

First insight into the genomes of the *Pulmonaria officinalis* group (Boraginaceae) provided by repeatome analysis and comparative karyotyping

Lucie Kobrlová

Palacký University

Jana Čížková

Institute of Experimental Botany

Veronika Zoulová

Institute of Experimental Botany

Kateřina Vejvodová

University of South Bohemia

Eva Hřibová

hribova@ueb.cas.cz

Institute of Experimental Botany

Research Article

Keywords: Comparative karyotyping, Genome size, *Pulmonaria*, Repeatome, Satellite DNA

Posted Date: March 25th, 2024

DOI: <https://doi.org/10.21203/rs.3.rs-4148849/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Additional Declarations: No competing interests reported.

Version of Record: A version of this preprint was published at BMC Plant Biology on September 13th, 2024. See the published version at <https://doi.org/10.1186/s12870-024-05497-4>.

Abstract

Background

The genus *Pulmonaria* (Boraginaceae) represents a taxonomically complex group of species in which morphological similarity contrasts with striking karyological variability. The presence of different numbers of chromosomes in the diploid state suggests multiple hybridization/polyploidization events followed by chromosome rearrangements (dysploidy). Unfortunately, the phylogenetic relationships and evolution of the genome, have not yet been elucidated. Our study focused on the *P. officinalis* group, the most widespread species complex, which includes two morphologically similar species that differ in chromosome number, i.e. *P. obscura* ($2n = 14$) and *P. officinalis* ($2n = 16$). Ornamental cultivars, morphologically similar to *P. officinalis* (garden escapes), whose origin is unclear, were also studied. Here, we present a pilot study on genome size and repeatome dynamics of these closely related species in order to gain new information on their genome and chromosome structure.

Results

Flow cytometry confirmed a significant difference in genome size between *P. obscura* and *P. officinalis*, corresponding to the number of chromosomes. Genome-wide repeatome analysis performed on partial Illumina sequencing data showed that retrotransposons were the most abundant repeat type, with a higher proportion of Ty3/Gypsy elements, mainly represented by the Tekay lineage. Comparative analysis revealed no species-specific retrotransposons or striking differences in their copy number between the species. A new set of chromosome-specific cytogenetic landmarks, represented by satellite DNAs, showed that the chromosome structure in *P. officinalis* was more variable compared to that of *P. obscura*. Comparative karyotyping strongly supported the hybrid origin of putative hybrids with $2n = 15$ collected from a mixed population of both species and outlined the origin of ornamental garden escapes, confirming their derivation from the *P. officinalis* complex.

Conclusions

Large-scale genome size analysis and repeatome characterization of the two morphologically similar species of the *P. officinalis* group improved our knowledge of the genome dynamics and differences in the karyotype structure. A new set of chromosome-specific cytogenetic landmarks was identified and used to reveal the origin of putative hybrids and ornamental cultivars morphologically similar to *P. officinalis*.

Background

The genus *Pulmonaria* L. (Boraginaceae, *sensu* [1]) is a taxonomically complex group of species in which the rather similar morphology contrasts with striking karyological variability. In total, 16 different somatic chromosome counts ranging from $2n = 14$ to $2n = 38$ are currently reported in the genus [Kobřlová *unpubl.*], with $x = 7$ proposed as the basic chromosome number [e.g. 2, 3, 4, 5, 6]. According to the traditional morphology- and karyology-based taxonomy [cf. 3, 6], about 30 taxa are recognized (16 species and 14 subspecies), all found in Europe, with the exception of the *P. mollis* group, which extends as far as north-eastern and eastern Asia [5, 7, 8, 9]. Although almost nothing is known about the mechanisms of the chromosomal evolution of the genus, such a karyological diversity is indicative of extensive chromosomal rearrangements, i.e. the hypothesis of ancient

episodes of polyploidization/hybridization and subsequent dysploidal chromosome reorganizations [cf. 6, 10]. Considerable variability in chromosome number and genome size has been reported within the Boraginaceae family [summarized in 11, 12], but its role in species diversification and evolution has mostly only been hinted at or discussed [but see e.g. 10, 13, 14, 15].

To date, only a few attempts have been made to explore the evolutionary history of the genus *Pulmonaria*, focusing on a limited or specific group of species [10, 16, 17, 18, 19]. These studies have highlighted the significant role of hybridization (e.g. the recent hybrid origin of *P. helvetica* Bolliger [19] or the *P. hirta* complex [10]), and provided evidences for the involvement of introgression and dysploidy in the speciation process [10, 18]. In horticulture, *Pulmonaria* species are known to be crossable (no geographical isolation, flowering synchrony, similar pollinator preferences), giving rise to a wide variety of cultivars [see 20]. Even in the contact zones of some taxa, intermediate chromosome numbers have rarely been documented [2, 3, 21, 22]. However, most of the evolutionary relationships are still unknown, and the question remains as to what lies behind the observed variability in their morphology and chromosome number.

Along with chromosome number, genome size is another relevant biodiversity trait that can indicate significant genomic events and evolutionary changes [23, 24]. Diversity in genome size is driven by multiple evolutionary processes such as polyploidy, the proliferation of repetitive DNA sequences and/or drastic reduction of non-genic DNA [23, 25, 26, 27]. Typically, post-polyploid diploidization is thought to be associated with extensive loss of DNA or 'genome downsizing' [28, 29, 30, 31, 32, 33, 34]. However, the mechanisms, rates and selection pressures driving these changes in DNA content remain unknown [cf. 33]. Genome size can also be a useful tool for the practical botany, i.e. distinguishing between cryptic taxa, identifying hybrids and resolving taxonomies in critical groups [e.g. 35, 36, 37, 38]. According to the available literature, the genome size values of *Pulmonaria* species have only been reported sporadically [summarized in 39], without any systematic study.

As already mentioned, karyological studies in *Pulmonaria* have mainly focused on determining the number of chromosomes and their significance for taxonomy [e.g. 3, 21]. A more detailed analysis of chromosome structure at the cytogenetic level has not yet been carried out. The most important technique in plant cytogenetic research is fluorescence *in situ* hybridization (FISH). The major limiting factor of the FISH techniques in non-model plant species is the lack of powerful DNA probes [reviewed in 40, 41]. Recent development of whole genome sequencing technologies and bioinformatics facilitates fast progress of cytogenomics. In non-model plants, the first step is identification of suitable cytogenetic landmarks, which provide specific pattern on chromosomes and can be used to identify individual chromosomes, to compare karyotypes and chromosomal rearrangements in closely related species and inter-specific hybrids, and to provide important information on the plant evolution through chromosome changes [e.g. 42, 43, 44, 45, 46]. The best performing cytogenetic probes include rRNA genes and satellite DNAs (satDNA). Both of them are present in high copy numbers of tandem organized repetitive units in the genomes, and very often provide species-specific patterns on chromosomes [e.g. 27, 47, 48, 49, 50].

To identify such cytogenetic probes, whole genome sequencing data can be utilized and further analyzed by the RepeatExplorer pipeline [51]. RepeatExplorer pipeline is based on a graph-based clustering algorithm and enables identification and characterization of repetitive DNA from small quantities of short read sequencing data [52, 53]. The pipeline has been used for repeat identification in many model and non-model plant species [54, 55]. The new version of RepeatExplorer2 provides *de novo* repeat identification and annotation of transposable elements, comparative repeat analysis in large sets of species, and identification of satellite DNA with the aim to develop

cytogenetic probes. The analyses are complemented by visualization of the results in the SeqGrapherR visualization tool [55].

Our study is focused on the *P. officinalis* group, the most widespread European species complex [3, 56, 57]. The *P. officinalis* group contains two morphologically relatively distinct species differing in chromosome numbers, namely *P. obscura* Dumort. ($2n = 14$) and *P. officinalis* L. ($2n = 16$ [e.g. 3, 22, 58, 59, 60]). Within the latter, two subspecies are sometimes recognized [61, 62, 63]. As the *P. officinalis* complex has been cultivated for medicinal and ornamental purposes for a very long time [64, 65], it has probably become even more widespread. In some regions it may also be non-native as a result of frequent cultivation and possible garden escapes [60, 65]. “Pure” *Pulmonaria* species have often been crossed in cultivation to produce more attractive varieties [cf. 20], which are rarely found in nature. Many of these plants are listed under the name *P. saccharata* in the horticultural trade. Their origin is not clear, but it can be assumed that some of them are derived from *P. officinalis* (esp. plants with distinctly white-spotted leaves, cordate at the base). However, they are not “true” *P. saccharata sensu* Miller, whose taxonomic status has long been debated, and are most likely related to the *P. hirta* complex [cf. 10, 66, 67].

In this work, we performed the analysis of genome size and comparative analysis of the repeatomes in closely related species of the *P. officinalis* group in order to shed light on the karyotype structure and variability. Specifically, we have analyzed repeatomes and karyotypes of *P. obscura* and *P. officinalis* s.str. individuals from pure separate populations, their putative hybrid accessions with $2n = 15$, which have been collected in a mixed population where *P. obscura* and *P. officinalis* grew together, and three populations of ornamental cultivars, morphologically similar to *P. officinalis*, that have escaped into the wild where they were collected (here listed as *P. saccharata*-like). Genome size has been studied and compared on a wider geographical scale. This pilot study can serve as a springboard for future cytogenetic and genomic studies, to understand the role of chromosomal rearrangements in the evolution of this genus.

Results

Genome size and chromosome number variation in the *P. officinalis* group

A total number of 196 accessions from 65 populations of the *Pulmonaria officinalis* group were collected throughout Europe and analyzed by flow cytometry (Fig. 1A). The analyses resulted in good quality histograms with distinct peaks and low coefficients of variation (below 5%) for both PI and DAPI staining (Supplementary Table 1). The flow cytometry data were confirmed by chromosome counts, some of which were obtained *de novo* (see below), but previously published reports were also revised. A total of 754 published chromosome records were found and revised for the *P. officinalis* complex (Supplementary Table 2). Of these, 287 chromosome reports with $2n = 14$ were related to *P. obscura* and 450 with $2n = 16$ to *P. officinalis* s.str., respectively (Fig. 1A). Five records were excluded as they probably belong to another species of the genus. In one case $2n = 17$ have been reported for *P. officinalis*. In addition to this, there are even karyological records (16 in total) that refer to hybrids of *P. obscura* and *P. officinalis*. In all cases, 15 chromosomes were counted in these plants (Supplementary Table 2).

P. obscura ($2n = 14$; mean $2C = 2.92 \pm 0.10$ pg) and *P. officinalis* s.str. ($2n = 16$; mean $2C = 3.21 \pm 0.13$ pg) differed significantly in holoploid genome size ($2C$ value, Fig. 2A) as well as in the genomic GC content (Fig. 2B). Their DNA amount ranges are listed in Table 1. The genome size value of the presumed hybrid plants (B481) with $2n = 15$ ($2C = 3.08 \pm 0.02$ pg) was positioned between their values (within the theoretically expected intermediate genome size

based on the genome sizes of *P. obscura* and *P. officinalis*). The GC content was also intermediate (Table 1). Among *P. saccharata*-like populations, B465.1 with $2n = 16$ had the genome size value ($2C = 3.13$ pg) in the range of *P. officinalis*. The remaining two populations (B15: mean $2C = 3.74 \pm 0.03$ pg; B472.1: $2C = 3.63$ pg) had larger genome sizes than *P. obscura* and *P. officinalis*. Interestingly, these populations differ in the number of chromosomes (B15: $2n = 15$ vs. B472.1: $2n = 16$). The similar pattern was detected in the genomic GC content, only in B472.1 the GC content was lower than in B15 (Table 1).

Table 1
Summary of flow cytometric analyses of the *Pulmonaria officinalis* group.

Taxon	$2n$	N pop	N ind	Mean 2C [pg]	Min	Max	1C [Mbp]	C/n [pg]	GC [%]
<i>P. obscura</i>	14	25	83	2.92 ± 0.10	2.73	3.08	1 428	0.21	35.89
P. officinalis s.str.	16	36	103	3.21 ± 0.13	2.95	3.48	1 570	0.20	35.09
P. obscura × P. officinalis (B481)	15	1	3	3.08 ± 0.02	3.06	3.09	1 511	0.21	35.42
P. saccharata-like (B15)	15	1	5	3.74 ± 0.03	3.71	3.78	1 829	0.25	35.64
P. saccharata-like (B465.1)	16	1	1	3.13	-	-	1 531	0.20	35.16
P. saccharata-like (B472.1)	16	1	1	3.63	-	-	1 775	0.23	35.19

Genome-wide repeatome analysis

To identify the major types of repetitive sequences and to compare their genome representation in *P. officinalis* group, the comparative mode of RepeatExplorer2 pipeline was used. The analysis was performed on partial Illumina sequencing data of *P. obscura* ($2n = 14$; B473.1, Fig. 1B), *P. officinalis* ($2n = 16$; B470.3, Fig. 1C), their putative interspecific hybrid ($2n = 15$; B481.1) and three *P. saccharata*-like accessions (B15.1; $2n = 15$, Fig. 1D and B465.1, B472.1; both $2n = 16$). The identified repetitive sequences accounted for 58.48% of the *P. obscura* genome and 47.86% of the *P. officinalis* genome. A similar and overall highest proportion of repetitive sequences was found in repeatomes of the putative interspecific hybrid B481.1 (64.10%) and *P. saccharata*-like accession B15.1 (64.53%), while the repetitive DNA content of the other two *P. saccharata*-like accessions (B465.1 and B472.1) was around 50% (Table 2, Fig. 3).

In all studied taxa, LTR retroelements were the most abundant repeats, accounting for 46.68% of *P. obscura* genome, 34.52% of *P. officinalis* genome and 33.80–49.96% of the other genomes. The genome proportion of Ty3/Gypsy elements, which were mainly represented by Tekay lineage, ranged from 23.90% in *P. saccharata*-like (B472.1) to 35.06% in putative interspecific hybrid (B481.1). Ty1/Copia elements were about twice less abundant compared to Ty3/Gypsy superfamily, and were mostly represented by elements of SIRE and Angela clades (Table 2, Fig. 3). DNA transposons and long interspersed nuclear elements (LINE elements) were found in low copy numbers in all analyzed groups, with genome proportions ranging from 0.14–0.23% and from 0.16–0.22%, respectively (Table 2). rDNA sequences accounted for about 3.00–5.00%, and other tandem organized repeats represented 1.25% to more than 4.00% of the *Pulmonaria* genomes (Table 2).

