Palacký University Olomouc

Faculty of Science

Department of Botany

And

Institute of Experimental Botany

Centre of Region Haná for Biotechnological and Agricultural Research

Štěpán Stočes

Study on gene expression in interspecific grass hybrids

Ph.D. Thesis

Supervisor: RNDr. David Kopecký, Ph.D.

Olomouc 2018

Acknowledgements:

I would like to thank prof. Ing. Jaroslav Doležel, DrSc for the opportunity to be a member of the Laboratory of Molecular Cytogenetics and Cytometry and the Centre of Plant Structural and Functional Genomics.

I also wish to thank my supervisor, RNDr. David Kopecký, Ph.D., for giving me the opportunity to pursue a PhD. I especially appreciate his support and guidance, his inspirational ideas, his patience and help.

My cordial thanks also go to Mgr. Jan Bartoš, Ph.D., Mgr. Eva Hřibová, Ph.D. and prof. Bruno Studer for their stimulating ideas. Further, I would like to express my thanks to all my colleagues, friends. Finally, I would like to thank my family for their patience and all their support.

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 Reference chapter.

……..

This work was supported by European Union (grant No. ED0007/01/01 Centre of the Region Hana´ for Biotechnological and Agricultural Research), Czech Science Foundation (award P501/11/0504), National Program of Sustainability (award no. LO1204), Internal Grant Agency of Palacký University (award no. Prf/2012/001) and Sciex-NMSch, a scientific exchange program between Switzerland and the new member states of the European Union (Project Code 14.099).

BIBLIOGRAPHIC IDENTIFICATION

Author's name: Mgr. Štěpán Stočes

Title: Study on gene expression in interspecific grass hybrids

Type of Thesis: Ph.D. thesis

Department: Botany

Supervisor: RNDr. David Kopecký, Ph.D.

The Year of Presentation: 2018

Abstract:

Interspecific hybrids are results of crossing two different biological species, frequently followed by polyploidization. Such allopolyploids are common among vascular plants and have played a key role in the evolution of angiosperms. Their origin is always characterized by subsequent changes in gene-expression due to a combination of homoeologous genes coming from parental species. The fate of these duplicated genes may follow different evolutionary paths. They can lose their function by genetic or epigenetic changes (Non-functionalization), acquire a new function (Neofunctionalization), or reduce the capacity to the level of expression of ancestral gene (Subfunctionalization). In practical breeding, the knowledge on the evolution of parental-specific genes and their expression changes will be beneficial for the development of elite cultivars of interspecific hybrids.

The advent of Next Generation Sequencing (NGS) enables the identification and utilization of a huge number of molecular markers and allows to study gene expression in resolution never achieved before. Single nucleotide polymorphisms (SNP) are the most abundant markers in the genome, relatively frequently distributed, easily identifiable and nowadays the most frequently used. In this study, the reference transcriptomes of *Festuca pratensis* and *Lolium multiflorum* have been reconstructed and thousands of SNPs distinguishing homoeologous genes have been identified. This enables to determine parental-specific gene expression in F1 and F2 hybrids based on the quantification of mapped reads to transcriptome references. The thesis is not only

comprehensive study for further research, but also a source of highly valuable markers, which can be used in grass breeding.

Key words: Interspecific hybrid, gene expression, Lolium multiflorum, Festuca pratensis, SNP, DArT, NGS, transcriptome

Number of pages/Apendices: 134/0

Language: English

BIBLIOGRAFICKÁ IDENTIFIKACE

Jméno: Mgr. Štěpán Stočes

Název: Study on gene expression in interspecific grass hybrids

Typ práce: Disertační práce

Obor: Botanika

Školitel: RNDr. David Kopecký, Ph.D.

Rok obhajoby: 2018

Abstrakt:

Mezidruhoví kříženci jsou výsledkem křížení dvou různých biologických druhů. Po takovéto hybridizaci často dochází ke zdvojení hybridního genomu a vzniku allopolyploidního jedince. Allopolyploidie hrála důležitou úlohu v evoluci vyšších rostlin a velká řada druhů je allopolyploidní. Tito kříženci kombinují genetickou informaci z obou rodičů a obsahují velké množství duplikovaných genů. Chování těchto genů v novém prostředí jde jen těžko odhadovat, ale existuje několik možností. Jeden z duplikovaných genů může ztratit svoji funkci, tj. být kompletně nebo částečně umlčen či odstraněn na základě genetických nebo epigenetických změn. Další možností je, že jeden duplikovaný gen získá úplně novou funkci. V posledním případě může dojít k redukci úrovně genové exprese u obou genů a výsledné množství proteinů přepisovaných z těchto genů se sníží na původní hladinu jako v rodičovském genomu. Pro šlechtění elitních odrůd mezidruhových kříženců by bylo žádoucí pochopit celý proces týkající se změn genové exprese rodičovských alel během F1 a následujících generací.

Identifikace a využití velkého množství molekulárních markerů se s příchodem nové generace sekvenačních strategií stala dostupnější a efektivnější. Díky vysokému rozlišení a dostatečné hloubce čtení se mimo jiné aplikace otevřely možnosti pro studium genové exprese. Nejrozšířenějšími molekulárními markery současnosti jsou jednonukleotidové polymorfismy (SNP), které se v genomu vyskytují v ohromujícím množství, jsou relativně pravidelně distribuované a relativně snadno detekovatelné.

V této práci byla rekonstruována sekvence transkriptomu kostřavy luční a jílku mnohokvětého a identifikovány SNP markery, které umožňují odlišit homeologní geny z obou druhů. To umožnilo identifikaci rodičovských variant genů v reciprokých F1 a F2 hybridech kostřavy a jílku (Festulolium) a stanovení druhově-specifické genové exprese u těchto kříženců. Tato práce není jen uceleným studiem změn genové expresse u prvních generací mezidruhových kříženců, ale je rovněž důležitým zdrojem markerů využitelných při cíleném šlechtění hybridních trav. Navrhovaný postup je aplikovatelný na studium genomového složení a genové exprese dalších hybridních a allopolyploidních druhů.

Klíčová slova: Mezidruhový kříženec, genová exprese, jílek mnohokvětý, kostřava luční, SNP, DArT, sekvenování nové generace (NGS), transcriptom

Počet stránek/příloh: 134/0

Jazyk: anglický

CONTENT

1. INTRODUCTION

Genetic variation is essential for the survival and adaptation of a species, because it is basic condition of natural selection and evolution. It is based on differences in alleles of genes mainly caused by random mutation like insertion, deletion for example resulting in a modification of the gene structure. Polyploidy is a major key in the evolution of plants (Wood *et al.,* 2009; Mayrose *et al.,* 2011). Allopolyploidy is the fusion of two different nuclei from different species to form a newly single egg with two complete sets of chromosomes (Van de Peer *et al.,* 2009). It is common that the contribution of parental genomes is not equal in allopolyploids resulting in the alterion of gene expression. The changes in gene expression may appear immediately after hybridization or later during the diversification of duplicated genes. Recently, the gene expression can be investigated using next generation sequencing (NGS) strategies. One group of suitable interspecific hybrids for studies of gene expression changes and interactions are forage grasses.

Grasses play an essential role in many ecosystems and they are widely used in the agricultural production. The climate change calls for the development of crops that will be resistant to biotic and abiotic stresses such as heat, cold and drought. High-yielding species of forage grasses provide high-quality forage for livestock, are used for sport and ornamental purposes, and play an important role in protecting natural environment (Beard and Green, 1993). New generation of grasses with superior characteristics could be obtained after interspecific hybridization of ryegrass (*Lolium* ssp.) and fescue (*Festuca* ssp.) species (Humphreys *et al.*, 2005). Ryegrasses are high-yielding and nutritious species with good competitiveness. However, they have low ability to survive summer drought and winter freezing (Straková *et al.*, 2007). On contrary, fescues are known for their tolerance to abiotic stresses. Fescues and ryegrasses are closely related and produce fertile hybrids called Festulolium with high level of homoeologous chromosome pairing and recombination. This allows combination of complementary agronomic traits of both genera (Kopecký *et al.*, 2005).

The aim of this thesis is to develop a simple, fast and inexpensive approach to produce new data on the gene expression in interspecific hybrids and their parental species. Bioinformatic approach has been used to reconstruct transcriptomes of Italian

ryegrass and meadow fescue. These transcriptomes were used for the identification of parental-specific gene-associated markers. These markers constitute an important resource and tool for genomic research of Italian ryegrass and meadow fescue and they are next step for detailed research to genome constitution and gene expression of Festulolium hybrids.

2. LITERATURE OVERVIEW

2.1 GENETIC VARIABILITY

Genetic variability is a key factor in the ability of the species to evolve and to survive stressed conditions (Bradshaw and Hardwick, 1989). The populations with high genetic variability are usually more flexible and have a better response to environmental changes. The source of genetic variability is based on polymorphisms at the DNA level. The main source of variation is a random process of mutations of DNA which can be deletion, insertion, inversion for example. Mutations can occur in coding or non-coding regions and thus, new mutations do not always mean new phenotype, especially in complex eukaryotic genomes. Similarly, degeneration of genetic code enables the appearance of so called silent mutations, which does not change the production of specific amino acid (Griffiths *et al.,* 2000). Another source of genetic variability is recombination. It is an exchange of the parts of non-sister chromatids of homologous chromosome pair. It is also called DNA crossover (Fig. 1). Recombination is also critical for repairing DNA damaged in mitosis (Chen *et al.*, 2008).

Genetic variation may be increased by recombination much faster than by mutations. However, recombination is limited in asexual organisms and bacteria (Griffiths *et al.,* 2000). Genetic variability can be also increased by the movement of individuals between populations. This phenomenon is called gene flow or gene migration. It changes the frequency of alleles in the populations and sometimes introduces new alleles. Moreover, genetic material can be transferred between organisms via horizontal gene transfer. This plays a major role in the evolution of many organisms like bacteria (Amábile-Cuevas and Chicurel 1993), but can be seen also in plants (Richardson and Palmer, 2007). Genetic variability can be also increased by polyploidy and interspecific hybridization. These processes will be further discussed in details.

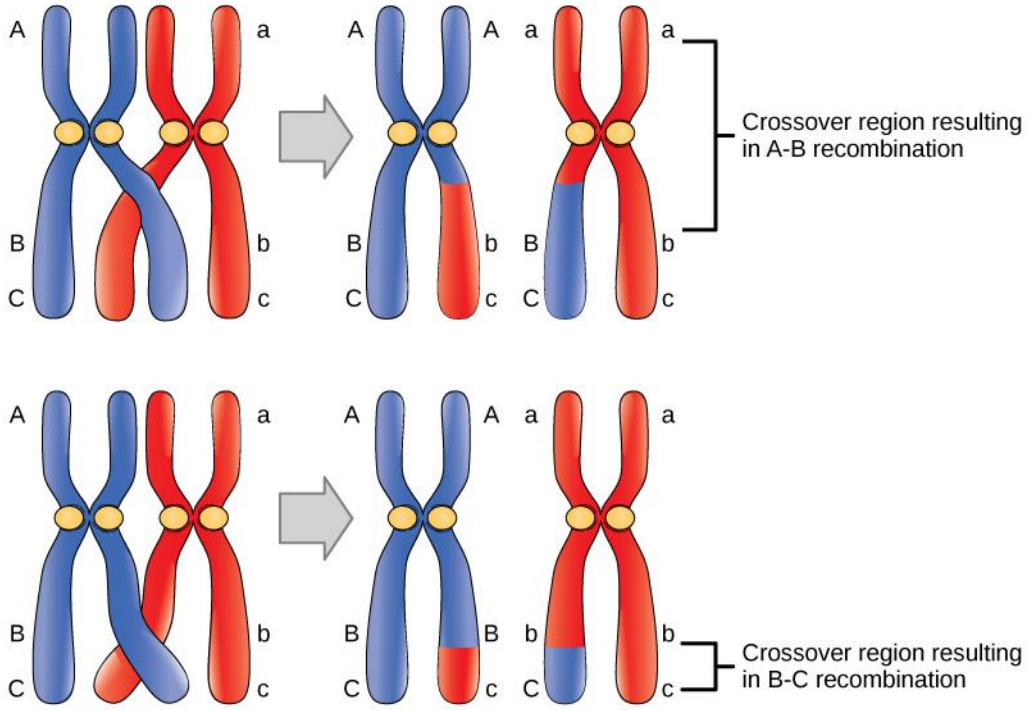


Fig. 1 *- DNA crossover. Recombination between genes A and B is more frequent than recombination between genes B and C because genes A and B are farther apart. (Shea, T., 2014)*

2.1.1 POLYPLOIDY

Polyploidy is a condition when organism has more than two complete sets of chromosomes. It is more common in plants, but also appears in animals (mainly fish and amphibians). In fact, a lot of species which are diploid underwent polyploidization event during their evolution (Van de Peer and Mayer, 2005). The most probable mechanism of polyploidization is based on incorrect meiotic formation of gametes. Unreduced gametes leading to polyploidy formation obtain complete set of chromosomes unlike a normal state (reduced gametes). If a diploid gamete fuses with another diploid gamete, tetraploid zygote is formed. In principle, it is possible to distinguish between two basic types of polyploids. Autopolyploids have chromosome sets from a single species and allopolyploids from two different species. The specific cases of allopolyploids are amphiploids, which contain complete chromosome sets from both parents (Rieger *et al.*, 1968). The typical example of allopolyploids is *Triticale*. Hexaploid *Triticale* has six chromosome sets, four from durum wheat, and two from rye.

The advantages of polyploids are heterosis, gene redundancy, frequent losing of self-incompatibility and gain of asexual reproduction. Heterosis involves the fixing of divergent parental genomes in allopolyploids. The offspring has twice more copies of genes which allow replacement of dysfunctional genes. The bigger amount of genes can be used in new way during time in evolutionary selection (Adams and Wendel, 2005). Another advantage of polyploidy is that polyploids are sometimes able to disrupt self-incompatibility system and become self-fertile (Comai *et al.*, 2000). On the other hand, polyploidy also faces to some disadvantages. The polyploid cell is forced to expend more energy during mitosis and meiosis and the bigger amount of DNA can also lead to the improper segregation of chromatids. Polyploidy also induces changes in gene expression and may lead to epigenetic instability (Mittelsten Scheid *et al.*, 1996). Moreover, structural genomic changes are reported to allopolyploid plants as wheat (Shaked *et al.*, 2001).

2.1.2 INTERSPECIFIC HYBRIDIZATION

Interspecific hybridization is defined as a reproduction between organisms of two different species resulting in a new hybrid individual. Thus, it is a mechanism prerequisite to the formation of allopolyploids. It therefore plays an important role in the evolution and speciation. Combination of diverse genotypes potentially produces genes that contribute to reproductive isolation and selection of individual alleles during many generations of recombination and help to constitute a new species (Barton and Gale 1993). The hybridization can occur between subspecies within one species (intraspecific hybrids), between species within one genus (interspecific hybrids) and between two genera (intergeneric hybrids). Hybridization between two species or between two genera is always rare on a per-individual basis, but very common on a per species basis. About 10 % of animal and 25 % of plant species hybridize regularly (Mallet, 2005). Interspecific hybridization is also frequently used in breeding. Hybrid cultivars are formed by crossing usually 2, 4 or more parental components with good combination abilities to achieve heterosis. The phenomenon of heterosis was first discovered by J.G. Kölreuter in 1763. Heterosis enhances fitness and vitality. Interspecific hybridization increases genetic variation much faster than the mutation and recombination and frequently leads to the origination of new species (Abbott *et al.*,

2013; Mallet, 2007). Despite the frequency of interspecific hybrids found in nature, their fertility is usually supressed due to many reproductive barriers.

2.1.3 HYBRIDIZATION IN PLANTS

Plants hybridize with higher frequency than animals. The importance of hybridization in the evolution of plant species was recognized by many botanists long time ago (reviewed in Grant 1981; Rieseberg 1997). There are two types of interspecific hybrids: homoploids (mainly for animals) and allopolyploids (mainly for plants). Homoploid hybridization is when the offspring have the same ploidy level as the two parental species, while allopolyploid hybridization is an addition of the chromosome number of two parents (Fig. 2). Most of the interspecific hybrids are allopolyploids (Fig. 3) (Tate *et al.,* 2005) and allopolyploidy was confirmed as a main mechanism of speciation in flowering plants (Leitch and Bennett, 1997). Allopolyploids seem to appear more frequently at places with dramatic climate change (Mayr, 1963; Coyne and Orr, 2004; Fawcett et al., 2009). Thus, allopolyploid formation is probably a response of plants on the abiotic (drought, extreme temperatures, high salinity, soil pollutants) or biotic stresses (viruses, parasites, fungi, bacterium) (Chen, 2007).

Allopolyploid formation is frequently followed with the changes in gene expression by the processes of sequence elimination, activation of transposons, methylation, gene silencing and intergenomic conversion (Comai, 2000; Wendel, 2000). Interspecific hybrids possess chromosome sets from two species in one nucleus. It frequently leads to inability of these chromosomes to pair in meiosis and thus, to partial or complete sterility. The fertility can be restored by duplication of chromosome sets and formation of allopolyploid. Then, each chromosome has its homologue and pair easily in meiosis resulting in viable gametes. Thus, allopolyploids are generally more fertile than homoploid hybrids.

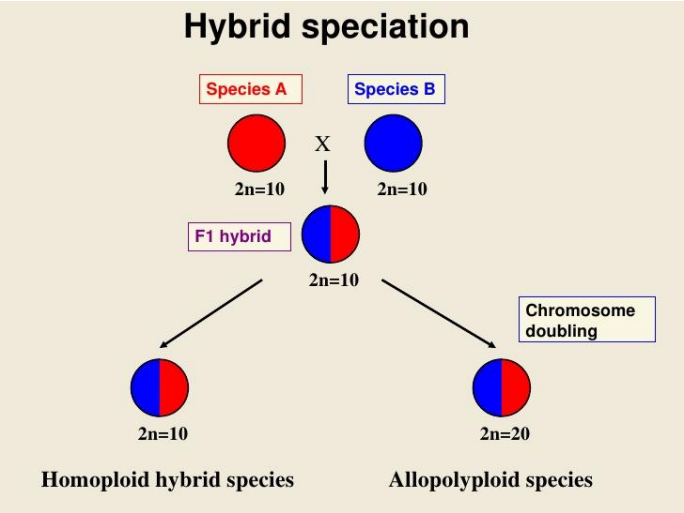


Fig. 2 *- Illustration of hybrid speciation. Members of two different species hybridize and produce F1 hybrid. Hybrid can remain the same chromosome number as parental species (2n=10) – homoploid hybrid – or whole genome duplication can appear, which results in the doubling of chromosome number and such hybrid is called allopolyploid or amphiploid. The picture is from Richard Abbott talk „Plant Introductions & Evolution: Hybrid Speciation and Gene Transfer"*

Besides ancient allopolyploids, newly formed hybrids occur in the nature as Tragopogon (Lipman *et al.,* 2013). As was mentioned earlier, allopolyploidy is also responsible for the formation of many crops including wheat and *Brassica* (Feldman and Levy, 2005; Chalhoub *et al.*, 2014). Interspecific hybridization is usually used in breeding to introgress new alleles from wild relatives and increase genetic variability of crop species. Such example is Festulolium, a hybrid of fescues (*Festuca* sp.) and ryegrasses (*Lolium* sp.). Fescues and ryegrasses hybridize in nature, but their hybrids are sterile. However, using tetraploid parental plants for initial cross leads to the development of fertile hybrids. Such approach has been used in breeding programs and many Festulolium cultivars have been released (Kopecký *et al.*, 2006). Festulolium became popular among farmers due to the combination of agricultural characteristics inherited from both parents (such as yield and nutrition characteristics from ryegrasses and drought tolerance and winter hardiness from fescues). Despite the widespread use of Festulolium and other allopolyploids and interspecific hybrids, their genomic constitution and expression of genes from parents is frequently unknown.

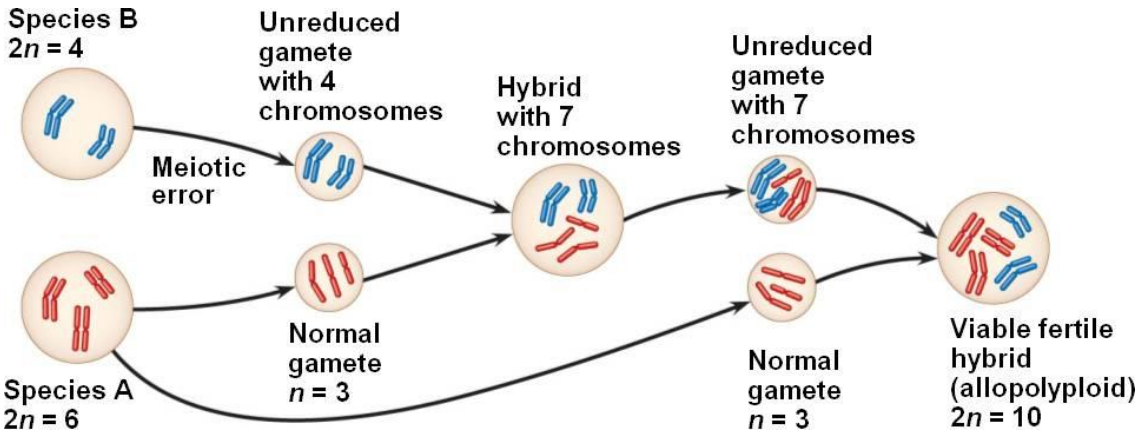


Fig 3. *– An example of origination of fertile allopolyploid*

[http://bio1151.nicerweb.com]

2.1.4 METHODS ENABLING IDENTIFICATION OF INTERSPECIFIC HYBRIDS

It is usually difficult to identify interspecific hybrids and their progenitors based on the morphological characteristics. On the other hand, genetic and genomic techniques provides excellent tool to study genomic composition, evolution and diversity of the plant species including allopolyploids and interspecific hybrids. In the next chapters, the most frequently used methods will be described in details with the focus on the grass hybrid identification.

2.1.4.1 Cytogenetic techniques

Cytogenetic methods as fluorescent *in situ* hybridization (FISH) and genomic *in situ* hybridization (GISH) enables determination of parental species of interspecific hybrids as well as analyses of genomic composition. *In situ* hybridization techniques are based on hybridization of labelled DNA or RNA sequences (probes) and target sequence in sample (mostly in a form of squashed chromosomes or nuclei on microscopic slide). The probes are prepared using radio-, fluorescent- or antigen-labelled bases and they are localized using autoradiography, fluorescence microscopy, or immunohistochemistry. GISH uses total genomic DNAs from potential parents labelled by fluorescent dyes (fluorochromes) as probes (Figure 4) (Schwarzacher *et al.*, 1989).

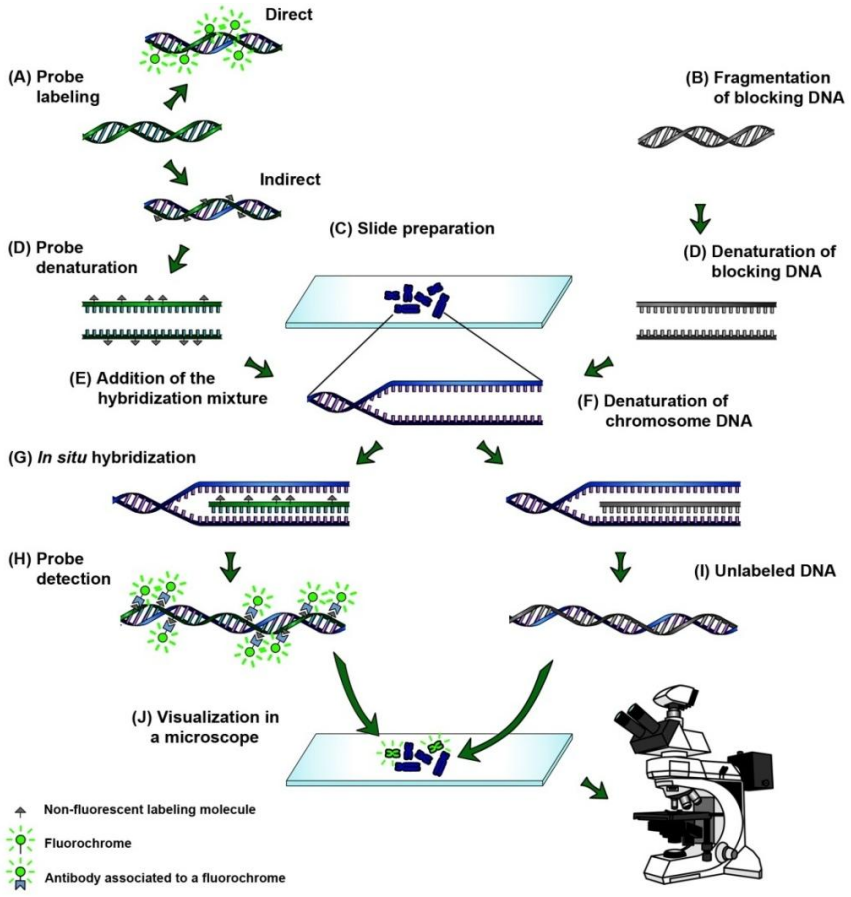


Figure 4 - *Overview of genomic in situ hybridization (GISH). (A) Direct or indirect probe labelling (from genomic DNA of one parent). (B) Fragmentation of the blocking DNA (from genomic DNA of the second parent). (C) Microscopic slide preparation with metaphase chromosomes. (D) Probe and blocking DNA denaturation in a hybridization mixture. (E) Addition of the hybridization mixture. (F) Denaturation of the chromosome DNA. (G) In situ hybridization of probe and blocking DNA with the complementary sequences of the chromosomes. (H) Detection of the probe in the metaphase chromosome spreads representing one of the hybrids. (I) Blocking DNA is hybridized with the complementary sequences of chromosomes and represents the genome of the second progenitor. (J) Visualization of hybridization signals associated to a probe (green). Unlabelled chromosomes are visualized with a counter-staining (blue). When the probe labelling is direct, the detection step of the GISH can be excluded. The fluorochromes are the signalling molecules and can be directly visualized in a fluorescence microscope with the appropriate filter (Brammer et al., 2013).*

In case of Festulolium hybrids, it is possible to distinguish chromatin of fescue from ryegrass in squashed metaphase chromosomes of hybrids fixed on microscopic slides. Besides characterization of genomic composition and karyotyping, it allows studying homoeologous chromosome pairing and recombination (Kopecký *et al.*, 2008). GISH is very effective and simple method for studying the structure and evolution of

hybrid genomes. However, it has limited resolution and low capacity, which prevents wider use of this method in practical breeding (Kopecký *et al.*, 2009).

2.1.4.2 Genetic Markers

Genetic marker is DNA sequence, which can be used to distinguish different populations, species and individuals. Its location in the genome is usually known and can be described as variation in genome. Genetic markers are used to study genetic distance between individuals, relationship between inherited diseases, they enable identification of recombination breakpoints, localization of specific genomic regions, genotyping, analysis of phylogenetic relationships and marker assisted selection (MAS). Ideally, markers are associated with specific trait or gene, are polymorphic and easily identifiable. The density of markers is usually essential for some applications. High density of markers for example enables genome-wide association studies (Hirschhorn and Daly, 2005). In the past, quantity of markers was limited. With the advent of next-generation sequencing (NGS), it is nowadays possible to discover not tens or hundreds, but thousands or millions of markers across any genome (Stapley *et al.,* 2010). The most frequently used markers will be described in detail in following chapters.

2.1.4.2.1 RESTRICTION FRAGMENT LENGTH POLYMORPHISM (RFLP)

RFLP markers are based on restriction fragment length polymorphism (Grodzicker et al., 1975). Homologous sections of DNA are restricted with specific enzymes. These restriction endonucleases are able to recognize specific sequence of DNA and subsequently cleave it. Once there is a polymorphism in this specific sequence, DNA is not cleaved. The fragments are separated and transferred to nylon membrane, where the hybridization of individual fragments with specific probes occurs (Southern blot). RFLP was firstly used for identification of DNA sequence polymorphisms of a temperature-sensitive mutation of adeno-virus serotypes (Grodzicker et al., 1975). Botstein *et al.* (1980) used RFLP for human genome mapping. Since then, these markers have been widely used in plant genome research (Helentjaris *et al.*, 1986; Weber and Helentjaris, 1989). The restriction enzymes are essential for RFLP analyses. The greatest resolution is obtained using 'four-cutters' (enzymes recognizing four base pair sequence) because there are many such sites in the genome (Semagn *et al.*, 2006). Most common ‘four-cutter’ enzymes are combined with 'six-

cutter' enzymes. The use of two enzymes instead of one increases the number of possible sites in the genome. The 'six-cutter' enzymes are easily available and the fragments of DNA are in the size range of 200bp to 20kb, thus ideal for electrophoresis. The advantage of RFLP markers is that the knowledge on sequence information is not necessary. Moreover, they are reproducible and well transferable between laboratories. On the other hand, using RFLP markers is time consuming and of low-capacity for which it is necessary to obtain high quality and pure DNA in sufficient quantities. The RFLP markers were used in forage grasses of *Festuca-Lolium* complex for phylogenetic and genetic analyses (Xu and Sleper 1994; Charmet *et al.*, 1997; Yamada and Kishida 2003).

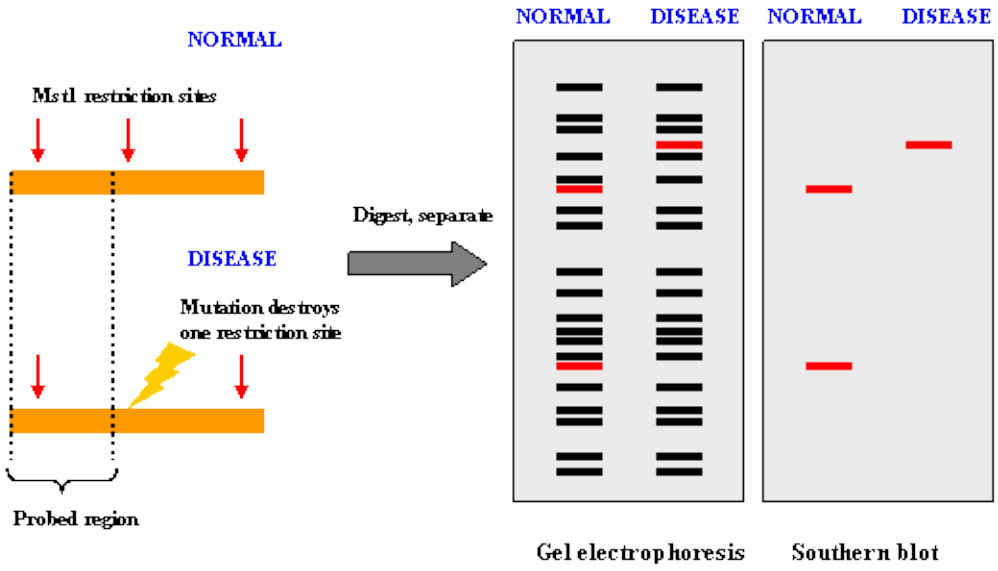


Figure 5 *– Normal and disease samples were digested on specific sites by Mst1 restriction endonuclease. One restriction site of disease sample has been destroyed by mutation. The bands with specific length were visualized using gel electrophoresis.*

[http://www.ncbi.nlm.nih.gov/probe/docs/techrflp/]

2.1.4.2.2 AMPLIFIED FRAGMENT LENGTH POLYMORPHISM (AFLP)

Amplified fragment length polymorphisms (AFLP) are markers based on restriction digestion and polymerase chain reaction (PCR). Total genomic DNA is digested using two restriction endonucleases and resulting fragments are amplified using PCR. Polymorphisms are identified by the presence or absence of DNA fragments unlike RFLP markers. They are visualized on agarose or polyacrylamide gel or using

automatic DNA sequencers. AFLP markers are more efficient than RFLPs. They are highly reproducible (Jones *et al.*, 1997) without necessity of DNA sequence information. One of the major advantages of AFLP is the capability to detect more than point mutations compared to RFLPs (Becker *et al.*, 1995). On the other hand, ALFP analysis is laborious. It needs template DNA free of inhibitor compounds that interferes with the restriction enzyme. AFLP markers are dominant and thus, they are unable to differentiate between dominant homozygotes and heterozygotes. This reduces the accuracy in population genetic analyses, genetic mapping, and marker assisted selection (MAS) (Semagn *et al.*, 2006). Another disadvantage is that AFLP markers have low reproducibility. AFLP markers were used for analysis of genetic diversity in perennial ryegrass (Guthridge *et al.* 2001; Roldan-Ruiz *et al.,* 2002)

2.1.4.2.3 RESTRICTION SITE ASSOCIATED DNA MARKERS (RAD)

RAD markers are based on the reduction of genomic representation in contrast with RFLP and AFLP. Genomic DNA is digested by restriction enzyme during the isolation (Baird *et al.,* 2008). The flanking DNA sequences are used around the restriction site, called RAD tag (Miller *et al.*, 2007). RAD tags are prepared so that DNA is digested by specific enzyme followed by the ligation of biotinylated adapters with overhangs. The DNA is randomly cut into fragments smaller than average distance between restriction sites and biotinylated fragments are isolated using streptavidin beads (Lewis *et al.*, 2007). This method was developed for preparing RAD tags to microarray analysis (Miller *et al.*, 2007). Isolation procedure has been modified and nowadays, it is used for high-throughput sequencing on the Illumina platform and known as Reduced-representation sequencing (RADseq) (Davey *et al.*, 2011). The mechanism of RADseq is also based on digestion of DNA by restriction enzymes. Adapter P1 with overhangs is ligated. DNA is randomly cut into fragments smaller than distance of restriction sites with blunt ends. Adapter P2 is ligated into blunt ends. The resulting fragments are amplified using PCR and sequenced by next generation sequencing methods (Figure 7).

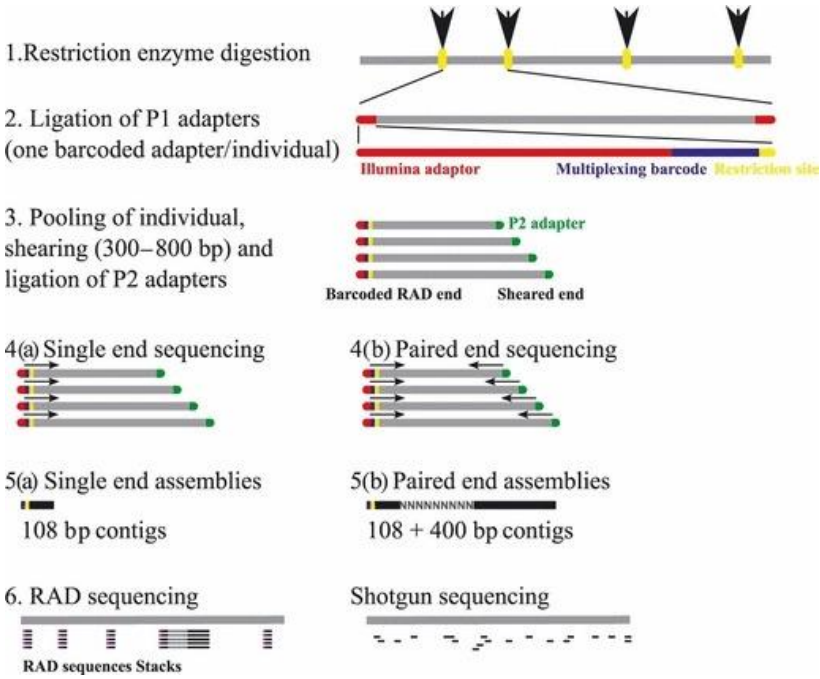


Figure 7 - *Overview of RAD sequencing. (1) DNA is digested with a restriction enzyme. (2) A modified adapter containing the Illumina P1 amplification and sequencing primer and a DNA barcode is ligated to the fragments. (3) Samples are pooled, sheared into 300- to 800-bp libraries (required for Illumina sequencing) and ligated to a second adapter P2. (4) Sequencing is performed either as (a) single end (one sequence of 36-108 bp per fragment) or (b) paired end (two sequences of 36-108 bp per fragment). (5a). Barcoded sequences are assembled into overlapping stacks. (5b) Given that restriction fragments are sheared randomly, paired-end sequencing allows the assembly of larger contigs on the sheared end of the fragment, whose size depends on the length of the Illumina sequences and the size of the sheared fragments isolated (see Etter et al. 2011 for more details and additional RAD sequencing strategies). (6a) Reduced representation of the genome through RAD sequencing, (b) as compared to shotgun sequencing. Partially re-drawn from (Etter et al. 2011).*

Brant *et al.* (2012) modified RADseq and developed double digest RADseq (ddRADseq). The ddRADseq method combines the throughput of sequencing and genotype analysis based on the short reads sequence data. Compare to RADseq, ddRADseq has higher flexibility and robustness in region recovery. RAD markers can be useful for QTL-mapping, association mapping, population genetics and phylogenetic studies.

2.1.4.2.4 VARIABLE NUMBER TANDEM REPEAT (VNTR)

Tandemly repeated sequences are dispersed by thousands of copies in all higher eukaryote genomes. The repeats are formed into patterns of nucleotide sequences

repeated directly to each other (tandemly organized) and oriented in the same direction. Tandem repeats can be expanded or reduced during recombination or replication as errors, leading to the variable number of tandem repeats (VNTR) (Nakamura *et al.*, 1998). Since the number of copies of individual tandem repeats is specific for individuals, species or populations, VNTRs can be used as genetic markers. Based on the size of the repeat motif (number of nucleotides being tandemly repeated) and number of copies, tandem repeats are divided into microsatellites, minisatellites and satellites. Tandem repeats having motifs of 1-5bp with about 5 to 50 copies are called microsatellites. Microsatellites occur at thousands of locations in the genome, but predominate in some regions such as telomeres (Richard *et al.*, 2008). Minisatellites have motif of 6-25 bp creating blocks up to 30kb length and are generally GC rich. They have been implicated as regulators of gene expression (Rose *et al.,* 2016). Minisatellites have been extensively used for DNA fingerprinting as genetic markers in population studies (Vergnoud and Denoeud, 2000). . The satellites (satDNA) are tandemly repeated DNA sequences with motif longer than 25 bp forming large blocks up to 100 Mb. They are frequently localized near to centromeres and telomeres in regions of heterochromatin.

Detection of VNTR (mainly minisatellites and microsatellites) is based on the amplification of specific tandem repeat by PCR. The products are separated on a gel electrophoresis and the size of fragments is determined by the number of repeating segments (Figure 8).

Tandem repeats are codominant, locus-specific, highly polymorphic and hypervariable markers and enable high-throughput screening (Tautz, 1989). Disadvantages of tandem repeats (and especially microsatellites) are occurrence of stutter bands, null alleles, homoplasy and too many alleles at specific region of genome, which demand very high sample size for analysis (Mohindra *et al.,* 2001).Microsatellites have been used to estimate genetic diversity of Festulolium (Pivoriene and Pašakinskiene, 2008).

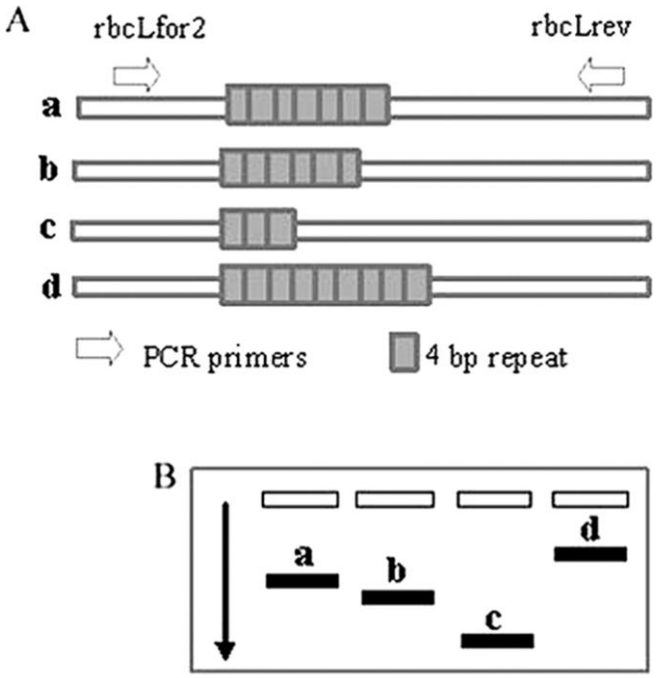


Figure 8 *- Visualization of DNA changes at the chloroplast microsatellite reporter. (A) Diagram of microsatellite indels with PCR primers. The original out-of-frame construct (OF28) contained seven 4-bp repeats (a). Deletion of one repeat (b) or four repeats (c) or insertion of two repeats (d) made the insertion in frame. (B) Diagram of differential mobility of PCR products in electrophoretic gel (GuhaMajumdar et al., 2014).*

2.1.4.2.5 DIVERSITY ARRAYS TECHNOLOGY (DArT)

DArT is based on microarray hybridization technique and enables detection of several hundreds to thousands polymorphic loci in the genome. The genomic representations are prepared using restriction enzymes (methyl-sensitive to achieve high frequency of markers from coding regions). The genomic DNA is digested and adapters are ligated. The genome complexity is reduced by PCR with primers complementary to the adapters. The fragments are cloned and inserts are amplified by vector specific primers. Subsequently, amplified inserts are purified and arranged on the solid support called microarray. The genomic representation is labelled and hybridized to the array (Jaccoud *et al*., 2001). The polymorphic clones show specific signal intensities for different individuals and they are assembled into genotyping array.

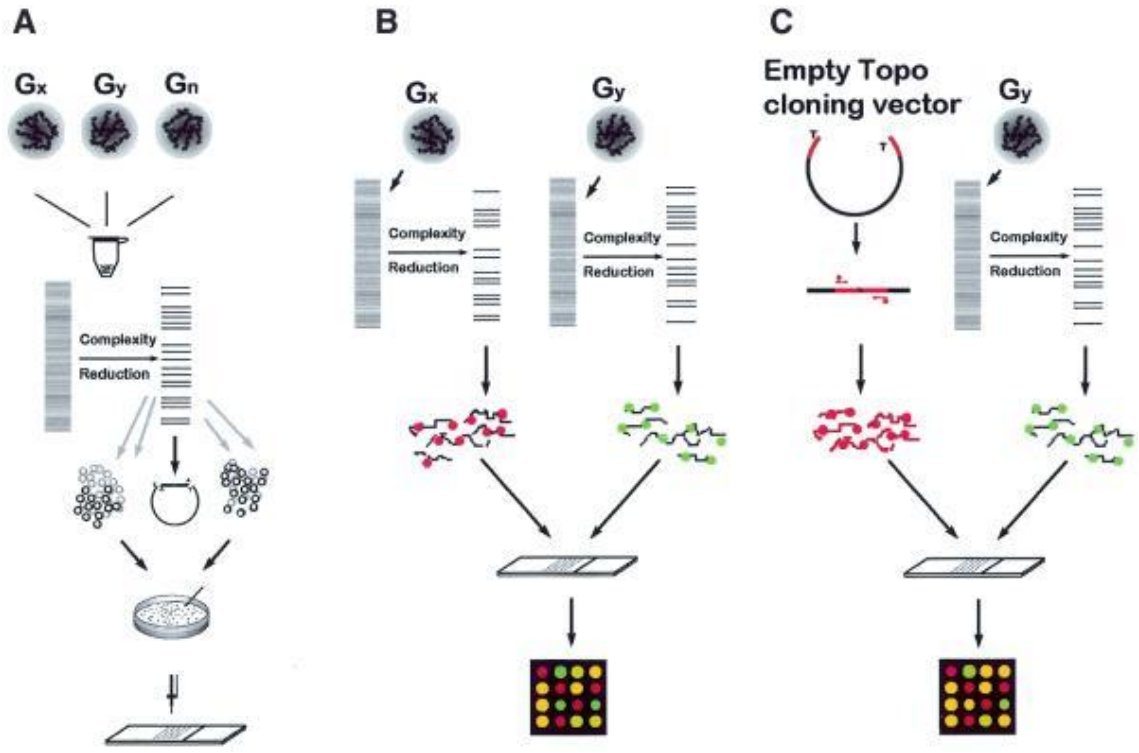


Figure 9 - *Schematic representation of DArT. (A) Generation of Diversity Panels. Genomic DNAs of specimens to be studied are pooled together. (B) Contrasting two samples using DArT. Two genomic samples are converted to representations using the same methods as in (A). (C) Genetic fingerprinting using DArT. The DNA sample for analysis is converted to a representation using the methods as in (A) and labelled with green fluorescent dye. Polymorphic spots are identified by binary distribution of signal ratios among input samples.*

As for RFLP and AFLP, DArT markers do not require any DNA sequence information. DArT is high-throughput, low-cost, quick and reproducible. However, the limitation of this method is that DArT markers are dominant. The original DArT is implemented on the microarray platform, which involves fluorescent labelling of genomic representations and their hybridization to dedicated DArT arrays. New version of this technique is called DArTseq. The DArTseq method combines genome complexity reduction with next generation sequencing. Kopecký *et al*. (2009) developed DArTFest array for *Festuca-Lolium* complex. It has been used to analyse genetic diversity (Kopecký et al., 2009), genetic mapping (Bartoš et al., 2011, Dierking et al., 2015) and the analysis on the genomic composition of Festulolium hybrids (Kopecký et al., 2011).

2.1.4.2.6 SINGLE NUCLOTIDE POLYMORPHISM (SNP)

The most frequently used markers of these days are single nucleotide polymorphisms (SNPs). SNP is a variation in a single nucleotide on specific location in genome and occurs to > 1% rate. SNPs are the most prevalent type of polymorphisms and thus informative on account to their wide dispersal across the genome and their high density especially in allogamous species such as fescues and ryegrasses. SNPs are present in both coding and non-coding regions of the genome. However, the genomic distribution of SNPs is heterogeneous and most of the SNPs are present in non-coding regions (Sachidanandam *et al*, 2001). The average frequency in human genome is 1 SNP per 1000 nucleotides (Chakravarti 1999). In generally, SNPs occur more frequently in regions where natural selection is acting. The development of SNP markers allows automatizing genotype analysis with increased effectiveness (Khlestkina and Salina, 2006). SNPs are used in association studies, haplotype mapping, linkage disequilibrium analyses and for many other applications (Cardon and Bell, 2001; Flint-Garcia *et al*., 2003). There are several methods used for SNP detection including single-strand conformation polymorphism (SSCP) (Goszczynski and Jooste, 2002), single base extension (Podini and Vallone, 2009), high resolution melting curve analysis (Birrer *et al*., 2014), capillary electrophoresis (Drabovich and Krylov, 2006) and primarily DNA sequencing (Altshuler *et al.*, 2000; Berthouly-Salazar *et al*., 2016). The major advantages of SNPs are that they are virtually unlimited, evenly distributed along the genome, bi-allelic and co-dominant. In past, massive SNP discovery was limited to a few species with available reference genome sequence, such as maize (Mammadov *et al*., 2010) and Arabidopsis (Weigel and Mott, 2009). Recently it is possible to identify SNPs without genome reference sequence using bioinformatics tools such as OGA (Ruttink *et al.*, 2013) However, the discovery of SNPs via next generation sequencing in complex genomes with high frequency of repeated sequences is still expensive. To overcome this limitation, modern approaches reducing genome complexity by methyl-filtration resulting in the enrichment of coding regions has been developed (Ersoz *et al.,* 2012, Blanca *et al.,* 2012). Many transcriptome studies lead to the development of SNP markers applicable for genetic mapping in cereals (Jones et al., 2009; Trebbi *et al.,* 2011) oil crops (Delourme *et al.*, 2013; Choi *et al.,* 2007), and model legumes (Loridon *et al.,* 2013; Yang *et al.,* 2011).

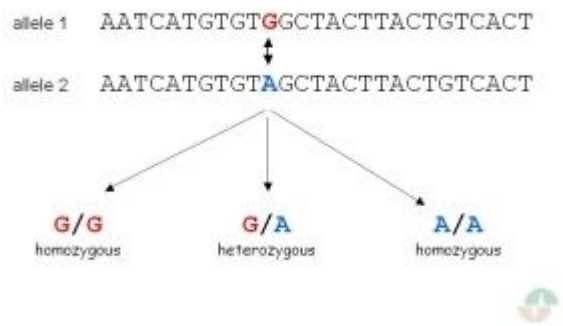


Figure 10 *- Single nucleotide polymorphism (SNP)*

[http://www.traitgenetics.com]

2.2 GENE EXPRESSION

The process of gene expression leads to the synthesis of functional gene products based on the gene coding sequence. The DNA is firstly transcribed to RNA and subsequently the RNA is translated to protein sequence. Transcription of eukaryotic genes produces primary messenger RNA (pre-mRNA), which is modified into a mature mRNA. In principle, non-coding regions in genes (introns) are removed from pre-mRNA (RNA splicing) (Collins, 2011). Since eukaryotic genes often contain more introns, there can exist more variants of pre-mRNA splicing resulting in different mRNAs (alternative splicing). The regulation of gene expression is more frequently at the transcription level (Lee *et al.*, 2002). Products of genes are most frequently proteins. However, different types of RNA (tRNA, small nuclear RNA and rRNA) can also be the product of non-protein coding genes. Variations in gene expression explain the phenotypic differences in cells. The volume of gene products is sensitive to a large number of subjects and presence of stochasticity in gene expression is also the source of bias in gene expression (Raser and O'Shea, 2005). Simultaneously, the expression of one gene can influence the expression of another.

2.2.1 MECHANISM OF GENE EXPRESSION

The mechanism of gene expression involves several consecutive steps. Genetic information on particular DNA regions is transcribed in RNA strand (mRNA), which is then most frequently translated into protein. The mRNA carries information about

protein coding sequence. The enzyme Helicase unwinds the DNA strands (transcription bubbles) and the enzyme RNA polymerase copies the DNA template into the pre-mRNA. (Alberts *et al.*, 2002). The strand containing the gene sequence is called the sense strand, while the complementary strand is the antisense strand. The pre-mRNA molecule is complementary to the antisense DNA strand and contains introns which are not required for protein synthesis. The pre-messenger RNA is modified, introns are removed during the process of splicing and only gene coding regions (exons) are kept (Black, 2003). In alternative splicing, exons are spliced to several different possible mRNA products. Thus, a single gene can have thousands of differentially spliced transcripts. Splicing is very important in regulation of gene expression (Bingham *et al.,* 2003). Finally, the 5' end of mRNA is capped by addition of 7-methylguanosine ("capping").and a poly(A) tail (about 250 adenine residues) is added to 3' end during the post-transcriptional process. The cap plays a role in the ribosomal recognition of messenger RNA while the poly(A) is important for the stability and the nuclear export of the mRNA (Brodsky and Silver, 2000). Then, mRNA is transported into the cytoplasm to the ribosomes, where translation is starting. Ribosomes consist from two subunits. These two subunits form together 80s particle in eukaryotic (Fig. 11) and 70s particle in prokaryotic cells, referring to their sedimentation coefficients in Svedberg units. Ribosomes contain ribosomal RNA (rRNA) and different types of proteins (Wool IG, 1979). The structure of ribosomes is conserved in all organisms with only small variations. Ribosomal RNA is responding for functionality of proper translation of mRNA to proteins. First, ribosomes assemble around the mRNA and transfer RNA (tRNA) binds at the start codon. The tRNA is composed from 76-90 nucleotides and serves as a link between mRNA and amino acid sequence of proteins. Each triplet corresponds to specific site (based on genetic code) for tRNA which carries one amino acid. The ribosomes are moving along mRNA codons and continuing with process of creating an amino acid chain (Campbell 1996). The large subunit of ribosome leads to the formation of a peptide bond in a nascent protein, polypeptide. Subsequently, the amino acids obtain three-dimensional structure using chaperones. Chaperones are specific proteins which help with the control of conformation (Ellis 1991). This is essential for the functionality of the proteins

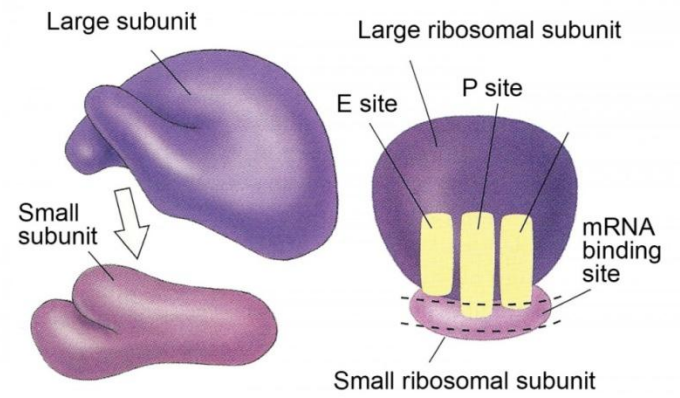


Figure 11 *– The structure of ribosome, large subunit on the top of the small subunit. [http://mol-biol4masters.masters.grkraj.org/html/Protein_Synthesis5-Ribosome_as_Translation_Machine.htm]*

2.2.1.1 Gene Expression in Interspecific hybrids

Gene expression is complicated process involving many molecular interactions. The genome of interspecific hybrids originates from the fusion of two or more different genome species (polyploidization event) and thus, their gene expression is highly altered compare to parental species (Riddle and Bichler, 2003). Many studies observed expression bias to one parental genome compare to the second (Flagel *et al.,* 2008; Wang *et al.*, 2006). One copy of duplicated genes can be decommissioned by genetic and epigenetic changes (non-functionalization), or gains novel functionality (neo-functionalization), or both copies of gene are reduced to the expression level of ancestral gene (sub-functionalization) (Ma and Gustafson, 2005). The genetic and epigenetic changes are frequently caused by activation of transposons and DNA methylation. The transposable elements form 50-80% of plant genome (Feschotte *et al.,* 2002) and in normal stage are mostly inactive. They are activated as a response to stress conditions (Jiang *et al.,* 2003) and genomic shock (McClintock 1984). Such genomic shock can be realized by the origination of interspecific hybrid combining two evolutionary divergent genomes. It has been discovered that some genes have higher expression in F1 hybrids than in the parental species in interspecific hybrids of Arabidopsis (Comai *et al.*, 2003), Drosophila (Ranz *et al.* 2004) and maize (Auger *et al.* 2005). This supports the hypothesis, that genetic regulatory elements are *cis-trans*-regulatory coadaptive and stress induced by presence of two incompatible genomes. Similarly to interspecific hybridization, polyploidization also alters gene expression

(Force *et al.*, 1999). Generally, more copies of one gene produce more gene product (dosage effect). According to duplication-degeneration-complementation (DCC) hypotheses, the duplicated genes obtain degenerative mutations, which reduce the gene activity on the level of ancestral gene (Force *et al.*, 1999).

2.2.2 Regulation of Gene Expression

The gene expression and its regulation ensure that correct complement of RNA and proteins is present in the right cell at the right time. It is a very complex set of interactions between genes, RNA molecules, proteins and other components, when and where specific genes are activated and in what quantities. Although the multicellular organisms are genetically homogeneous, each cell is structurally and functionally heterogeneous based on the difference in gene expression. The gene regulation needs the activity of trans-acting factors and *cis*-acting elements (Mignotte *et al.,* 1989). The *cis*-acting elements are localized in DNA sequences near to gene and they are required for gene expression. The *trans*-acting factors are considered to be proteins, which bind to the *cis*-acting sequences to control gene expression. These factors have influence on adaptation and divergence between species (Wray 2007; Fay and Wittkopp 2008). The *cis*-regulatory variants mostly have quantitatively larger effects on expression than *trans*-regulatory variants (Zhang et al. 2011; Gruber et al. 2012) with some exceptions (Emerson et al. 2010). Also, *cis*-regulatory variants have additive effect on expression. On the other hand, *trans*-regulatory variants have typically dominant or recessive effect (McManus et al. 2010, Zhang et al. 2011; Gruber et al. 2012). The typical examples of *cis*-regulatory elements are promoters and enhancers (Levin, 2010). The genes are activated or repressed by transcription factors (TF), which bind to specific regions on DNA (promoter regions). The regulation of genes is different between prokaryotes and eukaryotes. For prokaryotes, almost every regulation has negative character and turns the activity of genes off. In contrast, expression in eukaryotes is more complicated and moreover, tissue specific.

