universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Characterization of the repeatome and karyotype of *Capsicum flexuosum* (Solanaceae) with special emphasis on tandem repeats"

verfasst von / submitted by

Nadine Jany, BSc

angestrebter akademischer Grad / in partial fulfilment of the requirements for the degree of

Master of Science (MSc)

Wien, 2022 / Vienna, 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 877
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Masterstudium Genetik und Entwicklungsbiologie UG2002
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Hanna Schneeweiss, Privatdoz.

Acknowledgments

First, I want to express my deepest gratitude to Univ.-Prof. Mag. Dr. Hanna Schneeweiss for her patience and guidance during my time working on this thesis. Thank you for always having an open door and ear for me. I will never forget the wonderful time I had with you and the whole Plant Evolutionary Cytogenetics research group.

Furthermore, I want to thank my parents, my sister and my brother for their love and continuous support throughout my whole studies. Thank you for always having my back and believing in me.

Finally, I want to thank my partner. You motivated and encouraged me whenever I needed it. Thank you for always standing by my side.

TABLE OF CONTENTS

Part I – INTRODUCTION

1. Plant genome evolution

Plant genomes are very variable concerning their sizes, chromosome numbers and DNA composition. Although every species has a specific number of chromosomes, these vary widely among species, genera and families. Both increases and decreases in the chromosome numbers frequently accompany diversification and speciation of plants (Heslop-Harrison & Schwarzacher, 2011).

The chromosome numbers of angiosperms vary 160-fold with a range of $2n = 4$ up to $2n = 640$. The lowest chromosome number of $n = 2$ was reported in *Ornithogalum tenuifolium*/Hyacinthaceae, *Zingeria biebersteiniana* and *Colpodium versicolor*/Poaceae, *Brachycome dichromosomatica* and *Haplopappus gracilis*/Asteraceae (Cremonini, 2005) and *Rhynchospora tenuis*/Cyperaceae (Vanzela et al., 1996) whereas the highest number of $n = 320$ was found in *Sedum suaveolens* (Uhl, 1978). The majority of known angiosperms, however, have chromosome numbers ranging between $n = 7$ to $n = 20$ (Grant et al., 1982; Masterson et al., 1994; Weiss-Schneeweiss & Schneeweiss, 2013).

Plant genome sizes are highly variable as they span four orders of magnitude (Gregory et al., 2007). They range from 60 megabases (Mb) in the carnivorous corkscrew plant *Genlisea aurea* to the rare Japanese plant *Paris japonica* whose genome size amounts to 152000 Mb (Leitch et al., 2019). The most frequently reported plant genome size is about 500 Mb which suggests that small genome sizes in plants prevail (Michael, 2014). Despite such wide range of chromosome numbers and genome sizes observed among angiosperms, their genomes have rather similar gene content (Weiss-Schneeweiss & Schneeweiss, 2013).

The process of polyploidy changes both chromosome numbers and genome size. Genome size increases and decreases can additionally be caused by either amplification or reduction of copy numbers of various repetitive DNA types (discussed below), whereas dysploidy results in (step-wise) changes in chromosome numbers.

1.1. Structural rearrangements

Processes that alter the karyotype structure affect both the size and morphology of chromosomes. Each such change in chromosome structure usually starts with one or several double-strand breaks (DSBs). If a double-strand break is repaired correctly, the original structure of the DNA molecule is restored. However, if two or more DSBs are repaired at the same time, mis-repairs can occur and these can lead to an alteration of structure and DNA content of the chromosomes involved. When homologous sequences are used as templates to repair a DSB, the repair mechanism is most likely accurate. Broken chromosomal ends can also be joined directly without or only with very little homology. This very error-prone mechanism is called non-homologous end joining (NHEJ) and can lead to insertions, deletions, inversions and translocations. Non-allelic homologous recombination is another type of mis-repair. It can most often lead to deletions (removal of chromosomal segments) or inversions (of the segment between the breaking points) if intrastrand-recombination occurs. Unequal homologous recombination between sister chromatids initiates duplications and deletions. Reciprocal chromosomal translocations can occur when repeats of non-allelic homologous or non-homologous chromosomes are combined through mis-repair (Lysák & Schubert, 2013).

Structural chromosomal rearrangements can be either balanced or unbalanced. In the case of unbalanced rearrangements (e.g. insertions, deletions, duplications) a chromosomal fragment is either gained or lost and usually leads to severe consequences for the organism. Balanced rearrangements like translocation and inversions on the other hand don't lead to the loss or gain of genetic material in the individual in which they occur (Rieseberg, 2001; Lysák & Schubert, 2013).

1.2. Numerical chromosomal changes and accompanying rearrangements

Two main processes affect chromosome numbers in evolution: dysploidy and polyploidy. One of the most common processes involved in the evolution of angiosperm genomes is polyploidy. It results in the multiplication of entire sets of chromosomes and thus also in the instantaneous increase of genome size. Two main types of polyploidy can be distinguished:

allo- and autopolyploidy. Autopolyploidy involves the multiplication of whole sets of chromosomes within one species. Allopolyploidy, on the other hand, refers to the merger of the chromosome sets of two (sub) species that is accompanied by chromosome set multiplication (Stebbins, 1971; Guerra, 2008; Weiss-Schneeweiss & Schneeweiss, 2013). Cytogenetic and genomic analysis revealed that all land plants have experienced at least one whole genome duplication (Soltis et al., 2009; Van de Peer et al., 2009; Lysák & Schubert, 2013), and in many plant groups polyploidy is ongoing.

The second process contributing to chromosome number variation is dysploidy. It results in either an increase (ascending dysploidy) or decrease (descending dysploidy) of chromosome numbers through structural chromosomal rearrangements. Dysploidy, in contrast to polyploidy, does not result in the multiplication of whole chromosome sets. Instead, chromosome number changes stepwise via karyotype rearrangements, e.g., from $n = 7$ to $n = 6$, affecting the number of centromeres, but having little effect on the total amount of DNA. Several mechanisms can lead to structural chromosomal rearrangements in angiosperms and cause numerical changes in chromosome sets.

One of the processes that have been shown to be particularly involved in dysploidy is Robertsonian exchange/Robertsonian translocation. This kind of rearrangement occurs via a centric fusion-fission event, where the nonhomologous chromosomes break and/or reunite in a centromeric region (Friebe et al., 2005; Guerra, 2008). Translocations are often proceeded or accompanied by pericentric inversions.

In contrast to dysploidy, the loss or gain of entire chromosomes is referred to as aneuploidy (Guerra, 2008). This type of chromosomal mutation leads in most cases to a detrimental change in gene balance. Aneuploidy, therefore, does not play an important role in the evolution of plant genomes (Weiss-Schneeweiss & Schneeweiss, 2013).

2. Repetitive DNA

The eukaryotic genome is not only composed of unique coding sequences but also contains large (and varying) amounts of repetitive DNAs. In general, there are two classes of repetitive DNAs: dispersed and tandemly arranged DNAs. Dispersed repetitive DNAs comprise mainly DNA transposons, retrotransposons and retrotransposed sequences. Ribosomal RNA genes

and satellite DNAs which also include centromeric tandem repeats and telomeric DNA on the other hand are arranged in tandems (López-Flores & Garrido-Ramos, 2012; Biscotti et al., 2015b).

The discrepancy between the amount of DNA in a genome (C-value), in contrast to its complexity indicated by the number of coding sequences needed for organismal development and function, is referred to as the C-Value enigma (Gregory, 2005). This paradox can be explained by changes in composition and the amount of repetitive DNA families that can vary even between closely related species (Garrido-Ramos, 2017).

2.1. Satellite DNAs

Satellite DNA represents the tandemly arranged copies of repetitive DNA motif (monomer). Satellite DNAs are highly abundant in eukaryotic genomes and consist of tens to thousands of nucleotides long monomers of varying DNA sequences organized in continuous arrays that can reach a length of up to 100 Mbp (Kubis et al., 1998; Macas et al., 2002; Plohl et al., 2012; Garrido-Ramos, 2015). As satellite DNAs are one of the most dynamic components of the eukaryotic genome they are excellent markers to study the evolution of karyotypes in closely related species (Lim et al., 2000; Cuadrado & Jouve, 2002; Navrátilová et al., 2003; Pires et al., 2004). They are mainly found in the constitutive heterochromatin at pericentromeric and subtelomeric regions and additionally can occur in the centromeric region (Schmidt & Heslop-Harrison, 1998; Macas et al., 2000, 2007; Garrido Ramos, 2015). In some species, satellite DNAs are also found at interstitial chromosomal regions (Plohl et al., 2012).

Satellite DNAs are non-coding and are not subjected to strong selection. They are therefore evolving quickly within the genomes. Novel satellite DNA sequences are known to arise and amplify within plant genomes with high frequency and are therefore often species- or genus-specific (Weiss-Schneeweiss & Schneeweiss, 2013). Specific and new satellite DNA types often also arise within B- and sex-chromosomes (Sandery et al., 1990; Kejnovsky et al., 2009). Satellite DNAs are also known to contribute to genome restructuring following auto- and particularly allopolyploidization, as the parent-specific variants might spread to both parental genomes and new variants can arise within a short time (Skalicka et al. 2005;

Koukalova et al., 2010). Satellite DNAs, due to high levels of sequence and loci number variation are often used as cytological markers for chromosome identification.

The number of satellite DNA families, as well as their copy numbers, are highly variable among genomes. In plants, satellite DNAs can make up between 0.1% and 36% of the whole genome (Macas et al., 2000; de la Herrán et al., 2001; Hribová et al., 2010; Ambrožová et al., 2011; Cížková et al., 2013; Emadzade et al., 2014; Garrido-Ramos, 2015). This variation is one of the reasons for the significant differences in genome sizes between species that are related, but also between different cytotypes or populations of the same species (Macas et al., 2000; Ambrožová et al., 2011; Emadzade et al., 2014; Kopecna et al., 2014). Sometimes this difference can be traced back to one single satellite DNA that was amplified/reduced in various proportions in related species or populations, but it can also be caused by differential amplification of several satellite DNA families (Garrido-Ramos, 2017). The levels and dynamics of amplification/reduction of satellite DNAs of the genome, monomer sequence diversity and genomic/chromosomal localization can vary widely between different plant groups (Schmidt & Heslop-Harrison, 1998; Chester et al., 2010). Some plant genomes harbour many different satellite DNA families most of which are amplified only moderately, whereas other genomes contain only one or a few satellite DNA families, but highly amplified (Weiss-Schneeweiss & Schneeweiss, 2013). A combination of these two extremes is also possible.

The original view of satellite DNAs representing simply “junk or parasitic DNA” has changed over time. Satellite DNAs are now known to participate in some of the functional aspects of the genome. They are often present in the centromeres and the pericentromeric heterochromatin. It has also been proposed that satellite DNAs may play a role in the organization, pairing and segregation of chromosomes (Plohl et al., 2008, 2012, 2014). Additionally, the transcripts of satellite DNAs were shown to take part in controlling the elongation, capping and replication of telomeres, in assembling the kinetochore, in the transcriptional response to stress, in maintaining and establishing epigenetic regulation of heterochromatin and finally in the modulation of gene expression (Pezer et al., 2012; Biscotti et al., 2015a; Garrido-Ramos, 2017; Grenfell et al., 2017).

2.2. rDNA

Markers most commonly used for karyotype characterization in most of the eukaryotic genomes are 35S and 5S rRNA genes. Both of these rDNA types code for ribosomal RNAs that are necessary for building ribosomes and thus translation and are therefore very conserved in their coding regions. They are the only coding genomic sequences that are indispensable for the function of the cell and whole organism that are also classified as tandemly repetitive DNAs. 35S rDNA and 5S rDNA are organized in most plant families in unlinked arrangements (independent in loci number and location), which means that the numbers and chromosomal locations of these two loci types are not correlated. A few exceptions to this rule are in families Asteraceae and in Gymnosperms (Garcia et al., 2009) where both rDNA types are present in genomes in linked arrangement. Both rRNA gene types are present in the plant genomes in varying numbers of loci (minimum one locus of 5S rDNA and one locus of 35SrDNA; Małuszyńska et al., 1998). Each locus is composed of thousands of copies (monomers) in tandem arrangement. A single 5S rRNA monomer is organized fairly simply. It consists of the conserved coding unit of the 5S rRNA gene (c. 121 bp) and one non-transcribed spacer (NTS) of varying length and sequence. The 35S rDNA unit (also known as 45S rDNA) is organized in a more complex way. The transcribed monomer consists of conserved 18S, 5.8S and 25/28S rRNA genes, which are separated by two non-coding internal transcribed spacers (ITS1 + ITS2). Two other non-coding regions, external transcribed spacer (ETS) and intergenic spacer (IGS) flank the rDNA genic regions (Roa & Guerra, 2015). The 35S rRNA genes are found in the nucleolar organizing regions (NORs) which can be identified in chromosomes as they form secondary constrictions or satellites (Volkov et al., 2004). 35S rRNA genes are transcribed in the cell nucleus by RNA polymerase I, whereas 5S rRNA genes are transcribed independently by RNA polymerase III (Garcia & Kovařík, 2013).

Analyses of 784 species belonging to 175 genera of angiosperms and gymnosperms revealed that 54.4% of the taxa possessed one single 5S rDNA locus. In those species, the 5S rDNA locus mostly mapped to the short chromosomal arm (58.7%), in most cases in its proximal region (52.0%). This preferential positioning of the 5S rDNA suggests its non-random distribution in the genome (Roa & Guerra, 2015). A similar study showed that the number of 35S rDNA sites varied widely in angiosperm with monocentric chromosomes, from 2 to 32 sites (= signals in the diploid chromosome set), with an average number of 5.1 sites and a

median of 4 sites per genome (Roa & Guerra, 2012). This suggested the presence of, on average, two 35S rDNA loci per genome.

There are several reasons why the 5S and 35S rDNA loci are used as markers for karyotype characterization. First, their coding regions are highly conserved so they can be localized by FISH (fluorescence *in situ* hybridization) in chromosomes of numerous plant groups using standard, "universal" probes, without any further information on their individual genome composition. Second, in contrast to the conserved nature of their coding regions, the localization and number of loci of rDNA can change rather quickly during diversification and speciation. Third, sequences of spacer regions are used as first-choice molecular markers to infer a phylogenetic relationship within the taxa groups. Finally, rDNA loci exhibit a high inter- and intragenomic mobility (change in loci number and position), particularly, but not only following polyploidization during species differentiation (Weiss-Schneeweiss & Schneeweiss, 2013). All these provide valuable information for evolutionary comparisons.

2.3. Transposable elements

Transposable elements (TEs) are typically 100 bp to over 22000 bp long mobile DNA sequences that can replicate independently in the host genome (Arkhipova et al., 2019). TEs are very abundant in plant genomes. Their content in the genomes ranges from 14% in *Arabidopsis thaliana* (The Arabidopsis Genome Initiative, 2000) to 80% in maize (Schnable et al., 2009).

TEs replicate in the genome and insert their copies in various genomic locations. These, often high, numbers of insertions can affect genome structure and function via genomic rearrangements, unequal and ectopic recombination, movement and exon shuffling, gene capture and by providing new material for centromeric and intronic structures (Devos et al., 2002; Wright et al., 2003; Morgante et al., 2005; Lisch, 2012; Sharma et al., 2013; Bennetzen & Wang, 2014; Vitte et al., 2014). Direct insertions of TEs within or close to the genes can lead to a disruption of the gene function or promote new functions through alternative splicing (Varagona et al., 1992; Leprince et al., 2001) or alternative promoter control (Kashkush et al., 2002).

In general, transposable elements can be divided into two groups based on their transposition modes. Class I retroelements are replicated through RNA intermediates that are reverse transcribed back into DNA and the (many) new copies are inserted in new genomic locations. This mechanism is referred to as "copy-and-paste" as the original element stays intact. Class II DNA transposons, on the other hand, transpose via a DNA intermediate. As the element is excised (by transposase) from its original genomic location and inserted at a new location the mechanism is called "cut-and-paste" (Finnegan, 1989).

Long terminal repeat retroelements (LTR-retroelements)

Together with tyrosine recombinase mobilized elements and target primed non-LTR elements, long terminal repeat (LTR) retrotransposons make up the three major subclasses of Class I retroelements (Wells & Feschotte, 2020). In size, LTR elements range from 3500 to over 22000 bp (Kumar & Bennetzen, 1999). They are the most abundant repeats in the majority of plant genomes and can make up 70% or more of the genome in some species (Sanmiguel & Bennetzen, 1998; Kumar & Bennetzen 1999; Li et al. 2004).

LTR retroelements carry one coding region that is flanked upstream and downstream by identical long terminal repeats, each carrying a regulatory sequence. The internal coding domain encodes the *gag* and *pol* genes which are expressed as a single polycistronic RNA. The *pol* gene encodes for an RNAse H (RH), a retrotranscriptase (RT), a protease (PR) and an integrase (IN). The *gag* gene encodes for one protein called GAG (Kumar & Bennetzen 1999; Havecker et al., 2004). All these genes are necessary for the life cycle and movement of the element in the genome.

Two major distinct evolutionary superfamilies are distinguished within LTR retroelements: Ty3-*gypsy* and Ty1-*copia* elements (Wicker & Keller, 2007). These families can be further divided into more lineages. In plants, there were several distinct *gypsy* lineages identified: Athila, Tat, Galadriel, Reina, CRM/CR, and Del/Tekay. In the *Copia* superfamily at least seven lineages were identified: TAR/Tork, Angela/Tork, GMR/Tork, Maximus/Sire, Ivana/Oryco, Ale/Retrofit, and Bianca (Wicker & Keller, 2007; Du et al., 2010; Llorens et al., 2011). The difference between the two major superfamilies lies in the position of the integrase within the coding region. When the integrase is positioned downstream of the reverse transcriptase the element belongs to the *Gypsy* superfamily; when the integrase is found in an upstream position the element is identified as a member of the *Copia* superfamily (Neumann et al., 2019; Orozco-Arias et al., 2019).

3. Genus *Capsicum*

Capsicum belongs to the Solanaceae family and comprises approximately 35 to 40 species (Carrizo García et al., 2013, 2016). They are diploid self-pollinating plants and belong to the asteroid clade of eudicots, which includes more than 3000 other species from all over the world. *Capsicum* is closely related to tomato, tobacco, potato, eggplant and petunia (Kim et al., 2014).

The genus originated in the Americas. Currently, it can be found in tropical and temperate areas from central Argentina to Paraguay and Brazil to Mexico (Carrizo García et al., 2016). Five of the *Capsicum* species were domesticated by American natives (*C. annum var. annum, C. chinense, C. pubescens, C. frutescens, C. baccatum*; Moscone et al., 2007) and were brought to the rest of the world after Columbus set foot on the continent (Scaldaferro et al., 2006). *Capsicum* was one of the first crops that were domesticated in the Western Hemisphere (Aguilar-Melendez et al., 2009). Today, the cultivated species of the genus are economically very important as they are not only used as spice (hot peppers) but also have great significance in the human diet as vegetable (sweet peppers; Scaldaferro et al., 2006). Additionally, some of the *Capsicum* species are also used for medical and ornamental purposes (Hunziker, 1979; Pickersgill, 1991; Eshbaugh, 1993; Heiser, 1995).

3.1. Morphology

The majority of the plants of the genus *Capsicum* grow as perennial shrubs, although there are some exceptions like the species of *Capsicum annum* complex (*C. annum, C. chinense, C. frutescens*) which have a biennial herbaceous growth form. There are also some taxa like *C. eximium* and *C. caballeroi* which have shrubby, arborescent and herbaceous forms. *C. rhomboideum* and *C. parvifolium* can even grow as trees (Moscone et al., 2007).

For a long time, the colour of the corolla has been used to identify and characterize the different cultivated and wild *Capsicum* species. Previously, the species were subdivided into a "purple flowering group" which combines species with pink, violet and lilac flowers and a "white flowering group" (Pickersgill, 1991). For a few years now a third group is recognized that comprises species from the northwest of South America and Central America, the so-

called “yellow flowering group” (Moscone et al., 2007; Scaldaferro et al., 2013). *Capsicum* exhibits an exceptional feature shared only with a sister genus *Lycianthes* in Solanaceae. The plants possess an entire cup-shaped calyx with mostly five to ten teeth that function as nerve prolongations (Carrizo García et al., 2016).

An important character unique to species of the genus *Capsicum* is the pungency. The placenta of the fruits produces unique alkaloids: the capsaicinoids (Stewart et al., 2007). In super-hot chile peppers, these alkaloids can also be produced by the pericarp (Bosland et al., 2015). Some of the species, however, belonging to the most basal clade, like *C. geminifolium* or *C. rhomboideum* do not produce any pungent fruits (Stewart et al., 2007). The major gene responsible for pungency is the *Pun1* gene, the expression of which varies dramatically depending on environmental conditions and modifier genes that are epistatically affected by *Pun1* (IBPGR, 1983; Bosland and Votava, 2000; Hunziker, 2001; Stewart et al., 2005). Capsaicinoids are used in the human diet and are believed to have some health benefits. They can be used as an analgesic for arthritis and other types of pain (Fraenkel et al., 2004), they are tested for their ability to inhibit the growth of several forms of cancer (Surh et al., 2002; Mori et al., 2006) and they can promote weight loss (Lejeune et al., 2003; Westerterp-Plantenga et al., 2005; Ludy et al., 2012).

