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

Veronika Burešová

STUDY OF ORGANIZATION OF HYBRID AND POLYPLOID GENOMES

Ph.D. Thesis

Supervisor: RNDr. David Kopecký, Ph.D.

Olomouc

2017

Acknowledgements

I would like to thank to my supervisor, RNDr. David Kopecký, Ph.D., for his excellent ideas, patience and help. Further, thanks go to the head of laboratory, prof. Ing. Jaroslav Doležel, DrSc., for the opportunity to work in so successful research group and also to all members from the Centre of the Plant Structural and Functional Genomics who provided a friendly and motivating atmosphere for my research. Finally, I would like to thank to the foundation "Nadání Josefa, Marie a Zdeňky Hlávkových" and Erasmus for financial support during my internship.

This work was supported by the Ministry of Education, Youth and Sports of the Czech Republic, the European Regional Development Fund (Operational Programme Research and Development for Innovations No. ED0007/01/01) and Centre of region Haná for Biotechnological and Agricultural Research.

Declaration

I hereby declare that I elaborated this Ph.D. thesis independently under the supervision of RNDr. David Kopecký, Ph.D. using only information sources referred in the chapter References.

.....

BIBLIOGRAPHIC IDENTIFICATION

Author's name: Veronika Burešová

Title: Study of organization of hybrid and polyploid genomes

Type of Thesis: Ph.D. thesis

Department: Botany

Supervisor: RNDr. David Kopecký, Ph.D.

The Year of Presentation: 2017

Abstract:

Understanding structure and organization of plant genomes is one of the great challenges of current biology. Improvement of existing and introduction of new methods is necessary step for increasing our knowledge of mechanisms and processes which participate evolution of plant genomes. Molecular cytogenetics and microscopy methods enable the detailed analyses of the structure, evolution and variability of genomes. In addition, detailed understanding of plant genome structure can be beneficial for plant breeding.

The first part of the thesis focuses on the genome analysis of introgression lines of wheat-*Th. ponticum* with blue aleurone trait which exhibit significantly increased level of anthocyanins compare to the common varieties of wheat. The aim of our study was to characterize genomic constitutions of these elite lines using fluorescence *in situ* hybridization (FISH). Our results revealed large variation in genome composition of blue-aleurone wheats. We identified six different types of the *Th. ponticum* introgressions and this confirmed the hypothesis that alien chromatin from *Th. ponticum* activates the blue aleurone trait present, but inactivated, in common wheat lines.

Flow cytometry is the next important tool for analysis of complex hybrid and polyploid genomes. Dividing the genome into small defined units, chromosomes, makes the sequencing of large genomes easier. Nevertheless, chromosomes of the

majority of species have the same or very similar relative DNA content and thus, it is unable to sort them by conventional flow cytometry. Fortunately, the development of fluorescent *in situ* hybridization in suspension (FISHIS) is elegant solution of this situation. This method allows specific chromosome flow-sorting based on combination fluorescent signal of oligonucleotide SSR probe(s) and DAPI staining. The second part of the thesis deals with the optimization of FISHIS methodology for flow cytometric analysis in selected species of the *Triticeae* tribe, including *Triticum* or *Aegilops* species.

Keywords: genome, bread wheat, polyploidy, interspecific hybridization, FISH,

Number of Pages/Appendices: 86/VI

Language: English

BIBLIOGRAFICKÁ IDENTIFIKACE

Jméno: Veronika Burešová

Téma: Studium organizace hybridních a polyploidních genomů

Typ práce: Disertační práce

Obor: Botanika

Školitel: RNDr. David Kopecký, Ph.D.

Rok obhajoby: 2017

Abstrakt:

Porozumění struktuře a organizaci rostlinných genomů je jednou z velkých výzev současné biologie. Zdokonalování stávajících a zavádění nových metod je nezbytným krokem k rozšíření našich znalostí o mechanismech a procesech, které se účastní evoluce rostlinných genomů. Molekulární cytogenetika a mikroskopické techniky patří ke standartním metodám tohoto šetření. Tyto techniky umožňují detailní studium variability, struktury a evoluce rostlinných genomů. Takové poznání genomu rostlin může rovněž hrát zásadní roli ve šlechtění rostlin.

První část práce je zaměřena na analýzu genomu introgresních linií pšenice a pýru pontického s modrým aleuronem. Tyto linie vykazují výrazně vyšší hladinu antokyanů v porovnání s běžnými odrůdami pšenice. Cílem naší studie bylo charakterizovat genomové složení těchto linií pomocí fluorescenční *in situ* hybridizace (FISH). Naše výsledky odhalily velké rozdíly v genomovém složení u modrozrnných pšenic. Celkem jsme identifikovali šest různých typů introgrese pýru pontického. Toto zjištění tak potvrdilo hypotézu, že přítomnost cizího chromatinu, v tomto případě pýru, aktivuje gen pro modrý aleuron, který zůstává u běžné pšenice neaktivní.

Průtoková cytometrie je dalším důležitým nástrojem v analýze hybridních a polyploidních genomů. Rozdělení genomu na malé jasně definované jednotky, chromozomy, činí sekvenování velkých genomů daleko snazší. Nicméně

chromozomy většiny druhů mají stejný nebo velmi podobný relativní obsah DNA a tudíž třídění takových chromozomů běžnou průtokovou cytometrii není možné. Naštěstí vývoj fluorescenční *in situ* hybridizace v suspenzi (FISHIS) je elegantním řešením této situace. Tato metoda, založená na kombinaci fluorescenčního signálu oligonukleotidových mikrosatelitních sond a DAPI barvení, umožňuje třídění jednotlivých chromozomů. Druhá část této práce se zabývá optimalizací metodiky FISHIS pro průtokovou cytometrii vybraných rodů kmene *Triticeae* (*Triticum*, *Aegilops*).

Klíčová slova: genom, pšenice setá, polyploidie, intespecifická hybridizace, FISH

Počet stran/příloh: 86/VI

Jazyk: anglický

CONTENT

1 INTRODUCTION 11

2 LITERATURE OVERVIEW 12

2.1 GENOME 12

2.1.1 Plant nuclear genome composition 12

2.1.1.1 Eukaryotic genes 13

2.1.1.2 Repetitive DNA in eukaryotic genome 17

2.1.2 Chromatin 19

2.1.3 Organization of chromatin in eukaryotic genome 23

2.1.3.1 Chromosomes 23

2.1.3.2 Chromosome organization in the interphase nucleus 26

2.2 INTERSPECIFIC HYBRIDIZATION AND POLYPLOIDY AS TOOLS IN PLANT BREEDING 29

2.2.1 Plant hybridization 29

2.2.2 Polyploidy 31

2.2.3 Alterations associated with polyploidy 32

2.2.4 Introgression breeding 35

2.2.5 Wide hybridization of wheat 35

2.2.5.1 Introgressions between bread wheat and rye 36

2.2.5.2 Introgressions between bread wheat and barley 37

2.2.5.3 Introgressions between bread wheat and *Aegilops* (goat grasses) 38

2.2.5.4 Introgressions between bread wheat and *Thinopyrum* (syn. *Agropyron*) 38

2.2.5.5 Introgressions between bread wheat and other *Triticum* species 39

2.3 CHARACTERIZATION OF PLANT GENOMES USING CYTOGENETIC TECHNIQUES 40

2.3.1 Chromosome banding 40

2.3.2 Fluorescent *in situ* hybridization 42

2.4 REFERENCES 47

3 AIMS OF THE THESIS 67

4 RESULTS 68

4.1 PUBLISHED PAPERS 68

4.1.1 Variation in genome composition of blue-aleurone wheat 68

4.1.2 Advances in plant chromosome genomics 68

4.1.3 Molecular organization and comparative analysis of chromosome 5B of the wild wheat ancestor *Triticum dicoccoides* 68

4.1.4 Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their syntheny with wheat....68
4.2 PUBLISHED ABSTRACTS – POSTER PRESENTATIONS....74
4.2.1 Genomic constitutions of cereals with blue aleurone trait....74
4.2.2 Genomic constitution of wheat genotypes with blue aleurone....74
4.2.3 Genomic constitution of blue grained wheat genotypes....74
4.3 PUBLISHED ABSTRACTS – ORAL PRESENTATIONS....79
4.3.1 Nuclear organization in interspecific plant hybrids....79
4.3.2 Nuclear organization in interspecific plant hybrids revealed by 3D-FISH....79
5 CONCLUSION....82
5.1 CYTOGENETIC ANALYSIS OF INTROGRESSION LINES OF WHEAT-*THINOPYRUM PONTICUM* WITH BLUE ALEURONE TRAIT USING MULTICOLOUR FLUORESCENCE *IN SITU* HYBRIDIZATION (FISH)....82
5.2 OPTIMIZATION AND UTILIZATION OF FLUORESCENCE *IN SITU* HYBRIDIZATION IN SUSPENSION (FISHIS) FOR FLOW CYTOMETRIC ANALYSIS OF SELECTED SPECIES OF THE *TRITICEAE* TRIBE....82
6 LIST OF ABBREVIATIONS....84
7 LIST OF APPENDICIES....86

1 INTRODUCTION

Interspecific hybridization and polyploidy are driving forces in evolution and speciation of flowering plants. The study of hybrid and polyploid species by a range of current microscopy and molecular-cytogenetic techniques provides deeper view into structure, organization and function of complex plant genomes. The results of such studies are essential for effective plant breeding.

Bread wheat (*Triticum aestivum* L.) is one of the most important crops providing staple food for ~30% of the world population. It is an allohexaploid species originating from two interspecific hybridization events that gave rise to a highly redundant 17-Gb genome with three homoeologous sets of chromosomes. The long term domestication led to the significant decrease of genetic diversity of this important crop. One of the aims of modern breeding is to transfer genetic information (usually some agriculturally important traits) from wild relatives into wheat genome. This process, for which the term introgressive hybridization is used, is important tool for increasing the genepools and improving commercial varieties. So far, various types of introgression lines have been produced and numerous genes have been transferred into wheat germplasm. Various types of molecular-cytogenetics methods, including *in situ* hybridization techniques have been successfully employed for identification and localization of introgressed chromatin.

This thesis is focused on the characterization of genomic constitutions of elite lines of blue grained wheat using multicolour fluorescence *in situ* hybridization. Further, I aimed on optimization of fluorescence *in situ* hybridization in suspension (FISHIS) protocol used for flow cytometric analysis of selected species of the *Triticeae* tribe. These results are increasing our knowledge about the structure of complex plant genomes and will be prerequisite for subsequent genomic analyses.

2 LITERATURE OVERVIEW

2.1 Genome

The term "genome" was introduced by Hans Winkler in 1920 and represents all genetic material needed to build and maintain an organism (Winkler 1920). In eukaryotes, the majority of genetic information is located in the nucleus. Most eukaryotes have a single nucleus, but there are cell types, such as mammalian red blood cells that have no nuclei. On the other hand, some species of protozoa and fungi, as well as some types of human and plant tissue can have multinucleated cells (Zettler *et al*. 1997; Horton 2006). Besides nuclear DNA, a small part of genome is found in cytoplasm, primarily in organelles that contain their own DNA such as mitochondria and plastids of plants. Additionally, other components of the genome can be sequences from viruses, plasmids and other vectors having own genetic information.

The study of genomes by a range of microscopy and molecular methods provides deeper understanding into the evolutionary history of all living organisms (Leliaert *et al*. 2012). It was found that plant genomes are in general much more complex than genomes of other eukaryotes with extensive variation in genome size, chromosome number, ploidy level, arrangement of chromatin and number of genes (Heslop-Harrison 2000; Alberts 2002; Kellogg and Bennetzen 2004). The wide diversity of plant genomes is the result of highly active processes constantly influencing the development of species (Madlung 2013).

Next part of the thesis will be devoted to various types of genomic DNA sequences and packaging of the genome.

2.1.1 Plant nuclear genome composition

The plant nuclear genome consists of genes, regulatory sequences, and other non-coding sequences present in low copy number, and a various types of repetitive DNA (Figure 1) that makes up an abundant part of nuclear genome in most eukaryotes (Heslop-Harison and Schmidt 2007; Biscotti *et al*. 2015). The DNA sequences are evolutionarily very dynamic and their analysis provides an insight into

the genome structure, evolution and phylogenetics of species (Feschotte and Pritham 2007).

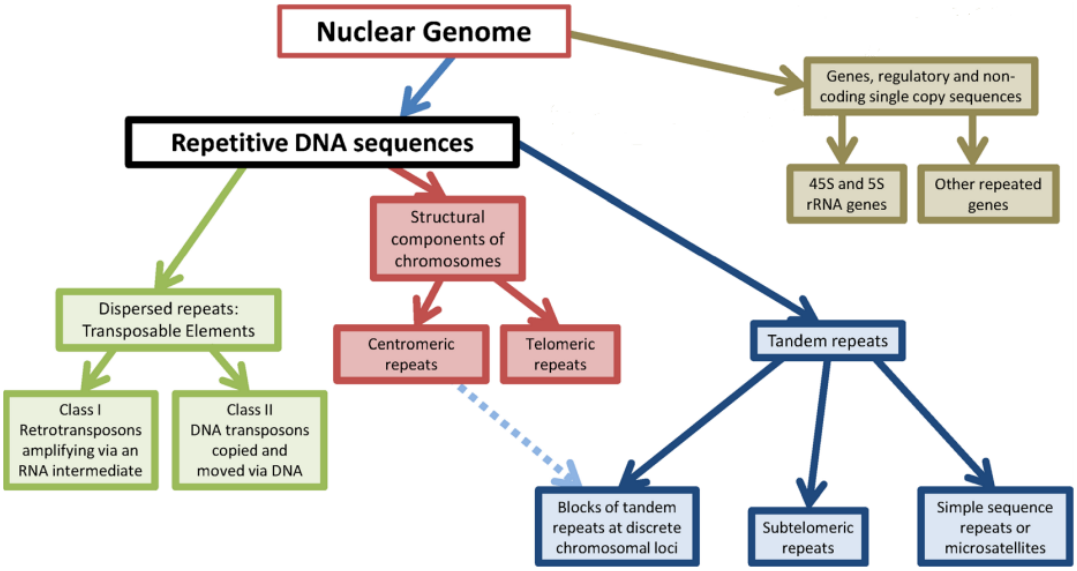


Figure 1: Scheme of major DNA components of the plant nuclear genome and their relationship (Biscotti *et al.* 2015).

2.1.1.1 Eukaryotic genes

In molecular terms, a gene can be defined as a segment of DNA that is the basic functional unit of inheritance controlling the transmission and expression of functional product, which may be either a polypeptide or a noncoding RNA (e.g., rRNA, tRNA). Genes can mutate in their sequences and form different variants, known as alleles. The variation in alleles of genes is necessary to maintain diversity among individuals and species. Most of the eukaryotic genes are present in the nucleus and are responsible for the existence of organisms (Colby 1996; Schlegel 2010).

Gene structure and gene expression in eukaryotes are more complex than in prokaryotes. Eukaryotic genes (Figure 2) consist of coding sequences (exons) interrupted by noncoding sequences (introns). Both exons and introns are transcribed into the precursor mRNA (pre-mRNA), thereafter, the introns are discarded through a process known as splicing. The remaining coding segments are ligated to form the messenger RNA (mRNA) strand. Before the mRNA is prepared for export from nucleus, it is modified by the addition of the 7-methylguanosine ‘cap’ to the 5′ end

and a poly(A)tail to the 3′ end. Prior to the protein-coding gene sequence is a promoter consisting of several short regions that together with specific proteins regulate transcription of a gene (Alberts *et al*. 2002). Besides the promoter, there are other sequences controlling gene expression such as enhancers that can be located far away from the open reading frame (ORF) of the gene (Maston *et al*. 2006; Pennacchio *et al*. 2013). Nevertheless, such remote regulation of transcription appears to be rare in plants (Kellogg and Bennetzen 2004). Antagonists of enhancers are silencers that bind proper transcription factors, called repressors, supressing the transcription (Maston *et al*. 2006). The promotor region is followed by non-coding sequence called 5′ untranslated region (5′ UTR). The coding part of the gene starts with the codon AUG marking the transcription initiation point which is followed by introns and exons. The gene is ended by the terminating exons which contain a non-coding stop codon that is followed by the 3′ untranslated region known as 3′ UTR (Alberts *et al*. 2002; Heslop-Harison and Schmidt 2012). The average length of a gene, including its regulatory components, is about 1-5 kb (Kellogg and Bennetzen, 2004).

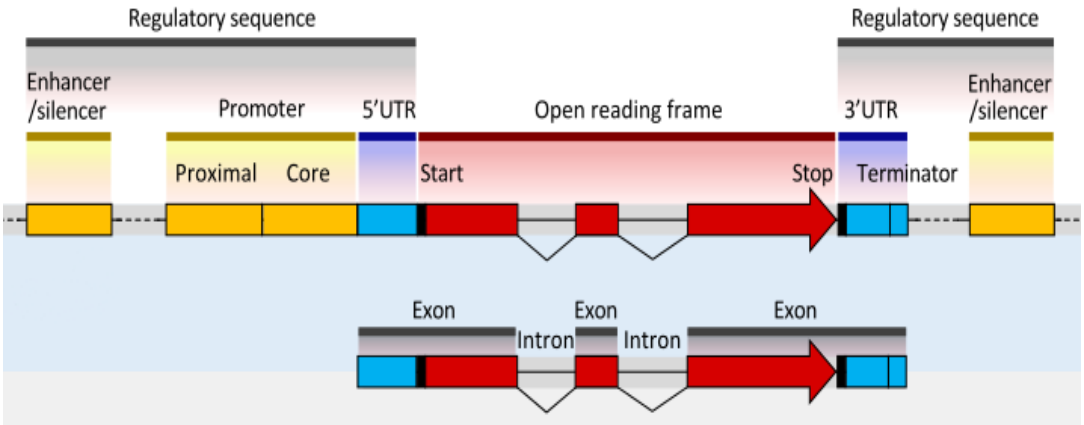


Figure 2: The structure of a typical eukaryotic protein-coding gene (adapted from http://www.wikiwand.com/en/Talk:Gene).

Genomic analyses revealed that most of the genes occur in groups of related genes, called gene families. The gene family can be defined as a group of genes originating from common ancestor. The genes derived from the same ancestor (homologous genes) can be divided, depending on the events taking place during

evolution, into orthologs and paralogs. Orthologous genes are results of speciation of ancestral gene and generally retain a similar function. Orthologs are present in different biological species. On the other hand, paralogs, present in a single organism, have diverged after a duplication event and have obtained a new function (Ohno 1970; Lynch and Force 2000; Armisén *et al*. 2008). Members of the same gene family may cluster within a region of DNA or are dispersed throughout the genome. It has been found that plant genes are relatively often clustered, even in large genomes such as maize and barley. However, the gene clusters in large genomes are constituted with only few genes and each cluster is usually isolated from another by long region of repetitive DNA (Kellogg and Bennetzen 2004; Armisén *et al*. 2008).

Gene duplication realized by various mechanisms results in multiple copies of genes that can obtain a new function. This process is called neo-functionalization (Armisén *et al*. 2008; Wu *et al*. 2011). "Orphan" genes are good example of coding sequences that diverged during evolution and lost homologues in other lineage, even in phylogenetically close relatives (Armisén *et al*. 2008; Tautz and Domazet-Lošo 2011). Sequence analyses of multitude genomes have revealed that orphans typically make up 10% to 30% of genes in all domains of life including viruses (Wissler *et al*. 2013; Arendsee *et al*. 2014). In some cases, multiple copies of genes are needed for production of non-coding RNAs or proteins required in large quantities. For example 45S and 5S rDNA genes coding various types of ribosomal RNA, which are essential components of the ribosomes of all eukaryotes, are present in extraordinarily high numbers. These multiple gene copies are arranged in long tandem arrays located in one or several chromosomal loci (Garcia and Kovařík 2013; Shcherban 2015; Biscotti *et al*. 2015).

The ultimate fate of a duplicated gene can be different. Besides neo-functionalization, the original function of the gene can portioned between both gene copies (sub-functionalization). Similarly, one gene copy can be silenced through inactivating mutations (non-functionalization). Copies of genes that are disabled in such manner are called non-processed or duplicated pseudogenes (Ohno 1970; Lynch and Force 2000). These duplicated genes usually have intact exon-intron structure. Nevertheless, they lost other components, e.g. promoters or enhancers, preventing their activity (Pink *et al*. 2011). In higher eukaryotes, there are also other types of pseudogenes which do not arise by gene duplication including processed or

retrotransposed pseudogenes arisen via the process of retrotransposition (Zhang *et al*. 2003; Pink *et al*. 2011). For very long time, processed pseudogenes were recognized as 'dead on arrival' elements. However, several last studies indicated that they are able to function as noncoding RNA genes (Graur *et al*. 1989; Sakai *et al*. 2007; Pink *et al*. 2011).

The number of genes varies between organisms and changed over time with altered definitions of gene and methods of their detection (Zhang 2002). Whole genome sequencing and gene annotations have shown that many prokaryotes have several thousands of genes and gene content is proportional to the genome size. Eukaryotic genomes, which often have larger size of genome than prokaryotes, contain only an order of magnitude more genes and the number of protein-coding genes does not correlate with their genome size. This phenomenon is known as the C-value paradox and has been commonly observed in higher plants (Thomas 1971; Gall 1981; Moore 1984; Eddy 2012). It is believed that the total number of protein-coding sequences is generally highly similar in various plant species with different genome size. Moreover, determining the proportion of protein-coding sequences can be limited not only by high amount of non-coding DNA but also by more complex regulation of gene expression and presence of alternative variants of the same genes (Mehrotra and Goyal 2014). Before completion of the Human Genome Project in 2001 (Venter *et al*. 2001), scientists guessed about 30,000 to 100,000 protein-coding genes (Pray 2008). However, the sequencing has shown that the total expressed gene number is significantly lower and the number of genes in the human genome is estimated from 20,000 to 25,000 (Schuler *et al*. 1996; Pray 2008; Pertea and Salzberg 2010). Ezkurdia *et al*. (2014) revised the previous estimates and suggested around 19,000 protein-coding genes in the human genome. Sequence analysis of flowering plant *Arabidopsis thaliana* (25,500 - 26,500 genes) showed that a majority of the genome is found in duplicated segments, which are the results of numerous whole, segmental and local duplications, and only 35% of the genes are present in single copies (AGI 2000; Knappe *et al*. 2003; Armisén *et al*. 2008; Krebs *et al*. 2013). The whole-genome shotgun sequencing of rice (*Oryza sativa* L.), which is about three times larger than Arabidopsis (150 Mb) and is the smallest of the cereal crops, revealed around 35,700 protein coding-genes (Yu *et al*. 2002; Ensembl Plants 2016). Nevertheless, most important crops have genomes much larger and complex than the

model organisms. For example sequencing of bread wheat revealed a high level of elasticity and a changed gene composition in all wheat subgenomes compare to their diploid ancestors (Staňková 2015). Sequence analysis of wheat chromosomes revealed that about 24% of genes are duplicated. The study of 3B chromosome pseudomolecule demonstrated even 37% of gene duplicates. The total number of predicted protein-coding wheat sequences is variable among studies and generally is estimated around 100,000 (Choulet *et al*. 2014; IWGSC 2014; Staňková 2015).

2.1.1.2 Repetitive DNA in eukaryotic genome

In fact, only a very small fraction of the eukaryotic genome is responsible for coding proteins. A substantial portion of nuclear DNA in most eukaryotic species consists of highly repetitive sequences (Figure 3). The repetitive DNA sequences have been accumulated into the genomes during evolution and cause variation in genome size among organisms (Pearce *et al*. 1996; Heslop-Harison and Schmidt 2012; Biscotti *et al*. 2015). During the last few decades, many repetitive DNA have been studied to gain more information on their genuine role in eukaryotic genomes. Nowadays, it is known that repetitive sequences are essential in numerous processes such as stabilization and maintenance of the chromosome integrity, their movement, pairing and recombination, karyotypic differentiation and evolution (Mehrotra and Goyal 2014). Moreover, they are involved in modifications of histone proteins, DNA or chromatin and, hence, genome regulation (Heslop-Harison and Schmidt 2012). Repetitive DNA is also very changeable in sequence composition and copy number during time. These structural changes increase the diversity and divergence of genomes and lead to the evolution of species (Cuadrado and Jouve 2002; Sýkorová *et al*. 2003; Heslop-Harison and Schmidt 2012; Mehrotra and Goyal 2014).

Figure 3: General distribution of repetitive sequences on a plant chromosome. Red, centromeric tandem repeats; blue, telomeric repeats; yellow, sub-telomeric tandem repeats; green, intercalary tandem repeats; brown, dispersed repeats; white, genes and low-copy sequences (Mehrotra and Goyal, 2014).

The repetitive DNA consists of sequence motifs with various sizes from single-nucleotide motifs to motifs longer than 10,000 bp that can be repeated hundreds or thousands of times in the genome (Heslop-Harison and Schmidt 2012). A division of the repeats into groups is not easy. Although, the majority of repeat motifs is possible to divide, there are always some intermediate forms which cannot be unequivocally classified. Based on the level of order, the repetitive sequences can be divided into two main groups: tandem repeats and dispersed repeats. In both cases, sequence motifs are located at heterochromatic regions, which are found mostly in chromosome domains such as centromeres and telomeres but also at intercalary positions (Mehrotra and Goyal 2014).

Tandem repeats are arranged one by one in monotonous arrays and according to the length of the repeated unit can be distinguished as microsatellite, minisatellite and satellite sequences. Microsatellites, or SSRs (simple sequence repeats) are composed of very short repetitive unit (1-5 bp) forming blocks around 150 bp. The SSRs are often used in plant genetics as cytogenetic and molecular markers for their high frequency in the eukaryotic genomes (Ramel 1997; Heslop-Harison and Schmidt 2012). Minisatellites have longer DNA motifs (mostly about 6-25 bp) creating blocks in length up to 30 kb. The number of minisatellite loci is estimated in the thousands in the human genome and they are remarkable for their extreme polymorphism and high mutation rate (Ramel 1997; Denoeud *et al*. 2003). Satellite DNA (satDNA) consists of a series of monomers longer than 25 bp that often form arrays up to 100 Mb. Satellite sequences are present in genetically inactive heterochromatin regions and represent a fast-evolving part of the eukaryotic genome (Ugarkovic 2005; Mehrotra and Goyal 2014; Shcherban 2015; Biscotti *et al*. 2015).

In contrast to tandemly organized repeats, dispersed repetitive sequences are scattered throughout the genome. The main sources of dispersed DNA repeats are transposable elements, also known as mobile genetic elements (Heslop-Harison and Schmidt 2012). Transposable elements can be divided into two main classes, class I and class II, differed from each other by the way of their transposition. Transposable elements of class I or retrotransposons (RNA transposons) spread by the mechanism "copy-and-paste" through an RNA intermediate, while transposable elements class II (DNA transposons) use "cut-and-paste" mechanism via a double-stranded DNA break. According to the presence and absence of long terminal repeats (LTRs), RNA

transposons are usually grouped into LTR retrotransposons with two main superfamilies *Gypsy* and *Copia* and non-LTR retrotransposons further classified into long and short interspersed nuclear elements, *LINEs* and *SINEs*, respectively (Wicker *et al*. 2007). Furthermore, the transposition can be classified for both classes as either "autonomous" with transposable elements encoding all proteins necessary for amplification, or "non-autonomous" elements depending, partially or completely, on other mobile genetic elements in the genome (Feschotte and Pritham 2007).

As mentioned above, different types of repetitive motifs are used in molecular genetic mapping of eukaryotic genomes that often contain very large amount of repetitive sequences (Pestsova *et al*. 2000). Mainly the microsatellites are highly abundant in genomes of all eukaryotes and can be widely dispersed or confined only to particular regions of the chromosomes. Furthermore, they have a high degree of length polymorphism and the possibility of a simple amplification by PCR. For different types of molecular cytogenetic analysis, di- (GC, AG, AC, AT), tri- (AAC, AAG, AGG, CAT) and tetranucleotide (GACA, GATA, GGAT) repeat motifs of synthetic SSRs are highly applied in the grass tribe *Triticeae*. Thanks to these properties and the increasing accessibility of genomic sequences in DNA databases, these markers are often used for study of genetic variation, genetic mapping, identification of chromosomes, characterization of plant genomes and in breeding (Cuadrado and Jouve 2007; Cuadrado *et al*. 2008a, 2008b; Mehrotra and Goyal 2014).

2.1.2 Chromatin

The genetic information is stored in the form of chromatin in the eukaryotic nucleus. Chromatin is highly dynamic and functionally organized structure constituted by complex of DNA, RNA, variety of histones and non-histone proteins forming the first structural level of DNA organization, a fibre known as nucleosome (e.g., Soutoglou and Misteli 2007; Woodcock and Ghosh 2010). This 11 nm fibre (Figure 4) is comprised of 147 bp of DNA wrapped around a histone octamer consists of two copies of the four histones H2A, H2B, H3 and H4 (Olins and Olins 1974; Kornberg 1974; Olins and Olins 2003). Each of these histone proteins is formed by structured core and unstructured amino(N)-terminal tail domain which can protrude

from their own nucleosome to other nucleosome. Only the histone H2A has a long C-terminal tail with a large interface with the histones H3 and H4 (Biswas *et al*. 2011). This compact form of the nucleosome is unwrapped during replication and gene expression of DNA. Histone H1 (H5) asymmetrically binds to linker DNA between nucleosomes and it is supposed to play a key role in chromatin condensation and formation of higher-order structures (Kouzarides 2007; Happel and Doenecke 2009; Rossetto *et al*. 2012).

Chromatin is accepted to be the basic unit of the genomic DNA organization (Olins and Olins 2003). Finch and Klug (1976) first observed 30-nm chromatin fibre called solenoid that is typically postulated as being the second structural level of DNA organization (Robinson *et al*. 2006). However, Maeshima *et al*. (2014) demonstrated that chromatin consists of dynamic and disordered 11 nm fibres, without 30 nm chromatin structure. These fibres are further organized into chromatin loops (250 nm) that connect genes and enhancers to large chromosomal domains (700 nm) and nuclear compartments. The chromatin fibres are functional subunit of chromosomes and dynamic structure regulating gene expression (Gibcus and Dekker 2013).

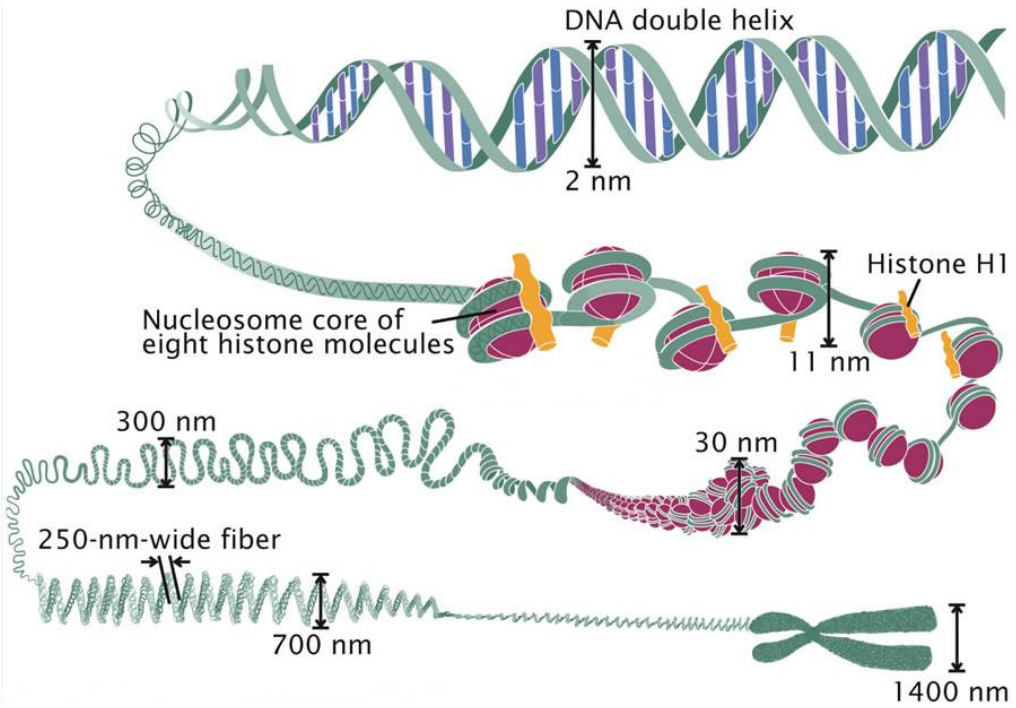


Figure 4: Packing the nuclear genome: from a DNA molecule to a mitotic chromosome (Pierce 2005).

Depending on level of compaction, chromatin is classified as euchromatin or heterochromatin. Euchromatin is an uncoiled form of chromatin that contains most of the single-copy DNA and is often transcriptionally active. Conversely, gene-poor heterochromatin is typically highly condensed and usually transcriptionally silent (Huisinga *et al*. 2006). Heterochromatin is often localized at the periphery of the nucleus and can be divided into two main categories: constitutive and facultative (Brown 1966). Constitutive heterochromatin is permanently condensed and transcriptionally inert. This type is often found around centromeres and telomeres and is important for correct chromosome segregation during cell division (Schueler and Sullivan 2006; Eymery *et al*. 2009). Facultative heterochromatin is viewed as a less static structure than constitutive heterochromatin formed at different chromosomal regions and becoming condensed or decondensed at some stage during development (Gilbert *et al*. 2003; Wegel and Shaw 2005).

Chromatin is very dynamic system which undergoes various structural changes during cell cycle (e.g., Woodcock and Ghosh 2010; Doenecke 2014). The dynamics of high-order chromatin structures play a key role in regulation of gene expression and other biological processes associated with DNA (Maeshima *et al*. 2014). The structure of chromatin is mainly regulated by histone proteins that pack and arrange the fibre and can be modified by many various post-translational modifications (PTMs) of histones that directly or indirectly influence altering of chromatin packing (Happel and Doenecke 2009; Rossetto *et al*. 2012). PTMs can occur in any histone, but they are largely concentrated in H3 and H4 histones with N-terminal tail domains sticking out from the nucleosome that are liable to PTMs. The cores of the histones H2A, H2B or H3 can also be post-transcriptionally modified (e.g., Bannister and Kouzarides 2011; Zhang *et al*. 2012; Christophorou *et al*. 2014). The most frequent and well characterized mechanisms that modify chromatin structure are acetylation, methylation, phosphorylation and ubiquitination.

Acetylation of histones is dependent on the action of enzyme called histone acetyl transferase (HAT). This enzyme suppresses the tendency of chromatin fibre to fold up into compact structure. Acetylated chromatin is more accessible to transcription apparatus. In contrast, deacetylation of histones, which is regulated by histone deacetylase (HDAC), leads to the repression of transcription (Sterner and Berger 2000; Chen *et al*. 2001; Eberharter and Becker 2002). Other histone

modification that is capable to activate or deactivate transcription of genes is methylation of histones. The most common methylation takes place on a specific lysine (K) that can exist in a mono-, di-, or tri-methylated state when trimethylation of H3 at lysine 4 (H3K4me3) activates transcription and demethylation of histone H3 at lysine 9 (H3K9me2) silences transcription of genes (Lan and Shi 2009; Gupta *et al*. 2010). Phosphorylation of histones is also highly dynamic process regulated by a number of protein kinases and phosphatases which phosphorylate or dephosphorylate the histone tail of all nucleosomal histones. The majority of histone phosphorylation occurred predominantly within the N-terminal tails of serines, threonines and tyrosines (Bannister and Kouzarides 2011; Rossetto *et al*. 2012). Phosphorylation of H2A(X) at serine 139 is the other important histone modification which takes place during cellular response to DNA damage and can be used as marker for DNA double strand breaks (Rogakou *et al*. 1998; Rossetto *et al*. 2012). Another example is phosphorylation of H3 at serine 10 that plays essential role during condensation and segregation of chromosomes in mitosis and meiosis (Wei *et al*. 1999; de la Barre *et al*. 2000). The phosphorylation is also associated with regulation of gene expression and apoptosis (Yeo *et al*. 2012). Acetylation, methylation and phosphorylation cause relatively small molecular changes on amino-acid side chains. In contrast, another important mechanism of PTMs is ubiquitylation that causes much larger covalent modification of histones (Bannister and Kouzarides 2011). During ubiquitylation, the protein of ubiquitin is linked to specific lysine residues via the sequential action of three enzymes of target proteins (E1-activating, E2-conjugating and E3-ligating enzymes). This reversible enzymatic PTM has a critical regulation function in a wide variety of cellular processes such as repression or activation of transcription, DNA repair, biogenesis of organelles, cell cycle and apoptosis (Finley *et al*. 2012; Duplan and Rivas 2014; Hongyong *et al*. 2014).

This large number of different types of histone modifications ensures one of the distinguished controls of chromatin structure. However, the overall regulation of chromatin is provided via complexity of cross-talks between various modifications and can occur by multiple mechanisms. PTMs are regulated by many chromatin-associated factors that interact with modified histones through many distinct and specific domains (Bartke *et al*. 2010; Bannister and Kouzarides 2011).

2.1.3 Organization of chromatin in eukaryotic genome

The overall structure and functional organization of chromatin depends on the stage of the cell cycle when chromatin undergoes various structural changes (Davis and Bardeen 2004). During interphase, the chromatin occurs in a relaxed state to allow access to polymerases that transcribe and replicate the DNA (Bannister and Kouzarides 2011). However, in the early stages of cell division, mitosis and meiosis, the chromatin fibre become more condensed and is arranged into highly organized structures called chromosomes.

2.1.3.1 Chromosomes

The term chromosome comes from the Greek words for color (chroma) and body (soma). Chromosomes were first observed by researchers during 19th century. Nevertheless, the connection of chromosomes with Mendel´s laws was discovered by Theodor Boveri and Walter Sutton in 1902 (Martins 1999). Each chromosome consists of a single enormously long linear DNA molecule, which is coiled tightly around proteins (Alberts 2002).

Chromosomes can be different in size and structure but their basic morphology remains the same in eukaryotes (Cooper 2000). Most chromosomes have centromeric region which divides chromosome into two arms – the shorter arm (p, S) and the longer arm (q, L). For primary characterization of chromosomes is very often used so called centromere index (CI), which is the ratio of short arm length to the total length of a chromosome. Based on the CI, there are four types of chromosomes: metacentric, submetacentric, acrocentric and telocentric (Figure 5). The centromere is important for correct segregation of chromosomes during the cell division and is the place for connection of protein complex which is called kinetochore. Two protein complexes, condensin and cohesin, regulate condensation of chromosome and sister chromatid cohesion, respectively. The condensin has the ability to reorganise chromosomes into highly compact structure, whereas the cohesion complex keeps replicated sister chromatids together until their separation at anaphase (Koshland and Strunnikov 1996; Kimura *et al*. 1999, 2001). The separation of sister chromatids is ensured by proteolytic enzyme separin, also known as separase. The complete division of chromatids is hampered by enzyme securine

that inactivate separine until the beginning of anaphase (Ciosk *et al*. 1998; Uhlmann *et al*. 1999).

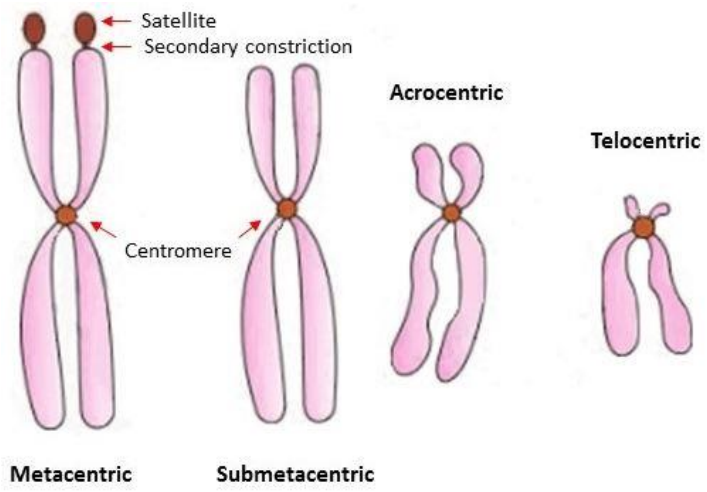


Figure 5: Types of chromosomes based on the position of centromere (adapted from http://shobhanapathak.blogspot.cz/2015/08/ncert-solutions-for-class-11th-ch-8.html).

Some chromosomes have a secondary constriction on the arms where genes for organization of nucleolus are presented. The satellite lies behind secondary construction. Each arm of eukaryotic chromosome is terminated by telomere. The telomere is a specific nucleoprotein repetitive region protecting the ends of chromosome against a shorting during replication and interacts with nuclear lamina. The morphology of chromosomes is defined by position of primary constriction, attendance and position of secondary constriction, length of chromosomes and other morphological characteristics such as knobs or chromomeres. In summary, all characteristics of chromosomes such as their appearance, number and function is expressed as karyotype constituted the complete set of chromosomes typical for a species or an individual organism.

In diploid eukaryote cells, the chromosomes are present in pairs of homologous chromosomes when one of each chromosome pair originating from the maternal and one from the paternal gamete. The diploid chromosome number is referred to as 2n. Whereas, in gametes is each chromosome only in one copy and these cells are haploid (n) with the haploid number of chromosomes. According to the number of chromosome sets, cells are monoploid (one set), diploid (two sets),

triploid (three sets), tetraploid (four sets), pentaploid (five sets), hexaploid (six sets), etc. It is possible that the ploidy level may increase in the germline, which can result in the origination of polyploid organism (Griffiths *et al*. 1999). The number of chromosomes in the ancestral set is known as monoploid number (x) and differs from the haploid number (n). For example, the nuclear genome of *Triticum aestivum* L. (2n = 6x = 42) consists of six sets of chromosomes in somatic cells, two sets from each of three different diploid species that are distant ancestors of wheat.

It is known that chromosomes are mobile structures of the cell nucleus. Prophase of the first division of meiosis is a period of dynamic chromosome behaviour (Tiang *et al*. 2012). During this phase, homologous chromosomes interact, pair and exchange genetic material (Figure 6; Alberts 2002). Chromosome pairing is process in which two homologous chromosomes find each other among all chromosomes in the nucleus and juxtapose (Tiang *et al*. 2012). The paired chromosomes are called bivalent or tetrad. At this point nonsister chromatids of homologous chromosomes may exchange some segments. The exchange between chromatids results in recombination of genetic information (Alberts 2002; Tiang *et al*. 2012). Recombination is the fundamental process essential for the production of new combination of alleles in population (e.g., Loeve and Hill 2010; Lobkovsky *et al*. 2015; Wilson *et al*. 2016). There are recombination hot spots on chromosomes, which exhibit elevated frequencies of crossovers. In wheat, barley, maize, and other members of *Poaceae*, recombination sites have been observed predominantly in distal regions of the chromosomes (e.g., Saintenac *et al*. 2008; Phillips *et al*. 2013). Phillips *et al*. (2015) observed that the sites of crossovers are malleable and can be shifted to proximal regions by elevating growth temperature.

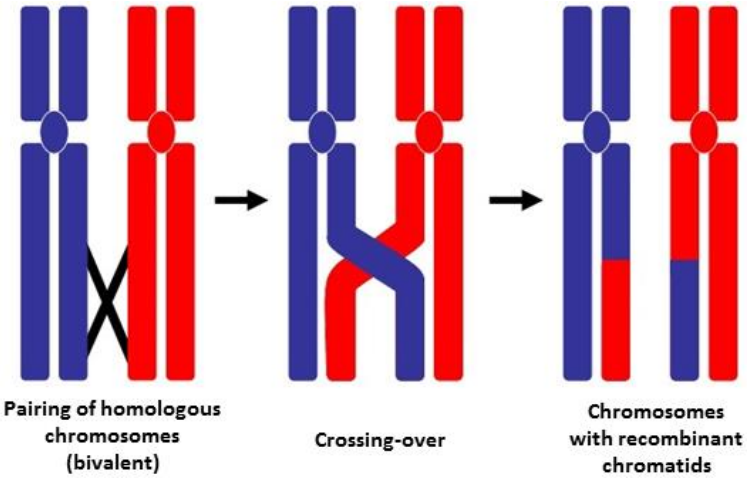


Figure 6: Chromosome dynamics in meiotic prophase 1 (adapted from http://slideplayer.cz/slide/2424733/).

Behaviour of chromosomes during pairing is under genetic control. However, it can be changed to induce homoeologous recombination between chromosomes that normally do not pair (Zhang *et al*. 2015). For instance, meiotic pairing between homoeologous chromosomes of hexaploid wheat is controlled by two genes (Sears 1976). The major locus *Ph1* (*Pairing homoeologous 1*) supressing homoeologous pairing in wheat is located on chromosome arm 5BL (Sears and Okamoto 1958; Riley and Chapman 1958). Later on, *Ph2* has been identified to play a role in homoeologous pairing recombination (Mello-Sampayo and Canas 1973). In wheat lines without chromosome 5B or possessing its mutant form (*ph1b*), chromosomes may pair with their wheat homoeologues and also with chromosomes originated from other species in the tribe *Triticeae* in case of interspecific or intergeneric hybrids (Riley and Chapman 1958; Dyck *et al*. 1990; Zhang *et al*. 2015).

2.1.3.2 Chromosome organization in the interphase nucleus

Organization of chromatin in interphase nuclei has been a subject of many speculations for several decades. With the development of molecular cytogenetics and microscopy methods, our knowledge about this topic has been significantly increased (Tiang *et al*. 2012). During last years, many studies on chromatin organization have been done also in plants. To the most important discoveries belong elucidating the interphase chromatin organization in Arabidopsis (e.g., Fransz *et al*. 2002; Pecinka *et al*. 2004; Schubert *et al*. 2014). However, there are still many

unanswered questions about the effect of the three dimensional folding of chromatin in the interphase nucleus on the regulation of gene expression, nuclear packaging or chromosomal dynamics (Kellogg and Bennetzen 2004).

During interphase, eukaryotic chromatin is in decondensed state and appears loosely distributed in the cell nucleus (Tiang *et al*. 2012). Nevertheless, several studies revealed (e.g., Cremer and Cremer 2001; Lysak *et al*. 2001; Pecinka *et al*. 2004) that each chromosome occupies a distinct sub-nuclear volume, known as chromosome territory (CT). For the first time, CTs were visualized by fluorescence *in situ* hybridization (FISH) using probe sets designed to paint entire mammalian interphase chromosomes, giving rise to the term "chromosome painting" (Bolzer *et al*. 2005; Gorkin *et al*. 2014). Cytological studies in different species have revealed that territorial arrangement of chromosomes is a general feature of eukaryotic interphase nuclei and is common in plants and animals (Leitch 2000; Pecinka *et al*. 2004; Cremer and Cremer 2010; Schwartz and Hakim 2014). CTs are spatially distinct with considerable intermingling between different chromosomes near the borders of CTs (Gorkin *et al*. 2014; Guo and Fang 2014). Due to the current technical advances in cytogenetic and genomic techniques, we can study genome architecture on three-dimensional level. The techniques of three dimensional fluorescence *in situ* hybridization (3D-FISH) and various modifications of chromatin conformation capture (3C, 4C, 5C and HiC) enabled the development of complex approaches for studies of the spatial arrangement of whole genomes, chromosomes, repetitive DNA sequences and genes within interphase nuclei (Dekker *et al*. 2002; Zhao *et al*. 2006; Dostie *et al*. 2006; van Berkum *et al*. 2010; Sequeira-Mendes and Gutierrez 2016).

Cytogenetic studies have indicated that larger chromosomes are located toward the periphery and smaller chromosomes are distributed significantly closer to the centre of the nucleus (e.g., Sun *et al*. 2000; Grob *et al*. 2014). Gene density can also influence the position of CTs, when gene-rich chromosomes occupy more central position of the nucleus and gene-poor chromosomes are located at the nuclear periphery (Cremer and Cremer 2001; Bolzer *et al*. 2005; Ferrai *et al*. 2010). Gene-rich CTs have irregular shape and gene-poor CTs appear to have more regular ellipsoidal shape (Sehgal 2014). The position of specific regions in CT are non-random and their movement in domains often correlate with their transcriptional activity (Sutherland and Bickmore 2009; Gorkin *et al*. 2014; Schwartz and Hakim

2014). Sanborn *et al*. (2015) proved that it is possible to influence the shift of chromatin loops and domains using targeted mutations as small as a single base pair.

In plants with large genomes such as wheat, rye and barley, CTs tend to display so called Rabl´s configuration (Figure 7; Rabl 1885) with centromeres and telomeres clustered at the opposite poles of cell nuclei (Pecinka *et al*. 2004; Fransz and de Jong 2011). The Rabl´s configuration is presumably a relic of anaphase-telophase arrangement and may probably function to maintain a certain level of chromosome order and integrity. However, Rabl´s configuration is not a general feature of large genomes as demonstrated in maize and sorghum (Dong and Jiang 1998). In small genome species, such as Arabidopsis, CTs do not exhibit the Rabl´s configuration, but have a different spatial distribution (Pecinka *et al*. 2004; Schubert and Shaw 2011). However, rather small genome of rice can be organized in Rabl´s and non-Rabl´s configurations depending on cell type (Santos and Shaw 2004). Interestingly, it seems that there is a way to induce Rabl´s configuration by DNA demethylation in those tissues of rice where Rabl´s configuration is normally absent (Santos *et al*. 2011).

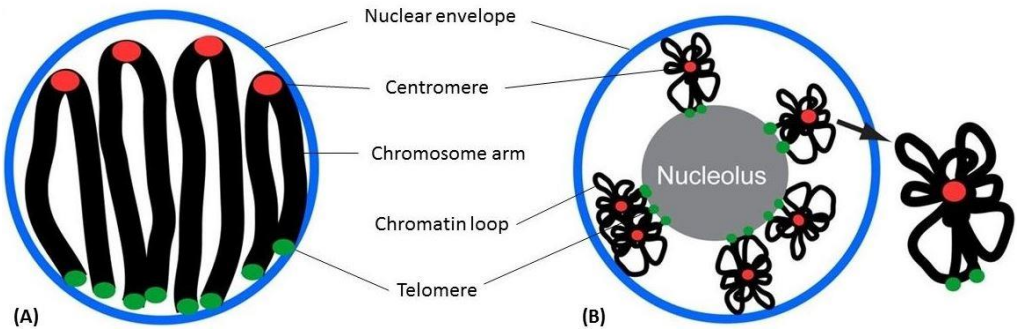


Figure 7: Patterns for chromosome arrangement in interphase nucleus. (A) Rabl configuration found in plants with large genome. (B) Rosette-like organization typical for small-genome plant species (adapted from Tiang *et al*. 2012).

Most of the current knowledge on nuclear architecture in plants comes from studies on model species – *Arabidopsis thaliana*. As mentioned above, its chromosomes do not exhibit Rabl´s configuration. Centromeres are randomly distributed at the nuclear periphery, while telomeres are gathered around the nucleolus (Dong and Jiang 1998). Centromeric heterochromatin forms distinct

chromocenters, where the majority of genes are located, while euchromatin is formed into loops getting out from the chromocenters. This arrangement of chromosomes is known as "rossete structure" (Fransz *et al*. 2002; Pecinka *et al*. 2004; Tiang *et al*. 2012; Schubert *et al*. 2014). Position of CTs and arrangement of heterochromatin domains is mostly random in differentiated as well as in meristematic tissues, except for the chromosomes bearing nucleolar organizing regions (NORs) associated with nucleolus (Pecinka *et al*. 2004; Berr and Schubert 2007).

The increase of our knowledge about 3D organization of chromatin in interphase nucleus would be highly valuable not only for understanding the function and evolution of plant genomes, their organization within the 3D space of nuclei and interaction of parental genomes after interspecific hybridization, but also for plant breeders who wish to introduce alien chromatin by interspecific hybridization (Schubert and Shaw 2011; Rey *et al*. 2015).

2.2 Interspecific hybridization and polyploidy as tools in plant breeding

Interspecific hybridization and polyploidy are driving forces in evolution, speciation and domestication of flowering plants (Levin 2002). In plant breeding, these processes go hand in hand and can be observed in the breeding history of many crops (Soltis and Soltis 2009; Weiss-Schneeweiss *et al*. 2013). Nevertheless, the details of changes that occur during and after hybridization and polyploidy are still poorly understood (López-Caamal and Tovar-Sánchez 2014). In the next chapters, processes of interspecific hybridization and polyploidy will be discussed in details.

2.2.1 Plant hybridization

Interspecific hybridization is a process of merging genetic material originated from two different biological species. Crossing may occur either at the same (homoploid hybridization), or various ploidal level (polyploid hybridization). Both types of hybridization can be potential genetic source for development of new species. It was found that interspecific hybridization is common among plant species (Soltis and Soltis 2009; Yakimowski and Rieseberg 2014).

Nevertheless, merging of two genetically different genomes may be limited by series of factors (or barriers) that can be divided into several groups. The first category is represented by morphological barriers where various structures of flowers do not allow successful pollination and thus, hamper interspecific hybridization. The forming of a new polyploid is further very often restricted by incompatibility of pollen with foreign stigma (physiological barriers). Disharmony in flowering time (phenological barriers) and other environmental factors, such as climate conditions, a habitat disturbance, and amount of nutrition play an important role in successful hybridization. Last but not least, cytological and genetic differences between two hybridized species, e.g. various number of chromosomes and chromosome sets, homology of genomes, constitute the next important factor affected fusion of their gametes and/or fertility of potential hybrids (Chen 2007; Soltis and Soltis 2009).

Cytological and genetic diversity between parents can lead to the formation of univalents or multivalents during meiotic chromosome pairing, and thus, production of unbalanced gametes resulting in a complete or partial sterility of hybrids. Nevertheless, the hybrids can restore the fertility by repeated backcrossing with one of the parental species or by chromosome doubling (Comai 2005; Soltis and Soltis 2009).

After successful crossing of two genetically different organisms, extensive modification of the genome and transcriptome occurs including novel expression patterns and regulatory interactions. The gene expression is massively altered immediately after the hybridization and is followed by long-term modifications of gene expression in subsequent generations (Flagel *et al*. 2008). This is probably responsible for the changes in genome constitution of successive generations of hybrids, especially in those without chromosome pairing control system. One of the genome frequently takes a lead and the other genome is either immediately eliminated (e.g. *Hordeum vulgare* x *H. bulbosum*) or continuously decreased. This is a case of *Festuca* x *Lolium* hybrids, where *Lolium* genome predominates in subsequent generations and *Festuca* genome is slowly, but continuously, decreasing (Zwierzykowski *et al*. 2006). Thus, we can speculate about the dominance of one genome over the other. On the other hand, interspecific hybrids with a presence of some chromosome pairing control system, such as *Ph1* in wheat preclude homoeologous pairing and recombination (either in hexaploid bread wheat or in

hybrids where wheat is one of the parent) and thus, keep both parental genomes separated in distinct parts of the nucleus.

Hybridization between individuals can result in interspecific gene flow and generate new hybrid species which have higher genetic diversity than their parents (Barton 1979; Wallace *et al*. 2011). Nevertheless, there are a number of factors preventing exchange of genetic material among species in nature and the gene-flow is therefore restricted (Urbanelli 2002). In artificial conditions, it is possible to overcome such barriers and utilize interspecific hybridization for targeted breeding (Kaneko and Bang 2014).

2.2.2 Polyploidy

Polyploidy or whole genome duplication (WGD) was initially described as evolutionary dead-end (Stebbins 1950), but nowadays it is viewed that this mechanism is an important process in evolution. Recent genomic studies indicate that probably all angiosperms, including crops (e.g. wheat, rice, maize, soybean, potato, sugarcane) have undergone at least one round of the genome doubling (Ramsey and Schemske 2002; Soltis and Soltis 2009; Weiss-Schneeweiss *et al*. 2013). The classification of polyploids has been long time debated. Kihara and Oho (1926) divided polyploids into two main groups: autopolyploids and allopolyploids. Autopolyploid can be defined as a polyploid with homologous chromosome sets derived from single species, while allopolyploid is polyploid with two or more sets of genetically diverse, but usually closely related, chromosomes obtained from different species.

Generally, polyploids can arise spontaneously by two main ways: a) by the fusion of unreduced gametes (sexual polyploidization) which have more than one set of chromosomes and are formed during rare meiotic failures, or b) by somatic genome doubling (somatic polyploidization) that may occur in zygotic, embryotic or meristematic tissue (Ramsey and Schemske 2002; McKain *et al*. 2016). In plants, formation via the fusion of unreduced gametes has been observed more frequently than somatic duplication. Polyploidy can be also induced artificially using natural (colchicine) or synthetic compounds (oryzalin, trifluralin). Their effectiveness

depends highly on the concentration applied, duration of treatment, and type of plant material (Ascough *et al*. 2008).

Although duplication of genomes is the shared property for auto- and allopolyploidy, the differences in their heredity and chromosome compositions have significant consequences. Autopolyploids generally exhibit polysomic inheritance and formation of multivalents from homologous chromosomes, while allopolyploids (e.g. wheat, tobacco) have usually disomic inheritance and homologous bivalent pairing is prevalent, resulting largely in the maintenance of two parental separated genomes in allopolyploid nuclei (Comai 2005; Chen 2007). Nevertheless, if the homoeologous chromosomes of allopolyploids have segments that are homologous, pairing may occur between the homoeologous chromosomes derived from different species, especially in the absence of chromosome pairing control system. This pairing is then connected with formation of multivalents and such polyploids are called segmental allopolyploids (Stebbins 1950; Soltis and Soltis 2009). Therefore, categorization of polyploids into two types is not ambiguous and the third category should be accepted.

2.2.3 Alterations associated with polyploidy

The successful merging of two diverse genomes and formation of allopolyploids is accompanied by extensive changes at the nuclear and cellular levels. Barbora McClintock (1984) believed that genomic incompatibilities following interspecific hybridization are among the causes of "genomic shock" which is a response to the extensive stress. Forming new genomes of polyploids activates repetitive elements and dynamic changes in genome size, genome structure and epigenetic control may occur. These changes reflecting genomic and functional plasticity of duplicate genes and genomes can lead to genotypic and phenotypic differentiation of polyploid organisms (Comai 2005; Soltis and Soltis 2009; Weiss-Schneeweiss *et al*. 2013).

One of the changes, which occur after hybridization and polyploidy, are changes in epigenetic regulation. The epigenetic mechanisms, such as DNA methylation, histone modifications or RNA interference, may influence activity of transposable elements, reprogram gene expression profile (gene activation/silencing)

and developmental patterns of new allopolyploids (Chen 2010; Madlung 2013; De Storme and Mason 2014). The severe stress induced by the interspecific hybridization and polyploidy is also accompanied by the extensive genetic and genomic changes such as various types of mutations, chromosomal rearrangements or DNA elimination. These processes lead to fundamental differences between individuals with varying degrees of ploidy (te Beest *et al*. 2012; Weiss-Schneeweiss *et al*. 2013).

Polyploidy is often associated with larger cells, because polyploid cells have to accommodate larger genomes (te Beest *et al*. 2012). The increasing cell size, in response to higher amount of genetic material, is accompanied by changes in cellular architecture such as disruption of relationship among the cellular components. For example altered interactions between heterochromatin and nuclear lamina, which form a dense fibrillar network that is associated with the inner membrane of the nuclear envelope may have a negative impact on cell vigour (Otto and Whiton 2000; Knight *et al*. 2005). Additionally, larger genomes need more time to replicate, which result in decreased cell growth rate (te Beest *et al*. 2012). These changes may potentially slow down metabolism and development, affect reproductive success, generation time, seed mass, ecological strategy and the type of habitats of polyploid species (Comai 2005; te Beest *et al*. 2012; Madlung 2013).

On the other hand, such extensive changes may confer key advantages to polyploids and ensure their evolutionary success. One of the most significant benefits of allopolyploidy is a phenomenon known as heterosis, or hybrid vigour, which is a result of genome-wide changes and interactions between different species (Chen 2010). Higher phenotypic plasticity of polyploids associated with increased number of alleles may increase fitness traits such as growth rate, reproductive output and biomass (te Beest *et al*. 2012). Effect of heterosis is used in plant breeding for a very long time; however, this phenomenon is still poorly explored. The improved understanding of the molecular regulation will help us to better exploit heterosis in modern breeding and agriculture (te Beest *et al*. 2012, Madlung 2013).

The complexity of epigenetic, genetic and genomic changes associated with hybridization and polyploidy modify morphological and physiological characteristics of plant species (te Beest *et al*. 2012). As mentioned above, polyploid

individuals usually have a larger cell size which in turn leads to enlarged plant organs. This phenomenon is called gigas effect (Figure 8; Ramsey and Schemske 2002; Saminathan *et al*. 2015). Polyploids are usually taller and more robust, with larger leaves, flowers and seeds than their diploid ancestors. Larger seed size often results in more robust seedlings which may have a competitive advantage compare to diploids. The type of flowers and flowering phenology are also substantially influenced by polyploidy. In general, flowers of polyploids are very often cleistogamic, allowing for more efficient self-fertilization, with prolonged or later period of flowering (te Beest *et al*. 2012). The slower metabolism caused by an increased cell size gives to polyploids the advantage of greater longevity and consequently shift from annual to perennial life cycle very often observed in polyploids (Ramsey and Schemske 2002; te Beest *et al*. 2012). Lower number but larger size of stomata per unit of leaf present in polyploids reduces transpiration rates and enables more efficient water management and better adaptation to dry conditions (Levin 2002; te Beest *et al*. 2012). It has been found that polyploid plants have greater tolerance to lower nutrition uptake, cold, drought and salinity, and can be more common at higher elevations and latitudes than diploids (Thompson *et al*. 2004; te Beest *et al*. 2012).

Figure 8: A comparison between the leaf (A) and the flower (B) of a diploid (CLD1) and induced teraploid (CLT1) watermelon illustrating the gigas effect (adapted from Saminathan *et al*. 2015).

2.2.4 Introgression breeding

Interspecific hybridization makes the transfer of a gene pool from one species into another possible. Repeated backcrossing of interspecific hybrids with one of the parental species has also contributed to the evolution and speciation of various species (Liu *et al*. 2014). This process, for which the term introgressive hybridization is now used, is besides its role in plant evolution and speciation also important tool in modern plant breeding (Morgan *et al*. 2011, Liu *et al*. 2014). Introgression of genomic regions from wild relatives into elite crop lines provides plant breeders an opportunity to improve the agricultural performance of commercial varieties (Zamir 2001; Molnár-Láng *et al*. 2015).

Moreover, chromosome(s) or chromosome segment(s) carrying gene(s) coding agriculturally important trait(s) such as higher tolerance to biotic and abiotic stresses may be transferred by series of backcrosses from wild relatives into crops (Anamthawat-Jónsson 2001; Molnár-Láng *et al*. 2014). To date, an extensive number of cases of targeted incorporation of alien segments have been done and many types of introgression lines have been produced (Molnár-Láng *et al*. 2014; Molnár-Láng *et al*. 2015).

