

**Plant Cytogenetics and Molecular Biology Group**  
**Institute of Biology, Biotechnology and Environmental Protection**  
**Faculty of Natural Sciences**  
**University of Silesia in Katowice**

**Dana Trunova**

**Repetitive DNA sequences and the evolution of the *Brachypodium* genus**

**Doctoral thesis**

**Supervisor: Dr hab. Bożena Kolano, profesor UŚ**  
**Auxiliary supervisor: Dr Natalia Borowska-Żuchowska**

**Katowice 2025**

**Part of the results presented in this dissertation have already been published in the following article:**

**Trunova, D.**, Borowska-Zuchowska, N., Mykhailik, S. *et al.* Does time matter? Intraspecific diversity of ribosomal RNA genes in lineages of the allopolyploid model grass *Brachypodium hybridum* with different evolutionary ages. *BMC Plant Biol* **24**, 981 (2024).  
<https://doi.org/10.1186/s12870-024-05658-5>

The author received financial support within the framework of a research project funded by the National Science Centre, Poland, project number **2020/39/O/NZ8/00184**.

## TABLE OF CONTENT

1. Introduction.....	3
1.1. The historical and evolutionary context of the grass family and its economic impact: insights from the model genus <i>Brachypodium</i> .....	3
1.2. <i>Brachypodium</i> genus: an overview .....	5
1.3. Repetitive DNA .....	9
1.3.1. Tandem repeats .....	10
1.3.2. Dispersed repeats.....	15
1.3.3. Evolution of repetitive DNA .....	18
1.3.4. Plant Genome and Repetitive DNA .....	21
1.3.5. Repetitive DNA in <i>Brachypodium</i> : the background.....	22
2. AIMS OF THE THESIS .....	25
3. Materials and Methods.....	26
3.1. Design of study .....	26
3.2. Characteristics of the plant material .....	26
3.3. Genomic DNA isolation and Illumina sequencing .....	28
3.4. Bioinformatic analyses .....	28
3.4.1. NGS data .....	28
3.4.2. Bioinformatic analysis of raw Illumina reads.....	30
3.5. PCR, Cloning and Sequencing.....	31
3.6. Restriction analyses (gCAPS).....	34
3.7. Southern blot hybridisation.....	34
3.8. Cytomolecular analyses .....	35
3.8.1. Material cultivation and preparation.....	35
3.8.2. Root meristem preparations.....	35
3.8.3. Preparation of DNA probes for FISH .....	36
3.8.4. Fluorescence <i>in situ</i> hybridisation .....	37
3.8.5. Image acquisition and processing.....	38
3.9. Phylogenetic analyses .....	38

3.9.1.	Preparation of the phylogenetic tree .....	38
3.10.	Reagents .....	39
4.	Results .....	42
4.1.	Phylogenetic relationships in the <i>Brachypodium</i> genus .....	42
4.2.	Comparative analyses of repetitive DNA elements in <i>B. hybridum</i> and its diploid progenitors .....	43
4.2.1.	Evolutionary pathways of rRNA genes in allotetraploid <i>Brachypodium hybridum</i> of different evolutionary age .....	46
4.2.2.	Satellite DNA landscape in allotetraploid <i>B. hybridum</i> and its putative diploid progenitors <i>B. distachyon</i> and <i>B. stacei</i> .....	49
4.2.3.	Composition and evolutionary dynamics of LTR retrotransposons in <i>B. hybridum</i> and its putative diploid progenitors .....	51
4.2.4.	Comparative analysis of repeatome composition in allotetraploid <i>Brachypodium hybridum</i> of different evolutionary age .....	53
4.3.	Characterisation of the repetitive DNA elements in selected diploid and allopolyploid species of <i>Brachypodium</i> genus .....	55
4.3.1.	Satellite DNA landscape across the <i>Brachypodium</i> genus .....	57
4.3.2.	Comparative analysis of 5S rDNA in <i>Brachypodium</i> sp .....	61
4.3.3.	Composition of LTR-retrotransposons in <i>Brachypodium</i> genus .....	63
4.3.4.	Comparative analysis of repeatome composition in <i>Brachypodium</i> genomes .....	65
4.4.	Comparative analysis of the repetitive DNA fraction in the phylogenetic backgrounds .....	66
5.	Discussion .....	67
5.1.	Phylogeny and basic chromosome number evolution .....	67
5.2.	Repetitive DNA sequences organisation in the allotetraploid <i>B. hybridum</i> .....	67
6.	Conclusions .....	78
7.	Summary .....	79
8.	STRESZCZENIE .....	80
9.	References .....	82
10.	Supplementary materials .....	103
	Supplementary material 1 .....	103



## 1. INTRODUCTION

### 1.1. The historical and evolutionary context of the grass family and its economic impact: insights from the model genus *Brachypodium*

The Poaceae (grasses) represents one of the most significant monocotyledonous plant families in terms of ecological and economic impact. Comprising approximately 780 genera and over 12,000 species, the Poaceae is globally ubiquitous, thriving in diverse habitats from tropical rainforests to arid deserts and temperate grasslands (Hodkinson, 2018, Kellogg, 2015). The majority of grass diversity is divided into two main clades, which are identified by acronyms of subfamilies: BEP (Bambusoideae, Ehrhartoideae (formerly Oryzoideae) and Pooideae) and PACMAD (Panicoideae, Arundinoideae, Chloridoideae, Micrairoideae, Aristidoideae and Danthonioideae; Soreng et al., 2015).

Among PACMAD, the subfamily Panicoideae includes some of the most economically important grasses, such as *Zea mays* (maize), *Saccharum officinarum* (sugarcane), *Sorghum bicolor* (great millet) and *Panicum miliaceum* (proso millet) (Hodkinson, 2018, Soreng et al., 2015). These species are crucial for global agriculture, providing staple foods and raw materials for various industries. Maize, in particular, is a major cereal crop with widespread use in food products, livestock feed, and biofuel production. Other subfamilies of PACMAD, though less economically prominent than Panicoideae, contribute to the biodiversity and stability, playing important roles in local ecosystems (Hodkinson, 2018, Soreng et al., 2015).

The BEP clade, another major clade within the Poaceae family, comprises three primary subfamilies: Bambusoideae, Ehrhartoideae, and Pooideae (Hodkinson, 2018, Soreng et al., 2015). Bambusoideae includes both woody and herbaceous bamboos, which are vital in various cultural, ecological, and economic contexts. Economically important species in this subfamily include *Phyllostachys edulis* (Moso bamboo) and *Bambusa vulgaris* (Hodkinson, 2018). Moso bamboo is extensively cultivated in China and other parts of Asia for its timber, which is used in construction, furniture making, and as a source of biomass for energy production. Additionally, bamboo shoots from various species are an important food source. Ehrhartoideae (formerly Oryzoideae) is best known for *Oryza sativa* (rice) (Hodkinson, 2018), one of the world's most critical staple crops. Rice feeds billions of people globally, particularly in Asia, and is central to economies of many countries. Pooideae is the largest subfamily within the BEP clade and contains many of the grasses adapted to temperate climates. This subfamily includes so called “core pooid” genera that contain majority economically significant cereals such as *Avena sativa* (oats) and the tribe Triticeae, that includes wheat (*Triticum*), barley (*Hordeum vulgare*) and rye (*Secale cereale*) (Fjellheim et al., 2014, Hodkinson, 2018). Wheat is a principal

global food crop, essential for bread, pasta, and numerous other food products. Barley is crucial for both food and beverage production, notably in brewing. Oats are valued for their nutritional benefits and are commonly used in food products like oatmeal and granola. Pooideae also includes important forage grasses such as *Lolium perenne* (perennial ryegrass) and *Festuca arundinacea* (tall fescue), which are extensively used in pasturelands and for turf (Hodkinson, 2018).

The studies of cereal genomes are challenging mainly due to (i) the high content of repetitive sequences that significantly impact the genome size and (ii) frequent polyploid origin (Draper et al., 2001). For these reasons many studies on grasses mainly concerned model species, which were closely related to crops; however, they have features that make them more amenable in research. The first model plant for monocots was rice (Izawa and Shimamoto, 1996, Havukkala, 1996). However, its utility as a model for temperate cereals and forage grasses was limited due its distant phylogenetic position, long life cycle, demanding growth requirements, and the lack of traits relevant to temperate crop, such as pathogen resistance and freezing tolerance (Draper et al., 2001). Therefore, *Brachypodium distachyon* emerged as a promising model species due to its close relationship with temperate cereals, its small and relatively simple genome size as well as short life cycle, and undemanding growth requirements (Draper et al., 2001). These features rendered it an optimal candidate for genome sequencing and functional genomics research (Draper et al., 2001, IBI, 2010).

*Brachypodium* is the only genus of the monotypic tribe Brachypodieae, that belongs to the Pooideae subfamily (Catalán et al., 1997, Draper et al., 2001). The separation of *Brachypodium* from its ancestral line occurred before the emergence of the "core pooids" group within the Pooideae, which encompasses temperate cereals and forage grasses (Catalán et al., 1997, Catalan et al., 2016, Draper et al., 2001). *B. distachyon* as a model has contributed significantly to genome assembly in a range of Triticeae species, including barley ( $2n = 2x = 14$ ; genome composition HH) (Mayer et al., 2012) and bread wheat (*Triticum aestivum*  $2n = 6x = 42$ ; AABBDD) (Choulet et al., 2014, Mayer et al., 2012, Pfeifer et al., 2014), *Triticum durum* ( $2n = 4x = 28$ ; AABB) (Maccaferri et al., 2019), *Triticum urartu* ( $2n = 2x = 14$ ; AA) (Ling et al., 2018), *Aegilops tauschii* ( $2n = 2x = 14$ ; DD) (Luo et al., 2017). This model grass also played an important role in the functional annotation of genes and proteins (Scholthof et al., 2018). *B. distachyon* genomic resources have been employed in comparative genomics, functional genomics, and comparative developmental studies in various other grasses, including foxtail millet (Bennetzen et al., 2012) and sugarcane (Garsmeur et al., 2018). Nowadays, *B. distachyon* constitutes a powerful research environment that includes diverse lines (wild, inbred, and mutants) and developed tools (databases, e.g., whole genome sequences, pangenome; highly

efficient transformation systems, etc.; Hasterok et al., 2022, Scholthof et al., 2018). The use of *Brachypodium* as a model system for the genomics of temperate cereals is now of limited interest due to the rise of new sequencing technologies, e.g., third-generation long-read sequencing (PacBio; OxfordNanopore; Hasterok et al., 2022, Scholthof et al., 2018). However, the whole genus *Brachypodium* is still a valuable model for the studies of polyploidy (annual *B. hybridum*) and perenniality (perennial *B. sylvaticum*) (Gordon et al., 2016, Scholthof et al., 2018, Steinwand et al., 2013).

## 1.2. *Brachypodium* genus: an overview

The genus *Brachypodium* includes approximately 20 distinct species (Catalan et al., 2016), which are predominantly found in temperate regions of the Northern Hemisphere, with a significant presence in Europe, Asia, and North Africa (Díaz-Pérez et al., 2018, Orozco-Arias et al., 2019b). *Brachypodium* grows in various habitats, ranging from open grasslands to forest groundcovers, illustrating its adaptability to diverse environmental conditions (Catalan et al., 2016). The genus consists of both annual and perennial species with different basic chromosome numbers ( $x = 5, 8, 9, 10$ ), more than 4-fold differences in genome size (0.28 – 1.29 pg./1C DNA) and several ploidy levels (from diploid to octoploid) (Díaz-Pérez et al., 2018, Gordon et al., 2016, Lusinska et al., 2019, Wolny et al., 2011).

*Brachypodium* genus includes three annual species: *B. distachyon*, *B. stacei*, and *B. hybridum* (Catalán et al., 2016). *B. distachyon* was initially considered as an autopolyploid series with  $2n = 10, 20$ , and  $30$  chromosomes (Robertson, 1981), but these were later classified as three different species, diploid *B. distachyon* ( $2n = 10$ ), diploid *B. stacei* ( $2n = 20$ ), and allotetraploid *B. hybridum* ( $2n = 30$ ) (Catalán et al., 2012). The multidisciplinary approach (cytogenetic, e.g., genomic *in situ* hybridisation (GISH) and fluorescent *in situ* hybridisation (FISH) and phylogenetic analyses of 35S rDNA internal transcribed spacer (ITS) and 35S rDNA external transcribed spacer (ETS) regions along with low-copy CAL, DGAT and GI genes) have shown that *B. hybridum* is an allotetraploid species resulted from a cross between two diploid species that resemble the modern *B. distachyon* ( $2n = 10$ ; DD) and *B. stacei* ( $2n = 20$ ; SS) (Catalán et al., 2016, Catalán et al., 2012, Hasterok et al., 2004, Hasterok et al., 2006a). Phylogenetic analyses utilizing the plastid DNA have indicated that *B. hybridum* originated from at least two separate hybridisation events between *B. distachyon* and *B. stacei* (Gordon et al., 2020). These crosses occurred bidirectionally, resulting in two distinct lineages of *B. hybridum*: one carrying plastid DNA inherited from *B. stacei* (S-plastome) and the other from *B. distachyon* (D-plastome) (Gordon et al., 2020, Mu et al., 2023). The *B. hybridum* lineages with ancestral D-plastome and S-plastome differ significantly in their time of origin,

approximately 1.4 and 0.14 Mya, respectively (Gordon et al., 2020, Mu et al., 2023). Latter studies revealed the existence of one more lineage of *B. hybridum* that was formed approximately 0.13 Mya and, has S-plastotype and is from the Eastern Mediterranean region (Mu et al., 2023).

*Brachypodium* genus also comprises approximately 17 perennial species. These perennials are very diverse in their phenotype, ploidy level and geographical distribution. From the American short-rhizomatous species *B. mexicanum* ( $2n = 40$ ), which bears more similarity to annuals than perennials, to the more recently evolved Eurasian and African long-rhizomatous diploid and polyploid species which belong to the core-perennial clade, i.e. *B. arbuscula* ( $2n = 18$ ), *B. boissieri* ( $2n = 42, 46$ ), *B. glaucovirens* ( $2n = 16$ ), *B. phoenicoides* ( $2n = 28$ ), *B. pinnatum* ( $2n = 16, 18, 28$ ), *B. retusum* ( $2n = 36, 38$ ), *B. rupestre* ( $2n = 18, 28$ ), and *B. sylvaticum* ( $2n = 18$ ) (Catalán et al., 2016, Catalan et al., 2016, Catalán et al., 2012, Catalán et al., 1995)

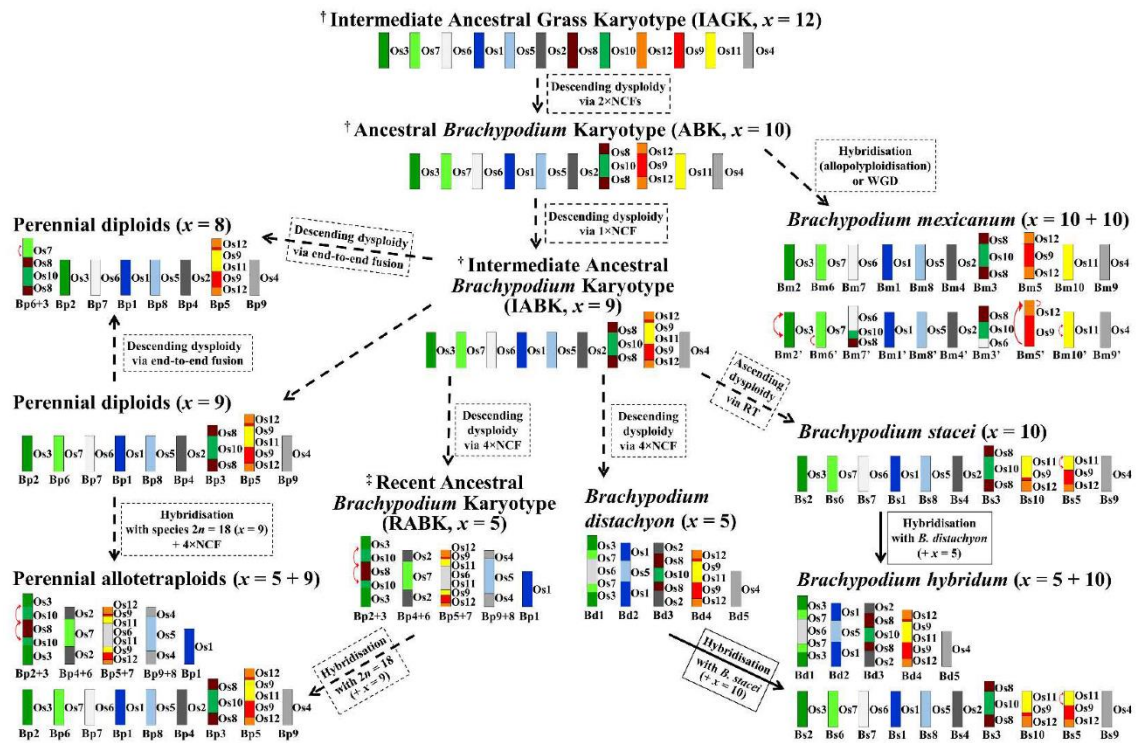
Phylogenetic analyses have placed the origin and divergence of the *Brachypodium* ancestor in the mid-Miocene, approximately 12.6 Mya (Catalan et al., 2016). The early divergence of the annual and short-rhizomatous lineages, i.e. *B. stacei*, *B. mexicanum*, and *B. distachyon* occurred in Miocene. Then, the perennial lineages experienced a burst of rapid diversification from the late Miocene through the Pleistocene, particularly within the Mediterranean and broader Eurasian regions. Radiation event gave rise to the clade known as the “core-perennials”. Within this group, *B. arbuscula* likely diverged first, followed by *B. retusum*, with later divergences giving rise to additional lineages, including *B. pinnatum*, *B. rupestre*, *B. glaucovirens*, and *B. sylvaticum* (Catalan et al., 2016). The evolutionarily youngest species is the allotetraploid *B. hybridum* (Catalan et al., 2016, Catalán et al., 2012).

Karyotypes of *Brachypodium* species have been shaped predominantly by two mechanisms: polyploidy and dysploidy (Betekhtin et al., 2014, Lusinska et al., 2019, Lusinska et al., 2018). Polyploidy refers to the presence of more than two sets of chromosomes due to whole-genome duplication, which is common in plants and can drive speciation and genetic diversity (Comai, 2005). Dysploidy, on the other hand, refers to chromosome number changes that do not involve complete sets but instead result from the fusion or fission of individual chromosomes (Mayrose and Lysak, 2020). This process can either increase (ascending dysploidy) or decrease (descending dysploidy) the chromosome number (Mayrose and Lysak, 2020).

In *Brachypodium*, both mechanisms have contributed to its chromosomal evolution, with polyploidy leading to speciation and genome complexity and dysploidy causing karyotype restructuring through changes in chromosome morphology and number (Betekhtin et al., 2014,

Lusinska et al., 2019, Lusinska et al., 2018). A model for karyotype evolution in *Brachypodium* was proposed (Lusinska et al., 2019, Lusinska et al., 2018). At first, an intermediate ancestral grass karyotype (IAGK,  $x = 12$ ) was proposed based on paleogenomic analyses (Pont et al., 2019, Murat et al., 2010). Next, descending dysploidy events gave rise to a hypothetical ancestral *Brachypodium* karyotype (ABK,  $x = 10$ ), followed by an emergence of an intermediate ancestral *Brachypodium* karyotype (IABK,  $x = 9$ ; Fig. 2) (Lusinska et al., 2019). Among the current grass species *Oryza sativa* possessed the most similar karyotype structure to the IAGK (Salse, 2016). That is why the chromosomes of the IAGK are named after the rice chromosomes (from Os1 to Os12). The ABK karyotype ( $x = 10$ ) evolved through two centric fusions (Os10 to Os8 and Os9 to Os12) (Lusinska et al., 2019). Later, the IABK karyotype ( $x = 9$ ) was formed through additional centric fusion (Os11 to combined chromosome Os9 and Os12) (Lusinska et al., 2019). *B. mexicanum* ( $2x = 10+10$ ) arose directly from ABK through polyploidisation and chromosome rearrangements. All other *Brachypodium* karyotypes are believed to have evolved from the IABK, and the proposed evolutionary mechanisms are shown in Fig. 1 (Lusinska et al., 2019).

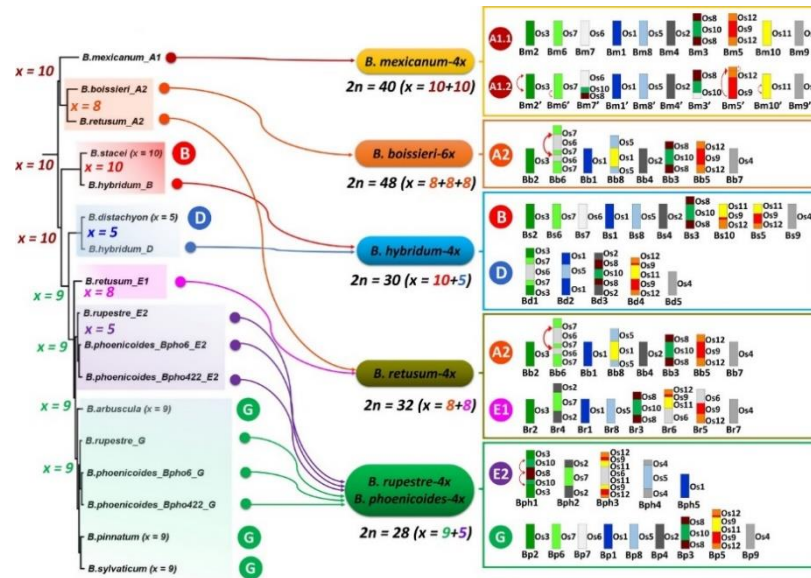
Nowadays, six different karyotypes were distinguished among the present-day *Brachypodium* species (Fig. 1) (Lusinska et al., 2019). Two different karyotypes with a basic chromosome number  $x = 5$  (*B. distachyon* and one subgenome of perennial allotetraploids with  $2n = 28$  chromosomes) and two karyotypes with  $x = 10$  (*B. stacei* and *B. mexicanum*) were distinguished (Lusinska et al., 2019). Among the diploid perennials, the most common basic karyotype has  $x = 9$  and two diploid karyotypes (*B. glaucovirens*; diploid cytotype of *B. pinnatum* (PI 185135) had  $x = 8$  (Lusinska et al., 2019). The studies on the chromosome structure of tetraploid *B. pinnatum* and *B. phoenicoides* ( $2n = 28$ ) showed that they are of an allotetraploid origin with one subgenome of  $x = 5$  (whose chromosomal pattern differ from this in the annual *B. distachyon*) and the second subgenome with  $x = 9$ , which is quite similar to the basic karyotype of the “core perennials” (Lusinska et al., 2019, Lusinska et al., 2018).



**Figure 1.** A model of karyotype evolution in the genus *Brachypodium* based on the intermediate ancestral grass karyotype. The model uses comparative chromosome barcoding data from perennial diploids, allotetraploids, *B. mexicanum*, and annual species *B. stacei* and *B. hybridum* (Lusinska et al., 2019, Lusinska et al., 2018). Os — ancestral rice chromosome equivalents (ARCEs). Genome/subgenome labels: Bd—*B. distachyon*, Bp — perennials, Bm, Bm' — *B. mexicanum*, Bs — *B. stacei*. Dashed arrows show hypothetical karyotype evolution, while solid arrows indicating the experimentally confirmed origin of *B. hybridum* [49]. Red arrows highlight minor intrachromosomal changes (inversions, translocations). † Extinct species. ‡ Extinct or unknown diploid species.

The figure is available in original article: Lusinska, J., Betekhtin, A., Lopez-Alvarez, D., Catalan, P., Jenkins, G., Wolny, E., & Hasterok, R. (2019). Comparatively Barcoded Chromosomes of *Brachypodium* Perennials Tell the Story of Their Karyotype Structure and Evolution. *International Journal of Molecular Sciences*, 20(22), 5557. <https://doi.org/10.3390/ijms20225557>. The original article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

The novel approach (phylogenomic subgenome detection; PhyloSD) has uncovered the homeologous subgenomes in current *Brachypodium* allopolyploids. Three of the subgenomes are similar to present diploids (B, D and G) and four of them have unknown origin (A1, A2, E1, E2) as shown in Figure 2 (Sancho et al., 2022).



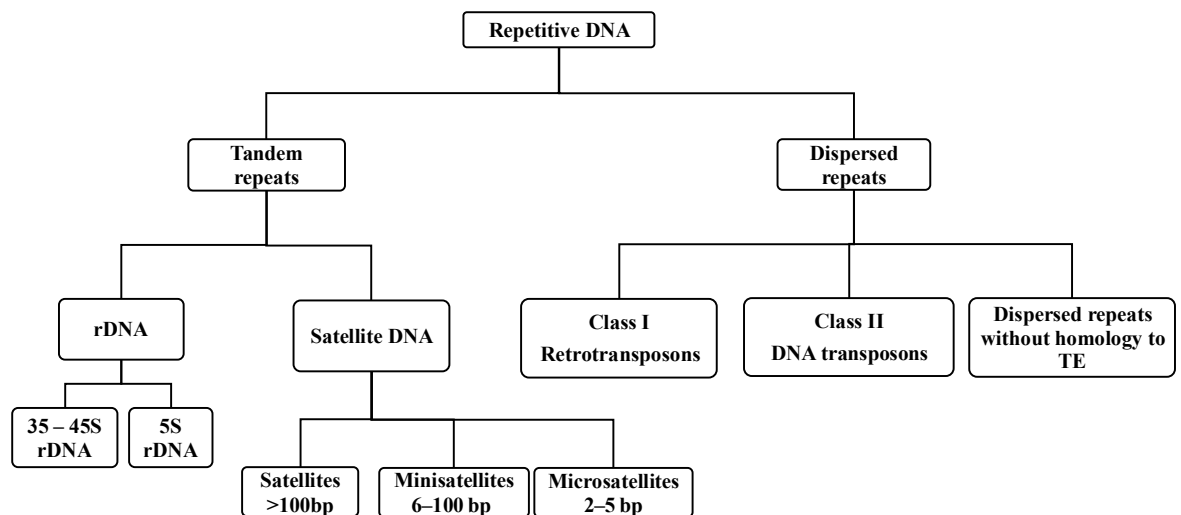
**Figure 2.** A comprehensive evolutionary framework for the origin of *Brachypodium* allopolyploids based on the combined phylogenomic and comparative chromosome barcoding (CCB) analyses. Colours indicate the different types of (sub)genomes that were retrieved in the phylogenomic analysis, and letters designate the karyotype profiles that were found in the diploids and polyploids. The arrows link the inferred (sub)genomes and karyotypes of each studied *Brachypodium* polyploid. The karyotype models are based on the CCB analysis of the *B. arbuscula* (2x), *B. retusum* (4x) and *B. boissieri* (6x) species that were analysed, and other *Brachypodium* representatives that had been previously studied (Gordon et al., 2020, Lusinska et al., 2019, Lusinska et al., 2018). Within the karyotypes, each chromosome or homeologous chromosome region corresponded to the relevant ancestral rice chromosome equivalents (Os1–Os12; Os – *Oryza sativa*; (IBI, 2010)). The basic chromosome numbers ( $x$ ) that were obtained for each genome and karyotype and inferred for the ancestors of the subgenomic tree are shown in the topology; their colours correspond to their respective (sub)genomic and karyotypic assignments. (Sub)genome designations: ‘A1’ – ancestral *B. mexicanum* (dark red); ‘A2’ – ancestral *B. boissieri* (orange); ‘B’ – *B. stacei* (red); ‘D’ – *B. distachyon* (blue); ‘E1’ – intermediate *B. retusum* (purple); ‘E2’ – intermediate *Brachypodium* core perennials (violet); ‘G’ – recent *Brachypodium* core perennials (green). Chromosome designations within the (sub)genomes: Bb – *B. boissieri*; Bd – *B. distachyon*; Bm, Bm’ – *B. mexicanum*; Bp – *Brachypodium* core perennials  $x = 9$ ; Bph – *Brachypodium* core perennials  $x = 5$ ; Br – *B. retusum*; Bs – *B. stacei*. CCB, comparative chromosome barcoding.

The figure is available in original article: Sancho, R., Inda, L.A., Díaz-Pérez, A., Des Marais, D.L., Gordon, S., Vogel, J.P., Lusinska, J., Hasterok, R., Contreras-Moreira, B. and Catalán, P. (2022), Tracking the ancestry of known and ‘ghost’ homeologous subgenomes in model grass *Brachypodium* polyploids. *Plant J*, 109: 1535–1558. <https://doi.org/10.1111/tpj.15650>. The original article is available under the [Creative Commons CC-BY-NC-ND](#) license and permits non-commercial use of the work as published, without adaptation or alteration provided the work is fully attributed.

### 1.3. Repetitive DNA

Repetitive DNA sequences once deemed non-functional or “junk DNA” are now recognized for their diverse roles in maintaining genome integrity, regulating gene expression, and driving evolutionary processes (Garrido-Ramos, 2015, Bennetzen and Wang, 2014, Mehrotra and

Goyal, 2014). These repetitive sequences are often categorized into two groups: (i) tandem repeats, such as satellite DNA (satDNA), and (ii) dispersed repeats, such as transposable elements (TEs) (Fig. 3) (Biscotti et al., 2015). In plants, the abundance of repetitive DNA varies considerably across species and is a key factor influencing genome size diversity (Borowska-Zuchowska et al., 2022). For example, transposable elements can account for up to 85.7% of the *Vicia sylvatica* genome (6.98 Gb/1C) (Macas et al., 2015), while in carnivorous *Genlisea nigrocaulis* (genome size 86 Mbp/1C) only 15.9% of the genome is represented by the repetitive DNA (Vu et al., 2015).

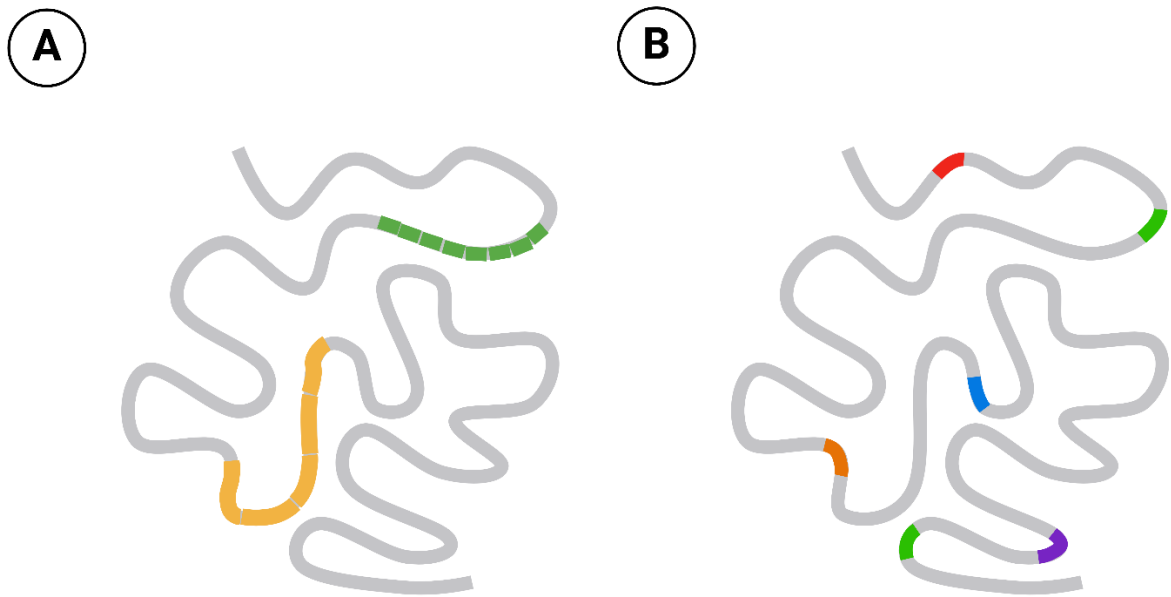


**Figure 3.** Classification of repetitive DNA sequences in plants, based on (Biscotti et al., 2015, Mehrotra and Goyal, 2014, Kubis et al., 1998).

### 1.3.1. Tandem repeats

Tandem repeats are organised in repeating units, where individual copies are arranged adjacent to each other to form tandem arrays of the monomer (Fig. 4A). Tandem repeats include ribosomal RNA genes (rRNA genes) and divergent families of satDNAs (Mehrotra and Goyal, 2014, Garrido-Ramos, 2015).



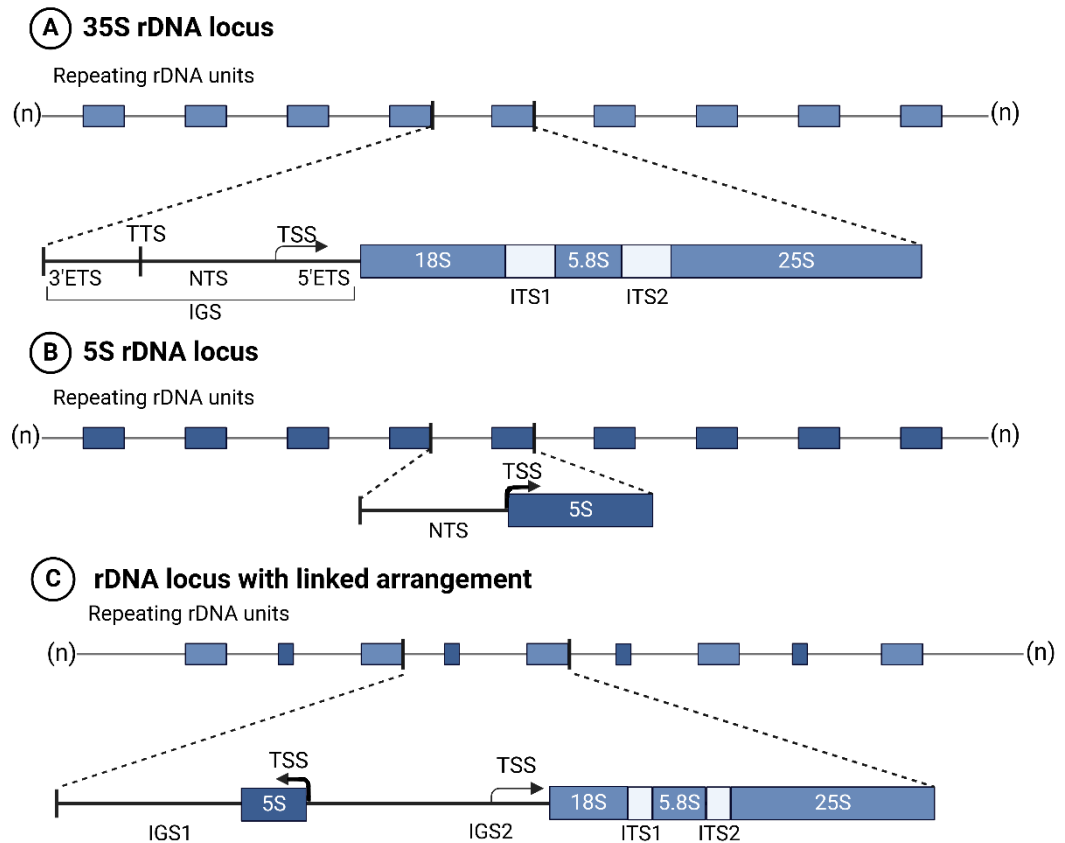


**Figure 4.** Graphical representation of genomic organization of tandem repeats (A) and dispersed repeats (B). Created with BioRender.com

#### 1.3.1.1. rRNA genes

The rRNA genes are housekeeping genes responsible for the synthesis of ribosomal RNA (rRNA) which is an essential component of ribosomes (Hemleben et al., 1988). The nuclear ribosomal DNA (rDNA) sequences (5S and 35-45S rDNA) are organized in tandem arrays located at one or multiple genomic loci (Volkov et al., 2007). Usually 35-45S rDNA and 5S rDNA are localised in separate loci (S-type rDNA arrangement) (Volkov et al., 2007, Garcia et al., 2009). However, in some species from the tribes Anthemideae and Gnaphalieae (Asteraceae), such as *Artemisia tridentata*, and in some gymnosperms (e.g., *Ephedra* L. and *Ginkgo biloba* L.), 5S rDNA units are inserted into the intergenic spacer (IGS) of the 35S rDNA, forming the so-called L-type rDNA arrangement (Garcia et al., 2010, Garcia et al., 2009, Garcia and Kovařík, 2013). Genes encoding for 18S, 5.8S, and 25S rRNAs are transcribed by DNA-dependent RNA polymerase I (PolI) together as a single operon, known as 35S-45S rDNA in plants, 48S rDNA in mammals, and 35S rDNA in yeast (Borowska-Żuchowska et al., 2023, Hori et al., 2023). The 35S rDNA unit also comprises external transcribed spacers (ETS), internal transcribed spacers (ITS1 and ITS2), and non-transcribed spacer (NTS) (Volkov et al., 2007). Within the IGS of the 35S rDNA, which includes the 3' ETS, NTS, and 5' ETS, the critical sites for transcription initiation (TSS) and transcription termination (TTS) are located (Fig. 5) (Volkov et al., 2007). The 5S rDNA unit includes 5S rRNA 120 bp coding sequence, as well as non-transcribed spacer (NTS) (Volkov et al., 2007). The 5S rRNA genes are transcribed

by DNA-dependent RNA polymerase III (PolIII). The NTS contains a TATA-like motif 28 bp upstream the transcribed region, necessary for transcription (Douet and Tourmente, 2007). This internal promoter is composed of box A, intermediate element (IE) and box C, which are recognized by PolIII. The downstream region contains the poly-T cluster used as transcription terminator (Garcia et al., 2010, Douet and Tourmente, 2007).



**Figure 5.** The structure of 5S and 35S rDNA units adapted from (Garcia et al., 2009, Volkov et al., 2007, Garcia et al., 2010). The transcription start site (TSS) marks the beginning of transcription, while the transcription termination site (TTS) indicates its end. The non-transcribed spacer (NTS) and the intergenic spacer (IGS) are shown. The external (ETS) and internal transcribed spacers (ITS1, ITS2) are marked. The linked arrangement (C) showed as in several species of *Artemisia* genus (Garcia et al., 2009, Volkov et al., 2007, Garcia et al., 2010). Created with BioRender.com

The 35S and 5S rDNA consist of highly conserved coding regions and more variable noncoding regions (Fig. 5) (Hemleben et al., 2004). The conserved regions make rDNA universal chromosomal markers for karyotype analysis (Hasterok et al., 2006b, Maragheh et al., 2019, Moreno et al., 2018), while the nuclear ribosomal internal transcribed spacer – nrITS (ITS1-5.8S-ITS2), ETS and 5S rDNA NTS are widely used in phylogenetic analyses due to their sequence variability, even among closely related species (Álvarez and Wendel, 2003, Feliner

and Rosselló, 2007). Among these, nrITS has emerged as a particularly useful marker (Feliner and Rosselló, 2007, Álvarez and Wendel, 2003). It is widely utilised in phylogenetic studies to investigate relationships at various taxonomic levels, including species, genus, and family (Ataei et al., 2020, Kolano et al., 2019, Wood et al., 2024, Chen et al., 2018).

#### **1.3.1.2. Satellite DNA**

The term satellite DNA originates from genomic DNA density gradient centrifugation experiments, which revealed additional bands of different densities when compared with the rest of the DNA sequences of the genome (Lower et al., 2018). Over time, it became clear that at least some tandem repeats play critical roles in maintaining the chromosome structure and function, notably in centromeres and telomeres (Garrido-Ramos, 2017, Garrido-Ramos, 2015). Satellite DNA is categorized into different types based on the length of the repeated units: microsatellites (2-5 bp), minisatellites (6-100 bp), and satellite DNA (>100 bp) (Mehrotra and Goyal, 2014, Biscotti et al., 2015).

Microsatellites, also known as simple sequence repeats (SSRs), are short tandem repeats consisting of 2-5 nucleotide motifs repeated multiple times (Mehrotra and Goyal, 2014). Microsatellite frequency differs among plant species, e.g., it is higher in *Arabidopsis* (the estimated frequency is 0.85%) and lower in species with comparatively larger genomes such as maize (0.37%) (Oliveira et al., 2006). Usually, SSRs are distributed on chromosomes in numerous small loci; however, in some species, such as wheat and rye, major loci may be observed after FISH, e.g., in intercalary or pericentromeric regions (Cuadrado and Schwarzacher, 1998). It was also shown that microsatellite motifs are highly clustered in the pericentromeric regions of chromosomes of tomato (Areshchenkova and Ganai, 1999), *Arabidopsis* (Brandes et al., 1997), and sugar beet (Schmidt and Heslop-Harrison, 1996). Due to their high level of polymorphism in copy number, microsatellites serve as useful molecular markers for genetic mapping, diversity studies, and breeding programs (Mehrotra and Goyal, 2014). For example, SSR markers have been employed in genome mapping in crop plants such as rice and bread wheat (Temnykh et al., 2000, Röder et al., 1995).

Minisatellites, with repeat units typically between 6 and 100 bp, are commonly found in the subtelomeric and pericentromeric regions of plant chromosomes (Garrido-Ramos, 2017). For example, PSAT3 and PSAT7 (51-bp monomer) were observed in pericentromeric chromosome region of some *Petunia* species (Alisawi et al., 2023). Another example of minisatellites are VicTR-B SatDNA subfamilies in *Vicia* species, which are characterised by subtelomeric and intercalary chromosomal localisations (Macas et al., 2006). The intercalary

distribution is not rare among minisatellite, e.g., families BvMSat08 and BvMSat09 show hybridisation signals in the intercalary heterochromatin in *Beta vulgaris* chromosomes (Zakrzewski et al., 2010). Telomeric repeats are also classified as minisatellites. Telomeres form protective caps at chromosome ends, preventing chromosome degradation by egzonucleases and fusion with other chromosomes (Aguilar and Prieto, 2021, Peska and Garcia, 2020). In plants, telomeric sequences are typically composed of the repeat motif (TTTAGGG)<sub>n</sub>, also called *Arabidopsis*-type telomeric repeat (Peska and Garcia, 2020). However, different motifs were reported for some species: (i) the vertebrate-type motif (TTAGGG)<sub>n</sub> in some plants, e.g., *Aloe* and in a few other species of *Asparagales* (Weiss and Scherthan, 2002, Puizina et al., 2003); (ii) (CTCGGTTATGGG)<sub>n</sub> was found in *Allium* (Fajkus et al., 2016); (iii) (TTTTTTAGGG)<sub>n</sub> in *Cestrum elegans* (Peška et al., 2015); and (iv) mixed sequence variants, such as (TTCAGG)<sub>n</sub> and (TTTCAGG)<sub>n</sub> in some *Genlisea* species (*Lentibulariaceae*) (Tran et al., 2015).

Satellite DNA (repeat length >100 bp) typically forms large arrays in heterochromatic regions of plant chromosomes (Mehrotra and Goyal, 2014). Satellite DNA is commonly found in subtelomeric regions of plant chromosomes and even might be telomere-associated, for example *PpTr* satellites in *Poa* species (Wei et al., 2020); PSAT1 in *Petunia* (Alisawi et al., 2023), and TcSAT1, TcSAT2, and TcSAT3 in *Tanacetum cinerariifolium* (Mlinarec et al., 2019). Rarely some satellite families might show certain conservativity through whole tribes/families of plants; for instance, the 120-bp repeat is widely distributed within the *Triticeae*, and some *Aveneae* species, and constitutes a large and evolutionarily old component of the genome (Contento et al., 2005).

The special types of satDNAs (with the monomer length usually from 150 to 180 bp) are centromeric tandem repeats, which are parts of centromere-kinetochore domain. These regions are essential for sister chromatid segregation in mitosis and the second meiotic division as well as segregation of homologous chromosomes during the first meiotic division (Biscotti et al., 2015, Garrido-Ramos, 2015, Garrido-Ramos, 2017, Naish and Henderson, 2024). The length of the satellite monomer in the centromere often corresponds to the length of DNA wrapped around one (150-180 bp) or two adjacent nucleosomes (300-360 bp) (Garrido-Ramos, 2017, Garrido-Ramos, 2015), for instance, *Arabidopsis thaliana* (CEN178 satellites are ~178 bp) (Naish et al., 2021, Wlodzimierz et al., 2023) *Arabidopsis lyrata* (CEN168 and CEN179 satellites are ~168 and ~179 bp) (Wlodzimierz et al., 2023), *Brassica rapa* (CentBr satellites are ~176 bp) (Zhang et al., 2023), *Vitis vinifera* (satellites are ~107 bp on 16 of 19 chromosomes) (Shi et al., 2023), *Oryza* species (CentO satellites are ~155 bp) (Cheng et al., 2002, Song et al., 2021). However, presence of untypically long centromeric satellites also were

reported, e.g., CEN455, CEN721, CEN1600 (with size 455, 721 and 1600 bp, respectively) in *Vigna unguiculata* (Yang et al., 2023), while untypically short motifs, e.g., GmCent (92 bp) were described in *Glycine max* (Liu et al., 2023). Centromeric repeats are often species-specific, and therefore their sequence and organization (e.g. number of copies, array length, chromosomal distribution) can differ significantly even between closely related taxa (Melters et al., 2013).

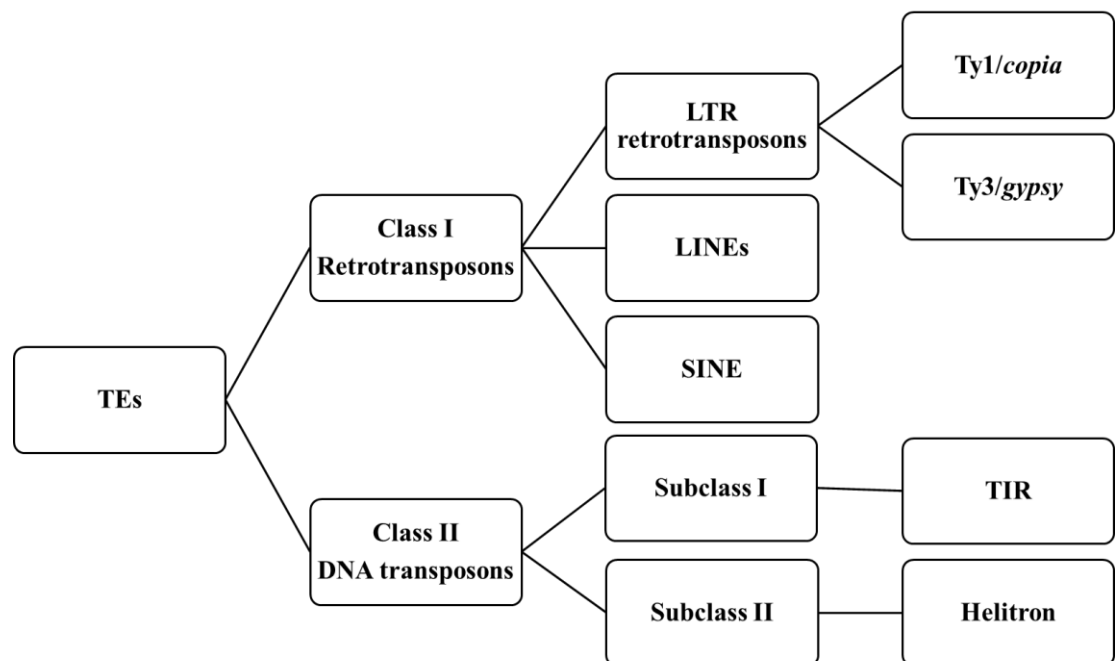
In diploids, centromeric regions often consist of a single satDNA family, arranged in arrays spanning millions of base pairs. For example, in *A. thaliana*, the 178-bp centromeric repeat (CEN180) is a prominent motif of centromeric region (Naish et al., 2021). In polyploids, centromeres may consist of different satDNA families inherited from ancestral genomes, reflecting their allopolyploid origin, for instance, in *Triticum aestivum* (common wheat), centromeres harbour diverse repeat motifs contributed by ancestral genomes (Su et al., 2019).

### 1.3.2. Dispersed repeats

Dispersed repeats are scattered throughout the genome rather than clustered in tandem arrays (Kubis et al., 1998). These repetitive sequences include predominantly transposable elements (TEs), which make up a large portion of most plant genomes (Biscotti et al., 2015). Besides those, there are dispersed repeats without homology to TEs (Kubis et al., 1998). Transposable elements are classified primarily into two classes based on their mechanism of transposition: class I (retrotransposons) and class II (DNA transposons) (Fig. 6) (Biscotti et al., 2015, Bourque et al., 2018). TEs can be either autonomous or nonautonomous (Wicker et al., 2007, Gao, 2023). Autonomous TEs encode the enzymatic machinery necessary for their own transposition, while the nonautonomous TEs rely on enzymes produced by autonomous elements (Wicker et al., 2007, Gao, 2023).

Typically, retrotransposons contain gene encoding a reverse transcriptase, and they move within the genome through a “copy-and-paste” mechanism (Almojil et al., 2021). The further classification of retrotransposons is based on the presence or absence of Long Terminal Repeats (LTR) (Fig.6) (Almojil et al., 2021). The LTRs frame the element from both ends. These regions are indispensable for transcription and the integration of the retrotransposon into the host genome. LTR-retrotransposon have two (sometimes three) open reading frames (ORF): *gag* which encodes structural proteins for the retrotransposon’s viral-like particle (Llorens et al., 2009), *pol* that encodes four proteins necessary for retrotransposition: proteinase (PROT), integrase (IN), reverse transcriptase (RT), and RNase H (RH) (Llorens et al., 2009) and *env* (envelope; in only some families of retrotransposons) encodes a protein similar to retroviruses

ones enabling the formation of a viral envelope required for cell-to-cell transmission (Llorens et al., 2009).



**Figure 6.** Classification of plant TEs following Rexdb and GyDB nomenclature. Adapted from (Schietgat et al., 2018, Llorens et al., 2010, Wicker et al., 2007, Neumann et al., 2019, Llorens et al., 2009).

The majority of LTR retrotransposons found in plant genomes belong to Ty1/*copia* and Ty3/*gypsy* superfamilies (Fig. 6) (Orozco-Arias et al., 2019a, Goodwin and Poulter, 2001, Schietgat et al., 2018, Llorens et al., 2009). These two superfamilies differ in the organisation of the *pol* ORF (Domingues et al., 2012, Llorens et al., 2009). The *pol* ORF of Ty3/*gypsy* elements is arranged as PR-RT-RH-IN, whereas in Ty1/*copia* elements, it is organized as PR-IN-RT-RH (Wicker et al., 2007, Neumann et al., 2019).

The Ty3/*gypsy* superfamily is further divided into chromovirus and non-chromovirus clades (Llorens et al., 2009, Llorens et al., 2010, Neumann et al., 2019). Chromovirus among other includes such families as CRM, Tekay, Reina (Neumann et al., 2019). Non-chromovirus includes superclade OTA (composed of elements related to retrotransposons Ogre, Tat/Retand and Athila) (Neumann et al., 2019). The Ty1/*copia* superfamily among other includes Ale, Ivana, Tork, Angela, SIRE and Tar families (Neumann et al., 2019).

The chromosomal distribution of LTR-retrotransposons depends on the element type and can be species-specific. For instance, chromoviruses (Ty3/*gypsy* subclade) are frequently localised in centromeric and pericentromeric regions of plant chromosomes (Sharma and Presting, 2014, Cheng et al., 2002). For example, the CRM family (centromeric

retrotransposons in maize) colocalised with centromere-specific histone H3 (CenH3). Since the presence of CenH3 is essential for assembly of centromere-kinetochore complex, the CRM family was proven to be a part of a functional centromere in maize (Sharma and Presting, 2014).

Non-LTR retrotransposons include long and short interspersed nuclear elements (LINEs and SINEs, respectively) (Wicker et al., 2007). In plant genomes, these elements are typically much less common than LTR retrotransposons (Schulman, 2012, Hartig et al., 2023). LINEs often possess all necessary genes for retrotransposition, while SINEs are non-autonomous retrotransposons that rely on the enzymatic machinery of LINEs for their mobility (Orozco-Arias et al., 2019a). In plant genomes, SINEs typically exhibit a dispersed chromosomal distribution (Wenke et al., 2011, Meng et al., 2020). For example, in potato, tomato, and tobacco, SolS-II elements are spread across all chromosomes except for 35S rDNA loci (Wenke et al., 2011).

Class 2 elements, also known as DNA transposons, are mobilised via a DNA intermediate, either directly through a “cut-and-paste” mechanism (Subclass I) (Bourque et al., 2018, Almojil et al., 2021) or a “peel-and-paste” (rolling-circle) replicative mechanism involving a circular DNA intermediate in the case of Helitrons (Subclass II) (Grabundzija et al., 2016, Bourque et al., 2018, Almojil et al., 2021). DNA transposons of the subclass I are flanked by terminal inverted repeats (TIRs). The size of TIRs differs among families of TIR transposons (e.g., 300 bp in Mutator or less than 50 bp in CACTA) (Feschotte and Pritham, 2007). Among TIR transposons several families (e.g., Mutator, CACTA, hAT, PIF/Harbinger) are frequently identified in plant genomes (Gao, 2023). Each TIR transposon family encodes a distinct transposase enzyme responsible for catalysing transposition, that differ in both sequence and domain structure (Gao, 2023). Most often TIR transposons contain a single open reading frame encoding transposase (Feschotte and Pritham, 2007, Gao, 2023), with an exception for PIF/Harbinger transposons, that contain two ORFs: one encoding a transposase and the other encoding a Myb-like protein, likely involved in transposition regulation (Feschotte and Pritham, 2007, Gao, 2023). Another type of DNA transposons are Miniature Inverted-Repeat Transposable Elements (MITEs), that were first discovered in maize (Bureau and Wessler, 1992). These small DNA transposons have TIRs and generally are less than 500 bp in length (Gao, 2023). Unlike traditional DNA transposons, which follow the “cut-and-paste” transposition model and are typically present in low copy numbers, MITEs are often found in extremely high copy numbers, sometimes exceeding 10,000 copies within certain families (Gao, 2023).

In plants, TIR DNA transposons demonstrate varied chromosomal localisation, often influenced by their specific family and functional characteristics (Feschotte and Pritham, 2007).

For instance, MULEs (Mutator-like elements) are widespread in the maize genome and tend to localise within gene-rich regions (Lisch, 2002, Lisch, 2015), while CACTA elements are predominantly localised in distal regions of chromosomes in wheat (Daron et al., 2014, Kunze and Weil, 2007).

### **1.3.3. Evolution of repetitive DNA**

The evolution of repetitive DNA elements has been a focal point of many scientific inquiries. Various hypotheses have been proposed to elucidate the mechanisms underlying the proliferation, elimination, and divergence of repetitive sequences within plant genomes.

#### **1.3.3.1. Hypotheses on tandem repeats evolution**

The Library Hypothesis posits that tandem repeats originate from a common ancestral pool of repeats shared among related species (Salser et al., 1976, Belyayev et al., 2020). This "library" of repeats serves as a genetic repository, where certain sequences are selectively amplified, lost or differentiating during evolution of closely related species (Salser et al., 1976, Belyayev et al., 2020). This model proposes a balance between conservation and divergence. Some repeats are preserved across taxa, while others evolve rapidly, increasing genome diversity. Evidence for this hypothesis is found in the conservation of satellite DNA families among closely related species, such as in grasses, where specific tandem repeats are shared across genera, reflecting a common ancestral origin (Mestrovic et al., 1998, Melters et al., 2013, Contento et al., 2005).

Concerted evolution and birth-and-death hypothesis offer complementary frameworks to understand the dynamics of repetitive DNA families within genomes, in particularly tandem repeats (Pinhal et al., 2011). Most of the data concerning concerted evolution and birth-and-death hypothesis has been derived from studies on rDNA. Concerted evolution describes the process in which tandem repeats, e.g., 35S rDNA, within a species are often homogenised over evolutionary time, maintaining sequence uniformity among repeats (Álvarez and Wendel, 2003, Thakur et al., 2021). This homogenisation is thought to occur primarily through mechanisms such as unequal crossing-over and gene conversion (Álvarez and Wendel, 2003, Lunerová et al., 2017). Unequal crossing-over involves misalignment of homologous chromosomes during meiosis, leading to unequal exchange of tandem repeat units (Álvarez and Wendel, 2003, Wendel et al., 1995). This process can amplify or reduce tandem repeats copy number, creating opportunities for intralocus homogenisation. Interlocus homogenisation, on the other hand, refers to the unification of tandem repeat sequences across different chromosomal loci (Wang et al., 2023, Lunerová et al., 2017). While unequal crossing-over primarily operates within a



locus, gene conversion can act between physically separated loci, facilitating sequence exchange and ensuring that all rDNA arrays within an organism retain a high degree of similarity (Wang et al., 2023, Sochorová et al., 2017). This phenomenon of concerted evolution has been shown in various plant species, with evidence suggesting that the homogenisation processes are both rapid and efficient, maintaining functionality despite extensive repeat proliferation (Álvarez and Wendel, 2003, Kovarik et al., 2005, Kovarik et al., 2004b, Bao et al., 2010).

In contrast, the birth-and-death hypothesis describes a dynamic process where tandem repeats are subjected to cycles of amplification (birth), loss (death), and divergence over evolutionary time (Rooney and Ward, 2005, Pinhal et al., 2011). According to this hypothesis, some tandem repeat variants are maintained, while others are eliminated, or new variants might arise (Pinhal et al., 2011, Rooney and Ward, 2005). The concerted evolution and birth-and-death hypothesis are not mutually exclusive but represent different scales and dynamics of tandem repeats evolution (Pinhal et al., 2011, Rooney and Ward, 2005). Concerted evolution dominates within species, ensuring sequence uniformity for functional needs, while the birth-and-death operates over longer evolutionary timescales, contributing to interspecific divergence and genome plasticity (Pinhal et al., 2011, Rooney and Ward, 2005).

#### **1.3.3.2. Mechanisms of Tandem Repeat Evolution**

One of the key mechanisms of tandem repeats, especially microsatellites, evolution is replication slippage, occurs when the DNA polymerase temporarily dissociates from the template strand and realigns incorrectly. This process can result in insertions, deletions and frameshift mutations, what can lead to the emergence and further expansion of microsatellite array (Levinson and Gutman, 1987, Paço et al., 2019, Louzada et al., 2020).