The colour and form of the fruits of the cultivated *Capsicum* taxa vary extremely due to human selection. Wild peppers, on the other hand, usually possess small spherical fruits, although some also develop elliptic or ovoid fruits. The colour of the fruits is mostly yellowish-green or red and in some rare cases orange (Pickersgill et al., 1988).

3.2. Evolution

The approximately 35-40 species of *Capsicum* can be classified into eleven well-supported clades of which four are monotypic. These are the eleven clades of *Capsicum* according to their position in a strict consensus tree (from most basal to most derived): Andean, Caatinga, Flexuosum, Bolivian, Longidentatum, Atlantic Forest, Purple Corolla, Pubescens, Tovarii, Baccatum and Annum clades. The diversification of these clades occurred stepwise and can be reconstructed in clockwise order. Starting from western-north-western South America they might have evolved over a gap in the lowlands of the Amazonas to south-eastern and central

Brazil. Following, the expansion spread back to western and central South America and up towards Central America (Carrizo García et al., 2016).

All species of the genus *Capsicum* are diploid. Two different base chromosome numbers have been documented in the genus, either $x = 12$ or $x = 13$. The chromosome number doesn't vary within a species. In the closely related sister genera and outgroups (*Dunalia brachyacantha, Lycianthes rantonnetii, L. lycioides* and *Saracha punctate*) a base chromosome number of $x = 12$ was observed. Thus, the ancestral base chromosome number of *Capsicum* has been suggested to be $x = 12$. The base chromosome number of $x = 13$ has been inferred to have evolved independently two times, in the Atlantic Forest and Andean clades (Carrizo García et al., 2016).

3.3. Heterochromatin and rDNA

Genomes of *Capsicum* species harbour a wide range of the amounts of constitutive heterochromatin. Heterochromatin content ranges from 1.72% to 38.91% of the haploid karyotype length (mean value of 10.90%). In most taxa, the amount of heterochromatin shows a positive correlation with the size of the genome and karyotype length (Moscone et al., 1996, 2003, 2007; Scaldaferro et al., 2013).

Among all *Capsicum* taxa, four different types of constitutive heterochromatin were observed: AT-rich, moderately GC-rich, highly GC-rich and mixed GC- and AT-rich (Moscone et al., 2007; Scaldaferro et al., 2013). Like in all other plant groups, the NOR-associated heterochromatin of *Capsicum* is GC-rich. Generally, GC-rich heterochromatin is highly abundant in *Capsicum* and occurs in structures like intercalary bands, macrosatellites and a number of distal bands (Scaldaferro et al., 2013).

Constitutive heterochromatin is generally located at subterminal positions of chromosomes in *Capsicum*, but occasionally can also be found in intercalary positions (Moscone et al., 1993, 1996, 2007; Scaldaferro et al., 2013). Taxa with $x = 13$ from eastern coastal Brazil and *C. flexuosum* ($x = 12$) are more likely to possess more intercalary GC-rich constitutive heterochromatin than other taxa with $x = 12$ chromosomes (Scaldaferro et al., 2013).

One to four active NORs at metaphase (haploid set of chromosomes) have been found among twelve wild and cultivated *Capsicum* taxa (Scaldaferro et al., 2016). Most of the NORs were

found on the short chromosomal arms (82.98%). A few species (17.02%) had a NOR located on the long chromosomal arms. The number of 35S rDNA loci on the other hand varied between one and 30 among all the species analysed. In all analysed species only one 5S rDNA locus was found located interstitially within the short arm on a larger metacentric/submetacentric chromosome (Scaldaferro et al., 2016).

3.4. C. *flexuosum*

The Flexuosum clade comprises *C. flexuosum* and *C. aff. flexuosum*, a peculiar sample of a single location. Spotted white corollas with greenish spots in the throat, black seeded small spherical red berries and pendant flowers with toothless calyces are typical characteristics of *C. flexuosum*. *C. aff. flexuosum* on the other hand shows different pigmentation patterns and corolla shapes. *C. flexuosum* grows as shrub as high as two meters (Carrizo García et al., 2013; Scaldaferro et al., 2013).

C. flexuosum is diploid with the base chromosome number of $x = 12$. The karyotype comprises eleven metacentric/submetacentric chromosome pairs of decreasing size and one subtelocentric pair. Satellites (NORs) were found on two chromosome pairs. With 16.82% *C. flexuosum* shows the highest heterochromatin content in all *Capsicum* taxa. The karyotype of *C. flexuosum* also differs in the position of constitutive heterochromatin from other *Capsicum* taxa. In most chile pepper species constitutive heterochromatin is mainly found at terminal positions of chromosomes, whereas in *C. flexuosum* it is mostly localized in intercalary positions. Two satellite regions comprise NOR-associated heterochromatin which spans the entire satellite region and extends to most of the short arms (Scaldaferro et al., 2013).

Like all other *Capsicum* species, *C. flexuosum* has a single 5S rDNA locus. The 5S rRNA genes were mapped on the short chromosomal arm of one of the metacentric/submetacentric chromosome pairs. Mapping of 35S rDNA resulted in the identification of 14 to 15 35S rDNA loci in *C. flexuosum,* of which two were major and twelve to 13 were small. This varying number is likely caused by a polymorphism found on one chromosome pair. In *C. flexuosum,* the 5S rDNA locus was found on the same chromosome as one of the 35S rDNA minor loci (Scaldaferro et al., 2016).

4. Aims of the study

The lack of information about composition, chromosomal location and genomic abundance of various repetitive DNA types in the economically important *Capsicum* genus requires further investigation. With the aid of novel analytic bioinformatics tools and next generation sequencing, it is possible to gain new information about type, distribution, evolution and variability of repetitive DNA elements in the genome of plants.

This study aims to analyse in detail the repetitive DNA fraction of the nuclear genome of *Capsicum flexuosum*, a wild *Capsicum* species that has been suggested to have one of the largest genomes and largest chromosomes of all chile peppers. Specifically, this study aims to: (1) measure genome size of *C. flexuosum*, (2) characterize repeat profiles of its nuclear genome, (3) identify all satellite DNAs and (4) map 5S and 35S rDNA loci, as well as loci of ten abundant satellite DNAs (some shared with other *Capsicum* species), in *C. flexuosum* using fluorescence *in situ* hybridization.

Ultimately, the results presented in this study not only provide novel information about the repeatome of chile peppers, but also provide new chromosomal markers to identify all individual chromosome pairs of *C. flexuosum*. This information will assist further genome-wide analyses of karyotype evolution of all species of the genus *Capsicum*.

Part II – MATERIAL AND METHODS

1. Plant material

Two *Capsicum flexuosum* individuals were used for all experiments. The plants were grown from seeds collected in Misiones, Candelaria Argentina (collectors and vouchers: Davina & Hofi 599). Both plants were cultivated over several years in pots in the Botanical Garden of the University of Vienna.

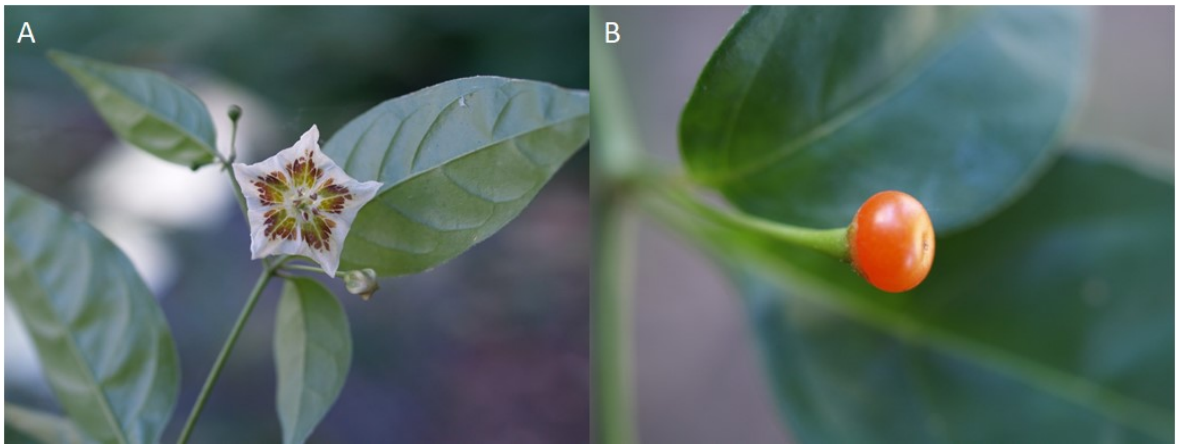

Fig. 1. Flower (A) and fruit (B) of *Capsicum flexuosum* (Photos: Vildana Suljevic).

2. Root pre-treatment and fixation of chromosomes

Plant material used for chromosome preparations, both for classical (Feulgen staining) and molecular cytogenetics (FISH, fluorescence *in situ* hybridization), had to be pre-treated and fixed. One day prior to fixation, the plants were thoroughly watered. The following day, the plants were carefully removed from the soil. Healthy root tips were cut off and placed into a small bottle containing tap water. The roots were carefully rinsed with water to remove any soil particles. Afterwards the roots were incubated in a 0.002 M solution of 8-hydroxyquinoline (Sigma-Aldrich, Vienna, Austria) which resulted in the inhibition of the polymerization of the microtubules of the mitotic spindle and thus arrest of the divisions in metaphase, condensation of the chromosomes and accumulation of metaphases. The incubation was performed in the dark at room temperature for three hours followed by incubation at 4°C for another three hours. The 8-hydroxyquinoline solution was removed and

the roots were washed and fixed in a solution containing ethanol and acetic acid (3:1). The material was subsequently stored in the freezer at -20°C until use.

3. Feulgen staining

The method of Feulgen staining was used for establishing the chromosome number and analyses of basic karyotype structure (Fukui & Nakayama, 1996). Several roots of the previously fixed material were removed from the fixative and incubated in 5N HCl (VWR, Vienna, Austria) in a glass vial for 30 min at room temperature. Afterwards, the HCl was removed with a pipette and the material was washed twice with tap water. The material was then drained of excess water and Schiff's reagent (Sigma-Aldrich, Vienna, Austria) was added. The material was stained in the darkness for 60 minutes. During this step, the free aldehyde groups of the DNA resulting from 5N HCl treatment bound to Schiffs reagent. This colorimetric reaction resulted in magenta-staining of the DNA in the chromosomes. Subsequently, individual root meristems were transferred onto pre-cleaned slides into a drop of 60% acetic acid. The meristems of the roots were dissected under a stereo microscope using entomological needles. Coverslips were carefully applied and the material was squashed. The slides were examined under the Axio-Imager A1 microscope (Carl Zeiss, Vienna, Austria). Complete, well-dispersed mitotic metaphase plates were photographed with the CCD camera and AxioVision software (Carl Zeiss, Vienna, Austria). For the karyotype analysis, the chromosomes were cut out and arranged using Corel Photo-Paint 2017 (Corel Corporation, Ottawa, Canada).

4. Preparation of chromosomes for FISH

The chromosomes of the previously fixed material were prepared following the protocol of Jang and Weiss-Schneeweiss (2015), with some modifications. The material was removed from the fixative and 5-10 roots were selected for further treatment. The material was incubated in a citric buffer pH 4.8 for 15 min. Subsequently, the material was incubated in pre-warmed enzyme mix (0.4% pectolyase (Sigma-Aldrich, Vienna, Austria), 0.4% cytohelicase (Sigma-Aldrich, Vienna, Austria), 1% cellulose (Serva, Heidelberg, Germany) in

citric buffer pH 4.8 at 37°C for 27 min to digest the cell walls. The enzyme mix was carefully removed and the material was washed with citric buffer for 10 minutes at room temperature. The citric buffer was then removed and replaced with 60% acetic acid. The meristems were transferred to a pre-cleaned slide into a drop of 60% acetic acid and dissected under a stereo microscope. Coverslip was carefully applied to the slide and the material was gently squashed to ensure the even distribution of the material. The quality of the slides was checked under a phase-contrast AxioPlan microscope (Carl Zeiss, Vienna Austria). The slides with well spread and complete metaphase plates were frozen at -80°C for at least 20 min. Afterwards, the coverslips were removed, the slides were dried at room temperature and transferred to 4°C where they were stored until use and for maximally a few days.

5. Fluorescence *in situ* hybridization

The chromosomes on the slides were refixed in a solution of ethanol and acetic acid (3:1) for 20 min, dehydrated for 30 min in 96% ethanol and air-dried at room temperature for 45 min. The slides were then washed three times for 3 min each in 2xSSC buffer. To remove the cytoplasm which could be covering the chromosomes, the slides were treated with 0.01% pepsin solution (pH2, Sigma-Aldrich, Vienna, Austria) solution at 37°C for 20 min. Afterwards, the slides were washed three times for 3 min in 2xSSC. The slides were then incubated in a Coplin jar in 4% paraformaldehyde (VWR, Vienna, Austria) for 5 min to re-fix the histones. The slides were then washed three more times, 3 min each change, with 2xSSC and dehydrated in 70% ice-cold ethanol and 96% ice-cold ethanol for 3 min each. The slides were then air-dried at room temperature for at least 40 min.

Hybridization mix containing 10% dextran sulphate (Promega, Madison, USA), 0.02xSSC, blocking SS (salmon sperm) DNA (1 ng/µl; Sigma-Aldrich, Vienna, Austria) and HPLC water was prepared. To each of the aliquots of 45 µl hybridization mix 2.5 µl of a probe labelled with biotin and 2.5 µl of a probe labelled with digoxygenin were added (10-100 ng of each labelled probe).

Probes used for FISH were as follows: (1) ten satellite DNAs (Table 1) identified in comparative RepeatExplorer analyses of 15 *Capsicum* species representing all major clades of the genus (see below; Weiss-Schneeweiss, unpublished), (2) 35S rDNA (18S/25S rDNA) from *Arabidopsis thaliana* in plasmid pSK+ and (3) conserved partial genic region of 5S

rDNA (Weiss-Schneeweiss et al., 2012). The 35S rDNA probes (18S or 25S rDNA) were labelled with biotin or digoxygenin using nick-translation kit (Roche, Vienna, Austria) according to manufacturer instructions. All satellite DNA oligonucleotide probes and the 5S rDNA probe were commercially labelled via end-labelling using either biotin or digoxygenin (IBA, Germany).

The hybridization mixes, each containing two differently labelled probes, were denatured in a thermoblock (Eppendorf, Vienna, Austria) at 98°C for 5 min and immediately incubated on ice for 5 min. 8 µl of the denatured hybridization mix were applied to each of the individual slides, covered with a 20x20 coverslip and sealed with Fixogum. The slides were denatured on a PCR *in situ* block (Peqlab GmbH, Erlangen, Germany) using the following cycling profile: 74°C for 4 min, followed by 1 min at 65°C, 1 min at 55°C, 1 min at 45°C and 1 min at 37°C. The slides were then transferred to an incubator and left to hybridize at 37°C overnight.

On the next day, the Fixogum was removed and the slides were placed in a Coplin jar containing 2xSSC where the coverslips were carefully removed. The slides were then washed in two changes of 2xSSC at 39°C for 5 min each and finally washed in 2xSSC-Tween20 (Sigma-Aldrich, Vienna, Austria). Blocking solution, containing bovine serum albumin (BSA; Sigma-Aldrich, Vienna, Austria) in 2xSSC-Tween20 was prepared and applied to each washed slide that were covered with foil and incubated in a wet box for 20 min at 37°C. The detection solution containing Streptavidin-Cy3 (Sigma-Aldrich, Vienna, Austria) for biotin detection and Anti-digoxygenin-FITC (Roche, Vienna, Austria) for digoxygenin detection in blocking solution was applied to each slide. The slides were covered with foil and incubated at 37°C for 60 min. The foil was then removed and the slides were washed two times in 2xSSC for 5 min each at 39°C and 5 min in 2xSSC-Tween20 at 39°C. Finally, the access of liquid was drained from the slides. A drop of antifade mounting medium Vectashield containing DAPI (4',6-diamidino-2-phenylindole, 2 ng/µl; Vector Laboratories, Burlingame, CA, USA) was applied to each slide, covered with a 24x24 coverslip and sealed with Fixogum. Slides were stored at 4°C overnight to stabilize the fluorescence. Preparations were examined under an AxioImager M2 epifluorescent microscope (Carl Zeiss, Vienna, Austria). Well spread, complete mitotic metaphase plates were photographed using a CCD camera and AxioVision 4.8 software (Carl Zeiss, Vienna, Austria). The images were contrasted in Adobe Photoshop and the chromosomes were cut out and arranged using Corel Photo-Paint 2017.

Table 1. Satellite DNA probes used for FISH (Weiss-Schneeweiss, unpubl.).

Satellite DNA name	Consensus monomer length	Consensus monomer sequence
CapSat1	37bp	ACAGAAAGTAAAATCCCATTTTTCAGGAGGGTCCTGA
CapSat3	44bp	TTAAGTGTAGTTACTTGAAATGTGAATTAGTGACTTGAAAATGG
CapSat5	75bp	TTGAATGTGGTTTGGTTGGTGACTTAGTTACTTGAAAATGGTTAAGTGTAGGGACTTTAAAGGGGATTTTTTGTT
CapSat6	41bp	TCGAGTCATTCCCTTTGATTCCTATAAGACGTACCTTTTAG
CapSat8	183bp	TGGCCTCATCCACACAAACGGCCCACACTGGGCCGCTCCCAAGCCTCCCTGGCACGCCTCGGTCGATTTTGGGCCCAAAAAATGCCCGAGCCTCGGGCCCACCCCCGGGCTCATAGGCGGCCGTACACGAGCCTGGCCTGAAAAGGGCCAGTTCGCCCCGTTTCGCCCGTGGGAACACCTAAT
CapSat10	136bp	ATGAGACATACATAGGTTGGATGGACCAATAATAGAACTTGGTTTCTTACATTACTACACGTGTATTACCTCCTAACATCTATGTATTCCATACTCGGATGGATTATCGGTGTTTCTTTGACTACTAAATCACTAA
CapSat11	183bp	CCCACGGAAAATCGCCAAAAAACTGCCGTTTTCACTCTTTCGCCCATGGGACGCCCTCAAAACCTCCCACCGACCCCCCGGAGCCCATGGATCCGCTTAACAGCTCGTCAGGGACGCCCCCGAGAGGCTCGGCCCAAAACCATTTCCGGAGCCAGCACCACCCCATCGATCCGTGGGCTAACA
CapSat12	76bp	AGGGACTTTAAAGGTGTTTTTTTGTTTTGAAAGTGGTTATGTTTGGTGACTTGGTAGGTTGAAAATGATTAAGTGT
CapSat13	227bp	AGTGTGTCATAATCCAGTTTAAGCTTGTCACCAATGTCTGTCTCTCCACTTCATTCTGCAACTCCTCCACTCTTAGTCTTAATTCTTCTATTGATTTCTGAGCACTTTCAGATCCTGCAATCCGCTGGTTCAGATGATTCTCCATTTCTTGGATTTGGATGTTCCGGGACATCAGTTCTGCACCAAGTGTCTGTACCTTAGCGTGAAGCTCATCTTTATCCGCCATT
CapSat15	44bp	AAGTCACTAATTCACATTTCAAGTGACTACACTTAACCATTTTC

6. Flow cytometry

To measure the size of the genome of *Capsicum flexuosum* the nuclei were isolated and stained with propidium iodide following the protocols of Temsch (2003) and Baranyi et al. (1996). Approximately 25 mg of young leaves were harvested from each plant. *Solanum pseudocapsicum* was used as an internal standard (1C = 1.2946 pg; Temsch et al., 2010). The leaves of both standard organism and *C. flexuosum* were co-chopped in 1 ml of extraction buffer (pH 1.5) using a razor blade in a plastic Petri dish. As an independent control, the leaf of the standard organism alone was chopped in an isolation buffer (Merck, Darmstadt, Germany) to enable the identification of the G_1 peak (Baranyi et al., 1996; Temsch, 2003). The chopped material was rinsed in another change of isolation buffer and filtered through a 30 µm nylon mesh (Saatilene Hitech, Sericol, Germany) to remove any larger material particles which could hamper the measurements. The filtrate was incubated in RNaseA solution (0.15 mg/ml; Sigma-Aldrich, Vienna, Austria) at 37°C in a water bath for 30 min to remove RNAs. The nuclei were incubated in buffer containing propidium iodide (1 mg/ml; Sigma-Aldrich, Vienna, Austria) for one hour at 4°C to stain the DNA.