For targeted introgression breeding, it is necessary to confirm hybridity of the plants originating from distant crosses (Liu *et al*. 2014). During the twentieth century, high number of chemical, genetic and molecular methods have been developed which have been successfully applied for identification of alien introgression including molecular markers (e.g. RFLP, AFLP, microsatellites) and *in situ* hybridization techniques (GISH, FISH) allowing accurate identification and localization of introgressed chromatin (Anamthawat-Jónsson 2001; Molnár-Láng *et al*. 2015).

With respect to aims of the thesis, the following chapter will be focused on the introgressive hybridization of bread wheat (*Triticum aestivum* L.), one of the major cereal crop species with its wild relatives.

2.2.5 Wide hybridization of wheat

Bread wheat (*Triticum aestivum* L.) belongs to the genus *Triticum*, the tribe *Triticeae*, the subfamily *Pooideae* and the *Poaceae* family (Löve 1984). It is an allohexaploid species (2n = 6x = 42, AABBDD) which arose from two spontaneous

hybridization between three different diploid progenitors classified in the genera *Triticum* and *Aegilops* (Feldman and Levy 2012). The hybridization events are estimated to have taken place around 0,5 million and 10,000 years ago in the region of Middle East (Zamir 2001; Dubcovsky and Dvorak 2007; Feldman and Levy 2012; Choulet *et al*. 2014). The first hybridization occurred between *Triticum urartu* (A genome) and an unknown species related to *Aegilops speltoides* (S genome related to the B genome of wheat). This interspecific hybridization resulted in tetraploid *Triticum turgidum* (AABB genome) which crossed with a wild diploid grass *Aegilops tauschii* (DD genome) resulted in the allohexaploid species *Triticum aestivum* (AABBDD; Feldman and Levy 2012; Choulet *et al*. 2014; Staňková 2015). Each of the three structurally similar (homoeologous) ancestral subgenomes is about 5,5 Gb in size and, therefore, results in a highly redundant 17 Gb/1C genome of bread wheat with more than 80% repetitive sequences (Smith and Flavell 1974; Feldman and Levy 2012; Choulet *et al*. 2014).

The domestication of the wheat has led to the decreasing of genetic variability among and within wheat cultivars. Thus, incorporation of new alleles via interspecific hybridization has been a special goal of the wheat breeders. So far, numerous agriculturally important genes have been incorporated into bread wheat from *Secale*, *Hordeum*, *Aegilops*, *Thinopyrum*, *Triticum*, and other genera of the *Triticeae* tribe (Molnár-Láng *et al.* 2014, Molnár-Láng *et al*. 2015).

2.2.5.1 Introgressions between bread wheat and rye

Genus *Secale* includes *S. cereale* which is widely cultivated species especially due to its tolerance to many abiotic and biotic stresses. Moreover, rye is high yielding in regions with poor soils and in harsher environments. The most famous hybrid of wheat and rye is triticale (Figure 9), the first man-made amphiploid combining entire genomes of diploid rye and either two (AB) or three (ABD) genomes of hexaploid wheat (Gregory 1987). The main objective of breeders was to combine the hardiness and disease resistance of rye with the milling and baking qualities of wheat.

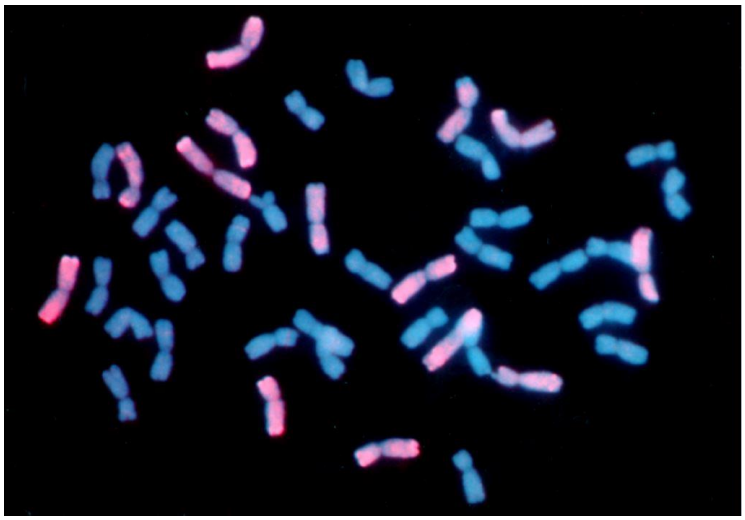


Figure 9: Genomic *in situ* hybridization (GISH) in a triticale cell (2n = 6x = 42), using *Secale cereale* DNA as probe (pink) and *Triticum aestivum* DNA as blocking DNA. Chromosomes were counterstained with DAPI (blue).

There are many types of introgressions from rye to bread wheat including addition, substitution and translocation lines. The first wheat-rye addition lines were produced by O´Mara (1940) who created several disomic wheat-rye addition lines. This set was completed by Driscoll and Sears (1971) and is still a useful tool in wheat and rye genetics (Lukaszewski 2015). Another widespread type of introgression is the 1RS.1BL translocation. This introgression introduced four loci responsible for the tolerance against various diseases located on the short arm of chromosome 1R of rye (Schlegel and Korzun 1997).

2.2.5.2 Introgressions between bread wheat and barley

Barley (*Hordeum vulgare* L.) is together with bread wheat one of the most important crops in the world. Introgressive hybridization enables incorporation of the major agronomical traits (e.g. earliness, tolerance to soil salinity and drought, nutrition quality) from barley to wheat genome. The first successful cross between these species was done by Kruse (1973). After few years later, addition lines (2H, 3H, 4H, 5H, 6H, and 7H) have been produced between the wheat cultivar Chinese Spring (CS) and the spring barley Betzes (Szakács and Molnár-Láng 2007; Molnár-Láng and Linc 2015). Wheat-barley chromosome addition lines are suitable starting material for production of translocation lines (e.g., 2DS.2DL-1HS, 3HS.3BL, 6BS.6BL-4HL) containing only small segment of barley with genes coding

agriculturally important traits (Molnár-Láng *et al*. 2014; Molnár-Láng and Linc 2015).

2.2.5.3 Introgressions between bread wheat and *Aegilops* (goat grasses)

Aegilops is the most closely related genus to *Triticum* and comprises from diploid, tetraploid and hexaploid species with six diverse genomes named C, D, M, N, S, and U. The members of this genus are also *A. tauschii* (2n=2x=14, DD) that is the donor of the hexaploid wheat D genome, and *Ae. speltoides* (2n=2x=14, SS) exhibiting the closest relationship to the B genome of wheat (Dvorak 1998; Molnár-Láng *et al*. 2014; Zhang *et al*. 2015). Generally, the genus *Aegilops* has many agriculturally useful traits, including high level of resistance to biotic (e.g. rusts, powdery mildew) and abiotic (cold, salinity, drought) stresses that may be transferred into bread wheat (Schneider *et al*. 2008; Molnár-Láng *et al*. 2014).

Selected diploid (e.g., *Ae. speltoides, Ae. umbellulata, Ae. caudata*) and polyploid (e.g., *Ae. peregrina, Ae. geniculata*) species from genus *Aegilops* (Schneider *et al*. 2008; Molnár-Láng *et al*. 2014) have been used for production of wheat-*Aegilops* addition lines. Nevertheless, these lines have no practical application in agriculture. For expression of genes from wild species it is desirable to incorporate alien segment directly into a wheat chromosome. The translocation of chromosome segments can be achieved either spontaneously or by irradiation (Zhang *et al*. 2015). The first translocation induced by irradiation was done by Sears (1956) who transferred a segment from *Ae. umbellulata* chromosome 6U carrying resistance to wheat leaf rust (*Lr9*) to chromosome 6BL wheat. This approach has been widely used and a large number of genes were successfully incorporated from *Aegilops* species into wheat and other crop species (Molnár-Láng *et al*. 2014; Zhang *et al*. 2015; Chaudhary *et al*. 2016; Rajpal *et al*. 2016).

2.2.5.4 Introgressions between bread wheat and *Thinopyrum* (syn. *Agropyron*)

Genus *Thinopyrum* (formerly *Agropyron*) can be also used for wheat improvement. This genus includes mostly allopolyploid species on diploid, tetraploid, hexaploid, octoploid and decaploid levels, with the J, E and St genomes (Wang 2011). The most valuable species for introgressive hybridization with wheat

are *Th. intermedium* and *Th. ponticum* that contain resistance genes and their genomes (E and St) are closely related to the A and D genomes of bread wheat.

Remarkable are introgression lines of wheat-*Th. ponticum* with blue aleurone trait. These lines exhibit significantly increased level of anthocyanins having antioxidant, anti-inflammatory, antimicrobial and anti-cancerogenic potential. Burešová *et al*. (2015) revealed large variation in genomic constitutions of blue-aleurone wheat genotypes. They observed six different types of the *Th. ponticum* introgressions (Figure 10), ranging from disomic additions to disomic substitutions, substitutions of whole chromosome arms and various translocations of distal parts of chromosome arm(s). Their results confirmed the hypothesis that alien chromatin from *Th. ponticum* activates the blue aleurone trait present, but inactivated, in common wheat lines.

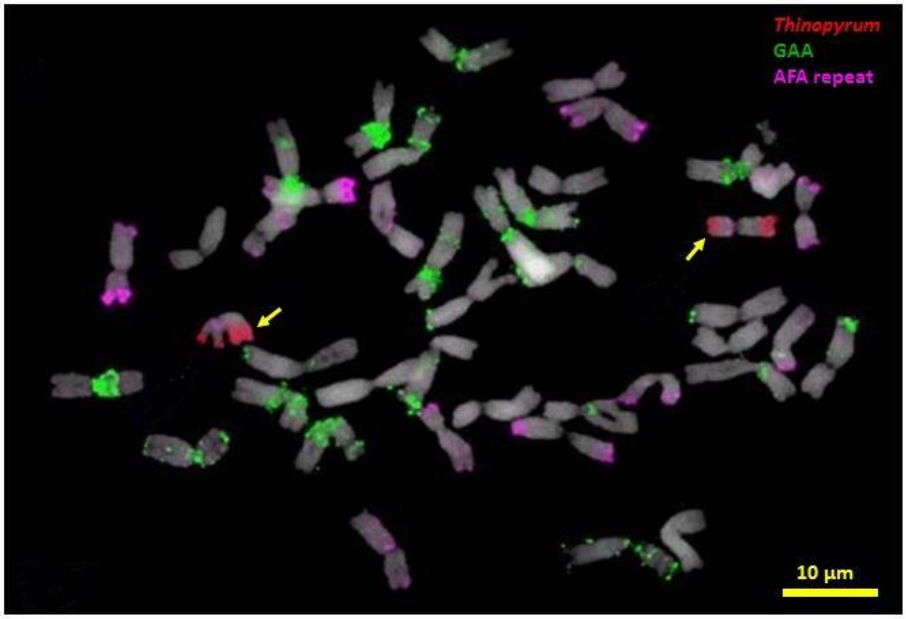


Figure 10: Multicolour fluorescent *in situ* hybridization (FISH) on *Triticum aestivum* cv. Xiao Yan (2n = 6x = 42) was performed using three specific probes. Chromosomes were counterstained by DAPI (grey pseudocolor; Burešová *et al*. 2015).

2.2.5.5 Introgressions between bread wheat and other *Triticum* species

The last mentioned alien introgressions are from the wild species of genus *Triticum* including diploid einkorn (e.g., *T. monococcum, T. urartu, T. boeticum)* and emmer wheats *(*e.g., *T. dicoccum, T. dicoccoides, T. durum*). A large number of

resistance genes has been transferred by introgressive hybridization of wild *Triticum* species and bread wheat (Fedak 2015).

So far, genes against stem and leaf rust, stripe rust, powdery mildew and *Fusarium* head blight have been successfully incorporated into wheat lines. For example, Kolmer *et al*. (2010) transferred the gene *Lr63* for leaf rust resistance from *T. monococcum* into chromosome 3AS of hexaploid wheat. After few years later, two mildew resistance genes from *T. boeticum* were successfully introgressed into wheat chromosome 7AL (Chhuneja *et al*. 2012; Fedak 2015).

2.3 Characterization of plant genomes using cytogenetic techniques

The application of broad spectrum of molecular, cytogenetic, genomic and phenomic methods is prerequisite for the analysis of structure, function, organization and evolution of large and complex hybrid and polyploid plant genomes. With the aim of the Ph.D. thesis, various cytogenetic methods and their modifications will be discussed in the following chapters.

2.3.1 Chromosome banding

The discovery of chromosomes has been early on followed by their characterization including defining their numbers and morphological features such as presence and location of primary and secondary constrictions. However, standard staining methods did not permit unequivocal identification of chromosomes mainly due to their small size and/or similar morphology in almost all species. Therefore, new methods allowing such identification have been needed (Fedak and Kim 2008). The chromosome banding techniques (Q-, G-, R-, C-, and N-banding), developed in the early 1970s, have enabled more precise characterization of individual chromosomes in many species including plants.

Caspersson *et al*. (1968) observed characteristic patterns on chromosomes after using fluorescent dye quinacrine that specifically interact with AT-rich regions (bright bands) in the chromatin. This technique is called Q-banding and has been used for the identification of chromosomes in many species. Equivalent to Q-banding is Giemsa banding (G-banding) producing, conversely to Q banding, the dark bands

in regions with high content of AT bases (Drets and Shaw 1971). Similarly, Dutrillaux and Lejeune (1971) developed R-banding producing the reverse pattern to G- banding. R-bands are GC rich and are helpful for visualization of telomeric ends. The most effective method in wheat karyotyping is C-banding (Figure 11), which stains areas of constitutive heterochromatin located mainly in centromeric regions (Gill and Kimber 1974; Linde-Laursen 1975). C-banding has been also used for identification of alien chromosome(s), characterization of translocations and other structural rearrangements at subchromosomal level (Gill and Kimber 1977; Lukaszewski and Gustafson 1983). The N-banding technique that selectively stains nucleolar organizing regions (NORs) was first applied by Matsui and Sasaki (1973) on mammalian chromosomes. However, it has been revealed that the localization of N-bands does not necessarily correspond to NORs and may occur in different specific regions (Hagele 1979). This method has been successfully used for chromosome mapping of members of the *Triticeae* tribe, including wheat, rye, and barley and various alien introgression lines (Islam 1980; Zeller *et al*. 1987).

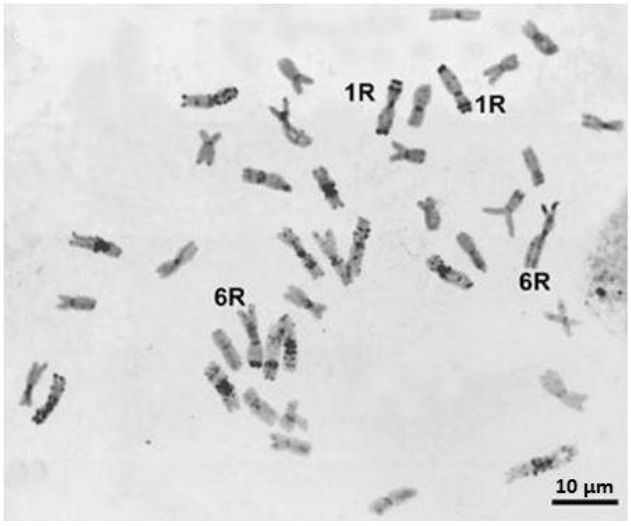


Figure 11: C-banding of mitotic metaphase cell of a double wheat-rye substitution line containing rye chromosomes 1R and 6R (adapted from Forsström *et al*. 2002).

There is no doubt that chromosome banding techniques have been widely used for the karyotyping and chromosome characterization in many plant species. However, the resolution of banding analysis has been frequently limited and does not always provide reliable results.

2.3.2 Fluorescent *in situ* hybridization

Introduction of *in situ* hybridization started a new era of the plant genome analyses (Volpi and Bridger 2008). *In situ* hybridization is a technique allowing the localization of DNA or RNA sequences directly in the nuclei, cytoplasm, organelles, chromosomes, or extended chromatin fibres. This method is used for the identification of chromosomes, chromosome segment(s), specific sequences or whole chromosome sets from different species (Kato *et al*. 2005). Generally, it is based on hybridization of two complementary DNA or RNA sequences: target sequences (usually in form of nuclei or chromosomes fixed on the microscopic slide) and labelled probe (Fuchs and Schubert 1998).

The initial *in situ* hybridization technique used probes labelled with radioisotopes (Gall and Pardue 1969). However, due to time-consuming exposition, instability and negative impact on human health, the radioactive probes are nowadays rarely used and have been replaced by fluorescence-labelled probes. Fluorescence *in situ* hybridization (FISH) and its variations became a valuable tool for the analyses of plant genomes (Ohmido *et al*. 2010). During years, sensitivity, specificity and resolution of FISH have been significant improved hand in hand with the advances in fluorescence microscopy and digital imaging (Volpi and Bridger 2008).

The principle of FISH is simple. Fluorescently labelled DNA sequence (probe) and a sample of interest (e.g., chromosomes, nuclei, chromatin fibres) fixed on slides are denatured to produce single stranded DNA. Denaturation is followed with the hybridization of probe with complementary site on the sample. After washing steps necessary to remove unbound DNA probe, the place(s) of hybridization probe and target sequences can be visualized using fluorescence microscopy (Fuchs and Schubert 1998).

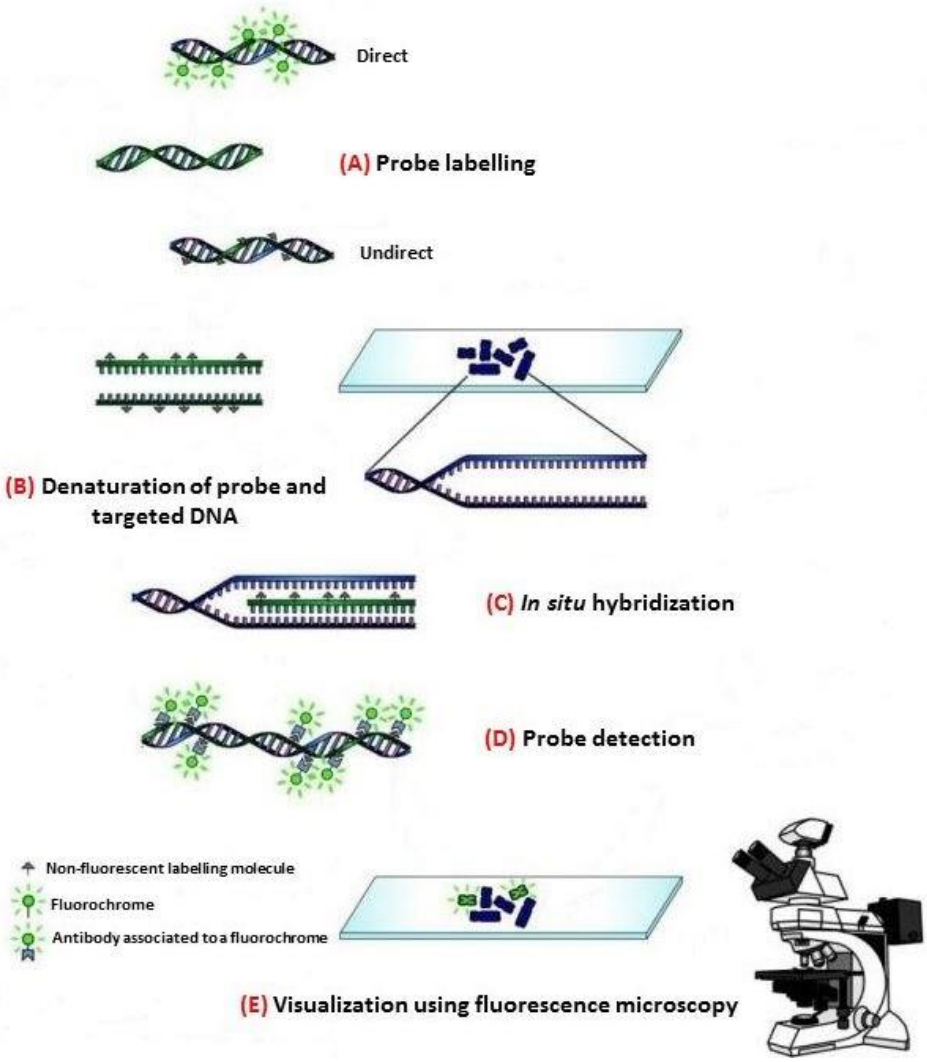


Figure 12: Scheme of fluorescence *in situ* hybridization (adapted from Brammer *et al*. 2013).

Probes for FISH can be prepared from various DNA (RNA) including cloned DNA sequences, genomic DNA, PCR products, synthetic oligonucleotides and inserts from DNA libraries cloned in plasmids, bacteriophages (e.g., lambda), cosmids, and bacterial or yeast artificial chromosome (BAC or YAC; Schwarzacher and Heslop-Harrison 2000).

Although great progress has been made in microscopy and fluorochrome chemistry during last years, the detection of genes and non-coding single or low copy sequences (only a few kilobase pairs in length) is still very difficult in plants. The visualization of so small probes requires sophisticated protocol and appropriate equipment for detection. BAC clones are suitable for localization of single copy sequences in plants with small genome as was shown in rice and cotton (Jiang *et al*.

1995; Hanson *et al*. 1996). However, this approach is not suitable for plants with large genomes due to high amount of repetitive sequences (Fuguerova *et al*. 2012). In this case, cDNAs can be used as FISH probes for localization of single-copy sequences as demonstrated by Danilova *et al*. (2014) in wheat.

Besides localization of single or low copy sequences, FISH has become a routine method for mapping of various types of repetitive sequences (Fuchs and Schubert 1998). FISH helps to define overall genomic distribution of repetitive sequences. Moreover, repeats have been found as useful probes to distinguish individual chromosomes in many plant species. The abundant rDNA gene clusters (45S rDNA, 5S rDNA), centromeric and telomeric repeats and microsatellites are widely used in karyotyping and chromosome identification (Jiang and Gill 2006; Fuguerova *et al*. 2012).

Probes for FISH can be labelled either directly, by incorporation of fluorescent nucleotides, or indirectly, by incorporation of reporter molecules that are subsequently detected using fluorescently labelled antibodies. There are several labelling methods including nick translation, random primer labelling (random priming) and PCR labelling. The protocol for nick translation (Rigby *et al*. 1977) is based on activity of DNase I and DNA polymerase I. The first enzyme does random ´nicks´ in each strand of the double-stranded DNA and the second enzyme incorporates the labelled nucleotides into gaps and resynthesizes DNA strand. This procedure is appropriate for labelling total genomic DNA and large cloned inserts. Random primer labelling developed by Feinberg and Vogelstein (1983) is method based on random hybridization of hexanucleotide primer mix to the single-stranded DNA using the Klenow fragment of DNA polymerase I. This procedure is suitable for labelling of short DNA fragments. The probes can be also prepared using polymerase chain reaction (PCR) with fluorescently labelled nucleotides (Schwarzacher and Heslop-Harrison 2000).

Various modifications of FISH have been developed since the original protocol. Several most widely used applications used in plant genome analyses will be further discussed.

GISH

Genomic *in situ* hybridization (Schwarzacher *et al*. 1989) is the modification of FISH using labelled total genomic DNA(s) as probe(s). GISH can be used for evolutionary studies of allopolyploids and analyses of their genomic constitution. Similarly, such identification of parental chromatin can be monitored in interspecific hybrids, either natural or synthetic. This is valuable in breeding of interspecific hybrids and enables tracking of even small introgressed segments in introgression lines (Schwarzacher *et al*. 1992; Fuchs and Schubert 1998). This technique has been used for clarification of allopolyploid nature of wheat including investigation of its parental subgenomes and many other members of the tribe *Triticeae* (Kato *et al*. 2005).

3D-FISH

3D-FISH is one of the methods to study spatial organization of plant genome and allows the three-dimensional visualization of chromosome territories, subchromosomal domains and various DNA sequences including individual genes during all stages of the cell cycle (Bass *et al*. 1997, 2014). Fixed interphase nuclei or chromosomes are preserved into polyacrylamide gel followed by the hybridization with probe(s) and analysed using confocal microscopy accompanied with specific software (e.g., Imaris, Image J). 3D-FISH can be also used for the visualization of various types of chromatin proteins (3D immuno-FISH). For instance, Phillips *et al*. (2010, 2013) were able to localize proteins of synaptonemal complex, such as ZYP1 and Asy1, in 3D nuclei of barley. 3D-FISH can be also used for the identification of the parental chromatin in nuclei of interspecific hybrids and allopolyploids using labelled total genomic DNAs from parental species as probes (3D-GISH). Such analyses will increase our knowledge on the effects of spatial organization of parental chromatins to the alteration in the gene expression frequently observed in newly developed hybrids (Rey *et al*. 2015).

FISHIS

FISH in suspension (Giorgi *et al*. 2013; Lucretti *et al*. 2014) is fast and cost-effective method using fluorescently label synthetic oligonucleotides as probes for separation of chromosomes by flow cytometry. Based on the hybridization pattern determined by the FISHIS probe(s) and DAPI staining, it is possible to sort chromosomes, which

have the same or very similar relative DNA content and thus, it is unable to sort them by conventional flow cytometry. This method has been used to sort chromosomes from various species of genus *Aegilops* (e.g., *Ae. comosa*, *Ae. umbelullata*, *Ae.speltoides*), *Triticum* (e.g., *T. monococcum*, *T. diccocoides and T. aestivum*) and *Agropyron* (Akpinar *et al*. 2015; Molnár *et al*. 2016).

PRINS

Primed *in situ* labelling is the method combining advantages of FISH and PCR. PRINS is based on annealing of specific oligonucleotide primers, followed by primer extension with Taq DNA polymerase in the presence of labelled nucleotides (Koch *et al*. 1989). In comparison with FISH, this technique is more specific and considerably faster. Moreover, application of directly labelled nucleotides may speed up the protocol (Gosden and Lawson 1995). In plant cytogenetics, PRINS has been used for mapping tandem repeat sequences in wheat, barley and bean. The major disadvantage of this method is relatively high background (Macas *et al*. 1995; Kubaláková *et al*. 1997).

Chromosome painting

Chromosomal *in situ* suppression (CISS) or chromosome painting is the method of hybridization of fluorescently labelled chromosomes or smaller chromosome segments with cytological preparations. This technique was developed independently by few research teams (Pinkel *et al*. 1988; Cremer *et al*. 1988; Lichter *et al*. 1988). Chromosome-specific probes may be derived from flow-sorted (Telenius *et al*. 1992) or microdissected chromosomes or chromosome regions (Lüdecke *et al*. 1989). In human biology, chromosome painting has become a versatile tool in clinical and evolutionary cytogenetics, radiation biology and nuclear topography (Ried *et al*. 1998). Application of CISS with probes from flow-sorted or dissected chromosomes has not been successful in plants. However, employment of cocktail of chromosome-specific BAC and YAC clones as probes delivered the resolution similar to classic CISS in small plant genomes (Jiang *et al*. 1995; Lysak *et al*. 2001). However, high amount of repetitive sequences which are present also in large insert clones hampers the utilization of this approach in large and complex genomes (Fuchs and Schubert 1998; Schubert *et al*. 2001).

2.4 References

AGI - Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 408(6814):796-815

Akpinar BA, Yuce M, Lucas S, Vrána J, Burešová V, Doležel J, Budak H (2015) Molecular organization and comparative analysis of chromosome 5B of the wild wheat ancestor *Triticum dicoccoides*. Sci Report 5:e10763

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) Molecular biology of the cell, 4th ed. Garland Science, New York

Anamthawat-Jónsson K (2001) Molecular cytogenetics of introgressive hybridization in plants. Methods Cell Sci 23(1-3):139-148

Arendsee ZW, Li L, Wurtele ES (2014) Coming of age: orphan genes in plants. Trends Plant Sci 19(11):698-708

Armisén D, Lecharny A, Aubourg S (2008) Unique genes in plants: specificities and conserved features throughout evolution. BMC Evol Biol 8:280

Ascough GD, Staden J, Erwin JE (2008) Effectiveness of colchicine and oryzalin at inducing polyploidy in *Watsonia lepida.* N E Brown Hort Science 43(7):2248-2251

Bannister AJ, Kouzarides T (2011) Regulation of chromatin by histone modifications. Cell Res 21:381-395

Bartke T, Vermeulen M, Xhemalce B, Robson SC, Mann M, Kouzarides T (2010) Nucleosome-interacting proteins regulated by DNA and histone methylation. Cell 143(3):470-484

Barton N (1979) The dynamics of hybrid zones. Heredity 43:341-359

Bass HW, Marshall WF, Sedat JW, Agard DA, Cande WZ (1997) Telomeres cluster *de novo* before initiation of synapsis: a three-dimensional spatial analysis of telomere positions before and during meiotic prophase. J Cell Biol 137:5-18

Bass HW, Wear EE, Lee TJ, Hoffman GG, Gumber HK, Allen GC, Thompson WF, Hanley-Bowdoin L (2014) A maize root tip system to study DNA replication programmes in somatic and endocycling nuclei during plant development. J Exp Bot 65(10):2747-2756

Berr A, Schubert I (2007) Interphase chromosome arrangement in *Arabidopsis thaliana* is similar in differentiated and meristematic tissues and shows a transient mirror symmetry after nuclei division. Genetics 176(2):853-863

Biscotti MS, Olmo E, Heslop-Harrison JS (2015) Repetitive DNA in eukaryotic genomes. Chromosome Res 23:415-420

Biswas M, Voltz K, Smith JC, Langowski J (2011) Role of histone tails in structural stability of the nucleosome. PLoS Comput Biol 7(12):e1002279

Bolzer A, Kreth G, Solovei I, Koehler D, Saracoglu K, Fauth C, Müller S, Eils R, Cremer C, Speicher MR, Cremer T (2005) Three-dimensional maps of all chromosomes in human male fibroblast nuclei and prometaphase rosettes. PLoS Biol 3(5):157

Brammer SP, Vasconcelos S, Poersch LB, Oliveira AR, Brasileiro-Vidal AC (2013) Genomic *in situ* hybridization in *Triticeae*: a methodological approach. In: Andersen SB (ed) Plant breeding from laboratories to fields. Agricultural and Biological Sciences

Brown SW (1966) Heterochromatin. Science 151(3709):417-425

Burešová V, Kopecký D, Bartoš J, Martinek P, Watanabe N, Vyhnánek T, Doležel J (2015) Variation in genome composition of blue-aleurone wheat. Theor Appl Genet 128(2):273-282

Casperson T, Farber S, Foley GE, Kudynowski J, Modest EJ, Simonsson E, Wagh V, Zech L (1968) Chemical differentiation along metaphase chromosomes. Exp Cell Res 49:219-222

Chaudhary HK, Kaila V, Rather SA, Jamwal NS, Badiyal A (2016) Chromosome engineering for high precision crop improvement. In: Rajpal VR, Rao SR, Raina SN (eds) Gene pool diversity and crop improvement. Volume 1. Sustainable development and biodiversity. pp 291-323 Springer Science+Business Media, Switzerland

Chen H, Tini M, Evans RM (2001) HATs on and beyond chromatin. Curr Opin Cell Biol 13(2):218-224

Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol 58:377-406

Chen ZJ (2010) Molecular mechanisms of polyploidy and hybrid vigor. Trends Plant Sci 15(2):57-71

Chhuneja P, Kumar K, Stirnweis D, Hurni S, Keller B, Dhaliwal HS, Singh K (2012) Identification and mapping of two powdery mildew resistance genes in *Triticum boeoticum* L. Theor Appl Genet 124(6):1051-1058

Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Le Gouis J, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Šafář J, Doležel J, Rogers J, Vandepoele K, Aury JM, Mayer K, Bergès H, Quesneville H, Wincker P, Feuillet C (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science 345(6194):1249721

Christophorou MA, Castelo-Branco G, Halley-Stott RP, Oliveira CS, Loos R, Radzisheuskaya A, Mowen KA, Bertone P, Silva JC, Zernicka-Goetz M, Nielsen ML, Gurdon JB, Kouzarides T (2014) Citrullination regulates pluripotency and histone H1 binding to chromatin. Nature 507(7490):104-108

Ciosk R, Zachariae W, Michaelis C, Shevchenko A, Mann M, Nasmyth K (1998) An Esp1/Pds1 complex regulates loss of sister chromatid cohesion at the metaphase to anaphase transition in yeast. Cell 93:1067-1076

Colby C (1996) Introduction to evolutionary biology. Available at http://www.talkorigins.org/faqs/faq-intro-to-biology.html

Comai L (2005) The advantages and disadvantages of being polyploid. Nat Rev Genet 6(11):836-846

Cooper GM (2000) The cell: a molecular approach, 2nd ed. Sunderland (MA): Sinauer Associates, Boston University

Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. Nat Rev Genet 2(4):292-301

Cremer T, Cremer M (2010) Chromosome territories. Cold Spring Harb Perspect Biol 2(3):a003889

Cremer T, Lichter P, Borden J, Ward DC, Manuelidis L (1988) Detection of chromosome aberrations in metaphase and interphase tumor cells by *in situ* hybridization using chromosome-specific library probes. Hum Genet 80(3):235-246

Cuadrado A, Cardoso M, Jouve N (2008a) Physical organisation of simple sequence repeats (SSRs) in *Triticeae*: structural, functional and evolutionary implications. Cytogenet Genome Res 120(3-4):210-219

Cuadrado A, Cardoso M, Jouve N (2008b) Increasing the physical markers of wheat chromosomes using SSRs as FISH probes. Genome 51(10):809-815

Cuadrado A, Jouve N (2002) Evolutionary trends of different repetitive DNA sequences during speciation in the genus *Secale*. J Hered 93:339-345

Cuadrado A, Jouve N (2007) The nonrandom distribution of long clusters of all possible classes of trinucleotide repeats in barley chromosomes. Chromosome Res 15(6):711-720

Danilova T, Friebe B, Gill BS (2014) Development of a wheat single gene FISH map for analysing homeologous relationship and chromosomal rearrangements within the *Triticeae*. Theor Appl Genet 127(3):715-730

Davis SK, Bardeen CJ (2004) The connection between chromatin motion on the 100nm length scale and core histone dynamics in live XTC-2 cells and isolated nuclei. Biophys J 86:555-564

de la Barre AE, Gerson V, Gout S, Creaven M, Allis CD, Dimitrov (2000) Core histone N-termini play an essential role in mitotic chromosome condensation. EMBO J 19(3):379-391

De Storme N, Mason A (2014) Plant speciation through chromosome instability and ploidy change: cellular mechanisms, molecular factors and evolutionary relevance. Curr Plant Biol 1:10-33

Dekker J, Rippe K, Dekker M, Kleckner N (2002) Capturing chromosome conformation. Science 295(5558):1306-1311

Denoeud F, Vergnaud G, Benson G (2003) Predicting human minisatellite polymorphism. Genome Res 13(5):856-867

Doenecke D (2014) Chromatin dynamics from S-phase to mitosis: contributions of histone modifications. Cell Tissue Res 356(3):467-475

Dong F, Jiang J (1998) Non-Rabl patterns of centromere and telomere distribution in the interphase nuclei of plant cells. Chromosome Res 6(7):551-558

Dostie J, Richmond TA, Arnaout RA, Selzer RR, Lee WL, Honan TA, Rubio ED, Krumm A, Lamb J, Nusbaum C, Green RD, Dekker J (2006) Chromosome conformation capture carbon copy (5C): a massively parallel solution for mapping interactions between genomic elements. Genome Res 16(10):1299-1309

Drets ME, Shaw MW (1971) Specific banding patterns of human chromosomes. Proc Natl Acad Sci U S A 68(9):2073-2077

Driscoll CJ, Sears ER (1971) Individual addition of the chromosomes of ‘Imperial‘ rye to wheat. Agron Abstr 1971:6

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

Duplan V, Rivas S (2014) E3 ubiquitin-ligases and their target proteins during the regulation of plant innate immunity. Front Plant Sci 5:42

Dutrillaux B, Lejeune J (1971) Sur une nouvelle technique d'analyse du caryotype humain. Comptes Rendus Hebdomadaires des Seances de l'Acad6mie des Sciences 272:2638-2640

Dvořák J, Luo MC, Yang ZL, Zhang HB (1998) The structure of the *Aegilops tauschii* genepool and the evolution of hexaploid wheat. Theor Appl Genet 97:657-670

Dyck PL, Kerber ER, Martens JW (1990) Transfer of a gene for stem rust resistance from *Aegilops caudata* to common wheat. Can J Plant Sci 70:931-934

Eberharter A, Becker PB (2002) Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. EMBO Rep 3(3):224-229

Eddy SR (2012) The C-value paradox, junk DNA and ENCODE. Curr Biol 22(21):898-899

Ensembl Plants (2016) Available at http://plants.ensembl.org/Oryza_sativa/Info/Annotation

Eymery A, Callanan M, Vourc'h C (2009) The secret message of heterochromatin: new insights into the mechanisms and function of centromeric and pericentric repeat sequence transcription. Int J Dev Biol 53(2-3):259-268

Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, Vazquez J, Valencia A, Tress ML (2014) Multiple evidence strands suggest that there may be as few as 19,000 human protein-coding genes. Hum Mol Genet 23(22):5866-5878

Fedak G (2015) Alien introgressions from wild *Triticum* species, *T. monococcum, T. urartu, T. turgidum, T. dicoccum, T. dicoccoides, T. carthlicum, T. araraticum, T. timopheevii,* and *T. miguschovae*. In: Molnár-Láng M, Ceoloni C, Doležel J (eds) Alien introgression in wheat: cytogenetics, molecular biology, and genomics. pp 191-219 Springer Science+Business Media, Switzerland

Fedak G, Kim NS (2008) Tools and methodologies for cytogenetic studies of plant chromosomes. Tsitol Genet 42(3):64-80

Feinberg AP, Vogelstein B (1983) A technique for radiolabelling DNA restriction endonuclease fragments to high specificactivity. Anal Biochem 132(1):6-13

Feldman M, Levy AA (2012) Genome evolution due to allopolyploidization in wheat. Genetics 192(3):763-774

Ferrai C, Castro IJ, Lavitas L, Chotalia M, Pombo A (2010) Gene positioning. Perspect Biol 2(6):a000588

Feschotte C, Pritham EJ (2007) DNA transposons and the evolution of eukaryotic genomes. Annu Rev Genet 41:331-368

Figueroa DM, Amarillo IE, Bass HW (2012) Cytogenetic mapping in plants. In: Bass HW, Birchler JA (eds) Plant cytogenetics. Plant genetics and genomics: crops and models 4. pp 79-119 Springer Science+Business Media, New York

Finch JT, Klug A (1976) Solenoidal model for superstructure in chromatin. Proc Natl Acad Sci U S A 73(6):1897-1901

Finley D, Ulrich HD, Sommer T, Peter Kaiser (2012) The ubiquitin-proteasome system of *Saccharomyces cerevisiae.* Genetics 192(2):319-360

Flagel L, Udall J, Nettleton D, Wendel J (2008) Duplicate gene expression in allopolyploid *Gossypium* reveals two temporally distinct phases of expression evolution. BMC Biol 6:16

Flavell RB, Bennett MD, Smith JB, Smith DB (1974) Genome size and the proportion of repeated nucleotide sequence DNA in plants. Biochem Genet 12:257-269

Forsström PO, Merker A, Schwarzacher T (2002) Characterisation of mildew resistant wheat-rye substitution lines and identification of an inverted chromosome by fluorescent *in situ* hybridization. Heredity 88(5):349-355

Fransz P, de Jong H (2011) From nucleosome to chromosome: a dynamic organization of genetic information. Plant J 66(1):4-17

Fransz P, Jong H, Lysak M, Castiglione MR, Schubert I (2002) Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. Proc Natl Acad Sci U S A 99(22):14584-14589

Fuchs J, Schubert I (1998) Characterization of plant genomes using fluorescence *in situ* hybridization. In Maluszynska J (ed) Plant cytogenetics. Proceedings of Spring Symposium, Cieszyn, Prace Naukowe Uniwersytetu Slaskiego 1698. pp 112-123 Katowice

Gall JG (1981) Chromosome structure and the C-value paradox. J Cell Biol 91(3 Pt 2):3-14

Gall JG, Pardue ML (1969) Formation and detection of RNA-DNA hybrid molecules in cytological preparations. Proc Natl Acad Sci U S A 63:378-383

Garcia S, Kovařík A (2013) Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. Heredity 111(1):23-33

Gibcus JH, Dekker J (2013) The hierarchy of the 3D genome. Mol Cell 49:776-782

Gilbert N, Boyle S, Sutherland H, Heras J, Allan J, Jenuwein T, Bickmore WA (2003) Formation of facultative heterochromatin in the absence of HP1. EMBO J 22(20):5540-5550

Gill BS, Kimber G (1974) The Giemsa C-banded karyotype of rye. Proc Natl Acad Sci U S A 71(4):1247-1249

Gill BS, Kimber G (1977) Recognition of translations and alien chromosome transfers in wheat by Giemsa C-banding technique. Crop Sci 17:264-266

Giorgi D, Farina A, Grosso V, Gennaro A, Ceoloni C, Lucretti S (2013) FISHIS: fluorescence *in situ* hybridization in suspension and chromosome flow sorting made easy. PLoS One 8(2):e57994

Gorkin DU, Leung D, Ren B (2014) The 3D genome in transcriptional regulation and pluripotency. Cell Stem Cell 14(6):762-775

Gosden J, Lawson D (1995) Instant PRINS: a rapid method for chromosome identification by detecting repeated sequences *in situ*. Cytogenel Cell Genet 68:57-60

Graur D, Shuali Y, Li WH (1989) Deletions in processed pseudogenes accumulate faster in rodents than in humans. J Mol Evol 28(4):279-285

Gregory RS (1987) Triticale breeding. In: Lupton FGH (ed) Wheat breeding: its scientific basis. pp 269-286 Chapman and Hall, London

Griffiths AJF, Gelbart WM, Miller JH, Lewontin RC (1999) Modern genetic analysis. W. H. Freeman and Company, New York

Grob S, Schmid MW, Grossniklaus U (2014) Hi-C analysis in *Arabidopsis* identifies the KNOT, a structure with similarities to the flamenco locus of *Drosophila*. Mol Cell 55(5):678-693

Guo T, Fang Y (2014) Functional organization and dynamics of the cell nucleus. Front Plant Sci 5:378

Gupta S, Kim SY, Artis S, Molfese DL, Schumacher A, Sweatt JD, Paylor RE, Lubin FD (2010) Histone methylation regulates memory formation. J Neurosci 30(10):3589-3599

Hagele K (1979) Characterization of heterochromatin in *Schistocerca gregaria* by C and N banding methods. Chromosoma 70:239-250

Hanson RE, Islam-Faridi MN, Percival EA, Crane CF, Ji Y, McKnight TD, Stelly DM, Price HJ (1996) Distribution of 5S and 18S-28S rDNA loci in a tetraploid cotton (*Gossypium hirsutum* L.) and its putative diploid ancestors. Chromosoma 105(1):55-61

Happel N, Doenecke D (2009) Histone H1 and its isoforms: contribution to chromatin structure and function. Gene 431(1-2):1-12

Heslop-Harrison JS (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. Plant Cell 12(5):617-636

Heslop-Harrison JS, Schmidt T (2012) Plant nuclear genome composition. Encyclopedia of life sciences. pp 1-8 John Wiley & Sons Ltd, Chichester

Hongyong Fu, Goring DR, Genschik P (2014) Reversible ubiquitylation in plant biology. Front Plant Sci 5:707

Horton TR (2006) The number of nuclei in basidiospores of 63 species of ectomycorrhizal *Homobasidiomycetes*. Mycologia 98(2):233-238

Huisinga KL, Brower-Toland B, Elgin SCR (2006) The contradictory definitions of heterochromatin: transcription and silencing. Chromosoma 115:110-122

Islam AKMR (1980) Identification of wheat barley addition lines with N-banding of chromosomes. Chromosoma 76:365-373

IWGSC - International Wheat Genome Sequencing Consortium (2014) A chromosome-based draft sequence of the hexaploid bread wheat (*Triticum aestivum*) genome. Science 345(6194):1251788

Jiang J, Gill B, Wang G-L, Ronald PC, Ward DC (1995) Metaphase and interphase fluorescence *in situ* hybridization mapping of the rice genome with bacterial artificial chromosomes. Proc Natl Acad Sci 92:4487-4491

Jiang J, Gill BS (2006) Current status and the future of fluorescence *in situ* hybridization (FISH) in plant genome research. Genome 49(9):1057-1068

Kaneko Y, Bang SW (2014) Interspecific and intergeneric hybridization and chromosomal engineering of *Brassicaceae* crops. Breed Sci 64(1):14-22

Kato A, Vega JM, Han F, Lamb JC, Birchler JA (2005) Advances in plant chromosome identification and cytogenetic techniques. Curr Opin Plant Biol 8(2):148-154

Kellogg EA, Bennetzen JL (2004) The evolution of nuclear genome structure in seed plants. Am J Bot 91(10):1709-1725

Kihara H, Ono T (1926) Chromosomenzahlen und systematische Gruppierung der Rumex-Arten. Zeitschr Zellf Mikro Anat 4:475-481

Kimura K, Cuvier O, Hirano T (2001) Chromosome condensation by a human condensin complex in *Xenopus* egg extracts. J Biol Chem 276:5417-5420

Kimura K, Rybenkov VV, Crisona NJ, Hirano T, Cozzarelli NR (1999) 13S condensin actively reconfigures DNA by introducing global positive writhe: implications for chromosome condensation. Cell 98:239-248

Knappe S, Flügge UF, Fischer K (2003) Analysis of the plastidic phosphate translocator gene family in *Arabidopsis* and identification of new phosphate translocator-homologous transporters, classified by their putative substrate-binding site. Plant Physiol 131(3):1178-1190

Knight CA, Molinari NA, Petrov DA (2005) The large genome constraint hypothesis: evolution, ekology and phenotype. Ann Bot 95:177-190

Koch J, Kolyraa S, Petersen KB, Gregersen N, Bolund L (1989) Oligonucleotide priming methods for the chromosome-specific labelling of alpha satellite DNA *in situ*. Chromosoma 98:259-265

Kolmer JA, Anderson JA, Flor JM (2010) Chromosome location, linkage with simple sequence repeat markers, and leaf rust resistance conditioned by gene *Lr63* in wheat. Crop Sci 50:2392-2395

Kornberg RD (1974) Chromatin structure: a repeating unit of histones and DNA. Science 184:868-871

Koshland D, Strunnikov A (1996) Mitotic chromosome condensation. Annu Rev Cell Dev Biol 12:305-333

Kouzarides T (2007) Chromatin modifications and their function. Cell 128:693-705

Krebs JE, Lewin B, Goldstein ES, Kilpatrick ST (2013) Lewin's essential genes, 3rd ed. Jones & Bartlett Learning, Burlington

Kruse A (1973) *Hordeum* x *Triticum* hybrids. Hereditas 73:157-161

Kubaláková M, Macas J, Doležel J (1997) Mapping of repeated DNA sequences in plant chromosomes by PRINS and C-PRINS. Theor Appl Genet 94:758-763

Lan F, Shi Y (2009) Epigenetic regulation: methylation of histone and non-histone proteins. Sci China C Life Sci 52:311-322

Leitch AR (2000) Higher levels of organization in the interphase nucleus of cycling and differentiated cells. Microbiol Mol Biol Rev 64(1):138-152

Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, Clerck O (2012) Phylogeny and molecular evolution of the green algae. Crit Rev Plant Sci 31:1-46

Levin DA (2002) The role of chromosomal change in plant evolution. Oxford University Press, New York

Lichter P, Cremer T, Borden J, Manuelidis L, Ward C (1988) Delineation of individual human chromosomes in metaphase and interphase cells by *in situ* suppression hybridization using chromosome specific library probes. Hum Genet 80:224-234

Linde-Laursen I (1975) Giemsa C-banding of the chromosomes of 'Emir' barley. Hereditas 81:285-289

Liu D, Zhang H, Zhang L, Yuan Z, Hao M, Zheng Y (2014) Distant hybridization: a tool for interspecific manipulation of chromosomes. In: Pratap A, Kumar J (eds) Alien gene transfer in crop plants. Volume 1. Innovations, methods and risk assessment. pp 25-42 Springer Science+Business Media, New York

Lobkovsky AE, Wolf YI, Koonin EV (2015) Evolvability of an optimal recombination rate. Genome Biol Evol 8(1):70-77

Loeve L, Hill WG (2010) The population genetics of mutations: good, bad and indifferent. Philos Trans R Soc Lond B Biol Sci 365(1544):1153-1167

López-Caamal A, Tovar-Sánchez E (2014) Genetic, morphological, and chemical patterns of plant hybridization. Rev Chil His Nat 87:16

Löve A (1984) Conspectus of the *Triticeae*. Feddes Repert 95:425-521

Lucretti S, Giorgi D, Farina A, Grosso V (2014) FISHIS: A new way in chromosome flow sorting makes complex genomes more accessible. In: Tuberosa R, Graner A, Frison E (eds) Genomics of plant genetic resources. Volume 1. Managing, sequencing and mining genetic resources. pp 319-348 Springer Science+Business Media, Dordrecht

Lüdecke HJ, Senger G, Claussen U, Horsthemke B (1989) Cloning defined regions of the human genome by microdissection of banded chromosomes and enzymatic amplification. Nature 23:348-350

Lukaszewski AJ (2015) Introgressions between wheat and rye. In: Molnár-Láng M, Ceoloni C, Doležel J (eds) Alien introgression in wheat: cytogenetics, molecular biology, and genomics. pp 163-189 Springer Science+Business Media, Switzerland

Lukaszewski AJ, Gustafson JP (1983) Translocations and modifications of chromosomes in triticale x wheat hybrids. Theor Appl Genet 64:239-248

Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalization. Genetics 154:459-473

Lysak MA, Fransz PF, Ali HB, Schubert I (2001) Chromosome painting in *Arabidopsis thaliana*. Plant J 28(6):689-697

Macas J, Doležel J, Gualberti G, Pich U, Schubert I, Lucretti S (1995) Primer-induced labelling of pea and field bean chromosomes *in situ* and in suspension. Biotechniques 19:402-404

Madlung A (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110(2):99-104

Maeshima K, Imai R, Tamura S, Nozaki T (2014) Chromatin as dynamic 10-nm fibers. 123(3):225-237

Martins LA-CP (1999) Did Sutton and Boveri propose the so-called Sutton-Boveri chromosome hypothesis? Genet Mol Biol 22(2):261-271

Maston GA, Evans SK, Green MR (2006) Transcriptional regulatory elements in the human genome. Annu Rev Genomics Human Genet **7**:29-59

Matsui S, Sasaki M (1973) Differential staining of nucleolus organisers in mammalian chromosomes. Nature 246:148-150

McClintock B (1984) The significance of responses of the genome to challenge. Science 226:792-801

McKain MR, Tang H, McNeal JR, Ayyampalayam S, Davis JI, dePamphilis CV, Givnish TJ, Pires JC, Stevenson DW, Leebens-Mack JH (2016) A phylogenomic assessment of ancient polyploidy and genome evolution across the *Poales*. Genome Biol Evol 8(4):1150-1164

Mehrotra S, Goyal V (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genomics Proteomics Bioinformatics 12(4):164-171

Mello-Sampayo T, Canas AP (1973) Suppressors of meiotic chromosome pairing in common wheat. In: Sears ER, Sears LMS (eds) Proceedings of the 4th international wheat genetics symposium. pp 703-713 College of Agriculture and University of Missouri Columbia

Molnár I, Vrána J, Burešová V, Cápal P, Farkas A, Darkó É, Cseh A, Kubaláková M, Molnár-Láng M, Doležel J (2016) Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their syntheny with wheat. Plant J:13266

Molnár-Láng M, Ceoloni C, Doležel J (2015) Alien introgression in wheat: cytogenetics, molecular biology, and genomics. Springer Science+Business Media, Swirzerland

Molnár-Láng M, Linc G (2015) Wheat-barley hybrids and introgression lines. In: Molnár-Láng M, Ceoloni C, Doležel J (eds) Alien introgression in wheat: cytogenetics, molecular biology, and genomics. pp 315-345 Springer Science+Business Media, Switzerland

Molnár-Láng M, Molnár I, Szakács É, Linc G, Bed Z (2014) Production and molecular cytogenetic identification of wheat-alien hybrids and introgression lines. In: Tuberosa R, Graner A, Frison E (eds) Genomics of plant genetic resources. Volume 1. Managing, sequencing and mining genetic resources. pp 255-2863 Springer Science+Business Media, Dordrecht

Moore GP (1984) The C-value paradox. BioScience 34(7):425-429

Morgan ER, Timmerman-Vaughan GM, Conner AJ, Griffin WB, Pickering R (2011) Plant interspecific hybridization: outcomes and issues at the intersection of species. Plant Breed Rev 34:161-220

O´Mara JG (1940) Cytogenetic studies on triticale: I. a method for determining the effects of individual *Secale* chromosomes on *Triticum*. Genetics 25:401-408

Ohmido N, Fukui K, Kinoshita T (2010) Recent advances in rice genome and chromosome structure research by fluorescence *in situ* hybridization (FISH). Proc Jpn Acad Ser B Phys Biol Sci 86(2):103-116

Ohno S (1970) Evolution by gene duplication. Springer-Verlag, New York

Olins AL, Olins DE (1974) Spheroid chromatin units (v bodies). Science 183:330-332

Olins DE, Olins AL (2003) Chromatin history: our view from the bridge. Nature Rev Mol Cell Biol 4:809-814

Otto SP, Whiton J (2000) Polyploid incidence and evolution. Annu Rev Genet 34:401-437

Pearce SR, Harrison G, Li D, Heslop-Harrison, Kumar A, Flavell AJ (1996) The *Ty1*-copia group retrotransposons in *Vicia* species: copy number, sequence heterogeneity and chromosomal localisation. Mol Gen Genet 250:305-315

Pecinka A, Schubert V, Meister A, Kreth G, Klatte M, Lysak MA, Fuchs J, Schubert I (2004) Chromosome territory arrangement and homologous pairing in nuclei of *Arabidopsis thaliana*. Chromosoma 113(5):258-269

Pennacchio LA, Bickmore W, Dean A, Nobrega MA, Bejerano G (2013) Enhancers: five essential questions. Nat Rev Genet 14(4):288-295

Pertea M, Salzberg SL (2010) Between a chicken and a grape: estimating the number of human genes. Genome Biol 11(5):206

Pestsova E, Ganal MW, Röder MS (2000) Isolation and mapping of microsatellite markers specific for the D genome of bread wheat. Genome 43(4):689-697

Phillips D, Jenkins G, Macaulay M, Nibau C, Wnetrzak J, Fallding D, Colas I, Oakey H, Waugh R, Ramsay L (2015) The effect of temperature on the male and female recombination landscape of barley. New Phytol 208(2):421-429

Phillips D, Nibau C, Ramsay L, Waugh R, Jenkins G (2010) Development of a molecular cytogenetic recombination assay for barley. Cytogenet Genome Res 129(1-3):154-161

Phillips D, Wnetrzak J, Nibau C, Barakate A, Ramsay L, Wright F, Higgins JD, Perry RM, Jenkins G (2013) Quantitative high resolution mapping of *HvMLH3* foci in barley pachytene nuclei reveals a strong distal bias and weak interference. J Exp Bot 64(8):2139-2154

Pierce BA (2005) Genetics: a conceptual approach, 2nd ed. W. H. Freeman and Company, New York

Pink RC, Wicks K, Caley DP, Punch EK, Jacobs L, Carter DR (2011) Pseudogenes: pseudo-functional or key regulators in health and disease? RNA 17(5):792-798

Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J, Gray J. (1988) Fluorescence *in situ* hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc Natl Acad Sci U S A 85: 9138-9142

Pray LA (2008) Eukaryotic genome complexity. Nat Educ 1(1):96. Available at http://www.nature.com/scitable/topicpage/eukaryotic-genome-complexity-437

Rabl C (1885) Über Zellteilung. Morph Jahrb 10:214-330

Rajpal VR, Rao SR, Raina SN (2016) Gene pool diversity and crop improvement. Volume 1. Sustainable development and biodiversity. Springer Science+Business Media, Switzerland

Ramel C (1997) Mini- and microsatellites. Environ Health Perspect 105(4):781-789

Ramsey J, Schemske DW (2002) Neopolyploidy in flowering plants. Annu Rev Ecol. Evol Syst 33:589-639

Rey E, Molnár I, Doležel J (2015) Genomics of wild relatives and alien introgressions. In: Molnár-Láng M, Ceoloni C, Doležel J (eds) Alien introgression in wheat: cytogenetics, molecular biology, and genomics. pp 347-381 Springer Science+Business Media, Switzerland

Ried T, Schröck E, Ning Y, Wienberg J (1998) Chromosome painting: a useful art. Hum Mol Genet 7(10):1619-1626

Rigby PW, Dieckmann M, Rhodes C, Berg P (1977) Labelling deoxyribonucleic acid to high specific activity *in vitro* by nick translation with DNA polymerase I. J Mol Biol 113(1): 237-251

Riley R, Chapman V (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat. Nature 182:713-715

Robinson PJ, Fairall L, Huynh VA, Rhodes D (2006) EM measurements define the dimensions of the ‘‘30-nm’’ chromatin fiber: evidence for a compact, interdigitated structure. Proc Natl Acad Sci U S A 103(17):6506-6511

Rogakou EP, Pilch DR, Orr AH, Ivanova VS, Bonner WM (1998) DNA double-stranded breaks induce histone H2AX phosphorylation on serine 139. J Biol Chem 273(10):5858-5868

Rossetto D, Avvakumov N, Côté J (2012) Histone phosphorylation: a chromatin modification involved in diverse nuclear events. Epigenetics 7(10):1098-1108

Saintenac C, Falque M, Martin OC, Paux E, Feuillet C, Sourdille P (2008) Detailed recombination studies along chromosome 3B provide new insights on crossover distribution in wheat (*Triticum aestivum* L.). Genetics 181(2):393-403

Sakai H, Koyanagi KO, Imanishi T, Itoh T, Gojobori T (2007) Frequent emergence and functional resurrection of processed pseudogenes in the human and mouse genomes. Gene 389(2):196-203

Saminathan T, Nimmakayala P, Manohar S, Malkaram S, Almeida A, Cantrell, Tomason Y, Abburi L, Rahman MA, Vajja VG, Khachane A, Kumar B, Rajasimha HK, Levi A, Wehner T, Reddy UK (2015) Differential gene expression and alternative splicing between diploid and tetraploid watermelon. J Exp Bot 66(5):1369-1385

Sanborn AL, Rao SS, Huang SC, Durand NC, Huntley MH, Jewett AI, Bochkov ID, Chinnappan D, Cutkosky A, Li J, Geeting KP, Gnirke A, Melnikov A, McKenna D, Stamenova EK, Lander ES, Aiden EL (2015) Chromatin extrusion explains key features of loop and domain in wild-type and engineered genomes. Proc Natl Acad Sci U S A 112(47):6456-6465

Santos AP, Ferreira L, Maroco J, Oliveira MM (2011) Abiotic stress and induced DNA hypomethylation cause interphase chromatin structural changes in rice rDNA loci. Cytogenet Genome Res 132(4):297-303

Santos AP, Shaw P (2004) Interphase chromosomes and the Rabl configuration: does the genome size matter? J Microsc 214(2):201-206

Schlegel R, Korzun V (1997) About the origin of 1RS.1BL wheat-rye chromosome translocations from Germany. Plant Breed 116:537-540

Schlegel RHJ (2010) Dictionary of plant breeding, 2nd ed. CRC Press. Taylor and Francis Group. Boca Raton, London, New York

Schneider A, Molnár I, Molnár-Láng M (2008) Utilisation of *Aegilops* (goatgrass) species to widen the genetic diversity of cultivated wheat. Euphytica 163:1-19

Schubert I, Fransz PF, Fuchs J, Jong H (2001) Chromosome painting in plants. Methods Cell Sci 23:57-69

Schubert I, Shaw P (2011) Organization and dynamics of plant interphase chromosomes. Trends Plant Sci 16(5):273-281

Schubert V, Rudnik R, Schubert I (2014) Chromatin associations in *Arabidopsis* interphase nuceli. Front Genet 5:389

Schueler MG, Sullivan BA (2006) Structural and functional dynamics of human centromeric chromatin. Annu Rev Genomics Hum Genet 7:301-313

Schuler GD, Boguski MS, Stewart EA, Stein LD, Gyapay G, Rice K, White RE, Rodriguez-Tomé P, Aggarwal A, Bajorek E, Bentolila S, Birren BB, Butler A,Castle AB, Chiannilkulchai N, Chu A, Clee C, Cowles S, Day PJ, Dibling T, Drouot N, Dunham I, Duprat S, East C, Edwards C, Fan JB, Fang N, Fizames C, Garrett C, Green L, Hadley D, Harris M, Harrison P, Brady S, Hicks A, Holloway E, Hui L, Hussain S, Louis-Dit-Sully C, Ma J, MacGilvery

A, Mader C,Maratukulam A, Matise TC, McKusick KB, Morissette J, Mungall A, Muselet D, Nusbaum HC, Page DC, Peck A, Perkins S, Piercy M, Qin F, Quackenbush J,Ranby S, Reif T, Rozen S, Sanders C, She X, Silva J, Slonim DK, Soderlund C, Sun WL, Tabar P, Thangarajah T, Vega-Czarny N, Vollrath D, Voyticky S,Wilmer T, Wu X, Adams MD, Auffray C, Walter NA, Brandon R, Dehejia A, Goodfellow PN, Houlgatte R, Hudson JR Jr, Ide SE, Iorio KR, Lee WY, Seki N,Nagase T, Ishikawa K, Nomura N, Phillips C, Polymeropoulos MH, Sandusky M, Schmitt K, Berry R, Swanson K, Torres R, Venter JC, Sikela JM, Beckmann JS, Weissenbach J, Myers RM, Cox DR, James MR, Bentley D, Deloukas P, Lander ES, Hudson TJ (1996) A gene map of the human genome. Science 274(5287):540-546

Schwartz M, Hakim O (2014) 3D view of chromosomes, DNA damage, and translocations. Curr Opin Genet Dev 25:118-125

Schwarzacher T, Anamthawat-Jónsson K, Harrison GE, Islam AK, Jia JZ, King IP, Leitch AR, Miller TE, Reader SM, Rogers WH, Shi M, Heslop-Harrison JS (1992) Genomic *in situ* hybridization to identify alien chromosomes and chromosome segments in wheat. Theor Appl Genet 84:778-786

Schwarzacher T, Heslop-Harrison JS (2000): Practical *in situ* hybridization. BIOS Scientific Publishers Limited, Oxford

Schwarzacher T, Leitch AR, Bennett MD, Heslop-Harrison JS (1989) *In situ* localization of parental genomes in a wide hybrid. Ann Bot 64:315-324

Sears ER (1956) The transfer of leaf rust resistance from *Aegilops umbellulata* to wheat. Brookhaven Symp Biol 9:1-21

Sears ER (1976) Genetic control of chromosome pairing in wheat. Annu Rev Genet 10:31-51

Sears ER, Okamoto M (1958) Intergenomic chromosome relationships in hexaploid wheat. Proc X Int Cong Genet 2:258-259

