

# DISSERTATION



Structure, organization, and evolution of  
satellite DNAs in species of the genera  
*Beta* and *Patellifolia*

DISSERTATION

Zur Erlangung des Doktorgrades (Dr. Rer. Nat.)

vorgelegt

der Fakultät Mathematik und Naturwissenschaften  
der Technischen Universität Dresden

von

M.Sc. Bich Hong Ha

Dresden 2018



## **1. Gutachter**

Prof. Dr. Thomas Schmidt

Lehrstuhl für Zell- und Molekularbiologie der Pflanzen

Fachrichtung Biologie

Technische Universität Dresden

Dresden

## **2. Gutachter**

Prof. Dr. Andreas Houben

Arbeitsgruppe Chromosomenstruktur und -funktion

Züchtungsforschung

Leibniz-Institut für Pflanzengenetik und Kulturpflanzenforschung

Gatersleben

## **Acknowledgement**

Firstly, I thank the scholarship program of Vietnam Government and the Graduate Academy of TUD for the financial supports.

I would like to thank my supervisor, Prof. Dr. Thomas Schmidt, for the opportunity to work in his research group at Institute of Botany, TUD, providing me all the facilities to access the work, and the valuable discussions. Without his support, I would not have enough strength to finish my thesis.

I am grateful for Prof. Dr. Andreas Houben agreeing to review my dissertation.

I would like to emphasize Dr. Tony Heitkam, who helped me step by step to be familiar with bioinformatics. She shared me about far-reaching topics could run, broaden my knowledge, and gave me valuable tips for my experiments. Her patience in reading and providing detailed feedback on my dissertation was outstanding.

My special thanks are to Dr. Beatrice Weber. She is my worthy supervisor who never minds spending time and giving me guidance for all my experiments as well as correction of my writing.

I would like to thank Dr. Falk Zakrzewski for the initial guidance in satellite DNA analysis and his support in my first presentation at the research group. I appreciate Dr. Gerhard Menzel for his discussion in the meetings.

My grateful thanks go to the technicians Susann Liedtke, Nadin Fliegner and Ines Walter for the technical help.

I also thank all colleagues in the research group for the stimulating, constructive atmosphere and for the fruitful cooperation.

Last but not least, I want to express my thankfulness to my family, my husband and my little daughter, my friends, who always support me by their love and beliefs.

Vielen herzlichen Dank!

Xin chân thành cảm ơn!

## TABLE OF CONTENTS

Acknowledgement .....	ii
TABLE OF CONTENTS .....	iii
LIST OF FIGURES .....	vii
LIST OF TABLES .....	xi
Abbreviations .....	xiii
1 Introduction .....	1
1.1 Repetitive DNAs in plant genomes .....	1
1.2 Satellite DNAs .....	4
1.3 Genomic analysis in <i>Beta</i> and <i>Patellifolia</i> species .....	7
1.3.1 Genus <i>Beta</i> and <i>Patellifolia</i> .....	7
1.3.2 Repeated DNA analysis in <i>Beta</i> and <i>Patellifolia</i> species .....	9
1.3.3 Fluorescent <i>in situ</i> hybridization in chromosome analysis .....	10
1.4 Aim of this thesis .....	13
2 Material and Methods .....	14
2.1 Material .....	14
2.1.1 Plant material .....	14
2.1.2 Chemicals and consumables .....	14
2.1.3 Culture media and antibiotics .....	18
2.1.4 Solutions and Buffers .....	19
2.1.5 Oligonucleotides .....	20

2.1.6 Sequence data bases .....	22
2.1.7 Software .....	23
2.2 Methods.....	24
2.2.1 Bioinformatics.....	24
2.2.1.1 Identification of putative satellite families and monomer size .....	24
2.2.1.2 Sequences analysis.....	25
2.2.1.3 Similarity search .....	25
2.2.1.4 Phylogenetic tree.....	25
2.2.2 Molecular methods.....	26
2.2.2.1 Isolation of DNA.....	26
2.2.2.2 Agarose gel electrophoresis .....	27
2.2.2.3 Polymerase chain reaction .....	27
2.2.2.4 Cloning.....	27
2.2.3 Molecular cytogenetic methods .....	32
2.2.3.1 Preparation of plant chromosomes.....	32
2.2.3.2 Labeling of DNA probes for FISH .....	33
2.2.3.3 Fluorescent <i>in situ</i> hybridization (FISH) .....	34
2.2.3.4 UV microscopy, photography and image processing .....	37
3 Results.....	38
3.1 Comparative identification of repetitive DNA in sugar beet and related wild beet species .....	38
3.2 Satellite landscape in <i>Beta lomatogona</i> .....	44



3.2.1 Generation of a satellite overview in <i>Beta lomatogona</i> .....	44
3.2.2 The satellite families B1Sat1, B1Sat5 and B1Sat6 form large, chromosome-specific arrays .....	47
3.2.2.1 Monomeric structure and organization based on bioinformatics and multimer cloning .....	47
3.2.2.2 Genomic organization of B1Sat1, B1Sat5 and B1Sat6 in <i>Beta lomatogona</i> .....	56
3.2.2.3 Methylation of B1Sat1, B1Sat5 and B1Sat6 satellites .....	58
3.2.2.4 Localization of B1Sat1, B1Sat5 and B1Sat6 along <i>Beta lomatogona</i> chromosomes.....	60
3.2.2.5 B1Sat1, B1Sat5 and B1Sat6 are assigned to <i>Beta lomatogona</i> chromosomes 3, 5, 6, 7, 8 and 9 .....	61
3.2.2.6 Distribution of <i>B. lomatogona</i> satellite B1Sat1, B1Sat5 and B1Sat6 in the genera <i>Beta</i> , <i>Patellifolia</i> , and related species .....	65
3.2.2.7 Sequence divergence of <i>B. lomatogona</i> satellite B1Sat1, B1Sat5 and B1Sat6 in the genus <i>Beta</i> .....	68
3.2.2.8 Comparative chromosomal localization of the satellite families B1Sat1, B1Sat5 along <i>Beta</i> chromosomes.....	75
3.2.3 The tandem repeats B1Sat2, B1Sat3 and B1Sat4 are sequence domains of retrotransposons.....	80
3.2.3.1 B1Sat2, B1Sat3 and B1Sat4 are associated with retrotransposons and show unconventional satellite features .....	80
3.2.3.2 Molecular structure and organization of satellite repeats in retrotransposons .....	81
3.2.3.2 Genomic organization of B1Sat2, B1Sat3, and B1Sat4 in <i>B. lomatogona</i> .....	89
3.2.3.3 Chromosomal localization of B1Sat2, B1Sat3 and B1Sat4 along <i>B. lomatogona</i> chromosomes .....	90

3.2.3.5 Distribution and sequence divergence of <i>B. lomatogona</i> tandem repeats B1Sat2, B1Sat3 and B1Sat4 in <i>Beta</i> species .....	92
3.2.4 Characterization of satellites pBC1418 and pHT36 in the <i>B. lomatogona</i> genome .....	97
3.2.5 <i>Beta lomatogona</i> satellite overview .....	101
3.3 Satellite DNAs in the genus <i>Patellifolia</i> .....	102
3.3.1 Comparative analysis of repetitive DNA sequences of <i>Patellifolia procumbens</i> and <i>Patellifolia patellaris</i> .....	103
3.3.2 Characterization of minisatellite PproSat1 and satellite PpatSat1 .....	106
3.3.2.2 Genome organization and satellite distribution in <i>Patellifolia</i> and related species ...	110
3.3.2.3 Chromosomal localization in <i>Patellifolia</i> species.....	115
4. Discussion .....	117
4.1 Repetitive DNAs are a major component of plant genomes.....	118
4.2 Abundance of satellite DNAs in <i>Beta</i> genomes.....	121
4.2.1 Organization of the conventional satellite families B1Sat1, B1Sat5, and B1Sat6 .....	121
4.2.2 The satellite families B1Sat2, B1Sat3 and B1Sat4 as tandem repeat domains in Ogre/Tat retrotransposons .....	129
4.3 Comparative analysis of genome composition revealed the relationship between <i>Patellifolia procumbens</i> and <i>Patellifolia patellaris</i> .....	131
4.4 Evolution of the satellite families in genus <i>Beta</i> and related species .....	134
5. Further work.....	141
6. Summary .....	142
7. Literatures .....	144
Supplementary figures .....	163

## LIST OF FIGURES

Figure 3.1: Repeat compositions of <i>Beta</i> , <i>Patellifolia</i> and <i>C. quinoa</i> species analyzed by mean of <i>RepeatExplorer</i> .....	39
Figure 3.2: Percentage in the satellite fractions of <i>Beta</i> , <i>Patellifolia</i> and <i>C. quinoa</i> species analyzed by <i>RepeatExplorer</i> .....	42
Figure 3.3: Circular and star-like graph shape and dotplot of putative satellite clusters. ....	47
Figure 3.4: Monomer consensus sequences of three conventional satellite families in the <i>B. lomatogona</i> genome .....	49
Figure 3.5: Ladder-like pattern in agarose gel after PCR with satellite-specific primers .....	50
Figure 3.6: Higher order structure of B1Sat5 satellite from analyzed clones .....	52
Figure 3.7: Sequence alignment of two subunits of satellite B1Sat5 .....	54
Figure 3.8: Dendrogram representing the phylogenetic relationship of the subunits sub1 and sub2 of the satellite family B1Sat5 .....	55
Figure 3.9: Genomic organization of <i>B. lomatogona</i> tandemly repeated sequences .....	56
Figure 3.10: Schematic restriction map of the two consecutive B1Sat5 monomers .....	58
Figure 3.11: Southern hybridization of three satellites to <i>B. lomatogona</i> genomic DNA restricted with the methylation sensitive enzymes <i>HpaII</i> and <i>MspI</i> .....	59
Figure 3.12: Chromosomal localization of classic satellites B1Sat1, B1Sat5 and B1Sat6 along chromosomes of <i>B. lomatogona</i> .....	61
Figure 3.13: Assignment of satellite DNA patterns to chromosomes by multicolor-FISH .....	63
Figure 3.14: Arrangement of all <i>B. lomatogona</i> chromosomes combined with B1Sat1, B1Sat5 and B1Sat6 satellites .....	64
Figure 3.15: Schematic karyogram of <i>B. lomatogona</i> .....	65
Figure 3.16: Abundance and genomic organization of three <i>B. lomatogona</i> satellites in genus <i>Beta</i> and <i>Patellifolia</i> .....	67

Figure 3.17: PCR with satellite-specific primers of B1Sat1, B1Sat5, and B1Sat6 .....	68
Figure 3.18: PCR from different <i>Beta</i> , <i>Patellifolia</i> and related species .....	70
Figure 3.19A: Dendrogram representation of relationship between B1Sat1 sequences in <i>Beta</i> species .....	73
Figure 3.19B: Dendrogram representation of relationship between B1Sat5 sequences in <i>Beta</i> species .....	74
Figure 3.19C: Dendrogram representation of relationship between B1Sat6 sequences in <i>Beta</i> species .....	75
Figure 3.20: Different signal patterns among <i>Beta</i> species of satellite B1Sat1 .....	77
Figure 3.21: Chromosomal localization of B1Sat5 in <i>Beta</i> sections <i>Corollinae</i> and <i>Nanae</i> ....	79
Figure 3.22: Supercluster with Ogre/Tat retrotransposon sequences containing the satellite families B1Sat2 and B1Sat3 .....	81
Figure 3.23: Circular/star-like graph shape and dotplot of tandem repeat clusters. ....	82
Figure 3.24: Bioinformatic monomer sequences of three satellite families in the <i>B. lomatogona</i> genome .....	84
Figure 3.25: Ladder-like pattern in agarose gel after PCR with satellite-specific primers of B1Sat2, B1Sat3 and B1Sat4 .....	85
Figure 3.26: Alignment of B1Sat3 subunits sub1 and sub2.....	88
Figure 3.27: Dendrogram representing structural relationship between subunit 1 (sub1) and subunit 2 (sub2) of B1Sat3 monomers.....	89
Figure 3.28: Genomic organization of <i>B. lomatogona</i> tandemly repeated sequences.....	90
Figure 3.29: Chromosomal localization of tandem repeats along chromosomes of <i>B. lomatogona</i> .....	91
Figure 3.30: PCR from different <i>Beta</i> and related species .....	93

Figure 3.31: Abundance and genomic organization of Blsat4 satellite family in genus <i>Beta</i> and <i>Patellifolia</i> .....	94
Figure 3.32: Dendrogram representating the relationship between BlSat4 sequence in <i>Beta</i> species.....	97
Figure 3.33: Monomer consensus of pBC1418 (A) and pHT36 (B) in <i>B. lomatogona</i> .....	99
Figure 3.34: Tandem organization of two satellite families pBC1418 and pHT36. ....	100
Figure 3.35: Chromosomal localization of the satellite pBC1418 and pHT36 along chromosomes of <i>B. lomatogona</i> .....	101
Figure 3.36: Comparative analysis of two species <i>P. procumbens</i> and <i>P. patellaris</i> based on the cluster fraction .....	103
Figure 3.37: Bioinformatic identification of monomer sequences of PproSat1 and PpatSat1 .....	106
Figure 3.38: PCR from (A) <i>P. procumbens</i> and (B) <i>P. patellaris</i> DNA for verifacation of the tandem repeat organization.....	107
Figure 3.39: Sequence alignment of (A) PproSat1 and (B) PpatSat1 monomers.....	109
Figure 3.40: Garden PCR of (A) PproSat1 and (B) PpatSat1 .....	110
Figure 3.41: Autoradiograph of Southern hybridization shows the abundance and organization of (A) minisatellite PproSat1 and (B) satellite family PpatSat1 in <i>Patellifolia</i> and related species.....	112
Figure 3.42: Sequence alignment of PproSat1 from <i>P. procumbens</i> , CqSat1 from <i>C. quinoa</i> , and pBC1447 from <i>B. corolliflora</i> . ....	113
Figure 3.43: Neighbor-joining clustering analysis of three minisatellite monomer sequences in <i>B. patula</i> , <i>P. procumbens</i> , and <i>C. quinoa</i> .....	114
Figure 3.44: Chromosomal localization of <i>P. procumbens</i> -specific satellite PproSat1 .....	115
Figure 3.45: Chromosomal localization of PpatSat1 satellite on (A) metaphase chromosomes of <i>P. patellaris</i> and (B) prometaphase chromosomes of <i>P. procumbens</i> . ....	116

Figure 4.1: Schematic representation of a possible evolution of the satellite families B1Sat05  
..... 125

Figure 4.2: Dendrogram presents the evolution of novel identified satellites as well as the  
main known satellites in two subfamilies *Betoideae* (left) and *Chenopodiaceae* (right) ..... 135

## LIST OF TABLES

Table 1.1: Taxonomy of genus <i>Beta</i> and <i>Patellifolia</i> .....	7
Table 1.2: Summary of satellite DNA families in genera <i>Beta</i> and <i>Patellifolia</i> . .....	11
Table 2.1: Species included in the analyses .....	14
Table 2.2: Chemicals and consumables.....	15
Table 2.3: Enzymes .....	17
Table 2.4: Kits .....	18
Table 2.5: Used oligonucleotides .....	21
Table 2.6: BACs used as probes for detection of chromosome in multi-color FISH analysis.	22
Table 2.7: DNA probes used for FISH and Southern hybridization .....	22
Table 2.8: Features of paired-end Illumina libraries from studied species .....	23
Table 2.9: Utilized Software and Internet services .....	23
Table 2.10: Maximum of excitation and emission of the used fluorochromes .....	37
Table 3.1: Sequence reads from reference species used for <i>RepeatExplorer</i> analysis.....	38
Table 3.2: Genome proportion of repetitive sequences in <i>Beta</i> , <i>Patellifolia</i> , and <i>C. quinoa</i> species analyzed by <i>RepeatExplorer</i> .....	40
Table 3.3: Satellite families in <i>Beta</i> , <i>Patellifolia</i> , and <i>C. quinoa</i> species analyzed by <i>RepeatExplorer</i> .....	43
Table 3.4: Satellites in the <i>B. lomatogona</i> genome. ....	45
Table 3.5: Features of novel satellite DNA families in <i>B. lomatogona</i> genome.....	46
Table 3.6: Clones of three satellites B1Sat1, B1Sat5, and B1Sat6 in <i>B. lomatogona</i> . ....	51

Table 3.7: Chromosomal distribution of B1Sat1, B1Sat5 and B1Sat6 along <i>B. lomatogona</i> chromosomes .....	65
Table 3.8: Variation of typical <i>B. lomatogona</i> satellite monomers in the genus <i>Beta</i> .....	71
Table 3.9: Summary of chromosomal distribution of B1Sat1 and B1Sat5 along <i>Beta</i> chromosomes .....	77
Table 3.10: Clones of three tandem repeats B1Sat2, B1Sat3, and B1Sat4 in <i>B. lomatogona</i> ....	86
Table 3.11: Variation of retrotransposon-associated <i>B. lomatogona</i> satellite monomers in the genus <i>Beta</i> .....	95
Figure 3.32: Dendrogram representating the relationship between B1Sat4 sequence in <i>Beta</i> species .....	97
Table 3.12: Mapping of six identified satellite sequences to read sequences of <i>Beta</i> and related species .....	102
Table 3.13: Species-enriched and -specific clusters from comparison of <i>P. procumbens</i> and <i>P. patellaris</i> genomes .....	105
Table 3.14: Clones of two satellite PproSat1 in <i>P. procumbens</i> and PpatSat1 in <i>P. patellaris</i>	108
Table 3.15: Chromosomal distribution of PproSat1 and PpatSat1 along <i>Patellifolia</i> chromosomes .....	116



## Abbreviations

2n	Diploid chromosome set
DNA	Deoxyribonucleic acid
rDNA	Ribosomal DNA
RNA	Ribonucleic acid
BLAST	Basic local alignment search tool
bp	Base pair
BAC	Bacterial artificial chromosome
BSA	Bovine serum albumine
Cy3	Cyanine3 fluorochrome
DAPI	4', 6-Diamidino-2-phenylindole
dATP	Deoxyadenosine triphosphate
dCTP	Deoxycytidine triphosphate
dGTP	Deoxyguanosine triphosphate
dTTP	Deoxythymidine triphosphate
dUTP	Deoxyuridine triphosphate
e.g.	Exempli gratia (for example)
EDTA	Ethylenediaminetetraacetic acid
FISH	Fluorescent <i>in situ</i> hybridization
FITC	Fluorecein isothiocyanate
GISH	Genomic <i>in situ</i> hybridization
HOR	Higher order repeat
i.e.	Id est (that is)
IPTG	Isopropyl $\beta$ -D-1-thiogalactopyranoside
kb	Kilobase
LTR	Long terminal repeat
Mb	Megabase
Mya	Million years ago
MUSCLE	Multiple sequence comparison by log-expectation
NOR	Nuclelus organizer region
PCR	Polymerase chain reaction
PVP	Polyvinylpyrrolidone
SDS	Sodium dodecyl sulfate
SSC	Saline sodium citrate buffer
TAE	Tris-Acetate-EDTA
X-gal	5-Bromo-4.chloro-3-indolyl- $\beta$ -galactopyranoside



## 1 Introduction

### 1.1 Repetitive DNAs in plant genomes

Plant nuclear genomes vary greatly in their sizes even between closely related species. The genome size (1C value) of an organism is the amount of DNA in a haploid set of chromosome. The C-value paradox indicated that there is no correlation between the genome size and the genome complexity in eukaryotic organisms (Thomas, 1971). The range of genome sizes in the angiosperms varies 2000-fold from 63 Mbp in *Genlisea tuberosa* (Fleischmann *et al.*, 2014) to 158,000 Mbp in *Paris japonica* (Pellicer *et al.*, 2010).

In angiosperms, most of the observed large-scale genome size variation is not due to different gene numbers or gene sizes (Bennetzen *et al.*, 2005). Genome duplication and polyploidization events have been seen as a major reason strongly influencing genome sizes. Leitch and Bennett (2004) postulated three patterns for changes in DNA amount in related species following polyploidy: (1) most polyploids show additivity in DNA amount relative to diploids, (2) many polyploids indicate a reduction in DNA amount relative to diploids and (3) a few polyploids reveal an increase in DNA amount relative to the respective diploids. However, molecular investigations of the plant nuclear DNA content have shown that most genome size variability is associated with differences in repetitive DNA content (Flavell *et al.*, 1974; Bennetzen *et al.*, 2005).

In polyploids and interspecific hybrids, the repetitive DNA can be eliminated, as shown for rDNA in tobacco (Volkov *et al.*, 1999). It has been reported that sequence elimination is one of the major and immediate responses of the genome to wide hybridization or allopolyploidy in wheat (Shaked *et al.*, 2001). This lead to a significant reduction of the genome size in allopolyploids in comparison to the expected value (Ozkan *et al.*, 2003). Genome downsizing is typically mediated by unequal recombination (Devos *et al.*, 2002). Loss of DNA in polyploids is a widespread phenomenon occurring in angiosperms (Leitch and Bennett, 2004).

The contribution of repetitive DNAs to genome size was reported for many plants and accumulation and proliferation of transposable elements are largely responsible for the different sizes of plant genomes (Sanmiguel *et al.*, 1996; Heslop-Harrison, 2000; Hawkins *et al.*, 2006; Vitte and Bennetzen, 2006). For example, the Norway spruce (*Picea abies*): despite having a similar amount of genes (28,354) like the model plant *Arabidopsis thaliana* (27,407), the Norway spruce

genome spanning 19,600 Mbp is approximately 127 times larger than the 154 Mbp *A. thaliana* genome. There is no evidence for a recent whole-genome duplication, hence the large genome size of Norway spruce is the result of accumulating retrotransposons as one major class of repetitive DNA (Nystedt *et al.*, 2013). Sugar pine (*Pinus lambertiana* Dougl) with 31,000 Mbp genome size is even 201 times bigger than that of *Arabidopsis thaliana* and here Ty3-gypsy LTR retrotransposons are considered as a main reason for the genome expansion (Steven *et al.*, 2016).

Repetitive DNAs are homologous sequence motifs of different size (2 to > 10,000 bp) which are organized in a repeated manner in the genome as tandemly arranged structures or dispersed localized sequences with most of the genes between these repeat blocks (Schmidt and Heslop-Harrison, 1998).

Transposable elements (TEs) are dispersed sequences, which are able to move from one location in the genome to another. TEs can be divided into two classes according to their mechanisms of transposition. Class I elements (retrotransposons) transpose via “copy and paste” mechanisms, where the element is transcribed into RNA and then upon reverse transcriptase enzymes the RNA sequences back into DNA, which is then reinserted into the target site (Joly-Lopez and Bureau, 2014). Retrotransposons can be further divided into two categories: long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons. LTR retrotransposons, including Ty3-copia and Ty3-gypsy retrotransposons, make up the majority of the transposable element classes in most plants (Pearce *et al.*, 1996; Weber *et al.*, 2010; Steven *et al.*, 2016). Non-LTR retrotransposons include long interspersed elements (LINEs) and short interspersed elements (SINEs). Class II elements (DNA transposons) transpose via “cut and paste” mechanisms, where the element transposes directly from DNA to DNA mediated by a transposon-encoded transposase (Joly-Lopez and Bureau, 2014). DNA transposons are grouped into superfamilies such as Tc1/mariner, hAT, and Mutator (Feschotte *et al.*, 2002). Additionally, both class I and class II TEs can be either autonomous or non-autonomous. Autonomous TEs can move on their own, while non-autonomous elements require the enzymes of other TEs in order to move. Examples of non-autonomous elements are SINE and MITEs (Miniature Inverted-repeat Transposable Elements).

Tandem repeats include satellite DNA consisting of numerous tandemly arranged repeats that are non-coding and mostly located in heterochromatic regions, micro- and minisatellites, telomeric repeats and ribosomal genes. Telomeric repeats are important tandem repeats which are localized

at the chromosomal ends of nearly all eukaryotic species to stabilize the chromosomes. This physical termination is formed in many plants by the highly conserved DNA sequence with the consensus (TTTAGGG)<sub>n</sub> to prevent the loss of terminal nucleotides during replication (Martinez and Blasco, 2011). The number of telomeric repeats is species-specific. Telomeric repeats extend over a length of 2-5 kb in *Arabidopsis* (Richard and Ausubel, 1988) while they are arranged over 60-160 kb in tobacco (Fajkus *et al.*, 1995) and between 10-18 kb in *G. pygmaea* (Tran *et al.*, 2015). Another functional and very important tandem repeat is the ribosomal DNA (rDNA or rRNA genes). These DNA sequences code for rRNAs that are in addition to several proteins, forming the ribosomes. In interphase nuclei, the rDNA sites are often visible as nucleoli which is a reason for their name nucleolus organizing regions (NOR) on one or more chromosomes (Lopez-Flores *et al.*, 2012). In all angiosperms, the rDNA is divided into two groups - the 5S rDNA and the 18S-5.8S-28S rDNA. The genes for the 18S-5.8S-25S rRNA are arranged as a transcription unit, consisting of highly conserved, coding sections and variable spacer regions arranged in a tandem manner (Hemleben and Zentgraf, 1994). The number of the 18S-5.8S-25S rRNA repeat units' ranges from 39 to 19,300 in animals and from 150 to 26,000 in plants (Prokopowich *et al.*, 2003). The 5S rRNA genes are spatially separated from the 18S-5.8S-25S rRNA loci. The tandemly arranged units of 200-900 bp consisting of a coding region of 120 bp and variable spacers are present in a number of 1000-50000 copies per haploid genome (Ellis *et al.*, 1988; Hemleben and Zentgraf, 1994). Garcia *et al.* (2017) analyzed the Plant rDNA database and showed that 18S-5.8S-28S rDNA loci mostly locate at terminal regions while 5S rDNA loci predominantly locate at interstitial or pericentromeric regions of chromosomes. Due to highly conservation in eukaryotic species, rDNA sequences are ideal probes usage for a wide spectrum of plant species.

Repetitive sequences are included in numerous processes and therefore play an important role in eukaryotic genomes. Repetitive DNA sequences are present in heterochromatin regions, likely contributed to the chromosome movement and pairing (Fajkus *et al.*, 2008), chromosome recombination (Nagy and Bennetzen, 2008), interaction of chromatin protein (Melters *et al.*, 2013; Rosic *et al.*, 2014; Kowar *et al.*, 2016), chromosome structural determination (Lopez-Flores, 2012; Plohl, 2014), and karyotypic evolution (Leitch and Bennett, 2004; Kelly *et al.*, 2015; Steven *et al.*, 2016). Furthermore, these sequences also implicated in epigenetic regulatory processes via siRNA

(Martienssen *et al.*, 2003; Ugarkovic, 2005) or cytosine methylation (Zakrzewski *et al.*, 2014, 2017).

## 1.2 Satellite DNAs

Satellite DNA is a non-coding genome component, which is arranged as long arrays of its underlying repeat units (monomers) in a tandem manner. For a long time, the function of satellites within the genome was unknown and satellite DNA was classified as 'junk DNA' (Graur *et al.*, 2015). However, recent researches demonstrated that satellite DNA is a major component of genomes occurring in essential chromosomal domains, such as the centromere, the intercalary heterochromatin, and in subtelomeric chromosome regions (Heslop-Harrison *et al.*, 1999; Melters *et al.*, 2013; Zakrzewski *et al.*, 2014).

The typical repeating unit of satellites is either 150-180 bp or 320-360 bp (Heslop-Harrison, 2000). These particular lengths seem to correlate with the size of a single nucleosome requiring approximately 146 bp of DNA to form the two turns around each nucleosome plus 25-30 bp of the linker DNA (Manuelidis and Chen, 1990), which may represent a preferred condition of satellite organization within an optimized chromatin structure (Schmidt and Heslop-Harrison, 1998; Jiang *et al.*, 2003; Plohl *et al.*, 2008; Zhang *et al.*, 2013; Melters *et al.*, 2013; Weiss-Schneeweiss *et al.*, 2015). Historically, satellites can be subdivided into microsatellites ( $\leq 10$  bp), minisatellites (10-100 bp), and conventional satellites (150-180 bp or 300-360 bp) (Mehrotra and Goyal, 2014). Today, the term "satellite DNA" is applied to any tandem repeats which are organized in long arrays (hundreds to thousands of repeat units) in the heterochromatin (Garrido-Ramos, 2017). For example, CajaSat1 satellite with monomer length of 43 bp constituting more than 11 % of the *Camellia japonica* genome forms very large arrays (Heitkam *et al.*, 2015). The term 'satellitome' has been proposed as the whole collection of different satellite DNA families in a genome (Ruiz-Ruano *et al.*, 2016).

As consequence of being a non-coding genome component, satellite DNA sequence evolve fast which leads to changes in sequence composition, distribution among species and abundance (Schmidt & Heslop-Harrison 1998, Macas *et al.*, 2002; Hemleben *et al.*, 2007; Plohl *et al.*, 2008; Palomeque and Lorite, 2008). The high divergence of satellites even between closely related species is characterized by the occurrence of species-specific satellite families or subfamilies, for

example, the pTS5 satellite in *P. procumbens* (Schmidt and Heslop-Harrison, 1996), pTa535 satellite in *Triticum acetivum* (Komuro *et al.*, 2013), and the BoR300 satellite from the olive fruit fly *Bactrocera oleae* (Tsoumani *et al.*, 2013). On the other hand, members of many satellite families show a remarkably high conservation, such as within a genus (Cafasso *et al.*, 2014), within a group of species in a genus (Martinsen *et al.*, 2009), within several genera from a family (Garrido-Ramos *et al.*, 1998), or even within two families (Vittorazzi *et al.*, 2014). This ambivalence is a key feature of repeats in genome evolution (Hall *et al.* 2003) and makes satellite DNAs as a useful tool in phylogenetic analyses. When taking into account a large number of plant satellite families only low sequence conservation was observed (Macas *et al.*, 2002). In addition, plant satellites show an over representation of the AA/TT dinucleotide as well as an enrichment of the CAAAA motif that is supposed to be involved in breakage-reunion of repeated sequences (Macas *et al.*, 2002; Mehrotra and Goyal, 2014), and in increasing DNA stability (Zhang *et al.*, 2013).

Satellite DNA is typically organized in blocks of large tandem arrays of several Mbp in size in centromeric, intercalary heterochromatin and subtelomeric chromosome regions. Many satellite DNAs occurring in centromeres of different plant species have been described, which implies a function in the centromere identity (Plohl *et al.*, 2008; Melters *et al.*, 2013). For example, it has been shown that the 180 bp satellite pAL1 is a key functional element of *Arabidopsis* centromeres, constituting between 2 and 5% of the *Arabidopsis* genome, and actively involved in the binding of *Arabidopsis*-specific CENH3 (Nagaki *et al.*, 2003). The Zcen1 satellite family was characterized in *Zingiber beibersteiniana* as a satellite-centromere interaction (Saunders and Houben, 2001), and the Cen8 in rice (Nagaki *et al.*, 2004). In *B. vulgaris*, the active centromeres comprise mainly of the pBV satellite and Beetle7 Ty3-gypsy retrotransposons (Kowar *et al.*, 2016).

Oftenly, higher-order structures are features of highly abundant and homogenized centromeric satellites as it is exemplarily described for the human alpha-satellite (Rudd *et al.*, 2006; Sevim *et al.*, 2016). Higher-order structures are the result of the simultaneous amplification and homogenization of two or more adjacent monomers (Plohl *et al.*, 2010).

The presence of satellite repeats was shown more than 50 years ago, but it is still today one of the most fascinating parts of the eukaryotic genome (Garrido-Ramos, 2017). Using CsCl density gradients centrifugation combined with renaturation, Green and Gordon (1967) described more than one satellite band being found in the genome of *Nicotiana tabacum* and *Tagetes patula*,

strongly indicating the presence of more than only one satellite family in plant genomes. Through the use of restriction enzymes, cloning and DNA sequencing, Southern hybridization, and cytogenetics, highly repetitive satellite DNAs have now been characterized at a molecular level for many plants (Kubis *et al.*, 1997; Dechyeva *et al.*, 2006; Suarez-Santiago *et al.*, 2007; Kolano *et al.*, 2011). Recently, with next-generation sequencing (NGS), more plant genome sequences become available. An efficient software enabling the analysis of repeats is *RepeatExplorer*, which performs graph-based clustering of sequence read similarities to identify repetitive sequences within genomes (Novak *et al.*, 2010, 2013). This bioinformatic method was applied to species such as *Orobanchaceae* (Piednoel *et al.*, 2012); *Rumex acetosa* (Steflova *et al.*, 2013); *Camelia* (Heitkam *et al.*, 2015), *B. vulgaris* (Kowar *et al.*, 2016), *Locusta migratoria* (Ruiz-Ruano *et al.*, 2016), and *Eragrostis tef* (Gebre *et al.*, 2016). *RepeatExplorer* is also applied to comparative repeat analysis in *Musaceae* (Novak *et al.*, 2014) and *Triatoma infestans* (Pita *et al.*, 2017).

Together with typical satellite DNAs, there are un-typical satellite repeats which are organized in short tandem arrays and integrated within transposable elements. Association of tandem repeats with transposable element has reported in both animals and plants (Satovic *et al.*, 2016). For example, two different MITEs, terMITE1 and terMITE2, were described in termites containing a variable number of internal tandem repeats of 16 and 114 bp long, respectively (Luchetti *et al.*, 2015). Four copies of 154 bp tandem repeats are present in the Helitron-2 transposon of *Drosophila virilis* (Abdurashitov *et al.*, 2013). The subtelomeric Ty3-*gypsy* retrotransposon-Retand element in *Silene* species was reported to contain internal tandem repeats (Kejnovsky *et al.*, 2006), and the PisTR-A satellite integrates into Ogre element in *Pisum sativum* (Neumann *et al.*, 2001; Macas *et al.*, 2009). Bioinformatic analysis of LTR retrotransposons from different plant genomes revealed the frequent occurrence of variable tandem repeats within 3' UTRs of the Tat lineage of mobile elements, some elements contain up to three different tandem repeats (Macas *et al.*, 2009). Therefore, it is possible that retrotransposons contribute to the evolution of satellite DNA by generating a library of short repeats that can later be distributed in the genome and eventually amplified to new satellites (Macas *et al.*, 2009).



### 1.3 Genomic analysis in *Beta* and *Patellifolia* species

#### 1.3.1 Genus *Beta* and *Patellifolia*

The genus *Beta* belongs to the subfamily *Betoideae* - a member of the *Amaranthaceae* family within the *Caryophyllales* order which diverged from other core eudicots approximately 110 Mya (million years ago). The phylogeny within the genus *Beta* has not been finalized. FordLloyd (2005) differentiates the four sections *Beta*, *Corollinae*, *Nanae*, and *Procumbentes*, while Hohmann *et al.* (2006) and Kadereit *et al.* (2006) based on molecular analysis recommend a separation of the section *Procumbentes* as an independent genus *Patellifolia*. Thulin *et al.* (2010) also confirmed that *Patellifolia* is a genus distinct from *Beta*. In this thesis, the taxonomic system as described by Hohmann *et al.* (2006) was used and *Procumbentes* was considered as genus *Patellifolia* (Table 1.1).

**Table 1.1: Taxonomy of genus *Beta* and *Patellifolia***

Genus	Section	Species	Diploid genome (2n)	Distribution
<i>Beta</i>	<i>Beta</i>	<i>Beta vulgaris ssp. vulgaris</i>	18	Coastal habitats from South-Western Norway to Cape Verde
		<i>Beta vulgaris ssp. maritima</i>	18	
		<i>Beta patula</i>	18	
		<i>Beta macrocarpa</i>	18	
<i>Beta</i>	<i>Corollinae</i>	<i>Beta corolliflora</i>	36	Highlands and mountains of Turkey, Armenia and the near-by lands
		<i>B. macrorhiza</i>	18	
		<i>Beta lomatogona</i>	18	
		<i>Beta macrorhiza</i>	18	
		<i>Beta trigyna</i>	45	
		<i>Beta intermedia</i>	36	East Europe to Asia
<i>Beta</i>	<i>Nanae</i>	<i>Beta nana</i>	18	Mountains in Greece
<i>Patellifolia</i>		<i>Patellifolia procumbens</i>	18	Canary Islands, coasts of North-West Africa
		<i>Patellifolia webbiana</i>	18	
		<i>Patellifolia patellaris</i>	36	

Section *Beta* comprises wild and cultivated beets of which the economically most important member is *Beta vulgaris subsp. vulgaris* (sugar beet, herein after referred to as *B. vulgaris*). *Beta maritima* is subspecies of *B. vulgaris* while *Beta patula* and *Beta macrocarpa* form separate species. Species of the section *Beta* are widely distributed along the Mediterranean and central and northern Atlantic coastlines, while wild beets of other sections have a more limited geographic distribution and are either found on European islands of the Atlantic Ocean or at coastal and inland locations from Greece to Iran (Ford-Lloyd and Williams, 1975; De Bock, 1986).

Section *Corollinae* is found across areas of the Balkan Peninsula, Turkey, Transcaucasia, and Iran, a distribution that falls within that of section *Beta* but is generally at higher altitudes. Clarification of the species in this section has been hampered by the occurrence of polyploidy and apomixes (Ford Lloyd, 2005), including three basic species *Beta corolliflora*, *Beta macrorhiza*, *Beta lomatogona*, and two hybrid species *Beta intermedia* and *Beta trigyna*. *B. lomatogona* is a drought-resistant species and it may be an important genetic breeding resource because of its relation to sugar beet with a wide range of phenotypic as well as genomic configurations. In this section, *B. corolliflora* is a tetraploid species. Its polyploid origin has been investigated but it is still unclear if *B. corolliflora* is an autotetraploid or allotetraploid species arisen from *B. lomatogona* and *B. macrorhiza* (Reamon-Buttner *et al.*, 1996).

The section *Nanae* has only one species, namely *Beta nana* which is diploid and has a very restricted distribution on a few mountain tops in Greece. This species was indicated to be more closely allied to section *Corollinae* than to any other section (Ford Lloyd, 2005), a conclusion also supported by Gao *et al.* (2000) using tandemly repetitive DNA.

Genus *Patellifolia* comprises three species *Patellifolia procumbens*, *Patellifolia patellaris* and *Patellifolia webbiana*. Curtis (1968), based on the material of unspecified origin, indicated that *P. patellaris* is tetraploid ( $2n = 36$ ) and self-compatible, whereas *P. procumbens* and *P. webbiana* are diploid and self-incompatible. *P. procumbens* and *P. webbiana* could also be crossed easily, whereas attempts of hybridization between these two species and *P. patellaris* failed. Santoni and Bervillé (1992) concluded on the basis of a study of rDNA unit types, that “the three *Procumbentes* species are closely related and could correspond to one species only”. Bramwell (2001) described that *P. patellaris* is as an annual with cordate leaves, whereas *P. procumbens* and *P. webbiana* are perennials with hastate or sagittate leaves. *P. webbiana* is further differentiated by having “leaves

more or less linear” versus “ovate or deltoid” in *P. procumbens*. From these observations, it remains unclear if the tetraploidy in *P. patellaris* results from an autopolyploidy from *P. procumbens* or an allopolyploidy between *P. procumbens* and an unknown species (Mesbah, 1997).

The genus *Spinacia* is associated with the subfamily *Chenopodioideae* and includes among others the species *Spinacia oleracea* (*S. oleracea*), which is widespread in Europe, Asia and North America. Another genus of the *Chenopodioideae* is *Chenopodium*, which includes among other species *Chenopodium quinoa* (*C. quinoa*), native to South America. Both *S. oleracea* and *C. quinoa* are related to the genus *Beta* and were used as an out group in this study.

### **1.3.2 Repeated DNA analysis in *Beta* and *Patellifolia* species**

Sugar beet *B. vulgaris* is an important crop for food and feed products and bioethanol. It accounts for nearly 30% of the world-wide sugar production (www.proplanta.de, 2014). During the last 200 years of sugar beet breeding, the sugar content has increased from 8% to 18% in today’s cultivars. Breeding has also actively selected for traits like resistance to viral and fungal diseases, improved taproot yield, and abiotic stress resistance (*Beta maritima*: The origin of beets, 2012).

Sugar beet is a diploid species encompassing  $2n = 18$  chromosomes. The haploid genome size is estimated to be 750 Mbp and the genome sequence is available (Dohm *et al.*, 2014). In *B. vulgaris* as well as in wild beets, several studies have been performed analyzing the abundance, genomic organization and evolution of tandemly repeated and dispersed repetitive DNA elements; including LTR retrotransposons (Schmidt *et al.*, 1995; Weber *et al.*, 2010, 2013; Wollrab *et al.*, 2012), non-LTR retrotransposons LINEs (Kubis *et al.*, 1998; Wenke *et al.*, 2009; Heitkam *et al.*, 2009, 2014), non-LTR retrotransposons SINEs (Schwichtenberg *et al.*, 2016), DNA transposons (Jacobs *et al.*, 2004; Menzel *et al.*, 2006, 2008), and satellites (Table 1.2).

Together with sugar beet, low coverage sequencing is also available for *B. lomatogona*, *B. nana*, *P. procumbens*, *P. patellaris* and *C. quinoa*. Therefore, many repeated families, especially satellite families, have been characterized. Satellite DNA families which are genus-, section- or species-specific isolated from cultivated and wild species of genera *Beta* and *Patellifolia* are summarized in Table 1.2. Additionally, detailed characterization of beet repeats regarding methylation of satellite families was carried out (Zakrzewski *et al.*, 2011; Schmidt *et al.*, 2014; Zakrzewski *et al.*,

2014), and interaction between repeat families and histone variants in sugar beet centromeres (Kowar *et al.*, 2016).

### 1.3.3 Fluorescent *in situ* hybridization in chromosome analysis

In plant genome analysis fluorescent *in situ* hybridization (FISH) was introduced for the first time in 1985 (Rayburn and Gill, 1985). This technique allows microscopic identification of chromosomes and localization and visualization of DNA sequences on chromosomes. In addition, chromosomes' corellation genetic linkage groups and physical maps can be performed with markers. The principle is based on the addition of labeled nucleic acid sequences (probes) to complementary target areas (targets) on chromosomes. By denaturation and renaturation of the chromosome in the presence of the probe, the latter is attached to homologous target regions and can be visualized by the label.

For FISH experiments, a variety of different fluorochromes can be used for labeling and multi-color experiments. These include, among others, FITC (fluorescein isothiocyanate, green fluorescence), Texas Red (red fluorescence), AMCA (aminomethylcoumaracetic acid, blue fluorescence), or Cyanine (e.g., Cy5 infrared or Cy3 red). In order to recognize the morphology of the non-hybridized chromatin structures, chromosomes and nuclei are counter-stained. The fluorochromes DAPI (4,6-diamidino-2-phenylindole) or propidium iodide, which emit blue or red fluorescence, respectively, are usually used for this purpose (Schawarazcher and Heslop-Harrison, 2000). Since fluorochromes with different emission spectra are available, simultaneous use of several DNA probes is possible (multi-color FISH, Leitch *et al.*, 1991; Lichter, 1997). The multi-color FISH allows the relative positioning of several DNA markers along chromosomes with up to four probes. An increase in the multi-color FISH was achieved by combinatorial labeling of five fluorochromes (Szinay *et al.*, 2008). The simultaneous detection of four major *B. vulgaris* satellite DNA families (pBV I, pBV VI, pEV and pAv34) revealed the chromosome-specific distribution patterns of each satellite arrays. Therefore, multi-color FISH was proven to be a feasible method for discrimination all nine *B. vulgaris* chromosomes (Päsold *et al.*, 2012).

**Table 1.2: Summary of satellite DNA families in genera *Beta* and *Patellifolia*.**

The distribution is based on Southern hybridization carried out in the indicated reference. The signal is indicated with plus mark (+), no signal is designated with a hyphen mark (-), not determined is marked with (nd). The order was sorted by year of publication.

Satellite family	Monomer size [bp]	Origin	Distribution				Chromosomal localization	Reference
			<i>Beta</i>	<i>Corollinae</i>	<i>Nana</i>	<i>Patellifolia</i>		
pEV1	160	<i>B. vulgaris</i>	+	+	-	+	intercalary	Schmidt <i>et al.</i> , 1991; Zakrzewski <i>et al.</i> , 2010
pBV1	327	<i>B. vulgaris</i>	+	-	-	-	pericentric	Schmidt <i>et al.</i> , 1991; Menzel <i>et al.</i> , 2008
pHT30	140	<i>B. trigyma</i>	+	+	+	-	nd	Schmidt <i>et al.</i> , 1993
pHT36	142	<i>B. trigyma</i>	+	+	+	-	nd	
pHT46	142	<i>B. trigyma</i>	+	+	+	-	nd	
pHC28	149	<i>B. corolliflora</i>	+	+	+	+	intercalary	
pTS5	158	<i>P. procumbens</i>	-	-	+	+	pericentric	Schmidt <i>et al.</i> , 1996
pTS4.1	312	<i>P. procumbens</i>	-	-	+	+	pericentric/intercalary	
pAN1	228	<i>B. nana</i>	+	+	+	-	pericentric/intercalary	Kubis <i>et al.</i> , 1997
pRN1	322	<i>B. nana</i>	+	+	+	-	pericentric/intercalary	
pAV34	363	<i>B. vulgaris</i>	+	+	+	+	subtelomeric	Jansen <i>et al.</i> , 1999
pBC216	322	<i>B. corolliflora</i>	-	+	-	-	intercalary	Gao <i>et al.</i> , 2000
pBC1418	378	<i>B. corolliflora</i>	nd	+	nd	nd	nd	
pBC1447	83	<i>B. corolliflora</i>	nd	+	nd	nd	centromere, on all chromosomes	
pHC8	161	<i>B. corolliflora</i>	+	+	+	-	pericentric on some and dispersed	Gindullis <i>et al.</i> , 2001b
pAp11-1	239	<i>P. procumbens</i>	+	+	-	+	pericentric/intercalary	Dechyeva <i>et al.</i> , 2003
pAp4-1	551	<i>P. procumbens</i>	-	-	-	+	dispersed	
pAp22	582	<i>P. procumbens</i>	-	-	-	+	dispersed	
pRn34	525	<i>B. nana</i>	+	+	+	+	subtelomeric	Dechyeva <i>et al.</i> , 2006
pRp34	352	<i>P. procumbens</i>	+	+	+	+	subtelomeric	
pAV34-1	357	<i>B. vulgaris</i>	+	+	+	+	subtelomeric	Dechyeva <i>et al.</i> , 2008
pAC34-1	357	<i>B. corolliflora</i>	+	+	+	+	subtelomeric	
<i>FokI</i> satellite/Dione	130	<i>B. vulgaris</i>	+	nd	nd	nd	dispersed	Zakrzewski <i>et al.</i> , 2010, 2014
<i>Hinfl</i> satellite/Tantalos	325	<i>B. vulgaris</i>	+	nd	nd	nd	dispersed	
<i>AluI</i> satellite/Niobe	173	<i>B. vulgaris</i>	+	nd	nd	nd	dispersed	
BvMSat1, 3, 4, 5, 6, 7, 8, 9, 10, 11	10-96	<i>B. Vulgaris</i>	+	nd	nd	nd	intercalary or weak signals in the centromere/pericentromere	Zakrzewski <i>et al.</i> , 2010
BvSat4	122	<i>B. Vulgaris</i>	+	+	+	-	nd	Zakrzewski, unpublished

The success and the efficiency of the detection of a nucleic acid sequence by FISH depends on the length of the hybridized target sequence, the frequency of the target sequence in the genome, the number of bound reporter molecules in the DNA probe used and the applied detection method. The development of different labeling methods, novel fluorochromes, new techniques of chromosome preparation and not least the further development of microscopy and image processing technology allow the detection of ever closer neighboring and ever shorter DNA fragments. The resolving power and the sensitivity of the FISH methods have thus been steadily improved. The degree of condensation of the target sequence is decisive for the resolution of the DNA-DNA FISH. Mitotic, highly condensed metaphase chromosomes have a resolution limit of approximately 1 Mb (Heiskanen *et al.*, 1996). Meiotic pachytene chromosomes in plants, on the other hand, are 10-50x longer than mitotic metaphase chromosomes (De Jong, 1981), and loci with distances of 50 kb can be differentiated (Florijn *et al.*, 1996, Raap *et al.*, 1996). The resolving power of fluorescent *in situ* hybridization varies between 2 Mb and 10 Mb and depends on the cytological targets, encompassing interphase nuclei, mitotic prometaphase and metaphase chromosomes, super-stretched mitotic metaphase chromosomes, meiotic pachytene chromosomes, and extended DNA fibers (Jiang and Gill, 2006).

The sensitivity is important for the application of FISH methods. The detection of unique target sequences of less than 1 kb has been achieved in human metaphase chromosomes for the first time in 1990 (Fan *et al.*, 1990) and has been repeated since then. In plants, the detection limit of 700 bp has been achieved 11 years later (Desel *et al.*, 2001). FISH on stretched chromatin threads (fibre-FISH) was performed by Fransz *et al.* (1996) with herbal material and is still a demanding procedure of FISH applications. Besides physical mapping different variants of FISH are applied in plants for chromosome identification (Pedersen and Langridge 1997; Dong *et al.* 2000; Kim *et al.* 2002; Lengerova *et al.* 2004; Szinay *et al.* 2008; Braz *et al.*, 2017), chromosome-arms identification (Päsold *et al.*, 2012), karyotyping (Badaeva *et al.* 2002; Han *et al.* 2008; Falistocco 2009; Amosova *et al.*, 2017), repeat analyses (Dechyeva and Schmidt 2006; Menzel *et al.* 2006; Han *et al.* 2008; Macas *et al.* 2009; Zakrzewski *et al.*, 2013; Heitkam *et al.*, 2015, Ruiz-Ruano *et al.*, 2016), and chromosome-specific painting (Lysak *et al.* 2001; Tang *et al.* 2008; Han *et al.*, 2015). Most recently, CRISPR-FISH has been applied to visualize telomere repeats in live leaf cells as well as telomere movements during interphase of *Nicotiana benthamiana* (Dreissig *et al.*, 2017).

Bacterial artificial chromosomes (BACs) have also been located on chromosomes by BAC-FISH. This method supports in the construction of contigs and positional cloning of important genes (Jiang *et al.* 1995, Gindullis *et al.* 2001a, Lysak *et al.* 2001, Suzuki *et al.* 2001, Cheng *et al.* 2002, Koornneef *et al.* 2003, Lengerova *et al.* 2004; Schulte *et al.*, 2006; Jacobs *et al.*, 2009). BAC-FISH on plant chromosomes was first described by Jiang *et al.* (1995) for the localization of a resistance locus in *Oryza sativa*. A number of other plant species followed, including barley (Lapitan *et al.*, 1997), onion (Suzuki *et al.*, 2001), tomato (De Jong *et al.*, 1999; Szinay *et al.*, 2008), and banana (De Capdeville *et al.*, 2008). BACs contain large genome sections of an organism and therefore also potentially repetitive DNA, which can lead to strong background signals so that the actual target area of the clone is no longer identifiable (Ohmido *et al.*, 1998). The development of a BAC set with chromosome-specific targeting regions allows the labeling and identification of all chromosomes in a cell and has been extended to crops such as *Medicago truncatula* (Kulikova *et al.*, 2001), Sorghum (Kim *et al.*, 2002), and sugar beet (Päsold *et al.*, 2012).

#### **1.4 Aim of this thesis**

In order to broaden the narrow gene pool of sugar beet, an overview of genomic composition, in which repetitive DNAs accounts for major genome proportion, of wild beets and related species is needed. Based on bioinformatic approaches, repetitive as well as satellite component of five species from genera *Beta* and *Patellifolia*, and a distal related species - *C. quinoa* will be characterized in detail.

As close relative to sugar beet section *Beta*, section *Corollinae* will be selected for in-depth analysis with *B. lomatogona* as representative species. The satellite landscape of *B. lomatogona* will be characterized by bioinformatics, molecular genetics, and cytogenetics. Furthermore, using chromosome arms-specific BACs from *B. vulgaris*, the distribution of the *B. lomatogona* chromosome-specific satellite DNA families will be determined by multi-color FISH and the hybridization patterns will be assigned to chromosomes.

Although *P. patellaris* species has been proven to be closely related *P. procumbens*, its origin is still an open question. Using comparative analysis of the genome sequences of *P. procumbens* and *P. patellaris* the overall genomic differences and species-specific as well as species-enriched clusters will be identified. Characterization of these clusters in *P. patellaris* and *P. procumbens* genomes will enable conclusions about the ploidy nature of *P. patellaris*.

## 2 Material and Methods

### 2.1 Material

#### 2.1.1 Plant material

The seeds were incubated in moist Whatman paper for 3-4 days at room temperature without light. The pre-germinated seeds were then planted into soil. The plants were grown in greenhouse conditions. All plants included in the analyses are listed in Table 2.1. For isolation of genomic DNA, leaves have been lyophilized and for preparation of chromosome slides, young leaves were fixed in fixative (methanol:glacial acid = 3:1).

**Table 2.1: Species included in the analyses**

Genus	Section	Species/ Sub-species	2n	Accession	Common name
<i>Beta</i>	<i>Beta</i>	<i>B. vulgaris ssp. vulgaris</i>	18	KWS 232	sugar beet
		<i>B. patula</i>	18	BETA 548	wild beet
		<i>B. lomatogona</i>	18	58258	wild beet
		<i>B. corolliflora</i>	36	17812	wild beet
	<i>Corollinae</i>	<i>B. intermedia</i>	36	BETA407	wild beet
		<i>B. trigyna</i>	45	BETA947	wild beet
	<i>Nanae</i>	<i>B. nana</i>	18	81FD26	wild beet
<i>Patellifolia</i>		<i>P. procumbens</i>	18	35336	wild beet
		<i>P. patellaris</i>	36	54753	wild beet
<i>Spinacia</i>		<i>S. oleracea</i>	12	Matador	-
<i>Chenopodium</i>		<i>C. quinoa</i>	36	CHEN 125	-

#### 2.1.2 Chemicals and consumables

The chemicals and consumables used are listed in Table 2.2, enzymes in Table 2.3 and kits in Table 2.4.



**Table 2.2: Chemicals and consumables**

<b>Product</b>	<b>Company</b>
Acetic acid	Roth, Karlsruhe
Ammonium sulfat	Roth, Karlsruhe
Agarose	Serva Electrophoresis GmbH, Heidelberg
Ampicillin	Roth, Karlsruhe
Bacto-Agar	Roth, Karlsruhe
Anti-digoxigenin antibody FITC	Roche Diagnostics GmbH, Mannheim
Biotin-16-dUTP	Thermo Fisher Scientific Inc. Waltham
Bovine serum albumine	Roth, Karlsruhe
Bromophenol blue	Roth, Karlsruhe
Citiflour AF1	Agar Scientific UK Ltd. Essex, UK
Citric acid	VWR International GmbH, Darmstadt
Cover slip	Menzel Gläser, Braunschweig
DAPI	Fluka Chemie GmbH, Buchs, Schweiz
Dextran sulfate	AppliChem, Darmstadt
dGTP, dTTP-Mix	Thermo Fisher Scientific Inc. Waltham
Digoxigenin-dUTP	Roche Diagnostics GmbH, Mannheim
dNTP-Mix	Roth, Karlsruhe
EDTA	Roth, Karlsruhe
Electroporation cuvettes	BioRad Laboratories GmbH, München
Ethanol	VWR International GmbH, Darmstadt
Ethidium bromide	Roth, Karlsruhe
Falcons 15ml, 50ml	Sarstedt AG & Co, Nümbrecht
Ficoll 400	AppliChem, Darmstadt
Formamide (37 %)	Roth, Karlsruhe
GeneRuler™ 50bp Plus DNA Ladder	Thermo Fisher Scientific Inc. Waltham
GeneRuler™ 100bp Plus DNA Ladder	Thermo Fisher Scientific Inc. Waltham
GeneRuler™ 1kb DNA Ladder	Thermo Fisher Scientific Inc. Waltham
Glucose	Roth, Karlsruhe
Glycerol	VWR International GmbH, Darmstadt
GoTaq-buffer	Promega Corporation, Madison, USA
Humid chamber	Roth, Karlsruhe
Hydrochloric acid	VWR International GmbH, Darmstadt
Hydroxyquinoline	Sigma Aldrich Chemie GmbH, Steinheim
Immersion oil 518F	Carl Zeiss, Oberkochen

**Table 2.2:** Continued

<b>Product</b>	<b>Company</b>
Isopropyl $\beta$ -D-1-thiogalactopyranoside (IPTG)	Thermo Fisher Scientific Inc. Waltham
Klenow buffer	Thermo Fisher Scientific Inc. Waltham
Magnesium chloride	Roth, Karlsruhe
Magnesium sulfate	Roth, Karlsruhe
Methanol	VWR International GmbH, Darmstadt
Monosodium phosphate	Roth, Karlsruhe
Nylon membrane (Hybond XL)	GE Healthcare UK Ltd. Chalfont St. Giles
Paper towels	CWS boco, Dreieich
Parafilm	Bemis flexible packing, Neenah, USA
Paraformaldehyde	Roth, Karlsruhe
Photographic film developer	ADEFO Chemie GmbH, Dietzenbach
Photographic film fixer	ADEFO Chemie GmbH, Dietzenbach
Photographic film Fujicolor Superia X-Tra 400	GE Healthcare UK Ltd. Chalfont St. Giles
Pipette tips	Sarstedt AG & Co, Nümbrecht
Polyvinylpyrrolidone	Roth, Karlsruhe
Potassium chloride	Roth, Karlsruhe
Random hexamer primer	Thermo Fisher Scientific Inc. Waltham
Rapid Ligation Buffer	Promega Corporation, Madison, USA
Restriction buffer CutSmart	New England Biolabs GmbH, Frankfurt am Main
Restriction buffer O	Thermo Fisher Scientific Inc. Waltham
Restriction buffer R	Thermo Fisher Scientific Inc. Waltham
Restriction buffer SuRE/Cut buffer L	Roche Diagnostics GmbH, Mannheim
Restriction buffer Tango	Thermo Fisher Scientific Inc. Waltham
Salmon sperm DNA	Roth, Karlsruhe
Sephadex G50	GE Healthcare UK Ltd. Chalfont St. Giles
Sodium dodecyl sulfate	Roth, Karlsruhe
Sodium chloride	VWR International GmbH, Darmstadt
Sodium citrate	Roth, Karlsruhe
Sodium hydroxide	VWR International GmbH, Darmstadt
Streptavidin-Cy3	Sigma Aldrich Chemie GmbH, Steinheim
Streptavidin-DY-547	Dyomics,
Superfrost slides	Menzel Gläser, Braunschweig
TRIS	Roth, Karlsruhe
Tryptone	Roth, Karlsruhe

**Table 2.2:** Continued

<b>Product</b>	<b>Company</b>
Tween 20	Roth, Karlsruhe
Whatman paper	GE Healthcare UK Ltd. Chalfont St. Giles
X-Gal	Thermo Fisher Scientific Inc. Waltham
X-ray cassettes	Roth, Karlsruhe
Xylene cyanole	Roth, Karlsruhe
Yeast extract	Roth, Karlsruhe
$\alpha$ -[ <sup>32</sup> P]-dATP	Perkin Elmer, Rodgau
$\alpha$ -[ <sup>32</sup> P]-dCTP	Perkin Elmer, Rodgau

**Table 2.3: Enzymes**

<b>Enzyme</b>	<b>Supply Source</b>
Cellulase ( <i>Aspergillus niger</i> ) C1184	Sigma Aldrich Chemie GmbH, Steinheim
Cellulase Onozuka 16419	Sigma Aldrich Chemie GmbH, Steinheim
Cytohelicase ( <i>Helix pomatia</i> ) C8274	Serva Electrophoresis GmbH, Heidelberg
DreamTaq polymerase	Thermo Fisher Scientific Inc. Waltham
GoTaq polymerase	Promega Corporation, Madison, USA
Klenow fragment	Thermo Fisher Scientific Inc. Waltham
Pectinase liquid P4716	Sigma Aldrich Chemie GmbH, Steinheim
Pectolyase ( <i>Aspergillus japonicus</i> ) P3026	Sigma Aldrich Chemie GmbH, Steinheim
Pepsin	Sigma Aldrich Chemie GmbH, Steinheim
Restriction endonuclease <i>AluI</i>	Thermo Fisher Scientific Inc. Waltham
Restriction endonuclease <i>BseGI</i>	Thermo Fisher Scientific Inc. Waltham
Restriction endonuclease <i>BsmI</i>	Thermo Fisher Scientific Inc. Waltham
Restriction endonuclease <i>HpaII</i>	Roche Diagnostics GmbH, Mannheim
Restriction endonuclease <i>NdeI</i>	New England Biolabs GmbH, Frankfurt am Main
Restriction endonuclease <i>MaeI</i>	New England Biolabs GmbH, Frankfurt am Main
Restriction endonuclease <i>MspI</i>	Roche Diagnostics GmbH, Mannheim
Ribonuclease A	AppliChem, Darmstadt
T4 DNA ligase	Promega Corporation, Madison, USA

**Table 2.4: Kits**

<b>Kit</b>	<b>Supply Source</b>
Biotin Nick Translations Kit	Roth, Karlsruhe
Digoxigenin Nick Translations Kit	Roth, Karlsruhe
GeneJET™ Plasmid Miniprep Kit	Thermo Fisher Scientific Inc. Waltham
GeneJet™ Gel Extraction and DNA Cleanup Micro Kit	Thermo Fisher Scientific Inc. Waltham
Nick Translations Kit	Roth, Karlsruhe
pGEM <sup>R</sup> -T Vector System	Promega Corporation, Madison, USA

The strain of *Escherichia coli* XL1 - Blue (Stratagene, La Jolla, USA) was used as a host for plasmid propagation.

For the cloning of PCR products, the high-copy plasmid pGEM-T (Promega) was used.

### 2.1.3 Culture media and antibiotics

All culture media were prepared using desalinated water, followed by autoclaving at 121°C and 2 bar for 20 min.

#### Luria-Bertani (LB) liquid medium

Bacto-Trypton	1	%
Yeast extract	0.5	%
NaCl	1	%

#### LB freezing medium

LB liquid medium with

K <sub>2</sub> HPO <sub>4</sub>	36	mM
KH <sub>2</sub> PO <sub>4</sub>	13.2	mM
sodium citrate	1.7	mM
MgSO <sub>4</sub>	0.4	mM
(NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	6.8	mM
Glycerine	4.4	% (v/v)

#### LB-Agar

LB liquid medium with 1.5 % bacto-agar

Indicator plates

LB-Agar with

IPTG	0.5	mM
X-Gal	0.004	%

SOC medium                      Storage: -20°C

Tryptone/Peptone	2.0	%
Yeast extract	0.5	%
Sodium chloride	0.05	%
Potassium chloride	2.5	mM
Adjust to pH 7.0 with sodium hydroxide		

Addition after autoclaving:

Magnesium chloride	10	mM
Magnesium sulphate	10	mM
Glucose	20	mM

Antibiotics

Ampicillin	100	µg/ml medium
------------	-----	--------------

**2.1.4 Solutions and Buffers**

CTAB (1x)

Tris/HCl (pH 8.0)	0.1	M
EDTA (pH 8.0)	10	mM
NaCl	0.7	M
CTAB	1	% (v/v)

Addition before usage:

β-Mercaptoethanol	0.2	% (v/v)
-------------------	-----	---------

Denhardt solution (100x)

PVP	2	%
BSA	2	%
Ficoll 400	2	%

Neutralization buffer

SSC	0.1	x
Tris HCl, pH 7.5	0.2	M
SDS	0.2	%

Denhardt buffer

Denhardt solution	5	x
SSC	5	x
SDS	0.5	%

Alkaline denaturation solution

Sodium hydroxide	0.4	M
Sodium chloride	1.5	M

<u>Stripping solution</u>			<u>Enzyme buffer</u>		
Sodium hydroxide	0.2	M	Citric acid (pH 4.5)	4	mM
SDS	0.1	%	Sodium citrate	6	mM
			1x enzyme-buffer		
<u>1 x TE buffer</u>			<u>Fixation solution</u>		
Tris/HCl	1	M	Methanol	75	%
EDTA	10	mM	Glacial acetic acid	25	%
pH 8.0					
<u>Enzyme solution (PINE)</u>			<u>SSC/Tween (4x)</u>		
Cellulase ( <i>A. niger</i> )	2	%	SSC	4	x
Cellulase Onozuka	4	%	Tween	0.2	%
Cytohelicase( <i>H.pomatia</i> )	2	%			
Pectolyase ( <i>A .japonicus</i> )	0.5	%	<u>Loading buffer (10x)</u>		
Pectinase liquid	5	%	TAE	1	x
1x enzyme-buffer			Glycerine	50	%
<u>SSC (20x)</u>			<u>Bromophenol blue</u>		
NaCl	3	M	Xylene cyanol	0.1	%
Sodium citrate	0.3	M			
			<u>TAE buffer (50x)</u>		
<u>DAPI solution</u>			Tris base	242	g
Stock: DAPI in H <sub>2</sub> O	100	µg/ml	EDTA (pH 8.0)	50	mM
Final: DAPI in McIlvaine	2	µg/ml	Glacial acetic acid	57.1	ml
			Add H <sub>2</sub> O	1000	ml

### 2.1.5 Oligonucleotides

Primer pairs were selected and analyzed with the internet service OligoAnalyzer 3.1 in consideration of the following criteria: melting temperature, GC content, length, self-dimer. Primers using in this thesis were synthesized by Eurofins GmbH and are listed in Table 2.5. These primers were used for PCR analysis and probe generation.

**Table 2.5: Used oligonucleotides**

Primer name	Sequence 5' – 3'	Length [bp]	GC content [%]	T <sub>m</sub> [°C]
EpiM13-F	GTA AAA CGA CGG CCA GTG	18	55.6	56
EpiM13-R	GGA AAC AGC TAT GAC CAT G	19	47.4	56
pBC1418-F	GTT TTG GGA AGT GAA ATA GC	20	40	56
pBC1418-R	GCT ACA TGT TAC ATA TAG GAG	21	38.1	56
pHT36-F	AAC ATG TGG CTA AAT GCG AG	20	45	49
pHT36-R	GGT CAT ATA TAG TTC CAA TAG G	22	36.4	49
BlSat01-F	ACG AGT AGT TTG ATG CAT G	19	42.1	52
BlSat01-R	CAT TTT CAA GTA AAA TGG CC	20	35	52
BlSat02-F	TAC TTG GAC GGT CAC CTT C	19	52.6	47
BlSat02-R	TGG ACG TTC TCC TTC TTT GG	20	50	47
BlSat03-F	ATA TAC TCA GCC AGA GGT GC	20	50	52
BlSat03-R	TTG GCC AAG TGG GTA CCT TG	20	55	52
BlSat04-F	ACT CCC CTT ATT GCC ATA TG	20	45	64
BlSat04-R	GCA TTA TAA AGT GAA CCC ATC	21	38.1	64
BlSat05-F	AGA TTC CTC AAG TCC GAA TG	20	45	64
BlSat05-R	GTT GAG CAT GGA AAA ATG CC	20	45	64
BlSat06-F	ACTACCACAACCCTTGGGTG	20	55	64
BlSat06-R	TTGGAAGGCACACTCATGCC	20	55	64
PpatSat01-F	CACATGACCAACTCCCGAAGG	21	57.1	62
PpatSat01-R	CGCGTCGCAATTCCGGTCGA	20	65	62
PproSat01-F	AAC ACA TTC AAA CAA AGC	18	33.3	52
PproSat01-R	GTG TTT GAC TTT CAT TTG	18	33.3	52

BACs as probes used in multi-color FISH experiments for discrimination of *B. lomatogona* chromosome-specific satellite families are listed in Table 2.6. These BACs which are *B. vulgaris* chromosome-specific arm were screened from BAC libraries of sugar beet (Päsold *et al.*, 2012).