2.2.2.1 Regulation of Transcription

Regulation of transcription is controlled primarily in the stage of initiation of transcription, before the stretch of DNA is transcribed into RNA molecule. However, transcription may be regulated also during the entire process (Hobert, 2008).

Transcription in eukaryotic cells is controlled by several proteins. They bind to specific sequences and modify the activity of RNA polymerases. In contrast, regulation of gene expression is modulated using a single protein in prokaryotes (Pulverer, 2005). Genes transcribed by RNA polymerase II have two promoter elements, the TATA box and the initiator motif (Inr) sequence. These elements served as binding sites for transcription factors (TFs). They are frequently located upstream of the TATA box (Smale and Kadonaga, 2003). The promoters are 100-1000 bp long and located a prior to the gene and can be regulated by elements many kilobases distant. Other regulatory sequences are called enhancers. Enhancers are short regions of DNA located up to 1Mbp far from the gene and they are sites for the binding of proteins called activators, which enables activating gene expression (Ren, 2010). The first enhancer was identified by Walter Schaffner in 1981 during studies of the promoter virus SV40. Enhancers, like promoters, serve as binding sites for TFs and thus, regulate RNA polymerase (Pennacchio *et al.,* 2013). This is possible because of DNA looping, which allows a transcription factor bound to a distant enhancer to interact with RNA polymerase or general transcription factors at the promoter (Mercer and Mattick, 2013). The binding of specific transcriptional regulatory proteins to enhancers is responsible for the control of gene expression during development and differentiation, as well as during the response of cells to hormones and growth factors. An important aspect of enhancers is that they usually contain multiple functional sequence elements that bind different transcriptional regulatory proteins. These proteins work together to regulate gene expression.

The TFs are divided by their function on activators and repressors. Transcriptional activators are consisted of two domains. One region of the protein specifically binds to DNA and activates transcription by binding to promoter or enhancer sequences determined by the specificity of their DNA-binding domains. The other activates transcription by interacting with other components of the transcriptional machinery. On the other hand, transcriptional repressors bind to specific DNA sequences and inhibit transcription. Some repressors contain the same DNA-binding domain as the activator but lack its activation domain (Fujimoto *et al.*, 2000). As a result, their binding to a promoter or enhancer blocks the binding of the activator, thereby inhibiting transcription.

2.2.2.2 Regulation of Translation

Gene expression is regulated at multiple levels, including the translation of mRNAs into proteins. Compared to transcriptional regulation, translational control of existing mRNAs allows for more rapid changes in cellular concentrations of the encoded proteins. It can be used for maintaining homeostasis and in addition to modulating more permanent changes in cell physiology (Holcik and Sonenberg, 2005). However, regulation of translation is less prevalent than control of transcription. Most regulation is exerted at the initiation of translation. The start codon is identified and decoded by the special methionyl tRNA (Met-tRNAi). The mRNA is activated after binding of initiation factors (eIFs). The eIFs recognize the mRNA's cap structure at the 5′ or the poly(A) tail at the 3′ end. This process can be down regulated by reducing the activities of the eIFs that stimulate Met-tRNAi recruitment to the 40S subunit (Sonenberg and Hinnebush, 2009). The chemical substances as toxins and antibiotics prevent the formation of the initiation complex of ribosome and mRNA (Leonard *et al.,* 1988; Swaney *et al,* 1998), also facilitate blocking the specific sites for binding tRNA (Slover *et al.,* 2007), blocking elongation of the polypeptide or causing premature release of the peptide chain (Levinson and Warren, 2008).

2.2.2.3 Regulation using epigenetic modification of DNA

The gene expression can be also regulated by epigenetic factors modifying the DNA and histones (Szyf, 2009). The epigenetic factors have influence on gene expression of organism without changes to the underlying DNA sequence. The epigenetic processes are essential for alterations in gene expression during development and cell proliferation, but they can arise in adult cells as random changes or under environmental influences. One of the most known epigenetic regulations is the histone modification. The modulation of chromatin structure by epigenetic modification of histones is key component of posttranslational regulation of gene expression. One of the basic modifications of histones is methylation. Methylation of histones can increase or decrease transcription of genes according the type of the methylated amino acid. Histone methylation is essential for assembling of heterochromatin. The histone can be modified by editing chains of amino acids. Methylation of lysine in histone H3 leads to transcriptionally silent chromatin (constitutive heterochromatin). Methylation of histones is carried out by histone methyltransferases (HMT) and histone demethylases

(HDM). The histone lysine methyltransferase (KTM) is responsible for this methylation activity in the pattern of histones H3 and H4 (Jenuwein *et al.,* 1998). In contrast, acetylation of histones is associated with an open chromatin conformation that facilitates transcription (Cheng and Blumenthal, 2010). The acetylated N termini protruding from the nucleosome core provide reduced affinity for the DNA, allowing the chromatin to adopt a more relaxed structure for the recruitment of the basic transcription machinery. The increase in histone acetylation is associated with decreased histone deacetylase (HDAC) activity. In contrast, the recruitment of HDAC leads to gene repression. Next epigenetic modification of histones is phosphorylation. Phosphorylation plays a major role in DNA damage response (Rosseto *et al*, 2010). DNA damage induces genome instability and promotes tumorigenesis. Cells have mechanisms responsible for recognition of DNA damage and activation of cell cycle checkpoints leading to DNA repair. DNA damage response (DDR) proteins are interacting with phosphorylated carboxy terminus of histone H2, which are designated place of damage (Stucki *et al*., 2005). H3 phosphorylation is involved in chromatin relaxation and regulation of gene expression. Phosphorylated histones are associated with gene expression often related with regulation of proliferative genes (Lau *et al.*, 2011). Last modification mentioned in this work is ubiquitination and deubiquitination. Ubiquitin is covalently attached to a target protein through an isopeptide bond to C-terminal of glycine. Histone ubiquitination is a reversible modification, like that of acetylation and phosphorylation. Ubiquitin plays an important role in regulating transcription either through proteasome-dependent destruction of transcription factors or proteasome-independent mechanisms (Conaway *et al*. 2002). The second way is modification of DNA by covalent addition of a methyl group from the methyl donor S-adenosylmethionine (SAM) to a cytosine base within the DNA. Methylated DNA is very often distributed on CpG dinucleotides (Laird, 2003). DNA methylation in the promoter region of a gene is associated with decreased transcriptional activity (Robertson and Wolffe, 2000), because methylated DNA inhibit the binding sites for TF. However, Suzuki and Bird (2008) observed gene transcription with highly methylated starches of DNA in the promoter. DNA methylation is a highly dynamic process that is fully reversible. Methylation of cytosines can also persist and being inherited from parents into the next generation (genetic imprinting).

2.2.2.4 Regulation by RNA processing

A specific type of regulation is undoubtedly the RNA-mediated epigenetic regulation of gene expression, called RNA interference. The RNA interference (RNAi) is a process of inhibition of gene expression by some types of RNA. There exist two main types of small RNA molecules having an ability of RNAi: microRNA (miRNA) and small interfering RNA (siRNA) (Ghildiyal and Zamore, 2009). The RNAi is usually performed by enzyme called dicer, which digests exogenous double-stranded RNA (dsRNA) into small fragments around 21 nucleotides long. The miRNAs are endogenous and purposefully expressed products of an organism's own genome, whereas siRNAs are primarily exogenous in origin, derived directly from the virus or transposon. Second, miRNAs process from stem-loop precursors with incomplete double-stranded character, whereas siRNAs were found to be excised from long, fully complementary dsRNAs.

The mechanism of gene expression regulation by siRNA is starting by unwinding dsRNA in two single-stranded RNAs (ssRNAs). Then ssRNA is incorporated into the RNA-induced silencing complex (RISC) (Figure 12). The RISC is ribonucleoprotein complex and the complete structure is still unknown (Sontheimer, 2005). Most often it contains proteins from family Argonaut, which are the catalytic centre of RISC (Hall, 2005). RISC uses the siRNA to bind target complementary region of mRNA transcripts and degrade it into the form of cytoplasmic bodies (P-bodies) (Sen and Blau, 2005). The miRNA is created in form of primary transcripts (pri-miRNA) from long gene coding RNA. These pri-miRNAs are processed to 60-80 nucleotide loop structure known as pre-miRNA. The pre-miRNA contains RNase III enzyme and dsRNA-binding protein (Voinnet, 2009). The pre-miRNA is cleaved by Dicer and final miRNA is integrated into the RISC complex. Then the regulating is the same as in case of siRNA.

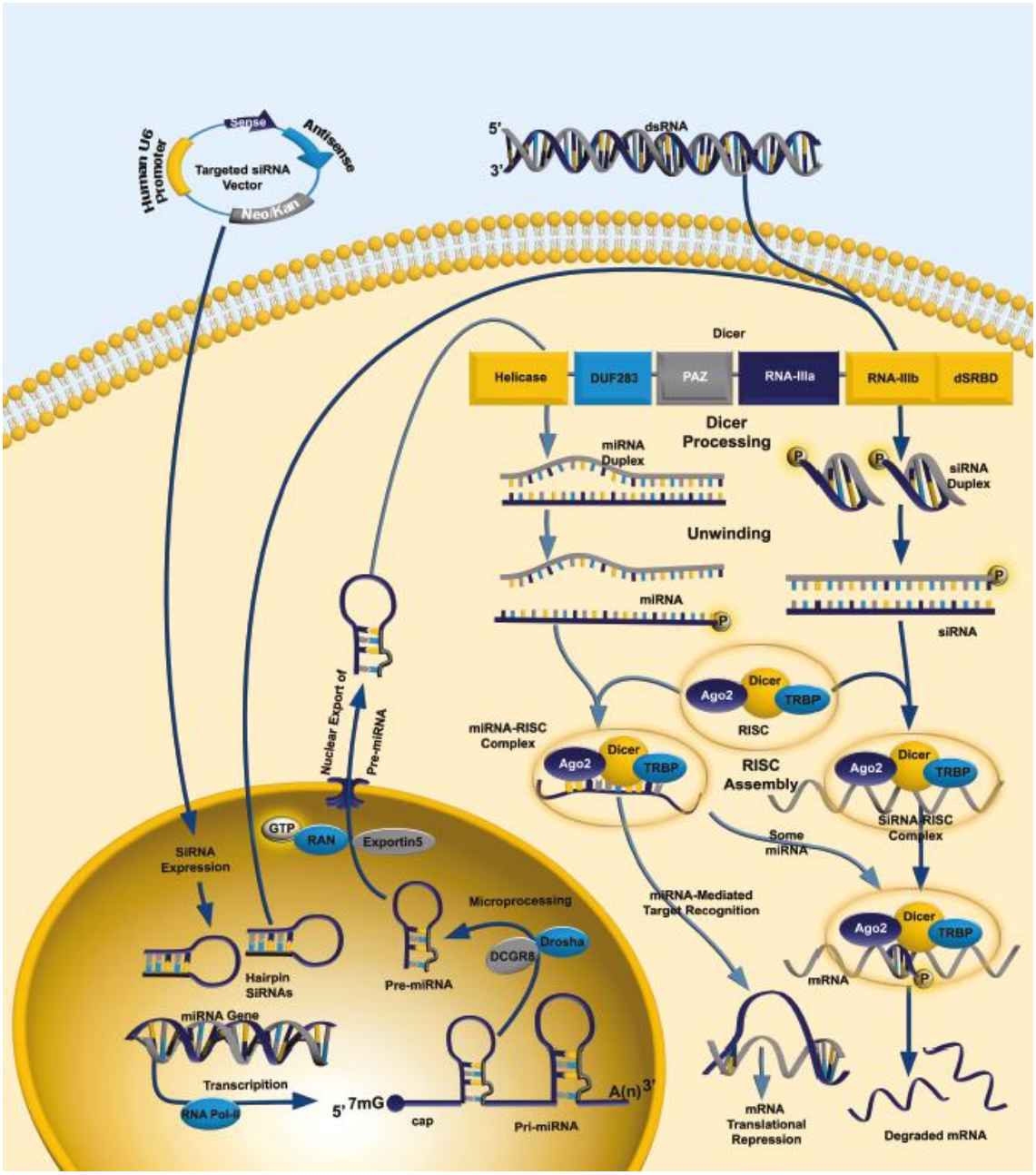


Figure 12 - *Long dsRNA molecules are cleaved to produce siRNA by Dicer. siRNA molecules are then incorporated into a multiprotein RNA-inducing silencing complex (RISC). The duplex RNA is unwound leaving the anti-sense strand to guide RISC to complementary mRNA for subsequent endonucleolytic cleavage*

[http://www.abcam.com].

2.2.3 MEASUREMENT OF GENE EXPRESSION

There are several methods evaluating the gene expression and its level. One of the simplest ways is to convert RNA to cDNA using reverse transcription, followed by amplification of cDNA by PCR and loading into gel. Differential expression is displayed (DD) on the gel as different intensity of the bands (Shen *et al.*, 1995). This approach does not require any knowledge about the genome. However, this method is not very precise and has low sensitivity. Only a fraction of transcripts can be analysed by this way in single reaction. The results contain indispensable amount of false positive cases and level of automation is limited.

2.2.3.1 Serial analysis of gene expression

The more sophisticated and accurate method is Serial analysis of gene expression (SAGE). In this method, cDNA is bound to Streptavidin beads via interaction with the biotin and then is cleaved using anchoring enzyme (AE). The cleavage sites are individual for each cDNA. The cleaved cDNA is discarded and the rest of fragments are split in two groups according to the one of two adapters. These adapters contain sticky ends with AE, a recognition site for target enzyme (TE) and primer sequence unique to adaptor. TE cleaved the bond between cDNA and beads after adapter ligation. Furthermore, fragments are ligated together with both adaptors. These "ditags" are PCR amplified using anchor adaptors and specific primers followed by another round of fission by AE and ligation with other ditags. This creates a cDNA concatemer (Figure 13). These concatemers are transformed into bacteria genome and replicated.

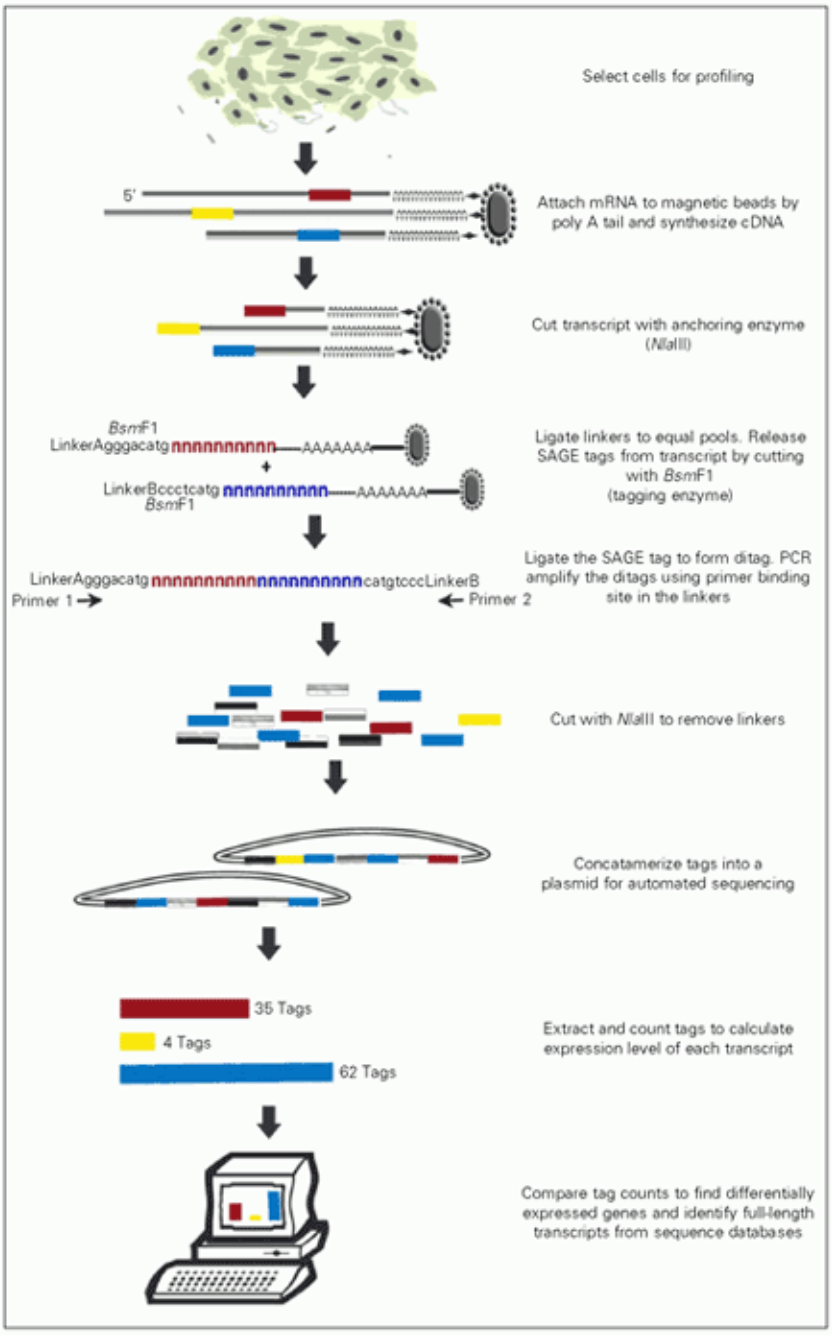


Figure 13 *- The principle of serial analysis of gene expression (SAGE). Gene expression profiles are determined from cells of interest by first capturing their mRNA using oligo-dT-coated beads and then preparing the cDNA. The anchoring enzyme NlaIII is used to cleave cDNA that remains attached to the beads. Linkers, which contain a site for tagging enzyme BsmF1 and primers, are ligated to the cDNA. The BsmF1 is used to release a short tag. These tags are paired into ditags, amplified by PCR, cut with NlaIII, ligated to form concatamers, and cloned into a sequencing vector for efficient counting on an automated sequencer. Tag counts from each tissue type are stored electronically and used for comparison to other cell populations. The relative fraction of each transcript can be calculated as well. Informatics is used to match the SAGE tag to a known gene or expressed sequence tag (Cerutti et al., 2003).*

The results are quantified using assessment of expressed gene by sequencing small specific sequence tags (Velculescu *et al.,* 1995). This method is suitable for the analysis of large number of genes, but lacks the flexibility to easily incorporate different subsets of the transcriptome into the analysis or to focus on smaller subsets of genes for more detailed analyses (Wang *et al.,* 1999). Therefore, new approach called rapid analysis of gene expression (RAGE) has been developed. It enables to analyse expression changes for most of detected genes in paired bitag preparations that differ in some biologically meaningful way. The cDNA is synthetized with biotinylated oligo as primer, digested with a restriction endonuclease and received fragments are bond to a streptavidin-coated bead. The adapter compatible with sticky end is ligated to cDNA fragments. These are digested with second restriction endonuclease and second adaptor is added by ligation (Figure 14). This reduces sequence complexity ~15-fold relative to the cDNA population. The RAGE is able to detect gene with expression level < 0.01% or ~30 molecules of mRNA per cell (Wang *et al.,* 1999).

2.2.3.2 Quantitative PCR in real time

Another way, how to measure gene expression is based on quantitative PCR in real time (RT-qPCR). In simple terms, RNA is reverse transcribed to cDNA. The cDNA molecules are monitored during PCR. The PCR products are detected in two ways. Fluorescent dye is intercalated into any cDNA. The fluorescence is measured during each cycle of PCR reaction (Pfaffl 2000). If the cDNA products increase during PCR, then fluorescence intensity also increases. The second way is to use sequence specific DNA probes, which are labelled with fluorescent reporter permitting detection after hybridization with complementary sequence. Fluorescent reporter monitors only one sequence in single experiment. This limitation can be circumvented using different-coloured labels for monitoring more sequences in the same tube. The PCR reaction follows standard way except the reporter probe is added. The cDNA amplification is initiated from the primers. The polymerase reaches probe and degrades, resulting in an increase in fluorescence. The increase in fluorescence is quantified in each cycle of PCR reaction. This method can be used for relative as well as absolute quantification of gene expression (Dhanasekaran *et al.,* 2010).

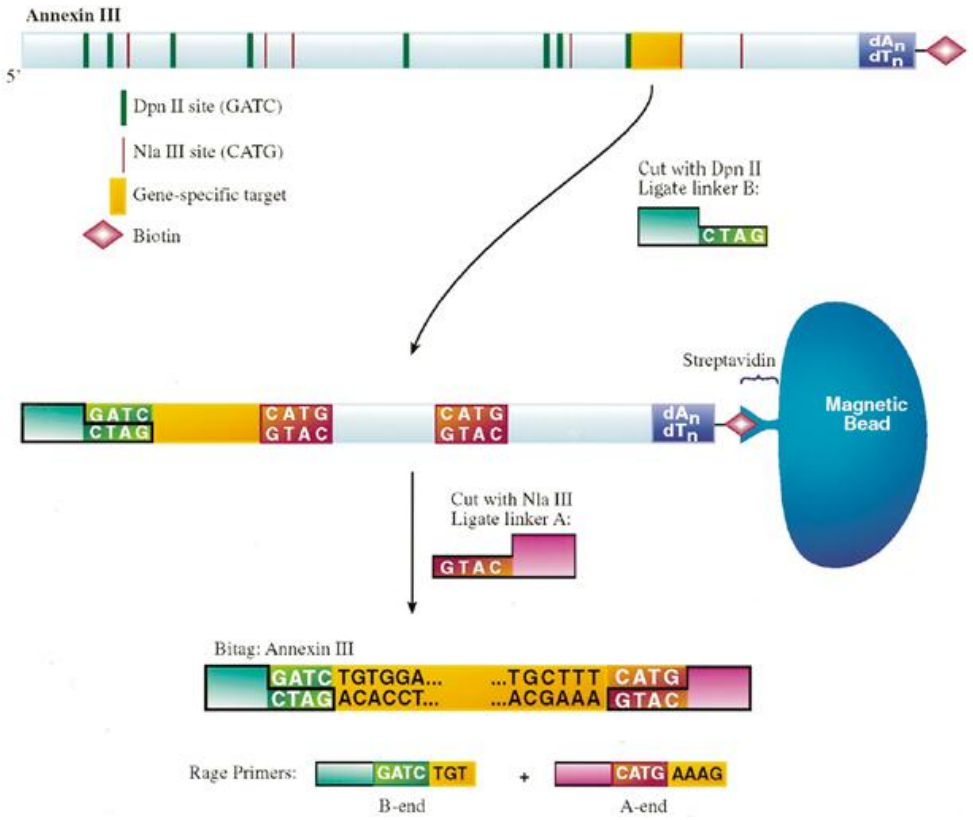


Figure 14 *– Preparation of bitag templates. The method for preparation of bitags is illustrated for the Anx3 gene. cDNA was synthesized with biotinylated oligo (dT) as first strand primer, digested with a frequent-cutting restriction endonuclease (DpnII), and the 3´ fragment recovered by binding to a streptavidin-coated bead. To provide a common priming sequence, a 16-bp adapter with a DpnII compatible sticky end (B-linker) was ligated onto the cDNA fragments. The fragments were then digested with a second frequent cutting restriction endonuclease (NlaIII) and a second common priming site (A-linker) was added by ligation. This procedure resulted in a template preparation (B/A bitags) that contained a single gene-specific target sequence from each cDNA. (Wang et al., 1999).*

Absolute quantification provides the exact number of cDNA molecules in reaction compare to DNA standards using a calibration curve. The precise amount of the message or template used for the curve is known. For accuracy of the procedure, it is necessary to have the same amplification efficiency of PCR for sample and standard. In contrast, the relative quantification is determined by fold changes in expression between two samples. The reference sample is simply known to contain the message of interest in high abundance, but its absolute amount is not necessarily known. The relative quantification calibration curve result for the gene of interest is normalized to that of a reference gene in the same sample, and then the normalized numbers are compared between samples to obtain a fold change. Reference gene has to be primarily

stable (Brunner *et al.,* 2004). Generally, RT-PCR is not ideal for multiplexing (10's to 100's of genes at maximum), requires high technical skill and expensive equipment.

2.2.3.3 Northern Blotting

Another method of measuring RNA abundance is called Northern Blotting and the principle is similar to the Southern Blotting, which is used for measuring DNA abundance (Trayhurn 1996). In Northern Blotting, mRNA is separated on agarose gel, transferred to a nitrocellulose membrane, denatured and hybridized with complementary DNA, which is labelled by radioactive isotopes (Figure 15). Finally, the intensity of band indicates the RNA abundance. Total RNA is extracted from cells, separated by size on gel or denatured into single stranded RNA. Then, it is transferred to nylon membrane (blotting membrane). The transfer takes place via buffer which usually contains formamide (Lee *et al.,* 1992). The RNA is immobilized on membrane. The membrane is treated with a pieces of RNA (probes), which are designed to have a complementary sequence to RNA sequence in the sample. This allows hybridization of the probe to a specific RNA fragment. The probe is labelled with radioactive isotopes (^{32}P) or with chemiluminescence (alkaline phosphatase or horseradish peroxidase). It permits to detect the RNA molecule of interest from other different RNA molecules on the membrane. The signal is detected by X-ray film and quantified by densitometry. The Northern Blotting is inexpensive and sensitive method quantifying transcript abundance with almost infinite dynamic range. On the other hand, Northern Blotting is not high-throughput and is nowadays replaced by other modern technologies as RNA microarrays.

2.2.3.4 RNA microarray

One of the most powerful approaches for studying gene expression is RNA microarray (sometimes also called RNA chip or biochip) (DeRisi *et al.,* 1996). RNA microarray is collection of microscopic RNA spots which are attached to a solid surface. Recently, there exist two main types of arrays, with microscopic RNA spots directly on a surface or on coded beads. The original type is solid-phase array containing thousands of microscopic spots. The surface of microarray is commonly from glass, plastic or silicon. The alternative type of array is based on beads. The bead array is a collection of microscopic beads, each with a specific probe and a ratio of two

dyes, which do not interfere with the fluorescent dyes used on the target sequence. The oligonucleotide fragments, cDNAs or small fragments of PCR products that correspond to mRNAs (probes) are attached in small amounts on surface solid-phase array (deposition strategy).

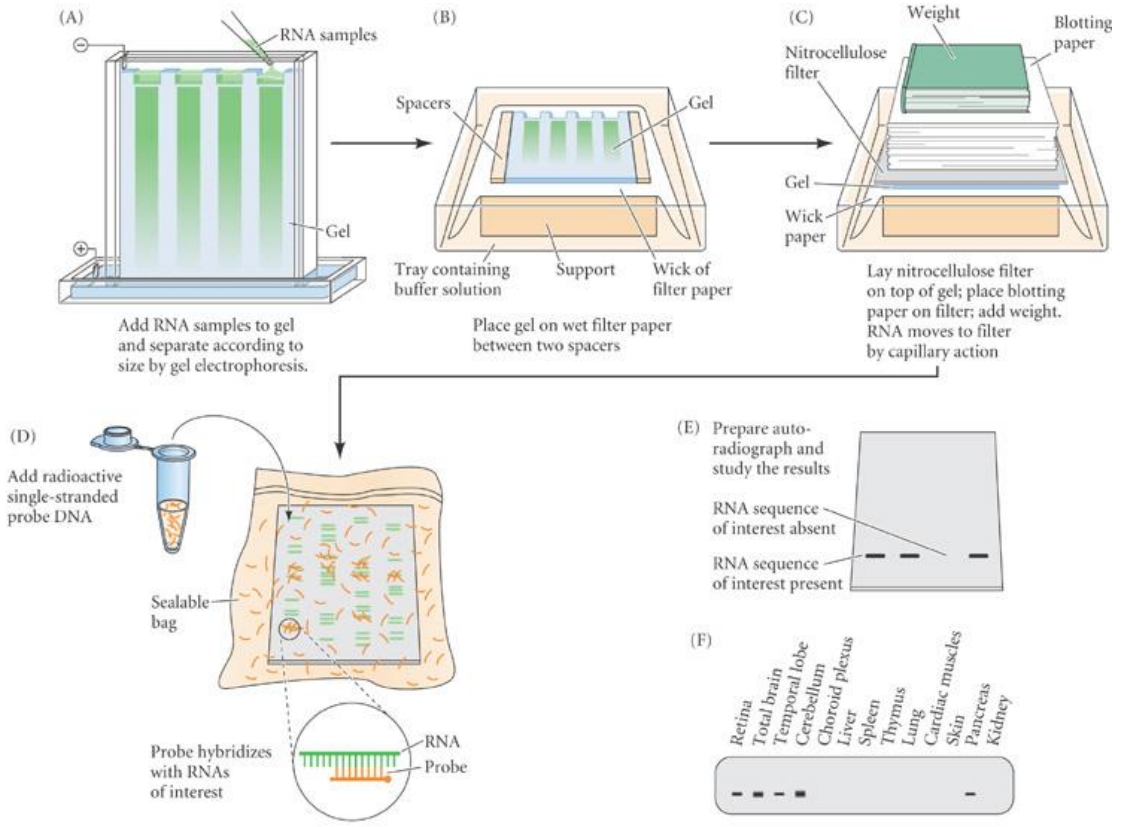


Figure 15 *- Northern blotting procedure. (A) RNA is isolated from various tissues and is separated by size using gel electrophoresis. (B) The gel is then placed on a paper wick, which absorbs an ionic solution from a trough. (C) A filter that traps RNA is placed above the gel, and blotting paper is placed above the filter. Capillary action draws the solution through the gel, trapping the RNA on the filter. (D) The filter is incubated with radioactive single-stranded DNA complementary to the mRNA of interest. (E) After any unbound DNA is washed off, mRNA in the samples is localized using autoradiography. (F) Drawing of a developmental Northern blot showing the presence of Pax6 mRNA in the eye, brain, and pancreas of a mammalian embryo.*

[http://10e.devbio.com/image.php?id=528]

These probes are often chosen from databases (GenBank, UniGene, dbESt…) or are selected randomly from any library of interest. The size of probes is in range from 0.5 to 2 kb and each probe usually represents a single gene. The mRNA to be analysed is purified, reverse transcribed, and amplified into single-stranded antisense RNA

(aRNA). Labelled nucleotides are incorporated during amplification and labelled aRNA is digested. Digested aRNAs are hybridized with single-stranded probes based on their complementarity. The unbound aRNAs are washed out. The labelled aRNAs are hybridized to probes and the microarray is scanned by laser and emission level of the fluorescence is measured.

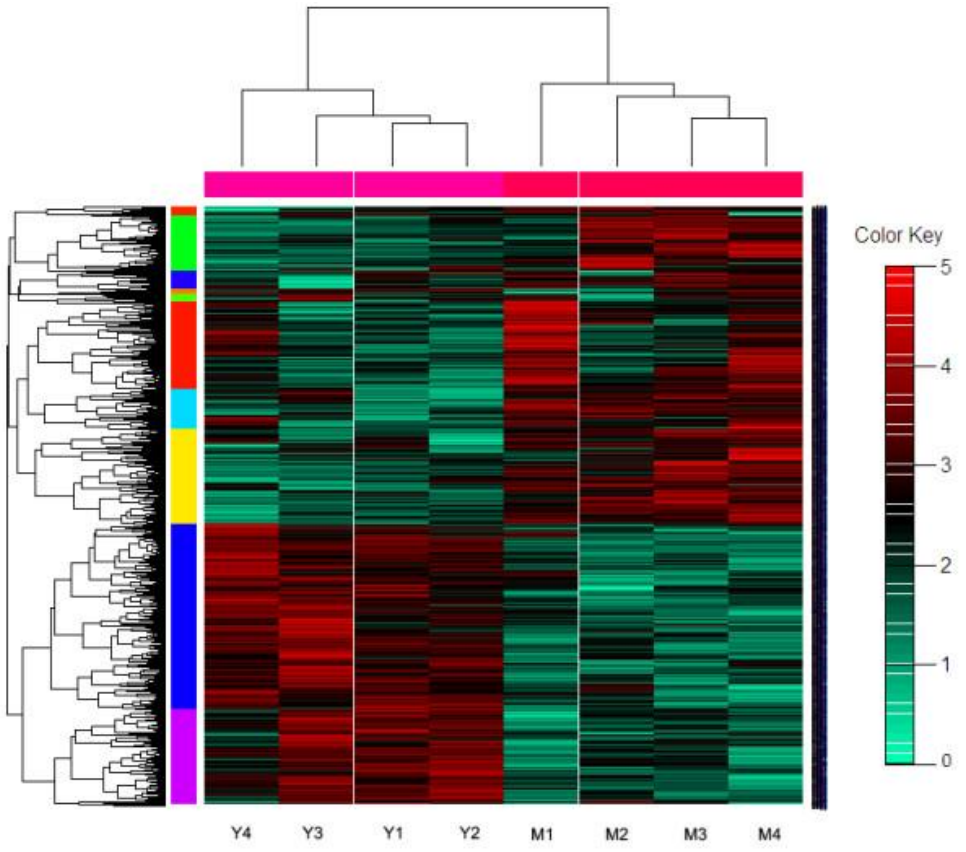


Figure 16 - *Hierarchical heat map of differential gene expression. A hierarchical cluster heat map showing the log2 transformed expression values for expression array following hybridization of mRNA prepared from samples Y and M. The individual differentially expressed genes of samples are shown horizontally. Green colour indicates lower gene expression and red colour indicates higher expression. The genes included are those with q value less than 0.1.*

[https://www.researchgate.net/figure/260376751_fig1_Hierarchical-heat-map-of-differential-gene-expression-as-determined-by-Affymetrix]

The intensity of the signal correlates with the amount of sample (aRNA) binding to the probes on spot. The resulting data are statistically evaluated and visualized as heatmap (Figure 16). Quantification is based on the comparison of intensities of fluorescent signals between samples from different conditions. Disadvantage of microarrays for analysis of gene expression is a potential presence of cross-

hybridization artefacts, poor quantification of lowly and highly expressed genes and the necessity of knowledge on the target sequence for designing probes (Kukurba and Monthomery, 2015).

2.2.3.5 Next generation sequencing technologies

The most up-to-date approach to study gene expression is provided by the advent of next generation sequencing technologies (NGS). The sequencing of RNA (RNA-seq) (Nagalakshmi, 2008) has ability to rapidly produce billions of bases of sequence data and analyse the whole transcriptome (whole transcriptome shotgun sequencing), post-transcriptional modifications, alternative gene spliced transcripts, gene fusion and changes in gene expression (Morin *et al.*, 2008; Maher *et al.*, 2009). There are several NGS platforms, among them Illumina (Solexa) sequencing (Bennett, 2004), Ion Torrent next-generation sequencing (Life Technologies), PacBio Sequencing (Rhoads and Au, 2015) and Oxford Nanopore's sequencing technology are the most commonly used. The huge amount of sequence data required (Mikheyev and Tin, 2014) development of new programs for sequence analysis and opened new interdisciplinary field of science called bioinformatics (Korf, 2013). Bioinformatics combines statistics, informatics and engineering to analyse biological data.

One of the most critical steps in the accuracy of gene expression analyses using RNAseq is a preparation of library. One of the most commonly used library for analyses of transcriptomes is RNA 'Poly(A)' library. The substrate (usually magnetic beads) contains covalently attached poly(T) oligonucleotides which hybridize with poly(A) tail of mRNA and the rest of RNA is washed off (Mortazavi *et al.*, 2008). Once the library is created and the RNAseq is realized, bioinformatics approach is needed for the analysis of gene expression. If annotated reference transcriptome or genome exists, a data analysis follows steps: a) read mapping, b) counting reads, c) normalization and d) detection of differentially expressed genes. The read alignment tools are used for read mapping. First, the tools make index of the reference, which enables quick retrieval of positions in the reference sequence. Then, the reads are aligned on the indexed reference based on the algorithms of alignment tools (Hatem *et al.*, 2013; Flicek and Birney, 2009). Since NGS produces sequencing errors, mapping tools (algorithms) has to enable imperfect alignments and hence, increase the percentage of mapped reads (Hatem *et al.*, 2013). This is essential for quantification of gene expression, because reduction of the

number of mapped reads may produce bias in expression estimates. Other difficulty is represented by the existence of alternative splicing, isoform abundance based on aligned reads, and duplicated and paralogous genes, which have high sequence similarity. The analysis of gene expression by RNA-seq is based on counting number of reads specifically mapped on transcriptome reference. The frequency of mapped reads is normalized by many strategies based on different approaches (Bullard *et al.*, 2010; Risso *et al.*, 2011; Dohm *et al.*, 2008; Li *et al.*, 2010). Another bias may be caused by the differential sequencing depth between samples. The sequencing depth is defined as the total number of sequenced or mapped reads (Finotello and Camillo, 2014). Bullard et al. 2010 applied a quantile normalization, which is frequently used for microarray preprocessing. Similarly, the 'Trimmed Mean of M-values' (TMM) normalization to account for differences in library composition between samples and novel normalization method that assumes a Poisson model has been applied (Robinson and Oshlack *et al.*, 2010; Li *et al.* 2012). However, these methods are not completely effective and therefore RNA spike-in is recommended to be used for calibration. The RNA spike-in has known value of expression and is sequenced with the transcripts of interest.

Besides bias with the read mapping, variation in the gene length can also influence the gene expression analysis and result in the incorrect conclusions. It has been discovered that longer genes produce higher number of reads than shorter ones. Two major ways to reduce this bias has been proposed. Trapnell *et al. (*2010) used for normalization of mapped reads method known as Fragments Per Kilobase of transcript per Million mapped reads (FPKM). Similarly, Mortazavi *et al.* (2008) applied method called Reads Per Kilobase of exon model per Million mapped reads (RPKM), which is based on dividing the number of reads aligned to exon by the total number of mapped reads. The analysis of gene expression is then carried out as statistic test, used for identified genes with statistically significant variation in gene expression in the compared conditions. Most approaches are based on the Poisson (Jiang and Wong *et* al., 2009) and Negative Binomial (NB) (Di *et al.*, 2011) distributions.

2.3 FORAGE CROPS AND THEIR IMPORTANCE

In this chapter, the group of plants used during Ph.D. study will be introduced. Historically, forage crops have always played a key role in human societies and civilizations. They were essential for domestication of wild grazing animals. Recently, they are still highly important as livestock feed that provides meat, milk and other commodities. Forage grasses and legumes are the most widely cultivated crops in Europe covering approximately 20 to 40 percent of utilised agricultural area (depending on the definition) (Source FAOSTAT, 2014). Legumes are used for food, feed grain and oil. Grasses are the main components of pastures and meadows, but are also used for leisure and ornamental purposes.

2.3.1 LEGUMES

Legumes belong into family Fabaceae and their typical representatives are alfalfa, white and red clovers, soybeans, peas, beans, lentils, peanuts, lupins, mesquite, carob and tamarind. The most widespread forage legume in temperate climates is alfalfa (*Medicago sativa* L.) followed by clovers (*Trifolium* spp.) (Russelle, 2001). These plants are cultivated primarily for their seeds. However, alfalfa and clover are also grown in pastures. They are important for their ability to fix atmospheric nitrogen (N) and for their nutritional value. Specifically, they have high fibre content and are also a good source of proteins. Legumes contain substantial amounts of B vitamins as well as the nutritionally important minerals, such as iron, calcium and potassium (Rebello *et al.* 2014).

2.3.2 GRASSES

Grasses are monocotyledonous plants classified into family Poaceae. Forage grasses belong to the subfamily Pooideae together with wheat and barley. Grassland covers more than 40 % of the terrestrial area (Suttie *et al.* 2005). They represent an essential source of nutrients for livestock and also have significant impact to ecosystems. Forage species are often allogamous with high degree of self-incompatibility (Yang *et al.,* 2008) and many of their cultivars are highly heterozygous (Posselt, 2010). The most predominant forage grass species of the temperate regions are ryegrasses (*Lolium* spp.) and fescues (*Festuca* spp.) together with bluegrasses

(Humphreys, 2005). Members of fescues and ryegrasses can intercross and produce interspecific hybrids known as Festulolium.

2.3.2.1 Ryegrasses

Ryegrass (*Lolium* L.) is a genus of eight diploid species native in Europe and northwest Asia. Ryegrasses are annual, biennial or perennial herbs that are widespread in most temperate regions, including Australia and New Zealand, where they have been introduced with the first settlers. The most agriculturally important species are perennial (*Lolium perenne* L., 2n = 2x = 14) and Italian ryegrasses (*Lolium multiflorum* Lam., 2n = 2x = 14).

Perennial ryegrass is a tussock-forming perennial species with a fibrous root system. The leaves grow to a length up to 30 cm, width up to 7 mm and they are dark green, hairless, flat with upper surface evenly ribbed and lower surface smooth and shiny. Young leaves usually folded in the bud (V-shaped cross-section), but occasionally rolled (spiral cross-section), particularly in young plants. The auricles are small, narrow and ligule is white, translucent and shorter than wide (Lamp *et al.*, 2001).

Italian ryegrass is annual to biennial, but some cultivars persist for more than two years. Leaf blades are green to dark green, hairless, flat with upper surface evenly ribbed and lower surface smooth and shiny. The length of leaves is up to 40 cm with width 5-12 mm. Young leaves are rolled in the bud, and auricles are small and narrow. Ligule is white, translucent, shorter than wide and leaf is sheath hairless with fine longitudinal ribs as in leaf blades, rounder at back (Lamp *et al.*, 2001).

Figure 17 *- Lolium perenne L. and Lolium multiflorum Lam*

[https://upload.wikimedia.org/wikipedia/commons/e/e8/Lolium_multiflorum_%E2%80%94_Flora_Batava_%E2%80%94_Volume_v15.jpg][https://upload.wikimedia.org/wikipedia/commons/c/ce/478_Lolium_perenne.jpg]

Both ryegrass species are popular for their high yield, digestibility and good palatability, rapid establishment from seed, relatively fine texture, good density and uniformity (Kopecký *et al.,* 2006). However, *Lolium* species are sensitive to abiotic and biotic stress in general. Genome size of Italian ryegrass is 1C=2.56 Gbp with karyotype 2n=2x=14. However, tetraploid forms have been developed by polyploidization in breeding programs (Kopecký *et al.*, 2010).

2.3.2.2 Meadow Fescue

Genus *Festuca* is composed from nearly 500 species with ploidy levels from diploid (2n = 2x) to dodecaploid (2n = 12x). The vast majority of fescues are allopolyploid (Loureiro *et al.*, 2007; Smarda *et al.*, 2008). Two main evolutionary lineages were defined based on leaf anatomy and supported by the phylogeny analyses using sequence of internal transcribed spacer (ITS): broad-leaved and fine-leaved

fescues. Meadow fescue (*Festuca pratensis* Huds., 2n = 2x = 14) is a typical representative of broad-leaved fescues of mesophilic and hygrophilic meadows and pastures. It is among the most valuable forage species with a range of cultivated varieties. It provides quality forage with yield potential comparable with perennial ryegrass. *F. pratensis* is tolerant against severe climatic conditions such as winter freezing and summer drought (Jauhar 1993). However, it is sensitive and less adapted in lowland coastal regions, in the temperate zone and in the locations with high doses of nitrogen, where this species is less competitive in comparison with weeds and other grasses (Straková *et al.*, 2007). This is the reason why meadow fescue occurs rarely in intensively managed pastures (Kölliker *et al.*, 1999) and thus, is preferentially used as one of the major component of permanent pastures and meadows in alpine, eastern and northern regions of Europe (Ergon *et al.*, 2006). The genome size of tall fescue is 1C=3.175 Gbp (Kopecký *et al.,* 2010), which is about 6-7 times more than the genomes of *Brachypodium distachyon* and rice (Leitch and Bennett, 2005).

Figure 18 - *Festuca pratensis Huds.*

[https://en.wikipedia.org/wiki/Festuca_pratensis#/media/File:Festuca_pratensis.jpg]

2.3.2.3 Festulolium

Festuca and *Lolium* ssp. are outbreeders with a gametophytic self-incompatibility (Lundqvist 1962; Cornish *et al.* 1979). Due to their close relationship, they can hybridize and several interspecific hybrids called Festulolium have been reported along riverbanks of southern England and Northwest France (Borrill 1975, Lewis 1975). However, these natural hybrids are sterile. The existence of natural hybrids inspired breeders to cross members of these two genera in the controlled conditions. Since 1960´s, over fifty cultivars have been released at several breeding stations. Festulolium hybrids combine beneficial agronomic traits (rapid growth in the spring, root length, forage quality and resistance to biotic and abiotic stresses) from both genera and therefore, it could become an ideal component of agricultural or turf grass systems (Ghesquiere *et al.*, 2010). There are two types of Festulolium hybrids. Amphiploids are developed by the intercross of F1 hybrids and keep both parental genomes in about 1:1 ratio. On the other hand, introgression forms originated from the backcross of F1 hybrids with one of the parental species. This type is typically highly similar to the parent used for backcrossing with one or several improved traits introgressed from the other parent (Buckner *et al.,* 1977, 1983). Festuloliums are widely used and the number of cultivars is increasing, but the genome structure and the expression of parental-specific genes is still not well known compare to other crop species.

3. AIM OF THE THESIS

I. Identification of DArT markers associated with freezing tolerance

II. Identification of Repeats as a source of new cytogenetic markers

III. Development of robust bioinformatics methods for identifying and sorting of SNPs

IV. Gene expression analysis of interspecific hybrids *Lolium multiflorum* × *Festuca pratensis*

4. RESULTS

4.1 ORIGINAL PAPERS

4.1.1 GENETIC MAPPING OF DArT MARKERS IN THE FESTUCA-LOLIUM COMPLEX AND THEIR USE IN FREEZING TOLERANCE ASSOCIATION ANALYSIS

Bartoš J, Sandve SR, Kölliker R, Kopecký D, Christelová P, Stočes Š, Østrem L, Larsen A, Kilian A, Rognli OA, Doležel J

Theor. Appl. Genet. 122: 1133-1147, 2011.

Theor Appl Genet (2011) 122:1133–1147
DOI 10.1007/s00122-010-1518-z

ORIGINAL PAPER

Genetic mapping of DArT markers in the *Festuca–Lolium* complex and their use in freezing tolerance association analysis

Jan Bartoš · Simen Rød Sandve · Roland Kölliker · David Kopecký · Pavla Christelová · Štěpán Stočes · Liv Østrem · Arild Larsen · Andrzej Kilian · Odd-Arne Rognli · Jaroslav Doležel

Received: 20 September 2010 / Accepted: 18 December 2010 / Published online: 7 January 2011

Abstract Species belonging to the *Festuca–Lolium* complex are important forage and turf species and as such, have been studied intensively. However, their out-crossing nature and limited availability of molecular markers make genetic studies difficult. Here, we report on saturation of *F. pratensis* and *L. multiflorum* genetic maps using Diversity Array Technology (DArT) markers and the DArTFest array.The 530 and 149 DArT markers were placed on genetic maps of *L. multiflorum* and *F. pratensis*, respectively, with overlap of 20 markers, which mapped in both species. The markers were sequenced and comparative sequence analysis was performed between *L. multiflorum*, rice and *Brachypodium*. The utility of the DArTFest array was then tested on a Festulolium population FuRs0357 in an integrated analysis using the DArT marker map positions to study associations between markers and freezing tolerance. Ninety six markers were significantly associated with freezing tolerance and five of these markers were genetically mapped to chromosomes 2, 4 and 7. Three genomic loci associated with freezing tolerance in the FuRs0357 population co-localized with chromosome segments and QTLs previously indentified to be associated with freezing tolerance. The present work clearly confirms the potential of the DArTFest array in genetic studies of the *Festuca–Lolium* complex. The annotated DArTFest array resources could accelerate further studies and improvement of desired traits in *Festuca–Lolium* species.

Communicated by M. Xu.

Electronic supplementary material The online version of this article (doi:10.1007/s00122-010-1518-z) contains supplementary material, which is available to authorized users.

J. Bartoš (✉) · D. Kopecký · P. Christelová · Š. Stočes · J. Doležel
Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Sokolovská 6, 77200 Olomouc, Czech Republic
e-mail: bartos@ueb.cas.cz

S. R. Sandve · O.-A. Rognli
Department of Plant and Environmental Sciences, Norwegian University of Life Sciences, Ch. Magnus Falsens vei 1, 1432 Ås, Norway

R. Kölliker
Agroscope Reckenholz-Tänikon Research Station ART, Reckenholzstrasse 191, 8046 Zurich, Switzerland

L. Østrem
Norwegian Institute for Agricultural and Environmental Research, Fureneset, 6997 Hellevik i Fjaler, Norway

A. Larsen
Graminor AS, c/o Bioforsk Nord Bodø, Torggården, 8049 Bodø, Norway

A. Kilian
Diversity Arrays Technology, 1 Wilf Crane Crescent, Yarralumla, ACT 2600, Australia

Introduction

The *Festuca–Lolium* species complex includes some of the world's most important forage grasses. Even though these species are closely related to major cereal crops and share many characteristics with wheat (*Triticum aestivum* L.), barley (*Hordeum vulgare* L.) and other well studied crops, the species of the *Festuca–Lolium* complex possess a number of biological and genomic features unique to this group; for example, the perenniality of many forage grasses as compared to the annuality of temperate cereals, and chromosome rearrangements specific for the *Festuca–Lolium* complex (Alm et al. 2003). The species within the

complex possess a range of complementary characteristics, which are often explored in grass hybrid breeding. Italian ryegrass (*Lolium multiflorum* Lam.) and perennial ryegrass (*L. perenne* L.) are two of the most important forage grasses of temperate regions. Italian ryegrass is especially valued for its high dry matter yield and its excellent forage quality. Although primarily used for hay and silage production in temporary leys, it is also a substantial component of permanent grassland (Peter-Schmid et al. 2008). On the contrary, perennial ryegrass is cultivated mainly for grazing. It forms the main component of productive permanent grassland mainly due to its incomparable perenniality. However, *Lolium* species often suffer from limited persistence under abiotic stress.

On the other hand, *Festuca* species generally have a better tolerance to abiotic stress when compared to *Lolium* species. Meadow fescue (*Festuca pratensis* Huds.) is a forage grass which constitutes a significant component of species-rich permanent pastures and hay fields in alpine and eastern regions of Europe. In Scandinavia, it is also a major component of intensively managed swards cut for silage (Rognli et al. 2010). Because of their close evolutionary relationships, introgression of *Festuca* genes into *Lolium* genomes through the development of *Festuca* × *Lolium* hybrids may allow to improve forage grass stress resistance (Humphreys et al. 2005). In this respect, *F. pratensis* is often used to improve freezing tolerance in *Festuca* × *Lolium* hybrids (Kosmala et al. 2006). Genome-wide molecular markers are essential for characterization of such hybrids and identification of introgressed chromosome segments.

*D*iversity *A*rrays *T*echnology (DArT, Jaccoud et al. 2001) is a high-throughput sequence independent genotyping method based on the reduction of genome complexity and DNA microarray hybridization. Through the past decade, it has become a valuable source of markers for genomes with limited sequence information. DArT arrays have been developed for crop species like banana (*Musa acuminata* Colla), cassava (*Manihot esculenta* Crantz), cereals as well as for model plants such as *Arabidopsis*, rice (*Oryza sativa* L.) and sorghum (*Sorghum bicolor* L.) (for a complete list of species see http://www.diversityarrays.com/genotypingserv.html). DArT markers were often used for diversity studies, and for construction and saturation of genetic maps (Wenzl et al. 2006; Tinker et al. 2009). DArT markers have also been used in the analysis of important agricultural traits, for example in an association study of tan spot resistance in hexaploid wheat (Singh et al. 2010).

Recently, Kopecký et al. (2009) developed a DArTFest array for the *Festuca–Lolium* complex. This printed microarray contains 7,680 probes derived from methyl-filtered genomic representations of three *Festuca* (*F. arundinacea* Schreb., *F. glaucescens* Boiss. and *F. pratensis*) and two *Lolium* (*L. multiflorum* and *L. perenne*) species. The DArTFest array has a potential to provide important insights into the genome structure and evolution of the species within the *Festuca–Lolium* complex. Many of the DArT markers cross-hybridize with several different species, which enables comparative genomic studies. DArT markers can also be sequenced and physically mapped to genomes of fully sequenced model species to further understand the syntenic relationships between chromosomes of forage grasses and model plants. This may in turn lead to more efficient identification of genes underlying important agricultural traits. To date, DarTFest array has been used to study genetic diversity (Kopecký et al. 2009) and to determine genomic constitution of Festulolium cultivars (Kopecký et al. 2011).

Following the breeding priorities, molecular dissection of agronomic traits in *L. multiflorum* has so far mainly been focused on resistance to diseases such as crown rust or bacterial wilt (Studer et al. 2006, 2007). In this context, a genetic linkage map based on 306 F_1 individuals has been established (*Xtg*-ART) and more recently been used for construction of a consensus linkage map in *Lolium* using EST-derived SSR markers (Studer et al. 2010). Alm et al. (2003) established a *F. pratensis* mapping population (HF2/7 × B14/16) consisting of 138 F_1 individuals and mapped 446 RFLP, AFLP and SSR markers. This population has been used for the analysis of freezing and drought tolerance, winter hardiness, and vernalization sensitivity (Alm et al., unpublished; Ergon et al. 2006). In this study, we (1) saturated and improved these genetic maps of *F. pratensis* and *L. multiflorum* with DArT markers, (2) sequenced the mapped markers to investigate the genomic origin of DArT markers and the syntenic relationships to model genomes, and (3) used these new DArTFest array resources in an integrated analysis of agriculturally important trait freezing tolerance in a *L. perenne* × *F. pratensis* hybrid population.

Materials and methods

Plant material

The HF2/7 × B14/16 mapping family consisting of 138 F_1-offspring was used for the construction of the *F. pratensis* map (Alm et al. 2003). For the *L. multiflorum* map, 288 F_1 individuals of the *Xtg*-ART mapping population (Studer et al. 2006) were used. In the association study, the Festulolium population (FuRs0357) was used. The FuRs0357 population (*L. perenne* × *F. pratensis*) originates from a wide genetic pool from several initial hybrids made from either *Festulolium* cv. Prior (LpFp, 4*x*) crossed with *L. perenne* (2*x*) or crosses between *L. perenne* (4*x*)

and *F. pratensis* (2*x*). The initial hybrids were backcrossed twice onto diploid *L. perenne* to obtain BC_2 progenies and then put through two generations of seed propagation. Twenty one plants belonging to either a high freezing tolerance (HFT) group (11 plants) or a low freezing tolerance (LFT) group (10 plants) were used in the study.

DArT screening

Genomic DNA was isolated from individual plants from both mapping populations and from the Festulolium population FuRs0357 using Invisorb Spin Plant Mini Kit (Invitek, Berlin, Germany). The previously developed DArTFest array (Kopecký et al. 2009) was hybridized with fluorescently labelled genomic representations of individual plants, which were prepared from genomic DNA by the same *Pst*I/*Taq*I complexity reduction method as used for preparing the array (for details see Akbari et al. 2006). Hybridization signals were converted into 0–1 scores using the DArTsoft software package developed at Diversity Arrays Technology Pty Ltd (DArT P/L, Yarralumla, Australia).