Table 2
Proportion of repetitive DNA sequences identified *de novo* in *Pulmonaria* taxa

Repeat		Lineage/class	Proportion of repeats in genomes [%]					
			POBS	POFF	HYBR	PSAC	PSAC	PSAC
			B473.1	B470.3	B481.1	B465.1	B472.1	B15.1
LTR retroelements	Ty1/Copia	SIRE	12.05	6.47	10.38	6.48	6.18	9.83
		Angela	2.63	2.18	3.01	2.17	2.21	3.07
		Tork	1.00	0.93	1.00	0.99	0.96	1.03
		TAR	0.03	0.03	0.03	0.02	0.03	0.03
		Ale	0.07	0.07	0.08	0.08	0.06	0.07
		Total Ty1/Copia	15.78	9.68	14.50	9.74	9.48	14.03
Ty3/Gypsy		Ogre	0.82	0.82	0.92	0.80	0.74	0.91
		Athila	0.04	0.09	0.08	0.10	0.08	0.07
		Tekay	29.69	23.41	33.97	24.84	23.00	32.67
		Chromovirideae	0.07	0.08	0.09	0.08	0.08	0.08
		Total Ty3/Gypsy	30.62	24.40	35.06	25.82	23.90	33.73
Unclassified LTR elements			0.28	0.44	0.40	0.43	0.42	0.47
Total LTR			46.68	34.52	49.96	35.99	33.8	48.23
Other	LINE		0.23	0.17	0.23	0.16	0.14	0.21
	DNA transposons		0.16	0.20	0.18	0.22	0.18	0.22
	Tandem repeats		3.78	1.25	4.33	3.33	2.12	3.60
	rRNA genes		3.78	4.36	2.78	4.52	4.73	5.09
Unclassified repeats			3.85	7.36	6.62	7.84	7.61	7.18
Total repetitive DNA content			58.48	47.86	64.1	52.06	48.58	64.53
POBS: <i>P. obscura</i> ; POFF: <i>P. officinalis</i> ; HYBR: <i>P. obscura</i> × <i>P. officinalis</i> , putative interspecific hybrid; PSAC: <i>P. saccharata</i> -like								

Variability in the Satellite DNAs and rDNA loci

Considering satellite (tandem organized) repeats as preferable cytogenetic landmarks, which can be species- or even subspecies-specific, we inspected the individual *P. obscura* and *P. officinalis* datasets by TAREAN program.

Five putative satellites (tandem organized repeats) were observed, three of them (PulTR01_29, PulTR03_308, PulTR05_70) were shared by both species. One satDNA (PulTR02_305) was only found in *P. obscura*, and another one (PulTR04_420) was only detected *in silico* in the *P. officinalis* genome. Highly abundant tandem organized repeat PulTR05_70 provided strong cluster signals distributed along whole chromosomes with enrichments in pericentromeric regions, and cannot be used as a chromosome-specific cytogenetic landmark (Suppl. Figure 1). The other tandem repeats produced chromosome-specific signals and were used together with rDNA sequences to create karyotypes of the *Pulmonaria* accessions studied.

Tandem organized character is also typical for rRNA genes, which are present in the genomes in high copy numbers and localized in specific chromosomal regions. The 45S rRNA gene unit is usually about 10 kb long and is often represented by several cluster graphs after the RepeatExplorer analysis, as it was observed in the case of all the *Pulmonaria* accessions analyzed. To reconstruct the whole 45S unit, we followed the procedure described in Kapustová *et al.* [68]. The unit length of the *Pulmonaria* 45S rDNA unit varied from 9.1 kb to 9.6 kb and was found to be highly conserved at the sequence level within all accessions studied (Suppl. Figure 2). The intergenic spacer (IGS) region contained two different tandemly organized repeats (with repetitive units of 117 and 150-nt long) which were identical in all the accessions studied, with the exception of a *P. saccharata*-like plant B472.1, which contained tandem regions with shorter repetitive units (79 and 150-nt long). A higher variability was found in the IGS, which contained two relatively large INDEL regions that differed between *P. obscura* and *P. officinalis* (Suppl. Figure 3A). To support this observation, we performed read mapping of *Pulmonaria* accessions to the assembled *P. obscura* 45S rDNA reference (Suppl. Figure 3B). The differences in read coverage along the 45S rDNA unit are evident in the IGS, indicating the variability at the sequence level. It should be pointed out that the assembly of the 45S rDNA unit was obtained from short Illumina sequences, so the resulting consensus sequence may represent the most abundant sequence type.

The 5S rRNA gene unit is typically up to 1kb in length in plant species [69], so the graph-based clustering algorithm of the RepeatExplorer2 pipeline allowed its complete reconstruction, even in the comparative analysis (Fig. 4). The graph region representing the 5S rRNA gene was shared by all *Pulmonaria* accessions studied (Fig. 4B). Three variable graph loops emanating from this conserved graph region correspond to three types of 5S rDNA unit, differing in the length of their IGS (Fig. 4A, B). Mapping of sample-specific sequencing reads onto the graph topology revealed the presence of different 5S rDNA units between *P. obscura* and *P. officinalis* (Fig. 4C, D). The cluster layout of B481.1, the putative hybrid between *P. obscura* and *P. officinalis*, was represented by all three loops specific to its presumed progenitors (Fig. 4E). A similar situation was observed for the *P. saccharata*-like accession B15.1 (Fig. 4F). The other *P. saccharata*-like individuals (B465.1 and B472.1) shared the same graph layout as *P. officinalis* (i.e. B470.3; Fig. 4G, H).

The analysis of 5S rDNA specific graph layouts observed by clustering analysis of individual accessions confirmed the observation of the comparative analysis. All *Pulmonaria* individuals studied were composed of at least two different types of 5S rDNA genes, which differed in the IGS (Suppl. Figure 4).

Comparative karyotyping in the *P. officinalis* group

Molecular karyotyping was performed using newly identified satellites (PulTR01_29, PulTR02_305, PulTR03_308 and PulTR04_420) and 5S and 45S rDNA sequences. In general, *in situ* hybridization confirmed the results obtained by repeatome analysis. FISH analysis with the probes for rDNAs and four satDNAs resulted in well visible cluster signals on specific chromosomes in the genome of the analyzed plants of the *P. officinalis* complex.

FISH analysis of *P. obscura* plants (all $2n = 14$) collected from three different populations (B467, B469, B473; Suppl. Table S1) provided highly consistent results. The only exception was observed for plant B473.1 in the number of the most abundant satDNA (PulTR01_29). The 45S rDNA was located into terminal NOR regions on four chromosome pairs (Fig. 5; Fig. 6A, B, D, E). 5S rDNA loci were detected on two chromosome pairs in pericentromeric regions. One chromosome pair contained signals of 5S, 45S rDNA and satellite PulTR02_305 (Fig. 5; Fig. 6B, E). In B473.1, the probe for PulTR01_29 provided very strong signals in subtelomeric regions only on one chromosome pair. The same chromosome pair also contained a signal of PulTR03_308 in the pericentromeric region (Fig. 6C). In B467.2 and B469.1, one additional subtelomeric signal of PulTR01_29 was found, located on a chromosome arm with 5S rDNA locus (Fig. 5; Fig. 6A, D, F). As expected from the results of the RepeatExplorer2 analysis, signals of PulTR04_420 were not detected in any of the *P. obscura* plants analyzed.

In general, the genome of *P. officinalis* contained one additional pair of chromosomes (i.e. $2n = 16$) compared to *P. obscura*. Molecular karyotype was studied in individuals from two different populations (B100, B470; Suppl. Table S1). A higher variability in the chromosomal distribution of the probes was detected within representatives of *P. officinalis* compared to *P. obscura* (Suppl. Figure 5). Terminal NORs (45S rDNA) were found only on three chromosome pairs in both analyzed individuals (B100.2, B470.1). These terminal 45S rDNA loci were often fragile and broken off from the chromosomes (Fig. 7A, F). Interstitial 45S rDNA clusters were detected on one chromosome pair, on the same arm which also contained signals of satDNA PulTR04_420 and PulTR01_29 (Fig. 5; Fig. 7B, F). The genome of B100.2 contained an additional weak signal of interstitial 45S rDNA (Fig. 5; Fig. 7D, E, F). Variability in the number of signals specific to 5S rDNA loci was detected. While 5S rDNA loci were identified in pericentromeric regions on seven chromosomes of B470.3, only six chromosomes were bearing these loci in B100.2. The signal of PulTR03_308 was detected on one NOR bearing chromosome pair in both individuals with various combinations of co-localization with PulTR04_420 and 5S rDNA (Fig. 5; Fig. 7C, E, G). Signals of PulTR02_305 were not detected on chromosomes of *P. officinalis*, supporting the results of comparative repeatome analysis by RepeatExplorer2. FISH analysis of the other representative of the same population B100 provided slightly different molecular karyotype (Suppl. Figure 5).

Karyotype analysis of the putative natural hybrid between *P. obscura* and *P. officinalis* (B481.1; Suppl. Table S1), collected in a mixed population of both species, confirmed the expected chromosome number $2n = 15$. In this case, eight 45S rDNA clusters were found, seven in terminal chromosomal regions and one in an interstitial position. 5S rDNA loci were detected on five chromosomes in pericentromeric regions. One chromosome pair contained signals of 45S rDNA and 5S rDNA, another individual chromosome contained 45S and 5S rDNA and PulTR02_305 (found only as one single locus in the genome). Additional two chromosomes contained one or two signals of 5S rDNA, respectively (Fig. 5; Fig. 8A, B). Five signals of the most abundant satellite PulTR01_29 were detected on five chromosomes, one of which co-localized with the 45S rDNA locus, one signal was detected on the chromosome bearing weak interstitial signal of 45S and one with signal of PulTR03_308 and PulTR04_420, respectively. One chromosome pair with the remaining two signals of PulTR01_29 also contained a signal of 5S rDNA. SatDNAs PulTR03_308 and PulTR04_420 co-localized on one chromosome pair, containing also PulTR01_29, or joint signal of 45S rDNA and PulTR01_29 (Fig. 8C).

Ornamental garden escapes, morphologically similar to *P. officinalis*

Karyotype analysis of *P. saccharata*-like plants from three populations, which were collected in the wild but apparently escaped from cultivation (B15, B465, B472; Suppl. Table S1), revealed variability in chromosome number. While B465.1 and B472.1 were characterized by $2n = 16$, $2n = 15$ was detected in B15.1. The karyotypes of B465.1 and B472.1 were very similar, containing terminal NORs on three chromosome pairs and one interstitial 45S rDNA cluster on one additional chromosome pair (Fig. 8G, I, K, L, M). Two chromosome pairs were bearing PulTR01_29 in subtelomeric regions and other two chromosome pairs contained 5S rDNA clusters in pericentromeric regions (Figure G, H, K). The signal of PulTR04_420 was found on two chromosome pairs (Fig. 8G, J, K, L). One of these chromosome pairs also contained signals of PulTR01_29 and an interstitial signal of 45S rDNA, while the other chromosome pair bearing terminal 45S rDNA and the PulTR04_420 co-localized with PulTR03_308 (Fig. 5; Fig. 7; Fig. 8G, J, K, L). However, these two accessions differed in the presence of an additional 5S rDNA and PulTR01_29 interstitial loci (Fig. 5; Fig. 7).

The karyotype of B15.1 was similar to that of the putative interspecific hybrid (B481.1) with the same number of chromosomes, indicating a similar hybrid character. 45S rDNA loci were located as strong signals in terminal regions of three chromosome pairs, one additional chromosome contained a weak terminal signal and another contained an interstitial signal (Fig. 8D, E). Odd number of signal localizations was detected for 5S rDNA (5 chromosomes with interstitial signals), PulTR01_29 (three chromosomes with the signals in terminal regions), PulTR04_420 (three chromosomes with interstitial signals), and for PulTR02_305 (found only on one chromosome) (Fig. 8D, E, F). Finally, PulTR03_308 provided signals on one chromosome pair co-localizing with PulTR04_420 (Fig. 5; Fig. 7).

Discussion

The genus *Pulmonaria* is karyologically highly variable [e.g. 3, 6], with 16 different somatic chromosome numbers reported (Kobrová *unpubl.*) and about 30 taxa recognized, growing in Europe and northeastern and eastern Asia [5, 7, 8, 9]. However, the mechanism, origin and evolutionary consequences of genome size and karyotype variability remain unexplored [but see 10]. Apparently, chromosomal rearrangements have played an important role in the evolution of this genus [cf. 10, 19, 39], but how and to what extent has never been clearly demonstrated.

In this work, we studied the *P. officinalis* complex (Fig. 1), which includes two morphologically similar, closely related species that differ in chromosome number [3, 10, 18, 59, 60]. The present study confirmed previously reported chromosomal data, which were consistent across the whole of Europe. The predominant $2n = 16$ reports correspond to *P. officinalis* s.l., whereas only $2n = 14$ were documented in *P. obscura* populations (Fig. 1). Only in one case, $2n = 17$ was documented from the “pure” population of *P. officinalis* [22], but no explanation was given. Despite the morphological and karyological differences observed, the origin of this species group remains unknown. To uncover the differences in the karyotype evolution, we used complex methodological approaches involving partial Illumina sequencing followed by bioinformatic analysis and characterization of repeatomes in the *P. officinalis* group. We identified a new set of chromosome-specific cytogenetic landmarks and performed comparative karyotyping within and between the two species, their putative natural hybrid from a population where both species occur, and ornamental cultivars morphologically similar to *P. officinalis*, which are also rarely found in nature.

Impact of DNA repeats dynamics on genome size

Genome size can reflect some aspects of the evolutionary history of taxa by allowing us to understand the influence of DNA gain/loss between related species [e.g. 70, 71]. Despite the apparent karyological variability of several Boraginaceae genera [see e.g. 14, 72, 73], there have been almost no complex analyses of genome size variation and the evolutionary pathways behind the observed diversity [but see 74, 75, 76]. Given the sparse DNA content records in the Boraginaceae family, the only comprehensive study has been published, providing the first genome size reports for most taxa [12].