In addition, transposon-mediated mechanisms also play a role in the origin and amplification of tandem repeats (Paço et al., 2019). Transposable elements can insert repetitive sequences into various parts of the genome, contributing to the expansion and diversification of tandem repeats over time (Zattera and Bruschi, 2022). Moreover, TEs can serve as a source of novel repetitive DNA. The insertion of TEs into specific genomic locations can lead to the formation of new tandem repeats (Zattera and Bruschi, 2022).

Recombination-based mechanisms play a crucial role in the evolution of tandem repeats, particularly through illegitimate recombination (Kuhn et al., 2009), unequal cross-over (Wang et al., 2023) and ectopic recombination (Kent et al., 2017). Illegitimate recombination refers to non-homologous recombination events that occurs between DNA sequences that do not share

a similar or identical sequence. Ectopic recombination, also known as non-allelic recombination, occurs between highly similar sequences located at different genomic loci (Kent et al., 2017). This process occurs when highly similar sequences located at non-allelic positions within the genome undergo recombination, often involving repetitive DNA elements such as tandem repeats, satellite sequences, and transposable elements (Kent et al., 2017). Both, ectopic recombination and illegitimate recombination can lead to structural rearrangements, including deletions, duplications, and inversions (Kent et al., 2017).

Extrachromosomal circular DNA (eccDNA) formation is linked to ectopic recombination events. Such recombination can result in the excision of tandem or interspersed repeats, leading to the generation of eccDNA. Notably, repetitive sequences, such as satellite DNA, rDNA, and transposable elements, are particularly susceptible to these recombination events (Cohen et al., 2008, Cohen et al., 2010). Once formed, eccDNA can persist as independent elements or reintegrate into the genome through recombination (Cohen et al., 2010, Cohen et al., 2008). It has been proposed that eccDNA contributes to the reduction of tandem repeat arrays by excising repeat units, whereas its reintegration at other chromosomal loci can lead to array expansion (Cohen et al., 2008, Cohen et al., 2010).

### **1.3.3.3. Evolution of transposable elements**

The evolution of TEs include cycles that consist of: (i) the amplification phase, where active elements increase their copy number through transposition; and (ii) the phase of elimination through recombination mechanisms (Almojil et al., 2021, Bourque et al., 2018, Vitte and Panaud, 2005). These mechanisms effectively remove internal coding sequences, leaving behind a truncated, non-autonomous remnant (Devos et al., 2002). The solo-LTR elements, which were described in many plant genomes, are thought to be remnants of eliminated retrotransposons (Vitte and Panaud, 2005). Solo-LTR formation represents a major route for the gradual reduction of retrotransposon content in plant genomes (Vitte and Panaud, 2005). In DNA transposons, internal deletions or the inactivation of transposase genes can lead to the accumulation of non-functional elements that can no longer mobilise (Feschotte, 2008).

As TE sequences degenerate over time, they lose their structural integrity and become dispersed repeats without detectable homology to their ancestral elements (Novák et al., 2020a). This process occurs through the accumulation of point mutations, small insertions, and deletions, eventually leading to their recognition as generic interspersed repeats rather than identifiable TEs (Novák et al., 2020a).

#### 1.3.4. Plant Genome and Repetitive DNA

Variation in plant's genome sizes is shaped by mechanisms such as repetitive sequence amplification/elimination and to lower extent whole genome duplication (WGD) (Borowska-Zuchowska et al., 2022). A **C-value paradox** (C-value enigma) describes the lack of correlation between genome size and organismal complexity. While it was initially assumed that larger genomes correspond to greater gene numbers and complexity, comparative studies have revealed significant genome size variation across species, independent of functional gene content (Elliott and Gregory, 2015). The primary driver of genome expansion is the accumulation of TEs and satellite DNA (Bennetzen and Kellogg, 1997, Borowska-Zuchowska et al., 2022, Novák et al., 2020a). In plants with small genomes, TEs are often tightly controlled and represent a minor fraction of the genome (Novák et al., 2020a). In contrast, species with large genomes exhibit extensive TE proliferation, contributing to genomic expansion (Novák et al., 2020a).

Plants with small genomes, such as *A. thaliana* (135 Mb), typically have low TE content, often constituting less than 20% of the genome (Bennetzen and Wang, 2014, Novák et al., 2020a). These species developed efficient TE epigenetic silencing pathways which actively suppress TE activity (Slotkin and Martienssen, 2007). The presence of efficient recombination-based removal processes, such as unequal homologous recombination and illegitimate recombination, facilitates TE elimination (Maumus and Quesneville, 2014). For species with small and medium-sized genomes (up to about 10 Gbp) the linear increase in repeat-sequence genome proportion with genome size was associated with a significant increase in the proportion of higher-copy repeats, e.g., *Vicia pannonica* (6.8 pg/1C DNA), in which a specific family of Ty3/gypsy Ogre retrotransposons comprises around 38% of the genome (Novák et al., 2020a, Macas et al., 2015). Increases in genome size are often linked to the proliferation of specific retrotransposon families, such as Ty3/gypsy in the *Hesperis* clade and Fabaceae (Hloušková et al., 2019, Macas et al., 2015) or in *Solanum*, *Helianthus* and *Passiflora* species (Gaiero et al., 2019, Mascagni et al., 2017, Sader et al., 2021). Tandem repeats, although less impactful overall, can also significantly influence genome size in some species, e.g., in *Oenothera biennis* (Novák et al., 2020a, Ågren et al., 2015).

In contrast, for species with genomes larger than 10 Gbp, the proportion of single-copy and low-copy sequences typically increases with genome size, while the proportion of highly abundant repetitive sequences decreases (Novák et al., 2020a). This pattern is associated with a high content of degenerated repeats that have accumulated point mutations, insertions/deletions (InDels), and structural rearrangements (McDonald et al., 2011). As a result, these sequences are now often recognized as unique ones.

#### 1.3.4.1. Polyploidy and the fate of repetitive sequences

Whole-genome duplications (WGDs) are significant events in plant evolution that lead to changes in genome structure, function, and repeat content. Following WGDs, several processes impact the maintenance and evolution of repetitive DNA. One of the first major consequences of whole-genome duplication is the reactivation of TEs. These elements, which are often silenced or suppressed in the diploids, can be reactivated in the newly formed polyploids (Chen et al., 2020). In some polyploids, genome size and repeat content showed additive value of progenitor species (McCann et al., 2018). This suggests that whole-genome duplication does not always lead to a significant increase in the abundance of repetitive elements, e.g. in *Dactylorhiza majalis* (Eriksson et al., 2022).

Another important phenomenon following WGD is the reorganisation of repetitive sequences including amplification/loss of repetitive sequences, homogenisation, and rise of new repeat families (Chen, 2007, Parisod et al., 2010, Moreno-Aguilar et al., 2022). The dominance of one subgenome over the other is another critical factor influencing repeat evolution after WGD (Alger and Edger, 2020). In many polyploid species, one of the subgenomes often exhibits greater genomic stability and retains more of the original repeat content, while the other subgenome may undergo more extensive sequence reorganisation (Alger and Edger, 2020, Bird et al., 2021). This process of subgenome dominance affects the long-term evolutionary fate of repetitive DNA in polyploid species, contributing to the differential retention and divergence of repeat families in the genome, e.g. in *Brassica rapa* and *Gossypium hirsutum* L (Cheng et al., 2016, Fang et al., 2017). Overall, the processes following WGD – such as TE reactivation, repeat amplification/loss, homogenisation and the rise of new repeat families– play a fundamental role in shaping the repetitive DNA landscape in polyploid organisms (Otto, 2007, Parisod et al., 2010, Parisod et al., 2012).

#### 1.3.5. Repetitive DNA in *Brachypodium*: the background

The knowledge of the organisation and evolution of the repetitive sequences in *Brachypodium* is still limited, most of the data have been obtained for *B. distachyon* and *B. hybridum*. In *B. distachyon*, transposable elements accounted for approximately 28.1% of the genome (IBI, 2010, Stritt et al., 2020). Retrotransposons were the most abundant ones, contributing 23.33%, while DNA transposons constituted 4.77% of the genome (IBI, 2010). Among retrotransposons, LTR retrotransposons, especially the Ty3/gypsy family, were particularly abundant (Stritt et al., 2020). Studies on the whole-genome assemblies of both lineages of *B. hybridum* revealed that

transposable element activity following whole-genome duplication occurs gradually rather than instantaneously (Scarlett et al., 2022).

Another study of repetitive DNA content in *Brachypodium* using RepeatExplorer2 revealed that LTR/gypsy retrotransposons represented the major fractions of the repeatome in all the studied *Brachypodium* genomes (Decena et al., 2024). The work of Decena et al. shed more light on the dynamics in repetitive DNA elements, that showed rather gradual evolutionary responses than genome shock in *Brachypodium* polyploids (Decena et al., 2024). The prevalence of two families: Tat/Retand and Tekay was observed as a general trend for LTR retrotransposons content among *Brachypodium* species (Decena et al., 2024).

A notable satellite DNA sequence, CentBd, has been identified in *B. distachyon*. This 156 bp repeat is exclusively localised in centromeres. Interestingly, it was shown that this satDNA family was associated with the centromeric histone CENH3 (Li et al., 2018). Later, another subfamily of centromeric satellite – CentBs was found in *B. stacei* (Chen et al., 2024). It was also found that both subfamilies CentBd and CentBs were maintained in recent lineage of *B. hybridum* (Chen et al., 2024). Furthermore, chromatin immunoprecipitation (ChIP) analyses revealed, that centromeres in *B. distachyon* contained retrotransposon-like sequences, which exhibit high similarity to Ty3/gypsy retrotransposons. Further analyses of centromere assemblies revealed that centromeric satellite arrays were primarily disrupted by CRM retrotransposons in *B. distachyon*, *B. stacei* and the recent lineage of *B. hybridum* (Chen et al., 2024).

Among tandem repeats, rRNA genes are relatively well characterised in *Brachypodium* species. Most diploid species possess one locus of 35S and 5S rDNA per genome, although some species exhibit polymorphism in the number of rDNA loci, such as *B. sylvaticum*, which showed intraspecific variation in the number of 35S rDNA loci (Wolny et al., 2011, Wolny and Hasterok, 2009). In polyploid species, the number of rDNA loci increased accordingly, with tetraploids typically harbouring two and hexaploids having three loci (Wolny and Hasterok, 2009, Wolny et al., 2011). Research on rDNA sequences in *Brachypodium* has significantly advanced understanding of its evolution and expression, particularly focusing on the allotetraploid species *B. hybridum* and its diploid progenitors, *B. distachyon* and *B. stacei* (Borowska-Zuchowska et al., 2020). In a comprehensive cytomolecular analysis, the evolution of 35S rDNA intergenic spacers (IGSs) was investigated within recent lineage *B. hybridum* and its ancestral species (Borowska-Zuchowska et al., 2016). Distinct structural differences in the IGS regions between the progenitors were shown, suggesting that these variations may contribute to the regulation of 35S rDNA expression in the allopolyploid species (Borowska-Zuchowska et al., 2016). Moreover, nucleolar dominance (ND) was examined in *B. hybridum*,

a phenomenon where 35S rDNA loci from one progenitor are preferentially silenced (Borowska-Zuchowska et al., 2020, Borowska-Żuchowska et al., 2023). It was discovered that rDNA loci inherited from *B. stacei* are predominantly silenced throughout most of the plant's life cycle, with their inactive state maintained by DNA methylation and specific heterochromatic histone modifications, e.g., demethylation of lysine 9 of histone H3 (Borowska-Zuchowska et al., 2020). In a subsequent study, the first evidence of tissue-specific regulation of ND in *B. hybridum* was revealed (Borowska-Żuchowska et al., 2023). It was observed that while *B. stacei*-derived rDNA loci are generally silenced, they can be reactivated in certain tissues at least in some *B. hybridum* genotypes, e.g. in adventitious roots (Borowska-Żuchowska et al., 2023).

## 2. AIMS OF THE THESIS

This work aimed to shed more light on organisation and evolution of repetitive DNA in *Brachypodium* genomes. The following hypotheses were tested:

1. The composition and organisation of repetitive DNA in selected *Brachypodium* representatives is correlated with their phylogeny.
2. After hybridisation and polyploidisation, the repetitive DNA sequences undergo reorganisation in the allotetraploid *B. hybridum*
3. Different repeat lineages have different rates and/or patterns of evolution.

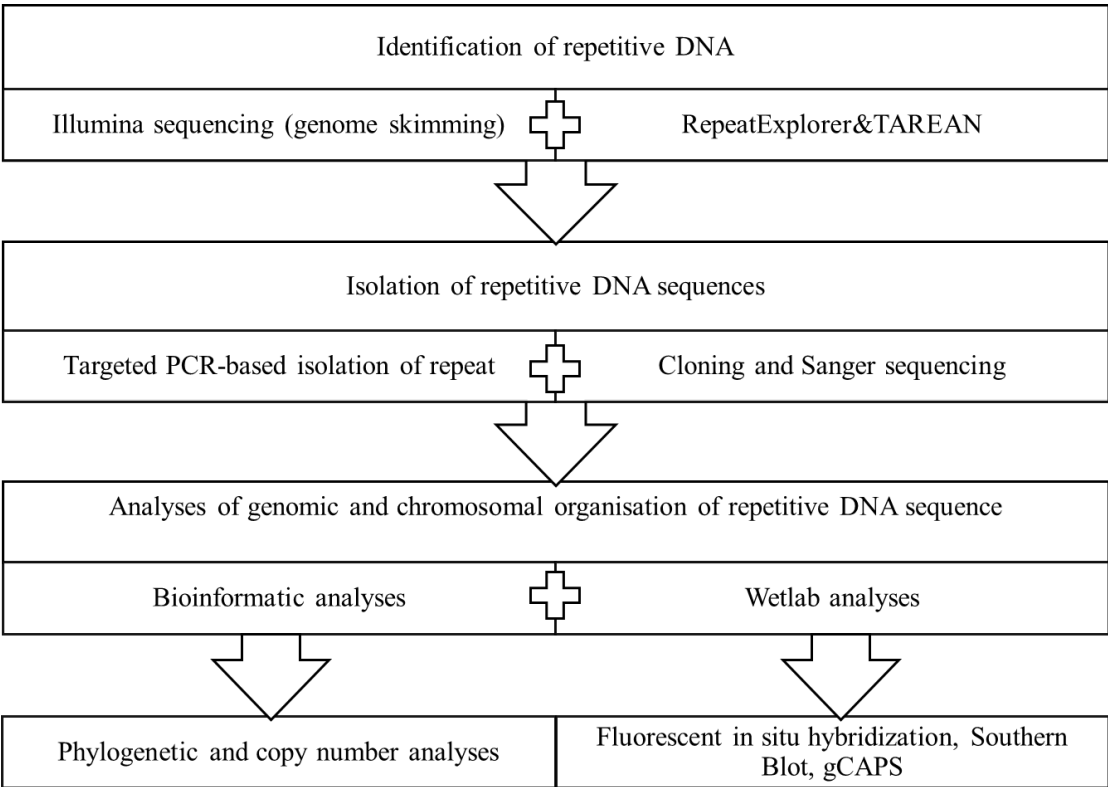
To test these hypotheses, the following tasks were planned:

- To assess the intra- and interspecific variations of the reconstructed monomers of various satellite and retrotransposon families at all levels of their organisation, i.e. DNA sequence, chromosomal localisation, genomic distribution and abundance.
- To investigate the dynamics of genome turnover in two *B. hybridum* genotypes with respect to their putative diploid progenitors and their age of origin (1.4 my and 0.14 my evolutionary lineages).
- To analyse, in a phylogenetic context, the evolutionary trajectories of various satDNAs and retrotransposon families in diploid and polyploid *Brachypodium* representatives. This work aimed to shed more light on the organisation and evolution of repetitive DNA in *Brachypodium* genomes.
- To test the origin of the selected perennial *Brachypodium* polyploids based on their repeatome organisation.

### 3. MATERIALS AND METHODS

#### 3.1. Design of study

Figure 7 presents a schematic overview of the research workflow. The diagram outlines the key steps involved in identifying and characterizing repetitive DNA elements in *Brachypodium* genomes, from genome selection and sequence pre-processing to repeat annotation, classification, and evolutionary analysis. This workflow integrates bioinformatic tools, sequence alignment strategies, and comparative genomics to elucidate the patterns and mechanisms underlying the evolution of tandem and dispersed repeats.



**Figure 7.** The scheme of the research workflow, including research phases, methods and their order.

#### 3.2. Characteristics of the plant material

18 genotypes of twelve *Brachypodium* species were used in this study (Table 1). Plants used for DNA isolation were grown from seeds in pots with soil mixed with vermiculite (3:1 v/v) at 20°C and 16/8 h photoperiod in the greenhouse of the University of Silesia in Katowice, Poland.



**Table 1.** Characteristics of *Brachypodium* species used in this study. Seeds of different species were sourced from (A) Department of Agriculture, National Plant Germplasm System, Beltsville, MD, USA; (B) Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK; (C) High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain; (D) Botanical Garden Berlin-Dahlem, Germany.

No	Species	Genotype	Chromosome Number / Basic chromosome number	Ploidy	Genome size (pg/1C DNA)	Origin	Source of seed/plants
1	<i>B. distachyon</i>	Bd21-3	$2n = 10$ ( $x = 5$ )	2x	0.32	Iraq	A
2		Bd21		2x	0.32	Iraq	A
3		ABR5		2x	0.32	Spain, Huesca	B
4	<i>B. stacei</i>	ABR114	$2n = 20$ ( $x = 10$ )	2x	0.28	Spain, Formentera	B
5		Bsta5		2x	0.28	Spain, Alicante	C
6	<i>B. hybridum</i>	ABR113	$2n = 30$ ( $2x = 5 + 10$ )	4x	0.63	Portugal, Lisbon	B
7		Bhyb26		4x	0.66	Spain, Jaen, La Cimbarrá	C
8	<i>B. arbuscula</i>	Barb502	$2n = 18$ ( $x = 9$ )	2x	0.35	Spain: Canary Isles	C
9	<i>B. sylvaticum</i>	Ain-1	$2n = 18$ ( $x = 9$ )	2x	0.43	Ain Draham, Tunisia	A
10		Bsyl434		2x	0.32	Ukraine, Crimea	A
11	<i>B. glaucovirens</i>	PI 4202	$2n = 16$ ( $x = 8$ )	2x	0.44	Greece, Crete	D
12	<i>B. pinnatum</i>	Bpin502	$2n = 16$ ( $x = 8$ )	2x	0.44	Iraq	A
13		Bpin514	$2n = 28$ ( $2x = 5 + 9$ )	4x	0.76	Turkey, Samsun	A
14	<i>B. phoenicoides</i>	Bpho513	$2n = 28$ ( $2x = 5 + 9$ )	4x	0.75	Spain,	A
15	<i>B. mexicanum</i>	Bmex347	$2n = 40$ ( $2x = 10 + 10$ )	4x	1.9	Mexico, Hidalgo	B
16	<i>B. rupestre</i>	Brup182	$2n = 38$	6x	0.91	Croatia, Istria	C
17	<i>B. boissieri</i>	Bbois10	$2n = 48$	6x	1.46	Spain, Granada	C
18	<i>B. retusum</i>	Bret551	$2n = 42$	6x	1.29	Spain, Málaga	C

### 3.3. Genomic DNA isolation and Illumina sequencing

Genomic DNA (gDNA) was extracted from young and healthy leaves using the CTAB (cetyltrimethylammonium bromide) method as described by (Doyle, 1991). Approximately 100 mg of leaves was frozen in liquid nitrogen and ground into a fine powder. The powdered tissue was transferred to a sterile 2 ml tube containing 750 µl of pre-heated to 60°C CTAB buffer, 2 µl of β-mercaptoethanol, and 3 µl of RNase (100 mg/ml) and thoroughly mixed. The mixture was incubated at 60°C for 1 h, with occasional mixing. Subsequently, 800 µl of chloroform: isoamyl alcohol (24:1) was added, mixed and incubated for 3 min. at RT. After centrifugation at 14,000 rpm for 10 min., the upper aqueous phase was transferred to a new 1.5 ml sterile tube. DNA was precipitated by adding 600 µl of ice-cold isopropanol, gently mixed, and centrifuged at 14,000 rpm for 15 min at 4 °C. The supernatant was discarded, and the DNA pellet was washed with 70% ice-cold ethanol, followed by another centrifugation at 14,000 rpm for 10 min. at 4 °C. The DNA pellet was air-dried at 37 °C for 20 min. and resuspended in 100–200 µl of DNase/RNase-free water. To remove residual RNA, RNase A was added to a final concentration of 0.1 mg/ml, and the mixture was incubated at 37 °C for 60 min. The quality of the isolated DNA was assessed via gel electrophoresis.

### 3.4. Bioinformatic analyses

#### 3.4.1. NGS data

The Illumina HiSeq data, available for annual species in The European Nucleotide Archive (ENA), were used for bioinformatic analyses: (i) for *B. distachyon* genotype Bd21-3 (SRR4236817); (ii) for *B. stacei* genotype ABR114 (SRR1800504); (iii) for *B. hybridum* genotype ABR113 (SRR3945061) and (iv) for *B. hybridum* genotype Bhyb26 (SRR4184872). Paired-end sequencing (2 × 150 bp, 350 bp insert size) of total genomic DNA of perennial species was performed using Illumina MiSeq sequencing system in the commercial facility (Genomed, Warsaw, Poland) to obtain 1×genome coverage (Table 2).

#### 3.4.2. Identification of repetitive DNA sequences

The web-based RepeatExplorer2 pipeline implemented in Galaxy server (<https://galaxy-elixir.cerit-sc.cz/>) was used to analyse repetitive DNAs in *Brachypodium* genomes (Novák et al., 2013, Novák et al., 2020c). Plastid sequences were filtered out to avoid non-nuclear DNA sequences, following standard protocols in RepeatExplorer analyses (Qiu and Ungerer, 2018). The *B. distachyon* plastid dataset (NCBI accession number: NC\_011032.1) was used to remove DNA sequences of plastid origin. Reads were then trimmed to a uniform length (150bp) by pre-processing and QC tools of RepeatExplorer2 (Novák et al., 2020c). The Illumina sequencing datasets used for clustering analysis of individual genotypes were sampled to obtain 0.1–1.0x coverage of the genome (Novák et al., 2020c). Then similarity-based clustering of Illumina

paired-end reads using the RepeatExplorer2 and TAREAN (Tandem Repeat Analyzer) pipelines were performed using default parameters (Novák et al., 2020c). The tandem repetitive nature of all clusters identified by TAREAN as putative satellite (Novák et al., 2020c) were confirmed with jDotter (Brodie et al., 2004).

**Table 2.** Characteristics of Illumina sequencing data.

Species	Genotype	Platform	Read length	Source
<i>B. distachyon</i>	Bd21-3	Illumina HiSeq 2500	250	SRR4236817
<i>B. stacei</i>	ABR114	Illumina HiSeq 2500	250	SRR1800504
<i>B. hybridum</i>	ABR113	Illumina HiSeq 2000	250	SRR3945061
	Bhyb26	Illumina HiSeq 2500	250	SRR4184872
<i>B. arbuscula</i>	Barb502	Illumina 1.9	150	This study
<i>B. sylvaticum</i>	Bsylv434	Illumina 1.9	150	This study
<i>B. glaucovirens</i>	PI 4202	Illumina 1.9	150	This study
<i>B. pinnatum</i>	Bpin502	Illumina 1.9	150	This study
	Bpin514	Illumina 1.9	150	This study
<i>B. phoenicoides</i>	Bpho513	Illumina 1.9	150	This study
<i>B. mexicanum</i>	Bmex347	Illumina 1.9	150	This study
<i>B. rupestre</i>	Brup182	Illumina 1.9	150	This study
<i>B. boissieri</i>	Bbois10	Illumina 1.9	150	This study
<i>B. retusum</i>	Bret551	Illumina 1.9	150	This study

The comparative analyses were performed for two datasets: (i) only annual species: two genotypes of *B. hybridum* (ABR113 and Bhyb26, representing two distinct evolutionary lineages) and their putative diploid progenitors *B. distachyon* (genotype Bd21-3) and *B. stacei* (genotype ABR114); (ii) both annual and perennial species (Table 2). The analyses were performed using settings recommended in the protocol (Novák et al., 2020c). For the first dataset (annuals only), 500,000 reads, obtained via random sampling using the tool “RepeatExplorer Utilities → Read sampling”, were used for each species. The results of comparative analyses were further normalised according to the genome sizes in order to establish the proportion of repetitive DNA in relation to genome sizes (Novák et al., 2020c). For the second dataset (annuals and perennials), the number of reads corresponding to 0.15× genome coverage was sampled for each species and then used for the analyses.

### 3.4.2. Bioinformatic analysis of raw Illumina reads

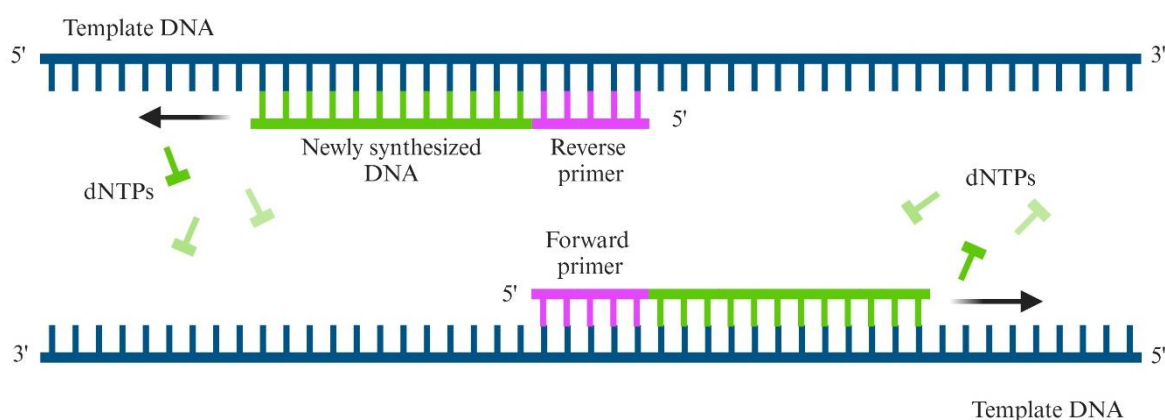
Raw Illumina reads were used to (i) determine rDNA copy number by reads count, according to (Borowska-Zuchowska et al., 2020) and (ii) to obtain the consensus sequence of IGS in Bhyb26. In order to assess the copy number of a particular repeat, at first, the 18S and 5S rDNA coding sequences were used as references in raw Illumina reads mapping using CLC Genomics Workbench (Qiagen, Hilden, Germany). The “Map Reads to Reference” tool was applied with parameters: mismatch cost 2, insertion and deletion costs 3, length fraction 0.5, and similarity fraction 0.8. The mapped reads were used to generate initial consensus sequences via the “Extract Consensus” tool. These consensus sequences were then used as references for a second round of mapping with the same parameters, producing refined consensus sequences. The copy number of either 18S or 5S rDNA units was calculated from the Illumina read count using the following formula: (i) the genome proportion (GP) was calculated from the number of mapped reads to either the 18S rDNA or 5S rDNA consensus sequences divided by the total number of reads in percentages; (ii) the genome space (GS) of 18S and 5S rDNA was determined using the formula  $GP \times \text{genome size in Mb}$ ; and (iii) the copy number was calculated as the GS value divided by the size of either 18S rDNA in Mb (0.00081 Mb) or 5S rDNA (0.000012 Mb).

The cloned IGS sequence from *B. hybridum* accession Bhyb26 served as a reference for mapping raw Illumina reads from the same accession using the ‘Map Reads to Reference’ function in CLC Genomics Workbench. Prior to alignment, low-quality reads—those containing ambiguous bases, shorter than 150 bp, or falling below a quality threshold of  $P = 0.05$ —were excluded using the software's ‘Trim’ tool. The mapping parameters were as described in the previous paragraph. Consensus sequence of the IGS was extracted and aligned with the *B. distachyon* and *B. stacei* IGSs. For further analysis, the 35S rDNA intergenic spacers were compared using DOTTER and Tandem Repeats Finder (Benson, 1999, Sonnhammer and

Durbin, 1995). Additionally, SNP identification was carried out using the 'Basic Variant Detector' function available in CLC Genomics Workbench. The analysis was conducted with the following criteria: a minimum coverage threshold of 300 reads, a minimum variant-supporting read count of 30, and a minimum allele frequency threshold of  $\geq 10\%$  for high-frequency variants, calculated as the proportion of reads supporting the alternative allele relative to the total number of informative reads at the variant site.

### 3.5. PCR, Cloning and Sequencing

The primers designed in this study were obtained using the Primer design tool in Geneious Prime software. Primers designed for satDNA had overlapping 5' ends as shown in Fig. 8.



**Figure 8.** Schematic representation of primer binding sites in consecutive monomers in satDNA repeats. Created with BioRender.com.

The PCR reactions were performed using Bio-Rad Thermal Cycler (Bio-Rad, USA). The PCR reactions were performed using Taq PCR Master Mix (EURx Ltd., Poland) according to manufacturer's instructions. PCR conditions were as follows: 95°C (3 min); 35 cycles of: 95°C (20 s), annealing temperature was adjusted for each primer pair according to manufacturer's recommendations (Table 3) and 72°C (80 s); followed by a final 72°C extension (10 min).

PCR products were separated by electrophoresis on a 1% agarose gel stained with GelRed (Sigma-Aldrich, Steinheim, Germany) and visualised with a UV transilluminator. PCR products (the bands of corresponding size) were cut from the gel and then gel-purified using the QIAquick Gel Extraction Kit (Qiagen, Germany). The amplicons were ligated using the pGEM-T Easy vector system (Promega, Madison, USA) following the manufacturer's instructions. The ligation mixture contained: 2.5 µl of reaction buffer, 0.5 µl of vector, 1.5 µl of isolated PCR product and 0.5 µl of T4 DNA ligase. The reaction mixture was incubated overnight at 4°C.

**Table 3.** The primers used in PCR reactions.

Sequence	Primer pair	Annealing temperature	Source
5S NTS region	<b>5S forward:</b> GAG TAG TAC TAG GAT GGG TGA CC <b>5S reverse:</b> ACT GCG GAG TTC TGA TGG GAT C	55°C	(Maughan et al., 2006)
35S nrITS	<b>18S dir:</b> CGT AAC AAG GTT TCC GTA GG <b>25S com:</b> AGC GGG TAG TCC CGC CTG A	55°C	(Venora et al., 2000)
18Sfor and 5.8Srev	<b>18S for:</b> GCG CTA CAC TGA TGT ATT CAA CGA G <b>5.8S rev:</b> CGC AAC TTG CGT TCA AAG ACT CGA	55°C	(Kovarík et al., 2005)
35S IGS	<b>IGS_Pr1F:</b> TTG CTG CCA CGA TCC ACT GAG <b>IGS_Pr1R:</b> CTA CTG GCA GGA TCA ACC AGG  <b>IGS_Pr2F:</b> GTC CAC TCG TGA CTT GTG ACT <b>IGS_Pr2R:</b> CGG ACG AGG CAG GAT TTC TG	54°C	(Chang et al., 2010)
			This study
CentBd, CentBs	<b>CentSat_F:</b> GGT GGA CCA TTG CTA GGG AG <b>CentSat_R:</b> ACT AGG ACG GAC AAT TCG GC	54°C	This study
Athila RT	<b>Athila_F:</b> CCT TCC AAA ACA GTG ACA GGA <b>Athila_R:</b> TAA TTG GAG GTG GCA GCT TG	52°C	This study

Tat/Retand RT	<b>Retand_F:</b> CGA ATG TGT GTC GAC TTT ACC <b>Retand_R:</b> CTC CGA ACG CGC ACT TCT	52°C	This study
Tekay RT	<b>Tekay_F:</b> AGG ATG TGT GTT GAC TAT CGT TCT <b>Tekay_R:</b> ACT CAC ATT TGC TTA ACT TGG CA	52°C	This study
CRM RT	<b>CRM_F:</b> TTC AAA GCT GGC CAA CTC CT <b>CRM_R:</b> TAG CGT CGA TCT TGG CCT TA	51°C	This study
1800bp SatFam	<b>1800_F:</b> CAA AGC GCC TTC CGA GAC TT <b>1800_R:</b> GCC ACA CCA CCA CTT AAT GC	53°C	This study
360bp SatFam	<b>360_F:</b> ACC TGA ACG GAT GGG GTA GA <b>360_R:</b> CTT GAT TGC CGC GAT TCT GG	53°C	This study

The transformation procedure followed the method described by (Hus et al., 2020). The ligation product (2 µl) was mixed with One Shot TOP10 Chemically Competent *Escherichia coli* cells (Thermo Scientific) and incubated on ice for 30 min. A heat shock at 42°C for 30 s was applied, followed by immediate cooling on ice for 2 min. The sample was then supplemented with 250 µl of pre-warmed SOC medium (Thermo Scientific) and incubated at 37 °C with shaking (250 rpm) for 1 h. Subsequently, 100 µl of the bacterial culture was spread on LB agar plates (Sigma-Aldrich) containing ampicillin (100 µg/ml, Duchefa Biochemie) and 2% X-Gal (Sigma-Aldrich) and incubated at 37°C for 16-17 h. After the blue/white screening, single white colonies were picked with sterile tips and transferred to tubes with 5 ml of LB liquid medium (Sigma-Aldrich) supplemented with ampicillin (100 mg/ml), followed by incubation at 37°C with shaking (250 rpm) for 16-17 h. Plasmids were extracted from bacterial cultures using the NucleoSpin® Plasmid kit (Qiagen). Bacterial cultures were transferred to 5 ml tubes, centrifuged at 8,000 rpm for 5 min. at RT, and the supernatant was discarded. The

pellet was resuspended in 250 µl of chilled resuspension buffer, followed by the addition of 250 µl of lysis buffer and gentle mixing by inversion. After a 4-min. incubation, 350 µl neutralization buffer was added, mixed, and centrifuged at 13,000 rpm for 10 min. The supernatant was transferred to a QIAprep spin column, centrifuged at 10,000 rpm for 30 s, and the filtrate was discarded. The column was washed with 500 µl binding buffer, followed by 750 µl washing buffer, each step including centrifugation at 10,000 rpm for 30 s. Residual washing buffer was removed by an additional centrifugation for 1 min. Finally, the plasmid DNA was eluted with 50 µl of elution buffer. Depending on the ploidy level, five to ten randomly chosen recombinant colonies were selected for further analyses. The universal M13 primer pair (forward: 5'-GCCAGGGTTTCCAGTCACGA-3' and reverse: 5'-CAGGAAACAGCTATGAC-3') was used for sequencing. Sanger sequencing was performed in a commercial facility (Macrogen, Amsterdam, Netherlands or Genomed, Warsaw, Poland). ClustalW alignment method with further manual correction was used to analyse and compare the sequences of repetitive DNA elements using BioEdit software (Hall, 1999).

### 3.6. Restriction analyses (gCAPS)

Genomic-Cleaved Amplified Polymorphic Sequence (gCAPS) analysis based on detection of polymorphism in the ITS1 region of *B. hybridum* genotypes was performed in order to verify the presence of both ancestral 35S rDNA families. The fragment of 35S rDNA containing ITS1 was amplified using PCR with primer pair 18Sfor and 5.8Srev (Table 3) originally designed by (Kovarík et al., 2005). PCR mixture consisted of 50 ng of genomic DNA, 2 µl of 10x PCR buffer (Sigma-Aldrich USA), 1.2 µl of MgCl<sub>2</sub> (Sigma-Aldrich USA), 0.8 µl of each dNTP, 0.8 µl of each primer, 0.2 µl Taq polymerase (Sigma-Aldrich USA) and ddH<sub>2</sub>O to reach the volume of 20 µl. The PCR cycling conditions included an initial denaturation at 95°C for 3 min., followed by 35 cycles of 95°C for 20 s, 60°C for 20 s, and 72°C for 80 s, with a final extension at 72°C for 10 min. A 5 µl of PCR product was digested with 5U of *Mlu*I restriction enzyme at 37°C for 2 h and subsequently visualized on a 1% agarose gel. The ITS1 region in *B. distachyon* contains a single *Mlu*I restriction site, whereas the ITS1 region in *B. stacei* remains uncut by the *Mlu*I enzyme. To assess digestion efficiency, PCR products from *B. stacei* and *B. distachyon* were included as controls.

### 3.7. Southern blot hybridisation

Genomic organization of selected repeats was analysed using Southern blot. Approximately, 2 µg of gDNA was subjected to restriction using the following conditions:



2 µg of genomic DNA  
1 µl restriction enzyme  
2 µl 10x restriction enzyme buffer  
ddH<sub>2</sub>O (to reach total volume 20 µl)

The digestion mixture was incubated for approximately 4 h at temperature recommended by manufacturer. The 20 µl of digestion products with 5 µl of loading dye were separated by electrophoresis on a 1.2% agarose gel stained with GelRed, using 40V for several hours. The gel was depurinated for 15 min. in 0.25 M HCl and then washed in denaturing buffer twice for 15 min, and then in neutralising buffer for 10 min. three times. The DNA was then blotted onto positively charged nylon membrane (Roche, Mannheim, Germany) using the VacuBlot System (Biometra Analytik Jena, Jena, Germany). DNA was UV cross-linked to the membrane using a CK-1000 Ultraviolet Crosslinker (Ultra-Violet Products, Cambridge, UK). The clones used as probes were labelled with alkali-labile digoxigenin-11-dUTP, with nick translation according to manufacturer's instructions (Roche, Basel, Switzerland). Hybridisation was performed using a DIG High Prime DNA Labelling and Detection Starter Kit II (Roche, Mannheim, Germany) according to the manufacturer's instructions using an HB-1000 Hybridizer (UltraViolet Products, Cambridge, UK). The hybridisation was conducted overnight at 37°C. Then, the membrane was washed in 1× SSC (saline sodium citrate buffer) with 0.1% SDS at 65°C. Hybridisation signals were visualised using ChemiDocXRS (BioRad, USA).

### **3.8. Cytomolecular analyses**

#### **3.8.1. Material cultivation and preparation.**

Primary roots were used for cytogenetic analyses in the case of annual species, while adventitious roots were used in cytogenetic studies on perennials. Seeds were germinated on a filter paper moistened with water for 3-5 days at RT in the dark. Whole seedlings with approximately 2 cm-long roots (annuals) or adventitious roots (perennials) were placed in ice-cold water for 24 h and then fixed in 3:1 (v/v) methanol:acetic acid.

#### **3.8.2. Root meristem preparations**

Cytogenetic preparations were carried out using primary and adventitious roots, following the protocol outlined by (Jenkins and Hasterok, 2007). Roots were initially washed in 0.01 M citrate buffer (pH 4.8) for 15 min at RT. Enzymatic digestion of the roots was then performed at 37°C for 1–1.5 h using a mixture containing 6% (v/v) pectinase (Sigma-Aldrich, St. Louis, MO, USA), along with 0.5% (w/v) cellulase (Sigma-Aldrich) and 0.5% (w/v) "Onozuka R-10" cellulase (Serva, Heidelberg, Germany). Following digestion, the meristems were washed in citrate buffer for 10 min. on ice before being dissected from the root tips. The dissected

meristems were squashed in 45% acetic acid, frozen on dry ice, and the coverslips were carefully removed. The resulting slides were air-dried and examined under a phase-contrast microscope. Only slides with a high number of metaphase plates and minimal cytoplasmic residue were selected for further analysis.

### 3.8.3. Preparation of DNA probes for FISH

For rDNA analysis two probes were used in the study: (i) a 410-bp-long clone pTa794 from *T. aestivum* (Gerlach and Dyer, 1980) was used to visualise the 5S rDNA loci; (ii) a 2.3 kb fragment of the 25S rRNA gene from *A. thaliana* (Untried and Gruendler, 1990) was used to detect the 35S rDNA loci. The following newly obtained and cloned satDNAs were used as probes: (i) a 156-bp-long clone CentBd from *B. distachyon*; (ii) 159-bp-long clone CentBp from *B. arbuscula*; and (iii) 359-bp-long clone 360bpSatFam from *B. arbuscula*. Clones of conserved domain of reverse transcriptase (RT), isolated in this study, were used to detect the chromosomal distribution of LTR-retrotransposons: (i) 519-bp-long fragment of Athila RT from *B. stacei*; (ii) 430-bp-long fragment of Tat/Retand RT from *B. distachyon*; (iii) 437-bp-long fragment of Tekay RT from *B. distachyon*; and (iv) 417-bp-long fragment of CRM RT from *B. distachyon*. The cloned sequences were labelled either with tetramethylrhodamine-5-dUTP (TAMRA-dUTP, Roche, Basel, Switzerland) or digoxigenin-11-dUTP (DIG-dUTP, Roche) using Nick Translation Mix labelling kit (Roche). The mixture used for the nick-translation reaction is provided in Table 4.

**Table 4.** Nick-translation mixture

Components	Volume
dATP (0.4 mM)	2.5 µl
dTTP (0.4 mM)	1.67 µl
dCTP (0.4 mM)	2.5 µl
dGTP (0.4 mM)	2.5 µl
DNA template (100 ng/µl)	6 µl
Tetramethylrhodamine-5-dUTP or digoxigenin-11-dUTP	0.83 µl
10x reaction buffer and nick-translation enzymes	4 µl
Total volume	20 µl

The nick-translation reaction was conducted in the thermocycler with the following conditions: 95 min at 15°C and 10 min at 60°C to terminate the reaction. The probes were

precipitated by adding 2  $\mu$ l of 3 M sodium acetate and 50  $\mu$ l of ice-cold 100% ethanol and incubated at -20°C overnight. The tubes were then centrifuged at 14,000 rpm for 30 min at 4°C, and the supernatant was removed. The pellet was washed twice with 70% ethanol and centrifuged at 14,000 rpm for 5 min at 4°C. After drying, the pellet was dissolved in 10  $\mu$ l of sterile distilled water. The probes were stored at -20°C until use.

#### 3.8.4. Fluorescence *in situ* hybridisation

The slides were incubated with RNase (100  $\mu$ g/ml) in 2 $\times$  SSC in a humid chamber for 1 h at 37°C. Then, the slides were washed three times with 2 $\times$  SSC buffer for 5 min each and fixed in 1% formaldehyde in 1 $\times$  PBS for 10 min at 37°C. Then, the slides were washed three times in 2 $\times$  SSC, dehydrated in ethanol series (70%, 90%, and 99%), and air-dried. The hybridisation mixture composition is provided in Table 5.

**Table 5.** FISH hybridisation mixture.

Components	Volume
100% deionised formamide	20 $\mu$ l
50% (w/v) dextran sulphate	8 $\mu$ l
20 $\times$ SSC	4 $\mu$ l
10% (w/v) SDS	2 $\mu$ l
DNA probe (75-200 ng/slide)	1-6 $\mu$ l
Sterile distilled water	0-5 $\mu$ l
Total volume	40 $\mu$ l

The hybridisation mixture was denatured for 10 min at 80°C and stabilised on ice for 10 min. Then, 40  $\mu$ l of the hybridisation mixture was applied to each slide and covered with plastic coverslips. The probe and chromosomes were denatured for 4.5 min at 75°C using an *in situ* Hybrid OmniSlide Thermal Cycler System (Fischer Thermo Scientific, Waltham, MA, USA). After denaturation, the slides were placed in a humid chamber and incubated at 37°C for approximately 48 h. After hybridisation, the coverslips were removed by incubating slides with 2 $\times$  SSC at 42°C. Then, stringent washes were performed. The slides were incubated twice in 15% formamide in 0.1 $\times$  SSC at 42°C for 5 min (corresponding to a stringency of 82%). The slides were washed three times in 2 $\times$  SSC at 37°C, followed by three washes in 2 $\times$  SSC at RT. Then, the slides were transferred to 4 $\times$  SSC with 0.2% Tween20 for 5 min. at RT. Next, 200  $\mu$ l of blocking reagent was applied to each slide and covered with plastic coverslips. After 20 min. of incubation at RT, 40  $\mu$ l of FITC-conjugated anti-DIG antibody (Roche) was applied to each

slide. The plastic coverslips were applied, and the slides were incubated in a humid chamber for 1-2 h at 37°C. Then, the slides were washed in 4× SSC with 0.2% Tween20 for 10 min. at 37°C. This step was repeated three times. Then, slides were dehydrated in ethanol series (70%, 90%, and 99%) and air-dried. The dried slides were sealed in Vectashield buffer (Vector Laboratories, Newark, CA, USA) with DAPI (4',6-diamidino-2-phenylindole, Serva) at a concentration of 2.5 µg/ml and stored in the dark at 4°C until use.

### 3.8.5. Image acquisition and processing

Photomicrographs were acquired using AxioImager.Z.2 (Zeiss, Oberkochen, Germany) wide-field epifluorescence microscope equipped with a monochromatic AxioCam HRm (Zeiss) camera. The brightness and contrast of images were adjusted uniformly. The images from each channel were taken separately and merged in Photoshop CS3 (Adobe, San Jose, CA, USA) and ZEN Lite (Zeiss).

## 3.9. Phylogenetic analyses

### 3.9.1. Preparation of the phylogenetic tree

To analyse the evolution of repetitive DNA within a phylogenetic context, a phylogenetic tree of the *Brachypodium* genus was constructed using *Gigantea* gene (GI) as a molecular marker. The sequences for 13 accessions used in the analyses (Table 6) were published earlier (Díaz-Pérez et al., 2018). The GI sequences for *B. hybridum* subgenomes and for *B. sylvaticum* Bsyl434 were assembled from Illumina reads using mapping to reference (as described in chapter 3.4.3 for IGS). For *B. hybridum* subgenomes GI sequences of its putative diploid progenitors *B. distachyon* and *B. stacei* were used as reference in mapping analysis. For *B. sylvaticum* Bsyl434, the GI sequence of different genotype – *B. sylvaticum* Ain1 was used as reference.

**Table 6.** *Gigantea* sequences used in phylogenetic analysis

Species	Genotype	GenBank accession number
<i>B. distachyon</i>	Bd21-3	JX666047
<i>B. stacei</i>	ABR114	HQ890969
<i>B. hybridum</i>	ABR113	reconstructed <i>de novo</i> from Illumina reads
	Bhyb26	reconstructed <i>de novo</i> from Illumina reads
<i>B. arbuscula</i>	Barb502	KP709902-KP709906

<i>B. sylvaticum</i>	Ain1	KP710037-KP710041
	Bsyl434	reconstructed <i>de novo</i> from Illumina reads
<i>B. glaucovirens</i>	PI 4202	KP709927
<i>B. pinnatum</i>	Bpin502	KP709985-KP709989
	Bpin514	KP709990-KP709992
<i>B. phoenicoides</i>	Bpho513	KP709960-KP709969
<i>B. mexicanum</i>	Bmex347	HQ890968-HQ890971
<i>B. rupestre</i>	Brup182	KP710023-KP710027
<i>B. boissieri</i>	Bbois10	KP709915-KP709920
<i>B. retusum</i>	Bret551	KP709997-KP710006

Multiple sequence alignments for analysis were performed using webPRANK (Löytynoja and Goldman, 2010) with further manual correction. Phylogenetic relationships were inferred using maximum likelihood analyses (ML) as implemented in IQ-TREE v.2 (Minh et al., 2020). The most appropriate model of sequence evolution for the ML analyses was determined using the Bayesian information criterion as implemented in IQ-TREE. The best-fit model was HKY+F. The significance of the inferred relationships was assessed via bootstrapping with 1000 replicates. *O. sativa* was used as an outgroup.

The unrooted phylogenetic trees were reconstructed to estimate the subfamilies' composition of tandemly repeated DNA. Multiple sequence alignment and the phylogenetic analysis were performed as described before. The best-fit model for analysis of *Brachypodium* cloned centromeric satellite was K3Pu+F+G4. The best-fit model for analysis of centromeric satellite based on consensus sequences was JC+G4. The best-fit model for analysis of 360bp satellite based on consensus sequences was K2P+G4. The analysis of the ancestral state of characters (genome size) was performed using the Maximum parsimony (MP) method as implemented in Mesquite (Maddison, 2008).

### 3.10. Reagents

#### Enzymatic mixture for slide preparation

6% pectinase (Sigma-Aldrich, St. Louis, MO, USA) 0.6 ml

0.5% Onozuka R-10 cellulase (Serva, Heidelberg, Germany) 0.05 g

0.5% w/v cellulase (Sigma-Aldrich) 0.05 g

#### Ampicillin (100 mg/ml)

100 mg of ampicillin dH<sub>2</sub>O - up to 1 ml

Sterilised through a 0.45 µm filter

### **CTAB buffer**

2% hexadecyltrimethylammonium bromide (CTAB)

20 mM EDTA

100 mM Tris-HCl, pH 8

1.4 M NaCl

### **LB Agar medium with ampicillin, X-Gal and IPTG**

8.75 g of LB Agar

dH<sub>2</sub>O to 250 ml

Sterilised in an autoclave.

After sterilisation, ampicillin (100 mg/ml) was added to the cooled medium at a ratio of 1 µl of antibiotic per 1 ml of LB medium. 30 ml of the medium was poured into each dish and left to solidify. 40 µl of 2% X-Gal (Sigma-Aldrich) and 7 µl of 20% IPTG (Sigma-Aldrich) were added to the solid medium. The mixtures were spread over the medium using a sterile spreader, and the dishes were then placed in an incubator at 37°C for 30 min.

### **LB Broth medium with ampicillin**

5 g of LB Broth

dH<sub>2</sub>O to 250 ml

Sterilised in an autoclave.

After sterilisation, ampicillin (100 mg/ml) was added to the cooled medium at a ratio of 1 µl of antibiotic per 1 ml of LB medium.

### **RNase (10mg/ml)**

10 mg RNase A

1 ml of a solution of 10 mM Tris-HCl + 15 mM NaCl

The RNase solution was stored at -20 °C. For slide pretreatment in FISH, the stock solution was diluted 1:1000 with 2× SSC

### **1% formaldehyde in 1× PBS**

6 ml formaldehyde (37%)

20 ml 10× PBS, pH 7.0

174 ml of dH<sub>2</sub>O

### **10× PBS, pH 7.0**

Solution (A): 0.1 M Na<sub>2</sub>HPO<sub>4</sub> + 1.4 M NaCl

Solution (B): 0.1 M NaH<sub>2</sub>PO<sub>4</sub> + 1.4 M NaCl

Solution A and B was mixed until pH 7.0 was reached and sterilised in an autoclave.

**20× SSC, pH 7.0**

175.3 g 3 M NaCl

88.3 g 0.3 M sodium citrate  $C_6H_5Na_3O_7 \times 2H_2O$

dH<sub>2</sub>O was added to the volume of 1 litre, and pH was adjusted with 1N HCl and sterilised in an autoclave.

**Maleic buffer pH 7.5**

9.28 g of maleic acid

7.008 g of NaCl

800 ml of dH<sub>2</sub>O

**10% sodium dodecyl sulfate (SDS)**

10 g SDS

100 ml of H<sub>2</sub>O

**Detection buffer pH 9.5**

3.152 g of 1M tris-HCl

1.168 g of 1M NaCl

200 ml of H<sub>2</sub>O

**Denaturing buffer (0.5M NaOH, 1.5M NaCl)**

20 g NaOH

87.66 g NaCl

dH<sub>2</sub>O was added to the volume of 1 L and sterilised in an autoclave.

**Neutralizing solution pH 7.4 (1.5M NaCl, 0.5M Tris)**

60.56 g Tris-base

87.66 g NaCl

dH<sub>2</sub>O was added to the volume of 1 L and sterilised in an autoclave.

**Citrate buffer (0.01M) pH 4.8**

Solution A (40 ml) and B (60 ml) was mixed and diluted 10 times

Solution A:

2.1 g citric acid monohydrate

100 ml of dH<sub>2</sub>O

Solution B:

2.94 g trisodium citrate dihydrate

100 of dH<sub>2</sub>O

## 4. RESULTS

### 4.1. Phylogenetic relationships in the *Brachypodium* genus

Phylogenetic reconstruction based on single-copy nuclear gene *Gigantea* (GI) was performed for 15 accessions representing 13 *Brachypodium* species (Table 1). To reconstruct the phylogeny of *Brachypodium* species, 496-bp-long multiple sequence alignment was created using GI sequences that were obtained from earlier published data (Díaz-Pérez et al., 2018) with exception for *B. sylvaticum* Bsyl434, for which consensus GI sequence was *de novo* reconstructed, using mapping of raw Illumina reads. The alignment contained 28 sequences, one per diploid species and 2-3 for allopolyploid species, representing all homeologous of GI gene. Due to the abundance of allopolyploid species two trees were reconstructed: the first tree (Figure 9A) included only diploid *Brachypodium* species; the second tree (Figure 9B) included all analysed diploid and polyploid accessions.

Based on the phylogenetic analysis of diploid *Brachypodium* species two main clades were distinguished, i.e. the clade with annual species and the second one with perennials (Figure 9A). In the clade of perennials two main subclades were distinguished: (i) subclade with closely related species *B. arbuscula*, *B. glaucovirens*, and diploid *B. pinnatum*, and (ii) *B. sylvaticum* created separate lineage. The phylogenetic analysis of GI sequences isolated from both diploid and polyploid species allow to distinguish three main clades: (i) clade comprising the sequences isolated only from polyploid species, according to (Gordon et al., 2020) these clade consists of the ancestral “ghost” sequences referred to allopolyploid subgenomes derived from unknown ancestor; Figure 9B); (ii) a clade of sequences isolated primarily from annual species that includes two subclades, one with sequences from *B. distachyon* and *B. hybridum* and the second one with sequences from *B. stacei*, *B. hybridum*, and additionally one sequence from perennial *B. mexicanum* (Figure 9B); (iii) core-perennials clade consisting of sequences isolated from perennial diploids and most perennial allopolyploids (except for *B. mexicanum*).



# Figure 9

## Figure 9

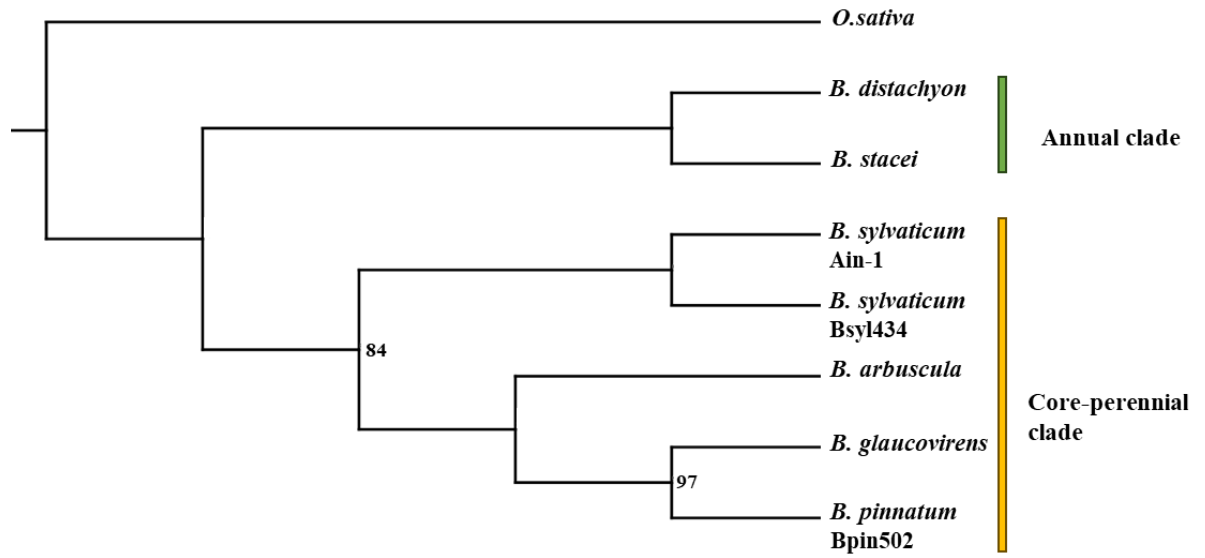
### Phylogenetic relationships of the *Brachypodium* species.

The phylogenetic trees resulting from analyses of the single-copy nuclear *Gigantea* (GI) gene sequences using the ML method, as implemented in IQ-tree2. The tree was rooted with *Oryza sativa*. Bootstrap scores are shown near the nodes. The BS values below 70 were excluded from figure.

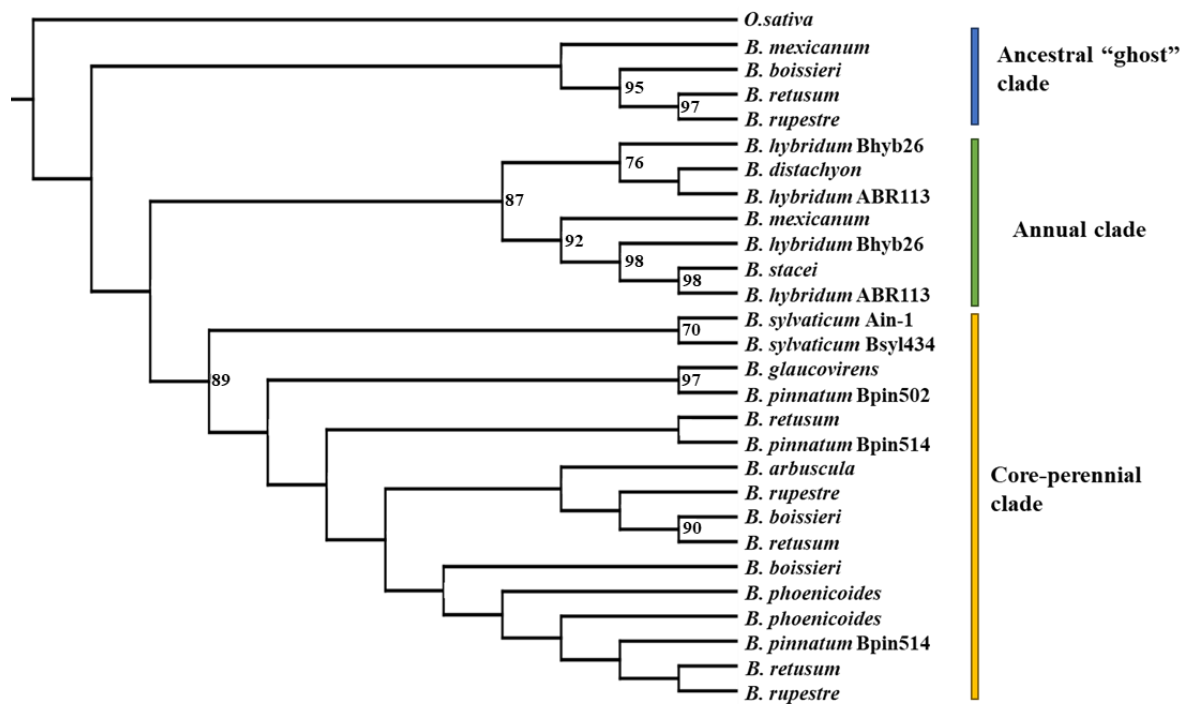
A      The phylogenetic tree of diploid *Brachypodium* species. Main clades, annuals and perennials, are marked with colour bars.

