

Evolutionary dynamics of transposable elements and satellite DNAs in polyploid Spartina species

Delphine Giraud, Oscar Lima, Virginie Huteau, Olivier Coriton, Julien Boutte, Ales Kovarik, Andrew R Leitch, Ilia J Leitch, Malika Aïnouche, Armel Salmon

▶ **To cite this version:**

Delphine Giraud, Oscar Lima, Virginie Huteau, Olivier Coriton, Julien Boutte, et al.. Evolutionary dynamics of transposable elements and satellite DNAs in polyploid Spartina species. Plant Science, 2021, 302, pp.110671. 10.1016/j.plantsci.2020.110671 . hal-03101741

HAL Id: hal-03101741
https://hal.science/hal-03101741

Submitted on 17 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Journal Pre-proof

Evolutionary dynamics of transposable elements and satellite DNAs in polyploid *Spartina* species

Delphine Giraud[a], Oscar Lima[a], Virginie Huteau[b], Olivier Coriton[b], Julien Boutte[a,c], Ales Kovarik[d], Andrew R. Leitch[e], Ilia J. Leitch[f], Malika Aïnouche[a], Armel Salmon[a]†

[a] UMR CNRS 6553 ECOBIO, Université de Rennes 1, F-35042 Rennes Cedex, France
delphine.giraud@univ-rennes1.fr
oscar.lima@univ-rennes1.fr
malika.ainouche@univ-rennes1.fr
armel.salmon@univ-rennes1.fr

[b] Plateforme de cytogénétique moléculaire végétale, INRAE, Université de Rennes 1, Agrocampus Ouest, IGEPP, F-35650 Le Rheu, France
olivier.coriton@inrae.fr
virginie.huteau@inrae.fr

[c] INRAE, Université de Rennes 1, Agrocampus Ouest, IGEPP, F-35650 Le Rheu, France
boutte.julien@gmail.com

[d] Institute of Biophysics, Academy of Sciences of the Czech Republic, Brno CZ-61265, Czech Republic
kovarik@ibp.cz

[e] School of Biological and Chemical Sciences, Queen Mary University of London, London, E1 4NS, UK
a.r.leitch@qmul.ac.uk

[f] Department of Comparative Plant and Fungal Biology, Royal Botanic Gardens, Kew TW9 3DS, UK
I.Leitch@kew.org

† Corresponding author: Armel Salmon (Email: armel.salmon@univ-rennes1.fr)
Université de Rennes 1
Centre National de la Recherche Scientifique
UMR 6553 Ecosystems-Biodiversity-Evolution
Campus de Beaulieu, Bâtiment 14A
263 avenue du Général Leclerc

F-35042 Rennes Cedex, France

Highlights

- First repeat annotations from high-throughput genome sequencing in polyploid Spartina
- Repeat dynamics among *Spartina* species is correlated with phylogenetic relationships
- Genome evolution is determined by divergence times rather than by ploidy levels.
- Repeat patterns of the tetraploid *S. spartinae* suggest relationships with hexaploids

Journal Pre-proof

- Burst of novel 370pb-satellite DNA representing 6.89% of the *S. alterniflora* genome

Abstract

Repeated sequences and polyploidy play a central role in plant genome dynamics. Here, we analyze the evolutionary dynamics of repeats in tetraploid and hexaploid *Spartina* species that diverged during the last 10 million years within the Chloridoideae, one of the poorest investigated grass lineages. From high-throughput genome sequencing, we annotated *Spartina* repeats and determined what sequence types account for the genome size variation among species. We examined whether differential genome size evolution correlated with ploidy levels and phylogenetic relationships. We also examined the tempo of repeat sequence dynamics associated with allopatric speciation over the last 3-6 million years between hexaploid species that diverged on the American and European Atlantic coasts and tetraploid species from North and South America. The tetraploid *S. spartinae*, whose phylogenetic placement has been debated, exhibits a similar repeat content as hexaploid species, suggesting common ancestry. Genome expansion or contraction resulting from repeat dynamics seems to be explained mostly by the contrasting divergence times between species, rather than by genome changes triggered by ploidy level change *per se*. One 370 bp satellite may be exhibiting 'meiotic drive' and driving chromosome evolution in *S. alterniflora*. Our results provide crucial insights for investigating the genetic and epigenetic consequences of such differential repeat dynamics on the ecology and distribution of the meso- and neopolyploid *Spartina* species.

Keywords: genome dynamics, polyploidy, transposable elements, satellite DNA, *Spartina*

1. Introduction

Repetitive sequences, including dispersed repeats such as transposable elements (TEs), or tandem repeats such as satellite DNA, represent an important fraction of plant genomes that impact evolutionary dynamics [1–3]. Together with polyploidy (whole genome duplication), repetitive sequences are the key drivers of genome size variation in plants, where up to c. 2400-fold variation may be observed [4–9]. In angiosperms, the repeat content, much of which comprises TEs, ranges from 18% in *Arabidopsis thaliana* to 80% in *Triticum aestivum* and 85% in *Zea mays* [10–14]. These elements are classified according to their transposition mode: retrotransposons (or Class I elements) replicate *via* a "copy-paste" process (resulting in genome size variation), while DNA transposons (or Class II elements) replicate *via* a "cut-paste" process [15–17]. Transposable elements may accumulate in various coding and non-coding genomic regions following transposition bursts, and may be subject to differential expansion or loss during species evolution [1,18–22]. Thus, they play an important role in the dynamics of expansion and contraction of plant genomes that may be observed even between closely related species [23–26]. Transposable elements not only shape genome structure, but can also substantially affect gene expression, when inserted in coding or regulatory regions [27–29]. To prevent TE proliferation and their putative consequences on gene regulation and function, TE expression is controlled by several epigenetic silencing pathways, such as mRNA cleavage, small RNA synthesis and DNA methylation (via transcriptional and post-transcriptional gene silencing) [30–33].

A notable feature that has emerged from comparative studies of TE activity is that TE amplification does not appear to be constant, with some TE lineages showing evidence of periods of transposition activity, alternating with periods of relative quiescence [26,34]. Several studies have suggested that TE activity is influenced by environmental factors, such as biotic or abiotic stress [35–38].

Allopolyploidy, which combines interspecific hybridization with polyploidy, is widespread in plants. Both the merging of the different species genomes into one nucleus and polyploidy event itself have the potential to

Journal Pre-proof

impact TE regulatory processes [26,39,40], although it remains unclear which process has the most significant effect on TE activity. Altered patterns of epigenetic and TE transcriptional activities from expectation (i.e. the sum of the activity seen in the diploid parents) have been observed in both experimentally resynthesized hybrids and allopolyploids (*e.g. Arabidopsis* [41,42], wheat [27,43], tobacco [44,45]) and natural hybrids and allopolyploids [46,47]. Most of these studies suggest that the immediate effects of allopolyploidy have the most impact on TE regulatory processes, potentially leading to bursts of TE activity, although disrupted patterns may stabilize within a few generations [48]. Nevertheless, allopolyploids of different ages (*i.e*. recent to millions of years), show evidence for differential TE amplification and elimination [49–53]. It might be expected that genome redundancy in polyploids would buffer against the mutagenic impact of TE activity [19,26,40]. The short-term responses of allopolyploid genomes to different TE composition inherited from their parental species is also expected to influence subgenome dominance and the fractionation process which contribute to the long-term divergence of polyploids [54,55].

The genus *Spartina* Schreb (Poaceae) offers excellent opportunities for comparative genomic analyses, enabling patterns and tempos of repetitive sequence evolution to be deciphered. Since the last whole genome duplication event (WGD) in the ancestry of family Poaceae [56,57](Van de Peer et al 2017, Zwaenepoel and Van de Peer 2019) polyploidy events in *Spartina* have given rise to 15 species, which range from tetraploid to dodecaploid, with evidence of extensive reticulate evolution in the species' ancestries (resulting from hybridization, introgression and allopolyploidy) [58–62]. No diploid *Spartina* species are known. Species in *Spartina* are perennial and mostly colonize coastal salt marshes, where they are considered as "ecosystem engineers" because they increase sediment accretion. *Spartina* species are particularly tolerant to salt stress, drought, anoxic conditions and pollution by heavy metals and hydrocarbons [63–67].

Phylogenetic analyses have revealed that *Spartina* forms a distinctive clade within genus *Sporobolus*, rendering *Sporobolus* polyphyletic [68]. Patterns of branching within the phylogenetic trees indicate that *Spartina* diversified within the last 12-20 million years [68–72] and comprises two main subclades, that diverged 6-10 million years ago [72]: (i) a tetraploid lineage (2n=4x=40) including New World species, and (ii) a hexaploid lineage (2n=6x=60, 62) which includes the North-American *S. alterniflora*, *S. foliosa* and the European species *S. maritima* (Fig. 1 and Table 1). The phylogenetic affinities of the South American tetraploid species *S. spartinae* (syn. *S. argentinensis*) remains unresolved, being placed alternatively as sister to the hexaploid lineage or sister to all other *Spartina* species [69,72]. Hybridization within and between these subclades has given rise to additional ploidy levels (7x, 8x, 9x, 12x, including a classical example of neopolyploid speciation reviewed in Ainouche *et al*. (2012).

Cytogenetic investigations reveal that *Spartina* genome sizes (estimated using flow cytometry) vary between and within ploidy levels. For example, amongst tetraploids (2n=40), genome sizes vary from 1.45 pg/2C in *S. bakeri* [73] to 1.56 pg/2C in *S. pectinata* [74] and 2.00 pg/2C in *S. spartinae* [75]. These data reveal lineage-specific genome size changes have occurred in the recent evolutionary history of these polyploid species.

To date, no exhaustive investigations have been undertaken to evaluate the nature and dynamics of repetitive sequences between different species of *Spartina* that diversified since tetra- and hexa-mesoploid events in the last 6-10MYs. Instead, previous studies have examined the dynamics of targeted TEs following recent allopolyploidy. For example, using Transposon-display approaches [76] it was shown that there have

Journal Pre-proof

been no transposition burst associated with the divergence of the neoallododecaploid *S. anglica* from its hexaploid parental species *S. alterniflora* and *S. maritima*. In contrast, methylation-sensitive AFLP analyses revealed that hybridization (in the F1 sterile hybrid *S. x townsendii*), was accompanied by substantial methylation changes, mostly in regions flanking TEs [47], that were inherited in the formation of *S. anglica* [77].

In this paper, we take advantage of high-throughput (Illumina) genome sequencing to explore the evolutionary dynamics of repetitive sequences in tetraploid and hexaploid *Spartina* species with contrasting genome sizes. Of particular interest to this analysis is that some of these species have been involved in the formation of recent hybrids and neoallopolyploids. We first evaluate the relative amounts of the repetitive compartment in the different polyploid lineages, and determine what sequence types account for genome size variation among species. This allowed us to examine whether the differential evolution of repetitive sequences is correlated with ploidy levels, phylogenetic relationships, or the known divergence times between lineages.

2. Material and methods

2.1 Plant material

Five *Spartina* species were selected for analysis representing the tetraploid (4x) and hexaploid (6x) parental lineages that gave rise to higher ploidy levels (7x, 8x, 9x 12x) observed in the genus [78]. The two sister hexaploid species *S. maritima* (Curtis) Fernald (2n=6x=60; Fig. 1, see Table 1 for synonyms and nomenclature) and *S. alterniflora* Loisel. (2n=6x=62) were selected for analysis because of their involvement in the recent (i.e. < c. 130 years ago) allopolyploid speciation event that gave rise to *S. anglica* [58,69,71,79]. Three tetraploid species (2n=4x=40) were also selected. These being two closely related-species *S. versicolor* (Fabre) and *S. bakeri* Merr., that diverged recently from each other (less than 2 million years ago (Mya), Fig. 1). A third, more distantly related tetraploid species, *S. spartinae* (Trin.) Merr., was studied because it exhibits the highest tetraploid genome size reported so far (1.97pg/2C, based on our own genome size estimate, see below).

The hexaploid species were sampled along the French Atlantic coast from their natural environment: *S. maritima* on the Etel river marshes (Presqu'île du Verdon, Morbihan, France) and *S. alterniflora* on the Faou river marshes (Le Faou, Finistère, France). *S. maritima* and *S. alterniflora* were also sampled in Southern England to detect putative variations in repeat accumulation between populations of the same species. *S. maritima* was collected at the Isle of Wight (UK) and *S. alterniflora* at Marchwood (UK). The tetraploid *S. versicolor* was collected on the Mediterranean coast at Hyères (Var, France), *S. bakeri* in Florida (Miami Botanical Garden, USA) and *S. spartinae* in the Santa Fe Province (Argentina). The plants were maintained in the greenhouse at the University of Rennes 1 before analysis.

2.2 Flow cytometry

The ploidy level and genome size of each species was assessed using flow cytometry at the Imagerie-Gif facilities (Institute for Integrative Biology of the Cell, UMR 9198 CNRS-CEA, Gif sur Yvette, France) using propidium iodide (50µg/ml) to stain the nuclei following established protocols [73]. Briefly, fresh leaves were chopped in buffer containing 45mM $MgCl_2$, 30mM sodium-citrate, 60mM MOPS (pH 7.0), 1% PVP-10, 0.1% Triton X-100 and 10mM sodium metabisulfite ($S_2O_5Na_2$). *Petunia hybrida* (2.85pg/2C) was used as the internal

Journal Pre-proof

calibration standard and the samples were analyzed on a CytoFLEX S flow cytometer (Beckman Coulter). Genome size measurements were conducted on the same plants that were used for sequencing.

2.3 Sequencing and *de novo* identification of repeated sequences

Repeated sequences for each species were *de novo* retrieved from Illumina genomic sequencing data. Genomic DNA was extracted from fresh leaves and Illumina libraries were prepared according to the Tru-Seq PCR-Free Protocol provided by Illumina. Libraries were sequenced (100-150 bp; paired-end with 500 bp-inserts) at the BGI (Beijing Genomics Institute) using the Hi-Seq X technology with a read depth ranging from 20x to 90x depending on species. Random samples of 1 million Illumina paired-end reads (total of 2 million reads by sample) were processed using the RepeatExplorer pipeline [80,81] to identify repeated sequences in each species in the absence of an assembled reference genome. This approach uses a similarity-based clustering of reads to build consensus sequences representing different types of repeated sequences. In *S. alterniflora*, the presence of satellite DNA sequences increased the computational and memory requirements for data analysis. Consequently, a random sampling of reads was performed on the 2 million reads sample and only a subset of 854 068 reads was analyzed by RepeatExplorer for this species (Table 2).

Prior to RepeatExplorer analyses, all reads deriving from chloroplast DNA and rDNA were removed from the samples. Reads of each species were aligned or mapped on their respective rDNA or chloroplast genome (previously assembled by Boutte et al. (2015)[82] and Rousseau-Gueutin et al. 2015 [72]) using BLAST+ (v.2.5.0; parameters: blastn with 90% identity along 90% of the reads length; [83]) and Bowtie2 (parameters: --score-min, local, G, 52, 8; [84]). Reads with more than 90% identity with chloroplast or ribosomal sequences were not used in the analyses described below.