Genetic mapping

Genetic maps of *F. pratensis* and *L. multiflorum* were constructed in JoinMap 4.0 using the Kosambi mapping algorithm. As DArT markers are dominant, marker files were coded for double haploid populations (DH) and maps were calculated for each parent separately. Only markers segregating close to the expected 1:1 ratio (>0.4 and <0.6) were used. Maps were then combined using bridging markers present in both parents. The inclusion of markers was decided by the following procedure. First, all markers were grouped into putative linkage groups based on Likelihood Ratio Odds (LOD) grouping. Next, initial maps were calculated and all markers with a Chi-square value of >3 (highly distorted segregation) were removed. Subsequent recalculation of maps was performed until no markers had a Chi-square score of >3. Finally, the two parental maps for each LG were combined into one. All final linkage maps were edited and finalized using the MapChart software (Voorrips 2002).

Association of DArT markers with freezing tolerance

Freezing tolerance was measured with the re-growth method (Larsen 1978). In brief, 300 genotypes from FuRs0357 were put through controlled freezing stress which resulted in differential freezing induced damage and survival. Plants were grouped randomly in five boxes using six replicates per genotype. The resulting HFT (high freezing tolerance) and LFT (low freezing tolerance) plants belonged to the 10% of phenotypes with the highest and lowest freezing tolerance, respectively. The HFT group had an average freezing tolerance score of 7.06 (±0.22) and the LFT group had an average freezing tolerance score of 2.59 (±0.61) [range; 0 (dead) to 9 (no damage)]. Associations between DArT marker genotypes and freezing tolerance groups were tested using Fisher's Exact test. The null hypothesis was that the DArT marker genotypes were not associated with freezing tolerance; hence we expected random distribution of genotypes in HFT and LFT groups. We calculated the corresponding q values from the Fisher Exact test p values (Storey 2002) to correct for multiple testing and false positives. A significance threshold of $q < 0.05$ was used. Statistical analyses were carried out in R (R Development Core Team 2009) using the "fisher.test" and "q value" functions in the q-value package.

Sequencing of selected clones/markers

All DArT markers, which were placed on the genetic maps of *F. pratensis* and *L. multiflorum* and/or significantly associated with freezing tolerance, were sequenced. Reaction mix for cycle sequencing was prepared using standard BigDye chemistry (BigDye® Terminator v3.1 Cycle Sequencing Kit, Applied Biosystems). The dilution of reaction components was scaled to a final volume of 10 µl containing 3.2 pmol of universal M13 (forward or reverse) primer and 20 ng of sequence-ready (DArT clones) template. The reaction products were purified using the CleanSEQ kit (Agencourt Bioscience Corp., Beckman Coulter Comp., Beverly, MA, USA) and analyzed on an ABI 3730xl DNA analyzer (Applied Biosystems, Foster City, CA, USA). The raw sequence data were assembled and edited using the DNA Baser software v.2 (Heracle-Software, http://www.DnaBaser.com). Vector and adaptor sequences were removed prior to further analysis. All sequences were deposited in GenBank under accession numbers HN266254-HN266937.

Sequence analysis

All mapped markers were compared to each other to determine the extent of redundancy in DArT clones. Stand alone blastn software was used for the search with *E* value set to e-10. Only reciprocal blast hits were taken into account. The map position of markers with significant hits was checked and markers which mapped to different map positions were not considered redundant. Groups of markers, which shared significant homology to each other and mapped to the same genetic position, are hereafter referred to as 'bins'. The sequences were further compared to known plant repeat sequences. To analyze repeat content, TREP Release 10 (http://wheat.pw.usda.gov/ITMI/Repeats/) was merged with

TIGR Plant Repeat Databases for Brassicaceae, Fabaceae, Poaceae and Solanaceae (Ouyang and Buell 2004). This database contains 8,432 repetitive elements from 18 different genera. Repeat analysis was performed using RepeatMasker software (http://repeatmasker.org) with CrossMatch search engine (http://www.phrap.org/phredphrapconsed.html) and default settings. We compared DArT sequences with non-redundant protein sequences (nr) and non-human, non-mouse ESTs database (est_others) at GenBank in order to estimate the number of DArT markers derived from expressed loci. The protein search (blastx) was performed using blastcl3 with default settings and BLOSUM62 scoring matrix. Nucleotide search (blastn) was also performed by blastcl3 with default settings but reward for a nucleotide match was set to 2. Only the best blast hits (with lowest *E* value) were taken into account.

Comparison of mapped DArTs to model genomes

The sequences of the mapped DArT markers were compared to the *Oryza sativa* ssp. *japonica* cv. 'Nipponbare' genome and the *Brachypodium distachyon* Bd21 genome using blastn with *E* value set to e-10. Twelve pseudomolecules for the rice chromosomes (Build5) were downloaded from the IRGSP website (http://rgp.dna.affrc.go.jp/E/IRGSP/Build5/build5.html). The Bd21 genome was downloaded from Brachybase.org (http://files.brachypodium.org/). Only the best blast hits with alignment lengths of at least 50 bp were taken into account.

GISH analysis of Festulolium plants

Twelve individuals from the Festulolium population FuRs0357 (six with high freezing tolerance, HFT, and six with low freezing tolerance, LFT) were used for the analysis of genomic constitution using genomic in situ hybridization (GISH). The other plants suffered under conditions optimal for preparation of chromosome spreads (root growth and accumulation of cells in metaphase). GISH was performed as described by Kopecký et al. (2005). Briefly, total genomic DNA of *F. pratensis* was labeled with digoxigenin using DIG-Nick Translation Kit (Roche Applied Science, Indianapolis, IN, USA) and used as a probe. Genomic DNA of *L. perenne* was sheared to ~200 bp fragments and used as blocking DNA. Sites of probe hybridization were detected by anti-DIG-FITC conjugate (Roche). Chromosomes were counterstained with 1.5 μg/ml 4′,6-diamidino-2-phenylindole (DAPI) prepared in Vectashield antifade solution (Vector Laboratories, Burlingame, USA). Observations were made using an Olympus AX70 microscope equipped with epi-fluorescence and SensiCam B/W camera. Scion-Image and Adobe Photoshop software were used for processing of color pictures.

Results

Genetic mapping

DArTFest array contains 2,257 probes positively scored in *F. pratensis* and 2,761 probes positively scored in *L. multiflorum* (Kopecký et al. 2009). In total we placed 659 markers on genetic maps, 149 markers in the *F. pratensis* population HF2xB14, and 530 markers in the *L. multiflorum* population *Xtg*-ART (Table 1; Fig. 1). The effectiveness of DArT mapping was 6.6 and 19.2% for *F. pratensis* and *L. multiflorum*, respectively. Only 20 markers could be mapped in both species. Out of them, 17 mapped to homologous linkage groups and three markers D355658, D557428, D558076 mapped to Fp4 and Lm1, Fp5 and Lm7, Fp1 and Lm7, respectively.

Sequencing and sequence analysis of mapped markers

Out of the 659 mapped markers, 620 yielded 302,808 bp of sequence with an average length of 488.4 bp. Sequencing of the remaining 39 DArT markers failed due to technical reasons. A total of 398 DArT markers (64.2%) were found to be singletons, while the 222 remaining markers were redundant and were assigned to 90 marker bins. The biggest bin consisted of six markers. Hence, 489 non-redundant DArT markers/bins were mapped when considering each bin a unique locus (Table 2). Nevertheless, markers belonging to one bin can slightly differ in their precise map position (see Fig. 1). This could be due to inaccuracy during the array hybridization leading to genotyping errors and variation in estimated numbers of recombinants, inaccuracies in the consensus map estimation, or duplicated genes in tandem arrays.

Table 1 Distribution of genetic markers among linkage groups

	Festuca pratensis		*Lolium multiflorum*	
	Non-DArT markers[a]	DArT	Non-DArT markers[b]	DArT
LG1	61	23	63	67
LG2	26	10	43	73
LG3	34	22	49	73
LG4	68	22	61	106
LG5	70	12	43	71
LG6	49	27	35	63
LG7	65	32	68	76
Total	373	148	352	529

[a] AFLP, RFLP, SSR markers and isozymes, see Alm et al. (2003) for details

[b] SSR and AFLP markers, see Studer et al. (2006) for details

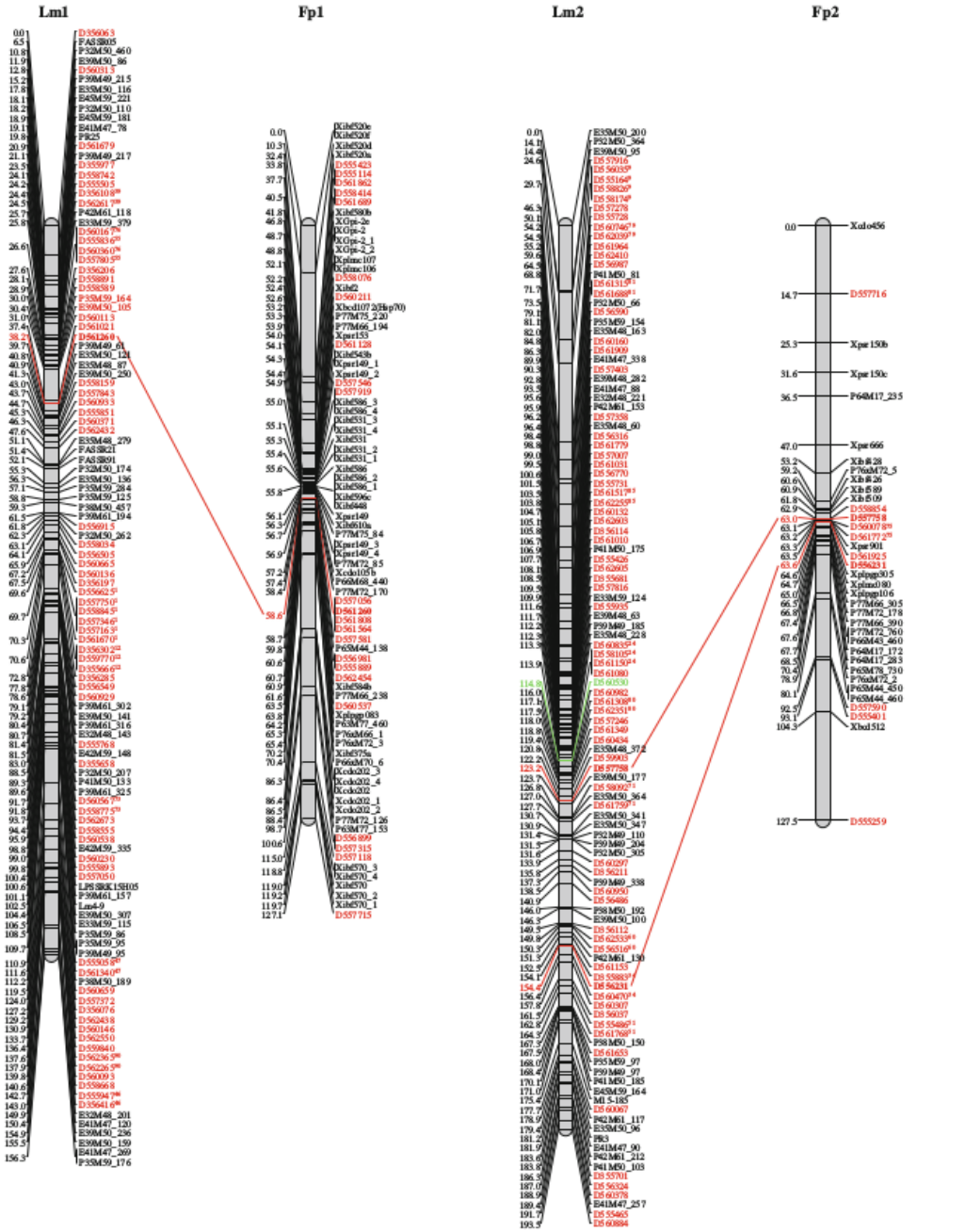


Fig. 1 Genetic maps of seven *F. pratensis* and *L. multiflorum* chromosomes. DArT markers are labeled with D and clone number (six digits) and are indicated by *red color*. Markers belonging to one redundant marker bin are indicated by the same *superscript*. Markers associated with freezing tolerance are indicated by *green color*

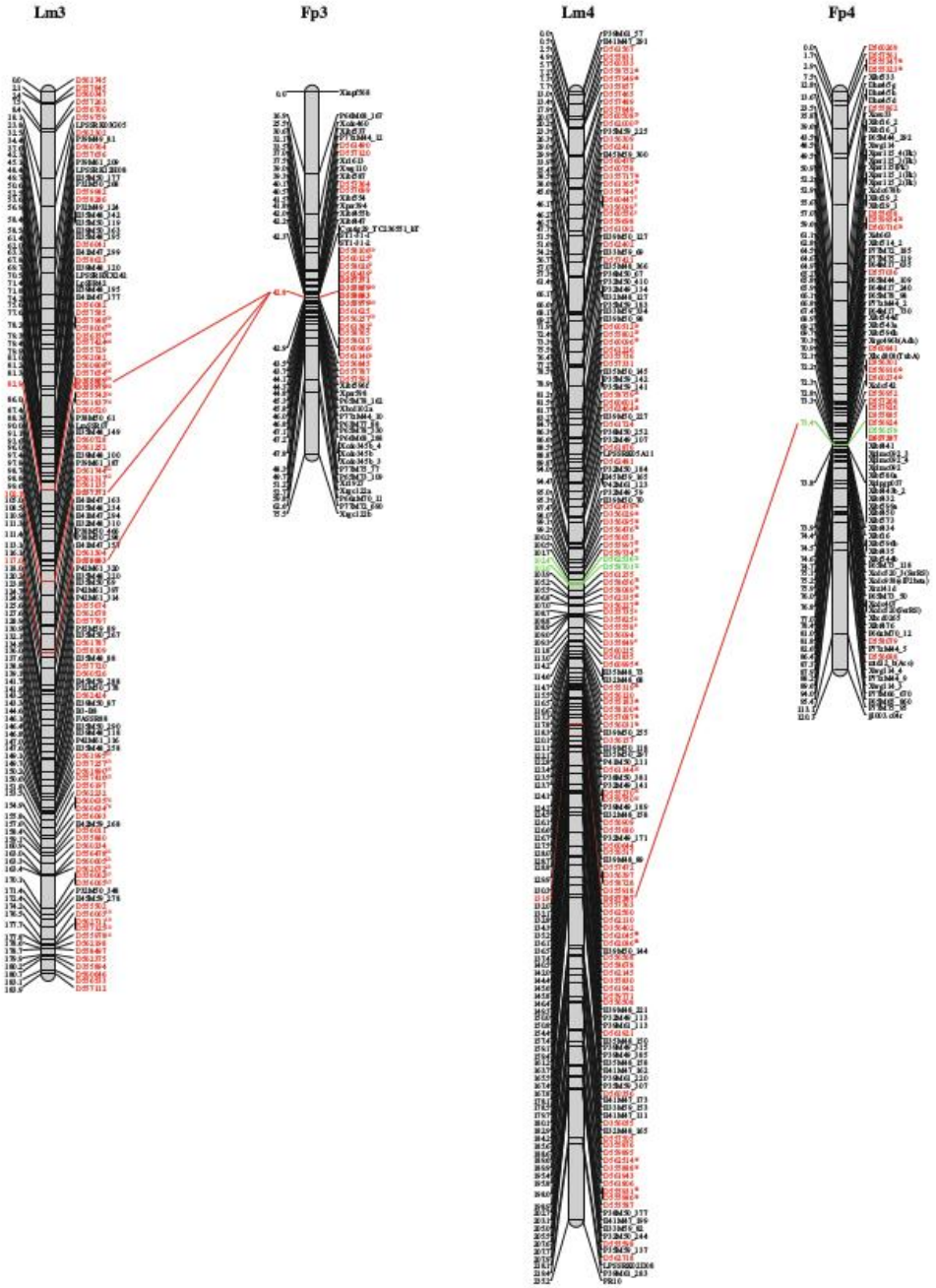


Fig. 1 continued

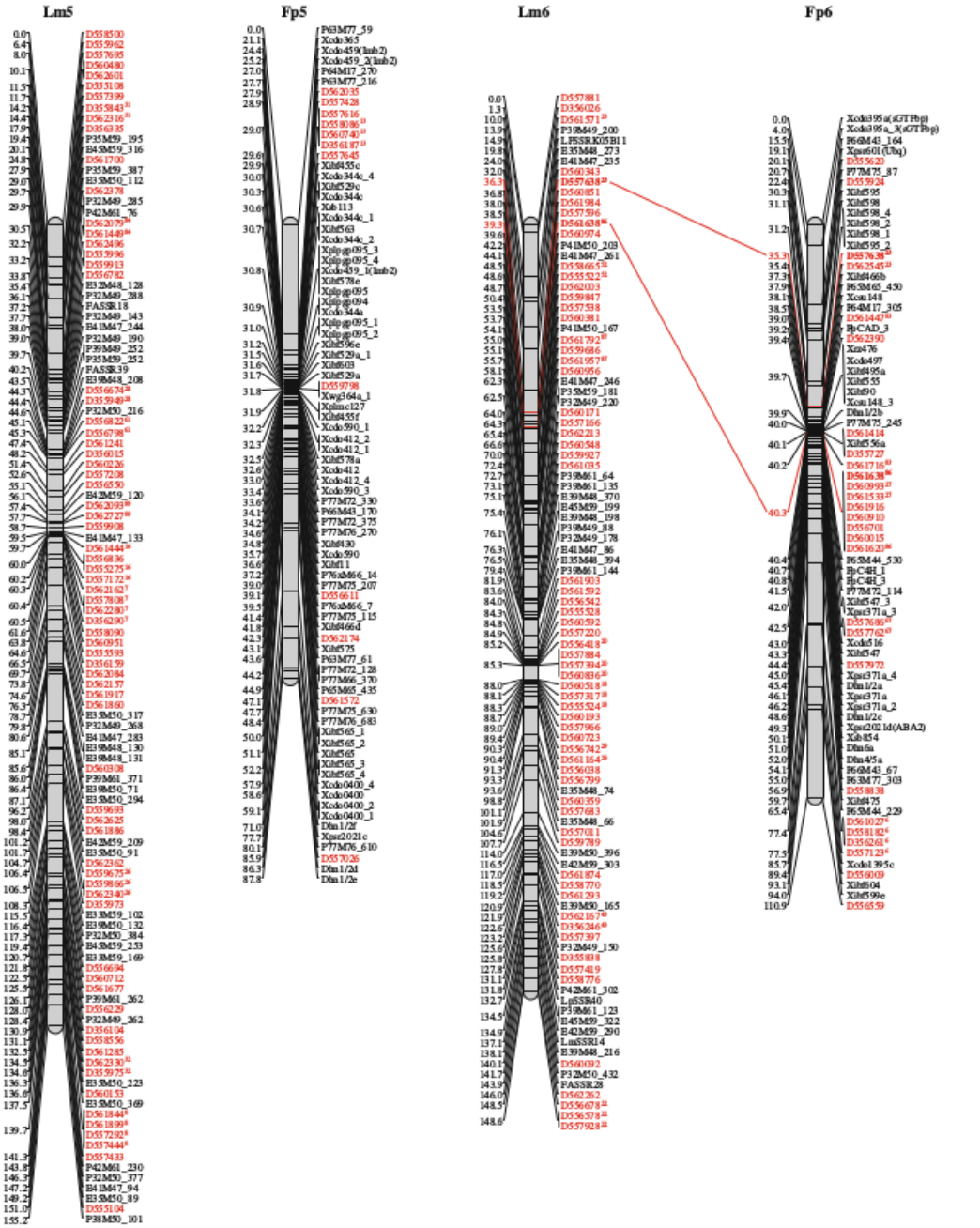


Fig. 1 continued

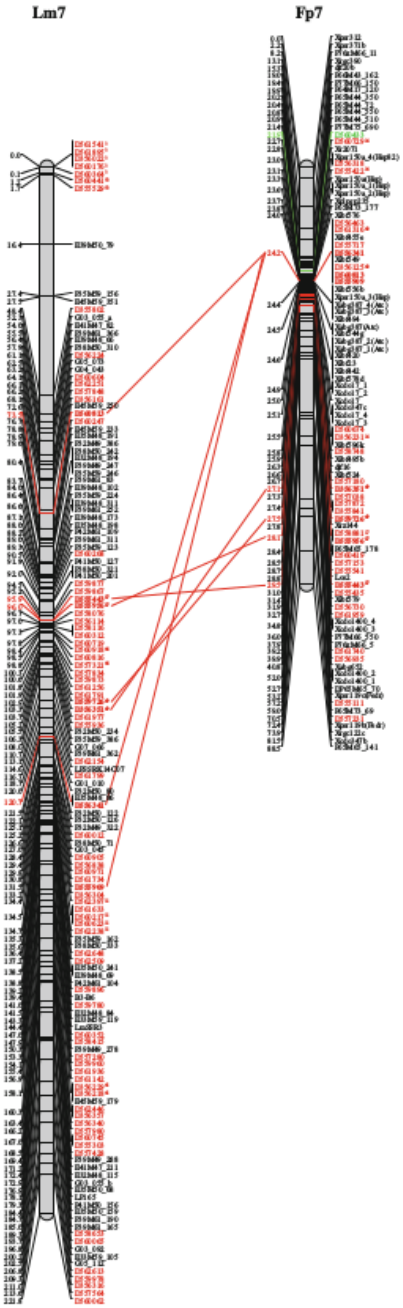


Fig. 1 continued

Table 2 Redundancy of sequenced DArT markers

Extent of redundancy	Number of bins	Number of markers	Frequency of markers (%)
Unique	399	399	64.3
Duplicates	64	128	20.6
Triplicates	15	45	7.2
4/bin	8	32	5.2
5/bin	1	5	0.8
6/bin	2	12	1.9
Total	489	621	100

To estimate the number of DArT markers derived from repetitive DNA elements, we performed blast search against an in-house built composite plant repeat database. Only 44 (7.1%) of the DArT markers contained repetitive elements supporting the notion that DArT markers represent hypo-methylated low copy genomic regions (Wenzl et al. 2006). In the repeat containing DArT sequences, the retrotransposons (27) dominated DNA-transposons (13). The remaining four hits in the plant repeat database were unclassified. To estimate the impact of repeat elements on marker redundancy, we compared DArT markers derived from repetitive elements with the marker bins. Interestingly, repetitive elements could explain the presence of redundant markers only for seven (7.8%) of the marker bins. This indicates that the majority of redundant markers in bins are either due to multiple cases of cloning the same sequence during the random marker development procedure, or they come from tandem duplicated genes.

In order to estimate the proportion of expressed sequences among our mapped DArT markers, we performed a blastn search against GenBank "est_others" database. This identified 368 (59.4%) DArT markers with significant homology to expressed sequences. Blastx search against non-redundant protein sequences (nr) revealed 162 (26.1%) DArT markers with significant homology to known and hypothetical proteins, and of these 152 were also identified as expressed sequences in blastn search. In total, 378 DArT markers (293 non-redundant bins) were identified as potentially gene-derived sequences.

Comparison of mapped DArTs to model genome(s)

The genetically mapped DArT markers were compared to the rice and *Brachypodium* genomes in order to map the sequences in silico and investigate the synteny between the model genomes and those of the *Festuca–Lolium* complex. Using the defined criteria for blast search, mapped DArT

markers in *L. multiflorum* produced 227 (162 non-redundant bins) and 299 (227) hits against rice and *B. distachyon* genomes, respectively. Of the markers mapped in *F. pratensis*, 50 (40) and 70 (54) produced blast hits in rice and *Brachypodium* genomes, respectively. Due to the low number of hits of the *F. pratensis* markers, the analysis of syntenic relationships was limited to comparisons between *L. multiflorum* and rice (Table 3) and *L. multiflorum* and *Brachypodium* (Table 4). All significant homologies are listed in Supplementary Tables 1 and 2.

We identified 11 syntenic regions with more than 5 markers shared between particular *L. multiflorum* and rice chromosomes. The highest degree of synteny was found between the chromosomes/linkage groups Lm3 and Os01, where 71.4% of markers presented in Lm3 had a homologous sequence on Os01. In comparison, 12 syntenic regions with more than 5 shared markers were found between the genomes of *L. multiflorum* and *Brachypodium*. The most conserved syntenic relationships were found between Lm5-Bd4 and Lm3-Bd2, with 76.9 and 76.5% of the *L. multiflorum* markers having homologues in corresponding *Brachypodium* chromosome, respectively.

Table 3 Synteny of *L. multiflorum* and rice as revealed by mapped DArT markers

	Lm1	Lm2	Lm3	Lm4	Lm5	Lm6	Lm7
Os01	1	1	**15**	1	1	1	–
Os02	2	–	1	3	2	**16**	–
Os03	–	1	–	**18**	–	–	2
Os04	2	**10**	2	2	2	1	–
Os05	**7**	–	1	1	–	–	–
Os06	–	–	–	2	–	–	**15**
Os07	–	**7**	–	1	–	2	–
Os08	–	2	–	–	–	–	**7**
Os09	–	–	–	–	**7**	1	–
Os10	**7**	–	–	–	1	1	–
Os11	1	–	1	**6**	1	2	1
Os12	–	–	1	–	3	–	1
Total	20	21	21	34	17	24	26

Numbers correspond to non-redundant bins mapped to each rice chromosome. Chromosome relationships with highest synteny are bold

Table 4 Synteny of *L. multiflorum* and *Brachypodium* as revealed by mapped DArT markers

	Lm1	Lm2	Lm3	Lm4	Lm5	Lm6	Lm7
Bd01	1	**11**	3	**30**	1	3	**22**
Bd02	**12**	1	**26**	2	4	2	1
Bd03	**10**	2	1	4	1	**20**	**8**
Bd04	1	1	3	**9**	**20**	**5**	1
Bd05	1	**15**	1	1	1	1	2
Total	25	30	34	46	27	31	34

Numbers correspond to non-redundant bins mapped to each *Brachypodium* chromosome. Chromosome relationships with highest synteny are bold

Analysis of markers associated with freezing tolerance

Out of 1,868 DArT markers segregating in the Festulolium FuRs0357 population, about 5% (96) had a significantly different distribution of genotype scores among HFT and LFT plants ($q < 0.05$). Unfortunately, only five markers (four non-redundant bins) could be mapped. These five markers were mapped to linkage groups Fp4 (D556159) and Fp7 (D560433) in *F. pratensis* and Lm2 (D560530) and Lm4 (D558701, D562536) in *L. multiflorum* (Table 5; Fig. 1). The FuRs0357 population was in fact derived from a cross between *L. perenne* and *F. pratensis*. Nevertheless, one can expect that map positions of DArT markers in *L. perenne* to be highly co-linear with their position in *L. multiflorum* due to the close evolutionary relationship (Catalán et al. 2004, Studer et al. 2010). Seventy two of the freezing tolerance-associated markers were sequenced for the purpose of studying their putative genomic origin and localization in model genomes in silico. Several markers associated with freezing tolerance were found to map Bd1, and nine of these markers were flanking *Vrn1* gene paralogs on Bd1 (Fig. 2). This further supports the presence of a freezing tolerance QTL on chromosome 4 in *L. perenne* and *F. pratensis* as Bd1 has syntenic relationship to chromosome 4 of species in the *Festuca–Lolium* complex. We performed functional analysis of all DArT markers associated with freezing tolerance using blast2go (Conesa et al. 2005). For 18 of these markers a gene functional class could be assigned based on the homology to known proteins and transcription is the most frequent one (4 markers). All markers associated with freezing tolerance are listed in Supplementary Table 3.

Table 5 Mapped markers having significantly (Fisher Exact test) different distributions among HFT (high freezing tolerance) and LFT (low freezing tolerance) plants

DArT ID	*q* value	Presence	LG	cM
D560433	0.028	LFT	LG7 *F. pratensis*	21
D556159	0.042	HFT	LG4 *F. pratensis*	73
D558701	0.042	LFT	LG4 *L. multiflorum*	103
D562536	0.042	LFT	LG4 *L. multiflorum*	102
D560530	0.044	HFT	LG2 *L. multiflorum*	115

"Presence" denotes if a marker presence signal is associated with the LFT or HFT group

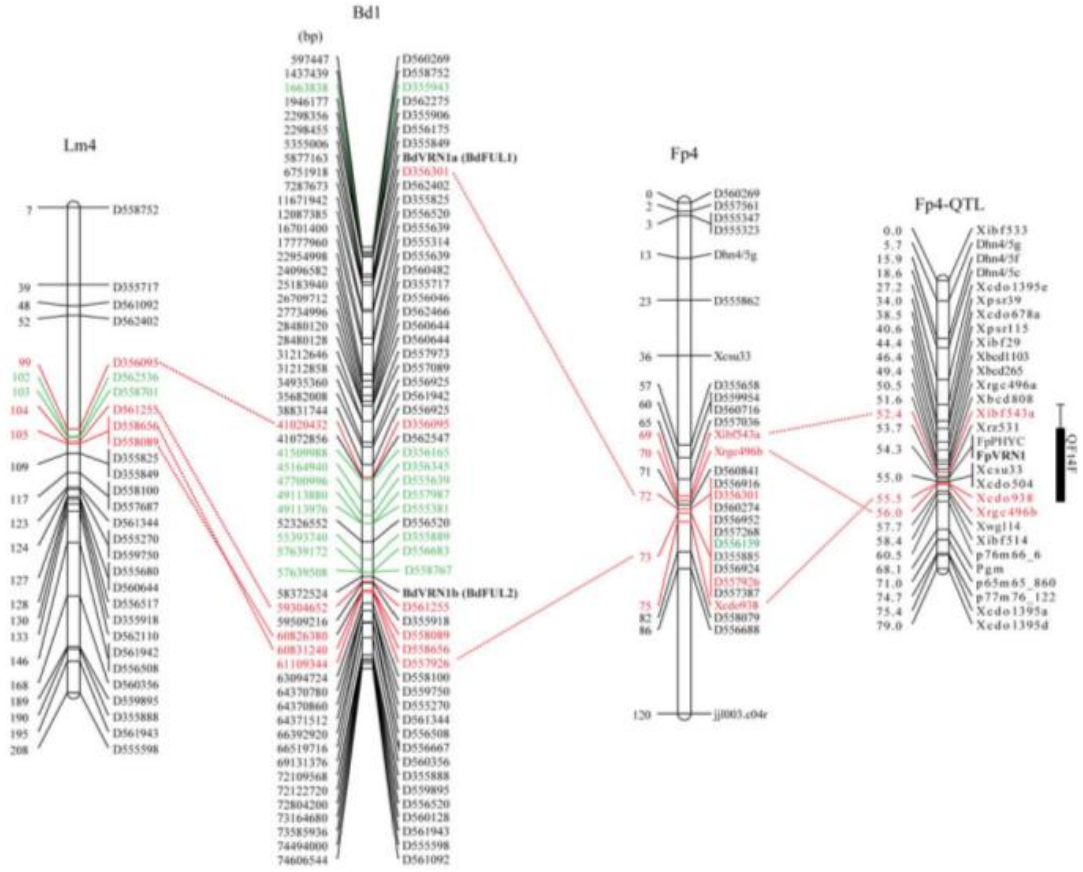


Fig. 2 Syntenic relationships between *F. pratensis*, *L. multiflorum*, and *Brachypodium* chromosome 4. Comparative map shows the syntenic relationships between the freezing tolerance QTL in *F. pratensis* and the DArT markers associated with freezing tolerance phenotype closely linked to VRN1 paralogs in the *Brachypodium* genome. The position of the DArT markers on Bd1 are extrapolated from the best blastn hit in the Bd genome. *Red color text* denotes bridging markers between maps. *Green color text* denotes DArT markers significantly associated with freezing tolerance phenotype flanking FUL1 and FUL2. For the Bd1 chromosome, the *left side numbers* are physical position in base pair. *Left side numbers* for all other maps are genetic distances in cM. The rightmost map is reproduced from Alm et al. (unpublished results) where the *black bar* represents the 95% confidence interval for a freezing tolerance QTL

GISH analysis of Festulolium plants

Microscopic analysis of mitotic metaphase plates of six LFT and six HFT plants revealed that the FuRs0357 Festulolium population contained plants with different ploidy levels. Four of the LFT plants were tetraploid, with the remaining two diploid, while all six HFT plants were diploid. Such a large difference is not likely to occur by chance (Fisher Exact test; $P = 0.045$), which indicates that the polyploidy levels could be functionally linked to freezing tolerance. The higher ploidy level could affect the proportion of markers which cannot be called as "0" or "1" when converting fluorescence signal since it increases signal to noise ratio. Exactly this pattern was observed in the FuRs0357 population. The average numbers of not called DArT marker genotypes per plant were 14 in HFT and 127 in LFT (*T* test; $P = 0.0003$), reflecting the higher ploidy level in the LFT group.

GISH analysis revealed *F. pratensis* chromatin only in a small subset of plants, no matter if they were freezing tolerant or not. Two recombined chromosomes carrying terminal *F. pratensis* segments were found in one tetraploid LFT plant (Fig. 3a). *Festuca pratensis* chromatin was not detected in any other LFT plant. Two HFT plants contain

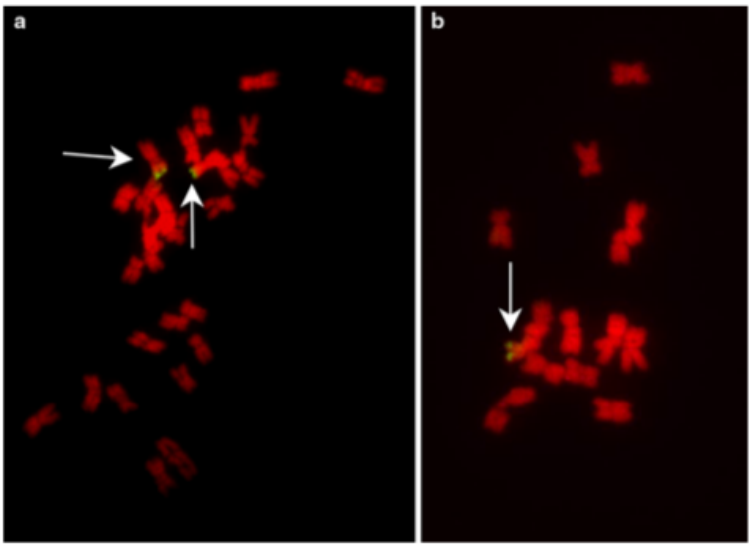


Fig. 3 Analysis of Festulolium plants with low (LFT) and high (HFT) freezing tolerance using genomic in situ hybridization (GISH). GISH in *L. perenne* × *F. pratensis* LFT149 (**a**) and HFT152 (**b**) plants. Total genomic DNA of *F. pratensis* was labeled with FITC and used as a probe (*green color*); genomic DNA of *L. perenne* was used to block hybridization of common sequences. Chromosomes were counterstained by DAPI (shown in red pseudocolor). Tetraploid LFT149 plant ($2n = 4x = 28$) has two recombined chromosomes with small terminal translocations of *F. pratensis* (*arrows*), while diploid HFT plant ($2n = 2x = 14$) has one recombined chromosome with small terminal *F. pratensis* translocation (*arrow*)

chromosomes carrying terminal *F. pratensis* segments (Fig. 3b). No *F. pratensis* chromatin was found in the remaining four HFT plants.

Discussion

Mapping of DArT markers

The DArT technology has become a valuable molecular tool for many plant species. More than thousand DArT markers have been genetically mapped in rye (*Secale cereale* L.) (Bolibok-Bragoszewska et al. 2009) and oat (*Avena sativa* L.) (Tinker et al. 2009), and more than 2,000 have been mapped in barley (Wenzl et al. 2006). Here we have used the recently developed DArTFest array (Kopecký et al. 2009) to map genetically hundreds of DArT markers in the *Festuca–Lolium* complex. Different numbers of markers could be mapped in the two species and compared to *F. pratensis*, four times more markers were mapped in *L. multiflorum*. Among the 2,257 probes detected in *F. pratensis*, 1,078 (47.8%) were previously found polymorphic within this species. Similarly, 2,184 out of 2,761 probes (79.1%) were found polymorphic among the same number of *L. multiflorum* accessions (Kopecký et al. 2009). Based on this, we expect significantly lower number of markers to be mapped in *F. pratensis*. The discrepancy in the number of markers mapped in *L. multiflorum* compared to *F. pratensis* could be also due to low genetic variation between the parents of the *F. pratensis* mapping population. A higher level of polymorphism in *L. multiflorum* ecotypes compared to those of *F. pratensis* was also observed when using co-dominant SSR markers (Peter-Schmid et al. 2008). Based on studies of cpDNA variation, Fjellheim et al. (2006) concluded that meadow fescue in Europe went through a bottleneck during or after the last glaciations. Thus, the number of genetically mapped markers may reflect the overall diversity within a particular species.

Unexpectedly, only 20 markers could be mapped in both species. To understand this, we used data collected by Kopecký et al. (2009) and examined markers mapped in *L. multiflorum* for their presence and polymorphism among *F. pratensis* accessions and vice versa. The detailed analysis revealed that the low number of shared markers was not due to the absence of markers in the other species but due to the lack of polymorphism in the other mapping population. These 20 shared markers represent 7.0% of markers mapped in *L. multiflorum* and present in *F. pratensis* and at the same time 18.3% of markers mapped in *F. pratensis* and present in *L. multiflorum*. Consequently, these findings are in agreement with overall performance of markers available at DarTFest array (for details see Table 6).

Table 6 Extent of DArT markers' polymorphism and effectiveness of genetic mapping

	Fp present[a]	Fp polymorphic	Fp mapped	Lm present[a]	Lm polymorphic	Lm mapped
DArTFest[b]	2,257	1,078 (47.8%)	149 (6.6%)	2,761	2,184 (79.1%)	530 (19.2%)
Fp mapped	–	–	–	109	77 (70.6%)	20 (18.3%)
Lm mapped	287	146 (50.9%)	20 (7.0%)	–	–	–

[a] Number of markers detected in at least one accession of particular species. Based on data collected by Kopecký et al. (2009)

[b] DArTFest array contains 7,680 probes derived from five species of *Festuca–Lolium* complex (Kopecký et al. 2009)

Sequence analysis of DArT markers

Among the sequenced DArT markers, 64.2% were found to be unique. In similar studies, Wittenberg et al. (2005) and Tinker et al. (2009) found 56.3 and 48.1% unique markers among DArTs developed for *Arabidopsis* and oat, respectively. The levels of redundancy are strikingly similar despite the great difference in genome size, ranging from 157 Mbp for *A. thaliana* to 12,961 Mbp for oat (Bennett and Leitch 2005), and contrasting repeat content in the three species. Sequence analysis revealed that only about 7% of the mapped DArT markers were of repetitive origin and the effect of repetitive DNA on marker redundancy seems to be negligible. Sources of DArT marker redundancy are most likely related to the PCR amplification involved in array development and possibly also the presence of duplicated genes in tandem arrays which are frequent in plant genomes (Rizzon et al. 2006; Hanada et al. 2008).

DArT markers are derived from genomic representations prepared with methyl-sensitive restriction enzymes and hence should represent low-copy genomic regions. As expected, a majority of the sequences were found to be derived from expressed parts of the genomes. About 60% of all markers had homology to the NCBI EST databases and one fourth had homology to the NCBI protein database. These groups are overlapping extensively, resulting in slightly more than 60% expressed non-redundant markers/bins. This is similar to the finding of Tinker et al. (2009) in oat. Interestingly, also most of the DArT markers with repetitive origin had blast hits to ESTs and proteins. This further supports the tendency of DArT markers to originate from expressed genomic loci.

Synteny with rice and Brachypodium

Genome-wide comparisons of *L. multiflorum* to rice and *Brachypodium* based on DArT markers revealed syntenic regions shared among the species. These regions corresponded to the established syntenic relationships between *L. multiflorum* and the model species for all chromosomes but Lm5 (see Tables 3, 4). Syntenic blocks between Lm5 and Os03 and Os12 could be expected (Bolot et al. 2009). We identified markers shared between Lm5 and Os12, however, not enough markers to meet our definition of a syntenic relationship, neither did we observe the expected syntenic relationship between Lm5 and Bd1 (The International Brachypodium Initiative 2010). The number of DArT markers used to infer syntenic regions in this paper is small, and many grass chromosomes have complex syntenic relationships (The International Brachypodium Initiative 2010). Thus, the lack of expected syntenic blocks in our study could be due to uneven and low marker coverage of *L. multiflorum* chromosomes. We also found an additional syntenic relationship between chromosomes Lm6 and Bd4, identified neither between wheat and *Brachypodium* nor between barley and *Brachypodium* (The International Brachypodium Initiative 2010). However, the number of markers shared is the lowest we have accepted and this syntenic relationship requires confirmation using additional markers.

The DArTFest resources and marker-trait analysis

As an example of how we can integrate genetic maps and DArT marker sequence information generated in this study, we performed a marker-trait association study in Festulolium populations with divergent freezing tolerance levels. We identified DArT markers associated with the freezing tolerance phenotype on chromosomes 2, 4 and 7, corresponding to QTLs and chromosome segments associated with freezing tolerance in previous studies.

One DArT marker (D556159) and one bin (D558701, D562536) associated with freezing tolerance mapped in the same genomic region of chromosome 4 in *F. pratensis* and *L. multiflorum*, respectively. This observation is in line with the results obtained by Kosmala et al. (2006) using GISH, who found that Festulolium hybrid plants carrying a large central part of *F. pratensis* chromosome 4 exhibited increased freezing tolerance. Three DArT markers mapped to a chromosomal region corresponding to a freezing tolerance QTL in *F. pratensis* closely linked to the *Vrn1* gene (Alm et al., unpublished). This QTL is also found in some species of *Triticeae* where it is referred to as Fr-1 (Galiba

et al. 2009). Figure 2 shows a comparative map between LG4 in forage grasses and the syntenic relationship to the *Brachypodium* chromosome 1. The existence of two *Vrn1* paralogs in grasses (Preston and Kellogg 2007) complicates the analysis of syntenic relationships. In rice, the two *Vrn1* paralogs (also referred to as FUL1 and FUL2) are situated on different chromosomes (Os03 and Os07), while both are found on Bd1 in *Brachypodium* (Higgins et al. 2010). Syntenic relationships between forage grass LG4 and both regions with *Vrn1* copies on Bd1 are revealed by blast analysis of the significant DArT markers (Fig. 2). Markers close to both *Vrn1* paralogs were significantly associated with freezing tolerance, and as many as nine significant DArT markers were clustering close to the BdVrn1b/FUL2. This strongly supports the notion that the significant markers on Lm4 and Fp4 reflect the same underlying QTL located close to the *Vrn1* gene.

It has been debated whether the gene underlying Fr-1 QTL is directly involved in freezing tolerance or if FR-1 QTL is a pleiotropic effect of the vernalization response gene *Vrn1*. *Vrn1* is a transcription factor involved in the transition from vegetative to reproductive phase in Pooideae grasses (Cattivelli et al. 2002; Trevaskis et al. 2007). Recently, polymorphisms in *Vrn1* were shown to exhibit a pleiotropic effect on freezing tolerance by directly or indirectly affected expression of genes involved in the cold acclimation pathway (e.g. CBF genes) (Dhillon et al. 2010). This might indicate that the DArT markers associated with freezing tolerance on LG4 in forage grasses are linked to differences in the vernalization control rather than differences in the ability to withstand freezing.

Another DArT marker associated with freezing tolerance (D560530) is mapped to central part of *L. multiflorum* chromosome 2. As in case of chromosome 4, also this chromosome was implicated to carry genes associated with freezing tolerance. Using GISH, Kosmala et al. (2006) identified hybrid Festulolium plants carrying terminal *F. pratensis* segment on the short arm of *L. multiflorum* chromosome 2 with increased tolerance. Moreover, Shinozuka et al. (2006) mapped a glycine-rich RNA binding protein putatively associated with freezing tolerance to the long arm of *L. perenne* chromosome 2, and this could also be a candidate QTL underlying the significant marker on Lm2. Because the genetic position of centromere on chromosome 2 is not known, we were unable to decide on which chromosome arm the marker D560530 resides. The last of DArT markers associated with freezing tolerance (D560433) are mapped to *F. pratensis* chromosome 7, about 1.5 cM proximal to the previously reported QTL on Fp7 (Alm et al., unpublished).

High freezing tolerance in Festulolium hybrids is usually explained by *F. pratensis* introgressions carrying desired alleles into *L. multiflorum* genomic background (Kosmala et al. 2006). We used GISH to investigate genomic constitution of Festulolium hybrids with the aim to of detecting *F. pratensis* introgressions in *L. perrenne* background. However, they were detected rarely in LFT and HFT plants. This observation indicated that the freezing tolerance is either not caused by *F. pratensis* introgressions, at least in this particular case, or that the introgressions were too small to be detected by GISH. This study, however, revealed that several LFT plants were tetraploid. A common notion is that polyploid plants have improved stress tolerance compared to their diploid progenitors (Jackson and Chen 2010, Zhang et al. 2010). However, the tolerance to physical stress as freezing may decrease. A tetraploid C2 population of *F. pratensis* was shown to be less freeze tolerant and display a lower genetic variation than the original diploid population (Larsen 1979, 1994). This is in agreement with our findings as the frequency of tetraploid individuals was significantly higher in (sub)population with low freezing tolerance. The Festulolium FuRs0357 population was developed from crosses between autotetraploid *L. perenne* and tetraploid or diploid *F. pratensis*, with further backcrossing to diploid *L. perenne* for two generations. This scheme should result in diploid hybrid plants with introgressed segments of *F. pratensis* chromosomes. Tetraploid hybrid plants among the BC2 progenies could have originated from unreduced gametes of tetraploid *L. perenne* parent, and if this is the case, they contain nearly pure *Lolium* genomes. It has been observed that autotetraploid *L. perenne* has lower cold tolerance than diploid *L. perenne* (Sugiyama 1998).

Conclusion

This work marks a significant advance in improving genetic maps of important grass species *F. pratensis* and *L. multiflorum*, which are also used frequently as parents to develop increasingly popular hybrid Festulolium cultivars. As the DArTFest array became available recently, we chose to use DArT markers, because of high-throughput and low cost per data point. Sequencing genetically mapped DArT markers not only revealed that a majority of them originated from transcribed low- and single-copy genomic regions, but also allowed to establish syntenic relations with sequenced genomes of rice and *Brachypodium*. All these resources and knowledge mark significant step forward in analyzing genome structure and evolution in important grass species and expand the range of molecular tools available for breeding. Identification of DArT markers associated with freezing tolerance may help in elucidating molecular mechanisms underlying this trait and contribute to the development of resistant varieties of forage grasses.

Acknowledgments We are grateful to Marie Seifertová, MSc. for excellent technical assistance, and to the team at Diversity Arrays Technology Pty for DArT genotyping. This work has been supported by the Ministry of Agriculture of the Czech Republic (grant award NAZV QH71267) and by European Union (grant No. ED0007/01/01 Centre of the Region Haná for Biotechnological and Agricultural Research).

References

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

Alm V, Fang C, Busso CS, Devos KM, Vollan K, Grieg Z, Rognli OA (2003) A linkage map of meadow fescue (*Festuca pratensis* Huds.). Theor Appl Genet 108:25–40

Bennett MD, Leitch IJ (2005) Plant DNA *C*-values database (release 4.0, Oct. 2005). http://www.kew.org/cvalues/

Bolibok-Bragoszewska H, Heller-Uszyńska K, Wenzl P, Uszyński G, Kilian A, Rakoczy-Trojanowska M (2009) DArT markers for the rye genome—genetic diversity and mapping. BMC Genomics 10:578

Bolot S, Abrouk M, Masood-Quraishi U, Stein N, Messing J, Feuillet C, Salse J (2009) The 'inner circle' of the cereal genomes. Curr Opin Plant Biol 12:119–125

Catalán P, Torrecilla P, Rodriguez JÁL, Olmstead RG (2004) Phylogeny of the festucoid grasses of subtribe Loliinae and allies (Poeae, Pooideae) inferred from ITS and trnL-F sequences. Mol Phylogenet Evol 31:517–541

Cattivelli L, Baldi P, Crosatti C, Di Fonzo N, Faccioli P, Grossi M, Mastrangelo AM, Pecchioni N, Stanca AM (2002) Chromosome regions and stress-related sequences involved in resistance to abiotic stress in Triticeae. Plant Mol Biol 48:649–665

Conesa A, Götz S, García-Gómez JM, Terol J, Talón M, Robles M (2005) Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. Bioinformatics 21:3674–3676

Dhillon T, Pearce SP, Stockinger EJ, Distelfeld A, Li C, Knox AK, Vashegyi I, Vágújfalvi A, Galiba G, Dubcovsky J (2010) Regulation of freezing tolerance and flowering in temperate cereals: the VRN-1 connection. Plant Physiol 153:1846–1858

Ergon Å, Fang C, Jørgensen Ø, Aamlid TS, Rognli OA (2006) Quantitative trait loci controlling vernalisation requirement heading time and number of panicles in meadow fescue (*Festuca pratensis* Huds.). Theor Appl Genet 112:232–242

Fjellheim S, Rognli OA, Fosnes K, Brochmann C (2006) Recent bottlenecking in the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. J Biogeogr 33:1470–1478

Galiba G, Vágújfalvi A, Li C, Soltész A, Dubcovsky J (2009) Regulatory genes involved in the determination of frost tolerance in temperate cereals. Plant Sci 176:12–19

Hanada K, Zou C, Lehti-Shiu MD, Shinozaki K, Shiu SH (2008) Importance of lineage-specific expansion of plant tandem duplicates in the adaptive response to environmental stimuli. Plant Physiol 148:993–1003

Higgins JA, Bailey PC, Laurie DA (2010) Comparative genomics of flowering time pathways using *Brachypodium distachyon* as a model for the temperate grasses. PLoS ONE 5:e10065

Humphreys J, Harper JA, Armstead IP, Humphreys MW (2005) Introgression-mapping of genes for drought resistance transferred from *Festuca arundinacea* var. *glaucescens* into *Lolium multiflorum*. Theor Appl Genet 110:579–587

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

Jackson S, Chen J (2010) Genomic and expression plasticity of polyploidy. Curr Opin Plant Biol 13:153–159

Kopecký D, Lukaszewski AJ, Doležel J (2005) Genomic constitution of Festulolium cultivars released in the Czech Republic. Plant Breeding 124:454–458

Kopecký D, Bartoš J, Lukaszewski AJ, Baird JH, Černoch V, Kölliker R, Rognli OA, Blois H, Caig V, Lübberstedt T, Studer B, Shaw P, Doležel J, Andrzej Kilian A (2009) Development and mapping of DArT markers within the Festuca–Lolium complex. BMC Genomics 10:473

Kopecký D, Bartoš J, Christelová P, Černoch V, Kilian A, Doležel J (2011) Genomic constitution of *Festuca* × *Lolium* hybrids revealed by the DArTFest array. Theor Appl Genet. doi:10.1007/s00122-010-1451-1

Kosmala A, Zwierzykowski Z, Gasior D, Rapacz M, Zwierzykowska E, Humphreys MW (2006) GISH/FISH mapping of genes for freezing tolerance transferred from *Festuca pratensis* to *Lolium multiflorum*. Heredity 96:243–251

Larsen A (1978) Freezing tolerance in grasses. Methods for testing in controlled environments. Meld Norg Landbr Høgsk 57:1–56

Larsen A (1979) Freezing tolerance in grasses. Variation within populations and responce to selection. Meld Norg Landbr Høgsk 58:1–28

Larsen A (1994) Breeding winter hardy grasses. Euphytica 77:231–237

Ouyang S, Buell CR (2004) The TIGR plant repeat databases: a collective resource for the identification of repetitive sequences in plants. Nucleic Acids Res 32(Database issue):D360–D363

Peter-Schmid M, Boller B, Kölliker R (2008) Habitat and management affect genetic structure of *Festuca pratensis* but not *Lolium multiflorum* ecotype populations. Plant Breed 127:510–517

Preston JC, Kellogg EA (2007) Conservation and divergence of *APETALA1/FRUITFULL*-like gene function in grasses: evidence from gene expression analyses. Plant J 52:69–81

R Development Core Team (2009) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL http://www.R-project.org

Rizzon C, Ponger L, Gaut BS (2006) Striking similarities in the genomic distribution of tandemly arrayed genes in *Arabidopsis* and rice. PLoS Comput Biol 2:e115

Rognli OA, Saha MC, Bhamidimarri S, van der Hejden S (2010) Fescues. In 'B. Boller et al. (eds) Fodder cops and Amenity grasses, handbook of plant breeding, vol 5. doi:10.1007/978-1-4419-0760-8_11, Springer Science + Business Media, pp 261–292

Shinozuka H, Hisano H, Yoneyama S, Shimamoto Y, Jones ES, Forster JW, Yamada T, Kanazawa A (2006) Gene expression and genetic mapping analyses of a perennial ryegrass glycine-rich RNA-binding protein gene suggest a role in cold adaptation. Mol Genet Genomics 275:399–408

Singh PK, Mergoum M, Adhikari TB, Shah T, Ghavami F, Kianian SF (2010) Genetic and molecular analysis of wheat tan spot resistance effective against Pyrenophora tritici-repentis races 2 and 5. Mol Breed 25:369–379

Storey JD (2002) A direct approach to false discovery rates. J R Stat Soc Ser B 64:479–498

Studer B, Boller B, Herrmann D, Bauer E, Posselt UK, Widmer F, Kölliker R (2006) Genetic mapping reveals a single major QTL

for bacterial wilt resistance in Italian ryegrass (*Lolium multiflorum* Lam.). Theor Appl Genet 113:661–671

Studer B, Boller B, Bauer E, Posselt UK, Widmer F, Kölliker R (2007) Consistent detection of QTLs for crown rust resistance in Italian ryegrass (*Lolium multiflorum* Lam.) across environments and phenotyping methods. Theor Appl Genet 115:9–17

Studer B, Kölliker R, Muylle H, Asp T, Frei U, Roldán-Ruiz I, Barre P, Barth S, Skøt L, Armstead IP, Dolstra O, Roulund N, Nielsen KK, Lübberstedt T (2010) EST-derived SSR markers used as anchor loci for the construction of a consensus linkage map in ryegrass (*Lolium* spp.). BMC Plant Biol 10:177

Sugiyama S (1998) Differentiation in competitive ability and cold tolerance between diploid and tetraploid cultivars in *Lolium perenne*. Euphytica 103:55–59

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

Tinker NA, Kilian A, Wight CP, Heller-Uszynska K, Wenzl P, Rines HW, Bjørnstad A, Howarth CJ, Jannink JL, Anderson JM, Rossnagel BG, Stuthman DD, Sorrells ME, Jackson EW, Tuvesson S, Kolb FL, Olsson O, Federizzi LC, Carson ML, Ohm HW, Molnar SJ, Scoles GJ, Eckstein PE, Bonman JM, Ceplitis A, Langdon T (2009) New DArT markers for oat provide enhanced map coverage and global germplasm characterization. BMC Genomics 10:39

Trevaskis B, Hemming MN, Dennis ES, Peacock WJ (2007) The molecular basis of vernalization-induced flowering in cereals. Trends Plant Sci 12:352–357

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

Wenzl P, Li H, Carling J, Zhou M, Raman H, Paul E, Hearnden P, Maier C, Xia L, Caig V, Ovesná J, Cakir M, Poulsen D, Wang J, Raman R, Smith KP, Muehlbauer GJ, Chalmers KJ, Kleinhofs A, Huttner E, Kilian A (2006) A high-density consensus map of barley linking DArT markers to SSR, RFLP and STS loci and agricultural traits. BMC Genomics 7:206

Wittenberg AH, van der Lee T, Cayla C, Kilian A, Visser RG, Schouten HJ (2005) Validation of the high-throughput marker technology DArT using the model plant *Arabidopsis thaliana*. Mol Genet Genomics 274:30–39

Zhang X-Y, Hu C-G, Yao J-L (2010) Tetraploidization of diploid Dioscorea results in activation of the antioxidant defense system and increased heat tolerance. J Plant Physiol 167:88–94

4.1.2 FLOW SORTING AND SEQUENCING MEADOW FESCUE CHROMOSOME 4F

Kopecký D, Martis M, Číhalíková J, Hřibová E, Vrána J, Bartoš J, Kopecká J, Cattonaro F, Stočes Š, Novák P, Neumann P, Macas J, Šimková H, Studer B, Asp T, Baird JH, Navrátil P, Karafiátová M, Kubaláková M, Šafář J, Mayer K, Doležel J

PLANT PHYSIOLOGY 163: 1323-1337, 2013

Flow Sorting and Sequencing Meadow Fescue Chromosome 4F[1][C][W]

David Kopecký*, Mihaela Martis, Jarmila Číhalíková, Eva Hřibová, Jan Vrána, Jan Bartoš, Jitka Kopecká, Federica Cattonaro, Štěpán Stočes, Petr Novák, Pavel Neumann, Jiří Macas, Hana Šimková, Bruno Studer, Torben Asp, James H. Baird, Petr Navrátil, Miroslava Karafiátová, Marie Kubaláková, Jan Šafář, Klaus Mayer, and Jaroslav Doležel

Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, 78371 Olomouc, Czech Republic (D.K., J.Č., E.H., J.V., J.B., J.K., Š.S., H.Š., P.Na., M.Ka., M.Ku., J.Š., J.D.); Munich Information Center for Protein Sequences/Institute of Bioinformatics and Systems Biology, Helmholtz Center Munich, 85764 Neuherberg, Germany (M.M., K.M.); Istituto di Genomica Applicata and Istituto di Genomica Applicata Technology Services, 33100 Udine, Italy (F.C.); Biology Centre Academy of Sciences of the Czech Republic, Institute of Plant Molecular Biology, 37005 Ceske Budejovice, Czech Republic (P.No., P.Ne., J.M.); Institute for Agricultural Sciences, 8092 Zurich, Switzerland (B.S.); Aarhus University, Department of Molecular Biology and Genetics, Research Centre Flakkebjerg, DK–4200 Slagelse, Denmark (T.A.); and Department of Botany and Plant Sciences, University of California, Riverside, California 92521 (J.H.B.)