Sehgal N, Fritz AJ, Morris K, Torres I, Chen Z, Xu J, Berezney R (2014) Gene density and chromosome territory shape. Chromosoma 123(5):499-513

Sequeira-Mendes J, Gutierrez C (2016) Genome architecture: from linear organisation of chromatin to the 3D assembly in the nucleus. Chromosoma 125(3):455-469

Shcherban AB (2015) Repetitive DNA sequences in plant genomes. Russian Journal of Genetics: Applied Research 5(3):159-167

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

Soltis PS, Soltis DE (2009) The role of hybridization in plant speciation. Annu Rev Plant Biol 60:561-588

Soutoglou E, Misteli T (2007) Mobility and immobility of chromatin in transcription and genome stability. Curr Opin Genet Dev 17(5):435-442

Staňková H (2015) Construction of physical map of 7DS wheat chromosome arm and its use for positing cloning. Dissertation. Palacký University Olomouc, Faculty of Science, Department of Botany, Olomouc

Stebbins GL (1950) Variation and evolution in plants. Columbia University Press, New York

Sterner DE, Berger SL (2000) Acetylation of histones and transcription-related factors. Microbiol Mol Biol Rev 64(2):435-459

Sun HB, Shen J, Yokota H (2000) Size-dependent positioning of human chromosomes in interphase nuclei. Biophys J 79(1):184-190

Sutherland H, Bickmore WA (2009) Transcription factories: gene expression in unions? Nat Rev Genet 10(7):457-466

Sýkorová E, Lim KY, Kunická Z, Chase MW, Bennett MD, Fajkus J, Leitch AR (2003) Telomere variability in the monocotyledonous plant order *Asparagales*. Proc Biol Sci 270:1893-1904

Szakács É, Molnár-Láng M (2007) Development and molecular cytogenetics identification of new winter wheat/winter barley (Martonvásári 9 kr1/Igri) disomic addition lines. Genome 50(1):43-50

Tautz D, Domazet-Lošo T (2011) The evolutionary origin of orphan genes. Nat Rev Genet 12(10):692-702

te Beest M, Le Roux JJ, Richardson DM, Brysting AK, Suda J, Kubesová M, Pysek P (2012) The more the better? The role of polyploidy in facilitating plant invasions. Ann Bot 109(1):19-45

Telenius H, Pelear AH, Tunnacliffe A, Carter NP, Behmel A, Ferguson-Smith MA, Nordenskjøld M, Pfragner R, Ponder BAJ (1992) Cytogenetic analysis by chromosome painting using DOP-PCR amplified flow sorted chromosomes. Genes Chromosomes Cancer 4:257-263

Thomas CA (1971) The genetic organization of chromosomes. Annu Rev Genet 5:237-256

Thompson JN, Nuismer SL, Merg K (2004) Plant polyploidy and the evolutionary ecology of plant/animal interactions. Biological J Linn Soc 82:511-519

Tiang CL, He Y, Pawlowski WP (2012) Chromosome organization and dynamics during interphase, mitosis, and meiosis in plants. Plant Physiol 158(1):26-34

Ugarkovic D (2005) Functional elements residing within satellite DNAs. EMBO Rep 6(11):1035-1039

Uhlmann F, Lottspeich F, Nasmyth K (1999) Sister chromatid separation at anaphase onset is promoted by cleavage of the cohesin subunit Scc1p. Nature 400:37-42

Urbanelli S (2002) Genetic divergence and reproductive isolation in the Ochthebius (*Calobius*) complex (Coleoptera: *Hydraenidae*). Heredity 88:333-341

van Berkum NL, Lieberman-Aiden E, Williams L, Imakaev M, Gnirke A, Mirny LA, Dekker J, Lander ES (2010) Hi-C: a method to study the three-dimensional architecture of genomes. J Vis Exp 39:1869

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Gabor Miklos GL, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M,Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R,Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan P, Heiman TJ, Higgins ME, Ji RR, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin X, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang Z, Wang A, Wang X, Wang J, Wei M, Wides R, Xiao C, Yan C,Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu S, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A,Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Center A, Cheng ML, Curry L, Danaher S,Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B,Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers YH, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E,Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwartz R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang YH, Coyne M,Dahlke C, Mays A, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B,Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan J, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Liu X, Lopez J, Ma D, Majoros W,McDaniel J, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, Zhu X (2001) The sequence of the human genome. Science 291(5507):1304-1351

Volpi EV, Bridger JM (2008) FISH glossary: an overview of the fluorescence *in situ* hybridization technique. Biotechniques 45(4):385-409

Wallace LE, Culley TM, Weller SG, Sakai AK, Kuenzi A, Roy T, Wagner WL, Nepokroeff M (2011) Asymmetrical gene flow in a hybrid zone of Hawaiian *Schiedea* (*Caryophyllaceae*) species with contrasting mating systems. PLoS One 6(9):e24845

Wang RR-C (2011) *Agropyron* and *Psathyrostachys*. In: Kole C (ed) Wild crop relatives: genomic and breeding resources, cereals. pp 77-109 Springer-Verlag, Berlin, Heidelberg

Wegel E, Shaw P (2005) Gene activation and deactivation related changes in the three-dimensional structure of chromatin. Chromosoma 114:331-337

Wei Y, Yu L, Bowen J, Gorovsky MA, Allis CD (1999) Phosphorylation of histone H3 is required for proper chromosome condensation and segregation. Cell 97(1):99-109

Weiss-Schneeweiss H, Emadzade K, Jang T-S, Schneeweiss GM (2013) Evolutionary consequences, constraints and potential of polyploidy in plants. Cytogenet Genome Res 140:137-150

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH (2007) A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8(12):973-982

Wilson BA, Garud NR, Feder AF, Assaf ZJ, Pennings PS (2016) The population genetics of drug resistance evolution in natural populations of viral, bacterial and eukaryotic pathogens. Mol Ecol 25(1):42-66

Winkler HL (1920) Verbreitung und Ursache der Parthenogenesis im Pflanzen und Tierreiche. Verlag von G Fischer, Jena

Wissler L, Gadau J, Simola DF, Helmkampf M, Bornberg-Bauer E (2013) Mechanisms and dynamics of orphan gene emergence in insect genomes. Genome Biol Evol 5(2):439-455

Woodcock CL, Ghosh RP (2010) Chromatin higher-order structure and dynamics. Cold Spring Harb Perspect Biol 2(5):a000596

Wu DD, Irwin DM, Zhang JP (2011) *De novo* origin of human protein-coding genes. PLoS Genet 7(11):e1002379

Yakimowski SB and Rieseberg LH (2014) The role of homoploid hybridization in evolution: a century of studies synthesizing genetics and ecology. Am J Bot 101(8):1247-1258

Yeo RW, Yang K, Li G, Lim SK (2012) High glucose predisposes gene expression and ERK phosphorylation to apoptosis and impaired glucose-stimulated insulin secretion via the cytoskeleton. PLoS One 7(9):e44988

Yu J, Hu S, Wang J, Wong GK, Li S, Liu B, Deng Y, Dai L, Zhou Y, Zhang X, Cao M, Liu J, Sun J, Tang J, Chen Y, Huang X, Lin W, Ye C, Tong W, Cong L,Geng J, Han Y, Li L, Li W, Hu G, Huang X, Li W, Li J, Liu Z, Li L, Liu J, Qi Q, Liu J, Li L, Li T, Wang X, Lu H, Wu T, Zhu M, Ni P, Han H, Dong W, Ren X, Feng X,Cui P, Li X, Wang H, Xu X, Zhai W, Xu Z, Zhang J, He S, Zhang J, Xu J, Zhang K, Zheng X, Dong J, Zeng W, Tao L, Ye J, Tan J, Ren X, Chen X, He J, Liu D,Tian W, Tian C, Xia H, Bao Q, Li G, Gao H, Cao T, Wang J, Zhao W, Li P, Chen W, Wang X, Zhang Y, Hu J, Wang J, Liu S, Yang J, Zhang G, Xiong Y, Li Z,Mao L, Zhou C, Zhu Z, Chen R, Hao B, Zheng W, Chen S, Guo W, Li G, Liu S, Tao M, Wang J, Zhu L, Yuan L, Yang H (2002) A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 296(5565):79-92

Zamir D (2001) Improving plant breeding with exotic genetic libraries. Nat Rev Genet 2(12):983-989

Zeller FJ, Cermeno MC, Friebe B (1987) Cytological identfication of telotrisomic and double ditelosomic lines in *Secale* cereale cv. Heines Hellkon by means of Giemsa C-banding patterns and crosses with wheatrye addition lines. Genome 29:58-62

Zettler LA, Sogin ML, Caron DA (1997) Phylogenetic relationships between the *Acantharea* and the *Polycystinea*: a molecular perspective on Haeckel's *Radiolaria*. Proc Natl Acad Sci U S A 94 (21):11411-11416

Zhang MQ (2002) Computational prediction of eukaryotic protein-coding genes. Nat Rev Genet 3(9):698-709

Zhang P, Dundas IS, McIntosh RA, Xu SS, Park RF, Gill BS, Friebe B (2015) Wheat-*Aegilops* introgressions. In: Molnár-Láng M, Ceoloni C, Doležel J (eds) Alien introgression in wheat: cytogenetics, molecular biology, and genomics. pp 221-243 Springer Science+Business Media, Switzerland

Zhang X, Wen H, Shi X (2012) Lysine methylation: beyond histones. Acta Biochim Biophys Sin 44(1):14-27

Zhang Z, Harrison PM, Liu Y, Gerstein M (2003) Millions of years of evolution preserved: a comprehensive catalog of the processed pseudogenes in the human genome. Genome Res 13(12):2541-2558

Zhao Z, Tavoosidana G, Sjölinder M, Göndör A, Mariano P, Wang S, Kanduri C, Lezcano M, Sandhu KS, Singh U, Pant V, Tiwari V, Kurukuti S, Ohlsson R (2006) Circular chromosome conformation capture (4C) uncovers extensive networks of epigenetically regulated intra- and interchromosomal interactions. Nat Genet 38(11):1341-1347

Zwierzykowski Z, Kosmala A, Zwierzykowska E, Jones N, Jokm W, Bocianowski J (2006) Genome balance in six successive generations of the allotetraploid *Festuca pratensis* × *Lolium perenne*. Theor Appl Genet 113:539-547

3 AIMS OF THE THESIS

I **Cytogenetic analysis of introgression lines of wheat-*Th. ponticum* with blue aleurone trait using multicolour fluorescence *in situ* hybridization (FISH)**

II **Optimization and utilization of fluorescence *in situ* hybridization in suspension (FISHIS) for flow cytometric analysis of selected species of the *Triticeae* tribe**

4 RESULTS

4.1 Published papers

4.1.1 Variation in genome composition of blue-aleurone wheat

(Appendix I)

4.1.2 Advances in plant chromosome genomics

(Appendix II)

4.1.3 Molecular organization and comparative analysis of chromosome 5B of the wild wheat ancestor *Triticum dicoccoides*

(Appendix III)

4.1.4 Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their synteny with wheat

(Appendix IV)

4.1.1 Variation in genome composition of blue-aleurone wheat

Burešová V, Kopecký D, Bartoš J, Martinek P, Watanabe N, Vyhnánek T, Doležel J

Theoretical and Applied Genetics, 128(2):273-282, 2015

IF: 3.9

Abstract:

Anthocyanins are of great importance for human health due to their antioxidant, anti-inflammatory, anti-microbial and anti-cancerogenic potential. In common wheat (*Triticum aestivum* L.) their content is low. However, elite lines with blue aleurone exhibit significantly increased levels of anthocyanins. These lines carry introgressed chromatin from wild relatives of wheat such as *Thinopyrum ponticum* and *Triticum monococcum*. The aim of our study was to characterize genomic constitutions of wheat lines with blue aleurone using genomic and fluorescence *in situ* hybridization. We used total genomic DNA of *Th. ponticum* and two repetitive DNA sequences (GAA repeat and the *Afa* family) as probes to identify individual chromosomes. This enabled precise localization of introgressed *Th. ponticum* chromatin. Our results revealed large variation in chromosome constitutions of the blue-aleurone wheats. Of 26 analysed lines, 17 carried an introgression from *Th. ponticum*; the remaining nine lines presumably carry *T. monococcum* chromatin undetectable by the methods employed. Of the *Th. ponticum* introgressions, six different types were present, ranging from a ditelosomic addition (cv. Blue Norco) to a disomic substitution (cv. Blue Baart), substitution of complete (homologous) chromosome arms (line UC66049) and various translocations of distal parts of a chromosome arm(s). Different types of introgressions present support a hypothesis that the introgressions activate the blue aleurone trait present, but inactivated, in common wheat germplasm.

4.1.2 Advances in plant chromosome genomics

Doležel J, Vrána J, Cápal P, Kubaláková M, Burešová V, Šimková H

Biotechnology advances, 32(1):122-136, 2014

IF: 9.848

Abstract:

Next generation sequencing (NGS) is revolutionizing genomics and is providing novel insights into genome organization, evolution and function. The number of plant genomes targeted for sequencing is rising. For the moment, however, the acquisition of full genome sequences in large genome species remains difficult, largely because the short reads produced by NGS platforms are inadequate to cope with repeat-rich DNA, which forms a large part of these genomes. The problem of sequence redundancy is compounded in polyploids, which dominate the plant kingdom. An approach to overcoming some of these difficulties is to reduce the full nuclear genome to its individual chromosomes using flow-sorting. The DNA acquired in this way has proven to be suitable for many applications, including PCR-based physical mapping, *in situ* hybridization, forming DNA arrays, the development of DNA markers, the construction of BAC libraries and positional cloning. Coupling chromosome sorting with NGS offers opportunities for the study of genome organization at the single chromosomal level, for comparative analyses between related species and for the validation of whole genome assemblies. Apart from the primary aim of reducing the complexity of the template, taking a chromosome-based approach enables independent teams to work in parallel, each tasked with the analysis of a different chromosome(s). Given that the number of plant species tractable for chromosome sorting is increasing, the likelihood is that chromosome genomics - the marriage of cytology and genomics - will make a significant contribution to the field of plant genetics.

4.1.3 Molecular organization and comparative analysis of chromosome 3B of the wild wheat ancestor *Triticum dicoccoides*

Akpinar BA, Yuce M, Lucas S, Vrána J, Burešová V, Doležel J, Budak H

Scientific reports, 5:10763, 2015

IF: 5.228

Abstract:

Wild emmer wheat, *Triticum turgidum* ssp. *Dicoccoides* is the wild relative of *Triticum turgidum*, the progenitor of durum and bread wheat, and maintains a rich allelic diversity among its wild populations. The lack of adequate genetic and genomic resources, however, restricts its exploitation in wheat improvement. Here, we report next-generation sequencing of the flow-sorted chromosome 5B of *T. dicoccoides* to shed light into its genome structure, function and organization by exploring the repetitive elements, protein-encoding genes and putative microRNA and tRNA coding sequences. Comparative analyses with its counterparts in modern and wild wheats suggest clues into the B-genome evolution. Syntenic relationships of chromosome 5B with the model grasses can facilitate further efforts for fine-mapping of traits of interest. Mapping of 5B sequences onto the root transcriptomes of two additional *T. dicoccoides* genotypes, with contrasting drought tolerances, revealed several thousands of single nucleotide polymorphisms, of which 584 shared polymorphisms on 228 transcripts were specific to the drought-tolerant genotype. To our knowledge, this study presents the largest genomics resource currently available for *T. dicoccoides*, which, we believe, will encourage the exploitation of its genetic and genomic potential for wheat improvement to meet the increasing demand to feed the world.

4.1.4 Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their syntheny with wheat

Molnár I, Vrána J, Burešová V, Cápal P, Farkas A, Darkó É, Cseh A, Kubaláková M, Molnár-Láng M, Doležel J

Plant Journal, doi:10.1111/tpj.13266, 2016

IF: 5.468

Abstract:

Goat grasses (*Aegilops* spp.) contributed to the evolution of bread wheat and are important sources of genes and alleles for modern wheat improvement. However, their use in alien introgression breeding is hindered by poor knowledge of their genome structure and a lack of molecular tools. The analysis of large and complex genomes may be simplified by dissecting them into single chromosomes via flow cytometric sorting. In some species this is not possible due to similarities in relative DNA content among chromosomes within a karyotype. This work describes the distribution of GAA and ACG microsatellite repeats on chromosomes of the U, M, S and C genomes of *Aegilops*, and the use of microsatellite probes to label the chromosomes in suspension by fluorescence *in situ* hybridization (FISHIS). Bivariate flow cytometric analysis of chromosome DAPI fluorescence and fluorescence of FITC-labelled microsatellites made it possible to discriminate all chromosomes and sort them with negligible contamination by other chromosomes. DNA of purified chromosomes was used as a template for PCR using COS markers with known positions on wheat A, B and D genomes. Wheat-*Aegilops* macrosyntenic comparisons using COS markers revealed significant rearrangements in the U and C genomes, while the M and S genomes exhibited structure similar to wheat. Purified chromosome fractions provided an attractive resource to investigate the structure and

evolution of the *Aegilops* genomes, and the COS markers assigned to *Aegilops* chromosomes will facilitate alien gene introgression into wheat.

4.2 Published abstracts – poster presentations

4.2.1 Genomic constitutions of cereals with blue aleurone trait

(Appendix V)

4.2.2 Genomic constitution of wheat genotypes with blue aleurone

(Appendix VI)

4.2.3 Genomic constitution of blue grained wheat genotypes

(Appendix VI)

4.2.1 Genomic constitutions of cereals with blue aleurone trait

Burešová V, Kopecký D, Šafář J, Vyhnánek T, Martinek P, Doležel J

In: Abstracts of the "Olomouc Biotech 2011. Plant Biology: Green for Good II". Olomouc, Czech Republic, 2013.

Abstract:

Anthocyanins are of great importance for human health due to their antioxidant potential. Their content is rather low in common varieties of wheat (*Triticum aestivum* L.). However elite lines with blue aleurone and introgressed chromatin from wild relatives exhibit significantly increased levels of anthocyanins. There is evidence that the donor of chromosome introgressions has been *Thinopyrum ponticum* (syn. *Agropyron elongatum*). The aim of our study was to characterize genomic constitution of selected cereals (wheat, barley and triticale) with blue aleurone using genomic *in situ* hybridization (GISH) and fluorescence *in situ* hybridization (FISH). We used total genomic DNA of *Th. ponticum* and two repetitive DNA sequences (GAA repeat, *Afa* family) as probes to identify individual chromosomes. This enabled precise localization of introgressed chromatin. Our results revealed large variation in genomic constitution of blue aleurone wheat genotypes. Out of 14 analysed lines, 11 lines carried an introgression from *Th. ponticum*. In the remaining three genotypes (cvs. Indigo, Skorpion, and line H83-952-1), we were unable to detect any introgressed chromosome segment. Six different types of introgressions were found, ranging from the addition of a telocentric chromosome pair (cv. Blue Norco) to substitution of one chromosome pair (cv. Blue Baart), substitution of complete (homologous) chromosome arms (line UC660-49) and various substitutions of distal parts of chromosome arm(s). Different types of introgressions observed in or work support a hypothesis that the introgressions activate the blue aleurone trait pathway, which is present, but deactivated in common wheat germplasm.

4.2.2 Genomic constitution of wheat genotypes with blue aleurone

Burešová V, Kopecký D, Bartoš J, Martinek P, Watanabe N, Vyhnánek T, Doležel J

In: Abstracts of the "Plant molecular cytogenetics in genomic and postgenomic era". Katowice, Poland, 2014.

Abstract:

Anthocyanins are recognized as health-enhancing components for human due to their antioxidant and anti-inflammatory activities. They can be found in fruits, vegetables and some cereals. Recently, wheat with different grain colours (especially blue and purple) has been identified as a new source of anthocyanins and several lines with blue aleurone layer have been developed. The blue colorization has been observed after the introgression of chromatin from wild relatives into wheat. At least three different donors have been identified including *Thinopyrum ponticum*, *Triticum monococcum* and *Th. bessarabicum*. We employed GISH/FISH to detect the introgression of *Th. ponticum* and identified individual wheat chromosome(s) carrying the introgression. Our results clearly demonstrate that there are at least six different types of introgression, ranging from the addition of entire chromosome (cvs. Blue Baart and Blue Norco) to substitution of chromosome arm (cv. UC66049) and chromosome segment(s) (cv. Xiao Yan). These introgressions were located on the wheat chromosomes of homoeologous group 4 and/or are in the form of disomic additions. In some lines (i.e. cvs. Skorpion and Tschermaks Blaukörniger Sommerweizen), we were unable to detect introgressed chromatin of *Th. ponticum* indicating different source of blue aleurone trait (presumably *T. monococcum*). Currently, we are optimizing protocol for sorting individual chromosomes with introgression using FISHIS (FISH in suspension). Our next goal is to identify the origin of *Th. ponticum* introgressions in various genotypes.

4.2.3 Genomic constitution of blue grained wheat genotypes

Burešová V, Kopecký D, Bartoš J, Martinek P, Watanabe N, Vyhnánek T, Doležel J

In: Sborník abstrakt, Bulletin České společnosti experimentální biologie rostlin, "6. Metodické dny". Seč, Czech Republic, 2014.

Abstract:

Anthocyanins determine red, purple and blue colouring in many species of fruits, vegetables, honey, olive oil, flowers and others. Nowadays, the interest in such pigment is increasing and it is not only for its natural colouring abilities but also for its beneficial properties for human health. Clinical studies revealed significant antioxidant, antimicrobial, anti-inflammatory and anti-carcinogenic effect of anthocyanins. Moreover, anthocyanins are proposed as a functional food component that may help to prevent heart diseases, stroke, obesity, diabetes and other lifestyle diseases. Anthocyanins have been identified also in some cereals. High content of anthocyanins has been detected in wheat with blue aleurone layer. It is known that blue colour of aleurone is determined by the presence of alien chromatin of wild relatives in common wheat. Three such donors of introgression: *Thinopyrum ponticum*, *Th. bessarabicum* and *Triticum monococcum* were identified up to date. We used GISH/FISH to detect introgression of *Th. ponticum* and determine wheat chromosome(s) carrying the introgression. Our analysis revealed large variation in genomic constitution of blue grained wheat genotypes. There are at least six different types of introgression. Among genotypes, we detected either the addition of the entire pair of chromosomes (cvs. Blue Baart and Blue Norco), or disomic substitution of chromosome arm (cv. UC66049) and chromosome segment(s) (cv. Xiao Yan). *Thinopyrum* substitutions were located on the wheat chromosomes of homoeologous group 4. In some genotypes (i.e. cvs. Skorpion and Tschermaks Blaukörniger Sommerweizen), we were unable to detect introgressed chromatin of *Th. ponticum* indicating different source of blue aleurone trait (presumably *T. monococcum*). Our next goal is flow sorting individual chromosomes with introgression using FISHIS

(FISH in suspension) and identification the origin of *Th. ponticum* introgressions in various genotypes.

4.3 Published abstracts – oral presentations

4.3.1 Nuclear organization in interspecific plant hybrids

4.3.2 Nuclear organization in interspecific plant hybrids revealed by 3D-FISH

4.3.1 Nuclear organization in interspecific plant hybrids

Burešová V, Kopecký D, Vrána J, Jenkins G, Phillips D, Doležel J

In: Abstracts of the "VI. Festulolium Working Group Workshop". Olomouc, Czech Republic, 2016.

Abstract:

The plant cell nucleus is enclosed within the nuclear envelope harbouring chromosome territories (CTs) and various nuclear bodies. It seems that the spatial organization of chromosomes is non-random and is characterized by many local and long-range contacts among genes and regulatory elements. Thus, it is evident that the architecture of interphase chromosomes plays a role in the regulation of gene expression. The introduction of sophisticated high-resolution microscopy and state-of-the-art genomics enables the complementary strategies to study CTs with high resolution. The aim of the project is to characterize spatial nuclear organization in interspecific hybrids and provide an insight into the positioning of chromatin from both parents using 3D-FISH. In our presentation, we will present the results of our pilot experiment on the positioning of chromosome domains in interspecific hybrids (rye-wheat disomic chromosome arm substitution lines). The combination of flow sorting, *in situ* hybridization and high resolution confocal microscopy accompanied with spatial software module allowed us to achieve high resolution and avoid potential bias caused by the evaluation of nuclei in various cell cycle stages. Our next goal will be to analyse *Festuca-Lolium* hybrids with various proportions of parental genomes.

4.3.2 Nuclear organization in interspecific plant hybrids revealed by 3D-FISH

Burešová V, Kopecký D, Vrána J, Jenkins G, Phillips D, Doležel J

In: Abstracts of the "Society for Experimental Biology". Brighton, United Kingdom, 2016.

Abstract:

The spatial organization of interphase chromosomes are non-random and occupy defined regions within the nucleus, which are termed chromosomal territories (CTs). CTs represent sites of local and long-range contact between genes and regulatory elements thus plays a key role in the regulation of gene expression. Interspecific hybrids are commonly used in plant breeding programmes and aim to combine agriculturally important traits from two species into elite crop cultivars. Despite the importance of interspecific hybridization in modern breeding, only little is known about the spatial organization of hybrid nuclei and its consequences to the parental allele-specific gene expression in hybrids. The main aims of our project are: (i) characterize spatial nuclear organization in interspecific hybrids, (ii) provide an insight into the positioning of chromatin from both parents (iii) investigate the effect of spatial localization of specific regions in CTs and entire nucleus on gene expression. We will present the results of our pilot experiment on the positioning of chromosome domains in interspecific hybrids (rye-wheat disomic chromosome arm substitution lines). The combination of flow sorting, *in situ* hybridization and high resolution confocal microscopy accompanied with spatial software module allowed us to achieve high resolution and avoid potential bias caused by the evaluation of nuclei in various cell cycle stages. Our preliminary data confirmed existence of Rabl configuration when the centromeres and telomeres occupy opposite poles of the nucleus. Surprisingly, two substituted homologous chromosome arms of rye seem to be located far from each other in 3D space of interphase nucleus of wheat.

5 CONCLUSION

One of the aims of this Ph.D. thesis was the genome analysis and characterization of genomic composition of introgression lines of wheat-*Th. ponticum* with blue aleurone trait using molecular cytogenetic techniques. Furthermore, optimization and utilization of fluorescence *in situ* hybridization in suspension (FISHIS) for genomic analyses of the genera *Triticum* and *Aegilops* has been other goal of the study.

5.1 Cytogenetic analysis of introgression lines of wheat-*Thinopyrum ponticum* with blue aleurone trait using multicolour fluorescence *in situ* hybridization (FISH)

Anthocyanins are of great importance for human health due to their antioxidant potential. Their content is rather low in common varieties of wheat (*Triticum aestivum* L.). However, elite lines with blue aleurone and introgressed chromatin from wild relatives exhibit significantly increased levels of anthocyanins. There is evidence that the donor of chromosome introgressions has been *Thinopyrum ponticum* (syn. *Agropyron elongatum*). The aim of our study was to characterize genomic constitution of selected wheat genotypes with blue aleurone with using multicolour fluorescence *in situ* hybridization. Our results revealed six different types of the *Th. ponticum* intorogressions. We suppose that introgressions of wild relatives such as *Th. ponticum* into the common lines of wheat activates the blue aleurone trait pathway, which is present, but deactivated in wheat germplasm.

5.2 Optimization and utilization of fluorescence *in situ* hybridization in suspension (FISHIS) for flow cytometric analysis of selected species of the *Triticeae* tribe

Flow cytometric analysis of fluorescence of DAPI stained chromosomes do not allow the discrimination of chromosomes which have the same or very similar DNA content. In order to overcome this obstacle was developed fast and reliable method, called fluorescence *in situ* hybridization in suspension (FISHIS). This

method makes use of fluorescent oligonucleotide SSR probes and DAPI fluorescence for flow cytometric analysis. To date, this biparametric analysis was successfully used for discrimination of similar chromosomes in genera *Aegilops* (e.g., *Ae. umbellulata*, *Ae.comosa*, *Ae. speltoides*) and *Triticum* (e.g. *T. aestivum*, *T. dicoccoides*). We believed that this method will be useful for sorting of chromosomes from other plant species.

6 LIST OF ABBREVIATIONS

3D	three dimensional
AFLP	amplified fragment length polymorphism
BAC	bacterial artificial chromosome
Cdna	complementary deoxyribonucleic acid
CI	centromeric index
CISS	chromosomal *in situ* suppression
CS	Chinese Spring
CT	chromosome territory
DAPI	4’6-diamidino-2-phenylindole
DNA	deoxyribonucleic acid
FISH	fluorescence *in situ* hybridization
FISHIS	fluorescence *in situ* hybridization in suspension
Gb	gigabase
GISH	genomic *in situ* hybridization
H	histone
HAT	histone acetyl transferase
HDAC	histone deacetylase
Kb	kilobase
LINE	long interspersed nuclear element
LTR	long terminal repeat
Mb	megabase
mRNA	messenger ribonucleic acid
n	haploid number
NGS	next generation sequencing
NOR	nuclear organizer region
ORF	open reading frame
PCR	polymerase chain reaction
Ph	pairing homoloeologous

pre-mRNA	precursor messenger ribonucleic acid
PRINS	primed *in situ* labelling
PTM	posttranscriptional modification
rDNA	ribosomal deoxyribonucleic acid
RFLP	restriction fragment length polymorphism
RNA	ribonucleic acid
rRNA	ribosomal ribonucleic acid
x	monoploid number

7 LIST OF APPENDICIES

Original papers

Appendix I Variation in genome composition of blue-aleurone wheat

Appendix II Advances in plant chromosome genomics

Appendix III Molecular organization and comparative analysis of chromosome 5B of the wild wheat ancestor *Triticum dicoccoides*

Appendix IV Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their syntheny with wheat

Published abstracts – poster presentations

Appendix V Genomic constitutions of cereals with blue aleurone trait

Appendix VI Genomic constitution of wheat genotypes with blue aleurone

Genomic constitution of blue grained wheat genotypes

APPENDIX I

Variation in genome composition of blue-aleurone wheat

Burešová V, Kopecký D, Bartoš J, Martinek P, Watanabe N, Vyhnánek T, Doležel J

Theoretical and Applied Genetics, 128(2):273-282, 2015

IF: 3.9

Theor Appl Genet (2015) 128:273–282
DOI 10.1007/s00122-014-2427-3

ORIGINAL PAPER

Variation in genome composition of blue-aleurone wheat

Veronika Burešová · David Kopecký · Jan Bartoš · Petr Martinek · Nobuyoshi Watanabe · Tomáš Vyhnánek · Jaroslav Doležel

Received: 11 August 2014 / Accepted: 8 November 2014 / Published online: 16 November 2014

Abstract
Key message **Different blue-aleurone wheats display major differences in chromosome composition, ranging from disomic chromosome additions, substitutions, single chromosome arm introgressions and chromosome translocation of *Thinopyrum ponticum*.**
Abstract Anthocyanins are of great importance for human health due to their antioxidant, anti-inflammatory, antimicrobial and anti-cancerogenic potential. In common wheat (*Triticum aestivum* L.) their content is low. However, elite lines with blue aleurone exhibit significantly increased levels of anthocyanins. These lines carry introgressed chromatin from wild relatives of wheat such as *Thinopyrum ponticum* and *Triticum monococcum*. The aim of our study was to characterize genomic constitutions of wheat lines with blue aleurone using genomic and fluorescence in situ hybridization. We used total genomic DNA of *Th. ponticum* and two repetitive DNA sequences (GAA repeat and the *Afa* family) as probes to identify individual chromosomes. This enabled precise localization of introgressed *Th. ponticum* chromatin. Our results revealed large variation in chromosome constitutions of the blue-aleurone wheats. Of 26 analyzed lines, 17 carried an introgression from *Th. ponticum*; the remaining nine lines presumably carry *T. monococcum* chromatin undetectable by the methods employed. Of the *Th. ponticum* introgressions, six different types were present, ranging from a ditelosomic addition (cv. Blue Norco) to a disomic substitution (cv. Blue Baart), substitution of complete (homologous) chromosome arms (line UC66049) and various translocations of distal parts of a chromosome arm(s). Different types of introgressions present support a hypothesis that the introgressions activate the blue aleurone trait present, but inactivated, in common wheat germplasm.

Communicated by Alan H. Schulman.

V. Burešová · D. Kopecký (✉) · J. Bartoš · J. Doležel
Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, 78371 Olomouc, Czech Republic
e-mail: kopecky@ueb.cas.cz

P. Martinek
Agrotest Fyto, Ltd., Havlíčkova 2787, 767 01 Kroměříž, Czech Republic

N. Watanabe
College of Agriculture, Ibaraki University, 3-21-1 Chuo, Ami, Inashiki, Ibaraki 300-0393, Japan

T. Vyhnánek
Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

Introduction

Anthocyanins are a group of intensely colored water-soluble pigments responsible for most of red, blue and purple colors of fruits, vegetables, flowers and other tissues. They are abundant in red, blue and purple-colored berries and their products (derivatives) such as red wine, and in seeds of some species (Mazza and Miniati 1993). Over 400 anthocyanins have been described so far. Of these, six are the most abundant in plant kingdom and are classified based on the number and position of hydroxyl and methoxyl groups on the flavan nucleus: cyanidin, the most widespread anthocyanidin in nature), delphinidin, pelargonidin, peonidin, petunidin and malvidin (Mazza 2007). Their presence in plants is beneficial as they attract animals, and thereby assist in pollination and seed dispersal (Harborne and Williams 2001) as well as offer protection against the ultraviolet-induced damage (Mazza and Miniati 1993).

Anthocyanins display a range of biological activities some of which are significant in human diet and health, such as antioxidant (Wang et al. 1997), anti-inflammatory (Wang and Mazza 2002), anti-microbial (Pisha and Pezzuto 1994) and anti-cancerogenic activities (reviewed in Wang and Stoner 2008; Bowen-Forbes et al. 2010), improvement of vision (Matsumoto et al. 2003; Lila 2004), induction of apoptosis (Katsube et al. 2003) and neuroprotective effects (Youdim et al. 2000). According to some reports, intake of anthocyanins may have protective effect against coronary heart disease, the leading cause of death in most developed countries (Anderson et al. 2000; Rechner and Kroner 2005). An intriguing question is the uptake of anthocyanins in humans after their ingestion. Their levels in human blood are far below the levels required to exhibit anti-cancerogenic effects in vitro (Wang and Stoner 2008). Thus, large and long-term intervention trials are needed for a definite proof of the potential human health benefits of these compounds (Mazza 2007). Anthocynanins can also serve as natural food colorants to prevent or decrease the usage of synthetic colors (Gao and Mazza 1994). Blue and purple corn grains are used for making blue and pink tortillas, and red rice is commonly used as a food colorant in bread, ice cream, and liquor (Yoshinaga 1986).

Consumable anthocyanins are found in fruits and vegetables. Their content varies considerably and is affected by genes and environmental conditions (Horbowicz et al. 2008). The highest total anthocyanin content was found in blueberries, chokeberries, elderberries, grapes and eggplants, exceeding 5,000 mg kg^{-1} (Clifford 2000). Besides fruits and vegetables, anthocyanins may also be present in substantial amounts in cereals such as purple corn, red and black rice and wheat with purple pericarp or blue aleurone (Abdel-Aal et al. 2006). Blue-grained wheat genotypes are of particular interest due to their relatively high total anthocyanin content. Furthermore, anthocyanin pigments can be concentrated by dry milling and fractionation processes to produce fractions with high anthocyanin levels (Abdel-Aal et al. 2006). Another reason to place blue-aleurone wheat into the focus is the relative composition of anthocyanins. In the plant kingdom, the most abundant anthocyanidin is cyanidin-3-glucoside, which is the main anthocyanin in fruits such as various berries and black currant, vegetables, red and black rice, cob corn and purple pericarp wheat (Escribano-Bailon et al. 2004). On the other hand, the major anthocyanidin of the blue-aleurone wheat is delphinidin-3-glucoside (Trojan et al. 2014). It is the most potent angiogenic inhibitor among anthocyanins and may be helpful in cancer prevention and treatment (Lamy et al. 2006). Delphinidin is also said to be more effective in the inhibition of tumorogenesis, by blocking the activation of the mitogen-activated protein kinase (Hou et al. 2004). Additionally, Afaq et al. (2007) investigated the photo-chemopreventive effect of delphinidin on UVB-induced biomarkers of skin cancer development.

The biochemical pathway of anthocyanins is well known (Ficco et al. 2014). The early steps of their synthesis are regulated by a cascade of enzymes including chalcone synthase (CHS), chalcone-isomerase (CHI), flavanone 3-hydroxylase (F3H) and dihydroflavonol-4-reductase (DFR). In wheat, the genes for CHS, F3H and DFR were cloned and mapped to the proximal region of the long arm of the homoeologous group 3 (DFR) (Yang et al. 2004; Himi and Noda 2004). All these genes were identified in the parental genotypes of the blue grain wheat—standard wheat and *Th. ponticum*, which both do not express blue aleurone phenotype. This implies that there must be some regulatory gene(s) that control the expression of these genes in developing seeds of standard and blue-grained wheats, but the regulatory pattern in the blue-grained seeds may not be the same as that in standard wheat and *Th. ponticum* (Yang et al. 2004). This is in agreement with studies on maize and other flowering plants, where at least eight structural genes and two families of regulatory genes controlling the flavonoid biosynthesis were identified (Gao et al. 2000). Moreover, the situation is complicated by the effect of the environment, where the level of expression in blue-aleurone wheats is influenced by temperature, light intensity, pH and other factors (Zeven 1991). Different levels of anthocyanin concentration were found along the developmental process with the maximum peak observed during the mid-grain-development stage (Knievel et al. 2009; Trojan et al. 2014).

As indicated above, the expression of blue coloration of the aleurone layer (*Ba*) in blue-grained wheat is associated with the presence of a chromosome or chromosome segment introgressed from alien species. Three genes involved in regulating the expression of blue coloration of the aleurone in wheat have been identified. They originate from different species: *Ba1* (syn. *Ba(b)*) is a dominant gene originating from *Thinopyrum ponticum* ($2n = 10x = 70$, $StStStStE^eE^eE^bE^bE^xE^x$ (previously designated as Ag); syn. *Lophopyrum ponticum*; *Elytrigia pontica*; *Agropyron elongatum*) and was physically mapped to region 0.71–0.8 of the long arm of 4Ag from centromere (Zheng et al. 2006a), *Ba2* (syn. *Ba(a)*) is an incompletely dominant gene mapped close to the centromere on long arm of $4A^m$ and to $4A^{bo}$ in *Triticum monococcum* and *T. boeoticum*, respectively (Dubcovsky et al. 1996; Singh et al. 2007), while *BaThb* is expressed by introgression of *Th. bessarabicum* ($2n = 2x = 14$, E^bE^b = JJ) into wheat and was mapped to chromosome 4J between centromere and FL0.52 (William and Mujeeb-Kazi 1993). *Th. bessarabicum* is the probable donor species that contributed the E^b genome to many polyploid wheatgrasses including *Th. ponticum* (Zhang et al. 1996). Thus, *Ba* genes from *Th. bessarabicum* (*BaThb*) and from *Th. ponticum* (*Ba1*) may have a common origin.

Genomic constitution of blue grain wheat genotypes is largely unknown. This is mainly due to limited information passed from one breeder to another and because most of the breeders in the early blue-grained wheat breeding programs are no longer active. Moreover, it is possible that as a consequence of the exchange of breeding materials, the same or closely related accessions were used at several research programs (Zeven 1991). However, it is known that in case of *Ba1*, substitution lines were developed by replacing wheat homoeologous chromosomes 4B and 4D by *Th. ponticum* chromosome 4Ag (Cermeno and Zeller 1986; Arbuzova et al. 2012). Similarly, in *Ba2* wheat genotypes, 4A and 4B were replaced by $4A^{bo}$ chromosome from *T. boeoticum* or *T. monococcum* (Zeven 1991).

Segregation ratios indicate that *Ba* is controlled by a single dominant gene (Zeven 1991; Dubcovsky et al. 1996) and that at least the *Ba1* allele expresses a strong xenia effect when endosperm traits are influenced by genes from the male parent (Keppenne and Baenziger 1990; Knievel et al. 2009). As the aleurone is part of triploid endosperm tissue, four combinations of alleles are possible. Three doses of *Ba1* produce dark blue seed, two doses give medium-blue seed, one dose gives light blue seed, and the absence of the gene results in the lack of blue color. Thus, *Ba1* shows a clear dosage effect (Knott 1958).

Here, we summarize the results of a comprehensive study on the genomic constitution of almost all publically available genotypes of blue grain wheat. We combined GISH and FISH to detect introgressions of *Th. ponticum* chromosomes and chromosome segments and to identify wheat chromosome(s) involved in substitutions.

Materials and methods

Plant material

Seed samples of blue grain wheat genotypes were obtained from Prof. Adam J. Lukaszewski, University of California, Riverside, USA; Prof. C.O. Qualset, University of California, Davis, USA; Prof. F.J. Zeller, Technical University of München, Freising-Weihenstephan, Germany; Dr. Robert Metzger, Oregon State University, Corvallis, Oregon, USA; Prof. A. Börner, Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany and from Genebank of the Crop Research Institute, Prague-Ruzyně, Czech Republic (Table 1).

In situ hybridization

Seeds were germinated on wet filter paper in Petri dishes, root tips were collected in ice water for 26–30 h and fixed in a mixture of absolute alcohol:glacial acetic acid (3:1) at 37 °C for 7 days. Cytological preparations and in situ hybridization with labeled DNA were made according to Masoudi-Nejad et al. (2002). In all experiments, genomic in situ hybridization (GISH) was done with a probe prepared from total genomic DNA of *Th. ponticum*. The probe was labeled with biotin by nick translation and detected with streptavidin-Cy3 using standard kits from Roche Applied Science following the manufacturer's instructions. The hybridization mix contained unlabeled genomic DNA of *T. aestivum* cv. Chinese Spring sheared to ca. 200–500 bp fragments at 1:150 ratio (probe:blocking DNA). Following the hybridization, preparations were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) in VectaShield antifade (Vector Laboratories) and observed under Zeiss Axio Imager.Z2 microscope. For identification of individual chromosomes, two additional probes were employed: A digoxigenin-labeled probe for GAA microsatellites, prepared using PCR with $(GAA)_7$ and $(CCT)_7$ primers and wheat genomic DNA as a template, and a probe for a 260-bp fragment of the *Afa* family repeat, prepared and labeled by Texas Red using PCR with primers AS-A and AS-B on wheat genomic DNA according to Kubaláková et al. (2005).

Results

Chromosome constitutions

We found large variation among karyotypes of the blue grain wheat genotypes (Fig. 1; Table 1). 'Xiao Yan' is a homozygote for translocation of both arms of wheat chromosome 4D (Fig. 1b) where the distal about one halves of the arms of 4D were replaced by (probably) their homoeologues from *Th. ponticum* (4AgS and 4AgL). In UC66049 (Qualset et al. 2005) and its derivatives, the entire 4BL arm was replaced by an arm of a *Thinopyrum* chromosome. We can only speculate that this translocation is 4BS.4AgL, more so that tetraploid UC66049/LD222 (B_6F_4) produced by backcrossing of durum wheat LD222 to UC66049 was disomic for the same translocation and both lines were fertile. Two lines UC66049/RU440-4 (B_3F_2) were produced by backcrossing the blue aleurone RU440-4, a sib line of Skorpion (RU 440-6), to UC66049. These two lines were created to combine the *Ba1* and *Ba2* genes. The presence of monosomic or disomic chromosome arm substitution of *Thinopyrum* (4BS.4AgL) indicates introgression of *Ba1*. The presence of chromosome 4A from *T. boeoticum* carrying *Ba2* could not be detected using the probes employed here, but could be detected using aneuploid lines.

In Sebesta Blue 3 (SB3) and four other genotypes, chromosome arm 4BL carries a *Thinopyrum* introgression (Fig. 1) covering about two-thirds of 4BL. Thus, these

Table 1 Genomic constitution and color intensity of blue aleurone genotypes

Accessions	Karyotype	Type of introgression	Provided by
Xiao Yan	$2n = 6x = 42$	Disomic substitution on 4DS and 4DL	F.J. Zeller
UC66049	$2n = 6x = 42$, 4BS.4AgL	Disomic chromosome arm substitution of 4BL	C.O. Qualset
UC66049/RU440-4 (B_3F_2)	$2n = 6x = 42$, 4BS.4AgL	Disomic chromosome arm substitution of 4BL	N. Watanabe
RU440-4/UC66049 (B_3F_2)	$2n = 6x = 42$, 4BS.4AgL	Monosomic or disomic chromosome arm substitution of 4BL	N. Watanabe
UC66049/LD222 (B_6F_4)	$2n = 4x = 28$, 4BS.4AgL	Disomic chromosome arm substitution of 4BL	N. Watanabe
EF02-54-9 (Sebesta Blue 3)	$2n = 6x = 42$, 4BS.4BL.4AgL	Disomic introgression on 4BL	Martinek (Šebesta)
H90-35-1 (Metzger Blue3)	$2n = 6x = 42$, 4BS.4BL.4AgL	Disomic introgression on 4BL	Martinek (Metzger)
M90-41	$2n = 6x = 42$, 4BS.4BL.4AgL	Disomic introgression on 4BL	Martinek (Metzger)
M90-41-1 (Metzger Blue8)	$2n = 6x = 42$, 4BS.4BL.4AgL	Disomic introgression on 4BL	Lukaszewski
M90-99-2[a] (Metzger Blue9)	$2n = 4x = 28$, 4BS.4BL.4AgL $2n = 4x = 28$	Disomic, monosomic or no introgression on 4BL	Martinek (Metzger)
EF02-5426-3 (Sebesta Blue 1)	$2n = 6x = 44$	Introgression on two pairs of *T. aestivum* chromosomes	Martinek (Šebesta)
EF02-5430-2 (Sebesta Blue 2)	$2n = 6x = 44$	Introgression on two pairs of *T. aestivum* chromosomes	Martinek (Šebesta)
48 M	$2n = 6x = 44$	Introgression on two pairs of *T. aestivum* chromosomes	Martinek (Woś)
H90-15-1 (Metzger Blue1)	$2n = 6x = 44$	Introgression on two pairs of *T. aestivum* chromosomes	Lukaszewski
H90-15-2 (Metzger Blue2)	$2n = 6x = 44$	Introgression on two pairs of *T. aestivum* chromosomes	Martinek (Metzger)
Blue Baart	$2n = 6x = 44$	Disomic addition of *Th. ponticum* chromosome (4 J?)	Martinek (Lukaszewski)
Blue Norco[a] (Metzger Blue5)	$2n = 6x = 42 + 2t$ $2n = 6x = 42 + 1t$ $2n = 6x = 42$	Monosomic or disomic addition of telosomic *Th. ponticum* chromosome	Martinek (Lukaszewski)
1066/91 amphiploid (Metzger Blue7)	$2n = 6x = 42$, 34T.a. + 8Th.p.	Eight *T. aestivum* chromosomes replaced by their *Th. ponticum* counterparts	Lukaszewski
Skorpion (RU 440-6)	$2n = 6x = 42$	Not detected	Martinek (Škorpík)
Tschermaks Blaukörniger Sommerweizen	$2n = 6x = 42$	Not detected	Martinek (Börner)
Barevna 9	$2n = 6x = 42$	Not detected	Martinek (Škorpík)
Barevna 11	$2n = 6x = 42$	Not detected	Genebank Ruzyně
Barevna 17	$2n = 6x = 42$	Not detected	Genebank Ruzyně
Barevna 23	$2n = 6x = 42$	Not detected	Genebank Ruzyně
Barevna 25	$2n = 6x = 42$	Not detected	Martinek (Škorpík)
H83-952-1	$2n = 6x = 42$	Not detected	Martinek (Metzger)

[a] Variation in genome composition has been detected in these genotypes

genotypes were developed by backcrossing of UC66049 to *T. aestivum*, where homoeologous recombination between 4AgL of UC66049 and 4BL of *T. aestivum* took place. Tetraploid M90-99-2 was probably produced by crossing of SB3 or its close relative with durum wheat. This genotype was unstable in genomic constitution with two, one, or no translocated segments present in individual plants (Fig. 2).

Sebesta Blue 2 (SB2) and its relatives, and Sebesta Blue 1 (SB1), have the most complicated karyotypes. SB2 and its relatives have 44 chromosomes of which 40 are normal wheat chromosomes with a chromosome pair (likely 4D) missing. We detected two pairs of translocated chromosomes involving wheat and *Th. ponticum* chromatin. One pair of these translocated chromosomes has the centromere and pericentromeric parts of *Th. ponticum* and a segment

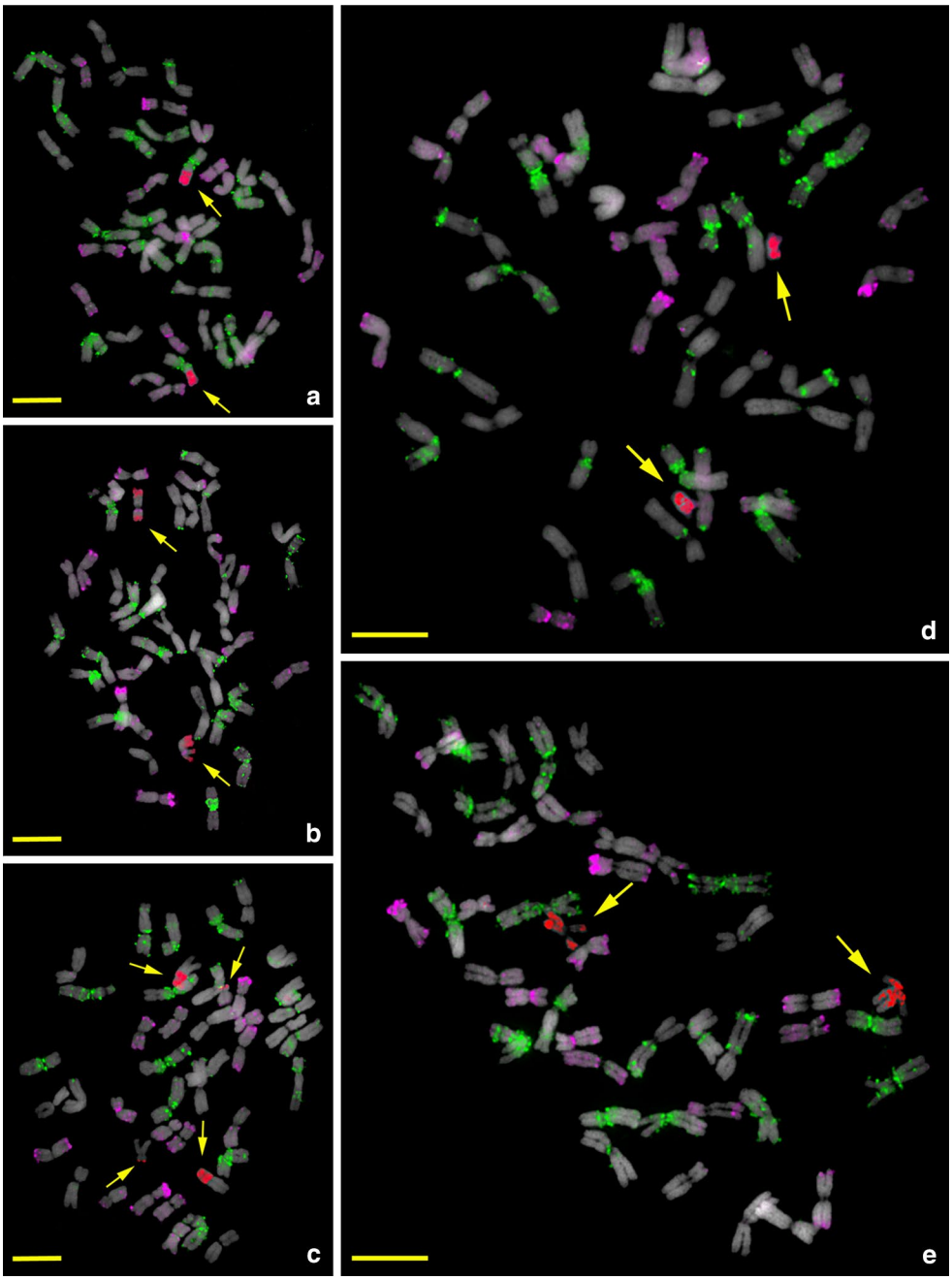


Fig. 1 Cytological analysis of blue-aleurone wheat genotypes. In situ hybridization on **a** M90-41, **b** Xiao Yan, **c** H90-15-2. **d** Blue Norco and **e** Blue Baart was performed using GAA (*green* color), *Afa* repeat (*purple* color) and total genomic DNA of *Th. ponticum* (*red* color) as probes. Chromosomes were counterstained by DAPI (*grey pseudocolor*). *Bar* 10 μm

from wheat chromosome (presumably 4DL) on one arm. The other translocated pair has one arm (presumably 4DS) and pericentromeric region of the second arm from wheat with a small terminal translocation from *Th. ponticum* (Fig. 1c). The karyotype of SB1 is even more complicated. It seems to have the same two translocations between wheat and *Thinopyrum* as SB2, but also 1–3 telocentrics from the B-genome of wheat.

Blue Baart is a disomic addition of a pair of *Th. ponticum* chromosomes (Fig. 1e), presumably 4Ag. Blue Norco is a ditelosomic addition from *Th. ponticum* (Fig. 1d). However, plants with only one *Thinopyrum* telocentric chromosome as well as plants without any *Thinopyrum* chromosome were also presented (see below). Amphiploid 1,066/91 has 42 chromosomes of which 34 are of *T. aestivum* and eight of *Th. ponticum* without any identifiable translocations (Fig. 3).

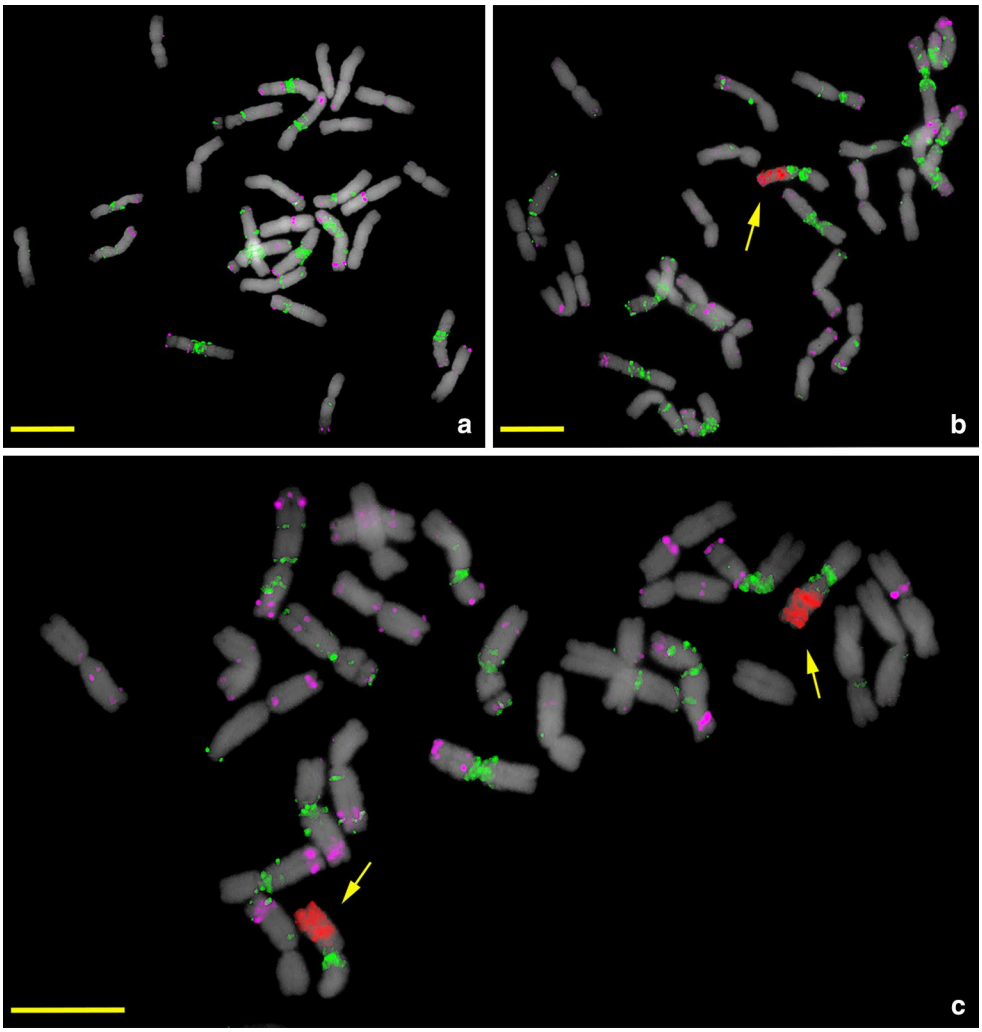


Fig. 2 Variation in genomic composition of M90-99-2 blue-aleurone wheat. In situ hybridization was performed using GAA (*green* color), *Afa* repeat (*purple* color) and total genomic DNA of *Th. ponticum* (*arrows*; *red* color) as probes. Chromosomes were counterstained by DAPI (*grey pseudocolor*). Intensity of *blue coloring* of seeds corresponds with the dose of *Th. ponticum* segment: seeds with standard (*red*) color had no *Th. ponticum* chromatin (**a**), monosomic substitution was detected in *light blue* seeds (**b**) and disomic substitution had seeds with *dark blue coloring*. *Bar* 10 μm

No detectable *Thinopyrum* chromatin was found in Skorpion, Barevna 9, Barevna 11, Barevna 17, Barevna 23, Barevna 25, Tschermaks Blaukörniger Sommerweizen and H83-952-1 and we can only speculate that the blue aleurone pigmentation is a consequence of a *T. monococcum* introgression and, therefore, represents *Ba2* locus. At least the first seven genotypes as listed have a common ancestor and belong to the legacy of Erich von Tschermak. It is in agreement with results of Zeller et al. (1991) who concluded, based on C-banding patterns and meiotic chromosome pairing in crosses of several European blue-grained wheat strains with double ditelosomic lines and other aneuploid lines of Chinese Spring that the *T. aestivum* Blaukorn strains "Berlin", "Probstdorf", "Tschermak", and "Weihenstephan" were chromosome substitutions of chromosome 4A from diploid *T. monococcum* or *T. boeoticum* for 4A of *T. aestivum*.

Dosage effects

In two genotypes, M90-99-2 and Blue Norco, we observed significant differences in color intensities within each sample (Fig. 4a–c). Thus, we selected 10 seeds each from the dark blue color, light blue color and the standard red color of grain. For Blue Norco, the intensity of the blue color correlated with the dosage of a telocentric *Th. ponticum* chromosome. All plants originating from dark blue kernels had a pair of telocentric chromosomes. In the light blue kernels, only one telocentric *Th. ponticum* was present,

while red seeds had a standard wheat karyotype with 42 chromosomes and no detectable *Thinopyrum* chromatin. In M90-99-2, the situation was more complicated. Among 10 dark seeds, five were homozygous for the 4BS.4BL.4AgL translocation, while the other half were heterozygous. Among 10 light blue kernels, one was homozygous for the translocation and nine were heterozygotes. All 10 red seeds had no *Thinopyrum* chromatin (Fig. 2). Thus, in this genotype 20 % (6/30) seeds were misclassified based on the aleurone color.

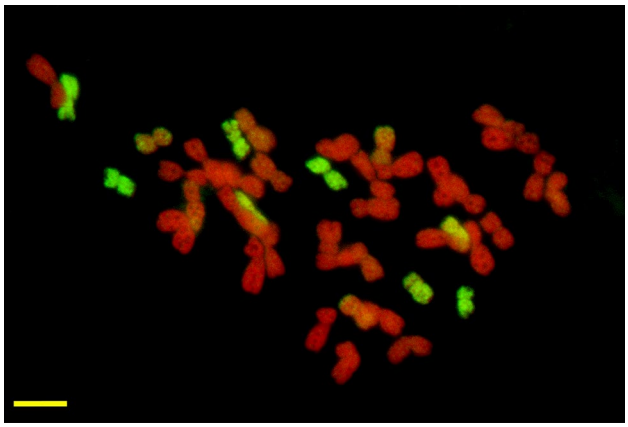


Fig. 3 Cytological analysis of 1066/91 amphiploid blue-aleurone wheat. Genomic in situ hybridization was performed using total genomic DNA of *Th. ponticum* (*green* color) as probe and blocking DNA of *T. aestivum*. Chromosomes were counterstained by DAPI (*red pseudocolor*). *Bar* 10 µm

Discussion

Blue-aleurone wheat is being heralded as a source of functional food due to high anthocyanin content. However, little is known on the variability of genomic constitution among and within various genotypes. Our results indicate that there are at least six different types of introgressions from *Th. ponticum* to bread wheat producing blue color of the aleurone layer. All of them appear to involve wheat chromosomes from the homoeologous group 4 (chromosomes 4B and 4D), or are disomic additions. Based on fertility of the analyzed lines, including tetraploid wheats, it can be assumed with some confidence that in all cases (translocations and additions), the *Th. ponticum* chromosome involved is its group-4 homoeologue. This is in agreement with previous reports. Jan et al. (1981) described UC66049 as 4BS.4AgL translocation, which was confirmed in our study. Similarly, complicated karyotypes were reported in the Sebesta Blue material by Morrison et al. (2004). Whelan (1989) described the karyotype of Blue Norco as disomic addition of a telocentric *Thinopyrum* chromosome. Here, we found variation in the blue color intensity of this genotype, which correlated well with the dosage of *Th. ponticum* chromosome (Fig. 4a–c). Dark blue kernels carry disomic addition of *Th. ponticum,* a monosomic addition generates light blue kernels, and the standard red kernels indicate the absence of any *Th. ponticum* introgression. A similar correlation has also been observed in M90-99-2 which exhibited variation in kernel coloration (data not shown).

Fig. 4 Seed samples of various blue-aleurone wheat genotypes. Note large variation in *blue* color intensity within cv. Blue Norco. Seeds with *dark blue* color were lately identified as disomic chromosome addition of *Th. ponticum* (**a**), *light blue colorizing* were in seeds with one chromosome of *Th. ponticum* (**b**), and in seeds with *red* color, we were unable to detect *Thinopyrum* chromatin. Similar variation was also found among genotypes with the same genomic constitution: **d** EF02-54-9 (Sebesta Blue 3), **e** M90-41 and **f** H90-35-1

Apart from the effect of chromosome instability, much variation in the blue color intensity was observed among the genotypes used in this study, ranging from dark blue seeds of Sebesta Blue genotypes to only slightly bluish kernels of Skorpion, Tschermaks Blaukörniger Sommerweizen and Barevna (data not shown). Generally, a lighter blue color was found among genotypes where no *Th. ponticum* chromatin could be detected and thus, probably carrying *Ba2* gene from *T. monococcum* or *T. boeoticum*. The exception was H83-952-1, which produces dark blue color kernels and has no detectable *Th. ponticum* chromatin. However, we cannot exclude a possibility that the introgression was too small to be detected by GISH. Some variation for color intensity was observed even among genotypes with the same confirmed chromosome constitution. The lines homozygous for the 4BS.4BL.4AgL translocation significantly vary in color from dark blue of Sebesta Blue 3 to light blue of H90-35-1 (Fig. 4d–f). This variation could be also due to different aging of seed samples donated for this study. Clear differences in color intensity between monosomic and disomic introgressions indicate a clear dosage effect. In this sense, color intensity (the amount of anthocyanin) could perhaps be further increased by a combination of two *Th. ponticum* introgressions on different wheat chromosomes.

