

# MASTERARBEIT | MASTER'S THESIS

Titel | Title

Evolution of the repeatome of several diploid and polyploid  
*Prospero autumnale* (Hyacinthaceae) genomes

verfasst von | submitted by  
Hilal Aktas BSc

angestrebter akademischer Grad | in partial fulfilment of the requirements for the degree of  
Master of Science (MSc)

Wien | Vienna, 2024

Studienkennzahl lt. Studienblatt | Degree  
programme code as it appears on the  
student record sheet:

UA 066 877

Studienrichtung lt. Studienblatt | Degree  
programme as it appears on the student  
record sheet:

Masterstudium Genetik und Entwicklungsbiologie

Betreut von | Supervisor:

Univ.-Prof. Mag. Dr. Hanna Schneeweiss Privatdoz.

# Inhaltsverzeichnis

<b>1. Introduction</b>	<b>3</b>
1.1. <i>Genome Evolution</i>	3
1.2. <i>Repetitive DNA</i>	4
1.2.1. Transposable elements	4
1.2.2. Satellite DNA	6
1.2.3. Ribosomal DNA	7
1.3. <i>Prospero autumnale complex</i>	8
1.3.1. Diploid cytotypes	9
1.3.2. Polyploid cytotypes	12
1.4. <i>Aim of the study</i>	15
<b>2. Material and Methods</b>	<b>17</b>
2.1. <i>Plant Material</i>	17
2.2. <i>Pre-treatment and fixation of chromosomes</i>	17
2.3. <i>Feulgen staining for karyotype analysis</i>	17
2.4. <i>Fluorescence in situ hybridization (FISH)</i>	18
2.5. <i>Genomic in situ hybridization (GISH)</i>	20
2.6. <i>Genome size estimation using flow cytometry</i>	22
2.7. <i>Analysis of repetitive DNA using RepeatExplorer2</i>	23
<b>3. Results</b>	<b>27</b>
3.1. <i>Chromosome number and karyotype analysis</i>	27
3.2. <i>Localization of satellite DNA and parental genome identification</i>	29
3.2.1. B <sup>6</sup> B <sup>6</sup> cytotype	29
3.2.2. B <sup>7</sup> B <sup>7</sup> cytotype	31
3.2.3. B <sup>6</sup> B <sup>7</sup> hybrid	31
3.2.4. B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group I	32
3.2.5. B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group II	34
3.2.6. B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group III	35
3.2.7. B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group IV	35
3.3. <i>Genome size estimation using flow cytometry</i>	38
3.4. <i>Analysis of repetitive DNA using RepeatExplorer2</i>	39
3.4.1. <i>Prospero autumnale</i> cytotype B <sup>6</sup> B <sup>6</sup>	39
3.4.2. <i>Prospero autumnale</i> cytotype B <sup>7</sup> B <sup>7</sup>	43
3.4.3. <i>Prospero autumnale</i> allotetraploid B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup>	44
3.5. <i>Comparative analysis of the repeatome using Repeat Explorer 2</i>	49
3.5.1. Comparative analysis of diploid cytotypes B <sup>6</sup> B <sup>6</sup> and B <sup>7</sup> B <sup>7</sup> of <i>Prospero autumnale</i>	49
3.5.2. Comparative analysis diploid cytotypes B <sup>6</sup> B <sup>6</sup> and B <sup>7</sup> B <sup>7</sup> , and allotetraploid B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> of <i>Prospero autumnale</i>	54
<b>4. Discussion</b>	<b>59</b>
4.1. <i>Karyotype analysis and genome size evolution</i>	59
4.2. <i>Repeatome analysis of the Prospero autumnale complex and identification of novel satellite DNAs</i>	60

4.3.	<i>Localization of satellite DNAs</i> .....	61
5.	<b>References</b> .....	65
6.	<b>Abstract</b> .....	70
7.	<b>Zusammenfassung</b> .....	71

# 1. Introduction

## 1.1. Genome Evolution

Comparing various features of the karyotype among closely related species or within a species complex, such as the *Prospero autumnale* complex, enables the assessment of genome evolution mechanisms that contribute to plant diversification and speciation (Schubert, 2007; Jang et al., 2013; Weiss-Schneeweiss & Schneeweiss, 2013). Key features of the karyotype that may be influenced during genome evolution include the number and size of chromosomes, genome size, the positioning of primary and secondary constrictions and more (Schubert, 2007; Weiss-Schneeweiss & Schneeweiss, 2013).

The chromosome numbers in angiosperms exhibit considerable variation, with the most common chromosome counts falling between  $2n = 14$  and  $2n = 40$  chromosomes (Grant, 1982; Masterson, 1994; Weiss-Schneeweiss & Schneeweiss, 2013). Chromosome numbers can also vary within a species. Such plant species encompass several cytotypes with different chromosome numbers, often without accompanying morphological variation, as observed in the species complex *Prospero autumnale* (Ainsworth et al., 1983; Weiss-Schneeweiss & Schneeweiss, 2013).

Various mechanisms can result in changes of the basic chromosome number. Structural changes, such as centric fusions and fissions, represent one example of mechanisms that can lead to alterations in the basic chromosome number (=dysploidy; Ainsworth et al., 1983; Vaughan et al., 1997; Weiss-Schneeweiss & Schneeweiss, 2013). Centric fusion involves the fusion of two chromosomes in the (peri)centromeric regions (often preceded by pericentric inversions to create subtelocentric chromosome), resulting in a decrease in chromosome number, whereas fission occurs when a chromosome breaks into two, leading to an increase in chromosome number (Schubert, 2007). Centric fusions and fissions are balanced rearrangements (Lysák & Schubert, 2013; Weiss-Schneeweiss & Schneeweiss, 2013). Another mechanism leading to an increase in chromosome number over time is polyploidy, either involving one species (=autopolyploidy) or different species (=allopolyploidy; Leitch & Leitch, 2008; Husband et al., 2013; Weiss-Schneeweiss et al., 2013). Unlike dysploidy or aneuploidy, polyploidy involves the multiplication of the complete set(s) of chromosomes (Leitch & Leitch, 2008). Both autopolyploids, once considered an evolutionary dead end, and allopolyploids have been shown to contribute significantly to plant diversification and speciation (Madlung, 2013; Weiss-Schneeweiss et al., 2013). Polyploidy is widespread in plants, particularly ferns and angiosperms, but different plant groups have

different propensities to form polyploids (Weiss-Schneeweiss & Schneeweiss, 2013; Weiss-Schneeweiss et al., 2013). It is widely accepted that all current angiosperms have undergone at least two whole-genome duplication rounds (WGD; Jiao et al., 2011).

Plant genomes vary not only in chromosome numbers, but also in genome sizes. Changes in genome size can result from combination of various processes such as polyploidization, the accumulation of repetitive DNA (leading to an increase in genome size), or the deletion of DNA (resulting in a decrease in genome size; Bennetzen, 2002; Hawkins et al., 2008; Weiss-Schneeweiss & Schneeweiss, 2013).

Another notable feature of the karyotype is the localization of primary and secondary constrictions. Primary constrictions represent the centromere, which is usually located at a specific position in the chromosome in most angiosperms (Weiss-Schneeweiss & Schneeweiss, 2013). Changes in centromere position may result from structural alterations affecting their location or through centromere deactivation and *de novo* formation (Han et al., 2006; Lysák & Schubert, 2013). Structural chromosomal changes such as pericentric inversions or translocations can also influence centromere position (Schubert, 2007). Secondary constrictions, known as nucleolar organizer regions (NORs), carry tandemly repeated 35S rRNA genes. The number of NORs can vary between taxa with a tendency toward reduction often observed following polyploidization events (Weiss-Schneeweiss & Schneeweiss, 2013). NORs can be located subterminally or interstitially within chromosomes (Schubert, 2007).

## 1.2. Repetitive DNA

A significant fraction of eukaryotic genomes is composed of repetitive DNA sequences, collectively referred to as the repeatome. Based on their organization, the repetitive DNAs can be divided into two major groups: dispersed repeats and tandem repeats. Dispersed repeats encompass transposable elements, whereas tandem repetitive DNA include satellite DNAs and rRNA genes (Figure 1; Weiss-Schneeweiss et al., 2015).

### 1.2.1. Transposable elements

Transposable elements (TEs) constitute a substantial portion of plant repeatomes and can transpose through various mechanisms. TEs are classified into two classes: Class I comprising retrotransposons, and Class II consisting of DNA transposons and Helitrons. The distinction between these classes lies in their mode of transposition (Weiss-Schneeweiss et al., 2015;

Bourque et al., 2018). Some transposable elements can significantly impact genome and chromosome size through their amplification (Weiss-Schneeweiss et al., 2015).

Class I TEs, retrotransposons, employ a "copy and paste" mechanism for transposition, facilitated by reverse transcriptase, which converts mRNAs into cDNA that is then integrated into new genomic locations (Weiss-Schneeweiss et al., 2015). Retrotransposons encompass two subclasses: LTR-retroelements (possessing long terminal repeats) and non-LTR retrotransposons (Weiss-Schneeweiss et al., 2015; Bourque et al., 2018). They differ in their integration mechanisms, involving either DNA cleavage and integration or target-primed reverse transcription (Weiss-Schneeweiss et al., 2015). LTR retrotransposons consist of two long terminal repeats flanking an open reading frame, which includes regions encoding proteins such as the gag protein, protease, integrase, reverse transcriptase, and RNase H (Neumann et al., 2019; Wells & Feschotte, 2020). These proteins are essential for retrotransposon's life cycle, increasing its copy numbers and integration into the host genome (Sanchez et al., 2017). LTR retrotransposons are further categorized into superfamilies such as Ty1-*copia* (Pseudoviridae) and Ty3-*gypsy* (Metaviridae) differing in the order of their coding domains, with either of these superfamilies being prevalent in different species (Weiss-Schneeweiss et al., 2015; Neumann et al., 2019). Non-LTR retrotransposons include long interspersed nuclear elements (LINEs) and short interspersed nuclear elements (SINEs). Unlike LTR retrotransposons, they lack the long terminal repeats. SINEs are non-coding RNA Pol III-transcribed nonautonomous retrotransposons, which rely on the reverse transcriptase and endonuclease activity provided by LINEs for transposition (Dewannieux et al., 2003; Elbarbary et al., 2016). In contrast, LINEs are autonomous retrotransposons transcribed by RNA Pol III, encoding the necessary proteins for transposition and capable of integrating into the genome independently (Beck et al., 2011; Elbarbary et al., 2016).

The Class II transposable elements (TEs), known as DNA transposons and Helitrons, employ a "cut and paste" or "peel and paste" mechanisms to relocate within the genome, respectively. Unlike retrotransposons, subclass II of Class II DNA transposons do not typically increase their copy number during transposition, hence they do not directly contribute to changes in genome size (Weiss-Schneeweiss et al., 2015; Hickman & Dyda, 2016). DNA transposons typically feature a transposase gene flanked by two terminal inverted repeats, crucial for excising and integrating the transposon into a new location. Examples of DNA transposon families include Tc1/mariner, PIF/Habinger, hAT, Mutator, and CACTA, among others (Feschotte & Pritham, 2007; Munoz-Lopez & Garcia-Perez, 2010; Hickman & Dyda, 2016). Helitrons, constituting the second subclass of DNA

transposons, utilize a "peel and paste" mechanism. Unlike subclass I DNA transposons, Helitrons generate a circular DNA intermediate, formed by peeling off the sense strand and synthesizing the antisense strand (Grabundzija et al., 2018). Helitrons lack the terminally inverted repeats present in DNA transposons (Kapitonov & Jurka, 2001). Autonomous Helitrons encode a Rep/Hel protein (DNA helicase and replication initiator), including a HUH nuclease domain, essential for their transposition (Kapitonov & Jurka, 2001).

### 1.2.2. Satellite DNA

Satellite DNAs (satDNAs) are tandemly repeated non-coding DNA motifs, consisting of tens to hundreds of thousands of monomers of varying lengths localized in few to many loci across the genome (Šatović-Vukšić & Plohl, 2023). These sequences are found in all eukaryotes, predominantly in heterochromatic regions, and are known to undergo dynamic changes over time. Therefore, studying the evolution of satellite DNA provides valuable insights into genome evolution processes (Schmidt & Heslop-Harrison, 1998; Šatović-Vukšić & Plohl, 2023). Although predominantly located in pericentromeric and subtelomeric heterochromatin, satellite DNAs can also be present in interstitial positions within the genome (Macas et al., 2007; Plohl et al., 2014; Šatović-Vukšić & Plohl, 2023). In addition to typical satellite DNAs, genomes harbor also mini- and microsatellites, which are typically located in euchromatin and differ in length of their monomers (microsatellites: 2-6 bp, minisatellites: 15-60 bp; Schmidt & Heslop-Harrison, 1998).

Satellite DNAs exhibit high diversity and are represented by various satellite DNA families, often species- or genus-specific, characterized by homogeneity of their monomer sequences across families. Different plant species harbor different sets of satDNA families (Schmidt & Heslop-Harrison, 1998; Weiss-Schneeweiss et al., 2015). The homogeneity of repeat sequences within satellite DNA families is achieved through concerted evolution, where novel variants of monomers are homogenized and fixed across the array (Elder & Turner, 1995; Hemleben et al., 2007; Šatović-Vukšić & Plohl, 2023). The length of satellite DNA monomers varies, with a preference for sequences ranging from 150-180 bp and 300-360 bp (Schmidt & Heslop-Harrison, 1998; Hemleben et al., 2007).

In plants, satDNAs can comprise from 0.1% to 36% of the genome, thus significantly influencing the genome size of species (Garrido-Ramos, 2015; Garrido-Ramos, 2017). While long regarded as "junk DNA," recent research has highlighted the involvement of satDNAs in crucial biological processes such as chromosome organization, chromosome pairing, genome

evolution, speciation, chromosome segregation and more (Grewal & Elgin, 2007; Plohl et al., 2012; Garrido-Ramos, 2015).

### 1.2.3. Ribosomal DNA

Ribosomal DNA (rDNA) is another type of tandemly arranged repetitive DNA, responsible for encoding ribosomal RNA. These rDNAs, including 35S rRNA and 5S rRNA genes, are ubiquitous in eukaryotic genomes and represent highly conserved segments, often utilized for karyotype analysis (Weiss-Schneeweiss & Schneeweiss, 2013; Jang et al., 2013; Weiss-Schneeweiss et al., 2015; Garcia et al., 2017; Jang et al., 2018a). The number and position of rRNA genes vary among species, with most having the 35S and 5S rRNA genes located on different chromosomes/loci (Garcia et al., 2017). The 35S rDNA monomer comprises three genes (18S, 5.8S, and 25/28S rRNA genes), interspersed with internal transcribed spacers (ITS 1 & 2) and flanked by an external transcribed spacer (ETS). These genes are organized in tandem arrays, separated by intergenic spacers (IGS), and can be found in numerous copies and one to many loci (Roa & Guerra, 2015; Weiss-Schneeweiss et al., 2015). The second type of rRNA, 5S rDNA, consists of the tandemly arranged 5S rRNA genes separated by NTS (non-transcribed spacers; Weiss-Schneeweiss et al., 2015; Garcia et al., 2017).

35S rDNA loci typically are localized subterminally within short arms of chromosomes (Roa & Guerra, 2012; Garcia et al., 2017). The number of 35S rDNA loci varies, with the majority of angiosperms having one 35S rDNA locus, followed closely by those with two loci of 35S rDNAs or more (Roa & Guerra, 2012). In comparison, most angiosperms (54.1%) have one pair of 5S rDNA loci, with 25.8% possessing two 5S rDNA loci (Roa & Guerra, 2015). The majority of 5S rDNA loci are located proximally within the short arms of chromosomes, irrespective of the number of loci (Roa & Guerra, 2015). However, the positioning of 5S rDNA loci appears to correlate with chromosome size (Roa & Guerra, 2015). 35S and 5S rDNA loci have different evolutionary trajectories. Both experience high rates of intra- and intergenomic mobility and are involved in genome diploidization (Clarkson et al., 2005; Raskina et al., 2008). However, 35S rDNAs are more prone to reduction of loci numbers due to silencing or loss of loci but also exhibit interlocus homogenization (Clarkson et al., 2005; Weiss-Schneeweiss et al., 2012; Roa & Guerra, 2015; Garcia et al., 2017).



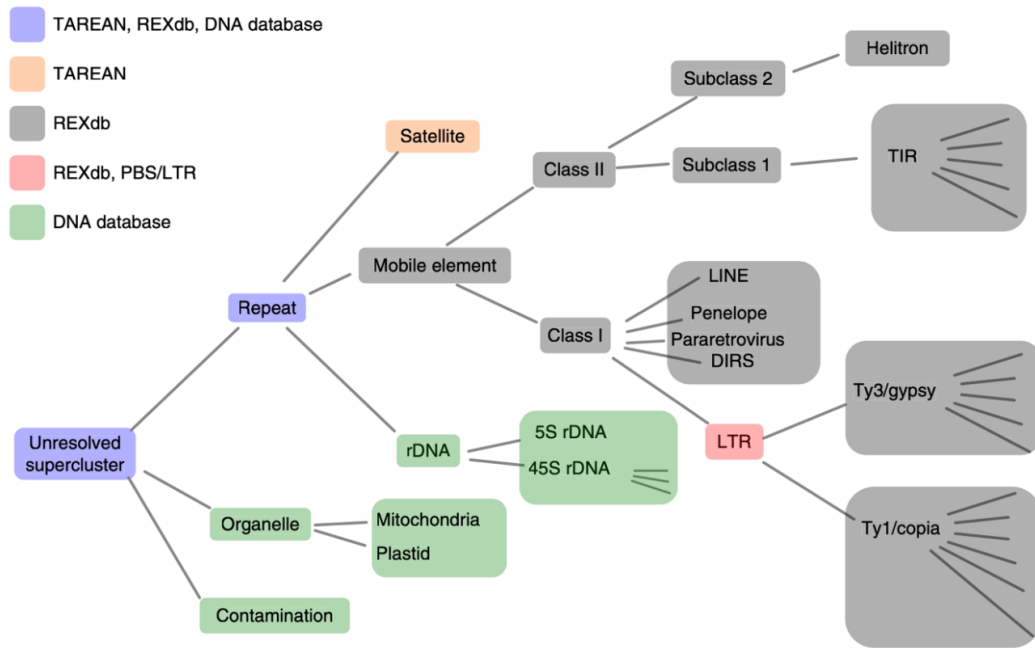


Figure 1. Classification of repetitive DNAs into different categories (Novák et al., 2020).

### 1.3. *Prospero autumnale* complex

The *Prospero autumnale* complex, formerly known as *Scilla autumnalis*, is autumn-flowering species of the genus *Prospero* Salisb., belonging to the Hyacinthaceae family (Speta, 1993; Jang et al., 2018a). The complex is distributed across the Mediterranean basin, the Caucasus, and western Europe (Figure 2; Speta, 1993; Jang et al., 2013). Comprising four diploid cytotypes and various genomic types of polyploids, it represents a diverse species complex (Jang et al., 2013; Jang et al., 2018a).

Diploid cytotypes within the *Prospero autumnale* complex differ in multiple features (Ainsworth et al., 1983; Ebert et al., 1996; Vaughan et al., 1997; Jang et al., 2013; Emadzade et al., 2014). Polyploidy, including both auto- and allopolyploidy, is prevalent in the *Prospero autumnale* complex, (Ainsworth et al., 1983; Speta, 1993; Vaughan et al., 1997; Weiss-Schneeweiss et al., 2013; Jang et al., 2018a). In addition to its regular A-chromosome set, the genome of *Prospero autumnale* harbors supernumerary genetic material, including B-chromosomes and supernumerary chromosomal segments (Jang et al., 2018b).



Figure 2. Distribution of the *Prospero autumnale* complex (POWO, 2024).

Genus *Prospero* provides an excellent opportunity to investigate the effects of chromosomal changes on plant diversification and speciation, given its substantial chromosomal variation. This is particularly evident in the species complex *Prospero autumnale*, which encompasses several cytotypes with varying basic chromosome numbers and numerous levels of polyploidy (Ainsworth et al., 1983; Parker et al., 1991; Vaughan et al., 1997; Jang et al., 2013). Thus, this species complex serves as the primary focus of analysis in this study.

### 1.3.1. Diploid cytotypes

The *Prospero autumnale* complex comprises four diploid cytotypes characterized by basic chromosome numbers ranging from  $x = 5$  to  $x = 7$ . These cytotypes include  $B^5B^5$  ( $x = 5$ ),  $B^6B^6$  ( $x = 6$ ),  $AA$  ( $x = 7$ ), and  $B^7B^7$  ( $x = 7$ ; Ainsworth et al., 1983; Jang et al., 2013). Significant differences are observed in the karyotype structure, as well as in the number and location of 5S/35S rDNA loci, genome size, and position of NOR among these cytotypes (Vaughan et al., 1997; Jang et al., 2013). Additionally, variation in the satDNA *PaB6* loci and copy numbers is evident across cytotypes (Emadzade et al., 2014). Despite these chromosomal differences, the cytotypes are morphologically very similar (Figure 3; Jang et al., 2013). Phylogenetic analyses indicated that the  $B^6B^6$ ,  $B^7B^7$  and  $AA$  form sister clades, while  $B^5B^5$  is derived from the  $B^7B^7$  cytotype (Jang et al., 2013; Jang et al., 2018a). This study will focus on the diploid cytotypes  $B^6B^6$ ,  $B^7B^7$  and their diploid hybrid,  $B^6B^7$ .



Figure 3. Morphological features of diploid cytotypes of the *Prospero autumnale* complex. 1-2:  $B^7B^7$  ( $2n = 2x = 14$ ), 3-4: AA ( $2n = 2x = 14$ ), 5-6:  $B^6B^6$  ( $2n = 2x = 12$ ), 7-8:  $B^5B^5$  ( $2n = 2x = 10$ ). Image modified from Jang, 2013.

#### 1.3.1.1. Cytotype $B^7B^7$ ( $2n = 2x = 14$ )

The  $B^7B^7$  cytotype, characterized by a basic chromosome number of  $x = 7$ , is distributed across the whole Mediterranean region, with the broadest range among all cytotypes (Figure 4; Parker et al., 1991; Jang et al., 2013; Jang et al., 2018a). Its karyotype comprises five submetacentric chromosomes, one subtelocentric chromosome and one near metacentric chromosome. The NOR/35SrDNA locus is located in the pericentromeric region of chromosome 3 (Jang et al., 2013). The DNA content of the  $B^7B^7$  cytotype ranges from 4.23 pg to 4.45 pg, depending on the number of 5S rDNA loci (Jang et al., 2013). The pericentric 35S rDNA locus is located on the long arm of chromosome 3, while the number and location of 5S rDNA loci vary. Some individuals/populations carry one 5S rDNA locus located on the long arm of chromosome 1, whereas the individuals with duplicated 5S rDNA locus carry two loci in close proximity to each other on the long arm of chromosome 1 (Jang et al., 2013). Additionally, there are either 12-14 moderate or six weak satDNA *PaB6* signals in the  $B^7B^7$  cytotype (Emadzade et al., 2014; Jang et al., 2018a). This cytotype is hypothesized to resemble the ancestral *Prospero autumnale* cytotype (Vaughan et al., 1997; Jang et al., 2013).

#### 1.3.1.2. Cytotype AA ( $2n = 2x = 14$ )

Another diploid cytotype with a basic chromosome number of  $x = 7$  is the AA cytotype, found in the south of the Iberian Peninsula and Morocco (Figure 4). The karyotype structure is identical to the  $B^7B^7$  cytotype, with five submetacentric chromosomes, one subtelocentric chromosome and one near metacentric chromosome. The DNA content is 7.85 pg. The 35S rDNA locus is located on the long arm of chromosome 3, while the 5S rDNA locus is in the pericentric region of chromosome 2. Only two weak satDNA *PaB6* signals were detected in the AA cytotype (Jang et al., 2013; Jang et al., 2018a).

#### 1.3.1.3. Cytotype $B^6B^6$ ( $2n = 2x = 12$ )

The  $B^6B^6$  cytotype, characterized by a basic chromosome number of  $x = 6$ , carries a fusion chromosome formed by the fusion of chromosomes 6 and 7 of cytotype  $B^7B^7$ , resulting in the reduction of the base chromosome number (Vaughan et al., 1997; Jang et al., 2013). This cytotype is endemic to Crete (Figure 4; Vaughan et al., 1997; Jang et al., 2013). Its karyotype comprises four submetacentric chromosomes, one subtelocentric and a submetacentric fusion chromosome F(6/7). One nucleolus organizer region (NOR) is typically found on chromosome 3 (Jang et al., 2013). The 1C DNA content of this cytotype is 6.27 pg. Two 5S rDNA loci are present in this cytotype, one on the short arm of chromosome 2 and another on the long arm of chromosome 1 (Jang et al., 2013). A total of 12 strong signals of satDNA *PaB6* were detected in this cytotype, with the loci being in the pericentromeric region on each chromosome (Emadzade et al., 2014; Jang et al., 2018a).

#### 1.3.1.4. Cytotype $B^5B^5$ ( $2n = 2x = 10$ )

The  $B^5B^5$  cytotype has a basic chromosome number of  $x = 5$  and is endemic to Libya (Figure 4). In addition to the fusion chromosome F(6/7), which is identical to the fusion chromosome in the  $B^6B^6$  cytotype, it carries another submetacentric fusion chromosome F(1/3), along with two submetacentric chromosomes (2 & 5) and one subtelocentric chromosome (4). The 35S rDNA locus is located on the short arm of the fusion chromosome F(1/3), while the 5S rDNA is on the other fusion chromosome F(6/7). The DNA content is  $4.86 \pm$  pg and eight satDNA *PaB6* signals were detected in this cytotype. Notably, the  $B^5B^5$  cytotype does not contribute to the formation of polyploids (Jang et al., 2013; Emadzade et al., 2014; Jang et al., 2018a).