Each sample of reads was analyzed independently with the RepeatExplorer pipeline using the Galaxy-web interface. All-to-all comparisons of reads and clustering were performed with default parameters (i.e. minimum of 40 nt overlap between reads with at least 80% identity in the overlapping region). Only clusters containing more than 200 reads (> 0.01% of the sampled reads) were analyzed further.

Reads assigned to each cluster were assembled with CAP3 [85] (included in the RepeatExplorer pipeline) to obtain consensus sequences (contigs) representing different TE families. CAP3 contigs were then co-assembled using Genome Assembler (Roche, v. 2.8; [86]) with the same settings as CAP3 (i.e. min. overlap of 40 nt with at least 80% identity in the overlapping region) to improve the assembly process and maximize contig length. Finally, repeated sequence databases for each species were established consisting of contigs obtained after Genome Assembler co-assembling with non-co-assembled CAP3 contigs.

2.4 Repeated sequence annotation and quantification

Different methods were used to annotate consensus sequences of each database. First annotations were performed using RepeatMasker (v. 4.0.0; http://repeatmasker.org). During this step, reads of each cluster were aligned via rmblastn (v. 2.2.23+; [83] against “Viridiplantae” sequences of Repbase (database version at 08/2016; [87,88]. To complete the annotation, assembled contigs were aligned with blastx against a TE-specific protein database (translation of Repbase “Viridiplantae” nucleic sequences into protein sequences). Satellite sequences were also identified with “self-to-self” sequence comparisons by carrying out dot-plot analyses using YASS [89].

Journal Pre-proof

For each species, the quantity of repeated sequences was estimated from the number of reads found within clusters representing different TE families (percentage of sampled reads). For validation of estimated abundances, a second evaluation was realized by mapping (i) the same sample of reads and (ii) a larger set of reads (approximately 330 millions reads per species) on the five *Spartina* repeated sequence databases using Bowtie2 (parameters: --score-min, local, G, 52, 8; [84]). Comparisons of both quantifications are given in Supplementary Table S1.

2.5 Southern blot hybridization

The relative abundance of two main satellite sequences (see Results) was checked by Southern blot hybridization in *S. alterniflora* and *S. maritima* (both from English populations). Genomic DNAs were digested with restriction endonucleases (2x 5 U /ug DNA, for 3 h each), separated by agarose gel electrophoresis and transferred to Hybond XL membranes (GE-Healthcare, Little Chalfont, UK) using alkaline capillary transfer. The probes were inserts of the Smar_CL1_1 and Salt_SALT28 clones. The Smar_CL_1 clone containing three units of a 167-bp satellite was obtained by PCR amplification of *S. maritima* DNA using primers derived from the *S. maritima* cluster 1 consensus sequence (Spartina_mar_sat_F: 5'CGGTCCCGTATATGAATTT3', Spartina_mar_sat_R: 5'GGAGTTTCGGTCAGTTTT3'). The Salt_SALT28 clone containing a single monomer of the 370-bp satellite was obtained by a classical method based on restriction digestion (BstNI) of *S. alterniflora* genomic DNA followed by gel separation, excision and purification of the 300-500 bp fraction. The fragments were ligated to a plasmid vector (pDrive, Qiagen) and clones screened by Southern hybridization with labeled *S. alterniflora* genomic DNA. Clones with strong signals were sequenced by the Sanger method using the SP6 primer. For labeling of probes, the inserts were amplified from plasmids using the universal Sp6 and T7 vector primers. About 100 ng of purified PCR product was radioactively labeled with ^{32}P-dCTP according to the manufacturer protocol (DecaLabel DNA Labeling Kit, Thermofisher, USA). Southern blot hybridization was carried out in a 0.25 M sodium phosphate buffer (pH 7.0) supplemented with 7% w/v sodium dodecyl sulphate (SDS) at 65°C. The membranes were washed with 2 x SSC (10 x SSC = 1.5 M NaCl, 0.15 M sodium citrate, pH 7.0), 0.1% SDS (twice for 5 min) and then with 0.2 x SSC and 0.1% SDS (twice for 15 min at 65°C). The membranes were exposed to a storage phosphor screen, scanned (Typhoon FLA7400, GEHealthcare), and the signal was quantified using Image Quant (GE-Healthcare).

2.6 Cytogenetic analyses

Cytogenetic analyses were conducted to determine the distribution of a highly repeated satellite sequence identified in the *S. alterniflora* genome (see Results). Primers were designed to amplify the repeat by PCR from *S. alterniflora* genomic DNA (5'-CCATGCGTGAGATGGTCATA and 5'-GCTTCTTCAGACTTGTGGGC). The product was labeled using random priming with biotin-16-dUTP and used as a probe in Fluorescence *in situ* hybridization (FISH) experiments. In addition, the probe, pTa 71 [90], which contains a 9-kb EcoRI fragment of the 35S rDNA repeat unit (18S-5.8S-26S genes and spacers) isolated from *Triticum aestivum*, was used as a control. This probe was labelled with Alexa-488 dUTP (Invitrogen, Life Technologies) by random priming.

Root tips from the five studied *Spartina* species were harvested and treated in the dark with 0.04% (w/v) 8-hydroxyquinoline for 2h at 4°C followed by 2 h at room temperature to accumulate metaphases. They were

Journal Pre-proof

then fixed in 3:1 ethanol-glacial acetic acid for 12 h at 4°C and stored in ethanol 70 % at -20 °C. Before use, root tips were washed in 0.01 M citric acid-sodium citrate pH 4.5 buffer for 15 min and digested in a solution of 5% Onozuka R-10 cellulase 1µM (Sigma) and 1% Y23 pectolyase 1µM (Sigma) at 37 °C for 3 h 45 min. The digested root tips were then carefully washed with distilled water for 30 min. One root tip was transferred to a slide and macerated with a drop of 3:1 fixative solution. The slides were dried at room temperature and stored at -20°C.

For FISH, chromosome preparations were incubated in RNase A (100ng/µL) and pepsin (0.05%) in 10 mM HCl, fixed with paraformaldehyde (4%), dehydrated in an ethanol series (70%, 90% and 100%) and air-dried. The hybridization mixture consisted of 50% deionized formamide, 10% dextran sulfate, 2 X SSC, 1% SDS, 100 ng of labeled probe targeting the repeat and 50 ng of labelled pTa71 targeting 35S rDNA. The mixture was denatured at 92°C for 6 min and transferred to ice. Chromosomes were denatured in a solution of 70% (v/v) formamide in 2X SSC at 70°C for 2 min. The denatured probe was placed on the slide and *in situ* hybridization carried out overnight in a moist chamber at 37°C. After hybridization, slides were washed for 5 min in 50% formamide in 2X SSC at 42°C, followed by several washes in 4X SSC-Tween. The biotinylated probe was immunodetected using Texas Red avidin DCS (Vector Laboratories) and the signal was amplified with biotinylated anti-avidin D (Vector Laboratories). The chromosomes were mounted and counterstained in Vectashield (Vector Laboratories) containing 2.5 µg/mL 4',6-diamidino-2-phenylindole (DAPI). Fluorescence images were captured using a CoolSnap HQ camera (Photometrics, Tucson, Ariz) on an Axioplan 2 microscope (Zeiss, Oberkochen, Germany) and analyzed using MetaVueTM (Universal Imaging corporation, Downington, PA).

2.7 Phylogenetics analyses

The evolutionary history of the most abundant TEs identified in *Spartina* (i.e. *Gypsy* and *Copia* retrotransposons, *LINE*, *CACTA* - see below) were investigated using phylogenetic analyses of protein sequences coding for the reverse transcriptase (RT) domain of retrotransposons or the transposase domain of *CACTA* elements. Sequences were retrieved from each *Spartina* TE database using the DANTE tool in the Galaxy version of RepeatExplorer [81]. Consensus sequences which showed homology with RT domains were translated into proteins and retained when they represented at least 50% of the RT domain size (87 and 125 amino acids for *Gypsy* and *Copia* elements, respectively). Similar analyses were performed to select sequences in other species used as outgroups in the phylogenetic analysis. A set of RT sequences from Repbase (v. 23.08; [87,88]) was selected for six species representing various plant lineages: *Pinus taeda*, *Amborella trichopoda*, *Arabidopsis thaliana*, *Medicago truncatula*, *Zea mays* and *Oryza sativa.* Only sequences without stop-codon were kept. *Gypsy* and *Copia* RT protein sequences were separately aligned using Clustal Omega [91]. Informative blocks in multiple sequence alignments were selected with the GBlocks package [92]. Phylogenetic analyses were performed with the IQ-TREE software [93] which enabled the testing of different models of molecular evolution via ModelFinder [94] and the construction of maximum-likelihood trees. The best-fit substitution model used was LG+R6 (determined according to BIC criteria) and robustness of branches was estimated after 10 000 Ultrafast Bootstraps [95].

2.8 Analyses of TE diversity

Journal Pre-proof

Diversity within each TE lineage was estimated with two complementary methods. First, divergence between reads coding for RT (or transposase) was determined by pairwise sequence comparisons. All reads from a sample of each species were aligned on the consensus sequences encoding RT domains using Bowtie2 (parameters: --score-min, local, G, 52, 8; [84]. The distance between each pair of aligned reads was then calculated under Kimura two parameters model [96] with MEGA7 [97]. Only distances between reads overlapping at least 50 nucleotides were considered.

Second, TE diversity was evaluated by detecting haplotypes using the pipeline previously developed [98]. This pipeline detects polymorphisms or SNPs (Single-Nucleotide Polymorphisms) from reads aligned to consensus TE sequences and constructs the different haplotypes based on shared and phased polymorphisms (setting parameters: min. read depth: 10; read-depth threshold for SNP detection: 10; min. number of shared SNP for haplotyping: 2). The number of haplotypes retrieved for each TE provides an estimation of the different TE copies.

3. Results

3.1 Annotation and quantitative evaluation of repeated sequences in *Spartina* genomes

Between 2.2-4.0% of reads (from samples of 2 million reads) showed sequence homology to plastid DNA, and 0.7-1.5% to ribosomal DNA, depending on the species (Table 2). Mapping of all Illumina read datasets to the *Spartina* repeated sequence databases confirmed that no sample bias was observed in the quantification of repeats by RepeatExplorer (see Supplementary Table S1).

RepeatExplorer forms clusters of reads that represent repeat sequences. The number of clusters (repeat types) that occupy > 1% of the genome ranged from 59 clusters in *S. maritima* to 74 clusters in *S. bakeri* (Table 2). In total, these clusters comprise repeated sequences that occupy more than 45% of the sampled genomes of both hexaploid species (Table 3). Amongst the tetraploids, while *S. spartinae* has a similar proportion of repeats (44% of the genome), the other two species have substantially lower proportions (33%). In all species, the most abundant TEs are Class I retrotransposons with the largest proportion of the repeat content being *Gypsy* (52-70% of total repeats) followed by *Copia* (16%-23%) elements (Table 3).

The two sister tetraploid species *S. versicolor* and *S. bakeri* have similar TE content predominantly comprising *Gypsy* elements belonging to *Tekay* and *Ogre* lineages (estimated to represent ~ 11% (74 Mb) and ~ 5.5% (37 Mb) of their genomes respectively), *Copia* elements in the *SIRE* lineage (~3.5% (24 Mb) of their genomes) and *CACTA* elements (~2.7% (18.5 Mb) of their genomes) (Fig. 2 and Supplementary Table S2). Overall, the total repetitive fraction of the genomes (including satellite repeats) is very similar, comprising 224 Mb (33%) in *S. versicolor* and 230 Mb (33%) in *S. bakeri*. In contrast, *S. spartinae*, which is also tetraploid, contains double the amount of repetitive sequences (~ 421 Mb). This difference is explained by larger amounts of all TE families particularly the *Ogre* and *Tekay* lineages which are, respectively, three and two times more abundant in *S. spartinae* than in *S. versicolor* and *S. bakeri*. Nevertheless, given the larger genome size of *S. spartinae* (963 Mb/1C) compared with *S. versicolor* (680 Mb/1C) and *S. bakeri* (689 Mb/1C), the proportion of the *S. spartinae* genome occupied by repeats (i.e. 43.7%) is only a third higher.

As with the tetraploid species, the most abundant repeats in the hexaploids are LTR-retrotransposons (Fig. 2 and Supplementary Table S2). In total, *Gypsy* and *Copia* elements comprise 35-37% of their genomes, with differences only in the abundance of individual element types (Supplementary Table S2). For example,

Journal Pre-proof

Chromovirus (*Tekay* and *CRM*) and *Ikeros* lineages are more abundant in *S. alterniflora* (estimated to 399 Mb and 25 Mb, respectively) than in *S. maritima* (314 Mb and 9 Mb). In contrast, *Ogre* and *SIRE* lineages are more abundant in *S. maritima* (74 Mb and 143 Mb respectively) than in *S. alterniflora* (37 Mb and 126 Mb).

In addition to the variation in abundance of LTR-retrotransposons, other TE lineages also exhibit interspecific variation. For example, the non-LTR retrotransposon LINE-RTE is less abundant in the tetraploid species (~3 Mb) than in the hexaploid species (14 Mb in *S. alterniflora* and 21 Mb in *S. maritima*). Amongst the Class II transposons, *CACTA* elements represent a more substantial component of the *S. maritima* genome (50 Mb) compared with *S. alterniflora* (26 Mb) and the tetraploid genomes (~ 18-26 Mb) (Fig. 2 and Supplementary Table S2).

3.2 Three highly repeated satellites in *Spartina*

Three highly repeated satellite DNA sequences were identified in the *Spartina* genomes (Fig. 3). One of these repeats, which has a monomer unit length of 167 bp, occupies less than 8 Mb of the tetraploid genomes (from 0.4 Mb in *S. spartinae* to 7.9 Mb in *S. versicolor*) corresponding to ~ 0.04% to 1.16% of their genomes, respectively, but it is more abundant in the hexaploid species (i.e. 17 Mb in *S. alterniflora* and 35 Mb in *S. maritima*), corresponding to ~ 1% and 2% of their genomes, respectively. The relative abundance of this satellite in each hexaploid species was confirmed by Southern blot hybridization, which showed strong probe hybridization to *S. maritima* DNA and weaker hybridization to *S. alterniflora* DNA (Fig. 4c) reflecting sequence divergence, as supported by the detection of two independent clusters of reads using RepeatExplorer (Fig. 4d). The detection of monomers (only in *S. maritima*), dimers and trimers in Southern blots (identification of band sizes corresponding to 167-bp repeats) also demonstrate the tandem arrangement of units. A second satellite repeat (monomer-length of 147 bp) is more abundant in *S. versicolor* (18 Mb) and *S. bakeri* (17 Mb) than in *S. spartinae* (6 Mb) and both hexaploids (1.3 and 10 Mb in *S. alterniflora* and *S. maritima*, respectively). These two satellite DNAs have sequence homology to SAT1 and CentC (genbank ID: M32521.1 and KT724912.1) which are two centromeric repeats identified in several grasses like rice, maize, *Sorghum* and *Brachypodium* [99–103].

