Satellite DNA landscapes after allotetraploidization of quinoa (Chenopodium quinoa) reveal unique A and B subgenomes

Tony Heitkam1,* 1Institute of Botany, Technische Universität Dresden, 01069 Dresden, Germany, and 2Institute of Biology, Martin-Luther-Universität Halle-Wittenberg, 06120 Halle (Saale), Germany

SUMMARY

If two related plant species hybridize, their genomes may be combined and duplicated within a single nucleus, thereby forming an allotetraploid. How the emerging plant balances two co-evolved genomes is still a matter of ongoing research. Here, we focus on satellite DNA (satDNA), the fastest turn-over sequence class in eukaryotes, aiming to trace its emergence, amplification, and loss during plant speciation and allopolyploidization. As a model, we used Chenopodium quinoa Willd. (quinoa), an allopolyploid crop with 2n = 4x = 36 chromosomes. Quinoa originated by hybridization of an unknown female American Chenopodium diploid (AA genome) with an unknown male Old World diploid species (BB genome), dating back 3.3–6.3 million years. Applying short read clustering to quinoa (AABB), C. pallidicaule (AA), and C. suecicum (BB) whole genome shotgun sequences, we classified their repetitive fractions, and identified and characterized seven satDNA families, together with the 5S rDNA model repeat. We show unequal satDNA amplification (two families) and exclusive occurrence (four families) in the AA and BB diploids by read mapping as well as Southern, genomic, and fluorescent in situ hybridization. Whereas the satDNA distributions support C. suecicum as possible parental species, we were able to exclude C. pallidicaule as progenitor due to unique repeat profiles. Using quinoa long reads and scaffolds, we detected only limited evidence of intergenomic homogenization of satDNA after allopolyploidization, but were able to exclude dispersal of 5S rRNA genes between subgenomes. Our results exemplify the complex route of tandem repeat evolution through Chenopodium speciation and allopolyploidization, and may provide sequence targets for the identification of quinoa’s progenitors.

Keywords: Chenopodium quinoa, polyploidy, tandem repeat, satellite DNA, 5S rDNA, fluorescent in situ hybridization, single molecule real-time reads.

INTRODUCTION

Quinoa (Chenopodium quinoa Willd.) is an allotetraploid crop, domesticated in the South-American Andes for at least 8000 years (Dillehay et al., 2007). Due to the ability to grow on marginal soils in short vegetation periods, the high tolerance for abiotic stresses such as cold and UV radiation, and the high nutritional value of its grains, quinoa is ranked as a ‘high potential’ crop with priority for sustainable agriculture (www.fao.org, Zurita-Silva et al., 2014). Having an allotetraploid origin with 2n = 4x = 36 chromosomes, quinoa was derived from hybridization of a female American Chenopodium diploid (A genome) with a male Old World diploid (B genome), dating back 3.3–6.3 million years (Kolano et al., 2012; Štorchová et al., 2015; Walsh et al., 2015; Jarvis et al., 2017; Maughan et al., 2019). Potential diploid progenitors with B genomes have c. 30% larger genomes than A genome species (c. 0.6 and 0.9 Gb; Kolano et al., 2016; Mandák et al., 2016). Thus, for the quinoa genome size of 1.45–1.5 Gb (Palomino et al., 2008), additivity of A and B genomes without genome upsizing or downsizing was postulated (Kolano et al., 2016). In ongoing sequencing efforts, three independent quinoa reference genomes have been generated (Yasui et al., 2016; Jarvis et al., 2017; Zou et al., 2017).

Phylogenetically, quinoa is a member of the Amaranthaceae (formerly Chenopodiaceae), which encompasses many crops. Some of these have available genome sequences, for example Beta vulgaris (sugar beet, Dohm et al., 2014), Spinacia oleracea (spinach, Xu et al., 2017), and Amaranthus hypochondriacus (amaranth, Clouse et al., 2016). Quinoa belongs to the subfamily Chenopodioidae,
which has recently been revised and split into multiple smaller genera such as \textit{Chenopodiumastrum}, \textit{Oxybasis}, \textit{Lipandra}, and \textit{Dysphania}, in addition to the previously recognized genera \textit{Chenopodium}, \textit{Atriplex}, \textit{Blitum}, and \textit{Spinacia} (Fuentes-Bazan et al., 2012a; Fuentes-Bazan et al., 2012b).

To assign specific chromosomes of hybrids or polyploids to their respective parental genomes as well as to physically map DNA sequences along chromosomes, fluorescent in situ hybridization (FISH) and genomic in situ hybridization (GISH) are valuable technologies (Schwarzacher et al., 1989; D’Hont, 2005; Chester et al., 2010; Mandáková et al., 2013; Jiang, 2019). In many \textit{Chenopodium} species, the 5S and 18S–5.8S–25S rRNA genes have already been sequenced and localized along chromosomes, including potential A-genome and B-genome progenitors of quinoa (Maughan et al., 2006; Kolano et al., 2012; Kolano et al., 2016; Kolano et al., 2019).

Fast-evolving tandem repeats such as satellite DNAs (satDNAs) can serve as chromosomal landmarks and are often specific for individual chromosomes or subgenomes (Shapiro and von Sternberg, 2005; Heslop-Harrison and Schwarzacher, 2011; Schmidt et al., 2019). SatDNAs are non-translated and tandemly organized repeated units, often uniform within large chromosomal arrays. Maintenance of genome integrity, epigenetic regulation of gene expression, and centromere formation are among their main functional roles (Melters et al., 2013; Zakrzewski et al., 2013; Zhang et al., 2013; Jagannathan et al., 2018).

On an evolutionary timescale, tandem repeats change rapidly and are among the first sequence classes to diversify in emerging species (Charlesworth et al., 1994; Oliver et al., 2013; Zhang et al., 2015; McCann et al., 2018; Bracewell et al., 2019). Accumulation of mutations (single nucleotide changes, indels) leads to the emergence of new satDNA variants, which may spread and displace existing variants and often form homogenized arrays (Plohl et al., 2012; Garrido-Ramos, 2015). Allopolyploidization of related species combines diverged repetitive DNA families (Koukalova et al., 2010; Vicent and Casacuberta, 2017), with a range of largely unpredictable consequences: in the new allopolyploid, repeats may be redistributed, reduced, replaced, newly combined, or selectively amplified. In quinoa, already some repetitive sequences have been identified, including also a satDNA family (Kolano et al., 2011; Orzechowska et al., 2018). However, genome-wide satDNA profiles of quinoa and potential parental species are still needed to deduce the effects of speciation and allopolyploidization on \textit{Chenopodium} repeat evolution.

Here, we analyzed the influence of speciation and polyploidization on the tandem repeat landscape within \textit{Chenopodium}. Next-generation genome sequences of allotetraploid \textit{C. quinoa} and two putative diploid progenitors, \textit{C. pallidicaule} (A genome) and \textit{C. suecicum} (B genome) were subjected to comparative read clustering. We identified the major tandem repeats, including satDNAs and 5S rRNA genes, and examined their amplification in A and B genomes. Their higher-order organization and the interlocus homogenization of A- and B-specific variants was investigated using \textit{C. quinoa} long single molecule real-time (SMRT) reads. Combining GISH and FISH, we showed that the satellite DNAs target A-specific and B-specific chromosomes in \textit{C. quinoa} genomes and reveal a surprisingly low rate of intersubgenome dispersion.

**RESULTS**

**Identification of tandem repeats in \textit{Chenopodium} genomes**

In order to measure the contributions of the A and B subgenomes on the \textit{C. quinoa} repeat composition, we compared the repeat fractions of \textit{C. pallidicaule} (AA diploid), \textit{C. suecicum} (BB diploid), and \textit{C. quinoa} (AABB tetraploid) by three-way read clustering with equal amounts of shot gun Illumina reads. We used c. 127.8 Mb of each genome as input for RepeatExplorer, enabling a representative read cluster analysis (Novák et al., 2013; Goubert et al., 2015; Weiss-Schnieweiss et al., 2015).

Based on the cluster data we estimate the fraction of repetitive DNA to be 76.9% for \textit{C. pallidicaule}, 78.1% for \textit{C. suecicum}, and 73.2% for \textit{C. quinoa}. However, as diverged repeats escape the clustering threshold of 90%, the real repeat content is likely to be higher. To quantify highly repetitive sequences, we comparatively analyzed the largest 150 clusters and, if possible, assigned the underlying sequences to Ty1-copia and Ty3-gypsy long terminal repeat (LTR) retrotransposons, pararetroviruses, long interspersed nuclear elements (LINEs), DNA transposons, rDNA, satDNA, and organellar DNA (Figure 1).

Based on star-like or ring-like cluster shapes and tandem arrangements within the consensus sequence, we selected eight clusters harbouring seven \textit{Chenopodium} satDNAs for further analysis. The sequences were grouped into families and subfamilies according to their sequence similarity, and designated ChenSat-1a, ChenSat-1b, ChenSat-2a, ChenSat-2b, ChenSat-2c, ChenSat-2d, and ChenSat-2e. Gaps in the bar plots indicated the absence from a genome, such as seen for the satDNA families ChenSat-1b, ChenSat-2a, and ChenSat-2c (Figure 1). The number of reads indicates a satDNA fraction of 2.4%, 1.5% and 5.4% for \textit{C. pallidicaule}, \textit{C. suecicum}, and \textit{C. quinoa} relative to their overall repeat fractions. As read cluster abundance can misrepresent the real satDNA contribution (Novák et al., 2010; Ruiz-Ruano et al., 2018), we analyzed the graphs to uncover clusters with masked satDNA (Figure S1, see explanation in Appendix S1). As a result, we directly inferred higher amplification in either A genomes (for ChenSat-2a, ChenSat-2b, and ChenSat-2e) or B genomes (for ChenSat-1b, ChenSat-2c, and ChenSat-2d).
Assignment of Chenopodium satDNA to major families

To analyze the identified tandem repeats, we derived the monomer consensus sequences by independent and iterative mapping of sequence reads from each of the three genomes. This resulted in subgenome-specific reference sequences for satellite monomers and 5S rRNA genes including the spacer (Data S1). The seven analyzed satDNA consensus sequences from the three Chenopodium genomes fall into only two satDNA families, ChenSat-1 and ChenSat-2, marked by similar monomer lengths, conserved sequence stretches, and moderate pairwise sequence identities between 40 and 60%.

For ChenSat-1a, the C. quinoa, C. pallidicaule and C. suecicum consensus monomers are highly similar with >95% sequence identity. Its 40 bp monomers are marked by a very low G/C content (26–27%) and make up a high genome proportion in all analyzed genomes (0.6–3.7%, Table 1). ChenSat-1a is the major satellite of Chenopodium genomes (Figure 2a) and has already been cloned from the C. quinoa genome (Kolano et al., 2011, accession HM641822). Surprisingly, ChenSat-1a also has a 87% similarity to the 40 bp satellite pBC1447 from the distantly related wild beet Beta corolliflora (Gao et al., 2000, accession AJ288880), an Amaranthaceae species from the Betoideae subfamily.