#### 1.3.1.5. Diploid hybrid B<sup>6</sup>B<sup>7</sup>

Two of the diploid cytotypes within the *Prospero autumnale* complex, B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup>, hybridize in natural populations and form diploid hybrids B<sup>6</sup>B<sup>7</sup>. This hybrid possesses a chromosome number of  $2n = 2x = 13$ , comprising six chromosomes derived from the B<sup>6</sup>B<sup>6</sup> cytotype and seven chromosomes derived from the B<sup>7</sup>B<sup>7</sup> cytotype (Jang et al., 2018a).

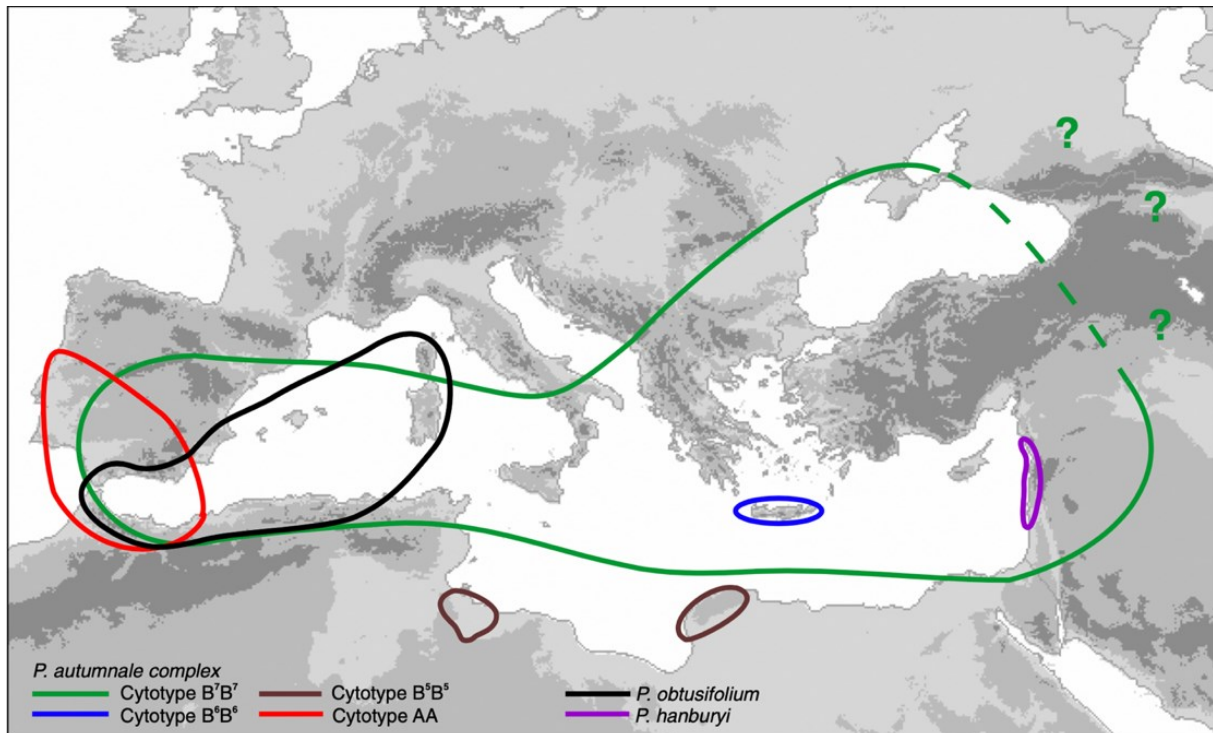


Figure 4. The geographical distribution patterns of the diploid cytotypes of the *Prospero autumnale* complex, along with two other species, *P. hanburyi* and *P. obtusifolium* (Jang et al., 2013).

#### 1.3.2. Polyploid cytotypes

Polyploids in the *Prospero autumnale* complex are formed by the diploid cytotypes B<sup>6</sup>B<sup>6</sup>, B<sup>7</sup>B<sup>7</sup>, and AA, with B<sup>5</sup>B<sup>5</sup> not contributing to polyploid formation. Autotetraploids are exclusively derived from the B<sup>7</sup>B<sup>7</sup> cytotype, while allopolyploids arise from combinations of the AA and B<sup>7</sup>B<sup>7</sup> or B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup> cytotypes (Figure 5; Ainsworth, 1980; Ebert, 1993; Taylor, 1997; Vaughan et al., 1997; Jang et al., 2018a). This study focuses on the tetraploids formed by the B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup> cytotypes.

##### 1.3.2.1. Autotetraploid B<sup>7</sup>B<sup>7</sup>B<sup>7</sup>B<sup>7</sup> ( $2n = 4x = 28$ )

The autotetraploid B<sup>7</sup>B<sup>7</sup>B<sup>7</sup>B<sup>7</sup> is the sole autotetraploid found in the *Prospero autumnale* complex and has  $2n = 28$  chromosomes. It has a broad geographic distribution, with individuals found as far north as England as far east as Greece and Turkey. One 35S rDNA

locus and either one or two 5S rDNA loci (depending on the number of 5S rDNA loci in the  $B^7$  parent) were detected. In individuals with duplicated 5S rDNA loci 15-26 weak satDNA *PaB6* signals were detected, while with two 5S rDNA loci one to two strong signals and few weak signals were detected. Notably, the 1C DNA content was not additive, with either larger (in individual with two 5S rDNA loci) or smaller (in individuals with four 5S rDNA loci) genome sizes than expected (Vaughan et al., 1997; Jang et al., 2018).

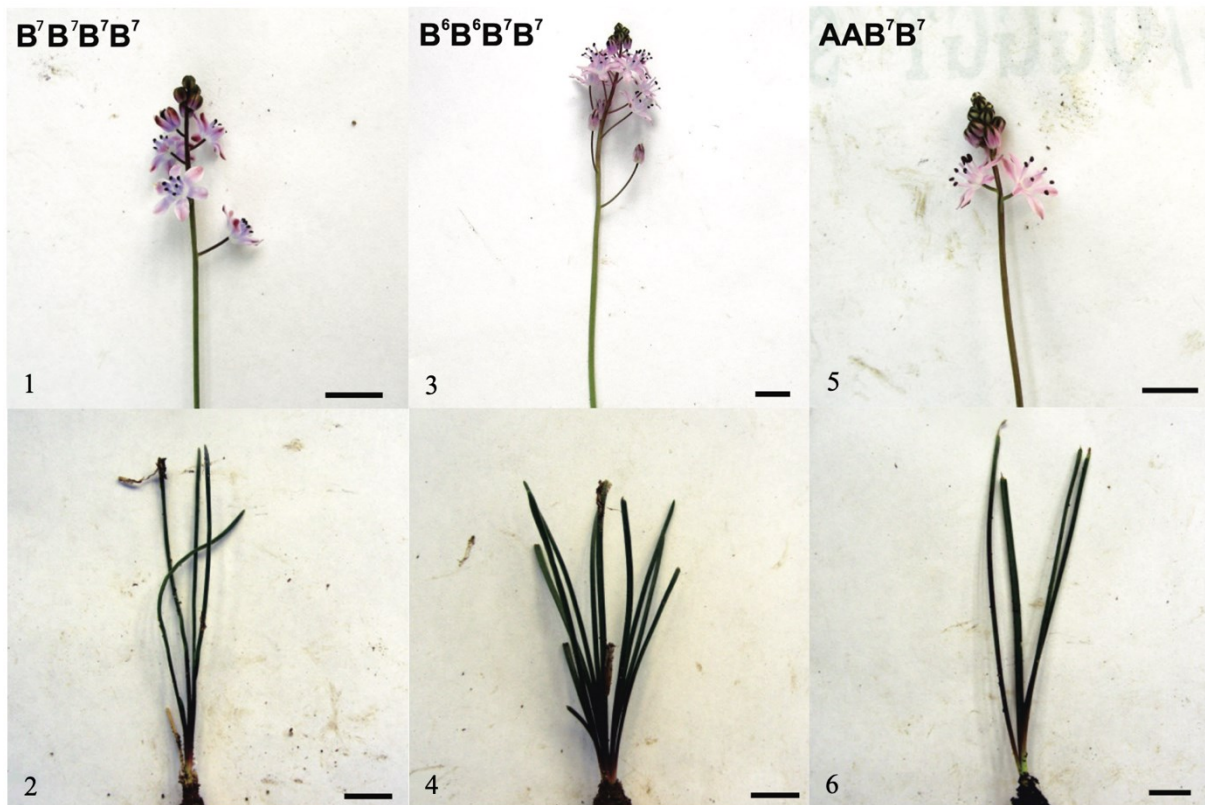


Figure 5. Morphological features of tetraploid cytotypes of the *Prospero autumnale* complex. 1-2:  $B^7B^7B^7B^7$  ( $2n = 4x = 28$ ), 3-4:  $B^6B^6B^7B^7$  ( $2n = 4x = 26$ ), 5-6:  $AAB^7B^7$  ( $2n = 4x = 28$ ). Image modified from Jang, 2013.

#### 1.3.2.2. Allotetraploid $AAB^7B^7$ ( $2n = 2x = 28$ )

The allotetraploid  $AAB^7B^7$  has  $2n = 28$  chromosomes, with each parental cytotype contributing equally to its genome. A distinguishing feature is the size difference between chromosomes of AA origin, which are larger in size than those of  $B^7B^7$  origin. This tetraploid is endemic to the Iberian Peninsula, reflecting the geographical restriction of one of its parents, the AA cytotype. Analysis revealed the presence of a single NOR of  $B^7B^7$  origin, and the loss of the NOR of AA origin. Additionally, 5S rDNA loci inherited from both parents were detected. The number of satDNA *PaB6* signals ranged between 12 and 14. The 1C DNA content was larger than expected (Vaughan et al., 1997; Jang et al., 2018a).

### 1.3.2.3. Allotetraploid $B^6B^6B^7B^7$ ( $2n = 4x = 25-28$ )

The allotetraploid  $B^6B^6B^7B^7$  individuals exhibit variation in chromosome numbers, ranging from  $2n = 25$  to 28 chromosomes, depending on the numbers of fusion chromosome F(6/7) and free chromosomes 6 and 7 in the genome. This type of allotetraploid is endemic to Crete and is classified into four distinct groups based on differences in the number and distribution of 35S and 5S rDNA loci, as well as the number of satDNA *PaB6* loci and the parental origin of chromosomes. The formation of the four distinct groups within the *Prospero autumnale* complex was facilitated by processes such as repeat homogenization, numerical convergence, and recurring cycles of hybridization (Jang et al., 2018a).

#### Group I allotetraploids ( $2n = 4x = 25-28$ )

Group I allotetraploids encompass individuals varying in chromosome numbers ( $2n = 25, 26, 27$  or 28) and with distinct genetic features. These include 35S rDNA locus on all four chromosomes 3, single 5S rDNA locus on all chromosomes 1 and 5S rDNA locus on all chromosomes 2. Genomic *in situ* hybridization (GISH) with parental genomic DNAs failed to differentiate the parental genomes. The presence of strong satDNA *PaB6* signals on all chromosomes suggests the spread of repetitive DNAs from one parental genome ( $B^6B^6$ ) to the other ( $B^7B^7$ ). The observed pattern of post-polyploidization genome homogenization towards one of the parents, specifically diploid cytotype  $B^6B^6$ , within Group I allotetraploids suggests that this process could be responsible for the failure to determine the parentage through GISH (Jang et al., 2018a).

#### Group II allotetraploids ( $2n = 4x = 28$ )

Group II allotetraploids emerged from a cross event between  $B^7B^7$  diploids and Group I allotetraploids with  $2n = 28$ . There have 28 chromosomes, and the F1 fusion chromosome is absent. In Group II tetraploids, 14 strong satDNA *PaB6* signals were consistently observed. Analysis revealed the presence of 35S rDNA locus on two chromosomes 3 of  $B^7$  origin, occasionally accompanied by a weaker third signal, and a distinct pattern of 5S rDNA loci distribution, with single and one duplicated loci in all chromosomes 1, and additional locus on two chromosomes 2 (Jang et al., 2018a).

### Group III allotetraploids ( $2n = 4x = 28$ )

Group III allotetraploids have  $2n = 28$  chromosomes, comprising 21 chromosomes inherited from one parent and seven from the other, originating from the parental genomes of Group I and Group II allotetraploids, respectively. Analysis revealed the presence of four 35S rDNA signals on all chromosomes 3, along with three single and one duplicated 5S rDNA signals on chromosomes 1 and three signals on three of four chromosomes 2. In Group III tetraploids, 21 strong satDNA *PaB6* signals were observed, and GISH differentiated parental genomes (Jang et al., 2018a).

### Group IV allotetraploids ( $2n = 4x = 28$ )

The Group IV allotetraploids originated from a cross between Group II allotetraploids and B<sup>7</sup>B<sup>7</sup> diploid cytotype, as confirmed by GISH, resulting in a chromosome number of  $2n = 28$ . They inherited seven chromosomes from one parent and 21 from the other, with no fusion chromosome present. This tetraploid carries four 35S rDNA signals on all chromosomes 3 as well as 5S rDNA loci on all four chromosomes 1, one of which is duplicated, and one 5S rDNA signal on one of four chromosomes 2. The Group IV tetraploid had seven strong signals of satDNA *PaB6* (Jang et al., 2018a).

#### 1.4. Aim of the study

The *Prospero autumnale* complex exhibits high levels of chromosomal variation, both on diploid and polyploid levels across its distribution range, making it an ideal candidate for studying the dynamics and patterns of chromosomal change and its significance for the genome evolution. With the advent of next generation sequencing (NGS) technologies and dedicated pipeline RepeatExplorer2, it is possible to analyze the repeatome composition of the any plant genome. This approach implemented with analyses of patterns of chromosomal distribution of satellite DNAs identified has a potential to provide new information to better understand the mechanisms and processes driving the evolution within the *Prospero autumnale* complex, that contribute to the high levels of chromosomal diversity observed in this species today.

The specific aims of this study are to 1) analyze the repeatome composition of two diploid parental cytotypes, B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup>, as well as Group I allotetraploid B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> of *Prospero autumnale*, and to identify potential novel putative satellite DNAs using Illumina DNA sequence data and the bioinformatics pipeline RepeatExplorer2, 2) measure the genome



sizes of selected *Prospero autumnale* plants, 3) map the three known putative satellite DNAs (satDNAs), namely satDNA *PaB6*, satDNA Pa138 and satDNA Pa147, in the chromosomes of B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup> cytotypes, as well as in their diploid hybrid and all four groups of the allotetraploids B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup>, employing fluorescence *in situ* hybridization (FISH) and 4) map the parental genomic DNAs in chromosomes in a diploid hybrid and all B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> allotetraploids of *Prospero autumnale* using genomic *in situ* hybridization (GISH).

This study provides novel insights into the composition and evolution of total repetitive DNA fractions of two diploid cytotypes of *Prospero* and in their diploid and allotetraploid hybrid genomes. It also offers new insight into the post-polyploidization dynamics of the evolution of novel tandem satellite DNA repeats in these genomes.

## 2. Material and Methods

### 2.1. Plant Material

16 plants from various cytotypes of *Prospero autumnale* (Hyacinthaceae) were analyzed, including several B<sup>6</sup>B<sup>6</sup> ( $2n = 2x = 12$ ; Crete, Greece) and B<sup>7</sup>B<sup>7</sup> ( $2n = 2x = 14$ ; Mediterranean) diploid individuals, B<sup>6</sup>B<sup>7</sup> diploid hybrids ( $2n = 2x = 13$ ; Crete, Greece) and several individuals of B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> allotetraploids ( $2n = 4x = 25-28$ ; Crete, Greece), which represent all four groups (Jang et al., 2018). All plants were cultivated in the Botanical Garden of the University of Vienna (Table 1).

### 2.2. Pre-treatment and fixation of chromosomes

One week prior to the harvesting of root tips, the outer root mass of the plant was removed, and the plant was watered. This process aimed to induce the proliferation of nascent roots, used later as a source of mitotic metaphase chromosomes for fluorescence *in situ* hybridization (FISH) experiments. Subsequently, the newly developed healthy root meristems were collected into a bottle containing water and washed to remove the access soil. Prior to the fixation of chromosomes, the roots were pretreated with a 0.05% colchicine solution (Sigma Aldrich, Vienna, Austria) for 4 hours at room temperature in darkness to increase the number of metaphase cells with well-condensed chromosomes. The employed chemical agent, colchicine, effectively prohibits spindle fibers polymerization, leading to chromosome condensation and allowing their effective spreading. Post-treatment, the roots were rinsed with tap water and fixed in mixture 3:1 mixture ethanol and glacial acetic acid (Merck KGaA, Darmstadt, Germany) for 12 hours at room temperature. This fixation process aimed to maintain the structural integrity of chromosomes while facilitating the extraction of various proteins.

### 2.3. Feulgen staining for karyotype analysis

For the karyotype analysis, Feulgen staining was applied following the protocol of Fukui and Nakayama (1996). The initial step in the Feulgen staining process involved the removal of fixed material from the fixative followed by the hydrolysis in 5N hydrochloric acid (HCl; Merck KGaA, Darmstadt, Germany) for 20 minutes at room temperature. The hydrochloric acid cleaved the DNA, rendering the aldehyde groups accessible for reaction with Schiff's reagent. Roots were rinsed with tap water to remove residual 5N HCl, and subsequently incubated in Schiff's reagent (Merck KGaA, Darmstadt, Germany) for 60 minutes at room

temperature in darkness. Schiff's reagent (basic leucofuchsin) reacts with aldehyde groups in the DNA, resulting in a distinctive pink staining of the nuclei in the meristem cells. The stained root tips were placed on a microscope slide in a drop of 60% acetic acid and the meristem was dissected using a stereomicroscope (Carl Zeiss Stemi 2000, Carl Zeiss, Vienna, Austria) and entomological needles. Subsequently, coverslip was applied and the meristem was squashed using a needle. The slides were examined under a microscope (Axio Imager M2, Carl Zeiss, Vienna, Austria), and images of the chromosomes were captured with a CCD camera (Axiocam 702 mono, Carl Zeiss, Vienna, Austria) using the software Zen pro 3.8 and used for subsequent karyotype analysis. Chromosomes were cut out using the software Corel Photo-Paint X8 (Cascade Parent Limited, Ottawa, Canada).

#### 2.4. Fluorescence *in situ* hybridization (FISH)

Fluorescence *in situ* hybridization was conducted following the established protocol (Jang & Weiss-Schneeweiss, 2015). The chromosomal spreads were prepared using enzymatic digestion. Fixed root tips (Table 1) were washed with citrate buffer, pH 4.8, for 10 minutes to remove the fixative. Subsequently, the roots were incubated in a prewarmed enzyme mix, comprising 0.4% pectolyase (Sigma Aldrich, Vienna, Austria), 0.4% cytohelicase (Sigma Aldrich, Vienna, Austria), and 1% cellulase (Serva, Heidelberg, Germany) in citrate buffer (pH 4.8), at 37°C for 20-25 minutes. Following the digestion of the cell walls, the material was washed in two changes of citrate buffer, pH 4.8. Chromosomal spreads were prepared by placing a single root meristem onto a microscope slide in a droplet of 60% acetic acid. The material was fragmented under a stereo microscope (Carl Zeiss Stemi 2000, Carl Zeiss, Vienna, Austria) and squashed under a coverslip. Chromosomes were gently squashed using the tip of a needle. The slide quality was assessed under a phase-contrast microscope (AxioImager M2, Carl Zeiss, Vienna, Austria). Slides with good quality chromosomal spreads were frozen to -80°C (FTS systems FlexiCool), and the coverslips were removed. Following air-drying, the slides were stored in the refrigerator at 4°C until further use.

The chromosomes were refixed in fixative (3:1 ethanol-acetic acid) for 15-20 minutes at room temperature and washed in 96% ethanol for 15 minutes, followed by air-drying for 1 hour at room temperature. To remove any residual RNAs that could increase the background, the slides were treated with DNase-free RNase (100 µg/ml in 2xSSC; Sigma Aldrich, Vienna, Austria) in a prewarmed wet chamber (2xSSC), in a water bath at 37°C for 1 hour. The slides were then washed three times 5 minutes in 2xSSC (pH 7.0) at room temperature. Afterwards the preparations were pretreated with 10mg/ml pepsin, pH 2 (Sigma Aldrich, Vienna,

Austria). The slides were incubated in pepsin solution for 20 minutes at 37°C in a waterbath. This step was performed to remove proteins of the cytoplasm. Following the treatment, the slides were washed three times 5 minutes each with 2xSSC at room temperature. The slides were then incubated in 4% paraformaldehyde (VWR, Vienna, Austria) for 5 minutes at room temperature to re-fix the remaining histones and washed in four changes of 2xSSC, 5 minutes each. Preparations were dehydrated in 70% cold ethanol and 96% cold ethanol for 3 minutes each at room temperature. The slides were then air-dried for an hour.

A hybridization mix (90 µl) was prepared, containing 10% (w/v) dextran sulfate (Sigma Aldrich, Vienna, Austria), 0.02xSSC and 100 ng/µl of blocking DNA SS (Sigma Aldrich, Vienna, Austria). 100 ng of each of the two probes (combinations of three probes were used; Table 2) were added to the mix. The DNA probes were labelled with haptens, either biotin or digoxigenin (synthesized and labelled commercially; Sigma Aldrich, Vienna, Austria). The hybridization mix containing the probes was denatured at 96°C for 2-5 minutes on the thermoblock, followed by a 5-minute incubation on ice. Subsequently, 10 µl of the denatured hybridization mix was applied to each slide, the coverslip was applied and the edges were sealed with Fixogum. Chromosomes and probes were co-denatured on the PCR *in situ* block (Peqlab, Vienna, Austria) following a program that included 4 minutes at 75°C, 1 minute at 65°C, 1 minute at 55°C, and 1 minute at 45°C. The slides were transferred to a prewarmed wet chamber (2xSSC) and hybridized overnight at 37°C. Following the overnight incubation at 37°C, the slides were washed in 2xSSC at room temperature to remove the coverslips. Subsequently, three stringent washes, 5 minutes each, were carried out at 39°C in a water bath in 2xSSC, 0.1xSSC, and 2xSSC-Tween20 (Sigma Aldrich, Vienna, Austria). After the final wash, 50 µl of the blocking solution (1% bovine serum albumin in 2xSSC-Tween20; Sigma Aldrich, Vienna, Austria) was applied to each slide and incubated in a wet chamber in the water bath for 20 minutes at 37°C to prevent unspecific binding of antibodies. Subsequently, 50 µl of the detection solution (streptavidin-Cy3 to detect biotin and anti-digoxigenin-FITC to detect digoxigenin; both Sigma Aldrich, Vienna, Austria) was applied to each slide and incubated in the water bath for 1 hour at 37°C. The slides were then washed twice with 2xSSC for 7 minutes each, and with 2xSSC-Tween20 for 7 minutes at 37°C. The excess buffer was removed from the slides and 8 µl of the antifade buffer (Vectashield containing 2 ng/µl DAPI; Vector Laboratories, Inc., Burlingame, USA) was added to each slide. The antifade buffer incorporated DAPI for nucleic acid staining and Vectashield to prevent photobleaching, thereby preserving the fluorescence. Coverslips were added and sealed with Fixogum. Slides were stored at 4°C overnight before they were examined using

an epifluorescence microscope (Carl Zeiss AxioImager M2, Carl Zeiss, Vienna, Austria). Images were taken with the CCD camera AxioCam HRm using the software Zen pro 3.8, contrasted using Adobe Photoshop 2021 (Adobe Inc., California, U.S.) and subsequently cut out using Corel Photo-Paint X8 (Cascade Parent Limited, Ottawa, Canada).

## 2.5. Genomic *in situ* hybridization (GISH)

The genomic *in situ* hybridization (GISH) protocol used was the same as that of FISH, except for the probes used. Genomic DNAs (gDNAs) of the two diploid parental genomes, B<sup>6</sup> and B<sup>7</sup>, were used as probes in the hybridization mix. The parental specific DNA probes used were H156 BIO (B<sup>6</sup>) and H464 DIG (B<sup>7</sup>). The total genomic DNA from the parental genomes were extracted using the CTAB method (Doyle and Doyle, 1987) and sheared and labeled according to Jang and Weiss-Schneeweiss (2015). Probes were purified using ethanol precipitation and resuspended in water (Jang and Weiss-Schneeweiss, 2015). The gDNA probes were incorporated into the hybridization mix (FISH protocol). The slides were examined under an epifluorescence microscope (Carl Zeiss AxioImager M2, Carl Zeiss, Vienna, Austria) and images were acquired with the AxioCam HRm using the software Zen pro 3.8. Subsequently the slides were reprobbed with satellite DNA probes. For reprobbed, the slides after GISH were incubated in 2xSSC at 37°C for 10 minutes to remove the coverslips, and then incubated in two changes of 4xSSC+Tween20 at room temperature for 20 minutes each, followed by an additional 10 minutes incubation in 2xSSC at room temperature. The slides were then dehydrated in 70% and 96% ethanol for 3 minutes each and air-dried. Subsequently new hybridization mix containing selected satellite DNA probes was applied using the FISH protocol (Table 1). Unlabeled blocking DNA (PCR-amplified *PaB6* monomers) was also added to the GISH hybridization mixture to block *PaB6* satellite DNA. This enabled to obtain good-quality labelling of whole parental chromosomes (Jang and Weiss-Schneeweiss, 2015).

Table 1. List of Prospero autumnale plants used for FISH and GISH (indicated with asterisk \*).

Plant number	Cytotype	Diploid chromosome number (2n)	Genome size (1C, pg)	Locality
H193	B <sup>6</sup> B <sup>6</sup>	12	-	Greece, Crete
H160	B <sup>6</sup> B <sup>6</sup>	12	6.12 ± 0.01	Greece, Crete
H44	B <sup>7</sup> B <sup>7</sup>	14	4.35 ± 0.04	Ukraine, Crimea
H47	B <sup>7</sup> B <sup>7</sup>	14	4.49 ± 0.03	Greece, Crete
H237	B <sup>6</sup> B <sup>7</sup>	13	-	Greece, Crete, Amariano
H77	B <sup>6</sup> B <sup>7</sup>	13	5.37 ± 0.05	Greece, Crete
H258*	B <sup>6</sup> B <sup>7</sup>	13	5.45 ± 0.07	Greece, Crete
H211	B <sup>6</sup> B <sup>7</sup>	13	5.40 ± 0.03	Greece, Crete
H347	B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group I	28	-	Greece, Crete
H331	B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group I	28	11.53 ± 0.03	Greece, Crete
H213	B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group I	27 + 1 B	12.19 ± 0.05	Greece, Crete
H434*	B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group II	28	10.06 ± 0.20	Greece, Crete
H354	B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group II	28	-	Greece, Karpathos
H388	B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group II	28	10.25 ± 0.03	Greece, Crete
H238*	B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group III	28	10.84 ± 0.07	Greece, Crete
H152*	B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group IV	28	9.25 ± 0.01	Greece, Crete

Table 2. List of satellite DNA probes used for FISH.