This study reveals the locations of *Th. ponticum* introgressions and confirms earlier reports. What remains unclear is the sources of introgressions. We still have no information as to which chromosome from decaploid *Th. ponticum* is involved and even if it is the same chromosome in all cases. As mentioned above, all indications point to a group-4 homoeologue(s). However, because we do not have a detailed karyotype of *Th. ponticum*, the identity of the chromosome or chromosomes involved remains unanswered. A possible solution to identify and compare introgressions from different accessions of the blue-aleurone wheat would be to flow-sort chromosomes with *Th. ponticum* substitutions or additions forms (such as Blue Norco and Blue Baart), sequence them and analyze their gene content. This approach has been successfully applied to characterize *T. militinae* introgression in bread wheat (Abrouk et al. 2014). Our pilot experiments indicate that using FISHIS (Giorgi et al. 2013), it should be possible to flow-sort individual chromosomes with introgression from at least four distinct genotypes (data not shown). This should help to uncover the origin of the *Th. ponticum* introgressions in wheat. Based on the cytogenetic analysis, we suspect that there were various sources of *Th. ponticum* introgressions. If this hypothesis is confirmed, the introgressions probably serve as an activator of anthocyanin biosynthetic pathway in aleurone layer.

Most of the existing blue-aleurone wheats carry large blocks of *Th. ponticum* chromatin. This may preclude the use of blue-grained wheat in agriculture as negative effects on grain yield and nutritive characteristics are to be expected. Smaller segment translocations carrying the *Ba* gene must be developed as they ought to minimize the linkage drag. This cannot be accomplished by simple backcrossing, because *Th. ponticum* and *T. aestivum* homoeologues are not expected to pair. Compensating translocation chromosomes with small *Th. ponticum* introgressions can be produced in the presence of the *ph1b* mutation of Sears (1981). Alternatively, very small non-compensating introgressions with minimal adverse effect on grain yield can be produced by irradiation (Zheng et al. 2006b) or hybridization with a gametocidal line (Shi and Endo 1999). The development of lines with the segment of minimal length and still carrying *Ba* gene will increase the potential of blue-aleurone wheats in breeding and agricultural use.

Author contributions DK, PM, TV, JB and JD designed the research; PM and NW provided seed material; VB and DK performed cytogenetic analyses; VB and DK drafted the manuscript. All authors read and approved the final version of the manuscript.

Acknowledgments We are grateful to Prof. Adam J. Lukaszewski for critical reading and valuable comments on the manuscript. We thank Prof. Adam J. Lukaszewski, Prof. C.O. Qualset, Prof. F.J. Zeller, Dr. Robert Metzger, Prof. A. Börner and Genebank of the Crop Research Institute, Prague-Ruzyně, Czech Republic for providing seed samples. This work has been partially supported by the Ministry of Agriculture of the Czech Republic (Project No. QJ1210257) and by the grant LO1204 from the National Program of Sustainability I.

References

Abdel-Aal ESM, Young JC, Rabalski I (2006) Anthocyanin composition in black, blue, pink, purple, and red cereal grains. J Agric Food Chem 54:4696–4704

Abrouk M, Šimková H, Pingault L, Martis M, Jarve K, Paux E, International Wheat Genome Sequencing Consortium, Dolezel J, Valarik M (2014) In silico identification and characterization of wheat 4AL—*T. militinae* introgression. Plant and Animal Genome XXII Web. https://www.pag.confex.com/pag/xxii/webprogram/Paper10789.html. Accessed 11 August 2014

Afaq F, Syed DN, Malik A, Hadi N, Sarfaraz S, Kweon MH, Khan N, Zaid MA, Mukhtar H (2007) Delphinidin, an anthocyaindin in pigmented fruits and vegetables, protects human HaCaT keratinocytes and mouse skin against UVB-mediated oxidative stress and apoptosis. J Invest Dermatol 127:222–232

Anderson JW, Hanna TJ, Peng X, Kryscio R (2000) Whole grain foods and heart disease risk. J Am Coll Nutr 19:291–299

Arbuzova VS, Badaeva ED, Efremova TT, Osadchaya TS, Trubacheeva NV, Dobrovolskaya OB (2012) Cytogenetic study of the blue grain line of the common wheat cultivar Saratovskaya 29. Russ J Genet 48:785–791

Bowen-Forbes CS, Zhang Y, Nair MG (2010) Anthocyanin content, antioxidant, anti-inflammatory and anticancer properties of blackberry and raspberry fruits. J Food Compos Anal 23:554–560

Cermeno MC, Zeller FJ (1986) Identity and cytological characterisation of alien chromosomes conferring blue aleurone colour in common wheat. 1st International. symp chromosome engineering in plants. Xian, China, pp 123–125

Clifford MN (2000) Anthocyanins—nature, occurrence and dietary burden. J Sci Food Agric 80:1063–1072

Dubcovsky J, Luo MC, Zhong GY, Bransteitter R, Desai A, Kilian A, Kleinhofs A, Dvorak J (1996) Genetic map of diploid wheat, *Triticum monococcum* L, and its comparison with maps of *Hordeum vulgare* L. Genetics 143:983–999

Escribano-Bailon MT, Santos-Buelga C, Rivas-Gonzalo JC (2004) Anthocyanins in cereals. J Chromatogr A 1054:129–141

Ficco DBM, Mastrangelo AM, Trono D, Borrelli GM, De Vita P, Fares C, Beleggia R, Platani C, Papa R (2014) The colours of durum wheat: a review. Crop Pasture Sci 65:1–15

Gao L, Mazza G (1994) Quantitation and distribution of simple and acylated anthocyanins and other phenolics in blueberries. J Food Sci 59:1057–1059

Gao JW, Liu JZ, Li B, Feng BS, Yu GQ, Li ZS (2000) Preliminary study on pigments in seed aleurone layer of blue-grained wheat. Acta Botanica Boreali-Occidentalia Sinica 20:936–941

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

Harborne JB, Williams CA (2001) Anthocyanins and other flavonoids. Nat Prod Rep 18:310–333

Himi E, Noda K (2004) Isolation and location of three homoeologous dihydroflavonol-4-reductase (DFR) genes of wheat and their tissue-dependent expression. J Exp Bot 55:365–375

Horbowicz M, Kosson R, Grzesiuk A, Debski H (2008) Anthocyanins of fruits and vegetables—their occurrence, analysis and role in human nutrition. Veg Crop Res Bull 68:5–22

Hou D-X, Fujii M, Terahara N, Yoshimoto M (2004) Molecular mechanisms behind the chemopreventive effects of anthocyanidins. J Biomed Biotechnol 2004:321–325

Jan CC, Dvorak J, Qualset CO, Soliman KM (1981) Selection and identification of a spontaneous alien chromosome translocation in wheat. Genetics 98:389–398

Katsube N, Iwashita K, Tsushida T, Yamaki K, Kobori M (2003) Induction of apoptosis in cancer cells by bilberry (*Vaccinium myrtillus*) and the anthocyanins. J Agric Food Chem 51:68–75

Keppenne VD, Baenziger PS (1990) Inheritance of the blue aleurone trait in diverse wheat crosses. Genome 33:525–529

Knievel DC, Abdel-Aal E-SM, Rabalski I, Nakamura T, Hucl P (2009) Grain color development and the inheritance of high anthocyanin blue aleurone and purple pericarp in spring wheat (*Triticum aestivum* L.). J Cereal Sci 50:113–120

Knott DR (1958) The inheritance in wheat of a blue endosperm color derived from *Agropyron elongatum*. Can J Bot 36:571–574

Kubaláková M, Kovářová P, Suchánková P, Číhalíková J, Bartoš J, Lucretti S, Watanabe N, Kianian SF, Doležel J (2005) Chromosome sorting in tetraploid wheat and its potential for genome analysis. Genetics 170:823–829

Lamy S, Blanchette M, Michaud-Levesque J, Lafleur R, Durocher Y, Moghrabi A, Barrette S, Gingras D, Beliveau R (2006) Delphidin, a dietary anthocyanidin, inhibits vascular endothelial growth factor receptor-2 phosphorylation. Cancerogenesis 27:989–996

Lila MA (2004) Anthocyanins and human health: an in vitro investigative approach. J Biomed Biotechnol 5:306–313

Masoudi-Nejad A, Nasuda S, McIntosh RA, Endo TR (2002) Transfer of rye chromosome segments to wheat by a gametocidal system. Chromosome Res 10:349–357

Matsumoto H, Nakamura Y, Tachibanaki S, Kawamura S, Hirayama M (2003) Stimulatory effect of cyanidin 3-glycosides on the regeneration of rhodopsin. J Agric Food Chem 51:3560–3563

Mazza G (2007) Anthocyanins and heart health. Annali dell'Istituto Superiore di Sanita 43:369–374

Mazza G, Miniati E (1993) Anthocyanins in fruits, vegetables and grains. CRC Press Incorporation, Boca Raton

Morrison LA, Metzger RJ, Lukaszewski AJ (2004) Origin of the Blue-Aleurone Gene in Sebesta blue wheat genetic stocks and a protocol for its use in apomixes screening. Crop Sci 44:2063–2067

Pisha E, Pezzuto JM (1994) Fruits and vegetables containing compounds that demonstrate pharmacological activity in humans. Econ Med Plant Res 6:189–233

Qualset CO, Sollman KM, Jan C-C, Dvorak J, McGuire PE, Vogt HE (2005) Registration of UC66049 *Triticum aestivum* blue aleurone genetic stock. Crop Sci 45:432

Rechner AR, Kroner C (2005) Anthocyanins and colonic metabolites of dietary polyphenols inhibit platelet function. Thromb Res 116:327–334

Sears ER (1981) Transfer of alien genetic material to wheat. In: Evans LT, Peacock WJ (eds) Wheat science—today and tomorrow. Cambridge University Press, New York, pp 75–89

Shi F, Endo TR (1999) Genetic induction of structural changes in barley chromosomes added to common wheat by a gametocidal chromosome derived from *Aegilops cylindrica*. Genes Genet Syst 74:49–54

Singh K, Ghai M, Garg M, Chhuneja P, Kaur P, Schnurbusch T, Keller B, Dhaliwal HS (2007) An integrated molecular linkage map of diploid wheat based on a *Triticum boeoticum* × *T. monococcum* RIL population. Theor Appl Genet 115:301–312

Trojan V, Musilová M, Vyhnánek T, Klejdus B, Hanáček P, Havel L (2014) Chalcone synthase expression and pigment deposition in wheat with purple and blue colored caryopsis. J Cereal Sci 59:48–55

Wang J, Mazza G (2002) Effects of anthocyanins and other phenolic compounds on the production of tumor necrosis factor alpha in LPS/IFN-gamma-activated RAW 264.7 macrophages. J Agric Food Chem 50:4183–4189

Wang L-S, Stoner GD (2008) Anthocyanins and their role in cancer prevention. Cancer Lett 269:281–290

Wang H, Cao GH, Prior RL (1997) Oxygen radical absorbing capacity of anthocyanins. J Agric Food Chem 45:304–309

Whelan EDP (1989) Transmission of an alien telocentric addition chromosome in common wheat that confers blue seed colour. Genome 32:30–34

William MDHM, Mujeeb-Kazi A (1993) *Thinopyrum bessarabicum*: biochemical and cytological markers for detection of genetic introgression in its hybrid derivatives with *Triticum aestivum* L. Theor Appl Genet 86:365–370

Yang GH, Li B, Gao JW, Liu JZ, Zhao XQ, Zheng Q, Tong YP, Li ZS (2004) Cloning and expression of two chalcone synthase and a flavonoid 3′5′-hydroxylase 3′end cDNAs from developing seeds of blue-grained wheat involved in anthocyanin biosynthetic pathway. Acta Bot Sin 46:588–594

Yoshinaga K (1986) Liquor with pigments of red rice. J Brew Soc Jpn 81:337–342

Youdim KA, Martin A, Joseph JA (2000) Incorporation of the elderberry anthocyanins by endothelial cells increases protection against oxidative stress. Free Rad Biol Med 29:51–60

Zeller FJ, Cermeno MC, Miller TE (1991) Cytological analysis on the distribution and origin of the alien chromosome pair conferring blue aleurone color in several European common wheat (*Triticum aestivum* L.) strains. Theor Appl Genet 81:551–558

Zeven AC (1991) Wheats with purple and blue grains: a review. Euphytica 56:243–258

Zhang XY, Dong YS, Wang RRC (1996) Characterization of genomes and chromosomes in partial amphiploids of the hybrid *Triticum aestivum* × *Thinopyrum ponticum* by in situ hybridization, isozyme analysis and RAPD. Genome 39:1062–1071

Zheng Q, Li B, Zhang XY, Mu SM, Zhou HP, Li ZS (2006a) Physical mapping of the blue-grained gene from *Thinopyrum ponticum* by GISH and FISH in a set of translocation lines with different seed colors in wheat. Genome 49:1109–1114

Zheng Q, Li B, Zhang XY, Mu SM, Zhou HP, Li ZS (2006b) Molecular cytogenetic characterization of wheat-*Thinopyrum ponticum* translocations bearing blue-grained gene(s) induced by r-ray irradiation. Euphytica 152:51–60

APPENDIX II

Advances in plant chromosome genomics

Doležel J, Vrána J, Cápal P, Kubaláková M, Burešová V, Šimková H

Biotechnology advances, 32(1):122-136, 2014

IF: 9.848

Biotechnology Advances 32 (2014) 122–136

Contents lists available at ScienceDirect

Biotechnology Advances

journal homepage: www.elsevier.com/locate/biotechadv

Research review paper

Advances in plant chromosome genomics ☆

Jaroslav Doležel *, Jan Vrána, Petr Cápal, Marie Kubaláková, Veronika Burešová, Hana Šimková

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

ARTICLE INFO

Available online 7 January 2014

Keywords:
BAC library
Chromosome sorting
Cytogenetics
DNA arrays
DNA marker development
Flow cytometry
Genome assembly
Mitotic metaphase chromosomes
Next generation sequencing
Physical mapping

ABSTRACT

Next generation sequencing (NGS) is revolutionizing genomics and is providing novel insights into genome organization, evolution and function. The number of plant genomes targeted for sequencing is rising. For the moment, however, the acquisition of full genome sequences in large genome species remains difficult, largely because the short reads produced by NGS platforms are inadequate to cope with repeat-rich DNA, which forms a large part of these genomes. The problem of sequence redundancy is compounded in polyploids, which dominate the plant kingdom. An approach to overcoming some of these difficulties is to reduce the full nuclear genome to its individual chromosomes using flow-sorting. The DNA acquired in this way has proven to be suitable for many applications, including PCR-based physical mapping, *in situ* hybridization, forming DNA arrays, the development of DNA markers, the construction of BAC libraries and positional cloning. Coupling chromosome sorting with NGS offers opportunities for the study of genome organization at the single chromosomal level, for comparative analyses between related species and for the validation of whole genome assemblies. Apart from the primary aim of reducing the complexity of the template, taking a chromosome-based approach enables independent teams to work in parallel, each tasked with the analysis of a different chromosome(s). Given that the number of plant species tractable for chromosome sorting is increasing, the likelihood is that chromosome genomics – the marriage of cytology and genomics – will make a significant contribution to the field of plant genetics.

Contents

1. Sequencing of plant genomes . . . 123
2. Reducing the complexity of the sequencing template . . . 123
3. Flow cytometry . . . 124
   3.1. Sample preparation . . . 125
   3.2. Analysis and sorting . . . 126
4. Uses of flow-sorted chromosomes . . . 128
   4.1. Physical mapping . . . 128
      4.1.1. Mapping by PCR . . . 128
      4.1.2. Construction of clone-based physical maps . . . 129
      4.1.3. Cytogenetic mapping . . . 130
   4.2. Genetic marker development . . . 130
      4.2.1. SSRs and ISBPs . . . 130
      4.2.2. DArT markers . . . 130
      4.2.3. Marker development from chromosome-specific shotgun sequences . . . 130
      4.2.4. Marker specificity . . . 131
   4.3. Sequencing . . . 131
      4.3.1. BAC clones . . . 131
      4.3.2. Whole chromosome sequencing using 454 technology . . . 131
      4.3.3. Whole chromosome sequencing using Illumina technology . . . 132
      4.3.4. Validation of whole genome assemblies . . . 133

* Corresponding author. Tel.: +420 585 238 703.
E-mail addresses: dolezel@ueb.cas.cz (J. Doležel), vrana@ueb.cas.cz (J. Vrána), capal@ueb.cas.cz (P. Cápal), kubalakova@ueb.cas.cz (M. Kubaláková), buresova@ueb.cas.cz (V. Burešová), simkovah@ueb.cas.cz (H. Šimková).

0734-9750/$ – see front matter © 2013 The Authors. Published by Elsevier Inc. All rights reserved.
http://dx.doi.org/10.1016/j.biotechadv.2013.12.011

5. Conclusions . . . 133
Acknowledgements . . . 133
References . . . 133

1. Sequencing of plant genomes

The last decade has seen a major leap in our understanding of plant genome structure, function and evolutionary dynamics. The main driver of this advance has been the elaboration of next generation sequencing (NGS) platforms, which allow for the parallel acquisition of huge numbers of reads, representing hundreds of billions of nucleotides; in concert, advances in bioinformatics have been necessary to enable this ocean of DNA sequence to be analyzed. The first plant genome to be fully sequenced was that of *Arabidopsis thaliana*, chosen for its small genome of ~150 Mb; although this represented a logistical challenge in the context of 1990s sequencing technology, it would no longer do so, given the capacity of modern instruments, which can generate up to 60 Gb of sequence per run. The *A. thaliana* genome was acquired using a clone-by-clone (CBC) strategy (The Arabidopsis Genome Initiative, 2000). The minimum set of clones to be sequenced, termed the "minimum tiling path" (MTP), is elaborated from the physical map, which is constructed on the basis of overlapping large-insert DNA clones. The second plant species to be sequenced was rice, using a similar strategy (Matsumoto et al., 2005). Apart from its importance as a crop species, rice was selected also because of its relatively small genome size (~400 Mb). The acquisition of these two whole genome sequences marked a new departure for plant genetics, allowing, for the first time, a holistic view of the entire genome. Since the beginning of the present century, the pace of sequencing has accelerated, so that by 2010, a number of important plant species had been sequenced.

A gradual shift in sequencing strategy, moving away from the CBC approach to a whole genome shotgun (WGS) one was already underway during the first phase of plant genome sequencing. The shotgun method was used for acquiring the genome sequences of poplar (Tuskan et al., 2006), grapevine (Jaillon et al., 2007) and sorghum (Paterson et al., 2009). The 2.5 Gb maize genome was published in 2009, but exceptionally relied on the CBC approach (Schnable et al., 2009). Since 2010, NGS technologies have become routine, and have greatly driven down both the price and effort required of genome sequencing. In this second phase of plant genome sequencing, already some 40 plant species have been sequenced, and the expectation is that not only reference genome sequences will be acquired for most of the economically and scientifically important plant species, but that the scale of re-sequencing will grow by orders of magnitude (The million plant and animal genomes project, 2013). Unlike *de novo* sequencing, which requires the assembly of the genome from short reads, re-sequencing is technically simpler, as the reads can be referenced to an available complete genome sequence. The quality of re-sequenced genomes is therefore determined by the quality of the reference genome sequence; the fuller the coverage of the reference sequence, the more correctly the re-sequenced contigs will be ordered. The feasibility of sequencing many individuals from the same species offers opportunities for population genetics analysis and genotype-based breeding (Long et al., 2013).

High quality reference genome sequences are particularly important for the analysis of the functional organization of DNA. The function of the nuclear genome cannot be understood without an understanding of its various components, as exemplified by the human genome ENCODE project (Gerstein et al., 2012). An unfortunate consequence of the widespread use of NGS shotgun sequencing is a drop in assembly quality, so that the highest quality genome sequences remain those of *A. thaliana*, rice and maize, which were acquired by the CBC method (Feuillet et al., 2011; Shangguan et al., 2013). Assembly is particularly problematical for large genome species such as Norway spruce (1C: ~20 Gb), where only some 25% of the genome was assemblable into scaffolds longer than 10 Kb (Nystedt et al., 2013); such issues can arise in smaller genomes too, for example in chickpea (1C: ~0.9 Gb), where the genome sequence presently comprises over 180,000 scaffolds (Jain et al., 2013). Of course, it is not always necessary to generate a gold standard sequence, since for some applications a rough genome draft is sufficient for the purpose. The difficulty arises when such draft genome assemblies are presented as reference sequences (Sierro et al., 2013). In some cases, projects relying on incomplete genome sequences may fail, and there are examples where funding proposals aimed at the acquisition of a high quality reference sequence have been declined as the donors believed that the work had already been done.

The power of NGS lies in its capacity to generate a huge volume of reads, but its weakness is that these reads are rather short. Plant genomes are populated by many families of repetitive DNA elements (Schmidt and Heslop-Harrison, 1998), and these can be impossible to resolve when only short reads are available. The problem of sequence redundancy is compounded in polyploids, which dominate the plant kingdom. Genome assembly from shotgun reads may not be straightforward even in compact genomes having a small content of repetitive DNA. A good example is the bladderwort *Utricularia gibba*, with a genome size of just 77 Mb, of which only 3% is repetitive; nevertheless an attempt at shotgun sequencing resulted in a set of >3800 sequence contigs arranged in over 1200 scaffolds (Ibarra-Laclette et al., 2013). Technical improvements in read length and/or the algorithms used for sequence assembly should in time, however, enable reference genome sequences to be produced by NGS shotgun methods (Roberts et al., 2013). NGS shotgun sequencing may be at present be of limited utility in acquiring gold standard reference sequences (Marx, 2013), but the technology is very powerful for simpler templates such as bacterial artificial chromosomes (BACs), which form the backbone of many physical maps (Feuillet et al., 2011). Incomplete sequence assembly is then limited to at most 100 Kb, the genomic location of which is known. BAC clones are commonly sequenced in pools to reduce cost (Sato et al., 2011; Steuernagel et al., 2009), and this requires a bar-coding strategy to attribute the resulting contigs to their specific BAC. The sequence redundancy typical of large and particularly of polyploid genomes, makes the construction of a physical map based on BAC clones difficult (Meyers et al., 2004; Paux et al., 2008); it is a task which would be greatly simplified if the template complexity could be reduced.

2. Reducing the complexity of the sequencing template

As both the CBC and the NGS shotgun sequencing strategies are compromised by sequence redundancy, any reduction in template complexity would be helpful. Breaking down the genome into its individual chromosomes represents an attractive option, especially for polyploid genomes, as this would abolish the problem of redundancy due to the presence of homoeologs (Fig. 1). Flow-sorting has been developed to achieve exactly this result, and this review outlines its potential for plant genome analysis and sequencing. Methods designed to simplify the assembly of shotgun sequence reads and to construct ready-to-sequence clone-based physical maps are described. Chromosome sorting is not, of course, the sole option available for reducing template complexity prior to DNA sequencing. The selection of DNA based on either its renaturation kinetics ("Cot filtration") (Peterson et al., 2002)

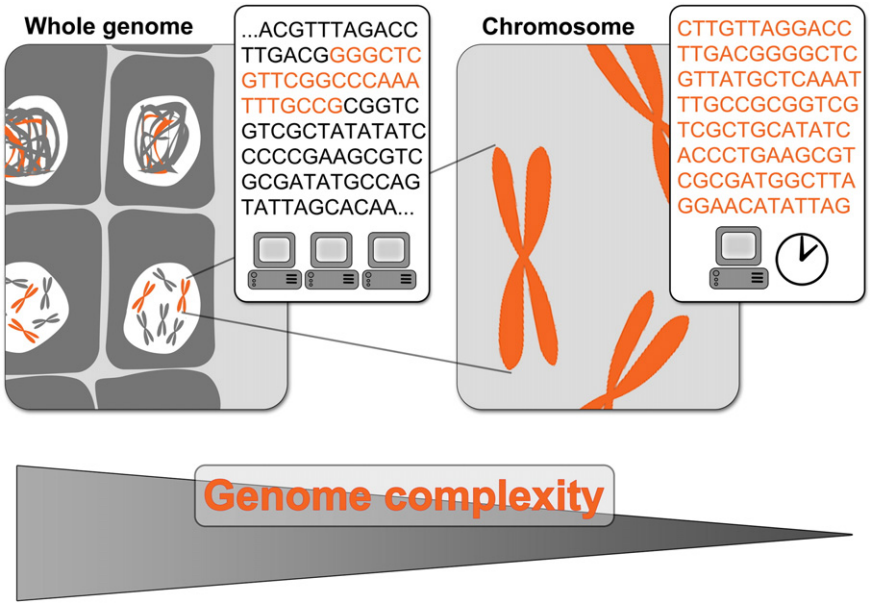


Fig. 1. Chromosome genomics: instead of treating the whole nuclear genome as a unit, single chromosomes are isolated and their DNA used as the template for genomic analyses. The reduction in template complexity achieved speeds mapping, sequencing and sequence analysis, and simplifies the necessary bioinformatics. In polyploids, interference from homoeologs is minimized.

or its methylation status (Rabinowicz et al., 2003) both were designed to eliminate much of the repetitive DNA component, leaving mainly low copy sequences. A complexity reduction step has also been incorporated into genotyping-by-sequencing, based on the use of methylation-sensitive restriction enzymes to eliminate the highly methylated repetitive component prior to sequencing (Elshire et al., 2011), and several other target-enrichment strategies have been developed (Mamanova et al., 2010). Inevitably, this sort of strategy, unlike one based on individual chromosomes, cannot deliver a complete genome sequence. Chromosome number is variable from species to species, but is typically in the range 5–20. Thus, complexity can in principle, be reduced by around an order of magnitude. For example, each barley or bread wheat chromosome harbors, on average, respectively about 14% and 5% of the full genome complement.

During most of an organism's life cycle, its chromosomes are extended and intimately intertwined with one another in interphase nuclei. The exceptions are during cell division, when the chromosomes become very much shortened and are physically separated from one another. Attempts have been made to isolate mitotic chromosomes using microdissection (Matsunaga et al., 1999; Stein et al., 1998). A clear advantage of this approach is that the chromosomes have already been attached to a fixed surface, where they can be optically identified prior to their mechanical isolation. However the process is highly labor-intensive, so only small populations of individual chromosomes can be isolated; while the resulting DNA can be amplified to provide template sufficient for sequencing, the required amplification imposes such a restriction on the length of the DNA recovered (Schondelmaier et al., 1993; Stein et al., 1998) that it become unsuitable for constructing the large insert libraries required to assemble a physical map. Moreover, extensive amplification inevitably introduces a bias. The alternative to micro-dissection is to isolate large populations of intact mitotic metaphase chromosomes in suspension. The methods required to achieve this necessitate not just the ability to prepare such suspensions, but also the means to physically separate a specific chromosome from the mass of non-homologs present. Attempts have been made to achieve this separation using gradient centrifugation (Stubblefield and Oro, 1982) or by capture on magnetic beads following hybridization with a labeled chromosome-specific probe (Dudin et al., 1988; Vitharana and Wilson, 2006); however, to date, the most successful method is flow-sorting (Doležel et al., 1994, 2007a, 2011). In what follows, we first explain the methodology involved in flow cytometric chromosome analysis and sorting (termed "flow cytogenetics") and then discuss current and potential applications of flow-sorted chromosomes in plant genomics ("chromosome genomics").

3. Flow cytometry

Flow cytometry was initially developed as an alternative to microscopy for counting blood cells; its advantage is its high throughput and potential for automation. The capacity to handle large numbers of individual cells enables the detection of rare mutants, and can deliver meaningful statistical data with respect to frequency. A typical flow cytometer does not capture images of the cells; rather the aim is to analyze light scatter and fluorescence. Flow cytometers need to be capable of measuring these properties simultaneously in real time, as they combine to provide a wealth of information (Rieseberg et al., 2001), specifically regarding cell viability, physiological status, apoptosis, ploidy and cell cycle status. Supported by a variety of fluorescent probes and antibodies, flow cytometry has developed into a ubiquitous tool in immunology, pathology, oncology and other areas of biomedical research (Shapiro, 2003). Although less commonly exploited in plant biology, these devices have found a number of fundamental research and industrial uses, the main ones being the estimation of genome size and ploidy level (Doležel et al., 2007b). The salient feature of flow cytometry is that the target particles are suspended in a narrow stream of liquid (typically saline); they are forced to move in a single file, where they can be made to interact one-by-one with an orthogonally oriented light beam (Fig. 2). Solid state lasers provide the most commonly used light source, and it is not unusual to install more than one laser, with each set up to excite a different fluorochrome incorporated into the particles. The flow rate is typically several thousand per second. To sort the particles into discrete sub-populations, the stream is broken into ~1 nL droplets. Those carrying a target particle are electrically charged and then deflected from the main stream of non-target particles by passage through an electrical field. Because the rate of droplet generation exceeds the particle flow rate, the majority of droplets are empty and very few droplets contain more than one particle. Clumps of particles tend to block the narrow orifice (typically < 100 μm) of the flow chamber, thereby disrupting laminar flow and compromising the analysis (Shapiro, 2003). Poor results are also obtained if the particles are

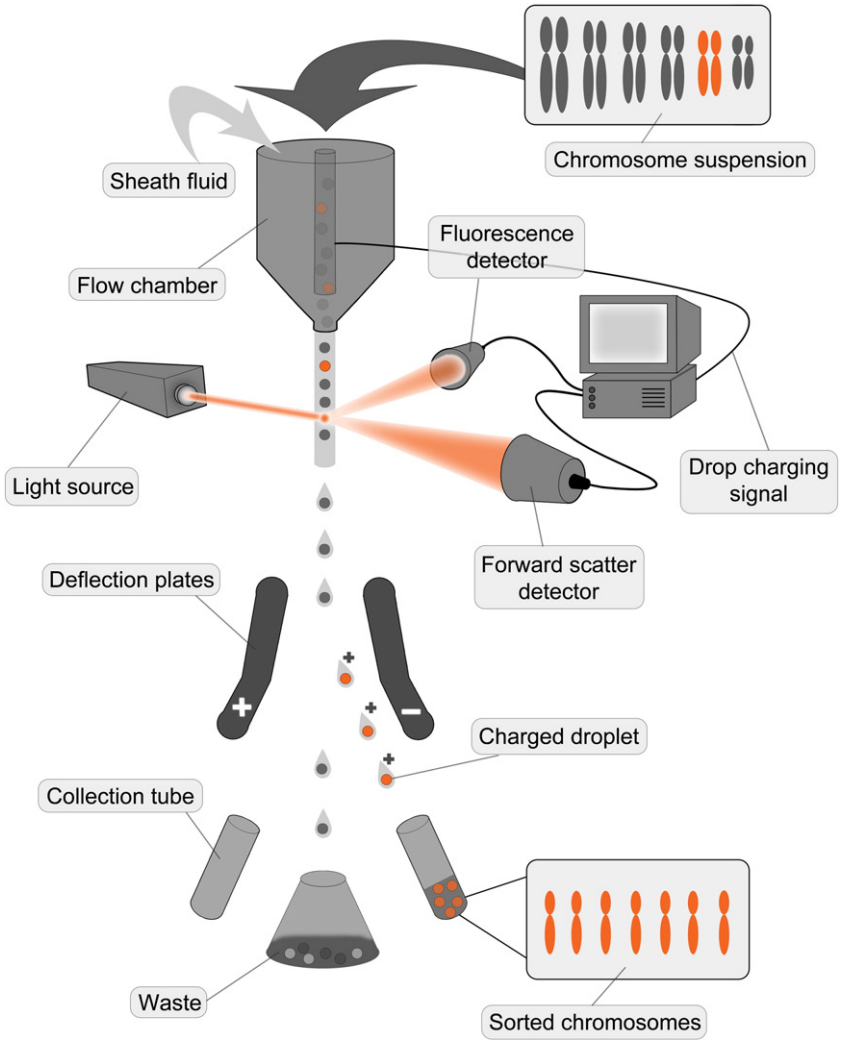


Fig. 2. The mechanics of flow-sorting. Chromosomes held in liquid suspension are stained by a fluorochrome and passed into a flow chamber containing sheath fluid. The geometry of the chamber forces the chromosome suspension into a narrow stream in which the chromosomes become aligned in a single file, and so are able to interact individually with an orthogonally directed laser beam(s). Pulses of scattered light and emitted fluorescence are detected and converted to electric pulses. If the chromosome of interest differs in fluorescence intensity from other chromosomes, it is identified and sorted. The sorting is achieved by breaking the stream into droplets and by electrically charging droplets carrying chromosomes of interest. The droplets are deflected during passage through electrostatic field between defection plates and collected in suitable containers.

mechanically damaged. Thus sample quality is of prime importance, and this is especially the case for chromosome analysis and sorting, where any reduction in resolution will produce unwanted contamination of a sorted chromosome by other chromosomes, chromosome fragments or aggregates. The elaboration of a robust method for preparing sufficiently high quality chromosome suspensions has been the most serious barrier to the development of flow cytogenetics in humans and animals, but particularly in plants.

3.1. Sample preparation

Since in general somatic tissue is easier to obtain than reproductive tissue, flow cytogenetics has largely concentrated on isolating mitotic metaphase chromosomes. At any given time, the majority of plant and animal cells in non-reproductive tissue are in interphase, so accessing mitotic metaphase chromosomes requires a pre-treatment to first encourage cell division, and then to arrest cells at mitotic metaphase. (Note that targeting meiotic chromosomes in microspores is in principle highly attractive, since cell division is well synchronized in these cells. However there are practical difficulties associated with the acquisition of a sufficient number of dividing cells.) Some technical issues surround the release of metaphase chromosomes into the liquid medium. Current protocols designed to prepare chromosome suspensions from human or animal cells are based on either synchronized cell lines or stimulated peripheral blood; the chromosomes are released by hypotonic lysis (Chen et al., 2008; Yang et al., 2011). Plant cells are less tractable, mainly because of their rigid cell wall. Synchronizing mitosis is also less straightforward than in animal cells. De Laat and Blaas (1984), who were the first to demonstrate the sorting of plant chromosomes, used hydroxyurea for synchronization and colchicine to arrest cells at metaphase. A similar approach was taken by Arumuganathan et al. (1991) in tomato and by Schwarzacher et al. (1997) in wheat. Although plant cells, like animal ones, can be cultured *in vitro,* such cultures are often karyologically unstable (Leitch et al., 1993; Schwarzacher et al., 1997), and their cell cycle is not well synchronized (typically not exceeding 35%, see Arumuganathan et al. (1994) and de Laat and Blaas (1984)). Following the animal cell protocols, hypotonic lysis was used in early experiments to release plant chromosomes, but this was only feasible if the cell walls were first digested enzymatically. While this step provides a non-disruptive means of releasing the chromosomes, it also introduces a time delay between metaphase arrest and the chromosome release, which lowers the chromosome yield due to the premature separation of sister chromatids and/or chromosome decondensation.

The release of chromosomes from leaf-derived protoplasts was described by Conia et al. (1987). The strategy adopted was to force arrest of the cells in the G1 phase, and then to transfer the cells into a

medium formulated to initiate cell cycling, so that they would enter mitosis in synchrony. Unfortunately the induction was not sufficiently effective, since only 10% of the cells divided. Additionally, chromosome release was hampered by a partial regeneration of the cell wall. A major advantage of sourcing chromosomes from live plant tissue (such as the leaf) as opposed to *in vitro* cultured cells is that their karyotype is normal. The choice of root tip meristems as a source of mitotic chromosomes is based on a naturally high rate of cell division, and (unlike leaf-derived cells), the ease of synchronizing mitosis, with rates above 50% being attainable (Doležel et al., 1992). A productive method of chromosome release from root tips, avoiding the need to digest the cell wall, was elaborated by Doležel et al. (1992). The material was first fixed in formaldehyde to render the chromosomes mechanically stable and resistant to shearing forces, and then homogenized. Apart from karyological stability, the advantage of using root tips is that seedlings can be obtained in a majority of plants and roots can be exposed to various treatments using a hydroponic system. The procedure can be extended to species which produce few (or no) seeds by inducing hairy root cultures (Neumann et al., 1998; Veuskens et al., 1995).

A typical root tip-based protocol (e.g., Vrána et al., 2012) involves seed germination, the exposure of roots of young seedlings to hydoxyurea (a DNA synthesis inhibitor) to arrest the cells at the G1/S interface, followed by recovery to synchronize the cell cycle through the S and G2 phases and into mitosis. Dividing cells are arrested at mitotic metaphase by treating with a mitotic spindle poison such as the herbicides amiprophos-methyl, oryzalin or trifluralin (Doležel et al., 1992; Guo et al., 2006; Vláčilová et al., 2002). In species where these compounds induce chromosome stickiness, a treatment with nitrous oxide (Kato, 1999) has proven to be efficacious (unpublished data). An option is an overnight exposure to ice water prior to fixation, a treatment which can improve the dispersion of chromosomes in the cytoplasm and thereby increase the chromosome yield (Vrána et al., 2000). The treated roots are then fixed in formaldehyde and the chromosomes released into the isolation buffer by chopping using a sharp scalpel or razor blade (Doležel et al., 1992). When working with small root tips, homogenization using a handheld homogenizer is both rapid and convenient (Gualberti et al., 1996). Of especial importance is the composition of the isolation buffer, as this ensures the maintenance of chromosome morphology and DNA integrity, as well as providing a compatible environment for DNA staining.

3.2. Analysis and sorting

To date, flow cytometry has been used to sort chromosomes in 24 plant species, belonging to 18 genera (Table 1). Staining chromosomal DNA with a fluorochrome (commonly either ethidium bromide (Li et al., 2004), Hoechst 333242 (Conia et al., 1987) or DAPI (Kubaláková et al., 2005)) results in a distribution of fluorescence signal intensity (the "flow karyotype"), in which, ideally, each chromosome can be recognized by a different peak. Formaldehyde fixation has been found to interfere with the stoichiometric binding of some fluorochromes to chromosomal DNA, and DAPI has been found to be the least sensitive of the fluorochromes in this respect (Doležel and Lucretti, 1995). The size of the peak is dependent on the DNA content, and it is common to find that the DNA content of two (or more) of the chromosomes is so similar that they are represented in the flow karyotype as a single, broad peak. Thus, for example, in the flow karyotype of chickpea, six of the eight chromosomes can be separated, while the other two form a single peak (Fig. 3A). In contrast, the bread wheat ($n = 21$) flow karyotype comprises only one single chromosome peak (chromosome 3B), with the other 20 chromosomes forming three composite peaks (Fig. 3B). Karyotype variation within wheat has allowed some additional chromosomes to be discriminated (Kubaláková et al., 2002), and the same is the case for chickpea (Vláčilová et al., 2002; Zatloukalová et al., 2011). The pattern of light scatter can be used to differentiate between chromosomes and cell detritus (Conia et al., 1987), while the width of the fluorescence pulse aids in the discrimination of chromosomes doublets (Lucretti et al., 1993).

The inability to discriminate each chromosome in the flow karyotype presents a serious limitation to the utility of flow cytometry, so substantial effort has been devoted to overcoming this problem. An early strategy was to simultaneously stain the material with two fluorochromes differing in their base pair preference (for instance Hoechst 33258 which binds preferentially to AT rich sequence and Chromomycin A3, which targets GC rich sequence). In the human karyotype, this method effectively discriminates almost every chromosome (Ferguson-Smith,

Table 1
List of plant species for which a flow cytometric analysis of mitotic chromosomes has been published.

Genus	Species	Common name	n	References
Aegilops	*biuncialis*	Goatgrasses	14	Molnár et al. (2011)
	comosa		7	Molnár et al. (2011)
	geniculata		14	Molnár et al. (2011)
	umbellulata		7	Molnár et al. (2011)
Avena	*sativa*	Oat	21	Li et al. (2001)
Cicer	*arietinum*	Chickpea	8	Vláčilová et al. (2002), Zatloukalová et al. (2011)
Dasypyrum	*villosum*		7	Grosso et al., 2012; Giorgi et al., 2013
Festuca	*pratensis*	Meadow fescue	7	Kopecký et al., 2013
Haplopappus	*gracilis*		2	de Laat and Blaas (1984), de Laat and Schel (1986)
Hordeum	*vulgare*	Barley	7	Lysák et al. (1999), Lee et al. (2000), Suchánková et al. (2006)
Lycopersicon	*esculentum*	Tomato	12	Arumuganathan et al. (1991)
	pennellii	Tomato	12	Arumuganathan et al. (1991, 1994)
Nicotiana	*plumbaginifolia*	Tobacco	10	Conia et al. (1989)
Oryza	*sativa*	Rice	12	Lee and Arumuganathan (1999)
Petunia	*hybrida*	Petunia	7	Conia et al. (1987)
Picea	*abies*	Norway spruce	12	Überall et al. (2004)
Pisum	*sativum*	Pea	7	Gualberti et al. (1996), Neumann et al. (1998, 2002)
Secale	*cereale*	Rye	7	Kubaláková et al. (2003)
Silene	*latifolia*	White campion	12	Veuskens et al. (1995), Kejnovský et al. (2001)
Triticum	*aestivum*	Bread wheat	21	Wang et al. (1992), Schwarzacher et al. (1997), Lee et al. (1997), Gill et al. (1999), Vrána et al. (2000), Kubaláková et al. (2002), Giorgi et al., 2013
	durum	Durum wheat	14	Kubaláková et al. (2005), Giorgi et al., 2013
Vicia	*faba*	Field bean	6	Lucretti et al. (1993), Doležel and Lucretti (1995), Lucretti and Doležel (1997)
	sativa	Common vetch	6	Kovářová et al. (2007)
Zea	*mays*	Maize	10	Lee et al. (1996, 2002), Li et al. (2001, 2004)

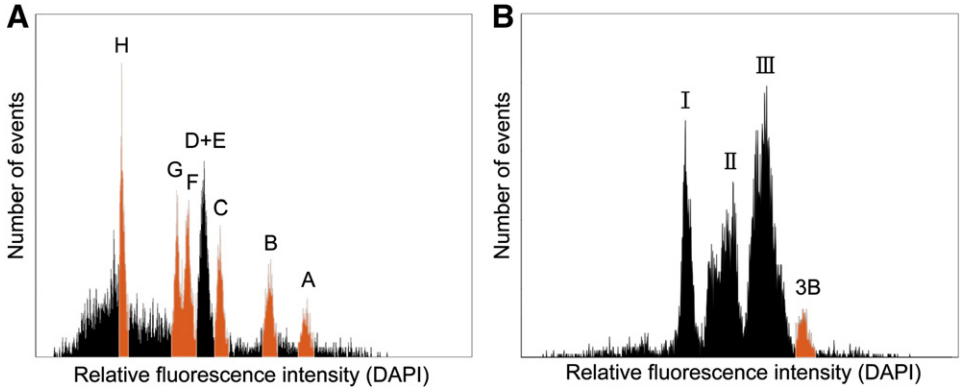


Fig. 3. Flow karyotyping in chickpea and bread wheat. The fluorescence intensity histograms (flow karyotypes) were obtained from DAPI-stained suspensions of mitotic chromosomes. (A) Chickpea cv. Frontier (2n = 2x = 16) forms seven peaks, six of which each represent a single chromosome (A–C and F–H). The seventh peak harbors both chromosomes D and E. (B) In the wheat cv. Chinese Spring (2n = 6x = 42) flow karyotype, only chromosome 3B forms a discrete peak. The remaining 20 chromosomes are dispersed into the composite peaks I–III.

1997; Langlois et al., 1982), but it has not been successful in plant genomes (Lee et al., 1997, 2000; Lucretti and Doležel, 1997; Schwarzacher et al., 1997), presumably because global variation in AT/GC ratio among the chromosomes is masked by the ubiquitous presence of repetitive DNA (Fuchs et al., 1996; Schubert et al., 2001). The approach taken attempted to exploit polymorphism in chromosome length resulting from deletions and translocations. Lucretti et al. (1993) were the first to show that reciprocal translocations in field bean could be used to identify a number of its chromosomes, and a similar success was recorded by Neumann et al. (1998) working with garden pea. In some cases, chromosome sorting has been facilitated by cryptic structural features (Kubaláková et al., 2002; Lee et al., 2002). The tolerance of polyploids to aneuploidy has been used to develop a plethora of true-breeding cytogenetic materials, especially in bread wheat. Of particular interest in the context of flow karyotyping are telocentric chromosomes (telosomes), in which an entire arm has been lost; a collection of these, involving most of the 42 chromosome arms of wheat was generated by Sears (Sears and Sears, 1978). The small size of telosomes means that their peaks become well separated from the rest of the flow karyotype, allowing them to be readily sorted (Gill et al., 1999; Guo et al., 2006; Kubaláková et al., 2002) (Fig. 4A).

The tolerance of polyploids to aneuploidy has also allowed for the production of stable lines in which a single chromosome pair from a related species can be maintained in isolation from the others. If this "alien" chromosome differs in DNA content from those of the host species, its peak should be recognizable, and can therefore be sorted. For example, in cereal rye, the only chromosome which can be successfully sorted from the other six is 1R, but the other six proved to be sortable when represented in a single chromosome addition line (Kubaláková et al., 2003) (Fig. 4B). In the case of barley, the peaks overlap with those of wheat, so the chromosome addition line approach is not fruitful. However, it has proved possible to discriminate and sort barley telosomes present as a single pair in a wheat background (Suchánková et al., 2006). The availability of such addition lines has been a boon for chromosome sorting in the wild relatives of wheat, which otherwise have proven difficult to purify (Grosso et al., 2012; Molnár et al., 2011). Like bread wheat, oat is also a hexaploid able to tolerate the addition of an alien chromosome pair, and this property has been used to sort certain maize chromosomes (Li et al., 2001). Some plant species possess so called B chromosomes, whose evolution, function and molecular organization have long been controversial (Jones, 1995; Jones and Houben, 2003). They are typically much smaller than the standard chromosomes, and therefore are amenable to sorting (Kubaláková et al., 2003; Martis et al., 2012). A further example is represented by the dioecious species white campion, which carries a sex chromosome which differs in size from the rest of the chromosome complement, and can thus be sorted (Kejnovský et al., 2001; Veuskens et al., 1995).

Sorting specific chromosomes using an addition line is a convenient means of isolating a portion of the donor genome. However, the development of these lines is very laborious, so they can only ever be generated from a limited number of donors. A similar consideration relates to translocation and deletion lines. Many applications, however, focus on a

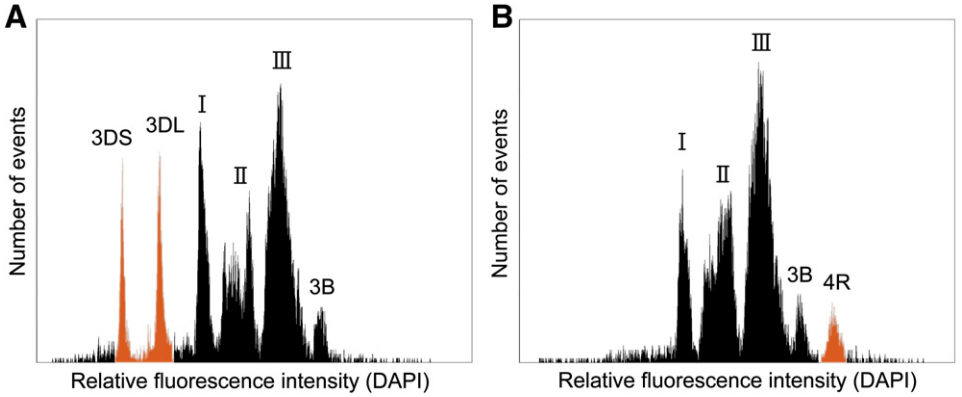


Fig. 4. The use of cytogenetic stocks to isolate particular wheat chromosomes. Flow karyotypes were obtained from DAPI-stained suspensions of mitotic chromosomes. (A) The double ditelosomic line dDt3D (20″ + t″3DS + t″3DL) carries the two arms of chromosome 3D in the form of two distinct telosomes, each of which is smaller than any of the 20 entire wheat chromosomes; these form discrete, sortable peaks. (B) The wheat-rye (Chinese Spring/Imperial) disomic addition line 4R (2n = 44; $21_W'' + 1_R''$) forms peaks I–III and 3B, and a discrete, sortable peak harboring rye chromosome 4R.

specific accession of the donor species, for example because it harbors a specific gene or allele. Currently, two approaches have been elaborated to discriminate chromosomes without recourse to specialized cytogenetic stocks. In the first, composite peaks are divided into sections and those which are enriched for the chromosome of interest are retained (Vrána et al., submitted for publication). Although the purity level attained is necessarily lower than is achievable from well discriminated peaks, fractions with a contamination level as low as 20% can be prepared from composite peaks in wheat. Importantly, a majority of wheat chromosomes sorted in this way have proven to be free of contamination by homoeologs, which greatly simplifies sequence analysis. The second approach relies on the differential labeling of chromosomes, based on the presence of repetitive sequences. The earliest attempts to achieve this goal, as described by Macas et al. (1995), involved a modification of the PRINS (primed *in situ* DNA labeling) technique. While this did lead to some successful results (Pich et al., 1995), it was plagued by poor reproducibility and by non-quantitative labeling DNA repeats (unpublished data). Both suspended rye and barley chromosomes were labeled with fluorochromes by Ma et al. (2005), but no attempt was made to apply flow cytometry to these preparations. Finally, Giorgi et al. (2013) developed a reproducible method termed FISHIS (FISH in suspension), which differentially labels chromosomes by hybridizing with oligonucleotide probes targeting specific microsatellite sequences (Fig. 5). The successful binding of these probes may well be related either to their ability, as small molecules, to easily invade the chromosomes, or be the result of the formation of alternative B-DNA structures (Cuadrado and Jouve, 2010). As yet, it has not been established to what extent (if any) the FISHIS procedure damages chromosomal DNA and proteins, and hence with which downstream applications FISHIS-labeled chromosomes will be compatible.

Given a sorting speed 5–40 chromosomes per second (Doležel and Lucretti, 1995; Vrána et al., 2012), it is feasible to recover some 200,000 chromosomes per working day using a commercial flow-sorter (Šafář et al., 2010), a number sufficient to acquire microgram quantities of chromosome-specific DNA. The two major factors influencing the yield of sorted chromosomes are the level of resolution achievable and the quality of the initial sample (specifically, the overall number of intact chromosomes present and the amount of debris). Where aneuploid material is the source, yields can be reduced because the target chromosome is not represented in the disomic state in every seedling. The assignation of chromosome identity to flow karyotype peak is most conveniently achieved using a chromosome-specific PCR assay (Lysák et al., 1999; Vrána et al., 2000), particularly as such assays only require a small amount of DNA as a template. PCR assays are not, however, capable of estimating peak purity; in principle, this could be achieved using a quantitative PCR assay based on a set of primers designed to specifically recognize each chromosome in the genome. More straightforwardly, the chromosomal content of a given peak can be inspected by conventional microscopy following a PRINS or FISH labeling protocol (Kubaláková et al., 2000, 2005). Such an analysis of course requires a prior characterization of the karyotype.

4. Uses of flow-sorted chromosomes

Because the morphology of flow-sorted chromosomes isolated from formaldehyde-fixed root tips is well preserved (Doležel et al., 1992), high molecular weight DNA is readily derivable. As a result, flow-sorted plant chromosomes have proven valuable for a range of applications, including cytogenetic analysis, physical and genetic mapping and whole genome sequencing (Fig. 6).

4.1. Physical mapping

4.1.1. Mapping by PCR

As the template requirement for PCR is small, sorted chromosomes have proven to represent an elegant means of chromosomally assigning a given DNA sequence. This approach was adopted to map vicillin genes in field bean, since these genes were difficult to map genetically due to a paucity of allelic variation (Macas et al., 1993); similarly, genes mapping to the sex chromosome in white campion were successfully identified (Kejnovský et al., 2001; Matsunaga et al., 2003, 2005), and the genetic and physical maps of both garden pea (Neumann et al., 2002) and chickpea (Vláčilová et al., 2002; Zatloukalová et al., 2011) were successfully integrated. Macas et al. (1993) and Neumann et al. (2002) exploited sorted reciprocal translocation chromosomes in field bean and garden pea to locate a number of DNA sequences to their sub-chromosomal region. More recently, PCR amplification of template consisting of flow-sorted chromosomes has been used to develop DNA markers to support positional cloning (Šimková et al., 2011a, 2011b). Such an approach is particularly useful in allopolyploid species, where the development (and subsequent mapping) of low copy sequences can be complicated by the presence of three homoeologs.

Physical mapping applications which require a larger quantity of DNA of course require a more prolonged chromosome sorting effort, although where high molecular weight DNA is not needed, this can be avoided by the amplification of template derived from a modest number of sorted chromosomes. Šimková et al. (2008a) showed that microgram quantities of chromosomal DNA with a majority of fragments between 5 and 30 Kb can be produced using a multiple displacement amplification (MDA) protocol based on ϕ29 DNA polymerase. Starting with a 10 ng aliquot of DNA derived from a population of 10,000 barley

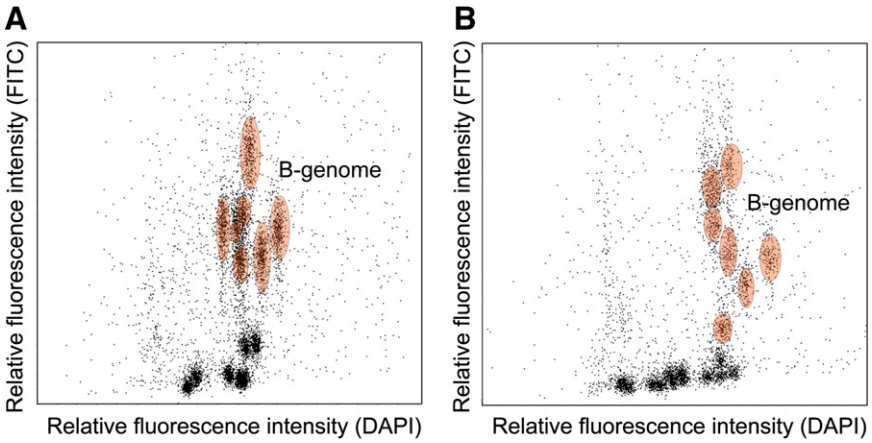


Fig. 5. Flow karyotypes of (A) tetraploid (cv. Creso, 2n = 28) and (B) hexaploid (cv. Chinese Spring, 2n = 42) wheat after the joint fluorescent labeling of GAA_n microsatellites and DAPI staining. The former was achieved by hybridization with a GAA_7-FITC probe, following the FISHIS procedure. The B genome chromosomes have a higher GAA content than either the A or D genome ones, and so can be discriminated on the basis of their higher FITC fluorescence (highlighted in orange).

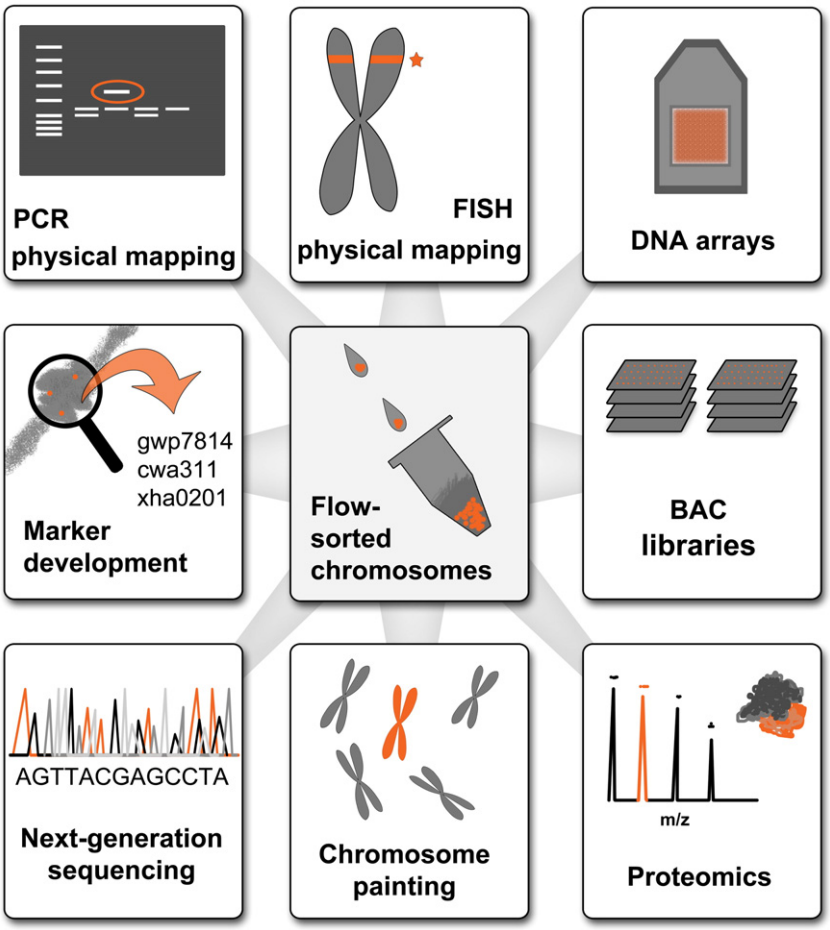


Fig. 6. Major current and potential uses of flow-sorted chromosomes.

chromosomes 1H, the amplification product proved to be very representative of the whole chromosome, since only 1.9% of SNP loci known to map to this chromosome failed to be recovered. On this basis, the chromosome 1H DNA pool was used infer a 1H location to 40 SNP loci which had hitherto not been mapped. When DNA was amplified in this way from each of the 12 individual arms of chromosomes 2H–7H maintained individually in wheat–barley telosome addition lines, 370 SNP loci which had not hitherto been genetically mapped were allocated a chromosome arm (Muñoz-Amatriaín et al., 2011). Prior to using 7H-specific simple sequence repeat (SSR) markers to characterize a spontaneous wheat–barley Robertsonian translocation, Cseh et al. (2011) were able to verify their chromosomal arm location by testing against a template of flow-sorted chromosome arms 7HS and 7HL. Chromosome sorting was also exploited for the positional cloning of a powdery mildew resistance gene located on wheat chromosome arm 4AL (Jakobson et al., 2012).

4.1.2. Construction of clone-based physical maps

The construction of a physical map as a template for either CBC sequencing or positional cloning requires large insert genomic DNA libraries, most commonly generated in the form of bacterial artificial chromosomes (BACs), which are able to accommodate an insert of up to several hundred Kb in length (Shizuya et al., 1992). The quantity of high molecular weight DNA required for this purpose is in the microgram range, so achieving this from flow-sorted material involved the elaboration of a customized protocol (Šimková et al., 2003). Using this protocol, Šafář et al. (2004) succeeded in constructing the first documented chromosome-specific BAC library of a eukaryotic organism; the chromosome involved was wheat 3B, and was sourced from a set of two million sorted chromosomes, prepared over 18 working days. The library comprised about 68,000 clones with a mean insert size of 103 Kb, and represented more than 6 x coverage of the chromosome. Further improvements to the protocol extended the coverage to >15x and the mean insert size to >120 Kb (IEB genomic resources database, 2013; Šafář et al., 2010). In addition to a number of wheat whole chromosome- and chromosome arm-specific BAC libraries, a library has also been constructed from the short arm of cereal rye chromosome 1R (Šimková et al., 2008b).

The International Wheat Genome Sequencing Consortium (IWGSC) has chosen a CBC chromosome-based strategy to produce a reference sequence of the wheat genome (Feuillet and Eversole, 2007) and so a chromosome-specific BAC library has been generated for each of the 21 chromosomes of the model cultivar Chinese Spring (IEB genomic resources database, 2013). The feasibility of constructing a physical map of each wheat chromosome based on such libraries was confirmed by the successful contig map of chromosome 3B produced by fingerprinting the 3B BAC library (Paux et al., 2008). The initial version of the map comprised just over 1000 contigs anchored with nearly 1500 molecular markers, and represented 82% of the chromosome. The lessons drawn from this exercise have been incorporated into the ongoing effort to establish a physical map for each of the remaining 20 wheat chromosomes (Lucas et al., 2013; Philippe et al., 2013; Sehgal et al., 2012). The sequencing of 13 of the 3B contigs involved over 150 BACs (Paux et al., 2008), and led to the annotation of >18 Mb of sequence. While the global gene density was found to be about one per 104 Kb, some 75% of the genes clustered into small groups (each containing on average three genes), and the density increased by two fold in regions close to the telomere, largely as a consequence of tandem and interchromosomal duplications. Using the same physical map, Rustenholz et al. (2011) were able to locate some 3,000 genes, distributed along the

whole chromosome, and a similar pattern of gene islands and greater gene density at the chromosome ends emerged. Most of the gene islands resulted from interchromosomal duplications specific to polyploid wheat and are enriched in genes sharing the same function or expression profile. Gene space organization and evolution proved to be similar on chromosome arm 1BL (Philippe et al., 2013). The definition of an MTP for both chromosome 3B and chromosome arm 3DS enabled Bartoš et al. (2012) to attempt a comparison of the molecular organization of these two homoeologs. What was revealed was a similar rate of non-collinear gene insertion, with the majority of duplications occurring prior to the divergence of the B and D genomes some 30 Mya. One third of insertions occurred during the past 2.5–4.5 My, leading to the suggestion that gene insertion was accelerated by allopolyploidisation. Pseudogenes appear to represent only a small fraction of the non-collinear genic sequence; for the most part, they seem to have arisen during the evolution of the polyploid wheat genome and not from insertion of non-functional genes.

Beyond their utility for acquiring the genome sequence of wheat, the chromosome-specific BAC libraries have found a number of other uses. The chromosome 7DL and 7DS libraries have been queried with markers linked to the aphid resistance genes *Dn2401* and *Gb3* (Šimková et al., 2011a, 2011b). Both PCR- and hybridization-based screening has demonstrated the gain in efficiency brought about by the reduction in complexity of the template. Thus, just three rounds of screening on three high density filters were sufficient to build a BAC contig spanning *Gb3*. To achieve the positional cloning of genes not present in cv. Chinese Spring (such as *Gb3*), other cultivars have been targeted for making chromosome-specific BAC libraries (IEB genomic resources database, 2013; Janda et al., 2006). Chromosome-specific BAC libraries are especially valuable in polyploids as they avoid the problem of homoeology. Additionally the necessary size of such libraries is an order of magnitude lower and so are more straightforward to store, handle and screen (Šimková et al., 2011a). Finally, the dissection of a large genome into its constituent chromosomes parts helps to structure collaborative projects where each of the various partners can be made responsible for the management of a specific chromosome(s), even though the BAC libraries have been generated centrally.