Our study represents the first large-scale investigation of interspecific genome size variation in *Pulmonaria*. As already shown in a pilot study by Koblířová & Hroneš [12], genome size is effective in delimiting morphologically similar taxa of the Boraginaceae, which is also true for the *P. officinalis* group. We found a significant difference in genome size between *P. obscura* and *P. officinalis*, corresponding to the number of chromosomes (Fig. 2), confirming previous results [cf. 12], but on a larger geographical scale. The suitability of using flow cytometry to revise the distribution of the *P. officinalis* group (i.e. relative genome size) has already been documented in the Bohemian Forest and adjacent foothills [60].

So far, the genome size has only been estimated for eight *Pulmonaria* taxa, including *P. officinalis* and *P. obscura*, ranging from 2.27 to 4.27 pg (i.e. very small/small genomes according to the categories of Leitch et al. [77], Supplementary Table 3). Only minor differences were observed when comparing previously analyzed genome sizes of *P. obscura* and *P. officinalis* with our data, most likely due to different methodologies used (i.e. nuclei isolation buffer, reference standard, plant organ [cf. 12, 39]). The only exception is the study by Šmarda *et al.* [78], where almost the same 2C values are presented for *P. obscura* and *P. officinalis*, probably as a consequence of taxa misidentification.

DNA transposons have been shown to be the major contributor to the enormous variation in genome size in plants [e.g. 27, 79, 80, 81, 82, 83]. To shed light on genome size dynamics and relationships between *P. officinalis* and *P. obscura* species, we performed a genome-wide comparison of their repeatomes. We found that most of the repetitive elements in the genomes of the *Pulmonaria* taxa studied were dispersed repeats represented by LTR retrotransposons [cf. 27, 81], with higher proportion of Ty3/Gypsy elements, which were twice more abundant than Ty1/Copia. Ty3/Gypsy retroelements were almost exclusively represented by Tekay retrotransposons (Chromoviridae clade), whereas SIRE elements were the most abundant types of the Ty1/Copia superfamily (Table 2). Ty3/Gypsy elements represent one of the major classes of LTR retrotransposons and are dominant in many plant groups, such as the family Poaceae [e.g. 82, 84, 85, 86] or the tribe Fabaeae [27, 87]. Unfortunately, a genome-wide analysis of DNA repeats and their impact on genome size has not been performed in any other species of the Boraginaceae family. However, the higher proportion of Ty3/Gypsy retroelements have also been found in genera of the closely related Solanaceae family, such as *Solanum*, *Nicotiana* and *Capsicum* [88, 89, 90, 91, 92, 93]. In contrast, recent studies in the genus *Salvia*, a member of the closely related Lamiaceae family, have shown that the nuclear genomes of different species contain different proportions of Ty3/Gypsy and Ty1/Copia retroelements [94, 95, 96], indicating a proliferation of different types of DNA repeats during the evolution of individual species. In comparison, the *Pulmonaria* species analyzed contained a similar proportion of the repeat lineages and individual clusters were represented by reads from all specimens analyzed. This indicates a high degree of genome homology within the *P. officinalis* complex, suggesting that the evolution of this species group was not accompanied by a dramatic diversification of DNA transposons, as previously shown in other plant species [e.g. 82]. To better understand the proliferation of DNA repeats during genome evolution and its impact on genome size variation and speciation, analysis of a larger data set of *Pulmonaria* species from different phylogenetic groups is required.

Satellite DNAs and their use in comparative karyotyping

Repeatome analysis using the RepeatExplorer pipeline also allows the identification of putative satellite DNAs, which together with tandem organized ribosomal genes are the best cytogenetic landmarks. SatDNAs are usually species- or subspecies-specific, provide chromosome-specific labeling patterns and can therefore be used not only to generate and compare molecular karyotypes, but also to identify putative chromosomal structural changes [42, 45, 97]. The use of the FISH technique revealed different patterns of chromosomal localization of the tandem repeats and rDNA loci examined, both between and within *P. obscura* and *P. officinalis* (Fig. 5). The almost identical cytogenetic pattern of satDNAs and rDNA sequences in *P. obscura* ($2n = 14$), collected from three different populations, suggests karyotype stability in this diploid species (Fig. 5). In comparison, the chromosome structure in *P. officinalis* appears to be more dynamic, as individuals from two different populations differ slightly in the cytogenetic pattern of the satDNAs and rDNA sequences. Odd number of signals of some satDNAs as well as of rDNA sequences, and interstitial 45S rDNA loci were found in both diploid accessions ($2n = 16$), indicating chromosomal structural changes involved in the origin and evolution of *P. officinalis* (Fig. 5). It is generally accepted that $n = 7$ is the basic chromosome number in *Pulmonaria* [e.g. 2, 3, 4, 5, 6], which raises the question of how the species represented by different chromosome numbers arose. In the case of *P. officinalis*, there are two possible scenarios. There could have been chromosome fission leading to $2n = 16$, or a more complex karyotype evolution with polyploidization and further diploidization by chromosome rearrangements. Large differences between the molecular karyotypes of *P. obscura* and *P. officinalis* suggest that diploid *P. obscura* most probably could not give a rise to *P. officinalis* by chromosome fissions (Fig. 5). Thus, we can only speculate, that the evolution and speciation in the *P. officinalis* group could be influenced by polyploidy and/or hybridization followed by post-polyploid diploidization process, which can result in numerous chromosomal rearrangements [see 98, 99, 100] and the origin of *Pulmonaria* species differing in their basic chromosome number. A similar evolutionary scenario involving chromosome multiplication, hybridization and, in particular, structural rearrangements of chromosomes (dysploidy) has been outlined by Liu *et al.* [10] for the origin of the *P. hirta* complex on the Italian peninsula and the Swiss endemic *P. helvetica* [19], respectively.

Evidence of hybridization within the *P. officinalis* complex

The evolution of the genus *Pulmonaria* remains unresolved, due to a lack of rigorous phylogenomic studies. However, several molecular studies have been published highlighting the important role of hybridization and introgression in the evolution of the genus [10, 18, 19]. The high level of chromosomal variation may also support the hypothesis of ancient hybridization events and subsequent chromosomal rearrangements [cf. 6, 7, 10, 18]. In addition, some species groups exhibit weak ecological and geographic isolation, near-synchronous phenology and pollinator sharing, all of which may facilitate the hybridization [cf. 18]. This is particularly true for the *P. officinalis* complex, which is widespread in Europe and therefore often in secondary contact with other *Pulmonaria* species (cf. 56, 101]. As the ranges of *P. obscura* and *P. officinalis* partly overlap (Fig. 1A), the co-occurrence of both species in the same habitat can be expected. Some authors have occasionally reported mixed populations, with morphological intermediates rarely observed [21, 22]. The extent of hybridization between these two species is still controversial. Nevertheless, several karyological data referring to as *P. obscura* $\times$ *P. officinalis* with an intermediate number of chromosomes (i.e. $2n = 15$), may provide convincing evidence of an ongoing hybridization between these two species [18, 21, 22].

In our study, we analyzed presumed hybrids from a mixed population (B481) of *P. obscura* and *P. officinalis*. Chromosome counting in all three analyzed individuals confirmed $2n = 15$, and their hybrid origin was also

supported by their genome sizes, halfway between those of the parents (Table 1). The cytogenetic mapping of the set of satDNAs and rDNA sequences also strongly supports the hybrid origin, by the presence of *P. obscura* and *P. officinalis* species-specific satDNAs (PulTR_305 and PulTR_420) in haploid state, and also their pattern on chromosomes (Fig. 7). The hybrid status of these individuals was further confirmed by detailed analysis of the 5S rDNA sequences. As recently shown, graph-based clustering of the RepeatExplorer pipeline enables reconstruction of complete 5S rDNA sequences from partial Illumina sequencing data and provides clues to the evolutionary history of interspecific hybrids and allopolyploids [69, 102]. The detailed analysis of the 5S rDNA cluster shapes of all *Pulmonaria* accessions examined revealed the presence of both *P. obscura*- and *P. officinalis*-specific 5S rDNA sequences in the genome of a putative hybrid clone B481.1 (Fig. 4).

Given the overlapping ranges of some *Pulmonaria* species groups [see 56, 101], the possibility of interspecific hybridization might be expected to be relatively common. So far, however, natural hybrids have only occasionally been identified by chromosome counting [3, 21] or distinguished on the basis of intermediate morphology [e.g. 103]. Questions also remain about the hybrid's fertility and their longevity in populations. Although the hybridization is apparently rare, hybrids can potentially persist and spread clonally at the locality [cf. 104], as *Pulmonaria* species partly reproduce by vegetative propagation (i.e. creeping rhizomes).

Origin of ornamental cultivars morphologically similar to *P. officinalis*

As a valuable medicinal and ornamental plant, *P. officinalis* is represented in horticulture by several cultivars and has also been used to generate new artificial hybrids [cf. 20]. This seems to be the case for plants with distinctly white-spotted leaves, cordate at the base, which are sometimes offered commercially as *P. saccharata*. However, they are not “true” *P. saccharata sensu* Miller [see 10, 66, 67]. These plants often escape into the wild and are sometimes confused with *P. officinalis*. The origin of these cultivars is unknown, they only resemble *P. officinalis* complex in their morphology (i.e. cordate bases of leaves, Fig. 1D).

In total, we analyzed three populations of these garden escapees (B15, B465 and B472). Our cytogenetic analysis and detailed examination of the reconstructed 5S rDNA sequence indicate that two analyzed *P. saccharata*-like accessions (B465.1 and B472.1) are derived from the *P. officinalis*. Both plants had $2n = 16$ as *P. officinalis*. They also shared the general cytogenetic pattern of satDNAs and rDNA sequences typical of *P. officinalis*, and contain only *P. officinalis*-specific, not *P. obscura*-specific, 5S rDNA sequence types (Fig. 4, 5). In addition, the genome size value of the B465.1 plant was also in the range of *P. officinalis*, only in the B472.1 plant was slightly larger (Table 1). This may confirm the assumption that these plants are derived from *P. officinalis*.

On the other hand, an interesting cytogenetic pattern was observed in the third *P. saccharata*-like plant analyzed (B15.1). The karyotype of this plant was similar to that of the interspecific hybrid B481.1, with the same chromosome number $2n = 15$ and *P. obscura*- and *P. officinalis*-specific 5S rDNA sequences in its genome (Fig. 4). Furthermore, even the cytogenetic pattern of the probes and the presence of *P. obscura*- and *P. officinalis*-specific satDNAs (PulTR_305 and PulTR_420) in the haploid state also suggest a hybrid origin (Fig. 5). In contrast, the genome size of B15.1 (the whole population, respectively) was considerably larger, the largest found in the whole data set presented (Table 1). However, unlike the population B481, population B15 was collected in the area where only *P. obscura* occurs naturally and where no population of *P. officinalis* has been confirmed (Kobrllová, *pers. obs.*). The morphology of the plants was also typical for cultivated *P. saccharata*-like plants. Their origin therefore requires further investigation, although the cytogenetic data presented suggest a hybrid origin between *P. obscura*

and *P. officinalis* (e.g. phylogenetic revision and analysis of a larger data set of *Pulmonaria* species from different phylogenetic groups). As this population is a garden escape, its geographical origin is unclear and it cannot be ruled out that it was originally collected from a mixed population of both species. Our data undoubtedly demonstrate that molecular karyotyping is a powerful method for identifying the mode of karyotype evolution and the hybrid origin of *Pulmonaria* taxa.

Conclusions

Our study provides comprehensive information on genome size variability and repeatome dynamics of the two morphologically similar species of the *P. officinalis* group. Large-scale genome size analysis using flow cytometry confirmed a significant difference in DNA content between *P. obscura* and *P. officinalis*, corresponding to the number of chromosomes. Partial sequencing of six accessions, including putative natural hybrid of *P. obscura* and *P. officinalis*, and ornamental garden escapes resembling *P. officinalis*, showed that a large proportion of their genomes is represented by various types of DNA transposons, with Ty3/Gypsy elements being the most abundant. Comparative analysis of the repeatomes revealed no species-specific retrotransposons or striking differences in their copy number between the species, suggesting a common evolutionary history. Cytogenetic analysis using probes for rDNA sequences and newly identified satellite DNAs allows the origin of the *Pulmonaria* accessions to be determined, or at least outlined. Thus, in combination with a robust phylogenetic framework, it can contribute to the elucidation of the evolutionary history of *Pulmonaria* relatives. In our case, comparative karyotyping strongly supported the hybrid origin of putative hybrids with $2n = 15$, collected from a mixed population of *P. obscura* and *P. officinalis*, and also outlined the origin of ornamental garden escapes, confirming their derivation from the *P. officinalis* complex. Finally, databases of repeats were created, which can be used for repeat identification (or masking) in future sequencing projects.

Materials and methods

Plant material

A total of 196 plants from 65 populations of the *Pulmonaria officinalis* group (Fig. 1A), representing typical populations of *P. obscura* and *P. officinalis* s. str. (Fig. 1B, 1C), including their potential hybrids and several garden escapes of cultivars morphologically similar to *P. officinalis* (in horticulture often referred to as *P. saccharata*, here listed as *P. saccharata*-like, Fig. 1D), were included in this study (see Supplementary Table 1). These samples were collected between 2014 and 2023 from natural populations across Europe, some of which were cultivated in the experimental garden of Palacký University in Olomouc, Czech Republic, or deposited in the Herbarium of Palacký University in Olomouc (OL).

Flow cytometry: genome size and GC content

Estimation of the nuclear DNA content, i.e. absolute genome size (AGS [105]), and DNA base composition (GC content [106]) were estimated using Partec PAS and Partec ML instruments, with PI (propidium iodide) and DAPI (4,6-diamidino-2-phenylindole) staining. The same methodology as in Kobrlová and Hroneš [12] was followed, using fresh, rarely silica dried, leaves for sample preparation. *Pisum sativum* L. 'Ctirad' ($2C = 9.09$ pg [107]; GC content = 38.5% [108]) was selected as a primary internal standard, since it has non-overlapping genome size with neither G1 and nor G2 phase of all studied samples. The conversion from picograms (pg) to base pairs (bp) followed Doležel *et al.* [109], using 1 pg DNA = 978 Mbp. DNA base content was estimated using the protocol and

GC content calculation tool of Šmarda *et al.* [106]. The data analyses were performed using the NCSS 9 statistical software [110]. The non-parametric Kruskal-Wallis test was used to test for differences between population means of genome size/GC content of *P. obscura* and *P. officinalis*.