With 48 bp, ChenSat-1b monomers are slightly longer than ChenSat-1a (Figure 2a). From the satDNAs studied here, ChenSat-1b has the highest G/C content with 53%, contributing many cytosine targets for potential DNA methylation (Table 1). It is only present in reads from B (sub)genome species, that is C. suecicum and C. quinoa, but absent from the A genome species C. pallidicaule (Figure 3). The short satDNA monomers of ChenSat-1a and ChenSat-1b contain stretches of sequence similarity and have an overall identity of 60%, indicating an evolutionary relationship.

Five of the Chenopodium satDNAs analyzed here are diverged subfamilies forming the ChenSat-2 family. We hypothesize that ChenSat-2a to ChenSat-2e are derived from the same satDNA precursor with diversification during speciation. As visualized by a multiple sequence alignment and an all-against-all dotplot (Figures 2b and S2), their monomer sequences share conserved residues over the whole length, with interfamily identities between 40.5
and 59.6%. ChenSat-2a to ChenSat-2e are highly amplified in only one of the two (sub)genomes, show G/C contents below 32%, and satDNA-typical monomer lengths of 170–171 bp (Table 1 and Figure 2b), presumably indicating selective constraints on the monomer size.

To test whether (sub)genome-specificity is restricted to non-coding tandem repeats, we extracted the consensus sequences of four 5S rDNA variants, one from *C. pallidicaule*, one from *C. suecicum*, and two from *C. quinoa*, derived from the A and B subgenomes, respectively. As expected, the 5S rRNA genes, including the regulatory boxes A, IE, and C, are 100% identical in all analyzed *Chenopodium* species (Figure S3), and differ only slightly from more distant plant 5S rRNA genes (e.g. 92% identity to pine). However, between the subgenomes, the 5S rDNA spacers accumulated differences, including point mutations and a 12 bp insertion in *C. pallidicaule* (Figure S3). This sequence divergence is sufficient to assign the 5S rDNA spacers from *C. quinoa* to either the A or the B subgenome.

**Higher-order structure, head-to-head junctions, and retrotransposon association of quinoa satellites**

Using 130 314 *C. quinoa* SMRT reads, we searched for higher-order arrangements and changes in repeat orientation. To achieve this, we identified tandem repeat-containing SMRT reads with a robust nHMM search (Table S1), retrieved 7319 sequences and prepared dotplots for visual inspection. We observed four different patterns of tandem repeat organization: continuous arrays, short arrays, inversions, and higher-order arrangements. For all *C. quinoa* tandem repeats (ChenSat-1a, ChenSat-1b, and ChenSat-2b to ChenSat-2e, 5S rDNA), exemplary dotplots of 5000 bp sequence stretches were shown, representative for each category (Figure S4). All analyzed tandem repeats occurred in both long and short arrays. Head-to-head organization was rare and only found in sequence reads containing ChenSat-1a, ChenSat-2b, and ChenSat-2e. Notably, most ChenSat-1a arrays identified were organized in higher order, with lengths of repeating units ranging from 1000 to 2000 bp, occasionally exceeding 5000 bp. Similar higher order repeat (HOR) structures have been identified for ChenSat-1b, ChenSat-2c, and ChenSat-2e.

For ChenSat-1a, we observed dotplots with array interruptions. Inspection of the interrupting sequences revealed the interspersion with LTR retrotransposons. Most strikingly, ChenSat-1a arrays have been interrupted by Ty3-gypsy retrotransposons of the chromovirus lineage in many cases (Figure S5). In order to assign these retroelements to a specific chromoviral clade, we compared its key enzyme, the reverse transcriptase (RT), with other chromovirus RTs from Neumann et al. (2011). We clearly assigned this *Chenopodium* retrotransposon to the CRM-type chromoviruses of group A (Figure S6a), often marked by an integration preference for the centromeric heterochromatin (Neumann et al., 2011). It has a high similarity to the centromeric chromoviruses *Beetle1*, *Beetle2* and *Beetle7* in the related genera *Beta* and *Patellifolia* (Weber and Schmidt, 2009; Weber et al., 2013), all known to colocalize with satDNA. An in-depth comparison of the integrase region enabled the identification of a C-terminal chromodomain with the CR-motif (Figure S6b), presumably conferring integration preference into centromeres (Novikova, 2009; Neumann et al., 2011). This retrotransposon

| Repeat family | Genome | Monomer length | GC content (%) | Pairwise identity (%)<sup>a</sup> | No. of analyzed reads<sup>b</sup> |
|---------------|--------|----------------|----------------|-------------------------------|-------------------------------|
| ChenSat-1a    | *C. pallidicaule* | 40  | 27  | 88  | 767  |
|               | *C. suecicum*     | 40  | 27  | 86  | 2479 |
|               | *C. quinoa*       | 40  | 26  | 88  | 12 758 |
| ChenSat-1b    | *C. suecicum*     | 48  | 53  | 85  | 149  |
|               | *C. quinoa*       | 48  | 53  | 84  | 387  |
| ChenSat-2a    | *C. pallidicaule* | 170 | 31  | 93  | 2338 |
| ChenSat-2b    | *C. pallidicaule* | 170 | 30  | 87  | 781  |
|               | *C. quinoa*       | 170 | 30  | 89  | 361  |
| ChenSat-2c    | *C. suecicum*     | 170 | 30  | 88  | 2077 |
|               | *C. quinoa*       | 170 | 30  | 86  | 679  |
| ChenSat-2d    | *C. suecicum*     | 171 | 32  | 90  | 931  |
|               | *C. quinoa*       | 171 | 31  | 85  | 212  |
| ChenSat-2e    | *C. pallidicaule* | 171 | 24  | 89  | 6942 |
|               | *C. quinoa*       | 171 | 29  | 84  | 848  |
| Chen5S        | *C. pallidicaule* (AA) | 331 | 48  | 96  | 994  |
|               | *C. suecicum* (BB) | 318 | 43  | 95  | 1041 |
|               | *C. quinoa* (A derived) | 319 | 39  | 93  | 235  |
|               | *C. quinoa* (B derived) | 317 | 35  | 99  | 348  |

<sup>a</sup>Pairwise identity represents the average identity of the reads to the consensus.

<sup>b</sup>Number of reads that map to the species-specific consensus from a total of 2 · 1.7 million paired-end reads.

---

© 2020 The Authors.

*The Plant Journal* published by Society for Experimental Biology and John Wiley & Sons Ltd, *The Plant Journal*, (2020), doi: 10.1111/tpj.14705
Subgenome-specific tandem repeats in quinoa

is highly associated with ChenSat-1a, also highly repetitive, and is represented by RepeatExplorer cluster CL39 (Figure 1).

Chenopodium A and B genomes have distinct satDNA profiles

In order to quantify the tandem repeat abundance in the A- and B-genome diploid species, we individually aligned reads from each of the three genomes against the monomeric consensus sequences. Individual read mapping counts of C. quinoa, C. suecicum, and C. pallidicaule monomers (Figure 3) showed that ChenSat-1a is present in all three species, with a five-fold to 16-fold copy number in C. quinoa as compared to the diploid genomes. ChenSat-2a has been exclusively detected in C. pallidicaule, and ChenSat-2b in C. pallidicaule and C. quinoa, suggesting their specificity for the A genome. Likewise, B-genome specificity was inferred for ChenSat-1b, ChenSat-2c, and ChenSat-2d, all present in C. suecicum and C. quinoa, but absent in C. pallidicaule. ChenSat-2e has been detected in all three species, however with large read count differences, with a particular high amplification in the A genome diploid C. pallidicaule (6942 reads) and the A subgenome of C. quinoa (848 reads). The 5S rDNA is present in all genomes, however, the A-genome- and B-genome-derived 5S rDNA spacers from quinoa, have been only detected in A and B genomes, respectively.

As next-generation sequencing is biased against G/C-rich and A/T-rich sequences, sometimes differing strongly from the genomic mean G/C values (Benjamini and Speed, 2020).
2012; Chen et al., 2013), quantification of satDNA based on read counts may misrepresent the genomic satDNA abundance. To verify the genome-specificity detected by bioinformatics (Figure 3), and to estimate the tandem repeat abundance in Chenopodium and related genera, we comparatively hybridized the ChenSat probes onto restricted genomic DNA (Figure 4). For each satDNA family, the chosen restriction enzyme is indicated. We tested 18 species (Table 2), including two A genome diploids (lanes 1 and 2), two B genome diploids (lanes 3 and 4), two allopolyploids containing A and B subgenomes (lanes 5 and 6), distantly related Chenopodium vulvaria with the diploid V genome (lane 7), further related species from all sections of the Chenopodiaceae (lanes 8–15), and three outgroups from the Betoideae (lanes 16–18). As satDNA is composed of regularly repeated sequences, restriction of genomic DNA cuts the array into fragments of similar length. Therefore, if a ladder pattern is revealed upon hybridization, this is clear evidence for an organization in long arrays, as summarized in Table 2.

ChenSat-1a occurs in all tested Chenopodium species (lanes 1–7) and in the closely related Atriplex hortensis (lane 8), producing ladder patterns with similar band sizes, consistent with the 40 bp monomer (Figure 4a). ChenSat-1a is highly abundant in A-genome-containing species (C. watsonii, C. pallidicaule, C. quinoa, C. album) and abundant, but in lower copy numbers, in the B sub-genome diploids (C. suecicum, C. ficifolium) and the more distantly related species (C. vulvaria, A. hortensis). Although read-based (Figure 3) and experimental (Figure 4a) quantifications differ, both methods provide strong evidence that ChenSat-1a plays a major role in all analyzed Chenopodium genomes.

In contrast, ChenSat-1b shows only signals in the species containing B genomes, such as C. suecicum, C. ficifolium, C. quinoa and C. album (Figure 4b, lanes 3–6). The hybridization generates similar restriction ladders for all four species, with the highest abundance in the diploid C. ficifolium (lane 4). The autoradiogram verifies the short ChenSat-1b monomer length (48 bp), with bands up to the hexamer and a smear ranging up to 500 bp.

ChenSat-2a hybridizes exclusively to C. pallidicaule (Figure 4c, lane 2), and is absent from the other A-genome-containing di- and polyploids tested. This is in line with its exclusion from the C. quinoa and C. suecicum genomes as detected by read mapping and clustering (Figures 3 and S1). Hybridization yields ladder patterns with strong monomeric bands, supporting the 170 bp monomer length, and signals up to the decamer, before falling together to form a smear.

Upon hybridization of ChenSat-2b, we detected a strong ladder pattern for both A genome diploids (Figure 4d, lanes 1, 2) and C. quinoa (lane 5). The C. vulvaria genome produces weak ladder hybridization patterns, indicating presence of ChenSat-2b (as the only repeat from the ChenSat-2 family) in more distantly related Chenopodium species. The banding pattern supports a conserved ChenSat-2b monomer length of 170 bp (lanes 1, 2, 5, 7). Interestingly, C. suecicum (lane 3), C. album (lane 6), and Lipandra polyspermum (lane 11) produce moderate and, in particular A. hortensis (lane 8) produces strong signals in the high molecular weight fraction of the DNA. This indicates the presence of diverged repeats lacking the conserved Haelll site in these genomes, presumably also belonging to ChenSat-2b or a closely related subfamily.
Similar to ChenSat-1b, the B-specific ChenSat-2c and ChenSat-2d probes hybridize exclusively to the B-genome-containing species *C. suecicum*, *C. ficifolium*, *C. quinoa*, and *C. album* (Figure 4b,e,f, lanes 3–6). ChenSat-2c hybridization generates ladder patterns in all four species, without monomer length variation (Figure 4e). We observe equally strong and similarly spaced ladder patterns up to the nonamer and higher. Hybridization with ChenSat-2d produces two superimposed restriction ladders for all B-subgenome-containing species, indicating presence of an internal *Hae*III site in some monomers (Figure 4f).

