ORIGINAL ARTICLE

Repeat turnover meets stable chromosomes: repetitive DNA sequences mark speciation and gene pool boundaries in sugar beet and wild beets

Nicola Schmidt[1], Katharina Sielemann[2,3], Sarah Breitenbach[1], Jörg Fuchs[4], Boas Pucker[5], Bernd Weisshaar[2], Daniela Holtgräwe*[2], and Tony Heitkam*[1]

[1]Faculty of Biology, Technische Universität Dresden, 01069 Dresden, Germany
[2]Genetics and Genomics of Plants, Center for Biotechnology (CeBiTec) & Faculty of Biology, Bielefeld University, 33615 Bielefeld, Germany
[3]Graduate School DILS, Bielefeld Institute for Bioinformatics Infrastructure (BIBI), Bielefeld University, 33615 Bielefeld, Germany
[4]Leibniz Institute of Plant Genetics and Crop Plant Research (IPK) Gatersleben, 06466 Stadt Seeland, Germany
[5]Plant Biotechnology and Bioinformatics, Institute of Plant Biology & Braunschweig Integrated Centre of Systems Biology (BRICS), TU Braunschweig, 38106 Braunschweig, Germany

*Manuscript corresponding authors:

Tony Heitkam
Faculty of Biology
Technische Universität Dresden
01069 Dresden
Germany
Phone: (+49) 351 463 39593
E-Mail: tony.heitkam@tu-dresden.de

Daniela Holtgräwe
CeBiTec & Faculty of Biology
Bielefeld University
33615 Bielefeld
Germany
(+49) 521 106 8724
dholtgra@cebitec.uni-bielefeld.de

Running title: Repeats in Beets

ABSTRACT

• *Background*

Sugar beet (*Beta vulgaris* subsp. *vulgaris*) and its crop wild relatives share a base chromosome number of nine and similar chromosome morphologies. Yet, interspecific breeding is impeded by chromosome and sequence divergence that is still not fully understood. Since repetitive DNA sequences represent the fastest evolving parts of the genome, they likely impact genomic variability and contribute to the separation of beet gene pools. Hence, we investigated if innovations and losses in the repeatome can be linked to chromosomal differentiation and speciation.

• *Results*

We traced genome- and chromosome-wide evolution across sugar beet and twelve wild beets comprising all sections of the beet genera *Beta* and *Patellifolia*. For this, we combined data from short and long read sequencing, flow cytometry, and cytogenetics to build a comprehensive data framework for our beet panel that spans the complete scale from DNA sequence to chromosome up to the genome.

Genome sizes and repeat profiles reflect the separation of the beet species into three gene pools. These gene pools harbor repeats with contrasting evolutionary patterns: We identified section- and species-specific repeat emergences and losses, e.g. of the retrotransposons causal for genome expansions in the section *Corollinae/Nanae*. Since most genomic variability was found in the satellite DNAs, we focused on tracing the 19 beetSat families across the three beet sections/genera. These taxa harbor evidence for contrasting strategies in repeat evolution, leading to contrasting satellite DNA profiles and fundamentally different centromere architectures, ranging from chromosomal uniformity in *Beta* and *Patellifolia* species to the formation of patchwork chromosomes in *Corollinae/Nanae* species.

• *Conclusions*

We show that repetitive DNA sequences are causal for genome size expansion and contraction across the beet genera, providing insights into the genomic underpinnings of beet speciation. Satellite DNAs in particular vary considerably among beet taxa, leading to the evolution of distinct chromosomal setups. These differences likely contribute to the barriers in beet breeding between the three gene pools. Thus, with their

isokaryotypic chromosome sets, beet genomes present an ideal system for studying the link between repeats, genome variability, and chromosomal differentiation/evolution and provide a theoretical basis for understanding barriers in crop breeding.

KEY WORDS

genome divergence; chromosomes; speciation; repetitive DNA; transposable elements; satellite DNAs; crop wild relatives; sugar beet; *Beta*; *Patellifolia*

Background

Interplay between chromosomal stability and genome evolution

Large evolutionary leaps in terms of speciation and genome divergence are occasionally accompanied by chromosome number changes caused by polyploidisation events and/or structural rearrangements (*Arabidopsis*: Yogeeswaran *et al.*, 2005; *Carex*: Escudero *et al.*, 2012 & 2015; *Cylicomorpha*: Rockinger *et al.*, 2016). In contrast, some genera display genomic divergence despite harboring karyotypes with stable base chromosome numbers across all known species (McCann *et al.*, 2020; Vitales *et al.*, 2020a; Pellicer *et al.*, 2021). That even extends to large intraspecific differences in genome size despite a consistent chromosome number (*Euphrasia*: Becher *et al.*, 2021). Therefore, the question arises how inter- and intraspecific genomic divergence accumulates while the chromosomal setup is maintained.

One of the links between the genome and its partitioning into chromosomes are repetitive DNA sequences, which provide structure to eukaryotic karyotypes. Despite their fast evolution, repetitive DNA sequences have a conserved function by providing sequence material to the main structural regions, such as centromeres and telomeres. Nevertheless, how the evolution of the repeatome itself relates to the global chromosome and genome evolution is still a matter of debate (e.g. centromere paradox; Henikoff *et al.*, 2001; Presting, 2018).

Species of the Beta *and* Patellifolia *genera are marked by dynamic genomes, but stable chromosome numbers*

We wondered how genomic diversity despite stability of chromosomal number is reflected in the repeat composition. To test this, we used species of the beet genera *Beta* and *Patellifolia* as examples – plant taxa that have been used to study repetitive DNA evolution for over thirty years (Schmidt *et al.*, 1990; and later publications from the lab of late Thomas Schmidt). The members of these two genera belong to chromosomally stable taxa, all marked by a base chromosome number of x=9 (despite some higher ploidies) and containing roughly equally-sized, metacentric chromosomes.

The genera *Beta* and *Patellifolia* comprise at least eleven species that are separated by up to 38.4 million years of evolution (Hohmann *et al.*, 2006). Species from the genus *Beta* are divided into two or three different sections (*Beta*, *Corollinae/Nanae*) depending on whether the endemic species *B. nana* is considered a separate section

(Kadereit *et al.*, 2006; Frese and Ford-Lloyd, 2020; Sielemann *et al.*, 2022). Cultivated beets such as sugar beet are varieties of *B. vulgaris* subsp. *vulgaris* within the section *Beta*. The presumed progenitor of all cultivated beets, *B. vulgaris* subsp. *maritima*, also belongs to this section (Frese and Ford-Lloyd, 2020). For better readability, *B. vulgaris* subsp. *vulgaris*, *B. vulgaris* subsp. *adanensis*, and *B. vulgaris* subsp. *maritima* will be hereafter referred to as *B. vulgaris*, *B. adanensis*, and *B. maritima*, respectively.

Sugar beet is a relatively young crop that went through an exceptionally narrow bottleneck during its 200 years of domestication (Fischer, 1989). This has led to very low genetic diversity and a loss of several valuable traits such as pathogen resistances and tolerance to adverse environmental conditions (e.g. drought, salinated soil; Panella *et al.*, 2020). Therefore, it is necessary to harvest genetic variation in the beet germplasm to improve cultivated beet varieties. However, crossing experiments of sugar beet with many of its crop wild relatives (CWRs) resulted in reproduction-defective offspring such as sterile and semi-fertile, as well as aneuploid and anorthoploid plants indicating that postzygotic isolation mechanisms, e.g. the lack of chromosome homology due to larger genomic differences, are causal for the limited gain of genetically improved seeds rather than prezygotic isolation mechanisms (Frese and Ford-Lloyd, 2020).

Based on the crossability with *B. vulgaris*, the wild beets are grouped into three different gene pools. These correspond to the three beet main taxa with section *Beta* species representing the primary gene pool, *Corollinae/Nanae* species representing the secondary gene pool, and *Patellifolia* species representing the tertiary gene pool. Given the fact that di- and polyploid species are present in all of these gene pools (Frese and Ford-Lloyd, 2020) and due to the chromosome uniformity across all beet clades, polyploidisation and restructuring of chromosomes seem to play a rather subordinate role in the emergence of genomic variety in *Beta* and *Patellifolia* species. Instead, differences at the DNA sequence level (Sielemann *et al.*, 2023a) may be the cause of flawed chromosome pairing, thus resulting in the observed crossing barriers.

The rapid evolution of repetitive elements can be a major factor in reduced pairing between homologous and homoeologous chromosomes (Dvorak, 1983) as it leads to varying repeat compositions in the genomes of even closely related species. Repetitive DNA sequences include transposable elements (TEs) and tandem repeats (TRs) that

differ in their origin, amplification mode, and sequence characteristics (Bennetzen, 2005; Bennetzen and Wang, 2014).

There are two classes of TEs: class I or retrotransposons with a 'copy-and-paste' transposition mechanism and an RNA intermediate, and class II or DNA transposons with a 'cut-and-paste' mechanism, thus DNA being the intermediate in replication. Class I is divided into two subclasses, those flanked by long terminal repeats (LTR retrotransposons) and those without (non-LTR retrotransposons; Finnegan, 1989; Wells and Feschotte, 2020). The classes and subclasses are further divided hierarchically into order, superfamily, family, subfamily, and lineage (as reviewed in Wicker *et al.*, 2007; Piégu *et al.*, 2015; Neumann *et al.*, 2019).

TRs on the other hand include ribosomal genes, telomeres and satellite DNAs. Satellite DNAs (satDNAs) are highly repetitive non-coding sequences that are arranged in large tandem arrays and contribute to structurally important chromosomal regions such as the centromeres (Lower *et al.*, 2018; Garrido-Ramos, 2021).

Beet genomes, in particular the sugar beet genome, are well characterized regarding retrotransposons (Wollrab *et al.*, 2012; Weber *et al.*, 2013; Heitkam *et al.*, 2014; Schwichtenberg *et al.*, 2016; Maiwald *et al.*, 2021), DNA transposons (Menzel *et al.*, 2006; Menzel *et al.*, 2012), and even related viral relics (Schmidt *et al.*, 2021). Additionally, at least 13 satDNA families are characterized (e.g. Dechyeva and Schmidt, 2006; Zakrzewski *et al.*, 2010). In beets, the most iconic repeat identified so far is the centromeric satDNA pBV (Schmidt and Metzlaff, 1991), which differs from the centromeric satDNA in more distant wild beets (pTS5, pTS4.1; Schmidt and Heslop-Harrison, 1996; Gindullis *et al.*, 2001). However, there is not yet a full picture of the repeat landscape across beet genomes that is needed to identify the evolutionary relationships across time and across species. Expanding the knowledge about the repeat landscapes may help to explain the different crossabilities between the beet gene pools.

To understand the impact of repeats on crossing barriers, beet genomic resources are needed. So far there are three published reference genome sequence assemblies for sugar beet at our disposal (consecutive RefBeet versions: Dohm *et al.*, 2014; consecutive EL10 versions: Funk *et al.*, 2018 and McGrath *et al.*, 2020; KWS2320ONT_v1.0: Sielemann *et al.*, 2023b) and the number of studies comparing the wild beet germplasm to cultivated varieties is continuously increasing (Sielemann *et al.*, 2022, 2023a; Wascher *et al.*, 2022). Hence, sugar beet and its CWRs are well-suited

to serve as a reference for maintaining stable karyotypes despite global changes across the repetitive genome.

To illuminate how the karyotypically stable wild beet genomes differ from another, we focused on repetitive DNA sequences as one of the most impactful contributors to genomic variability. We asked how repeat evolution across wild beets can be linked to speciation and, hence, to crossing barriers. We further investigated if the wild beet germplasm's splits into species, sections, and genera (i.e. primary, secondary, and tertiary gene pools) are mirrored by broad genomic innovations in the repeatome. For this, we chose 17 beet accessions, including all major beet taxa, measured their genome sizes, and determined their chromosome numbers and karyotypes. To add a (pan-)genomic layer, we generated low-pass whole genome shotgun data for all accessions as well as long reads for selected beet genomes. After assessing the phylogenetic placements of these accessions (Sielemann *et al.*, 2022, 2023a), we estimated their repetitive DNA content and finely classified their repetitive DNA sequences in the respective TE and TR hierarchies. This not only allows following repetitive DNA evolution comprehensively across the beet genera, determining repeat gains, losses and replacements, but also linking it back to chromosomal location and karyotypic stability.

RESULTS

Corollinae *species have the largest monoploid genome sizes among sugar beet and wild beets*

To provide the foundation for later analyses of genome and chromosome evolution, we first determined the chromosome configuration and ploidy levels of all 17 cultivated and wild beet accessions, and then estimated their genome sizes by flow cytometry (Additional file 1: Fig. S1; Table 1). Building onto the plastome-based phylogenetics framework (Sielemann *et al.*, 2022), our 17-beet-species-panel allows comprehensive investigation of the *Beta* and *Patellifolia* germplasm.

The beet genome sizes are in accordance with the respective ploidy level (Table 1; Additional file 1: Fig. S1). Diploid members of the section *Beta* have 1C genomes sizes of approx. 700 Mbp, whereas the genome of the sole tetraploid species within this section, *B. macrocarpa*, is nearly twice as large. In contrast, diploid members of the section *Corollinae* achieve considerably higher values with an average 1C genome size of almost 940 Mbp. Also, the polyploid species *B. corolliflora* and *B. intermedia* have

genome sizes slightly larger than mathematically expected (2027 Mbp instead of 1880 Mbp, and approx. 2470 Mbp instead of 2350 Mbp). With a value of approx. 760 Mbp, the genome size of the diploid *B. nana* is more similar to that of the diploids from the section *Beta* than from the section *Corollinae*. Regarding the sister genus, diploid *Patellifolia* species have genome sizes similar to that of the diploids from the section *Beta*, whereas the genome of the tetraploid *P. patellaris* is again slightly larger than mathematically expected (1502 Mbp instead of 1435 Mbp).