Satellite DNA Name	Consensus length	Sequence
Satellite DNA <b>PaB6</b>	249 bp	CCCTAACCCCTAATTGGAACCTGGCCAAAAACCTTAACCCT GACTAGTGGAAACCGTACACATCCCAATAACTCTAACCC TAATCAGAACTGGCCTCCATCCAAAAAACCAAACCCTA ATCGGGGGAATTGTCCACTTTCATTAACCCTAACCCCTA ACTAGTCGAAGCGTCCACTAACTCAAATACTAACCCCT AATTGGAATCGACCACCAACCCAAACACTAACCCAGGGA ACAAGACACCTTCGAAAAA
Satellite DNA <b>Pa138</b>	34 bp	GATGTGATTCTTGAATGGAAATCTAAGAAGATGT
Satellite DNA <b>Pa147</b>	76 bp	TCAGAGTTAGATTAGAAACAAATCAAATAATTGATAAAGAA TTACAAGAACAAATCAAGAAAGATACAAGAACAAC

## 2.6. Genome size estimation using flow cytometry

Flow cytometry is a method utilized to examine various characteristics of cells in a liquid solution as they pass through a laser beam, producing fluorescent and scattered light signals that are then detected and processed (McKinnon, 2018). The flow cytometry procedure used followed the protocols of Temsch (2003) and Baranyi and Greilhuber (1996) and was employed to determine the genome size of four *Prospero autumnale* plants. *Pisum sativum* “Kleine Rheinländerin” (1C = 4.42 pg, Greilhuber and Ebert, 1994) and *Solanum pseudocapsicum* (1C = 1.29 pg, Temsch et al., 2010) were used as internal standards. All analyzed plants were cultivated in the Botanical Garden of the University of Vienna, Austria.

Leaf material from *Solanum pseudocapsicum*, *Pisum sativum* and analyzed *Prospero autumnale* samples were collected from the living plants. Material of the sample and appropriate standard were co-chopped using a razor blade in 600 µl of ice-cold isolation buffer in a Petri dish until the cell suspension turned green, indicating the breakdown of the cell walls and membranes. Additional 600 µl isolation buffer was added to collect the majority of the nuclei in the suspension, which was then filtered using a 50 µm mesh (Saatile Hitech, Sericol GmbH, Bottrop, Germany) to prevent larger leaf material from blocking the flow chamber. The filter was removed and 50 µl of RNase A (Sigma Aldrich, Vienna, Austria) was added to the suspension to digest RNAs at 37°C for 30 minutes in the water bath. Subsequently, the samples were incubated in 6% propidium iodide (PI; Sigma Aldrich, Vienna, Austria) in 0.4 M Na<sub>2</sub>PO<sub>4</sub> at 4°C for 1 hour, to stain the isolated nuclei.

The flow cytometer (CyFlow ML, Partec, Muenster, Germany) equipped with a green laser (5432 nm, 100 mW, Cobolt Samba, Cobolt AB, Stockholm, Sweden) was used to measure the samples. The flow cytometer was first adjusted using the reference standard, followed by measurements of the relative fluorescent intensity of *Prospero autumnale* samples containing that specific standard. A total of 3333 particles were measured per run and the results were visualized using the flow cytometry data acquisition and analysis software FloMax (Partec, Muenster, Germany). Each sample was measured three times. The coefficients of variation of the G<sub>1</sub> peak of the reference standards were expected to be below 3% (Temsch et al., 2021), while approximately 5% was considered acceptable for *Prospero autumnale* samples. A histogram of fluorescence intensity was acquired, depicting the G<sub>1</sub> and G<sub>2</sub> peaks of the reference standard and *Prospero autumnale* samples. Because the 1C value of the reference organism is known, the 1C value of the sample can be calculated according to the following equation: (adapted from Temsch et al., 2021)

$$1C \text{ value (Prospero)} = \frac{G1 \text{ peak position (Prospero)}}{G1 \text{ peak position (Standard)}} \times 1C \text{ value (Standard)}$$

## 2.7. Analysis of repetitive DNA using RepeatExplorer2

The genomic DNAs of three *Prospero autumnale* individuals representing diploid B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup> cytotypes as well as an allotetraploid of B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> (Group I; Table 3; Jang et al., 2018a) were isolated using CTAB protocol (Jang and Weiss-Schneeweiss, 2015). The concentration of the extracted DNAs was measured with Quant-iT Picogreen dsDNA assay kit (PeqLab, Austria) and the extracts were checked on agarose gel. DNA fragmentation (500-700 bp), library preparation and sequencing were performed at CSF-NGS (Vienna Biocenter, Austria). Each of the three samples were sequenced using next generation (NGS) 150 bp paired-end Illumina HiSeq2500 technology (Illumina, San Diego, CA, USA). The analysis of repeat composition of three *Prospero autumnale* cytotypes was conducted using a RepeatExplorer2 pipeline (<https://repeatexplorer-elixir.cerit-sc.cz/>; Novák et al., 2020). The repeatome analyses were performed for each of the cytotypes individually (diploid B<sup>6</sup>B<sup>6</sup>, B<sup>7</sup>B<sup>7</sup>, and an allotetraploid of B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup>), followed by comparative analyses of both diploids and then all three cytotypes following the protocols in Novák et al. (2020).

Briefly, repetitive DNA types and families were identified by clustering via all-to-all DNA sequence comparison of NGS datasets and assembled into clusters using a graph-based clustering algorithm. Clusters representing the same repeat type were grouped into superclusters (Novák et al., 2020). This approach enabled the identification and quantification



of various repetitive DNA types, including dispersed mobile genetic elements and tandem repeats (including satellite DNAs) present in the genome.

For the repeat composition analysis, two fastq files containing each either forward or reverse reads were uploaded to the Galaxy Server. Subsequently, the quality of the reads was assessed using the Fastqc tool. This examination aimed to identify any contamination by adapter sequences and any sequence bias (Figure 6). The reads were then pre-processed using the 'Pre-processing of FASTQC paired-end reads' tool in Repeat Explorer Utilities. The first 11 base pairs of all reads were trimmed, and the reads were filtered based on quality, discarding all single reads. Finally, the two forward and reverse reads were interlaced. Upon completion of the pre-processing, two new files were generated, one with all interlaced paired-end reads and the other providing information about the nucleotide composition after filtering (Figure 7).

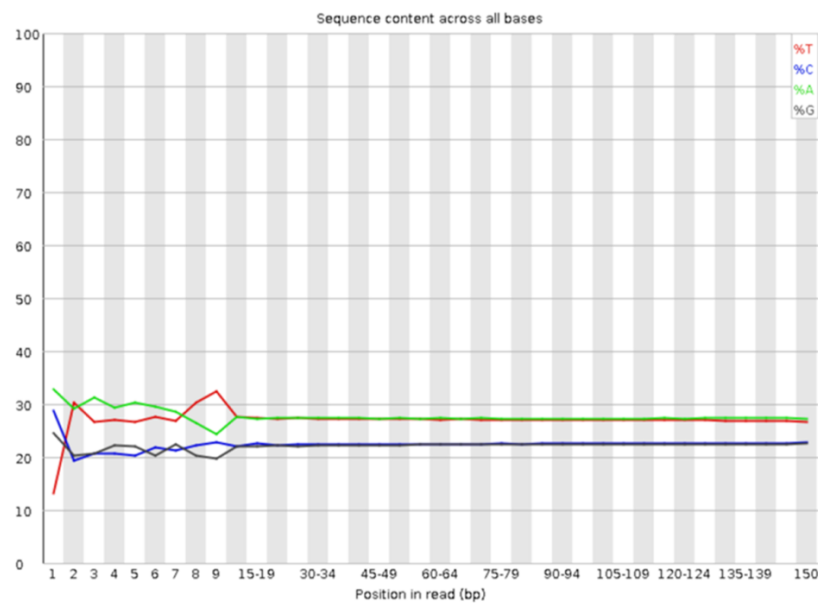


Figure 6. Nucleotide composition before filtering. Contamination with adapter sequences detected in the first 11 base pairs.

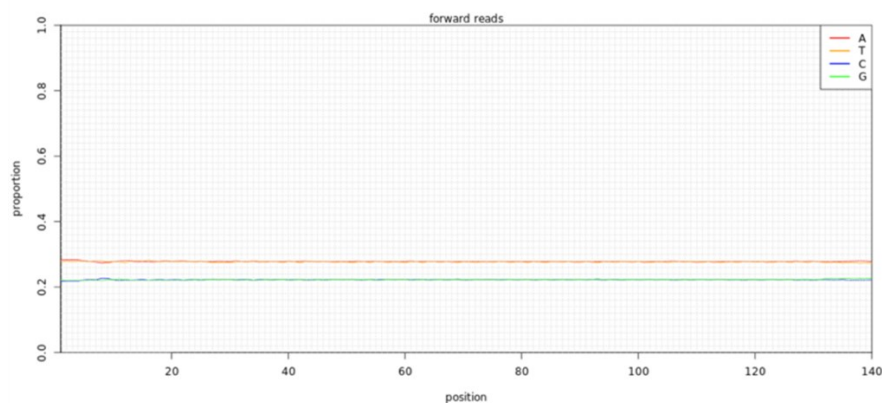


Figure 7. Nucleotide composition after filtering and interlacing.

Table 3. Prospero autumnale plants used for repeat composition analysis using RepeatExplorer2.

Plant ID	Cytotype	Diploid chromosome number (2n)	Genome Size (1C, pg)	Genome size (Mbp)	Locality
H427	B <sup>6</sup> B <sup>6</sup>	12	6.38	6253	Greece, Crete
H424	B <sup>7</sup> B <sup>7</sup>	14	4.34	4253	Montenegro
H14	B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup> Group I	26	11.81	11574	Greece, Crete

The overlap of mates in pairs was analyzed using Repeat Explorer Utilities. Ideally, the sequencing library should consist of DNA fragments longer than the sum of the length of the forward and reverse sequence reads. Therefore, the analysis focused on identifying reads with significant overlap, indicative of fragments that are too short for the analysis that can introduce bias in estimating the proportions of repeat families in the genome. However, overlaps can also result from presence of tandemly repeated monomers of a length shorter than read length. Finally, the analysis was executed using 'Repeat Explorer clustering' in RepeatExplorer2, specifying the use of paired-end reads, selecting either single or comparative analysis mode, and ensuring the selection of the appropriate taxon and protein domain databases (Viridiplante 3.0) for automatic repeat annotation. The output of the analysis comprised three files: an HTML report, an archive for downloading, and a log file.

Comparative repeat analysis was conducted on the previously individually analyzed B<sup>6</sup>B<sup>6</sup>, B<sup>7</sup>B<sup>7</sup>, and B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> read datasets. The datasets of individual cytotypes were merged (concatenated) and subjected to repeat clustering, where individual clusters represented the same type of repeat present/shared in different genomes. This approach provides insights into shared repetitive DNA families as well as sample-specific repetitive DNA sequences. All interlaced files of individual samples were uploaded to RepeatExplorer2. The quality of the datasets was assessed using the Fastqc tool. Following that, the Fasta read name affixer tool under Repeat Explorer Utilities was used to modify the read IDs. A prefix (species code) was added to distinguish reads originating from different cytotypes. Prior to merging the datasets, a predetermined number of reads was sampled from each dataset. For the comparative analysis of diploid cytotypes and the allotetraploid 2,000,000 reads were sampled from each dataset. For the comparative analysis of diploid parental cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup>, 500,000 reads were sampled per cytotype. The merging of all datasets was performed using the

Concatenate Dataset tool under Text Manipulation. The final analysis was executed using the merged datasets, specifying the use of paired-end reads.

The files from unpacked archives were used for final repeat annotation. The automatic annotation of each cluster, based on similarity to reference databases and structure of the graphs, was manually verified, and the final annotation was performed, taking into consideration additionally the biology of the elements (Novák et al., 2020). Putative mobile genetic elements were largely identified using a database of conserved protein domains (REXdb) (Novák et al., 2020). Putative satellite DNAs were identified using TAREAN and validated using Dotter (Sonnhammer and Durbin, 1995). Finally, the genome proportions of various repeat types were calculated.

### 3. Results

#### 3.1. Chromosome number and karyotype analysis

Feulgen staining was used to examine the karyotypes of 10 *Prospero autumnale* plants representing  $B^6B^6$  and  $B^7B^7$  cytotypes as well as their diploid and tetraploid hybrids. Following the staining, images of metaphase chromosomes were used to cut out and assemble the karyotypes. Karyotypes of individuals of two diploid cytotypes  $B^6B^6$  ( $2n = 2x = 12$ ; H538) and  $B^7B^7$  ( $2n = 2x = 14$ ; H438), the hybrid  $B^6B^7$  ( $2n = 2x = 13$ ; H237) as well as all Group I plants of allotetraploid  $B^6B^6B^7B^7$  ( $2n = 4x = 25$ , H208;  $2n = 4x = 26$ , H178;  $2n = 4x = 27$ , H207;  $2n = 4x = 28$ , H331), of Group II ( $2n = 2x = 28$ ; H356), Group III ( $2n = 2x = 28$ ; H238) and Group IV allotetraploid ( $2n = 2x = 28$ ; H152) were created.

The karyotype of the  $B^6B^6$  cytotype comprises four pairs of submetacentric chromosomes (1, 2, 3, and 5), one pair of subtelocentric chromosome (4) and one pair of submetacentric fusion chromosome (6; designated as F(6/7)). A secondary constriction (nucleolar organizer region (NOR)) is present in the long arm in both homologues of chromosome 3 (Figure 8A). The diploid complement of the  $B^7B^7$  individual is composed of five pairs of submetacentric chromosomes (1, 2, 3, 5, and 6), one pair of subtelocentric chromosomes (4) and one pair of nearly metacentric chromosomes (7). The secondary constrictions are in the long arm of chromosome 3 (Figure 8B). The karyotype of  $B^6B^7$  consists of six chromosomes derived from the  $B^6B^6$  parental genome and seven chromosomes from the  $B^7B^7$  parental genome. The karyotype structure of chromosomes of  $B^6B^6$  and  $B^7B^7$  origin corresponds to that in the parent genomes. NORs are in the long arms of chromosomes 3 (Figure 8C).

The Group I  $B^6B^6B^7B^7$  allotetraploid individuals can have different chromosome numbers, ranging from  $2n = 4x = 25$  to 28. This variation arises from the presence of different number of fusion chromosomes F(6/7) and free chromosomes 6 and 7. Group I allotetraploids with  $2n = 4x = 25$  chromosomes have three submetacentric fusion chromosomes F(6/7) and only one of each free submetacentric chromosome 6 and nearly metacentric chromosome 7. The rest of the chromosomes are either submetacentric (chromosomes 1, 2, 3 and 5) or subtelocentric (chromosome 4). The NORs are in the long arm of all chromosomes 3 (Figure 8D). The number of fusion chromosomes F(6/7) decreases with the increasing chromosome number in group I individuals. Individuals with  $2n = 4x = 26$  chromosomes have two fusion chromosomes and two of each chromosomes 6 and 7. Individuals with  $2n = 4x = 27$  chromosomes have only one fusion chromosome F(6/7) and three of each free chromosomes 6

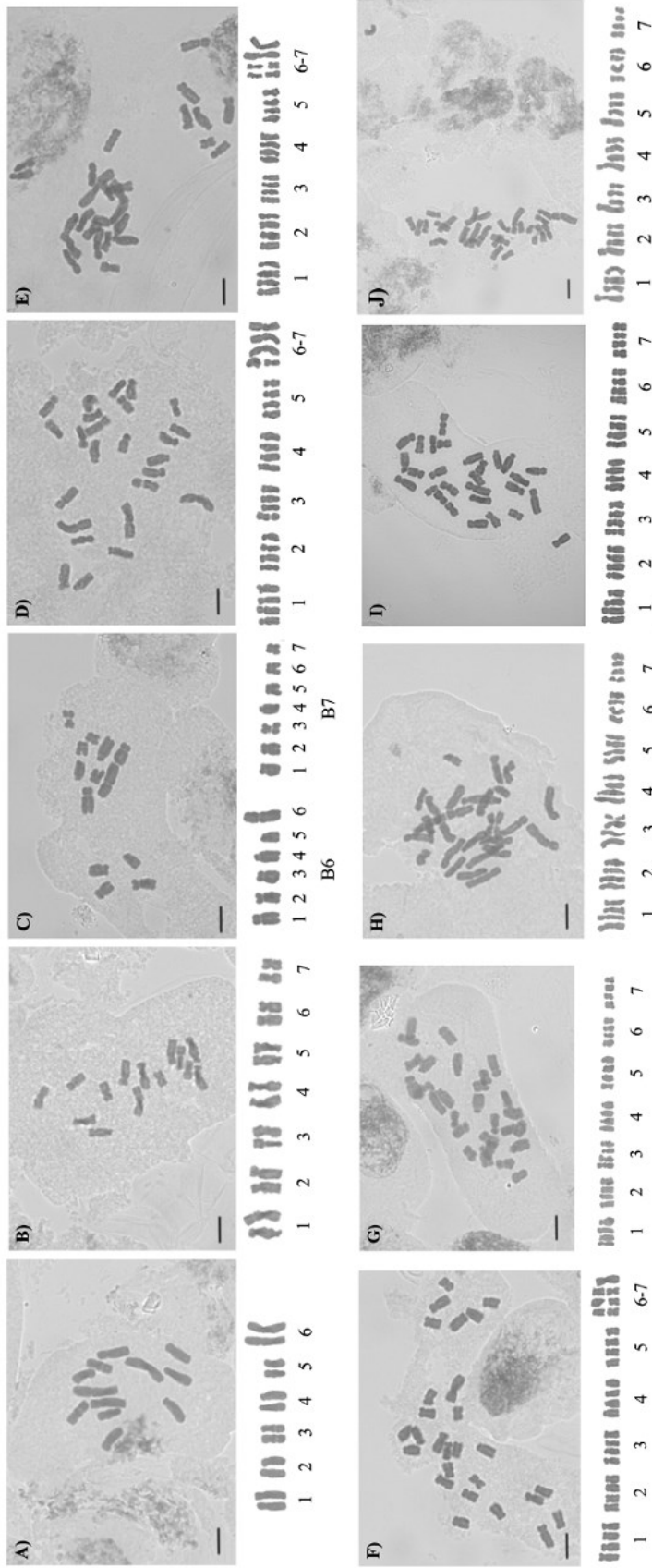


Figure 8. Mitotic metaphase chromosomes and cut-out karyotypes of analyzed individuals of *Prospero autumnale*. **A)** the  $B^6B^6$  cytotype (H538;  $2n = 2x = 12$ ), **B)** the  $B^7B^7$  cytotype (H438;  $2n = 2x = 14$ ), **C)** the  $B^6B^7$  hybrid (H237) of *Prospero autumnale* ( $2n = 2x = 13$ ), **D)** Group I allotetraploid  $B^6B^6B^7B^7$  (H208) of *Prospero autumnale* ( $2n = 4x = 25$ ), **E)** Group I allotetraploid  $B^6B^6B^7B^7$  (H178;  $2n = 4x = 26$ ), **F)** Group I allotetraploid  $B^6B^6B^7B^7$  (H207;  $2n = 4x = 27$ ), **G)** Group I allotetraploid  $B^6B^6B^7B^7$  (H331;  $2n = 4x = 28$ ), **H)** Group II allotetraploid  $B^6B^6B^7B^7$  (H356;  $2n = 4x = 28$ ), **I)** Group III allotetraploid  $B^6B^6B^7B^7$  (H238;  $2n = 4x = 28$ ), **J)** Group IV allotetraploid  $B^6B^6B^7B^7$  (H152;  $2n = 4x = 28$ ). Scale bar = 5  $\mu\text{m}$ .

and 7. Individuals with  $2n = 4x = 28$  chromosomes completely lack the fusion chromosomes and have all four chromosomes 6 and all four chromosomes 7 (Figures 8E-G). In all the individuals of Group I, chromosomes 1, 2, 3, and 5 are submetacentric, while chromosome 4 is subtelocentric. The fusion chromosomes F(6/7) and the free chromosome 6 are submetacentric, whereas the free chromosome 7 is nearly metacentric.

The karyotype of the Group II allotetraploids ( $2n = 4x = 28$ ) consists of two smaller and two larger homoeologues of each chromosome type. Chromosomes 1, 2, 3, 5 and 6 are submetacentric, chromosome 7 is nearly metacentric. Both homoeologues of chromosomes 4 are subtelocentric (Figure 8H). Group III allotetraploids ( $2n = 4x = 28$ ) have three larger and one smaller homoeologues per each chromosome type, whereas Group IV ( $2n = 4x = 28$ ) individuals have one larger and three smaller homoeologues of each chromosome (Figures 8I-8J). The structure of individual chromosome types in Group III and Group IV corresponds to that of Group II. The NORs in all Groups are on the long arms of chromosomes 3.

### 3.2. Localization of satellite DNA and parental genome identification

Three distinct satellite DNA sequences, satDNA *PaB6*, satDNA Pa138, and satDNA Pa147, were localized in chromosomes of multiple individuals of the *Prospero autumnale* complex using fluorescence *in situ* hybridization (FISH). The captured images were used to generate karyograms and ideograms. The satellite DNA sequences were mapped in the diploid cytotypes B<sup>6</sup>B<sup>6</sup> (individuals H160 and H193) and B<sup>7</sup>B<sup>7</sup> (H44 and H47), their hybrid B<sup>6</sup>B<sup>7</sup> (H77, H211, H237 and H258), as well as Group I (H213, H331 and H347), Group II (H354 and H388), Group III (H238) and Group IV (H152) allotetraploids. SatDNA *PaB6* and Pa147 were labelled with biotin and satDNA Pa138 was labeled with digoxigenin. The only exception was in H77, where satDNA *PaB6* was labeled with digoxigenin and satDNA Pa138 with biotin. The chromosomes were counterstained with DAPI. Genomic *in situ* hybridization (GISH) was used to identify the parental chromosomes in the hybrid B<sup>6</sup>B<sup>7</sup>, and allotetraploids Group II-IV.

#### 3.2.1. B<sup>6</sup>B<sup>6</sup> cytotype

SatDNA *PaB6* was detected in the pericentromeric region of all 12 chromosomes (H160). The detected signals were strong and formed larger blocks, with the exception of chromosome 1, where one homologue carried a weaker signal (Figures 9A & 10A).

SatDNA Pa138 was mapped in two individuals, H160 and H193. Dot-like satDNA Pa138 signals were detected in six out of 12 chromosomes in plant H160. Long arms of

chromosomes 1, 2 and 5 carried one locus each (Figures 9A & 10A). SatDNA Pa138 was detected in four of six homologous chromosomes in individual H193 (Figures 9B & 10B). This individual carried single loci on the long arms of chromosomes 1, 2 and 5, and additional locus on the short arm of chromosome 3, close to the centromere.

SatDNA Pa147 was detected in five out of 12 chromosomes in individual H193. One locus each was detected on the long arms of chromosomes 1 and 5 (only in one of the homologues of chromosome 5). Chromosome 3 carried one locus in the short arm, next to the centromere and very close to Pa138 locus (Figures 9B & 10B).

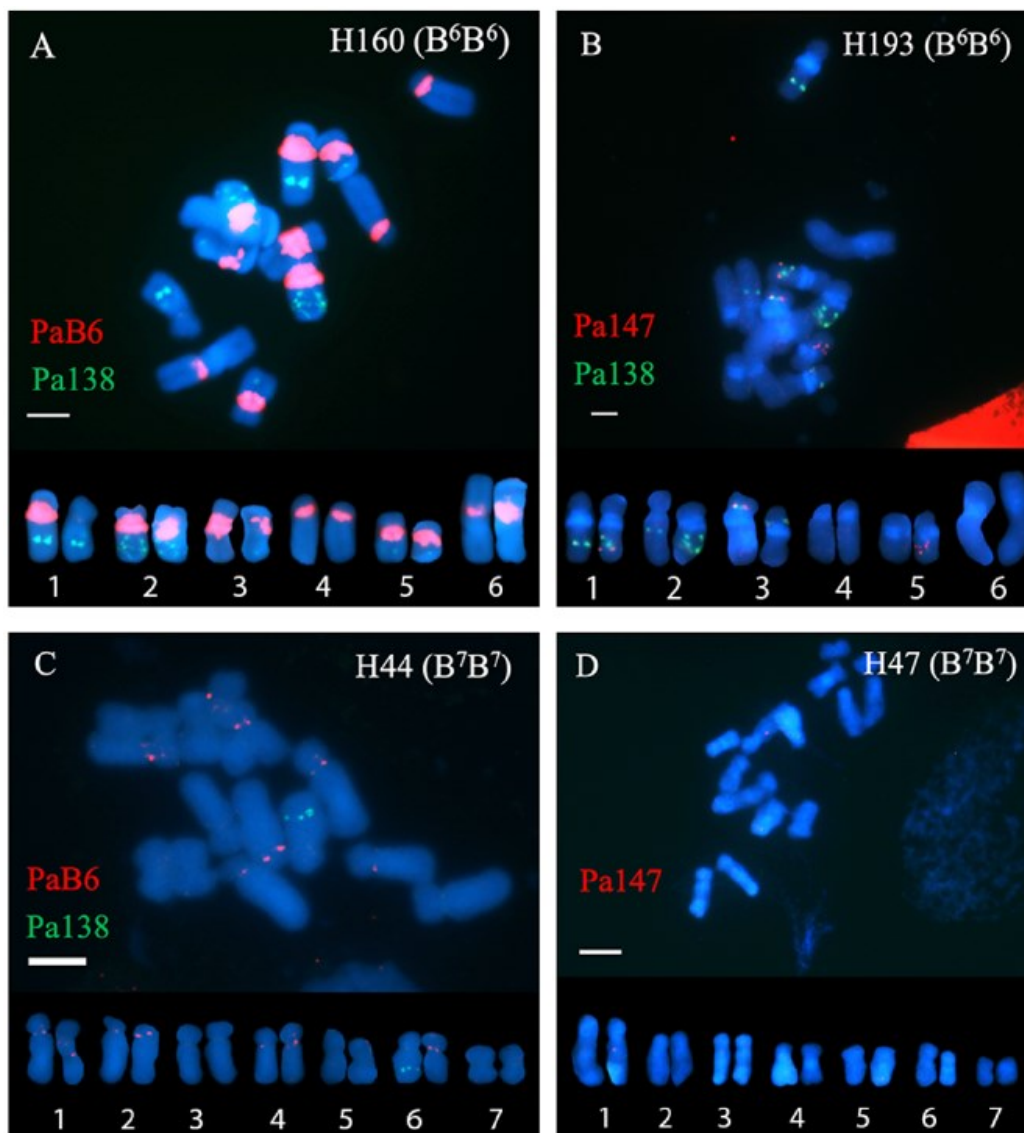


Figure 9. Localization of satDNA PaB6, Pa138 and Pa147 in mitotic metaphase chromosomes of the diploid cytotypes  $B^6B^6$  and  $B^7B^7$ . A: Individual H160: metaphase plate and cut-out karyotype with satDNA PaB6 (red) and satDNA Pa138 (green). B: Individual H193: metaphase plate and karyotype with satDNA Pa138 (green) and Pa147 (red) signals. C: Individual H144: metaphase plate and cut-out karyotype with satDNA PaB6 (red) and satDNA Pa138 (green). D: Individual H47: metaphase plate and cut-out karyotype with satDNA Pa147 (red). Scale bar = 5  $\mu$ m.