Indeed, apart from a single canonical *Hae*III site, the ChenSat-2d consensus sequence contains five additional positions, in which a single nucleotide mutation could result in an intact 5-GGCC-3 restriction site (Figure 2). *C. quinoa* (lane 5) shows a weaker hybridization pattern, likely to be caused by reduced ChenSat-2d abundance and consistent with the read mapping.

For ChenSat-2e, we detected strong ladder hybridization for the A genome diploids (lanes 1 and 2) and for *C. quinoa* (lane 5) after short exposure (data not shown). After long exposure (72 h), weaker signals with reduced ladder

---

**Figure 4.** Abundance, specificity and genomic organization of *Chenopodium* tandem repeats in the genus *Chenopodium* and in related plant genomes. Genomic DNA of 18 plants was cut enzymatically as indicated in each panel and separated by gel electrophoresis. The autoradiograms after comparative Southern hybridization of ChenSat-1a (a), ChenSat-1b (b), and ChenSat-2a to ChenSat-2e (c–g) are shown. The ladder patterns indicate the organization into long arrays and strong signal strengths point to high abundance. Exposure times ranged between 9 h and 2 weeks as indicated below the autoradiographs.

(h) Selected plant species, their corresponding lanes, and their relationship according to published trnL-F and matK/trnK phylogenies (Fuentes-Bazan et al., 2012a). Branch lengths are not to scale.
patterns in B diploids were detectable (Figure 4g), indicating very low abundance in B-containing genomes.

Summarizing, for the allotetraploid *C. quinoa*, and the diploids *C. palidicaule* and *C. suecicum*, we collected computational and experimental evidence for their organization in long arrays. We showed that most satDNAs are differentially amplified in A and B genomes and provide an in-depth view into their distribution in Chenopodiaceae genomes.

**Satellite DNA allows the assignment of chromosomes to the A and B subgenome**

To determine the satDNA localization along chromosomes we hybridized the 5S and the 18S–5.8S–25S rDNA as well as the newly identified tandem repeats to *C. palidicaule* (AA diploid), *C. ficifolium* (BB diploid) and *C. quinoa* (AABB tetraploid) metaphase spreads (Figure 5).

To enable a comparison with our *in silico* analyses, we also performed FISH to *C. suecicum* (BB diploid), included as Figure S7. For assignment of the *C. quinoa* chromosomes to either the A or the B subgenome, we re-hybridized five metaphases with genomic DNA of B-derived *C. suecicum* by GISH (Figure 5e,i,k,m,o, green).

Regarding the 5S rRNA genes, in *C. palidicaule*, we observed a major signal on two chromosomes in the interstitial region (Figure 5a,f,h,n, blue), in line with previous reports for the 5S rDNA of *C. palidicaule* (Kolano et al., 2012). In *C. ficifolium*, hybridization of the 5S rRNA genes resulted in a distal signal on one chromosome pair (Figure 5b,d,j,l, blue). In tetraploid *C. quinoa*, we observed strong 5S rDNA signals on four chromosomes (Figure 5c,e,i,k,m,o, blue). Whereas the two B-derived chromosomes
carried a distal signal, the two A-derived chromosomes harboured the 5S rDNA in the interstitial region (Figure 5e, i, o), also in line with prior reports (Maughan et al., 2006; Kolano et al., 2012). Hybridization of the 18S–5.8S–25S rDNA genes revealed a single chromosome pair with a distal signal on one chromosome arm for all three genotypes, despite C. quinoa’s tetraploidy (Figure 5, panels b–f, h–o, orange). GISH to C. quinoa metaphases indicated the origin of the 18S–5.8S–25S rDNA genes from B genomes (Figure 5e, i, k, m, o), corroborating previous reports (Maughan et al., 2006; Kolano et al., 2016).

As ChenSat-1a has been detected in all three genotypes as visible by Southern hybridization (Figure 4a), it was hybridized to chromosome spreads of each of the three accessions (Figure 5a–5c). In the C. pallidicaule A genome, six strong, pericentromeric ChenSat-1a signals (Figure 5a, red) were detected on metaphase spreads, including the chromosomes harbouring the 5S rRNA genes (Figure 5a, blue signals). ChenSat-1a is positioned in close vicinity and partially overlapping to the 5S rRNA gene array in the pericentromeric region on one chromosome (Figure 5a, arrowed). In the B genome of C. ficifolium ChenSat-1a produces eight strong signals, often localizing in the pericentromeric regions and spreading into intercalary regions of the remaining chromosomes (Figure 5b, red). Co-localization of ChenSat-1a with 5S and 18S–5.8S–25S rRNA genes has not been observed (Figure 5b, blue, orange).

ChenSat-1a was detected in the pericentromeric chromatin of all C. quinoa chromosomes with 24 major, eight moderate and four minor signals (Figure 5c, red). We observed co-hybridization of ChenSat-1a and two of the 5S rRNA gene signals (Figure 5c, blue, arrows), verified by the presence of both repeats on identical SMRT reads (Table S2). The chromosome pair carrying the distal 18S–5.8S–25S rDNA contained pericentromeric, moderate ChenSat-1a signals.

In C. ficifolium, the chromosome pair with the 18S–5.8S–25S rRNA genes (orange) carried two major ChenSat-1b signals in the intercalary and pericentromeric regions (Figure 5d, arrows). The distal 5S rRNA genes (blue) marked an additional chromosome pair with major, intercalary ChenSat-1b signals on one chromosome arm (Figure 5d, red, arrowheads). We observed further major signals in intercalary and pericentromeric regions of a third chromosome pair (Figure 5d). Similarly, ChenSat-1b was localized along distal and pericentromeric regions of the 18 chromosomes of the C. quinoa B subgenome (Figure 5e, green), with two major and four moderate signals, indicating large ChenSat-1b arrays (Figure 5e). Similar to C. ficifolium (Figure 5d), the two major signals are located on the 18S–5.8S–28S rDNA-carrying chromosome pair, whereas two of the four moderate signals can be found in the intercalary regions of the chromosome arms carrying the distal 5S rDNA (Figure 5e, arrowheads, orange and arrows, blue signals).

ChenSat-2a is present on all C. pallidicaule chromosomes with varying intensities (Figure 5f). It is localized mainly in the interstitial, but also in the distal chromosomal areas. At the higher resolution of interphase nuclei, we observed that ChenSat-2a is largely excluded from the strongly 4,6-diamidino-2-phenylindole (DAPI)-positive heterochromatin, but without complete exemption from the chromocentres (Figure 5g).

In C. pallidicaule, the ChenSat-2b signals are dispersed on all chromosomes with varying signal intensities, mostly in interstitial and intercalary regions (Figure 5h). In C. quinoa, ChenSat-2b is dispersed on many of the 18 A-genome-derived chromosomes (Figure 5i). Two chromosomes, presumably a pair, carry a major intercalary signal on one arm (arrow), suggesting the presence of large ChenSat-2b arrays.

ChenSat-2c produces weak-to-moderate signals and is distributed in the intercalary and distal regions of the C. ficifolium chromosomes (Figure 5j, red). Only two chromosomes showed stronger hybridization. As indicated by FISH and GISH on C. quinoa metaphases, ChenSat-2c is present on all 18 B subgenome-derived chromosomes (Figure 5k). These signals are present in low-to-moderate signal strength in intercalary and pericentromeric regions. We observed ChenSat-2c co-localization with the distal and B-genome-derived 5S rDNA (arrows).

On chromosomes of C. ficifolium, ChenSat-2d is widely dispersed with low-to-moderate intensities (Figure 5l). In C. quinoa only B-genome-derived chromosomes (Figure 5m green) showed ChenSat-2d arrays of different size distributed in the intercalary and distal chromosome regions, with exclusion from the pericentromeric regions. At C. pallidicaule metaphases, ChenSat-2e produced two very strong hybridization sites on the 5S rDNA-carrying chromosome pair (Figure 5n, blue, arrowed). Twelve chromosomes show weak-to-moderate signals and four chromosomes give only faint signals. Similarly, in C. quinoa, a chromosome pair carries a major signal (Figure 5o, arrows), however, without co-localization to the 5S rDNA genes. GISH revealed that these two major signals localized on A-genome-derived chromosomes.

Intermingling of tandem repeat families along C. quinoa chromosomes as detected on SMRT reads, scaffolds and by FISH

To infer evolutionary relationships and potential exchange between homoeologous chromosomes, we analyzed the physical neighbourhood and the intermingling of tandem repeats on C. quinoa SMRT reads, scaffolds, and (pseudo) chromosomes. Using a tandem repeat nHMM analysis, we found that 201 out of 130 314 C. quinoa SMRT reads
harboured arrays from at least two different satellite repeats (Table S2). We focused on two combinations:

1 ChenSat-1a and ChenSat-1b were detected on six reads (Figure 6a). As both repeats share considerable sequence identity (Figure 2a) and are potentially related, we verified their co-localization by multicolour FISH with ChenSat-1a and ChenSat-1b probes. We identified arrays of both tandem repeats in the pericentromeric region of two metaphase chromosomes (Figure 6c, arrows) and showed interspersion along stretched chromatin fibres (Figure 6d).

2 Combinations of ChenSat-2c and ChenSat-2e were detected on 91 SMRT reads (Figure 6b) with both satellites organized in short arrays. ChenSat-2e is strongly enriched in the A genome, whereas ChenSat-2c is restricted to B-derived regions. Using dual-colour FISH, we identified both repeats in close proximity on two chromosomes (Figure 6e, arrows), corroborating their association on the SMRT reads.

Figure 6. Co-localization of satDNAs in the *C. quinoa* genome investigated on long reads and along chromosomes.
(a) ChenSat-1a and ChenSat-1b arrays were detected on six single molecule real-time (SMRTs) reads. Here, a dotplot of a representative 10 kb region from a SMRT read is shown.
(b) Similarly, short arrays of ChenSat-2c and ChenSat-2e have been detected next to each other on SMRT reads in 91 cases. A dotplot of a 10 kb region of a representative read is shown.
(c) Dual-colour fluorescent in situ hybridization (FISH) on *C. quinoa* metaphase chromosomes provides practical evidence that ChenSat-1a (red) and ChenSat-1b (green) co-occur on two chromosomes (arrows). We showed 4',6-diamidino-2-phenylindole (DAPI)-stained chromosomes and fibres in grey, and additionally probed the 18S–5.8S–25S rDNA (orange) for easier chromosome allocation.
(d) Hybridization of ChenSat-1a (red) and ChenSat-1b (green) to stretched fibres further supports their interspersed arrangement.
(e) Dual-colour FISH of the B-specific ChenSat-2c (red) and the A-enriched ChenSat-2e (green) on *C. quinoa* metaphases. The arrows mark two chromosomes showing both weak ChenSat-2c and weak ChenSat-2e signals, a potential indication of intergenomic exchange. Scale bars correspond to 5 μm.
As the 5S rRNA gene variants from the *C. quinoa* A and B genomes differ strongly in their spacer sequences (Figure 7a), they can also be used to detect intergenomic homogenization. However, the nHMM approach did not detect any co-occurrence of A-derived and B-derived 5S rRNA genes. To verify this, we exemplarily selected the eight longest *C. quinoa* SMRT reads, which were completely covered by 5S rDNA tandem repeats, and extracted 283 genes with spacer. The 5S rDNA monomers were aligned and their relationship was visualized by a dendrogram (Figure 7b). All monomers fall into one of two groups, belonging to either the A or B subgenome. For three reads, all 110 monomers group exclusively with the A subgenome reference, whereas the remaining five reads (173 monomers) were assigned to the B subgenome group. We did not find evidence for intermingling of 5S rDNA variants, and no indication of interlocus recombination between homoeologous chromosomes.

