Comparative analysis of DNA repeats and identification of novel Fesreba centromeric element in fescues and ryegrasses

CURRENT STATUS: Under Review

BMC Plant Biology

Jana Zwyrtková, Alžběta Němečková, Jana Čížková, Kateřina Holušová, Veronika Kapustová, Radim Svačina, David Kopecký, Bradley John Till, Jaroslav Doležel, Eva Hřibová

Jana Zwyrtková
Ustav Experimentalní Botaniky Akademie ved Ceske republiky

Alžběta Němečková
Ustav Experimentalní Botaniky Akademie ved Ceske republiky

Jana Čížková
Ustav Experimentalní Botaniky Akademie ved Ceske republiky

Kateřina Holušová
Ustav Experimentalní Botaniky Akademie ved Ceske republiky

Veronika Kapustová
Ustav Experimentalní Botaniky Akademie ved Ceske republiky

Radim Svačina
Ustav Experimentalní Botaniky Akademie ved Ceske republiky

David Kopecký
Ustav Experimentalní Botaniky Akademie ved Ceske republiky

Bradley John Till
Centro de Genomica Nutricional Agroacuicola

Jaroslav Doležel
Ustav Experimentalní Botaniky Akademie ved Ceske republiky

Eva Hřibová
IEB AS CR

hribova@ueb.cas.cz Corresponding Author

ORCiD: https://orcid.org/0000-0002-6868-4344
Prescreen

10.21203/rs.2.24364/v2

Subject Areas

*Plant Molecular Biology and Genetics*

Keywords

*Festuca spp.*, *Lolium spp.*, *Illumina sequencing*, *repetitive DNA*, *centromere organization*
Abstract

Background Cultivated grasses are an important source of food for domestic animals worldwide. Better knowledge of their genomes can speed up the development of new cultivars with better quality and resistance to biotic and abiotic stresses. The most widely grown grasses are tetraploid ryegrass species (Lolium spp.) and diploid and hexaploid fescue species (Festuca spp.). In this work we characterized repetitive DNA sequences and their contribution to genome size in five fescue and two ryegrass species, as well as one fescue and two ryegrass cultivars.

Results Partial genome sequences produced by Illumina technology were used for genome-wide comparative analyses using RepeatExplorer pipeline. Retrotransposons were found to be the most abundant repeat types in all seven grass species. Athila element of Ty3/gypsy family showed the most striking differences in copy number between fescues and ryegrasses. The sequence data enabled the assembly of an LTR element Fesreba, which is highly enriched in centromeric and (peri)centromeric regions in all species. A combination of FISH with a probe specific to Fesreba element and immunostaining with CENH3 antibody showed their colocalization and indicated a possible role of Fesreba in centromere function.

Conclusions Comparative repeatome analysis in a set of fescues and ryegrasses provided new insights into their genome organization and divergence, including the assembly of LTR element Fesreba. A new LTR element Fesreba was identified and found abundant in centromeric regions of the fescues and ryegrasses. It may have a role in the function of their centromeres.

Background

Grasses (Poaceae) are an important source of food for domestic animals worldwide and perform important ecological and environmental functions. The tribe Poeae is the largest tribe in family Poaceae and the species from its largest subtribe, Loliinae, grow on a range of habitats, including wetlands, dry areas, regions with cold and temperate climate and some are well adapted to extreme conditions in mountain, arctic and sub-antarctic regions [1]. The subtribe Loliinae comprises a cosmopolitan genus Festuca and its satellite genera [2, 3]. Festuca is the largest genus of the family Poaceae, containing more than 600 species and Torrecilla and Catalán [4] discriminate its two main evolutionary lines: ‘broad-leaved’ and ‘fine-leaved’ (Figure 1). Broad-leaved Festuca species (hereafter called fescues), includes subgenus Schedonorus, which gave rise to Lolium species (hereafter called ryegrasses), a sister group of fescues (Figure 1) [1]. The evolution of grasses, including Loliinae, was accompanied by frequent polyploidization and hybridization events, and about 70% of grass species are polyploid [5]. The species of Loliinae have large genomes ranging from 2.6 Gbp/1C to 11.8 Gbp/1C [6, 7].

This study focuses on species from subgenus Schedonorus, a complex of species with various ploidy levels [6, 8], which includes important species widely used for forage and turf. Although, some Schedonorus species are diploid, such as Festuca pratensis Huds. (2n = 2x = 14) and Lolium multiflorum Lam. (2n = 2x = 14) and L. perenne L. (2n = 2x = 14), a majority of species are allopolyploid [9, 10] and include tetraploid F. glaucescens Boiss. (2n = 4x = 28) and F. mairei St. Yves (2n = 4x = 28), hexaploid F. arundinacea Schreb. (2n = 6x = 42) and F. gigantea (L.) Vill. (2n = 6x = 42) [3, 10]. Fescues are more tolerant than ryegrasses to abiotic stresses, provide high quality forage for livestock and are grown especially for turf purposes. On the other hand, ryegrasses are characterized by high yield and excellent nutritional value and are mostly cultivated as pasture. Artificial inter-generic hybrids between fescues and ryegrasses species have been developed combining favorable characters of both genera [11-13].

Even though fescues and ryegrasses are intensively studied, their evolution and the origin of most of polyploid representatives remains obscure [10, 14, 15]. Like in other species with large genomes, nuclear genomes of fescues and ryegrasses include a large number and a variety of repetitive DNA sequences [16, 17]. Their amplification in the genome, accompanied by interspecific hybridization and polyploidization, lead to genome size expansion [18-23]. However, these processes were probably counterbalanced by recombination-based
mechanisms which removed substantial parts of nuclear genomes [24–26].

Repetitive DNA elements may have different roles in a nuclear genome. Tandem organized ribosomal RNA genes and telomeric sequences are the key components of nucleolar organizing regions and chromosome termini, respectively. Centromeric regions in Arabidopsis, Brachypodium, rice and maize, are partly formed from specific satellite DNAs with ~130 bp long units [27–30], while in other plant species, including cereals, these regions are formed by large blocks of Ty3/gypsy retrotransposons containing chromodomain [28, 31–33]. In F. pratensis, a putative LTR element localizing preferentially to centromeric regions was identified [34]. In addition to understanding the molecular organization of chromosome domains, characterization of repetitive parts of nuclear genomes helps to develop cytogenetic markers [20, 34, 35]. Repetitive DNA sequences are also used extensively in studies of genetic diversity and to study processes of genome evolution and speciation [36–39].

The main goal of the present work was to elucidate repetitive landscape and its impact on genome size and genome divergence in closely related land grasses, including natural polyploid species. We characterized repetitive DNA sequences in nuclear genomes of ten representatives of fescues and ryegrasses. We performed global analysis of repetitive DNA sequences and characterized their abundance and variability after partial Illumina sequencing. Apart from global characterization of repetitive parts of fescue and ryegrass genomes, we characterized and assembled DNA sequence of an LTR element, which is highly enriched in centromeric and (peri)centromeric chromosome regions in all ten genotypes. Co-localization of centromere-specific histon H3 variant CENH3 with the LTR element indicated its role in centromere function.

Results

Genome size estimation

The amount of nuclear DNA was estimated after flow-cytometric analysis of propidium iodide-stained nuclei (Figure 2). Due to large differences in genome size between the analysed species, two internal reference standards were used, Pisum sativum cv. Citrad; 2C = 9.09 pg DNA [40] and Secale cereale cv. Dankovske; 2C = 16.19 pg DNA [40]. All analyses resulted in histograms of relative DNA content with two dominant peaks corresponding to G1 nuclei of the sample and the standard. The 2C nuclear DNA content thus determined ranged from 5.32 pg in L. multiflorum to 20.17 pg in F. gigantea. Monoploid genome size (1Cx) ranged from 2.43 in F. mairei to 3.36 pg in F. gigantea (Table 1). The remaining representatives of fescues and ryegrasses had similar monoploid genome sizes (1Cx ~ 2.7 Gb).