ORCID ID: 0000-0002-2834-1734 (D.K.).

The 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 has been used in intergeneric hybridization with ryegrasses (*Lolium* spp.) to produce Festulolium cultivars. In this work, we describe a new approach to analyze the large genome of meadow fescue, which involves the 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 approximately 50× coverage. This provided, to our knowledge, the first insight into the composition of the fescue genome, enabled the construction of the virtual gene order of the chromosome, and facilitated detailed comparative analysis with the sequenced genomes of rice (*Oryza sativa*), *Brachypodium distachyon*, sorghum (*Sorghum bicolor*), and barley (*Hordeum vulgare*). Using GenomeZipper, we were able to confirm the collinearity of chromosome 4F with barley chromosome 4H and the long arm of chromosome 5H. Several new tandem repeats were identified and physically mapped using fluorescence in situ hybridization. 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 the 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 of chromosome structure and evolution at a resolution never achieved before.

Meadow fescue (*Festuca pratensis*) is a grass typically found in hygrophylic and mesophylic 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 3 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 lowland coastal regions in the temperate zone. In such climates, it is being replaced by interspecific hybrids of meadow fescue (or tall fescue) with ryegrasses (*Lolium* spp.), especially with *Lolium multiflorum*.

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 meadow fescue is lagging far behind most other crops. Genomic studies in meadow fescue are complicated by the 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 meadow fescue are limited.

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

* Address correspondence to kopecky@ueb.cas.cz.

The author responsible for distribution of materials integral to the findings presented in this article in accordance with the policy described in the Instructions for Authors (www.plantphysiol.org) is: David Kopecký (kopecky@ueb.cas.cz).

[C] Some figures in this article are displayed in color online but in black and white in the print edition.

[W] The online version of this article contains Web-only data.
www.plantphysiol.org/cgi/doi/10.1104/pp.113.224105

The only existing genetic map was developed by Alm et al. (2003) and consists of 466 RFLP, amplified fragment length polymorphism (AFLP), and isozyme markers with a total length of 658.8 centimorgan (cM). This map was recently enriched by 149 Diversity Arrays Technology (DArT) markers (Bartoš et al., 2011). So far, two bacterial artificial chromosome (BAC) libraries have been developed for meadow fescue. Donnison et al. (2005) constructed a five-genome coverage library and used it to clone candidate orthologous sequences to the CONSTANS-like rice (*Oryza sativa*) Heading date1 (*Hd1*; Photosensitivity1 [*Se1*]) gene in *Lolium perenne* and meadow fescue, while Kopecký et al. (2010) used a partial genomic BAC library (about one 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 7,680 probes derived from methyl-filtered genomic representations of five species from the *Festuca-Lolium* species complex (*L. perenne*, *L. multiflorum*, *F. pratensis*, *Festuca arundinacea*, and *Festuca 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 meadow fescue. 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 protein 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 meadow fescue has been reported by Vigeland et al. (2013).

Recent advances in next-generation sequencing (NGS) technologies, which enable the sequencing of 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 is not available. In such cases, sequencing flow-sorted chromosomes was found to be 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 (*Hordeum vulgare*). This approach is based on the comparisons of chromosomal shotgun sequences against model genomes (rice, sorghum [*Sorghum bicolor*], and *Brachypodium distachyon*) to detect syntenic regions. Similarly, Vitulo et al. (2011), Wicker et al. (2011), and Hernandez et al. (2012) used GenomeZipper for comparative studies of wheat (*Triticum aestivum*) homologous group 1 and wheat chromosomes 5A and 4A with rice, sorghum, *B. distachyon*, 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 (*Secale cereale*; 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 fluorescence in situ hybridization (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 meadow fescue. We have developed flow cytogenetics in this species and used flow cytometric sorting to dissect the meadow fescue genome into individual chromosomes. NGS of isolated chromosomes provided, to our knowledge, the first insights into the genome structure of this species and permitted the 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 the 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 removal from the DNA synthesis inhibitor (recovery time).

Chromosome Analysis and Sorting

Flow cytometry of mitotic metaphase chromosomes permitted the separation of meadow fescue chromosomes based on their size (relative 4,6-diamino-2-phenylindole [DAPI] fluorescence intensity). The histogram of the chromosome DAPI fluorescence (flow karyotype) of meadow fescue ($n = 7$) 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 the identification of particles sorted from individual peaks on flow

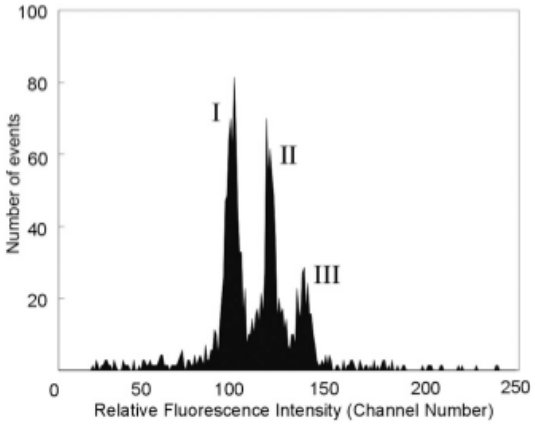


Figure 1. Flow karyotype. The distribution of relative chromosome fluorescence intensity (flow karyotype) was obtained after analysis of a DAPI-stained suspension of mitotic chromosomes of meadow fescue ($2n = 2x = 14$). The flow karyotype consists of two composite peaks, I and II, representing chromosomes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

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 8×10^3 copies of chromosome 4F from a sample prepared from approximately 150 root tips. For DNA amplification and subsequent sequencing, we collected about 40,000 4F chromosomes. The 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 of Chromosome 4F

An equivalent of 50 ng of DNA was obtained by flow cytometric sorting of 40,000 copies of chromosome 4F. This yielded 25 ng of 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) of DNA. For sequencing, individual samples of amplified DNA were combined to reduce a possible bias introduced by the MDA. A total of 4 μg of MDA DNA from chromosome 4F was used for Illumina HiSEquation 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 approximately 50× 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). The assembly with the best L50 (a weighted median statistic such that 50% of the entire assembly is contained in contigs equal to or larger than this value; 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 (1,623 bp). This suggests that the nonrepetitive 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 *B. distachyon*, rice, and sorghum were made. We used stringent BLASTX searches that required 75% or greater (*B. distachyon*)/70% (rice, sorghum) sequence similarity over at least 30 amino acids (Table I). Between 2,629 and 3,056 significant matches were found. A total count of cumulative, nonredundant 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. It should be mentioned that this number does not consider the putatively nonhomologous genes.

Table I. *GenomeZipper results*

General overview of the GenomeZipper results of the 4F chromosome. All numbers given are nonredundant.

Parameter	Chromosome 4F
No. of anchored gene loci	4,672
Lolium spp. SNP markers with match to syntenic genes	109 (out of 154)
No. of anchored barley full-length complementary DNAs	1,697
No. of anchored 4F contigs via first best hit	20,613
No. of anchored 4F contigs via bidirectional BLAST hit	3,700
No. of anchored *Festuca* spp. ESTs	1,444
No. of anchored *B. distachyon* genes	3,056
No. of anchored rice genes	2,629
No. of anchored sorghum genes	2,826

A Virtual Gene Map of Chromosome 4F

Because of the lack of suitable genetic markers for meadow fescue, 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 a closely related species that evolved from the genus *Festuca* about 2.2 million years ago (Polok, 2007). Out of the 154 *L. perenne* markers, 109 (70.8%) could be associated to genes with conserved synteny under these conditions: (1) the best first hit, with (2) a minimal alignment length of 30 amino acids, and (3) an identity of 75% (*B. distachyon*)/70% (rice, sorghum). A virtual ordered gene map was developed using the GenomeZipper protocol (Mayer et al., 2011). With this approach, we identified 4,672 loci, of which 4,626 correspond to regions with conserved synteny in *B. distachyon*, rice, and sorghum (Table I). 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 *B. distachyon* genes reflects a closer phylogenetic relationship of meadow fescue to *B. distachyon* than to rice and sorghum. Furthermore, 1,697 barley full-length complementary DNAs were uniquely associated to either a marker sequence or at least one syntenic gene (Table I).

Comparative Analysis of Chromosome 4F with Relatives

Using GenomeZipper, we performed a detailed comparative analysis of chromosome 4F with sequenced genomes of *B. distachyon*, rice, sorghum, and barley. An ancient translocation differentiates chromosome 4F from 4H of barley (Fig. 3). Chromosome 4F is collinear

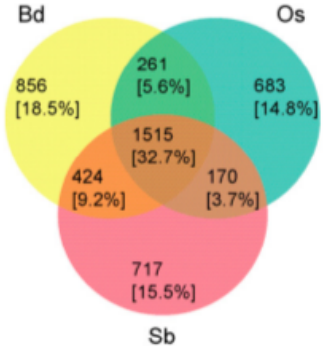


Figure 2. Model species genes anchored to 4F. The Venn diagram shows the number of *B. distachyon* (Bd), rice (Os), and sorghum (Sb) genes that are anchored in the virtual gene map of meadow fescue chromosome 4F. Overlaps of the circles show the number of genes anchored at the same locus. [See online article for color version of this figure.]

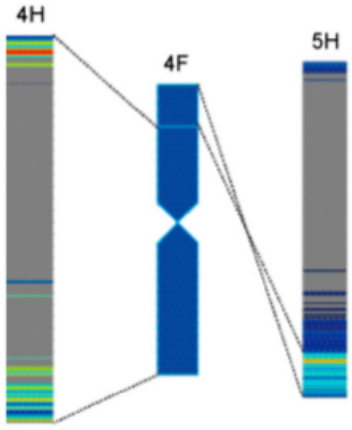


Figure 3. Collinearity of 4F with the barley genome. Meadow fescue 4F Illumina contigs were mapped on the physical map of barley chromosomes 4H and 5H. Syntenic regions are colored yellow-red, whereas nonsyntenic regions are colored dark blue. This approach clearly identifies the collinearity of chromosome 4F with barley chromosome 4H and the terminal part of 5HL. The connectors used indicate the orientations of the particular parts.

with the entire chromosome 4H but also with the terminal region of the long arm of chromosome 5H (5HL). Based on the position of genetic markers, the 5H syntenic segment is in an inverted orientation. For sequences generated from 4F, we identified in silico syntenic regions on *B. distachyon* chromosomes 1 and 4 (Fig. 4), rice chromosomes 3 and 11 (Supplemental Fig. S1), and sorghum chromosomes 1 and 5 (Supplemental Fig. S2).

Comparative mapping results between physical maps of barley chromosomes 4H and 5H, *Festuca* chromosome 4F, *Lolium* chromosome 4L, and wheat chromosome 4A show that the terminal segment originally from the ancestral chromosome 4 and now located at 5HL is not identical to the segment involved in the 4AL/5AL translocation (Naranjo et al., 1987; Supplemental 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* spp. 4F and *Lolium* spp. 4L chromosomes (both carrying ancestral type chromosome 4), while in wheat, the long arm of chromosome 4A is involved (Supplemental Fig. S3). The 5H regions syntenic with 4F and 4A are from 127.96 and 159.79 cM to the telomere (196.85 cM), respectively. While *Lolium* spp. 4L and *Festuca* spp. 4F appear structurally alike, a comparison of *Festuca* spp. 4F against wheat 4A indicates numerous inversions and translocations (Fig. 5). The extensive collinearity between *Lolium* and *Festuca* spp. along with the known genome structure of *Lolium* spp. (Pfeifer et al., 2013) suggest that the collinear 5HL segment on *Festuca* spp. chromosome 4 represents a

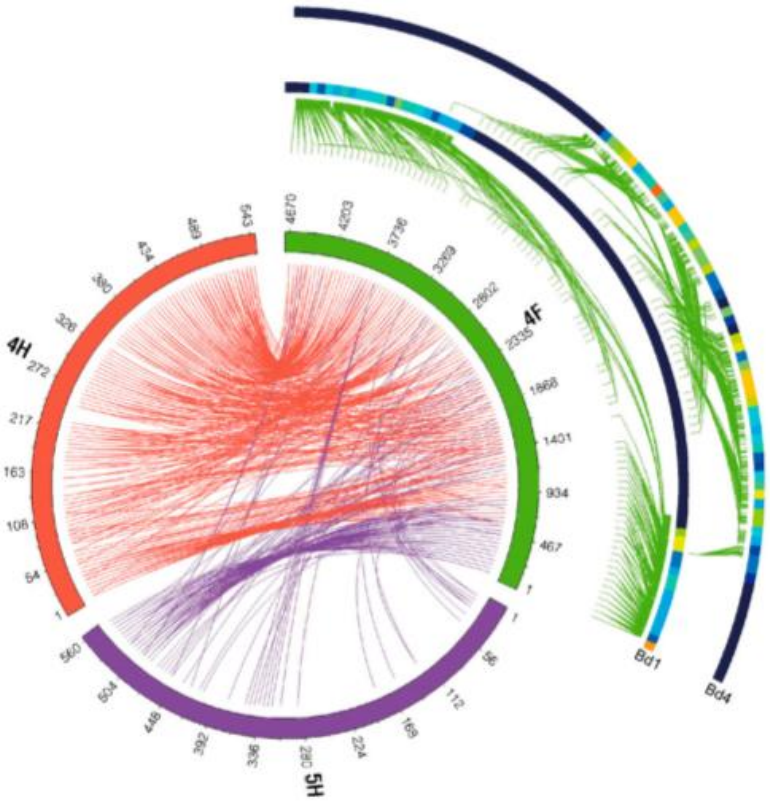


Figure 4. Collinearity of 4F with *B. distachyon*. High-density comparative analyses show the linear gene order of the meadow fescue GenomeZipper versus the sequenced genomes of barley and *B. distachyon*. The concentric circles are as follows. The inner circle represents chromosome 4F (green) and collinear chromosomes 4H (red) and 5H (purple) of barley. The outer circles represent homologous chromosomes of *B. distachyon* (Bd1 and Bd4). The heat maps illustrate the density of genes hit by the contigs from the 4F chromosome. Syntenic regions are colored red, whereas nonsyntenic regions are colored 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.

unidirectional translocation in barley rather than a reciprocal translocation. In the case of reciprocal translocation, we would expect to observe some sequences being present in the terminal region of the short arm of barley chromosome 4H (4HS) and absent in 4F (Figs. 3 and 5A). However, this was not the case, and all sequences detected in the terminal region of 4HS were also present on chromosome 4F. This interpretation is consistent with observations made in *Lolium* spp. (Pfeifer et al., 2013). Since *Festuca* (and *Lolium*) spp. 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* genera split.

The linear ordered gene maps of *Festuca* spp. chromosome 4F, wheat chromosome 4A, and barley chromosome 4H allow one to analyze the number of syntenically conserved genes against the reference genomes of *B. distachyon*, rice, and sorghum (Fig. 6). The proportion of shared syntenic genes on 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* spp. 4F but not for barley 4H and wheat 4A (30%–47%) were found illustrating the higher degree of syntenic conservation and, at least in part, a closer evolutionary relationship of *Festuca* spp. 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 1× coverage of chromosome 4F (550 Mb). Graph-based clustering resulted in 252,144 clusters containing 4,224,228 sequence reads (76%). A total of 855 clusters contained at least 100 sequences, from which 450 clusters contained more than 555 sequences (0.01% of analyzed reads; Supplemental 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

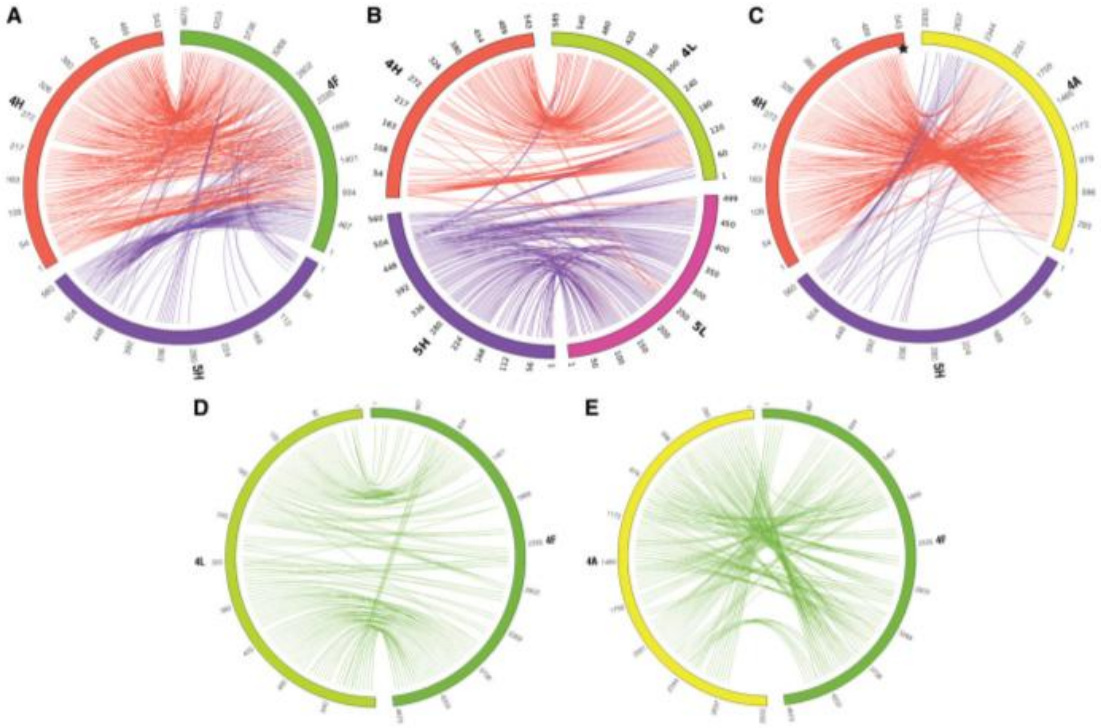


Figure 5. Chromosome 4/5 translocation. The top three circles (A–C) show the chromosome 4/5 translocation in three different grass chromosomes (*Festuca* spp. 4F, wheat 4A, and *Lolium* spp. 4L) in comparison with the corresponding homologous chromosomes 4 and 5 in barley. The location of the segment from 5HL for both *Festuca* (A) and *Lolium* (B) spp. is 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* spp. rather than a reciprocal translocation, as found in wheat (Alm et al., 2003). The bottom two circles depict the collinearity between meadow fescue chromosome 4F and *L. perenne* chromosome 4L (D) and wheat chromosome 4A (E). While *Lolium* and *Festuca* spp. are highly collinear, the synteny between wheat and *Festuca* spp. is interrupted by several inversions and translocations.

species-specific and/or chromosome-specific cytogenetic markers.

Annotation of the largest clusters led to the classification of approximately 53% of the 4F chromosome sequence. The 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 that account for more than 8% of the analyzed sequence reads of chromosome 4F. These sequences can represent unknown DNA repeats and/or whole-genome amplification artifacts (Fig. 7).

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 (Supplemental Fig. S5). The analysis revealed the presence of tandem or semitandem organized repeats in 43 clusters, which together represented about 2.71% of the 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* spp. genomes (Supplemental Table S1). A set of primers specific for tandem units was designed and used for the preparation of probes for FISH and for Southern hybridization. PCR products obtained after amplification on genomic DNA of meadow fescue as the template DNA were sequenced using dideoxy-chain termination reaction to check their authenticity. Moreover, PCR was done on flow-sorted chromosomes of

Figure 6. Conserved syntenic regions in the virtual linear ordered maps of *Festuca* spp., barley, wheat, and *Lolium* spp. chromosome 4 as defined by comparison with model grass genomes. The intersecting circles of the Venn diagrams depict the percentage of syntenic conserved reference genes (*B. distachyon*, rice, sorghum) shared by meadow fescue chromosome 4F (red), wheat chromosome 4A (orange), and barley chromosome 4H (blue). [See online article for color version of this figure.]

meadow fescue 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 meadow fescue flow karyotype and that positive PCR products were obtained for all 15 selected tandem repeats also on *Lolium* spp. genomic DNA (Supplemental 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 that were successfully mapped on *Festuca* spp. mitotic chromosomes using FISH (see below). A ladder-like pattern that is typical for tandem organized repetitive units was obtained for all eight putative tandem repeats that gave visible signals after FISH (see below; Fig. 9). As shown in Figure 10, the tandem repeats identified in sequences of chromosome 4F of meadow fescue are also present in the nuclear genome of *L. perenne*, although 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; Supplemental Table S1). All of them produced hybridization patterns also on other chromosomes. Additionally, the other three fpTRs mapped on chromosomes other than 4F (fpTR6, fpTR11, and fpTR12; Supplemental Table S1). This might be caused by the sensitivity of FISH. None of the fpTRs produced signal

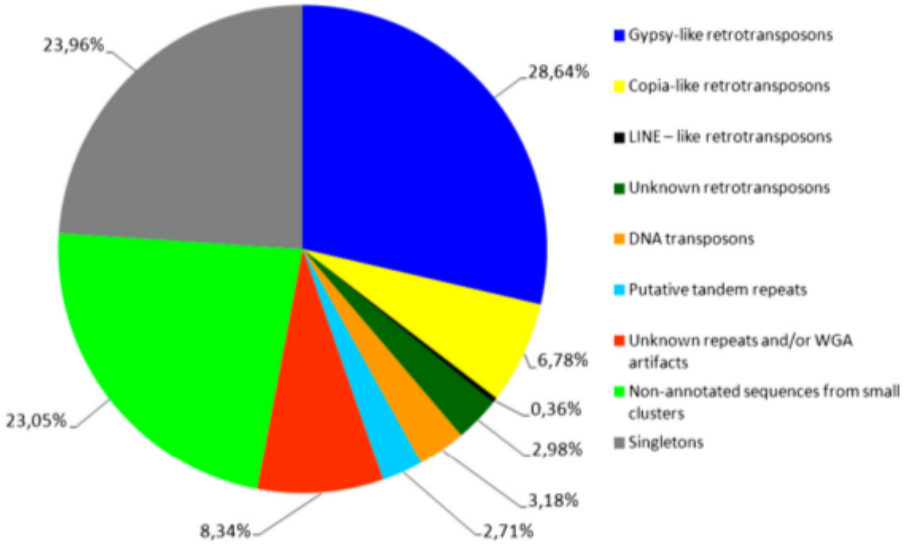


Figure 7. Repeats on 4F. The proportion of different repetitive elements in sequence reads of 4F chromosome is shown. [See online article for color version of this figure.]

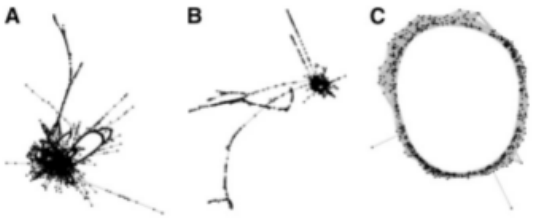


Figure 8. Graph layouts visualized by SeqGrapheR. Examples of graph layouts were calculated using the Fruhterman-Reingold algorithm for different types of tandem organized repeats: fpTR1 (A), fpTR5 (B), and fpTR15 (C). Dots and lines represent sequence reads and similarity hits between them, respectively.

exclusively on 4F; thus, they cannot be considered as chromosome specific. On the other hand, our observation opens the 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* × meadow fescue (Fig. 9), the repeats may find even broader use as probes for FISH in cytogenetics of the *Festuca-Lolium* spp. complex.

Using FISH, we were able to localize a telomeric repeat on *Festuca* spp. 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 on 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 behind 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 for *L. perenne* and 1C = 2,567 Mb for *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 that avoid sequencing of the repetitive parts of genomes. Of these, sequencing of complementary DNA to generate ESTs 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; Peterson 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. In hexaploid wheat, individual chromosomes represent only 3.6% to 5.9% of the entire genome (Doležel et al., 2009), and even in a diploid species such as meadow fescue, isolation of individual chromosomes dissects the 3.2-Gb genome into 373- to 543-Mb units, each representing 11.7% to 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*-meadow fescue, 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 chromosome sequences is a powerful tool for comparative studies. It significantly increases the resolution level relative to genetic markers. In the previous report

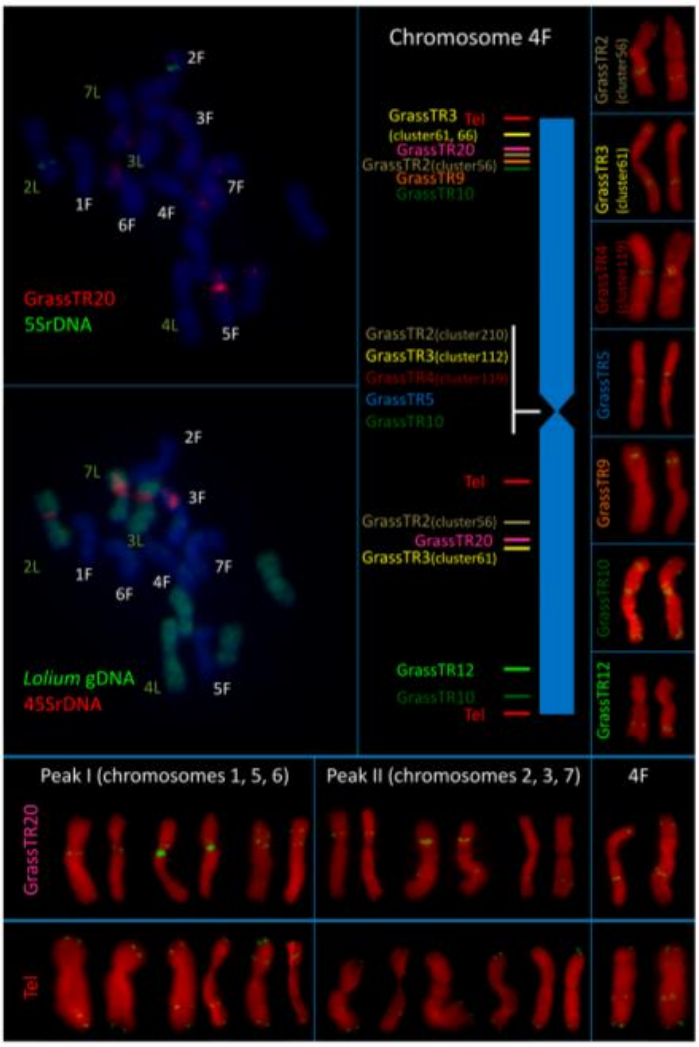


Figure 9. Cytogenetic mapping of repeats. Cytogenetic mapping shows newly identified tandem repeats and telomeric repeats on sorted chromosomes of meadow fescue and the F1 hybrid of *L. perenne* × meadow fescue.

on the collinearity of the meadow fescue genome with rice, sorghum, *Lolium* spp., oat (*Avena sativa*), maize (*Zea mays*), and Triticeae, a limited number of markers were used (Alm et al., 2003). Prior to this study, only 36 markers were available from chromosome 4F, with only a subset (seven to 30) suitable for any pairwise comparison. This contrasts with thousands of sequences from individual chromosomes (4F here) of model plant species (rice, sorghum, *B. distachyon*, and barley).

Estimation of Gene Content

The GenomeZipper approach permitted the estimation of gene content in meadow fescue and, by implication, all related *Festuca* spp. Assuming the estimated size of chromosome 4F at 543 Mb and the 4F gene density representative for the entire genus *Festuca* genome (Kopecký et al., 2010), the total genus *Festuca* genome contains about 35,000 genes. This is a slightly higher number that the estimation for barley (32,000 or more 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* spp. and barley (28,000 genes; Hernandez et al., 2012). In *L. perenne*, the most closely related species for which DNA sequence data are available, almost 25,700 matches with *B. distachyon* genes were obtained (Byrne et al., 2011).

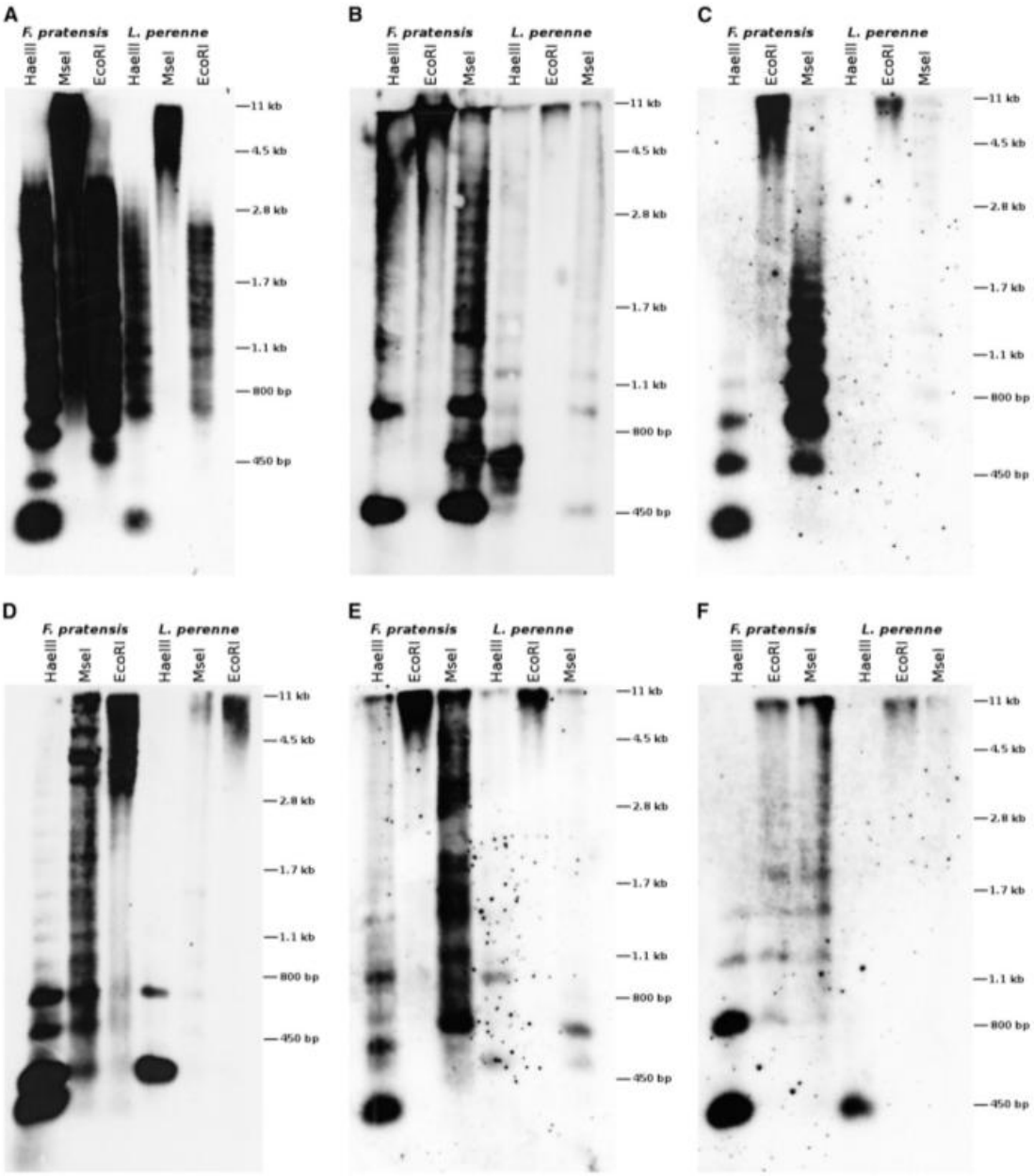


Figure 10. Southern hybridization with the probes for tandem repeats. Examples show Southern hybridization of meadow fescue and *L. perenne* genomic DNA digested using three different restriction endonucleases with the probes for tandem organized repeats fpTR1 (A), fpTR4 (B), fpTR5 (C), fpTR7 (D), fpTR12 (E), and fpTR15 (F). The ladder-like pattern indicates the presence of tandem organized repetitive units and supports the results of dotter analysis as well as the FISH analysis.

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 gene 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 simple sequence repeat, presence/absence variation, and single-nucleotide polymorphism (SNP), and help in developing SNP platforms for marker-assisted breeding of this agriculturally valuable species.

4F Is an Ancestral Chromosome toward Modern Cereals

Based on the comparative analysis, chromosome 4F appears collinear with the entire barley chromosome 4H as well as with the terminal region of 5HL (Fig. 3). The observed homology of 4F to chromosome arms 4S, 4L, and 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, meadow fescue chromosome 4F is likely more ancient than barley chromosome 4H. In cereal genomes (the A, B, and D genomes of wheat and the H genome of barley), the direct ancestral precursor that 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 the *Festuca* and *Lolium* genera from the grass lineage between 25 million years ago (divergence of Triticeae and Poeae) and 11.6 million years ago (divergence of barley and wheat genomes; Chalupska et al., 2008). In *Festuca* and *Lolium* spp., 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* spp., 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 4-fold increase in recombination frequency was observed in the region surrounding the translocation break point (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 the physical localization of a sequence of interest directly on chromosomes. Cytogenetic mapping has many applications, ranging from the study of structural chromosome changes (Mandáková and Lysák, 2008) to the determination of the positions and orientation of unassembled BAC contigs to support the 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 sources of useful cytogenetic markers. Satellites are tandemly repeated sequences with a repeat unit greater than 25 bp. 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* spp. (Katsiotis et al., 1997), banana (*Musa* spp.; Hřibová et al., 2007), and other plant species (Sharma and Raina, 2005) because they tend to form clusters, which facilitates their detection and positive chromosome identification. In this study, 11 newly characterized satellites provided localized signals on one or more chromosomes of meadow fescue and could serve as new cytogenetic markers in this species and probably also in ryegrasses.

Chromosomes of *Festuca* and *Lolium* spp. can be readily discriminated in hybrids using genomic in situ hybridization (Thomas et al., 1994; Kopecký et al., 2008a, 2008b). This method enables the 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 the presence or signal intensity of individual fpTRs. Hence, none or just a few of the 15 tested appears responsible for distinguishing parental genomes. This indicates that tandem repeats are not responsible for the discrimination capacity of genomic in situ hybridization. However, newly identified tandem repeats provided new sources of cytogenetic markers and significantly increased levels of chromosome identification.

CONCLUSION

To our knowledge, this is the first report on the 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 the

genome organization of the species. The sequence data were 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 (*B. distachyon*, 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 NGS technologies is a powerful approach that provides insights on chromosome structure and evolution at superior resolution.

MATERIALS AND METHODS

Plant Material

For chromosome sorting, seeds of meadow fescue (*Festuca pratensis* 'Fure'; $2n = 2x = 14$) were obtained from Dr. Arild Larson (Graminor). Seeds were germinated in the dark at 25°C distributed in petri dishes filled with distilled water for 5 to 7 d to achieve optimal root length (approximately 2–3 cm). For additional in situ hybridization experiments, *Lolium perenne* 'SR4220', meadow fescue cv Fure, and a diploid F1 hybrid of *L. perenne* cv SR4220 × meadow fescue cv Skawa were used. The plants was transferred to a hydroponic culture of Hydroponex at 0.9 g L^{-1} (Hu-Ben); after 5 to 7 d, actively growing root tips were collected to ice water for approximately 28 h, fixed in a 3:1 mixture of absolute ethanol and glacial acetic acid at 37°C for 7 d, stained in 1% acetocarmine for 2 h, 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 the 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 nutrient solution (Gamborg and Wetter, 1975) with 1, 1.5, or 2 mM HU. After 18 h of incubation, the solution was replaced by a HU-free Hoagland solution. Samples of root tips were taken at 1-h intervals for up to 6 h, and the cell cycle synchrony and mitotic activity were examined with a microscope. To accumulate cells at metaphase, the seedlings were treated for 2 h with Hoagland 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 (approximately 2 mm long) were cut and transferred to a vial tube containing 750 μL of LB01 buffer (Doležel et al., 1989). The chromosomes were isolated after homogenization with a Polytron PT1300 homogenizer (Kinematica) 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 DAPI adjusted to a final concentration of 2 μg mL^{-1}.

Chromosome Analysis and Sorting

Chromosome analysis and sorting were done on a FACSAria II SORP flow cytometer (BD Biosciences) equipped with a 100-mW, 488-nm laser for scattered light detection and a 100-mW, 355-nm laser for DAPI excitation. Suspension of chromosomes was analyzed at rates of 800 to 1,000 events per second. Approximately 20,000 to 50,000 chromosomes were analyzed in each sample. DAPI fluorescence was collected through a 450/50-nm band-pass filter. To discriminate doublets, dot plots of DAPI fluorescence area versus DAPI fluorescence width were used. To ensure high purities of the sorted chromosomes, a tight gate was created around the chromosome 4F population, and chromosomes were sorted at rates of four to six chromosomes per second. To evaluate the purity of chromosome sorting, approximately 1,000 chromosomes were sorted onto microscope 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 the Purity of Sorted Chromosomes Using FISH

To identify flow-sorted chromosomes and determine the extent of contamination of the 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 meadow fescue (Kopecký et al., 2010). The DNA was labeled by the DIG-Nick Translation Kit or the biotin-Nick Translation Kit (Roche Applied Science) 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-fluorescein isothiocyanate (Roche) and streptavidin-Cy3 (Amersham), and counterstaining was with DAPI in the Vectashield antifade solution (Vector Laboratories). Preparations were screened with an Olympus AX70 microscope with epifluorescence and a SensiCam B/W camera.

Amplification and Sequencing of Chromosome 4F

Flow-sorted chromosomes were treated with proteinase, and their DNA was subsequently purified using Microcon YM-100 columns (Millipore) as described by Šimková et al. (2008) with minor modifications. Chromosomal DNA was amplified by MDA using the Illustra GenomiPhi V2 DNA Amplification Kit (GE Healthcare Bio-Sciences). Sequencing of amplified chromosomal DNA was performed with HiSEquation 2000 (Illumina). Four micrograms of MDA-amplified DNA was used to create the corresponding shotgun DNA-seq library. The library for the 4F chromosome was run in a single lane at Istituto di Genomica Applicata Technology Services.

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 version 8.6.2 repeat library by applying the following parameters: 70% identity cutoff, 100-bp minimal length, seed length 14, exdrop 5, and e-value 0.001. The Munich Information Center for Protein Sequences repeat library contains known grass transposons from the Triticeae Repeat Database (http://wheat.pw.usda.gov/ITMI/Repeats) as well as de novo-detected long terminal repeat 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 on meadow fescue chromosome 4, the repeat-filtered contigs were aligned against the protein sequences of *Brachypodium distachyon* (version 1.2), rice (*Oryza sativa*; rice RAP-DB genome build 4), and sorghum (*Sorghum bicolor*; version 1.4) by BLASTX. The following stringent filtering criteria were applied: (1) only the first best hit with (2) a minimal alignment length of 30 amino acids and (3) a minimal sequence identity of 75% (*B. distachyon*)/70% (rice, sorghum). A sliding-window approach (window size of 0.5 Mb, shift size of 0.1 Mb) was used to identify segments with conserved gene order, based on the density of homology matches between *Festuca* spp. and the reference genomes.

The extracted conserved genes from the three model grass genomes and the corresponding *Festuca* spp. 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 subsequently 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, *B. distachyon*) get highest

priority for ordering, and subsequently, the additional reference genomes are overlaid onto this scaffold. Due to the absence of a high-quality genetic marker map of meadow fescue, 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* spp. and barley (*Hordeum vulgare*), 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 meadow fescue and barley (at least 85% identity and a minimum alignment length of 100 bp).

To test whether the observed *Festuca* spp. chromosome 4/5 translocation is structurally identical to the 4/5 translocation found in Triticeae (Naranjo et al., 1987), the *Festuca* spp. 4F contigs, ordered gene scaffolds for wheat (*Triticum aestivum*) 4A and *Lolium* spp. 4L, and barley chromosomes 4H and 5H were compared by bidirectional 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* spp. 4F, the *B. distachyon*, rice, and sorghum homologs contained in the individual linear ordered gene maps were compared with 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 1× coverage of chromosome 4F (550 Mb) was extracted from the Illumina data and used for the reconstruction of repetitive elements using a graph-based method according to Novák et al. (2010) using the Louvain clustering algorithm (Blondel et al., 2008). Resulting clusters of sequence reads were manually annotated using several sources, including similarity searches with RepeatMasker, and BLASTX and BLASTN programs were used for similarity searches 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 (Sonnhammer and Durbin, 1995) 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 the 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 meadow fescue, we used a 5S ribosomal DNA (rDNA) probe accompanied by reprobing of slides with a 45S rDNA probe and a probe made from genomic DNA of *L. perenne*. The latter two probes were made using the biotin-Nick Translation Kit and the DIG-Nick Translation Kit (Roche), respectively. For the 45S rDNA probe, DNA clone pTa71 (Gerlach and Bedbrook, 1979) containing a 9-kb *Eco*RI fragment of wheat rDNA, which carries the 18S-5.8S-26S cluster of ribosomal RNA genes, was used. The probe for 5S rDNA was prepared using PCR with a pair of specific primers (RICRGAC1 and RICRGAC2) that 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-digoxigenin-fluorescein isothiocyanate and avidin-Cy3 conjugates, and counterstaining was by 0.2 mg mL^{-1} DAPI in Vectashield antifade solution (Vector Laboratories). Slides were evaluated with an Olympus AX70 microscope equipped with epifluorescence and a SensiCam B/W camera. ScionImage and Adobe Photoshop software were used for processing of color images.

Southern Hybridization

Aliquots of genomic DNA samples corresponding to 3×10^8 of the nuclear genomes of meadow fescue cv Fure and *L. perenne* cv SR4220 were digested using restriction enzymes *Hin*dIII, *Mse*I, and *Eco*RI, size fractionated by 1.2% agarose gel electrophoresis, and transferred to Hybond N^+ nylon membranes (Amersham). Probes specific for selected putative tandem repeats were amplified using specific primers (Table I) and labeled with biotin. The Southern hybridization was done at 68°C overnight, and signals were detected using the BrightStar BioDetect kit according to the manufacturer's instructions to 90% stringency (Ambion), incubated with chemiluminescent substrate (CDP-Star; Amersham Biosciences), and exposed on x-ray film.

Sequence data from this article can be found in the GenBank/EMBL data libraries under accession numbers JX624129 to JX624136.

Supplemental Data

The following materials are available in the online version of this article.

Supplemental Figure S1. Collinearity of 4F with rice.

Supplemental Figure S2. Collinearity of 4F with sorghum.

Supplemental Figure S3. Collinearity of 4F, 4A of wheat, and 5H of barley.

Supplemental Figure S4. Distribution of chromosome 4F reads in sequence clusters.

Supplemental Figure S5. Dot plot of five tandem repeats.

Supplemental Table S1. Summary of the detection of tandem repeats.

ACKNOWLEDGMENTS

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

Received June 26, 2013; accepted October 4, 2013; published October 4, 2013.

LITERATURE CITED

Alm V, Fang C, Busso CS, Devos KM, Vollan K, Grieg Z, Rognli OA (2003) A linkage map of meadow fescue (*Festuca pratensis* Huds.) and comparative mapping with other Poaceae species. Theor Appl Genet **108:** 25–40

Baird JH, Kopecký D, Lukaszewski AJ, Green RL, Bartoš J, Doležel J (2012) Genetic diversity of turf-type tall fescue using diversity arrays technology. Crop Sci **52:** 408–412

Bartoš J, Sandve SR, Kölliker R, Kopecký D, Christelová P, Stočes S, Østrem L, Larsen A, Kilian A, Rognli OA, et al (2011) Genetic mapping of DArT markers in the *Festuca-Lolium* complex and their use in freezing tolerance association analysis. Theor Appl Genet **122:** 1133–1147

Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E (2008) Fast unfolding of communities in large networks. J Stat Mech **2008:** P10008

Brandes A, Röder MS, Ganal MW (1995) Barley telomeres are associated with two different types of satellite DNA sequences. Chromosome Res **3:** 315–320

Byrne S, Panitz F, Hedegaard J, Bendixen C, Studer B, Farrell JD, Swain SC, Armstead I, Caccamo M, Asp T (2011) *De novo* genome sequencing of perennial ryegrass (*Lolium perenne*). *In* S Barth, D Milbourne, eds, Abstracts of the EUCARPIA 29th Fodder Crops and Amenity Grasses Section Meeting. Teagasc, Dublin, p 23

Chalupska D, Lee HY, Faris JD, Evrard A, Chalhoub B, Haselkorn R, Gornicki P (2008) Acc homoeoloci and the evolution of wheat genomes. Proc Natl Acad Sci USA **105:** 9691–9696

Choulet F, Wicker T, Rustenholz C, Paux E, Salse J, Leroy P, Schlub S, Le Paslier M-C, Magdelenat G, Gonthier C, et al (2010) Megabase level sequencing reveals contrasted organization and evolution patterns of the wheat gene and transposable element spaces. Plant Cell **22:** 1686–1701

Doležel J, Bartoš J, Voglmayr H, Greilhuber J (2003) Nuclear DNA content and genome size of trout and human. Cytometry A **51:** 127–128, author reply 129

Doležel J, Binarová P, Lucretti S (1989) Analysis of nuclear DNA content in plant cells by flow cytometry. Biol Plant **31:** 113–120

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

Doležel J, Šimková H, Kubaláková M, Šafář J, Suchánková P, Číhalíková J, Bartoš J, Valárik M (2009) Chromosome genomics in the Triticeae. *In* C Feuillet, GJ Muehlbauer, eds, Genetics and Genomics of the Triticeae. Plant Genetics and Genomics: Crops and Models 7. Springer Science +Business Media, New York, pp 285–316

Donnison IS, O'Sullivan DM, Thomas A, Canter P, Moore B, Armstead I, Thomas H, Edwards KJ, King IP (2005) Construction of a *Festuca pratensis* BAC library for map-based cloning in *Festulolium* substitution lines. Theor Appl Genet **110:** 846–851

Egan AN, Schlueter J, Spooner DM (2012) Applications of next-generation sequencing in plant biology. Am J Bot **99:** 175–185

Fukui K, Kamisugi Y, Sakai F (1994) Physical mapping of 5S rDNA loci by direct-cloned biotinylated probes in barley chromosomes. Genome **37:** 105–111

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

Gamborg OL, Wetter LR (1975) Plant Tissue Culture Methods. National Research Council of Canada, Saskatoon, Canada

Gerlach WL, Bedbrook JR (1979) Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res **7:** 1869–1885

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

Hřibová E, Dolezelová M, Town CD, Macas J, Doležel J (2007) Isolation and characterization of the highly repeated fraction of the banana genome. Cytogenet Genome Res **119:** 268–274

Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res **29:** E25

Katsiotis A, Hagidimitriou M, Heslop-Harrison JS (1997) The close relationship between the A and B genomes in *Avena* L. (Poaceae) determined by molecular cytogenetic analysis of total genomic, tandemly and dispersed repetitive DNA sequences. Ann Bot (Lond) **79:** 103–109

Kopecký D, Bartoš J, Christelová P, Černoch V, Kilian A, Doležel J (2011) Genomic constitution of *Festuca* × *Lolium* hybrids revealed by the DArTFest array. Theor Appl Genet **122:** 355–363

Kopecký D, Bartoš J, Lukaszewski AJ, Baird JH, Černoch V, Kölliker R, Rognli OA, Blois H, Caig V, Lübberstedt T, et al (2009) Development and mapping of DArT markers within the *Festuca-Lolium* complex. BMC Genomics **10:** 473

Kopecký D, Havránková M, Loureiro J, Castro S, Lukaszewski AJ, Bartoš J, Kopecká J, Doležel J (2010) Physical distribution of homoeologous recombination in individual chromosomes of *Festuca pratensis* in *Lolium multiflorum*. Cytogenet Genome Res **129:** 162–172

Kopecký D, Lukaszewski AJ, Doležel J (2008a) Cytogenetics of Festulolium (Festuca × Lolium hybrids). Cytogenet Genome Res **120:** 370–383

Kopecký D, Lukaszewski AJ, Doležel J (2008b) Meiotic behaviour of individual chromosomes of *Festuca pratensis* in tetraploid *Lolium multiflorum*. Chromosome Res **16:** 987–998

Krzywinski M, Schein J, Birol I, Connors J, Gascoyne R, Horsman D, Jones SJ, Marra MA (2009) Circos: an information aesthetic for comparative genomics. Genome Res **19:** 1639–1645

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

Luo R, Liu B, Xie Y, Li Z, Huang W, Yuan J, He G, Chen Y, Pan Q, Liu Y, et al (2012) SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience **1:** 18

Mandáková T, Lysák MA (2008) Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). Plant Cell **20:** 2559–2570

Martienssen RA, Rabinowicz PD, O'Shaughnessy A, McCombie WR (2004) Sequencing the maize genome. Curr Opin Plant Biol **7:** 102–107

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

Mayer KFX, Martis M, Hedley PE, Simková H, Liu H, Morris JA, Steuernagel B, Taudien S, Roessner S, Gundlach H, et al (2011) Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell **23:** 1249–1263

Mayer KF, Taudien S, Martis M, Simkova H, Suchankova P, Gundlach H, Wicker T, Petzold A, Felder M, Steuernagel B, et al (2009) Gene content and virtual gene order of barley chromosome 1H. Plant Physiol **151:** 496–505

Mayer KF, Waugh R, Brown JW, Schulman A, Langridge P, Platzer M, Fincher GB, Muehlbauer GJ, Sato K, Close TJ, et al (2012) A physical, genetic and functional sequence assembly of the barley genome. Nature **491:** 711–716

Naranjo T, Roca A, Goicoecha PG, Giraldz R (1987) Arm homoeology of wheat and rye chromosomes. Genome **29:** 873–882

Novák P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics **11:** 378

Paux E, Sourdille P, Salse J, Saintenac C, Choulet F, Leroy P, Korol A, Michalak M, Kianian S, Spielmeyer W, et al (2008) A physical map of the 1-gigabase bread wheat chromosome 3B. Science **322:** 101–104

Pedrosa-Harand A, Kami J, Gepts P, Geffroy V, Schweizer D (2009) Cytogenetic mapping of common bean chromosomes reveals a less compartmentalized small-genome plant species. Chromosome Res **17:** 405–417

Peterson DG, Wessler SR, Paterson AH (2002) Efficient capture of unique sequences from eukaryotic genomes. Trends Genet **18:** 547–550

Pfeifer M, Martis M, Asp T, Mayer KFX, Lübberstedt T, Byrne S, Frei U, Studer B (2013) The perennial ryegrass GenomeZipper: targeted use of genome resources for comparative grass genomics. Plant Physiol **161:** 571–582

Polok K (2007) Molecular Evolution of the Genus *Lolium* L. Studio Poligrafii Komputerowej, Olsztyn, Poland

Rabinowicz PD, Schutz K, Dedhia N, Yordan C, Pamell LD, Stein L, McCombie WR, Martienssen RA (1999) Differential methylation of genes and retrotransposons facilitates shotgun sequencing of the maize genome. Nat Genet **23:** 305–308

Rayburn AL, Gill BS (1986) Molecular identification of the D-genome chromosomes of beat. J Hered **77:** 253–255

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

Schwarzacher T, Heslop-Harrison P (2000) Practical in Situ Hybridization. BIOS Scientific Publishers, Oxford

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

Sharma S, Raina SN (2005) Organization and evolution of highly repeated satellite DNA sequences in plant chromosomes. Cytogenet Genome Res **109:** 15–26

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

Sonnhammer ELL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene **167:** GC1–GC10

Studer B, Byrne S, Nielsen RO, Panitz F, Bendixen C, Islam MS, Pfeifer M, Lübberstedt T, Asp T (2012) A transcriptome map of perennial ryegrass (*Lolium perenne* L.). BMC Genomics **13:** 140

Suchánková P, Kubaláková M, Kovárová P, Bartoš J, Číhalíková J, Molnár-Láng M, Endo TR, Doležel J (2006) Dissection of the nuclear genome of barley by chromosome flow sorting. Theor Appl Genet **113:** 651–659

Thiel T, Graner A, Waugh R, Grosse I, Close TJ, Stein N (2009) Evidence and evolutionary analysis of ancient whole-genome duplication in barley predating the divergence from rice. BMC Evol Biol **9:** 209

Thomas HM, Morgan WG, Meredith MR, Humphreys MW, Thomas H, Leggett JM (1994) Identification of parental and recombined chromosomes in hybrid derivatives of *Lolium multiflorum* × *Festuca pratensis* by genomic *in situ* hybridization. Theor Appl Genet **88:** 909–913

Tomaszewski C, Byrne S, Foito A, Kildea S, Kopecký D, Doležel J, Heslop-Harrison JS, Stewart D, Barth S (2012) A genetic linkage map of an F2 perennial ryegrass population highly enriched with DArT markers suitable for mapping of crown rust resistence QTL. Plant Breed **131:** 345–349

Tsujimoto H, Mukai Y, Akagawa K, Nagaki K, Fujigaki J, Yamamoto M, Sasakuma T (1997) Identification of individual barley chromosomes based on repetitive sequences: conservative distribution of *Afa*-family repetitive sequences on the chromosomes of barley and wheat. Genes Genet Syst **72:** 303–309

Vigeland MD, Spannagl M, Asp T, Paina C, Rudi H, Rognli OA, Fjellheim S, Sandve SR (2013) Evidence for adaptive evolution of low-temperature stress response genes in a Pooideae grass ancestor. New Phytol **199:** 1060–1068

Vitulo N, Albiero A, Forcato C, Campagna D, Dal Pero F, Bagnaresi P, Colaiacovo M, Faccioli P, Lamontanara A, Šimková H, 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, Simková H, Cíhalíková J, Lysák MA, Doležel J (2000) Flow sorting of mitotic chromosomes in common wheat (*Triticum aestivum* L.). Genetics **156:** 2033–2041

Wenzl P, Suchánková P, Carling J, Simková H, Huttner E, Kubaláková M, Sourdille P, Paul E, Feuillet C, Kilian A, et al (2010) Isolated chromosomes as a new and efficient source of DArT markers for the saturation of genetic maps. Theor Appl Genet **121:** 465–474

Wicker T, Mayer KFX, Gundlach H, Martis M, Steuernagel B, Scholz U, Šimková H, Kubaláková M, Choulet F, Taudien S, et al (2011) Frequent gene movement and pseudogene evolution is common to the large and complex genomes of wheat, barley, and their relatives. Plant Cell **23:** 1706–1718

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

4.1.3 REPETITIVE DNA: A VERSATILE TOOL FOR KARYOTYPING IN *FESTUCA PRATENSIS* HUDS.

Anna Křivánková, David Kopecký, Štěpán Stočes, Jaroslav Doležel and Eva Hřibová

Cytogenetic and Genome Research 151:96-105, DOI:10.1159/000462915, 2017

Original Article

Cytogenetic and Genome Research

Cytogenet Genome Res 2017;151:96–105
DOI: 10.1159/000462915

Accepted: October 18, 2016
by M. Schmid
Published online: March 24, 2017

Repetitive DNA: A Versatile Tool for Karyotyping in *Festuca pratensis* Huds.

