

**Cytogenetic analyses of the genus *Genlisea*,
which is characterized by striking genome plasticity**

Dissertation

zur Erlangung des
Doktorgrades der Naturwissenschaften (Dr. rer. nat.)

der

Naturwissenschaftlichen Fakultät I – Biowissenschaften –

der Martin-Luther-Universität
Halle-Wittenberg,

vorgelegt

von Herrn Tran Duc Trung

geb. am March 08th, 1983 in Hanoi, Vietnam

Gutachter:

1. **Prof. Dr. Ingo Schubert**
IPK, Gatersleben, Germany
2. **Prof. Dr. Gunter Reuter**
Martin-Luther-Universität, Halle-Wittenberg, Germany
3. **Prof. Dr. Thomas Schmidt**
Institut für Botanik, TU Dresden, Dresden, Germany

Tag der Promotion: 11.8.2015

Acknowledgements

This work was funded by the Deutsche Forschungsgemeinschaft (DFG) and supported in part by scholarship from the Ministry of Education and Training (MOET) of Vietnam. All experiments, if is not further notified, were carried out at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben.

First and foremost, I would like to express my deepest appreciation to my supervisor, Prof. Dr. Ingo Schubert, who gave me a great opportunity to be a part of his research group and pursue my academic goals. Without his invaluable guidance, fruitful discussion and constant encouragement during the research period, this work would have never been accomplished. His patience in reading each chapter and providing detailed feedback for my dissertation was outstanding. I am deeply grateful for great support he gave to me over the past four years. I lack words to describe the support of my co-supervisor, Dr. Joerg Fuchs. His generosity, advice and friendship have been an enormous inspiration of mine. I would like to show my sincere gratitude for his consistent enthusiasm during last years.

On this occasion, I would like to show my tremendous gratefulness to Dr. Giang T.H. Vu for introducing me to Prof. Dr. Ingo Schubert and for her step-by-step instruction in BAC library construction. I also would like to acknowledge Dr. Gabriel Jotvchev, who guided me to the colorful microscopic world of cytogenetic.

I appreciate the members of the former research group Karyotype Evolution, Dr. Hieu X. Cao, Dr. Inna Lermontova and Dr. Veit Schubert, for their insightful contribution to this work. Many thanks go to Martina Kuehne, Andrea Kunze and Joachim Bruder, who always gave me excellent technical assistance whenever I need.

Within IPK, I would like to send my thanks to Ines Walde (research group Genome Diversity) for her generous help in preparation of spotted membranes for *G. margaretae* BAC library, to Dr. Renate Schmidt and members of research group Genome Plasticity for their delicate work in dot-blot DNA hybridization, and to Dr. Andreas Houben and members of research group Chromosome Structure and Function for their helpful supports and suggestions. I also would like to thank Dr. Ha T.M. Pham (research group Yeast Genetics) for her helpful comments and suggestions during my dissertation preparation.

My grateful thanks go to Dr. Jiri Macas and his group (Institute of Plant Molecular Biology, Ceske Budejovice, Czech Republic) for their brilliant bioinformatic results, to Prof. Dr. Jiri Fajkus and his group (Masaryk University, Brno, Czech Republic) for their amazing data on telomere work, and to Dr. Hana Simkova and her group (Institute of Experimental Botany, Olomouc, Czech Republic) for their valuable outcome in BAC library construction.

PhD student's life would not go smoothly if it was only filled with academic work. I would like to thank my colleagues and friends, who shared with me enjoyable and precious moments at IPK-Gatersleben, and who encouraged and supported me a lot in scientific work. I also gratefully thank Dr. Brit Leps for all of her helps in administrative business, which made my stay at IPK very comfortable.

I wish to express my thanks to Assoc. Prof. Dr. La Tuan Nghia and my colleagues at home country institution (Plant Resources Center, VAAS, Vietnam) for their encouragements as well as facilitation for my abroad study.

My warmest thanks go to my parents, my mother-in-law, my sister and brother, and my nephews for their love and encouragement. It would be a great pride to say that, as a family, you backed me up and I am forever grateful for this.

Last but not least, this is the time for me to express my thankfulness to my wife and my little son, who continuously encouraged me during the last four years by their endless loves, unconditional beliefs and deep empathies. This dissertation stands as a dedication to their un-recompensable sacrifices.

Tran Duc Trung

TABLE OF CONTENTS

List of figures	i
List of tables.....	ii
Abbreviations	iii
1. INTRODUCTION	1
1.1. THE NUCLEAR GENOME STRUCTURE AND ORGANIZATION	1
1.1.1. Nuclear genome composition.....	1
1.1.1.1. Tandem repetitive DNA sequences	1
1.1.1.2. Dispersed repetitive DNA sequences	2
1.1.2. Chromosome structure and organization	3
1.1.2.1. Chromosomes in interphase nucleus	4
1.1.2.2. Metaphase chromosomes.....	5
1.1.3. Genome size variation in plants	7
1.1.3.1. Genome size enlargement by recurrent whole genome duplication (WGD) and massive accumulation of transposable elements	8
1.1.3.2. Genome size shrinkage through accumulating deletions	9
1.1.4. Karyotype alteration	10
1.2. THE CARNIVOROUS GENUS <i>GENLISEA</i> - SUBJECT OF THE STUDY.....	12
1.2.1. Geographic distribution and infrageneric relationship	12
1.2.2. Unique morphological and molecular features of <i>Genlisea</i> species	14
1.3. GENOME SIZE AND KARYOTYPE EVOLUTION IN <i>LENTIBULARIACEAE</i>	16
1.4. AIMS OF THE DISSERTATION	18
2. MATERIALS AND METHODS	19
2.1. PLANT MATERIAL AND CULTIVATION	19
2.2. GENOMIC DNA ISOLATION AND CYTOLOGICAL PREPARATIONS	19
2.3. GENOME SIZE MEASUREMENT	21
2.4. BAC LIBRARY CONSTRUCTION FOR <i>G. MARGARETAE</i> AND COLONY MEMBRANE PREPARATION.....	21
2.5. SCREENING FOR REPEAT-FREE BAC CLONES OF <i>G. MARGARETAE</i> USING DOT-BLOT DNA HYBRIDIZATION	22
2.6. IN-SILICO ANALYSIS FOR THE IDENTIFICATION OF REPETITIVE ELEMENTS AND SINGLE-COPY SEQUENCES	22
2.7. IMMUNOSTAINING	24
2.8. FISH PROBE GENERATION AND LABELING	25
2.9. FLUORESCENCE IN SITU HYBRIDIZATION (FISH).....	27
2.10. MICROSCOPY AND IMAGE PROCESSING	28
2.11. ANALYSIS OF TELOMERE LENGTHS AND <i>BAL31</i> -SENSITIVITY.....	28
2.12. TELOMERE REPEAT AMPLIFICATION PROTOCOL (TRAP).....	29
2.13. SEQUENCE ANALYSIS	30
3. RESULTS AND DISCUSSION.....	31
3.1. DIFFERENCES IN CYTOLOGICAL FEATURES REFLECT THE GENOME SIZE PLASTICITY IN THE SUBGENUS <i>GENLISEA</i>	31

3.1.1. The variation of nuclear DNA content reveals one of the largest genome size ranges for a flowering plant genus.....	31
3.1.2. Nuclear distribution of epigenetic methylation marks depends not only on genome size but also on the dispersion of repetitive DNA	33
3.1.3. The chromosome number is highly variable in groups with miniaturized genomes but likely constant in the group with larger genomes.....	35
3.1.4. Number and chromosomal distribution of ribosomal DNA loci are similar within, but different between groups characterized by small or large genomes.....	38
3.1.5. Conclusion regarding the polymorphism of cytological features in correlation with genome size plasticity in the subgenus <i>Genlisea</i>	40
3.2. THE VARIABILITY OF REPETITIVE DNA BETWEEN <i>G. NIGROCAULIS</i> AND <i>G. HISPIDULA</i> IS REFLECTED BY THE ALTERATIONS IN CENTROMERE AND TELOMERE SEQUENCES	41
3.2.1. FISH revealed a different chromosomal distribution of repetitive sequences in <i>G. nigrocaulis</i> and <i>G. hispidula</i>	41
3.2.2. A 161 bp tandem repeat is a centromeric repeat in <i>G. nigrocaulis</i> whereas centromere regions of <i>G. hispidula</i> are predominantly occupied by four centromeric retrotransposons	46
3.2.3. The <i>Arabidopsis</i> -type telomeric minisatellite is conserved in <i>G. nigrocaulis</i> and its close relative <i>G. pygmaea</i>	51
3.2.4. In <i>G. hispidula</i> the <i>Arabidopsis</i> -type telomere repeat is replaced by two sequence variants.....	54
3.2.5. Telomeric repeat variation in <i>Genlisea</i> is the first example for an intrageneric switch.....	56
3.3. FISH-BASED CHROMOSOME IDENTIFICATION, A PREREQUISITE TO STUDY KARYOTYPE EVOLUTION IN <i>GENLISEA</i>	58
3.3.1. In combination with rDNA probes, single-copy FISH discriminated 11 chromosome pairs of <i>G. nigrocaulis</i> and revealed homeologous chromosomes of <i>G. pygmaea</i>	59
3.3.2. Thirteen chromosome pairs were distinguishable by different types of tandem repeats in <i>G. hispidula</i>	65
3.3.3. FISH-based chromosome identification using repeat-free BAC clones identified 15 of the 19 chromosome pairs of <i>G. margaretae</i>	68
3.3.4. Conclusions regarding karyotyping in <i>Genlisea</i>	76
4. SUMMARY	78
References	80
Publications related to this dissertation	89
Declaration about personal contributions	90
Curriculum vitae	91
Declarations	92

List of figures

Figure 1: Dysploid alterations of chromosome number.	11
Figure 2: Phylogenetic relationship and general morphology of the three Lentibulariaceae genera.	12
Figure 3: Two species of the subgenus <i>Genlisea</i> and the molecular phylogenetic tree of the genus <i>Genlisea</i>	13
Figure 4: Structure of <i>G. nigrocaulis</i> chromosomes after two sequential FISH experiments.	28
Figure 5: The striking difference in genome size within <i>Genlisea</i>	32
Figure 6: Nuclear heterochromatin organization and distribution of methylation marks in <i>Genlisea</i> species in comparison to <i>A. thaliana</i>	34
Figure 7: Chromosome numbers of six <i>Genlisea</i> species representing three sections of the subgenus <i>Genlisea</i>	37
Figure 8: Different loci number and chromosomal distribution of 5S and 45S rDNA in <i>Genlisea</i> species.	39
Figure 9: Nuclear and chromosomal distribution of different repetitive sequences in <i>G. nigrocaulis</i>	42
Figure 10: Nuclear and chromosomal distributions of different repetitive sequences in <i>G. hispidula</i>	44
Figure 11: Cross-FISH using repetitive probes identified from <i>G. nigrocaulis</i> and <i>G. hispidula</i> on nuclei or chromosomes of <i>G. pygmaea</i> and <i>G. subglabra</i> , respectively.	45
Figure 12: Characterization of the four putative centromeric retrotransposons in <i>G. hispidula</i>	47
Figure 13: Immunostaining-FISH combinations validating centromeric position of putative repetitive elements.	48
Figure 14: Sequence organization of centromeric GnCent repeats in <i>G. nigrocaulis</i> and <i>G. pygmaea</i>	50
Figure 15: Localization of centromere- and telomere-specific repeats in <i>G. nigrocaulis</i> and <i>G. pygmaea</i>	52
Figure 16: Telomere sequence confirmation and telomerase activity analysis in <i>G. pygmaea</i>	53
Figure 17: Cytogenetic characterization of two novel telomere variants of <i>G. hispidula</i>	54
Figure 18: Confirmation of the telomere identity of two novel sequence variants in <i>G. hispidula</i>	55
Figure 19: Cytogenetic characterization of telomeric repeats in <i>G. subglabra</i> , <i>G. margaretae</i> and <i>G. aurea</i>	57

Figure 20: Simplified phylogenetic tree of green plants indicating telomeric sequence deviations from TTAGGG, the most wide-spread telomeric repeat among plants.	58
Figure 21: Localization of ten single-copy fragments by FISH on metaphase plates of <i>G. nigrocaulis</i>	60
Figure 22: FISH-based chromosome identification using unique and rDNA probes in <i>G. nigrocaulis</i>	62
Figure 23: Cross-FISH using nine of ten single-copy probes derived from <i>G. nigrocaulis</i> to chromosomes of <i>G. pygmaea</i>	64
Figure 24: Cross-FISH revealed the loss of a homeologous locus in tetraploid <i>G. pygmaea</i>	65
Figure 25: FISH-based chromosome identification using tandem repeat probes in <i>G. hispidula</i>	66
Figure 26: Sequence exchange between homeologous chromosomes after interspecific hybridization in <i>G. hispidula</i>	67
Figure 27: Identification and characterization of repeat-free BAC clones of <i>G. margaretae</i>	69
Figure 28: Chromosomal localization of <i>G. margaretae</i> repeat-free BACs as revealed by sequential three-color FISH (BAC pools 1-4, subset 1).	73
Figure 29: Chromosomal localization of <i>G. margaretae</i> repeat-free BACs as revealed by sequential three-color FISH (BAC pools 5-8, subset 2).	74
Figure 30: Sequential FISH using repeat-free BAC probes and 45S rDNA identify 15 of the 19 chromosome pairs of <i>G. margaretae</i>	75

List of tables

Table 1: Genome sizes and chromosome numbers of some <i>Genlisea</i> species.	15
Table 2: List of <i>Genlisea</i> species used in this study	19
Table 3: List of primers used to generate probes.....	23
Table 4: List of primary antibodies	25
Table 5: Sequences of primers used in analyses of telomerase activity in tissues of <i>G. hispidula</i> and <i>G. pygmaea</i>	30
Table 6 : Established procedures for cytological preparation of <i>Genlisea</i> species using different types of tissue.	36
Table 7: Estimated sizes and FISH signal patterns for 60 putatively repeat-free BAC clones of <i>G. margaretae</i> selected by dot-blot DNA hybridization.	70

Abbreviations

BSA	Bovine serum albumin
BAC	Bacterial artificial chromosome
bp	Base pair
CRM	Centromeric retrotransposon of maize
DAPI	4',6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
DSB	Double-strand break
dNTP	Deoxynucleotide triphosphate
dUTP	Deoxyuridine triphosphate
EDTA	Ethylenediaminetetraacetic acid
FISH	Fluorescence <i>in situ</i> hybridization
GISH	Genomic <i>in situ</i> hybridization
HOR	Higher order repeat
kbp	kilo base pair
LTR	Long terminal repeat
Mbp	Mega base pair
Mya	Million years ago
NOR	Nucleolus organizer region
PCR	Polymerase chain reaction
PFGE	Pulsed-field gel electrophoresis
PVP	Polyvinylpyrrolidone
TE	Transposable element
TRAP	Telomere repeat amplification protocol
TRF	Terminal restriction fragment
WGD	Whole genome duplication
μm	micrometer

1. INTRODUCTION

1.1. The nuclear genome structure and organization

The genome of eukaryotic organisms carries the genetic information encoded in complementary double-stranded DNA molecules. The genome is contained in three different compartments. The major one is the nuclear genome harboring heredity material within distinct, linear chromosomes in the cell nucleus. The organellar genomes of plastids and mitochondria derived from prokaryotic genomes are much smaller and mostly circular.

1.1.1. Nuclear genome composition

Genomes of all eukaryotes consist of two categories of DNA sequences regarding their abundance: *i*) single- or low-copy sequences comprising genes (exons, introns) and regulatory elements, and *ii*) high-copy or repetitive sequences. In plants, the annotation of more than 50 sequenced genomes revealed approximately 20,000 to 94,000 predicted genes with an estimated average number of ~32,000. These protein-coding genes are commonly scattered throughout the genome and flanked by non-coding, repetitive sequences which were found to be highly variable between genomes of all sequenced species (Michael & Jackson, 2013). Previously, repetitive sequences were generally considered as “junk” or “selfish” DNA because no beneficial function to the host genome was recognizable. However, later insights from genomic studies revealed that some repetitive sequences, in different manners, may influence gene regulation and genome structure (for review see Shapiro & von Sternberg, 2005). According to their genomic organization, repetitive sequences are classified into two main groups: tandem repetitive and dispersed repetitive DNA (Lopez-Flores & Garrido-Ramos, 2012).

1.1.1.1. Tandem repetitive DNA sequences

Some repetitive portions of eukaryotic genomes have a particular nucleotide composition deviating from the average of the species-specific GC content towards either more AT or more GC. These genome fractions thus form “satellite” bands separating from that of bulk genomic DNA during density gradient centrifugation (Thiery *et al.*, 1976). The sequences of these “satellite” bands mostly represent tandem repeats. Based on the length of basic repeat unit, microsatellite, minisatellite and satellite DNA are the three major types of tandem repetitive DNA sequences. Whereas microsatellite units (usually less than 9 bp) are found in arrays of about 1 kbp distributed throughout chromosomes in both non-coding and coding regions, minisatellite units (from 9 to 100 bp) may extend up to several kbp and cluster in subtelomeric, pericentromeric or interstitial regions of chromosomes.

Satellite DNAs with a monomer length ranging from 100 to 400 bp may constitute Mbp-long arrays (Lopez-Flores & Garrido-Ramos, 2012; Mehrotra & Goyal, 2014).

High evolutionary dynamics is a notable characteristic of tandem repetitive DNA sequences. Micro- and minisatellites show high mutation rates (in copy number rather than in sequence of the repeated units) and are therefore unstably inherited and often polymorphic even between individuals of a population. On the other hand, satellite DNAs differing in unit length, nucleotide sequence and/or in copy number may be species- and/or chromosome-specific. Mitotic sister chromatid exchange, meiotic crossing-over and gene conversion were proposed as the main molecular mechanisms that generate but also maintain the intra- and interspecific polymorphism of tandem repetitive sequences which can be used to distinguish chromosomes within a species and between related species (Hemleben *et al.*, 2007; Lopez-Flores & Garrido-Ramos, 2012).

Among tandem repetitive DNA sequences, centromeric and telomeric repeats and ribosomal genes (rDNA) are well-characterized. Similar to centromeric and telomeric repeats, ribosomal genes belong to the highly repetitive DNA and may be arrayed in hundreds to ten thousands of copies (Rogers & Bendich, 1987). Unlike centromeric repeats that greatly differ between species, telomeric repeats and rDNA sequences are more conserved (see section 1.1.2.2). Thus 45S and 5S rDNA which usually display a species-specific cluster distribution are frequently used as markers for karyotyping by FISH (for review see Garcia *et al.*, 2014a).