B      The phylogenetic tree of all analysed diploid and polyploid accessions. All homeologous of GI gene of allopolyploid species are included. Main clades are marked with colour bars.

**A**



**B**



#### 4.2. Comparative analyses of repetitive DNA elements in *B. hybridum* and its diploid progenitors.

The repetitive sequence content was annotated in the individual analyses for two diploids (*B. distachyon*, genome composition DD and *B. stacei*, genome composition SS) and in two genotypes of allotetraploid *B. hybridum* (ABR113 and Bhyb26, genome composition DDSS). In *B. distachyon*, approximately 25% of reads were included in large clusters (each above the threshold of 0.01% of the genome), representing highly repetitive sequences (Figure 10A). The other 43% of reads fall into small clusters, which include low-amplified repeats (Figure 10A). Among the reads included in large clusters, about 8.7% were classified as LTR-retrotransposons with Ty3/gypsy being the most abundant (5.8%), followed by Ty1/copia (2.8%), rDNA (1.5%), DNA transposons (1.9%) and satDNA (0.7%; Table 7). The LINE elements reached 0.8% of the genome, while pararetroviruses as little as 0.09% (Table 7). Approximately 11.7% of reads in large clusters could not have been classified as any of the major groups of repeats (Figure 10A).

In *B. stacei* only around 18% of reads were included in large clusters (Figure 10A). The other 38% of reads fall into small clusters (Figure 10A). Among the reads included into large clusters about 7.6% were classified as LTR-retrotransposons with Ty3/gypsy being the most abundant (4.4%), followed by Ty1/copia (3.2%), satDNA (1.5%), rDNA (0.6%) and DNA transposons (0.5%; Table 7). The LINE elements and pararetroviruses constituted 0.06% and 0.03% of the genome, respectively (Table 7). Approximately 7.9% of reads in large clusters could not have been classified as any of the major groups of repeats (Figure 10A).

In *B. hybridum* ABR113 (evolutionary younger allotetraploid lineage), around 23% of reads were included in large clusters (Figure 10A). The other 40% of reads fall into small clusters (Figure 10A). Among the reads included into large clusters about 7.7% were classified as LTR-retrotransposons with Ty3/gypsy being the most abundant (5.2%), followed by Ty1/copia (2.5%), satDNA (2.3%), rDNA (1.3%) and DNA transposons (0.8%; Table 7). The LINE elements and pararetroviruses constituted around 0.1% and 0.07% of the genome, respectively (Table 7). Approximately 10.7% of reads in large clusters could not have been classified as any of the major groups of repeats (Figure 10A).

In *B. hybridum* Bhyb26 (evolutionary older allotetraploid lineage) around 21% of reads were included in large clusters (Figure 10A). The other 38% of reads fall into small clusters (Figure 10A). Among the reads included into large clusters about 7.9% were classified as LTR-retrotransposons with Ty3/gypsy being the most abundant (5%; Table 7), followed by Ty1/copia (2.8%), satDNA (2%), rDNA (0.7%) and DNA transposons (0.8%; Table 7). The LINE elements and pararetroviruses constitutes 0.3% and 0.09% of the genome, respectively (Table 7).

Approximately 9.1% of reads in large clusters could not have been classified as any of the major groups of repeats (Figure 10A). The repeat content analysis revealed notable variation between the two diploids; it was found that *B. distachyon* has a higher repeat content due to its bigger genome size. Both accessions of the allotetraploid *B. hybridum* showed repeatome coverage roughly equal to the additive value of the two diploid progenitor species. The relative composition of the repeatome was also analysed across annual species to complement the absolute quantification of repetitive DNA content. The proportions of distinct repeat types are shown as fractions of the total repeatome (100%), thereby allowing a direct comparison of repeatome structure independently of genome size (Figure 10B). A closer look into the repetitive fraction of annual *Brachypodium* species showed that Ty3/gypsy LTR-retrotransposons were the most abundant repeat type in both diploids as well as in both allopolyploid genotypes, followed by Ty1/copia LTR-retrotransposons, tandem repeats and DNA-transposons. Comparing the proportion of different repetitive DNA fractions in repeatome, Ty1/copia and satDNA proportions were notably higher in *B. stacei*, than in *B. distachyon* (Figure 10B). While proportions of LINEs and DNA transposons were higher in *B. distachyon* (Figure 10B). Among allotetraploid genotypes, in Bhyb26 (evolutionary older allotetraploid lineage) the proportion of retroelement was higher than in ABR113 (evolutionary younger allotetraploid lineage). On the other hand, the proportion of rDNA in Bhyb26 appeared to be much lower than in ABR113 (Figure 10B).

**Table 7.** Genome proportion (percentage) of repeats estimated by RepeatExplorer2 for individual *Brachypodium* samples.

Species	Genotype	Transposable elements																	Tandem repeats		
		Class I																	Class II		
		LTR- retrotransposons																			
		Ty1/ <i>Copia</i>								Ty3/ <i>Gypsy</i>											
		Tar	SIRE	Angela	Ivana	Tork	Bianca	Ikeros	Ale	Athila	Ogre	Tat/Retand	Tekay	CRM	Reina	Pararetrovirus	LINE	DNA- transposons	SatDNA	rDNA	Unclassified
<i>B. distachyon</i>	Bd21-3	0.38	0.50	0.49	0.10	0.36	0.41	0.16	0.42	0.03	0.13	3.71	1.23	0.88	0.12	0.10	0.82	1.51	0.73	0.38	0.50
<i>B. stacei</i>	ABR114	0.25	1.07	1.29	0.08	0.26	0.10	0.14	0.04	1.40	-	1.00	1.30	0.68	0.04	0.04	0.06	0.57	1.54	0.25	1.07
<i>B. hybridum</i>	ABR113	0.33	0.44	1.04	0.07	0.23	0.22	0.14	0.07	0.29	-	2.54	0.92	1.01	0.08	0.08	0.11	0.83	2.36	0.33	0.44
	Bhyb26	0.35	0.80	0.83	0.09	0.24	0.28	0.17	0.01	0.41	-	2.41	1.11	1.16	-	0.10	0.35	0.84	2.01	0.35	0.80
<i>B. arbuscula</i>	Barb502	0.25	2.17	0.26	0.14	0.26	0.27	0.33	0.02	0.16	-	5.70	1.83	0.93	-	0.14	0.22	2.33	4.66	0.25	2.17
<i>B. sylvaticum</i>	Bsyl434	0.51	1.78	0.42	0.06	0.16	0.18	0.36	0.01	0.23	-	5.82	8.97	1.83	-	-	0.13	2.91	3.02	0.51	1.78
<i>B. glaucovirens</i>	PI 4202	0.72	2.28	0.27	0.18	0.17	0.32	0.41	0.05	0.24	0.01	8.65	10.50	1.22	-	0.21	0.22	4.59	3.03	0.72	2.28
<i>B. pinnatum</i>	Bpin502	0.44	2.01	0.57	0.37	0.16	0.21	0.39	0.05	0.29	0.39	8.81	10.52	0.45	-	0.34	0.24	4.71	2.72	0.44	2.01
	Bpin514	0.58	1.77	0.13	0.28	0.16	0.21	0.21	0.05	0.01	-	5.49	1.14	0.60	-	0.02	0.13	1.97	3.99	0.58	1.77
<i>B. phoenicoides</i>	Bpho513	0.61	1.83	0.10	0.05	0.26	0.13	0.16	0.04	0.09	-	4.36	1.90	0.03	-	0.10	0.22	1.82	2.94	0.61	1.83
<i>B. mexicanum</i>	Bmex347	0.89	3.34	2.45	0.18	0.09	0.33	0.54	0.15	0.78	1.86	17.39	19.77	1.42	-	0.06	0.07	5.04	0.65	0.89	3.34
<i>B. rupestre</i>	Brup182	0.52	1.83	0.07	-	0.20	0.16	0.20	0.01	0.16	-	4.31	0.74	0.86	-	0.01	0.26	1.97	3.86	0.52	1.83
<i>B. boissieri</i>	Bbois10	0.44	1.27	0.18	0.04	0.24	0.19	0.15	-	0.01	-	9.76	1.27	0.50	-	0.12	0.19	1.16	5.36	0.44	1.27
<i>B. retusum</i>	Bret551	0.43	1.86	0.12	0.07	0.20	0.17	0.18	-	0.01	-	5.75	1.20	0.75	-	0.14	0.20	1.53	4.23	0.43	1.86

# Figure 10

## Figure 10

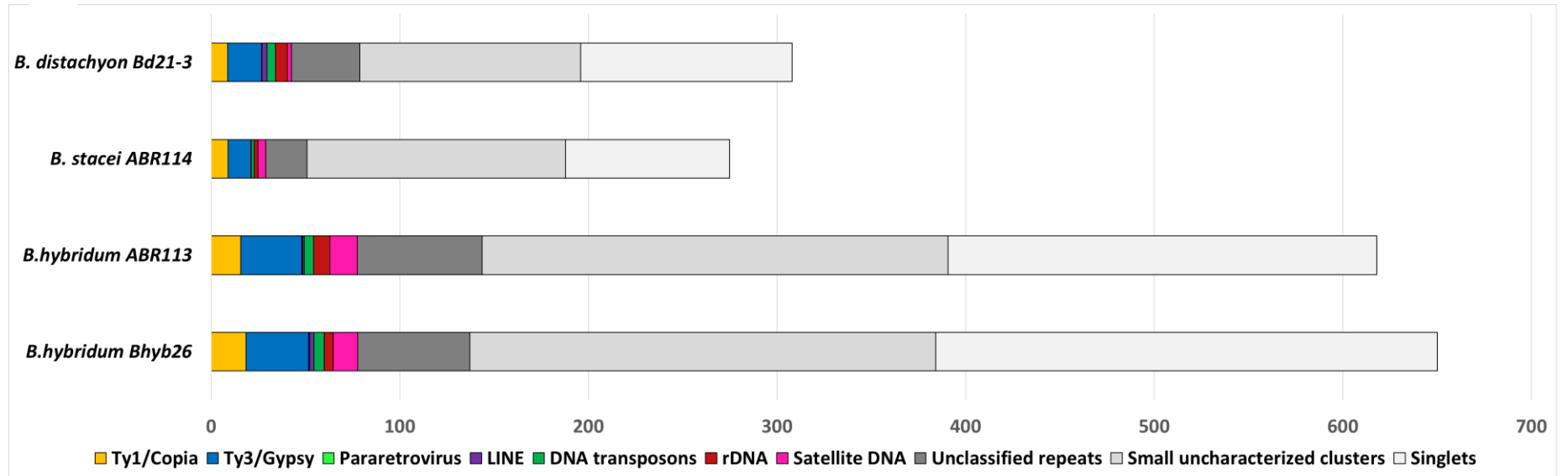
**RepeatExplorer2 estimation of repetitive DNA content in two lineages of *B. hybridum*: evolutionary younger ABR113 ( $n=2x=5+10$ , 616 Mbp/1C DNA, age of origin 0,14 Mya) and evolutionary older Bhyb26 ( $n=2x=5+10$ , 645 Mbp/1C DNA, age of origin 1,4 Mya), and their putative diploid progenitors: *B. distachyon* ( $n=x=5$ , 312Mbp/1C DNA) and *B. stacei* ( $n=x=10$ , 273Mbp/1C DNA). Colour codes for repeat subfamilies are indicated in the chart.**

A      The repetitive DNA content in relation to the genome size. Scale on axis represents genome size in Mbp/1C DNA.

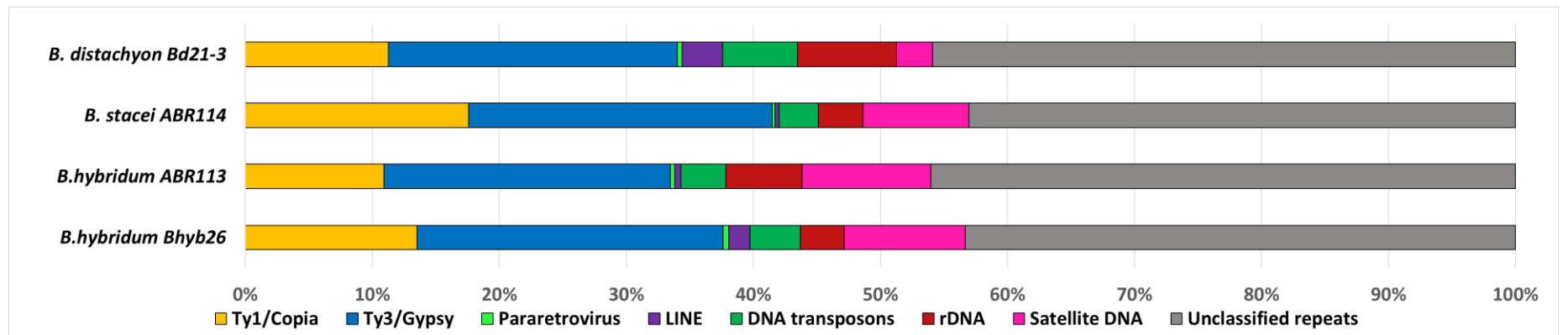
B      Relative abundance of major repeat types in genomes of annual *Brachypodium* species. The repetitive sequences from small clusters (constituted less than 0.01% of genome) and singlets were excluded from figure. The total content of the repetitive sequences from large clusters has been designated as 100%



A



B



#### 4.2.1. Evolutionary pathways of rRNA genes in allotetraploid *Brachypodium hybridum* of different evolutionary age

The part of the research concerning rDNA evolution in *Brachypodium* annual species has already been published (Trunova et al., 2024). The reprint of the original manuscript is provided in the Supplementary material 1 (Trunova et al., 2024).

##### 4.2.1.1. 35S and 5S rDNA loci number and chromosomal localisation

Chromosome numbers were confirmed for all three studied *Brachypodium* species: *B. distachyon* exhibited  $2n = 10$  chromosomes, while *B. stacei*  $2n = 20$  chromosomes, notably smaller in size. Both *B. hybridum* accessions displayed  $2n = 30$  chromosomes, consisting of 10 larger chromosomes inherited from *B. distachyon* and twenty smaller chromosomes originating from *B. stacei*. In *B. distachyon*, a single locus of 5S rDNA was located proximally on the long arm of chromosome Bd4, and a single 35S rDNA locus was found distally on the short arm of Bd5 (Supplementary materials 1. (Trunova et al., 2024); Fig. 1A, E, F). Similarly, *B. stacei* harboured one 5S rDNA locus on the short arm of Bs5 and one 35S rDNA locus proximally on the short arm of Bs6 (Supplementary materials 1. (Trunova et al., 2024); Fig. 1B, E, F). The recently evolved *B. hybridum* genotype ABR113 retained an additive rDNA loci pattern when compared with the putative ancestral genomes (Supplementary materials 1. (Trunova et al., 2024); Fig. 1C, E, F), whereas the evolutionary older Bhyb26 genotype exhibited a reduced number of rDNA loci. The 35S rDNA locus was inherited from the D-subgenome, while the 5S rDNA originated from the S-subgenome (Supplementary materials 1. (Trunova et al., 2024); Fig. 1D, E, F).

##### 4.2.1.2. *In silico* detection and quantification of rDNA repeat elements in the genome

Sequences corresponding to 35S and 5S rDNA were identified using the RepeatExplorer2 pipeline, and their relative abundance in the genome was estimated from raw Illumina sequencing data. Each analysed lineage yielded a single cluster containing reads complementary to the 5S rDNA. For both diploid progenitors and the *B. hybridum* Bhyb26 genotype, a simple circular cluster graph was reconstructed, indicating the presence of one variant of 5S rDNA in each genotype (Supplementary materials 1. (Trunova et al., 2024); Fig. 2A, B, and D). In contrast, the ABR113 genotype produced a more complex graph consisting of two loops connected by a junction region that included the 5S rDNA coding sequence, suggesting the presence of two distinct non-transcribed spacer (NTS) variants (Supplementary

materials 1, (Trunova et al., 2024); Fig. 2C). Mapping of raw reads enabled estimation of 5S rDNA copy number at approximately 1367 units/1C DNA in ABR113 (1C DNA  $\approx$  618 Mb) and 2570 units/1C DNA in Bhyb26 (1C DNA  $\approx$  650 Mb) (Supplementary materials 1. (Trunova et al., 2024); Table 2). Differences were also observed between the *B. hybridum* genotypes in terms of 35S rDNA content, with ABR113 containing around 1356 units/1C DNA and Bhyb26 approximately 747 units/1C DNA (Supplementary materials 1, (Trunova et al., 2024); Table 2).

#### 4.2.1.3. 5S rDNA structure

Genomic DNA from both *Brachypodium* diploid species and *B. hybridum* accessions was subjected to PCR using primers targeting the coding region of 5S rDNA, enabling the amplification of the entire non-transcribed spacer along with 102 bp long fragment of the coding sequence. *B. distachyon* yielded the longest amplicon, approximately 350 bp in size, whereas the product from *B. stacei* was notably shorter, around 250 bp. In *B. hybridum* ABR113, two distinct bands - corresponding to  $\sim$ 350 bp and  $\sim$ 250 bp - were detected, indicating the presence of two divergent 5S rDNA families. In contrast, the older *B. hybridum* Bhyb26 displayed a single band of roughly 220 bp in length (Supplementary materials 1. (Trunova et al., 2024); in Figure 2E).

Southern blot analysis showed differences in NTS length between variants of 5S rDNA in the genomes of *B. distachyon*, *B. stacei* and *B. hybridum* Bhyb26 (evolutionary older allotetraploid lineage). *In silico* analysis of the 5S rDNA sequences revealed a single restriction site for *MseI* in *B. stacei* and in both *B. hybridum* accessions, whereas the sequence from *B. distachyon* contained two *MseI* restriction sites (Supplementary materials 1. (Trunova et al., 2024); in Figure 2G). After Southern blot hybridisation, a single fragment of approximately 240 bp was detected in *B. stacei*. In *B. distachyon*, the 5S rDNA probe hybridised to the full-length monomer ( $\sim$ 370 bp) and to two smaller fragments of about 240 bp and 130 bp, according to restriction profile of *MseI* enzyme (Supplementary materials 1, (Trunova et al., 2024); Figure 2E and F). For the more recently evolved *B. hybridum* accession ABR113, both 240 bp and 370 bp fragments were present, indicating the retention of 5S rDNA variants from both ancestral genomes. The presence of a 370-bp fragment suggested the loss of the second *MseI* site in the D-genome-derived rDNA variant in ABR113. In contrast, the older *B. hybridum* lineage Bhyb26 exhibited a single band of approximately 240 bp, shorter than either parental variant.

The ML phylogenetic analysis was conducted based on cloned sequences of the 5S rDNA non-transcribed spacer (NTS) obtained from two *B. hybridum* accessions and the diploid annual species *B. distachyon* and *B. stacei*. The analysed sequences ranged in length from 221

to 356 bp, and the final alignment used for phylogenetic analyses spanned 417 bp (including gaps) and contained 70 parsimony-informative sites. The resulting phylogenetic tree (Supplementary materials 1, (Trunova et al., 2024); Figure 2H) revealed three distinct clades of 5S rDNA NTS sequences. The first group, supported by a high bootstrap value (BS 97), contained longer sequence variants typical for *B. distachyon* as well as isolated from the younger allotetraploid *B. hybridum* ABR113, indicating their origin from the D-subgenome. The second group (BS82, Supplementary materials 1, (Trunova et al., 2024); Figure 2H) encompassed sequences of approximately 250 bp derived from *B. stacei* and *B. hybridum* ABR113, inherited from S-subgenome. Sequences from the older allotetraploid *B. hybridum* Bhyb26 formed a distinct, well-supported third clade (BS100, Supplementary materials 1, (Trunova et al., 2024); Figure 2H). The *B. distachyon*-like clade split further into two subclades: one composed exclusively of *B. distachyon* sequences, and the other containing *B. hybridum*-derived sequences (Supplementary materials 1, (Trunova et al., 2024); Figure 2H). The *B. stacei*-like clade was subdivided into four subgroups, two of which consisted of sequences from *B. hybridum* ABR113, while the remaining two included sequences from two different *B. stacei* accessions (Supplementary materials 1, (Trunova et al., 2024); Figure 2H).

#### 4.2.1.4. 35S rDNA structure

Both bioinformatic and molecular methods were employed to investigate the structure of the ITS1 and intergenic spacer (IGS) regions of the 35S rDNA homoeologues in the younger and older *B. hybridum* lineages. Due to sequence variation in the ITS1 region inherited from both ancestors, it was possible to differentiate the homoeologues 35S rDNA variants using the gCAPS (genomic Cleaved Amplified Polymorphic Sequences) technique. A unique *Mlu*I restriction site was detected within the ITS1 of the D-subgenome, whereas this site was absent in the ITS1 of the S-subgenome (Supplementary materials 1 in Figure 3B). A fragment covering the 3' end of the 18S rDNA and the full ITS1 region was PCR-amplified and subsequently digested with *Mlu*I. As a result, the *B. distachyon*-derived ITS1 was cleaved into two fragments of approximately 308 and 370 bp, while the *B. stacei*-like variant remained intact (678 bp). In the more recent *B. hybridum* accession ABR113, all three expected bands were observed, indicating the presence of both parental variants. In contrast, only the *B. distachyon*-like fragment was detected in the older *B. hybridum* lineage Bhyb26 (Supplementary materials 1, (Trunova et al., 2024); Figure 3A). The intergenic spacer (IGS) region from the older *B. hybridum* genotype Bhyb26 was cloned and sequenced, then compared with the corresponding sequences from *B. stacei* and *B. distachyon*. The IGS of Bhyb26 was 2519 bp long. Sequence alignment revealed an overall similarity (86%) between Bhyb26 and *B. distachyon* IGS

sequences, although two subregions (51 bp and 29 bp in length) were more similar to *B. stacei*, while three regions (120 bp, 12 bp, and 35 bp) appeared to be unique to Bhyb26 (Supplementary materials 1, (Trunova et al., 2024); Figure S5). Analyses using dot matrix plots and Tandem Repeats Finder indicated the presence of internal repetitive elements within the Bhyb26 IGS that were structurally similar to those found in the *B. distachyon* 35S rDNA IGS (Supplementary materials 1, (Trunova et al., 2024); Figure. 3E and Figure S6). Several TATA box motifs were detected, corresponding to both genic and spacer promoter regions. The homogeneity of the Bhyb26 IGS sequence was further confirmed by mapping Bhyb26 raw Illumina reads to the cloned IGS, generating a consensus sequence, and conducting SNP analysis. Only three single nucleotide polymorphisms were detected in the external transcribed spacer (ETS) region, with a SNP frequency threshold of  $\geq 10\%$  (Supplementary materials 1, (Trunova et al., 2024); Figure S6).

#### **4.2.2. Satellite DNA landscape in allotetraploid *B. hybridum* and its putative diploid progenitors *B. distachyon* and *B. stacei***

In *B. distachyon*, four clusters containing satellite DNA were identified, while in *B. stacei* two clusters only (Figure 11A). Among these satDNA repeats, two satellite families (156-157 bp family and 1800 bp family) characteristic for both diploids were found: in *B. distachyon* (Figure 11B 1 and 4) and in *B. stacei* (Figure 11B 5 and 6). Additionally, two species-specific satDNA families were observed in *B. distachyon* (Figure 11B 2 and 3).

*B. hybridum* ABR113 (evolutionary younger allotetraploid) exhibited three clusters containing satDNA, while for Bhyb26 (evolutionary older allotetraploid), four clusters with satDNA were revealed (Figure 11A). Among identified satDNA clusters, two (containing 156-157 bp family) were common to both tetraploid genotypes. Besides that, one genotype-specific family of minisatellite was observed in ABR113 (Figure 11A and B 9), while in Bhyb26 other genotype-specific family of minisatellite was observed (Figure 11A and B 12). Bhyb26 also harboured one more satellite family (118 bp) that was not found in other analysed accessions (Figure 11A and B 13). Only one satDNA family (156-157 bp) was common for all analysed accessions (Figure 11A and B, positions 1, 5, 7, 8, 10 and 11). Only this family was possible to be detected on chromosomes using FISH.

# Figure 11

## Figure 11

**Overview of satDNA families identified across the analysed *Brachypodium* annual accessions: two diploids (*B. distachyon* and *B. stacei*) and two genotypes of allotetraploid *B. hybridum* (ABR113 and Bhyb26).**

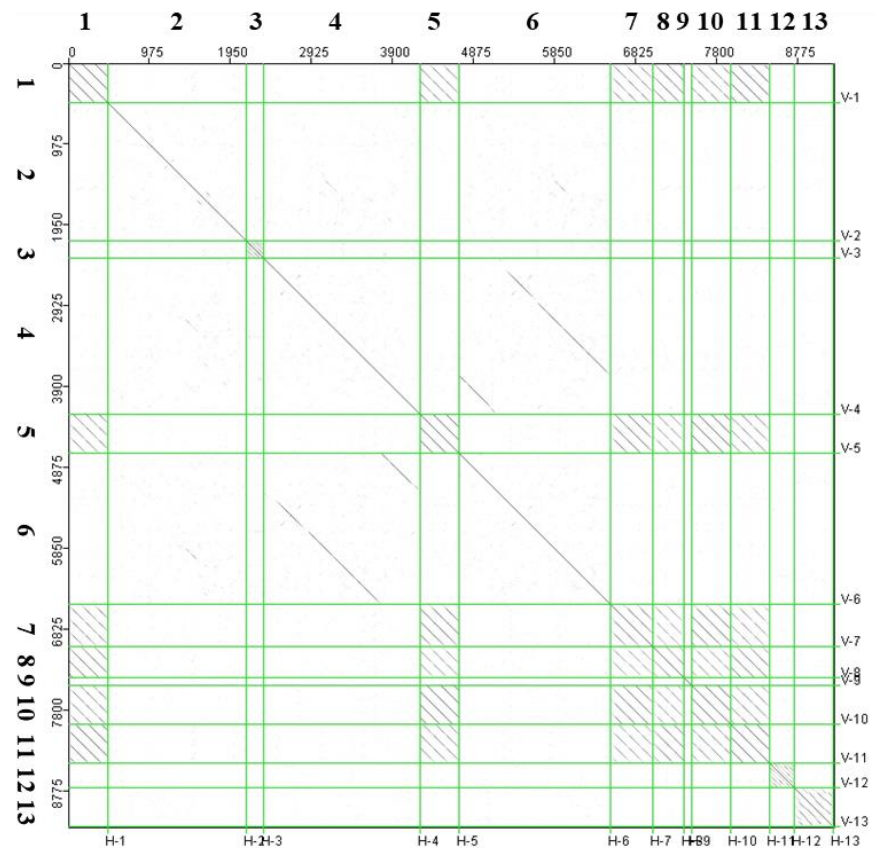
A     The table summarises all satDNA families detected in diploid and allotetraploid accessions, including their cluster ID, monomer length, and abundance.

B     The dot-plot graph visualises the internal structure of representative satDNA monomers. The presence of diagonal lines within the same cluster confirms their tandem arrangement, while cross-clusters diagonal lines indicate sequence homology among different satDNA families. Repeats with high internal similarity appear as continuous and regular patterns, supporting their classification as tandemly organised satellite sequences.

**A**

	Name	Proportion of genome (%)	Consensus size (bp)
<i>B. distachyon</i> (Bd21-3) 2x			
1	BdCl1 (CentBd)	0.680	156
2	BdCl191	0.026	1675
3	BdCl251	0.011	209
4	BdCl264	0.010	1874
<i>B. stacei</i> (AABR114) 2x			
5	BsCl1 (CentBs)	2.1	157
6	BsCl91 (Bsta1818)	0.044	1818
<i>B. hybridum</i> (ABR113) 4x			
7	ABR113Cl1 (CentBs)	2.1	157
8	ABR113Cl22 (CentBd)	0.26	156
9	ABR113Cl67	0.11	69
<i>B. hybridum</i> (Bhyb26) 4x			
10	Bhyb26Cl1 (CentBs)	1.2	157
11	Bhyb26Cl2 (CentBd)	0.76	156
12	Bhyb26Cl104	0.046	52
13	Bhyb26Cl136	0.025	118

**B**





#### 4.2.2.1. Structure and dynamics of 156-157 bp satDNA family in annual *Brachypodium* species

Analysis of the 156-157 bp satDNA family typical for all analysed accessions showed that there are two distinct subfamilies (with similarity 86%; Figure 12E): one typical for *B. distachyon* (Figure 12A) and one typical for *B. stacei* (Figure 12B). In both diploids, only one cluster containing this satDNA family was found. In contrast, both allotetraploid accessions harboured two clusters (Figure 12C and D) containing two subfamilies corresponding to both ancestors.

Southern blot hybridisation with the sequence of 156-157 bp satDNA isolated from *B. distachyon* as a probe to *Ava*II-digested genomic DNA from *B. distachyon* (Bd21 and ABR5), *B. stacei* (ABR114), *B. hybridum* ABR113 and Bhyb26 (Figure 12G) showed a typical for tandem repeats ladder-like pattern of hybridisation signals (Figure 12G). The shortest band representing the monomer size of the sequence (~150 bp) was obtained for all analysed accessions (Figure 12G). After FISH with 156 bp satDNA sequence isolated from *B. distachyon* (CentBd) as a probe showed the hybridisation signals in primary constrictions of all chromosomes in all analysed accessions (Figure 13).

ML analysis was performed using cloned monomer sequences of centromeric satDNA family from both *B. hybridum* lineages and the two diploid *Brachypodium* annuals (Figure 14). The length of the analysed sequences ranged from 155 bp to 159 bp, and the final alignment, used for phylogenetic analyses, was 160 bp long (including gaps) with 80 parsimony informative sites. Among the analysed sequences, two main groups were revealed: (i) the first clade consists of sequences isolated from *B. distachyon* and *B. hybridum* (both ABR113 and Bhyb26); (ii) the second clade comprised sequences isolated from *B. stacei* and *B. hybridum* (both ABR113 and Bhyb26) (Figure 14).

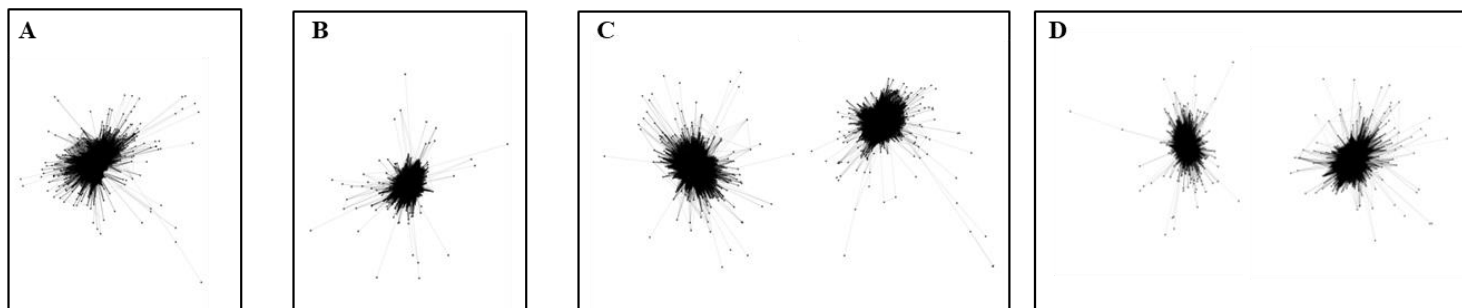
The presence of two 156 and 157 bp subfamilies was also confirmed by the comparative analysis using RepeatExplorer2. In this analysis, raw Illumina reads from both diploid putative ancestors and two allotetraploid *B. hybridum* genotypes were used. Additionally, the *in silico* allotetraploid that was constructed as the additive value of reads representing both putative ancestors was used (Figure 12F). In both *B. hybridum* genotypes, these two satDNA subfamilies were present, however, the comparative analysis of natural *B. hybridum* accessions with artificial *in silico* allotetraploid revealed a potential proliferation of *B. distachyon*-like subfamily in the evolutionary older Bhyb26 (Figure 12F). In younger ABR113 the abundance of both subfamilies was close to the additive value shown by *in silico* allotetraploid (Figure 12F).

# Figure 12

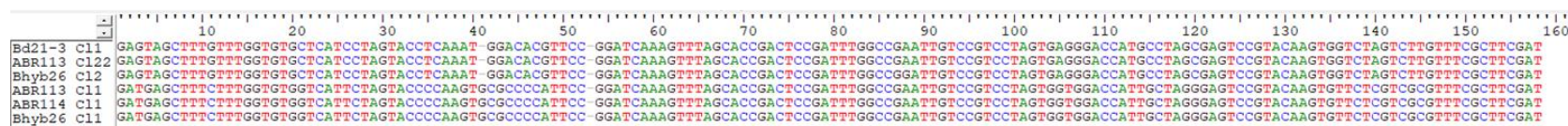
## Figure 12

**Structure of 156-157 bp satellite DNA family in *B. hybridum* and its diploid progenitors *B. distachyon* and *B. stacei*.**

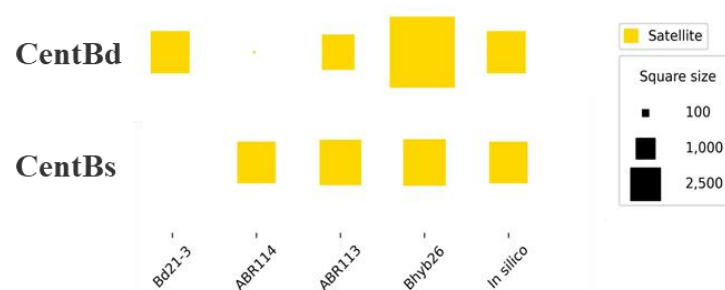
- A-D** Repeat cluster graphs of 156-157 bp satDNA family constructed in RepeatExplorer2 analysis for *B. distachyon* (A); *B. stacei* (B); *B. hybridum* ABR113 (C); *B. hybridum* Bhyb26 (D).
- E** Alignment of consensus sequences of 156 – 157 bp satDNA family, derived from clusters of individual RepeatExplorer2 analysis.
- F** The graph representing the comparative analysis of 156-157 bp satDNA family. The subfamilies are indicated as CentBd – for subfamily typical for *B. distachyon* (156 bp) and CentBs – for subfamily typical for *B. stacei* (157 bp).
- G** Southern blot hybridisation of isolated 156 bp satDNA to genomic DNA from *B. distachyon* (Bd21 – 1, ABR5 – 5), *B. stacei* (ABR114 – 3), *B. hybridum* ABR113 (2) and Bhyb26 (4) genotypes digested with *Ava*II. The restriction map is shown above the photograph of the membrane.



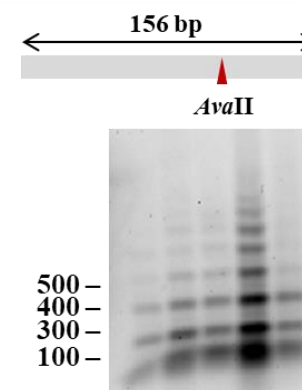
E



F



G

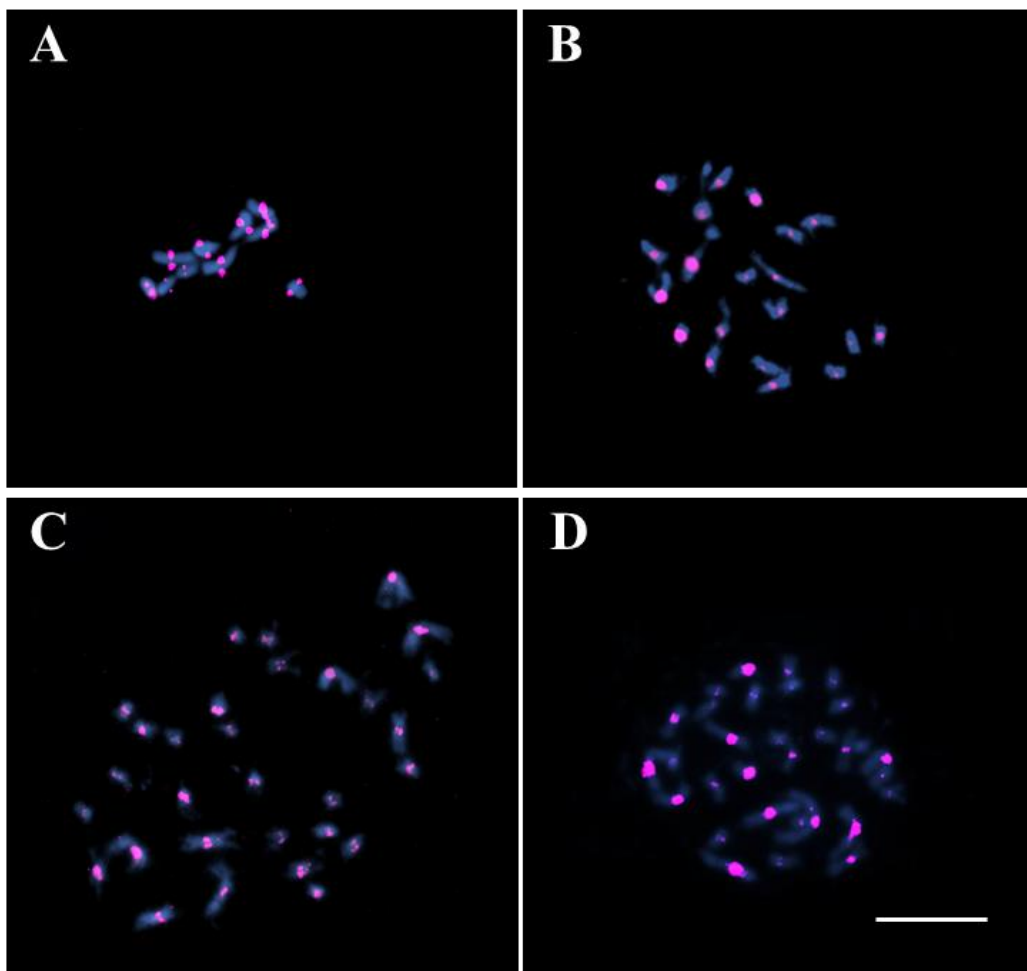


# Figure 13

## Figure 13

**Distribution of the 156-157 bp satDNA (purple signals) on the mitotic metaphase chromosomes stained with DAPI (blue fluorescence) of the studied *Brachypodium* accessions.**

Fluorescence *in situ* hybridisation (A-D) of the CentBd (purple signals) probe to (A) *B. distachyon* Bd21 ( $2n = 10$ ;  $x = 5$ ), (B) *B. stacei* ABR114 ( $2n = 20$ ;  $x = 10$ ), (C) *B. hybridum* ABR113 ( $2n = 30$ ;  $2x = 5 + 10$ ) and (D) *B. hybridum* Bhyb26 ( $2n = 30$ ;  $2x = 5 + 10$ ) chromosomes. The scale bars: 10  $\mu\text{m}$ .



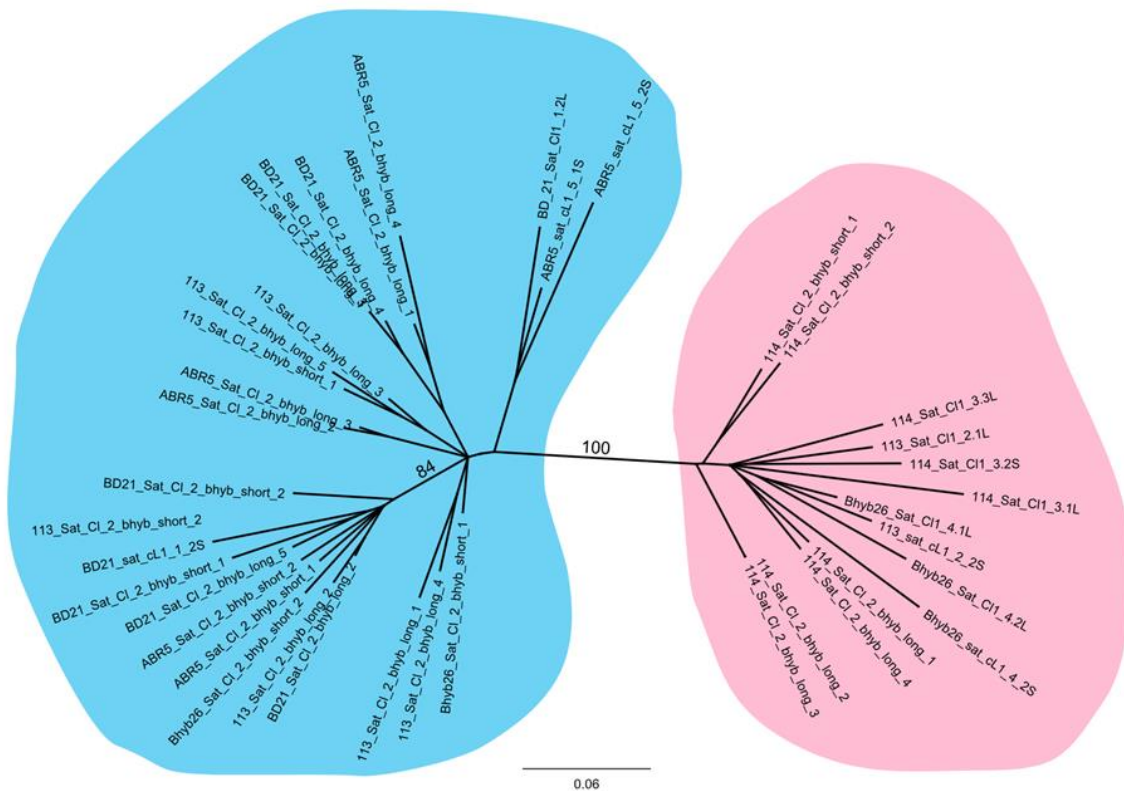
# Figure 14



## Figure 14

**The unrooted phylogenetic tree of the cloned 156–157 bp satDNA family sequences.**

The phylogenetic tree obtained using ML analysis based on sequences isolated from the studied annual *Brachypodium* species were. Bootstrap scores are shown near the nodes. The BS values below 70 were excluded from figure.



***B. distachyon* (ABR5, BD-21)**  
**and *B. hybridum* (ABR113, Bhyb26)**

***B. stacei* (ABR114)**  
**and *B. hybridum* (ABR113, Bhyb26)**

#### 4.2.3. Composition and evolutionary dynamics of LTR retrotransposons in *B. hybridum* and its putative diploid progenitors

In the analysed annual *Brachypodium* accessions, the most highly amplified repeats forming large clusters (exceeding 0.01% of the genome) were predominantly LTR-retrotransposons. A total of 14 major LTR-retrotransposon families were identified. These included six families from the Ty3/*gypsy* superfamily: Athila, Ogre, Tat/Retand, Tekay, CRM, and Reina and eight families from the Ty1/*copia* superfamily: Tar, SIRE, Angela, Ivana, Tork, Bianca, Ikeros, and Ale (Figure 15). Differences in the composition and relative abundance of specific retroelement families were observed between the diploid species *B. distachyon* and *B. stacei* (Figure 15). Within the Ty3/*gypsy* superfamily in *B. distachyon*, the Tat/Retand family was identified as the most abundant – 3.71% of the genome (Table 7), which corresponds to approximately 60% of all identified Ty3/*gypsy* LTR-retrotransposons (Figure 15A). The second major fraction of Ty3/*gypsy* LTR-retrotransposons was represented by Tekay family – 1.23% of the genome (Table 7), which constituted around 20% of all identified Ty3/*gypsy* LTR-retrotransposons in *B. distachyon* (Figure 15A). CRM family comprises 0.88% (Table 7) of the genome, which constituted nearly 15% of all identified Ty3/*gypsy* LTR-retrotransposons (Figure 15A). Ogre and Reina families were rather low abundant in *B. distachyon* (Table 7). The Athila family in *B. distachyon* was the least abundant with 0.03% of the genome (Table 7). In contrast, Athila family was the most abundant among Ty3/*gypsy* in *B. stacei* (1.40% of the genome; Table 7), which constituted approximately 32% of all identified Ty3/*gypsy* LTR-retrotransposons (Figure 15A), followed by Tekay – 1.30% of the genome (Table 7), which constituted 32% of all identified Ty3/*gypsy* LTR-retrotransposons (Figure 15A) and Tat/Retand – 1% of the genome (around 22% of all Ty3/*gypsy*) and CRM – 0.68% of the genome (Table 7), which constituted 15% of all identified Ty3/*gypsy* LTR-retrotransposons (Figure 15A).

In both accessions of *B. hybridum*, the dominant family was Tat/Retand with 2.54% and 2.41% of ABR113 and Bhyb26 genomes, respectively (Table 7), that corresponds to approximately 50% of all identified Ty3/*gypsy* LTR-retrotransposons (Figure 15A), followed by Tekay and CRM with around 1% of genome content (nearly 20% of all Ty3/*gypsy*; Table 7). Interestingly, the Ogre family was not identified in younger *B. hybridum* ABR113, while in the older Bhyb26 two Ty3/*gypsy* families (Ogre and Reina) were not identified (Figure 15A). Reina family was present in their putative diploid progenitors, while Ogre was present only in *B. distachyon* (Figure 15A).

In *B. distachyon* the Ty1/*copia* superfamily showed higher diversity compared to Ty3/*gypsy*. Six different families were identified (Tar, SIRE, Angela, Tork, Bianca and Ale), which constituted from 0.35% to 0.5% of the genome (Table 7). The contribution of each of

these families to the entire Ty1/*copia* fraction varied from 12% to 18% (Figure 15B). The most abundant was SIRE with 0.5% of the genome (Table 7), and the least abundant was Ivana with 0.1% of the genome (Table 7). In contrast, in *B. stacei* different landscape of Ty1/*copia* elements was observed with two dominating families: (i) Angela approximately 40% of all Ty1/*copia* (Figure 15B), that corresponds to 1.29% of the genome (Table 7); (ii) SIRE nearly 32% of all Ty1/*copia* (Figure 15B), which correspond to 1.07% of the genome (Table 7). The least abundant was the Ale family with only 0.04% of the genome (Table 7).

In evolutionary younger *B. hybridum* ABR113 the most abundant family was Angela with 1.04% of the genome (Table 7), which corresponded to 40% of all Ty1/*copia* (Figure 15B). In older Bhyb26 Angela comprised 0.83% of the genome (Table 7) and constituted 30% of all Ty1/*copia*. Among other Ty1/*copia* families SIRE was highly abundant in *B. hybridum*: in ABR113 SIRE comprised approximately 0.44% of the genome (Table 7), which constituted around 18% of all Ty1/*copia* (Figure 15B), while in the older Bhyb26 SIRE comprised nearly 0.8% of the genome (Table 7) and up to 30% of all Ty1/*copia* (Figure 15B). Other Ty1/*copia* families (Tar, Ivana, Tork, Bianca and Ikeros) were rather low abundant in both genotypes of *B. hybridum* (Figure 15B). The least abundant Ty1/*copia* family was Ale, that constituted 0.07% of the genome in ABR113 and only 0.01% of the genome in Bhyb26 (Table 7).

The most abundant LTR-retrotransposons families (Tat/Retand, CRM, Tekay and Athila) were selected for further *in situ* hybridisation analyses. The FISH with the sequence of the conserved domain of reverse transcriptase (*rt*) of CRM retrotransposon as a probe revealed its localisation in centromeric/pericentromeric regions (Figure 16) in mitotic metaphase chromosomes. In *B. distachyon*, the FISH signals were present in all chromosomes (Figure 16A), while in *B. stacei*, only a few chromosomes exhibited strong signals (Figure 16B). In both accessions of *B. hybridum*, the chromosomal distribution of CRM was similar to that of the putative ancestral species (Figure 16C and D).

After FISH with the *rt* sequence of Tekay as a probe, dispersed signals were observed on all chromosome arms in *B. distachyon* (Figure 17A), while in *B. stacei*, only a few chromosomes exhibited signals, mostly in the pericentromeric region (Figure 17B). In both accessions of *B. hybridum*, the chromosomal distribution of CRM resembled the pattern exhibited by both putative ancestors (Figure 17C and D). FISH with *rt* sequence of Tat/Retand was also performed. This family, which represented the most amplified LTR-retrotransposon family in *B. distachyon* and *B. hybridum*, showed a dispersed distribution across all chromosome arms in *B. distachyon* (Figure 18A). Strikingly, no hybridisation signals were detected on the chromosomes of *B. stacei* (Figure 18B), suggesting a low abundance of this family in the *B. stacei* genome. However, FISH signals were clearly observed on *B. stacei*-

derived chromosomes in both *B. hybridum* accessions (ABR113 and Bhyb26) (Figure 18C and D).

A comparative analysis of two *B. hybridum* genotypes, differing in their evolutionary age, alongside an *in silico* allotetraploid (constructed by combining reads from the diploid progenitors), revealed that even within a single LTR-retrotransposon family, individual subfamilies (present in separate clusters in RepeatExplorer2) can follow distinct evolutionary trajectories. For instance, within Ty3/gypsy Tat/Retand family two main trends were observed: (i) the abundance in both ABR113 and Bhyb26 was close to *in silico* allotetraploid (Figure 19 A2), (ii) abundance was lower in younger ABR113, comparing to *in silico* allotetraploid, but higher in older Bhyb26 (Figure 19 A1). On the other hand, CRM family exhibited two different trends: (i) higher abundance in ABR113 than in *in silico* allotetraploid, but in Bhyb26 the abundance was significantly lower (Figure 19 B1); (ii) the abundance of this element is higher in the ABR113 when compared to the *in silico* allotetraploid, and even more abundant in the evolutionarily older Bhyb26 genotype (Figure 19 B2). Tekay elements showed two other trajectories: (i) the abundance was slightly elevated in the ABR113 when compared to the *in silico* allotetraploid, while a reduction was observed in the evolutionarily older Bhyb26 when compared with ABR113 (Figure 19 C1); (ii) a slight, progressive increase in abundance was observed in both ABR113 and Bhyb26 genotypes (Figure 19 C2). In all families, there were subfamilies (clusters) that did not show significant changes in their abundance (e.g., Figure 19 B1, C1, C2, yellow lines).

#### **4.2.4. Comparative analysis of repeatome composition in allotetraploid *Brachypodium hybridum* of different evolutionary age**

To explore the dynamics of repetitive DNA in allotetraploid *B. hybridum*, the comparative RepeatExplorer2 analysis was performed across diploid progenitors (*B. distachyon* and *B. stacei*) and their derived allotetraploids (*B. hybridum* genotypes ABR113 and Bhyb26) (Figure 20). The results revealed substantial differences in the abundance and composition of repetitive elements between the two diploid species. *B. distachyon* and *B. stacei*. They showed distinct repeatome profiles, with variation in both the proportion of specific repeat families and their relative contributions to the genome (Figure 20). Due to differences in genome size and repetitive DNA profiles between the diploid progenitors, direct comparison of their repeatomes to that of the derived allotetraploid *B. hybridum* was challenging. To overcome this, an *in silico* allotetraploid was generated, providing an additive baseline for comparison with the repeatome in natural *B. hybridum* (Figure 20). The *in silico* allotetraploid here is a sum of reads derived

from the putative diploid progenitors. Repetitive DNA clusters representing at least 0.01% each of the analysed reads were included in the comparative analysis (Figure 20).

Overall, the repeatome profiles of both *B. hybridum* accessions closely resembled the additive value represented by the *in silico* allotetraploid (Figure 20). Further analysis showed slight differences in the amplification of specific repeat families between two *B. hybridum* accessions when compared to *in silico* tetraploid (Figure 20). For instance, the prominent reduction in copy number of 35S rDNA (CI23, CI24, CI15, CI42; red arrows, Figure 20) was observed in older Bhyb26, while in younger ABR113 the value was close to the *in silico* one (Figure 20). Except the cluster CI57 (red star, Figure 20), which contains *B. distachyon*-like 35S rDNA IGS, which showed a decrease in copy number in younger ABR113. On the other hand, analysis of the 5S rDNA showed that in both accessions of natural allotetraploid, the amplification level was lower than in the *in silico* allotetraploid (CI90, red diamond, Figure 20). In contrast, different tandem repeat family (centromeric satellite) showed striking proliferation of *B. distachyon*-like subfamily (CI14) in older Bhyb26, compared to younger ABR113 and *in silico* allotetraploid (yellow arrow, Figure 20), while *B. stacei*-like subfamily showed in both accessions of *B. hybridum* a value close to that presented by an *in silico* allotetraploid (CI1, yellow diamond, Figure 20). Interestingly, the burst of one LINE lineage was observed in older Bhyb26 (CI65, fuchsia arrow, Figure 20). Among other TE families different patterns were observed, e.g., Angela family of Ty1/*copia* LTR-retrotransposons showed two different patterns of amplification levels: (i) the abundance was higher in older allotetraploid Bhyb26, comparing to both *in silico* allotetraploid and younger ABR113 (CI21, CI51, CI55, orange arrows, Figure 20); (ii) abundance was higher in younger ABR113, comparing to both *in silico* allotetraploid and older Bhyb26 (CI20, CI22, CI90, orange diamonds, Figure 20).

Summarising, several main trends were observed: (i) the abundance of some LTR retroelements in both *B. hybridum* genotypes (ABR113 and Bhyb26) remains similar to the *in silico* allotetraploid (Figure 20); (ii) in others, abundance increases progressively from the younger ABR113 to the older Bhyb26 (Figure 20); and (iii) in some cases abundance was higher in ABR113 compared to the *in silico* allotetraploid, but decreased in the older Bhyb26 (Figure 20).

# Figure 15

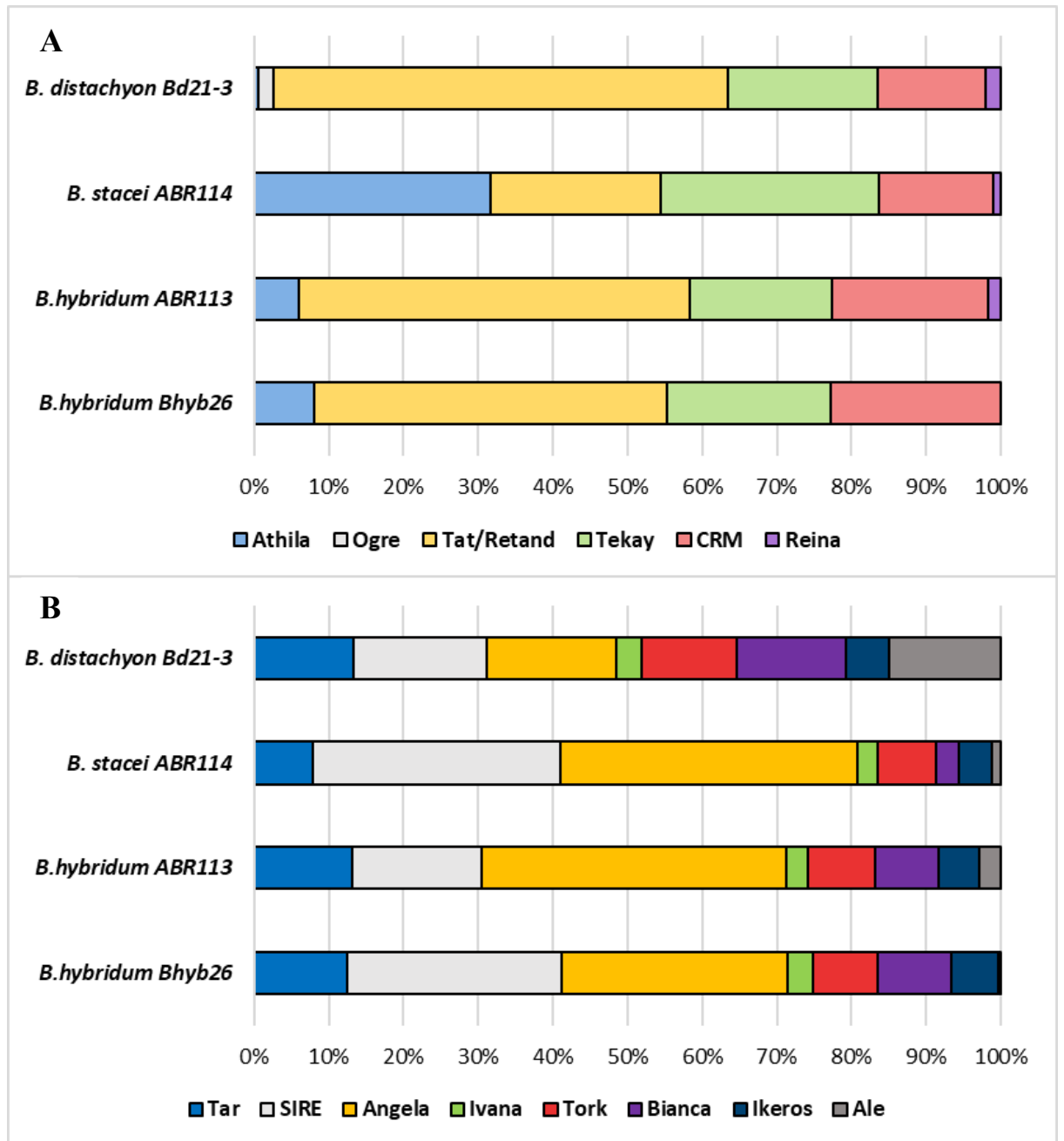
## Figure 15

**The composition of LTR-retrotransposons of two genotypes of *B. hybridum* (evolutionary younger ABR113 and evolutionary older Bhyb26) and their putative diploid progenitors, *B. distachyon* and *B. stacei*.**

The total content of the LTR-retrotransposons superfamily from large clusters has been designated as 100%.