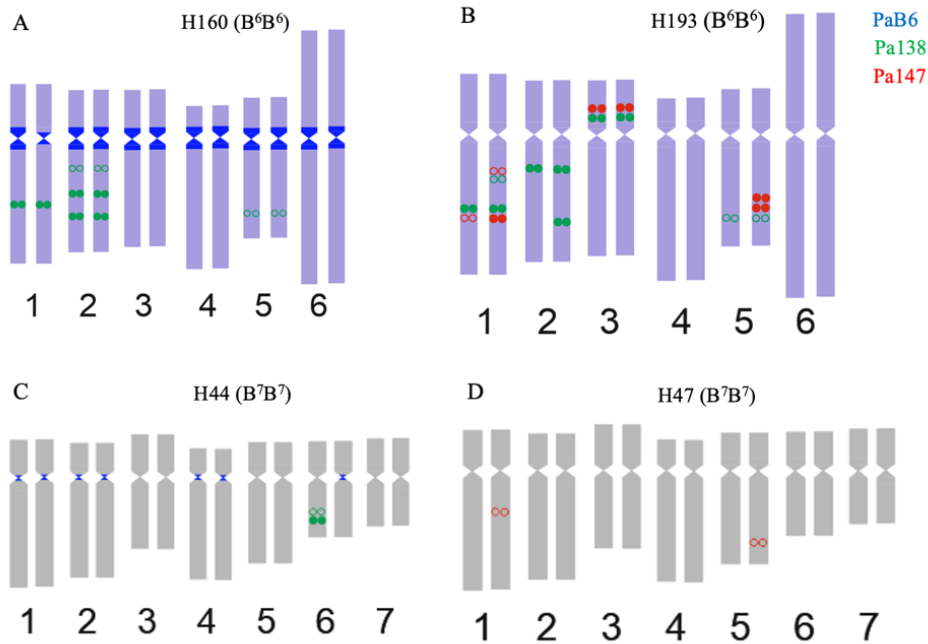


Figure 10. Ideograms of diploid cytotypes  $B^6B^6$  and  $B^7B^7$  (satDNA *PaB6* in dark blue; satDNA *Pa138* in green; satDNA *Pa147* in red; open circle indicates variable signals; full circle indicates constant signals).

### 3.2.2. $B^7B^7$ cytotype

SatDNA *PaB6* was localized in the chromosomes of individual H44. Seven out of 14 chromosomes (1, 4, 5 and one homologue of chromosome 6) carried satDNA *PaB6* loci in the pericentric regions (Figures 9C & 10C). One or two dot-like and very small satDNA *Pa138* signals were detected in only one homologue of chromosome 6 in individual H44 (Figures 9C & 10C). SatDNA *Pa147* was mapped in the individual H47 (Figures 9D & 10D). Only one dot-like satDNA *Pa147* locus each was detected in one homologue of chromosomes 1 and 5. The signals were located in the long arms of the chromosomes.

### 3.2.3. $B^6B^7$ hybrid

SatDNA *PaB6* was mapped in the chromosomes of three homoploid hybrid individuals (H77, H237 and H258; Figures 11A, B, E & 12 A, B, D). In all three individuals six out of 13 chromosomes carried strong signals of satDNA *PaB6* in the pericentromeric regions. In individuals H77 and H237 the signal on chromosome 1 was weaker and dot-like. The fusion chromosome F(6/7) in individual H77 carried an additional larger signal in the long arm. All satDNA *PaB6* signals were detected in chromosomes inherited from the  $B^6$  parental cytotype, as confirmed by GISH (Figure 15A).



SatDNA Pa138 was mapped in the chromosomes of individuals H77, H211, H237 and H258. All signals were detected in chromosomes inherited from the B<sup>6</sup> parental cytotype. In individual H77, four out of 13 chromosomes carried dot-like satDNA Pa138 signals (Figures 11E & 12D). Long arms of chromosomes 1, 2, 3 and 5 carried one locus each. SatDNA Pa138 signals were also detected in the short arms of chromosome 3. In individuals H211 and H237, the same satDNA Pa138 loci were detected as in H77, except for the signals on the long arm of chromosome 3. In total, four out of 13 chromosomes carried satDNA Pa138 in these two individuals (Figures 11B-C & 12 B-C). In individual H258, chromosomes 1 and F(6/7) carried the satDNA Pa138 loci, but in different positions than other individuals. One locus on chromosome 1 was located on the short arm, and two adjacent loci on fusion chromosome were located interstitially within its long arm (Figures 11A & 12A).

SatDNA Pa147 was detected in three out of 13 chromosomes in individual H211 (Figures 11C & 12C). These three chromosomes were inherited from the B<sup>6</sup> parental cytotype. One locus was detected in the long arms of chromosome 5, two loci in long arm of chromosome 2 and one in the short arm of chromosome 3.

#### 3.2.4. B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> Group I

SatDNA *PaB6* was detected in the chromosomes of individuals H347 and H213 (Figures 13A-B & 14A-B). All 28 chromosomes carried satDNA *PaB6* loci in the pericentric regions in both individuals. In individual H213, an additional signal in the short arm of one homologue of chromosome 3 was detected. In the B-chromosome (individual H213), satDNA *PaB6* signals were distributed across the whole chromosome.

In individual H213, 14 out of 27 chromosomes carried satDNA Pa138 loci (Figures 13A & 14A). Long arms of one homologue of chromosome 3, two homologues of chromosome 4, all four chromosomes 5, all three chromosomes 6 and one homologue of chromosome 7 carried satDNA Pa138 loci. In three homologues of chromosome 2, satDNA Pa138 loci were detected in the short arms, close to the centromere. SatDNA Pa138 was detected in 13 out of 28 chromosomes in individual H347 (Figures 13B & 14B). All detected signals were dot-like and all were localized in the long arms of chromosomes 2, 5, 6 and one homologue of chromosome 7.

SatDNA Pa147 was mapped in the chromosomes of individuals H347 (Figures 13C & 14A). Dot-like signals were located in 12 out of 28 chromosomes, all in the long arms of one homologue of chromosome 1, two homologues of chromosome 4, chromosome 5, three homologues of chromosome 6 and two homologues of chromosome 7.

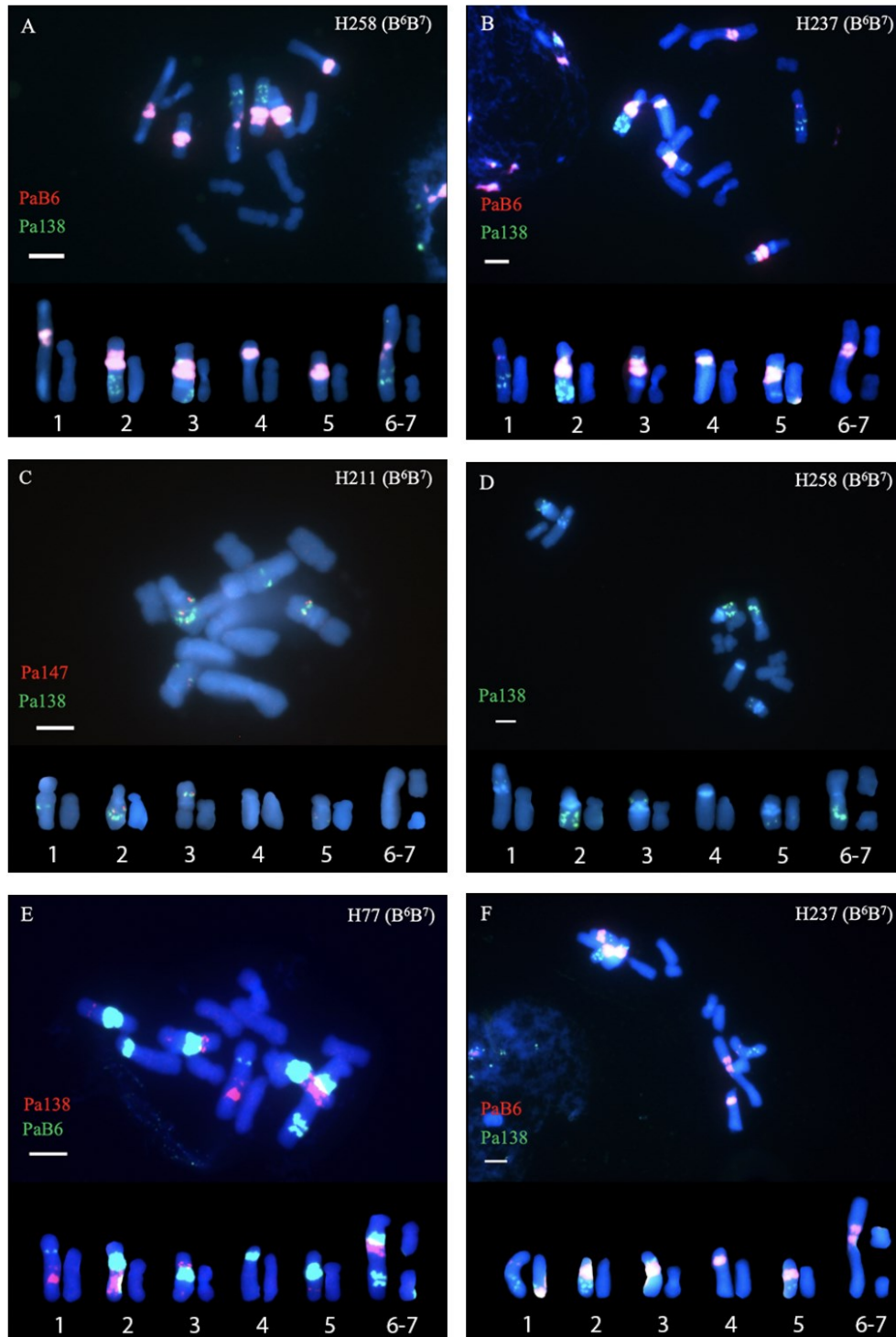


Figure 11. Localization of satDNA PaB6, Pa138 and Pa147 in mitotic metaphase chromosomes of the diploid hybrid  $B^6B^7$ . A: Individual H258: metaphase mitotic chromosomes and karyotype with satDNA PaB6 (red) and satDNA Pa138 (green) signals. B: Individual H237: metaphase mitotic chromosomes and karyotype with satDNA PaB6 (red) and satDNA Pa138 (green) signals. C: Individual H211: metaphase mitotic chromosomes and karyotype with satDNA Pa147 (red) and satDNA Pa138 (green) signals. D: Individual H258: metaphase mitotic chromosomes and karyotype with satDNA Pa138 (green) signals. E: Individual H77: metaphase mitotic chromosomes and karyotype with satDNA PaB6 (green) and satDNA Pa138 (red) signals. F: Individual H237: metaphase mitotic chromosomes and karyotype with satDNA PaB6 (red) and satDNA Pa138 (green) signals. Scale bar = 5  $\mu$ m.

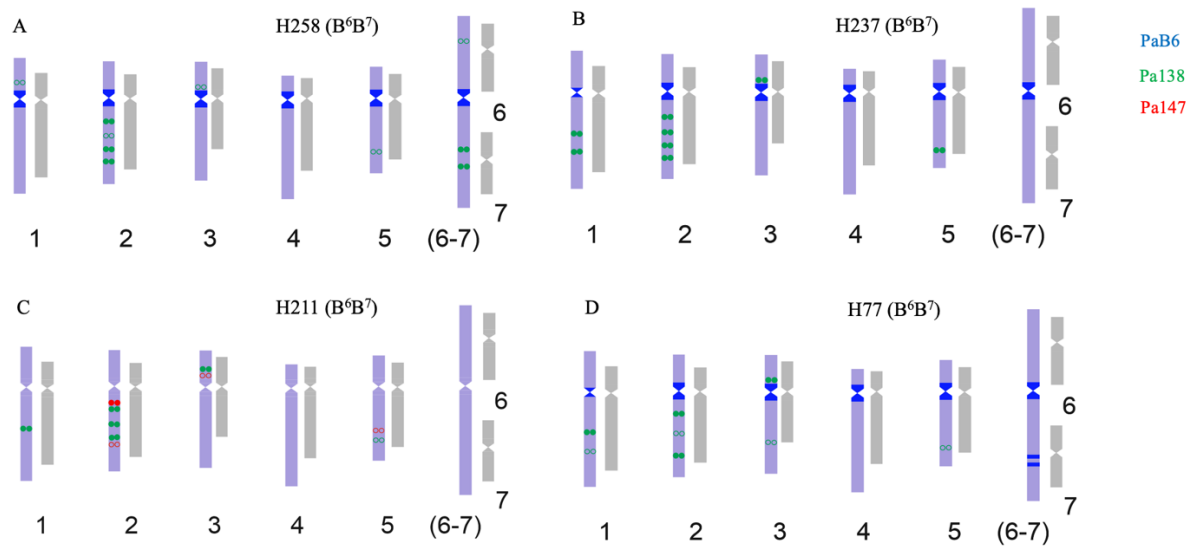


Figure 12. Ideograms of diploid hybrid  $B^6B^7$  individuals (*satDNA* PaB6 in dark blue; *satDNA* Pa138 in green; *satDNA* Pa147 in red; open circle indicates variable signals; full circle indicates constant signals).

### 3.2.5. $B^6B^6B^7B^7$ Group II

*SatDNA* PaB6 was detected in the pericentric regions of 14 out of 28 chromosomes in individual H354 (Figures 13D & 14C). These chromosomes were inherited from the Group I parental cytotype (Figure 15B).

*SatDNA* Pa138 was mapped in the chromosomes of individuals H354 and H388 (Figures 13D-E & 14C-D). Long arms of ten chromosomes in H354 carried *satDNA* Pa138 loci. *SatDNA* Pa138 was detected in one homologue of chromosome 2, two homologues of chromosome 4, chromosome 5, three homologues of chromosome 6. In individual H388, *satDNA* Pa138 loci were detected in nine chromosomes. Two homologues of chromosome 2, chromosome 5 and three homologues of chromosome 6. All signals were located in the long arms of the chromosomes and most in chromosomes of Group I origin, with only chromosomes 5 and 6 of B7 origin carrying small signals of this satellite DNA.

*SatDNA* Pa147 was mapped in individual H388 (Figures 13E & 14D). Two homologues of chromosome 1, one homologue of chromosome 3, two homologues of chromosome 5 and one homologue of chromosome 6 carried *satDNA* Pa138 signals. All signals, except for those on chromosome 3, were on the long arm of the chromosomes. Both parental chromosomes carried the loci in different configurations.

### 3.2.6. B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> Group III

SatDNA *PaB6* was detected in 21 out of 28 chromosomes, all of which were inherited from the Group I parental cytotype (H238; Figures 13F, 14E, 15C). The signals were localized all in the pericentric regions of the chromosomes.

SatDNA Pa138 was detected in 11 out of 28 chromosomes (Figures 13F & 14E). Signals were detected in all four chromosomes 2, three homologues of chromosome 5 (of Group II origin), three homologues of chromosome 6 (of Group II origin) and one homologue of chromosome 7 (of Group II origin). All signals were dot-like and in the long arms of the chromosomes.

SatDNA Pa147 was detected in seven chromosomes. All signals were detected in the long arms of the chromosomes (Figures 13G & 14E). Signals were located in two of chromosomes 1 (one in each Group II and B<sup>7</sup> parental origin), two chromosomes 4 (one in each Group II and B<sup>7</sup> parental origin) and three chromosomes 6 (of Group II origin).

### 3.2.7. B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> Group IV

SatDNA *PaB6* was detected in seven out of 28 chromosomes inherited from the Group II parental cytotype (H152; Figures 13H, 14F, 15D). The signals were located in the pericentric regions. SatDNA Pa138 loci were detected in one homologue of chromosome 1 (B<sup>7</sup> origin), one homologue of chromosome 2 (Group II origin) and one homologue of chromosome 6 (Group II origin; Figures 13I & 14F). SatDNA Pa147 loci were detected in two chromosomes 1 (one of Group II and one of B<sup>7</sup> origin), one chromosome 2 and one chromosome 4 (both of B<sup>7</sup> origin). The signals were located in the long arms of the chromosomes, except for chromosome 4 (Figures 13I & 14F).

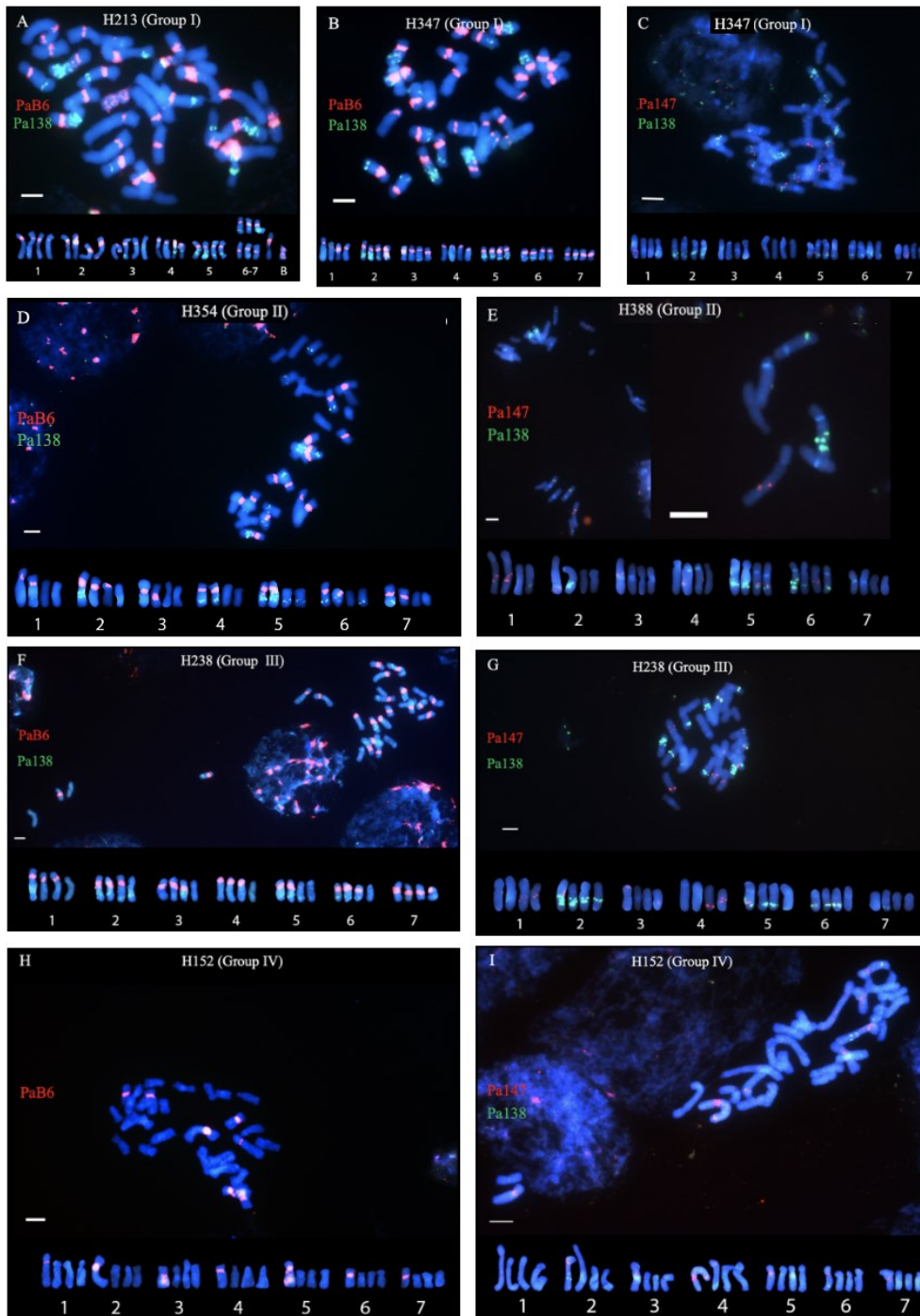


Figure 13. Localization of *satDNA* PaB6, Pa138 and Pa147 in mitotic metaphase chromosomes of Groups I-IV. A: Individual H213: metaphase mitotic chromosomes and karyotype with *satDNA* PaB6 (red) and *satDNA* Pa138 (green) signals. B: Individual H347: metaphase mitotic chromosomes and karyotype with *satDNA* PaB6 (red) and *satDNA* Pa138 (green) signals. C: Individual H347: metaphase mitotic chromosomes and karyotype with *satDNA* Pa147 (red) and *satDNA* Pa138 (green) signals. D: Individual H354: metaphase mitotic chromosomes and karyotype with *satDNA* PaB6 (red) and *satDNA* Pa138 (green) signals. E: Individual H388: metaphase mitotic chromosomes and karyotype with *satDNA* Pa147 (red) and *satDNA* Pa138 (green) signals. F: Individual H238: metaphase mitotic chromosomes and karyotype with *satDNA* PaB6 (red) and *satDNA* Pa138 (green) signals. G: Individual H238: metaphase mitotic chromosomes and karyotype with *satDNA* Pa147 (red) and *satDNA* Pa138 (green) signals. H: Individual H152: metaphase mitotic chromosomes and karyotype with *satDNA* PaB6 (red) signals. I: Individual H152: metaphase mitotic chromosomes and karyotype with *satDNA* Pa147 (red) and *satDNA* Pa138 (green) signals. Scale bar = 5  $\mu$ m.

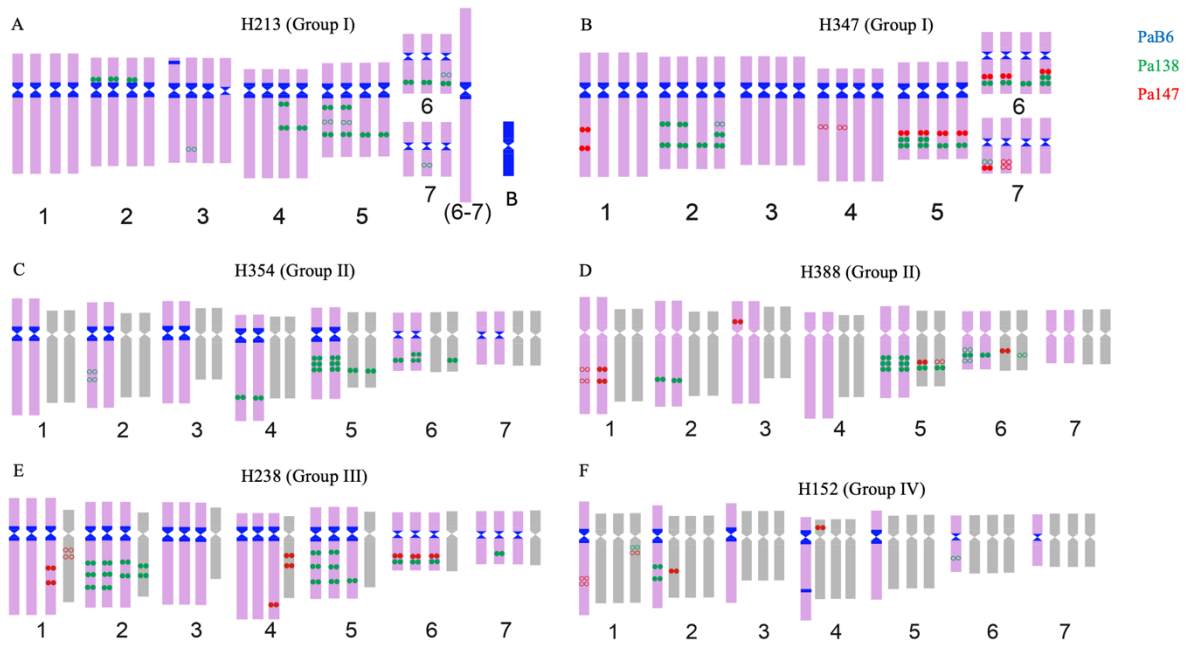


Figure 14. Ideograms of allotetraploids of Groups I-IV (satDNA PaB6 in dark blue; satDNA Pa138 in green; satDNA Pa147 in red; open circle indicates variable signals; full circle indicates constant signals).

