Genomics of meadow fescue chromosome 4F

David Kopecký

Centre of the Region Haná for Biotechnological and Agricultural Research

Institute of Experimental Botany

Šlechtitelů 31

78371 Olomouc

Czech Republic

Phone: +420 585 238 723

Email: kopecky@ueb.cas.cz

Journal research area: Genetics, Genomics, and Molecular Evolution
Flow sorting and sequencing meadow fescue chromosome 4F

David Kopecký¹, Mihaela Martis², Jarmila Číhalíková¹, Eva Hříbová¹, Jan Vrána¹, Jan Bartoš¹, Jitka Kopecká¹, Federica Cattonaro³, Štěpán Stočes¹, Petr Novák⁴, Pavel Neumann⁴, Jiří Macas⁴, Hana Šimková¹, Bruno Studer⁵, Torben Asp⁶, James H. Baird⁷, Petr Navrátil¹, Miroslava Karafiátová¹, Marie Kubaláková¹, Jan Šafář¹, Klaus Mayer², Jaroslav Doležel¹

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

² Munich Information Center for Protein Sequences/Institute of Bioinformatics and Systems Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany

³ Istituto di Genomica Applicata and IGA Technology Services, 33100 Udine, Italy

⁴ Biology Centre ASCR, Institute of Plant Molecular Biology, Branišovská 31, 37005 České Budějovice, Czech Republic

⁵ Institute for Agricultural Sciences, Universitätstrasse 2, 8092 Zürich, Switzerland

⁶ Aarhus University, Department of Molecular Biology and Genetics, Research Centre Flakkebjerg, DK-4200 Slagelse, Denmark

⁷ Department of Botany and Plant Sciences, University of California, Riverside, CA 92521, USA
Financial Source

This work has been supported by the Czech Science Foundation (grant award P501/11/0504), the Ministry of Education, Youth and Sports of the Czech Republic (grant award OC10037) and the European Regional Development Fund (Operational Programme Research and Development for Innovations No. ED0007/01/01).

Corresponding author

David Kopecký

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

Analysis of large genomes is hampered by a high proportion of repetitive DNA, which makes the assembly of short sequence reads difficult. This is also the case in meadow fescue (*Festuca pratensis*), which is known for good abiotic stress resistance and it has been used in intergeneric hybridization with ryegrasses to produce *Festulolium* cultivars. In this work we describe a new approach to analyze the large genome of meadow fescue, which involves reduction of sample complexity without compromising information content. This is achieved by dissecting the genome to smaller parts - individual chromosomes and groups of chromosomes. As the first step, we flow-sorted chromosome 4F and sequenced it by Illumina with ~50x coverage. This provided the first insight into the composition of the fescue genome, enabled the construction of virtual gene order of the chromosome, and facilitated detailed comparative analysis with the sequenced genomes of rice, *Brachypodium*, *Sorghum* and barley. Using GenomeZipper, we were able to confirm collinearity of chromosome 4F with barley chromosome 4H and the long arm of chromosome 5H (5HL). Several new tandem repeats were identified and physically mapped using FISH. They were found as robust cytogenetic markers for karyotyping of meadow fescue and ryegrass species and their hybrids. The ability to purify chromosome 4F opens a way for more efficient analysis of genomic loci on this chromosome underlying important traits, including freezing tolerance. Our results confirm that next generation sequencing of flow-sorted chromosomes enables an overview on chromosome structure and evolution at a resolution never achieved before.
Introduction

Meadow fescue (Festuca pratensis Huds.) is a grass typically found in hygrophylic and mesophytic meadows and pastures. It provides high-quality forage in pure stands and in grass-clover mixtures. Besides its usage on temporary meadows and pastures, this species is preferentially used as a component of permanent grasslands where it produces forage during the first three years. Its outstanding winter survival is why this species is the major component of grasslands in Northern Europe and Canada, and for the same reason it is used widely in mountain regions of Central and Western Europe (Germany, Austria, Switzerland, Czech Republic and Slovakia). However, meadow fescue is less adapted and competitive in low-land coastal regions in the temperate zone. In such climates, it is being replaced by interspecific hybrids of meadow fescue (or tall fescue) with ryegrasses and especially, with Lolium multiflorum Lam.

Despite the agronomic importance of meadow fescue, either as a stand-alone crop or as a parent in intergeneric hybridization, the progress in genetics and genomics of F. pratensis is lagging far behind most other crops. Genomic studies in F. pratensis are complicated by large genome size (3,175 Mb / 1C) (Kopecký et al., 2010) and a rather symmetrical karyotype. Relative to cereals, and even to closely related ryegrasses, genetic and genomic resources of F. pratensis are limited. The only existing genetic map was developed by Alm et al. (2003) and consists of 466 RFLP, AFLP and isozyme markers with a total length of 658.8 cM. This map was recently enriched by 149 DArT markers (Bartoš et al., 2011). So far, two BAC libraries have been developed for meadow fescue. Donnison et al. (2005) constructed a 5x genome coverage library and used it to clone candidate orthologous sequences to the CONSTANS-like rice Hd1(Se1) gene in Lolium perenne and F. pratensis, while Kopecký et al. (2010) used a partial genomic BAC library (about 1x genome coverage) as a source of cytogenetic markers. The only existing DNA array was developed by Kopecký et al. (2009) using DArT technology (Jaccoud et al., 2001) and it consists of 7680 probes derived from methyl-filtered genomic representations of five species from the Festuca-Lolium complex (L. perenne, L. multiflorum, F. pratensis, F. arundinacea and F. glaucescens). It has been used for diversity studies (Kopecký et al., 2009, 2011; Baird et al., 2012), genetic mapping (Bartoš et al., 2011; Tomaszewski et al., 2012) analysis of genomic composition of Festulolium cultivars (Kopecký et al., 2011) and association analysis of agronomic traits (Bartoš et al., 2011). Sequencing 620 genetically mapped DArT markers provided the first large source of
DNA sequences for *F. pratensis*. However, DArT markers are believed to originate primarily from genic regions due to a methyl-filtration step in the preparation of genomic representations. This was proved by blast analysis against EST and hypothetical proteins databases (Bartoš et al., 2011). Thus, the DArT markers did not provide a complete insight into the genome composition. Recently, 454 transcriptome sequencing of two mapping parents of *F. pratensis* has been reported by Vigeland et al. (2013).

Recent advances in next generation sequencing (NGS) technologies, which enable sequencing entire genomes, make it possible to analyze genome structure at high resolution and perform detailed comparative studies across species. NGS data have been widely used for molecular marker development, phylogenetic and ecological studies and analysis of transcriptomes using RNA-sequencing (Egan et al., 2012). Unfortunately, the NGS of whole genomes has limited use in comparative genomic studies in species with no or poor genetic maps and in species for which a reference genome sequence are not available. In such cases, sequencing flow-sorted chromosomes was found as an extremely valuable approach. For example, NGS of sorted chromosomes has been used to study structural genome changes in cereals during their evolution and speciation. Mayer et al. (2009; 2011) used a novel approach called “GenomeZipper” to deduce the putative gene order in barley. This approach is based on the comparisons of chromosomal shotgun sequences against model genomes (rice, *Sorghum* and *Brachypodium*) to detect syntenic regions. Similarly, Wicker et al. (2011), Vitulo et al. (2011) and Hernandez et al. (2012) used GenomeZipper for comparative studies of wheat homoeologous group 1, and wheat chromosomes 5A and 4A, with rice, *Sorghum*, *Brachypodium* and a virtual barley genome. This analysis revealed considerable restructuring of wheat chromosome 4A and confirmed the utility of this approach.