**Table 2.6: BACs used as probes for detection of chromosome in multi-color FISH analysis**

BAC probe	Linkage group	Chromosome-arm specificity	Insert size [kb]
6G2	LG1	North arm of chromosome 1	112
33K7	LG2	South arm of chromosome 2	112
65L12	LG3	North arm of chromosome 3	130.5
45L24	LG4	North arm of chromosome 4	97
4C10	LG5	North arm of chromosome 5	112
54M10	LG6	South arm of chromosome 6	97
19H23	LG7	South arm of chromosome 7	90
26O24	LG8	North arm of chromosome 8	179
92N14	LG9	South arm of chromosome 9	130.5

**Table 2.7: DNA probes used for FISH and Southern hybridization**

Target sequences	Probe	Length [bp]
BlSat1	BlSat1_1	204
BlSat2	BlSat2_11	539
BlSat3	BlSat3_1	566
BlSat4	BlSat4_7	761
BlSat5	BlSat5_4	645
BlSat6	BlSat6_6	624
PproSat1	PproSat1_1	200
PpatSat1	PpatSat1_2	335
5S rRNA gene of <i>B. lomatogona</i>	pXV2	348
18S-5.8S-25S rRNA genes of <i>B. vulgaris</i> (Päsold <i>et al.</i> , 2012)	pZR18S	8.500

### 2.1.6 Sequence data bases

Paired-end Illumina libraries were used for analysis of repetitive DNA in the genome of *Beta lomatogona* and related species, which are listed in Table 2.8. These sequences were originated from the sugar beet sequencing project (Dohm *et al.*, 2014).



**Table 2.8: Features of paired-end Illumina libraries from studied species**

Species	Insert size [bp]	Number of reads	Size of library [Gb]
<i>B. lomatogona</i>	100	2 x 3.000.000	6.0
<i>P. procumbens</i>	100	2 x 7.000.000	14.0
<i>P. patellaris</i>	100	2 x 7.000.000	14.0
<i>B. vulgaris</i>	100	2 x 3.000.000	6.0
<i>B. patula</i>	100	2 x 3.000.000	6.0
<i>B. nana</i>	100	2 x 3.000.000	6.0
<i>S. oleracea</i>	100	2 x 3.000.000	6.0
<i>C. quinoa</i>	100	2 x 3.000.000	6.0

### 2.1.7 Software

The data was analyzed, visualized and edited by served computer programs and internet services. The detail information of those was listed in Table 2.9.

**Table 2.9: Utilized Software and Internet services**

Program/Service	Function	Reference	Website
Adobe Photoshop 7.0	Editing of autoradiographs and microscope images	---	<a href="http://www.adobe.com">www.adobe.com</a>
BioEdit	Sequence storage and management	Hall (1999)	<a href="http://www.mbio.ncsu.edu/BioEdit/bioedit.html">www.mbio.ncsu.edu/BioEdit/bioedit.html</a>
BLAST	Homology search	Altschul <i>et al.</i> (1990)	<a href="http://www.ebi.ac.uk/Tools/sss/">www.ebi.ac.uk/Tools/sss/</a> <a href="http://blast.ncbi.nlm.nih.gov/Blast.cgi">http://blast.ncbi.nlm.nih.gov/Blast.cgi</a>
Case Data Manager Expo 4.5.0.28	Analysis of images obtained by the fluorescent microscope	---	<a href="http://www.spectral-imaging.com">www.spectral-imaging.com</a>
Geneious 6.0	Alignments, assemblies, detection of tandem repeats	Drummond <i>et al.</i> (2010)	<a href="http://www.geneious.com">www.geneious.com</a>

**Table 2.9:** Continued

Program/Service	Function	Reference	Website
MUSCLE	Multiple Alignment	Edgar (2004)	<a href="http://www.ebi.ac.uk/Tools/msa/muscle">www.ebi.ac.uk/Tools/msa/muscle</a>
MEGA 7	Generation of phylogenetic tree	Kumar (2016)	
OligoAnalyzer 3.1	Analysis of oligonucleotides	---	<a href="http://Eu.idtdna.com/analyzer/Applications/OligoAnalyzer">Eu.idtdna.com/analyzer/Applications/OligoAnalyzer</a>
RepeatExplorer	Detection of repetitive sequences	Novák <i>et al.</i> (2010)	<a href="http://www.repeatexplorer.org">http://www.repeatexplorer.org</a>
SeqGrappheR	Visualization and analysis of graphical representation of repeats	Novák (2012)	
Tandem Repeat Finder	Detection of tandem repeats in DNA sequences	Benson (1999)	

## 2.2 Methods

### 2.2.1 Bioinformatics

#### 2.2.1.1 Identification of putative satellite families and monomer size

High-through put genome sequencing data were used as an input for *RepeatExplorer* (Novák *et al.*, 2010) to perform graph-bases clustering analysis of sequence read similarities to identify repetitive elements. In this study, three million of the paired-end Illumina reads from *Beta lomatogona* were analyzed as a single *RepeatExplorer* running, seven million of the paired-end Illumina reads from *Patellifolia patellaris* and *Patellifolia procumbens* were analyzed as a comparative *RepeatExplorer*. The *RepeatExplorer* clusters were examined with special attention to clusters whose graph showed circular or star shape, these clusters indicated a composition of satellite repeats.

Based on their graphical representation, candidate tandem repeat clusters were selected. The corresponding contigs were analyzed by the program *Geneious* (Drummond *et al.*, 2011) in order to detect tandem repeat monomers. Besides, the contigs also were used as an input for the program Tandem Repeat Finder (Benson, 1999) to determine tandem repeat monomers. The outputs of the

two programs were compared to receive final results. The identified monomers then were used as the template for the mapping of three million reads against those monomers. The mappings were repeated with artificial dimers to investigate the read coverage over the full length of the monomers. From these assemblies, the consensus sequences of the monomers were isolated to obtain representative sequences of the repeat monomers.

In the other hand, the most representative contigs of each candidate cluster were analyzed using *SeqGrapheR* tool which is designed to complement *RepeatExplorer* in order to visualize graph representation of repeats and investigate sequence variability within repeat families.

#### 2.2.1.2 Sequences analysis

After receiving the sequencing results, plasmid insert sequences were detected using the program *Geneious* which discriminate them from the known sequences of plasmids. The plasmid insert sequences were aligned together and with the monomer consensus sequences to examine whether the inserts are suitable as probes for Southern hybridization as well as for fluorescent *in situ* hybridization.

#### 2.2.1.3 Similarity search

A basic local alignment search tool (BLAST) analysis was performed to investigate if other species also carry repeats similar to the identified ones within their genome or these repeats are really novel ones. Therefore, the query sequences based on the isolated monomer sequences were uploaded to the NCBI BLAST service and a search with the Blastn algorithm in the database of nucleotide collection was performed.

The mapping of sequence reads from related species against the identified monomers also was performed to investigate the relative proportion of each monomer in those species' genome.

#### 2.2.1.4 Phylogenetic tree

The alignment of selected sequences was used as input for phylogeny tool in MEGA7. The default parameter was applied, only changed complete deletion into pairwise deletion in Gap/Missing data treatment.

## 2.2.2 Molecular methods

### 2.2.2.1 Isolation of DNA

#### **Isolation of plant DNA**

Genomic DNA was isolated from young lyophilized leaves using the CTAB standard protocol (SaghaiMaroof *et al.*, 1984) with few modifications.

- The leaf material was dried for 2 - 3 days in a vacuum chamber at -60°C and 0.2 mbar, and was stored in a freezer at -20°C until using.
- Freeze-dried leaves were pulverized by smashing them with metal beads.
- Addition of 12.5 ml of pre-warmed CTAB buffer including 18 µl β-mercaptoethanol, gently incubated for 30 min at 65°C.
- Centrifugation for 30 min at 3200 g and 4°C.
- The upper phase was transferred into a new 50 ml tube; addition of 1 volume chloroform : isoamyl alcohol (24:1 v/v) and incubation for 10 min in an overhead mixer.
- Centrifugation for 30 min at 3200 g, and 4°C.
- The upper phase was transferred into a new 50 ml tube; addition of 50 µl RNase A (10 mg/ml) and incubation for at least 30 min at 37°C.
- Incubation of the sample on ice for at least 5 min.
- Addition of 0.7 volumes cold isopropanol and careful inversion of the tube in order to precipitate DNA.
- DNA was transferred into a new 2 ml tube and 76 % ethanol was added for washing step.
- Washing step was repeated one more time with 76 % ethanol.
- After air-drying of DNA, 100 to 500 µl water was added.
- The pellet was dissolved overnight.

### Isolation of bacterial plasmid DNA

DNA of the high copy number plasmid pGEM<sup>R</sup>-T was isolated with the GeneJet<sup>TM</sup> Plasmid Miniprep Kit following the instruction of the user's manual with a modified elution step as DNA was eluted in deionized water.

#### 2.2.2.2 Agarose gel electrophoresis

Gel electrophoresis was carried out with agarose gels from 1.0 to 2.0 % agarose concentration. For the preparation of gels, agarose was dissolved in 1x TAE buffer and the DNA stain ethidium bromide was added with a final concentration of 0.004 %. The separation took place with 1-5 V/cm.

#### 2.2.2.3 Polymerase chain reaction

Polymerase chain reaction (PCR) was carried out with repeat-specific primers (Table 2.5). The PCR reaction was composed as follows:

DNA template	50.0	ng
PCR buffer (5x)	10.0	μl
dNTP	0.2	mM
Primer forward	10.0	pmol
Primer reverse	10.0	pmol
<i>Taq</i> polymerase (5U/ μl)	0.5	μl
ddH <sub>2</sub> O ad	50.0	μl

For satellite B1Sat02, B1Sat03 and B1Sat05, PCR reactions were optimized by adding DMSO in final concentration of 10%.

#### PCR program

Pre-denaturation	95°C	3 min	
Denaturation	95°C	30 sec	} 35 cycles
Annealing	47-64°C	20 sec	
Elongation	72°C	60 sec	
Final elongation	72°C	5 min	

The Annealing temperature varied depending on the primers' base composition.

#### 2.2.2.4 Cloning

##### Elution of DNA fragments from agarose gels

After PCR, the DNA fragments were separated by gel electrophoresis and cut out of the gel. The GeneJet™ Gel Extraction and DNA Cleanup Micro Kit were used in order to extract the fragments from the gel slices. The instructions of the user's manual were followed apart from the final step, where DNA was eluted in deionized water.

### **Ligation of DNA**

The method is based on the ability of bacterial cells to maintain and replicate plasmids. Cloning vectors are specialized artificial plasmids allowing to transfer and accumulate the desired DNA fragments in the host bacteria. The vectors contain selectable markers, antibiotic resistance, replication origin and multiple cloning sites/polylinkers. In the experiments described here, the PCR fragments were cloned into the pGEM<sup>R</sup>-T cloning vector following the instruction of the user's manual.

### **Preparation of electrocompetent *E. Coli* cells**

*E.coli* XL1-Blue fresh cells are plated on LB-agar plate and incubated overnight at 37°C.

Day 1: Selection of a single colony of *E. Coli* from fresh LB plate and inoculation of a 5 ml starter culture of LB. Grow culture at 37°C in continuously shaking at 220 rpm overnight.

Day 2:

- Inoculate 1 liter pre-warmed LB liquid medium with the starter culture and separate the culture on two flasks.
- Incubate the cells for 3-4 hours at 37°C and 220 rpm until they have reached an optical density of OD<sub>600</sub> = 0.6, immediately put the cells on ice.
- Place centrifuge bottles on the ice at this time.
- Split the culture into four ice-cold centrifuge bottles and centrifuge at 5800 rpm (Beckman JA-10 rotor) for 10 minutes at 4°C to harvest the cells.
- Decant the supernatant and resuspend each pellet in 10 ml of ice-cold 10 % glycerol and then combine resuspension into two centrifuge bottles and fill up to 200 ml with ice-cold 10 % glycerol.
- Harvest the cells by centrifugation at 5800 rpm (Beckman JA-10 rotor) for 10 minutes at 4°C. Decant the supernatant and resuspend each pellet in 10 ml of ice-cold 10 % glycerol and fill up to 200 ml.

- Repeat the centrifuge step above one time. At this step, chill two 50 ml conical tubes on ice.
- Decant the supernatant and resuspend each pellet in 10 ml of ice-cold 10 % glycerol. Transfer each suspension to a 50 ml conical tube and fill up to 50 ml.
- Centrifugation of tubes at 5000 rpm (Beckman JA-10 rotor) for 10 minutes at 4°C.
- Carefully resuspend each pellet in 1 ml of ice-cold 10 % glycerol by gently swirling. Fill 50 µl aliquots into sterile 1.5 ml microfuge tubes (already chilled on ice) and immediately snap freeze the aliquots with liquid nitrogen.
- Store frozen cells in a -80°C freezer.

Transformation efficiency was tested by the following steps:

- Thaw 2 aliquots of electrocompetent cells on ice. One aliquot was used to transform with 1 µl vector plasmid-DNA (10pg) and one was used as un-transformed control (DNA transformation see below).
- After incubation in 1 ml SOC medium for 1 hour at 37°C (220 rpm), the cell suspension was plated to indicator plates including the corresponding antibiotic to calculate transformation efficiency and to plate with other antibiotic plates to test contamination.
- Incubation of plates over night at 37°C.

### **DNA transformation**

The competent cells of the *E. coli* strain XL1-Blue were transformed with the insert-carrying vector via electroporation. After adding 1-3 µl of the ligation mix to a thawed *E. coli* aliquot of 50 µl, the mix was transferred to a 2 mm electroporation cuvette and exposed to a voltage of 2500 V. Immediately after the electroporation, prewarmed SOC medium was added. After incubation for 1 h at 37°C the cells were plated on indicator plates and incubated over night at 37°C.

### **Blue/white-screening**

After incubation of the indicator plates, recombinant white colonies were detected as the insert prevent the expression of the *β-galactosidase* gene. Those colonies were picked and transferred to the colony PCR-mix as well as to LB-medium (including Ampicillin 100 µg/ml). For colony PCR, the PCR reaction described in section 2.2.2.3 was modified by replacing the DNA with the inoculum of the colony and the use of vector-specific EpiM13 primers. The PCR program

mentioned in section 2.2.3.2 was modified by extended initial denaturation 6 minute at 94°C, shortening of the denaturation step to 20 seconds and of the elongation step to 1 minute.

### **Plasmid isolation and sequencing**

Colonies which were indicated by PCR to carry plasmids with inserts were chosen for sequencing. The plasmids were isolated with the GeneJet™ Plasmid Miniprep Kit following the instruction of the user's manual with a modified elution step as DNA was eluted in deionized H<sub>2</sub>O. Sequencing of plasmid inserts was carried out by the company Eurofins Genomic GmbH.

#### 2.2.2.5 Southern hybridization

### **Preparation of Southern membranes**

The agarose gel after restriction and electrophoretic separation was exposed to UV light for 1 min to take gel images with ruler and marker. Subsequently, the DNA was transferred by alkaline capillary transfer in 0.4 M NaOH/ 1.5 M NaCl onto positively charged Hybond N<sup>+</sup> membrane. The membrane was washed in 2x SSC for 5 min at RT and fixed for 2 hours at 80°C.

### **Preparation of probes**

The probes for the Southern hybridization were prepared by PCR using the PCR reaction and PCR program described in section 2.2.2.3 with EpiM13 primers. For the preparation of the probes, the plasmids whose inserts had the highest similarity to the repeat consensus sequence were chosen.

After PCR, the probes were purified by extraction of DNA fragments from the gel slices, using The GeneJet™ Gel Extraction and DNA Cleanup Micro Kit. The instructions of the user's manual were followed apart from the final step where DNA was eluted in deionized H<sub>2</sub>O.

### **Random prime labeling of DNA probes**

DNA probes were labeled with <sup>32</sup>P as follows:

- 50-100 ng of the DNA was resuspended in water to a final volume of 76 µl.
- The probe was denatured for 10 min at 95°C and quickly chilled on ice for at least 5 minutes.
- The following reagents were added to the probe: 10 µl of 10 x Klenow buffer, 5 µl random primer, 5 µl of 0.5 mM dGTP/ dTTP, 1.5 µl of α-<sup>32</sup>P-dATP, 1.5 µl of α-<sup>32</sup>P-dCTP, 2 u/µl of Klenow polymerase.



- The mixture was incubated at least 1 hour at 37°C.
- The labeled probe was purified from unincorporated radionucleotides via Sephadex G-50 column by centrifuging at 1000 g for 2 minutes. The column was washed with 100 µl 1 x TE and centrifuged for additional 2 minutes.
- Finally, the DNA probe was denatured for 10 minutes at 95°C and directly used for hybridization.

### **Southern hybridization**

The Southern hybridization was performed as follows:

- The Hybond N<sup>+</sup> membrane was pre-hybridized in 50 ml of Denhardt buffer (containing 5 x Denhardt solution with 5 x SSC and 0.5 % SDS) and 1 ml denatured salmon sperm – DNA (10 mg/ml) for 2 hours at 60°C.
- The membrane was transferred into hybridization tubes containing 25 ml hybridization solution and the labeled heat-denatured probe was added.
- The membrane was hybridized at 60°C overnight in a hybridization oven to achieve the desired stringency.
- The membrane was washed once for 10 minutes in 2 x SSC/0.1 % SDS at 60°C and once for 10 minutes in 1 x SSC/0.1 % SDS at 60°C.
- The membrane was transferred into the foil to avoid drying out.
- The autoradiogram was taken on the double-coated X-Ray film Hyperfilm-MP using individual exposure times at – 80°C incubation.

### **Stripping of membrane**

Stripping of the membrane was performed by washing for 15 minutes in stripping solution at 60°C, followed by washing for 10 minutes with demineralized H<sub>2</sub>O and 20 minutes in neutralization buffer. Finally, the membrane was washed in 2x SSC and dried at 80 °C. After this step, the membrane was ready for further hybridization experiments.

### **2.2.3 Molecular cytogenetic methods**

#### **2.2.3.1 Preparation of plant chromosomes**

##### **Fixation of plant material**

The plant material was pre-treated and fixed as follows:

- The flower and leaf material was collected 4-5 h after dawn.
- Flowers were fixed directly in fixative. Leaves were pre-treated with 2 mM 8-hydroxyquinoline for 2.5-3.5 h depending on the desirable rate of chromosome condensation and transferred into fresh fixative.
- The fixative was changed ones after a 3-4 h incubation at 4°C. The fixed material could be stored at 4°C for several months.

##### **Preparation of mitotic chromosomes**

The dropping method enabled to prepare mitotic chromosome with a large number of microscopy slides of uniform quality. It was applied for the chromosome preparation from young leaves and root tips according to Schwarzacher & Heslop-Harrison (2000) with modifications.

- Fixed plant material was washed once for 5 min in water and twice for 5 min in citrate buffer at RT.
- The material was transferred to the appropriate enzyme solution (PINE) in citrate buffer (see 2.1.4). Leaves were incubated for 3 h at 37°C or overnight at RT.
- Afterwards, the material was macerated with the forceps and preparative needle, mixed carefully with a 200 µl pipette and incubated again for 10-15 min at 37°C. Removal of underground particles.
- The material was washed twice with citrate buffer by centrifugation for 5 min at 4000 rpm, RT.
- The buffer was replaced twice with fresh fixative after centrifugation, for 5 min at 4000 rpm, RT.

- After the final centrifugation for 5 min at 4500 rpm, RT, the supernatant was carefully removed with a Pasteur pipette leaving only 100 - 200  $\mu$ l of the nuclei suspension in the tube. The walls of the tube were carefully rinsed with another 50-100  $\mu$ l of fresh fixative.
- 13  $\mu$ l of the mixed material was dropped onto an acid-cleaned glass slide from a height of 50 cm. The slide was shaken off to release the nuclei from the cytoplasm.
- Slides were examined with the phase-contrast microscope Zeiss Axioscop 40 at magnifications x 10 and x 40 to select slides with chromosome spreads clear of cytoplasm. These slides could be stored at 4 °C for a few months.

#### 2.2.3.2 Labeling of DNA probes for FISH

In order to detect specific DNA sequences on plant chromosomes, the corresponding probes were labeled with biotin/digoxigenin and detected immunologically with the antibodies coupled to fluorescent fluorochromes (indirect labeling) or fluorochrome (direct labeling).

#### **Labelling of DNA probes by PCR**

Labelling by PCR was suitable for DNA probes less than 3 kb long and was performed as follows:

##### - PCR reaction

Template DNA	20-50	ng
M13 Forward primer	20	pM
M13 Reverse primer	20	pM
10 x PCR buffer	5.0	$\mu$ l
dNTPs	10	mM
Digoxigenin-11-dUTP	1.75	nM
or Biotin-16-dUTP	3.5	nM
<i>GoTaq</i> DNA polymerase	2.5	units
Total volume	50	$\mu$ l

##### - PCR program

Pre-denaturation:	94°C	3 min	
Denaturation:	94°C	30 sec	} 35 cycles
Annealing:	56°C	30 sec	
Elongation:	72°C	90 sec	
Final elongation:	72°C	5 min	

The quality of the labeling was checked by agarose gel electrophoresis of an aliquot. The labeled probe migrates slower than the unlabeled control PCR product and is visible in the gel as a shifted band.

### **Labeling of DNA probes for FISH by nick translation**

The nick translation method is based on the ability of the DNase I to introduce randomly distributed breaks of a single strand, or nicks, into DNA. The nicks are then filled by DNA *polymerase* I, which replaces the removed nucleotides with digoxigenin- or biotin-labeled ones.

Labeling of probes by nick translation was applied for DNA probes larger than 3 kb.

Labeling by nick translation was performed with DIG-Nick Translation and Biotin-Nick Translation kits following the manufacturer's instructions.

The labeled probes were purified from the unincorporated label by ethanol precipitation: 0.1 volume of 4M LiCl and 2.5 volume of absolute ethanol were added into mixture. After incubation at -20°C overnight (or -80°C for 30 minutes), the mix was centrifuged for 20 min at 4°C with maximum speed. The supernatant was discarded and the air-dried pellet was resuspended in 20 µl deionized H<sub>2</sub>O.

#### 2.2.3.3 Fluorescent *in situ* hybridization (FISH)

Fluorescent *in situ* hybridization is a method allowing the visualization of fluorescent DNA labeled probes on chromosomes under the UV-microscope. The procedure consists of the pre-treatment of chromosome spreads, hybridization, post-hybridization washes and the detection of the probes.

### **Fluorescent *in situ* hybridization on chromosome spreads**

*In situ* hybridization and probe detection was performed according to Heslop-Harrison *et al* (1991) modified for beet by Schmidt *et al.* (1994). During the whole FISH procedure, the preparations should be treated very carefully and once wet, they should not dry out. All washing steps have been performed in Coplin jars, while all incubation steps have been carried out at 37°C in a humid chamber with a plastic stripe covered.

Pre-treatment of chromosome preparations:

- The selected slides were aged overnight at 37°C in an incubator. The area containing chromosome spreads was indicated with a diamond pen.

- 2 µg of RNase A (10 µg/µl) in 200 µl of 2x SSC were applied per slide, the preparations were covered with plastic cover slips and incubated in a humid chamber for 1 hour at 37°C.
- After the incubation, the cover slips were carefully removed and the slides were washed three times for 5 minutes with 2x SSC at RT.
- Slides were equilibrated in 0.01 N HCl for 1 minute, and 10 µg pepsin in 200 µl of 0.01N HCl was applied per slide and incubated for 15 minutes at 37°C.
- The cover slips were carefully removed and the slides were washed three times for 5 minutes with 2x SSC.
- The preparations were fixed in freshly prepared 4 % formaldehyde solution for 10 minutes, followed by three washing steps with 2x SSC for 5 minutes.
- The slides were dehydrated in 70 %; 90 % and 100 % ethanol for 3 minutes and air-dried.

*In situ* hybridization:

- The hybridization solution contained:
 

15 µl	Formamide (100 %)
6 µl	Dextran sulphate (50 %)
0.5 µl	SDS (10 %)
2 µl	salmon sperm DNA (1 µg/ µl)
3 µl	20 x SSC
X µl	labeled probes (0.5 – 2 ng/µl)
Y µl	H <sub>2</sub> O ad 30 µl final volume

Hybridization solution was pre-denature at 70°C for 10 minutes and immediately chilled on ice.

This composition had stringency of 76 % at 37°C.

- 30 µl of the hybridization solution was applied in small drops onto dried slides, the preparations were covered with plastic cover slips, denatured and stepwise drilled using the *in situ* thermocycler Touchdown.

The denaturation program was:	70°C	8 min
	55°C	5 min
	50°C	2 min
	45°C	3 min
	37°C	overnight

- The slides were transferred to a humid chamber and hybridized overnight at 37°C.

Post-hybridization washing:

- The cover slips were carefully removed by submerging the slides in 2 x SSC
- Two washing steps were performed in 20 % formamide in 0.1x SSC (79 % stringency) for 5 minutes at 42°C.
- The washing solution was removed by rinsing twice in 2x SSC for 5 minutes at 42°C and once in 2x SSC for 5 minutes at 37°C.

For probes labeled with direct fluorochromes proceed to counterstaining and mounting step.

Signal detection (for indirectly labeled probes):

- The slides were equilibrated in 4x SSC/0.2 % Tween for 5 minutes at 37°C.
- 200 µl of 5% BSA in 4x SSC/0.2 % Tween was applied per slide, followed by an incubation step for 30 minutes at 37°C.
- 100 µl of the appropriate antibody dilution in 3 % BSA in 4x SSC/0.2 % Tween were applied per slide and the slides were incubated under the same plastic cover slips for 1 hour (or more) at 37°C.

Antibody dilutions: for digoxigenin labeled probes Anti-DIG-FITC 1:75 (v:v) and for biotin labelled probes Streptavidin-Cy3 (Sigma) 1:200 (v:v).

- After the detection, cover slips were removed and unbound antibody was washed off for 10 minutes three times in 4x SSC/0.2 % Tween at 37°C.

Counterstaining and Mounting:

- The slides were drained off excess 4x SSC/0.2 % Tween and submerged in 2x SSC to avoid drying out.
- Finally, 2 - 4 µg/ml DAPI solution and a drop of antifade solution (CityFluor AF1) were applied, the preparations were covered with glass cover slips (24 x 32 mm) and squeezed out the excess liquids using filter paper. The slides could be stored at 4°C overnight to obtain more stable signals.

### Multicolor *in situ* hybridization with BAC probe

Multicolor FISH was performed according to the standard FISH protocol above, except for hybridization solution where salmon sperm DNA was replaced by Cot-100 DNA. A maximum of four probes were used in one FISH experiment.

#### 2.2.3.4 UV microscopy, photography and image processing

The slides after *in situ* hybridization were analyzed with a Zeiss Axioplan 2 imaging UV-fluorescence microscope. If fluorescent dyes are excited with light of a certain wavelength, the emission can be made visible with corresponding optical filters. The used fluorophores with the maximum excitation, the maximum emission, and the filter name are shown in Table 2.10.

The individual filters were recorded using the Applied Spectral Imaging software (ASI) 3.3 and the coupled CCD camera BV300-20A in grey values using a magnification of 1600x. The pseudo-colorization, brightness and contrast optimization as well as the superimposition of the individual channels were also carried out with the ASI software. The following image processing was performed with Photoshop 7.0 with tools which only change the photographs uniformly.

**Table 2.10: Maximum of excitation and emission of the used fluorochromes**

Fluorochrome	Maximum of excitation [nm]	Maximum of emission [nm]	Color at max of emission	Filter name
Cy3	550	570	Yellow	Zeiss 15
DAPI	358	461	Blue	Zeiss 02
DY-415	418	467	Blue	AHF F36-544 HC
DY-495	493	521	Green	AHF F36-720 HC
DY-547	557	574	Yellow	AHF F36-730 HC
DY-647	653	672	Dark red	Zeiss 26
FITC	492	520	Green	Zeiss 09

### 3 Results

#### 3.1 Comparative identification of repetitive DNA in sugar beet and related wild beet species

In order to compare, characterize, and quantify the repetitive portion of selected species in the beet genera, graph-based read clustering was performed, using *RepeatExplorer* software. Partly, results from this thesis have been obtained in cooperation with other projects; including the repeat analysis from *B. vulgaris* (Kowar *et al.*, 2016), *B. nana* (Bannack, 2017), and *C. quinoa* (Ost, 2016).

The number of sequence reads from six species used for *RepeatExplorer* analysis is shown in Table 3.1. In *RepeatExplorer* only clusters making up more than 0.01% of the genome of these species were taken into account. These clusters were used as queries to search against databases which include all known Beta repeats, rDNA, telomere DNA, plastid DNA and protein databases. The type of repetitive DNAs was approved when more than 10% of the reads of the cluster generated a positive hit.

**Table 3.1: Sequence reads from reference species used for *RepeatExplorer* analysis**

Species	Genome size [Mb]	Number of sequence reads	Genome coverage [%]
<i>B. vulgaris</i>	750	3.000.000	30
<i>B. lomatogona</i>	] Not determined	3.000.000	-
<i>B. nana</i>		3.000.000	-
<i>P. procumbens</i>		7.000.000	-
<i>P. patellaris</i>		7.000.000	-
<i>C. quinoa</i>	1390	3.000.000	20

The classification of major repeats from *RepeatExplorer* outputs were summarized in Figure 3.1. Among all classes of repetitive DNAs, DNA-transposon, Ty3-gypsy and Ty1-copia retrotransposons, rDNA and satellite repeats are high and moderately repeated sequences in all analyzed *Beta*, *Patellifolia*, and *Chenopodium* genomes. A small proportion of LINE and SINE was identified in all genomes. Plastid DNA was annotated with 11.1% in *B. vulgaris* whereas this repeated sequence was not detected in other genomes. Satellite DNA was observed in all analyzed genomes, ranging from 6.6% in *B. lomatogona* to 22.7% in *B. vulgaris*. In each genome, a large proportion of repeat clusters is unknown (14.1% - 56.9%).



Analysis of *RepeatExplorer* outputs revealed that repetitive DNA sequences form the largest proportion of the total DNA (the bar charts in Figure 3.2). The analyzed clusters which correspond to repetitive DNAs account for nearly 50% (47.9% - 49.9%) of each genome, except for *B. lomatogona* (65.6%) and *C. quinoa* (34.4%). In each bar chart, green clusters representing satellite families were indicated in more detail in the pie charts in Figure 3.2. All repeated classes in six species were summarized in Table 3.2.

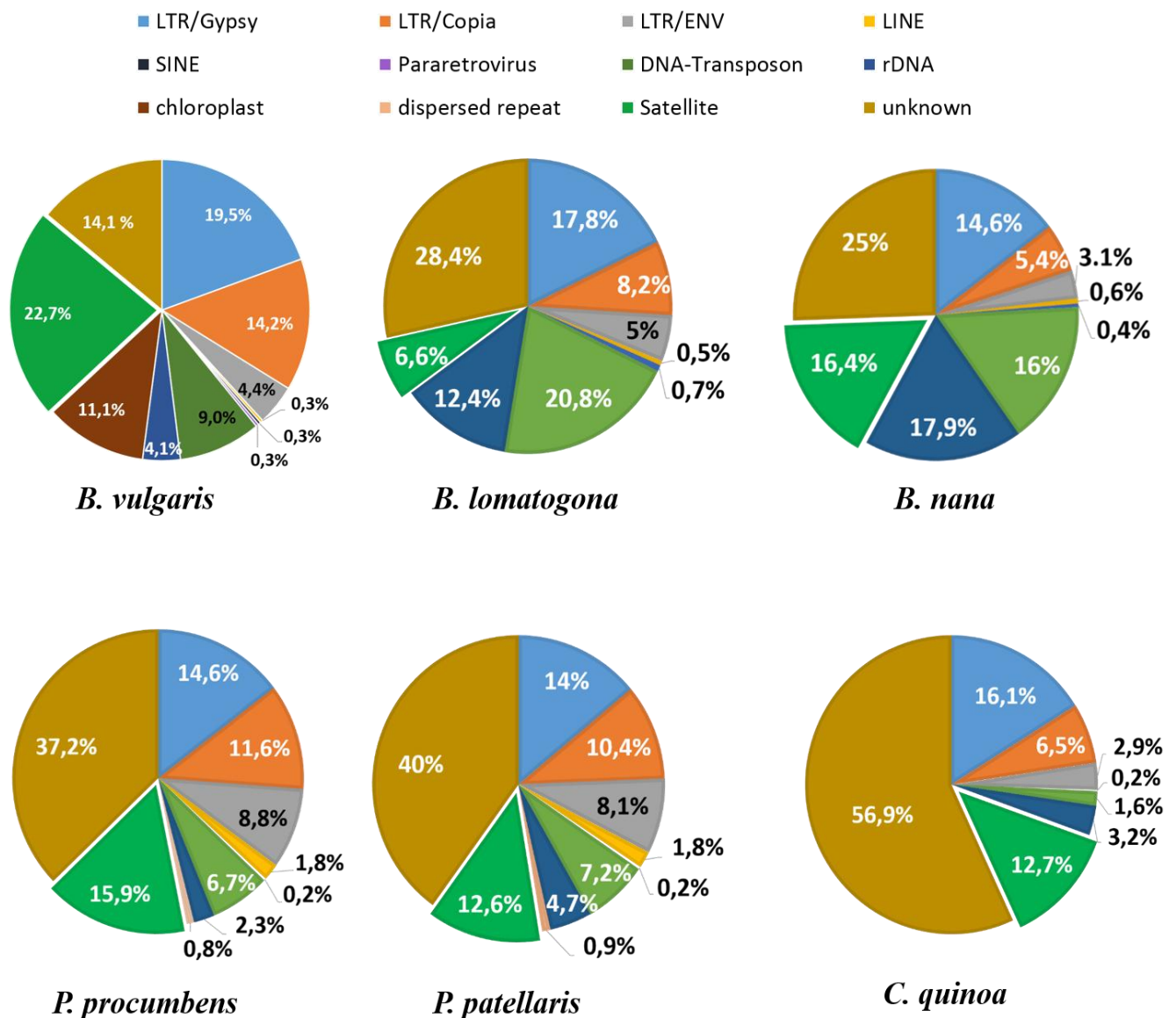


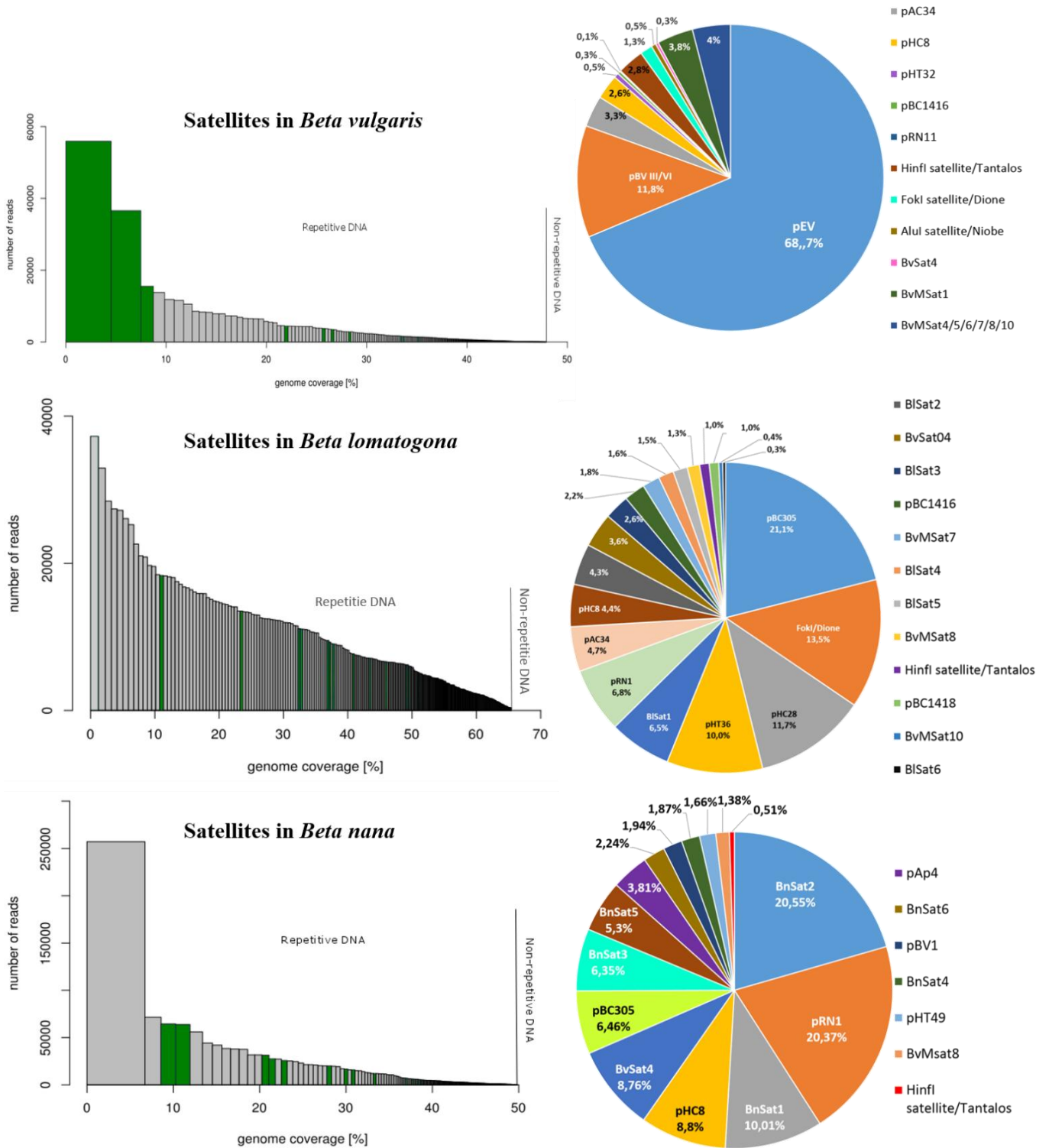
Figure 3.1: Repeat compositions of *Beta*, *Patellifolia* and *C. quinoa* species analyzed by mean of *RepeatExplorer*.

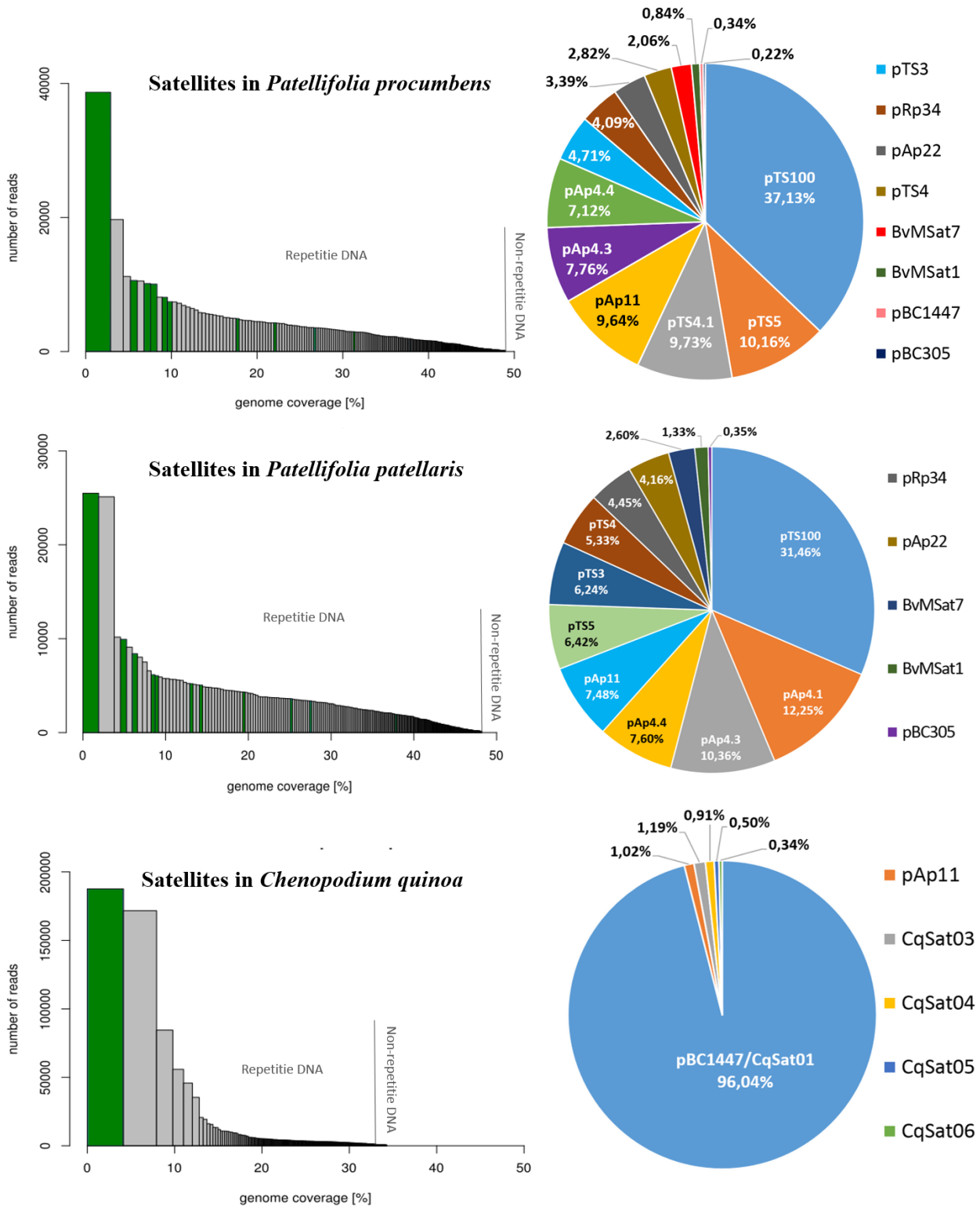
**Table 3.2: Genome proportion of repetitive sequences in *Beta*, *Patellifolia*, and *C. quinoa* species analyzed by RepeatExplorer**

Repeat class	Genome proportion [%]					
	<i>B. vulgaris</i>	<i>B. lomatogona</i>	<i>B. nana</i>	<i>P. procumbens</i>	<i>P.</i>	<i>C. quinoa</i>
LTR/ <i>Gypsy</i>	19.5	17.8	14.6	14.6	14.0	16.1
LTR/ <i>Copia</i>	14.2	8.2	5.4	11.6	10.4	6.5
LTR/ENV	4.4	5.0	3.1	8.8	8.1	2.9
Non-LTR/LINE	0.3	0.5	0.6	1.8	1.8	0.2
Non-LTR/SINE	0.3	0.7	0.4	0.2	0.2	0.2
Pararetrovirus	0.3	-	-	-	-	-
DNA-transposon	9.0	20.8	16.0	6.7	7.2	1.6
rDNA	4.1	12.4	17.9	2.3	4.7	3.2
Chloroplast	11.1	-	-	-	-	-
Dispersed repeat	-	-	-	0.8	0.9	-
<b>Satellite</b>	<b>22.7</b>	<b>6.6</b>	<b>16.4</b>	<b>15.9</b>	<b>12.6</b>	<b>12.7</b>
Unknown	14.1	28.4	25.0	37.2	40	56.9

Note: “-“ not detected.

The satellite families of each analyzed genome were shown in the pie chart (Figure 3.2) and Table 3.3. The pBV and pEV satellites were the most abundant satellite families in sugar beet *B. vulgaris*, which together constituted nearly 80.5% of all known satellites. pAC34, pHC8, *HinfI* satellite/Tantalos, and BvMSat1 accounted for 3.3%, 2.6%, 2.8%, and 3.8%, respectively, while the six remaining satellites (pHT32, pBC1416, pRN11, *FokI* satellite/Dione, *AluI* satellite/Nobe, and BvSat4) and six minisatellites (BvMSat4/5/6/7/8/10) constituted 7%. *B. lomatogona*. *B. nana* genome shared similar major satellite families with *B. lomatogona* genome, including pBC305, pHC8, pRN1, *HinfI*/Tantalos, BvSat4 and minisatellites BvMSat8. The proportion of *FokI* satellite/Dione and pAC34 was 13.5% and 4.7% of known satellites in *B. lomatogona*, respectively, however, they were not presented in *B. nana*. In this thesis, six new satellite families (BISat1-BISat6) were identified in *B. lomatogona* representing 16.9% of the all *B. lomatogona* satellites. The satellites pTS3, pTS4, pTS5, pTS100, pAp4, pAp11, pBC305, pRp34 and minisatellites BvMSat1, BvMSat7 were identified as the majority of satellite families in *P. procumbens* and *P. patellaris*. *C. quinoa* is a distant relative species compared with *Beta* and *Patellifolia* species, therefore, a different set of satellite families was identified. In particular, pBC1447 (Gao *et al.*, 2000) was the most abundant satellite family with 96.04% of all known satellites in *C. quinoa* and pAp11(Dechyeva *et al.*, 2003) presented very small proportion (1.02%). Four new satellite families (CqSat3-CqSat6) also formed the small proportion of *C. quinoa* satellites of nearly 3.0% (Ost, 2016).





**Figure 3.2: Percentage in the satellite fractions of *Beta*, *Patellifolia* and *C. quinoa* species analyzed by RepeatExplorer.**

Analyzed clusters representing repeat content of each species was shown on the left, green bars indicated clusters representing satellite families. Satellite families and their proportion were indicated on the right.

Table 3.3: Satellite families in *Beta*, *Patellifolia*, and *C. quinoa* species analyzed by *RepeatExplorer*

Satellite family	Satellite proportion [%]					
	<i>B. vulgaris</i>	<i>B. lomatogona</i>	<i>B. nana</i>	<i>P. procumbens</i>	<i>P. patellaris</i>	<i>C. quinoa</i>
pBV I/III/VI	11.8	-	1.94	-	-	-
pEV	68.7	-	-	-	-	-
pHC8	2.6	4.4	8.8	-	-	-
pHC28	-	11.7	-	-	-	-
pHT32	0.5	-	-	-	-	-
pHT36	-	10.0	1.7	-	-	-
pHT49	-	-	1.66	-	-	-
pRN1	-	6.8	20.37	-	-	-
pRN11	0.1	-	-	-	-	-
BvSat04	0.3	3.6	8.8	-	-	-
<i>Hinf</i> I satellite/Tantalos	2.8	1.0	0.51	-	-	-
<i>Fok</i> I satellite/Dione	1.3	13.5	-	-	-	-
<i>Alu</i> I satellite/Niobe	0.5	-	-	-	-	-
BvMSat1	3.8	-	-	0.8	1.3	-
BvMSat4	0.4	-	-	-	-	-
BvMSat5	1.2	-	-	-	-	-
BvMSat6	0.4	-	-	-	-	-
BvMSat7	1.4	1.8	-	2.1	2.6	-
BvMSat8	0.3	1.3	1.38	-	-	-
BvMSat10	0.3	0.3	-	-	-	-
pAp4	-	-	3.81	24.6	30.2	-
pAp11	-	-	-	9.6	7.5	1.02
pAp22	-	-	-	3.4	4.2	-
pAC34/ pRp34	3.3	4.7	-	4.1	4.5	-
pBC305	-	21.1	6.46	0.22	0.35	-
pBC1416	0.3	2.2	-	-	-	-
pBC1418	-	1.0	-	-	-	-
pBC1447	-	-	-	0.3	-	96.04
pTS3	-	-	-	4.71	6.24	-
pTS4	-	-	-	2.82	5.33	-
pTS5	-	-	-	10.16	6.42	-
pTS100	-	-	-	37.13	31.46	-
BISat1	<b>0.28</b>	<b>6.5</b>	<b>0.11</b>	-	-	-
BISat2	-	<b>4.3</b>	<b>0.003</b>	-	-	-
BISat3	<b>0.0065</b>	<b>2.6</b>	<b>0.13</b>	-	-	-
BISat4	<b>0.05</b>	<b>1.6</b>	<b>0.18</b>	<b>0.02</b>	<b>0.03</b>	<b>0.005</b>
BISat5	<b>0.02</b>	<b>1.5</b>	<b>1.83</b>	-	-	-
BISat6	<b>0.8</b>	<b>0.4</b>	<b>0.11</b>	-	-	-
BnSat1	<b>0.01</b>	<b>1.1</b>	<b>10.01</b>	-	-	-
BnSat2	-	<b>0.3</b>	<b>20.55</b>	-	-	-
BnSat3	<b>0.02</b>	<b>0.02</b>	<b>6.35</b>	-	-	-
BnSat4	<b>0.02</b>	<b>0.75</b>	<b>1.87</b>	-	-	-
BnSat5	<b>0.02</b>	<b>0.9</b>	<b>5.3</b>	-	-	-
BnSat6	<b>0.07</b>	<b>0.3</b>	<b>2.24</b>	-	-	-
CqSsat3	-	-	-	-	-	<b>1.19</b>
CqSsat4	-	-	-	<b>0.001</b>	<b>0.07</b>	<b>0.91</b>
CqSsat5	-	-	-	<b>0.0005</b>	<b>0.04</b>	<b>0.5</b>
CqSsat6	-	-	-	-	-	<b>0.34</b>

Note: “-“ not detected by *RepeatExplorer* or mapping. The proportion of the new satellite families BISat1-BISat6, BnSat1-BnSat6, and CqSat3-CqSat6 in other species (bold number) was estimated by mapping of sequence reads of reference species against each satellite sequence.

## 3.2 Satellite landscape in *Beta lomatogona*

### 3.2.1 Generation of a satellite overview in *Beta lomatogona*

Satellite DNA is an important component of plant genomes. In *B. lomatogona* this component makes up 6.57% of the genome. Published satellites such as pBC305 (Gao *et al.*, 2000), pHC28 and pHT36 (Schmidt *et al.*, 1993), FokI satellite/Dione (Zakrzewski *et al.*, 2010), pRN1 (Kubis *et al.*, 1997) were the most abundant *B. lomatogona* satellites. The six novel satellites identified in this thesis accounted for 16.9% of satellite repeats and 0.8% of the genome (Table 3.4).

The similarity search of the *RepeatExplorer* clusters using *Beta*-specific repeat database was performed resulting in 13 known satellites in *B. lomatogona* (Table 3.4, un-color field). These known satellite families were identified in 16 clusters (Table 3.4). The top three abundant satellite families in the *B. lomatogona* genome are pBC305 (~ 1%), FokI satellite/Dione (~ 0.6%) and pHC28 (~ 0.5%), followed by pHT36, pRN1, pAC34, pHC8, and BvSat4. Less than 0.1% of *B. lomatogona* genome, there are satellite families pBC1416, BvMSat7, BvMSat8, pBC11418, and *Hinf*I satellite/Tantalos. Most of satellite families' distributions are in all three sections of the genus *Beta*.

For an in-depth characterization of tandem repeat clusters, only clusters covering more than 0.01% of the genome of *B. lomatogona* were used. Read clusters with circular or star-like graphs indicated satellite repeats and were selected for further analysis. Within each cluster, corresponding reads were assembled to form of contigs. In most cases, the largest contigs correspond to representative satellite multimers. The largest contig was extracted and plotted against itself. If several monomers are present in a contig, parallel lines appear in the dotplot. According to the cluster graphs in *RepeatExplorer* output, a total of six clusters; including CL76, CL126, CL166, CL214, CL222, and CL345; were selected and analyzed in detail. For generation of satellite consensus sequence, a total of approximately 3 million *B. lomatogona* raw reads were mapped against the complete monomer of each satellite family. Six annotated satellite families were designated BISat1 – BISat6 (Beta lomatogona Satellite 1-6) in decreasing abundance and their features were summarized in Table 3.5.

**Table 3.4: Satellites in the *B. lomatogona* genome.** The names and sizes of satellite families as well as the plant species where the satellite was first described are given. The numbers of the clusters are sorted in ascending order in the *RepeatExplorer* output. The distribution of the satellites was marked with (+), non-occurrence was indicated with the hyphen (-). The localization of the satellites on chromosomes was investigated using FISH experiments. Yellow fields indicated the knowledge gaps which were fulfilled in this thesis.

Satellite family	Species	Monomer size [bp]	RE output (CL)	Genome proportion [%]	Distribution				Chromosomal position	Reference
					<i>Beta</i>	<i>Corollinae</i>	<i>Nana</i>	<i>Patellifolia</i>		
pBC305	<i>B. corolliflora</i>	448	CL14, CL60	0.969	-	+	-	-	dispersed	Gao <i>et al.</i> , 2000
FokI/Dione	<i>B. vulgaris</i>	130	CL61, CL87	0.617	+	+	+	-	dispersed	Zakrzewski <i>et al.</i> , 2010, 2014
pHC28	<i>B. corolliflora</i>	149	CL74, CL98	0.538	+	+	+	+	intercalary	Schmidt <i>et al.</i> , 1993
pHT36	<i>B. trigyna</i>	142	CL38	0.444	+	+	+	-	pericentromere, 4 chr pairs	Schmidt <i>et al.</i> , 1993; this thesis
pRN1	<i>B. nana</i>	322	CL73	0.313	+	+	+	-	pericentric/intercalary	Kubis <i>et al.</i> , 1997
pAC34	<i>B. corolliflora</i>	357	CL109	0.218	+	+	+	+	subtelomeric	Dechyeva <i>et al.</i> , 2008
pHC8	<i>B. corolliflora</i>	161	CL124	0.204	+	+	+	-	pericentric/ dispersed	Gindullis <i>et al.</i> , 2001
BvSat4	<i>B. vulgaris</i>	122	CL135	0.166	+	+	+	-	-	Zakrzewski, unpublished
pBC1416	<i>B. corolliflora</i>	266	CL182	0.099	-	+	-	-	-	Gao <i>et al.</i> , 2000
BvMSat07	<i>B. vulgaris</i>	30	CL203	0.081	+	+	+	-	intercalary	
BvMSat08	<i>B. vulgaris</i>	32	CL239	0.058	+	+	+	-	intercalary	Zakrzewski <i>et al.</i> , 2010
pBC1418	<i>B. corolliflora</i>	378	CL260	0.045	-	+	-	-	centromere, 2 chr pairs	Gao <i>et al.</i> , 2000; this thesis
HinfI/Tantalos	<i>B. vulgaris</i>	325	CL261	0.045	+	+	+	-	terminal regions	Zakrzewski <i>et al.</i> , 2010, 2014
BlSat1	<i>B. lomatogona</i>	171	CL76	0.296	+	+	+	-	pericentromere, chr 3,5,6,9	
BlSat2	<i>B. lomatogona</i>	90	CL126	0.198	-	+	+	-	dispersed	
BlSat3	<i>B. lomatogona</i>	190	CL166	0.121	-	+	+	-	dispersed	
BlSat4	<i>B. lomatogona</i>	276	CL214	0.074	+	+	+	+	dispersed	This thesis
BlSat5	<i>B. lomatogona</i>	313	CL222	0.07	-	+	+	-	pericentromere, chr 3,5,7	
BlSat6	<i>B. lomatogona</i>	315	CL345	0.013	+	+	+	-	subtelomere, chr 8	

**Table 3.5: Features of novel satellite DNA families in *B. lomatogona* genome**

Satellite family	Cluster	Monomer length [bp]	AT content [%]	Genome proportion [%]	Average identity to consensus [%]	No. of reads analyzed
Typical satellites						
BlSat1	CL76	171	74.8	0.296	91	1782
BlSat5	CL222	313	61.7	0.07	84.5	445
BlSat6	CL345	315	42.5	0.013	89.6	167
Non-typical satellites						
BlSat2	CL126	90	48.9	0.198	92.2	741
BlSat3	CL166	190	52.5	0.121	84.9	433
BlSat4	CL214	276	50	0.074	92.2	216

An overview of the known satellite as well as new satellite families in *B. lomatogona* is shown in Table 3.4, including the species in which the tandem repeat was first described, the size of monomer, the distribution in *Beta* and *Patellifolia* genera, and the chromosomal localization. In particular, the knowledge gaps were marked by yellow fields and these gaps have filled with this thesis.

Six satellite families BlSat1-BlSat6 are new and have not been characterized before. Out of the six, three are typical satellite repeats and three are associated with dispersed repeats, therefore these satellite families are described separately in section 3.2.2 and 3.2.3, respectively.

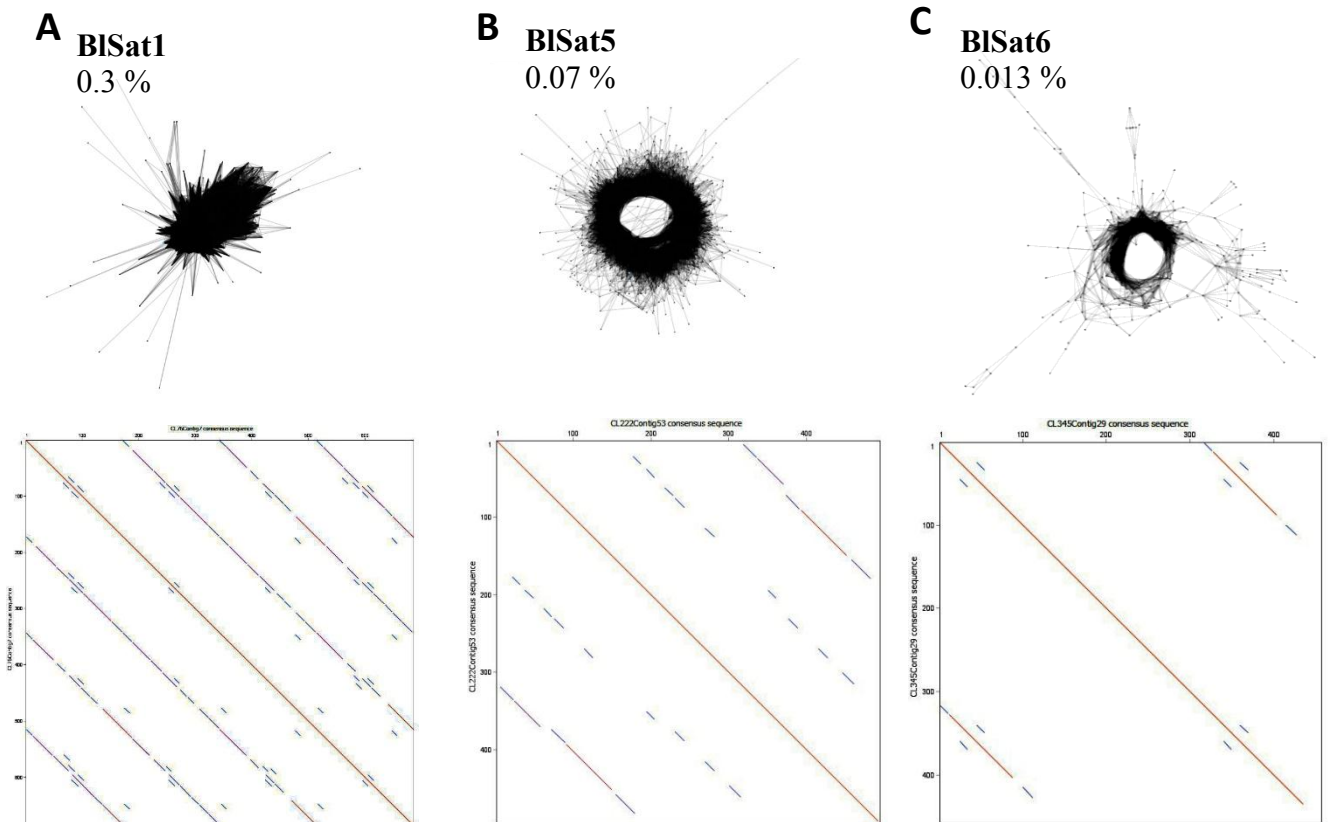
The satellite families pBC1418 and pHT36 are known satellites (Schmidt *et al.*, 1993; Gao *et al.*, 2000). However, the completion of the missing information regarding the distribution of these satellites in *B. lomatogona* genome is of interest (section 3.2.4).



### 3.2.2 The satellite families BISat1, BISat5 and BISat6 form large, chromosome-specific arrays

#### 3.2.2.1 Monomeric structure and organization based on bioinformatics and multimer cloning

BISat1, BISat5, and BISat6 indicated typical satellites supporting by star-like and circular shape of graphs (Figure 3.3). Cluster CL76 representing satellite family BISat1 contains reads equaling to nearly 0.3% of the *B. lomatogona* genome and analysis of the largest contig resulted in three complete monomers with the size of 171 bp (Figure 3.3A). The satellite BISat5 represented 0.07% of the genome and only one complete monomer could be annotated with its size of 313 bp (Figure 3.3B). The smallest selected cluster CL345, representing the satellite family BISat6, makes up only 0.013% of the genomic DNA. In the largest contig, only one complete monomer was identified with 315 bp in length (Figure 3.3C).



**Figure 3.3: Circular and star-like graph shape and dotplot of putative satellite clusters.**

(A) Star-like graph shape and dotplot of cluster CL76 representing BISat1, (B) Circular graph shape and dotplot of cluster CL222 representing BISat5, (C) Circular graph shape and dotplot of cluster CL345 representing BISat6. The continuous lines in the dotplot correspond to a high homolog value. Nucleotide positions are recorded on the X and Y axes of the dotplots allow the estimation of the monomer length.

Schematic representations of the bioinformatic monomer consensus sequences are shown in Figure 3.4 with primer binding sites and selected restriction enzymes that can be used for the isolation of satellite monomers. The software *Oligo Analysis* was employed to determine the potential restriction enzymes cutting only once in each monomer. The information of AT content and identity value from Table 3.5 were combined to characterize three satellite families B1Sat1, B1Sat5, and B1Sat6.

The 171 bp B1Sat1 monomer has highest AT content (74.8%). Restriction enzyme *MaeII* with recognized sequence CTAG was found to cut once in B1Sat1 monomer. In the mapping of reads to consensus sequence, high identity was observed (91%). The B1Sat1 monomer contains duplications of two heptamers (AAGTTGT and TTGAAAA) and one octamer (GGCCATTT) (Figure 3.4A).

The B1Sat5 monomer is 313 bp long. This satellite has a moderate AT content (61.7%), where identity value is lowest (84.5%). B1Sat5 is comprised of two subunits, subunit 1 of 156 bp and subunit 2 of 157 bp, which show 69.4% identity. *AluI* was found to cut once in each subunit, but at different positions (Figure 3.4B).

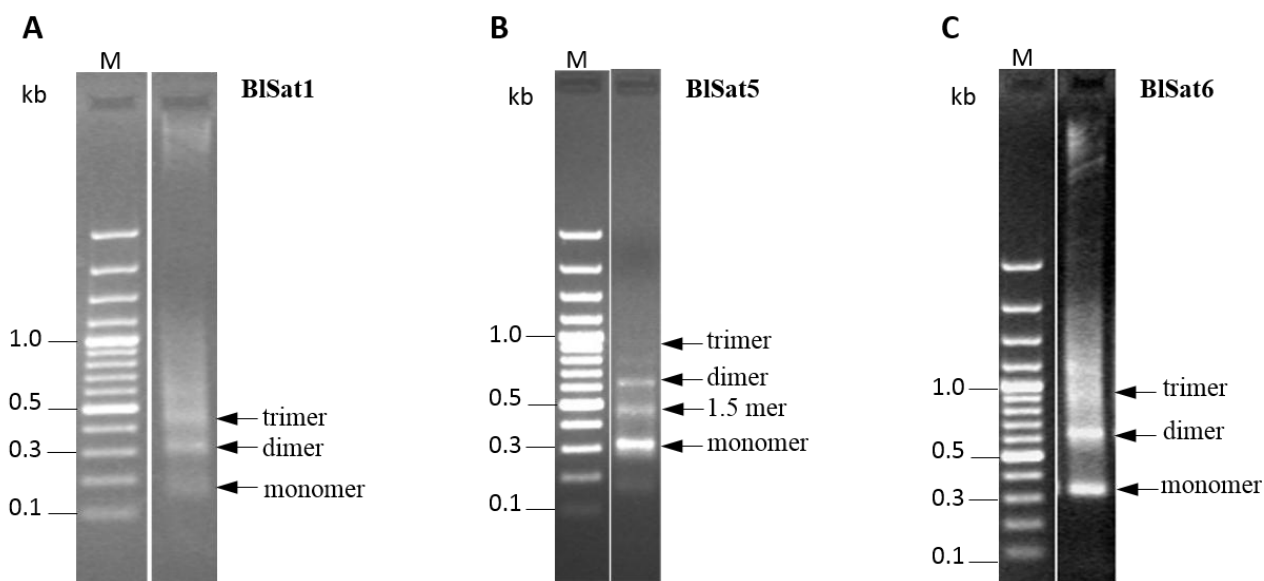
The satellite family B1Sat6 has a simple structure where the monomer length is 315 bp (Figure 3.4C). The lowest AT content of 42.5% and identity of 89.6% were observed for this family. There are three CCGG sequences which can be used to investigate methylation. It was obtained that *BsmI* is a suitable restriction enzyme resulting in a ladder-like pattern in Southern hybridization

In addition, the sequence motif CCGG was shown as the recognition site of methylation-sensitive isoschizomers *HpaII* and *MspI*. *HpaII* only cuts unmethylated CCGG motifs, whereas *MspI* cuts CCGG and also CC<sup>met</sup>GG. Therefore, the *MspI/HpaII* isoschizomer pair has been used to investigate the differences between methylation state of the inter cytosine in CCGG motif. This sequence motif was detected in the satellite family B1Sat6 (black rectangle in Figure 3.4C).



putative satellite (Figure 3.5). The amplicon of multimers is observed only when monomers of the repeat are arranged in the head to tail fashion. Products up to the tetramer were visible for BISat1 and BISat6. For BISat5 not only monomer, dimer and trimer bands, but also submonomer bands, i.e. a half, one and a half, two and a half monomer bands were obtained.

The lengths of PCR fragment correspond to the respective satellite monomers, as each satellite-specific primer pair was designed adjacently in the opposite direction. For example, the PCR fragment of the monomer of BISat5 is equal to the computationally detected monomer. PCR fragments are approximately 7 bp and 3 bp shorter than the monomers of BISat1 and BISat6, respectively, because of the distance between forward and reverse primers.