1.1.1.2. Dispersed repetitive DNA sequences

Transposable elements (TEs) are the most abundant dispersed repetitive DNA sequences. Two major classes of TEs were characterized based on their structural features and mechanisms of transposition. Retrotransposons (or class I elements) transpose *via* a “copy and paste” mechanism by means of an RNA intermediate. Class I elements are further divided into two subclasses: long terminal repeat (LTR) retrotransposons and non-LTR retrotransposons, the latter includes long and short interspersed nuclear elements (LINEs and SINEs, respectively). Both LINEs and SINEs are considered as precursors or ancestors of LTR-retrotransposons (Schmidt, 1999). DNA transposons (or class II elements) use the “cut and paste” mechanism without an RNA intermediate to move to (a) new chromosomal positions (for review see Wicker *et al.*, 2007). The abundance and diversity of TEs within the genome are variable among eukaryotes. Although sequenced genomes revealed a similar number of TEs families, not all of them proliferated to high-copy numbers in individual plant species. For example, in *Oryza brachyantha* and *Brachypodium distachyon*, characterized by similar genome sizes (~300 Mbp/1C) and repetitive DNA proportions (29.2 and 21.4%, respectively), the *Mutator*

family represents the most abundant DNA transposon in the former (7.5%; Chen *et al.*, 2013) but occupies only 0.63% of the genome in the latter (International Brachypodium, 2010). In all plant species investigated so far, the most ubiquitous dispersed DNA elements belong to two superfamilies, *Ty1/copia* and *Ty3/gypsy*, of LTR-retrotransposons (Wicker *et al.*, 2007; Zhao & Ma, 2013). In some species such as maize and barley, LTR elements may occupy up to 75% of the genome and scatter throughout most of chromosomes (Baucom *et al.*, 2009; Mayer *et al.*, 2012).

Retroelements not only contribute to genome expansion (see section 1.1.3) but may also function as transcriptional enhancers or silencers regulating the expression of host genes. In amphiploid wheat derived from an interspecific hybridization followed by chromosome doubling, the *Wis2-1A* retroelements were found transcriptionally active. Depending on position (downstream or upstream the nearby genes), the LTRs of *Wis2-1A* via transcriptional interference altered the transcript synthesis of either antisense or sense strand which is associated with silencing or activation of adjacent genes, respectively (Kashkush *et al.*, 2003). Another example is the insertion of a *Tcs1*-like *Ty1/copia* element adjacent to the *Ruby* gene, a transcriptional activator of anthocyanin production, altering the fruit color expression of *Citrus* species (Butelli *et al.*, 2012). Moreover, retroelements may induce, via mis-repair of DNA double strand breaks (DSBs), chromosomal mutations, such as deletions, translocations and inversions, and thus reconstruct the karyotype of the host organisms (Schubert *et al.*, 2004) (see section 1.1.4).

1.1.2. Chromosome structure and organization

The DNA of the eukaryotic nuclear genome together with special proteins forms the hierarchically structured chromatin and is packaged into distinct linear chromosomes. During the cell cycle chromosomes adopt different levels of compaction. The basic unit of chromatin is the nucleosome chain. An octamer of four histones (H3, H4, H2A and H2B) constitutes the core of a nucleosome wrapped by ~146 bp of a DNA double helix and characterized by certain dynamic histone modifications (Kouzarides, 2007). The array of nucleosomes connected by 20 - 60 bp of linker DNA forms the approximate 11-nm diameter “beads-on-a-string” fiber, the first level of chromatin organization. Binding to the linker DNA and the nucleosome as well, the linker histone (H1 or H5) helps to stabilize the presumed more condensed 30-nm chromatin filament, the second structural level of chromatin as supposed by some researchers (for review see Li & Reinberg, 2011). However, in contrast to the “beads-on-a-string” fiber, it is uncertain whether the 30-nm filament exists and according to which model it is further structured at the higher order levels of chromatin organization up to mitotic chromosomes (Joti *et al.*, 2012; Ausio, 2015). From the cytological and molecular view, chromatin is divided into two classes including the less condensed, potentially transcriptionally active

euchromatin and the more condensed, transcriptionally mostly silenced heterochromatin. Two types of heterochromatin were characterized in mammals and plants. Whereas constitutive heterochromatin regions are mostly composed of highly tandem repetitive DNA and largely transcriptionally inactive, facultative heterochromatin regions containing TEs and genes are reversibly inactivated such as the inactive X chromosome of female mammalian organisms (for review see Trojer & Reinberg, 2007).

The wrapping of the DNA molecule around the nucleosomes and the higher folding structure of chromatin prevent transcription, because the two DNA strands need to be temporarily separated allowing the access of essential components such as polymerases. Acetylation, phosphorylation and methylation of histones, particularly at some amino acids of their N-terminal tails, are the main epigenetic modifications altering chromatin structure and thus facilitating transcription, DNA replication, repair and recombination. For instance, acetylation neutralizes the positive charges of lysine residues, thus loosening the histone-DNA binding of chromatin for synthesis activities. The acetylation level of lysine residues of different chromatin regions can vary during cell cycle and between species. Phosphorylation, mostly at serine, threonine and tyrosine residues, also may alter chromatin condensation *via* charge changes. Some phosphorylated serine and threonine residues, such as H3S10ph and H3S28ph in plants, strongly associate with chromatin condensation during mitosis and meiosis (see section 1.1.2.2). Methylation does not alter the charge of histone protein. Frequently found on lysine and arginine, methylation has extra levels of complexity (lysine residues can be mono-, di-, or tri-methylated whereas arginine residues can be mono- or di- (asymmetric or symmetric) methylated) (for review see Fuchs *et al.*, 2006; Kouzarides, 2007; Bannister & Kouzarides, 2011). Methylated lysine residues of histone H3 and H4 are stable during cell cycle and considered as signals for high or low transcription potential of eu- and heterochromatin, respectively (see section 1.1.2.1.).

Directly regulating the transcriptional activity of DNA, the cytosine DNA methylation (5-methylcytosine or 5mC) occurs in CpG, CpHpG, and CpHpH context in plants, where H represents any nucleotide but guanine (He *et al.*, 2011). Since (peri)centromeric sequences as well as other repetitive elements are heavily methylated, cytosine methylation together with some histone methylations are considered as heterochromatic marks.

1.1.2.1. Chromosomes in interphase nucleus

During interphase of cell cycle, the stage between two nuclear divisions, important genetic activities such as replication, transcription and DNA repair take place. The interphase chromosomes are rather

decondensed and occupy individual territories which together are surrounded by the nuclear envelope. The interphase nucleus reveals eu- and heterochromatic regions. While euchromatin promotes gene expression, heterochromatin including (peri)centromeres and (sub)telomeres assembles into densely stained regions such as the chromocenters of *A. thaliana* (Fransz *et al.*, 2002). The nuclear arrangement of chromosome territories can be determined by FISH using chromosome-specific probes. In some organisms the so-called Rabl orientation with centromeres clustered at one nuclear pole and telomeres at the other is maintained from the late anaphase throughout interphase as shown in wheat, barley or rye (Dong & Jiang, 1998). However in other species the Rabl orientation is not recognizable. Instead, random nuclear arrangement of chromosome territories appears after chromosome painting by FISH with chromosome-specific BAC clones in *Arabidopsis* species (Berr *et al.*, 2006). The content and distribution of repetitive sequences, organization of eu- and heterochromatin and epigenetic modifications may influence the nuclear arrangement of chromosomes (for review see Wako & Fukui, 2010; Schubert & Shaw, 2011).

The patterns of DNA methylation and histone post-translational modifications, particularly methylation, further characterize the chromatin organization within interphase nuclei. In nuclei of *A. thaliana*, cytosine methylation preferentially accumulates at heterochromatic chromocenters mainly comprising (peri)centromeric repeats (Fransz *et al.*, 2002). Houben *et al.* (2003) observed an apparent dependence of subnuclear and chromosomal distribution of di-methylation of lysine 9 of histone H3 (H3K9me₂) from nuclear DNA content. H3K9me₂ was found preferentially accumulated in heterochromatic chromocenters of species with a genome size of less than 500 Mbp/1C, while it was nearly homogeneously dispersed in nuclei of species possessing larger genomes. On the other hand, the euchromatin-specific modification H3K4me₂ was observed in all species to be homogeneously distributed except at heterochromatic chromocenters regardless of the nuclear DNA content (Houben *et al.*, 2003). In the most intensively investigated *A. thaliana*, other histone methylations were found either at heterochromatin (H3K9me_{1, 2}; H3K27me_{1, 2} and H4K20me₁) or at euchromatin (H3K4me_{1, 2, 3}; H3K9me₃; H3K27me₃; H3K36me_{1, 2, 3} and H4K20me_{2, 3}) (Fuchs *et al.*, 2006). In spite of some exceptional chromatin specificities such as euchromatin-associated H3K9me₂ in maize (Shi & Dawe, 2006), most of histone and also DNA methylations show conserved chromatin-specific patterns in angiosperms (for review see Fuchs & Schubert, 2012).

1.1.2.2. Metaphase chromosomes

The linear chromosomes reach their highest condensations during the nuclear division when they become microscopically visible. Monocentric eukaryotic chromosomes at metaphase consist of two sister chromatids cohering at the primary constriction, the centromere, where the kinetochore

proteins assemble and spindle microtubules attach to move the sister chromatids apart during anaphase. The two physical ends of each chromatid are protected from degradation and fusion by telomere structures and counteract chromosome end shortening during DNA replication by telomerase activity. Despite the highly conserved function of centromeres, telomeres and most of their proteins, DNA sequences in particular those of centromeres are less conserved among eukaryotic organisms.

Tandem repeats are the most common centromeric DNA sequences reported for plants and animals. A typical characteristic of these centromeric repeats is the rapid divergence (Henikoff *et al.*, 2001; Ma *et al.*, 2007). A large scale comparative analysis of centromeric repeats of hundreds of plant and animal species revealed a very low overall homology (Melters *et al.*, 2013). Variation in sequence and length of repeat units of centromeric repeats has been detected *e.g.* within *A. thaliana* (centromeric repeat family pAL1; Martinez-Zapater *et al.*, 1986), *Oryza sativa* (CentO, formerly RCS2 family; Dong *et al.*, 1998; Cheng *et al.*, 2002) and barley (minisatellite AGGGAG; Hudakova *et al.*, 2001). Another frequent component of centromeric DNA are retroelements. Centromeric retrotransposons (CR), belonging to the chromovirus clade of *T3/gypsy* LTR-retroelements, are abundant, especially in centromeres of grass species (Neumann *et al.*, 2011). Centromeric tandem repeats and CRs, either alone or together, contribute to the structure of centromeres which may span from hundreds of kbp to several Mb in different species, and vary regarding their extension even among chromosomes of a complement (Hosouchi *et al.*, 2002; Jiang *et al.*, 2003; Plohl *et al.*, 2014). However only a part of such extended and complex sequence blocks forms the functional centromere which contains the centromeric histone variant CenH3 and binds kinetochore proteins (for review see Houben & Schubert, 2003; Jiang *et al.*, 2003; Fukagawa & Earnshaw, 2014). Further (peri)centromere-specific histone marks are histone H2A phosphorylated at threonine 133 in maize (Dong & Han, 2012) or at threonine 120 (Demidov *et al.*, 2014), and histone H3 phosphorylated at serine 10 and 28 (Houben *et al.*, 1999) in all tested plants.

Compared to centromeric sequences, the telomere DNA sequences show a higher degree of conservation within eukaryotic phyla. In plants, the heptanucleotide repeat (TTTAGGG) which was first discovered in *A. thaliana* (Richards & Ausubel, 1988) is the most common telomere sequence (Fuchs *et al.*, 1995). Nevertheless, within families of the monocotyledonous order Asparagales this plant-type telomeric repeat is replaced completely or partially by the vertebrate-type (TTAGGG) or the *Tetrahymena*-type (TTGGGG) (Adams *et al.*, 2001; Weiss & Scherthan, 2002; Sykorova *et al.*, 2003c; 2006), or by unknown sequences in the genus *Allium* (Pich & Schubert, 1998). Sequence alterations of the telomeric repeats were also detected within species of three closely related genera

of the dicotyledonous family Solanaceae. Whereas a non-canonical (TTTTTAGGG) repeat was characterized to be maintained by telomerase activity of *Cestrum elegans* (Peska *et al.*, 2015), telomeric repeats that protect chromosome ends of *Vestia* and *Sessea* species are unknown (Sykorova *et al.*, 2003a).

In addition to a centromere and two telomeres, metaphase chromosomes usually possess one or more characteristic heterochromatic blocks which on large chromosomes often can be visualized by different banding techniques (Schreck & Disteche, 2001). Such gene-poor heterochromatic blocks are “hot spots” for chromosome rearrangements due to mis-repair of DSBs, when ectopic instead of allelic homologous sequences are used as template for homologous recombination repair (Schubert *et al.*, 2004).

Number, shape and size of chromosomes are characteristic for each eukaryotic organism. The specific chromosome complement of an organism is called karyotype. In plants, the chromosome number can range from $2n = 4$ (*Haplopappus gracilis*; Jackson, 1959) to $2n = 1,440$ (*Ophioglossum reticulatum*; Khandelwal, 1990). Nevertheless, there is no consistent correlation between chromosome number and genome size. The primary constriction divides the monocentric chromosome into two arms. The arm length ratio further determines the shape of chromosome as metacentric, sub-metacentric, acrocentric or telocentric (Levan *et al.*, 1964). In combination with chromosome-specific bands, number, shape and size of a chromosome complement are representative features of a karyotype (Levin, 2002).

1.1.3. Genome size variation in plants

Swift (1950) demonstrated the apparent constancy of the haploid amount of nuclear DNA, which he termed as “C-value”, among cells of *Zea mays* and *Tradescantia* species. Standing for nuclear genome size or “holoploid genome size” (Greilhuber *et al.*, 2005), the C-value is defined as the number of base pairs in the double-helical DNA molecules of an un-replicated, basic chromosome set of gametes (Soltis *et al.*, 2003). The base pair number can be reversibly converted to mass (1 pg of double strand DNA equals 978 Mbp) as calculated by Dolezel *et al.* (2003). So far, genome size estimates of nearly 9,000 angiosperm species revealed an up to 2,440-fold genome size difference with the smallest genome (~61 Mbp/1C) claimed for carnivorous *Genlisea tuberosa* (Fleischmann *et al.*, 2014) and the largest genome of *Paris japonica* (up to 150 Gbp/1C; Pellicer *et al.*, 2010). This variation shows neither relation to the morphological complexity nor to the gene number of different species. Thomas (1971) described this phenomenon as “C-value paradox” which later on was called “C-value enigma” by Gregory (2001). A long-standing question concerns the causes, mechanisms and

biological significance of genome size variation in different groups of organism. Bennetzen and Kellogg (1997) supposed a unidirectional genome size evolution, saying that plants “have a one way ticket to larger genome sizes” through polyploidization (whole genome duplication) and accumulation of transposable elements. However, recent research supports also the opposite evolutionary trend, namely genome size reduction by the active removing of dispensable DNA.

1.1.3.1. Genome size enlargement by recurrent whole genome duplication (WGD) and massive accumulation of transposable elements

WGD or polyploidy, which results from either multiplication of one genome (autopolyploidy) or combination of two or more divergent genomes (allopolyploidy) *via* interspecific hybridization, is widespread among flowering plants and is one of the major mechanisms that account for the large genome size differences. Considering the occurrence of WGD in the context of evolutionary history, polyploid species may be classified as paleo-, meso-, or neopolyploidy. Paleopolyploid species experienced one or more archaic WGD events during their evolution. The homeologous regions of these old duplications are no longer cytogenetically detectable, and paleopolyploid WGD events such as those causing about 30–36-fold duplication of ancestral angiosperm genes in *Gossypium* species (Paterson *et al.*, 2012) can only be traced *via* comparative bioinformatic analysis of orthologous sequences. There is genomic evidence that all extant seed plants have a paleopolyploid ancestry ~192 Mya (Jiao *et al.*, 2011). But also later WGDs happened in several lineages of flowering plants. Evolutionary younger WGD events followed by dysploid chromosome rearrangements resulted in mesopolyploidy. Although mesopolyploid species have a diploid-like chromosome number, WGD events are still detectable, for instance, by comparative genomic analysis as in *Brassica rapa* (Wang *et al.*, 2011) or by comparative chromosome painting as in other Brassicaceae species (Lysak *et al.*, 2005; Mandakova *et al.*, 2010). On the other hand, recent WGD events create neopolyploid species with multiple genome sizes and chromosome numbers. However, “genome diploidization” frequently occurs after WGD events and gradually restores during the course of evolution a diploid-like state through *i*) elimination of redundant and non-essential sequences, *ii*) divergence of homeologous sequences *via* mutations and *iii*) chromosomal rearrangements (Renny-Byfield & Wendel, 2014). Genome diploidization generated, for instance, the extant “diploid” genome of *A. thaliana*, which underwent at least three rounds of ancient WGD (Eckardt, 2004). The diploidisation of neopolyploid genomes may start already in the first generations after of WGD such as in the ~100 year old allopolyploid *Tragopogon miscellus* (Tate *et al.*, 2009).

As a consequence of ancient WGDs and subsequent diploidization processes, numbers of protein-coding gene in sequenced genomes vary, but do not linearly correlate with genome sizes

(Michael, 2014). By contrast, repetitive sequences appear in all sequenced genomes and their proportions apparently correlate with genome size, for instance TEs comprise more than 80% genome of barley but occupy about ~20% of the 18-fold smaller genome of *Branchypodium distachyon* (Michael & Jackson, 2013). Although larger and smaller genomes often harbor similar numbers of LTR-retroelement families, some of them show explosive proliferation toward thousands of copies within larger genomes but display only few copies in smaller genomes (Bennetzen & Wang, 2014). For example, three LTR retroelements, the *Ty1/copia RIRE1* and two *Ty3/gypsy* elements *Kangourou* and *Wallabi*, occupy ~60% genome and thus account for the two-fold larger genome size of *O. australiensis* (965 Mbp/1C; Piegu *et al.*, 2006) compared to that of rice. Even horizontal transfer of TEs was detected among 40 sequenced plant genomes (El Baidouri *et al.*, 2014). Interestingly, these alien TEs remained actively transposing and could cause a transpositional burst, and thus genome expansion. However, in order to keep the genome stable, most TEs are silenced transcriptionally by suppressing epigenetic chromatin modifications, or post-transcriptionally *via* transcript degradation mediated by small interfering RNA (Kazazian, 2004). In longer terms, silenced TEs can even be removed from genomes by deletions that counteract genome expansion (Bennetzen & Wang, 2014).

1.1.3.2. Genome size shrinkage through accumulating deletions

In contrast to genome enlargement, mechanisms behind genome shrinkage are less known. To counteract genome expansion, unequal homologous and illegitimate recombinations were claimed as causes of genome size reduction in plants (Bennetzen *et al.*, 2005; Bennetzen & Wang, 2014; Michael, 2014). LTR retroelements, the driver of genome expansion, are also a main target of DNA removal. Ectopic homologous recombination of two LTR retrotransposons may result in solo-LTRs (Shirasu *et al.*, 2000). Because LTR retrotransposons can be organized as nested arrangements, unequal homologous recombination can remove many elements and thus decrease genome size. The solo-LTR formation through unequal homologous recombination for three *Ty3/gypsy*-like elements, *hopi*, *Retrosat1*, and *RIRE3*, was supposed to contribute to genome size decrease in rice (Vitte & Panaud, 2003). Illegitimate recombination during mis-repair of DNA DSBs without a requirement of homologous sequences can ubiquitously occur, and is frequently associated with small or even large deletions. Such DNA depletions gradually downsize the genome (Bennetzen & Wang, 2014). The accumulation of small deletions, caused by illegitimate recombination, shrank the genome of *Arabidopsis* and rice (Bennetzen *et al.*, 2005). Alternatively, the rapid loss of the retroelement *Gorge3* from small genomes led to three-fold genome size difference in the genus *Gossypium* (Hawkins *et al.*, 2009). Furthermore, the frequency and extension of DNA deletions seem to be

inversely correlated with genome size, since the prevalence and average length of deletions that occurred during DNA DSB repair were found to be larger in *Arabidopsis* than in tobacco (Kirik *et al.*, 2000).