Repeat abundance and genome size correlate positively for all major repeat types, except for tandem repeats

We estimated repeat proportions in the genomes of all species through individual as well as comparative clustering using the RepeatExplorer2 pipeline. Combined, the repeats identified for each species represent between 53% (Bptu) and 68% (Bint) of the total genome (Additional file 1: Table S1). For better comparability and to minimize the dependence of the genome size from the ploidy level, polyploid species were also analyzed with downsampled read sets so that they correspond to a diploid chromosome set ('ploidy corrected genome size'). In general, we observe a high correlation between repeat proportion (absolute values) and genome size, with $r^2 = 0.996$ ($p < 8.94e\text{-}17$) for the ploidy corrected genome sizes (Fig. 1A). Further, we observe a clustering of species with larger genome sizes and higher repeat contents (*Corollinae* members) vs. species with smaller genome sizes and lower repeat contents (section *Beta*, *B. nana*, and the genus *Patellifolia*; Fig. 1A).

When individual repeat classes (i.e. LTR retrotransposons, DNA transposons, and satDNAs) are considered, the distinct taxonomic groups can be resolved according to their genome sizes and respective repeat contents (absolute values; Fig. 1B-D, shadings). Again, a positive correlation can be observed when the overall proportion of LTR retrotransposons ($r^2 = 0.948$, $p = 7.32e\text{-}9$) and DNA transposons ($r^2 = 0.798$, $p = 1.22e\text{-}4$) is plotted against the ploidy corrected genome sizes, respectively (Fig. 1B, C). However, when specifically focusing on the proportion of satDNAs, an extraordinarily high amount (up to roughly 14%, see Additional file 1: Table S1) is found in the beet species with small genome sizes, in particular *B. vulgaris* and *B. adanensis*, whereas the beet species with the largest genome sizes from the section *Corollinae/Nanae* (with the exception of *B. macrorhiza*) show the lowest contents of satDNAs (less than 5%). This

results in a negative trend regarding the relation between the satDNA proportion and the genome size (r2 = -0.166, p > 0.52; Fig. 1D). The same observations apply for the overall and individual repeat fractions calculated as proportions (Additional file 1: Fig. S2).

Genomic TEs differences across sugar beet and wild beets reflect the separation into beet sections and genera

The read clustering results based on a 0.1× coverage of the ploidy corrected genome size for each species are analyzed. The repetitive fraction of all analyzed genomes is composed mainly of LTR retrotransposons (14-30% of the genome) with twice the relative content in *B. corolliflora* and *B. intermedia* (*Corollinae* section) compared to species of the section *Beta* (see Additional file 1: Table S1). With the exception of *B. vulgaris*, in which Ty1-*copia* and Ty3-*gypsy* retrotransposons show a rather balanced relation to each other, in most other genomes, the Ty3-*gypsy* elements dominate (Fig. 2A-C, see Additional file 1: Table S1). This observation is most pronounced in *P. webbiana*, where the proportion of Ty3-*gypsy* retrotransposons is four times bigger than that of Ty1-*copia* retrotransposons.

As a substantial difference between species from the section *Beta* and the remaining taxa (section *Corollinae/Nanae* and genus *Patellifolia*), their most abundant LTR retrotransposons belong to different superfamilies: In section *Beta*, SIRE/Maximus retrotransposons from the Ty1-*copia* superfamily contribute the highest TE share (3-4%), whereas the genomes of the other species consist largely of Athila retrotransposons from the Ty3-*gypsy* superfamily (4-6% in *Corollinae* species; 8-9% in *Patellifolia* species). However, *B. nana* stands out as the proportion of Athila elements (2%) within its genome is clearly surpassed by the proportion of Ogre (3%) and Tekay elements (5%). In this respect, *B. nana* resembles the species from the section *Beta* in which Tekay elements are the most abundant Ty3-*gypsy* retrotransposons as well (2-4% of the genome, with approximately as many Tekay elements as SIRE/Maximus elements in the tetraploid *B. macrocarpa*).

Another notable observation is that beets from the section *Corollinae/Nanae* are characterized by a quite high amount of non-autonomous LTR retrotransposons (3-5%, see Additional file 1: Table S1) compared to the other species. The investigation of the

corresponding clusters revealed that this increased share can be traced back to one single, so far uncharacterized element.

Taking a look at the DNA transposon fraction, species from the genus *Beta* are dominated by EnSpm/CACTA terminal inverted repeat (TIR) transposons, whereas the most abundant DNA transposons in *Patellifolia* species belong to the hAT superfamily (Fig. 2D). We find a higher genome proportion of DNA transposons in species with larger genome sizes, with *Patellifolia* species generally having fewer DNA transposons compared to beets of the genus *Beta*.

TEs are generally conserved across the beet genomes and show section-specific abundances

An all-to-all comparison across the beets (accession 1 of every species; see Table 1) revealed that the overall most abundant repeat, a so far unknown LTR retrotransposon of the Athila/Errantivirus lineage, occurs in all analyzed genomes, representing a genome proportion of 0.89% (Bnan1) up to 8.08% (Pweb1; Additional file 1: Fig. S3). However, the high Athila/Errantivirus proportion in *Corollinae* species is caused by another LTR retrotransposon originally described as five distinct dispersed repeats by Gao *et al.* (2000; pBC1054, pBC227, pBC305, pBC507/169, and pBC537). In general, about half of the repeat families are found in all analyzed beet genomes. The other half consists of species-, section-, and genus-specific repeats. There are fewer *Patellifolia*-specific repeats (33 out of 395 clusters) than *Beta*-specific repeats (172 out of 395 clusters), resulting in a lower overall repeat diversity in the genus *Patellifolia*. However, the *Patellifolia*-specific repeats are highly abundant, whereas all repeats in the *Beta* genomes (with the exception of some satDNAs, see paragraph below) are rather moderately abundant. Yet, we also find section-specific repeats with regard to the sections *Beta* (8 out of 172 genus Beta-specific clusters) and *Corollinae/Nanae* (11 out of 172 genus *Beta*-specific clusters). Several LTR retrotransposons (Athila, Tekay, Tat, and Ale/Retrofit elements) seem to be re-amplified in the beet genomes of the *Corollinae/Nanae* section, whereas an increased abundance of SIRE/Maximus elements was observed in the beet genomes of the *Beta* section. Furthermore, a so far unknown Ogre LTR retrotransposon, as well as the non-autonomous LTR retrotransposon mentioned in the paragraph above, set *Corollinae/Nanae* apart from all other analyzed beets (see Fig. 2C and Additional file 1: Fig. S3).

Beet satDNAs break ranks (1): Beta *and* Patellifolia *satDNAs are few, but highly amplified, whereas* Corollinae/Nanae *satDNAs are diverse and lowly abundant*

Comparing *Beta* and *Patellifolia* genomes, the greatest repeatome difference is found in the satDNAs (Fig. 3). They contribute roughly 6% to the genomes of *Patellifolia* species and only 4% to the genomes of the *Corollinae/Nanae* species (with the exception of *B. macrorhiza*; see Additional file 1: Table S1). In contrast, in genomes of the section *Beta*, satDNAs contribute up to 14%, which is about as much as the LTR retrotransposon fraction.

The reason for this divergence is the extraordinarily high abundance of two iconic sugar beet satDNAs (see Additional file 1: Fig. S3), the centromeric beetSat01-pBV and the intercalary beetSat02-pEV (Schmidt and Metzlaff, 1991; Schmidt *et al.*, 1991). BeetSat01-pBV and beetSat02-pEV are highly abundant in all species of the section *Beta* contributing up to 8% and 5% of the genomes, respectively (Fig. 3; see Additional file 1: Table S2). However, beetSat02-pEV also occurs in beets of the section *Corollinae/Nanae* and within the sister genus *Patellifolia* with low or moderate abundance (Fig. 3). In contrast, canonical beetSat01-pBV satDNA arrays (as described by Schmidt and Metzlaff, 1991; Zakrzewski *et al.*, 2011) are strictly limited to the beets of the section *Beta*.

Although genomes of the *Patellifolia* genus contain less satDNAs than those of the section *Beta*, they are also dominated by abundant satDNA families (Fig. 3; see Additional file 1: Table S2): Again, the most prominent satDNAs in the *Patellifolia* species are those that constitute the centromeres: the centromeric beetSat03-pTS5 and the pericentromeric beetSat04-pTS4.1 (Schmidt & Heslop-Harrison, 1996). However, in sequence and monomer length they are substantially different from beetSat01-pBV. In addition, we identified another highly abundant and genus-specific satDNA (beetSat05). The satDNA designated as beetSat06 is the only satDNA that is distributed quite equally among all beet genomes (Fig. 3; Additional file 1: Table S2). This subtelomeric satDNA ubiquitously occurs on all chromosomes and was previously described under different names depending on the respective plant species (e.g. in *B. vulgaris* it is known as pAv34; Dechyeva & Schmidt, 2006). However, Dechyeva and Schmidt (2006) found that, in the wild beet *B. nana*, this satDNA is restricted to one single pair

of chromosomes which is consistent with the reduced abundance of beetSat06 that we observed in *B. nana* using the comparative repeat analysis (Fig. 3).

Generally, within the *Corollinae/Nanae*, the specific satDNA quantities set *B. nana* slightly apart from the other species within this section (Fig. 3; Additional file 1: Table S2). This applies for other known satDNAs such as beetSat07-pHC8 (Gindullis *et al.*, 2001; Fig. 4G, H) or beetSat10-pRN (Kubis *et al.*, 1997; Fig. 4I, J), as well as for the satDNAs newly identified during the comparative repeat analysis (beetSat15–18; Additional file 1: Data S1 and Data S2). These satDNAs are most pronounced in beet genomes of the section *Corollinae/Nanae* without the conspicuously high abundances observed in the section *Beta* and the sister genus *Patellifolia*. Among the satDNAs that define the *Corollinae/Nanae* genomes, only beetSat13 is prominent, especially in *B. macrorhiza* (Fig. 3). This satDNA is also known as pBC1447 in *B. corolliflora* and ChenSat-1a in the related *Chenopodium quinoa* (Gao *et al.*, 2000; Heitkam *et al.*, 2020).

Overall, regarding the interplay of satDNA diversity and abundance, we conclude that *Beta* and *Patellifolia* contain relatively few satDNA families that can reach high copy numbers. In contrast, species within the *Corollinae/Nanae* accumulate a wide satDNA variety with only low amplification levels (Fig. 3).

Beet satDNAs break ranks (2): Species- and section-specific array expansions lead to the emergence of unique satDNA landscapes across the beet genomes

As presented above, individual satDNAs can occur in high abundance in one beet species and in low abundance in another. To understand, if beet satDNA organization is retained across species and abundance patterns, we investigated their tandem arrangement across the three main beet taxa. By using long reads from *B. vulgaris* (section *Beta*), *B. corolliflora* (section *Corollinae*), and *P. procumbens* (genus *Patellifolia*) we showed that nearly all identified satDNAs occur in long tandem arrays in at least one of the three beet taxa (see dotplot visualization in Additional file 1: Fig. S4). Only beetSat18 does not show any long arrays in neither of the three species. This is easily explained as this satDNA is specific for *B. nana*, with head-to-tail arrangements identified on *B. nana* short reads.

We noted variation in monomer length and sequence (Additional file 1: Fig. S5), often with section specificity. Sugar beet's main satDNA beetSat01-pBV is especially intriguing: It occurs only scarcely in *Patellifolia* genomes, forming no tandem arrays.

However, in *B. corolliflora*, the beetSat01-pBV monomer is part of a different, tandemly arranged repeat, resulting in a much longer monomer size of approx. 1800 bp (Additional file 1: Fig. S4). This larger tandem repeat shows no further sequence similarity to publicly available nucleotide or protein database entries. Indeed, it may serve as a starting point to understand formation of one of the largest satDNA families in any plant genome. Similar potential start points of satDNA emergence were detected for beetSat10, beetSat11, beetSat12, beetSat17, and beetSat18 (Additional file 1: Fig. S4; arrows), suggesting a general pattern.

Further, monomer length variations were detected, indicating section specificity, but none as striking as for beetSat1-pBV. For instance, beetSat02-pEV variants diverge slightly in monomer length and sequence among the three beet taxa (Additional file 1: Fig. S5), whereas beetSat06 shows variation in higher order arrangement.

Overall, regarding satDNA amplification and array formation, we conclude that different satDNAs amplify in different beet species. Whereas large arrays can form in one wild beet genome, the same satDNA can occur only as a relic in the next. Similarly, emergence of species-/section-specific satDNA variants occurs. We observe that the respective beet satDNA landscapes are characterized by different evolutionary mechanisms: Whereas few *Beta* and *Patellifolia* satDNAs underwent amplification and homogenization, the satDNAs landscape of *Corollinae/Nanae* species is still rather dynamic, mirrored by the emergence of several new satDNAs.

Beet satDNAs break ranks (3): Beta *and* Patellifolia *satDNAs constitute major parts of all chromosomes, whereas* Corollinae/Nanae *satDNAs are restricted to chromosome subsets*

To understand the structural chromosome makeup across the beets, we again focused on the three species *B. vulgaris*, *B. corolliflora* and *P. procumbens* to represent the three main beet taxa. First, we investigated the five main satDNAs: beetSat01-pBV to beetSat05.

Our two-color FISH onto sugar beet chromosomes shows that beetSat01-pBV (green) constitutes the centromeres, whereas beetSat02-pEV (red) builds large intercalary blocks along all *B. vulgaris* chromosomes (Fig. 4A, B; Kubis *et al.*, 1998). In contrast, beetSat02-pEV is restricted to only a subset of chromosomes and/or comparatively small intercalary regions along *B. corolliflora* and *P. procumbens* chromosomes

(Fig. 4C, D). This indicates that beetSat02-pEV is not only specific in its monomer sequence (see above), but also in its chromosomal localization among the three beet taxa.

In *P. procumbens*, beetSat03-pTS5 marks most, but not all centromeres (Fig. 4E, green signals; see also Gindullis *et al.*, 2001). BeetSat05 resides at as many chromosomes (10-14 centromeres; Fig. 4E, blue signals). Thus, each centromere contains either beetSat03-pTS5 or beetSat05, or both. All centromeres are flanked by beetSat04-pTS4.1 that accompanies the centromeric beetSat03-pTS5 and beetSat05 signals, but also occupies some distal locations (Fig. 4E, red signals). All five main satDNAs are restricted to the DAPI-positive heterochromatin in interphase nuclei (Fig. 4B, F). Concluding, we note that chromosomes of *B. vulgaris* and *P. procumbens* – despite their phylogenetic distance – are organized similarly, with main satDNAs building the centromeres and intercalary regions.