After the staining was completed, each sample was loaded to a CyFlow Space Cytometer (Sysmex Austria GmbH, Vienna, Austria) equipped with a green laser (5432 nm, 100 mW, Cobolt Samba, Cobolt AB, Stockholm, Sweden). The fluorescence emitted by each particle in the tube was detected and recorded by appropriate software. The results were visualised as a histogram using FloMax Software (Partec, Münster, Germany). In total, 3300 particles were measured per sample. The Gaussian model was applied to calculate the mean fluorescence intensity of each sample. The known genome size of the internal standard organism allowed to calculate the size of the genome of the samples using the following formula:

$$1\text{C value (pg)} = \frac{\text{mean } G_1 \text{ peak position of sample}}{\text{mean } G_1 \text{ peak position of standard}} * \text{standard C1 value}$$

7. Next generation sequencing and repeat analyses

The genomic DNA of a *Capsicum flexuosum* was isolated following the CTAB protocol (Jang & Weiss-Schneeweiss, 2015). The quality of the extracted DNA was checked on an agarose gel and the concentration of DNA was measured using a NanoDrop–fluorospectrophotometer and Quant-iT Picogreen dsDNA assay kit (PeqLab, Austria). Genomic DNA was fragmented into 600-800 bp long fragments and the library was constructed at CSF-NGS (Vienna Biocenter, Austria). The DNA was then sequenced using 150 bp paired-end Illumina HiSeq2500 technology (Illumina, San Diego, CA, USA). Two independently prepared libraries (technical replicates) were sequenced for *C. flexuosum*.

The reads were processed and analysed using the RepeatExplorer pipeline (Novák et al., 2013, 2020; http://repeatexplorer.org). The reads were checked for quality, and the adaptors, PhiX spike-in DNA (Illumina), and organellar DNAs were removed. The remaining good-quality interlaced read pairs were clustered using default settings in RepeatExplorer (Novák et al., 2013; http://repeatexplorer.org). The reads of the two libraries were pooled for analysis. Clustering was based on an all-to-all comparison of reads. The resulting clusters and superclusters represented all types of repeats, comprising sequences that were very similar and thus highly connected (Novák et al., 2010).

The automatic annotation of the clusters provided by RepeatExplorer was subsequently manually verified. The resulting data allowed for the characterization of the total repeat content and proportions of *C. flexuosum* genome. It also allowed for the identification of tandem repeats (satellite DNAs) through the use of TAREAN (included in RepeatExplorer pipeline) which were manually verified via Dotter (Sonnhammer & Durbin, 1995).

Part III – RESULTS

1. Chromosome number and karyotype

Feulgen staining was performed to establish the exact chromosome number and karyotype structure of analysed *C. flexuosum* plants. Two well-spread mitotic metaphase plates of *C. flexuosum* were selected. The cut-out karyotypes and the corresponding ideograms are presented in Fig. 2 and Fig. 3, with the black dots indicating satellites.

Both analysed individuals had chromosome number of $2n = 2x = 24$ (Fig. 2A; Fig. 3A). The diploid karyotype consisted of eleven metacentric/submetacentric chromosome pairs and one subtelocentric chromosome pair. Two of the chromosome pairs (one metacentric and one submetacentric) carried satellites (NORs; Fig. 2B; Fig. 3B). The total length of the haploid chromosome set was 65.06 µm (SD ± 2.2).

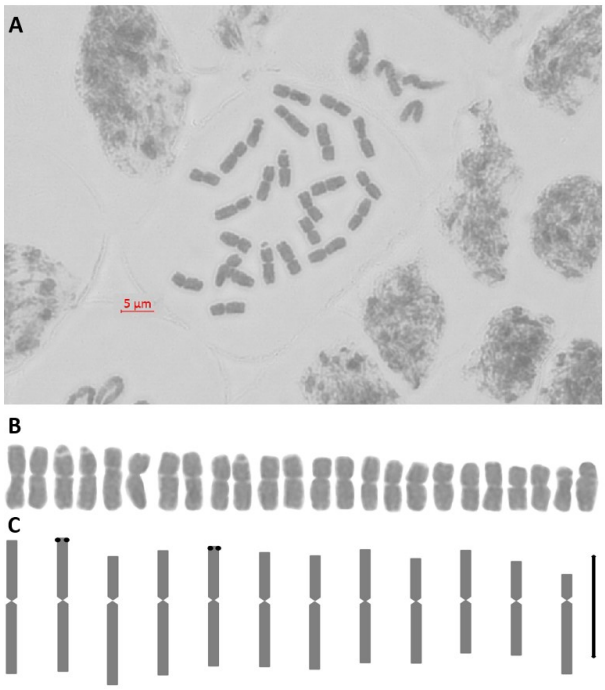

Fig. 2. Metaphase plate with $2n = 2x = 24$ chromosomes (A), karyotype (B) and ideogram with satellites (NORs) indicated as black dots (C) of *C. flexuosum* individual 1a; scale bar, 5µm.

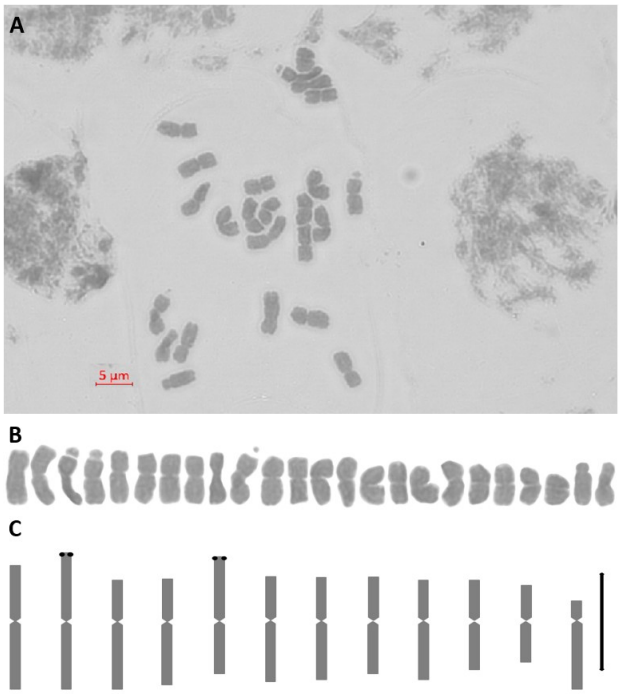

Fig. 3. Metaphase plate with $2n = 2x = 24$ chromosomes (A), karyotype (B) and ideogram with satellites (NORs) indicated as black dots (C) of *C. flexuosum* individual 1b; scale bar, 5µm.

2. Genome size

The genome size of *C. flexuosum* was measured in two replicates using flow cytometry. The G_1 value of *C. flexuosum* was calculated using the G_1 value of the standard organism *Solanum pseudocapsicum*.

The 1C value (DNA content of unreplicated haploid nuclei) of the two measurements was 7.51 pg and 7.31 pg. Thus, the mean 1C value of both measurements was 7.41 pg (SD ± 0.1). The data output of the FloMax programme for one of the measurements is shown in Fig. 4. In the lower part of the figure the settings used for the measurement are shown. Fig. 4A depicts the resulting histogram with three peaks representing the intensity of the measured fluorescence. The first peak (G_1 of *Solanum pseudocapsicum*; abbreviated as SP) was manually set to 104.81. The next peak represents replicated G_2 nuclei of *Solanum pseudocapsicum* at 211.85 and the third peak shows the G_1 nuclei of *C. flexuosum* at 608.31.

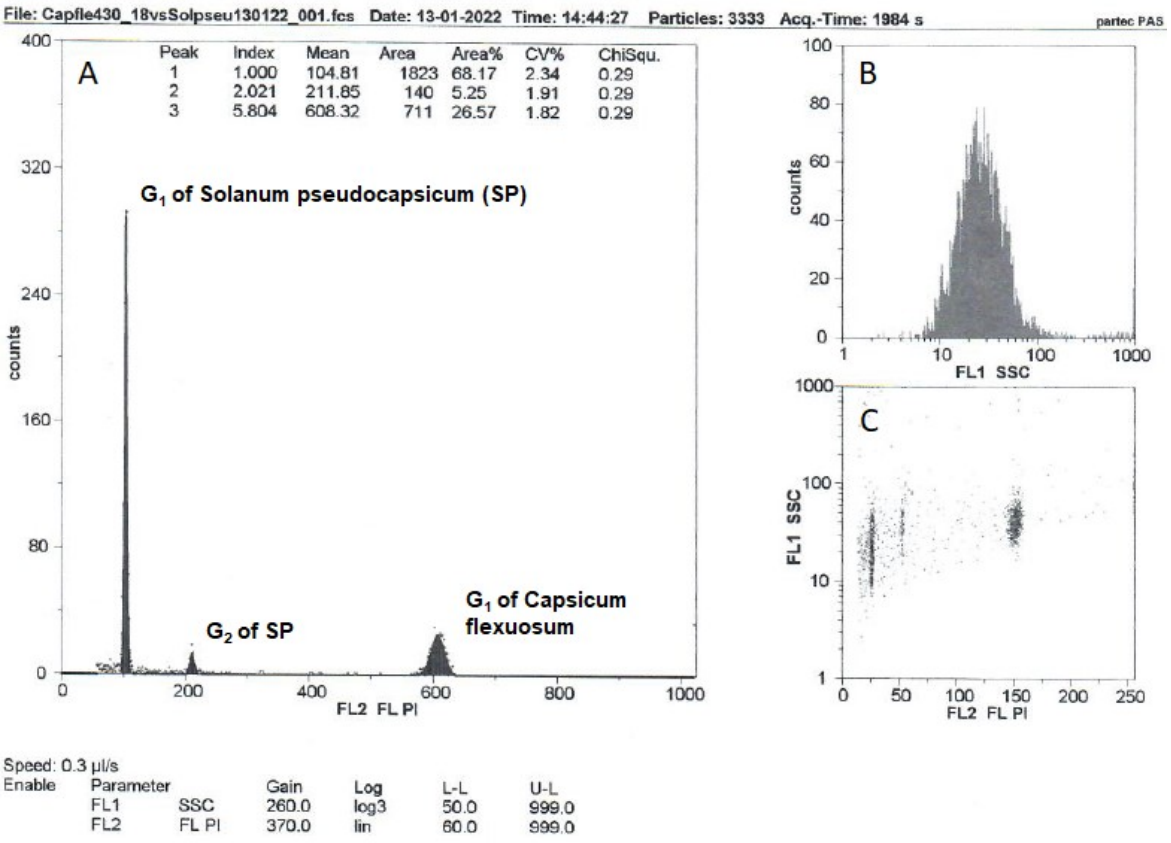

Fig. 4. Example of flow cytometry histogram used for the calculation of genome size. Histogram of DNA content of *C. flexuosum* measured together with internal standard organism of *Solanum pseudocapsicum* (A), side scatter histogram (B) and cytogram fluorescence (FL1) vs side scatter (C).

3. Repeatome profile analyses of *C. flexuosum*

RepeatExplorer pipeline (Novák et al., 2020) was used to analyse the repeat profiles of the genome of *Capsicum flexuosum*. The genome proportions of all types of repeats identified in *C. flexuosum* genome are shown in Table 2 and Fig. 5. In total, repeats were found to represent 64.73% of the analysed genome.

Table 2. Estimates of genome proportions (%) of various repeat types identified in *C. flexuosum.*

Type	Superfamily	Family	Percentage	No. of reads
Retrotransposons			**49.12**	**2055664**
	Ty3-*gypsy*		33.19	1388848
		Tekay	27.17	1136804
		TAT/Ogre	3.02	126217
		Athila	1.75	73122
		Other	0.98	41103
		CRM	0.20	8502
		Galadriel	0.07	3100
	Ty1-*copia*		2.61	109373
		TAR	0.96	40361
		Tork	0.59	24809
		Bianca	0.51	21304
		SIRE	0.25	10521
		Ikeros	0.13	5559
		Ale	0.07	3101
		Angela	0.07	2953
		Ivana	0.02	765
	other/non-LTR		13.32	557443
		Other LTR	12.30	514922
		Pararetrovirus	0.49	20587
		LINE	0.41	17279
		Non-LTR	0.11	4655
DNA transposons			**0.95**	**39719**
		MuDR	0.47	19636
		hAT	0.18	7489
		CACTA	0.16	6627
		Other	0.14	5967
Tandem repeats			**10.51**	**439872**
	Satellite DNAs		6.69	279790
	rDNA		3.83	160082
		35S	3.80	159110
		5S	0.02	972
Unclassified			**4.14**	**173346**
Total repeats			**64.73**	**2708601**
Low-copy			**35.27**	**1476100**

Class I dispersed repeats, represented mostly by retrotransposons, were the most abundant repeats in *C. flexuosum* genome representing 49.12% (Table 2). Within retrotransposons, LTR-retroelements (LTR-long terminal repeat-) made up 48.10% of the genome, followed by pararetroviruses (0.49%), non-LTR elements represented by LINEs (long interspersed nuclear elements) with 0.41% and other, unclassified LTR elements with 0.11%. Over all repeat groups, the copy number of Ty3-*gypsy* LTR elements was the highest with 33.19%. An example of a cluster graph representing Ty3-*gypsy* TAT/Ogre element as reconstructed by RepeatExplorer with all protein domains is shown in Fig. 6. The overwhelming majority of the Ty3-*gypsy* elements belonged to the Tekay family (27.17%). Ty1-*copia* retroelements were in minority representing only 2.61% of the genome. An example of a cluster graph representing Ty3-*copia* Tork element as reconstructed by RepeatExplorer with all protein domains is shown in Fig. 7. Class II dispersed repeats represented by DNA transposons made up only a very small portion of *C. flexuosum* genome (0.95%), with the Mutator element being most abundant (0.47%).

Tandem repeats were represented in *C. flexuosum* genome by rDNAs and satellite DNAs. Satellite DNAs represented 6.69% of the genome, whereas 3.83% of the repeats were identified as rDNAs. The majority of the rDNAs represented 35S rDNA (3.80%). This rDNA type was reconstructed as supercluster (SCL5) consisting of eleven individual clusters (Fig. 8). Each cluster consisted of reads representing only part of the rDNA consensus monomer. The copy number of 5S rDNA monomers was much lower in the genome of *C. flexuosum* (0.02%) and it was reconstructed as single cluster (Fig. 9). Unclassified repeats represented 4.14% of the genome and 35.27% of the reads were classified as low-copy sequences (non-repeats).

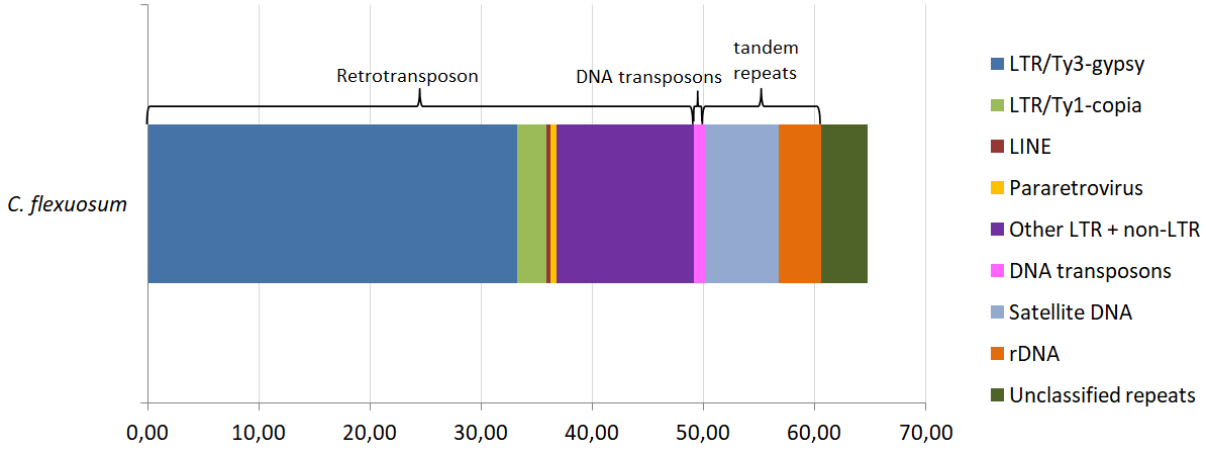

Fig. 5. Proportions (in %) of repetitive DNA elements in the genome of *C. flexuosum*.

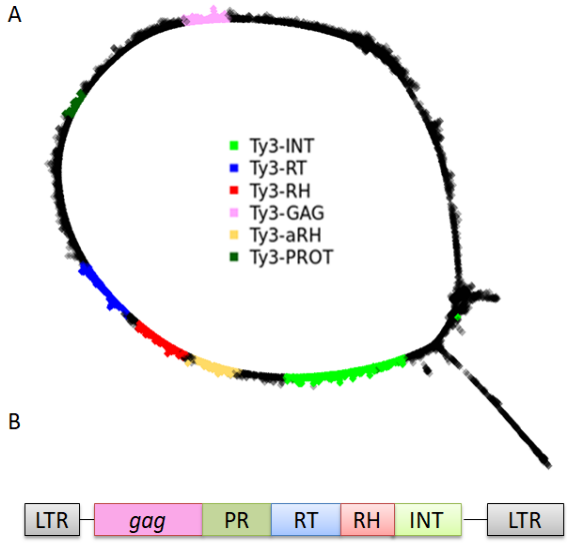

Fig. 6. Graph of cluster representing Ty3-*gypsy* TAT/Ogre element (cluster CL97, A) and the structure of the element with all protein domains (B).

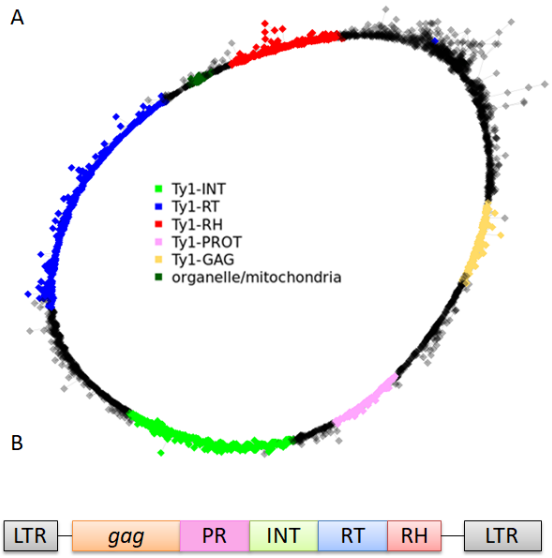

Fig. 7. Graph of cluster representing Ty1-*copia* Tork element (CL137; A) and the structure of the element with all protein domains (B).

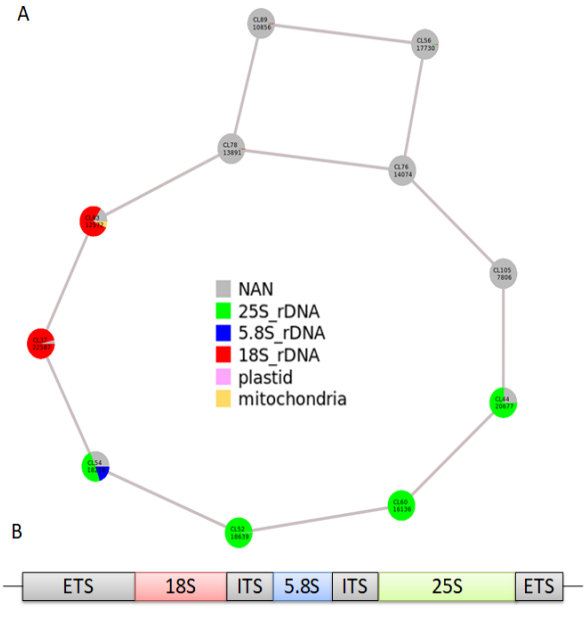

Fig. 8. Supercluster representing 35S rDNA monomer (A) and schematic representation of the structure of the monomer with all rDNA coding domains and spacers (B).

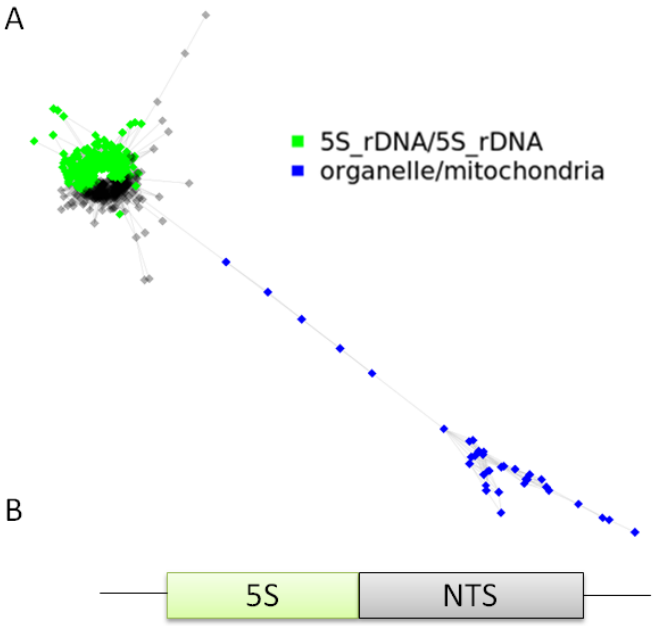

Fig. 9. Graph of cluster representing 5S rDNA monomer (CL206; A) and the structure of the element with gene and spacer (B).