Despite the fact that most of the ~3% of the 335,200 angiosperms with estimated genome size (Garcia *et al.*, 2014b) possess genome sizes less than 500 Mbp/1C (Michael, 2014), the efficiency of deletion bias as main driver of genome shrinkage (for review see Petrov, 2002; Gregory, 2004; Kuo & Ochman, 2009) is still a matter of controversy. The same is true for the question whether DNA loss is adaptive or neutral (Wolf & Koonin, 2013).

1.1.4. Karyotype alteration

During the evolutionary history chromosome alteration, together with genome size evolution, is most likely a cause rather than a consequence of speciation because such alterations may establish fertility barriers. The chromosome number variation of congeneric species may be a result of ploidy mutations (autopolyploidy or allopolyploidy) which are frequently observed in plants.

Mis-segregation (aneuploidy) or structural chromosome rearrangements are further mechanisms changing chromosome complements. Size and structure of chromosomes can be altered primarily by duplication, deletion, inversion or translocation. The chromosome structure can be altered also by secondary rearrangements. For instance, in individuals that are doubly heterozygous for two primary rearrangements with one chromosome involved in both, meiotic crossing over between homologous regions flanked by non-homologous regions leads to gametes with a new karyotype, and to complementary gametes displaying a re-established wild type chromosome complement (for review see Schubert, 2007; Schubert & Lysak, 2011).

The chromosome number may be changed *via* dysploid chromosome rearrangements accompanied by little or negligible changes in genome size (Schubert & Lysak, 2011). Chromosome number reductions may occur by reciprocal translocations between two chromosomes. When such an event yields a large and a very small chromosome; the small one tends to get lost during meiosis if it does not carry essential genes (Figures 1A, B). Three dysploid rearrangements were accounted for the chromosome number reduction from eight to five during evolution of *A. thaliana* (Lysak *et al.*, 2006a). By contrast, chromosome number can be increased by “fission” within a centromeric region of a chromosome splitting it into two new telocentric ones (Figure 1C). The two new chromosomes may survive when either a telomere array pre-exists within the split centromere site (Schubert *et al.*, 1995) or telomere sequences are *de novo* added to the break ends (Nelson *et al.*, 2011). The chromosome number increase from $n = 8$ to $n = 14$ within the monkeyflowers genus *Mimulus* was

assumed to result from at least eight “fission” plus two “fusion” events as suggested by comparative linkage mapping with gene-based markers (Fishman *et al.*, 2014). Alternatively, ascending and descending dysploidy may simultaneously occur by two translocations between three chromosomes (one involved in both translocations) followed by segregation that leads to gametes containing one chromosome more than parental lines and gametes having one chromosome less (Figure 1C). Fertilization of gametes of the same dysploid karyotype thereafter establishes homozygous progenies with chromosome number increase or decrease in comparison to their parents as experimentally proven in *Vicia faba* (Schubert & Rieger, 1985).

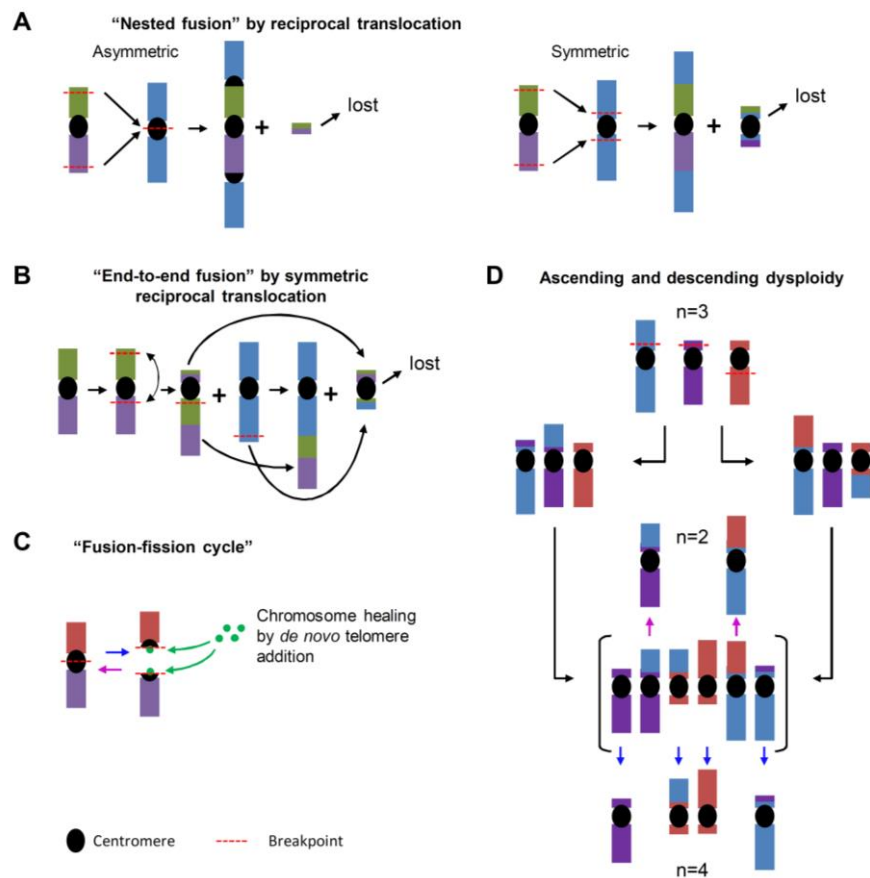


Figure 1: Dysploid alterations of chromosome number.

(A) “Nested fusion” by asymmetric (requires three DSBs) or symmetric (requires four DSBs) reciprocal translocation combines two chromosomes without significant loss of genetic material. The small acentric fragment (of the asymmetric translocation) is lost during mitosis, and the centric fragment (of the symmetric translocation) might get lost during meiosis. (B) “End-to-end fusion” by symmetric reciprocal translocation involving at least one acrocentric chromosome, which may be derived from a metacentric chromosome by pericentromeric inversion, yields a large and a small monocentric chromosome, the latter prone to get lost during meiosis. (C) Chromosome number variation by “fusion-fission cycles”. By asymmetric reciprocal translocation with break points within telomeric sequence arrays at centric ends, two telocentric chromosomes generate a large meta(di)centric chromosome (magenta arrow). A break within the centromere of a metacentric chromosome generates two telocentric ones (blue arrow), which require for stabilization *de novo* addition of telomeric sequences (if the break does not occur within remnant of telomere from a previous “fusion” event). (D) Descending and ascending dysploid karyotypes can be the result of mis-segregation from

meiotic hexavalents (in bracket) of individuals doubly heterozygous for two translocations involving three chromosomes (one of them is metacentric), when the two metacentric translocation chromosomes segregate to one pole (magenta arrows) and the four acrocentric chromosomes to the other (blue arrows), respectively. This figure was redrawn based on Schubert and Lysak (2011).

1.2. The carnivorous genus *Genlisea* - subject of the study

The carnivorous family Lentibulariaceae belongs to the high core clade of the order Lamiales (Schaferhoff *et al.*, 2010) and comprises three monophyletic genera owning distinct morphology: *Pinguicula* (butterworts), *Utricularia* (bladderworts) and *Genlisea* (corkscrew plants) (Mueller *et al.*, 2003; Mueller *et al.*, 2006; Fleischmann, 2012). Interestingly, each of the three genera developed a peculiar trapping mechanism. *Pinguicula* species use sticky, glandular leaves (flypaper traps) to catch small insects. *Utricularia* species have subterranean leaves forming unique bladder-shaped suction traps to catch mainly aquatic animals and phytoplankton. The genus *Genlisea* developed lobster pot traps from root-like subterranean and chlorophyll-free leaves to attract and entrap a wide spectrum of soil-borne microscopic organisms. Taxonomic and phylogenetic treatments established *Genlisea* and *Utricularia* as sister genera closer related to each other than to the genus *Pinguicula* (Figure 2) (Lloyd, 1942; Jobson & Albert, 2002; Jobson *et al.*, 2003).

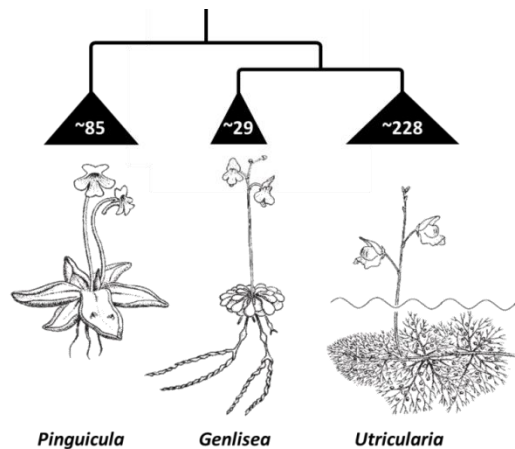


Figure 2: Phylogenetic relationship and general morphology of the three Lentibulariaceae genera.

The estimated species number is shown. Sources: Mueller *et al.* (2006).

1.2.1. Geographic distribution and infrageneric relationship

The first specimens of *Genlisea*, collected from Brazil, were described by Auguste de Saint-Hilaire, a French botanist and naturalist in 1833 (Lloyd, 1942). To date, at least 29 *Genlisea* species native to Neotropics and the tropical Africa (including Madagascar) are recognized. All of them have similar habitats of seasonally wet to waterlogged, nutrient-poor soil with low vegetation cover. The geographic distribution is clearly different between species, some only occur in a narrow geographical area while others are more widespread, but no species can be found in both continents (Fleischmann, 2012; Fleischmann *et al.*, 2014).

Using the morphological character of capsule dehiscence, Elza Fromm-Trinta divided the genus into two main sections: *Tayloria* characterized by longitudinal capsule splitting and *Genlisea* recognized by capsules that open spirally or create curved valves (Taylor, 1991). The two sections were later on promoted as two subgenera, *Tayloria* and *Genlisea*, based on a revision of Africa-originated *Genlisea* species (Fischer *et al.*, 2000). The phylogenetic relationships of almost all recognized *Genlisea* species were recently reconstructed using combined *trnK*, *rps16* and *trnQ-rps16* datasets (Fleischmann *et al.*, 2010; Fleischmann *et al.*, 2014). In accordance with the morphological treatment, the molecular phylogenetic analysis clearly confirmed two subgenera which previously were discriminated by capsule dehiscence. Furthermore, the phylogenetic tree also displays the geographic distribution of *Genlisea* species. All eight species belonging to the subgenus *Tayloria* are endemic to South America (Brazil). The subgenus *Genlisea* is further divided into three geographic-specific sections namely *Africanae* (six species) and *Recurvatae* (three species) comprising the tropical Africa and Madagascar-native species and *Genlisea* consisting of 12 South America-native species (Figure 3C) (Fleischmann *et al.*, 2010).

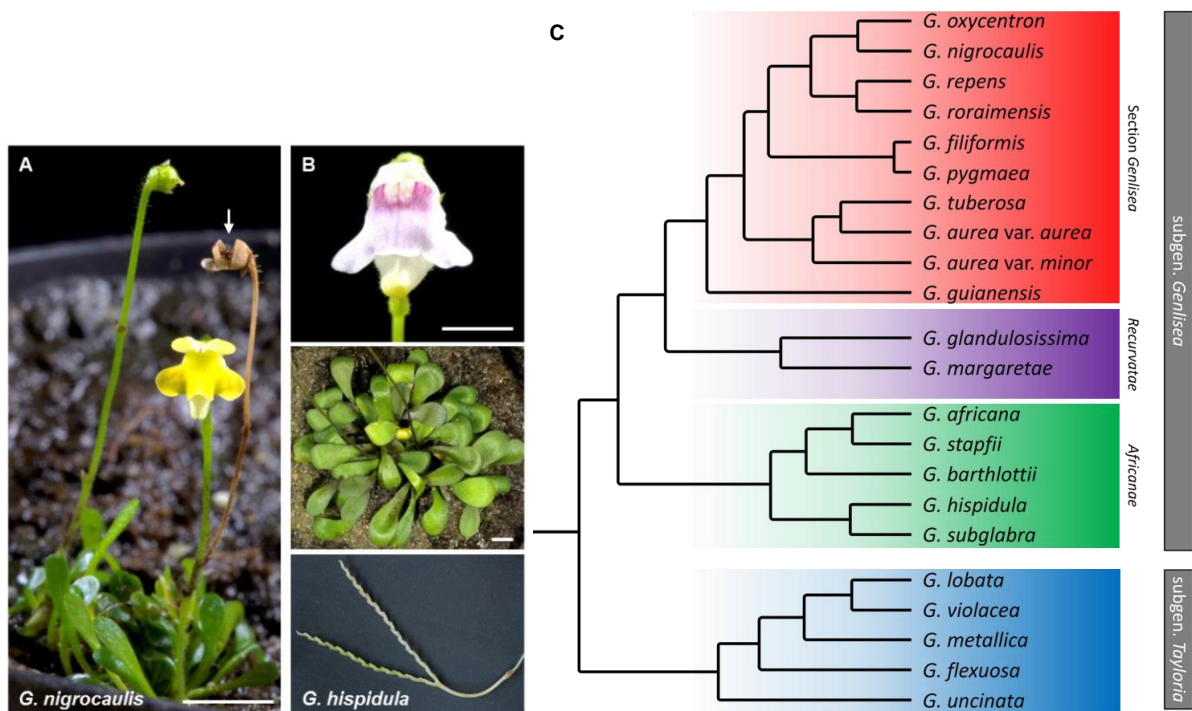


Figure 3: Two species of the subgenus *Genlisea* and the molecular phylogenetic tree of the genus *Genlisea*.

(A) A flowering *G. nigrocaulis* plant (section *Genlisea*). The white arrow denotes the opened capsule. (B) Flower (above), rosette (middle) and the achlorophyllous, highly modified underground leaf (eel trap, below) of *G. hispidula* (section *Africanae*). Both species are grown in the greenhouse of IPK. Plants and flowers of *Genlisea* were photographed by H. Ernst (IPK). The trap photo is taken from: <http://forum.carnivoren.org/index.php?/topic/35704-begleitpflanzen-cephalotus-heliamphora-nepenthes-und-co/>). (C) Phylogenetic tree based on sequence comparison of three chloroplast loci of the genus *Genlisea* (adapted from Fleischmann *et al.*, 2014). Not all of 29 so far described species are included. Bars represent 0.5 cm.

In combination with the geographic separation of extant *Genlisea* taxa (no species occurs in both continents) and the estimated age of the order Lamiales, the phylogenetic clustering suggested a bidirectional trans-Atlantic dispersion with two subsequent colonization events for *Genlisea* (Fleischmann *et al.*, 2010). From South America where the genus *Genlisea* as well as its sister genus *Utricularia* originated (Jobson *et al.*, 2003), the ancestor of the subgenus *Genlisea* migrated to Africa (the first colonization). After the radiation leading to the section *Africanae*, an ancestral member of the clade *Recurvatae* and *Genlisea* re-colonized South America and gave rise to the section *Genlisea* (the second colonization).

1.2.2. Unique morphological and molecular features of *Genlisea* species

The unique trapping mechanism is one of the distinct characteristics of the genus *Genlisea*. Since the speculation of Darwin about the carnivorous nature of *Genlisea* (Lloyd, 1942), the unique eel-trap became a fascinating characteristic of this genus. The complex architecture of the *Genlisea* trap was studied in detail (for review see Fleischmann, 2012). *Genlisea* does not have true roots, instead root-like organs appear that represent fragile, subterranean, leave-derived traps (so-called rhizophylls). In general the achlorophyllous rhizophyll connects to the rosette by a footstalk which extends and attaches to the vesicle, a “lobster pot”-like chamber where the prey is digested and released nutrients are absorbed. The vesicle is linked downstream with the tubular neck with successive rows of inwardly pointing hairs to guide prey toward the vesicle. The tubular neck ends by trap mouth at branching zone from which two trap arms, helically twisting in opposite directions, elongate (the corkscrew). These two arms and trap mouth are the entrances of prey (for review see Reut, 1993; Plachno *et al.*, 2007; Fleischmann, 2012). Barthlott *et al.* (1998) speculated that *Genlisea* attract, trap and digest protozoa in the vesicles of its subterranean leaves. The organic compounds derived from the prey marked with the isotope sulphur-35 were traced in the rosette leaves after two days of “feeding” indicating that *Genlisea* is able to absorb nutrient released by prey digestion. Plachno *et al.* (2005) and Darnowski and Fritz (2010) provided further hints for the ability of *Genlisea* to passively and chemotactically attract a wide range of prey, from protozoa to small crustaceans. A broader spectrum of prey was inferred from the transcriptome of traps of *G. nigrocaulis* and *G. hispidula* (Cao *et al.*, 2015)

Another peculiar feature of *Genlisea* is the exceptionally high DNA substitution rate. Comparing the chloroplast gene *rbcl*, the mitochondrial gene *coxI* and the nuclear 5.8S rDNA gene, the relative mutation frequency of *Genlisea* and *Utricularia* lineages is four to 14 times higher than that of *Pinguicula* (Jobson & Albert, 2002). Moreover, in comparison with about 300 angiosperm genera representing 200 families, *Genlisea* together with *Utricularia* displayed the highest nucleotide

substitution rate of the *matK* gene (Mueller *et al.*, 2003). Such high mutation frequencies were speculated to facilitate the carnivorous specialization during evolution of *Genlisea* and *Utricularia* (Mueller *et al.*, 2006).

The scientific interest in the genus *Genlisea* increased rapidly since the striking discovery of new ultra-small genomes (less than 100 Mbp/1C), supporting the presumed minimum size for genome of a free-living angiosperm (~50 Mbp/1C; Bennett & Leitch, 2005), was reported by Greilhuber *et al.* (2006). These authors claimed that *G. margaretae* (63.4 Mbp/1C) and *G. aurea* (63.6 Mbp/1C) possess the smallest nuclear genome sizes among flowering plants, less than half of that of *A. thaliana* (157 Mbp/1C; Bennett *et al.*, 2003). Another species, *G. hispidula* (1,510 Mbp/1C), was found having an about 24-times bigger genome. More recently, the genomes of *G. tuberosa* (61 Mbp/1C) and *G. lobata* (1,722.4 Mbp/1C) were considered as the smallest and largest genomes, respectively, not only for the genus but also for family Lentibulariaceae (Fleischmann *et al.*, 2014) (Table 1).

Table 1: Genome sizes and chromosome numbers of some *Genlisea* species.

Genome sizes were measured either by Feulgen densitometry (a-Greilhuber *et al.*, 2006) or by flow-cytometry (b-Veleba *et al.*, 2014; c-Fleischmann *et al.*, 2014). The colors marked on the left indicate sections/subgenus according to the phylogenetic tree in figure 3C.