The third satellite DNA sequence (with a monomer length of 370 bp) is the most abundant satellite in the hexaploid species *S. alterniflora*, representing up to 135 Mb (~7%) of its genome. In contrast, it was only detected in 0.15% of *S. maritima* reads. The lower abundance was also supported by the Southern blot hybridization (Fig. 4a) and FISH (Fig. 5) analyses. A comparative clustering analysis suggests that this satellite repeat is diverging between the two hexaploid species, given the observed spatial separation of the *S. alterniflora* and *S. maritima* sequences in the graphical depiction of three repeat clusters using Repeat Explorer (Fig. 4b). A comparison of the 370-bp satellite sequence in the NCBI database failed to find any evidence of homology between this satellite and centromeric repeats identified in other species.

Southern blotting (Fig. 4a) and FISH (Fig. 5) indicate that this 370 bp-long sequence is tandemly repeated and is physically located in subtelomeric positions on four chromosomes in *S. maritima* and on 42 chromosomes in *S. alterniflora*. It is notable that on two of the chromosomes in *S. alterniflora* the repeat was localized in the subtelomeric region of both chromosome arms (in contrast to all other chromosomes which had signal on just one of the chromosome arms). Cytogenetic analyses suggested that 14 chromosomes of the tetraploid *S. spartinae* also carried this repeat on one chromosome arm, even though this sequence was not detected in this

Journal Pre-proof

species by *in silico* sequence analyses using RepeatExplorer. Semiquantitative PCR further revealed that the satellite sequence was also present (Supplementary Fig. S1), in low copy numbers in the tetraploids *S. versicolor* and *S. bakeri* and in high copy numbers in the hexaploid *S. foliosa*, as it occurs in *S. alterniflora*, its sister species.

3.3 Evolutionary dynamics of TEs

Phylogenetic analyses of the two *main Gypsy* lineages (*Tekay* and *Ogre*) detected in *Spartina* reveals contrasting evolutionary histories depending on species (Fig. 6a,b). *Spartina versicolor* and *S. bakeri* exhibit similar divergence profiles for *Gypsy Tekay* and *Ogre* elements, although some TE lineages were differentially enriched (Fig. 6a,b). For example, *Tekay* elements of lineage III (as determined from phylogenetic analyses of the RT domain, Fig. 6) are more abundant and divergent in *S. bakeri* (54.2 Mb; 177 haplotypes) than in *S. versicolor* (35 Mb; 68 haplotypes). Conversely, *Tekay* elements in lineage IV occurred in greater quantity but lower sequence diversity in *S. versicolor* (37.7 Mb; 36 haplotypes) than in *S. bakeri* (21.8 Mb; 71). Similar differential accumulation of TEs between tetraploid species is also observed for *Gypsy Ogre* elements (Fig. 6b). Despite these differences, both tetraploid species contain similar overall amounts of *Tekay* elements (72.8 Mb and 76.3 Mb in *S. versicolor* and *S. bakeri*, respectively) and *Ogre* elements (37.3 Mb and 36.3 Mb, respectively) (Supplementary Table S2).

In *S. spartinae*, pairwise sequence comparisons show higher divergence between reads coding for the reverse transcriptase (RT) domain of *Tekay* elements than in other tetraploid species (average distance of 0.41 in *S. spartinae* versus 0.22 and 0.35 in *S. versicolor* and *S. bakeri*, respectively; Fig. 6a), reflecting an older expansion of *Tekay* elements in this species. *Spartina spartinae* also displays the highest diversity of haplotypes: 803 haplotypes in *Tekay* lineage III and 417 haplotypes in lineage IV. This higher level of diversity in *S. spartinae* was typically found in each of the most abundant TE families identified (Fig. 6a-f, Supplementary Fig. S2). Interestingly, while all TEs were found to occur in intermediate quantities in *S. spartinae* compared with the other tetraploids and the two hexaploids, *Ogre* elements have accumulated in *S. spartinae* (104 Mb) compared with the other species (from 23 Mb to 74 Mb; Supplementary Table S2). In addition, elements from *Ogre* lineage III occupy a larger genome proportion in *S. spartinae* (68.7 Mb) than in other species (~10-14 Mb).

The two hexaploids, *S. alterniflora* and *S. maritima*, share similar divergence patterns for *Gypsy Tekay* elements, but these differ from the tetraploid species. Thus, the *Tekay* element profiles in hexaploids showed two distinct peaks, representing two predominant levels of divergence between RT domains: one with little average sequence divergence (0.13) and the other showing greater sequence divergence (0.51). These are likely to corresponds to two independent bursts during their common evolutionary history. In contrast, there is only a single peak of sequence divergence apparent in the tetraploids. The abundance and diversity of *Tekay* elements also differ between the two hexaploids species. For example, there is twice the amount of *Tekay* lineage IV derived reads in *S. alterniflora* (206 Mb) compared with *S. maritima* (106 Mb). In contrast, *Tekay* lineage III elements are more abundant in *S. maritima* (187 Mb) than in *S. alterniflora* (136Mb). There were no reads from *S. alterniflora* in *Ogre* lineages III and VI, whilst 64 haplotypes and 18 haplotypes were detected respectively of these lineages in *S. maritima*.

Journal Pre-proof

Phylogenetic analysis of *Copia SIRE* elements (Fig. 6c) revealed three distinct lineages in all five *Spartina* species. The sister tetraploid species *S. versicolor* and *S. bakeri* show similar quantities and diversities for each lineage (lineage I: 14-16 Mb / 69-92 haplotypes; lineage II: 0.2-0.8 Mb / 5-7 haplotypes; lineage III: 7-8 Mb / 35-38 haplotypes). The hexaploid species *S. alterniflora* has approximately twice as many elements from lineage II (21 Mb) and more diversity (40 haplotypes) than *S. maritima* (10.8 Mb / 18 haplotypes), whereas *S. maritima* has about twice as many elements from lineage III (78 Mb) than *S. alterniflora* (35.8 Mb). For each *SIRE* lineage, *S. spartinae* showed an intermediate abundance of repeats compared with the other tetraploid and the hexaploid species (lineage I: 27 Mb; lineage II: 0.5 Mb; lineage III: 6 Mb), but this species had the greatest diversity of lineage I elements. The estimated divergence between reads showed a bimodal curve in all species with one peak closed to 0 (between 0 and 0.4) and another peak around 0.7. Compared to the patterns observed with *Gypsy*-derived sequences, most pairwise sequence comparisons in *SIRE* elements showed lower levels of divergence. This is also true for the *Copia Ivana* elements (Fig. 6d) where the curve decreases exponentially close to 0 (average distance between 0.05 and 0.09) in all species except *S. spartinae*, where reads tend to be slightly more divergent (average distance of 0.15).

The non-LTR LINE-RTE elements identified in *Spartina* species belong to a single lineage (Fig. 6e). As noted above for LTR-retrotransposons, the number of estimated haplotypes was not correlated with the abundance of this lineage in the investigated genomes. In the tetraploid species, the number of haplotypes varies from 31-37 in *S. versicolor* and *S. bakeri* to 135 in *S. spartinae*, whilst the amount of this repeat in the genome is typically around 3 Mb in all species. Amongst the hexaploids, *S. alterniflora* has a high level of diversity and a considerably larger amount of these repeats compared with tetraploids (129 haplotypes for 13 Mb). The repeat amount is even higher in *S. maritima* and yet the number of haplotypes is not as large as in *S. alterniflora* (38 haplotypes / 21 Mb). Divergence distance analyses confirm a remarkably low divergence between RT reads of LINE-RTE elements in all species.

Finally, six distinct lineages of CACTA elements were detected (Fig. 6f). All lineages contain sequences from all five species with the exception of lineage IV which does not appear to include sequences from *S. alterniflora*. Each lineage is similar in diversity and quantity in the two tetraploids *S. versicolor* and *S. bakeri*. In contrast, *S. spartinae* has a higher diversity of elements in all lineages than the two tetraploid species. Amongst the hexaploids, *S. maritima* contains more *CACTA* elements in each lineage than *S. alterniflora* (lineage I to V). Pairwise sequence comparisons in CACTA elements showed the same low divergence in all species with only one peak around 0.1-0.2.

Phylogenetic and divergence analyses of the other identified TE-families are available in Supplementary Fig. S3.

4. Discussion

4.1 Repeat content in *Spartina* and comparisons with other genomes in Poaceae

By comparing genomic sequences in related tetraploid and hexaploid *Spartina* species, we shed light on the evolutionary dynamics of the repetitive sequence component comprising their genomes over the last 6-10 million years (estimated time since *Spartina* diverged from its last common ancestor in subfamily Chloridoideae [72]). The genome sizes of tetraploids are smaller (1.39-1.97 pg/2C, 8 species) than hexaploids (3.67 – 4.05 pg/2C, 3 species). This difference is not only accounted for by the higher ploidy levels, since dividing the

Journal Pre-proof

genome size by ploidy level reveals that the monoploid genome sizes (1Cx-value; [104]) of the tetraploid species are also considerably smaller (0.34-0.49 pg/1Cx) than those of the hexaploid species (0.61-0.68 pg/1Cx) (Table 1). Thus, there is likely to have been genome upsizing in the hexaploids and/or downsizing in the tetraploids associated with species radiations.

Spartina belongs to subfamily Chloridoideae, one of the poorest investigated lineages in the grass family. Our analyses have allowed us to identify the repeat fraction of the genomes in five *Spartina* species. These genomes are only ~ 33% - 45% repetitive depending on species, which indicates limited amplification of repeats, or more efficient removal of repeats, compared with other grasses, such as *Zea mays*, *Hordeum vulgare* and *Triticum aestivum* which are estimated to be ~ 80% repetitive [10,14,105]). The estimated abundances of repeats in *Spartina* are similar to those reported for the allotetraploid *Eragrostis tef* (27%) which belongs to the same subfamily as *Spartina* (i.e. Chloridoideae) and whose genome size of 730 Mb/1C [106,107] is broadly similar to the tetraploid species *S. versicolor* and *S. bakeri.* Potentially older more degraded TEs occur in the genomes of *Spartina*, but they remain undetected using the homology parameters used here (80% of min. identity).

The majority of repeated sequences are LTR-retrotransposons representing between 24% and 38% of the genome depending on *Spartina* species (Table 3). Amongst LTR-retrotransposons, *Gypsy* elements are more abundant (17% to 30%) than *Copia* elements (from 6% to 10%). This is commonly observed in Poaceae, for example, *Zea mays* (76% of *Gypsy* vs. 9% of *Copia*), *Sorghum bicolor* (19% vs. 5%) and *Setaria italica* (22% vs. 7%) [10,108,109]. The non-LTR retrotransposon *LINE* elements and DNA transposons represent only a small fraction of *Spartina* genomes (less than 1% of *LINEs* and from 1.5% to 4.4% of DNA transposons) as is also typically reported for many plant genomes analyzed so far [110].

Within *Spartina*, different repeat families evolved through differential expansion or deletion in the tetraploid and hexaploid lineages. Thus, each investigated species is now characterized by a distinctive repeat profile, with the most closely related species showing the most similar profiles. For several repeat elements, pairwise sequence comparisons revealed a bimodal profile of sequence similarities (Fig. 6). This could represent two independent bursts of repeat activity during their evolution within the genomes of different species - one wave of activity that is still ongoing giving rise to a population of repeats which have high levels of sequence similarity, and another more ancient burst of repeat activity generating a peak of more degenerate copies.

4.2 Evolutionary dynamics of TEs in tetraploid *Spartina* species

The two closely related tetraploids species *S. versicolor* and *S. bakeri* are sister taxa which are thought to have diverged less than 2 mya [72] and have similar genome sizes (680 Mb/1C and 689 Mb/1C, respectively). Their genomes contain 33% of repeats (~230 Mb) comprising approximately 17% *Gypsy*, 6% *Copia*, 3-4% DNA transposons and 4-5% of satellite repeats (Table 3, Fig. 2, 3, 6). Similar genome sizes and repeat profiles indicate no major changes in TE content since their divergence from their common ancestor ~ 2 mya. In contrast, the tetraploid *S. spartinae*, whose phylogenetic relationship to other *Spartina* species is still debated [69,72], exhibited more divergent features of its repeat landscape, which comprises 44% of its genome (~421 Mb). Not only does this species have the largest tetraploid genome analyzed (1C = 963 Mb) with about two-fold more repetitive DNA than the other analyzed tetraploids, but the diversity of repeats in terms of the number of

Journal Pre-proof

haplotypes in each family is also high, resembling more closely that seen for the hexaploids. For example, the *Ogre* lineage III constituted 7% of the *S. spartinae* genome (i.e. 69 Mb) but represented only a limited fraction of the *S. versicolor* (1.5%) and *S. bakeri* (2.1%) genomes (i.e. 10 and 14 Mb respectively). In addition, the number of *Ogre* lineage III haplotypes is greater in *S. spartinae* (294 haplotypes vs 158 and 127 haplotypes in *S. versicolor* and *S. bakeri* respectively).

Biological traits such as mode of reproduction, the degree of selfing versus outcrossing have all been suggested to impact TE dynamics [111]. *Spartina* species are all perennial, wind pollinated plants. Asexual reproduction is significant for the propagation and invasive success of *S. versicolor* (syn. *S. patens*) in introduced European populations, where seed production is scarce [112,113]. However, little is known about the relative importance of sexual versus asexual reproduction in the reproductive biology of *S. bakeri* and *S. spartinae*.

The diversity and quantity of repeats in *S. spartinae* suggests different patterns of repeat proliferation and divergence compared with the other tetraploids, or the early divergence of this species from the other tetraploids. However, the precise phylogenetic placement of *S. spartinae* is unresolved from phylogenetic studies published so far [69,71,72], with some evidence suggesting that it may have diverged early in *Spartina*, while other data indicating that it is sister to the hexaploid clade (Fig. 1). Our data might support the latter assertion, but it will be necessary to explore the TE content of other tetraploid *Spartina* species (Fig. 1), such as *S. arundinaceae* (1C = 782 Mb) and *S. pectinata* (1C = 763 Mb), to better resolve the patterns of gains and/or losses of repeats from their common ancestor. Given the uncertain phylogenetic position of *S. spartinae*, several alternative hypotheses about the past dynamics of TEs could explain the repeat patterns observed in *S. spartinae*. (i) The high TE divergence in *S. spartinae* could reflect ongoing mutation accumulation of old repeat elements inherited from the ancestral genome of *Spartina*. The lower amount and diversity of TEs observed in *S. versicolor* and *S. bakeri* would thus reflect a loss of these elements associated with genome downsizing from their ancestor. Studies have shown that in some cases genome downsizing can occur rapidly and drastically following polyploidy [23,114–117], but it has also been shown to occur more progressively in other species [24,118], as may be apparent amongst some of the tetraploid *Spartina* species. (ii) The high diversity of repeats observed in *S. spartinae* could be explained by a hybrid origin of this species. This hypothesis could be consistent with the conflicting phylogenetic signal over its placement. Considering its current geographic range, *S. spartinae* has a disjunct distribution in South America (northern Argentina and Paraguay) and North America (Gulf coast in USA, and Mexico to Costa Rica) where several *Spartina* species such as *S. alterniflora*, *S. patens*, *S. bakeri* and *S. cynosuroides* are encountered [58,119]. A complex reticulate history of *S. spartinae* may therefore have occurred, involving, for example, introgression and/or horizontal gene transfers [120,121]. It is therefore possible that conflicts between TEs of multiple evolutionary origins and their repressors could have contributed to increasing the expansion of some TE lineages. Such processes could certainly have given rise to a complex repeat profile as a result of repeats being inherited from two or more species. (iii) A non-exclusive likely hypothesis is that current TEs in *S. spartinae* reflect TE accumulation following bursts of activity in the common ancestor of *S. spartinae* and the hexaploid species, which also have a greater diversity of repeats than the other tetraploids. Such arguments would reinforce the hypothesis of a close phylogenetic relationship between the hexaploids and *S. spartinae*, which would then represent one of the tetraploid ancestor lineages involved in the formation of the hexaploid *Spartina* clade.