Anna Křivánková David Kopecký Štěpán Stočes Jaroslav Doležel
Eva Hřibová

Institute of Experimental Botany, Centre of the Region Haná for Biotechnical and Agricultural Research, Olomouc, Czech Republic

Keywords
Fluorescence in situ hybridization · Karyotyping · Meadow fescue · Repetitive DNA · Tandem organized repeats

Abstract
FISH is a useful method to identify individual chromosomes in a karyotype and to discover their structural changes accompanying genome evolution and speciation. DNA probes for FISH should be chromosome specific and/or exhibit specific patterns of distribution along each chromosome. Such probes are not available in many plants including meadow fescue (*Festuca pratensis* Huds.), an important forage grass species. In the present study, various DNA repeats identified in Illumina shotgun sequences specific to chromosome 4F of *F. pratensis* were used as probes for FISH to develop the molecular karyotype of meadow fescue and to reveal a long-range molecular organization of its chromosomes. Five tandem repeats produced specific patterns on individual chromosomes. Their use in combination with probes for rRNA genes enabled the establishment of the molecular karyotype of meadow fescue. Most of the mobile genetic elements were dispersed along all the chromosomes except for the DNA transposon CACTA, which was localized preferentially to telomeric and subtelomeric regions, and a putative LTR element, which was localized to (peri)centromeric regions. Cytogenetic mapping of the 5 tandem repeats in other accessions of meadow fescue showed a highly similar distribution and confirmed the versatility and robustness of these probes.

Fescue (*Festuca* L.) is an economically important genus containing over 500 species of perennial grasses, widely distributed over all continents except Antarctica. Fescues are grown for ornamental and turf purposes, provide high quality forage for livestock, and are also used to control soil erosion. In contrast to numerous fescues, the closely related ryegrass (*Lolium* L.) genus contains only 8 species. These annual, biennial, and perennial grasses are native to Eurasia but expanded to the temperate regions of all continents. Ryegrasses are mostly cultivated as pastures, frequently in mixtures with fescues, and are also grown as lawns around houses and in sport fields. Artificial hybrids between fescues and ryegrasses called *Festulolium* are popular as forage grasses and turfs due to superior agronomical traits. *Festulolium* combines high yield and excellent nutritional value of ryegrasses with the tolerance against abiotic stresses of fescues.

KARGER

E-Mail karger@karger.com
www.karger.com/cgr

Eva Hřibová
Institute of Experimental Botany
Centre of the Region Haná for Biotechnical and Agricultural Research
Šlechtitelů 31, CZ–78371 Olomouc (Czech Republic)
E-Mail hribova@ueb.cas.cz

Meadow fescue (*Festuca pratensis* Huds.) is an outbreeding diploid species (2n = 2x = 14) which has a well-developed root system and efficient water and nutrient capture and is widely used as a pasture crop in northern temperate regions [Ergon et al., 2006]. In lowland coastal regions, meadow fescue is usually replaced by *Festulolium* which has better forage quality but lower tolerance to abiotic stress. The nuclear genome size of meadow fescue is relatively high with 1C ~ 3.175 Gb [Kopecký et al., 2010]. Similarly to other species with large genomes, a high proportion of the meadow fescue genome is represented by different types of repetitive DNA sequences. The ubiquity of repetitive elements in genomes of higher eukaryotes and the fact that they evolve more rapidly than coding sequences make them a unique source of molecular markers to analyze genetic diversity and to study processes of genome evolution and speciation [Koo et al., 2010; Mehrotra and Goyal, 2014; Nybom et al., 2014]. DNA repeats have been used as cytogenetic probes to study the long-range molecular organization of plant genomes [Alkhimova et al., 2004; Cuadrado and Jouve, 2007; Hřibová et al., 2010], to develop molecular karyotypes [Jiang and Gill, 2006; Fradkin et al., 2013; Shibata et al., 2013; Zhang et al., 2013], and to provide comparative analyses of closely related species [Nybom et al., 2014; Dodsworth et al., 2015].

The karyotype of meadow fescue possesses 7 chromosomes collinear to those of Triticeae [Alm et al., 2003]. For several decades, meadow fescue chromosomes were identified based on their morphological attributes such as total length, arm ratio, and presence or absence of a secondary constriction [Malik and Thomas, 1966]. More recently, Thomas et al. [1997] used FISH to localize probes for 5S and 45S ribosomal genes on metaphase chromosomes of various diploid, tetraploid, and hexaploid fescues. In addition to comparing evolutionary relationships among the species, FISH with the 2 probes enabled the identification of 2 meadow fescue chromosomes. The short arm of chromosome 2 was found to bear the 5S rDNA locus, while the 45S rDNA locus was localized on the short arm of chromosome 3.

With the aim to develop additional markers and to discriminate individual chromosomes, Kopecký et al. [2008] screened a panel of 74 BAC clones from a partial BAC library of *F. pratensis*. Out of them, only BAC clone 1G18, when used as a probe in FISH, gave specific hybridization patterns on individual chromosomes. Unfortunately, the discrimination of 3 meadow fescue chromosomes (chromosomes 1, 5, and 6) was not unambiguous in several meadow fescue cultivars. Motivated by the lack of reliable cytogenetic markers for fescues, we decided to analyze repetitive DNA sequences obtained by next-generation sequencing (NGS) of chromosome 4F of *F. pratensis* and utilized them for cytogenetic mapping and karyotyping.

NGS technologies are fast and effective high-throughput methods that enable massive parallel sequencing. Due to their affordable costs, they permit obtaining an enormous amount of short reads from whole genome or chromosome sequences. Recently, NGS was used to provide deeper insights into the genomes of fescues and ryegrasses [Studer et al., 2012; Kopecký et al., 2013; Pfeifer et al., 2013]. In fescues, flow cytometric chromosome sorting [Doležel et al., 2014] was used to purify the largest chromosome 4F, amplify its DNA, and sequence it on an Illumina platform. Using the GenomeZipper approach [Mayer et al., 2011], a comparative analysis of 4F with sequenced genomes of *Brachypodium distachyon*, rice, sorghum, and barley was performed, and a virtual gene map of chromosome 4F was constructed. Apart from genic sequences, repetitive DNA elements were identified and characterized using a graph-based clustering method [Novák et al., 2010]. Different types of mobile genetic elements as well as tandemly organized repeats were mapped by FISH on flow-sorted chromosomes [Kopecký et al., 2013].

In the present work, we took advantage of the availability of the fescue chromosome 4F sequence data and performed a detailed analysis of repetitive DNA sequences. The use of tandemly organized repetitive elements as probes for FISH permitted the identification of individual chromosomes and to establish the molecular karyotype of *F. pratensis* and perform comparative analysis of the karyotype in a set of meadow fescue accessions. Apart from karyotyping, localization of different types of mobile genetic elements on mitotic chromosomes provided new insights into the long-range organization of the meadow fescue genome.

Methods

Plant Material and Genomic DNA Extraction

Six diploid cultivars of *Festuca pratensis* ("Fure," "Dedinovskaya," "Kolumbus," "Merifest," "Pardus," and "Tomosakae") were obtained as seeds from Vladimír Černoch (DLF Seeds, Hladké Životice, Czech Republic). The seeds were germinated, and plants were grown in a greenhouse. Cultivar "Fure" was selected as a model genotype, and the other meadow fescue cultivars representing a wide range of geographical distribution were used for comparative study. For identification of individual chromosomes, we used disomic *Festuca-Lolium* substitution lines developed in our previous work [Kopecký et al., 2008], where 2 homologous chro-

Table 1. Mobile genetic elements used as probes for FISH mapping

Repetitive DNA sequence	Primer name	Primer sequence (5′–3′)	Localization on chromosomes of *F. Pratensis* "Fure"
Ty1/copia			
Angela	FP_CL21Contig31	F: GGAAGCGGTATGGATGAAGA R: CTTTGCCCTAGGGAGAGCTT	dispersed signal over all chromosomes, weaker in centromeric regions
Sire/Maximus	FP_CL24Contig504	F: GTTGGGAAAATTGGAGAGCA R: CAGTTCCCCTAACCCAACAA	dispersed signal over all chromosomes, weaker in centromeric regions
Sire/Maximus	FP_CL120Contig43	F: GTACCCATGCAAAACCCATC R: TCGTCTTTTGCTTGCATTTG	dispersed signal over all chromosomes, weaker in centromeric regions
Ty3/gypsy			
Athila	FP_CL5Contig345	F: AGGGCTCTACATCACCATGC R: ATTCAATGGAGGCATGAACC	dispersed signal over all chromosomes, weaker in centromeric regions
Athila	FP_CL6Contig166	F: TGTGACTCCAAGAGGGGGTA R: GTCCGAACCCCTTCCAGTAT	dispersed signal over all chromosomes, weaker in centromeric regions
Athila	FP_CL40Contig411	F: ACTTGCCCACTTGTTTGAGG R: TGCTGATGGTGAGACCTCATT	dispersed signal over all chromosomes, weaker in centromeric regions
Chromovirideae	FP_CL44Contig43	F: TGAGTGGTGCTGCTGTTTTC R: CACGCAAAACATGGTTCATC	dispersed signal over all chromosomes, weaker in centromeric regions
Chromovirideae	FP_CL189Contig25	F: CATCTCGTTTGACGGAGGTT R: TGCTGTCCTCATGCAAGAAC	dispersed signal over all chromosomes, weaker in centromeric regions
Ogre/Tat	FP_CL7Contig476	F: GGAGTCGAGGGACAAAGAAA R: ACGCCTCTAATTCCCACAGA	dispersed signal over all chromosomes, weaker in centromeric regions
Ogre/Tat	FP_CL9Contig568	F: ACTCTGTGCGAATGCTTCCT R: TCACGCACATAGAGCCACTT	dispersed signal over all chromosomes, weaker in centromeres
Ogre/Tat	FP_CL22Contig298	F: GAGAACCTTGGGTGCAAAAA R: CTCCTGCTCCCAAATCACAT	weaker dispersed signal over all chromosomes, excluding centromeric regions
LTR-putative	FP_CL30Contig994	F: TGGCTGGTGCTAGGTTCTCT R: AGAAGCACCCATGGAATCAG	dispersed signal over all chromosomes, weaker in centromeric regions
LTR-putative	FP_CL38Contig72	F: TTCTGGGTCGGTTATTTTCG R: GATTGATTCCTGTCGCCAAT	strong signals in pericentromeric regions
DNA transposons			
CACTA	FP_CL4Contig141	F: ATGGAACGCAATGGAGAAAG R: CCACACACGTACACCAGGTC	strong signals, preferentially in subtelomeric regions and interstitial part of chromosomes

The primer name contains information about the cluster (CL) in which the repetitive element was identified by Kopecký et al. [2013].

mosomes of *L. multiflorum* are substituted by 2 homologous chromosomes of *F. pratensis* in a tetraploid *L. multiflorum*. Thus, the specific hybridization pattern obtained after FISH with different probes was anchored to individual *F. pratensis* chromosomes.

Genomic DNA of meadow fescues was extracted from lyophilized young leaves, which were grinded using mixer mill MM 301 (Retsch, Haan, Germany). DNA was extracted from 10–15 mg of leaf powder by NucleoSpin® Plant II kit (Macherey-Nagel, Düren, Germany) following the manufacturer's recommendations.

DNA Sequence Data

Illumina sequence data of chromosome 4F of *F. pratensis* cv. "Fure" were obtained in our previous project [Kopecký et al., 2013]. Random data set representing 1× sequence coverage of chromosome 4F (550 Mb) was used to reconstruct repetitive elements by graph-based method of Novák et al. [2010] using Louvain clustering algorithm [Blondel et al., 2008] on RepeatExplorer server [Novák et al., 2012]. The contigs representing different types of repeats used in this study are publicly available on our web site (http://olomouc.ueb.cas.cz/projects/Festuca_pratensis).

For phylogenetic analysis of the CACTA DNA transposon, protein sequences of DNA transposases found in the respected contigs were implemented into the dataset created by Buchmann et al. [2014]. Protein sequences were realigned using MAFFT program v7.029 (–localpair –maxiterate 1000) [Katoh and Toh, 2008]. SeaView v4.2.1. [Gouy et al., 2010] was used for a graphical view of alignment and for phylogenetic analysis using the maximum likelihood model. The resulting tree was drawn and edited using FigTree (http://tree.bio.ed.ac.uk/software/figtree/) program.

DNA Cloning, Sequencing, and Sequence Data Analysis

Individual sequences representing different types of mobile elements were amplified from genomic DNA using PCR with specific primers (Table 1). The PCR reaction mix contained 10 ng

Table 2. Tandemly organized repeats used as probes for FISH mapping

Tandem repeat	Primer name	Primer sequence (5′–3′)	Localization on chromosomes of *F. pratensis* "Fure"	GenBank accession
fpTR1	CL46/Contig120	F: TGCAATAGCTCAGAGCGAAA R: CGCATGTCATCCTCTCAGAA	dispersed signals on all chromosomes	JX624129
fpTR4	CL206/Contig107	F: GCCACACCTGGTTTTCCATA R: GCGAGGGTGAACCAATGTA	strong subtelomeric signal on 4L; additional weak signals on 1S and 7L	JX624130
fpTR5	CL231/Contig50	F: GCCAATGTGTAGACCCCTTC R: GCCTGAACATGGGTTTTAGG	dispersed signals on all chromosomes	JX624131
fpTR6	CL286/Contig19	F: GCACTTCTTGGGCAAATCAAG R: GATGCCAAGGGTAGCAGTGT	strong signal in pericentromeric region of 2L and interstitial part of 5L; additional weak signals on 1L, 3L, 5S, 7S, and 7L	JX624132
fpTR7	CL294/Contig18	F: GGGAAGGGCTAACCAGAGG R: TGGCATTGTGACTTCGTTTG	strong pericentromeric signal on 1L and 5L; additional weak signals on 1L, 2S, 2L, 3L, and 6L	JX624133
fpTR11	CL411/Contig20	F: ATCATCCATCGCGTTGTCTT R: TACCGAGAAGGGTGGCTCAT	dispersed signals on all chromosomes	JX624134
fpTR12	CL414/Contig18	F: GCCCCAGCTATGGATGTCTA R: ATGCTTTCTAAACGGCCTCA	strong telomeric and subtelomeric signal on 3S, 4L, and 6L; polymorphic signal on 4L	JX624135
fpTR15	CL437/Contig2	F: AAAACCCATGGCTTGAAGTG R: CCTTGCCTAGCTACCTCACG	strong pericentromeric signal on 6L; additional weak signals on 1S, 3S, 4S, 4L, 5S, 7S, and 7L	JX624136

Primer name contains information about cluster (CL) in which the repetitive element was identified [Kopecký et al. 2013].

genomic DNA, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, 1 μM primers, 1× PCR buffer, and 2 U/100 of Dynazyme™ II DNA polymerase (Finnzymes, Espoo, Finland). The amplification was performed as follows: 94°C for 5 min (1 cycle), 94°C for 50 s, 60°C for 60 s, 72°C for 50 s (35 cycles), and 72°C for 10 min (1 cycle). PCR products were purified using Agencourt AMPure XP (Beckman Coulter, Brea, Calif., USA), ligated into pCR-XL-TOPO vector, and transformed into One Shot TOP10 electrocompetent *E. coli* (Invitrogen Life Technologies, Carlsbad, Calif., USA). For each repeat type, 16 cloned PCR products were sequenced using the BigDye Terminator v3.1 Cycle Sequencing kit (Applied Biosystems, Foster City, Calif., USA) according to the manufacturer's instructions on an ABI 3730xl DNA analyzer (Applied Biosystems).

Nucleotide sequences were edited using Staden package [Staden, 1996]. Multiple sequence alignments were done using MAFFT program v6.717–1 (–globalpair –maxiterate 1000) [Katoh and Toh, 2008] and graphically displayed in SeaView v4.2.1. [Gouy et al., 2010]. Sequenced clones which showed the highest similarity to the sequence contigs obtained from Illumina sequencing of chromosome 4F were kept at –80°C until used for preparation of FISH probes.

Chromosome Preparations

For cytogenetic analysis, seeds were germinated in a Petri dish with moist filter paper for 5–6 days at room temperature (RT). Actively growing roots were collected into ice water (28 h) and fixed in a 3:1 mixture of absolute ethanol and glacial acetic acid for 1 week at 37°C. After staining in 1% acetocarmine for 2 h at RT, root tips were dissected and squashed in a drop of 45% acetic acid according to Mirzaghaderi [2010]. Coverslips were separated from slides after freezing on a block of dry ice for 1 h, and the slides were immediately submerged in 45% acetic acid at RT. Subsequently, slides were washed in preheated 45% acetic acid for 3 min at 50°C and air dried at RT. The quality of chromosome spreads was evaluated under a Primo Star Microscope (Carl Zeiss MicroImaging, Göttingen, Germany). The best preparations were chosen for in situ hybridization and post-fixed for 10 min in 4% formaldehyde solution and dehydrated in an increasing ethanol series (70, 90, and 100% ethanol, 2 min each). Each slide was prepared from one root dissected from individual seedlings. All cytogenetic analyses were done in at least 10 plants per genotype and 3 metaphases per plant.

Probes for FISH

Probes for tandem repeats (Table 2) identified by Kopecký et al. [2013] as well as for mobile elements (Table 1) were prepared by PCR with specific primers and biotin- or digoxigenin-labeled dUTPs (Roche Applied Science, Penzberg, Germany). PCR was performed in a 25 μL volume and the mix contained 15 ng of meadow fescue genomic DNA, 1 mM PCR buffer, 1.5 mM $MgCl_2$ (ThermoFisher Scientific/Finnzymes, Vantaa, Finland), 0.2 mM biotin- or digoxigenin-labeled dNTPs (ThermoFisher Scientific/Fermentas), 1.0 mM primers, and 1U Dynazyme TM II DNA polymerase (ThermoFisher Scientific/Finnzymes). Reaction conditions were as follows: 5 min at 94°C, followed by 35 cycles of denaturation

(94°C, 50 s), annealing (specific temperature for each primer pair, 50 s), extension (72°C, 50 s), and final extension (72°C, 5 min). The 5S rDNA probe was prepared from meadow fescue DNA using PCR with specific primers RICRGAC1 and RICRGAC2 [Fukui et al., 1994] and the aforementioned labeled nucleotides. The 45S rDNA probe was prepared from clone pTa71 [Gerlach et al., 1979] using a biotin- or digoxigenin-Nick Translation Kit (Roche Applied Science).

Fluorescence in situ Hybridization and Microscopy

The hybridization mix consisting of 40% formamide, 10% dextrane sulphate, 10 µg salmon sperm DNA (Sigma-Aldrich, Schnelldorf, Germany), 1× SSC buffer, and 1–3 µg/mL of each labeled probe was applied onto slides and covered with a glass coverslip. The chromosomes with hybridization mixture were denatured at 80°C for 3 min, and the slides were incubated overnight in a moist chamber at 37°C. Washing steps were performed according to Masoudi-Nejad et al. [2002].

Biotin-labeled and digoxigenin-labeled probes were detected using Cy3-labeled streptavidin (Life Technologies/Invitrogen, Camarillo, Calif., USA) and anti-dig FITC (Roche Applied Science), respectively. The antibodies were applied in concentrations recommended by the manufacturers. Chromosomes were counterstained with 4′,6′-diamidino-2-phenylindole (DAPI) in Vectashield antifade solution (Vector Laboratories, Burlingame, Calif., USA). The preparations were evaluated using an Axio Imager Z.2 Zeiss microscope (Zeiss, Oberkochen, Germany) equipped with a Cool Cube 1 (Metasystems, Altlussheim, Germany) camera and appropriate filters. The capture of fluorescence signals and merging the layers were performed with ISIS software (Metasystems), and the final image adjustment was done in Adobe Photoshop 12.0.

Identification of Individual Chromosomes of F. pratensis

To anchor a specific hybridization signal to individual *F. pratensis* chromosomes, we employed disomic substitution *F. pratensis-L. multiflorum* lines developed previously for all 7 chromosomes of *F. pratensis*. We labeled the genomic DNA of *L. multiflorum* with digoxigenin using dig-Nick translation mix (Roche Applied Science) and used it as a probe in addition to a biotin-labeled probe specific for different tandem repeats. Total genomic DNA of *F. pratensis* was sheared and used as blocking DNA. Probe-to-blocking DNA ratio was 1:150 with minor variations. The conditions for in situ hybridization were the same as described above.

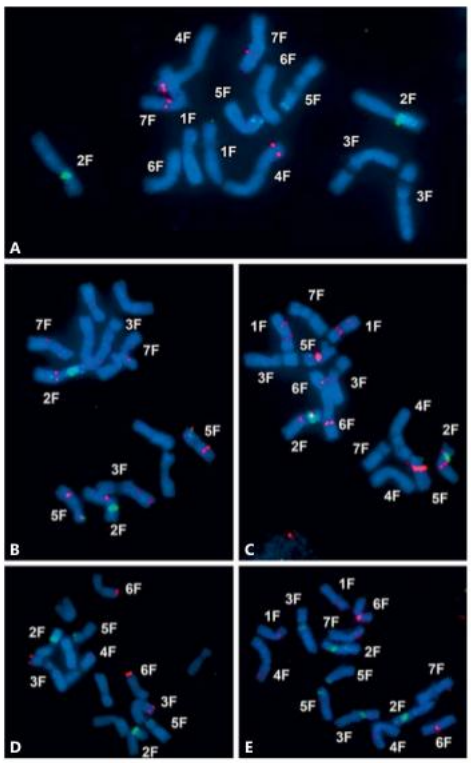


Fig. 1. Chromosome distribution of tandemly organized repeats. Mitotic metaphase spreads of *F. pratensis* cv. "Fure" (2n = 14) after FISH with probes for 5 tandemly organized repeats and rRNA genes. fpTR4 is shown in red and 5S rDNA in green (**A**), fpTR6 in red and 5S rDNA in green (**B**), fpTR7 in red and 5S rDNA in green (**C**), fpTR12 in red and 5S rDNA in green (**D**), and fpTR15 in red and 5S rDNA and 45S rDNA in green (**E**). The chromosomes were counterstained with DAPI (blue).

Results

Karyotyping Using Tandemly Organized Repeats

In order to establish the karyotype of *F. pratensis*, the diploid cultivar "Fure," which was used in our previous study, was selected [Kopecký et al., 2013]. In total, 8 tandemly organized repeats were used as probes for FISH (Table 2; GenBank accession numbers: JX624129–JX624136). Out of them, 3 (fpTR1, fpTR5, and fpTR11) gave dispersed signals on all meadow fescue chromosomes. The 5 remaining tandem repeats (fpTR4, fpTR6, fpTR7, fpTR12, and fpTR15) provided clustered signals on metaphase chromosomes, and their distribution had a chromosome-specific pattern (Fig. 1).

Tandem repeat fpTR4 gave a strong signal in the subtelomeric region on the short arm of chromosome 4 and additional weak signals on the short arm of chromosome 1 and the long arm of chromosome 7. The fpTR6 repeat

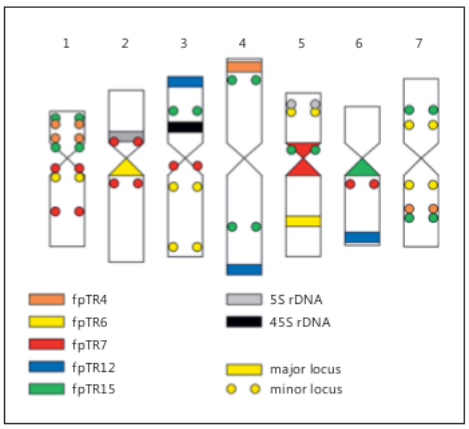


Fig. 2. Idiogram of *F. pratensis* cv. "Fure" displaying the 7 chromosomes of *F. pratensis* with marked locations of 5S rDNA (grey), 45S rDNA (black), and 5 tandem organized repeats: fpTR4 (orange), fpTR6 (yellow), fpTR7 (red), fpTR12 (blue), and fpTR15 (green). The unique combination of hybridization patterns for each chromosome enables the identification of all chromosome in *F. pratensis*.

localized to the pericentromeric region of the long arm of chromosome 2 and the interstitial part of the long arm of chromosome 5. Weaker signals of fpTR6 were obtained on chromosome arms 1L, 3L, 5S, 7S, and 7L. Tandem repeat fpTR7 gave a strong signal in the centromeric and pericentromeric region of chromosome 5 and additional weak signals on chromosome arms 1L, 2S, 2L, 3L, and 6L. Strong signals were observed after FISH with fpTR12 in telomeric and subtelomeric regions of chromosome arms 3S, 4L, and 6L. However, polymorphism was observed in case of the signal on chromosome arm 4L, where some plants had a signal on both homologues, some displayed a signal on only one 4L chromosome arm, and no signal was visible on 4L in some plants. Tandem repeat fpTR15 localized in the pericentromeric region of the long arm of chromosome 6 and also showed signals on chromosome arms 1S, 3S, 4S, 4L, 5S, 7S, and 7L. Apart from the tandem repeats, we used rDNA as probes for FISH. The 45S rDNA locus was localized on the short arm of chromosome 3, whereas the short arm of chromosome 2 carried the 5S rDNA locus. A minor 5S rDNA locus was identified in the distal part of the short arm of chromosome 5. This work resulted in the establishment of the molecular karyotype of *F. pratensis* cv. "Fure" (Fig. 2; Table 2).

In order to analyze intraspecific karyotypic variation, we used FISH with 2 tandemly organized repeats (fpTR7 and fpTR15) on the remaining 5 cultivars of *F. pratensis*. These cultivars were selected based on the wide range of their geographical distribution with the aim to include potentially genetically distant genotypes. Despite this, the observed variability of FISH patterns was marginal. The most variable FISH pattern was obtained for chromosome 1. Tandem repeat fpTR7 gave a signal in the pericentromeric region of the long arm of chromosome 1 in cultivars "Fure" and "Merifest." In cultivar "Columbus," fpTR7 was localized in the interstitial part of the short arm of chromosome 1, and in cultivar "Tomosakae" this tandem repeat localized in the interstitial part of the long arm of chromosome 1. Only a weak signal was obtained in the subtelomeric region of the short arm of chromosome 1 in cultivar "Pardus," and no visible hybridization signal of fpTR7 was detected on chromosome 1 of cultivar "Dedinovskaya." The presence of fpTR15 was found on the short arm of chromosome1 in cultivars "Dedinovsakaya," "Kolumbus," and "Pardus" as well as in the pericentromeric region of the long arm of chromosome 1 in cultivar "Tomosakae" (online suppl. Fig. 1; see www.karger.com/doi/10.1159/000462915 for all online suppl. material). No variation in major loci has been observed among the cultivars, among the plants of the same cultivar, and among the metaphases of the same plant.

Long-Range Organization of Mobile Genetic Elements

Besides the tandem repeats, different types of *Ty1/copia*, *Ty3/gypsy* retroelements and DNA transposons identified in 4F sequence data were selected for detailed characterization. With the aim to confirm their presence in the meadow fescue genome, specific primers were designed for their coding domains (if present) or for the regions with the highest sequencing coverage and were used for PCR amplification with DNA from *F. pratensis* cv. "Fure" as template. PCR products were cloned into the TOPO vector and sequenced using Sanger sequencing. Sequence analysis of a majority of transposable elements resulted in highly similar sequences to those identified by Illumina data of chromosome 4F and suggested a low diversity of their coding regions. Clones which gave the highest homology with contigs assembled from the sequencing data specific for chromosome 4F [Kopecký et al., 2013] were used as probes for FISH.

Fourteen different mobile genetic elements (Table 1) were used as probes and localized on mitotic metaphase chromosomes using FISH. Most of them resulted in signals dispersed over all chromosomes with lower inten-

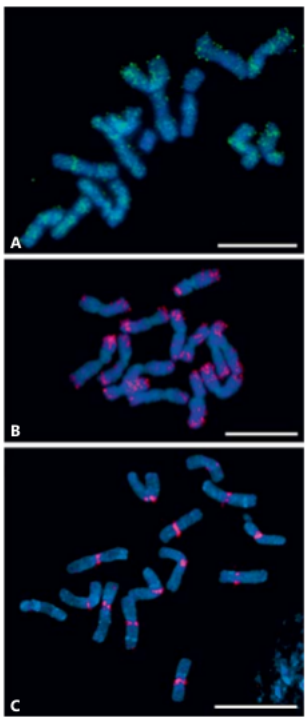


Fig. 3. Chromosome distribution of different classes of transposable elements on mitotic metaphase spreads of *F. pratensis* cv. "Fure" (2n = 14) after FISH. **A** A probe for the *Ty3/gypsy* Athila retrotransposon (CL6) gave a dispersed signal on all chromosomes. **B** The CACTA DNA transposon (CL4) is preferentially localized in (sub)telomeric regions of mitotic chromosomes. **C** The putative LTR transposon (CL38) is localized in centromeric regions. The chromosomes were counterstained with DAPI (blue). Scale bars, 10 µm.

sities in (peri)centromeric and (sub)telomeric regions (Fig. 3). Signal intensities varied among individual types of mobile elements. One of the strongest dispersed signals was observed for the *Athila*-like retroelement (CL6; Table 1) from the *Ty3/gypsy* family suggesting high abundance in the genome. Out of 14 examined probes for mobile genetic elements, only 2 resulted in a specific FISH pattern on metaphase chromosomes. The DNA transposon CL4 (Table 1) gave strong signals in subtelomeric and telomeric regions of all 7 pairs of *F. pratensis* chromosomes. This DNA transposon was identified as a member of the CACTA family. Phylogenetic analysis of the transposase domains found in 5 contigs reconstructed from sequencing data specific for chromosome 4F [Kopecký et al., 2013] revealed that this DNA transposon was a *Caspar*-like element (online suppl. Fig. 2).

Another FISH probe that gave a specific pattern on chromosomes of *Festuca* was prepared from putative LTR retrotransposon CL38 (Table 1) identified in chromosome 4F sequence data [Kopecký et al., 2013]. FISH with a probe for CL38 resulted in strong signals in the centromeric regions of all chromosomes (Fig. 3). Unfortunately, the assembly of Illumina sequence reads representing this element resulted in relatively short contigs (<750 bp) which did not contain any coding domains that would allow a more detailed characterization of this repetitive element.

Discussion

DNA repeats are ubiquitous and occupy the major part of large plant genomes [Neumann et al., 2006; Wicker et al., 2007; D'Hont et al., 2012]. They evolve more rapidly than genic sequences, and mobile genetic elements are considered as main contributors to interspecific divergence [Charlesworth et al., 1994; Piednoël et al., 2012; Novák et al., 2014]. Repetitive DNA sequences represent a good source of molecular markers including cytogenetic landmarks [Paesold et al., 2012; Čížková et al., 2013]. As previously shown, some repeats might be species- or even chromosome-specific and can be used as probes for FISH in karyotyping, providing new insights into the evolutionary origins of species [Hobza et al., 2006; Paux et al., 2006; Dodsworth et al., 2015]. Moreover, when used as probes for FISH, they enable the identification of individual chromosomes. Despite the previous studies which provided information on chromosome pairing, genome constitution of *Festuca-Lolium* hybrids, and characterization of introgression lines [Kopecký et al., 2008, 2010; Akiyama et al., 2012], the karyotype of *Festuca* spp. has not been fully described, and only chromosomes 2 and 3 bearing 5S and 45S loci, respectively, could be unambiguously identified in *F. pratensis*. In the present study, different types of DNA repeats identified in our previous work [Kopecký et al., 2013] were used for karyotyping and analysis of the long-range organization of the genome structure in *F. pratensis* Huds. The combined use of FISH probes for 5 tandem repeats and rDNA

sequences enabled the identification of all 7 chromosomes of meadow fescue.

In situ investigation of mobile genetic elements can provide new insights into chromosome structure and composition. In this work, most of the mobile DNA elements were dispersed along all fescue chromosomes. This agrees with previous reports in other plant species [Wicker et al., 2003; Hřibová et al., 2010; Sergeeva et al., 2010; Akiyarma et al., 2012]. The strongest hybridization signals were obtained for the *Athila* retroelement (CL6) from the *Ty3/gypsy* family, indicating that it is the most abundant DNA repeat in the meadow fescue genome. Only 2 transposable elements provided unique distribution patterns. FISH mapping revealed that *Caspar*-like DNA transposon (CL4) clusters predominantly in subtelomeric regions of all meadow fescue chromosomes. *Caspar*-like elements, members of the CACTA family of DNA transposons, were identified in high copy number in different *Triticeae* genomes including *Hordeum* sp., *Triticum* spp., *Aegilops* spp., and *Secale* sp. [Wicker et al., 2003, 2009; Zhang et al., 2004; Sergeeva et al., 2010]. While several CACTA DNA transposons are dispersed in *Triticeae* chromosomes [Raskina et al., 2004; Altinkut et al., 2006], *Caspar* elements predominantly localize in subtelomeric regions [Sergeeva et al., 2010]. An active utilization of recombination processes for transposition was proposed as the most probable mechanism responsible for the increased abundancy of the *Caspar* element in subtelomeric regions of *Triticeae* genomes [Sergeeva et al., 2010]. In contrast to the *Caspar*-like element, one putative LTR element identified previously in 4F sequence data [Kopecký et al., 2013] preferentially localized to centromeric regions. Unfortunately, the assembly of Illumina reads from chromosome 4F resulted in relatively short contigs which did not contain any coding domains essential for detailed characterization of this repetitive element. Nevertheless, based on the presence of LTR sequences in several contigs, we can speculate that the repeat belongs to the CRM clade of Chromovirideae elements of *Ty3/gypsy* retrotransposons. DNA elements of this clade were found in different plant species, and one group of these elements containing the CR motif is preferentially localized in centromeric regions in different plant species, including monocots as well as eudicots [Nagaki et al., 2005; Bao et al., 2006; Liu et al., 2008; Neumann et al., 2011]. The relatively short assembled contigs, which do not contain coding domains, can be a consequence of sequencing DNA amplified from chromosome 4F using whole genome amplification [Šimková et al., 2008], which is not representative in terms of copy number. Another explanation why the assembly of sequencing data from chromosome 4F [Kopecký et al., 2013] produced only short contigs not containing coding regions can be the non-autonomous character of this LTR element, either on chromosome 4F or even in the whole genome of meadow fescue. Sequencing whole genomic DNA of *F. pratensis* will be needed to resolve this issue.

Cytogenetic mapping indicated that different types of mobile genetic elements are the most abundant types of DNA repeats dispersed throughout the genome, whereas tandemly organized repeats displayed specific patterns of distribution. Nevertheless, out of the 8 putative tandem repeats, 3 gave dispersed signals over all *F. pratensis* chromosomes. This observation may indicate that they evolved from LTR sequences and/or that the (retro)transposition plays a role in their evolution and movement throughout the genome [Macas et al., 2009].

Five tandem repeats, which exhibited a ladder-like pattern after Southern blot hybridization, gave clustered signals typical for satellite DNA sequences after FISH on mitotic chromosomes. Satellite DNA is a typical component of subtelomeric and centromeric chromosome regions in various plant species [Galasso et al., 1995; Cheng et al., 2002; Jiang et al., 2003; Macas et al., 2007; Han et al., 2008]. The lack of tandemly organized DNA repeats specifically localized in centromeric regions of all meadow fescue chromosomes may indicate that the centromeric regions in this species are likely to be composed of various types of retrotransposons.

Although none of the 5 tandem repeats gave a specific signal on a single chromosome, their combined use as probes for FISH together with rDNA sequences permitted the identification of all meadow fescue chromosomes. Each mitotic chromosome of meadow fescue contained at least 1 strong signal of a tandem repeat, except for chromosome 3, which contained only weak signals of fpTR6, 7, 12, and 15. On the other hand, chromosome 3 bears a secondary constriction with the 45S rDNA locus and thus can be identified unambiguously. Comparative FISH analysis of 6 accessions of meadow fescue showed a high level of homogeneity in their karyotypes, suggesting a low level of polymorphism in the distribution of DNA tandem repeats. In our previous study, all tandemly organized repeats were also mapped on flow-sorted chromosomes [Kopecký et al., 2013]. Their chromatin is more decondensed due to a relatively short fixation step during the preparation of the chromosome suspension, and thus, a FISH pattern of tandem repeats on flow sorted chromosomes is not always the same as a FISH pattern on mitotic metaphase spreads. In some cases, flow-sorted chro-

mosomes can display a higher level of resolution after FISH mapping. On the other hand, using a different chemical component during cell cycle synchronization [Kopecký et al., 2013] can inflict changes in chromatin structure followed by a very low accessibility of probes during FISH.

Conclusions

This work reports the cytogenetic mapping of a set of repetitive DNA elements in *F. pratensis* and improves the knowledge on its nuclear genome organization. The Athila element from the *Ty3/gypsy* subclass was found to be the most abundant DNA repeat in the *Festuca* genome and is localized along all chromosomes similar to most of other retroelements from *Ty3/gypsy* and *Ty1/copia* families. Subtelomeric regions were preferentially occupied by a Caspar-like DNA transposon, while a putative LTR element CL38 was preferentially localized in pericentomeric regions of all chromosomes of *F. pratensis*. Apart from transposons, tandemly organized repeats were mapped by FISH to chromosomes in different accessions of *F. pratensis*. Despite the fact that none of the tandem repeats was chromosome- or chromosome-arm specific, their co-localization and use in combination with probes for rDNA sequences permitted the unambiguous identification of all 7 chromosomes of *F. pratensis*. The results obtained will be also useful for the analysis of genome and chromosome organization of related grass species, such as ryegrasses and interspecific hybrids including *Festulolium* (*Festuca* × *Lolium* hybrids).

Acknowledgements

We thank Eva Jahnová, Marie Seifertová, and Radka Tušková for excellent technical assistance. This work has been supported by the National Program of Sustainability I (grant award LO1204). The computing was supported by the National Grid Infrastructure MetaCentrum (grant No. LM2010005 under the programme Projects of Large Infrastructure for Research, Development, and Innovations).

References

Akiyama Y, Kimura K, Yamada-Akiyama H, Kubota A, Takahara Y, et al: Genomic characteristics of a diploid F_4 *Festulolium* hybrid (*Lolium multiflorum* × *Festuca arundinacea*). Genome 55:599–603 (2012).

Alkhimova O, Mazurok N, Potapova T, Zakian S, Heslop-Harrison J, Vershinin A: Diverse patterns of the tandem repeats organization in rye chromosomes. Chromosoma 113:42–52 (2004).

Alm V, Fang C, Busso CS, Devos KM, Grieg Z, Rognli OA: A linkage map of meadow fescue (*Festuca pratensis* Huds.) and comparative mapping with other *Poaceae* species. Theor Appl Genet 108:25–40 (2003).

Altinkut A, Raskina O, Nevo E, Belyayev A: En/Spm-like transposons in *Poaceae* species: transposase sequence variability and chromosomal distribution. Cell Mol Biol Lett 11:214–230 (2006).

Bao W, Zhang W, Yang Q, Zhang Y, Han B, et al: Diversity of centromeric repeats in two closely related wild rice species, *Oryza officinalis* and *Oryza rhizomatis*. Mol Genet Genomics 275:421–430 (2006).

Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E: Fast unfolding of communities in large networks. J Stat Mech 2008:P10008 (2008).

Buchmann JP, Löytynoja A, Wicker T, Schulman AH: Analysis of CACTA transposases reveals intron loss as major factor influencing their exon/intron structure in monocotyledonous and eudicotyledonous hosts. Mob DNA 5:24 (2014).

Charlesworth B, Sniegowski P, Stephan W: The evolutionary dynamics of repetitive DNA in eukaryotes. Nature 371:215–220 (1994).

Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, et al: Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell 14:1691–1704 (2002).

Čížková J, Hřibová E, Humplíková L, Christelová P, Suchánková P, Doležel J: Molecular analysis and genomic organization of major DNA satellites in banana (*Musa* spp.). PLoS One 8:e54808 (2013).

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

D'Hont A, Denoeud F, Aury JM, Baurens FC, Carreel F, et al: The banana (*Musa acuminata*) genome and the evolution of monocotyledonous plants. Nature 488:213–217 (2012).

Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, et al: Genomic repeat abundances contain phylogenetic signal. Syst Biol 64:112–126 (2015).

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

Ergon Å, Fang C, Jørgensen Ø, Aamlid TS, Rognli OA: Quantitative trait loci controlling vernalisation requirement, heading time and number of panicles in meadow fescue (*Festuca pratensis* Huds.). Theor Appl Genet 112:232–242 (2006).

Fradkin M, Ferrari MR, Espert SM, Ferreira V, Grassi E, et al: Differentiation of triticale cultivars through FISH karyotyping of their rye chromosomes. Genome 56:267–272 (2013).

Fukui K, Ohmido N, Klush GS: Variability in rDNA loci in the genus *Oryza* detected through fluorescence in situ hybridization. Theor Appl Genet 87:893–899 (1994).

Galasso I, Schmidt T, Pignone D, Heslop-Harrison JS: The molecular cytogenetics of *Vigna unguiculata* (L.) Walp: the physical organization and characterization of 18s-5.8s-25s rRNA genes, 5s rRNA genes, telomere-like sequences, and a family of centromeric repetitive DNA sequences. Theor Appl Genet 91:928–935 (1995).

Gerlach WL, Bedbrook JR: Cloning and characterization of ribosomal RNA genes from wheat and barley. Nucleic Acids Res 7:1869–1885 (1979).

Gouy M, Guindon S, Gascuel O: SeaView version 4: a multiple graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol 27:221–224 (2010).

Han YH, Zhang ZH, Liu JH, Lu JY, Huang SW, Jin WW: Distribution of the tandem repeat sequences and karyotyping in cucumber (*Cucumis sativus* L.) by fluorescence in situ hybridization. Cytogenet Genome Res 122:80–88 (2008).

Hobza R, Lengerová M, Svoboda J, Kubeková H, Kejnovský E, Vyskot B: An accumulation of tandem DNA repeats on the Y chromosome in *Silene latifolia* during early stages of sex chromosome evolution. Chromosoma 115: 376–382 (2006).

Hřibová E, Neumann P, Matsumoto T, Roux N, Macas J, Doležel J: Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. BMC Plant Biol 10:204 (2010).

Jiang J, Birchler JA, Parrott WA, Kelly Dawe R: A molecular view of plant centromeres. Trends Plant Sci 8:570–575 (2003).

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

Katoh K, Toh H: Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinform 9:286–298 (2008).

Koo D-H, Nam Y-W, Choi D, Bang J-W, de Jong H, Hur Y: Molecular cytogenetic mapping of *Cucumis sativus* and *C. melo* using highly repetitive DNA sequences. Chromosome Res 18:325–336 (2010).

Kopecký D, Lukaszewski AJ, Doležel J: Meiotic behaviour of individual chromosomes of *Festuca pratensis* in tetraploid *Lolium multiflorum*. Chromosome Res 16:987–998 (2008).

Kopecký D, Havránková M, Loureiro J, Castro S, Lukaszewski AJ, et al: Physical distribution of homoeologous recombination in individual chromosomes of *Festuca pratensis* in *Lolium multiflorum*. Cytogenet Genome Res 129: 162–172 (2010).

Kopecký D, Martis M, Čihalíková J, Hřibová E, Vrána J, et al: Flow sorting and sequencing meadow fescue chromosome 4F. Plant Physiol 163:1323–1337 (2013).

Liu Z, Yue W, Li D, Wang RR-C, Kong X, et al: Structure and dynamics of retrotransposons at wheat centromeres and pericentromeres. Chromosoma 117:445–456 (2008).

Macas J, Neumann P, Navrátilová A: Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. BMC Genomics 8:427 (2007).

Macas J, Koblížková A, Navrátilová A, Neumann P: Hypervariable 3′ UTR region of plant LTR-retrotransposons as a source of novel satellite repeats. Gene 448:198–206 (2009).

Malik CP, Thomas PT: Karyotypic studies in some *Lolium* and *Festuca* species. Caryologia 19:167–196 (1966).

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

Mayer KFX, Martis M, Hedley PE, Šimková H, Liu H, et al: Unlocking the barley genome by chromosomal and comparative genomics. Plant Cell 23:1249–1263 (2011).

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

Mirzaghaderi G: A simple metaphase chromosome preparation from meristematic root tip cells of wheat for karyotyping or in situ hybridization. Afr J Biotechnol 9:314–318 (2010).

Nagaki K, Neumann P, Zhang D, Ouyang S, Buell CR, et al: Structure, divergence, and distribution of the CRR centromeric retrotransposon family in rice. Mol Biol Evol 22:845–855 (2005).

Neumann P, Koblížková A, Navrátilová A, Macas J: Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement. Genetics 173:1047–1056 (2006).

Neumann P, Navrátilová A, Koblížková A, Kejnovský E, Hřibová E, et al: Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mob DNA 2:4 (2011).

Novák P, Neumann P, Macas J: Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics 11:378 (2010).

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J: RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. BMC Genomics 13:722 (2012).

Novák P, Hřibová E, Neumann P, Koblížková A, Doležel J, Macas J: Genome-wide analysis of repeat diversity across the family *Musaceae*. PLoS One 9:e98918 (2014).

Nybom H, Weising K, Rotter B: DNA fingerprinting in botany: past, present, future. Investig Genet 5:1 (2014).

Paesold S, Borchardt D, Schmidt T, Dechyeva D: A sugar beet (*Beta vulgaris* L.) reference FISH karyotype for chromosome and chromosome-arm identification, integration of genetic linkage groups and analysis of major repeat family distribution: Sugar beet FISH karyotype. Plant J 72:600–611 (2012).

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

Pfeifer M, Martis M, Asp T, Mayer KFX, Lübberstedt T, et al: The perennial ryegrass GenomeZipper: targeted use of genome resources for comparative grass genomics. Plant Physiol 161:571–582 (2013).

Piednoël M, Aberer AJ, Schneeweiss GM, Macas J, Novák P, et al: Next-generation sequencing reveals the impact of repetitive DNA across phylogenetically closely related genomes of *Orobanchaceae*. Mol Biol Evol 29:3601–3611 (2012).

Raskina O, Belyayev A, Nevo E: Activity of the En/Spm-like transposons in meiosis as a base for chromosome repatterning in a small, isolated, peripheral population of *Aegilops speltoides* Tausch. Chromosome Res 12:153–161 (2004).

Sergeeva EM, Salina EA, Adonina IG, Chalhoub B: Evolutionary analysis of the CACTA DNA-transposon Caspar across wheat species using sequence comparison and in situ hybridization. Mol Genet Genomics 284:11–23 (2010).

Shibata F, Nagaki K, Yokota E, Murata M: Tobacco karyotyping by accurate centromere identification and novel repetitive DNA localization. Chromosome Res 21:375–381 (2013).

Šimková H, Svensson JT, Condamine P, Hřibová E, Suchánková P, et al: Coupling amplified DNA from flow-sorted chromosomes to high-density SNP mapping in barley. BMC Genomics 9:294 (2008).

Staden R: The Staden sequence analysis package. Mol Biotechnol 5:233–241 (1996).

Studer B, Byrne S, Nielsen RO, Panitz F, Bendixen C, et al: A transcriptome map of perennial ryegrass (*Lolium perenne* L.). BMC Genomics 13:140 (2012).

Thomas HM, Harper JA, Meredith MR, Morgan WG, King IP: Physical mapping of ribosomal DNA sites in *Festuca arundinacea* and related species by in situ hybridization. Genome 40: 406–410 (1997).

Wicker T, Guyot R, Yahiaoui N, Keller B: CACTA transposons in *Triticeae*. A diverse family of high-copy repetitive elements. Plant Physiol 132:52–63 (2003).

Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, et al: A unified classification system for eukaryotic transposable elements. Nat Rev Genet 8:973–982 (2007).

Wicker T, Taudien S, Houben A, Keller B, Graner A, et al: A whole-genome snapshot of 454 sequences exposes the composition of the barley genome and provides evidence for parallel evolution of genome size in wheat and barley. Plant J 59:712–722 (2009).

Zhang H, Bian Y, Gou X, Dong Y, Rustgi S, et al: Intrinsic karyotype stability and gene copy number variations may have laid the foundation for tetraploid wheat formation. Proc Natl Acad Sci USA 110:19466–19471 (2013).

Zhang P, Li W, Fellers J, Friebe B, Gill BS: BAC-FISH in wheat identifies chromosome landmarks consisting of different types of transposable elements. Chromosoma 112:288–299 (2004).

4.1.4 ORTHOLOGY GUIDED TRANSCRIPTOME ASSEMBLY OF ITALIAN RYEGRASS AND MEADOW FESCUE FOR SINGLE NUCLEOTIDE POLYMORPHISMS DISCOVERY

Štěpán Stočes[1], Tom Ruttink[2], Jan Bartoš[1], Bruno Studer[3], Steven Yates[3], Zbigniew Zwierzykowski[4], Michael Abrouk[1], Isabel Roldán-Ruiz[2], Tomasz Książczyk[3], Elodie Rey[1], Jaroslav Doležel[1] and David Kopecký[1]

The Plant Genome Volume 9, doi: 10.3835/plantgenome2016.02.0017.

Published August 4, 2016

ORIGINAL RESEARCH

Orthology Guided Transcriptome Assembly of Italian Ryegrass and Meadow Fescue for Single-Nucleotide Polymorphism Discovery

Štěpán Stočes, Tom Ruttink, Jan Bartoš, Bruno Studer, Steven Yates, Zbigniew Zwierzykowski, Michael Abrouk, Isabel Roldán-Ruiz, Tomasz Książczyk, Elodie Rey, Jaroslav Doležel, and David Kopecký*

Abstract

Single-nucleotide polymorphisms (SNPs) represent natural DNA sequence variation. They can be used for various applications including the construction of high-density genetic maps, analysis of genetic variability, genome-wide association studies, and map-based cloning. Here we report on transcriptome sequencing in the two forage grasses, meadow fescue (*Festuca pratensis* Huds.) and Italian ryegrass (*Lolium multiflorum* Lam.), and identification of various classes of SNPs. Using the Orthology Guided Assembly (OGA) strategy, we assembled and annotated a total of 18,952 and 19,036 transcripts for Italian ryegrass and meadow fescue, respectively. In addition, we used transcriptome sequence data of perennial ryegrass (*L. perenne* L.) from a previous study to identify 16,613 transcripts shared across all three species. Large numbers of intraspecific SNPs were identified in all three species: 248,000 in meadow fescue, 715,000 in Italian ryegrass, and 529,000 in perennial ryegrass. Moreover, we identified almost 25,000 interspecific SNPs located in 5343 genes that can distinguish meadow fescue from Italian ryegrass and 15,000 SNPs located in 3976 genes that discriminate meadow fescue from both *Lolium* species. All identified SNPs were positioned in silico on the seven linkage groups (LGs) of *L. perenne* using the GenomeZipper approach. With the identification and positioning of interspecific SNPs, our study provides a valuable resource for the grass research and breeding community and will enable detailed characterization of genomic composition and gene expression analysis in prospective *Festuca* × *Lolium* hybrids.

Published in Plant Genome
Volume 9. doi: 10.3835/plantgenome2016.02.0017

Core Ideas

- Transcriptomes of *F. pratensis* and *L. multiflorum* were sequenced and assembled
- We present a catalogue of SNPs for ancestry analysis and future breeding of grasses
- We defined interspecific SNPs to study parental-genome specific gene expression in hybrids
- We positioned SNPs on linkage groups for high-resolution genome constitution analysis

GRASSES of the genera *Festuca* and *Lolium* belong to the Pooideae subfamily within the grass family (Poaceae) and are widely cultivated in temperate regions. They include species such as perennial ryegrass, Italian ryegrass, meadow fescue, and tall fescue (*F. arundinacea* Schreb.). Perennial ryegrass is a highly tillering species

Š. Stočes, J. Bartoš, M. Abrouk, E. Rey, J. Doležel, and D. Kopecký, Institute of Experimental Botany, Centre of the Region Haná for Biotechnological and Agricultural Research, Šlechtitelů 31, 78371 Olomouc, Czech Republic; T. Ruttink and I. Roldán-Ruiz, Institute for Agricultural and Fisheries Research (ILVO), Plant Sciences Unit–Growth and Development, Caritasstraat 39, 9090 Melle, Belgium; B. Studer, and S. Yates, Institute of Agricultural Sciences, Forage Crop Genetics, ETH Zurich, Universitätstrasse 2, 8092 Zurich, Switzerland; Z. Zwierzykowski and T. Książczyk, Dep. of Environmental Stress Biology, Institute of Plant Genetics of the Polish Academy of Sciences, Strzeszyńska 34, 60479 Poznań, Poland. Received 16 Feb. 2016. Accepted 6 June 2016. *Corresponding author (kopecky@ueb.cas.cz).

Abbreviations: AD, minimal allele count score; CDS, coding DNA sequence; DBG, De Bruijn graph; DP, total read depth; GQ, genotype quality score; HQ, high quality; HRM, high-resolution melting curve analysis; LG, linkage group; OGA, Orthology Guided Assembly; PCR, polymerase chain reaction; PGAS, parental genome allele specific; QTL, quantitative trait loci; RNA-seq, RNA sequencing; SNP, single-nucleotide polymorphism.

suitable for permanent pastures and meadows and is a dominant species for turf applications in Europe. Italian ryegrass is used for hay and silage production in temporary grasslands. Both ryegrasses are highly nutritious and high seed yielding (reviewed in Humphreys et al., 2010). However, they are sensitive to abiotic stress conditions. Meadow fescue is widely grown as a pasture crop in Northern temperate regions (Ergon et al., 2006) because of its good persistence and high tolerance to environmental stresses such as cold, freezing, and drought. Tall fescue is an important cool-season perennial forage and turf grass species with excellent drought tolerance (Rognli et al., 2010).