**Figure 3.5: Ladder-like pattern in agarose gel after PCR with satellite-specific primers**

The agarose gel electrophoresis was carried out with gels of 2 % agarose. The positions of up to trimer are marked with arrows. (A) Ladder-like pattern of the BISat1 satellite, (B) BISat5 satellite, and (C) BISat6 satellite, M: marker GeneRuler 100 bp Plus DNA Ladder.

In order to obtain the exact sequence of the determined satellite monomers as well as to prepare probes for Southern and FISH hybridization experiments, PCR fragments corresponding to monomers, dimers and/or trimers were cloned and sequenced.

All plasmids carrying an insert of the expected size that were selected for sequence analysis. The clones with high identity to the bioinformatically determined consensus satellite sequences were chosen as probes for Southern and FISH hybridizations. In general, the similarity between clones

and satellite consensus sequence was high, ranging from 78.5% to 93.9% (Table 3.6 and Supplementary Figure S1, S2, S3). All probes contain at least one complete monomer.

**Table 3.6: Clones of three satellites B1Sat1, B1Sat5, and B1Sat6 in *B. lomatogona*.**

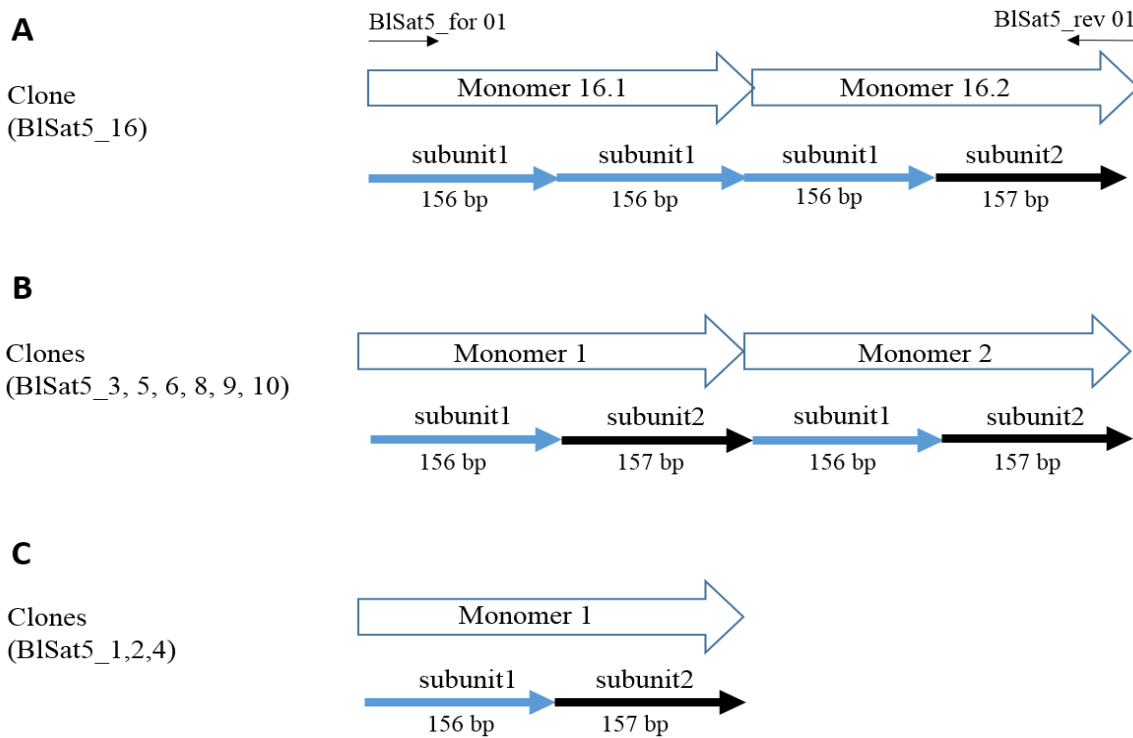
For each satellite family, the clone used for hybridization experiments was marked with asterisk

Satellite family	Clone name	Length of insert [bp]	Number of complete monomers	Identity to consensus [%]
<b>B1Sat1</b>	B1Sat1_1*	204	1	93.9
<b>B1Sat5</b>	B1Sat5_1	311	1	85.0
	B1Sat5_2	467	1	87.1
	B1Sat5_3	628	2	87.4
	B1Sat5_4*	645	1	89.4
	B1Sat5_5	627	2	92.9
	B1Sat5_6	621	2	92.0
	B1Sat5_8	628	2	86.3
	B1Sat5_9	626	2	91.4
	B1Sat5_10	622	2	90.8
	B1Sat5_16	628	2	69.8
<b>B1Sat6</b>	B1Sat6_1	612	2	83.3
	B1Sat6_2	892	3	80.1
	B1Sat6_3	627	2	82.5
	B1Sat6_4	623	2	81.9
	B1Sat6_5	926	3	78.5
	B1Sat6_6*	624	2	85.5
	B1Sat6_7	941	3	82.7

For the satellite family B1Sat1, only one clone with a complete monomer was obtained. The sequence identity of the monomer and the B1Sat1 consensus (B1Sat1\_in silico) is 93.9%. A recognition sequence for restriction enzyme *MaeII* (CTAG) is located at the nucleotide position 16-20 (Supplementary Figure S1).

A total of 17 complete monomers of B1Sat5 were evaluated from the 10 selected clones and showed an identity of 84.4% in a paired comparison (Supplementary Figure S2). The first recognition site of *AluI* (AGCT) is at position 72-75 and this site is conserved. However, the second recognized position of *AluI* at position 166-169 varies significantly between monomers.

As indicated in Figure 3.4B the satellite family B1Sat5 showed a higher order structure. With the monomer sequences, it is possible to achieve more detailed insight into the higher order structure of this satellite family. In a total of 10 analyzed clones, the clone B1Sat5\_16 shows difference in the order of the subunits, in which one monomer (Monomer 16.1) comprises two subunits sub1 and the other monomer (Monomer 16.2) was structured from two subunits sub1 and sub2 (Figure 3.6A). Nine remaining clones were observed with higher order structure of two subunits sub1 and sub2 (Figure 3.6B, C). This result indicated that there are two variants of higher order structure in B1Sat5 satellite family and the monomers including subunits sub1 and sub2 are the predominant monomer structure in the *B. lomatomogona* genome.



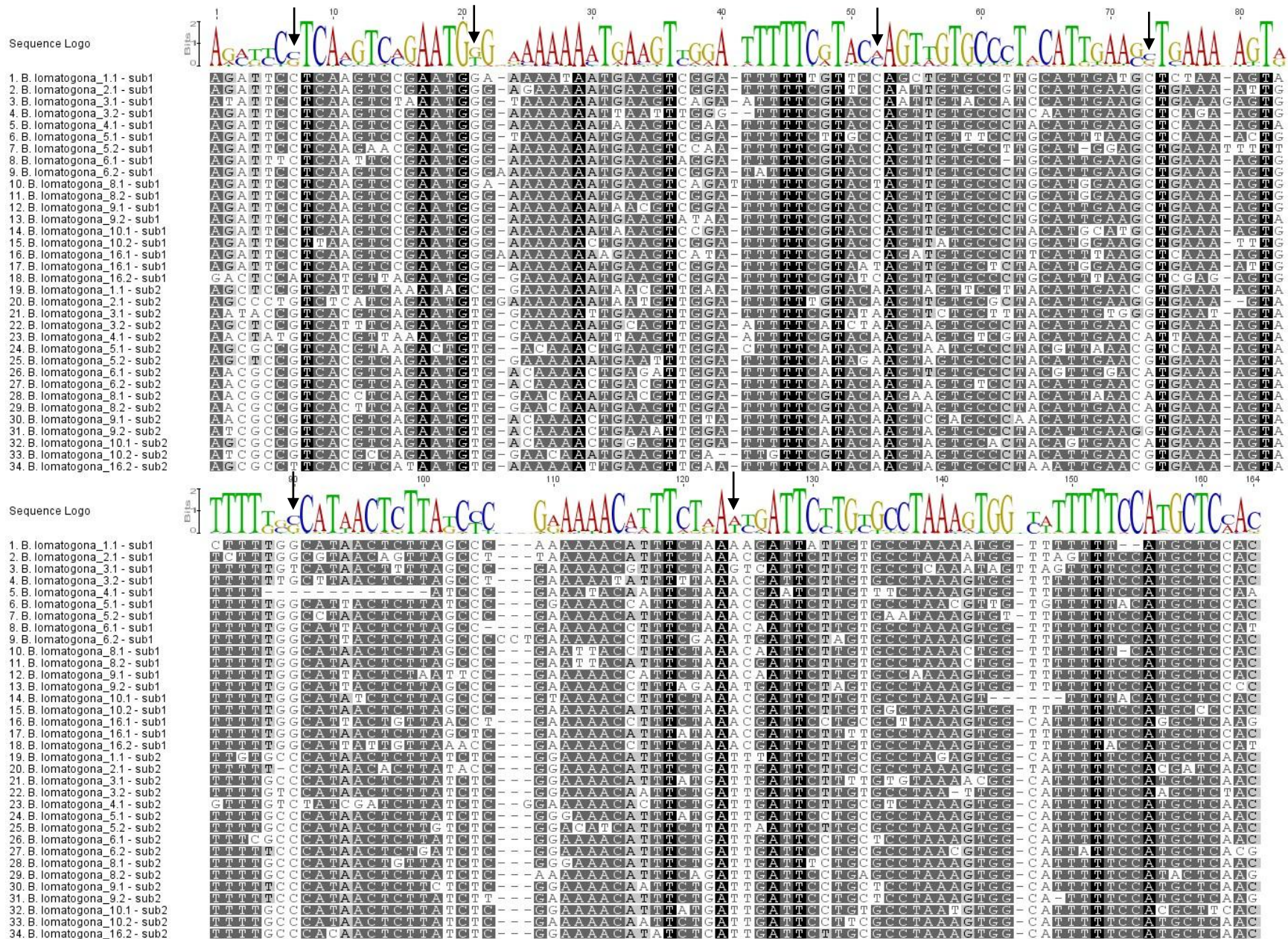
**Figure 3.6: Higher order structure of B1Sat5 satellite from analyzed clones**

The monomers were indicated with big arrows whereas subunit1 and subunit2 were marked with linear blue arrows and black arrows, respectively.

In this satellite, subunit 1 and subunit 2 have the size of 156bp and 157 bp, respectively. The sequence similarity between the subunits sub1 was 83.6%, whereas that between the subunits sub2 was slightly lower (83.5%). The comparison of the subunit1 and subunit 2 from all 17 B1Sat5 monomers showed lower similarity of 77.4%. There are 6 nucleotide positions differing between subunit 1 and subunit 2 (arrows in Figure 3.7), such as position 7 where subunit 1 includes C and

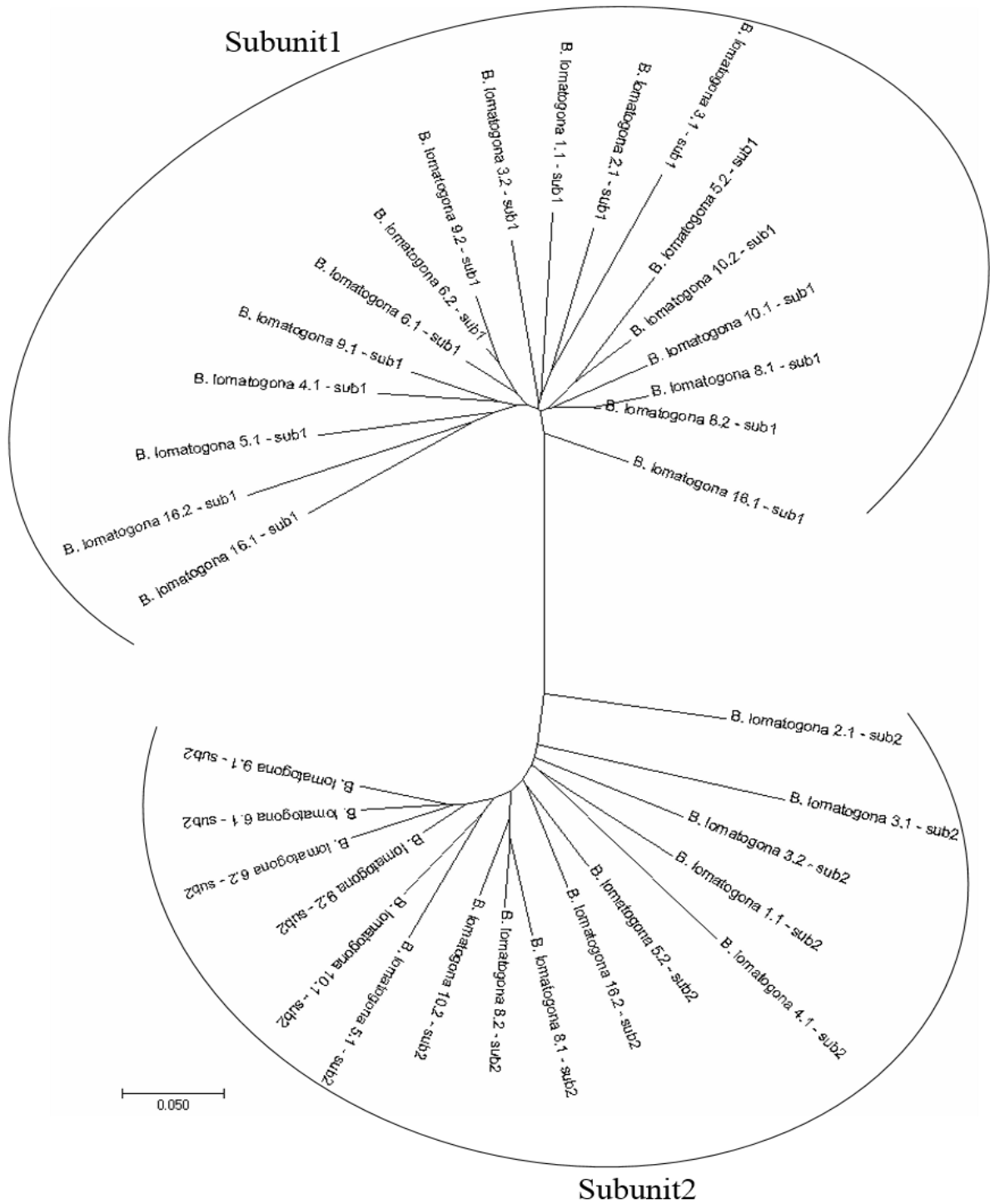
A, and subunit 2 includes G and T. This differences reveal the divergence of subunit 1 and subunit 2 in the evolution of the B1Sat5 satellite. To further investigate the relationship between two B1Sat5 subunits, a neighbour-joining analysis was carried out. The result clearly revealed two major variants (Figure 3.8). One included all subunits sub1, and the other is comprised of all subunits sub2.

Sequenced monomers of satellite B1Sat6 showed a pairwise identity of 79.8%, which is the lowest identity of the six identified satellites. Of 17 monomers analyzed in seven clones, the recognition sequence GAATGCN of restriction endonuclease *BsmI* is quite conserved between monomers (Supplementary Figure S3).



**Figure 3.7: Sequence alignment of two subunits of satellite BISat5.** All subunits (subunit 1 and subunit 2) extracted from monomers of BISat5 satellite were aligned using the MUSCLE algorithm in *Geneious*. The higher value of sequence identity, the darker the filled shading.



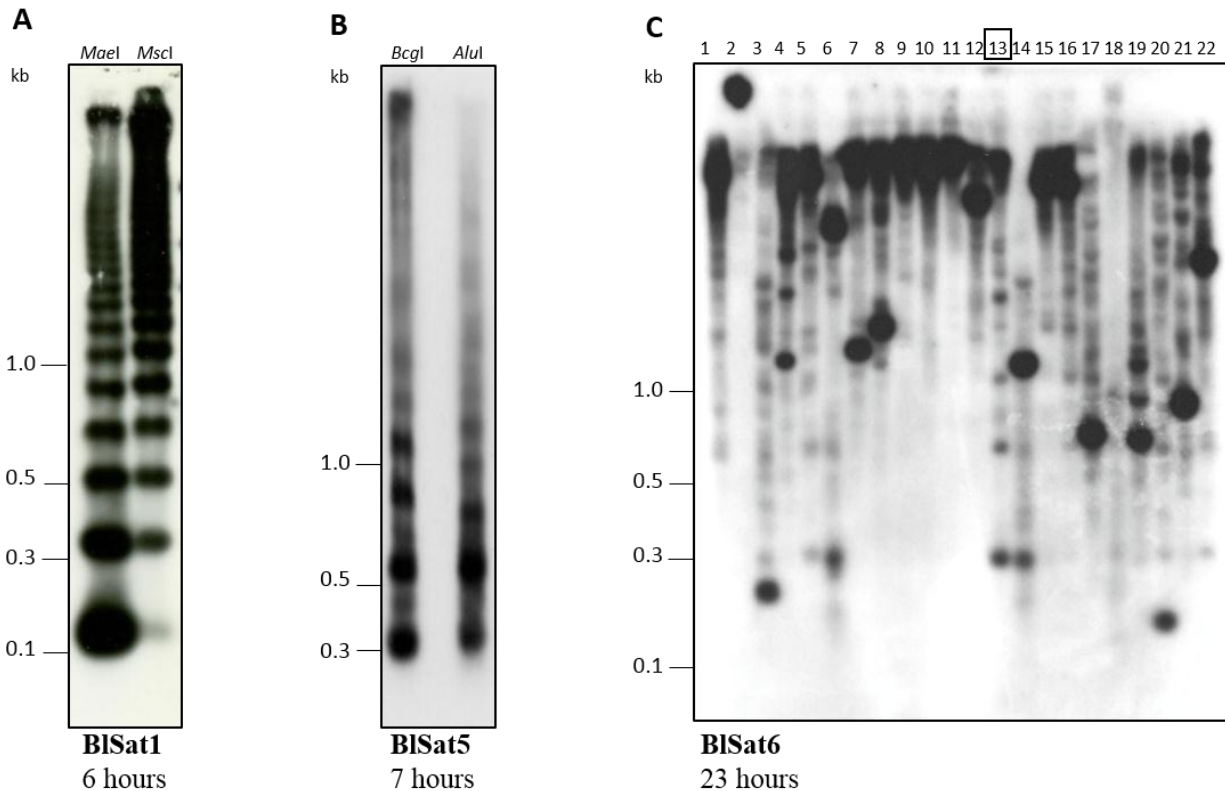


**Figure 3.8: Dendrogram representing the phylogenetic relationship of the subunits sub1 and sub2 of the satellite family BISat5.**

The neighbor-joining tree demonstrates that the subunits clearly fall into two distinct variants.

### 3.2.2.2 Genomic organization of B1Sat1, B1Sat5 and B1Sat6 in *Beta lomatogona*

In order to verify the tandem arrangement of B1Sat1, B1Sat5, and B1Sat6, their genomic organization in *B. lomatogona* was investigated by Southern hybridization (Figure 3.9). For B1Sat1 and B1Sat5 satellites, restriction enzymes were chosen based on restriction sites identified in the consensus sequences. For B1Sat6, this did not lead to the desired outcome. Therefore, a list of 22 enzymes was tested. The restricted DNA was separated, blotted and hybridized with satellite-specific probes.



**Figure 3.9: Genomic organization of *B. lomatogona* tandemly repeated sequences**

The digested DNA was separated in blot gels with agarose concentration of 1.2 % for (B) and 2.0 % for (A), (C). The exposition time was 6 hours for B1Sat1 (A), 7 hours for B1Sat5 (B), and 23 hours for B1Sat6 (C). The following restriction enzymes were used to digest genomic DNA of *B. lomatogona*: (1) *Bsm*AI, (2) *Mae*II, (3) *Mse*I, (4) *Msp*I, (5) *Apa*LI, (6) *Bst*NI, (7) *Stu*I, (8) *Nsi*I, (9) *Msc*I, (10) *Nde*I, (11) *Xho*I, (12) *Dra*I, (13) *Bsm*I, (14) *Bse*GI, (15) *Bam*HI, (16) *Apa*I, (17) *Hin*FI, (18) *Fok*I, (19) *Mae*I, (20) *Rsa*I, (21) *Mbo*I, and (22) *Alu*I.

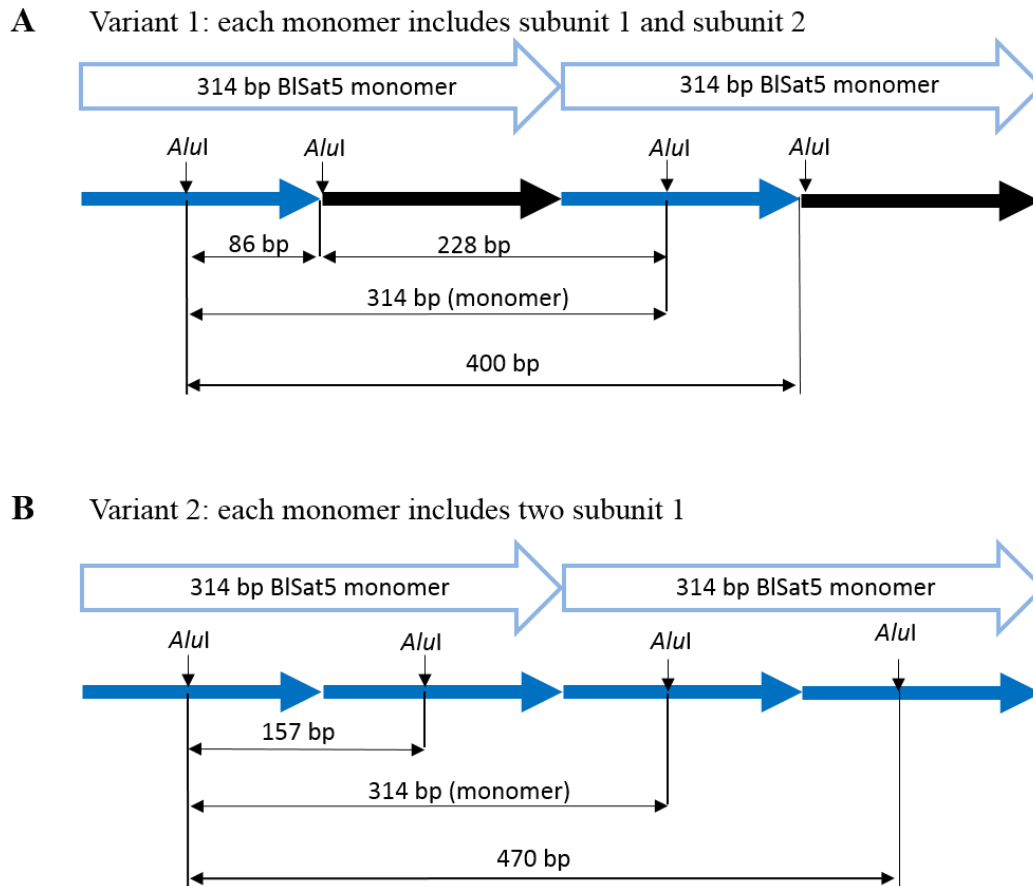
For B1Sat1, two enzymes were tested and both enzymes resulted in the typical ladder-like pattern for satellite DNA. *B. lomatogona* genomic DNA restricted by *Mae*I showed a strong ladder-like banding pattern where at least a 15-mer was observed. *Msc*I endonuclease gave weak signal corresponding to monomer size and stronger signal for multimers and high molecular weight

fragments (Figure 3.9A) after 6 hours of exposure. Accordingly, *MaeI* was chosen for further investigation of this satellite in the genus *Beta*.

A ladder-like pattern was obtained after hybridization of B1Sat5 probe with *B. lomatogona* genomic DNA restricted by *BcgI* and *AluI* (Figure 3.9B). Interestingly, a higher order structure was reflected in B1Sat5 autoradiograph. As described in section 3.2.2.1, B1Sat5 monomer is composed of two subunits. *BcgI* cuts once in subunit 2 while *AluI* cuts once in each subunit. Therefore, in addition to main bands of monomers and multimers, several intermediate bands were visible. The *AluI* restricted map of two B1Sat5 variants is shown in Figure 3.10. The variant 1 including one subunit sub1 and one subunit sub2 after restricted with *AluI* resulted in monomer band (314 bp) and two additional bands (86 bp and 228 bp) (Figure 3.10A). The variant 2 including two subunits 1 after restricted with *AluI* released monomer band (314 bp) and one additional band (157 bp). This was clearly observed in Figure 3.15B, in which there were three additional bands between monomer and multimer bands.

With the restriction endonuclease *BsmI* (Figure 3.9C, lane 13) the tandem-like organization of the satellite family B1Sat6 was easily observed. The strongest signal is approximately 300 bp, corresponding to the monomer size. Further signals from dimer up to pentamer were also visible. The ladder-like pattern of B1Sat6 was observed clearer in Figure 3.15C.

The abundance of tandem repeats is proportional to the signal strength in relation to exposition time and loaded DNA. However, many other factors can affect the estimation of the abundance, such as the age of radioactively labeled nucleotides dATP and dCTP or the AC content in the satellites. Of three conventional satellites, B1Sat1 is indicated to be the most abundant satellite of which very strong signals were detected after only 6-hour exposition. It was followed by B1Sat5 with a 7-hour exposure. B1Sat6 is not the prominent satellite in the genome of *B. lomatogona*, as the signals were visible after the longer exposure time of 23 hours.



**Figure 3.10: Schematic restriction map of the two consecutive BISat5 monomers**

The subunit 1 and subunit 2 are indicated with blue arrow and black arrow, respectively. BISat5 monomers with internal subunit 1 and subunit 2 (A) and with only subunit 1 (B) were restricted with *AluI* resulting in monomer band of 3114 bp and three additional bands of 86 bp, 157 bp, and 228 bp.

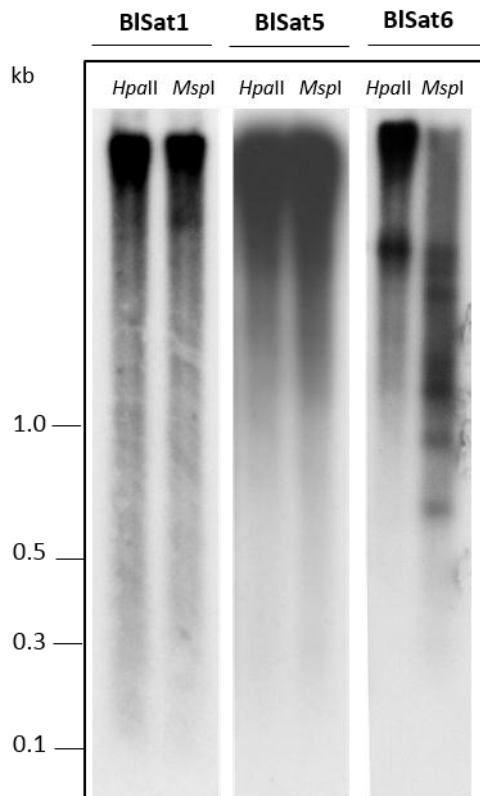
### 3.2.2.3 Methylation of BISat1, BISat5 and BISat6 satellites

In order to investigate the degree of methylation of *B. lomatogona* satellites, genomic *B. lomatogona* DNA was restricted with the methylation-sensitive enzymes *HpaII* and *MspI*. These enzymes are isoschizomers with the restriction site CCGG which are presented in two of three typical satellite consensus sequences, except for BISat5. Nevertheless, BISat5 satellite family also contains CCGG sites in diverged monomers. *HpaII* cuts only unmethylated CCGG sequences, whereas *MspI* is able to tolerate methylation of the internal cytosine.

Among the three satellites, the difference between *HpaII* and *MspI* restriction was only observed in the BISat6 family (Figure 3.11). *MspI* restriction is able to release a mean ladder pattern, while *HpaII* restriction is reduced with a strong signal at a high molecular weight and a moderate signal at approximately 2 kb long fragment corresponding to the satellite heptamer. This revealed that

BISat6 has a high degree of inner cytosine methylation of CCGG sites. Furthermore, genomic DNA restricted with *MspI* released DNA bands corresponding to the satellite multimer instead of the monomer, this indicates that the internal methylation cytosine is not present in every BISat6 monomer.

BISat1 and BISat5 satellites showed similar hybridization pattern, both enzymes produced a smear, but no ladder-like pattern was released. This result indicated that dispersed unmethylated CCGG sites are present in the sequences of these satellites. They may contain CCGG sites with the methylation of both inner and outer cytosine. BISat5 hybridization produces relatively short smear in both enzymes, this suggested a less frequent presence of CCGG sites in the satellite, and these CCGG sites are unmethylated and partly methylated in few monomers.



**Figure 3.11: Southern hybridization of three satellites to *B. lomatogona* genomic DNA restricted with the methylation sensitive enzymes *HpaII* and *MspI*.**

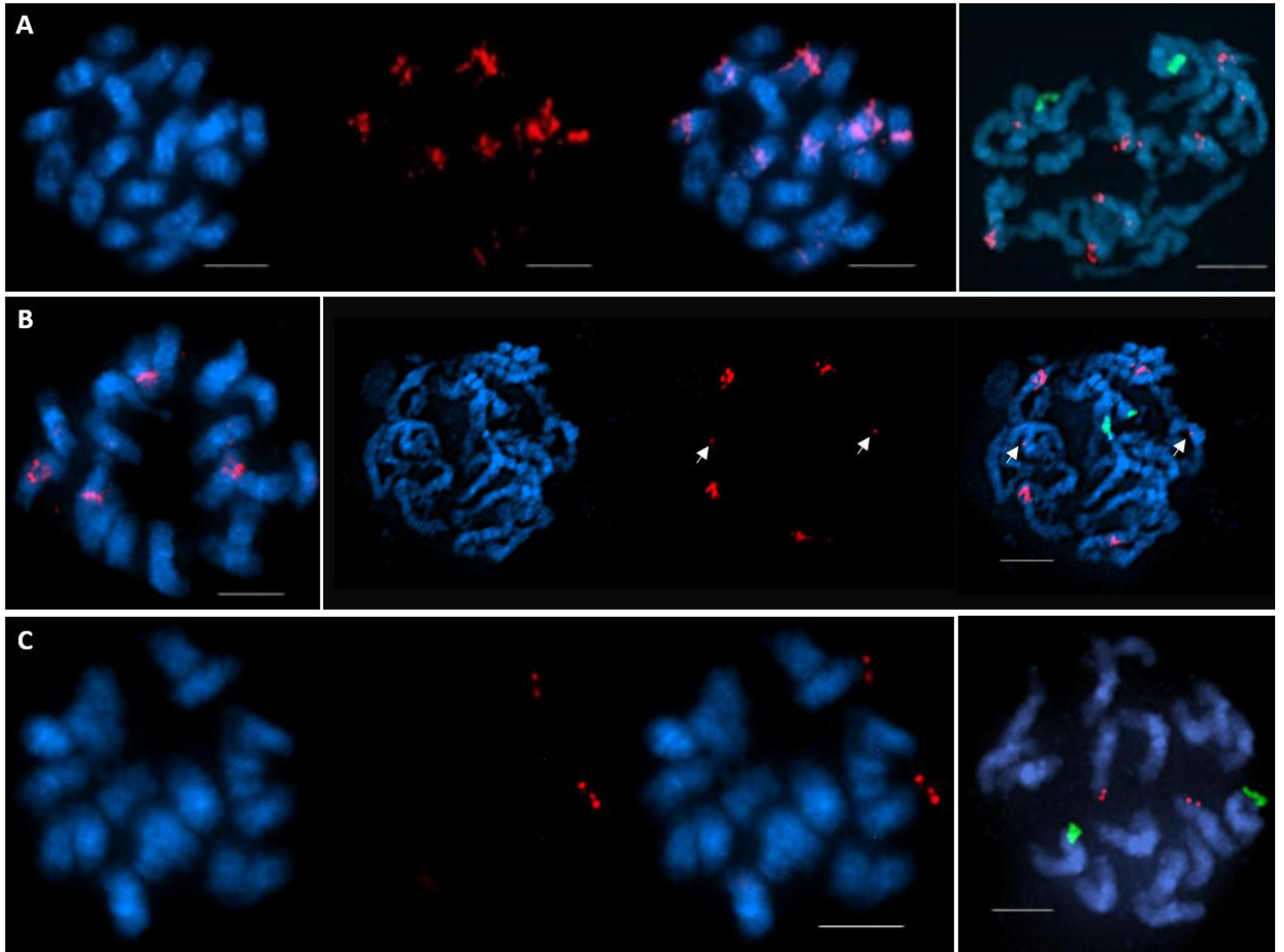
#### 3.2.2.4 Localization of B1Sat1, B1Sat5 and B1Sat6 along *Beta lomatogona* chromosomes

In order to localize the identified satellites along the *B. lomatogona* chromosomes, metaphase preparations were hybridized with biotin-labeled B1Sat1, B1Sat5 and B1Sat6 probes. Ribosomal genes serve as chromosomal landmarks for chromosome 1 and 4 in *Beta* species (Schmidt *et al.*, 1994). Therefore, preparations were hybridized with 18S-5.8S-25S rDNA probe to obtain more information of chromosome-specific *B. lomatogona* satellites. *B. lomatogona* has the chromosome number of  $n = 9$ , and 18S-5.8S-25S rRNA genes locate on the short arm of the chromosome 1 (green signals in Figure 3.12).

The FISH images of chromosome preparation hybridized with B1Sat1-specific probe show strong signals on two chromosome pairs and weaker signals on two other pairs (Figure 3.12A, left panel). At the higher resolution of a prometaphase spread, B1Sat1 signals are visible in the pericentromeric and intercalary regions of the chromosomes (Figure 3.12A, right panel). Two chromosomes show signal in the pericentromeric regions. The signals of B1Sat1 do not localize on chromosome 1. Very strong signals indicate the large satellite-typical arrays of B1Sat1. However, the signals strength was in a broad range, which indicates that the B1Sat1 arrays are not uniform in size on the four chromosomes.

Satellite family B1Sat5 is likely localized in the pericentromeric regions of three chromosome pairs. The B1Sat5 probe were hybridized strongly on two pairs (Figure 3.12B, left panel), of which the strong signals on one pair show an additional adjacent small signal. Weak signals were detectable on one other pair (minor site indicated by arrows in Figure 3.12B, right panel). This satellite might associate with the centromere, however it can not conclude before performing Chromatin immunoprecipitation (CHIP) experiments. B1Sat5 does not co-localize with the 18S-5.8S-25S rRNA genes on chromosome 1.

B1Sat6 is localized in the subtelomeric regions of one chromosome pair (Figure 3.12C). Similarly, this satellite does not co-localize with the 18S-5.8S-25S rRNA genes on chromosome 1, however specifically marks a chromosome arm in *B. lomatogona*.



**Figure 3.12: Chromosomal localization of classic satellites BISat1, BISat5 and BISat6 along chromosomes of *B. lomatogona*.**

Blue fluorescence shows DAPI-stained DNA, whereas red fluorescence indicates satellite DNA. Green signals reveal the position of 18S-5.8S-25S rRNA genes. (A) Signals of BISat1 on metaphase and prometaphase nuclei; (B) The centromeric localization of BISat5, arrows show the minor signals; (C) The subtelomeric satellite BISat6. Scale bar is 5  $\mu$ m.

### 3.2.2.5 BISat1, BISat5 and BISat6 are assigned to *Beta lomatogona* chromosomes 3, 5, 6, 7, 8 and 9

In order to determine the chromosome of chromosome-specific satellites, co-localization with chromosome-specific BAC probes was performed. The number of probes simultaneously hybridized should not exceed four because of the limitation of the separately detectable fluorochromes. The signals were not always clearly distinguishable from one to another even with highly selective bandpass filters. In principle, if probe signals are strong and have a relatively broad emission spectrum, it can break into other channels and thus lead to background fluorescence or even to the masking of other specific signals. The fluorochromes for labeling and detecting of the simultaneously hybridized probes were selected to avoid the overlaps of the emitted light. A

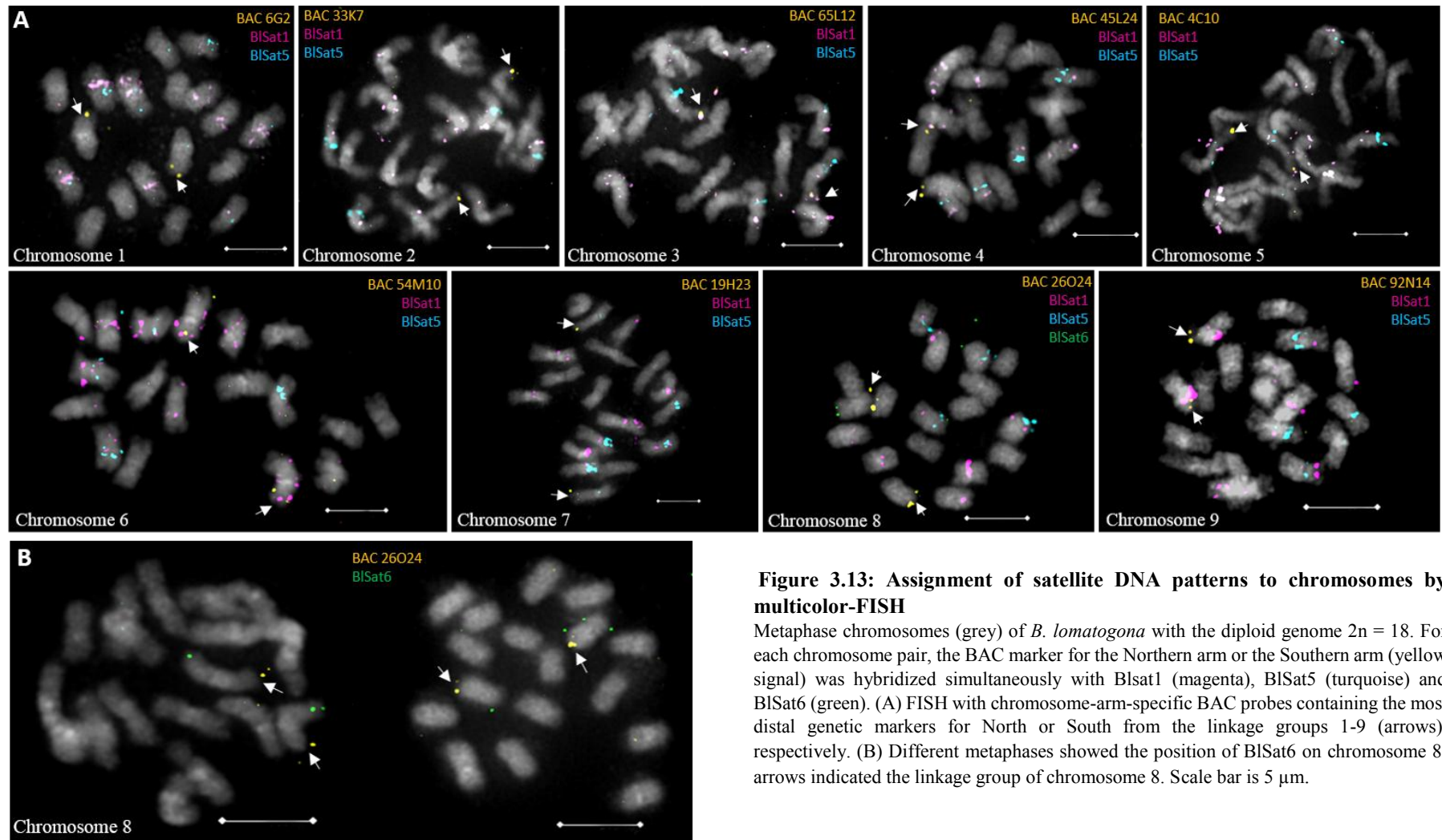
collection of nine chromosome arm-specific BAC probes was selected, and an individual BAC probe combined with the three satellite probes was hybridized on metaphase chromosomes in one FISH experiment (Figure 3.13 and Figure 3.14).

The signals of the probe B1Sat1 are detectable on four pairs of chromosomes which are 3, 5, 6, and 9. The hybridization pattern on chromosome 9 shows the strongest signal in pericentromeres indicating large arrays of B1Sat1. On chromosome 6, the signals of B1Sat1 occur at the pericentromeric and distal regions of both arms. Weaker signals on chromosomes 3 and 5 also present in the pericentromeric region, and additional stronger signals are detected in distal region of the North arm of chromosome 3 (Figure 3.14, magenta).

The signal intensities of the family B1Sat5 can be detected on three pairs of chromosomes 3, 5, and 7. Strong signals were visible in the pericentromeric regions of chromosomes 3 and 5. Interestingly, both B1Sat5 and B1Sat1 localize on two pairs of chromosomes 3 and 5. The signals of B1Sat1 are most likely in the pericentromeric and intercalary regions whereas the hybridization patterns of B1Sat5 tend to be associated in the centromeres. Therefore, the signals of B1Sat1 and B1Sat5 are not overlapping. A weak signal of B1Sat5 was detected in the pericentromeric heterochromatin of chromosome 7 (Figure 3.14, turquoise).

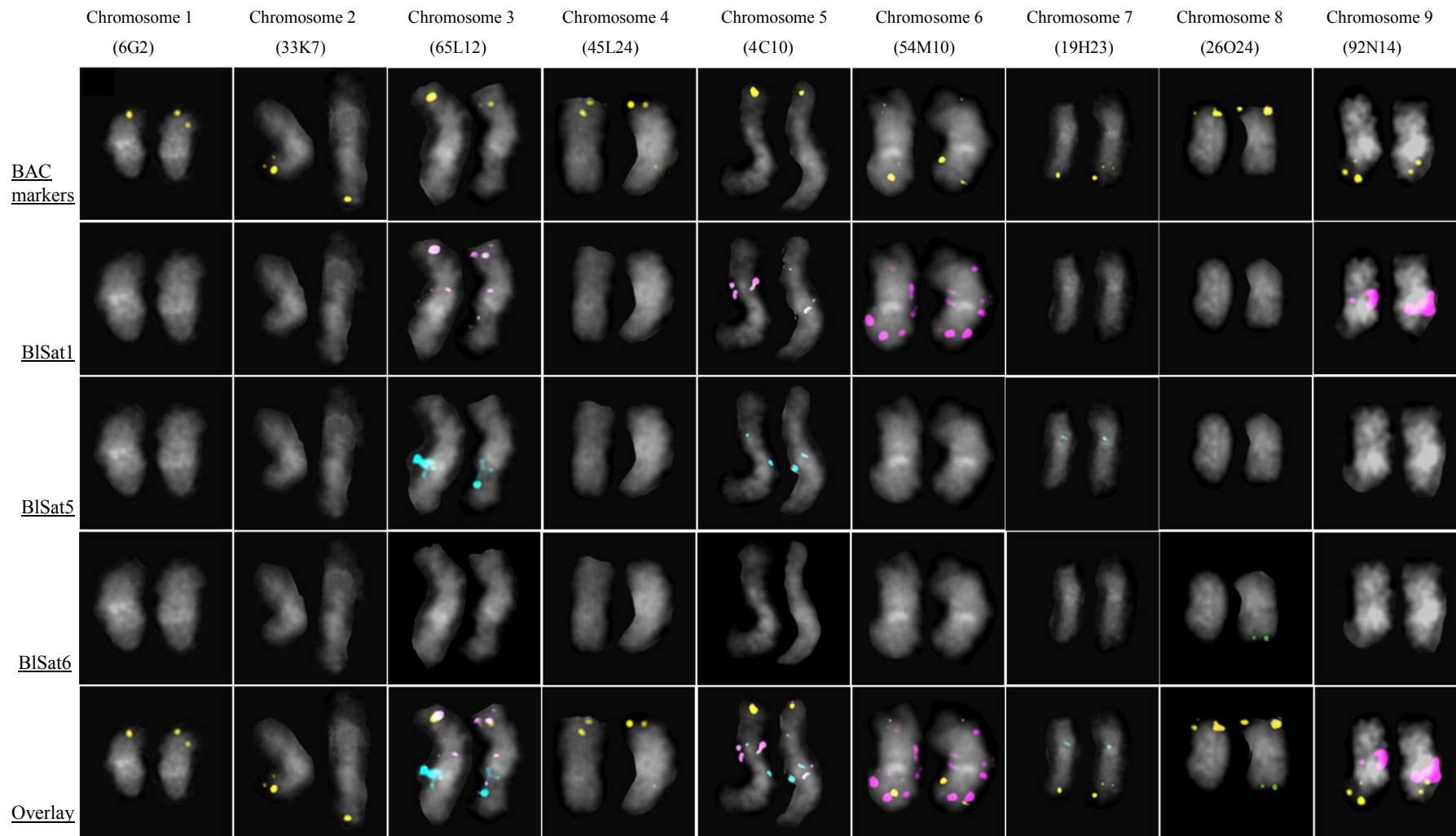
The signals of B1Sat6 were detected in subtelomeric heterochromatin of chromosome 8. In Figure 3.13A only one chromosome of the pair carries the B1Sat6 signal and the signal occurs in the South arm of chromosome 8. Therefore, two metaphases with a double signal of B1Sat6 were shown as evidence for B1Sat6 position (Figure 3.13B).





**Figure 3.13: Assignment of satellite DNA patterns to chromosomes by multicolor-FISH**

Metaphase chromosomes (grey) of *B. lomatogona* with the diploid genome  $2n = 18$ . For each chromosome pair, the BAC marker for the Northern arm or the Southern arm (yellow signal) was hybridized simultaneously with BIsat1 (magenta), BIsat5 (turquoise) and BIsat6 (green). (A) FISH with chromosome-arm-specific BAC probes containing the most distal genetic markers for North or South from the linkage groups 1-9 (arrows), respectively. (B) Different metaphases showed the position of BIsat6 on chromosome 8, arrows indicated the linkage group of chromosome 8. Scale bar is 5  $\mu\text{m}$ .



**Figure 3.14: Arrangement of all *B. lomatogona* chromosomes combined with BI Sat1, BI Sat5 and BI Sat6 satellites**

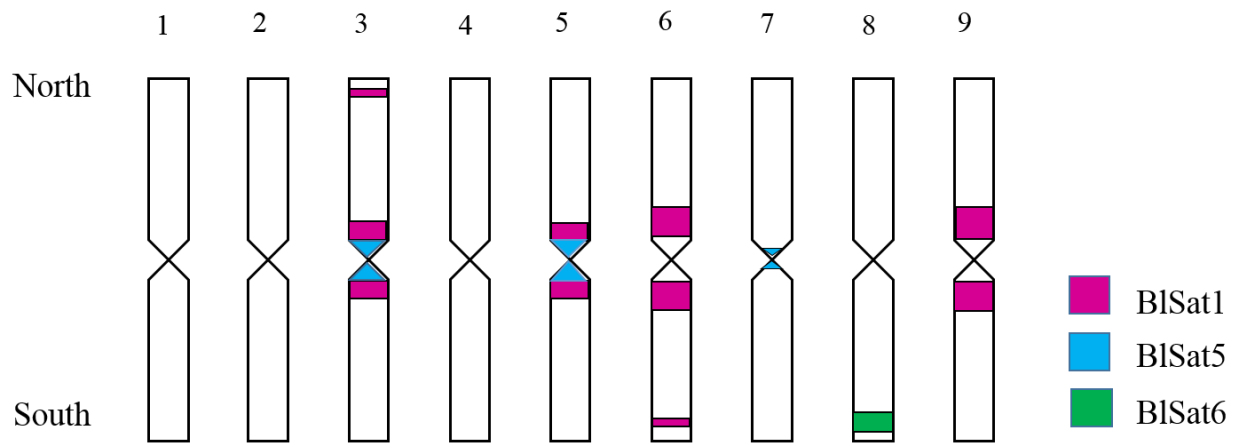
*B. lomatogona* chromosome pairs labeled with the North or South BACs representing a set of 18 chromosomes, combining with BI Sat1 (magenta), BI Sat5 (turquoise) and BI Sat6 (green). The single chromosomes shown are derived from different metaphases, so that the sequence intensities among them are only partially comparable.

The chromosomal distribution of three satellites B1Sat1, B1Sat5 and B1Sat6 is summarized in Table 3.7 and Figure 3.15. The signal strength of satellite family B1Sat1 varies, indicating that B1Sat1 is organized in both large and small arrays. Satellite family B1Sat5 is likely associated with the centromeres of chromosome 3, 5, and 7, whereas B1Sat6 is South-arm-specific of chromosome 8.

**Table 3.7: Chromosomal distribution of B1Sat1, B1Sat5 and B1Sat6 along *B. lomatogona* chromosomes**

The signal strength is indicated as strong (+++), moderate (++) and weak (+). No signal is designated with the hyphen mark (-)

Satellite	Chr1	Chr2	Chr3	Chr4	Chr5	Chr6	Chr7	Chr8	Chr9
B1Sat1	-	-	++	-	++	+++	-	-	+++
B1Sat5	-	-	+++	-	++	-	+	-	-
B1Sat6	-	-	-	-	-	-	-	++	-



**Figure 3.15: Schematic karyogram of *B. lomatogona***

The positions of the satellite families B1Sat1, B1Sat5, and B1Sat6 were indicated, but are not to scale.

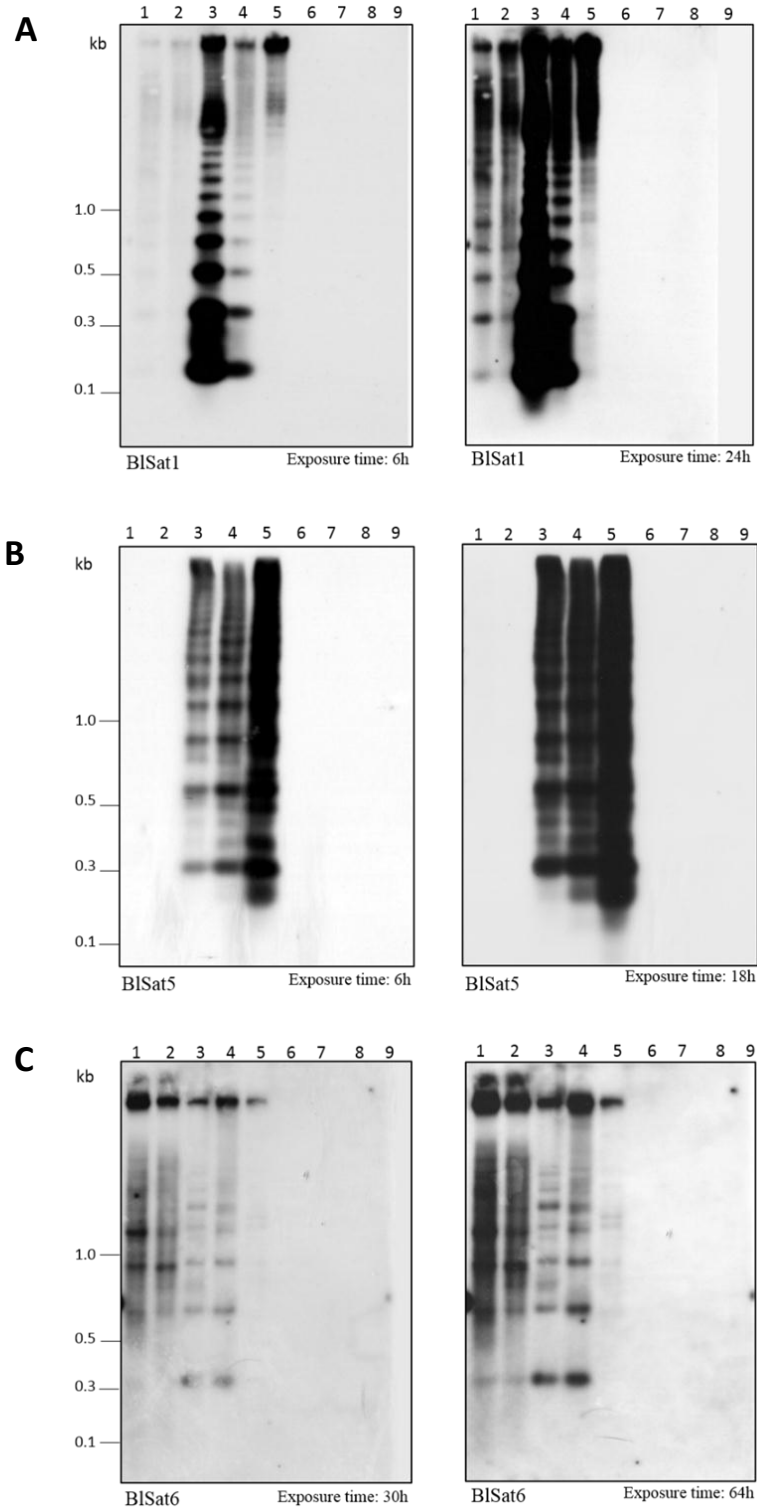
### 3.2.2.6 Distribution of *B. lomatogona* satellite B1Sat1, B1Sat5 and B1Sat6 in the genera *Beta*, *Patellifolia*, and related species

The abundance and distribution of the *B. lomatogona* satellite families in related species was investigated by Southern hybridizations, using *B. lomatogona* satellite-specific probes. A total of nine species were selected for comparative Southern hybridizations, including five species of the genus *Beta*, two species of the genus *Patellifolia*, and *S. oleracea* and *C. quinoa* as out group species. The autoradiograms from hybridization of the three satellite probes show a variable abundance of this repeats in the genus *Beta* (Figure 3.16).

BlSat1 strongly hybridized to the *B. lomatogona* genome, resulting in a ladder-like pattern after six hours of exposition (Figure 3.16A). Weaker signals were also detected in the *B. corolliflora* genome. In *B. vulgaris*, *B. patula* and *B. nana* genomes, hybridization signals are visible after longer exposure, indicating that BlSat1 occurs in much lower abundance or higher divergence in these species. In *Patellifolia* as well as *S. oleracea* and *C. quinoa* species no signal was detected. Satellite family BlSat1 is an exclusive satellite of genus *Beta*.

In comparative Southern hybridization, BlSat5 signals were detected in species of sections *Corollinae* and *Nanae*, no signal in species of section *Beta* as well as species of the genus *Patellifolia* and two out group species (Figure 3.16B). The abundance of BlSat5 is not uniform in the two sections *Corollinae* and *Nanae*. The strongest ladder-like signals were observed in the *B. nana* genome after six hours of exposition. The signal pattern is very similar but the signal intensity decreases in the following order *B. nana*, *B. corolliflora*, and *B. lomatogona*, which reveals the same organization of BlSat5 in three species but different abundance. Interestingly, the higher order structure of this satellite is visible on the autoradiogram, indicated by the additional signals between monomer and multimer signals. Hypothetically, there may be two variants of BlSat5 in these genomes, where one includes both subunit 1 and subunit 2, and the other composes only one subunit (sub1 or sub2).

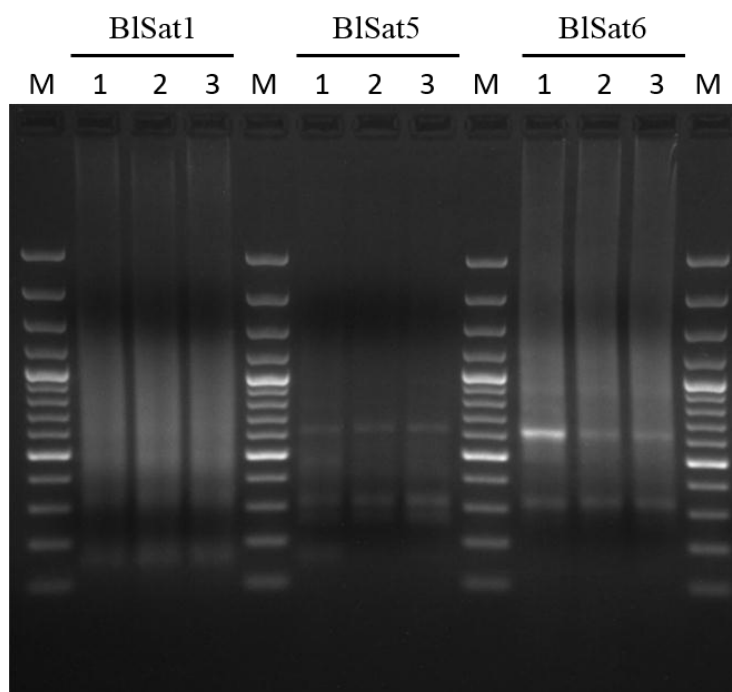
In contrast to BlSat1 and BlSat5, BlSat6 is much less abundant in *Beta* genomes (Figure 3.16C). After 30 hours of exposition, the satellite repeat BlSat6 hybridized to *Beta* species with a ladder-like pattern. A similar signal intensity was detected for *B. lomatogona* and *B. corolliflora*, it is likely that BlSat6 is present in the same abundance in both species. However, there is a shift which indicates a reduced monomer size in *B. vulgaris* and *B. patula*. This observation will be characterized in detail by sequence analysis in the following section. In addition, several faint signals were detectable among strong signals of monomers and multimers, this might result from random mutation inside BlSat6 monomers. No signals were detected in *Patellifolia* as well as *C. quinoa* and *S. oleracea* species after longer exposition (64 hours).



**Figure 3.16: Abundance and genomic organization of three *B. lomatogona* satellites in genus *Beta* and *Patellifolia***

Southern hybridizations of BISat1 probe to genomic DNA restricted with *MaeI* (A), BISat5 to genomic DNA restricted with *AluI* (B), and BISat6 to genomic DNA restricted with *BsmI* (C). The following species were tested: (1) *B. vulgaris*; (2) *B. patula*; (3) *B. lomatogona*; (4) *B. corolliflora*; (5) *B. nana*; (6) *P. procumbens*; (7) *P. patellaris*; (8) *C. quinoa*; and (9) *S. oleracea*.

The occurrence of the satellite families B1Sat1, B1Sat5, and B1Sat6 was also confirmed in *B. macrorhiza* of section *Corollinae* by PCR (Figure 3.17). This species has been considered as one progenitor of *B. corolliflora*. Therefore, it was selected for investigation of satellite distribution giving a helpful hint to trace the tetraploid origin of *B. corolliflora*. A ladder pattern up to trimer was observed in *B. macrorhiza* for all three satellite families. The pattern of each satellite family is similar in three species *B. lomatogona*, *B. macrorhiza*, and *B. corolliflora*. This result indicates the presence of three satellite families in *B. macrorhiza* and again confirms the close relationship between species in section *Corollinae*.



**Figure 3.17: PCR with satellite-specific primers of B1Sat1, B1Sat5, and B1Sat6**

The following species were tested: (1) *B. lomatogona*, (2) *B. macrorhiza*, (3) *B. corolliflora*, (M) Marker GeneRuler 100 bp Plus DNA Ladder.

### 3.2.2.7 Sequence divergence of *B. lomatogona* satellite B1Sat1, B1Sat5 and B1Sat6 in the genus *Beta*

In order to investigate sequence divergence of three *B. lomatogona* satellite families in *Beta*, *Patellifolia* and related species, garden PCR was performed (Figure 3.18), followed by cloning and sequencing.

PCR of B1Sat1 shows its presence in three sections *Beta*, *Corollinae* and *Nanae*. In sections *Corollinae* and *Nanae*, the amplicons up to tetramer were visible, while in section *Beta* only trimer

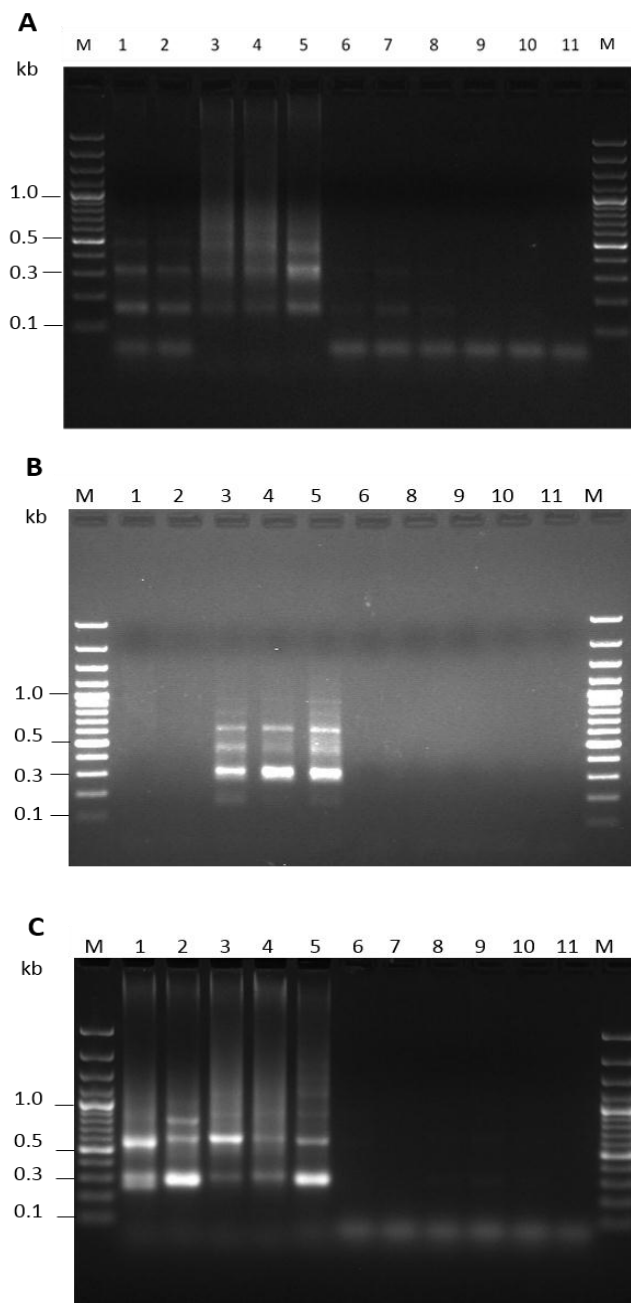
bands were present. Very weak PCR amplicons were observed in the two species of the genus *Patellifolia* (Figure 3.18A; lane 6, 8) while no signal was detected in the Southern hybridization (Figure 3.16A; lane 6, 7). The satellite units could not be amplified with BLSat1 primers in the related species *C. quinoa* and *S. oleraceae*. The result indicated that BLSat1 is specific for genus *Beta* and this satellite family may be organized in longer arrays as well as be more abundant in the genomes of *Corollinae* and *Nanae* sections.

The amplification of genomic sequences with BLSat5 primers resulted in a ladder-like banding pattern for both sections *Corollinae* and *Nanae*. The DNA can be well differentiated up to a trimer band. Among strong bands of monomer, dimer and trimer, there are additional bands corresponding to the size of one, three and five subunits (Figure 3.18B).

Figure 3.18C showed ladder banding pattern of BLSat6 at all tested species of *Beta*, *Corollinae* and *Nanae* sections. The same banding pattern was observed in *B. lomatogona*, *B. corolliflora* and *B. nana* whereas the banding pattern of *B. vulgaris* and *B. patula* tended to be slightly lower.

PCR products amplified from *Beta* species using satellite-specific primers were cloned and sequenced. Sequence analysis as well as their alignment were shown in Table 3.8 and Figure S4, Figure S5, and Figure S6 of the Supplement.

Accordingly, eight complete BLSat1 monomers of *B. vulgaris*, *B. patula*, *B. lomatogona*, and *B. nana* have the length of 171 bp, which was equal to the bioinformatic monomer (BLSat1\_in silico). Eight monomers of *B. vulgaris*, *B. patula*, and *B. nana* have the length of 161 bp and only one monomer of *B. patula* has the length of 159 bp (Table 3.8, BLSat1). The pairwise identity between monomers is 87.5%. The variation of BLSat1 monomer size is due to a deletion event, in particular the monomer sizes of 161 bp and 159 bp resulted from an internal deletion of 10 and 12 nucleotides, respectively (Supplementary Figure S4).



**Figure 3.18: PCR from different *Beta*, *Patellifolia* and related species**

The genomic DNA was amplified with satellite-specific primers for the satellite families B1Sat1 (A); B1Sat5 (B); and B1Sat6 (C). The following species were tested: (1) *B. vulgaris*, (2) *B. patula*, (3) *B. lomatogona*, (4) *B. corolliflora*, (5) *B. nana*, (6) *P. procumbens*, (7) *P. webiana*, (8) *P. patellaris*, (9) *S. oleracea*, (10) *C. quinoa*, and (11) Negative control (no genomic DNA). (M) Marker GeneRuler 100 bp Plus DNA Ladder.



Sequence analysis of B1Sat5 clones from *B. lomatogona*, *B. corolliflora*, and *B. nana* resulted in various monomer sizes, which ranges from 309 bp to 318 bp (Table 3.8, B1Sat5). Most of monomers (23 out of 28 monomers) have their size between 312 and 315 bp, and 314 bp is the prominent monomer size (11 out of 23 monomers). Compared to B1Sat1, B1Sat5 monomers are more diverged with a pairwise identity of 84.9% (Figure S5 of the Supplement). Analysis of subunits indicated that most of subunit 1 and subunit 2 have the size of 156-157 bp. The pairwise identity between subunit 1 is slightly lower than that between subunit 2 (84.5% and 85.2%, respectively) (Figure S7 of the Supplement), and subunits from different species were not separated into species-specific clusters by Neighbor-joining analysis (data not shown).

The monomer size of B1Sat6 varies between *Beta* species (Table 3.8, B1Sat6) and their sequences are also divergent (75.4% of pairwise identity). In particular, there are two monomer variants in *B. vulgaris*, with the monomer size of 261 bp and 312 bp. This explains for the shift of B1Sat6 signals of this species in autoradiogram (Figure 3.16C, lane 1). The variability in the monomer length can be assigned to the deletion of sequence motifs (Supplementary Figure S6).