Journal Pre-proof

4.3 Differential evolution of TEs in hexaploid *Spartina* species

The hexaploid species *S. alterniflora* and *S. maritima* diverged around 2-4 Mya [72] on the American and Euro-African Atlantic coasts, respectively, and they display distinct repeat profiles. Both species have a similar fraction of their genome being repetitive (i.e. 45.5%, three-fold more than previously reported from BAC-end sequencing [122]). However, there is a greater quantity of TEs in *S. maritima* (estimated to 751 Mb) than in *S. alterniflora* (724 Mb) and a greater quantity of satellite DNAs in *S. alterniflora* (164 Mb) than in *S. maritima* (66 Mb). This is consistent with previous transposon display analyses (Parisod et al. 2009) who suggested greater accumulation of TEs in *S. maritima* than *S. alterniflora*.

Within TEs, *Tekay* and *CRM* lineages were more abundant in *S. alterniflora* while *Ogre* and *SIRE* lineages were more abundant in *S. maritima* (Fig. 2 and Supplementary Table S1). All *Gypsy, SIRE and CACTA* lineages (representing the majority of TEs) displayed high diversity (i.e. many haplotypes) with similar divergence patterns in the two species (Fig. 6) suggesting that the TEs in *S. maritima* and *S. alterniflora* were derived from the same transposition bursts before they diverged 2-4 mya but have since evolved independently.

However, the number of detected copies in the *Copia Ivana* and *LINE-RTE* (non-LTR retrotransposon) lineages was lower in *S. maritima* than *S. alterniflora.* Pairwise sequence comparisons suggest in addition, high sequence similarity between many of the *S. maritima* reads indicative of recent bursts of amplification in this species. Similar results were observed for *Copia Ikeros* in *S. alterniflora* (Supplementary Fig. S3). A recent study of the evolution of ra-siRNAs (repeat-associated small-interfering RNA) in *S. alterniflora* and *S. maritima* identified many ra-siRNAs targeting preferentially *Ivana, Ikeros* and *LINE-RTE* lineages [123]. Such results are consistent with the hypothesis that these elements (i.e. *Ivana* and *LINE-RTE* elements in *S. maritima* and *Ikeros* elements in *S. alterniflora*) have recently been and/or are still active, with their activity being regulated *via* lineage-specific siRNAs. Further work is needed to determine if the elements are transcriptionally active and if they are able to insert into the genome.

4.4 Expansion of satellite sequences in *S. alterniflora*

Three different satellite DNA sequences were identified in *Spartina* that range in abundance from 1.6% (15.6 Mb) in the genome of *S. spartinae* to 8.4% (164.3 Mb) in *S. alterniflora*. Two of them (a 167 bp-long monomer and a 147 bp-long monomer) were found to be similar to centromeric repeats previously characterized in grasses, including in *Zea mays*, *Oryza*, *Sorghum* and *Brachypodium* [100–103]. The 147 bp monomer repeat was more abundant in the two closely related tetraploid *Spartina* species (*S. versicolor* and *S. bakeri*), while the 167 bp monomer sequence was more abundant in the hexaploid species (validated by Southern blotting, Fig. 4). Out of the three tetraploids, the genome of *S. spartinae* had the smaller fraction of satellite DNA overall (only 1.6% *vs.* 4-5% in the other tetraploids; Table 3), highlighting that in contrast to the TEs discussed above, these satellite repeats were not contributing to its larger genome size.

A third remarkable satellite sequence with a monomer size of 370 bp was found to be especially abundant in the hexaploid *S. alterniflora*, where it represents ~7% of the analyzed genomic sequences. This agrees with previous cytogenetic observations [62] suggesting an abundance of heterochromatin on *S. alterniflora* chromosomes. No sequence similarity to other known satellite DNAs was observed, even to satellite sequences reported in a related Chloridoideae species *Eragrostis tef* [107]. Southern blotting allowed us to confirm the

Journal Pre-proof

presence of this 370-bp tandem repeat to be in higher abundance in *S. alterniflora* (135 Mb) than *S. maritima* (~3 Mb). The sequence was localized by FISH to subtelomeric positions on 42 chromosomes of *S. alterniflora* (2n = 62), and to just four chromosomes in *S. maritima* (2n = 60) (Fig. 4). Interestingly, this repeat was mostly localized to the subtelomeric region of just one of the two chromosome arms in 40 out of 42 chromosomes of *S. alterniflora* and the four chromosomes of *S. maritima*. The remaining two chromosomes of *S. alterniflora* with FISH signal showed the 370-bp satellite was localized in the subtelomeric regions of both chromosome arms. These chromosomes might represent a fixed pair of supernumerary or B chromosomes in *S. alterniflora*, whose karyotype is 2n = 60+2 compared with *S. maritima* which is 2n = 60 [58]. These two extra chromosomes are inherited together in hybrids involving *S. alterniflora* (e.g. *S.* × *townsendii* (2n = 60+2) [124] and *S.* × *neyrautii* (2n = 60+2) [125]). Potentially this satellite repeat has a key function in chromosome segregation during meiosis, "driving" the supernumerary chromosomes into the next generation, as found for similar satellite DNA sequences on B or supernumerary chromosomes in other species [126,127].

Semi-quantitative PCR indicated this 370-bp satellite was also present in the hexaploid *S. foliosa* (2n = 62; 2C = 4.05pg) in similar abundance to *S. alterniflora* (Supplementary Fig. S1). This is consistent with the phylogenetic relationships between these two species as they are considered to be sister species within the hexaploid *Spartina* subclade (Fig. 2) [69,71,72]. Thus, it is likely that the evolution of this satellite DNA occurred prior to the divergence of these species. If the sequence shows 'meiotic drive' as the results from *S. alterniflora* suggest, its occurrence more widely amongst *Spartina* species may be important for the evolution of linked alleles.

The 370-bp satellite was not detected *in silico* in the genomic sequence datasets from the tetraploid species. Nevertheless, FISH and semi-quantitative PCR analyses indicated it was present in low amounts in *S. versicolor* and *S. bakeri*, and on 14 chromosomes in *S. spartinae* (Fig. 5). These data suggest that this satellite may have been present in the common ancestor of *Spartina*. Various studies have shown that satellite DNA sequences can evolve rapidly and their amount can vary considerably, even between closely related species [3,115,127–129].

In conclusion, our comparative analyses of five polyploid species representative of the main *Spartina* clades, reveal contrasting TE dynamics and satellite DNA accumulation within both tetraploid and hexaploid lineages. Repetitive sequence dynamics reflects the long-term evolution associated with allopatric speciation between hexaploids that diverged 3-6 MYA between the American (*S. alterniflora*) and European (*S. maritima*) Atlantic coasts respectively, or between North-American (*S. bakeri* *S. versicolor* sub-clade) and South-American (*S spartinae*) tetraploids, in agreement with their phylogenetic history. Genome expansion or contraction resulting from repeat evolution seems to be mostly determined by the divergence times between species, rather than by genomic changes triggered by ploidy level change *per se*, as suggested by our earlier studies on *Spartina* neopolyploids, where the immediate response to allopolyploidy affects epigenetic regulation of repeats.

The results obtained here, which include repeat content evaluation and characterization between *Spartina* lineages, provide crucial insights for future investigations into the genetic, epigenetic and transcriptomic consequences of differential repeat dynamics in the meso- and neopolyploid species.

Journal Pre-proof

Funding

This work was financed under the International Associated Laboratory "Ecological Genomics of Polyploidy" supported by CNRS (INEE, UMR CNRS 6553 Ecobio), University of Rennes 1 and the Iowa State University (USA). The analyses benefited from the Molecular Ecology (UMR CNRS 6553 Ecobio) and Genouest (Biogenouest) facilities. Delphine Giraud benefited from a PhD scholarship from the University of Rennes 1.

Conflicts of interest

The authors declare that they have no conflict of interest

Authors' contributions

Malika Aïnouche and Armel Salmon designed the experiment. Oscar Lima extracted the DNA for sequencing, set up PCR amplifications. Virginie Huteau and Olivier Coriton performed FISH analyses. Ales Kovarik realized Southern blot hybridizations. Julien Boutte developed the IlluHaplotyper pipeline for the haplotype detection. Delphine Giraud performed RepeatExplorer analyses, annotations, phylogenies, haplotyping, pairwise sequence comparisons. Delphine Giraud, Malika Aïnouche and Armel Salmon analyzed the results and wrote the article. Ales Kovarik, Andrew R. Leitch, Ilia J. Leitch discussed and contributed to the writing of the manuscript.

Acknowledgements

We thank Dr. A. Baumel (IMBE, Aix-Marseille University, France), Dr. T. Columbus (Rancho Santa Ana Botanical Garden, CA USA) for their help in getting *Spartjna* plant material, as well as T. Fontaine and R. Bodiguel (Experimental Ecology platform, UMR CNRS Ecobio, University of Rennes 1, France) for maintaining the *Spartina* living collection in the Greenhouse. M. Bourges and N. Valentin (Institute of Integrative Biology of the Cell, Gif s/Yvette) are thanked for flow cytometry analyses. We acknowledge the perspicacity of late Dr. K.Y. Lim (Queen Mary University of London) who early suggested from cytogenetic observations the abundance of satellite repeats in *S. alterniflora*.

References

[1] K.R. Oliver, W.K. Greene, Transposable elements: powerful facilitators of evolution, BioEssays. 31 (2009) 703–714. https://doi.org/10.1002/bies.200800219.

[2] J.L. Bennetzen, H. Wang, The contributions of transposable elements to the structure, function, and evolution of plant genomes, Annu. Rev. Plant Biol. 65 (2014) 505–530. https://doi.org/10.1146/annurev-arplant-050213-035811.

[3] S. Mehrotra, V. Goyal, Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function, Genomics Proteomics Bioinformatics. 12 (2014) 164–171. https://doi.org/10.1016/j.gpb.2014.07.003.

[4] J. Pellicer, M.F. Fay, I.J. Leitch, The largest eukaryotic genome of them all?, Bot. J. Linn. Soc. 164 (2010) 10–15. https://doi.org/10.1111/j.1095-8339.2010.01072.x.

[5] J. Pellicer, O. Hidalgo, S. Dodsworth, I. Leitch, Genome size diversity and its impact on the evolution of land plants, Genes. 9 (2018) 88. https://doi.org/10.3390/genes9020088.

[6] S. Dodsworth, A.R. Leitch, I.J. Leitch, Genome size diversity in angiosperms and its influence on gene space, Curr. Opin. Genet. Dev. 35 (2015) 73–78. https://doi.org/10.1016/j.gde.2015.10.006.

[7] T.P. Michael, Plant genome size variation: bloating and purging DNA, Brief. Funct. Genomics. 13

Journal Pre-proof

(2014) 308–317. https://doi.org/10.1093/bfgp/elu005.

[8] L. Ren, W. Huang, E.K.S. Cannon, D.J. Bertioli, S.B. Cannon, A mechanism for genome size reduction following genomic rearrangements, Front. Genet. 9 (2018). https://doi.org/10.3389/fgene.2018.00454.

[9] P.S. Soltis, D.B. Marchant, Y. Van de Peer, D.E. Soltis, Polyploidy and genome evolution in plants, Curr. Opin. Genet. Dev. 35 (2015) 119–125. https://doi.org/10.1016/j.gde.2015.11.003.

[10] P.S. Schnable, D. Ware, R.S. Fulton, J.C. Stein, F. Wei, S. Pasternak, C. Liang, J. Zhang, L. Fulton, T.A. Graves, P. Minx, A.D. Reily, L. Courtney, S.S. Kruchowski, C. Tomlinson, C. Strong, K. Delehaunty, C. Fronick, B. Courtney, S.M. Rock, E. Belter, F. Du, K. Kim, R.M. Abbott, M. Cotton, A. Levy, P. Marchetto, K. Ochoa, S.M. Jackson, B. Gillam, W. Chen, L. Yan, J. Higginbotham, M. Cardenas, J. Waligorski, E. Applebaum, L. Phelps, J. Falcone, K. Kanchi, T. Thane, A. Scimone, N. Thane, J. Henke, T. Wang, J. Ruppert, N. Shah, K. Rotter, J. Hodges, E. Ingenthron, M. Cordes, S. Kohlberg, J. Sgro, B. Delgado, K. Mead, A. Chinwalla, S. Leonard, K. Crouse, K. Collura, D. Kudrna, J. Currie, R. He, A. Angelova, S. Rajasekar, T. Mueller, R. Lomeli, G. Scara, A. Ko, K. Delaney, M. Wissotski, G. Lopez, D. Campos, M. Braidotti, E. Ashley, W. Golser, H. Kim, S. Lee, J. Lin, Z. Dujmic, W. Kim, J. Talag, A. Zuccolo, C. Fan, A. Sebastian, M. Kramer, L. Spiegel, L. Nascimento, T. Zutavern, B. Miller, C. Ambroise, S. Muller, W. Spooner, A. Narechania, L. Ren, S. Wei, S. Kumari, B. Faga, M.J. Levy, L. McMahan, P. Van Buren, M.W. Vaughn, K. Ying, C.-T. Yeh, S.J. Emrich, Y. Jia, A. Kalyanaraman, A.-P. Hsia, W.B. Barbazuk, R.S. Baucom, T.P. Brutnell, N.C. Carpita, C. Chaparro, J.-M. Chia, J.-M. Deragon, J.C. Estill, Y. Fu, J.A. Jeddeloh, Y. Han, H. Lee, P. Li, D.R. Lisch, S. Liu, Z. Liu, D.H. Nagel, M.C. McCann, P. SanMiguel, A.M. Myers, D. Nettleton, J. Nguyen, B.W. Penning, L. Ponnala, K.L. Schneider, D.C. Schwartz, A. Sharma, C. Soderlund, N.M. Springer, Q. Sun, H. Wang, M. Waterman, R. Westerman, T.K. Wolfgruber, L. Yang, Y. Yu, L. Zhang, S. Zhou, Q. Zhu, J.L. Bennetzen, R.K. Dawe, J. Jiang, N. Jiang, G.G. Presting, S.R. Wessler, S. Aluru, R.A. Martienssen, S.W. Clifton, W.R. McCombie, R.A. Wing, R.K. Wilson, The B73 maize genome: complexity, diversity, and dynamics, Science. 326 (2009) 1112–1115. https://doi.org/10.1126/science.1178534.

[11] M.I. Tenaillon, J.D. Hollister, B.S. Gaut, A triptych of the evolution of plant transposable elements, Trends Plant Sci. 15 (2010) 471–478. https://doi.org/10.1016/j.tplants.2010.05.003.