	Species	Genome size (Mbp/1C)	Chromosome number (2n)
Sec. <i>Genlisea</i>	<i>G. oxycentron</i>	74.6 ^c	
	<i>G. nigrocaulis</i>	73 – 80 ^b	
	<i>G. repens</i>	77 ^b , 78 – 149.7 ^c	
	<i>G. pygmaea</i>	161 ^b	
	<i>G. tuberosa</i>	61 - 65 ^c	
	<i>G. aurea</i> var. <i>aurea</i>	63.6 ^a – 83 ^c	~52 ^a
	<i>G. aurea</i> var. <i>minor</i>	117 ^c - 131 ^b	
	<i>G. guianensis</i>	298.1 ^c	~40 ^c
Sec. <i>Recurvatae</i>	<i>G. glandulosissima</i>	169 ^b , 189.3 ^c	~38 ^c
	<i>G. margaretae</i>	63.4 ^a , 168 ^b , 113 – 195 ^c	36 or 38 ^c
Sec. <i>Africanae</i>	<i>G. hispidula</i>	1417 ^b , 1510 ^a	32 ^c
	<i>G. subglabra</i>	1471 ^b	32 ^c
Subgen. <i>Tayloria</i>	<i>G. lobata</i>	1200 ^b , 1277 ^a , 1722.4 ^c	16 ^a
	<i>G. violacea</i>	460 ^b , 1005 ^a , 1609 ^c	16 ^c
	<i>G. metalica</i>	1056 ^c	16 ^c
	<i>G. flexuosa</i>	1121 ^b , 1140.3 ^c	16 ^c
	<i>G. uncinata</i>	995 ^a	16 ^a

In phylogenetic context, both the ultra-small and the ~24-fold larger genomes are exclusively found within the subgenus *Genlisea* (sections *Genlisea* and *Africanae*, respectively), but not in the subgenus *Tayloria*. Additionally, polyploidy was assumed to occur within *Genlisea* species on the basis of the nuclear DNA content described for *G. aurea* (Albert *et al.*, 2010; Veleba *et al.*, 2014) and *G. repens* (Fleischmann *et al.*, 2014). Most species of *Pinguicula* possess genome sizes ranging from 400 to 800 Mbp/1C, while most *Utricularia* species possess genomes of less than 400 Mbp/1C. Particularly, *U. gibba* (88.3 Mbp/1C) and *U. purpurea* (79 Mbp/1C) possess ultra-small genomes (Greilhuber *et al.*, 2006; Veleba *et al.*, 2014). Thus, high genome size plasticity seems to be another representative feature of the Lentibulariaceae family, especially of the genus *Genlisea*, a promising subject to investigate the mechanisms behind genome evolution in plants.

1.3. Genome size and karyotype evolution in Lentibulariaceae

In order to investigate the genome size plasticity within Lentibulariaceae and mechanisms behind, several studies were performed recently based on either genome size data in a phylogenetic context or on whole genome sequencing data.

Considering genome size information in a phylogenetic context, Veleba *et al.* (2014) proposed a general model of genome evolution within the Lentibulariaceae family. The authors observed different genome size evolution patterns in each of three genera. While *Pinguicula* genomes showed a slight but consistent tendency for expansion, those of the evolutionary younger sister genera *Utricularia* and *Genlisea* were found remarkably miniaturized (in all three clades of the former and in the two most derived sections of the latter). Ultra-small genomes, more than four-fold smaller than that estimated for the common ancestor of the family (~400 Mbp/1C), occurred in the section *Genlisea*. Moreover a drastic genome enlargement occurred in the basal clades of the subgenus *Tayloria* and the section *Africanae* with some members possessing the largest genomes in the whole family. In another study, Fleischmann *et al.* (2014) combined genome size data with chromosome numbers and hypothesized a karyotype evolution model. According to that, after divergence from the common ancestor shared by the subgenus *Tayloria* with a basic chromosome number $x = 8$, a tetraploidization was speculated for the subgenus *Genlisea* which resulted in larger genome size and chromosome number in members of the section *Africanae*. Subsequently dysploid chromosome rearrangement was assumed for the younger sections *Recurvatae* and *Genlisea* accompanied by genome shrinkage.

To clarify the molecular mechanisms behind the ultra-small genomes of Lentibulariaceae, whole genome sequencing study was performed for *Utricularia gibba* (~80 Mbp/1C; Ibarra-Laclette

et al., 2013) and *G. aurea* (63 Mbp/1C; Leushkin *et al.*, 2013). The reduction of the *U. gibba* genome was presumed to be driven by strong deletion bias apparently removing most of LTR-retrotransposons and other redundant DNA after at least three rounds of WGD found in the *U. gibba* genome since its divergence from the common ancestor of tomato and grape. This assumption was supported by the extremely low proportion of (often truncated) TEs, the shorter and fewer introns, in comparison to that of *A. thaliana*, per gene (all as results of numerous microdeletions) and by the presence of most of LTR retrotransposons as solo elements (results of ectopic recombination) (Ibarra-Laclette *et al.*, 2013). On other hand, for the miniature *G. aurea* genome, about 21,000 genes and gene fragments were claimed. Compared to *Mimulus guttatus*, *G. aurea* displayed shorter introns and intergenic regions due to shrinkage involving non-coding sequences. Nevertheless, intron loss did not seem to occur in *G. aurea* since the observed intron number per gene is typical for angiosperms (Leushkin *et al.*, 2013). Although both studies provide some insights into genome shrinkages of Lentibulariaceae, data for congeneric species with large genome size difference are lacking to further specify the highly dynamic genome evolution in this family, especially within the genus *Genlisea*. Based on a comparative whole-genome study Vu *et al.* (2015) supposed a bidirectional genome size evolution which led to more than 18-fold difference in nuclear DNA content between *G. nigrocaulis* and *G. hispidula*. Whereas the smaller genome of *G. nigrocaulis* was characterized by a very low proportion of TE (9.7 % genome) and usually short introns in 15,550 predicted 'high confidence' gene, the larger genome of *G. hispidula* apparently resulted from WGD (presumed by *i*) gene number, *ii*) SNP-based allele frequency and *iii*) homologous gene pair comparison between two species) and retrotransposon proliferation (up to 41.6% of genome). Because of the similar morphology as well as natural habitats of species of the *G. nigrocaulis* and the *G. hispidula* lineages, divergent genome size of these species was assumed to be a selectively neutral feature during evolution of these lineages (Vu *et al.*, 2015).

Taken together, the current genome analyses shed some light on molecular mechanisms behind the genome size divergence in Lentibulariaceae by either genome expansion or shrinkage in closely related *Genlisea* species. Up to now, the cytological data available for Lentibulariaceae are restricted to counting or estimation of the chromosome number. The obtained genomic data, thus, provided the useful basis for the hitherto lacking molecular cytogenetic and comparative analyses aiming to elucidate the karyotype evolution within the genus *Genlisea*.

1.4. Aims of the dissertation

This dissertation was directed toward the establishment of a cytological basis for studies of genome and karyotype evolution within the carnivorous genus *Genlisea* utilizing sequence data from the ultra-small genome of *G. nigrocaulis* and the 18-fold larger one of *G. hispidula* (Vu *et al.*, 2015). There were three main tasks to be focused on:

First, we measured for species of the subgenus *Genlisea* the nuclear DNA content, characterized the nuclear phenotype regarding to DNA and histone H3 methylation, counted the chromosome number, and investigated the chromosomal distribution of ribosomal DNA clusters. The obtained data served as a prerequisite for further cytological analysis in this subgenus.

Second, using the genomic sequence data, we focused on chromosomal distribution of repetitive sequences characterized for *G. nigrocaulis* and *G. hispidula*. The specific chromosomal location of putative centromeric sequences and telomeric repeat variants was also examined for these two species and some close relatives.

Third, employing either single-copy sequences, tandem repeat sequences and repetitive-free BAC clones, we karyotyped three species as representatives for three sections of the subgenus *Genlisea*.

2. MATERIALS AND METHODS

2.1. Plant material and cultivation

Plants of studied species were obtained from different sources shown in table 2 and cultivated in a greenhouse under conditions recommended for *Genlisea*: pots were submerged in containers filled with rainwater and kept at 20 - 25°C under normal light condition (Fleischmann, 2012).

Table 2: List of *Genlisea* species used in this study

(n.a.: not available, CZ: Czech Republic)

Voucher	Species	Cultivated	Origin
n.a.	<i>Genlisea aurea</i> A. St.-Hil.	n.a.	BestCarnivorousPlants K. Pasek, Ostrava-Poruba, CZ
GAT 7857	<i>Genlisea hispidula</i> Stapf	Gatersleben (green house)	LE 294, A. Fleischmann, Ludwig- Maximilians-Universität München
GAT 7858	<i>Genlisea hispidula</i> Stapf	Gatersleben (green house)	BestCarnivorousPlants K. Pasek, Ostrava-Poruba, CZ
GAT 7859	<i>Genlisea hispidula</i> Stapf	Gatersleben (green house)	BestCarnivorousPlants K. Pasek, Ostrava-Poruba, CZ
n.a.	<i>Genlisea margaretae</i> Hutch	Gatersleben (green house)	BestCarnivorousPlants K. Pasek, Ostrava-Poruba, CZ
GAT 7444	<i>Genlisea nigrocaulis</i> Steyerm.	Gatersleben (green house)	Gartenbau Th. Carow, Nüdlingen; Germany
GAT 7445	<i>Genlisea nigrocaulis</i> Steyerm.	Gatersleben (green house)	Carnivores and more Chr. Klein, Merzig, Germany
GAT 23586	<i>Genlisea pygmaea</i> A.St.-Hil.	Gatersleben (green house)	BestCarnivorousPlants K. Pasek, Ostrava-Poruba, CZ
n.a.	<i>Genlisea subglabra</i> Stapf	Gatersleben (green house)	Carnivores and more Chr. Klein, Merzig, Germany

2.2. Genomic DNA isolation and cytological preparations

Genomic DNA of studied species was isolated using the DNeasy® Plant Mini kit (Qiagen). For each sample, about 100 mg of fresh and healthy leaves were harvested, cleaned in distilled water, frozen in liquid nitrogen and disrupted using a TissueLyser II (Qiagen). The DNA isolation afterward was performed according to manufacturer's instruction. Concentration and quality of the DNA were estimated using a NanoDrop spectrophotometer (Thermo Scientific) and by 1% (w/v) agarose-gel electrophoresis.

For isolation of interphase nuclei, about 20 mg freshly collected *Genlisea* leaves were fixed in 4% Formaldehyde in Tris buffer [100 mM Tris-HCl, 5 mM MgCl₂, 85 mM NaCl, 0.1% (v/v) Triton X-100,

pH 7.0] (RT, 20 min). After washing in Tris buffer (2x5 min), the leaves were chopped by a clean razor blade in 1 ml nucleus isolation buffer [15 mM Tris-HCl, 2 mM EDTA, 80 mM KCl, 20 mM NaCl, 0.5 mM Spermine, 0.1% (v/v) β -mercaptoethanol, 0.1% (v/v) Triton X-100, pH 7.5] to get a homogeneous suspension before filtering through a 35 μ m mesh. The nuclei suspension was stained with DAPI (1 μ g/ml), flow-sorted using a FACSAria instrument (BD Bioscience). About 4,000 flow-sorted nuclei were dropped together with 10 μ l sucrose buffer [100 mM Tris, 50 mM KCl, 2 mM MgCl₂, 0.05% (v/v) Tween 20, 5% (w/v) sucrose] onto a microscopic slide. After overnight air-drying, the sample region was marked on slide and slides were stored at -20°C until use.

The extended chromatin fibers were prepared according to Li *et al.* (2005). About 50 mg of fresh leaves were used to isolate interphase nuclei as described above (without Formaldehyde fixation). Five μ L of nuclei suspension were dropped on one end of a microscopic slide. After incubation with 20 μ l nucleus lysis buffer [0.5% (w/v) SDS, 5 mM EDTA, 100 mM Tris, pH 7.0] (RT, 10 min), a cleaned coverslip was used to drag the nuclei droplet along the slide to stretch interphase chromatin. The slides were then fixed in ethanol: glacial acetic acid (3:1) (RT, 2 min) before baked at 60°C for 30 min and stored at -20°C until use.

Genlisea chromosome preparations for FISH were performed according to Lysak *et al.* (2006b) with some modifications. In brief, plant material was treated in 20 mM 8-hydroxyquinoline at RT for 2 h and then at 4°C for 4 h to arrest mitosis and to accumulate dividing cells. After a short wash in distilled water, the material was fixed in ethanol: glacial acetic acid (3:1) at RT for at least 24 h. Fixed tissue was used immediately or stored at 4°C for several days. After three washes in citrate buffer [10 mM Na-citrate, pH 4.5] for 5 min each, the plant tissue was softened in 2 ml PC enzyme mixture [2% (w/v) pectinase, 2% (w/v) cellulase in citrate buffer] at 37°C, for ~10 min (young leaves) or ~15 min (flower buds). Combination of enzyme mixture and digestion time was adjusted empirically to get optimal results for the different types of tissue of different species. The digestion was stopped by adding ice-cold citrate buffer. Chromosome spreading was performed by dispersing and flattening the enzymatically softened material between a microscopic slide and a coverslip in a drop of 75% glacial acetic acid (squashing method). Alternatively, 5 μ L of cell suspension made from the digested material in 75% glacial acetic acid were dropped on a slide and squashed under a coverslip to obtain spread metaphase chromosomes (drop-squashing method) (Schwarzacher & Leitch, 1994). After freezing in liquid nitrogen, slides were rinsed in 2 \times SSC [300 mM Na-citrate, 30 mM NaCl, pH 7.0], dehydrated in an ethanol series (70, 90 and 96%) and air-dried. The quality of spreading was evaluated under an epifluorescence microscope after DAPI staining (10 μ g/ml). Slides harboring more than 10 well-spread metaphases were stored at 4°C until use for FISH.

2.3. Genome size measurement

Genome size measurements were performed according to Fuchs *et al.* (2008) using either a FACStar^{PLUS} or a FACSaria IIu flow sorter (BD Biosciences). For *G. aurea*, *G. margaretae*, *G. nigrocaulis*, and *G. pygmaea*, *Arabidopsis thaliana*, ecotype 'Columbia' (2C = 0.32 pg; Bennett *et al.*, 2003) and for *G. hispidula*, *G. subglabra*, *Raphanus sativus* 'Vorán' (IPK gene bank accession number RA 34; 2C = 1.11 pg; Schmidt-Lebuhn *et al.*, 2010), was used as internal reference standard. The absolute DNA contents were calculated based on the values of the G1 peak means and genome sizes were estimated based on the calculation proposed by Dolezel *et al.* (2003). This part was done by Dr. J. Fuchs.

Categories of genome size were defined as very small (less than 1,300 Mbp), small (less than 3,400 Mbp), medium (between 3,400 and 13,000 Mbp), large (more than 13,000 Mbp) and very large (more than 34,000 Mbp) (Leitch *et al.*, 2005). In this study, to differentiate genome size classes of Lentibulariaceae species, the term "ultra-small" was specified for nuclear DNA amount less than 100 Mbp/1C, "small" and "medium" for those less than 400 and 800 Mbp/1C, and "large" for that more than 1,000 Mbp/1C.

2.4. BAC library construction for *G. margaretae* and colony membrane preparation

The BAC library for *G. margaretae* was constructed by Dr. H. Simkova using flow-sorted nuclei as source for high molecular weight DNA isolation (Simkova *et al.*, 2003). The *Hind*III- partially digested genomic DNA fragments were cloned into the dephosphorylated pIndigoBAC-5 vector (Epicentre) according to Simkova *et al.* (2011). The BAC library was arranged into 384-well microtiter plates filled with freezing media [1% (w/v) *bacto-tryptone*, 0.5% (w/v) *bacto-yeast extract*, 1% (w/v) *NaCl*, 36 mM *K₂HPO₄*, 13.2 mM *KH₂PO₄*, 1.7 mM *Na-citrate*, 6.8 mM *(NH₄)₂SO₄*, 4.4% (w/v) *glycerol*, 0.4 mM *MgSO₄*, 12.5 µg/ml *chloramphenicol*, pH 7.5] using a Q-bot apparatus (Genetix).

Using a replicator, the BAC clones of each 384-well plate were stamped on a Hybond N+ membrane (GE Healthcare) laid on a LB agar plate [1% (w/v) *actotryptone*, 0.5% (w/v) *bacto yeast extract*, 1% (w/v) *NaCl*, 1.5% (w/v) *agar*, 12.5 µg/ml *chloramphenicol*, pH 7]. After incubation at 37°C for 16-18 h, the membrane was detached from the LB agar plate. The on-membrane isolation and fixation of BAC DNA were performed according to the following steps using blotting papers wetted with corresponding solutions: [10% (w/v) *SDS*] for 10 sec; air-drying; [0.5 M *NaOH*, 1.5 M *NaCl*] for 3 min; air-drying; [1.5 M *NaCl*, 0.5 M *Tris-HCl*, pH 7.4] for 5 min; UV-crosslinking (automatic mode, 125 kJ); [2× *SSC*] for 5 min and drying at 80°C for 30 min. The membranes of the whole library were prepared by I. Walde; I was partly involved in this work.

2.5. Screening for repeat-free BAC clones of *G. margaretae* using dot-blot DNA hybridization

The *G. margaretae* BAC library spotted on membranes was screened for colonies harboring repeat-free inserts by DNA hybridization. For probe preparation, genomic DNA of *G. margaretae* was fragmented to a size of 0.5-2 kbp by a Bioruptor® Standard sonication device (Diagenode). A 167 bp fragment amplified from the pIndigo BAC-5 vector by the primer combination LACZB/T7A (Table 3) was used as control probe for signal normalization of the different BAC clones. Prior to labeling, DNA probes were denatured (95°C for 10 min) and briefly chilled on ice. To generate 25 µl of randomly labeled radioactive probes, about 50 ng DNA was added into a 1.5 ml screw-cap tube and mixed with labeling mixture [5 mM Tris-HCl (pH 8.0), 5 mM MgCl₂, 0.072% (v/v) β-mercaptoethanol, 200 mM HEPES (pH 6.6), 150 µg/ml random hexadeoxy ribonucleotides, 20 µM of each unlabeled d(A-T-G)TP], 3 µl [α-³²P]-dCTP (10 µCi/µl) and 1 unit of Klenow fragment. The reaction was carried out at 37°C for 30 min and then the labeled probe was denatured at 95°C for 10 min and chilled on ice.

For DNA hybridization, the treated membranes were pre-hybridized with 30 ml hybridization buffer A [5× SSPE (900 mM NaCl, 50 mM sodium phosphate, 5 mM EDTA pH 7.7), 5× Denhardt's solution (0.1% (w/v) Ficoll® 400, 0.1% (w/v) PVP, 0.1% (w/v) BSA), 0.5% (w/v) SDS] in hybridization tubes (65°C for 2-4 h). After exchanging the buffer by 15 ml hybridization buffer B [0.5 M sodium phosphate pH 7.2, 7% (w/v) SDS, 1 mM EDTA pH 8.0], the radioactively labeled probe was added and the hybridization was carried out at 60°C overnight. The membranes were then washed to remove the unbound probes and excessive radio-labeled nucleotides as follows: [4× SSPE, 0.5% (w/v) SDS] at RT for 5 min; [2× SSPE, 0.5% (w/v) SDS] at RT for 5 min; [2× SSPE, 0.1% (w/v) SDS] at 60°C for 30 min; [0.5× SSPE, 0.1% (w/v) SDS] at 60°C for 30 min and finally 2× SSC at RT for 5 min. After that, the membranes were exposed to X-ray film to develop hybridization signal facilitating the selection of putatively repeat-free BAC clones. This part was done by Dr. R. Schmidt.