Putative satellite DNAs identified by RepeatExplorer were verified using dot plot (dotter) where the contig sequences of clusters representing putative satellite DNAs were plotted against themselves to find the strings of tandem repeats. Seven satellite DNA families were identified in *Capsicum flexuosum* (Fig. 10). All seven satellite DNAs newly identified in *C. flexuosum* were also plotted against the satellite DNAs identified previously in comparative genus-wide repeat analyses of *Capsicum* (Weiss-Schneeweiss, unpublished; Fig. 11). Five of the seven satellite DNAs identified in *Capsicum flexuosum* represented satellite DNAs previously reported in other species of the genus *Capsicum* (CapSat1, CapSat5, CapSat8a, CapSat8b, CapSat14; Fig. 11). Two remaining identified satellite DNAs were novel to the genome of *C. flexuosum* and are referred to as CapSat17 and CapSat18 (Table 3). These two satellite DNAs represented only a very small proportion of the genome and need to be verified by FISH as tandem repeats. Monomer sequences, graph layouts and monomer sequence logos of all seven satellite DNAs identified in *C. flexuosum* are shown in Table 3 and Figs. 12-18. The logo of satellite DNA CapSat8b has not been generated as RepeatExplorer did not recognize this cluster as putative satellite DNA.

Table 3. Characterization of seven satellite DNAs identified by RepeatExplorer in *C. flexuosum* genome. Two novel satellite DNAs identified in *C. flexuosum* are indicated in red.

Satellite DNA name	Cluster	Consensus monomer length	Genome proportion	Consensus monomer sequence
CapSat8a	CL1	183bp	5.56%	TCCCTGGCATGCCTCGGTCGATTTTGGGCCCAAAA AATGCCCGAGCCTCGGGCCCACCCCCGGGCTCAT AGGCGGCCGTACACGAGCCTGGCCTGAAAAGGG CCAGTTCGCCCCGTTTCGCCCGTGGGAACACCTAA TTGGCCTCATCCACACAAACGGCCCACACTGGGCC ACTCCCAAGCC
CapSat17	**CL76**	**24bp**	**0.34%**	**GCACCATGGCACCGTGGCACCAAG**
CapSat8b	CL92	183bp	0.25%	TCCCTGGCATGCCCCGGTCGATTTTTGGCCCACAG CACGCCCGGACCCCGGGCCCACCCCAGAAACCGT GGGCTATAGCACACCGATTTGGCCCGAAAATGCC CCGTTCGCGCCGTTTCGCCCGTTTGTCCACCCAAA TGGCCACAAAAATTTAAACGGACCACACTGGGAC TATCCCCAACC
CapSat5	CL98	76bp	0.21%	CAAAGAATTTCCCTTTAAAGTCCCTAAACTTAACC ATTTTCAACAAACTAAGTCACCTAACAAACCACAT TCAAAA
CapSat14	CL121	76bp	0.14%	ATTTGCTTGAAAATGCTTAAGTGTTGGTGACTTTA AAAGTGAATTAGTGTTTAGAAGGTGATGTTGTTG GTGACTT
CapSat1	CL133	37bp	0.10%	GAAAAATGGGATTTTACTTTAAGTTCAGGACCCTC CT
CapSat18	**CL149**	**173bp**	**0.08%**	**GGAACGCAATACCGCCTCATCATGCAATTGCCAA AAGACTCAATTGAAGTTGTAGTTGACCCTATGAG AACTCACCCCTAGTCCTAGATTATTCAATTAAGGT ATTAGTCTAACATTTGAGGTAGTAAGAATCGGTT CGGCTCAGAGTGATCGATCGACGTCTCTGGTCGG GA**

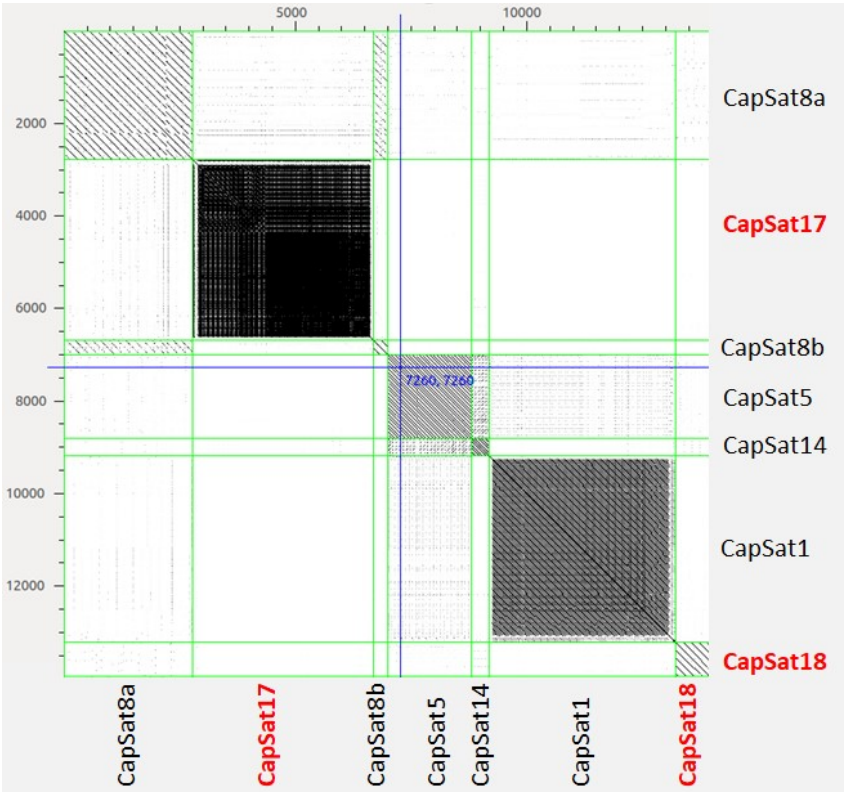

Fig. 10. Dot plot of satellite DNAs identified in the genome of *Capsicum flexuosum*. Novel satellite DNAs are indicated in red.

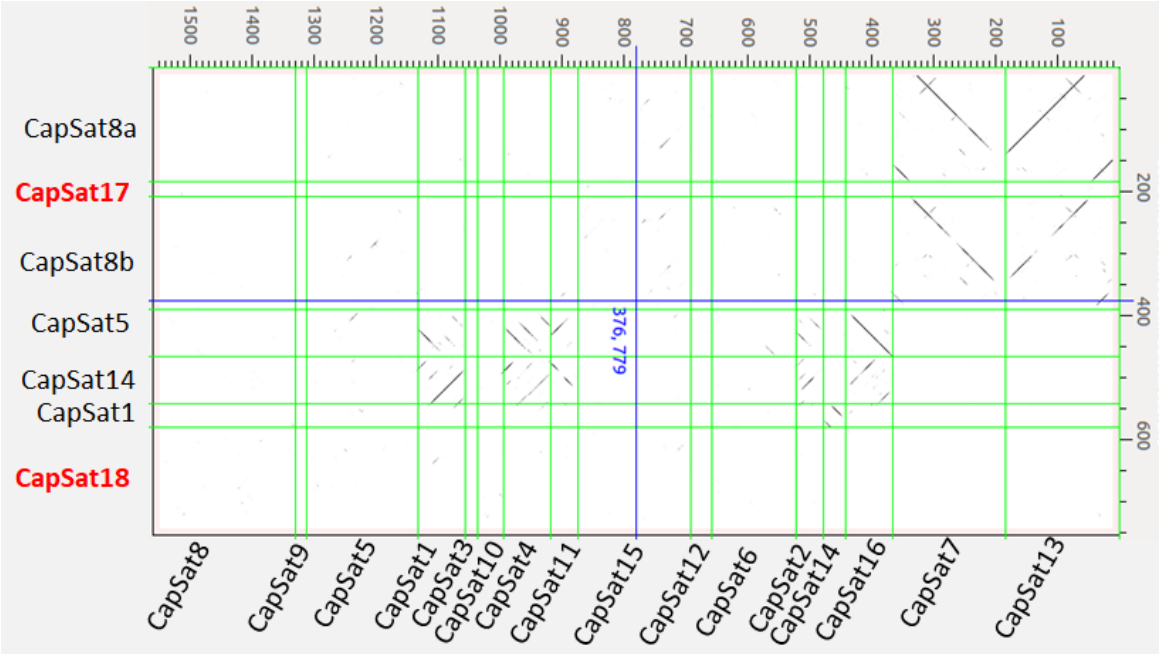

Fig. 11. Dot plot of satellite DNAs monomers identified in the genome of *Capsicum flexuosum* (y-axis) against satellite DNAs previously identified in other *Capsicum* species (Weiss-Schneeweiss, unpubl.; x-axis) and used for FISH. Novel satellite DNAs are indicated in red.

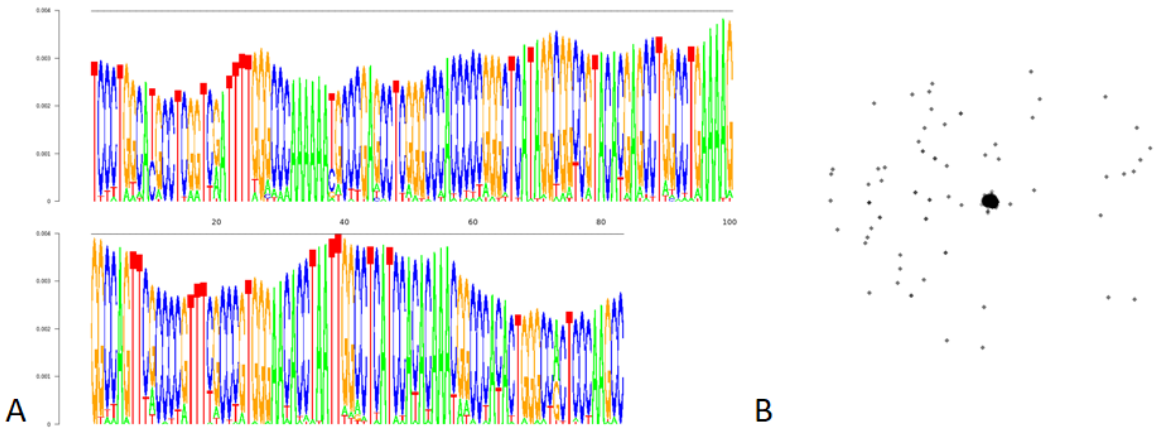

Fig. 12. Monomer sequence logo (A) and graph layout (B) of CapSat8a identified by RepeatExplorer in *C. flexuosum* genome.

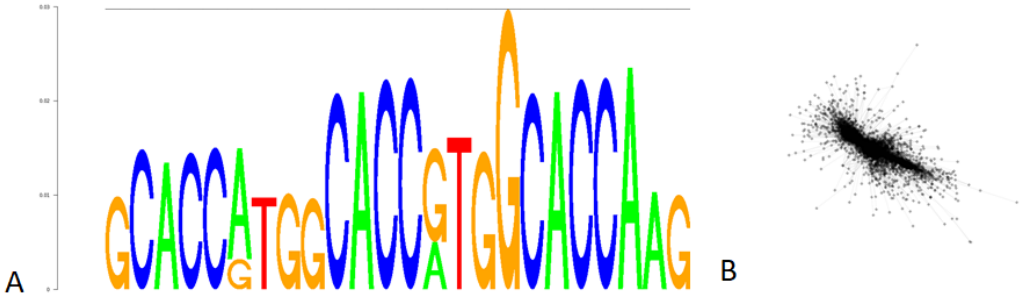

Fig. 13. Monomer sequence logo (A) and graph layout (B) of CapSat17 identified by RepeatExplorer in *C. flexuosum* genome.

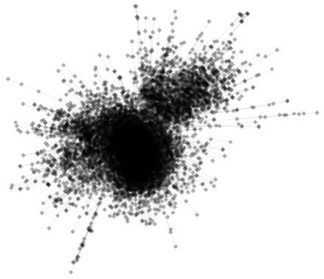

Fig. 14. Graph layout of CapSat8b identified by RepeatExplorer in *C. flexuosum* genome.

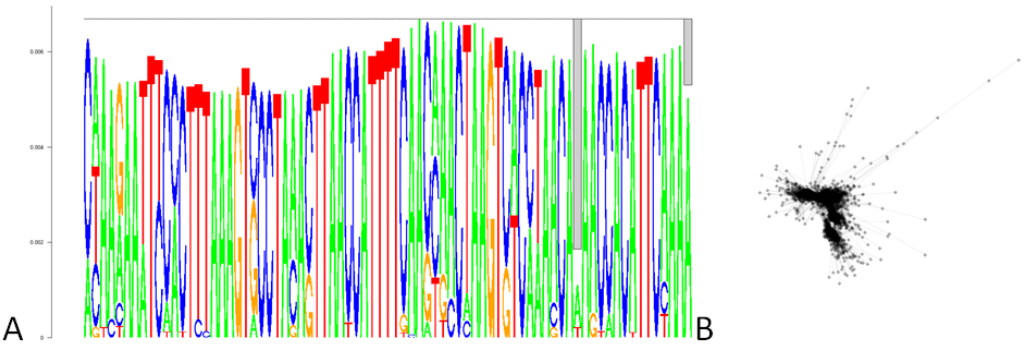

Fig. 15. Monomer sequence logo (A) and graph layout (B) of CapSat5 identified by RepeatExplorer in *C. flexuosum* genome.

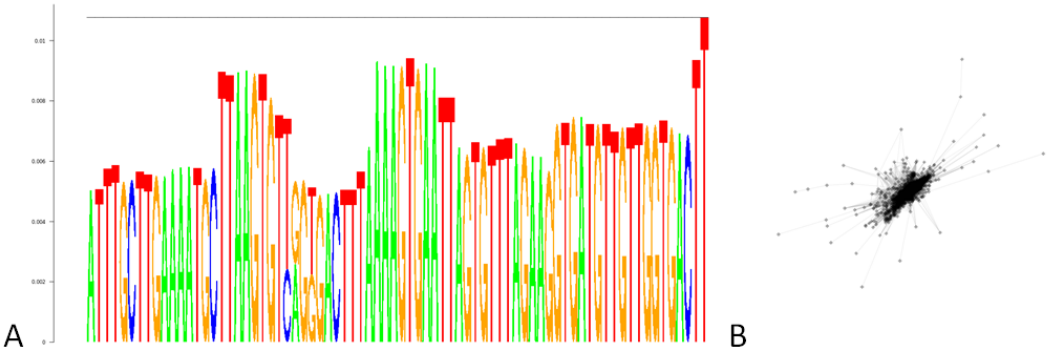

Fig. 16. Monomer sequence logo (A) and graph layout (B) of CapSat14 identified by RepeatExplorer in *C. flexuosum* genome.

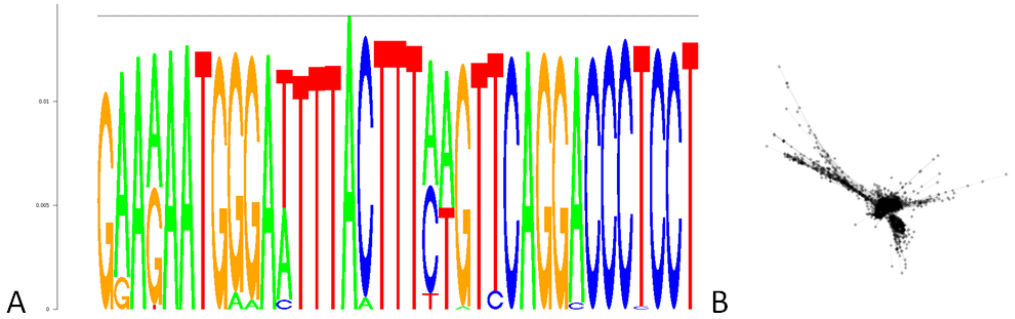

Fig. 17. Monomer sequence logo (A) and graph layout (B) of CapSat1 identified by RepeatExplorer in *C. flexuosum* genome.

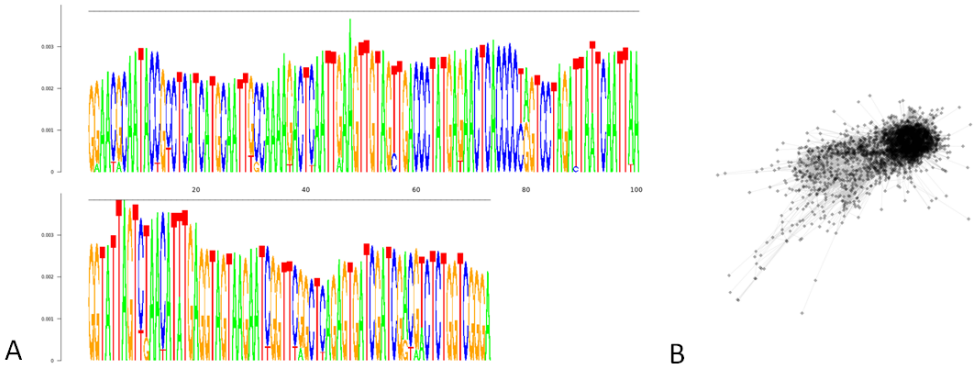

Fig. 18. Monomer sequence logo (A) and graph layout (B) of CapSat18 identified by RepeatExplorer in *C. flexuosum* genome.

4. Localization of tandem repeats in *C. flexuosum* chromosomes

Ten satellite DNAs (CapSat1, CapSat3, CapSat5, CapSat6, CapSat8, CapSat10, CapSat11, CapSat12, CapSat13 and CapSat15) and two rDNA probes (5S and 35S rDNA) were mapped in chromosomes of C. *flexuosum* using fluorescence *in situ* hybridization (FISH). The karyotypes were cut-out and chromosome pairs were always arranged according to their size (largest to smallest). The order of chromosome pairs was also assisted by presence of heterochromatic blocks and NOR (where visible). In order to map all the probes and be able to compare their localization in chromosomes nine different combinations of the satellite DNAs and rDNA probes were used for two-colour FISH (Fig. 19-27). Always two probes were mapped simultaneously for each individual preparation. Most of the probes were used in combination with at least two other probes (used on different preparations) which allowed for the correlation of localization of all probes relative to one another (e.g., probe CapSat8 was mapped with CapSat5, but also with CapSat1; CapSat1 was also mapped with CapSat13, etc.). All cut-out karyotypes were then arranged in the same way relative to one another taking into consideration marker loci of all satellite DNAs mapped in various combinations. Reprobing of the preparations was attempted but the quality of the chromatin deteriorated quickly and resulted in suboptimal quality of reprobed chromosomes.

The karyotype of *C. flexuosum* consists of several pairs of chromosomes that are very similar and thus difficult to be distinguished from one another based on chromosome morphology alone. Therefore, all the signals of tandem repeats mapped by FISH were plotted onto simplified ideograms where all chromosomes (of a haploid complement) were represented by identical-sized bars, regardless of the real size of the corresponding chromosomes (shown in each figure as a cut-out). Centromeres were also not indicated (and thus chromosomal arms were not marked) due to difficulties in establishing the exact position of the centromeres in FISH-preparations and thus inability to precisely determine the length of chromosomal arms. The strategy used here is therefore very conservative but results in less speculative data interpretation.

Loci were classified by their size (i.e., the size of chromosomal segments carrying the signal) as major (large blocks; indicated by rectangles on ideograms) and minor loci (small signals, indicated as filled circles on ideograms). Additionally, when mapping individual satellite DNAs, the signal strength of some of the loci differed between chromosomal spreads (even within the same preparation), most often between the two homologous chromosomes of the

pair, a phenomenon known to commonly occur during FISH experiments due to technical issues, e.g., accessibility of the chromosomal target DNA (FISH is semi-quantitative method). Thus, the loci for which the signals detected varied in strength or presence among homologous chromosomes of a cell, among preparations and/or among individuals were marked in the ideograms with an asterisk (Figs. 19-28). In the final step a combined karyotype of *C. flexuosum* has been created where all probes are mapped in all 12 chromosome pairs (Fig. 29) and the data were summarized in Table 4.

rDNA localization

Only one 5S rDNA locus was detected at a subterminal position of the short arm of chromosome 5 (all chromosome numbers refer to simplified ideogram; Fig. 19). In contrast, a very high number of 35S rDNA loci (22 in total: 16 major and 6 minor) was detected in *C. flexuosum*; of those, 14 loci were detected in all plates and eight only in some (Fig. 19, 20, 23). Chromosomes 1 and 4 carried one major locus on the long arm and one minor locus on the short chromosomal arm. Chromosomes 2, 3, 5, 6 and 9 had two major loci, one in each of the two chromosomal arms. 35SrDNA signals in the short arms of chromosomes 2 and 5 co-localized with the secondary constrictions (NORs). Two minor loci were found on both chromosomal arms of chromosome 7. Chromosome 8 had one major locus on the short arm and two major loci on the long arm. Chromosomes 10 and 11 carried each one minor locus on the short arm. Chromosome 12 carried one major locus on the long chromosomal arm.