[12] K.R. Oliver, J.A. McComb, W.K. Greene, Transposable elements: powerful contributors to Angiosperm evolution and diversity, Genome Biol. Evol. 5 (2013) 1886–1901. https://doi.org/10.1093/gbe/evt141.

[13] N.-S. Kim, The genomes and transposable elements in plants: are they friends or foes?, Genes Genomics. 39 (2017) 359–370. https://doi.org/10.1007/s13258-017-0522-y.

[14] T. Wicker, H. Gundlach, M. Spannagl, C. Uauy, P. Borrill, R.H. Ramírez-González, R. De Oliveira, K.F.X. Mayer, E. Paux, F. Choulet, Impact of transposable elements on genome structure and evolution in bread wheat, Genome Biol. 19 (2018). https://doi.org/10.1186/s13059-018-1479-0.

[15] A. Kumar, J.L. Bennetzen, Plant retrotransposons, Annu. Rev. Genet. 33 (1999) 479–532. https://doi.org/10.1146/annurev.genet.33.1.479.

[16] C. Feschotte, E.J. Pritham, DNA transposons and the evolution of eukaryotic genomes, Annu. Rev. Genet. 41 (2007) 331–368. https://doi.org/10.1146/annurev.genet.40.110405.090448.

[17] T. Wicker, F. Sabot, A. Hua-Van, J.L. Bennetzen, P. Capy, B. Chalhoub, A. Flavell, P. Leroy, M. Morgante, O. Panaud, E. Paux, P. SanMiguel, A.H. Schulman, A unified classification system for eukaryotic transposable elements, Nat. Rev. Genet. 8 (2007) 973–982. https://doi.org/10.1038/nrg2165.

[18] C. Vitte, O. Panaud, H. Quesneville, LTR retrotransposons in rice (*Oryza sativa*, L.): recent burst amplifications followed by rapid DNA loss, BMC Genomics. 8 (2007). https://doi.org/10.1186/1471-2164-8-218.

[19] P. Civáň, M. Švec, P. Hauptvogel, On the coevolution of transposable elements and plant genomes, J. Bot. 2011 (2011) 1–9. https://doi.org/10.1155/2011/893546.

[20] S.E. Staton, B.H. Bakken, B.K. Blackman, M.A. Chapman, N.C. Kane, S. Tang, M.C. Ungerer, S.J. Knapp, L.H. Rieseberg, J.M. Burke, The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements, Plant J. 72 (2012) 142–153. https://doi.org/10.1111/j.1365-313X.2012.05072.x.

[21] M. El Baidouri, O. Panaud, Comparative genomic paleontology across plant kingdom reveals the dynamics of TE-driven genome evolution, Genome Biol. Evol. 5 (2013) 954–965. https://doi.org/10.1093/gbe/evt025.

Journal Pre-proof

[22] K. Wang, G. Huang, Y. Zhu, Transposable elements play an important role during cotton genome evolution and fiber cell development, Sci. China Life Sci. 59 (2016) 112–121. https://doi.org/10.1007/s11427-015-4928-y.

[23] A.R. Leitch, I.J. Leitch, Genomic plasticity and the diversity of polyploid plants, Science. 320 (2008) 481–483. https://doi.org/10.1126/science.1153585.

[24] S. Renny-Byfield, J.F. Wendel, Doubling down on genomes: polyploidy and crop plants, Am. J. Bot. 101 (2014) 1711–1725. https://doi.org/10.3732/ajb.1400119.

[25] J.F. Wendel, S.A. Jackson, B.C. Meyers, R.A. Wing, Evolution of plant genome architecture, Genome Biol. 17 (2016). https://doi.org/10.1186/s13059-016-0908-1.

[26] C.M. Vicient, J.M. Casacuberta, Impact of transposable elements on polyploid plant genomes, Ann. Bot. 120 (2017) 195–207. https://doi.org/10.1093/aob/mcx078.

[27] K. Kashkush, M. Feldman, A.A. Levy, Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat, Nat. Genet. 33 (2003) 102–106. https://doi.org/10.1038/ng1063.

[28] D. Lisch, How important are transposons for plant evolution?, Nat. Rev. Genet. 14 (2013) 49–61. https://doi.org/10.1038/nrg3374.

[29] J.F. Pereira, P.R. Ryan, The role of transposable elements in the evolution of aluminium resistance in plants, J. Exp. Bot. 70 (2019) 41–54. https://doi.org/10.1093/jxb/ery357.

[30] C.M. Vicient, Transcriptional activity of transposable elements in maize, BMC Genomics. 11 (2010) 601. https://doi.org/10.1186/1471-2164-11-601.

[31] M.J. Axtell, Classification and comparison of small RNAs from plants, Annu. Rev. Plant Biol. 64 (2013) 137–159. https://doi.org/10.1146/annurev-arplant-050312-120043.

[32] D. Fultz, S.G. Choudury, R.K. Slotkin, Silencing of active transposable elements in plants, Curr. Opin. Plant Biol. 27 (2015) 67–76. https://doi.org/10.1016/j.pbi.2015.05.027.

[33] M.J. Sigman, R.K. Slotkin, The first rule of plant transposable element silencing: location, location, location, Plant Cell. 28 (2016) 304–313. https://doi.org/10.1105/tpc.15.00869.

[34] J. Macas, P. Novák, J. Pellicer, J. Čížková, A. Koblížková, P. Neumann, I. Fuková, J. Doležel, L.J. Kelly, I.J. Leitch, In depth characterization of repetitive DNA in 23 plant genomes reveals sources of genome size variation in the legume tribe *Fabeae*, PLOS ONE. 10 (2015). https://doi.org/10.1371/journal.pone.0143424.

[35] R. Kalendar, J. Tanskanen, S. Immonen, E. Nevo, A.H. Schulman, Genome evolution of wild barley (*Hordeum spontaneum*) by *BARE-1* retrotransposon dynamics in response to sharp microclimatic divergence, Proc. Natl. Acad. Sci. 97 (2000) 6603–6607. https://doi.org/10.1073/pnas.110587497.

[36] V.V. Cavrak, N. Lettner, S. Jamge, A. Kosarewicz, L.M. Bayer, O. Mittelsten Scheid, How a retrotransposon exploits the plant's heat stress response for its activation, PLoS Genet. 10 (2014) e1004115. https://doi.org/10.1371/journal.pgen.1004115.

[37] M.-A. Grandbastien, LTR retrotransposons, handy hitchhikers of plant regulation and stress response, Biochim. Biophys. Acta BBA - Gene Regul. Mech. 1849 (2015) 403–416. https://doi.org/10.1016/j.bbagrm.2014.07.017.

[38] H. Ito, J.-M. Kim, W. Matsunaga, H. Saze, A. Matsui, T.A. Endo, Y. Harukawa, H. Takagi, H. Yaegashi, Y. Masuta, S. Masuda, J. Ishida, M. Tanaka, S. Takahashi, T. Morosawa, T. Toyoda, T. Kakutani, A. Kato, M. Seki, A stress-activated transposon in *Arabidopsis* induces transgenerational abscisic acid insensitivity, Sci. Rep. 6 (2016) 23181. https://doi.org/10.1038/srep23181.

[39] B. McClintock, The significance of responses of the genome to challenge, Science. 226 (1984) 792–801. https://doi.org/10.1126/science.15739260.

[40] C. Parisod, K. Alix, J. Just, M. Petit, V. Sarilar, C. Mhiri, M. Ainouche, B. Chalhoub, M.-A. Grandbastien, Impact of transposable elements on the organization and function of allopolyploid genomes, New Phytol. 186 (2010) 37–45. https://doi.org/10.1111/j.1469-8137.2009.03096.x.

[41] A. Madlung, A.P. Tyagi, B. Watson, H. Jiang, T. Kagochi, R.W. Doerge, R. Martienssen, L. Comai, Genomic changes in synthetic *Arabidopsis* polyploids, Plant J. 41 (2005) 221–230. https://doi.org/10.1111/j.1365-313X.2004.02297.x.

Journal Pre-proof

[42] U. Göbel, A.L. Arce, F. He, A. Rico, G. Schmitz, J. de Meaux, Robustness of transposable element regulation but no genomic shock observed in interspecific *Arabidopsis* hybrids, Genome Biol. Evol. 10 (2018) 1403–1415. https://doi.org/10.1093/gbe/evy095.

[43] B. Yaakov, K. Kashkush, Mobilization of *Stowaway*-like MITEs in newly formed allohexaploid wheat species, Plant Mol. Biol. 80 (2012) 419–427. https://doi.org/10.1007/s11103-012-9957-3.

[44] M. Petit, K.Y. Lim, E. Julio, C. Poncet, F. Dorlhac de Borne, A. Kovarik, A.R. Leitch, M.-A. Grandbastien, C. Mhiri, Differential impact of retrotransposon populations on the genome of allotetraploid tobacco (*Nicotiana tabacum*), Mol. Genet. Genomics. 278 (2007) 1–15. https://doi.org/10.1007/s00438-007-0226-0.

[45] M. Petit, C. Guidat, J. Daniel, E. Denis, E. Montoriol, Q.T. Bui, K.Y. Lim, A. Kovarik, A.R. Leitch, M.-A. Grandbastien, C. Mhiri, Mobilization of retrotransposons in synthetic allotetraploid tobacco, New Phytol. 186 (2010) 135–147. https://doi.org/10.1111/j.1469-8137.2009.03140.x.

[46] J. Beaulieu, M. Jean, F. Belzile, The allotetraploid *Arabidopsis thaliana–Arabidopsis lyrata* subsp. *petraea* as an alternative model system for the study of polyploidy in plants, Mol. Genet. Genomics. 281 (2009) 421–435. https://doi.org/10.1007/s00438-008-0421-7.

[47] C. Parisod, A. Salmon, T. Zerjal, M. Tenaillon, M.-A. Grandbastien, M. Ainouche, Rapid structural and epigenetic reorganization near transposable elements in hybrid and allopolyploid genomes in *Spartina*, New Phytol. 184 (2009) 1003–1015. https://doi.org/10.1111/j.1469-8137.2009.03029.x.

[48] M. Ha, J. Lu, L. Tian, V. Ramachandran, K.D. Kasschau, E.J. Chapman, J.C. Carrington, X. Chen, X.-J. Wang, Z.J. Chen, Small RNAs serve as a genetic buffer against genomic shock in *Arabidopsis* interspecific hybrids and allopolyploids, Proc. Natl. Acad. Sci. 106 (2009) 17835–17840. https://doi.org/10.1073/pnas.0907003106.

[49] G. Hu, J.S. Hawkins, C.E. Grover, J.F. Wendel, The history and disposition of transposable elements in polyploid *Gossypium*, Genome. 53 (2010) 599–607. https://doi.org/10.1139/G10-038.

[50] V. Sarilar, A. Marmagne, P. Brabant, J. Joets, K. Alix, BraSto, a Stowaway MITE from *Brassica*: recently active copies preferentially accumulate in the gene space, Plant Mol. Biol. 77 (2011) 59–75. https://doi.org/10.1007/s11103-011-9794-9.

[51] S. Renny-Byfield, A. Kovarik, L.J. Kelly, J. Macas, P. Novak, M.W. Chase, R.A. Nichols, M.R. Pancholi, M.-A. Grandbastien, A.R. Leitch, Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences, Plant J. 74 (2013) 829–839. https://doi.org/10.1111/tpj.12168.

[52] N. Senerchia, F. Felber, C. Parisod, Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats, New Phytol. 202 (2014) 975–985. https://doi.org/10.1111/nph.12731.

[53] H. Cheng, G. Sun, S. He, W. Gong, Z. Peng, R. Wang, Z. Lin, X. Du, Comparative effect of allopolyploidy on transposable element composition and gene expression between *Gossypium hirsutum* and its two diploid progenitors, J. Integr. Plant Biol. 61 (2019) 45–59. https://doi.org/10.1111/jipb.12763.

[54] P.P. Edger, R. Smith, M.R. McKain, A.M. Cooley, M. Vallejo-Marin, Y. Yuan, A.J. Bewick, L. Ji, A.E. Platts, M.J. Bowman, K.L. Childs, J.D. Washburn, R.J. Schmitz, G.D. Smith, J.C. Pires, J.R. Puzey, Subgenome dominance in an interspecific hybrid, synthetic allopolyploid, and a 140-year-old naturally established neo-allopolyploid monkeyflower, Plant Cell. 29 (2017) 2150–2167. https://doi.org/10.1105/tpc.17.00010.

[55] S. Bottani, N.R. Zabet, J.F. Wendel, R.A. Veitia, Gene expression dominance in allopolyploids: hypotheses and models, Trends Plant Sci. 23 (2018) 393–402. https://doi.org/10.1016/j.tplants.2018.01.002.

[56] Y. Van de Peer, E. Mizrachi, K. Marchal, The evolutionary significance of polyploidy, Nat. Rev. Genet. 18 (2017) 411–424. https://doi.org/10.1038/nrg.2017.26.

[57] A. Zwaenepoel, Y. Van de Peer, Inference of ancient whole-genome duplications and the evolution of gene duplication and loss rates, Mol. Biol. Evol. 36 (2019) 1384–1404. https://doi.org/10.1093/molbev/msz088.

[58] C.J. Marchant, Evolution in *Spartina* (Gramineae): III. Species chromosome numbers and their taxonomic significance, J. Linn. Soc. Lond. Bot. 60 (1968) 411–417. https://doi.org/10.1111/j.1095-8339.1968.tb00097.x.

[59] A. Baumel, M.L. Ainouche, M.T. Misset, J.-P. Gourret, R.J. Bayer, Genetic evidence for hybridization

Journal Pre-proof

between the native *Spartina maritima* and the introduced *Spartina alterniflora* (Poaceae) in South-West France: *Spartina x neyrautii* re-examined, Plant Syst. Evol. 237 (2003) 87–97. https://doi.org/10.1007/s00606-002-0251-8.

[60] D.R. Ayres, E. Grotkopp, K. Zaremba, C.M. Sloop, M.J. Blum, J.P. Bailey, C.K. Anttila, D.R. Strong, Hybridization between invasive *Spartina densiflora* (Poaceae) and native *S. foliosa* in San Francisco Bay, California, USA, Am. J. Bot. 95 (2008) 713–719. https://doi.org/10.3732/ajb.2007358.

[61] M.L. Ainouche, P.M. Fortune, A. Salmon, C. Parisod, M.-A. Grandbastien, K. Fukunaga, M. Ricou, M.-T. Misset, Hybridization, polyploidy and invasion: lessons from *Spartina* (Poaceae), Biol. Invasions. 11 (2009) 1159–1173. https://doi.org/10.1007/s10530-008-9383-2.

[62] S. Renny-Byfield, M. Ainouche, I.J. Leitch, K.Y. Lim, S.C. Le Comber, A.R. Leitch, Flow cytometry and GISH reveal mixed ploidy populations and *Spartina* nonaploids with genomes of *S. alterniflora* and *S. maritima* origin, Ann. Bot. 105 (2010) 527–533. https://doi.org/10.1093/aob/mcq008.