2.6. In-silico analysis for the identification of repetitive elements and single-copy sequences

From whole genome sequencing data of *G. nigrocaulis* and *G. hispidula* (Vu *et al.*, 2015), repetitive elements were identified using similarity-based clustering of unassembled sequence reads (Novak *et al.*, 2010) and further characterized using the RepeatExplorer pipeline (Novak *et al.*, 2013). The consensus sequences of identified repetitive elements were used to design primers for probe generation by Dr. J. Macas, Dr. P. Neumann and Dr. P. Novak (Table 3).

The *de novo* whole genome shotgun assembly and scaffolding were performed for *G. nigrocaulis* (Vu *et al.*, 2015). The consensus sequences of these “pseudo-molecules” were screened by Dr. H. Cao for single-copy regions as potential sources for single-copy FISH probes (Table 3).

Table 3: List of primers used to generate probes.

Gn: *G. nigrocaulis*; Gh: *G. hispidula*; Ath: *A. thaliana*; TEL: telomere; Ta: annealing temperature; F: forward primer; R: reverse primer.

Primer	Primer sequence (5'-3')	Ta (°C)	Product
LACZB	CAACTGTTGGGAAGGGCGATC	60	pIndigoBAC-5 vector-specific fragment
T7A	CCAATTCGCCCTATAGTGAG	60	fragment
Gn_v4c12_p4	F: TGAGTGGTCAAAGAAGACAGGAAG R: ATTTCCGTTAGCGTAGATTCAAGC	67	Single-copy fragment of 8.3 kbp
Gn_v4c202_p4	F: TTCGATTCTGGATGATAATTGACTG R: AGTTCAAGCTTCGACGAGTATGTG	69	Single-copy fragment of 9.2 kbp
Gn_v4s2_p6	F: GCCGAAGCGTCATTTACTCACTAC R: CAATCCTCTCCAACGCATCTCTTAC	67	Single-copy fragment of 10.7 kbp
Gn_v4s15_p4	F: GGTATAATTACGGAAGTCGATCC R: GAAACCTGTTTCGAGAAATCACT	66	Single-copy fragment of 10.4 kbp
Gn_v4s17_p2	F: ACTCAATCCGGTTCCTGTAAGTTC R: AGTTCATCCTCTGATGGCCTTAAC	65	Single-copy fragment of 10.3 kbp
Gn_v4s19_p2	F: CCCAGATGAGAGCAATTTGTATTG R: AACGCATTTATAGATGAGGATTG	65	Single-copy fragment of 8.5 kbp
Gn_v4s56_p44	F: CGTCTGTAGAATTTGAGCAGCGAG R: GCTACTTCATTTGCGGGTGGATAAG	66	Single-copy fragment of 10.5 kbp
Gn_v4.2s58_4	F: AGTGATGGAAGTGACTCCAGTGAG R: TAATTCGCTCTCTTGCTGCATAC	69	Single-copy fragment of 9.3 kbp
Gn_v4s130_z1	F: TACGCTCTGCATTGGGAGTC R: TACGGAAACACCGAACACAA	65	Single-copy fragment of 9.1 kbp
Gn_v4s196_b1	F: GCAGAGCAAAATCCGAAAC R: GGCTTCGGCTAATGGACTTG	68	Single-copy fragment of 8.3 kbp
Gn7c161	F: GCCTTATTATGCATCAAATAGCTTC R: GCAATTGGATCCTTTAATAACCTC	55	Tandem repeat of 161 bp motif
Gn44c19	F: TTTATTATTCAGTGTCGGAATGAC R: AATATACGTCATGGAATCAAGATAATG	55	Tandem repeat of 144 bp motif
Gn10c83	F: GTATATATGTACCGCTTGTGCTCAG R: AACTATATCGTTCAGGCATATGAAAC	55	<i>Bianca</i> -like <i>Ty1/copia</i> element (1,730 bp)
Gh14c16	F: ATAAACTGATTTCTACCCACCA R: ATGAGTTCTTACACTGATTTCTACCTG	55	Tandem repeat of 60 bp motif
Gh250c46	F: GAGCTCGTTCCTGATCAGTCC R: ACTGGAAGAATCTTCCGATCTC	55	Tandem repeat of 74 bp motif
Gh45c31	F: TCGAAGAGATCGGATAGATAGAATC R: GTTTGTTTCAGTTCAACATTTGAGG	55	Tandem repeat of 110 bp motif
Gh80c174	F: TTGAGCTCGATCAGTTTCACC R: GAGATCAAATAGATTGAATCATCCAG	55	Tandem repeat of 112 bp motif

Primer	Primer sequence (5'-3')	Ta (°C)	Product
Gh336c35	F: ACCACGTGGCCGATCTGT R: AGATCGGTCTAGGCGTGGAA	55	Tandem repeat of 146 bp motif
Gh19c56	F: GTTTTCGGTAAGTAATCCAATG R: TGCAATAATTCGACTACGAAATCAC	55	Tandem repeat of 900 bp motif
Gh338c4	F: GAATCTTCATCTTCTTGATTCYCTTG R: CATCTCACTTGAAGGTTTAGGAGCA	64	<i>Ty3/gypsy</i> retroelement (8 kbp)
GhCR01	F: GCTTTGACAGAACCCCTTAATC R: AGCTTCCTTGGTCTCCACTT	66	Putative centromeric retrotransposon element (4.8 kbp)
GhCR02	F: TGTCCTCAAGTAAAGATAAAGAAAAA R: TCCACCTCAAATGAGTTCATAA	64	Putative centromeric retrotransposon element (4.8 kbp)
GhCR03	F: TCTAGATAGCAAGATGATCCTTGAG R: TTAGGCTCCTCGAATGTGAT	68	Putative centromeric retrotransposon element (4.8 kbp)
GhCR04	F: ATGTCCAAGGATAAGGTGATTG R: CTTGAGTGTTCTTGCTTGAT	68	Putative centromeric retrotransposon element (5.2 kbp)
Atha_TEL	F: TTTAGGTTTAGGGTTTAGGGTTTAGGG R: CCCTAAACCCTAAACCCTAAACCCTAAA	55	Arabidopsis-type telomere
Gh_TEL6	F: TTCAGGTTTCAGGTTTCAGGTTTCAGG R: CCTGAACCTGAACCTGAACCTGAA	55	Short telomeric variant
Gh_TEL7	F: TTTTCAGGTTTCAGGTTTCAGGTTTCAGG R: CCTGAAACCTGAAACCTGAAACCTGAAA	55	Long telomeric variant
5S rDNA	F: GTGCGATCATACCAGCRKTAATRCACCGG R: GAGGTGCAACACGAGGACTTCCCAGGRGG	55	<i>Genlisea</i> -specific 5S rDNA

2.7. Immunostaining

Immunostaining experiments were performed using either flow-sorted nuclei or squashed chromosome preparations. Slides with flow-sorted nuclei stored at -20°C were baked at 60°C for 15 min, rinsed in 1x PBS [137 mM NaCl, 2.7 mM KCl, 11.9 mM phosphates, pH 7.3-7.5] (RT, 2x5 min) and incubated in 4% (w/v) Paraformaldehyde in 1x PBS (RT, 20 min) for fixation. Slides were afterward washed in 1x PBS (RT, 3x5 min) and subjected to the blocking step (see below). Mitotic chromosome preparation for immunostaining was performed as described for FISH with some modifications. For all washing steps, 1x PBS was used. The material was fixed in ice-cold 4% (w/v) Paraformaldehyde in 1x PBS (20 min) and softened in 2 ml PCPM enzyme mixture [2.5% (w/v) of each *Pectinase*, *Cellulase onozuka*, *Pectolyase* and *Macerozyme* in 1x PBS] (37°C, 45 min). Squashing was performed in a drop of 0.5% (v/v) Triton X-100 in 1x PBS. After freezing in liquid nitrogen slides were kept in ice-cold 1x PBS and used for immunostaining experiment at the day of preparation.

To prevent unspecific immunobinding, slides were incubated in blocking buffer [8% (w/v) BSA, 0.1% (v/v) Triton X-100 in 1x PBS] at 37°C for 1 h, shortly washed in 1x PBS and incubated with a primary antibody at 4°C for 16 h in a humid chamber. All used primary antibodies were diluted in

antibody buffer [1% (w/v) BSA in 1× PBS] at dilution ratios listed in table 4. Subsequently, slides were washed in 1× PBS (RT, 3x10 min) and then incubated with Cy3-conjugated anti-rabbit or Alexa 488-conjugated anti-mouse antibody (Life Technologies, dilution 1:200 in antibody buffer) in a humid chamber at 37°C for 1 h. After final washes in 1× PBS (RT, 3x10 min), slides were blotted for excessive buffer and counter-stained with DAPI (10 µg/ml) in VectaShield (Vector Laboratories).

Table 4: List of primary antibodies

Antibody	Host	Company/Cat-number	Dilution
5-methylcytosine	mouse	Eurogentec (MMS-900P-A)	1:100
H3K4me1	rabbit	Upstate (07-436)	1:400
H3K4me2	rabbit	Upstate (07-030)	1:200
H3K4me3	rabbit	Upstate (07-473)	1:400
H3K9me1	rabbit	Upstate (07-395)	1:400
H3K9me2	rabbit	Upstate (07-441)	1:100
H3K9me3	rabbit	Upstate (07-523)	1:100
H3K27me1	rabbit	Upstate (07-448)	1:500
H3K27me2	rabbit	Upstate (07-452)	1:200
H3K27me3	rabbit	Upstate (07-449)	1:500
H3S10ph	mouse	Upstate (06-570)	1:50
H2AThr121ph	rabbit	MyBioSource (MBS004447)	1:700

2.8. FISH probe generation and labeling

Beside cytological preparation, probe generation and labeling is the crucial step that critically influences the outcome of a FISH experiment. In this study, depending on the nature of target sequences, probes were labeled by either nick-translation or PCR using fluorochrome (direct labeling) or biotin/digoxigenine (indirect labeling) modified dUTP.

All repetitive DNA probes were PCR amplified from genomic DNA of either *G. nigrocaulis* or *G. hispidula* using a GoTag Kit (Promega) or, for fragments larger than 4 kbp (TEs), a Phusion High Fidelity DNA Polymerase Kit (Thermo Scientific). PCR products were analyzed by 1% (w/v) agarose-gel electrophoresis and cloned by pGEM-T Easy Vector Systems (Promega) according to the protocol of the manufacturer. Plasmids with fragments of interest were isolated using a QIAprep Spin Miniprep Kit (Qiagen) and confirmed by Sanger sequencing.

The *A. thaliana* BAC clone T15P10 (Arabidopsis Biological Resource Center, USA) was used as 45S rDNA probe. The 5S rDNA gene was PCR-amplified from genomic DNA of *G. nigrocaulis* using a degenerate primer pair designed by Dr. H. Cao (Table 3). The amplicon of expected size (117 bp) was extracted from 0.8% (w/v) agarose-gel using a QIAquick Gel Extraction Kit (Qiagen), cloned by pGEM-T Easy Vector Systems (Promega) and confirmed by Sanger sequencing. The cloned 5S rDNA gene was used as template for probe generation. Telomere-specific probes were generated by PCR in the absence of template DNA according to Ijdo *et al.* (1991) using tetramers of the corresponding telomere repeats (Table 3).

For single-copy probes, the uniqueness of identified single-copy fragments was confirmed by PCR amplification using a Phusion High-Fidelity DNA Polymerase kit followed by FISH analysis on flow-sorted nuclei. For each primer pair the optimal annealing temperature was determined by a temperature gradient test (Table 3). All single-copy fragments that could be specifically amplified (as single band with the expected size) were used to generate FISH probes.

For BAC probes, BAC DNA was isolated according to Farrar and Donnison (2007). The insert size of putatively repeat-free BAC clones selected from dot-blot DNA hybridization was analyzed by pulsed-field gel electrophoresis (PFGE) using a CHEF-DR II apparatus (Bio-Rad) under the following condition: 1% (w/v) agarose-gel in 0.5× TAE, 6 V/cm, ramping time 5-25 sec, 14°C for 14 h.

In order to increase the labeling efficiency for the short 5S rDNA gene (117 bp), this FISH probe was labeled during the PCR amplification in the presence of biotin-dUTP or digoxigenine-dUTP (Roche) using a GoTaq Kit (Promega). About 0.1 ng of plasmid harboring 5S rDNA gene as DNA template were mixed with 10 µl 5× Colorless PCR buffer; 2.5 µl of 2 mM d(AGC)TP mixture; 2.5 µl of 1 mM labeled-dUTP; 10 pmol of each forward and reverse primer (Table 3) and 5 units GoTaq Polymerase in 50 µl PCR volume. The following amplification program was used: 95°C/2 min; 30 cycles of 95°C/45 sec, 55°C/30 sec, 72°C/30 sec; final extension 72°C/5 min.

Other FISH probes were labeled by nick translation. For a 50 µl of nick-translation volume, about 1 µg of probe DNA and 5 µl of each 10× nick translation buffer [0.1 M MgSO₄, 1 mM dithiothreitol, 500 µg/ml BSA in 0.5 M Tris-Cl (pH 7.2)], 0.1 M mercaptoethanol, 2 mM d(AGC)TP mixture were added into a 0.5 ml tube. For direct labeling, 2 µl of 1 mM Cy3-dUTP (Amersham) or either 0.8 µl of 1 mM TexasRed-dUTP or 1 mM Alexa 488-dUTP (Life Technology) were added. For indirect labeling, 2 µl of either 1 mM biotin-dUTP or 1 mM digoxigenine-dUTP (Roche) were added. After adding 3 µl DNase I [4 µg/ml in 0.15 M NaCl/50% (w/v) glycerol] and 10 units DNA polymerase I (Fermentas) the tube was gently mixed and incubated at 15°C for 90-120 min until the size of

fragments reached 200~500 bp, controlled by 1% (w/v) agarose-gel electrophoresis. The enzymatic activity was inactivated by incubation at 65°C for 10 min. The labeled probe was then stored at -20°C.

For FISH, 3-5 µl (about 60-100 ng for repetitive DNA or BAC clone) or 8-10 µL (about 160-200 ng for single-copy sequence) of labeled probe were used per slide. Based on experiment design, single or pooled probes were prepared. Labeled probe(s) was transferred into a 0.5 ml tube per slide. To precipitate probe, 0.1 volume of 3 M Sodium acetate and 2.5 volume of ice-cold 96% ethanol were added and the tube was briefly mixed and centrifuged. After an overnight incubation at -20°C, the tube was centrifuged at 14,000 rpm/4°C for 30 min. Probe pellet was washed with ice-cold 70% ethanol, air-dried for 5 min and dissolved in 20 µl hybridization buffer [50% (v/v) Formamide, 10% (w/v) Dextran sulfate, 50 mM Sodium phosphate in 2× SSC, pH 7] at 37°C for at least 1 hour. The ready-to-use FISH probes were stored at -20°C.

2.9. Fluorescence in situ hybridization (FISH)

FISH was performed according to Lysak *et al.* (2006b) with some modifications. Prior to hybridization, preparations were rinsed in 2× SSC (2x5 min), treated with pepsin [100 µg/ml in 0.01 N HCl] at 37°C for 2 or 7-10 min (for flow-sorted nuclei or squashing preparation, respectively), and washed in 2× SSC (RT, 2x5 min). Afterwards, slides were post-fixed in 4% (v/v) Formaldehyde in 2× SSC (RT, 10 min), washed in 2× SSC (RT, 2x5 min), dehydrated in 70, 90, 96% ethanol and air-dried. For each slide, 20 µl of prepared FISH probes containing hybridization mixture were pre-denatured (95°C for 5 min, then chilled on ice for 10 min) and transferred to the slide. The preparation was covered with a coverslip, sealed by rubber gum, denatured on a heat plate (80°C, 2 min) and incubated in a moist chamber at 37°C for 15-18 h.

Post-hybridization washing was performed as described (Lysak *et al.*, 2006b). For probes labeled indirectly (some single-copy and 5S rDNA probes) a fluorescent detection was required. Slides were incubated with blocking buffer [5% (w/v) BSA, 0.2% (v/v) Tween 20 in 4× SSC] in a moist chamber at 37°C for 1h. Afterwards, fluorescent detection of biotin with Fluorescein-conjugated or DyLight®594-conjugated streptavidin (1:200, Vector Laboratories) or of digoxigenin with DyLight®488-conjugated goat anti-digoxigenin (1:200, Vector Laboratories) in antibody buffer [0.5% (w/v) BSA, 150 mM NaCl, 100 mM Tris-HCl, pH 7.5] was carried out in a moist chamber at 37°C for 45 min. Then slides were washed in TNT buffer [100 mM Tris-HCl, pH 7.5, 150 mM NaCl, 0.05% (v/v) Tween-20] at 42°C for 3x5 min. Probes labeled directly with fluorescent-dUTPs did not require a detection step. The post-washing and detection steps for single-copy probes were carried out at high

stringent condition (50°C). Slides were finally dehydrated in an ethanol series, air-dried and counterstained with DAPI (10 µg/ml) in VectaShield.

For several purposes, hybridized slides were reused for further FISH experiments. In these cases, the hybridized probes were removed according to Shibata *et al.* (2009). In brief, slides were rinsed in 0.1× SSC at RT for 3×10 min to detach the coverslip and to clean DAPI solution. The old probes were then stripped by washing in probe stripping solution [0.2% (v/v) Tween-20, 50% (v/v) Formamide in 0.1× SSC] at 45~50°C for 2×7 min. Finally, slides were washed in 0.1× SSC at RT for 2×5 min and dehydrated in an ethanol series before using for further FISH experiments. The stringency of probe-stripping varied depending on the nature of the probe and the specimen. For instance, *G. nigrocaulis* chromosomes were usually deteriorated after a treatment at 45°C or higher (Figure 4) while chromosomes of *G. margaretae* and *G. hispidula* could be treated at 45°C and 50°C, respectively, for several times with only minor damage of chromosome morphology.

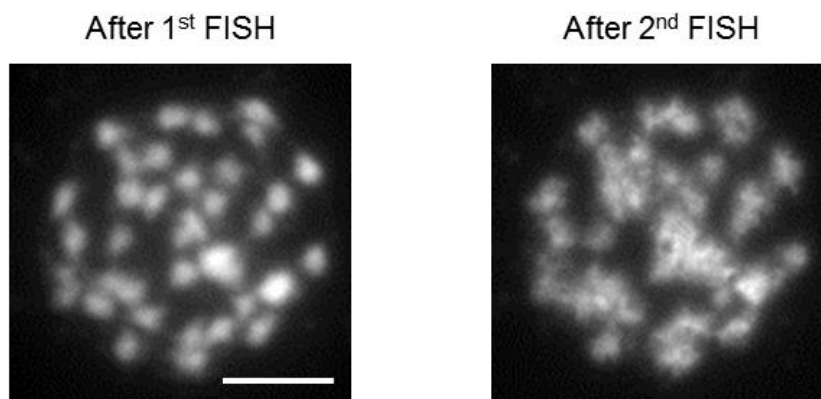


Figure 4: Structure of *G. nigrocaulis* chromosomes after two sequential FISH experiments.

Probe-stripping treatment after the 1st FISH was performed at 45°C. Conditions of two FISH experiments were identical. Bar represents 5 µm.

2.10. Microscopy and image processing

FISH and immunostaining preparations were analyzed using a Zeiss Axioplan 2 epifluorescence microscope (Zeiss, Germany) equipped with a cooled CCD camera (Diagnostic instruments, USA). Fluorescence images for each fluorochrome were captured separately using an appropriate filter and a 100× objective lens. The images were processed (brightness and contrast adjustment only), pseudo-colored and merged using Adobe Photoshop software (Adobe Systems).

