple sequence repeat (SSR) markers (*gpw356*, *gpw1142*, *gpw2139*, *gpw3030*, *gpw4153*, *gpw5095*, *gpw7051*, *barc52*, *barc78*, *barc153*, *barc1172*, *gwm160*, *gwm350* and *cfd31*) mapped previously in chromosome 4AL (Somers et al. 2004; Sourdille et al. 2004) were used to saturate the *YrAW1* carrying genomic region. Primer sequences of SSR markers were obtained from the GrainGenes 2.0 database (http://wheat.pw.usda.gov).

A set of 24 expressed sequenced tags (ESTs) were selected from the 4AL4-0.80-1.00 deletion bin (http://wheat.pw.usda.gov/cgi-bin/weSTSql/map_locus.cgi). ESTs amplifying 4AL-specific alleles were selected by comparing amplification profile images of each EST. Forty-four EST-based sequence tagged site (eSTS) markers were designed from selected ESTs using Primer3 software (http://frodo.wi.mit.edu/) and tested on resistant and susceptible bulks together with parents. Resistance-linked markers were subsequently genotyped on the entire RIL population to generate linkage map. In addition, 136 gene-based markers (Xue et al. 2008; Jakobson et al. 2012)

mapped in chromosome 4AL were also tested on bulks and parents. Polymorphic markers were genotyped on RIL population.

To further saturate the region using more targeted approach, we used 454 survey sequence of chromosome 4A and virtual ordering of identified coding sequences using synteny with barley EST map and genomic sequences of rice, *Brachypodium* and *Sorghum*—the 4A genome zipper (Hernandez et al. 2012). Additionally, during the construction of physical map of the *QPm.tut-4A* gene region, the *gwm160* locus was anchored to the zipper (Jakobson et al. 2012 and unpublished data) using sequences of *psr160*, *psr119*, and *cdo454* markers (GrainGenes 2.0, http://wheat.pw.usda.gov) which are flanking the region (Paillard et al. 2003). For example, on collinear rice chromosome 6 (R6), the region encompasses 0.74 Mb and on the 4AL zipper it contains 89 genes. Twenty-eight of these genes were selected for marker development to cover the *Yr51* region. Homologous wheat 4AL sequence scaffolds from the 4A survey sequence were selected using genes from the zipper syntenic region. The selected scaffolds were annotated and used to develop a set of primers using exon–exon and exon-intergenic sequence approach to enhance chromosome specificity of resultant markers (unpublished). Primers were designed using Primer3 software (http://frodo.wi.mit.edu/) and the markers were designated as "sun" (Sydney University) and "owm" (Olomouc Wheat Marker).

### PCR amplification

For PCR amplification of SSR and sequence tagged site STS markers, assays were performed in 10-μl reaction mixture containing 0.2-mM dNTPs, 1× Immolase PCR buffer (Bioline), 0.2 mM each of forward and reverse primer, 50 ng of genomic DNA, and 0.2 U of Immolase DNA polymerase (Bioline). Following an initial denaturing step of 95 °C for 10 min, PCR amplifications were performed for 40 cycles with the touchdown profile: 30 s at 92 °C, 30 s at 65 °C, and 30 s at 72 °C. Following the first cycle, the annealing temperature was reduced by 1 °C per cycle for the next five cycles. A final extension step at 72 °C for 7 min was performed.

The amplified PCR products were resolved in 2.5 % agarose (Amresco) gel stained with GelRed™ (Biotium) and scanned under UV gel documentation system (UVP-GelDoc-It). GeneRuler™ 1-Kb ladder (Fermentas) was used to determine allele sizes. Markers that did not show polymorphism on agarose gels were resolved in 8 % denaturing gel [19:1 (acrylamide: bis-acrylamide solution), 1× TBE, 8 M Urea], stained with 1× solution of SYBR®Gold (Invitrogen) in ddH$_2$O, and visualized in UV gel documentation system. The Quick-Load® 50-bp DNA ladder (New England Biolabs) was used to determine allele sizes.

**Data analyses and genetic mapping**

Chi-squared analyses were performed to determine the goodness-of-fit of observed segregation with the expected genetic ratios (1:2:1 and 1:1 in $F_3$ and RIL population, respectively) and to detect marker-trait linkages. The genotypic status of each RIL with respect to the resistance gene under study was deduced from seedling stripe rust response data. Recombination fractions were calculated with the MAP MANAGER version QTXb20 (Manly et al. 2001) and converted to centimorgans (cM) using the Kosambi mapping function (Kosambi 1944). Logarithm of odds (LOD) score of 3.0 was used to determine significance of genetic linkages. MapChart software (Voorrips 2002) was used to construct and align three genetic maps for a visual inspection of map order.

**Results**

**Inheritance studies**

*YrAW1* produced infection type (IT) ;1-nn and a relatively higher IT2C was observed in some experiments. The MSP#5515 was tested at the seedling stage against the Pst pathotype 134E16A+Yr17+Yr27+. It was classified into three categories, namely: homozygous resistant (;1-nn), segregating (;1-n, 3+), and homozygous susceptible (3+). Monogenic segregation of *YrAW1* in MSP#5515 was confirmed (Table 2). The segregating families (*Yr51yr51*) included a low proportion of resistant individuals indicating the recessive mode of inheritance of resistance. The susceptibility of F1 plants from crosses of *YrAW1* stock (AUS91456) with susceptible cultivars confirmed

the recessive nature of this gene. MSP#5515-derived RIL population was tested at the seedling stage, and RILs were classified as homozygous resistant (;1-n) and homozygous susceptible (3+). Chi-squared analysis of stripe rust response variation conformed to single gene ratio (Table 2).

Multi-pathotype tests

Resistant RILs (*YrAW1YrAW1*) produced IT ;n–;1-nn and susceptible RILs (*yrAW1yrAW1*) produced IT 3+ against six Pst pathotypes 134 E16A+, 134 E16A+Yr17+, 134 E16A+Yr17+Yr27+, 110 E143A+, 108 E141A+, and 104 E137+ (Fig. 1). These results supported the effectiveness of *YrAW1* against a range of Australian Pst pathotypes carrying virulence for stripe rust resistance genes present in the global wheat germplasm.

Molecular mapping

*Chromosome location of YrAW1*

DArT markers based BSA identified association of 14 DArT markers with *YrAW1* in the long arm of chromosome 4A. List of linked DArT markers and their map locations on the consensus DArT map (Diversity Array Technology Pty Ltd, Australia, personal communication) are given in Table 3. Linked DArT markers were converted into STS

**Table 2** Frequency distribution of AUS27858/Westonia-derived MSP#5515 and MSP#5515-derived RIL population when tested against Pst pathotype 134 E16A+Yr17+Yr27+ at the seedling stage

| Genotype | Number of families | | $\chi^2_{(1:2:1)}$ |
|---|---|---|---|
| | Observed | Expected | |
| MSP#5515 | | | |
| *YrAW1YrAW1* | 24 | 22 | 0.18 |
| *YrAW1yrAW1* | 43 | 44 | 0.02 |
| *yrAW1yrAW1* | 21 | 22 | 0.05 |
| Total | 88 | 88 | 0.25 |
| MSP#5515-derived RIL population | | | $\chi^2_{(1:1)}$ |
| *YrAW1YrAW1* | 42 | 42.5 | 0.006 |
| *yrAW1yrAW1* | 43 | 42.5 | 0.006 |
| Total | 85 | 85 | 0.012 |

Table value of $\chi^2_{(1:2:1)}$ at $P = 0.05$ and 2df $= 5.99$ and $\chi^2_{(1:1)}$ at $P = 0.05$ and 1df $= 3.84$



**Fig. 1** Infection types produced by a homozygous resistant line carrying *Yr51* with Pst pathotypes *1*) 134 E16A+, *2*) 134 E16A+Yr17+, *3*) 134 E16A+Yr17+Yr27+, *4*) 110 E143A+, *5*) 108 E141A+, and *6*) 104 E137A+ and the susceptible control Morocco

**Table 3** List of STS markers derived from DArT clone sequences and their locations on DArT consensus map (Diversity Array Technology Pty Ltd, Australia, personal communication)

| STS markers | DArT clones | DArT consensus map (cM) |
|---|---|---|
| sun108 | wPt-5003 | 87.86 |
| sun105 | wPt-3795 | 87.86 |
| sun103 | wPt-0150 | 99.79 |
| sun111 | wPt-6966 | 100.32 |
| sun106 | wPt-4487 | 102.74 |
| sun102 | rPt-7987 | 102.74 |
| sun112 | wPt-731166 | 103.08 |
| sun114 | wPt-742051 | 103.09 |
| sun109 | wPt-5172 | 104.60 |
| sun107 | wPt-4620 | 104.60 |
| sun104 | wPt-763 | 104.60 |
| sun101 | rPt-0238 | 104.60 |
| sun110 | wPt-6176 | 105.83 |
| sun113 | wPt-731374 | 106.90 |

markers (*sun101*, *sun102*, *sun103*, *sun104*, *sun105*, *sun106*, *sun107*, *sun108*, *sun109*, *sun110*, *sun111*, *sun112*, *sun111* and *sun114*) and tested on contrasting bulks and parents. Two STS markers (*sun104* and *sun106*) that generated repeatable polymorphisms between parents and contrasting bulks were tested on the entire MSP#5515 RIL population. Both STS markers behaved as dominant markers and amplified products in one of the parents only. Marker *sun106* did not amplify any product in AUS27858, and on the other hand parent Westonia was null for marker *sun104*. Markers *sun104* and *sun106* were mapped 2.5 cM and 1.8 cM distal and proximal to *YrAW1*, respectively. These results confirmed the location of *YrAW1* in chromosome 4AL. Since there was no other stripe rust resistance located in chromosome 4AL, *YrAW1* was formally named *Yr51*.