Besides the use of flow-sorted chromosomes for NGS, they can be used for several other applications. A number of chromosome-specific BAC libraries were constructed for hexaploid wheat and rye (Šafář et al., 2010) and many of these libraries are used for physical mapping (Paux et al., 2008). Moreover, sorted chromosomes are valuable templates for cytogenetic mapping using FISH (Kubaláková et al., 2003; Suchánková et al., 2006). Sorted chromosomes were also used as an efficient source of genetic markers for the saturation of genetic maps (Wenzl et al., 2010).

In this study, we set out to explore the possibilities of applying chromosome genomics in *F.*
pratensis. We have developed flow cytogenetics in this species and used flow cytometric sorting to dissect the meadow fescue genome into individual chromosomes. Next generation sequencing of isolated chromosomes provided the first insights into the genome structure of this species and permitted analysis of collinearity at a high resolution level.

Results

Cell cycle synchronization and preparation of chromosome suspension
The experiments performed to optimize the induction of cell cycle synchrony and accumulation of metaphases in meristem root tips were based on our previous experience with cereals and legumes. The optimal concentration of hydroxyurea (HU) was 1.5 mM; lower concentration (1 mM) did not completely block the cycle. On the other hand, the cell cycle did not recover after blockage by 2 mM HU. The highest metaphase index (55%) was reached 5 h after the removal from the DNA synthesis inhibitor (recovery time).

Chromosome analysis and sorting
Flow cytometry of mitotic metaphase chromosomes permitted separation of F. pratensis chromosomes based on their size (relative DAPI fluorescence intensity). The histogram of chromosome DAPI fluorescence (flow karyotype) of F. pratensis (n=7) consisted of three distinct peaks (Fig 1). All seven chromosomes of meadow fescue can be individually identified based on their specific patterns of hybridization with various DNA probes (Kopecký et al., 2010). This permitted identification of particles sorted from individual peaks on flow karyotype. Thus, chromosomes 1, 5 and 6 were sorted as a group from peak I, chromosomes 2, 3 and 7 formed peak II and, chromosome 4 formed peak III. On average, we were able to collect about 8×10³ copies of chromosome 4F from a sample prepared from ca 150 root tips. For DNA amplification and subsequent sequencing, we collected about 40,000 4F chromosomes. Purity of isolated chromosomes was estimated by FISH. Based on screening 300 sorted chromosomes from two randomly chosen samples, the average purity was 92.4%, with contamination by almost all remaining chromosomes, in low frequencies (chromosome 1 – 0.36%; chromosome 2 – 2.18%, chromosome 3 – 2.54%, chromosome 5 – 0.36% and chromosome 7 – 2.18%).

Amplification and sequencing chromosome 4F
An equivalent of 50 ng DNA was obtained by flow cytometric sorting of 40,000 copies of
chromosome 4F. This yielded 25 ng purified DNA, which was used in three independent multiple displacement amplification (MDA) reactions providing a total of 12.1 µg (4.1 µg + 4.2 µg + 3.8 µg) DNA. For sequencing, individual samples of amplified DNA were combined to reduce a possible bias introduced by the MDA. A total of 4 µg MDA DNA from chromosome 4F were used for Illumina HiSeq 2000 sequencing. A total of 85,351,865 paired-end sequence reads with lengths of 101 bp were generated, yielding 28,553 Mb of sequence. Considering the size of chromosome 4F at 543 Mb (Kopecký et al., 2010), this represented ca 50x coverage.

The assembly of the Illumina paired-end reads was performed using SOAPdenovo (Luo et al., 2012) with different k-mer sizes (i.e. 21 – 83 k-mer sizes). The assembly with the best L50 (70-mer, 132 bp) and a maximum contig length of 15,919 bp was chosen for further analyses. The low L50 value, the large number of short contigs and the high amount of repetitive elements indicate a highly fragmented assembly. Removing the repetitive sequences from the assembly, the L50 value increases considerably (1623 bp). This suggests that the non-repetitive genome space is well assembled and provides a useful resource for chromosome structure analysis, gene detection and construction of a virtual ordered gene map.

Gene content of chromosome 4F
To estimate the number of genes on 4F, sequence comparisons against the genomes of Brachypodium, rice and Sorghum were made. We used stringent BLASTX searches that required ≥75% (Brachypodium) / 70% (rice, Sorghum) sequence similarity over at least 30 amino acids (Table 1). Between 2,629 and 3,056 significant matches were found. A total count of cumulative, non-redundant matches on 4F was 4,626. Considering the estimated molecular length of 4F at 543 Mb and the gene density on 4F representative for the entire meadow fescue genome (Kopecký et al., 2010), this would scale up to 27,048 genes. There should be mentioned that this number does not consider the putatively non-homologous genes.

A virtual gene map of chromosome 4F
Because of the lack of suitable genetic markers for F. pratensis, 154 genetic markers from chromosome 4 of L. perenne were used as a backbone to construct a virtual gene map of chromosome 4F. L. perenne is closely related species, which evolved from Festuca about 2.2 MYA (Polok, 2007). Out of the 154 L. perenne markers, 109 (70.8%) could be associated to
genes with conserved synteny under these conditions: (a) the best first hit with (b) a minimal alignment length of 30 amino acids and (c) an identity of 75% (Brachypodium) /70% (rice, Sorghum). A virtual ordered gene map was developed using GenomeZipper protocol (Mayer et al., 2011). With this approach, we identified 4,672 loci from which 4,626 correspond to regions with conserved synteny in Brachypodium, rice and Sorghum (Table 1). Among these, 1,515 (32.7%) loci were supported by genes in all three reference organisms, 855 (18.5%) in two of them and 2,256 (48.8%) genetic loci were supported by only one gene with conserved synteny (Fig 2). The higher number of syntenic Brachypodium genes reflects a closer phylogenetic relationship of meadow fescue to Brachypodium than to rice and Sorghum. Furthermore, 1,697 barley full length cDNAs were uniquely associated to either a marker sequence or at least one syntenic gene (Table 1).

**Comparative analysis of chromosome 4F with relatives**

Using GenomeZipper, we performed a detailed comparative analysis of chromosome 4F with sequenced genomes of Brachypodium, rice, Sorghum and barley. An ancient translocation differentiates chromosome 4F from 4H of barley (Fig 3). Chromosome 4F is collinear with entire chromosome 4H, but also with the terminal region of 5HL. Based on the position of genetic markers, the 5H syntenic segment on is in inverted orientation. For sequences generated from 4F, we identified *in silico* syntenic regions on Brachypodium chromosomes 1 and 4 (Fig 4), rice chromosomes 3 and 11 (Fig S1) and Sorghum chromosomes 1 and 5 (Fig S2).