4.1.3. Cytogenetic mapping

Ordering and orienting BAC and sequence contigs is an important step in, respectively, building a clone-based physical map and assembling a shotgun sequence. Genetic markers are seldom helpful in proximal chromosome regions because these are associated with a low frequency of recombination. An alternative means of ordering is to apply FISH to mitotic or meiotic chromosomes (Karafiátová et al., 2013; Tang et al., 2009). A development of this idea is to apply FISH to mechanically stretched (by 100 fold), flow-sorted mitotic metaphase chromosomes (Valárik et al., 2004). The stretching greatly improves the achievable level of spatial resolution, to an extent where the individual probes can be ordered.

4.2. Genetic marker development

The reduction in template complexity achieved by targeting individual chromosomes has been beneficial for genetic marker development. Markers are a critical resource for the construction of genetic linkage maps, the understanding of trait inheritance, the assembly of physical maps and DNA shotgun sequences, and positional cloning. An array of marker types have been developed, the most ubiquitously used of which in plant genetics are SSRs, DArTs (diversity array technology), ISBPs (insertion site based polymorphisms) and SNPs (single nucleotide polymorphisms) (Poczai et al., 2013). Various strategies have been employed to base marker development on chromosome-specific libraries.

4.2.1. SSRs and ISBPs

An initial approach began by cloning the DNA derived from sorted chromosomes (generally following an amplification step) to generate small insert DNA libraries (Macas et al., 1996). The earliest significant marker type was hybridization-based, such as the RFLPs (restriction fragment length polymorphisms) generated on tomato chromosome 2 by Arumuganathan et al. (1994). With the advent of PCR, attention switched to SSRs. An enrichment was carried out on the initial library to bias the recovery of a target microsatellite motif (Koblížková et al., 1998). Požárková et al. (2002) developed a set of SSR markers from chromosome 1 of field bean, and some of these were later used to elaborate a genetic map of the species (Román et al., 2004). Kofler et al. (2008) employed a similar approach to develop 57 SSR markers from MDA-amplified DNA of rye chromosome arm 1RS, a source of a number of agronomically important genes for wheat (Lukaszewski, 1990). In addition to developing SSR markers from amplified 1RS DNA, Kofler et al. (2008) also developed 138 SSR assays from 2778 BAC end sequences (BES) obtained from the 1RS-specific BAC library The same set of BES included 249 transposable element junctions which could be exploited to produce 64 ISBP markers, of which 12 were 1RS specific (Bartoš et al., 2008). BES derived from chromosome-specific libraries have proven informative for marker development in wheat itself as well. For example, Lucas et al. (2012) identified 433 potential SSRs and 9,338 potential ISBP sequences from ~13,500 BES generated from chromosome arm 1AL. About one half of the putative ISBP markers tested proved to be functional. Similarly, among ~10,000 3AS BES, Sehgal et al. (2012) identified over 1,000 potential SSR and nearly 8,000 potential ISBP sequences, of which an estimated 18% and 29%, respectively, marked loci on 3AS.

4.2.2. DArT markers

Wenzl et al. (2010) demonstrated how useful chromosome sorting is to develop DArT markers to significantly increase saturation of linkage maps at specific genome regions. Using DNA of chromosome 3B and chromosome arm 1BS of wheat, the authors developed DArT arrays with 2,688 and 384 clones, respectively. Out of 711 polymorphic 3B-derived markers, 553 (78%) mapped to chromosome 3B, while 59 of 68 polymorphic 1BS-derived markers (87%) mapped to chromosome arm 1BS. Hence a majority of markers were specific to target chromosomes. The 3B DArT array was used in development of a new consensus genetic map of the chromosome, leading to doubling the number of genetically distinct loci on 3B. The efficiency of chromosome targeting can be estimated by comparing the 510 polymorphic 3B markers obtained by screening 2,688 3B-derived clones with 269 polymorphic markers identified by screening approximately 70,000 whole genome-derived clones (Wenzl et al., 2010). Coupling chromosome sorting with the DArT platform is straightforward, as the DNA requirement is only ~5 ng, a quantity which can be recovered in less than one hour of flow-sorting.

4.2.3. Marker development from chromosome-specific shotgun sequences

The combination of MDA-generated chromosome-specific DNA and high throughput sequencing platforms offers an efficient route towards whole genome shotgun sequencing and the *in silico* identification of genetic markers. The development of a SNP map of wheat chromosome 3B serves as a good example of the power of this approach (Shatalina et al., 2013). A set of 737 gene-containing contigs harboring chromosome 3B SNPs between the two cultivars Arina and Forno was selected, and a subset of 96 of these SNPs used to genotype an Arina x Forno recombinant inbred line population; of these, 70 mapped to the expected chromosome. The 454-derived sequence of rye chromosome arm 1RS allowed Fluch et al. (2012) to identify >4000 potential SSR loci, and similarly Nie et al. (2012) used Illumina-derived sequence of wheat chromosome arm 7DL to identify >16,000 putative SSR loci. When a random set of 33 of the latter was tested by PCR, 18 proved to be informative across a panel of 20 cultivars. Similarly, the 454-derived sequence produced from wheat chromosome arm 1BL (Wicker et al., 2011) was used by Philippe et al. (2013) to identify nearly 19,000 putative ISBPs and 200

SSRs. Finally, a comparison of homoeologous group 7 sequences across four Australian wheat cultivars located some 900,000 informative SNP loci (Berkman et al., 2013).

4.2.4. Marker specificity

A feature of the chromosome-based strategy is that it can save a substantial volume of screening effort, particularly in polyploid species. Thus, for example, Požárková et al. (2002) were able to use flow-sorted fractions as a PCR template to verify the chromosome specificity of SSR markers in filed bean. Michalak de Jimenez et al. (2013) used a radiation hybrid approach to map wheat chromosome 1D, exploiting DNA amplified from the homoeologous group 1 chromosomes as a source of 1D-specific markers. Shotgun sequences of each chromosome of barley (Mayer et al., 2011), rye (Martis et al., 2013) and bread wheat (K Eversole, pers. comm.) have now been acquired using either the Illumina or the 454 platform; thus it should be in future possible to rapidly verify chromosome-specificity *in silico* in these species.

4.3. Sequencing

4.3.1. BAC clones

NGS technology has the capacity to shotgun-sequence whole genomes, but the quality of genome assembly in large genome species is poor compared to that obtained using the CBC method, as used to derive the reference sequences of *A. thaliana*, rice and maize (Shangguan et al., 2013). Handling a genomic BAC library of a large genome species is cumbersome, because of the number of clones involved. Particular problems are associated with the presence of homoeology in polyploid genomes. A chromosome-based strategy at present represents the most promising one in these cases, and has been adopted for the acquisition of the hexaploid wheat genome sequence (The International Wheat Genome Sequencing Consortium, 2013); so far it has generated a 1 Gb reference sequence of chromosome 3B after sequencing its MTP using a combination of Roche 454 and Illumina technologies (Choulet et al., submitted for publication). The project of the International Wheat Genome Sequencing Consortium involves the construction of a full set of chromosome-specific BAC libraries, the definition of an MTP for each, and the CBC-sequencing of the MTP using NGS.

The availability of a number of chromosome-specific BAC libraries has already provided some interesting research opportunities. Bartoš et al. (2008) end-sequenced a random set of 1,536 clones from a BAC library specific for the short arm of rye chromosome 1R (1RS). The analysis of repeat content indicated a similar fraction of repeats as in the B genome of wheat (84%). However, as the rye genome is much larger (almost 8 Gb/1C vs. ~5.6 Gb/1C), a lower than expected proportion of repeats was probably due to insufficient representation of rye repeats in DNA sequence databases that were searched to identify repeats. Since only 0.9% of the 1RS derived BES were classified as genic sequences, it was estimated that the arm harbored about 2000 genes. A similar analysis of ~10,000 3AS BACs led to an estimate that the proportion of repetitive DNA present was 79% (Sehgal et al., 2012). About 1.4% of the DNA was estimated to represent coding sequence, producing an estimated 2,850 genes as present on the arm, the length of which is just 0.8 times the size of the entire rice genome, which is estimated to harbor over 45,000 genes (Yu et al., 2002). An increase in gene density towards the telomere was noted, and for up to 30% of the genes, synteny was not maintained with the rice, sorghum and *B. distachyon* genomes. Similarly, Lucas et al. (2012) used >13,000 1AL BES to characterize the composition of this chromosome arm, producing an estimate of ~1.0% for the proportion of the arm's DNA which represented coding sequence and a gene number of 4700. The analysis confirmed the presence of two known major synteny blocks (Mayer et al., 2009), as well as three smaller blocks not previously identified.

4.3.2. Whole chromosome sequencing using 454 technology

The combination of NGS technology and chromosome sorting currently represents the most affordable means of obtaining the sequence composition of single chromosomes. Generally, MDA-amplified DNA, which typically generates fragments in the size range 5–30 Kb (Šimková et al., 2008a), is suitable for this purpose. However, it is unsuitable for constructing paired-end and mate-pair libraries with insert sizes >3 Kb (Belova et al., 2013). If longer insert sequencing libraries are needed, the amplification step should be avoided and a larger number of chromosomes need to be sorted. Amplified chromosomal DNA from barley chromosome 1H was sequenced using the 454 technology by Mayer et al. (2009). Comparison of the sequences with genes of rice and sorghum and with EST datasets of barley and wheat identified 5400 genes. Based on the integration with synteny data from the two grass model species, the authors proposed a virtually ordered inventory of 1987 genes and their work increased the number of 1H anchored genes by 6-fold compared to previous map resources. Mayer et al. (2011) exploited the same approach by adding low-pass 454-acquired sequence from the other barley chromosomes, incorporating at the same time all available full length cDNA sequence and DNA microarray hybridization data. The result was a sequence-based gene map of barley capturing an estimated 86% of the total gene content. This so-called "Genome Zipper" approach is illustrated in Fig. 7, and has succeeded in precisely localizing six of the seven barley centromeres, and established gene order in the poorly recombining proximal chromosome regions. Due to its relative simplicity, Genome Zipper is an attractive approach for all species, whose genomes have not been sequenced and in which chromosomes can be isolated by flow-sorting.

The possibility of sequencing all six arms of the wheat group 1 homoeologs allowed Wicker et al. (2011) to make structural comparisons at the single chromosomal level. Analysis of sequences from low-pass sequencing with Roche 454 technology (1.3- to 2.2x chromosome coverage) indicated that all three wheat subgenomes have similar sets of genes that are syntenic with model grass genomes. However, the number of genic sequences that have their homologs outside the group 1 syntenic region in the grass models outnumbers the syntenic ones. Further analysis indicated that a large proportion of the genes that are found in only one of the three homoeologous wheat chromosomes were most probably pseudogenes resulting from transposon activity and double strand break repair. The 1A sequences were later used by Lucas et al. (2013) to produce a virtual gene order along chromosome arm 1AL, adopting the Genome Zipper approach, and this was readily integrated into a physical map of the arm. The analysis confirmed the presence of non-syntenic genes and identified some putative translocations.

Vitulo et al. (2011) characterized the content of wheat chromosome 5A by acquiring 454-derived sequence from each arm. Their estimate was that coding sequence represented 1.1% of 5AS and 1.3% of 5AL, leading to the prediction that the whole chromosome harbors just over 5,000 genes. Similarly, Hernandez et al. (2012) analyzed chromosome 4A, a chromosome which has undergone a major series of evolutionary re-arrangements (Devos et al., 1995). Application of the Genome Zipper method produced a virtual gene map capturing at least 85% of the chromosome's estimated gene content. A comparison with the maps of barley chromosomes 4H, 5H and 7H identified and ordered five distinct regions (Fig. 8), the gene content and order within each of which being inferred from synteny. A 454-derived sequence of both arms of chromosome 3A recognized over 3500 contigs (Akhunov et al., 2013). A comparison with the equivalent sequences of the model grass genomes detected that some 35% of genes had experienced structural rearrangements leading to a variety of mis-sense and non-sense mutations. In particular, 38% of these genes were affected by a premature stop codon, which is on line with other studies indicating ongoing pseudogenization of the wheat genome. Alternative splicing patterns were diverse between homoeologs, perhaps an effect of the genetic redundancy resulting from polyploidy.

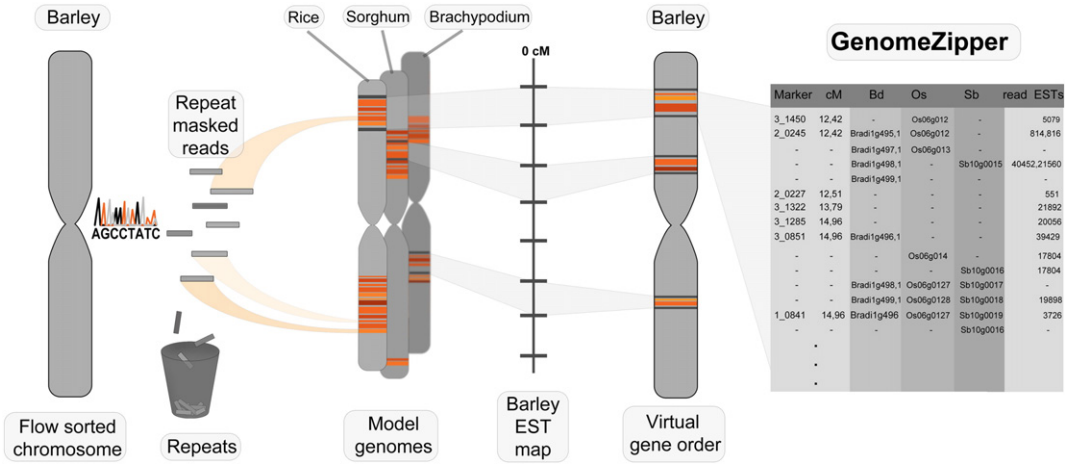


Fig. 7. Genome Zipper analysis in barley chromosome. MDA-amplified DNA of flow-sorted barley chromosomes was sequenced by 454 technology. Repeat-masked sequence reads identified syntenic regions in the rice, sorghum and *B. distachyon* genomes. Genes located in these regions were then aligned with a EST-based barley map of barley, which served as a scaffold to anchor collinear segments derived from the non-barley genomes. Genic sequence reads of barley were integrated and ordered by assuming collinearity within syntenic regions, leading to the derivation of a virtual gene map of barley.

The 454 technology has also been applied to sequence 1RS chromosome arm of rye, revealing the presence of just over 3000 gene loci and identifying syntenic regions in model genomes of rice and brachypodium, and in barley chromosome 1H (Fluch et al., 2012). The subsequent 454-based sequencing of all chromosomes of rye established their virtual linear gene order models (genome zippers) comprising over 22,000 or 72% of the detected set of ~31,000 rye genes (Martis et al., 2013). The study revealed six major translocations that shaped the modern rye genome in comparison to a putative Triticeae ancestral genome. Moreover, the results indicated that introgressive hybridizations and/or a series of whole-genome or chromosome duplications played a role in rye speciation and genome evolution.

A very attractive application of flow cytometric sorting is to isolate specialized chromosomes such as sex chromosomes and supernumerary B chromosomes. Since B chromosomes act as a selfish genetic element, they have been proposed as a vehicle for chromosome-mediated gene transfer (Birchler et al., 2008). The structure of rye B chromosomes has been elucidated by sequencing flow-sorted material using the 454 platform (Martis et al., 2012). Although they have long been considered to be gene poor (Jones, 1995; Jones and Houben, 2003), a sequence alignment with rice, *B. distachyon*, sorghum and barley genomic sequence identified the presence of almost 5000 putative gene fragments. A strong indication was that their DNA probably originated from both chromosome arm 3RS and chromosome 7R, although the sequence appears to have been subjected to complex rearrangement. Molecular clock-based dating of the rye B chromosomes' origin places it at 1.1–1.3 Mya, which is not long after the formation of the genus *Secale* (1.7 Mya).

4.3.3. Whole chromosome sequencing using Illumina technology

The initial attempts at shotgun sequencing of flow-sorted plant chromosomes were based on the 454 platform, which generates read lengths of several hundred nucleotides. With the development of the Illumina platform, Berkman et al. (2011) were able to demonstrate that short read sequencing technology could equally be used for chromosome shotgun sequencing and subsequent assembly. Thus, a coverage of >30× was achieved for chromosome arm 7DS, and the subsequent assembly comprised over 550,000 contigs (up to 32.6 Kb in length) with an N50 of 1159 bp. The coverage represented approximately 40% of the whole arm, since much of the repetitive DNA collapsed into a single contig. A comparison with the *B. distachyon* sequence identified nearly 1,500 genes, of which about one in three were non-syntenous. A comparison with bin-mapped wheat ESTs (Qi et al., 2004) highlighted possible erroneous allocations, with the result that the 7DS assembly probably captured all or nearly all of the arm's gene content. The same approach was used to sequence and assemble chromosome arm 7BS (Berkman et al., 2012). A comparison between the assemblies of 7DS, 7BS and 4AL recognized the known evolutionary translocation between chromosomes 7B and 4A and closely defined its break-point. The level of collinearity between 7BS and 7DS was 84%, while that between the wheat and *B. distachyon* was 60%. Extending the approach to cover the whole of the group 7 homoeologs showed that there has been more gene loss in 7A and 7B than in 7D (Berkman et al., 2013).

Micro RNAs (miRNAs) are an important component of post-transcriptional gene regulation, so their distribution at the chromosome

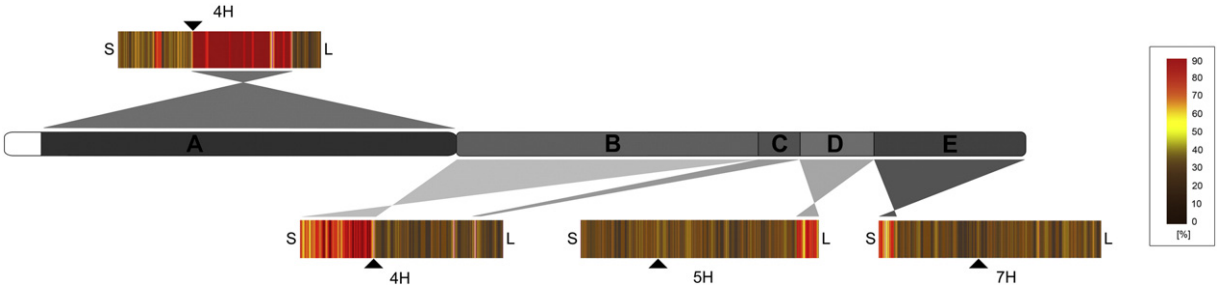


Fig. 8. The 4A shotgun sequence of barley. Repeat-masked 4AS and 4AL shotgun sequence reads were compared with the sequence of virtual barley chromosomes (Mayer et al., 2011). Syntenic regions on chromosomes 4H, 7H and 5H are shown in red, and non-syntenic regions in brown. The centromeres are indicated by black triangles and the chromosome arms are labeled S and L. Connectors indicate corresponding segments and the orientation of the individual segments. Taken with permission from Hernandez et al., Plant Journal 2012;69:377–86, John Wiley & Sons Ltd. Modified.

level is expected to provide novel insights into genome organization and function. Vitulo et al. (2011) used chromosome shotgun sequences to identify 195 candidate miRNA precursors belonging to 16 miRNA families on chromosome 5A, while Kantar et al. (2012), focusing on chromosome 4A, found 68 different miRNAs of which 37 had not been observed previously in wheat. The two chromosome arms differed with respect to both the variety and representation of miRNAs. Among the 62 putative targets identified, 24 were found to give hits to expressed sequences.

4.3.4. Validation of whole genome assemblies

Many genomes have already been sequenced using NGS shotgun approach, and it is not realistic to expect that they will be sequenced again following CBC strategy. Additional approaches are therefore needed to improve the assemblies. These may include improved bioinformatics tools for whole genome assembly, incorporation of sequences obtained using methods resulting in longer reads (Roberts et al., 2013), optical mapping (Dong et al., 2013) and mapping on nanochannel arrays (Hastie et al., 2013; Lam et al., 2012). Cytogenetic mapping has a role to play in the verification of sequence assemblies (Febrer et al., 2010; Islam-Faridi et al., 2009). However, a powerful option is to sequence isolated chromosomes using NGS and compare chromosome-derived sequences with whole genome assemblies. Preliminary results obtained with genome assemblies of two types of chickpea (Jain et al., 2013; Varshney et al., 2013) highlighted regions that appear to have been mis-assembled and provided the basis for genome assembly improvement (R. Varshney and D. Edwards, pers. comm.). Thus, chromosome genomics can be employed in genome sequencing projects to validate and assist in the accurate sequence assemblies obtained by NGS shotgun.

5. Conclusions

The recent past few years have witnessed marked progress in chromosome genomics, a technology which has rapidly established itself as a facilitator of mapping and sequencing of plant genomes. The number of species tractable to flow-sorting has expanded, confirming the broad applicability of suspensions of intact chromosomes obtained from synchronized root tips (Doležel et al., 1992). The development of the FISHIS technique (Giorgi et al., 2013) should expand the reach of flow-sorting, since it provides a powerful means of discriminating between chromosomes which are similar in size, thereby easing the dissection of complete genomes into their individual chromosome components. There has also been a notable increase in the number and variety of applications using flow-sorted chromosomes, driven most importantly by the step change in sequencing power achieved by NGS technologies, but also by the possibility of producing microgram quantities of chromosomal DNA via MDA. Chromosome genomics has been especially useful in species lacking a reference genome sequence. The analysis of sequence at the single chromosome level has provided new insights into the structure of complex, and particularly polyploid genomes, where comparisons between homoeologs has informed the process of genome evolution in a polyploid setting. Sequencing single chromosomes has been highly productive in the context of marker development and validation. Finally, chromosome-specific shotgun sequences are proving to represent a convenient means of verifying genome sequence assemblies, of identifying candidate genes and of analyzing the organization and evolution of specialized chromosomes such as sex chromosomes and supernumerary B chromosomes.

The chromosome genomics approach has been particularly fruitful in the wheat genome, the analysis of which using a whole genome approach is hampered by the size of the genome and the presence of homoeologs. The current international effort coordinated by IWGSC to sequence the wheat genome has therefore been largely based on the construction of ready-to-sequence chromosome arm-specific BAC libraries. The experience gained in this task already suggests that chromosome genomics can contribute materially to the analysis of genomes lacking a high quality reference sequence. A number of potential applications still remain to be addressed. A prime example is chromosome mapping on nanochannel arrays (Lam et al., 2012), the availability of which would ease the initial assembly and validation of genome sequences. The organization of the chromosomes during interphase and their behavior during most of both mitosis and meiosis are difficult to unravel in large genome species in the absence of chromosome painting probes; isolated single chromosomes would certainly offer an excellent opportunity to develop these. As the function of the nuclear genome is intimately linked to DNA organization and the architecture of the interphase nucleus, there is also a need to study chromatin proteins and their dynamics. A proteomic analysis of flow-sorted chromosomes should represent an attractive approach to study chromatin free of contaminating cytoplasmic components.

Acknowledgements

We thank Drs. Eva Hřibová and David Kopecký for their help with FISHIS. We appreciate the useful comments made by the three anonymous reviewers. This research was supported by the Czech Science Foundation (awards P501/10/1740, P501/10/1778 and P501/12/G090) and by the Ministry of Education, Youth and Sports of the Czech Republic and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. ED0007/01/01).

References

Akhunov ED, Sehgal S, Liang H, Wang S, Akhunova AR, Kaur G, et al. Comparative analysis of syntenic genes in grass genomes reveals accelerated rates of gene structure and coding sequence evolution in polyploid wheat. Plant Physiol 2013;161:252–65.
Arumuganathan K, Slattery JP, Tanksley SD, Earle ED. Preparation and flow cytometric analysis of metaphase chromosomes of tomato. Theor Appl Genet 1991;82:101–11.
Arumuganathan K, Martin GB, Telenius H, Tanksley SD, Earle ED. Chromosome 2-specific DNA clones from flow-sorted chromosomes of tomato. Mol Gen Genet 1994;242:551–8.
Bartoš J, Paux E, Kofler R, Havránková M, Kopecký D, Suchánková P, et al. A first survey of the rye (*Secale cereale*) genome composition through BAC end sequencing of the short arm of chromosome 1R. BMC Plant Biol 2008;8:95.
Bartoš J, Vlček Č, Choulet F, Džunková M, Cviková K, Šafář J, et al. Intraspecific sequence comparisons reveal similar rates of non-collinear gene insertion in the B and D genomes of bread wheat. BMC Plant Biol 2012;12:155.
Belova T, Zhan B, Wright J, Caccamo M, Asp T, Šimková H, et al. Integration of mate pair sequences to improve shotgun assemblies of flow-sorted chromosome arms of hexaploid wheat. BMC Genomics 2013;14:222.
Berkman PJ, Skarshewski A, Lorenc MT, Lai K, Duran Ch, Ling EYS, et al. Sequencing and assembly of low copy and genic regions of isolated *Triticum aestivum* chromosome arm 7DS. Plant Biotechnol J 2011;9:768775.
Berkman PJ, Skarshewski A, Manoli S, Lorenc MT, Stiller J, Smits L, et al. Sequencing wheat chromosome arm 7BS delimits the 7BS/4AL translocation and reveals homoeologous gene conservation. Theor Appl Genet 2012;124:423–32.
Berkman PJ, Visendi P, Lee HC, Stiller J, Manoli S, Lorenc MT, et al. Dispersion and domestication shaped the genome of bread wheat. Plant Biotechnol J 2013;11:564–71.
Birchler JA, Yu W, Han F. Plant engineered minichromosomes and artificial chromosome platform. Cytogenet Genome Res 2008;120:228–32.
Chen W, Kalscheuer V, Tzschach A, Menzel C, Ullmann R, Schulz MH, et al. Mapping translocation breakpoints by next-generation sequencing. Genome Res 2008;18:1143–9.
Choulet J, Alberti A, Theil S, Glover N, Barbe V, Daron J, et al. A reference sequence of wheat chromosome 3B reveals structural and functional compartmentalization. Science 2014. (submitted for publication).
Conia J, Bergounioux C, Perennes C, Muller P, Brown S, Gadal P. Flow cytometric analysis and sorting of plant chromosomes from *Petunia hybrida* protoplasts. Cytometry 1987;8:500–8.
Conia J, Muller P, Brown S, Bergounioux C, Gadal P. Monoparametric models of flow cytometric karyotypes with spreadsheet software. Theor Appl Genet 1989;77:295–303.
Cseh A, Kruppa K, Molnár I, Rakszegi M, Doležel J, Molnár-Láng M. Characterization of a new 4BS.7HL wheat–barley translocation line using GISH, FISH, and SSR markers and its effect on the β-glucan content of wheat. Genome 2011;54:795–804.
Cuadrado A, Jouve N. Chromosomal detection of simple sequence repeats (SSRs) using nondenaturating FISH (ND-FISH). Chromosoma 2010;119:495–503.
de Laat AMM, Blaas J. Flow-cytometric characterization and sorting of plant chromosomes. Theor Appl Genet 1984;67:463–7.
de Laat AMM, Schel JHN. The integrity of metaphase chromosomes of *Haplopappus gracilis* (Nutt) Gray isolated by flow cytometry. Plant Sci 1986;47:145–51.

Devos KM, Dubcovsky J, Dvorak J, Chinoy CN, Gale MD. Structural evolution of wheat chromosomes 4A, 5A, and 7B and its impact on recombination. Theor Appl Genet 1995;91:282–8.
Doležel J, Lucretti S. High-resolution flow karyotyping and chromosome sorting in *Vicia faba* lines with standard and reconstructed karyotypes. Theor Appl Genet 1995;90: 797–802.
Doležel J, Číhalíková J, Lucretti S. A high-yield procedure for isolation of metaphase chromosomes from root tips of *Vicia faba* L. Planta 1992;188:93–8.
Doležel J, Lucretti S, Schubert I. Plant chromosome analysis and sorting by flow cytometry. Crit Rev Plant Sci 1994;13:275–309.
Doležel J, Kubaláková M, Paux E, Bartoš J, Feuillet C. Chromosome-based genomics in cereals. Chromosome Res 2007a;15:51–66.
Doležel J, Greilhuber J, Suda J. Flow cytometry with plants: an overview. In: Doležel J, Greilhuber J, Suda J, editors. Flow cytometry with plant cells. Weinheim: Wiley-VCH Verlag GmbH & Co. KGaA; 2007b. p. 41–65.
Doležel J, Kubaláková M, Číhalíková J, Suchánková P, Šimková H. Chromosome analysis and sorting using flow cytometry. In: Birchler JA, editor. Plant chromosome engineering. Methods and protocols. New York: Springer; 2011. p. 221–38.
Dong Y, Xie M, Jiang Y, Xiao N, Du X, Zhang W, et al. Sequencing and automated whole-genome optical mapping of the genome of a domestic goat (*Capra hircus*). Nat Biotechnol 2013;31:135–41.
Dudin G, Steegmayer EW, Vogt P, Schnitzer H, Diaz E, Howell KE, et al. Sorting of chromosomes by magnetic separation. Hum Genet 1988;80:111–6.
Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Edward S, et al. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. PLoS One 2011;6:e19379.
Febrer M, Goicoechea JL, Wright J, McKenzie N, Song X, Lin J, et al. An integrated physical, genetic and cytogenetic map of *Brachypodium distachyon*, a model system for grass research. PLoS One 2010;5:e13461.
Feuillet C, Eversole K. Physical mapping of the wheat genome: a coordinated effort to lay the foundation for genome sequencing and develop tools for breeders. Isr J Plant Sci 2007;55:307–13.
Feuillet C, Leach JE, Rogers J, Schnable PS, Eversole K. Crop genome sequencing: lessons and rationales. Trends Plant Sci 2011;16:77–88.
Ferguson-Smith MA. Genetic analysis by chromosome sorting and painting: phylogenetic and diagnostic applications. Eur J Hum Genet 1997;5:253–65.
Fluch S, Kopecky D, Burg K, Šimková H, Taudien S, Petzold A, et al. Sequence composition and gene content of the short arm of rye (*Secale cereale*) chromosome 1. PLoS One 2012;7:e30784.
Fuchs J, Houben A, Brandes A, Schubert I. Chromosome 'painting' in plants—A feasible technique? Chromosoma 1996;104:315–20.
Gerstein MB, Kundaje A, Hariharan M, Landt SG, Yan KK, Cheng C, et al. Architecture of the human regulatory network derived from ENCODE data. Nature 2012;489:91–100.
Gill KS, Arumuganathan K, Le JH. Isolating individual wheat (*Triticum aestivum*) chromosome arm by flow cytometric analysis of ditelosomic lines. Theor Appl Genet 1999;98:1248–52.
Giorgi D, Farina A, Grosso V, Gennaro A, Ceoloni C, Lucretti S. FISHIS: fluorescence *in situ* hybridization in suspension and chromosome flow sorting made easy. PLoS One 2013;8:e57994.
Grosso V, Farina A, Gennaro A, Giorgi D, Lucretti S. Flow sorting and molecular cytogenetic identification of individual chromosomes of *Dasypyrum villosum* L. (*H. villosa*) by a single DNA probe. PLoS One 2012;7:e50151.
Gualberti G, Doležel J, Macas J, Lucretti S. Preparation of pea (*Pisum sativum* L.) chromosome and nucleus suspensions from single root tips. Theor Appl Genet 1996;92: 744–51.
Guo DW, Ma YZ, Li LC, Chen YF. Flow sorting of wheat chromosome arms from the ditelosomic line 7BL. Plant Mol Biol Rep 2006;24:23–31.
Hastie AR, Dong L, Smith A, Finklestein J, Lam ET, Naxin Huo N, et al. Rapid genome mapping in nanochannel arrays for highly complete and accurate *de novo* sequence assembly of the complex *Aegilops tauschii* genome. PLoS One 2013;8:e55864.
Hernandez P, Martis M, Dorado G, Pfeifer M, Gálvez S, Schaaf S, et al. Next-generation sequencing and syntenic integration of flow-sorted arms of wheat chromosome 4A exposes the chromosome structure and gene content. Plant J 2012;69:377–86.
Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L, Chang TH, et al. Architecture and evolution of a minute plant genome. Nature 2013;498:94–8.
Islam-Faridi MN, Nelson CD, DiFazio SP, Gunter LE, Tuskan GA. Cytogenetic analysis of Populus trichocarpa–ribosomal DNA, telomere repeat sequence, and marker-selected BACs. Cytogenet Genome Res 2009;125:74–80.
Jaillon O, Aury JM, Noel B, Policriti A, Clepet C, Casagrande A, et al. The grapevine genome sequence suggests ancestral hexaploidization in major angiosperm phyla. Nature 2007;449:463–7.
Jain M, Misra G, Patel RK, Priya P, Jhanwar S, Khan AW, et al. A draft genome sequence of the pulse crop chickpea (*Cicer arietinum* L.). Plant J 2013;74:715–29.
Jakobson I, Reis D, Tiidema A, Peusha H, Timofejeva L, Valárik M, et al. Fine mapping, phenotypic characterization and validation of non-race-specific resistance to powdery mildew in a wheat–*Triticum militinae* introgression line. Theor Appl Genet 2012;125:609–23.
Janda J, Šafář J, Kubaláková M, Bartoš J, Kovářová P, Suchánková P, et al. Advanced resources for plant genomics: BAC library specific for the short arm of wheat chromosome 1B. Plant J 2006;47:977–86.
Jones RN. Tansley review no. 85. B chromosomes in plants. New Phytol 1995;131:411–34.
Jones N, Houben A. B chromosomes in plants: escapees from the A chromosome genome? Trends Plant Sci 2003;8:417–23.
Kantar M, Akpınar BA, Valárik M, Lucas SJ, Doležel J, Hernández P, et al. Subgenomic analysis of microRNAs in polyploid wheat. Funct Integr Genomics 2012;12:465–79.
Karafiátová M, Bartoš J, Kopecký D, Ma L, Sato K, Houben A, et al. Mapping nonrecombining regions in barley using multicolor FISH. Chromosome Res 2013;21:739–51.
Kato A. Air drying method using nitrous oxide for chromosome counting in maize. Biotech Histochem 1999;74:160–6.
Kejnovský E, Vrána J, Matsunaga S, Souček P, Široký J, Doležel J, et al. Localization of male-specifically expressed MROS genes of *Silene latifolia* by PCR on flow-sorted sex chromosomes and autosomes. Genetics 2001;158:1269–77.
Koblížková A, Doležel J, Macas J. Subtraction with 3′ modified oligonucleotides eliminates amplification artifacts in DNA libraries enriched for microsatellites. Biotechniques 1998;25:32–8.
Kofler R, Bartoš J, Gong L, Stift G, Suchánková P, Šimková H, et al. Development of microsatellite markers specific for the short arm of rye (*Secale cereale* L.) chromosome 1. Theor Appl Genet 2008;117:915–26.
Kopecký D, Martis M, Číhalíková J, Hřibová E, Vrána J, Bartoš J, et al. Genomics of meadow fescue chromosome 4 F. Plant Physiol 2013;163:1323–37.
Kovářová P, Navrátilová A, Macas J, Doležel J. Chromosome analysis and sorting in *Vicia sativa* using flow cytometry. Biol Plant 2007;51:43–8.
Kubaláková M, Lysák MA, Vrána J, Šimková H, Číhalíková J, Doležel J. Rapid identification and determination of purity of flow-sorted plant chromosomes using C-PRINS. Cytometry 2000;41:102–8.
Kubaláková M, Vrána J, Číhalíková J, Šimková H, Doležel J. Flow karyotyping and chromosome sorting in bread wheat (*Triticum aestivum* L.). Theor Appl Genet 2002;104: 1362–72.
Kubaláková M, Valárik M, Bartoš J, Vrána J, Číhalíková J, Molnár-Láng M, et al. Analysis and sorting of rye (*Secale cereale* L.) chromosomes using flow cytometry. Genome 2003;46:893–905.
Kubaláková M, Kovářová P, Suchánková P, Číhalíková J, Bartoš J, Lucretti S, et al. Chromosome sorting in tetraploid wheat and its potential for genome analysis. Genetics 2005;170:823–9.
Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol 2012;30:771–6.
Langlois RG, Yu LC, Gray JW, Carrano AV. Quantitative karyotyping of human chromosomes by dual beam flow cytometry. Proc Natl Acad Sci U S A 1982;79:7876–80.
Lee JH, Arumuganathan K. Metaphase chromosome accumulation and flow karyotypes in rice (*Oryza sativa* L.) root tip meristem cells. Mol Cells 1999;9:436–9.
Lee JH, Arumuganathan K, Kaeppler SM, Papa CM, Kaeppler HF. Cell synchronization and isolation of metaphase chromosomes from maize (*Zea mays* L.) root tips for flow cytometric analysis and sorting. Genome 1996;39:697–703.
Lee JH, Arumuganathan K, Yen Y, Kaeppler S, Kaeppler H, Baenziger PS. Root tip cell cycle synchronization and metaphase-chromosome isolation suitable for flow sorting in common wheat (*Triticum aestivum* L.). Genome 1997;40:633–8.
Lee JH, Arumuganathan K, Chung YS, Kim KY, Chung WB, Bae KS, et al. Flow cytometric analysis and chromosome sorting of barley (*Hordeum vulgare* L.). Mol Cells 2000;10:619–25.
Lee JH, Arumuganathan K, Kaeppler SM, Park SW, Kim KY, Chung YS, et al. Variability of chromosomal DNA contents in maize (*Zea mays* L.) inbred and hybrid lines. Planta 2002;215:666–71.
Leitch AR, Schwarzacher T, Wang ML, Leitch IJ, Surlan-Momirovich G, Moore G, et al. Molecular cytogenetic analysis of repeated sequences in a long term wheat suspension culture. Plant Cell Tissue Organ Cult 1993;33:287–96.
Li IJ, Arumuganathan K, Rines HW, Phillips RL, Riera-Lizarazu O, Sandhu D, et al. Flow cytometric sorting of maize chromosome 9 from an oat-maize chromosome addition line. Theor Appl Genet 2001;102:658–63.
Li L, Arumuganathan K, Gill KS, Song Y. Flow sorting and microcloning of maize chromosome 1. Hereditas 2004;141:55–60.
Long Q, Rabanal FA, Meng D, Huber CD, Farlow A, Platzer A, et al. Massive genomic variation and strong selection in *Arabidopsis thaliana* lines from Sweden. Nat Genet 2013;45:884–90.
Lucas SJ, Šimková H, Šafář J, Jurman I, Cattonaro F, Vautrin S, et al. Functional features of a single chromosome arm in wheat (1AL) determined from its structure. Funct Integr Genomics 2012;12:173–82.
Lucas SJ, Akpınar BA, Kantar M, Weinstein Z, Aydınoglu F, Šafář J, et al. Physical mapping integrated with syntenic analysis to characterize the gene space of the long arm of wheat chromosome 1A. PLoS One 2013;8:e59542.
Lucretti S, Doležel J. Bivariate flow karyotyping in broad bean (*Vicia faba*). Cytometry 1997;28:236–42.
Lucretti S, Doležel J, Schubert I, Fuchs J. Flow karyotyping and sorting of *Vicia faba* chromosomes. Theor Appl Genet 1993;85:665–72.
Lukaszewski AJ. Frequency of 1RS. 1AL and 1RS. 1BL translocations in United States wheats. Crop Sci 1990;30:1151–3.
Lysák MA, Číhalíková J, Kubaláková M, Šimková H, Künzel G, Doležel J. Flow karyotyping and sorting of mitotic chromosomes of barley (*Hordeum vulgare* L.). Chromosome Res 1999;7:431–44.
Ma YZ, Lee JH, Li LC, Uchiyama S, Ohmido N, Fukui K. Fluorescent labeling of plant chromosomes in suspension by FISH. Genes Genet Syst 2005;80:35–9.
Macas J, Doležel J, Lucretti S, Pich U, Meister A, Fuchs J, et al. Localization of seed protein genes on flow-sorted field bean chromosomes. Chromosome Res 1993;1:107–15.
Macas J, Doležel J, Gualberti G, Pich U, Schubert I, Lucretti S. Primer-induced labelling of pea and field bean chromosomes in situ and in suspension. Biotechniques 1995;19:402–8.
Macas J, Gualberti G, Nouzová M, Samec P, Lucretti S, Doležel J. Construction of chromosome-specific DNA libraries covering the whole genome of field bean (*Vicia faba* L.). Chromosome Res 1996;4:531–9.
Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods 2010;7: 111–8.

Martis MM, Klemme S, Banaei-Moghaddam AM, Blattner FR, Macas J, Schmutzer T, et al. Selfish supernumerary chromosome reveals its origin as a mosaic of host genome and organellar sequences. Proc Natl Acad Sci U S A 2012;109:13343–6.
Martis MM, Zhou R, Haseneyer G, Schmutzer T, Vrána Jan, Kubaláková M, et al. Reticulate Evolution of the Rye Genome. Plant Cell 2013;25:3685–98.
Marx V. The genome jigsaw. Nature 2013;501:263–8.
Matsumoto T, Wu JZ, Kanamori H, et al. The map-based sequence of the rice genome. Nature 2005;436:793–800.
Matsunaga S, Kawano S, Michimoto T, Higashiyama T, Nakao S, Sakai A, et al. Semi-automatic laser beam microdissection of the Y chromosome and analysis of Y chromosome DNA in a dioecious plant, *Silene latifolia*. Plant Cell Physiol 1999;40:60–8.
Matsunaga S, Isono E, Kejnovský E, Vyskot B, Doležel J, Kawano S, et al. Duplicative transfer of a MADS box gene to a plant Y chromosome. Mol Biol Evol 2003;20:1062–9.
Matsunaga S, Lebel-Hardenack S, Kejnovský E, Vyskot B, Grant SR, Kawano S. An anther- and petal-specific gene SlMF1 is a multicopy gene with homologous sequences on sex chromosomes. Genes Genet Syst 2005;80:395–401.
Mayer KFX, Taudien S, Martis M, Šimková H, Suchánková P, Gundlach H, et al. Gene content and virtual gene order of barley chromosome 1H. Plant Physiol 2009;151:496–505.
Mayer KFX, Martis M, Hedley PE, Šimková H, Liu H, Morris JA, et al. Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 2011;23:1249–63.
Meyers BC, Scalabrin S, Morgante M. Mapping and sequencing complex genomes: let's get physical. Nat Rev Genet 2004;5:578–88.
Michalak de Jimenez MK, Bassi FM, Ghavami F, Simons K, Dizon R, Seetan RI, et al. A radiation hybrid map of chromosome 1D reveals synteny conservation at a wheat speciation locus. Funct Integr Genomics 2013;13:19–32.
Molnár I, Kubaláková M, Šimková H, Cseh A, Molnár-Láng M, Doležel J. Chromosome isolation by flow sorting in Aegilops umbellulata and Ae. comosa and their allotetraploid hybrids Ae. biuncialis and Ae. geniculata. PLoS One 2011;6:e27708.
Muñoz-Amatriaín M, Moscou MJ, Bhat PR, Svensson JT, Bartoš J, Suchánková P, et al. An improved consensus linkage map of barley based on flow-sorted chromosomes and single nucleotide polymorphism markers. Plant Genome 2011;4:238–49.
Neumann P, Lysák M, Doležel J, Macas J. Isolation of chromosomes from *Pisum sativum* L. hairy root cultures and their analysis by flow cytometry. Plant Sci 1998;137:205–15.
Neumann P, Požárková D, Vrána J, Doležel J, Macas J. Chromosome sorting and PCR-based physical mapping in pea (*Pisum sativum* L.). Chromosome Res 2002;10:63–71.
Nie X, Li B, Wang L, Liu P, Biradar SS, Li T, et al. Development of chromosome-arm-specific microsatellite markers in *Triticum aestivum* (Poaceae) using NGS technology. Am J Bot 2012;99:369–71.
Nystedt B, Street NR, Wetterbom A, Zuccolo A, Lin YC, Scofield DG, et al. The Norway spruce genome sequence and conifer genome evolution. Nature 2013;497:579–84.
Paterson AH, Bowers JE, Bruggmann R, Dubchak I, Grimwood J, Gundlach H, et al. The *Sorghum bicolor* genome and the diversification of grasses. Nature 2009;457:551–6.
Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, et al. A physical map of the 1-gigabase bread wheat chromosome 3B. Science 2008;322:101–4.
Philippe R, Paux E, Bertin I, Sourdille P, Choulet F, Laugier C, et al. A high density physical map of chromosome 1BL supports evolutionary studies, map-based cloning and sequencing in wheat. Genome Biol 2013;14:R64.
Pich U, Meister A, Macas J, Doležel J, Lucretti S, Schubert I. Primed in situ labelling facilitates flow sorting of similar sized chromosomes. Plant J 1995;7:1039–44.
Peterson DG, Wessler SR, Paterson AH. Efficient capture of unique sequences from eukaryotic genomes. Trends Genet 2002;18:547–50.
Poczai P, Varga I, Laos M, Cseh A, Bell N, Valkonen JPT, et al. Advances in plant gene-targeted and functional markers: a review. Plant Methods 2013;9:6.
Požárková D, Koblížková A, Román B, Torres AM, Lucretti S, Lysák M, et al. Development and characterization of microsatellite markers from chromosome 1-specific DNA libraries of *Vicia faba*. Biol Plant 2002;45:337–45.
Qi LL, Echalier B, Chao S, Lazo GR, Butler GE, Anderson OD, et al. A chromosome bin map of 16,000 expressed sequence tag loci and distribution of genes among the three genomes of polyploid wheat. Genetics 2004;168:701–12.
Rabinowicz PD, McCombie WR, Martienssen RA. Gene enrichment in plant genomic shotgun libraries. Curr Opin Plant Biol 2003;6:150–6.
Rieseberg M, Kasper C, Reardon KF, Scheper T. Flow cytometry in biotechnology. Appl Microbiol Biotechnol 2001;56:350–60.
Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. Genome Biol 2013;14:405.
Román B, Satovic Z, Požárková D, Macas J, Doležel J, Cubero JI, et al. Development of a composite map in *Vicia faba*, breeding applications and future prospects. Theor Appl Genet 2004;108:1079–88.
Rustenholz C, Choulet F, Laugier C, Šafář J, Šimková H, Doležel J, et al. A 3000-loci transcription map of chromosome 3B unravels the structural and functional features of gene islands in hexaploid wheat. Plant Physiol 2011;157:1596–608.
Šafář J, Bartoš J, Janda J, Bellec A, Kubaláková M, Valárik M, et al. Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. Plant J 2004;39:960–8.
Šafář J, Šimková H, Kubaláková M, Číhalíková J, Suchánková P, Bartoš J, et al. Development of chromosome-specific BAC resources for genomics of bread wheat. Cytogenet Genome Res 2010;129:211–23.
Sato K, Motoi Y, Yamaji N, Yoshida H. 454 sequencing of pooled BAC clones on chromosome 3H of barley. BMC Genomics 2011;12:246.
Schnable PS, Ware Doreen, Fulton Robert S, Stein Joshua C, Wei Fusheng, Pasternak Shiran, et al. The B73 maize genome: complexity, diversity, and dynamics. Science 2009;326:1112–5.
Schmidt T, Heslop-Harrison JS. Genomes, genes and junk: the large-scale organization of plant chromosomes. Trends Plant Sci 1998;3:195–9.
Schondelmaier J, Martin R, Jahoor A, Houben A, Graner A, Koop HU, et al. Microdissection and microcloning of the barley (*Hordeum vulgare* L.) chromosome 1HS. Theor Appl Genet 1993;86:629–36.
Schubert I, Fransz PF, Fuchs J, de Jong JH. Chromosome painting in plants. Methods Cell Sci 2001;23:57–69.
Schwarzacher T, Wang ML, Leitch AR, Miller N, Moore G, Heslop-Harrison JS. Flow cytometric analysis of the chromosomes and stability of a wheat cell-culture line. Theor Appl Genet 1997;94:91–7.
Sears ER, Sears LMS. The telocentric chromosomes of common wheat. In: Ramanujams S, editor. Proc. 5th Int. Wheat Genetics Symp. New Delhi: Indian Agricultural Research Institute; 1978. p. 389–407.
Sehgal SK, Li W, Rabinowicz PD, Chan A, Šimková H, Doležel J, et al. Chromosome arm-specific BAC end sequences permit comparative analysis of homoeologous chromosomes and genomes of polyploid wheat. BMC Plant Biol 2012;12:64.
Shangguan L, Han J, Kayesh E, Sun X, Zhang C, Pervaiz T, et al. Evaluation of genome sequencing quality in selected plant species using expressed sequence tags. PLoS One 2013;8:e69890.
Shapiro HM. Practical Flow Cytometry. 4th ed. Inc.: John Wiley & Sons; 2003
Shatalina M, Wicker T, Buchmann JP, Oberhaensli S, Šimková H, Doležel J, et al. Genotype-specific SNP map based on whole chromosome 3B sequence information from wheat cultivars Arina and Forno. Plant Biotechnol J 2013;11:23–32.
Shizuya H, Birren B, Kim UJ, Mancino V, Slepak T, Tachiiri Y, et al. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. Proc Natl Acad Sci U S A 1992;89:8794–7.
Sierro N, Battey JND, Ouadi S, Bovet L, Goepfert S, Bakaher N, et al. Reference genomes and transcriptomes of *Nicotiana sylvestris* and *Nicotiana tomentosiformis*. Genome Biol 2013;14:R60.
Šimková H, Číhalíková J, Vrána J, Lysák MA, Doležel J. Preparation of HMW DNA from plant nuclei and chromosomes isolated from root tips. Biol Plant 2003;46:369–73.
Šimková H, Svensson JT, Condamine P, Hřibová E, Suchánková P, Bhat PR, et al. Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. BMC Genomics 2008a;9:294.
Šimková H, Šafář J, Suchánková P, Kovářová P, Bartoš J, Kubaláková M, et al. A novel resource for genomics of Triticeae: BAC library specific for the short arm of rye (*Secale cereale* L.) chromosome 1R (1RS). BMC Genomics 2008b;9:237.
Šimková H, Šafář J, Kubaláková M, Suchánková P, Číhalíková J, Robert-Quatre H, et al. BAC libraries from wheat chromosome 7D: efficient tool for positional cloning of aphid resistance genes. J Biomed Biotechnol 2011a;2011:302543.
Šimková H, Staňková H, Valárik M, Šafář J, Lapitan N, Luo M-Ch, et al. Chromosome genomics facilitates cloning of a Russian wheat aphid resistance gene. Abstracts of the 21st International Triticeae Mapping Initiative Workshop; 2011b. p. 17. [Mexico City].
Stein N, Ponelies N, Musket T, McMullen M, Weber G. Chromosome micro-dissection and region-specific libraries from pachytene chromosomes of maize (*Zea mays* L.). Plant J 1998;13:281–9.
Steuernagel B, Taudien S, Gundlach H, Seidel M, Ariyadasa R, Schulte D, et al. De novo 454 sequencing of barcoded BAC pools for comprehensive gene survey and genome analysis in the complex genome of barley. BMC Genomics 2009;10:547.
Stubblefield E, Oro J. The isolation of specific chicken macrochromosomes by zonal centrifugation and flow sorting. Cytometry 1982;5:273–81.
Suchánková P, Kubaláková M, Kovářová P, Bartoš J, Číhalíková J, Molnár-Láng M, et al. Dissection of the nuclear genome of barley by chromosome flow sorting. Theor Appl Genet 2006;113:651–9.
Tang X, De Boer JM, Van Eck HJ, Bachem C, Visser RG, De Jong H. Assignment of genetic linkage maps to diploid *Solanum tuberosum* pachytene chromosomes by BAC-FISH technology. Chromosome Res 2009;17:899–915.
The Arabidopsis Genome Initiative. Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. Nature 2000;408:796–815.
. [Internet]The IEB genomic resources database. The Centre of Plant Structural and Functional Genomics of the Institute of Experimental Botany, Olomouc, Czech Republic. [cited 2013 Dec 20. Available from]http://olomouc.ueb.cas.cz/dna-libraries/cereals.
. [Internet]The International Wheat Genome Sequencing Consortium (IWGSC). [cited 2013 Dec 20. Available from]http://www.wheatgenome.org/.
. [Internet]The million plant and animal genomes project. BGI. [cited 2013 Dec 20. Available from]http://www.genomics.cn/en/navigation/show_navigation?nid=5657.
Tuskan GA, DiFazio S, Jansson S, Bohlmann J, Grigoriev I, Hellsten U, et al. The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray). Science 2006;15:1596–604.
Überall I, Vrána J, Bartoš J, Šmerda J, Doležel J, Havel L. Isolation of chromosomes from *Picea abies* L. and their analysis by flow cytometry. Biol Plant 2004;48:199–203.
Valárik M, Bartoš J, Kovářová P, Kubaláková M, de Jong H, Doležel J. High-resolution FISH on super-stretched flow-sorted plant chromosomes. Plant J 2004;37:940–50.
Varshney RK, Song C, Saxena RK, Azam S, Yu S, Sharpe AG, et al. Draft genome sequence of kabuli chickpea (*Cicer arietinum*): genetic structure and breeding constraints for crop improvement. Nat Biotechnol 2013;31:240–6.
Veuskens J, Marie D, Brown SC, Jacobs M, Negrutiu I. Flow sorting of the Y sex chromosome in the dioecious plant *Melandrium album*. Cytometry 1995;21:363–73.
Vitharana SN, Wilson GS. Fractionation of chromosome 15 with an affinity-based approach using magnetic beads. Genomics 2006;87:158–64.
Vitulo N, Albiero A, Forcato C, Campagna D, Dal Pero F, Bagnaresi P, et al. First survey of the wheat chromosome 5A composition through a next generation sequencing approach. PLoS One 2011;6:e26421.
Vláčilová K, Ohri D, Vrána J, Číhalíková J, Kubaláková M, Kahl G, et al. Development of flow cytogenetics and physical genome mapping in chickpea (*Cicer arietinum* L.). Chromosome Res 2002;10:695–706.
Vrána J, Kubaláková M, Šimková H, Číhalíková J, Lysák MA, Doležel J. Flow-sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). Genetics 2000;156:2033–41.

Vrána J, Šimková H, Kubaláková M, Číhalíková J, Doležel J. Flow cytometric chromosome sorting in plants: the next generation. Methods 2012;57:331–7.
Vrána J, Kubaláková M, Číhalíková J, Valárik M, Doležel J. Preparation of subgenomic fractions enriched for particular chromosomes in plants. Cytometry 2014. submitted for publication.
Wang ML, Leitch AR, Schwarzacher T, Heslop-Harrison JS, Moore G. Construction of a chromosome-enriched HpaII library from flow-sorted wheat chromosomes. Nucleic Acids Res 1992;20:1897–901.
Wenzl P, Suchánková P, Carling J, Šimková H, Huttner E, Kubaláková M, et al. Isolated chromosomes as a new and efficient source of DArT markers for the saturation of genetic maps. Theor Appl Genet 2010;121:465–74.
Wicker T, Mayer KFX, Gundlach H, Martis M, Steuernagel B, Scholz U, et al. Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. Plant Cell 2011;23:1706–18.
Yang H, Chen X, Wong WH. Completely phased genome sequencing through chromosome sorting. Proc Natl Acad Sci U S A 2011;108:12–7.
Yu J, Hu S, Wang J, Wong KGS, Li S, Liu B, et al. A draft sequence of the rice genome (*Oryza sativa* L. ssp. *indica*). Science 2002;296:79–92.
Zatloukalová P, Hřibová E, Kubaláková M, Suchánková P, Šimková H, Adoración C, et al. Integration of genetic and physical maps of the chickpea (*Cicer arietinum* L.) genome using flow-sorted chromosomes. Chromosome Res 2011;19:729–39.

APPENDIX III

Molecular organization and comparative analysis of chromosome 3B of the wild wheat ancestor *Triticum dicoccoides*

Akpinar BA, Yuce M, Lucas S, Vrána J, Burešová V, Doležel J, Budak H

Scientific reports, 5:10763, 2015

IF: 5.228

SCIENTIFIC REPORTS

OPEN

Molecular organization and comparative analysis of chromosome 5B of the wild wheat ancestor *Triticum dicoccoides*

Received: 30 January 2015
Accepted: 28 April 2015
Published: 18 June 2015

Bala Ani Akpinar[1], Meral Yuce[1], Stuart Lucas[1], Jan Vrána[2], Veronika Burešová[2], Jaroslav Doležel[2] & Hikmet Budak[1,3]

Wild emmer wheat, *Triticum turgidum* ssp. *dicoccoides* is the wild relative of *Triticum turgidum*, the progenitor of durum and bread wheat, and maintains a rich allelic diversity among its wild populations. The lack of adequate genetic and genomic resources, however, restricts its exploitation in wheat improvement. Here, we report next-generation sequencing of the flow-sorted chromosome 5B of *T. dicoccoides* to shed light into its genome structure, function and organization by exploring the repetitive elements, protein-encoding genes and putative microRNA and tRNA coding sequences. Comparative analyses with its counterparts in modern and wild wheats suggest clues into the B-genome evolution. Syntenic relationships of chromosome 5B with the model grasses can facilitate further efforts for fine-mapping of traits of interest. Mapping of 5B sequences onto the root transcriptomes of two additional *T. dicoccoides* genotypes, with contrasting drought tolerances, revealed several thousands of single nucleotide polymorphisms, of which 584 shared polymorphisms on 228 transcripts were specific to the drought-tolerant genotype. To our knowledge, this study presents the largest genomics resource currently available for *T. dicoccoides*, which, we believe, will encourage the exploitation of its genetic and genomic potential for wheat improvement to meet the increasing demand to feed the world.

With an annual global production of more than 700 million tons across over 200 million hectares, wheat is the most widely grown crop worldwide (http://faostat.fao.org/). While the allohexaploid bread wheat (*Triticum aestivum*, 2n = 6x = 42, AABBDD genome) and the allotetraploid durum wheat (*Triticum turgidum* ssp. *durum*, 2n = 4x = 28, AABB genome) account for almost all global production, wild diploid and tetraploid wheat species and their relatives are still grown around the Fertile Crescent where they originated. For decades, wild species have been attractive sources of genetic diversity to be introduced into the narrow gene pool of modern cultivated wheats[1]. Introgression of genes and alleles from wild relatives is gaining increasing attention due to the urgent need to increase global wheat production[2,3].

Recent research indicates that the evolution of bread wheat involved three hybridization events[4]. Following the divergence of the *Triticum* and *Aegilops* lineages from a common ancestor ~6.5 million years ago, the first of these events is thought to involve A and B genome lineages which eventually gave rise to the diploid wheat D-genome progenitor, *Aegilops tauschii* (2n = 2x = 14, DD genome). The second event is dated back to a few hundred thousand years ago and resulted in the formation of the tetraploid AABB genome of *Triticum turgidum*, through the hybridization between *Triticum urartu* (A genome)

[1]Sabanci University Nanotechnology Research and Application Centre (SUNUM), Sabanci University, Orhanlı, 34956 Tuzla, Istanbul, Turkey. [2]Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Slechtitelu 31, CZ-78371 Olomouc, Czech Republic. [3]Faculty of Engineering and Natural Sciences, Molecular Biology, Genetics and Bioengineering, Sabanci University, Orhanlı, 34956 Tuzla, Istanbul, Turkey. Correspondence and requests for materials should be addressed to H.B. (email: budak@sabanciuniv.edu)

and a close relative of *Aegilops speltoides* (B genome), followed by whole genome duplication, probably via the production of non-reduced gametes[5]. Although several *T. turgidum* subspecies cultivated for thousands of years have lost their importance along the agricultural history, durum wheat, *T. turgidum* ssp. *durum* remains an important crop[1]. Finally, the most recent hybridization, dating back to ~10.000 years, involved *T. turgidum* and *Ae. tauschii*, and resulted in the emergence of the allohexaploid bread wheat, *Triticum aestivum*[1,4]. Domestication and, more recently, intensive breeding programs for better agricultural gain have considerably depleted the genetic diversity in today's elite cultivars. Fortunately, this diversity is still maintained in wild populations, which are adapted to a range of environmental conditions[2,6].

Wild emmer wheat, *Triticum turgidum* ssp. *dicoccoides* (2n = 4x = 28, AABB genome), is the wild relative of durum wheat and is capable of producing fertile offspring with both tetraploid and hexaploid wheat cultivars[2,7]. Some wild emmer genotypes exhibit remarkable tolerance against drought, the major abiotic stress factor responsible for severe yield losses worldwide[6]. One such genotype, TR39477, exhibits a strong and consistent tolerance against shock and prolonged drought stress, sharply contrasting with another genotype, TTD-22, highly sensitive to drought[8,9]. Such genetic diversity found within the natural populations of wild emmer wheat might provide clues into the key players of the drought response which may be targeted for introgression into the elite cultivars[7,10].

Wild emmer wheat genotypes are also recognized for high grain micronutrient content, tolerance against herbicides and resistance genes against biotic stresses, particularly against powdery mildew[2,7]. While the great potential that *T. dicoccoides* holds for wheat improvement has been recognized for decades, this potential remains largely unexploited to date. The rich gene pool and direct ancestry of *T. dicoccoides* enable the transfer of beneficial traits into elite cultivars relatively easily; however, 'linkage drag', caused by the co-transfer of chromosomal segments with negative effects on crop performance, complicates the introgression of such traits. If possible at all, the elimination of these undesirable segments, thereby minimizing the linkage drag, may take years of back-crosses[1,2]. While extensive genetic and genomic resources can largely circumvent these difficulties through marker-assisted selection or transgenic approaches[2,6], such resources are currently very limited for *T. dicoccoides*.

Advances in chromosome genomics, in particular, flow-cytometric isolation of individual chromosomes or chromosome arms enabling the construction of chromosome-specific Bacterial Artificial Chromosome (BAC) libraries or shotgun sequencing of isolated chromosomes by Next Generation Sequencing (NGS) methods have been pivotal in wheat genomics research[11–13]. Recently, draft sequences of all 21 bread wheat chromosomes have been published[14]. This important advance followed the publication of the draft genome sequences of A and D genome progenitors, *T. urartu*[15] and *Ae. tauschii*[16], altogether providing valuable insights into the genome organization and evolution of wheat. These sequencing efforts are likely to extend into the wild relatives of wheat, not only to complement and further broaden the comparative evolutionary genomics studies, but also to explore and exploit these rich sources for the benefit of the humankind.

In this study, we report the flow-cytometric sorting and sequencing of chromosome 5B of *Triticum turgidum* ssp. *dicoccoides*, which is known to harbor genes encoding resistance against powdery mildew disease, as well as quantitative trait loci for grain protein and mineral content[7]. As the first genomics study carried out on wild emmer wheat, the large-scale sequence information on chromosome 5B should enable the development of molecular markers linked to beneficial traits and facilitate gene transfer to support bread and durum wheat improvement.

Results

Flow-sorting, sequencing and assembly of Tdic5B. Flow cytometric analysis of fluorescence of DAPI-stained chromosomes alone did not permit the discrimination of chromosome 5B from other chromosomes of wild emmer wheat, *Triticum dicoccoides* variety 26676. Thus, biparametric analysis of GAA microsatellite content and DAPI fluorescence intensity was employed. This approach enabled the discrimination of all wild emmer wheat chromosomes and permitted sorting of chromosome 5B (Fig. 1). Fluorescence *in situ* hybridization (FISH) with probes for GAA microsatellites and *Afa* repeat family indicated an average purity of 95.24% from three independent samples. As obtaining sufficient amounts of DNA for direct sequencing by flow-cytometry is prohibitively resource-intensive, three flow-sorted 5B chromosome fractions were amplified by three independent rounds of Multiple Displacement Amplification (MDA) that yielded a total of 12.56 μg of *T. dicoccoides* 5B chromosome (Tdic5B hereafter) DNA.