DNA extraction and sequencing

Genomic DNA was isolated using alkyltrimethylammonium bromide (MATAB) lysis: after sorbitol washes, the ground plant material was incubated in 2% (w/v) MATAB for 20 min at 65°C, immediately after the incubation, the same volume of chloroform:isoamyl alcohol (24:1) was added, gently but thoroughly mixed and centrifuged at 10,000 *g* for 3 min at 4°C. After centrifugation, aqueous upper phase was collected to a new tube and this step was repeat until the upper phase was clear. Genomic DNA was precipitated by adding 0.7 volume of isopropanol, centrifuged at 10,000 *g* for 3 min at 4°C. Finally, the pellet was washed by cold 70% and 96% ethanol, air dried and diluted in TE buffer, pH 8.

Genomic DNA was sheared by Bioruptor Plus (Diagenode, Liege, Belgium) to achieve an insert size of about 500 bp. Libraries for sequencing were prepared from 2 µg of fragmented DNA using TruSeq® DNA PCR-free kit (Illumina) and sequenced on a NovaSeq 6000 (Illumina), producing 2 × 100-bp or 2 × 150-bp paired-end reads to achieve at least 3Gb of nucleotide sequence per each genotype. Raw data were trimmed for low-quality bases and adapter sequences and to the same length using fastp v.0.20.1 [111].

Analysis and characterization of DNA repeats

Random datasets corresponding to 0.1× coverage of the individual accessions were used for reconstruction and characterization of DNA repeats using RepeatExplorer2 [55], that includes TAREAN analysis tool for identification of tandemly organized repeats [53]. RepeatExplorer2 and TAREAN analyses were also used to perform comparative analysis of *Pulmonaria* repeatomes on a merged dataset containing all studied individuals (1 mil. reads per accession), marked by specific prefixes. In both cases, the resulting clusters of repeats were characterized by various tools, including BLASTN and BLASTX, and phylogenetic analysis of the repetitive elements' coding domains [52, 112]. The presence of tandemly organized repeats within the clusters identified by TAREAN was confirmed with Dotter [113].

The results of the clustering were then used to create repetitive databases. Databases of Illumina reads were deposited in the Sequence Read Archive (project number: PRJNA1076467). Assembled contigs from different types of repetitive DNA elements are publicly available online (<https://olomouc.ueb.cas.cz/en/content/dna-repeats>). The sequences of newly identified tandem organized repeats and 5S rDNA which were used as cytogenetic markers were deposited in GenBank (accessions: PP457292–PP457296). Cluster graphs of 5S rDNA sequences were visualized using SeqGrapheR visualization tool [55].

Preparation of chromosome spreads

Mitotic metaphase chromosome spreads were prepared from root meristems by a dropping method according to Šimoníková *et al.* [44]. Briefly, actively growing root tips of *Pulmonaria* were collected and pre-treated in 0.05% (w/v) colchicine for three hours at room temperature, fixed in 3:1 ethanol:acetic acid fixative overnight at 4°C and stored in 70% ethanol at – 20°C. Chromosome preparations were prepared using the drop technique according to Kato *et al.* [114, 115], with minor modifications: After washing in 75 mM KCl and 7.5 mM EDTA (pH 4), root tip segments were digested in a mixture of 2% (w/v) cellulase and 2% (w/v) pectinase in 75 mM KCl and 7.5 mM

EDTA (pH 4) for 45 min at 37°C. The cell suspension was dropped onto glass slides in a box lined with wet paper towels and let dried.

Probe design and fluorescence in situ hybridization

Consensus sequences of TAREAN analysis which contained tandemly organized repeats were used for specific primer design using the Primer3 program [116]. Probes for newly identified tandem repeats were labeled by PCR either directly with Cy5 fluorochrome (Thermo Fisher Scientific), DEAC (Jena Biosciences, Jena, Germany), or indirectly with biotin-dUTP or digoxigenin-dUTP (Sigma Aldrich/Roche Applied Science, Mannheim, Germany) using primers listed in Supplementary Table S4 and *P. obscura* DNA as template. The 25 µl of PCR mix contained 30 ng of genomic DNA, 200 µM dNTPs including directly- or indirectly-labeled dUTP, 1 µM primers and 0.5 U of Q5 High-Fidelity DNA polymerase and appropriate reaction buffer (New England Biolabs, Massachusetts, USA). Plasmid pTa71 (45S rDNA) containing 9-kb fragment from *Triticum aestivum* with 18S-5.8S-26S rDNA and intergenic spacers [117] was labeled by nick translation (Sigma Aldrich) using Cy5 fluorochrome (Thermo Fisher Scientific).

Hybridization mixture containing 50% (v/v) formamide, 10% (w/v) dextran sulfate in 2 × SSC and 10 ng/µl of labeled probes was added onto slide and denatured for 30s at 80°C, followed by overnight hybridization performed in a humid chamber at 37°C. If the chromosome structure was damaged after the denaturation step, the slides with chromosome spreads were post-fixed in 4% (v/v) formaldehyde in 2 × saline-sodium citrate (SSC) for 10 min at room temperature, washed in 2 × SSC for 2 × 5 min, and dehydrated using ethanol series. The sites of digoxigenin- and biotin-labeled probes were detected using anti-digoxigenin-FITC (Sigma Aldrich/Roche Applied Science) and streptavidin-Cy3 (Thermo Fisher Scientific/Invitrogen, Carlsbad, CA, USA), respectively. Chromosomes were counterstained with DAPI and mounted in Vectashield Antifade Mounting Medium (Vector Laboratories, Burlingame, CA, USA).

Microscopic and image analysis

Slides were examined using Axio Imager Z.2 Zeiss microscope (Zeiss, Oberkochen, Germany) equipped with a Cool Cube 1 camera (Metasystems, Altlußheim, Germany) and appropriate optical filters, and a PC running ISIS software 5.4.7 (Metasystems). The final image adjustment was performed in Adobe Photoshop CS5, and idiograms and final pictures were created in Adobe Photoshop CS5 and GIMP (GNU Image Manipulation Program) v2.10.34. A minimum of ten preparations with mitotic metaphase chromosome spreads and different probe combinations were used for the final karyotype reconstruction of each genotype.

Declarations

Availability of data and materials

The datasets supporting the results of this article are included within the article and its supplementary information files. Databases of Illumina reads are deposited in the Sequence Read Archive (project number: PRJNA1076467). Assembled contigs from different types of repetitive DNA elements are publicly available online (<https://olomouc.ueb.cas.cz/en/content/dna-repeats>). The sequences of newly identified tandem organized repeats were deposited in GenBank (accessions: PP457292–PP457296).

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

Competing interests

The authors declare that they have no competing interests.

Funding

This study was financially supported by an internal grant from Palacký University, Olomouc (IGA PrF-2024-001).

Authors' contributions

E.H. and L.K. conceived the project. K.V. provided plant material from the B481 population. L.K. performed flow cytometry and analyzed the results. J.Č., L.K. and E.H. provided and analyzed sequencing data. J.Č. and V.Z. performed the cytogenetic part of the work. L.K., J.Č. and E.H., wrote the original draft, and V.Z. and K.V. revised and edited the manuscript. All authors have read and approved the final manuscript.

Acknowledgements

We thank Eva Jahnová, Radomíra Tušková, and Petr Navrátil for excellent technical assistance. The computing was supported by the e-INFRA CZ project (ID:90254), supported by the Ministry of Education, Youth and Sports of the Czech Republic, and the ELIXIR-CZ project (ID:90255), part of the international ELIXIR infrastructure.

References

1. Chacón J, Luebert F, Hilger HH, Ovchinnikova S, Selvi F, Cecchi L, et al. The borage family (Boraginaceae s.s.): a revised infrafamilial classification based on new phylogenetic evidence, with emphasis on the placement of some enigmatic genera. *Taxon* 2016;65(3):523-546. doi:10.12705/653.6
2. Merxmüller H, Grau J. Dysploidie bei *Pulmonaria*. *Rev Roum Biol-Botanique*. 1969;14(1):57-63.
3. Sauer W. Karyo-systematische Untersuchungen an der Gattung *Pulmonaria* (Boraginaceae). *Chromosomen-Zahlen, Karyotyp-Analysen und allgemeine Hinweise auf die Entwicklungsgeschichte*. *Bibliot Bot*. 1975;131:1-85.
4. Sauer W. A phylogenetic model for North-Eurasiatic Tubifloras on the base of comparative morphology, cytogenetics and ecology. *Acta Botanica Yunnanica* 1986;8(4):383-395.
5. Májovský J, Murín A. Cytotaxonomy of the genus *Pulmonaria* L. in Slovakia. *Bol Soc Brot Ser. 2*. 1980;53:25-739.
6. Bolliger M. Die Gattung *Pulmonaria* in Westeuropa. Liechtenstein, Vaduz: J. Cramer; 1982.
7. Sauer W. The *Pulmonaria dacica* group: its affinities with central and south-east European allies and with the genus *Paraskeria* (Boraginaceae). *Plant Syst Evol*. 1987;55:257-276. doi:10.1007/BF00936303
8. Ge-ling Z, Riedl H, Kamelin R. Boraginaceae. In: Wu ZY, Raven PH, eds. *Flora of China*, Vol. 16: Gentianaceae through Boraginaceae. China, Beijing: Science Press and USA, St. Louis: Missouri Botanical Garden Press; 1995. p. 329-427.

9. Baasanmunkh S, Oyuntsetseg B, Tsegmed Z, Oyundelger K, Urgamal M, Gantuya B, et al. Distribution of vascular plants in Mongolia –I Part. *Mong J Biol Sci.* 2022;20(1):3-28. doi:10.22353/mjbs.2022.20.01
10. Liu L, Astuti G, Coppi A, Peruzzi L. Different chromosome numbers but slight morphological differentiation and genetic admixture among populations of the *Pulmonaria hirta* complex (Boraginaceae). *Taxon* 2022;71(5):1025-1043. doi:10.1002/tax.12721
11. Weigend M, Selvi F, Thomas DC, Hilger HH. Boraginaceae. In: Kadereit JW, Bittrich V, eds. The families and genera of vascular plants, Vol. 14. Flowering plants. Eudicots. Aquifoliales, Boraginales, Bruniales, Dipsacales, Escalloniales, Garryales, Paracryphiales, Solanales (except Convolvulaceae), Icacinaceae, Metteniusaceae, Vahliaceae. China, Cham: Springer; 2016. p. 41-102.
12. Kobrlová L, Hroneš M. First insights into the evolution of genome size in the borage family—complete data set of Boraginaceae in the Czech Republic. *Bot J Linn Soc.* 2019;189(2):115-131. doi:10.1093/botlinnean/boy079
13. Selvi F, Bigazzi M. Chromosome studies in Turkish species of *Nonea* (Boraginaceae): The role of polyploidy and descending dysploidy in the evolution of the genus. *Edinb J Bot.* 2002; 59(3):405-420.
14. Selvi F, Coppi A, Bigazzi M. Karyotype variation, evolution and phylogeny in *Borago* (Boraginaceae), with emphasis on subgenus *Buglossites* in the Corso-Sardinian system. *Ann Bot.* 2006;98:857-868. doi:10.1093/aob/mcl167.
15. Kolarčík V, Zozomová-Lihová J, Ducár E, Mártonfi P. Evolutionary significance of hybridization in *Onosma* (Boraginaceae): analyses of stabilized hemisexual odd polyploids and recent sterile hybrids. *Biol J Linn Soc.* 2014;112:89-107. doi:10.1111/bij.12270
16. Kirchner D. Molekulare Phylogenie und Biogeographie der Gattung *Pulmonaria* L. (Boraginaceae). Germany, Aachen: Verlag Mainz; 2004.
17. Şuteu D, Băcilă I, Coldea G. Detection of barcode markers able to differentiate the endemic taxon *Pulmonaria filarzskyana* Jáv. from *Pulmonaria rubra* Schott. *Contrib Bot.* 2016;51:7-16.
18. Meeus S, Janssens S, Helsen K, Jacquemyn H. Evolutionary trends in the distylous genus *Pulmonaria* (Boraginaceae): Evidence of ancient hybridization and current interspecific gene flow. *Mol Phylogenet Evol.* 2016;98:63-73. doi:10.1016/j.ympev.2015.11.022
19. Grünig S, Fischer M., Parisod Ch. Recent hybrid speciation at the origin of the narrow endemic *Pulmonaria helvetica*. *Ann Bot.* 2021;127(1):21-31. doi:10.1093/aob/mcaa145
20. Bennett M. *Pulmonarias and the Borage family.* UK, London: BT Batsford; 2003.
21. Bolliger M. Die *Pulmonaria obscura-officinalis*-Gruppe in der Schweiz. *Ber Schweiz Bot Ges.* 1978;88(1/2):30-62.
22. Dersch G. Über *Pulmonaria officinalis* L., *Pulmonaria obscura* Dum. und ihren Bastard in Nordhessen und Südniedersachsen. *Hessische floristische Briefe* 1994;43:49-63.
23. Pellicer J, Hidalgo O, Dodsworth S, Leitch IJ. Genome size diversity and its impact on the evolution of land plants. *Genes* 2018;9:88. doi:soltis10.3390/genes9020088
24. Soltis PS, Soltis DE. Plant genomes: Markers of evolutionary history and drivers of evolutionary change. *Plants, People, Planet* 2021;3(1):74-82. doi:10.1002/ppp3.10159
25. Wendel JF. Genome evolution in polyploids. *Plant Mol Biol.* 2000;42(1):225-249. Doi:
26. Ibarra-Laclette E, Lyons E, Hernández-Guzmán G, Pérez-Torres CA, Carretero-Paulet L., Chang T-H, et al. Architecture and evolution of a minute plant genome. *Nature* 2013;498:94-98. doi:10.1038/nature12132