Comparative mapping results between physical maps of barley chromosomes 4H and 5H, Festuca 4F, Lolium 4L and wheat 4A show that the terminal segment originally from the ancestral chromosome 4 and now located at chromosome arm 5HL is not identical to the segment involved in the 4AL/5AL translocation (Naranjo et al., 1987; Fig S3). Our observation is similar to that of Alm et al (2003). The location of the corresponding segment from 5HL is found on short arms of Festuca 4F and Lolium 4L chromosomes (both carrying ancestral-type chromosome 4), while in wheat, the long arm of chromosome 4A is involved (Fig S3). The 5H regions syntenic with 4F and 4A are from 127.96 cM and from 159.79 cM to the telomere (196.85 cM), respectively. While Lolium 4L and Festuca 4F appear structurally alike, a comparison of Festuca 4F against wheat 4A indicates numerous inversions and translocations (Fig 5). The extensive collinearity between Lolium and Festuca along with the known genome structure of Lolium (Pfeifer et al., 2013), suggest that the collinear 5HL segment on Festuca chromosome 4 represents a unidirectional translocation in
barley, rather than a reciprocal translocation. In case of reciprocal translocation, we would expect to observe some sequences being present in terminal region of 4HS and absent in 4F (Fig 3; Fig 5A). However, this was not a case, and all sequences detected in the terminal region of 4HS were also present in chromosome 4F. This interpretation is consistent with observations made in *Lolium* (Pfeifer et al, 2013). Since *Festuca* (and *Lolium*) chromosome 4 is collinear with rice chromosome 3, which indicates more ancestral chromosomal structure, we hypothesize that this unidirectional translocation occurred in the Triticeae branch after the *Festuca/Lolium* split.

The linear ordered gene maps of *Festuca* chromosome 4F, wheat chromosome 4A and barley chromosome 4H allows to analyze the number of syntenically conserved genes against the reference genomes of *Brachypodium*, rice and *Sorghum* (Fig 6). The proportion of shared syntenic genes in all three chromosomes ranges between 23% and 28%, while between 19% and 24% of the genes are found for two of the chromosomes and between 49% and 57% are found for only one of the chromosomes analyzed. Strikingly higher percentages of genes conserved in *Festuca* 4F but not for barley 4H and wheat 4A (30-47%) were found illustrating the higher degree of syntenic conservation on one hand and, at least in part, a closer evolutionary relationship of *Festuca* against the reference genomes used.

**Repetitive DNA of chromosome 4F**

With the aim to determine the major repetitive DNA constituents of chromosome 4F, repeat reconstruction was done on Illumina data representing 1x coverage of chromosome 4F (550 Mbp). Graph-based clustering resulted in 252,144 clusters containing 4,224,228 sequence reads (76%). 855 clusters contained at least 100 sequences, from which 450 clusters contained more than 555 sequences (0.01% of analyzed reads) (Fig S4). The largest clusters were manually annotated using a combination of multiple approaches based on similarity searches against sequence databases and graph layout analysis. The aim was to identify tandem organized repeats useful as species specific and/or chromosome specific cytogenetic markers. Annotation of the largest clusters led to classification of ~ 53% of the 4F chromosome sequence. Largest clusters contained mainly Ty3/Gypsy-like elements, which were found to be the most abundant repeats (28.64% of analyzed sequence reads) of chromosome 4F. Ty1/Copia-like elements represented less than 7% and DNA transposons represented about 3% of annotated sequences (Fig 7). Interestingly, we were not able to characterize sequences within several large clusters which account for more than 8% of the analyzed sequence reads of chromosome 4F. These sequences can represent unknown DNA repeats and/or WGA
Based on the graph layout (Fig 8), contigs derived from clusters containing putative tandem repeats were identified and further investigated using similarity dot-plot analysis (Fig S5). The analysis revealed presence of tandem or semi-tandem organized repeats in 43 clusters, which together represented about 2.71% of chromosome 4F sequence.

**Experimental investigation of tandem organized repeats**

Out of 43 *in silico* identified tandem repeats, sequences of 15 putative tandem repeats containing repetitive units of different length were investigated in more detail to study their organization in the *Festuca* and *Lolium* genomes (Table S1). A set of primers specific for tandem units was designed and used for preparation of probes for FISH and for Southern hybridization. PCR products obtained after amplification on genomic DNA of *F. pratensis* as the template DNA were sequenced using dideoxy-chain termination reaction to check their authenticity. Moreover, PCR was done on flow-sorted chromosomes of *F. pratensis* and on *L. perenne* genomic DNA to verify species and/or chromosome specificity. The results showed that all tandem repeats were present in all chromosome peaks of *F. pratensis* flow karyotype and positive PCR products were obtained for all 15 selected tandem repeats also on *Lolium* genomic DNA (Table S1).

To confirm the tandem character of the repeats within the clusters, Southern hybridization was used to study the organization of repeats from the clusters which were successfully mapped on *Festuca* mitotic chromosomes using FISH (see below). A ladder-like pattern which is typical for tandem organized repetitive units was obtained for all 8 putative tandem repeats which gave visible signals after FISH (see below) (Fig 9). As shown in Fig 10, the tandem repeats identified in sequences of chromosome 4F of *F. pratensis* are also present in the nuclear genome of *L. perenne*, even though at lower copy number.

**Physical localization of repeats using FISH**

Out of 15 tandem repeats tested (fpTR1 – fpTR15), we were able to physically map five fpTRs on chromosome 4F using FISH (Fig 9; Table S1). All of them produced hybridization patterns also on other chromosomes. Additionally, the other three fpTRs mapped on other chromosomes other than 4F (fpTR6, fpTR11 and fpTR12) (Table S1). This might be caused by the sensitivity of FISH. None of the fpTRs produced signal exclusively on 4F and thus, they cannot be considered as chromosome-specific. On the other hand, our observation opens
a way for precise and robust molecular karyotyping of this species. As most of the fpTRs produced visible signals also on mitotic chromosomes of *L. perenne* and F1 hybrid *L. perenne × F. pratensis* (Fig 9), the repeats may find even broader use as probes for FISH in cytogenetics of the *Festuca – Lolium* complex.

Using FISH, we were able to localize telomeric repeat on *Festuca* chromosomes. Surprisingly, it produced signals not only in the telomeric regions of both chromosome arms of 4F, but also interstitially (Fig 9). Such interstitial signals were also detected in chromosome arms 2FL, 5FL and 6FS. In general, FISH done on sorted chromosomes displayed higher sensitivity and produced more bands with higher resolution on individual chromosomes than on standard squashed preparations (Fig 9).