![Dendrogram](image)

Figure 7. The analyzed *Chenopodium* genomes contain homogeneous 5S rDNA arrays with either A-genome or B-genome variants.

(a) Divergence of the spacer from the 5S rDNA allows the monomer assignment to the A or B subgenome from *C. quinoa*. Dotplot visualization of the four 5S rDNA consensuses shows high similarity among the A-derived (pal and quiA) as well as among B-derived (sue and quiB) 5S rDNA monomers. The conserved gene (turquoise) and the variable spacer region (ochre) are indicated by shading. Comparison between (sub-)genomes shows the accumulation of single nucleotide polymorphisms and indels, resulting in lower identities, which range between 81 and 87%. The dotplot was generated using a word size of 10 with tolerance of 1 mismatch. For the corresponding multiple sequence alignment see Figure S3.

(b) Homogeneity of 5S rDNA arrays was analyzed by extraction of 283 5S rDNA monomers from eight *C. quinoa* single molecule real-time (SMRT) reads. We considered each SMRT read as an individual array. The monomers have been aligned and grouped by a neighbour-joining analysis. The position of reference monomers from *C. pallidicaule* and *C. suecicum* was indicated by magenta and green arrowheads, respectively. As 5S rDNA monomers derived from A and B subgenomes differ vastly in their spacer regions, the resulting dendrogram forms two major branches, each representative for A-derived and B-derived variants. For three reads, all 110 monomers group exclusively with the A subgenome reference, whereas the remaining five reads (173 monomers) were assigned to the B subgenome group. We did not find evidence for intermingling of 5S rDNA variants, and no indication of interlocus recombination between homoeologous chromosomes.

As the 5S rRNA gene variants from the *C. quinoa* A and B genomes differ strongly in their spacer sequences (Figure 7a), they can also be used to detect intergenomic homogenization. However, the nHMM approach did not detect any co-occurrence of A-derived and B-derived 5S rRNA genes. To verify this, we exemplarily selected the eight longest *C. quinoa* SMRT reads, which were completely covered by 5S rDNA tandem repeats, and extracted 283 genes with spacer. The 5S rDNA monomers were aligned and their relationship was visualized by a dendrogram (Figure 7b). All monomers fall into one of two groups, belonging to either the A or B genome. All A-genome-specific monomers have been derived from three of eight reads, whereas all B-genome-specific monomers have been extracted exclusively from the remaining five SMRT reads (see magenta and green line colour in Figure 7b, respectively). This sample already indicates maintenance of the 5S rRNA gene arrays without exchange...
between homoeologous chromosomes of the quinoa A and B genome.

Although some SMRT reads can span up to 50 kb, they do not scale to a chromosome. For a larger overview, we used 619 high-quality scaffolds from the *C. quinoa* genome sequence (Jarvis et al., 2017) to provide evidence of the interspersion and intermingling of tandem repeats. Based on gene data, 226 of the 619 scaffolds were assigned to either the A or B genome (Jarvis et al., 2017). If our genome-specific satDNAs were used as markers, 51 previously unassigned quinoa scaffolds have now been classified as either belonging to the A (24) or B (27) genome (Table S3). In six cases, we mapped both A-genome- and B-genome-derived satDNAs to the same scaffold, indicating exchange between quinoa A and B genome. However, as these six cases could not be verified on the shorter SMRT reads, we cannot rule out incorrect genome assembly.

Taking together, we detected intermingling of several satDNA families along the chromosomes, using long reads, detection on scaffolds, and FISH. We co-localized both ChenSat-1 families in *C. quinoa*, which may be an indication of their common decent. For the ChenSat-2 subfamilies, sequence data may indicate DNA exchange between A and B subgenomes, whereas we can confidently exclude dispersal of the 5S rRNA genes in the respective other subgenome.

**DISCUSSION**

Next-generation and third-generation sequence reads give an overview on satDNA landscapes in *Chenopodium*

By clustering of next-generation sequence reads, we detected and classified the repetitive genome proportion of allotetraploid *C. quinoa* and the two potential progenitor genomes *C. palidicaule* and *C. suecicum*, revealing repeat fractions of 73.2, 76.9, and 78.1%, respectively. As repeats may have escaped their detection by accumulation of mutations over time, recombination, diversification, reshuffling, and decay (Ma et al., 2004; Wollrab et al., 2012; Elliott and Gregory, 2015; Sanchez et al., 2017; Bourque et al., 2018), we consider these values as underrepresentations. From the extracted repeat dataset, we identified and characterized seven *Chenopodium* satDNAs and the 5S rRNA genes. Six out of the seven ChenSat repeats were specific or strongly enriched in either the *Chenopodium A* (ChenSat-2a, ChenSat-2b, ChenSat-2e) or B genome (ChenSat-1b, ChenSat-2c, ChenSat-2d), respectively. Only ChenSat-1a has been detected in both A and B genomes in high abundance.

Read cluster analyses of high-throughput data have already been used to gain access to tandem repeat families in a number of plants, such as pea, bean, onion, *Camellia*, crocus, pepper, and fern (Macas et al., 2007; Heitkam et al., 2015; Kirov et al., 2017; Ávila Robledillo et al., 2018; Schmidt et al., 2019; Zhou et al., 2019), with genomic satDNA fractions ranging from 0.1 to 36% (Garrido-Ramos, 2017). In the allopolyploid *C. quinoa*, satDNA accounts for 5.4% of the repeated DNA fraction and for 4.0% of the genome. Compared to the closely related sugar beet with a satDNA proportion of 11.15% (Kowar et al., 2016), the satDNA content of quinoa is low. However, close taxonomic relationship between species is not correlated with similarly sized repeat fractions; even within species of a single genus, satDNA families can amplify with vast differences as observed between *Fritillaria* species (Kelly et al., 2015).

Genome assemblies are error-prone, in particular in repetitive regions. For example, the human reference genome is one of best-studied genome assemblies, but still contains many inaccuracies regarding satDNA array length and higher-order organization (Miga, 2015; Jain et al., 2018). For non-model organisms, assemblies often contain even less information making the study of repetitive regions laborious (Peona et al., 2018). The availability of third-generation long reads opens the way to solve genomic and evolutionary questions targeting satellite and ribosomal DNAs, such as array length, abundance and organization in higher-order structures or head-to-head arrangements (Sevim et al., 2016; Khost et al., 2017; Symonová et al., 2017; Lower et al., 2018; Cechova et al., 2019; Vondrak et al., 2020). For *C. quinoa*, we identified arrays of the satellite families ChenSat-1a, ChenSat-1b, ChenSat-2b, ChenSat-2c, ChenSat-2d, and ChenSat-2e on SMRT reads; all were arranged in short or long homogeneous arrays, but also in structures of higher order (ChenSat-1a, ChenSat-1b, ChenSat-2c, and ChenSat-2e) and head-to-head arrangements (ChenSat-1a, ChenSat-2b, and ChenSat-2e). Similar to observations in the pike genome (Symonová et al., 2017), we also detect inversions in the 5S rDNA, and long as well as short arrays. Higher-order arrangement of ChenSat-1a has also been detected in clones of multimers as reported very recently (Belyayev et al., 2019).

**Seven *Chenopodium* satDNAs fall into only two major satDNA families**

We identified seven *Chenopodium* satDNAs, belonging to one of two families, the ChenSat-1 family with short 40–48 bp monomers or the ChenSat-2 family with 170–171 bp monomers.

The most abundant satDNA family in the analyzed *Chenopodium* genomes is ChenSat-1a, initially published as *C. quinoa* clone 12–13P (Kolano et al., 2011; Orzechowska et al., 2018). Its short monomer size of 40 bp is unusual for a tandem repeat of high abundances, which generally consist of 160–180 or 320–360 bp monomers (Hemleben et al., 2007). Historically, satDNAs have been split into micro- (2–5 bp monomer), mini- (6–100 bp monomer), and conventional satellites (>100 bp monomer) (Vergnaud and...
Denoeud, 2000; Mehrotra and Goyal, 2014). In recent times, the focus has shifted to recognize not only the satDNA’s monomer size, but its genomic organization of repeating units into long arrays as the main molecular hallmark of satDNA (Richard et al., 2008). These long arrays have been verified for ChenSat-1a and ChenSat-1b, and we therefore do not consider them as minisatellites, but as canonical satellite DNA. Major satDNA arrays, made up of short monomers <100 bp, have been occasionally observed in other plants: In *Ricinus communis* a satellite with 39 bp monomers constitutes 8.2% of the genome (Melters et al., 2013). Likewise, a 43 bp satellite from *Camellia japonica* occupies more than 11% of the genome and is located distally on all chromosomes (Heitkam et al., 2015). An extreme case is the fern *Vandenboschia speciosa*, in which the numerous satDNAs have short monomer lengths (between 33 and 141 bp, Ruiz-Ruano et al., 2018). This underpins that satellites with short monomer lengths can form very large arrays as observed here for ChenSat-1a and ChenSat-1b.

Despite the short monomer length, we observed a ChenSat-1a organization with and without higher order, indicating an evolution towards longer and more complex repeated ChenSat-1a patterns (as reviewed by Plohl et al., 2012). The ChenSat-1a higher-order repetitions were not of conserved lengths, but differed in size on the different long reads. This may be explained by Rudd et al.’s hypothesis (2006) that higher-order and monomeric satDNAs evolve at different rates, leading to less conserved higher-order repeat units compared with monomers.

The large expansion of the ChenSat-1a arrays, often with distinct organization in higher order, indicated that amplification by recombination played a large role in ChenSat-1a evolution. Its monomers are generally low in guanine/cytosine (28%), and do not include any symmetrical DNA methylation targets. As DNA methylation may hinder recombination, as reported in *Arabidopsis thaliana* (Melamed-Bessudo and Levy, 2012), low DNA methylation may have stimulated recombination and thus ChenSat-1a abundance. Strikingly, although we observed vast expansions of ChenSat-1a across *C. quinoa* chromosomes, *C. quinoa* genome size (between 1.45 and 1.5 Gb, Palomino et al., 2008) remained largely additive of the genome sizes of possible AA and BB diploid progenitors (between 0.6 and 0.9 Gb, Kolano et al., 2016; Mandák et al., 2016). This indicates satDNA array expansion may have been counterbalanced by genome size reduction, leading to relatively equal genome sizes.