27. Macas J, Novák P, Pellicer J, Čížková J, Koblížková A, Neumann P, et al. In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe Fabeae. *PLoS One* 2015;10(11):e0143424. doi:10.1371/journal.pone.0143424
28. Leitch IJ Bennet MD. Genome downsizing in polyploid plants. *Biol J Linn Soc.* 2004;82(4):651-663. doi:10.1111/j.1095-8312.2004.00349.x
29. Bennetzen JL, Ma J, Devos KM. Mechanisms of Recent Genome Size Variation in Flowering Plants. *Ann Bot.* 2005;95:127-132. doi:10.1093/aob/mci008
30. Renny-Byfield S, Kovarik A, Kelly LJ, Macas J, Novak P, Chase MW, et al. Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences. *Plant J.* 2013;74(5):829-39. doi:10.1111/tpj.12168.
31. Zenil-Ferguson R, Ponciano JM, Burleigh JG. Evaluating the role of genome downsizing and size thresholds from genome size distributions in angiosperms. *Am J Bot.* 2016;103(7):1175-1186. doi:10.3732/ajb.1500408
32. Simonin KA, Roddy AB. Genome downsizing, physiological novelty, and the global dominance of flowering plants. *PLoS Biol.* 2018;16(1):e2003706. doi:10.1371/journal.pbio.2003706
33. Wang X, Morto JA, Pellicer J, Leitch IJ, Leitch AR. Genome downsizing after polyploidy: mechanisms, rates and selection pressures. *Plant J.* 2021;107:1003-1015. doi:10.1111/tpj.15363
34. Chase MW, Samuel R, Leitch AR, Guignard MS, Conran JG, Nollet F, et al. Down, then up: non-parallel genome size changes and a descending chromosome series in a recent radiation of the Australian allotetraploid plant species, *Nicotiana* section *Suaveolentes* (Solanaceae). *Ann Bot.* 2023;131(1):123-142. doi:10.1093/aob/mcac006
35. Prančl J, Kaplan Z, Trávníček P, Jarolímová V. Genome size as a key to evolutionary complex aquatic plants: polyploidy and hybridization in *Callitriche* (Plantaginaceae). *PLoS ONE.* 2014;9(9):e105997. doi:10.1371/journal.pone.0105997.
36. Prančl J, Koutecký P, Trávníček P, Jarolímová V, Lučanová M, Koutecká E, et al. Cytotype variation, cryptic diversity and hybridization in *Ranunculus* sect. *Batrachium* revealed by flow cytometry and chromosome numbers. *Preslia.* 2018;90(3):195-223. doi:10.23855/preslia.2018.195
37. Popelka O, Sochor M, Duchoslav M. Reciprocal hybridization between diploid *Ficaria calthifolia* and tetraploid *Ficaria verna* subsp. *verna*: evidence from experimental crossing, genome size and molecular markers. *Bot J Linn Soc.* 2019;189(3):293-310. doi: 10.1093/botlinnean/boy085
38. Urfusová R, Mahelka V, Krahulec F, Urfus T. Evidence of widespread hybridization among couch grasses (*Elymus*, Poaceae). *JSE* 2021;59(1):113-124. doi:10.1111/jse.12563
39. Koprivý L, Fráková V, Kolarčík V, Mártonfióvá L, Dudáš M, Mártonfi P. Genome size and endoreplication in two pairs of cytogenetically contrasting species of *Pulmonaria* (Boraginaceae) in Central Europe. *AoB PLANTS* 2022;14(5):plac036. doi:10.1093/aobpla/plac036
40. Jiang JM, Gill BS. Current status and the future of fluorescence *in situ* hybridization (FISH) in plant genome research. *Genome* 2006;49:1057-1068. doi:10.1139/g06-076
41. Jiang J. Fluorescence in situ hybridization in plants: Recent developments and future applications. *Chromosome Res.* 2019;27:153-165. doi:10.1007/s10577-019-09607-z
42. Jang TS, Emadzade K, Parker J, Temsch EM, Leitch AR, Speta F, et al. Chromosomal diversification and karyotype evolution of diploids in the cytologically diverse genus *Prospero* (Hyacinthaceae). *BMC Evol Biol.* 2013;3:136. doi:10.1186/1471-2148-13-136

43. Said M, Hřibová E, Danilova TV, Karafiátová M, Čížková J, Friebe B, et al. The *Agropyron cristatum* karyotype, chromosome structure and cross-genome homoeology as revealed by fluorescence in situ hybridization with tandem repeats and wheat single-gene probes. *Theor Appl Genet.* 2018;131(10):2213-2227. doi:10.1007/s00122-018-3148-9
44. Šimoníková D, Němečková A, Čížková J, Brown A, Swennen R, Doležel J, et al. Chromosome painting in cultivated bananas and their wild relatives (*Musa* spp.) reveals differences in chromosome structure. *Int J Mol Sci.* 2020;21:7915. doi:10.3390/ijms21217915.
45. Waminal NE, Pellerin RJ, Kang SH, Kim HH. Chromosomal mapping of tandem repeats revealed massive chromosomal rearrangements and insights into *Senna tora* dysploidy. *Front Plant Sci.* 2021;12:629898. doi:10.3389/fpls.2021.629898
46. Doležalová A, Sládek L, Šimoníková D, Holušová K, Karafiátová M, Varshney RK, et al. Karyotype differentiation in cultivated chickpea revealed by oligopainting fluorescence in situ hybridization. *Front Plant Sci.* 2022;12:791303. doi:10.3389/fpls.2021.791303
47. Čížková J, Hřibová E, Christelová P, Van den Houwe I, Häkkinen M, Roux N, et al. Molecular and cytogenetic characterization of wild *Musa* species. *PLoS One* 2015;10(8):e0134096. doi:10.1371/journal.pone.0134096
48. Jang TS, McCann J, Parker JS, Takayama K, Hong SP, Schneeweiss GM, et al. rDNA loci evolution in the genus *Glechoma* (Lamiaceae). *PLoS One* 2016;11(11):e0167177. doi:10.1371/journal.pone.0167177
49. Křivánková A, Kopecký D, Stočes Š, Doležel J, Hřibová E. Repetitive DNA: A versatile tool for karyotyping in *Festuca pratensis* Huds. *Cytogenet Genome Res.* 2017;151(2):96-105. doi: 10.1159/000462915
50. Luo J, Liao R, Duan Y, Fu S, Tang Z. Variations of subtelomeric tandem repeats and rDNA on chromosome 1RS arms in the genus *Secale* and 1BL.1RS translocations. *BMC Plant Biol.* 2022;22(1):212. doi:10.1186/s12870-022-03598-6
51. Novák P, Neumann P, Macas J. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinform.* 2010;11:378. doi:10.1186/1471-2105-11-378
52. Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics.* 2013;29(6):792-793. doi:10.1093/bioinformatics/btt054
53. Novák P, Ávila Robledillo L, Koblížková A, Vrbová I, Neumann P, Macas J. TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res.* 2017;45(12):e111. doi:10.1093/nar/gkx257
54. Weiss-Schneeweiss H, Leitch AR, McCann J, Jang TS, Macas J. Employing next generation sequencing to explore the repeat landscape of the plant genome. In: Hörandl E, Appelhans M, eds. *Next Generation Sequencing in Plant Systematics*, Vol. 158. Germany, Oberreifenberg: Koeltz Scientific Books; 2015. p. 155-179.
55. Novák P, Neumann P, Macas J. Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat Protoc.* 2020;15(11):3745-3776. doi:10.1038/s41596-020-0400-y
56. Merxmüller H, Sauer W. *Pulmonaria* L. In: Tutin TG, Heywood VH, Burges NA, Moore DM, Valentine DH, Walters SM, Webb DA, eds. *Flora Europaea*, Vol. 3. UK, Cambridge: Cambridge University Press; 1972. p. 100-102.
57. Hultén E, Fries M. *Atlas of North European vascular plants north of the Tropic of Cancer*. Germany, Königstein: Koeltz Scientific Books; 1986.

58. Sauer W, Gruber G. Beitrag zur Kenntnis der Gattung *Pulmonaria* in Bayern: Kritische Sippen, Verbreitung und allgemeine Hinweise auf die Arealbildung. Bayerische Botanische Gesellschaft zur Erforschung der Heimischen Flora 1979;50:127-160.
59. Hroneš M, Koblíková L. 12. *Pulmonaria* L. In: Kaplan Z, Danihelka J, Chrtek J Jr, Kirschner J, Kubát K, Štech M et al., eds. Key to the Flora of the Czech Republic. Ed.2. CZ, Praha: Academia; 2019. p. 815-816.
60. Vejvodová K, Kúr P, Koutecký P, Diewald W, Grulich V, Máchalová K, et al. Distribution and identification of *Pulmonaria officinalis* and *P. obscura* in the Bohemian Forest and its adjacent foothills. Silva Gabreta 2021;27:97-119.
61. Astuti G, Cristofolini G, Peruzzi L, Pupillo P. A new subspecies of *Pulmonaria officinalis* (Boraginaceae) from the southern Alps. Phytotaxa 2014;186(3):148-157. doi: 10.11646/phytotaxa.186.3.3
62. Cecchi L, Selvi F. Synopsis of Boraginaceae subfam. Boraginoideae tribe Boragineae in Italy. Plant Biosyst. 2015;149(4):630-677. doi:10.1080/11263504.2015.1057261
63. POWO. "Plants of the World Online" Kew (UK): The Royal Botanic Gardens. 2024. <http://www.plantsoftheworldonline.org/>. Accessed 3 March 2024.
64. Scott P. Physiology and Behaviour of Plants. UK, Chichester: John Wiley & Sons Ltd.; 2008.
65. Meeus S, Brys R, Honnay O, Jacquemyn H. Biological Flora of the British Isles: *Pulmonaria officinalis*. J Ecol. 2013;101(5):1353-1368. doi:10.1111/1365-2745.12150
66. Puppi G, Cristofolini G. Sul significato del binomio *Pulmonaria saccharata* Miller. Webbia. 1991;45(2):221-234.
67. Puppi G, Cristofolini G. Systematics of the complex *Pulmonaria saccharata*, *Pulmonaria vallarsae* and related species (Boraginaceae). Webbia. 1996;51(1):1-20.
68. Kapustová V, Tulpová Z, Toegelová H, Novák P, Macas J, Karafiátová M, et al. The dark matter of large cereal genomes: Long tandem repeats. Int J Mol Sci. 2019;20(10):2483. doi:10.3390/ijms20102483
69. Garcia S, Wendel JF, Borowska-Zuchowska N, Ainouche M, Kuderova A, Kovarik A. The utility of graph clustering of 5S ribosomal DNA homoeologs in plant allopolyploids, homoploid hybrids, and cryptic introgressants. Front Plant Sci. 2020;11:41. doi:10.3389/fpls.2020.00041
70. Leitch IJ, Soltis DE, Soltis PS, Bennett MD. Evolution of DNA amounts across land plants (Embryophyta). Ann Bot. 2005;95(1):207-17. doi:10.1093/aob/mci014
71. Hawkins JS, Proulx SR, Rapp RA, Wendel JF. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. Proc Natl Acad Sci USA 2009;106(42):17811-6. doi:10.1073/pnas.0904339106
72. Bigazzi M, Selvi F. Karyotype morphology and cytogeography in *Brunnera* and *Cynoglottis* (Boraginaceae). Bot J Linn Soc 2001;136(4):365-378. doi:10.1111/j.1095-8339.2001.tb00580.x
73. Bigazzi M, Selvi F. Chromosome variation in Anatolian species of *Nonea* Medik. (Boraginaceae), with special reference to endemics and *N. persica*. Caryologia 2003;56(4):509-519. doi:10.1080/00087114.2003.10589364
74. Koblíková L, Hroneš M, Koutecký P, Štech M, Trávníček B. *Symphytum tuberosum* complex in central Europe: cytogeography, morphology, ecology and taxonomy. Preslia 2016;88:77-112.
75. Koblíková L, Duchoslav M, Hroneš M. Morphological, ecological and geographic differences between diploids and tetraploids of *Symphytum officinale* (Boraginaceae) justify both cytotypes as separate species. AoB Plants 2022;144:plac028. doi:10.1093/aobpla/plac028

76. Kolarčík V, Kocová V, Čákovice D, Kačmárová T, Piovár J, Mártonfi P. Nuclear genome size variation in the allopolyploid *Onosma arenaria* – *O. pseudoarenaria* species group: methodological issues and revised data. *Botany* 2018;96:397-410. doi:10.1139/cjb-2017-0164
77. Leitch IJ, Chase MW, Bennett MD. Phylogenetic analysis of DNA C-values provides evidence for a small ancestral genome size in flowering plants. *Ann Bot.* 1998;82:85-94.
78. Šmarda P, Knápek O, Březinová A, Horová L, Grulich V, Danihelka J, et al. Genome sizes and genomic guanine+cytosine (GC) contents of the Czech vascular flora with new estimates for 1700 species. *Preslia* 2019;91(2):117-142. doi:10.23855/preslia.2019.117
79. Piednoel M, Aberer AJ, Schneeweiss GM, Macas J, Novak P, Gundlach H, et al. Next-generation sequencing reveals the impact of repetitive DNA across phylogenetically closely related genomes of Orobanchaceae. *Mol Biol Evol.* 2012;29(11):3601-3611 doi:10.1093/molbev/mss168
80. Novák P, Hřibová E, Neumann P, Koblízková A, Doležel J, Macas J. Genome-wide analysis of repeat diversity across the family *Musaceae*. *PLoS One* 2014;9(6):e98918. doi:10.1371/journal.pone.0098918.
81. McCann J, Macas J, Novák P, Stuessy TF, Villasenor JL, Weiss-Schneeweiss H. Differential genome size and repetitive DNA evolution in diploid species of *Melampodium* sect. *Melampodium* (Asteraceae). *Front. Plant Sci.* 2020;11:362. doi:10.3389/fpls.2020.00362
82. Zwyrteková J, Němečková A, Čížková J, Holušová K, Kapustová V, Svačina R, et al. Comparative analyses of DNA repeats and identification of a novel Fesreba centromeric element in fescues and ryegrasses. *BMC Plant Biol.* 2020;20(1):280. doi:10.1186/s12870-020-02495-0.
83. Ferraz ME, Ribeiro T, Sader M, Nascimento T, Pedrosa-Harand A. Comparative analysis of repetitive DNA in dysploid and non-dysploid *Phaseolus* beans. *Chromosome Res.* 2023;31(4):30. doi:10.1007/s10577-023-09739-3
84. International Rice Genome Sequencing Project, Sasaki T. The map-based sequence of the rice genome. *Nature* 2005;436(7052):793-800. doi:10.1038/nature03895
85. Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, et al. The B73 maize genome: complexity, diversity, and dynamics. *Science.* 2009;326(5956):1112-1115. doi:10.1126/science.1178534.
86. Papon N, Lasserre-Zuber P, Rimbart H, De Oliveira R, Paux E, Choulet F. All families of transposable elements were active in the recent wheat genome evolution and polyploidy had no impact on their activity. *Plant Genome* 2023; 16(3):e20347. doi:10.1002/tpg2.20347.
87. Kreplak J, Madoui MA, Cápál P, Novák P, Labadie K, Aubert G, et al. A reference genome for pea provides insight into legume genome evolution. *Nat Genet.* 2019;51:1411-1422. doi:10.1038/s41588-019-0480-1
88. Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee AH, et al. Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat Genet.* 2014;46: 270-278. doi:10.1038/ng.2877
89. Sierro N, Battey J, Ouadi S, Bakaher N, Bovet L, Willig A, et al. The tobacco genome sequence and its comparison with those of tomato and potato. *Nat Commun.* 2014;5:3833. doi:10.1038/ncomms4833
90. Dodsworth S, Jang TS, Struebig M, Chase MW, Weiss-Schneeweiss H, Leitch AR. Genome-wide repeat dynamics reflect phylogenetic distance in closely related allotetraploid *Nicotiana* (Solanaceae). *Plant Syst Evol.* 2017;303:1013-1020. doi:10.1007/s00606-016-1356-9
91. Gaiero P, Vaio M, Peters SA, Schranz ME, de Jong H, Speranza PR. Comparative analysis of repetitive sequences among species from the potato and the tomato clades. *Ann Bot.* 2019;123(3):521-532. doi:10.1093/aob/mcy186.