Repeat composition and comparative analysis of repetitive DNA sequences

Inter-specific comparisons, reconstruction and quantification of major repeat families were done using RepeatExplorer pipeline [41]. The process involved grouping of orthologous repeat families from all analyzed species in the same cluster and facilitated the assembly, identification and quantification of individual repeat elements.

In all accessions, LTR retroelements were found to be the most abundant nuclear genome component (Table 2, Figure 3). Out of them, Ty3/gypsy elements were more than four times more abundant than Ty1/copia retrotransposons (Table 2). The biggest difference in copy number between fescues and ryegrasses was revealed for an LTR element from the Athila clade. While nuclear genomes of both Lolium species were enriched for the element, which accounts for ~25 – 30% of their genomes, the orthologous Athila element accounted for only ~5 – 7% of nuclear genomes in fescues (Table 2). A relatively large part of the genomes was represented by unclassified LTR sequences, indicating high frequency of unique LTR sequences. DNA transposons and LINE elements were found in low copy numbers, and tandem repeats accounted for 1.5 % to more than 8 % of the genome sequences (Table 2, Figure 3).

Comparative analysis by RepeatExplorer showed that most clusters of orthologous repeat families contained
reads from all accessions, and that a large number of similar sequences was identified in fescues and ryegrasses. Within the fescues, *F. mairei* and *F. glaucescens* showed the lowest similarity in DNA repeats as compared to other fescues. The composition as well as the abundance of DNA repeats in ryegrasses were found to be highly conserved. Tandem organized repeats were the most diverged elements among the studied fescues and ryegrasses and some of the repeats were found to be species-specific (Figure 4, Additional file 1: Table S1). In addition to tandem repeats, some of small sequence clusters contained reads from only a few species. Species-specific variants of a majority of repetitive elements within and between fescues and ryegrasses were identified only after detailed analysis of individual repeat clusters using the SeqGrapheR program (Figure 5A–C). A detailed analysis revealed the presence of species-specific DNA contigs, which may be used to develop molecular and cytogenetic markers.

To confirm the differences determined in silico, selected repetitive DNA elements were analysed using Southern hybridization. For those DNA repeats, which seemed to have species-specific variants, specific probes were designed. A probe for Ty3/gypsy Athila element, which was reconstructed in cluster CL1 and which showed the largest copy number variation between fescues and ryegrasses (Table 2), gave strong hybridization signals on genomic DNA from ryegrasses, with no or weak signals on DNA from fescues (Figure 5D). Similarly, a probe for Ty3/gypsy Athila element, which was reconstructed in cluster CL38 and contained mostly Festuca sequence reads (Figure 5B), provided strong visible signals only with fescue genomic DNA (Figure 5E). Finally, Southern hybridization was done with a probe for Ty3/gypsy Ogre-Tat retrotransposon, identified in cluster CL20. The probe was designed from contigs representing fescues (Figure 5C) provided strong hybridization signals on all analysed fescues and low intensity signals in ryegrasses (Figure 5F). In general, the signal intensities obtained after Southern hybridizations corresponded to copy numbers identified in silico.

**Centromere composition**

Partial genome sequence data obtained using Illumina technology made it possible to reconstruct nearly complete centromeric LTR elements in all ten accessions of fescues and ryegrasses. Detailed characterization of the element called Fesreba confirmed that it belongs to Ty3/gypsy Chromoviridae lineage. Phylogenetic analysis of its reverse transcriptase domain showed close relationship with the Cereba element (Figure 6), which was identified earlier in barley (*Hordeum vulgare*) [42].

Southern hybridization with a probe for reverse transcriptase (RT) domain of Fesreba and a probe for its LTR region [34] showed their presence in all fescues and ryegrasses included in this work (Additional file 2: Figure S1). Similar hybridization patterns indicated sequence conservation between Fesreba repetitive DNA elements in these species. The results were supported by in silico data, which showed high similarity at DNA sequence level (most abundant copies of Fesreba shared at least 92% similarity at DNA level within and between fescues and ryegrasses), but lower abundance in ryegrasses. To confirm the differences in Fesreba copy number, quantification was performed for RT domain and LTR sequence using droplet digital PCR (ddPCR). The results confirmed two-fold higher copy number of Fesreba in fescues as compared to ryegrasses (Additional file 3: Table S2). The assay also showed that a majority of analysed genotypes contained five to fifty times more copies of LTR region of Fesreba as compared to its coding region (Additional file 3: Table S2).

To confirm preferential localization of Fesreba to centromeric chromosome regions, FISH on mitotic metaphase plates was conducted with probes derived from its RT domain and LTR region. In all fescues and ryegrasses, both probes localized preferentially to centromeric regions of all chromosomes (Figure 7). Whilst the hybridization signals of RT domain were observed almost exclusively in centromeric regions, a probe derived from non-coding LTR region resulted in stronger signals in centromeric and/or pericentromeric regions and weak signals along the chromosomal arms, as previously shown in *F. pratensis* [34]. Weak signals of LTR part of Fesreba in distal parts of chromosomes indicate the presence of unique LTRs spread over the genome and correspond to higher copy number of LTR non-coding part of Fesreba as compared to its coding sequence.

In addition to the fescues and ryegrasses included in this study, FISH was done with the same probes on mitotic metaphase plates from related grass species oats, barley, rye, bread wheat and *Aegilops tauschii*. High
homology of RT coding domain resulted in successful in situ localization in all species. On the other hand, the probe specific to LTR region of Fesreba provided visible signals only in A. sativa (Additional file 4: Figure S2). Finally, immunostaining with centromere-specific histone H3 variant CENH3 [43] in combination with FISH with probes for RT domain and for LTR region of Fesreba resulted in overlapping signals in all studied fescues and ryegrasses (Figure 8, Additional file 5: Figure S3).

Discussion

Due to a genome shock, monoploid genome size (1Cx) of polyploid species is often, but not always lower as compared to that of their progenitors [24, 44]. In this study, we performed comparative analysis of repeatomes and analyzed the impact of DNA repeats on genome size in a set of Festuca and Lolium species differing in ploidy. The set comprised hexaploids F. arundinacea subsp. arundinacea and F. gigantea, tetraploids F. glaucescens and F. mairei, and artificial autotetraploids F. pratensis cv. Westa, L. multiflorum cv. Mitos and L. perenne cv. Neptun developed in breeding programs. We estimated nuclear DNA amounts using flow cytometry and a test of normality confirmed that the dataset had normal distribution. Our study suggested possible genome changes in hexaploid F. arundinacea and tetraploid ryegrasses compared to their probable progenitors. Although the difference between monoploid genome size of natural polyploid F. arundinacea and its probable parents (F. pratensis and F. glaucescens) are small, they are still statistically significant (P < 0.01). The same was true for tetraploid ryegrass cultivars obtained after polyploidization. Genome downsizing was detected in case of F. arundinacea (≈ 2 % difference between expected and estimated value) and tetraploid L. perenne (≈ 1 % decrease). In tetraploid cultivar of L. multiflorum, slight increase of genome size (≈ 4 %) was detected, corresponding with previous study of Kopecký et al. [47]. In case of tetraploid fescue cultivar obtained after polyploidization, statistically significant difference in 1Cx value was not found (P > 0.01).