**Discussion**

Despite their considerable economic importance, forage and turf grasses lag far beyond many other crops in terms of genetics, genomics and bioinformatics. The delay is due to many factors, but outcrossing character, population-based breeding and frequent aneuploidy do not help. Moreover, genomes of species within *Festuca-Lolium* are large and complex. The genome of meadow fescue was estimated at 1C = 3,175 Mb by DNA flow-cytometry and is similar in size to that of human (Doležel et al., 2003; Kopecký et al., 2010). The ryegrass genome is just a little smaller (1C = 2,623 Mb of *L. perenne* and 1C = 2,567 Mb of *L. multiflorum*). Because of the genome complexity, any approach reducing sample complexity is welcome in genomic studies.

**Partitioning the grass genomes**

Analysis of a complex genome where a majority of DNA is represented by repeats can be simplified by several approaches. There are methods available which avoid sequencing of the repetitive parts of genomes. Of these, sequencing of cDNA to generate ESTs (Expressed Sequence Tags) is among the most successful. However, ESTs fail to sample rare or conditional transcripts (Martienssen et al., 2004) and other methods were proposed to target the gene space in large and complex genomes, such as Cot fractionation and methyl filtration (Rabinowicz et al., 1999; Petersen et al., 2002). Unfortunately, the latter two methods did not meet the expectations.

Another alternative to tackle the complex genomes is to dissect them into smaller elements and sequence these elements individually. Working with naturally uniform and independent units – such as chromosomes – is perhaps the most powerful approach.
individual chromosomes represent only 3.6-5.9% of the entire genome (Doležel et al., 2009) and even in a diploid species such as meadow fescue, isolation of individual chromosomes dissect the 3.2 Gb genome into 373-543 Mb units, each representing 11.7-17.1% of the entire genome (Kopecký et al., 2010).

Individual chromosomes can be isolated in two ways: by microdissection and flow-sorting. Microdissection gives access to any chromosome or a chromosome segment. However, this is a very tedious approach and for all practical purposes the total yield is limited to only a few copies of a particular chromosome (Zhou and Hu, 2007). On the other hand, flow cytometry can isolate high copy numbers of the same chromosome with purity usually exceeding 90%.

Flow sorting relies on differences in chromosome size or, rather, on the difference in relative fluorescence intensity. The output of flow cytometric analysis is a histogram of relative chromosome fluorescence intensity (reflecting chromosome size), which is called a flow karyotype (Fig 1). Ideally, each chromosome is represented by a single peak on the flow karyotype. However, similarities in chromosome size within a genome usually result in the appearance of composite peaks representing two or more chromosomes. According to Doležel et al. (2009), there has to be at least a 10% difference in chromosome size to generate a separate peak on a flow karyotype. Unfortunately, in most plant species, chromosomes are not that different in size. In hexaploid wheat (21 chromosome pairs), the flow karyotype consists of only four peaks, and only one of those contains a single chromosome, 3B (Vrána et al., 2000). Similarly, the genome of meadow fescue was dissected here into three peaks, where only one peak represented a single chromosome type (chromosome 4F). However, the plasticity of plant genomes (especially in polyploids) makes it possible to develop special cytogenetic stocks with reconstructed karyotypes. Single chromosome substitution lines of *L. multiflorum – F. pratensis*, developed in our previous work, may provide an option for sorting other chromosomes of meadow fescue due to a sufficient difference in chromosome length (Kopecký et al., 2010).

**Virtual gene order on chromosome 4F**

The GenomeZipper approach applied to individual chromosomes sequences is a powerful tool for comparative studies. It significantly increases the resolution level relative to genetic markers. In the previous report on the collinearity of the meadow fescue genome with rice, *Sorghum*, *Lolium*, oat, maize and Triticeae, a limited number of markers was used (Alm et al., 2003). Prior to this study, only 36 markers were available from chromosome 4F, with only a subset (7-30) suitable for any pairwise comparison. This contrasts with thousands of
sequences from individual chromosomes (4F here) of model plant species (rice, *Sorghum*, *Brachypodium* and barley).

**Estimation of gene content**

The GenomeZipper approach permitted estimation of gene content in *F. pratensis*, and by implication, all related *Festuca* species. Assuming the estimated size of chromosome 4F at 543 Mb and the 4F gene density representative for the entire *Festuca* genome (Kopecký et al., 2010), the total *Festuca* genome contains about 35,000 genes. It is slightly higher number that the estimation for barley (≥32,000 genes; (Mayer et al., 2011) and less than that for the B-genome of wheat (38,000 genes; Choulet et al., 2010). The recent estimate for the A-genome of wheat is below that for *Festuca* and barley (28,000 genes; Hernandez et al., 2012). In *Lolium perenne*, the most closely related species for which DNA sequence data is available, almost 25,700 matches with *Brachypodium* genes were obtained (Byrne et al., 2011).

Apart from estimating gene content, the annotated sequences represent an invaluable genome resource for molecular biologists and breeders of fescue. For example, the sequences can be screened for the presence of resistance genes analogs. Because of a general lack of molecular markers for fescue this new resource opens avenues for *in silico* identification of sequences suitable as DNA markers, including SSR, presence/absence variation (PAV) and SNP and help in developing SNP platforms for marker assisted breeding of this agriculturally valuable species.

**4F is an ancestral chromosome towards modern cereals**

Based on the comparative analysis, chromosome 4F appears collinear with entire barley chromosome 4H as well as with the terminal region of the long arm of chromosome 5H (5HL) (Fig 3). The observed homology of 4F to chromosome arms 4S, 4L, 5L from Triticeae is consistent with previous observations by Alm et al. (2003) based on mapping of molecular markers. Chromosome 4F is collinear with the entire chromosome 3Os of rice (including the centromeric introgression of chromosome 11Os). On the other hand, 3Os is collinear with chromosome 4H and also with the terminal region of 5HL (Gale and Devos, 1998). Thus, *F. pratensis* chromosome 4F is likely more ancient than barley chromosome 4H. In cereal genomes (A, B and D genomes of wheat and H genome of barley), the direct ancestral precursor which is syntenic with rice chromosome 3Os was broken and the terminal part of the short arm was translocated to the distal part of the long arm of chromosome 5, resulting in
the present-day chromosome 5H. This happened after the divergence of *Festuca* and *Lolium* from the grass lineage between 25 MYA (divergence of Triticeae and Poeae) and 11.6 MYA (divergence of barley and wheat genomes) (Chalupska et al., 2008). In *Festuca* and *Lolium*, the direct ancestral precursor syntenic with 3Os of rice remained intact and resulted in the present-day 4F and 4L chromosomes of *Festuca* and *Lolium*, respectively, as originally proposed by Alm et al. (2003). Therefore, chromosome 4F, in its current form, can be assumed as a potential transient chromosome 4H' proposed by Thiel et al. (2009). Interestingly, a four-fold increase in recombination frequency was observed in the region surrounding the translocation breakpoint (Kopecký et al., 2010). This may indicate an unstable region of the chromosome more prone to chromosome breakage.

**Newly identified repeats are a valuable source of cytogenetic markers**

Cytogenetic mapping involves physical localization of a sequence of interest directly on chromosomes. Cytogenetic mapping has many applications, ranging from a study of structural chromosome changes (Mandáková and Lysák, 2008) to determination of positions and orientation of unassembled BAC contigs to support development of physical maps (Pedrosa-Harand et al., 2009). In our specific case, cytogenetic markers capable of precise and unambiguous identification of chromosomes are helpful for purity checks of sorted chromosome fractions.