92. Song B, Song Y, Fu Y, Kizito EB, Kamenya SN, Kabod PN, et al. Draft genome sequence of *Solanum aethiopicum* provides insights into disease resistance, drought tolerance, and the evolution of the genome. *Gigascience* 2019;8(10):giz115. doi:10.1093/gigascience/giz115
93. Li N, He Q, Wang J, Wang B, Zhao J, Huang S, et al. Super-pangenome analyses highlight genomic diversity and structural variation across wild and cultivated tomato species. *Nat Genet.* 2023;55:852-860. doi:10.1038/s41588-023-01340-y
94. Song Z, Lin C, Xing P, Fen Y, Jin H, Zhou C, et al. A high-quality reference genome sequence of *Salvia miltiorrhiza* provides insights into tanshinone synthesis in its red rhizomes. *Plant Genome.* 2020;13(3):e20041. doi:10.1002/tpg2.20041.
95. Jia KH, Liu H, Zhang RG, Xu J, Zhou S-S, Jiao S-Q, et al. Chromosome-scale assembly and evolution of the tetraploid *Salvia splendens* (Lamiaceae) genome. *Hortic Res.* 2021;8:177. doi:10.1038/s41438-021-00614-y
96. Muravenko OV, Yurkevich OY, Kalnyuk JV, Samatadze TE, Zoshchuk SA, Amosova AV. Integration of Repeatomic and Cytogenetic Data on Satellite DNA for the Genome Analysis in the Genus *Salvia* (Lamiaceae). *Plants* 2022; 11(17):2244. doi:10.3390/plants11172244.
97. Jiang X, Jiang C, Yuan W, Zhang M, Fang Z, Li Y, et al. A universal karyotypic system for hexaploid and diploid *Avena* species brings oat cytogenetics into the genomics era. *BMC Plant Biol.* 2021;21:213. doi:10.1186/s12870-021-02999-3
98. Lysák MA, Berr A, Pečinka A, Schmidt R, McBreen K, Schubert I. Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc Natl Acad Sci USA* 2006;103: 5224-5229. doi:10.1073/pnas.0510791103
99. Mandáková T, Lysák MA. Post-polyploid diploidization and diversification through dysploid changes. *Curr Opin Plant Biol.* 2018;42:55-65. doi:10.1016/j.pbi.2018.03.001.
100. Mandáková T, Li Z, Barker MS, Lysák MA. Diverse genome organization following 13 independent mesopolyploid events in Brassicaceae contrasts with convergent patterns of gene retention. *Plant J.* 2017;91(1): 3-21. doi: 10.1111/tpj.13553
101. Meusel H, Jäger EJ, Rauschert S, Weinert E. *Vergleichende Chorologie der zentraleuropäischen Flora. Band II.* Germany, Jena: Gustav Fischer; 1978.
102. Vozárová R, Herklotz V, Kovařík A, Tynkevich YO, Volkov RA, Ritz CM, et al. Ancient origin of two 5S rDNA families dominating in the genus *Rosa* and their behavior in the Canina-type meiosis. *Front Plant Sci.* 2021;12:643548. doi:10.3389/fpls.2021.643548
103. Gams H. *Pulmonaria* L. In: Hegi G, ed. *Illustrierte Flora von Mitteleuropa, Vol V/3.* Germany, München: J.F. Lehmanns Verlag; 1927. p. 2209-2221.
104. Meeus S, Honnay O, Brys R, Jacquemyn H. Biased morph ratios and skewed mating success contribute to loss of genetic diversity in the distylous *Pulmonaria officinalis*. *Ann Bot.* 2012;109:227-235. doi:10.1093/aob/mcr272
105. Greilhuber J, Doležel J, Lysák M, Bennett MD: The origin, evolution and proposed stabilization of the terms 'Genome Size' and 'C-Value' to describe nuclear DNA contents. *Ann Bot.* 2005;95(1):255-266. doi:10.1093/aob/mci019
106. Šmarda P, Bureš P, Horová L, Foggi B, Rossi G. Genome size and GC content evolution of *Festuca*: ancestral expansion and subsequent reduction. *Ann Bot.* 2008;101(3):421-433. doi:10.1093/aob/mcm307

107. Doležel J, Greilhuber J, Lucretti S, Meister A, Lysák MA, Nardi L, et al. Plant genome size estimation by flow cytometry: inter-laboratory comparison. *Ann Bot.* 1998;82(Supplement A):17-26. doi: 10.1093/oxfordjournals.aob.a010312
108. Barow M, Meister A. Lack of correlation between AT frequency and genome size in higher plants and the effect of nonrandomness of base sequences on dye binding. *Cytometry A* 2002;47(1):1-7. doi:10.1002/cyto.10030
109. Doležel J, Bartoš J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size of trout and human. *Cytometry A.* 2003;51(2):127-8; author reply 129. doi:10.1002/cyto.a.10013.
110. Hintze J. NCSS 9.0.15. 2013. Available at: <https://www.ncss.com/>
111. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34(17):i884-i890. doi:10.1093/bioinformatics/bty560.
112. Neumann P, Novák P, Hošťáková N, Macas J. Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mob DNA* 2019;10:1. doi:10.1186/s13100-018-0144-1
113. Sonnhammer EL, Durbin R. A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 1995;167(1-2):GC1-GC10. doi:10.1016/0378-1119(95)00714-8
114. Kato A, Lamb JC, Birchler JA. Chromosome painting using repetitive DNA sequences as probes for somatic chromosome identification in maize. *Proc Natl Acad Sci USA* 2004;101(37):13554-13559. doi:10.1073/pnas.0403659101
115. Kato A, Albert PS, Vega JM, Birchler JA. Sensitive fluorescence in situ hybridization signal detection in maize using directly labeled probes produced by high concentration DNA polymerase nick translation. *Biotech Histochem.* 2006;81(2-3):71-78. doi:10.1080/10520290600643677.
116. Untergasser A, Cutcutache I, Koressaar T, Ye J, Faircloth BC, Remm M, et al. Primer3--new capabilities and interfaces. *Nucleic Acids Res.* 2012;40(15):e115. doi:10.1093/nar/gks596.
117. Gerlach WL, Bedbrook JR. Cloning and characterization of ribosomal-RNA genes from wheat and barley. *Nucleic Acids Res.* 1979;7(7):1869-1885. doi:10.1093/nar/7.7.1869

Figures

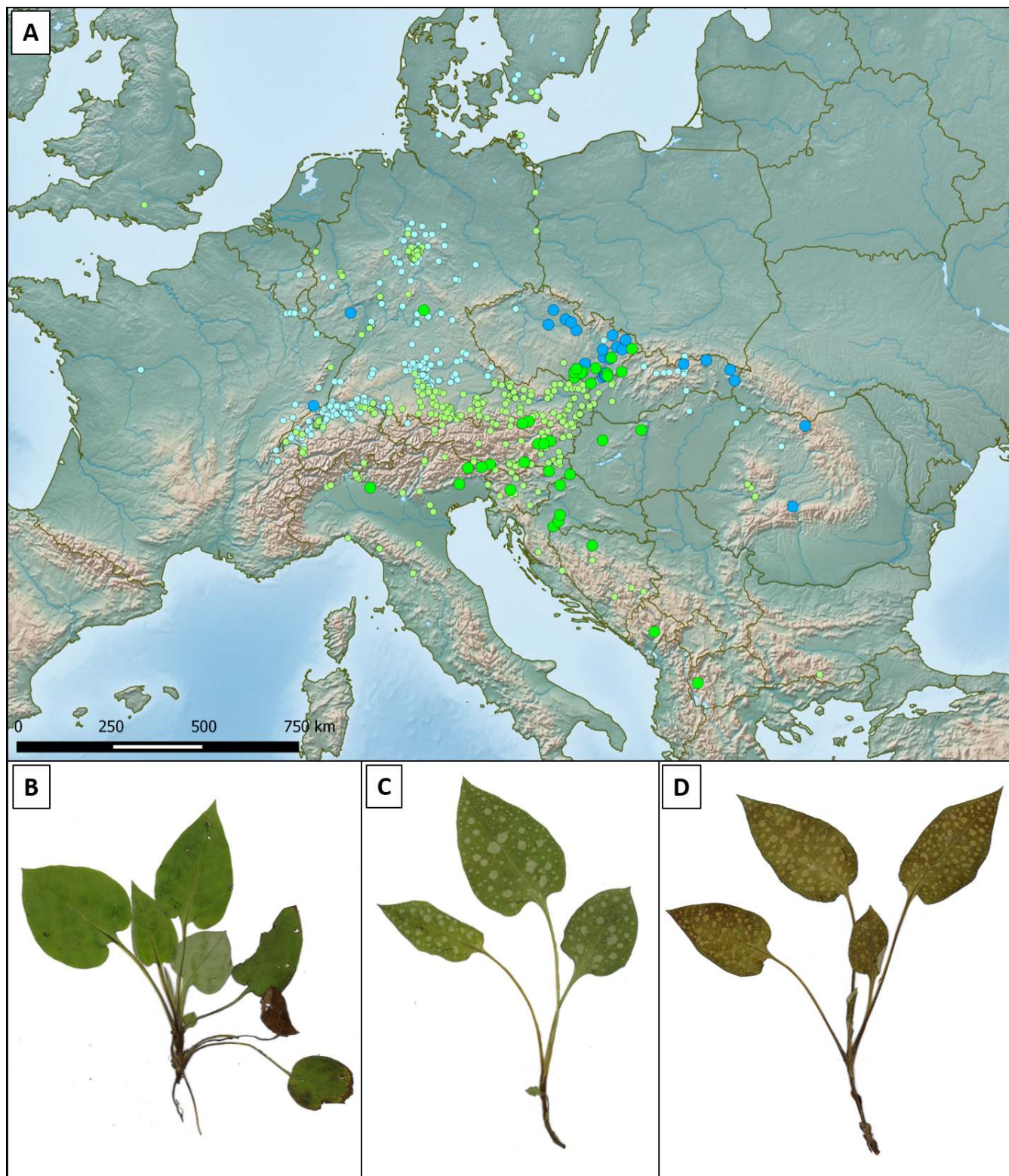


Figure 1

Distribution map of the *Pulmonaria officinalis* group. (A) Distribution map of the *P. officinalis* group sampled and analyzed in this study (large dots; see Supplementary Table 1), including published chromosome reports (744 records in total, small dots; see Supplementary Table 2): *P. obscurain* green and *P. officinalis* s.str. in green. Illustrative images of (B) *P. obscura* (B473.1), (C) *P. officinalis* (B470.1) and (D) *P. saccharata*-like accession (B15.1).