- A      Composition of the Ty3/*gypsy* LTR retrotransposon.
- B      Composition of the Ty1/*copia* LTR retrotransposon.





# Figure 16

## Figure 16

**Distribution of the CRM LTR-retrotransposon on the mitotic metaphase chromosomes of the studied *Brachypodium* accessions.**

Fluorescence *in situ* hybridisation (A-D) with the CRM\_BsCl4 probe (green fluorescence) isolated from *B. stacei*.

A1 – A4      *B. distachyon* Bd21 ( $2n = 10$ ;  $x = 5$ )

B1 – B4      *B. stacei* ABR114 ( $2n = 20$ ;  $x = 10$ )

C1 – C4      *B. hybridum* ABR113 ( $2n = 30$ ;  $2x = 5 + 10$ )

D1 – D4      *B. hybridum* Bhyb26 ( $2n = 30$ ;  $2x = 5 + 10$ )

A1 – D1      Mitotic metaphase chromosomes (blue fluorescence)

A2 – D2      FISH signals corresponding to the CRM\_BsCl4 probe (green fluorescence)

A3 – D3      FISH signals corresponding to the 5S rDNA probe (red fluorescence)

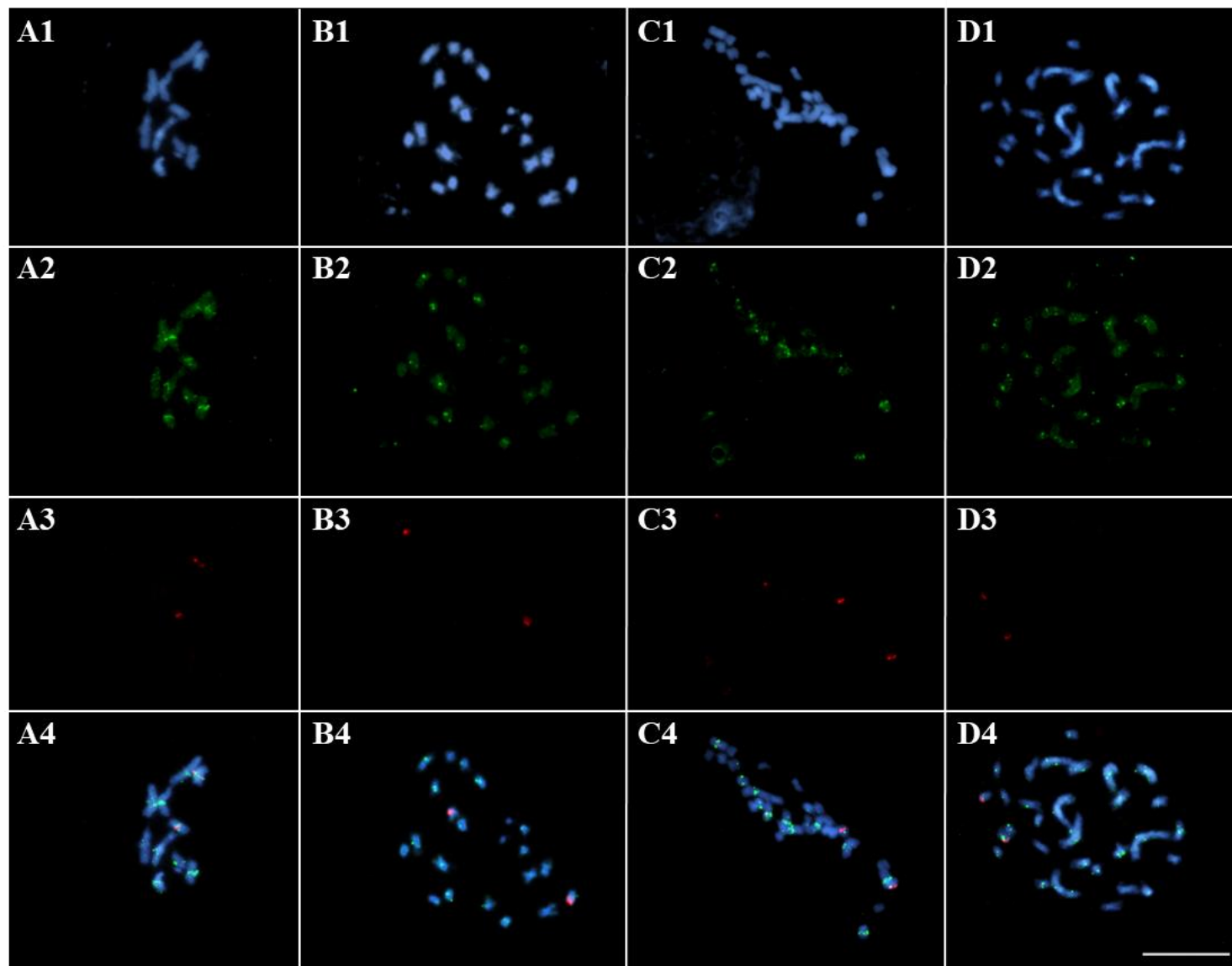
A4            Superimposed channels A1 – A3

B4            Superimposed channels B1 – B3

C4            Superimposed channels C1 – C3

D4            Superimposed channels C1 – C3

Scale bar: 10  $\mu\text{m}$ .



# Figure 17

## Figure 17

**Distribution of the Tekay LTR-retrotransposon on the mitotic metaphase chromosomes of the studied *Brachypodium* accessions.**

Fluorescence *in situ* hybridisation (A-D) with the Tekay\_BdCI5 probe (**green fluorescence**) isolated from *B. distachyon*.

A1 – A4      *B. distachyon* Bd21 ( $2n = 10$ ;  $x = 5$ )

B1 – B4      *B. stacei* ABR114 ( $2n = 20$ ;  $x = 10$ )

C1 – C4      *B. hybridum* ABR113 ( $2n = 30$ ;  $2x = 5 + 10$ )

D1 – D4      *B. hybridum* Bhyb26 ( $2n = 30$ ;  $2x = 5 + 10$ )

A1 – D1      mitotic metaphase chromosomes (**blue fluorescence**)

A2 – D2      FISH signals corresponding to the Tekay\_BdCI5 probe (**green fluorescence**)

A3 – D3      FISH signals corresponding to the 5S rDNA probe (**red fluorescence**)

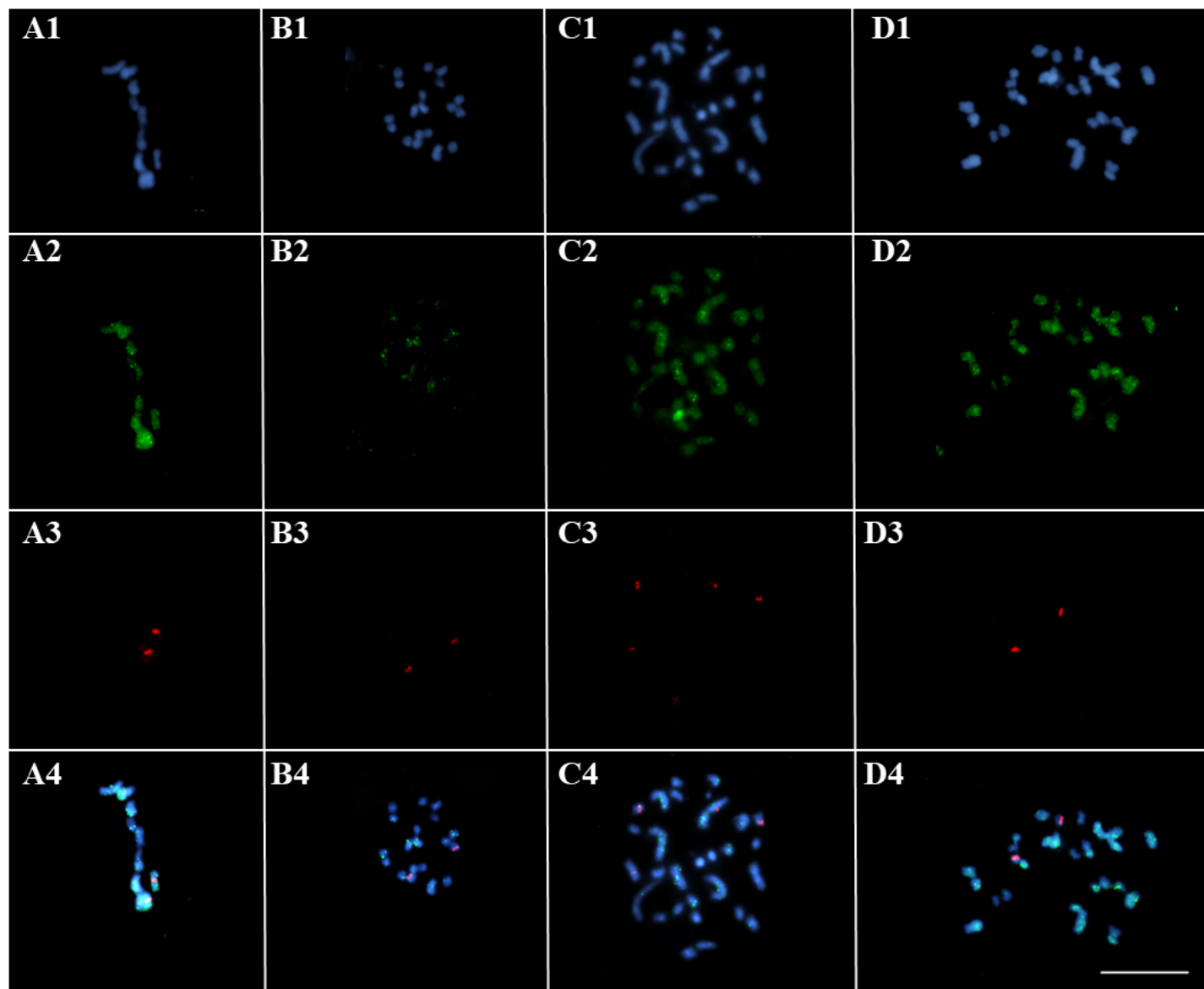
A4              Superimposed channels A1 – A3

B4              Superimposed channels B1 – B3

C4              Superimposed channels C1 – C3

D4              Superimposed channels C1 – C3

Scale bar: 10  $\mu\text{m}$ .



# Figure 18



## Figure 18

**Distribution of Tat/Retand LTR-retrotransposon on the mitotic metaphase chromosomes of the studied *Brachypodium* accessions.**

Fluorescence *in situ* hybridisation (A-D) with the Tat/Retand\_BdCl24 probe (**green fluorescence**) isolated from *B. distachyon*.

A1 – A4      *B. distachyon* Bd21 ( $2n = 10$ ;  $x = 5$ )

B1 – B4      *B. stacei* ABR114 ( $2n = 20$ ;  $x = 10$ )

C1 – C4      *B. hybridum* ABR113 ( $2n = 30$ ;  $2x = 5 + 10$ )

D1 – D4      *B. hybridum* Bhyb26 ( $2n = 30$ ;  $2x = 5 + 10$ )

A1 – D1      Mitotic metaphase chromosomes (**blue fluorescence**)

A2 – D2      FISH signals corresponding to the Tat/Retand\_BdCl24 probe (**green fluorescence**)

A3 – D3      FISH signals corresponding to the 5S rDNA probe (**red fluorescence**)

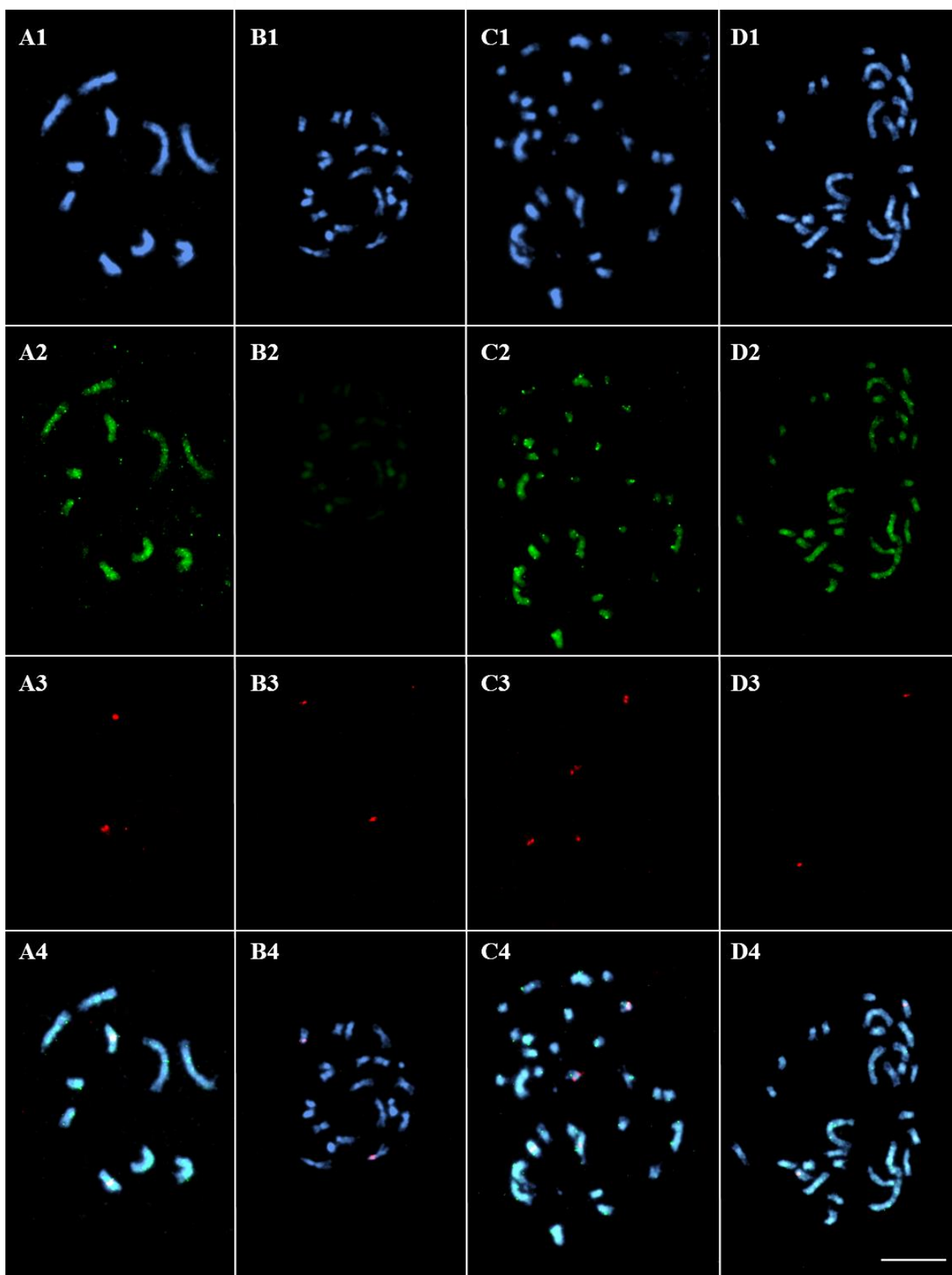
A4      Superimposed channels A1 – A3

B4      Superimposed channels B1 – B3

C4      Superimposed channels C1 – C3

D4      Superimposed channels C1 – C3

Scale bar: 10  $\mu$ m.



# Figure 19

## Figure 19

**Comparative analysis of the abundance of selected Ty3/gypsy LTR-retrotransposon families in two *B. hybridum* genotypes with different evolutionary age.**

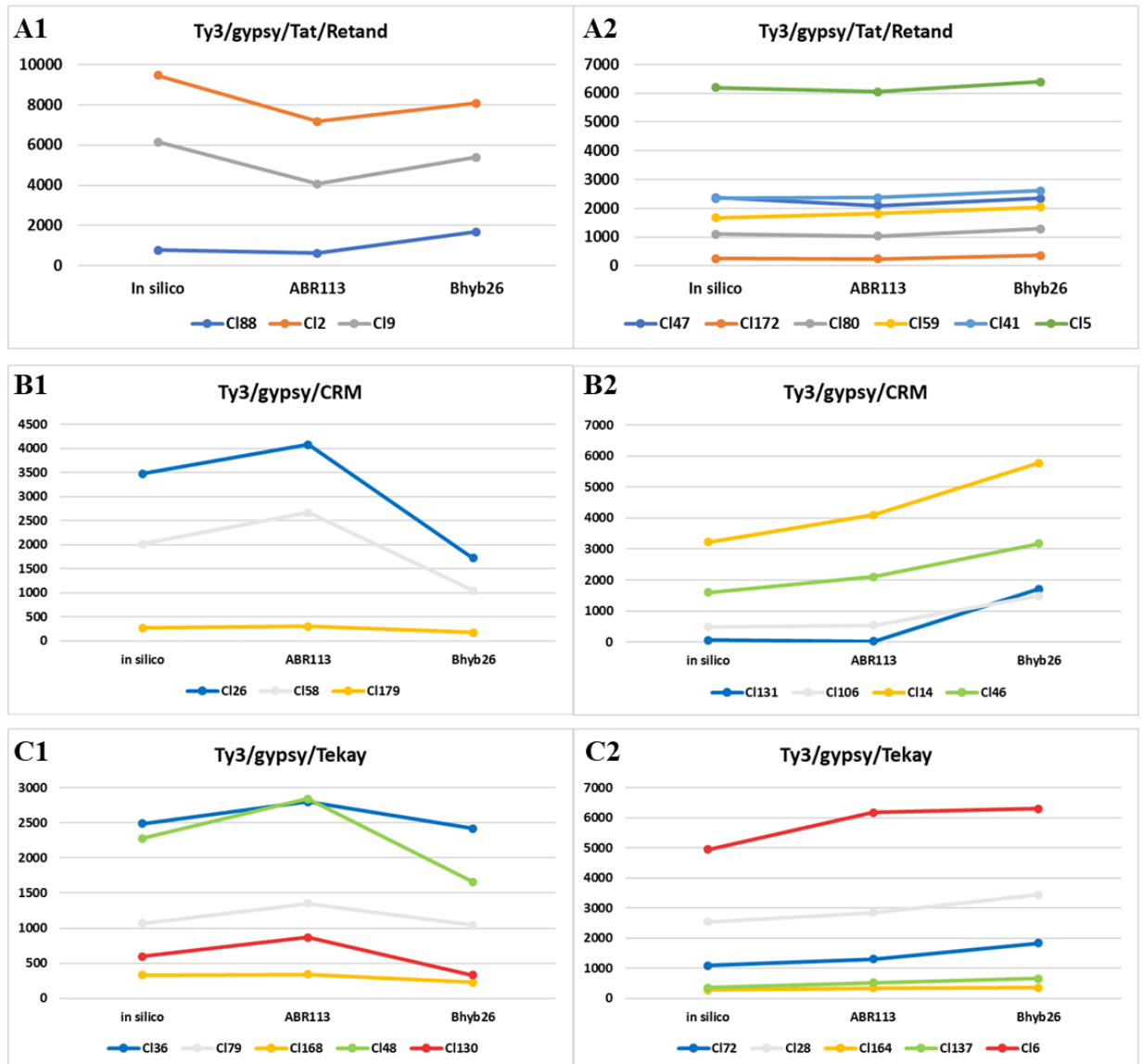
The younger ABR113 and older Bhyb26 are compared to their putative diploid progenitors, represented here by an *in silico* allotetraploid.

The Y-axis reflects the number of reads that belong to the cluster. Two genotypes of *B. hybridum* and *in silico* allotetraploid are presented on the X-axis. Graphs marked with 1 and 2 show different trends within one LTR-retrotransposon family.

A1 – A2      Tat/Retand family

B1 – B2      CRM family

C1 – C2      Tekay family

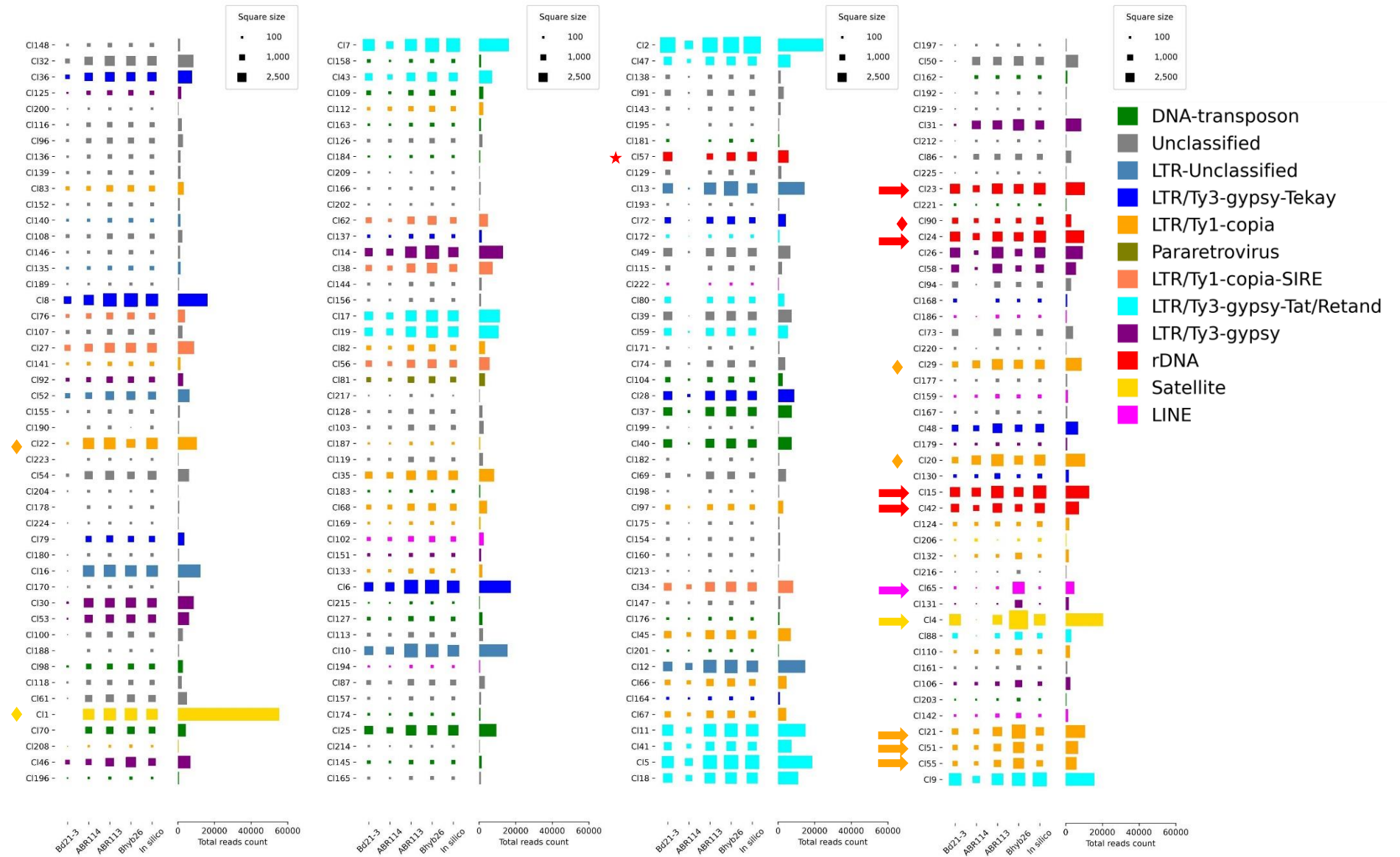


# Figure 20

## Figure 20

### **Comparative analysis of repeat composition among annual *Brachypodium* species.**

The chart illustrates the distribution of repeat clusters (comprising at least 0.01% of analysed reads) in analysed genotypes. The total read count per cluster is indicated by bar length, while square size reflects the genomic abundance in the studied accessions. Bars and squares are coloured according to the type of repetitive element. The *in silico* allotetraploid is a sum of reads derived from putative diploid progenitors.





#### 4.3. Characterisation of the repetitive DNA elements in selected diploid and allopolyploid species of *Brachypodium* genus

The repetitive DNA content across the *Brachypodium* genus was annotated in the individual analyses across a representative set of annual and perennial accessions, including diploids and polyploids (Figure 21). While annual diploid species *B. distachyon* and *B. stacei* displayed the lowest repetitive DNA content, perennial diploid species (*B. sylvaticum*, *B. arbuscula*, *B. glaucovirens*, and diploid *B. pinnatum*) exhibited higher repeat content (Figure 21). While their genomes remained compact, there was a noticeable elevation in Ty3/gypsy elements and unclassified repeats compared to annual diploids. Annual allotetraploid accessions (*B. hybridum* ABR113 and Bhyb26) displayed intermediate levels of repetitive DNA content (Figure 21). In contrast, perennial polyploids exhibited higher abundance of the repetitive sequences. This was particularly striking in *B. mexicanum*, where a significant genome size increase was primarily driven by massive amplification of Ty3/gypsy elements and a large fraction of unclassified repeats (Figure 21).

In diploid *B. sylvaticum*, around 33% of reads were included in large clusters (Figure 21). Another 45% of reads fall into small clusters (Figure 21). Among the reads included in large clusters, Ty3/gypsy elements were the most abundant (nearly 16.5% of the genome, Figure 21). Ty1/copia elements were present at a much lower but significant level (3.4% of the genome, Figure 21), followed by satellite DNA (3.02%) and DNA transposons (2.91%). Other classes of repetitive DNA, including rDNA (0.51%) and LINEs (0.13%), were represented in small proportions (Table 7, Figure 21).

The overall proportion of repetitive DNA in diploid *B. arbuscula* genome was lower than in *B. sylvaticum*. Around 25% of reads were included in large clusters (Figure 21). Approximately 50% of reads fall into small clusters (Figure 21). However, the repeat composition was similar to that of *B. sylvaticum*, with Ty3/gypsy (8.6% of the genome) and Ty1/copia (3.7% of the genome) constituting the majority of repetitive sequences (Figure 21). In contrast, satDNA abundance was higher, representing 4.66% of the genome in *B. arbuscula*. DNA-transposons constituted 2.33% of the genome (Table 7, Figure 21). rDNA and LINEs were present in lower amounts, 0.25% and 0.22% of the genome, respectively.

In diploid *B. glaucovirens*, around 40% of reads were included in large clusters (Figure 21). Approximately 45% of reads fall into small clusters (Figure 21). Ty3/gypsy and Ty1/copia repeats were the most abundant with 19.8% and 4.4% of the genome, respectively (Table 7, Figure 21). The DNA transposons were also highly abundant in *B. glaucovirens*, constituting 4.59% of the genome, while the satellite DNA accounted for 3.03% of the genome (Table 7,

Figure 21). Other classes of repetitive DNA, including rDNA (0.72%) and LINEs (0.21%), were represented in small proportions (Table 7, Figure 21).

In diploid *B. pinnatum*, nearly 40% of reads were included in large clusters (Figure 21), while the other 45% of reads fall into small clusters (Figure 21). The diploid *B. pinnatum* exhibited a repeat composition similar to *B. glaucovirens*, Ty3/*gypsy* and Ty1/*copia* elements were the most abundant accounting for 20% and 4.4% of the genome, respectively. The DNA transposons abundance was the highest among diploids and made up to 4.71% of the genome, while the satellite DNA accounted for 2.72% of the genome (Table 7, Figure 21). Other classes of repetitive DNA, including rDNA (0.44%) and LINEs (0.24%), were represented in small proportions (Table 7, Figure 21).

In the allotetraploid *B. pinnatum*, approximately 25% of reads were included in large clusters (Figure 21), while the other 45% of reads fell into small clusters (Figure 21). Ty3/*gypsy* and Ty1/*copia* elements were the most abundant comprising 6.6% and 3.3% of the genome, respectively. The satellite DNA represented nearly 4% of the genome (Table 7, Figure 21). The DNA transposons constituted 1.97% of the genome, while other classes of repetitive DNA, including rDNA (0.58%) and LINEs (0.13%), were present in small proportions (Table 7, Figure 21).

*B. phoenicoides*, another allotetraploid, exhibited a similar repeat profile to most perennial *Brachypodium* allopolyploids (Figure 21). Nearly 25% of reads were included in large clusters (Figure 21), while approximately 50% of reads fell into small clusters (Figure 21). Ty3/*gypsy* and Ty1/*copia* elements comprised the majority of repetitive sequences (6.3% and 3.1% of the genome, respectively), the satellite DNA constituted 2.94% of the genome (Table 7, Figure 21). The DNA transposons accounted for 1.82% of the genome (Table 7, Figure 21). Other classes of repetitive DNA, including rDNA (0.61%) and LINEs (0.22%), were present in small proportions (Table 7, Figure 21).

The allotetraploid *B. mexicanum* had the largest genome among all analysed species and the highest repetitive DNA content (Table 7, Figure 21). Approximately 60% of the reads were included in large clusters (Figure 21). Another 20% fell into small clusters (Figure 21). Ty3/*gypsy* elements were most abundant, accounting for 41% of the genome (Table 7, Figure 21). Ty1/*copia* elements also made a substantial contribution, accounting for nearly 8% of the genome (Table 7, Figure 21). A remarkable quantity of DNA transposons, accounting for nearly 5% of the genome, was observed (Table 7, Figure 21), alongside lower levels of rDNA (0.89%), satDNA (0.65%) and LINEs (0.07%) (Table 7, Figure 21).

Allohexaploid *B. rupestre* exhibited a moderate genome size. Nearly 27% of reads were included in large clusters (Figure 21), while the other 50% of reads fall into small clusters

(Figure 21). Ty3/gypsy and Ty1/copia elements were the most abundant repetitive sequences (6% and 3% of the genome, respectively), the satellite DNA made up 3.86% of the genome (Table 7, Figure 21). The DNA-transposons accounted for 1.97% of the genome (Table 7, Figure 21). Other classes of repetitive DNA, including rDNA (0.52%) and LINEs (0.26%), were presented in smaller proportions (Table 7, Figure 21).

In allohexaploid *B. boissieri*, the repeatome followed the typical pattern of *Brachypodium* perennial polyploids. Approximately 35% of the reads were included in large clusters (Figure 21). Another 30% fell into small clusters (Figure 21). Ty3/gypsy elements were the most abundant (11.5% of the genome), followed by satellite DNA (5.36% of the genome; Table 7, Figure 21). The Ty1/copia elements were also highly abundant and made up 2.5% of the genome (Table 7, Figure 21). The DNA-transposons constituted 1.16% of the genome (Table 7, Figure 21). Other classes of repetitive DNA, including rDNA (0.44%) and LINEs (0.19%), were in smaller proportions (Table 7, Figure 21).

In allohexaploid *B. retusum*, nearly 28% of the reads were included in large clusters (Figure 21). Approximately 40% of the reads fell into small clusters (Figure 21). Ty3/gypsy elements were the most abundant (7.7% of the genome), followed by satellite DNA (4.23% of the genome; Table 7, Figure 21). The Ty1/copia elements represented 3% of the genome (Table 7, Figure 21). The DNA transposons were less abundant with 1.53% of the genome (Table 7, Figure 21). Other classes of repetitive DNA, including rDNA (0.43%) and LINEs (0.2%), were present in smaller proportions (Table 7, Figure 21).

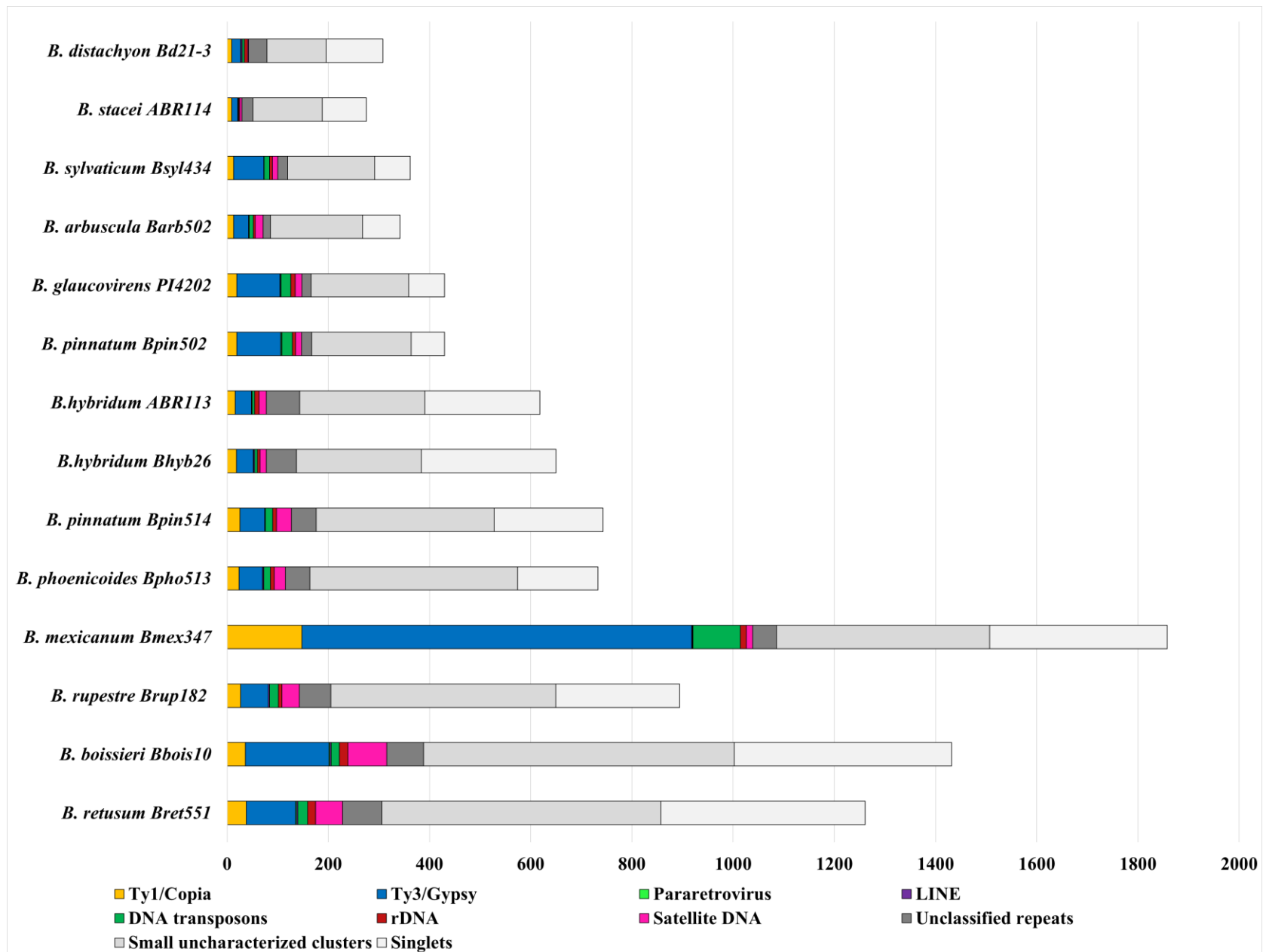
The closer look into composition of repetitive DNA within large clusters (Figure 22) showed that across all examined species, Ty3/gypsy LTR-retrotransposons consistently represent the most abundant class of repetitive DNA (Figure 22). This pattern is particularly pronounced in perennial diploid species such as *B. pinnatum* and *B. glaucovirens* and polyploid species *B. mexicanum* (4x), *B. boissieri*, and *B. retusum*, where Ty3/gypsy occupy a substantial fraction of the repeatome (Figure 22). Ty1/copia elements are the second most abundant class in most species, with the most striking proliferation in *B. mexicanum* (Figure 22). Two of the most variable classes of repetitive DNA were satellites and DNA-transposons (Figure 22). Diploids: *B. sylvaticum*, *B. arbuscula*, *B. glaucovirens*, and *B. pinnatum* showed higher overall repeat content compared to annual diploids (Figure 21 and 22).

# Figure 21

## Figure 21

**Estimation of repetitive DNA content retrieved from the individual RepeatExplorer2 analyses in all analysed *Brachypodium* accessions.**

Colour codes for repeat subfamilies are indicated in the chart. The repetitive DNA content in relation to the genome size. Scale on axis represents genome size in Mbp/1C DNA.



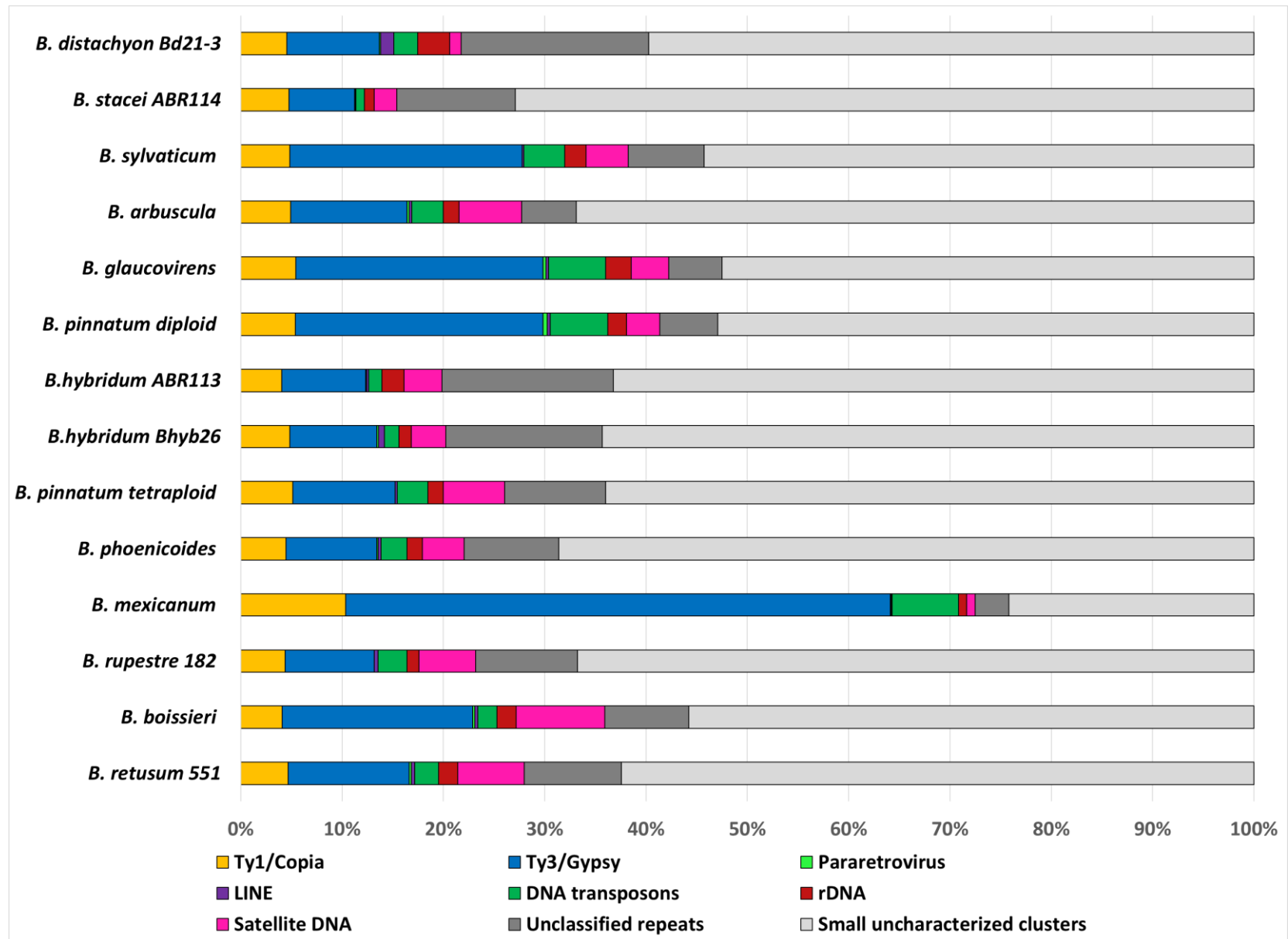
# Figure 22

## Figure 22

**Relative abundance of major repeat types in the genomes of *Brachypodium* species.**

The repetitive sequences from small clusters (constituting less than 0.01% of the genome) and singlets were excluded from the figure. The total content of the repetitive sequences from large clusters has been designated as 100%.





#### 4.3.1. Satellite DNA landscape across the *Brachypodium* genus

Five major satDNA families were identified across the *Brachypodium* genus (Table 8): the centromeric satellite family, the 360 bp family, the 118 bp family, the 1800 bp family, and the 1827 bp family. In several accessions, additional species- or genotype-specific satellite repeats were also detected (Table 8). Among the identified families, the centromeric satellite and the 360 bp families were the most abundant. More detailed analysis of the centromeric satellite family revealed the presence of two subfamilies in annual species: CentBd, specific to *B. distachyon*, and CentBs, characteristic of *B. stacei*. In perennial species, CentBd and the third subfamily (CentBp) were additionally identified (Table 8, Figure 23B). Interestingly, in annual species, each diploid accession harboured only a single centromeric satellite subfamily, while two (CentBd and CentBs) subfamilies were detected in the allotetraploid accessions, reflecting inheritance from their respective diploid progenitors. In contrast, among perennial species, regardless of ploidy level, nearly all accessions contained two subfamilies: CentBd and CentBp (Table 8, Figure 23A). An exception to this pattern was observed in *B. mexicanum*, an allotetraploid perennial species, which harboured only a single centromeric satellite subfamily, CentBs (Table 8, Figure 23A). Comparison of consensus sequences showed that within particular subfamilies, the similarity varied between 97-100%, while the similarity between subfamilies varied in the range 82-87% (Figure 23B).

Southern hybridisation with a sequence of centromeric satDNA isolated from *B. distachyon* as a probe was performed to *Ava*II-digested genomic DNA from *Brachypodium* species. (Figure 23C). In most analysed species, only one restriction site of *Ava*II was present (restriction map, Figure 23C). For perennial diploids (*B. arbuscula*, *B. pinnatum*, *B. glaucovirens*) and some allopolyploids *B. phoenicoides* (4x), *B. pinnatum* (4x), *B. retusum* (6x) and *B. rupestre* (6x), the typical for tandem repeats ladder-like pattern of hybridisation signals was observed (Figure 23C). However, in allopolyploids *B. phoenicoides* (4x), *B. pinnatum* (4x) and *B. retusum* (6x), except for the ladder-like pattern, the smear was also present (Figure 23C). Only smear and lack of ladder-like pattern were observed in allohexaploid *B. boissieri*, due to the presence of a second restriction site (restriction map b) in 156 bp satDNA sequence (Figure 23C), while a barely visible signal was observed for allotetraploid *B. mexicanum* and diploid *B. sylvaticum* Ain-1 (Figure 23C).

ML analysis was performed using two datasets (i) consensus sequences obtained from RepeatExplorer2 analyses (Figure 24A); (ii) cloned sequences of centromeric satDNA family from all analysed *Brachypodium* sp. (Figure 24B). The length of the consensus sequences ranged from 156 bp to 159 bp, and the final alignment, used for phylogenetic analyses, was 160 bp long (including gaps). The length of cloned monomer sequences ranged from 152 bp to 159

bp and the final alignment, used for phylogenetic analyses, was 163 bp long (including gaps). The analysis of consensus sequences resulted in three clades, representing all three subfamilies (Figure 24A): (i) first clade consists of consensus sequences from *B. distachyon*, *B. hybridum* (both ABR113 and Bhyb26) and all perennial accessions, besides *B. mexicanum*; (ii) second clade comprised consensus sequences of *B. stacei*, *B. hybridum* (both ABR113 and Bhyb26) and *B. mexicanum*; (iii) third clade included only consensus sequences of all perennial species, besides *B. mexicanum*.

However, the analysis of cloned sequences showed a more complex picture. While *stacei*-like sequences (isolated from *B. stacei*, *B. hybridum* ABR113 and Bhyb26) were included in one well-supported group (Figure 24B, highlighted with blue, BS100). The second well-supported group (BS95) contained sequences isolated from *B. mexicanum*, which showed similarity to *B. stacei*-like variants (BS97). The third largest group (BS100) consist of perennial and *distachyon*-like sequences isolated from all analysed species, besides *B. stacei* (Figure 24B; highlighted in green and olive/yellow strips). In this group, a subgroup consisting of sequences isolated from the *B. distachyon*, both *B. hybridum* accessions and one clone from *B. mexicanum* (highlighted in green) could be distinguished. The analysis of the nucleotide sequence of this *B. mexicanum* clone showed that it is much more similar (89%) to the *B. distachyon*-like variant than the *B. stacei*-like (80%) variant (Figure 24B).

The 360 bp satellite family was only found in perennial species. Besides *B. mexicanum* all analysed perennial accessions harboured this satDNA family. The detailed analysis of this family showed the presence of two subfamilies (360A and 360B) with 84% similarity between subfamilies (Figure 25A and E). Most of the species were characterised by the presence of 360A subfamily (Figure 25A and E). Interestingly, two different diploid genotypes of *B. sylvaticum* appeared to harbour different subfamilies. In genotype Ain-1, subfamily 360B was identified, while in Bsyl434, subfamily 360A was found (Table 8). Besides *B. sylvaticum* Ain-1, two allohexaploid species (*B. boissieri* and *B. retusum*) showed the presence of 360B subfamily (Figure 25A and E). ML analysis of consensus sequences resulted in two groups (BS100), representing both subfamilies (Figure 25A). The first group included diploid *B. sylvaticum* Ain-1 and two allohexaploid species *B. boissieri* and *B. retusum* (Figure 25A). The second group included all other perennials: diploids (*B. sylvaticum* Bsyl434, *B. arbuscula*, *B. glaucovirens*, *B. pinnatum*) and polyploids (*B. pinnatum*, *B. phoenicoides*, *B. rupestre*; Figure 25A). FISH with 360 bp satDNA isolated from *B. arbuscula* as a probe, revealed its localization predominantly in subtelomeric regions of chromosomes. In diploid *B. pinnatum* ( $2n = 16$ ) three chromosome pairs exhibited the signal in subtelomeric region of one chromosome arm, interstitial signals were observed on one chromosome pair, while four chromosome pairs

exhibited the lack of signal (Figure 25 B). In diploid *B. sylvaticum* Bsyl434 ( $2n = 18$ ) four chromosome pairs exhibited the signal in subtelomeric region of one chromosome arm, while five chromosome pairs exhibited the lack of signal (Figure 25 C). In diploid *B. arbuscula* ( $2n = 18$ ) four chromosome pairs exhibited the signal in subtelomeric region of both chromosome arms, three chromosome pairs exhibited the signal in subtelomeric region of only one chromosome arm, interstitial signals were observed on one chromosome pair, while only one chromosome pairs exhibited the lack of signal (Figure 25 D). Interestingly, in *B. pinnatum*, the chromosome pair bearing the 5S rDNA signals also carried signals of the 360 bp satDNA (Figure 25 B). In contrast, in *B. sylvaticum* and *B. arbuscula*, the chromosomes with 5S rDNA did not show any 360 bp satDNA signals (Figure 25 C and D). Notably, in *B. arbuscula*, the chromosome pair with 5S rDNA was the only one lacking 360 bp satellite signals (Figure 25 D).

1800 bp satDNA family was identified with RepeatExplorer2 only in a few accessions: in annual diploids, in diploid *B. sylvaticum* and in allotetraploid *B. pinnatum* (Figure 26 A). However, this satellite family was also detected in evolutionary older annual allotetraploid *B. hybridum* Bhyb26 and in perennial allopolyploids *B. phoenicoides*, *B. rupestre* and *B. boissieri* with Southern hybridisation (Figure 26 B). In annual diploid *B. stacei* the band representing the monomer size (~1800 bp) was obtained (Figure 26B), while in *B. distachyon*, due to the presence of the second restriction site for *AvaII* in this repeat, a significantly weaker band representing shorter sequence (~1600 bp) was observed (Figure 26B). In *B. hybridum* Bhyb26, similarly to *B. distachyon* the band representing a sequence of approximately 1600 bp was obtained, while in *B. hybridum* ABR113, there was no signal (Figure 26B). In perennial diploid *B. sylvaticum* and allohexaploids *B. rupestre*, similarly to *B. distachyon*, the band representing a sequence of approximately 1600 bp was obtained (Figure 26B). In *B. boissieri*, the band was slightly shorter, while allotetraploid *B. phoenicoides* and allotetraploid *B. pinnatum* were characterised by the presence of ~1800 bp band (Figure 26B). The abundance of the 1800 bp satellite family was too low to be detected by FISH.

Among satDNA families with lower abundance, 118 bp family was identified in most analysed accessions (Table 8). In annuals it was only identified in evolutionary older *B. hybridum* Bhyb26; however, among perennials it was identified in all accessions, besides both genotypes of *B. sylvaticum* and allotetraploid *B. pinnatum* (Table 8). Interestingly, this satDNA family was also identified in *B. mexicanum* (Table 8). 1827 bp satDNA family was identified only in perennial diploids *B. glaucovirens*, *B. pinnatum* and in allotetraploid *B. pinnatum*; however, it was rather low abundant (Table 8).

**Table 8.** Major satDNA families identified in single species RepeatExplorer2 analyses of *Brachypodium* species.

Accession	Centromeric satellite family			360 bp family		118bp family	1800bp family	1827bp family	Species-specific
	CentBd	CentBs	CentBp	360A	360B				
<i>B. distachyon</i>	0.68%						0.01%		2
<i>B. stacei</i>		1.5%					0.044%		-
<i>B. hybridum</i> ABR113	0.26%	2.1%							1
<i>B. hybridum</i> Bhyb26	0.76%	1.2%				0.025%	detected in SB		1
<i>B. sylvaticum</i> Ain-1	0.087%		0.2%		0.18%		0.034%		2
<i>B. sylvaticum</i> Bsyl434	0.23%		1.7%	0.9%			0.022%		2
<i>B. arbuscula</i>	0.975%		0.981%	2.1%		0.032%			-
<i>B. glaucovirens</i>	0.23%		1.2%	1.5%		0.035%		0.016%	1
<i>B. pinnatum</i> 2x	0.29%		1.2%	0.97%		0.038%		0.021%	-
<i>B. pinnatum</i> 4x	1%		2.6%	0.2%			0.015%	0.033%	1
<i>B. phoenicoides</i>	0.77%		1.8%	0.23%		0.031%	detected in SB		-
<i>B. mexicanum</i>		0.49%				0.45%			5
<i>B. boissieri</i>	0.19%		4.2%		0.17%	0.8%	detected in SB		-
<i>B. rupestre</i> 182	1.1%		2.4%	0.082%		0.02%	detected in SB		-
<i>B. retusum</i> 551	0.64%		3%		0.076%	0.11%			1

# Figure 23

## Figure 23

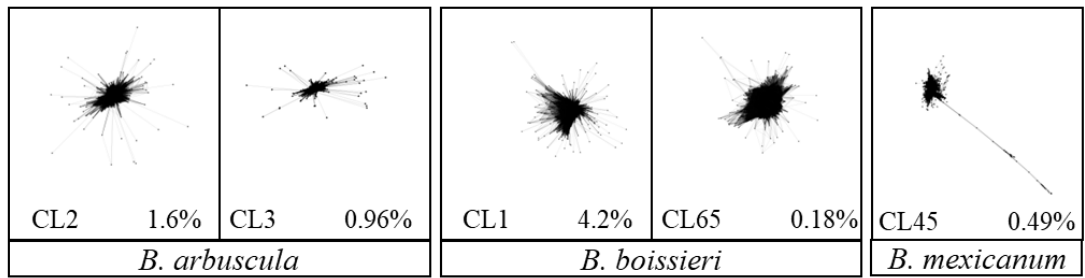
### **Analyses of centromeric satellite family in studied *Brachypodium* species.**

A      Examples of repeat cluster graphs of centromeric satellite family constructed in RepeatExplorer2 and TAREAN analysis for diploid *B. arbuscula*, allohexaploid *B. boissieri*, and allotetraploid *B. mexicanum*.

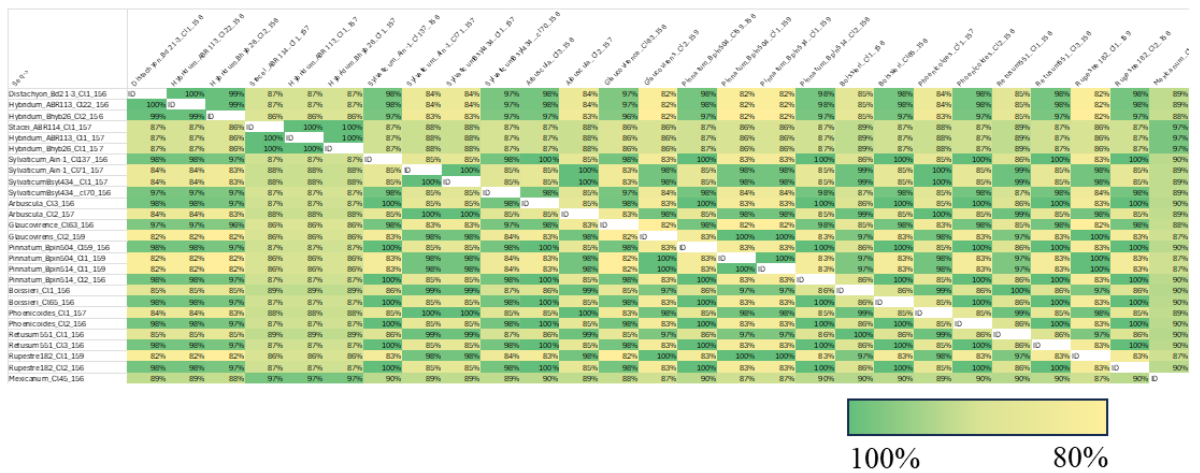
B      Identity matrix of centromeric satellite family consensus sequences, obtained from RepeatExplorer2 clustering analysis. The similarity of sequences showed in percentages. The colour represents the similarity level.

C      Southern hybridisation of isolated centromeric satDNA to genomic DNA from *Brachypodium* species restricted with *Ava*II. The restriction map is shown above the photo of a membrane.

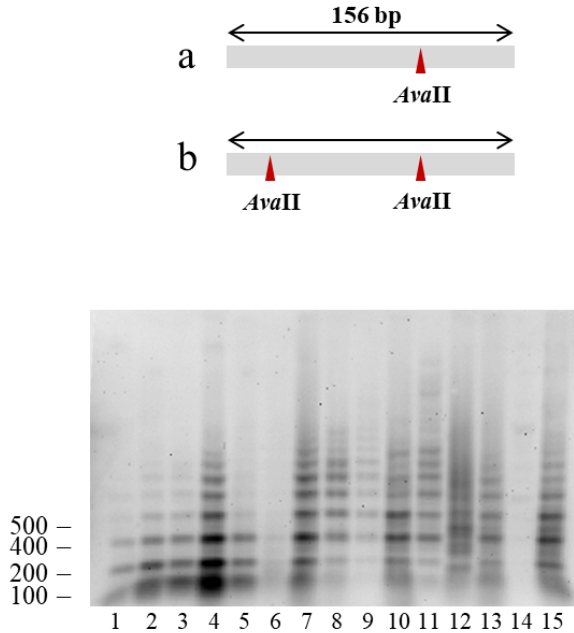
A



B



C



- 1 – *B. distachyon* (Bd21)
- 2 – *B. hybridum* (ABR 113)
- 3 – *B. stacei* (ABR 114)
- 4 – *B. hybridum* (Bhyb26)
- 5 – *B. distachyon* (ABR 5)
- 6 – *B. sylvaticum* Ain-1
- 7 – *B. arbuscula*
- 8 – *B. pinnatum* (2x)
- 9 – *B. glaucovirens*
- 10 – *B. phoenicoides*
- 11 – *B. rupestre*
- 12 – *B. boissieri*
- 13 – *B. retusum*
- 14 – *B. mexicanum*
- 15 – *B. pinnatum* (4x)



# Figure 24

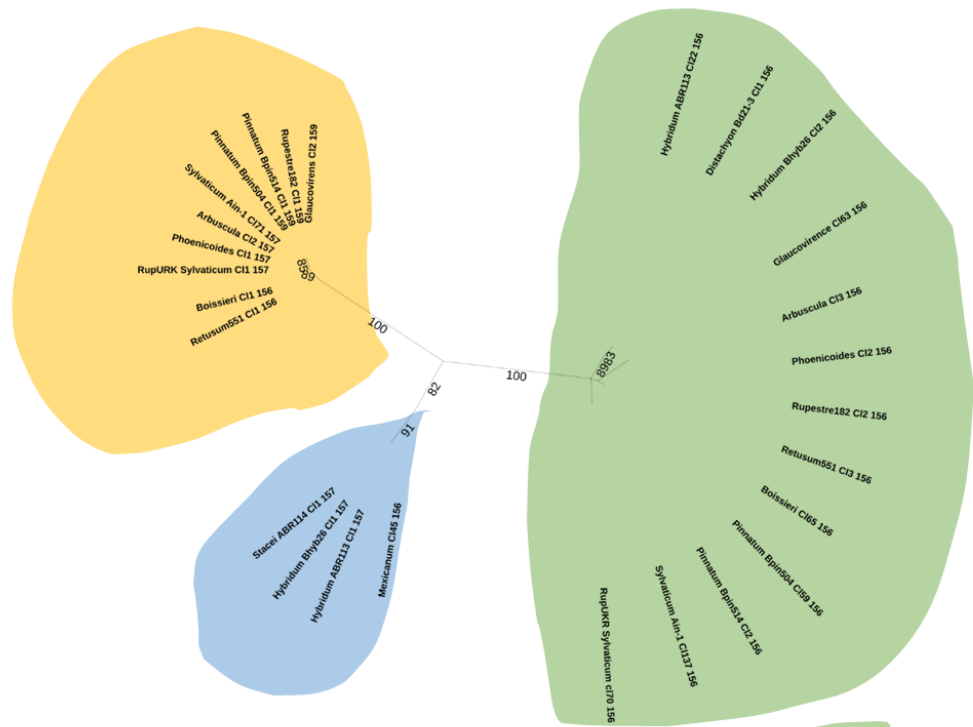
## Figure 24

### **Phylogenetic analyses of the centromeric satellite sequences isolated from studied *Brachypodium* species.**

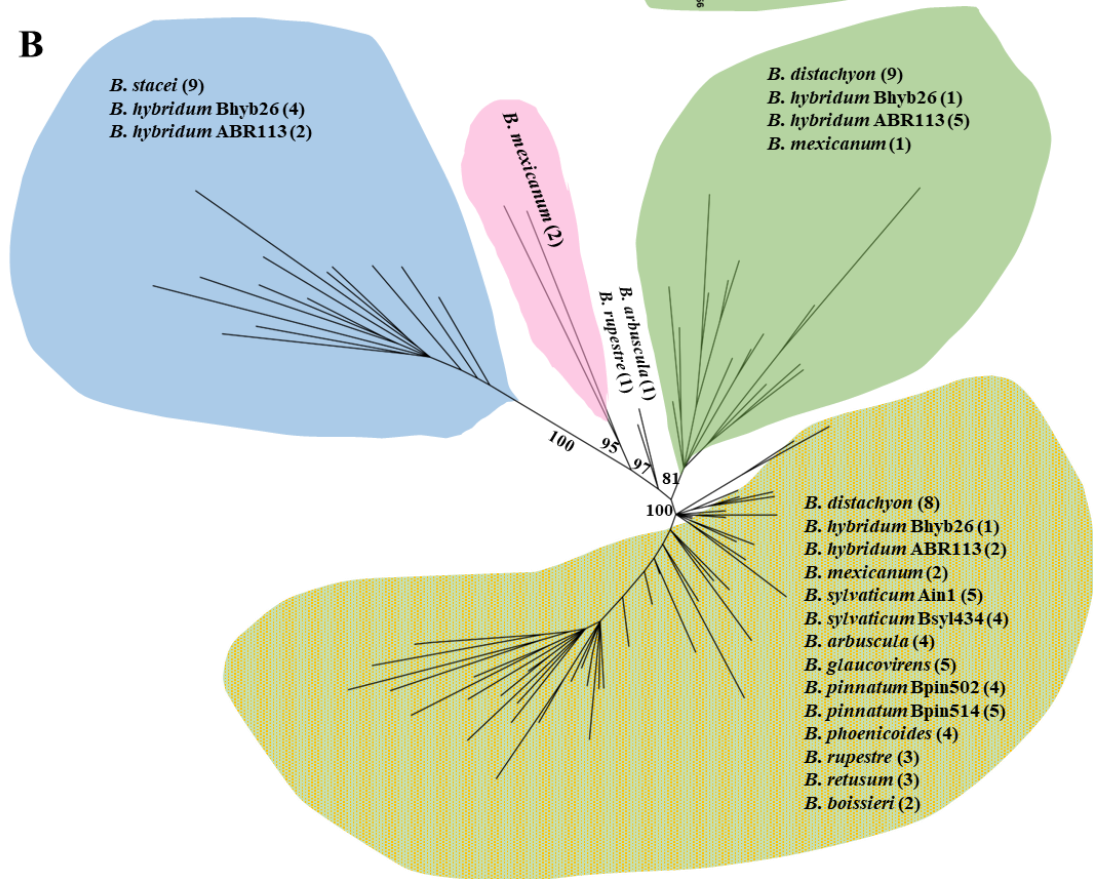
A      The ML unrooted phylogenetic tree of the centromeric satellite consensus sequences reconstructed for studied *Brachypodium*. Bootstrap scores are shown near the nodes. The values below 70 were excluded from figure.