Intergeneric hybrids between *Festuca* and *Lolium* species, referred to as Festuloliums, combine the beneficial agronomic traits from both genera (Ghesquiere et al., 2010). During the last decades, Festuloliums have become a valuable source of new grasses for cultivation (Yamada et al., 2005). They are highly productive, nutritious, and resilient and, as such, are widely used for agricultural and amenity purposes (reviewed in Ghesquiere et al., 2010). Festuloliums exhibit significant diversity for many interesting traits, mainly as a result of their variable genome constitution, which differs markedly even among plants of the same cultivar (Kopecký et al., 2011). Identification of parental chromosomes and characterization of genomic constitution of these interspecific hybrids is typically done using cytogenetic methods such as fluorescence and genomic in situ hybridization (FISH and GISH) (Zwierzykowski et al., 1998a,b; Canter et al., 1999; Kopecký et al., 2005, 2006; Książczyk et al. 2015). However, these methods only allow the detection of large chromosomal segments and are not applicable for high-throughput screening (Kopecký et al., 2006).

In contrast to cytogenetic approaches, DNA sequence polymorphisms offer the opportunity to determine genomic constitution of Festuloliums at much higher resolution. Various DNA marker types distinguishing *Lolium* and *Festuca* genomes have been reported such as microsatellites (Pašakinskiene et al., 2000; Studer et al., 2006) and Diversity Array Technology markers (Kopecký et al., 2009). However, SNPs are the most prevalent type of polymorphisms and thus informative on account of their relatively even dispersal across the genome and their high density especially in allogamous species such as fescues and ryegrasses (Birrer et al., 2014; Wang et al., 2014; Czaban et al., 2015). Single-nucleotide polymorphisms within one species (here referred to as intraspecific SNPs) can be used for quantitative trait loci (QTL) mapping, map-based cloning, genome-wide association studies, and marker-assisted or genomic selection. Single-nucleotide polymorphisms distinguishing the progenitor species (here referred to as interspecific SNPs) can be used to study genomic constitution of interspecific hybrids such as Festuloliums or hybrid ryegrasses with a high resolution. Moreover, if identified in transcribed genome regions, interspecific SNPs can also be used to profile parental genome allele specific (PGAS) expression levels in subsequent generations of interspecific hybrids. Thus, the underlying causes of the superior phenotype could be better understood and perhaps used to improve *Festuca*, *Lolium*, and Festulolium cultivars.

RNA sequencing (RNA-seq, Mortazavi et al., 2008) is suitable to identify SNPs in the gene space of a given species, particularly in the absence of a complete genome sequence (Denoeud et al., 2008; Varshney et al., 2009; Garvin et al., 2010). While a draft set of scaffolds of the *L. perenne* genome has recently been published (Byrne et al., 2015), there is no high-quality genome sequence available, which could serve as a reference for read mapping and identification of polymorphisms in Italian ryegrass and meadow fescue, the two species used in this study.

Considering the available genomic resources, we set out to create a reference transcriptome for Italian ryegrass and meadow fescue based on RNA-seq on a small panel of genotypes. Illumina sequencing technology has been widely used for transcriptome sequencing because of its high throughput and low error rate (Gonçalves da Silva et al., 2014; Peng et al., 2014; Wu et al., 2014). This technology yields short reads (up to 300 bp), which can be used to de novo reconstruct the transcriptome. Several algorithms exist for de novo assembly of short Illumina reads based on De Bruijn graphs (DBG), including Trans-ABySS (Simpson et al., 2009; Robertson et al., 2010), Trinity (Grabherr et al., 2011), Velvet (Zerbino and Birney 2008), Oases (Schulz et al., 2012), and CLCBio Genomics Workbench (CLC Bio-Qiagen). However, a high degree of heterozygosity, typical for outcrossing species such as fescues and ryegrasses (Modrek and Lee 2002; Ruttink et al., 2013; Farrell et al., 2014), can result in transcript fragmentation and allelic redundancy in contigs that are assembled using DBG algorithms. This problem can be resolved by a second round of clustering with overlap–layout–consensus assemblers such as CAP3 (Huang and Madan, 1999). A quick and targeted strategy to perform such secondary clustering on a gene-by-gene basis for an entire transcriptome is known as Orthology Guided Assembly (OGA; Ruttink et al., 2013). The OGA strategy uses the proteome of a closely related species to guide the clustering procedure and to generate a nonredundant and annotated consensus sequence per gene (Ruttink et al., 2013).

The purpose of the experiments presented here was to identify a large and robust set of SNP markers useful to determine genomic constitution of Italian ryegrass × meadow fescue hybrids. By focusing on the transcriptome, we circumvented the lack of a reference genome sequence from the two species. The newly developed set of SNPs will also be useful to quantify PGAS expression levels in interspecific hybrids. In this paper we report on (i) the reconstruction of reference transcriptomes for meadow fescue and Italian ryegrass and their comparison with the previously published perennial ryegrass transcriptome assembled with the same OGA strategy; (ii) the identification of transcript-anchored intraspecific and interspecific SNPs in the three species; (iii) the identification of SNPs that discriminate parental genome

specific alleles in pairwise crosses between meadow fescue and Italian ryegrass; and (iv) the identification of putative genomic locations of the SNPs, based on the perennial ryegrass GenomeZipper (Pfeifer et al., 2013). This large-scale marker set we have obtained constitutes an important resource to advance genomic research in meadow fescue and Italian ryegrass, two of the most important forage grass species, and their hybrids.

Materials and Methods

Plant Material and RNA Extraction

Six genotypes of Italian ryegrass (Lm) were used for transcriptome sequencing: two genotypes of the tetraploid cultivar 'Mitos' (Lm51 and Lm52) and one genotype of each of the diploid Italian ryegrass cultivars 'Abercomo' (Lm53), 'Fox' (Lm54), 'Prolog' (Lm55) and 'Sikem' (Lm56). Six meadow fescue (Fp) genotypes were used for transcriptome sequencing: two genotypes of the tetraploid cultivar 'Westa' (Fp52 and Fp53), one genotype of the tetraploid cultivar 'Patra' (Fp54), and one genotype of each of the diploid cultivars 'Skawa' (Fp51), 'Fure' (Fp55), and 'WSC' (Fp56). The plants were grown in a growth chamber (Weiss-Gallenkamp) at 50% relative humidity (day, 14 h of 20,000 lux light intensity at 24°C; night, 10 h at 20°C). After establishment, the plants were cut to a height of ~5 cm above ground level to stimulate tillering. After 3 wk of regrowth, 100 mg of young leaves were harvested and frozen in liquid nitrogen. Three independent samples were collected from the genotypes Lm51, Lm52, Fp52, and Fp53 and one sample from each of the remaining genotypes. Total RNA was extracted from ground samples using the RNeasy Plant Mini Kit (Qiagen, Inc.). The quality, quantity, and integrity of RNA were checked using an RNA Pico 6000 chip on a Bioanalyzer 2100 (Agilent Technologies, Inc.). Samples with RNA integrity number >7 were used for RNA sequencing.

RNA Sequencing Library Construction and Illumina Sequencing

High-throughput sequencing of Italian ryegrass and meadow fescue libraries was performed at Istituto di Genomica Applicata (IGA Technology Services, Udine, Italy). RNA sequencing libraries were generated using the TruSeq RNA Sample Prep kit v2 according to the manufacturer's protocol (Illumina Inc.). Adapters were ligated to the complementary DNA and 200 ± 25 bp fragments were gel purified and amplified by polymerase chain reaction (PCR). DNA libraries were quantified using a Bioanalyzer 2100 (Agilent Technologies), pooled in equimolar amounts for sequencing using Illumina HiSeq2000 or HiSeq2500 instruments to produce paired-end 101 bp reads.

De Novo Transcriptome Assembly

The FASTX-Toolkit (http://hannonlab.cshl.edu/fastx_toolkit/index.html) was used to trim the 5-bp index and to discard reads with a Phred quality score <25. Reads of each sample were then used for independent de novo assembly using CLCBio Genomics Workbench v8.4. Only contigs longer than 200 bp were retained. Scaffolding was switched off to avoid long stretches of Ns in the contigs, which would interfere with the construction of a consensus sequence by CAP3 during downstream OGA assembly. The total number of reads, number of contigs, and N50 length are listed in Supplemental Table S1.

Orthology Guided Assembly

To resolve transcript fragmentation and allelic redundancy, an OGA strategy was performed combining the contigs from each of the 10 independent de novo assemblies per species. Briefly, we used 26,632 *Brachypodium distachyon* (L.) Beauv. proteins (The International Brachypodium Initiative, 2010) as a reference to assemble consensus transcript sequence as described previously (Ruttink et al., 2013). Transcripts matching particular *B. distachyon* proteins were assembled using CAP3. We then trimmed the transcripts to the coding DNA sequence (CDS), determined the remaining redundancy per *B. distachyon* protein, selected the longest representative sequence per *B. distachyon* protein (minimal CDS length of 102 bp), and thus created novel, nonredundant reference transcriptome assemblies for meadow fescue and Italian ryegrass. All transcripts were assigned to orthologous protein families of the PLAZA 3.0 Monocots database (Proost et al., 2014) as an entry point for full comparative genomics and functional annotation. The previously assembled reference transcriptome from perennial ryegrass (Ruttink et al., 2013) was included in the subsequent steps to compare the effect of using a reference transcriptome from a closely related species on read mapping and SNP calling.

For more accurate and cross-species analyses, we also identified near full-length transcripts (length of the transcript is between 85 and 115% of the respective orthologous *B. distachyon* CDS). An overview per orthologous *B. distachyon* protein is available in Supplemental Table S2. The OGA assembled transcriptome sequences are available as FASTA file and GFF with annotations in Zenodo (http://dx.doi.org/10.5281/zenodo.55304).

Read Mapping and Single-Nucleotide Polymorphism Calling

Here, we describe the analysis procedure for a single reference transcriptome. However, all analyses were performed separately on the two other reference transcriptomes in parallel. Quality trimmed RNA-seq reads of all genotypes were aligned against one of the three nonredundant reference transcriptomes using BWA (Li and Durbin, 2009). After local realignment with GATK IndelRealigner, duplicated reads were removed with MarkDuplicates. In parallel to the new reads obtained from meadow fescue and Italian ryegrass, the previously published RNA-seq reads of 14 genotypes of perennial rygrass (Supplemental Table S1) were mapped onto the reference and all BAM files were used for combined SNP analyses

across the three species. The SNPs were called using the UnifiedGenotypeCaller of GATK (McKenna et al., 2010; DePristo et al., 2011). The identified genotype calls per sample were further filtered using custom Python scripts. Only biallelic SNPs with genotype quality score GQ $\geq$ 30 and minimal allele count AD $\geq$ 4 were retained. We used different total read depth for tetraploid (DP $\geq$ 16) and for diploid (DP $\geq$ 10) genotypes as a trade-off between assuring a high probability of a correct genotype call and retrieving enough data for a similar number of genes given varying total library size per sample. For some perennial ryegrass samples, the library size was much lower than the average library size of the novel obtained data from meadow fescue and Italian ryegrass. Although missing data in some of the 14 perennial ryegrass genotypes may lead to some underestimation of the number of SNPs (see Fig. 4C of Ruttink et al., 2013), we required SNP call data in at least 10 out of 14 perennial ryegrass genotypes during interspecific SNP identification. With these settings, over 7000 genes exceeded the threshold number of genotypes required to call interspecific SNP in each of the species (six in Fp, six in Lm, and 10 in Lp).

For each of the tetraploid genotypes Fp52, Fp53, Lm51, and Lm52, we tested the robustness of SNP calling by comparing the genotype calls across their three biological replicates. Genotype calls present in at least two of three replicates were compared after quality filtering (GQ $\geq$ 30, DP $\geq$ 16, AD $\geq$ 4). If all genotype calls per SNP were consistent (two identical calls and one missing call or three identical calls), these were merged into a single genotype call per sample and used for all downstream analyses. Otherwise, the genotype call for that SNP was labeled as missing data.

Classification of Intraspecific and Interspecific Single-Nucleotide Polymorphisms

All analyses were performed in parallel on the three nonredundant reference transcriptomes of meadow fescue, Italian ryegrass, or perennial ryegrass, respectively. For each reference transcriptome, SNP calls from the different meadow fescue, Italian ryegrass, or perennial ryegrass genotypes were first classified as homozygous reference allele, heterozygous, homozygous alternative allele, or missing data. Next, for each SNP, the following criteria were used to classify it as one of following three classes: (i) intraspecific SNPs (INTRA), (ii) interspecific SNPs in two-way comparison (INTER-2W), and (iii) interspecific SNPs in three-way comparison (INTER-3W) (See Fig. 1 for the classification scheme).

(i) If two or more alleles of a given SNP were detected in any number of genotypes within a species (i.e., at least one heterozygous genotype or at least one genotype each for the homozygous reference allele and the homozygous alternative allele), the SNP position was classified as an intraspecific SNP and indexed for that species (e.g., $INTRA_{Fp}$).

(ii) If a SNP position was monomorphic in all genotypes of one species and carried monomorphic alternative allele in all genotypes of other species, the SNP position was called as interspecific SNP of type INTER-2W for two-way species comparison (e.g., $INTER\text{-}2W_{Fp\text{-}Lm}$). Such interspecific SNPs were identified for all three combinations of two-way species comparison (Fp–Lm, Fp–Lp, and Lm–Lp). In case of meadow fescue and Italian ryegrass, all six genotypes had to have a SNP call. In case of perennial ryegrass, some of the 14 genotypes had low read depth (see Supplemental Table S1) resulting in a relatively low number of SNPs with calls in all 14 genotypes. Thus, we requested at least 10 out of 14 genotypes to have SNP calls.

(iii) If all six Fp genotypes and all six Lm genotypes or at least 10 out of 14 Lp genotypes were monomorphic in one species and carried the monomorphic alternative allele in the two other species, the SNP position was called interspecific SNP of type INTER-3W for three-way species comparison (e.g., $INTER\text{-}3W_{Fp}$ is different in Fp vs. both Lm and Lp; see Fig. 1B).

Classification of Interspecific Single-Nucleotide Polymorphisms in Pairwise Comparisons

Pairwise comparisons were made between tetraploid genotypes of meadow fescue and tetraploid genotypes of Italian ryegrass, to simulate biparental hybrid crosses. If the meadow fescue genotype contained a homozygous reference allele, and the Italian ryegrass genotype contained the homozygous alternative allele (or vice versa), the SNP was called interspecific SNP of INTER-PW type for pairwise comparison (e.g., $INTER\text{-}PW_{Fp52\text{-}Lm51}$). Such pairwise comparisons were made for all six combinations of three tetraploid meadow fescue genotypes (Fp52, Fp53, and Fp54) and two tetraploid Italian ryegrass genotypes (Lm51 and Lm52). For each pairwise combination, a unique HQ gene set (which is different from the near full-length gene set described above) was selected after three filtering steps (see Fig. 1C). Each OGA-assembled transcript in the separate transcriptomes is annotated to the central *B. distachyon* gene set, thus delineating orthologs across the three transcriptomes. Therefore, the effect of the reference bias can be analyzed on a gene-per-gene level. For Filter 1, we used only 8094 near full-length transcripts (85–115% of the *B. distachyon* CDS length in both species) shared between the meadow fescue and Italian ryegrass references. For Filter 2, to avoid read mapping bias, we retained only transcripts for which the ratio of average read depth for meadow fescue reference transcript and Italian ryegrass reference transcript varied between 0.5 and 2 for both meadow fescue and Italian ryegrass reads mapped on both respective transcriptomes. For Filter 3, we retained only those genes where the difference in the number of SNPs identified in meadow fescue vs. Italian ryegrass orthologous reference transcript was $\leq$2 or <20% of the average. Together, these three filters strongly select for transcripts with little or no reference sequence specific bias and allow studying the PGAS expression in interspecific hybrids using any of the two reference transcriptomes for read mapping. For each

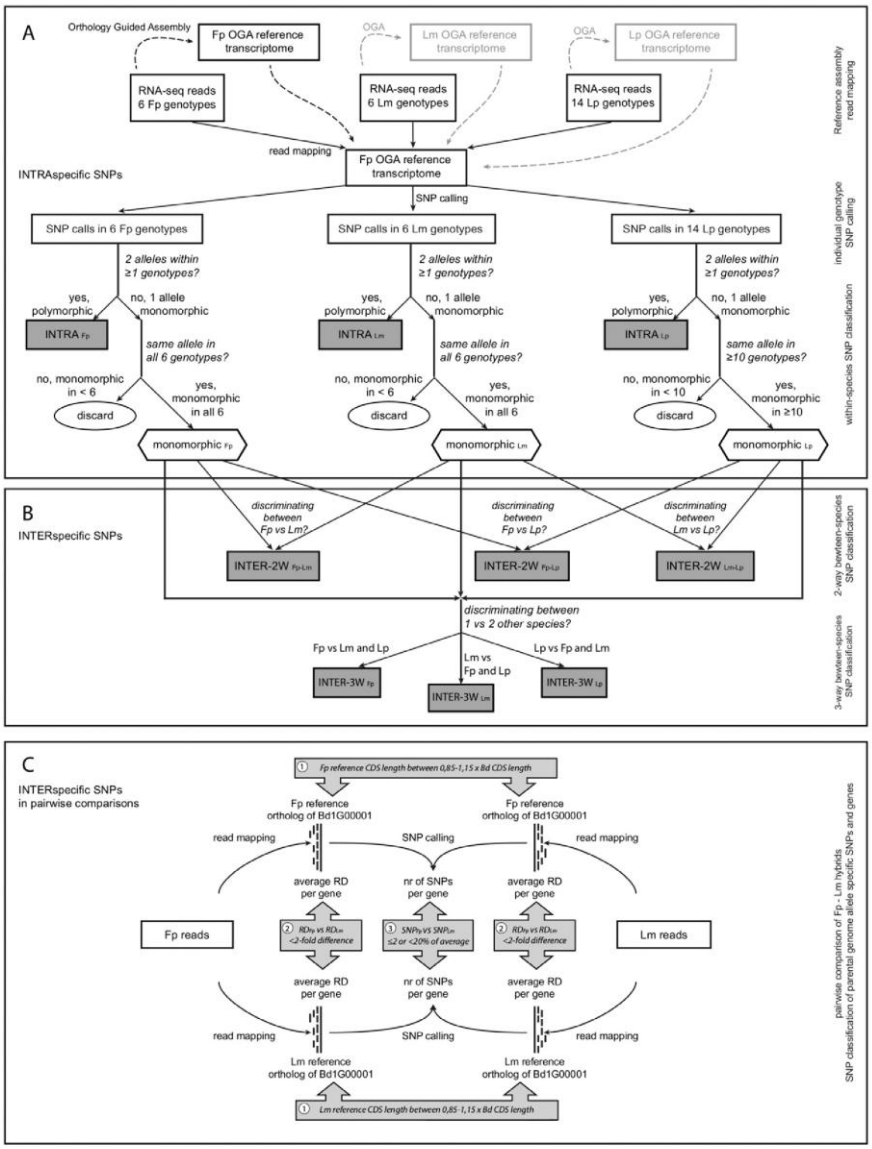


Fig. 1. Classification of intraspecific and interspecific single-nucleotide polymorphisms (SNPs). Selection criteria are indicated in italics. (A) De novo assembly and Orthology Guided Assembly generate a reference transcriptome for each of the three species. Intraspecific SNPs (INTRA) are identified by mapping all RNA-seq data from *Festuca pratensis*, *Lolium multiflorum*, and *L. perenne* onto a single reference transcriptome. This analysis is performed in parallel for each reference transcriptome. (B) Interspecific SNPs are identified between two species (INTER-2W) or by comparing all three species (INTER-3W). (C) Filtering scheme for the selection of high-quality reference transcripts in pairwise genotype comparisons (INTER-PW).

pairwise comparison, a specific HQ set of transcripts was defined because these depend on the specific read depth (Filter 2) and SNP sets (Filter 3) for a given combination of parental genotypes (see Fig. 1C).

SNPhylo

A phylogenetic tree comparing SNP profiles of the 26 genotypes using the meadow fescue reference transcriptome was calculated in SNPhylo (v.20140430; Lee et al., 2014) using 100 bootstraps. Only SNPs without missing data in any of the 34 samples (including RNA-Seq replicates) and with minor allele frequency >10% (default settings) were used over the whole set of genotypes. The SNP sets identified with the three different reference transcriptomes gave nearly identical results (data not shown).

Positioning of Genes using the Perennial Ryegrass GenomeZipper

We positioned orthologous genes of meadow fescue, Italian ryegrass, and perennial ryegrass containing various classes of SNPs on seven LGs of perennial ryegrass using the GenomeZipper approach (Mayer et al., 2009). This approach is based on the collinearity of the gene order shared between a model species (e.g., *B. distachyon*) and the three species of our interest. The genetic map of perennial ryegrass (Studer et al., 2012) was used as a backbone to determine syntenic chromosome blocks between *B. distachyon* and perennial ryegrass. A similar virtual gene order was previously published for perennial ryegrass (Pfeifer et al., 2013). We placed additional genes onto the existing virtual gene order of perennial ryegrass based on orthologous relationships with *B. distachyon* genome using in-house Perl scripts. Previously unanchored meadow fescue, Italian ryegrass, and perennial ryegrass transcripts were positioned between orthologs of neighboring *B. distachyon* genes with a known position in the GenomeZipper.

We used Circos (Krzywinski et al., 2009) for circular visualization of orthologous blocks of *B. distachyon* chromosomes with perennial ryegrass LGs, distribution of all Italian ryegrass transcripts, distribution of near full-length transcripts assembled in all three species, the number of intraspecific SNPs per gene $INTRA_{Fp}$, $INTRA_{Lm}$, $INTRA_{Lp}$, and two examples of interspecific SNPs in pairwise comparisons of potential Italian ryegrass × meadow fescue hybrids.

Validation of Single-Nucleotide Polymorphism by High-Resolution Melting Curve Analysis

High-resolution melting curve analysis (HRM) was employed to validate the SNPs identified in silico. Ten SNPs (from 10 genes) discriminating the alleles of meadow fescue and Italian ryegrass were selected. Primers for HRM analysis were designed to amplify 43 to 67 bp DNA fragments with a melting temperature of 60°C. High-resolution melting curve analysis–PCR was performed as described by Studer et al. (2009) in a total volume of 10 μL using 1× LightScanner High Sensitivity Genotyping MasterMix containing LCGreen PLUS (Idaho Technology, Inc.). Each reaction contained 20 ng of genomic DNA and 0.10 μM of forward and reverse primers and was covered with 15 μL mineral oil to avoid sample evaporation during PCR and the melting process. The PCR amplification was conducted in a Thermoblock 96 Cycler (SensoQuest) under the following conditions: initial denaturation for 2 min at 95°C, 40 cycles of denaturation for 30 s at 94°C, annealing for 30 s at T_a (Supplemental Table S3), and elongation for 30 s at 72°C. The melting curve analysis was performed at a temperature range from 60 to 95°C in steps of 0.05°C using a LightScanner Instrument and the LightScanner and Call-IT software modules (Idaho Technology).

Table 1. Number of intraspecific single-nucleotide polymorphisms (italic) in *Festuca pratensis* (Fp), *Lolium multiflorum* (Lm), and *L. perenne* (Lp) and corresponding genes (bold) identified using three different reference transcriptomes with all genes or the 6475 near full-length gene set (in parentheses).

Reference transcriptome	$INTRA_{Fp}$	$INTRA_{Lm}$	$INTRA_{Lp}$
Fp	*250,929 (113,934)*	*715,572 (370,524)*	*529,367 (288,958)*
	12,910 (5,552)	**13,864 (5,966)**	**13,472 (5,940)**
Lm	*248,271 (115,390)*	*750,463 (385,592)*	*556,027 (303,890)*
	13,420 (5,750)	**15,320 (6,279)**	**14,717 (6,258)**
Lp	*249,469 (114,545)*	*786,791 (388,114)*	*634,918 (314,057)*
	13,670 (5,768)	**15,368 (6,288)**	**15,617 (6,321)**

Results

De Novo Transcriptome Assembly in Meadow Fescue and Italian Ryegrass

Transcriptome sequencing was performed on 10 samples (corresponding to six genotypes) of meadow fescue and 10 samples (corresponding to six genotypes) of Italian ryegrass, yielding 38.6 million reads per sample on average (ranging from 14.7 to 73.5 million). After quality trimming, reads of each sample were used separately for de novo assembly using CLCBio Genomics Workbench v8.4. This yielded between 41,110 and 100,314 contigs per sample with N50 ranging from 558 to 1025 bp (Supplemental Table S1). The N50 value was markedly lower in Italian ryegrass (666 bp on average) than meadow fescue (941 bp on average) suggesting a higher level of sequence variation and heterozygosity in Italian ryegrass. Using the OGA approach, 18,952 nonredundant Italian ryegrass transcripts were assembled by combining the contigs of all six genotypes based on orthology with the *B. distachyon* proteome. Similarly, 19,036 nonredundant Italian ryegrass transcripts were assembled and annotated (Table 1).

In total, 17,455 orthologous transcripts were shared between the transcriptomes of meadow fescue and Italian ryegrass. Out of these, 16,613 transcripts overlapped with the previously published perennial ryegrass transcriptome containing 19,279 nonredundant transcripts (Ruttink et al., 2013) generated using the same OGA

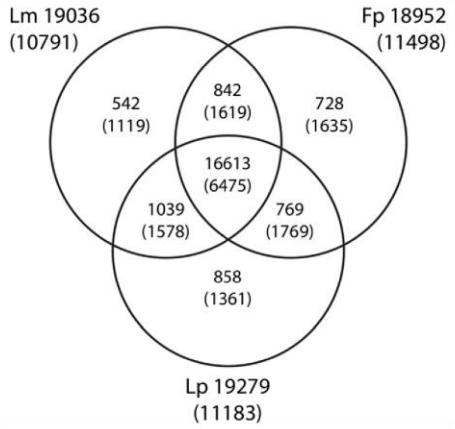


Fig. 2. Venn diagram of the overlap in the number of transcripts in three different reference transcriptomes (*Festuca pratensis*, *Lolium multiflorum*, and *L. perenne*). The number of near full-length transcripts is given in parentheses.

procedure. Only 728, 542, and 858 genes were uniquely assembled in the meadow fescue, Italian ryegrass, and perennial ryegrass reference transcriptomes, respectively (Fig. 2). Most likely, this was due to low expression levels in the other species. From a total of 11,498; 10,791; and 11,183 near full-length transcripts (length of the transcript CDS varied between 85 and 115% of the respective orthologous *B. distachyon* CDS) in meadow fescue, Italian ryegrass, and perennial ryegrass, respectively, 8094 were shared between Italian ryegrass and meadow fescue, and 6475 transcripts were common for all three species (Fig. 2). Estimation of the number of SNPs per transcript depends on the length of the assembled transcript, which varied between species. Thus, to make accurate comparisons of SNP counts and SNP densities between species, we only compared orthologous transcripts that were near full-length in all three references (linked through their OGA based annotation). This is represented by a set of 6475 genes per reference transcriptome (Table 1, 2).

Single-Nucleotide Polymorphism Discovery and Robustness of Genotype Calling

The robustness of SNP calling was tested by comparing genotype calls across the three biological replicates of two genotypes of meadow fescue and two genotypes of Italian ryegrass. The results indicated high-quality SNP data as 98.6, 98.9, 99.1, and 99.2% of the genotype calls for the tetraploid genotypes Fp52, Fp53, Lm51, and Lm52, respectively, were identical among replicates.

Phylogenetic relationships between all genotypes were determined using SNPhylo. Expectedly, genotypes from the three species formed separate clusters and both Italian and perennial ryegrass clustered relatively close to each other. The lowest level of genetic diversity was found among the meadow fescue genotypes (Fig. 3) and the highest among the 14 perennial ryegrass genotypes, which originate from three cultivars and 11 natural accessions from across Europe.

Intraspecific Single-Nucleotide Polymorphisms per Species

We first identified SNPs that were polymorphic within species irrespective of their genotype calls in the other two species (Fig. 1A; Table 1). Similar numbers of genes with at least one SNP were identified in all three species (Table 1), while Italian ryegrass displayed the highest number of SNPs per gene (49.8 SNPs per gene), followed by perennial ryegrass (37.8 SNPs per gene) and meadow fescue (18.5 SNPs per gene) using the Italian ryegrass reference transcriptome. There were only slight differences among the results using three different reference transcriptomes (Supplemental Fig. S2). Because the perennial ryegrass reference transcriptome contains slightly more genes, higher numbers of SNPs and corresponding genes were identified in all three species using this reference transcriptome.

Interspecific Single-Nucleotide Polymorphisms in Group Comparisons

Interspecific SNPs (Fig. 1B) were identified by two-way comparison (when just two species have to be distinguishable) and by three-way comparison (when one species carries a different allele at a given position as compared with the other two species). This analysis was performed separately per reference transcriptome, and data was presented on either the complete reference transcriptome or on the subset of 6475 near full-length genes per reference transcriptome (Table 2) for comparisons among references.

In total, 21,992 SNPs located in 4962 genes of the Italian ryegrass reference transcriptome, which can distinguish meadow fescue from Italian ryegrass, were identified. Of these, 13,373 SNPs were located in 2668 genes of the 6475 near full-length Italian ryegrass gene set. This set of SNPs can be used to develop a platform for genotyping a broader genepool of meadow fescue and Italian ryegrass to validate its potential to discriminate between species. Validated SNPs can further be used to characterize genomic constitution of meadow fescue × Italian ryegrass hybrids without a prior knowledge of parents from which they derived. Similar number of SNPs was identified for perennial ryegrass–meadow fescue combinations (Table 2). In contrast, only 354 interspecific SNPs located in 109 genes of the Italian ryegrass reference transcriptome that consistently discriminate perennial ryegrass from Italian ryegrass were identified. Similarly, a three-way comparison identified 13,260 SNPs located in 3611 genes of the Italian ryegrass reference transcriptome, which discriminated meadow fescue from both *Lolium* species (Table 2). Because of the low number of interspecific SNPs distinguishing perennial ryegrass from Italian ryegrass, only 122 SNPs located in 45 genes of the Italian ryegrass reference transcriptome were

Table 2. Number of interspecific single-nucleotide polymorphisms (italic) in *Festuca pratensis* (Fp), *Lolium multiflorum* (Lm), and *L. perenne* (Lp) and corresponding genes (bold) in two-way and three-way comparisons identified using three different reference transcriptomes with all genes or the 6475 near full-length gene set (in parentheses).

Reference transcriptome	Fp–Lm	Fp–Lp	Lm–Lp	Fp-specific	Lm-specific	Lp-specific
Fp	*20,840 (12,517)*	*20,223 (12,607)*	*358 (231)*	*11,603 (7,341)*	*134 (83)*	*130 (82)*
	4,554 (2,510)	**4,172 (2,348)**	**106 (69)**	**3,183 (1,843)**	**43 (26)**	**58 (40)**
Lm	*21,992 (13,373)*	*22,540 (14,181)*	*354 (241)*	*13,260 (8,406)*	*122 (85)*	*127 (83)*
	4,962 (2,668)	**4,648 (2,564)**	**109 (66)**	**3,611 (2,030)**	**45 (25)**	**62 (39)**
Lp	*24,765 (13,761)*	*26,397 (14,952)*	*472 (274)*	*15,204 (8,720)*	*179 (100)*	*178 (98)*
	5,343 (2,712)	**5,138 (2,641)**	**139 (81)**	**3,976 (2,084)**	**59 (33)**	**80 (47)**

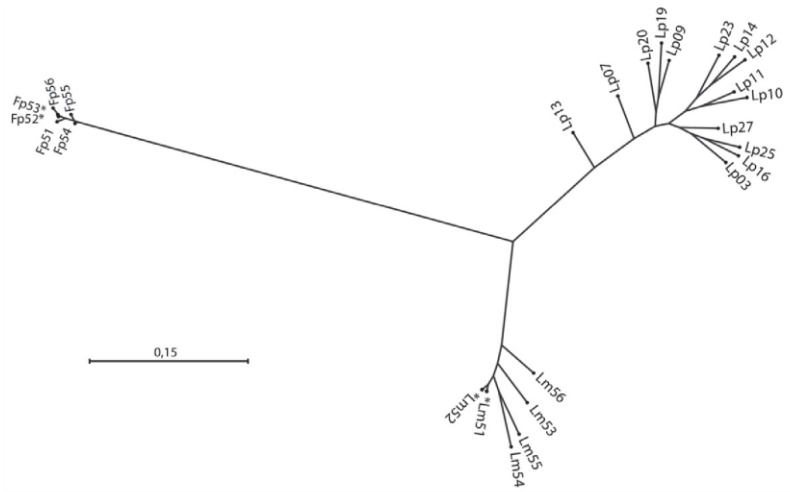


Fig. 3. SNPhylo phylogenetic tree based on 33,725 single-nucleotide polymorphisms shared among 26 genotypes of *Lolium perenne*, *L. multiflorum*, and *Festuca pratensis*. Asterisk (*) indicates overlapping positions in the phylogenetic tree of the three replicates for Fp52, Fp53, Lm51, and Lm52.

classified as species-specific for Italian ryegrass. One hundred twenty-seven SNPs located in 62 genes of the Italian ryegrass reference transcriptome were classified as species-specific for perennial ryegrass.

Because all three reference transcriptomes were annotated based on orthology with the *B. distachyon* proteome, we intersected the gene sets based on common *B. distachyon* orthologs and found that many of the genes with interspecific SNPs were commonly identified using the alternative reference transcriptomes. The numbers of SNPs identified in the set of 6475 near full-length orthologous transcripts present in all three reference transcriptomes were consistent (Table 2), indicating a limited bias of the reference transcriptomes in these analyses. A detailed comparison at the gene-by-gene level further showed high consistency in the number of SNPs called per gene across the 6475 near full-length gene sets (Supplemental Fig. S3). It is important to note that these correlations indicate reproducibility of read mapping and subsequent SNP calling and do not concern the general sequence similarity between orthologous reference sequences and also do not take into account the number of intraspecific SNPs per gene. Taken together, these analyses illustrate that, in general, the three reference sequences per gene can be used interchangeably to identify interspecific SNPs.

Interspecific Single-Nucleotide Polymorphisms in Pairwise Comparisons

Single-nucleotide polymorphisms and corresponding genes that enabled allele discrimination in Italian ryegrass × meadow fescue hybrids derived from known tetraploid parental genotypes were identified in a set of pairwise genotype comparisons. For each pair of parents, two complete reference transcriptomes (meadow fescue and Italian ryegrass) were used. The 62,052 interspecific SNPs located in 8370 genes in the combination Fp54–Lm51 and 93,967 interspecific SNPs located in 10,918 genes in

Table 3. The number of interspecific single-nucleotide polymorphisms (italic) and corresponding genes (bold) in pairwise comparison of selected *Festuca pratensis* (FP) and *Lolium multiflorum* (Lm) genotypes (Fp52, Fp53, or Fp54 vs. Lm51 or Lm52) identified using two different transcriptome references with all genes or with the HQ (high-quality) gene sets per pairwise comparison (in parentheses).

Reference transcriptome		Fp52	Fp53	Fp54
Fp	Lm51	*88,660 (49,261)*	*83,215 (46,518)*	*60,592 (33,228)*
		10,072 (5,090)	**9,667 (4,951)**	**7,737 (4,026)**
	Lm52	*79,744 (43,818)*	*75,810 (41,814)*	*61,957 (34,096)*
		9,277 (4,688)	**8,988 (4,587)**	**7,753 (4,021)**
Lm	Lm51	*93,967 (49,920)*	*86,697 (46,638)*	*62,052 (33,050)*
		10,918 (5,090)	**10,422 (4,951)**	**8,370 (4,026)**
	Lm52	*85,456 (44,633)*	*80,180 (42,332)*	*63,687 (34,046)*
		10,168 (4,688)	**9,775 (4,587)**	**8,374 (4,021)**

the combination Fp52–Lm51 were identified using the Italian ryegrass reference transcriptome (Table 3). Thus, the pairwise comparison strategy enabled scoring higher numbers of interspecific SNPs and respective genes than general group comparisons (INTER-2W and INTER-3W). However, these SNPs were specific to a particular combination of parents. Slightly higher numbers of SNPs were detected when the Italian ryegrass reference was used as compared with the meadow fescue reference.

Because interspecific SNPs in pairwise comparisons discriminated between alleles derived from specific parental genomes, they can be used to quantify PGAS expression and to analyze genes preferentially expressed from one parental genome in hybrid progeny. Since our genotypes were highly polymorphic (Table 1), the possibility was raised that intraspecific SNPs interfere with read mapping to the common consensus reference obtained from different genotypes during OGA. Furthermore, read mapping bias depends on the combination of the reference and the particular hybrid genotype under study as each genotype carries different alleles leading to genotype-specific read–reference mismatch patterns. We tested this potential bias on a gene-by-gene basis (see Fig. 1C) and for each biparental combination. We found that the number of discriminatory SNPs per gene identified using a given meadow fescue reference transcript was not always equal to the number of SNPs identified using the orthologous Italian ryegrass reference transcript (Fig. 4). Therefore, after filtering, as described in Material and Methods and Fig. 1C, only sets of genes without signs of read mapping bias, the so-called HQ gene sets, were kept. This reduced the number of SNPs and corresponding genes to 52 to 56% and 46 to 52%, respectively (Table 3).

Validation of Interspecific Single-Nucleotide Polymorphisms

A set of 10 in silico identified interspecific SNPs for pairwise comparisons located in 10 nuclear genes were validated by HRM analyses. Amplification products sufficiently specific for HRM analysis were obtained for all 10 SNPs. The HRM results were in agreement with the results of the in silico SNP discovery pipeline. The only exception was SNP 1G15070_322, where a clear assignment of the melting profiles into two groups failed, possibly from the presence of an intron between the two primer pairs (Supplemental Table S3).

Genome Location of Genes Containing Various Classes of Single-Nucleotide Polymorphisms

Using the *Lolium* GenomeZipper, 17,768 genes were computationally ordered into the seven perennial ryegrass LGs based on the presumed synteny of perennial ryegrass with *B. distachyon* and using the genetic map of perennial ryegrass as a backbone (Pfeifer et al., 2013). Here, we positioned additional genes into the virtual gene order of perennial ryegrass using the GenomeZipper with *B. distachyon* as reference species. This expanded the GenomeZipper to 23,501 anchored genes distributed across all seven perennial ryegrass chromosomes, ranging from 2488 genes in LG6 to 4276 genes in LG2. Subsequently, we could position various classes of SNPs and their corresponding genes onto the seven perennial ryegrass LGs. This gave an overview of the global spatial distribution of various SNPs (Fig. 5). Only small differences in the frequency of intraspecific, as well as interspecific SNPs, were detected along the chromosomes. Moreover, we did not observe any SNP hotspots or large blocks of genes without intra- or interspecific SNPs. This indicates that any part of the genome can be targeted in breeding or in genomic studies and also enables the detection of even very small introgressions in Festulolium hybrids. Additional information on the position of SNPs and corresponding genes is provided in Supplemental Table S2.

Discussion

The advent of next-generation sequencing technologies has been a driving force in the expansion of genomic resources and the knowledge on plant genome structure, function, evolution, and diversity. While next-generation sequencing technologies readily provide vast amounts of short-read sequence data, it is challenging to assemble sequence reads into longer contigs especially in large and complex genomes with a high fraction of repetitive DNA (Alkan et al., 2011). RNA-seq targets transcribed sequences and avoids sequencing repetitive parts of the genome. Thus, this approach supports gene discovery and directs the identification of SNP markers toward the functional genes. In this study, we extensively sequenced transcriptomes of six meadow fescue and six Italian ryegrass genotypes and applied an OGA strategy for transcriptome assembly and annotation. The transcriptomes were compared with previously published RNA-seq data from 14 perennial ryegrass genotypes and used for large-scale transcript-anchored SNP discovery.

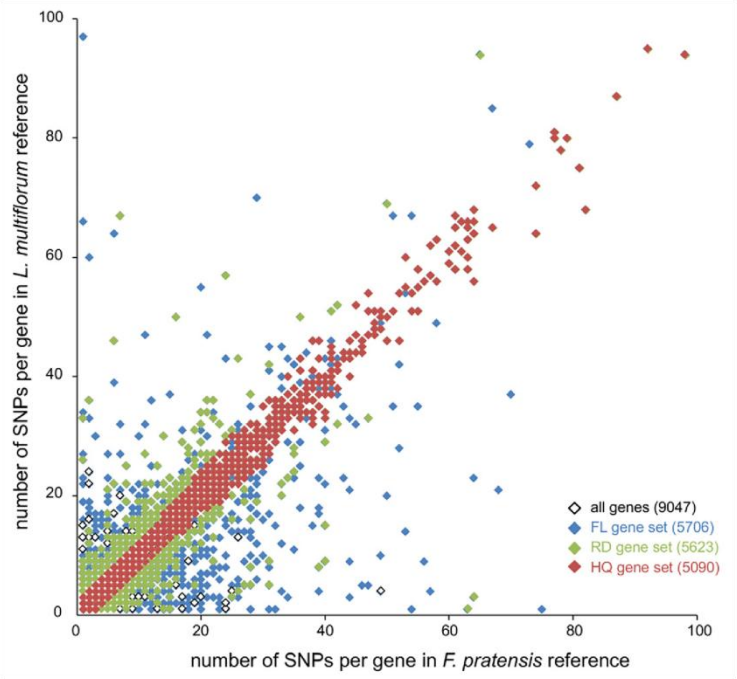


Fig. 4. Selection of single-nucleotide polymorphisms (SNPs) suitable for discrimination between parental genome-specific alleles in progeny derived from a cross between *Festuca pratensis* genotype Fp52 and *Lolium multiflorum* genotype Lm51. Comparison of selected genes sets carrying such SNPs using various filtering criteria. FL, filtered for near full-length transcript length; RD, filtered for average read depth per transcript; HQ (high quality), filtered for number of SNPs per transcript. The numbers of transcripts that pass the filtering criteria are given in parentheses. Only transcripts with at least one SNP per gene in both transcript references are shown.

De Novo Assembly of the Transcriptome of Meadow Fescue, Italian Ryegrass, and Perennial Ryegrass

Using OGA, we assembled 18,952 and 19,036 nonredundant transcripts of meadow fescue and six Italian ryegrass genotypes, respectively, which represent orthologs of ~70% of the *B. distachyon* proteome. This is highly consistent with the previously assembled perennial ryegrass transcriptome (Ruttink et al., 2013). Out of these, 11,498; 10,791; and 11,183 represent near full-length transcripts in meadow fescue, Italian ryegrass, and perennial ryegrass, respectively. The numbers are greater than the 9646 and 10,430 near full-length transcripts of perennial ryegrass (Czaban et al., 2015) but lower than 28,455 gene models identified by Byrne et al. (2015) for perennial ryegrass. This might be at least partially explained by the fact that we used only one tissue for RNA isolation, and thus we did not capture the entire transcriptome.

The main goal of this study was to develop a robust strategy to identify gene-anchored, intra- and interspecific SNPs across three closely related species. However, previous studies on genetic diversity in *Festuca* and *Lolium* revealed great variation in SNP densities across different genes within a species (Ruttink et al., 2013) as well as varying degrees of evolutionary conservation across different genes between species (Czaban et al., 2015). Clearly, SNP identification, as well as read-count-based quantification of allele-specific expression levels, critically depends on the way the reads are mapped onto a representative reference sequence. For this reason, we used multiple SNP and gene filtering to avoid read-mapping bias. The numbers of SNPs identified using different references were compared on a gene-by-gene basis. We proved that application of appropriate filtering ensures comparable results independent on the reference used.

Genetic Variability within Species Predicted by Intraspecific Single-Nucleotide Polymorphisms

A two- to three-fold difference in the number of intraspecific SNPs per gene was found between the species. Compared with both ryegrasses, meadow fescue had the lowest number of intraspecific SNPs per gene, indicating

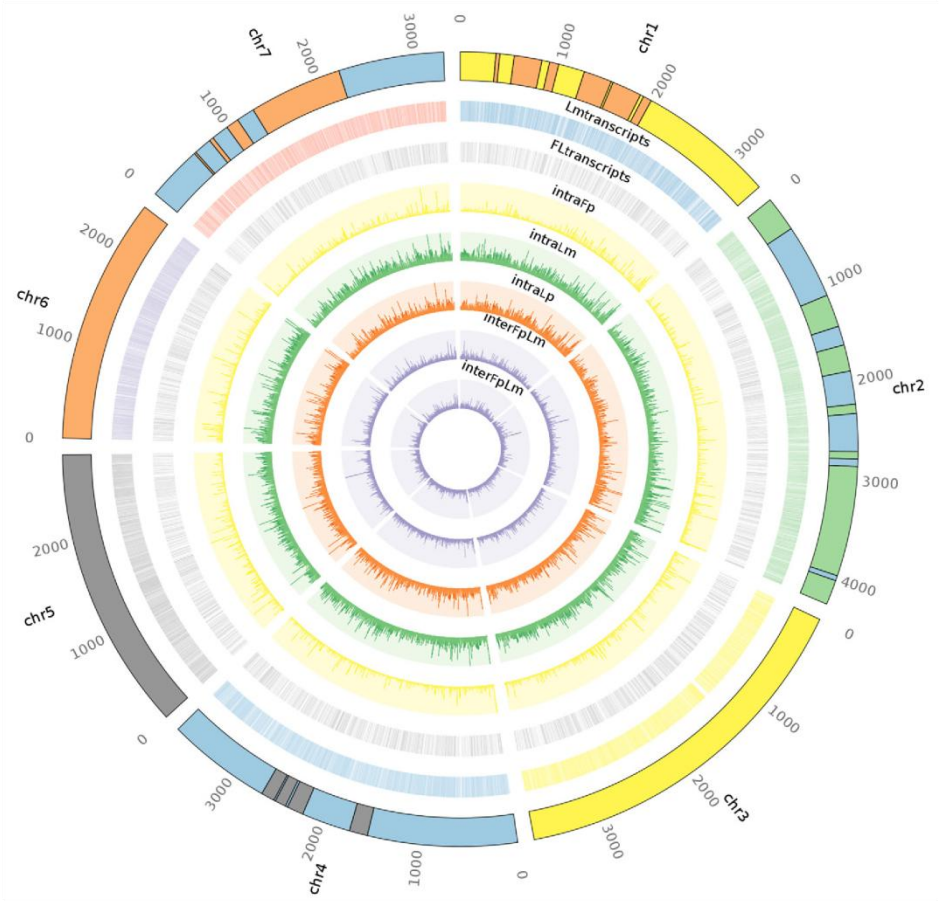


Fig. 5. Circular map of the seven *Lolium perenne* genetic linkage groups. From outer to inner circle: orthologous blocks of *Brachypodium distachyon* chromosomes with *Lolium perenne* (each color represents one *B. distachyon* chromosome: Bd1 blue, Bd2 yellow, Bd3 orange, Bd4 gray, and Bd5 green); distribution of all *L. multiflorum* transcripts; distribution of 6475 near full-length (FL) transcripts assembled in all three species; the number of intraspecific SNPs per gene $INTRA_{Fp}$ (yellow), $INTRA_{Lm}$ (green), $INTRA_{Lp}$ (orange) with maximum 400 SNPs per gene; the number of interspecific SNPs per gene in pairwise comparisons (violet) with a maximum of 150 SNPs per gene ($INTER\text{-}PW_{Fp52\text{-}Lm51}$ and $INTER\text{-}PW_{Fp54\text{-}Lm51}$). Note that the circular maps show linear order of genes but not physical distances between them.

lower genetic variability within this species. This observation, although based on the analysis of a relatively limited number of genotypes per species, correlates with previous findings of Kopecký et al. (2009) and Kölliker et al. (1999). The low level of intraspecific genetic variability is probably a consequence of the genetic bottleneck that meadow fescue has undergone since the last glaciation (Fjellheim et al., 2006).

Interspecific Single-Nucleotide Polymorphisms Enable the Analysis of the Genomic Composition and Gene Expression Analysis of Interspecific Hybrids

Although not novel, SNP genotyping to discriminate plant species and distinguish interspecific hybrids and allopolyploids remains a challenging approach. Curk et al. (2015) employed 454 amplicon sequencing of 57 gene

fragments to mine 105 SNPs distinguishing parental species of modern *Citrus* varieties. High numbers of species-specific SNPs (~33,000) identified in the Illumina Infinium *Brassica* SNP 60K array enabled discrimination of the A and C genomes in *Brassica* allopolyploids (Mason et al., 2015).

The approach presented here allowed us to identify transcript-anchored SNPs that are spread evenly across the genome and can be used to distinguish between the three species (i.e., interspecific SNPs). Moreover, interspecific SNPs can be used to characterize the genomic constitution of interspecific Festulolium hybrids at a high level of resolution. Most of the recent Festulolium cultivars originated from crosses made decades ago followed by several steps of polycrosses and backcrosses (Ghesquiere et al., 2010). Consequently, the information on parental genotypes is lacking for most Festulolium cultivars. For those cultivars, thousands of interspecific SNPs that are monomorphic within species and polymorphic between species (Table 2) can be instrumental to reconstruct their ancestry and relatedness. If new interspecific crosses are made with known parental genotypes, like the ones suggested here, the interspecific SNPs in pairwise comparisons (Table 3) will enable to study genome stability and fine-scale chromosomal rearrangements in subsequent generations of Festulolium (Le Scouarnec and Gribble, 2012). The great advantage of our approach is the even spread of SNP markers across the genomes and a possibility to predict the expected resolution in a given chromosomal region based on the analysis of parental transcriptomes. This may guide the choice of specific parents for interspecific crosses. Another application of selected SNPs in specific HQ gene sets (Table 3) is PGAS expression analysis. We present evidence of gene-specific read mapping bias and resolve this issue by suggesting well-defined controls to avoid reference transcriptome dependent mapping bias, which may affect read-count-based estimations of expression levels (Li and Jiang 2012).

Genotyping Platform for Large-Scale Screening

To facilitate large-scale screening of thousands to tens of thousands of plants in grass breeding programs, a high-throughput genotyping platform needs to be developed with a set of highly informative markers. Customized array-based genotyping platforms were developed for *Lolium* and used to construct genetic maps in biparental crosses (Studer et al. 2012), identify cultivars of Italian and perennial ryegrasses (Wang et al. 2014), and study genetic diversity and the demographic history of natural accessions of perennial ryegrass (Blackmore et al., 2015). Our SNP sets contain complete positional information (putative genome location, neighboring SNPs) on three different types of SNPs suitable for the development of novel probe sets for genes spread across the genome.

Alternatively, targeted resequencing of selected regions of a genome containing intra- and interspecific SNPs could decrease the cost of genotyping. Several methods for target enrichment of genomic DNA have been developed (reviewed in Teer and Mullikin, 2010). Because of a high level of target and sample multiplexing, amplicon resequencing seems to be the most promising approach to genotype hundreds of loci in thousands of plants involved in breeding programs, as it allows simultaneous screening for known SNPs and to discover novel polymorphisms.

Apart from the results from a recent *Lolium–Festuca* study (Czaban et al., 2015), our data represent a valuable resource for ecological and evolutionary genomic research in these species. The two *Lolium–Festuca* transcriptome studies complement each other well. Czaban et al. (2015) used RNA-seq for comparative gene family analysis, phylogenetic analysis, and identification of genes under positive selection pressure. In contrast, our study provides a valuable tool for precise characterization of genomic constitution and gene expression in prospective Festulolium hybrids. Once efficient screening methods have been developed, our set of intraspecific SNPs may be used for QTL fine mapping, map-based cloning of genes involved in traits of agricultural, ecological, and evolutionary interest; genome-wide association studies; demographic history and selection analyses; ecopopulation profiling; and seed contamination screening. Thus, this study provides an outstanding resource for the grass research community and forage grass breeders. Moreover, the versatility of our approach makes it suitable for use in genomic and evolutionary studies of any interspecific hybrids and allopolyploids.

Supplemental Information Available

Supplemental information is available with the online version of this manuscript.

Acknowledgments

This work was supported by Czech Science Foundation (award P501/11/0504), National Program of Sustainability (award no. LO1204), IGA (award no. Prf/2012/001) and Sciex-NMS^{ch}, a scientific exchange program between Switzerland and the new member states of the European Union (Project Code 14.099). Access to computing and storage facilities owned by parties and projects contributing to the National Grid Infrastructure MetaCentrum, provided under the program "Projects of Large Infrastructure for Research, Development, and Innovations" (LM2010005) is greatly appreciated. We thank Sabine van Glabeke for excellent bioinformatics support.