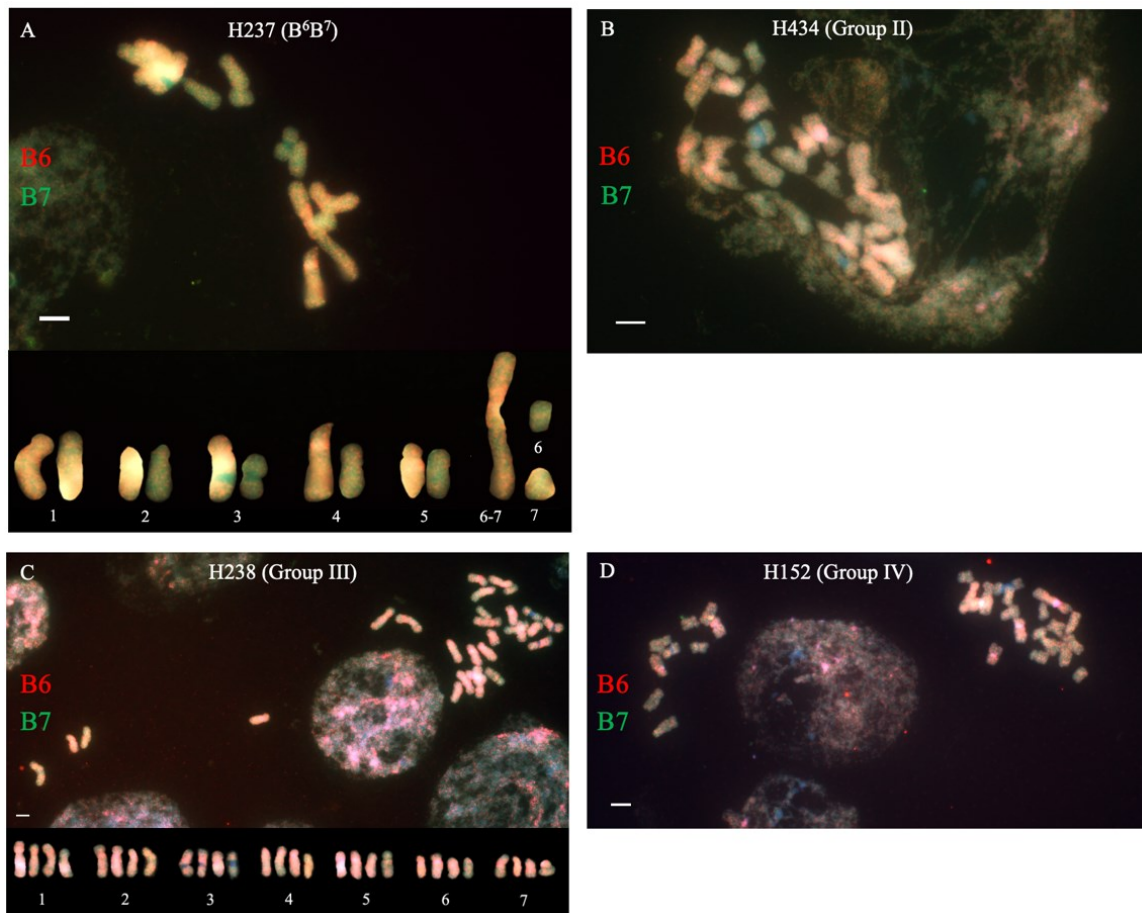


Figure 15. Mitotic metaphase chromosomes and cut-out karyotype after GISH (chromosomes inherited by B<sup>7</sup> parental cytotype in green and chromosomes inherited by the B<sup>6</sup> cytotype in red).

### 3.3. Genome size estimation using flow cytometry

The genome size of three *Prospero autumnale* plants (H44, H493 and H228) was measured using flow cytometry. In order to calculate the genome sizes of H44 and H493, the samples were measured along with the standard organism *Solanum pseudocapsicum*, while the standard organism *Pisum sativum* was used for H228. The G<sub>1</sub> value of the standard organism and sample can be seen in the flow cytometry histogram (Figure 16). The genome size of each plant was measured three times, and the mean was used as the final genome size. The 1C value for H44 was 4.42 pg, while H493 and H228 showed larger genome sizes (8.43 pg and 11.35 pg, respectively; Table 4).

Table 4. Measured genome sizes of H44, H493 and H228 (1C-values).

Cytotype (Plant ID)	2n	Genome size (pg)	Standard Deviation	Genome size (Mbp)	Locality
B <sup>7</sup> B <sup>7</sup> (H44)	14	4.42	0.0608	4332	Ukraine, Crimea
B <sup>7</sup> B <sup>7</sup> B <sup>7</sup> B <sup>7</sup> (H493)	28	8.43	0.1439	8262	Montenegro
B <sup>7</sup> B <sup>7</sup> B <sup>7</sup> B <sup>7</sup> B <sup>7</sup> B <sup>7</sup> (H228)	42	11.35	0.0781	11 123	Greece, mainland

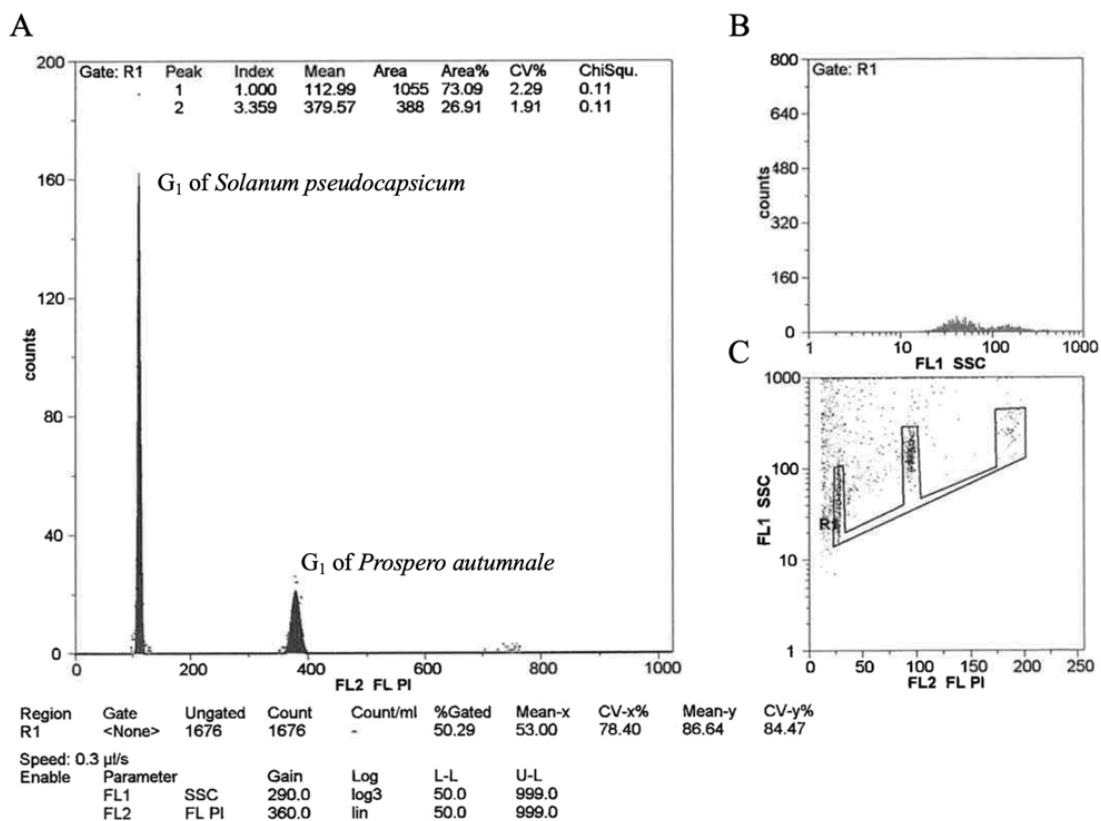


Figure 16. Flow cytometry histogram of *Prospero autumnale* cytotype B<sup>7</sup>B<sup>7</sup> (H44) depicting the two peaks for the G<sub>1</sub> nuclei of *Solanum pseudocapsicum* and *Prospero autumnale* (A), The side scatter plot (B) and plot depicting the side scatter vs. the cytogram fluorescence (C).

### 3.4. Analysis of repetitive DNA using RepeatExplorer2

The repeatomes of two distinct diploid cytotypes of *Prospero autumnale*, B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup>, as well as an allotetraploid B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> of Group I were individually examined using RepeatExplorer2. To validate the presence of satellite DNAs, dot plots were generated using Dotter.

#### 3.4.1. *Prospero autumnale* cytotype B<sup>6</sup>B<sup>6</sup>

In total, 1 568 434 reads (0.07x coverage) were analyzed. Of these, 14 932 reads were identified as plastid DNA or contamination and were excluded from the subsequent analysis of repetitive DNA types. Among the remaining 1 553 502 reads, 75.72% were assigned to the Top clusters, while 7% were assigned to small clusters and 17% were singlets (Figure 17 and Table 5). Retrotransposons comprised 70% of the identified repetitive DNA, with 41.14% of reads assigned to the Ty1-*copia* retrotransposon superfamily, 27.8% to the Ty3-*gypsy* retrotransposon superfamily, and 1.06% representing other LTR retrotransposons. The three most abundant retrotransposon families within the repeatome were Ty1-*copia* families Ikeros (18.55%) and Tork (14.03%), and Ty3-*gypsy* family Tekay (13.81%; Figure 18).

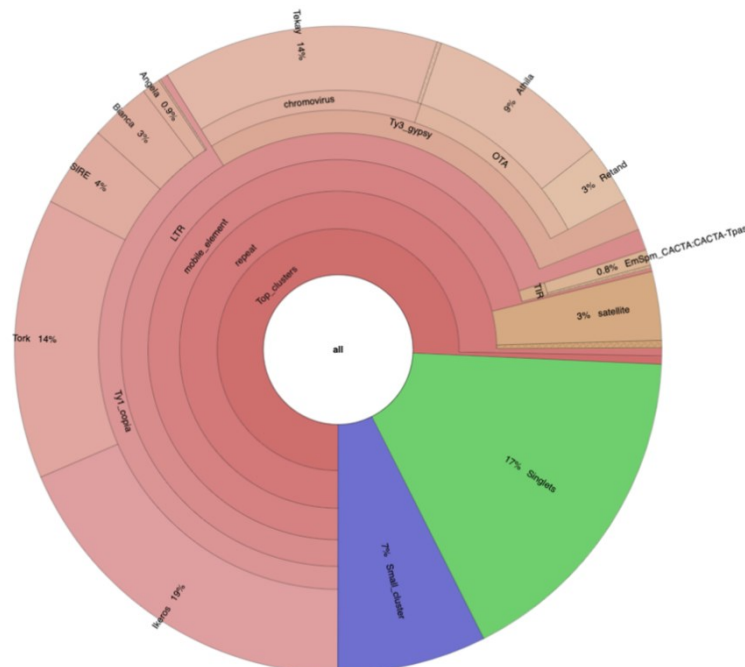


Figure 17. Multi-level pie chart illustrating the proportions (in %) of various repetitive DNA types identified in the repeatome analysis of the diploid cytotype B<sup>6</sup>B<sup>6</sup> of *Prospero autumnale*. The image was created using Krona.



Table 5. Proportions (in %) of repetitive DNA types in the diploid cytotype B<sup>6</sup>B<sup>6</sup> of *Prospero autumnale*.

Type	Superfamily	Family	Percentage %	No. of reads
Retrotransposons			<b>70.00</b>	<b>1087464</b>
	Ty3- <i>gypsy</i>		<u>27.80</u>	<u>431922</u>
		Reina	0.22	3419
		Tekay	13.81	214569
		Athila	9.20	142854
		Retand	2.97	46176
		Unclassified	1.60	24904
	Ty1- <i>copia</i>		<u>41.14</u>	<u>639066</u>
		Ale	0.11	1650
		Angela	0.93	14385
		Bianca	3.04	47277
		Ikeros	18.55	288163
		SIRE	4.11	63805
		Tork	14.03	217959
		Unclassified	0.38	5827
	other/non-LTR		<u>1.06</u>	<u>16476</u>
		Other LTR	1.06	16476
DNA transposons			<b>0.93</b>	<b>14492</b>
	Subclass I		<u>0.93</u>	<u>14492</u>
		EnSpm_CACTA	0.79	12239
		MuDR_Mutator	0.08	1239
		PIF_Harbinger	0.03	499
		MITE	0.03	515
Tandem repeats			<b>3.79</b>	<b>58849</b>
	Satellite DNA		<u>3.42</u>	<u>53164</u>
	rDNA		<u>0.37</u>	<u>5685</u>
		45S rDNA	0.37	5685
Unclassified repeats			<b>0.59</b>	<b>9134</b>
<b>Total repeats</b>			<b>75.31</b>	<b>1169939</b>
Unclassified			<b>0.41</b>	<b>6343</b>
<b>Small Clusters</b>			<b>7.43</b>	<b>115427</b>
<b>Singlets</b>			<b>16.85</b>	<b>261793</b>
<b>TOTAL nuclear genome</b>			<b>100</b>	<b>1553502</b>

DNA transposons were only detected in a significantly lower proportion (0.93%), with the majority attributed by the EnSpm-CACTA family (0.79%). Among tandem repeats, 3.42% of the analyzed reads were identified as satellite DNAs, while 0.37% corresponded to rDNAs. A further 0.59% of the reads could not be classified into specific repetitive DNA types, and 0.41% were unclassified reads in Top clusters.

RepeatExplorer2 assigned potential satellite DNAs reads to four distinct clusters. Subsequent verification via Dotter, a tool that generates dot plots by aligning sequences of clusters against themselves, confirmed tandem arrangements of three out of the four putative satellite DNAs. Additionally, Dotter allowed for identification of a fourth satellite DNA in a cluster that was initially not classified as such by RepeatExplorer2. The identified satellite DNAs in this cytotype include satellite DNA *PaB6*, and three novel satellite DNAs, here referred to as Pa138, Pa160 (=identical to satDNA Pa160 in comparative analysis, therefore named Pa160 instead of Pa123) and Pa124 (Figure 19). Among these, satellite DNA PaB6 was most abundant (3.16%; Table 6). The remaining three satellite DNAs were much less abundant (0.03-0.02%).

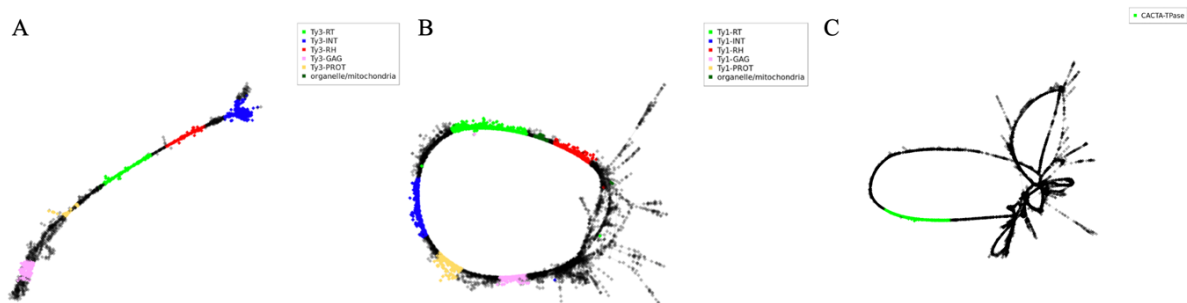
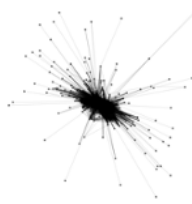
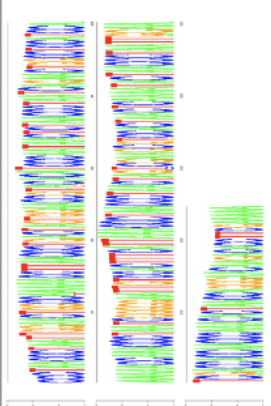

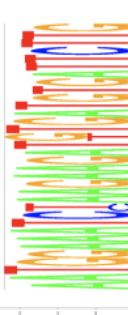


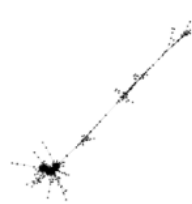


Figure 18. Cluster graphs constructed by RepeatExplorer2 representing the two different retrotransposon families Ty1-copia and Ty3-gypsy and the DNA transposon family EnSpm\_CACTA. A: Retrotransposon family Tekay (Cluster 100) belonging to the superfamily Ty3-gypsy identified by RepeatExplorer2. B: Retrotransposon family Tork (Cluster 64) belonging to the superfamily Ty1-copia identified by RepeatExplorer2. C: DNA transposon family EnSpm\_CACTA. Domains: RT = Reverse Transcriptase, INT = Integrase, RH = RNase H, GAG = Group specific antigen, PROT = Protease.

Table 6. List of putative satellite DNAs identified by Dotter in the diploid cyto type B<sup>6</sup>B<sup>6</sup> of Prospero autumnale.

Satellite DNA	Monomer length	Genome proportion	DNA sequence	Cluster Graph	Sequence logo
<b>PaB6</b>	249 bp	3.16%	CCCTAACCCCTAATTGGAACTGG CCAAAAACCTTAACCCCTGACTA GTGGAAACCGTACACATCCCAAT AACTTAACCCTAATCAGAACT GGCTCCATCCAAAAAACCAA ACCCTAATCGGGGAATTGTCC ACTTTCATTAAACCCCTAACCCCT AACTAGTCGAAGCGTCCACTAA CTCAAAATACTAACCCCTAATTG GAATCGACCACCAACCCCAAC ACTAACCCAGGGAACAAGACAC CTTCGAAAAA		
<b>Pa138</b>	34 bp	0.20%	AATGGAAATCTAAGAAGATGT GATGTGATTCCTTG		
<b>Pa160</b>	213 bp	0.03%	ATATTGGTTTCACTCCGAGTTA GCAAGGAGCCGAGCTAAGATG GCGGCCGTGCCCGTTGCGTT GTACGGAGATGGGTCCACTCC GAGTGAAGCGGGGGACGG ATTTTGAGCAGGCTGAGTCGG TCGATGTACAGAGATGGGTCC ACTCCGAGTGAATGCGACGGA CGAGGCATAAATCCCGCTAGTC GGTCTAATGCTTAACCG		
<b>Pa124</b>	142 bp	0.03%	TCCGAGTTCGGAGCTAAGTCC GGCAGAGTTTTGCGAGCTAG GGCCGATTGCGTTACGGATATA ATTTTCACTCCGAGTTGGCAAT CGGTAGGTACTAGACCCGCCAC GGTTGTTAGTCGCTTACGGATA TAATTTTCAC		

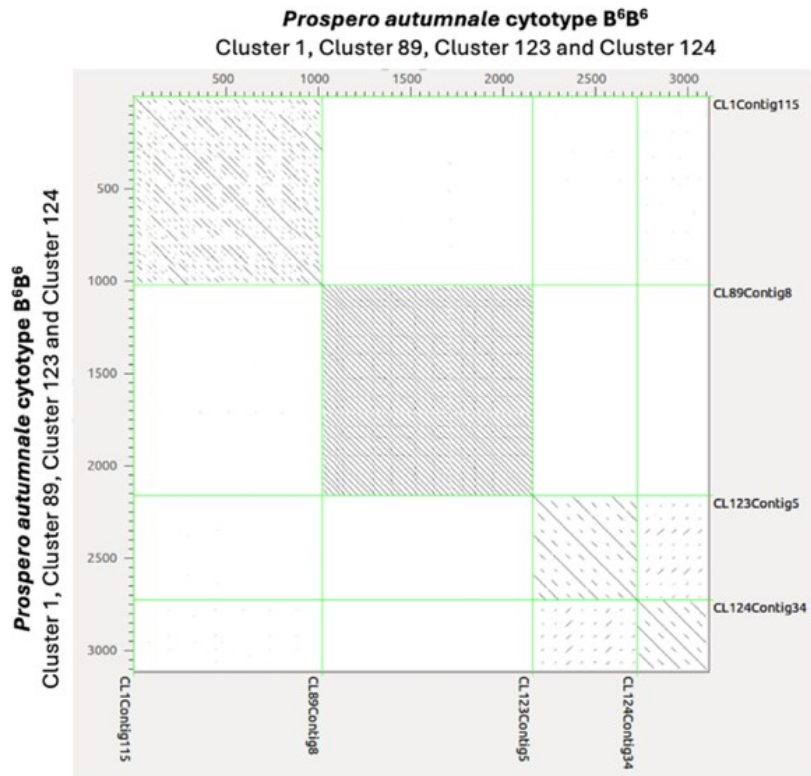


Figure 19. Dot plot of the four putative satellite DNAs identified in the diploid cytotype B<sup>6</sup>B<sup>6</sup> of *Prospero autumnale*. X and Y axes: Most abundant contig sequence of clusters 1, 89, 123 and 124, containing the reads of putative satellite DNAs. The image was created using Dotter.

### 3.4.2. *Prospero autumnale* cytotype B<sup>7</sup>B<sup>7</sup>

In total 3 388 153 reads (0.22x coverage) were analyzed. Of these, 39 341 reads were identified as plastid DNA or contamination and were consequently excluded from the subsequent analysis of repetitive DNA types. Among the remaining 3 348 812 reads, 74.41% were assigned to the Top clusters, while 10% were assigned to small clusters and 16% of the reads represented singlets (Figure 20). Retrotransposons comprised 71.03% of the identified repetitive DNA, with 42.31% of the reads assigned to the Ty1-*copia* retrotransposon superfamily, 25.73% to the Ty3-*gypsy* retrotransposon superfamily and 2.99% representing other LTR or non-LTR retrotransposons. Tekay (Ty3-*gypsy* 11.03%), Ikeros (Ty1-*copia*, 17.78%) and Tork (Ty1-*copia*, 17.10%) were the most prevalent retrotransposon families within the repeatome. Additionally, 1.49% of the analyzed reads were identified as DNA transposons, with EnSpm-CACTA being the most prevalent family (0.96%). Analyzed reads classified as tandem repeats represented only rDNA reads (0.50%), no satDNAs were identified. 0.98% of the reads could not be classified into specific repetitive DNA types, and 0.41% were unclassified top cluster reads (Table 7).

While RepeatExplorer2 analysis identified three clusters as putative satellite DNA reads, further examination of the RepeatExplorer2 results and analysis with Dotter did not confirm the presence of any satellite DNAs in the diploid cytotype B<sup>7</sup>B<sup>7</sup> of *Prospero autumnale*.

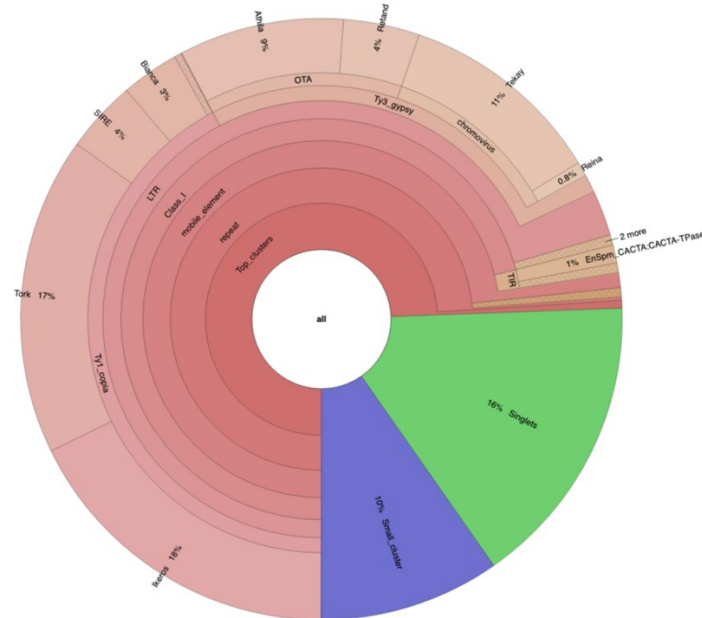


Figure 20. Multi-level pie chart illustrating the proportions (in %) of various repetitive DNA types identified in the repeatome analysis of the diploid cytotype B<sup>7</sup>B<sup>7</sup> of *Prospero autumnale*. The image was created using Krona.

### 3.4.3. *Prospero autumnale* allotetraploid B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup>

In total 2 975 029 (0.07x coverage) reads were analyzed. Of these, 31 783 reads were identified as plastid DNA or contamination and were consequently excluded from the subsequent analysis. 80.86% of the remaining 2 943 246 reads were attributed to Top clusters, while 11% were small cluster reads and 9% were singlets (Figure 21). In the allotetraploid of *Prospero autumnale*, 69.69% of the reads were identified as retrotransposons. 37.28% of these reads were classified as Ty1-*copia* and 18.59% as Ty3-*gypsy*. Additionally, 10.73% were classified as other LTR or non-LTR retrotransposons, while 3.08% could not be further classified into specific families. Only 0.25% of the reads could be assigned to DNA transposons. Satellite DNAs collectively constituted 0.49% of the genome. Moreover, 1.59% of the reads were assigned to rDNAs, 2.23% could not be further classified into repetitive DNA families, and 6.61% were unclassified top cluster reads (Table 8).

Table 7. Proportions (in %) of repetitive DNA types in the diploid cytotype B<sup>7</sup>B<sup>7</sup> of *Prospero autumnale*.