DNA retrotransposons are a major contributor to the variation in nuclear genomes in plants (e.g. [23, 45, 46]). Various approaches and tools have been developed to study this important part of nuclear genomes, one of them being RepeatExplorer, which facilitates de novo repeat identification and characterization [41, 47]. The pipeline uses graph-based clustering and is suitable for the analysis of next generation sequencing data to reconstruct and characterize DNA repeats in a particular species, or to compare DNA repeat composition in different genotypes [22, 23, 48, 49]. The pipeline has been used frequently to reconstruct DNA repeats in diversity studies, to create repeat databases for repeat masking [18, 45, 47] and to identify tandem organized repeats suitable as probes for molecular cytogenetics [34, 51, 52].

Our work revealed that Ty3/gypsy elements had the highest impact on genome size in fescues and ryegrasses. Ty3/gypsy elements were found most abundant also in other Poaceae species, including wheat, rice, maize and barley [7,53–55]. In barley, about 50% of the genome is made of fifteen high copy TE families with the elements of the Angela lineage (Ty1/copia family) being the most abundant and representing almost 14% of the genome [55]. The Ty3/gypsy superfamily was 1.5-fold more abundant than Ty1/copia superfamily [55].

Festuca and Lolium genera comprise closely related complexes of species and in line with this, a high homology of DNA repeats was observed in this work. The main difference being the copy number. In Lolium species, Ty3/gypsy Athila LTR retroelement accounted for ≈25% of their nuclear genomes, while in fescues it accounted for ≈ 0.7% in tetraploids F. glaucescens and F. mairei, and for ≈ 6% in other analysed fescues. This observation indicates a burst of Athila LTR element linked with Lolium speciation. It is known, that activation and integration of transposable elements may occur, e.g., due to environment change, lead to a rapid burst in a species-specific manner [45, 46, 56] and impact the evolution and speciation [45, 57]. In some species, rapid increase of lineage-specific retroelements can also result in significant genome upsizing [23, 57-59], which was not observed in fescues and ryegrasses included in our study.

Species-specific DNA elements identified in this work were represented by tandem organized repeats (Additional file 1: Table S1). Unique tandem repeats were found also in other plant species and thanks to their genus- or species-specificity they have been used widely in molecular cytogenetics, e.g., for identification of chromosomes using FISH (e.g. [60–63]). Tandem repeats originally identified in F. pratensis chromosome 4F were found useful
as probes for FISH to identify individual chromosomes of the species [17, 34] and in comparative karyotype analysis of its cultivars. The present work resulted in identification of other putative tandem organized repeats, either genus- or species-specific (Additional file 1: Table S1). These observations expand the number of potential cytogenetic markers for comparative karyotyping and identification of chromosomes in other fescue and ryegrass species.

Although relatively high number of tandem repeats was revealed after sequencing F. pratensis chromosome 4F, none of them localized to chromosome centromeric regions [17, 34]. However, mapping of other types of DNA repeats on mitotic metaphase chromosomes showed preferential localization of one, uncharacterized DNA element CL38 to centromeric regions of F. pratensis chromosomes [34]. In this work, the entire DNA element homologous to CL38 repeat was reconstructed and its nature was clarified. Phylogenetic analysis of its coding domains (Figure 6) confirmed close relationships with other plant centromeric elements of Ty3/gypsy Chromoviridae lineage, such as the Cereba-like elements [42]. Preferential localization of Cereba element to centromeric regions of barley chromosomes was showed by Hudakova et al. [32] and more complex study of centromere specific element representing CRM lineage of Ty3/gypsy family in larger set of plant species followed [19, 33]. These studies imply a role of transposable elements at the structural level and their impact on centromere structure. Li et al. [64] showed a strong association of Cereba element with histone H3 variant CENH3, which plays a role in centromere function. Co-localization of centromere specific element Fesreba, reconstructed in this work with histone CENH3 (Figure 8, Additional file 5: Figure S3) indicates a role for this element in the function of fescue and ryegrass centromeres as well.

Conclusions

Partial sequencing of genomes in ten fescues and ryegrasses revealed various types of retrotransposons as the most abundant repeat type. A comparative repeatome analysis improved the knowledge of genome organization in fescues and ryegrasses and confirmed close relationships of Festuca and Lolium. The most striking difference was observed for the Athila element, which is ~5 times more abundant in Lolium as compared to Festuca. Highly diverged DNA repeats were represented by tandem organized repeats, which are candidates for species-specific cytogenetic markers. In addition to tandem repeats, other species-specific variants of a majority of repetitive DNA sequences within and between fescues and ryegrasses were identified. A nearly complete LTR element Fesreba was assembled and was found to be highly enriched in centromeric and (peri)centromeric chromosome regions in all species. A combination of FISH with a probe for Fesreba and immunostaining with CENH3 antibody showed their co-localization and indicated a possible role of Fesreba in centromere function.

Material And Methods

Plant material

L. perenne GR3320 (2n = 2x = 14), F. arundinacea subsp. arundinacea (2n = 6x = 42), F. gigantea GR11759 (2n = 6x = 42), F. mairei GR610941 (2n = 4x = 28) were obtained as seeds from Leibniz Institute of Plant Genetics and Crop Plant Research (Gatersleben, Germany) gene bank. Seeds of F. pratensis cv. Fure (2n = 2x = 14) were obtained from Dr. Arild Larson (Graminor, Norway), L. perenne cv. Neptun (2n = 4x = 28) and L. multiflorum cv. Kurl1 (2n = 2x = 14), from Dr. Vladimír Černoch (DLF Seeds, Czech Republic). Plants of L. multiflorum cv. Mitos (2n = 4x = 28), F. pratensis cv. Westa (2n = 4x = 28) and F. glaucescens (2n = 4x = 28) were provided by one of us (DK).

Seeds of barley (Hordeum vulgare) cv. Morex, rye (Secale cereale) cv. Dánkowskie Diament, oats (Avena sativa) cv. Atego were obtained from Leibniz Institute of Plant Genetics and Crop Plant Research gene bank. Seeds of Triticum aestivum cv. Chinese Spring were obtained from Prof. Takashi R. Endo (Kyoto University, Japan) and seeds of Aegilops tauschii were provided by Dr. Valárik (Institute of Experimental Botany, Czech Republic). Seeds of pea (Pisum sativum cv. Ctirad) and rye (Secale cereale cv. Dankovske), which served as internal reference standards in flow cytometric analysis, were provided by one of us (JD).
Estimation of nuclear genome size

Nuclear DNA amounts were determined according to Doležel et al. [65] following the two-step procedure of Otto [66] with modifications. Samples of isolated nuclei stained by propidium iodide were analysed using Sysmex CyFlow Space flow cytometer (Sysmex Partec GmbH, Münster, Germany) equipped with a 532-nm laser. Two reference standards were used to estimate DNA amounts in absolute units. Pea (Pisum sativum cv. Ctirad; 2C = 9.09 pg DNA, [40]) served as an internal standard for DNA content estimation in all accessions with the exception of F. mairei, for which rye (Secale cereale cv. Dankovske; 2C = 16.19 pg DNA, [40]) was used. Three plants were measured per accession and each plant was analysed three times on three different days. At least 5000 nuclei per sample were analysed. Nuclear DNA content was then calculated from individual measurements following the formula:

\[
2C \text{ nuclear DNA content [pg]} = 2C \text{ nuclear DNA content of reference standard } \times \frac{\text{sample } G_1 \text{ peak mean}}{\text{standard } G_1 \text{ peak mean}}
\]

Mean nuclear DNA content (2C) was calculated for each plant. Genome size (1C value) was then determined considering 1 pg DNA equal to 0.978×10^9 bp [67]. Statistical significance of differences between monoploid (1Cx) genome sizes were determined using one-way ANOVA. The analysis was performed using NCSS 97 statistical software (Statistical Solutions Ltd., Cork, Ireland). The significance level \(\alpha = 0.01\) was used.