**Table 3.8: Variation of typical *B. lomatogona* satellite monomers in the genus *Beta***

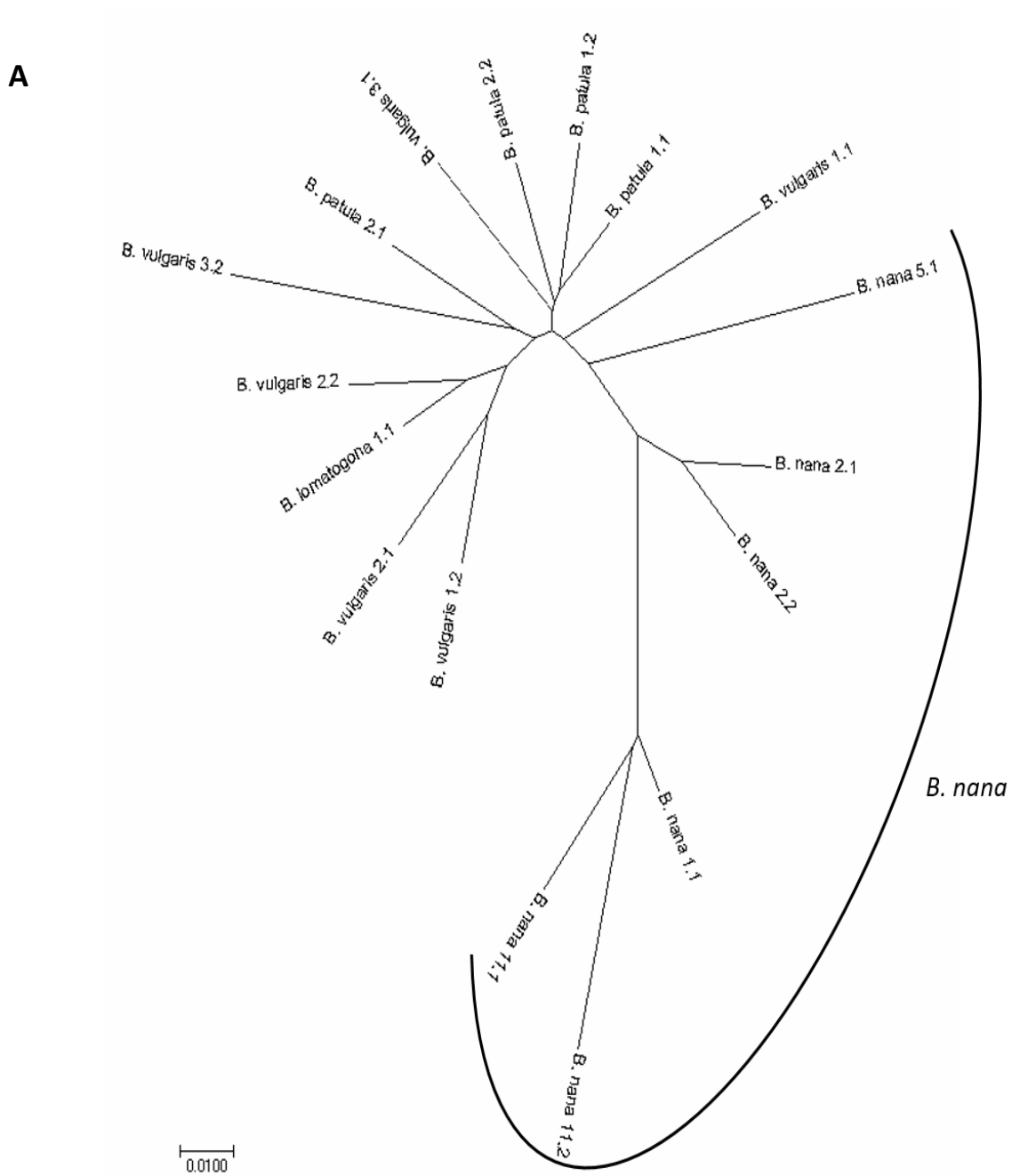
Satellite family	Clone name	Identity to consensus [%]	Type of clone	Monomer sizes of a simple repeat [bp]
<b>B1Sat1</b>	B1Sat1_in silico	100	monomer	171
	<i>B. vulgaris</i> _1	87.6	dimer	161 + 161
	<i>B. vulgaris</i> _2	92.5	dimer	171 + 171
	<i>B. vulgaris</i> _3	91.6	dimer	171 + 171
	<i>B. patula</i> _1	92.5	dimer	171 + 161
	<i>B. patula</i> _2	89.0	dimer	161 + 161
	<i>B. lomatogona</i> _1	93.9	monomer	171
	<i>B. nana</i> _1	85.3	monomer	161
	<i>B. nana</i> _2	90.2	dimer	171 + 159
	<i>B. nana</i> _5	91.0	monomer	171
	<i>B. nana</i> _11	82.7	dimer	161 + 161
<b>B1Sat5</b>	B1Sat5_in silico	100	monomer	313
	<i>B. lomatogona</i> _1	85.0	monomer	312
	<i>B. lomatogona</i> _2	87.1	monomer	313
	<i>B. lomatogona</i> _3	87.4	dimer	316 + 312
	<i>B. lomatogona</i> _4	89.4	monomer	313
	<i>B. lomatogona</i> _5	92.9	dimer	313 + 314
	<i>B. lomatogona</i> _6	92.0	dimer	312 + 318
	<i>B. lomatogona</i> _8	86.3	dimer	314 + 314
	<i>B. lomatogona</i> _9	91.4	dimer	314 + 313

Table 3.8: Continued

Satellite family	Clone name	Identity to consensus [%]	Type of clone	Monomer size of simple repeat [bp]	
BISat5	B. lomatogona_10	90.8	dimer	309 + 313	
	B. lomatogona_16	69.8	dimer	315 + 315	
	B. corolliflora_1	91.3	monomer	314	
	B. corolliflora_4	89.4	monomer	314	
	B. corolliflora_7	91.9	monomer	314	
	B. corolliflora_10	91.6	monomer	313	
	B. nana_1	87.2	monomer	314	
	B. nana_2	91.3	monomer	314	
	B. nana_3	91.1	monomer	318	
	B. nana_5	92.6	dimer	312 + 314	
	B. nana_7	91.9	dimer	315 + 314	
	BISat6	BISat6_in silico	100	monomer	315
		B. vulgaris_1	74.4	dimer	312 + 261
B. vulgaris_2		74.5	dimer	312 + 261	
B. vulgaris_3		74.2	dimer	312 + 261	
B. vulgaris_7		74.5	dimer	312 + 261	
B. vulgaris_8		74.5	dimer	312 + 261	
B. lomatogona_1		83.3	dimer	316 + 300	
B. lomatogona_2		80.1	trimer	311 + 300 + 285	
B. lomatogona_3		82.5	dimer	315 + 316	
B. lomatogona_4		81.9	dimer	312 + 315	
B. lomatogona_5		78.5	trimer	315 + 317 + 298	
B. lomatogona_6		84.5	dimer	313 + 315	
B. lomatogona_7		82.7	trimer	313 + 316 + 316	
B. corolliflora_3		82.4	dimer	315 + 313	
B. corolliflora_4		80.9	dimer	314 + 295	
B. corolliflora_7		82.0	dimer	287 + 314	
B. nana_1		82.8	dimer	311 + 315	
B. nana_2		80.9	dimer	316 + 295	
B. nana_4		80.9	dimer	313 + 312	
B. nana_5		80.6	dimer	316 + 299	
B. nana_6	81.4	dimer	300 + 299		

The neighbor-joining analysis was performed to investigate the relationship of each satellite in different *Beta* species, using the alignments of the three typical satellite monomer sequences in *Beta* species. The result revealed an incomplete clustering of all three satellite families. In particular, the BISat1 sequences originating from *B. vulgaris*, *B. patula*, and *B. lomatogona* could not be resolved. Only BISat1 sequences from *B. nana* were arranged in a loose branch (Figure 3.19A). BISat5 and BISat6 sequences from *Beta* species could not be also resolved, forming a

mixture of branches (Figure 3.19B, C). This indicates that each satellite sequence even from different species are highly similar.



**Figure 3.19A: Dendrogram representation of relationship between BISat1 sequences in *Beta* species**

B

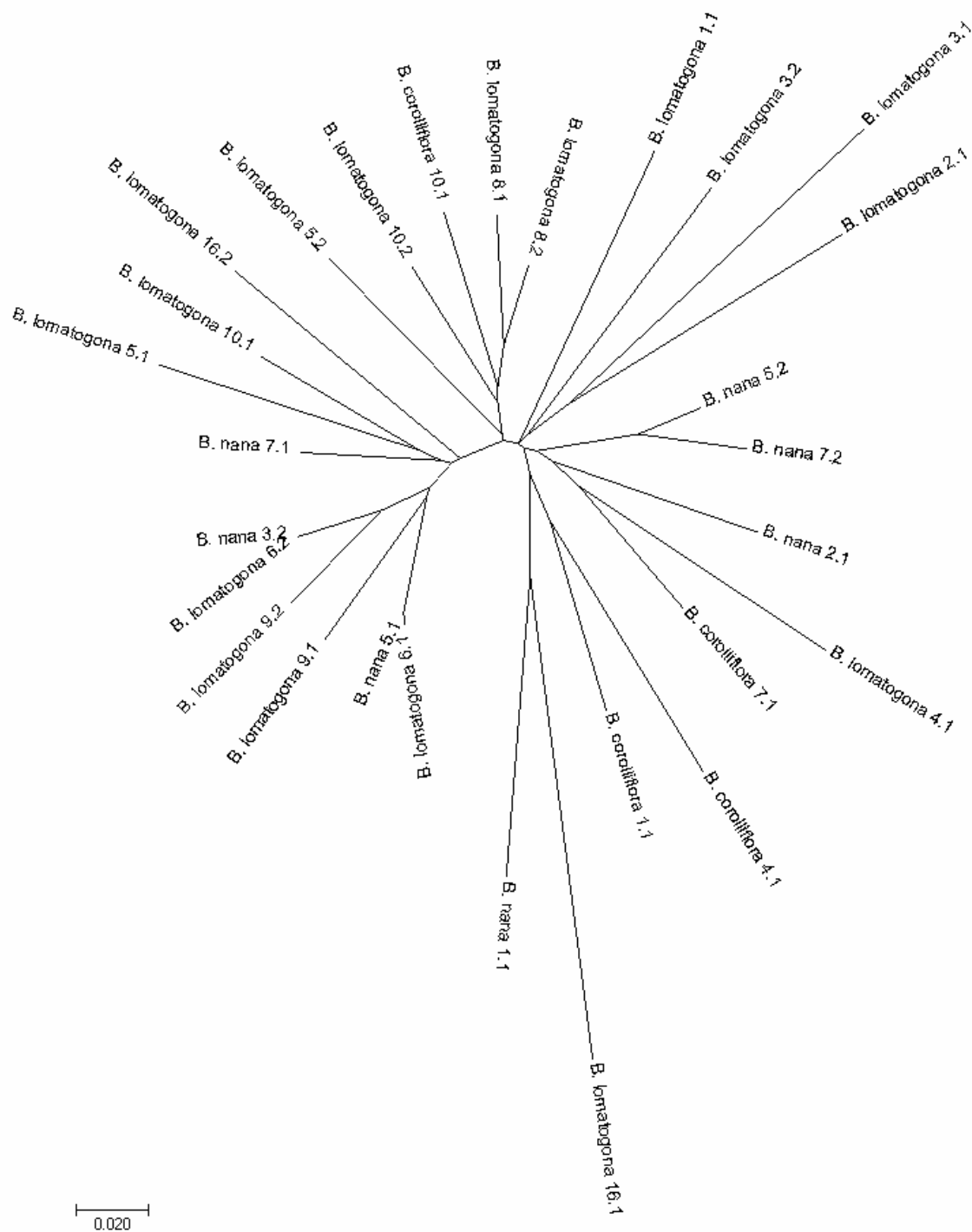
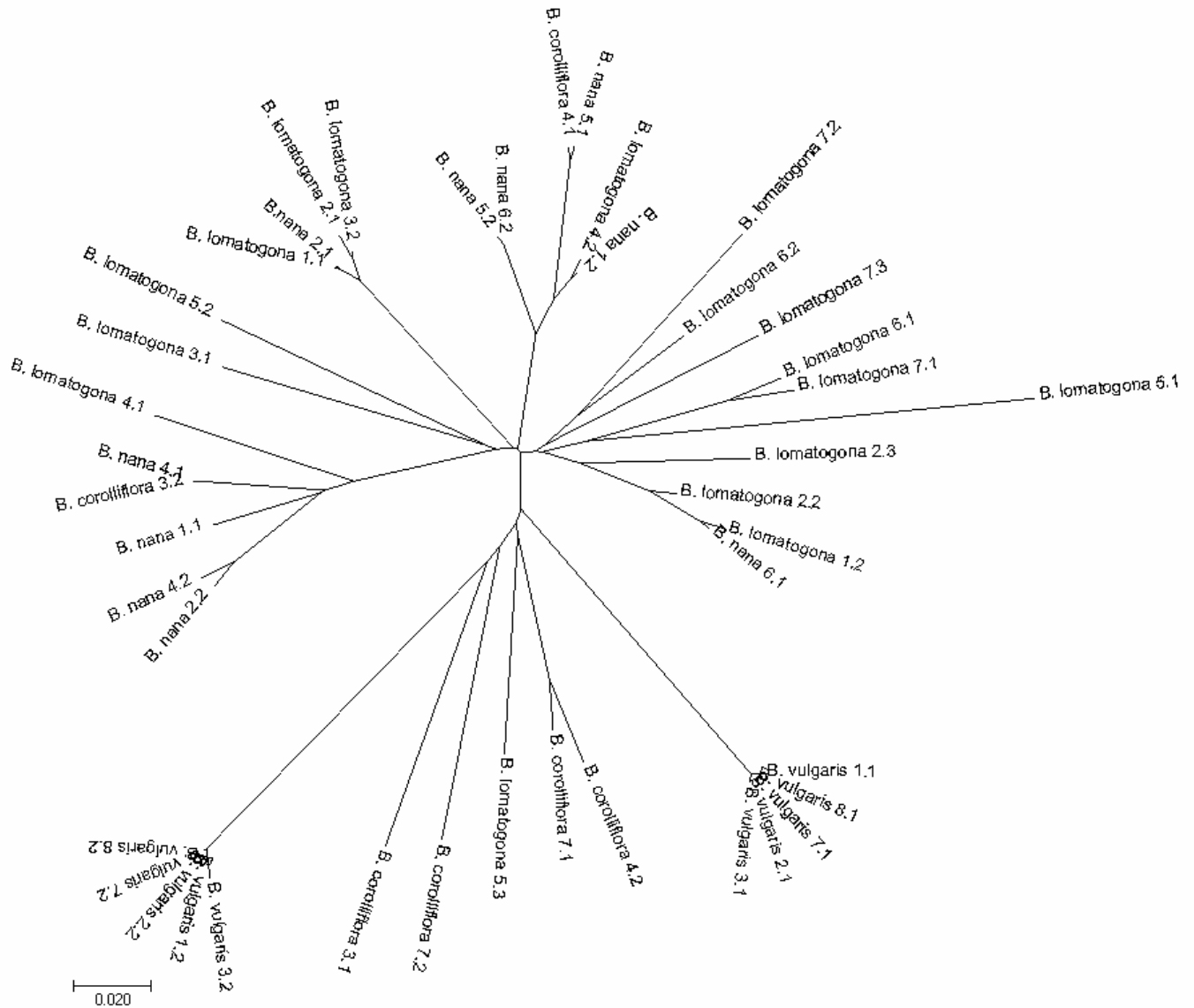


Figure 3.19B: Dendrogram representation of relationship between BISat5 sequences in *Beta* species

C



**Figure 3.19C: Dendrogram representation of relationship between BISat6 sequences in *Beta* species**

### 3.2.2.8 Comparative chromosomal localization of the satellite families BISat1, BISat5 along *Beta* chromosomes

In order to investigate satellite evolution through the genus *Beta*, tandem repeats that produce long satellite-typical arrays, including BISat1 and BISat5, were localized on the metaphases of additional species. The tested species were chosen based on the Southern hybridization signals in Figure 3.16A, B.

It was shown that ladder-like pattern in Southern hybridization experiment B1Sat1 occurs in the sections *Beta*, *Corollinae* and *Nanae*. Therefore, biotin-labeled B1Sat1 probes were hybridized to the metaphase chromosomes of *B. vulgaris* (section *Beta*); *B. corolliflora*, *B. intermedia*, *B. trigyna* (section *Corollinae*); and *B. nana* (section *Nanae*). Species in section *Beta* and *Nanae* are diploid with 18 chromosomes ( $2n = 2x = 18$ ). Dissimilarly, section *Corollinae* is more complex with different polyploidy, including diploid *B. lomatogona* ( $2n = 2x = 18$ ), tetraploid *B. corolliflora* and *B. intermedia* ( $2n = 4x = 36$ ), and pentaploid *B. trigyna* ( $2n = 5x = 45$ ).

The localization of B1Sat1 along *Beta* chromosomes was summarized in Table 3.9. Signal strength differs strongly between species belonging to different sections (Figure 3.20). B1Sat1 localizes in the distal regions of four *B. vulgaris* chromosomes, the signal hybridization is also amplified on four *B. nana* chromosomes in pericentromeric position and two other chromosomes in distal position. Although four species of section *Corollinae* are at differently polyploid level, B1Sat1 is still localized on eight chromosomes. This indicates that satellite family B1Sat1 is not differentially amplified through section *Corollinae*, the number of chromosomes having signals is conserved and only signal strength differs. This satellite is less amplified in sections *Beta* and *Nanae*.

Compared to B1Sat1 satellite family B1Sat5 is less divergent and only occurs in the species of sections *Corollinae* and *Nanae* (Figure 3.21). The hybridization pattern was detectable on six chromosomes of *B. lomatogona* (Figure 3.12B), but in the other tested species of sections *Corollinae* and *Nanae* the hybridization pattern of this satellite was amplified to a higher number of chromosomes. In particular, signal pattern of B1Sat5 was detected in the pericentromeric position of ten *B. corolliflora* chromosomes. Signal strength varied for different chromosomes, four chromosomes have stronger signals than the others. In *B. intermedia* and *B. trigyna*, detectable signal occurs in the pericentromeric region of 16 chromosomes. Strong signals on four chromosomes and quite weak signals on the others were noticed, indicating the presence of small and large satellite arrays.

As indicated in the Southern hybridization experiment (Figure 3.16B) and the mapping of B1Sat5 sequence to read sequences of *Beta* species (Table 3.3), satellite B1Sat5 is abundant in the *B. nana* genome. In FISH experiment, the same result was obtained, eight chromosomes were labeled with the B1Sat5 satellite probe in the pericentromeric regions of *B. nana* chromosomes.

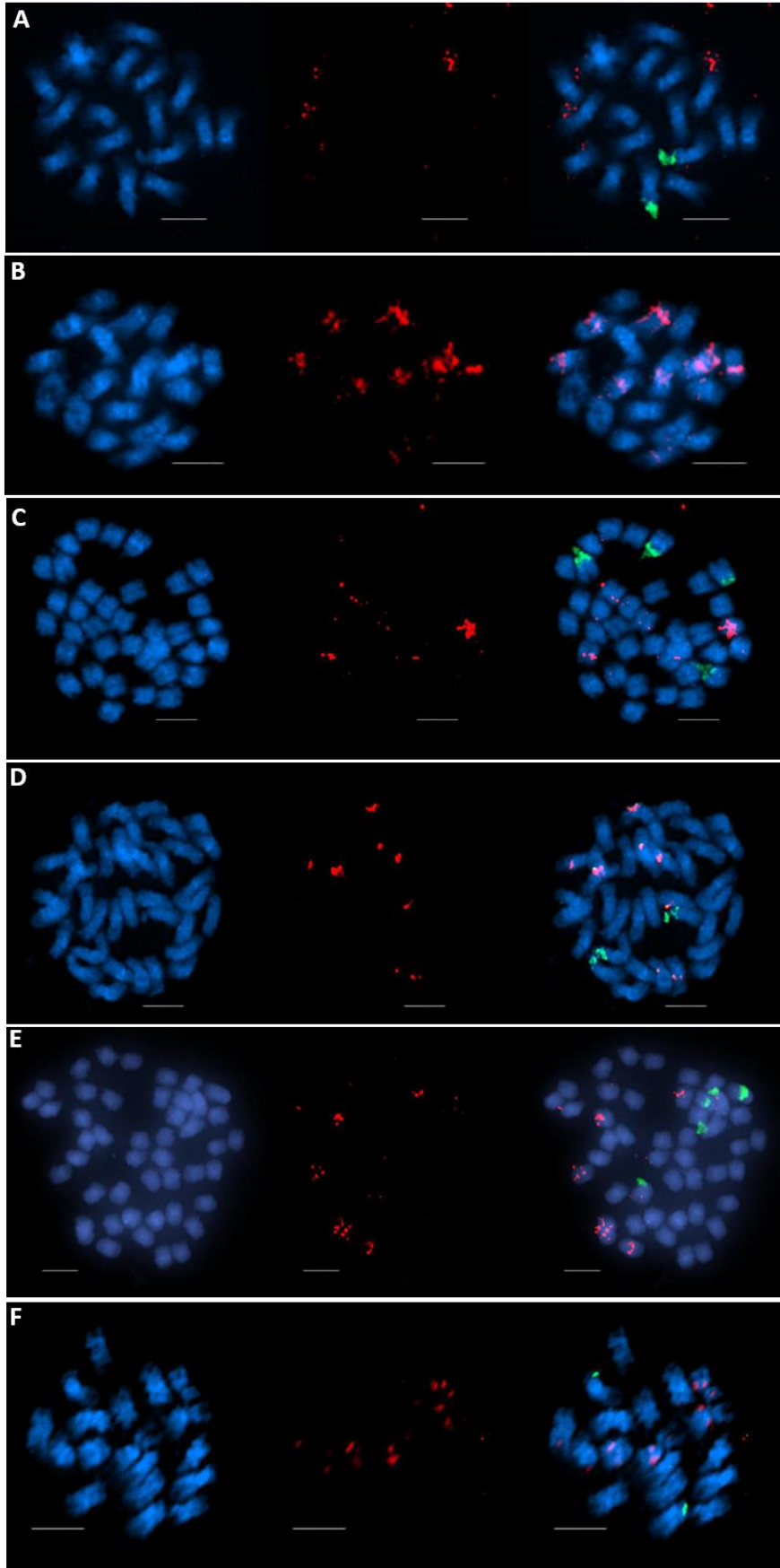
In general, the number of signals as well as the supposed number of chromosome pairs from B1Sat1 and B1Sat5 satellite families were shown in Table 3.9. The B1Sat1 satellite showed a conserved chromosomal localization in the species of section *Corollinae* despite of different polyploid levels between these species. Eight chromosomes, presumably four chromosome pairs, have signals in the pericentromeric heterochromatin of *Corollinae* species, while this number is six in *B. nana*, presumably three chromosome pairs. In *B. vulgaris*, only four chromosomes, presumably two pairs, have signals in the distal regions. The number of B1Sat5 signals increased with the higher polyploidy levels of species in section *Corollinae*. In *B. lomatogona*, the FISH signals were detected on three chromosome pairs where two strong and one faint signals located at pericentromeric regions. The number of signals is ten in *B. corolliflora*, whereas in two species, *B. intermedia* and *B. trigyna*, the number of signals were 16, presumably 8 chromosome pairs. The signals were also detected on four *B. nana* chromosome pairs.

**Table 3.9: Summary of chromosomal distribution of B1Sat1 and B1Sat5 along *Beta* chromosomes**

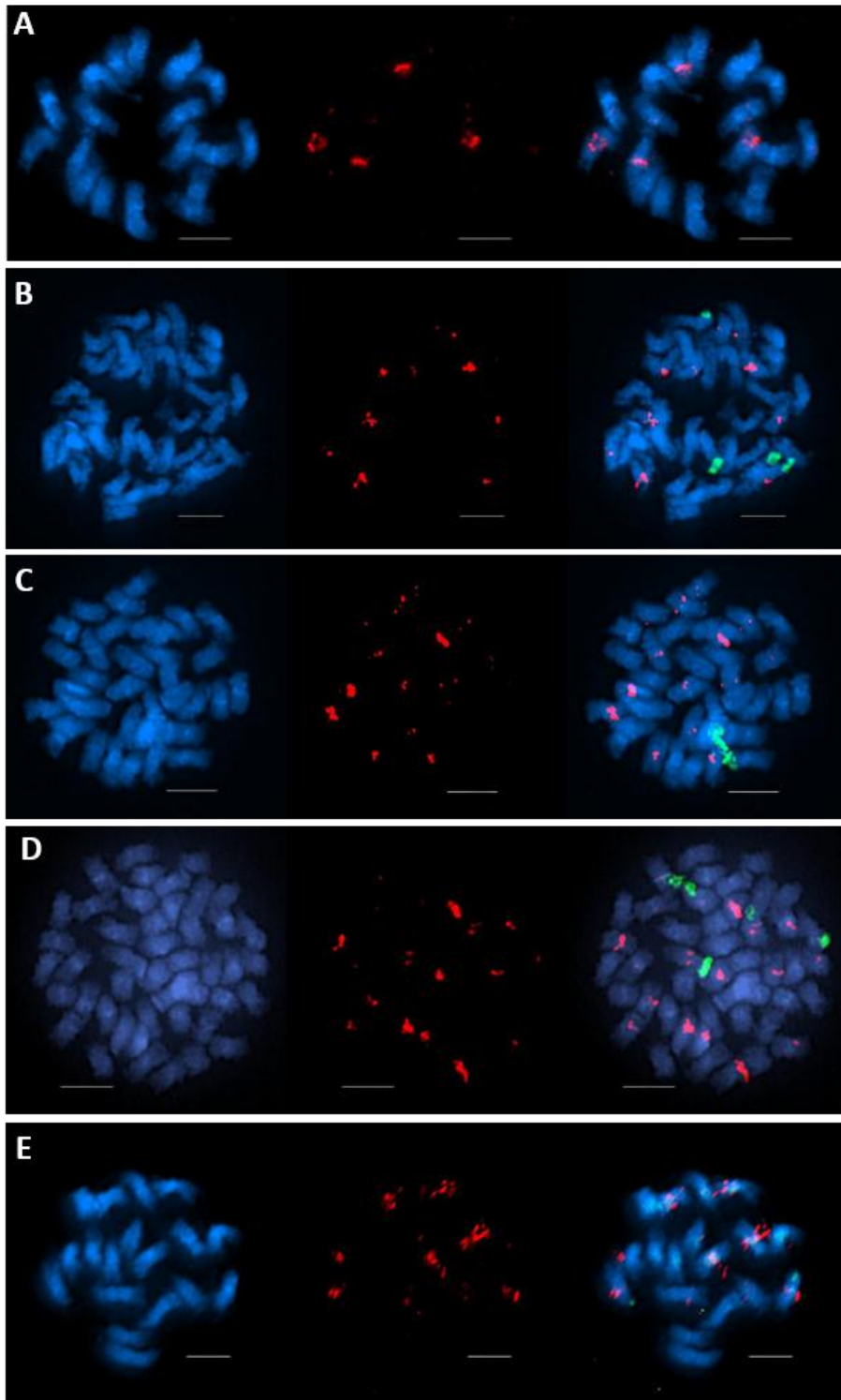
Satellite family	Section	Tested species	Genome type	No. of signals	No. of chromosome pairs with signals
<b>B1Sat1</b>	<i>Beta</i>	<i>B. vulgaris</i>	2n = 2x = 18	4	2
		<i>Corollinae</i>	<i>B. lomatogona</i>	2n = 2x = 18	8
		<i>B. corolliflora</i>	2n = 4x = 36	8	4
		<i>B. intermedia</i>	2n = 4x = 36	8	4
		<i>B. trigyna</i>	2n = 5x = 45	8	4
	<i>Nanae</i>	<i>B. nana</i>	2n = 2x = 18	6	3
<b>B1Sat5</b>	<i>Corollinae</i>	<i>B. lomatogona</i>	2n = 2x = 18	6	3
		<i>B. corolliflora</i>	2n = 4x = 36	10	5
		<i>B. intermedia</i>	2n = 4x = 36	16	8
		<i>B. trigyna</i>	2n = 5x = 45	16	8
	<i>Nanae</i>	<i>B. nana</i>	2n = 2x = 18	8	4

**Figure 3.20: Different signal patterns among *Beta* species of satellite B1Sat1.**

Blue and red fluorescence shows DAPI-stained DNA and satellite DNA, respectively. Green signals reveal the position of 18S-5.8S-25S rRNA genes. Scale bar is 5  $\mu$ m. (A) *B. vulgaris*; (B) *B. lomatogona*; (C) *B. corolliflora*; (D) *B. intermedia*; (E) *B. trigyna*; and (F) *B. nana*.







**Figure 3.21: Chromosomal localization of B1Sat5 in *Beta* sections *Corollinae* and *Nanae***

Blue and red fluorescence shows DAPI-stained DNA and satellite DNA, respectively. Green signals reveal the position of 18S-5.8S-25S rRNA genes. Scale bar is 5  $\mu\text{m}$ . The following species were tested: (A) *B. lomatogona*; (B) *B. corolliflora*; (C) *B. intermedia*; (D) *B. trigyna*; and (E) *B. nana*.

### 3.2.3 The tandem repeats B1Sat2, B1Sat3 and B1Sat4 are sequence domains of retrotransposons

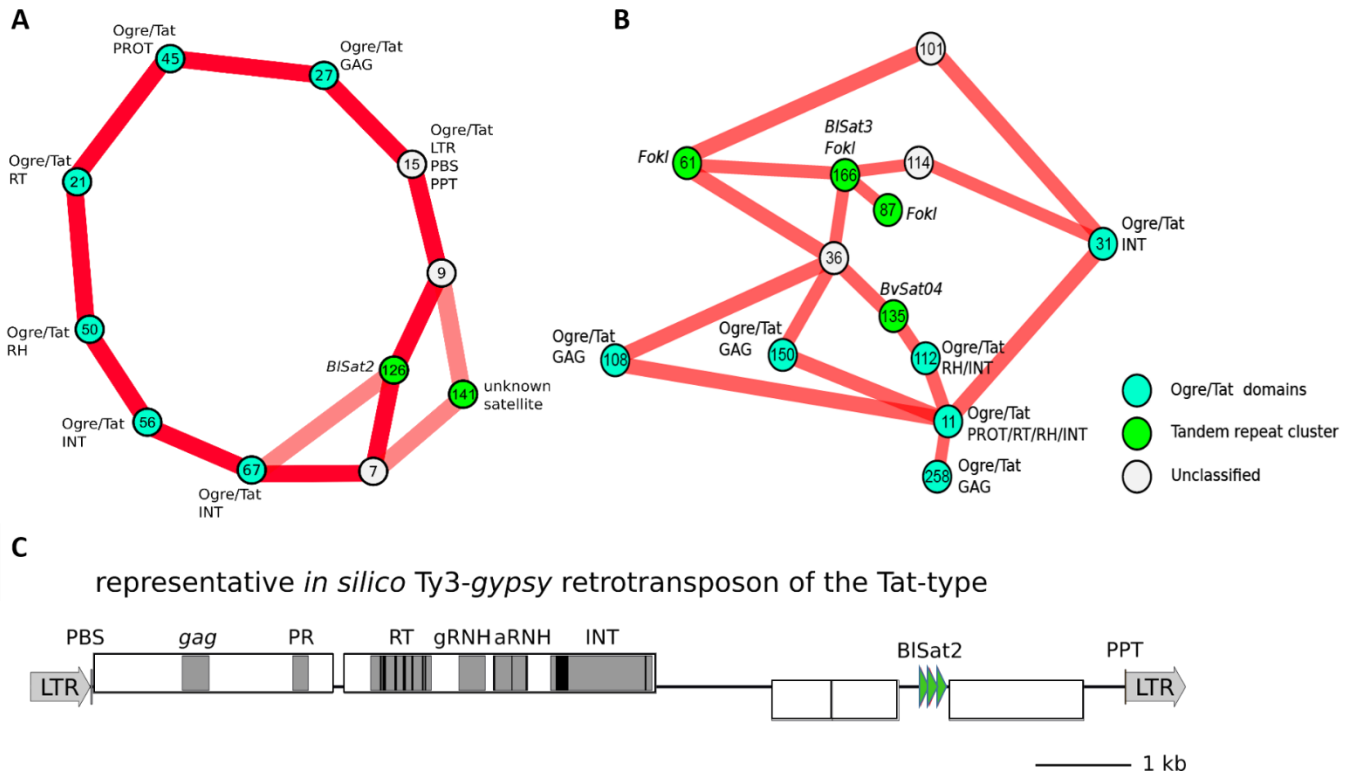
#### 3.2.3.1 B1Sat2, B1Sat3 and B1Sat4 are associated with retrotransposons and show unconventional satellite features

The graphs of typical satellite clusters are usually star-like or circular shape. However, the graphs of three clusters CL126, CL166, and CL214 representing B1Sat2, B1Sat3, and B1Sat4, respectively, have the long outliers of less similar sequences. Therefore, super-clustering was performed to assign them to other clusters. Link between B1Sat2 and B1Sat3 clusters and Ogre/Tat-typical protein domains was indicator of these clusters associated with Ogre/Tat retrotransposons (Figure 3.22). There have been some evidences of the integration between two satellites B1Sat2 and B1Sat3 and Ogre clade of LTR retrotransposon (Hoffman, 2017; Bannack, 2017). This may a main reason for the dispersed chromosomal localization of these satellites. B1Sat4 was not detected in any supercluster although its graph also contained the long outliers of less similar sequences.

The satellite family B1Sat2 (cluster CL126) together with a yet unknown satellite (cluster CL141) were associated with Ogre/Tat retroelement (Figure 3.22A). Characteristic LTR retrotransposon domains such as protease (cluster CL45) and reverse transcriptase (cluster CL21), RNaseH (cluster CL50) and integrase (clusters CL56 and CL67) and gag protein (cluster CL27) were found in this supercluster. Cluster CL15 contains the long terminal repeats (LTR), primer binding site (PBS) and polypurine tract (PPT). The complete Ogre/Tat retrotransposon element was successful reconstructed, which includes a tandem array of three B1Sat2 monomers (Figure 3.22C).

Similarly, satellite family B1Sat3 was found in an Ogre/Tat supercluster (Figure 3.22B). In cluster CL166, two satellite families, B1Sat3 and the *FokI* satellite/Dione, were annotated. The *FokI* satellite/Dione satellite, a known satellite isolated from *B. vulgaris* (Zakrzewski *et al.*, 2010), was found in cluster CL61 and cluster CL87, whereas BvSat4 satellite (Zakrzewski *et al.*, 2010), was annotated in cluster CL135. Additionally, various sequences were found in this supercluster, in particular, *gag* sequences in clusters CL108, CL150 and CL258; DNA sequences of protease, reverse transcriptase (RT), RNaseH and integrase in cluster CL11, integrase sequences in clusters CL31 and CL112, and RNaseH sequences in cluster CL112. However, clusters with sequences from the terminal repeats (LTRs), the primer binding site (PBS) and polypurine tract (PPT) could not be identified. Clusters CL36, CL101 and CL114 were uncharacterized.

Although no supercluster containing BISat4 could be identified in the *RepeatExplorer* output, the similar graph shape of cluster CL214 (including BISat4) indicates that this cluster might be associated with LTR retrotransposon. With the longer and higher quality sequence reads from *B. lomatogona*, the re-clustering may be performed resulting in a supercluster of LTR retrotransposon.

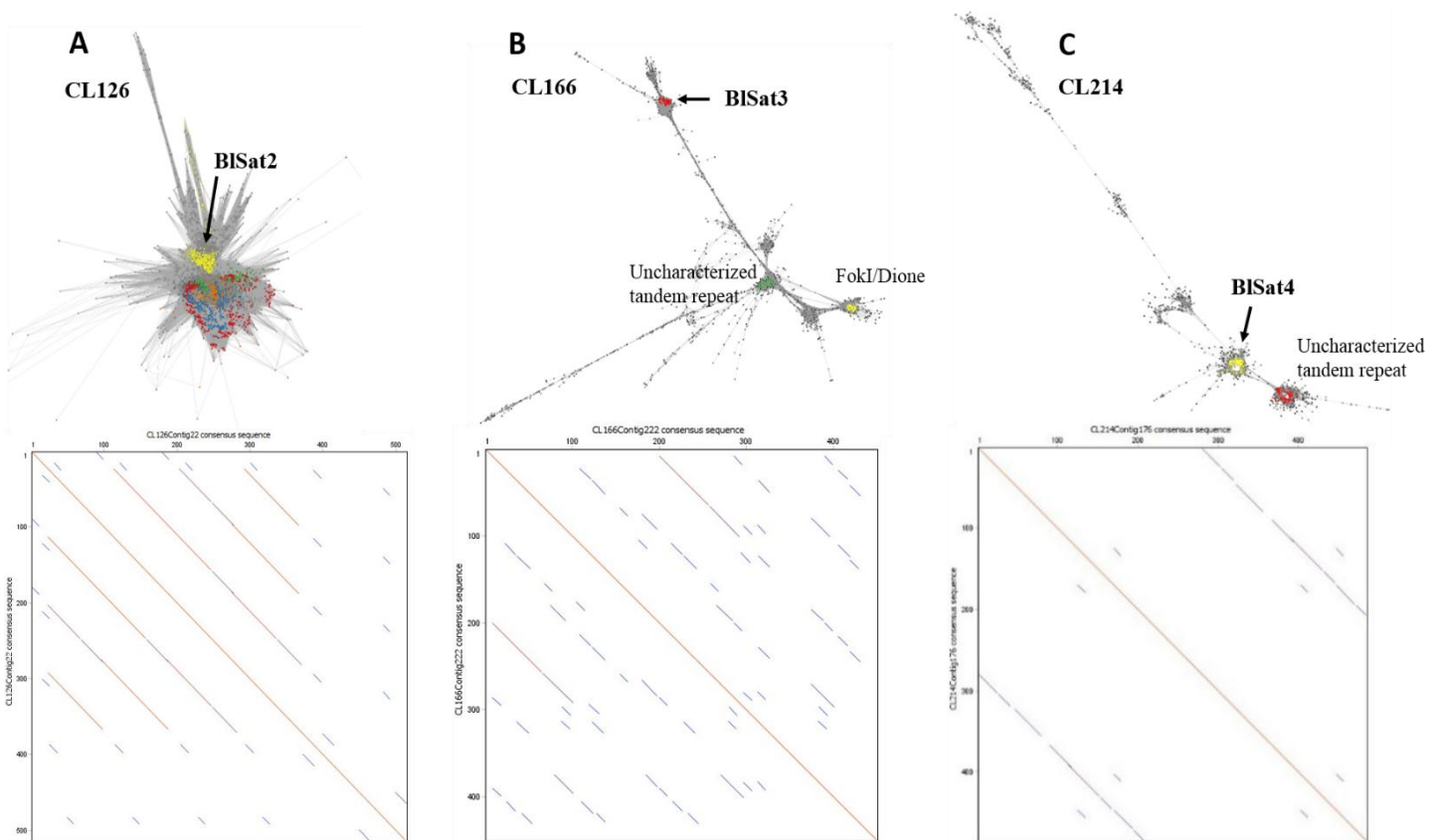


**Figure 3.22: Supercluster with Ogre/Tat retrotransposon sequences containing the satellite families BISat2 and BISat3**

The numbers of the linked clusters were given. The LTR retrotransposon domains (turquoise blue) and the associated tandem repeats (green) can be identified. Uncharacterized clusters were not colored. The satellite families BISat2 (A) and BISat3 (B) associate with supercluster Ogre/Tat retrotransposon. (C) Complete Ogre/Tat retrotransposon element including a tandem array of three BISat2 monomers.

### 3.2.3.2 Molecular structure and organization of satellite repeats in retrotransposons

The three tandem repeats BISat2, BISat3, and BISat4 were characterized as un-typical satellite DNA due to their positions in complex graph shapes. Apart from the satellite “coil”, they have long outliers of less similar sequences (Figure 3.23). There were several dominant contigs in each cluster (colors in each graph), of which one contig representing a novel satellite family was selected and analyzed in detail.



**Figure 3.23: Circular/star-like graph shape and dotplot of tandem repeat clusters.**

Star-like graph shape and dotplot of cluster CL126 (A) and cluster CL166 (B), and circular graph shape and dotplot of cluster CL214 (C). The colors in graphs indicate the most representative contigs in each cluster. The arrows show the position of contigs corresponding to satellites BISat2, BISat3 and BISat4. The continuous lines in dotplot correspond to high homologous value. Nucleotide positions are recorded on the X and Y axes of the dotplots allow the estimation of the monomer length.

In cluster CL126, the first five largest contigs were analyzed. They were mapped onto the cluster graph using *SeqGraper*. Out of five, three contigs (contig 22 - red, contig 14 – green, and contig 6 - orange) were tandemly arranged with a monomer of 90 bp length. Contig 22 containing most sequence reads was analyzed further and designated as BISat2 (Figure 3.23A, arrow). Its dotplot showed four parallel lines, indicating three complete monomers (Figure 3.23A).

The three major contigs 148 (yellow), 222 (red) and 191 (green) were analyzed for cluster CL166. A larger contig (contig 148) was annotated as satellite *FokI* satellite/Dione (Zakrzewski *et al.*, 2010) with the monomer length of 131 bp. Contig 222 of which the monomer size was 190 bp was annotated as a novel satellite and designated BISat3 (Figure 3.23B, arrow). 278 sequence reads (7.6% of the reads within the cluster) correspond to BISat3. Its dotplot showed two parallel lines

corresponding to two monomers of 190 bp (Figure 3.23B). Contig 191 indicated as uncharacterized tandem repeat with basic unit of 157 bp in length.

The two major contigs together incorporate 34% of the reads within the cluster CL214 are contig 176 and contig 75. Both contigs show tandem arrangement, but only contig 176 was analyzed and contig 75 was uncharacterized. Contig 176 is the largest contig representing BISat4 (Figure 3-23C, arrow) with monomer size of 276 bp. Two clear parallel lines observed in the dotplot indicated one complete monomer of satellite family BISat4 ((Figure 3.23C).

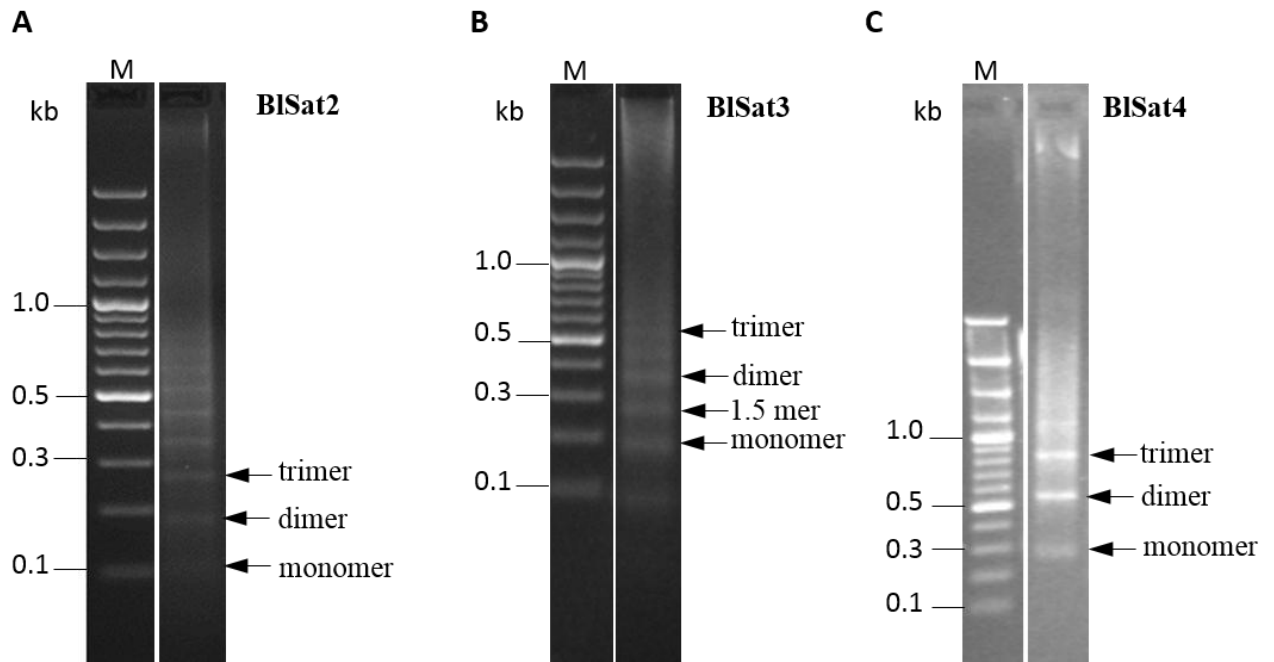
Schematic representations of the bioinformatically derived monomer consensus sequences are shown in Figure 3.24. Together with Table 3.5, the information of AT content and identity value are present.

BISat2 has a simple structure with monomer length of 90 bp and moderate AT content (48.9%). This satellite shows high identity (92.2%) of reads to the consensus sequence derived from bioinformatic analysis. The sequence motif of TTGG was repeated four times in BISat2 monomer (blue color in Figure 3.24A). A suitable restriction enzyme *BseGI* in Figure 3.24A resulting in the ladder-like pattern in Southern hybridization was detected.

The 190 bp BISat3 satellite shows AT content of 52.5% and lower conservation in monomer sequence, indicated by identity value of 84.9%. This satellite family consists repetitions of two smaller subunits. The length of one subunit sequence (sub1) was 104 bp and of other (sub2) was 86 bp. Their alignment showed that the subunit sub2 has an internally deleted sequence of 17 bp (indicated by gap in subunit 2, Figure 3.24B). The similarity between subunits sub1 and sub2 is 77.7%. The restriction enzyme cutting once in monomer was *BseGI*.



between main bands of monomer and multimers. This might support to the hypothesis that there is a higher order structure of BISat3.



**Figure 3.25: Ladder-like pattern in agarose gel after PCR with satellite-specific primers of BISat2, BISat3 and BISat4**

The agarose gel electrophoresis was carried out with gels of 2 % agarose. The monomer positions are marked with arrows. (A) Ladder-like pattern of the BISat2 satellite, (B) BISat3 satellite, and (C) BISat4 satellite, M: marker GeneRuler 100 bp Plus DNA Ladder.

The inserts corresponding to putative monomers and/or multimers were clones and sequenced to serve for sequence analysis. All monomers from each satellite were aligned together and to the consensus monomer sequence. From the sequencing results, the probes for BISat2, BISat3 and BISat4 were selected for Southern and FISH hybridization experiments were marked with asterisk in Table 3.10. The probe of BISat2 showed the highest identity of 94.7%, followed by probe of BISat3 with 93.1%. The probe of BISat4 was slightly lower similar with 90%. All three probes are multimers. The alignments were shown in Figure S6 of the Supplement.

**Table 3.10: Clones of three tandem repeats B1Sat2, B1Sat3, and B1Sat4 in *B. lomatogona***

For each satellite family, the clone used for hybridization was marked with asterisk

Satellite family	Clone name	Length of insert [bp]	Number of complete monomers	Identity to consensus [%]
B1Sat2	B1Sat2_3	496	4	87.1
	B1Sat2_4	289	3	94.5
	B1Sat2_5	181	2	91.4
	B1Sat2_6	361	3	89.3
	B1Sat2_11*	539	5	94.7
B1Sat3	B1Sat3_1*	566	3	93.1
	B1Sat3_4	378	2	89.3
	B1Sat3_5	358	2	87.9
	B1Sat3_6	265	1	79.7
	B1Sat3_7	350	2	92.5
	B1Sat3_8	275	1	94.2
	B1Sat3_9	274	1	93.9
	B1Sat3_10	657	1	91.4
	B1Sat3_11	271	1	91.8
	B1Sat3_12	653	3	92
	B1Sat3_14	670	3	93.2
	B1Sat3_15	583	2	89.8
B1Sat4	B1Sat4_1	542	2	96.3
	B1Sat4_4	503	2	90.1
	B1Sat4_6	542	2	98.2
	B1Sat4_7*	761	3	90
	B1Sat4_9	503	2	88.9

The sequencing of five clones with inserts carrying sequences of B1Sat2 satellite resulted in nucleotide sequences of 17 complete monomers. The pairwise identity of these monomers is 88.0%. There is a *Bse*GI site at nucleotide position 43-47, but this region is not conserved in all monomers, only *B. lomatogona\_5* clone showed the *Bse*GI recognition site (Supplementary Figure S8).

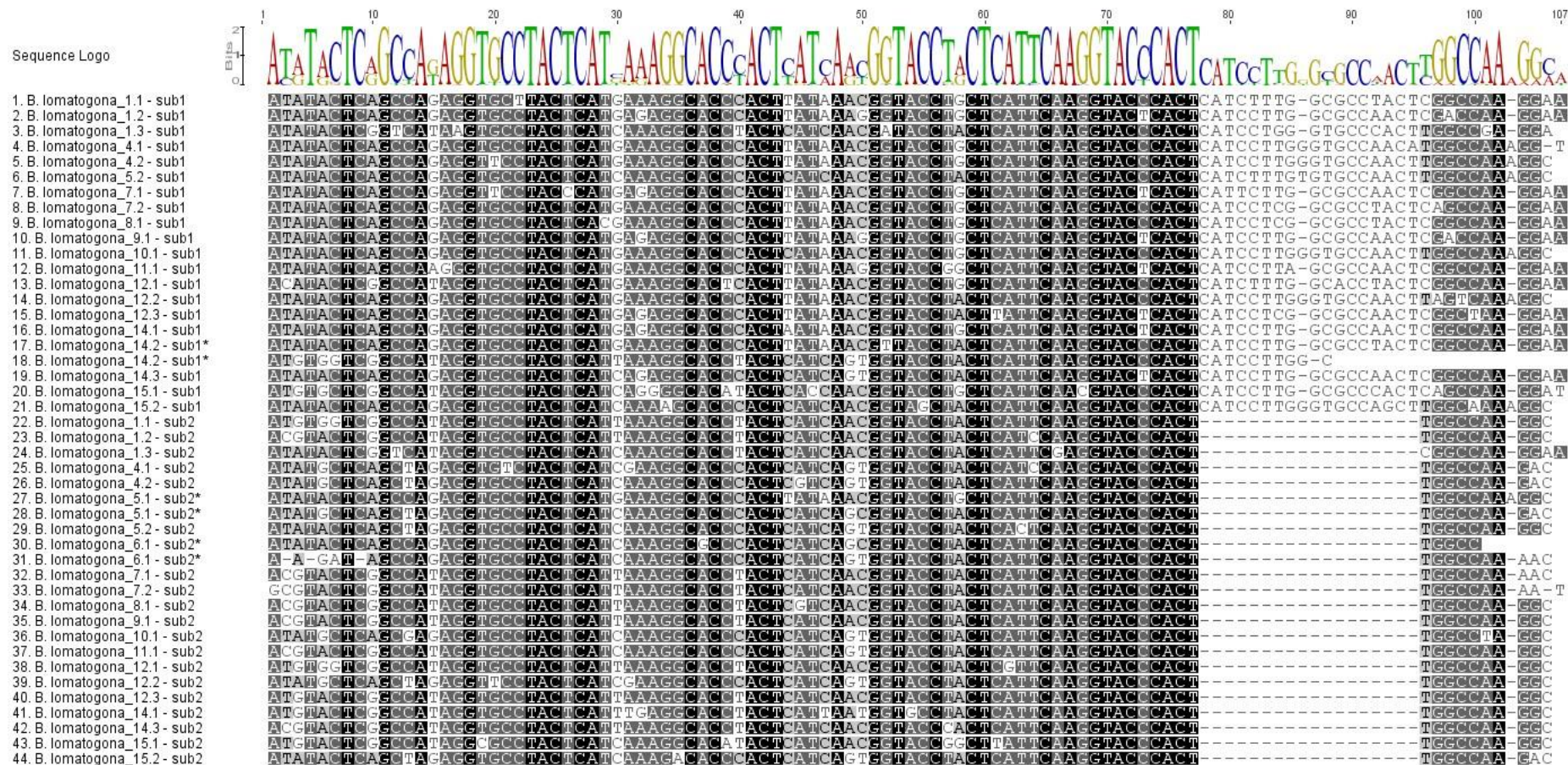
The sequencing of 12 clones with inserts of satellite family B1Sat3 yielded sequences of 22 complete monomers (Supplementary Figure S9). The pairwise identity between these monomers is 87.9%. At position 77-81 there is a recognition site of *Bse*GI whose recognized sequence is



CATCCNN. This sequence motif is relatively conserved between monomers, only five monomers showed variation. There are two monomers with shorter length caused by internal deletions.

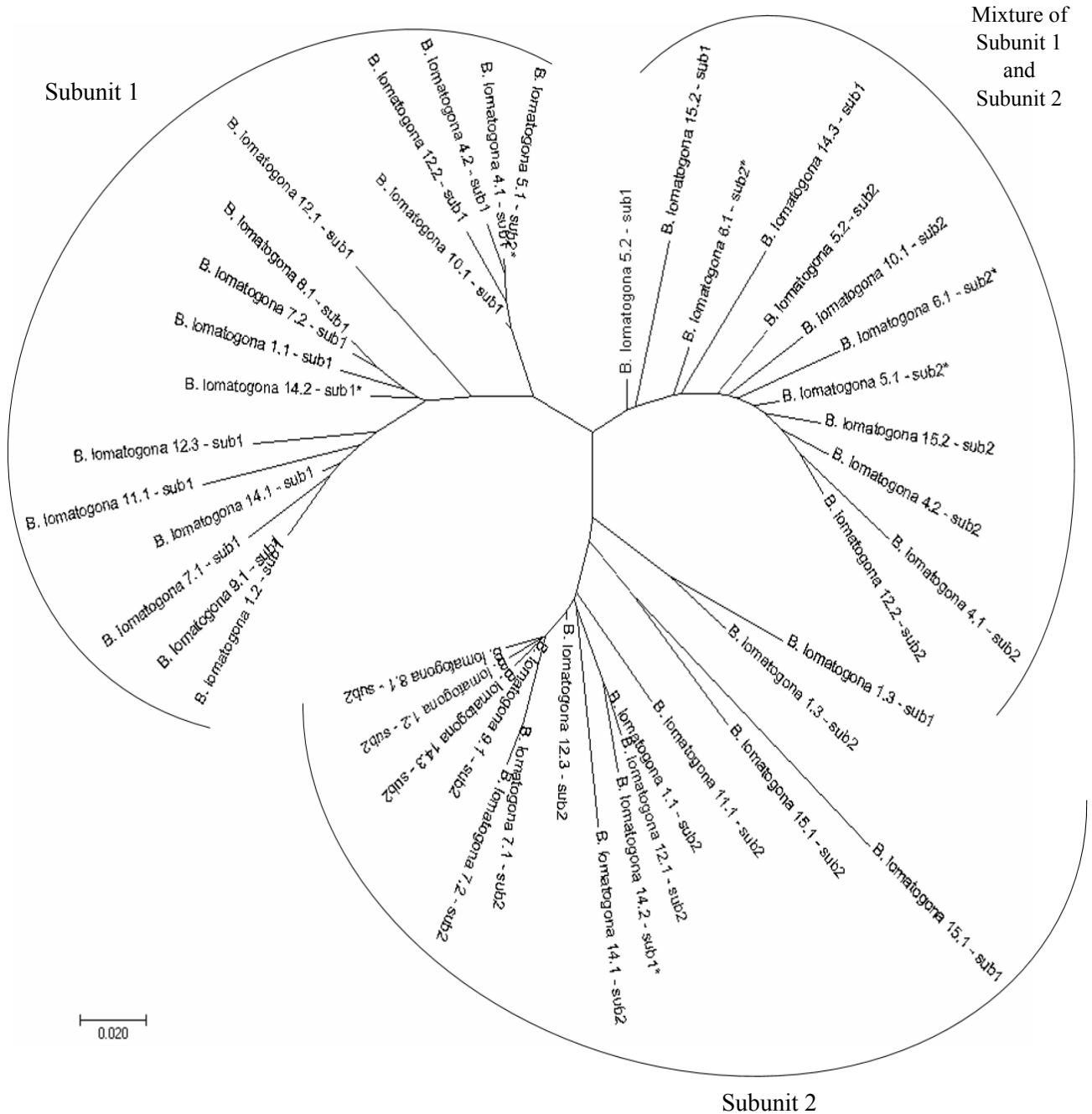
According to bioinformatic analysis, satellite B1Sat3 has a monomer size of 190 bp and contains two subunits (Figure 3.24B). Subunit 1 is 104 bp in length while subunit 2 is 86 bp. The alignment of two bioinformatic subunits shows 77.7% of identity, subunit 2 presents a deletion of 17 nucleotides. Sequencing of clones results in 12 complete monomers and nine of these monomers include two subunits sub1 and sub2. There are two monomers (*B. lomatogona\_5.1* and *B. lomatogona\_6.1*) which include two subunits sub2 in monomer sequences resulting shorter monomer size of 172 bp, and the other monomer *B. lomatogona\_14.2* is composed of two subunits sub1 (subunits were marked with an asterisk in Figure 3.26). As subunit 1 and subunit 2 were defined by absence and presence of 17 bp deletion, respectively, the alignment between all subunits was generated to clearly identify the differences between subunit 1 and subunit 2 (Figure 3.26). The alignment of all subunit 1 showed a pairwise identity of 88.8% while a slightly higher similarity between subunit 2 was observed (89.5%). The pairwise identity of all subunit 1 and subunit 2 only showed 79.8%. The alignment showed that variable point mutations do not correlate with presence or absence of the deletion, which resulted in a heterogeneous array. This was reflected in incomplete sorting by Neighbor-joining clustering (Figure 3.27). Most of the subunits sub1 were grouped into one clade and the other clade including most of the subunits sub2 was observed. However, there was a mediate group that comprised four subunits sub1 and eight subunits sub2.

The sequences of 11 monomers were analyzed in five clones of satellite B1Sat4. The pairwise identity is 88.4%. A recognition sequence for *NdeI* (CATATG) is in a conserved region, at position 15-20 (Supplementary Figure S10).



**Figure 3.26: Alignment of B1Sat3 subunits sub1 and sub2**

All subunits (sub1 and sub2) extracted from monomers of B1Sat3 satellite were aligned using the MUSCLE algorithm in *Geneious*. The higher value of sequence identity, the darker the filled shading.



**Figure 3.27: Dendrogram representing structural relationship between subunit 1 (sub1) and subunit 2 (sub2) of BISat3 monomers**

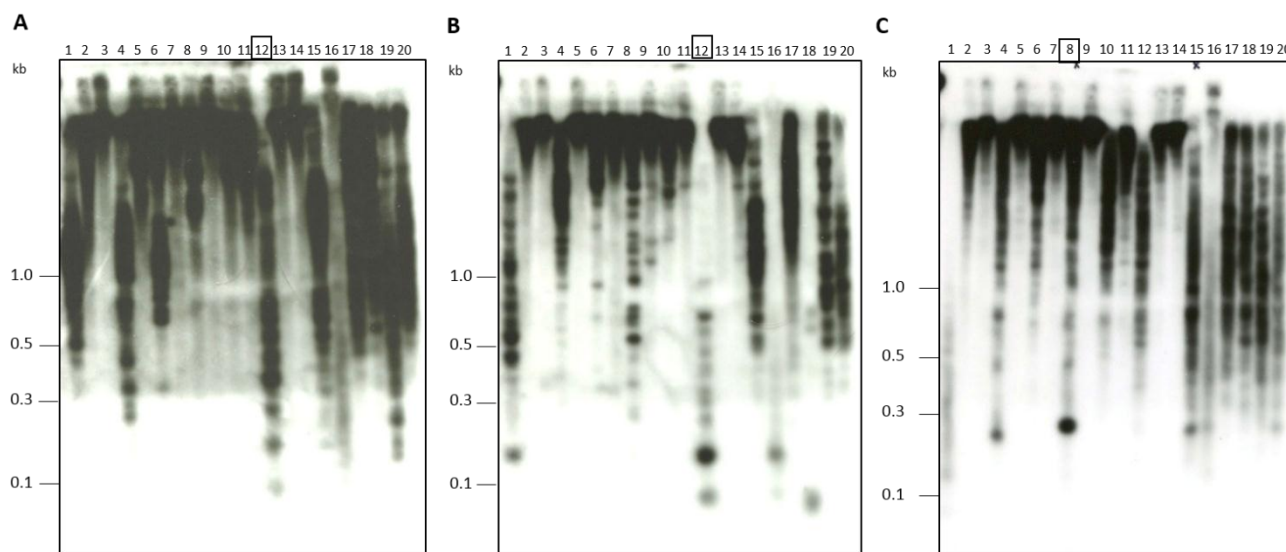
### 3.2.3.2 Genomic organization of BISat2, BISat3, and BISat4 in *B. lomatogona*

In order to investigate the genomic organization of the satellites BISat2, BISat3 and BISat4 Southern hybridizations were performed with different 20 restriction enzymes (Figure 3.28).

The autoradiogram of B1Sat2 after a 20-hour exposition showed a ladder-like pattern after restriction with *BseGI*. *BseGI* cuts the B1Sat2 monomer once, monomeric and multimeric (up to octamer) bands were visible of which signal intensity increases with the fragment size (Figure 3.28A, lane 12). This suggests that the restriction site is not conserved in this tandem repeat.

Similar to B1Sat2, a ladder-like pattern of *BseGI* was also observed in the autoradiogram of B1Sat3. The genomic DNA digested with *MseI* and *FokI* showed monomer bands of which the size is equal to the bioinformatically identified monomer (Figure 3.28B; lane 1, 16).

As shown in the autoradiogram of B1Sat4 there is a characteristic band pattern in lane 10, which corresponds to the restriction endonuclease *NdeI*. The signal is very strong at monomer size because of conserved *NdeI* recognized site in the B1Sat4 monomers (Figure 3.28C, lane 8).

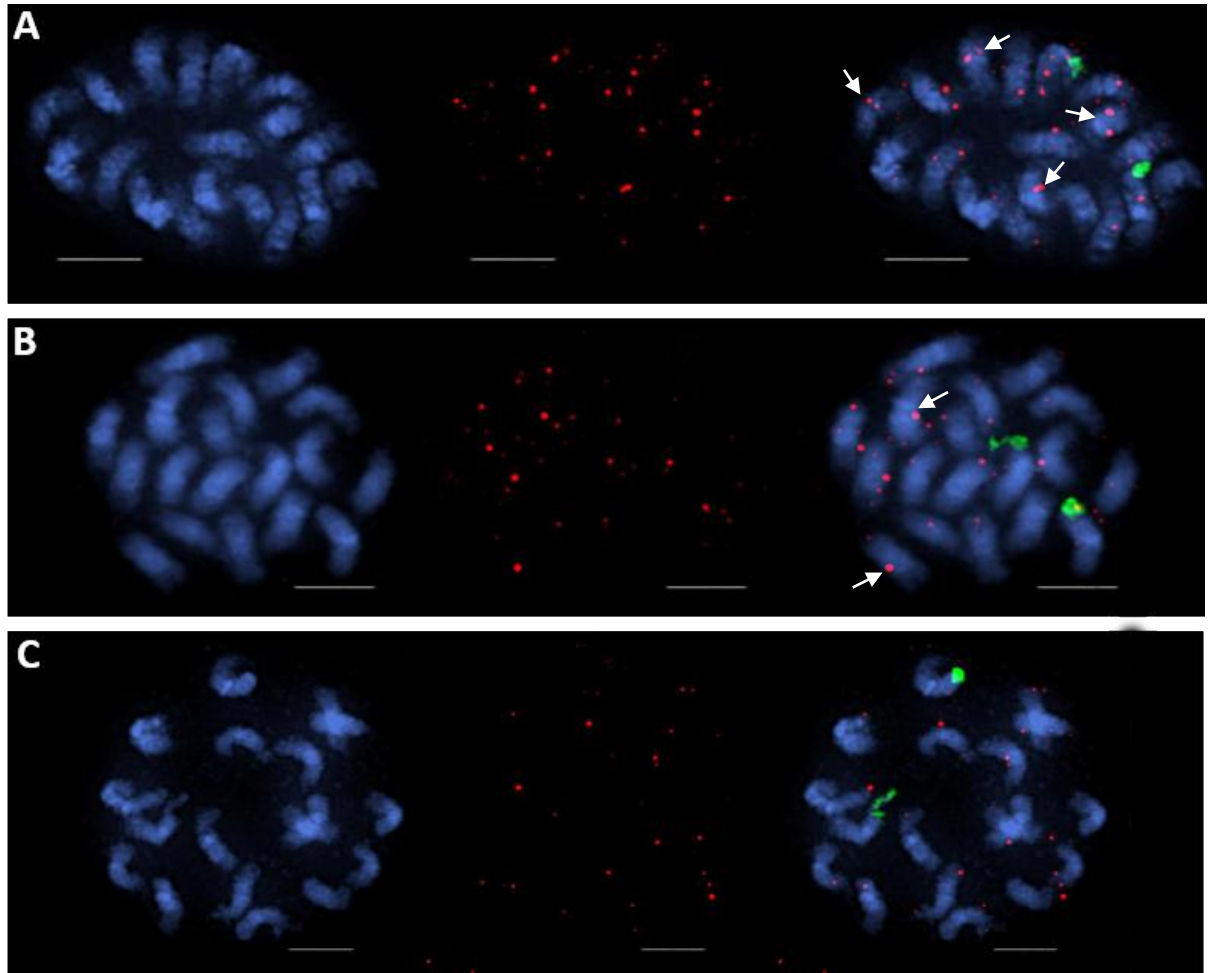


**Figure 3.28: Genomic organization of *B. lomatogona* tandemly repeated sequences**

The digested DNA was separated in blot gels with agarose of which concentration is 2.0 %. The exposition time are 20 hours for B1Sat2 (A), 30 hours for B1Sat3 (B) and 14 hours for B1Sat4 (C). The following restriction enzymes were used to digest genomic DNA of *B. lomatogona*: (1) *MseI*, (2) *MspI*, (3) *ApaLI*, (4) *BstNI*, (5) *StuI*, (6) *NsiI*, (7) *MscI*, (8) *NdeI*, (9) *XhoI*, (10) *DraI*, (11) *BsmI*, (12) *BseGI*, (13) *BamHI*, (14) *ApaI*, (15) *HinfI*, (16) *FokI*, (17) *MaeI*, (18) *RsaI*, (19) *MboI*, (20) *AluI*.

### 3.2.3.3 Chromosomal localization of B1Sat2, B1Sat3 and B1Sat4 along *B. lomatogona* chromosomes

B1Sat2, B1Sat3, and B1Sat4 are not localized in satellite-typical arrays but tend to a dispersed localization on different number *B. lomatogona* chromosomes and signal strength is not strong (Figure 3.29). This result indicates that these satellites were organized in short arrays.



**Figure 3.29: Chromosomal localization of tandem repeats along chromosomes of *B. lomatogona*.** (A) Dispersed localization of B1Sat2 is visible on metaphase chromosomes; (B) The dispersed localization of B1Sat3; (C) The signals of B1Sat4 in distal regions. Blue fluorescence shows DAPI-stained DNA, whereas red fluorescence indicates satellite DNA. Green signals reveal the position of 18S-5.8S-25S rRNA genes. Scale bar is 5  $\mu\text{m}$ .

Satellite family B1Sat2 localizes on 18 chromosomes at distal region. Signal intensity is moderately on ten chromosomes and weakly on eight remaining chromosomes (Figure 3.29A). In addition, signals of B1Sat2 are also identified at pericentromeric region of four chromosomes (arrows in Figure 3.29A).

B1Sat3 signals were observed at distal regions of 16 chromosomes but its strength varies from moderately to very weakly (Figure 3.29B). In which two chromosomes include additional signal at central position (arrows in Figure 3.29B).