Saturation of 4AL map

Simple sequence repeat, eSTS, and gene-based markers were used to saturate the *Yr51* carrying region of chromosome 4AL. Of 14 SSR markers tested, five markers *gpw7051*, *gwm855*, *gwm160*, *mag3273* and *barc78* were mapped distal to *Yr51* (Fig. 2). Four eSTS markers (*sun139*, *sun140*, *sun154*, and *sun155*) flanked *Yr51*. The markers *sun139* and *sun140* were mapped 8.2-cM distal to *Yr51*, and *sun154* and *sun155* were mapped at 1.8-cM proximal to *Yr51* (Fig. 2).

Of 136 gene-based markers, six showed polymorphism between parents and contrasting bulks. These markers were genotyped on the entire MSP#5515 RIL population. Marker *owm45F3R3* mapped 1.2-cM proximal to *Yr51*.

**4AL**
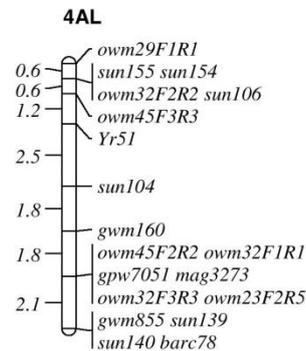


**Fig. 2** Genetic linkage map of chromosome 4AL showing location of stripe rust resistance gene *Yr51* based on DArT-derived STS, SSR, eSTS and the 4AL zipper-derived *owm* markers in the MSP#5515 RIL population

A linkage map consisting of 18 markers (2 STS, 4 eSTS, 5 SSR, and 7 gene-based markers) was constructed using phenotypic and genotypic data of MSP#5515 RIL population (Fig. 2). The sequences of markers (except SSR) used in the linkage map are given in Table 4. The linkage map spanned over a total genetic distance of 10.6 cM.

Validation of *Yr51*-linked markers

Since *Yr51* is not present in modern wheat genotypes, positive validation was not feasible. Markers *owm45F3R3* and *sun104* were genotyped on a set of 27 Australian and 13 Indian wheat lines to check the absence of *Yr51*-linked alleles of these markers, often referred to as negative validation (Table 5). Marker *sun104* amplified 225 bp in resistant parent AUS27858 and null in susceptible parent Westonia. All test cultivars did not amplify the *Yr51*-linked 225 bp allele indicating the usefulness of this marker in marker-assisted selection of this gene in these backgrounds. We did not get meaningful results with marker *owm45F3R3*, presumably due to differences in chromosomal rearrangements in this region. Therefore, *sun104* can be used for marker-assisted selection of *Yr51* in wheat genotypes lacking the resistance-linked 225-bp allele.

Discussion

Intensive cereal improvement and global spread of elite wheat germplasm led to a decrease in genetic diversity (Feuillet et al. 2008). To replenish the gene pool of modern varieties, landraces and uncultivated wheat relatives can serve as a valuable source of genetic variation. The

**Table 4** Primers polymorphic in the *Yr51* region and designed in this study using different genomic resources

| Marker | Forward sequence | Reverse sequence |
|---|---|---|
| Gene-based markers | | |
| *owm23F2R5 (Os06g0107600)*[a] | CATGGTGTCCCTCGTCAAG | AGGTAGAGCGTCTCGTGCAG |
| *owm29F1R1 (Os06g0106100)* | CATCACAGGCTCTTTCAGCA | GCTCGTGGAGAGACCAAGAC |
| *owm32F1R1 (Os06g0105800)* | ACGGTCTTCCTTCGTGGGTA | ACGCTCACGACATCGCTAAT |
| *owm32F2R2 (Os06g0105800)* | GGATCTCCTACGCTCTCGTG | TTGATCCAGATACAACAGGACAT |
| *owm32F3R3 (Os06g0105800)* | CGCCCCCAAGAAAGTTGTAT | TGCAAACGAGGACACATTTC |
| *Owm45F2R2 (Os06g0107700)* | GGCTCGTCTACACCAACGAC | TTGGGGTCTTTAGGCATGAG |
| *Owm45F3R3 (Os06g0107700)* | CGCAACAGGGACCGGTAT | GAGCTGCTGGTCGGAACTC |
| DArT-STS markers | | |
| *sun104 (wPt-763)* | TGCTATGTGCGTGATGATGA | TTACATGCTCCAGCGACTTG |
| *sun106 (wPt-4487)* | TGCACACAAGGAGAGGAGTG | AGAGGACAGTGCCCGTGTAG |
| eSTS markers | | |
| *sun139 (BF483646.1)* | TTTGGTCGGTTGGTTTGTTT | CCCCGACATCATCCTTTTTA |
| *sun140 (BF483646.2)* | CCGCACATATACATATAACCTCAA | CCTCCCTGTGCACAAACATA |
| *sun154 (BE444404.1)* | ATATTAGGGGCAAGCAAGCA | TCTCCCCAAGAACACCAAAC |
| *sun155 (BE444404.2)* | GTTTGGTGTTCTTGGGGAGA | ATTCCAACCTGCCCTGTATG |

[a] Owm markers developed from syntenic region of rice genome

**Table 5** Validation of *Yr51*-linked marker *sun104 (wPt-763)* on diverse wheat genotypes

| Cultivars/RIL | Allele size (bp) |
|---|---|
| AUS27858 and *Yr51* carrying resistant RIL (AUS91456) | 225 bp |
| Westonia and susceptible RIL | Null |
| Australian genotypes | |
| Braewood, Calingiri, Camm, Carinya, Carnamah, Derrimut, Diamondbird, EGA Bonnie Rock, EGA Gregory, Ellison, Frame, Giles, Gladius, Goldmark, H45, Halberd, Kellalac, Kukri, QAL2000, Rubric, Sunsoft 98, Sunlin, Sunvale, Sunzell, Tatiara, Ventura, Wyalkatchem | Null |
| Indian genotypes | |
| HD2402, PBW502, PBW343, PBW533, PBW550, FLW2, FLW6, K9107, HD2733, WH542, DBW17, PBW343 + *Lr24* + *Lr28*, PBW343*2/Kukuna | Null |

transfer of favorable genes from wild relatives of wheat often accompany with unwanted genes, whereas use of landraces in wheat improvement has not shown such disadvantages.

Isolation of *YrAW1* in MSP#5515 singly enabled the confirmation of its monogenic inheritance and was located in chromosome 4AL through BSA using DArT markers. It was named *Yr51* and shown to be effective against key Australian Pst pathotypes tested. Using SSR markers from the wheat composite map (wheat.pw.usda.gov/GG2/index.shtml), the precise location of *Yr51* (Fig. 2a) in the most distal deletion bin 4AL4-0.80-1.00 of chromosome 4A was determined.

During the evolution of common wheat, chromosome 4AL has undergone translocations and inversions. Two reciprocal translocations events, pericentric and paracentric inversions in 4AL, have been previously reported (Devos et al. 1995). First translocation occurred at the diploid level between chromosome 4AL and 5AL. Then, a pericentric inversion took place before another translocation between 4AL and 7BS at the tetraploid stage. Paracentric inversion resulted in modern 4AL chromosome containing segments of 7BS, 5AL, ancestral 4AL and proximal segment of the ancestral 4AS (Naranjo et al. 1987; Devos et al. 1995; Miftahudin et al. 2004; Hernandez et al. 2012). Berkman et al. (2012) reported that 13 % genes has been translocated from 7BS to 4AL, and 13 genes in chromosome 7BS appear to have originated from 4AL. Due to complex composition of the chromosome 4A, mapping is a challenging task. To saturate the *Yr51* region, several marker resources were explored. The public domain markers included SSR, EST (GrainGenes 2.0, http://wheat.pw.usda.gov) and "*mag*" markers (Xue et al. 2008). In addition, to identify markers closely linked with *Yr51* we utilized a large synteny study of 4A chromosome specific survey sequence with barley, rice, *Brachypodium*

and *Sorghum* genomes—the 4A genome zipper (Hernandez et al. 2012).