Various types of tandem repeats are valuable source of useful cytogenetic markers. Satellites are tandemly repeated sequences with a repeat unit >25bp. Satellites of the *Afa* family have been used for identification of the D-genome chromosomes in wheat (Rayburn and Gill, 1986) and of barley chromosomes (Tsujimoto et al., 1997). Other satellites were used for karyotyping in barley (Brandes et al., 1995), *Avena* (Katsiotis et al., 1997), banana (Hřibová et al., 2007) and other plant species (Sharma and Reina, 2005) because they tend to form clusters, which facilitates their detection and positive chromosome identification. In this study, eleven newly characterized satellites provided localized signals on one or more chromosomes of *F. pratensis* and could serve as new cytogenetic markers in this species and probably also in ryegrasses.

Chromosomes of *Festuca* and *Lolium* can be readily discriminated in hybrids using genomic *in situ* hybridization (Thomas et al., 1994; Kopecký et al., 2008b). This method enables visualization of parental chromatin in any stage of the mitotic cycle in natural allopolyploids or synthetic wide hybrids (Schwarzacher et al., 1989). It is assumed that this discrimination is based on the presence of abundant species-specific repetitive elements. From this perspective
alone it is of considerable interest to analyze the frequency and distribution of major repeats in parental species of a wide hybrid, to determine which of them and in what proportions, contribute to the discrimination of parental genomes in hybrids. From observations using PCR made in this study, all newly identified tandem repeats were present in both genera – Festuca and Lolium. Southern hybridization and FISH revealed slight differences in presence or in signal intensity of individual fpTRs. Hence none or just a few of the 15 tested appears responsible for distinguishing parental genomes. It indicates that tandem repeats are not responsible for the discrimination capacity of genomic in situ hybridization. However, newly identified tandem repeats provided new source of cytogenetic markers and significantly increased level of chromosome identification.

Conclusions
This is the first report on dissection of a complex and large forage grass genome using chromosome sorting. We sorted chromosome 4F of meadow fescue and sequenced it by Illumina. This provided a unique and rich resource to study genome organization of the species. The sequence data was used to estimate gene content, construct virtual gene order and characterize repetitive elements. This approach provided an opportunity to describe synteny between chromosome 4F and genomes of model species (Brachypodium, rice, Sorghum and barley). Finally, we demonstrated the potential of sorted chromosomes for cytogenetic mapping of various repeats. Our results demonstrate that coupling the chromosome sorting and next generation sequencing technologies is a powerful approach, which provides insights on chromosome structure and evolution at superior resolution.

Material and Methods

Plant material
For chromosome sorting, seeds of meadow fescue (Festuca pratensis Huds., 2n=2x=14) cv. Fure were obtained from Dr. Arild Larson, Graminor AS, Norway. Seeds were germinated in the dark at 25°C distributed in Petri dishes filled with distilled water for 5-7 days to achieve optimal root length (ca 2-3 cm). For additional in situ hybridization experiments, L. perenne cv. SR4220, F. pratensis cv. Fure, and a diploid F1 hybrid of L. perenne SR4220 × F. pratensis cv. Skawa were used. The plants was transferred to a hydroponic culture of Hydroponex at 0.9 g/l (Hu-Ben, Čerčany, Czech Republic); after 5 to 7 days, actively growing root tips were collected to ice water for ~28 hours, fixed in a 3:1 mixture of absolute ethanol...
and glacial acetic acid at 37°C for seven days, stained in 1% acetocarmine for two hours and squashed in a drop of 45% acetic acid on clean microscope slides (Masoudi-Nejad et al., 2002).

Cell cycle synchronization and accumulation of metaphases
For chromosome flow sorting, synchronization of cell cycle and accumulation of metaphases in root tips were performed according to Vrána et al. (2000). Distilled water in Petri dishes was replaced by Hoagland’s nutrient solution (Gamborg and Wetter, 1975) with 1, 1.5 or 2 mM hydroxyurea (HU). After 18 h incubation, the solution was replaced by a HU-free Hoagland’s solution. Samples of root tips were taken at 1-h intervals for up to 6 hrs and the cell cycle synchrony and mitotic activity were examined under a microscope. To accumulate cells at metaphase, the seedlings were treated for 2 hrs with Hoagland’s solution containing 5 μM oryzalin after recovery from HU. Mitotic activity and metaphase frequency were analyzed on Feulgen-stained squash preparations.

Preparation of chromosome suspension
Chromosome suspension was prepared according to Doležel et al. (1992) with minor modifications. Seedlings were rinsed in deionized water and fixed in 2% v/v formaldehyde fixative made in Tris buffer supplemented with 0.1% w/v Triton X-100 at 5 °C for 20 min. After washing in Tris buffer, about 150 root tips (ca 2 mm long) were cut and transferred to a vial tube containing 750 μl LB01 buffer (Doležel et al., 1989). The chromosomes were isolated after homogenization with a Polytron PT1300 homogenizer (Kinematica AG, Littau, Switzerland) at 20 000 rpm for 15 s. To remove large cellular fragments, the suspension was passed through a 50-μm pore size nylon mesh. Prior to flow cytometry, isolated chromosomes were stained with 4,6-diamino-2-phenylindole (DAPI) adjusted to a final concentration of 2μg/ml.

Chromosome analysis and sorting
Chromosome analysis and sorting were done on a FACSaria II SORP flow cytometer (BD Biosciences, San José, USA) equipped with a 100 mW 488nm laser for scattered light detection and a 100 mW 355 nm laser for DAPI excitation. Suspension of chromosomes was analyzed at rates of 800-1000 events/sec. Approximately 20,000 – 50,000 chromosomes were analyzed in each sample. DAPI fluorescence was collected through 450/50 band-pass filter.
To discriminate doublets, dot-plots of DAPI fluorescence area versus DAPI fluorescence width were used. To ensure high purities of sorted chromosomes, tight gate was created around the chromosome 4F population and chromosomes were sorted at rates of 4-6 chromosomes/sec. To evaluate purity of chromosome sorting, approximately 1,000 chromosomes were sorted onto microscopic slides and tested by FISH. For DNA amplification, 40,000 chromosomes were collected into individual tubes and stored at -20 °C until used.

**Test of purity of sorted chromosomes using fluorescence in situ hybridization (FISH)**

To identify flow-sorted chromosomes and determine the extent of contamination of sorted chromosome fractions, we performed FISH with probes for the 5S and 45S ribosomal DNA and two BAC clones (1G18 and 2N14). The two BAC clones were selected from a partial BAC library of *F. pratensis* (Kopecký et al., 2010). The DNA was labeled by DIG-Nick Translation Kit or biotin-Nick Translation Kit (Roche Applied Science, Indianapolis, IN, USA) or by PCR labeling with biotin- or digoxigenin-labeled nucleotides (Roche). In situ hybridization was performed according to Kubaláková et al. (2003). Detection of hybridization sites was by Anti-digoxigenin-FITC (Roche) and streptavidin-Cy3 (Amersham, Piscataway, NJ, USA) and counterstaining was with DAPI in the Vectashield antifade solution (Vector Laboratories, Burlingame, CA, USA). Preparations were screened under Olympus AX70 microscope with epi-fluorescence and a SensiCam B/W camera.