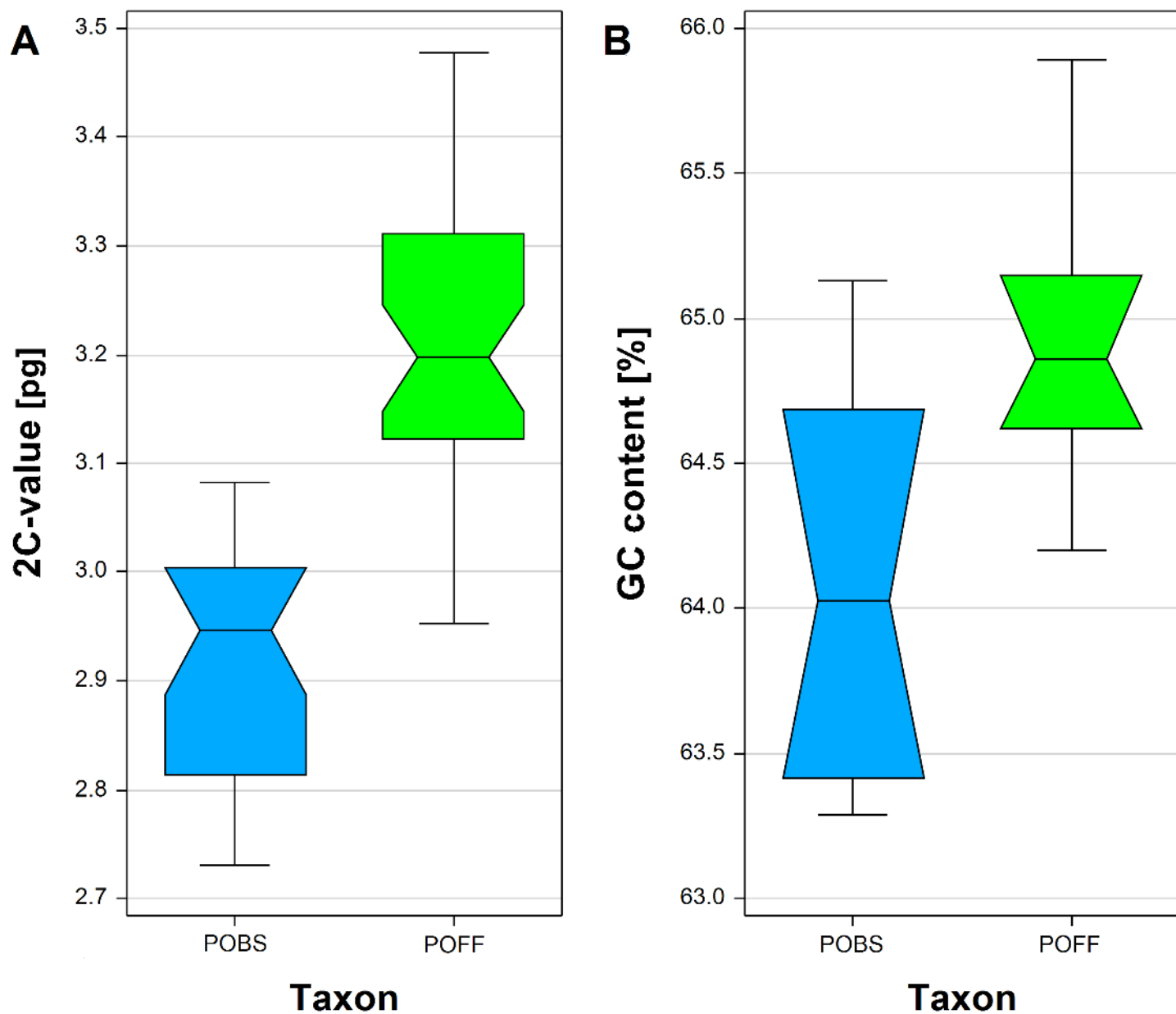


Figure 2

Genome size and GC content variation in *Pulmonaria obscura* and *P. officinalis*. (A) Variation in absolute genome size (2C-value); and (B) genomic GC content detected in the *Pulmonaria officinalis* group: *P. obscura*(POBS), *P. officinalis* (POFF). Rectangles define the 25th and 75th percentiles, horizontal lines show median values, whiskers are 10–90 percentiles.

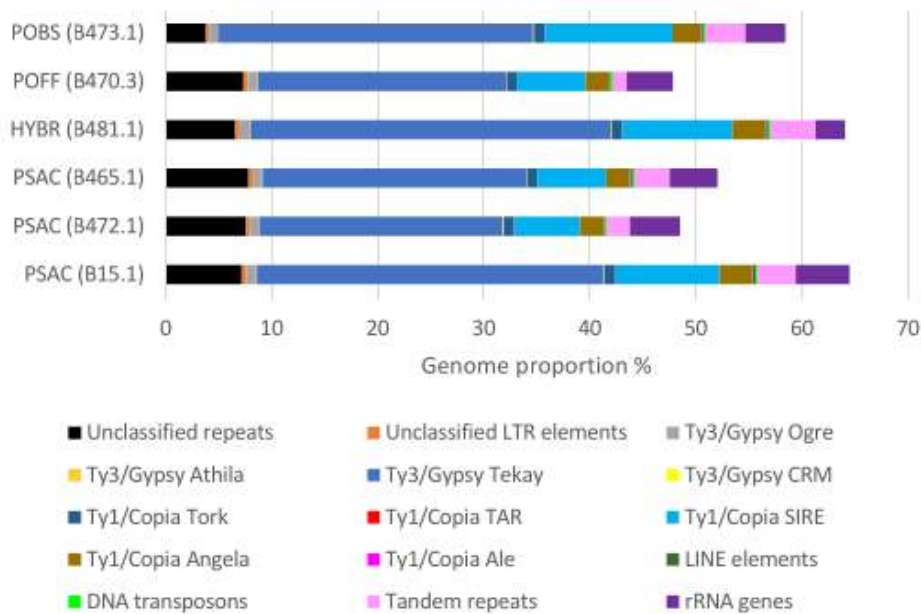


Figure 3

Repeatome composition in analyzed *Pulmonaria* accessions. Genome proportion of individual repeat type was obtained as the ratio of reads specific to the individual repeat type to all reads used for the clustering analysis by the RepeatExplorer2 pipeline. *P. obscura* (POBS; B473.1); *P. officinalis* (POFF; B470.3); an interspecific natural hybrid *P. obscura* × *P. officinalis* (HYBR; B481.1); *P. saccharata*-like accessions (PSAC1–PSAC3; B465.1; B472.1; B15.1).

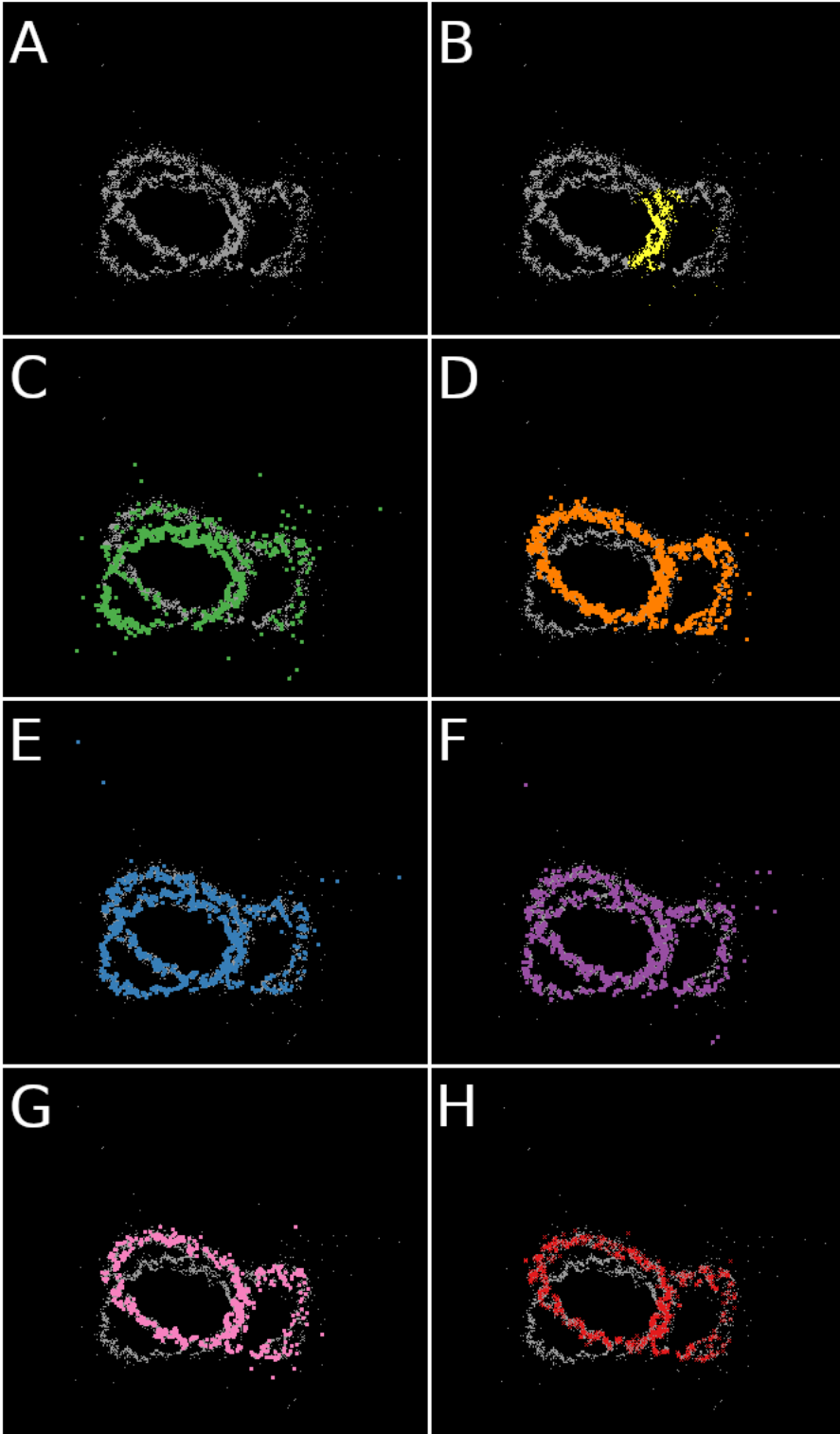


Figure 4

Graph structure of 5S rDNA sequence reads from the comparative analysis of RepeatExplorer2. (A) Graph structure obtained from all sequence reads homologous to the 5S rDNA. (B) The position of the 5S genic region on the graph topology is highlighted in yellow. (C–H) Cluster graph with annotated read origin: *P. obscura* in green (C); *P. officinalis* in orange (D); reads specific for *P. obscura* × *P. officinalis* (B481.1) in blue (E); reads specific for *P. saccharata*-like accession B15.1 are highlighted in purple (F), B465.1 in pink (G), and B472.1 in red (H).

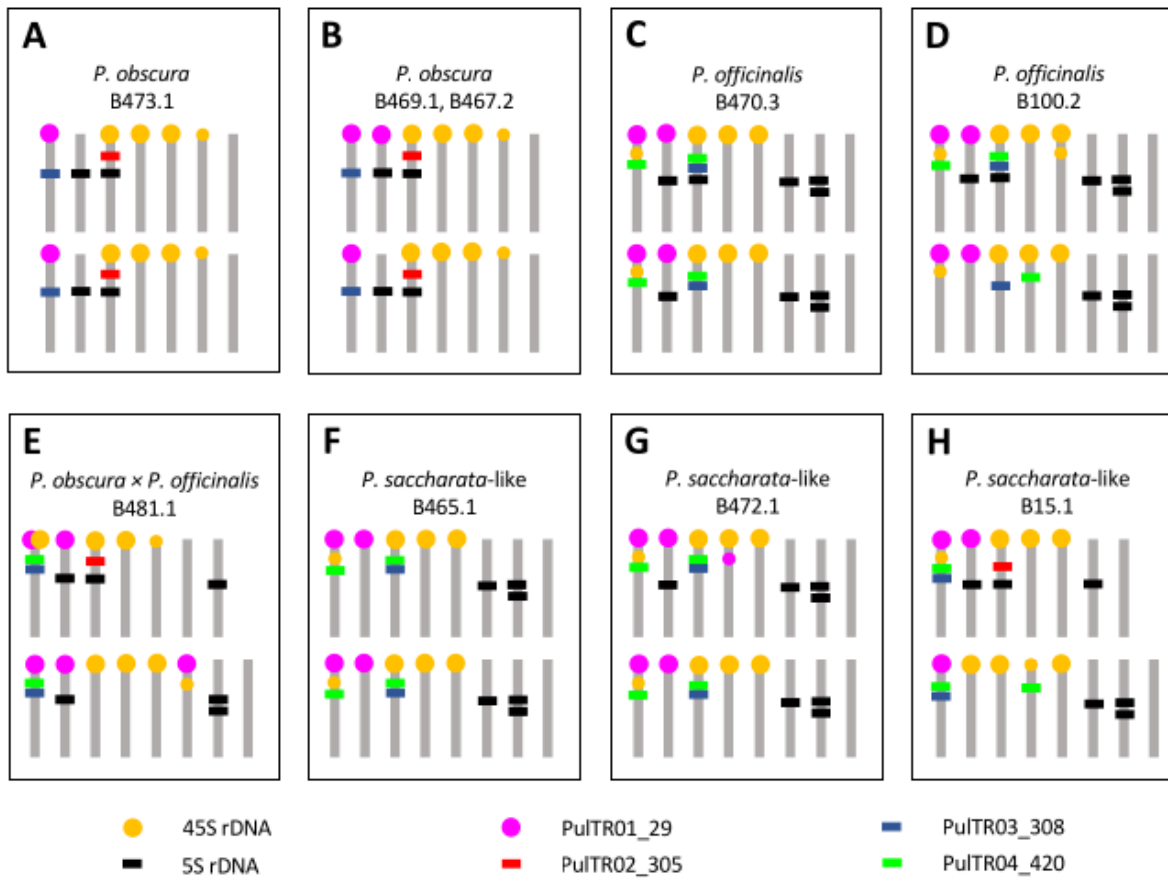


Figure 5

Idiograms of analyzed *Pulmonaria* accessions.