The targeted marker development approach using the 4A genome zipper and 4AL survey sequence resulted in saturation of the *Yr51* region with eight additional gene-based markers. Similarly, synteny-based approach using rice genome was used in high-density mapping and positional cloning projects in wheat (e.g., Distelfeld et al. 2004; Yan et al. 2004; Valárik et al. 2006). However, in many cases micro-collinearity in the region of interest was interrupted (Distelfeld et al. 2004; Valárik et al. 2006). In case of the 4A genome zipper utilizing four syntenic genomes, the breaks in collinearity of one genome could be bypassed by synteny in the other (Hernandez et al. 2012). In addition, the use of wheat sequence scaffolds for primer design increases effectiveness of PCR and specificity of products (Staňková et al. unpublished results). On the other hand, designing multiple primers pairs for single gene revealed multiple locations of genes from which markers *owm32* and *owm45* were developed (Fig. 2b). This observation could account for frequent gene duplication events and pseudogene evolution in wheat as described by Wicker et al. (2011). Markers *owm45F3R3* and *sun104* flanked *Yr51* at a genetic distance of 1.2 and 2.5 cM on the proximal and distal sides, respectively.

The closely linked marker *sun104* was negatively validated in a set of 40 genetically diverse wheat genotypes. Although the marker *owm45F3R3* mapped more closer to *Yr51*, it was not successfully validated in the absence of its resistance-linked allele among these 40 genotypes. Comparative sequence data from 4A Zipper (data not presented) indicated chromosomal rearrangements in this region. Marker *sun104* can be used in marker-assisted pyramiding of *Yr51* with other genes for which markers are available.

This project is part of the Australia–India collaboration, and therefore *Yr51* is currently being backcrossed into Australian and Indian wheat cultivars through marker-assisted selection. Recurrent parents carry marker-tagged stem rust and leaf rust resistance genes. Care will be taken to select triple rust resistant backcross derivatives for use as donors in wheat breeding programs in Australian, India, and elsewhere. Seed of genetic stock carrying *Yr51* singly has been deposited with the Australian Winter Cereal Collection Tamworth and it has been accessioned as AUS 91456.

## References

Akbari M, Wenzl P, Caig V, Carling J, Xia L, Yang S, Uszynski G, Mohler V, Lehmensiek A, Kuchel H, Hayden MJ, Howes N, Sharp P, Vaughan P, Rathmell B, Huttner E, Kilian A (2006) Diversity arrays technology (DArT) for high-throughput profiling of the hexaploid wheat genome. Theor Appl Genet 113:1409–1420

Bariana HS (2003) Breeding for disease resistance. In: Thomas B, Murphy DJ, Murray BG (eds) Encyclopedia of applied plant sciences. Harcourt, Academic Press, pp 244–253

Bariana HS, McIntosh RA (1993) Cytogenetic studies in wheat XV. Location of rust resistance genes in VPM1 and their genetic linkage with other resistance genes in chromosome 2A. Genome 36:476–482

Berkman PJ, Skarshewski A, Manoli S, Lorenc MT, Stiller J, Smits L, Lai K, Campbell E, Kubaláková M, Simková H, Batley J, Doležel J, Hernandez P, Edwards D (2012) Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. Theor Appl Genet 124:423–432

Devos K, Dubcovsky J, Dvořák J, Chinoy CN, Gale MD (1995) Structural evolution of wheat chromosomes 4A, 5A and 7B and its impact on recombination. Theor Appl Genet 91:282–288

Distelfeld A, Uauy C, Olmos S, Schlatter AR, Dubcovsky J, Fahima T (2004) Microcolinearity between a 2-cM region encompassing the grain protein content locus Gpc-6B1 on wheat chromosome 6B and a 350-kb region on rice chromosome 2. Funct Integr Genomics 4:59–66

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

Hernandez P, Martis M, Dorado G, Pfeifer M, Galvez S, Schaaf S, Jouve N, Simkova H, Valarik M, Dolezel J, Mayer KFX (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

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res 29:e25

Jakobson I, Reis D, Tiidema A, Peusha H, Timofejeva L, Valárik M, Kladivová M, Šimková H, Doležel J, Järve K (2012) Fine mapping, phenotypic characterization and validation of non-race-specific resistance to powdery mildew in a wheat-*Triticum militinae* introgression line. Theor Appl Genet 125:609–623

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

Manly KF, Cudmore RH, Meer JM (2001) Map Manager QTX, cross platform software for genetic mapping. Mamm Genome 12:930–932

McIntosh RA, Dubcovsky J, Rogers J, Morris C, Appels R, Xia X (2011) Catalogue of gene symbols for wheat: 2011 supplement

Miftahudin RK, Ma XF, Mahmoud AA, Layton J, Milla MA, Chikmawati T, Ramalingam J, Feril O, Pathan MS, Momirovic GS, Kim S, Chema K, Fang P, Haule L, Struxness H, Birkes J, Yaghoubian C, Skinner R, McAllister J, Nguyen V, Qi LL, Echalier B, Gill BS, Linkiewicz AM, Dubcovsky J, Akhunov ED, Dvorák J, Dilbirligi M, Gill KS, Peng JH, Lapitan NL, Bermudez-Kandianis CE, Sorrells ME, Hossain KG, Kalavacharla V, Kianian SF, Lazo GR, Chao S, Anderson OD, Gonzalez-Hernandez J, Conley EJ, Anderson JA, Choi DW, Fenton RD, Close TJ, McGuire PE, Qualset CO, Nguyen HT, Gustafson JP (2004) Analysis of expressed sequence tag loci on wheat chromosome group 4. Genetics 168:651–663

Naranjo T, Roca P, Goicoechea PG, Giraldez (1987) Arm homoeology of wheat and rye chromosomes. Genome 29:873–882

Paillard S, Schnurbusch T, Winzeler M, Messmer M, Sourdille P, Abderhalden O, Keller B, Schachermayr G (2003) An integrative genetic linkage map of winter wheat (*Triticum aestivum* L.). Theor Appl Genet 107:1235–1242

Poland JA, Brown PJ, Sorrells ME, Jannink JL (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. PLoS One 7:e32253. doi:10.1371/journal.pone.0032253

Somers D, Isaac P, Edwards K (2004) A high-density microsatellite consensus map for bread wheat (*Triticum aestivum* L.). Theor Appl Genet 109:1105–1114

Sourdille P, Singh S, Cadalen T, Brown-Guedira GL, Gay G, Qi L, Gill BS, Dufour P, Murigneux A, Bernard M (2004) Microsatellite-based deletion bin system for the establishment of genetic-physical map relationships in wheat (*Triticum aestivum* L.). Funct Integr Genomics 4:12–25

Valárik M, Linkiewicz AM, Dubcovsky J (2006) A microcolinearity study at the earliness per se gene Eps-A m1 region reveals an ancient duplication that preceded the wheat-rice divergence. Theor Appl Genet 112:945–957

Voorrips RE (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. J Hered 93:77–78

Wicker T, Mayer KFX, Gundlach H, Mihaela M, Burkhard S, Uwe S, Simkova H, Kubalakova M, Choulet F, Taudien S, Platzer M, Feuillet C, Fahima T, Budak H, Dolezel Y, Keller B, Stein N (2011) Frequent genetic movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. Plant Cell 23:1706–1718

Xue S, Zhang Z, Lin F, Kong Z, Cao Y, Li C, Yi H, Mei M, Zhu H, Wu J, Xu H, Zhao D, Tian D, Zhang C, Ma Z (2008) A high-density intervarietal map of the wheat genome enriched with markers derived from expressed sequence tags. Theor Appl Genet 117:181–189

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

# APPENDIX V

**Sequence ready physical map of bread wheat chromosome 4A**

Klocová B, Abrouk M, Frenkel Z, Kumar A, Kianian SF, Šimková H, Šafář J, Hu Y, Luo M, Carling J, Kilian A, Korol AB, Wang S, Akhunov E, Doležel J, Valárik M