Satellite DNAs localization

Three minor CapSat1 loci were found in three chromosome pairs, all in subterminal position within the long arms. The loci on chromosomes 4 and 7 were detected in most, but not all plates/homologous chromosomes of a pair whereas the locus on chromosome 8 was detected consistently in all spreads (Fig. 21, 26).

Four minor CapSat3 loci were detected. Loci in chromosomes 3 and 9 were localized within the short chromosomal arms. Two loci were detected on chromosome 4, one on each of the chromosomal arms (Fig. 22).

CapSat5 was present in 18 loci in *C. flexuosum* chromosomes: 14 minor and four major ones. Chromosomes 1, 7, 8 and 9 each had two minor loci, one on each chromosomal arm. Chromosomes 10 and 11 each possessed one minor locus within the short arm, whereas

chromosome 5 had one minor locus on its long arm. Chromosome 2 had two minor loci on the short arm. One major locus of CapSat5 was found on short arms of chromosomes 6 and 12. Chromosome 3 carried one major locus on the long arm. One major locus on the short arm and one minor locus on the long arm were detected on chromosome 4 (Fig. 23, 24).

CapSat6 was detected in 19 loci, twelve of which were minor and seven were major. All analysed chromosomal spreads carried 13 of the loci, whereas six loci were only found in some of the spreads. Three minor loci were detected on chromosome 1, one on the long arm and two on the short arm. Chromosomes 2, 6, 7 and 8 carried one minor locus each, always on the short arm. Chromosome 3 and 5 each had two major loci, one on each chromosomal arm. One minor locus (short arm) and one major locus (long arm) were found on chromosome 4. Chromosomes 9 and 10 both carried two minor loci: chromosome 9 had one locus on each of the two arms whereas chromosome 10 had both loci on the short arm. Chromosomes 11 and 12 had one major locus each, chromosome 11 within the long arm and chromosome 12 on the short arm (Fig. 27).

Eight CapSat8 loci were detected in total. All loci were major and located on the long chromosomal arms. One locus was detected in all plates on each of the chromosomes 1, 3, 4, 6, 8, 9 and 12 whereas the locus on chromosome 5 was detected only in some plates/homologous chromosomes of the pair (Fig. 20, 24, 26).

Eight CapSat10 loci were found in *C. flexuosum* chromosomes. All loci were major and located on the long chromosomal arms. They were found on chromosomes 1, 3, 4, 5, 6, 8, 9 and 12 (Fig. 22).

Eight loci of CapSat11 were detected in the analysed species. All loci were major and located on the long chromosomal arms. The loci on chromosomes 1, 3, 5, 6, 8, 9 and 12 were found consistently in all plates. The locus on chromosome 4 was found in most plates/homologous chromosomes of a pair (Fig. 25).

Nine CapSat12 loci were detected in eight chromosome pairs. All loci were major and located on the long chromosomal arms. Chromosome 5 had two loci in close proximity to one another. Chromosomes 1, 3, 4, 6, 8, 9 and 12 each carried one locus (Fig. 27).

Nine major CapSat13 loci were detected in *C. flexuosum*. All loci were located on the long chromosomal arms. Chromosomes 1, 3, 4, 6, 8, 9 and 12 each carried one locus whereas chromosome 5 had two CapSat13 loci in close proximity to one another (Fig. 21).

Nine CapSat15 loci were detected in eight chromosome pairs. Six loci were minor and three were major. Five of the loci were detected in all cells and all homologs whereas four loci were observed in majority (but not all) homologs. Chromosomes 1, 3 and 9 possessed one minor locus on the short chromosomal arm, whereas chromosomes 8 and 11 had one minor locus each on the long arm. Chromosomes 4 and 5 both had one major locus on the short arm whereas chromosome 2 had one minor and one major locus within its short arm (Fig. 27).

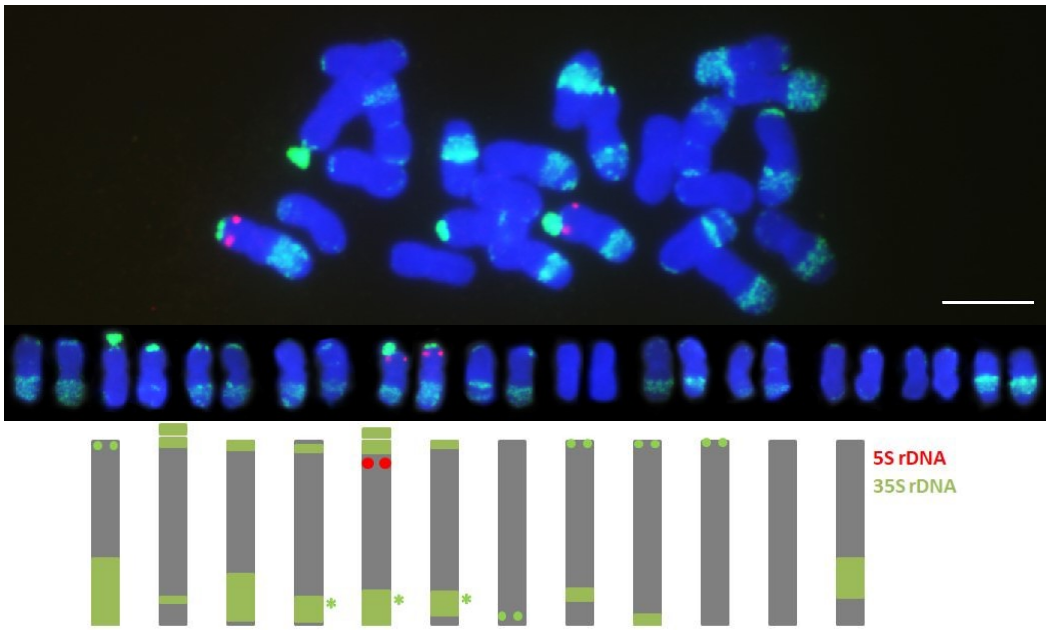

Fig. 19. Localization of 5S rDNA (red) and 35S rDNA (green) in chromosomes of *C. flexuosum* (A), cut-out karyotype (B) and corresponding ideogram (C). Cut-out chromosome pairs are arranged according to size and correlated with all other karyotypes across all figures (Fig19-27). Major loci are indicated as filled rectangles and minor loci as filled circles on ideograms. Asterisks indicate loci that were detected in most, but not all homologs/chromosomal spreads. Scale bar, 5µm.

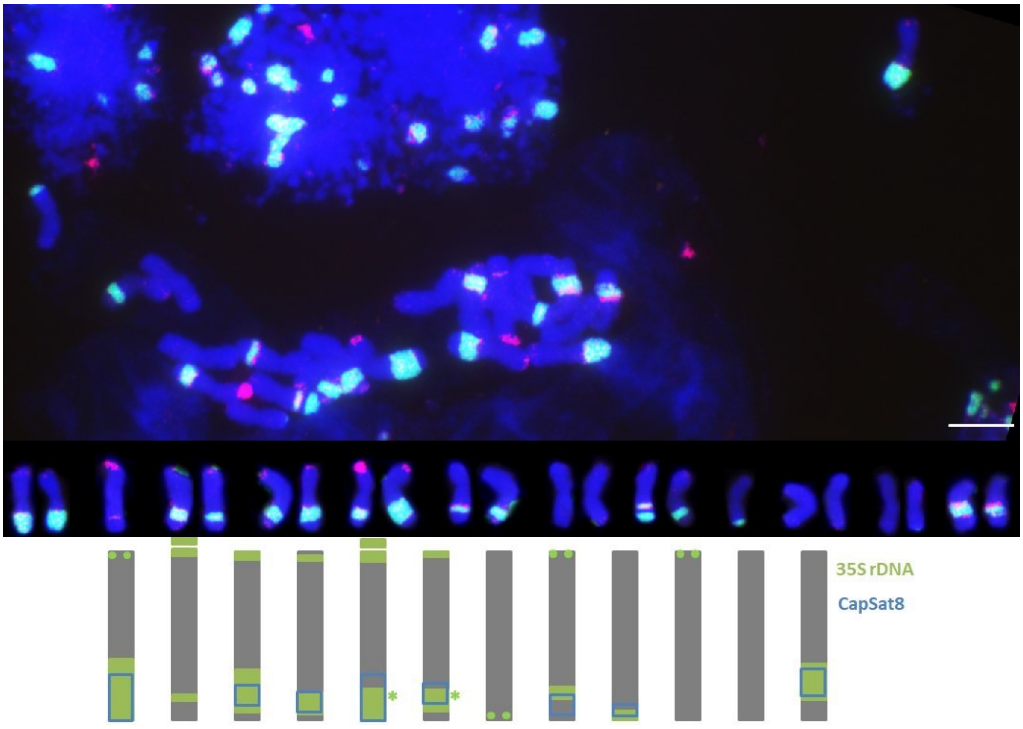

Fig. 20. Localization of 35S rDNA (red in photo, green in ideogram) and CapSat8 (green in photo, blue in ideogram) in chromosomes of *C. flexuosum* (A), cut-out karyotype (B) and corresponding ideogram (C). Cut-out chromosome pairs are arranged according to size and correlated with all other karyotypes across all figures (Fig19-27). Major loci are indicated as rectangles and minor loci as filled circles on ideograms. CapSat8 major loci are indicated as not-filled rectangles for better visualisation of those loci that overlapped with 35S rDNA loci. Asterisks indicate loci that were detected in most, but not all homologs/chromosomal spreads. Scale bar, 5µm.

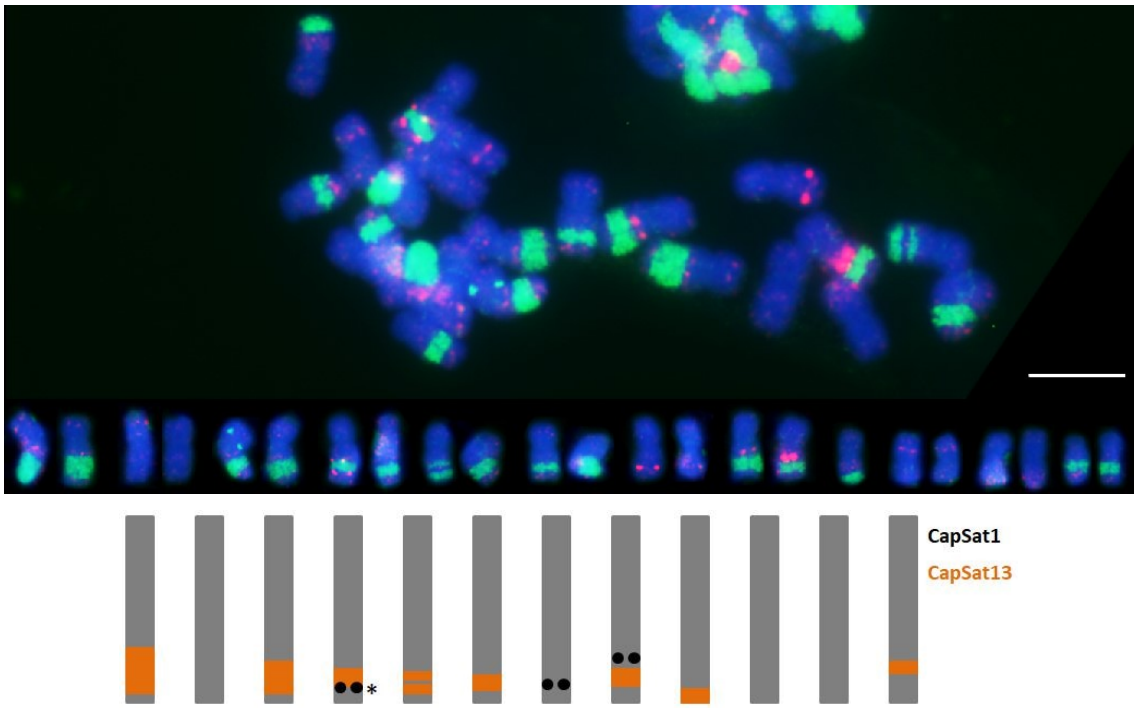

Fig. 21. Localization of CapSat1 (red in picture, black in ideogram) and CapSat13 (green in picture, orange in ideogram) in chromosomes of *C. flexuosum* (A), cut-out karyotype (B) and corresponding ideogram (C). Cut-out chromosome pairs are arranged according to size and correlated with all other karyotypes across all figures (Fig19-27). Major loci are indicated as filled rectangles and minor loci as filled circles on ideograms. Asterisks indicate loci that were detected in most, but not all homologs/chromosomal spreads. Scale bar, 5µm.

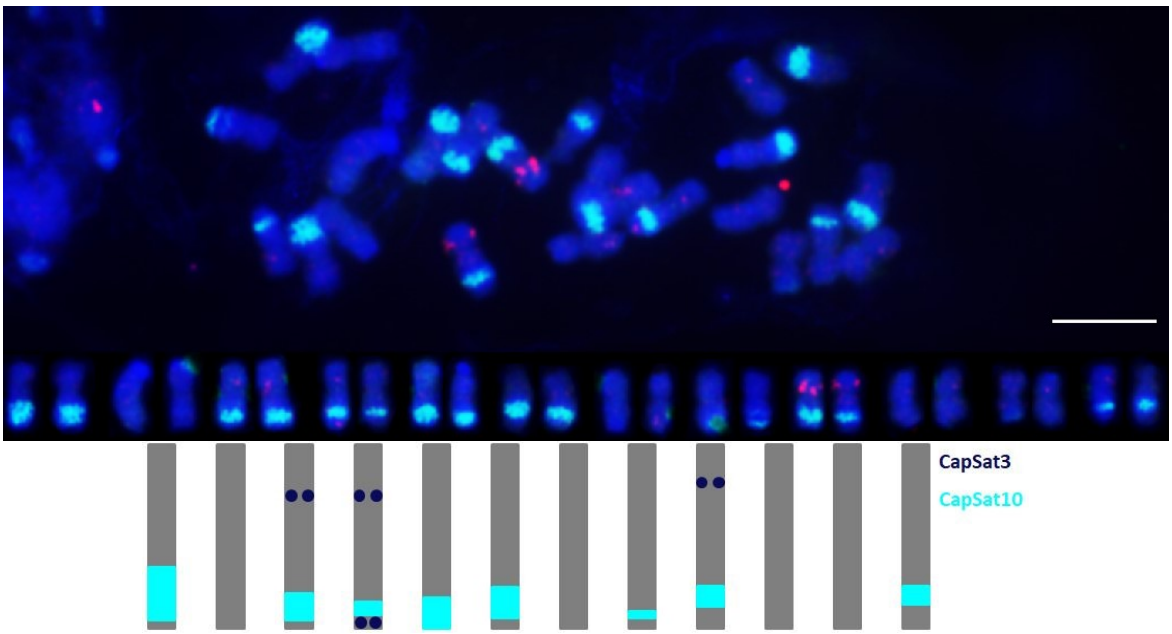

Fig. 22. Localization of CapSat3 (red in picture, dark blue in ideogram) and CapSat10 (green in picture, turquois in ideogram) in chromosomes of *C. flexuosum* (A), cut-out karyotype (B) and corresponding ideogram (C). Cut-out chromosome pairs are arranged according to size and correlated with all other karyotypes across all figures (Fig19-27). Major loci are indicated as filled rectangles and minor loci as filled circles on ideograms. Scale bar, 5µm.

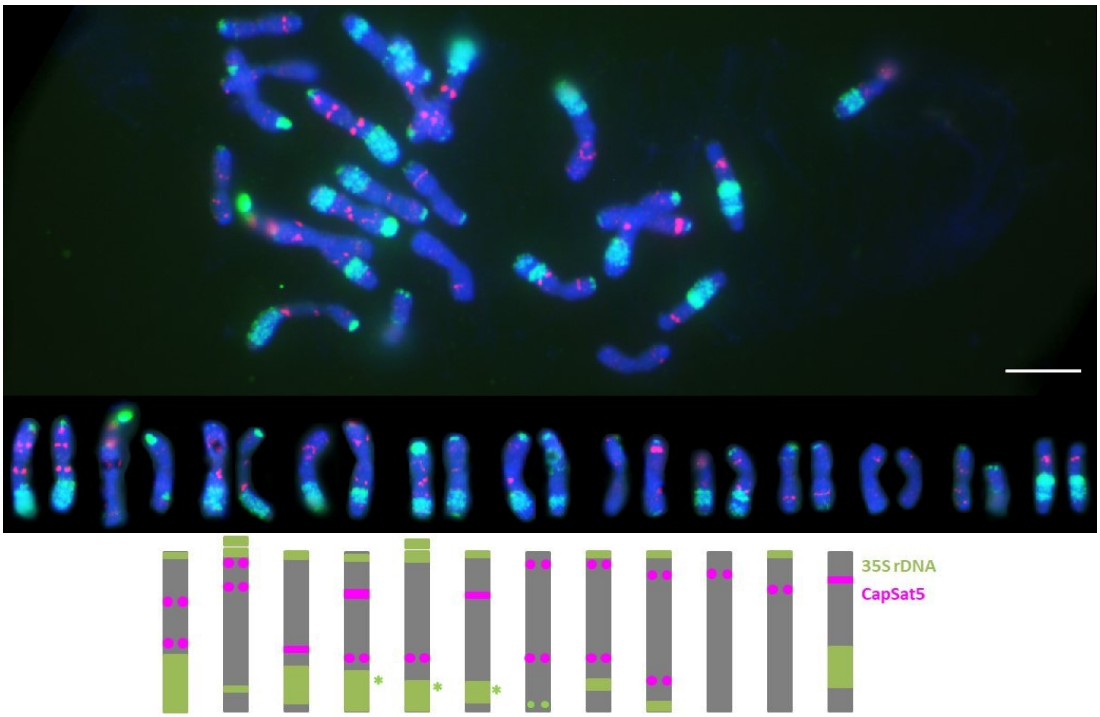

Fig. 23. Localization of CapSat5 (red in picture, pink in ideogram) and 35S rDNA (green in picture, green in ideogram) in chromosomes of *C. flexuosum* (A), cut-out karyotype (B) and corresponding ideogram (C). Cut-out chromosome pairs are arranged according to size and correlated with all other karyotypes across all figures (Fig19-27). Major loci are indicated as filled rectangles and minor loci as filled circles on ideograms. Asterisks indicate loci that were detected in most, but not all homologs/chromosomal spreads. Scale bar, 5µm.

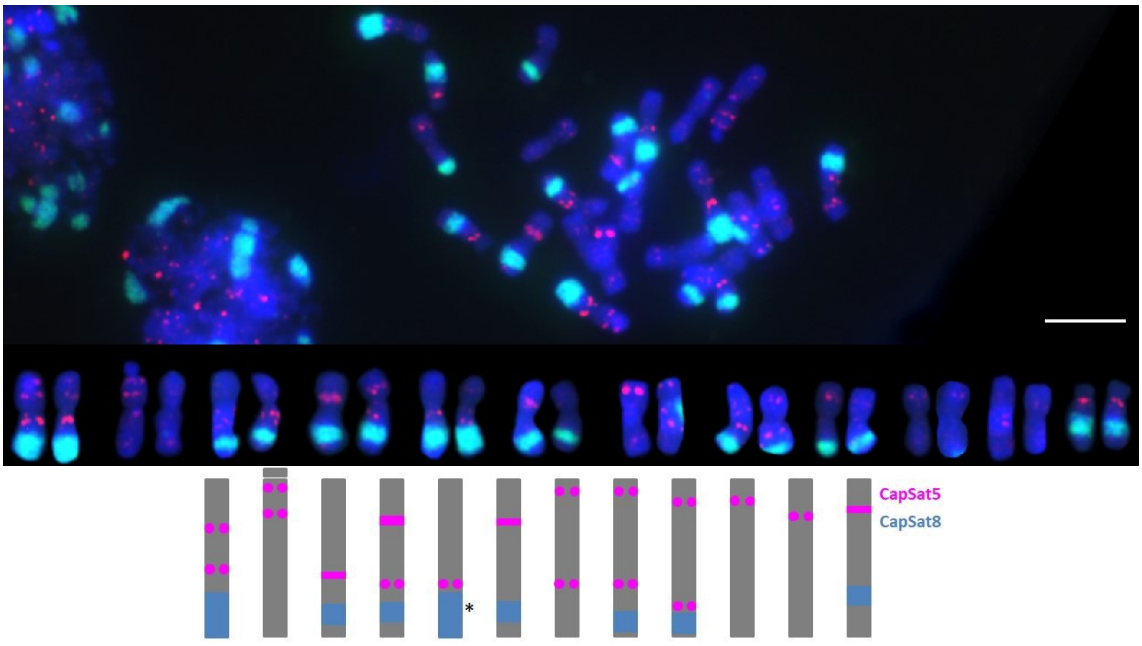

Fig. 24. Localization of CapSat5 (red in picture, pink in ideogram) and CapSat8 (green in picture, blue in ideogram) in chromosomes of *C. flexuosum* (A), cut-out karyotype (B) and corresponding ideogram (C). Cut-out chromosome pairs are arranged according to size and correlated with all other karyotypes across all figures (Fig19-27). Major loci are indicated as filled rectangles and minor loci as filled circles on ideograms. Asterisks indicate loci that were detected in most, but not all homologs/chromosomal spreads. Scale bar, 5µm.