B      The ML unrooted phylogenetic tree of the cloned centromeric satellite sequences isolated from the studied *Brachypodium* species. The number of clones belonging to the group is shown in brackets. Bootstrap scores are shown near the nodes. The values below 70 were excluded from figure.

**A**



**B**



# Figure 25

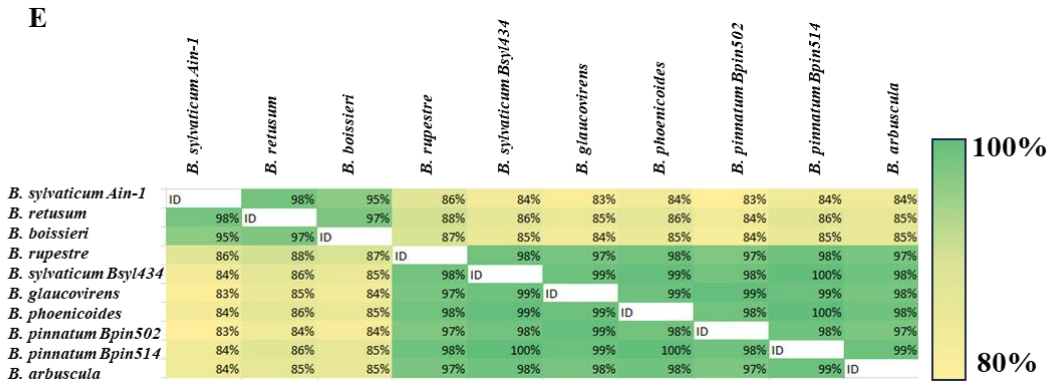
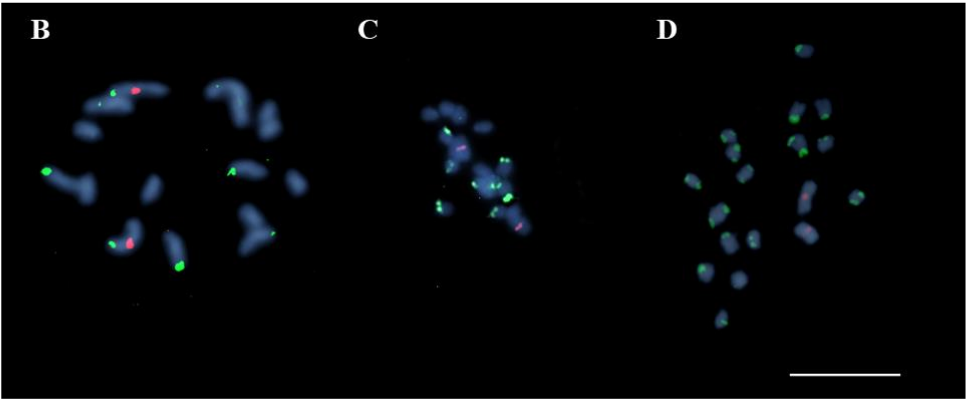
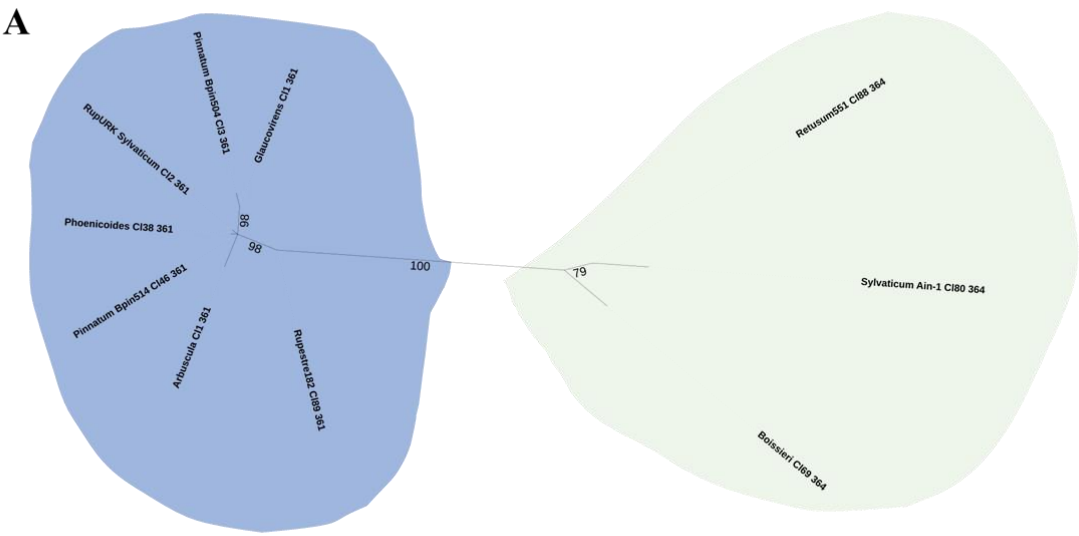
## Figure 25

### 360 bp satellite family in studied *Brachypodium* species.

A The ML unrooted phylogenetic tree of the 360 bp satDNA consensus sequences reconstructed for the studied *Brachypodium* species. The 360A bp subfamily is highlighted in blue, while the 360B bp subfamily is highlighted in mint. Bootstrap scores are shown near the nodes. The values below 70 were excluded from figure.

B-D Distribution of the 360 bp satellite (**green fluorescence**) and the 5S rDNA probe (**red fluorescence**) on the mitotic metaphase chromosomes counterstained with DAPI (**blue fluorescence**); (**B**) *B. pinnatum* ( $2n = 16; x = 8$ ), (**C**) *B. sylvaticum* Bsyl434 ( $2n = 18; x = 9$ ), (**D**) *B. arbuscula* ( $2n = 18; x = 9$ ).

E Identity matrix of 360 bp satellite family consensus sequences, obtained from RepeatExplorer2 clustering analysis. The similarity of sequences is shown in percentages. The colour represents the similarity level.



# Figure 26

## Figure 26

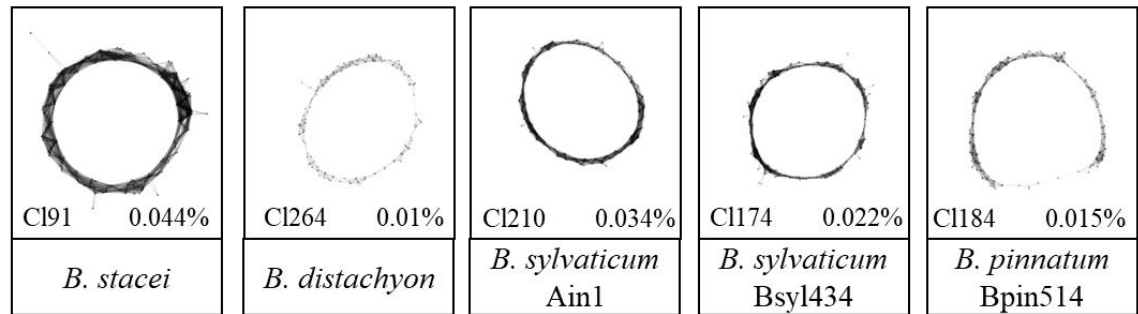
### 1800 bp satellite family in the studied *Brachypodium* species.

A Cluster graphs of 1800 bp satellite family reconstructed in RepeatExplorer2 and TAREAN analysis for annual diploids *B. distachyon* and *B. stacei*, perennial diploid *B. sylvaticum* Ain1 and Bsyl434, and allotetraploid *B. pinnatum*. The cluster number and proportion of the genome in percentages are shown below the graphs.

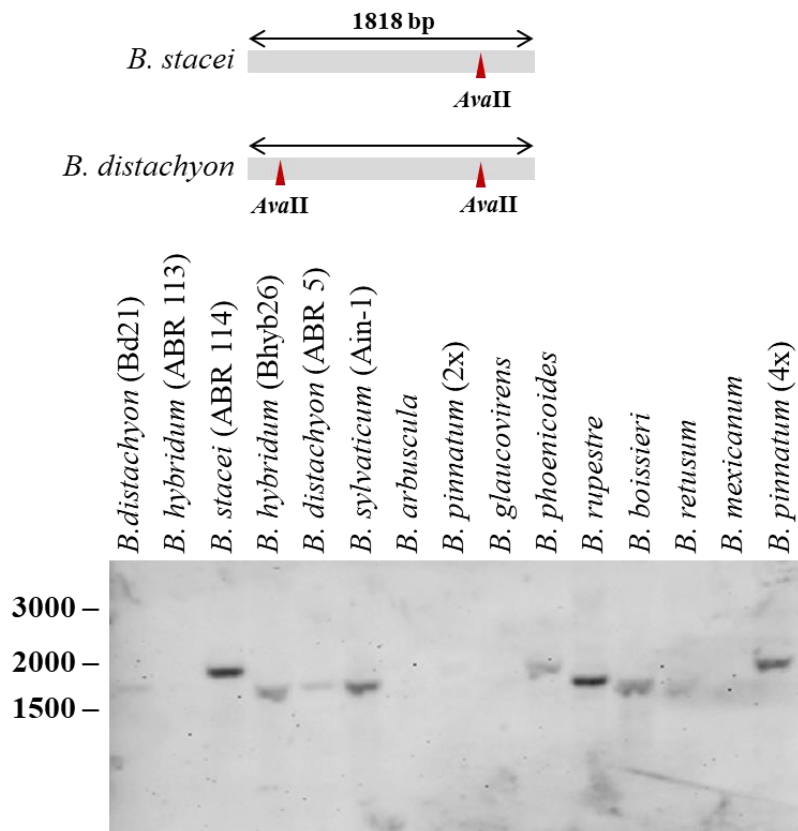
B Southern hybridisation of isolated 1800 bp satDNA to genomic DNA digested with *Ava*II from *Brachypodium* species. The restriction map of the 1800 bp repeat is shown above the photo of a membrane.



**A**



**B**



#### 4.3.2. Comparative analysis of 5S rDNA in *Brachypodium* sp.

Sequences corresponding to 5S rDNA were identified using RepeatExplorer2. Each analysed accession yielded a single cluster containing reads complementary to the 5S rDNA. A simple circular cluster graph was reconstructed for diploids (*B. distachyon*, *B. stacei*, *B. arbuscula*, *B. sylvaticum*, *B. glaucovirens* and *B. pinnatum*) and for some allopolyploids (evolutionary older *B. hybridum* Bhyb26 (4x), *B. pinnatum* (4x), *B. boissieri* (6x), indicating the presence of one variant of 5S rDNA in each species (Figure 27A). For other allopolyploids (evolutionary younger *B. hybridum* ABR113 (4x), *B. mexicanum* (4x), *B. phoenicoides* (4x), *B. retusum* (6x) and *B. rupestre* (6x) more complex graph structures consisting of two loops connected by a junction region that included the 5S rDNA coding sequence were reconstructed, suggesting the presence of two distinct non-transcribed spacer (NTS) variants (Figure 27A).

The comparative RepeatExplorer2 analysis resulted with one cluster corresponding to 5S rDNA (Figure 27B). The complex graph consisted of 8 NTS loops (representing similar variants of NTS sequence) interconnected by a junction section (coding region of the 5S rDNA): 1) loop shared by *B. distachyon* (light pink) and evolutionary younger *B. hybridum* ABR113 (red); 2) loop shared by allotetraploid *B. phoenicoides* (maroon) and allohexaploid *B. retusum* (peach); 3) loop shared by *B. stacei* (light green) and evolutionary younger *B. hybridum* ABR113 (red); 4) loop represented only by allotetraploid *B. mexicanum* (dark blue); 5) loop represented only by evolutionary older *B. hybridum* Bhyb26 (neon green); 6) loop represented allohexaploid *B. boissieri* (blue) and allohexaploid *B. retusum* (peach); 7) loop represented only allohexaploid *B. rupestre* (olive); 8) loop included all perennial diploids (*B. arbuscula* (yellow), *B. glaucovirens* (dark green), *B. pinnatum* (navy blue), *B. sylvaticum* (aqua) and polyploids (*B. pinnatum* (fuchsia), *B. retusum* (peach), *B. rupestre* (olive), *B. phoenicoides* (maroon) (Figure 27B).

Southern blot analysis was conducted to confirm the genomic organisation of 5S rDNA in *Brachypodium* species. Single restriction site for *MseI* enzyme of the 5S rDNA sequence was present in *B. stacei* and both allotetraploid genotypes of *B. hybridum*. In contrast, there were two sites for this enzyme in the corresponding sequence of *B. distachyon* (Figure 27C). Therefore, after the Southern hybridisation, a single band of approximately 270 bp was observed in *B. stacei*, while in *B. distachyon*, the 5S rDNA probe hybridised to two shorter *MseI* fragments of approximately 240 bp and 130 bp (Figure 27C). For most perennials (diploid *B. arbuscula*, *B. sylvaticum*, *B. glaucovirens*, *B. pinnatum* and polyploid *B. rupestre* (6x), *B. retusum* (6x) and *B. pinnatum* (4x) the 5S rDNA probe hybridised to two *MseI* fragments of approximately 240 bp and 130 bp (as shown in annual *B. distachyon*; Figure 27C). A single 250

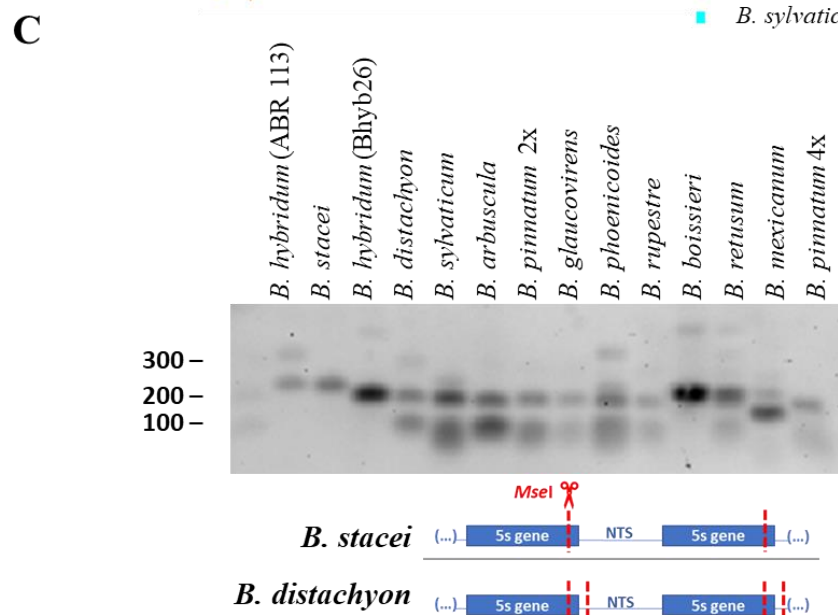
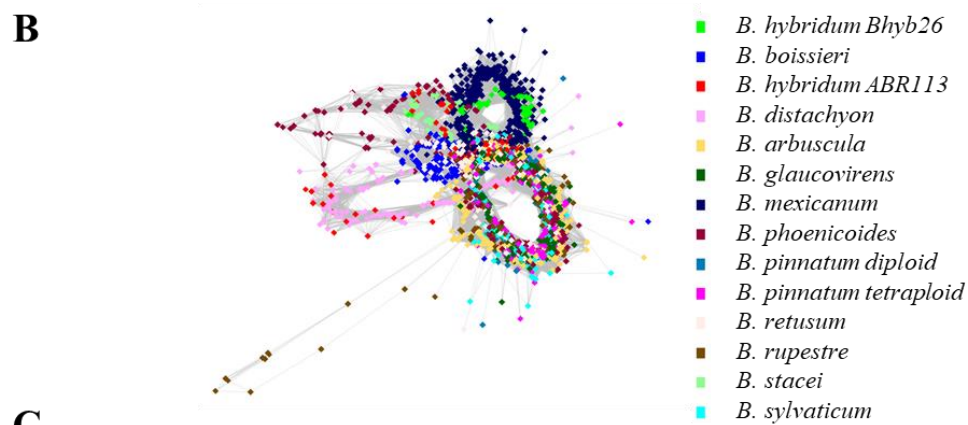
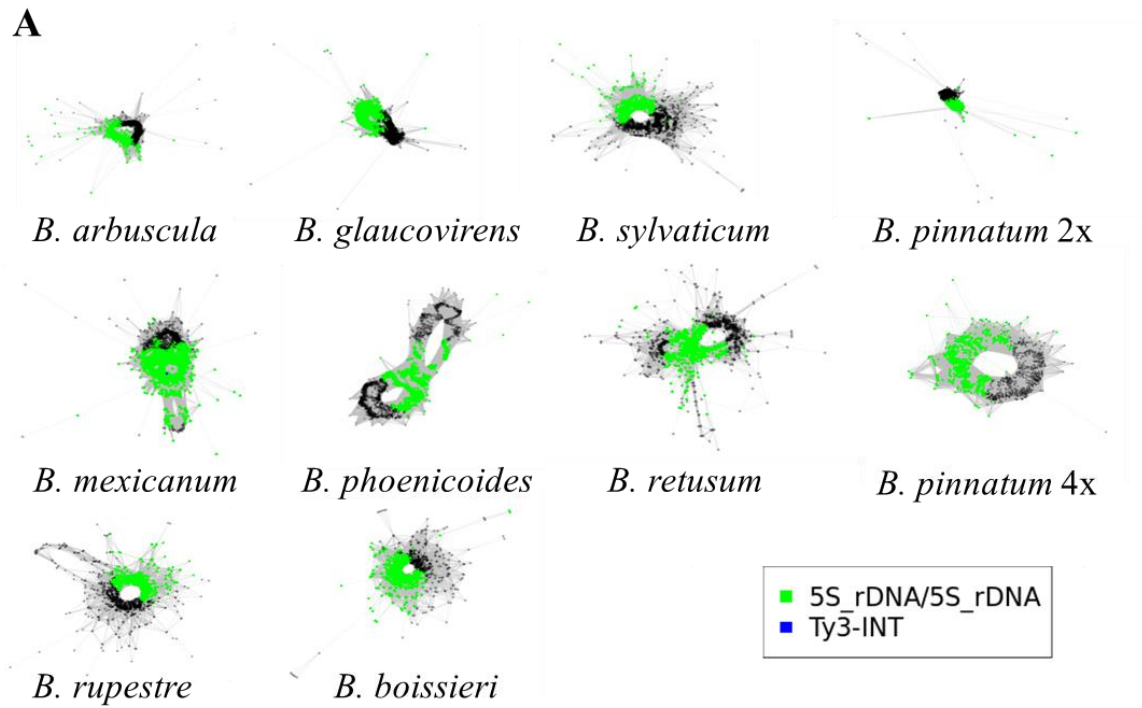
bp long band was observed for allohexaploid *B. boissieri* (Figure 27C). For *B. mexicanum* two fragments of approximately 200 bp and 150 bp were observed (Figure 27C).

# Figure 27

## Figure 27

### Structure of 5S rDNA in analysed *Brachypodium* species.

- A Cluster graphs resulted in single species RepeatExplorer2 analysis for *Brachypodium* *perennials*.
- B Cluster graph resulted from comparative RepeatExplorer2 analysis for all analysed *Brachypodium* accessions.
- C Southern hybridisation of 5S rDNA probe to genomic DNA from all analysed *Brachypodium* accessions restricted with *Mse*I. The restriction map is showed below the membrane photography.



#### 4.3.3. Composition of LTR-retrotransposons in *Brachypodium* genus

The genomic composition of LTR-retroelements of annual *Brachypodium* accessions was described in detail in chapter 4.2.3. In the analysed perennial *Brachypodium* accessions, the most highly amplified repeats within large clusters (constituting more than 0.01% of the genome) were predominantly LTR-retrotransposons. In total, 13 major LTR-retrotransposon families were identified. These included five families from the Ty3/gypsy superfamily: Athila, Ogre, Tat/Retand, Tekay and CRM and eight families belonged to the Ty1/copia superfamily: Tar, SIRE, Angela, Ivana, Tork, Bianca, Ikeros, and Ale (Figure 28). In diploid *B. sylvaticum* among Ty3/gypsy, Tekay family was the most abundant (ca. 45% of all identified Ty3/gypsy LTR-retroelements), followed by Tat/Retand (nearly 35%) and CRM (nearly 10%; Figure 28A). Among Ty1/copia superfamily SIRE elements were the most abundant (nearly 40% of all identified Ty1/copia LTR retrotransposons), followed by Angela and Tar, each contributing around 15% of all identified Ty1/copia LTR-retroelements (Figure 28B). Other families of LTR-retrotransposons were rather low abundant. Similar to *B. sylvaticum*, in diploids *B. pinnatum* and *B. glaucovirens*, the Ty3/gypsy Tekay family was the most abundant (ca. 45% of all identified Ty3/gypsy LTR-retroelements), followed by Tat/Retand (40%) and CRM (nearly 5%, Figure 28A). Among the Ty1/copia elements SIRE family was the most abundant (nearly 45%), followed by Tar (nearly 15%; Figure 28B). Other families of LTR-retrotransposons were rather low abundant. In contrast, in diploid *B. arbuscula* Tat/Retand was the most amplified family (nearly 65% of all identified Ty3/gypsy LTR-retrotransposons), followed by Tekay, that made up approximately 20% of all identified Ty3/gypsy LTR-retrotransposons (Figure 28A). Among Ty1/copia elements SIRE family was the most amplified (nearly 45%), followed by Ikeros (nearly 10%). Other families of LTR-retrotransposons were rather low abundant (Figure 28B).

In allotetraploid *B. mexicanum* Tat/Retand and Tekay each constituted approximately 40-45% of all identified Ty3/gypsy retrotransposons, while other families were rather low abundant (Figure 28A). Among Ty1/copia SIRE and Angela were the most abundant accounting for approximately 40% and 30%, respectively (Figure 28B). In other polyploid accessions (*B. pinnatum*, *B. phoenicoides*, *B. retusum*, *B. boissieri* and *B. rupestre*), Tat/Retand family was the most abundant among Ty3/gypsy retrotransposons, with approximately 70 – 75% of all identified Ty3/gypsy elements (Figure 28A). The abundance of Tekay and CRM families showed variability in the analysed polyploid accessions (Figure 28A). In allotetraploid *B. phoenicoides*, Tekay family made up approximately 30% of all identified Ty3/gypsy LTR-retrotransposons, while in other polyploids Tekay and CRM accounted for approximately 5 – 15% of all identified Ty3/gypsy LTR-retrotransposons (Figure 28A). Other Ty3/gypsy families

were rather low abundant in polyploids (Figure 28A). Among Ty1/*copia* in all analysed polyploids SIRE family was the most abundant with approximately 55-60% of all identified Ty1/*copia* elements, followed by Tar with approximately 15-20% (Figure 28B). Other families of LTR-retrotransposons were rather less abundant (Figure 28B).



# Figure 28

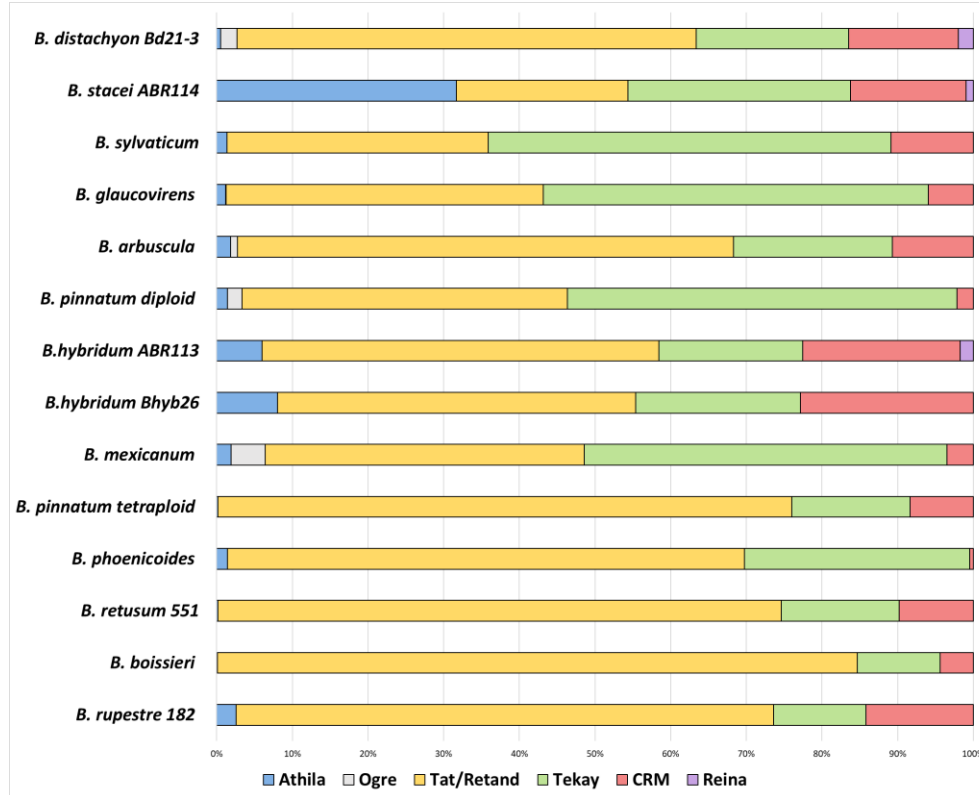
## Figure 28

**The composition of LTR-retrotransposons in the genomes of all analysed *Brachypodium* species.**

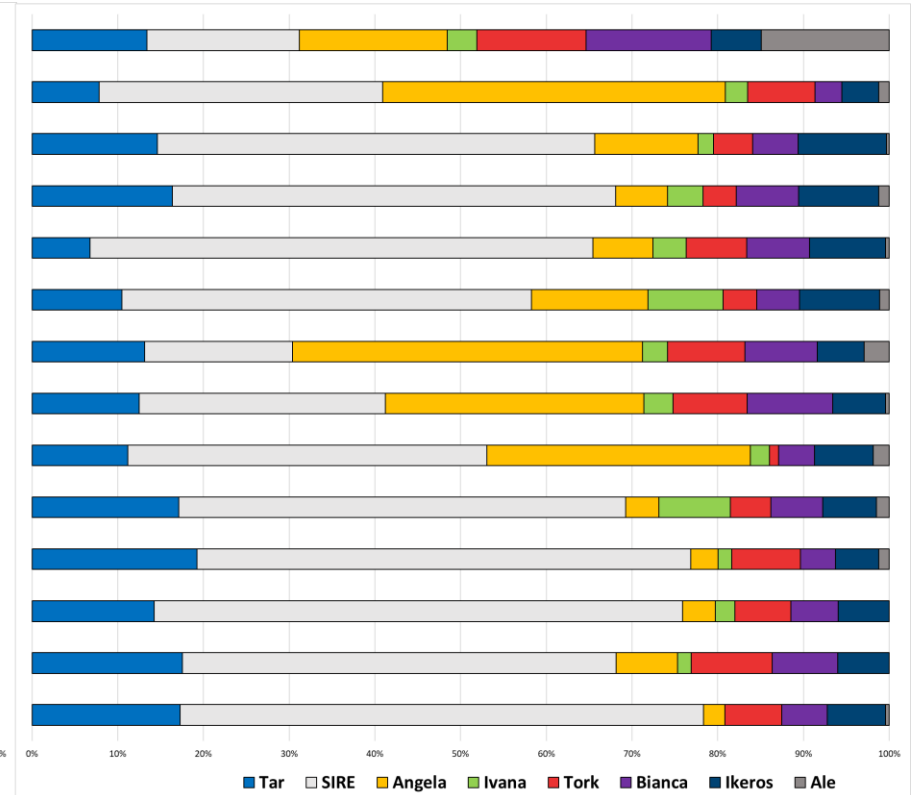
The total content of the LTR-retrotransposons superfamily within large clusters (each constituting more than 0.01% of genome) has been designated as 100%.

- A      Composition of the Ty3/*gypsy* LTR retrotransposon.
- B      Composition of the Ty1/*copia* LTR retrotransposon.

A



B



#### 4.3.4. Comparative analysis of repeatome composition in *Brachypodium* genomes

To assess the diversity of repetitive DNA across the *Brachypodium* genus, the comparative RepeatExplorer2 analysis was performed for 14 *Brachypodium* accessions of different ploidy levels, including both annual and perennial species (Figure 29). LTR-retrotransposons were the most abundant class of repeats in all species, with Ty3/gypsy elements being the most amplified (Figure 29). With the exception of *B. mexicanum*, the overall composition of the repetitive DNA landscape was rather similar across the analysed *Brachypodium* species. Notable divergence was observed primarily in the distribution and abundance of satDNAs, which displayed more variation among species, especially between annuals and perennials (yellow squares, Figure 29). For instance, comparative analysis showed the presence of CentBp (C11, green diamond, Figure 29) and 360 bp (C133, yellow diamond, Figure 29) satDNA families in most of the perennials; however, these families were almost absent in annuals and *B. mexicanum* (Figure 29). Annual *Brachypodium* accessions, both diploids and allotetraploids, mostly exhibited a lower total abundance of repetitive elements compared to perennials (Figure 29). The same families of repeats were present within large clusters in all analysed accessions; however, the amplification level was different (Figure 29). Especially, in allohexaploid perennials, increases in repeat content were observed (green arrow, Figure 29). For example, allohexaploid *B. boissieri* displayed notable higher amount of the 360 bp satellite DNA (C133, yellow diamond, Figure 29) family as well as the Ty3/gypsy Tat/Retand elements (aqua squares, Figure 29). The most outstanding repetitive DNA profile was revealed for allotetraploid *B. mexicanum* (blue arrow, Figure 29). High amplification of several Tat/Retand (aqua squares, Figure 29) lineages was specific for *B. mexicanum* (blue arrow, Figure 30). It has also exhibited massive amplification of DNA-transposons and LTR-retrotransposon lineages, especially within Ty3/gypsy Tekay (dark blue squares, Figure 29). It was also revealed that satDNA landscape of *B. mexicanum* differs from all other analysed accessions (yellow squares, Figure 29): three satDNA families were amplified only in *B. mexicanum* (C1232, C1266, C1287), while the 360 bp family (C133, yellow diamond, Figure 29) amplification was at the very low level (Figure 29).

# Figure 29

## Figure 29

### Comparative analysis of repeat composition among *Brachypodium* species.

The chart illustrates the distribution of repeat clusters (comprising at least 0.01% of analysed reads) among analysed accessions. The total read count per cluster is indicated by bar length, while square size reflects the genomic abundance in the respective accession. Bars and squares are coloured according to the type of repetitive element. The species names are represented with codes:

ARB – *B. arbuscula*

GLA – *B. glaucovirens*

MEX – *B. mexicanum*

PHO – *B. phoenicoides*

PI2 – *B. pinnatum* Bpin502

PI4 – *B. pinnatum* Bpin514

BHY – *B. hybridum* Bhyb26

ABR – *B. hybridum* ABR113

BDD – *B. distachyon*

STA – *B. stacei*

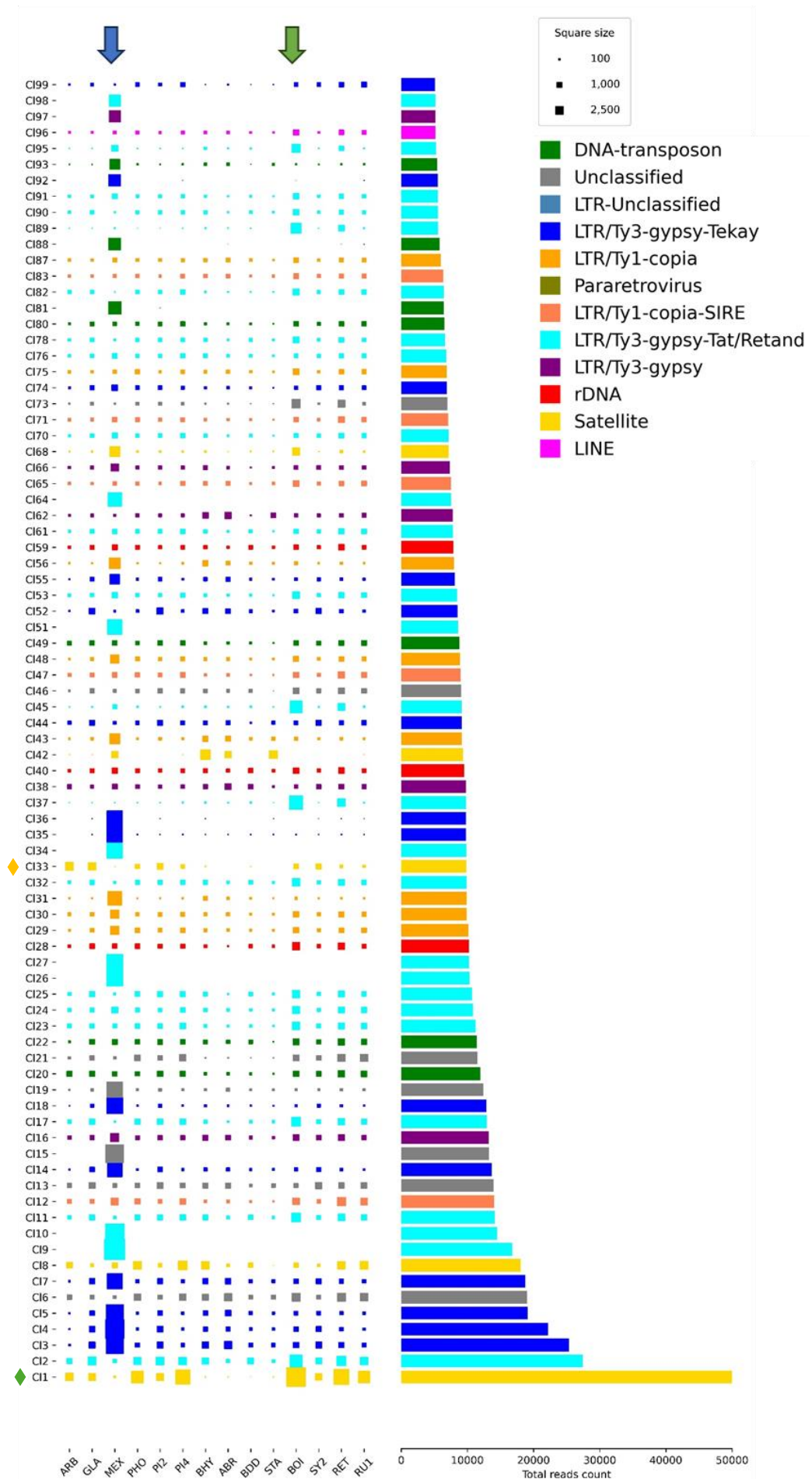
BOI – *B. boissieri*

SY2 – *B. sylvaticum* Bsyl434

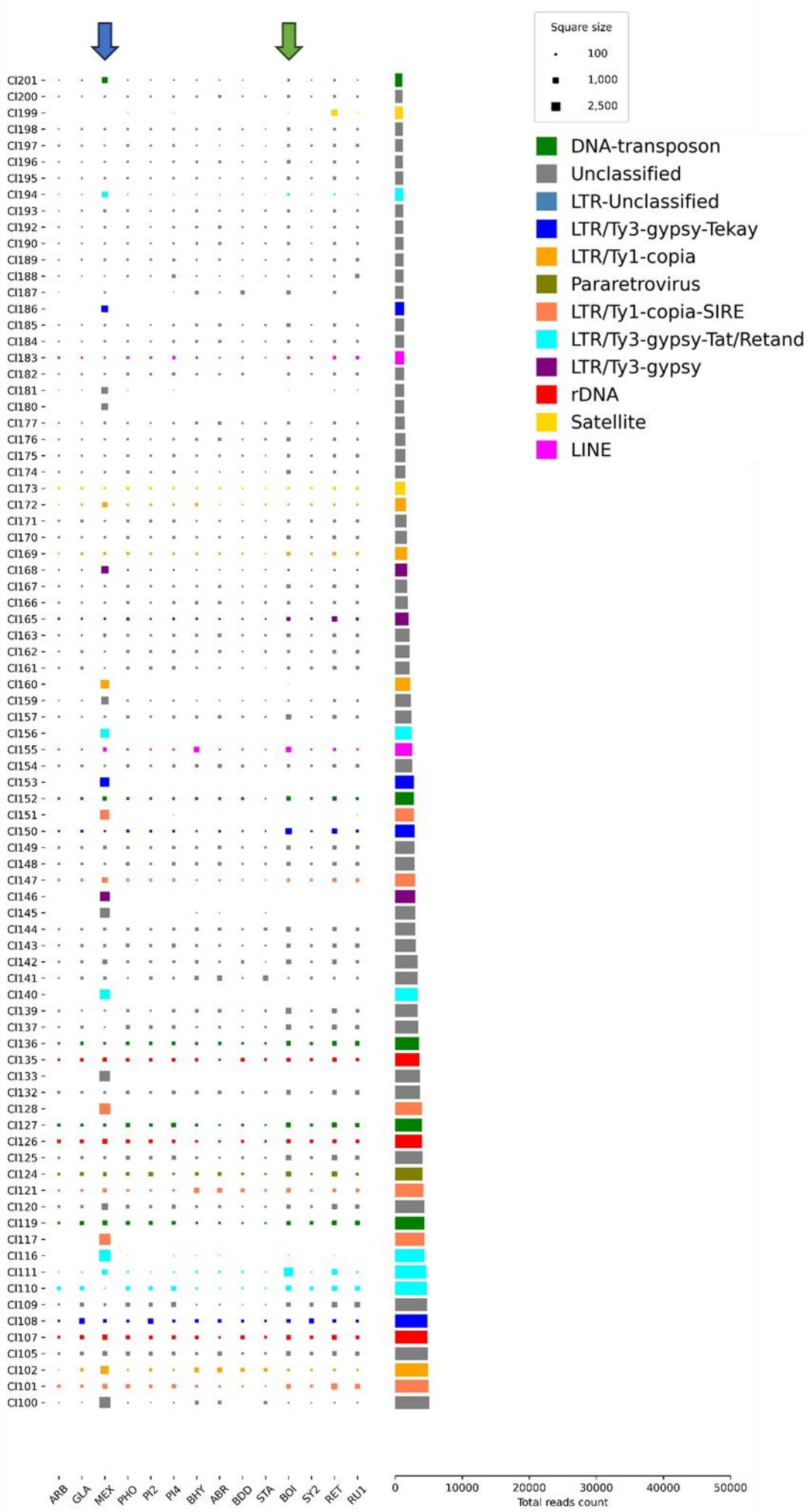
RET – *B. retusum*

RU1 – *B. rupestre*.

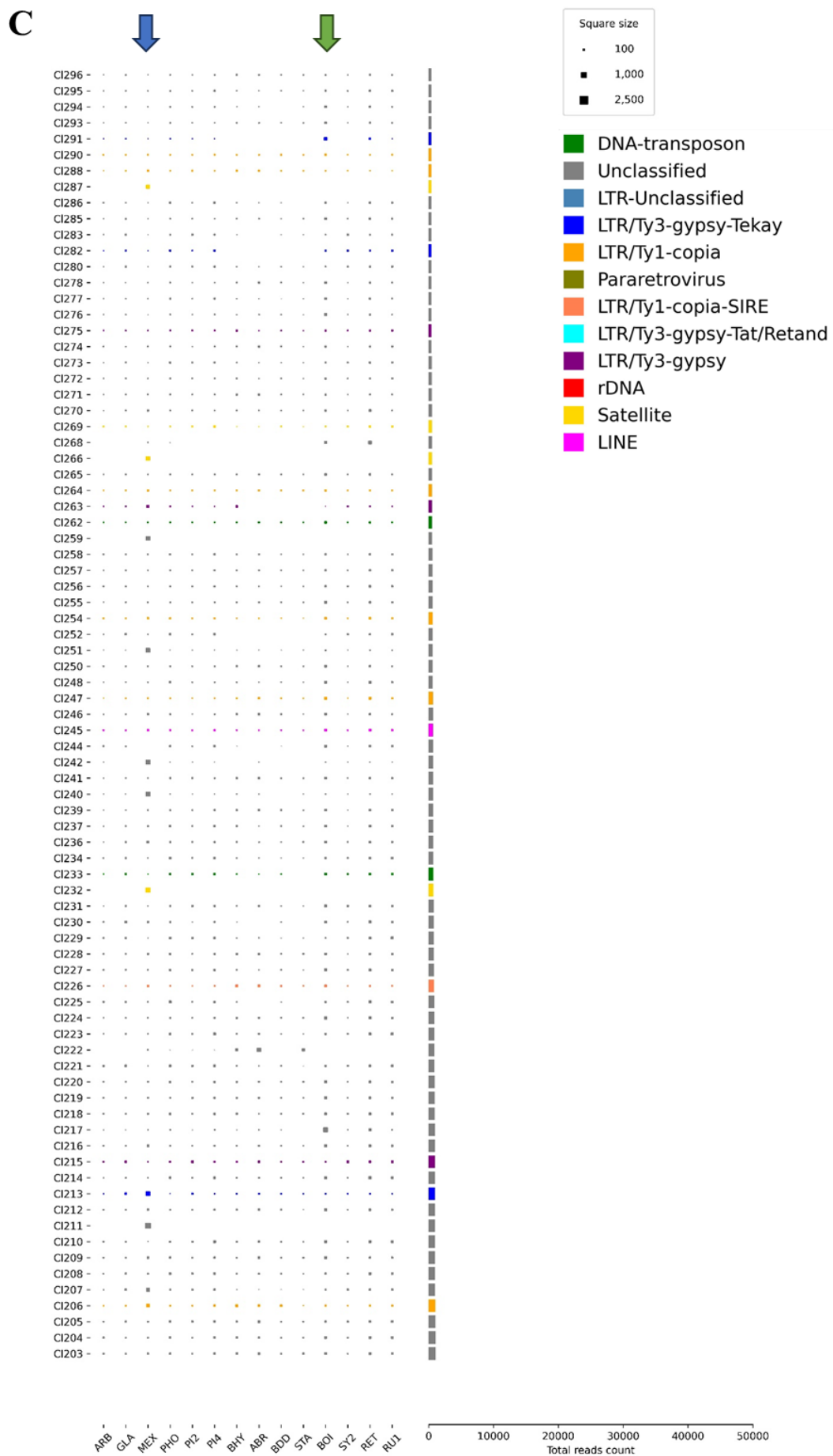
A



**B**







#### 4.4. Comparative analysis of the repetitive DNA fraction in the phylogenetic backgrounds

The genome sizes of analysed species were mapped on the ML phylogenetic tree using maximum parsimony method (Figure 30). A genome size of 0.33 pg/1C DNA was inferred as the ancestral state for the *Brachypodium* genus when only diploid species were taken into analysis (Figure 30). The analysis showed that the diversification of annual diploid species was accompanied by a genome downsizing (Figure 30). While evolution of perennial diploid *Brachypodium* species was characterised by increases in genome size. The noticeable genome size increase has happened after the split of *B. arbuscula* and *B. glaucovirens* – *B. pinnatum* group.

The presence/absence of satDNA repeats was analysed in the phylogenetic background. Several events of amplification of satDNAs were inferred when only diploid species were taken into analysis. In the common ancestor of annual (*B. distachyon* and *B. stacei*) and perennial species, two satDNA amplification events were inferred – CentBd (blue upward arrow, Figure 30) and 1800 bp (aqua upward arrow, Figure 30) families. In *B. distachyon*, both the CentBd and 1800 bp were present; however, during the speciation of *B. stacei* CentBs (green arrow, Figure 30) underwent amplification, while CentBd was eliminated (blue downward arrow, Figure 30).

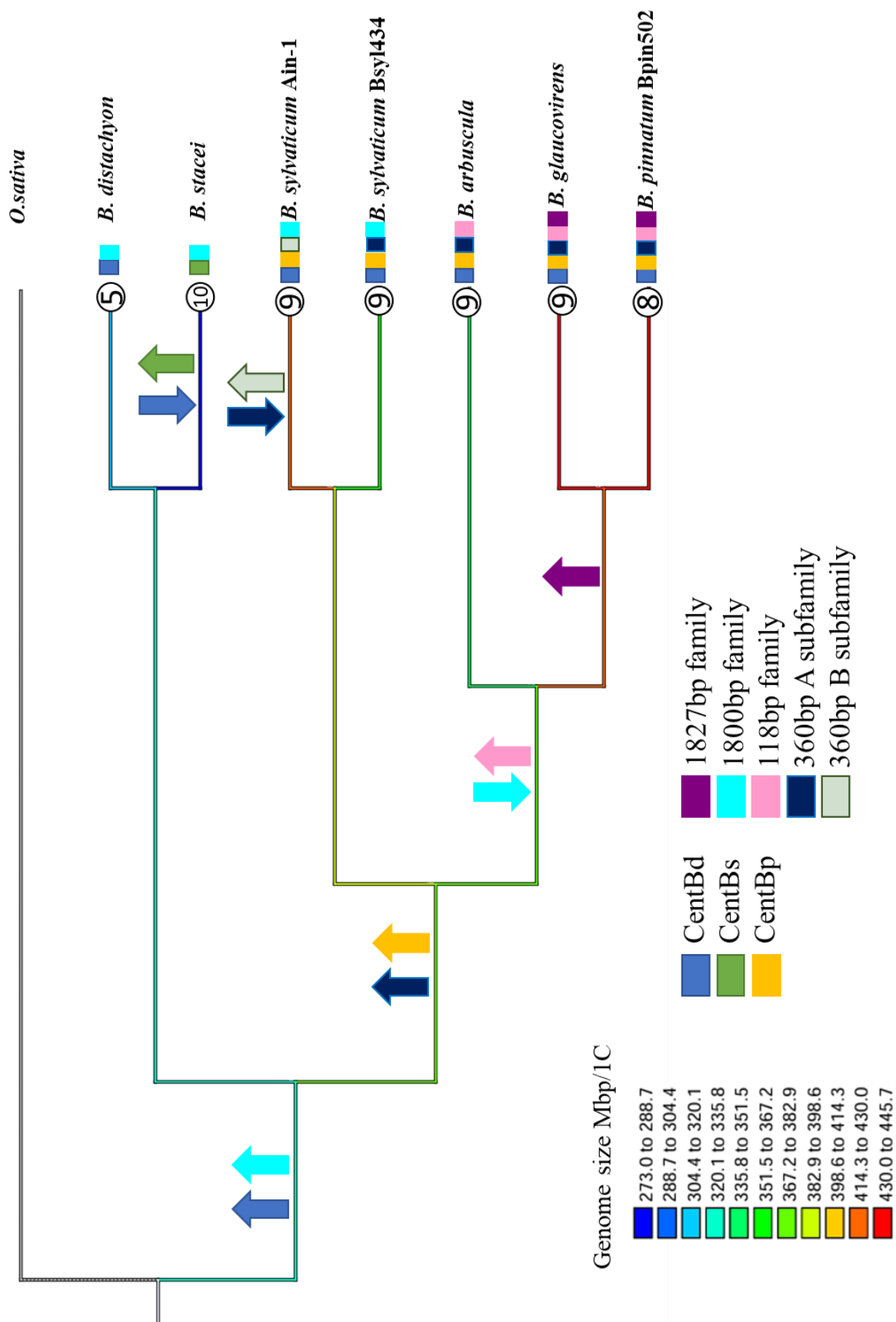
In a common ancestor of perennial *Brachypodium* species, the amplification of CentBp and 360 bp satellite families took place (yellow and dark blue arrows, respectively, Figure 30). In *B. sylvaticum*, CentBd and CentBp subfamilies of centromeric satellite remained; however, the Ain-1 accession experienced independent amplification of 360B subfamily and elimination of 360A subfamily (light green upward and dark blue downward arrows, respectively, Figure 30). In the common ancestor of *B. arbuscula*, *B. glaucovirens* and diploid *B. pinnatum*, the amplification of 118 bp satDNA family (pink upward arrow, Figure 30) and elimination of 1800 bp family (aqua downward arrow, Figure 30) were inferred. The *B. glaucovirens* – *B. pinnatum* group is marked by the amplification of the 1827 bp family (purple upward arrow, Figure 30). When we combine the data on genome size evolution with inferred events of satDNA amplification/elimination, the results suggest that changes in genome size (both increases and decreases) were accompanied by reorganization of satDNA in *Brachypodium* diploids (Figure 30).

# Figure 30

## Figure 30

### **The evolution of analysed satDNAs in diploid *Brachypodium* species.**

Genome sizes were mapped on the ML tree of *Gigantea* sequences using the maximum parsimony method as implemented in Mesquite. The colour of the branches shows the genome size. Numbers in circles represent basic chromosome number. The coloured rectangles indicate the presence of satDNA family in accessions. The arrows indicate the amplification/elimination of copies of satDNA families.



## 5. DISCUSSION

### 5.1. Phylogeny and basic chromosome number evolution

In this study, the sequences of the nuclear gene *Gigantea* (GI), a conserved gene involved in photoperiodic flowering, were used to reconstruct the phylogenetic relationships within the *Brachypodium* genus. Phylogenetic reconstruction using the GI gene supported the division of *Brachypodium* species into three major clades, consistent with previous studies using multilocus nuclear datasets (Díaz-Pérez et al., 2018, Sancho et al., 2022) and supported the allopolyploid origin of the analysed polyploid *Brachypodium* species, as was previously shown (Díaz-Pérez et al., 2018, Sancho et al., 2022).

Chromosome number has traditionally been considered a fundamental cytogenetic character, often used to delimit species and trace evolutionary histories. However, in plants, intraspecific polymorphisms in chromosome number are not rare (Leitch and Leitch, 2008), for example, in *Festuca* (Martínez-Sagarra et al., 2021), *Poa* (Spinnler and Stöcklin, 2018), and *Prospero* (Jang et al., 2013). In the genus *Brachypodium*, we observed significant intraspecific variation in chromosome number. Several *Brachypodium* species exhibit cytotypic diversity, where different chromosome numbers are found within a single species. For instance, *B. pinnatum* is known to exist in both diploid and polyploid forms, with reported chromosome counts of  $2n = 16$  and  $18$  in diploids and  $2n = 28$  in allopolyploid (Catalan et al., 2016, Díaz-Pérez et al., 2018, Lusinska et al., 2019). In this work, we analysed two cytotypes of *B. pinnatum* with  $2n = 16$  and  $2n = 28$ . Polyploidy is a prominent feature in the evolution of *Brachypodium*, where several species exhibit multiple ploidy levels. For instance, *B. retusum* and *B. rupestre* exist as allotetra- and allohexaploids. In this work, we analysed allohexaploid accessions of these species. Previously, *B. distachyon* was also considered to have diploid, autotetraploid and autohexaploid cytotypes ( $2n = 10, 20, 30$ ), which were later described as separate species: diploid *B. distachyon* ( $2n = 10$ ;  $x = 5$ ), diploid *B. stacei* ( $2n = 20$ ;  $x = 10$ ) and their derived allotetraploid *B. hybridum* ( $2n = 30$ ;  $2x = 5 + 10$ ) (Catalán et al., 2012, Hasterok et al., 2004). The distinction of these three species was made possible through phylogenetic and cytogenetic studies as well as detailed phenotypic analyses (Catalán et al., 2012). Also, the repeatome analyses presented in this work support this distinction.

### 5.2. Repetitive DNA sequences organisation in the allotetraploid *B. hybridum*

Previous studies based predominantly on single-copy sequences showed subgenomic stasis in *B. hybridum* (Gordon et al., 2020). However, repetitive DNA often undergoes reorganisation in allopolyploids through processes such as elimination of repeats, reactivation of transposable

elements through epigenetic repatterning that ultimately leads to diploidisation of the polyploid nuclei (Renny-Byfield et al., 2013, Parisod et al., 2012, Feldman and Levy, 2005). In this study, a comprehensive comparative analysis of repetitive DNA elements was conducted for two *B. hybridum* genotypes differing in evolutionary age: ABR113 and Bhyb26, estimated to have originated 0.14 and 1.4 Mya, respectively (Gordon et al., 2020). These genotypes were compared alongside their putative diploid progenitors, represented by the extant species *B. distachyon* and *B. stacei*. The aim was to assess whether hybridisation and subsequent polyploidisation led to the reorganisation of repetitive DNA sequences in the allotetraploid *B. hybridum*.

Previous studies of the repetitive DNA landscape of *B. hybridum* have suggested limited post-polyploidisation reorganisation of the repeatome, pointing instead to a relatively stable evolutionary trajectory characterised by subgenome stasis (Decena et al., 2024, Gordon et al., 2020, Mu et al., 2023). Other studies have provided evidence of TEs proliferation following whole-genome duplication (WGD), though it was primarily observed in the evolutionarily older lineage of *B. hybridum* (Bhyb26) (Scarlett et al., 2022). However, in this study, the use of comparative RepeatExplorer2 analysis combined with wet-lab techniques made it possible to directly compare and track specific repeat families across the genomes of different *Brachypodium* species. Additionally, repeatome the analysis of both lineages of *B. hybridum* alongside their putative diploid progenitors was presented here for the first time. This approach allows for a detailed comparison of abundance of individual repetitive DNA family/subfamily across accessions, providing novel insights into repeatome dynamics following allopolyploidisation in *B. hybridum*. It is also important to acknowledge a key limitation in this comparative approach: while *B. distachyon* and *B. stacei* serve as extant representatives of the ancestral subgenomes of *B. hybridum*. They may differ from the true progenitor species involved in the original allopolyploidisation events since the diploids also underwent evolution and the allotetraploid has shown to have recurrent origin (Gordon et al., 2020, Mu et al., 2023).

The concept of “genomic shock” predicts widespread transposable element (TE) mobilisation following allopolyploid formation (de Tomás and Vicent, 2023, McClintock, 1950, Otto, 2007). In several species, extensive TE reactivation has indeed been documented to follow hybridisation and/or WGD, e.g., in *Helianthus* and *Capsella bursa-pastoris* (Ågren et al., 2016, Ungerer et al., 2009). However, in the majority of the studied allopolyploids, post-hybridisation TE activity is limited to one or a few specific families of repeats (Danilova et al., 2017, Shams and Raskina, 2018, Eriksson et al., 2022). Our results indicated that notable alterations in the repeatome of the allotetraploid *B. hybridum* have occurred within specific repeat families when compared to putative diploid progenitors. However, these changes appear

to be accumulated in a gradual manner, since the repeatome rearrangements were more striking in the evolutionary older lineage Bhyb26. Thus, the rapid and extensive genomic reorganisation typically associated with the “genome shock” hypothesis seems not to be true in *B. hybridum*.