ChenSat-1a localizes at the centromeric constriction of all *C. quinoa* centromeres. This is in line with the observation, that in most organisms, the main satellite is expected at the centromere (Jiang et al., 2003; Meleters et al., 2013). Its arrays are interspersed with full-length chromoviruses of the CRM lineage, as described for many plant centromeres (Cheng et al., 2002; Weber et al., 2013). This chromovirus family was different from a previously characterized partial *C. quinoa* chromovirus RT of the Tekay clade (Kolano et al., 2013). Instead, it represents a canonical CRM chromovirus, containing a chromomdomain of the CR-type, typical for centromeric retrotransposons (Novikova, 2009; Weber and Schmidt, 2009; Neumann et al., 2011), thus providing support for ChenSat-1a’s association with centromeres.

Similar to ChenSat-1a, the related family ChenSat-1b has also a short monomer length of 48 bp. High sequence identities of 60% over the whole length indicate a common decent. As ChenSat-1b has a different species distribution, limited solely to B-genome-derived *Chenopodium* genomes, we assumes that ChenSat-1a is more ancient and likely to be the progenitor. Only in a few instances, we localized ChenSat-1b to the *C. quinoa* pericentromeric regions, making a role in the formation of active centromeres unlikely. This contrasts with observations for some plants, such as the common bean, in whose genomes active centromeres are formed by two alternative satDNAs, CentPv1 or CentPv2 (Iwata et al., 2013).

The satDNA subfamilies ChenSat-2a, ChenSat-2b, ChenSat-2c, ChenSat-2d, and ChenSat-2e were characterized by the conventional monomer lengths of 170 and 171 bp, containing conserved sequence stretches, and pairwise sequence identities of about 60%. Therefore, we considered these as subfamilies of the ChenSat-2 family, and postulated a common origin. All ChenSat-2 satDNA families formed long arrays, as verified by Southern hybridization and SMRT read analysis.

**Different repeat landscapes emerged in *Chenopodium A* and B genomes during speciation**

Repeats are subject to rapid changes during adaptation to new environments, contributing to rapid genome evolution and speciation (Oliver et al., 2013; Stapley et al., 2015; Serrato-Capuchina and Matute, 2018). After allopolyplidization, it is assumed that an imbalance of repeats and their epigenetic impact drive the polyploidization life circle and often lead to genome size shrinkage, diploidization, and subgenome dominance (Edger et al., 2017; Vicent and Casacuberta, 2017; Mhiiri et al., 2019). We traced the satellite DNA evolution in the genus *Chenopodium* and follow the evolutionary history of *C. quinoa*, as summarized in our evolutionary scenario (Figure 8).

For the ChenSat-1 family, we presented evidence that the subfamilies ChenSat-1a and ChenSat-1b are related. First, both satDNAs have short monomer sizes below 50 bp and about 60% sequence similarity. Second, they occur close to each other, as detected on long reads and in the reference genome assembly, and also confirmed by multicolour FISH with ChenSat-1a and ChenSat-1b probes. As ChenSat-1a occurs ubiquitously in *Chenopodium* species, whereas
ChenSat-1b is restricted to B genomes, we assumed that the B-specific ChenSat-1b emerged from ChenSat-1a by accumulation of mutations in a B genome precursor. The increase in monomer size may have occurred by replication slippage, as has been suggested for microsatellites (Viguera et al., 2001), by unequal crossing over, or by repair of double-stranded breaks. ChenSat-1b emerged without full replacement of ChenSat-1a, indicating incomplete homogenization. This is striking, as newly emerging families often replace the progenitor (Dover, 1982; Plohl et al., 2012). This may point to a structurally important role of ChenSat-1a, potentially forming the active centromere.

Figure 8. Scenario for the evolution of tandemly repeated DNA during the history of Chenopodium speciation and allotetraploidization. We detected ChenSat-1a in all Chenopodium species tested, as well as in Atriplex hortensis, indicating the presence of ChenSat-1a in the common ancestor of Chenopodium and Atriplex. Similarly, as members of the ChenSat-2 superfamily have been detected in all Chenopodium and Atriplex species tested, we suggested the presence of a common ChenSat-2 progenitor family, most likely to be similar to ChenSat-2b or at least closely related. The ChenSat-2b repeat persisted in A genomes, but was reduced in the B genomes. In addition, ChenSat-2e emerged in A genome diploids, whereas ChenSat-2c and ChenSat-2d evolved in B genomes. ChenSat-2a is most likely to be the youngest family, as it was absent in most A-containing and B-containing species, except C. pallidicaule. We can therefore exclude C. pallidicaule as the parental species for C. quinoa. Moreover, ChenSat-1b evolved in B genomes, presumably from divergence of ChenSat-1a monomers. Dating back at least 3 million years ago (Ma), allotetraploidization of maternal A and paternal B genomes has led to C. quinoa (Kolano et al., 2016), containing ChenSat-1a, ChenSat-1b, and ChenSat-2b to ChenSat-2e tandem repeats. For the rDNAs, after Chenopodium speciation, the SS and 18S–5.8S–25S rDNA spacers began to diverge and form A-specific and B-specific Chenopodium variants. The two A-derived and two B-derived 5S rDNA major sites are added to generate four 5S rDNA major sites in C. quinoa. In contrast, only B-derived 18S–5.8S–25S rDNA was detectable in C. quinoa (Maughan et al., 2006; Kolano et al., 2016), whereas the A variant had been lost. The time axis includes estimates of the stem age of the Atripliceae (Kadereit et al., 2010), the early splits of Atriplex and Chenopodium as well as C. vulvaria (Mandák et al., 2018), and the allotetraploidization event leading to quinoa (Jarvis et al., 2017).
For the ChenSat-2 family, we observed diversification, with at least five ChenSat-2-derived subfamilies in different A and B genomes of Chenopodium. All ChenSat-2 subfamilies are most likely to be derived from a presumed common progenitor. Among other hallmarks, we observed a highly conserved monomer size (170–171 bp) for all ChenSat-2 subfamilies, likely to be important for DNA phasing (Melters et al., 2013), and an indication of selective constraints. We used several approaches to corroborate the different abundance patterns and genome specificities, such as read mapping, quantification in the assembled C. quinoa subgenomes, Southern hybridization, and FISH to allotetraploid C. quinoa. The restriction to individual (sub)genomes indicates a rapid and species-specific ChenSat-2 evolution leading to a variety of different subfamilies. Similar observations have been reported for various plants and animals (Kopecna et al., 2012; Cai et al., 2014; Liu et al., 2019). A theoretical model explaining this mode of satDNA evolution has already been proposed 50 years ago (Fry and Salser, 1977): It suggests that related species shared a library of conserved satDNAs in low copy number, some of which may expand differently during speciation. This classical satellite library hypothesis serves well for explaining the patchy, species-specific evolution of the ChenSat-2 superfamily.

As ChenSat-2b is most widespread with presence in distantly related C. vulvaria, A. hortensis, and even L. polyspermum, we suggested ChenSat-2b or a precursor ChenSat-2b variant as the progenitor sequence. However, reduced ChenSat-2b signals in B-genome diploids indicates an incomplete elimination from these species, explainable for example by molecular drive (Dover, 1982; Dover, 2002).

ChenSat-2a is exclusively present in C. pallidicaule, but absent in other A-containing genomes such as C. quinoa, explainable by two scenarios: (i) ChenSat-2a may have emerged in C. pallidicaule after speciation, thus effectively excluding C. pallidicaule as potential parent of C. quinoa. This is consistent with data from genome sequencing considering C. pallidicaule as an unlikely progenitor of C. quinoa (Jarvis et al., 2017; Mangelson et al., 2019). (ii) Alternatively, ChenSat-2a may have been eliminated from other A-containing Chenopodium genomes analyzed here. This has been observed for example in natural and synthetic Nicotiana tabacum allotriploids, in which continuous NicCL3 satDNA arrays specific for the diploid progenitor N. tomentosiformis have been lost (Renny-Byfield et al., 2012). Nevertheless, multiple losses are necessary to explain the observed ChenSat-2a distribution. Therefore, we consider the emergence of ChenSat-2a in C. pallidicaule as the more likely scenario.

We did not observe the rise of new satDNA families in allotetraploid quinoa, as documented in other polyploids. In Nicotiana allotriploids which are older than 5 million years (N. nesophila, N. stocktonii, and N. repanda), new satDNAs have evolved, and sometimes have amplified to replace the parental satDNAs (Koukalova et al., 2010).

**Tandem repeats may provide targets for recombination between homoeologous chromosomes in allopolyploid quinoa**

With two distinct repeat landscapes of the A- and B-derived subgenomes, quinoa is well suited to investigate the invasion of satDNA into the respective other subgenome. In plants with a similar genome composition such as Nicotiana allotriploids older than 1 million years (N. quadrivalvis and N. clevelandii), an exchange of satellite sequences between homoeologous chromosomes was already suspected (Koukalova et al., 2010). In the allotetraploid C. quinoa genome, recombination between homoeologous chromosomes are assumed to be rare, however some incidents were detected already 2 decades ago (Ward, 2000). Using the most current reference genome assembly, only a small number of homoeologous gene pairs (3.1%) has been mapped within the same subgenome, suggesting that recombination and chromosomal rearrangements have occurred between the A and B subgenomes to a small extent (Jarvis et al., 2017). Accordingly, we mapped A-derived and B-derived satDNAs to the same scaffold in only six cases (Table S3), possibly indicating intergenomic recombination. However, using single molecule long reads originating from a single genomic region, we did not identify co-localization of A-specific and B-specific Chenopodium rDNA variants. Nevertheless, we provide evidence that short ChenSat-2e arrays, enriched in the A genome, co-occur with the B-genome-derived ChenSat-2c satDNA family, and confirmed co-localization on the same chromosome by multicolour FISH and long read data.

Taken together, using short and long read bioinformatics as well as Southern and fluorescent in situ hybridization, we traced seven satDNAs through Chenopodium speciation and allotriploidization. We observed satDNA diversification, replacement, reduction, and identified repeat families highly amplified in either the A-genome or B-genome diploids. After re-unification of both genomes in the allotriploid quinoa, four of the seven satDNAs were subgenome specific. We observed intermingling of satDNA families, which may point to homoeologous exchange of the ChenSat sequences. However, for the 5S rRNA genes, we can confidently suggest a strict separation of sequences on the A and B subgenomes.

**EXPERIMENTAL PROCEDURES**

**Read clustering and tandem repeat identification**

To identify the satDNAs from tetraploid C. quinoa and its diploid relatives, we used RepeatExplorer in comparative mode (Novák et al., 2010; Novák et al., 2013). We analyzed the reads from the quinoa genome projects (Yasui et al., 2016; Jarvis et al., 2017) deposited at the NCBI sequence read archive: Chenopodium...
pallicidae (SRR442539), Chenopodium suecicum (SRR4425238), and Chenopodium quinoa (DRR057249). Read pre-treatment and interlacing was performed with custom scripts accompanying the local RepeatExplorer installation (paired_-fastq_filtering.R and fasta_interlacer followed by seqclust). The reads were quality-trimmed to include only sequences with a Phred score ≥ 10 over 95% of the read length. Overlapping read pairs have been excluded. Before comparative clustering, we randomly sampled 1.7 million paired shotgun pre-treated reads for C. pallicidae, C. suecicum, and C. quinoa each, from which RepeatExplorer automatically chose 1,265,058, 1,263,518, and 1,265,808 reads, respectively. The resulting clusters have been classified by similarity searches against the Conserved Domain Database for the functional annotation of proteins (Marchler-Bauer et al., 2011), RepBase Update (Jurka et al., 2005), the REXdb database (Neumann et al., 2019), and a custom library containing common plant sequences (e.g. ribosomal, telomeric, and plastid sequences). Clusters connected by mates and with matching annotations have been combined manually to super-clusters. Clusters with satellite-typical star-like and circular graph representations (Novák et al., 2010) were selected for further analysis. The RepeatExplorer-derived contigs were assembled and putative monomers were detected using Tandem Repeats Finder (Benson, 1999).