2.11. Analysis of telomere lengths and BAL31-sensitivity

High molecular weight DNA of *Genlisea* plants was prepared in agarose plugs as described (Sykorova *et al.*, 2006). The terminal position of probed sequences was tested by BAL31 nuclease sensitivity. DNA samples in agarose plugs (~2 µg) were equilibrated in 200 µl of BAL31 nuclease buffer (NEB) for

15 min and then digested in buffer with 3 units of BAL31 nuclease in a thermomixer (Eppendorf) at 30°C for 20 or 60 min. The reaction was stopped by buffer exchange with 50 mM EGTA pH 8.0, and BAL31 was irreversibly inactivated by incubation at 58°C for 30 min. Then the plugs were washed in 0.1× TE buffer (3×15 min) and equilibrated in appropriate restriction buffer for subsequent restriction enzyme digestion to measure telomere lengths by terminal restriction fragment (TRF) analysis (Fajkus *et al.*, 1998). After digestion, the low-molecular-mass fraction of digested DNA was precipitated with ethanol and dissolved in TE for analysis by conventional agarose-gel electrophoresis and DNA hybridization. High-molecular-mass fractions were retained in the agarose plugs, and analyzed by PFGE using the Chef Mapper instrument (BioRad) under the following conditions: 1% (w/v) FastLane agarose-gel in 0.5× TBE buffer, 6 V/cm, ramping time 0.5-26 sec, 14°C for 20 h. Conventional and PFGE gels were alkali blotted and hybridized with radioisotope-labeled probes (Fojtova *et al.*, 2010). Hybridization signals were visualized with a FLA-7000 phosphorimaging system (Fuji Film). This part was done by Prof. J. Fajkus and Dr. M. Fojtova.

2.12. Telomere repeat amplification protocol (TRAP)

Leaves and shoots of *Genlisea* species were manually homogenized in extraction buffer according to Fitzgerald *et al.* (1996). Crude extracts obtained after centrifugation were 5× and 10× diluted for analysis of telomerase activity. Primers usually applied in TRAP assays for plants (Fajkus *et al.*, 1998) were used for analysis of *G. pygmaea*. Additional primers and primer combinations corresponding to putative telomere repeats in *G. hispidula* were tested (Table 5). One µl of 10 µM substrate primer (TS21 for *G. pygmaea* or 47F for *G. hispidula*; Fojtova *et al.*, 2002) was mixed with 1 µl of diluted crude protein extract and elongation of the primer by the telomerase proceeded in 25 µl reaction buffer at 26°C for 45 min (Fitzgerald *et al.*, 1996). After heat inactivation of telomerase (94°C, 5 min), 1 µl of 10 µM reverse primer (TelPr for *G. pygmaea*; HisPr short or HisPr long for *G. hispidula*) and 2 units of DyNAzymell DNA polymerase (Finnzymes) were added and extension products were amplified (35 cycles of 95°C/30 sec, 65°C/30 sec, 72°C/30 sec; final extension 72°C/5 min). Aliquot samples of TRAP reactions were analyzed on 12.5% (w/v) polyacrylamide gel in 0.5× TBE buffer. Gels were stained by GelStar Nucleic Acid Gel Stain (LONZA) and signals were visualized using the LAS-3000 system (FujiFilm). Products of TRAP were cloned using a TOPO-TA cloning kit (Invitrogen) and sequenced to characterize sequences added by telomerase. This part was done by Prof. J. Fajkus and Dr. M. Fojtova.

Table 5: Sequences of primers used in analyses of telomerase activity in tissues of *G. hispidula* and *G. pygmaea*.

S – substrate primer; R – reverse primer.

Primer	Primer sequence (5'-3')	Note
TS21	GACAATCCGTCGAGCAGAGTT	S, <i>G. pygmaea</i>
TelPr	CCGAATTCAACCCTAAACCCTAAACCCTAAACCC	R, <i>G. pygmaea</i>
47F	CGCGGTAGTGATGTGGTTGTGTT	S, <i>G. hispidula</i>
HisPr long	CCGAATTCAAACCTGAAACCTGAAACCTGAAACC	R, <i>G. hispidula</i>
HisPr short	CCGAATTCTGAACCTGAACCTGAACCTGAACC	R, <i>G. hispidula</i>

2.13. Sequence analysis

Sequences of repetitive elements cloned from genomic DNA of *G. nigrocaulis*, *G. pygmaea* and *G. hispidula* (section 2.8) were processed and analyzed using the BioEdit software version 7.2.5 (Hall, 1999). For centromeric tandem repeats of *G. nigrocaulis* and *G. pygmaea*, repetitive units were aligned by the ClustalW program, and calculated for sequence similarity matrix. An un-rooted phylogenetic tree of centromeric repeat units was obtained by the DNA Maximum Likelihood program with default parameters. For centromeric retrotransposons of *G. hispidula*, the reverse transcriptase domain was extracted from sequence of analyzed elements using the BLAST program with rice *RIRE* element as a query and combined with the data set of other CR elements provided by Neumann *et al.* (2011). Phylogenetic analysis of CR elements was performed using the Protdist to Neighbor phylogenetic tree program. All phylogenetic trees were drawn and edited using the FigTree program [<http://tree.bio.ed.ac.uk/software/figtree/>].

3. RESULTS AND DISCUSSION

3.1. Differences in cytological features reflect the genome size plasticity in the subgenus *Genlisea*

The subgenus *Genlisea* is the most diverse clade of the carnivorous genus *Genlisea* regarding not only its geographic distribution and species number but also, and specially, its genome size variation. In this first part, the basic cytological features at the nuclear and the chromosomal level of representative species of the three sections within this subgenus will be described.

3.1.1. The variation of nuclear DNA content reveals one of the largest genome size ranges for a flowering plant genus

During the last decade, flow-cytometry became the preferred method for genome size measurement in plants. Besides the easiness of sample preparation and high throughput, the capability to estimate genome size, ploidy level, nuclear replication state and endopolyploidy are advanced features of this method in comparison with other approaches such as Feulgen densitometry or genome sequencing (Dolezel *et al.*, 2007).

With suitable internal standards (see Materials and Methods), nuclear genome sizes of six species of the subgenus *Genlisea* were estimated based on at least four different measurements for each species: for the section *Genlisea*, *G. nigrocaulis* (86 Mbp/1C), *G. pygmaea* (179 Mbp/1C) and *G. aurea* (133 Mbp/1C); for the section *Recurvatae*, *G. margaretae* (184 Mbp/1C) and for the section *Africanae*, *G. hispidula* (1,555 Mbp/1C) and *G. subglabra* (1,622 Mbp/1C) (Figure 5).

In general, our genome size measurements of species from sections *Genlisea*, *Recurvatae* and *Africanae* are in accordance with those previously published by Greilhuber *et al.* (2006), Veleba *et al.* (2014) and Fleischmann *et al.* (2014). Our values, except that of *G. margaretae*, deviate by not more than 12% from the average values of previous findings (Figure 5). These variations can be explained by the use of different reference standards and different instruments. Contradictory results were reported for *G. margaretae*. Greilhuber *et al.* (2006) once claimed this species to possess the smallest angiosperm genome size. However, our measurement suggested a mis-identification of the *G. margaretae* sample used by Greilhuber *et al.* (2006) as later on was confirmed by Fleischmann *et al.* (2014). Within the section *Tayloria*, considerably different values were reported for *G. lobata* and *G. violacea* (Figure 5). It remains to be elucidated to which extent these differences reflect unrecognized taxonomic diversities (Veleba *et al.*, 2014), wrong taxonomic determinations of morphologically similar species or the occurrence of populations with different ploidy levels.

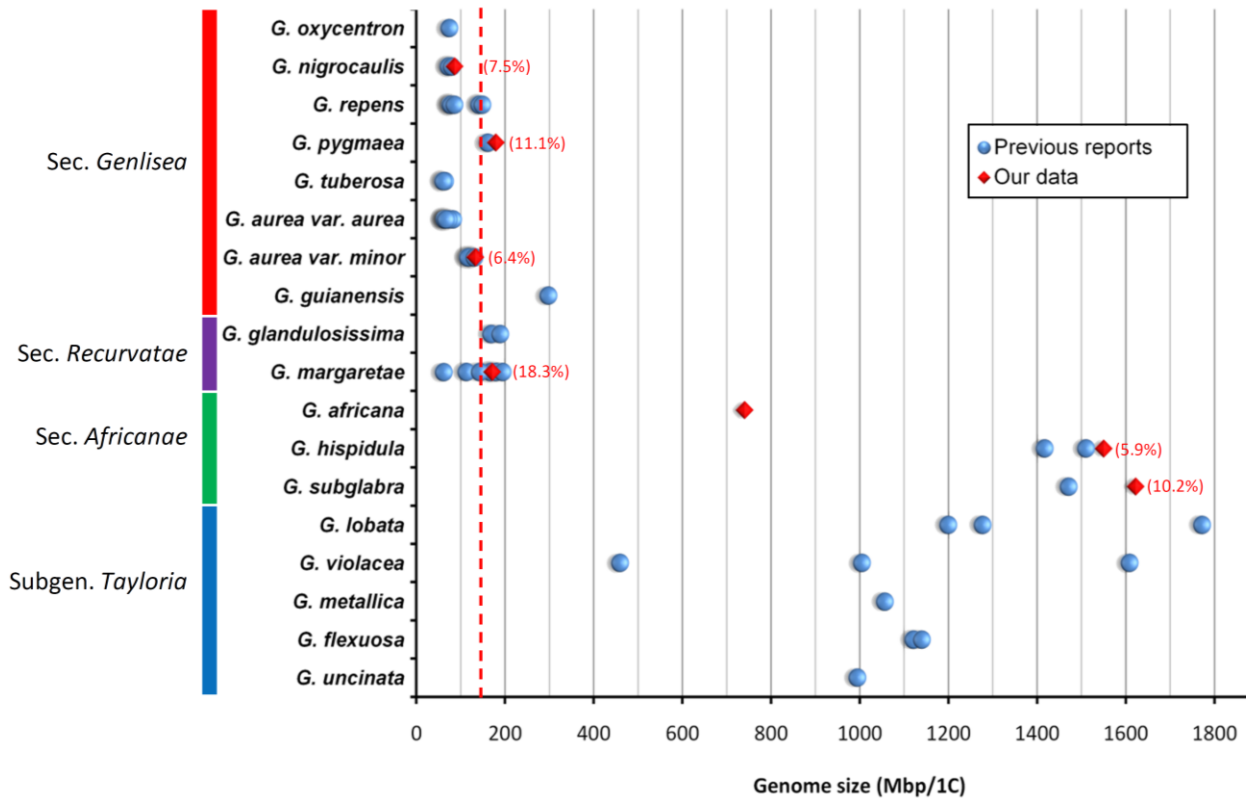


Figure 5: The striking difference in genome size within *Genlisea*.

The order of taxa is accordant with their phylogenetic positions with sections/subgenus marked by colors as in figure 3C. The vertical red dash line represents the genome size of *A. thaliana* (Bennett *et al.*, 2003). Blue dots represent genome size data that were collected from literature (Greilhuber *et al.*, 2006; Fleischmann *et al.*, 2014; Veleba *et al.*, 2014). Genome sizes of taxa marked with red diamonds are our own results with numbers indicating the percentage deviations from the average values of reported data. The value of *G. africana* is based on a single measurement. For other taxa, genome sizes are not available.

In the phylogenetic context, the genome sizes measured so far for *Genlisea* species revealed a tendency to decrease in the more derived taxa. Species of the basal section *Africanae* of the subgenus *Genlisea* have larger genomes than species belonging to the evolutionary younger sections *Recurvatae* and *Genlisea*. This evolutionary genome size switch most likely started latest in the common ancestor of *Recurvatae* and *Genlisea* after branching from *Africanae*. Since all miniaturized genomes were found in the section *Genlisea* (Fleischmann *et al.*, 2014), the trend of genome size shrinkage seems to be more intensive with evolutionary advancement in this section. Additionally, genome size data also suggest polyploid populations within *G. aurea*, (Albert *et al.*, 2010), *G. repens* (Fleischmann *et al.*, 2014) and polyploidy of *G. pygmaea* within the section *Genlisea*. The observed genome size plasticity indicates a bi-directional genome size evolution within the genus *Genlisea* which probably started from an intermediate size (400~800 Mbp/1C) at the ancestral basis of the genus. The wide range of genome sizes is supposed to be selection neutral because of similar natural

habitats and morphological features of species of the lineages with ultra-small and those with manifold larger genomes (*G. nigrocaulis* and *G. hispidula*) (Vu *et al.*, 2015).

The nuclear DNA content of more than 50 shrubby *Oxalis* species varies ~44-fold (from 362 to 16,135 Mbp/1C; De Azkue & Martinez, 1988; Emshwiller, 2002) which represents the largest fold-range of genome size within an angiosperm genus. The estimated genome sizes in *Genlisea* range from ultra-small ($1C \leq 100$ Mbp) to small ($1C \leq 3,400$ Mbp) categories (Greilhuber *et al.*, 2006) and reveal with a ~25-fold (within the subgenus *Genlisea*) or ~27-fold (within the entire genus *Genlisea*) genome size differences, and thus one of the largest fold-ranges of genome size recorded so far within angiosperm genera.

3.1.2. Nuclear distribution of epigenetic methylation marks depends not only on genome size but also on the dispersion of repetitive DNA

The interphase phenotypes in correlation with genome size and DNA composition of *Genlisea* species were examined using flow-sorted nuclei. Despite the low content of repetitive DNA (16.6%; Vu *et al.*, 2015) nuclei of *G. nigrocaulis* showed strongly DAPI stained chromocenters. Similar microscopically detectable chromocenters were observed in nuclei of *G. pygmaea* and *G. aurea* (Figure 6A). This nuclear phenotype is coupled with nuclear chromatin organization which, usually, is a characteristic of small genomes with low repetitive DNA content such as in *A. thaliana* (Fransz *et al.*, 2002). Conversely, the interspersed active genes with many and mainly silenced TEs is responsible for a more homogenous chromatin organization in larger genome species (Houben *et al.*, 2003). Thus the nearly homogeneous DAPI staining pattern without conspicuous heterochromatin clusters in nuclei of *G. hispidula* is apparently due to its larger genome size, containing ~64% of mainly dispersed repetitive sequences (most of them are TEs) (Vu *et al.*, 2015). Interestingly, *G. margaretae* showed a similar nuclear phenotype, although possessing a small genome as *G. pygmaea* and *A. thaliana* (Figure 6A).

The nuclear chromatin organization of the five *Genlisea* species was further investigated by DNA and histone H3 methylation marks. For testing DNA methylation, immunolabelling of interphase nuclei using antibodies against 5-methylcytosine (5mC) was performed. While nuclei of *G. nigrocaulis* as well as of *G. pygmaea* and *G. aurea* showed an accumulation of signals at the heterochromatic chromocenters (Figure 6B), nuclei of *G. hispidula* and surprisingly also of *G. margaretae* revealed a dispersed signal distribution throughout the entire nucleus (Figure 6B). A similar clustering of signal at the chromocenters was obtained in nuclei of *G. nigrocaulis*, *G. pygmaea* and *G. aurea* using antibodies against methylation at lysine residues of histone H3 (H3K9me2, H3K27me1,2) (Figure 6C),

modifications that were previously identified as conserved heterochromatin-associated marks in plants (Fuchs & Schubert, 2012). An antibody against H3K9me1, a heterochromatin-specific mark in *A. thaliana*, additionally labeled the nuclear euchromatic region in these three species (Figure 6D). On the other hand the typical euchromatin-associated marks H3K4me1,2,3, H3K9me3 and H3K27me3 preferentially labeled in *G. nigrocaulis*, *G. pygmaea* and *G. aurea* the euchromatic regions (Figure 6E). In nuclei of *G. hispidula* and *G. margaretae* all histone methylation marks were found nearly homogeneously distributed over the entire chromatin (Figure 6E).

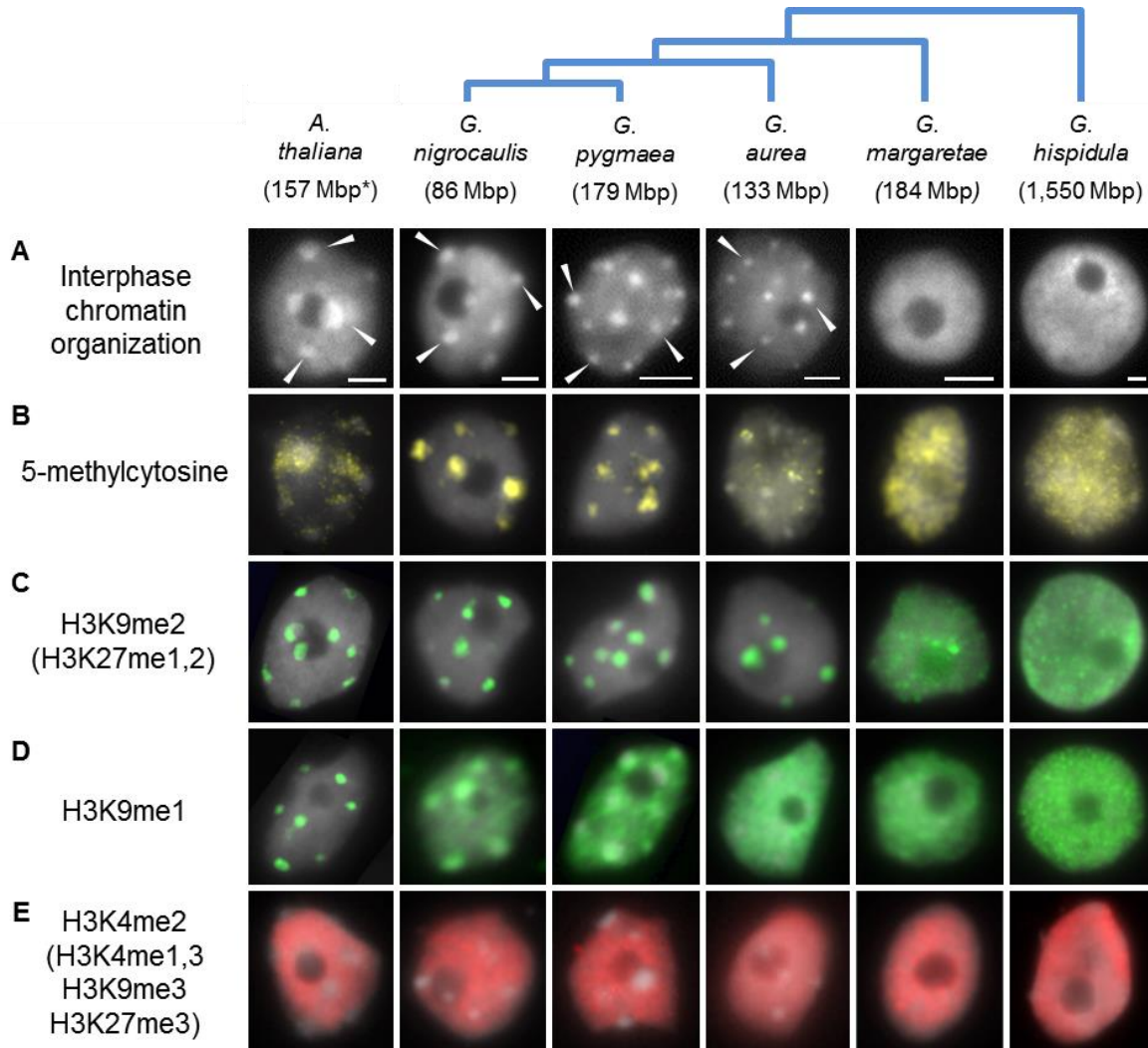


Figure 6: Nuclear heterochromatin organization and distribution of methylation marks in *Genlisea* species in comparison to *A. thaliana*.