# Sequence-ready physical map of bread wheat chromosome 4A

Barbora Klocová[1], Michael Abrouk[1], Zeev Frenkel[2], Ajay Kumar[3], Shahryar F. Kianian[3,4], Hana Šimková[1], Jan Šafář[1], Yuqin Hu[5], Mingcheng Luo[5], Jason Carling[6], Andrzej Kilian[6], Abraham Korol[2], Shichen Wang[7], Eduard Akhunov[7], Jaroslav Doležel[1], Miroslav Valárik[1]

[1]Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Slechtitelu 31, Olomouc, CZ- 78371, Czech Republic,
[2]Institute of Evolution, University of Haifa, Haifa 31905, Israel
[3]Department of Plant Sciences, North Dakota State University, Fargo, ND 58108, USA
[4]USDA-ARS Cereal Disease Laboratory, University of Minnesota, St. Paul, MN 55108, USA
[5]Department of Plant Sciences, University of California, Davis, CA 95616, USA
[6]Diversity Arrays P/L, 1 Wilf Crane Crescent, Yarralumla, Canberra, ACT 2600, Australia
[7]Department of Plant Pathology, Kansas State University, Manhattan, KS 66506-5502, USA

## Introduction

Despite the progress in Next Generation Sequencing technologies (NGS), sequencing of large and complex plant genomes like bread wheat (*Triticum aestivum* L.) remains a challenge. Size (17 Gb), complexity (allohexaploid), and repetitive element content (over 80 %) of the genome are the main obstructions for obtaining a reference genome sequence. Dissecting the wheat genome to chromosomes and/or chromosome arms is a powerful tool to overcome some of difficulties. As a part of coordinated effort by the **International Wheat Genome Sequencing Consortium (IWGSC)**, we have constructed physical maps of chromosome 4A using BAC libraries from flow-sorted short (4AS) and long (4AL) arms. Newly developed tools and mapping resources like GenomeZipper, chromosome-specific Radiation Hybrid panels (RH), super-clustering tool based on LTC software, 90K Infinium wheat SNP chip, DArT markers and other high-throughput mapping platforms and advanced mapping populations supports physical map anchoring and improvement.

## Result

- The physical maps of 4AS and 4AL were assembled using LTC software and the super-clustering tool based on LTC software (Table 1).
- The GenomeZipper was constructed from 4A survey sequence assembly. It identified and tentatively ordered 2 395 genes. As a backbone map, the consensus wheat map consisting of 1 780 DArT markers, was used to identify 612 loci that were found in the 4A survey sequence (Figure 2).

- A set of 4A-specific deletion lines and the 4A RH panel were genotyped with the 90K Infinium wheat SNP chip.
- MTPs with expected 50% overlap of BACs for physical maps of 4AS and 4AL were selected and reordered.
- 3D pools of MTP clones for both arms were Illumina sequenced and used for physical map anchoring to the survey sequence and GenomeZipper. 99 % of contigs with more than 4 BACs had at least one marker mapped. Markers mapped to more than two contigs were not considered.



**Figure 1: 90K infinium deletion SNP map integrated with the 4A GenomeZipper**

Linkage between deletion SNP markers and putative position of genes in GenomeZipper is represented by lines. Red lines represent markers of 4AS arm. The blue and green lines represent linkage for 4AL arm. Individual bins of the deletion map are represented by blocks with different shades of red for 4AS and green and blue for 4AL arms. 2 683 SNPs were successfully genotyped to 14 out of 17 bins of the 4A deletion map. 750 markers were specific for 4AS and 1842 markers were specific for 4AL. 45 markers mapping to both arms represent markers from centromeric region. Bin_2 from 4AS and bin_9, bin_4 from 4AL had no marker mapped. The most probable explanation could be the size of bins or lack of markers for the bins. The deletion map was used to improve resolution and order of the GenomeZipper.

### Table 1: Summary of 4AS and 4AL arm specific physical maps

| Chromosome | 4AS | 4AL |
|---|---|---|
| Number of BACs in contigs | 24 040 | 44 319 |
| Number of contigs | 250 | 924 |
| N50* (kb) | 9 246 | 6 692 |
| L50* | 63 | 99 |
| MTP clones | 4 422 | 8 256 |
| Estimated coverage[b] | 94 % | 87 % |
| Number of DArT markers | 117 | 495 |
| GenomeZipper (Number of gene markers) | 653 | 1 742 |
| Number of Survey sequence based markers | 54 125 | 62 656 |
| Anchored contigs by survey sequences | 100 % | 99 % |

* based on a CB unit size of 1.5 kb
[b] Based on estimated size of chromosome arms (318 and 540 Mb for 4AS and 4AL arm respectively)

## Conclusion

- Physical maps for both arms of chromosome 4A were assembled using LTC software

- High-density GenomeZipper of 4A chromosome was constructed based on 612 markers consensus Chinese Spring DArT map

- 4A deletion map containing 14 bins was constructed using 90K Infinium wheat SNP chip with 2 683 SNPs mapped

- Gene order of the GenomeZipper was refined by the deletion map

- 100 % of the 4AS BAC contigs were anchored to 4AS survey sequence and 86 % of them were anchored to the GenomeZipper

- 99 % of the 4AL BAC contigs were anchored to the 4AL survey sequence and 67 % of them to the GenomeZipper



**Figure 2: GenomeZipper and physical map anchoring**

The GenomeZipper is represented by the consensus 4A DArT map and syntenic gene order based on *Brachypodium* (Bd), rice (Os) and sorghum (Sb) genes. The pink labelled area represents example of a GenomeZipper part with physical map anchored. The table provides a detailed information on linkage between the GenomeZipper and physical maps. 86 % and 67 % of the 4AS and 4AL BAC contigs, respectively, were anchored to the GenomeZipper. The red highlighted lines in the table represent multicopy markers, which were not considered. The orange highlighted lines represent markers specific for two contigs and may indicate merging of contigs. The remaining lines represent highly reliable single copy markers.

## Future

Based on the acquired information and genomic resources, the 4A physical maps will be further improved by contigs merging, anchoring and orientation. For this goal the SNP mapping data on Radiation Hybrid map will be used. Finally, all 4A MTP clones will be sequenced through BAC by BAC approach using Illumina and reference sequence will be acquired.

## Acknowledgements

## References

Frenkel, Z.; Paux, E.; Mester, D.; et al., BMC Bioinformatics S (2010) DOI: 10.1186/1471-2105-11-584; Kumar et al. BMC Genomics, 13:597, (2012), www.biomedcentral.com/1471-2164/13/597; Mayer, K.F.X., et al. Plant Physiol. (2009) DOI: 10.1104/pp.109.142612; Raats et al., 2013. Genome Biology.2013, 14:R138; IWGSC: www.wheatgenome.org

# APPENDIX VI

**Construction and characterization of wheat 4A chromosome specific physical map as a base step for the chromosome sequencing**

Klocová B, Kladivová M, Frenkel Z, Kumar A, Kianian S F, Šimková H, Šafář J, Hu Y, Zhang Y, You F M, Luo M, Korol A, Doležel J, Valárik M

# Construction and characterization of wheat 4A chromosome specific physical map as a base step for the chromosome sequencing.

Barbora Klocová[1], Monika Kladivová[1], Zeev Frenkel[2], Ajay Kumar[3], Shahryar F. Kianian[3], Hana Šimková[1], Jan Šafář[1], Yuqin Hu[4], Yong Zhang[4], Frank M. You[4], Mingcheng Luo[4], Abraham Korol[2], Jaroslav Doležel[1], Miroslav Valárik[1]

[1]Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, Olomouc – Holice  783 71, Czech Republic
[2]Institute of Evolution, University of Haifa, Haifa 31905, Israel
[3]Department of Plant Sciences, North Dakota State University, Fargo, Loftsgard Hall 470G,ND 58108, USA
[4]Department of Plant Sciences, University of California, Davis, CA 95616, USA

## Introduction

*Triticum astivum* L. (bread wheat) has large and complex genome. The genome size is approximately 17 Gb with 80% of repetitive sequences and is composed of three homeologous genomes A, B and D. All of this makes sequencing, mapping, and **Marker Assisted Breeding** (**MAB**) a challenging task. Also, anchoring and orienting contigs of the physical maps are difficult.
To overcome all of these problems, the genome of cv. Chinese Spring is dissected to individual chromosomes and chromosomal arms by flow cytometry and chromosome specific BAC libraries are constructed. This allows to coordinate the collaboration of several laboratories from around the world in the framework of the **International Wheat Genome Sequencing Consortium** (**IWGSC**, http://www.wheatgenome.org) to succeed in wheat genome sequencing.
Our laboratory is coordinating development of sequence-ready chromosome arm specific physical maps for both arms of chromosome 4A.