Generation of representative consensus monomers and tandem repeat quantification by read mapping

To quantify the repeat abundance as well as to derive species-specific consensus sequences, we independently mapped 2× 1.7 million paired random short reads from each genome (C. quinoa, C. pallicidae, and C. suecicum), and repeated this until the consensus remained stable. For this, we used the Geneious 6.1.8 mapping tools with ‘medium sensitivity’ parameters (Kearse et al., 2012). We extracted the number of reads mapping to the species-specific consensus (Table 1) and visualized the mapping counts as a bubble chart with R and ggplot2 (Wickham, 2016) in Figure 3.

Sequence comparison

Multiple sequence alignments were generated with the MAFFT (Katoh and Standley, 2013) and MUSCLE (Edgar, 2004) local alignment tools. They have been manually refined and used for the calculation of pairwise sequence identities with MEGA X (Kumar et al., 2018). We explored and visualized sequences with the multi-purpose software Geneious 6.1.8 (Kearse et al., 2012). Dotplots were generated with FlexiDot (Seibt et al., 2018) with word sizes as indicated in the respective figure legends.

Computational localization of tandem repeats along the C. quinoa scaffolds

Tandem repeat positions on the C. quinoa pseudochromosomes and scaffolds (Jarvis et al., 2017) were deduced by local BLASTn of a tandem repeat dimer. We retained only hits with an e-value ≤ 10-10, and transferred the hits into gff3 format. For scaffolds, we counted the number of hits for the specific satDNA (sub)family and thus deduced the scaffold’s origin from the A or B subgenome.

Detection of higher-order arrangements and interlocus homogenization

We analyzed higher-order arrangements of tandem repeats on available C. quinoa SMRT reads from accession number DRR057268. To account for the sequence error of the long reads, we used a nucleotide Hidden Markov Model (nHMM)-based approach. For each tandem repeat, we generated an nHMM from Illumina reads mapped to the respective consensus and used nhammer (Wheeler and Eddy, 2013) to infer monomers along the SMRT reads. Hits were filtered individually for each satDNA family with parameters indicated in Table S2. Local monomer organization was analyzed visually using FlexiDot self dotplots (Seibt et al., 2018).

For detection of 5S rDNA interlocus homogenization, individual monomers were retrieved from the respective SMRT reads with the highest monomer count. The monomers were aligned using MAFFT (Katoh and Standley, 2013), grouped by neighbour-joining analysis and visualized with the ETE toolkit library (Huerta-Cepas et al., 2010).

Plant material, DNA isolation, polymerase chain reaction and cloning

Plant material was obtained from sources indicated in Table 2. The plants were grown in a greenhouse under long day conditions. We isolated DNA as described (Arumuganathan and Earle, 1991) using 2× cetyl trimethyl ammonium bromide buffer and the additive polyvinylpyrrolidone (PVP). Especially for C. quinoa DNA, it was essential to retrieve the DNA immediately after isopropanol precipitation without centrifugation to avoid contamination with metabolites.

From the C. quinoa reference monomers, outward facing primers were designed (Table S4). For amplification of satellite DNA probes for Southern hybridization and FISH, polymerase chain reaction (PCR) was carried out with specific primer pairs. PCR reactions with 50 ng plasmid template were performed in 50 µl volume containing 10× DreamTaq buffer and 2.5 units of DreamTaq polymerase (Promega, www.promega.de). Standard PCR conditions were 94°C for 5 min, followed by 35 cycles of 94°C for 1 min, primer-specific annealing temperature for 30 sec, 72°C for 1 min and a final incubation time at 72°C for 5 min. We cloned and sequenced the PCR products for each satDNA family, and selected the one with the highest identity to the reference monomer as the probe for subsequent hybridizations.

Southern hybridization

For comparative Southern blots, restriction enzymes specific for each tandem repeat were selected based on bioinformatics and practical tests of different enzymes. Genomic DNA was restricted, separated on 2% agarose gels and transferred onto Hybrid-N+ nylon membranes (GE Healthcare, www.gehealthcare.de) by alkaline transfer. Hybridizations were performed according to standard protocols using probes labelled with 32P by random priming (Sambrook et al., 1989). Filters were hybridized at 60°C and washed at 60°C for 10 min in 2× SSC/0.1% SDS. Signals were detected by autoradiography.

Probe labelling, metaphase preparation, genomic, and fluorescent in situ hybridization

The satellite-specific probes were labelled by PCR in the presence of biotin-16-dUTP (Roche, www.roche.de). We used nick translation to mark the probes for the ribosomal genes. The probe pZR18S containing a 5066 bp fragment of the sugar beet 18S–5.8S–25S rRNA gene (HESt8879, Paesold et al., 2012) was labelled with DY-415 or DY-647-dUTP (Dyomics, dyomics.com), whereas probe pXV1 (Schmidt et al., 1994) for the 5S rRNA gene was labelled with digoxigenin-11-dUTP. For GISH with B-subgenome-specific probes, genomic DNA of C. suecicum was heated to 99°C.
for 10 min before labelling with digoxigenin-11-dUTP, also by nick translation.

We prepared mitotic chromosomes from the meristem of young leaves. Before fixation in methanol/glacial acetic acid (3:1), leaves were incubated for 2.5-3 h in 2 M 8-hydroxyquinoline. Fixed plant material was digested for 4.5 h at 37°C in an enzyme mixture consisting of 4% (w/v) cellulase Onozuka R10 (Sigma 16419, www.sigmaaldrich.com) and 20% (v/v) pectinase from Aspergillus niger (Sigma P4716) in citrate buffer (4 M citric acid and 6 M sodium citrate) according to Kolano et al. (2011). After maceration, the mix was incubated for an additional 25 min, before chromosome spreading by dropping according to Schwarzacher and Healop-Harrison (2000) with modifications for beet (Schmidt et al., 1994).

We used FISH and GISH procedures described previously (Heslop-Harrison et al., 1991) with modifications for Amaranthaceae plants (Schmidt et al., 1994). Chromosome preparations were counterstained with DAPI and mounted in antifade solution (CitiFluor, www.citifluor.co.uk). We examined the slides with a Zeiss Axioimager M1 UV epifluorescence microscope with the appropriate filters, and equipped with an ASI BV300-20A camera coupled with Applied Spectral Imaging software. The images were processed using Adobe Photoshop C5 software (Adobe Systems, San Jose, CA, USA) using only contrast optimization, Gaussian, and channel overlay functions affecting the whole image equally.

ACKNOWLEDGEMENTS

If I should die and leave you here a while, be not like others sore undone, who keep long vigils by the silent dust. For my sake turn to life and smile, nerving thy heart and trembling hand to undone, who keep long vigil by the silent dust. For my sake turn to life and smile, nerving thy heart and trembling hand to

CONFLICTS OF INTEREST

The authors declare no conflicts of interest.

AUTHOR CONTRIBUTIONS

TH wrote the paper, made the figures, performed the sequence analyses and contributed to study design. BW analyzed the chromovirus and assigned it to the CRM clade. IW and SL hybridized the probes along the chromosomes. IW performed the Southern hybridizations, and encouraged a stronger focus on the polyploid genomes. CO initially identified and cloned the tandem repeat sequences, as well as selected enzymes suitable for restriction. TS contributed to the research design and paper writing.

DATA AVAILABILITY STATEMENT

All data to availability and reproduce our findings are available as Supporting Information alongside the manuscript. Specifically, consensus sequences of the tandem repeats are available as Data S1. We have re-analyzed publically available sequence data as indicated in the Experimental procedures section. These reads are stored in the European Bioinformatics Institute (EBI) short read archive (accessions SRR4425239, SRR4425238, DRR057249, DRR057268). We deposited sequences of satellite hybridization probes online at the European Nucleotide Archive under the accessions LR215734 to LR215739 (study: http://www.ebi.ac.uk/ena/data/view/PRJEB31131).

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this article.

REFERENCES

Arumuganathan, K. and Earle, E. (1991) Nuclear DNA content of some important plant species. Plant Mol. Biol. Rep. 9, 209-218.
Avila Robledillo, L., Koblitzkova, A., Novak, P., Bottinger, K., Vrbova, I., Neumann, P., Schubert, I. and Macas, J. (2018) Satellite DNA in Vicia faba is characterized by remarkable diversity in its sequence composition, association with centromeres, and replication timing. Sci. Rep. 8, 5838.
Belyaev, A., Josefová, J., Jandová, M., Kalendar, R., Krak, K. and Mandák, B. (2019) Natural history of a satellite DNA family: from the ancestral genome component to species-specific sequences, concerted and non-concerted evolution. Int. J. Mol. Sci. 20, 1201.

Benjamini, Y. and Speed, T.P. (2012) Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic Acids Res. 40, e72.

Benson, G. (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Res. 27, 573-580.

Bourque, G., Burns, K.H., Gehring, M. et al. (2018) Ten things you should know about transposable elements. Genome Biol. 19, 199.

Bracewell, R., Chatla, K., Nalley, M.J. and Bachtrog, D. (2020) Evolution of repetitive DNA in eukaryotes. Mol. Biol. Evol. 37, 1691-1704.

Cechova, M., Harris, R.S., Tomaszkiewicz, M., Arbeithuber, B., Chiaromonte, F. and Makova, K.D. (2019) High satellite repeat turnover in great apes studied with short- and long-read technologies. Mol. Biol. Evol. 36, 2415-2431.

Charlesworth, B., Sniegowski, P. and Stephan, W. (1994) The evolutionary dynamics of repetitive DNA in eukaryotes. Nature, 371, 215-220.

Chen, Y.C., Liu, T., Yu, C.H., Chiang, T.Y. and Hwang, C.C. (2013) Effects of GC bias in next-generation-sequencing data on de novo genome assembly. PLoS ONE, 8, e62856.

Cheng, Z., Dong, F., Langdon, T., Ouyang, S., Buell, C.R., Gu, M., Blattner, F.R. and Jiang, J. (2002) Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon. Plant Cell, 14, 2679-2692.

Chester, M., Leitch, A.R., Soltis, P.S. and Soltis, D.E. (2010) Review of the application of modern cytogenetic methods (FISH/GISH) to the study of reticulation (polyploidy/hybridisation). Genes, 1, 168.

Clouse, J.W., Adhikary, D., Page, J.T., Ramaraj, T., Deyholos, M.K., Uddal, J.A., Fairbanks, D.J., Jellen, E.N. and Maughan, P.J. (2016) The amaranth genome: genome, transcriptome, and physical map assembly. Plant Genome, 9, 1-14.