This is in sharp contrast to the chromosome organization in *B. corolliflora*, a representative of the *Corollinae/Nanae* section: Out of the many satDNAs with varying degrees of amplification and homogenization (see above), we selected six for hybridizations (beetSat07-pHC8, beetsSat08, beetSat10-pRN, beetSat13, beetSat15, and beetSat17; Figure 4G-N). None of them is exclusively localized at the centromeres. Instead, intercalary signals are common as well and (peri-)centromeric signals are restricted to only some chromosomes, often being less prominent than centromeric *Beta/Patellifolia* signals. The comparison of two of these satDNAs at *B. vulgaris* and *B. corolliflora* chromosomes reveals that the signal patterns are more (beetSat07-pHC8) or less (beetSat10-pRN) similar in both genomes (Fig. 4G-J):

BeetSat07-pHC8 resides on all *B. vulgaris* chromosomes in intercalary and distal regions, but also close to at least two centromeres (Fig. 4G). In *B. corolliflora*, beetSat07-pHC8 resides in two centromeric and several intercalary regions as well. Major signals are pronounced on six chromosomes (presumably three chromosome pairs; Fig. 4H), including the two centromeric signals.

The second satDNA, beetSaat10-pRN, is lowly abundant in *B. vulgaris* (see Fig. 3), thus producing only few, faint and scattered signals (Fig. 4I). In contrast, beetSat10-pRN hybridized strongly to 6-8 chromosomes of *B. corolliflora*, predominantly in the centromeres (Fig. 4J).

To better understand how the patchy satDNA emergence/amplification patterns in the *Corollinae/Nanae* affect the chromosomes, we also localized beetSat08, beetSat13, beetSat15, and beetSat17 (Figure 4K-N):

BeetSat08 resides on nearly all *B. corolliflora* chromosomes (Fig. 4K). Two major and two minor signals mark the centromeres of four chromosomes, whereas the majority of signals was rather weak and distributed over the intercalary and distal regions.

BeetSat13, which is known to be associated with the centromeres of *C. quinoa* chromosomes (ChenSat-1a; Heitkam *et al.*, 2020), can also be found at the centromeres of at least ten *B. corolliflora* chromosomes (Fig. 4L). BeetSat13 signals near the centromere were detected on seven additional chromosomes. The remaining chromosomes did not hybridize with the beetSat13 probe or showed intercalary beetSat13 signals.

All six major sites of beetSat15 hybridization are associated with *B. corolliflora* centromeres as well (Fig. 4M).

The comparatively low abundance of beetSat17 (see Fig. 3) resulted in only few hybridization signals in *B. corolliflora* (Fig. 4N): Signals were detected on twelve chromosomes with at least two of them showing weak centromeric signals.

Overall, regarding the chromosomal impacts of the vastly different satDNA landscapes in *Beta* and *Patellifolia* versus *Corollinae/Nanae*, we note that the different evolutionary patterns affect chromosome structure, especially the centromeres. *Beta* and *Patellifolia* centromeres are made up of few, highly abundant, homogenized satDNAs. Instead, the *Corollinae/Nanae* harbor 'patchwork centromeres': We identified at least 28 (peri-)centromeric signals with six different satDNAs (beetSat07-pHC8, beetSat08, beetSat10-pRN, beetSat13, beetSat15, beetSat17) on *B. corolliflora* chromosomes. From this, we conclude that *Corollinae/Nanae* centromeres are constituted by a multitude of different satDNAs, whereas *Beta* and *Patellifolia* centromeres are constituted by one or two main satDNAs, respectively.

DISCUSSION

A beet species panel to understand the evolving genome under karyotypic stability

We leverage a comprehensive repeatome study across 17 accessions of cultivated and wild beet species, spanning two sister genera, *Beta* and *Patellifolia*. We target all major species across all sections, including higher polyploids. Building onto the plastome-

433 based phylogenetics framework (Sielemann *et al.*, 2022, 2023a), our 17-beet-species-
434 panel represents a well-characterized sugar beet and CWR panel with regard to tracing
435 genome evolution, chromosome stability, pangenomics, and taxonomy.
436
437 *Beet and wild beet genome size variation results from repeat content fluctuations*
438 Considering the respective ploidy level, the beet genome sizes show relatively moderate
439 variation: 1.4-fold variation among the diploids, 1.5-fold variation among the
440 tetraploids, and also 1.5-fold variation among the ploidy corrected genome sizes.
441 Meanwhile, intraspecific differences are minimal, with almost no variation. For taxa
442 with high chromosomal variability, e.g. across *Euphrasia* individuals (1.3-fold
443 variation), structural changes such as the loss or gain of chromosome fragments are
444 likely responsible for genome size variations (Becher *et al.*, 2021). In contrast, we
445 hypothesize that genome size variation within plant taxa with stable chromosome
446 setups, such as the analyzed beet species, depends strongly on genomic repeats. This
447 can be demonstrated by a clear correlation between genome size and repeat fraction. For
448 the beet species, we have found such a correlation not only between the overall repeat
449 content, but especially between the LTR retrotransposons and the respective genome
450 size, indicating that the amplification and elimination of LTR retrotransposons in
451 particular is causal for genome size differences between beets of the same ploidy. In
452 comparison to other angiosperms, the determined correlation is at least as high (e.g.
453 compared to *Fabeae* sp.; Macas *et al.*, 2015), if not higher (e.g. compared to *Eleocharis*
454 sp., *Solanum* sp., *Hesperis*-clade sp.; Zedek *et al.*, 2010; Michael, 2014; Gaiero *et al.*,
455 2019; Hloušková *et al.*, 2019; Gantuz *et al.*, 2021), pointing to a particularly strong
456 impact of repeats (i.e. LTR retrotransposons) on beet genome sizes.
457 Specific amplification and elimination of repeats may explain why the genome size of
458 *B. nana* rather resembles those of the section *Beta*, even though this species is
459 repeatedly considered a member of the section *Corollinae* (Kadereit *et al.*, 2006; Frese
460 and Ford-Lloyd, 2020; Sielemann *et al.*, 2022). However, genome sizes of beets from
461 the genus *Patellifolia* are most similar to those of the section *Beta*, suggesting that
462 repeat amplification and/or acquisition caused a genome upsizing in *Corollinae/Nanae*
463 species.
464

Repeat abundances mirror the phylogenies of the beet and wild beet sections

With 53-68%, the repetitive fraction of beet genomes represents typical values compared to other Amaranthaceae (approx. 76% in quinoa: Heitkam *et al.*, 2020; approx. 51% in spinach: Li *et al.*, 2021). Discrepancies in repeat abundances to a previous RepeatExplorer2 analysis in *B. vulgaris* (Kowar *et al.*, 2016) are moderate and result from the exclusion of organellar DNA from our read set as well as the improvement of the repeat annotation by including a more comprehensive, custom repeat database. In general, the repeat content results not only from the number of repetitive sequences, but also from their type: A repeat with a longer element structure accounts for a larger share of the genome than a shorter repeat. Within the beet genomes, the most frequent repeats belong to the LTR retrotransposons, namely the Ty3-*gypsy* retrotransposons, which include the longest known TEs in plants (i.e. Ogre elements with >23 kb in length; Orozco-Arias *et al.*, 2019). The predominance of Ty3-*gypsy* retrotransposons within the repeat fraction was observed for other angiosperms as well (Kelly *et al.*, 2015; Macas *et al.*, 2015; Gaiero *et al.*, 2019; Hloušková *et al.*, 2019; Dodsworth *et al.*, 2020). However, the rather balanced share of Ty3-*gypsy* and Ty1-*copia* retrotransposons in *B. vulgaris* may reflect different TE dynamics in the domesticated beet cultivars, whereas the independent amplification and/or acquisition of Ty3-*gypsy* retrotransposons may have increased their abundance within the CWRs.

We found that most TE families are distributed across all analyzed beet species, indicating that the present beet repeat set can be traced back to the last common ancestor of both genera *Beta* and *Patellifolia*. After speciation from this ancestor, a few new TE families emerged, though rather sparsely. Instead, specific TE amplification/elimination has led to genus- and section-specific TE abundances. This global trend in beet genomes is mirrored in the ups and downs of individual TE families/lineages (e.g. SIRE/Maximus elements: Weber *et al.*, 2010; chromoviruses: Weber *et al.*, 2013; LINEs: Heitkam *et al.*, 2014; non-autonomous LTR retrotransposons: Maiwald *et al.*, 2021). Differences in the repeat profiles support the relationships of the current beet phylogenies (Frese and Ford-Lloyd, 2020; Sielemann *et al.*, 2022) with characteristic repeatomes for species from the sections *Beta*, *Corollinae/Nanae*, and from the genus *Patellifolia*, respectively. However, whereas the amplification of TE families/lineages (in particular Ogre and non-autonomous LTR retrotransposons) has probably led to an increase in genome size of *Corollinae/Nanae*

498 species, accumulation of specific TEs in the genomes of *Patellifolia* or section *Beta*
499 species seem to have taken place at a comparable level, so that no striking differences in
500 genome size are apparent between those beet taxa. To be more precise, there has been a
501 strong amplification of few repeats in *Patellifolia* genomes, while many repeats have
502 been amplified rather moderately in genomes of the section *Beta*.
503 Since the genomic shock of polyploidisation may stimulate rapid and dynamic genomic
504 changes such as the activation of retrotransposons and viral elements (McClintock,
505 1984; Lopez-Gomollon *et al.*, 2021), genome size may subsequently increase. The
506 overall repeat content is indeed greater within the polyploid beet genomes compared to
507 their closest diploid relatives, mainly due to an increased abundance of Ty3-*gypsy*
508 retrotransposons. For most polyploid beet species, we measured larger genome sizes
509 than mathematically expected, which can thus be attributed to a general accumulation of
510 all kinds of present Ty3-*gypsy* retrotransposons instead of the targeted amplification of
511 distinct TEs. During the 'cycle of polyploidy', polyploid plants usually undergo genome
512 downsizing (Wendel, 2015), which may be the case for *B. macrocarpa* (since its
513 genome size is smaller than mathematically expected) but not for the other polyploid
514 beet species, indicating that these are relatively young polyploids (less than 0.9-1.4
515 million years; Romeiras *et al.*, 2016).
516 It is often observed that genomic repeat profiles contain a phylogenetic signal
517 (Dodsworth *et al.*, 2015; McCann *et al.*, 2020; Vitales *et al.*, 2020b; Herklotz *et al.*,
518 2021). In the case of *Beta* and *Patellifolia* species, this is only partly true: Overall, the
519 repeatomes enable to separate the beet species into genera and sections. However,
520 repeatome differences at the subspecies level do not reflect the currently proposed
521 relationships of the subspecies taxa: As an example, *B. vulgaris* is considered to be
522 more closely related to *B. maritima* accessions, regardless of their origin, than to
523 *B. adanensis* (Wascher *et al.*, 2022). According to kmer-based genomic distance,
524 *B. adanensis* may even be considered a distinct species rather than a subspecies
525 (Wascher *et al.*, 2022). In contradiction with this report, focusing on the three
526 subspecies *B. adanensis*, *B. vulgaris*, and *B. maritima*, the repeat profiles of the first two
527 are most similar to each other. This may be explained by the *B. maritima* accessions
528 used here: both Bmar1 and Bmar2 are from the Atlantic coast and are therefore less
529 closely related to *B. vulgaris* than its presumed progenitor *B. maritima* from the
530 Mediterranean area. From this, we conclude that the repeat profiles are not suited to

address still debated phylogenetic questions regarding the (sub-)species level of beets. This also concerns the non-uniform treatment of *P. procumbens* and *P. webbiana* as distinct or as the same species.

The abundance of a distinct repeat depends on several factors that either cause an increase, a reduction, or that keep the current copy number stable. Those factors include an enhanced amplification (as mentioned above), the targeted or non-targeted elimination from the genome (e.g. by defense mechanisms and/or the loss of whole chromosomes or chromosome fragments), and selective pressure for the maintenance of specific genomic regions. An example in which some of these factors come into play is the section-specific abundance of endogenous pararetroviruses (EPRVs) found among the beets. There are three to five times more EPRVs within the genomes of wild beets from the section *Corollinae* compared to beets from the sister genus *Patellifolia*, the section *Beta*, and even *B. nana*. Such differences were also found between wild and cultivated potato species (Gaiero *et al.*, 2019) and it was assumed that the EPRVs underwent either an increased amplification in wild potatoes or a selective bias in cultivated potatoes. As for sugar beet, we know that there is a trade-off between the targeted inactivation and a simultaneous preservation of EPRV sequences (Schmidt *et al.*, 2021). Therefore, we believe that *Corollinae* species accumulated pararetroviral sequences by the gain of further EPRVs in comparison to the remaining beet species. Similar section-specific emergence and loss occurred also for other TE families/lineages, especially among the LTR retrotransposons.

SatDNAs emerge, amplify and vanish without impacting the global genome size

The only repeat type for which no correlation with the beet genome size was found are the satDNAs. Strikingly, of all repeats, the satDNAs show the most pronounced specificity in abundance and distribution among the beet genomes, pointing to the fact that satDNAs are the most dynamic repeats (Garrido-Ramos, 2021). With this study, we present the first all-encompassing account of every present *Beta/Patellifolia* satDNA, involving all known as well as so far unpublished satDNAs. By comparison to publicly available databases, we determined that, with the exception of beetSat02-pEV (also found in quinoa: Schmidt *et al.*, 2014), beetSat06-pAv34 (pRs34; also found in spinach: Dechyeva and Schmidt, 2006), and beetSat13 (ChenSat-1a; also found in quinoa:

Heitkam *et al.*, 2020), the satDNAs are specific for and restricted to beet species from the genera *Beta* and *Patellifolia*.

Usually, the repeatome of plants is constituted mostly by LTR retrotransposons, accounting for up to 80% of the plant genome size (Orozco-Arias *et al.*, 2019). Only few plants are known in which the proportion of satDNAs is comparable to that of TEs (e.g. olive: Barghini *et al.*, 2014; radish: He *et al.*, 2015; *Fritillaria affinis*: Kelly *et al.*, 2015). Here, we present sugar beet (*B. vulgaris*) as another plant species with peculiar high amounts of satDNAs. However, with diminishing phylogenetic relationship to sugar beet, the proportion of satDNAs decreases in its CWRs. In addition, each beet section/genus has its own distinct repeat profile, which sets it apart from the other (Fig. 5). Thus, we observed satDNAs specific for the genus *Patellifolia*, as well as for the sections *Corollinae/Nanae* (three different satDNAs, each) and *Beta* (one satDNA). Even the sole two satDNAs that are present in all analyzed beet species (i.e. beetSat02-pEV, beetSat06-pAV34) show variability in abundance as well as sequence and monomer length, resulting in genus- and section-specific variants that may initiate homogenization processes and sequence shifts in the future.