## Results

Chromosome arm specific BAC libraries were constructed for wheat chromosome 4A. Both libraries were fingerprinted and assembled into contigs using both, **LTC** (**Linear Topology Contig**) and **FPC** (**FingerPrint Contigs**) software. The **LTC** assembly was enhanced by newly developed algorithm for super-scaffolding (Frenkel et al., 2010). Applying the super-scaffolding resulted in significant improvement of contig length (Table 1).
The **4AS** library was assembled into **416** contigs and **MTP** (**Minimum Tiling Path**) consist of **4 422** clones. After superclustering, there were 99 supercontigs with more than 100 clones in one contig. The longest supercontig comprised **582** clones.

Also, **4AL** physical map was assembled into **1129** contigs and **MTP** contains of **8 369** clones. After additional superclustering there were **115** supercontigs with more than 100 clones in one contig. The longest supercontig comprised **997** clones.
Moreover, the supercontigs of both physical maps contains more than half of informative clones. The newly established connections between contigs in supercontigs were evaluated by molecular markers (Figure 1).
3D pools of **MTP** clones were prepared to simplify marker anchoring to the maps. To the physical map of 4AL were anchored 20 markers, four of this markers were found in one supercontig.

### Table 1.: Summary of 4AS and 4AL arm specific physical maps

| Chromosome | 4AS | | | 4AL | | |
|---|---|---|---|---|---|---|
| assembly method | FPC | LTC | LTC (super-scaffolding) | FPC | LTC | LTC (super-scaffolding) |
| Number of all useful BACs | | 32 944 | | | 60 140 | |
| Number of BACs  in contigs > 5 clones | 24 633 | 24 040 | 24 040 | 39 285 | 41 097 | 41 254 |
| Number of clusters>100 clones | 76 | 73 | 99 | 75 | 105 | 115 |
| Number of clusters 100- 50 clones | 171 | 106 | 31 | 258 | 151 | 64 |
| Number of clusters < 50 clones | 203 | 238 | 118 | 972 | 872 | 745 |
| Number of clones in clusters> 50 clones | 20 212 | 18 563 | 21 544 | 23 103 | 29 634 | 32 856 |
| Number of clones in clusters< 50 clones | 4 421 | 5 477 | 2 496 | 16 182 | 11 463 | 8 398 |
| Number of singletons | 8 311 | 8 904 | 8 904 | 20 855 | 19 043 | 18 886 |
| MTP by LTC | 4 433 | 4 422 | 4 422 | 7 291 | 8 369 | 8 526 |
| Estimated coverage (%)[a,b] | 83 | 94 | 94 | 72 | 87 | 87 |

[a] based on the expected size of chromosome arms (318 and 540 Mb for 4AS and 4AL arms, respectively)
[b] based on the total number of MTP clones and expected overlap 30% and 50% for the LTC and FPC, respectively and average insert size 131 and 128 kb for 4AS and 4AL BAC libraries, respectively

## Conclusion and prospects

- **4AS** physical map has 248 contigs after super-scaffolding. This contigs consist of 24 040 clones.
- **4AL** physical map has 924 contigs after super-scaffolding. This contigs consist of 41 254 clones.
- **4AS** and **4AL MTP**  clones covers **94 %** and **87 %** length of arm, respectively.
- After sequencing of 3D pools of **MTP** BAC clones high-capacity physical map anchoring will be performed *in silico*.
- Endosperm radiation hybrid panel comprising 1 100 lines will be used for contig ordering and orientation.

## References

- Frenkel, Z.; Paux, E.; Mester, D.; et al., BMC Bioinformatics S (2010) DOI: 10.1186/1471-2105-11-584
- Xue S., Zhang Z., Lin F., et al., Theor Appl Genet (2008) DOI 10.1007/s00122-008-0764-9

### Figure 1: Super-contig algorithm verification

The super-contig SuperCtg5 comprises 294 BAC clones and represents about 3.6 Mb of chromosome 4A. It was put together by end-to-end merging of five standard LTC contigs and one singleton (202C08) via lower significant clone overlaps (without using markers).

The nodes correspond to clones; edges correspond to significant clone overlaps; colors of edges reflect p-value; nodes corresponding to the same contig have the same color and geometric shape.

Connecting clone 202C08 was not included into initial assembly because it is not proven by overlaps of parallel clones (probably putative chimerical). More detailed analysis detected low significant overlap between clones from the ends of Ctg16496 and Ctg16615 (yellow edge) proving this clone (202C08).

The connection between these two contigs was verified by molecular marker MAG2139 (Xue et al., 2008). Nodes corresponding to clones carrying this marker are marked by red arrows.

**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**

Barbora Balcárková

# Physical map of the wheat chromosome 4AL and positional cloning of a gene for yield

**P1527 Biology - Botany**

**Summary of Ph.D. Thesis**

**Olomouc 2017**

Ph.D. candidate:      **Mgr. Barbora Balcárková (Klocová)**


Supervisor:           **Mgr. Miroslav Valárik, Ph.D.**
                      Institute of Experimental Botany, Olomouc

Reviewers:            **Prof. RNDr. Milan Navrátil, CSc.**
                      Palacký University, Olomouc

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

                      **Mgr. Zdeněk Kubát, Ph.D.**
                      Institute of Biophysics, Brno

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 Department 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.) is one of the most important crops world-wide. Wheat provides basic food for 35 % of the human population and is grown in a wide range of environments, mainly in temperate zone. Its importance can be compared only with rice or maize. Fast growing human population dramatically increases importance of food production including the production of wheat.

Breeding of highly yielding wheat cultivars resistant to biotic and abiotic stresses is the main challenge for current breeding programs. Achieving goals to generate more yielding cultivars is hampered by polyploid nature of wheat genome. The hexaploid genome of bread wheat is result of two successive spontaneous hybridization events between three different diploid progenitors' species (*T. urartum*, AA; the most likely one of *Aegilops speltoides* from the *Sitopsis* group BB; goat grass *Ae. tauschii*, DD) from *Poaceae* family. The hybridization resulted in huge 17 Gb allohexaploid genome (Shewry, 2009; Matsuoka, 2011; Peng *et al*., 2011). All of the sub-genomes are closely related and similar to each other. Moreover, the majority of the genome (over 80 %) comprises repetitive elements (Moore 1995; Choulet *et al*., 2014).

Breeding, mapping, marker development, Marker Assisted Selection (MAS) and genome sequencing in such large and complex genome is still challenging (Feuillet *et al*., 2008). Reduction of the complexity by chromosome flow-sorting offers a solution. This approach offers genome reduction up to fifty times (Doležel *et al*., 2007). The flow-sorting of chromosomes or chromosomal arms provides sufficient quantity of high molecular weight DNA for survey sequencing (IWGSC, 2014) and BAC library construction (Šimková *et al*., 2011). Such approach also allowed distribution of the workload for the wheat genome sequencing to different laboratories organized in the International Wheat Genome Sequencing Consortium (IWGSC; Feuillet *et* Eversole, 2008).

The main strategy for wheat genome sequencing was construction of physical maps for all wheat chromosomes or chromosome arms (Eversole *et al*., 2014). The chromosome specific BAC libraries were fingerprinted (Luo *et al*., 2003) and obtained high informative fingerprints were assembled into physical maps using either FPC (Soderlund *et al*., 2000) or LTC (Frenkel *et al*., 2010) softwares. BAC clones of the Minimal Tiling Path (MTP) were selected and used for sequencing, marker development, positional cloning of genes, and anchoring of physical map contigs onto chromosomes.

The anchoring was mediated through genetic maps or virtual gene order (GenomeZippers, GZ; Mayer *et al*., 2011). Only, MTPs of oriented and anchored physical maps can provide full utilization of their potential (Meyers *et al*., 2004; Paux *et al*., 2008a; Ariyadasa *et* Stein, 2012). However, genome regions with limited recombination lack information about marker order and their utilization in physical map anchoring is limited. Even the GZ uses for prediction of gene order synteny with sequenced grass genomes of rice, sorghum and *Brachypodium*, as a back bone of GZ was used recombination map (Mayer *et al*., 2011). The GZs for all wheat chromosomes are currently available (IWGSC, 2014) but their use is limited because at least one–third of the wheat genome, especially in centromeric and pericentromeric regions, is located in recombination poor regions. Recombination mapping in such regions does not offer sufficient resolution for marker ordering (Akhunov *et al*., 2003). Ideal solution of this limitation offers Radiation hybrid (RH) mapping which is based on radiation-induced deletions thus markers ordering is independent from recombination (Michalak de Jimenez *et al*., 2013). RH mapping was originally used to map human genome (Goss *et* Harris, 1975) and adapted to wheat (Hossain *et al*., 2004). At present, the RH mapping together with high-throughput genotyping techniques (e.g. SNP chips, DArTs, POPSEQ, …) becomes a powerful tool for high-resolution mapping in wheat. RH maps were already used for anchoring of 3B and 6B physical maps (Paux *et al*., 2008b; Kobayashi *et al*., 2015).