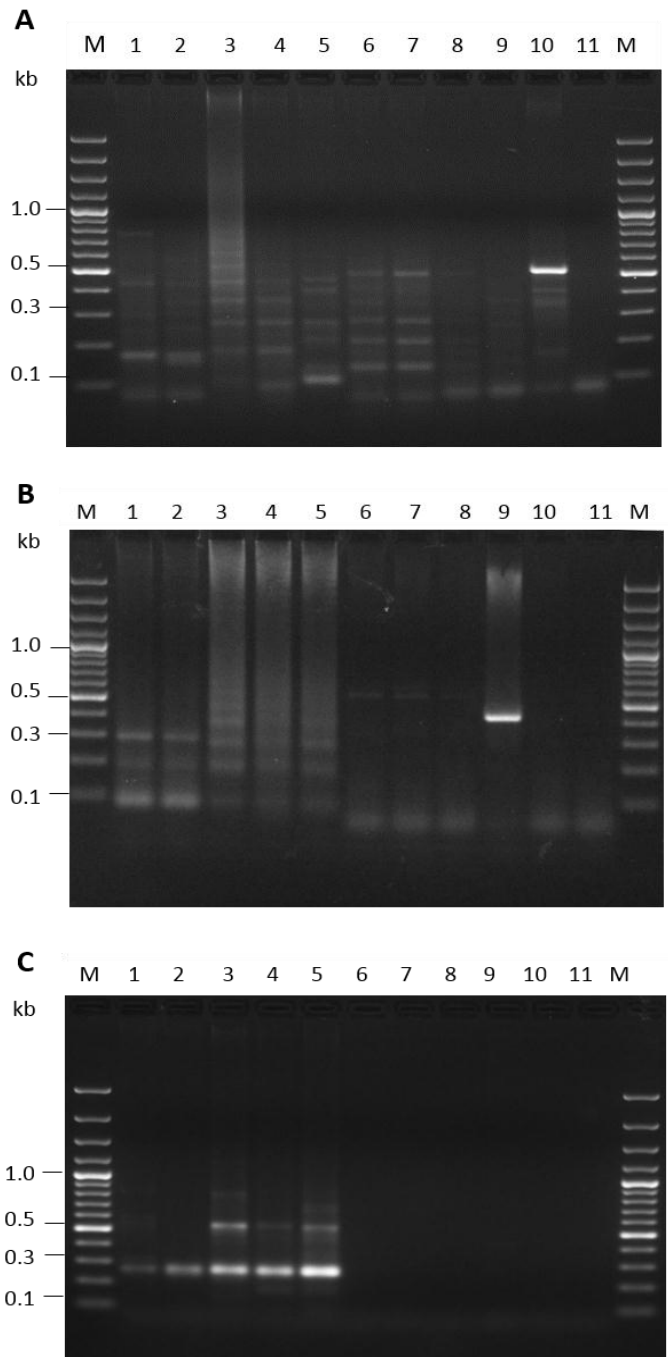
The B1Sat4 satellite family occurs on 14 chromosomes, in which two chromosomes show signal at terminal region of both arms and 12 chromosomes show signal at terminal region of one arm (Figure 3.29C). In the four chromosomes without B1Sat4 signals, there are two of chromosome 1.

#### 3.2.3.5 Distribution and sequence divergence of *B. lomatogona* tandem repeats B1Sat2, B1Sat3 and B1Sat4 in *Beta* species

The B1Sat2 satellite family showed amplification in all tested species (Figure 3.30A) but the DNA bands are not uniform. Sequencing of 30 clones from ten tested species revealed the tandem arrangement of B1Sat2 in four species, including *B. lomatogona*, *B. corolliflora*, *P. procumbens* and *P. webbiana*. Sequences of clones isolated from the six remaining species (*B. vulgaris*, *B. patula*, *B. nana*, *P. patellaris*, *C. quinoa*, and *S. oleracea*) show very low similarity with B1Sat2 sequence and do not arrange in tandem array. The alignment of complete monomers from four species is shown in Figure S11A of the Supplement. The B1Sat2 monomers in *B. lomatogona* have the size of 90 bp, whereas those in *B. corolliflora*, *P. procumbens* and nine monomers in *P. webbiana* have the size of 73 bp. The five monomers of *P. webbiana* have the size of 70 bp (Table 3.11, B1Sat2). These monomers are variable in sequence supported by a quite low identity of 60.3%.

A PCR with the primers of B1Sat3 results in a ladder-like pattern of separated amplicons in *B. lomatogona*, *B. corolliflora* and *B. nana* forming bands up to the trimer. In addition to monomer, dimer and trimer bands, there are sub-bands of which corresponds to a half monomer, one and a half monomer. The sugar beet *B. vulgaris* and wild beet *B. patula* also showed an amplicon ladder. One faint DNA band was observed in three species of *Patellifolia* section and one strong DNA band appeared in *S. oleracea* (Figure 3.30B). Sequencing was only successful for the clones of *B. lomatogona* and *B. corolliflora*, and resulted in 22 complete monomers of *B. lomatogona* and four complete monomers of *B. corolliflora* (Table 3.11, B1Sat3). The alignment of 26 monomers indicated the conserved monomer size of 190 bp, only two monomers of *B. lomatogona* have shorter length because of the variation in the B1Sat3 structure (Figure S12A of the Supplement). As described in section 3.2.3.2 there are two monomers of *B. lomatogona*, which include only subunit sub2 resulting in the monomer length of 172bp, instead of 190 bp. In *B. corolliflora*, three monomers including only subunit sub1 (104 bp) were also identified. These results might contribute to the conclusion that there are several variants of B1Sat3 satellite family in the genome

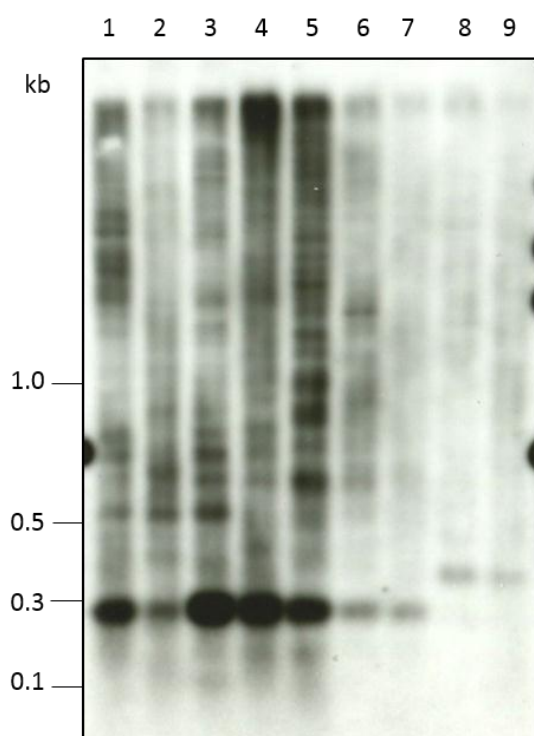
of *B. lomatogona* and *B. corolliflora*. The alignment of the subunits sub1 as well as the subunits sub2 extracted from all monomers of two species *B. lomatogona* and *B. corolliflora* was shown in Figure S14 of the Supplement indicating high similarity in each group of subunits sub1 and subunits sub2 (88.6% and 89.6%, respectively).



**Figure 3.30: PCR from different *Beta* and related species**

The genomic DNA was amplified with satellite-specific primers for satellite families BISat2 (A), BISat3 (B), and BISat4 (C), subsequently separated by gel electrophoresis. The following species were tested: (1) *B. vulgaris*, (2) *B. patula*, (3) *B. lomatogona*, (4) *B. corolliflora*, (5) *B. nana*, (6) *P. procumbens*, (7) *P. webiana*, (8) *P. patellaris*, (9) *S. oleracea*, (10) *C. quinoa*, and (11) Negative control (no genomic DNA). (M) Marker GeneRuler 100 bp Plus DNA Ladder.

BISat4 family showed the amplification of *Beta Corollinae* and *Nanae* species and gave ladder-like banding patterns. The monomer, dimer and trimer units were visible for these species. *B. vulgaris* and *B. patula* have the weak amplicons of monomer and dimer units. The BISat4 units could not be amplified in species of genus *Patellifolia* as well as in *C. quinoa* and *S. oleracea* (Figure 3.30C). The sequencing of 11 clones from *B. vulgaris*, *B. patula*, *B. lomatogona*, *B. corolliflora*, and *B. nana* results in 22 complete monomers. The length of these monomers ranges from 226 bp to 280 bp (Table 3.11, BISat4). The shorter monomer sizes compared to a normal monomer size of 276 bp resulted from a sequence deletion. Two monomers in *B. nana* have a longer monomer size (280 bp) because of an insertion of four nucleotides (CATG). The pairwise similarity of all monomers is 89.0% and the first 30 nucleotides are very conserved (Figure S13 of the Supplement).



**Figure 3.31: Abundance and genomic organization of BISat4 satellite family in genus *Beta* and *Patellifolia***

Restricted genomic DNA with *NdeI* was analyzed by comparative Southern hybridization using probes from the BISat4. The following species were tested: (1) *B. vulgaris*; (2) *B. patula*; (3) *B. lomatogona*; (4) *B. corolliflora*; (5) *B. nana*; (6) *P. procumbens*; (7) *P. patellaris*; (8) *C. quinoa*; and (9) *S. oleracea*.

Although BISat4 is associated with LTR retrotransposons, this satellite family does not have any link to protein domains of LTR retrotransposons. Therefore, BISat4 was selected to investigate its genomic organization as well as abundance in different species (Figure 3.31). BISat4 is mostly uniformly and dispersed distributed in the plants of genus *Beta*. In the genus *Patellifolia*, weaker



signals are detectable. Furthermore, a pattern shift indicates an increased monomer size in the *S. oleracea* and *C. quinoa* genomes. This indicates that BlSat4 monomer length is conserved in the genera *Beta* and *Patellifolia*. The strong signals were observed at monomer size of four species and the signals corresponding to multimers were weaker, indicating the conserved *NdeI* recognized sites in different species. It is also assumed that this satellite family does not organize in long tandem arrays. This may be an additional indicator for association of Blsat4 and LTR retrotransposons.

**Table 3.11: Variation of retrotransposon-associated *B. lomatogona* satellite monomers in the genus *Beta***

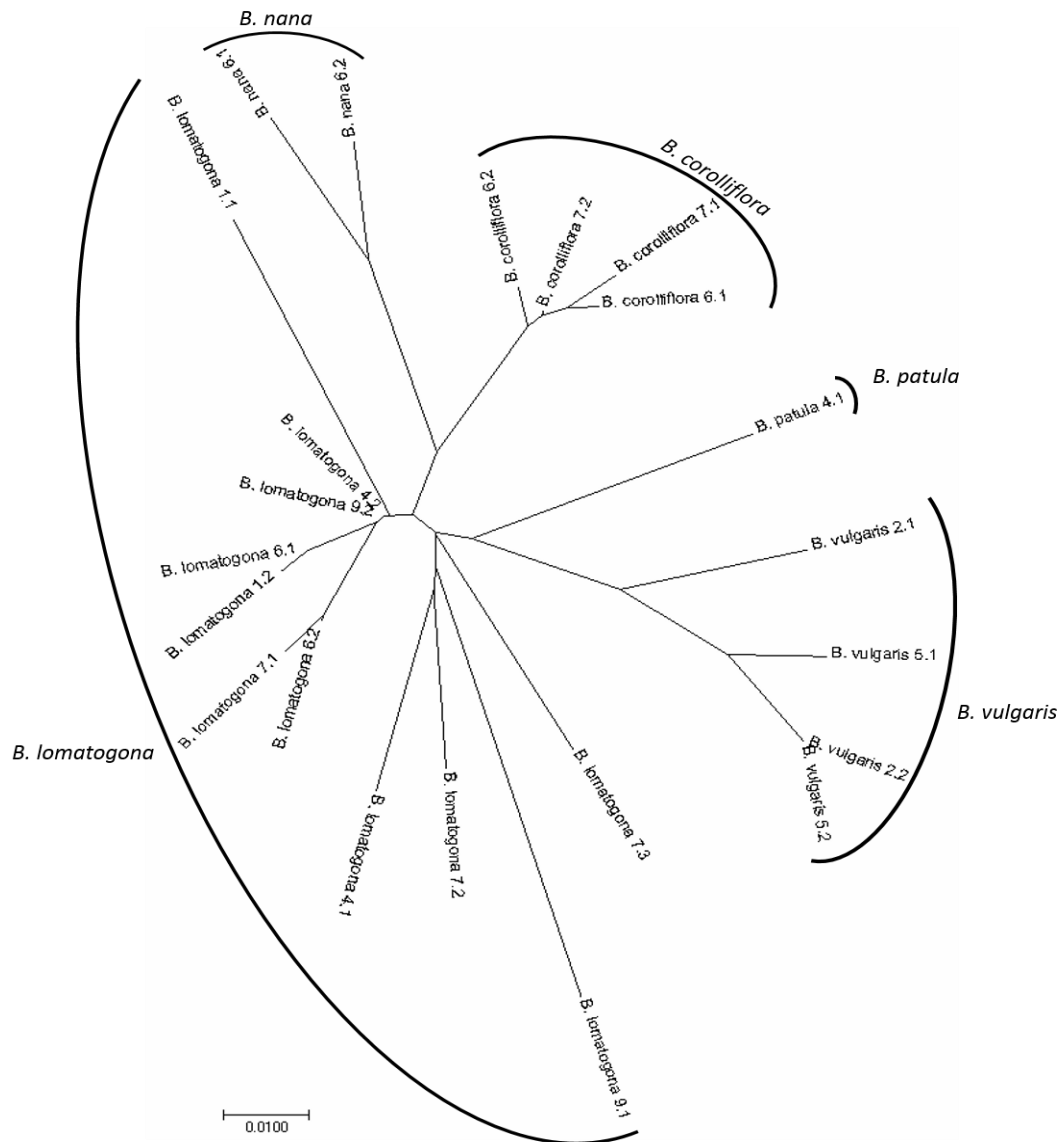
Satellite family	Clone name	Identity to consensus [%]	Type of clone	Monomer sizes of a repeat [bp]
<b>BlSat2</b>	BlSat2_in silico	100	monomer	90
	B. lomatogona_3	87.1	tetramer	89 + 89 + 89 + 91
	B. lomatogona_4	94.5	trimer	90 + 89 + 90
	B. lomatogona_5	91.4	dimer	90 + 91
	B. lomatogona_6	89.3	trimer	90 + 90 + 90
	B. lomatogona_11	94.7	pentamer	90 + 90 + 90 + 90 + 89
	B. corolliflora_1	51.1	monomer	73
	B. corolliflora_5	51.1	monomer	73
	P. procumbens_1	46.8	monomer	73
	P. procumbens_7	47.8	dimer	73 + 73
	P. webbiana_1	46.8	monomer	73
	P. webbiana_3	46	hexamer	73 + 73 + 70 + 70 + 70 +
	P. webbiana_4	45.7	monomer	73
	P. webbiana_5	46.6	hexamer	73 + 73 + 70 + 70 + 73 +
<b>BlSat3</b>	BlSat3_in silico	100	monomer	190
	B. lomatogona_1	93.1	trimer	190 + 190 + 190
	B. lomatogona_4	89.3	dimer	191 + 191
	B. lomatogona_5	87.9	dimer	173 + 190
	B. lomatogona_6	79.7	monomer	164
	B. lomatogona_7	92.5	dimer	190 + 190
	B. lomatogona_8	94.2	monomer	190
	B. lomatogona_9	93.9	monomer	190
	B. lomatogona_10	91.4	monomer	191
	B. lomatogona_11	91.8	monomer	190
	B. lomatogona_12	92.0	trimer	190 + 191 + 190
	B. lomatogona_14	93.2	trimer	190 + 190 + 190

**Table 3.11:** Continued

Satellite family	Clone name	Identity to consensus [%]	Type of clone	Monomer sizes of a repeat [bp]
	B. lomatogona_15	89.8	dimer	190 + 191
	B. corolliflora_2	88.7	monomer	190
	B. corolliflora_3	94.2	dimer	190 + 191
	B. corolliflora_4	89.3	monomer	191
<b>BISat4</b>	BISat4_in silico	100	monomer	276
	B. vulgaris_2	93.5	dimer	275 + 275
	B. vulgaris_5	93.2	dimer	275 + 275
	B. patula_4	84.8	monomer	245
	B. lomatogona_1	96.3	dimer	276 + 276
	B. lomatogona_4	90.1	trimer	275 + 238
	B. lomatogona_6	98.2	dimer	276 + 276
	B. lomatogona_7	90.0	trimer	276 + 273
	B. lomatogona_9	88.9	dimer	275 + 238 + 226
	B. corolliflora_6	95.5	dimer	276 + 276
	B. corolliflora_7	95.5	dimer	276 + 276
	B. nana_6	93.0	dimer	279 + 280

To further investigate the relationship of BISat2, BISat3, and BISat4 sequences in *Beta* species, a neighbor-joining analysis was carried out. The BISat2 sequences were grouped into two distinct clades, one includes sequences from *P. procumbens* and *P. webbiana*, and the other clade comprises sequences from *B. lomatogona* and *B. corolliflora*. In the latter clade, subdivision into separated species was observed, in which a branch of *B. corolliflora* sequences separated from *B. lomatogona* sequences (Figure S11B of the Supplement). This result might suggest that BISat2 sequences were fixed and homogenized in two species of section *Corollinae* as well as two species of the genus *Patellifolia*. The BISat3 sequences originating from *B. lomatogona* and *B. corolliflora* were not be resolved, reflecting by a mixture of clusters (Figure S12B of the Supplement). Interestingly, BISat4 sequences from five *Beta* species were group into species-specific sequences (Figure 3.32). The BISat4 sequences from *B. vulgaris* were arranged more closely to the sequence from *B. patula*. Branches of BISat4 sequences from *B. corolliflora* and *B. nana* are close together and formed a separated clade. The sequences from *B. lomatogona* seem to be more divergent, there are two groups, one is close to the clade of *B. vulgaris* and *B. patula*, the other is close to the clade of *B. corolliflora* and *B. nana*. In this case, BISat4 reflects the relationship among five respective

species in the genus *Beta*, which is in line with current taxonomy of the genus *Beta* (FordLloyd, 2005).



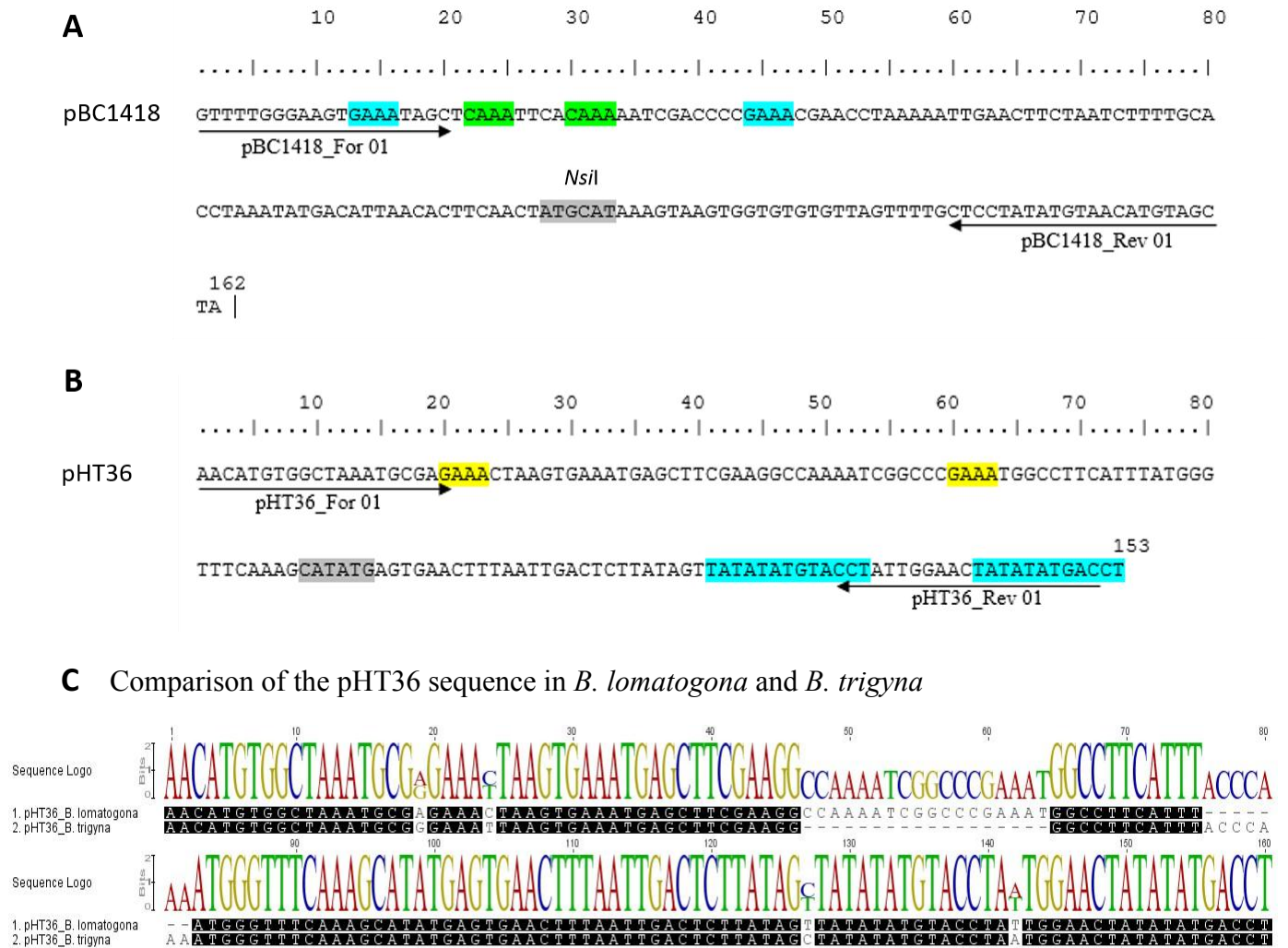
**Figure 3.32: Dendrogram representating the relationship between BISat4 sequence in *Beta* species**

### 3.2.4 Characterization of satellites pBC1418 and pHT36 in the *B. lomatogona* genome

The satellite families pBC1418 and pHT36 identified in *B. corolliflora* and *B. trigyna*, respectively (Gao *et al.*, 2000; Shmidt *et al.*, 1993) are known satellite but their characterization is not complete. The first characterization of pBC1418 was conducted in *B. corolliflora* and this satellite is distributed on four chromosomes, in the close vicinity to the 5S rRNA genes (Gao *et al.*, 2000).

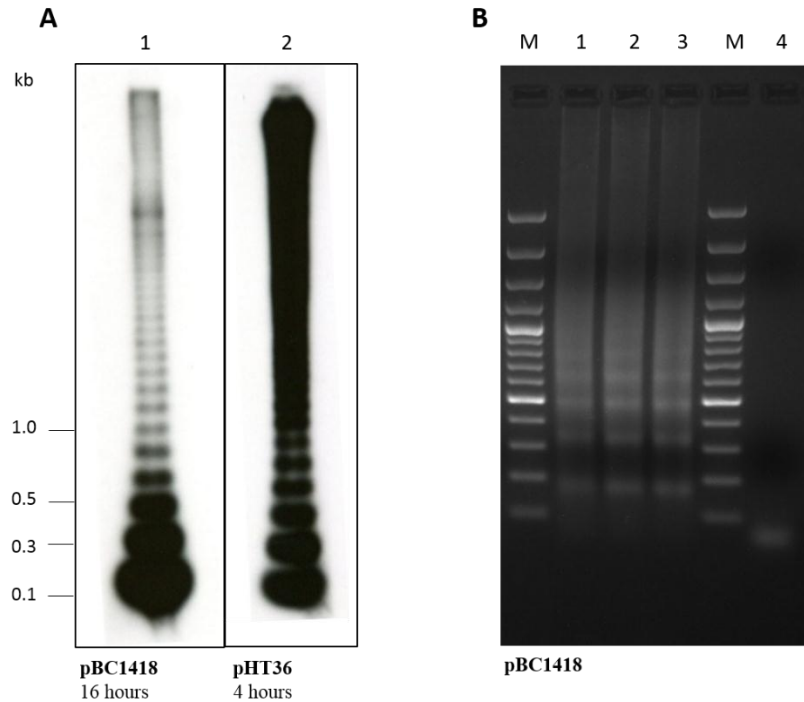
However, comparison of 5S rDNA and pBC1418 sequences revealed that they are not related. The comparative Southern hybridization showed that the satellite family pBC1418 is a satellite specific for the section *Corollinae* (Gao *et al.*, 2000). The probe of this satellite family used in comparative Southern experiment was 378 bp in length, but the monomer size of pBC1418 was not reported. Similarly, satellite family pHT36 was firstly identified in *B. trigyna* – a species of section *Corollinae* with monomer size of 142 bp (Schmidt *et al.*, 1993). The comparative hybridization of the pHT36 satellite probe with other species revealed the occurrence of this satellite in all three sections of genus *Beta*, while no hybridization signal was detected in the genus *Patellifolia* (Schmidt *et al.*, 1993). However, the chromosomal distribution of pHT36 has not been investigated in any *Beta* species. In this section, these two satellite families pBC1418 and pHT36 have been characterized in *B. lomatogona*.

In *B. lomatogona*, pBC1418 has the monomer size of 162 bp (Figure 3.33A). There are duplications of two sequence motifs (blue and green colors in Figure 3.33A) in this monomer sequence. The hybridization of pBC1418 to *B. lomatogona* DNA restricted with endonuclease *NsiI* revealed a ladder-like pattern characteristic for satellite DNA (Figure 3.34A, lane 1). Very strong signals at monomer, dimer and trimer size were observed after 16 hours exposure, indicating a conserved *NsiI* restriction sites in this satellite. This family localizes in the centromeric regions (Figure 3.35A). Strong FISH signals detected in the centromere or near the centromere heterochromatin of one chromosome pair indicated that pBC1418 is arranged in very long arrays. These signals might be on chromosome 4 as Gao *et al.* (2000) described in *B. corolliflora*. Additional faint signals also occurred in the centromeric regions of an additional other chromosome pair. Moreover, PCR with pBC1418-specific primers indicated that this satellite family is also present in the *B. macrorhiza* genome with the same ladder pattern as in *B. lomatogona* and *B. corolliflora* (Figure 3.34B).



**Figure 3.33: Monomer consensus of pBC1418 (A) and pHT36 (B) in *B. lomatogona***

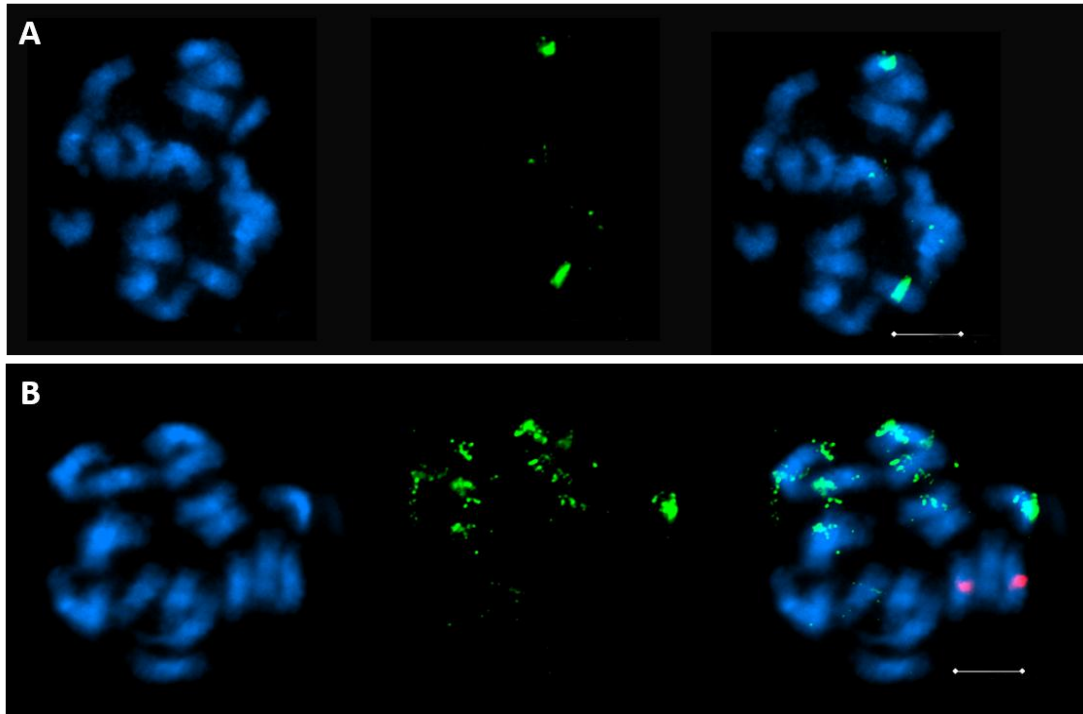
For each monomer the position of the primers was marked with arrows below the sequences (the arrow head represents the 3' end of the primer). The restriction sequence sites of enzyme used to release typical ladder-like pattern were marked by grey boxes. The yellow, blue and green colors indicated the repeated sequence motifs. Comparison of the pHT36 satellite sequence between *B. lomatogona* and *B. trigyna* was shown (C).



**Figure 3.34: Tandem organization of two satellite families pBC1418 and pHT36.**

(A) Genomic DNA of *B. lomatogona* was restricted with *Nsi*I (1) and *Nde*I (2) and hybridized with pBC1418 (1) and pHT36 (2). (B) PCR with pBC1418 satellite-specific primers in *B. lomatogona* (1), *B. macrorhiza* (2), and *B. corolliflora* (3), (4) negative control without genomic DNA, (M) marker GeneRuler 100 bp Plus DNA Ladder.

In the *RepeatExplorer* output of *B. lomatogona*, pHT36 is the third abundant satellite and makes up 8.78% of satellite component. Compared with 142 bp in *B. trigyna* (Schmidt *et al.*, 1993), the monomer size of this family in *B. lomatogona* with 153 bp is longer (Figure 3.33C) and they share a similarity of 81.9%. In the pHT36 monomer sequence, GAAA motif (yellow color in Figure 3.33B) and TATATATGTACCT motif (blue color in Figure 3.32B) were repeated twice. Southern hybridization of pHT36 probe with *B. lomatogona* genomic DNA restricted by *Nde*I showed strong signals at both monomer and multimers after a 4-hour exposition (Figure 3.34A, lane 2), indicating the high abundance in the *B. lomatogona* genome. This satellite is amplified in both large and small arrays in the pericentromeric regions of eight chromosomes (Figure 3.35B). In particular, four chromosomes show strong signals in the pericentromere and also have additional signals in distal regions on one arm of these chromosomes. Weaker signals are present in the centromeric regions of four other chromosomes. All the FISH signals were not co-localized with 5S rDNA on chromosome 4.



**Figure 3.35: Chromosomal localization of the satellite pBC1418 and pHT36 along chromosomes of *B. lomatogona***

Blue fluorescence shows DAPI-stained DNA, whereas green fluorescence indicates satellite DNA. Scale bar is 5  $\mu$ m. (A) Green signals of pBC1418 on metaphase chromosomes. (B) Satellite family pHT36 with green signals on four chromosome pairs. Red signals indicated 5S rDNA.

### 3.2.5 *Beta lomatogona* satellite overview

In conclusion, molecular and cytogenetic results give evidence that BlSat1, Blsat5 and Blsat6 were typical satellites. Dissimilarly, the other three repeats (BlSat2, BlSat3, and BlSat4) were non-typical satellites of which two were associated with Ogre/Tat retroelements.

The AT-content of the repeats ranges from 42.5% to 74.8% where most of them are high enough to be characterized as rich AT satellite DNA. BlSat1 has the highest AT-content (74.8%), followed by BlSat5 (61.7%). BlSat2 and BlSat6 contain the low AT-content of 48.9% and 42.5%, respectively. 50.0% of AT-content was observed in BlSat3 and BlSat4. Compared with the known satellite families pEV1 (AT content of 59%), pBV1 (69%), pAv34 (62%) in *B. vulgaris*, the AT content of *B. lomatogona* has a wider range. Additionally, it was noticed that the highest sequence variability belongs to BlSat3 and BlSat5 (approximate 84.0% of identity to consensus monomer sequence), followed by BlSat6 (89.6%), BlSat1, BlSat2 and BlSat4 exhibit lower variation (more than 90.0% of identity).

In addition to the Southern hybridization and PCR results, mapping of read sequences of reference species (Table 3.1) against satellite monomer sequences was also a useful method to investigate the abundance of six identified satellite families. Compared to the Southern hybridization results, this mapping shows the expected results (Table 3.12). The first four satellite families B1Sat1-B1Sat4 have the largest proportion in *B. lomatogona* genome. The satellite B1Sat5 is particularly noticeable because this satellite is very abundant in *B. nana*. B1Sat6 is also a slightly more abundant in *B. nana* compared with that in *B. lomatogona*.

The identified satellite families also occur in other species of genus *Beta*. B1Sat1, B1Sat3, B1Sat5 and B1Sat6 show a small proportion in the genome of *Beta*, *Corollinae* and *Nanae* species. Family B1Sat2 is more abundant in *B. lomatogona* genome, accounting for nearly 0.2%, but scarce in *B. nana* (making up only 0.00054%). In two species of genus *Patellifolia* and two out group species (*C. quinoa* and *S. oleracea*), these satellites do not occur. B1Sat4 satellite appears in the genome of all tested species.

**Table 3.12: Mapping of six identified satellite sequences to read sequences of *Beta* and related species**

The values in each field are the percentage of reads mapped to satellite, which correspond to the genome proportion of each satellite in respective species. Dark grey highlighted fields indicate the highest genome proportion of satellites, while greyish highlighted fields are corresponding to a lower proportion.

Satellite family	<i>B. vulgaris</i>	<i>B. patula</i>	<i>B. lomatogona</i>	<i>B. nana</i>	<i>P. procumbens</i>	<i>P. patellaris</i>	<i>C. quinoa</i>	<i>S. oleracea</i>
B1Sat1	0.0254	0.025	0.296	0.0178	0	0	0	0
B1Sat2	0	0	0.198	0.00054	0	0	0	0
B1Sat3	0.0006	0.0025	0.121	0.0214	0	0	0	0
B1Sat4	0.0046	0.0048	0.074	0.03	0.0031	0.0037	0.00065	0.0034
B1Sat5	0.002	0.0025	0.07	0.3	0	0	0	0
B1Sat6	0.0075	0.0056	0.013	0.0186	0	0	0	0

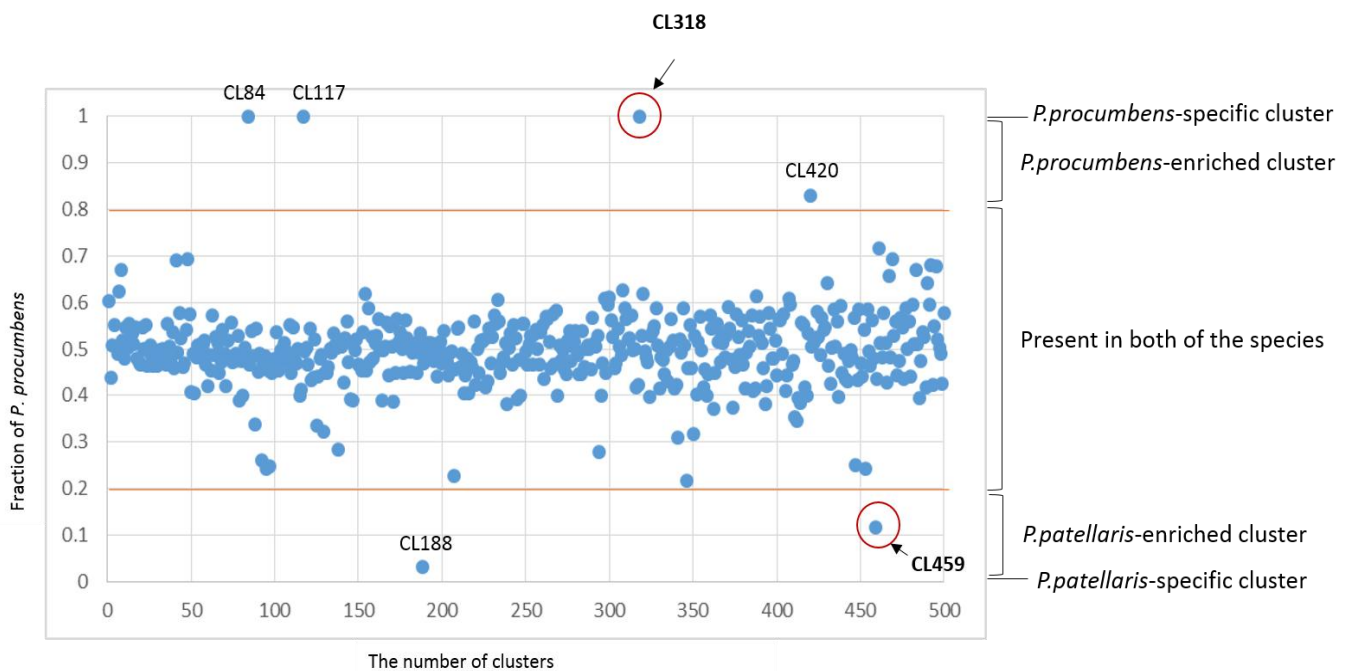
### 3.3 Satellite DNAs in the genus *Patellifolia*

In this part of the thesis, comparative read clustering of Illumina sequence reads from *P. procumbens* and *P. patellaris* by using *RepeatExplorer* was performed. The aim of this work is to identify species-specific and species-enriched repeats in each species, from which it is possible to have insight into the polyploid origin of *P. patellaris*.



### 3.3.1 Comparative analysis of repetitive DNA sequences of *Patellifolia procumbens* and *Patellifolia patellaris*

Seven million paired sequence reads from each species (from the group research of Cell and Molecular biology in Plants, TU Dresden) were used as input for comparative read clustering. The first 500 clusters were separated into three categories: Sequences similarly abundant in both species, enriched sequences and species-specific (Figure 3.36). Although most of the clusters are shared by both of species, there are some species-specific or species-enriched clusters, which can be used to discriminate the two species (Table 3.13).



**Figure 3.36: Comparative analysis of two species *P. procumbens* and *P. patellaris* based on the cluster fraction** X axis showed the cluster number. Y axis indicated the *P. procumbens* fraction that was calculated by the proportion of *P. procumbens* reads in total reads of the cluster. The value of 1 means that all reads in the cluster are originated from *P. procumbens* and the cluster is specific for the *P. procumbens* genome. In contrary, the value of 0 means that all reads in the cluster originated from *P. patellaris* and the cluster is specific for the *P. patellaris* genome. The values between the two red lines (from 0.2 to 0.8) indicated the similarity comparison of sequence reads from two species.




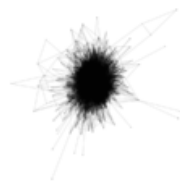
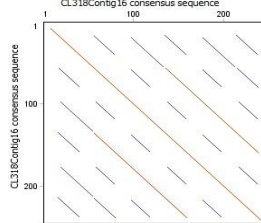

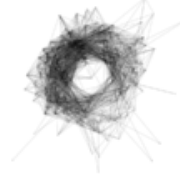
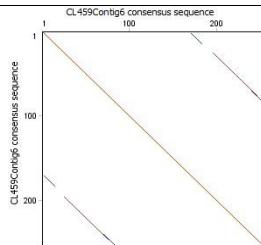
Three clusters (CL84, CL117, and CL318) are specific for *P. procumbens* genome, one cluster (CL420) showed an abundant distribution in *P. procumbens*, and two clusters (CL188, CL459) are the most abundant in *P. patellaris* genome. Based on graph shapes and consensus sequences, no annotation could be assigned to the clusters CL84, CL117 and CL420. The cluster CL188 was annotated as Beon element, a Ty3-gypsy retrotransposon detected in *B. vulgaris* (Weber *et al.*,

2013), with 114 hits (6.42%). The other clusters including clusters CL318 and CL459 were annotated as the minisatellite pBC1447 (Gao *et al.*, 2000) and one new satellite, respectively. Within the scope of the thesis, the two satellite families (corresponding to CL318 and CL459) were analyzed in detail.

Cluster CL318 included 359 sequence reads grouped into 23 contigs and all reads originated from the *P. procumbens* genome (accounted for 0.03% of the genome). It means that this cluster is presumably species-specific. The analysis of the three largest contigs in cluster CL318 resulted in a satellite family with the monomer size of 40 bp. This satellite was designated as minisatellite PproSat1 (*Patellifolia procumbens* Satellite 1). The star-like graph shape of this cluster as well as the repetitive character illustrated by parallel lines in the dotplot supports classification as tandem repeat.

Cluster CL459 comprises 10 contigs with the total of 163 sequence reads, from which 144 reads originated from the *P. patellaris* genome (made up 88.3% of the total reads of the cluster) and only 19 reads from *P. procumbens* genome (11.7%). Based on the circular graph shape and on the dotplot of the largest contig, this cluster is a candidate for a new satellite family designated PpatSat1.

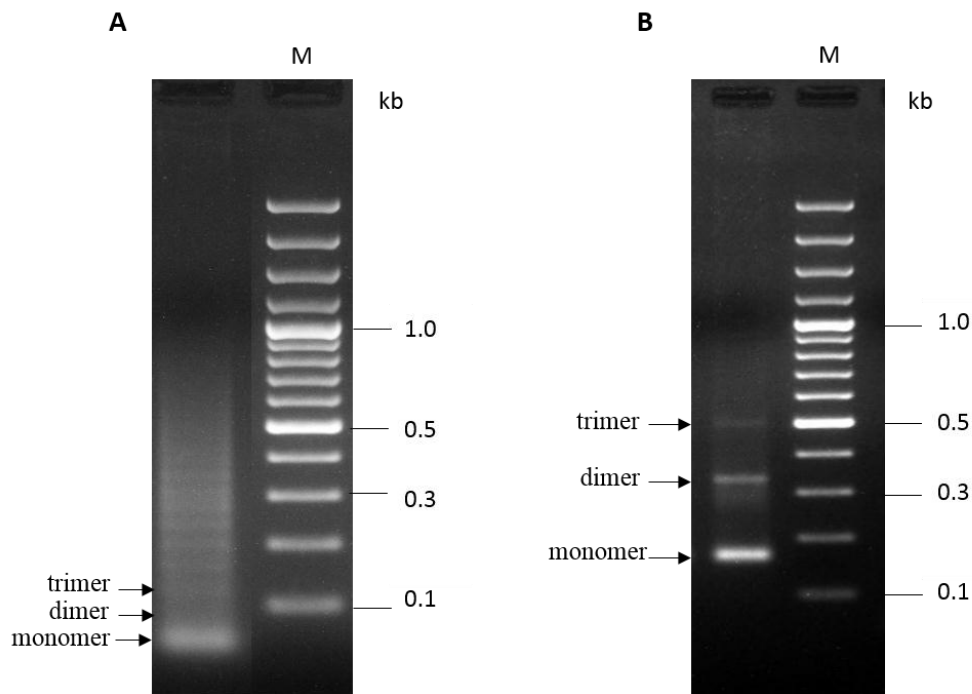
Table 3.13: Species-enriched and -specific clusters from comparison of *P. procumbens* and *P. patellaris* genomes

Cluster	No. of reads in <i>P. procumbens</i>	No. of reads in <i>P. patellaris</i>	Graph layout	Dotplot	Annotation
CL84	5446	0		-	unclassified
CL117	3863	0		-	unclassified
CL188	47	1429		-	LTR retrotransposon Ty3-gypsy, chromovirus, Beon family (Weber <i>et al.</i> , 2013)
<b>CL318</b>	<b>359</b>	<b>0</b>			<b>Satellite</b> <b>PproSat1</b> Similarity to pBC1447 (Gao <i>et al.</i> , 2000)
CL420	161	33		-	unclassified
<b>CL459</b>	<b>19</b>	<b>144</b>			<b>Satellite</b> <b>PpatSat1</b>



(Figure 3.39A). The homology between the clone monomers and bioinformatic monomer sequence is 87.2 %. The shorter monomers resulted from nucleotide deletions, which are 38 bp and 39 bp in length (Table 3.14).

With PpatSat1-specific primer pairs, the amplicons corresponding to the monomer, dimer and trimer of satellite family PpatSat1 were amplified in *P. patellaris* (Figure 3.38B). The strongest band at the monomer size of 170 bp was indicated with an arrow. Monomer, dimer and trimer bands were purified from the agarose gel before being cloned into pGEM-T vector. Eight clones including at least one complete monomer were obtained and analyzed (Table 3.14). A total of 12 monomers was aligned and showed a pairwise identity of 87.4% (Figure 3.39B). The monomer size ranges from 167 bp to 174 bp. The variation in monomer length is due to insertions and deletions.



**Figure 3.38: PCR from (A) *P. procumbens* and (B) *P. patellaris* DNA for verification of the tandem repeat organization**

The agarose gel electrophoresis was carried out with gels of 2 % agarose concentration. The monomer positions are marked with arrows. (A) Ladder-like patterns of the PproSat1 minisatellite, and (B) PpatSat1 satellite, M: marker GeneRuler 100 bp Plus DNA Ladder.

**Table 3.14: Clones of two satellite PproSat1 in *P. procumbens* and PpatSat1 in *P. patellaris***

For each satellite family, the clone used for hybridization was marked with asterisk

Satellite family	Clone name	Length of insert [bp]	Number of monomers	Monomer length [bp]	Identity to consensus [%]
PproSat1	PproSat1_1	200	5	40 + 39 + 39 + 39 + 38	92.1
	PproSat1_2	200	5	40 + 39 + 39 + 39 + 38	92.1
	PproSat1_4*	200	5	40 + 39 + 39 + 39 + 38	92.1
PpatSat1	PpatSat1_1	172	1	171	93
	PpatSat1_2	172	1	171	93
	PpatSat1_3	333	1	168	93.5
	PpatSat1_4	333	1	169	93.5
	PpatSat1_5	501	2	167 + 169	92.7
	PpatSat1_7	325	1	168	92.4
	PpatSat1_8	510	3	170 + 169 + 174	90.9
	PpatSat1_10*	335	2	169 + 167	93.5

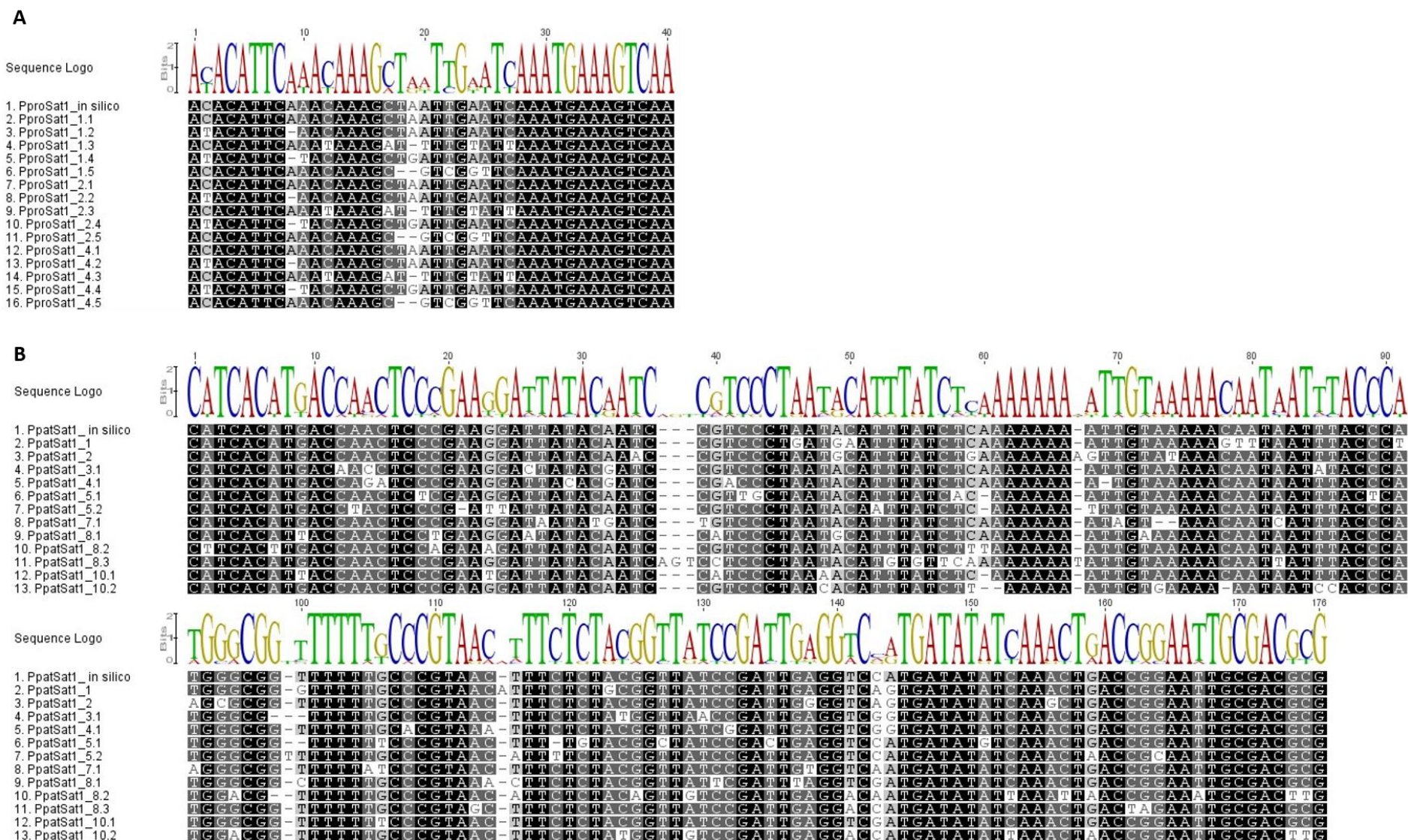
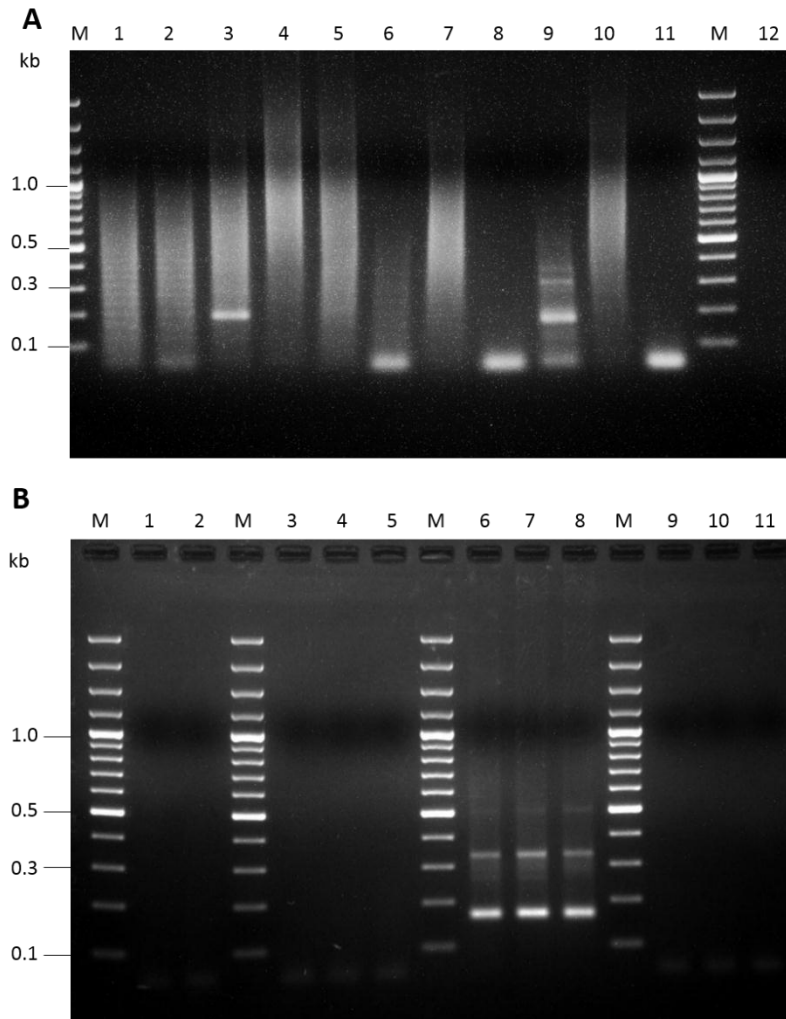


Figure 3.39: Sequence alignment of (A) PproSat1 and (B) PpatSat1 monomers.

### 3.3.2.2 Genome organization and satellite distribution in *Patellifolia* and related species

In order to confirm genome specificity and to gain insight into distribution of PproSat1 and PpatSat1 among related species, PCR was performed with genomic DNA of *Beta*, *Patellifolia* species, *C. quinoa* and *S. oleracea* (Figure 3.40).



**Figure 3.40: Garden PCR of (A) PproSat1 and (B) PpatSat1**

The genomic DNA was amplified with satellite-specific primers and subsequently separated by gel electrophoresis. (A) Primers for PproSat1; and (B) Primer for PpatSat1. The following species were tested: (1) *B. vulgaris*, (2) *B. patula*, (3) *B. lomatogona*, (4) *B. corolliflora*, (5) *B. nana*, (6) *P. procumbens*, (7) *P. webbiana*, (8) *P. patellaris*, (9) *S. oleracea*, (10) *C. quinoa*, (11) Negative control (no genomic DNA), and (12) PCR mix without primers. (M) Marker GeneRuler 100 bp Plus DNA Ladder.

A ladder-like pattern of PproSat1 is visible in *B. vulgaris*, *B. patula* and *P. procumbens*. In *B. corolliflora*, *B. nana*, *P. webbiana*, and *C. quinoa* only smears were observed. The PCR results verified that the PproSat1 monomer is 40 bp. Due to short monomer size it is difficult to separate

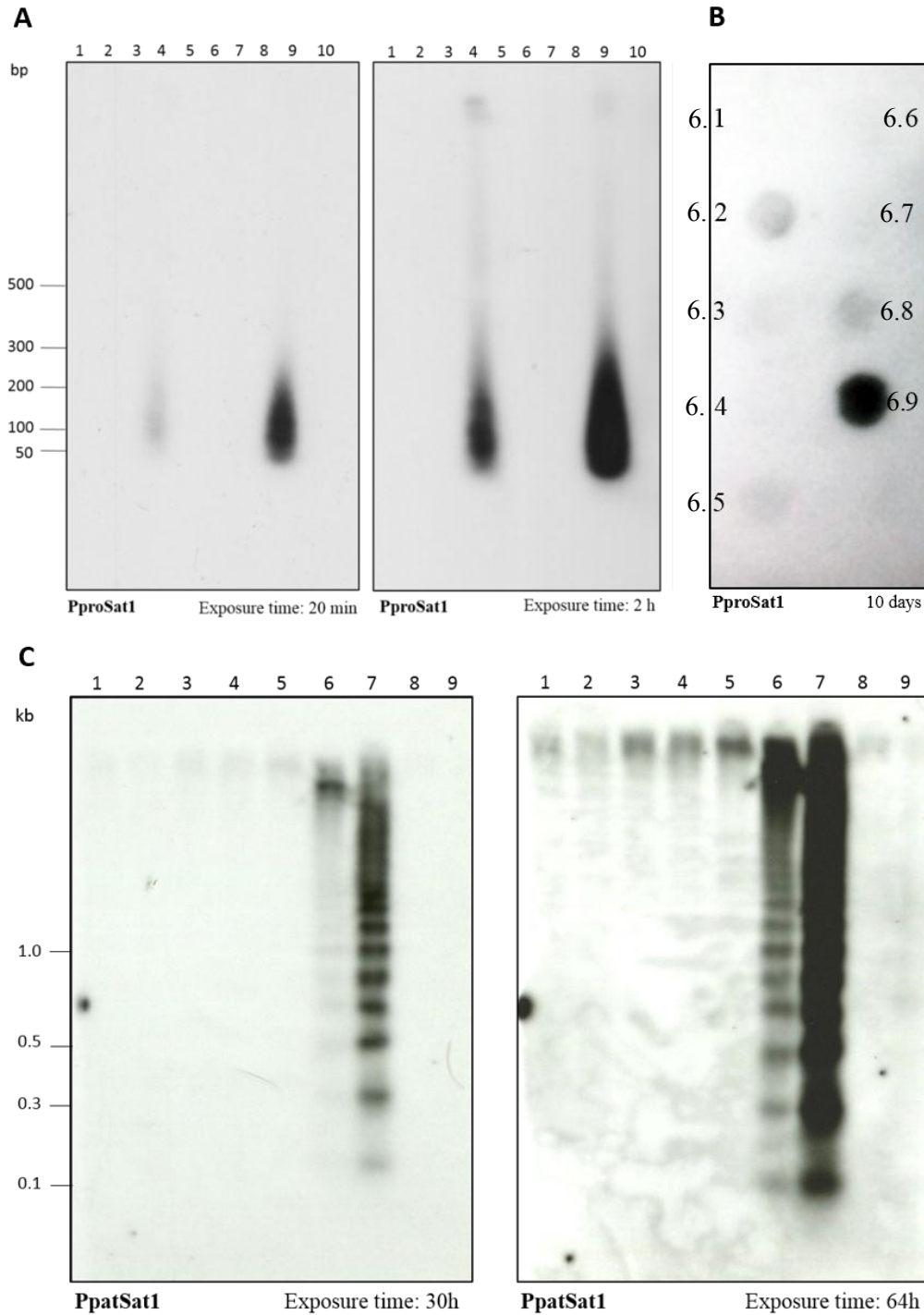


individual DNA bands by gel electrophoresis. Some specific bands were amplified in *B. lomatogona* and *S. oleracea* (Figure 3.40A). As PproSat1 has a short monomer size and high AT content, possibilities for primer design were limited. The chosen primers overlap in their 5' region for five nucleotides. Therefore, in order to test for self-primers binding PCR reactions without genomic DNA (Figure 3.40A, lane 11) and without PproSat1-primer pair (Figure 3.40A, lane 12) were performed. No PCR amplification was observed in both lanes 11 and 12, indicating PproSat-specific primers are not self-binding and PCR amplicons in the respective species resulted from genomic DNA. Interestingly, the absence of PproSat1 family in the *P. patellaris* genome was also reflected by PCR (Figure 3.40A, lane 8).

PCR with PpatSat1-specific primers resulted in a ladder pattern in all three species of the genus *Patellifolia* (Figure 3.40B, lane 6, 7, 8), no signal was observed in the genus *Beta* as well as two out group species. This is an indication that the satellite family PpatSat1 is genus *Patellifolia*-specific. The amplicon of up to the trimer bands was visible.

Subsequently, comparative Southern hybridization was performed to investigate the abundance of the two satellite families in *Patellifolia* and related species. Probe PproSat1\_1 was hybridized to *AluI* digested genomic DNA of five *Beta* species, three *Patellifolia* species and the out group species, *C. quinoa* and *S. oleracea*, in order to investigate the genomic organization as well as the species distribution of minisatellite PproSat1. A ladder pattern observed in *C. quinoa* after only 20 minutes of exposition indicated a high abundance of PproSat1 in this genome (Figure 3.41A, lane 9). After 2-hour exposure, the ladder-like pattern was visible in *B. corolliflora* (Figure 3.41A, lane 4). The pattern was similar in both species with the strongest bands corresponding to dimers (80 bp) and trimers (120 bp). For longer exposition time of 6 days, the ladder pattern was still not observed in *P. procumbens* (data not shown), but PCR with PproSat1-specific primers (Figure 3.38A) and dot blot hybridization with PproSat1 probe (Figure 3.41B; dot 6.2, 6.5, 6.8, and 6.9) indicated the occurrence of PproSat1 satellite family in *P. procumbens*.

Comparative Southern hybridization results indicated that satellite family PpatSat1 are tandemly arranged in the *Patellifolia* genomes. The pattern was similar in both tested species and the ladder started at 170 bp corresponding to monomer size. No hybridization signal was observed in *Beta* species as well as in *C. quinoa* and *S. oleracea* (Figure 3.41C).



**Figure 3.41: Autoradiograph of Southern hybridization shows the abundance and organization of (A) minisatellite PproSat1 and (C) satellite family PpatSat1 in *Patellifolia* and related species.**

The following species were tested in (A) and (C): (1) *B. vulgaris*; (2) *B. patula*; (3) *B. lomatogona*; (4) *B. corolliflora*; (5) *B. nana*; (6) *P. procumbens*; (7) *P. webbiana*; (8) *P. patellaris*; (9) *C. quinoa*; and (10) *S. oleracea*. (B) The dot blot hybridization between the PproSat1 probe and different *P. procumbens* genomic DNA concentrations.

Although the different abundance of this family was not clear in PCR experiment, but *RepeatExplorer* output and Southern experiment strongly indicated that PpatSat1 is more abundant in *P. patellaris* compared with *P. procumbens* (Figure 3.41B). In this case, restriction endonuclease *MspI* resulted in a ladder pattern in both *P. procumbens* and *P. patellaris*. However, its isochizomer *HpaII* did not result in any ladder pattern or smear (data not shown), which indicated the methylations in the internal cytosine of CCGG motif in PpatSat1 satellite. In *P. patellaris*, hybridization signal at monomer is weaker than signals at multimers, this is an indication of not very conserved *MspI* recognized sites in PpatSat1 sequence.

Comparison with known repeats revealed high identity of PproSat1 to minisatellite pBC1447 (from *B. corolliflora*, Gao *et al.*, 2000). In addition, CqSat1 (from *C. quinoa*, Ost, 2017) was also annotated as minisatellite pBC1447. An alignment of the three minisatellite families indicated high similarity (92.5% - 95% identity), only five nucleotide positions showed a variation in the total of 40 nucleotides (Figure 3.42). This indicated that although three minisatellites were identified from distantly related species, they might belong to the same repeat family.

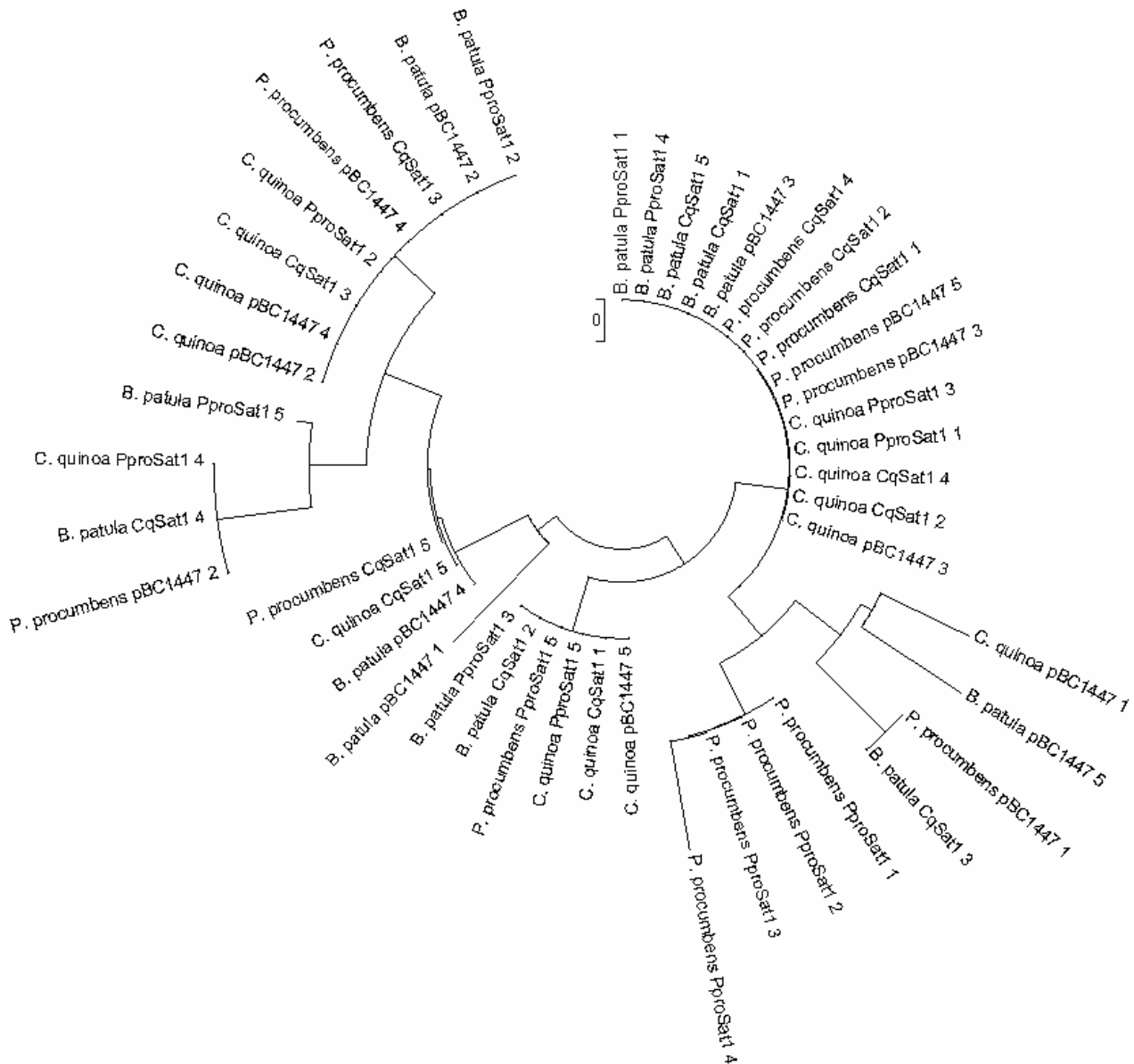


**Figure 3.42:** Sequence alignment of PproSat1 from *P. procumbens*, CqSat1 from *C. quinoa*, and pBC1447 from *B. corolliflora*.

Mapping of available sequence reads of different species (Table 3.1) against the three minisatellite sequences revealed a relative amplification pattern of each minisatellite in these species. All three minisatellites are very abundant in the genome of *C. quinoa* (accounting for 3.9% of the genome), whereas only approximately 0.03% represented in the *P. procumbens* genome. Interestingly, the mapping also indicated appearance of these minisatellite families in *B. patula* with 0.01%, even though no Southern hybridization signal of Pprosat1 was observed in this species.