As beetSat02-pEV and beetSat06-pAV34 are distributed among all analyzed beets (Fig. 5) and as variants of them are also present in more distantly related plant species (beetSat02-pEV in quinoa: Schmidt *et al.*, 2014; beetSat06-pAV34 in spinach: Dechyeva and Schmidt, 2006), we assume that these satDNAs were already present in the common ancestor of both genera *Beta* and *Patellifolia*. During the subsequent beet speciation, a great variety of new satDNAs emerged and some of them accumulated section-specifically by remarkable re-amplification. Such a restriction to individual sections or species is sometimes known for TEs, but is particularly pronounced for the fast evolving satDNAs (Orozco-Arias *et al.*, 2019, Garrido-Ramos, 2021). For example, beetSat13's specificity in abundance is visible even at the level of beet accessions (Bmrh1 vs. Bmrh2). Moreover, this satDNA can be found in all *Corollinae/Nanae* species as well as in the more distantly related quinoa (ChenSat-1a: Heitkam *et al.*, 2020), but not in beet genomes from the genus *Patellifolia* and the section *Beta* (Fig. 5). The patchy distribution of beetSat13 among Amaranthaceae members may be either explained by its loss in most of the beet genomes except those from the section *Corollinae/Nanae*, or by an independent acquisition into the genome of the *Corollinae/Nanae* ancestor. It was argued that in quinoa this satDNA emerged from a

CACTA-like TE (Belyayev *et al.*, 2020). Thus, an independent acquisition/generation of beetSat13 into/within the genome of the *Corollinae/Nanae* ancestor may have occurred similarly by spawning from a TE or even by the transmission of extrachromosomal circular molecules consisting of this satDNA (Navrátilová *et al.*, 2008).

SatDNAs impact beet chromosome architecture in a gene pool-dependent manner

Apart from their role as intercalary heterochromatin (beetSat02-pEV) and (sub-)telomere contributors (beetSat06-pAV34), satDNAs often provide the DNA backbone onto which the centromeres are formed (Melters *et al.*, 2013; Garrido-Ramos, 2021). Strikingly, this iconic location is not occupied by the same satDNA among beets, but by a peculiarly high number of different satDNAs (Fig. 5). It is not unusual for the centromeric satDNAs to differ even in very closely related species (Henikoff *et al.*, 2001; Melters *et al.*, 2013). However, whereas the centromeric satDNA in beets from the section *Beta* is well defined and conserved across the whole karyotype (Schmidt and Metzlaff, 1991; Zakrzewski *et al.*, 2011), beets from the section *Corollinae/Nanae* and the genus *Patellifolia* seem to differ in their satDNA composition even from chromosome to chromosome. This phenomenon was previously observed in potatoes (Gong *et al.*, 2012) and especially in legumes (Avila Robledillo *et al.*, 2020), as well as in animal species (chicken: Shang *et al.*, 2010). Usually, the most abundant satDNA of a species is the one constituting the centromeres (Melters *et al.*, 2013). Yet, while this is true for beets from the section *Beta*, there is no clearly predominant satDNA in beets from the section *Corollinae/Nanae* and the genus *Patellifolia*, going along with the observation that the variety of similarly abundant satDNAs is spread over the (peri-)centromeric regions of subsets of chromosomes. This is particularly remarkable as these satDNAs (i.e. beetSat07-pHC8 to beetSat11, beetSat13, beetSat15, and beetSat17 in *Corollinae/Nanae* species; beetSat03-pTS5 and beetSat05 in *Patellifolia* species) differ considerably in their monomer lengths and/or sequences and yet chromosomes maintain an error-free cell division. Furthermore, centromeric satDNA families of other closely related plant taxa (e.g. *Arabidopsis* species), albeit species-specific, nevertheless derived from a common ancestor (summarized by Garrido-Ramos, 2021), which does not seem to be true for the beets: Monomer sequence and length differences are too large to be explained by high mutation rates, even though some satDNAs (e.g.

beetSat01-pBV) indeed evolve towards quite considerable changes in monomer sequence and length (see Additional file 1: Fig. S4). The centromere drive model (Henikoff *et al.*, 2001) does not seem to apply either, as the presence of multiple centromeric satDNAs with different sequences precludes any sequence-dependent coevolution with the kinetochore complex (Avila Robledillo *et al.*, 2020). Considering the evolutionary mode of satDNAs, which are thought to typically evolve in a concerted manner (Šatović-Vukšić and Plohl, 2023), it may be that *Corollinae/Nanae* and *Patellifolia* species are only at an intermediate stage on the way to centromere and satDNA homogenization. This is supported by the fact that beet species from the section *Corollinae/Nanae* in particular belong to a highly variable hybrid complex (Frese and Ford-Lloyd, 2020). Hence, hybridization events between genotypes with different satDNA profiles may frequently disrupt any ongoing satDNA homogenization processes. This may have led to the observed variety of lowly amplified satDNAs, which were then selectively retained to form unique satDNA mixes at each centromere within one nucleus. This would then result in the observed combination of different chromosome sets with different satDNA architectures. To account for their unique chromosome make-up, we refer to the *Corollinae/Nanae* chromosomes as patchwork chromosomes (Fig. 5; middle). Alternatively, keeping a set of different centromeres may hold an advantage for these beets, especially in chromosome recognition during meiotic pairing. For the *Corollinae/Nanae*, we cannot yet determine if there is chromosome-complementarity among individual satDNAs. Nevertheless, multicolor FISH indicates a possible chromosome-complementarity of beetSat03-pTS and beetSat05 in the genus *Patellifolia*. Taking together, we detect vastly different satDNA landscapes that vary in satDNA family number, abundance and location. These provide the DNA backbone for the formation of overall different chromosome architectures within a nucleus, leading from overall similar chromosome/centromere structures (section *Beta*), over conceptually similar chromosomes (*Patellifolia*) towards completely different chromosomal/centromeric setups (*Corollinae/Nanae*).

Conclusion

In the use of wild germplasm for breeding, crossing barriers often impede the sustainable introgression of target traits from wild species into crops. This is also the case for cultivated and wild beet species, which are taxa with isokaryotypic

chromosomes, but which nevertheless exhibit strong crossing borders between gene pools. We found that global differences between beet genomes can be attributed primarily to their repeatomes, especially to the specific composition of satDNAs, whereas the base number and morphology of chromosomes is rather consistent. After the divergence from the common beet ancestor, genus- and section-specific repeat variants propagated by independent transposition and centromeres evolved distinctly, corresponding to the beet phylogeny and the categorization into gene pools. Based on their repeatome, beet species divide into three different groups (section *Beta* as primary gene pool, section *Corollinae/Nanae* as secondary gene pool, and genus *Patellifolia* as tertiary gene pool), supporting the repeated handling of *B. nana* as a member of the section *Corollinae*. In comparison to the other beet species, this section is characterized by a peculiarly high number of different satDNAs with comparable abundances. We found that these satDNAs contribute to the *Corollinae/Nanae* centromeres in a composite manner, resulting in a highly variable chromosome composition (patchwork chromosomes). These insights are valuable for future beet breeding as they help to face the challenge of overcoming postzygotic crossing barriers and provide unique insights into genome, repeatome, and chromosome evolution in karyotypically stable taxa.

METHODS

Plant material, genome size measurement, and DNA extraction

The 17 investigated beet accessions cover the whole spectrum of species within the genera *Beta* and *Patellifolia* (see Table 1). Seeds of the *B. vulgaris* subsp. *vulgaris* genotype KWS2320 were obtained from KWS Saat SE, Einbeck, Germany. The seeds of all other accessions were obtained from the Leibniz Institute of Plant Genetics and Crop Plant Research Gatersleben, Germany. The material of the KWS SAAT SE & Co. KGaA, Einbeck and IPK Gatersleben was transferred under the regulations of the standard material transfer agreement (SMTA) of the International Treaty. The plants were grown under long day conditions in a greenhouse.

Nuclei extraction and staining were performed using the CyStain PI OxProtect reagent kit from Sysmex. Nuclear DNA was measured by flow cytometry, using *Raphanus sativus* (2C = 1.11 pg DNA) or *Pisum sativum* (2C = 9.07 pg DNA; Doležel *et al.*, 1992) as internal standard. Four DNA estimations were carried out for each plant (5000 nuclei per analysis) on at least two different days. Nuclear DNA content (2C value in

[pg]) was calculated as: sample peak mean/standard peak mean × 2C DNA content of the standard. DNA amounts in picograms were converted to the number of base pairs using the conversion factor 1 pg DNA = 0.978 × 109 bp (Doležel *et al.*, 2003). The mean nuclear DNA content was then calculated for each plant as 1C value (in [Mbp]; Table 1).

Genomic DNA was extracted from approximately 100 mg of fresh leaf tissue samples using the NucleoSpin® Plant II protocols from Macherey-Nagel. Libraries were prepared using the TruSeq Nano DNA library preparation kit (Fa. Illumina) and were sequenced on an Illumina HiSeq1500 sequencer at the CeBiTec Sequencing Core Facility, Bielefeld University, Germany (250 bp paired-end reads).

Repeat classification and quantification

The Illumina reads were trimmed to a length of 100 bp using Trimmomatic (v0.39; Bolger *et al.*, 2014). High quality of the trimmed reads was ensured by FastQC examination (v0.11.5; Andrews, 2020). Bowtie2 (v2.2.6; Langmead and Salzberg, 2012) was applied to identify and subsequently remove reads representing organellar DNA from the sequence data by mapping them to a database containing publicly available chloroplast and mitochondrial DNAs of *Beta* species (https://ncbi.nlm.nih.gov/nuccore). The sequence reads were then subsampled to reduce the genome coverage to 0.1× for all species, and different numbers of paired-end reads sampled depending on the genome size. To standardize all read pre-treatments, we have used the mentioned tools embedded in the preparation module of the ECCsplorer pipeline (Mann *et al.*, 2022). For the identification and quantification of repetitive sequences, the similarity-based read clustering method was applied as described by Novák *et al.* (2010) and implemented in the RepeatExplorer2 pipeline (Novák *et al.*, 2013, 2020). We used the pipeline default settings (90% similarity over 55% of the read length) and included our custom database of Betoideae repeats (https://zenodo.org/record/8255813) for individual analyses (each beet accession distinctly) as well as an all-vs-all comparison across all species (accession 1 of every species; see Table 1). When comparing different accessions of the same species (Bmar1/Bmar2, Bmrh1/Bmrh2, Bint1/Bint2, and Bnan1/Bnan2), the RepeatExplorer2 pipeline produced highly similar results for the respective repeat compositions (see Additional file 1: Table S1). Any discrepancies resulted from the read sampling as we

confirmed by using re-sampled read datasets as input. Hence, only one accession per species was included in the comparative RepeatExplorer2 analysis.

For this all-vs-all comparison, each read set was downsampled to represent 4% of the respective genome (coverage of 0.04×) based on 'ploidy corrected genome sizes' (1C values [Table 1] divided by 2 [for the tetraploid species *B. macrocarpa*, *B. corolliflora*, and *P. patellaris*] and 2.5 [for the pentaploid species *B. intermedia*], respectively). Nevertheless, the pipeline reached a limit, retrieving 3,985,090 reads. Although automatic annotation of read clusters was improved by the inclusion of our custom Betoideae repeat database, not all clusters were classified. The cluster graph shapes of all unassigned clusters that represented at least 0.001% of the investigated genomes, were examined manually and consensus sequences, if provided, were searched using publicly available databases (https://blast.ncbi.nlm.nih.gov/Blast.cgi). Finally, the relative abundance of each repeat was calculated based on the number of reads in the respective cluster, excluding remaining organellar reads.

Sequence comparison

Multiple sequence alignments were generated with the MAFFT (v7.017; Katoh and Standley, 2013) and MUSCLE (v3.8.31; Edgar, 2004) local alignment tools. They have been manually refined and used for the calculation of pairwise sequence identities. We explored and visualized sequences with the multipurpose software Geneious 6.1.8 (Kearse *et al.*, 2012). Dotplots were generated with FlexiDot (Seibt *et al.*, 2018) with a word size of 20 bp using long reads of *B. vulgaris* (PacBio; accession number SRX3402137; Funk *et al.*, 2018), *B. corolliflora* (ONT; accession number ERS13530775; Sielemann *et al.*, 2023a), and *P. procumbens* (ONT; accession number ERS13530778; Sielemann *et al.*, 2023a).

Generation of satDNA probes

The probe for the *Patellifolia*-specific satDNA beetSat05 was ordered as EXTREmer oligonucleotides synthesized by *Eurofins Genomics* based on the RepeatExplorer2 consensus sequence (Additional file 1: Data S1).

Standard PCR reactions of genomic *B. corolliflora* (BETA 408) and *B. nana* (BETA 546) DNA were performed using primer pairs designed for two further beet-specific satDNAs (beetSat15 and beetSat17; Additional file 1: Table S3). The PCR

conditions were 94 °C for 3 min followed by 35 cycles of 94 °C for 30 s, primer-specific annealing temperature for 30 s, 72 °C for 25 sec (beetSat15) and 45 s (beetSat17), respectively, and a final incubation at 72 °C for 5 min. PCR fragments were purified, cloned and commercially sequenced. Sequenced inserts spanning two (beetSat17; 279 bp) and six (betSat15; 303 bp) monomers of the respective satDNA were used as probes for the following hybridization experiments. Their identity to the respective reference sequence (RepeatExplorer2 consensus sequence) was 96.4% (beetSat15) and 93.7% (beetSat17), respectively.

Preparation of chromosome spreads and fluorescent in situ hybridization (FISH)

The meristem of young leaves was used for the preparation of mitotic chromosomes. The plant tissues were treated as described by Schmidt *et al.* (2023). Accumulation of metaphases was achieved by a combination of an incubation with 2 mM 8-hydroxyquinoline for 3 h and nitrous oxide for 30 min. Fixed plant material was digested using the 'leaf enzyme solution II' (Schmidt *et al.*, 2023).