In both evolutionary lineages of *B. hybridum*, the patterns of abundance of many dispersed repeat families closely resembled an additive inheritance model, reflecting contributions from both ancestral subgenomes. Nevertheless, numerous deviations from this pattern were observed among tandem repeats as well as some dispersed repeats. Notably, the organisation and abundance of rDNA (e.g., new 5S rDNA NTS family in Bhyb26), satellite DNA (e.g., centromeric satellite), and certain lineages of TEs (e.g., Tat/Retand and Angela families) exhibited a non-additive pattern, suggesting repeat-specific reorganisation events such as reduction/amplification or rise of new repeat families. For example, extensive proliferation of one LINEs family was observed only in *B. hybridum* Bhyb26. As for LTR-retrotransposons several main trends were observed even within one family. For instance, within Ty1/*copia* Angela family, two different patterns of amplification were observed: (i) the abundance was higher in older allotetraploid Bhyb26, comparing to both *in silico* allotetraploid and younger ABR113; (ii) abundance was higher in younger ABR113, comparing to both *in silico* allotetraploid and older allotetraploid Bhyb26. Such variation suggests that TE dynamics are both genotype- and TE-lineage-specific, as previously was described across different species, e.g., *Hydrangea*, *Amomum* and *Dactylorhiza* (Hlavatá et al., 2024, Ishiguro et al., 2025, Eriksson et al., 2022). For instance, in wild wheats, it was demonstrated that multiple LTR-retrotransposon families exhibit contrasting evolutionary trajectories following independent allopolyploidisation events (Senerchia et al., 2014).

Another example of post-polyploidisation TE activity in *B. hybridum* is the conquest of the *B. stacei*-like subgenome by the Ty3/*gypsy* Tat/Retand family from *B. distachyon*-like subgenome. This retrotransposon, originally more abundant in *B. distachyon*, likely has expanded from *B. distachyon*-like subgenome into the *B. stacei*-like subgenome in *B. hybridum*, suggesting intergenomic retrotransposition. Such asymmetric TE dynamics following polyploidisation have been documented in other plant systems, e.g., in *Gossypium* (cotton), in which a post-polyploid activation and movement of retrotransposons from maternal A-genome to paternal D-genome resulted in significant genomic restructuring (Zhao et al., 1998, Fang et al., 2017).

Tandem repeats often exhibit relatively fast rates of turnover in allopolyploids (McCann et al., 2018, Heitkam et al., 2020, Koukalova et al., 2010, Garrido-Ramos, 2017). In *B. hybridum*, the most prominent changes were observed in the context of tandem repeats. Notably, within centromeric satellite DNA, two distinct subfamilies showed distinct evolutionary



trajectories. In the evolutionarily older lineage Bhyb26, the CentBd subfamily underwent marked amplification, when compared to *in silico* allotetraploid, whereas in the younger lineage ABR113, CentBd was reduced when compared to *in silico* allotetraploid. Such differences can be explained either by the recurrent origin of *B. hybridum* and lineage-specific evolution of the centromeric repeats or by changes in the trajectory of repeat amplification/reduction over the evolutionary time. On the other hand, the abundance of the CentBs subfamily was not changed when compared to the *in silico* allotetraploid in neither of the studied genotypes. Both *B. hybridum* genotypes not only differ in their estimated time of origin but also in the direction of hybridisation between the diploid progenitors, as reflected by their distinct plastome origin (D- and S-plastomes; Gordon et al., 2020). This divergence raises the possibility that the direction of the cross may influence the evolutionary dynamics of certain repeat families, including centromeric satDNA. It was hypothesised that paternal genome might undergo more prominent changes than the maternal one (Song et al., 1995). However, according to many studies on different plant species, this hypothesis was not supported (Weiss-Schneeweiss et al., 2012, Wendel et al., 1995).

It was shown in many plant systems that certain repetitive families may undergo reduction or even elimination, while others may arise, following the ‘birth and death’ hypothesis (Ruiz-Ruano et al., 2016, Lower et al., 2018, Nei and Rooney, 2005). The elimination of satDNA families from allopolyploids as well as rise of new families are well-documented across different genera, e.g., *Nicotiana* (Koukalova et al., 2010, Lim et al., 2007) and *Beta* (Schmidt et al., 2024). This also corroborates the evolutionary trajectories of some satellite families in *B. hybridum*. Exemplary, satDNA family of approximately 1800 bp of monomer length, displayed pronounced evolutionary changes in *B. hybridum*. According to RepeatExplorer2 analysis, this satDNA was eliminated from both *B. hybridum* genotypes; however, Southern blot showed low abundance of similar repeat in Bhyb26. Additionally, each *B. hybridum* genotype showed presence of genotype-specific minisatellites that were not present in either of the progenitors, suggesting *de novo* amplification of these families in allotetraploids, similarly to the rise of new 5S rDNA NTS variant in Bhyb26 (Trunova et al., 2024).

The most significant reorganisation in *B. hybridum* was observed for rDNA. Despite the fact that in the evolutionary recent genotype ABR113, both parental variants of 35S rDNA were maintained, a notable reduction of the *B. stacei*-inherited 35S rDNA units was previously shown using *in silico* analysis, FISH and Southern hybridisation (Borowska-Zuchowska et al., 2020, Borowska-Zuchowska et al., 2021). Moreover, it was shown that *B. stacei*-like 35S rDNA locus was preferentially silenced *via* nucleolar dominance phenomenon (Borowska-Zuchowska et al., 2020). Interestingly, the *B. stacei*-inherited 35S rDNA locus was eliminated in evolutionary

older Bhyb26 genotype as was shown in this work (Trunova et al., 2024). Thus, a gradual elimination of the silenced S-subgenome 35S rDNA took place in *B. hybridum*. So far, in all studied *B. hybridum* accessions, a decrease in the S-subgenome 35S rDNA units was observed (Borowska-Zuchowska et al., 2020).

In *B. hybridum*, the reduction/elimination of 35S rDNA from the S-subgenome has been observed regardless of the maternal or paternal origin of the *B. stacei*-inherited subgenome (Trunova et al., 2024). Such uniparental reduction/elimination of 35S rDNA has been widely reported as a hallmark of diploidisation in allopolyploid species (Wendel et al., 1995, Guo and Han, 2014, Guggisberg et al., 2008, Kovarik et al., 2004a, Sochorová et al., 2017). Complete locus loss has also been documented in genera like *Chenopodium* (Kolano et al., 2016), *Paspalum* (Vaio et al., 2019), and *Melampodium* (Weiss-Schneeweiss et al., 2012).

Most polyploids studied so far exhibit additive inheritance of parental 5S rDNA variants (Weiss-Schneeweiss et al., 2012, Tynkevich et al., 2022, Kolano et al., 2016, Kolano et al., 2019). Indeed, in evolutionary younger *B. hybridum* ABR113, 5S rDNA loci inherited from both ancestors were present, as was shown by *in silico* analysis, FISH and Southern hybridisation. However, in the older one, the elimination of *B. distachyon*-like 5S rDNA locus was reported (Trunova et al., 2024). Two pathways of 5S rDNA diploidisation have been described: (i) locus loss, as in *Melampodium strigosum* (Weiss-Schneeweiss et al., 2012), and (ii) variant replacement without locus loss, as in *Anemone baldensis* (Mlinarec et al., 2012, Mlinarec et al., 2016). However, such an interlocus concerted evolution of 5S rDNA loci in allopolyploids seems rare since it was described only in *Anemone baldensis*. The situation observed in *B. hybridum* Bhyb26 is more complex, since except the uniparental 5S rDNA locus elimination, the rise of a new 5S rDNA NTS was shown (Trunova et al., 2024). Most of the studies showed that homogenisation of 5S rDNA seems to occur within a single array with no (or negligible) exchange between the loci (Mahelka et al., 2013, Cronn et al., 1996, Kellogg and Appels, 1995). Therefore, *B. hybridum* Bhyb26 appear as a valuable model of rDNA sequence evolution.

In summary, rRNA genes in *B. hybridum* exhibit a strong trend toward diploidisation, with distinct patterns of rDNA loss differentiating younger and older lineages. The 35S and 5S rDNA loci behave differently across subgenomes, with the older lineage showing preferential elimination of 35S rDNA from the S-subgenome and 5S rDNA from the D-subgenome. These findings indicate that in Bhyb26, 5S and 35S rRNA genes are likely expressed from different subgenomes, underscoring the intricate regulatory mechanisms that shape rRNA gene expression in polyploid genomes (Trunova et al., 2024).

### 5.3. Trajectories of repetitive DNA evolution in *Brachypodium* genus

Our results highlight the significant role of the repeatome in shaping genome size variation across the analysed *Brachypodium* species, in which repetitive elements comprise a major fraction of the genome. That is well-established in plant genome biology to attribute genome size variation to the differential accumulation and loss of repetitive DNA (Wang et al., 2021, Borowska-Zuchowska et al., 2022, Novák et al., 2020a). Both genome size and repetitive DNA content were lower in annual *Brachypodium* species, compared to perennials. In a comprehensive study of species and subspecies within the *Asteraceae* family, it was found that annual species have significantly smaller genomes than perennials (Garcia et al., 2013). This supports the hypothesis that smaller genomes are favoured in annual plants due to selective pressures for shorter life cycles and rapid development (Garcia et al., 2013). The genome size diversity among analysed *Brachypodium* species ranged from the smallest one in annual diploid *B. stacei* (0.28 pg/1C DNA) to the largest one in perennial allotetraploid *B. mexicanum* (1.9 pg/1C DNA). This diversity was related not only to polyploidy but primarily due to the high content of repetitive DNA in *B. mexicanum*. Such striking genome size increase in *B. mexicanum* was mainly caused by the proliferation of Ty3/gypsy LTR-retrotransposons, what is commonly found in plants, for instance in the *Hesperis* clade (Brassicaceae) and the tribe Fabeae (Fabaceae) (Hloušková et al., 2019, Macas et al., 2015). In all analysed *Brachypodium* species Ty3/gypsy LTR-retrotransposons were the major contributors to repetitive DNA content. The comparative analysis using the RepeatExplorer2 pipeline revealed that all analysed *Brachypodium* accessions share the same major families of LTR-retrotransposons. This finding suggests that a conserved set of repetitive elements forms the core of the repeatome within the genus, likely inherited from a common ancestor. Such conservation of dominant retrotransposon families is common for closely related species from such genera like *Allium* and *Prunus* (Evans et al., 1983, Wang et al., 2022). Among the studied *Brachypodium* species, only *B. mexicanum* showed a strikingly different abundance of retrotransposons, which may be explained by both the geographical isolation of this species and its early split from other species that belong to the genus (Díaz-Pérez et al., 2018)

Besides the transposable elements, the tandemly repetitive sequences that include rDNA sequences and satDNAs were also analysed in *Brachypodium* species. The cluster graph analysis of the 5S rDNA variants in RepeatExplorer2 was proved to be a useful tool in identifying hybrid or allopolyploid plant species based on the structural characteristics of their 5S rDNA repeats (Garcia et al., 2020). Therefore, such an analysis was also performed in both diploid and polyploid *Brachypodium* species. As expected, the diploid *Brachypodium* species were characterised by simple circular cluster graphs reflecting the presence of one type of 5S

rDNA (*B. stacei*,  $x = 10$ ; *B. distachyon*,  $x = 5$ ; *B. arbuscula*, *B. glaucovirens*, *B. sylvaticum*,  $x = 9$ ; and *B. pinnatum*,  $x = 8$ ). Two allotetraploids: *B. phoenicoides* and *B. mexicanum* possess two parental variants of 5S rDNA, while other allopolyploids showed elimination of one or even two parental variants. Exemplary, simple circular graphs were obtained for *B. pinnatum* ( $2x = 5 + 9$ ) and *B. boissieri* ( $3x = 8+8+8$ ), suggesting that either their diploid progenitors had very similar 5S rDNA NTS variants or homogenisation occurred towards one parental variant during diploidisation, or new 5S rDNA NTS type arose, as it was shown in *B. hybridum* Bhyb26 (Trunova et al., 2024, Weiss-Schneeweiss et al., 2012). However, FISH analysis with rDNA did not reveal any evidence of locus loss in the studied allopolyploids (Wolny and Hasterok, 2009). In allotetraploid *B. pinnatum*, two loci of 5S rDNA and two loci of 35S rDNA were detected, and in the allohexaploid *B. boissieri*, three loci of 5S rDNA and three loci of 35S rDNA were observed (Wolny and Hasterok, 2009).

The patterns of satDNA evolution observed in the *Brachypodium* genus can be effectively interpreted in the context of the library hypothesis, which posits that related species share a common set, i.e. “library” of satellite sequences inherited from a common ancestor, with differences among species resulting primarily from differential amplification, reduction, or elimination of particular satellite families (Mestrović et al., 1998, Melters et al., 2013, Contento et al., 2005). The comparative analysis revealed that most of *Brachypodium* species share several major satellite DNA families, suggesting a common ancestral library. However, the abundance and chromosomal distribution of these satDNAs varied significantly across species.

Previously, it was reported that two subfamilies (CentBd and CentBs) of centromeric satellite were present in annual *Brachypodium* species (Chen et al., 2024). The CentBd subfamily was found in diploid *B. distachyon*, while CentBs was characteristic of diploid *B. stacei*. Their derived allotetraploid *B. hybridum* inherited both subfamilies of centromeric satellite (Chen et al., 2024). RepeatExplorer2 analysis revealed the presence of CentBd subfamily in all *Brachypodium* perennials with the exception of *B. mexicanum*. However, traces of CentBd subfamily were revealed via cloning experiment in *B. mexicanum* as well, suggesting its presence within the genome in low copy number. Analysis of centromeric satDNA subfamilies in phylogenetic context in *Brachypodium* showed that amplification of CentBd subfamily most likely occurred in the common ancestor of annuals and perennials. Subfamily CentBs was present in diploid *B. stacei* and allotetraploid *B. mexicanum* (Figure 31), suggesting its presence in a common ancestor of *Brachypodium* species and an elimination of CentBs subfamily from other perennial species. The third subfamily CentBp arose in a common ancestor of perennial *Brachypodium* species. In our analysis, new centromeric satellite subfamily CentBp, characteristic only for perennial species, was identified (Figure 30).

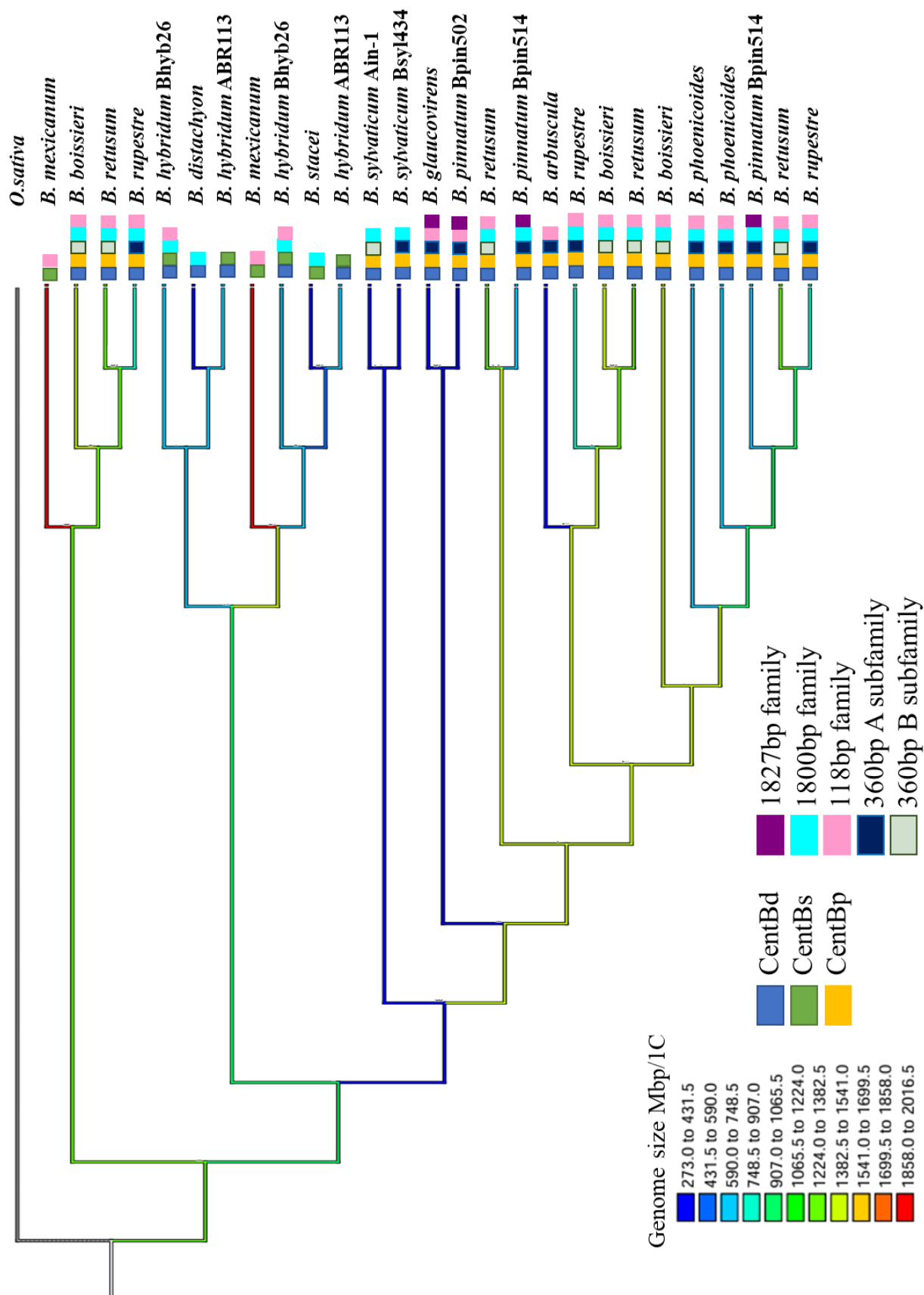
Remarkably, all analysed perennial species (diploids and polyploids), except *B. mexicanum*, showed the presence of two subfamilies (CentBd and CentBp). The amplification of CentBp subfamily most likely occurred after split of perennial and annual species (Figure 30). The observation of more than one centromeric satellite family or subfamily within a diploid species was earlier described in some *Arabidopsis* species (Kawabe and Nasuda, 2005). These findings align with the library hypothesis, indicating that centromeric satellite family evolution may involve the reduction and amplification of elements from a common ancestral set of repeats.

# Figure 31

## Figure 31

### **The evolution of analysed satDNAs in genomes of *Brachypodium* species.**

The phylogenetic tree is based on analysis of *Gigantea* marker. Genome sizes were mapped on the ML tree of *Gigantea* sequences using the maximum parsimony method implemented in Mesquite. The colour of the branches represents the genome size. The coloured rectangles indicate the presence of satDNA family in a particular accession.





Another highly amplified satDNA family, 360 bp, was identified in all perennial species, except *B. mexicanum*. Two distinct subfamilies within the 360 bp satDNA family were found. However, regardless of ploidy level, each species harboured only one cluster with this satellite. Interestingly, two distinct subfamilies A and B of the 360 bp satDNA family were found in two genotypes, Bsyl434 and Ain-1, of *B. sylvaticum*, respectively (Figure 30). This might be related to their geographical separation. Genotype Ain-1 was collected in Ain Draham, Tunisia, while Bsyl434 in Crimea, Ukraine. This observation aligns with patterns observed in other plant species, in which satDNA diversification correlates with geographic and evolutionary divergence. For instance, in the genus *Deschampsia* (Poaceae), several satDNA families were identified. They exhibited interspecific variation in both abundance and chromosomal distribution, which appears to be related to geographical distances (González et al., 2021). Another study revealed population specificity of satellite family pAge2 in *Arabidopsis halleri* and *Arabidopsis lyrata* (Kawabe and Charlesworth, 2007). This variation was attributed to differential amplification and elimination of satellite DNA. Interestingly, in allohexaploid species *B. boissieri* and *B. retusum* 360B subfamily was identified similarly to Ain-1 genotype of *B. sylvaticum* (Figure 31). Presence of lineage-specific variants that are restricted to individual species or genotype, supports the hypothesis that tandem repeats may evolve as described by a birth-and-death hypothesis, according which differential amplification, sequence divergence, and eventual elimination of tandem repeat families lead to genome-specific profiles of repeats (Pinhal et al., 2011, Rooney and Ward, 2005). Such evolutionary dynamics was shown, for instance, in Cardueae (Carduoideae Asteraceae), in which satDNA families underwent differential amplification/elimination and divergence, leading to the rise of species-specific variants (del Bosque et al., 2014). The rise of new satDNA variants and the simultaneous elimination of others are hallmarks of the evolutionary pathway according to “birth and death” hypothesis (Pinhal et al., 2011, Rooney and Ward, 2005).

The evolutionary dynamics of satellite DNA are often shaped by the process of concerted evolution, which promotes sequence homogenisation within genomes and can lead to species- or lineage-specific repeat profiles (Wang et al., 2023, Lunerová et al., 2017). The homogenisation is thought to occur primarily through recombination, which tend to homogenise repeat sequences within species, while allowing divergence between species (Wang et al., 2023, Lunerová et al., 2017). Further speciation, differential amplification and homogenisation in distinct lineages lead to rise of distinct subfamilies of 360 bp satDNA (Figure 31).

The analysis of another satDNA family (118 bp) in a phylogenetic context, considering only diploid species, showed that the amplification of this family likely occurred after the

divergence of the *B. arbuscula*–*B. glaucovirens*–*B. pinnatum* group from *B. sylvaticum*, as it was not detected in the annual diploids or in *B. sylvaticum* (Figure 30). However, the evolution of this satDNA family is more complex. It was also found in *B. mexicanum*, therefore 118 bp family was most likely inherited from common ancestor before the split of *B. mexicanum* and European *Brachypodium* species and eliminated during the speciation of some lineages (*B. sylvaticum*, annuals, allotetraploid *B. pinnatum*; Figure 31). Such complicated evolutionary dynamics was previously observed in *Beta* genus, where distinct satDNA families within the genus, revealing differential amplification and distribution patterns (Schmidt et al., 1991, Dechyeva et al., 2003, Dechyeva and Schmidt, 2006). Since 118 bp family was not detected in annual diploid species as well as in recent lineage of *B. hybridum* ABR113, but was found only in the evolutionary older Bhyb26, two hypotheses can be proposed:

- (i) 118 bp satDNA family was present in the common ancestor of the annual clade and was present in the exact diploid ancestors of an ancient lineage of *B. hybridum* Bhyb26. Then it was eliminated during the evolution of diploid *B. distachyon* and *B. stacei*. For that reason, this family is absent in the recent lineage of allotetraploid *B. hybridum* ABR113.
- (ii) The abundance of 118 bp family was significantly reduced in the ancestor of *Brachypodium* annuals, and later reamplified as a result of post-polyploid evolution of Bhyb26.

The 1800 bp family was found in majority of analysed species. The analysis of this family in phylogenetic context suggests that most likely this family was amplified in a common ancestor of *Brachypodium* species and then was eliminated in *B. arbuscula*–*B. glaucovirens*–*B. pinnatum* subclade, based on the analysis of diploid species (Figure 30). This family was absent also in *B. mexicanum*, which suggests that either amplification of 1800 bp family occurred after the split of *B. mexicanum* from other *Brachypodium* species or was eliminated during *B. mexicanum* speciation (Figure 31). Interestingly, this repeat was eliminated from the genome of recent lineage ABR113, despite its presence in the putative diploid progenitors and evolutionary older lineage of *B. hybridum* (Bhyb26).

The 1827 bp satDNA family was detected in allotetraploid *B. pinnatum* ( $2n = 4x = 28$ ,  $2x = 5 + 9$ ) and in two diploids: *B. pinnatum* ( $x = 8$ ) and *B. glaucovirens* ( $x = 9$ ), which may suggest that either of these two diploids could serve as an ancestor of allotetraploid *B. pinnatum*. Previous cytogenetic studies that employed GISH also pointed to diploid *B. pinnatum* as a possible progenitor of the corresponding subgenome in allotetraploid *B. pinnatum* (Wolny and Hasterok, 2009).

One of the objectives of this study was to test the origins of selected perennial *Brachypodium* polyploid species through comparative analysis of their repeatome composition. While repetitive DNA has proven to be a valuable genomic marker for distinguishing parental contributions and inferring evolutionary relationships, e.g., in *Nicotiana* (Dodsworth et al., 2014) and Fabaceae (Vitales et al., 2020). However, our analyses revealed a high degree of repeatome similarity among *Brachypodium* species, several changes in their repeatome organisation, which seems to accompany the speciation of the *Brachypodium* species. This finding complicates efforts to trace polyploid origins solely based on the repetitive DNA. The evolutionary dynamics of repetitive DNA in allopolyploid genomes present a complex and often ambiguous picture. One of the primary challenges lies in distinguishing whether particular repeat families were inherited from diploid progenitors or have arisen through amplification during the post-polyploid phase of genome evolution. This difficulty stems from the dynamic nature of repetitive sequences, which are subject to rapid turnover, expansion, and contraction over evolutionary timescales (Garrido-Ramos, 2017). Interpretation is further complicated by the often-uncertain phylogenetic origins of many allopolyploid species. In several cases, the precise identity of the diploid progenitors remains unresolved, and in some instances, evidence suggests that these progenitor species may now be extinct (Catalan et al., 2016, Díaz-Pérez et al., 2018, Sancho et al., 2022). Such historical gaps hinder our ability to trace the origin and trajectory of specific repeat families across lineages. These challenges underscore the importance of integrating comparative repeatome analyses with robust phylogenetic frameworks to better understand the evolutionary fate of the repetitive DNA in polyploid genomes.

## 6. CONCLUSIONS

1. Annual diploid species (*B. distachyon*, *B. stacei*) with the smallest genome sizes, display lowest proportions of repetitive DNA among analysed species.
2. *B. hybridum*, an allotetraploid derived from *B. distachyon* and *B. stacei*, showed an overall additive pattern of dispersed repeats, whereas tandem repeats exhibited substantial deviation from the expected additive values observed in its putative diploid ancestors.
3. Both analysed *B. hybridum* accessions, which differ significantly in their time of origin, exhibited divergent landscapes of repetitive DNA sequences, reflecting the gradual evolutionary changes that shaped their genomes.
4. Uniparental loss of rDNA loci occurred in *B. hybridum* Bhyb26 (with an estimated time of origin of 1.4 Mya). While the 35S rDNA locus was eliminated from the S-subgenome, the 5S rDNA locus was lost from the D-subgenome. This phenomenon resulted in the formation of chimeric ribosomes with rRNA molecules derived from distinct parental origins.
5. While an additive pattern of 5S rDNA loci was observed in the evolutionarily recent *B. hybridum*, the evolutionarily older lineage exhibited not only the loss of one 5S rDNA locus but also the emergence of a new 5S rDNA NTS family - a phenomenon that has been very rarely described to date.
6. The landscape of repetitive sequences in the *Brachypodium* genus, along with the reconstructed trajectories of repeat reorganisation, supports the ‘library hypothesis’ as a model of satDNA evolution in this genus.
7. Taking into account the presence of two or more lineage-specific subfamilies of satDNA, their evolution according to the ‘birth-and-death’ hypothesis in the *Brachypodium* genus cannot be ruled out.
8. Despite the global genomic stasis of unique sequences reported in annual *B. hybridum*, the repetitive genome fraction, especially tandem repeats, appears to be a much more dynamic part of the genome.

## 7. SUMMARY

Repetitive DNA constitutes a major component of plant genomes, playing crucial roles in genome structure, function, and evolution. Despite its significance, the repetitive fraction of the genome has remained relatively understudied in the model grass genus *Brachypodium*. This genus, with its diverse range of diploid and polyploid species and well-characterised evolutionary relationships, offers an excellent system for investigating the dynamics of repetitive DNA evolution.

To address this gap, a comprehensive comparative analysis of the *Brachypodium* repeatome was conducted, integrating both bioinformatic and cytogenetic approaches. Next generation sequencing (Illumina) data were analysed using RepeatExplorer2&TAREAN – repeat identification and quantification tools. Fluorescence *in situ* hybridisation (FISH) and Southern Blot hybridisation were applied to study chromosomal and genomic organisation of selected repetitive elements.

The study focused on characterising the structure, diversity, and evolutionary dynamics of various classes of repetitive DNA, including transposable elements and tandem repeats. The composition and distribution of major repetitive DNA elements were analysed in two accessions of the natural allotetraploid *Brachypodium hybridum*, which originated at different times (0.14 and 1.4 Mya), and in their putative diploid progenitors (*B. distachyon* and *B. stacei*). *B. hybridum* exhibited an overall additive pattern of dispersed repeats. The most prominent changes were observed in the context of tandem repeats. Both analysed *B. hybridum* accessions exhibited divergent landscapes of repetitive DNA sequences, reflecting the gradual evolutionary changes that have shaped their genomes.

Comparative analyses of repetitive sequences in annual and perennial *Brachypodium* species revealed both lineage-specific elements and repeats shared across most of the analysed species. Contrasting evolutionary dynamics were observed among different repeat types. Several conserved satellite DNA families showed variable amplification among related species. Phylogenetic assessment of satellite DNAs across *Brachypodium* species provided insights into the evolutionary dynamics of tandem repeats. LTR-retrotransposon profiles varied among taxa, indicating lineage-specific patterns of proliferation and accumulation. These findings underscore the utility of *Brachypodium* as a model system and provide new insights into the contribution of repetitive elements to genome diversification, polyploid evolution, and structural genome organization in grasses.

## 8. STRESZCZENIE

Powtarzalne sekwencje DNA, stanowiące istotny składnik genomów roślin, odgrywają kluczową rolę w strukturze, funkcjonowaniu i ewolucji genomu. Frakcja powtarzalnego DNA w genomach modelowego rodzaju traw *Brachypodium* pozostaje jednak stosunkowo słabo poznana. Rodzaj ten, dobrze scharakteryzowany pod względem zależności filogenetycznych i obejmujący zarówno gatunki diploidalne, jak i poliploidalne, stanowi doskonały model do badań nad ewolucyjną dynamiką powtarzalnych sekwencji DNA.

W ramach niniejszej pracy przeprowadzono dogłębną analizę porównawczą repeatomów przedstawicieli rodzaju *Brachypodium*, wykorzystującą zarówno metody bioinformatyczne, jak i cytogenetyczne. Dane uzyskane za pomocą sekwencjonowania nowej generacji (Illumina) zostały przeanalizowane z wykorzystaniem narzędzi RepeatExplorer2 i TAREAN, służących do identyfikacji i ilościowej analizy sekwencji powtarzalnych. Przy pomocy fluorescencyjnej hybrydyzacji *in situ* (FISH) oraz hybrydyzacji metodą Southern badano organizację chromosomową i genomową wybranych sekwencji powtarzalnych.

Przeprowadzono charakterystykę struktury, zróżnicowania oraz dynamiki ewolucyjnej różnych typów sekwencji powtarzalnych, w tym transpozonów i powtórzeń tandemowych. Wykonano porównawczą analizę sekwencji powtarzalnych dla dwóch genotypów naturalnego allotetraploida *B. hybridum*, różniących się czasem powstania (0,14 oraz 1,4 mln lat temu) oraz jego hipotetycznych diploidalnych gatunków ancestralnych (*B. distachyon* i *B. stacei*). Amplifikacja rozproszonych sekwencji powtarzalnych w genomach badanych genotypów *B. hybridum* najczęściej wykazywała addytywny charakter względem liczby kopii tych sekwencji w genomach diploidalnych przodków. Z kolei sekwencje tandemowo powtarzalne charakteryzowały się większą dynamiką zmian, wykazując - w zależności od rodziny sekwencji - zarówno amplifikację, jak i redukcję liczby powtórzeń w porównaniu do hipotetycznych gatunków ancestralnych. Każdy z analizowanych genotypów *B. hybridum* charakteryzował się specyficznym wzorem amplifikacji i/lub organizacji sekwencji powtarzalnych, wskazujące na to, że genomy *B. hybridum* kształtowane są przez stopniowe zmiany ewolucyjne.

Analizy porównawcze sekwencji powtarzalnych w jednorocznych i wieloletnich gatunkach rodzaju *Brachypodium*, ujawniły obecność zarówno elementów specyficznych dla poszczególnych linii ewolucyjnych, jak i rodzin sekwencji powtarzalnych wspólnych dla większości analizowanych gatunków. Zaobserwowano zróżnicowaną dynamikę zmian ewolucyjnych w zależności od typu powtórzenia. Rodziny sekwencji satelitarnych wykazywały różny poziom amplifikacji, nawet wśród blisko spokrewnionych gatunków. Również profile

amplifikacji retrotranspozonów znacząco różniły się pomiędzy taksonami, wskazując na liniowo-specyficzne wzorce proliferacji i akumulacji. Uzyskane wyniki potwierdzają, że *Brachypodium* stanowi cenny system modelowy do badań nad ewolucją sekwencji powtarzalnych oraz ich rolą w różnicowaniu się genomów i ewolucji poliploidów.

## 9. REFERENCES

- Ågren JA, Greiner S, Johnson MTJ, Wright SI. 2015. No evidence that sex and transposable elements drive genome size variation in evening primroses. *Evolution*, 69: 1053-1062.
- Ågren JA, Huang H-R, Wright SI. 2016. Transposable element evolution in the allotetraploid *Capsella bursa-pastoris*. *American Journal of Botany*, 103: 1197-1202.
- Aguilar M, Prieto P. 2021. Telomeres and Subtelomeres Dynamics in the Context of Early Chromosome Interactions During Meiosis and Their Implications in Plant Breeding. *Frontiers in Plant Science*, 12.
- Alger EI, Edger PP. 2020. One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Curr Opin Plant Biol*, 54: 108-113.
- Alisawi O, Richert-Pöggeler KR, Heslop-Harrison JS, Schwarzacher T. 2023. The nature and organization of satellite DNAs in *Petunia hybrida*, related, and ancestral genomes. *Frontiers in Plant Science*, 14.
- Almojil D, Bourgeois Y, Falis M, Hariyani I, Wilcox J, Boissinot S. 2021. The Structural, Functional and Evolutionary Impact of Transposable Elements in Eukaryotes. *Genes (Basel)*, 12.
- Álvarez I, Wendel JF. 2003. Ribosomal ITS sequences and plant phylogenetic inference. *Molecular phylogenetics and evolution*, 29: 417-434.
- Areshchenkova T, Ganai MW. 1999. Long tomato microsatellites are predominantly associated with centromeric regions. *Genome*, 42: 536-44.
- Ataei N, Schneeweiss GM, García MA, Krug M, Lehnert M, Valizadeh J, Quandt D. 2020. A multilocus phylogeny of the non-photosynthetic parasitic plant *Cistanche* (Orobanchaceae) refutes current taxonomy and identifies four major morphologically distinct clades. *Molecular phylogenetics and evolution*, 151: 106898.
- Bao Y, Wendel JF, Ge S. 2010. Multiple patterns of rDNA evolution following polyploidy in *Oryza*. *Molecular Phylogenetics and Evolution*, 55: 136-142.
- Belyayev A, Jandová M, Josefiová J, Kalendar R, Mahelka V, Mandák B, Krak K. 2020. The major satellite DNA families of the diploid *Chenopodium album* aggregate species: Arguments for and against the "library hypothesis". *PLoS One*, 15: e0241206.
- Bennetzen JL, Kellogg EA. 1997. Do Plants Have a One-Way Ticket to Genomic Obesity? *The Plant Cell*, 9: 1509-1514.



- Bennetzen JL, Schmutz J, Wang H, Percifield R, Hawkins J, Pontaroli AC, Estep M, Feng L, Vaughn JN, Grimwood J. 2012. Reference genome sequence of the model plant *Setaria*. *Nature biotechnology*, 30: 555-561.
- Bennetzen JL, Wang H. 2014. The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. *Annual Review of Plant Biology*, 65: 505-530.
- Benson G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic acids research*, 27: 573-580.
- Betekhtin A, Jenkins G, Hasterok R. 2014. Reconstructing the Evolution of *Brachypodium* Genomes Using Comparative Chromosome Painting. *PLOS ONE*, 9: e115108.
- Bird KA, Niederhuth CE, Ou S, Gehan M, Pires JC, Xiong Z, VanBuren R, Edger PP. 2021. Replaying the evolutionary tape to investigate subgenome dominance in allopolyploid *Brassica napus*. *New Phytologist*, 230: 354-371.
- Biscotti MA, Olmo E, Heslop-Harrison JS. 2015. Repetitive DNA in eukaryotic genomes. *Chromosome Research*, 23: 415-420.
- Borowska-Zuchowska N, Kovarik A, Robaszkiewicz E, Tuna M, Tuna GS, Gordon S, Vogel JP, Hasterok R. 2020. The fate of 35S rRNA genes in the allotetraploid grass *Brachypodium hybridum*. *The Plant Journal*, 103: 1810-1825.
- Borowska-Zuchowska N, Kwasniewski M, Hasterok R. 2016. Cytomolecular Analysis of Ribosomal DNA Evolution in a Natural Allotetraploid *Brachypodium hybridum* and Its Putative Ancestors—Dissecting Complex Repetitive Structure of Intergenic Spacers. *Frontiers in Plant Science*, Volume 7 - 2016.
- Borowska-Zuchowska N, Mykhailik S, Robaszkiewicz E, Matysiak N, Mielańczyk Ł, Wojnicz R, Kovarik A, Hasterok R. 2023. Switch them off or not: selective rRNA gene repression in grasses. *Trends in Plant Science*, 28.
- Borowska-Zuchowska N, Robaszkiewicz E, Mykhailik S, Wartini J, Pinski A, Kovarik A, Hasterok R. 2021. To Be or Not to Be Expressed: The First Evidence of a Nucleolar Dominance Tissue-Specificity in *Brachypodium hybridum*. *Frontiers in Plant Science*, Volume 12 - 2021.
- Borowska-Zuchowska N, Senderowicz M, Trunova D, Kolano B. 2022. Tracing the Evolution of the Angiosperm Genome from the Cytogenetic Point of View. *Plants*, 11: 784.
- Bourque G, Burns KH, Gehring M, Gorbunova V, Seluanov A, Hammell M, Imbeault M, Izsvák Z, Levin HL, Macfarlan TS, Mager DL, Feschotte C. 2018. Ten things you should know about transposable elements. *Genome Biology*, 19: 199.

- Brandes A, Thompson H, Dean C, Heslop-Harrison JS. 1997. Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis thaliana* L. *Chromosome Research*, 5: 238-246.
- Bureau TE, Wessler SR. 1992. Tourist: a large family of small inverted repeat elements frequently associated with maize genes. *The Plant Cell*, 4: 1283-1294.
- Catalán P, Kellogg EA, Olmstead RG. 1997. Phylogeny of Poaceae Subfamily Pooideae Based on ChloroplastndhF Gene Sequences. *Molecular Phylogenetics and Evolution*, 8: 150-166.
- Catalán P, López-Álvarez D, Bellosta C, Villar L. 2016. Updated taxonomic descriptions, iconography, and habitat preferences of *Brachypodium distachyon*, *B. stacei*, and *B. hybridum* (Poaceae). *Anales del Jardín Botánico de Madrid*, 73: e028.
- Catalan P, López-Álvarez D, Díaz-Pérez A, Sancho R, López-Herránz ML. 2016. Phylogeny and Evolution of the Genus *Brachypodium*. In: Vogel JP, ed. *Genetics and Genomics of Brachypodium*. Cham: Springer International Publishing.
- Catalán P, Müller J, Hasterok R, Jenkins G, Mur LA, Langdon T, Betekhtin A, Siwinska D, Pimentel M, López-Alvarez D. 2012. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Annals of Botany*, 109: 385-405.
- Catalán P, Shi Y, Armstrong L, Draper J, Stace CA. 1995. Molecular phylogeny of the grass genus *Brachypodium* P. Beauv. based on RFLP and RAPD analysis. *Botanical Journal of the Linnean Society*, 117: 263-280.
- Chang K-D, Fang S-A, Chang F-C, Chung M-C. 2010. Chromosomal conservation and sequence diversity of ribosomal RNA genes of two distant *Oryza* species. *Genomics*, 96: 181-190.
- Chen C, Wu S, Sun Y, Zhou J, Chen Y, Zhang J, Birchler JA, Han F, Yang N, Su H. 2024. Three near-complete genome assemblies reveal substantial centromere dynamics from diploid to tetraploid in *Brachypodium* genus. *Genome Biology*, 25: 63.
- Chen H, Al-Shehbaz IA, Yue J, Sun H. 2018. New insights into the taxonomy of tribe Euclidieae (Brassicaceae), evidence from nrITS sequence data. *PhytoKeys*: 125-139.
- Chen ZJ. 2007. Genetic and epigenetic mechanisms for gene expression and phenotypic variation in plant polyploids. *Annu Rev Plant Biol*, 58: 377-406.
- Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J, Plott C, Lovell J, Lin Y-M, Vaughn R, Liu B, Simpson S, Scheffler BE, Wen L, Saski CA, Grover CE, Hu G, Conover JL, Carlson JW, Shu S, Boston LB, Williams M, Peterson DG, McGee K, Jones DC, Wendel JF, Stelly DM, Grimwood J, Schmutz J. 2020. Genomic diversifications of five *Gossypium*

- allopolyploid species and their impact on cotton improvement. *Nature Genetics*, 52: 525-533.
- Cheng F, Sun C, Wu J, Schnable J, Woodhouse MR, Liang J, Cai C, Freeling M, Wang X. 2016. Epigenetic regulation of subgenome dominance following whole genome triplication in *Brassica rapa*. *New Phytol*, 211: 288-99.
- Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J. 2002. Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. *The Plant Cell*, 14: 1691-1704.
- Choulet F, Alberti A, Theil S, Glover N, Barbe V, Daron J, Pingault L, Sourdille P, Couloux A, Paux E. 2014. Structural and functional partitioning of bread wheat chromosome 3B. *Science*, 345: 1249721.
- Cohen S, Agmon N, Sobol O, Segal D. 2010. Extrachromosomal circles of satellite repeats and 5S ribosomal DNA in human cells. *Mobile DNA*, 1: 11.
- Cohen S, Houben A, Segal D. 2008. Extrachromosomal circular DNA derived from tandemly repeated genomic sequences in plants. *The Plant Journal*, 53: 1027-1034.
- Comai L. 2005. The advantages and disadvantages of being polyploid. *Nature Reviews Genetics*, 6: 836-846.
- Contento A, Heslop-Harrison JS, Schwarzacher T. 2005. Diversity of a major repetitive DNA sequence in diploid and polyploid Triticeae. *Cytogenetic and Genome Research*, 109: 34-42.
- Cronn RC, Zhao X, Paterson AH, Wendell JF. 1996. Polymorphism and concerted evolution in a tandemly repeated gene family: 5S ribosomal DNA in diploid and allopolyploid cottons. *Journal of Molecular Evolution*, 42: 685-705.
- Cuadrado A, Schwarzacher T. 1998. The chromosomal organization of simple sequence repeats in wheat and rye genomes. *Chromosoma*, 107: 587-594.
- Danilova TV, Akhunova AR, Akhunov ED, Friebe B, Gill BS. 2017. Major structural genomic alterations can be associated with hybrid speciation in *Aegilops markgrafii* (Triticeae). *The Plant Journal*, 92: 317-330.
- Daron J, Glover N, Pingault L, Theil S, Jamilloux V, Paux E, Barbe V, Mangenot S, Alberti A, Wincker P, Quesneville H, Feuillet C, Choulet F. 2014. Organization and evolution of transposable elements along the bread wheat chromosome 3B. *Genome Biology*, 15: 546.
- de Tomás C, Vicient CM. 2023. The Genomic Shock Hypothesis: Genetic and Epigenetic Alterations of Transposable Elements after Interspecific Hybridization in Plants. *Epigenomes*, 8.

- Decena MÁ, Sancho R, Inda LA, Pérez-Collazos E, Catalán P. 2024. Expansions and contractions of repetitive DNA elements reveal contrasting evolutionary responses to the polyploid genome shock hypothesis in *Brachypodium* model grasses. *Frontiers in Plant Science*, 15: 1419255.
- Dechyeva D, Gindullis F, Schmidt T. 2003. Divergence of satellite DNA and interspersion of dispersed repeats in the genome of the wild beet *Beta procumbens*. *Chromosome Research*, 11: 3-21.
- Dechyeva D, Schmidt T. 2006. Molecular organization of terminal repetitive DNA in *Beta* species. *Chromosome Research*, 14: 881-897.
- del Bosque MEQ, López-Flores I, Suárez-Santiago VN, Garrido-Ramos MA. 2014. Satellite-DNA diversification and the evolution of major lineages in Cardueae (Carduoideae Asteraceae). *Journal of Plant Research*, 127: 575-583.
- Devos KM, Brown JK, Bennetzen JL. 2002. Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome research*, 12: 1075-1079.
- Díaz-Pérez A, López-Álvarez D, Sancho R, Catalán P. 2018. Reconstructing the origins and the biogeography of species' genomes in the highly reticulate allopolyploid-rich model grass genus *Brachypodium* using minimum evolution, coalescence and maximum likelihood approaches. *Molecular Phylogenetics and Evolution*, 127: 256-271.
- Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novák P, Piednoël M, Weiss-Schneeweiss H, Leitch AR. 2014. Genomic Repeat Abundances Contain Phylogenetic Signal. *Systematic Biology*, 64: 112-126.
- Domingues DS, Cruz GMQ, Metcalfe CJ, Nogueira FTS, Vicentini R, de S Alves C, Van Sluys M-A. 2012. Analysis of plant LTR-retrotransposons at the fine-scale family level reveals individual molecular patterns. *BMC Genomics*, 13: 137.
- Douet J, Tourmente S. 2007. Transcription of the 5S rRNA heterochromatic genes is epigenetically controlled in *Arabidopsis thaliana* and *Xenopus laevis*. *Heredity*, 99: 5-13.
- Doyle J. 1991. DNA protocols for plants. *Molecular techniques in taxonomy*: Springer.
- Draper J, Mur LA, Jenkins G, Ghosh-Biswas GC, Bablak P, Hasterok R, Routledge AP. 2001. *Brachypodium distachyon*. A new model system for functional genomics in grasses. *Plant Physiol*, 127: 1539-55.
- Elliott TA, Gregory TR. 2015. What's in a genome? The C-value enigma and the evolution of eukaryotic genome content. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 370: 20140331.

- Eriksson M, Mandáková T, McCann J, Temsch E, Chase M, Hedrén M, Weiss-Schneeweiss H, Paun O. 2022. Repeat Dynamics across Timescales: A Perspective from Sibling Allotetraploid Marsh Orchids (*Dactylorhiza majalis s.l.*). *Molecular Biology and Evolution*, 39.
- Evans IJ, James AM, Barnes SR. 1983. Organization and evolution of repeated DNA sequences in closely related plant genomes. *J Mol Biol*, 170: 803-26.
- Fajkus P, Peška V, Sitová Z, Fulnečková J, Dvořáčková M, Gogela R, Sýkorová E, Hapala J, Fajkus J. 2016. *Allium* telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG) n is synthesized by telomerase. *The Plant Journal*, 85: 337-347.
- Fang L, Guan X, Zhang T. 2017. Asymmetric evolution and domestication in allotetraploid cotton (*Gossypium hirsutum L.*). *The Crop Journal*, 5: 159-165.
- Feldman M, Levy AA. 2005. Allopolyploidy – a shaping force in the evolution of wheat genomes. *Cytogenetic and Genome Research*, 109: 250-258.
- Feliner GN, Rosselló JA. 2007. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Molecular phylogenetics and evolution*, 44: 911-919.
- Feschotte C. 2008. Transposable elements and the evolution of regulatory networks. *Nature Reviews Genetics*, 9: 397-405.
- Feschotte C, Pritham EJ. 2007. DNA transposons and the evolution of eukaryotic genomes. *Annu Rev Genet*, 41: 331-68.
- Fjellheim S, Boden S, Trevaskis B. 2014. The role of seasonal flowering responses in adaptation of grasses to temperate climates. *Frontiers in Plant Science*, 5.
- Gaiero P, Vaio M, Peters SA, Schranz ME, de Jong H, Speranza PR. 2019. Comparative analysis of repetitive sequences among species from the potato and the tomato clades. *Annals of Botany*, 123: 521-532.
- Gao D. 2023. Introduction of Plant Transposon Annotation for Beginners. *Biology*, 12: 1468.
- Garcia S, Hidalgo O, Jakovljević I, Siljak-Yakovlev S, Vigo J, Garnatje T, Vallès J. 2013. New data on genome size in 128 Asteraceae species and subspecies, with first assessments for 40 genera, 3 tribes and 2 subfamilies. *Plant Biosystems*, 147.
- Garcia S, Kovařík A. 2013. Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. *Heredity*, 111: 23-33.
- Garcia S, Lim K, Chester M, Garnatje T, Pellicer J, Vallès J, Leitch A, Kovarik A. 2009. Linkage of 35S and 5S rRNA genes in *Artemisia* (family Asteraceae): First evidence from angiosperms. *Chromosoma*, 118: 85-97.

- Garcia S, Panero J, Siroky J, Kovarik A. 2010. Repeated reunions and splits feature the highly dynamic evolution of 5S and 35S ribosomal RNA genes (rDNA) in the Asteraceae family. *BMC plant biology*, 10: 176.
- Garcia S, Wendel JF, Borowska-Zuchowska N, Aïnouche M, Kuderova A, Kovarik A. 2020. The Utility of Graph Clustering of 5S Ribosomal DNA Homoeologs in Plant Allopolyploids, Homoploid Hybrids, and Cryptic Introgressants. *Frontiers in Plant Science*, Volume 11 - 2020.
- Garrido-Ramos MA. 2015. Satellite DNA in Plants: More than Just Rubbish. *Cytogenetic and Genome Research*, 146: 153-170.
- Garrido-Ramos MA. 2017. Satellite DNA: An Evolving Topic. *Genes (Basel)*, 8.
- Garsmeur O, Droc G, Antonise R, Grimwood J, Potier B, Aitken K, Jenkins J, Martin G, Charron C, Hervouet C. 2018. A mosaic monoploid reference sequence for the highly complex genome of sugarcane. *Nature communications*, 9: 2638.
- Gerlach WL, Dyer TA. 1980. Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. *Nucleic Acids Res*, 8: 4851-65.
- González ML, Chiapella JO, Urdampilleta JD. 2021. Chromosomal Differentiation of *Deschampsia* (Poaceae) Based on Four Satellite DNA Families. *Front Genet*, 12: 728664.
- Goodwin TJD, Poulter RTM. 2001. The DIRS1 Group of Retrotransposons. *Molecular Biology and Evolution*, 18: 2067-2082.
- Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Tartaglio VS, Session A, Martin J, Cartwright A, Katz A, Singan VR, Goltsman E, Barry K, Dinh-Thi VH, Chalhoub B, Diaz-Perez A, Sancho R, Lusinska J, Wolny E, Nibau C, Doonan JH, Mur LAJ, Plott C, Jenkins J, Hazen SP, Lee SJ, Shu S, Goodstein D, Rokhsar D, Schmutz J, Hasterok R, Catalan P, Vogel JP. 2020. Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nature Communications*, 11: 3670.
- Gordon SP, Liu L, Vogel JP. 2016. The Genus *Brachypodium* as a Model for Perenniality and Polyploidy. In: Vogel JP, ed. *Genetics and Genomics of Brachypodium*. Cham: Springer International Publishing.
- Grabundzija I, Messing SA, Thomas J, Cosby RL, Bilic I, Miskey C, Gogol-Döring A, Kapitonov V, Diem T, Dalda A, Jurka J, Pritham EJ, Dydá F, Izsvák Z, Ivics Z. 2016. A Helitron transposon reconstructed from bats reveals a novel mechanism of genome shuffling in eukaryotes. *Nature Communications*, 7: 10716.

- Guggisberg A, Baroux C, Grossniklaus U, Conti E. 2008. Genomic Origin and Organization of the Allopolyploid *Primula egaliksensis* Investigated by *in situ* Hybridization. *Annals of Botany*, 101: 919-927.
- Guo X, Han F. 2014. Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. *The Plant Cell*, 26: 4311-4327.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic acids symposium series*: Oxford.
- Hartig N, Seibt K, Heitkam T. 2023. How to start a LINE: 5' switching rejuvenates LINE retrotransposons in tobacco and related *Nicotiana* species. *The Plant journal : for cell and molecular biology*, 115.
- Hasterok R, Catalan P, Hazen SP, Roulin AC, Vogel JP, Wang K, Mur LA. 2022. *Brachypodium*: 20 years as a grass biology model system; the way forward? *Trends in Plant Science*, 27: 1002-1016.
- Hasterok R, Draper J, Jenkins G. 2004. Laying the Cytotaxonomic Foundations of a New Model Grass, *Brachypodium distachyon* (L.) Beauv. *Chromosome Research*, 12: 397-403.
- Hasterok R, Marasek A, Donnison IS, Armstead I, Thomas A, King IP, Wolny E, Idziak D, Draper J, Jenkins G. 2006a. Alignment of the Genomes of *Brachypodium distachyon* and Temperate Cereals and Grasses Using Bacterial Artificial Chromosome Landing With Fluorescence in Situ Hybridization. *Genetics*, 173: 349-362.
- Hasterok R, Wolny E, Hosiawa M, Kowalczyk M, Kulak-Ksiazczyk S, Ksiazczyk T, Heneen WK, Maluszynska J. 2006b. Comparative analysis of rDNA distribution in chromosomes of various species of Brassicaceae. *Annals of Botany*, 97: 205-216.
- Havukkala IJ. 1996. Cereal genome analysis using rice as a model. *Current Opinion in Genetics & Development*, 6: 711-714.
- Heitkam T, Weber B, Walter I, Liedtke S, Ost C, Schmidt T. 2020. Satellite DNA landscapes after allotetraploidization of quinoa (*Chenopodium quinoa*) reveal unique A and B subgenomes. *The Plant Journal*, 103: 32-52.
- Hemleben V, Ganai M, Gerstner J, Schiebel K, Torres R. 1988. Organization and length heterogeneity of plant ribosomal RNA genes. *Architecture of eukaryotic genes*: 371-383.
- Hemleben V, Volkov RA, Zentgraf U, Medina FJ. 2004. Molecular Cell Biology: Organization and Molecular Evolution of rDNA, Nucleolar Dominance, and Nucleolus Structure. In: Esser K, Lüttge U, Beyschlag W, Murata J, eds. *Progress in Botany: Genetics Physiology Systematics Ecology*. Berlin, Heidelberg: Springer Berlin Heidelberg.

- Hlavatá K, Závěská E, Leong-Škorníčková J, Pouch M, Poulsen AD, Šída O, Khadka B, Mandáková T, Fér T. 2024. Ancient hybridization and repetitive element proliferation in the evolutionary history of the monocot genus *Amomum* (Zingiberaceae). *Frontiers in Plant Science*, Volume 15 - 2024.
- Hloušková P, Mandáková T, Pouch M, Trávníček P, Lysak MA. 2019. The large genome size variation in the *Hesperis* clade was shaped by the prevalent proliferation of DNA repeats and rarer genome downsizing. *Annals of Botany*, 124: 103-120.
- Hodkinson TR. 2018. Evolution and taxonomy of the grasses (Poaceae): A model family for the study of species-rich groups. *Annual plant reviews Online*: 255-294.
- Hori Y, Engel C, Kobayashi T. 2023. Regulation of ribosomal RNA gene copy number, transcription and nucleolus organization in eukaryotes. *Nature Reviews Molecular Cell Biology*, 24: 414-429.
- Hus K, Betekhtin A, Pinski A, Rojek-Jelonek M, Grzebelus E, Nibau C, Gao M, Jaeger KE, Jenkins G, Doonan JH, Hasterok R. 2020. A CRISPR/Cas9-Based Mutagenesis Protocol for *Brachypodium distachyon* and Its Allopolyploid Relative, *Brachypodium hybridum*. *Front Plant Sci*, 11: 614.
- IBI. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature*, 463: 763-768.
- Ishiguro S, Taniguchi S, Schmidt N, Jost M, Wanke S, Heitkam T, Ohmido N. 2025. Repeatome landscapes and cytogenetics of hortensias provide a framework to trace *Hydrangea* evolution and domestication. *Annals of botany*, 135.
- Izawa T, Shimamoto K. 1996. Becoming a model plant: The importance of rice to plant science. *Trends in Plant Science*, 1: 95-99.
- Jang TS, Emadzade K, Parker J, Temsch EM, Leitch AR, Speta F, Weiss-Schneeweiss H. 2013. Chromosomal diversification and karyotype evolution of diploids in the cytologically diverse genus *Prospero* (Hyacinthaceae). *BMC Evol Biol*, 13: 136.
- Jenkins G, Hasterok R. 2007. BAC 'landing' on chromosomes of *Brachypodium distachyon* for comparative genome alignment. *Nature Protocols*, 2: 88-98.
- Kawabe A, Charlesworth D. 2007. Patterns of DNA Variation Among Three Centromere Satellite Families in *Arabidopsis halleri* and *A. lyrata*. *Journal of Molecular Evolution*, 64: 237-247.
- Kawabe A, Nasuda S. 2005. Structure and genomic organization of centromeric repeats in *Arabidopsis* species. *Molecular Genetics and Genomics*, 272: 593-602.
- Kellogg EA. 2015. *Flowering plants. Monocots: Poaceae*: Springer.



- Kellogg EA, Appels R. 1995. Intraspecific and interspecific variation in 5S RNA genes are decoupled in diploid wheat relatives. *Genetics*, 140: 325-343.
- Kent TV, Uzunović J, Wright SI. 2017. Coevolution between transposable elements and recombination. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 372: 20160458.
- Kolano B, McCann J, Orzechowska M, Siwinska D, Temsch E, Weiss-Schneeweiss H. 2016. Molecular and cytogenetic evidence for an allotetraploid origin of *Chenopodium quinoa* and *C. berlandieri* (Amaranthaceae). *Molecular Phylogenetics and Evolution*, 100: 109-123.
- Kolano B, McCann J, Oskędra M, Chrapek M, Rojek M, Nobis A, Weiss-Schneeweiss H. 2019. Parental origin and genome evolution of several Eurasian hexaploid species of *Chenopodium* (Chenopodiaceae). *Phytotaxa*, 392: 163–185-163–185.
- Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A. 2010. Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years. *New Phytol*, 186: 148-60.
- Kovarik A, Matyasek R, Lim K, Skalická K, Koukalova B, Knapp S, Chase M, Leitch A. 2004a. Concerted evolution of 18–5.8–26S rDNA repeats in *Nicotiana* allotetraploids. *Biological Journal of the Linnean Society*, 82: 615-625.
- Kovarik A, Matyasek R, Lim KY, Skalická K, Koukalová B, Knapp S, Chase M, Leitch AR. 2004b. Concerted evolution of 18–5.8–26S rDNA repeats in *Nicotiana* allotetraploids. *Biological Journal of the Linnean Society*, 82: 615-625.
- Kovarik A, Pires JC, Leitch AR, Lim KY, Sherwood AM, Matyasek R, Rocca J, Soltis DE, Soltis PS. 2005. Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. *Genetics*, 169: 931-44.
- Kubis S, Schmidt T, Heslop-Harrison JS. 1998. Repetitive DNA Elements as a Major Component of Plant Genomes. *Annals of Botany*, 82: 45-55.
- Kuhn GCS, Teo CH, Schwarzacher T, Heslop-Harrison JS. 2009. Evolutionary dynamics and sites of illegitimate recombination revealed in the interspersion and sequence junctions of two nonhomologous satellite DNAs in cactophilic *Drosophila* species. *Heredity*, 102: 453-464.
- Kunze R, Weil CF. 2007. The hAT and CACTA superfamilies of plant transposons. *Mobile DNA ii*: 565-610.
- Leitch AR, Leitch IJ. 2008. Genomic plasticity and the diversity of polyploid plants. *Science*, 320: 481-3.