**Amplification and sequencing of chromosome 4F**

Flow-sorted chromosomes were treated with proteinase and their DNA subsequently purified using Microcon YM-100 columns (Millipore Corporation, Bedford, USA) as described by Šimková et al. (2008) with minor modifications. Chromosomal DNA was amplified by multiple displacement amplification (MDA) using Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Bio-Sciences Corp., Piscataway, USA). Sequencing of amplified chromosomal DNA was performed with the HiSeq 2000 (Illumina Inc., San Diego, CA, USA). Four micrograms of MDA-amplified DNA were used to create the corresponding shotgun DNA-seq library. The library for the 4F chromosome was run in a single lane at IGA Technology Services Srl (Udine, Italy).

**Assembly of chromosome 4F sequences and Genome Zipper analysis**
The Illumina paired-end reads were assembled by SOAPdenovo (Luo et al., 2012) using different k-mer sizes. The result of the 70-mer run provided the assembly with the best sequence coverage and L50. Repetitive DNA content was identified using Vmatch (http://www.vmatch.de) against the MIPS-REdat Poaceae v8.6.2 repeat library by applying the following parameters: 70% identity cut-off, 100 bp minimal length, seed length 14, exdrop 5 and e-value 0.001. The MIPS repeat library contains known grass transposons from the Triticeae Repeat Database (http://wheat.pw.usda.gov/ITMI/Repeats), as well as de novo detected LTR retrotransposon sequences from various grass species. After repeat masking and filtering out the repeats, 26.6% (1,081,325 sequences) of the sequences remained and were considered for the subsequent steps.

To assess the number of syntenic conserved genes present in the _F. pratensis_ chromosome 4, the repeat-filtered contigs were aligned against the protein sequences of _Brachypodium_ (v1.2), rice (rice RAP-DB genome build 4) and _Sorghum_ (v1.4) by BLASTX. The following stringent filtering criteria: (i) only the first best hit with (ii) a minimal alignment length of 30 amino acids and (iii) a minimal sequence identity of 75% (_Brachypodium)/70% (rice, _Sorghum_) were applied. A sliding window approach (windows size 0.5 Mbp, shift size 0.1 Mbp) was used to identify segments with conserved gene order, based on the density of homology matches between _Festuca_ and the reference genomes.

The extracted conserved genes from the three model grass genomes and the corresponding _Festuca_ contigs were structured and ordered to a virtual linear gene model by using the genomeZipper approach (Mayer et al., 2011). The approach uses a genetic marker scaffold to compare and subsequentially integrate and order syntenic conserved, homologous genes (orthologs) along the scaffold. Thereby intervals defined by the genetic markers serve as anchor points to define the corresponding intervals from the reference genomes. Genes from the least distant reference genome (in this case _Brachypodium_) get highest priority for ordering and subsequentially the additional reference genomes are overlayed onto this scaffold. Due to the absence of a high-quality genetic marker map of _F. pratensis_, genetic markers of _L. perenne_ chromosome 4 were used as a scaffold to anchor the collinear segments from the model genomes (Studer et al., 2012).

To analyze chromosomal rearrangements between _Festuca_ and barley, chromosome 4F contigs were compared against the physical map of the barley genome (Mayer et al., 2012) using BLASTX. Filtering criteria were adjusted to the evolutionary distance between _F. pratensis_ and _H. vulgare_ (at least 85% identity and a minimum alignment length of 100 bp).

To test whether the observed _Festuca_ chromosome 4/5 translocation is structurally identical
as the 4/5 translocation found in Triticeae (Naranjo et al., 1987), the Festuca 4F contigs, ordered gene scaffolds for wheat 4A and Lolium 4L and barley chromosomes 4H and 5H were compared by bi-directional BLAST hits and visualized using Circos (Krzywinski et al., 2009). To estimate the number of conserved syntenic genes shared by wheat 4A, barley 4H and Festuca 4F, the Brachypodium, rice and Sorghum homologs contained in the individual linear ordered gene maps were compared to each other. A Venn diagram to illustrate shared genes in all three chromosomes was generated for each reference genome.

Repeat reconstruction and annotation
A random data set representing 1x coverage of chromosome 4F (550 Mb) was extracted from the Illumina data and used for reconstruction of repetitive elements using graph-based method according to Novák et al. (2010) using Louvain clustering algorithm (Blondel et al., 2008). Resulting clusters of sequence reads were manually annotated using several sources, including similarity searches with RepeatMasker; blastx and blastn programs were used for similarity search against public databases and also against the database of domains derived from plant mobile elements (http://repeatexplorer.umbr.cas.cz). Clusters represented as graphs were also analyzed using the SeqGraphR program (http://w3lamc.umbr.cas.cz/lamc/resources.php). Dotter (30) was used to confirm the presence of tandem organized repeats.

Physical localization of repeats using FISH
FISH on flow-sorted chromosomes was done according to Kubaláková et al. (2003) while the protocol of Masoudi-Nejad et al. (2002) was used for metaphase plates prepared from root tips of F1 hybrid. Probes were prepared for 15 putative tandem repeats (fpTR1 – fpTR15) identified in sequence data using PCR labeling with biotin- or digoxigenin-labeled nucleotides (Roche) and pairs of specific primers. For chromosome identification in metaphase of F1 hybrids, L. perenne and F. pratensis, we used a 5SrDNA probe accompanying with reprobing of slides with a 45SrDNA probe and a probe made from genomic DNA of L. perenne. The latter two probes were made using biotin-Nick Translation Kit and DIG-Nick Translation Kit (Roche), respectively. For the 45SrDNA probe, DNA clone pTa71 (Gerlach and Bedbrook, 1979) containing a 9 kb EcoRI fragment of wheat ribosomal DNA, which carries the 18S-5.8S-26S cluster of ribosomal RNA genes was used. The probe for 5SrDNA was prepared using PCR with a pair of specific primers (RICRGAC1, RICRGAC2), which amplify 303 bp in rice (Fukui et al., 1994), using rice genomic DNA as a template. Additionally, a probe for
telomeric repeats was prepared using PCR with (AGGGTTT)$_3$ and (CCCTAAA)$_5$ primers without a template. Reprobing was done as described by Schwarzacher and Heslop-Harrison (2000). Probe hybridization signals were detected by Anti-DIG-FITC and avidin-Cy3 conjugates, and counterstaining was by 0.2 mg/ml DAPI in Vectashield antifade solution (Vector Laboratories). Slides were evaluated with Olympus AX70 microscope equipped with epi-fluorescence and a SensiCam B/W camera. ScionImage and Adobe Photoshop software were used for processing of color pictures.