The probes for the chromosomal localization of several satDNAs using FISH were labeled directly as well as indirectly: The probes 'pBV I' (Schmidt and Metzlaff, 1991; Kubis *et al.*, 1998) and 'beetSat05' (this manuscript; Additional file 1: Data S1) for specific centromeric satDNA families were directly labeled with DY647-dUTP (Dyomics). The probe 'pEV I' (similar to Schmidt *et al.*, 1991; accession number OY726583) marking an intercalary satDNA family (Kubis *et al.*, 1998) was directly labeled with DY415-dUTP (Dyomics). The probe 'pTS4.1' for the pericentromeric satDNA in *Patellifolia* species (Schmidt *et al.*, 1990; Schmidt and Heslop-Harrison, 1996) was indirectly labeled by PCR in the presence of digoxigenin-11-dUTP detected by antidigoxigenin-fluorescein isothiocyanate (FITC; both from Roche Diagnostics). The remaining satDNAs ('pTS5': Schmidt *et al.*, 1990, Schmidt and Heslop-Harrison, 1996; 'pHC8': Gindullis *et al.*, 2001; 'pRN1': Kubis *et al.*, 1997; beetSat08: accession number OY726584 [corresponding to 'BlSat1' from Ha, 2018]; beetSat13 [corresponding to 'ChenSat-1a']: Heitkam *et al.*, 2020; beetSat15 and beetSat17: this manuscript, accession numbers OY726585 and OY726586) were indirectly labeled by PCR in the presence of biotin-16-dUTP (Roche Diagnostics) detected by streptavidin-Cy3 (Sigma–Aldrich). The hybridization procedure was performed as described previously (Schmidt *et al.*, 1994) with a stringency of 82%. Chromosomes were

counterstained with DAPI (4',6'-diamidino-2-phenylindole; Böhringer, Mannheim) and mounted in an antifade solution (CitiFluor). Slides were examined with a fluorescence microscope (Zeiss Axioplan 2 imaging) equipped with appropriate filters. Images were acquired directly with the Applied Spectral Imaging v. 3.3 software coupled to the high-resolution CCD camera ASI BV300-20A. After separate capture for each fluorochrome, the individual images were combined computationally and processed using Adobe Photoshop CS5 software (Adobe Systems, San Jose, CA, USA). We used only contrast optimization, Gaussian and channel overlay functions affecting all pixels of the image equally.

DECLARATIONS

Ethics approval and consent to participate

The material of the KWS SAAT SE & Co. KGaA, Einbeck and IPK Gatersleben was transferred under the regulations of the standard material transfer agreement (SMTA) of the International Treaty. Plants were grown in accordance with German legislation.

Consent for publication

Not applicable.

Availability of data and materials

Illumina whole genome sequence data are available at EBI under accession numbers as indicated in Table 1. ONT reads for *B. corolliflora* and *P. procumbens* are available at EBI under the accession numbers ERS13530775 and ERS13530778. RepeatExplorer2 outputs have been made available at ZENODO (doi: 10.5281/zenodo.7821055; https://zenodo.org/record/7821055). Satellite DNA consensus sequences and the sequences used as FISH probes are available in Additional file 1: Data S1 and Data S2. Furthermore, the cloned sequences of the satDNAs beetSat02-pEV and beetSat08, as well as the newly identified satDNAs beetSat15 and beetSat17 were submitted to ENA (accession numbers OY726583-OY726586).

Competing interests

The authors declare no competing interests.

Funding

This work was supported by the German Federal Ministry of Education and Research (call „Epigenetics: Opportunities for Plant Research“, grant 031B1221A). KS was funded by Bielefeld University.

Authors' contributions

NS, BW, DH, and TH designed the study. NS selected and cultivated the plants. NS, KS, and BP performed DNA extraction. NS and JF performed flow cytometry. BP performed sequencing. NS and SB performed FISH. NS implemented the bioinformatic methodology, analyzed the data and prepared the figures and tables. NS, KS, and TH wrote the manuscript. All authors read and approved the final manuscript.

Acknowledgements

We acknowledge the KWS SAAT SE & Co. KGaA, Einbeck and the Genbank Gatersleben for providing seeds and data for the investigated accessions. Computational resources for RepeatExplorer2 analysis were provided by the ELIXIR-CZ project (LM2015047), part of the international ELIXIR infrastructure.

SUPPLEMENTARY DATA

Additional supporting information may be found in the online version of this article and consists of the following. Fig. S1: DAPI-stained chromosomes of 17 different beet accessions. Fig. S2: Correlation between repeat proportion and genome size in *Beta* and *Patellifolia* species. Fig. S3: Comparative repeat composition among *Beta* and *Patellifolia* species. Fig. S4: Arrangement of beet satDNA monomers on long reads of *B. vulgaris* (Bvul), *B. corolliflora* (Bcor), and *P. procumbens* (Ppro). Fig. S5: Comparative nucleotide alignment of the three section-specific beetSat02-pEV variants. Table S1: Genome proportion [%] of different repeat classes and superfamilies (repeat proportion) in *Beta* and *Patellifolia* species. Table S2: Genome proportion [%] of different satDNAs (and tandem repeats) across *Beta* and *Patellifolia* species. Table S3: Primer sequences for the amplification of beet-specific satellite DNAs. Data S1: beetSat nucleotide sequences derived as RepeatExplorer2 consensuses from the comparative analysis. Data S2: beetSat sequences used as FISH probes.

LITERATURE CITED

Andrews S. FastQC: a quality control tool for high throughput sequence data. 2020. http://www.bioinformatics.babraham.ac.uk/projects/fastqc.

Avila Robledillo L, Neumann P, Koblizkova A, Novak P, Vrbova I, Macas J. Extraordinary sequence diversity and promiscuity of centromeric satellites in the legume tribe *Fabeae*. Mol Biol Evol. 2020;37:2341–2356.

Barghini E, Natali L, Cossu RM, Giordani T, Pindo M, Cattonaro F, et al. The peculiar landscape of repetitive sequences in the olive (*Olea europaea* L.) genome. Genome Biol Evol. 2014;6:776–791.

Becher H, Powell RF, Brown MR, Metherell C, Pellicer J, Leitch IJ, Twyford AD. The nature of intraspecific and interspecific genome size variation in taxonomically complex eyebrights. Ann Bot. 2021;128:639–651.

Belyayev A, Josefiová J, Jandová M, Mahelka V, Krak K, Mandák B. Transposons and satellite DNA: on the origin of the major satellite DNA family in the *Chenopodium* genome. Mob DNA. 2020;11:20.

Bennetzen JL. Mechanisms of recent genome size variation in flowering plants. Ann Bot. 2005;95:127–132.

Bennetzen JL, Wang H. The contributions of transposable elements to the structure, function, and evolution of plant genomes. Annu Rev Plant Biol. 2014;65:505–530.

Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30:2114–2120.

Dechyeva D, Gindullis F, Schmidt T. Divergence of satellite DNA and interspersion of dispersed repeats in the genome of the wild beet *Beta procumbens*. Chromosome Res. 2003;11:3–21.

Dechyeva D, Schmidt T. Molecular organization of terminal repetitive DNA in *Beta* species. Chromosome Res. 2006;14:881–897.

Dodsworth S, Chase MW, Kelly LJ, Leitch IJ, Macas J, Novák P, Piednoël M, Weiss-Schneeweiss H, Leitch AR. Genomic repeat abundances contain phylogenetic signal. Syst Biol. 2015;64:112–126.

Dodsworth S, Kovarik A, Grandbastien MA, Leitch IJ, Leitch AR. Repetitive DNA dynamics and polyploidization in the genus *Nicotiana* (Solanaceae). In: Ivanov NV, Sierro N, Peitsch MC, editors. The tobacco plant genome. Compendium of Plant Genomes. Springer, Cham; 2020. p. 85–99.

893 Dohm JC, Minoche AE, Holtgräwe D, Capella-Gutiérrez S, Zakrzewski F, Tafer H, et
894 al. The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*).
895 Nature 2014;505:546–549.
896 Doležel J, Sgorbati S, Lucretti S. Comparison of three DNA fluorochromes for flow
897 cytometric estimation of nuclear DNA content in plants. Physiol Plant. 1992;85:625–
898 631.
899 Doležel J, Bartoš J, Voglmayr H, Greilhuber J. Nuclear DNA content and genome size
900 of trout and human. Cytom Part A. 2003;51:127–128.
901 Dvorak J. Evidence of genetic suppression of heterogenetic chromosome pairing in
902 polyploid species of *Solanum*, sect. *Petota*. Can J Genet Cytol. 1983;25:530–539.
903 Edgar RC. MUSCLE: a multiple sequence alignment method with reduced time and
904 space complexity. BMC Bioinformatics 2004;5:113.
905 Escudero M, Hipp AL, Waterway MJ, Valente LM. Diversification rates and
906 chromosome evolution in the most diverse angiosperm genus of the temperate zone
907 (*Carex*, Cyperaceae). Mol Phylogenet Evol. 2012;63:650–655.
908 Escudero M, Maguilla E, Loureiro J, Castro M, Castro S, Modesto L. Genome size
909 stability despite high chromosome number variation in *Carex* gr. *laevigata*. Am J Bot.
910 2015;102:233–238.
911 Finnegan DJ. Eukaryotic transposable elements and genome evolution. Trends Genet.
912 1989;5:103–107.
913 Fischer HE. Origin of the ‘Weisse Schlesische Rübe’ (white Silesian beet) and
914 resynthesis of sugar beet. Euphytica. 1989;41:75–80.
915 Frese L, Ford-Lloyd B. Taxonomy, phylogeny, and the genepool. In: Biancardi E,
916 Panella L, McGrath J, editors. *Beta maritima*. Springer, Cham; 2020. p. 121–151.
917 Funk A, Galewski P, McGrath JM. Nucleotide-binding resistance gene signatures in
918 sugar beet, insights from a new reference genome. Plant J. 2018;95:659–671.
919 Gaiero P, Vaio M, Peters SA, Schranz ME, de Jong H, Speranza PR. Comparative
920 analysis of repetitive sequences among species from the potato and the tomato clades.
921 Ann Bot. 2019;123:521–532.
922 Gantuz M, Marfil CF, Masuelli RW. Transposable elements and genome expansion in
923 cultivated and wild potato and tomato species. In: Carputo D, Aversano R, Ercolano
924 MR, editors. The wild *Solanums* genomes. Compendium of Plant Genomes. Springer,
925 Cham, 2021. p. 201–214.

926 Gao D, Schmidt T, Jung C. Molecular characterization and chromosomal distribution of
927 species-specific repetitive DNA sequences from *Beta corolliflora*, a wild relative of
928 sugar beet. Genome. 2000;43:1073–1080.
929 Garrido-Ramos MA. The genomics of plant satellite DNA. In: Ugarković Đ, editor.
930 Satellite DNAs in physiology and evolution. Progress in Molecular and Subcellular
931 Biology, vol 60. Springer, Cham. 2021. p. 103–143.
932 Ghaffari R, Cannon EK, Kanizay LB, Lawrence CJ, Dawe RK. Maize chromosomal
933 knobs are located in gene-dense areas and suppress local recombination. Chromosoma.
934 2013;122:67–75.
935 Gindullis F, Desel C, Galasso I, Schmidt T. The large-scale organization of the
936 centromeric region in *Beta* species. Genome Res. 2001;11:253–265.
937 Gong Z, Wu Y, Koblízková A, Torres GA, Wang K, Iovene M, Neumann P, Zhang W,
938 Novák P, Buell CR, Macas J, Jiang J. Repeatless and repeat-based centromeres in
939 potato: implications for centromere evolution. Plant Cell. 2012;24:3559–3574.
940 Ha BH. Structure, organization, and evolution of satellite DNAs in species of the genera
941 *Beta* and *Patellifolia*. Doctoral dissertation, Technische Universität Dresden. 2018.
942 https://nbn-resolving.org/urn:nbn:de:bsz:14-qucosa-238083.
943 He Q, Cai Z, Hu T, Liu H, Bao C, Mao W, Jin W. Repetitive sequence analysis and
944 karyotyping reveals centromere-associated DNA sequences in radish (*Raphanus sativus*
945 L.). BMC Plant Biol. 2015;15:105.
946 Heitkam T, Holtgräwe D, Dohm JC, Minoche AE, Himmelbauer H, Weisshaar B,
947 Schmidt T. Profiling of extensively diversified plant LINEs reveals distinct plant-
948 specific subclades. Plant J. 2014;79:385–397.
949 Heitkam T, Weber B, Walter I, Liedtke S, Ost C, Schmidt T. Satellite DNA landscapes
950 after allotetraploidization of quinoa (*Chenopodium quinoa*) reveal unique a and B
951 subgenomes. Plant J. 2020;103:32–52.
952 Henikoff S, Ahmad K, Malik HS. The centromere paradox: stable inheritance with
953 rapidly evolving DNA. Science. 2001;293:1098–1102.
954 Herklotz V, Kovařík A, Wissemann V, Lunerová J, Vozárová R, Buschmann S, et al.
955 Power and weakness of repetition – Evaluating the phylogenetic signal from repeatomes
956 in the family Rosaceae with two case studies from genera prone to polyploidy and
957 hybridization (*Rosa* and *Fragaria*). Front Plant Sci. 2021;12:738119.