[63] R. Bedre, V.R. Mangu, S. Srivastava, L.E. Sanchez, N. Baisakh, Transcriptome analysis of smooth cordgrass (*Spartina alterniflora* Loisel), a monocot halophyte, reveals candidate genes involved in its adaptation to salinity, BMC Genomics. 17 (2016). https://doi.org/10.1186/s12864-016-3017-3.

[64] K.I. Paredes-Páliz, E. Mateos-Naranjo, B. Doukkali, M.A. Caviedes, S. Redondo-Gómez, I.D. Rodríguez-Llorente, E. Pajuelo, Modulation of *Spartina densiflora* plant growth and metal accumulation upon selective inoculation treatments: a comparison of gram negative and gram positive rhizobacteria, Mar. Pollut. Bull. 125 (2017) 77–85. https://doi.org/10.1016/j.marpolbul.2017.07.072.

[65] M. Robertson, A. Schrey, A. Shayter, C.J. Moss, C. Richards, Genetic and epigenetic variation in *Spartina alterniflora* following the *Deepwater Horizon* oil spill, Evol. Appl. 10 (2017) 792–801. https://doi.org/10.1111/eva.12482.

[66] M. Alvarez, J. Ferreira de Carvalho, A. Salmon, M.L. Ainouche, A. Cavé-Radet, A. El Amrani, T.E. Foster, S. Moyer, C.L. Richards, Transcriptome response of the foundation plant *Spartina alterniflora* to the *Deepwater Horizon* oil spill, Mol. Ecol. 27 (2018) 2986–3000. https://doi.org/10.1111/mec.14736.

[67] A. Cavé-Radet, A. Salmon, O. Lima, M.L. Ainouche, A. El Amrani, Increased tolerance to organic xenobiotics following recent allopolyploidy in *Spartina* (Poaceae), Plant Sci. 280 (2019) 143–154. https://doi.org/10.1016/j.plantsci.2018.11.005.

[68] P.M. Peterson, K. Romaschenko, Y.H. Arrieta, J.M. Saarela, A molecular phylogeny and new subgeneric classification of *Sporobolus* (Poaceae: Chloridoideae: Sporobolinae), Taxon. 63 (2014) 1212–1243. https://doi.org/10.12705/636.19.

[69] A. Baumel, M.L. Ainouche, R.J. Bayer, A.K. Ainouche, M.T. Misset, Molecular phylogeny of hybridizing species from the genus *Spartina* Schreb. (Poaceae), Mol. Phylogenet. Evol. 22 (2002) 303–314. https://doi.org/10.1006/mpev.2001.1064.

[70] M.L. Ainouche, A. Baumel, A. Salmon, G. Yannic, Hybridization, polyploidy and speciation in *Spartina* (Poaceae), New Phytol. 161 (2003) 165–172. https://doi.org/10.1046/j.1469-8137.2003.00926.x.

[71] P.M. Fortune, K.A. Schierenbeck, A.K. Ainouche, J. Jacquemin, J.F. Wendel, M.L. Ainouche, Evolutionary dynamics of *Waxy* and the origin of hexaploid *Spartina* species (Poaceae), Mol. Phylogenet. Evol. 43 (2007) 1040–1055. https://doi.org/10.1016/j.ympev.2006.11.018.

[72] M. Rousseau-Gueutin, S. Bellot, G.E. Martin, J. Boutte, H. Chelaifa, O. Lima, S. Michon-Coudouel, D. Naquin, A. Salmon, K. Ainouche, M. Ainouche, The chloroplast genome of the hexaploid *Spartina maritima* (Poaceae, Chloridoideae): comparative analyses and molecular dating, Mol. Phylogenet. Evol. 93 (2015) 5–16. https://doi.org/10.1016/j.ympev.2015.06.013.

[73] P.M. Fortune, K. Schierenbeck, D. Ayres, A. Bortolus, O. Catrice, S. Brown, M.L. Ainouche, The enigmatic invasive *Spartina densiflora* : a history of hybridizations in a polyploidy context, Mol. Ecol. 17 (2008) 4304–4316. https://doi.org/10.1111/j.1365-294X.2008.03916.x.

[74] S. Kim, A.L. Rayburn, A. Parrish, D.K. Lee, Cytogeographic distribution and genome size variation in prairie cordgrass (*Spartina pectinata* Bosc ex Link), Plant Mol. Biol. Report. 30 (2012) 1073–1079. https://doi.org/10.1007/s11105-012-0414-x.

[75] N. Baisakh, P.K. Subudhi, K. Arumuganathan, A.P. Parco, S.A. Harrison, C.A. Knott, M.D. Materne, Development and interspecific transferability of genic microsatellite markers in *Spartina spp.* with different genome size, Aquat. Bot. 91 (2009) 262–266. https://doi.org/10.1016/j.aquabot.2009.07.007.

Journal Pre-proof

[76] A. Baumel, M. Ainouche, R. Kalendar, A.H. Schulman, Retrotransposons and genomic stability in populations of the young allopolyploid species *Spartina anglica* C.E. Hubbard (Poaceae), Mol. Biol. Evol. 19 (2002) 1218–1227. https://doi.org/10.1093/oxfordjournals.molbev.a004182.

[77] A. Salmon, M.L. Ainouche, J.F. Wendel, Genetic and epigenetic consequences of recent hybridization and polyploidy in *Spartina* (Poaceae), Mol. Ecol. 14 (2005) 1163–1175. https://doi.org/10.1111/j.1365-294X.2005.02488.x.

[78] M. Ainouche, H. Chelaifa, J. Ferreira, S. Bellot, A. Ainouche, A. Salmon, Polyploid evolution in *Spartina*: dealing with highly redundant hybrid genomes, in: P.S. Soltis, D.E. Soltis (Eds.), Polyploidy Genome Evol., Springer Berlin Heidelberg, Berlin, Heidelberg, 2012: pp. 225–243. http://link.springer.com/10.1007/978-3-642-31442-1_12 (accessed October 12, 2016).

[79] A. Baumel, M.L. Ainouche, J.E. Levasseur, Molecular investigations in populations of *Spartina anglica* C.E. Hubbard (Poaceae) invading coastal Brittany (France), Mol. Ecol. 10 (2001) 1689–1701. https://doi.org/10.1046/j.1365-294X.2001.01299.x.

[80] P. Novák, P. Neumann, J. Macas, Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data, BMC Bioinformatics. 11 (2010). https://doi.org/10.1186/1471-2105-11-378.

[81] P. Novák, P. Neumann, J. Pech, J. Steinhaisl, J. Macas, RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads, Bioinformatics. 29 (2013) 792–793. https://doi.org/10.1093/bioinformatics/btt054.

[82] J. Boutte, B. Aliaga, O. Lima, J. Ferreira de Carvalho, A. Ainouche, J. Macas, M. Rousseau-Gueutin, O. Coriton, M. Ainouche, A. Salmon, Haplotype detection from next-generation sequencing in high-ploidy-level species: 45S rDNA gene copies in the hexaploid *Spartina maritima*, G3. 6 (2015) 29–40. https://doi.org/10.1534/g3.115.023242.

[83] S.F. Altschul, W. Gish, W. Miller, E.W. Myers, D.J. Lipman, Basic local alignment search tool, J. Mol. Biol. 215 (1990) 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

[84] B. Langmead, S.L. Salzberg, Fast gapped-read alignment with Bowtie 2, Nat. Methods. 9 (2012) 357–359. https://doi.org/10.1038/nmeth.1923.

[85] X. Huang, A. Madan, CAP3: a DNA sequence assembly program, Genome Res. 9 (1999) 868–877. https://doi.org/10.1101/gr.9.9.868.

[86] M. Margulies, M. Egholm, W.E. Altman, S. Attiya, J.S. Bader, L.A. Bemben, J. Berka, M.S. Braverman, Y.-J. Chen, Z. Chen, S.B. Dewell, L. Du, J.M. Fierro, X.V. Gomes, B.C. Godwin, W. He, S. Helgesen, C.H. Ho, G.P. Irzyk, S.C. Jando, M.L.I. Alenquer, T.P. Jarvie, K.B. Jirage, J.-B. Kim, J.R. Knight, J.R. Lanza, J.H. Leamon, S.M. Lefkowitz, M. Lei, J. Li, K.L. Lohman, H. Lu, V.B. Makhijani, K.E. McDade, M.P. McKenna, E.W. Myers, E. Nickerson, J.R. Nobile, R. Plant, B.P. Puc, M.T. Ronan, G.T. Roth, G.J. Sarkis, J.F. Simons, J.W. Simpson, M. Srinivasan, K.R. Tartaro, A. Tomasz, K.A. Vogt, G.A. Volkmer, S.H. Wang, Y. Wang, M.P. Weiner, P. Yu, R.F. Begley, J.M. Rothberg, Genome sequencing in microfabricated high-density picolitre reactors, Nature. 437 (2005) 376–380. https://doi.org/10.1038/nature03959.

[87] J. Jurka, V.V. Kapitonov, A. Pavlicek, P. Klonowski, O. Kohany, J. Walichiewicz, Repbase Update, a database of eukaryotic repetitive elements, Cytogenet. Genome Res. 110 (2005) 462–467. https://doi.org/10.1159/000084979.

[88] W. Bao, K.K. Kojima, O. Kohany, Repbase Update, a database of repetitive elements in eukaryotic genomes, Mob. DNA. 6 (2015). https://doi.org/10.1186/s13100-015-0041-9.

[89] L. Noe, G. Kucherov, YASS: enhancing the sensitivity of DNA similarity search, Nucleic Acids Res. 33 (2005) W540–W543. https://doi.org/10.1093/nar/gki478.

[90] W.L. Gerlach, J.R. Bedbrook, Cloning and characterization of ribosomal RNA genes from wheat and barley, Nucleic Acids Res. 7 (1979) 1869–1885. https://doi.org/10.1093/nar/7.7.1869.

[91] F. Madeira, Y. mi Park, J. Lee, N. Buso, T. Gur, N. Madhusoodanan, P. Basutkar, A.R.N. Tivey, S.C. Potter, R.D. Finn, R. Lopez, The EMBL-EBI search and sequence analysis tools APIs in 2019, Nucleic Acids Res. 47 (2019) W636–W641. https://doi.org/10.1093/nar/gkz268.

[92] J. Castresana, Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis, Mol. Biol. Evol. 17 (2000) 540–552. https://doi.org/10.1093/oxfordjournals.molbev.a026334.

[93] L.-T. Nguyen, H.A. Schmidt, A. von Haeseler, B.Q. Minh, IQ-TREE: A fast and effective stochastic

Journal Pre-proof

algorithm for estimating maximum-likelihood phylogenies, Mol. Biol. Evol. 32 (2015) 268–274. https://doi.org/10.1093/molbev/msu300.

[94] S. Kalyaanamoorthy, B.Q. Minh, T.K.F. Wong, A. von Haeseler, L.S. Jermiin, ModelFinder: fast model selection for accurate phylogenetic estimates, Nat. Methods. 14 (2017) 587–589. https://doi.org/10.1038/nmeth.4285.

[95] D.T. Hoang, O. Chernomor, A. von Haeseler, B. Quang Minh, L. Sy Vinh, Ufboot2: improving the ultrafast bootstrap approximation, Mol. Biol. Evol. 32 (2017) 518–522. https://doi.org/10.5281/zenodo.854445.

[96] M. Kimura, A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences, J. Mol. Evol. 16 (1980) 111–120.

[97] S. Kumar, G. Stecher, K. Tamura, MEGA7: Molecular Evolutionary Genetics Analysis version 7.0 for bigger datasets, Mol. Biol. Evol. 33 (2016) 1870–1874. https://doi.org/10.1093/molbev/msw054.

[98] J. Boutte, J. Ferreira de Carvalho, M. Rousseau-Gueutin, J. Poulain, C. Da Silva, P. Wincker, M. Ainouche, A. Salmon, Reference transcriptomes and detection of duplicated copies in hexaploid and allododecaploid *Spartina* species (Poaceae), Genome Biol. Evol. 8 (2016) 3030–3044. https://doi.org/10.1093/gbe/evw209.

[99] E.S. Dennis, W.J. Peacock, Knob heterochromatin homology in maize and its relatives, J. Mol. Evol. 20 (1984) 341–350. https://doi.org/10.1007/BF02104740.

[100] L. Aragón-Alcaide, T. Miller, T. Schwarzacher, S. Reader, G. Moore, A cereal centromeric sequence, Chromosoma. 105 (1996) 261–268.

[101] J. Jiang, S. Nasuda, F. Dong, C.W. Scherrer, S.-S. Woo, R.A. Wing, B.S. Gill, D.C. Ward, A conserved repetitive DNA element located in the centromeres of cereal chromosomes, Proc. Natl. Acad. Sci. 93 (1996) 14210–14213. https://doi.org/10.1073/pnas.93.24.14210.

[102] E.V. Ananiev, R.L. Phillips, H.W. Rines, Chromosome-specific molecular organization of maize (*Zea mays* L.) centromeric regions, Proc. Natl. Acad. Sci. 95 (1998) 13073–13078. https://doi.org/10.1073/pnas.95.22.13073.

[103] F. Dong, J.T. Miller, S.A. Jackson, G.-L. Wang, P.C. Ronald, J. Jiang, Rice (*Oryza sativa*) centromeric regions consist of complex DNA, Proc. Natl. Acad. Sci. 95 (1998) 8135–8140. https://doi.org/10.1073/pnas.95.14.8135.

[104] J. Greilhuber, Intraspecific variation in genome size in Angiosperms: identifying its existence, Ann. Bot. 95 (2005) 91–98. https://doi.org/10.1093/aob/mci004.

[105] M. Mascher, H. Gundlach, A. Himmelbach, S. Beier, S.O. Twardziok, T. Wicker, V. Radchuk, C. Dockter, P.E. Hedley, J. Russell, M. Bayer, L. Ramsay, H. Liu, G. Haberer, X.-Q. Zhang, Q. Zhang, R.A. Barrero, L. Li, S. Taudien, M. Groth, M. Felder, A. Hastie, H. Šimková, H. Staňková, J. Vrána, S. Chan, M. Muñoz-Amatriaín, R. Ounit, S. Wanamaker, D. Bolser, C. Colmsee, T. Schmutzer, L. Aliyeva-Schnorr, S. Grasso, J. Tanskanen, A. Chailyan, D. Sampath, D. Heavens, L. Clissold, S. Cao, B. Chapman, F. Dai, Y. Han, H. Li, X. Li, C. Lin, J.K. McCooke, C. Tan, P. Wang, S. Wang, S. Yin, G. Zhou, J.A. Poland, M.I. Bellgard, L. Borisjuk, A. Houben, J. Doležel, S. Ayling, S. Lonardi, P. Kersey, P. Langridge, G.J. Muehlbauer, M.D. Clark, M. Caccamo, A.H. Schulman, K.F.X. Mayer, M. Platzer, T.J. Close, U. Scholz, M. Hansson, G. Zhang, I. Braumann, M. Spannagl, C. Li, R. Waugh, N. Stein, A chromosome conformation capture ordered sequence of the barley genome, Nature. 544 (2017) 427–433. https://doi.org/10.1038/nature22043.