The physical maps can significantly facilitate positional cloning of agronomically important genes because the final step in the cloning procedure is construction of physical map of cloned gene locus (Keller *et al*. 2005). The contigs of physical maps are also valuable source of new markers for map saturation of cloned genes loci (Feuillet *et al*. 2003). Such markers could be also used in breeding process for MAS of desirable genes in breeding materials and their progenies.

This Ph.D. work is focused on construction of wheat 4A chromosome specific physical map, its use in cloning and marker development of tightly linked markers to agronomically important genes affecting yield (*Qyi-4A-bga*), resistances to powdery mildew (*QPm-tut-4A*), rust (*Yr51*) and pre-harvest sprouting (*Phs-A1*).

The yield is complex trait composed of several components and is significantly influenced by biotic and abiotic stresses. Recently, QTL affecting Thousand Kernel Weight (*Qyie-4A-bga*) was identified. Effect of the QTL was particularly stable in different environments and contributes between 5 and 17 % to the overall phenotypic variation. The QTL was identified on chromosome 4A in 50 cM region between DNA markers *wmc161* and *wmc760*. Using the 4A physical map, the *Qyi-4A-*

*bga* gene locus was reduced to 0.5 cM and the flanking markers are used by breeders in pre-breeding and breeding process (unpublished results).

Fungus *Blumeria graminis* f. sp. *tritici* is a causal agent of wheat powdery mildew disease. The powdery mildew is accounted for up to 40 % of yield and quality loses. Recently, a major race nonspecific resistance locus affecting resistance in seedling and adult plant was introgressed to 4A chromosome. The resistance locus *QPm-tut-4A* was introgressed from *T. militinae* to the approximately 10 cM locus on 4AL chromosome arm (Jakobson *et al*, 2006). Physical map was used to saturate the gene region with new markers and reduce the locus to 0.02 cM.

Another fungus (*Puccinia striiformis* f. sp. *tritici)* causing stripe rust disease has also negative effect on yield. A resistance gene *Yr51* exhibiting high levels of resistance against Australian pathotypes of *P. striiformis* f. sp. *tritici* was identified on chromosome 4A. Using 4A physical map the gene locus was delimited to 3.7 cM (Randhawa *et al*., 2014). The flanking markers are currently being used for marker-assisted improvement of wheat breeding germplasms of Australian and Indian origin.

With changing climate additional yield and quality loss are often caused by high rain falls during harvest time. These loses are accounted to germination of cereal grains before harvest (also known as pre-harvest sprouting). The pre-harvest sprouting resistance gene *Phs-A1* was localised on chromosome 4AL. The 4A physical map was used for saturation of the region with molecular markers and candidate genes identification (Shorinola *et al*., 2016).

# 2   Aims of the thesis

### I.   Construction of physical map of the wheat chromosome 4A

The main aims of the thesis were construction of physical map of wheat chromosome 4A, the minimal tilling path (MTP) selection, preparation and sequencing of MTP 3D pools, and anchoring of physical map contigs to chromosome through available mapping and genomic resources.

### II.   Construction of high density RH map of the chromosome 4A

The second aim was construction of RH map which resolution is independent from natural recombination and allows marker ordering and the physical map anchoring in low recombining regions.

### III.   Application of the physical map for positional cloning of important wheat genes

The additional aim of the study was application of the created genomic resources for high density mapping and positional cloning of agronomically important genes.

# 3 Material and Methods

## 3.1 Construction of physical map of the wheat chromosome 4A

**Plant material**

Wheat chromosome arms 4AS and 4AL were flow-sorted from double-ditelosomic line of bread wheat (*Triticum aestivum* L.) cv. 'Chinese Spring' (CS) carrying both arms of chromosome 4A as telosomes.

**Construction of the 4A chromosome specific physical map**

Chromosome specific BAC libraries of 4AS and 4AL chromosomal arms were prepared as previously described by Šimková *et al*. (2011). BAC libraries were fingerprinted using SNaPShot kit to produce high-information-content fingerprints (HICF; Luo *et al*. 2003). Assembly of fingerprinted clones was performed using LTC software (Frenkel *et al*. 2010) to produce BAC contigs. MTP, which represents minimal number of clones covering chromosome, was selected from both physical maps. MTP clones were pooled in to 3D-pools according Paux *et al*. (2008b). 3D-pools of chromosome 4A were sequenced using Illumina Nextera kit and HiSeq 2000 NGS Illumina sequencer. Physical map was improved by advance tool of LTC named super-scaffolding. This tool allows end-to-end merging and contig elongation and helps reduces number of physical contigs.

**_In silico_ anchoring of physical map**

Physical map of chromosome 4A was *in silico* anchored to three new genomic recourses to improve its informative value. Firstly, sequences of 3D-pools were used to anchored physical map to CSS (Chromosome Survey Sequences; IWGSC, 2014) utilizing deconvolution approach described by Cviková *et al*. (2015). Secondary, physical map was anchored to 4A-specific RH map (Balcárková *et al*., 2017) through CSS. Abrouk *et al*. (2016) developed GenomeZipper especially for chromosome 4A. This GZ was used as third anchoring genomic resource.

## 3.2 Construction of high density RH map of chromosome 4A

**Plant material**

Nulli-tetrasomic lines (NT4A4B and NT4A4D) and standard bread wheat cv. Chinese Spring (Sears *et* Sears, 1978). 414 lines of 4A RH panel developed by crossing of irradiated CS pollen and the NT4A4B and NT4A4D lines.

Selected 15 deletion lines of 4A deletion stock (Endo *et* Gill 1996) containing different fragments of chromosome 4A were used for verification of RH map.

**Endosperm radiation hybrid map**

4A-specific endosperm RH (ERH) panel for each NT lines were prepared according to Tiwari *et al*. (2012). From 414 lines, 119 lines were selected equally covering the NT lineas and irradiation doses. The selected lines were genotyped by 90K iSelect wheat SNP chip (Wang *et al*., 2014) together with 15 deletion lines (Endo *et* Gill 1996). Genotyping data were manually analyzed using Illumina GenomeStudio v2011.1 software. RH map was constructed using combination of MultioPoind (Mester *et al*., 2003, Ronin *et al*., 2010) and LTC approach (Frenkel *et al*., 2010). Resulting 2 316 4A specific SNP markers were ordered in RH map comprising 1 080 mapping bins was verified by 4A deletion map (2 695 markers). The final RH map was used for *in silico* anchoring and orientation of 4A physical contigs especially in peri-centromeric and centromeric region.

## 3.3 Application of the physical map for positional cloning of important wheat genes

**Plant material**

Genetic mapping of different genes was carried out in various mapping populations.

- Powdery mildew resistance locus (*QPm-tut-4A*) and introgression studies

Development of mapping population for mapping of *Pm* resistance locus (*QPm.tut-4A*) and for introgression studies was described in Jakobson *et al*. (2006, 2012).

- Pre-harvest sprouting (*Phs-A1*)

Mapping of Pre-harvest sprouting gene (*Phs-A1*) was performed in two mapping population derived from UK winter wheat (*T. aestivum*) cultivars Option x Claire and Alchemy x Robigus. Details of mapping population development were described in Shorinola *et al.* (2016).

- Yellow rust resistance gene (*Yr51*)

Mapping population developed from bread wheat (*T. aestivum*) population labelled as HSB#5515 was used to evolution of advance mapping population of *Yr51* resistance gene. Procedure was in detail described in Randhawa *et al.* (2014).

- Locus influencing yield (*Qyie-4A-bga*)

3 000 lines of RIL F8 mapping population of bread wheat (*T. aestivum*) derive from cultivars CS x Renan and three NIL mapping populations produced using three parents, Aztec, Autan and Cezanne were used for *Qyie-4A-bga* loci mapping. The mapping populations were provided by international collaborator the Biogemma (France).

# 4    Summary of results

This thesis focuses on construction and utilization of genomic resources from the wheat chromosome 4A. The first goal of this work was to construct and anchored 4A physical map. The second part of the thesis is dedicated to construction of 4A-specific RH map for anchoring of physical contigs. The last part of the thesis was focused to application of anchored physical map to saturate loci of cloned genes with new markers.

## 4.1   Construction of physical map of the wheat chromosome 4A

Besides gene cloning and markers development, construction of the physical map represents an ideal base for genome sequencing and the sequence validation. With the ultimate aim to aid the wheat genome sequencing the physical maps of long and short arm of the 4A chromosome were constructed.