D’Hont, A. (2005) Unraveling the genome structure of polyploids using FISH and GISH; examples of sugarcane and banana. Cytogenet. Genom. Res. 109, 27-33.

Diliberto, T.J., Rossen, J., Andres, T.C. and Williams, D.E. (2007) Precarismatic adoption of peanut, squash, and cotton in northern Peru. Science, 316, 1890-1893.

Dohm, J.C., Minoche, A.E., Holtgraves, D. et al. (2014) The genome of sugar beet (Beta vulgaris) a recently domesticated crop. Nature, 505, 546-549.

Dover, G. (1982) Molecular drive: a cohesive mode of species evolution. Nature, 299, 111-117.

Dover, G. (2002) Molecular drive. Trends Genet. 18, 587-589.

Edgar, R. (2004) MUSCLE: a multiple sequence alignment method with reduced time and space complexity. BMC Bioinformatics, 5, 113.

Edgar, P.P., Smith, R., McKain, M.R. et al. (2017) Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower. Plant Cell, 29, 2150-2167.

Elliott, T.A. and Gregory, T.R. (2015) Do larger genomes contain more diverse transposable elements? BMC Evol. Biol. 15, 69.

Fry, K. and Salser, W. (1977) Nucleotide sequences of HS-alpha satellite DNA from kangaroo rat Dipodomys ordi and characterization of similar sequences in other rodents. Cell, 12, 1069-1084.

Fuentes-Bazan, S., Mansion, G. and Borsch, T. (2012a) Towards a species level tree of the globally diverse genus Chenopodium (Chenopodiaceae). Mol. Phylog. Evol. 62, 359-374.

Fuentes-Bazan, S., Uotila, P. and Borsch, T. (2012b) A novel phylogeny-based generic classification for Chenopodium sensu lato, and a tribal rearrangement of Chenopodiaceae (Chenopodiaceae). Willdenowia, 42, 5-24.

Gao, D., Schmidt, T. and Jung, C. (2000) Molecular characterization and chromosomal distribution of species-specific repetitive DNA sequences from Beta corolliflora, a wild relative of sugar beet. Genome, 43, 1073-1080.

Garrido-Ramos, M.A. (2015) Satellite DNA in plants: more than just rubbish. Cytogenet. Genome Res. 146, 153-170.

Garrido-Ramos, A.M. (2017) Satellite DNA: an evolving topic. Genes, 8, 1-41.

Goubert, C., Modolo, L., Vieira, C., Valiente-Moro, C., Mavingui, P. and Boulesteix, M. (2015) De novo assembly and annotation of the Asian tiger mosquito (Aedes albopictus) repertoire with dnapipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (Aedes aegypti). Genome Biol. Evol. 7, 1192-1205.

Heslop-Harrison, J.S. and Schwarzacher, T. (2011) Organization of the plant genome in chromosomes. Plant J. 66, 18-33.

Heslop-Harrison, J.S., Schwarzacher, T., Ananthawat-Jonsson, K., Leitch, A.R., Shi, M. and Leitch, J.I. (1991) In-situ hybridization with automated chromosome denaturation. Technique, 3, 109-116.

Huerta-Cepas, J., Dopazo, J. and Gabaldon, T. (2010) ETE: a python environment for tree exploration. BMC Bioinformatics, 11, 24.

Iwata, A., Tek, A.L., Richard, M.M. et al. (2016) The amaranth genome: a recently domesticated crop. Plant Cell Rep. 35, 117-136.

Iwata, A., Tek, A.L., Richard, M.M. et al. (2018) Linear assembly of a human centromere on the Y chromosome. Nat. Biotechnol. 36, 321.

Javors, D.E., Ho, Y.S., Lightfoot, D.J. et al. (2017) The Genome of Chenopodium quinoa. Nature, 542, 307-312.

Jiang, J. (2019) Fluorescence in situ hybridization in plants: recent developments and future applications. Chromosome Res. 27, 153-165.

Jiang, J., Birchler, J.A., Parrott, W.A. and Dawe, R.K. (2003) A molecular view of plant centromeres. Trends Plant Sci. 8, 570-575.

Jurka, J., Kapitonov, V.V., Pavlicek, A., Klonowski, P., Kohany, O. and Waliuchewicz, J. (2005) Repeatbase update, a database of eukaryotic repetitive elements. Cytogenet Genome Res. 110, 462-467.

Kaderet, G., Mavrodiev, E.V., Zacharias, E.H. and Sukhorukov, A.P. (2010) Molecular phylogeny of Atripliceae (Chenopodiaceae, Chenopodiaceae): implications for systematicatics, biogeography, flower and fruit evolution, and the origin of C4 photosynthesis. Am. J. Bot. 97, 1664-1687.

Katoch, K. and Standley, D.M. (2013) MAFFT multiple sequence alignment software version 7: improvements in performance and usability. Mol. Biol. Evol. 30, 712-780.

Kearse, M., Moir, R., Wilson, A. et al. (2012) Geneious basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics, 28, 1467-1469.

Kelly, L.J., Renny-Byfield, S., Pellicer, J. et al. (2015) Analysis of the giant genomes of Fritillaria (Liliaceae) indicates that a lack of DNA removal characterizes extreme expansions in genome size. New Phytol. 208, 596-607.

Khosh, D.E., Eickbush, D.G. and Larracuente, A.M. (2017) Single-molecule sequencing resolves the detailed structure of complex satellite DNA loci in Drosophila melanogaster. Genome Res. 27, 709-721.

Kirov, I.V., Kiseleva, A.V., Van Luere, K., Van Roy, N. and Khrustaleva, L.I. (2017) Tandem repeats of Allium fistulosum associated with major chromosomal landmarks. Mol. Genet. Genomics, 292, 453-464.

Kolano, B., Gardunia, B.W., Michalska, M., Bonifacio, A., Fairbanks, D., Maughan, P.J., Coleman, C.E., Stevens, M.R., Jellen, E.N. and Maluszynska, S. (2011) Chromosomal localization of two novel repetitive sequences isolated from the Chenopodium quinoa Willd. Genome, 54, 710-717.

Kolano, B., Tomczak, H., Molewska, R., Jellen, E.N. and Maluszynska, J. (2012) Distribution of SS and 35S rRNA gene sets in 34 Chenopodium species (Amaranthaceae). Bot. J. Linn. Soc. 170, 220-231.

Kolano, B., Bednara, E. and Weiss-Schneeweiss, H. (2013) Isolation and characterization of reverse transcriptase fragments of LTR retrotransposons from the genome of Chenopodium quinoa (Amaranthaceae). Plant Cell Rep. 32, 1575-1589.
Kolano, B., McCann, C., Orzechowska, M., Siwińska, D., Temsch, E. and Weiss-Schneeweiss, H. (2016) Molecular and cytogenetic evidence for an allotetraploid origin of Chenopodium quinoa and C. berlandieri (Amaran- 
thaceae). Mol. Phylogenet. Evol. 100, 109–123.
Kolano, B., McCann, J., Oskseda, M., Chrapień, M., Rojek, M., Nobis, A. and Weiss-Schneeweiss, H. (2018) Parental origin and genome evolution of several Eurasian hexaploid species of Chenopodium (Cenopodiaceae). Phytotaxa, 392, 23.
Kopećna, O., Kubičková, S., Cernohorská, K., Havelová, J. and Kovarík, A. (2016) Isoforms and paralogs in C20 genes. Nature Commun. 7, 1547–1549.
Liu, G., Li, X., Zhou, X., Li, M., Zhang, F., Schwarzer, M. and FastHap- 
ris, J.S. (2019) The repetitive DNA landscape in Avena (Poaceae): chro- 
mosome and genome evolution defined by major repeat classes in whole-genome sequence reads. BMC Plant Biol. 19, 226.
Lower, S.S., McGurrin, M.P., Clark, A.G. and Barbas, D.A. (2018) Satellite DNA evolution: old ideas, new approaches. Curr. Opin. Genet. Dev. 49, 70–78.
Ma, J., Devos, K.M. and Bennetzen, J.L. (2004) Analyses of LTR-retrotrans- 
poson structures reveal recent and rapid genomic DNA loss in rice. Mol. Biol. Evol. 21, 860–869.
Ma, J., Neumann, P. and Navrátlová, A. (2007) Repetitive DNA in the pea (Pisum sativum L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and Medicago truncatula. BMC Genomics, 8, 427.
Mandak, B., Krak, K., Vit, P., Pavičiková, Z., Lomonosova, M.N., Habibi, F., Wang, L., Jellen, E.N. and Douda, J. (2016) How genome size variation is linked with evolution within Chenopodium sensu lato. Perspect. Plant Ecol. Evol. Syst. 23, 18–32.
Mandak, B., Krak, K., Vit, P., Lomonosova, M.N., Belyaev, A., Habibi, F., Wang, L., Douda, J. and Storchova, H. (2018) Hybridization and polyploidyization within the Chenopodium album aggregate analysed by means of cytologi- 
cal and molecular markers. Mol. Phylogenet. Evol. 129, 189–201.
Mandáková, T., Kovář, A., Zozomová-Lihová, J., Shimizu-Itatsugi, R., Shi- 
mizu, K.K., Mummennhoff, K., Marhold, K. and Lysák, M.A. (2013) The more the merrier: recent hybridization and polyploidy in Cardaminaceae. Plant Cell, 25, 3290–3295.
Mangelson, H., Jarvis, D.E., Mellingo, P., Rollano-Penaloz, O.M., Palma- 
Encinas, V.D., Gomez-Pando, L.R., Jellen, E.N. and Maughan, P.J. (2019) The genome of Chenopodium pallidicaule: an emerging Andean super 
grain. Appl. Plant Sci 7, e11300.
Marchler-Bauer, A., Lu, S., Anderson, J.B. et al. (2011) CDD: a Conserved 
Domain Database for the functional annotation of proteins. Nucleic Acids Res. 39, D225–D229.
Maughan, P.J., Kolano, B.A., Maluszczyńska, J. et al. (2006) Molecular and cytological characterization of ribosomal RNA genes in Chenopodium quinoa and Chenopodium berlandieri. Genome, 49, 825–839.
Maughan, P.J., Chaney, L., Lightfoot, D.J., Cox, B.J., Tester, M., Jellen, E.N. and Jarvis, D.E. (2019) Mitochondrial and chloroplast genomes provide insights into the evolutionary origins of quinoa (Chenopodium quinoa). Wildl. Sci. Rep. 9, 185.
McCann, J., Jang, T.-S., Macas, J., Schneeweiss, G.M., Matzke, N.J., Novák, P., Stuessy, T.F., Villasenor, J.L. and Weiss-Schneeweiss, H. (2018) Dat- 
ing the species network: Allopolyploidy and repetitive DNA evolution in American daisies (Melampodium sect. Melampodium, Asteraceae). Syst. Biol. 67, 1010–1024.
Mehrotra, S. and Goyal, V. (2014) Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. Genomics Proteomics Bioinformatics, 12, 164–171.
Melamed-Bessudo, C. and Levy, A.A. (2012) Deficiency in DNA methylation increases meiotic crossovers rates in euchromatic but not in heterochro- 
matic regions in Arabidopsis. Proc. Natl Acad. Sci. USA, 109, E891.
Melters, D.P., Bradnam, K.R., Young, H.A. et al. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biol. 14, R10.
Mihiri, C., Paridos, C., Daniel, J., Petit, M., Lim, K.Y., Dorihac de Borne, F., Kovařík, A., Leitch, A.R. and Grandbastien, M.A. (2019) Parental trans- 
posable element loads influence their dynamics in young Nicotiana hybrids and allotetraploids. New Phytol. 221, 1619–1633.
Miga, K.H. (2015) Completing the human genome: the progress and chal- 
lenge of satellite DNA assembly. Chromosome Res. 23, 421–426.
Neumann, P., Navratilová, A., Koblízková, A., Kejnovský, E., Hřibová, E., Hobza, R., Widmer, A., Dolezel, J. and Macas, J. (2011) Plant cen- 
 tromeric retrotransposons: a structural and cytogenetic perspective. Mol. DNA, 2, 4.
Neumann, P., Novák, P., Hostáková, N. and Macas, J. (2019) Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classi- 
fication. Mol. DNA, 19, 1–17.
Novák, P., Neumann, P. and Macas, J. (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. BMC Bioinformatics, 11, 378.
Novák, P., Neumann, P., Pech, J., Steinhaist, J. and Macas, J. (2013) Repea 
tExplorer: a galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Mol. Bioinformatics, 29, 792–793.
Novikova, O. (2009) Chromodomains and LTR retrotransposons in plants. Commun. Integr. Biol. 2, 158–162.
Olivier, K.R., McBimb, J.A. and Greene, W.K. (2013) Transposable elements: powerful contributors to angiosperm evolution and diversity. Genome Biol. 5, 1886–1901.
Orzechowska, M., Majka, M., Weiss-Schneeweiss, H., Kovařík, A., Bor- 
owska-Zuchowska, N. and Kolano, B. (2018) Organization and evolution of two repetitive sequences, ‘18-24J and 12-13P’, in the genome of Che- 
 nopodium (Amaranthaceae). Genome, 61, 643–652.
Paesold, S., Borchardt, D., Schmidt, T. and Dechyeva, D. (2012) A sugar beet (Beta vulgaris L.) reference FISH karyotype for chromosome and chromosome-arm identification, integration of genetic linkage groups and analysis of major repeat family distribution. Plant J. 72, 600– 
611.
Palomino, G., Hernández, L.T. and de la Cruz Torres, E. (2008) Nuclear gene 
nome size and chromosome analysis in Chenopodium quinoa and C. ber- 
landieri subsp. nuttalliae. Euphytica, 164, 221.
Peona, V., Weissenstein, M.H. and Suh, A. (2018) How complete are ‘com- 
plete’ genome assemblies? – An avian perspective. Mol. Ecol. Resour. 18, 1188–1195.
Pichlová, M., Mestrovic, N. and Mravincová, B. (2012) Satellite DNA evolution. Genome Dyn. 7, 126–152.
Renny-Byfield, S., Kovář, A., Chester, M., Nichols, R.A., Macas, J., Novák, P. and Leitch, A.R. (2012) Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allotetraploids of Nicotiana taba- 
la. Nature Commun. 3, 1891.
Ruiz-Ruano, F.J., Navarro-Dominguez, B., Camacho, J.P.M. and Garrido- 
Ramos, M.A. (2016) Characterization of the satelliteome in lower vascular 
plants: the case of the endangered fern Vandenboschia speciosa. Ann. Bot. 123, 587–599.
Sambrook, J., Fritsch, E.F. and Maniatis, T. (1989) Molecular Cloning: A Lab-