References

Alkan, C., S. Sajjadian, and E.E. Eichler. 2011. Limitations of next-generation genome sequence assembly. Nat. Methods 8:61–65. doi:10.1038/nmeth.1527

Birrer, M., R. Kölliker, Ch. Manzanares, T. Asp, and B. Studer. 2014. A DNA marker assay based on high-resolution melting curve analysis for distinguishing species of the *Festuca–Lolium* complex. Mol. Breed. 8:421–429. doi:10.1007/s11032-014-0044-0

Blackmore, T., I. Thomas, R. McMahon, W. Powell, and M. Hegarty. 2015. Genetic–geographic correlation revealed across a broad European ecotypic sample of perennial ryegrass (*Lolium perenne*) using array-based SNP genotyping. Theor. Appl. Genet. 128:1917–1932. doi:10.1007/s00122-015-2556-3

Byrne, S.L., I. Nagy, M. Pfeifer, I. Armstead, S. Swain, B. Studer, et al. 2015. A synteny-based draft genome sequence of the forage grass *Lolium perenne*. Plant J. 84:816–826. doi:10.1111/tpj.13037

Canter, P.H., I. Pašakinskiene, R.N. Jones, and M.W. Humphreys. 1999. Chromosome substitutions and recombination in the amphiploid

Lolium perenne × *Festuca pratensis* cv. Prior (2*n* = 4*x* = 28). Theor. Appl. Genet. 98:809–814. doi:10.1007/s001220050087
Curk, F., G. Ancillo, F. Ollitrault, X. Perrier, J.P. Jacquemoud-Collet, A. Garcia-Lor, et al. 2015. Nuclear species-diagnostic SNP markers mined from 454 amplicon sequencing reveal admixture genomic structure of modern citrus varieties. PLoS One 10:e0125628. doi:10.1371/journal.pone.0125628
Czaban, A., S. Sharma, S.L. Byrne, M. Spannagl, K.F.X. Mayer, and T. Asp. 2015. Comparative transcriptome analysis within the *Lolium/Festuca* species complex reveals high sequence conservation. BMC Genomics 16:249. doi:10.1186/s12864-015-1447-y
Denoeud, F., J.M. Aury, C. Da Silva, B. Noel, O. Rogier, M. Delledonne, et al. 2008. Annotating genomes with massive-scale RNA sequencing. Genome Biol. 9:R175. doi:10.1186/gb-2008-9-12-r175
DePristo, M.A., E. Banks, R. Poplin, K.V. Garimella, J.R. Maguire, C. Hartl, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498. doi:10.1038/ng.806
Ergon, A., C. Fang, O. Jorgensen, T.S. Aamlid, and O.A. Rognli. 2006. Quantitative trait loci controlling vernalisation requirement, heading time and number of panicles in meadow fescue (*Festuca pratensis* Huds.). Theor. Appl. Genet. 112:232–242. doi:10.1007/s00122-005-0115-z
Farrell, J.D., S. Byrne, C. Paina, and T. Asp. 2014. *De Novo* Assembly of the perennial ryegrass transcriptome using an RNA-seq strategy. PLoS One 9:e103567. doi:10.1371/journal.pone.0103567
Fjellheim, S., O.A. Rognli, K. Fosnes, and C. Brochmann. 2006. Phylogeographical history of the widespread meadow fescue (*Festuca pratensis* Huds.) inferred from chloroplast DNA sequences. J. Biogeogr. 33:1470–1478. doi:10.1111/j.1365-2699.2006.01521.x
Garvin, M.R., K. Saitoh, and A.J. Gharrett. 2010. Application of single nucleotide polymorphisms to non-model species: A technical review. Mol. Ecol. Resour. 10:915–934. doi:10.1111/j.1755-0998.2010.02891.x
Ghesquiere, M., M.W. Humphreys, and Z. Zwierzykowski. 2010. Festulolium. In: B. Boller, U.K. Posselt, and F. Veronesi, editors, Fodder crops and amenity grasses, book series: Handbook of plant breeding 5. Springer Science+Business Media, Berlin. p. 293–316.
Gonçalves da Silva, A., W. Barendse, J.W. Kijas, W.C. Barris, S. McWilliam, R.J. Bunch, et al. 2014. SNP discovery in nonmodel organisms: Strand bias and base-substitution errors reduce conversion rates. Mol. Ecol. Resour. 15:723–736. doi:10.1111/1755-0998.12343
Grabherr, M.G., B.J. Haas, M. Yassour, J.Z. Levin, D.A. Thompson, I. Amit, et al. 2011. Full-length transcriptome assembly from RNA-seq data without a reference genome. Nat. Biotechnol. 29:644–652. doi:10.1038/nbt.1883
Huang, X., and A. Madan. 1999. CAP3: A DNA sequence assembly program. Genome Res. 9:868–877. doi:10.1101/gr.9.9.868
Humphreys, M., U. Feuerstein, M. Vandewalle, and J. Baert. 2010. Ryegrasses. In: B. Boller, U.K. Posselt, and F. Veronesi, editors, Fodder crops and amenity grasses, book series: Handbook of plant breeding 5. Springer Science+Business Media, Berlin. p. 211–260.
Kölliker, R., F.J. Stadelmann, B. Reidy, and J. Nösberger. 1999. Genetic variability of forage grass cultivars: A comparison of *Festuca pratensis* huds., *Lolium perenne* L. and *Dactylis glomerata* L. Euphytica 106:261–270. doi:10.1023/A:1003598705582
Kopecký, D., J. Bartoš, P. Christelová, V. Černoch, A. Kilian, J. Doležel. 2011. Genomic constitution of *Festuca* × *Lolium* hybrids revealed by the DArTFest array. Theor. Appl. Genet. 122:355–363. doi:10.1007/s00122-010-1451-1
Kopecký, D., J. Bartoš, A.J. Lukaszewski, J.H. Baird, V. Černoch, R. Kölliker, et al. 2009. Development and mapping of DArT markers within the *Festuca–Lolium* complex. BMC Genomics 10:473. doi:10.1186/1471-2164-10-473
Kopecký, D., J. Loureiro, Z. Zwierzykowski, M. Ghesquière, and J. Doležel. 2006. Genome constitution and evolution in *Lolium* × *Festuca* hybrid cultivars (Festulolium). Theor. Appl. Genet. 113:731–742. doi:10.1007/s00122-006-0341-z
Kopecký, D., A.J. Lukaszewski, and J. Doležel. 2005. Genomic constitution of Festulolium cultivars released in the Czech Republic. Plant Breed. 124:454–458. doi:10.1111/j.1439-0523.2005.01127.x
Krzywinski, M., J. Schein, I. Birol, J. Conners, R. Gascoyne, D. Horsman, S.J. Jones, and M.A. Marra. 2009. Circos: An information aesthetic for comparative genomics. Genome Res. 19:1639–1645. doi:10.1101/gr.092759.109
Książczyk, T., E. Zwierzykowska, K. Molik, M. Taciak, P. Krajewski, Z. Zwierzykowski. 2015. Genome-dependent chromosome dynamics in three successive generations of the allotetraploid *Festuca pratensis* × *Lolium perenne* hybrid. Protoplasma 252:985–996. doi:10.1007/s00709-014-0734-9
Li, H., and R. Durbin. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760. doi:10.1093/bioinformatics/btp324
Li, W., and T. Jiang. 2012. Transcriptome assembly and isoform expression level estimation from biased RNA-Seq reads. Bioinformatics 28:2914–2921. doi:10.1093/bioinformatics/bts559
Lee, T.H., H. Guo, X. Wang, C. Kim, and A.H. Paterson. 2014. SNPhylo a pipeline to construct a phylogenetic tree from huge SNP data. BMC Genomics 15:162. doi:10.1186/1471-2164-15-162
Le Scouarnec, S., and S.M. Gribble. 2012. Characterising chromosome rearrangements: Recent technical advances in molecular cytogenetics. Heredity 108:75–85. doi:10.1038/hdy.2011.100
Mason, A.S., J. Takahira, C. Atri, B. Samans, A. Hayward, W.A. Cowling, J. Batley, and M.N. Nelson. 2015. Microspore culture reveals complex meiotic behaviour in a trigenomic *Brassica* hybrid. BMC Plant Biol. 15:173. doi:10.1186/s12870-015-0555-9
Mayer, K.F.X., S. Taudien, M. Martis, H. Šimková, P. Suchánková, H. Gundlach, et al. 2009. Gene content and virtual gene order of barley chromosome 1H. Plant Physiol. 151:496–505. doi:10.1104/pp.109.142612
McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20:1297–1303. doi:10.1101/gr.107524.110
Modrek, B., and C. Lee. 2002. A genomic view of alternative splicing. Nat. Genet. 30:13–19. doi:10.1038/ng0102-13
Mortazavi, A., B.A. Williams, K. McCue, L. Schaeffer, and B. Wold. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. Nat. Methods 5:621–628. doi:10.1038/nmeth.1226
Pašakinskiene, I., C.M. Griffiths, A.J.E. Bettany, V. Paplauskiene, and M.W. Humphreys. 2000. Anchored simple-sequence repeats as primers to generate species-specific DNA markers in *Lolium* and *Festuca* grasses. Theor. Appl. Genet. 100:384–390. doi:10.1007/s001220050050
Peng, Y., X. Gao, R. Li, and G. Cao. 2014. Transcriptome sequencing and de novo analysis of *Youngia japonica* using the Illumina platform. PLoS One 9:e90636. doi:10.1371/journal.pone.0090636
Pfeifer, M., M. Martis, T. Asp, K.F.X. Mayer, T. Lübberstedt, S. Byrne, U. Frei, and B. Studer. 2013. The perennial ryegrass GenomeZipper: Targeted use of genome resources for comparative grass genomics. Plant Physiol. 161:571–582. doi:10.1104/pp.112.207282
Proost, S., M. Van Bel, D. Vaneechoutte, Y. Van de Peer, D. Inzé, B. Mueller-Roeber, and K. Vandepoele. 2014. PLAZA 3.0: An access point for plant comparative genomics. Nucleic Acids Res. 43:D974–D981. doi:10.1093/nar/gku986
Robertson, G., J. Schein, R. Chiu, R. Corbett, M. Field, S.D. Jackman, et al. 2010. De novo assembly and analysis of RNA-seq data. Nat. Methods 7:909–912. doi:10.1038/nmeth.1517
Rognli, O.A., M.C. Saha, S. Bhamidimarri, and S. van der Heijden. 2010. Fescues. In: B. Boller, U.K. Posselt, and F. Veronesi, editors, Fodder crops and amenity grasses, book series: Handbook of plant breeding 5. Springer Science+Business Media, Berlin. p. 261–292.
Ruttink, T., L. Sterck, A. Rohde, C. Bendixen, P. Rouzé, T. Asp, Y. Van de Peer, and I. Roldan-Ruiz. 2013. Orthology guided assembly in highly heterozygous crops: Creating a reference transcriptome to uncover genetic diversity in *L. perenne*. Plant Biotechnol. J. 11:605–617. doi:10.1111/pbi.12051
Schulz, M.H., D.R. Zerbino, M. Vingron, and E. Birney. 2012. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092. doi:10.1093/bioinformatics/bts094
Simpson, J.T., K. Wong, S.D. Jackman, J.E. Schein, S.J.M. Jones, and Î. Birol. 2009. ABySS: A parallel assembler for short read sequence data. Genome Res. 19:1117–1123. doi:10.1101/gr.089532.108
Studer, B., B. Boller, D. Herrmann, E. Bauer, U.K. Posselt, F. Widmer, R. Kölliker. 2006. Genetic mapping reveals a single major QTL for

bacterial wilt resistance in Italian ryegrass (*Lolium multiflorum* Lam.). Theor. Appl. Genet. 113:661–671. doi:10.1007/s00122-006-0330-2

Studer, B., S. Byrne, R.O. Nielsen, F. Panitz, C. Bendixen, M.S. Islam, M. Pfeifer, T. Lübberstedt, and T. Asp. 2012. A transcriptome map of perennial ryegrass (*Lolium perenne* L.). BMC Genomics 13:140. doi:10.1186/1471-2164-13-140

Studer, B., L.B. Jensen, A. Fiil, and T. Asp. 2009. "Blind" mapping of genic DNA sequence polymorphisms in *Lolium perenne* L. by high resolution melting curve analysis. Mol. Breed. 24:191–199. doi:10.1007/s11032-009-9291-x

Teer, J.K., and J.C. Mullikin. 2010. Exome sequencing: The sweet spot before whole genomes. Hum. Mol. Genet. 19:R145–R151. doi:10.1093/hmg/ddq333

The International Brachypodium Initiative. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. Nature 463:763–768. doi:10.1038/nature08747

Varshney, R.K., S.N. Nayak, G.D. May, and S.A. Jackson. 2009. Next generation sequencing technologies and their application for crop genetics and breeding. Trends Biotechnol. 27:522–530. doi:10.1016/j.tibtech.2009.05.006

Wang, J., L.W. Pembleton, R.C. Baillie, M.C. Drayton, M.L. Hand, M. Bain, et al. 2014. Development and implementation of a multiplexed single nucleotide polymorphism genotyping tool for differentiation of ryegrass species and cultivars. Mol. Breed. 33:435–451. doi:10.1007/s11032-013-9961-6

Wu, T., S. Luo, R. Wang, Y. Zhong, X. Xu, Y. Lin, X. He, B. Sun, H. Huang. 2014. The first Illumina-based de novo transcriptome sequencing and analysis of pumpkin (*Cucurbita moschata* Duch.) and SSR marker development. Mol. Breed. 34:1437–1447. doi:10.1007/s11032-014-0128-x

Yamada, T., J.W. Forster, M.W. Humphreys, and T. Takamizo. 2005. Genetics and molecular breeding in *Lolium/Festuca* grass species complex. Grassland Science 51:89–106. doi:10.1111/j.1744-697X.2005.00024.x

Zerbino, D.R., and E. Birney. 2008. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. Conserv. Genet. Resour. 18:821–829.

Zwierzykowski, Z., A.J. Lukaszewski, A. Lesniewska, and B. Naganowska. 1998a. Genomic structure of androgenic progeny of pentaploid hybrids, *Festuca arundinacea* × *Lolium multiflorum*. Plant Breed. 117:457–462. doi:10.1111/j.1439-0523.1998.tb01973.x

Zwierzykowski, Z., R. Tayyar, M. Brunell, and A.J. Lukaszewski. 1998b. Genome recombination in intergeneric hybrids between tetraploid *Festuca pratensis* and *Lolium multiflorum*. J. Hered. 89:324–328. doi:10.1093/jhered/89.4.324

4.1.5 HIPPIES NAŠICH LUK: PŘÍBĚH MEZIRODOVÝCH KŘÍŽENCŮ TRAV

Štěpán Stočes a David Kopecký

Živa 8/2016, str. 158-161

Štěpán Stočes, David Kopecký

Nové poznatky v genetice rostlin I. Hippies našich luk: příběh mezirodových kříženců trav

Pozn. redakce: V tomto a následujícím ročníku bude vycházet volný seriál článků věnovaný směrům výzkumu Centra strukturní a funkční genomiky rostlin, olomouckého pracoviště Ústavu experimentální botaniky AV ČR, v. v. i., kde nejmodernějšími metodami cytogenetiky, molekulární biologie a genomiky studují strukturu a funkci dědičné informace rostlin. Centrum se řadí ke světově uznávaným pracovištím a účastní se významných mezinárodních projektů (viz např. Živa 2012, 4: 155–157); svými poznatky přispívá ke šlechtění nových odrůd zemědělských plodin. Jeho pracovníci získali mnohá domácí i mezinárodní ocenění. Od r. 2015 se podílí na koordinaci výzkumného programu Potraviny pro budoucnost v rámci Strategie AV21 (blíže v Živě 2015, 1: 2–3.)

Tento příspěvek jsme zaměřili na rod jílek (*Lolium*), který je se svými 8 druhy mezi ostatními zástupci velké a početné čeledi lipnicovitých (*Poaceae*) tak trochu chudým příbuzným. Přesto si troufáme tvrdit, že se s jílky většina z nás setkává téměř každý den. Dva druhy – jílek vytrvalý (*L. perenne*) a j. mnohokvětý (*L. multiflorum*) totiž patří mezi nejčastější trávy našich luk, pastvin a trávníků. V článku přiblížíme objevování podstaty genetické informace těchto významných trav.

Jílek mnohokvětý je ceněn pro vysoký výnos suché hmoty a vynikající kvalitu píce. Přestože se pěstuje na dočasných loukách a využívá především k produkci sena a siláže na orné půdě, tvoří rovněž významnou součást trvalých travních porostů. Jílek vytrvalý se naproti tomu pěstuje především k pastvě a zejména díky své vytrvalosti patří mezi hlavní složky stálých travních ploch. Poskytuje ještě kvalitnější píci než jílek mnohokvětý. Tento druh se navíc používá jako klíčová součást našich trávníků. Najít ho můžeme ve sportovních (na fotbalových nebo golfových hřištích) i okrasných trávnících (na veřejných prostranstvích, kolem rodinných domů), ale také podél silniční a dálniční sítě a v okolí letišť.

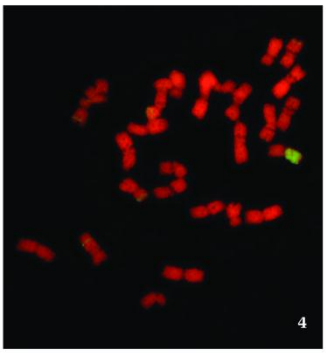
Cestu na výsluní si jílky nehledaly snadno. Samotné slovo *Lolium* pochází z latiny a poprvé bylo zmíněno ve Vergiliově Georgice neboli Zpěvech rolnických jakožto obtížný plevel. Jílky mají svou domovinu v Evropě, severní Africe a mírném pásu Asie, ale dnes jsou rozšířené v mírném pásu všech kontinentů. Podle způsobu rozmnožování lze tento rod rozdělit na dvě hlavní vývojové větve – samosprašné druhy *L. persicum*, jílek oddálený (*L. remotum*), j. cizí (*L. loliaceum*) a j. mámivý (*L. temulentum*), a druhy cizosprašné, kam řadíme již zmíněné jílky mnohokvětý a vytrvalý a také j. tuhý (*L. rigidum*). Posledním zástupcem rodu je endemit Kanárských ostrovů *L. canariense*, který představuje intermediální formu rozmnožování.

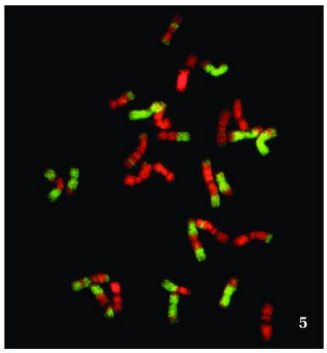
Osud jílků se začal psát někdy před necelými třemi miliony let, kdy se větev samosprašných druhů oddělila pravděpodobně od předchůdce dnešní kostřavy luční (*Festuca pratensis*). Cizosprašné druhy se od kostřavy oddělily před 2,35 milionu let a zhruba před jedním milionem let došlo k oddělení jílku mnohokvětého a j. vytrvalého. Molekulární analýzy prokázaly, že v porovnání s kostřavou luční vykazují oba druhy jílku podstatně větší genetickou variabilitu. U kostřavy luční zřejmě došlo buď v průběhu, nebo krátce po skončení

poslední doby ledové k tzv. efektu hrdla láhve (bottleneck effect), kdy byla genetická variabilita dramaticky redukována snížením počtu jedinců v populaci. Přesto má i kostřava luční šlechtitelům a zemědělcům co nabídnout. Jde o výrazně mrazuvzdorný druh uplatňující se na loukách a pastvinách Skandinávie, střední a východní Evropy a v alpských regionech. U nás se využívá především v druhově bohatých trvalých travních porostech – na pastvinách a loukách k produkci sena. Kostřava luční se podílela na evoluci dalšího agronomicky významného druhu – kostřavy rákosovité (*F. arundinacea*). Ta vznikla mezidruhovým křížením diploidní k. luční (2n = 2x = 14, kde 2n je diploidní počet chromozomů, x značí základní chromozomové číslo) a tetraploidní *F. glaucescens* (2n = 4x = 28), druhu v současnosti známého z vyšších poloh Pyrenejí a Alp. Vzhledem k tomu, že kostřava rákosovitá je hexaploidní, došlo zřejmě okamžitě po hybridizaci ke zdvojení počtu chromozomů – tzv. alopolyploidizaci (amfidiploidii), což umožnilo párování homologních chromozomů v průběhu meiózy a tudíž stabilizaci genomu. Tato kostřava mohla vzniknout také splynutím neredukovaných gamet, které vznikají u této skupiny trav poměrně běžně. Kostřava rákosovitá tvoří složku intenzivně hnojených travních porostů a přidává se do trvalých porostů k prodloužení pastevního období. Tento druh se vyznačuje velmi bohatým kořenovým systémem, který zřejmě podmiňuje jeho vynikající odolnost k suchu.

Kostřavy bývají společně s jílky a dalšími travními druhy (zástupci rodů lipnice – *Poa*, bojínek – *Phleum*, srha – *Dactylis*) používány ve směsích speciálně sestavených k různým účelům v zemědělství (pastevní, luční nebo siláţní směsi) i jinde (např. pro fotbalové trávníky). Uplatnění mezidruhových nebo mezirodových hybridů, kteří v sobě kombinují pozitivní vlastnosti rodičů, by mohlo složení směsí zjednodušit. Mezirodová hybridizace se přímo nabízí u kostřav a jílků, kdy kombinace vysokého výnosu a nutričních vlastností jílků s výbornou adaptací kostřav přežívat abiotické stresy by mohla vytvořit trávy schopné nejenom konkurovat rodičovským druhům, ale v řadě znaků je překonat a lépe se přizpůsobit měnícímu se klimatu.

V přírodě se kostřavy a jílky kříží, ale vzniklí hybridi jsou většinou sterilní. První nálezy kříženců pocházejí z jihovýchodní Anglie a severozápadní Francie z 30. let 20. stol. Hybridita rostlin byla v té době určována na základě intermediárního typu květenství – tedy něco mezi lichoklasem jílku a latou kostřavy (obr. 2 a 3). Tento znak se však ukázal jako nespolehlivý. Většina kříženců vykazuje květenství jílkového typu – klas. Kříţence v přírodě nacházíme stále častěji a např. v jižní Anglii okupují hybridi diploidního jílku vytrvalého (2n = 2x = 14; LpLp – se dvěma sadami chromozomů j. vytrvalého; uvedené kódy, i u dalších taxonů, jsou vždy odvozeny z počátečních písmen latinského jména) a diploidní kostřavy luční (2n = 2x = 14; FpFp) břehy řek, kde výborně přežívají pravidelné zaplavování a vytěsňují rodičovské druhy. Cytologické analýzy prokázaly, že zmíněné rostliny jsou skutečně hybridní a v jejich





1 Pokusná školka F_2 generace kříženců kostřavy luční (*Festuca pratensis*) a jílku mnohokvětého (*Lolium multiflorum*) na pozemku Ústavu experimentální botaniky AV ČR, v. v. i., v Olomouci
2 Květenstvím jílků je klas. Detail květu jílku vytrvalého (*L. perenne*)
3 Větvení laty u kostřavy luční, většinou není u jejích kříženců s jílky patrné.
4 U introgresní odrůdy 'Bečva', vzniklé křížením jílku mnohokvětého a kostřavy luční a následným zpětným křížením F_1 hybrida s jílkem, došlo téměř k úplné eliminaci kostřavové DNA (zelená barva) a obrovské převaze jílkové DNA (červená).
5 Cytogenetická analýza amfiploidní odrůdy 'Elmet' pocházející z křížení jílku mnohokvětého s kostřavou luční odhalila četné výměny (vzniklé homeologními rekombinacemi) mezi genomem kostřavy (zelená barva) a jílku (červená).

karyotypu se vyskytuje jedna sada chromozomů kostřavy a jedna jílku (2n = 2x = 14; LpFp), mezi nimiž dochází k četným výměnám, tzv. homeologním rekombinacím. Překvapivé bylo zjištění, že mezi očekávanými diploidy se nacházely rovněž rostliny triploidní (2n = 3x = 21). Tito triploidi pravděpodobně vznikli spojením redukované gamety jednoho rodiče a nередukované gamety rodiče druhého. Cytogenetická analýza odhalila, že mezi triploidy se vyskytují oba typy – LpLpFp i LpFpFp, a tudíž schopnost vytvářet nередukované gamety mají oba druhy. Na rozdíl od sterilních diploidů vykazují triploidní hybridní rostliny částečnou samičí fertilitu a teoreticky mohou zpětným křížením s jedním z rodičovských druhů tvořit potomstvo. Pro jakékoli křížence mezi kostřavami a jílky se vžilo souhrnné označení ×Festulolium či prostě Festulolium.

Imitace matky přírody na šlechtitelských stanicích

V 60. a 70. letech minulého století se s rozvojem biotechnologických metod otevřela cesta k alternativním způsobům šlechtění. Šlechtitelé trav začali křížit rostliny napříč druhy či dokonce rody a kultivací nezralých embryí byli schopni získat životaschopné kříženec. Patrně první úspěch na poli mezirodové hybridizace jílků a kostřav zaznamenala skupina prof. Roberta C. Bucknera ve Spojených státech amerických, který křížil diploidní jílek mnohokvětý s hexaploidní kostřavou rákosovitou. Tetraploidní hybrid byl poté několikrát zpětně křížen s k. rákosovitou. Takto získaná odrůda 'Kenhy' byla zaregistrována jako kostřava rákosovitá a v jejím genomu se vyskytuje dědičná informace jílku jen sporadicky. Přesto se stala komerčně úspěšnou a po dlouhou dobu se pěstovala v monokulturách nebo směsích s několika odrůdami kostřavy rákosovité. Zpětného křížení a vytvoření tzv. introgresních odrůd využil i český šlechtitel Antonín Fojtík. Postupoval podobně jako R. C. Buckner a rovněž vytvořil kříženec jílku mnohokvětého a kostřavy rákosovité. Zpětné křížení ale prováděl s oběma rodičovskými druhy a časem vytvořil tetraploidní odrůdy jílkového typu 'Lofa' a 'Bečva' (obr. 4, během šlechtění došlo k několikanásobným zpětným křížením s jílkem) a také řadu odrůd kostřavového typu podobných odrůdě 'Kenhy' (zde se zpětné křížení provádělo s k. rákosovitou). Introgresní odrůdy Festulolium se staly velmi žádanými na trhu s travními semeny a odrůda 'Lofa' vlajkovou lodí šlechtitelské stanice v Hladkých Životicích na severní Moravě, kde A. Fojtík pracoval. Tento typ odrůd představuje jakési obohacení jednoho z rodičovských druhů několika málo vlastnostmi druhého rodiče. Pomocí molekulárních a cytologických přístupů odhalujeme v genomu těchto introgresních odrůd jen malé úseky chromozomů jednoho rodiče s absolutní převahou DNA rodiče druhého. U některých odrůd proběhla dokonce úplná eliminace jednoho z rodičovských genomů.

Vedle introgresních odrůd Festulolií se šlechtí i tzv. alopolyploidní (či amfiploidní) odrůdy, kdy po vytvoření F_1 kříženců dojde k volnému sprášení mezi rostlinami této generace. Jak bylo zmíněno výše, diploidní kříženci vykazují kompletní samčí sterilitu. Pro její překonání se proto kříží autotetraploidní rostliny jílku (LpLpLpLp) a kostřavy (FpFpFpFp) a získaní alotetraploidní kříženci mají genom LpLpFpFp. V případě křížení kostřavy luční s autotetraploidním jílkem mnohokvětým (LmLmLmLm) mají kříženci konstituci LmLmFpFp.

První úspěchy s amfiploidy zaznamenali šlechtitelé ve Walesu, když koncem 70. let vytvořili a zaregistrovali tetraploidní odrůdy trav 'Prior' (LpLpFpFp) a 'Elmet' (LmLmFpFp; obr. 5). Tyto dvě odrůdy se však na trhu neuplatnily a v dnešní době již nejsou nabízeny. Přesto právě ony otevřely amfiploidním Festulolиím cestu do povědomí šlechtitelů a farmářů. V řadě zemí se poté začala amfiploidní Festulolia produkovat a A. Fojtík hrál znovu klíčovou roli ve šlechtění těchto kříženců. Na šlechtitelské stanici v Hladkých Životicích (dnes DLF Seeds, s. r. o., která je součástí nadnárodní dánské semenářské firmy DLF Seeds A/S) se do dnešního dne podařilo vyšlechtit a zaregistrovat přes dvě desítky odrůd Festulolií, přičemž řada z nich je na semenářském trhu velmi úspěšná. Povedlo se u nich totiž zkombinovat značný výnos semene z jílku s vytrvalostí a odolností k abiotickým stresům (především k chladu a mrazu) získanou z kostřavy luční.

V kříženci je jílek králem

Jak jsme již zmínili, agronomicky významné kostřavy a jílky jsou cizosprašné druhy a touto vlastností se vyznačují i jejich kříženci. To má za následek značnou variabilitu genomového složení, kdy je každá rostlina v podstatě jedinečná. Zastoupení DNA obou rodičů v rostlinách hybridních odrůd lze určit pomocí genomové hybridizace *in situ* (GISH). Tato cytogenetická technika se zakládá na značení DNA dvou potenciálních rodičovských druhů dvěma různými fluorochromy (fluorescenčními barvivy) a následné hybridizaci značených DNA (sond) s mitotickými chromozomy kříženec na mikroskopickém preparátu.

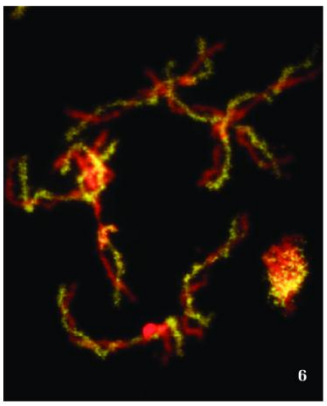
Sekvence DNA sondy a chromozomů, které jsou k sobě komplementární, spolu hybridizují a ve fluorescenčním mikroskopu pak pozorujeme chromozomy, nebo jejich části, které „svítí“ jednou, nebo druhou barvou a indikují tak jejich příslušnost k rodičovskému druhu (viz obr. 4 a 5).

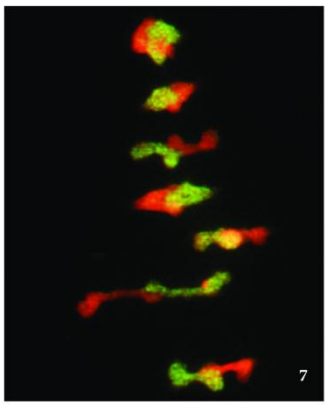
V naší laboratoři jsme ve spolupráci se šlechtitelskou stanicí v Hladkých Životicích takto určili genomové složení všech dostupných odrůd Festulolií. Výsledky byly poměrně překvapivé, když jsme téměř u všech odrůd zjistili převahu jílkového genomu nad kostřavovým. Výjimku tvořily pouze odrůdy introgresní formy, jejichž šlechtění zahrnovalo zpětná křížení s kostřavou rákosovitou. U ostatních byla variabilita mezi odrůdami (ale i v rámci jednotlivých odrůd) značná – od těch, které vykazovaly poměr rodičovských genomů téměř 1 : 1 s jen velmi mírnou dominancí jílkového genomu nad kostřavovým, až po odrůdy, kde pouze několik rostlin obsahovalo malý úsek chromozomu kostřavového původu. U pěti odrůd jsme dokonce nebyli schopni detekovat žádnou DNA z kostřavy. Dominanci jílkového genomu nad kostřavovým pozorovali rovněž naši kolegové z Institute of Plant Genetics v polské Poznani. Studovali genomové složení u 6 následných generací kříženců (F_1 až F_6) kostřavy luční s jílkem mnohokvětým a všimli si plynulého posunu směrem k jílku z generace na generaci. To by mohlo být částečně vysvětleno přednostním výběrem rostlin, které vykazují vlastnosti jílku. Takové jedince totiž šlechtitelé preferují – s vysokým výnosem, rychlým vývojem na jaře a dobrými nutričními vlastnostmi. Překvapení ale nastalo, když polští vědci zkoumali rostliny následných generací vybrané náhodně. I zde se ukázal plynulý posun v genomovém složení hybridů v následných generacích směrem k jílkovému genomu. Příčinu neznáme, ale řada experimentů nám pomáhá tento fenomén osvětlit.

Jílek boduje i na úrovni RNA

Samotná přítomnost DNA z jílku a z kostřavy v genomu křížence ještě neznamená, že se všechny geny obou rodičů projeví. V jednom z našich posledních projektů jsme se zaměřili na studium genové exprese u prvních generací kříženců. Nejdříve jsme vytvořili reciproké tetraploidní křížence jílku mnohokvětého a kostřavy luční tak, že jednou byl jako mateřská rostlina použit jílek a podruhé kostřava. Ze získaných F_1 kříženců jsme jejich volným sprášením odvodili další generace a pomocí metody GISH vybrali rostliny, které měly stejný počet rodičovských chromozomů (14 z jílku a 14 z kostřavy) a nevykazovaly žádné výměny částí rodičovských chromozomů mezi sebou.

Napříč všemi rostlinnými druhy, a především blízce příbuznými, se nacházejí velmi podobné sestavy genů. Geny různých druhů, jež se vyvinuly ze stejného genu společného předka, mající (téměř) stejnou sekvenci DNA a vytvářející (téměř) stejný produkt, nazýváme ortologní geny či ortology. To umožňuje jejich identifikaci u druhů s neznámou sekvencí genomu, jako v případě kostřav a jílků. Tímto způsobem jsme určili ortologní geny, které spolu sdílejí jílek mnohokvětý, kostřava luční a blízce příbuzný druh trávy – válečka *Brachypodium distachyon*, jejíž malý genom byl už přečten. U všech rostlin našich tetraploidních kříženců má každý takový gen dvě kopie z jílku a dvě kopie z kostřavy. Protože však v sekvencích genů došlo během evoluce po oddělení obou druhů (nebo rodů) k několika bodovým mutacím (záměnám jednotlivých nukleotidů), můžeme rodičovské varianty určitého genu identifikovat pomocí jednonukleotidových polymorfismů (SNP – Single Nucleotide Polymorphism; viz též Živa 2016, 2: 61–63 nebo 2011, 6: 262–263). Po izolaci a sekvenování RNA z obou rodičů a reciprokých hybridů jsme takto byli schopni zjistit, jak se hladina RNA transkriptů jako míra exprese genů liší mezi rodičovskými druhy a k jakým změnám dochází u kříženců v první a následných generacích. Analýza ca 12 tisíc ortologních genů u rodičovských druhů ukázala, že se pouze velmi malá část exprimuje rozdílně. U F_1 generace kříženců byla situace odlišná. Většina genů se v kříženci sice stále exprimovala stejně jako u obou rodičů, avšak nemalý počet (ca 800–900) vykazoval zvýšenou expresi


6


7

6 a 7 Promiskuitní párování (blíže v textu) chromozomů diploidního křížence kostřavy luční (zeleně) a jílku mnohokvětého (červeně) v profázi (obr. 6) a metafázi (7) prvního meiotického dělení

8 Typické místo výskytu tetraploidní kostřavy apeninské (*F. apennina*) ve Švýcarsku (Küblisbühlschwand)

9 Lokalita Le Moléson ve Švýcarsku, kde se nachází hybridní zóna diploidní kostřavy luční a tetraploidní kostřavy apeninské v nadmořské výšce 1 200 až 1 800 m. Ve výšce pod 1 200 m n. m. roste pouze kostřava luční a na místech nad 1 800 m n. m. jen k. apeninská.

10 Kostřava apeninská se vyskytuje do výšek přesahujících až 2 000 m n. m., jako na úbočí švýcarské hory La Para. Snímky D. Kopeckého


8


9

jílkových variant genů a jen několik genů pocházejících z kostřavy bylo exprimováno přednostně. Téměř stejné výsledky jsme získali u reciprokých hybridů a nezaznamenali jsme výraznější vliv směru křížení. Genovou expresi tedy neovlivnilo, zda byl jílek použit jako otcovská, nebo mateřská rostlina. V následných generacích narůstal počet genů exprimovaných přednostně z jílkové varianty.

V další naší práci budeme studovat vliv stárnutí rostlin a působení stresových podmínek (odolnost proti chladu a mrazu) na genovou expresi. Tyto faktory by mohly prospívat kostřavovému genomu. Již jsme uvedli, že kostřava je vytrvalejší a mnohem lépe přežívá stresové podmínky. Chtěli bychom zjistit, zda působení stresu může zvrátit genovou expresi na úrovni celého genomu ve prospěch kostřavy. Jsme si vědomi, že studujeme pouze část ze všech genů a že geny, které jsou specifické pro jeden rodičovský druh, zatímco v druhém rodiči se nevyskytují, nejsme technicky schopni v naší práci obsáhnout. Přesto by nové výsledky mohly přispět k rozšíření poznatků o příčinách dominance jílku v kříženci a k pochopení funkce hybridního genomu.

Volná láska během meiózy

Postupná převaha jílku by mohla být zapříčiněna i tím, že během meiotického dělení dochází k nepřesnostem v párování chromozomů a kostřavové chromozomy jsou během tohoto procesu postupně eliminovány. U naprosté většiny mezidruhových kříženců se chromozomy rodičovských druhů mezi sebou nepárují. Pokud mají svého homologa (v případě tetraploidního kříženec máme dvě sady homologních chromozomů z jednoho rodiče a dvě sady z druhého rodiče), párují se s ním. V opačném případě, např. když je hybrid diploidní a nese po jedné saně chromozomů od každého rodiče, se chromozomy nepárují a meiózou procházejí osamoceně jako univalenty. To samozřejmě vede ke sterilitě kříženec, ale může být způsobena dvěma faktory. Prvním z nich je skutečnost, že homeologní chromozomy (tedy chromozomy, které se vyvinuly ze stejného chromozomu společného předchůdce) nemají dostatečnou sekvenční podobnost a nedochází k jejich synapsi (spojení) na počátku zygotene (druhého stadia profáze prvního meiotického dělení), kdy se chromozomy začínají párovat. Druhé vysvětlení spočívá v přítomnosti genu regulujícího párování chromozomů. Takový systém byl popsán u tetraploidní a hexaploidní pšenice (je nazván *Ph* z anglického termínu pairing homoeologous) – dovoluje párování pouze homologních chromozomů a zaručuje striktní tvorbu bivalentů. Podobný stav nacházíme u polyploidních kostřav. U hexaploidní kostřavy rákosovité a dalších planých příbuzných se vyskytuje systém, který umožňuje párování chromozomů pouze s jejich homologními partnery.

Jak jsme již zmínili, diploidní kříženci kostřav a jílků jsou rovněž sterilní a předpokládalo se, že se chromozomy obou rodičů v hybridech nepárují. Cytologické analýzy ale prokázaly bezproblémové párování chromozomů jílku s chromozomy kostřavy (obr. 6 a 7). Promiskuitní chování chromozomů v průběhu meiózy hybridů


10

nás inspirovalo k názvu tohoto článku, který odkazuje na volnomyšlenkářské hnutí 60. let 20. stol. Sterilita diploidních hybridů tedy musí mít původ jinde a zřejmě souvisí s nějakým typem postzygotické bariéry. Vraťme se ale k párování chromozomů. U diploidních kříženců nemají chromozomy kostřavy partnery – pokud se chtějí párovat, musejí se spokojit s chromozomy jílku, což také dělají. U tetraploidních kříženců je však situace jiná. Zde má každý jílkový i kostřavový chromozom partnera (svůj homolog). Překvapivě ani zde si však chromozomy nevybírají a berou zavděk čímkoli. Dochází tak k homeologním párování, které vede k početným výměnám mezi chromozomy pocházejícími z obou rodičů. Studium distribuce homeologních rekombinací prokázalo, že je v podstatě možné jakýkoli úsek z genomu kostřavy přenést do genomu jílku. Jde o velmi důležité zjištění pro šlechtitele, kteří tak mohou vnést určitý agronomicky významný gen či geny z kostřavy do jílku. Po takové introgresi by však měla nastat stabilizace genomového složení hybridů, aby se tento znak v následujících generacích neztratil vlivem dalších homeologních rekombinací. U alopolyploidních odrůd jsou homeologní rekombinace nežádoucí již od samého začátku. U diploidních jílků ani kostřavy luční se však žádný regulační systém podobný pšeničnému *Ph* nevyskytuje.

Plané druhy jako nový příslib stabilních kříženců

Řešení pro stabilizaci genomového složení Festulolií by mohly představovat planě rostoucí druhy příbuzné agronomicky významným kostřavám. Jak bylo uvedeno výše, hexaploidní kostřava rákosovitá a některé další polyploidní druhy kostřav disponují systémem, který nedovoluje párování chromozomů pocházejících ze dvou různých druhů. Zdá se, že se tento systém vyvinul pouze jednou a do polyploidních druhů se dostal při jejich vzniku mezidruhovou hybridizací. Od systému, který se nachází v pšenici, se liší. Jeho funkčnost je totiž podmíněna přítomností dvou kopií, na rozdíl od pšeničného *Ph*, kde stačí pouze jedna kopie. Nutno podotknout, že gen nebo geny podmiňující tento systém nebyly dosud u kostřav identifikovány.

Největší potenciál představují plané druhy blízce příbuzné kostřavě luční a k. rákosovité. Jedním z nich je kostřava apeninská (*F. apennina*), která se vyskytuje ve vyšších polohách Alp, Apenin a Karpat (obr. 8 a 10). Tento tetraploidní druh se někdy uvádí jako poddruh – kostřava luční italská (*F. pratensis* subsp. *apennina*). Kostřava luční obývá nižší polohy do ca 1 800 m n. m., zatímco k. apeninskou pod 1 200 m n. m. nenacházíme (obr. 9). Z toho je patrné, že oba druhy či poddruhy rostou sympatricky v poměrně širokém pásu. Molekulární a cytogenetické experimenty odhalily, že kostřava apeninská velmi pravděpodobně vznikla křížením k. luční a dalšího dosud neurčeného druhu. Neznámý rodič však zřejmě nesl regulační systém, který se vyskytuje v kostřavě apeninské, na rozdíl od k. luční. Tento neidentifikovaný druh se nejspíš podílel i na evoluci dalších dvou tetraploidních kostřav – kostřavy atlasské (*F. mairei*) a *F. glaucescens* (druhu, který se společně s k. luční podílel na vzniku k. rákosovité). Příbuznost všech tří druhů s kostřavou luční, k. rákosovitou i oběma jílky dává naději na jejich úspěšnou implementaci do šlechtitelského procesu. Vedle systému regulujícího párování chromozomů v meióze mají tyto tři druhy další agronomicky významné vlastnosti – především schopnost přežít podmínky během dlouhého období mrazu a sucha. Nové mezirodové křížení již bylo provedeno a příští roky ukáží, zda použití těchto druhů povede k vytvoření odrůd trav vhodných pro měnící se klimatické podmínky. Šlechtění založené na uplatnění nejnovějších poznatků genetiky má – na rozdíl od evoluce – šanci urychlit proces vývoje nových odrůd tak, aby hospodářsky využitelné trávy vznikaly souběžně s měnícím se klimatem a ne pomalým tempem přirozeného výběru v reakci na tyto změny.

Výzkumné projekty podpořila Grantová agentura České republiky (P501/11/0504), Národní program udržitelnosti I (LO1204), Interní grantová agentura Univerzity Palackého v Olomouci (Prf/2012/001) a projekt Sciex-NMSch (14.099).

Doporučená literatura a vysvětlivky termínů uvedeny na webu Živy.

4.2 PUBLISHED ABSTRACTS

4.2.1 CHROMOSOME GENOMICS: NEW MILESTONE IN GENOME ANALYSIS OF FORAGE AND TURF GRASSES

David Kopecký, Jarmila Číhalíková, Jan Vrána, Jitka Kopecká, **Štěpán Stočes**, Miroslava Havránková, Jan Bartoš, Jan Šafář, Hana Šimková, Jaroslav Doležel

Grasses are among the most important and widely cultivated plants. On arable land they are grown for grassland and silage production. Meadow fescue (Festuca pratensis Huds.) is one of the predominant grass species especially in temperate and Northern regions because of its ability to survive under freezing conditions. Moreover, it has been used widely in intergeneric hybridization with various ryegrass species, and several superior Festulolium cultivars have been produced. Because of the large genome (1C=3175Mbp) and high proportion of repetitive DNA, genomic studies are difficult. In this work, we describe a possibility to dissect meadow fescue genome to smaller parts – individual chromosomes and groups of chromosomes. Following the chromosome sorting strategy originally developed for legumes and cereals we have developed a procedure for grasses and currently we are able to sort *F. pratensis* chromosome 4 and two groups of chromosomes: chromosomes 2, 3 and 7 and chromosomes 1, 5 and 6.As the first step we will sequence chromosome 4 by Illumina with the expected 20x coverage and assemble low copy and genic regions. This will facilitate detailed comparative analysis using sequenced genomes of rice, Brachypodium and sorghum and provide the first insight into the genome composition of this species. The possibility to purify chromosome 4 opens the way for more efficient analysis of genomic loci on this chromosome underling important traits such as freezing tolerance. This work has been supported by the Czech Science Foundation (grant award P501/11/0504) 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. CZ.1.05/2.1.00/01.0007).

Keywords: fescue, chromosome sorting, flow cytometry, grasses, sequencing

4.2.2 GENE EXPRESSION ANALYSIS IN FESTULOLIUM HYBRIDS USING RNASEQ

Stepan Stoces, Tom Ruttink, David Kopecky, Jan Bartos and Jaroslav Dolezel

The advent of Next Generation Sequencing permitted sequencing whole genomes and transcriptomes with a depth and resolution never achieved before. One of the exciting opportunities has been the analysis of gene expression in wide hybrids. The hybridization brings homo(eo)logous genes of parental genomes under one roof and these pairs of genes may evolve according to different scenarios during the first generations after hybrid creation. First, one gene copy may become non-functional by genetic and/or epigenetic changes (non-functionalization). The other option is that one copy may acquire a novel, usually beneficial function and will be preserved by natural selection with the other copy retaining the original function (neo-functionalization). The last scenario for duplicated genes is that both copies become partially compromised by mutation accumulation to the point where their total capacity has been reduced to the level of the single-copy ancestral gene (sub-functionalization). The aim of our project is to study processes described above in initial generations (F1-F3) of hybrids of ryegrasses (*Lolium*) and fescues (*Festuca*) – Festuloliums. We set out to develop SNP platform based on RNAseq data to reconstruct genomic composition of Festulolium hybrids and to analyze the expression of parental genes in hybrid progenies. We first performed Illumina RNAseq on six *F. pratensis* and six *L. multiflorum* parental genotypes and mapped sequence reads from each genotype separately onto a common transcriptome reference sequence of 19,345 *L. perenne* genes. This led to identification of 600,000 to 900,000 putative SNPs per genotype. In each pair-wise *F. pratensis* x *L. multiflorum* combination, we selected SNPs that were homozygous within each genotype, and polymorphic between species. This yielded 30,000 to 90,000 species-specific SNPs, depending on the sequence depth and particular pair-wise combination of *F. pratensis* × *L. multiflorum* genotypes. We found that the SNPs were localized in 6,000 to 9,000 genes distributed more or less evenly across the genome. The next step is to analyze RNAseq data from reciprocal F1, F2 and F3 hybrids, to quantify gene expression from the respective parental genomes and to assess the putative loss and silencing of parental alleles in hybrid progenies.

This work was supported by grant awards P501/11/0504 and CZ.1.05/2.1.00/01.0007.

4.2.3 GENE EXPRESSION CHANGES IN INTERGENERIC *FESTUCA* X *LOLIUM* HYBRIDS (XFESTULOLIUM)

Stepan Stoces, Tom Ruttink, David Kopecky, Michael Abrouk, Jan Bartoš, Jaroslav Dolezel

Polyploidy has been recognized a major force in the evolution of flowering plants. Allopolyploidy involves a merge of distant genomes followed by whole-genome duplication. Although recently formed allopolyploids retain duplicated copies of a majority of genes on homoeologous chromosomes, it appears that the contribution of parental genomes doesn't need to be equal and that altered gene expression is common. Such a divergence is believed to develop in two phases. The first takes place immediately after the hybrid formation, while the second is a gradual evolution of gene expression mediated by diversification of duplicated genes. We have sequenced transcriptomes of two grass species, *Festuca pratensis* and *Lolium multiflorum*. The sequences were used to identify species-specific SNPs, which were located in the parental genomes. Developed SNP platform was used to analyze gene expression in newly synthesized Festuca × Lolium hybrids. This provided novel data on genome interactions and silencing of genes belonging to the parental genomes.

This work was supported by grants P501/11/0504 and LO1204.

4.2.4 GRASS BREEDING MEETS GENOMICS

David Kopecky, Stepan Stoces, Bruno Studer

Forage and turf grasses are of major importance for agricultural production and provides valuable ecosystem services and its impact is likely to rise in changing climatic and socio-economic environments. Despite their economic and ecological importance, genomic resources available for these species are lagging far behind those of model and major crops species. Understanding the genome structure and function of grassland species, however, provides opportunities to accelerate crop improvement and thus to mitigate the future challenges of increased feed and food demand, scarcity of natural resources such as water and nutrients, and high product qualities. In this presentation, a selection of technological developments providing new insights into the structure, evolution and function of plant genomes will be discussed. Many of these technologies were originally developed in human or animal science and are now increasingly applied in plant genomics. Main goal is to highlight the benefits of using these technologies for forage and turf grass genome research and breeding and to discuss their potentials and limitations.

This work has been supported by the Ministry of Education, Youth and Sports of the Czech Republic (grant No. LO1204 from the National Program of Sustainability I).

4.2.5 GENE EXPRESSION IN INTERSPECIFIC HYBRIDS

Štěpán Stočes, Steven Yates, Tom Ruttink, Jan Bartoš, Zbigniew Zwierzykowski, Michael Abrouk, Bruno Studer, Isabel Roldán-Ruiz, Tomasz Ksiazczyk, Elodie Rey, Jaroslav Doležel and David Kopecký

Interspecific hybrids result from the cross of two different, but usually closely related species and thus, carry two or even sometimes more subgenomes. Interspecific hybridization frequently followed by polyploidization play a key role in plant evolution. Moreover, artificial interspecific hybrids (such as Triticale, hybrid ryegrass and many others) made in breeding programs became popular among farmers and are frequently used in recent agriculture. However, only little is known about the mechanisms underlying interactions of two parental genomes present in hybrid individual.

In our project, we studied gene expression changes in reciprocal *Festuca pratensis* x *Lolium multiflorum* hybrids. We sequenced transcriptomes of both parents and using the Orthology Guided Assembly (OGA) strategy; we assembled and annotated 19,036 and 18,952 transcripts. Thereafter, we identified interspecific SNPs, which allowed discrimination of *L. multiflorum* and *F. pratensis* alleles in hybrids and were used for monitoring differential (DE) and allele specific (ASE) expression in parents and F1 and F2 hybrid generations.

The level of gene expression is highly similar in both parents for most of the genes and only several genes display different level of expression. Similarly, in most of the genes, there is an equal expression from both parental alleles in hybrids. Our data suggests that the ASE is heritable but is dependent upon the maternal plant. We have also discovered that *cis* elements are more prevalent then selection which manipulates *trans* elements. We speculate that the evolution favours small incremental changes per loci rather than manipulating *trans* elements whose effects would be broader and therefore less targeted.

4.3 PUBLISHED POSTERS

4.3.1 THE DEVELOPMENT OF AN SNP PLATFORM FOR RAPID SCREENING OF THE GENOMIC CONSTITUTION OF XFESTULOLIUM HYBRIDS AND GENE EXPRESSION ANALYSIS

Štěpán Stočes , Tom Ruttink, David Kopecký, Jan Bartoš and Jaroslav Doležel

Vienna 2013

The development of an SNP platform for rapid screening of the genomic constitution of xFestulolium hybrids and gene expression analysis

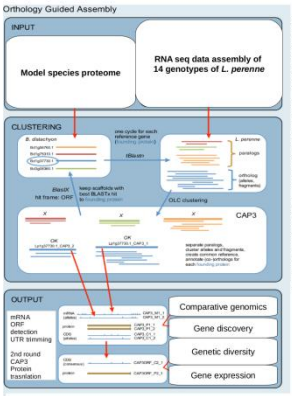
Štěpán Stočes[1], Tom Ruttink[2], David Kopecký[1], Jan Bartoš[1] and Jaroslav Doležel[1]

1) Centre of the Region Haná for Biotechnological and Agricultural Research, Institute of Experimental Botany, Šlechtitelů 31, CZ-783 71, Olomouc, Czech Republic
2) Institute for Agricultural and Fisheries Research, Caritasstraat 21, 9090, Melle, Belgium

Introduction

- xFestulolium is the superior grass developed by breeders using intergeneric cross of ryegrass and fescue species. It combines complementary agronomic attributes of both genera – high yield and nutrition of ryegrasses and tolerance to abiotic stresses of fescues.
- Despite the increasing popularity of these hybrids among seed companies and farmers, there is a lack of knowledge on the genomic constitution and gene expression. There are several methods, which have been used for discrimination of genome composition, but they generally suffer by low-throughput and limited resolution.
- On the other hand, Next Generation Sequencing technologies enables production of large sequence datasets, which can be used for the analysis of genome composition, molecular marker development, phylogenetic and ecological studies and analysis of transcriptomes using RNA-sequencing.
- The aim of our project is to study genomic constitution and gene expression in F1-F3 generations of xFestulolium hybrids using Illumina RNAseq technology.

B) Development of coding sequences reference



Since lack of fully annoted reference sequence of *L. multiflorum* (Lm) and *F. pratensis* (Fp), we used existing reference sequence of *L. perenne* (Lp) made from 14 genotypes of this species for *de novo* assembly of our RNA-seq data.

Lp reference sequence was developed using *Brachypodium distachyon* protein set for identification of orthologs and candidate genes. All contigs from 14 Lp genotypes with a significant similarity to the candidate *B.distachyon* proteins were selected and clustered with CAP3 software.

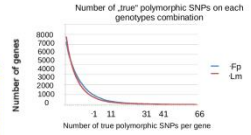
To distinguish between alleles and paralogs, it compared CAP3-contigs (BLASTx) against all *B.distachyon* proteins and retained only those with a best BLAST hit with the original candidate gene. Finally, were identified single orthologs and was constructed non-redundant reference transcriptome sequence.

Results

Out of the total number of 600,000 to 900,000 SNPs, we identified 30,000 to 90,000 highly species-specific SNPs based on the combination of *F. pratensis* × *L. multiflorum* genotypes. These SNPs are localized on 6000 to 9000 genes distributed more or less evenly over all chromosomes.

Tab.1 - Number of all and potentially polymorphic SNPs between reference sequence and our genotypes

Genotypes	total SNP	potential "P" SNP
Fp51	696 555	436867
Fp52	845 168	629 622
Fp53	587 351	452 110
Fp54	795 953	546 360
Fp55	814 781	572 569
Fp56	745 826	483 592
Lm51	638 082	313 567
Lm52	664 712	490 588
Lm53	900 556	696 801
Lm54	800 478	524 964
Lm55	758 922	469 496
Lm56	622 605	474 160
total	9 184 194	6 090 696

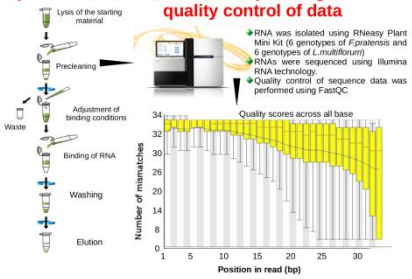


Tab.2 - Number of true polymorphic SNPs for the combinations of parental genotypes used for development of F1 hybrids

Genotypes	Fp51	Fp52	Fp53	Fp54	Fp55	Fp56
Lm51		37926				
Lm52			64397			
Lm53			127364			
Lm54						
Lm55						92726
Lm56						

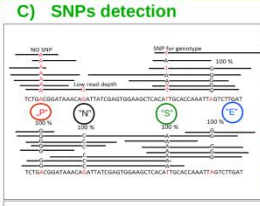
Bioinformatics pipeline

A) Isolation RNA, Illumina sequencing and quality control of data



- RNA was isolated using RNeasy Plant Mini Kit (6 genotypes of *F.pratensis* and 6 genotypes of *L.multiflorum*)
- RNAs were sequenced using Illumina RNA technology.
- Quality control of sequence data was performed using FastQC

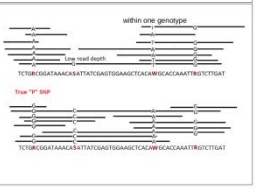
C) SNPs detection



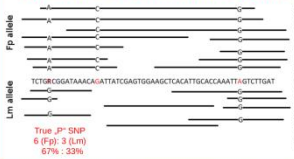
Our RNASeq data from 6 and 6 genotypes of parental species were trimmed and mapped on CDS reference. Thereafter, we detected all possible SNPs under low stringency conditions.

SNPs were cathegorized into the following groups:
"P" - polymorphic between Lm and Fp
"N" - not detected, low read depth
"S" - polymorphic within a species (either Lm or Fp)
"E" - nonpolymorphic between Fp and Lm, but polymorphic between samples and reference (Lp)

Reference was annotated with detected SNPs and transformed into the IUPAC code in location of these SNPs.
Reads from Fp and Lm were mapped on the reference including IUPAC codes and another round of SNPs detection was achieved. True polymorphic SNPs were selected for next steps. These true SNPs (1) include polymorphism between parental species, (2) are nonpolymorphic within one species and (3) have sufficient read depth.
IUPAC codes in reference sequence were changed to the original bases except true polymorphic SNPs.



D) Gene expression in F1 generations



Reads from F1 hybrids were mapped on the reference with the IUPAC codes and SNPs were detected under low stringency conditions. The ratio of parental variants (Lm and Fp) for each SNP was obtained. This provided the insight on the change of gene expression in F1 hybrids. The analysis of F2 and F3 hybrids are in progress.

Conclusions

- Bioinformatics pipeline for the study of gene expression in intergeneric grass hybrids was developed
- This pipeline was found as an excellent tool for the analysis of gene expression in intergeneric hybrids.
- It offers a versatile tool for the gene expression studies of various intergeneric and interspecific hybrids.