The physical map of 4A short arm comprise 415 contigs and 4 422 MTP clones were selected and pooled into 3D-pools. 4AS physical contigs were then end-to-end elongate in to 130 scaffolds and 120 contigs. The assembly of 4A long arm contains 1 127 contigs and 8 369 MTP clones were selected and pooled into 3D-pools. The physical map was subsequently improved and elongated and number of contigs decreased to 746 contigs and 178 long scaffolds.

Both physical maps were anchored in three levels. CSS were used as first level. Overall, 250 contigs and scaffolds of 4AS physical map were anchored with at least one CSS which represents 100 % of physical map. Likewise, in the 4AL physical map 911 contigs and scaffolds were anchored with at least one CSS. This represents 98.6 % of 4AL physical map. RH map was used as second level of physical map anchoring. Total of 199 and 496 contigs and scaffolds representing 79.6 % and 54.4 % of 4AS and 4AL physical map were anchored to 4A-specific RH map, respectively. As third level of physical map anchoring was used 4A-specific GZ (Abrouk *et al.*, 2016). Overall, 212 and 615 contigs and scaffolds representing 84.8 % and 67.5 % of 4AS and 4AL physical maps were anchored to GZ, respectively.

Finally, 88.4 % and 73.5 % of 4AS and 4AL physical map, respectively, is fully anchored, ordered and oriented along 4A chromosome. Remaining 11.6 % and 25.1 % of physical maps is anchored only to CSS.

An interactive internet application was developed to open the anchored physical map to public (https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_4AL_v1/).

## 4.2 Construction of high density RH map of chromosome 4A

ERH panel consist of 414 lines from which were 119 lines selected for genotyping by wheat 90K iSelect SNP array. Genotyping yielded 2 711 SNP markers specific for 4A chromosome. The final RH map comprises 2 317 markers and spanning 6 550.9 centi Ray (cR).

Deletion lines were also genotyped by 90K iSelect wheat chip. Verification deletion map was constructed from 2 687 SNP identified as 4A specific. Deletion map was used as control of marker order in RH map.

Sequences of SNP markers were aliened and anchored to CSS. Only 14 SNP markers cannot be anchored to any 4A CSS.

## 4.3 Application of the physical map for positional cloning of important wheat genes

Anchored physical map significantly facilitated saturation of region of all mapped loci. Physical map allowed development of new specific markers and saturated or narrowed down gene locus.

- Powdery mildew resistance locus (*QPm-tut-4A*) and introgression studies

Physical map facilitated development of specific markers for *QPm-tut-4A* locus saturation and narrowed it down to 0.02 cM. Mapping population used for mapping of this gene was derived from *T. militinae* introgression line 8,1 (Jakobson *et al.*, 2012) and resources created in this study were also used for development of new method for introgression characterization (Abrouk *et al.*, 2016).

- Pre-harvest sprouting (*Phs-A1*)

Region of possible gene locus for *Phs-A1* gene was narrowed down to 0.3 cM and candidate genes were identified. Physical map was used for identification of new markers and the gene locus saturation. The BAC sequences of contig spanning the locus were annotated and causal gene identified (Shorinola *et al.*, 2016)

- Yellow rust resistance gene (*Yr51*)

Region of gene resistance *Yr51* was delimited by 3.7 cM between markers *owm45F3R3* and *sun104*. The physical map and GZ were used for new marker de-

velopment in the gene region. The markers *owm45F3R3* and *sun104* are now successfully used for marker-assisted selection (Randhawa *et al.*, 2014).

- <u>Locus influencing yield (*Qyie-4A-bga*)</u>

Using the 4A physical map, the *Qyi-4A-bga* gene locus was reduced to 0.5 cM and the flanking markers are being used in pre-breeding and breeding process (unpublished results).

# 5 Conclusions

The main aim of presented study was construction a physical map of bread wheat 4A chromosome and development of 4A-specific RH map for anchoring and physical map use in positional cloning of agronomically important genes and ultimately as a base for sequencing of the chromosome.

In the first part of this thesis, two 4A arm specific physical map were constructed. MTP clones were selected and pooled in to 3D-pools. Then, 3D-pools were Illumina sequenced and were used for physical map anchoring and marker development. Anchoring of physical map was done on tree levels using available CS chromosome specific survey sequences, improved 4A specific GenomeZipper and 4A-specific RH map. Combination of these tree genomic recourses allowed anchoring and orientation of 104 % and 92 % of total length of 4AS and 4AL chromosome, respectively. The developed genomic resources were made available to public through scientific papers and internet application.

In the second part of the thesis, 4A-specific ERH panel was developed to construct recombination independent 4A-specific RH map. 119 lines were genotyped by 90K wheat SNP chip. Final RH map comprised of 2 317 markers and spans 6 550.9 cR. RH map allowed anchoring of 79.6 % and 54.4 % of 4AS and 4AL physical maps, respectively.

In the third part of the thesis, physical map of 4AL chromosome was involved in positional cloning of agronomically important genes. Physical map was successfully engaged in identification of candidate genes for pre-harvest sprouting and delimitation of alien introgression to 4AL chromosome arm. Additionally, the map was successfully used in saturation of loci and identification of tightly linked marker for effective MAS of resistance genes *Yr51* and *QPm-tut-4A,* and yield influencing gene *Qyie-4A-bga*.

# 6    References

Abrouk M, Balcárková B, Šimková H, Komínková E, Martis MM, Jakobson I, Timofejeva L, Vrána J, Kilian A, Järve K, Doležel J, Valárik M (2016): *In-Silico* Identification and Characterization of *Triticum militinae* Introgression into Bread Wheat. Plant Biotechnology Journal, 15(2):249-256

Akhunov ED, Goodyear AW, Geng S, Qi L, Echalier B, Gill BS, Miftahudin, Gustafson JP, Lazo G, Chao S, Anderson OD, Linkiewicz AM, Dubcovsky J, La Rota M, Sorrells ME, Zhang D, Nguyen HT, Kalavacharla V, Hossain K, Kianian, SF, Peng J, Lapitan NLV, Gonzalez-Hernandez JL, Anderson JA, Choi D, Close TJ, Dilbirligi M, Gill KS, Walker-Simmons MK, Steber C, McGuire PE, Qualset, CO, Dvorak J (2003): The organization and rate of evolution of wheat genomes are correlated with recombination rates along chromosome arms. Genome research 13:753-763

Ariyadasa R, Stein N (2012): Advances in BAC-Based Physical Mapping and Map Integration Strategies in Plants. Journal of biomedicine and biotechnology 2012:184854

Balcárková B, Frenkel Z, Škopová M, Abrouk M, Kumar A, Chao S, Kianian SF, Akhunov E, Korol A, Doležel J, Valárik M (2017): Endosperm Radiation Hybrid panel specific for 4A chromosome of hexaploid wheat. Front Plant Sci., 7:2063

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. Pp 405-438, Springer

Cviková K, Cattonaro F, Alaux M, Stein N, Mayer KFX, Doležel J, Bartoš J (2015): High-throughput physical map anchoring via BAC-pool sequencing. BMC Plant Biology 15:99

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

Endo TR, Gill BS (1996): The deletion stocks of common wheat. J Hered 87:295-307

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

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 U S A., 100(25):15253-8.

Feuillet C, Eversol K (2008): Physical mapping of the wheat genome: A coordinated effort to lay the foundation for genome sequencing and develop tools for breeders. Israel Journal of Plant Sciences

Feuillet C, Langridge P, Waugh R (2008): Cereal breeding takes a walk on the wild side, Trends Genet. 24(1):24-32.

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

Goss SJ, Harris H (1975): New method for mapping genes in human chromosome. Nature 255:680-684

Hossain KG, Riera-Lizarazu O, Kalavacharla V, Vales MI, Maan SS, Kianian SF (2004): Radiation hybrid mapping of the species cytoplasm-specific (*scsae*) gene in wheat. Genetics. 168(1):415-23

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

Jakobson I, Peusha H, Timofejeva L, Järve K (2006): Adult plant and seedling resistance to powdery mildew in a Triticum aestivum x Triticum militinae hybrid line, Theor Appl Genet. 112(4):760-9.