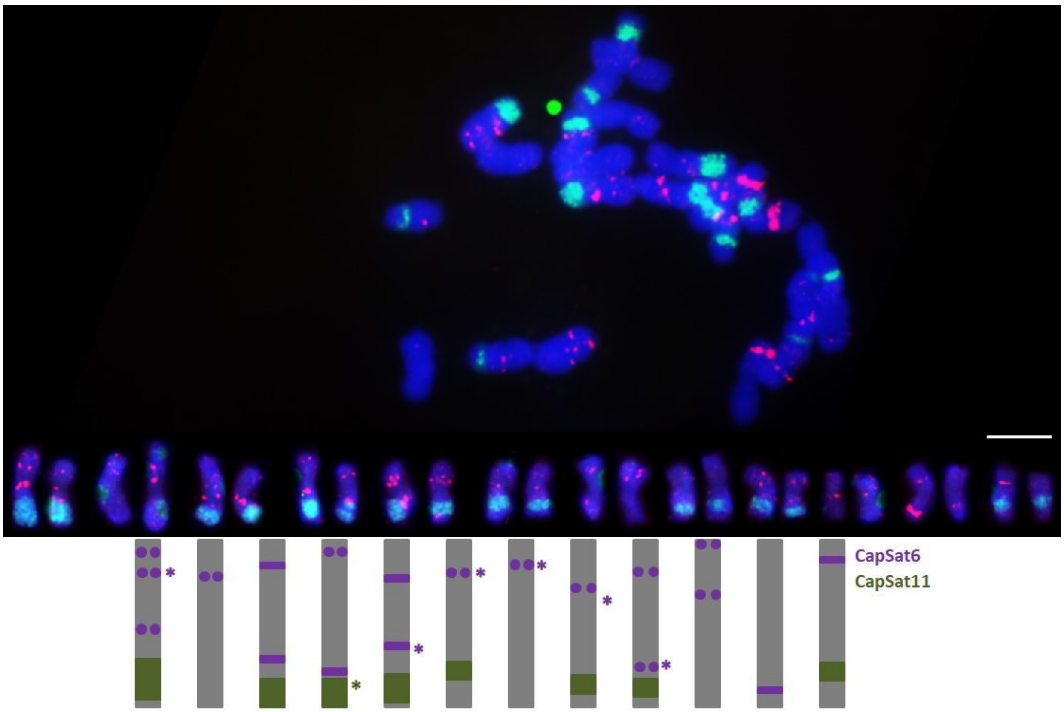

Fig. 25. Localization of CapSat6 (red in picture, purple in ideogram) and CapSat11 (green in picture, dark green in ideogram) in chromosomes of *C. flexuosum* (A), cut-out karyotype (B) and corresponding ideogram (C). Cut-out chromosome pairs are arranged according to size and correlated with all other karyotypes across all figures (Fig19-27). Major loci are indicated as filled rectangles and minor loci as filled circles on ideograms. Asterisks indicate loci that were detected in most, but not all homologs/chromosomal spreads. Scale bar, 5µm.

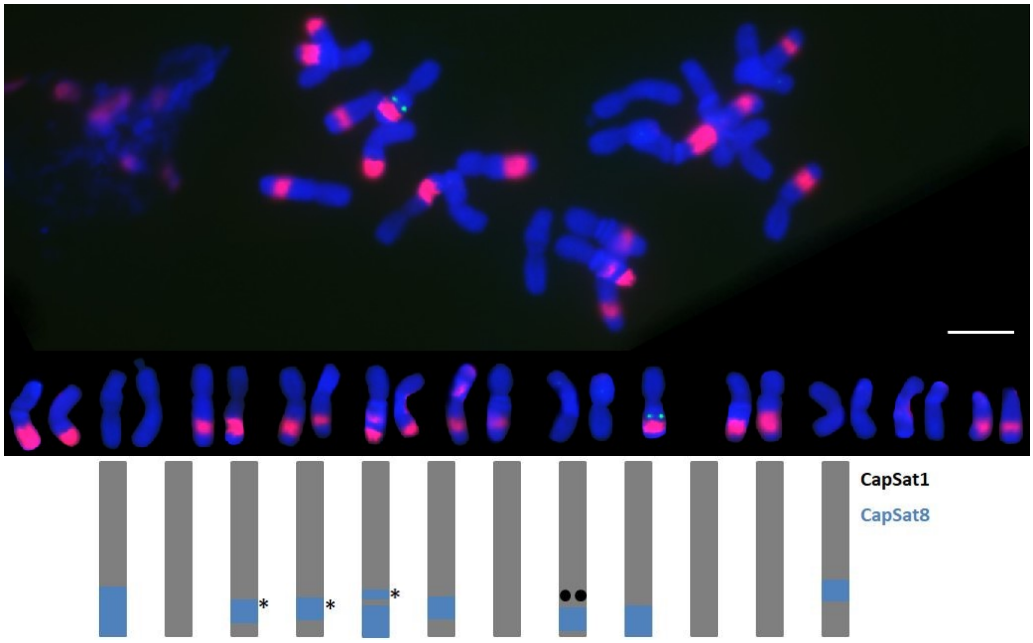

Fig. 26. Localization of CapSat8 (red in picture, blue in ideogram) and CapSat1 (green in picture, black in ideogram) in chromosomes of *C. flexuosum* (A), cut-out karyotype (B) and corresponding ideogram (C). Cut-out chromosome pairs are arranged according to size and correlated with all other karyotypes across all figures (Fig19-27). Major loci are indicated as filled rectangles and minor loci as filled circles on ideograms. Asterisks indicate loci that were detected in most, but not all homologs/chromosomal spreads. Scale bar, 5µm.

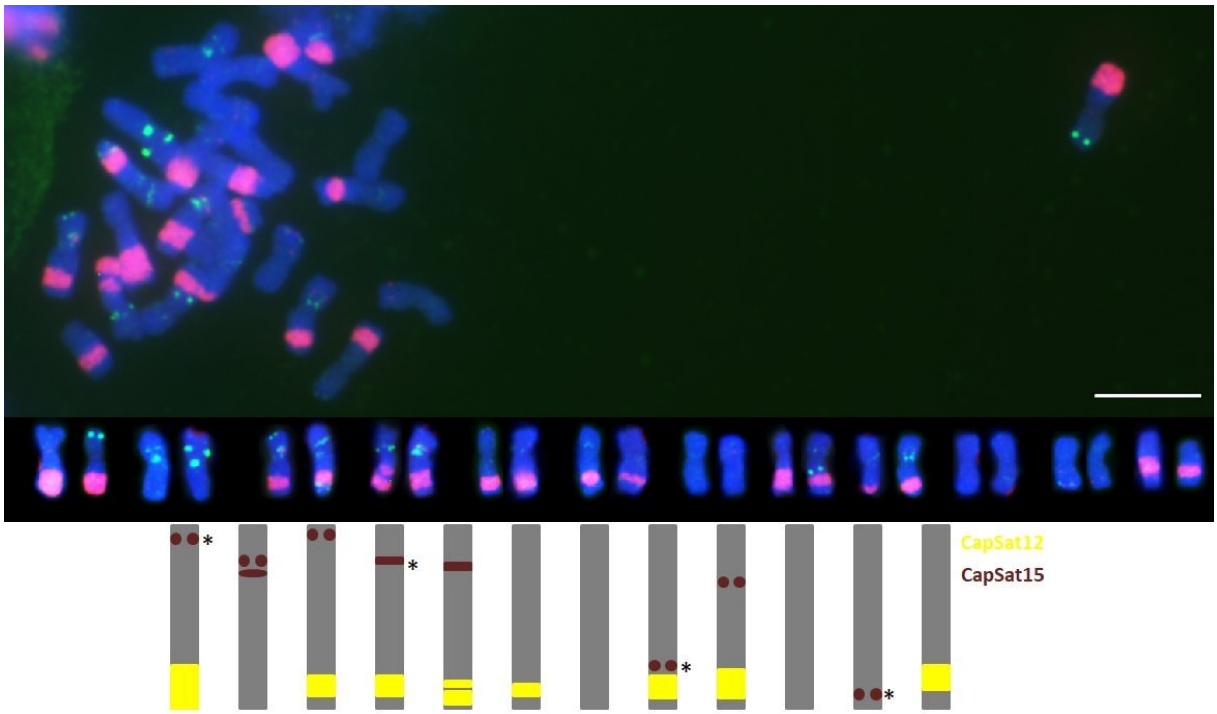

Fig. 27. Localization of CapSat12 (red in picture, yellow in ideogram) and CapSat15 (green in picture, dark red in ideogram) in chromosomes of *C. flexuosum* (A), cut-out karyotype (B) and corresponding ideogram (C). Cut-out chromosome pairs are arranged according to size and correlated with all other karyotypes across all figures (Fig19-27). Major loci are indicated as filled rectangles and minor loci as filled circles on ideograms. Asterisks indicate loci that were detected in most, but not all homologs/chromosomal spreads. Scale bar, 5µm.

Table 4. Number of loci of 5S and 35S rDNAs and ten satellite DNAs in *C. flexuosum* chromosomes. Numbers in brackets refer to number of major loci detected.

	Chr. 1	Chr. 2	Chr. 3	Chr. 4	Chr. 5	Chr. 6	Chr. 7	Chr. 8	Chr. 9	Chr. 10	Chr. 11	Chr. 12	total number
5S					1 (0)								1 (0)
35S	2 (1)	2 (2)	2 (2)	2 (1)	2 (2)	2 (2)	2 (0)	3 (3)	2 (2)	1 (0)	1 (0)	1 (1)	22 (16)
CapSat1				1 (0)			1 (0)	1 (0)					3 (0)
CapSat3			1 (0)	2 (0)					1 (0)				4 (0)
CapSat5	2 (0)	2 (0)	1 (1)	2 (1)	1 (0)	1 (1)	2 (0)	2 (0)	2 (0)	1 (0)	1 (0)	1 (1)	18 (4)
CapSat6	3 (0)	1 (0)	2 (2)	2 (1)	2 (2)	1 (0)	1 (0)	1 (0)	2 (0)	2 (0)	1 (1)	1 (1)	19 (7)
CapSat8	1 (1)		1 (1)	1 (1)	1 (1)	1 (1)		1 (1)	1 (1)			1 (1)	8 (8)
CapSat10	1 (1)		1 (1)	1 (1)	1 (1)	1 (1)		1 (1)	1 (1)			1 (1)	8 (8)
CapSat11	1 (1)		1 (1)	1 (1)	1 (1)	1 (1)		1 (1)	1 (1)			1 (1)	8 (8)
CapSat12	1 (1)		1 (1)	1 (1)	2 (2)	1 (1)		1 (1)	1 (1)			1 (1)	9 (9)
CapSat13	1 (1)		1 (1)	1 (1)	2 (2)	1 (1)		1 (1)	1 (1)			1 (1)	9 (9)
CapSat15	1 (0)	2 (1)	1 (0)	1 (1)	1 (1)			1 (0)	1 (0)		1 (0)		9 (3)

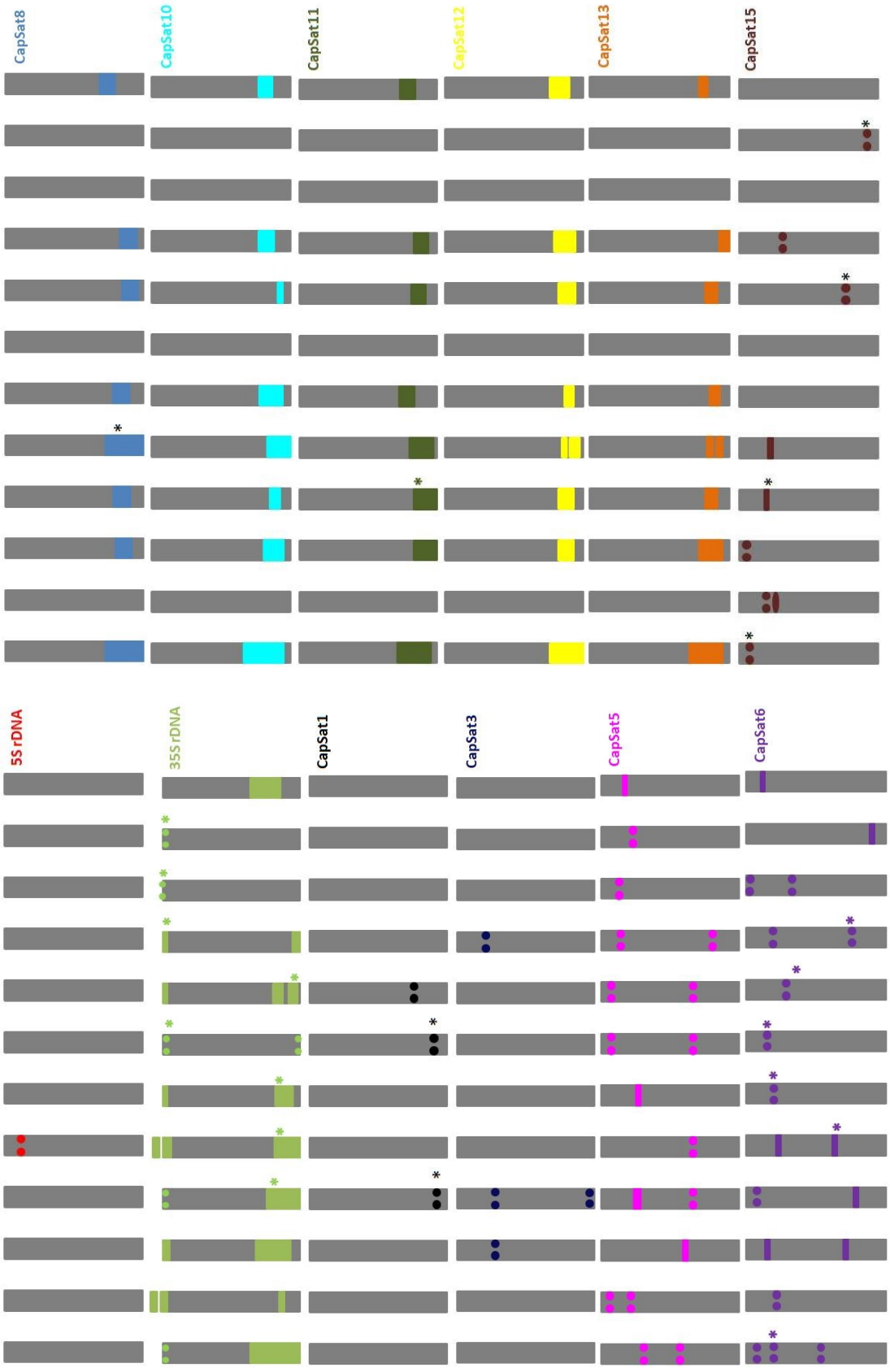

Fig. 28. Localization of all tandem repeats mapped in *C. flexuosum* chromosomes. The probes used are: 5S rDNA (red), 35S rDNA (green), CapSat1 (black), CapSat3 (dark blue), CapSat5 (pink), CapSat6 (purple), CapSat8 (blue), CapSat10 (turquois), CapSat11 (dark green), CapSat12 (yellow), CapSat13 (orange) and CapSat15 (dark red). Asterisks indicate loci that were detected in most, but not all plates/homologous chromosomes of a pair. Major loci are indicated as filled rectangles and minor loci as filled circles.

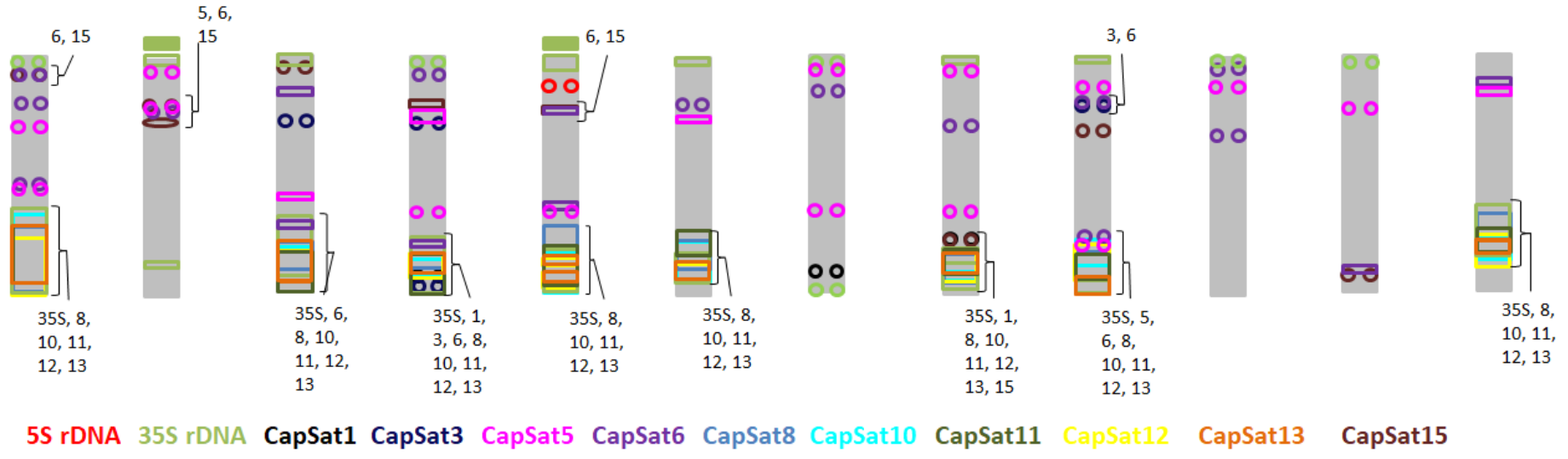

Fig. 29. Summary ideogram of all detected loci of 5S rDNA (red), 35S rDNA (green), CapSat1 (black), CapSat3 (dark blue), CapSat5 (pink), CapSat6 (purple), CapSat8 (blue), CapSat10 (turquois), CapSat11 (dark green), CapSat12 (yellow), CapSat13 (orange) and CapSat15 (dark red) in *C. flexuosum*. Numbers next to/below chromosomes indicate the probes mapped within composite blocks harbouring signals of multiple probes.

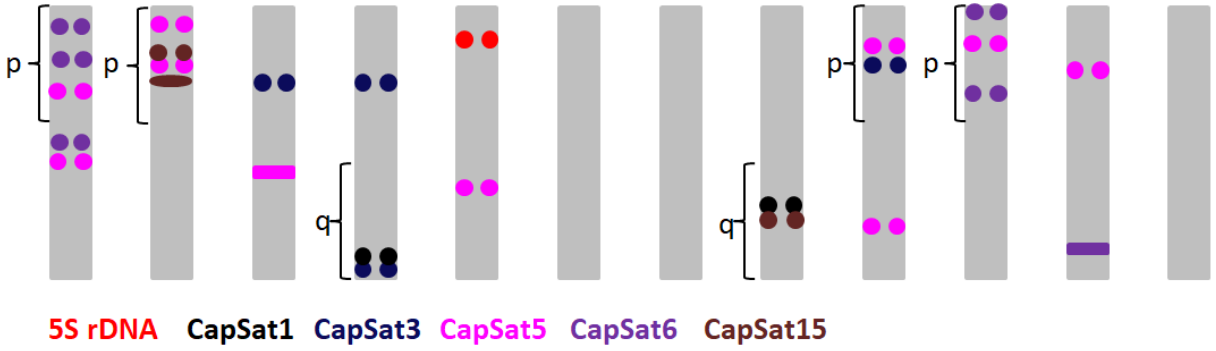

Fig. 30. Ideogram indicating probe combinations that unambiguously identify nine out of twelve chromosomes in *C. flexuosum*. Signals found on the long chromosomal arm (q) and the short chromosomal arm (p) are indicated.