958 Hloušková P, Mandáková T, Pouch M, Trávníček P, Lysak MA. The large genome size
959 variation in the *Hesperis* clade was shaped by the prevalent proliferation of DNA
960 repeats and rarer genome downsizing. Ann Bot. 2019;124:103–120.
961 Hohmann S, Kadereit JW, Kadereit G. Understanding Mediterranean-Californian
962 disjunctions: molecular evidence from Chenopodiaceae-Betoideae. Taxon. 2006;55:67–
963 78.
964 Kadereit G, Hohmann S, Kadereit JW. A Synopsis of Chenopodiaceae Subfam.
965 Betoideae and Notes on the Taxonomy of *Beta*. Willdenowia. 2006;Bd. 36, H. 1(Special
966 Issue: Festschrift Werner Greuter):9–19.
967 Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7:
968 improvements in performance and usability. Mol Biol Evol. 2013;30:772–780.
969 Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious
970 Basic: an integrated and extendable desktop software platform for the organization and
971 analysis of sequence data. Bioinformatics. 2012;28:1647–1649.
972 Kelly LJ, Renny-Byfield S, Pellicer J, Macas J, Novák P, Neumann P, et al. Analysis of
973 the giant genomes of *Fritillaria* (Liliaceae) indicates that a lack of DNA removal
974 characterizes extreme expansions in genome size. New Phytol. 2015;208:596–607.
975 Kowar T, Zakrzewski F, Macas J, Kobližková A, Viehoever P, Weisshaar B, Schmidt
976 T. Repeat composition of CenH3-chromatin and H3K9me2-marked heterochromatin in
977 sugar beet (*Beta vulgaris*). BMC Plant Biol. 2016;16:120.
978 Kubis S, Heslop-Harrison J, Schmidt T. A family of differentially amplified repetitive
979 DNA sequences in the genus *Beta* reveals genetic variation in *Beta vulgaris* subspecies
980 and cultivars. J Mol Evol. 1997;44:310–320.
981 Kubis S, Schmidt T, Heslop-Harrison JS. Repetitive DNA elements as a major
982 component of plant genomes. Ann Bot. 1998;82:45–55.
983 Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods.
984 2012;9:357–359.
985 Li N, Li X, Zhou J, Yu L, Li S, Zhang Y, Qin R, Gao W and Deng C. Genome-wide
986 analysis of transposable elements and satellite DNAs in *Spinacia* species to shed light
987 on their roles in sex chromosome evolution. Front Plant Sci. 2021;11:575462.
988 Lopez-Gomollon S, Müller SY, Baulcombe DC. Endogenous virus SRNA regulates
989 gene expression following genome shock in tomato hybrids. bioRxiv. 2021.
990 https://doi.org/10.1101/2021.09.20.461014.

991 Lower SS, McGurk MP, Clark AG, Barbash DA. Satellite DNA evolution: old ideas,
992 new approaches. Curr Opin Genet Dev. 2018;49:70–78.
993 Macas J, Novak P, Pellicer J, Čížková J, Koblížková A, Neumann P. In depth
994 characterization of repetitive DNA in 23 plant genomes reveals sources of genome size
995 variation in the legume tribe Fabeae. PLoS ONE 2015;10:1–23.
996 Maiwald S, Weber B, Seibt KM, Schmidt T, Heitkam T. The Cassandra retrotransposon
997 landscape in sugar beet (*Beta vulgaris*) and related Amaranthaceae: recombination and
998 re-shuffling lead to a high structural variability. Ann Bot. 2021;127:91–109.
999 Mann L, Seibt KM, Weber B, Heitkam T. ECCsplorer: a pipeline to detect
1000 extrachromosomal circular DNA (eccDNA) from next-generation sequencing data.
1001 BMC bioinform. 2022;23:1–15.
1002 McCann J, Macas J, Novák P, Stuessy TF, Villaseñor JL, Weiss-Schneeweiss H.
1003 Differential genome size and repetitive DNA evolution in diploid species of
1004 *Melampodium* sect. *Melampodium* (Asteraceae). Front Plant Sci. 2020;11:362.
1005 McClintock B. The significance of responses of the genome to challenge. Science.
1006 1984;226:792–801.
1007 McGrath JM, Funk A, Galewski P, Ou S, Townsend B, Davenport K, et al. A
1008 contiguous *de novo* genome assembly of sugar beet EL10 (*Beta vulgaris* L.). DNA Res.
1009 2023;30:dsac033.
1010 Melters DP, Bradnam KR, Young HA, Telis N, May MR, Ruby JG, et al. Comparative
1011 analysis of tandem repeats from hundreds of species reveals unique insights into
1012 centromere evolution. Genome Biol. 2013;14:1–20.
1013 Menzel G, Dechyeva D, Keller H, Lange C, Himmelbauer H, Schmidt T. Mobilization
1014 and evolutionary history of miniature inverted-repeat transposable elements (MITEs) in
1015 *Beta vulgaris* L.. Chromosome Res. 2006;14:831–844.
1016 Menzel G, Krebs C, Diez M, Holtgräwe D, Weisshaar B, Minoche AE. Survey of sugar
1017 beet (*Beta vulgaris* L.) *hAT* transposons and MITE-like *hATpin* derivatives. Plant Mol
1018 Biol. 2012;78:393–405.
1019 Michael TP. Plant genome size variation: bloating and purging DNA. Brief Funct
1020 Genomics. 2014;13:308–317.
1021 Navrátilová A, Koblizkova A, Macas J. Survey of extrachromosomal circular DNA
1022 derived from plant satellite repeats. BMC Plant Biol. 2008;8:1–13.

1023 Neumann P, Novák P, Hoštáková N, Macas J. Systematic survey of plant LTR-
1024 retrotransposons elucidates phylogenetic relationships of their polyprotein domains and
1025 provides a reference for element classification. Mob DNA. 2019;10:1–17.
1026 Novák P, Neumann P, Macas J. Graph-based clustering and characterization of
1027 repetitive sequences in next-generation sequencing data. BMC Bioinform. 2010;11:1–
1028 12.
1029 Novák P, Neumann P, Pech J, Steinhaisl J, Macas J. RepeatExplorer: a Galaxy-based
1030 web server for genome-wide characterization of eukaryotic repetitive elements from
1031 next generation sequence reads. Bioinformatics. 2013;29:792–793.
1032 Novák P, Neumann P, Macas J. Global analysis of repetitive DNA from unassembled
1033 sequence reads using RepeatExplorer2. Nat Protoc. 2020;15:3745–3776.
1034 Orozco-Arias S, Isaza G, Guyot R. Retrotransposons in plant genomes: structure,
1035 identification, and classification through bioinformatics and machine learning. Int J Mol
1036 Sci. 2019;20:3837.
1037 Panella LW, Stevanato P, Pavli O, Skaracis G. Source of useful traits. In: Biancardi E,
1038 Panella L, McGrath J, editors. *Beta maritima*. Springer, Cham; 2020. p. 167–218.
1039 Pellicer J, Fernández P, Fay MF, Michálková E, Leitch IJ. Genome size doubling arises
1040 from the differential repetitive DNA dynamics in the genus *Heloniopsis*
1041 (Melanthiaceae). Front Genet. 2021;12:726211.
1042 Piégu B, Bire S, Arensburger P, Bigot Y. A survey of transposable element
1043 classification systems - A call for a fundamental update to meet the challenge of their
1044 diversity and complexity. Mol Phylogenet Evol. 2015;86:90–109.
1045 Presting GG. Centromeric retrotransposons and centromere function. Curr Opin Genet
1046 Dev. 2018;49:79–84.
1047 Rockinger A, Carvalho FA, Renner SS. Chromosome number reduction in the sister
1048 clade of *Carica papaya* with concomitant genome size doubling. Am J Bot.
1049 2016;103:1082–1088.
1050 Romeiras MM, Vieira A, Silva DN, Moura M, Santos-Guerra A, Batista D, et al.
1051 Evolutionary and biogeographic insights on the Macaronesian *Beta-Patellifolia* species
1052 (Amaranthaceae) from a time-scaled molecular phylogeny. PLoS ONE.
1053 2016;11:e0152456.
1054 Šatović-Vukšić E, Plohl M. Satellite DNAs – From localized to highly dispersed
1055 genome components. Genes. 2023;14:742.

1056 Schmidt N, Seibt KM, Weber B, Schwarzacher T, Schmidt T, Heitkam T. Broken,
1057 silent, and in hiding: tamed endogenous pararetroviruses escape elimination from the
1058 genome of sugar beet (*Beta vulgaris*). Ann Bot. 2021;128:281–299.
1059 Schmidt N, Weber B, Klekar J, Liedtke S, Breitenbach S, Heitkam T. Preparation of
1060 mitotic chromosomes using the dropping technique. In: Heitkam T, Garcia S, editors.
1061 Walker J, series editor. Methods in Molecular Biology: Plant Cytogenetics and
1062 Cytogenomics. Humana New York, US. 2023. p. 151–162.
1063 Schmidt M, Hense S, Minoche AE, Dohm JC, Himmelbauer H, Schmidt T, Zakrzewski
1064 F. Cytosine methylation of an ancient satellite family in the wild beet *Beta procumbens*.
1065 Cytogenet Genome Res. 2014;143:157–167.
1066 Schmidt T, Junghans H, Metzlaff M. Construction of *Beta procumbens*-specific DNA
1067 probes and their application for the screening of *B. vulgaris* x *B. procumbens* (2n = 19)
1068 addition lines. Theoret Appl Genetics. 1990;79:177–181.
1069 Schmidt T, Jung C, Metzlaff M. Distribution and evolution of two satellite DNAs in the
1070 genus *Beta*. Theoret Appl Genetics. 1991;82:793–799.
1071 Schmidt T, Metzlaff M. Cloning and characterization of a *Beta vulgaris* satellite DNA
1072 family. Gene. 1991;101:247–250.
1073 Schmidt T, Heslop-Harrison JS. Variability and evolution of highly repeated DNA
1074 sequences in the genus *Beta*. Genome. 1993;36:1074–1079.
1075 Schmidt T, Schwarzacher T, Heslop-Harrison JS. Physical mapping of rRNA genes by
1076 fluorescent *in-situ* hybridization and structural analysis of 5S rRNA genes and
1077 intergenic spacer sequences in sugar beet (*Beta vulgaris*). Theoret Appl Genetics.
1078 1994;88:629–636.
1079 Schmidt T, Heslop-Harrison JS. High-resolution mapping of repetitive DNA by *in situ*
1080 hybridization: molecular and chromosomal features of prominent dispersed and
1081 discretely localized DNA families from the wild beet species *Beta procumbens*. Plant
1082 Mol Biol. 1996;30:1099–1113.
1083 Schwichtenberg K, Wenke T, Zakrzewski F, Seibt KM, Minoche A, Dohm JC, et al.
1084 Diversification, evolution and methylation of short interspersed nuclear element
1085 families in sugar beet and related Amaranthaceae species. Plant J. 2016;85:229–244.
1086 Seibt KM, Schmidt T, Heitkam T. FlexiDot: highly customizable, ambiguity-aware
1087 dotplots for visual sequence analyses. Bioinformatics. 2018;34:3575–3577.

1088 Shang WH, Hori T, Toyoda A, Kato J, Popendorf K, Sakakibara Y, et al. Chickens
1089 possess centromeres with both extended tandem repeats and short non-tandem-repetitive
1090 sequences. Genome Res. 2010;20:1219–28.
1091 Sielemann K, Pucker B, Schmidt N, Viehöver P, Weisshaar B, Heitkam T, Holtgräwe
1092 D. Complete pan-plastome sequences enable high resolution phylogenetic classification
1093 of sugar beet and closely related crop wild relatives. BMC Genomics 2022;23:113.
1094 Sielemann K, Schmidt N, Guzik J, Kalina N, Pucker B, Viehöver P, et al. Pangenome of
1095 cultivated beet and crop wild relatives reveals parental relationships of a tetraploid wild
1096 beet. bioRxiv. 2023a; https://doi.org/10.1101/2023.06.28.546919.
1097 Sielemann K, Pucker B, Orsini E, Elashry A, Schulte L, Viehoever P, et al. Genomic
1098 characterization of a nematode tolerance locus in sugar beet. bioRxiv. 2023b;
1099 https://doi.org/10.1101/2023.06.22.546034.
1100 Vitales D, Álvarez I, Garcia S, Hidalgo O, Feliner GN, Pellicer J, et al. Genome size
1101 variation at constant chromosome number is not correlated with repetitive DNA
1102 dynamism in *Anacyclus* (Asteraceae). Ann Bot. 2020a;125:611–623.
1103 Vitales D, Garcia S, Dodsworth S. Reconstructing phylogenetic relationships based on
1104 repeat sequence similarities. Mol Phylogenet Evol. 2020b;147:106766.
1105 Wascher FL, Stralis-Pavese N, McGrath JM, Schulz B, Himmelbauer H, Dohm JC.
1106 Genomic distances reveal relationships of wild and cultivated beets. Nat Commun.
1107 2022;13:2021.
1108 Weber B, Schmidt T. Nested Ty3-*gypsy* retrotransposons of a single *Beta procumbens*
1109 centromere contain a putative chromodomain. Chromosome Res. 2009;17:379–396.
1110 Weber B, Wenke T, Frömmel U, Schmidt T, Heitkam T. The Ty1-*copia* families
1111 SALIRE and Cotzilla populating the *Beta vulgaris* genome show remarkable
1112 differences in abundance, chromosomal distribution, and age. Chromosome Res.
1113 2010;18:247–263.
1114 Weber B, Heitkam T, Holtgräwe D, Weisshaar B, Minoche AE, Dohm JC, et al. Highly
1115 diverse chromoviruses of *Beta vulgaris* are classified by chromodomains and
1116 chromosomal integration. Mobile DNA. 2013;4:1–16.
1117 Wells JN, Feschotte C. A field guide to eukaryotic transposable elements. Annu Rev
1118 Genet. 2020;54:539–561.
1119 Wendel JF. The wondrous cycles of polyploidy in plants. Am J Bot. 2015;102:1753–
1120 1756.