From the mapping, five monomers of each minisatellite family were extracted in *B. patula*, *P. procumbens*, and *C. quinoa*. A total of 45 monomers was aligned resulting in high pairwise identity

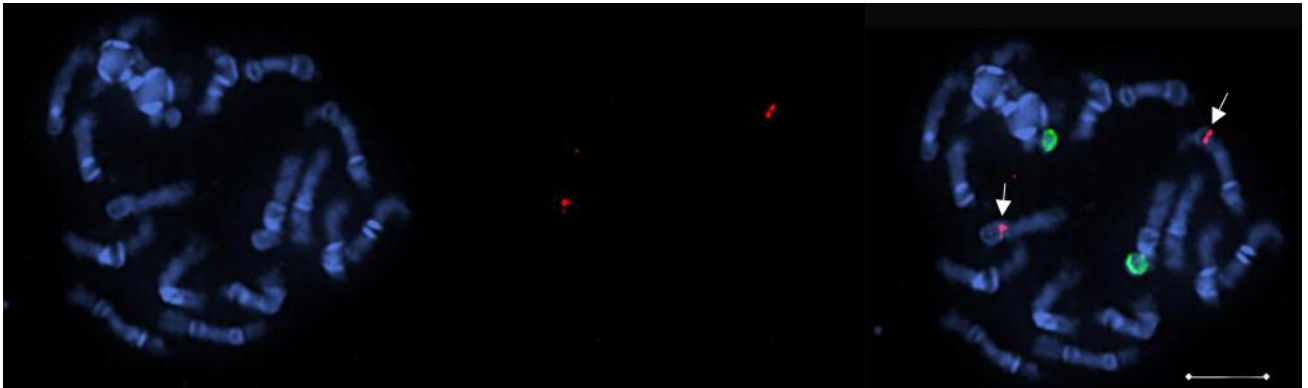
with 93.7% (Figure S15 of the Supplement). The variation among these monomers might result from random point mutations, this was also shown in Neighbor-joining clustering where no species-specific as well as satellite-specific cluster was observed (Figure 3.43). The three minisatellite families have formed a mixture of clusters, only four of five PproSat1 monomers from *P. procumbens* were grouped in a separated cluster. This result also revealed that PproSat1, CqSat1, and pBC1447 are the same repeat family, but they were amplified differently in species.



**Figure 3.43:** Neighbor-joining clustering analysis of three minisatellite monomer sequences in *B. patula*, *P. procumbens*, and *C. quinoa*

### 3.3.2.3 Chromosomal localization in *Patellifolia* species

The chromosomal location of minisatellite PproSat1 in the *P. procumbens* genome is shown in Figure 3.43. PproSat1 was detectable on one chromosome pair (shown by the red signals indicated with arrows in Figure 3.44) at intercalary regions of one chromosome arm. The two 18S rDNA signals are not co-localized with two signals of PproSat1.

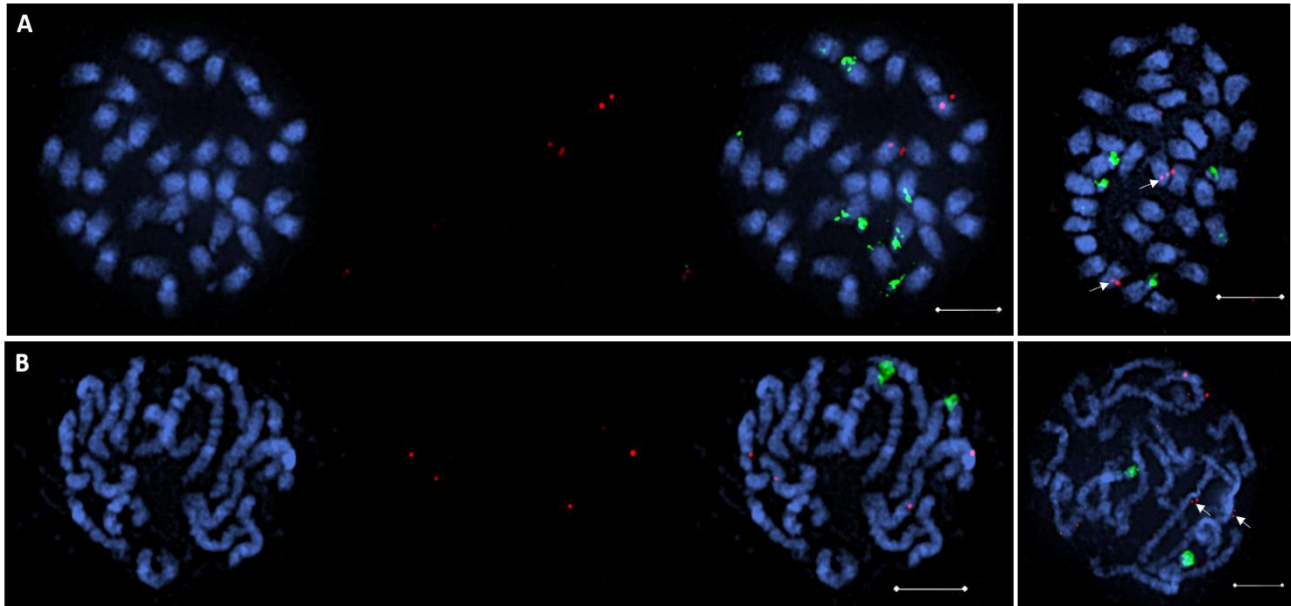


**Figure 3.44: Chromosomal localization of *P. procumbens*-specific satellite PproSat1**

Blue and red fluorescence shows DAPI-stained DNA and minisatellite family PpatSat1, respectively. Scale bar is 5  $\mu\text{m}$ .

The chromosomal localization of PpatSat1 was studied on metaphases of *P. procumbens* ( $2n = 2x = 18$ ) and *P. patellaris* ( $2n = 4x = 36$ ) by FISH. The satellite showed weak signals in the pericentromeric regions of four *P. procumbens* chromosomes (Figure 3.45B), while prominent clustering of this family was observed in the pericentromeric heterochromatin of two *P. patellaris* chromosomes (Figure 3.45A). In the right panel of Figure 3.44, arrows indicated similar chromosomal position of satellite family PpatSat1 in *P. procumbens* and *P. patellaris*. Through polyploid process this satellite was differently amplified, but the chromosomal position has been retained.

Table 3.15 summarized the number of signals as well as the supposed number of chromosome pairs from PproSat1 and PpatSat1 in two *Patellifolia* species.



**Figure 3.45: Chromosomal localization of PpatSat1 satellite on (A) metaphase chromosomes of *P. patellaris* and (B) prometaphase chromosomes of *P. procumbens*.**

Blue fluorescence shows DAPI-stained DNA, red fluorescence indicates satellite family PpatSat1. Green signals reveal the position of 18S-5.8S-25S rRNA genes. The pericentromeric PpatSat1 signals were marked with arrows. Scale bar is 5  $\mu\text{m}$ .

**Table 3.15: Chromosomal distribution of PproSat1 and PpatSat1 along *Patellifolia* chromosomes**

Satellite family	Genus	Species	Genome type	No. of signals	No. of chromosome pairs with signals
PproSat1	<i>Patellifolia</i>	<i>P. procumbens</i>	$2n = 2x = 18$	2	1
PpatSat1	<i>Patellifolia</i>	<i>P. procumbens</i>	$2n = 2x = 18$	4	2
		<i>P. patellaris</i>	$2n = 4x = 36$	2	1

#### 4. Discussion

Higher plant genomes contain a large fraction of repetitive sequences, these sequences are a major factor of variability in genome size and reflect the evolutionary divergence between species (Mehrotra and Goyal, 2014). Therefore, knowledge of repetitive DNAs in general and satellite DNAs in particular is valuable resources for genome studies.

In this study, the repetitive fraction of six genomes; including *B. vulgaris*, *B. lomatogona*, *B. nana*, *P. procumbens*, *P. patellaris*, and *C. quinoa*; was selected and analyzed using graph-based clustering of sequence reads. The proportion of repetitive DNA in these species was estimated at approximately 50% (34.4% – 65.6%), with LTR retrotransposons as the most abundant repeats. This is also observed in most higher plants, such as sorghum and maize (Lopez-Flores and Garrido-Ramos, 2012). Here, the focus of this thesis was the identification and annotation of the satellite families of the six species of two genera *Beta* and *Patellifolia*. Comparison of satellite abundance and localization patterns gives a unique insight into genome evolution.

Six novel satellite families were identified in the genome of *B. lomatogona*, a basal species of the section *Corollinae*. These families were investigated regarding their structure, abundance, distribution, and evolution. They were classified according to their distribution in the genome and thus assigned either to typical or to non-typical satellite families. Typical satellite families, including BISat1, BISat5 and BISat6, were defined as tandemly arranged in long arrays at specific heterochromatic loci, while non-typical satellite families, including BISat2, BISat3 and BISat4, were arranged in short array repetition in dispersed positions. The association of tandem repeats with transposable elements has been illustrated for both plants and animals (Meštrović *et al.*, 2015), and in this study the three non-typical satellite families are characterized as parts of Ogre/Tat LTR retrotransposons.

Two species *P. procumbens* and *P. patellaris* of the genus *Patellifolia* were selected for in-depth analysis because of their resistance to drought, soil salinity and beet cyst nematode *Heterodera schachtii* (Van Geyt *et al.*, 1990). As closely related to sugar beet, these species are a valuable genetic resource to improve the agronomic traits of sugar beet. However, the polyploid origin of *P. patellaris* is still under discussion although the question has been initially posed for over hundred years ago. Comparative *RepeatExplorer*-mediated read clustering was performed to identify the genetic difference between two genomes *P. procumbens* and *P. patellaris*. Two species-specific/enriched satellite families were identified, in which one is a *P. procumbens*-specific satellite and the other is enriched in *P. patellaris*. Their distribution

suggests that *P. patellaris* perhaps is allotetraploid species, with a half of its chromosome set derived from *P. procumbens*.

#### **4.1 Repetitive DNAs are a major component of plant genomes**

In this work, repeat clustering methodology was employed for six species, including species representing two genera, *Beta* and *Patellifolia*, as well as a closely related species of these genera, *C. quinoa*. In each species, the repeat proportion were counted from clusters which make up more than 0.01% of the genome. In addition, because of the high clustering threshold of 90%, the repetitive fraction detected by the graph-based clustering software does not include all repeats. Plant genomes may include evolutionarily old repeats which can be diverged (Heitkam *et al.*, 2014), truncated (Wenke *et al.*, 2009) or reshuffled (Wollrab *et al.*, 2012) and therefore difficult to identify. Thus, the repetitive fraction of analyzed species is likely larger. Compared with the small genome of *Arabidopsis thaliana* which has only 15% of repetitive DNA (Brandes *et al.*, 1997), this value in analyzed species is much higher. However, compared with other higher plants, such as poplar (42%) (Tuskan *et al.*, 2006), papaya (51.9%) (Tuskan *et al.*, 2006), apple (42.4%) (Velasco *et al.*, 2010), and African oil palm (57%) (Singh *et al.*, 2013), the repetitive fraction of six analyzed species is within a usual range.

Among repeated classes, LTR retrotransposons are the major repetitive DNA in all analyzed *Beta*, *Patellifolia*, and *Chenopodium* genomes. LTR retrotransposon fraction ranges from 18% to 35% of repeats in *B. vulgaris* and *P. procumbens*, respectively. This repeat class is also observed as the prominent fraction of higher plant genomes, such as it accounts for 29.95% - 38.3% in four species of *Nicotiana* section *Repandae* (Renny-Byfield *et al.*, 2013), 52.2% in *Camellia japonica* (Heitkam *et al.*, 2015), and even 67% in *Pinus lambertiana* (Steven *et al.*, 2016). There are examples indicating the increases in genome size have been driven by the amplification of LTR retrotransposon families. In maize, 50% of the genome constitute of retrotransposons, in which nearly 25% of the genome composes of five major classes of LTR retrotransposons (SanMiguel *et al.*, 1996). In *Vicia pannonica* a single family of Ty3-gypsy retrotransposons accounts for 38% of the genome and it is considered to increase the genome size by 50% (Neumann *et al.*, 2006). In the large genome of sugar pine, the genome expansion is also considered to be a result of the accumulation of Ty3-gypsy retrotransposons (Steven *et al.*, 2016). Therefore, the observed prominence of LTR retrotransposons in the fraction of highly repetitive sequences is a common feature of higher plant genomes where



retrotransposons represent one of the major forces driving genome size evolution (Neumann *et al.*, 2006; Tenaillon *et al.*, 2011; Nystedt *et al.*, 2013; Steven *et al.*, 2016).

Within LTR retrotransposons, the proportion of Ty1-*cop* and Ty3-*gypsy* retrotransposons do not differ greatly among the analyzed genomes, small differences among these species might result from differential amplification of repeat families present in the ancestor. The two non-LTR retrotransposon, including LINEs and SINEs, together constitute a small proportion of the six analyzed genomes (from 0.6% to 2%). This repeat fraction ranges from 0.05% to 0.65% of total repetitive DNA in the palm genomes (Ferreira Filho *et al.*, 2017). In Musaceae species, the proportion of LINEs is higher, accounting for approximately 0.5% to 1.2% of the genome (Novak *et al.*, 2014).

The other repeat class, DNA transposons, vary significantly among analyzed species (1.6% - 20.8%) and the variable proportion is not correlated with polyploidy level of genomes. Transposable proportion in a genome is not correlated with genome size, but the specific genome may be a factor for transposable element dynamics. For example, the differential transposable element dynamics between *Arabidopsis thaliana* and *Arabidopsis lyrata* indicated that genome size is the result of transposable element elimination (Hu *et al.*, 2011). Another example is *Phelipanche* and *Orobanche* species, despite having 1.3-3x larger genomes than *Orobanche* species, *Phelipanche* species have lower proportions of high-copy repetitive DNA (Piednoël *et al.*, 2012).

Plastid DNA is only present in *B. vulgaris*, this sequence may be either originated from contamination of the genomic DNA used for sequencing (Zakrzewski *et al.*, 2010) or integrated into nuclear DNA (Cullis *et al.*, 2009). The first detection of plastid DNA in nuclear DNA was in spinach (Scott *et al.*, 1984), here used as out group and related species to *Beta*. The proportion of organelle-derived DNA in *Zea mays* is 1.7% (Yuan *et al.*, 2003), and in *Sorghum bicolor* is 10% (Peterson *et al.*, 2002), which are lower as in *B. vulgaris*. With the high proportion of plastid DNA in *B. vulgaris* (11.1%), the latter scenario is very likely and has been reviewed.

Satellite DNA sequences comprise a large proportion of the repetitive DNA in genomes of the *Beta* and *Patellifolia* species, and *C. quinoa*, ranging from 6.6% in *B. lomatogona* (equivalent to 4.1% of the *B. lomatogona* genome) to 22.7% in *B. vulgaris* (equivalent to 10.8% of the *B. vulgaris* genome). These sequences account for 12.24% of the *Camellia japonica* genome (Heitkam *et al.*, 2015), 2.4% of the *Locusta migratoria* genome (Ruiz-Ruano *et al.*, 2016), and less than 1% of the Musaceae genomes (Novak *et al.*, 2014). This revealed that the fraction of satellite DNA can vary enormously, even in closely related species (Plohl *et al.*, 2012).

The high number of satellite families shared between species reflects the close relationship between those species. In particular, seven satellite families are shared between *B. lomatogona* and *B. nana* while only four satellite families are shared between *B. lomatogona* and *B. vulgaris*, which reflects *B. lomatogona* is closer relative to *B. nana* than to *B. vulgaris*. The close relationship between *Corollinae* and *Nanae* species was also reported when analysing tandem repeats (Gao *et al.*, 2000). *P. procumbens* and *P. patellaris* also share 13 of total 14 satellite families (Table 3.3), reflecting the very close relationship between these two species. This result is in agreement with the “satellite library” hypothesis, that related species may share a library of satellite DNA families and each species may have one or several predominant satellite families (Garrido-Ramos, 2017). For examples, the PSUB and PRAT satellite families are present in diverse taxa of Insecta, but these families are abundant in the beetle *Palorus subdepressus* and distributed in a low number of copies among the other species of Insecta (Mravinac *et al.*, 2005). Macas *et al.* (2007) identified a set of 15 satellite families in the *Pisum sativum* genome, in which the known satellite PisTR-B (Neumann *et al.*, 2001) constitutes 0.44% and the 13 remaining satellite families make up 1.15% of the genome proportion. There are up to 62 satellite families identified in *Locusta migratoria* (Ruiz-Ruano *et al.*, 2016), it has been known as the largest number of satellite families identified in a species.

The three sections of the genus *Beta* represented by three species share a set of satellite families, including pHC8, *Hinf*I satellite/Tantalos, BvSat4, and BvMSat8. The set of satellite families between *Patellifolia* species is definitely different, which includes pAp11, pTS3, pTS4, pTS5, and pTS100. This is an additional evidence supporting the separation of *Patellifolia* from the genus *Beta*. According to these set of satellite families from six species, the close relationship of *Beta* and *Patellifolia* species and the distal relationship of these two genera to *C. quinoa* are confirmed.

Although many satellite families have already been published in *Beta* as well as *Patellifolia* species, some of them are underrepresented in the *RepeatExplorer* results. This might be due to very divergent sequences that can not be detected, or their very low genome proportion (< 0.01%) (Piednoël *et al.*, 2012), and/or non-random read coverage of reference genomic sequence reads used as input for *RepeatExplorer* (Minoche *et al.*, 2011). For example, the pRn34 family, a variant of pAV34 in *B. vulgaris*, was identified in *B. nana* (Dechyeva *et al.*, 2006) but is not present in the *RepeatExplorer* output.

In six species, only genome size of *B. vulgaris* and *C. quinoa* were estimated at 750 Mb and 1390 Mb, respectively (Dohm *et al.*, 2014; Jarvis *et al.*, 2017). Therefore, based on the high

similarities between the three representative species of the genus *Beta*, the genome size of *B. lomatogona* and *B. nana* might be estimated relatively similar to *B. vulgaris* genome size.

#### **4.2 Abundance of satellite DNAs in *Beta* genomes**

Six novel identified satellite families in this study make up only 0.8% of total 6.6% satellite DNAs in the *B. lomatogona* genome, however, they are assigned either to typical or to non-typical satellite families because of different genomic organization in the genome. BISat1, BISat5, and BISat6 produced clear ladders in Southern hybridization as well as their localization in heterochromatin regions, which is an indication of typical satellite families. In contrary, BISat2, BISat3, and BISat4 are considered to be non-typical satellite families because of their dispersed chromosomal distribution.

The abundance of satellite families in the whole genome can vary significantly. In *RepeatExplorer* results of this study, the satellite family pEV is the most abundant in the *B. vulgaris* genome (4.2%) and the satellite pBC1447 accounts for the largest portion of satellite fraction in the *C. quinoa* genome (7.5%). In the *Vicia sativa* genome, the VicTR-B satellite DNA family presents approximately 25% (Macas *et al.*, 2000). The FriSAT1 satellite makes up to 36% of the *Fritillaria falcata* genome but only 0.1% of several genomes in the genus *Fritillaria* (Ambrožová *et al.*, 2011).

##### **4.2.1 Organization of the conventional satellite families BISat1, BISat5, and BISat6**

Satellite DNAs are characterized mainly by their nucleotide sequence, the length of the monomer sequence, AT content, the abundance, and the chromosomal localization. All these characteristics can vary widely from one satellite family to another (Plohl *et al.*, 2008, 2012, Mehrotra and Goyal, 2014).

Satellite DNA sequences vary significantly in sequence composition. The sequence variation of a satellite family is usually lower in one species and higher among different species as described in “concerted evolution” hypothesis (Dover and Tautz, 1986; Garrido-Ramos, 2015). The satellite DNA accumulates mutations over time, however, as consequence of non-coding DNA, these changes do not entail any serious effects. Therefore, satellite DNAs have been seen as the most dynamic repeat class. Together with point mutations, deletions and insertions also contribute to the sequence variation. The other feature of typical satellite DNAs is usually a high AT content (Macas *et al.*, 2002). However, the AT-rich is not a general rule for satellite

DNAs, because centromeric satellite DNAs investigated in hundreds of species do not appear to have a preference for AT- or GC-rich (Melters *et al.*, 2013).

BISat1 can be seen as an AT-rich satellite with AT content of 74.8%, and BISat5 exhibit a moderate AT content of 61.7%. Compared with satellite families pEV1 (AT content of 59%), pBV1 (AT content of 69%), and pAv34 (AT content of 62%) from *B. vulgaris*, these AT content are not significantly different. The high AT content results from deamination of 5-methylcytosine to thymine (Montero *et al.*, 1992; Hendrich *et al.*, 1999). However, the AT content of BISat6 is significantly decreased with an average AT content of 42.5%. Therefore, the monomer length and the GC/AT distribution are specific properties of the respective satellite family. The distribution of AT dinucleotides in monomers is important for nucleosome packing of satellites (Jiang *et al.*, 2003).

The three satellite families BISat1, BISat5, and BISat6 are considered as typical satellite DNA due to their monomer length as well as genomic and chromosomal distribution. The monomer length of all three satellite families falls into typical monomer size range of 150-180 bp or 300-360 bp (Heslop-Harrison, 2000). These particular lengths seem to be preferred for the structure of a single nucleosome and chromatin higher order structure (Schmidt and Heslop-Harrison, 1998; Jiang *et al.*, 2003; Plohl *et al.*, 2008; Zhang *et al.*, 2013; Melters *et al.*, 2013; Weiss-Schneeweiss *et al.*, 2015). In addition, BISat1 includes several repeat motifs in its sequence, which is not observed in the two remaining satellite families. It might be explained that BISat1 underwent a complex evolutionary process from shorter repetitions and the repeat motifs in the current sequence is remnants of events in the past (Garrido-Ramos, 2017).

### **The satellite family BISat1 is mainly amplified in the pericentromeric region**

BISat1 produced very clear ladders in Southern hybridization after a short exposure, which is considered as evidence of a tandem arrangement. This family is the most abundant satellite from the six identified satellite families in *B. lomatogona*, although its estimated genome proportion is only 0.3%. The abundance of BISat1 is also reflected in large arrays on four *B. lomatogona* chromosomes and smaller arrays on four other chromosomes. The signal positions are likely in the pericentromeric heterochromatin. Using chromosome-arm-specific BAC probes, the chromosomes carrying BISat1 arrays have been identified, which are chromosome 3, 5, 6, and 9. Usually, satellite DNA is found on all chromosomes, such as the pericentromeric satellite pEV in *B. vulgaris* (Schmidt *et al.*, 1991) and pRN1 in *B. nana* (Kubis *et al.*, 1997). The

variable array size indicated that BISat1 was amplified independently from each chromosome and its evolution may be still ongoing.

Comparative Southern hybridization revealed signals for BISat1 exclusively in the genus *Beta*. However, hybridization signal strength varies between tested species, decreasing in the following order: *B. lomatogona*, *B. corolliflora*, *B. nana*, *B. vulgaris*, *B. patula*. This indicates that BISat1 is amplified independently in each genome and it could spread to all chromosomes. Based on the presence of signals in comparative Southern hybridization experiment, FISH experiments were performed with *B. vulgaris* of section *Beta*, *B. nana* of section *Nanae*, and four species (*B. lomatogona*, *B. corolliflora*, *B. intermedia*, and *B. trigyna*) of section *Corollinae*. The extension to four species of section *Corollinae* is to investigate the BISat1 evolution in different polyploidy levels in this section. FISH signals appear on four *B. vulgaris* chromosomes at distal regions, which is completely different in chromosomal position from species of the two remaining sections. This again consolidated the closer relationship between two sections *Corollinae* and *Nanae*, compared to section *Beta*. Interestingly, the number of BISat1 signals as well as its chromosomal position was maintained in species of section *Corollinae* despite different polyploidy levels among these species. The divergence of BISat1 distribution among *Beta* species is discussed together with BISat5 below to have more information regarding polyploid origin of *B. corolliflora*.

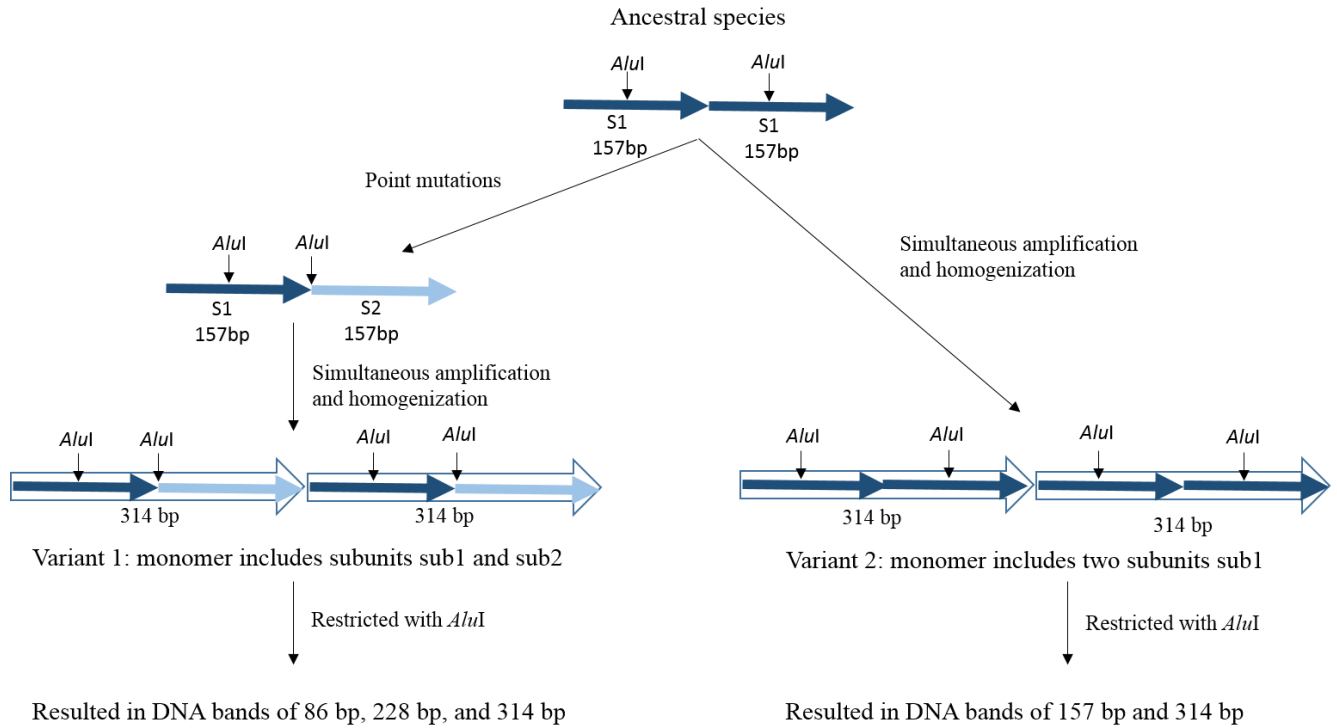
### **The pericentromeric satellite BISat5**

BISat5 localizes in the pericentromeric heterochromatin with differing array lengths on three *B. lomatogona* chromosome pairs. The largest BISat5 array localizes in the centromere of the chromosome 3, followed by arrays on chromosome 5 and chromosome 7. Although centromeric/pericentromeric satellite DNA is usually present on all chromosomes, the presence of pericentromeric satellite on only some of the chromosomes of a complement has been also described for the pTS5 satellite in *P. procumbens* (Gindullis *et al.*, 2001b).

The centromeric satellites are known as the most abundant satellite families in plant genomes (Ma *et al.*, 2007; Plohl *et al.*, 2014). For example, CentO centromeric satellite DNA of rice is composed of repeats of 155 bp, CentC centromeric satellite of maize have a length of 156bp (Mehrotra and Goyal, 2014; Plohl, 2014), or 327 bp centromeric satellite pBV1 of *B. vulgaris* (Schmidt *et al.*, 1991; Menzel *et al.*, 2008; Kowar *et al.*, 2016). Satellites may evolve to stabilize CENH3 nucleosomes, helping to prevent the loss of CENH3 nucleosomes against the pulling forces they undergo during chromosome segregation and to facilitate the formation of the

kinetochore (Garrido-Ramos, 2015). In addition, centromeric satellite repeats are usually hypomethylated (Zhang *et al.*, 2008) similar to pBV1 centromeric satellite in *B. vulgaris* (Zakrzewski *et al.*, 2013) and the 178 bp centromeric repeat in *A. thaliana* (Zhang *et al.*, 2008). In Southern hybridization, no ladder-like banding pattern was detected for B1Sat5 in *B. lomatogona* genomic DNA restricted with the methylation-sensitive enzymes *HpaII* and *MspI*, but the smear is started at approximately 1 kb. This indicates that there are many un-methylated cytosine in the B1Sat5 family.

The autoradiogram of hybridization of the B1Sat5 probe to *B. lomatogona* genomic DNA restricted with *AluI* (Figure 3.10B) revealed higher order structure of this family. The prominent B1Sat5 monomer is 314 bp in length and structured in two subunits of 157 bp, the subunit size probably corresponds to nucleosomal length. The higher order structure of B1Sat5 was observed in species of both sections *Corollinae* and *Nanae*. The Southern blot analysis also showed that all satellite members have a conserved length of approximately 314 bp corresponding to monomer size, however, three smaller bands of about 86 bp, 157 bp, and 228 bp are also detectable (Figure 3.15B). This indicated that there are at least two variants of the B1Sat5 family in the *Corollinae* and *Nanae* genomes. The appearance of two variants is supported by sequence analysis of clones from *B. lomatogona* (Figure 3.8), in which nine clones include complete monomers comprising of subunit sub1 and subunit sub2, only one clone comprises a monomer of two subunits sub1. Organization of centromeric satellite in higher order structure has also been reported for the alpha satellite in human (Rudd *et al.*, 2006; Sevim *et al.*, 2016).



**Figure 4.1: Schematic representation of a possible evolution of the satellite families BISat05**  
Subunit sub1 and sub2 were indicated by S1 (blue arrow) and S2 (light blue arrow), respectively.

Higher order structure of satellite DNAs probably results from simultaneous amplification and homogenization of two or more adjacent monomers (Plohl *et al.*, 2010). Therefore, the model of possible evolution of BISat5 is given (Figure 4.1). In ancestral species, satellite family with its unit length was 157 bp, but overtime point mutations were accumulated resulted in a new variant of 157 bp unit and different *AluI* recognized site. Both old and new units were simultaneously amplified forming a new satellite family (Variant 1 in Figure 4.1) with higher order structure. Together with variant 1, variant 2 including both subunit sub1 was also formed. The appearance of both variants in the genomes of *B. lomatogona*, *B. corolliflora*, and *B. nana* explained for the three small DNA bands with their length of 86 bp, 157 bp, and 228 bp. A satellite with higher order structure similar to BISat5 was described for the genus *Trifolium*, where the basic unit of the satellite TrR350 is 350 bp long and includes an internal direct repeat of 156 bp flanked by unrelated sequences (Ansari *et al.*, 2004), and for the 360 bp pAV34/pAC34 satellite family in the genus *Beta* with subunit length of 180 bp (Dechyeva *et al.*, 2008).

Comparative Southern hybridization revealed signals for BISat5 exclusively in section *Corollinae* and *Nanae*, whereas in *B. nana* the family is more abundant. This finding could be affirmed in comparative FISH of BISat5 in *Corollinae* and *Nanae* species, where BISat5 signals

were detected on six *B. lomatogona* chromosomes, 12 *B. corolliflora* chromosomes, and 16 *B. nana* chromosomes. In this case, it can be assumed that B1Sat5 has been evolved faster in the *B. nana* genome compared to that in *B. corolliflora* and *B. lomatogona*. The evolution of this satellite family might be ongoing.

### **The subtelomeric satellite B1Sat6**

B1Sat6 is only detected in the subtelomeric region of the chromosome 8 of *B. lomatogona*. The localization on unique chromosome of the B1Sat6 satellite makes this satellite family useful as a marker for the identifying *B. lomatogona* chromosome 8. Subtelomeric repeats, occurring on some or all chromosome arms, have been identified cytologically in many plants, such as maize (Li *et al.*, 2009), *B. vulgaris* (Dechyeva and Schmidt, 2006), and *Solanum tuberosum* (Torres *et al.*, 2011; Tang *et al.*, 2014), *Camellia japonica* (Heitkam *et al.*, 2015). Jain *et al.* (2010) proposed subtelomeric repeats as a supplemental part of chromosome end stability in the absence of canonical telomeric repeats. It is rare, that satellite repeats are chromosome-specific, like NUNSSP subtelomeric satellite from *Nicotiana undulata* (Lim *et al.*, 2005), and pBC216 from *B. corolliflora* (Gao *et al.*, 2000). Southern analysis with *HpaII/MspI* enzyme pair differing in methylation sensitivity did not show differences in cytosine methylation at CCGG sites in B1Sat1 and B1Sat5 sequences, but the difference can be observed in B1Sat6 sequence. A ladder pattern in genomic DNA restricted with *MspI* revealed more internal cytosine methylation at CCGG sites and this is a consequence of five CCGG sites in B1Sat6 sequence (Figure 3.4C). In plants, satellite sequences often present low level of cytosine methylation. For example, low level of cytosine methylation was detected in the pBV1 and pEV1 satellite families in sugar beet (Zakrzewski *et al.*, 2011), or reduced levels of cytosine methylation was observed in *Arabidopsis* (Zhang *et al.*, 2008), *jute* (Begum *et al.*, 2013), and *P. procumbens* (Schmidt *et al.*, 2014). The low cytosine methylation might result from the deamination of 5mC which is the most common nucleotide mutation and might lead to higher AT contents (Hendrich *et al.*, 1999).

### **Multi-color FISH reveals the *B. lomatogona* chromosome-specific distribution patterns of three typical satellites B1Sat1, B1Sat5, and B1Sat6**

Similar to sugar beet, *B. lomatogona* chromosomes are relatively small (about 2  $\mu\text{m}$  at condensed mitotic metaphase) and discrimination of chromosomes has not been possible accurately. A set of chromosome-arm-specific BACs selected to identify all nine linkage group



of sugar beet (Päsold *et al.*, 2012) was applied successfully to *B. lomatogona*. The mixture of four different labeled probes (the three satellite probes in combination with one chromosome-arm-specific BAC) was hybridized on metaphase chromosome spreads and revealed a unique distribution of satellite BISat1, BISat5, and BISat6 arrays (Figure 3.14).

The application of a BAC set with chromosome-specific repetitive sequences was also applied in *Medicago truncatula* (Kulikova *et al.*, 2001) and Sorghum (Kim *et al.*, 2002). Three satellite families BISat1, BISat5, and BISat6 localize in different chromosomal positions, and show the typical distribution for satellite DNAs, such as in pericentromeric and subtelomeric regions. This observation is in agreement with the content of the principle of the equilocal distribution of heterochromatin, that if a satellite family is located in one given chromosomal region, pericentromerically for example, it is arranged in differentially located arrays and this arrangement is maintained in all the chromosomes of the chromosome set or a particular subset of chromosome (Garrido-Ramos, 2017). This principle might explain the localization of more than one satellite family on the same chromosome as well as in one genome.

In multi-color FISH experiments, the three satellite probes were indirectly labeled and this method might reduce hybridization efficiency and make more background. Therefore, signals of satellite families might not exactly like that in FISH using direct labeling method. The application of BAC marker for the chromosome 1 North for *B. maritima*, *B. patula*, *B. corolliflora*, and *P. procumbens* revealed the conserved position on the homoeologous chromosome of all four wild beets (Päsold *et al.*, 2012). However, the result was limited to only chromosome 1. The karyotype of *B. lomatogona* is a result which shows the potential application of the BAC markers for other wild beets.

### **The distribution of BISat1, BISat5, and pBC1418 revealed the allotetraploid origin of *B. corolliflora***

Section *Corollinae* comprises two diploid species *B. lomatogona* and *B. macrorhiza*, two tetraploid species *B. corolliflora* and *B. intermedia*, and one pentaploid species *B. trigyna*. In these species, *B. intermedia* and *B. trigyna* are considered as hybrid species (Reamon-Büttner *et al.*, 1996), only the origin of *B. corolliflora* is still unclear. *B. corolliflora* is either an autotetraploid species resulting from *B. lomatogona* or *B. macrorhiza*, or an allotetraploid species resulting from the hybridization of *B. lomatogona* and *B. macrorhiza*. Hybridization of BISat1 and BISat5 probes on metaphase chromosomes of different *Corollinae* species was performed to trace the genome evolution within *Corollinae* section, especially in *B. corolliflora*.

BISat1 signals were detected on eight chromosomes of *B. lomatogona* as well as *B. corolliflora*, this might suggest that *B. lomatogona* is not the only species that participated in *B. corolliflora* polyploidization. This means *B. corolliflora* may be allotetraploid. In addition, PCR using BISat1-specific primers resulted in amplicons in *B. macrorhiza* indicating the occurrence of this satellite family in *B. macrorhiza*. From this evidence, a scenario can be happen, that the unknown species is *B. macrorhiza*, but after polyploidization BISat1 arrays in one of two parental species were degenerated, resulting in only eight *B. corolliflora* chromosomes carrying BISat1. Leitch and Bennett (2004) reported that a reduction of DNA amount relative to the respective diploid parents and satellite DNA sequences appeared to be at least to some degree responsible for this reduction.

BISat5 probe hybridized along *B. lomatogona* and *B. corolliflora* chromosomes showed six signals in *B. lomatogona* and ten signals in *B. corolliflora*. PCR amplification of BISat5 was also observed in three species, which are *B. lomatogona*, *B. corolliflora*, and *B. macrorhiza*. This FISH result support to autotetraploid origin of *B. corolliflora*. However, combined both FISH results of BISat1 and BISat5, a possible explanation for the origin of *B. corolliflora* is that this species is allopolyploidy as hybridization result of *B. lomatogona* and *B. macrorhiza*. In the case of BISat5, contrasted to BISat1, the BISat5 arrays from both parental species have equal contribution resulting in a double number of *B. corolliflora* chromosomes carrying BISat5.

Similar to BISat5, the pBC1418 satellite localized on eight *B. corolliflora* chromosomes (Gao *et al.*, 2000) but only on four *B. lomatogona* chromosomes (Figure 3.5A). The PCR result also confirmed occurrence of pBC1418 in *B. macrorhiza*.

Reamon-Büttner and Wricke (1993) proposed *B. corolliflora* as an autotetraploid species due to tetrasomic inheritance. Reamon-Büttner and her colleagues (1996) concluded that *B. corolliflora* is much more related to *B. macrorhiza* than *B. lomatogona*. Based on isozyme alleles analysis the author hypothesized that if *B. corolliflora* is an autotetraploid, *B. macrorhiza* is the progenitor. However, from the distribution of BISat1, BISat5, and pBC1418, it is more likely that *B. corolliflora* as an allotetraploid species. Analysis of all satellite families in related species is a powerful method to study the origin of polyploidy genomes.

Further investigation regarding the chromosomal distribution of BISat1, BISat5, and pBC1418 in *B. macrorhiza* should be performed, which is essential for understanding the contribution of *B. macrorhiza* in genome evolution of *B. corolliflora*.

#### 4.2.2 The satellite families B1Sat2, B1Sat3 and B1Sat4 as tandem repeat domains in Ogre/Tat retrotransposons

The three un-typical satellites B1Sat2, B1Sat3, and B1Sat4 are characterized by a moderate AT content (48.9% - 52.5%) and a monomer length that does not fall into the conventional range of 150-180 bp or 300-360 bp. The variation in nucleotide composition of these satellite repeats is mostly due to point mutations, but the identity among monomers is still high (on average of 88%). The chromosomal localization of the three tandem repeats B1Sat2, B1Sat3, and B1Sat4 differ from the three typical satellite families as these repeat clusters do not show localization at specific loci such as centromere, pericentromere, intercalary, or subtelomere. The satellite arrays are localized differently in variable number on all 18 chromosomes of *B. lomatogona*.

The B1Sat2 tandem repeat is present in distal regions of all 18 *B. lomatogona* chromosomes with additional signals in pericentromere of four chromosomes. The same chromosomal distribution was observed in B1Sat3, but only on 16 chromosomes, of which at least two chromosomes have additional signals near the centromere. This may reflect the appearance of both separated B1Sat2 and B1Sat3 arrays and short B1Sat2 and B1Sat3 as internal component of LTR retrotransposon elements like in the case of the PisTR-A tandem repeat (Neumann *et al.*, 2001). The B1Sat4 also localizes in a distal region of 14 *B. lomatogona* chromosomes, of which 12 chromosomes were detected with signals on one arm and two chromosomes with signals on both arms. Presumably, these satellite families localize mainly in the distal position of one chromosome arm, the weak signals indicate a small number of repeating units.

Bioinformatic analysis revealed that both tandem repeats B1Sat2 and B1Sat3 are a part of Ogre/Tat retrotransposons. A complete Ogre/Tat element was reconstructed from related clusters in *B. lomatogona*, in which B1Sat2 is present with three monomers. The contig analysis of the super-cluster including B1Sat3 tandem repeat also showed two additional satellite families *FokI* satellite/Dione and BvSat04 of *B. vulgaris* (Zakrzewski *et al.*, 2010). In addition, a study on the Ogre/Tat retrotransposon Hodor03 in *B. vulgaris* showed that *FokI* satellite/Dione and minisatellite BvMSat02 coexists in the element (Hoffmann, 2017). This supports the association of B1Sat3 with an Ogre/Tat retrotransposon. Although B1Sat4 has not shown a connection with any retrotransposon elements, the cluster graph as well as chromosomal distribution of this tandem repeats are similar to B1Sat2 and B1Sat3. In addition, the Southern hybridization pattern of B1Sat4 is similar to the pattern of pDvul2 dispersed repeat in *B. vulgaris* (Menzel *et al.*, 2008). It has been unclear whether pDvul2 is associated with a transposable element or not, but its genomic and chromosomal distribution is typical for some

Ty1-*copia* retrotransposons (Menzel *et al.*, 2008). Therefore, it is proposed that BISat4 is also likely associated with a LTR retrotransposon elements.

The dispersed distribution of the *FokI* satellite/Dione and BvMSat02 (Zakrzewski *et al.*, 2010,) was described without knowledge of physical linkage between these arrays and transposable elements. The appearance of short *FokI* satellite/Dione and BvMSat02 (Zakrzewski *et al.*, 2010) arrays in Hodor03 retrotransposon sequences of *B. vulgaris* (Hoffmann, 2017) might be a reason for their dispersed distribution. Together with a result from Bannack (2017) who demonstrated the association of four new *B. nana* satellite families with Ogre/Tat retrotransposons, tandem repeats as parts of LTR retrotransposons is now considered as the main reason of their dispersed distribution along *B. nana* chromosomes.

For further investigation towards the location of these three tandem repeats, it would be interesting to hybridize BISat2, BISat3, and BISat4 simultaneously on *B. lomatogona* metaphases and prophases. Relations between satellite positions could be an evidence for coexisting the three repeats in one retrotransposon element.

The association of tandem repeats with transposable elements has been illustrated for both animals and plants (Satovic *et al.*, 2016). It is particularly pronounced for Ty3-*gypsy* retrotransposons of plants in which 41.2% of all elements have internal tandem repeats (Macas *et al.*, 2009). It provides a library of short repeat arrays and that can be dispersed to various genomic loci via replication and transposition of the element itself, these repeats may then be subsequently expanded into separated satellite DNA at genomic loci where appropriate molecular mechanisms are active (Macas *et al.*, 2009).

A representative example of the association between tandem repeats and Ogre/Tat retrotransposons is the PisTR-A repeat (Neumann *et al.*, 2001). This repeat showed the ladder-like pattern in Southern hybridization like satellite DNA but occurred in a dispersed pattern in FISH experiments (Neumann *et al.*, 2001). The reason for the dispersed pattern was found eight years later: the PisTR-A satellite is associated with an Ogre/Tat element (Macas *et al.*, 2009). This repeat occurs both as an internal tandem repeat in the 3'-UTR of the Ogre/Tat retrotransposon and separately in long satellite arrays. The isolated satellite arrays resulted in strong signals at several genomic loci, whereas a dispersed organization was derived from the hybridization of the TE internal repeat variant (Macas *et al.*, 2009).

Other examples of internal tandem repeats into transposable elements are the subtelomeric Ty3-*gypsy* retrotransposons-Retand amplified in *Silene* species having up to 12 copies of a 67 bp

tandem repeat (Kejnovsky *et al.*, 2006), GM1 and GM2 from *Glycine max* or Hv1 from *Hordeum vulgare* also carry short tandem repeat arrays (Macas *et al.*, 2009). The Helitron-2 transposon of *Drosophila virilis* includes four copies of a 154 bp tandem repeat (Abdurashitov *et al.*, 2013). It is assumed that Ogre/Tat retrotransposons with internal tandem repeats and a common phenomenon due to the fact that Ogre/Tat retrotransposon elements are widespread in plant genomes and present a major repeat class in many species (Neumann *et al.*, 2006; Macas *et al.*, 2007; Zuccolo *et al.*, 2007). Furthermore, Ogre/Tat elements show complex structures with up to several different tandem repeats within one family.

With the improvement of sequencing techniques and bioinformatic tools, more repetitive sequences will be shed light on, in which the investigation of the complex structure of retrotransposon elements is an important objective.

#### **4.3 Comparative analysis of genome composition revealed the relationship between *Patellifolia procumbens* and *Patellifolia patellaris***

Species in the genus *Patellifolia* are important wild beets because of their resistance to the beet cyst nematode. Attempts to transfer beet cyst nematode resistance into cultivated beet were successfully (Jung and Wricke 1987; Van Geyt *et al.* 1988; Reamon-Ramos and Wricke 1992). However, in the tetraploid *P. patellaris*, it is unclear whether the genome doubling arose from *P. procumbens* (autopolyploidy) or from *P. procumbens* and an unknown species (allopolyploidy). Comparison of these two genomes using *RepeatExplorer* can provide useful information which contributes to the understanding of the polyploid origin of *P. patellaris*. *RepeatExplorer* is a useful tool for identification of repetitive DNA sequences as well as comparative analysis of multiple genomes (Novak *et al.*, 2014). Analysis of the first 500 clusters from the *RepeatExplorer* output revealed that two genomes share 98.8% similarity. 1.2% of the difference between these two genomes comprises two satellite families, in which PproSat1 is *P. procumbens*-specific and PpatSat1 is enriched in the *P. patellaris* genome. The use of combined data set can be more efficient in detecting low-copy repeat families which undetected in the individual genome screenings (Piednoël *et al.*, 2012), and in this case satellite family PpatSat1 was detected with very low genome proportion below 0.01%.

Both PCR and FISH experiments using the PproSat1 probe indicated that PproSat1 is a species-specific satellite of *P. procumbens*. This result seems contrasting because one haploid set of *P. patellaris* chromosomes is most likely derived from *P. procumbens*. Nevertheless, polyploidization is a complex process associated with either selective amplification or loss of

repetitive DNA (Parisod *et al.*, 2010), but genome downsizing is the general trend (Leitch and Bennett 2004; Leitch and Leitch 2008). A reduction of DNA amount relative to the respective diploid parents (Leitch & Bennett, 2004) lies within a range of 10 - 15%, with satellite DNA sequences appearing to be at least to some degree responsible for this reduction. Genome downsizing following polyploid formation is a widespread biological phenomenon, it was reported for many angiosperm species, such as *Brassica*, *Sedum*, *Ranunculus* (review by Leitch and Bennett, 2004), *Oryza sativa* (Ma *et al.*, 2004), *Nicotiana tabacum* (Renny-Byfield *et al.*, 2011), *Orobancha gracilis* (Piednoël *et al.*, 2012). However, not all satellite DNA types tend to change in response to allopolyploidy-induced 'genomic stress'. For example, stable inheritance of major centromeric and subtelomeric satellite DNAs was observed in early allotetraploids of *Tragopogon* (Pires *et al.*, 2004). Also in natural and synthetic tobacco plants some satellites, such as the subtelomeric HRS60 and centromeric NTS9 repeats, remained unchanged with respect to copy number (Hemleben, 2007). Repetitive DNA can be removed from the genome via homologous and illegitimate recombination (Fedoroff, 2012). In addition, a deletion at differential rates is thought to be the reason for genome size reduction. These deletions were often small, but numerous and common in non-coding and repetitive regions, including within transposable element (Hu *et al.*, 2011).

Genome reduction can explain the absence of the satellite family PproSat1 in the *P. patellaris* genome. It is unlikely that PproSat1 has been newly evolved in the *P. procumbens* genome, since the satellite PproSat1 occurs not only in *P. procumbens* but also in the other two species *B. corolliflora* and *C. quinoa*. Furthermore, the FISH signals of PproSat1 are only present in the distal region of one *P. procumbens* chromosome pair, this position may be easy to be lost by chromosomal breakage. However, in distal regions genes are amplified and concentrated, deletions of these regions may be lethal.

Genome reduction can explain the absence of the satellite family PproSat1 in the *P. patellaris* genome. Furthermore, the FISH signals of PproSat1 are only present in the distal region of one *P. procumbens* chromosome pair, this position may be easy to be lost by chromosomal breakage. It is unlikely that PproSat1 has been newly evolved in the *P. procumbens* genome, since the satellite PproSat1 occurs not only in *P. procumbens* but also in the other two species *B. corolliflora* and *C. Quinoa*. Another possible explanation for this phenomenon could be introgression of PproSat1 sequence into *P. procumbens* by interspecific crossings. The introgression of novel alleles into a species was reported for cotton species and this process has been applied successfully for cotton improvement (Chee *et al.*, 2016). However, up to date no

natural hybrid between *P. procumbens* and closely related species has been reported although many artificial interspecific hybridizations have been made for the purpose of introgressing disease resistance traits into the cultivated beet. Therefore, copy number reduction of PproSat1 sequence during polyploidization seems to be reasonable.

In contrast to PproSat1, PpatSat1 is not a species-specific satellite family, and its occurrence was observed in both *P. procumbens* and *P. patellaris*, respectively. However, PpatSat1 is much more abundant in *P. patellaris* compared with that in *P. procumbens*. This observation was also observed by comparative Southern hybridization and FISH. Interestingly, the chromosomal position of PpatSat1 still remains in the pericentromeric region after polyploidization. This again supports that *P. procumbens* is one haploid set of chromosomes of the *P. patellaris*, and PpatSat1 was amplified after polyploidization. The amplification of repetitive sequences after polyploidization was reported in *Orobancha gracilis* (Piednoël *et al.*, 2013). However, the effect on the particular repetitive family may also depend on the parental species (Vicent and Casacuberta, 2017), for example, the Sabine retrotransposon that is amplified in particular wheat polyploids, but is massively eliminated in others (Senerchia *et al.*, 2014).

The evolution of PproSat1 and PpatSat1 together might supply information regarding the polyploidization of *P. patellaris*. In the genus *Patellifolia*, PproSat1 is species-specific in *P. procumbens*, not present in *P. patellaris*. However, comparative Southern hybridization showed that PproSat1 also occurs in *B. corolliflora* and in the distally related species *C. quinoa*. Comparison of PproSat1 with previously known repeats revealed high identity of 92.5% and 95% to pBC1418 (from *B. corolliflora*, Gao *et al.*, 2000) and CqSat1 (from *C. quinoa*, Ost, master thesis, 2016), respectively. In comparative Southern hybridization experiment, no signal pattern of CqSat1 was observed in *B. lomatogona* and *P. patellaris* (Heitkam, unpublished). The distribution of one satellite family in a certain species of a section reflects a complex evolution of satellite sequence. Dramatic deletions of satellite DNAs might be a result of unequal crossover or large-scale changes (Ma and Jackson, 2006; Navrátilová *et al.*, 2008; Plohl *et al.*, 2012).

PpatSat1 is only present in species of the genus *Patellifolia* with variable abundance, this family might appear after the separation of the genera *Beta* and *Patellifolia*. Library hypothesis can be used to explain the different amplification of PpatSat1 in the genus *Patellifolia*.

Dechryeva and Schmidt (2009) indicated the differences in patterns of pTS5 and pRp34 satellite families and assumed that perhaps *P. patellaris* is allotetraploid. The signal pattern of the centromeric satellite family pTS5 was observed on 12 chromosomes of both species, *P.*

*procumbens* and *P. patellaris*. Together with pTS5, pRp34 signals were detected proximal to additional 18S rDNA signals in *P. procumbens*, but no visible signal on the *P. patellaris* chromosome end carrying the 18S rRNA genes. The authors presumed that a half of *P. patellaris* chromosome set is derived from *P. procumbens*.

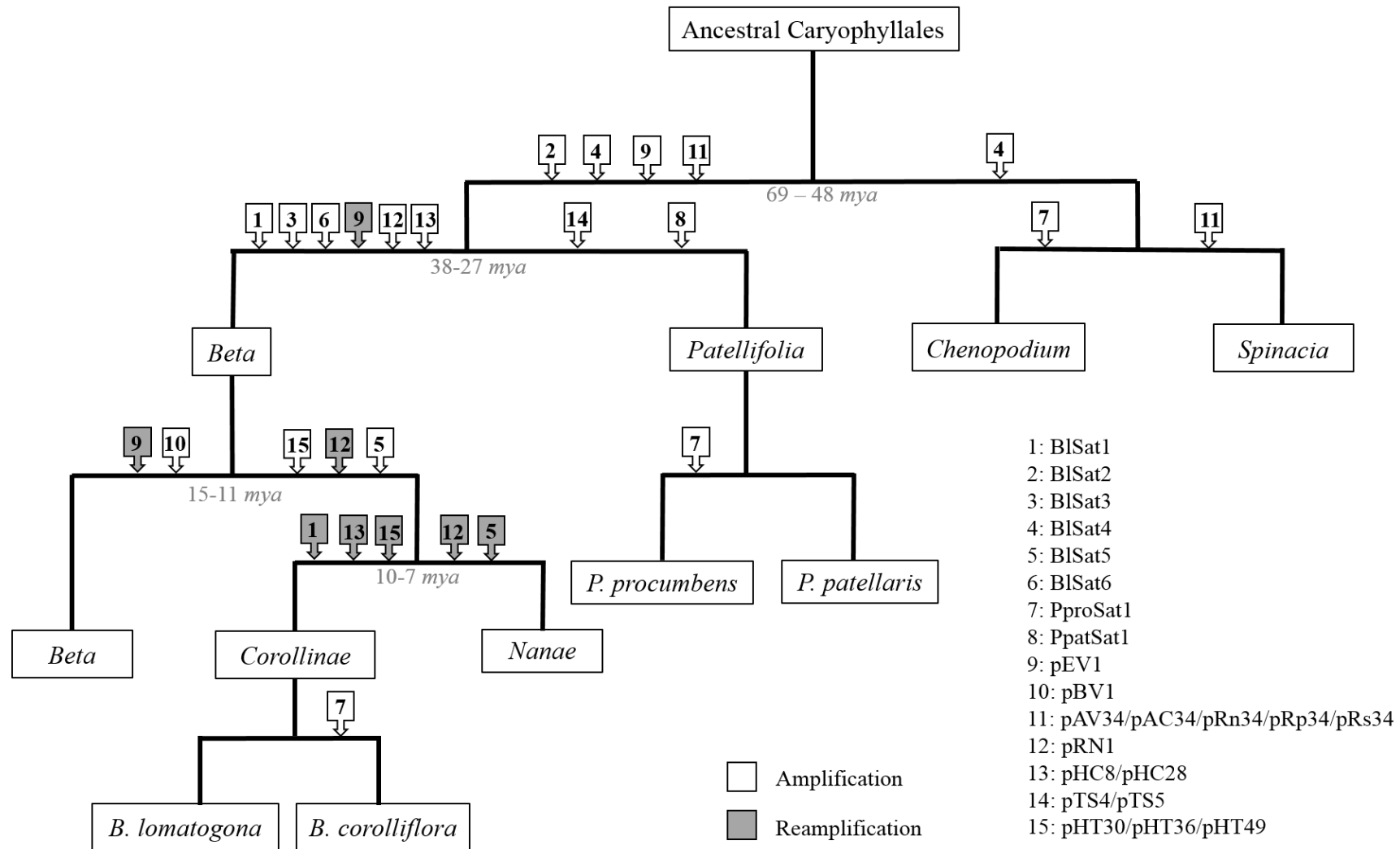
From these evidence, it might be assumed that *P. patellaris* is allotetraploid and *P. procumbens* is one of its parents. During the *P. patellaris* polyploidization, the PproSat1 satellite family was lost and the PpatSat1 has been accumulated. The other parent of *P. patellaris* has not been identified yet, therefore, further investigations regarding identification of specific repetitive families in *P. patellaris* would be interesting. The occurrence of *P. patellaris*-specific repeats in other species would be a useful hint for determining unknown parent.

The high similarity (98.8%) between the *P. procumbens* and *P. patellaris* genomes revealed that *P. patellaris* might be quite young species.

#### **4.4 Evolution of the satellite families in genus *Beta* and related species**

Satellite evolution may take place in both ways: on one hand satellite DNA families are able to quickly evolve by a change of copy number, on the other hand, they maintain the long-term stability of homogeneous arrays (Meštrović *et al.*, 2015). Using results from Southern hybridization, FISH, sequence analysis, and mapping of reference genome sequence reads against each satellite, it is possible to obtain knowledge about the evolution of the six *B. lomatogona* satellite families, two satellite families in *Patellifolia* species as well as the seven major known satellite families in the genus *Beta*. Figure 4.2 shows the suggested first appearance as well as possible reamplification of these satellite repeats. There are satellite families with distribution throughout the genera, besides, there are also section-specific or species-specific satellite families.





**Figure 4.2:** Dendrogram presents the evolution of novel identified satellites as well as the main known satellites in two subfamilies *Betoideae* (left) and *Chenopodiaceae* (right). The time scale is given by Hohmann *et al.* (2006).

BISat1 are present in all examined species of the genus *Beta* based on Southern hybridization signals. Mapping of sequence read data of reference species (Table 3.1) against BISat1 sequence produced hits in *Beta* species with the highest genome proportion belonging to *B. lomatogona*. Sequence analysis of BISat1 monomer in *Beta* species showed that this family is strongly conserved in *B. patula* (91.9% pairwise identity), followed by *B. vulgaris* (88.6% identity) and *B. nana* (85.5% identity). In *B. lomatogona*, it is not possible to draw any conclusion about sequence variability because only one monomer clone was obtained. However, the high identity of the unique clone sequence to consensus sequence as well as strong signals in Southern hybridization could indicate conserved monomers. The neighbor-joining analysis of BISat1 monomer sequences in *Beta* species does not reveal species-specific clusters, except for a cluster of BISat1 monomer sequences in *B. nana*. From these observations, it could be suggested that amplification of BISat1 happened after the radiation of the two genera *Beta* and *Patellifolia*, but high conservation throughout the genus *Beta*. This family was also reamplified in *B. lomatogona*. According to “concerted evolution” hypothesis (Dover and Tautz, 1986; Plohl *et al.*, 2010; Garrido-Ramos, 2015), BISat1 homogenization and fixation in each *Beta* species may be still going on. Furthermore, there were four weak signals of BISat1 detectable on *B. vulgaris* chromosomes, which is also consistent with Southern hybridization results. However, striking was the position of the BISat1 signals in *B. vulgaris*, which is not in the pericentromeric region as seen in *Corollinae* species but in the distal region of one chromosome arm. It is possible that the BISat1 satellite family changed its position in a species of the section *Beta* by chromosome arm inversion or other rearrangements. Evidence of arm inversions and translocations were reported when comparing the genomes of tomato to potato (Tanksley *et al.*, 1992), which might result in interspersed telomeric DNA. The presence of telomeric DNA arrays in intercalary and centromeric regions was also illustrated in *Pinus elliotii* (Schmidt *et al.*, 2000). The intercalary position of the pRs34 satellite family in *S. oleracea* was also considered as a consequence of chromosome arm inversion of the subtelomeric satellite present in the ancestral species (Dechyeva *et al.*, 2006).

As shown by comparative Southern hybridization, BISat5 only occurred in species of the sections *Corollinae* and *Nanae*, no signals were ascertained in section *Beta* as well as the genus *Patellifolia*. Mapping of the BISat5 sequence to reference sequence reads could partly confirm this occurrence, i.e. BISat5 makes up 0.3% of *B. nana* genome proportion while the proportion is 0.07% in *B. lomatogona*. The mapping results also showed very low abundance of BISat5 in *B. vulgaris* and *B. patula* (0.002% and 0.0025%, respectively) of the section *Beta*. This can be

explained by the fact that B1Sat5 is not in section *Beta* or only in very low, not detectable or mapping could be artefacts. B1Sat5 sequence variation among three species resulted in incomplete sorting by neighbor-joining clustering, in which all B1Sat5 monomers from different species are not grouped into species-specific clusters. It can be suggested that B1Sat5 appeared before the separation of sections *Corollinae* and *Nanae* but after differentiation of section *Beta* and the two sections *Corollinae* and *Nanae*. In particular, this family was significantly reamplified in the section *Nanae*. This could be explained by the rapid amplification of B1Sat5 subsequent to the division of *Corollinae* and *Nanae*. It also agrees with the phylogeny from Ford Lloyd (2005) that section *Nane* is more closely related to section *Corollinae* than to section *Beta*.

The differences of amplification of the same satellite families in closely related species could be explained by the satellite library hypothesis (Salser *et al.*, 1976; Ugarkovic and Plohl, 2002). This hypothesis proposes that a group of related species share a satellite DNA library. When a satellite DNA family is amplified differentially in one species, low-copy counterparts of it are found in other related species. The main mechanisms for the spreading of satellite DNA families may be “breakage and reunion” (Bedbrook *et al.* 1980), “slippage replication” (Levison and Gutman, 1987), unequal crossing-over (Smith, 1976; Schueler *et al.*, 2001), gene conversion (Dvorak *et al.*, 1987; Orel *et al.*, 2003), homologous recombination for the sequences containing direct repeats (Siebert and Puchta, 2002) and rolling-circle amplification (Cohen *et al.*, 2005).

Mapping of reference sequence reads against B1Sat6 sequence indicated presence of B1Sat6 in the genus *Beta*. This was confirmed by comparative Southern hybridization where signals were detected in *B. vulgaris*, *B. patula*, *B. lomatogona*, *B. corolliflora*, and *B. nana*. However, the abundance of B1Sat6 in *B. nana* differs between the mapping and the Southern hybridization, this may be due to sequence divergence of B1Sat6 in *B. nana* resulting in low hybridization efficiency. Analysis of B1Sat6 monomer sequences confirmed the divergence of this satellite in *Beta* species. Nevertheless, these differences are not species-specific resulting in a mixture of clusters in neighbor-joining analysis.

B1Sat2, B1Sat3, and B1Sat4 are known as short tandem repeats which are sequence domains of LTR retrotransposons. Mapping of reference sequence reads to the B1Sat2 monomer predicted an occurrence of the tandem repeat in *B. lomatogona* and *B. nana*. However, this repeat was amplified as multimers by PCR in *B. lomatogona*, *B. corolliflora*, *P. procumbens*, and *P. webbiana*. This indicated that B1Sat2 could occur in species of the genera *Beta* and *Patellifolia*, but in each species this repeat was either amplified in tandem arrays or changed into divergent

sequences. The BnSat03 tandem repeat isolated from *B. nana* (Bannack, 2017) and B1Sat2 showed a sequence identity of 70.8% although the BnSat03 monomer is shorter. The comparison of B1Sat2, BnSat03 and a 45 bp long tandem repeat which occurs in the Hodor7 retrotransposon in *B. vulgaris* (Hoffmann, 2017) showed a pairwise identity of 41.5%. This revealed the divergence of the ancestral sequence into different *Beta* and *Patellifolia* species. Monomer sequence variation of B1Sat2 was used for neighbor-joining analysis, resulting in species-specific clusters which reflects the separation of species *B. lomatogona*, *B. corolliflora* in one group and the other distantly related group including *P. procumbens* and *P. webbiana*. This is in line with current phylogeny (Hohmann *et al.*, 2006; Kadereit *et al.*, 2006; Thulin *et al.*, 2010).

Mapping of sequence reads of reference species (Table 3.1) against the B1Sat3 sequence showed the presence of this tandem repeat in all sections of the genus *Beta*, however the section *Beta* showed very low abundance (0.0006 – 0.0025%). This was also affirmed in PCR experiment where a ladder-like pattern of B1sat3 was observed in all examined species of the genus *Beta*. These results suggest that B1Sat3 appears after the separation of *Beta* and *Patellifolia*.

Comparative Southern hybridization indicated the occurrence of B1Sat4 in all examined species, even in distantly related species such as *C. quinoa* and *S. oleracea* (Figure 3.30). A similar result was also obtained using mapping of sequence reads of different species against B1Sat4 monomer sequence. Analysis of B1Sat4 monomer sequences in *B. vulgaris*, *B. patula*, *B. lomatogona*, *B. corolliflora*, and *B. nana* revealed that this sequence is not highly divergent within species. The variation of B1Sat4 sequences among examined species resulted in species-specific clusters in neighbor-joining tree, in which monomer sequences from each species were grouped together. This observation could be an example of concerted evolution which initially proposed by Dover and Tautz (1986). This hypothesis means that members of a satellite DNA family would show a high degree of intra-specific similarity and inter-specific divergence (Plohl *et al.*, 2010; Garrido-Ramos, 2015). The explanation of this concept is that non-reciprocal DNA exchange causes continual fluctuations in the sequences copy-number and, as a consequence, promotes the gradual and contiguous spread of a variant throughout a DNA family (homogenization) and throughout a population (fixation) as a dual process. B1Sat4 seems to be the oldest repeat among the six new identified satellite repeats and it is estimated to appear before differentiation of *Betoideae* and *Chenopodiaceae* subfamilies.

The other two satellite families in this thesis are PproSat1 and PpatSat1, which were identified in the genome of *P. procumbens* and *P. patellaris*, respectively. Comparative Southern

hybridization experiment indicated the occurrence of PproSat1 in *C. quinoa*, *B. corolliflora*, and dot blot confirmed presence of this satellite in *P. procumbens*. Mapping of short reads of different species against PproSat1 not only affirmed appearance of PproSat1 in *B. corolliflora*, *P. procumbens*, and *C. quinoa*, but also indicated very small proportion in *B. patula*. Although PproSat1 sequences extracted from sequence reads of *B. patula* showed high similarity with those in *C. quinoa* and *P. procumbens* (93.9% and 94.1%, respectively), no signal of PproSat1 was detected in *B. patula* by Southern experiment. This could be explained by the limited number of repeats in the *B. patula* genome. From these results, it could be concluded that PproSat1 is only amplified in specific species of distantly related genera, in particular in *B. corolliflora* species of *Beta* section *Corollinae*, *P. procumbens* species of the genus *Patellifolia*, and in *C. quinoa* species of the genus *Chenopodium*. The high similarity of PproSat1 sequences from distantly related species could be explained by the presence of this family already in an ancestral species early in the phylogeny, during the geographical separation this satellite sequence remains stable in nucleotide sequence. Such a high degree of conservation could point to a functional role of this satellite repeat, it is present on all chromosomes of *C. quinoa* (Heitkam, unpublished) but only present on one chromosome pair of *P. procumbens*.

PpatSat1 is genus-specific satellite family of the genus *Patellifolia*. This was confirmed by PCR, Southern hybridization, and FISH. The satellite PpatSat1 might have occurrence after the separation of the genera *Beta* and *Patellifolia*. The ladder-like pattern in both PCR and Southern hybridization of PpatSat1 was uniform in *Patellifolia* species, this might indicate the conserved monomer size of this family.