ary Manual, 2nd edn. New York, NY: Cold Spring Harbor Laboratory Press.
Sanchez, D.H., Gaubert, H., Drost, H.-G., Zabet, N.R. and Paszkowski, J. (2017) High-frequency recombination between members of an LTR retrotranspon family during transposition bursts. Nature Commun. 8, 1283.
Schmidt, T., Schwarzer, T. and Heslop-Harrison, J.S. (1994) Physical mapping of RNA genes by fluorescent in-situ hybridization and

© 2020 The Authors. The Plant Journal published by Society for Experimental Biology and John Wiley & Sons Ltd, The Plant Journal, (2020), doi: 10.1111/tjpl.14705
structural analysis of SS rRNA genes and intergenic spacer sequences in sugar beet (Beta vulgaris). Theor. Appl. Genet. 88, 629-636.

Schmidt, T., Heitkam, T., Liedtke, S., Schubert, V. and Menzel, G. (2019) Adding color to a century-old enigma: multi-color chromosome identification unravels the autotriploid nature of saffron (Crocus sativus) as a hybrid of wild Crocus cartwrightianus cytotypes. New Phytol. 222, 1965-1980.

Schwarzacher, T. and Heslop-Harrison, P. (2000) Practical in situ Hybridization. Oxford: BIOS Scientific Publishers.

Schwarzacher, T., Leitch, A.R., Bennett, M.D. and Heslop-Harrison, J.S. (1989) In situ localization of parental genomes in a wide hybrid. Annu. Bot. 64, 315-324.

Selbt, K.M., Schmidt, T. and Heitkam, T. (2018) FlexiDot: highly customizable, ambiguity-aware dotplots for visual sequence analyses. Bioinformatics, 34, 3575-3577.

Serrato-Capuchina, A. and Matute, D.R. (2018) The role of transposable elements in speciation. Genes, 9, 254.

Sevilm, V., Bashir, A., Chin, C.-S. and Miga, K.H. (2016) Alpha-CENTAUR: assessing novel centromeric repeat sequence variation with long read sequencing. Bioinformatics, 32, 1921-1924.

Shapiro, J.A. and von Sternberg, R. (2005) Why repetitive DNA is essential to genome evolution. Biol. Rev. 80, 227-250.

Stapley, J., Sautre, A.W. and Dennis, S.R. (2015) Transposable elements as agents of rapid adaptation may explain the genetic paradox of invasive species. Mol. Ecol. 24, 2241-2252.

Storchova, H., Drabosova, J., Chab, D., Kolar, J. and Jellen, E.N. (2015) The introns in FLOWERING LOCUS T LIKE (FTL) genes are useful markers for tracking paternity in tetraploid Chenopodium quinoa. Genet. Resour. Crop Evol. 62, 913-925.

Symonova, R., Ocalewicz, K., Kirtiklis, L., Dalmastro, G.B., Pelikánova, S., Garcia, S. and Kovarik, A. (2017) Higher-order organization of extremely amplified, potentially functional and massively methylated 5S rDNA in European pikes (Esox sp.). BMC Genomics, 18, 391.

Vergnaud, G. and Denoeud, F. (2000) Minisatellites: mutability and genome architecture. Genome Res. 10, 899-907.

Vicient, C.M. and Casacuberta, J.M. (2017) Impact of transposable elements on polyploid plant genomes. Annu. Bot. 120, 195-207.

Viguera, E., Caneille, D. and Ihrlich, S. (2001) Replication slippage involves DNA polymerase pausing and dissociation. EMBO J. 20, 2567-2595.

Vondrak, T., Avila Robledillo, L., Novák, P., Kobližková, A., Neumann, P. and Macas, J. (2020) Characterization of repeat arrays in ultra-long nanopore reads reveals frequent origin of satellite DNA from retrotransposon-derived tandem repeats. Plant J. 101, 484-500

Walsh, B.M., Adhikary, D., Maughan, P.J., Emshwiller, E. and Jellen, E.N. (2015) Chenopodium polyploidy inferences from Salt Overly Sensitive 1 (SOS1) data. Am. J. Bot. 102, 533-543.

Ward, S.M. (2000) allotetraploid segregation for single-gene morphological characters in quinoa (Chenopodium quinoa Willd.). Euphytica, 116, 11-16.

Weber, B. and Schmidt, T. (2009) Nested Ty3-gypsy retrotransposons of a single Beta procombens centromere contain a putative chromodomain. Chromosome Res. 17, 379-396.

Weber, B., Heitkam, T., Holtgrawe, D., Weisshaar, B., Minoche, A.E., Dohm, J.C., Himmelbauer, H. and Schmidt, T. (2013) Highly diverse chromoviruses of Beta vulgaris are classified by chromodomains and chromosomal integration. Mol. DNA, 4, 8.

Weiss-Schneeweiss, H., Leitch, A.R., McCann, J., Jang, T.-S. and Macas, J. (2015) Employing next generation sequencing to explore the repeat landscape of the plant genome. In Next-Generation Sequencing in Plant Systematics (Horandl, E. and Appelhans, M.S. eds), Königstein, Germany: International Association for Plant Taxonomy, pp. 1-25.

Wheeler, T.J. and Eddy, S.R. (2013) nhmmer: DNA homology search with profile HMMs. Bioinformatics, 29, 2487-2489.

Wickham, H. (2016) ggplot2: Elegant graphics for data analysis. New York, NY: Springer-Verlag.

Wolrab, C., Heitkam, T., Holtgrawe, D., Weisshaar, B., Minoche, A.E., Dohm, J.C., Himmelbauer, H. and Schmidt, T. (2012) Evolutionary reshuffling in the Errantivirus lineage Elbe within the Beta vulgaris genome. Plant J. 72, 636-651.

Xu, C., Jiao, C., Sun, H. et al. (2017) Draft genome of spinach and transcriptome diversity of 120 Spinacia accessions. Nat. Commun. 8, 15275.

Yasui, Y., Hirakawa, H., Oikawa, T. et al. (2016) Draft genome sequence of an inbred line of Chenopodium quinoa, an allotetraploid crop with great environmental adaptability and outstanding nutritional properties. DNA Res. 23, 535-546

Zakrzewski, F., Weber, B. and Schmidt, T. (2013) A molecular cytogenetic analysis of the structure, evolution, and epigenetic modifications of major DNA sequences in centromeres of Beta species. In Plant Cen tromere Biology (Jiang, J. and Birchler, J.A. eds). Ames, IA: John Wiley & Sons, Inc, pp. 39-55.

Zhang, T., Talbort, P.B., Zhang, W., Wu, Y., Yang, Z., Henikoff, J.G., Henikoff, S. and Jiang, J. (2013) The CentOx satellite confers translational and rotational shading on CenH3 nucleosomes in rice centromeres. Proc. Natl. Acad. Sci. USA 110, E4875-E4883.

Zhang, Y., Cheng, C., Li, L., Yang, S., Wang, Y., Li, Z., Chen, J. and Lou, Q. (2016) Chromosomal structures and repetitive sequences divergence in Cucumis species revealed by comparative cytogenetic mapping. BMC Genomics, 16, 730.

Zhou, H.C., Waminal, N.E. and Kim, H.H. (2019) In silico mining and FISH mapping of a chromosome-specific satellite DNA in Capsicum annuum L. Genes Genomics, 41, 1001-1006.

Zou, C., Chen, A., Xiao, L. et al. (2017) A high-quality genome assembly of quinoa provides insights into the molecular basis of salt bladder-based salinity tolerance and the exceptional nutritional value. Cell Res. 27, 1327.

Zúñiga-Silva, A., Fuentes, F., Zamora, P., Jacobsen, S.-E. and Schwemmer, A.R. (2014) Breeding quinoa (Chenopodium quinoa Willd.): potential and perspectives. Mol. Breed. 34, 13-30.