Phylogenetic analysis

Phylogenetic analysis of Loliinae subtribe was done based on data published by Catalán et al. [3]. Sequence of ITS regions were downloaded from the NCBI GenBank (GB codes: AF303401-407, AF303410-416, AF303418-419, AF303421-425, AF303428, AF478475-476, AF478478-491, AF478493, AF478498-499, AF519975-981, AF519983, AF532937, AF532939-948, AF532951-952, AF532954, AF532956-960, AF532962-963, AF543514, AF548028, AJ240143, AJ240146, AJ240148, AJ240153, AJ240155-157, AJ240160, AJ240162, AY099007, AY118087-088, AY118090-092, AY118094-096, AY228161). Brachypodium distachyon (GB code AF303339) was used as an outgroup species. The sequences were aligned by MAFFT program v7.029 (--localpair --maxiterate 1000) [68] and phylogenrams were constructed by PhyML 3.0 [69] implemented in SeaView v5.0.2 [70]. Approximate likelihood ratio test [71] was performed to assess the branch support. Phylogenetic trees were drawn and edited using FigTree program (http://tree.bio.ed.ac.uk/software/figtree/).

Illumina sequencing and data analysis

Genomic DNA was isolated using the NucleoSpin Plant II kit (Macherey-Nagel GmbH & Co. KG, Düren, Germany) following the manufacturer's recommendations and used for preparation of Illumina libraries using Nextera® DNA Sample Preparation Kit (Illumina, San Diego, USA). 50 ng of DNA was fragmented, purified and amplified according to the protocol. DNA concentration in individual libraries was measured using a Qubit fluorometer, adjusted to an equal molar concentration and pooled prior to sequencing. DNA sequencing was done with an Illumina MiSeq using either single or paired end sequencing to produce up to 500 base pair reads. Sequences reads were deposited in the Sequence Read Archive (BioProject ID: PRJNA601325, accessions SAMN13866227, SAMN13866228, SAMN13866229, SAMN13866230, SAMN13866231, SAMN13866232, SAMN13866233, SAMN13866234, SAMN13866235, SAMN13866236).

Illumina reads were trimmed for adapters and for quality using FASTX-toolkit [-q 20 -p 90] (http://hannonlab.cshl.edu/fastx_toolkit/index.html). Detailed characterization of repeat families was performed using stand-alone version of RepeatExplorer pipeline [36] running on IBM server with 16 processors, 100Gb of RAM and 17Tb of disk space. In the first step, comparative analysis of repetitive parts of the genomes was performed using the RepeatExplorer pipeline according to Novák et al. [48]. Random data sets represented the
same amount of reads 0.5× coverage of individual accessions and used to reconstruct repetitive elements using graph-based method according Novák et al. [47]. The assembled sequences within each individual cluster were characterized based on the homology searches and other tools useful for repeat characterization (e.g. BLASTN and BLASTX programs, phylogenetic analysis). Tandem organized repeats were identified using Dotter [72].

In the second step, RepeatExplorer pipeline was applied on a merged dataset containing all species marked by specific prefixes to perform comparative analysis [48]. The results of the clustering were then used to create repetitive databases. Databases of Illumina reads and assembled contigs from different types of repetitive DNA elements are publicly available on web site https://olomouc.ueb.cas.cz/en/content/dna-repeats.

**Southern hybridization**

Genomic DNA corresponding to $3 \times 10^6$ copies of a monoploid (1Cx) nuclear genome was digested by HaellI enzyme (New England Biolabs, Ipswick, Massachusetts, USA). DNA fragments were size-fractionated by electrophoresis in 1.2 % agarose gel and then transferred onto HybondTM N+ nylon membranes (GE Healthcare, Chicago, Illinois, USA). Probes were prepared using *F. pratensis* genomic DNA as template and PCR with biotin-labelled dUTP (Roche, Mannheim, Germany) and specific primers (Table 3). Southern hybridization was performed at 68°C overnight and hybridization signals were detected by Chemiluminescent Nucleic Acid Module (Thermo Fisher Scientific, Waltham, Massachusetts, USA) according to manufacturer's recommendations with 90 % stringency. Hybridization signals were visualized by chemiluminiscent substrate on Medical X-Ray Film Blue (Agfa HealthCare NV, Mortsel, Belgium).

**Droplet digital PCR**

Based on the assembled DNA contigs from Fesreba retrotransposon, two restriction endonucleases with unique restriction site in the retrotransposon (*Hpa*I and *Hpa*II) were identified and used for further analysis. 3 µg of genomic DNA was digested according manufacturer's recommendations (Bio-Rad Laboratories, Hercules, California, USA) and then diluted 1,000-fold to reach starter concentration of 0.06 ng/µl. Droplet Digital PCR experiment was performed using QX200 Droplet Digital PCR machine (Bio-Rad Laboratories) following manufacturer's recommendations using EvaGreen Supermix (Bio-Rad Laboratories), template DNA and specific primers for Fesreba (Additional file 6: Table S3). Three independent replicates were done for every analyzed accession.

**Cytogenetic mapping and immunostaining**

Cytogenetic mapping of selected repeats was done by fluorescence *in situ* hybridization (FISH) on mitotic metaphase plates. Chromosome spreads were prepared according to Křivánková et al. [34] and immunostaining was done according to Neumann et al. [73]. Root tips were collected into ice water for 28 h, washed in LB01 buffer [74], fixed in 3.7% formaldehyde for 25 min and digested using 2 % cellulase, 2 % pectinase and 2 % cytohelicase in 1x PBS for 90 min at 37 °C. After squashing the meristem and coverslip removal, the slides were washed in 1x PBS and then in PBS-Trition buffer (1x PBS, 0.5% Trition X-100, pH 7.4) for 25 min and then again in 1x PBS. For incubation with anti-grass CENH3 primary antibody [75], the slides were washed in PBS-Tween buffer (1x PBS, 0.1 % Tween 20, pH 7.4) for 25 min and then incubated with anti-grass CENH3 primary antibody (diluted 1 : 200 in PBS-Tween) overnight at 4°C. Next day slides were washed 1x PBS, the CENH3 antibody was detected using the anti-Rabbit Alexa Fluor 546 secondary antibody (ThermoFisher Scientific/Invitrogen) diluted 1 : 250 in PBS-Tween buffer, for 1 h at room temperature and washed 1x PBS. Before the FISH procedure, immunofluorescent signals were stabilized using ethanol : acetic acid (3 : 1) fixative and 3.7% formaldehyde for 10 min at room temperature. FISH was performed after three washes in 1 x PBS.

Probes for FISH, derived from RT and LTR regions of Fesreba element, were labelled by digoxigenin-11-dUTP or
biotin-16-dUTP (Roche Applied Science) using PCR with specific primers (Table 3). Hybridization mixture consisting of 50% formamide, 10% dextran sulfate in 1× SSC and 1 μg/ml of each labelled probe was added onto slides and denatured at 80°C for 3 min. The hybridization was carried out at 37°C overnight. The sites of probe hybridization were detected using anti-digoxigenin-FITC (Roche Applied Science) and streptavidin-Cy3 (Thermo Fisher Scientific, Waltham, Massachusetts, USA), the chromosomes were counterstained with DAPI and mounted in Vectashield (Vector Laboratories). The slides were examined with Axio Imager.Z2 microscope (Carl Zeiss, Oberkochen, Germany) equipped with Cool Cube 1 (Metasystems, Altlussheim, Germany) camera and appropriate optical filters. The capture of fluorescence signals and merging the layers were performed with ISIS software 5.4.7 (Metasystems) and the final image adjustment was done in Adobe Photoshop 12.0.

Declarations

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Availability of data and materials

All relevant supporting datasets are included within the article and its additional files.

Competing interests

The authors declare that they have no competing interests.