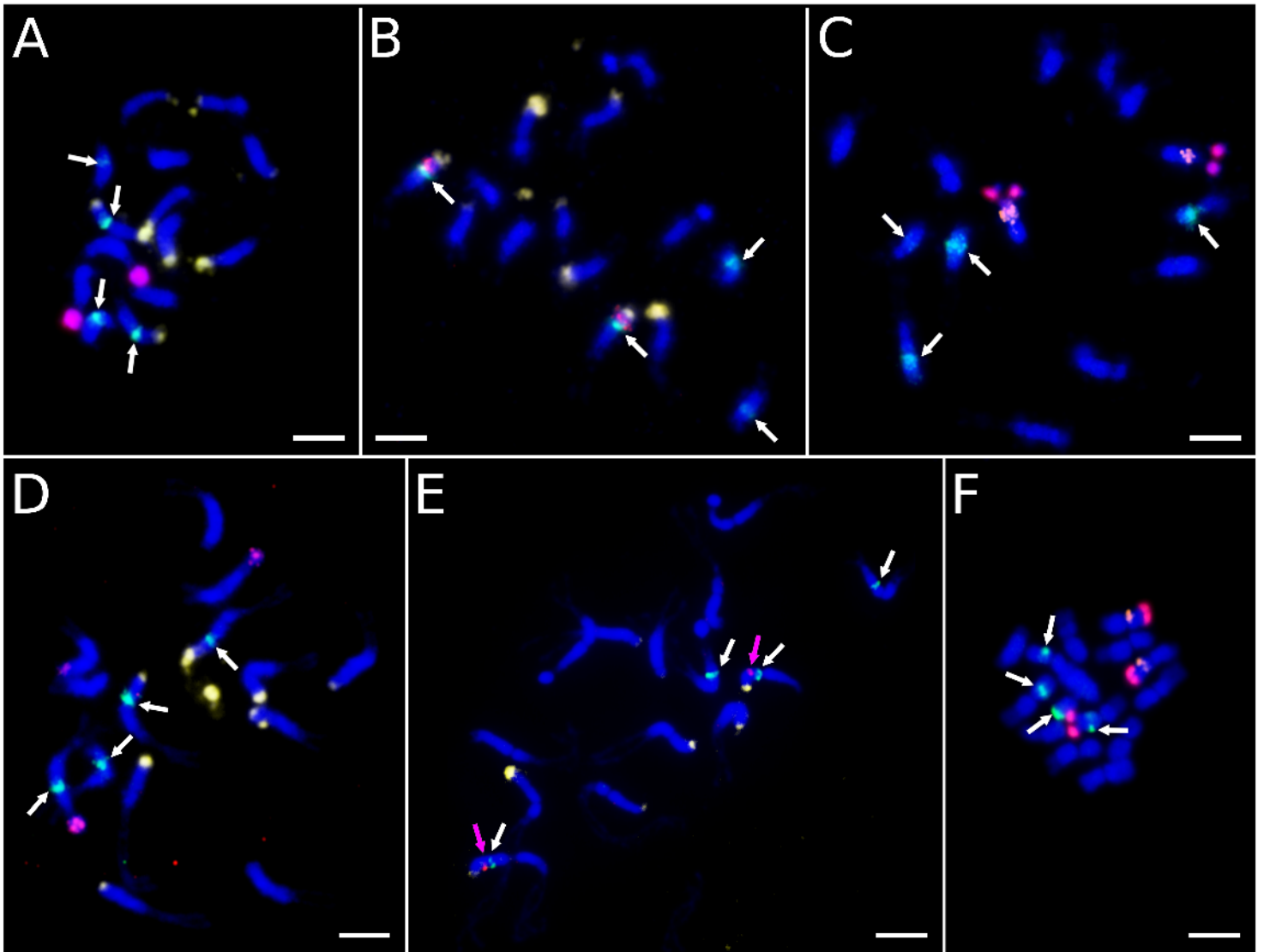


Figure 6

Chromosomal localization of newly identified satDNAs and rDNA sequences in *P. obscura* ($2n = 14$). (A, B, C) *P. obscura* (B473.1) with probes for: (A) 45S rDNA (yellow), PulTR01_29 (red) and 5S rDNA (green); (B) 45S rDNA (yellow), PulTR02_305 (red) and 5S rDNA (green); (C) PulTR01_29 (red), PulTR03_308 (orange), and 5S rDNA (green). (D, E, F) *P. obscura* (B467.2) with probes for: (D) 45S rDNA (yellow), PulTR01_29 (red) and 5S rDNA (green); (E) 45S rDNA (yellow), PulTR02_305 (red) and 5S rDNA (green); and (F) PulTR01_29 (red), PulTR03_308 (orange), and 5S rDNA (green). The white arrows indicate 5S rDNA loci and red arrows in (E) indicate loci of PulTR02_305. Chromosomes were counterstained with DAPI (blue). Bars = 5 μ m.

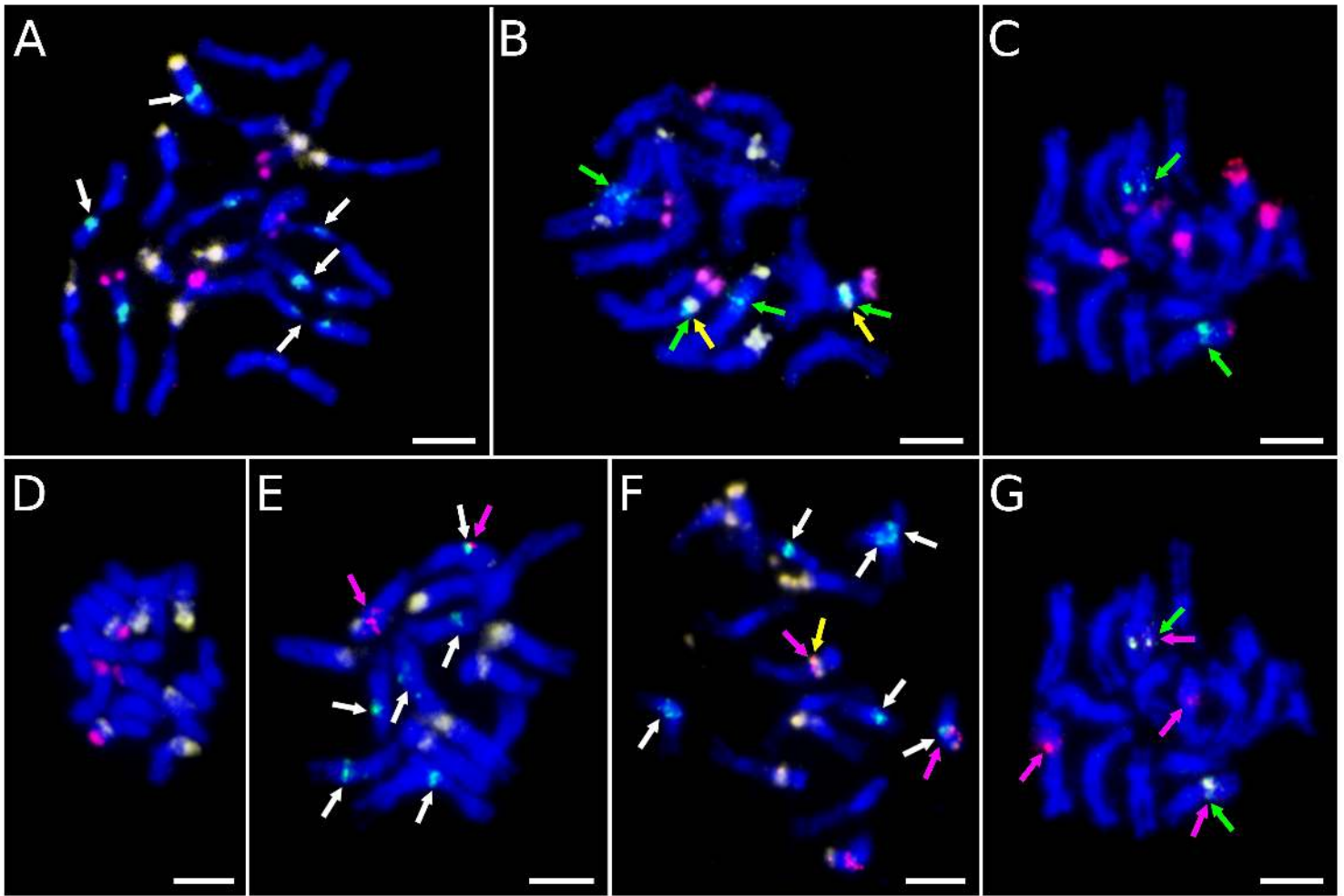


Figure 7

Chromosomal localization of newly identified satDNAs and rDNA sequences in *P. officinalis* ($2n = 16$). (A, B, C, G) *P. officinalis* (B470.3) with probes for: (A) 45S rDNA (yellow), PulTR01_29 (red) and 5S rDNA (green); (B) 45S rDNA (yellow), PulTR01_29 (red) and PulTR04_420 (green): green arrows indicate the loci of PulTR04_420 and; (C, G) the same plate with the signals for (C) 45S rDNA (red) and PulTR03_308 (green), and (G) co-localization of PulTR03_308 (green) and PulTR04_420 (red): green arrows indicate PulTR03_308, and red arrows indicate PulTR04_420. (D, E, F) *P. officinalis* (B100.2) with probes for: (D) 45S rDNA (yellow) and PulTR01_29 (red); (E) 45S rDNA (yellow), PulTR03_308 (red) and 5S rDNA (green): red arrows indicate PulTR03_308; and (F) 45S rDNA (yellow), PulTR04_420 (red), and 5S rDNA (green): red arrows indicate PulTR04_420. The white arrows indicate the 5S rDNA loci and yellow arrows point at minor interstitial loci of 45S rDNA. Chromosomes were counterstained with DAPI (blue). Bars = 5 μ m.

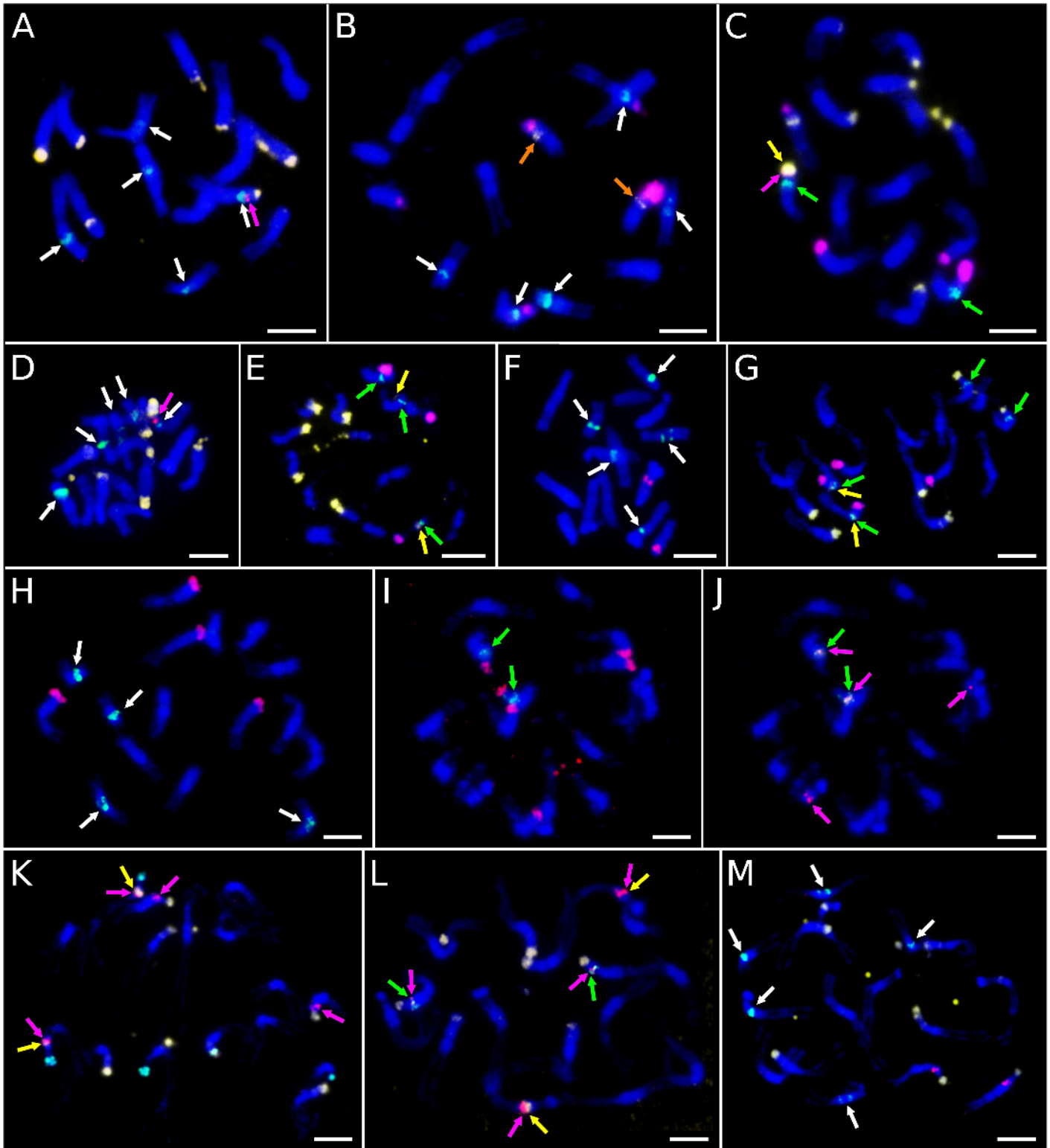


Figure 8

Chromosomal localization of new satDNAs and rDNA sequences in natural hybrid and *P. saccharata*-like accessions. (A, B, C) Putative natural hybrid between *P. obscura* and *P. officinalis* (B481.1; $2n = 15$) with probes for: (A) 45S rDNA (yellow), PulTR02_305 (red) and 5S rDNA (green): red arrow points at the locus of PulTR02_305; (B) PulTR01_29 (red), PulTR03_308 (orange) and 5S rDNA (green): orange arrows indicate loci of PulTR03_308; (C) 45S rDNA (yellow), PulTR01_29 (red) and PulTR04_420 (green): green arrows indicate PulTR04_420; red arrow points to PulTR01_29 and yellow arrow points at 45S rDNA locus indicating co-localization of these two probes.

(D, E, F,) *P. saccharata*-like accession B15.1 ($2n = 15$) with probes for: (D) 45S rDNA (yellow), PulTR02_305 (red) and 5S rDNA (green): red arrow points at a signal of PulTR02_305; (E) 45S rDNA (yellow), PulTR01_29 (red) and PulTR04_420 (green): green arrows point at PulTR04_420; (F) PulTR03_308 (red) and 5S rDNA (green). (G, H, I, J) *P. saccharata*-like accession B465.1 ($2n = 16$) with probes for: (G) 45S rDNA (yellow), PulTR01_29 (red) and PulTR04_420 (green): green arrows indicate PulTR04_420; (H) PulTR01_29 (red) and 5S rDNA (green); (I) 45S rDNA (red) and PulTR03_308 (green): green arrows point at PulTR03_308; and (J) PulTR03_308 (green) and PulTR04_420 (red): green arrows point at PulTR03_308 and red arrows indicate PulTR04_420. (K, L, M) *P. saccharata*-like accession B472.1 ($2n = 16$) with probes for: (K) 45S rDNA (yellow), PulTR01_29 (green) and PulTR04_420 (red): red arrows point at PulTR04_420; (L) 45S rDNA (yellow), PulTR03_308 (green) and PulTR04_420 (red): green arrows point at PulTR03_308, and red arrows indicate PulTR04_420; (M) 45S rDNA (yellow), PulTR03_308 (red) and 5S rDNA (green). The white arrows indicate the 5S rDNA loci and yellow arrows point at minor interstitial loci of 45S rDNA. Chromosomes were counterstained with DAPI (blue). Bars = 5 μ m.

Supplementary Files

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

- [SuppFigure1.tiff](#)
- [SupplFigure2.tiff](#)
- [SupplFigure3.tiff](#)
- [SupplFigure4.tiff](#)
- [SupplFigure5.tiff](#)
- [SupplTableS1S4.xlsx](#)