Type	Superfamily	Family	Percentage	No. of reads
Retrotransposons			<b>71.03</b>	<b>2378706</b>
	Ty3-gypsy		<u>25.73</u>	<u>861579</u>
		Reina	0.77	25754
		Tekay	11.03	369315
		Athila	8.86	296799
		Retand	4.10	137140
		Unclassified	0.97	32571
	Ty1-copia		<u>42.31</u>	<u>1416801</u>
		Ale	0.19	6232
		Angela	0.17	5638
		Bianca	3.11	104268
		Ikeros	17.78	595488
		SIRE	3.89	130119
		Tork	17.10	572793
		TAR	0.02	753
		Unclassified	0.05	1510
	other/non-LTR		<u>2.99</u>	<u>100326</u>
		LINE	0.01	373
		Other LTR	2.98	99953
DNA transposons			<b>1.49</b>	<b>49918</b>
	Subclass I		<u>1.49</u>	<u>49918</u>
		EnSpm_CACTA	0.96	32238
		hAT	0.02	745
		MuDR_Mutator	0.26	8704
		PIF_Harbinger	0.16	5441
		MITE	0.08	2790
Tandem repeats			<b>0.50</b>	<b>16780</b>
	rDNA		<u>0.50</u>	<u>16780</u>
		45S rDNA	0.50	16780
Unclassified repeats			<b>0.98</b>	<b>32901</b>
<b>Total repeats</b>			<b>74.00</b>	<b>2478305</b>
Unclassified			<b>0.41</b>	<b>13636</b>
<b>Small Clusters</b>			<b>9.71</b>	<b>325301</b>
<b>Singlets</b>			<b>15.87</b>	<b>531570</b>
<b>TOTAL nuclear genome</b>			<b>100</b>	<b>3348812</b>

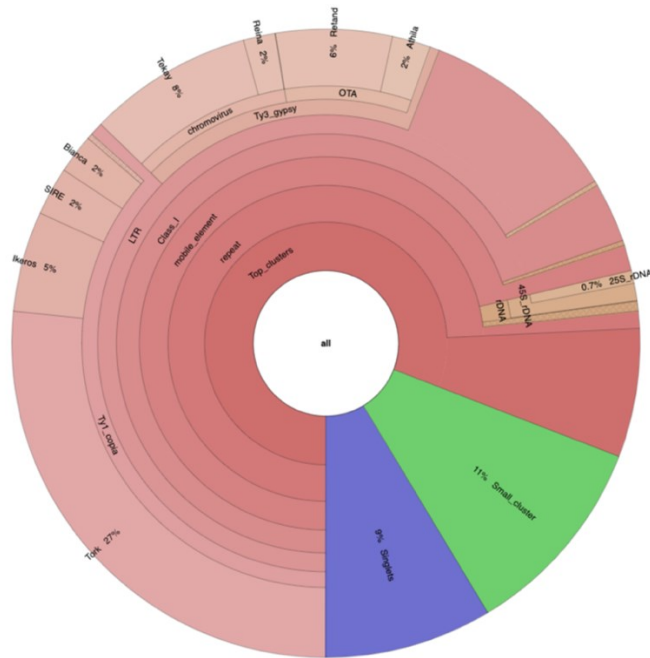


Figure 21. Multi-level pie chart illustrating the proportions (in %) of various repetitive DNA types identified in the repeatome analysis of the allotetraploid  $B^6B^6B^7B^7$  of *Prospero autumnale*. The image was created using Krona.

RepeatExplorer2 assigned potential putative satellite DNAs reads to five distinct clusters. Verification through Dotter confirmed tandem arrangements of two out these putative satellite DNAs. The identified satellite DNAs in this cytotype include satellite DNA *PaB6* and one novel satellite DNA referred to as Pa204 (=identical to satDNA Pa204 in comparative analysis, thus named Pa204 instead of Pa249; Figure 22). The most abundant satellite DNA was satellite DNA *PaB6* with 0.47% of the total 0.49% satellite DNAs identified (Table 9).

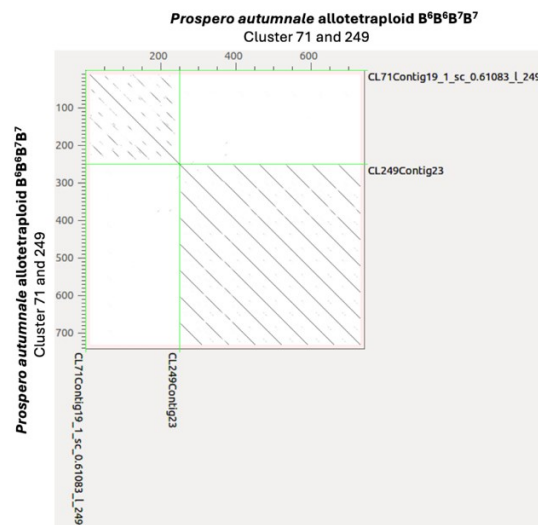


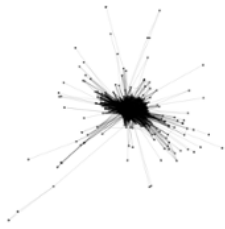
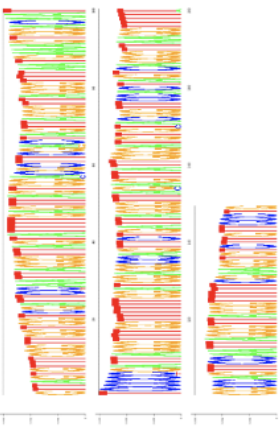
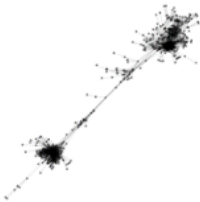
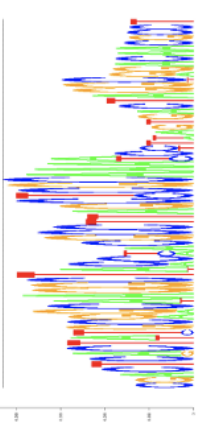
Figure 22. Dot plot of the two putative satellite DNAs identified in the allotetraploid  $B^6B^6B^7B^7$  of *Prospero autumnale*. X and Y axes: Most abundant contig sequence of clusters 71 and 249, containing the reads of putative satellite DNAs. The image was created using Dotter.

Table 8. Proportions (in %) of repetitive DNA types in the allotetraploid  $B^6B^6B^7B^7$  of *Prospero autumnale*.

Type	Superfamily	Family	Percentage	No. of reads
Retrotransposons			<b>69.69</b>	<b>2051132</b>
	Ty3-gypsy		<u>18.59</u>	<u>547251</u>
		Reina	1.63	48047
		Tekay	8.49	249871
		Athila	1.98	58193
		Galadriel	0.02	618
		Retand	5.98	176029
		Unclassified	0.49	14493
	Ty1-copia		<u>37.28</u>	<u>1097373</u>
		Ale	0.26	7640
		Angela	0.03	777
		Bianca	2.03	59664
		Ikeros	5.17	152120
		SIRE	2.49	73251
		Tork	26.64	783943
		Unclassified	0.68	19978
	other/non-LTR		<u>10.73</u>	<u>315876</u>
		LINE	0.24	7105
		Other LTR	10.49	308771
	Unclassified		<u>3.08</u>	<u>90632</u>
DNA transposons			<b>0.25</b>	<b>7367</b>
	Subclass I		<u>0.25</u>	<u>7367</u>
		EnSpm_CACTA	0.02	643
		MuDR_Mutator	0.07	2051
		PIF_Harbinger	0.06	1855
		MITE	0.10	2818
Tandem repeats			<b>2.08</b>	<b>61208</b>
	Satellite DNA		<u>0.49</u>	<u>14449</u>
	rDNA		<u>1.59</u>	<u>46759</u>
		5S rDNA	0.04	1125
		45S rDNA	1.55	45634
Unclassified repeats			<b>2.23</b>	<b>65558</b>
<b>Total repeats</b>			<b>74.25</b>	<b>2185265</b>
Unclassified			<b>6.61</b>	<b>194687</b>
<b>Small Clusters</b>			<b>10.52</b>	<b>309718</b>
<b>Singlets</b>			<b>8.62</b>	<b>253576</b>
<b>TOTAL nuclear genome</b>			<b>100</b>	<b>2943246</b>



Table 9. List of putative satellite DNAs identified in the allotetraploid  $B^6B^6B^7B^7$  of *Prospero autumnale*.

Satellite DNA	Monomer length	Genome proportion	DNA Sequence	Cluster Graph	Sequence logo
<b>PaB6</b>	249 bp	0.47%	<p>GTTAGTGTGGGGTTGGTGGTCCG            ATTCCAAATTAGGGTTAGTATTTT            GAGTTAGTGGACGCTTCGACTA            GTTAGGGTTAGGGTTTAATGAA            AGTGGACAATCCCCCGATTAG            GGTTTGGTTTTTGGATGGAGGC            CAGTTCTGATTAGGGTTAGAGTT            ATTGGGATGTGACGGTCCACT            AGTCAGGGTTAAGGTTTTGGCC            AGTTCCAAATTAGGGTTAGGGTTT            TTCGAAGGTGCTGTGCCCTG</p>		
<b>Pa204</b>	70 bp	0.02%	<p>CGACTCACTATGCGGCAAGGCT            ACCTACGCCTTAGCTCGCAAAA            CCTTAGTGACTAGGCCGGACAAC            GCCT</p>		

### 3.5. Comparative analysis of the repeatome using Repeat Explorer 2

Two separate comparative analyses were conducted using RepeatExplorer2: (1) comparative analysis of the repeatomes of the two diploid cytotypes, B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup> of *Prospero autumnale*, and (2) analysis of repeatomes of the two parental diploid genomes and their allotetraploid progenitor, B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup>. To validate the presence of satellite DNAs, dot plots were generated using Dotter.

#### 3.5.1. Comparative analysis of diploid cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup> of *Prospero autumnale*

500 000 reads from each cytotype (0.02x coverage for B<sup>6</sup>B<sup>6</sup>; 0.03x coverage for B<sup>7</sup>B<sup>7</sup>) were included in the comparative analysis of the repeatomes. Out of the total 1 000000 reads 6 773 reads were identified as plastid DNA or contamination and were excluded from the subsequent analysis. 993 227 reads were analyzed, with 498 152 reads from the B<sup>7</sup>B<sup>7</sup> genome and 495 075 from the B<sup>6</sup>B<sup>6</sup> genome. Genome proportions of the individual clusters and repetitive DNA families were normalized to the genome size of the two analyzed species (Figure 23, Table 10).

Similar proportions of repetitive DNA types were observed in both species for most of the repeat types. The proportions of Ty1-retroelements were very similar in both genomes, whereas Ty3-*gypsy* retrotransposons were in higher proportion in the B<sup>6</sup>B<sup>6</sup> genome (11.73% in B<sup>6</sup>B<sup>6</sup> vs 9.69% in B<sup>7</sup>B<sup>7</sup>; Figure 24) and the number of satellite DNA reads. The genome of the B<sup>6</sup>B<sup>6</sup> cytotype consisted of 1.72% satellite DNA reads, while in the B<sup>7</sup>B<sup>7</sup> cytotype this number was considerably lower (0.02%).

RepeatExplorer2 identified four putative satellite DNA clusters. Subsequent verification through Dotter confirmed the identity of four satellite DNAs. The identified satellite DNAs included satellite DNAs *PaB6*, Pa138, Pa160 and Pa124 (Figure 25). The proportions of the identified satellite DNAs differed significantly between the two cytotypes. In the B<sup>6</sup>B<sup>6</sup> cytotype satellite DNA *PaB6* represented 1.59% of the total genome (Table 11). The other three satellite DNAs only made up 0.13% of the genome. In the cytotype B<sup>7</sup>B<sup>7</sup> the satellite DNAs in total represented only 0.02% of the genome. Satellite DNA Pa138 was not detected in the B<sup>7</sup>B<sup>7</sup> cytotype. Using Dotter to plot the most abundant contig sequence of each cluster identified as satDNA showed that the four satellite DNAs correspond to the previously detected satellite DNAs in the cytotype B<sup>6</sup>B<sup>6</sup> (Figure 25).

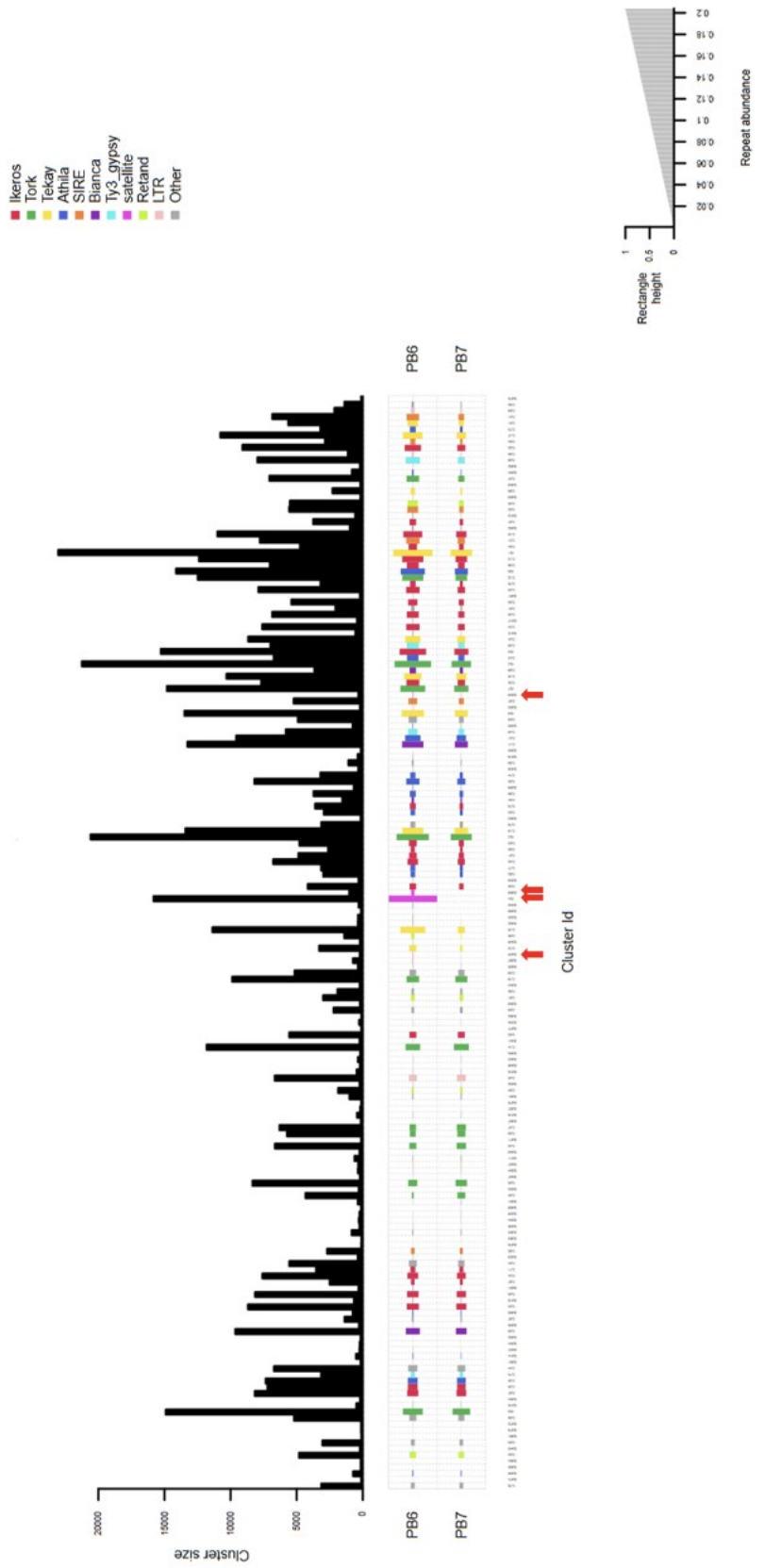


Figure 23. Summary of the comparative analysis of the repeats of the diploid cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup> of Prospero autumale depicting the number of reads in the individual clusters (top graph) and the genome proportion of the reads in the genome of the cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup>. PB6: Prospero autumale cytotypotype B<sup>6</sup>B<sup>6</sup>. PB7: Prospero autumale cytotypotype B<sup>7</sup>B<sup>7</sup>. Red arrows indicate clusters identified as satellite DNAs (Cluster 4, Cluster 100, Cluster 129 and Cluster 145).

Table 10. Proportions (in %) of repetitive DNA types in the comparative analysis of *Prospero autumnale* cytotypes  $B^6B^6$  and  $B^7B^7$ .

Type	Superfamily	Family	Genome proportion $B^6B^6$	No. of reads $B^6B^6$	Genome proportion $B^7B^7$	No. of reads $B^7B^7$
			$B^6B^6$		$B^7B^7$	
Retrotransposons			<b>32.46</b>	<b>322419</b>	<b>30.63</b>	<b>304190</b>
	Ty3-gypsy		<u>11.73</u>	<u>116459</u>	<u>9.69</u>	<u>96269</u>
		Reina	0.04	413	0.06	580
		Tekay	5.90	58632	4.41	43777
		Athila	3.61	35854	3.29	32628
		Retand	0.86	8554	0.84	8372
		Unclassified	1.31	13006	1.10	10912
	Ty1-copia		<u>20.24</u>	<u>201062</u>	<u>20.36</u>	<u>202180</u>
		Ale	0.05	476	0.07	711
		Angela	0.43	4276	0.54	5383
		Bianca	1.45	14375	1.38	13700
		Ikeros	9.39	93303	8.73	86662
		SIRE	1.81	17968	1.39	13799
		Tork	6.69	66443	7.86	78106
		Unclassified	0.42	4221	0.38	3819
	other/non-LTR		<u>0.49</u>	<u>4898</u>	<u>0.58</u>	<u>5741</u>
		Other LTR	0.49	4898	0.58	5741
DNA transposons			<b>0.43</b>	<b>4269</b>	<b>0.55</b>	<b>5428</b>
	Subclass I		<u>0.43</u>	<u>4269</u>	<u>0.55</u>	<u>5428</u>
		EnSpm_CACTA	0.38	3818	0.48	4778
		MuDR_Mutator	0.02	162	0.03	263
		PIF_Harbinger	0.02	159	0.02	156
		MITE	0.01	130	0.02	231
Tandem repeats			<b>1.91</b>	<b>18928</b>	<b>0.29</b>	<b>2832</b>
	Satellite DNA		<u>1.72</u>	<u>17127</u>	<u>0.02</u>	<u>244</u>
	rDNA		<u>0.18</u>	<u>1801</u>	<u>0.26</u>	<u>2588</u>
		45S rDNA	0.18	1801	0.26	2588
Unclassified repeats			<b>0.47</b>	<b>4690</b>	<b>0.60</b>	<b>5961</b>
Unclassified			<b>1.01</b>	<b>10010</b>	<b>1.08</b>	<b>10733</b>
<b>Total Top Clusters</b>			<b>36.28</b>	<b>360316</b>	<b>33.14</b>	<b>329144</b>

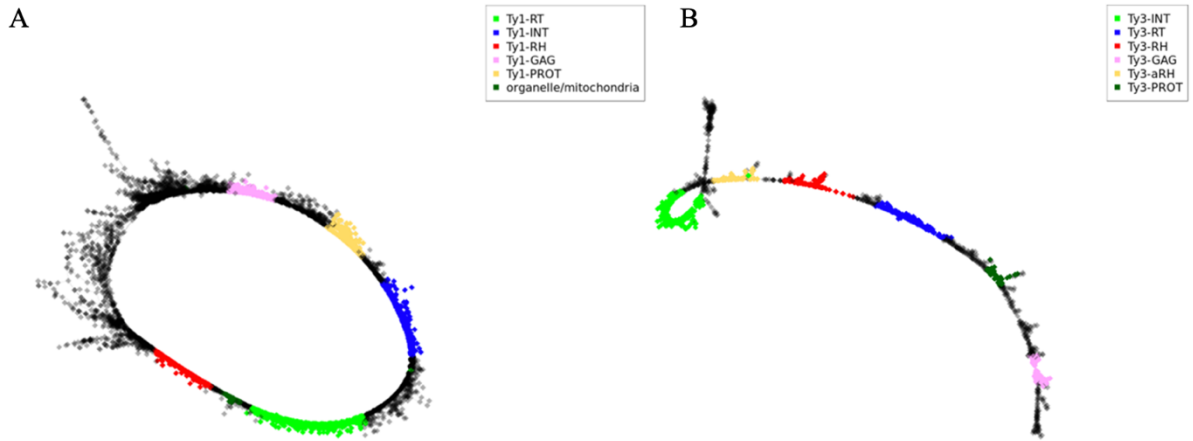


Figure 24. Cluster graphs constructed by RepeatExplorer2 visualizing two different retrotransposon families and their protein domains. Domains: RT = Reverse Transcriptase, INT = Integrase, RH = RNase H, GAG = Group specific antigen, PROT = Protease. A: Retrotransposon family Tork (Cluster 47) belonging to the superfamily Ty1-copia identified by RepeatExplorer2. B: Retrotransposon family Retand (Cluster 95) belonging to the superfamily Ty3-gypsy identified by RepeatExplorer2.

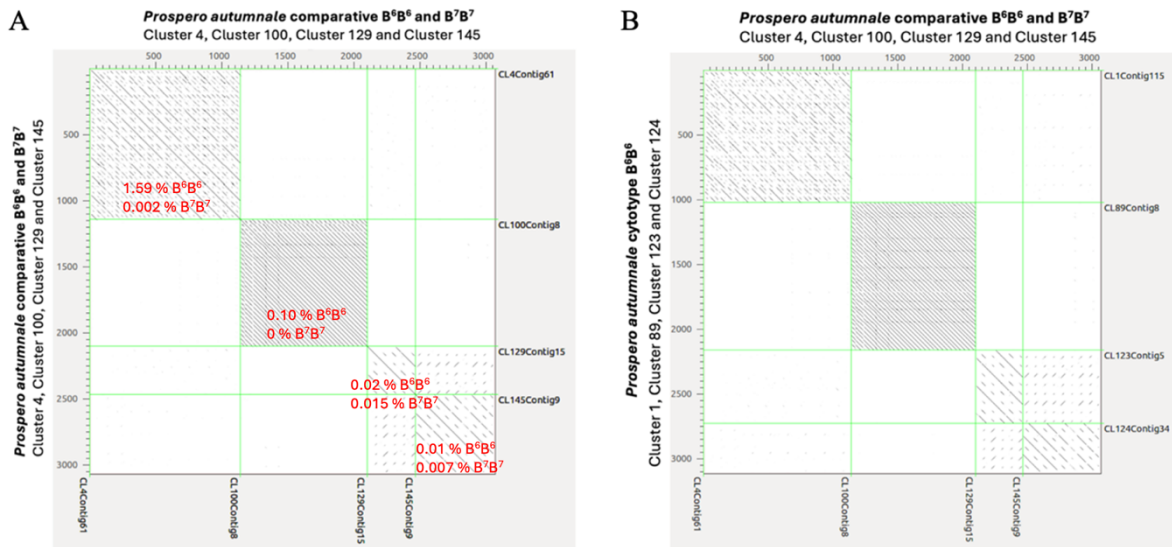
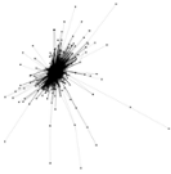
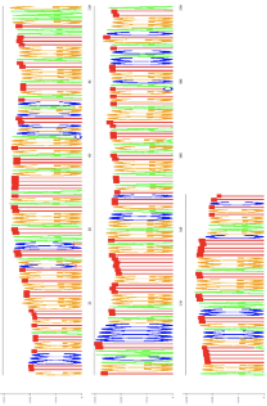
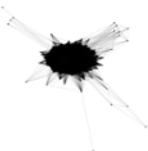
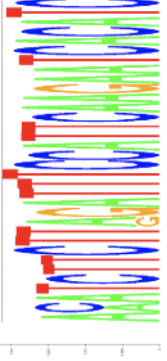
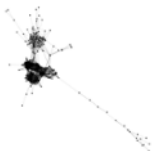
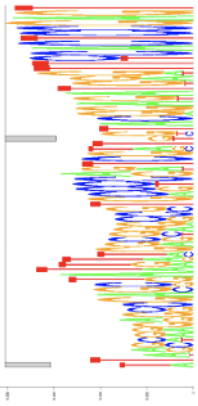

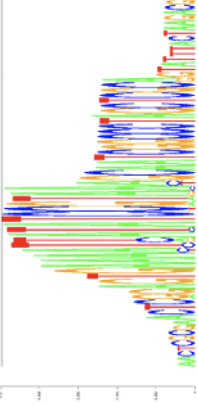


Figure 25. A: Dot plot of the four putative satellite DNAs identified in the comparative analysis of *Prospero autumnale* diploid cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup>. X and Y axes: Most abundant contig sequence of Cluster 4, Cluster 100, Cluster 129 and Cluster 145, containing the reads of putative satellite DNAs. B: Dot plot of the putative satellite DNAs identified in the comparative analysis of the cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup> of *Prospero autumnale* plotted against putative satellite DNAs identified in the cytotypes B<sup>6</sup>B<sup>6</sup>. X-axis: Most abundant contig sequence of Cluster 4, Cluster 100, Cluster 129 and Cluster 145, containing putative satellite DNA reads identified in the comparative analysis. Y-axis: Most abundant contig sequence of Cluster 1, Cluster 89, Cluster 123 and Cluster 124, containing putative satellite DNA reads identified in *Prospero autumnale* diploid cytotypes B<sup>6</sup>B<sup>6</sup>. The images were created using Dotter.

Table 11. List of putative satellite DNAs identified in the comparative analysis of repetitive DNA in diploid cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup> of *Prospero autumnale*.

Satellite DNA	Monomer length	Genome proportion B <sup>6</sup>	Genome Proportion B <sup>7</sup>	Consensus sequence	Cluster Graph	Sequence logo
<b>PaB6</b>	249 bp	1.59%	0.002%	GTTCCCTGGTTAGTGTITGGG TTGGTGTTCGATTCCAAITAG GGTTAGTATTTTGTAGTITAGIG GACGCTTCGACTAGTTAGGGT TAGGGTTTAAATGAAAGTGGAC AATCCCCCGATTAGGGTTTG GTTTTTGGATGGAAGGCCAGT TCTGATTAGGGTTAGAGTTAT TGGGATGTGTACGGTTCACAT AGTCAGGGTTAAGGTTTTTGG CCAGTTCCAAATTAGGGTITAGG GTTTTTCGAAAGGTGICTT		
<b>Pa138</b>	34 bp	0.10%	0%	ACATCTTCTTAGATTCCATTC AAGAATCACATC		
<b>Pa160</b>	71 bp	0.02%	0.015%	TTAAGCGGGAGCCGAGCTAAT TTGGCGGGCCGTGCCCCAGTC GTTGTACGGAGATGGGTTCCTCA CTCCGAGT		
<b>Pa124</b>	72 bp	0.01%	0.007%	AACTCGCGAACTCGGAGTGA AAATTATATCCGTAAGCAAT CGGCCCTAGCTGCAGTATTT ACTAAGACGT		

### 3.5.2. Comparative analysis diploid cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup>, and allotetraploid B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> of *Prospero autumnale*

2 000 000 reads from each cytotype (0.09x coverage for B<sup>6</sup>B<sup>6</sup>; 0.13x coverage B<sup>7</sup>B<sup>7</sup>; 0.05x coverage for B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup>) were included for the comparative analysis of the repeatome. Out of the total 2 787 663 reads analyzed, 27 796 reads were identified as plastid DNA or contamination and were consequently excluded from the subsequent analysis. 2 759 867 reads were used for the analysis, with 919 021 reads from the allotetraploid B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> genome, 921 413 from the B<sup>7</sup>B<sup>7</sup> genome and 919 432 from the B<sup>6</sup>B<sup>6</sup> genome. Genome proportions of the individual clusters and repetitive DNA families were normalized to the genome sizes of the analyzed species.