Funding

This work was supported from ERDF project “Plants as a tool for sustainable global development“ (No. CZ.02.1.01/0.0/0.0/16_019/0000827). The funders had no role in the study design, data analysis and interpretation, and manuscript writing, but just provided the financial.

Authors contributions

JZ prepared DNA for sequencing, analysed Illumina sequence data, performed DNA repeat reconstruction and further analysis of repeats. JC performed flow cytometric estimation of genome size, KH, JB and BT performed Illumina sequencing, JZ, VK and AN performed cytogenetic analysis, including immuno-FISH. RS and JZ performed ddPCR and DK and MV provided plant materials. EH and JD made an intellectual contribution to the concept of the study and revised the manuscript critically for important intellectual content. All authors read and approved the final manuscript.

Acknowledgements
We thank to H. Tvardíková and E. Jahnová for technical support, P. Navrátil for IT support, and V. Černoch (DLF Seeds, Czech Republic) and M. Valárik (Institute of Experimental Botany, Czech Republic) for providing plant material. The computing was supported by the National Grid Infrastructure MetaCentrum (grant No. LM2010005 under the program Projects of Large Infrastructure for Research, Development, and Innovations).

| List Of Abbreviations       |
|-----------------------------|
| 1C  | Holoploid genome          |
| 1Cx | Monoploid genome          |
| 2C  | Nuclear DNA amount in G1 nucleus prior to DNA replication |
| 4F  | Chromosome 4 of *Festuca pratensis* cv. Fure |
| bp  | Base pairs                |
| CENH3 | Centromeric histone H3    |
| CL  | Cluster of orthologous sequences obtained by RepeatExplorer analysis |
| Cy3 | Cy3 fluorescent dye       |
| DAPI | 4′,6-diamidino-2-phenyldindeole |
| ddPCR | Droplet digital polymerase chain reaction |
| DNA | Deoxyribonucleic acid      |
| dUTP | 2′-deoxyuridine 5′-triphosphate |
| FAR | *Festuca arundinacea* Schreb. subsp. *arundinacea* |
| FGI | *Festuca gigantea* L. GR11759 |
| FGL | *Festuca arundinacea* Schreb. subsp. *glaucens*
| FISH | Fluorescence in situ hybridization |
| FITC | Fluorescein isothiocyanate |
| FMA | *Festuca mairei* GR610941  |
| FPF | *Festuca pratensis* Huds. cv. Fure |
| FPW | *Festuca pratensis* Huds. cv. Westa |
| G1  | G1 phase of cell cycle     |
| Gbp | Gigabase pairs             |
| ID  | Identity number            |
| LINE | Long interspersed nuclear element |
| LM2 | *Lolium multiflorum* cv. Kuri1 |
| LMM | *Lolium multiflorum* Lam. cv. Mitos |
LP2  *Lolium perenne* L. GR3320

LPN  *Lolium perenne* L. cv. Neptun

LTR  Long terminal repeat

µl  Microliter

NCBI  National Center for Biotechnology Information

ng  Nanogram

PBS  Phosphate-buffered saline

PCR  Polymerase chain reaction

pg  Picogram

pH  Potential of hydrogen

rDNA  Ribosomal DNA, DNA with ribosomal RNA genes

rRNA  Ribosomal RNA, RNA involved in structure of ribosomes and proteosynthesis

RT  Reverse transcriptase domain

SSC  Saline sodium citrate

TE  Transposable element

**References**

1. Inda LA, Segarra-Moragues JG, Müller J, Peterson PM, Catalán P. Dated historical biogeography of the temperate *Loliinae* (*Poaceae, Pooidae*) grasses in the northern and southern hemispheres. *Mol Phyloge net Evol.* 2008;46:932–957. doi: 10.1016/j.ympev.2007.11.022.

2. Watson L, Dawitz MJ. The grass genera of the world. C. A. B. International, Wallingford, oxon, UK. 1992.

3. Catalán P, Torrecilla P, López Rodríguez JA, Olmstead RG. Phylogeny of the festucoid grasses of subtribe *Loliinae* and allies (*Poaea, Pooidae*) inferred from ITS and trnL-F sequences. *Mol Phylogen et Evol.* 2004;31(2):517-541. doi:10.1016/j.ympev.2003.08.025

4. Torrecilla P, Catalán P. Phylogeny of broad-leaved and fine-leaved *Festuca* lineages (*Poaceae*) based on nuclear ITS sequences. *Syst Bot.* 2002;27(2):241-252. doi:1043/0363-6445-27.2.241

5. Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, et al. A worldwide phylogenetic classification of the *Poaceae* (*Gramineae*). *J Syst Evol.* 2015;53:117-137. doi:10.1111/jse.12150

6. Šmarda P, Bureš P, Horová L, Faggi B, Rossi G. Genome size and GC content evolution of *Festuca*: Ancestral Expansion and subsequent reduction. *Ann Bot.* 2008;101(3):421–433. doi:10.1093/aob/mcm307

7. Kopecký D, Havránková M, Loureiro J, Castro S, Lukaszewski AJ, Bartoš J, et al. Physical distribution of homoeologous recombination in individual chromosomes of *Festuca pratensis* in *Lolium multiflorum*. *Cytogenet Genome Res.* 2010;129(1-3):162–72. doi:1159/000313379

8. Kopecký D, Lukaszewski AJ, Doležel J. Cytogenetics of *Festulolium* (*Festuca* × *Lolium* hybrids). *Cytogenet Genome Res.* 2008;120(3-4):370–83. doi:1159/000121086

9. Loureiro J, Kopecký D, Castro S, Santos C, Silveira P. Flow cytometric and cytogenetic analyses of Iberian Peninsula *Festuca* Plant Syst Evol. 2007;269:89–105. doi: 10.1007/s00606-007-0564-8

10. Hand ML, Cogan NO, Stewart AV, Forster JW. Evolutionary history of tall fescue morphotypes inferred from molecular phylogenetics of the *Lolium-Festuca* species complex. *BMC Evol Biol.* 2010;10:303.
11. Humphreys J, Harper JA, Armstead IP, Humphreys MW. Introggression-mapping of genes for drought resistance transferred from Festuca arundinacea var. glaucescens into Lolium multiflorum. Theor Appl Genet. 2005;110:579-587. doi:10.1007/s00122-004-1879-2

12. Kopecký D, Bartoš J, Christelová P, Černoch V, Kilian A, Doležel J. Genomic constitution of Festuca x Lolium hybrids revealed by the DAeTFest array. Theor Appl Genet. 2011; 122(2):355-363.

13. Kosmala A, Zwierzykowski Z, Gasior D, Rapacz M, Zwierzykowska E, Humphreys MW. GISH/FISH mapping of genes for freezing tolerance transferred from Festuca pratensis to Lolium multiflorum. Heredity. 2006;96:243-251. doi:10.1038/sj.hdy.6800786

14. Ezquerro-López D, Kopecký D, Inda Luis A. Cytogenetic relationships within the Maghrebian clade of Festuca Schedonorus (Poaceae), using flow cytometry and FISH. Anales del Jardín Botánico de Madrid. 2017;74(1):e052. doi:10.3989/ajbm.2455

15. Czaban A, Sharma S, Byrne SL, Spannagl M, Mayer KF, Asp T. Comparative transcriptome analysis within the Lolium/Festuca species complex reveals high sequence conservation. BMC Genomics. 2015;16(1):249.

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

17. 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. 2010;10:204.