[106] M. Ayele, J. Doleže, M. Van Duren, H. Brunner, F.J. Zapata-Arias, Flow cytometric analysis of nuclear genome of the Ethiopian cereal tef [*Eragrostis tef* (Zucc.) *Trotter*], Genetica. 98 (1996) 211–215. https://doi.org/10.1007/BF00121369.

[107] Y.G. Gebre, E. Bertolini, M.E. Pè, A. Zuccolo, Identification and characterization of abundant repetitive sequences in *Eragrostis tef* cv. Enatite genome, BMC Plant Biol. 16 (2016). https://doi.org/10.1186/s12870-016-0725-4.

[108] A.H. Paterson, J.E. Bowers, R. Bruggmann, I. Dubchak, J. Grimwood, H. Gundlach, G. Haberer, U. Hellsten, T. Mitros, A. Poliakov, J. Schmutz, M. Spannagl, H. Tang, X. Wang, T. Wicker, A.K. Bharti, J. Chapman, F.A. Feltus, U. Gowik, I.V. Grigoriev, E. Lyons, C.A. Maher, M. Martis, A. Narechania, R.P. Otillar, B.W. Penning, A.A. Salamov, Y. Wang, L. Zhang, N.C. Carpita, M. Freeling, A.R. Gingle, C.T. Hash, B. Keller, P. Klein, S. Kresovich, M.C. McCann, R. Ming, D.G. Peterson, Mehboob-ur-Rahman, D. Ware, P. Westhoff, K.F.X. Mayer, J. Messing, D.S. Rokhsar, The *Sorghum bicolor* genome and the diversification of grasses, Nature. 457 (2009) 551–556. https://doi.org/10.1038/nature07723.

Journal Pre-proof

[109] J.L. Bennetzen, J. Schmutz, H. Wang, R. Percifield, J. Hawkins, A.C. Pontaroli, M. Estep, L. Feng, J.N. Vaughn, J. Grimwood, J. Jenkins, K. Barry, E. Lindquist, U. Hellsten, S. Deshpande, X. Wang, X. Wu, T. Mitros, J. Triplett, X. Yang, C.-Y. Ye, M. Mauro-Herrera, L. Wang, P. Li, M. Sharma, R. Sharma, P.C. Ronald, O. Panaud, E.A. Kellogg, T.P. Brutnell, A.N. Doust, G.A. Tuskan, D. Rokhsar, K.M. Devos, Reference genome sequence of the model plant *Setaria*, Nat. Biotechnol. 30 (2012) 555–561. https://doi.org/10.1038/nbt.2196.

[110] E. Kejnovsky, J.S. Hawkins, C. Feschotte, Plant transposable elements: biology and evolution, in: J.F. Wendel, J. Greilhuber, J. Dolezel, I.J. Leitch (Eds.), Plant Genome Divers. Vol. 1, Springer Vienna, Vienna, 2012: pp. 17–34. https://doi.org/10.1007/978-3-7091-1130-7_2.

[111] J.A. Ågren, Evolutionary transitions in individuality: insights from transposable elements, Trends Ecol. Evol. 29 (2014) 90–96. https://doi.org/10.1016/j.tree.2013.10.007.

[112] A. Baumel, M. Rousseau-Gueutin, C. Sapienza-Bianchi, A. Gareil, N. Duong, H. Rousseau, O. Coriton, R. Amirouche, S. Sciandrello, B. Duarte, I. Caçador, J.M. Castillo, M. Ainouche, *Spartina versicolor* Fabre: another case of *Spartina* trans-Atlantic introduction?, Biol. Invasions. 18 (2016) 2123–2135. https://doi.org/10.1007/s10530-016-1128-z.

[113] J.M. Sánchez, C. Sánchez, L. Navarro, Can asexual reproduction by plant fragments help to understand the invasion of the NW Iberian coast by *Spartina patens*?, Flora. 257 (2019) 151410. https://doi.org/10.1016/j.flora.2019.05.009.

[114] S. Renny-Byfield, M. Chester, A. Kovarik, S.C. Le Comber, M.-A. Grandbastien, M. Deloger, R.A. Nichols, J. Macas, P. Novak, M.W. Chase, A.R. Leitch, Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs, Mol. Biol. Evol. 28 (2011) 2843–2854. https://doi.org/10.1093/molbev/msr112.

[115] S. Renny-Byfield, A. Kovařík, M. Chester, R.A. Nichols, J. Macas, P. Novák, A.R. Leitch, Independent, rapid and targeted loss of highly repetitive DNA in natural and synthetic allopolyploids of *Nicotiana tabacum*, PLoS ONE. 7 (2012). https://doi.org/10.1371/journal.pone.0036963.

[116] B. Koukalova, A.P. Moraes, S. Renny-Byfield, R. Matyasek, A.R. Leitch, A. Kovarik, Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years, New Phytol. 186 (2010) 148–160. https://doi.org/10.1111/j.1469-8137.2009.03101.x.

[117] I.J. Leitch, L. Hanson, K.Y. Lim, A. Kovarik, M.W. Chase, J.J. Clarkson, A.R. Leitch, The ups and downs of genome size evolution in polyploid species of *Nicotiana* (Solanaceae), Ann. Bot. 101 (2008) 805–814. https://doi.org/10.1093/aob/mcm326.

[118] J.F. Wendel, D. Lisch, G. Hu, A.S. Mason, The long and short of doubling down: polyploidy, epigenetics, and the temporal dynamics of genome fractionation, Curr. Opin. Genet. Dev. 49 (2018) 1–7. https://doi.org/10.1016/j.gde.2018.01.004.

[119] D.G. Mobberley, Taxonomy and distribution of the genus *Spartina*, (1956).

[120] M. El Baidouri, M.-C. Carpentier, R. Cooke, D. Gao, E. Lasserre, C. Llauro, M. Mirouze, N. Picault, S.A. Jackson, O. Panaud, Widespread and frequent horizontal transfers of transposable elements in plants, Genome Res. 24 (2014) 831–838. https://doi.org/10.1101/gr.164400.113.

[121] J. Peccoud, R. Cordaux, C. Gilbert, Analyzing horizontal transfer of transposable elements on a large scale: challenges and prospects, BioEssays. 40 (2018). https://doi.org/10.1002/bies.201700177.

[122] J. Ferreira de Carvalho, H. Chelaifa, J. Boutte, J. Poulain, A. Couloux, P. Wincker, A. Bellec, J. Fourment, H. Bergès, A. Salmon, M. Ainouche, Exploring the genome of the salt-marsh *Spartina maritima* (Poaceae, Chloridoideae) through BAC end sequence analysis, Plant Mol. Biol. 83 (2013) 591–606. https://doi.org/10.1007/s11103-013-0111-7.

[123] A. Cavé-Radet, D. Giraud, O. Lima, A. El Amrani, M. Aïnouche, A. Salmon, Evolution of small RNA expression following hybridization and allopolyploidization: insights from *Spartina* species (Poaceae, Chloridoideae), Plant Mol. Biol. (2019). https://doi.org/10.1007/s11103-019-00931-w.

[124] H. Groves, J. Groves, *Spartina townsendii* , 1880.

[125] Foucaud, Un Spartina inédit, Ann Soc Sci Nat Char Inf. 32 (1897) 220–22.

[126] M.A. Garrido-Ramos, Satellite DNA in plants: more than just rubbish, Cytogenet. Genome Res. 146 (2015) 153–170. https://doi.org/10.1159/000437008.

[127] M. Garrido-Ramos, Satellite DNA: an evolving topic, Genes. 8 (2017) 230.

Journal Pre-proof

https://doi.org/10.3390/genes8090230.

[128] M. Plohl, A. Luchetti, N. Meštrović, B. Mantovani, Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin, Gene. 409 (2008) 72–82. https://doi.org/10.1016/j.gene.2007.11.013.

[129] M. Plohl, N. Mestrovic, B. Mravinac, Satellite DNA evolution, in: M.A. Garrido-Ramos (Ed.), Genome Dyn., S. KARGER AG, Basel, 2012: pp. 126–152. https://doi.org/10.1159/000337122.

[130] J. Dolezel, J. Bartos, H. Voglmayr, J. Greilhuber, Nuclear DNA content and genome size of trout and human, Cytometry. 51A (2003) 127–128. https://doi.org/10.1002/cyto.a.10013.

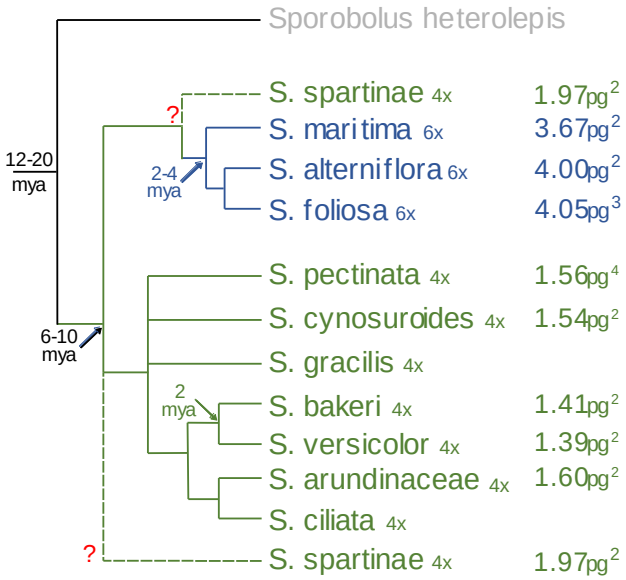


Fig. 1 Phylogenetic relationships between tetraploid (2n = 4x = 40; in green) and hexaploid (2n = 6x = 60 or 62; in blue) species of *Spartina* Schreb. Question marks represent the two putative phylogenetic positions of *S. spartinae*. Divergence times (millions of years ago - mya) represent upper limits estimated. Genome sizes (pg/2C) are indicated for each species. Superscript letters next to the genome size estimates correspond to the original source of the data (see Table 1).

Journal Pre-proof

Journal Pre-proof

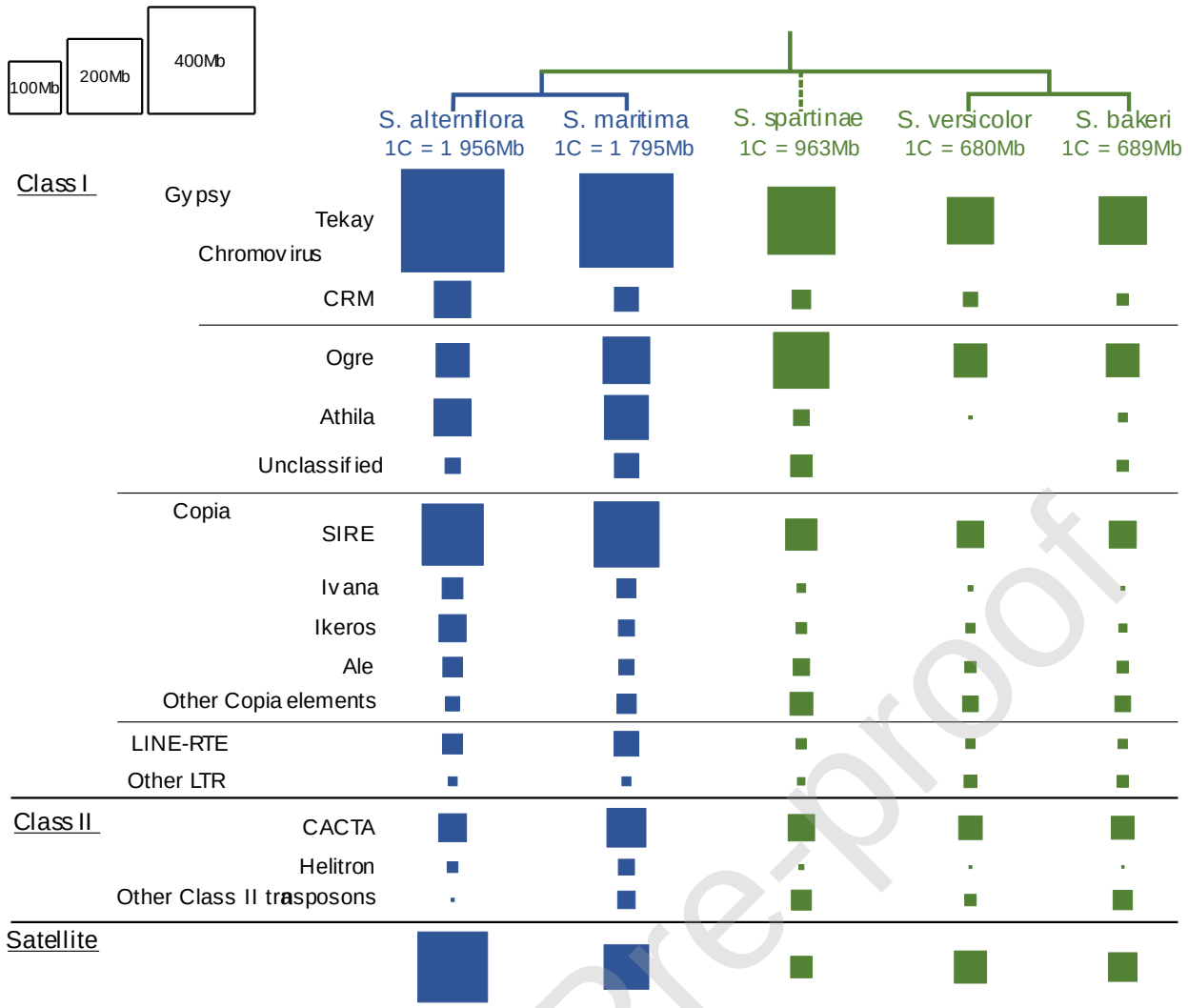


Fig. 2 Proportion of TEs and satellite sequences in the genomes of the five studied *Spartina* species. Tetraploid species are coloured in green and hexaploid species in blue.

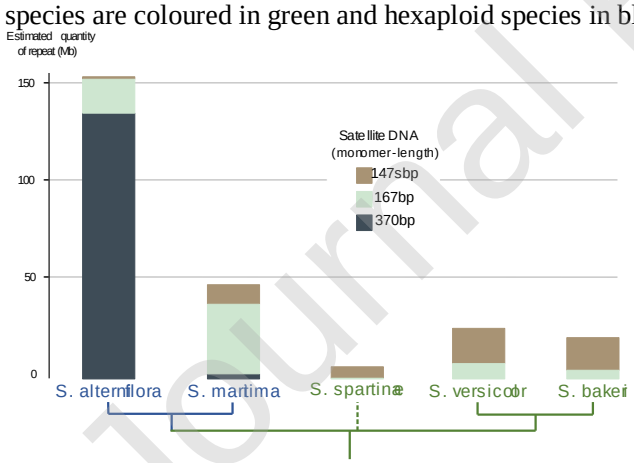


Fig. 3 Amount of satellite sequences found in hexaploid and tetraploid *Spartina* species. Three different satellite repeats were detected, two of which (i.e. 147bp and 167 bp repeats) showed sequence homology with centromeric repeats of grasses. The third satellite repeat of 370bp was abundant in *S. alterniflora*, where it represents approximately 7% of the genome.