Jakobson I, Reis D, Tidema A, Peusha H, Timofejeva L, Valárik M, Kladivová M, Šimková H, Doležel J, Järve K (2012): Fine mapping phenotypic characterization and validation of non-race-specific resistance to powdery mildew in a wheat-Triticum militinae introgression line. Theor Appl Genet 125(3):609-23

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 JCM, 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

Keller B, Feuillet C, Yahiaoui N (2005): Map-based isolation of disease resistance genes from bread wheat: cloning in a supersize genome. Genet Res., 85(2):93-100

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

14

Matsuoka Y (2011): Evolution of Polyploid *Triticum* Wheats under Cultivation: The Role of Domestication Natural Hybridization and Allopolyploid Speciation in their Diversification. Plant Cell Physiol 52(5):750-764

Mayer KF, Martis M, Hedley PE, Simková 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, Dolezel J, Waugh R, Stein N (2011): Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1246-1263

Mester D, Ronin Y, Minkov D, Nevo E, Korol A (2003): Constructing Large-Scale Genetic Maps Using an Evolutionary Strategy Algorithm. Genetics 165(4):2269-2282

Meyer BC, Scalabrin S, Morgane M (2004): Mappimg and sequencing complex genomes: Let's get physical! Nature reviews 5:578-588

Michalak de Jimmenez MK, Bassi FM, Ghavami F, Simons K, Dizon R, Seetan RI, Alnemer LM, Denton AM, Doğramaci M, Šimková H, Doležel J, Seth K, Luo M, 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

Moore G (1995): Cereal genome evolution: pastoral pursuits with 'Lego' genomes. Curr Opin Genet Dev, 5(6):717-24.

Paux E, Legeai F, Guilhot N, Adam-Blomdon A, Alaux M, Salse J, Sourdille P, Leroy P, Feuillet C (2008a): Physical mapping in large genomes: accelerating anchoring of BAC contigs to genetic maps through in silico analysis. Funct Integr Genomics 8:29-32

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 (2008b): A physical map of the 1-gigabase bread wheat chromosome 3B. Science 322: 101-104

Peng JH, Sun D, Nevo E (2011): Domestication evolution genetics and genomics in wheat. Mol Breeding 28:281-301

Randhawa M, Bansal U, Valárik M, Klocová B, Doležel J, Bariana H (2014): Molecular mapping of stripe rust resistance gene Yr51 in chromosome 4AL of wheat. Theor Appl Genet 127: 317–324

Ronin Y, Mester D, Minkov D, Korol A (2010): Building reliable genetic maps: different mapping strategies may result in different maps. Natural Science 2(6): 576-589

Sears ER, Sears L (1978):"Tetelocentric chromosomes of common wheat," in *Proceedings of the 5th International Wheat Genetics symposium*, ed.S. Ramanujams (New Delhi: Indian Agricultural Reseach Institute), 389–407.

Shewry PR (2009): Wheat. Journal of Experimental Botany 60 (6):1537-1553

Shorinola O, Bird N, Simmonds J, Berry S, Henriksson T, Jack P, Werner P, Gerjets T, Scholefield D, Balcárková B, Valárik M, Holdsworth M, Flintham J, Uauy C (2016): The wheat Phs-A1 pre-harvest sprouting resistance locus delays dormancy loss during seed after-ripening and maps 03 cM distal to the PM19 genes in UK germplasm. Journal Ex Bot

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 V47. Genome Res 10:1772-1787

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 a Measure of γ Ray-Induced Chromosome Breaks. PLOS ONE 7:11

Wang S, Wong D, Forrest K, Allen A, Chao S, Huang BE, Maccaferri M, Salvi S, Milner SG, Cattivelli L, Mastrangelo AM, Whan A, Stephen S, Barker G, Wieseke R, Plieske J, IWGSC, Lillemo M, Mather D, Appels R, Dolferus R, Browm-Guedira G, Korol A, Akhunova AR, Feuillet C, Salse J, Morgante M, Pozniak C, Luo MC, Dvorak J, Morell M, Dubcovsky J, Ganal M, Tuberosa R, Lawley C, Mikoulitch I, Cavanagh C, Edwards KJ, Hayden M, Akhunov E (2014): Characterization of polyploid wheat genomic diversity using a high-density 90 000 single nucleotide polymorphism array. Plant Biotechnology Journal (12) pp 787-796

# 7 List of author's publications

## 7.1 Original papers

Balcárková B, Frenkel Z, Škopová M, Abrouk M, Kumar A, Chao S, Kianian SF, Akhunov E, Korol A, Doležel J, Valárik M: A High Resolution Radiation Hybrid Map of Wheat Chromosome 4A. – Front Plant Sci., 2017

Abrouk M, Balcárková B, Šimková H, Komínková E, Martis MM, Jakobson I, Timofejeva L, Rey E, Vrána J, Kilian A, Järve K, Doležel J, Valárik M: The *in silico* identification and characterization of a bread wheat/*Triticum militinae* intrigression line. – Plant Biotechnology Journal, 2016

Shorinola O, Bird N, Simmonds J, Berry S, Henriksson T, Jack P, Werner P, Gerjets T, Scholefield D, Balcárková B, Valárik M, Holdsworth M, Flintham J, Uauy C: The wheat Phs-A1 pre-harvest sprouting resistance locus delays dormancy loss during seed after-ripening and maps 0.3 cM distal to the PM19 genes in UK germplasm. – Journal Ex. Bot., 2016

Randhawa M, Bansal U, Valárik M, Klocová B, Doležel J, Bariana H: Molecular mapping of stripe rust resistance gene Yr51 in chromosome 4AL of wheat. – Theor. Appl. Genet. 127: 317–324, 2014.

## 7.2 Published abstracts – poster presentation

Klocová B, Abrouk M, Frenkel Z, Kumar A, Kianian SF, Šimková H, Šafář J, Hu Y, Luo M, Carling J, Kilian A, Korol AB, Wang S, Akhunov E, Doležel J, Valárik M: Sequence ready physical map of bread wheat chromosome 4A. – In: Book of Abstracts "EUCARPIA Cereals Section & ITMI Conference". P. 321., Wernigerode, 2014.

Klocová B, Kladivová M, Frenkel Z, Kumar A, Kianian S F, Šimková H, Šafář J, Hu Y, Zhang Y, You F M, Luo M, Korol A, Doležel J, Valárik M: Construction and characterization of wheat 4A chromosome specific physical map as a base step for the chromosome sequencing. – In: Book of Abstracts "Olomouc Biotech 2013 - Plant Biotechnology: Green for Good II", Olomouc, 2013

# 8    Summary (in Czech)

### Fyzická mapa chromosomu 4AL pšenice a poziční klonování genu pro výnos

Pšenice setá (*Triticum aestivum* L.) představuje jednu z nejvýznamnějších zemědělských plodin, jež poskytuje základní 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 chromosomů 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 chromosomu 4A. V rámci práce byla zkonstruována fyzická mapa obou ramen chromosomu 4A. Ty posloužily k výběru tzv. minimal tilling path (MTP), což představuje minimální sestavu klonů, které pokrývají celou délku obou ramen. Klony z MTP byly sloučeny do tzv. 3D-poolů a následně osekvenovány. Tyto sekvence umožnily ukotvení fyzické mapy na chromosom a její orientaci pomocí tří různých genomických zdrojů. Nejprve byla fyzická mapa propojena se „survey" sekvencemi (CSS) což dále umožnilo ukotvení fyzické mapy k 4A specifické radiačně hybridní mapě a k „GenomeZipperu" (GZ).

Součástí práce byl i vývoj radiační hybridní (RH) mapy, která představuje alternativní variantu genetického mapování především v chromosomových oblastech s nízkým nebo dokonce nulovým výskytem rekombinace. Radiační hybridní mapa je založena na mapování delecí indukovaných radiací a tudíž nezávislá na rekombinaci. Výsledná RH mapa obsahuje 2 317 SNP markerů a je dlouhá 6 550,9 cR. Pomocí RH mapy bylo ukotveno 79,6 % fyzické mapy krátkého a 54,4 % dlouhého ramene 4A chromosomu. Plně zorientováno a ukotveno k RH mapě a GZ je 88,4 % fyzické mapy krátkého ramene a 73,5 % fyzické mapy dlouhého ramene. Zbývajících 11,6 % fyzické mapy krátkého ramene a 25,1 % dlouhého ramene je ukotveno jen k CSS. Výsledné ukotvené fyzické mapy jsou veřejné dostupné přes webovou aplikaci (https://urgi.versailles.inra.fr/gb2/gbrowse/wheat_phys_4AL_v1/).

Kromě sekvenování je fyzická mapa nástrojem usnadňujícím poziční klonování agronomicky významných genů. Fyzická mapa chromosomu 4A byla úspěšně využita pro pozičním klonování lokusů ovlivňující výnos (*Qyi-4A-bga*), resistenci k padlí travnímu (*QPm-tut-4A*), rzi travní (*Yr51*) a genu ovlivňujícího klíčení semen v klasu před sklizní (*Phs-A1*).