Repetitive DNA types in the three analyzed species were present in similar proportions, with the only notable difference being the number of identified satellite DNA reads (1.15% in B<sup>6</sup>B<sup>6</sup> vs 0.19% in B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> vs 0.02% in B<sup>7</sup>B<sup>7</sup>; Table 12, Figure 26). RepeatExplorer2 identified five putative satellite DNA clusters. Subsequent verification through Dotter confirmed the identity of three of these five potential satellite DNAs. Additionally, a fourth satellite DNA was identified using Dotter. The identified satellite DNAs were satellite DNA *PaB6*, the satellite DNA Pa138 and two novel satellite DNAs, Pa160 and Pa204 (Figure 27). The proportions of the identified satellite DNAs were different between all three cytotypes, with B<sup>7</sup>B<sup>7</sup> and B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> being more similar. In the B<sup>6</sup>B<sup>6</sup> cytotype satellite DNA *PaB6* constituted the majority with 1.067% of the total 1.15% satellite DNAs (Table 13). The other three satellite DNAs only made up 0.08%. In the cytotype B<sup>7</sup>B<sup>7</sup> satellite DNAs made up only 0.02% of the genome. In the allotetraploid 0.19% was represented by satellite DNAs, with the most abundant satellite DNA *PaB6* (0.16%). Using Dotter to plot the most abundant contig sequence of each cluster identified as satDNA showed four different satellite DNAs across all cytotypes (Figure 27A). Plotting against satellite DNAs identified in the B<sup>6</sup>B<sup>6</sup> cytotype revealed that they were the same satellite DNAs identified in the B<sup>6</sup>B<sup>6</sup> cytotype (Figure 27B). Plotting those against the B<sup>7</sup>B<sup>7</sup> cytotype revealed that the allotetraploids share two satellite DNAs with the parental cytotype B<sup>7</sup>B<sup>7</sup> (Figure 27C).

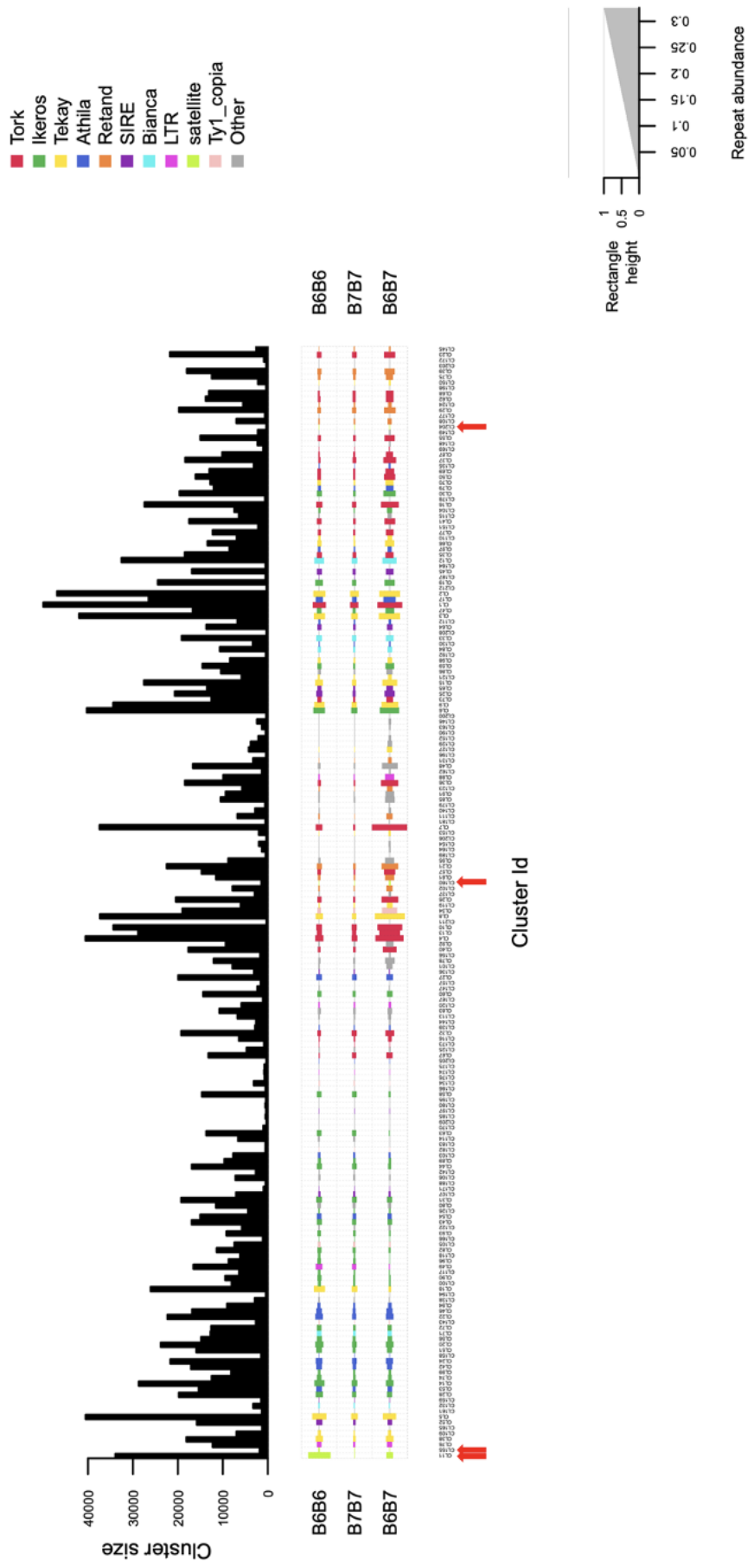


Figure 26. Summary of the comparative analysis of the repeatome of the diploid cytotypes  $B^6B^6$  and  $B^7B^7$  and the allotetraploid  $B^6B^6B^7B^7$  of Prospero autumnale depicting the number of reads in the individual clusters (top graph) and the genome proportion of the reads in the genome of the cytotypes  $B^6B^6$ ,  $B^7B^7$  and  $B^6B^6B^7B^7$ . B6B6: Prospero autumnale cytotype  $B^6B^6$ . B7B7: Prospero autumnale cytotype  $B^7B^7$ . B6B7: Prospero autumnale allotetraploid  $B^6B^6B^7B^7$ . Red arrows indicate clusters identified as satellite DNAs (Cluster 11, Cluster 155, Cluster 160 and Cluster 204).



Table 12. Proportions (in %) of repetitive DNA types in the comparative analysis of *Prospero autumnale* diploid cytotypes  $B^6B^6$ ,  $B^7B^7$  and the allotetraploid  $B^6B^6B^7B^7$ .

Type	Superfamily	Family	Genome proportion $B^6B^6$	No. of reads $B^6B^6$	Genome proportion $B^7B^7$	No. of reads $B^7B^7$	Genome proportion $B^6B^6B^7B^7$	No. of reads $B^6B^6B^7B^7$
			$B^6B^6$		$B^7B^7$		$B^6B^6B^7B^7$	
Retrotransposons			<b>23.45</b>	<b>647230</b>	<b>22.27</b>	<b>614543</b>	<b>23.72</b>	<b>654513</b>
	Ty3-gypsy		<u>8.77</u>	<u>241968</u>	<u>7.66</u>	<u>211475</u>	<u>8.90</u>	<u>245735</u>
		Reina	0.16	4449	0.23	6338	0.49	13650
		Tekay	4.59	126679	3.43	94736	4.36	120349
		Athila	2.90	79934	2.78	76706	1.93	53395
		Retand	1.12	30906	1.22	33695	2.11	58341
	Ty1-copia		<u>14.04</u>	<u>387500</u>	<u>14.03</u>	<u>387116</u>	<u>14.30</u>	<u>394681</u>
		Ale	0.04	1095	0.06	1608	0.08	2171
		Angela	0.24	6761	0.30	8223	0.06	1763
		Bianca	1.07	29616	0.99	27420	0.77	21289
		Ikeros	6.31	174186	5.88	162315	3.38	93310
		SIRE	1.36	37536	1.08	29780	0.97	26663
		Tork	4.65	128306	5.35	147694	8.58	236799
		Unclassified	0.36	10000	0.37	10076	0.46	12686
	other/non-LTR		<u>0.64</u>	<u>17762</u>	<u>0.58</u>	<u>15952</u>	<u>0.51</u>	<u>14097</u>
		LINE	0.01	395	0.01	330	0.06	1679
		Other LTR	0.63	17367	0.57	15622	0.45	12418
DNA transposons			<b>0.10</b>	<b>2777</b>	<b>0.12</b>	<b>3401</b>	<b>0.08</b>	<b>2090</b>
	Subclass I		<u>0.10</u>	<u>2777</u>	<u>0.12</u>	<u>3401</u>	<u>0.08</u>	<u>2090</u>
		EnSpm_CACTA	0.04	1010	0.05	1312	0.01	354
		MuDR_Mutator	0.03	868	0.04	1068	0.03	750
		PIF_Harbinger	0.02	644	0.02	613	0.03	905
		MITE	0.01	255	0.01	408	0.00	81
Tandem repeats			<b>1.27</b>	<b>35103</b>	<b>0.19</b>	<b>5119</b>	<b>0.71</b>	<b>19705</b>
	Satellite DNA		<u>1.15</u>	<u>31652</u>	<u>0.02</u>	<u>476</u>	<u>0.19</u>	<u>5373</u>
	rDNA		<u>0.13</u>	<u>3451</u>	<u>0.17</u>	<u>4643</u>	<u>0.52</u>	<u>14332</u>
		5S rDNA	0.00	38	0.00	78	0.01	353
		45S rDNA	0.12	3413	0.17	4565	0.51	13979
Unclassified repeats			<b>0.45</b>	<b>12455</b>	<b>0.55</b>	<b>15158</b>	<b>0.66</b>	<b>18128</b>
Unclassified			<b>0.81</b>	<b>22247</b>	<b>0.76</b>	<b>20906</b>	<b>1.27</b>	<b>34980</b>
<b>Total Top Clusters</b>			<b>26.09</b>	<b>719812</b>	<b>23.89</b>	<b>659127</b>	<b>26.43</b>	<b>729416</b>

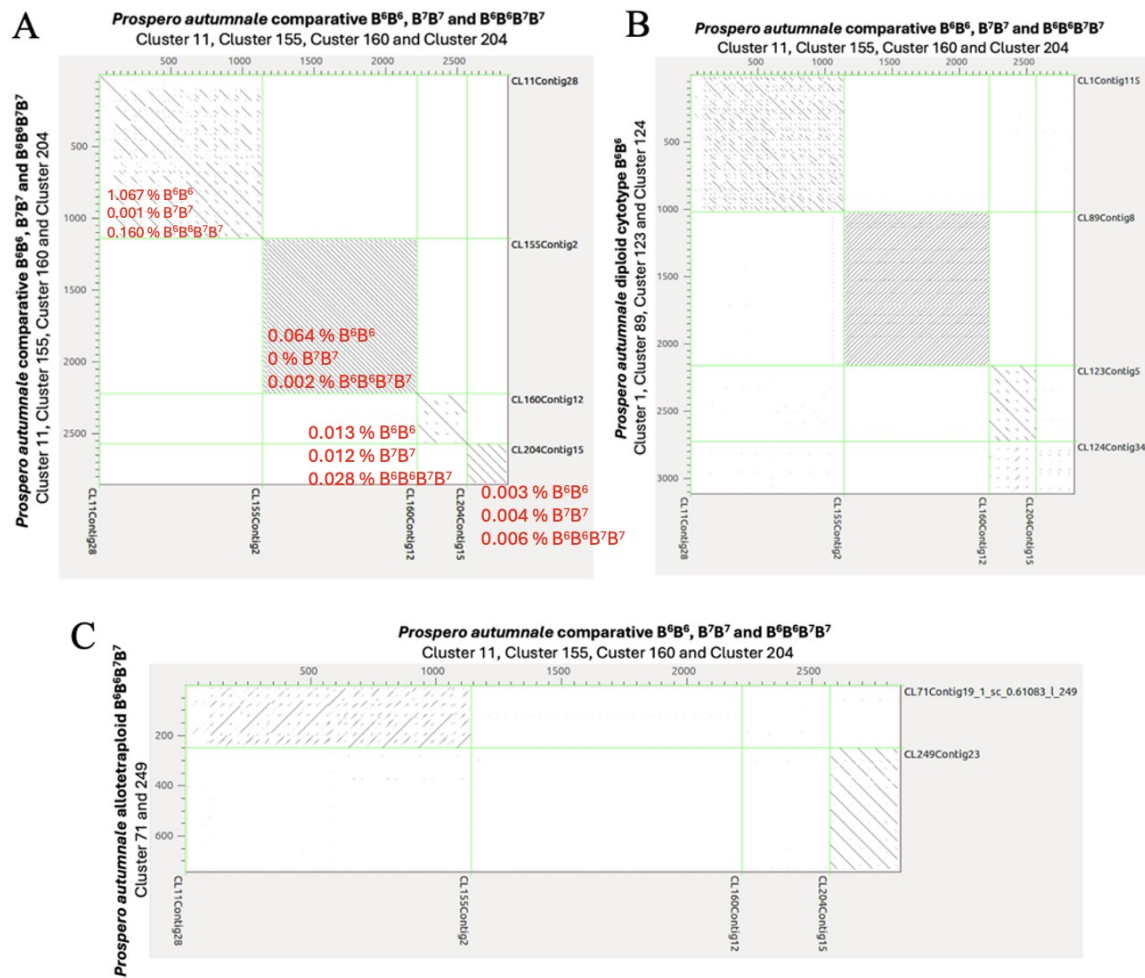


Figure 27. A: Dot plot of four putative satellite DNAs identified in the comparative analysis of the *Prospero autumnale* diploid cytotypes  $B^6B^6$ ,  $B^7B^7$  and allotetraploid  $B^6B^6B^7B^7$ . X and Y axes: Most abundant contig sequence of Cluster 11, Cluster 155, Cluster 160 and Cluster 204, containing the reads of putative satellite DNAs. B: Dot plot of the putative satellite DNAs identified in the comparative analysis of the *Prospero autumnale* diploid cytotypes  $B^6B^6$ ,  $B^7B^7$  and allotetraploid  $B^6B^6B^7B^7$  plotted against the putative DNAs identified in the cytotype  $B^6B^6$ . X axis: Most abundant contig sequence of Cluster 4, Cluster 100, Cluster 129 and Cluster 145, containing putative satellite DNA reads identified in the comparative analysis. Y axis: Most abundant contig sequence of Cluster 1, Cluster 89, Cluster 123 and Cluster 124, containing putative satellite DNA reads identified in the cytotype  $B^6B^6$ . C: Dot plot of the putative satellite DNAs identified in the comparative analysis of the *Prospero autumnale* diploid cytotypes  $B^6B^6$ ,  $B^7B^7$  and allotetraploid  $B^6B^6B^7B^7$  plotted against the putative DNAs identified in the cytotype  $B^7B^7$ . X-axis: Most abundant contig sequence of Cluster 4, Cluster 100, Cluster 129 and Cluster 145, containing putative satellite DNA reads identified in the comparative analysis. Y-axis: Most abundant contig sequence of Cluster 71 and Cluster 249, containing putative satellite DNA reads identified in the cytotype  $B^7B^7$ . The image was created using *Dotter*.

Table 13. List of putative satellite DNAs identified in the comparative analysis of repetitive DNA in diploid cytotypes B<sup>6</sup>B<sup>6</sup>, B<sup>7</sup>B<sup>7</sup> and the allotetraploid B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> of *Prospero autumnale*.

Satellite DNA	Monomer length	Genome proportion B <sup>6</sup> B <sup>6</sup>	Genome proportion B <sup>7</sup> B <sup>7</sup>	Genome proportion B <sup>6</sup> B <sup>6</sup> B <sup>7</sup> B <sup>7</sup>	Consensus Sequence	Cluster Graph	Sequence logo
<b>PaB6</b>	249 bp	1.067%	0.001%	0.160%	GTTAGTGTTGGGTTGGTGG TCGATTCCAAATTAGGGTTA GTATTTTGAGTTAGTGGAC GCTTCGACTAGTTAGGGTT AGGGTTTAAAGAAAGTGA CAATTCGCCGATTAAGGTT TGGTTTTTGGATGGAGGCC AGTTCGATTAGGGTTAGA GTTATTGGGATGTACGG TTCCACTAGTCAGGGTTAA GGTTTTGGCCAGTCCCAAT TAGGGTTAGGGTTTTTCGA AGGTGCTTGTTCCTG		
<b>Pa138</b>	34 bp	0.064%	0.000%	0.002%	CTTGAATGGAAATCTAAGA AGATGTGATGIGAIT		
<b>Pa160</b>	71 bp	0.013%	0.012%	0.028%	GTTTAGCAAGGAGCGGAGC TATTTTGGCGGCCGTGCC CGGTCGATGTACGGAGATG AGTTCACCTCCGA		
<b>Pa204</b>	70 bp	0.003%	0.004%	0.006%	CGCAAATCCTTAGTACTA GGCGGACAAACGCTCGACT CACTATGGGCAAGGCTAC CTACGCCTTAGCT		

## 4. Discussion

This study presents a detailed analysis of the repeatome of the *Prospero autumnale* complex and analysis of the localization of tandemly repeated satellites DNA in the chromosomes. The repeat profiles of the two diploid cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup>, as well as Group I allotetraploid were analyzed to gain insight to the content of repetitive DNA fraction and specifically the diversity and evolution of satellite DNAs in the *Prospero autumnale* complex. The previously identified satellite DNA *PaB6* and two novel satellite DNAs, Pa138 and Pa147, were also mapped in the chromosomes of both diploid cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup>, their homoploid diploid hybrid B<sup>6</sup>B<sup>7</sup> and all four groups of the B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> allotetraploid. The karyotypes and genome sizes were analyzed for the purpose of understanding the changes which take place during the evolution of their genomes.

### 4.1. Karyotype analysis and genome size evolution

The karyotype analysis of the *Prospero autumnale* complex revealed the presence of two base chromosome numbers in with the two analyzed diploid cytotypes, B<sup>7</sup>B<sup>7</sup> and B<sup>6</sup>B<sup>6</sup>, having  $x = 7$  and  $x = 6$  chromosomes, respectively. The diploid homoploid hybrid of these two cytotypes, B<sup>6</sup>B<sup>7</sup>, had  $2x = 2n = 13$  chromosomes. These results correspond to previous analysis of the *Prospero autumnale* complex (Jang et al., 2013). The current analysis also confirmed the varying chromosome numbers of Group I allotetraploids ( $2n = 4x = 25-28$ ) and the chromosome numbers of Group II, Group III and Group IV ( $2n = 4x = 28$ ) allotetraploids (Jang et al., 2018a). The structure of the analyzed chromosomes in different cytotypes only differed regarding the number of fusion chromosome F(6/7) and free chromosomes 6 and 7. The fusion chromosome was present in the B<sup>6</sup>B<sup>6</sup> cytotype, and its presence also resulted in the reduced diploid chromosome numbers in the hybrid B<sup>6</sup>B<sup>7</sup> and Group I allotetraploids with  $2n = 4x = 25-27$  chromosomes. Group I allotetraploids with 28 chromosomes lacked the fusion chromosome entirely. This in turn explained the lack of the fusion chromosome in Groups II-IV, which originated from a cross between the B<sup>7</sup>B<sup>7</sup> and Group I ( $2n = 4x = 28$ ) parental cytotypes, the latter only representing individuals lacking fusion chromosome. This was consistent with results published before (Jang et al., 2013; Jang et al., 2018a). The karyotypes of the diploid hybrid individuals and some allotetraploids (Groups II-IV) possessed two classes (sizes) of chromosomes, with chromosomes inherited from the B<sup>6</sup>B<sup>6</sup> cytotype being larger than those of B<sup>7</sup>B<sup>7</sup> (Jang et al., 2018a). Due to possible differences in chromosome condensation, these size changes were not always pronounced in the analyzed genomes.

The genome sizes of one diploid and two polyploid plants of the  $B^7B^7$  cytotype ( $2n = 2x = 14$ ,  $2n = 4x = 28$ ,  $2n = 6x = 42$ ) were measured using flow cytometry. Previous analysis reported mean 1C values ranging from 4.23 pg to 4.54 pg for the diploid  $B^7B^7$  (Jang et al., 2013; Vestek et al., 2019). For the autotetraploid and autohexaploid the reported mean values were between 7.67 pg and 9.15 pg and between 11.0 pg and 11.28 pg (Vestek et al., 2019). The genome size measurements of all three ploidy levels in this study (diploid: 4.42 pg, tetraploid: 8.43 pg, hexaploid: 11.35 pg) corresponded to previous measurements. The genome size in the analyzed polyploid plants was not additive in comparison to lower ploidy levels due to the phenomenon of genome downsizing that has often been observed in various other families in angiosperms (Leitch & Bennet, 2004; Vestek et al., 2019). In the autotetraploid this reduction of the genome size is only relatively small (-0.4 pg), but the genome size of the autohexaploid was already 1.91 pg smaller than expected. The reason for genome downsizing in some but not in other polyploids still remains unknown (Leitch & Bennet, 2004; Wang et al., 2021), although it often is connected to cytological diploidization.

#### 4.2. Repeatome analysis of the *Prospero autumnale* complex and identification of novel satellite DNAs

The repeat profiles of the diploid cytotypes  $B^6B^6$  and  $B^7B^7$ , as well as Group I  $B^6B^6B^7B^7$  allotetraploid were analyzed using RepeatExplorer2. In all three analyzed genomes retrotransposons represented the majority of repeats, an average 70%. The Ty1-*copia* superfamily was dominant in all three cytotypes, with 41% in  $B^6B^6$ , 42% in  $B^7B^7$  and 37% in Group I. In both  $B^6B^6$  and  $B^7B^7$  the most abundant retrotransposon family was Ty1-*copia* Ikeros (18%), while in Group I the Ty1-*copia* family Tork was dominant (up 27%). The genome proportion of Ty3-*gypsy* superfamily in all three cytotypes varied, with the smallest amount in Group I (18%) and highest in  $B^6B^6$  (28%). The most abundant Ty3-*gypsy* family was Tekay with 14% for  $B^6B^6$ , 11% for  $B^7B^7$  and 8% for Group I. Other retrotransposons made up quite a large portion of the Group I genome with 14% suggesting different genome dynamics after polyploidization. DNA transposons (0.25-1%), as well as rRNA genes (0.37 - 1.59%) made up only a very small portion of the genomes of all three cytotypes.

Individual analysis revealed that as much as 3.42% of  $B^6B^6$  genome was represented by satellite DNA reads, whereas these repeats only constituted 0.49% of the genome of Group I. In the repeatome of the  $B^7B^7$  cytotype no satellite DNAs were detected. The analysis of the repeatomes using RepeatExplorer2 and annotation verification via dotter lead to the identification of five satellite DNAs in the *Prospero autumnale* complex in total. One of

those, satellite DNA *PaB6*, was already previously reported (Emadzade et al., 2014). The comparative analysis of the repeatome revealed that satellite DNA *PaB6* was present in all three analyzed genomes but was represented by highest copy number in the B<sup>6</sup>B<sup>6</sup> cytotype (1.067% vs. 0.001% in B<sup>7</sup>B<sup>7</sup> and 0.16% in Group I). In addition, four novel satellite DNAs were identified. These were named Pa124 (identified in the B<sup>6</sup>B<sup>6</sup> genome), Pa138 (identified in the B<sup>6</sup>B<sup>6</sup> and B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> genome), Pa160 (identified in the B<sup>6</sup>B<sup>6</sup> genome) and Pa204 (identified in the B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> genome; Figure 28). Pa138 was only present in the B<sup>6</sup>B<sup>6</sup> cytotype (0.064%) and Group I allotetraploid (0.002%). Satellite DNAs Pa160 and Pa204 were present in all three analyzed genomes but only in very low copy numbers. Satellite DNA Pa124 was only detected in the genome of the B<sup>6</sup>B<sup>6</sup> cytotype but showed some degree (albeit low) of similarity to the satellite DNA Pa160. Satellite DNA Pa147, which was identified in cytotype AA of the *Prospero autumnale* complex (H. Weiss-Schneeweiss, unpubl.), was not detected in repeat analyses of any of the three analyzed genomes, although some copies could be detected in the chromosomes (see below). Amplification/reduction of satellite DNAs copy numbers and/or emergence of novel satellite DNAs during diversification and speciation, as well as accompanying and following allopolyploidy was also observed in other plant genera, such as *Nicotiana* (Koukalova et al., 2010; Weiss-Schneeweiss & Schneeweiss, 2013). The detection of satellite DNA in NGS sequence data is not unbiased due to different, often AT or CG rich composition of their monomers that are more or less prone to fragmentation during library preparation. Therefore, the copy numbers of satellite DNAs identified via RepeatExplorer2 need to be verified in wet lab experiments (e.g., dot blots; Emadzade et al., 2014).