**Southern hybridization**

Aliquots of genomic DNA samples corresponding to 3 x 10$^8$ of the nuclear genomes of *F. pratensis* Huds. cv. Fure and *L. perenne* cv. SR4220 were digested using restriction enzymes *Hae*III, *Mse*I and *Eco*RI, size-fractionated by 1.2% agarose gel electrophoresis and transferred to Hybond N+ nylon membranes (Amersham, USA). Probes specific for selected putative tandem repeats were amplified using specific primers (Table 1) and labeled with biotin. The Southern hybridization was done at 68°C over night and signals were detected using BrightStar® BioDetectTM kit under manufacturer's instructions to 90% stringency (Ambion, Austin, Texas), incubated with chemiluminescent substrate (CDP-Star, Amersham Biosciences) and exposed on X-ray film.

**Acknowledgements**

We are grateful to Dr. Arild Larsen and Dr. Vladimír Černoch for providing plant material. We thank Dr. Jarmila Číhalíková, Romana Nováková, Bc., and Zdeňka Dubská for assistance with chromosome sorting and DNA amplification. Special thanks belong to Prof. Adam J. Lukaszewski for critical reading and valuable comments.

**Literature Cited**


**Gamborg OL, Wetter LR** (1975) Plant tissue culture methods. National Research Council of Canada, Saskatoon, Saskatchewan, Canada


plant species. Chromosome Res 17: 405-417


evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice. BMC Evol Biol 9: 209


Zhou RN, Hu ZM (2007) The development of chromosome microdissection and
microcloning technique and its applications in genomic research. Curr Genomics 8: 67–72

Figures

Figure 1, Flow karyotype. Distribution of relative chromosome fluorescence intensity (flow karyotype) obtained after analysis of DAPI-stained suspension of mitotic chromosomes of meadow fescue (Festuca pratensis, 2n=2x=14). The flow karyotype consists of two composite peaks I and II representing chromosomes 1F, 5F and 6F and chromosomes 2F, 3F and 7F, respectively, and a peak III representing chromosome 4F. This chromosome can be discriminated and sorted individually.

Figure 2, Model species’ genes anchored to 4F. Venn diagram showing the number of Brachypodium (Bd), rice (Os) and Sorghum (Sb) genes which are anchored in the virtual gene map of Festuca pratensis chromosome 4F. Overlaps of the circles show the number of genes anchored at the same locus.

Figure 3, Collinearity of 4F with barley genome. Festuca pratensis 4F Illumina contigs mapped on the physical map of Hordeum vulgare chromosome 4H and 5H. Syntenic regions are painted in yellow-red color tones, whereas non-syntenic regions are painted in dark blue. This approach clearly identifies collinearity of chromosome 4F with barley chromosome 4H and terminal part of long arm of 5H chromosome. The connectors used indicate orientation of the particular parts.

Figure 4, Collinearity of 4F with Brachypodium. High-density comparative analyses of the linear gene order of the Festuca pratensis genome zipper versus the sequenced genome of barley and Brachypodium distachyon. The figure includes sets of concentric circles. The inner circle represents the chromosome 4F (green) and collinear chromosomes 4H (red) and 5H (purple) of barley. The outer circles represent homologous chromosomes of Brachypodium – Bd1 and Bd4. The heat maps illustrate the density of genes hit by the contigs from the 4F chromosome. Syntenic regions are painted in red color tones, whereas non-syntenic regions are painted in blue. Putative orthologs between 4F and barley chromosomes 4H and 5H are
connected with lines. The scale is given in gene loci for chromosome 4F and in megabase pairs for barley chromosomes 4H and 5H.

**Figure 5, Chromosome 4/5 translocation.** The first three figures (A,B,C) show the chromosome 4/5 translocation in three different grass chromosomes (*Festuca* 4F, wheat 4A and *Lolium* 4L) in comparison to the corresponding homoeologous chromosomes 4 and 5 in barley. The location of the segment from 5HL is for both *Festuca* (A) and *Lolium* (B) found on the short chromosome arm, while in wheat (C) the long arm of chromosome 4A is involved. The observed chromosome 4/5 translocation is a unidirectional translocation in *Festuca* and *Lolium* rather than a reciprocal translocation as found in wheat (Alm et al., 2003). The last two circles depict the collinearity between *Festuca pratensis* chromosome 4F and *Lolium perenne* chromosome 4L (D) and wheat chromosome 4A (E), respectively. While *Lolium* and *Festuca* are highly collinear, the synteny between wheat and *Festuca* is interrupted by several inversions and translocations.

**Figure 6, Conserved syntenic regions in the virtual linear ordered maps of Festuca, barley, wheat and Lolium chromosome 4 as defined by comparison to model grass genomes.** The intersecting circles of the Venn diagrams depict the percentage of syntenic conserved reference genes (*Brachypodium* distachyon, *Oryza sativa*, *Sorghum bicolore*) shared by *Festuca pratensis* chromosome 4F (red), *Triticum aestivum* chromosome 4A (orange) and *Hordeum vulgare* chromosome 4H (blue), respectively.

**Figure 7, Repeats on 4F.** Proportion of different repetitive elements in sequence reads of 4F chromosome.

**Figure 8, Graph layouts visualized by SeqGrapheR.** Examples of graph layouts calculated using Fruhterman-Reingold algorithm for different types of tandem organized repeats – fpTR1 (A), fpTR5 (B) and fpTR15 (C). Dots and lines represent sequence reads and similarity hits between them, respectively.

**Figure 9, Cytogenetic mapping of repeats.** Cytogenetic mapping of newly identified tandem repeats and telomeric repeat on sorted chromosomes of meadow fescue and F1 hybrid of *L. perenne × F. pratensis*. 
**Figure 10, Southern hybridization with the probes for tandem repeats.** Examples of Southern hybridization of *F. pratensis* and *L. perenne* genomic DNA digested using three different restriction endonucleases with the probes for tandem organized repeats (A) fpTR1; (B) fpTR4; (C) fpTR5; (D) fpTR7; (E) fpTR12 and (F) fpTR15. The ladder-like pattern indicates presence of tandem organized repetitive units and supports the results of dotter analysis as well as the FISH analysis.

**Tables**

**Table 1 - GenomeZipper results**

General overview of the GenomeZipper results of 4F chromosome. All given numbers are non-redundant.

<table>
<thead>
<tr>
<th>Number of …</th>
<th>Chromosome 4F</th>
</tr>
</thead>
<tbody>
<tr>
<td>No. of anchored gene loci</td>
<td>4,672</td>
</tr>
<tr>
<td><em>Lolium</em> SNP marker with match to syntenic genes</td>
<td>109 (out of 154)</td>
</tr>
<tr>
<td>No. of anchored barely flcDNAs</td>
<td>1,697</td>
</tr>
<tr>
<td>No. of anchored 4F contigs via first-best hit</td>
<td>20,613</td>
</tr>
<tr>
<td>No. of anchored 4F contigs via bi-directional blast hit</td>
<td>3,700</td>
</tr>
<tr>
<td>No. of anchored <em>Festuca</em> ESTs</td>
<td>1,444</td>
</tr>
<tr>
<td>No. of anchored <em>Brachypodium</em> genes</td>
<td>3,056</td>
</tr>
<tr>
<td>No. of anchored rice genes</td>
<td>2,629</td>
</tr>
<tr>
<td>No. of anchored <em>Sorghum</em> genes</td>
<td>2,826</td>
</tr>
</tbody>
</table>
Additional files

Supplementary Data File

Table S1 – Summary on detection of tandem repeats
The results of PCR, FISH and Southern Hybridization detection of 15 tandem repeats (fpTRs) in *F. pratensis* and *L. perenne*.