Three sequencing runs on GS FLX Titanium platform were performed on two Tdic5B libraries, giving a total of 1.57 Gb of good-quality sequence data (Table 1). Assuming that the size of Tdic5B is similar to its modern counterpart, the 840 Mbp-long *T. durum* 5B chromosome[17], the sequence data obtained in this study represent a coverage of 1.87x, with the probability of any given position being represented as least once in this dataset being 0.799.

Repetitive elements comprise a notable fraction of *Triticeae* genomes[18] and interfere with the accurate assembly of genomic sequences. Hence, reads identified as repetitive elements, together with the reads exhibiting significant similarities to ribosomal RNA and chloroplast/mitochondrial DNA, deemed to have derived from contaminants of sorted chromosome fractions, were excluded. The remaining sequence reads were assembled using gsAssembler tool of Newbler 2.6 software. This assembly, referred

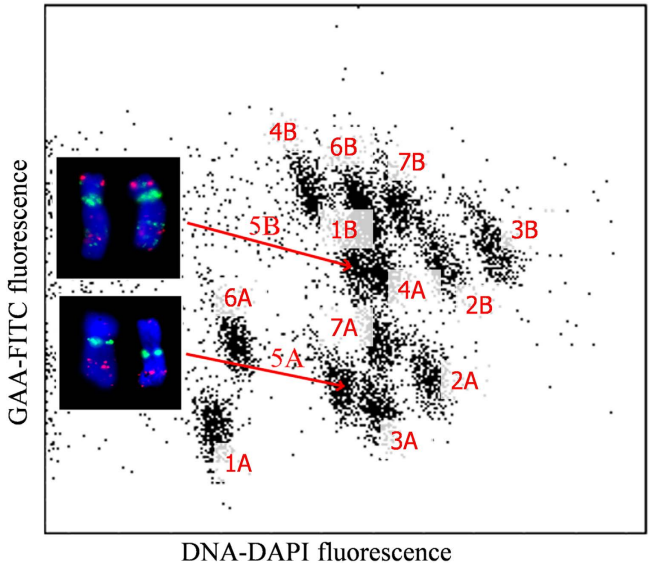


Figure 1. Biparametric flow karyotype of chromosomes isolated from T. dicoccoides. Prior to the analysis, GAA microsatellites were labeled by FITC and chromosomal DNA was stained by DAPI. FITC fluorescence was acquired at logarithmic scale, while DAPI fluorescence was measured at linear scale. This approach permitted separation from other chromosomes in the karyotype, including its homoeolog 5A. Insets: Images of flow-sorted chromosomes 5A and 5B. The chromosomes were identified after FISH with probes for GAA microsatellites (yellow-green) and for *Afa* family repeats (red).

Sequencing library	No.of reads N	Mean read length L (bp)	Total read length (Mb)	Sequencing Coverage[a]	Probability[b]
Tdic5B-1	953,680	294.8	281.2		
Tdic5B-2	1,694,938	357.2	605.4		
Tdic5B-2	1,640,921	419.2	687.9		
Combined	4,289,539	357.1	1,574.4	1.87	0.799

Assembly statistics	No. of reads/contigs	Mean length (bp)	Total length (Mb)	Length (% of chromosome)	N50 contig size (bp)
Filtered reads	501,177	357	100.8	12	
LCN assembly:					
Large contigs	14,302	1045	14.9	1.77	1117
All contigs	26,225	697	18.3	2.18	
Singletons	256,685	322	82.6	9.83	

Table 1. Sequencing and assembly metrics for Tdic5B. [a]Sequencing coverage was calculated using a chromosome size estimate of 840 Mbp[17]. [b]The probability of representation of any position in the dataset was calculated as follows: $P = [1 - (1 - L/S)^{N \times Purity}]$, where S is the chromosome size and L & N are as listed in the table.

to as Low-Copy Number (LCN) assembly hereafter, is comprised of 26,225 contigs and 256,685 singletons (Table 1), which is expected to cover majority of the non-repetitive regions of Tdic5B. The contigs of the LCN assembly had a peak depth of 2.1, close to the sequencing depth, indicating the accuracy of the contig construction. The cumulative length of the assembly was 100.9 Mb, shorter than the estimated 127 Mb non-repetitive loci, based on the repetitive fraction of the chromosome described below.

Repetitive content of Tdic5B. Repeat masking of Tdic5B sequences against known *Poaceae* repeat elements revealed that 84.9% of all Tdic5B sequences were repetitive, largely dominated by Long Terminal Repeat (LTR) retroelements (67.8% of all sequences). Within the LTR retroelements, Gypsy superfamily repeats had a marked abundance, accounting for over half of LTR elements, while the second most abundant Copia superfamily comprised 13% of all repeat elements (Fig. 2a). DNA transposons were mainly represented by En-Spm/CACTA repeats, which made up 17% of all repeats. Despite the predominance of

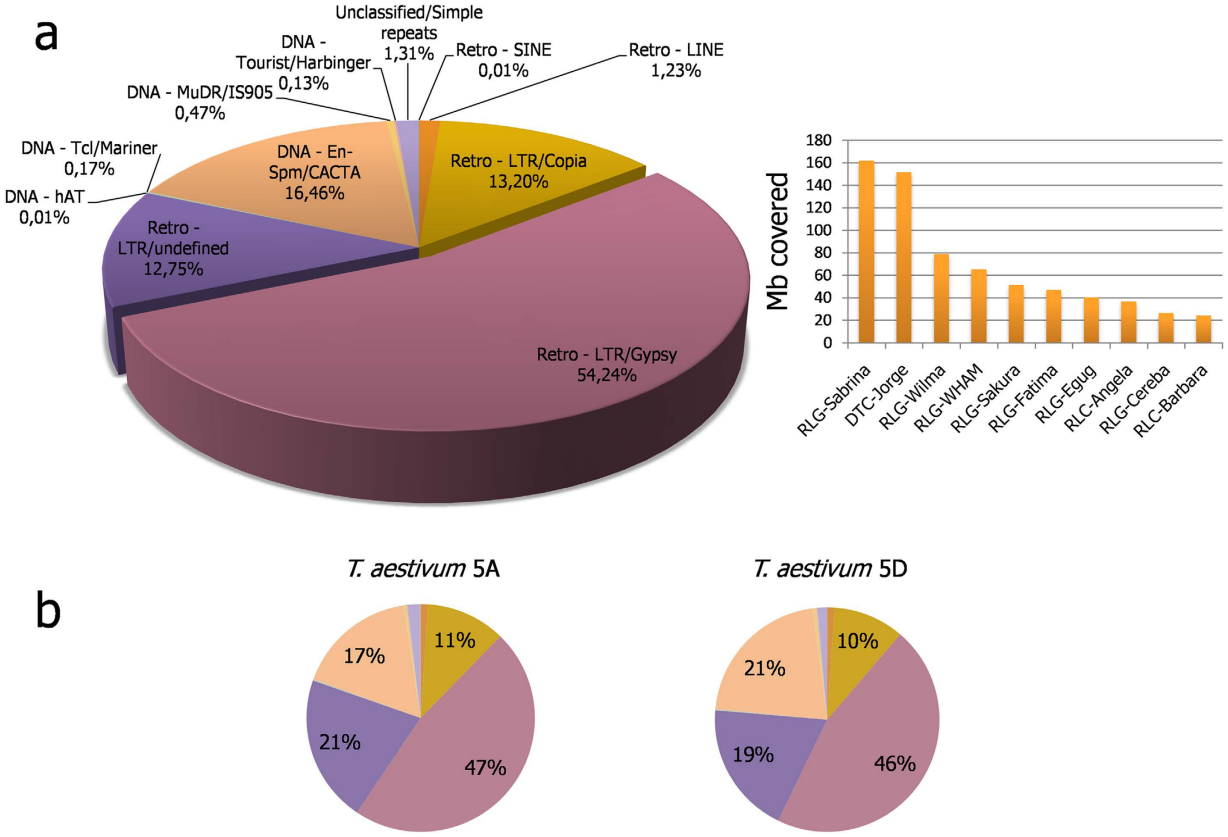


Figure 2. Repetitive element composition of Tdic5B. a. Repeat fractions by superfamily (left) and the cumulative sizes covered by the most abundant repeat families (right) of Tdic5B. **b.** Repeat fractions of *T. aestivum* 5A and 5D chromosomes by superfamily as in (**a**). DTC = DNA transposon, CACTA; RLG = retroelement, LTR, Gypsy; RLC = retroelement, LTR, Copia.

LTR retroelements among Tdic5B repeats, the DTC-Jorge family of En-Spm/CACTA superfamily DNA transposons had a notable coverage of the chromosome (Fig. 2a).

In order to compare the repeat content and distribution of Tdic5B with its modern and wild counterparts, raw sequences from the 5D chromosomes of *T. aestivum*[19] and *Ae. tauschii*[20] and the 5A chromosome of *T. aestivum*[13], obtained with the same NGS platform, were retrieved and masked against the same *Poaceae* repeat element database. The repeat content of Tdic5B was comparable to that of 5D chromosomes of *T. aestivum* and *Ae. tauschii* (82% and 81.1%, respectively)[19,20], while *T. aestivum* 5A chromosome contained fewer repetitive elements (72.8%)[13]. This is highly intriguing as the recently published chromosome-based draft sequences of bread wheat suggested repeat contents 5A > 5B > 5D[14]. However, the reference sequencing of chromosome 3B[21], the only bread wheat chromosome sequenced to this quality so far, reported a much higher repeat content than assessed by its draft sequence[14]. These inconsistencies may result from either different sequencing platforms being prone to different kinds of errors or amplification biases caused by MDA, both of which may have profound effects on the interpretation of low coverage NGS data. Therefore, a firm comparison of the repeat contents between group 5 chromosomes of the wheat ancestry may await reference sequencing of these chromosomes.

Despite the inconsistencies on the overall repeat content estimates, the chromosome-based draft genome sequence of bread wheat revealed higher abundance for class I retroelements for the A subgenome compared to B and D subgenomes (A > B > D), and an opposite trend for the class II DNA transposons (D > B > A)[14], in accordance with our observations for *T. aestivum* chromosomes 5A and 5D which were applied the same procedure as Tdic5B (Fig. 2b). As the undefined LTR elements, presumably representing older repeats, were the scarcest in the B genome, Mayer and his colleagues hypothesized that the modern B genome had undergone extensive transposon activity following polyploidization, giving rise to a higher retrotransposon content representing more recent proliferations[14]. In fact, this would be consistent with the repeat element distribution of Tdic5B, where undefined LTRs make up only 13% of all repeat annotations. It is tempting to speculate that, following tetraploidization, certain LTR families, in particular, those belonging to the Gypsy superfamily might have been proliferated in Tdic5B (Fig. 2a). Indeed, the repeat distribution of *Ae. tauschii* 5D chromosome suggests that the modern wheat D genome has undergone an expansion of the specific LTR retroelements coupled with the reduction of the relative contribution of DNA transposons compared to its progenitor[20].Since transposable elements

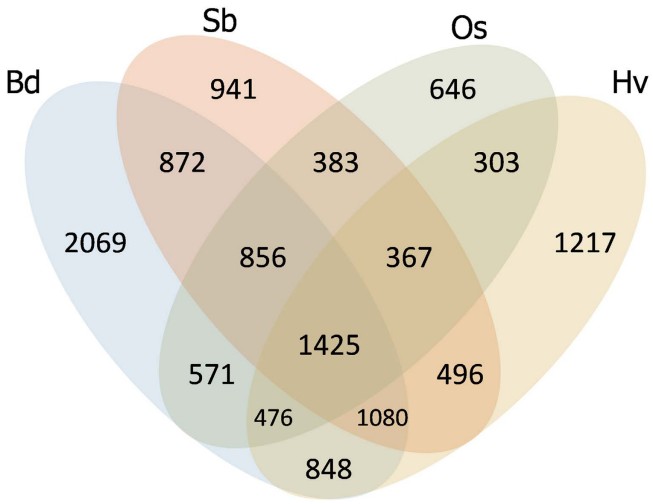


Figure 3. Venn diagram exhibiting Tdic5B sequence reads matching *Brachypodium* (Bd: *Brachypodium distachyon*), sorghum (Sb: *Sorghum bicolor*), rice (Os: *Oryza sativa*), and barley (Hv: *Hordeum vulgare*) proteins.

are known to have family-specific and species-specific evolutionary trajectories[22], which repeat families might have expanded in Tdic5B remains elusive at the time. All repeat annotations with regard to repeat families are given in Supplementary Table 1.

Gene content and conservation. To explore the gene content and conservation of Tdic5B, the LCN assembly was compared against the fully annotated proteomes of model grasses *Brachypodium distachyon*[23], rice[24] and sorghum[25], in addition to the high-confidence proteins of its close relative, barley[26], and wheat UniGene and UniProt sequences. A total of 19,669 sequences from the LCN assembly (5,635 contigs and 14,034 singletons) were deemed as gene-associated, as suggested by significant matches to related grass proteins and UniGene/UniProt sequences (Supplementary Table 2). Over half of these sequences, 3,161 contigs and 9,389 singletons, retrieved matches from at least one related grass proteome, indicating 'conserved' genes among grasses (Fig. 3). Among these, 2,555 contigs and 4,850 singletons were also supported by matches from wheat UniGene and UniProt sequences. A total of 1,425 sequences of the LCN assembly retrieved matches from all four proteomes, which possibly correspond to highly conserved genes, suggestive of central cellular processes, or, of a shared ancient origin (Fig. 3). Considering the fully annotated proteomes of model grasses, LCN assembly sequences matching *Brachypodium* proteins (8,197) outnumbered that of rice (5,027) and sorghum (6,420), as would be expected from the evolutionary distances, although the high number of matches with sorghum proteins is intriguing. In addition to these 'conserved' gene-associated sequences, 2,474 contigs and 4,645 singletons were found to have significant matches to only wheat UniGene or UniProt sequences indicating a collection of gene fragments, pseudogenes and a number of putatively *Triticum*-specific genes; for simplicity, these are collectively referred as 'non-conserved' gene-associated sequences. Due to the prevalence of pseudogenes in polyploid wheat genomes[27], several of these non-conserved gene-associated sequences are suspected to represent non-functional gene copies which might have undergone extensive rearrangements or accumulated too many mutations through the wheat genme evolution. To estimate the total genic content of Tdic5B and interpolating the estimate to the entire genome, *Brachypodium*, rice, sorghum and barley proteins exhibiting significant similarities to the LCN assembly were used as references onto which masked Tdic5B sequences were mapped. This approach merged non-overlapping sequences of the LCN assembly that matched the same query protein, and resulted in the construction of 4,818 'conserved gene models' for the Tdic5B (Supplementary File 1). Assuming an average coding sequence length of 2000 bases[13] and a chromosome length of 840 Mbp[17], the genic fraction (~9.63 Mb estimated coding length) of Tdic5B equals to 1.15%, similar to that of *Triticum aestivum* 5A (~1.23%)[13] and 5D (~1.15%)[19] chromosomes, but considerably lower than *Aegilops tauschii* 5D chromosome (2.1-2.9%)[20]. At a size of approximately 12 Gbp, this genic fraction corresponds to a total estimate of over 68,800 genes for the entire genome of *T. dicoccoides*. At the whole genome level, this estimate is consistent with both diploid wheat progenitors *Ae. tauschii*[16] and *T. urartu*[15], for which ~35,000 protein-coding loci were predicted, while considerably lower than the sum of high-confidence gene loci reported for the A and B genomes of *T. aestivum* (40,253 for the A genome and 44,523 for the B genome)[14]. The actual number of genes may be slightly higher than estimated for *T. dicoccoides*, as a fraction of the non-conserved gene associated sequences that did not match any of the four related grass proteomes likely represents genuine *Triticum*-specific genes. Additionally, the cumulative length of the LCN assembly being shorter than the coding length estimated by repeat annotations (100.9 Mb vs. 127 Mb) suggest that some paralogous loci might have

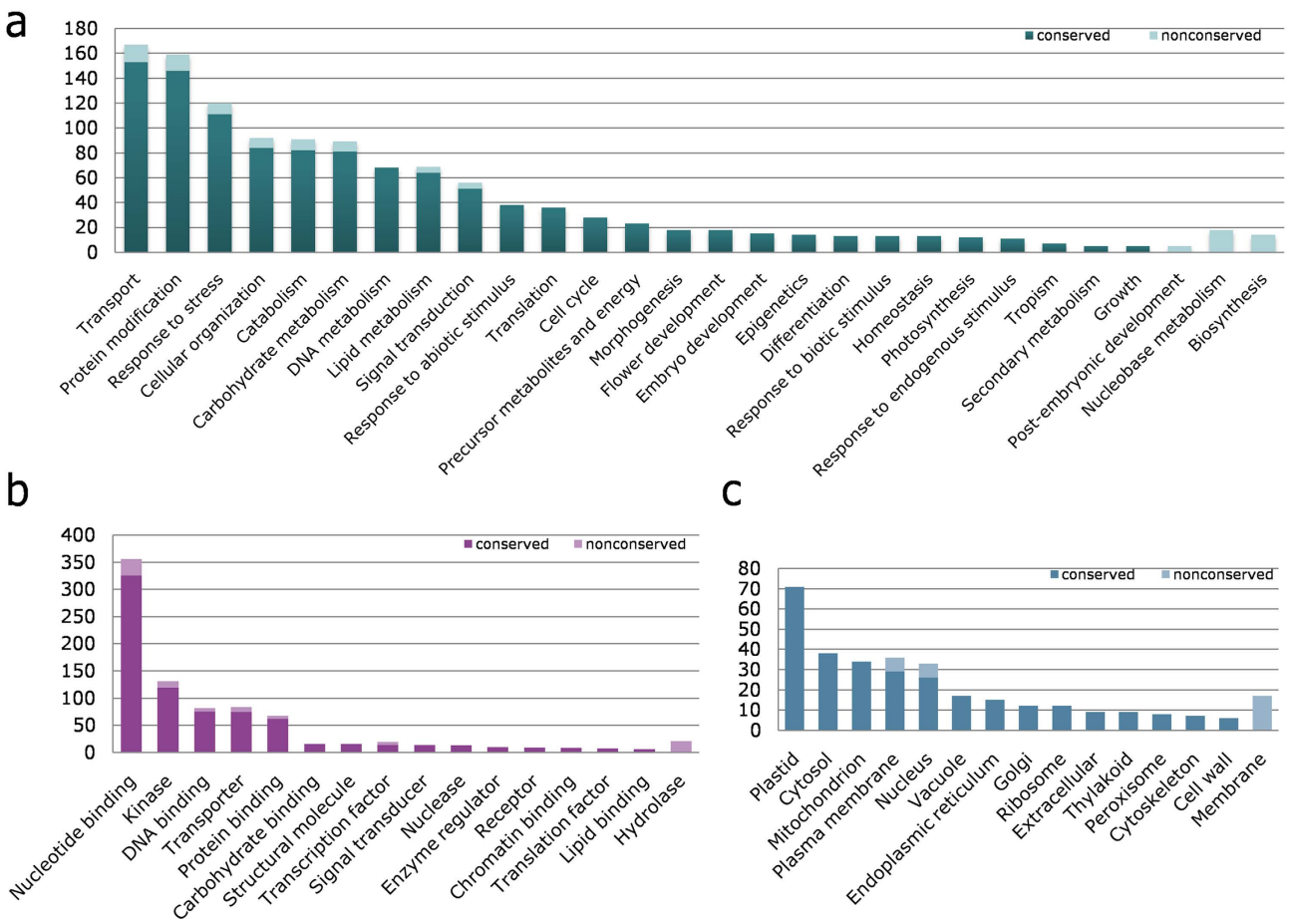


Figure 4. Gene-Ontology annotations of Tdic5B conserved and non-conserved genes in terms of, **a.** Biological Process, **b.** Molecular Function, **c.** Cellular Component.

been collapsed into single contigs in the LCN assembly, causing a slight underestimate of the coding fraction of the chromosome. All conserved gene models for Tdic5B are given in Supplementary File 1.

To gain insight into the functional gene space of Tdic5B, the LCN assembly contigs and singletons corresponding to the 7,612 putative conserved and 4,011 putative non-conserved gene associated sequences were annotated based on *Viridiplantae* proteins. Gene Ontology (GO) annotation of these sequences with regard to Biological Process (BP), Molecular Function (MF) and Cellular Component (CC) suggested a variety of GO terms (Fig. 4). Among BP terms, 'transport' and 'protein modification' processes were the most prominent, with a significant share of 'response to stress', for which wild progenitors are generally attributed (Fig. 4a). In terms of MF, 'nucleotide binding' and 'kinase' activities together, essential to all central pathways, accounted for more than half of all annotations (Fig. 4b). 'Transporter' function was also evident among MF terms, possibly in connection to the 'transport' process in BP terms. Although the LCN assembly was filtered against cpDNA and mtDNA sequences, 'plastid' terms alone took up almost a quarter of all CC annotations, (Fig. 4c). Similarly, mitochondrion-related sequences were also abundant among CC terms. Since energy or photosynthesis-related processes or functions were not among top terms for BP and MF, these abundances in CC terms were not expected. Interestingly, more than 72% of GO annotations related to either plastid or mitochondrion were observed to be hypothetical or predicted proteins, suggesting that the unusual abundance of these CC terms may be due to mis-annotations. Despite a number of leading terms in each classification, Tdic5B annotations revealed an array of processes, functions and components in general. This observation is, in fact, in accordance with the transcriptional autonomy of wheat sub-genomes[14], such that Tdic5B appears to encode a variety of genes capable to carrying out diverse functions.

Syntenic relationships. Conserved genes between Tdic5B and model grasses *Brachypodium*, rice and sorghum were observed to be organized into large-scale syntenic blocks on *Brachypodium* chromosomes 1 and 4 (Bd1 & Bd4), rice chromosomes 3, 9 and 12 (Os3, Os9 & Os12), and sorghum chromosomes 1 and 2 (Sb1 & Sb2) (Supplementary Fig. 1, 2). These syntenic blocks defined three groups of syntenic relationships between the model grass genomes, in accordance with the previous findings[23] (Supplementary Fig. 2, ribbons). The first syntenic group involved proximal ends of Bd1 and Sb1 and the

distal end of Os3 and, conversely, involved distal ends of Bd1 and Sb1 and the proximal end of Os3. The second syntenic group connected the distal ends of Bd4, Os9 and Sb2. Finally, the third group involved only *Brachypodium* and rice, in which the proximal end of the Bd4 was connected to the distal end of Os12. Syntenic genes conserved within these blocks are likely to be found in syntenic blocks along Tdic5B. As indicated by the dark red histograms in Supplementary Fig. 2, conserved genes of Tdic5B were usually found at the telomeric regions of model grass chromosomes, in accordance with the overall gene density trends along these chromosomes (light blue and light gray histograms flanking chromosomes for genes on '+' and '−' strands, respectively). Furthermore, these conserved genes were widely supported by barley homologues (Supplementary Fig. 2, light red histograms), implying that these are indeed functional genes.

Among the non-syntenic Tdic5B sequences (matching *Brachypodium*, rice or sorghum genes on non-orthologous chromosomes), 69 contigs and 206 singletons were found to match genes that were syntenic within *Brachypodium*, rice or sorghum genomes. Considering the evolutionary relationships between *Brachypodium*, rice or sorghum, a gene that is found on a non- colinear position in *Brachypodium*, but on colinear positions in rice and sorghum, is deemed as 'moved' (i.e. rearranged) specifically in the *Brachypodium* genome[28]. Similarly, non-syntenic Tdic5B sequences matching *Brachypodium*, rice and sorghum genes that are syntenic with each other indicate genes that are rearranged in the wheat lineage. Of such sequences (69 contigs and 206 singletons), 64 contigs and 191 singletons could be annotated based on *Viridiplantae* proteins, although 113 of these were hypothetical/predicted proteins (Supplementary Table 2). Intriguingly, 20 of these sequences did not match any known *Viridiplantae* proteins, a subset of which may actually correspond to pseudogenes or gene fragments that have lost their functionality through extensive rearrangements.

Putative tRNA and miRNA repertoire of Tdic5B. The analysis of Tdic5B sequences for putative tRNA genes revealed that the LCN assembly and the unmasked reads encode up to 78 and 875 tRNA genes, respectively, with a marked abundance for $tRNA^{Lys}$ species among unmasked reads (Supplementary Fig. 3a). This marked abundance was also reported for the unmasked low coverage sequences from *T. aestivum* 6B[29] and 5D[19] chromosomes, as well as *Ae. tauschii* 5D chromosome[20], and, is generally attributed to a Transposable Element (TE)-driven capture and subsequent co-proliferation. Targeted insertion of transposable elements into high copy small RNA genes have been observed previously, and, implicated as a potential tool for gene delivery[30]. Consistent with these observations, repetitive sequences predicted to contain putative $tRNA^{Lys}$ genes belonged almost exclusively to the LTR/Gypsy superfamily. Conversely, putative tRNA genes encoded by the non-repetitive LCN assembly were slightly less than that of *T. aestivum* and *Ae. tauschii* 5D chromosomes, as well as, much smaller orthologous *Brachypodium* chromosomes 1 & 4, indicating that tRNA genes are not likely expanded in *T. dicoccoides* (Supplementary Fig. 3b).

MicroRNAs (miRNAs) are an important subclass of small RNAs and carry out crucial functions in growth, development and stress responses by regulating gene expression[31]. The LCN assembly of Tdic5B identified 217 genomic loci for 64 miRNAs, based on sequence homology to known *Viridiplantae* miRNAs (miRBase, Release 21) and secondary structure preservation (Supplementary Table 3). The minimal folding free-energy index (MFEI) of miRNA precursors is generally higher than other types of RNAs, such as tRNAs (0.64), rRNAs (0.59), and mRNAs ($0.62-0.66$), and, thus, is utilized in computational miRNA prediction approaches[32]. Accordingly, MFEI values of miRNA precursors predicted from Tdic5B assembly were 0.95 ± 0.13. Among the predicted miRNAs, over half (54.8%) belonged to the miR2118 family. Three other miRNA families with well-established roles in plants, miR167, miR169 and miR399, were also prominent (10.1%, 6.9% and 7.8%, respectively) among miRNAs putatively encoded by Tdic5B. Computational prediction of miRNAs from the LCN assemblies constructed from raw 454 sequences of *T. aestivum* chromosomes 5A[13] and 5D[19], using the same procedure as Tdic5B, suggested that 9 miRNA families detected from Tdic5B are not present in these chromosomes (Fig. 5), although experimental validation is required for a firm conclusion.

To explore the functional networks regulated by the miRNAs predicted from Tdic5B sequences, miRNA-targeted genes were predicted from the transcriptome sequences assembled from RNA-Sequencing (RNA-Seq) of five wheat tissues (http://wheat-urgi.versailles.inra.fr/Seq-Repository/RNA-Seq)[14,33]. The wheat transcriptome assembly provided a comprehensive source for target genes, as reflected by the one-third of transcripts that could not be annotated based on known plant proteins (Supplementary Table 3). These, along with the hypothetical and predicted proteins, together comprising over two-thirds of all targets, suggest that our knowledge on miRNA-target interactions is going to evolve as more wheat miRNAs and proteins are annotated and characterized. Disease resistance-associated proteins alone comprised over 10% of all annotations, emphasizing the abundance of biotic stress related loci on Tdic5B. The remaining annotations revealed proteins involved in a variety of biological pathways; multiple targets regulated by the same miRNA, or, conversely, common targets of a number of different miRNAs point out to a complex and intermingled network of miRNA-regulated gene expression.

Single Nucleotide Polymorphisms on Tdic5B. Despite the rich allelic diversity maintained among wild wheat populations, saturated genetic maps to exploit this diversity are scarce. Therefore, Tdic5B sequences were mapped against the transcriptome sequences of two different *T. dicoccoides* varieties,

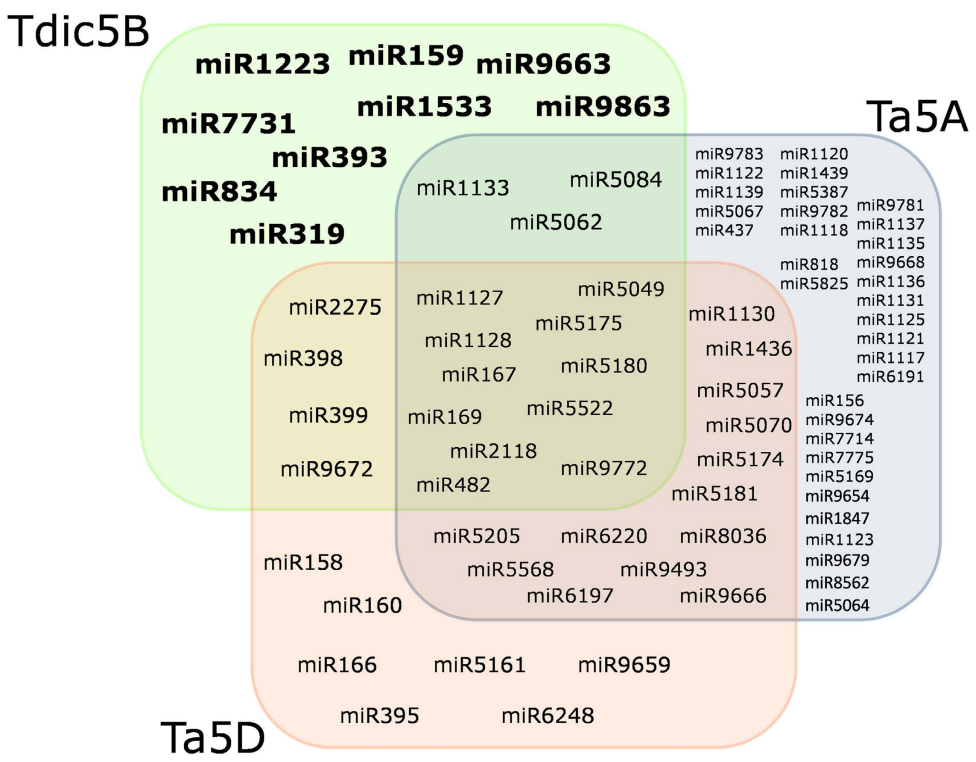


Figure 5. Predicted miRNA repertoires of *T. aestivum* 5A & 5D (Ta5A & Ta5D) and Tdic5B.

TR39477 and TTD-22, assembled from RNA-Seq data obtained recently (Budak *et al.*, in review), to reveal Single Nucleotide Polymorphisms (SNPs), following the pipeline proposed by You *et al.*[34]. In the absence of a reference-quality genome sequence, You *et al.*[34] recently suggested a methodology to discover potential SNPs, by mapping short reads generated by NGS technologies on relatively longer reads or sequence assemblies, such as full length cDNA sequences or transcriptome assemblies. Using this approach, RNA-Seq sequences from drought-treated and control root tissues of TR39477 and TTD-22 varieties were assembled to generate longer transcriptome sequences to be used as reference. To minimize false alignments with transcripts from elsewhere in the genome, 5B-related transcripts were retrieved by blast searches against Tdic5B sequences. Despite the high stringency used to filter out 5B related sequences, it should be noted that a small number of highly similar homoeologous sequences from Tidc5A chromosome or paralogous loci from elsewhere may not be excluded and remain among the filtered transcripts. Unmasked Tdic5B reads were then mapped onto these 5B-related transcripts and sequence variations within positive alignments were filtered against depth and SNP proximity[34]. Consequently, a total of 9275, 10034, 8913 and 9242 SNPs in 1827, 1879, 2064 and 2137 5B-related transcripts from drought-treated TR39477, control TR39477, drought-treated TTD-22, and control TTD-22 samples, were identified, respectively (Supplementary Table 4). These corresponded to the average SNP frequencies of 1,043.4 bases/SNP (1,047.8 for drought, 1,038.9 for control) for TR39477, and, 1,368.3 bases/SNP (1,370.6 for drought, 1,365.9 for control) for TTD-22 varieties, considering the total length of all respective 5B-related transcripts.

The two *T. dicoccoides* varieties used to discover potential SNPs exhibited contrasting levels of drought tolerance, consistent across different drought exposures. TR39477 is characterized by its high tolerance against drought, compared to highly sensitive TTD-22[8,9]. Transcripts from the drought-treated TR39477 roots were further examined, as SNPs within these transcripts may be utilized in breeding programmes if linked to drought stress tolerance. Of the 1,827 SNP-containing transcripts from the drought-treated TR39477 transcriptome, 507 exhibited significant similarities to transcripts from control TR39477, drought-treated TTD-22, and control TTD-22 transcriptomes. On these 507 transcripts, positions corresponding to SNPs identified in TR39477 samples were examined across other samples through pair-wise alignments and only those that are covered by transcripts from both control and drought-treated samples and that are consistent (having the same nucleotide) in control and drought-treated samples of the same variety were recorded. A total of 584 SNPs in 228 transcripts identified in TR39477 had the same nucleotide in TTD-22 transcripts as in Tdic5B sequences (for instance, C in TR39477 but T in TTD-22 and Tdic5B; "Group 1" in Supplementary Table 5). Conversely, 1,092 SNPs in 290 transcripts had the same nucleotide in TTD-22 and TR39477, but differed in Tdic5B ("Group 2" in Supplementary File 5). Interestingly, 3 SNPs on 3 transcripts identified in TR39477 had a different nucleotide in each of the three varieties. For instance, the transcript c23780_g2_i1 from TR39477 drought sample had the

base 'Thymine' at position 967 (as well as the corresponding transcript from TR39477 control sample). However, the corresponding position in corresponding transcripts from TTD-22 control and drought samples had 'Guanine' instead, while the Tdic5B sequences mapping to this position had 'Cytosine' ("Group 3" in Supplementary Table 5). As these transcripts can be readily differentiated based on SNPs in all three genotypes, phenotypic traits conferred by these transcripts can also be readily screened using linked molecular markers. However, functional annotations of these transcripts through the comparison against known *Viridiplantae* proteins revealed sequence similarities to only hypothetical proteins with currently unknown functions. Functional characterization of these transcripts and physiological characterization of *T. dicoccoides* 26676 variety used in this study, particularly against drought stress conditions, may provide candidate genes for wheat improvement, for which SNP-based molecular markers for gene cloning and transfer can then be designed and implemented in breeding programs.

Discussion

Domestication and breeding for modern agriculture have narrowed gene pools within crop populations for improved yield, rendering crops susceptible to stress factors. Wild germplasms adapted to a range of environments maintain a rich genetic diversity and are a promising source for crop breeding programmes. Wild emmer wheat, *Triticum turgidum* ssp. *dicoccoides* is the wild relative of the tetraploid durum wheat progenitor, *Triticum turgidum*. The potential that *T. dicoccoides* holds for wheat improvement has been recognized for almost a century; accordingly, a number of genes associated with abiotic and biotic stress tolerance, grain protein and micronutrient content have been mapped to several wild emmer chromosomes. A subset of these genes have also been introgressed into modern wheat cultivars[7]. A majority of genes introgressed from *T. dicoccoides* into modern cultivars comprised disease-resistance genes, particularly against powdery mildew and rust. Fine mapping and characterization of additional resistance genes, including powdery mildew resistance, continue as pathogen evolution necessitates the identification of novel alleles against novel pathogen strains[35–39]. A few loci controlling important agronomic traits, such as grain protein and micronutrient content have also been mapped to 5B chromosome *T. dicoccoides*[7,35,36,40]. In addition, *T. dicoccoides* exhibits allelic variation for the *Ph1* locus located on the long arm of 5B chromosome. This locus is responsible for the suppression of homoeologous chromosome pairing during meiosis, extending the utility of studying this chromosome beyond agronomically relevant traits[41–43].

Despite its rich genetic diversity and direct ancestry to durum and bread wheat, genomic resources are highly limited for *T. dicoccoides*, restricting its exploitation in wheat improvement. In this study, we present the next-generation sequencing of flow-sorted *T. dicoccoides* 5B chromosome to 1.87x coverage, enabling us to explore its repeat content and composition, conserved protein-coding and tRNA-encoding genes, miRNA repertoire and nucleotide variations with two related genotypes with contrasting levels of drought tolerance. To our knowledge, the sequence information generated in this study is currently the largest genomics resource available for *T. dicoccoides*, providing an in-depth view into its genome structure and organization.

Comparison of Tdic5B sequences against the known *Poaceae* repeats revealed that repetitive sequences make up 84.9% of the chromosome, consistent with the highly repetitive nature of *Triticeae* genomes. Recently, low-coverage 454 sequencing of *T. aestivum* 5B chromosome has been reported[44]. Despite representing only 7% of the chromosome (61 Mb of sequence data, thus, not included in the main comparative analyses), *T. aestivum* 5B sequences, which were applied the same repeat-masking procedure, suggested a repeat content of 83.7%, similar to Tdic5B (Supplementary Fig. 4). Repeat superfamily distribution of Tdic5B suggested recent amplification of the Gypsy superfamily, as suggested by the comparative analysis of the recent draft chromosome sequences of bread wheat[14]. Repetitive element distributions revealed from the limited 454 sequencing data, from the *T. aestivum* 5B chromosome support this view (Supplementary Fig. 4). Differential expansion of high-copy and low-copy elements following polyploidization and diploidization is a known phenomenon[45]; however, due to the highly dynamic profileration profiles of repetitive elements in different backgrounds, which family members of the Gypsy superfamily might have expanded in Tdic5B could not be determined with the present data on its counterparts. The genic fraction of Tdic5B (1.15%) assessed from a total of 4,818 conserved gene models was comparable to that of *T. aestivum* 5A and 5D chromosomes. Recently, over 5,500 functional gene or gene-models were reported for the reference sequence of the 3B chromosome including the unanchored scaffolds[21]. The estimated gene content is considerably lower for Tdic5B, which is largely attributable to the differences in chromosome sizes (1 Gb for 3B vs. estimated 840 Mb for Tdic5B), and to a lesser extent, can be explained by the exclusion of non-conserved gene associated sequences in Tdic5B gene content estimation. The reference sequence of 3B chromosome revealed ~27% pseudogenic loci among all identified coding loci[21]. As distinguishing pseudogenic loci from genuine genes at this level of coverage would be impractical, these non-conserved gene associated sequences were excluded from gene estimation. Accordingly, the actual gene content of Tdic5B is expected to slightly exceed 4,818 gene models constructed in this study. The LCN assembly of Tdic5B matched 7,612 conserved genes from model grasses, *Brachypodium*, rice and sorghum, and revealed 3 syntenic blocks, involving, (1) Bd1-Os3-Sb1, (2) Bd4-Os9-Sb2, (3) Bd4-Os12 chromosomes (Supplementary Fig. 2), consistent with the previous observations[23]. The presence of large syntenic blocks and colinearity within these blocks is crucial, especially for species with limited genetic mapping data. Indeed, fine mapping of a number of traits in *T. dicoccoides* relied heavily on the syntenic

relationships and colinearity[35,38,39,46]. In addition to protein-coding loci, Tdic5B was observed to contain slightly fewer putative tRNA genes and miRNAs, compared to its modern counterparts *T. aestivum* chromosomes 5A and 5D, for which raw sequence data obtained with the same NGS platform were retrieved and processed using the same procedures as Tdic5B. While NGS data for *T. aestivum* 5B chromosome is available from two sources[14,44], these could not be used for direct comparisons due to either limited data size[44] or different sequencing technology[14].

Homology-based miRNA prediction identified 64 unique miRNAs putatively encoded by Tdic5B. Among the predictions, miR2118 family were the most abundant, representing over half of the putative miRNA-coding genomic loci. Additionally, Tdic5B was found to encode 11 members of miR167 family, 10 members of miR169 family and 6 members of miR399 family. Remarkably, the precursors of miR2118 and miR169 have been experimentally verified to be specific to the 5D chromosome of modern bread wheat[47]. miR2118 family was also reported to be represented by 42 family members in *Ae. tauschii* draft genome[16]. It is tempting to speculate that the coding regions for miR2118 and miR169 on ancient B-genome might have been lost through wheat genome evolution due to functional redundancy in homoeologous genomes, while these regions are still retained in the B-genomes of tetraploid wild populations. miR169 has been identified as an abiotic stress-responsive miRNA family in plants, specifically targeting NF-YA subunit of Nuclear Transcription Factor Y (NF-Y)[48]. Consistently, target annotations of wheat transcriptome sequences identified several NF-Y subunits exclusively targeted by miR169 and miR2118 (Supplementary Table 3). miR2118 has also been implicated to target NBS-LRR disease resistance genes[49], as reflected in the target annotations of putative Tdic5B miR2118 family. Intriguingly, these observations indicate that several putative miR2118-targeted wheat transcriptome sequences assembled from RNA-Seq of five different wheat tissues[14,33] that lacked an annotation or annotated as hypothetical proteins may actually correspond to biotic or abiotic stress-related genes. The lack of an apparent sequence similarity to known *Viridiplantae* proteins implies that these transcripts may code for novel or highly diverged proteins and their further characterization may reveal new candidates for wheat improvement.

Mapping of Tdic5B reads onto 5B-related transcriptome sequences of two *T. dicoccoides* varieties, TR39477 and TTD-22, revealed one SNP in every 1,043.4 and 1,368.3 bases on average, respectively. It should be noted that, however, some of these SNPs may arise from highly similar homoeologous Tdic5A sequences or, to a lesser extent, highly similar paralogous loci elsewhere in the genome, which could not be differentiated from 5B-related transcripts computationally, despite the highly stringent filtering criteria. Recently, Brenchley and her colleagues could differentiate homoeologous sequences with high precision for 66% of gene assemblies obtained from 5X coverage sequences of the entire bread wheat genome[50]. Similarly, among approximately 30% of the transcriptome assemblies of TR39477 and TTD-22 that are probably highly similar, transcripts that differ by 2% of less by sequence composition on the homoeologous 5A chromosome are likely to be retained among the 5B-related transcripts used for SNP analyses. Thus, it is important to implement SNPs reported in this study cautiously for functional studies, until they are verified experimentally. The SNP frequencies observed in this study imply that coding regions carry more sequence divergence between 26676 and TR39477 genotypes, which may be utilized to design SNP-based markers, particularly for traits linked to the remarkable drought tolerance of the TR39477 genotype. The contrasting drought tolerances of TR39477 and TTD-22 potentiates the use of the SNPs for novel molecular marker design to aid in genetic and physical mapping of genomic drought-resistance loci. Through effective genotyping of wild populations these SNPs could be useful for gene discovery and mapping, as demonstrated by the SNP-based genome-wide association mapping of stripe rust resistance reported recently[51]. NGS mediated discovery of SNPs was previously utilized for the fine mapping of a grain protein content locus in durum wheat[52]. Although the SNP frequencies reported here are relatively lower than the study of Sela *et al.*[51], and, another study reporting SNP discovery via NGS in two *Ae. tauschii* accessions[34], the accumulation of high-throughput NGS data is likely to play pivotal role in gene discovery and mapping in wild emmer wheat that can further be implemented into wheat improvement.

Methods

Flow-sorting, sequencing and assembly of Tdic5B. Seeds of *Triticum dicoccoides* accession 26676 were kindly provided by Dr. Etienne Paux (INRA, France). The seeds were germinated and their primary roots used for preparation of aqueous suspensions of intact mitotic metaphase chromosomes[11]. GAA microsatellites of chromosomes in suspension were labeled by FITC[53], chromosomal DNA was stained by DAPI at 2 μg/ml and the samples were analyzed using FACSAria SORP (BD Biosciences, San José, USA) at rate of 1,500 chromosomes/sec. Blue laser (488 nm, 100 mW) was used to excite FITC fluorescence of GAA microsatellites, while UV laser (355 nm, 100 mW) was used for DAPI excitation. Biparametric flow karyotypes of FITC fluorescence (logarithmic scale) and DAPI fluorescence (linear scale) were obtained after analyzing 20,000 chromosomes. In order to sort chromosome 5B, sort window was set up on the dot plot and the chromosome was sorted at rate of 20 chromosomes/sec. In order to assess contamination of the sorted fraction by other chromosomes, 2,000 chromosomes were sorted into a drop of P5 buffer[54]and air-dried. FISH with probes for GAA microsatellites and *Afa* repeat family was used to facilitate identification of chromosomes, which were counterstained by DAPI and observed by fluorescence microscopy. Three independent samples were prepared and average purity of sorted fraction was determined. To

produce the required amounts of chromosomal DNA for sequencing, 30,000 chromosomes (equivalent to 50 ng DNA) were sorted into PCR tube filled with 40 μl deionized water in three batches, and their DNA was amplified by isothermal multiple displacement amplification (MDA)[55].

Sequencing Tdic5B DNA was carried out on GS FLX Titanium platform (Roche 454 Life Sciences, Branford, CT, USA), following manufacturer's instructions. Two shotgun libraries were prepared from 0.5 μg of amplified Tdic5B and sequenced in three rounds. Raw reads are submitted to the EBI Sequence Read Archive under the primary accession number PRJEB8079.

All sequence reads were compared against MIPS Repeat Element Database v9.3 p for *Poaceae* (ftp://ftp-mips.helmholtz-muenchen.de/plants/REdat/)[56], using RepeatMasker v.3.3.0 software (http://www.repeatmasker.org/) to identify repetitive elements. Organellar genome and rRNA associated reads were identified through BLAST searches against *Triticum turgidum* ssp. *dicoccoides* TA0073 (GenBank: KJ614400.1), TA0060 (GenBank: KJ614401.1), TA1133 (GenBank: KJ614402.1) chloroplast, complete genome (1E-15, -dust "no"); *Triticum aestivum* mitochondrion, complete genome (NC_007579.1, 1E-15, -dust "no"); all *Triticum* rRNA sequences (419 sequences on 08.09.14) deposited in NCBI Nucleotide database (1E-05, -dust "no"). Sequence reads identified as repetitive or organellar genome/rRNA-associated were excluded from the sequence assembly. The remaining sequences were used to construct a Low Copy-Number (LCN) assembly using gsAssembler software (Newbler 2.6, Roche 454 Life Sciences, Branford, CT, USA) with the "Large and complex genome", "Heterozygotic genome", "Extend low-depth overlaps" options and a minimum overlap identity of 98%[20]. Sequencing and assembly metrics are given in Table 1.

For comparative analyses, raw sequences for *T. aestivum* 5A[13], 5B[44] and 5D[19] chromosomes, and, *Ae. tauschii* 5D[20] chromosome, all of which were obtained with GS FLX Titanium as Tdic5B, were retrieved, and the same procedures and criteria were applied using the same databases as Tdic5B.

Identification of protein-coding genes, putative tRNAs and miRNAs. Protein-coding gene-associated reads of the LCN assembly were identified using BLAST searches against the fully annotated *Brachypodium distachyon* (v1.2, http://mips.helmholtz-muenchen.de/plant/brachypodium)[23], *Oryza sativa (*assembly IRGSP-1.0, http://rapdb.dna.affrc.go.jp/download/irgsp1.html)[24], *Sorghum bicolor* (v1.4, http://mips.helmholtz-muenchen.de/plant/sorghum/)[25] proteins (1E-6, -length 30, -ppos 75); high-confidence *Hordeum vulgare* proteins (http://mips.helmholtz-muenchen.de/plant/barley/)[26] (1E-6, -length 30, -ppos 90); *Triticum aestivum* UniGenes (Build#63, ftp://ftp.ncbi.nih.gov/repository/UniGene/Triticum_aestivum/, 1E-30, -length 90, -pident 98) and *Triticum* UniProt sequences (14,4397 entries, http://www.uniprot.org/, (1E-6, -length 30, -ppos 100). The blast parameters were essentially adopted from previous studies to ensure consistency[13,19,20] and similarity/identity cutoffs were increased for the close relatives, barley and wheat species. To increase stringency, 'Best Reciprocal Hit' approach was applied for protein queries, where BLAST searches were performed as blastx and tblastn, and only reciprocal best hits were retained. For all BLAST searches, redundant LCN assembly singletons covering the exact same portion of a protein or gene query were eliminated to avoid amplification bias deriving from MDA. BLAST+ stand-alone toolkit, version 2.2.25[57] were used for all BLAST searches. Gene models were constructed by mapping masked Tdic5B reads onto the coding sequences of *Brachypodium*, rice, sorghum and barley proteins that exhibited significant similarities to the LCN assembly through BLAST searches. If an LCN contig or singleton is associated with multiple hits from the grass proteomes through BLAST searches the reference sequence is picked by this precedence: *Brachypodium*, rice, sorghum and barley. Mapping was performed using gsMapper software (Newbler 2.6, Roche 454 Life Sciences, Branford, CT, USA) with default settings, except for All Contig Threshold=40. Mapping results were processed with an in-house Perl script which merged non-overlapping sequences mapping to the different sections of the same reference sequence and filled the gaps (where no Tdic5B sequence was mapped) by strings of 'n'.

Circle plots and heatmaps demonstrating gene conservation and syntenic relationships were visualized using Circos software[58] and MATLAB R2010b, respectively. Ribbons in Circos image were generated with >100 members along 1 Mb intervals. Gene densities were counted on 500 kb intervals (light blue & light grey). Heatmaps were drawn with a sliding window approach of 50 kb step size and the genomic positions of annotated proteins were retrieved from MIPS database of plants (http://mips.helmholtz-muenchen.de/plant/genomes.jsp). All functional annotations were performed on BLAST2GO[59] using locally run BLAST results against *Viridiplantae* proteins (1E-6, -outfmt 5, -max_target_seq 1).

The tRNAscan-SE 1.21 program[60] was run locally with the default parameters for eukaryotic genomes to predict putative tRNA genes. Pseudogenes and other undetermined annotations were not evaluated.

Prediction of putative miRNAs was performed using two in-house Perl scripts, SUmirFind and SUmirFold. Mature miRNA sequences for *Viridiplantae* were retrieved from miRBase Release 21 (http://mirbase.org/) and used as query for homology searches. Hairpin structures were evaluated for miRNA characteristics as previously reported[47]. Potential miRNA targets were predicted online using psRNA-Target (http://plantgrn.noble.org/psRNATarget/) among transcriptome assemblies from RNA-Seq data of five *T. aestivum* tissues (http://wheat-urgi.versailles.inra.fr/Seq-Repository/RNA-Seq)[14,33].

Discovery of Single Nucleotide Polymorphisms. Single Nucleotide Polymorphisms (SNPs) were investigated essentially following You *et al.*[34]. RNA-Sequencing data from drought-treated and control roots of *T. dicoccoides* varieties TR39477 and TTD-22 were assembled using Trinity pipeline (http://

trinityrnaseq.sourceforge.net/). The assembled transcriptome sequences were blasted against Tdic5B reads to identify 5B-related transcripts (1E-30, -pident 98). The 5B-related transcripts sequences were then separately used as reference onto which Tdic5B unmasked reads were mapped using gsMapper software (Newbler 2.6, Roche 454 Life Sciences, Branford, CT, USA) with default settings. Nucleotide variations on single positions were retained and filtered for mapped read depth ($3 \leq \text{depth} \leq 10$) and SNP proximity ($>3$ bp between SNPs). To identify shared SNP positions, drought-treated TR39477 transcripts were blasted against remaining three sets of transcriptome sequences (1E-30, -pident 98) and positions corresponding to SNPs in drought-treated TR39477 transcripts were manually evaluated through pair-wise sequence alignments on NCBI Blast (http://blast.ncbi.nlm.nih.gov/BlastAlign.cgi).

References

1. Feuillet, C., Langridge, P. & Waugh, R. Cereal breeding takes a walk on the wild side. *Trends Genet.* **24,** 24–32 (2008).
2. Nevo, E. & Chen, G. Drought and salt tolerances in wild relatives for wheat and barley improvement. *Plant. Cell Environ.* **33,** 670–85 (2010).
3. Paux, E., Sourdille, P., Mackay, I. & Feuillet, C. Sequence-based marker development in wheat: advances and applications to breeding. *Biotechnol. Adv.* **30,** 1071–88 (2012).
4. Marcussen, T. *et al.* Ancient hybridizations among the ancestral genomes of bread wheat. *Science* **345,** 1250092–1250092 (2014).
5. Hao, M. *et al.* QTug.sau-3B is a major quantitative trait locus for wheat hexaploidization. *G3 (Bethesda).* **4,** 1943–53 (2014).
6. Budak, H., Kantar, M. & Kurtoglu, K. Y. Drought tolerance in modern and wild wheat. *ScientificWorld Journal.* **2013,** 548246 (2013).
7. Xie, W. & Nevo, E. Wild emmer: genetic resources, gene mapping and potential for wheat improvement. *Euphytica* **164,** 603–614 (2008).
8. Ergen, N. Z. & Budak, H. Sequencing over 13 000 expressed sequence tags from six subtractive cDNA libraries of wild and modern wheats following slow drought stress. *Plant. Cell Environ.* **32,** 220–36 (2009).
9. Ergen, N. Z., Thimmapuram, J., Bohnert, H. J. & Budak, H. Transcriptome pathways unique to dehydration tolerant relatives of modern wheat. *Funct. Integr. Genomics* **9,** 377–96 (2009).
10. Budak, H., Akpinar, B. A., Unver, T. & Turktas, M. Proteome changes in wild and modern wheat leaves upon drought stress by two-dimensional electrophoresis and nanoLC-ESI-MS/MS. *Plant Mol. Biol.* **83,** 89–103 (2013).
11. Kubaláková, M., Vrána, J., Číhalíková, J., Simková, H. & Dolezel, J. Flow karyotyping and chromosome sorting in bread wheat (Triticum aestivum L.). *Theor. Appl. Genet.* **104,** 1362–1372 (2002).
12. Safár, J. *et al.* Development of chromosome-specific BAC resources for genomics of bread wheat. *Cytogenet. Genome Res.* **129,** 211–23 (2010).
13. Vitulo, N. *et al.* First survey of the wheat chromosome 5A composition through a next generation sequencing approach. *PLoS One* **6,** e26421 (2011).
14. Mayer, K. F. X. *et al.* A chromosome-based draft sequence of the hexaploid bread wheat (Triticum aestivum) genome. *Science.* **345,** 1251788–1251788 (2014).
15. Ling, H.-Q. *et al.* Draft genome of the wheat A-genome progenitor Triticum urartu. *Nature* **496,** 87–90 (2013).
16. Jia, J. *et al.* Aegilops tauschii draft genome sequence reveals a gene repertoire for wheat adaptation. *Nature* **496,** 91–5 (2013).
17. Venora, G. *et al.* Chromatin organisation and computer aided karyotyping of Triticum durum Desf. cv. Timilia. *Caryologia* **55,** 91–98 (2002).
18. Smith, D. B. & Flavell, R. B. Characterisation of the wheat genome by renaturation kinetics. *Chromosoma* **50,** 223–242 (1975).
19. Lucas, S. J. *et al.* Next-generation sequencing of flow-sorted wheat chromosome 5D reveals lineage-specific translocations and widespread gene duplications. *BMC Genomics* **15,** 1080 (2014).
20. Akpinar, B. A., Lucas, S. J., Vrána, J., Doležel, J. & Budak, H. Sequencing chromosome 5D of Aegilops tauschii and comparison with its allopolyploid descendant bread wheat (Triticum aestivum). *Plant Biotechnol. J.* doi:10.1111/pbi.12302 (2014).
21. Choulet, F. *et al.* Structural and functional partitioning of bread wheat chromosome 3B. *Science* **345,** 1249721–1249721 (2014).
22. Senerchia, N., Felber, F. & Parisod, C. Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytol.* **202,** 975–985 (2014).
23. International Brachypodium Initiative. Genome sequencing and analysis of the model grass Brachypodium distachyon. *Nature* **463,** 763–8 (2010).
24. Tanaka, T. *et al.* The Rice Annotation Project Database (RAP-DB): 2008 update. *Nucleic Acids Res.* **36,** D1028–D1033 (2008).
25. Paterson, A. H. *et al.* The Sorghum bicolor genome and the diversification of grasses. *Nature* **457,** 551–6 (2009).
26. Mayer, K. F. X. *et al.* A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491,** 711–6 (2012).
27. Wicker, T. *et al.* Frequent Gene Movement and Pseudogene Evolution Is Common to the Large and Complex Genomes of Wheat, Barley, and Their Relatives. *Plant Cell* **23,** 1706–1718 (2011).
28. Wicker, T., Buchmann, J. P. & Keller, B. Patching gaps in plant genomes results in gene movement and erosion of colinearity. *Genome Res.* **20,** 1229–37 (2010).
29. Tanaka, T. *et al.* Next-generation survey sequencing and the molecular organization of wheat chromosome 6B. *DNA Res.* **21,** 103–14 (2014).
30. Kojima, K. K. & Jurka, J. A superfamily of DNA transposons targeting multicopy small RNA genes. *PLoS One* **8,** e68260 (2013).
31. Budak, H., Khan, Z. & Kantar, M. History and current status of wheat miRNAs using next-generation sequencing and their roles in development and stress. *Brief. Funct. Genomics* doi:10.1093/bfgp/elu021 (2014).
32. Kantar, M., Unver, T. & Budak, H. Regulation of barley miRNAs upon dehydration stress correlated with target gene expression. *Funct. Integr. Genomics* **10,** 493–507 (2010).
33. Pfeifer, M. *et al.* Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science* **345,** 1250091–1250091 (2014).
34. You, F. M. *et al.* Annotation-based genome-wide SNP discovery in the large and complex Aegilops tauschii genome using next-generation sequencing without a reference genome sequence. *BMC Genomics* **12,** 59 (2011).
35. Zhang, H. *et al.* Genetic and comparative genomics mapping reveals that a powdery mildew resistance gene Ml3D232 originating from wild emmer co-segregates with an NBS-LRR analog in common wheat (Triticum aestivum L.). *Theor. Appl. Genet.* **121,** 1613–21 (2010).
36. Xue, F., Ji, W., Wang, C., Zhang, H. & Yang, B. High-density mapping and marker development for the powdery mildew resistance gene PmAS846 derived from wild emmer wheat (Triticum turgidum var. dicoccoides). *Theor. Appl. Genet.* **124,** 1549–60 (2012).
37. Abdollahi Mandoulakani, B. *et al.* Development of IRAP- and REMAP-derived SCAR markers for marker-assisted selection of the stripe rust resistance gene Yr15 derived from wild emmer wheat. *Theor. Appl. Genet.* doi:10.1007/s00122-014-2422-8 (2014).
38. Ouyang, S. *et al.* Fine physical and genetic mapping of powdery mildew resistance gene MlIW172 originating from wild emmer (Triticum dicoccoides). *PLoS One* **9,** e100160 (2014).

39. Wang, Z. *et al.* Comparative genetic mapping and genomic region collinearity analysis of the powdery mildew resistance gene Pm41. *Theor. Appl. Genet.* **127,** 1741–51 (2014).
40. Blanco, A. *et al.* Molecular mapping of the novel powdery mildew resistance gene Pm36 introgressed from Triticum turgidum var. dicoccoides in durum wheat. *Theor. Appl. Genet.* **117,** 135–42 (2008).
41. Qi, L., Friebe, B. & Gill, B. S. Meiotic metaphase I pairing behavior of a 5BL recombinant isochromosome in wheat. *Chromosome Res.* **8,** 671–6 (2000).
42. Griffiths, S. *et al.* Molecular characterization of Ph1 as a major chromosome pairing locus in polyploid wheat. *Nature* **439,** 749–52 (2006).
43. Bhullar, R. *et al.* Silencing of a metaphase I-specific gene results in a phenotype similar to that of the Pairing homeologous 1 (Ph1) gene mutations. *Proc. Natl. Acad. Sci. USA* **111,** 14187–92 (2014).
44. Sergeeva, E. M. *et al.* Common Wheat Chromosome 5B Composition Analysis Using Low-Coverage 454 Sequencing. *Plant Genome* **7,** 1–16 (2013).
45. Renny-Byfield, S. *et al.* Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *Plant J.* **74,** 829–39 (2013).
46. Wu, H. *et al.* Comparative high-resolution mapping of the wax inhibitors Iw1 and Iw2 in hexaploid wheat. *PLoS One* **8,** e84691 (2013).
47. Kurtoglu, K. Y., Kantar, M., Lucas, S. J. & Budak, H. Unique and conserved microRNAs in wheat chromosome 5D revealed by next-generation sequencing. *PLoS One* **8,** e69801 (2013).
48. Luan, M. *et al.* Family-wide survey of miR169s and NF-YAs and their expression profiles response to abiotic stress in maize roots. *PLoS One* **9,** e91369 (2014).
49. Shivaprasad, P. V *et al.* A microRNA superfamily regulates nucleotide binding site-leucine-rich repeats and other mRNAs. *Plant Cell* **24,** 859–74 (2012).
50. Brenchley, R. *et al.* Analysis of the bread wheat genome using whole-genome shotgun sequencing. *Nature* **491,** 705–10 (2012).
51. Sela, H. *et al.* Linkage disequilibrium and association analysis of stripe rust resistance in wild emmer wheat (Triticum turgidum ssp. dicoccoides) population in Israel. *Theor. Appl. Genet.* **127,** 2453–63 (2014).
52. Trick, M. *et al.* Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. *BMC Plant Biol.* **12,** 14 (2012).
53. Giorgi, D. *et al.* FISHIS: fluorescence *in situ* hybridization in suspension and chromosome flow sorting made easy. *PLoS One* **8,** e57994 (2013).
54. Kubaláková, M. *et al.* Analysis and sorting of rye (Secale cereale L.) chromosomes using flow cytometry. *Genome* **46,** 893–905 (2003).
55. Simková, H. *et al.* Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. *BMC Genomics* **9,** 294 (2008).
56. Nussbaumer, T. *et al.* MIPS PlantsDB: a database framework for comparative plant genome research. *Nucleic Acids Res.* **41,** D1144–51 (2013).
57. Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10,** 421 (2009).
58. Krzywinski, M. *et al.* Circos: an information aesthetic for comparative genomics. *Genome Res.* **19,** 1639–1645 (2009).
59. Conesa, A. & Götz, S. Blast2GO: A comprehensive suite for functional analysis in plant genomics. *Int. J. Plant Genomics* **2008,** 619832 (2008).
60. Lowe, T. M. & Eddy, S. R. tRNAscan-SE: A Program for Improved Detection of Transfer RNA Genes in Genomic Sequence. *Nucleic Acids Res.* **25,** 0955–964 (1997).

Acknowledgments

We thank Dr. Etienne Paux (INRA, Clermont Ferrand) for providing seeds of *Triticum dicoccoides* accession 26676. We are grateful to our colleagues Dr. Marie Kubaláková, Dr. Jarmila Číhalíková, Bc. Romana Šperková and Zdeňka Dubská for the preparation of chromosome samples and determination of purities in sorted chromosome fractions. This work was supported by Sabanci University, by the Czech Science Foundation (award no. P501/12/G090), and by the Ministry of Education, Youth and Sports of the Czech Republic (grant LO1204 from the National Program of Sustainability I).