Journal Pre-proof

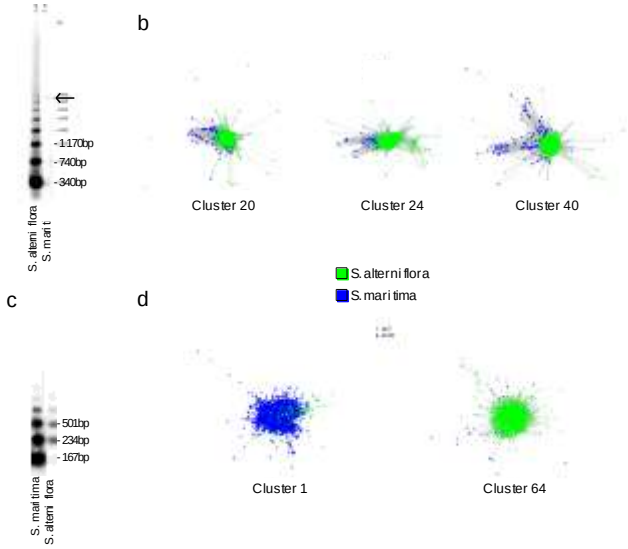


Fig. 4 Southern blot hybridization of genomic DNA from *S. alterniflora* and S. *maritima* with the 370-bp satellite (A) and the 167-bp satellite (C). (A) Regularly spaced hybridization signals in *S. alterniflora* show that the 370-bp satellite was abundant and tandemly repeated in this species. In contrast, in *S. maritima* has only a weak signal (see arrow) which confirms its presence, but in limited quantity. (B) Comparative graph genomic analysis of the 370-bp satellite (partitioned in 3 clusters). Nodes symbolize reads (green for *S. alterniflora* and blue for *S. maritima*) and the sizes of edges are proportional to the divergence between two related reads. Reads from 370-bp satellite exhibit a relatively low divergence in *S. alterniflora* (tightly clustered reads) compared with *S. maritima*, indicating a burst of divergence in the latter. (C) Southern blot hybridization confirms that the 167-bp satellite is more abundant in *S. maritima* genome than in *S. alterniflora*. (D) Graphical analyses indicate a genome specific diversity of the 167-bp satellite (each cluster specific to one species).

Journal Pre-proof

Journal Pre-proof

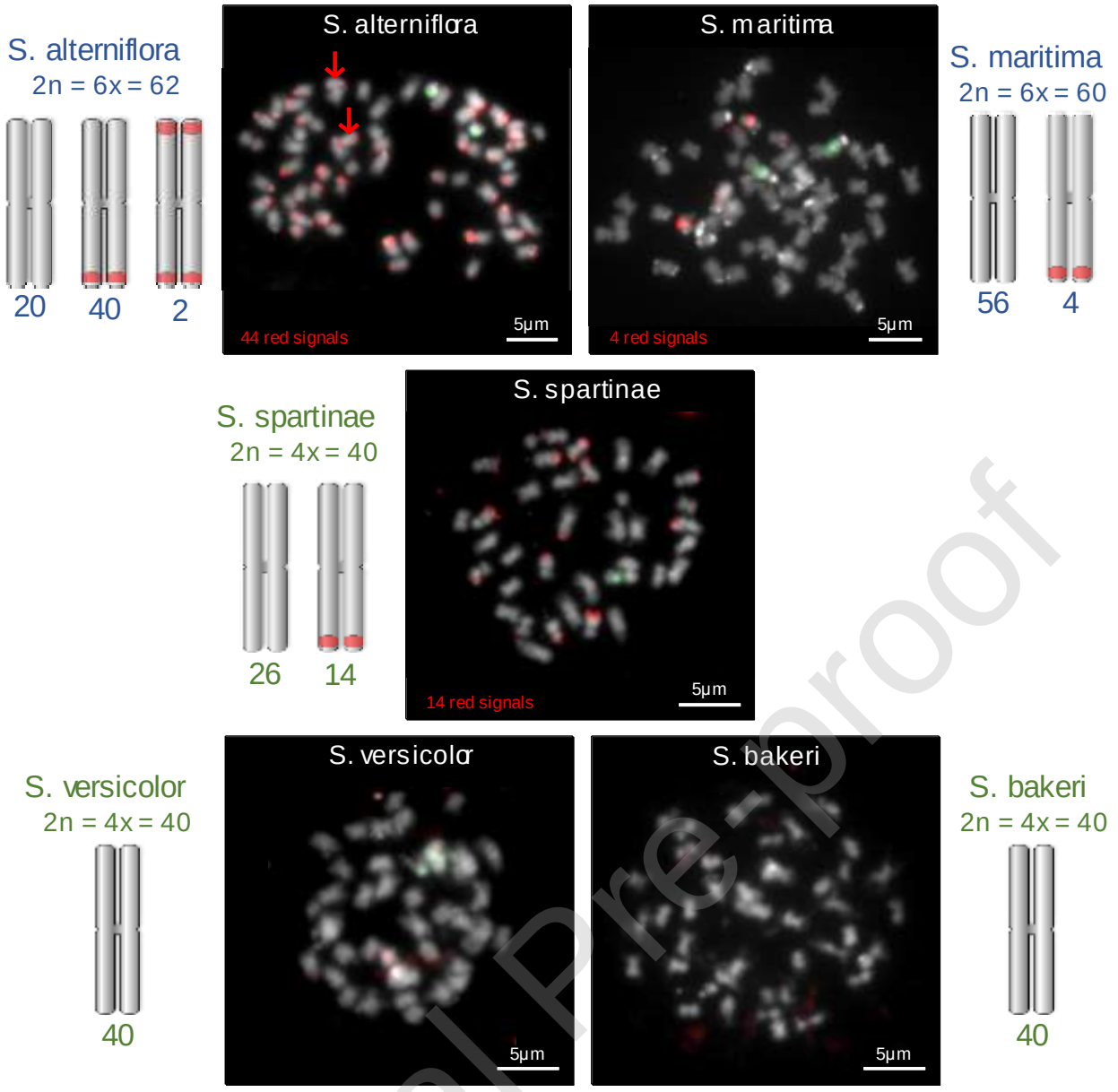
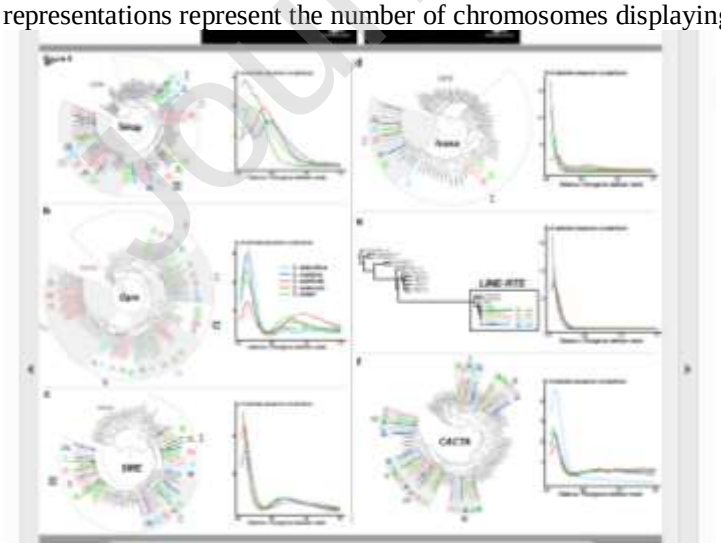


Fig. 5 FISH analyses of the newly identified satellite DNA (370 bp) on *Spartina* chromosomes. The satellite sequence was labeled with Texas red (red signal), rDNA was labeled with Alex-488 (green signal), and chromosomes were counterstained with DAPI (grey). Red arrows indicate the two chromosomes in *S. alterniflora* which contain the DNA satellite on both chromosome arms. The numbers beneath the chromosome representations represent the number of chromosomes displaying none, single or two red signals.



Journal Pre-proof

Fig. 6 Phylogenetic analyses and pairwise sequence comparisons (Kimura 2-parameters distance) of the main TEs found in *Spartina* species based on reverse transcriptase coding sequences for Class I elements (a-e) or transposase coding sequences for Class II *CACTA* elements (f). According to phylogenetic relationships between sequences, TEs were classified into different monophyletic lineages (coded in roman numbers). The estimated diversity (number of detected haplotypes) and estimates of their genome contribution (in Mb) are indicated for each lineage. The sequences are color coded to species as shown in (b).

Journal Pre-proof

Journal Pre-proof

Table 1 Ploidy levels, chromosome numbers and genome sizes estimated in *Spartina* species. Species names are mentioned according to Mobberley, 1953 [119] and from the revised nomenclature by Peterson et al.[68]. Genome sizes were converted from picograms (pg) to Megabases (Mb) using 1 pg = 978 Mbp [130].

Species	**Synonyms**	**2n**	**Genome size**	
			2C in pg	**1C in Mb**
Spartina maritima (Curtis) Fernald	***Sporobolus maritimus*** (Curtis) P.M.Peterson & Saarela, comb. nov.	6x = 60[1]	3.67[2]	1 795
Spartina alterniflora Loisel.	***Sporobolus alterniflorus*** (Loisel.) P.M.Peterson & Saarela, comb. nov.	6x = 62[1]	4.00[2]	1 956
Spartina foliosa Trin.	***Sporobolus foliosus*** (Trin.) P.M.Peterson & Saarela, comb. nov.	6x = 62[3]	4.05[3]	1 980
Spartina spartinae (Trin.) Merr.	***Spartina argentinensis*** Parodi ***Sporobolus spartinus*** (Trin.) P.M.Peterson & Saarela, comb. nov.	4x = 40[1]	1.97[2]	963
Spartina pectinata Link.	***Sporobolus michauxianus*** (Hitchc.) P.M.Peterson & Saarela, comb. nov.	4x, 6x, 8x = 40, 60, 80[4]	4x: 1.56[4] 6x: 2.33[4] 8x: 3.06[4]	763 1 139 1 496
Spartina cynosuroides (L.) Roth	***Sporobolus cynosuroides*** (L.) P.M.Peterson & Saarela, comb. nov.	4x = 40[1]	1.54[2]	753
Spartina bakeri Merr.	***Sporobolus bakeri*** (Merr.) P.M.Peterson & Saarela, comb. nov.	4x = 40[1]	1.41[2]	689
Spartina gracilis Trin.	***Sporobolus hookerianus*** P.M.Peterson & Saarela, nom. nov.	4x = 40[1]	-	-
Spartina versicolor (Fabre)	***Spartina patens*** (Aiton) Muhl. ***Sporobolus versicolor*** (Fabre) P.M.Peterson & Saarela, comb. nov. ***Sporobolus pumilus*** (Roth) P.M.Peterson & Saarela, comb. nov.	4x = 40[1]	1.39[2]	680
Spartina arundinacea (Thouars) Carmich.	***Sporobolus mobberleyanus*** P.M.Peterson & Saarela, nom. nov.	4x = 40[1]	1.60[2]	782
Spartina ciliata Brongn.	***Sporobolus coarctatus*** (Trin.) P.M.Peterson & Saarela, comb. nov.	4x = 40[1]	-	-

Journal Pre-proof

[1][58]; [2]This study, [3][60]; [4][74].

Table 2 Genome sizes and main characteristics of the sequence datasets analyzed for each species.

			S. alterniflora	*S. maritima*	*S. spartinae*	*S. versicolor*	*S. bakeri*
Ploidy level and chromosome number			2n = 6x = 62	2n = 6x = 60	2n = 4x = 40	2n = 4x = 40	2n = 4x = 40
Estimated genome size (1C in Mb)			1 956	1 795	963	680	689
Sample size (number of reads processed by RepeatExplorer)			854 068	2 000 000	2 000 000	2 000 000	2 000 000
Main features of the sample	**Plastid DNA**		45 769 reads	44 596 reads	79 956 reads	56 988 reads	75 966 reads
	Nuclear DNA	**Repetitive DNA**	67 clusters 388 003 reads	59 clusters 911 400 reads	72 clusters 874 800 reads	63 clusters 658 400 reads	74 clusters 659 800 reads
		Ribosomal DNA	13 539 reads	21 481 reads	30 719 reads	16 757 reads	18 791 reads
		Non-repetitive DNA	427 755 reads	1 042 929 reads	1 049 448 reads	1 286 553 reads	1 262 079 reads

Journal Pre-proof

Journal Pre-proof

Table 3 Annotation and estimated amounts of major TE lineages found in the five investigated *Spartina* species

Order	Superfamily	***S. alterniflora***			***S. maritima***			***S. spartinae***			***S. versicolor***			***S. bakeri***		
		1C = 1,956 Mb			1C = 1,795 Mb			1C = 963 Mb			1C = 680 Mb			1C = 689 Mb		
		# cluster	Estimated quantity in the genome (Mb)	% of genome	# cluster	Estimated quantity in the genome (Mb)	% of genome	# cluster	Estimated quantity in the genome (Mb)	% of genome	# cluster	Estimated quantity in the genome (Mb)	% of genome	# cluster	Estimated quantity in the genome (Mb)	% of genome
Class I Retrotransposon		**39**	**693.5**	**35.48%**	**38**	**681.8**	**37.99%**	**37**	**367.5**	**38.17%**	**30**	**165.8**	**24.37%**	**34**	**172.0**	**24.55%**
LTR	*Gypsy*	19	491.0	25.11%	23	473.4	26.38%	15	294.9	30.62%	10	117.2	17.22%	15	124.0	17.58%
	Copia	15	186.2	9.53%	12	185.1	10.32%	18	67.0	6.97%	14	40.0	5.89%	14	40.5	5.89%
	Other LTR	2	1.7	0.09%	1	1.8	0.10%	1	1.2	0.13%	2	3.8	0.56%	2	3.1	0.45%
LINE-RTE		1	13.6	0.7%	1	20.7	1.15%	1	3.7	0.38%	1	3.1	0.45%	1	2.9	0.42%
SINE		2	1.0	0.05%	1	0.8	0.04%	2	0.7	0.07%	3	1.7	0.25%	2	1.5	0.21%
Class II DNA-transposons		**8**	**30.3**	**1.55%**	**10**	**69.4**	**3.87%**	**21**	**38.0**	**3.95%**	**13**	**23.4**	**3.45%**	**13**	**30.6**	**4.43%**
TIR	*CACTA*	6	26.0	1.33%	5	50.7	2.82%	9	23.5	2.44%	6	18.9	2.78%	6	18.0	2.61%
	Other TIR	1	0.4	0.02%	5	10.2	0.57%	10	14.0	1.43%	6	4.3	0.64%	6	12.4	1.79%
Helitron		1	3.9	0.2%	1	8.5	0.48%	2	0.8	0.08%	1	0.2	0.03%	1	0.2	0.03%
Tandem repeats / satellite		**20**	**164.3**	**8.40%**	**11**	**66.4**	**3.71%**	**14**	**15.6**	**1.62%**	**20**	**34.7**	**5.10%**	**27**	**27.7**	**4.01%**
TOTAL Repeat DNA		**67**	**888.1**	**45.43%**	**60**	**817.6**	**45.57%**	**72**	**421.1**	**43.74%**	**63**	**223.9**	**32.92%**	**74**	**230.3**	**33.40%**

Journal Pre-proof

Journal Pre-proof

Journal Pre-proof