18. Leitch IJ, Bennett MD. Genome downsizing in polyploid plants. Biol J Linn Soc. 2004;82:651–663. doi:10.1111/j.1095-8312.2004.00349.x

19. Mandáková T, Joly S, Krzywinski M, Mummenhoff K, Lysák MA. Fast diploidization in close mesopolyploid relatives of Arabidopsis. Plant Cell. 2010;22(7):2277–90. doi:10.1105/tpc.109.074526

20. Renny-Byfield S, Kovařík A, Kelly LJ, Macas J, Novák P, Chase MW, et al. Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. Plant J. 2013;74(5):829–39. doi:10.1111/tpj.12168

21. Ananiev EV, Phillips RL, Rines HW. Chromosome-specific molecular organization of maize (Zea mays) centromeric regions. Proc. Natl Acad. Sci. USA. 1998;95:13073-13078. doi:10.1073/pnas.95.22.13073

22. Kumekawa N, Hosouchi T, Tsuruoka H, Kotani H. The size and sequence organization of the centromeric region of Arabidopsis thaliana chromosome 4. DNA Res. 2001;8:285–290. doi:10.1093/dnares/8.6.285

23. Li Y, Zuo S, Zhang Z, Li Z, Han J, Chu Z, et al. Centromeric DNA characterization in the model grass Brachypodium distachyon provides insights on the evolution of the genus. Plant J. 2018;93(6):1088–1101. doi:10.1111/tpj.13832

24. Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, et al. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell. 2002;14(8):1691-704. doi:10.1105/tpc.003079
31. Gorinšek B, Gubenšek F, Kordiš D. Phylogenomic analysis of chromoviruses. Cytogenet Genome Res. 2005;110(1-4):543–552. doi:10.1159/00008487

32. Hudakova S, Michalek W, Presting GG, ten Hoopen R, dos Santos K, Jasencakova Z, et al. Sequence organization of barley centromeres. Nucleic Acids Res. 2001;29:5029–5035. doi:10.1093/nar/29.24.5029

33. Neumann P, Navrátilová A, Kobližková A, Kejnovský E, Hřibová E, Hobza R, et al. Plant centromeric retrotransposons: a structural and cytogenetic perspective. Mob DNA. 2011;2(1):4. doi:1186/1759-8753-2-4

34. Křivánková A, Kopecký D, Stočes Š, Doležel J, Hřibová E. Repetitive DNA: A Versatile Tool for Karyotyping in Festuca pratensis Cytogenet Genome Res 2017;151(2):96–105. doi:10.1159/000462915

35. Paux E, Roger D, Badaeva E, Gay G, Bernard M, Sourdille P, et al. Characterizing the composition and evolution of homoeologous genomes in hexaploid wheat through BAC-end sequencing on chromosome 3B. Plant J. 2006;48(3):463–74. doi:10.1111/j.1365-313X.2006.02891.x

36. Fu K, Guo Z, Zhang X, Fan Y, Wu W, Li D, et al. Insight into the genetic variability analysis and cultivar identification of tall fescue by using SSR markers. Hereditas. 2016;153:9. doi:1186/s41065-016-0013-1

37. Koo DH, Nam YW, Choi D, Bang JW, de Jong H, Hur Y. Molecular cytogenetic mapping of Cucumis sativus and melo using highly repetitive DNA sequences. Chromosome Res. 2010;18(3):325–36. doi:10.1007/s10577-010-9116-0

38. Mehotra S, Goyal V. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genom Proteom Bioinf. 2014;12(4):164–71. doi:1016/j gpbb.2014.07.003

39. Nybom H, Weising K, Rotter B. DNA fingerprinting in botany: past, present, future. Investig Genet. 2014;5(1):1. doi:1186/2041-2223-5-1

40. Doležel J, Greilhuber J, Lucretti S, Meister A, Lysák MA, Nardi L, et al. Plant Genome Size Estimation by Flow Cytometry: Inter-laboratory Comparison. Ann Bot. 1998;82:17–26. doi:10.1093/aob/mci005

41. Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eucharyotic repetitive elements from next-generation sequence reads. Bioinformatics. 2013;29(6):792–3. doi:1093/bioinformatics/btt054

42. Presting GG, Malysheva L, Fuchs J, Schubert I. A Ty3/gypsy retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. 1998;16(6):721–728. doi:10.1111/j.1365-313X.1998.00341.x

43. Sanei M, Pickering R, Kumke K, Nasuda S, Houben A. Loss of centromeric histone H3 (CENH3) from centromeres precedes uniparental chromosome elimination in interspecific barley hybrids. Proc. Natl Acad. Sci. USA. 2011;108:E498-E505. doi:10.1073/pnas.1103190108

44. Murray BG, De Lange PJ, Ferguson AR. Nuclear DNA variation, chromosome numbers and polyploidy in the endemic and indigenous grass flora of New Zealand. Ann Bot. 2005;96(7):1293–305. doi:10.1093/aob/mci005

45. Pering GG, Malysheva L, Fuchs J, Schubert I. A Ty3/gypsy retrotransposon-like sequence localizes to the centromeric regions of cereal chromosomes. Plant J. 1998;16(6):721–728. doi:10.1111/j.1365-313X.1998.00341.x

46. Zhang Q-J, Gao L-I. Rapid and recent evolution of LTR retrotransposons drives rice genome evolution during the speciation of AA-genome Oryza G3 (Betscheda). 2017;7(6). doi:10.1534/g3.116.037572

47. McCann J, Macas J, Novák P, Stuessy TF, Villasenor JL, Weiss-Schneweiss H. Differential genome size and repetitive DNA evolution in diploid species of Melampodium Melampodium (Asteraceae). Front Plant Sci. 2020;11:362. doi:10.3389/fpls.2020.00362

48. Novák P, Neumann P, Macas J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics. 2010;11:378. doi:1186/1471-2105-11-378

49. Novák P, Hřibová E, Neumann P, Kobližková A, Doležel J, Macas J. Genome-wide analysis of repeat diversity across the family Musaceae. PLoS One 2014;9(6):e98918. doi:1371/journal.pone.0098918

50. McCann J, Jang TS, Macas J, Schneweiss GM, Matzke NJ, Novák P, et al. Dating the Species Network: Allopolyploidy and Repetitive DNA Evolution in American Daisies (Melampodium Melampodium, Asteraceae). Syst Biol. 2018; 67(6):1010–1024. doi:10.1093/sysbio/syy024

51. Renny-Byfield S, Kovařík A, Chester M, Nichols RA, Macas J, Novák P, et al. Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of Nicotiana tabacum. PLoS One. 2012;7(5):e36963. doi:1371/journal.pone.0036963

52. Macas J, Kejnovský E, Neumann P, Novák P, Kobližková A, Vyskot B. Next generation sequencing-based analysis of repetitive DNA in the model dioecious plant Silene latifolia. PLoS One. 2011;6(11):e27335. doi:1371/journal.pone.0027335