In order to have a broader view about satellite evolution in beet and related species, the major known satellite families, including pEV1, pBV1, pAV34/pAC34/pRn34/pRp34/pRs34, pRN1, pHC8/pHC28, pTS4/pTS5, pHT30/pHT36/pHT49, were taken into account. The intercalary satellite pEV1 (Schmidt *et al.*, 1991) is the most abundant satellite family in the *B. vulgaris* genome (68.7% of the satellite fraction in this species). This family is present in sections *Beta*, *Corollinae*, and *Nanae* of the genus *Beta*, and its subfamily pAp11 (Dechyeva *et al.*, 2003) occurs in species of the genus *Patellifolia*. Therefore, the pEV1 satellite family was most likely already present in ancestral species of the genus *Beta* and *Patellifolia*, during the species radiation this family was significantly diverged in the section *Corollinae* and *Nanae* but conserved in the section *Beta* and the genus *Patellifolia*. The pericentric/centromeric satellite family pBV1 was also identified in the *B. vulgaris* genome and this family is present in all species of the section *Beta*. It can be assumed that the pBV1 family is relatively young and its

amplification might occur after the separation of the section *Beta* and *Corollinae*. The subtelomeric satellite pAV34 was first described in *B. vulgaris* by Jansen *et al.* (1999) and then its subfamilies including pAC34 (from *B. corolliflora*), pRn34 (from *B. nana*), pRp34 (from *P. procumbens*), and pRs34 (from *S. oleracea*) were reported by Dechyeva *et al.* (2006, 2008). This family is distributed in both genera *Beta* and *Patellifolia* as well as in the genus *Spinacia*. The pRN1 satellite family is present in all species of the genus *Beta* (Kubis *et al.*, 1997). The two satellite families were identified in *B. corolliflora* are pHC8 (Gindullis *et al.*, 2001b) and pHC28 (Schmidt *et al.*, 1993). Both satellite families are distributed in species of the genus *Beta*, however, pHC28 is also present in species of the genus *Patellifolia* as a single DNA fragment (Schmidt *et al.*, 1993). The three satellites pHT30, pHT36, and pHT49 were firstly identified in the *B. trigyna* genome, these satellite families are present in species of the genus *Beta* (Schmidt *et al.*, 1993). Similar to pRN1, these repeats might be amplified before the radiation of *Beta* species, but after the separation of the genus *Beta* and *Patellifolia*. The genus-specific satellite families pTS4 and pTS5 were first described in *P. procumbens* (Schmidt *et al.*, 1991, 1996). These pericentric satellite families are present only in species of the genus *Patellifolia*, therefore, their appearance can be predicted after the separation of two genera *Beta* and *Patellifolia*. All these published satellite families are typical satellite DNA because they were identified based on restriction analysis and they might display important function in the genome, such as centromeric and subtelomeric formation and maintenance.

It is widely assumed that satellite sequences can evolve in a concerted manner to form species-specific satellite families, however, closely related species still share a set of satellite families in which one satellite family can be significantly amplified in specific species, but only few copies in other species. These are typically observed in tandem repeats in beet and related species.

## 5. Further work

Although the characterization of the satellite families in this thesis has been investigated, there are still some questions need to be answered.

The association between tandem repeats BISat2, BISa3, and BISat4 and Ogre retrotransposons needs to be further analyzed by using *B. lomatogona* sequence reads of good quality to allow the reconstruction of a complete Ogre/Tat retrotransposon elements with tandem repeat integration. The generation of LTR and RT probes from these Ogre elements and the investigation of their chromosomal distribution along *B. lomatogona* chromosomes, combined with the chromosomal distribution of the un-typical satellite families could shed light on the integration between the tandem repeats and the Ogre/Tat retrotransposons.

In order to have an insight into the contribution of the *B. macrorhiza* genome in the tetraploid genome of *B. corolliflora*, it would be interesting to investigate chromosomal distribution of the satellite families BISat1, BISat5, BISat6, and pBC1418 along *B. macrorhiza* chromosomes. Combined with their localization along *B. lomatogona* as well as *B. corolliflora* chromosomes, the tetraploid origin of *B. corolliflora* may be discovered.

Regarding the origin of *P. patellaris*, the other species-specific clusters from *P. procumbens* and *P. patellaris* need to be characterized not only in these two species but also in other available species including *P. webbiana*. Occurrence of these repeats in any species would be a useful hint for the investigation of allopolyploid origin of *P. patellaris*.

## 6. Summary

Genomes of higher plants comprise a large proportion of repetitive DNAs, where one major class is satellite DNA. Satellite DNA is organized in tandem arrays of basic repeating units, which often occurs in heterochromatin of centromeric/pericentromeric and intercalary as well as subtelomeric regions. Besides these typical satellite repeats, there are also non-typical satellite DNAs, which are organized in short tandem arrays and integrated into a transposable element. The chromosomal localization of non-typical satellites is not in large regions of heterochromatin, but tend to be dispersed along chromosomes. This thesis describes the identification of the major repeat classes including major satellite content in six beet and related species. The focus was on identification and characterization of new satellite families in the beet genomes.

In this study, the information regarding repetitive DNA as well as satellite families fraction in six beet and related species was gained based on graph-based clustering of next generation sequenced short sequence reads. The repeat proportion of the six analyzed species ranges from 34.4% in *C. quinoa* to 65.6% in *B. lomatogona*, in which the portion of nearly 50% belongs to *B. vulgaris*, *B. nana*, *P. procumbens*, and *P. patellaris*. Among all classes of repetitive DNAs, LTR retrotransposons are the most abundant repeat type in all analyzed genomes, which is a common feature of higher plant genomes. The other repeat sequences are DNA transposons, rDNA, and satellite DNA with variable portions in different species. A set of satellite families in each species was analyzed in detail, and reflects the relationship between six species. The closely related relationship between *B. lomatogona* and *B. nana* as well as between *P. procumbens* and *P. patellaris* is affirmed by seven and 13 satellite families shared between two species, respectively. Similarly, the closer relationship between *B. vulgaris* and two species *B. lomatogona* and *B. nana* than between *B. vulgaris* and two species *P. procumbens* and *P. patellaris* from the sister-genus *Patellifolia* is also confirmed. *C. quinoa* is a distantly related species and this is reflected by vastly different satellite content. Therefore, satellite DNA analysis might be a useful tool to trace species evolution.

In the *B. lomatogona* genome, by the application of *RepeatExplorer* tool, six novel tandemly repeated DNA sequences were identified and designated B1Sat1-B1Sat6. The three typical satellite families B1Sat1, B1Sat5, and B1Sat6 are organized in tandem arrays in large heterochromatic blocks. B1Sat1 is mainly localized in pericentric region of the chromosome 3, 5, 6, and 9, while B1Sat5 is amplified in pericentromeric region of the chromosome 3, 5, and 7. B1Sat6 is a chromosome-specific satellite and is located in the subtelomeric region on



the south arm of the chromosome 8. The other three satellite families B1Sat2, B1Sat3, and B1Sat4 are characterized as non-typical satellite DNA because of their dispersed distribution along chromosomes. B1Sat2 and B1Sat3 are identified as a tandem repeat domain in OGRE/Tat retrotransposons. The occurrence of one or several short tandem arrays in a transposable element is a common phenomenon in both animals and plants. These short repeats are considered to be continuously evolving and eventually amplifying to new satellite families.

Furthermore, the distribution of the six new satellite families in beet and related species was confirmed by comparative PCR, comparative Southern hybridization, and mapping of sequence reads from referent species against each satellite sequence. The B1Sat1 and B1Sat6 satellite families are specific for the genus *Beta*, while B1Sat5 is only amplified in two sections *Corollinae* and *Nanae* of the genus *Beta*. B1Sat4 is an ancient satellite family which exists in all tested species belonging to the genera *Beta*, *Patellifolia*, *Chenopodium*, and *Spinacia*, whereas B1Sat2 and B1Sat3 might have evolved before the separation of the genus *Beta* and *Patellifolia* but their sequences have been lost or heavily diverged during the species radiation.

Comparison of two wild beet genomes *P. procumbens* and *P. patellaris* was performed aiming to address the open question whether *P. patellaris* is auto- or allotetraploid. The high similarity between these two genomes indicates their close relationship. However, the genetic difference between two genomes, in particular the molecular characteristics as well as the chromosomal localization of two satellite families PproSat1 and PpatSat1, might support a hypothesis that *P. patellaris* is allotetraploid species with a half of its chromosome set derived from *P. procumbens*.

The results obtained in this work might provide comprehensive information of the repetitive classes as well as satellite families in the genomes of beets and related species. The results can be used as the species-specific and chromosome-specific markers in beet genome studies.

## 7. Literatures

Abdurashitov M. A., Gonchar D. A., Chernukhin V. A., Tomilova J. E., Schostak N. G., Zatsepina O. G., Zelentsova E. S., Evgen'ev M. B. (2013) 'Medium-sized tandem repeats represent an abundant component of the *Drosophila virilis* genome', *BMC Genomics*, 14(1).

Ambrožová K., Mandáková T., Bureš P., Neumann P., Leitch I. J., Koblížková A., Macas J., Lysak M. A. (2011) 'Diverse retrotransposon families and an AT-rich satellite DNA revealed in giant genomes of *Fritillaria lilies*', *Annals of Botany*, 107(2), pp. 255–268.

Amosova A. V., Bolsheva N. L., Zoshchuk S. A., Twardovska M. O., Yurkevich O. Y., Andreev I. O., Samatadze T. E., Badaeva E. D., Kunakh V. A., Muravenko O. V. (2017) 'Comparative molecular cytogenetic characterization of seven *Deschampsia* (Poaceae) species', *PLoS ONE*, 12(4), pp. 1–17.

Ansari H. A., Ellison N. W., Griffiths A. G., Williams W. M. (2004) 'A lineage-specific centromeric satellite sequence in the genus *Trifolium*', *Chromosome Research*, 12, pp. 357–367.

Badaeva E. D., Amosova A. V., Muravenko O. V., Samatadze T. E., Chikida N. N., Zelenin A. V., Friebe B., Gill B. S. (2002) 'Genome differentiation in Aegilops: Evolution of the D-genome cluster', *Plant Systematics and Evolution*, 231, pp 163–190.

Bedbrook J. R., Jones J., O'Dell M., Thompson R. D., Flavell R. B. (1980) 'A molecular description of telomeric heterochromatin in *Secale* species', *Cell*, 19, pp. 545-560.

Begum R., Alam S. S., Menzel G., Schmidt T. (2013) 'Comparative molecular cytogenetic analyses of a major tandemly repeated DNA family and retrotransposon sequences in cultivated jute *Corchorus* species (Malvaceae)', *Annals of Botany*, 112(1), pp. 123–134.

Bennetzen J. L., Ma J., Devos K. M. (2005) 'Mechanisms of recent genome size variation in flowering plants', *Annals of Botany*, 95(1), pp. 127–132.

Bramwell D., Bramwell S. (2001) 'Wild flowers of the Canary Islands', ed. 2. – Madrid.

Brandes A., Thompson H., Dean C., Heslop-Harrison J. S. (1997) 'Multiple repetitive DNA sequences in the paracentromeric regions of *Arabidopsis thaliana* L.', *Chromosome Research*, 5(4), pp. 238–246.

Braz G. T., He L., Zhao H., Zhang T., Semrau K. (2017) 'Comparative oligo-FISH mapping: an efficient and powerful methodology to reveal karyotypic and chromosomal evolution', *Genetics*, doi: 10.1534/genetics.117.300344/-/DC1.1.

Cafasso D., Chinali G. (2014) 'An ancient satellite DNA has maintained repetitive units of the original structure in most species of the living fossil plant genus *Zamia*', *Genome*, 57, pp125–135.

- Chee W. P., Paterson H. A., Udall A. J., Wenden F. J. (2016) 'Interspecific hybridization for Upland cotton improvement', In "Polyploidy and Hybridization for Crop Improvement" Ed. Mason A.
- Cheng Z., Dong F., Langdon T., Ouyang S., Buell C.R., Gu M., Blattner F.R., Jiang J. (2002) 'Functional rice centromeres are marked by a satellite repeat and a centromere-specific retrotransposon', *Plant Cell*, 14, 1691–1704.
- Cohen S., Agmon N., Yacobi K., Mislovati M., Segal D. (2005) 'Evidence for rolling circle replication of tandem genes in *Drosophila*'. *Nucleic Acids Research*, 33, pp 4519–4526.
- Cullis C. A., Vorster B. J., Van Der Vyver C., Kunert K. J. (2009) 'Transfer of genetic material between the chloroplast and nucleus: How is it related to stress in plants?', *Annals of Botany*, 103(4), pp. 625–633.
- Curtis G. J. (1968) 'Observations of fruit shape and other characters in the species of the section *Patellares*, genus *Beta*', *Euphytica*, 17(3), pp. 485–491.
- De Bock T. S. M. (1986) 'The genus *Beta*: domestication, taxonomy and interspecific hybridization for plant breeding', *Acta Hortic*, 182.
- De Capdeville G., Souza J., Manoel T., Szinay D., Diniz L. E. C., Wijnker E., Swennen R., Kema G. H. J., De Jong H. (2009) 'The potential of high-resolution BAC-FISH in banana breeding', *Euphytica*, 166(3), pp. 431–443.
- De Jong H. (1981) 'Investigation into chromosome morphology of sugar beet and related wild species', PhD thesis, University of Amsterdam.
- De Jong J. H., Fransz P., Zabel P. (1999) 'High resolution FISH in plants - techniques and applications', *Trends in Plant Science*, 4, pp258-263.
- Dechyeva D. and Schmidt T. (2006) 'Molecular organization of terminal repetitive DNA in *Beta* species', *Chromosome Research*, 14, pp.881-897.
- Dechyeva D., Schmidt T. (2009) 'Molecular cytogenetic mapping of chromosomal fragments and immunostaining of kinetochore proteins in *Beta*', *International Journal of Plant Genomics*, 2009. doi: 10.1155/2009/721091.
- Dechyeva D., Gindullis F., Schmidt, T. (2003) 'Divergence of satellite DNA and interspersions of dispersed repeats in the genome of the wild beet *Beta procumbens*', *Chromosome Research*, 11(1), pp. 3–21.
- Desel C., Jung C., Cai D., Kleine M., Schmidt T. (2001) 'High-resolution mapping of YACs and the single-copy gene *Hs1pro-1* on *Beta vulgaris* chromosomes by multi-colour fluorescence in situ hybridization', *Plant Molecular Biology*, 45(1), pp. 113–122.
- Devos K. M. (2002) 'Genome Size Reduction through Illegitimate Recombination Counteracts

Genome Expansion in *Arabidopsis*', *Genome Research*, 12(7), pp. 1075–1079.

Dohm J. C., Minoche A. E., Holtgräwe D., Capella-Gutiérrez S., Zakrzewski F., Tafer H., Rupp O., Sørensen T. R., Stracke R., Reinhardt R., Goesmann A., Kraft T., Schulz B., Stadler P. F., Schmidt T., Gabaldón T., Lehrach H., Weisshaar B., Himmelbauer H. (2014) 'The genome of the recently domesticated crop plant sugar beet (*Beta vulgaris*).', *Nature*, 505(7484), pp. 546–9.

Dong F., Song J., Naess S. K., Helgeson J. P., Gebhardt C., Jiang J. (2000) 'Development and applications of a set of chromosome-specific cytogenetic DNA markers in potato', *Theoretical and Applied Genetics*, 101, pp. 1001–1007.

Dover G. A. (1986) 'Molecular drive: a cohesive mode of species evolution', *Nature*, 299, pp. 111–117.

Dover G. A., Tautz D. (1986) 'Conservation and divergence in multigene families: alternatives to selection and drift', *Philosophical Transactions of the Royal Society B: Biological Science*, 312, pp. 275-289.

Dreissig S., Schiml S., Schindele P., Weiss O., Rutten T., Schubert V., Gladilin E., Mette M. F., Puchta H., Houben A. (2017) 'Live-cell CRISPR imaging in plants reveals dynamic telomere movements', *Plant Journal*, 91(4), pp. 565–573.

Dvorak J., Jue D., Lassner M. (1987) 'Homogenization of tandemly repeated nucleotide sequences by distance-dependent nucleotide sequence conversion', *Genetics*, 116, pp. 487- 498.

Ellis T. N., Lee D., Thomas C. M., Simpson P. R., Cleary W. G., Newman M. A., Burcham K. W. G. (1988) '5S rRNA genes in pisum: Sequence, long range and chromosomal organization', *Molecular and General Genetics*, 214, pp. 333–342.

Fajkus J., Kovarik A., Kralovics R., Bezdek M. (1995) 'Organization of telomeric and subtelomeric chromatin in the higher plant *Nicotiana tabacum*', *Molecular and General Genetics*, 247, pp. 633–638.

Fajkus J., Leitch A. R., Chester M., Sýkorová E. (2008) 'Evolution, composition and functions of telomeres and subtelomeres: lessons from plants', In: Nosek J, Tomáška L'u eds. Origin and Evolution of Telomeres, Neuveden, USA: *Landes Biosciences*, pp. 114–127.

Falstocco E. (2009) 'Presence of triploid cytotypes in the common fig (*Ficus carica* L.)', *Genome*, 52(11), pp. 919-25.

Fan Y. S., Davis L. M., Show T. B. (1990) 'Mapping small DNA sequences by fluorescence *in situ* hybridization directly on banded metaphase chromosomes', *Proceeding of the National Academy of Sciences of the United States of America*, 87(16), pp. 6223-6227.

- Feschotte C., Jiang N., Wessler S.R. (2002) 'Plant transposable elements: where genetics meets genomics', *Nature Rev. Genet.*, 3, pp. 329–341.
- Filho J. A. F., de Brito L. S., Leão A. P., Alves A. A., Formighieri E. F., Júnior M. T. S. (2017) 'In Silico Approach for Characterization and Comparison of Repeats in the Genomes of Oil and Date Palms.', *Bioinformatics and biology insights*, 11. doi: 10.1177/1177932217702388.
- Flavell R. B., Bennett M. D., Smith J. B., Smith, D. B. (1974) 'Genome size and the proportion of repeated nucleotide sequence DNA in plants', *Biochemical Genetics*, 12(4), pp. 257–269.
- Fleischmann A., Michael T. P., Rivadavia F., Sousa A., Wang W., Tensch E. M., Greilhuber J., Müller K. F., Heubl G. (2014) 'Evolution of genome size and chromosome number in the carnivorous plant genus *Genlisea* (*Lentibulariaceae*), with a new estimate of the minimum genome size in angiosperms', *Annals of Botany*, 114(8), pp. 1651–1663.
- Florijn R. J., van de Rijke F. M., Vrolijk H., Blonden L. A., Hofker M. H., den Dunnen J. T., Tanke H. J., van Ommen G. J., Raap A. K. (1996) 'Exon mapping by fiber-FISH or LR-PCR', *Genomics*, 38, pp. 277–282.
- Ford Lloyd (2005) *Genetics and breeding of sugar beet*.
- Ford-Lloyd B. V., Williams J. T. (1975) 'A revision of *Beta* section *Vulgares* (Chenopodiaceae), with new light on the origin of cultivated beets', *Botanical Journal of the Linnaean Society*, 71, pp. 89-102.
- Fransz P. F., Alonso-Blanco C., Liharska T., Peeters A. J. M., Zabel P., De Jong J. H. (1996) 'High resolution physical mapping in *Arabidopsis thaliana* and tomato by fluorescence in situ hybridisation to extended DNA fibres', *Plant Journal*, 9, pp. 421-430.
- Gao D., Schmidt T., Jung C. (2000) 'Molecular characterization and chromosomal distribution of species-specific repetitive DNA sequences from *Beta corolliflora*, a wild relative of sugar beet', *Genome*, 43(6), pp. 1073–1080.
- Garcia S., Kovařík A., Leitch A. R., Garnatje T. (2017) 'Cytogenetic features of rRNA genes across land plants: analysis of the Plant rDNA database', *Plant Journal*, 89(5), pp. 1020–1030.
- Garrido-Ramos M. A., de la Herran R., Ruiz Rejón M., Ruiz Rejón C. A. (1998) 'Satellite DNA of the Sparidae family (Pisces, Perciformes) associated with telomeric sequences', *Cytogenet. Cell Genet.* 83, pp. 3–9.
- Garrido-Ramos M. A. (2015) 'Satellite DNA in plants: More than just rubbish', *Cytogenetic and Genome Research*. doi: 10.1159/000437008.
- Garrido-Ramos M. A. (2017) 'Satellite DNA: An evolving topic', *Genes*, 8(9). doi: 10.3390/genes8090230.
- Gedamu Gebre Y., Bertolini E., Pè M. E., Zuccolo A. (2017) 'Identification and

characterization of abundant repetitive sequences in *Eragrostis tef* cv. Enatite genome', *BMC Plant Biology*, 16, pp. 39.

Gindullis F., Dechyeva D., Schmidt T. (2001a) 'Construction and characterization of a BAC library for the molecular dissection of a single wild beet centromere and sugar beet (*Beta vulgaris*) genome analysis', *Genome*, 44, pp. 846-855.

Gindullis F., Desel C., Galasso I., Schmidt T. (2001b) 'The large-scale organization of the centromeric region in *Beta* species', *Genome Research*, 11(2), pp. 253-265.

Graur D., Zheng Y., Azevedo R. B. R. (2015) 'An evolutionary classification of genomic function'. *Genome Biology and Evolution*, 7, pp. 642-645.

Green B. R., Gordon M. P. (1967) 'The satellite DNA's of some higher plants', *Biochim. Biophys. Acta - Nucleic Acids Protein Synth.*, 145, pp. 378-390.

Hall S. E., Kettler G., Preuss D. (2003) 'Centromere satellites from Arabidopsis populations: Maintenance of conserved and variable domains', *Genome Research*, 13, pp. 195-205.

Han Y., Zhang T., Thammaphichai P., Weng Y., Jiang J. (2015) 'Chromosome-specific painting in Cucumis species using bulked oligonucleotides', *Genetics*, 200(3), pp. 771-779.

Han Y., Zhang Z. H., Liu J. H., Lu J. Y., Huang S. W., Jin W. W. (2008) 'Distribution of the tandem repeat sequences and karyotyping in cucumber (*Cucumis sativus* L.) by fluorescence in situ hybridization', *Cytogenetic and Genome Research*, 122(1), pp. 80-88.

Hawkins J. S., Kim H., Nason J. D., Wing R. A., Wendel J. F. (2006) 'Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium Differential lineage-specific amplification of transposable elements is responsible for genome size variation in Gossypium', (515), pp. 1252-1261.

Heiskanen M., Peltonen L., Palotie A. (1996) 'Visual mapping by high resolution FISH', *Trends in Genetics*, 12 (10), pp. 379-382.

Heitkam T. and Schmidt, T. (2009) 'BNR - A line family from beta vulgaris- contains a rrm domain in open reading frame 1 and defines a 11 sub-clade present in diverse plant genomes', *Plant Journal*, 59(6), pp. 872-882.

Heitkam T., Holtgräwe D., Dohm J. C., Minoche A. E., Himmelbauer H., Weisshaar B., Schmidt T. (2014) 'Profiling of extensively diversified plant LINES reveals distinct plant-specific subclades', *Plant Journal*, 79(3), pp. 385-397.

Heitkam T., Petrasch S., Zakrzewski F., Kögler A., Wenke T., Wanke S., Schmidt T. (2015) 'Next-generation sequencing reveals differentially amplified tandem repeats as a major genome component of Northern Europe's oldest *Camellia japonica*', *Chromosome Research*. doi: 10.1007/s10577-015-9500-x.

- Hemleben V., Kovarik A., Torres-Ruiz R. A., Volkov R. A., Beridze T. (2007) 'The Natural History Museum Plant highly repeated satellite DNA : molecular evolution , distribution and use for identification of hybrids', *Systematics and Biodiversity*, 5 (3), pp. 277–289.
- Hemleben V., Zentgraf U. (1994) 'Structural organization and regulation of transcription by RNA polymerase I of plant nuclear ribosomal RNA genes.', *Plant Promoters and Transcription Factors*, 20, pp. 3–24.
- Hendrich B., Hardeland U., Ng H.H., Jiricny J., Bird A. (1999) 'The thymine glycosylase MBD4 can bind to the product of deamination at methylated CpG sites', *Nature*, 401, pp. 301–304.
- Heslop-Harrison J.S., Murata M., Ogura Y., Schwarzacher T., Motoyoshi F. (1999) 'Polymorphisms and genomic organization of repetitive DNA from centromeric regions of *Arabidopsis* chromosomes', *Plant Cell*, 11, pp. 31–42.
- Heslop-Harrison J. S. (2000) 'Comparative Genome Organization in Plants: From Sequence and Markers to Chromatin and Chromosomes', *The Plant Cell Online*, 12(5), pp. 617–636. d
- Hohmann S., Kadereit J. W., Kadereit G. (2006) 'Understanding Mediterranean-Californian disjunctions: Molecular evidence from Chenopodiaceae-Betoideae', *Taxon*, 55(1), pp. 67–78.
- Hu T. T., Pattyn P., Bakker E. G., Cao J., Cheng J. F., Clark R. M., Fahlgren N., Fawcett J. A., Grimwood J., Gundlach H. *et al.* (2011) 'The *Arabidopsis lyrata* genome sequence and the basis of rapid genome size change', *Nature Genet.*, 43(5), pp. 476–481.
- Jacobs G., Dechyeva D., Wenke T., Weber B., Schmidt T. (2009) 'A BAC library of *Beta vulgaris* L. for the targeted isolation of centromeric DNA and molecular cytogenetics of *Beta* species', *Genetics*, 135, pp.157–167.
- Jacobs, G., Dechyeva D., Menzel G., Dombrowski C., Schmidt T. (2004) 'Molecular characterization of Vulmar1, a complete mariner transposon of sugar beet and diversity of mariner- and En/Spm-like sequences in the genus *Beta*.', *Genome / National Research Council Canada = Genome / Conseil national de recherches Canada*, 47, pp. 1192–1201.
- Jacobs, G., Dechyeva D., Wenke T., Weber B., Schmidt T. (2009) 'A BAC library of *Beta vulgaris* L. for the targeted isolation of centromeric DNA and molecular cytogenetics of *Beta* species', *Genetica*, 135(2), pp. 157–167.
- Jain D., Hebden A. K., Nakamura T. M., Miller K. M., Cooper J. P. (2010) 'HAATI survivors replace canonical telomeres with blocks of generic heterochromatin', *Nature*, 467, pp. 223–227.
- Jarvis D. E., Ho Y. S., Lightfoot D. J., Schmöckel S. M., Li B. *et al.* (2017) 'The genome of *Chenopodium quinoa*', *Nature*, 542(7641), pp. 307–312. doi: 10.1038/nature21370.

- Jiang J., Gill B. S. (2006) 'Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research', *Genome*, 49, pp. 1057–1068.
- Jiang J., Birchler J. A., Parrott W. A., Dawe R. K. (2003) 'A molecular view of plant centromeres', *Trends in Plant Science*, 8(12), pp. 570–575.
- Jiang, J., Gill. B. S., Wang G. L., Ronald P., Ward D.C. (1995) 'Metaphase and interphase fluorescence in situ hybridization mapping of the rice genome with bacterial artificial chromosomes', *Proceeding of the National Academy of Sciences of the United States of America*, 92, pp. 4487–4491.
- Joly-Lopez, Z. and Bureau, T. E. (2014) 'Diversity and evolution of transposable elements in *Arabidopsis*', *Chromosome Research*, 22(2), pp. 203–216.
- Jung C., Wricke G. (1987) 'Selection of diploid nematode-resistant sugar beet from monosomic addition lines', *Plant Breeding*, 98, pp. 205-214.
- Kadereit G., Hohmann S., Kadereit J. W. (2006) 'A Synopsis of Chenopodiaceae Subfam . Betoideae and Notes on the Taxonomy of *Beta*', Source : Willdenowia , Bd . 36 , H . 1 , Special Issue : Festschrift Werner Greuter ( April 20), *Willdenowia - Annals of the Botanic Garden and Botanical Museum Berlin-Dahlem*, pp. 9–19.
- Kejnovsky E., Kubat Z., Macas J., Hobza R., Mracek J., Vyskot B. (2006) 'Retand: a novel family of gypsy-like retrotransposons harboring an amplified tandem repeat', *Molecular Genetics and Genomics*, 276, pp.254–263.
- Kelly L. J., Renny-Byfield S., Pellicer J., Macas J., Novák P., Neumann P., Lysak M. A., Day P. D., Berger M., Fay M. F., Nichols R. A., Leitch A. R., Leitch I. J. (2015) 'Analysis of the giant genomes of *Fritillaria* (*Liliaceae*) indicates that a lack of DNA removal characterizes extreme expansions in genome size', *New Phytologist*. doi: 10.1111/nph.13471.
- Kim J. S., Childs K. L., Islam-Faridi M. N., Menz M. A., Klein R. R., Klein P. E., Price H. J., Mullet J. E., Stelly D. M. (2002) 'Integrated karyotyping of sorghum by in situ hybridization of landed BACs', *Genome*, 45, pp. 402–412.
- Kolano, B., Gardunia B. W., Michalska M., Bonifacio A., Fairbanks D., Maughan P. J., Coleman C. E., Stevens M. R., Jellen E. N., Maluszynska J. (2011) 'Chromosomal localization of two novel repetitive sequences isolated from the *Chenopodium quinoa* Willd. genome.', *Genome / National Research Council Canada Genome / Conseil national de recherches Canada*, 54(9), pp. 710–717.
- Komuro S., Endo R., Shikata K., Kato A. (2013) 'Genomic and chromosomal distribution patterns of various repeated DNA sequences in wheat revealed by a fluorescence *in situ* hybridization procedure', *Genome*, 56: 131–137.
- Koornneef M., Fransz P., De Jong H., (2003) 'Cytogenetic tools for *Arabidopsis thaliana*', *Chromosome Research*, 11, pp. 183–194.



- Kowar, T., Zakrzewski F., Macas J., Kobližková A., Viehoveer P., Weisshaar B., Schmidt T. (2016) 'Repeat Composition of CenH3-chromatin and H3K9me2-marked heterochromatin in Sugar Beet (*Beta vulgaris*)'. *BMC plant biology*, 16, pp. 120.
- Kubis, S., Heslop-Harrison, J. S. and Schmidt, T. (1997) 'A family of differentially amplified repetitive DNA sequences in the genus *Beta* reveals genetic variation in *Beta vulgaris* subspecies and cultivars', *Journal of Molecular Evolution*, 44, pp. 310–320.
- Kubis, S., Schmidt, T. and Heslop-Harrison, J. S. (1998) 'Repetitive DNA elements as a major component of plant genomes', *Annals of Botany*, 82, pp. 45–55.
- Kulikova O., Gualtieri G., Geurts R., Kim D. J., Cook D., Huguet T., de Jong J. H., Fransz P. F., Bisseling T. (2001) 'Integration of the FISH pachytene and genetic maps of *Medicago truncatula*', *Plant Journal*, 27, pp. 49–58.
- Lapitan L. V., Brown S. E., Kennard W., Stephens J. L., Knudson D. L. (1997) 'FISH physical mapping with barley BAC clones', *Plant Journal*, 11, pp. 149–156.
- Leitch A. R., Leitch I. J. (2008) 'Genomic Plasticity and the Diversity of Polyploid Plants', *Science*, 25: Vol. 320, Issue 5875, pp. 481-483.
- Leitch I. J., Leitch A. R., Heslop-Harrison J. S. (1991) 'Physical mapping of plant DNA sequences by simultaneous *in situ* hybridization of two differently labelled fluorescent probes', *Genome*, 34, pp. 329-333.
- Leitch, I. J. and Bennett, M. D. (2004) 'Genome downsizing in polyploid plants', in *Biological Journal of the Linnean Society*. doi: 10.1111/j.1095-8312.2004.00349.x.
- Lengerova M., Kejnovsky E., Hobza R., Macas J., Grant S. R., Vyskot B. (2004) 'Multicolor FISH mapping of the dioecious model plant, *Silene latifolia*', *Theoretical and Applied Genetics*, 108, pp. 1193–1199.
- Lengerova M., Kejnovsky E., Hobza R., Macas J., Grant S. R., Vyskot B. (2004) 'Multicolor FISH mapping of the dioecious model plant, *Silene latifolia*', *Theoretical and Applied Genetics*, 108, pp. 1193-1199.
- Levison G., Gutman G. A. (1987) 'Slipped-strand mispairing: a major mechanism for DNA sequence evolution', *Mol Biol Evol*, 4, pp. 202-221.
- Lichter P. (1997) 'Multicolor FISHing: what's the catch', *Trends Genetics*, 13(12), pp. 475-479.
- Lichter P., Joos S. (1996) 'Non-isotopic *in situ* Hybridization to Metaphase Chromosomes and Interphase Nuclei', EMBO Practical Course at the German Cancer Research Center, Heidelberg.

- Lim K. Y., Matyasek R., Kovarik A., Fulnecek J., Leitch A. R. (2005) 'Molecular cytogenetics and tandem repeat sequence evolution in the allopolyploid *Nicotiana rustica* compared with diploid progenitors *N. paniculata* and *N. undulata*', *Cytogenet Genome Research*, 109, pp. 298-309.
- Lopez-Flores I., Garrido-Ramos M. A. (2012) 'The repetitive DNA content of eukaryotic genomes', *Genome Dynamics*, 7, pp. 1–28.
- Luchetti A. (2015) 'terMITEs: miniature inverted-repeat transposable elements (MITEs) in the termite genome (Blattodea: Termitoidea)', *Molecular Genetics and Genomics*, 290, pp.1499–1509.
- Lysak M. A., Fransz P. F., Ali H. B. M., Schubert I. (2001) 'Chromosome painting in *Arabidopsis thaliana*', *Plant Journal*, 28, pp. 689–697.
- Ma J. X., Devos K. M., Bennetzen J. L. (2004) 'Analyses of LTR- retrotransposon structures reveal recent and rapid genomic DNA loss in rice', *Genome Research*, 14, pp. 860–869.
- Ma J., Jackson S. A. (2006) 'Retrotransposon accumulation and satellite amplification mediated by segmental duplication facilitate centromere expansion in rice', *Genome Research*, 16, pp. 251–259.
- Ma J., Wing R. A., Bennetzen J. L., Jackson S. A. (2007) 'Plant centromere organization: A dynamic structure with conserved functions', *Trends Genetics*, 23, pp. 134–139.
- Macas J., Mészáros T., Nouzová M. (2002) 'PlantSat: a specialized database for plant satellite repeats', *Bioinformatics*, 18, pp. 28–35.
- 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', *Molecular and General Genetics*, 263(5), pp. 741–751.
- Macas J., Koblížková A., Navrátilová A., Neumann P. (2009) 'Hypervariable 3' UTR region of plant LTR-retrotransposons as a source of novel satellite repeats', *Gene*, 448(2), pp. 198–206.
- Macas J., Neumann P., and 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, pp. 1–16.
- Manuelidis L., Chen T. L. (1990) 'A unified model of eukaryotic chromosomes', *Cytometry*, 11(1), pp. 8–25.
- Martienssen R. A. (2003) 'Maintenance of heterochromatin by RNA interference of tandem repeats', *Nature Genet*, 35, pp. 213–214.
- Martínez P., Blasco M. A. (2011) 'Telomeric and extra-telomeric roles for telomerase and the telomere binding proteins', *Nature Rev Cancer*, 11, pp. 161–176.

- Martinsen L., Venanzetti F., Johnsen A., Sbordoni V., Bachmann L. (2009) 'Molecular evolution of the pDo500 satellite DNA family in Dolichopoda cave crickets (Rhaphidophoridae)', *BMC Evolutionary Biology*, 9, pp. 301.
- Mehrotra, S. and Goyal, V. (2014) 'Repetitive Sequences in Plant Nuclear DNA: Types, Distribution, Evolution and Function', *Genomics, Proteomics and Bioinformatics*. Beijing Institute of Genomics, Chinese Academy of Sciences and Genetics Society of China, 12(4), pp. 164–171.
- Melters, D. P., Bradnam K. R., Young H. A., Telis N., May M. R., Ruby J. G., Sebra R., Peluso P., Eid J., Rank D., Garcia J. F., Derisi J. L., Smith T., Tobias C., Ross-Ibarra J., Korf I., Chan S. W. (2013) 'Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution', *Genome Biology*, 14, p. R10.
- Menzel G., Dechyeva D., Keller H., Lange C., Himmelbauer H., Schmidt T. (2006) 'Mobilization and evolutionary history of miniature inverted-repeat transposable elements (MITEs) in *Beta vulgaris* L.', *Chromosome Research*, 14(8), pp. 831–844.
- Menzel G., Dechyeva D., Wenke T., Holtgräwe D., Weisshaar B., Schmidt T. (2008) 'Diversity of a complex centromeric satellite and molecular characterization of dispersed sequence families in sugar beet (*Beta vulgaris*)', *Annals of Botany*, 102(4), pp. 521–530.
- Mesbah (1997) 'Characterisation of alien chromosomes in monosomic additions of *Beta*', Wageningen Agricultural University.
- Meštrović N., Mravinac B., Pavlek M., Vojvoda-Zeljko T., Šatović E., Plohl M. (2015) 'Structural and functional liaisons between transposable elements and satellite DNAs', *Chromosome Research*, 23, pp. 583-596.
- Minoche A. E., Dohm J. C., Himmelbauer H. (2011) 'Evaluation of genomic high-throughput sequencing data generated on Illumina HiSeq and Genome Analyzer systems', *Genome Biology*, 12, pp. R112.
- Montero L. M., Filipski J., Gil P., Capel J., Martinez-Zapater J. M., Salinas J. (1992) 'The distribution of 5-methylcytosine in the nuclear genome of plants', *Nucleic Acids Research*, 20(12), pp. 3207-3210.
- Mravinac, B., Plohl, M. and Ugarković, D. (2005) 'Preservation and high sequence conservation of satellite DNAs suggest functional constraints', *Journal of Molecular Evolution*, 61(4), pp. 542–550.
- Nagaki K., Talbert P. B., Zhong C. X., Dawe R. K., Henikoff S., Jiang J. (2003) 'Chromatin immunoprecipitation reveals that the 180-bp satellite repeat is the key functional DNA element of *Arabidopsis thaliana* centromeres', *Genetics*, 163(3), pp. 1221–1225.

- Nagaki K., Cheng Z., Ouyang S., Talbert P. B., Kim M., Jones K. M., Henikoff S., Buell C. R., Jiang J. (2004) 'Sequencing of a rice centromere uncovers active genes', *Nature Genetics*, 36(2), pp. 138–145.
- Nagy, E. D. and Bennetzen, J. L. (2008) 'disease resistance gene cluster Pathogen corruption and site-directed recombination at a plant disease resistance gene cluster', *Genome Research*, pp. 1918–1923.
- Navrátilová A., Koblízková A., Macas J. (2008) 'Survey of extrachromosomal circular DNA derived from plant satellite repeats', *BMC Plant Biology*, 8, pp. 90.
- Neumann P., Nouzová M., Macas J. (2001) 'Molecular and cytogenetic analysis of repetitive DNA in pea (*Pisum sativum* L.)', *Genome*, 44(4), pp. 716–28.
- Neumann P., Koblízková A., Navrátilová A., Macas J. (2006) 'Significant expansion of *Vicia pannonica* genome size mediated by amplification of a single type of giant retroelement', *Genetics*, 173(2), pp. 1047–1056.
- Novak P., Neumann P., Macas J. (2010) 'Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data', *BMC Bioinformatics*, 11, pp. 378.
- Novak 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, pp. 792–793.
- Novák P., Hříbová E., Neumann P., Koblízková A., Doležel J., Macas J. (2014) 'Genome-wide analysis of repeat diversity across the family musaceae', *PLoS ONE*, 9(6).
- Nystedt B., Street N. R., Wetterbom A., Zuccolo A., Lin Y. C., Scofield D. G., Vezzi F., Delhomme N. *et al.* (2013) 'The Norway spruce genome sequence and conifer genome evolution', *Nature Publishing Group*, 497(7451), pp. 579–584.
- Ohmido N., Akiyama Y., Fukui K. (1998) 'Physical mapping of unique nucleotide sequences on identified rice chromosomes', *Plant Molecular Biology*, 38, pp. 1043–1052.
- Orel N., Kyryk A., Puchta H. (2003) 'Different pathways of homologous recombination are used for the repair of double-strand breaks within tandemly arranged sequences in the plant genome', *Plant Journal*, 35, pp. 604–612.
- Ozkan H., Tuna M., Arumuganathan K. (2003) 'Nonadditive changes in genome size during allopolyploidization in the wheat (*aegilops-triticum*) group', *Journal of Heredity*, 94, pp. 260–264.
- Päsold S., Borchardt D., Schmidt T., Dechyeva D. (2012) 'A sugar beet (*Beta vulgaris* L.) reference FISH karyotype for chromosome and chromosome-arm identification, integration of

genetic linkage groups and analysis of major repeat family distribution', *Plant Journal*, 72(4), pp. 600–611.

Palomeque T., Lorite P. (2008) 'Satellite DNA in insects: a review', *Heredity*, 100, pp. 564–573.

Parisod C., Holderegger R., Brochmann C. (2010) 'Evolutionary consequences of autopolyploidy', *New Phytologist*, 186, pp. 5 – 17.

Parisod C., Mhiri C., Lim K. Y., Clarkson J., Chase M. W., Leitch A. R., Grandbastie, M. A. (2012) 'Differential dynamics of transposable elements during long-term diploidization of *Nicotiana* section *Repandae* (*Solanaceae*) allopolyploid genomes', *PLoS ONE*, 7, e50352.

Pearce S. R., Harrison G., Li D., Heslop-Harrison J. S., Kumar A., Flavell A. J. (1996) 'The Ty1-copia group retrotransposons in *Vicia* species: Copy number, sequence heterogeneity and chromosomal localisation', *Molecular and General Genetics*, 250(3), pp. 305–315.

Pedersen C., Langridge P. (1997) 'Identification of the entire chromosome complement of bread wheat by two-colour FISH', *Genome*, 40, pp. 589–593.

Pellicer J., Fay M. F., Leitch I. J. (2010) 'The largest eukaryotic genome of them all?', *Botanical Journal*, 164, pp. 10–15.

Peterson D. G., Schulze S. R., Sciara E. B., Lee S. A., Bowers J. E., Nagel A., Jiang N., Tibbitts D. C., Wessler S. R., Paterson A. H. (2002) 'Integration of Cot Analysis , DNA Cloning , and High-Throughput Sequencing Facilitates Genome Characterization and Gene Discovery', *Genome Research*, 795, pp. 795–807.

Piednoël M., Aberer A. J., Schneeweiss G. M., Macas J., Novak P., Gundlach H., Temsch E. M., Renner S. S. (2012) 'Next-generation sequencing reveals the impact of repetitive dna across phylogenetically closely related genomes of orobanchaceae', *Molecular Biology and Evolution*, 29(11), pp. 3601–3611.

Piednoël M., Carrete-Vega G., Renner S. S. (2013) 'Characterization of the LTR retrotransposon repertoire of a plant clade of six diploid and one tetraploid species', *Plant Journal*, 75(4), pp. 699–709.

Pires C., Lim K., Kovarik A., Matyasek R., Boyd A., Leitch A., Leitch I., Bennett M., Soltis P., Soltis D. (2004) 'Genome size and distribution of tandem repetitive DNA in allopolyploid *Tragopogon* (*Asteraceae*)', *American Journal of Botany*, 91, pp. 1022-1035.

Pita S., Panzera F., Mora P., Vela J., Cuadrado A., Sánchez A., Palomeque T., Lorite P. 'Comparative repeatome analysis on *Triatoma infestans* Andean and Non-Andean lineages, main vector of Chagas disease', *PLoS ONE*, 12, e0181635.

- Plohl M., Luchetti A., Meštrović N., Mantovani B. (2008) 'Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin', *Gene*, 409, pp. 72–82.
- Plohl M., Meštrović N., Mravinac B. (2012) 'Satellite DNA evolution', in Garrido-Ramos M. A. (ed): Repetitive DNA, *Genome Dynamics*, 7: pp 126–152 (Karger, Basel 2012).
- Plohl M. (2010) 'Those mysterious sequences of satellite DNAs', *Periodicum Biologorum*.
- Plohl M., Luchetti A., Meštrović N., Mantovani B. (2008) 'Satellite DNAs between selfishness and functionality: Structure, genomics and evolution of tandem repeats in centromeric (hetero)chromatin', *Gene*, 409(1–2), pp. 72–82.
- Plohl M., Meštrović N., Mravinac B. (2010) 'Satellite DNA evolution.', *Genome Dynamics*, 7, pp. 126–52.
- Plohl M., Meštrović N., Mravinac, B. (2014) 'Centromere identity from the DNA point of view', *Chromosoma*, 123, pp. 313–325.
- Prince J. P., Pochard E., Tanksley S. D. (1993) 'Construction of a molecular linkage map of pepper and a comparison of synteny with tomato', *Genome*, 36, pp. 404–417.
- Prokopowich C. D., Gregory T. R., Crease T. J. (2003) 'The correlation between rDNA copy number and genome size in eukaryotes', *Genome*, 46, pp. 48– 50.
- Raap A. K., Florijn R. J., Blonden L. A. J., Wiegant J., Vaandrager J. W., Vrolijk H., den Dunnen J., Tanke H. J., van Ommen G. J. (1996) 'Fiber FISH as a DNA Mapping Tool', *Methods*, 9, pp. 67–73.
- Rayburn A. L., Gill B. S. (1985) 'Use of biotin-labeled probes to map specific DNA sequences on wheat chromosomes', *Journal of Heredity*, 76, pp. 78–81.
- Reamon-Büttner S. M., Wricke G., Frese L. (1996) 'Interspecific relationship and genetic diversity in wild beets in section *Corollinae* genus *Beta*: Isozyme and RAPD analyses', *Genetic Resources and Crop Evolution*, 43, pp. 261-274.
- Reamon-Ramos S. M., Wricke G. (1992) 'A full set of monosomic addition lines in *Beta vulgaris* from *Beta webbiana*: morphology and isozyme markers', *Theoretical and Applied Genetics*, 84, pp. 411–418.
- Renny-Byfield S., Chester M., Kovarik A., Le Comber S. C., Grandbastien M. A., Deloger M., Nichols R. A., Macas J., Novák P., Chase M. W., Leitch A. R. (2011) 'Next generation sequencing reveals genome downsizing in allotetraploid *Nicotiana tabacum*, predominantly through the elimination of paternally derived repetitive DNAs', *Molecular Biology and Evolution*, 28, pp. 2843–2854.

- Renny-Byfield S., Kovarik A., Kelly L. J., Macas J., Novak P., Chase M. W., Nichols R. A., Pancholi M. R., Grandbastien M. A., Leitch A. R. (2013) 'Diploidization and genome size change in allopolyploids is associated with differential dynamics of low- and high-copy sequences', *Plant Journal*, 74(5), pp. 829–839.
- Richards E. J., Ausubel F. M. (1988) 'Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*', *Cell*, 53, pp. 127–136.
- Rošić S., Köhler F., Erhardt S. (2014) 'Repetitive centromeric satellite RNA is essential for kinetochore formation and cell division', *Journal of Cell Biology*, 207(3), pp. 335–349.
- Rudd M. K., Wray G. A., Willard H. F. (2006) 'The evolutionary dynamics of alpha-satellite', *Genome Research*, 16, pp. 88–96.
- Ruiz-Ruano F. J., López-León M. D., Cabrero J., Pedro J., Camacho M (2016) 'High-throughput analysis of the satellitome illuminates satellite DNA evolution', *Scientific reports*, 6 (28333).
- Salser W., Bowen S., Browne D., El-Adli F., Fedoroff N., Fry K., Heindell H., Paddock G., Poon R., Wallace B., Whitcome P. (1976) 'Investigation of the organization of mammalian chromosomes at the DNA sequence level', *Federal proceedings*, 35, pp. 23-35.
- Sanmiguel P., Tikhonov A., Jin Y. K., Motchoulskaia N., Zakharov D., Al. Et (1996) 'Nested retrotransposon in the intergenic regions of the maize genome', *Science*, 274, pp. 765–768.
- Santoni S., Berville A. (1992) 'Characterization of the nuclear ribosomal DNA units and phylogeny of beta L.wild forms and cultivated beets', *Theoretical and Applied Genetics*, 83(9157), pp. 533–542.
- Satović E., Vojvoda Zeljko T., Luchetti A., Mantovani B., Plohl M. (2016) 'Adjacent sequences disclose potential for intra-genomic dispersal of satellite DNA repeats and suggest a complex network with transposable elements', *BMC Genomics*, 17(1), p. 997. doi: 10.1186/s12864-016-3347-1.
- Saunders V. A., Houben A. (2001) 'The pericentromeric heterochromatin of the grass *Zingeria biebersteiniana* (2n=4) is composed of Zbcen1-type tandem repeats that are inter-mingled with accumulated dispersedly organized sequences', *Genome*, 44, pp. 955–961.
- Schmidt A., Doudrick R. L., Heslop-Harrison J. S., Schmidt T. (2000) 'The contribution of short repeats of low sequence complexity to large conifer genomes', *Theoretical and Applied Genetics*, 101, pp. 7-14.
- Schmidt M., Hense S., Minoche A. E., Dohm J. C., Himmelbauer H., Schmidt T., Zakrzewski F. (2014) 'Cytosine methylation of an ancient satellite family in the wild beet *beta procumbens*', *Cytogenetic and Genome Research*, 143(1–3), pp. 157–167.

- Schmidt T., Kubis S., Heslop-Harrison J. S. (1995) 'Analysis and chromosomal localization of retrotransposons in sugar beet (*Beta Vulgaris L.*): LINEs and Ty1-copia-like elements as major components of the genome', *Chromosom Research*, 3, pp. 335–345.
- Schmidt T., Heslop-Harrison J. S. (1996) 'High-resolution mapping of repetitive DNA by in situ hybridization: molecular and chromosomal features of prominent dispersed and discretely localized DNA families from the wild beet species *Beta procumbens*', *Plant Molecular Biology*, 30(6), pp. 1099–1113.
- Schmidt T., Heslop-Harrison J. S. (1998) 'Genomes, genes and junk: The large-scale organization of plant chromosomes', *Trends in Plant Science*, 3(5), pp. 195–199.
- Schmidt T., Jung C., Metzclaff M. (1991) 'Distribution and evolution of two satellite DNAs in the genus *Beta*', *Theoretical and Applied Genetics*, 82(6), pp. 793–799.
- Schueler M. G., Higgins A. W., Rudd M. K., Gustashaw K., Willard H. F. (2001) 'Genomic and genetic definition of a functional human centromere', *Science*, 29, pp. 109–115.
- Schulte D., Cai D., Kleine M., Fan L., Wang S., Jung C. (2006) 'A complete physical map of a wild beet (*Beta procumbens*) translocation in sugar beet', *Molecular Genetics and Genomics*, 275(5), pp. 504–511.
- Schwarzacher T., Heslop-Harrison P. (2000) 'Practical *In Situ* Hybridization', Oxford: BIOS Scientific Publishers Ltd.
- Schwichtenberg K., Wenke T., Zakrzewski F., Seibt K. M., Minoche A., Dohm J. C., Weisshaar B., Himmelbauer H., Schmidt T. (2016) 'Diversification, evolution and methylation of short interspersed nuclear element families in sugar beet and related Amaranthaceae species', *Plant Journal*, 85(2), pp. 229–244.
- Senerchia N., Felber F., Parisod C. (2014), 'Contrasting evolutionary trajectories of multiple retrotransposons following independent allopolyploidy in wild wheats', *New Phytologist*, 202, pp. 975–985.
- Scott N. S., Timmis J. N. (1984) 'Homologies between nuclear and plastid DNA in spinach', *Theoretical and Applied Genetics*, 67(2–3), pp. 279–288.
- Sevim V., Bashir A., Chin C. S., Miga K. H. (2016) 'Alpha-CENTAURI: Assessing novel centromeric repeat sequence variation with long read sequencing', *Bioinformatics*, 32(13), pp. 1921–1924.
- Shaked H., Kashkush K., Ozkan H., Feldman M., Levy A. A. (2001) 'Sequence Elimination and Cytosine Methylation Are Rapid and Reproducible Responses of the Genome to Wide Hybridization and Allopolyploidy in Wheat', *The Plant Cell*, 13(8), p. 1749.



- Siebert R., Puchta H. (2002) 'Efficient repair of genomic double-strand breaks by homologous recombination between directly repeated sequences in the plant genome', *Plant Cell.*, 14, pp. 1121-1131.
- Singh R., Ong-Abdullah M., Low E. T., et al. (2013) 'Oil palm genome sequence reveals divergence of interfertile species in Old and New worlds', *Nature*, 500, pp. 335–339.
- Smith G. P. (1976) 'Evolution of repeated DNA sequences by unequal crossover', *Science*, 19, pp. 528–535.
- Steflova P., Tokan V., Vogel I., Lexa M., Macas J, et al (2013) 'Contrasting patterns of transposable element and satellite distribution on sex chromosomes (XY 1 Y 2) in the dioecious plant *Rumex acetosa*', *Genome Biology and Evolution*, 5, pp. 769–782.
- Stevens, K. A., Wegrzyn J. L., Zimin A., Puiu D., Crepeau M., Cardeno C., Paul R., Gonzalez-Ibeas D., Koriabine M., Holtz-Morris A. E., Martínez-García P. J., Sezen U. U., Marçais G., Jermstad K., McGuire P. E., Loopstra C. A., Davis J. M., Eckert A., de Jong P., Yorke J. A., Salzberg S. L., Neale D. B., Langley C. H. (2016) 'Sequence of the sugar pine megagenome', *Genetics*, 204(4), pp. 1613–1626.
- Suárez-Santiago V. N., Blanca G., Ruiz-Rejón M., Garrido-Ramos M. A. (2007) 'Satellite-DNA evolutionary patterns under a complex evolutionary scenario: the case of *Acrolophus* subgroup (*Centaurea L.*, *Compositae*) from the western Mediterranean', *Gene*, 404, pp. 80–92.
- Suzuki G., Ura A., Saito N., Do G. S., Seo B. B., Yamamoto M., Mukai Y. (2001) 'BAC FISH analysis in *Allium cepa*', *Genes and Genetic Systems*, 76, pp. 251–255.
- Szinay D., Chang S. B., Khrustaleva L., Peters S., Schijlen E., Bai Y., Stiekema W. J., Van Ham R. C. H. J., De Jong H., Klein Lankhorst R. M. (2008) 'High-resolution chromosome mapping of BACs using multi-colour FISH and pooled-BAC FISH as a backbone for sequencing tomato chromosome 6', *Plant Journal*, 56(4), pp. 627–637.
- Tang X., de Boer J. M., van Eck H. J., Bachem C., Visser R. G., de Jong H. (2008) 'Assignment of genetic linkage maps to diploid *Solanum tuberosum* pachytene chromosomes by BAC-FISH technology', *Chromosome Research*, 17, pp. 899–915.
- Tenaillon M. I., Hufford M. B., Gaut B. S., Ross-Ibarra J. (2011) 'Genome size and transposable element content as determined by high-throughput sequencing in maize and *Zea luxurians*', *Genome Biology and Evolution*, 3(1), pp. 219–229.
- Thomas C. A. (1971) 'The Genetic Organization of Chromosomes', *Annual Review of Genetics*, 5(1), pp. 237–256.
- Thulin M., Rydberg A., Thiede, J. (2010) 'Identity of *Tetragonia pentandra* and taxonomy and distribution of *Patellifolia* (Chenopodiaceae)', *Willdenowia - Annals of the Botanic Garden and Botanical Museum Berlin-Dahlem*, 40(2006), pp. 5–11.

- Torres G. A., Gong Z., Iovene M., Hirsch C. D., Buell C. R., *et al* (2011) 'Organization and evolution of subtelomeric satellite repeats in the potato genome', *G3 (Bethesda)* 1, pp. 85–92.
- Tran T. D., Cao H. X., Jovtchev G., Neumann P., Novák P., Fojtová M., Vu G. T.H., Macas J., Fajkus J., Schubert I., Fuchs J. (2015) 'Centromere and telomere sequence alterations reflect the rapid genome evolution within the carnivorous plant genus *Genlisea*', *Plant Journal*, 84(6), pp. 1087–1099.
- Tsoumani K. T., Drosopoulou E., Mavragani-Tsipidou P., Mathiopoulos K. D. (2013) 'Molecular characterization and chromosomal distribution of a species-specific transcribed centromeric satellite repeat from the olive fruit Fly, *Bactrocera oleae*', *PLoS ONE*, 8, pp. 1–11.
- Tuskan G. A., DiFazio S., Jansson S., *et al.* (2006) 'The genome of black cottonwood, *Populus trichocarpa* (Torr. & Gray)', *Science*, 313, pp. 1596–1604.
- Ugarković D., Plohl M. (2002) 'Variation in satellite DNA profiles – causes and effects', *EMBO Journal*, 21, pp. 5955–5959.
- Ugarkovic D. (2005) 'Functional elements residing within satellite DNAs', *EMBO reports*, 6(6), pp. 1035–1039.
- Van Geyt J. P. C., Lange W., Oleo M., De Bock T. S. M. (1990) 'Natural variation within the genus *Beta* and its possible use for breeding sugar beet: A review', *Euphytica*, 49, pp. 57-76.
- Van Geyt J. P. C., Oleo M., Lange W., and De Bock T.S.M. (1988) 'Monosomic additions in beet (*Beta vulgaris*) carrying extra chromosome of *Beta procumbens*, Identification of the alien chromosomes with the help of isozyme markers', *Theoretical and Applied Genetics*, 76, pp. 577-586.
- Velasco R., Zharkikh A., Affourtit J., *et al.* (2010) 'The genome of the domesticated apple (*Malus × domestica* Borkh.)', *Nature Genet.*, 42, pp. 833–839.
- Vicient C. M., Casacuberta, J. M. (2017) 'Impact of transposable elements on polyploid plant genomes', *Annals of Botany*, 120(2), pp. 195–207.
- Vitte C., Bennetzen J. L. (2006) 'Analysis of retrotransposon structural diversity uncovers properties and propensities in angiosperm genome evolution', *Proceedings of the National Academy of Sciences*, 103(47), pp. 17638–17643.
- Vittorazzi S. E., Lourenço, L. B., Recco-Pimentel, S. M. (2014) 'Long-time evolution and highly dynamic satellite DNA in leptodactylid and hylodid frogs', *BMC Genetics*, 15(1).
- Volkov R. A., Borisjuk N. V., Panchuk I. I., Schweizer D., Hemleben V. (1999) 'Elimination and rearrangement of parental rDNA in the allotetraploid *Nicotiana tabacum*', *Molecular Biology and Evolution*, 16(3), pp. 311–320.

- Weber B., Wenke T., Frömmel U., Schmidt T., Heitkam T. (2010) 'The Ty1-*copia* families SALIRE and Cotzilla populating the *Beta vulgaris* genome show remarkable differences in abundance, chromosomal distribution, and age', *Chromosome Research*, 18(2), pp. 247–263.
- Weber B., Heitkam T., Holtgräwe D., Weisshaar B., Minoche A. E., Dohm J. C., Himmelbauer H., Schmidt T. (2013) 'Highly diverse chromoviruses of *Beta vulgaris* are classified by chromodomains and chromosomal integration', *Mobile DNA*, 4(1).
- Weiss-schneeweiss H., Leitch A. R., Mccann J., Jang T., Macas J. (2015) 'Employing next generation sequencing to explore the repeat landscape of the plant genome', *Next-Generation Sequencing in Plant Systematics*, pp. 1–26.
- Wenke T., Holtgräwe D., Horn A. V., Weisshaar B., Schmidt T. (2009) 'An abundant and heavily truncated non-LTR retrotransposon (LINE) family in *Beta vulgaris*', *Plant Molecular Biology*, 71(6), pp. 585–597.
- Wollrab C., Heitkam T., Holtgräwe D., Weisshaar B., Minoche A. E., Dohm J. C., Himmelbauer H., Schmidt T. (2012) 'Evolutionary reshuffling in the Errantivirus lineage Elbe within the *Beta vulgaris* genome', *Plant Journal*, 72(4), pp. 636–651.
- Yuan Y., SanMiguel P. J., Bennetzen J. L. (2003) 'High-Cot sequence analysis of the maize genome', *Plant Journal*, 34(2), pp. 249–255.
- Zakrzewski F., Wenke T., Holtgräwe D., Weisshaar B., Schmidt T. (2010) 'Analysis of a c0t-1 library enables the targeted identification of minisatellite and satellite families in *Beta vulgaris*.', *BMC plant biology*, 10 (8), pp. 1–14.
- Zakrzewski F., Weisshaar B., Fuchs J., Bannack E., Minoche A. E., Dohm J. C., Himmelbauer H., Schmidt T. (2011) 'Epigenetic profiling of heterochromatic satellite DNA', *Chromosoma*, 120(4), pp. 409–422.
- Zakrzewski F., Schubert V., Viehoveer P., Minoche A. E., Dohm J. C., Himmelbauer H., Weisshaar B., Schmidt T. (2014) 'The CHH motif in sugar beet satellite DNA: A modulator for cytosine methylation', *Plant Journal*, 78(6), pp. 937–950.
- Zakrzewski F., Weber B., Schmidt T. (2013) 'A Molecular Cytogenetic Analysis of the Structure, Evolution, and Epigenetic Modifications of Major DNA Sequences in Centromeres of *Beta* Species', *Plant Centromere Biology*, pp. 39–55.
- Zakrzewski F., Schmidt M., Van Lijsebettens M., Schmidt T. (2017) 'DNA methylation of retrotransposons, DNA transposons and genes in sugar beet (*Beta vulgaris* L.)', *Plant Journal*, 90(6), pp. 1156–1175.
- Zhang T., Talbert P. B., Zhang W., Wu Y., Yang Z., Henikoff J. G., Henikoff S., Jiang J. (2013) 'The CentO satellite confers translational and rotational phasing on cenH3 nucleosomes in rice centromeres', *Proceeding of the National Academy of Sciences of the United States of America*,

110, pp. 4875–83.

Zhang W., Lee H. R., Koo D. H., Jiang J. (2008) ‘Epigenetic Modification of Centromeric Chromatin: Hypomethylation of DNA Sequences in the CENH3-Associated Chromatin in *Arabidopsis thaliana* and Maize’, *the Plant Cell Online*, 20(1), pp. 25–34.

Zuccolo A., Sebastian A., Talag J., Yu Y., Kim H. R., Collura K., Kudrna D., Wing R. A. (2007) ‘Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*’, *BMC Evolutionary Biology*, 7, pp. 1–15.