#### 4.3. Localization of satellite DNAs

Satellite DNA *PaB6* and two novel satellite DNAs, Pa138 and Pa147, were physically mapped in the chromosomes of several individuals of diploid cytotypes B<sup>6</sup>B<sup>6</sup> and B<sup>7</sup>B<sup>7</sup>, their diploid homoploid hybrid B<sup>6</sup>B<sup>7</sup>, as well as all four groups (Group I-IV) of the allotetraploids B<sup>6</sup>B<sup>6</sup>B<sup>7</sup>B<sup>7</sup> and the results are summarized in Fig. 29. SatDNA *PaB6* was detected in all analyzed diploid cytotypes and allotetraploids. All chromosomes of the B<sup>6</sup>B<sup>6</sup> cytotype and those in the hybrid individuals inherited from the B<sup>6</sup>B<sup>6</sup> parent carried satDNA *PaB6* in the pericentromeric regions, with one signal in B<sup>6</sup>B<sup>6</sup> cytotype being occasionally weaker than others. Additional novel loci were detected in the long arm of the fusion chromosome F(6/7) in the B<sup>6</sup>B<sup>7</sup> hybrid, not detected in previous analysis (Emadzade et al., 2014). In contrast, satDNA *PaB6* was detected as very small loci in only in seven out of 14 chromosomes in the

$B^7B^7$  cytotype. *PaB6* was, however, not detected in  $B^7B^7$ -derived chromosomes in the diploid hybrids. These results were consistent with the repeatome analysis, with the abundance of satellite DNA *PaB6* very low in the  $B^7B^7$  cytotype. SatDNA *PaB6* was mapped in two Group I allotetraploid plants with differing chromosome numbers. Signals were detected in every single chromosome, with one or two being weaker than the rest. Additionally, the plant H213, which carried a supernumerary B-chromosome, showed dispersed satDNA *PaB6* signals along this entire chromosome, consistent with high levels of amplification of *PaB6* observed in other B chromosomes in the complex (Jang et al., 2015). In Groups II-IV satDNA *PaB6* was detected in the pericentric regions of all chromosomes inherited from Group I or Group II parental cytotype, with one signal being occasionally weaker than the rest.

In the comparative repeatome analysis, satDNA Pa138 was only identified in the  $B^6B^6$  and Group I allotetraploid, whereas satDNA Pa147 was not identified in any of the cytotypes. However, both these satDNAs were detected in the chromosomes of all analyzed cytotypes, albeit as very small and few loci. The loci of satDNAs Pa138 and Pa147 were detected in all three diploid cytotypes. In  $B^6B^6$  and  $B^6B^7$ , both satDNA Pa138 and Pa147 were located on the same chromosomes, with some polymorphisms detected. These two satellite DNAs did not reveal any common pattern in the  $B^7B^7$  diploid and all allotetraploid cytotypes, with each having unique distribution patterns and high levels of polymorphism. In polyploids, signals were detected in chromosomes inherited from both parent cytotypes, again, with high levels of polymorphisms.

The analyses performed in this study indicated similarities in the repeatome of the two diploid cytotypes and the allotetraploid. The genome size and chromosome numbers of the allotetraploid were additive. The primary difference could be seen in the proportion and number of identified putative satellite DNAs. The few identified satellite DNAs made up only a small portion of the genome, something which was perceived in earlier studies (Weiss-Schneeweiss & Schneeweiss, 2013).

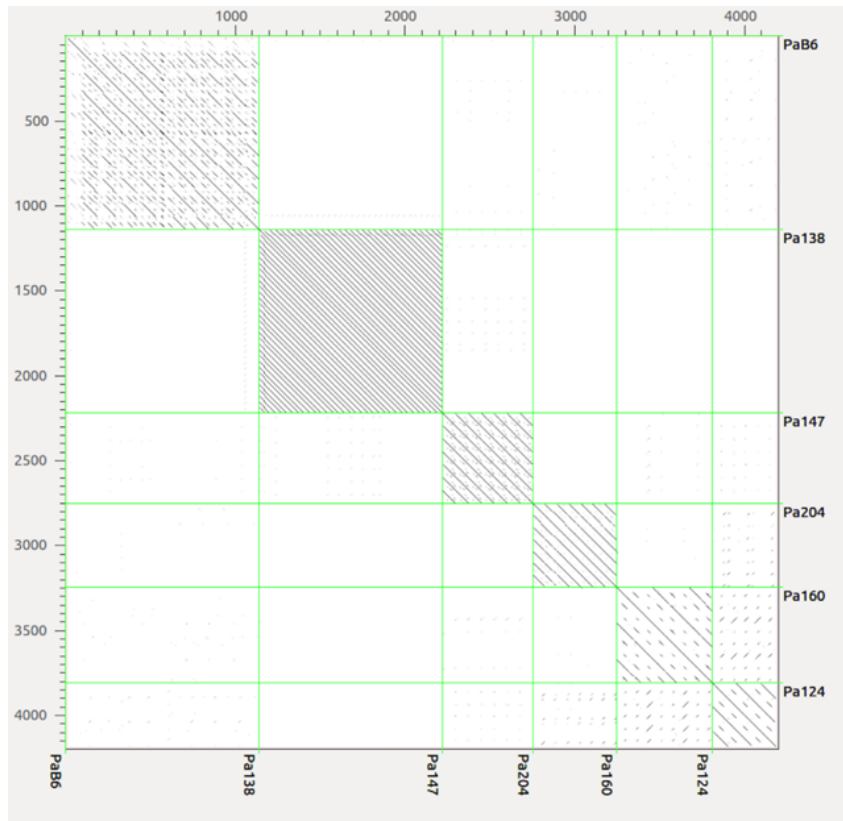


Figure 28. Dot-plot of all six putative satellite DNAs identified in the *Prospero autumnale* complex. X and Y axes: Most abundant contig sequence of PaB6, Pa138, Pa147 (H. Weiss-Schneeweiss, unpubl.), Pa204, Pa160 and Pa124, containing the reads of putative satellite DNAs.



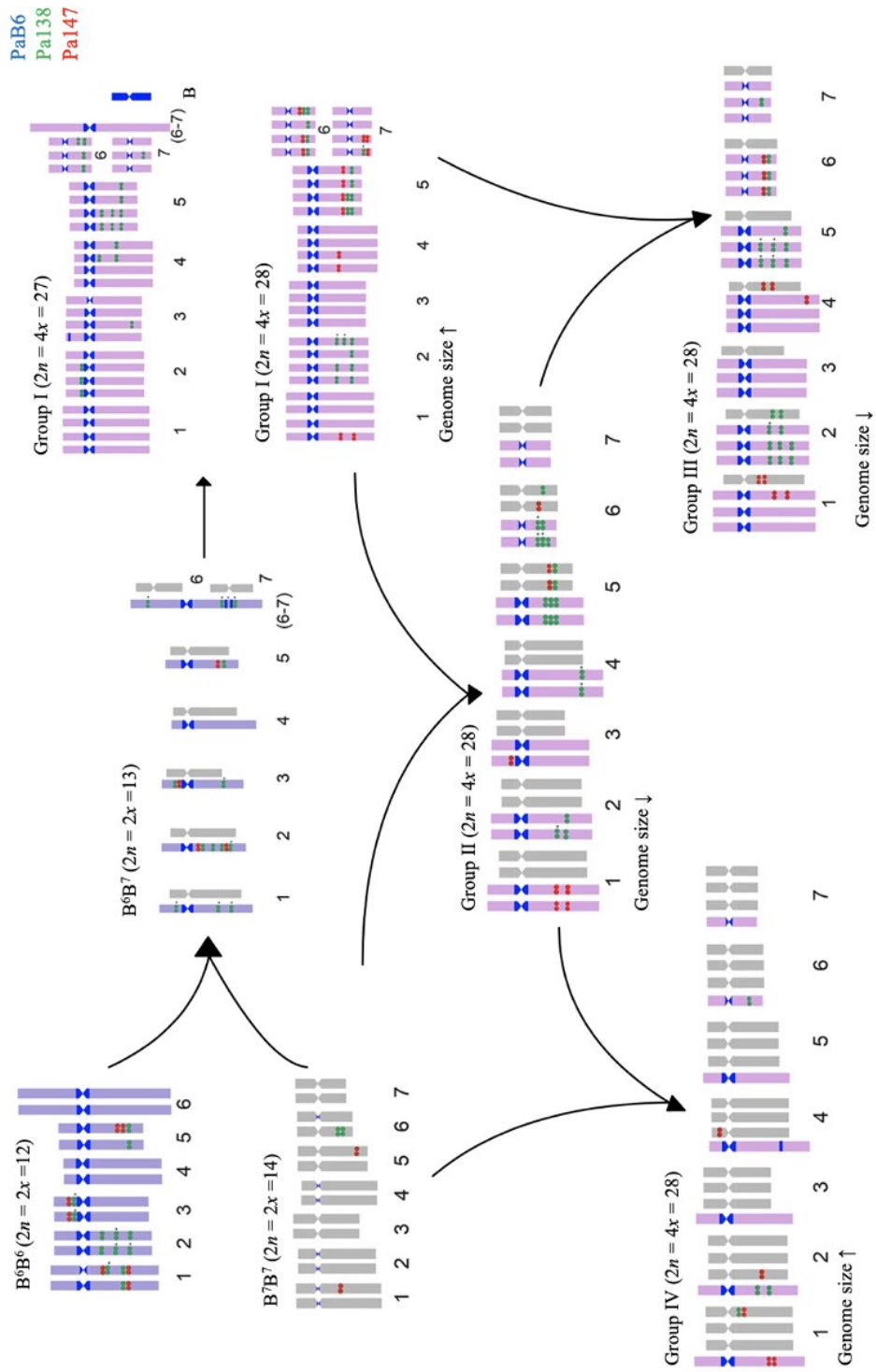


Figure 29. Evolution of tandem repeats PaB6, Pa138 and Pa147 in chromosomes of the analyzed diploid and allotetraploid cyto types of the Prospero autumnale complex. SatDNA PaB6 in dark blue, satDNA Pa138 in green, satDNA Pa147 in red. Arrows indicate origin of hybrids and polyploids. Asterisks indicate variable signals.

## 5. References

- Ainsworth CC. (1980) The population cytology of *Scilla autumnalis*. Ph.D. Thesis, University of London, London, UK.
- Ainsworth CC, Parker JS, Horton DM. (1983) Chromosome variation and evolution in *Scilla autumnalis*. In: Brandham PE, Bennett MD (eds) *Kew Chromosome Conference II*, London: Allen & Unwin, pp. 261—268.
- Baranyi M, Greilhuber J. (1995) Flow cytometric analysis of genome size variation in cultivated and wild *Pisum sativum* (Fabaceae). *Plant Systematics and Evolution* **194**: 231—239.
- Beck CR, Garcia-Perez JL, Badge RM, Moran JV. (2011) LINE-1 elements in structural variation and disease. *Annual Review of Genomics and Human Genetics* **12**: 187—215.
- Bennetzen JL. (2002) Mechanisms and rates of genome expansion and contraction in flowering plants. *Genetica* **115**: 29—36.
- 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.
- Clarkson JJ, Lim KY, Kovarik A, Chase MW, Knapp S, Leitch AR. (2005) Long-term genome diploidization in allopolyploid *Nicotiana* section *Repandae* (Solanaceae). *The New Phytologist* **168**: 241—252.
- Dewannieux M, Esnault C, Heidmann T. (2003) LINE-mediated retrotransposition of marked *Alu* sequences. *Nature Genetics* **35**: 41—48.
- Doyle JJ, Doyle JL. (1987) A rapid DNA isolation procedure for small quantities of fresh leaf tissue. *Phytochemical Bulletin* **9**: 11—15.
- Ebert I. (1993) Systematische Karyologie und Embryologie von *Prospero* Salisb. und *Barnardia* Lindl. (Hyacinthaceae). Ph.D. Thesis, University of Vienna, Vienna, Austria.
- Ebert I, Greilhuber J, Speta F. (1996) Chromosome banding and genome size differentiation in *Prospero* (Hyacinthaceae): diploids. *Plant Systematics and Evolution* **203**: 143—177.
- Elbarbary RA, Lucas BA, Maquat LE. (2016) Retrotransposons as regulators of gene expression. *Science* **351**: aac7247.
- Elder JF Jr, Turner BJ. (1995) Concerted evolution of repetitive DNA sequences in eukaryotes. *The Quarterly Review of Biology* **70**: 297—320.
- Emadzade K, Jang TS, Macas J, Kovařík A, Novák P, Parker J, Weiss-Schneeweiss H. (2014) Differential amplification of satellite *PaB6* in chromosomally hypervariable *Prospero autumnale* complex (Hyacinthaceae). *Annals of Botany* **114**: 1597—1608.
- Feschotte C, Pritham EJ. (2007) DNA transposons and the evolution of eukaryotic genomes. *Annual Review of Genetics* **41**: 331—368.

- Fukui K, Nakayama S. (1996) *Plant Chromosomes: Laboratory Methods*. Boca Raton, U. S. A: CRC Press.
- Garcia S, Kovařík A, Leitch AR, Garnatje T. (2017) Cytogenetic features of rRNA genes across land plants: analysis of the plant rDNA database. *The Plant Journal* **89**: 1020—1030.
- 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* **8**: 230.
- Grabundzija I, Hickman AB, Dyda F. (2018) Helraiser intermediates provide insight into the mechanism of eukaryotic replicative transposition. *Nature Communications* **9**: 1278.
- Grant V. (1982) Chromosome number patterns in primitive angiosperms. *Botanical Gazette* **143**: 390—394.
- Greilhuber J, Ebert I. (1994) Genome size variation in *Pisum sativum*. *Genome* **37**: 646—655.
- Grewal SI, Elgin SC. (2007) Transcription and RNA interference in the formation of heterochromatin. *Nature* **447**: 399—406.
- Han F, Lamb JC, Birchler JA. (2006) High frequency of centromere inactivation resulting in stable dicentric chromosomes of maize. *Proceedings of the National Academy of Sciences of the U. S. A.* **103**: 3238—3243.
- Hawkins JS, Grover CE, Wendel JF (2008) Repeated big bangs and the expanding universe: directionality in plant genome size evolution. *Plant Science* **174**: 557—562.
- Hemleben V, Kovařík A, Torres-Ruiz RA, Volkov RA, Beridze T. (2007). Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. *Systematics and Biodiversity* **5**: 277—289.
- Hickman AB, Dyda F. (2016) DNA transposition at work. *Chemical Reviews* **116**: 12758—12784.
- Husband BC, Baldwin SJ, Suda J. (2013) The incidence of polyploidy in natural plant populations: major patterns and evolutionary processes. In: Leitch IJ, Greilhuber J, Dolezel J, Wendel JF (eds.) *Plant Genome Diversity: Physical structure, behaviour and evolution of plant genomes*, vol 2. Springer-Verlag, Wien, pp 255—276.
- Jang TS. (2013) Chromosomal evolution in *Prospero autumnale* complex. Ph.D. Thesis, University of Vienna, Vienna, Austria.
- Jang TS, Emadzade K, Parker J, Tensch EM, Leitch AR, Speta F, Weiss-Schneeweiss H. (2013) Chromosomal diversification and karyotype evolution of diploids in the cytologically diverse genus *Prospero* (Hyacinthaceae). *BMC Evolutionary Biology* **13**: 136.
- Jang TS, Weiss-Schneeweiss H. (2015) Formamide-free genomic *in situ* hybridization allows unambiguous discrimination of highly similar parental genomes in diploid hybrids and allopolyploids. *Cytogenetic and Genome Research* **146**: 325—31.

- Jang TS, Parker JS, Emadzade K, Tensch EM, Leitch AR, Weiss-Schneeweiss H. (2018a) Multiple origins and nested cycles of hybridization result in high tetraploid diversity in the monocot *Prospero*. *Frontiers in Plant Science* **9**: 433.
- Jang TS, Parker JS, Weiss-Schneeweiss H. (2018b) Euchromatic supernumerary chromosomal segments-remnants of ongoing karyotype restructuring in the *Prospero autumnale* complex? *Genes* **9**: 468.
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, Soltis DE, Clifton SW, Schlarbaum SE, Schuster SC, Ma H, Leebens-Mack J, dePamphilis CW. (2011) Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**: 97—100.
- Kapitonov VV, Jurka J. (2001) Rolling-circle transposons in eukaryotes. *Proceedings of the National Academy of Sciences of the U. S. A.* **98**: 8714—8719.
- Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovarik A. (2009) Fall and rise of satellite repeats in allopolyploidy of *Nicotiana* over c. 5 million years. *New Phytologist* **186**: 148—60.
- Leitch IJ, Bennett MD. (2004) Genome downsizing in polyploid plants. *Biological Journal of the Linnean Society* **82**: 651—663.
- Leitch AR, Leitch IJ. (2008) Genomic plasticity and the diversity of polyploid plants. *Science* **320**: 481—483.
- Lysák MA, Schubert I. (2013) Mechanisms of chromosome rearrangements. In: Leitch IJ, Greilhuber J, Dolezel J, Wendel JF (eds.) *Plant Genome Diversity: Physical structure, behaviour and evolution of plant genomes*, vol 2. Springer-Verlag, Wien, pp 137—147.
- Macas J, Neumann P, Navrátilová A. (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* **8**: 427.
- Madlung A. (2013) Polyploidy and its effect on evolutionary success: old questions revisited with new tools. *Heredity* **110**: 99—104.
- Masterson J. (1994) Stomatal size in fossil plants: evidence for polyploidy in majority of angiosperms. *Science* **264**: 421—424.
- McKinnon KM. (2018) Flow cytometry: an overview. *Current Protocols in Immunology* **120**: 5.1.1—5.1.11.
- Muñoz-López M, García-Pérez JL. (2010) DNA transposons: nature and applications in genomics. *Current Genomics* **11**: 115—128.
- 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, Neumann P, Macas J. (2020) Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nature Protocols* **15**: 3745—76.
- Parker JS, Lozano R, Taylor S, Ruiz-Rejon, M. (1991) Chromosomal structure of populations of *Scilla autumnalis* in the Iberian Peninsula. *Heredity* **67**: 287—297.
- Plohl M, Meštrović N, Mravinac B. (2012) Satellite DNA evolution. In: Garrido-Ramos MA. (ed.) *Genome Dynamics*. Karger, Basel, pp. 126—152.
- Plohl M, Meštrović N, Mravinac B. (2014) Centromere identity from the DNA point of view. *Chromosoma* **123**: 313—325.
- POWO (2024). Plants of the world online. Facilitated by the Royal Botanic Gardens, Kew. Published on the Internet; <http://www.plantsoftheworldonline.org/> Retrieved 14 March 2024.
- Raskina O, Barber JC, Nevo E, Belyayev A. (2008) Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenetic and Genome Research* **120**: 351—357.
- Roa F, Guerra M. (2012) Distribution of 45S rDNA sites in chromosomes of plants: structural and evolutionary implications. *BMC Evolutionary Biology* **12**: 225.
- Roa F, Guerra M. (2015) Non-random distribution of 5S rDNA sites and its association with 45S rDNA in plant chromosomes. *Cytogenetic and Genome Research* **146**: 243—249.
- Sanchez DH, Gaubert H, Drost HG, Zabet NR, Paszkowski J. (2017) High-frequency recombination between members of an LTR retrotransposon family during transposition bursts. *Nature Communications* **8**: 1283.
- Šatović-Vukšić E, Plohl M. (2023) Satellite DNAs—from localized to highly dispersed genome components. *Genes* **14**: 742.
- Schmidt T, Heslop-Harrison JS. (1998) Genomes, genes and junk: the large scale organization of plant chromosomes. *Trends in Plant Science* **3**:195—199.
- Schubert I. (2007) Chromosome evolution. *Current Opinion in Plant Biology* **10**: 109—115.
- Sonnhammer EL, Durbin R. (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* **167**: GC1—10.
- Speta F. (1993) The autumn-flowering squills of the mediterranean region. In: *Proceedings of the 5th Optima Meeting, Istanbul, Turkey*, University of Istanbul, Istanbul, Turkey, pp. 109—124.
- Taylor S. (1997) Chromosomal evolution of *Scilla autumnalis*. Ph.D. Thesis, University of London, London, UK.
- Temsch E. 2003. Genome size variation in plants with special reference to the genus *Arachis* (Fabaceae), and the methodological improvement by application of a new optical immersion-gel in flow cytometry. Ph.D Thesis, University of Vienna, Vienna, Austria.

- Temsch EM, Greilhuber J, Krisai R. (2010) Genome size in liverworts. *Preslia* **82**: 63—80.
- Temsch EM, Koutecký P, Urfus T, Šmarda P, Doležel J. (2021) Reference standards for flow cytometric estimation of absolute nuclear DNA content in plants. *Cytometry (A)* **9**: 710—724.
- Vaughan HE, Taylor S, Parker JS. (1997) The ten cytological races of the *Scilla autumnalis* species complex. *Heredity* **79**: 371—379.
- Vestek A, Slovák M, Weiss-Schneeweiss H, Temsch EM, Luković J, Kučera J, Anačkov G. (2019) Morpho-anatomical differentiation and genome size variation in three ploidy levels within the B7 cytotype of *Prospero autumnale* (Hyacinthaceae) complex from the Balkan Peninsula and Pannonian Basin. *Plant Systematics and Evolution* **305**: 597—609.
- Wang X, Morton JA, Pellicer J, Leitch IJ, Leitch AR. (2021) Genome downsizing after polyploidy: mechanisms, rates and selection pressure. *The Plant Journal* **107**: 1003—1015.
- 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-Schneeweiss H, Emadzade K, Jang TS, Schneeweiss GM. (2013) Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenetic and Genome Research* **140**: 137—150.
- Weiss-Schneeweiss H, Schneeweiss GM. (2013) Karyotype diversity and evolutionary trends in angiosperms. In: Greilhuber J, Doležel J, Wendel JF (eds.) *Plant Genome Diversity: Physical structure, behaviour and evolution of plant genomes*, vol 2. Springer-Verlag, Vienna, pp. 209–230.
- Weiss-Schneeweiss H, Leitch A, McCann J, Jang TS, Macas J. (2015) Employing next generation sequencing to explore the repeat landscape of the plant genome. In: E. Hörandl E, Appelhans M. (eds.), *Next Generation Sequencing in Plant Systematics*. Koeltz Scientific Books, Germany, pp.1—25.
- Wells JN, Feschotte C. (2020) A field guide to eukaryotic transposable elements. *Annual Review of Genetics* **54**: 539—561.

## 6. Abstract

The *Prospero autumnale* complex is chromosomally very diverse encompassing several diploid cytotypes with varying basic chromosome numbers ( $x = 5, 6,$  and  $7$ ) and genome sizes as well as many auto- and allopolyploids, thus making it an ideal system to analyze the evolution of genomes and the role of chromosomal change in species diversification. Analysis of the repeat profiles of two diploid cytotypes,  $B^6B^6$  ( $x = 6$ ) and  $B^7B^7$  ( $x = 7$ ), and their Group I allotetraploids ( $2n = 25-28$ ) revealed similar and high proportions of repetitive DNA in all genomes (c. 75%). The most abundant repetitive DNA type in all three genomes was the retrotransposon superfamily Ty1-*copia* (c. 40%). DNA transposons and tandem repeats made up a small portion of the genomes. Analysis allowed for identification of satellite DNA *PaB6* and four novel satellite DNAs, all of which made up only a small proportion of the genomes. The mapping of satDNA *PaB6* and two novel satDNAs, Pa138 and Pa147, in *Prospero autumnale* diploid genomes as well as diploid homoploid hybrid and Group I-IV allotetraploids of different genomic origin revealed many polymorphisms. The localization of satDNA *PaB6* was consistent with previous analysis. It was only detected in chromosomes inherited from the  $B^6B^6$  and Group I parents. In contrast, the other two satDNAs had unique loci distribution patterns and were polymorphic. The evolution of the *Prospero autumnale* complex is not accompanied by morphological or karyotypic changes, but by changes in the repeatome, specifically satDNAs, most dynamic components of their genomes.

## 7. Zusammenfassung

Der *Prospero autumnale* Komplex ist chromosomal sehr vielfältig und umfasst mehrere diploide Cytotypen mit unterschiedlichen Basischromosomenzahlen ( $x = 5, 6$  und  $7$ ) und Genomgrößen sowie zahlreiche Auto- und Allopolyploide. Dadurch eignet er sich ideal zur Analyse der Genomevolution und der Rolle chromosomaler Veränderungen in der Artbildung. Die Analyse der repetitive DNA-Profile von zwei diploiden Cytotypen,  $B^6B^6$  ( $x = 6$ ) und  $B^7B^7$  ( $x = 7$ ), und ihren Gruppe I Allotetraploiden ( $2n = 25-28$ ) ergab in allen Genomen ähnliche und hohe Anteile repetitiver DNA (ca. 75%). Der am häufigsten vorkommender Typ repetitiver DNA in allen drei Genomen war die Retrotransposon Superfamilie Ty1-*copia* (ca. 40%). DNA Transposone und tandemartige Wiederholungen machten nur einen kleinen Teil der Genome aus. Die Analyse ermöglichte die Identifizierung der Satelliten-DNA *PaB6* und vier neuer Satelliten-DNAs, die alle nur einen kleinen Teil der Genome ausmachten. Die Kartierung der Satelliten-DNA *PaB6* und zwei neuer Satelliten-DNAs, Pa138 und Pa147, in Genomen von diploiden *Prospero autumnale* sowie diploiden homoploiden Hybriden und Gruppe I-IV-Allotetraploiden unterschiedlicher genomischer Herkunft zeigte viele Polymorphismen. Die Lokalisierung der Satelliten-DNA *PaB6* war konsistent mit früheren Analysen. Sie wurde nur in Chromosomen nachgewiesen, die von den Eltern  $B^6B^6$  und Gruppe I geerbt wurden. Im Gegensatz dazu hatten die anderen beiden satDNAs einzigartige Loci-Verteilungsmuster und waren polymorph. Die Evolution des *Prospero autumnale* Komplexes wird nicht von morphologischen oder karyotypischen Veränderungen begleitet, sondern von Veränderungen repetitiven DNA-Elementen, insbesondere den Satelliten-DNAs, den dynamischsten Komponenten ihrer Genome.