53. Said M, Hřibová E, Danilova TV, Karafiátová M, Čížková J, Friebe B, et al. The Agropyron cristatum
karyotype, chromosome structure and cross-genome homoeology as revealed by fluorescence in situ hybridization with tandem repeats and wheat single-gene probes. Theor Appl Genet. 2018;131(10):2213–2227. doi:10.1007/s00122-018-3148-9
53. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. Science. 2009;326(5956):1112–1115. doi:10.1126/science.1178534
54. International Rice Genome Sequencing Project, Sasaki T. The map-based sequence of the rice genome. Nature. 2005;436(7052):793– doi:1007/s00122-005-00714-8
55. International Barley Genome Sequencing Consortium, Mayer KF, Waugh R, Brown JW, Schulman A, Langridge P, et al. A physical, genetic and functional sequence assembly of the barley genome. 2012;491(7426):711–6. doi:10.1038/nature11543
56. Grandbastien MA, Audeon C, Bonivard E, Casacuberta JM, Chalhoub B, Costa A-PP, et al. Stress activation and genomic impact of Tnt1 retrotransposon in Solanaceae. Cytogenet Genome Res. 2005;110(1-4):229-241. doi:10.1159/000084957
57. Lee J, Waminal NE, Choi HI, Perumal S, Lee SC, Nguyen VB, et al. Rapid amplification of four retrotransposon families promoted speciation and genome size expansion in the genus Panax. Sci Rep. 2017;7(1):17986. doi:10.1126/science.1178534
58. Otto F. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA, in Crissman HA, Darzynkiewicz Z (eds): Methods in Cell Biology, Vol 33, pp 105-110. Acad Press, New York. 1990.
59. Koo DH, Tiwari VK, Hřibová E, Doležel J, Friebe B, Gill BS. Molecular cytogenetic mapping of satellite DNA sequences in Aegilops geniculata and wheat. Cytogenet Genome Res. 2016;148(4):314–21. doi:1159/000447471
60. Li B, Choulet F, Heng Y, Hao W, Paux E, Liu Z, et al. Wheat centromeric retrotransposons: the new ones take a major role in centromeric structure. Plant J. 2013;73(6):952-65. doi:1111/tpj.12086
61. Doležel J, Greilhuber J, Suda J. Estimation of nuclear DNA content in plants using flow cytometry. Nature Prot. 2007;2(9):2233–44. doi:10.1038/nprot.2007.310
62. Otto F. DAPI staining of fixed cells for high-resolution flow cytometry of nuclear DNA, in Crissman HA, Darzynkiewicz Z (eds): Methods in Cell Biology, Vol 33, pp 105-110. Acad Press, New York. 1990.
63. Doležel J, Bartoš J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. Cytometry A. 2003;51:127–128. doi:10.1002/cyto.a.10013
64. Katoh K, Toh H. Recent developments in the MAFFT multiple sequence alignment program. Brief Bioinf. 2008;9:286-298. doi:1093/bib/bbn013
65. Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performace of PhyML 3.0. Syst Biol. 2010;59(3):307-321. doi:10.1093/sysbio/syq010
66. Gouy M, Guindon S, Gascuel O. SeaView version 4: a multiplatform graphical user interface for sequence alignment and phylogenetic tree building. Mol Biol Evol. 2010;27(2):221-224. doi:1093/molbev/msp259
67. Anisimova M, Gascuel O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. Syst Biol. 2006;55:539-552. doi:10.1080/10635150600755453
68. Sonnhammer EL, Durbin R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. Gene. 1995;167(1-2):GC1-10. doi:10.1016/0378-1119(95)00714-8
69. Neumann P, Schubert V, Fuková I, Manning JE, Houben A, Macas J. Epigenetic histone marks of extended
meta-polycentric centromeres of Lathyrus and Pisum Front Plant Sci. 2016;7:234. doi:10.3389/fpls.2016.00234

74. Doležel J, Binarová P, Lucretti S. Analysis of nuclear DNA content in plant cells by flow cytometry. Biol Plantarum. 1989;31:113–120. doi:10.1007/BF02907241

75. Nagaki K, Cheng Z, Ouyang S, Talbert PB, Kim M, Jones KM, et al. Sequencing of a rice centromere uncovers active genes. Nat Genet. 2004;36(2):138-145. doi:10.1038/ng1289

Supplementary Information

Additional file 1: Table S1. List of clusters containing putative tandem repeats identified in Festuca and Lolium. (Additional_file_1_Table_S1.docx)

Additional file 2: Figure S1. Southern blots for RT domain and non-coding LTR part of the Fesreba element. (Additional_file_2_Figure_S1.tiff)

Southern blots were made with the probe for reverse transcriptase domain (A), and non-coding LTR region of the Fesreba element (B). Lanes contained genomic DNA digested by HaeIII restriction endonuclease. Lane 1: diploid F. pratensis cv. Fure; lane 2: tetraploid F. pratensis cv. Westa; lane 3: hexaploid F. arundinacea subsp. arundinacea; lane 4: hexaploid F. gigantea; lane 5: tetraploid F. glaucescens; lane 6: tetraploid F. mairei; lane 7: tetraploid L. multiflorum cv. Mitos; lane 8: diploid L. multiflorum cv. Kuri1; lane 9: tetraploid L. perenne cv. Neptun and lane 10: diploid L. perenne.

Additional file 3: Table S2. Representation of RT domain and non-coding part of LTR region of the Fesreba element estimated by ddPCR. (Additional_file_3_Table_S2.docx)

Copy number estimation of reverse transcriptase (RT) domain and non-coding part of LTR region of Fesreba element was done by droplet digital PCR. The values are averages of three independent experiments with standard deviation.

Additional file 4: Figure S2. Localization of centromeric LTR retrotransposon Fesreba on mitotic chromosomes by fluorescence in situ hybridization. (Additional_file_4_Figure_S2.tiff)

Mitotic metaphase plates were hybridized with a probe for reverse transcriptase domain of Fesreba element (A, C, E, G, I); and with a combination of probes for non-coding LTR part of Fesreba element and a probe for 45S rDNA, which served as control (B, D, F, H, J). (A, B) Avena sativa cv. Atego (2n = 2x = 14); (C, D) Secale cereale cv. Dánkowskie Diament (2n = 2x = 14); (E, F) Hordeum vulgare cv. Morex (2n = 2x = 14); (G, H) Triticum aestivum cv. Chinese Spring (2n = 6x = 42); and (I, J) Aegilops tauschii (2n = 2x = 14). Signals corresponding to 45S rDNA loci are marked by arrows. Hybridization signals of a probe for LTR region of Fesreba element are absent in all related species (D, F, H, J), except of A. sativa (B). Chromosomes were counterstained with DAPI (blue). Bar corresponds to 10 µm.

Additional file 5: Figure S3. Co-localization of CENH3 with Fesreba element in three Festuca and three Lolium species. (Additional_file_5_Figure_S3.tif)

Immunolocalization of histone H3 variant CENH3 (red) and FISH with probes derived from reverse transcriptase (RT) domain and non-coding LTR part of Fesreba element (green). F. gigantea (FGI); F. glaucescens (FGL); F. pratensis Westa (FPW); L. multiflorum Lm2 (LM2); L. perenne Neptun (LP2); and L. perenne (LPN). Column 1
shows merged images; column 2 shows CENH3 signals (red); and column 3 shows FISH signals corresponding to Fesreba element. In all accessions, the signals of CENH3 and FISH probes are overlapping. Nuclei were counterstained with DAPI (blue). Bar corresponds to 10 µm.

**Additional file 6: Table S3. Primers used for PCR amplification of DNA repeats.**
*(Additional_file_6_Table_S3.docx)*

### Table 1. Flow cytometric estimation of nuclear genome size.

| Species                      | Accession name | Code | Ploidy level |
|------------------------------|----------------|------|--------------|
| *Festuca pratensis*          | Fure           | FPF  | 2n = 2x = 14 |
| *Festuca pratensis*          | Westa          | FPW  | 2n = 4x = 28 |
| *Festuca arundinacea ssp. arundinacea* | Dulcia       | FAR  | 2n = 6x = 42 |
| *Festuca arundinacea ssp. glaucescens* | ---           | FGL  | 2n = 4x = 28 |
| *Festuca gigantea*           | GR 11759       | FGI  | 2n = 6x = 42 |
| *Festuca mairei*             | GR 610941      | FMA  | 2n = 4x = 28 |
| *Lolium multiflorum*         | Lm2            | LM2  | 2n = 2x = 14 |
| *Lolium multiflorum*         | Mitos          | LMM  | 2n = 4x = 28 |
| *Lolium perenne*             | GR 3320        | LP2  | 2n = 2x = 14 |
| *Lolium perenne*             | Neptun         | LPN  | 2n = 4x = 28 |
Table 2. Proportion of repetitive DNA sequences identified de novo.