Part IV – DISCUSSION

This study presents first detailed analyses of the repeatome in general and tandem repeats in particular of *C. flexuosum*, a wild chile pepper species with the largest known genome among all species of the genus *Capsicum*. The data allows to gain insight into patterns, diversity and dynamics of tandem repeats in this taxon and provides novel markers for individual chromosome identification in *Capsicum*.

Repeat profile of the genome of chile pepper species with largest genome

C. flexuosum was found to have a high amount of repetitive DNA (~60%). In some other plant genera like *Spinacia* a similar proportion of repetitive DNA was found (~50%) whereas in genera like *Beta* only a small fraction of the genome (~15%) consisted of repeats (Li et al., 2021). The present study showed that retrotransposons made up the biggest part of the genome in *C. flexuosum* (~50%) which was similar to other plant groups like soybean (~50%) and rice (~40%) but was lower than for example in maize (~80%; Vitte et al., 2014). Ty3-*gypsy* were the most abundant retroelements found in *C. flexuosum* (~30%) which correlates with similar proportions found in other genera like sugar beet, spinach (Li et al., 2021), sunflower (Staton et al., 2012) and onion (Fu et al., 2019). However, the repeat composition of even related genomes can vary significantly (Weiss-Schneeweiss et al., 2015).

In other genera of the Solanaceae family like *Solanum*, *Nicotiana* and *Petunia* retrotransposons also made up the biggest part of the repetitive DNA with LTR retroelements being the biggest contributor (Bombarely et al., 2016). The genome of hot pepper (*C. annuum*) was found to have 12-fold more Ty3-*gypsy* than Ty1-*copia* elements whereas this difference was not observed in the genome of tomato (Kim et al., 2014). This lead to the conclusion that the expansion of the pepper genome was due to the proliferation of Ty3-*gypsy* elements (Kim et al., 2014). These findings correlate with the present study where the amount of Ty3-*gypsy* and Ty1-*copia* retrotransposons were recovered in similar proportions, with Ty3-*gypsy* elements dominating.

Ty3-*gypsy* elements were shown to make up the majority of repetitive DNA in *C. annum, C. baccatum* and *C. chinense* while Ty1-*copia* elements were represented in low numbers (de Assis et al., 2020). In another study using BAC-clone sequencing, LTR elements were also

found to be the most abundant repeat element type in the genome of *C. annuum*. When compared with the tomato genome the study found LTR elements to be present at 22-times higher frequency in *Capsicum* genome compared to tomato. Similar to the present study Ty3-*gypsy* were found to be the most abundant repeat type in *C. annuum* with more than half of the repeats annotated as Ty3-*gypsy* (Park et al., 2011).

C. flexuosum was identified in this study as a species with the largest genome size in the genus *Capsicum* (1C = 7.14 pg). The genome sizes of nine different *Capsicum* species reported so far ranged from 3.35 pg (*C. chacoense*) to 5.77 pg (*C. parvifolium*) with a mean 1C value of 3.94 pg (Moscone et al., 2003). Thus, genome size of *C. flexuosum* is nearly twice as big as the mean genome size of the genus, despite the absence of polyploidy.

Tandem repeat diversity and characterization

Five tandem repeats identified in this study in the genome of *C. flexuosum* were used for karyotype characterization via FISH, together with two rDNA probes (5S and 35S rDNAs) and five satellite DNAs identified earlier in comparative analyses of *Capsicum* species representing all major evolutionary lineages (Weiss-Schneeweiss, unpubl.). In a study that mapped CapSat1 and CapSat5 to the chromosomes of *C. annum* and *C. pubescenes* via FISH, a high variation of loci number was found between these two species (Simon, 2020), similarly to the data presented in this study for *C. flexuosum*. This high variability in number and position of satellite DNA loci between closely related species is consistent with the fast evolution of satellite DNAs. Many of the satellite DNA families are shared among species and their monomer sequences are very similar among taxa, but their copy numbers and distribution in the genomes undergo dynamic changes.

Identification of novel satellite DNAs in *C. flexuosum*

Analyses of NGS DNA sequence data allowed not only for repeatome profile analysis, but also for identification of seven tandem repeats identified as satellite DNAs in the genome of *C. flexuosum*. Five of these satellites were previously identified in comparative analyses of 15 species of the genus (Weiss-Schneeweiss, unpubl.), whereas two tandem repeats were identified as putative novel satellite DNAs (CapSat17 and CapSat18). These two novel

repeats were present in the genome in very low proportions and analyses of their chromosomal localization are needed to confirm their satellite DNA status.

Variation in location and number of rDNA loci in *C. flexuosum*

35S and 5S rDNAs were so far mapped in 22 *Capsicum* species and varieties using FISH (Scaldaferro et al., 2016). The presence of one 5S rDNA locus was very conserved in all taxa. The number of 35S rDNA loci, in contrast, varied and was to a large extent correlated with phylogenetic grouping of the species. *C. flexuosum* was reported to have 14-15 35S rDNA loci and one 5S rDNA locus (Scaldaferro et al., 2016). These data roughly correspond to the number of loci detected in the present study where one 5S rDNA locus and 14-22 loci of 35S rDNA were recovered. The difference in number of 35S rDNA loci recovered could result from technical approaches, reflecting the various degrees of FISH resolution. In a current study protocol ff-GISH has been used which has been shown to be more sensitive in detecting minor loci (Jang & Weiss-Schneeweiss, 2015). Nearly all of the major 35S rDNA loci found in this study were also reported in the study of Scaldaferro et al. (2016). 35S rDNA loci that were only detected in this study represented minor signals, mostly found in subterminal position.

Two NORs pairs were reported in the present study and also in previous studies (Scaldaferro et al., 2016), which most likely represent the only functional 35S rDNA loci. Other signals detected using 35S rDNA probe likely represented either (1) a satellite DNA derived from small repeats present in IGS region or (2) cross-hybridization of GC-rich 35S rDNA probe to GC-rich satellite DNA, without any significant homology.

Although one 5S rDNA locus was consistently reported earlier and in the current study, the position of the 5S rDNA locus varied. In the present study 5S rDNA was reported to be located on the short arm of a chromosome that also carried one of the NOR regions (and thus a major 35S rDNA locus). In the study of Scaldaferro et al. (2016) the 5S rDNA locus did not map to either of the two chromosome pairs that carried NORs in *C. flexuosum*. Additionally, two variants of the chromosome carrying 5S rDNA were reported in the study of Scaldaferro et al. (2016), one that only had one 35S rDNA locus and another one with an additional subterminal 35S rDNA signal within the same chromosomal arm.

None of the 22 previously analysed samples of 14 different *Capsicum* species carried 5S rDNA locus on a chromosome that also carried NOR (Scaldaferro et al., 2016). Further analyses are required to clarify if the variation of location of the 5S rDNA locus is found in more samples of *C. flexuosum* and thus represent populational/individual polymorphisms.

Satellite DNAs as chromosomal markers

The localization of all twelve tandem repeat probes allowed for the identification of all individual 12 chromosome pairs in *C. flexuosum*. Some loci were consistently detected in all cells/homologous chromosomes while other loci were only detected in some but not all analysed cells/homologous chromosomes. These loci were almost always minor. Such variation might at least in part reflect technical difficulties as FISH is a semi-quantitative method and the level of detection depends on the target-DNA accessibility to the probes. However, some of the variation between the individuals/populations might reflect the existing polymorphisms. Tandem repeats with low copy numbers per locus are also more difficult to detect especially if the monomer sequences are short. Additionally, the degree of condensation of the chromosomes and their quality of preparation might have also likely influenced the ability of the probes to bind to target chromosomal DNA.

The present study identified 6.69% of the genome of *C. flexuosum* as satellite DNAs via analyses of repeat profiles using RepeatExplorer NGS DNA sequence data analyses. While this is already a relatively high percentage of the genome, the data obtained from FISH suggested that even a bigger part of the genome could consist of satellite DNAs. The discrepancies in accurate estimates of satellite DNAs amount in the genomes via NGS-data analyses versus wet-lab based analyses (Southern blotting or dot blots) have been addressed in other studies (Emadzade et al., 2014) and result, among others, from technical issues related to DNA fragmentation. Thus, analyses of copy numbers of the individual satellite DNA families using dot blotting are needed to provide more accurate estimates of their copy number in *C. flexuosum*.

The results of the present study indicated that some satellite DNA probes localized to similar chromosomal regions. Specifically, satellite DNAs CapSat8, CapSat10, CapSat11, CapSat12 and CapSat13 were mapped to the same regions of chromosomes as 35S rDNA probe (with the exception of NOR regions). A previous study reported mapping of genus-specific IGS

regions in *Capsicum annum* (Jo et al., 2011) and suggested that the IGS regions might have given rise to a satellite DNA. This previously unnamed satellite DNA might correspond to one or more of the satellite DNAs found here in similar chromosomal regions as interstitial 35S rDNA signals.

Several of the satellite DNAs loci (CapSat8, CapSat10, CapSat11, CapSat12 and CapSat13) co-localized not only with 35S rDNA but also with heterochromatic blocks in at least some chromosomes of *C. flexuosum*. CapSat8 clearly was the only satellite DNA that co-localized with almost all of heterochromatic blocks but its appearance was "diffused" within blocks. Thus, it is likely that the large heterochromatic blocks are composed of several types of interspersed satellite DNAs. Mapping of these satellite DNA probes in stretched DNA fibers (fiber FISH) will allow for more conclusions about their arrangement in individual heterochromatic segments of all chromosomes.

Localization of two rRNA genes and ten satellite DNAs allowed for characterization and identification of all twelve chromosome pairs in *C. flexuosum*

Satellite DNAs and rDNAs are often used as cytological markers. They can be useful to study and trace the evolution of karyotypes within a species as well as in related species (Cuadrado & Jouve, 2002; Lim et al., 2000; Navrátilová et al., 2003; Pires et al., 2004). This study used two rDNAs and ten satellite DNAs to identify all twelve chromosome pairs in *C. flexuosum* (Fig. 29).

All twelve chromosome pairs showed unique combinations of signals and could thus be clearly identified. Chromosomes 1, 3, 4, 5, 6, 8, 9 and 12 all carried a 35S rDNA, CapSat8 and CapSat10-13 locus on the long chromosomal arms covering more or less the same heterochromatic region. This landmark clearly distinguished them from the remaining four chromosomes. Chromosome 1 could be clearly identified as it was the only chromosome that had two CapSat5 loci. Chromosomes 3 and 9 both had one single CapSat3 locus, but could be distinguished by their CapSat5 loci number and distribution. Chromosome 4 and 8 both had single CapSat1 locus, but could be distinguished by the presence/absence of CapSat3 satellite DNA. Chromosome 5 carried the only 5S rDNA locus in the genome and possessed a 35S rDNA/NOR region which made its identification the easiest. Chromosome 6 and 12 mapped

the same number of satellite DNAs loci, but could be distinguished by the different number of 35S rDNA loci (Fig. 29).

Chromosomes 2, 7, 10 and 11 were distinguishable from the other chromosomes because they did not carry any loci of CapSat8 and CapSat10-13 on the long chromosomal arms. Chromosome 2 was the only chromosome in this group that could be distinguished by the presence of a NOR region, in addition to CapSat5 and CapSat15 loci on the short chromosomal arm. Chromosome 7 was distinguishable as it was the only one with a CapSat1 locus in this group. The only chromosome in this group with two CapSat6 loci was chromosome 10. Chromosome 11 was the only one that carried single locus of 35S rDNA, CapSat5, CapSat6 and CapSat15 (Fig. 29).

Despite the fact that all chromosomes carried unique probe combinations, the use of ten satellite DNAs and two rDNA probes for quick chromosome identification would not be very practical. Therefore, sets of two probes were identified that allowed for fast identification of several individual chromosomes (Fig. 30). Nine of the twelve chromosome pairs were clearly distinguishable using only two probes per chromosome and six probes in total. Chromosome 1 carried two CapSat6 loci as well as one CapSat5 locus on the short chromosomal arm and one CapSat6 and CapSat5 locus on the long arm. Chromosome 2 mapped two CapSat5 loci and two CapSat15 loci on the short chromosomal arm. Chromosome 3 had one CapSat3 locus on the short and one CapSat5 locus on the long chromosomal arm. Chromosome 4 carried a CapSat1 and CapSat3 locus on the long and one CapSat3 locus on the short chromosomal arm. Chromosome 5 was clearly distinguishable as it was the only chromosome with a 5S rDNA locus which mapped to the short chromosomal arm, additionally it can be identified by its CapSat5 locus on the long chromosomal arm. Chromosome 8 mapped a CapSat1 and a CapSat15 locus on the long chromosomal arm. Chromosome 9 carried a CapSat5 and a CapSat3 locus on the short arm and one CapSat5 locus on the long chromosomal arm. Chromosome 10 mapped one CapSat5 locus flanked by two CapSat6 loci on the short chromosomal arm. Chromosome 11 mapped one CapSat5 locus on the short and one CapSat6 locus on the long chromosomal arm. Chromosome 6, 7 and 12 were not able to be identified by only two probes. Such probe combinations will aid quick identification of most of the chromosomes of the species and allow for individual chromosome microdissection planned for further analyses of karyotype evolution in the genus *Capsicum*.

References

Aguilar-Meléndez A, Morrell PL, Roose ML, Kim SC (2009) Genetic diversity and structure in semiwild and domesticated chiles (*Capsicum annuum*; Solanaceae) from Mexico. *Am. J. Bot.* 96: 1190–1202.

Ambrožová K, Mandáková T, Bureš P, Neumann P, Leitch IJ, Koblížková A, Macas J et al. (2011) Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria lilies*. *Ann. Bot.* 107: 255–268.

Arkhipova IR, Yushenova IA, Angert E (2019) Giant transposons in eukaryotes: is bigger better? *Genome Biol. Evol.* 11(3): 906–918.

Baranyi M, Greilhuber J, Święcicki WK (1996) Genome size in wild *Pisum* species. *Theor. Appl. Genet* 93(5-6): 717–721.

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

Biscotti MA, Canapa A, Forconi M, Olmo E, Barucca M (2015a) Transcription of tandemly repetitive DNA: functional roles. *Chromosome Res.* 23: 463–477.

Biscotti MA, Olmo E, Heslop-Harrison JS (2015b) Repetitive DNA in eukaryotic genomes. *Chromosome Res.* 23: 415–420.

Bombarely A, Moser M, Amrad A, Bao M, Bapaume L, Barry CS, Bliek M et al. (2016) Insight into the evolution of the Solanaceae from the parental genomes of *Petunia hybrida*. *Nat. Plants* 2(6): 16074.

Bosland PW, Coon D, Cooke PH (2015) Novel formation of ectopic (nonplacental) capsaicinoid secreting vesicles on fruit walls explains the morphological mechanism for super-hot chile peppers. J. *Am. Soc. Hortic. Sci.* 140: 253–256.

Bosland PW, Votava EJ (2000) Peppers: Vegetable and Spice *Capsicums*. Crops Production Science in Horticulture 12. CAB International Publishing, Wallingford, UK, 204.

Carrizo García C, Sehr EM, Barfuss MHJ, Barboza GE, Samuel R, Moscone EA, Ehrendorfer F (2016) Phylogenetic relationships, diversification and expansion of chili peppers (*Capsicum*, Solanaceae). *Ann. Bot.* 118: 35–51.

Carrizo García C, Sterpetti M, Volpi P, Ummarino M, Saccardo F (2013) Wild *Capsicums*: identification and in situ analysis of Brazilian species. In: *Breakthroughs in the genetics and breeding of* Capsicum *and eggplant.* Lanteri S, Rotino GL, eds. Proceedings of the XV EUCARPIA Meeting on Genetics and Breeding of *Capsicum* and Eggplant: pp. 205–213.

Chester M, Leitch AR, Soltis PS, Soltis DE (2010) Review of the application of modern cytogenetic methods (FISH/GISH) to the study of reticulation (polyploidy/hybridization). *Genes* 1: 166–192.

Cížková J, Hribová A, Humplíková L, Christelová P, Suchánková P, Doležel J (2013) Molecular analysis and genomic organization of major DNA satellites in banana (*Musa spp*.). PLoS ONE 8(1): e54808.

Cremonini R (2005) Low chromosome number angiosperms. *Caryologia* 58: 403–409.

Cuadrado A, Jouve N (2002) Evolutionary trends of different repetitive DNA sequences during speciation in the genus *Secale*. *J. Hered.* 93: 339–345.

de Assis R, Baba VY, Cintra LA, Gonçalves LSA, Rodrigues R, Vanzela ALL (2020) Genome relationships and LTR-retrotransposon diversity in three cultivated *Capsicum* L. (Solanaceae) species. *BMC Genomics* 21(1): 237.

de la Herrán R, Robles F, Cuñado N, Santos JL, Ruiz Rejón M, Garrido-Ramos MA, Ruiz Rejón C (2001) A heterochromatic satellite DNA is highly amplified in a single chromosome of *Muscari* (Hyacinthaceae). *Chromosoma* 110: 197–202.

Devos KM, Brown JKM, Bennetzen JL (2002) Genome size reduction through illegitimate recombination counteracts genome expansion in *Arabidopsis*. *Genome Res.* 12: 1075–1079.

Du J, Tian Z, Hans CS, Laten HM, Cannon SB, Jackson SA, Shoemaker RC et al. (2010) Evolutionary conservation, diversity and specificity of LTR-retrotransposons in flowering plants: insights from genome-wide analysis and multi-specific comparison. *Plant J.* 63: 584–598.

Emadzade K, Jang TS, Macas J, Kovařík A, Novák P, Parker J, Weiss-Schneeweiss H (2014) Differential amplification of satellite PaB6 in chromosomally hypervariable *Prospero autumnale* complex (Hyacinthaceae). *Ann. Bot.* 114: 1597–1608.

Eshbaugh WH (1993) Peppers: history and exploitation of a serendipitous new crop discovery. In: *New crops*. Janick J, Simon JE, eds.. New York: Wiley: 132–139.

Finnegan DJ (1989) Eukaryotic transposable elements and genome evolution. *Trends Genet.* 5: 103–107.

Fraenkel L, Bogardus ST Jr, Concato J, Wittink DR (2004) Treatment options in knee osteoarthritis: the patient's perspective. *Arch. Intern. Med.* 164: 1299–1304.

Friebe B, Zhang P, Linc G, Gill BS (2005) Robertsonian translocations in wheat arise by centric misdivision of univalents at anaphase I and rejoining of broken centromeres during interkinesis of meiosis II. *Cytogenet. Genome Res.* 109: 293–297.

Fu J, Zhang H, Guo F, Ma L, Wu J, Yue M, Zheng X et al. (2019) Identification and characterization of abundant repetitive sequences in *Allium cepa*. *Sci. Rep.* 9(1): 16756.

Fukui K, Nakayama S, eds. (1996) Plant chromosomes:laboratory methods. Boca Raton, FL, USA: CRC Press.

Garcia S, Kovařík A (2013) Dancing together and separate again: gymnosperms exhibit frequent changes of fundamental 5S and 35S rRNA gene (rDNA) organisation. *Heredity*. 111(1): 23–33.

Garcia S, Lim KY, Chester M, Garnatje T, Pellicer J, Vallès J, Leitch AR et al. (2009) Linkage of 35S and 5S rRNA genes in *Artemisia* (family Asteraceae): first evidence from angiosperms. *Chromosoma* 118(1): 85–97.

Garrido-Ramos MA (2015) Satellite DNA in plants: more than just rubbish. *Cytogenet. Genome Res.* 146: 153–170.

Garrido-Ramos MA (2017) Satellite DNA: an evolving topic. *Genes* 8(9): 230.

Grant V (1982) Chromosome number patterns in primitive angiosperms. *Bot. Gaz.* 143: 390–394.

Gregory TR (2005) Genome size evolution in animals. In: *The Evolution of the Genome*. Gregory TR, ed.; Elsevier: Burlington, NJ, USA, pp. 3–87.

Gregory TR, Nicol JA, Tamm H, Kullman B, Kullman K, Leitch IJ, Murray BG et al. (2007) Eukaryotic genome size databases. *Nucleic Acids Res.* 35: D332–338.

Grenfell AW, Strzelecka M, Heald R (2017) Transcription brings the complex(ity) to the centromere. Cell Cycle 16: 235–236.

Guerra M (2008) Chromosome numbers in plant cytotaxonomy: concepts and implications. *Cytogenet. Genome Res.* 120: 339–350.

Havecker ER, Gao X, Voytas DF (2004) The diversity of LTR retrotransposons. *Genome Biol.* 5(6): 225.

Heiser CB (1995) Peppers: *Capsicum* (Solanaceae) In: Smartt J, Sommonds NW, eds. *Evolution of crop plants.* Essex: Longman, pp. 449–451.