Author Contributions

H.B. conceived the idea of the study, designed the experiments, supervised all analyses and drafted the manuscript. B.A.A. and S.J.L. carried out D.N.A. sequencing. B.A.A. and M.Y. carried out the assembly. B.A. carried out analyses regarding repeat content, gene space, tRNA/miRNA repertoires and syntenic relationships and drafted the manuscript. J.V., V.B. and J.D. purified flow-sorted chromosome 5B of *T. dicoccoides* and amplified its DNA for sequencing. J.D. edited the manuscript. All authors read and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at http://www.nature.com/srep

Competing financial interests: The authors declare no competing financial interests.

How to cite this article: Ani Akpinar, B. *et al.* Molecular organization and comparative analysis of chromosome 5B of the wild wheat ancestor *Triticum dicoccoides*. *Sci. Rep.* **5**, 10763; doi: 10.1038/srep10763 (2015).

APPENDIX IV

Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their syntheny with wheat

Molnár I, Vrána J, Burešová V, Cápal P, Farkas A, Darkó É, Cseh A, Kubaláková M, Molnár-Láng M, Doležel J

Plant Journal, doi:10.1111/tpj.13266, 2016

IF: 5.468

the plant journal

SEB Society for Experimental Biology

The Plant Journal (2016) doi: 10.1111/tpj.13266

RESOURCE

Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their synteny with wheat

István Molnár[1,*], Jan Vrána[2], Veronika Burešová[2], Petr Cápal[2], András Farkas[1], Éva Darkó[1], András Cseh[1,†], Marie Kubaláková[2], Márta Molnár-Láng[1] and Jaroslav Doležel[2]

[1]*Agricultural Institute, Centre for Agricultural Research, Hungarian Academy of Sciences, Brunszvik u. 2, H-2462 Martonvásár, Hungary, and*

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

Received 20 February 2016; revised 1 July 2016; accepted 6 July 2016.

*For correspondence (e-mail molnar.istvan@agrar.mta.hu).

[†]Present address: Division of Plant and Crop Sciences, School of Biosciences, The University of Nottingham, Sutton Bonington Campus, Loughborough, UK.

SUMMARY

Goat grasses (*Aegilops* spp.) contributed to the evolution of bread wheat and are important sources of genes and alleles for modern wheat improvement. However, their use in alien introgression breeding is hindered by poor knowledge of their genome structure and a lack of molecular tools. The analysis of large and complex genomes may be simplified by dissecting them into single chromosomes via flow cytometric sorting. In some species this is not possible due to similarities in relative DNA content among chromosomes within a karyotype. This work describes the distribution of GAA and ACG microsatellite repeats on chromosomes of the U, M, S and C genomes of *Aegilops*, and the use of microsatellite probes to label the chromosomes in suspension by fluorescence *in situ* hybridization (FISHIS). Bivariate flow cytometric analysis of chromosome DAPI fluorescence and fluorescence of FITC-labelled microsatellites made it possible to discriminate all chromosomes and sort them with negligible contamination by other chromosomes. DNA of purified chromosomes was used as a template for polymerase chain reation (PCR) using Conserved Orthologous Set (COS) markers with known positions on wheat A, B and D genomes. Wheat–*Aegilops* macrosyntenic comparisons using COS markers revealed significant rearrangements in the U and C genomes, while the M and S genomes exhibited structure similar to wheat. Purified chromosome fractions provided an attractive resource to investigate the structure and evolution of the *Aegilops* genomes, and the COS markers assigned to *Aegilops* chromosomes will facilitate alien gene introgression into wheat.

Keywords: *Aegilops umbellulata*, *Aegilops comosa*, *Aegilops speltoides*, *Aegilops markgrafii*, flow cytometric chromosome sorting, fluorescence *in situ* hybridization, conserved orthologous set markers.

INTRODUCTION

Bread wheat (*Triticum aestivum* L., 2*n* = 6*x* = 42, AABBDD genome) plays a fundamental role in the human diet. The pressure to produce enough food for the growing world population under a changing climate underlines urgent need for new high-yielding varieties with improved stress tolerance and quality-related traits. Breeding such varieties may be facilitated by employing new biotechnological tools and utilizing the extant genetic diversity among the wild relatives of wheat (Feuillet *et al.*, 2008).

The genus *Aegilops* (goatgrass) belongs to the tribe *Triticeae* and comprises 11 diploid, 10 tetraploid and two hexaploid species (Van Slageren, 1994). The U, M, S and C genomes were identified in 19 (eight diploid and 11 polyploid) *Aegilops* species (Kilian *et al.*, 2011). These species represent a rich source of genes and gene complexes that can be utilized in wheat improvement via chromosome-mediated gene transfer. For example, *Ae. umbellulata* Zhuk. (2*n* = 2*x* = 14, UU) and *Ae. comosa* Sm. in Sibth. &

Sm. ($2n = 2x = 14$, MM) are known sources of important agronomic traits such as tolerance to biotic (BYDV, Cereal cyst nematode, Hessian fly, Leaf rust, Stripe rust, Tan spot, and Powdery mildew) and abiotic stresses (Drought, Frost, Heat, Salt, Zn-deficiency), nutritional and bread-making quality (Molnár *et al.*, 2004; Schneider *et al.*, 2008; Kozub *et al.*, 2011; Dulai *et al.*, 2014; Farkas *et al.*, 2014).

Ae. speltoides Tausch. ($2n = 2x = 14$, SS) is the closest relative to the wheat B-genome (Dvorak *et al.*, 1998) and is an attractive source of genes providing tolerance against Leaf rust, Stem rust and Powdery mildew and for other traits, such as grain hardness protein, heat tolerance and tolerance to manganese toxicity (Schneider *et al.*, 2008; Kilian *et al.*, 2011). The genome of *Ae. markgrafii* (Greuter) Hammer ($2n = 2x = 14$, CC) codes for resistance genes against Leaf rust and Powdery mildew, genes for high protein and lysine content, and alleles affecting bread-making quality (Friebe *et al.*, 1992; Potz *et al.*, 1996; Liu *et al.*, 2003; Riar *et al.*, 2012).

Over the past decades, efforts were made to transfer *Aegilops* chromatin into wheat, resulting in addition, substitution and translocation lines containing chromosomes and chromosome segments from *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides* and *Ae. markgrafii* (Jiang *et al.*, 1994; Friebe *et al.*, 1996; Schneider *et al.*, 2008; Kilian *et al.*, 2011). Despite the valuable genetic variation within the wild relatives of wheat, and successful introgression of some favorable genes, the potential of alien gene transfer has been largely underutilized in wheat breeding.

The use of wild genes and alleles in breeding programs is hampered by laborious and time-consuming development of alien introgression lines. The main tools for their selection and characterization are low-throughput cytogenetic methods, such as C-banding (Friebe *et al.*, 1996), fluorescence *in situ* hybridization (FISH, Rayburn and Gill, 1985; Schwarzacher and Heslop-Harrison, 2000; Schneider *et al.*, 2005) and genomic *in situ* hybridization (GISH, Schwarzacher *et al.*, 1989; Le *et al.*, 1989). However, the potential of FISH to identify alien chromosomes and their segments is limited by small number of suitable probes, low throughout and inability to detect very small introgressions.

The efficiency of introgression breeding and the development of high-density genetic maps of *Aegilops* is limited by small number of molecular markers suitable for high-throughput screening (Zhang *et al.*, 1998). In recent decades, wheat-specific RFLP (Restriction Fragment Length Polymorphism), SSR (Simple Sequence Repeat), AFLP (Amplified Fragment Length Polymorphism) and Conserved Orthologous Set (COS) markers were tested in *Aegilops* species (Peil *et al.*, 1998; Schneider *et al.*, 2010; Rey *et al.*, 2015). Nagy *et al.* (2006) used S-SAP (Sequence-Specific Amplification Polymorphism) technology to produce 14 and 30 genome-specific markers for *Ae. umbellulata* and *Ae. biuncialis* ($2n = 4x = 28$, $U^bU^bM^bM^b$), respectively. More recently, Diversity Arrays Technology (DArT) markers and microarray hybridization-based sequence-independent marker systems were used to develop a high-density genetic map of wheat × wild emmer (Peleg *et al.*, 2008). The advent of next generation sequencing (NGS) technologies led to the development of SNP-based platforms for wheat genotyping (Rey *et al.*, 2015). However, low representation of wild wheat relatives in the SNP design may limit the utility of these platforms in alien introgression breeding (Winfield *et al.*, 2016) and new genomic resources need to be generated from wild relatives of wheat.

Poor knowledge of syntenic relationships between wheat and *Aegilops* chromosomes is another obstacle hampering the use of wild genetic diversity in wheat breeding. Collinearity between the homoeologous wheat and alien chromosomes may be interrupted as a consequence of evolutionary chromosome rearrangements in the *Aegilops* genomes (Devos *et al.*, 1993; Zhang *et al.*, 1998). Thus, genes on alien chromosome segments do not compensate for the loss of wheat genes and this may have a negative effect on agricultural performance of the wheat–alien translocations. Clearly, better knowledge on the genome organization of wild crop relatives and the development of new molecular resources and tools are needed if the extant genetic diversity of wild *Aegilops* species is to be better utilized.

The analysis of large *Triticeae* genomes can be simplified by dissecting them into individual chromosomes by flow cytometric sorting (Doležel *et al.*, 2007). As demonstrated in bread wheat, barley and rye, flow-sorted chromosomes are suitable for NGS to establish linear gene order and assess gene synteny with other species (Mayer *et al.*, 2011; Martis *et al.*, 2013; IWGSC 2014). High purity of flow-sorted chromosome fractions makes them an ideal template for PCR-based analyses and to assign molecular markers to *Aegilops* chromosomes (Molnár *et al.*, 2011b). Using gene-based COS markers and chromosomes flow-sorted from wheat–*Aegilops* introgression lines, Molnár *et al.* (2013) assigned 132 and 156 loci to the M- and U-genome chromosomes, respectively, of *Ae. comosa*, *Ae. umbellulata*, *Ae. biuncialis* and *Ae. geniculata.* The genomic position of orthologue unigene EST-contigs, which were used to design the COS markers, made it possible to investigate syntenic relationships between the U and M genomes of *Aegilops* and wheat using *Brachypodium* and rice as references. Unfortunately, in some species, flow cytometric chromosome analysis and sorting based on DAPI fluorescence alone fails to discriminate and sort all chromosomes. Thus, only chromosomes 1U, 3U and 6U could be purified from *Ae. umbellulata* and only $1U^b$ from *Ae. biuncialis*, while the remaining chromosomes could only be sorted in groups (Molnár *et al.*, 2011b). This limitation prevented a detailed comparative

analysis with wheat and hampered the use of the chromosome-based approach to sequence the genomes of wild relatives of wheat chromosome by chromosome.

To overcome this problem, Giorgi *et al.* (2013) developed a method termed FISHIS (FISH in suspension), which fluorescently labels specific microsatellite sequences on chromosomes in suspension. Some microsatellites, such as GAA and ACG motifs, form large clusters on chromosomes of *Aegilops* species and are detectable on mitotic metaphase spreads using FISH (Molnár *et al.*, 2011a), providing an opportunity to employ these repeats for fluorescent labelling of chromosomes prior to flow cytometry. Encouraged by the results obtained by genomics analyses of chromosomes flow-sorted from cereal crops, and motivated by the need to support alien introgression breeding of wheat, we set out to expand chromosome genomics in *Aegilops* and develop molecular tools and resources.

Here we report on the use of two microsatellite repeats, GAA and ACG, as probes for FISH to identify mitotic chromosomes of *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides* and *Ae. markgrafii.* The same microsatellite repeat probes were used to fluorescently label chromosomes in suspension prior to flow cytometric analysis to facilitate sorting all chromosomes from diploid progenitors of the U, M, S and C genomes of *Aegilops.* DNA amplified from flow-sorted chromosomes was used for PCR with COS markers to obtain insights into the macrosyntenic relationships between the genomes of *Aegilops* and bread wheat at chromosome level.

RESULTS

Chromosomal distribution of GAA and ACG repeats

In order to investigate the potential of GAA and ACG repeats as probes for fluorescent labelling chromosomes in suspension and to provide additional chromosomal landmarks for identification of *Aegilops* chromosomes and chromosome segments, sequential FISH was carried out on mitotic metaphase plates of *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides* and *Ae. markgrafii* using probes for the two microsatellites and probes for tandem repeats pSc119.2, Afa family and 18S rDNA (Figure 1). The karyotypes obtained are shown in Figure 2 and detailed in Table S1. Only minor differences in fluorescent labelling patterns were observed between this work and the results obtained by Badaeva *et al.* (1996a,b) (Table S1), and we could identify all chromosomes in the diploid *Aegilops* species. The labelling efficiency (i.e. the number and intensity of hybridization signals) of the microsatellite probes (Table S2), showed significant intragenomic differences among the four *Aegilops* species. No differences in FISH labelling patterns were observed between the two accessions of *Ae. markgrafii* (MvGB428 and MvGB607).

Flow sorting of mitotic chromosomes after FISHIS

When suspensions of mitotic chromosomes from diploid *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides and Ae. markgrafii* were analyzed for the distribution of DAPI fluorescence intensity (flow karyotypes), narrow peaks were obtained, giving better chromosome resolution as

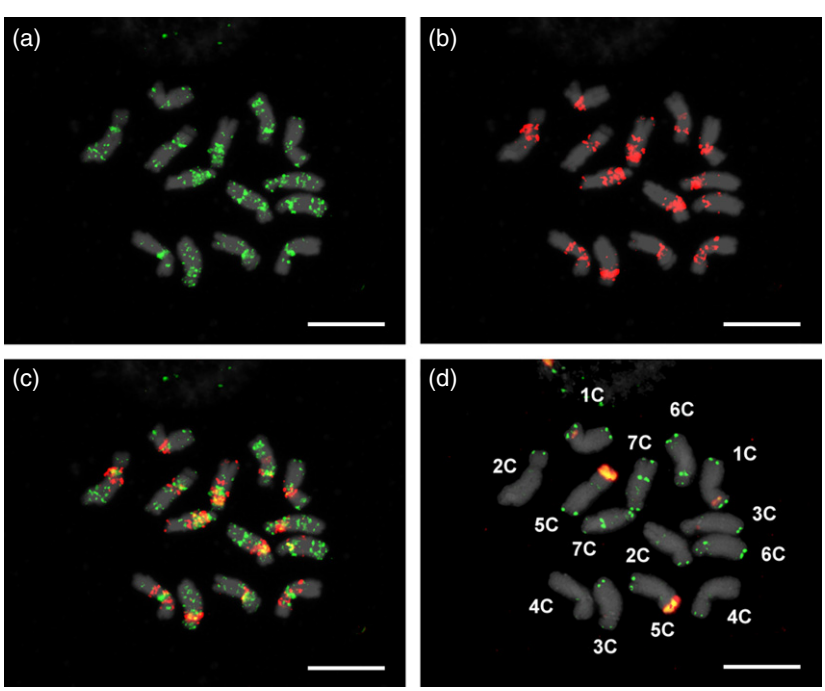


Figure 1. Fluorescence *in situ* hybridization on mitotic metaphase plates of *Aegilops markgrafii* with probes for GAA (green) and ACG (red) microsatellites (a–c), and with probes for 18S rDNA (yellow) and pSc119.2 repeat (green) (d). Chromosomes were counterstained by DAPI (grey). Bar = 10 μm.

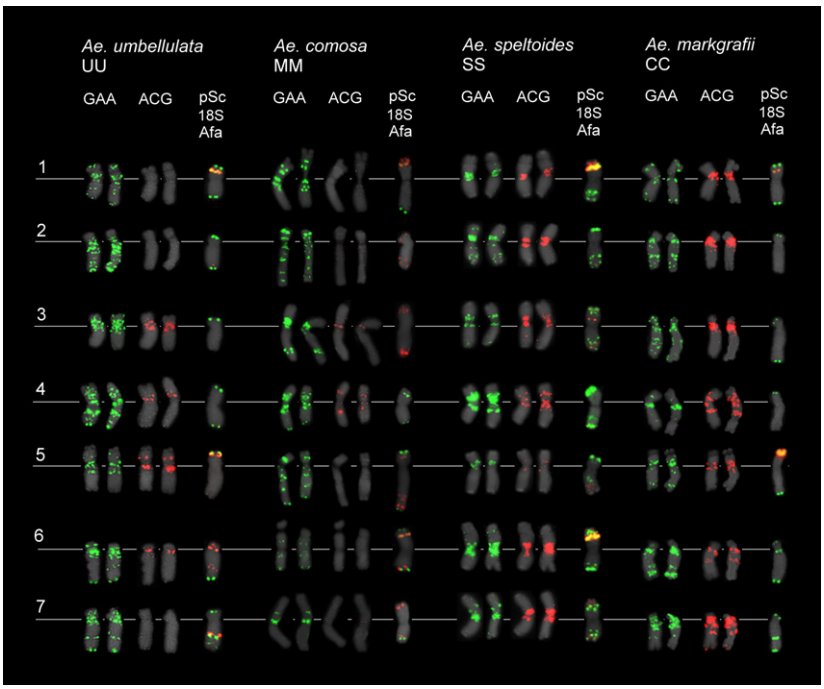


Figure 2. Representative karyotypes of *Aegilops umbellulata* (AE740/03), *Ae. comosa* (MvGB1039), *Ae. speltoides* (MvGB905) and *Ae. markgrafii* (MvGB428) after fluorescence *in situ* hybridization with repetitive DNA probes.
The signals of GAA and ACG probes were visualized as green and red, respectively, while the probes for 18S rDNA (yellow), Afa family repeat (red) and pSc119.2 repeat (green) were detected simultaneously. Chromosomes were counterstained by DAPI (grey).

compared to our previous work (Molnár *et al.*, 2011b, 2014, 2015). This was probably due to the fact that we used a BD FACSAria II SORP flow sorter in this study, which employs a gel-coupled flow cell instead of the classic jet-in-air system of BD FACSVantage flow sorter. The former system is more efficient in collecting fluorescence light pulses and provides better stability of the fluid stream.

Monovariate flow karyotype of *Ae. umbellulata* consisted of peaks I–III representing chromosomes 1U, 6U and 3U, respectively, and one composite peak IV containing the chromosomes 2U, 4U, 5U and 7U (Figure 3a). The bivariate flow karyotype obtained after FISHIS with a probe for GAA motif consisted of seven clearly separated populations corresponding to the seven chromosomes of *Ae. umbellulata* (Figure 3b). The chromosomes were assigned to the chromosome populations by FISH with probes for pSc119.2, Afa family and 18S rDNA on chromosomes flow-sorted onto microscope slides (Table S3). Better resolution of chromosome populations after bivariate flow karyotyping resulted in high purity (88–98%) of sorted chromosome fractions (Table 1).

Bivariate flow karyotyping in *Ae. comosa* after FISHIS with a probe for GAA (Figure 4a) revealed three chromosome populations (IV, VI and VII) representing chromosomes 6M, 3M and 7M, respectively (Figure S1). The three chromosomes could be sorted with a purity of 96.7%, 94.2% and 93.3%, respectively. On the other hand, populations of 1M and 4M, and 2M and 5M overlapped, resulting in lower purities (1M: 44.8%, 4M: 53.8%, 5M: 86.5%, 2M: 62.6%). To improve chromosome discrimination, double FISHIS was employed with probes for GAA and ACG (Figure 4b). This resulted in better separation of the chromosome populations and allowed chromosomes 1M, 2M, 4M and 5M to be sorted at purities of 79.6, 73.6, 78.4 and 90.2%, respectively (Figure S1 and Table S3). Importantly, the purity of the sorted 3M, 6M and 7M fractions also improved (Table 1).

As the combined use of GAA and ACG microsatellite repeats for FISHIS had a positive effect on bivariate flow karyotyping in *Ae. comosa*, the same approach was used in *Ae. speltoides* and *Ae. markgrafii.* Differences in the abundance of GAA and ACG motifs between chromosomes were large enough to allow separation of all S- and C-genome chromosomes (Figure 5). FISH analysis on flow-sorted chromosomes of *Ae. speltoides* showed that the populations of chromosomes 1S, 3S and 5S, on which GAA and ACG repeats are less abundant (Figure 2), were allocated in regions III, V and IV of the bivariate flow karyotype, characterized by lower FITC fluorescence intensity (Figures 5a and S2 and Table S3). Conversely, chromosome 4S, which has strong and complex GAA and ACG hybridization patterns, was assigned to the population with the highest level of FITC fluorescence (Figure 5a; region I).

Two accessions of *Ae. markgrafii* (MvGB428 and MvGB607) were used to secure enough seed to allow replications of the experiments. FISH on flow-sorted chromosome fractions showed that chromosomes 4C, 6C and 7C, which had complex, strong microsatellite hybridization patterns (Figure 2), were represented by populations VII, III and I, respectively, on bivariate flow karyotype (Figure 5b), while chromosomes 1C, 2C, 3C and 5C, which had lower

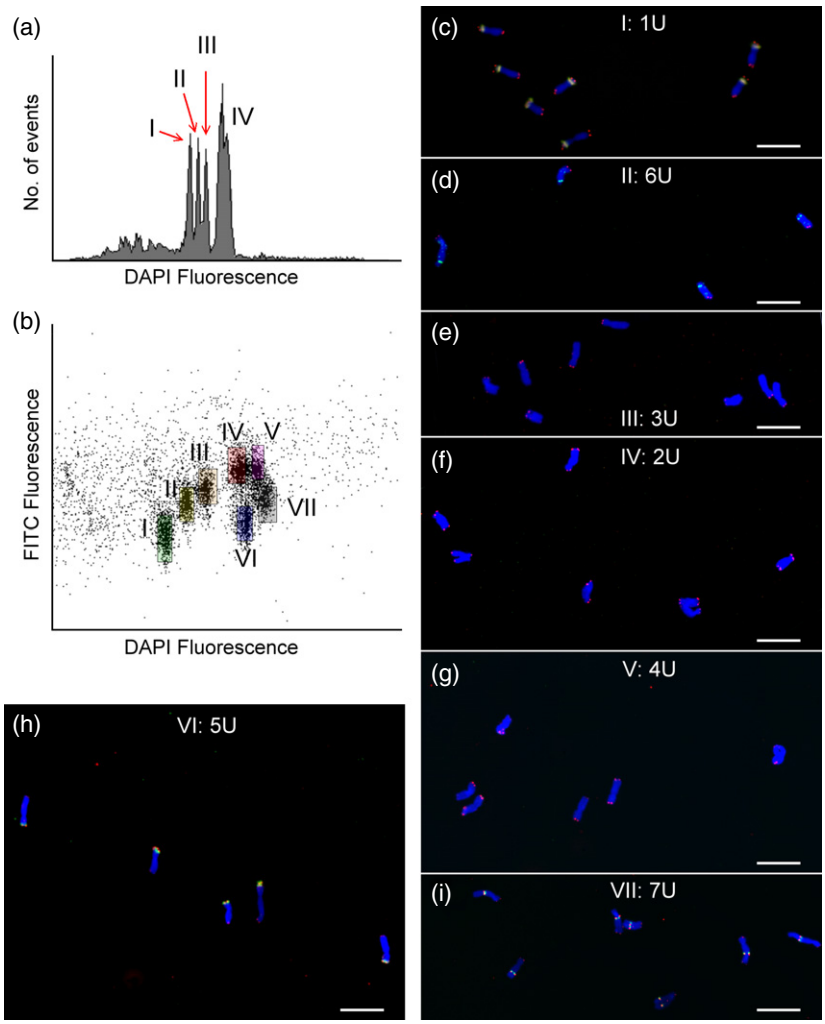


Figure 3. Flow cytometric analysis and sorting of *Ae. umbellulata* chromosomes.
(a) Distribution of fluorescence intensity (flow karyotype) obtained after the analysis of DAPI-stained suspensions of mitotic chromosomes. Monovariate flow karyotype comprises peaks I–III representing chromosomes 1U, 6U and 3U, respectively, and a composite peak of the remaining four chromosomes.
(b) Bivariate (DAPI versus GAA-FITC) flow karyotyping and sorting in *Ae. umbellulata*. FISHIS with probes for GAA resolved seven chromosome groups (I–VII colored regions).
(c–i) Chromosomes were flow-sorted from the colored regions I–VII onto microscope slides and identified by fluorescence *in situ* hybridization with probes for DNA repeats pSc119.2 (red), Afa family (green) and 18S rDNA (yellow). All seven chromosomes of *Ae. umbellulata* could be sorted at purities 88–98%. Bar = 20 μm.

Table 1 The effect of flow cytometric chromosome analysis method on the number of discriminated and sorted chromosomes and purity in flow-sorted chromosome fractions

	Ae. umbellulata			*Ae. comosa*			*Ae. speltoides*			*Ae. markgrafii*		
Method	Chr	% of the genome	Purity (%)	Chr.	% of the genome	Purity (%)	Chr.	% of the genome	Purity (%)	Chr.	% of the genome	Purity (%)
Monoparametric (DAPI)	1U	12.9	98.9[a]	–[a]			5S	13.8	89.8[b]	4C	12.4	91.3[c]
	3U	13.3	86.4[a]	–			–			–		
	6U	13.4	74.1[a]	–			–			–		
Biparametric (DAPI + FITC)	1U	12.9	98.9	1M	14.7	79.6	1S	13.8	98.8	1C	13.1	91.8
	2U	14.3	88.7	2M	13.1	73.6	2S	15.2	84.4	2C	15.8	94.4
	3U	13.3	96.4	3M	15.6	96.7	3S	15.5	95.7	3C	15.1	89.6
	4U	15.5	90.1	4M	12.6	78.0	4S	13.1	93.0	4C	12.4	97.9
	5U	15.1	93.2	5M	14.2	90.2	5S	13.8	99.2	5C	15.5	90.7
	6U	13.4	94.2	6M	13.6	99.6	6S	13.4	97.1	6C	12.0	91.9
	7U	15.2	98.0	7M	15.8	98.4	7S	14.9	99.0	7C	15.7	80.1

[a,b,c]Data from Molnár *et al.*, 2011b, 2014 and 2015, respectively.

GAA and ACG content, were assigned to populations with lower FITC fluorescence intensity (Figures 5b and S3). With the exception of chromosomes 2S and 7C, which could be sorted at purities of 84.4% and 80.9%, respectively, bivariate flow cytometry after FISHIS with probes for GAA and ACG permitted complete sets of chromosomes from

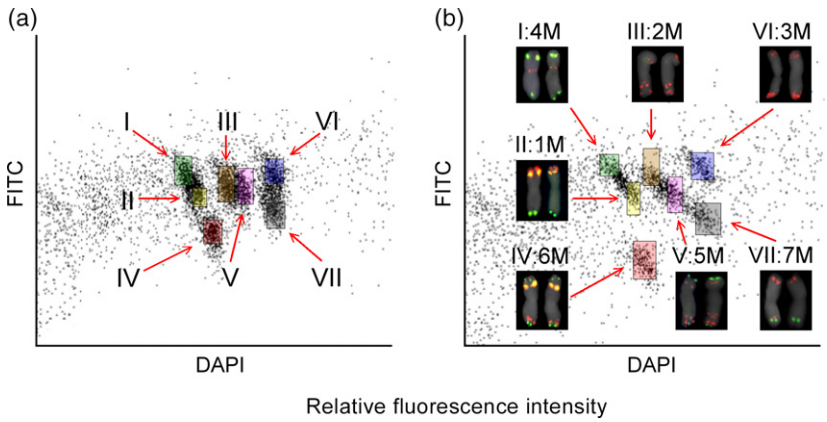


Figure 4. Bivariate flow karyotyping and flow sorting of *Ae. comosa* chromosomes.
(a) FISHIS with probes for GAA resolved only three chromosome groups (IV, VI and VII colored regions) specific for chromosomes 3M, 6M and 7M.
(b) Dual FISHIS with probes for GAA and ACG resolved all seven M-genome chromosomes of *Ae. comosa*, which could be flow-sorted at purities of 73–99%. Chromosomes were assigned to the colored regions by fluorescence *in situ* hybridization using probes for 18S rDNA (yellow), Afa family (red) and pSc119.2 (green). Chromosomes were counterstained by DAPI (grey).

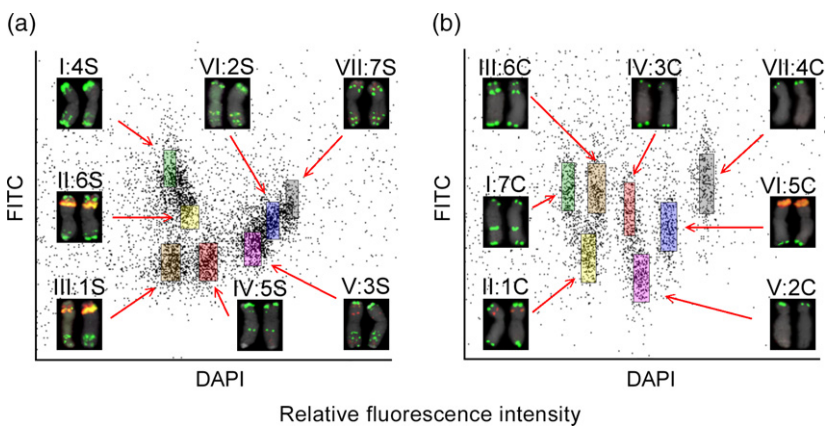


Figure 5. Bivariate flow karyotyping and flow sorting of chromosomes from (a) *Ae. speltoides* and (b) *Ae. markgrafii*.
Dual FISHIS with probes for GAA and ACG resolved all S-genome and C-genome chromosomes, which could be flow-sorted at purities of 84–99% and 80–97%, respectively. Chromosomes were assigned to the colored regions by fluorescence *in situ* hybridization using probes for 18S rDNA (yellow), Afa family (red) and pSc119.2 (green). Chromosomes were counterstained by DAPI (grey).

Ae. speltoides and *Ae. markgrafii* to be sorted at purities exceeding 93% and 90%, respectively (Table 1, Table S3).

Sorting chromosome arms after FISHIS

Stimulated by the positive results, we checked the utility of bivariate flow cytometry to purify chromosome arms of *Aegilops* from wheat–*Ae. umbellulata* ditelosomic addition lines. Chromosome suspensions of wheat (*T. aestivum* cv. Chinese Spring)-*Ae. umbellulata* double ditelosomic addition lines CSDtA2US (Figure 6a), CSDtA2UL (Figure 6b) and CSDtA7UL (Figure 6c) were labelled by FISHIS with a probe for GAA. Chromosome arms 2US, 2UL and 7UL of *Ae. umbellulata* could be easily discriminated from wheat chromosomes on bivariate flow karyotypes (Figures 6a–c and S4), allowing these arms to be sorted at high purities ranging from 88 to 94%.

Assignment of COS markers to U, M, S and C chromosomes

COS markers designed from wheat expressed sequence tags (ESTs) for which chromosome deletion bin map positions are known were assigned to *Aegilops* U-, M-, S- and C-genome chromosomes using PCR, with DNA amplified from flow-sorted chromosomes as a template (Table S4). Of the 123 COS markers, 100 amplified PCR products from genomic DNA of at least one of the four *Aegilops* species (Data S1). The 100 markers resulted in a total of 544 PCR products in the four *Aegilops* species (137, 131, 127 and 142 amplicons in *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides* and in the two accessions of *Ae. markgrafii*, respectively).

Because each of the *Aegilops* chromosomes has a major location in one of the populations on bivariate flow karyotype (Tables 1 and S3), the highest amount of PCR product obtained with a COS marker identified the population with the locus-carrying chromosome (Data S1 and Table S3). However, if the amounts of PCR product were similar in two different chromosome populations, it was not possible to discriminate between the intragenomic duplication and a false positive chromosomal assignment. Thus, COS markers which gave differences of less than 10% between the PCR product amounts of two different chromosome populations were excluded from further analysis. In total, 466 PCR products (225 polymorphic and 241 non-polymorphic with respect to wheat) were assigned to *Aegilops* chromosomes (Data S1).

Out of 118 loci assigned to U-genome chromosomes of diploid *Ae. umbellulata* (Table S5), 63 loci (53.38%) were polymorphic relative to wheat cv. GK Öthalom. In *Ae. comosa*, where 114 loci were mapped to M-genome chromosomes, 53 loci (46.49%) were polymorphic. Of the 120 loci

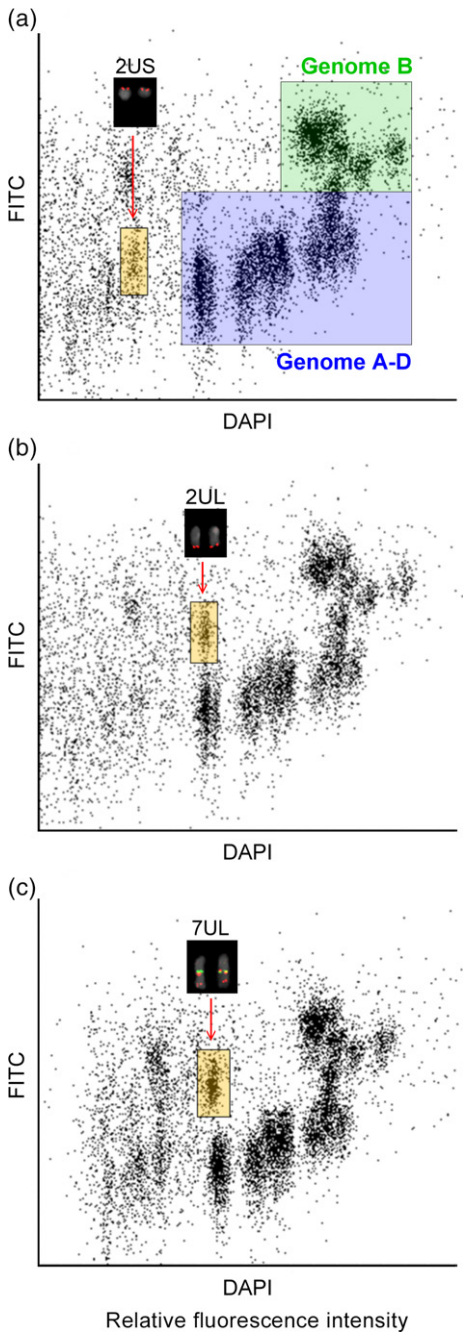


Figure 6. Bivariate flow karyotyping after FISHIS with a probe for GAA and flow sorting *Ae. umbellulata* chromosome arms from wheat (*T. aestivum* cv. Chinese Spring)–*Ae. umbellulata* double ditelosomic addition lines CSDtA2US, CSDtA2UL and CSDtA7UL.
(a–c) (a) FISHIS allowed discrimination of the homoeologous genomes A, D and B of hexaploid wheat (blue and green boxes, respectively) and populations representing 2US (a), 2UL (b) and 7UL (c). Chromosome arms 2US, 2UL and 7UL were identified using fluorescence *in situ* hybridization with probes for Afa family (green) and pSc119.2 (red) and could be sorted at purities of 94.9, 90.3 and 88.3%, respectively. Chromosomes were counterstained by DAPI (grey).

assigned to S-genome chromosomes of *Ae. speltoides*, 56 loci (46.66%) showed size polymorphism. Finally, 53 (46.49%) of the 114 loci mapped to C-genome chromosomes of *Ae. markgrafii* were polymorphic. Chromosome-specific COS markers with significant (≥5 bp) length polymorphism between wheat cv. GK Öthalom and *Aegilops* species (Table 2) will be suitable for marker-assisted selection of wheat–*Aegilops* introgression lines.

Wheat–*Aegilops* homology at chromosome level

Using the genetic map data and the deletion bin positions of the source ESTs (Supplementary Data S2), the 100 COS markers assigned to *Aegilops* chromosomes were physically mapped on wheat B, A and D genomes (Figures 7, S5 and S6). This provided an overview of the genome relationships between wheat and *Aegilops* species (Figures 7, S5 and S6).

The coverage of wheat B-genome chromosomes 3B, 5B, 6B and 7B with COS markers (16, 15, 15 and 20 markers/chromosome, respectively) was better as compared with the remaining chromosomes (1B, 2B and 4B with 12, 11 and 10 markers, respectively). Similar results were obtained for the A-genome chromosomes and to some extent for the D-genome chromosomes, where 17, 15 and 20 markers were specific for chromosomes 3D, 6D and 7D, respectively (Figures S5 and S6). Based on the presence or absence of COS markers on the same homoeologous group chromosomes in wheat and *Aegilops*, genetic relationships were quantified using the Jaccard similarity coefficients (Table S6) (Kosman and Leonard, 2005).

At the whole genome level, the structures of the S-genome chromosomes of *Ae. speltoides* and the M-genome of *Ae. comosa* were the most similar to wheat, followed by the U-genome of *Ae. umbellulata*, while the structure of the C-genome in *Ae. markgrafii* differed considerably. At chromosome level, the group 1 and group 5 chromosomes of *Aegilops* species generally showed greater macrosynteny with wheat than the remaining chromosome groups (Table S6).

The chromosomal locations of orthologous genes revealed structural relationships between the U-genome chromosomes of *Ae. umbellulata* and the A, B and D genomes of wheat. For example, COS marker *c746642*, specific for wheat (W) chromosome group 2 (W2), was located on chromosome 6U, COS marker *c755442* specific for W3 was located on 7U, four markers indicated homology between the short arms of W4 and 6U, while two markers indicate that intercalary part of the long arm of W6 is related to 4U. Another part of the W6 long arm, represented by five markers, was found to be homologous to 2U (Figures 7, S5 and S6).

Chromosomes of *Ae. comosa* exhibited greater synteny with wheat than those of *Ae. umbellulata.* However, some

Table 2 COS markers showing polymorphic (≥5 bp) PCR amplicons between wheat and *Aegilops* species, which are considered suitable for identification of introgressions of the U-, M-, S- and C-genome chromosomes from *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides* and *Ae. markgrafii* into hexaploid wheat. The size (in bp) of the chromosome-specific loci is shown in brackets

Homoeologous group in *Aegilops*	*Ae. umbellulata* (UU)	*Ae. comosa* (MM)	*Ae. speltoides* (SS)	*Ae. markgrafii* (CC)
1	*c757212* (244); *c735941* (238); *c743018* (298, 310); *c726029* (418); *c743346* (275); *c737520* (327); *c744747* (320); *c758392* (379, 390)	*c757212* (285); *c735941* (238); *c743346* (277); *c737520* (327); *c744747* (317)	*c757212* (280); *c735941* (227, 239); *c743018* (305, 317); *c743346* (278); *c737520* (330); *c744747* (317)	*c757212* (285); *c735941* (237); *c743018* (298, 310); *c743346* (274); *c737520* (327); *c744747* (320); *c751053* (498); *c765452* (357)
2	*c740970* (207); *c757237* (190, 194); *c767104* (443); *c741435* (201); *c760549* (430); *c742110* (194, 198); *c742079* (374)	*c740970* (207); *c757237* (230, 233); *c762599* (267, 269)	*c720763* (323, 326)	*c756721* (307); *c765220* (298, 302, 310); *c744766* (239); *c747871* (655); *c724406* (628); *c741435* (588); *c760549* (428); *c753637* (442); *be496986* (629); *c771657* (888); *c748987* (260); *c754211* (288, 291)
3	*c752137* (399, 410); *c805553* (442, 451); *c772427* (371); *c757460* (633); *c756279* (308); *c755305* (263)	*c805553* (450); *c772427* (371); *c751053* (502); *c752685* (597); *c771860* (374); *c740781* (413); *c756279* (285); *c761505* (1374); *c750237* (517); *c732202* (232); *c740257* (280); *c748987* (260)	*c757237* (228); *c746642* (654); *c805553* (450); *c751053* (595); *c739776* (323); *c741435* (468)	*c767104* (422); *c805553* (442, 451); *760830* (300, 305); *bf484254* (556); *c747342* (655); *c745166* (243); *c740257* (280)
4	*c759427* (557, 552); *c765452* (310, 322); *c724406* (633); *be496986* (716)	*c743018* (298, 310); *c733078* (458); *c765452* (310, 322); *c760004* (697); *bf484254* (536)	*c770094* (432); *c742110* (561)	*c740970* (207); *c757237* (225, 228); *c757460* (654)
5	*c762599* (269); *c743567* (588); *c758334* (630); *c728956* (340); *c756721* (308); *c771643* (370); *c748436* (873); *c749645* (354, 362); *c765220* (300, 304, 313); *c732202* (322)	*c743567* (585); *c756721* (295); *c748436* (745); *c749645* (316, 326); *c765220* (297, 301, 309); *c732202* (254)	*c762599* (267, 269); *c743567* (585); *c758334* (630); *c756721* (311); *c744654* (328); *c748436* (810); *c724685* (674); *c749645* (348, 356); *c765220* (299, 304, 312)	*c762599* (264, 269); *c743567* (585); *c758334* (622); *c748436* (795); *c749645* (339, 348)
6	*c746642* (673); *c771614* (286); *c760004* (690); *c744766* (238); *c747871* (657); *c753637* (424); *c760754* (430); *c771657* (836); *c754211* (281, 287); *c743137* (478)	*c744766* (254); *c747871* (660); *c724406* (700); *c760549* (430); *c753637* (424); *be496986* (647)	*c740781* (412); *c765452* (304, 308); *c760004* (177); *c737067* (470); *c744766* (251); *c747871* (660); *c724406* (694); *c760549* (428); *c753637* (514); *be496986* (633)	*c743137* (514)
7	*c760830* (300, 305); *bf484254* (568); *c759439* (849); *c747342* (663); *c745166* (243)	*c760830* (300, 305); *be494425* (531); *c759439* (851); *c747342* (668); *c754211* (281, 287, 290); *c743137* (514)	*c760830* (300, 305); *bf484254* (568); *c732202* (644); *c771657* (819); *c741119* (760); *c747342* (696); *c745166* (243); *c740257* (280); *c769080* (349); *c753911* (165); *c754211* (289, 292); *c743137* (515)	*c720763* (308, 311); *c746642* (694); *c744070* (215); *c765452* (309, 313, 321); *c760004* (685)

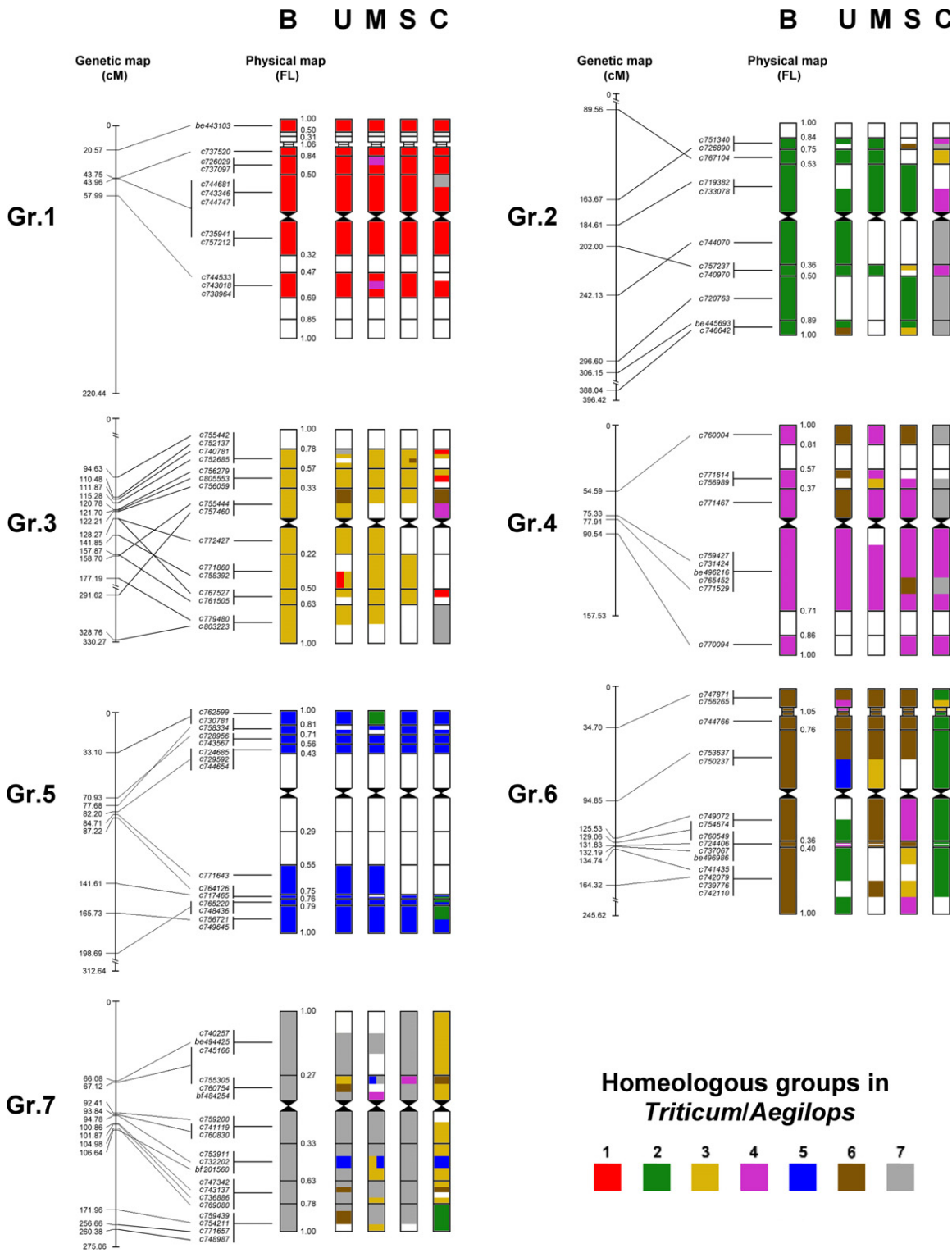


Figure 7. Visualization of wheat–*Aegilops* orthologous relationships from the perspective of wheat B-genome chromosomes.
Genetic map positions of the source ESTs of the COS markers are indicated on the left, while the physical positions on the deletion bin map are indicated on the right. Each marker assigned to chromosomes of *Ae. umbellulata* (U), *Ae. comosa* (M), *Ae. speltoides* (S) and *Ae. markgrafii* (C) is positioned to its known bin position and ordered within each chromosomal bin by the cM value of the marker-containing scaffold obtained from the Genome Zipper of the corresponding wheat chromosome arm. The wheat deletion bins were divided into windows according to the number of markers and each window was color-coded to visualize the marker position on the homoeologous groups of *Triticum*/*Aegilops* chromosomes. When a marker mapped to two chromosomes within A-genome, the marker-window was double color-coded. Marker windows and chromosome bins without markers were colored white.

 The Plant Journal, (2016), doi: 10.1111/tpj.13266

Table 3 Genomic regions conserved between hexaploid wheat and U, M, S and C-genome-chromosomes of *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides* and *Ae. Markgrafii*, respectively. The number of COS markers representing wheat homoeologous regions is shown in parentheses

Homoeologous group in *Aegilops* chromosomes		*Ae. umbellulata*	*Ae. comosa*	*Ae. speltoides*	*Ae. markgrafii*
1		W1 (12); W3 (1)	W1 (10)	W1 (12)	W1 (10); W3 (4)
2	2US	W2 (2); W6 (5)	W2 (7); W5 (1)	W2 (4); W3 (2)	W5 (3); W6 (12); W7 (4)
	2UL	W2 (5)			
3		W3 (10); W7 (1)	W3 (15); W4 (1); W6 (1); W7 (4)	W2 (2); W3 (13); W6 (2)	W2 (1); W3 (2); W6 (1); W7 (11)
4		W4 (5); W6 (3)	W1 (2); W4 (7); W7 (1)	W4 (7); W6 (3); W7 (1);	W2 (4); W3 (1); W4 (4);
5		W5 (14); W6 (1); W7 (1)	W5 (12); W7 (2)	W5 (14)	W5 (10); W7 (1)
6		W2 (1); W3 (1); W4 (3); W6 (3); W7 (4)	W6 (10);	W2 (1); W3 (1); W4 (2); W6 (8)	W3 (1); W7 (2);
7	7US	W7 (1)	W7 (12);	W7 (18)	W1 (1); W2 (5); W3 (2); W4 (5)
	7UL	W3 (1); W7 (11)			

rearrangements were observed relative to wheat. One COS marker indicated presence of a W5 fragment on 2M and four markers suggested a homology between W7 and 3M (Figures 7, S5 and S6). As expected, the S-genome of *Ae. speltoides* was closely related to wheat. However, two COS markers indicated genome rearrangements between W2 and 3S, while two markers specific for W4 were found on 6S. Homology between the long arm of W6 and 4S was indicated by three markers and between W6 and 3S by two markers (Figures 7, S5 and S6).

In *Ae. markgrafii*, chromosomes 1C and 5C exhibited the greatest synteny with wheat homoeologous groups, although three markers indicated the presence of a W5-specific region on chromosome 2C. It seems that the long arms of 2B and 3B, and the short arm of 4B are related to 7C. Five markers located on the long arm of 4B and four markers specific for different parts of 2B were detected on chromosome 4C, indicating their homology. Twelve markers specific for 6B were located on chromosome 2C, while 11 markers indicated homology between 7B and 3C.

Table 3 provides a complete list of conserved genomic regions between hexaploid wheat genomes and chromosomes from the U, M, S and C genomes of diploid *Aegilops* species as identified in the present work.

DISCUSSION

The exploitation of *Aegilops* species for wheat improvement has been the subject of research for more than a century. Yet, with a few exceptions, the large genetic diversity of *Aegilops* remains untapped (Schneider *et al.*, 2008; Kilian *et al.*, 2011). The present work aims to contribute to the efforts to change this by developing approaches to simplify the analysis of *Aegilops* genomes, describing relationships between (sub)genomes of bread wheat and genomes of four *Aegilops* species, and developing markers to facilitate exploitation of important traits in wheat breeding programs.

We demonstrate that it is possible to dissect the large U, M, S and C genomes of *Aegilops* into individual chromosomes representing 12.0–15.8% of the whole genome. This should facilitate the analysis and mapping these complex genomes whose 1C values exceed 4 Gbp (U: ~4938 Mbp, M: ~6044 Mbp, S: ~5036 Mbp, C: ~4528 Mbp), and which comprise high proportion of repetitive DNA (57 and 61% for *Ae. speltoides* and *Ae. tauschii*, respectively) (Kilian *et al.*, 2011; Shangguan *et al.*, 2013). Slicing the genomes into single chromosomes provides a powerful approach to perform structural and functional genome analysis (Doležel *et al.*, 2014; Rey *et al.*, 2015).

Chromosome samples are traditionally stained by DAPI and classified according to their relative DNA content using flow cytometry. Only chromosomes whose DAPI fluorescence intensity differs from other chromosomes in a karyotype can be discriminated and purified (Doležel *et al.*, 1992). As many species have chromosomes of similar size, individual chromosomes cannot be easily discriminated based on DAPI staining alone. Thus, only group 5 chromosomes could be sorted from *Ae. tauschii* and *Ae. speltoides* (Molnár *et al.*, 2014), chromosome 4C from *Ae. markgrafii* (Molnár *et al.*, 2015) and chromosomes 1U, 3U and 6U from *Ae. umbelulata* (Molnár *et al.*, 2011b).

To overcome the difficulty to sort particular chromosomes, Vrána *et al.* (2015) suggested dissecting composite chromosome peaks representing several chromosomes into smaller sections enriched for the chromosomes of interest, while Cápal *et al.* (2015) developed a protocol for sequencing single flow-sorted chromosomes. While useful for certain applications, these approaches do not allow particular chromosomes to be sorted at high purity and/or in large numbers. Conversely, labelling specific DNA

sequences by FISH should facilitate discrimination of otherwise indistinguishable chromosomes and their sorting in large numbers (Lucretti *et al.*, 2014). The present results show that the distribution of GAA and ACG hybridization signals differs within the U, M, S and C genomes. These results are on line with previous observations that microsatellite trinucleotide repeats (GAA, AAC, ACG) provide diagnostic landmarks to identify chromosomes in cereals such as wheat, barley and rye (Kubaláková *et al.*, 2005; Cuadrado *et al.*, 2008) and in *Aegilops* species with the U and M genomes (Molnár *et al.*, 2011a). The GAA and ACG karyotypes obtained in the present study show that the microsatellites provide useful chromosomal landmarks also in *Ae. speltoides* and *Ae. markgrafii.*

Motivated by the results of FISH on mitotic metaphase chromosomes, we used FISHIS (Giorgi *et al.*, 2013) to label the microsatellite repeats on chromosomes in suspension to improve chromosome discrimination and facilitate chromosome sorting in *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides* and *Ae. markgrafii.* Relative positions of populations representing individual chromosomes on bivariate flow karyotypes DAPI versus microsatellite-FITC agreed well with the number and intensity of GAA or ACG bands observed on mitotic metaphases.

In *Ae. comosa*, *Ae. speltoides* and *Ae. markgrafii*, FISHIS with the GAA probe alone did not discriminate the complete chromosome complements. This was achieved by dual FISHIS with probes for GAA and ACG, which increased the FITC signal diversity and improved discrimination of individual chromosomes. These results indicate that FISHIS with an appropriate mix of probes for microsatellite repeats may improve discrimination of individual chromosomes, even if the probes are labeled with the same fluorochrome. This approach could increase the potential of chromosome genomics in Triticeae and perhaps also in other species.

Contamination of sorted chromosome fractions by other chromosomes or chromosome fragments is common in flow cytometric chromosome sorting (Lysák *et al.*, 1999; Vitulo *et al.*, 2011; Doležel *et al.*, 2012). The present results demonstrated that bivariate flow karyotyping after FISHIS not only increased the number of *Aegilops* chromosomes that could be discriminated and sorted, but also increased the purity in flow-sorted fractions. This situation is in line with the observations of Giorgi *et al.* (2013).

The range of applications of flow-sorted chromosomes keeps expanding (Doležel *et al.*, 2012), and includes physical mapping using FISH (Valárik *et al.*, 2004), construction of large-insert DNA libraries (Šafář *et al.*, 2004), optical mapping (Staňková *et al.*, 2016), development of DNA markers (Bartoš *et al.*, 2008), and physical mapping on DNA arrays (Mayer *et al.*, 2011). Shot-gun NGS represents a particularly important application of flow-sorted chromosomes and has been the foundation of many international genome sequencing projects, including barley, rye and bread wheat (Mayer *et al.*, 2011; Martis *et al.*, 2013; The International Wheat Genome Sequencing Consortium (IWGSC), 2014).

The ability to purify chromosomes from the U, M, S and C genomes of *Aegilops* and production of microgram DNA amounts from them opens avenues for the application of chromosome genomics in *Aegilops* to support alien introgression breeding. For example, Tiwari *et al.* (2014) flow-sorted short arm of chromosome $5M^g$ from a wheat-*Ae. geniculata* ditelosomic addition line and sequenced it by Illumina technology. Out of the 2178 $5M^gS$-specific SNPs identified, 44 were validated by KASP assay and used to identify $5M^gS$-specific chromosome segments in released wheat germplasm lines. These results highlighted the importance of DNA samples derived from wild wheat relatives and their suitability for NGS and development of high-throughput genotyping assays to identify alien introgressions.

Alien gene transfer induced by homoeologous recombination (Riley and Chapman, 1958; Sears, 1977) depends on chromosome collinearity and may be hampered by irregularities in meiotic pairing of alien chromosomes with their wheat homoeologues due to structural rearrangements (Ceoloni *et al.*, 1988; Devos *et al.*, 1993; Cuadrado *et al.*, 1997; Lukaszewski *et al.*, 2004). The lack of knowledge on the evolutionary relationships between wheat and *Aegilops* hampers alien gene transfer, for example due to non-compensating translocations, (Friebe *et al.*, 1996; Ceoloni and Jauhar, 2006). The knowledge of wheat–*Aegilops* macrosyntenic relationships is also important to support targeted development of molecular markers specific for *Aegilops* chromosome regions potentially responsible for agronomic traits of interest (Burt and Nicholson, 2011) and to minimize the amount of undesirable alien chromatin.

Wheat–*Ae. umbellulata* macrosynteny was investigated using the RFLP-based genetic map of *Ae. umbellulata* (Zhang *et al.*, 1998; Devos and Gale, 2000) and at least 11 rearrangements were found that differentiated U-genome chromosomes from the D-genome of wheat. Later, Molnár *et al.* (2013) used wheat-specific COS markers on wheat–*Aegilops* addition lines and flow-sorted chromosomes to describe relationships between wheat genome and the U- and M-genomes of diploid and polyploid *Aegilops.* The present work extends the comparative analysis of wheat and *Aegilops* to the S and C genomes of *Ae. speltoides* and *Ae. markgrafii.* We used complete sets of chromosome-derived DNA samples to assign COS markers to *Aegilops* chromosomes and compare the structure of the *Aegilops* U-, M-, S- and C-genomes with the A-, B- and D-genomes of hexaploid wheat. Polymorphic markers assigned to U-, M-, S- or C-genome chromosomes will be useful to support the transfer of alien chromosomes or chromosome arms into wheat.

The U-genome–wheat homoeologouos relationships observed in this work were similar to those reported by Zhang *et al.* (1998) and Gale and Devos (1998). We found that 1U was related mainly to W1 which was also true for *Aegilops* group 1 chromosomes 1M, 1S and 1C. Danilova *et al.* (2014) used FISH to map full-length cDNA clones to wheat chromosomes. With two to six probes per chromosome arm, the authors observed close relationship between chromosomes 1U, 1C and W1. According to Zhang *et al.* (1998), the distal part of the long arm of W1 (represented by three RFLP markers) was related to chromosome 6U. In our work, relatively large distal bins on the long arm of W1 were represented by 1, 0 and 3 COS markers in 1A, 1B and 1D, respectively. Presumably these COS markers were located more proximally on the long arm of W1 than the RFLP markers used by Zhang *et al.* (1998) and thus failed to detect the 6U-specific region.

In the present work, all group 2 COS markers were located on 2U, except for marker *c746642* in the terminal bin of W2L, which was located on 6U in agreement with Zhang *et al.* (1998). According to Gale and Devos (1998), W3 was homoeologous to 3U (represented by 8 RFLP markers) and 7U (based on two RFLP markers). We also detected most of the W3 markers (10 COS markers) on 3U. However, one marker specific for the terminal part of the short arm of W3 was located on 7U. According to Zhang *et al.* (1998) and Gale and Devos, 1998), the short arm of W4 was related to 6U, while the long arm to 4U and 5U. In this work, COS markers specific for the short arm of W4 were also located on 6U, while those specific for the intercalary bin of the long arm were assigned to 4U. However, in contrast to Zhang *et al.* (1998), we did not detect any W4 COS markers on chromosome 5U.

We detected COS markers from W5 on 5U, but unlike Gale and Devos (1998), we did not observe homoeology with 4U as the most distal part of the long arm of W5 was not represented by COS markers. W5 was also found to be closely related to chromosome 5M of *Ae. comosa*, while one marker suggested a relationship with 2M. A homoeology between W5 and $5M^g$ of *Ae. geniculata* was also observed by Tiwari *et al.* (2015) who showed that approximately 72% of the annotated $5M^g$ genes had sequence identity to wheat genes on chromosomes 5A, 5B and 5D. Chromosomes 5S and 5C were also found to be homoeologous with W5 in the present work, while three markers on the long arm of W5, were detected on 2C.

Homoeologous chromosome group 6, and chromosomes 6A and 6D in particular, have segmental homoeology to the short arm of *Ae. umbellulata* chromosome 6U, and long arms of 4U and 6U (Gale and Devos, 1998; Zhang *et al.*, 1998). In general, the present work confirmed the previous observations (three W6 COS markers were detected on each of 6U and 4U), but unlike the earlier results, five W6 markers suggested a relationship with 2U. Mapping the group 6 COS markers revealed significant homoeology of W6 chromosomes to chromosome 6M of *Ae. comosa*, and less pronounced homoeology to chromosome 6S of *Ae. speltoides.* On the other hand, W6 was related to 2C in *Ae. markgrafii.*

Gale and Devos (1998) noted that the short arm and a significant part of the long arm of W7 was homoeologous to 7U, the distal part of W7 long arm was related to 6U, while the terminal part was homoeologous to chromosome 4U of *Ae. umbellulata.* On line with these observations we detected three of the five W7 short arm markers, and nine of the thirteen W7 long arm markers on 7U, while three markers from the distal bins of W7 long arm were found on 6U. For the group 7 chromosomes, the wheat–*Aegilops* macrosynteny was highest in *Ae. speltoides*, and lower in *Ae. comosa*, while no synteny was found between W7 and the chromosome 7C of *Ae. markgrafii.*

We have detected previously unknown wheat-*Ae. umbellulata* genome relationships. For example, COS marker *c755444* specific for the proximal bin of the W3 long arm was assigned to 6U and W6 marker *c750237* was assigned to 5U. We detected such local breaks in the wheat–*Aegilops* genome relationships also in *Ae. comosa*, *Ae. speltoides* and *Ae. markgrafii.* These results are consistent with the observations of Dobrovolskaya *et al.* (2011) who observed local synteny perturbations between *Ae. speltoides* and wheat. However, 76 out of 90 markers mapped in *Ae. speltoides* were assigned to chromosomes homoeologous with wheat, confirming that the species is highly syntenic with wheat (The International Wheat Genome Sequencing Consortium (IWGSC), 2014).

According to Jaccard similarity coefficients estimated in this work, the S-genome of *Ae. speltoides* and the M-genome of *Ae. comosa* are structurally similar to the wheat genomes, while the U-genome of *Ae. umbellulata* and the C-genome of *Ae. markgrafii* in particular, are significantly different. These results are on line with previous phylogenetic studies in which *Ae. umbellulata* and *Ae. markgrafii* formed a closer sub-cluster on the *Aegilops-Triticum* clade, indicating greater genetic similarity, relative to *Ae. comosa* and *Ae. speltoides* (Petersen *et al.*, 2006; Mahelka *et al.*, 2011).

Evolutionary genome rearrangements in *Ae. markgrafii* relative to wheat as described in the present study indicate a need to rename four C-genome chromosomes. As twelve out of nineteen W6 COS markers identified homology between chromosomes 2C and W6 ($J_{W6,2C}$: 0.800), we suggest renaming chromosome 2C to 6C. Eleven out of fifteen markers indicated homology between 3C and W7 ($J_{W7,3C}$: 0.611), and thus we suggest renaming 3C to 7C. Similarly, five markers mapped to chromosome 7C were specific to W2 ($J_{W2,7C}$: 0.454), and five to W4, so chromosome 7C could be renamed 2C. Finally, out of

three markers identified on chromosome 6C, two were related to W7 and one to W3 indicating a need to rename 6C to 7C or 3C. However, we note that the low number of markers per chromosome allowed only macro-level comparisons and a more detailed comparative analysis is needed before changing the chromosome nomenclature of *Ae. markgrafii.* Sequencing DNA from flow-sorted U, M, S and C-genome chromosomes and comparison of their gene content with that of wheat chromosomes (The International Wheat Genome Sequencing Consortium (IWGSC), 2014) could provide detailed information about the synteny between *Aegilops* genomes and wheat.