Figure S1, Collinearity of 4F with rice. High-density comparative analyses of the linear gene order of the *Festuca pratensis* genome zipper versus the sequenced genome of barley and *Oryza sativa*. The figure includes sets of concentric circles. The inner circle represents the chromosome 4F (green) and collinear chromosomes 4H (red) and 5H (purple) of barley. The outer circles represent homologous chromosomes of rice – Os3 and Os11. The heat maps illustrate the density of genes hit by the contigs from the 4F chromosome. Syntenic regions are painted in red color tones, whereas non-syntenic regions are painted in blue. Putative orthologs between 4F and barley chromosomes 4H and 5H are connected with lines. The scale is given in gene loci for chromosome 4F and in megabase pairs for barley chromosomes 4H and 5H.

Figure S2, Collinearity of 4F with *Sorghum bicolor*. High-density comparative analyses of the linear gene order of the *Festuca pratensis* genome zipper versus the sequenced genome of barley and *Sorghum bicolor* (3). The figure includes sets of concentric circles. The inner circle represents the chromosome 4F (green) and collinear chromosomes 4H (red) and 5H (purple) of barley. The outer circles represent homologous chromosomes of *Sorghum bicolor* – Sb1, Sb2, Sb5 and Sb8. The heat maps illustrate the density of genes hit by the contigs from the 4F chromosome. Syntenic regions are painted in red color tones, whereas non-syntenic regions are painted in blue. Putative orthologs between 4F and barley chromosomes 4H and 5H are connected with lines. The scale is given in gene loci for chromosome 4F and in megabase pairs for barley chromosomes 4H and 5H.

Figure S3, Collinearity of 4F, 4A of wheat and 5H of barley. *Festuca pratensis* 4F Illumina contigs and *Triticum aestivum* 4A contigs mapped on the physical map of *Hordeum vulgare* chromosome 5H. Syntenic regions are painted in yellow-red color tones, whereas non-
syntenic regions are painted in dark blue. This approach clearly identifies collinearity of chromosomes 4F and 4A with barley chromosome 5H. The connectors used indicate orientation of the particular parts and the different positions of the breakpoints.

**Figure S4, Distribution of chromosome 4F reads in sequence clusters.** Histograms show the results of clustering based on the hierarchical agglomeration algorithm of sequence reads representing 1x coverage of chromosome 4F. The Y-axis shows number of reads in the clusters and the X-axis shows the percentage of the reads in the clusters and singletons.

**Figure S5, Dot-plot of five tandem repeats.** Dot-plot analysis of five tandem organized repeats which gave visible signals on chromosome 4F after FISH.
Figure 1, Flow karyotype. Distribution of relative chromosome fluorescence intensity (flow karyotype) obtained after analysis of DAPI-stained suspension of mitotic chromosomes of meadow fescue (*Festuca pratensis*, 2n=2x=14). The flow karyotype consists of two composite peaks I and II representing chromosomes 1F, 5F and 6F and chromosomes 2F, 3F and 7F, respectively, and a peak III representing chromosome 4F. This chromosome can be discriminated and sorted individually.
Figure 2, Model species' genes anchored to 4F. Venn diagram showing the number of Brachypodium (Bd), rice (Os) and sorghum (Sb) genes which are anchored in the virtual gene map of Festuca pratensis chromosome 4F. Overlaps of the circles show the number of genes anchored at the same locus.
Figure 3, Collinearity of 4F with barley genome. *Festuca pratensis* 4F Illumina contigs mapped on the physical map of *Hordeum vulgare* chromosome 4H and 5H. Syntenic regions are painted in yellow-red color tones, whereas non-syntenic regions are painted in dark blue. This approach clearly identifies collinearity of chromosome 4F with barley chromosome 4H and terminal part of long arm of 5H chromosome. The connectors used indicate orientation of the particular parts.
Figure 4, Collinearity of 4F with Brachypodium. High-density comparative analyses of the linear gene order of the *Festuca pratensis* genome zipper versus the sequenced genome of barley and *Brachypodium distachyon*. The figure includes sets of concentric circles. The inner circle represents the chromosome 4F (green) and collinear chromosomes 4H (red) and 5H (purple) of barley. The outer circles represent homologous chromosomes of Brachypodium – Bd1 and Bd4. The heat maps illustrate the density of genes hit by the contigs from the 4F chromosome. Syntenic regions are painted in red color tones, whereas non-syntenic regions are painted in blue. Scale of orthology between 4F and barley chromosomes 4H and 5H are connected with lines. The scale is given in gene foci for chromosome 4F and in megabase pairs for barley chromosomes 4H and 5H.
Figure 5. Chromosome 4/5 translocation. The first three figures (A, B, C) show the synteny region of chromosome 4 and 5 in barley. The variant of translocation in wheat (C) is different from those in other species (A, B) in comparison to the corresponding homoeologous chromosomes 4 and 5 in barley. The location of the star in the third circle (C) marks the syntenic conserved region between barley chromosome 4H and wheat chromosome 4A (unpublished data).
Figure 6, Conserved syntenic regions in the virtual linear ordered maps of Festuca, barley, wheat and Lolium chromosome 4 as defined by comparison to model grass genomes. The intersecting circles of the Venn diagrams depict the percentage of syntenic conserved reference genes (Brachypodium distachyon, Oryza sativa, Sorghum bicolor) shared by Festuca pratensis chromosome 4F (red), Triticum aestivum chromosome 4A (orange) and Hordeum vulgare chromosome 4H (blue), respectively.
Figure 7, Repeats on 4F. Proportion of different repetitive elements in sequence reads of 4F chromosome.
Figure 8, Graph layouts visualized by SeqGrapheR. Examples of graph layouts calculated using Fruhterman-Reingold algorithm for different types of tandem organized repeats – fpTR1 (A), fpTR5 (B) and fpTR15 (C). Dots and lines represent sequence reads and similarity hits between them, respectively.
Figure 9, Cytogenetic mapping of repeats. Cytogenetic mapping of newly identified tandem repeats and telomeric repeat on sorted chromosomes of meadow fescue and F1 hybrid of L. perenne × F. pratensis.
Figure 10, Southern hybridization with the probes for tandem repeats. Examples of Southern hybridization of F. pratensis and L. perenne genomic DNA digested using three different restriction endonucleases with the probes for tandem organized repeats (A) fpTR1; (B) fpTR4; (C) fpTR5; (D) fpTR7; (E) fpTR12 and (F) fpTR15. The ladder-like pattern indicates presence of tandem organized repetitive units and supports the results of dotter analysis as well as the FISH analysis.