| Repeat                  | Lineage/class          | FPF  | FPW  | FAR  | FGI  |
|-------------------------|------------------------|------|------|------|------|
| LTR retroelements       | Ty1/Copia              |      |      |      |      |
|                         | Maximus-SIRE           | 1.72 | 1.65 | 1.69 | 1.78 |
|                         | Angela                 | 4.43 | 4.53 | 3.33 | 4.86 |
|                         | TAR (Tont)             | 0.3  | 0.27 | 0.28 | 0.30 |
|                         | Tork (Tnt)             | 0.05 | 0.04 | 0.05 | 0.05 |
|                         | Ale (Hopscotch)        | 0.1  | 0.07 | 0.07 | 0.07 |
|                         | Ivana-Oryoco           | 0.05 | 0.05 | 0.03 | 0.07 |
|                         | Total Ty1/Copia        | 6.65 | 6.61 | 5.45 | 7.13 |
| Ty3/Gypsy               | Athila                 | 6.32 | 6.88 | 6.73 | 6.02 |
|                         | Chromovirideae         | 9.6  | 9.57 | 7.97 | 7.40 |
|                         | Ogre-Tat               | 12.61| 12.03| 8.65 | 8.40 |
|                         | Total Ty3/Gypsy        | 28.53| 28.48| 23.35| 21.82|
| Unclassified LTR elements |                       |      |      |      |      |
|                         |                       | 5.51 | 5.15 | 6.35 | 4.43 |
| Other                   | LINE                   | 0.26 | 0.27 | 0.29 | 0.37 |
|                         | DNA transposons        | 2.35 | 2.16 | 1.95 | 1.81 |
|                         | Tandem repeats         | 5.52 | 5.53 | 3.41 | 14.63|
|                         | rRNA genes             | 1.13 | 1.07 | 0.57 | 0.50 |
| Unclassified repeats    |                        | 13.79| 13.94| 10.82| 12.76|
Figure 1
Phylogenetic tree of Loliinae subtribe (Figure_1.pdf) Phylogeny of subtribe Loliinae with Brachypodium distachyon used as an outgroup. The tree was constructed from ITS sequence regions of Loliinae species and B. distachyon using PhyML implemented in SeaView [70]. Detailed phylogeny of subgenus Schedonorus is depicted and shows the relationships of fescue and ryegrass species within this lineage (highlighted in light yellow).
Estimation of nuclear genome size. (Figure_2.tiff) Histograms of relative nuclear DNA content obtained after flow-cytometric analysis of propidium iodide-stained nuclei isolated from (A) F. arundinacea subsp. arundinacea; (B) F. mairei; (C) Lolium multiflorum cv. Mitos and (D) L. perenne cv. Neptun. Pismum sativum cv. Ctripad (2C = 9.09 pg) and Secale cereale cv. Dankovsky (2C = 16.19 pg), respectively, were used as internal reference standards. The ratio of relative G1 peak positions was used to calculate DNA amounts of the fescue and ryegrass accessions.

Figure 3

Genome proportion of the most abundant DNA repeats (Figure_3.tiff) Genome proportion of individual repeat types was obtained as a ratio of reads specific to individual repeat type to all reads used for clustering analysis done by RepeatExplorer pipeline. Diploid Festuca pratensis cv. Fure (FPF); tetraploid F. pratensis cv. Westa (FPW); hexaploid F. arundinacea subsp. arundinacea (FAR); hexaploid F. gigantea (FGI); tetraploid F. glaucescens (FGL); tetraploid F. mairei (FMA); diploid cv. Kurl1 of L. multiflorum (LM2); tetraploid cv. Mitos of L. multiflorum (LMM); diploid L. perenne (LP2) and tetraploid cv. Neptun of L. perenne (LPN).
Figure 4

Tandem organized repeat sequences identified in cluster CL102. (Figure_4.tif) (A) Graphical layout of the cluster CL102. (B) Dot-plot analysis shows the presence of homologous tandem organized units (parallel lines) of DNA repeat identified in cluster CL102 in all species, except of F. glaucescens, in which the assembled sequence...
contigs do not represent tandem organized sequences.

Figure 5
Graphical layouts of selected DNA repeats and their validation by Southern blots. (Figure_5.tiff) Graphical layouts
were obtained after clustering analysis done by RepeatExplorer. (A) Graphical layout of cluster CL1 and (B) Cluster CL38 containing Ty3/gypsy Athila element; (C) Cluster CL20 containing Ty3/gypsy Ogre/Tat elements. Sequencing reads from Festuca species are drawn in pink while sequencing reads from Lolium species are shown in yellow. (D – F) Validation of clustering results by Southern hybridization with sequences derived from clusters CL1, CL38 and CL20. Lanes contained genomic DNA digested by HaeIII restriction endonuclease. Lane 1: F. pratensis cv. Fure (2n = 2x = 14); lane 2: F. pratensis cv. Westa (2n = 4x = 28); lane 3: F. arundinacea subsp. arundinacea (2n = 6x = 42); lane 4: F. gigantea (2n = 6x = 42); lane 5: F. glaucescens (2n = 4x = 28); lane 6: F. mairei (2n = 4x = 28); lane 7: L. multiflorum cv. Mitos (2n = 4x = 28); lane 8: L. multiflorum cv. Kuri1 (2n = 2x = 14); lane 9: L. perenne cv. Neptun (2n = 4x = 28) and lane 10: L. perenne GR 3320 (2n = 2x = 14).
Figure 6
Phylogenetic tree of Chromoviridae elements. (Figure_6.tif) The tree was constructed from a Jukes-Cantor distance matrix of the reverse transcriptase domains of Ty3/gypsy Chromoviridae elements described in Neumann et al. [33] and Fesreba elements identified in the present work using BioNJ implemented in Seaview [70]. The tree was rooted on Ty3/gypsy/Tat element. Sub-clade of Cereba element (in red) and other closely
related elements identified in different plant species are labelled in blue. Fesreba elements identified in fescues and ryegrasses are also marked in blue.
Localization of centromeric LTR retrotransposon Fesreba on mitotic metaphase chromosomes. (Figure_7.tiff)
Localization was performed in Festuca and Lolium species by fluorescence in situ hybridization (yellow-green or violet signals) using a probe for reverse transcriptase domain of Fesreba element. (A) F. arundinacea subsp. arundinacea (2n = 6x = 42); (B) F. gigantea (2n = 6x = 42); (C) F. mairei (2n = 4x = 28); (D) F. pratensis cv. Westa (2n = 4x = 28); (E) L. perenne GR 3320 (2n = 2x = 14); and (F) L. multiflorum cv. Mitos (2n = 4x = 28). Chromosomes were counterstained with DAPI (blue). Bar corresponds to 10 µm.
Co-localization of CENH3 with Fesreba element in Festuca and Lolium. (Figure_8.tiff) Combination of immunolocalization of histone H3 variant CENH3 (red) and FISH on interphase nuclei with the probes for reverse transcriptase (RT) domain and non-coding LTR part of Fesreba element (green). F. pratensis cv. Fure (FPF); F.
arundinacea subsp. arundinacea (FAR); F. mairei (FMA); and L. multiflorum cv. Mitos (LMM). Column 1 shows merged images; column 2 shows CENH3 signals (red); and column 3 shows FISH signals corresponding to Fesreba element. In all accessions, the signals of CENH3 and FISH probes are overlapping. Nuclei were counterstained with DAPI (blue). Bar corresponds to 10 µm.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile5FigureS3rev.tif
- Additionalfile3TableS2rev.docx
- Additionalfile6TableS3rev.docx
- Additionalfile2FigureS1rev.tif
- Additionalfile4FigureS2rev.tif
- Additionalfile1TableS1rev.docx