This work was supported by grants P501/11/0504 and CZ.1.05/2.1.00/01.0007 and IGA PrF/2012/001

5. DISCUSSION

This Ph.D. thesis deals with the analysis on the structure and function of interspecific grass hybrids. Despite three relatively distinct projects I worked on during my PhD study, all the work had two main characters: (1) interest in the increasing knowledge on the interspecific hybridization and (2) using bioinformatics tools for these studies. Interspecific hybridization is one of the main mechanisms of plant origination and some of the main crops including wheat and Brassicas are allopolyploids (Song *et al.,* 1995; Liu *et al.*, 1998; Ozkan *at al.*, 2001). Besides ancient allopolyploids, hybridization of plants from distinct species is frequently used in plant breeding to introgress specific traits from wild relative to elite crop cultivar or to merge two species into a single organism to combine beneficial characters from both parental species. This is a case of Festulolium, artificial hybrid of fescues and ryegresses (reviewed in Ghesquiere *et al.*, 2010).

5.1 IDENTIFICATION OF DArT MARKERS ASSOCIATED WITH FREEZING TOLERANCE

One of the scientific projects I was involved in was the usage of Diversity Arrays Technology (DArT) markers for genetic mapping of Italian ryegrass (Lolium multiflorum) and meadow fescue (Festuca pratensis) and a case study on the usage of DArT markers in breeding. I was responsible for the identification of DArT markers which could be linked with freezing tolerance in grasses. I performed a marker-trait association study in Festulolium plants diverged in freezing tolerance. I found DArT markers associated with the freezing tolerance phenotype on chromosomes 2, 4 and 7. This observation corresponds with the results of Kosmala *et al.* (2006). They used GISH for analysis of Festulolium hybrid plants and found that plants carrying a large pericentromeric region of *F. pratensis* chromosome 4 exhibited increased freezing tolerance. Another DArT marker associated with freezing tolerance mapped to central part of *L. multiflorum* chromosome 2. Similarly to chromosome 4, chromosome 2 was also identified to carry genes associated with freezing tolerance. Kosmala *et al.* (2006) identified by GISH Festulolium plants with terminal *F. pratensis* segment on the short arm of *L. multiflorum* chromosome 2 with increased frost tolerance. Moreover, Shinozuka *et al.* (2006) mapped a glycine-rich RNA binding protein putatively associated with freezing tolerance to the long arm of *L. perenne* chromosome 2.

5.2 IDENTIFICATION OF REPEATS AS A SOURCE OF NEW CYTOGENETIC MARKERS

Festulolium hybrids are highly variable in their genomic composition and even the plants from the same cultivar differ significantly in the genomic constitution and proportion of parental genomes (Kopecký *et al.*, 2006). This is a consequence of out-crossing type of reproduction and pairing of parental chromosomes (those from Festuca with those from Lolium) and frequent homeologous recombination. Hence, it is highly beneficial to know the genomic composition of the plants used in breeding process. One of the simplest ways is to employ genomic in situ hybridization (GISH) and "paint" parental chromatin in the squashed metaphase spreads of hybrid plants (Thomas *et al.*, 1994). However, it will not provide the information on the chromosomes present in the individual hybrid plants. Thus, the aim of our laboratory was to develop molecular cytogenetic markers, which could enable precise karyotyping and identification of individual chromosomes. Using Illumina sequence reads from flow-sorted chromosome 4F of meadow fescue (Kopecký *et al.*, 2013), I was able to identify tandem repeats (satellites), which provided unequivocal identification of all seven chromosomes of meadow fescue (Křivánková *et al.*, 2017). Tandem repeats are generally valuable source of useful cytogenetic markers in plant species. In wheat and barley, satellites of the *Afa* family have been used for identification of the D-genome chromosomes (Rayburn and Gill, 1986; Tsujimoto *et al.*, 1997). Similarly, various satellites were employed 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). Besides tandem repeats, we analyzed mobile elements in meadow fescue genome. Most of the mobile DNA elements were dispersed along all fescue chromosomes. This agrees with previous reports in other plant species (Wicker *et al.*, 2003; Hřibová *et al.*, 2010; Sergeeva *et al.*, 2010; Akiyama *et al.*, 2012). The strongest hybridization signals were obtained for the Athila retroelement (CL6) from the Ty3/gypsy family, indicating that it is the most abundant DNA repeat in the meadow fescue genome.

5.3 DEVELOPMENT OF ROBUST BIOINFORMATICS MEHTODS FOR IDENTIFYING AND SORTING OF SNPs

Despite the increasing interest of grass breeders and farmers in Festulolium and number of studies on the structure and evolution of Festulolium genomes (reviewed in

Kopecký *et al.*, 2007), any information on the expression of parental alleles in hybrid plants is lacking. Main aim of my Ph.D. thesis was to develop a platform for the analysis of gene expression in grass hybrids and the pilot study on the gene expression of first generations of reciprocal xFestulolium hybrids. This task was challenging due to an absence of the reference sequence of any of the parental species used for the development of L. multiflorum x F. pratensis hybrids. I used Illumina RNA-sequencing and Orthology Guided Assembly (OGA; Ruttink *et al.*, 2013) to develop robust single nucleotide polymorphisms (SNP) based platform for hybrid gene expression studies. I identified 13,260 SNPs, which can differentiate parental alleles of about 3 611 genes in Festulolium hybrids (Stočes *et al.*, 2016). To facilitate large-scale screening of thousands to tens of thousands of plants in grass breeding programs, a high-throughput genotyping platform needs to be developed with a set of highly informative markers. Array-based genotyping platforms were developed for ryegrasses, and used to construct genetic maps (Studer *et al.* 2012), identify Italian and perennial ryegrasses cultivars (Wang *et al.* 2014), and study genetic diversity of *L. perenne* ecotypes (Blackmore *et al.*, 2015). The main advantage of our approach compare to the above mentioned studies is that our SNP sets contain complete positional information (putative genome location, neighboring SNPs) on three different types of SNPs suitable for development of novel probe sets for genes spread across the genome.

The possibility to produce large amount of transcriptomic sequence data from RNAseq enables the more accurate analysis of gene expression on genome-wide basis. We focused on the gene expression analysis from the newly formed allopolyploids with known parental plants and having the potential to be used in breeding programs. For that, first two generations of *Festuca* × *Lolium* hybrids were used to evaluate changes in gene expression following genome merge. Our parental-alleles specific gene expression analysis revealed that the expression of homeologs in the hybrids is inherited from their parents in most of the genes (73.8% and 77.7% of the genes in *Festuca* × *Lolium* and *Lolium* × *Festuca* hybrids, respectively). This percentage is in line with previously published studies (Wu *et al.*, 2016), reporting that 79% of such genes in triploid rice hybrids. It is slightly higher compared to the study of (Yoo *et al.*, 2013) that reported that 59.4-70.9% of the genes maintained the ratio between the parental specific gene expression levels in the F1 interspecific hybrids and allopolyploids of cotton. We further detected that over 20% of the analyzed genes showed novel bias in hybrids. This

bias was mostly caused by the overexpression of *Lolium* homeolog when the expression was at the same level in the parents (19.2% and 18.2% in *Festuca* × *Lolium* and *Lolium* × *Festuca* hybrids, respectively). The overexpression of *Festuca* homeolog in hybrids was much less frequent (5.4% and 3.2% in *Festuca* × *Lolium* and *Lolium* × *Festuca* hybrids, respectively). Such *Lolium* genome dominance in Festulolium hybrids revealed here on transcriptomic level could be responsible for the shift in genomic composition towards *Lolium* genome detected in hybrids. Zwierzykowski *et al.* (2006) evidenced the progressive replacement of *Festuca* chromosomes by *Lolium* chromosomes in successive generations of hybrids. Gene expression studies reported genome dominance (even though with lower frequencies) also in other allopolyploids, such as *Glycine max* (Ilut *et al.*, 2012).

6. GENERAL CONCLUSIONS

The aims of this thesis are consistent with long-term focus of the Centre of Plant Structural and Functional Genomics: to apply results from fundamental research in the practical breeding. First aim of the thesis was to identify markers associated with freezing tolerance, which could significantly speed up the process of selection during breeding process was the second aim. Several tens of markers, which differ in their occurrence between populations with high frost tolerance and low frost tolerance, have been identified and some of them positioned on linkage groups. The identification of markers which can be used for karyotyping and thus, for the screening of genomic composition of Festuca x Lolium hybrids, was the second aim of the thesis. Several tandem repeats from flow sorted and sequenced 4F chromosome have been proved as robust cytogenetic markers, which enable parental chromosome identification in interspecific hybrids.

The last, but not least aim of the thesis - analysis of gene expression in Festuca x Lolium hybrids using the most recent methods like next generation sequencing (NGS) increased our knowledge on the evolution and function of hybrid genomes. We described reconstruction of transcriptomes of meadow fescue and Italian ryegrass and their hybrids and identification of SNPs, which discriminate parental alleles in hybrids and thus, allowed tracking changes in gene expression in F1 and successive generations of hybrids.

The thesis significantly increased our knowledge about the structure, evolution and function of allopolyploid and hybrid genomes, and of plant genomes in general. The results are also applicable in breeding of forage and turf grasses.

7. REFERENCES

1. Abbott R, Albach D, Ansell S, Arntzen JW, Baird SJE et al., (2013) Hybridization and speciation. Journal of Evolutionary Biology **26**, 229 - 246.
2. Adams KL, Wendel JF (2005) Polyploidy and genome evolution in plants. Current Opinion in Plnat Biology **8**, 135-141.
3. Akiyama Y, Kimura K, Yamada-Akiyama H, Kubota A, Takahara Y, et al: (2012) Genomic characteristics of a diploid F_4Festulolium hybrid (*Lolium multiflorum* × *Festuca arundinacea*). Genome **55**:599-603
4. Alberts B, Johnson A, Lewis J, *et al.* 2002 Molecular Biology of the Cell. 4th edition. New York: Garland Science. Available from: http://www.ncbi.nlm.nih.gov/books/NBK21054/
5. Altshuler D, Pollara VJ, Cowles CR, Van Etten WJ, Baldwin J, Linton L, Lander ES (2000) An SNP map of the human genome generated by reduced representation shotgun sequencing. Nature **407** (6803): 513–6.
6. Amábile-Cuevas CF, Chicurel ME (1993) Horizontal Gene Transfer. American Scientist **81**(4), 332–341.
7. Barton NH, Gale KH (1993) Genetic analysis of hybrid zones. In: Harrison RG, editor. Hybrid zones and the evolutionary process. New York: Oxford University Press. p. 13–45.
8. Bartoš J, Sandve RS, Kölliker R, Kopecký D, Christelová P, Stočes Š, Østrem L, Larsen A, Killian A, Rognli AO, Doležel J (2011) Genetic mapping of DArT markers in the *Festuca–Lolium* complex and their use in freezing tolerance association analysis. Theor. Appl. Genet. **122**(6),1133-1147.
9. Beard JB, Green RL (1993) The Role of Turfgrasses in Environmental Protection and Their Benefits to Humans. Journal of Environmental Quality **23** (3), 452-460.
10. Becker J, Vos P, Kuiper M, Salamini F, Heun M (1995) Combined mapping of AFLP and RFLP markers in barley. Mol Gen Genet **249**, 65–73.
11. Bennett M, Leitch I (2005) Nuclear DNA amounts in angiosperms: Progress, problems and prospects. Annals of Botany **95**, 45-90.
12. Bennett S (2004) Solexa Ltd. Pharmacogenomics **5**(4), 433-438.

13. Berthouly-Salazar C, Mariac C, Couderc M, Pouzadoux J, Floc'h JB, Vigouroux Y (2016) Genotyping-by-Sequencing SNP Identification for Crops without a Reference Genome: Using Transcriptome Based Mapping as an Alternative Strategy. Front. Plant Sci. **7**, 777.
14. Bingham PM, Chou TB, Mims I, Zachar Z (2003) On/off regulation of gene expression at the level of splicing. Trends in Genetics **4**(5), 134-138.
15. Birrer M, Kolliker R, Manzanares C, Asp T, Studer B (2014) A DNA marker assay based on high-resolution melting curve analysis for distinguishing species of the *Festuca – Lolium* complex. Mol Breeding **34**, 421–429
16. Black DL (2003) Mechanisms of Alternative Pre-Messenger RNA Splicing Annual Review of Biochemistry **72**, 291-336.
17. Blackmore T, Thomas I, McMahon R, Powell W, Hegarty M (2015) Genetic–geographic correlation revealed across a broad European ecotypic sample of perennial ryegrass (*Lolium perenne*) using array-based SNP genotyping. Theor. Appl. Genet. **128**:1917–1932
18. Blanca J, Esteras C, Ziarsolo P, Perez D, FernAndez-Pedrosa V, Collado C, RodrAguez De Pablos R, Ballester A, Roig C, Canizares J, *et al* (2012) Transcriptome sequencing for SNP discovery across Cucumis melo. BMC Genomics. **13**(1), 280.
19. Botstein D, White RL, Skolnick M, Davis RW (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. Am. J. Hum. Genet. **32**, 314 – 331.
20. Borrill M (1975) *Festuca* L. In: C.A. Stace (Ed.), Hybridization and the Flora of the British Isles, pp. 543–547. Academic Press, London.
21. BradShaw AD, Hardwick K (1989) Evolution and stress-genotypic and phenotypic components. Biological Journal of the Linnean Society **37**, 137-155.
22. Brammer SP, Vasconcelos S, Poersch LB, Oliveira AR, Brasileiro-Vidal AC (2013) Genomic *in situ* Hybridization in *Triticeae*: A Methodological Approach, Plant Breeding from Laboratories to Fields, Prof. Sven Bode Andersen (Ed.), Available from: http://www.intechopen.com/books/plant-breeding-from-laboratories-to-fields/genomic-in-situ-hybridization-in-triticeae-a-methodological-approach

23. Brandes A, Röder MS, Ganal MW (1995) Barley telomeres are associated with two different types of satellite DNA sequences. Chromosome Res. 1995 Aug;**3**(5):315-20.
24. Brodsky AS, Silver PA (2000) Pre-mRNA processing factors are required for nuclear export. RNA **6**, 1737-1749.
25. Buckner RC, Boling JA, Burrus II PB, Bush LP, Hemken RA (1983) Registration of "Johnstone" tall fescue. Crop. Sci. **23,** 399–400.
26. Buckner RC, Burrus II PB, Bush LP (1977) Registration of "Kenhy" tall fescue. Crop Sci. **17,** 672–673.
27. Bullard J, Purdom E, Hansen K, et al. (2010) Evaluation of statistical methods for normalization and differential expression in mRNA-Seq experiments. BMC Bioinformatics **11**, 94.
28. Campbell N (1996) Biology; Fourth edition. The Benjamin/Cummings Publishing Company. p. 309 – 310.
29. Cardon LR and Bell JI (2001) Association study designs for complex diseases. Nat. Rev. Genet. **2**, 91–99.
30. Cerutti JM, Riggins GJ, Souza SJ (2003) What can digital transcript profiling reveal about human cancers? Brazilian Journal of Medical and Biological Research **36**(8), 975.
31. Collins LJ (2011) THE RNA INFRASTRUCTURE: An Introduction to ncRNA Networks. In Rna Infrastructure and Networks, L.J. Collins, ed, pp. 1-19.
32. Comai L, *et al.* (2000) Phenotypic instability and rapid gene silencing in newly formed *Arabidopsis* allotetraploids. Plan Cell **12**, 1551-1568.
33. Comai L (2000) Genetic and epigenetic interactions in allopolyploid plants. Plant Mol. Biol. **43**, 387–399
34. Comai, L, Madlung A, Josefsson C, Tyagi A (2003) Do the different parental 'heteromes' cause genomic shock in newly formed allopolyploids? Philos. Trans. R. Soc. Lond. B Biol. Sci. **358**: 1149–1155.
35. Conaway RC, Brower CS, Conaway JW (2002) Emerging roles of ubiquitin in transcription regulation. Science **296,** 1254-1258.
36. Cornish MA, Hayward MD, Lawrence MJ (1979) Self-incompatibility in ryegrass. I. Genetic control in diploid *Lolium perenne*. Heredity **43**, 95—106.

37. Coyne JA, Orr HA (2004) Speciation. MA: Sinauer Associates ISBN: 978-0-87893-089-0.
38. Davey JW, Hohenlohe PA, Etter PD, Boone JQ, Catchen JM, Blaxter ML (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. Nature Reviews Genetics **12**, 499-510.
39. Delourme R, Falentin C, Fomeju B, Boillot M, Lassalle G, André I, Duarte J, Gauthier V, Lucante N, Marty A, *et al* (2013) High-density SNP-based genetic map development and linkage disequilibrium assessment in Brassica napus L. BMC Genomics. **14**(1): 1-18.
40. DeRisi J, Penland L, Brown PO, Bitter ML, Meltzer PS, Ray M, Chen Y, Su YA, Trent JM (1996) Use of a cDNA microarray to analyse gene expression patterns in human cancer. Nature Genetics **14**, 457-460.
41. Dhanasekaran S, Doherty TM, Kenneth J, TB Trials Study Group (2010) Comparison of different standards for real-time PCR-based absolute quantification. Immunol Methods. **354** (1–2): 34–9.
42. Di Y, Schafer DW, Cumbie JS, et al. (2011) The NBP negative binomial model for assessing differential gene expression from RNA-seq. Stat Appl Genet Mol Biol **10**, 24.
43. Dierking, R., Azhaguval, P., Kallenbach, R., Saha, M. C., Bouton, J., & Chekhovskiy, K., Andrew. (2015). Linkage maps of a mediterranean × continental tall fescue population and their comparative analysis with other poaceae species. The Plant Genome **8**, 1-18.
44. Dohm JC, Lottaz C, Borodina T, et al. (2008) Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. Nucleic Acids Res **36**,e105.
45. Drabovich AP, Krylov SN (2006) Identification of base pairs in single-nucleotide polymorphisms by MutS protein-mediated capillary electrophoresis. Analytical Chemistry **78** (6), 2035–8.
46. Elgin SC (1996) Heterochromatin and gene regulation in Drosophila. Current Opinion in Genetics & Development **6(2)**, 193–202.
47. Ellis RJ (1991) Molecular Chaperones. Annual Review of Biochemistry **60**, 321-347.

48. Ergon A, Fang C, Jorgensen O, *et al.* (2006) Quantitative trait loci controlling vernalisation requirement, heading time and number of panicles in meadow fescue (*Festuca pratensis* Huds.). Theoretical and Applied Genetics **112**, 232–242.
49. Ersoz ES, Wright MH, Pangilinan JL, Sheehan MJ, Tobias C, Casler MD, Buckler ES, Costich DE (2012) SNP Discovery with EST and NextGen Sequencing in Switchgrass (*Panicum virgatum* L.). PLoS One. **7**(9), e44112-10.1371/journal.pone.0044112.
50. Etter OD, Basshan S, Hohenlohe PA, Johnoson EA, Cresko WA (2011) SNP discovery and genotyping for evolutionary genetics using RAD sequencing. Methods Mol. Biol. **772**, 157-178.
51. Fay JC, Wittkopp PJ (2008) Evaluating the role of natural selection in the evolution of gene regulation. Heredity **100**(2),191-199
52. Fawcett JA, Maere S, Van de Peer Y (2009). From the Cover: Plants with double genomes might have had a better chance to survive the Cretaceous-Tertiary extinction event. Proceedings of the National Academy of Sciences **106**, 5737-5742.
53. Feldman M, Levy AA (2005) Allopolyploidy--a shaping force in the evolution of wheat genomes. Cytogenetic and genome research, **109**(1-3), 250-258.
54. Feschotte C, Jiang N, Wessler SR (2002) Plant transposable elements: where genetics meets genomics. Nat Rev Genet. **3**,329–341.
55. Finotello F, Camillo DB (2014) Measuring differential gene expression with RNA-seq: challenges and strategies for data analysis. Briefings in Functional Genomics doi:10.1093/bfgp/elu035.
56. Flagel L1, Udall J, Nettleton D, Wendel J (2008) Duplicate gene expression in allopolyploid Gossypium reveals two temporally distinct phases of expression evolution. BMC Biol. **16**(6), 16.
57. Flicek P, Birney E (2009) Sense from sequence reads: methods for alignment and assembly. Nat Methods **6** Suppl. **11**, S6-S12.
58. Flint-Garcia SA, Thornsberry JM and Buckler ES (2003) Structure of linkage disequilibrium in plants. Annu. Rev. Plant Biol. **54**, 357–374.
59. Force A, Lynch M, Pickett FB, Amores A, Yan YL, Postlethwait J (1999) Preservation of duplicate genes by complementary, degenerative mutations. Genetics **151**(4), 1531–1545.

60. Fujimoto SY, Ohta M, Usui A, Shinshi H, Ohme-Takagi M (2000) Arabidopsis Ethylene-Responsive Element Binding Factors Act as Transcriptional Activators or Repressors of GCC Box–Mediated Gene Expression. The Plant Cell **12**, 393–404.
61. Ghesquiere M, Humphreys MW, Zwierzykowski Z (2010) Festulolium. *In:* Boller B, Posselt UK, Veronesi F (eds.) Fodder Crops and Amenity Grasses, Book Series: Handbook of Plant Breeding **5,** 293-316. Springer Science+Business Media.
62. Ghildiyal M, Zamore PD (2009) Small silencing RNAs: an expanding universe. Nature Reviews Genetics **10**, 94-108.
63. Goszczynski DE, Jooste AEC (2002) The application of single-strand conformation polymorphism (SSCP) technique for the analysis of molecular heterogeneity of grapevine virus A. Vitis **41**, 77– 82.
64. Grant V. (1981) 2nd edn. Columbia University Press; New York, NY. Plant speciation.
65. Griffiths AJF, Miller JH, Suzuki DT, et al. (2000) An Introduction to Genetic Analysis. 7th edition. New York: W. H. Freeman. Sources of variation. Available from: http://www.ncbi.nlm.nih.gov/books/NBK22012/
66. Grodzicker T, Williams J, Sharp P, Sambrook J (1975) Physical mapping of temperature sensitive mutants of adenovirus. Cold Spring Harbor Symp Quant Biol **39**,439-446
67. Gruber JD, Vogel K, Kalay G, Wittkopp PJ (2012) Contrasting properties of gene-specific regulatory, coding, and copy number mutations in *Saccharomyces cerevisiae*: frequency, effects, and dominance. PLoS Genet. **8**(2):e1002497.
68. GuhaMajumdar M, Dawson-Baglien E, Sears BB (2014) Creation of a Chloroplast Microsatellite Reporter for Detection of Replication Slippage in *Chlamydomonas reinhardtii*. DNA Res **21**(4), 417-427.
69. Guthridge KM, Dupal MD, Kolliker R, Jones ES, Smith KF, Forster JW (2001) AFLP analysis of genetic diversity within and between populations of perennial ryegrass (*Lolium perenne* L.). Euphytica **122**, 191-201.
70. Hall TM (2005) Structure and function of Argonaute proteins. Cell. **13**(10), 1403–1408.
71. Hatem A, Bozda D, Toland AE, et al. (2013) Benchmarking short sequence mapping tools. BMC Bioinformatics **14**, 184.

72. Helentjaris T, Slocum M, Wright S, Schaefer A, Nienhuis J (1986) Construction of genetic linkage maps in maize and tomato using restriction fragment length polymorphisms. Theor. Appl. Genet. **61**, 650 – 658.
73. Hirschhorn NJ, Daly JM (2005) Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics **6**, 95-108.
74. Hobert O (2008) Gene Regulation by Transcription Factors and MicroRNAs. Science **319**, 1785.
75. Holcik M, Sonenberg N (2005) Translation control in stress and apoptosis. Nature Reviews Molecular Cell Biology **6**, 318-327)
76. Hřibová, E., Doleželová, M., Town, C.D., Macas, J., Doležel, J (2007) Isolation and characterization of the highly repeated fraction of the banana genome. CYTOGENETIC AND GENOME RESEARCH **119** [3-4], 268-274
77. Hřibová, E., Neumann, P., Matsumoto, T., Roux, N., Macas, J., Doležel, J (2010) Repetitive part of the banana (Musa acuminata) genome investigated by low-depth 454 sequencing. BMC PLANT BIOLOGY **10**: 204
78. Humphreys MO (2005) Genetic improvement of forage crops – past, present and future. Journal of Agricultural Science **143**, 441–448.
79. Humphreys J, Harper JA, Armstead IP, Humphreys M.W. (2005) Introgression-mapping of genes for drought resistance transferred from Festuca arundinacea var. glaucescens into Lolium multiflorum. Theor Appl Genet **110**:579-587.
80. Chakravarti A (1999) Population genetics-making sense out of sequence. Nature Genetics **21**, 56–60.
81. Chalhoub B, Denoeud F, Liu S, Perkin IAP, Tang H, *et al.,* (2014) Early allopolyploid evolution in the post-Neolithic Brassica napus oilseed genome. Science **345**, 950-953.
82. Charmet G, Ravel C, Balfourier F (1997) Phylogenetic analysis in the *Festuca-Lolium* complex using molecular and ITS rDNA. Theor. Appl. Genet. **94**, 1038-1046.
83. Cheng X, Blumenthal RM. (2010) Coordinated chromatin control: structural and functional linkage of DNA and histone methylation. Biochemistry **49**:2999-3008.
84. Chen Z, Yang H, Pavletich NP (2008) Mechanism of homologous recombination from the RecA-ssDNA/dsDNA structures. Nature **453,** 489-484.

85. Chen ZJ (2007) Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. Annu Rev Plant Biol. **58**, 377-406.
86. Choi I-Y, Hyten DL, Matukumalli LK, Song Q, Chaky JM, Quigley CV, Chase K, Lark KG, Reiter RS, Yoon M-S, *et al* (2007) A soybean transcript map: gene distribution, haplotype and single-nucleotide polymorphism analysis. Genetics. **176**(1), 685-696.
87. Ilut D.C., Coate J.E., Luciano A.K., Owens T.G., May G.D., Farmer A.D., Doyle J.J. (2012). A comparative transcriptomic study of an allotetraploid and its diploid progenitors illustrates the unique advantages and challenges of RNA-seq in plant species. Am. J. Bot. **99**: 383–396
88. Jaccoud D, Peng K, Feinstein D, Kilian A (2001) Diversity arrays: a solid state technology for sequence information independent genotyping. Nucleic Acids Res. 15;**29**(4),E25
89. Jauhar PP (1993) Cytogenetics of the *Festuca-Lolium* complex,relevance to reeding, monographs on theoretical and applied genetics, vol 18. Springer, Berlin Heidelberg New York, p. 255
90. Jenuwein T, Laible G, Dorn R, Reuter G (1998). SET domain proteins modulate chromatin domains in eu- and heterochromatin. Cell. Mol. Life Sci. **54** (1), 80–93.
91. Jiang H, Wong WH (2009) Statistical inferences for isoform expression in RNA-Seq. Bioinformatics **25**, 1026-32.
92. Jiang N, Bao Z, Zhang X, Hirochika H, Eddy SR, et al. An active DNA transposon family in rice. Nature **421**,163–167.
93. Jones CJ, Edwards KJ, Castaglione S, Winfield MO, Sala F, van de Wiel C, et al. (1997) Reproducibility testing of RAPD, AFLP, and SSR markers in plants by a network of European laboratories. Mol. Breed. **3**, 381 – 390.
94. Jones E, Chu W-C, Ayele M, Ho J, Bruggeman E, Yourstone K, Rafalski A, Smith O, McMullen M, Bezawada C, *et al* (2009) Development of single nucleotide polymorphism (SNP) markers for use in commercial maize (*Zea mays L.*) germplasm. Mol Breeding. **24**(2), 165-176.
95. Katsiotis A, Hagidimitriou M, Heslop-Harrison JS (1997) The close relationship between the A and B genomes in *Avena* L. (Poaceae) determined by molecular cytogenetic analysis of total genomic, tandemly and dispersed repetitive DNA sequences. Annals of Botany **79**: 103-109.

96. Khlestkina EK, Salina EA (2006) SNP markers: Methods of analysis, ways of development, and comparison on an example of common wheat. Russian Journal of Genetics **42**(6), 585-594.
97. Korf I. (2013) Genomics: the state of the art in RNA-seq analysis. Nat Methods **10**,1165-6.
98. Kopecký D, Zwierzykowski Z, Doležel J (2005) Genomic constitution of Festulolium cultivars released in the Czech Republic. Plant Breeding **124**, 454-458.
99. Kopecký D, Loureiro J, Zwierzykowski Z, Ghesquière M, Doležel J (2006) Genome constitution and evolution in *Lolium* × *Festuca* hybrid cultivars (Festulolium). Theor Appl Genet **13**, 731–742.
100. Kopecký D, Lukaszewski AJ, Doležel J (2008) Cytogenetics of Festulolium (*Festuca* X *Lolium* hybrids). Cytogenet Genome Res **120**, 370-383.
101. Kopecký D, Bartoš J, Lukaszewski AJ, Baird JH, Černoch V, Kölliker R, Rognli AO, Blois H, Caig V, Lübberstedt T, Studer B, Shaw P, Doležel J, Kilian A (2009) Development and mapping of DArT markers within the *Festuca* - *Lolium* complex. BMC Genomics **10**,473.
102. Kopecký D., Havránková M., Loureiro J., Castro S., Lukaszewski A. J., Bartoš J., Kopecká J., Doležel J. (2010) Physical distribution of homoeologous recombination in individual chromosomes of *Festuca pratensis* in *Lolium multiflorum*. Cytogenetic and genome research **129**, 162-172.
103. Kopecký D, Bartoš J, Christelová P, Cernoch V, Kilian A, Doležel J (2011) Genomic constitution of Festuca × Lolium hybrids revealed by the DArTFest array. Theor Appl Genet. **122**(2),355-63.
104. Kosmala A, Zwierzykowski Z, Gasior D, Rapacz M, Zwierzykowska E, Humphreys MW (2006) GISH/FISH mapping of genes for freezing tolerance transferred from Festuca pratensis to Lolium multiflorum. Heredity **96**:243–251.
105. Křivánková A, Kopecký D, Stočes Š, Doležel J, Hřibová E. Repetitive DNA: A Versatile Tool for Karyotyping in Festuca pratensis Huds. Cytogenet Genome Res. 2017;**151**(2):96-105
106. Kölliker R, Stadelmann FJ, Reidy B, Nösberger J (1999) Genetic variability of forage grass cultivars: A comparison of *Festuca pratensis* Huds., *Lolium perenne* L., and *Dactylis glomerata* L. Euphytica **106**, 261–270.

107. Kukurba KR, Monthomery SB (2015) RNA Sequencing and Analysis. Cold Spring Harbor Protocols **11**, 951 – 969.

108. Laird PW (2003) The power and the promise of DNA methylation markers. Nature Rev Cancer **3**, 253–266.

109. Lamp CA, Forbes SJ, Cade JW (2001) Grasses of temperate Australia - A field guide. Inkata Press (1st Edition) and CH Jerram & Associates Science Publishers (Revised Edition).

110. Lau AT, Lee SY, Xu YM, Zheng D, Cho YY, Zhu F, *et al.* (2011) Phosphorylation of histone H2B serine 32 is linked to cell transformation. J Biol Chem. **286**,26628–26637.

111. Lee FJ, Moss J, Lin LW (1992) A simplified procedure for hybridization of RNA blots. Biotechniques **13(6)**, 844–846.

112. Lee TI, Rinaldi NJ, Robert F, Odom DT, Bar-Joseph Z, Gerber GK, Hannett NM, Harbison CT, Thompson CM, Simon I, Zeitlinger J, Jennings EG, Murray HL, Gordon DB, Ren B, Wyrick JJ, Tagne JB, Volkert TL, Fraenkel E, Gifford DK, Young RA (2002) Transcriptional Regulatory Networks in *Saccharomyces cerevisiae.* Science **25**, 799-804.

113. Leitch IJ, Bennett MD (1997) Polyploidy in angiosperms. Trends in Plant Science **2**, 470–476.

114. Leonard JE, Grothaus CD, Taetle R (1988) Ricin binding and protein synthesis inhibition in human hematopoietic cell lines. Blood. **72**(4), 1357–1363.

115. Lewis EJ, Tyler BF, Chorlton KH (1973) Development of Lolium-Festuca hybrids. Report Welsh Plant Breeding Station for 1972, pp. 34–37.

116. Lewis EJ (1975) *Festuca* L. × *Lolium* L. = *Festulolium* Aschers and Graebn. In: C.A. Stace (Ed.), Hybridization and the flora of the British Isles, pp. 547–552. Academic Press, London.

117. Lewis ZA, Shiver AL, Stiffler N, Miller MR, Johnson EA, Selker EU (2007) High density detection of restriction site associated DNA (RAD) markers for rapid mapping of mutated loci in Neurospora. Genetics. **177**(2), 1163-1171.

118. Li J, Jiang H, Wong WH. (2010) Modeling non-uniformity in short-read rates in RNA-Seq data. Genome Biol **11**, R25.

119. Li J, Witten DM, Johnstone IM, et al. (2012) Normalization, testing, and false discovery rate estimation for RNA-sequencing data. Biostatistics **13**, 523-38.

120. Lipman MJ, Chester M, Soltis PS, Soltis DE (2013) Natural hybrids between *Tragopogon mirus* and *T. Miscellus* (Asteraceae): a new perspective on karyotypic Changes following hybridization at the polyploidy level. American Journal of Botany **100**(10), 2016-2022.

121. Liu B, Vega JM and Feldman M (1998) Rapid genome changes in newly synthesized amphiploids of Triticum and Aegilops. II. Changes in low-copy coding DNA sequences. Genome **41**: 535–542.

122. Lohe AR, Hilliker AJ, Roberts PA (1993) Mapping Simple Repeated DNA Sequences in Heterochromatin of Drosophila Melanogaster. Genetics **134 (4)**, 1149–74.

123. Loridon K, Burgarella C, Chantret N, Martins F, Gouzy J, Prospéri J-M, Ronfort J (2013) Single-nucleotide polymorphism discovery and diversity in the model legume Medicago truncatula. Mol Ecol Resour. **13**(1), 84-95.

124. Loureiro J, Kopecký D, Castro S, Santos C, Silveira P (2007) Flow cytometric and cytogenetic analyses of Iberian Peninsula *Festuca* spp. Plant Systematics and Evolution **269**, 89-105.

125. Lundqvist A (1962) The nature of the two loci incompatibility system in grasses. I. The hypothesis of a duplicative origin. Hereditas **48**, 153-168.

126. Maher CA, Kumar-Sinha C, Cao X, et al. (2009) Transcriptome sequencing to detect gene fusions in cancer. Nature **458** (7234), 97–101.

127. Mallet J (2005) Hybridization as an invasion of the genome. Trends in Ecology and Evolution **20**, 229-237.

128. Mallet J (2007) Hybrid speciation. Nature **446**, 279 -283.

129. Mammadov J, Chen W, Ren R, Pai R, Marchione W, Yalçin F, Witsenboer H, Greene T, Thompson S, Kumpatla S (2010) Development of highly polymorphic SNP markers from the complexity reduced portion of maize [Zea mays L.] genome for use in marker-assisted breeding. Theor Appl Genet. **121** (3): 577-588.

130. Mayer E (1963) Animal species and evolution. Book: The Belknap press of Harvard University press, Cambtidge, Massachusetts. ISBN 9780674865327

131. Mayrose, I, Zhan SH, Rothfels CJ, Magnuson-Ford K, Barker MS, Rieseberg LH, Otto SP (2011). Recently Formed Polyploid Plants Diversify at Lower Rates. Science **333**, 1257-1257

132. Ma X-F, Gustafson JP (2005) Genome evolution of allopolyploids: a process of cytological and genetic diploidization. Cytogenet Genome Res **109**, 236–249.

133. McClintock B (1984) The significance of responses of the genome to challenge. Science **226**, 792–801.

134. McManus CJ, Coolon JD, Duff MO, Eipper-Mains J, Graveley BR, Wittkopp PJ (2010) Regulatory divergence in Drosophila revealed by mRNA-seq. Genome Res. **20**(6), 816-825.

135. Mercer, TR; Mattick, JS (2013) Understanding the regulatory and transcriptional complexity of the genome through structure. Genome Research. **23**(7), 1081–1088.

136. Mikheyev A S, Tin MMY (2014) A first look at the Oxford Nanopore MinION sequencer. Molecular Ecology Resources. **14** (6), 1097–102.

137. Miller MR, Dunham JP, Amores A, Cresko WA, Johnson EA (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. Genome Research **17**(2):240-248.

138. Mignotte V, Eleouet JF, Raich N, Romeo PH (1989) Cis- and Trans-Acting Elements Involved in the Regulation of the Erythroid Promoter of the Human Porphobilinogen Deaminase Gene. Proceedings of the National Academy of Sciences of the United States of America **86**(17), 6548-6552.

139. Mittelsten Scheid O, Jakovleva L, Afsar K, Maluszynska J, Paszkowski J (1996). A change of ploidy can modify epigenetic silencing. Proc. Natl. Acad. Sci. USA **93**, 7114–7119.

140. Mohindra V, Mishra A, Palanichamy M, Ponniah AG (2001) Cross-species amplification of *Catla catla* microsatellite locus in *Labeo rohita.* Indian Journal of Fisheries **48**(1), 103–108.

141. Morin RD, O'Connor MD, Griffith M, Kuchenbauer F, Delaney A, Prabhu A, Zhao Y, McDonald H, et al. (2008) Application of massively parallel sequencing to microRNA profiling and discovery in human embryonic stem cells. Genome Res. **18**, 610 – 621

142. Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B (2008) Mapping and quantifying mammalian transcriptomes by RNA-seq. Nature Methods **5** (7), 621–628.

143. Nagalakshmi U, Wang Z, Waern K, et al. (2008) The transcriptional landscape of the yeast genome defined by RNA sequencing. Science **320**, 1344-9.

144. Nakamura Y, Koyama K, Matsushima M (1998) VNTR (variable number of tandem repeat) sequences as transcriptional, translational, or functional regulators. Journal of Human Genetics **43**(3), 149 – 152.

145. Ozkan H, Levy AA and Feldman M (2001) Allopolyploidyinduced rapid genome evolution in the wheat (Aegilops-Triticum) group. Plant Cell **13**: 1735–1747.

146. Pennacchio LA, Bickmore W, Dean A, Nobrega, MA, Bejerano G (2013) Enhancers: Five essential questions. Nature Reviews Genetics. **14** (4), 288–95.

147. Pfaffl M (2001) Development and Validation of an Externally Standardised Quantitative Insulin-like Growth Factor-1 RT-PCR Using LightCycler SYBR Green I Technology. Biochemia No. 3 [http://gene-quantification.org/biochemica-3-2000.pdf]

148. Pivoriene O, Pašakinskiene I (2008) Genetic diversity assessment in perennial ryegrass and Festulolium by ISSR fingerprinting. Agriculture **95**, 125-133.

149. Podini D, Vallone PM (2009) SNP genotyping using multiplex single base primer extension assays. Methods Mol Biol. **578**, 379-91.

150. Posselt U (2010) Breeding methods in cross-pollinated species. In: Boller B, Posselt U, Veronesi F, (eds.) In book: Fodder Crops and Amenity Grasses, pp.39-87 Springer Science+Business Media.

151. Pulverer B (2005) Sequence-specific DNA-binding transcription factors. Nature Milestones doi:10.1038/nrm1800.

152. Raser JM, O'Shea EK (2005) Noise in Gene Expression: Origins, Consequences, and Control. Science (New York, N.Y.), **309**(5743), 2010–2013.

153. Ranz MJ, Namgyal K, Gibson G, Hartl LD (2004) Anomalies in the Expression Profile of Interspecific Hybrids of *Drosophila melanogaster* and *Drosophila simulans*. Genome Res. **14**(3), 373-9.

154. Rayburn AL and Gill BS (1986) Molecular identification of the D-genome chromosomes of wheat. Journal of Heredity, **77**(4), 253-255

155. Rebello CJ, Greenway FL, Finley JW (2014) A review of the nutritional value of legumes and their effects on obesity and its related co-morbidities. Obesity Reviews, **15**, 392–407.

156. Ren B (2010) Enhancers make non-coding RNA. Nature **465**, 173-174.

157. Riddle NC, Birchler JA (2003) Effects of reunited diverged regulatory hierarchies in allopolyploids and species hybrids. Trends Genet. **19**, 597-600

158. Rieger R, Michaelis A, Green MM (1968) A glossary of genetics and cytogenetics: Classical and molecular. ISBN 978-3-662-01012-9

159. Rieseberg LH (1997) Hybrid origins of plant species. Annu. Rev. Ecol. Syst. **28**,359–389.

160. Richard G-F, Kerrest A, Bujon B (2008) Comparative genomics and molecular dynamics of DNA repeats in Eukaryotes. Micr. Mol. Bio. Rev **72** (4), 686–727.

161. Richardson OA, Palmer JD (2007) Horizontal gene transfer in plants. Journal of Experimental Botany **58**(1), 1–9

162. Risso D, Schwartz K, Sherlock G, et al. (2011) GC-content normalization for RNA-Seq data. BMC Bioinformatics **12**,480.

163. Rhoads A, Au KF (2015) PacBio sequencing and its Applications. Genomics, Proteomics & Bioinformatics **13**(5), 278-289.

164. Robertson KD, Wolffe AP (2000) DNA methylation in health and disease. Nat Rev Genet **1**, 11–19.

165. Robinson MD, Oshlack A (2010) A scaling normalization method for differential expression analysis of RNA-seq data. Genome Biol **11**, R25.

166. Roldan-Ruiz I, Dendauw J, Van Dockstaele E, Depicker A, De Loose M (2002) AFLP markers reveal high polymorphic rates in ryegrasses (*Lolium* spp.) Mol. Breed. **6**, 125-134.

167. Rose AM, Shah AZ, Venturini G, Krishna A, Chakravarti A., Rivolta C, Bhattacharya SS (2016) Transcriptional regulation of PRPF31 gene expression by MSR1 repeat elements causes incomplete penetrance in retinitis pigmentosa. Sci Rep. **6**, 19450.

168. Rossetto D, Truman AW, Kron SJ, Côté J (2010) Epigenetic modifications in double-strand break DNA damage signaling and repair. Clin Cancer Res. **16**,4543–4552.

169. Russelle M (2001) Alfalfa. Am Sci **89**, 252–259.

170. Ruttink T, Sterck L, Rohde A, Bendixen C, Rouzé P, Asp T, Van de Peer Y, Roldan-Ruiz I (2013) Orthology guided assembly in highly heterozygous crops: Creating a reference transcriptome to uncover genetic diversity in *L. perenne*. Plant Biotechnol. J. **11**:605–617

171. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G *et al*. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature **409**: 928–933.

172. Saksouk N, Simboeck E, Déjardin J (2015) Constitutive heterochromatin formation and transcription in mammals. Epigenetics & Chromatin **8**:3

173. Sen GL, Blau HM (2005) Argonaute2/RISC resides in sites of mammalian mRNA decay known as cytoplasmic bodies. Nature Cell Biology **7**(6), 633–636.

174. Sergeeva EM, Salina EA, Adonina IG, Chalhoub B (2010) Evolutionary analysis of the CACTA DNA-transposon Caspar across wheat species using sequence comparison and in situ hybridization. Mol Genet Genomics. 2010 Jul;**284**(1):11-23

175. Shaked H, *et al.,* (2001) Sequence elimination and cytosine methylation are rapid and reproducible responses of the genome to wide hybridization and allopolyploidy in wheat. Plant Cell **13**, 1749-1759.

176. Shea, T. (2014). Mendel and Genetics: Thinking Outside the Box. Retrieved from http://www.trunity.net/lifeonthisrocksample/view/article/534eb0430cf226e0bdbfc91e

177. Shen R, Su Z, Olsson CA, Fisher PB (1995) Identification of the human prostatic carcinoma oncogene PTI-1 by rapid expression cloning and differential RNA display. Proceedings of the National Academy of Sciences of the United States of America **92**, 6778-6782

178. Shinozuka H, Hisano H, Yoneyama S, Shimamoto Y, Jones ES, Forster JW, Yamada T, Kanazawa A (2006) Gene expression and genetic mapping analyses of a perennial ryegrass glycine-rich RNA-binding protein gene suggest a role in cold adaptation. Mol Genet Genomics **275**:399–408

179. 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

180. Semagn K, Bjornstad A, Ndjiondjop MN (2006) An overview of molecular marker methods for plants. African Journal of Biotechnology **5**(25), 2540-2568.

181. Slover CM, Rodvold KA, Danziger LH (2007) Tigecycline: a novel broad-spectrum antimicrobial. Ann Pharmacother. **41**(6). 965–972.

182. Smale T, Kadonaga, T (2003) The RNA polymerase II core promoter. Annual Review of Biochemistry. **72**, 449–479.

183. Smarda P, Bureš P, Horová L, Foggi B, Rossi G (2008) Genome size and GC content evolution of Festuca: ancestral expansion and subsequent reduction. Annals of Botany **101**, 421-433.

184. Sonenberg N, Hinnebush AG (2009) Regulation of Translation Initiation in Eukaryotes: Mechanisms and Biological Targets. Cell **136**, 731–745.

185. Sontheimer EJ (2005). Assembly and function of RNA silencing complexes. Nature Reviews Molecular Cell Biology **6**(2), 127–138.

186. Song K, Lu P, Tang K and Osborn TC (1995) Rapid genome change in synthetic polyploids of Brassica and its implications for polyploid evolution. Proc. Natl. Acad. Sci. USA **92**: 7719–7723.

187. Stucki M, Clapperton JA, Mohammad D, Yaffe MB, Smerdon SJ, Jackson SP (2005) "MDC1 directly binds phosphorylated histone H2AX to regulate cellular responses to DNA double-strand breaks," Cell **123**(7), 1213–1226,

188. Stapley, J. *et al.* (2010) Adaptation genomics: the next generation. Trends Ecol. Evol. **25**, 705–712.

189. Stočes, Š., Ruttink, T., Bartoš, J., Studer, B., Yates, S., Zwierzykowski, Z., Abrouk, M., Roldán-Ruiz, I., Książczyk, T., Rey, E., Doležel, J., Kopecký, D (2016) Orthology Guided Transcriptome Assembly of Italian ryegrass and meadow fescue for single-nucleotide polymorphism discovery. THE PLANT GENOME **9**: 1-14.

190. Straková M., Straka J., Michalikova L., Plevova K. (2007) Kapesni atlas trav. str.7-12 Agrotis Travniky, s.r.o.

191. Studer B, Byrne S, Nielsen RO, Panitz F, Bendixen C, Islam MS, Pfeifer M, Lübberstedt T, Asp T (2012) A transcriptome map of perennial ryegrass (*Lolium perenne* L .). BMC Genomics **13**:140

192. Suttie JM, Reynholds SG, Batello C (2005) Introduction. In: Suttie JM, Reynolds SG, Batello C, editors. Grasslands of the world. Rome: FAO.

193. Suzuki MM, Bird A (2008) DNA methylation landscapes: provocative insights from epigenomics. Nat Rev Genet **9**, 465–476.

194. Swaney SM, Aoki H, Ganoza MC, Shinabarger DL (1998) The Oxazolidinone Linezolid Inhibits Initiation of Protein Synthesis in Bacteria. Antimicrob. Agents Chemother. **42**(12), 3251–3255.

195. Szyf M (2009) Epigenetics, DNA methylation, and Chromatin Modifiying Drugs. Annual Review of Pharmacology and Toxicology **49**, 243-263.

196. Tate JA, Soltis DE, Soltis PS (2005) Polyploidy in plants. In: Gregory T.R, editor. The evolution of the genome. Elsevier Science & Technology, Academic Press; San Diego, CA, pp. 371–426.

197. Tautz D (1989) Hypervariability of simple sequences as a general source for polymorphic DNA markers. Nucleic Acids Res. **17**, 6463–6471.

198. Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ; Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol **28** (5), 511–515.

199. Trayhurn P (1996) Northern Blotting. Pro. Nutrition Soc. **55**,583–589.

200. Trebbi D, Maccaferri M, Heer P, Sørensen A, Giuliani S, Salvi S, Sanguineti M, Massi A, Vossen E, Tuberosa R (2011) High-throughput SNP discovery and genotyping in durum wheat (*Triticum durum* Desf.). Theor Appl Genet. **123**(4), 555-569.

201. Tsujimoto H, Mukai Y, Akagawa K, Nagaki K, Fujigaki J, Yamamoto M, Sasakuma T. (1997) Identification of individual barley chromosomes based on repetitive sequences: conservative distribution of Afa-family repetitive sequences on the chromosomes of barley and wheat. Genes Genet Syst. 1997 Oct;**72**(5):303-9.

202. Van de Peer Y, Meyer A (2005) Chapter6: Large-scale gene and ancient genome duplication. In The Evolution of the Genome, ed. T.R. Gregory, 330-363.

203. Van de Peer Y, Maere S, Meyer A (2009) The evolutionary significance of ancient genome duplications. Nature Reviews Genetics **10**, 725-732.

204. Velculescu VE, Zhang L, Vogelstein B, Kinzler KW (1995) Serial analysis of gene expression. Science **270** (5235): 484–7.

205. Vergnoud G, Denoeud F (2000) Minisatellites: Mutability and Genome Architecture. Genome resourses **10**, 899-907.

206. Voinnet, O. (2009). Origin, Biogenesis, and Activity of Plant MicroRNAs. Cell **136**, 669-687.

207. Wang HJ, Cunnold DM, Froidevaux L, Russell JM (1999) A reference model for middle atmospheric ozone in 1992–1993, J. Geophys. Res., **104**, 21,629–21,643.

208. Wang J, Tian L, Madlung A, Lee HS, Chen M, Lee JJ, Watson B, Kagochi T, Comai L, Chen ZJ (2004) Stochastic and epigenetic changes of gene expression in Arabidopsis polyploids. Genetics. **167**(4), 1961-73.

209. Wang J, Tian L, Lee H-S, Wei NE, Jiang H, Watson B, Madlung A, Osborn TC, Doerge RW, Comai L, Chen ZJ (2006) Genomewide nonadditive gene regulation in Arabidopsis allotetraploids. Genetics **172**, 507-517.

210. Wang J, Pembleton LW, Baillie RC, Drayton MC, Hand ML, Bain M, *et al.* (2014) Development and implementation of a multiplexed single nucleotide polymorphism genotyping tool for differentiation of ryegrass species and cultivars. Mol. Breed. **33**:435–451

211. Warren L (2008) Review of medical microbiology and immunology. New York: McGraw-Hill Medical. ISBN 978-0-07-149620-9.

212. Weber D, Helentjaris T (1989) Mapping RFLP loci in maize using B-A translocations. Genetics **121**, 583 – 590.

213. Weigel D, Mott R (2009) The 1001 genomes project for Arabidopsis thaliana. Genome Biol. **10** (5): 107-110.

214. Wendel JF (2000) Genome evolution in polyploids. Plant Mol. Biol. **42**, 225–249.

215. Wicker T, Guyot R, Yahiaoui N, Keller B (2003) You have access CACTA Transposons in Triticeae. A Diverse Family of High-Copy Repetitive Elements. Plant Physiol. Vol. **132**, 52-63.

216. Wood, TE, Takebayashi, N, Barker, MS, Mayrose, I, Greenspoon, PB, Rieseberg, LH (2009). The frequency of polyploid speciation in vascular plants. Proceedings of the National Academy of Sciences **106**, 13875-13879.

217. Wool IG (1979) The structure and function of eukaryotic ribosomes. Annual Review of Biochemistry **48**, 719-754.

218. Wray GA (2007) The evolutionary significance of cis-regulatory mutations. Nat Rev Genet. **8**(3), 206-216.

219. Wu X, Liu J, Li D, Liu CM (2016) Rice caryopsis development II: Dynamic changes in the endosperm. J Integr Plant Biol **58**: 786–798

220. Xu WW, Sleper DA (1994) Phylogeny of tall fescue and related species using RFLPs. Theor. Appl. Gener **88**, 909-913.

221. Yamada T, Kishida T (2003) Genetic analysis of forage grasses based on heterologous RFLP markers detected by rice cDNAs. Plant Breed. **122**, 57-60.

222. Yang B, Thorogood D, Armstead ISB (2008) How far are we from unravelling self-incompatibility in grasses? New Phytologist **178**, 740–753.

223. Yang S, Tu Z, Cheung F, Xu W, Lamb J, Jung H-J, Vance C, Gronwald J (2011) Using RNA-Seq for gene identification, polymorphism detection and transcript profiling in two alfalfa genotypes with divergent cell wall composition in stems. BMC Genomics **12**(1), 1-19.

224. Yoo MO, Szadkowski E and Wendel JF (2013) Homoeolog expression bias and expression level dominance in allopolyploid cotton. Heredity volume **110**, pages 171–180

225. Zhang X, Cal AJ, Borevitz JO (2011) Genetic architecture of regulatory variation in *Arabidopsis thaliana*. Genome Res. **21**(5), 725-33.

226. Zwierzykowski Z, Kosmala A, Zwierzykowska E, Jones N, Jokś W, Bocianowski J (2006) Genome balance in six successive generations of the allotetraploid Festuca pratensis x Lolium perenne. Theor Appl Genet. 2006 Aug;**113**(3):539-47.

8. LIST OF ABBREVIATIONS

^{32}P	radioactive isotope of phosphor
3D	three dimensions
AE	anchoring enzyme
AFLP	amplified fragment length polymorphisms
aRNA	antisense RNA
BsmF1	enzyme
cDNA	complementary DNA
cm	centimetres
CpG	CpG is shorthand for 5'—C—phosphate—G—3'
DArT	diversity arrays technology
DCC	duplication-degeneration-complementation hypotheses
DD	differential display
DDR	DNA damage response
ddRADseq	modified reduced-representation sequencing
DNA	deoxyribonucleic acid
DpnII	restriction endonuclease
dsDNA	double-stranded DNA
dsRNA	double-stranded RNA
dT	biotinylated oligo
F1	first filial hybrid generation
F2	second filial hybrid generation
FISH	fluorescent *in situ* hybridization
FPKM	fragments Per Kilobase of transcript per Million mapped reads
Gbp	giga base pairs
GC	guanine and cytosine
GenBank	sequence database
GISH	genomic *in situ* hybridization
H3	histon H3
H4	histon H4

ITS	internal transcribed spacer
kb	kilo base pairs
KMT	lysine methyltransferase
lncRNA	long noncoding RNA
M	methionine
MAS	marker assisted selection
miRNA	microRNA
mm	millimetres
mRNA	messenger RNA
Mst1	restriction endonuclease
n	haploid chromosome dosage
NB	negative binomical distribution
NGS	next-generation sequencing strategies
NlaIII	enzyme
OF28	out-of-frame construct
P1	adapter in reduced-representation sequencing
P2	adapter in reduced-representation sequencing
P-bodies	cytoplasmic bodies
PCR	polymerase chain reaction
pre-mRNA	primary messenger RNA
pri-miRNA	primary miRNA
QTL	quantitative trait locus
RAD	restriction site associated DNA markers
RADseq	reduced-representation sequencing
RAGE	rapid analysis of gene expression
RFLP	restriction fragment length polymorphism
RISC	RNA-induced silencing complex
RNA	ribonucleic acid
RNAi	interference RNA
RNA-seq	sequencing of RNA
RPKM	reads Per Kilobase of transcript per Million mapped reads
rRNA	ribosomal RNA

RT-qPCR	quantitative real time PCR
S	serin
SAGE	serial analysis of gene expression
SAM	S-adenosylmethionine
sdESt	sequence database
siRNA	small interfering RNA
SNP	single nucleotide polymorphisms
snRNA	small nuclear RNA
SSCP	single-strand conformation polymorphism
SSR	microsatellite
ssRNA	single-stranded RNA
STR	short tandem repeat
TE	target enzyme
TERC	telomerase RNA component
TF	transcription factors
TMM	trimmed Mean of M-values normalization method
tRNA	transfer RNA
UniGene	sequence database
VNTR	variable number tandem repeat