(A) Interphase chromatin organization of *Genlisea* species in conjunction with genome size (*; Bennett *et al.*, 2003) and phylogenetic relationship. Pronounced heterochromatic chromocenters (white arrow heads) can be observed in small genome species as in *A. thaliana*. Bars represent 2.5 μm . (B-D) Distribution of heterochromatin-specific marks (DNA methylation and H3K9me2 in (B) and (C), respectively) in small genomes resembles that of *A. thaliana* except H3K9me1 (D). (E) Dispersed distribution of euchromatin-specific histone marks except at heterochromatic chromocenters in small genomes. Note that *G. hispidula* and even *G. margaretae* show more or less dispersed signal patterns resembling that of larger genomes. Due to lack of material, nuclei of *G. subglabra* were not investigated.

Histone and DNA methylation are highly conserved among eukaryotes. In plants, the nuclear distribution of these modifications in different species is known to depend on their genome sizes. In genomes ≤ 500 Mbp/1C the distribution of heterochromatin-specific marks are usually restricted to cytologically detectable heterochromatic regions (Houben *et al.*, 2003). The expected immunolabelling patterns of DNA methylation and heterochromatin-associated histone H3 methylation marks were observed in nuclei of *G. nigrocaulis* and other small genome species such as *G. pygmaea* and *G. aurea*. In larger genomes (>500 Mb), heterochromatin-associated methylation marks are often distributed more uniformly because of a higher density of mobile elements which are silenced and scattered throughout the entire genome. Additionally, in some medium sized genomes the lack of pronounced heterochromatic chromocenters may occur (Houben *et al.*, 2003). In both respects, the nuclear phenotype of *G. hispidula* is in concordance with the expectation.

Interestingly the homogenous nuclear phenotype and distribution of DNA as well as histone methylations found in nuclei of *G. margaretae* does not correlate with the expectation for such a small genome. A similar nuclear distribution of hetero- and euchromatin-specific marks were also observed in nuclei of *Oryza sativa* (Houben *et al.*, 2003), which possesses a 430 Mbp genome with about 40% repetitive sequences (Kawahara *et al.*, 2013). The more homogenous distribution of repetitive DNA within the rice genome rather than its genome size seems to influence the nuclear chromatin organization. A similar assumption probably fits for *G. margaretae*.

On the other hand, the additional labeling of euchromatic regions by the heterochromatin-specific mark H3K9me1 shown for *G. nigrocaulis*, *G. pygmaea* and *G. aurea* is another deviation in the nuclear distribution of methylation marks. Deviating distributions of histone methylation marks were also reported for other species. In maize, the heterochromatic marks H3K9me1 and H3K27me1 do not preferentially label the heterochromatic knobs but are more or less uniformly present in hetero- and euchromatic regions and H3K9me2, which is heterochromatin-specific in many tested plants, was found to associate with euchromatic region rather than with heterochromatic knobs (Shi & Dawe, 2006). Braszewska-Zalewska *et al.* (2010, 2012) reported a similar observation that H3K9me2, exclusively enriched in the chromocenters of *Brassica rapa* nuclei (782 Mbp/1C), labeled homogeneously the nuclei of *B. oleracea*, *B. nigra*, *B. napus*, *B. juncea* and *B. carinata* independent of their genome sizes. Although such deviation of heterochromatin-specific methylation marks can be interpreted as the labeling of temporarily inactive genes (Shi & Dawe, 2006), the irregular nuclear patterns of H3K9me1,2, H3K27me1 in maize, of H3K9me2 in *Brassica* and of H3K9me1 in *Genlisea* are apparently species-specific rather than correlated with genome size.

3.1.3. The chromosome number is highly variable in groups with miniaturized genomes but likely constant in the group with larger genomes

Preparation of well-spread and morphologically-intact chromosomes is an essential prerequisite for cytogenetic analysis. For plants that markedly differ in size and number of chromosomes, several technical aspects such as type of tissue and suitable treatments, cell wall digestion and chromosome spreading methods need to be optimized.

Except those of the subgenus *Tayloria* which is characterized by a quite low number of relative large chromosomes (Fleischmann *et al.*, 2014), chromosome counting in *Genlisea* is hampered not only by large numbers of small chromosomes but also by the difficulty in obtaining suitable tissues in some species. To establish a routine cytological preparation protocol for the six *Genlisea* species studied, different types of tissue (flower buds, tips of very young leaves and tips of young, early developed traps) and pretreatment conditions for the accumulation of dividing cells (ice-cold water or 20 mM 8-Hydroxyloquinoline) were tested using either standard squashing or drop-squashing methods. The optimal procedure was determined for each species empirically. For *G. nigrocaulis*, *G. pygmaea*, *G. aurea* and *G. margaretae*, flower buds were found to be the most suitable tissue for chromosome preparation by squashing. Additionally, the tips of very young leaves of *G. nigrocaulis* and *G. margaretae*, slowly softened by a low concentration of enzymes, are another source for chromosome spreading. Similarly, very young leaf tips of *G. hispidula* and *G. subglabra* were preferentially used to prepare metaphase chromosomes using drop-squashing (Table 6).

Table 6: Established procedures for cytological preparation of *Genlisea* species using different types of tissue.

Species	Tissue	Cell wall digestion (PC digestion at 37°C, min)	Chromosome spreading method	Cytoplasm treatment* (0.1 mg/mL Pepsin in 0.01N HCl at 37°C)
<i>G. nigrocaulis</i>	Flower buds	2% PC for 15 min	Squashing	10 min
	Leaf tips	1% PC for 30 min		7 min
<i>G. pygmaea</i>	Flower buds	2% PC for 15 min	Squashing	10 min
<i>G. aurea</i>	Flower buds	2% PC for 15 min	Squashing	10 min
<i>G. margaretae</i>	Flower buds	2% PC for 15 min	Squashing	10 min
	Leaf tips	1% PC for 35 min		7 min
<i>G. hispidula</i>	Leaf tips	2% PC for 10 min	Drop-squashing	10 min
<i>G. subglabra</i>	Leaf tips	2% PC for 10 min	Drop-squashing	10 min

(*) Treatment before FISH

Using optimized protocols well-spread chromosome complements facilitating an unambiguous counting were obtained for all six *Genlisea* species (Figure 7). In the section *Genlisea*

characterized by species with small genomes, $2n = 40$ was counted for *G. nigrocaulis*, $2n = 80$ for *G. pygmaea* and $2n \approx 104$ for *G. aurea*. The lowest number was observed in *G. margaretae* as member of the section *Recurvatae* with $2n = 38$ chromosomes. Both members of the section *Africanae* characterized by a large genome, *G. subglabra* and *G. hispidula*, possess $2n = 40$ chromosomes. The chromosome numbers determined for the six *Genlisea* species do not correlate with their genome size. Due to the high number, the chromosomes in studied *Genlisea* species are tiny ranging from less than $1 \mu\text{m}$ in *G. aurea* to $\sim 5 \mu\text{m}$ in *G. subglabra* which impede further identification of primary constrictions and other morphologic characteristics.

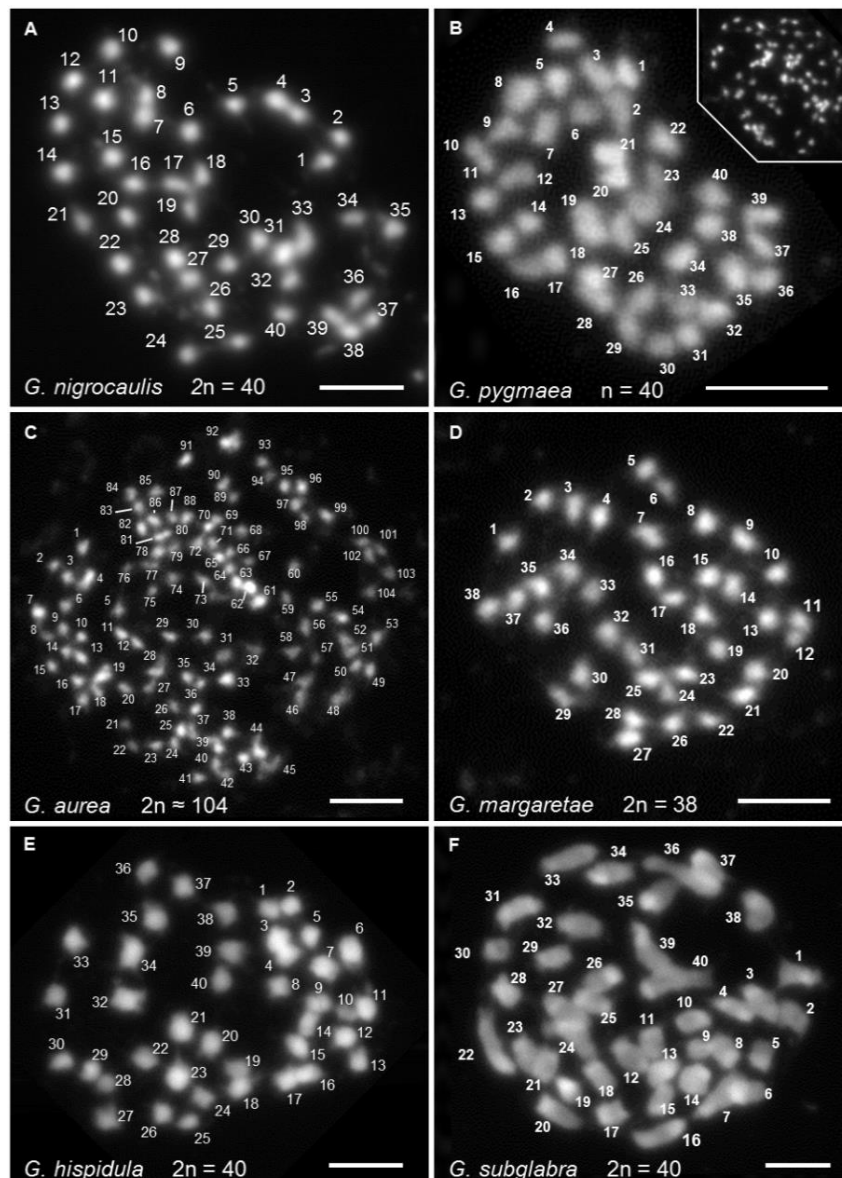


Figure 7: Chromosome numbers of six *Genlisea* species representing three sections of the subgenus *Genlisea*. (A) Mitotic spreading showing $2n = 40$ chromosomes in *G. nigrocaulis*. (B) 40 meiotic metaphase II bivalents of *G. pygmaea* (the inset shows mitotic chromosomes). (C) $2n \approx 104$ chromosomes in *G. aurea*. (D) $2n = 38$ chromosomes in *G. margaretae*. (E-F) $2n = 40$ chromosomes in *G. hispidula* (E) and *G. subglabra* (F). Bars represent $5 \mu\text{m}$.

Together with genome size evolution, chromosome number alteration in eukaryotes is an important driving force in speciation. Chromosome number alteration may involve the whole chromosome complement (polyploidy) or individual chromosomes (aneuploidy, dysploidy) (Schubert, 2007; Schubert & Lysak, 2011). In *Genlisea*, five of eight species of the subgenus *Tayloria* showed a constant chromosome numbers of $2n = 16$ (Greilhuber *et al.*, 2006; Fleischmann *et al.*, 2014). Fleischmann *et al.* (2014) claimed that this might be the basic chromosome number of the genus. However the calculation of basic number is difficult if data on ancestral karyotypes are missing. We found $2n = 40$ as the most frequent and possibly, the basic chromosome number for the small genome clade of the subgenus *Genlisea*. The divergent chromosome number emerging in sections *Recurvatae* (*G. margaretae*) and *Genlisea* (*G. aurea*) may be the results of numerical chromosome rearrangements. Polyploidization also occurred within the section *Genlisea*. Two populations of *G. aurea* were previously described to possess a nearly 2-fold difference in genome size and were considered as either two ploidy variants (Albert *et al.*, 2010; Veleba *et al.*, 2014) or two intraspecific taxa (Fleischmann *et al.*, 2014). The chromosome number and the estimated genome size suppose the investigated *G. aurea* (133 Mbp/1C, $2n \approx 104$) to be the tetraploid variant of the diploid taxon reported by Greilhuber *et al.* (2006) (~ 63 Mbp/1C, $2n \approx 52$). Likewise, the unexpected chromosome number of *G. pygmaea* ($2n = 80$) and its twice as large genome (179 Mbp/1C) in comparison to the close relative *G. nigrocaulis* (86 Mbp/1C, $2n = 40$) may also indicate tetraploidy. Additionally, the comparative analysis between sequenced genomes of *G. hispidula* and *G. nigrocaulis* suggested that the former and probably its close relative *G. subglabra* are tetraploid (Vu *et al.*, 2015). The identical chromosome number ($2n = 40$) shared by *G. nigrocaulis* and *G. hispidula* suggests a dysploid chromosome number reduction in *G. hispidula*. Thus the karyotype evolution of the genus *Genlisea*, particularly within the subgenus *Genlisea*, is apparently more complicated than that was proposed by Fleischmann *et al.* (2014).

3.1.4. Number and chromosomal distribution of ribosomal DNA loci are similar within, but different between groups characterized by small or large genomes

Ribosomal genes (rDNA) characterized by conserved sequences and variable locus numbers as well as chromosomal distributions are informative FISH markers not only for species discrimination but also for karyotype evolution analysis as shown for Brassicaceae (Ali *et al.*, 2005; Hasterok *et al.*, 2006b; Mandakova & Lysak, 2008). For the six *Genlisea* species, two distinct distribution categories of 45S and 5S rDNA loci were revealed by FISH connected to the small and large genome sizes. For species with small genomes, both rDNA probes revealed one locus each on metaphase plates of *G. nigrocaulis*, *G. pygmaea* and *G. margaretae* (Figures 8A, B, D), while they were found on two

chromosome pairs in *G. aurea* (Figure 8C). In the large genome species, *G. hispidula* and *G. subglabra*, the 45S rDNA probe showed two loci and the 5S rDNA probe was localized on three and four chromosome pairs, respectively. One of these chromosome pairs harbors two 5S rDNA loci in both species (Figures 8E, F).

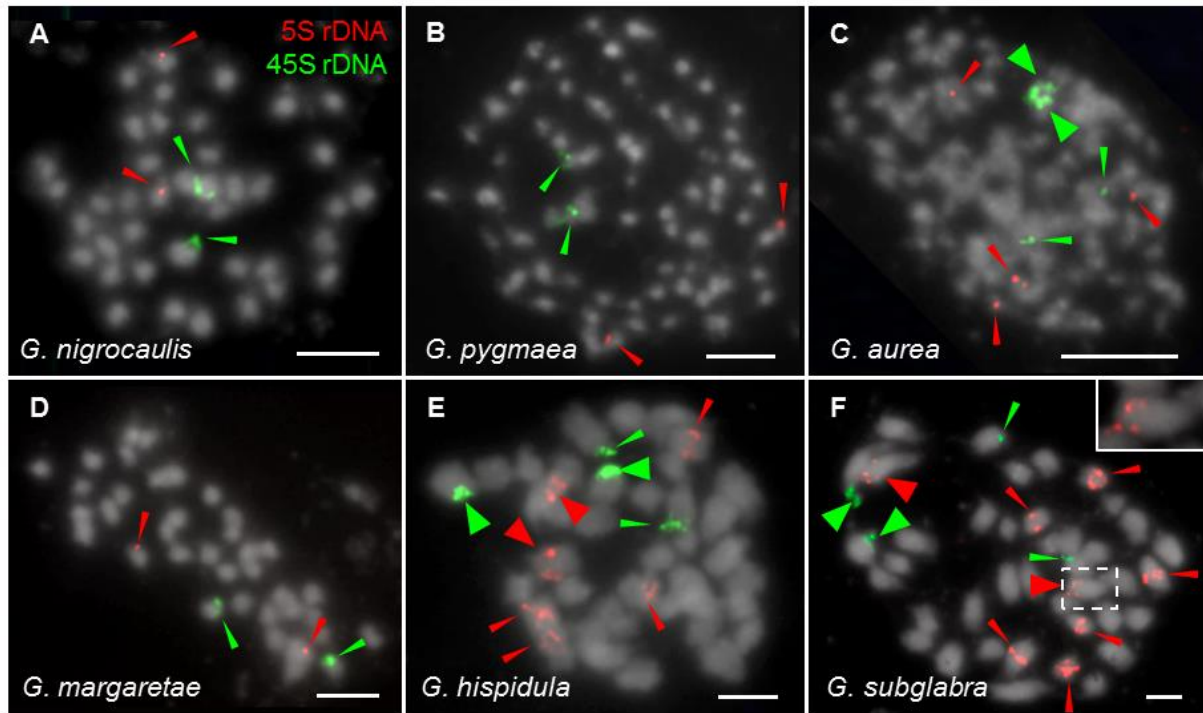


Figure 8: Different loci number and chromosomal distribution of 5S and 45S rDNA in *Genlisea* species.

Bigger arrow heads denote chromosomes bearing two 5S rDNA loci (in red) or stronger 45S rDNA FISH signals (in green) in *G. aurea* (C), *G. hispidula* (E) and *G. subglabra* (F). The inset in (F) shows a magnified *G. subglabra* chromosome bearing two 5S loci. Bars represent 5 μ m.

One locus of each 45S and 5S rDNA, as revealed in the phylogenetically more basal (*G. margaretae*) or the more derived (*G. nigrocaulis*) species, is possibly shared by the sections *Genlisea* and *Recurvatae* with small genomes. The pattern of two loci of each rDNA cluster is apparently the consequence of tetraploidy in the investigated *G. aurea* accession. Multiplication of rDNA loci, however, is not a universal indicator for polyploidy. In polyploids or after dysploid chromosome number reduction, rDNA loci might be lost and thus deviate from the expected numbers. The elimination of excess or dispensable sequences after polyploidy formation contributes to “genome diploidization” and is considered as one of the main causes leading to the loss of rDNA clusters in polyploids (Weiss-Schneeweiss *et al.*, 2013). Such a rDNA loss was observed in autotetraploid *Raphanus sativus* (Hasterok *et al.*, 2006b) and even in the recently formed allotetraploid *A. suecica* (Pontes *et al.*, 2004). The presence of only one locus of each 45S and 5S rDNA in the presumed tetraploid *G. pygmaea* is likely the consequence of a similar process linked with the extensive genome shrinkages in the section *Genlisea* involving not only repetitive sequences but also

redundant genes (Leushkin *et al.*, 2013; Vu *et al.*, 2015). On the other hand, the appearance of stronger FISH signal at one 45S rDNA locus compared to the remaining one may indicate ongoing copy number reduction in the latter rDNA cluster of the tetraploid *G. aurea*, resembling observations in different polyploid cytotypes of *Solanum elaeagnifolium* (Chiarini, 2014) and in synthetic polyploid wheat (Guo & Han, 2014).