1121 Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, et al. A unified
1122 classification system for eukaryotic transposable elements. Nat Rev Genet. 2007;8:973–
1123 982.
1124 Wollrab C, Heitkam T, Holtgräwe D, Weisshaar B, Minoche AE, Dohm JC, et al.
1125 Evolutionary reshuffling in the Errantivirus lineage Elbe within the *Beta vulgaris*
1126 genome. Plant J. 2012;72:636–651.
1127 Yogeeswaran K, Frary A, York TL, Amenta A, Lesser AH, Nasrallah JB. Comparative
1128 genome analyses of *Arabidopsis* spp.: inferring chromosomal rearrangement events in
1129 the evolutionary history of *A. thaliana*. Genome Res. 2005;15:505–515.
1130 Zakrzewski F, Wenke T, Holtgräwe D, Weisshaar B, Schmidt T. Analysis of a c_0t*-1*
1131 library enables the targeted identification of minisatellite and satellite families in *Beta*
1132 *vulgaris*. BMC Plant Biol. 2010;10:1–14.
1133 Zakrzewski F, Weisshaar B, Fuchs J, Bannack E, Minoche AE, Dohm JC, et al.
1134 Epigenetic profiling of heterochromatic satellite DNA. Chromosoma. 2011;120:409–
1135 422.
1136 Zakrzewski F, Weber B, Schmidt T. A molecular cytogenetic analysis of the structure,
1137 evolution, and epigenetic modifications of major DNA sequences in centromeres of
1138 *Beta* species. In: Jiang J, Birchler JA, editors. Plant centromere biology. Oxford: John
1139 Wiley & Sons. 2013. p. 39–55.
1140 Zakrzewski F, Schubert V, Viehoever P, Minoche AE, Dohm JC, Himmelbauer H, et al.
1141 The CHH motif in sugar beet satellite DNA: a modulator for cytosine methylation. Plant
1142 J. 2014;78:937–950.
1143 Zedek F, Šmerda J, Šmarda P, Bureš P. Correlated evolution of LTR retrotransposons
1144 and genome size in the genus *Eleocharis*. BMC Plant Biol. 2010;10:1–10.
1145
1146

TABLES

Table 1: Sampled beet and wild beet species, including five-letter code, accession details, ploidy level / chromosome number, and genome size. Gene pools are marked according to Frese and Ford-Lloyd, 2020.

beet gene pools	species	code	accession	ENA accession no.	chromosome configuration	genome size 1C [Mbp]
primary	*Beta vulgaris* subsp. *vulgaris*	Bvul1	KWS2320	ERR6110425	2n=2x=18	712
	Beta vulgaris subsp. *adanensis*	Bada1	BETA 1233	ERR6110427	2n=2x=18	709
	Beta vulgaris subsp. *maritima*	Bmar1	BETA 1101	ERR6110429	2n=2x=18	703
	Beta vulgaris subsp. *maritima*	Bmar2	BETA 2322	ERR6110431	2n=2x=18	694
	Beta patula	Bptu1	BETA 548	ERR6110433	2n=2x=18	706
	Beta macrocarpa	Bmca1	BETA 881	ERR6110435	2n=4x=36	1370
secondary	*Beta lomatogona*	Blom1	BETA 674	ERR6110449	2n=2x=18	929
	Beta macrorhiza	Bmrh1	BETA 830	ERR6110451	2n=2x=18	936
	Beta macrorhiza	Bmrh2	BETA 576	ERR6110453	2n=2x=18	946
	Beta corolliflora	Bcor1	BETA 408	ERR6110441	2n=4x=36	2027
	Beta intermedia	Bint1	BETA 431	ERR6110455	2n=5x=45	2468
	Beta intermedia	Bint2	BETA 923	ERR6110457	2n=5x=45	2471
	Beta nana	Bnan1	BETA 546	ERR6110459	2n=2x=18	766
	Beta nana	Bnan2	BETA 570	ERR6110461	2n=2x=18	754
tertiary	*Patellifolia procumbens*	Ppro1	BETA 951	ERR6110463	2n=2x=18	712
	Patellifolia webbiana	Pweb1	BETA 526	ERR6110465	2n=2x=18	723
	Patellifolia patellaris	Ppat1	BETA 892	ERR6110467	2n=4x=36	1502

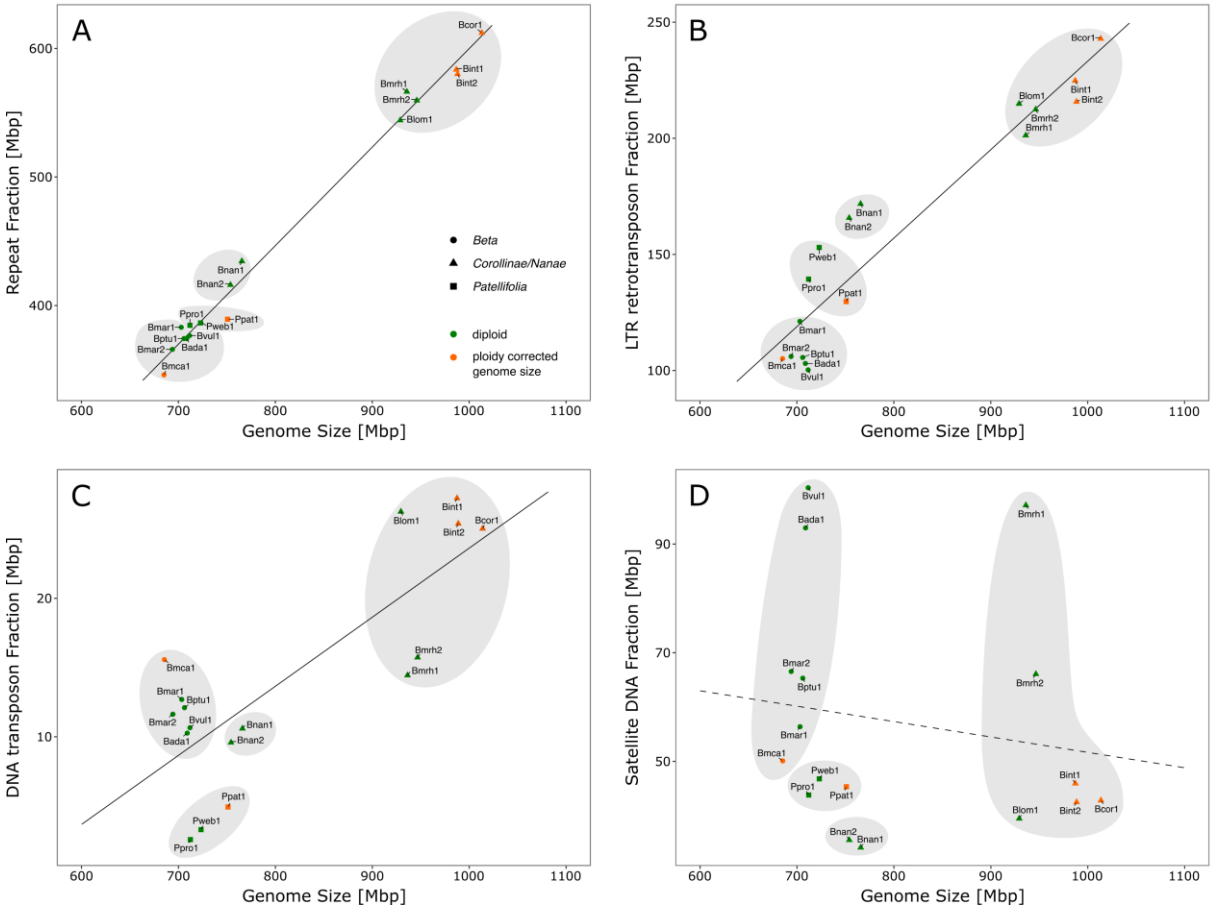


Fig. 1: Correlation between repeat fraction and genome size in *Beta* and *Patellifolia* species. For better comparability with natural diploids (green), polyploid species were downsampled as artificial diploids (orange). Shapes and shades indicate the different beet sections and genera, respectively. Samples are abbreviated according to Table 1. (A) Positive correlation between the total repeat fraction and genome size. (B) Positive correlation between LTR retrotransposon fraction and genome size. (C) Positive correlation between DNA transposon fraction and genome size. (D) The satellite DNA fraction and genome size tend to relate negatively.

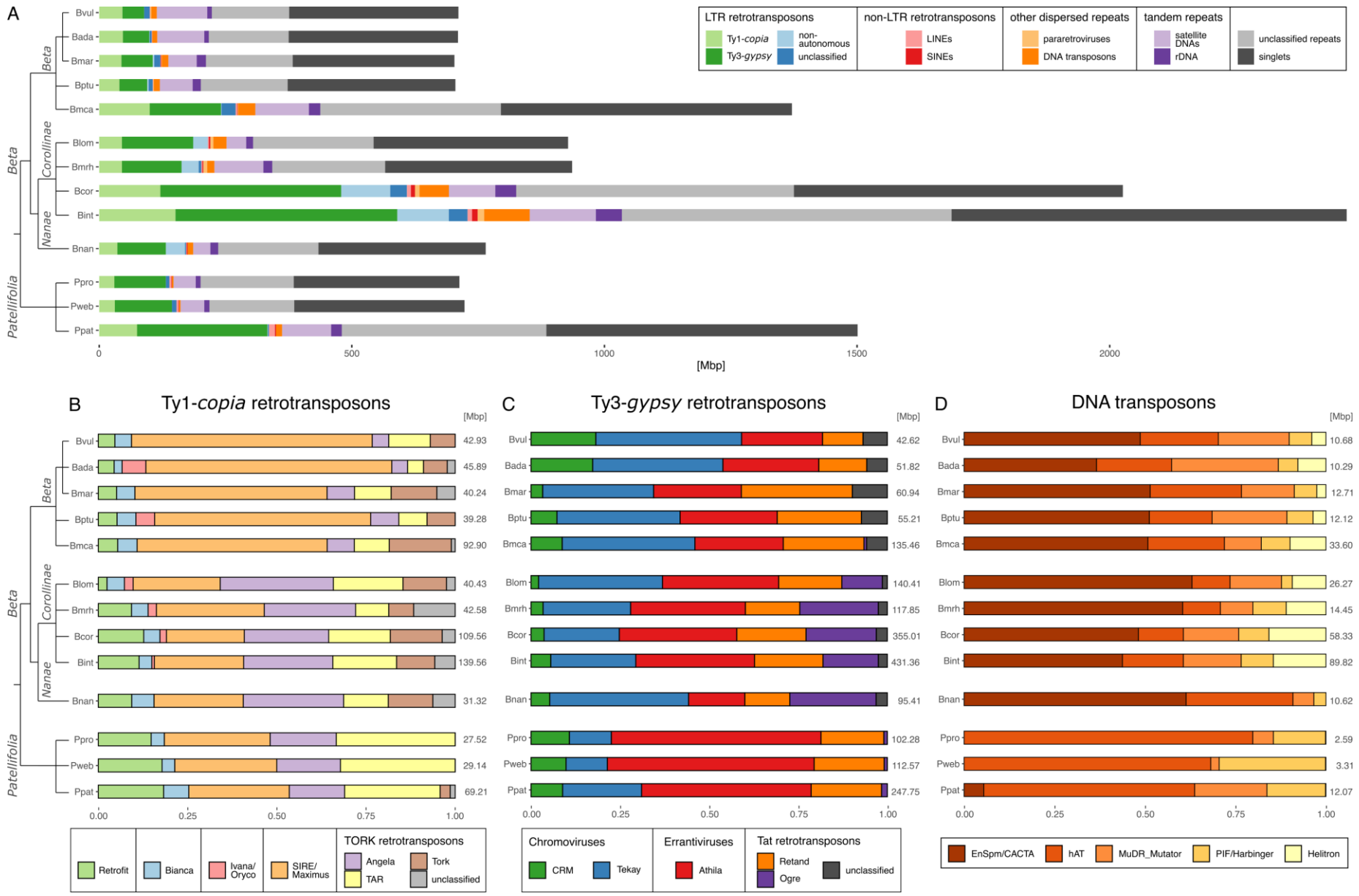

A
Bvul
Bada
Bmar
Bptu
Bmca
Blom
Bmrh
Bcor
Bint
Bnan
Ppro
Pweb
Ppat
Beta
Corollinae
Nanae
Patellifolia
LTR retrotransposons
Ty1-copia
Ty3-gypsy
non-autonomous
unclassified
non-LTR retrotransposons
LINEs
SINEs
other dispersed repeats
pararetroviruses
DNA transposons
tandem repeats
satellite DNAs
rDNA
unclassified repeats
singlets
0
500
1000
1500
2000
[Mbp]
B
Ty1-copia retrotransposons
[Mbp]
42.93
45.89
40.24
39.28
92.90
40.43
42.58
109.56
139.56
31.32
27.52
29.14
69.21
0.00
0.25
0.50
0.75
1.00
Retrofit
Bianca
Ivana/ Oryco
SIRE/ Maximus
TORK retrotransposons
Angela
Tork
TAR
unclassified
C
Ty3-gypsy retrotransposons
[Mbp]
42.62
51.82
60.94
55.21
135.46
140.41
117.85
355.01
431.36
95.41
102.28
112.57
247.75
Chromoviruses
CRM
Tekay
Errantiviruses
Athila
Tat retrotransposons
Retand
Ogre
unclassified
D
DNA transposons
[Mbp]
10.68
10.29
12.71
12.12
33.60
26.27
14.45
58.33
89.82
10.62
2.59
3.31
12.07
EnSpm/CACTA
hAT
MuDR_Mutator
PIF/Harbinger
Helitron

Fig. 2: Repeat composition of several *Beta* and *Patellifolia* species. The analyzed samples and abbreviations accord to accession 1 of the respective species (see Table 1) as the repeat compositions of different accessions from the same species are highly similar (see Additional file 1: Table S1). (A) Proportion of major repeat classes shown in relation to the total genome size. (B) Composition of the Ty1-*copia* LTR retrotransposon fraction. (C) Composition of the Ty3-*gypsy* LTR retrotransposon fraction. (D) Composition of the DNA transposon fraction.