This work represents an important step forward in developing chromosome genomics for wild relatives of wheat. FISH karyotypes will facilitate identification of *Aegilops* chromatin transferred to wheat. Bivariate flow karyotyping after FISHIS makes it possible to dissect the genomes of four important gene sources for cultivated wheat, *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides* and *Ae. markgrafii* into single chromosomes. This provides an opportunity for detailed characterization of their genomes, including gene content, allele discovery and targeted development of gene-based markers from specific genomic regions. The knowledge of homoeologous relationships between wheat and *Aegilops* species at chromosome level will be an important guide for targeted development of markers and for planning introgression breeding programs. COS markers assigned to chromosomes of the *Aegilops* species will be useful in pre-breeding programs to select chromosome segments carrying agronomically useful genes in *T. aestivum–Aegilops* recombinant lines. Altogether, these results promise to accelerate genomic studies on wild relatives of bread wheat and support pre-breeding studies that are required to meet the future challenges of food security and sustainable agriculture.

EXPERIMENTAL PROCEDURES

Plant material

Seeds of *Aegilops umbellulata* Zhuk. accession AE740/03 ($2n = 2x = 14$; UU) were kindly provided by the Institute of Plant Genetics and Crop Plant Research (Gatersleben, Germany). The accessions of *Ae. comosa* Sm. in Sibth. & Sm. MvGB1039 ($2n = 2x = 14$, MM), *Ae. speltoides* Tausch. MvGB905 ($2n = 2x = 14$, SS) and *Ae. markgrafii* (Greuter) Hammer MvGB428 and MvGB607 ($2n = 2x = 14$, CC) are maintained at the Martonvásár Cereal GenBank (Hungary). Wheat (*Triticum aestivum* L.) cv. Chinese Spring–*Ae. umbellulata* ditelosomic addition lines 2US, 2UL and 7UL (Friebe *et al.*, 1995) were kindly provided by Dr. Bernd Friebe (Wheat Genetics Resource Center, Kansas State University, USA). Accessions of *Secale cereale* L. cv. '*Petkus*', *Ae. tauschii* Coss. MvGB605, *Oryza sativa* L. cv. 'Bioryza' and *T. aestivum* L. cv. 'GK Öthalom' were also used in the present study and were obtained from the Cereal Research Non-Profit Company, Szeged, Hungary

Flow cytometric chromosome analysis and sorting

Suspensions of intact mitotic metaphase chromosomes were prepared from synchronized root tips of young seedlings following Vrána *et al.* (2000) and Kubaláková *et al.* (2005). The chromosome samples were fluorescently labelled by FISHIS using oligonucleotides 5′-FITC-GAA_7-FITC-3′ and/or 5′-FITC-ACG_7-FITC-3′ (Sigma-Aldrich, St. Louis, MO, USA) and counterstained by DAPI (4′,6-diamidino-2-phenylindole) as described by Giorgi *et al.* (2013). Bivariate flow karyotyping and chromosome sorting were done on a FACSAria II SORP flow cytometer and sorter (Becton Dickinson Immunocytometry Systems, San José, USA). Chromosome samples were analyzed at rates of 1500–2000 particles per second, and bivariate flow karyotypes FITC versus DAPI fluorescence were acquired. Sort windows were set on dotplots FITC versus DAPI, and chromosomes were sorted at rates of 15–20/sec. Flow-sorted chromosomes were identified and the purity in sorted chromosome fractions was determined according to Molnár *et al.* (2011b). Briefly, approximately one thousand chromosomes were sorted from each chromosome population identified on bivariate flow karyotype into a 15 μl drop of PRINS buffer supplemented with 5% (w/v) sucrose on a microscope slide (Kubaláková *et al.*, 1997). The slides were air-dried and used for FISH with probes for pSc119.2, pTa71 and Af*a* family repetitive DNA sequences.

Amplification of chromosomal DNA

Three batches of 30 000 chromosomes each were sorted from each chromosome population identified on bivariate flow karyotypes. The chromosomes were treated with proteinase K, after which their DNA was purified and amplified by multiple displacement amplification (MDA) using an Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare, Chalfont St. Giles, United Kingdom) as described by Šimková *et al.* (2008). Three independent MDA products from each sorted chromosome fraction were pooled into one sample to reduce amplification bias (Table S1) and used as template for PCR reaction with primers for COS markers.

Fluorescence *in situ* hybridization (FISH)

pSc119.2 and Afa family repeats were amplified from genomic DNA of *S. cereale* and *Ae. tauschii* and labelled with biotin-16-dUTP (Roche, Mannheim, Germany) and digoxigenin-11-dUTP (Roche), respectively, using PCR (Nagaki *et al.*, 1995; Contento *et al.*, 2005). 18S unit of 45S ribosomal RNA gene was amplified using PCR from genomic DNA of rice (Chang *et al.*, 2010) and labelled with 50% biotin-16-dUTP and 50% digoxigenin-11-dUTP. GAA and ACG microsatellites were amplified from genomic DNA of *T. aestivum* and labelled with digoxigenin-11-dUTP (Roche) and biotin-16-dUTP (Roche), respectively, using PCR. Digoxigenin and biotin were detected using anti-digoxigenin-rhodamine Fab fragments (Roche) and streptavidin–FITC (Roche), respectively.

FISH was performed on chromosomes flow-sorted onto microscopic slides and on slides prepared by squashing meristem root tips (Molnár *et al.*, 2011a). The pretreatment and stringent washing steps were omitted in experiments on flow-sorted chromosomes. Chromosome preparations were examined under a Zeiss AxioImager M2 fluorescence microscope system equipped with an AxioCam MRm CCD camera (Zeiss, Oberkochen, Germany), and the images were compiled with AxioVision v4.8 software (Zeiss) as described by Mikó *et al.* (2015). After capturing FISH signals on metaphase plates, the slides were washed and re-hybridized with GAA and ACG microsatellite probes at 42°C using the protocol described above.

COS marker analysis

Genomic DNA was prepared according to Cseh *et al.* (2013) from *Ae. umbellulata*, *Ae. comosa*, *Ae. speltoides* and *Ae. markgrafii* MvGB428 and MvGB607, which were also used for flow cytometric analyses, and from wheat cv. 'GK Öthalom'. PCR with primers for 123 COS markers (Quraishi *et al.*, 2009; Data S1) specific for wheat homoeologous groups I–VII, was performed in 12 μl reaction volumes as described by Molnár *et al.* (2014) using a touchdown reaction profile: 94°C (2 min); 10 cycles of 94°C (0.5 min), Ta +5°C (0.5 min) decreased in 0.5°C increments for every subsequent set of cycles, 72°C (1 min); 30 cycles of 94°C (0.5 min), Ta°C (0.5 min), 72°C (1 min); hold at 72°C (2 min). PCR products were separated using a Fragment Analyzer Automated CE System equipped with a 96-Capillary Array Cartridge (effective length 33 cm) (Advanced Analytical Technologies, Ames, USA) and analyzed with PROsize v2.0 software. The annealing temperature (Ta) for each COS marker, together with data on the PCR amplicons, are included in Data S1.

DNA sequence analysis

A deletion bin map was constructed for each wheat chromosome showing positions of the COS markers (Quraishi *et al.*, 2009). To order the markers along the chromosomes, EST sequences of the COS markers (Quraishi *et al.*, 2009, Supplementary Data S2) were used as queries in BLASTn searches to identify the scaffold containing the EST in the assembled chromosome survey sequences of hexaploid wheat (https://urgi.versailles.inra.fr/blast/blast.php; The International Wheat Genome Sequencing Consortium (IWGSC), 2014) Throughout the study, BLAST hits with E-values smaller than $2.8e^{-08}$, identity % >58.44 and alignment length >100 bp were considered significant (Data S2). The relative order and genetic distance (in cM) of the EST-specific scaffolds were obtained by searching the scaffold IDs in the GenomeZipper (v.5) of wheat chromosome arms (https://urgi.versailles.inra.fr/download/iwgsc/zipper/; The International Wheat Genome Sequencing Consortium (IWGSC), 2014) (Data S2).

Visualization of wheat–*Aegilops* orthologous relationships

In order to visualize wheat–*Aegilops* homoeologous relationships, a genetic map and physical deletion bin map of wheat were constructed showing positions of the mapped COS markers. Separate maps were drawn for the B, A, and D genomes of wheat (Figures 7, S5 and S6). The deletion bins were divided into as many parts as the number of COS markers located in the bins. The marker-specific bin parts were color-coded to show the homoeologous group location of the markers. For each homoeologous group (1–7), five wheat chromosome bin maps were displayed, one for a wheat genome (B or A and D) and one each for the *Aegilops* genomes U, M, S and C. This allowed to visualize the homoeologous group positions of the relevant wheat chromosome segments in the genomes of wheat and *Aegilops.* Moreover, a table was assembled showing the number of wheat homoeologous group-specific COS markers located on each of the *Aegilops* chromosome (Table 3). This highlighted wheat genomic regions related to a given chromosome in *Aegilops.*

Calculation of Jaccard similarity coefficients

Pairwise similarity between the structure of chromosomes within the same homoeologous groups of wheat and *Aegilops* species was determined using Jaccard's coefficient $J_{(i1,i2)} = a/(a + b + c)$ (Kosman and Leonard, 2005). For a given homoeologous group A, a = the number of markers present on group A chromosomes for both wheat and a corresponding *Aegilops* species; b = the number of markers where species i_1 (i.e. wheat) has a band on the group A chromosome, but i_2 (i.e. *Aegilops*) does not; c = the number of markers where the *Aegilops* species i_2 has a band on the group A chromosome, but i_1 (wheat) does not. Jaccard's coefficients were calculated for each homoeologous groups I–VII between wheat and each *Aegilops* species, and the similarity values are given in Table S6.

ACKNOWLEDGEMENTS

We thank Zdeňka Dubská and Romana Šperková for technical assistance with chromosome sorting. This work was funded by the Hungarian National Research Fund (K112226, K116277), by (TÁMOP-4.2.2.A-11-1-KONV-2012-0008), by a János Bolyai Research Scholarship from the Hungarian Academy of Sciences (for MI), by an OECD fellowship (TAD/CRP JA00079297), and by the Ministry of Education, Youth and Sports of the Czech Republic (grant LO1204 from the National Program of Sustainability I). AC gratefully acknowledges support from a Marie Curie Fellowship Grant (FP7-PEOPLE-2013-IEF-625671) under the seventh framework programme of the European Union. The authors declare that there is no conflict of interest.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

Figure S1. Identification of chromosomes flow-sorted from *Ae. comosa* using FISH.

Figure S2. Identification of chromosomes flow-sorted from *Ae. speltoides* using FISH.

Figure S3. Identification of chromosomes flow-sorted from *Ae. markgrafii* using FISH.

Figure S4. Identification of chromosome arms 2US, 2UL and 7UL flow-sorted from wheat-*Ae. umbellulata* ditelosomic addition lines using FISH.

Figure S5. Wheat–*Aegilops* orthologous relationships from the genomic perspective of A-genome chromosomes.

Figure S6. Wheat–*Aegilops* orthologous relationships from the genomic perspective of D-genome chromosomes.

Table S1. Karyotypic description of *Aegilops* chromosomes with probes pSc119.2, Afa family and 18S rDNA.

Table S2. Labelling efficiency of GAA and ACG repeats for *in situ* hybridisation on the chromosomes of *Aegilops*

Table S3. Chromosome assignment to populations on bivariate flow karyotypes of *Aegilops umbellulata*, *Ae. comosa*, *Ae. speltoides* and *Ae. markgrafii.*

Table S4. DNA yields after the multiple displacement amplification of DNA from flow-sorted chromosome fractions.

Table S5. The number of COS marker loci assigned to U, M, S and C-genome-chromosomes of *Aegilops* species.

Table S6. Jaccard similarity coefficients (J) calculated between the same homoeologous group chromosomes in wheat and *Aegilops* species.

Data S1. PCR products of COS markers amplified from wheat and *Aegilops* species.

Data S2. BLASTn search results and Genome Zipper data used for ordering COS markers on wheat chromosomes.

REFERENCES

Badaeva, E.D., Friebe, B. and Gill, B.S. (1996a) Genome differentiation in *Aegilops*. 1. Distribution of highly repetitive DNA sequences on chromosomes of diploid species. *Genome*, **39**, 293–306.

Badaeva, E.D., Friebe, B. and Gill, B.S. (1996b) Genome differentiation in *Aegilops*. 2. Physical mapping of 5S and 18S-26S ribosomal RNA gene families in diploid species. *Genome*, **39**, 1150–1158.

Bartoš, J., Paux, E., Kofler, R. *et al.* (2008) A first survey of the rye (*Secale cereale*) genome composition through BAC end sequencing of the short arm of chromosome 1R. *BMC Plant Biol.* **8**, 95–106.

Burt, C. and Nicholson, P. (2011) Exploiting co-linearity among grass species to map the *Aegilops ventricosa*-derived *Pch1* eyespot resistance in wheat and establish its relationship to *Pch2*. *Theor. Appl. Genet.* **123**, 1387–1400.

Cápal, P., Blavet, N., Vrána, J., Kubaláková, M. and Doležel, J. (2015) Multiple displacement amplification of the DNA from single flow-sorted plant chromosome. *Plant J.* **84**, 838–844.

Ceoloni, C. and Jauhar, P.P. (2006) Chromosome engineering of the durum wheat genome: strategies and applications of potential breeding value. In *Genetic Resources, Chromosome Engineering, and Crop Improvement, 2: Cereals* (Singh, R.J. and Jauhar, P.P. eds). Boca Raton: CRC Press, pp. 27–59.

Ceoloni, C., Del Signore, G., Pasquini, M. and Testa, A. (1988) Transfer of mildew resistance from *Triticum longissimum* into wheat by *ph1*-induced homoeologous recombination. In *7th International Wheat Genetics Symposium*. (Miller, T. E. and Koebner, R. M. D., eds). Cambridge, UK: Institute of Plant Science Research, pp. 221–226.

Chang, K.D., Fang, S.A., Chang, F.C. and Chung, M.C. (2010) Chromosomal conservation and sequence diversity of ribosomal RNA genes of two distant *Oryza* species. *Genomics*, **96**, 181–190.

Contento, A., Heslop-Harrison, J.S. and Schwarzacher, T. (2005) Diversity of a major repetitive DNA sequence in diploid and polyploid Triticeae. *Cytogenet. Genome Res.* **109**, 34–42.

Cseh, A., Soós, V., Rakszegi, M., Türkösi, E., Balázs, E. and Molnár-Láng, M. (2013) Expression of *HvCslF9* and *HvCslF6* barley genes in the genetic background of wheat and their influence on the wheat β-glucan content. *Ann. Appl. Biol.* **163**, 142–150.

Cuadrado, A., Vitellozzi, F., Jouve, N. and Ceoloni, C. (1997) Fluorescence in situ hybridization with multiple repeated DNA probes applied to the analysis of wheat-rye chromosome pairing. *Theor. Appl. Genet.* **94**, 347–355.

Cuadrado, A., Cardoso, M. and Jouve, N. (2008) Physical organisation of simple sequence repeats (SSRs) in Triticeae: structural, functional and evolutionary implications. *Cytogenet. Genome. Res.* **120**, 210–219.

Danilova, T.V., Friebe, B. and Gill, B.S. (2014) Development of a wheat single gene FISH map for analyzing homoeologous relationship and chromosomal rearrangements within the Triticeae. *Theor. Appl. Genet.* **127**, 715–730.

Devos, K.M. and Gale, M.D. (2000) Genome relationships: the grass model in current research. *Plant Cell*, **12**, 637–646.

Devos, K.M., Atkinson, M.D., Chinoy, C.N., Francis, H.A., Harcourt, R.L., Koebner, R.M.D., Liu, C.J., Masojć, P., Xie, D.X. and Gale, M.D. (1993) Chromosomal rearrangements in the rye genome relative to that of wheat. *Theor. Appl. Genet.* **85**, 673–680.

Dobrovolskaya, O., Boeuf, C., Salse, J., Pont, C., Sourdille, P., Bernard, M. and Salina, E. (2011) Microsatellite mapping of *Ae. speltoides* and map-based comparative analysis of the S, G, and B genomes of Triticeae species. *Theor. Appl. Genet.* **123**, 1145–1157.

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

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

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

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

Dulai, S., Molnár, I., Szopkó, D., Darkó, É., Vojtkó, A., Sass-Gyarmati, A. and Molnár-Láng, M. (2014) Wheat-*Aegilops biuncialis* amphiploids have efficient photosynthesis and biomass production during osmotic stress. *J. Plant Physiol.* **171**, 509–517.

Dvorak, J., Luo, M.C. and Yang, Z.L. (1998) Restriction fragment length polymorphism and divergence in the genomic regions of high and low recombination in self-fertilizing and cross-fertilizing *Aegilops* species. *Genetics*, **148**, 423–434.

Farkas, A., Molnár, I., Dulai, S., Rapi, S., Oldal, V., Cseh, A., Kruppa, K. and Molnár-Láng, M. (2014) Increased micronutrient content (Zn, Mn) in the 3 M^b(4B) wheat - *Aegilops biuncialis* substitution and 3 M^b.4BS translocation identified by GISH and FISH. *Genome*, **57**, 61–67.

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

Friebe, B., Schubert, V., Bluthner, W.D. and Hammer, K. (1992) C-banding pattern and polymorphism of *Aegilops caudata* and chromosomal constitutions of the amphiploid *T. aestivum–Ae. caudata* and six derived chromosome additional lines. *Theor. Appl. Genet.* **83**, 589–596.

Friebe, B., Jiang, J., Tuleen, N. and Gill, B.S. (1995) Standard karyotype of *Triticum umbellulatum* and the characterization of derived chromosome addition and translocation lines in common wheat. *Theor. Appl. Genet.* **90**, 150–156.

Friebe, B., Jiang, J., Raupp, W.J., McIntosh, R.A. and Gill, B.S. (1996) Characterization of wheat alien translocations conferring resistance to diseases and pests: current status. *Euphytica*, **71**, 59–87.

Gale, M.D. and Devos, K.M. (1998) Comparative genetics in the grasses. *Proc. Natl Acad. Sci. USA* **95**, 1971–1974.

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.

Jiang, J., Friebe, B. and Gill, B.S. (1994) Recent advances in alien gene transfer in wheat. *Euphytica*, **73**, 199–212.

Kilian, B., Mammen, K., Millet, E., Sharma, R., Graner, A., Salamini, F., Hammer, K. and Özkan, H. (2011) *Aegilops*. In *Wild Crop Relatives: Genomic and Breeding Resources*. (Kole, C., ed). Heidelberg: Dordrecht, London, New York, Springer-Verlag, pp. 1–76.

Kosman, E. and Leonard, K.J. (2005) Similarity coefficients for molecular markers in studies of genetic relationships between individuals for haploid, diploid, and polyploid species. *Mol. Ecol.* **14**, 415–442.

Kozub, N.A., Sozinov, I.A., Xynias, I.N. and Sozinov, A.A. (2011) Allelic variation at high-molecular-weight glutenin subunit loci in *Aegilops biuncialis* Vis. *Russ. J. Genet.* **47**, 1078–1083.

Kubaláková, M., Macas, J. and Doležel, J. (1997) Mapping of repeated DNA sequences in plant chromosomes by PRINS and C-PRINS. *Theor. Appl. Genet.* **94**, 758–763.

Kubaláková, M., Kovářová, P., Suchánková, P., Číhalíková, J., Bartoš, J., Lucretti, S., Watanabe, N., Kianian, S.F. and Doležel, J. (2005) Chromosome sorting in tetraploid wheat and its potential for genome analysis. *Genetics*, **170**, 823–829.

Le, H.T., Armstrong, K.C. and Miki, B. (1989) Detection of rye DNA in wheat-rye hybrids and wheat translocation stocks using total genomic DNA as a probe. *Plant Mol. Biol. Rep.* **7**, 150–158.

Liu, Z., Yan, Z., Wan, Y., Liu, K., Zheng, Y. and Wang, D. (2003) Analysis of HMW glutenin subunits and their coding sequences in two diploid *Aegilops* species. *Theor. Appl. Genet.* **106**, 1368–1378.

Lucretti, S., Giorgi, D., Farina, A. and Grosso, V. (2014) FISHIS: A new way in chromosome flow sorting makes complex genomes more accessible. In *Genomics of Plant Genetic Resources* (Tuberosa, R., Graner, A. and Frison, E. eds). Dordrecht: Springer Science+Business Media, pp. 319–348.

Lukaszewski, A.J., Rybka, K., Korzun, V., Malyshev, S.V., Lapinski, B. and Whitkus, R. (2004) Genetic and physical mapping of homoeologous recombination points involving wheat chromosome 2B and rye chromosome 2R. *Genome*, **47**, 36–45.

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

Mahelka, V., Kopecký, D. and Paštová, L. (2011) On the genome constitution and evolution of intermediate wheatgrass (*Thinopyrum intermedium:* Poaceae, Triticeae). *BMC Evol. Biol.* **11**, 127.

Martis, M.M., Zhou, R., Haseneyer, G. *et al.* (2013) Reticulate evolution of the rye genome. *Plant Cell*, **25**, 3685–3698.

Mayer, K.F.X., Martis, M., Hedley, P.E. *et al.* (2011) Unlocking the barley genome by chromosomal and comparative genomics. *Plant Cell*, **23**, 1249–1263.

Mikó, P., Megyeri, M., Farkas, A., Molnár, I. and Molnár-Láng, M. (2015) Molecular cytogenetic identification and phenotypic description of a new synthetic amphiploid, *Triticum timococcum* ($A^tA^tGGA^mA^m$). *Genet. Res. Crop Evol.* **62**, 55–66.

Molnár, I., Gáspár, L., Sárvári, É., Dulai, S., Hoffmann, B., Molnár-Láng, M. and Galiba, G. (2004) Physiological and morphological responses to water stress in *Aegilops biuncialis* and *Triticum aestivum* genotypes with differing tolerance to drought. *Funct. Plant Biol.* **31**, 1149–1159.

Molnár, I., Cifuentes, M., Schneider, A., Benavente, E. and Molnár-Láng, M. (2011a) Association between SSR-rich chromosome regions and intergenomic translocation breakpoints in natural populations of allopolyploid wild wheats. *Ann. Bot.* **107**, 65–76.

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

Molnár, I., Šimková, H., Leverington-Waite, M., Goram, R., Cseh, A., Vrána, J., Farkas, A., Doležel, J., Molnár-Láng, M. and Griffiths, S. (2013) Syntenic relationships between the U and M genomes of *Aegilops*, wheat and the model species *Brachypodium* and rice as revealed by COS markers. *PLoS ONE*, **8**, e70844.

Molnár, I., Kubaláková, M., Šimková, H., Farkas, A., Cseh, A., Megyeri, M., Vrána, J., Molnár-Láng, M. and Doležel, J. (2014) Flow cytometric chromosome sorting from diploid progenitors of bread wheat, *T. urartu*, *Ae. speltoides* and *Ae. tauschii*. *Theor. Appl. Genet.* **127**, 1091–1104.

Molnár, I., Vrána, J., Farkas, A., Kubaláková, M., Cseh, A., Molnár-Láng, M. and Doležel, J. (2015) Flow sorting of C-genome chromosomes from wild relatives of wheat *Aegilops markgrafii*, *Ae. triuncialis* and *Ae. cylindrica*, and their molecular organization. *Ann. Bot.* **116**, 189–200.

Nagaki, K., Tsujimoto, H., Isono, K. and Sasakuma, T. (1995) Molecular characterization of a tandem repeat, Afa family, and its distribution among *Triticeae*. *Genome*, **38**, 479–486.

Nagy, E.D., Molnár, I., Schneider, A., Kovács, G. and Molnár-Láng, M. (2006) Characterisation of chromosome-specific S-SAP markers and their use to study genetic diversity in *Aegilops* species. *Genome*, **49**, 289–296.

Peil, A., Korzun, V., Schubert, V., Schumann, E., Weber, W.E. and Röder, M.S. (1998) The application of wheat microsatellites to identify disomic *Triticum aestivum–Aegilops markgrafii* addition lines. *Theor. Appl. Genet.* **96**, 138–146.

Peleg, Z., Saranga, Y., Suprunova, T., Ronin, Y., Röder, M.S., Kilian, A., Korol, A.B. and Fahima, T. (2008) High density genetic map of durum wheat x wild emmer wheat based on SSR and DArT markers. *Theor. Appl. Genet.* **117**, 103–115.

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

Potz, H., Schubert, V., Houben, A., Schubert, I. and Weber, W.E. (1996) Aneuploids as a key for new molecular cloning strategies: development of DNA markers by microdissection using *Triticum aestivum–Aegilops markgrafii* chromosome addition line B. *Euphytica*, **89**, 41–47.

Quraishi, U.M., Abrouk, M., Bolot, S. *et al.* (2009) Genomics in cereals: from genome-wide conserved orthologous set (COS) sequences to candidate genes for trait dissection. *Funct. Integr. Genom.* **9**, 473–484.

Rayburn, A.L. and Gill, B.S. (1985) Repeated DNA sequences in *Triticum* (*Poaceae*): chromosomal mapping and its bearing on the evolution of B and G genomes. *Plant Syst. Evol.* **159**, 229–235.

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

Riar, A.K., Kaur, S., Dhaliwal, H.S., Singh, K. and Chhuneja, P. (2012) Introgression of a leaf rust resistance gene from *Aegilops caudata* to bread wheat. *J. Genet.* **91**, 155–161.

Riley, R. and Chapman, V. (1958) Genetic control of the cytologically diploid behaviour of hexaploid wheat. *Nature* **182**, 713–715.

Šafář, J., Bartoš, J., Janda, J. *et al.* (2004) Dissecting large and complex genomes: flow sorting and BAC cloning of individual chromosomes from bread wheat. *Plant J.* **39**, 960–968.

Schneider, A., Linc, G., Molnár, I. and Molnár-Láng, M. (2005) Molecular cytogenetic characterization of *Aegilops biuncialis* and its use for the identification of 5 derived wheat–*Aegilops biuncialis* addition lines. *Genome*, **48**, 1070–1082.

Schneider, A., Molnár, I. and Molnár-Láng, M. (2008) Utilisation of *Aegilops* (goatgrass) species to widen the genetic diversity of cultivated wheat. *Euphytica*, **163**, 1–19.

Schneider, A., Molnár, I. and Molnár-Láng, M. (2010) Selection of U and M genome-specific wheat SSR markers using wheat–*Aegilops biuncialis* and wheat–*Ae. geniculata* addition lines. *Euphytica*, **175**, 357–364.

Schwarzacher, T. and Heslop-Harrison, J.S. (2000) *Practical in situ Hybridization*. Bios: Oxford. pp. 203.

Schwarzacher, T., Leitch, A.R., Bennett, M.D. and Heslop-Harrison, J.S. (1989) In situ localization of parental genomes in a wide hybrid. *Ann. Bot.* **64**, 315–324.

Sears, E.R. (1977) An induced mutant with homoeologous pairing in common wheat. *Can. J. Gen. Cytol.* **19**, 585–593.

Shangguan, L., Han, J., Kayesh, E., Sun, X., Zhang, C., Pervaiz, T., Wen, X. and Fang, J. (2013) Evaluation of genome sequencing quality in selected plant species using expressed sequence tags. *PLoS ONE*, **8**, e69890.

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

Staňková, H., Hastie, A.R., Chan, S. *et al.* (2016) BioNano genome mapping of individual chromosomes supports physical mapping and sequence assembly in complex plant genomes. *Plant Biotechnol. J.* **14**, 1523–1531.

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

Tiwari, V.K., Wang, S., Sehgal, S. *et al.* (2014) SNP discovery for mapping alien introgressions in wheat. *BMC Genom.*, **15**, 273.

Tiwari, V.K., Wang, S., Danilova, T. *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.

Valárik, M., Bartoš, J., Kovářová, P., Kubaláková, M., de Jong, H. and Doležel, J. (2004) High-resolution FISH on super-stretched flow-sorted plant chromosomes. *Plant J.* **37**, 940–950.

Van Slageren, M.W. (1994) *Wild Wheats: A Monograph of Aegilops L. and Amblyopyrum (Jaub. & Spach) Eig (Poaceae)*. Wageningen, the Netherlands: Agricultural University, Aleppo, Syria: International Center for Agricultural Research in Dry Areas.

Vitulo, N., Albiero, A., Forcato, C. *et al.* (2011) First survey of the wheat chromosome 5A composition through a Next Generation Sequencing approach. *PLoS ONE*, **6**, e26421.

Vrána, J., Kubaláková, M., Šimková, 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.

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

Winfield, M.O., Allen, A.M., Burridge, A.J. *et al.* (2016) High-density SNP genotyping array for hexaploid wheat and its secondary and tertiary gene pool. *Plant Biotechnol. J.* **14**, 1195–1206.

Zhang, H., Jia, J., Gale, M.D. and Devos, K.M. (1998) Relationships between the chromosomes of *Aegilops umbellulata* and wheat. *Theor. Appl. Genet.* **96**, 69–75.

APPENDIX V

Genomic constitutions of cereals with blue aleurone trait

Burešová V, Kopecký D, Šafář J, Vyhnánek T, Martinek P, Doležel J

In: Abstracts of the "Olomouc Biotech 2011. Plant Biology: Green for Good II".
Olomouc, Czech Republic, 2013.

GENOMIC CONSTITUTION OF CEREALS WITH BLUE ALEURONE TRAIT

Veronika Burešová[1], David Kopecký[1], Jan Šafář[1], Tomáš Vyhnánek[2], Petr Martinek[3], Jaroslav Doležel[1]

[1]) Institute of Experimental Botany, Centre of Region Haná for Biotechnological and Agronomical Research, Šlechtitelů 31, 783 71 Olomouc, Czech Republic
[2]) Mendel University in Brno, Zemědělská 1, 613 00 Brno, Czech Republic
[3]) Agrotest Fyto, Ltd., Havlíčkova 2787, 767 01 Kroměříž, Czech Republic

Email: kopecky@ueb.cas.cz

BACKGROUND

Anthocyanins are of great importance for human health due to their antioxidant potential. Their content is rather low in common varieties of wheat (*Triticum aestivum* L.). However, elite lines with blue aleurone and introgressed chromatin from wild relatives exhibit significantly increased levels of anthocyanins. There is evidence that the donor of chromosome introgressions has been *Thinopyrum ponticum* (syn. *Agropyron elongatum*). The aim of our study was to characterize genomic constitution of selected wheat genotypes with blue aleurone.

PLANT MATERIAL

In our study we used fourteen genotypes of *Triticum aestivum* L. with blue aleurone (Fig. 1). They were analyzed by genomic *in situ* hybridization (GISH). This specific colouring is caused by presence of antocyanins in aleurone layer of caryopsis (Fig. 2). Next we used an uncultivated species of grass *Th. Ponticum*, which is supposed donor of the blue colouring.

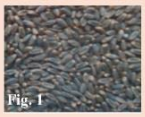


Fig. 1

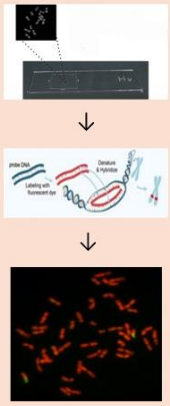

Fig. 2

METHODS

Young root tips were synchronized by iced water, fixed in 3:1 Carnoy I solution in 37°C for 7 days and squashed in 45% acetic acid. Slides with good quality were used for analysis.

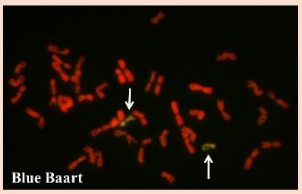
Genomic *in situ* hybridization (GISH) was performed using labelled genomic DNA of *Thinopyrum ponticum* as a probe. The chromosomes were counterstained with DAPI.

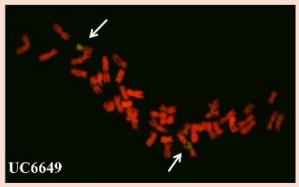
All observations were made on Zeiss Axio Imager Z2 fluorescence microscope equipped by CCD camera.

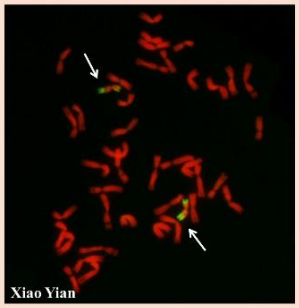


RESULTS

Our results revealed large variation in genomic constitution of blue aleurone wheat genotypes. Out of 14 analyzed lines, 11 lines carried an introgression from *Th. ponticum*. In the remaining three genotypes (cv. Indigo, Skorpion, and line H83-952-1), we were unable to detect any introgressed chromosome segment. Six different types of introgressions were found, ranging from the addition of a telocentric chromosome pair (cv. Blue Norco) to substitution of one chromosome pair (cv. Blue Baart), substitution of complete (homologous) chromosome arms (line UC660-49) and various substitutions of distal parts of chromosome arm(s).


Blue Baart


UC6649


Xiao Yian

CONCLUSIONS AND FUTURE WORK

Different types of introgressions observed in our work support a hypothesis that the introgressions activate the blue aleurone trait pathway, which is present, but deactivated in common wheat germplasms. In future, we plan to identify wheat chromosome carrying the alien introgression in the current set of wheat with blue aleurone. For fluorescence *in situ* hybridization (FISH) we would like to use total genomic DNA of *Th. ponticum* and repetitive DNA sequences (GAA repeat, *Afa* family) enabling the identification of individual wheat chromosomes. Furthermore, we plan to analyse other grains with blue aleurone trait (barley and Triticale).

This work was supported by grants P501/11/0504 and CZ.1.05/2.1.00/01.0007 and IGA PrF/2012/001

APPENDIX VI

Genomic constitution of wheat genotypes with blue aleurone

Burešová V, Kopecký D, Bartoš J, Martinek P, Watanabe N, Vyhnánek T, Doležel J

In: Abstracts of the "Plant molecular cytogenetics in genomic and postgenomic era". Katowice, Poland, 2014.

Genomic constitution of blue grained wheat genotypes

Burešová V, Kopecký D, Bartoš J, Martinek P, Watanabe N, Vyhnánek T, Doležel J

In: Sborník abstrakt, Bulletin České společnosti experimentální biologie rostlin, "6. Metodické dny". Seč, Czech Republic, 2014.

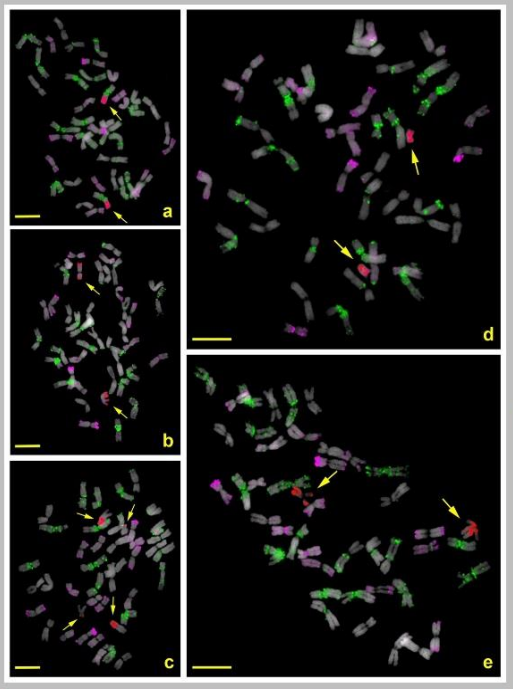
GENOMIC CONSTITUTION OF WHEAT GENOTYPES WITH BLUE ALEURONE

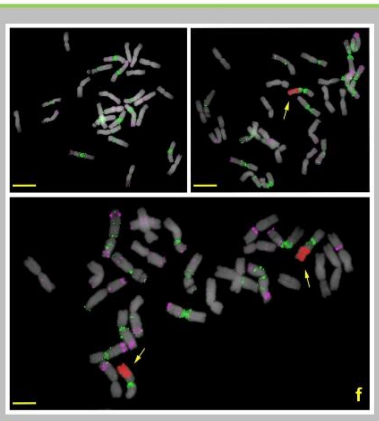
Veronika Burešová[1], David Kopecký[1], Jan Bartoš[1], Petr Martinek[2], Nobuyoshi Watanabe[3], Tomáš Vyhnánek [4], Jaroslav Doležel[1]

[1]) Institute of Experimental Botany, Centre of Region Haná for Biotechnological and Agronomical Research, Šlechtitelů 31, 783 71, Olomouc, Czech Republic, buresova@ueb.cas.cz
[2]) Agrotest Fyto, Ltd., Havlíčkova 2787, 767 01 Kroměříž, Czech Republic
[3]) College of Agriculture, Ibaraki University, 3-21-1 Chuo, Ami, Inashiki, Ibaraki 300-0393, Japan
[4]) Mendel University of Agriculture and Forestry in Brno, Zemědělská 1, 613 00 Brno, Czech Republic

BACKGROUND:

Anthocyanins belong to the most widely used natural pigments and determine red, purple and blue colouring in many plant species. Recently, the interest in such substances is increasing and it is not only for natural colouring abilities but also for its beneficial properties for human health. Clinical studies revealed many significant positive effects of anthocyanins such as anti-cancerogenic, anti-microbial and anti-inflammatory effects. Among others, they have been identified also in cereals. High content of these pigments has been revealed in wheat with blue aleurone layer. The blue colorization has been observed after the introgression of chromatin from wild relatives into wheat. At least three different donors have been identified including *Thinopyrum ponticum, Triticum monococcum* and *Th. bessarabicum*. The aim of our study was to characterize genomic constitutions of blue grained wheat using multicolour fluorescence *in situ* hybridization.





- GAA
- genomic DNA of *Th. ponticum*
- *Afa* family
- Bars, 10μm

RESULTS:

Our analysis revealed large variation in genomic constitution of blue grained wheat genotypes. Of 26 examined lines, 17 carried an introgression from *Th. ponticum*. We detected at least five different types of introgressions: disomic substitution of chromosome arm (cv. **M90-41**, **fig. a**) and chromosome segments (cv. **Xiao Yan**, **fig. b**) and two pairs of translocated chromosomes involving wheat and *Th. ponticum* chromatin (**H90-15-2**, **fig. c**) and the addition of the entire pair of chromosomes (cvs. **Blue Norco**, **fig. d**, and **Blue Baart**, **fig. e**), *Thinopyrum* substitutions were located on the wheat chromosomes of homoeologous group 4. In case of the remaining nine lines (i.e. cvs. Skorpion, Barevná 9), we were unable to detect introgressed chromatin of *Th. ponticum*, which indicates different source of blue aleurone trait (presumably *T. monococcum*). Two lines were influenced by dosage effect manifested with significant differences in colour of seeds (**Blue Norco**, **fig. g**), where two copies of the segment resulted in blue coloring, one copy was present in light blue seed and standard (red) coloring were in seeds without any *Thinopyrum* chromatin (**M90-99-2**, **fig. f**).

CONCLUSIONS AND FUTURE WORK:

There was detected a large variation of *Th. ponticum* introgressions in wheat with blue aleurone trait. We suppose that introgressions of wild relatives such as *Th. ponticum* or *Triticum monococcum* into the common lines of wheat activates the blue aleurone trait pathway, which is present, but deactivated in wheat germplasm. Our next goal is to flow sort individual chromosomes with introgressions using FISHIS (FISH in suspension) and to identificate the origin of *Th. ponticum* introgressions in various lines of blue grained wheat.

This work has been supported by the European Regional Development Fund (Operational Programme Research and Development for Innovations No. CZ.1.05/2.1.00/01.0007).

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

Veronika Burešová

Study of organization of hybrid and polyploid genomes

P1527 Biology - Botany

Summary of Ph.D. Thesis

Olomouc 2017

Ph.D. thesis was carried out at the Department of Botany, Faculty of Science, Palacký University Olomouc, between the years 2012-2017.

Ph.D. candidate: **Mgr. Veronika Burešová**

Supervisor: **RNDr. David Kopecký, Ph.D.**
Institute of Experimental Botany, Olomouc

Reviewers: **Doc. Mgr. Martin Lysák, Ph.D.**
Central European Institute of Technology, Brno

Ing. Pavel Neumann, Ph.D.
Institute of Plant Molecular Biology, České Budějovice

Doc. RNDr. Vladan Ondřej, Ph.D.
Palacký University, Olomouc

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

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

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

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

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

CONTENT

1 INTRODUCTION 5

2 AIMS OF THE THESIS 7

3 MATERIAL AND METHODS 8

3.1 Cytogenetic analysis of introgression lines of wheat-*Th. ponticum* with blue aleurone trait using multicolour fluorescence *in situ* hybridization (FISH) 8

3.2 Optimization and utilization of fluorescence *in situ* hybridization in suspension (FISHIS) for flow cytometric analysis of selected species of the *Triticeae* tribe 9

4 SUMMARY OF RESULTS 10

4.1 Cytogenetic analysis of introgression lines of wheat-Th. *ponticum* with blue aleurone trait using multicolour fluorescence *in situ* hybridization (FISH) 10

4.2 Optimization and utilization of fluorescence *in situ* hybridization in suspension (FISH IS) for flow cytometric analysis of selected species of the *Triticeae* tribe 10

5 CONCLUSION 11

6 REFERENCES 12

7 LIST OF AUTHOR´S PUBLICATIONS 14

7.1 Original papers 14

7.2 Published abstracts – poster presentation..... 14

7.3 Published abstracts – oral presentation..... 15

8 SUMMARY (IN CZECH) 16

1 INTRODUCTION

Understanding structure, organization and function of eukaryotic genomes provides deeper view into the evolutionary history of all living organisms (Leliaert *et al*. 2012). Plant genomes are in general much more complex than genomes of other eukaryotes and display extensive variation in genome size, chromosome number, ploidy level, arrangement of chromatin and number of genes (Heslop-Harrison 2000; Alberts 2002; Kellogg and Bennetzen 2004). Wide diversity of plant genomes is a consequence of highly active processes constantly influencing the development of species (Madlung 2013).

It turned out that many crop species are allopolyploids, resulting from events of interspecific hybridization and polyploidy (Renny-Byfield and Wendel 2014). Hybridization between individuals can lead to interspecific gene flow and generate new hybrid species which can exhibit higher genetic diversity than their parents (Barton 1979; Wallace *et al*. 2011). Nevertheless, there are a number of factors preventing exchange of genetic material among species in nature and the gene-flow is therefore restricted (Urbanelli 2002). In artificial conditions, it is possible to overcome such barriers and utilize interspecific hybridization for targeted breeding.

Bread wheat (*Triticum aestitvum* L.) is one of the most important crops providing staple food for ~30% of the world population (Choulet *et al*. 2014). It is an allohexaploid species originating from two interspecific hybridization events that gave rise to a highly redundant 17-Gb genome with three homoeologous sets of chromosomes (Feldman and Levy 2012; Choulet *et al*. 2014). The long term domestication led to the significant decreasing of genetic diversity of this important crop (e.g., Molnár-Láng *et al.* 2014; Choulet *et al.* 2014; Rey *et al*. 2015; Zhang *et al*. 2015). Fortunately, numerous agriculturally important genes have been incorporated from wild relatives into wheat genome by introgression breeding. These introgressions originate from various genera of the *Triticeae* tribe such as *Secale*, *Hordeum*, *Aegilops*, *Thinopyrum*, or *Triticum* (Molnár-Láng *et al.* 2014, Molnár-Láng *et al*. 2015).

For successful introgression breeding, it is necessary to confirm hybridity of the plants originating from distant crosses (Liu *et al*. 2014). During the twentieth century, high number of chemical, genetic and molecular methods have been developed and successfully applied for identification of alien introgression(s). Flow cytometry and fluorescent *in situ* hybridization (FISH) are suitable techniques for characterization of interspecific hybrids and allopolyploids (Molnár-Láng *et al*. 2015). In addition, introduction of new modifications of these methods makes the analysis of complex plant genomes more efficient. For example,

development of technique called FISHIS (fluorescence *in situ* hybridization in suspension) enables combination of flow sorting and FISH. This allows sorting of individual chromosomes based on the fluorescent signal of oligonucleotide SSR probe(s) and DAPI staining (Giorgi *et al*. 2013; Lucretti *et al*. 2014).

2 AIMS OF THE THESIS

I **Cytogenetic analysis of introgression lines of wheat-*Th. ponticum* with blue aleurone trait using multicolour fluorescence *in situ* hybridization (FISH)**

II **Optimization and utilization of fluorescence *in situ* hybridization in suspension (FISHIS) for flow cytometric analysis of selected species of the *Triticeae* tribe**

3 MATERIAL AND METHODS

3.1 Cytogenetic analysis of introgression lines of wheat-*Th. ponticum* with blue aleurone trait using multicolour fluorescence *in situ* hybridization (FISH)

Plant material

Seed samples of blue grain wheat genotypes were obtained from Prof. A. J. Lukaszewski (University of California, Riverside, USA), Prof. C.O. Qualset (University of California, Davis, USA), Prof. F.J. Zeller (Technical University of Munchen, Freising-Weihenstephan, Germany), Dr. Robert Metzger (Oregon State University, Corvallis, Oregon, USA), Prof. A. Borner (Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany) and from Genebank of the Crop Research Institute, Prague-Ruzyně, Czech Republic.

Fluorescent *in situ* hybridization

Seeds were germinated on wet filter paper in Petri dishes, root tips were collected in ice water for 26–30 h and fixed in a mixture of absolute alcohol:glacial acetic acid (3:1) at 37°C for 7 days. Cytological preparations and *in situ* hybridization with labelled DNA were made according to Masoudi-Nejad *et al*. (2002). In all experiments, genomic *in situ* hybridization (GISH) was done with a probe prepared from total genomic DNA of *Th. ponticum*. The probe was labelled with biotin by nick translation and detected with streptavidin-Cy3 using standard kits from Roche Applied Science following the manufacturer's instructions. The hybridization mix contained unlabelled genomic DNA of *T. aestivum* cv. Chinese Spring sheared to ca. 200–500 bp fragments at 1:150 ratio (probe:blocking DNA). Following the hybridization, preparations were counterstained with 4′,6-diamidino-2-phenylindole (DAPI) in VectaShield antifade (Vector Laboratories) and observed under Zeiss Axio Imager.Z2 microscope. For identification of individual chromosomes, two additional probes were employed: A digoxigenin-labelled probe for GAA microsatellites, prepared using PCR with (GAA)7 and (CCT)7 primers and wheat genomic DNA as a template, and a probe for a 260-bp fragment of the *Afa* family repeat, prepared and labelled by Texas Red using PCR with primers AS-A and AS-B on wheat genomic DNA according to Kubaláková *et al*. (2005).

3.2 Optimization and utilization of fluorescence *in situ* hybridization in suspension (FISHIS) for flow cytometric analysis of selected species of the *Triticeae* tribe

Plant material

Seeds of *Triticum dicoccoides* accession 26676 were provided by Dr. E. Paux (INRA, France). Seeds of *Aegilops umbellulata* Zhuk. accession AE740/03 (2n=2x=14; UU) were provided by the Institute of Plant Genetics and Crop Plant Research (Gatersleben, Germany). The accessions of *Ae. comosa* Sm. in Sibth. & Sm. MvGB1039 (2n=2x=14, MM), *Ae. speltoides* Tausch. MvGB905 (2n=2x=14, SS) and *Ae. markgrafii* (Greuter) Hammer MvGB428 and MvGB607 (2n=2x=14, CC) are maintained at the Martonvásár Cereal Genebank (Hungary). Wheat (*Triticum aestivum* L.) cv. Chinese Spring-*Ae. umbellulata* ditelosomic addition lines 2US, 2UL and 7UL (Friebe *et al*., 1995) were kindly provided by Dr. Bernd Friebe (Wheat Genetics Resource Center, Kansas State University, USA). Accessions of *Secale cereale* L. cv. *'Petkus', Ae. tauschii* Coss. MvGB605, *Oryza sativa* L. cv. 'Bioryza' and *T. aestivum* L. cv. 'GK Öthalom' were also used in the present study and were obtained from the Cereal Research Non-Profit Company, Szeged, Hungary.

Flow cytometry chromosome analysis and sorting with utilization of fluorescence *in situ* hybridization in suspension (FISHIS)

Suspensions of intact mitotic metaphase chromosomes were prepared from synchronized root tips of young seedlings following Vrána *et al*. (2000) and Kubaláková *et al*. (2005). The chromosome samples were fluorescently labelled by FISHIS using oligonucleotides 5'-FITC-GAA7-FITC-3' and/or 5'-FITC-ACG7-FITC-3' (Sigma) and counterstained by DAPI (4´,6-diamidino 2-phenylindole) as described by Giorgi *et al*. (2013). Bivariate flow karyotyping and chromosome sorting were done on a FACSAria II SORP flow cytometer and sorter (Becton Dickinson Immunocytometry Systems, San José, USA). Chromosome samples were analysed at rates of 1500–2000 particles per second, and bivariate flow karyotypes FITC vs. DAPI fluorescence were acquired. Sort windows were set on dotplots FITC vs. DAPI, and chromosomes were sorted at rates of 15 - 20 / sec. Flow-sorted chromosomes were identified and the purity in sorted chromosome fractions was determined according to Molnár *et al*. (2011). Briefly, approximately one thousand chromosomes were sorted from each chromosome population identified on bivariate flow karyotype into a 15 µl drop of PRINS buffer supplemented with 5% (w/v) sucrose on a microscope slide (Kubaláková *et al*. 1997). The slides were air-dried and used for FISH analysis with specific probes.

4 SUMMARY OF RESULTS

This thesis is focused on the characterization of genomic constitutions of blue grained wheat genotypes using multicolour fluorescence *in situ* hybridization. Furthermore, I aimed to optimize fluorescence *in situ* hybridization in suspension (FISHIS) protocol used for flow cytometric analysis of selected species of the *Triticeae* tribe. The results helped to increase our knowledge about the structure of complex plant genomes and will be prerequisite for subsequent genomic analyses.

4.1 Cytogenetic analysis of introgression lines of wheat-Th. *ponticum* with blue aleurone trait using multicolour fluorescence *in situ* hybridization (FISH)

Anthocyanins are of great importance for human health due to their antioxidant potential. Their content is rather low in common varieties of wheat (*Triticum aestivum* L.). However, elite lines with blue aleurone and introgressed chromatin from wild relatives exhibit significantly increased levels of anthocyanins. There is evidence that the donor of chromosome introgressions has been *Thinopyrum ponticum* (syn. *Agropyron elongatum*). The aim of our study was to characterize genomic constitution of selected wheat genotypes with blue aleurone with using multicolour fluorescence *in situ* hybridization. Our results revealed six different types six different types of the *Th. ponticum* intorogressions. We suppose that introgressions of wild relatives such as *Th. ponticum* into the common lines of wheat activates the blue aleurone trait pathway, which is present, but deactivated in wheat germplasm.

4.2 Optimization and utilization of fluorescence *in situ* hybridization in suspension (FISHIS) for flow cytometric analysis of selected species of the *Triticeae* tribe

Flow cytometric analysis of fluorescence of DAPI stained chromosomes do not allow the discrimination of chromosomes which have the same or very similar DNA content. In order to overcome this obstacle was developed fast and reliable method, called fluorescence *in situ* hybridization in suspension (FISHIS). This method makes use of fluorescent oligonucleotide SSR probes and DAPI fluorescence for flow cytometric analysis. To date, this biparametric analysis was successfully used for discrimination of similar chromosomes in genera *Aegilops* (e.g., *Ae. umbellulata*, *Ae.comosa*, *Ae. speltoides*) and *Triticum* (e.g. *T. aestivum*,

T. dicoccoides). We believed that this method will be useful for sorting of chromosomes from other plant species.

5 CONCLUSION

Within the framework of this thesis, I focused on the study of genome composition of blue-aleurone wheat and on the optimization and utilization of fluorescence *in situ* hybridization in suspension (FISHIS) for flow cytometric analysis of selected species of the *Triticeae* tribe.

The first part of this thesis focuses on the genome analysis of introgression lines of wheat-*Th. ponticum* with blue aleurone trait which exhibit significantly increased level of anthocyanins compare to the common wheat cultivars. The aim of the study was to characterize genomic constitutions of these lines using genomic and fluorescence *in situ* hybridization. The results revealed large variation in chromosome constitutions of blue-aleurone wheats. Six different types of the *Th. ponticum* introgressions were identified and this confirmed the hypothesis that alien chromatin from *Th. ponticum* activates the blue aleurone trait present, but inactivated, in common wheat lines.

Flow cytometry is the next important tool for analysis of complex hybrid and polyploid genomes. Dividing the genome into small defined units, chromosomes, makes the sequencing of large genomes easier. Nevertheless, chromosomes of the majority of species have the same or very similar relative DNA content and thus, it is unable to sort them by conventional flow cytometry. Fortunately, the utilization of fluorescent *in situ* hybridization in suspension (FISHIS) is elegant solution of this situation. This method allows specific chromosome flow-sorting based on the combination of fluorescent signal of oligonucleotide SSR probe(s) and DAPI staining. The second part of the thesis deals with the optimization of FISHIS methodology for flow cytometric analysis in selected species of the *Triticeae* tribe, including *Triticum* or *Aegilops* species.

6 REFERENCES

Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) Molecular biology of the cell, 4th ed. Garland Science, New York

Barton N (1979) The dynamics of hybrid zones. Heredity 43:341-359

Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E, Leroy P, Mangenot S, Guilhot N, Le Gouis J, Balfourier F, Alaux M, Jamilloux V, Poulain J, Durand C, Bellec A, Gaspin C, Šafář J, Doležel J, Rogers J, Vandepoele K, Aury JM, Mayer K, Bergès H, Quesneville H, Wincker P, Feuillet C (2014) Structural and functional partitioning of bread wheat chromosome 3B. Science 345(6194):1249721

Feldman M, Levy AA (2012) Genome evolution due to allopolyploidization in wheat. Genetics 192(3):763-774

Friebe B, Jiang J, Tuleen, Gill BS (1995) Standard karyotype of *Triticum umbellulatum* and the characterization of derive d chromosome addition and translocation lines in common wheat. Theor Appl Genet 90:150-156

Giorgi D, Farina A, Grosso V, Gennaro A, Ceoloni C, Lucretti S (2013) FISHIS: fluorescence *in situ* hybridization in suspension and chromosome flow sorting made easy. PLoS One 8(2):e57994

Heslop-Harrison JS (2000) Comparative genome organization in plants: from sequence and markers to chromatin and chromosomes. Plant Cell 12(5):617-636

Kellogg EA, Bennetzen JL (2004) The evolution of nuclear genome structure in seed plants. Am J Bot 91(10):1709-1725

Kubaláková M, Kovářová P, Suchánková P, Číhalíková J, Bartoš J, Lucretti S, Watanabe W, Kianian SF, Doležel J (2005) Chromosome sorting in tetraploid wheat and its potential for genome analysis. Genetics 170(2):823–829

Kubaláková M, Macas J, Doležel J (1997) Mapping of repeated DNA sequences in plant chromosomes by PRINS and C-PRINS. Theor Appl Genet 94:758-763

Leliaert F, Smith DR, Moreau H, Herron MD, Verbruggen H, Delwiche CF, Clerck O (2012) Phylogeny and molecular evolution of the green algae. Crit Rev Plant Sci 31:1-46

Liu D, Zhang H, Zhang L, Yuan Z, Hao M, Zheng Y (2014) Distant hybridization: a tool for interspecific manipulation of chromosomes. In: Pratap A, Kumar J (eds) Alien gene transfer in crop plants. Volume 1. Innovations, methods and risk assessment. pp 25-42 Springer Science+Business Media, New York

Lucretti S, Giorgi D, Farina A, Grosso V (2014) FISHIS: A new way in chromosome flow sorting makes complex genomes more accessible. In: Tuberosa R, Graner A, Frison E (eds) Genomics of plant genetic resources. Volume 1. Managing, sequencing and mining genetic resources. pp 319-348 Springer Science+Business Media, Dordrecht

Madlung A (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. Heredity 110(2):99-104

Masoudi-Nejad A, Nasuda S, McIntosh RA, Endo TR (2002) Transfer of rye chromosome segments to wheat by a gametocidal system. Chromosome Res 10(5):349-357

Molnár I, Cifuentes M, Schneider A, Benavente E, Molnár-Láng M (2011) Association between simple sequence repeat-rich chromosome regions and intergenomic translocation breakpoints in natural populations of allopolyploid wild wheats. Ann Bot 107(1):65-76

Molnár-Láng M, Linc G (2015) Wheat-barley hybrids and introgression lines. In: Molnár-Láng M, Ceoloni C, Doležel J (eds) Alien introgression in wheat: cytogenetics, molecular biology, and genomics. pp 315-345 Springer Science+Business Media, Switzerland

Molnár-Láng M, Molnár I, Szakács É, Linc G, Bed Z (2014) Production and molecular cytogenetic identification of wheat-alien hybrids and introgression lines. In: Tuberosa R, Graner A, Frison E (eds) Genomics of plant genetic resources. Volume 1. Managing, sequencing and mining genetic resources. pp 255-2863 Springer Science+Business Media, Dordrecht

Renny-Byfield S and Wendel JF (2014) Doubling down on genomes: polyploidy and crop plants. Am J Bot. 101(10):1711-1725

Rey E, Molnár I, Doležel J (2015) Genomics of wild relatives and alien introgressions. In: Molnár-Láng M, Ceoloni C, Doležel J (eds) Alien introgression in wheat: cytogenetics, molecular biology, and genomics. pp 347-381 Springer Science+Business Media, Switzerland

Urbanelli S (2002) Genetic divergence and reproductive isolation in the *Ochthebius* (*Calobius*) complex (Coleoptera: *Hydraenidae*). Heredity 88:333-341

Vrána J, Kubaláková M, Šimková H, Číhalíková J, Lysák MA, Doležel J (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). Genetics 156(4):2033-2041

Wallace LE, Culley TM, Weller SG, Sakai AK, Kuenzi A, Roy T, Wagner WL, Nepokroeff M (2011) Asymmetrical gene flow in a hybrid zone of Hawaiian *Schiedea* (*Caryophyllaceae*) species with contrasting mating systems. PLoS One 6(9):e24845

Zhang P, Dundas IS, McIntosh RA, Xu SS, Park RF, Gill BS, Friebe B (2015) Wheat-*Aegilops* introgressions. In: Molnár-Láng M, Ceoloni C, Doležel J (eds) Alien introgression in wheat: cytogenetics, molecular biology, and genomics. pp 221-243 Springer Science+Business Media, Switzerland

7 LIST OF AUTHOR´S PUBLICATIONS

7.1 Original papers

Burešová V, Kopecký D, Bartoš J, Martinek P, Watanabe N, Vyhnánek T, Doležel J (2015) Variation in genome composition of blue-aleurone wheat. Theor Appl Genet 128(2):273-282

Doležel J, Vrána J, Cápal P, Kubaláková M, Burešová V, Šimková H (2014) Advances in plant chromosome genomics. Biotechnol Adv 32(1):122-136

Akpinar BA, Yuce M, Lucas S, Vrána J, Burešová V, Doležel J, Budak H (2015) Molecular organization and comparative analysis of chromosome 5B of the wild wheat ancestor *Triticum dicoccoides*. Sci Report 5:10763

Molnár I, Vrána J, Burešová V, Cápal P, Farkas A, Darkó É, Cseh A, Kubaláková M, Molnár-Láng M, Doležel J (2016) Dissecting the U, M, S and C genomes of wild relatives of bread wheat (*Aegilops* spp.) into chromosomes and exploring their syntheny with wheat. Plant J. doi:10.1111/tpj.13266

7.2 Published abstracts – poster presentation

Burešová V, Kopecký D, Šafář J, Vyhnánek T, Martinek P, Doležel J (2013) Genomic constitution of cereals with blue aleurone trait. In: Abstracts of the "Olomouc Biotech 2011. Plant Biology: Green for Good II. Olomouc, Czech Republic

Burešová V, Kopecký D, Bartoš J, Vyhnánek T, Martinek P, Doležel J (2014) Genomic constitution of wheat genotypes with blue aleurone. In: Abstracts of the "Plant molecular cytogenetics in genomics and postgenomics era". Katowice, Poland

Burešová V, Kopecký D, Bartoš J, Martinek P, Watanabe N, Vyhnánek T, Doležel J (2014) Genomic constitution of blue grained wheat genotypes. In: Sborník abstract, Bulletin České společnosti experimentální biologie rostlin, "6. Metodické dny". Seč, Czech Republic

7.3 Published abstracts – oral presentation

Burešová V, Kopecký D, Vrána J, Doležel J, Jenkins G, Phillips D (2016) Nuclear organization in interspecific plant hybrids. In: Abstracts of the "VI. Festulolium Working Group Workshop". Olomouc, Czech Republic

Burešová V, Kopecký D, Vrána J, Doležel J, Jenkins G, Phillips D (2016) Nuclear organization in interspecific plant hybrids revealed by 3D-FISH. In: Abstracts of the "Society for Experimental Biology". Brighton, United Kingdom

8 SUMMARY (in Czech)

Studium organizace hybridních a polyploidních genomů

Porozumění struktuře a organizaci rostlinných genomů je jednou z velkých výzev současné biologie. Zdokonalování stávajících a zavádění nových metod je nezbytným krokem k rozšíření našich znalostí o mechanismech a procesech, které se účastní evoluce rostlinných genomů. Molekulární cytogenetika a mikroskopické techniky patří ke standartním metodám tohoto šetření. Tyto techniky umožňují detailní studium variability, struktury a evoluce rostlinných genomů. Takové poznání genomu rostlin může rovněž hrát zásadní roli ve šlechtění rostlin.

První část práce je zaměřena na analýzu genomu introgresních linií pšenice a pýru pontického s modrým aleuronem. Tyto linie vykazují výrazně vyšší hladinu antokyanů v porovnání s běžnými odrůdami pšenice. Cílem naší studie bylo charakterizovat genomové složení těchto linií pomocí fluorescenční *in situ* hybridizace (FISH). Naše výsledky odhalily velké rozdíly v genomovém složení u modrozrnných pšenic. Celkem jsme identifikovali šest různých typů introgrese pýru pontického. Toto zjištění tak potvrdilo hypotézu, že přítomnost cizího chromatinu, v tomto případě pýru, aktivuje gen pro modrý aleuron, který zůstává u běžné pšenice neaktivní.

Průtoková cytometrie je dalším důležitým nástrojem v analýze hybridních a polyploidních genomů. Rozdělení genomu na malé jasně definované jednotky, chromozomy, činí sekvenování velkých genomů daleko snazší. Nicméně chromozomy většiny druhů mají stejný nebo velmi podobný relativní obsah DNA a tudíž třídění takových chromozomů běžnou průtokovou cytometrii není možné. Naštěstí vývoj fluorescenční *in situ* hybridizace v suspenzi (FISHIS) je elegantním řešením této situace. Tato metoda, založená na kombinaci fluorescenčního signálu oligonukleotidových mikrosatelitních sond a DAPI barvení, umožňuje třídění jednotlivých chromozomů. Druhá část této práce se zabývá optimalizací metodiky FISHIS pro průtokovou cytometrii vybraných rodů kmene *Triticeae* (*Triticum*, *Aegilops*).