Heslop-Harrison JS, Schwarzacher T (2011) Organisation of the plant genome in chromosomes. *Plant J.* 66(1): 18–33.

Hribová E, Neumann P, Matsumoto T, Roux N, Macas J, Dolezel J (2010) Repetitive part of the banana (*Musa acuminata*) genome investigated by low-depth 454 sequencing. *BMC Plant Biol.*10, 204.

Hunziker AT (1979) South American Solanaceae: a synoptic survey. In: *The biology and taxonomy of the Solanaceae.* Hawkes JG, Lester RN, Skelding AD, eds. Linnean Society Symposium, London: Adademic Press 7: 49–85.

Hunziker AT (2001) *Genera Solanacearum. The Genera of Solanaceae Illustrated, Arranged According to a New System.* Gantner Verlag, Ruggell, Liechtenstein.

International Board for Plant Genetic Resources (IBPGR) (1983) *Genetic Resources of Capsicum, a Global Plan of Action.* IBPGR Executive Secretariat, Rome, Italy.

Jang TS, Weiss-Schneeweiss H (2015) Formamide-free genomic *in situ* hybridization allows unambiguous discrimination of highly similar parental genomes in diploid hybrids and allopolyploids. *Cytogenet. Genome Res.* 146(4): 325–331.

Jo SH, Park HM, Kim SM, Kim HH, Hur CG, Choi D (2011) Unraveling the sequence dynamics of the formation of genus-specific satellite DNAs in the family solanaceae. *Heredity* 106(5): 876–885.

Kashkush K, Feldman M, Levy AA (2002) Transcriptional activation of retrotransposonsalters the expression of adjacent genes in wheat. *Nat. Genet.* 33: 102–106.

Kejnovsky E, Hobza R, Cermak T, Kubat Z, Vyskot B (2009) The role of repetitive DNA in structure and evolution of sex chromosomes in plants. *Heredity* 102: 533–541.

Kim S, Park M, Yeom SI, Kim YM, Lee JM, Lee HA, Seo E et al. (2014) Genome sequence of the hot pepper provides insights into the evolution of pungency in *Capsicum* species. *Nat. Genet.* 46: 207–278.

Kopecna O, Kubickova S, Cernohorska H, Cabelova K, Vahala J, Martinkova N, Rubes J (2014) Tribe-specific satellite DNA in non-domestic Bovidae. *Chromosome Res.* 22: 277–291.

Koukalova B, Moraes AP, Renny-Byfield S, Matyasek R, Leitch AR, Kovařík A (2010) Fall and rise of satellite repeats in allopolyploids of *Nicotiana* over c. 5 million years. *New Phytol.* 186: 148–160.

Kubis S, Schmidt T, Heslop-Harrison JS (1998) Repetitive DNA elements as a major component of plant genomes. *Ann. Bot.* 82: 45–55.

Kumar A, Bennetzen JL (1999) Plant retrotransposons. *Annu. Rev. Genet.* 33: 479–532.

Leitch IJ, Johnston E, Pellicer J, Hidalgo O, Bennett MD (2019) Angiosperm DNA C-values database (release 9.0, Apr 2019) https://cvalues.science.kew.org/. accessed: 8.6.2022.

Lejeune MP, Kovacs EM, Westerterp-Plantenga MS (2003) Effect of capsaicin on substrate oxidation and weight maintenance after modest body-weight loss in human subjects. *Br. J. Nutr.* 90: 651–659.

Leprince AS, Grandbastien MA, Meyer C (2001) Retrotransposons of the Tnt1B family are mobile in *Nicotiana plumbaginifolia* and can induce alternative splicing of the host gene upon insertion. *Plant Mol. Biol.* 47: 533–541.

Li N, Li X, Zhou J, Yu L, Li S, Zhang Y, Qin R et al. (2021) Genome-wide analysis of transposable elements and satellite DNAs in *Spinacia* species to shed light on their roles in sex chromosome evolution. *Front. Plant Sci.* 11: 575462.

Li W, Zhang P, Fellers JP, Friebe B, Gill BS (2004) Sequence composition, organization, and evolution of the core Triticeae genome. *Plant J.* 40: 500–511.

Lim KY, Matyasek R, Lichtenstein CP, Leitch AR (2000) Molecular cytogenetic analyses and phylogenetic studies in the *Nicotiana* section *Tomentosae*. *Chromosoma* 109: 245–258.

Lisch D (2012) How important are transposons for plant evolution? *Nat. Rev. Genet.* 14: 49–61.

Llorens C, Futami R, Covelli L, Domínguez-Escribá L, Viu JM, Tamarit D, Aguilar-Rodríguez J et al. (2011) The *Gypsy* Database (GyDB) of mobile genetic elements: release 2.0. *Nucleic Acids Res.* 39: D70–D74.

López-Flores I, Garrido-Ramos MA (2012) The repetitive DNA content of eukaryotic genomes. *Genome Dyn.* 7: 1–28.

Ludy MJ, Moore GE, Mattes RD (2012) The effects of capsaicin and capsiate on energy balance: critical review and meta-analyses of studies in humans. *Chem. Senses* 37: 103–121.

Lysák MA, Schubert I (2013) Mechanisms of chromosome rearrangements. In: *Plant genome diversity, vol 2, Physical structure, behaviour and evolution of plant genomes.* Leitch IJ, Greilhuber J, Dolezel J, Wendel JF, eds. Springer-Verlag, Wien: pp. 137–147.

Macas J, Mészáros T, Nouzová M (2002) PlantSat: a specialized database for plant satellite repeats. *Bioinformatics* 18(1): 28–35.

Macas J, Neumann P, Navrátilová A (2007) Repetitive DNA in the pea (*Pisum sativum* L.) genome: comprehensive characterization using 454 sequencing and comparison to soybean and *Medicago truncatula*. *BMC Genomics* 8: 427.

Macas J, Požárková D, Navrátilová A, Nouzová M, Neumann P (2000) Two new families of tandem repeats isolated from genus *Vicia* using genomic self-priming PCR. *Mol. Gen. Genet.* 263: 741–751.

Małuszyńska J, Hasterok R, Weiss H (1998) rRNA genes: their distribution and activity in plants. In: *Plant cytogenetics*. Małuszyńska J, ed. Silesian University Press, Katowice: pp. 75–79.

Masterson J (1994) Stomatal size in fossil plants–evidence for polyploidy in majority of angiosperms. *Science* 264: 1759–1763.

Michael TP (2014) Plant genome size variation: bloating and purging DNA. *Brief Funct. Genomics* 13(4): 308–317.

Morgante M, Brunner S, Pea G, Fengler K, Zuccolo A, Rafalski A (2005) Gene duplication and exon shuffling by helitron-like transposons generate intraspecies diversity in maize. *Nat. Genet.* 37: 997–1002.

Mori A, Lehmann S, O'Kelly J, Kumagai T, Desmond JC, Pervan M, McBride WH et al. (2006) Capsaicin, a component of red peppers, inhibits the growth of androgen-independent, p53 mutant prostate cancer cells. *Cancer Res.* 66: 3222–3229.

Moscone EA (1993) Estudios cromosomicos en *Capsicum* (Solanaceae) II. Analisis cariotipico de *C. parvifolium* y *C. annuum var. annuum*. *Kurtziana* 22: 9–18.

Moscone EA, Baranyi M, Ebert I, Greilhuber J, Ehrendorfer F, Hunziker AT (2003) Analysis of nuclear DNA content in *Capsicum* (Solanaceae) by flow cytometry and Feulgen densitometry. *Ann. Bot.* 92: 21–29.

Moscone EA, Lambrou M, Ehrendorfer F (1996) Fluorescent chromosome banding in the cultivated species of *Capsicum* (Solanaceae). *Plant Syst. Evol.* 202: 37–63.

Moscone EA, Scadalferro MA, Gabriele M, Cecchini NM, Sánchez García Y, Jarret R, Daviña JR et al. (2007) The evolution of chili peppers (*Capsicum* – Solanaceae): a cytogenetic perspective. *Acta Hortic.* 745: 137–169.

Navrátilová A, Neumann P, Macas J (2003) Karyotype analysis of four *Vicia* species using *in situ* hybridization with repetitive sequences. *Ann. Bot*. 91: 921–926.

Neumann P, Novák P, Hoštáková N, Macas J (2019) Systematic survey of plant LTR-retrotransposons elucidates phylogenetic relationships of their polyprotein domains and provides a reference for element classification. *Mob. DNA* 10(1): 1.

Novák P, Neumann P, Macas J (2010) Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinform.*11: 378.

Novák P, Neumann P, Macas J (2020) Global analysis of repetitive DNA from unassembled sequence reads using RepeatExplorer2. *Nat. Protoc.* 15(11): 3745–3776.

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. *Bioinformatics*, 29 (6): 792–793.

Orozco-Arias S, Isaza G, Guyot R (2019) Retrotransposons in plant genomes: structure, identification, and classification through bioinformatics and machine learning. *Int. J. Mol. Sci.* 20(15): 3837.

Park M, Jo S, Kwon JK, Park J, Ahn JH, Kim S, Lee YH et al. (2011) Comparative analysis of pepper and tomato reveals euchromatin expansion of pepper genome caused by differential accumulation of Ty3/*Gypsy*-like elements. *BMC Genomics* 12: 85.

Pezer Z, Brajkovic J, Feliciello I, Ugarkovic D (2012) Satellite DNA-mediated effects on genome regulation. *Genome Dyn.* 7: 153–169.

Pickersgill B (1988) The genus *Capsicum*: a multidisciplinary approach to the taxonomy of cultivated and wild plants. *Biol Zent Bl* 107: 381–389.

Pickersgill B (1991) Cytogenetics and evolution of *Capsicum* L. In: *Chromosome engineering in plants: genetics, breeding, evolution*. Tsuchyia T, Gupta PK, eds. Amsterdam: Elsevier: pp. 139–160.

Pires JC, Lim KY, Kovařík A, Matyasek R, Boyd A, Leitch AR, Leitch IJ et al. (2004) Molecular cytogenetic analysis of recently evolved *Tragopogon* (Asteraceae) allopolyploids reveal a karyotype that is additive of the diploid progenitors. *Am. J. Bot.* 91: 1022–1035.

Plohl M, Luchetti A, Meštrovic N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin. *Gene* 409: 72–82.

Plohl M, Meštrovic N, Mravinac B (2012) Satellite DNA evolution. *Genome Dyn.* 7: 126–152.

Plohl M, Meštrovic N, Mravinac B (2014) Centromere identity from the DNA point of view. *Chromosoma* 123: 313–325.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends Ecol Evol.* 16(7): 351–358.

Roa F, Guerra M (2012) Distribution of 45S rDNA sites in chromosomes of plants: structural and evolutionary implications. *BMC Evol. Biol.* 12: 225.

Roa F, Guerra M (2015) Non-random distribution of 5S rDNA sites and its association with 45S rDNA in plant chromosomes. *Cytogenet. Genome Res.* 146(3): 243–249.

Sandery MJ, Forster JW, Blunden R, Jones N (1990) Identification of a family of repeated sequences on the rye B chromosome. *Genome* 33: 908–913.

Sanmiguel P, Bennetzen JL (1998) Evidence that a recent increase in maize genome size was caused by the massive amplification of intergene retrotransposons. *Ann. Bot.* 82: 37–44.

Scaldaferro MA, da Cruz MV, Cecchini NM, Moscone EA (2016) FISH and AgNor mapping of the 45S and 5S rRNA genes in wild and cultivated species of *Capsicum* (Solananceae). *Genome* 59(2): 95–113.

Scaldaferro MA, Grabiele M, Moscone EA (2013) Heterochromatin type, amount and distribution in wild species of chili peppers (*Capsicum*, Solanaceae). *Genet. Resour. Crop Evol.* 60: 693–709

Scaldaferro MA, Seijo JG, Acosta MC, Barboza GE, Ducasse DA, Moscone EA (2006) Genomic characterization of the germplasm in peppers (*Capsicum* – Solanaceae) by fluorescent *in situ* hybridization. *Plant Sci.* 43: 291–297.

Schmidt T, Heslop-Harrison JS (1998) Genomes, genes and junk: the large-scale organization of plant chromosomes. *Trends Plant Sci.* 3: 195–199.

Schnable PS, Ware D, Fulton RS, Stein JC, Wei F, Pasternak S, Liang C et al. (2009) The B73 maize genome: complexity, diversity, and dynamics. *Science* 326(5956): 1112–1115.

Sharma A, Wolfgruber TK, Presting GG (2013) Tandem repeats derived from centromeric retrotransposons. *BMC Genom.* 14: 142.

Simon AU (2020) Evolution of selected satellite DNAs in cultivated *Capsicum* species. M.Sc. thesis, Department of Botany and Biodiversity Research, Faculty of Life Sciences, University of Vienna, Vienna, Austria.

Skalicka K, Lim KY, Matyasek R, Matzke M, Leitch AR, Kovařík A (2005) Preferential elimination of repeated DNA sequences from the paternal, *Nicotiana tomentosiformis* genome donor of a synthetic, allotetraploid tobacco. *New Phytol*. 166: 291–303.

Soltis DE, Albert VA, Leebens-Mack J, Bell CD, Paterson AH, Zheng CF, Sankoff D et al. (2009) Polyploidy and angiosperm diversification. *Am. J. Bot.* 96: 336–348.

Sonnhammer ELL, Durbin R (1995) A dot-matrix program with dynamic threshold control suited for genomic DNA and protein sequence analysis. *Gene* 167: GC1-10.

Staton SE, Bakken BH, Blackman BK, Chapman MA, Kane NC, Tang S, Ungerer MC et al. (2012) The sunflower (*Helianthus annuus* L.) genome reflects a recent history of biased accumulation of transposable elements. *Plant J.* 72: 142–153.

Stebbins GL (1971) *Chromosomal evolution in higher plants.* Edward Arnold, London.

Stewart C, Kang BC, Liu K, Mazourek M, Moore SL, Yoo EY, Kim BD et al. (2005) The Pun1 gene for pungency in pepper encodes aputative acyltransferase. *Plant J.* 42: 675–688.

Stewart C, Mazourek M, Stellari GM, O'Connell M, Jahn M (2007) Genetic control of pungency in *C. chinense* via the Pun1 locus. *J.Exp. Bot.* 58: 979–991.

Surh YJ (2002) More than spice: capsaicin in hot chili peppers makes tumor cells commit suicide. *J. Natl. Cancer Inst.* 94: 1263–1265.

Temsch EM (2003) Genome size variation in plants with special reference to the genus *Arachis* (Fabaceae), and the methodological improvement by application of a new optical immersion-gel in flow cytometry. Ph.D. Thesis, Department of Systematic and Evolutionary Botany, University of Vienna, Vienna, Austria.

Temsch EM, Greilhuber J, Krisai R (2010) Genome size in liverworts. *Preslia* 82: 63–80.

The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408: 796–815.

Uhl CH (1978) Chromosomes of Mexican Sedum II. Section *Pachysedum*. *Rhodora* 80: 491–512.

Van de Peer Y, Fawcett JA, Proost S, Sterck L, Vandepoele K (2009)The flowering world: a tale of duplications. *Trends Plant Sci.* 14: 680–688.

Vanzela ALL, Guerra M, Luceno M (1996) *Rhynchospora tenuis* Link (Cyperaceae), a species with the lowest number of holocentric chromosomes. *Cytobios* 88: 219–228.

Varagona MJ, Purugganan M, Wessler SR (1992) Alternative splicing induced by insertion of retrotransposons into the maize waxy gene. *Plant Cell* 4: 811–820.

Vitte C, Fustier MA, Alix K, Tenaillon MI (2014) The bright side of transposons in crop evolution. *Brief. Funct. Genomics* 13: 276–295.

Volkov RA, Medina FJ, Zentgraf U, Hemleben V (2004) Organization and molecular evolution of rDNA, nucleolar dominance, and nucleolus structure. *Prog. Bot.* 65:106–146.

Weiss-Schneeweiss H, Blöch C, Turner B, Villaseñor JL, Stuessy TF, Schneeweiss GM (2012) The promiscuous and the chaste: frequent allopolyploid speciation and its genomic consequences in American daisies (*Melampodium sect. Melampodium*; Asteraceae). Evolution. 66(1):211–28.

Wiss-Schneeweiss H, Leitch AR, McCann J, Jang T-S, Macas J (2015) Employing next generation sequencing to explore the repeat landscape of the plant genome. In: *Next Generation Sequencing in Plant Systematics.* Hörandl E & Appelhans M, eds. Regnum Vegetabile 157 (ISSN 0080-0694). Koeltz Scientific Books, Königstein, Germany.

Weiss-Schneeweiss H, Schneeweiss GM (2013) Karyotype diversity and evolutionary trends in angiosperms. In: *Plant Genome Diversity Volume 2*. Greilhuber J., Dolezel J., Wendel J, eds. Springer, Vienna.

Wells JN, Feschotte C (2020) A field guide to eukaryotic transposable elements. *Annu. Rev. Genet.* 54: 539–561.

Westerterp-Plantenga MS, Smeets A, Lejeune MPG (2005) Sensory and gastrointestinal satiety effects of capsaicin on food intake. *Int. J. Obes.* 29: 682–688.

Wicker T, Keller B (2007) Genome-wide comparative analysis of *copia* retrotransposons in Triticeae, rice, and *Arabidopsis* reveals conserved ancient evolutionary lineages and distinct dynamics of individual *copia* families. *Genome Res.* 17:1072–1081.

Wright SI, Agrawal N, Bureau TE (2003) Effects of recombination rate and gene density on transposable element distributions in *Arabidopsis thaliana*. *Genome Res.* 13: 1897–1903.

Abstract

Five of nearly 40 chile pepper species are important vegetable and spice crops consumed worldwide. Given the economic importance of the genus, little is known about the genome and its composition especially of wild *Capsicum* species. All pepper species are diploids, but their genome sizes vary significantly, implying active role of repetitive DNAs in genome evolution of the genus. Satellite DNAs and rDNAs in particular are widely used as chromosomal markers to gain better insight into the evolution of karyotypes and genomes in many plant groups.

This study aimed to characterize karyotype and repetitive DNA fraction of one of the wild species of the genus *Capsicum*, *C. flexuosum*, possessing the largest genome among all chile pepper species. Genome skimming and RepeatExplorer pipeline were used to characterize the repeatome profile of *C. flexuosum*. The genome of this species was found to be mostly populated by Ty3-*gypsy* retrotransposons, similarly to other Solanaceae genera. Seven satellite DNAs were identified making up 7% of the genome, two of which were novel and unique to this taxon.

Two 35S and 5S rDNAs as well as ten earlier identified satellite DNAs were mapped in the chromosomes of *C. flexuosum* using fluorescence *in situ* hybridization. Most of these tandem repeats showed some level of variation and polymorphisms, with the exception of 5S rDNA. Most of the satellite DNAs were mapped within the heterochromatic blocks. The mapping of all twelve repeats provided combinations of probes for unequivocal identification of all twelve chromosome pairs in *C. flexuosum*.

Zusammenfassung

Fünf von fast 40 Chili Arten sind wichtige Gemüse und Gewürzpflanzen die weltweit konsumiert werden. Trotz der wirtschaftlichen Wichtigkeit der Gattung ist relativ wenig über die Komposition des Genoms von speziell wilden *Capsicum* Arten bekannt. Alle *Capsicum* Arten sind diploid, ihre Genomgröße variiert jedoch stark, was auf eine aktive Rolle von repetitiver DNA in der Evolution der Gattung hindeutet. Satelliten DNAs und besonders rDNAs werden oft als chromosomale Marker genutzt um einen Einblick in die Evolution des Karyotyps und Genoms von vielen Pflanzengruppen zu gewinnen.

Das Ziel dieser Arbeit ist die Charakterisierung des Karyotyps und der repetitiven DNA Fraktion einer wilden *Capsicum* Art, *C. flexuosum*, welche das größte Genom von allen *Capsicum* Arten vorweist. Genome-Skimming und die RepeatExplorer Pipeline wurden verwendet um das Repeatom von *C. flexuosum* zu charakterisieren. Großteils bestand das Genom aus Ty3-*gypsy* Retrotransposons, ähnlich wie in anderen Solanaceae Gattungen. Sieben Satelliten DNAs, welche 7% des Genoms ausmachten, wurden identifiziert. Zwei von diesen waren neu und einzigartig für das Taxon.

Zwei 35S und 5S rDNAs und zehn früher identifizierte Satelliten DNAs wurden auf den Chromosomen von *C. flexuosum* durch fluorescence *in situ* hybridization gefunden. Die meisten Tandem Repeats variierten zu teilen und wiesen Polymorphismen auf, mit der Ausnahme von 5S rDNA. Die meisten Satelliten DNAs wurden in Heterochromatin Blöcken gefunden. Alle zwölf Chromosomenpaare von *C. flexuosum* konnten durch die Kartierung der zwölf Repeats eindeutig identifiziert werden.