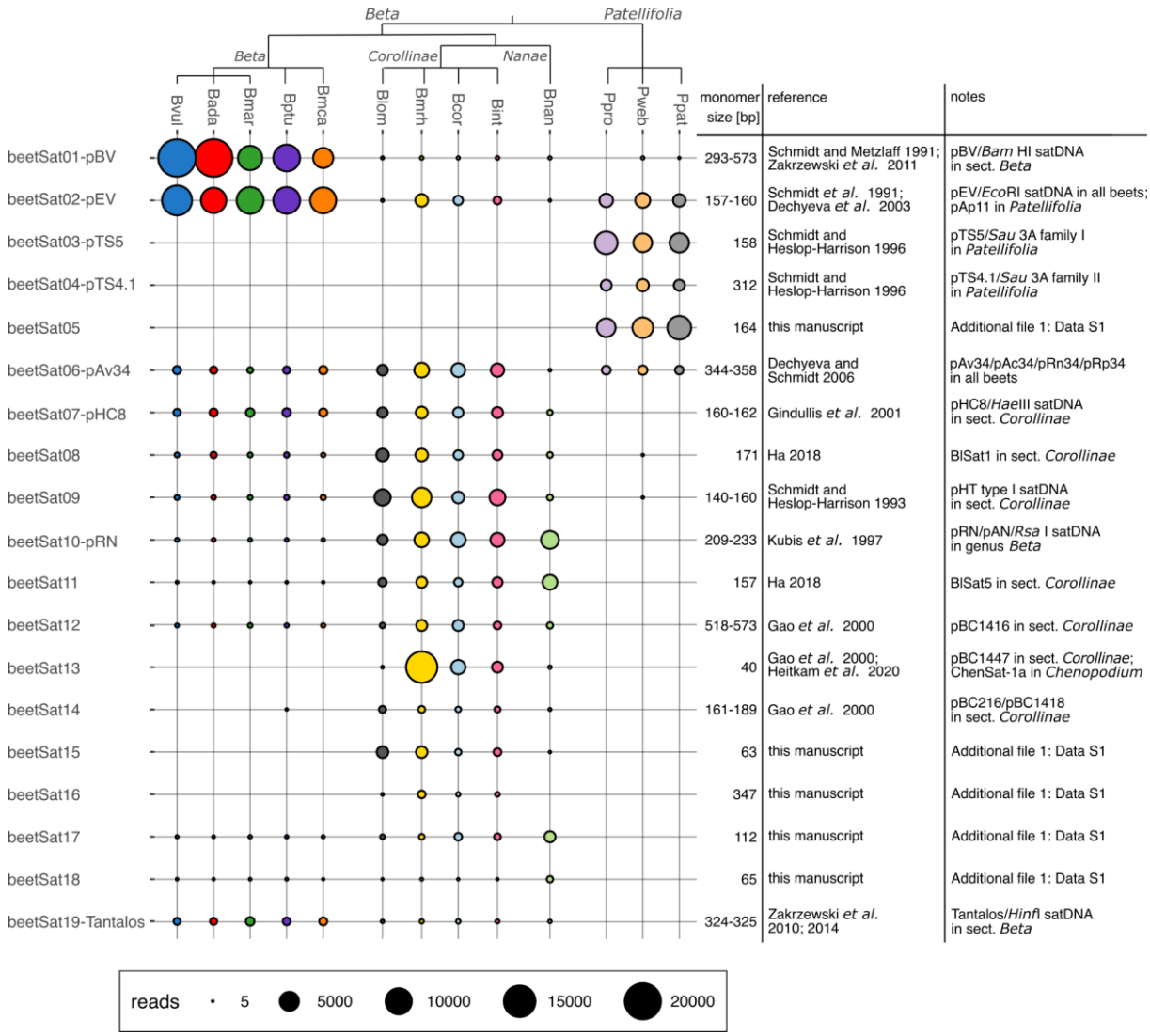


Fig. 3: Quantification of satDNAs within different *Beta* and *Patellifolia* genomes. The relative satDNA abundance is indicated by the size of the circles, which corresponds to the number of short reads within the respective satDNA cluster. The monomer sizes as well as relevant references are provided. Further previously described 'satDNAs' and minisatellites (i.e. Dione/*Fok*I-satellite and Niobe/*Alu*I-satellite: Zakrzewski et al. 2010, 2014; BvMSats: Zakrzewski et al. 2010; BvuSats: Li et al. 2021) were not included as we found that these repeats are not arranged in long arrays and therefore we consider them tandem repeats rather than satDNA. The monomer sequences of newly identified satDNAs can be found in Additional file 1: Data S1.

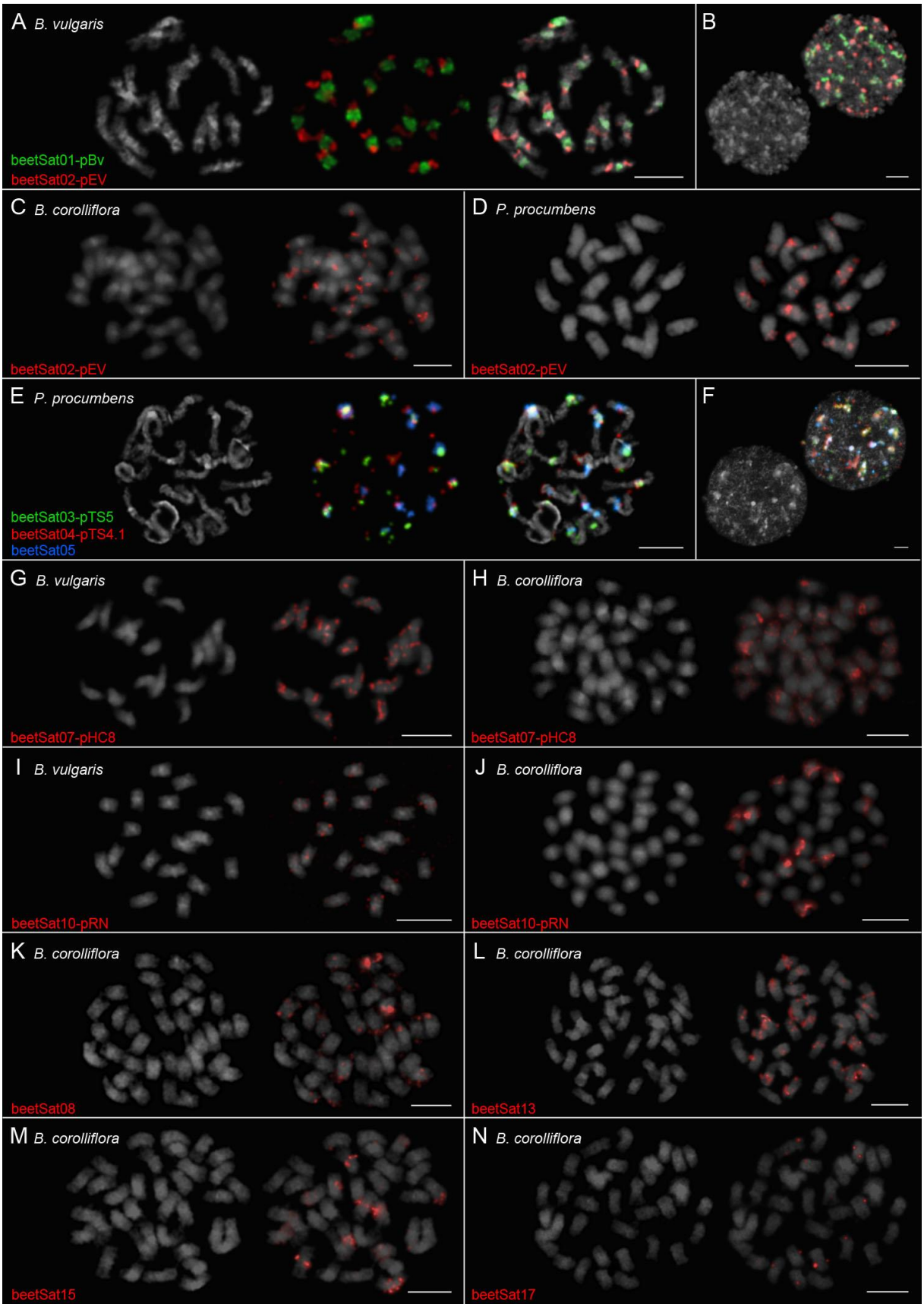
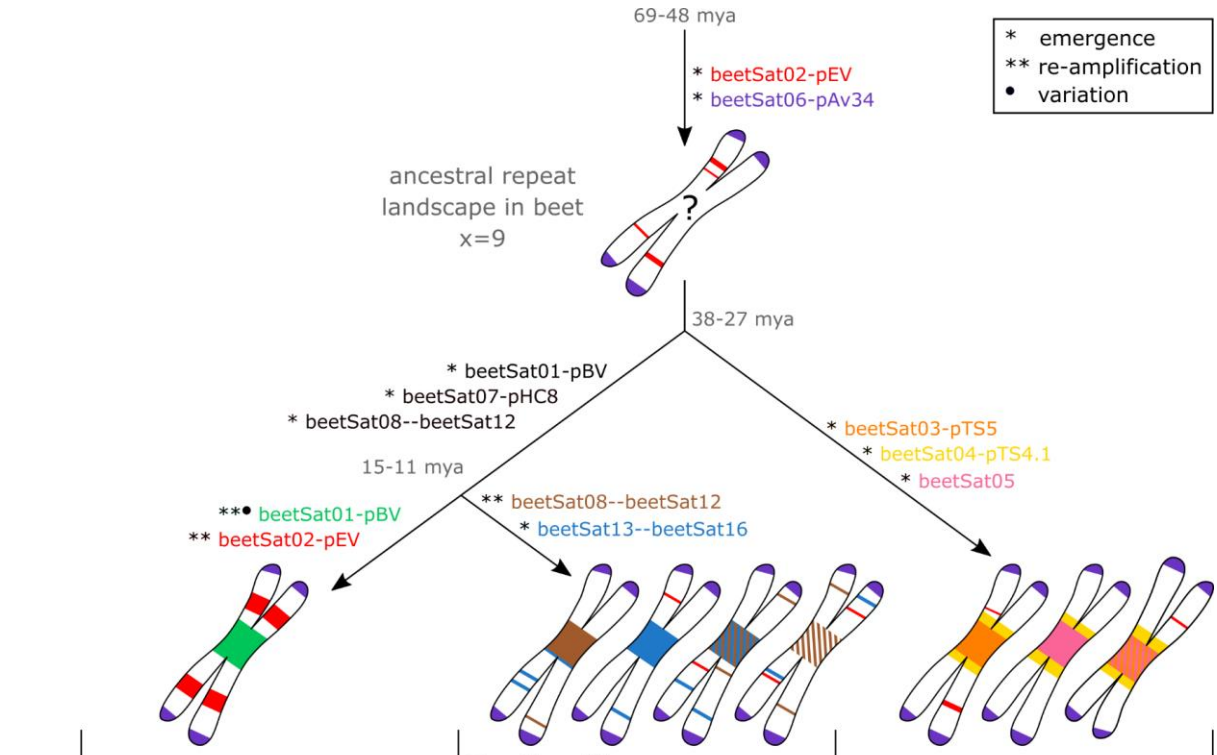


Fig. 4: Multicolor fluorescent *in situ* hybridization to chromosome spreads of *B. vulgaris*, *B. corolliflora*, and *P. procumbens*. DAPI-stained mitotic chromosomes and interphase nuclei are shown in gray. (A) The centromeric and intercalary localization of the satDNAs beetSat01-pBV (green signals) and beetSat02-pEV (red signals), respectively, become visible on prometaphase chromosomes

of *B. vulgaris*. (B) The interphase nucleus of *B. vulgaris* shows that beetSat01-pBV (green signals) and beetSat02-pEV (red signals) are located in heterochromatic regions. (C, D) The satDNA beetSat02-pEV can also be found on almost all chromosomes of *B. corolliflora* (C) and *P. procumbens* (D) in intercalary as well as distal regions, forming large as well as small arrays. Whereas the signals are mostly restricted on one chromosome arm each in *B. corolliflora* (C), they can be often found on both chromosome arms in *P. procumbens* (D). (E) The centromeric and pericentromeric localization of the satDNAs beetSat03-pTS5 (green signals) and beetSat04-pTS4.1 (red signals), respectively, become visible on prometaphase chromosomes of *P. procumbens*. Most centromeres (14 out of 18) also show large beetSat05 arrays (blue signals); in particular those that are not constituted by beetSat03-pTS5. (F) The interphase nucleus of *P. procumbens* shows that beetSat03-pTS5 (green signals), beetSat04-pTS4.1 (red signals), and beetSat05 (blue signals) are located in heterochromatic regions. (G, H) The satDNA beetSat07-pHC8 was detected on almost all arms of all chromosomes of *B. vulgaris* (G) and *B. corolliflora* (H): Signals were present in intercalary as well as distal regions, forming large as well as small arrays. Moreover, two centromeric beetSat07-pHC8 signals were detected in both species. (I, J) Hybridization sites for beetSat10-pRN are few in *B. vulgaris* (I), whereas large signal clusters were detected on most chromosomes of *B. corolliflora* (J). (I) The low beetSat10-pRN abundance in *B. vulgaris* is revealed by faint signals mostly on both arms of the respective chromosome. (J) The major hybridization sites of the beetSat10-pRN probe on *B. corolliflora* chromosomes are associated with the centromeres. Intercalary hybridization sites can be found mostly on one arm of further chromosomes. (K) The centromere of four *B. corolliflora* chromosomes showed signals (two major, two minor) after hybridization with the beetSat08 probe, whereas almost all other chromosomes showed intercalary and distal signals. (L) BeetSat13 can also be found at the centromeres of at least seventeen *B. corolliflora* chromosomes (ten major and seven minor signals). Eight further chromosomes hybridized in intercalary regions. (M) The beetSat15 hybridization resulted in at least six *B. corolliflora* chromosomes with signals at the centromeres and again eight chromosomes with intercalary signals. (N) BeetSat17 signals were detected on twelve *B. corolliflora* chromosomes with at least two of them being near the centromere. Information on probe labeling and detection can be found in the methods section. Scale bars = 5 µm.



	Beta	*Corollinae/Nanae*	*Patellifolia*
today's satDNA composition	beetSat01-pBV beetSat02-pEV beetSat06-pAv34 beetSat07-pHC8 beetSat08--beetSat12	beetSat01-pBV beetSat02-pEV beetSat06-pAc34 beetSat07-pHC8 beetSat08--beetSat16	beetSat02-pEV beetSat03-pTS5 beetSat04-pTS4.1 beetSat05 beetSat06-pRp34
impact on chromosomes	emergence of similar chromosome structure across all 9 chromosomes	emergence of patchwork chromosomes	emergence of 9 chromosomes with conceptually similar structure but three variants
amplification and homogenization	few satDNAs are strongly amplified and homogenized	many satDNAs emerged that are moderately amplified; homogenization incomplete: no satDNA reaches high amplification and overall satDNA abundance remains low	few satDNAs are strongly amplified and homogenized
biology	the number of species within the genus is low and their distribution is restricted to limited environments (except for the widespread *B. maritima*, the progenitor of *B. vulgaris*)	relatively many species within the genus (some probably already extinct) with usually wide distribution areas; frequent hybridization and polyploidy	very few species within the genus, yet the widest distribution among the beets

Fig. 5: Scenario for the evolution of satDNAs during the history of beet speciation. As the satDNAs beetSat02-pEV and beetSat06 are also present in the spinach/quinoa outgroup (Dechyeva and Schmidt, 2006; Schmidt *et al.*, 2014), we assume an emergence of these satDNAs even before the split of the beet crown. All other identified satDNAs seem to have appeared earliest in the direct beet ancestor with the subsequent emergence of genus- and section-specific satDNAs and different amplification and differentiation patterns. The age of the respective beet taxa has been estimated in million years ago (mya) by Hohmann *et al.* (2006).