- Levinson G, Gutman GA. 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. *Mol Biol Evol*, 4: 203-21.
- Li Y, Zuo S, Zhang Z, Li Z, Han J, Chu Z, Hasterok R, Wang K. 2018. Centromeric DNA characterization in the model grass *Brachypodium distachyon* provides insights on the evolution of the genus. *Plant J*, 93: 1088-1101.
- Lim KY, Kovarik A, Matyasek R, Chase MW, Clarkson JJ, Grandbastien MA, Leitch AR. 2007. Sequence of events leading to near-complete genome turnover in allopolyploid *Nicotiana* within five million years. *New Phytologist*, 175: 756-763.
- Ling H-Q, Ma B, Shi X, Liu H, Dong L, Sun H, Cao Y, Gao Q, Zheng S, Li Y. 2018. Genome sequence of the progenitor of wheat A subgenome *Triticum urartu*. *Nature*, 557: 424-428.
- Lisch D. 2002. Mutator transposons. *Trends in Plant Science*, 7: 498-504.
- Lisch D. 2015. Mutator and MULE Transposons. *Microbiol Spectr*, 3: Mdna3-0032-2014.
- Liu Y, Yi C, Fan C, Liu Q, Liu S, Shen L, Zhang K, Huang Y, Liu C, Wang Y. 2023. Pan-centromere reveals widespread centromere repositioning of soybean genomes. *Proceedings of the National Academy of Sciences*, 120: e2310177120.
- Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, Aguilar-Rodríguez J, Vicente-Ripolles M, Fuster G, Bernet GP, Maumus F, Munoz-Pomer A, Sempere JM, Latorre A, Moya A. 2010. The Gypsy Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Research*, 39: D70-D74.
- Llorens C, Muñoz-Pomer A, Bernad L, Botella H, Moya A. 2009. Network dynamics of eukaryotic LTR retroelements beyond phylogenetic trees. *Biology Direct*, 4: 41.
- Louzada S, Lopes M, Ferreira D, Adegá F, Escudeiro A, Gama-Carvalho M, Chaves R. 2020. Decoding the Role of Satellite DNA in Genome Architecture and Plasticity—An Evolutionary and Clinical Affair. *Genes*, 11: 72.
- Lower SS, McGurk MP, Clark AG, Barbash DA. 2018. Satellite DNA evolution: old ideas, new approaches. *Curr Opin Genet Dev*, 49: 70-78.
- Löytynoja A, Goldman N. 2010. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC bioinformatics*, 11: 1-7.
- Lunerová J, Renny-Byfield S, Matyasek R, Leitch A, Kovarik A. 2017. Concerted evolution rapidly eliminates sequence variation in rDNA coding regions but not in intergenic spacers in *Nicotiana tabacum* allotetraploid. *Plant Systematics and Evolution*, 303.
- Luo M-C, Gu YQ, Puiu D, Wang H, Twardziok SO, Deal KR, Huo N, Zhu T, Wang L, Wang Y. 2017. Genome sequence of the progenitor of the wheat D genome *Aegilops tauschii*. *Nature*, 551: 498-502.

- Lusinska J, Betekhtin A, Lopez-Alvarez D, Catalan P, Jenkins G, Wolny E, Hasterok R. 2019. Comparatively barcoded chromosomes of *Brachypodium* perennials tell the story of their karyotype structure and evolution. *International journal of molecular sciences*, 20: 5557.
- Lusinska J, Majka J, Betekhtin A, Susek K, Wolny E, Hasterok R. 2018. Chromosome identification and reconstruction of evolutionary rearrangements in *Brachypodium distachyon*, *B. stacei* and *B. hybridum*. *Annals of Botany*, 122: 445-459.
- Macas J, Navrátilová A, Koblízková A. 2006. Sequence homogenization and chromosomal localization of VicTR-B satellites differ between closely related *Vicia* species. *Chromosoma*, 115: 437-47.
- Macas J, Novák P, Pellicer J, Čížková J, Koblízková A, Neumann P, Fuková I, Doležel J, Kelly LJ, Leitch IJ. 2015. In Depth Characterization of Repetitive DNA in 23 Plant Genomes Reveals Sources of Genome Size Variation in the Legume Tribe Fabeae. *PLOS ONE*, 10: e0143424.
- Maccaferri M, Harris NS, Twardziok SO, Pasam RK, Gundlach H, Spannagl M, Ormanbekova D, Lux T, Prade VM, Milner SG. 2019. Durum wheat genome highlights past domestication signatures and future improvement targets. *Nature genetics*, 51: 885-895.
- Maddison WP. 2008. Mesquite: a modular system for evolutionary analysis. *Evolution*, 62: 1103-1118.
- Mahelka V, Kopecký D, Baum BR. 2013. Contrasting patterns of evolution of 45S and 5S rDNA families uncover new aspects in the genome constitution of the agronomically important grass *Thinopyrum intermedium* (Triticeae). *Molecular Biology and Evolution*, 30: 2065-2086.
- Maragheh FP, Janus D, Senderowicz M, Haliloglu K, Kolano B. 2019. Karyotype analysis of eight cultivated *Allium* species. *Journal of applied genetics*, 60: 1-11.
- Martínez-Sagarra G, Castro S, Mota L, Loureiro J, Devesa JA. 2021. Genome Size, Chromosome Number and Morphological Data Reveal Unexpected Intraspecific Variability in *Festuca* (Poaceae). *Genes (Basel)*, 12.
- Mascagni F, Giordani T, Ceccarelli M, Cavallini A, Natali L. 2017. Genome-wide analysis of LTR-retrotransposon diversity and its impact on the evolution of the genus *Helianthus* (L.). *BMC genomics*, 18: 1-16.
- Maughan PJ, Kolano BA, Maluszynska J, Coles ND, Bonifacio A, Rojas J, Coleman CE, Stevens MR, Fairbanks DJ, Parkinson SE, Jellen EN. 2006. Molecular and cytological characterization of ribosomal RNA genes in *Chenopodium quinoa* and *Chenopodium berlandieri*. *Genome*, 49: 825-839.

- Maumus F, Quesneville H. 2014. Deep Investigation of *Arabidopsis thaliana* Junk DNA Reveals a Continuum between Repetitive Elements and Genomic Dark Matter. *PLOS ONE*, 9: e94101.
- Mayer K, Waugh R, Langridge P, Close T, Wise R, Graner A, Matsumoto T, Sato K, Schulman A, Muehlbauer G. 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature*, 491: 711-716.
- Mayrose I, Lysak MA. 2020. The Evolution of Chromosome Numbers: Mechanistic Models and Experimental Approaches. *Genome Biology and Evolution*, 13.
- McCann J, Jang T-S, Macas J, Schneeweiss GM, Matzke NJ, Novák P, Stuessy TF, Villaseñor JL, Weiss-Schneeweiss H. 2018a. Dating the Species Network: Allopolyploidy and Repetitive DNA Evolution in American Daisies (*Melampodium* sect. *Melampodium*, Asteraceae). *Systematic Biology*, 67: 1010-1024.
- McClintock B. 1950. The origin and behavior of mutable loci in maize. *Proceedings of the National Academy of Sciences*, 36: 344-355.
- McDonald MJ, Wang WC, Huang HD, Leu JY. 2011. Clusters of nucleotide substitutions and insertion/deletion mutations are associated with repeat sequences. *PLoS Biol*, 9: e1000622.
- Mehrotra S, Goyal V. 2014. Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function. *Genomics, Proteomics & Bioinformatics*, 12: 164-171.
- Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, Sebra R, Peluso P, Eid J, Rank D, Garcia JF, DeRisi JL, Smith T, Tobias C, Ross-Ibarra J, Korf I, Chan SWL. 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biology*, 14: R10.
- Meng H, Feng J, Bai T, Jian Z, Chen Y, Wu G. 2020. Genome-wide analysis of short interspersed nuclear elements provides insight into gene and genome evolution in citrus. *DNA Research*, 27.
- Mestrović N, Plohl M, Mravinac B, Ugarković D. 1998. Evolution of satellite DNAs from the genus *Palorus* - experimental evidence for the "library" hypothesis. *Molecular biology and evolution*, 15: 1062-1068.
- Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, Von Haeseler A, Lanfear R. 2020. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Molecular biology and evolution*, 37: 1530-1534.
- Mlinarec J, Franjević D, Bočkor L, Besendorfer V. 2016. Diverse evolutionary pathways shaped 5S rDNA of species of tribe Anemoneae (Ranunculaceae) and reveal phylogenetic signal. *Botanical Journal of the Linnean Society*, 182: 80-99.

- Mlinarec J, Šatović Z, Malenica N, Ivančić-Baće I, Besendorfer V. 2012. Evolution of the tetraploid *Anemone multifida* ( $2n = 32$ ) and hexaploid *A. baldensis* ( $2n = 48$ ) (Ranunculaceae) was accompanied by rDNA loci loss and intergenomic translocation: evidence for their common genome origin. *Annals of Botany*, 110: 703-712.
- Mlinarec J, Skuhala A, Jurković A, Malenica N, McCann J, Weiss-Schneeweiss H, Bohanec B, Besendorfer V. 2019. The Repetitive DNA Composition in the Natural Pesticide Producer *Tanacetum cinerariifolium*: Interindividual Variation of Subtelomeric Tandem Repeats. *Frontiers in Plant Science*, 10.
- Moreno-Aguilar MF, Inda LA, Sánchez-Rodríguez A, Arnelas I, Catalán P. 2022. Evolutionary Dynamics of the Repeatome Explains Contrasting Differences in Genome Sizes and Hybrid and Polyploid Origins of Grass Loliinae Lineages. *Frontiers in Plant Science*, Volume 13 - 2022.
- Moreno R, Castro P, Vrána J, Kubaláková M, Cápál P, García V, Gil J, Millán T, Doležel J. 2018. Integration of genetic and cytogenetic maps and identification of sex chromosome in garden asparagus (*Asparagus officinalis* L.). *Frontiers in Plant Science*, 9: 1068.
- Mu W, Li K, Yang Y, Breiman A, Yang J, Wu Y, Zhu M, Wang S, Catalan P, Nevo E, Liu J. 2023. Subgenomic Stability of Progenitor Genomes During Repeated Allotetraploid Origins of the Same Grass *Brachypodium hybridum*. *Molecular Biology and Evolution*, 40.
- Murat F, Xu J-H, Tannier E, Abrouk M, Guilhot N, Pont C, Messing J, Salse J. 2010. Ancestral grass karyotype reconstruction unravels new mechanisms of genome shuffling as a source of plant evolution. *Genome research*, 20: 1545-1557.
- Naish M, Alonge M, Wlodzimierz P, Tock AJ, Abramson BW, Schmücker A, Mandáková T, Jamge B, Lambing C, Kuo P. 2021. The genetic and epigenetic landscape of the *Arabidopsis* centromeres. *Science*, 374: eabi7489.
- Naish M, Henderson IR. 2024. The structure, function, and evolution of plant centromeres. *Genome Res*, 34: 161-178.
- Nei M, Rooney AP. 2005. Concerted and birth-and-death evolution of multigene families. *Annu Rev Genet*, 39: 121-52.
- Neumann P, Novák P, Hošťáková N, Macas J. 2019. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mobile DNA*, 10: 1.
- Novák P, Guignard MS, Neumann P, Kelly LJ, Mlinarec J, Koblížková A, Dodsworth S, Kovařík A, Pellicer J, Wang W. 2020a. Repeat-sequence turnover shifts fundamentally in species with large genomes. *Nature Plants*, 6: 1325-1329.

- Novák P, Neumann P, Macas J. 2020c. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nature Protocols*, 15: 3745-3776.
- Oliveira EJ, Pádua JG, Zucchi MI, Vencovsky R, Vieira MLC. 2006. Origin, evolution and genome distribution of microsatellites. *Genetics and Molecular Biology*, 29: 294-307.
- Orozco-Arias S, Isaza G, Guyot R. 2019a. Retrotransposons in Plant Genomes: Structure, Identification, and Classification through Bioinformatics and Machine Learning. *International Journal of Molecular Sciences*, 20: 3837.
- Orozco-Arias S, Núñez-Rincón AM, Tabares-Soto R, López-Álvarez D. 2019b. Worldwide co-occurrence analysis of 17 species of the genus *Brachypodium* using data mining. *PeerJ*, 6: e6193.
- Otto SP. 2007. The Evolutionary Consequences of Polyploidy. *Cell*, 131: 452-462.
- Paço A, Freitas R, Vieira-da-Silva A. 2019. Conversion of DNA Sequences: From a Transposable Element to a Tandem Repeat or to a Gene. *Genes (Basel)*, 10.
- Parisod C, Holderegger R, Brochmann C. 2010. Evolutionary consequences of autopolyploidy. *New Phytologist*, 186: 5-17.
- Parisod C, Mhiri C, Lim KY, Clarkson JJ, Chase MW, Leitch AR, Grandbastien MA. 2012. Differential dynamics of transposable elements during long-term diploidization of *Nicotiana* section *Repandae* (Solanaceae) allopolyploid genomes. *PLoS One*, 7: e50352.
- Peška V, Fajkus P, Fojtová M, Dvořáčková M, Hapala J, Dvořáček V, Polanská P, Leitch AR, Sýkorová E, Fajkus J. 2015. Characterisation of an unusual telomere motif (TTTTTTAGGG)<sub>n</sub> in the plant *Cestrum elegans* (Solanaceae), a species with a large genome. *The Plant Journal*, 82: 644-654.
- Peska V, Garcia S. 2020. Origin, diversity, and evolution of telomere sequences in plants. *Frontiers in plant science*, 11: 117.
- Pfeifer M, Kugler KG, Sandve SR, Zhan B, Rudi H, Hvidsten TR, Consortium IWGS, Mayer KF, Olsen O-A. 2014. Genome interplay in the grain transcriptome of hexaploid bread wheat. *Science*, 345: 1250091.
- Pinhal D, Yoshimura TS, Araki CS, Martins C. 2011. The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. *BMC Evolutionary Biology*, 11: 151.
- Pont C, Wagner S, Kremer A, Orlando L, Plomion C, Salse J. 2019. Paleogenomics: reconstruction of plant evolutionary trajectories from modern and ancient DNA. *Genome Biology*, 20: 29.

- Puizina J, Weiss-Schneeweiss H, Pedrosa-Harand A, Kamenjarin J, Trinajstić I, Riha K, Schweizer D. 2003. Karyotype analysis in *Hyacinthella dalmatica* (Hyacinthaceae) reveals vertebrate-type telomere repeats at the chromosome ends. *Genome*, 46: 1070-1076.
- Renny-Byfield S, Kovarik A, Kelly LJ, Macas J, Novak P, Chase MW, Nichols RA, Pancholi MR, Grandbastien M-A, Leitch AR. 2013. Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *The Plant Journal*, 74: 829-839.
- Robertson IH. 1981. Chromosome numbers in *Brachypodium* Beauv.(Gramineae). *Genetica*, 56: 55-60.
- Röder MS, Plaschke J, König SU, Börner A, Sorrells ME, Tanksley SD, Ganai MW. 1995. Abundance, variability and chromosomal location of microsatellites in wheat. *Molecular and General Genetics MGG*, 246: 327-333.
- Rooney AP, Ward TJ. 2005. Evolution of a large ribosomal RNA multigene family in filamentous fungi: Birth and death of a concerted evolution paradigm. *Proceedings of the National Academy of Sciences*, 102: 5084-5089.
- Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM. 2016. High-throughput analysis of the satellitome illuminates satellite DNA evolution. *Scientific Reports*, 6: 28333.
- Sader M, Vaio M, Cauz-Santos LA, Dornelas MC, Vieira MLC, Melo N, Pedrosa-Harand A. 2021. Large vs small genomes in *Passiflora*: the influence of the mobilome and the satellitome. *Planta*, 253: 86.
- Salse J. 2016. Ancestors of modern plant crops. *Current Opinion in Plant Biology*, 30: 134-142.
- Salser W, Bowen S, Browne D, El-Adli F, Fedoroff N, Fry K, Heindell H, Paddock G, Poon R, Wallace B. 1976. Investigation of the organization of mammalian chromosomes at the DNA sequence level. *Federation proceedings*.
- Sancho R, Inda LA, Díaz-Pérez A, Des Marais DL, Gordon S, Vogel JP, Lusinska J, Hasterok R, Contreras-Moreira B, Catalán P. 2022. Tracking the ancestry of known and ‘ghost’ homeologous subgenomes in model grass *Brachypodium* polyploids. *The Plant Journal*, 109: 1535-1558.
- Scarlett VT, Lovell JT, Shao M, Phillips J, Shu S, Lusinska J, Goodstein DM, Jenkins J, Grimwood J, Barry K, Chalhoub B, Schmutz J, Hasterok R, Catalán P, Vogel JP. 2022. Multiple origins, one evolutionary trajectory: gradual evolution characterizes distinct lineages of allotetraploid *Brachypodium*. *Genetics*, 223.

- Schietgat L, Vens C, Cerri R, Fischer CN, Costa E, Ramon J, Carareto CMA, Blockeel H. 2018. A machine learning based framework to identify and classify long terminal repeat retrotransposons. *PLOS Computational Biology*, 14: e1006097.
- Schmidt N, Sielemann K, Breitenbach S, Fuchs J, Pucker B, Weisshaar B, Holtgräwe D, Heitkam T. 2024. Repeat turnover meets stable chromosomes: repetitive DNA sequences mark speciation and gene pool boundaries in sugar beet and wild beets. *The Plant Journal*, 118: 171-190.
- Schmidt T, Heslop-Harrison JS. 1996. The physical and genomic organization of microsatellites in sugar beet. *Proc Natl Acad Sci U S A*, 93: 8761-5.
- Schmidt T, Jung C, Metzlaff M. 1991. Distribution and evolution of two satellite DNAs in the genus *Beta*. *Theoretical and Applied Genetics*, 82: 793-799.
- Scholthof K-BG, Irigoyen S, Catalan P, Mandadi KK. 2018. *Brachypodium*: A Monocot Grass Model Genus for Plant Biology. *The Plant Cell*, 30: 1673-1694.
- Schulman AH. 2012. Hitching a ride: nonautonomous retrotransposons and parasitism as a lifestyle. *Plant transposable elements: impact on genome structure and function*: 71-88.
- Senerchia N, Felber F, Parisod C. 2014. Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats. *New Phytologist*, 202: 975-985.
- Shams I, Raskina O. 2018. Intraspecific and intraorganismal copy number dynamics of retrotransposons and tandem repeat in *Aegilops speltoides* Tausch (Poaceae, Triticeae). *Protoplasma*, 255: 1023-1038.
- Sharma A, Presting GG. 2014. Evolution of Centromeric Retrotransposons in Grasses. *Genome Biology and Evolution*, 6: 1335-1352.
- Shi X, Cao S, Wang X, Huang S, Wang Y, Liu Z, Liu W, Leng X, Peng Y, Wang N. 2023. The complete reference genome for grapevine (*Vitis vinifera* L.) genetics and breeding. *Horticulture Research*, 10: uhad061.
- Slotkin RK, Martienssen R. 2007. Transposable elements and the epigenetic regulation of the genome. *Nature Reviews Genetics*, 8: 272-285.
- Sochorová J, Coriton O, Kuderová A, Lunerová J, Chevre A-M, Kovařík A. 2017. Gene conversion events and variable degree of homogenization of rDNA loci in cultivars of *Brassica napus*. *Annals of botany*, 119: 13-26.
- Song J-M, Xie W-Z, Wang S, Guo Y-X, Koo D-H, Kudrna D, Gong C, Huang Y, Feng J-W, Zhang W. 2021. Two gap-free reference genomes and a global view of the centromere architecture in rice. *Molecular plant*, 14: 1757-1767.



- Song K, Lu P, Tang K, Osborn TC. 1995. Rapid genome change in synthetic polyploids of *Brassica* and its implications for polyploid evolution. *Proceedings of the National Academy of Sciences*, 92: 7719-7723.
- Sonnhammer EL, Durbin R. 1995. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*, 167: GC1-GC10.
- Soreng RJ, Peterson PM, Romaschenko K, Davidse G, Zuloaga FO, Judziewicz EJ, Filgueiras TS, Davis JI, Morrone O. 2015. A worldwide phylogenetic classification of the Poaceae (Gramineae). *Journal of Systematics and Evolution*, 53: 117-137.
- Spinnler F, Stöcklin J. 2018. DNA-content and chromosome number in populations of *Poa alpina* in the Alps reflect land use history. *Flora*, 246-247: 102-108.
- Steinwand MA, Young HA, Bragg JN, Tobias CM, Vogel JP. 2013. *Brachypodium sylvaticum*, a model for perennial grasses: transformation and inbred line development. *PLoS One*, 8: e75180.
- Stritt C, Wyler M, Gimmi EL, Pippel M, Roulin AC. 2020. Diversity, dynamics and effects of long terminal repeat retrotransposons in the model grass *Brachypodium distachyon*. *New Phytologist*, 227: 1736-1748.
- Su H, Liu Y, Liu C, Shi Q, Huang Y, Han F. 2019. Centromere Satellite Repeats Have Undergone Rapid Changes in Polyploid Wheat Subgenomes. *The Plant Cell*, 31: 2035-2051.
- Temnykh S, Park WD, Ayres N, Cartinhour S, Hauck N, Lipovich L, Cho YG, Ishii T, McCouch SR. 2000. Mapping and genome organization of microsatellite sequences in rice (*Oryza sativa* L.). *Theoretical and Applied Genetics*, 100: 697-712.
- Thakur J, Packiaraj J, Henikoff S. 2021. Sequence, Chromatin and Evolution of Satellite DNA. *International Journal of Molecular Sciences*, 22: 4309.
- Tran TD, Cao HX, Jovtchev G, Neumann P, Novák P, Fojtová M, Vu GT, Macas J, Fajkus J, Schubert I. 2015. Centromere and telomere sequence alterations reflect the rapid genome evolution within the carnivorous plant genus *Genlisea*. *The Plant Journal*, 84: 1087-1099.
- Trunova D, Borowska-Zuchowska N, Mykhailiyk S, Xia K, Zhu Y, Sancho R, Rojek-Jelonek M, Garcia S, Wang K, Catalan P, Kovarik A, Hasterok R, Kolano B. 2024. Does time matter? Intraspecific diversity of ribosomal RNA genes in lineages of the allopolyploid model grass *Brachypodium hybridum* with different evolutionary ages. *BMC Plant Biology*, 24: 981.

- Tynkevich YO, Shelyfist AY, Kozub LV, Hemleben V, Panchuk II, Volkov RA. 2022. 5S Ribosomal DNA of Genus *Solanum*: Molecular Organization, Evolution, and Taxonomy. *Frontiers in Plant Science*, Volume 13 - 2022.
- Ungerer MC, Strakosh SC, Stimpson KM. 2009. Proliferation of Ty3/gypsy-like retrotransposons in hybrid sunflower taxa inferred from phylogenetic data. *BMC Biology*, 7: 40.
- Untied I, Gruendler P. 1990. Nucleotide sequence of the 5.8S and 25S rRNA genes and of the internal transcribed spacers from *Arabidopsis thaliana*. *Nucleic Acids Research*, 18: 4011-4011.
- Vaio M, Mazzella C, Guerra M, Speranza P. 2019. Effects of the diploidisation process upon the 5S and 35S rDNA sequences in the allopolyploid species of the Dilatata group of *Paspalum* (Poaceae, Paniceae). *Australian Journal of Botany*, 67: 521-530.
- Venora G, Blangiforti S, Frediani M, Maggini F, Gelati MT, Castiglione MR, Cremonini R. 2000. Nuclear DNA contents, rDNAs, chromatin organization, and karyotype evolution in *Vicia* sect. *faba*. *Protoplasma*, 213: 118-125.
- Vitales D, Garcia S, Dodsworth S. 2020. Reconstructing phylogenetic relationships based on repeat sequence similarities. *Molecular Phylogenetics and Evolution*, 147: 106766.
- Vitte C, Panaud O. 2005. LTR retrotransposons and flowering plant genome size: emergence of the increase/decrease model. *Cytogenetic and Genome Research*, 110: 91-107.
- Volkov RA, Komarova NY, Hemleben V. 2007. Ribosomal DNA in plant hybrids: Inheritance, rearrangement, expression. *Systematics and Biodiversity*, 5: 261-276.
- Vu G, Schmutzer T, Bull F, Cao H, Fuchs J, Tran T, Jovtchev G, Pistrick K, Stein N, Pecinka A, Neumann P, Novák P, Macas J, Dear PH, Blattner F, Scholz U, Schubert I. 2015. Comparative Genome Analysis Reveals Divergent Genome Size Evolution in a Carnivorous Plant Genus. *The Plant Genome*.
- Wang D, Zheng Z, Li Y, Hu H, Wang Z, Du X, Zhang S, Zhu M, Dong L, Ren G. 2021. Which factors contribute most to genome size variation within angiosperms? *Ecology and evolution*, 11: 2660-2668.
- Wang L, Wang Y, Zhang J, Feng Y, Chen Q, Liu Z-S, Liu C-L, He W, Wang H, Yang S-F, Zhang Y, Luo Y, Tang H-R, Wang X-R. 2022. Comparative Analysis of Transposable Elements and the Identification of Candidate Centromeric Elements in the *Prunus* Subgenus *Cerasus* and Its Relatives. *Genes*, 13: 641.
- Wang W, Zhang X, Garcia S, Leitch AR, Kovařík A. 2023. Intragenomic rDNA variation - the product of concerted evolution, mutation, or something in between? *Heredity*, 131: 179-188.

- Wei L, Liu B, Zhang C, Yu Y, Yang X, Dou Q, Dong Q. 2020. Identification and characterization of satellite DNAs in *Poa L.* *Molecular Cytogenetics*, 13: 47.
- Weiss-Schneeweiss H, Blösch C, Turner B, Villaseñor JL, Stuessy TF, Schneeweiss GM. 2012. The promiscuous and the chaste: frequent allopolyploid speciation and its genomic consequences in American daisies (*Melampodium* sect. *Melampodium*; Asteraceae). *Evolution*, 66: 211-228.
- Weiss H, Scherthan H. 2002. Aloe spp.—plants with vertebrate-like telomeric sequences. *Chromosome Research*, 10: 155-164.
- Wendel JF, Schnabel A, Seelanan T. 1995. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proceedings of the National Academy of Sciences*, 92: 280-284.
- Wenke T, Döbel T, Sörensen TR, Junghans H, Weisshaar B, Schmidt T. 2011. Targeted Identification of Short Interspersed Nuclear Element Families Shows Their Widespread Existence and Extreme Heterogeneity in Plant Genomes. *The Plant Cell*, 23: 3117-3128.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews Genetics*, 8: 973-982.
- Wlodzimierz P, Rabanal FA, Burns R, Naish M, Primetis E, Scott A, Mandáková T, Gorringer N, Tock AJ, Holland D. 2023. Cycles of satellite and transposon evolution in *Arabidopsis* centromeres. *Nature*, 618: 557-565.
- Wolny E, Hasterok R. 2009. Comparative cytogenetic analysis of the genomes of the model grass *Brachypodium distachyon* and its close relatives. *Ann Bot*, 104: 873-81.
- Wolny E, Lesniewska K, Hasterok R, Langdon T. 2011a. Compact genomes and complex evolution in the genus *Brachypodium*. *Chromosoma*, 120: 199-212.
- Wood JRI, Muñoz-Rodríguez P, Wells T, Espinel-Ortiz DA, Romoleroux K, Cerón Martínez CE, Cornejo X, Scotland RW. 2024. Notes on *Ipomoea* (Convolvulaceae) from Ecuador. *Kew Bulletin*, 79: 915-939.
- Yang Y, Wu Z, Wu Z, Li T, Shen Z, Zhou X, Wu X, Li G, Zhang Y. 2023. A near-complete assembly of asparagus bean provides insights into anthocyanin accumulation in pods. *Plant Biotechnology Journal*, 21: 2473-2489.
- Zakrzewski F, Wenke T, Holtgräwe D, Weisshaar B, Schmidt T. 2010. Analysis of a c0t-1 library enables the targeted identification of minisatellite and satellite families in *Beta vulgaris*. *BMC Plant Biology*, 10: 8.

- Zattera ML, Bruschi DP. 2022. Transposable Elements as a Source of Novel Repetitive DNA in the Eukaryote Genome. *Cells*, 11: 3373.
- Zhang L, Liang J, Chen H, Zhang Z, Wu J, Wang X. 2023. A near-complete genome assembly of *Brassica rapa* provides new insights into the evolution of centromeres. *Plant biotechnology journal*, 21: 1022-1032.
- Zhao XP, Si Y, Hanson RE, Crane CF, Price HJ, Stelly DM, Wendel JF, Paterson AH. 1998. Dispersed repetitive DNA has spread to new genomes since polyploid formation in cotton. *Genome Res*, 8: 479-92.

## **10. SUPPLEMENTARY MATERIALS**

### **Supplementary material 1.**

RESEARCH

Open Access



# Does time matter? Intraspecific diversity of ribosomal RNA genes in lineages of the allopolyploid model grass *Brachypodium hybridum* with different evolutionary ages

Dana Trunova<sup>1</sup>, Natalia Borowska-Zuchowska<sup>1</sup>, Serhii Mykhailiyk<sup>1</sup>, Kai Xia<sup>2</sup>, Yuanbin Zhu<sup>2</sup>, Ruben Sancho<sup>3</sup>, Magdalena Rojek-Jelonek<sup>1</sup>, Sònia Garcia<sup>4</sup>, Kai Wang<sup>2</sup>, Pilar Catalan<sup>3</sup>, Ales Kovarik<sup>5</sup>, Robert Hasterok<sup>1</sup> and Bozena Kolano<sup>1\*</sup>

## Abstract

**Background** Polyploidisation often results in genome rearrangements that may involve changes in both the single-copy sequences and the repetitive genome fraction. In this study, we performed a comprehensive comparative analysis of repetitive DNA, with a particular focus on ribosomal DNA (rDNA), in *Brachypodium hybridum* ( $2n=4x=30$ , subgenome composition DDSS), an allotetraploid resulting from a natural cross between two diploid species that resemble the modern *B. distachyon* ( $2n=10$ ; DD) and *B. stacei* ( $2n=20$ ; SS). Taking advantage of the recurrent origin of *B. hybridum*, we investigated two genotypes, Bhyb26 and ABR113, differing markedly in their evolutionary age (1.4 and 0.14 Mya, respectively) and which resulted from opposite cross directions. To identify the origin of rDNA loci we employed cytogenetic and molecular methods (FISH, gCAPS and Southern hybridisation), phylogenetic and genomic approaches.

**Results** Unlike the general maintenance of doubled gene dosage in *B. hybridum*, the rRNA genes showed a remarkable tendency towards diploidisation at both locus and unit levels. While the partial elimination of 35S rDNA units occurred in the younger ABR113 lineage, unidirectional elimination of the entire locus was observed in the older Bhyb26 lineage. Additionally, a novel 5S rDNA family was amplified in Bhyb26 replacing the parental units. The 35S and 5S rDNA units were preferentially eliminated from the S- and D-subgenome, respectively. Thus, in the more ancient *B. hybridum* lineage, Bhyb26, 5S and 35S rRNA genes are likely expressed from different subgenomes, highlighting the complexity of polyploid regulatory networks.

**Conclusion** Comparative analyses between two *B. hybridum* lineages of distinct evolutionary ages revealed that although the recent lineage ABR113 exhibited an additive pattern of rDNA loci distribution, the ancient lineage Bhyb26 demonstrated a pronounced tendency toward diploidisation manifested by the reduction in the number of

\*Correspondence:  
Bozena Kolano  
bozena.kolano@us.edu.pl

Full list of author information is available at the end of the article



© The Author(s) 2024. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

both 35S and 5S loci. In conclusion, the age of the allopolyploid appears to be a decisive factor in rDNA turnover in *B. hybridum*.

**Keywords** rDNA loci, *Brachypodium hybridum*, 5S rDNA NTS, nrITS, 35S rDNA IGS, FISH

## Introduction

Polyploidy is among the major forces driving plant evolution and is often linked to the origin of key innovations found in many plant lineages [1, 2]. An ancient whole genome duplication (WGD) preceded angiosperm diversification, and further polyploidy events can be traced back to the origin and diversification of the major flowering plant lineages [3, 4]. Among polyploids, two groups are distinguished: (i) autopolyploids, which have more than two of the same genomes in their nuclei (e.g., autotetraploid *Arabidopsis arenosa*) [5], and (ii) allopolyploids, which have a hybrid origin and contain two or more duplicated different genomes, e.g., allohexaploid *Triticum aestivum* and allotetraploid *Chenopodium quinoa* [6, 7]. Both autopolyploid and allopolyploid taxa may arise several times from the same or highly similar parents [8–12]. Multiple origins of the same species can result in karyotypic, genomic, transcriptomic and phenotypic variation across lineages, as demonstrated in recently originated allotetraploid species of *Tragopogon* and *Achillea* (Asteraceae) [11, 13]. Several model plant polyploid systems have been used to study different aspects of post-polyploidisation diploidisation, which leads to rapid and extensive genomic changes, such as genome downsizing, structural chromosome rearrangements, amplifications and/or reactivation of repetitive elements, modification of the gene expression patterns, and concerted evolution of multigene families, e.g., ribosomal RNA (rRNA) genes [14–20].

A gradual evolution of a polyploid genome has been shown for the model grass system of the natural allotetraploid *Brachypodium hybridum* ( $2n=4x=30$ , subgenome composition DDSS) and two modern diploid species that likely resemble its ancestors: *B. distachyon* ( $2n=2x=10$ , DD) and *B. stacei* ( $2n=2x=20$ , SS) [21–24]. The phylogenetic studies based on the chloroplast DNA of *B. hybridum* revealed that at least two independent hybridisation events have occurred between *B. distachyon* and *B. stacei* [22]. Crosses between progenitor species occurred in both directions; therefore, two lineages of *B. hybridum* were distinguished: (i) plastotype with the chloroplast DNA (cpDNA) derived from *B. stacei* (S-plastome), and (ii) plastotype with the cpDNA derived from *B. distachyon* (D-plastome) [22, 25–27]. The *B. hybridum* lineages with ancestral D-plastome and S-plastome were formed at approximately 1.4 and 0.14 Mya, here represented by accessions Bhyb26 and ABR113, respectively. Multiple lines of evidence suggest that *B. hybridum* plants with D- and S-plastomes are

reproductively isolated [22]. A high level of collinearity between the genomes of the evolutionary younger *B. hybridum* (S-plastome) and the diploid *B. distachyon* and *B. stacei* was shown [22, 26–28], while the older lineage (D-plastotype) was characterised by some gene loss and few genome rearrangements when compared with the younger lineage (*B. hybridum* ABR113) and diploid species [22, 27]. No trace of homoeologous recombination and significant mobile element activation was found for both hybrid lineages [26]. However, there was evidence implying post-polyploidisation transposon activity in them (*B. hybridum*) [26, 27]. Many analyses reported that allopolyploids often preferentially retain genes from one dominant subgenome [29]. When considering the single-copy genes only, neither of the *B. hybridum* genotypes showed genome dominance at the transcriptional level; however, a biased gene loss appeared stronger in the older lineage [22, 27], although it may have resulted from the lack of contemporary progenitor genomes for an accurate comparative analysis [26].

In eukaryotes, housekeeping genes include the tandemly organised rRNA genes (35S rRNA genes and 5S rRNA genes) [30]. Each unit of 35S rDNA contains conserved genic regions encoding for 18S–5.8S–25S and intergenic spacer (IGS) and two internal transcribed spacers (ITS1 and ITS2) [31]. The 5S rDNA units are usually distributed in separate locus/loci in the genome and consist of the conserved genic region and variable non-transcribed spacer (5S rDNA NTS) [31, 32]; however, the linked rDNA configuration (in which the 5S rDNA is inserted into the 35S rDNA IGS) was also described [33, 34]. The evolutionary pathways of rRNA genes may not be the same as the single-copy ones, as was shown indeed in *B. hybridum* ABR113, where a significant reduction of the S-subgenome rDNA units was observed [35] while there was a balanced composition and expression of single-copy genes [22, 26]. Previous studies also revealed that the *B. stacei*-like 35S rDNA loci were transcriptionally silenced in all studied tissues of this genotype [36–38], showing a strong nucleolar dominance (ND) towards the D-subgenome rDNA. Until now, stable ND has been confirmed in a wide range of *B. hybridum* genotypes originating from different climatic zones [35]. Recent studies on another *B. hybridum* genotype (3-7-2; new name: Bhyb3\_7\_2) revealed tissue-specificity of ND in this genotype [39]. The evolutionary patterns of 35S rRNA genes have not been studied yet in *B. hybridum* Bhyb26. Also, little is known about the 5S rDNA evolution in this species. In most genotypes, both ancestral 5S rDNA loci are

present [35, 40, 41]. However, previous work detected a significant reduction of S-genome 5S rDNA copies in one of the genotypes using a bioinformatics approach [40].

The current study draws a complex picture of rRNA gene evolutionary trajectories in two *B. hybridum* lineages of distinct evolutionary ages and cross directions. We aimed to address the following questions: (i) are the evolutionary trajectories of 35S and 5S rDNA homoeologues similar?; (ii) is there a relationship between the direction of the cross and rDNA homoeologues loss in *B. hybridum*?; (iii) do the rRNA genes in the studied *B. hybridum* lineages of different evolutionary ages follow the same evolutionary pathways?

Materials and methods

Plant material

Seven genotypes of the three annual *Brachypodium* species were analysed (Table 1). Plants used for DNA isolation were grown from seeds in pots with soil mixed with vermiculite (3:1 v/v) at 20 °C and 16/8 h photoperiod in the greenhouse of the University of Silesia in Katowice, Poland. For cytogenetic analyses, the primary roots of the seedlings were used. Seeds were germinated on a filter paper moistened with water for 3–5 days at 20–22 °C in the dark. Whole seedlings with approximately two cm-long roots were placed in ice-cold water overnight and fixed in 3:1 (v/v) methanol: glacial acetic acid.

NGS data and in silico identification of rDNA repeats

The raw Illumina sequencing data for *B. distachyon* Bd21-3 (SRR4236817), *B. stacei* ABR114 (SRR1800504), *B. hybridum* ABR113 (SRR3945061) and *B. hybridum* Bhyb26 (SRR4184872), available in the European Nucleotide Archive (ENA), were utilised to identify rDNA repeats. Depending on the sequencing library, the paired-end read length ranged from 150 to 250 bp. The Illumina reads were trimmed to a length of 150 bp, and the high quality of the trimmed reads was ensured by the QC tools of RepeatExplorer2 [42, 43]. All libraries were sampled at random to obtain 1× coverage of the genome. A graph-based clustering of paired-end reads was performed to

identify, characterise and quantify repetitive sequences with the web-based RepeatExplorer2 pipeline implemented in the Galaxy server [https://galaxy-elixir.cerit-sc.cz/; 42]. The reads were analysed using default parameters (90% similarity, minimum overlap=55; cluster size threshold=0.01%) in each *Brachypodium* genotype separately [42, 43].

DNA isolation, polymerase chain reaction and cloning

Total genomic DNA (gDNA) was isolated from young and healthy leaves of 1-month-old plants using a cetyltrimethylammonium bromide-based (CTAB) method [44]. To analyse the ancestral contribution of rDNA variants in the allotetraploid, the nrITS, IGS of 35S rDNA and 5S rDNA NTS were cloned and sequenced. nrITS (ITS1-5.8S rDNA-ITS2) was amplified using a primer pair anchored in 18S rDNA and 25S rDNA (18Sdir and 25Scom; Table S1) [45] as earlier described [46]. The PCR product was treated with *E. coli* Exonuclease I and FastAP Thermosensitive Alkaline Phosphatase (Thermo Fisher Scientific, Waltham, USA) according to the manufacturer’s instruction, and the cycle sequencing was performed using a 3730xl DNA Analyzer (Applied Biosystems; USA) in a commercial facility (Macrogen, Amsterdam, Netherlands). The 35S rDNA intergenic spacer of *B. hybridum* genotype Bhyb26 was amplified from the gDNA using a primer pair designed to match the conserved regions of the 18S and 25S rRNA genes (Table S1) [47] as described [36]. Amplification products were separated by electrophoresis on a 1% agarose gel stained with GelRed (Sigma-Aldrich, Steinheim, Germany) and visualised with a UV transilluminator. PCR products were gel-purified using the QIAquick Gel Extraction Kit (Qiagen, Germany). The amplicons were cloned using the pGEM-T Easy vector system (Promega, Madison, USA) following the manufacturer’s instructions. Four clones were sequenced using the Sanger method and the primer-walking strategy in a commercial facility (Genomed, Warsaw, Poland). The 5S rDNA non-transcribed spacer (5S rDNA NTS) was amplified using a primer pair anchored in the 5S rRNA genic region (Table S1) [48] as earlier described [7].

**Table 1** General characteristics of the *Brachypodium* species used in this study

Species	Genotype	2n	x	Genome /subgenome designation	Origin	Source
<i>Brachypodium distachyon</i> (L.) P. Beauv.	Bd21-3	10	5	DD	Iraq	A
	Bd21	10	5	DD	Iraq	A
	ABR5	10	5	DD	Spain, Huesca	B
<i>Brachypodium stacei</i> Catalán, Joch. Müll., Mur & Langdon	ABR114	20	10	SS	Spain, Formentera	B
	Bsta5	20	10	SS	Spain, Alicante	C
<i>Brachypodium hybridum</i> Catalán, Joch. Müll., Hasterok & Jenkins	ABR113	30	10 + 5	DDSS	Portugal, Lisbon	B
	Bhyb26	30	10 + 5	DDSS	Spain, Jaen, La Cimbarrá	C

A, US Department of Agriculture, National Plant Germplasm System, Beltsville, MD, USA; B, Institute of Biological, Environmental and Rural Sciences, Aberystwyth University, Aberystwyth, UK; C, High Polytechnic School of Huesca, University of Zaragoza, Huesca, Spain



PCR products were gel-purified as described above, and ligated into pGEM-T Easy vectors (Promega, Madison, USA) following the manufacturer's instructions. Depending on the ploidy level, five to ten randomly chosen recombinant colonies were selected for further analyses, and the universal primer M13 was used for sequencing. Plasmid DNA was isolated using a QIAGEN Plasmid Mini Kit according to the manufacturer's instructions. Sanger sequencing was performed in a commercial facility (Macrogen, Amsterdam, Netherlands). All the newly generated sequences were deposited in GenBank: 5S rDNA NTS (PP339782- PP339812, PP339815, PP339816, PP339825), 35S rDNA IGS (PP339781) and nrITS (PP317522; Table S2).

### Bioinformatic and phylogenetic analyses

Comparative analyses were performed to assess the contribution of ancestral variants in both analysed allo-tetraploid lineages and to reveal the direction of rDNA sequence evolution in this tetraploid. Newly obtained sequences were assembled using CLC Genomics Workbench (Qiagen, Hilden, Germany). Multiple sequence alignments were performed in Geneious Prime 2022.0.1 (<https://www.geneious.com>) and manually adjusted. Comparative analysis of 35S rDNA IGS among analysed genotypes was done using newly obtained sequences and those published earlier (GenBank accession numbers: KX263276, KX263278, KX26327). Multiple sequence alignments for the 35S rDNA IGS were manually adjusted in Geneious Prime 2022.0.1.

The sequence of the cloned IGS from Bhyb26 was used as the reference sequence in the *B. hybridum* Bhyb26 raw Illumina reads mapping using the 'Map Read Reference' tool in the CLC Genomics Workbench. All reads containing ambiguous nucleotides, i.e. that were shorter than 150 bp and/or failed to pass a quality score limit of  $P=0.05$ , were removed using the 'TRIM' command in CLC Genomics Workbench. The mapping parameters were as follows: mismatch cost, 2; insertion cost, 3; deletion cost, 3; with the length fraction set at 0.5 and the similarity fraction set at 0.8. The consensus IGS sequences were then extracted from the mapped reads. Additionally, the 35S rDNA IGSs were compared using the DOTTER and Tandem Repeats Finder [49, 50].

Multiple sequence alignments for the 5S rDNA NTS dataset (in total 44 sequences; Table S2) were performed using webPRANK [51] and manually adjusted. Phylogenetic relationships for 5S rDNA NTS regions were inferred using maximum likelihood (ML), as implemented in IQ-TREE version 2.2.2 with default parameters [52, 53]. The best model of sequence evolution for the ML analysis, K2P, was determined using the Bayesian information criterion as implemented in IQ-TREE. The significance of the inferred relationships was assessed

via bootstrapping with 1000 replicates. The resulting phylogenetic tree was visualised using FigTree v.1.3.1. The copy numbers of the 35S rDNA and 5S rDNA units were calculated from the NGS read count using the following scheme: (i) the genome proportion (GP) was calculated from the number of mapped reads to either the 18S or 5S rDNA consensus sequence divided by the total number of reads in percentages; (ii) the genome spaces (GS) of particular rDNAs were determined using the following formula:  $GP \times \text{size of the genome in Mb}$  (618 Mb and 650 Mb for ABR113 and Bhyb26, respectively); (iii) the copy numbers of rDNAs were calculated as follows: GS value divided by the size of a either 18S rDNA in Mb (0.00181 Mb) or 5S rDNA coding region (0.000119 Mb).

### Restriction analyses (gCAPS) and Southern blot hybridisation

In order to determine the ancestral contribution of rDNA variants in studied lineages of *B. hybridum*, the following approaches have been performed: (i) the genomic cleaved amplified polymorphic sequence (gCAPS) of ITS1 region, (ii) Southern hybridisation with 25S rDNA and (iii) 5S rDNA as a probe. The genomic cleaved amplified sequence polymorphism analyses (gCAPS) were performed as described in [35]. Shortly, the ITS1 region of the 35S rRNA gene was amplified using a primer pair 18S-for and 5.8S-rev [Supplementary Table S1; 54]. The PCR products were then digested with a *MluI* restriction enzyme (New England Biolabs, Ipswich, USA) and size separated on 1% (w/v) agarose gel.

Southern blot hybridisation with 5S rDNA was performed using a non-radioactive method. The genomic DNA (2 µg) from *Brachypodium* spp. were subjected to digestion with *MseI* enzyme (New England Biolabs, Ipswich, USA) at 37 °C for 4 h. The digested gDNAs were separated on 1.2% (w/v) agarose gel and blotted onto positive-charged nylon membranes (Roche, Mannheim, Germany) using VacuBlot System (Biometra Analytik Jena, Jena, Germany). DNA was UV cross-linked to the membrane using a CK-1000 Ultraviolet Crosslinker (Ultra-Violet Products, Cambridge, UK). The 5S rDNA monomer of *Arabidopsis thaliana* [54] labelled with digoxigenin-11-dUTP, alkali-labile, using nick translation according to manufacturer's instructions (Roche, Basel, Switzerland) was used as a probe. Hybridisation was performed using a DIG High Prime DNA Labeling and Detection Starter Kit II (Roche, Mannheim, Germany) according to the manufacturer's instructions using an HB-1000 Hybridizer (UltraViolet Products, Cambridge, UK). Hybridisation signals were documented and quantified using ChemiDocXRS (BioRad, USA).

The Southern blot hybridisation of 35S rDNA was carried out following the methodology earlier described [55]. The genomic DNA (1 µg) was treated with *BglII*

restriction enzyme (New England Biolabs, Ipswich, USA) at 37°C for 2.5 h. Afterwards, the samples were loaded onto a 1% agarose gel, separated through electrophoresis, and transferred to a nylon membrane (Amersham Hybond, GE Healthcare, USA). The membrane was then hybridised using a 25S rRNA gene labelled with [ $\alpha$ - $^{32}$ P] dCTP as a probe, a 220 bp PCR product derived from the 3' end of the 25S rDNA from *Nicotiana tabacum*. The hybridisation was visualised using a phosphorimager (Typhoon 9500, GE Healthcare, USA), and the intensity of the 35S rDNA signal was measured using ImageQuant software (GE Healthcare, USA). Two biological replicates were performed for each probe.

#### Chromosome preparation and fluorescence in situ hybridisation (FISH)

The FISH method with 25S and 5S rDNA as probes was used to analyse the rRNA gene loci number and chromosomal localisation in the studied *Brachypodium* genotypes. The chromosome preparations were made following the published protocol [41]. A 2.3-kb fragment of the 25S rDNA coding region of *A. thaliana* [56] labelled with digoxigenin-11- dUTP (Roche, Basel, Switzerland) was used to detect 35S rDNA loci. The 5S rDNA monomer isolated from *T. aestivum* [54] and labelled with tetramethyl-rhodamine-5-dUTP (Roche, Basel, Switzerland) was used to detect the 5S rDNA loci. Both probes were labelled using nick translation according to the manufacturer's instructions (Roche, Basel, Switzerland). FISH was performed as previously described [41]. The hybridisation mixture consisting of 50% deionised formamide, 10% dextran sulphate, 2× SSC, 0.5% SDS (sodium dodecyl sulphate) and 100 ng of each labelled DNA probe was denatured for 10 min at 80 °C and immediately cooled down on ice. The denaturation of the slides and the hybridisation mixture were performed on an Omnislide Thermal cycler (ThermoHybaid, Franklin, MA, USA) at 75 °C for 4 min. Hybridisation was conducted for 20 h at 37 °C in a humid chamber. Post-hybridisation washes (10% deionised formamide in 0.1× SSC at 42 °C; stringency 76%) were followed by the immunodetection of digoxigenated probes using FITC-conjugated anti-digoxigenin antibodies (Roche, Basel, Switzerland). The slides were mounted in Vectashield (Vector Laboratories, Newark, CA, USA) containing 2.5 ng/μl of DAPI (4',6-diamidino-2-phenylindole dihydrochloride). All images were acquired using a Zeiss AxioImager.Z.2 fluorescent microscope equipped with an AxioCamHMR camera (Zeiss, Oberkochen, Germany). The images were processed uniformly using ZEN 2.3 Pro (Zeiss).

#### Genome size determination

The genome size of *B. hybridum* Bhyb26 was determined by flow cytometry. Sixteen Bhyb26 individuals, grown from seeds, were analysed and each sample was measured twice. All analyses were performed by the same operator when the plants were 4–5 weeks old; a maximum of five plants were analysed in one working day. The youngest fully developed leaves of the studied *B. hybridum* plants and the internal standard *Lycopersicon esculentum* Mill. cv. Stupicke (2C DNA=1.96 pg) [57] were chopped together in a Petri dish in 500 μl of a nuclei extraction buffer using a razor blade (Sysmex CyStain PI OxProtect, 05-5027-P01). The nuclei suspension was filtered through a 30 μm mesh (CellTrics, Sysmex, Kobe, Japan) and stained with a staining buffer containing propidium iodide and RNase (Sysmex CyStain PI OxProtect, 05-5027-P01) according to the manufacturer's instructions. The samples were incubated for 45 min in the dark and then analysed using a flow cytometer (CyFlow Space, Sysmex, Kobe, Japan) equipped with a 532 nm green laser. At least 10,000 nuclei were analysed for each sample. The sizes of the nuclear genomes were calculated as the linear relationship between the ratio of the 2C DNA peaks of a sample and the standard according to the formula:

$$\begin{aligned} \text{Sample 2C value (DNA pg or Mb)} \\ = \text{Reference 2C value} \times \frac{\text{sample 2C mean peak position}}{\text{reference 2C peak position}} \end{aligned}$$

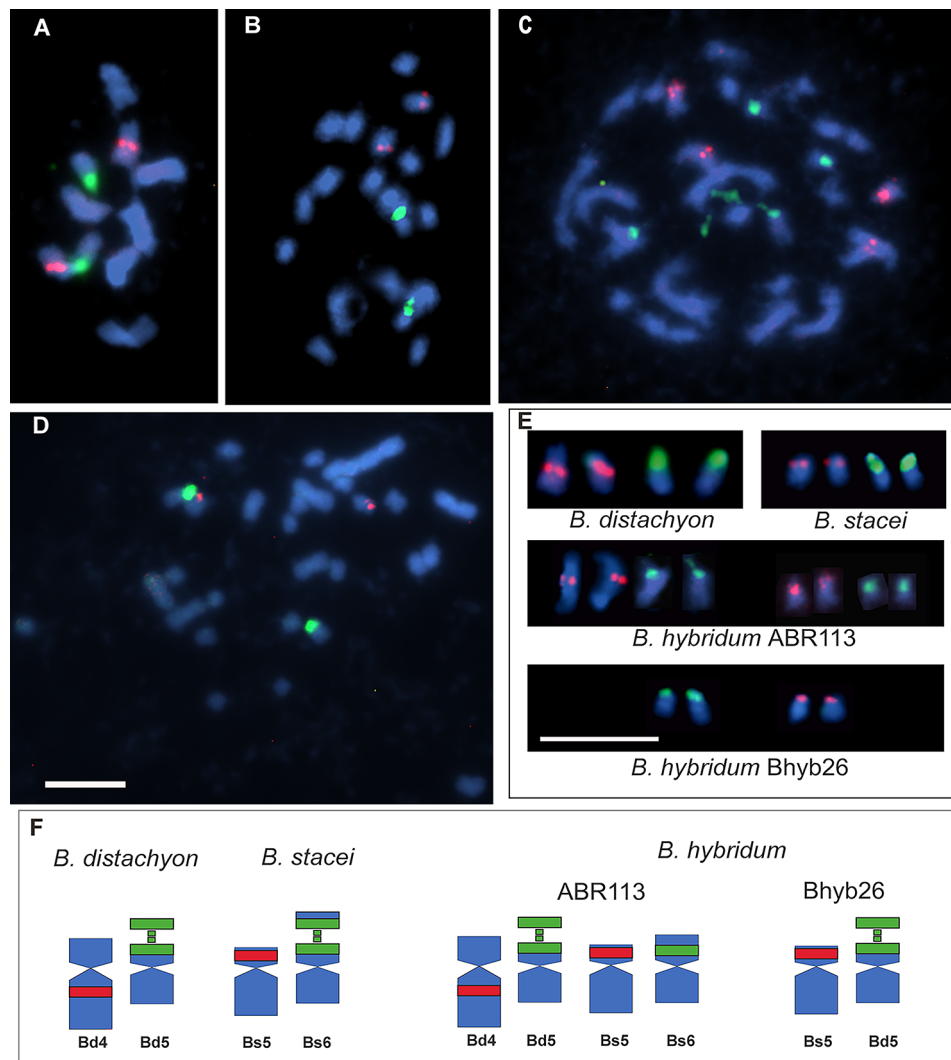
The conversion of pg to Mb was made according to the formula:

$$\text{DNA content (pg)} = \text{genome size (bp)} / (0.978 \times 10^9) \text{ [58]}.$$

## Results

#### Number and chromosomal distribution of 35S and 5S rDNA loci

The chromosome numbers were verified for all three analysed species. As expected,  $2n=10$  was observed in meristematic cells of *B. distachyon* and  $2n=20$ , with significantly smaller chromosomes, in *B. stacei*. Both genotypes of *B. hybridum* were characterised with  $2n=30$ , including ten larger chromosomes originating from the *B. distachyon*-like ancestor and 20 smaller ones derived from the *B. stacei*-like ancestor, as previously described [28]. In agreement with previous reports [28, 41], FISH with 5S rDNA and 25S rDNA as probes confirmed that *B. distachyon* has one 5S rDNA locus in the proximal region of the long arm of chromosome Bd4 and one 35S rDNA locus at the distal region of the short arm of chromosome Bd5 (Fig. 1A, E, F). An identical number of loci was confirmed in *B. stacei*, i.e. one 5S rDNA locus in the short arm of chromosome Bs5 and one 35S rDNA locus in the proximal region of the short arm of chromosome Bs6



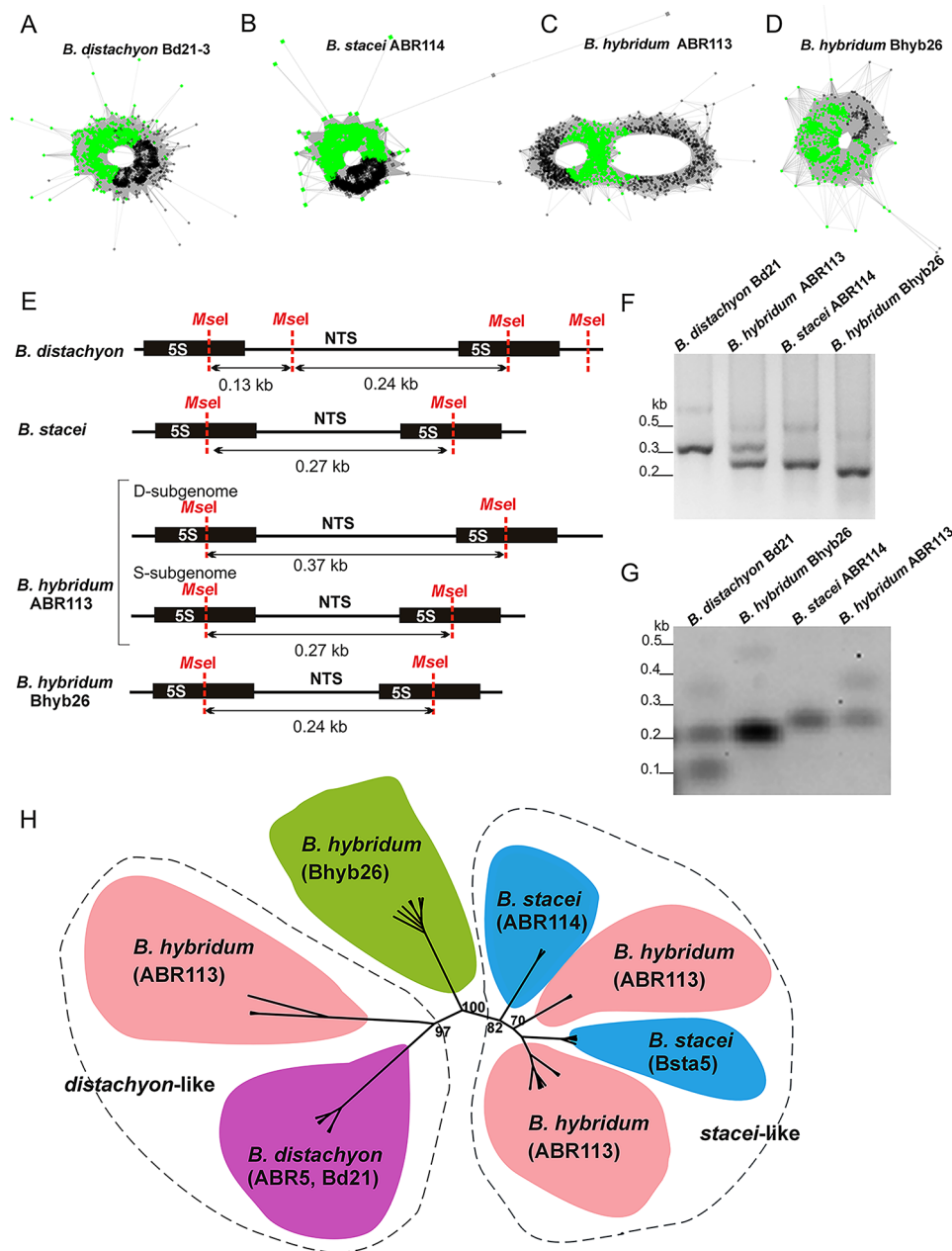
**Fig. 1** Distribution of the 35S and 5S rDNA loci on the mitotic chromosomes of the studied *Brachypodium* accessions. Fluorescence in situ hybridisation (A–E) of the 25S rDNA (green fluorescence) and 5S rDNA (red fluorescence) probes to (A) *B. distachyon* Bd21 ( $2n=10$ ;  $x=5$ ), (B) *B. stacei* ABR114 ( $2n=20$ ;  $x=10$ ), (C) *B. hybridum* ABR113 ( $2n=30$ ;  $x=5+10$ ) and (D) *B. hybridum* Bhyb26 ( $2n=30$ ;  $x=5+10$ ) chromosomes. (E) Chromosomal pairs that bear rDNA loci and (F) their idiograms. All bars: 5  $\mu$ m

(Fig. 1B, E and F) [22]. While the evolutionary younger *B. hybridum* genotype ABR113 had an additive number of 35S and 5S rDNA loci (Fig. 1C, E and F) derived from both ancestors, the evolutionary older Bhyb26 lineage, studied here for the first time, revealed only two chromosomal pairs bearing rDNA loci: one derived from the D-subgenome with the 35S rDNA site, and another pair originated from S-subgenome with the 5S rDNA site (Fig. 1D, E and F).