## Supplementary figures

Figure S1: Sequence divergence of B1Sat1 satellite monomers in *B. Lomatogona*

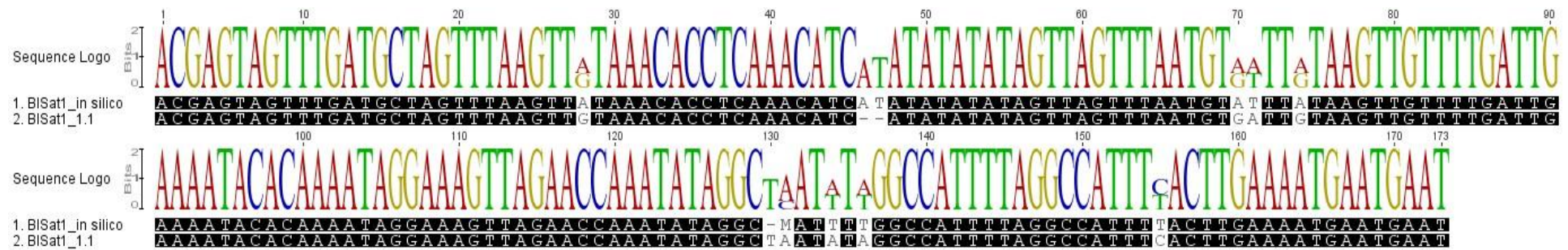
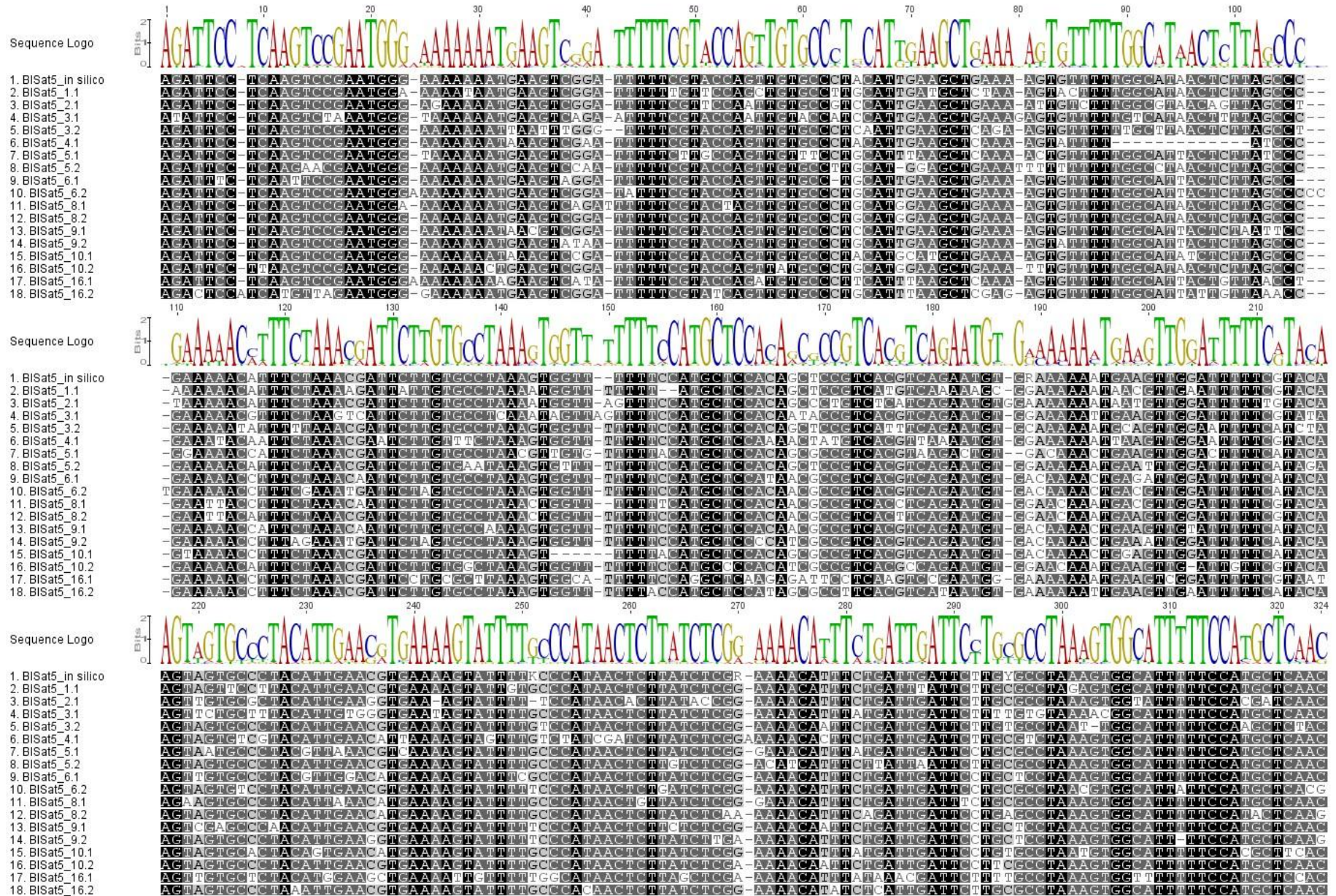


Figure S2: Alignment of BISat5 monomers in *B. Lomatogona* (A) and Distances between monomers (B)

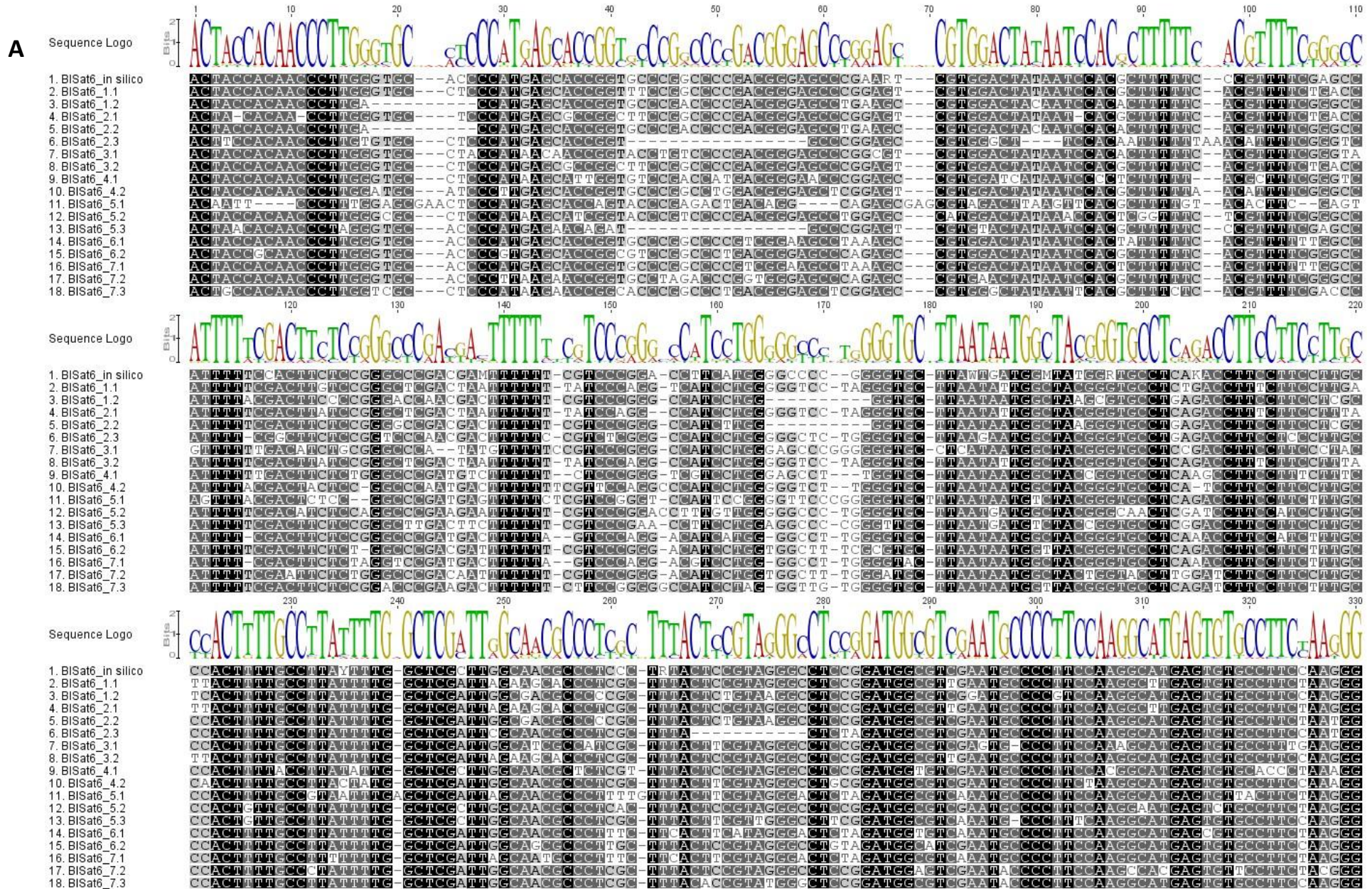
A



**B**

	BlSat5_in silico	BlSat5_1.1	BlSat5_2.1	BlSat5_3.1	BlSat5_3.2	BlSat5_4.1	BlSat5_5.1	BlSat5_5.2	BlSat5_6.1	BlSat5_6.2	BlSat5_8.1	BlSat5_8.2	BlSat5_9.1	BlSat5_9.2	BlSat5_10.1	BlSat5_10.2	BlSat5_16.1	BlSat5_16.2
BlSat5_in silico		89.5%	88.4%	86.4%	90.3%	85.1%	88.5%	91.4%	91.4%	91.5%	91.7%	94.1%	91.4%	91.4%	90.7%	93.3%	83.8%	89.5%
BlSat5_1.1	89.5%		82.2%	79.4%	82.8%	79.0%	81.8%	84.8%	83.8%	84.0%	85.1%	86.0%	82.8%	83.4%	81.8%	85.4%	75.9%	83.2%
BlSat5_2.1	88.4%	82.2%		80.4%	82.5%	77.2%	79.4%	82.6%	83.5%	82.4%	82.6%	84.4%	82.5%	82.2%	81.3%	83.8%	77.2%	80.7%
BlSat5_3.1	86.4%	79.4%	80.4%		80.1%	75.1%	79.7%	80.7%	82.3%	81.3%	82.0%	83.2%	81.3%	81.3%	80.7%	82.3%	76.3%	79.5%
BlSat5_3.2	90.3%	82.8%	82.5%	80.1%		80.0%	82.2%	84.6%	84.7%	84.3%	83.5%	86.6%	85.0%	84.1%	84.7%	85.7%	78.1%	82.9%
BlSat5_4.1	85.1%	79.0%	77.2%	75.1%	80.0%		78.7%	80.1%	80.0%	79.6%	80.1%	82.5%	81.3%	79.4%	78.1%	81.0%	72.8%	78.8%
BlSat5_5.1	88.5%	81.8%	79.4%	79.7%	82.2%	78.7%		83.5%	86.6%	85.8%	85.7%	86.0%	86.6%	85.4%	85.7%	85.7%	77.1%	85.1%
BlSat5_5.2	91.4%	84.8%	82.6%	80.7%	84.8%	80.1%	83.5%		86.3%	85.6%	85.8%	87.9%	85.4%	86.7%	84.8%	88.9%	77.8%	84.5%
BlSat5_6.1	91.4%	83.8%	83.5%	82.3%	84.7%	80.0%	86.6%	86.3%		90.6%	88.9%	89.8%	91.4%	91.4%	87.6%	88.2%	79.4%	87.0%
BlSat5_6.2	91.5%	84.0%	82.4%	81.3%	84.3%	79.6%	85.8%	85.6%	90.6%		87.8%	89.6%	90.3%	92.8%	87.1%	88.1%	78.9%	86.2%
BlSat5_8.1	91.7%	85.1%	82.6%	82.0%	83.5%	80.1%	85.7%	85.8%	88.9%	87.8%		93.3%	87.3%	86.7%	86.9%	89.5%	77.5%	84.2%
BlSat5_8.2	94.1%	86.0%	84.4%	83.2%	86.6%	82.5%	86.0%	87.9%	89.8%	89.6%	93.3%		88.5%	89.5%	87.9%	92.7%	79.7%	86.0%
BlSat5_9.1	91.4%	82.8%	82.5%	81.3%	85.0%	81.3%	86.6%	85.4%	91.4%	90.3%	87.3%	88.5%		90.1%	87.3%	88.5%	78.7%	86.7%
BlSat5_9.2	91.4%	83.4%	82.2%	81.3%	84.1%	79.4%	85.4%	86.7%	91.4%	92.8%	86.7%	89.5%	90.1%		86.6%	88.5%	79.7%	85.7%
BlSat5_10.1	90.7%	81.8%	81.3%	80.7%	84.7%	78.1%	85.7%	84.8%	87.6%	87.1%	86.9%	87.9%	87.3%	86.6%		86.3%	79.0%	83.8%
BlSat5_10.2	93.3%	85.4%	83.8%	82.3%	85.7%	81.0%	85.7%	88.9%	88.2%	88.1%	89.5%	92.7%	88.5%	88.5%	86.3%		79.0%	85.7%
BlSat5_16.1	83.8%	75.9%	77.2%	76.3%	78.1%	72.8%	77.1%	77.8%	79.4%	78.9%	77.5%	79.7%	78.7%	79.7%	79.0%	79.0%		79.4%
BlSat5_16.2	89.5%	83.2%	80.7%	79.5%	82.9%	78.8%	85.1%	84.5%	87.0%	86.2%	84.2%	86.0%	86.7%	85.7%	83.8%	85.7%	79.4%	

Figure S3: Alignment of BISat6 monomers in *B. Lomatogona* (A) and Distances between monomers (B)





**B**

	BlSat6_in silico	BlSat6_1.1	BlSat6_1.2	BlSat6_2.1	BlSat6_2.2	BlSat6_2.3	BlSat6_3.1	BlSat6_3.2	BlSat6_4.1	BlSat6_4.2	BlSat6_5.1	BlSat6_5.2	BlSat6_5.3	BlSat6_6.1	BlSat6_6.2	BlSat6_7.1	BlSat6_7.2	BlSat6_7.3
BlSat6_in silico		85.6%	84.3%	84.3%	85.9%	75.9%	82.4%	85.9%	81.0%	86.1%	70.8%	86.1%	83.7%	85.4%	87.2%	84.5%	84.2%	84.5%
BlSat6_1.1	85.6%		80.7%	95.9%	82.0%	74.5%	81.8%	97.5%	79.7%	84.0%	70.6%	81.7%	78.2%	82.6%	85.4%	82.6%	82.3%	83.0%
BlSat6_1.2	84.3%	80.7%		79.4%	96.3%	74.8%	78.6%	81.0%	77.1%	82.0%	67.6%	78.5%	74.1%	80.6%	82.3%	80.3%	79.4%	81.0%
BlSat6_2.1	84.3%	95.9%	79.4%		80.6%	72.6%	79.9%	97.2%	77.8%	81.8%	70.0%	79.8%	76.9%	80.1%	84.2%	80.1%	80.1%	82.3%
BlSat6_2.2	85.9%	82.0%	96.3%	80.6%		75.8%	79.6%	81.6%	79.3%	82.3%	68.2%	81.1%	75.3%	82.5%	83.5%	81.9%	81.3%	82.3%
BlSat6_2.3	75.9%	74.5%	74.8%	72.6%	75.8%		73.4%	74.8%	71.1%	75.9%	66.5%	72.1%	76.7%	76.0%	77.0%	74.8%	74.2%	76.2%
BlSat6_3.1	82.4%	81.8%	78.6%	79.9%	79.6%	73.4%		81.8%	78.0%	80.9%	67.9%	80.6%	76.7%	78.3%	80.8%	77.7%	80.5%	80.3%
BlSat6_3.2	85.9%	97.5%	81.0%	97.2%	81.6%	74.8%	81.8%		79.4%	84.3%	70.3%	81.1%	78.2%	82.0%	86.1%	82.0%	81.3%	83.9%
BlSat6_4.1	81.0%	79.7%	77.1%	77.8%	79.3%	71.1%	78.0%	79.4%		80.4%	67.0%	79.5%	74.1%	79.4%	81.0%	79.4%	78.5%	78.2%
BlSat6_4.2	86.1%	84.0%	82.0%	81.8%	82.3%	75.9%	80.9%	84.3%	80.4%		72.2%	81.1%	77.7%	83.0%	85.8%	82.7%	81.4%	83.3%
BlSat6_5.1	70.8%	70.6%	67.6%	70.0%	68.2%	66.5%	67.9%	70.3%	67.0%	72.2%		68.3%	65.3%	72.8%	73.0%	73.4%	68.2%	71.3%
BlSat6_5.2	86.1%	81.7%	78.5%	79.8%	81.1%	72.1%	80.6%	81.1%	79.5%	81.1%	68.3%		77.6%	81.7%	81.4%	81.1%	80.8%	81.7%
BlSat6_5.3	83.7%	78.2%	74.1%	76.9%	75.3%	76.7%	76.7%	78.2%	74.1%	77.7%	65.3%	77.6%		76.6%	78.8%	76.3%	79.1%	76.7%
BlSat6_6.1	85.4%	82.6%	80.6%	80.1%	82.5%	76.0%	78.3%	82.0%	79.4%	83.0%	72.8%	81.7%	76.6%		87.0%	95.5%	82.9%	81.3%
BlSat6_6.2	87.2%	85.4%	82.3%	84.2%	83.5%	77.0%	80.8%	86.1%	81.0%	85.8%	73.0%	81.4%	78.8%	87.0%		87.3%	87.3%	86.1%
BlSat6_7.1	84.5%	82.6%	80.3%	80.1%	81.9%	74.8%	77.7%	82.0%	79.4%	82.7%	73.4%	81.1%	76.3%	95.5%	87.3%		82.6%	80.7%
BlSat6_7.2	84.2%	82.3%	79.4%	80.1%	81.3%	74.2%	80.5%	81.3%	78.5%	81.4%	68.2%	80.8%	79.1%	82.9%	87.3%	82.6%		83.3%
BlSat6_7.3	84.5%	83.0%	81.0%	82.3%	82.3%	76.2%	80.3%	83.9%	78.2%	83.3%	71.3%	81.7%	76.7%	81.3%	86.1%	80.7%	83.3%	

Figure S4: Sequence divergence of BISat1 satellite monomers from *Beta* species

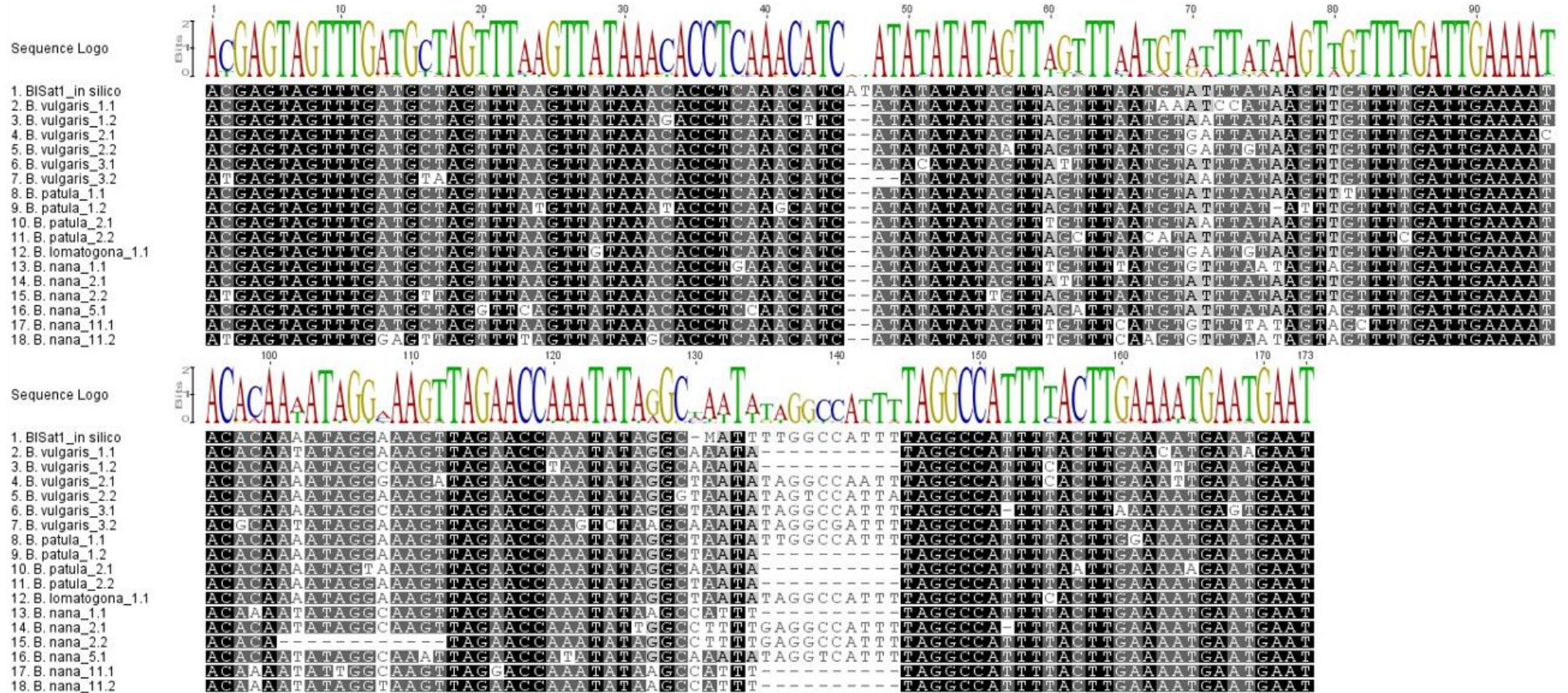


Figure S5: Sequence divergence of BISat5 satellite monomers from *Beta* species

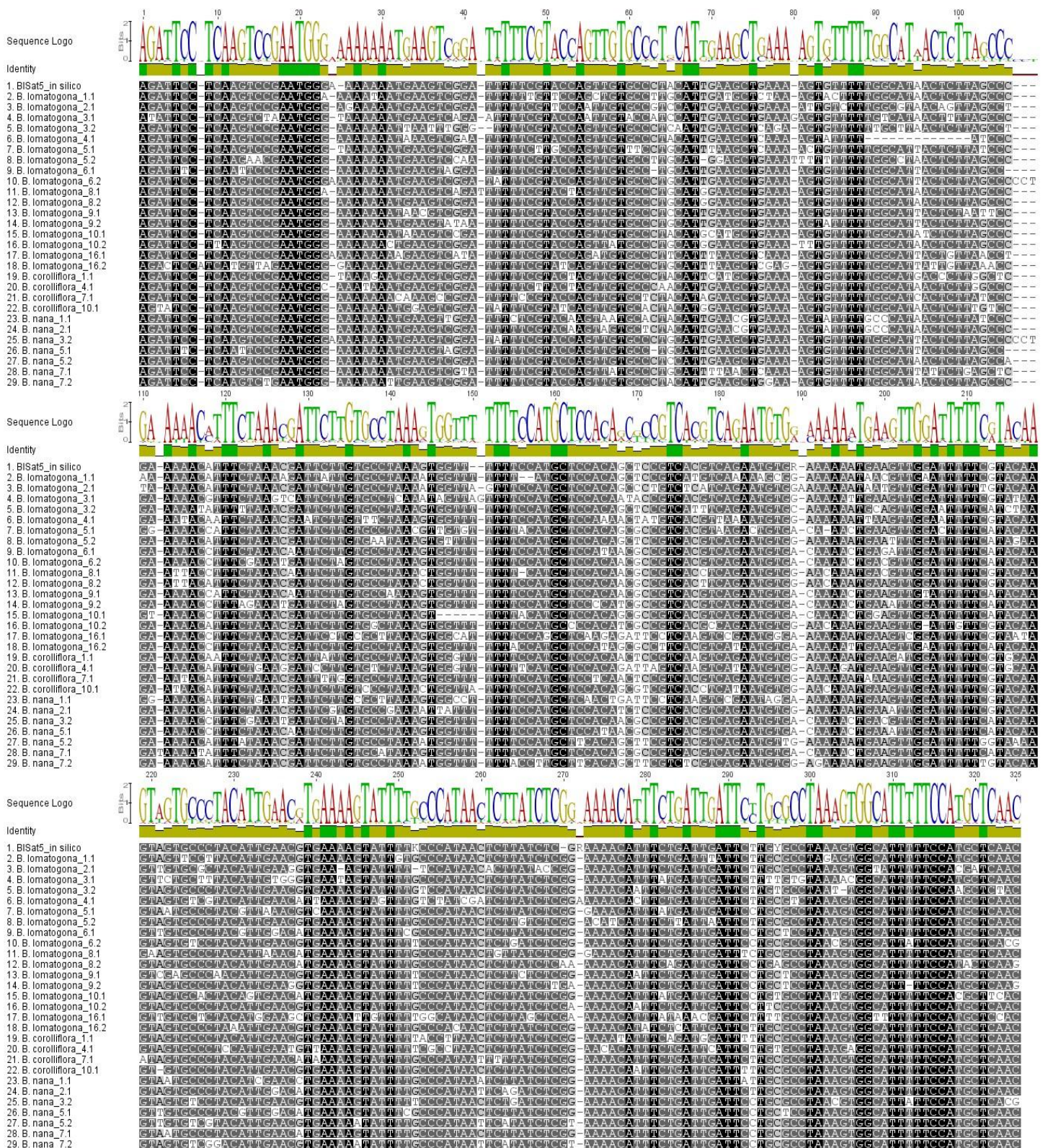


Figure S6: Sequence divergence of BISat6 satellite monomers from *Beta* species

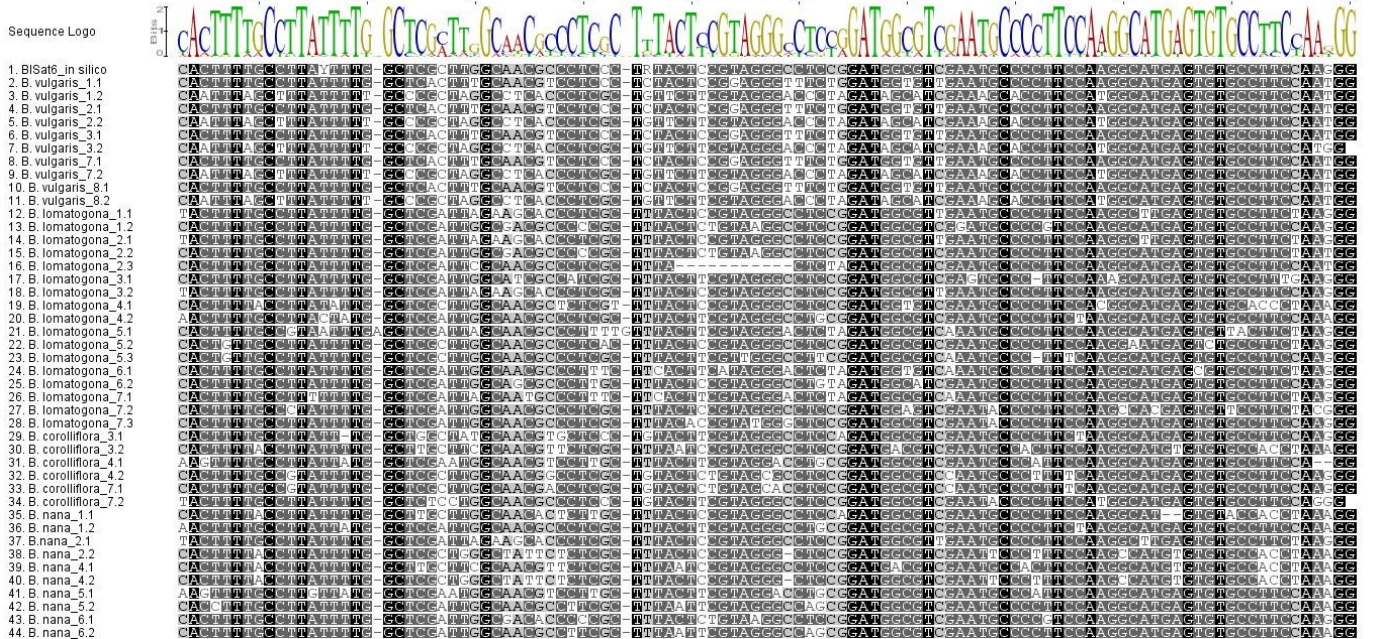
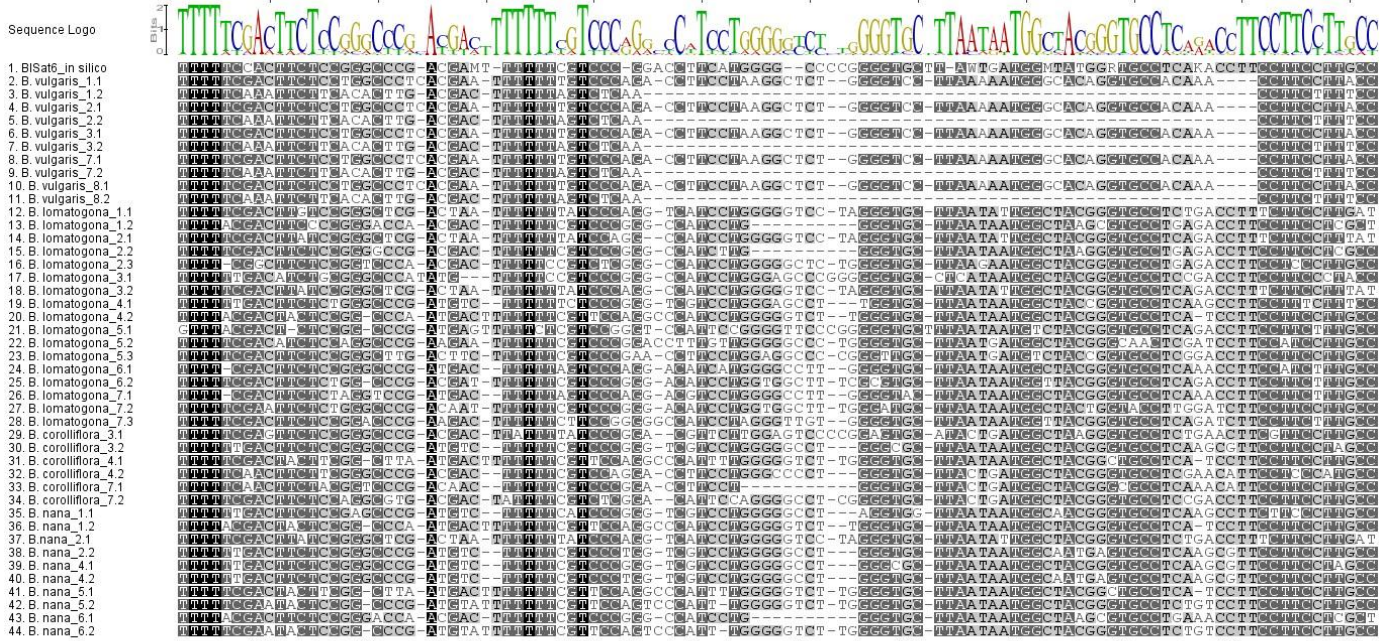
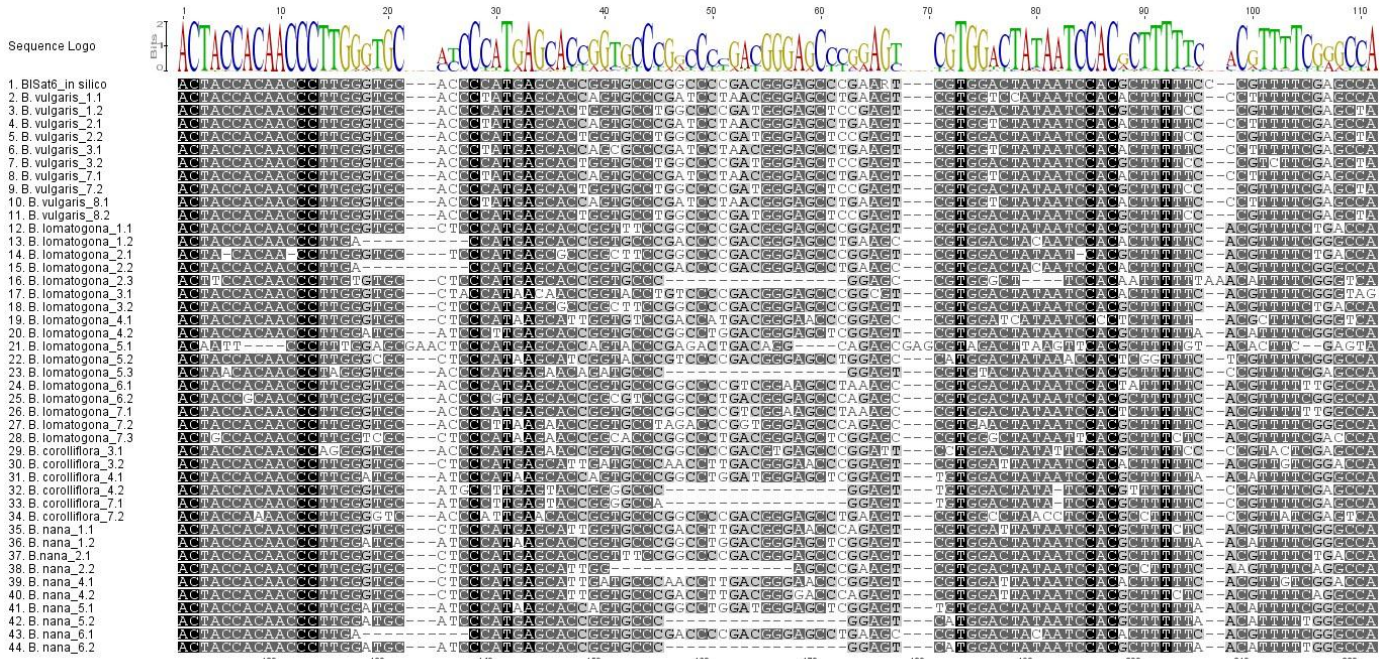
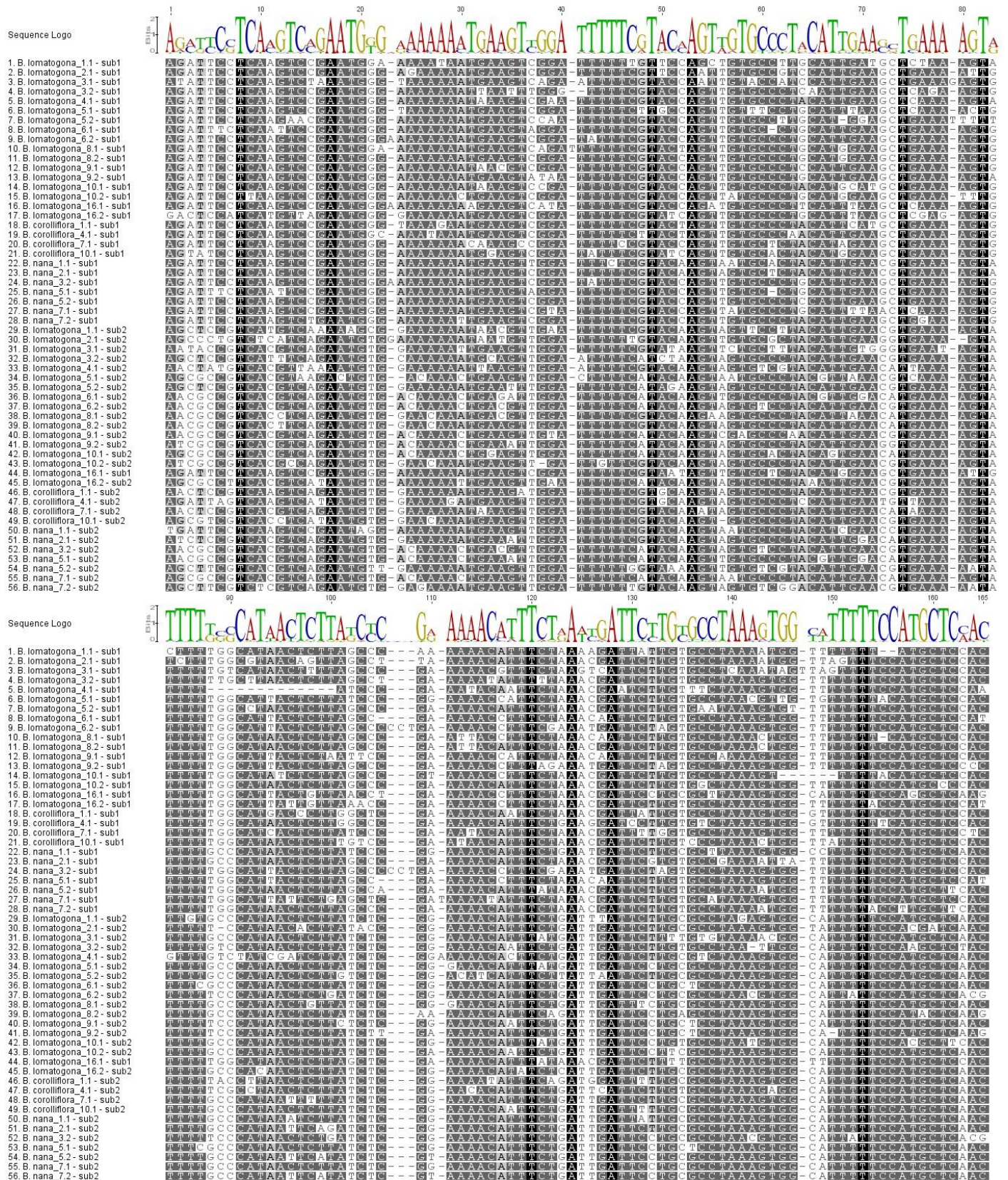


Figure S7: Alignment of subunits sub1 and sub2 from all BISat5 monomers in *Beta* species



**Figure S8: Sequence alignment of BiSat2 satellite monomers from *B. Lomatogona* (A) and distance between the monomers (B)**

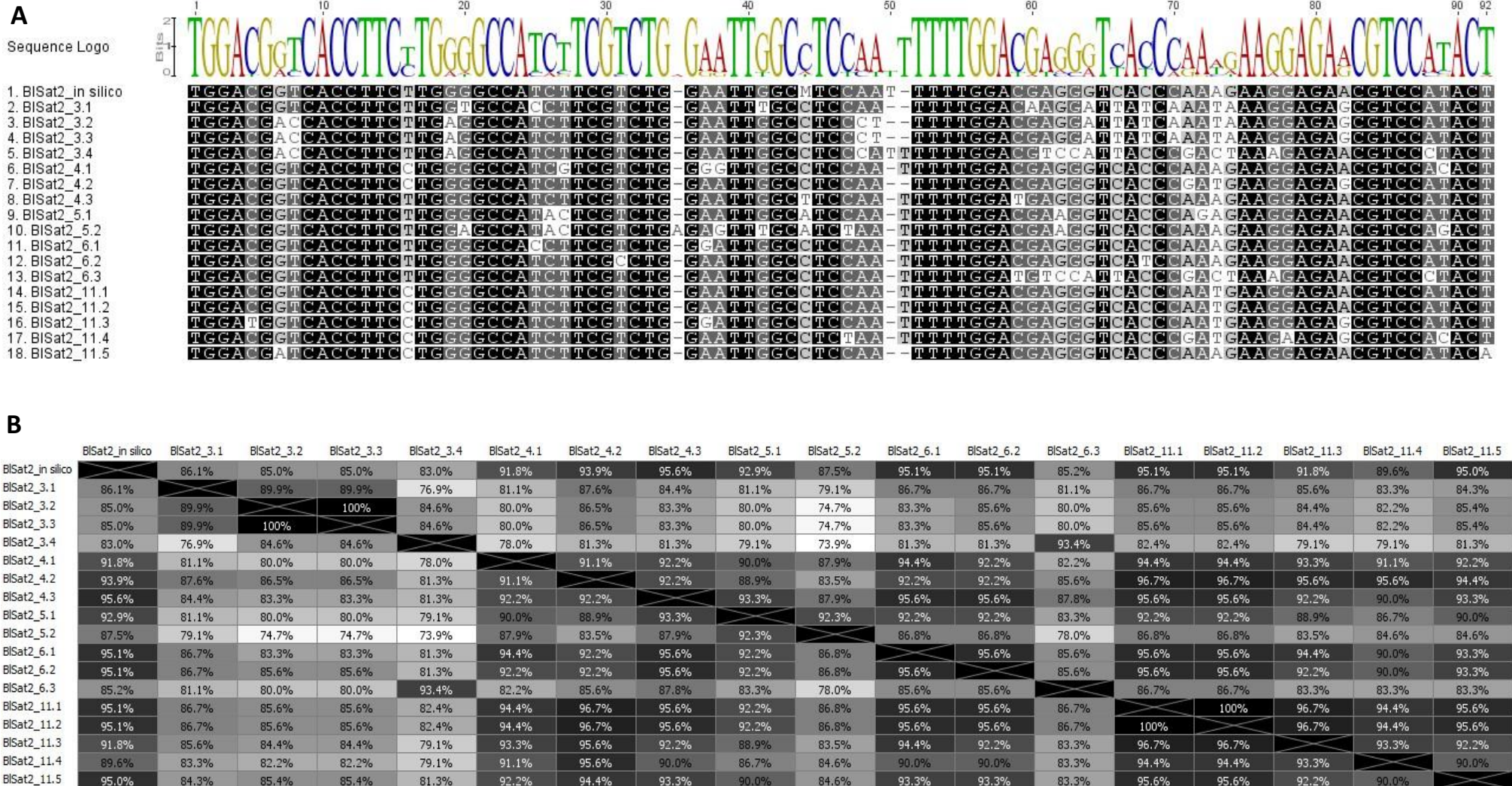
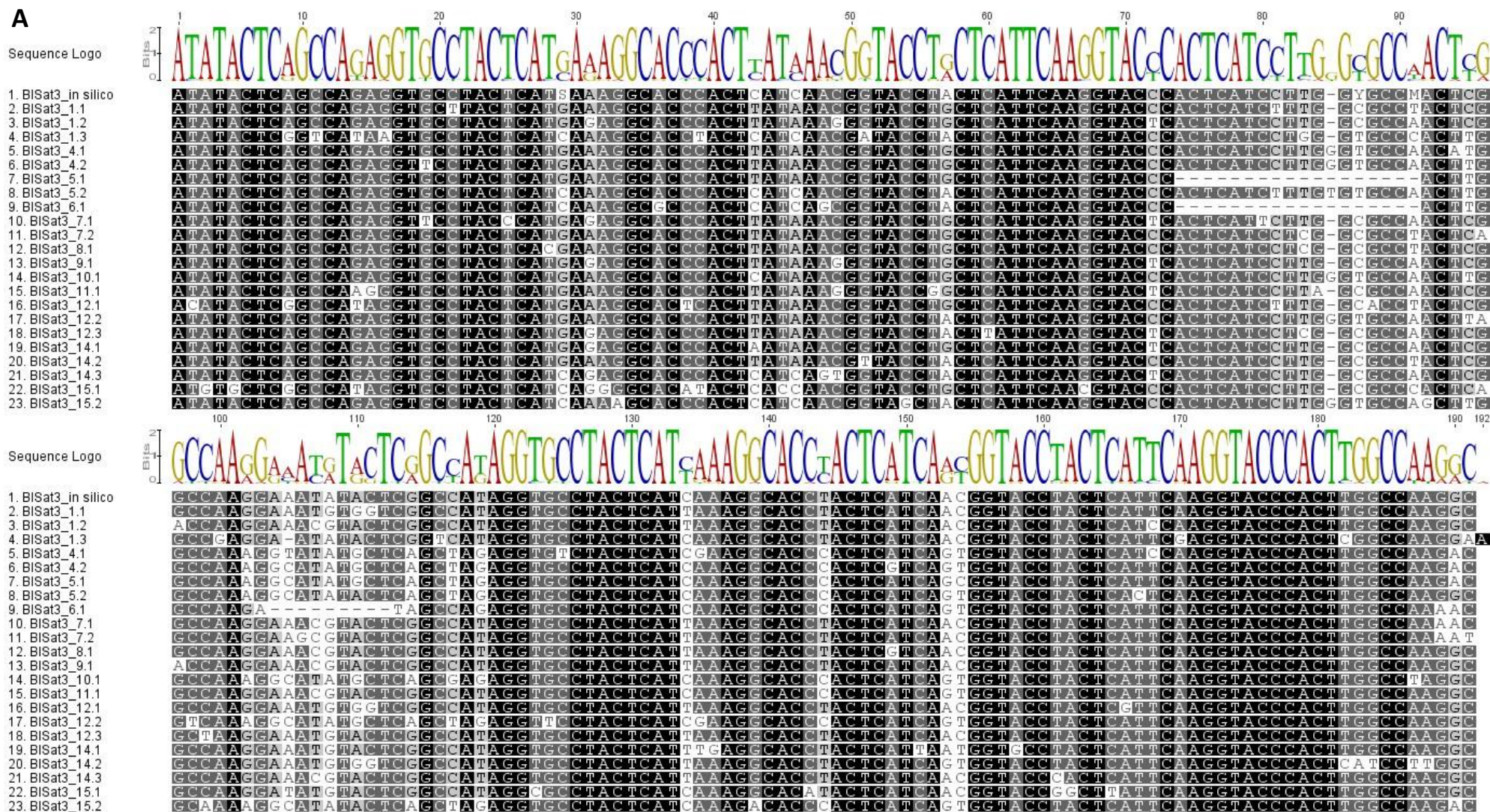


Figure S9: Sequence alignment of BISat3 satellite monomers from *B. Lomatogona* (A) and distance between the monomers (B)

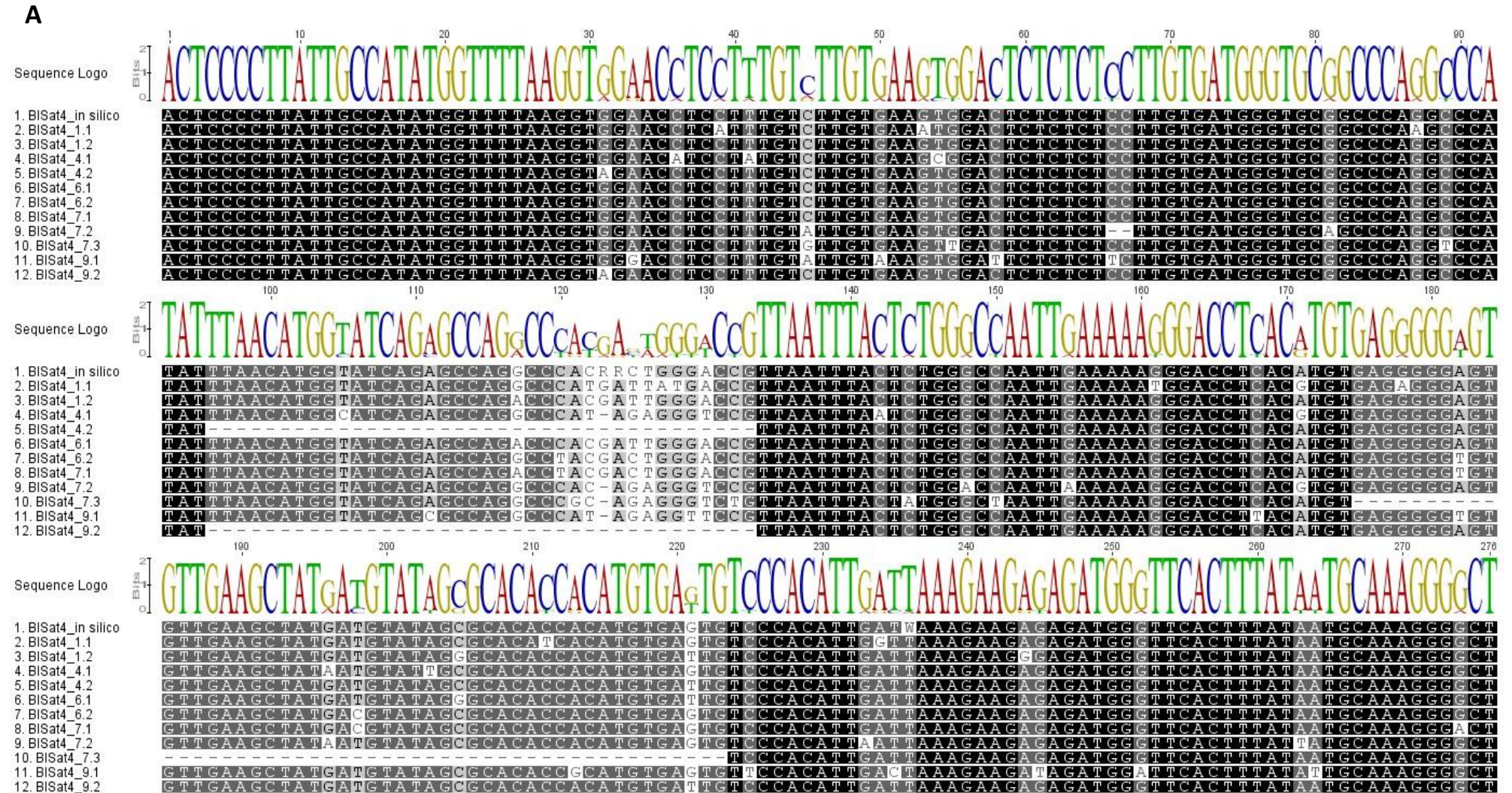


**B**

	BlSat3_j...	BlSat3_1.1	BlSat3_1.2	BlSat3_1.3	BlSat3_4.1	BlSat3_4.2	BlSat3_5.1	BlSat3_5.2	BlSat3_6.1	BlSat3_7.1	BlSat3_7.2	BlSat3_8.1	BlSat3_9.1	BlSat3_1...	BlSat3_1...	BlSat3_1...	BlSat3_1...	BlSat3_1...	BlSat3_1...	BlSat3_1...	BlSat3_1...	BlSat3_1...	
BlSat3_in silico		94.2%	93.4%	91.8%	88.7%	89.8%	83.4%	92.4%	79.7%	92.4%	92.6%	94.2%	93.9%	91.4%	91.8%	91.6%	89.8%	94.5%	92.9%	91.6%	95.0%	89.2%	90.3%
BlSat3_1.1	94.2%		94.2%	86.3%	88.0%	89.0%	83.7%	88.0%	76.8%	93.2%	94.2%	95.8%	94.7%	89.5%	91.6%	96.3%	88.0%	94.2%	93.2%	94.2%	92.1%	86.3%	84.8%
BlSat3_1.2	93.4%	94.2%		85.3%	87.4%	87.4%	81.1%	86.4%	74.7%	95.8%	93.7%	95.3%	99.5%	88.0%	94.2%	91.6%	86.4%	95.8%	94.7%	90.5%	95.3%	86.3%	84.3%
BlSat3_1.3	91.8%	86.3%	85.3%		82.7%	83.8%	77.9%	86.9%	75.1%	84.2%	85.8%	86.8%	85.8%	85.3%	84.2%	85.8%	83.8%	86.8%	84.7%	85.3%	87.4%	85.8%	84.8%
BlSat3_4.1	88.7%	88.0%	87.4%	82.7%		96.3%	87.4%	92.7%	78.0%	86.4%	86.4%	86.9%	86.9%	95.3%	88.0%	85.3%	95.3%	86.4%	86.9%	86.4%	84.3%	80.1%	92.1%
BlSat3_4.2	89.8%	89.0%	87.4%	83.8%	96.3%		89.0%	94.2%	79.1%	88.5%	87.4%	89.0%	88.0%	96.9%	89.0%	86.4%	95.8%	87.4%	87.4%	87.4%	85.3%	80.6%	93.7%
BlSat3_5.1	83.4%	83.7%	81.1%	77.9%	87.4%	89.0%		86.4%	87.9%	81.6%	82.1%	82.6%	81.6%	88.0%	82.1%	81.6%	86.9%	81.6%	80.0%	80.5%	78.9%	74.7%	84.8%
BlSat3_5.2	92.4%	88.0%	86.4%	86.9%	92.7%	94.2%	86.4%		81.2%	85.3%	85.3%	86.9%	86.9%	95.3%	88.0%	85.3%	94.2%	87.4%	86.9%	86.4%	88.5%	82.7%	95.3%
BlSat3_6.1	79.7%	76.8%	74.7%	75.1%	78.0%	79.1%	87.9%	81.2%		76.3%	77.4%	76.3%	75.3%	79.6%	76.8%	74.7%	78.0%	75.8%	74.7%	75.8%	77.9%	71.1%	79.6%
BlSat3_7.1	92.4%	93.2%	95.8%	84.2%	86.4%	88.5%	81.6%	85.3%	76.3%		94.7%	94.2%	96.3%	86.9%	92.1%	90.5%	85.3%	94.7%	93.7%	89.5%	93.7%	85.3%	84.3%
BlSat3_7.2	92.6%	94.2%	93.7%	85.8%	86.4%	87.4%	82.1%	85.3%	77.4%	94.7%		96.3%	94.2%	86.9%	91.1%	91.6%	86.4%	93.7%	91.6%	90.5%	91.6%	85.8%	84.3%
BlSat3_8.1	94.2%	95.8%	95.3%	86.8%	86.9%	89.0%	82.6%	86.9%	76.3%	94.2%	96.3%		95.8%	88.5%	92.6%	93.2%	86.9%	95.3%	93.2%	92.1%	93.2%	86.3%	84.8%
BlSat3_9.1	93.9%	94.7%	99.5%	85.8%	86.9%	88.0%	81.6%	86.9%	75.3%	96.3%	94.2%	95.8%		88.5%	94.7%	92.1%	86.9%	96.3%	95.3%	91.1%	95.8%	86.8%	84.8%
BlSat3_10.1	91.4%	89.5%	88.0%	85.3%	95.3%	96.9%	88.0%	95.3%	79.6%	86.9%	86.9%	88.5%	88.5%		89.5%	86.9%	95.8%	88.0%	88.5%	89.0%	86.9%	82.2%	93.7%
BlSat3_11.1	91.8%	91.6%	94.2%	84.2%	88.0%	89.0%	82.1%	88.0%	76.8%	92.1%	91.1%	92.6%	94.7%	89.5%		89.5%	88.0%	92.1%	92.1%	90.0%	91.6%	84.2%	85.9%
BlSat3_12.1	91.6%	96.3%	91.6%	85.8%	85.3%	86.4%	81.6%	85.3%	74.7%	90.5%	91.6%	93.2%	92.1%	86.9%	89.5%		85.3%	91.6%	90.5%	91.6%	89.5%	86.3%	82.2%
BlSat3_12.2	89.8%	88.0%	86.4%	83.8%	95.3%	95.8%	86.9%	94.2%	78.0%	85.3%	86.4%	86.9%	86.9%	95.8%	88.0%	85.3%		87.4%	86.9%	87.4%	85.3%	80.6%	92.7%
BlSat3_12.3	94.5%	94.2%	95.8%	86.8%	86.4%	87.4%	81.6%	87.4%	75.8%	94.7%	93.7%	95.3%	96.3%	88.0%	92.1%	91.6%	87.4%		94.7%	91.6%	94.7%	86.3%	85.9%
BlSat3_14.1	92.9%	93.2%	94.7%	84.7%	86.9%	87.4%	80.0%	86.9%	74.7%	93.7%	91.6%	93.2%	95.3%	88.5%	92.1%	90.5%	86.9%	94.7%		90.5%	93.2%	85.8%	84.8%
BlSat3_14.2	91.6%	94.2%	90.5%	85.3%	86.4%	87.4%	80.5%	86.4%	75.8%	89.5%	90.5%	92.1%	91.1%	89.0%	90.0%	91.6%	87.4%	91.6%	90.5%		89.5%	82.6%	84.3%
BlSat3_14.3	95.0%	92.1%	95.3%	87.4%	84.3%	85.3%	78.9%	88.5%	77.9%	93.7%	91.6%	93.2%	95.8%	86.9%	91.6%	89.5%	85.3%	94.7%	93.2%	89.5%		87.9%	86.4%
BlSat3_15.1	89.2%	86.3%	86.3%	85.8%	80.1%	80.6%	74.7%	82.7%	71.1%	85.3%	85.8%	86.3%	86.8%	82.2%	84.2%	86.3%	80.6%	86.3%	85.8%	82.6%	87.9%		80.6%
BlSat3_15.2	90.3%	84.8%	84.3%	84.8%	92.1%	93.7%	84.8%	95.3%	79.6%	84.3%	84.3%	84.8%	84.8%	93.7%	85.9%	82.2%	92.7%	85.9%	84.8%	84.3%	86.4%	80.6%	



Figure S10: Sequence alignment of BISat4 satellite monomers from *B. Lomatogona* (A) and distance between the monomers (B)



**B**

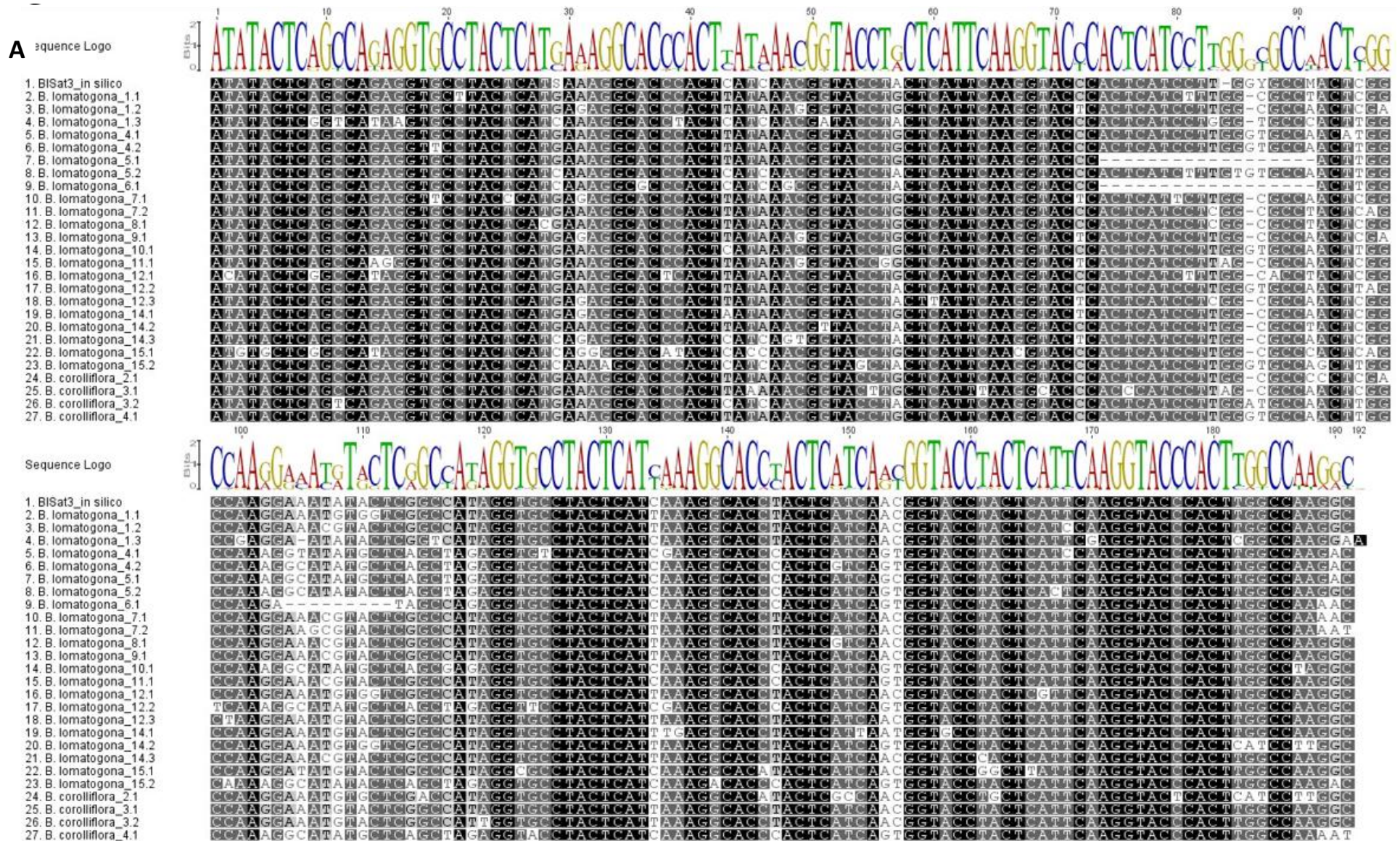
	BlSat4_in silico	BlSat4_1.1	BlSat4_1.2	BlSat4_4.1	BlSat4_4.2	BlSat4_6.1	BlSat4_6.2	BlSat4_7.1	BlSat4_7.2	BlSat4_7.3	BlSat4_9.1	BlSat4_9.2
BlSat4_in silico		95.1%	97.6%	94.9%	85.3%	98.0%	98.4%	97.6%	94.6%	77.9%	92.4%	85.3%
BlSat4_1.1	95.1%		94.6%	92.8%	82.6%	94.9%	94.6%	93.8%	91.7%	75.0%	89.5%	82.6%
BlSat4_1.2	97.6%	94.6%		93.8%	85.1%	99.6%	97.1%	97.1%	93.5%	77.5%	91.3%	85.1%
BlSat4_4.1	94.9%	92.8%	93.8%		83.3%	94.2%	94.2%	93.5%	94.5%	77.1%	91.6%	83.3%
BlSat4_4.2	85.3%	82.6%	85.1%	83.3%		85.5%	84.8%	84.4%	82.2%	66.5%	81.1%	100%
BlSat4_6.1	98.0%	94.9%	99.6%	94.2%	85.5%		97.5%	97.5%	93.8%	77.9%	91.7%	85.5%
BlSat4_6.2	98.4%	94.6%	97.1%	94.2%	84.8%	97.5%		99.3%	93.8%	77.9%	92.4%	84.8%
BlSat4_7.1	97.6%	93.8%	97.1%	93.5%	84.4%	97.5%	99.3%		93.1%	77.2%	91.7%	84.4%
BlSat4_7.2	94.6%	91.7%	93.5%	94.5%	82.2%	93.8%	93.8%	93.1%		76.7%	91.6%	82.2%
BlSat4_7.3	77.9%	75.0%	77.5%	77.1%	66.5%	77.9%	77.9%	77.2%	76.7%		74.9%	66.5%
BlSat4_9.1	92.4%	89.5%	91.3%	91.6%	81.1%	91.7%	92.4%	91.7%	91.6%	74.9%		81.1%
BlSat4_9.2	85.3%	82.6%	85.1%	83.3%	100%	85.5%	84.8%	84.4%	82.2%	66.5%	81.1%	





**Figure S12: Sequence alignment of BISat3 satellite monomers from *Beta* species**

(A) The divergent monomer sequences of BISat2 in the genus *Beta* and (B) Dendrogram representation of relationship among BISat2 sequences from *Beta* and *Patellifolia* species



**B**

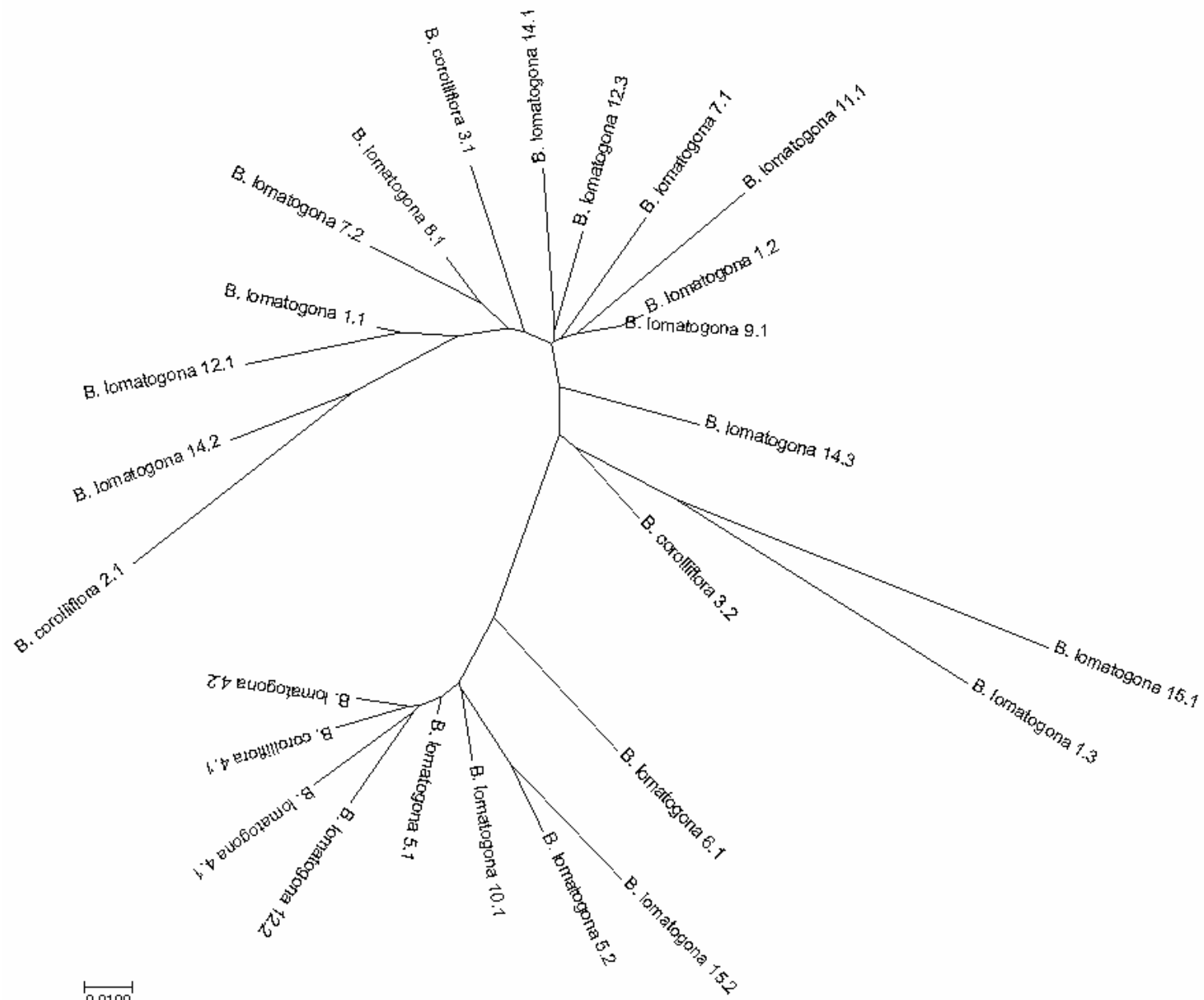
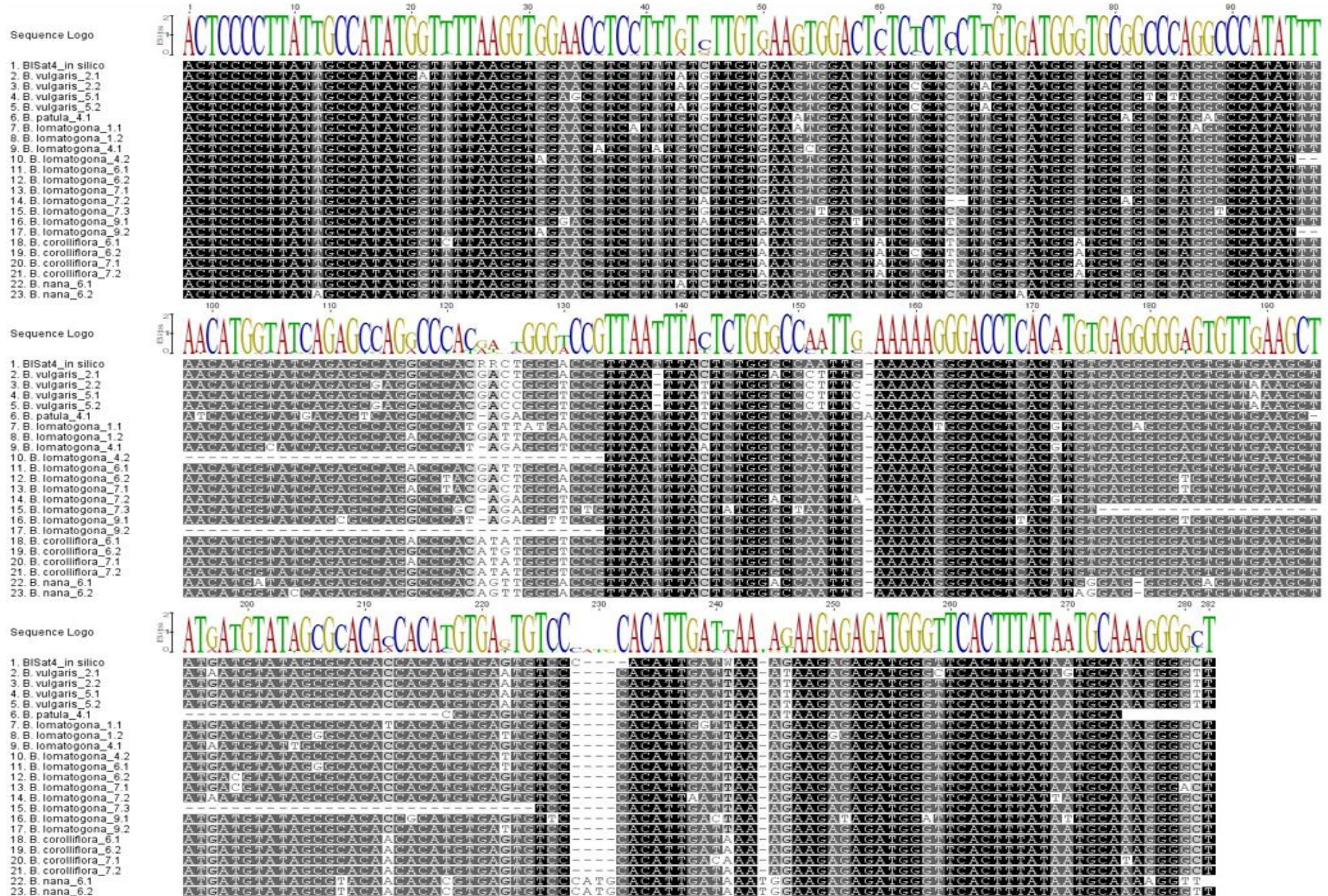


Figure S13: Sequence divergence of BISat4 satellite monomers from *Beta* species













## **Erklärung**

Hiermit versichere ich, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer als der angegebenen Hilfsmittel angefertigt habe; die aus fremden Quellen direkt oder indirekt übernommenen Gedanken sind als solche kenntlich gemacht. Die Arbeit wurde bisher weder im Inland noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfungsbehörde vorgelegt.

Die vorliegende Dissertation wurde unter wissenschaftlicher Betreuung durch Herrn Prof. Thomas Schmidt am Botanik Institut.

Ich versichere, dass diese Promotionsordnung in aktuelle Fassung anerkannt wird.

Bich Hong Ha