#### In silico identification and genome proportion of rDNA repeats

The reads complementary to 35S and 5S rDNA sequences were identified using RepeatExplorer2, and the genome proportion of both rDNA sequences was assessed from the raw Illumina reads. One cluster with reads

complementary to 5S rDNA was retrieved for each analysed lineage. A simple circular graph was reconstructed for the 5S rDNA cluster in both diploids and *B. hybridum* Bhyb26 (Fig. 2A, B and D). In contrast, in the case of ABR113, a complex graph with two loops interconnected by a junction region (composed of the 5S rDNA coding region) was revealed, indicating the presence of two different NTS variants (Fig. 2C). Based on the raw Illumina reads mapping, the copy number of 5S rDNA was estimated as 1367 units/1C and 2570 units/1C for the younger ABR113 (1C $\approx$ 618 Mb) and the older Bhyb26 (1C $\approx$ 650 Mb), respectively (Table 2). The analysed lineages of *B. hybridum* also differed in the genome proportion of 35S rDNA repeats. 35S rDNA accounted for 1356 units/1C and 747 units/1C in ABR113 and Bhyb26, respectively (Table 2).



**Fig. 2** Structure of 5S rDNA in the studied *Brachypodium* accessions. 5S rDNA sequence reads organised in graph structures from the RepeatExplorer2 graphical output of (A) *B. distachyon* Bd21-3, (B) *B. stacei* ABR114, (C) *B. hybridum* ABR113 and (D) *B. hybridum* Bhyb26. 5S rDNA coding sequences and intergenic spacers are highlighted as green and grey dots, respectively. Note, simple circle graphs indicate a single 5S rDNA family, and complex two-loop graphs indicate multiple families. (E) Length and *MseI* restriction maps of the 5S rDNA units of the analysed *Brachypodium* accessions. (F) Length analysis of 5S rDNA variants. Agarose gel electrophoresis of PCR products consisting of fragments of 5S rDNA coding region (102 bp) and NTSs (the original gel is shown in Supplementary Figure S1). (G) Southern blot hybridisation of the *MseI*-restricted genomic DNA from *B. distachyon*, *B. stacei*, *B. hybridum* ABR113 and Bhyb26 genotypes digested with *BglII* and hybridised with the 5S rDNA probe (the original membrane is shown in Supplementary Figure S2). (H) The unrooted phylogenetic tree of the studied annual *Brachypodium* species inferred through maximum likelihood analysis of the 5S rDNA NTS alignment. Bootstrap support scores  $\geq 70\%$  are shown near branches

### 5S rDNA structure

A PCR with primers anchored in coding 5S rDNA sequences was performed on genomic DNA to determine the lengths of 5S rDNA units in *Brachypodium* diploids (parental species) and *B. hybridum* lineages (Fig. 2E and

F). The amplicons contained the entire non-transcribed spacer and the 102-bp-long fragment of the coding sequence. The longest PCR product of about 350 bp was obtained for *B. distachyon*, while the *B. stacei* one was much shorter, approximately 250 bp. Two prominent



**Table 2** Copy number of the rDNA units in the studied *Brachypodium hybridum* accessions determined from the raw Illumina reads

Species	Genotype	Read archive accession	Mapped reads <sup>a</sup>	Total number of reads	GP [%] <sup>b</sup>	GS <sup>c</sup>	Copies <sup>d</sup>
<b>18S rDNA</b>							
<i>B. hybridum</i>	ABR113	SRR3945061	7902	2,000,000	0.21	2.44	1356
<i>B. hybridum</i>	Bhyb26	SRR4184872	4136	2,000,000	0.40	1.34	747
<b>5S rDNA</b>							
<i>B. hybridum</i>	ABR113	SRR3945061	531	2,000,000	0.03	0.16	1367
<i>B. hybridum</i>	Bhyb26	SRR4184872	949	2,000,000	0.05	0.31	2570

<sup>a</sup> The length of the 18S rDNA consensus sequence used as reference was 1.81 kb, and this of 5S rDNA was 119 bp  
<sup>b</sup> The genome proportion (GP) was calculated from the number of mapped reads to the consensus sequence divided by the total number of reads in percentages  
<sup>c</sup> The genome space (GS) was calculated as (genome size in Mb × GP × 100<sup>-1</sup>)  
<sup>d</sup> The copy number of the rDNA units was calculated as GS/size of a single 18S rDNA unit (0.00181 Mb) or 5S rDNA unit (0.000119 Mb). The following genome sizes were considered: *B. hybridum* ABR113 0.633 pg/1C ≈ 618 Mb (Catalán et al. 2012); *B. hybridum* Bhyb26 0.655 pg/1C ≈ 650 Mb (present study). The conversion of pg to Mb was made according to Dolezel et al. (2003)

products of about 350 and 250 bp were amplified from the younger *B. hybridum* ABR113, suggesting the presence of two 5S rRNA gene families in this genotype. In contrast, only one prominent band of approximately 220 bp was observed for the older Bhyb26 (Fig. 2F).

The contribution of different length variants of 5S rDNA NTS in genomes of analysed *Brachypodium* lineages was further supported by Southern hybridisation. Bioinformatic analyses of the 5S rDNA sequence revealed the presence of a single restriction site for *MseI* in *B. stacei* and both allotetraploid genotypes of *B. hybridum* (Fig. 2E). In contrast, there were two sites for this enzyme in the corresponding sequence of *B. distachyon*. After the Southern hybridisation, a single band of approximately 270 bp was observed in *B. stacei*. In *B. distachyon*, the 5S rDNA probe hybridised to two shorter *MseI* fragments of approximately 240 bp and 130 bp (Fig. 2E and G). Two bands (270 bp and 370 bp) were recovered for *B. hybridum* ABR113 and only a single 240-bp-long band was observed in *B. hybridum* Bhyb26 (Fig. 2E and G). No shorter D-subgenome-specific *MseI* fragments were visualised in any of the *B. hybridum* accessions.

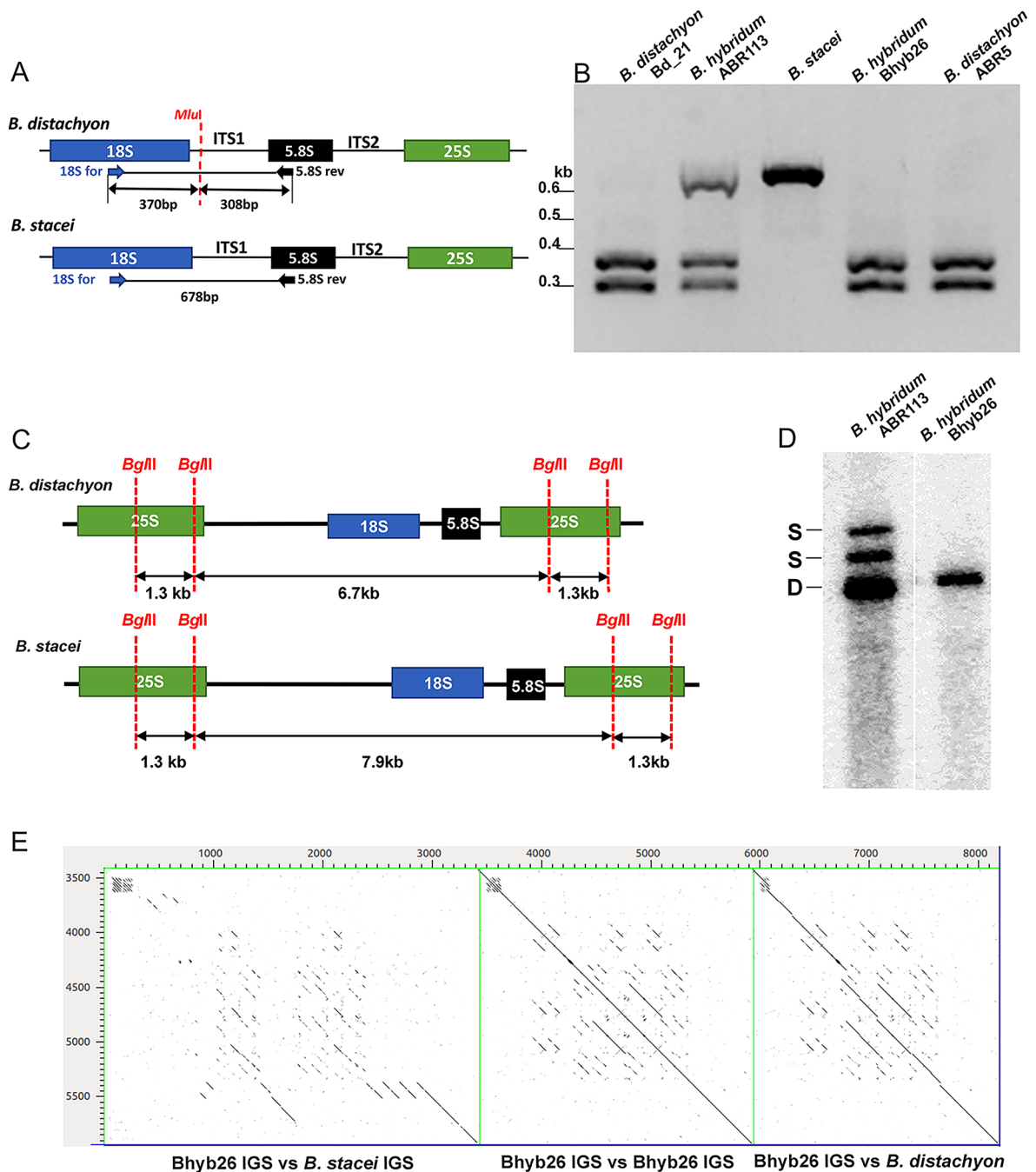
Maximum likelihood analysis was performed using cloned 5S rDNA NTS from both *B. hybridum* lineages and the two diploid *Brachypodium* annuals. The length of the analysed region ranged from 221 to 356 bp, and the final alignment, used for phylogenetic analyses, was 417 bp long (including gaps) with 70 parsimony informative sites. Among the analysed sequences, three main groups of 5S rDNA NTS were revealed (Fig. 2H). The first well-supported clade (Bootstrap Support 97; BS 97; *distachyon*-like) consisted of sequences isolated from *B. distachyon* and *B. hybridum* ABR113 (the sequences representing the longer variants of 5S rDNA NTS; ~ 353 bp). The second clade (BS 82; *stacei*-like) comprised approximately 250-bp-long sequences isolated from *B. stacei* and *B. hybridum* ABR113. The 5S rDNA sequences isolated from *B. hybridum* Bhyb26 were grouped into the third clade (~ 221 bp; BS 100; Fig. 2H). The *distachyon*-like clade could be further divided into two subclades: (i) a subclade

composed of sequences isolated from *B. distachyon* and (ii) sequences isolated from *B. hybridum* ABR113. The *stacei*-like clade could be further divided into four subclades. Two of them consisted of sequences isolated from *B. hybridum* ABR113, and two other subclades included sequences isolated from two different accessions of *B. stacei*.

**35S rDNA structure**

The ITS1 and IGS structures of the 35S rDNA homoeologues were analysed in *B. hybridum* younger ABR113 and older Bhyb26 lineages using bioinformatic and molecular biology methods. The polymorphism in the ITS1 of parental diploids allowed to distinguish between the progenitor ITS1 variants in allopolyploids by the gCAPS approach. A single *MluI* restriction site was found within the ITS1 of the D-subgenome, while there was no *MluI* site in the S-subgenome ITS1 (Fig. 3A). Region containing part of the 3' end of 18S rDNA together with the whole ITS1 was amplified and subjected to digestion with *MluI*. Thus, only the *B. distachyon*-like ITS1 was cut, resulting in two fragments, of approximately 300 and 370 bp, while the *B. stacei* one remained uncut (~ 680 bp). In the younger ABR113, an additive pattern of three expected bands was observed. Interestingly, the older Bhyb26 genotype, showed only the *B. distachyon*-like ITS1 profile (Fig. 3B).

Taking advantage of the length polymorphisms in the progenitor *B. distachyon* and *B. stacei* IGS, it was possible to distinguish between the homoeologous rDNA variants in *B. hybridum* [36]. Bioinformatic analysis showed two restriction sites for *BglII* within the 25S rDNA coding region of the 35S rDNA consensus sequences of D- and S-subgenomes [Figs. 3C and 35]. In the case of the younger ABR113 genotype, three *BglII* fragments were detected: (i) a fast migrating fragment of 6.7 kb corresponding to D-subgenome rDNA, (ii) a fragment of 7.9 kb corresponding to S-subgenome rDNA, and (iii) a 9.2 kb long fragment representing the complete 35S rDNA unit from the S-subgenome in which a single *BglII*



**Fig. 3** Structure of 35S rDNA in the studied *Brachypodium* accessions. **(A)** *MluI* restriction maps of the ITS1 region of the analysed *Brachypodium* accessions. **(B)** The gCAPS analysis of ITS1 sequences amplified from annual *Brachypodium* samples. The fragments were digested with *MluI* (the original gel is shown in Supplementary Figure S3). **(C)** *BglII* restriction maps of the 35S rDNA units of the analysed *Brachypodium* samples. **(D)** Southern blot hybridisation of genomic DNA from allotetraploid *B. hybridum* lineages digested with *BglII* and hybridised with the 25S rDNA probe (the original membrane is shown in Supplementary Figure S4). After hybridisation with 25S rDNA probes, one band (6.7 kb) was shown for the D-subgenome while the S-subgenome was characterised by two bands (9.2 kb and 7.9 kb); a longer one corresponding to the complete repeat (the SNP analysis revealed that one *BglII* restriction site could be mutated in some units; [35]) and a shorter one corresponding to the repeat cut two times with *BglII*. **(E)** Comparison of the 35S rDNA gene intergenic spacer sequences between *B. hybridum* Bhyb26 and diploid *B. stacei* ABR114 or *B. distachyon* Bd21 on dot matrix plots

site was mutated thus not cut. Only one *BglII* fragment of *B. distachyon* origin was observed in the evolutionary older Bhyb26 genotype, showing the elimination of S-subgenome rDNA homoeologues (Fig. 3D).

The IGS from the older *B. hybridum* Bhyb26 genotype was cloned, sequenced and compared with the corresponding regions from *B. stacei* and *B. distachyon*. The length of the IGS in Bhyb26 was 2519 bp. Sequence

comparison has shown 86% identity between the IGS regions from *B. distachyon* and Bhyb26; however, two 51 bp- and 29 bp-long subregions of IGS showed similarities to corresponding fragments in *B. stacei*, and three subregions (120 bp-, 12 bp- and 35 bp-long fragments) were Bhyb26-specific (Figure S5). Sequence analysis with both dot matrix plot and Tandem Repeats Finder revealed that Bhyb26 IGS contained internal repetitive motifs that resemble the *B. distachyon* 35S rDNA IGS, both in the sequence and organisation context (Fig. 3E and Figure S6). It is worth noting that the IGS contained several TATA box elements corresponding to genic and spacer promoters. The putative transcription initiation site immediately downstream of the TATA element showed high identity with its counterparts from diploid *Brachypodium* species. The sequence conservation of the Bhyb26 IGS was in silico verified by Bhyb26 raw Illumina reads mapping to the Bhyb26 cloned IGS, followed by the extraction of the consensus IGS sequence and re-mapping of reads and SNP analysis. Interestingly, only three SNPs were found in the ETS (SNP frequency set as  $\geq 10\%$ , Figure S6).

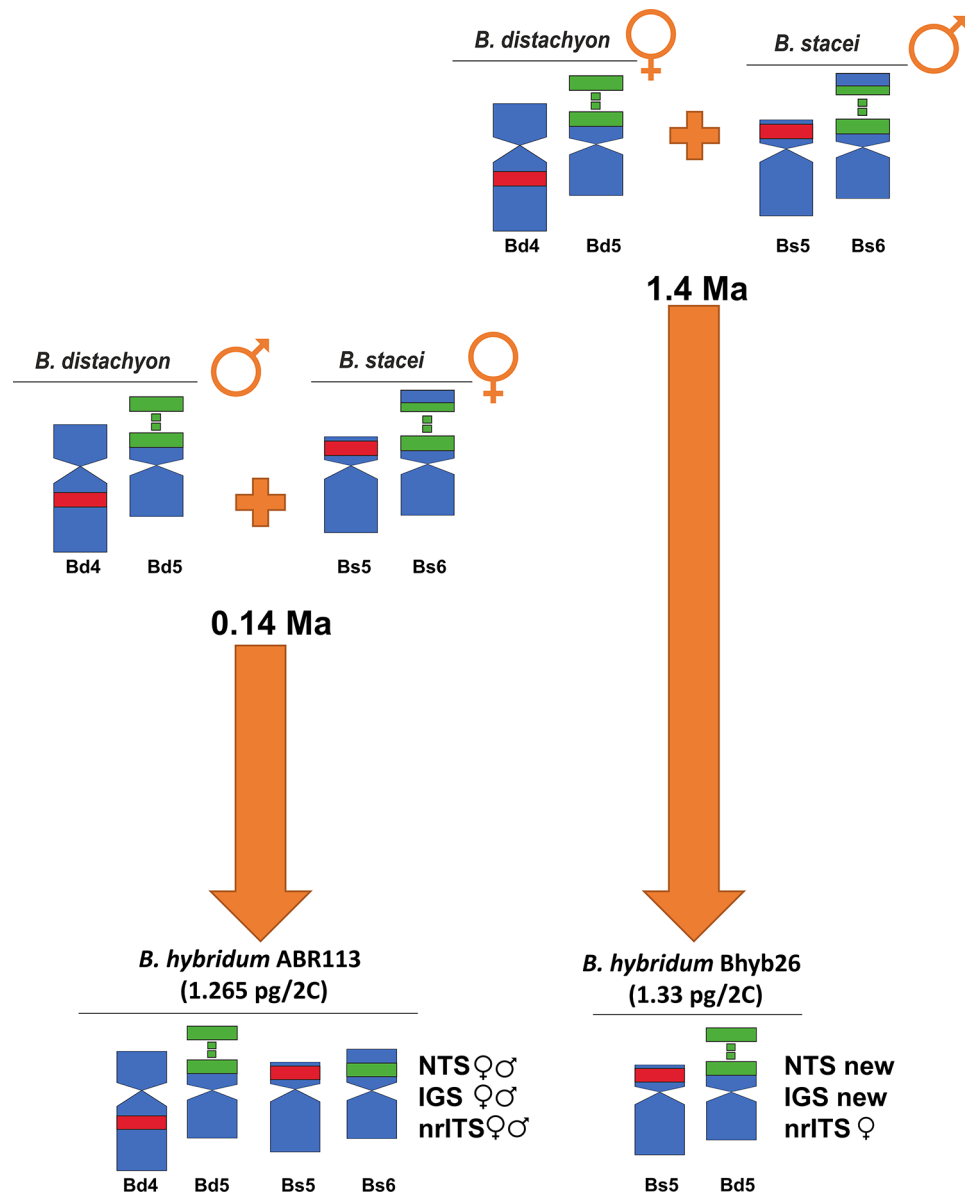
## Discussion

The presence of two distinct evolutionary lineages of *B. hybridum* representing different evolutionary ages constitutes a convenient model system for studying the fate of tandemly repetitive rRNA genes in allopolyploids. We demonstrated retention of both ancestral 35S rDNA homoeologues in the evolutionary younger ABR113 genotype, while the evolutionary older Bhyb26 genotype showed the complete elimination of the S-subgenome 35S rDNA units (Fig. 4). It is necessary to mention that the trend of gradual elimination of *B. stacei*-like rDNA units was already seen in some younger *B. hybridum* lineages [39] while the extent has never been so pronounced as in the case of Bhyb26. Uniparental loss of 35S rDNA repeats is considered a part of the diploidisation process that accompanied the evolution of plant allopolyploids. This phenomenon was also observed in different genera, e.g., *Gossypium* [59], *Triticum* [60], *Primula* [61], *Nicotiana* [62], *Atropa* [63] and *Brassica* [64]. Uniparental elimination of 35S rDNA units is not always connected with the reduction of loci number since the rDNA units of one parental subgenome can be overwritten by the units originating from the other [59]. However, uniparental loss of the entire 35S rDNA locus (or several loci) has been reported in a number of allopolyploids, e.g., *Chenopodium*, *Paspalum* and *Melampodium* [7, 65, 66] likely reflecting diploidisation processes. Well-documented examples of uniparental elimination of either maternal [7, 62] or paternal rDNA loci have been described [67, 68]. A complete elimination of 35S rDNA units from the S-subgenome (paternal parent) was observed in *B.*

*hybridum* Bhyb26. Partial elimination of 35S rDNA units from the S-subgenome was also reported in *B. hybridum* lineages that arose from the opposite crossing direction [39]. In those cases, the 35S rDNA units were partially eliminated from the maternal subgenome. Thus, regardless of the cross direction, the 35S rDNA units were lost from the S-subgenome, indicating that the direction of the cross plays little or no role in the tendency towards rDNA elimination in *B. hybridum*.

Despite the 35S rDNA coding sequences being highly conserved, the ITS and the IGS region evolve faster, resulting in their utility in evolutionary studies [69, 70]. These regions can be species-specific, allowing the distinction of the rDNA units that belong to different subgenomes in allopolyploids since the 35S rDNA IGS interlocus homogenisation is significantly less frequent than intralocus one [71]. This is also reflected in the evolutionary patterns of rDNA in the evolutionary younger *B. hybridum* ABR113 where both the D- and S-subgenome 35S rDNA homoeologues are maintained, and no trace of interlocus homogenisation was reported [35, 36]. In contrast, in the evolutionary older *B. hybridum* Bhyb26 a different scenario has been found, with only one ancestral ITS variant (belonging to the D-subgenome one) and an emerging new IGS family (Figs. 3 and 4). This new IGS variant showed high similarity to the *B. distachyon*-like IGS. There were two motives similar to *B. stacei* and three subregions unique to Bhyb26 only (Figure S5). Our results suggested that the new IGS variant arose by reorganising the *B. distachyon*-like unit coupled with (the first, to our knowledge) rDNA intergenomic (D vs. S) recombination in the hybrid nucleus. Thus, about 1.4 Ma were needed to reconstruct the IGS by intergenomic homogenisation in Bhyb26. These observations are congruent with the hypothesis of rare homologous recombination events perhaps occurring in the nucleolus [72, 73]. Interestingly, *B. stacei* plastome-specific insertions and SNPs were also observed in the Bhyb26 D-plastotype (*B. distachyon*-type plastome) [22]. These results suggest potential S and D genomic recombinations at both nuclear and plastome levels in the ancestral Bhyb26, a finding consistent with the detection of putative heteroplasmy and chloroplast capture events in several *B. distachyon* lineages [22, 74].

5S rDNA NTS is considered a convenient marker of polyploidy in plants mainly due to its negligible level of intergenomic homogenisation [75, 76]. Several reports showed that concerted evolution operates mainly within separate 5S rDNA arrays, with little, if any, exchange between different loci [77, 78]. Usually, the 5S rDNA loci number is additive and therefore, all ancestral variants were present in allopolyploids [7, 66, 79, 80]. A similar additive pattern of 5S rDNA loci was also revealed in ABR113, where both ancestral variants of 5S rDNA NTS



**Fig. 4** Evolution of rDNA loci in two allotetraploid lineages of *B. hybridum*, ABR113 and Bhyb26. Chromosome carrying 35S (green) or 5S (red) rDNA loci and their putative origin (♀ for maternal and ♂ for paternal parent, respectively), and types of ITS, IGS and 5S rDNA non-transcribed spacer (NTS) sequences (♀ for maternal, ♂ for paternal variants and new for a new Bhyb26-specific variant, respectively)

are present (Figs. 2 and 4). However, different evolutionary patterns of ancestral 5S rDNA homoeologues might also exist, where a significant reduction of *B. stacei*-like 5S rDNA NTS was observed in another S-plastotype *B. hybridum* accession (e.g., Bhyb127), as reflected in RepeatExplorer cluster analysis [40]. Thus, the ancestral 5S rDNAs followed different evolutionary scenarios even though both genotypes share the same cross direction and are of comparable evolutionary age [22]. Therefore, it is even more interesting when considering evolutionary older Bhyb26 (*B. distachyon* as maternal parent) where progenitor 5S rDNA types were apparently replaced by a new 5S rDNA variant (Figs. 2E-H and 4).

It is well-known that the intensity of the hybridisation signal on the chromosome is, to some extent, linked with the rDNA unit number [81]. Taking into account different FISH experiments with 5S rDNA as a probe on various *B. hybridum* genotypes, it can be assumed that at least three distinct 5S rDNA hybridisation signal patterns were observed: (i) the signals of comparable intensity in both ancestral rDNA-bearing D and S chromosomes (e.g., *B. hybridum* ABR107) [35]; (ii) the FISH signals corresponding to 5S rDNA of higher intensity in the S-subgenome locus than the D-subgenome one (e.g., *B. hybridum* ABR113 and *B. hybridum* 3-4-2; new name: Bhyb3\_4\_2) [35, 41]; and (iii) the 5S rDNA hybridisation



signals of higher intensity in the D-subgenome than the S-subgenome one (e.g., *B. hybridum* 19-13-2; new name: Bhyb19\_13\_2) [35]. The differences in copy number in locus may suggest a gradual elimination of one parental variant from *B. hybridum*, which could finally lead to the elimination of the one whole ancestral 5S rDNA locus. Two different pathways of 5S rDNA diploidisation have been described up to date; one scenario involves the elimination of one entire parental locus (e.g., *Melampodium strigosum*) [66] while the second scenario implies that one ancestral variant may be overwritten by another (*Anemone baldensis*) [67, 82]. However, such an inter-locus concerted evolution of 5S rDNA loci in allopolyploids seems rare since it was described only in *Anemone baldensis*. In this allohexaploid, the D-subgenome variant was replaced by the B-subgenome variant, although the 5S rDNA loci number did not change [67, 82].

Much effort has been put into deciphering the global evolutionary patterns of the ancestral subgenomes in the polyploid *B. hybridum* as it serves as a convenient model for understanding the evolutionary fates of allopolyploids. Molecular cytogenetic data [28] and the analyses of the whole genome sequences from multiple genotypes [22, 26, 27] have pointed out a lack of significant genomic rearrangements accompanying hybridisation and WGD in this species. In *Brachypodium* genomes no significant, large genomic changes have occurred during its evolution; on the contrary, there has probably been a slow, gradual process of genomic change across evolutionary time. However, rDNA sequences represent more dynamic genome components than single-copy sequences and transposable elements [22, 26, 27, 35]. Such massive changes in rDNA sequences were also described in *Arabidopsis suecica* [83], in which, as in *B. hybridum*, relative stasis in other genome components was shown. Therefore, an interesting question arises whether tandem repetitive sequences other than rDNA display such dynamic evolutionary pathways.

In conclusion, rRNA genes in *B. hybridum* demonstrate a pronounced tendency towards diploidisation, with distinct patterns of rDNA elimination observed between younger and older lineages. Notably, the 35S and 5S rDNA loci exhibit differential behaviour across subgenomes, with preferential loss of 35S rDNA from the S-subgenome and 5S rDNA from the D-subgenome in the older lineage. These results suggest that 5S and 35S rRNA genes are likely expressed from separate subgenomes in Bhyb26, highlighting the complex regulatory dynamics governing rRNA gene expression in polyploid genomes.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12870-024-05658-5>.

Supplementary Material 1  
Supplementary Material 2  
Supplementary Material 3  
Supplementary Material 4  
Supplementary Material 5  
Supplementary Material 6  
Supplementary Material 7  
Supplementary Material 8

## Acknowledgements

Computational resources for RepeatExplorer analysis were provided by the ELIXIR-CZ project (LM2015047), part of the international ELIXIR infrastructure. Genome data were generated by the Joint Genome Institute (DOE, USA) Community Science Program project 503504.

## Author contributions

B. K. conceived and designed the research framework. D. T., N. B-Z., S. M., K.X., Y. Z., R.S., M. R-J., K.W. and A.K. collected and analysed the data. D. T., N. B-Z. and B. K. wrote the original draft manuscript. D. T., N. B-Z., S. M., R.S., S.G., K.W., P. C., A.K., R.H., and B.K. revised and edited the final manuscript. All authors have read and agreed to the published version of the manuscript.

## Funding

This research was funded by the National Science Centre Poland (project number 2020/39/O/NZ8/00184) to B.K. and the Research Excellence Initiative of the University of Silesia in Katowice. P.C. and R.S. were funded by the Spanish Ministry of Science and Innovation grants TED2021-131073B-I00, PDC2022-133712-I00 and PID2022-140074NB-I00. The work was partially supported by the Czech Science Foundation (grant no. 22-16826S).

## Data availability

Sequence data that support the findings of this study have been deposited in the NCBI with the primary accession code: PP339786, PP339788, PP339790, PP339792, PP339787, PP339789, PP339791, PP339793, PP339793, PP339812 - PP339815, PP339808 - PP339811, PP339782 - PP339785, PP339794 - PP339807, PP317522, PP339781, PP339816 - PP339825.

## Declarations

### Ethics approval and consent to participate

Not applicable.

### Consent for publication

Not applicable.

### Competing interests

The authors declare no competing interests.

### Author details

<sup>1</sup>Plant Cytogenetics and Molecular Biology Group, Institute of Biology, Biotechnology and Environmental Protection, Faculty of Natural Sciences, University of Silesia in Katowice, Katowice 40-032, Poland

<sup>2</sup>School of Life Sciences, Nantong University, Nantong, Jiangsu 226019, China

<sup>3</sup>Department of Agricultural and Environmental Sciences, High Polytechnic School of Huesca, University of Zaragoza, Huesca 22071, Spain

<sup>4</sup>Institut Botànic de Barcelona IBB (CSIC-CMNCB), Barcelona, Catalonia 08038, Spain

<sup>5</sup>Department of Molecular Epigenetics, Institute of Biophysics, Czech Academy of Sciences, Brno CZ- 61200, Czech Republic

Received: 15 April 2024 / Accepted: 1 October 2024

Published online: 18 October 2024

## References

- Akagi T, Jung K, Masuda K, Shimizu KK. Polyploidy before and after domestication of crop species. *Curr Opin Plant Biol*. 2022;69:102255.
- Van de Peer Y, Mizrahi E, Marchal K. The evolutionary significance of polyploidy. *Nat Rev Genet*. 2017;18(7):411–24.
- Project AG. The *Arabidopsis* genome and the evolution of flowering plants. *Science*. 2013;342(6165):1241089.
- Van de Peer Y, Ashman TL, Soltis PS, Soltis DE. Polyploidy: an evolutionary and ecological force in stressful times. *Plant Cell*. 2021;33(1):11–26.
- Hollister JD, Arnold BJ, Svedin E, Xue KS, Dilkes BP, Bomblies K. Genetic adaptation associated with genome-doubling in autotetraploid *Arabidopsis arenosa*. *PLoS Genet*. 2012;8(12):e1003093.
- Feldman M, Levy AA. Genome evolution in allopolyploid wheat—a revolutionary reprogramming followed by gradual changes. *J Genet Genomics*. 2009;36(9):511–8.
- Kolano B, McCann J, Orzechowska M, Siwinska D, Tlemsch E, Weiss-Schnee-weiss H. Molecular and cytogenetic evidence for an allotetraploid origin of *Chenopodium quinoa* and *C. berlandieri* (Amaranthaceae). *Mol Phylogenet Evol*. 2016;100:109–23.
- Dillenberger MS, Wei N, Tennesen JA, Ashman T-L, Liston A. Plastid genomes reveal recurrent formation of allopolyploid *Fragaria*. *Am J Bot*. 2018;105(5):862–74.
- Jang T-S, Parker JS, Emadzade K, Tlemsch EM, Leitch AR, Weiss-Schnee-weiss H. Multiple origins and nested cycles of hybridization result in high tetraploid diversity in the monocot *Prosopis*. *Front Plant Sci* 2018;9.
- Matyášek R, Kuderová A, Kutlíková E, Kučera M, Kovařík A. Intragenomic heterogeneity of intergenic ribosomal DNA spacers in *Cucurbita moschata* is determined by DNA minisatellites with variable potential to form non-canonical DNA conformations. *DNA Res*. 2019;26(3):273–86.
- Mavrodiev E, Chester M, Suárez-Santiago V, Visger C, Rodriguez R, Susanna A, Baldini R, Soltis P, Soltis D. Multiple origins and chromosomal novelty in the allotetraploid *Tragopogon castellanus* (Asteraceae). *New Phytol*. 2015;206(3):1172–83.
- Soltis DE, Buggs RJ, Barbazuk WB, Schnable PS, Soltis PS. On the origins of species: does evolution repeat itself in polyploid populations of independent origin? *Cold Spring Harb Symp Quant Biol*. 2009;74:215–23.
- Chen D, Yan P-C, Guo Y-P. Imprints of independent allopolyploid formations on patterns of gene expression in two sibling yarrow species (*Achillea*, Asteraceae). *BMC Genomics*. 2021;22(1):264.
- Chen ZJ, Sreedasyam A, Ando A, Song Q, De Santiago LM, Hulse-Kemp AM, Ding M, Ye W, Kirkbride RC, Jenkins J, et al. Genomic diversifications of five *Gossypium* allopolyploid species and their impact on cotton improvement. *Nat Genet*. 2020;52(5):525–33.
- Levy AA, Feldman M. The impact of polyploidy on grass genome evolution. *Plant Physiol*. 2002;130(4):1587–93.
- Li Z, McKibben MTW, Finch GS, Blischak PD, Sutherland BL, Barker MS. Patterns and processes of diploidization in land plants. *Annu Rev Plant Biol*. 2021;72(1):387–410.
- Lim KY, Soltis DE, Soltis PS, Tate J, Matyasek R, Srubarova H, Kovarik A, Pires JC, Xiong Z, Leitch AR. Rapid chromosome evolution in recently formed polyploids in *Tragopogon* (Asteraceae). *PLoS ONE*. 2008;3(10):e3353.
- Ma XF, Gustafson JP. Genome evolution of allopolyploids: a process of cytological and genetic diploidization. *Cytogenet Genome Res*. 2005;109(1–3):236–49.
- Renny-Byfield S, Kovarik A, Kelly LJ, Macas J, Novak P, Chase MW, Nichols RA, Pancholi MR, Grandbastien MA, Leitch AR. 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.
- Soltis PS, Marchant DB, Van de Peer Y, Soltis DE. Polyploidy and genome evolution in plants. *Curr Opin Genet Dev*. 2015;35:119–25.
- Catalán P, Müller J, Hasterok R, Jenkins G, Mur LAJ, Langdon T, Betekhtin A, Siwinska D, Pimentel M, López-Alvarez D. Evolution and taxonomic split of the model grass *Brachypodium distachyon*. *Ann Bot*. 2012;109(2):385–405.
- Gordon SP, Contreras-Moreira B, Levy JJ, Djamei A, Czedik-Eysenberg A, Tartaglio VS, Session A, Martin J, Cartwright A, Katz A et al. Gradual polyploid genome evolution revealed by pan-genomic analysis of *Brachypodium hybridum* and its diploid progenitors. *Nat Commun*. 2020;11(1).
- López-Alvarez D, López-Herranz ML, Betekhtin A, Catalán P. A DNA barcoding method to discriminate between the model plant *Brachypodium distachyon* and its close relatives *B. stacei* and *B. hybridum* (Poaceae). *PLoS ONE*. 2012;7(12):e51058.
- Scholtzof K-BG, Irigoyen S, Catalan P, Mandadi KK. *Brachypodium*: a monocot grass model genus for plant biology. *Plant Cell*. 2018;30(8):1673–94.
- Hasterok R, Catalan P, Hazen SP, Roulin AC, Vogel JP, Wang K, Mur LAJ. *Brachypodium*: 20 years as a grass biology model system; the way forward? *Trends Plant Sci*. 2022;27(10):1002–16.
- Mu W, Li K, Yang Y, Breiman A, Yang J, Wu Y, et al. Subgenomic stability of progenitor genomes during repeated allotetraploid origins of the same grass *Brachypodium hybridum*. *Mol Biol Evol*. 2023;40(12).
- Scarlett VT, Lovell JT, Shao M, Phillips J, Shu S, Lusinska J, Goodstein DM, Jenkins J, Grimwood J, Barry K et al. Multiple origins, one evolutionary trajectory: gradual evolution characterizes distinct lineages of allotetraploid *Brachypodium*. *Genetics* 2023;223(2).
- Lusinska J, Majka J, Betekhtin A, Susek K, Wolny E, Hasterok R. Chromosome identification and reconstruction of evolutionary rearrangements in *Brachypodium distachyon*, *B. stacei* and *B. hybridum*. *Ann Bot*. 2018;122(3):445–59.
- Alger EI, Edger PP. One subgenome to rule them all: underlying mechanisms of subgenome dominance. *Curr Opin Plant Biol*. 2020;54:108–13.
- Borowska-Zuchowska N, Mykhailik S, Robaszkiewicz E, Matysiak N, Mielanczyk L, Wojnicz R, Kovarik A, Hasterok R. Switch them off or not: selective rRNA gene repression in grasses. *Trends Plant Sci*. 2023;28(6):661–72.
- Volkov R, Medina F, Zentgraf U, Hemleben V. Organization and molecular evolution of rDNA nucleolar dominance and nucleolus structure. In: Esser K, Lutttge U, Beyschlag W, Murata J, editors. *Progress in botany*. Vol 65. Berlin Heidelberg New York: Springer; 2004.
- Roa F, Guerra M. Non-random distribution of 5S rDNA sites and its association with 45S rDNA in plant chromosomes. *Cytogenet Genome Res*. 2015;146(3):243–9.
- García S, Kovařík A. Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. *Heredity*. 2013;111(1):23–33.
- García S, Lim KY, Chester M, Garnatje T, Pellicer J, Vallès J, Leitch AR, Kovařík A. Linkage of 35S and 5S rRNA genes in *Artemisia* (family Asteraceae): first evidence from angiosperms. *Chromosoma*. 2009;118(1):85–97.
- Borowska-Zuchowska N, Kovarik A, Robaszkiewicz E, Tuna M, Tuna GS, Gordon S, Vogel JP, Hasterok R. The fate of 35S rRNA genes in the allotetraploid grass *Brachypodium Hybridum*. *Plant J*. 2020;103(5):1810–25.
- Borowska-Zuchowska N, Kwasniewski M, Hasterok R. Cytomolecular analysis of ribosomal DNA evolution in a natural allotetraploid *Brachypodium hybridum* and its putative ancestors—dissecting complex repetitive structure of intergenic spacers. *Front Plant Sci* 2016;7.
- Borowska-Zuchowska N, Robaszkiewicz E, Wolny E, Betekhtin A, Hasterok R. Ribosomal DNA loci derived from *Brachypodium stacei* are switched off for major parts of the life cycle of *Brachypodium hybridum*. *J Exp Bot*. 2019;70(3):805–15.
- Idziak D, Hasterok R. Cytogenetic evidence of nucleolar dominance in allotetraploid species of *Brachypodium*. *Genome*. 2008;51(5):387–91.
- Borowska-Zuchowska N, Robaszkiewicz E, Mykhailik S, Martini J, Pinski A, Kovarik A, et al. To be or not to be expressed: the first evidence of a nucleolar dominance tissue-specificity in *Brachypodium hybridum*. *Front Plant Sci*. 2021;12.
- García S, Wendel JF, Borowska-Zuchowska N, Ainouche M, Kuderova A, Kovarik A. The utility of graph clustering of 5S ribosomal DNA homoeologs in plant allopolyploids, homoploid hybrids, and cryptic introgressants. *Front Plant Sci*. 2020;11.
- Hasterok R, Draper J, Jenkins G. Laying the cytotoxic foundations of a new model grass, *Brachypodium distachyon* (L.) Beauv. *Chromosome Res*. 2004;12(4):397–403.
- Novák P, Neumann P, Macas J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat Protoc*. 2020;15(11):3745–76.
- 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. *Bioinformatics*. 2013;29(6):792–3.
- Doyle J. DNA protocols for plants. In: Hewitt GM, Johnston AWB, Young JPW, editors. *Molecular techniques in Taxonomy*. Berlin, Heidelberg: Springer Berlin Heidelberg; 1991. pp. 283–93.
- Venora G, Blangiforti S, Frediani M, Maggini F, Gelati M, Castiglione MR, Cremonini R. Nuclear DNA contents, rDNAs, chromatin organization, and karyotype evolution in *Vicia* sect. faba. *Protoplasma*. 2000;213:118–25.

46. Kolano B, Siwinska D, McCann J, Weiss-Schneeweiss H. The evolution of genome size and rDNA in diploid species of *Chenopodium* s.l. (Amaranthaceae). *Bot J Linn Soc*. 2015;179(2):218–35.
47. Chang KD, Fang SA, Chang FC, Chung MC. Chromosomal conservation and sequence diversity of ribosomal RNA genes of two distant *Oryza* species. *Genomics*. 2010;96(3):181–90.
48. Maughan PJ, Kolano BA, Maluszynska J, Coles ND, Bonifacio A, Rojas J, Coleman CE, Stevens MR, Fairbanks DJ, Parkinson SE, et al. Molecular and cytological characterization of ribosomal RNA genes in *Chenopodium quinoa* and *Chenopodium berlandieri*. *Genome*. 2006;49(7):825–39.
49. Benson G. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res*. 1999;27(2):573–80.
50. Sonnhhammer EL, Durbin R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene*. 1995;167(1–2):1–10.
51. Löytynoja A, Goldman N. webPRANK: a phylogeny-aware multiple sequence aligner with interactive alignment browser. *BMC Bioinformatics*. 2010;11(1):579.
52. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. *Mol Biol Evol*. 2017;35(2):518–22.
53. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, von Haeseler A, Lanfear R. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol*. 2020;37(5):1530–4.
54. Gerlach WL, Dyer TA. Sequence organization of the repeating units in the nucleus of wheat which contain 5S rRNA genes. *Nucleic Acids Res*. 1980;8(21):4851–65.
55. Kovarik A, Pires JC, Leitch AR, Lim KY, Sherwood AM, Matyasek R, Rocca J, Soltis DE, Soltis PS. Rapid concerted evolution of nuclear ribosomal DNA in two *Tragopogon* allopolyploids of recent and recurrent origin. *Genetics*. 2005;169(2):931–44.
56. Unfried I, Gruendler P. Nucleotide sequence of the 5.8S and 25S rRNA genes and of the internal transcribed spacers from *Arabidopsis thaliana*. *Nucleic Acids Res*. 1990;18(13):4011.
57. Doležel J, Doleželová M, Novák FJ. Flow cytometric estimation of nuclear DNA amount in diploid bananas (*Musa acuminata* and *M. balbisiana*). *Biol Plant*. 1994;36(3):351–7.
58. Doležel J, Bartoš J, Voglmayr H, Greilhuber J. Letter to the editor. *Cytometry Part A*. 2003;51A(2):127–8.
59. Wendel JF, Schnabel A, Seelanan T. Bidirectional interlocus concerted evolution following allopolyploid speciation in cotton (*Gossypium*). *Proc Natl Acad Sci*. 1995;92(1):280–4.
60. Guo X, Han F. Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. *Plant Cell*. 2014;26(11):4311–27.
61. Guggisberg A, Baroux C, Grossniklaus U, Conti E. Genomic origin and organization of the allopolyploid *Primula egaliksensis* investigated by *in situ* hybridization. *Ann Botany*. 2008;101(7):919–27.
62. Kovarik A, Matyasek R, Lim KY, Skalická K, Koukalová B, Knapp S, Chase M, Leitch AR. Concerted evolution of 18–5.8–26S rDNA repeats in *Nicotiana* allotetraploids. *Biol J Linn Soc Lond*. 2004;82(4):615–25.
63. Volkov RA, Panchuk II, Borisjuk NV, Hosiawa-Baranska M, Maluszynska J, Hemleben V. Evolutional dynamics of 45S and 5S ribosomal DNA in ancient allohexaploid *Atropa belladonna*. *BMC Plant Biol*. 2017;17(1):21.
64. Sochorová J, Coriton O, Kuderová A, Lunerová J, Chèvre AM, Kovařík A. Gene conversion events and variable degree of homogenization of rDNA loci in cultivars of *Brassica napus*. *Ann Bot*. 2017;119(1):13–26.
65. Vaio M, Mazzella C, Guerra M, Speranza P. Effects of the diploidisation process upon the 5S and 35S rDNA sequences in the allopolyploid species of the Dilatata group of *Paspalum* (Poaceae, Paniceae). *Aust J Bot*. 2019;67(7):521–30.
66. Weiss-Schneeweiss H, Blösch C, Turner B, Villaseñor JL, Stuessy TF, Schneeweiss GM. The promiscuous and the chaste: frequent allopolyploid speciation and its genomic consequences in American daisies (*Melampodium* sect. *Melampodium*; Asteraceae). *Evolution*. 2012;66(1):211–28.
67. Mlinarec J, Šatović Z, Malenica N, Ivančić-Baće I, Besendorfer V. Evolution of the tetraploid *Anemone multifida* (2n = 32) and hexaploid *A. baldensis* (2n = 48) (Ranunculaceae) was accompanied by rDNA loci loss and intergenomic translocation: evidence for their common genome origin. *Ann Bot*. 2012;110(3):703–12.
68. Zozomová-Lihová J, Mandáková T, Kovaříková A, Mühlhausen A, Mummenhoff K, Lysak MA, et al. When fathers are instant losers: homogenization of rDNA loci in recently formed *Cardamine x schulzii* trigeneric allopolyploid. *New Phytol*. 2014;203(4):1096–108.
69. Feliner GN, Rossello JA. Better the devil you know? Guidelines for insightful utilization of nrDNA ITS in species-level evolutionary studies in plants. *Mol Phylogenet Evol*. 2007;44(2):911–9.
70. Hu X, Yu F, Huang Y, Sun L, Li X, Yang S, Chen K, Huang F, Zeng K, Zhang M, et al. Characterization analysis of the 35S rDNA intergenic spacers in *Erianthus Arundinaceus*. *Gene*. 2019;694:63–70.
71. Wang W, Zhang X, Garcia S, Leitch AR, Kovařík A. Intragenomic rDNA variation - the product of concerted evolution, mutation, or something in between? *Heredity*. 2023; 131(3):179–88.
72. Sims J, Rabanal FA, Elgert C, von Haeseler A, Schlöglhofer P. It is just a matter of time: balancing homologous recombination and non-homologous end joining at the rDNA locus during meiosis. *Front Plant Sci*. 2021;12.
73. Kovarik A, Dadejova M, Lim YK, Chase MW, Clarkson JJ, Knapp S, Leitch AR. Evolution of rDNA in *Nicotiana* allopolyploids: a potential link between rDNA homogenization and epigenetics. *Ann Bot*. 2008;101(6):815–23.
74. Sancho R, Cantalapiedra CP, López-Alvarez D, Gordon SP, Vogel JP, Catalán P, Contreras-Moreira B. Comparative plastome genomics and phylogenomics of *Brachypodium*: flowering time signatures, introgression and recombination in recently diverged ecotypes. *New Phytol*. 2018;218(4):1631–44.
75. Baum BR, Johnson DA. Molecular confirmation of the genomic constitution of *Douglasdeweya* (Triticeae: Poaceae): demonstration of the utility of the 5S rDNA sequence as a tool for haplome identification. *Mol Genet Genomics*. 2008;279(6):621–8.
76. Mahelka V, Kopecký D, Baum BR. Contrasting patterns of evolution of 45S and 5S rDNA families uncover new aspects in the genome constitution of the agronomically important grass *Thinopyrum intermedium* (Triticeae). *Mol Biol Evol*. 2013;30(9):2065–86.
77. Pinhal D, Yoshimura TS, Araki CS, Martins C. The 5S rDNA family evolves through concerted and birth-and-death evolution in fish genomes: an example from freshwater stingrays. *BMC Evol Biol*. 2011;11(1):151.
78. Yang CR, Baum BR, Johnson DA, Zhang HQ, Zhou YH. Molecular diversity of the 5S nuclear ribosomal DNA in *Campeiostrachys* with StHY haplome constitution. *J Syst Evol*. 2020;58(1):69–76.
79. Kolano B, McCann JAMIE, Oskędra M, Chrapek M, Rojek M, Nobis A, Weiss-Schneeweiss H. Parental origin and genome evolution of several eurasian hexaploid species of *Chenopodium* (Chenopodiaceae). *Phytotaxa*. 2019;392(3):163–85.
80. Tynkevich YO, Shelyfist AY, Kozub LV, Hemleben V, Panchuk II, Volkov RA. 5S ribosomal DNA of Genus *Solanum*: molecular organization, evolution, and taxonomy. *Front Plant Sci*. 2022;13.
81. Maluszynska J, Heslop-Harrison JS. Physical mapping of rDNA loci in *Brassica* species. *Genome*. 1993;36(4):774–81.
82. Mlinarec J, Franjević D, Bočkor L, Besendorfer V. Diverse evolutionary pathways shaped 5S rDNA of species of tribe Anemoneae (Ranunculaceae) and reveal phylogenetic signal. *Bot J Linn Soc*. 2016;182(1):80–99.
83. Burns R, Mandáková T, Gunis J, Soto-Jiménez LM, Liu C, Lysak MA, et al. Gradual evolution of allopolyploidy in *Arabidopsis suecica*. *Nat Ecol Evol*. 2021;5(10):1367–81.

## Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Katowice,  
dnia.....10.02.2025.....

mgr Dana Trunova  
imię i nazwisko kandydata

ul. Paderewskiego 32/103, 40-282, Katowice  
adres korespondencyjny

+48798444456

telefon kontaktowy

dana.trunova@us.edu.pl

e-mail

mgr Dana Trunova

First and last name of co-author of the publication

Uniwersytet Śląski w Katowicach


Affiliation

#### A STATEMENT OF THE APPLICANT'S AUTHOR OF THEIR CONTRIBUTION TO THE WORK

I declare that for the following work:

Trunova D., Borowska-Zuchowska N., Mykhailik S., Xia K., Zhu Y., Sancho R., Rojek-Jelonek M., Garcia S., Wang K., Catalan P., Kovarik A., Hasterok R., Kolano B.; Does time matter? Intraspecific diversity of ribosomal RNA genes in lineages of the allopolyploid model grass *Brachypodium hybridum* with different evolutionary ages; BMC Plant Biology; 2024; 24: 981

My participation involved cultivation of plants, performance of laboratory analyses such as: performance and analyses of results obtained with RepeatExplorer2&TAREAN software, DNA isolation, amplification of the rDNA sequences, preparation of cytogenetic preparations of meristematic tissues, preparation of probes for fluorescence *in situ* hybridization (FISH), performance of fluorescence *in situ* hybridization, photographic documentation of cytogenetic preparations, performance of 5S rDNA Southern Blot hybridization, performance of 35S rDNA gCAPs analysis, performance of phylogenetic analyses, participation in the results interpretation, writing the manuscript and preparing figures.

  
Signature of the author of the publication

A STATEMENT OF THE APPLICANT'S CO-AUTHOR OF THEIR CONTRIBUTION TO THE WORK

Katowice, dnia 10/06/2025

dr Natalia Borowska-Zuchowska  
First and last name of co-author of the publication

Uniwersytet Śląski w Katowicach  
Affiliation

**STATEMENT**

I declare that for the following work:

Trunova D., Borowska-Zuchowska N., Mykhailik S., Xia K., Zhu Y., Sancho R., Rojek-Jelonek M., Garcia S., Wang K., Catalan P., Kovarik A., Hasterok R., Kolano B.; Does time matter? Intraspecific diversity of ribosomal RNA genes in lineages of the allopolyploid model grass *Brachypodium hybridum* with different evolutionary ages; BMC Plant Biology; 2024; 24: 981

My participation consisted of planning and supervising the research, interpreting the results, writing and editing of the manuscript, and preparation of figures.

Borowska-Zuchowska N.

Signature of the co-author of the publication



A STATEMENT OF THE APPLICANT'S CO-AUTHOR OF THEIR CONTRIBUTION TO THE WORK

Katowice, dnia...10.06.2025...

dr hab. Bożena Kolano, prof. UŚ  
First and last name of co-author of the publication

Uniwersytet Śląski w Katowicach  
Affiliation

**STATEMENT**

I declare that for the following work:

Trunova D., Borowska-Zuchowska N., Mykhailik S., Xia K., Zhu Y., Sancho R., Rojek-Jelonek M., Garcia S., Wang K., Catalan P., Kovarik A., Hasterok R., Kolano B.; Does time matter? Intraspecific diversity of ribosomal RNA genes in lineages of the allopolyploid model grass *Brachypodium hybridum* with different evolutionary ages; BMC Plant Biology; 2024; 24: 981

My contribution included obtaining project funding, planning and supervising the research, interpreting the results, writing and editing of the manuscript, preparation of figures. I am the corresponding author.

Bożena Kolano.....

Signature of the co-author of the publication

A STATEMENT OF THE APPLICANT'S CO-AUTHOR OF THEIR CONTRIBUTION TO THE WORK

Katowice, dnia 11.06.25

dr Serhii Mykhailyk  
First and last name of co-author of the publication

Uniwersytet Śląski w Katowicach  
Affiliation

**STATEMENT**

I declare that for the following work:

Trunova D., Borowska-Zuchowska N., Mykhailyk S., Xia K., Zhu Y., Sancho R., Rojek-Jelonek M., Garcia S., Wang K., Catalan P., Kovarik A., Hasterok R., Kolano B.; Does time matter? Intraspecific diversity of ribosomal RNA genes in lineages of the allopolyploid model grass *Brachypodium hybridum* with different evolutionary ages; BMC Plant Biology; 2024; 24: 981

My participation involved performance of 35S rDNA Southern Blot hybridization, participation in writing the manuscript and preparing figures.



Signature of the co-author of the publication

Załącznik nr 10  
do pisma okólnego nr 2  
Prorektora ds. nauki i finansów  
z dnia 19 lutego 2024 r.

## A STATEMENT OF THE APPLICANT'S CO-AUTHOR OF THEIR CONTRIBUTION TO THE WORK

Katowice, 11.06.2025

Robert Hasterok  
First and last name of co-author of the publication

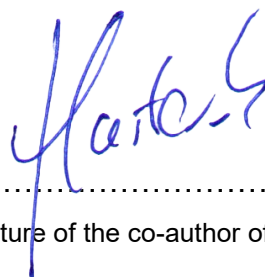
University of Silesia in Katowice, Poland  
Affiliation

### STATEMENT

I declare that for the following work:

Trunova D., Borowska-Zuchowska N., Mykhailiyk S., Xia K., Zhu Y., Sancho R., Rojek-Jelonek M., Garcia S., Wang K., Catalan P., Kovarik A., Hasterok R., Kolano B.; Does time matter? Intraspecific diversity of ribosomal RNA genes in lineages of the allopolyploid model grass *Brachypodium hybridum* with different evolutionary ages; BMC Plant Biology; 2024; 24: 981

My contribution included involvement in the revision and editing of the final manuscript.



.....  
Signature of the co-author of the publication