Other positional and numerical polymorphisms in rDNA pattern were observed in connection with TE-mediated chromosomal rearrangement (for review see Raskina *et al.*, 2008). The activity of TEs that flank or intercalate rDNA cluster was suspected for the inter-chromosomal mobility of NORs in *Allium cepa* and its interspecific hybrids (Schubert & Wobus, 1985) and for rDNA transposition during the evolution of the genus *Nemesia* (Datson & Murray, 2006). Likewise, the high activity of *Enhancer/Suppressor-mutator (En/Spm)* transposons is possibly linked with the amplification of 5S rDNA loci in *Aegilops speltoides* (Raskina *et al.*, 2004). WGD and proliferation of TEs were determined as mechanisms expanding the genome of *G. hispidula* (Vu *et al.*, 2015). Thereby, either polyploidization or TE-induced rearrangement involving rDNA clusters or even both might be responsible for the observed patterns of 45S and 5S rDNA loci in *G. hispidula* and its close relative *G. subglabra*. The lack of rDNA pattern data for other species makes the inferences of basic rDNA loci within the section *Africanae* and their evolutionary alteration in *G. hispidula* and *G. subglabra* ambiguous.

3.1.5. Conclusion regarding the polymorphism of cytological features in correlation with genome size plasticity in the subgenus *Genlisea*

The dynamic genome evolution in *Genlisea*, particularly within the subgenus *Genlisea*, is represented by the smallest genome estimated so far for *G. tuberosa* (Fleischmann *et al.*, 2014) and by the up to 25-fold larger genome of *G. subglabra*. While polyploidizations were accounted for the largest intrageneric genome size range recorded for *Oxalis* (Emshwiller, 2002), WGD in combination with repetitive DNA accumulation on the one hand and extensive sequence depletion on the other were assumed to be the main evolutionary processes driving the bi-directional genome size evolution within the genus *Genlisea* (Vu *et al.*, 2015).

Nuclear and chromosomal features were determined for the five *Genlisea* species. The nuclear chromatin organization characterized by DNA and histone methylations was found in concordance with those described for plants with small and large genomes. Unexpected chromatin distributions were that *i)* all methylation marks labeled nearly homogeneously the nuclei of *G. margaretae* as in large genomes, probably due to a higher proportion of dispersed repeats in its

genome and *ii*) the histone mark H3K9me1 additionally labeled the euchromatin regions in nuclei of small genome species *G. nigrocaulis*, *G. aurea* and *G. pygmaea*, possibly a typical pattern for the section *Genlisea*.

The optimized procedures for chromosome preparation facilitated for the first time a successful cytological analysis in *Genlisea*. The unambiguous chromosome counts revealed different chromosome number alterations within the subgenus *Genlisea*. Assuming $n = 20$ as basic chromosome number of the subgenus; these variations include chromosome reduction (*G. margaretae*) or increase (*G. aurea*), tetraploidization (*G. aurea*, *G. pygmaea*), and also dysploid chromosome number reduction after polyploidisation (*G. hispidula*, *G. subglabra*). Moreover, two patterns of rDNA distribution were observed for the groups possessing either small or larger genomes. These data, taken together, reveal a polymorphism in cytological features which reflect the genome size plasticity and provides a basis for further karyotype evolution analyses within the genus *Genlisea*.

3.2. The variability of repetitive DNA between *G. nigrocaulis* and *G. hispidula* is reflected by the alterations in centromere and telomere sequences

In order to understand the evolutionary mechanisms behind the genome sizes plasticity in *Genlisea*, the genomes of *G. nigrocaulis* and *G. hispidula* were sequenced for a comparative genome analysis (Vu *et al.*, 2015). In this part the remarkable variation in proportion and composition of repetitive profiles which are in accordance with the 18-fold genome size difference of these two species were cytogenetically investigated.

3.2.1. FISH revealed a different chromosomal distribution of repetitive sequences in *G. nigrocaulis* and *G. hispidula*

Repetitive DNA including tandem repeats and dispersed transposable elements is characterized as a ubiquitous component which may occupy a significant proportion of the entire genome. For *G. nigrocaulis* and *G. hispidula* a randomly selected proportion of whole genome shotgun sequence reads had been subjected to a graph-based clustering analysis (Vu *et al.*, 2015). Based on the consensus sequences of the highly abundant clusters, different classes of repetitive DNA of the two species were identified and used as FISH probes to investigate their nuclear and chromosomal distribution.

Repetitive DNA sequences constitute 15.9% of the genome of *G. nigrocaulis*, among these 2.3% were characterized as tandem repeats (Vu *et al.*, 2015). Two major clusters of tandem repeats were identified. The most abundant one, Gn7c161, is a 161bp repeat that showed a strong

accumulation at chromocenters in nuclei and was found to cluster in one signal focus per chromosome of *G. nigrocaulis* (Figure 9A). Such signal pattern suggests this sequence to be a centromere-specific repeat (see next section). The second abundant tandem repeat, Gn44c19, revealed up to ten signals per nucleus and metaphase plate (Figure 9B). Only 0.07% of the genome of *G. nigrocaulis* represent *Ty3/gypsy* elements whereas 2.5% belong to the *Ty1/copia* superfamily of LTR-retrotransposons. However, most of them were characterized by a very low copy number. Among the *Ty1/copia* retroelements, the *Bianca*-like lineage was the most abundant in both genomic and transcriptomic reads (Vu *et al.*, 2015). This element showed a dispersed FISH signal pattern throughout interphase nuclei and on all chromosomes of *G. nigrocaulis* (Figure 9C).

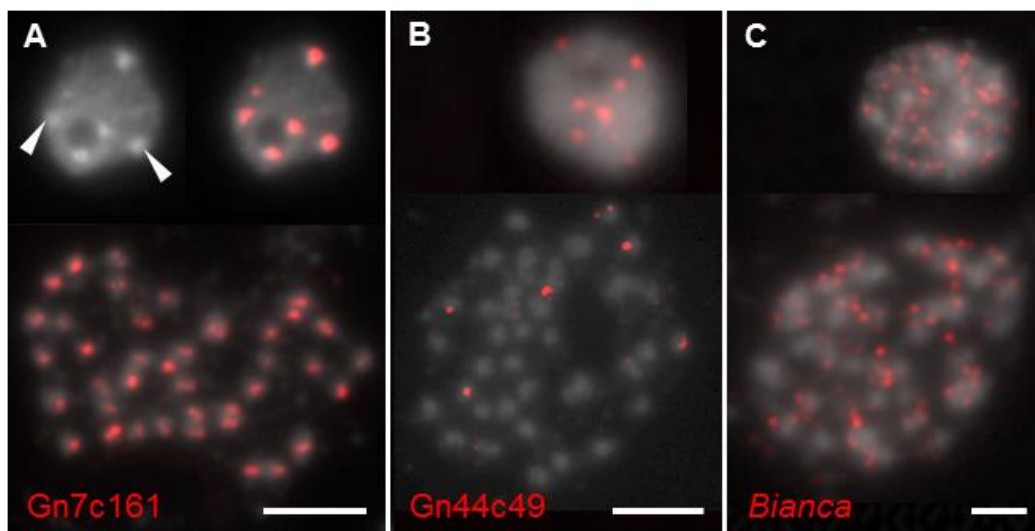


Figure 9: Nuclear and chromosomal distribution of different repetitive sequences in *G. nigrocaulis*.

(A) The most abundant tandem repeat Gn7c161 shows clustered signals on chromocenters (white arrow heads) in interphase nuclei and on all chromosomes. (B) Another repeat Gn44c49 reveals up to ten signals on nuclei as well as on metaphase chromosomes. (C) FISH revealed a dispersed distribution of the most abundant and transcribed *Ty1/copia* retroelement *Bianca* on nuclei and chromosomes of *G. nigrocaulis*. Bars represent 5 μ m.

By contrast, 64.1% of the *G. hispidula* genome were characterized as repetitive sequences (Vu *et al.*, 2015). Among only 1.96% of tandem repeats, the six most abundant motifs, Gh14c16, Gh250c46, Gh45c31, Gh80c174, Gh336c35 and Gh19c56 with monomer lengths of 60, 74, 110, 112, 146 and 900 bp respectively, were identified and used for FISH. Two of them revealed dispersed signals. Whereas Gh14c16 yielded signals on each chromosome pair with a tendency to accumulate towards the chromosome ends (Figure 10A, red arrow heads), Gh336c35 showed a high accumulation on 10 of 20 chromosome pairs (Figure 10B) preferably in regions lacking enrichment of Gh14c16 (Figure 10C). Shibata and Hizume (2002) and Kolano *et al.* (2011) also reported the labeling a half of chromosome complement by repetitive probe in polyploid *Allium* and *Chenopodium* species. This kind of FISH pattern, including that for Gh336c35, resembles those frequently obtained

by GISH for allopolyploids with genomic DNA of a species related to one ancestor as probe, and thus support the assumption of allotetraploidy for *G. hispidula* (Vu *et al.*, 2015). Three tandem repeats were found exclusively in the chromosome subset with low abundance of Gh336c35: Gh19c56 and Gh80c174 revealed distinct signals on one chromosome pair each (Figures 10D, E); Gh45c31 showed signals on two chromosome pairs, one signal pair slightly stronger than the other (Figure 10F, big and small red arrow heads, respectively). Gh250c46 revealed signals on two chromosome pairs, one with high and one with low Gh336c35 abundance (Figure 10G, big and small red arrow heads, respectively). *Ty3/gypsy* retroelements occupy 24.64% of the genome of *G. hispidula*. The full-length *Ty3/gypsy* retroelement Gh338c4 entirely labeled interphase nuclei as well as metaphase chromosomes (Figure 10H). Four other abundant *Ty3/gypsy* retroelements in the genome of *G. hispidula* were characterized as members of the CRM clade (centromeric retrotransposon of maize) (Vu *et al.*, 2015). FISH using these elements as probe revealed scattered signals on nuclei and various degrees of clustered signals on different chromosomes (Figures 10I-L) suggesting the four CRM elements to be constituents of the centromeres in *G. hispidula* (see next section).

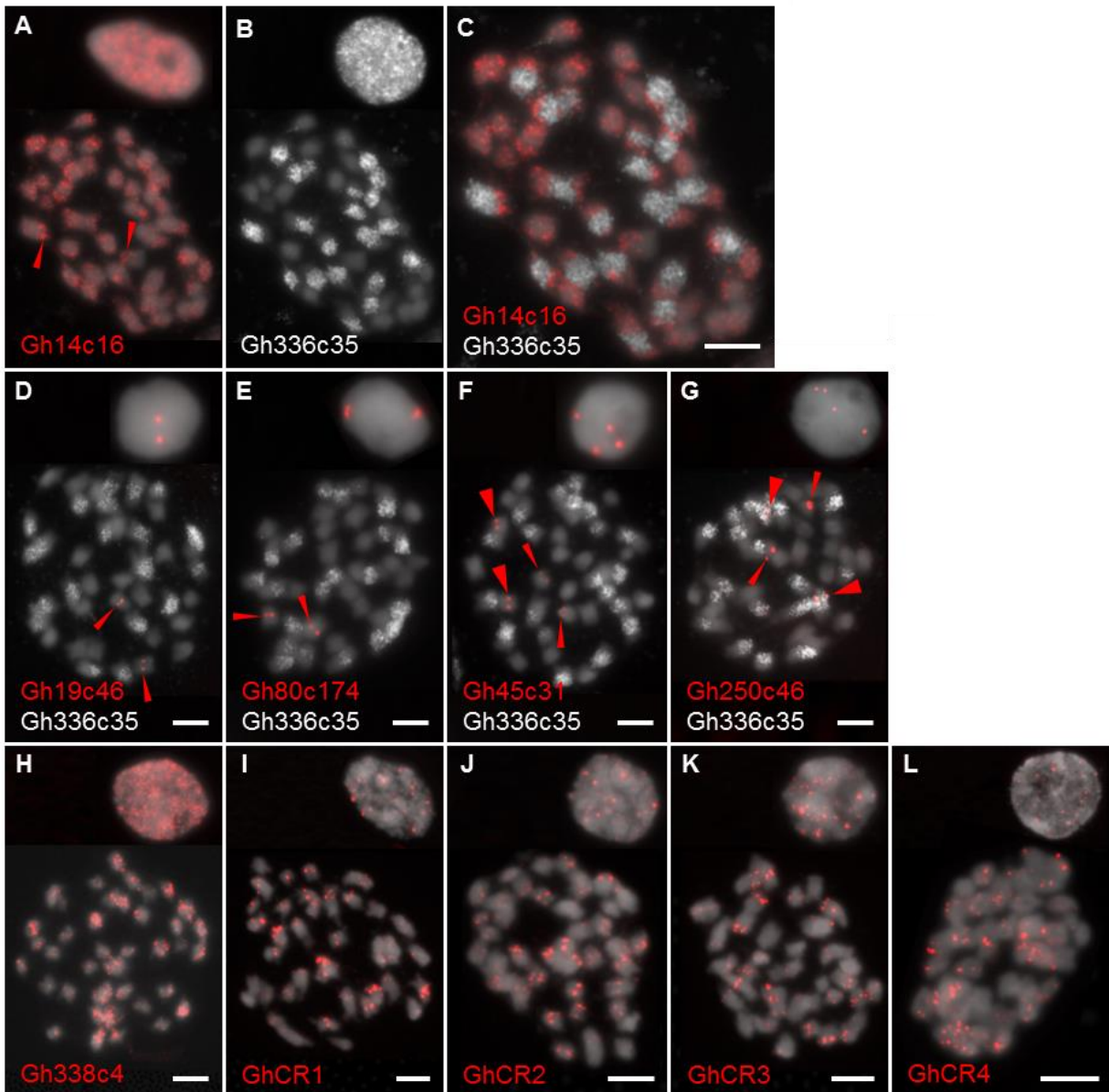


Figure 10: Nuclear and chromosomal distributions of different repetitive sequences in *G. hispidula*.

(A - C) Two tandem repeats yielding dispersed FISH signals: Gh14c16 labeled mostly in terminal regions of each chromosome pair (red arrow heads, **A**), and Gh336c35 accumulated on 10 of 20 chromosome pairs (**B**) preferably at regions that lack the Gh14c16 repeat (**C**). Other tandem repeats showed clustered signals on one (Gh19c46 in **D** and Gh80c174 in **E**) or two (Gh45c31 in **F** and Gh250c46 in **G**) chromosome pairs. With the exception of one locus of Gh250c46 (big red arrow heads, **G**), all loci were found on chromosomes with low Gh336c35 accumulation. (**H**) Gh338c4, a *Ty3/gypsy*-like retroelement, revealed dispersed FISH signals throughout the entire nucleus and metaphase chromosomes. (**I - L**) Clustered FISH signals on chromosomes of four CRM elements. Bars represent 5 μ m.

The sharing of investigated repetitive sequences across the six *Genlisea* species studied was validated by PCR amplification and FISH analysis. The three *G. nigrocaulis*-specific repetitive sequences were amplifiable by PCR from genomic DNA of *G. pygmaea* but not from other *Genlisea* species. Cross-FISH using these three probes yielded signal patterns on nuclei of *G. pygmaea* (Figure

11A) similar to those on nuclei of *G. nigrocaulis*. Neither nuclei of other species with small genomes (*G. aurea*, *G. margaretae*) nor of species with large genomes (*G. hispidula*, *G. subglabra*) showed hybridization signals of these probes. On the other hand, PCR amplification revealed the presence of repetitive sequences identified from the genome of *G. hispidula* in its relative *G. subglabra* but not in *Genlisea* species with small genomes. This co-occurrence was confirmed by FISH analysis for some tandem repeats on nuclei or metaphase chromosomes of *G. subglabra* (Figure 11B).

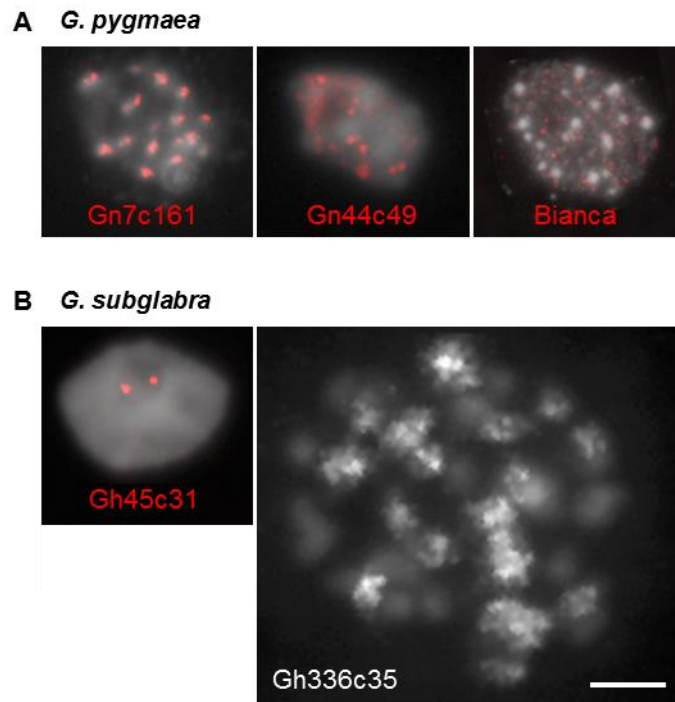


Figure 11: Cross-FISH using repetitive probes identified from *G. nigrocaulis* and *G. hispidula* on nuclei or chromosomes of *G. pygmaea* and *G. subglabra*, respectively.

(A) Three repetitive probes show FISH signal patterns on nuclei of *G. pygmaea* resembling those revealed in *G. nigrocaulis* (Figure 9). (B) Two repetitive probes of *G. hispidula* yielded comparable FISH signal patterns on nuclei and chromosomes of *G. subglabra*. For the tandem repeat Gh45c31, only two distinct signals instead of four as in *G. hispidula* were observed. Bar represents 5 μ m.

Repetitive sequences, especially tandem repeats, evolve rapidly regarding motif sequence, copy number and chromosomal localization which often make them species-specific (Lopez-Flores & Garrido-Ramos, 2012; Mehrotra & Goyal, 2014). The positive results of cross-FISH using repetitive probes of *G. nigrocaulis* in *G. pygmaea* and of *G. hispidula* in *G. subglabra* therefore indicate their relatedness. The presence of the most abundant and dispersed *Ty1/copia* retroelement *Bianca* in *G. nigrocaulis* (Figure 9C) and *G. pygmaea* (Figure 11A) suggests that this retroelement probably invaded the common ancestor of both related species *via* horizontal transfer which seems to occur more frequently than previously assumed (El Baidouri *et al.*, 2014). On the other hand, the similar

chromosomal distribution of the Gh336c35 repeat (Figure 11B) combining with the identical chromosome number ($2n = 40$) indicate that *G. subglabra*, as *G. hispidula*, is allotetraploid.

3.2.2. A 161 bp tandem repeat is a centromeric repeat in *G. nigrocaulis* whereas centromere regions of *G. hispidula* are predominantly occupied by four centromeric retrotransposons

Although the number and size of chromosomes may vary even between closely related species, linear eukaryotic chromosomes invariably possess a centromere as essential constituent which is required for sister chromatid cohesion until it becomes involved in chromosome segregation during nuclear divisions. Across eukaryotes, the function of centromeres is conserved but their sequences are divergent with tandem repeats and retroelements as main constituents.

For *G. nigrocaulis* and *G. hispidula*, candidates for centromere-associated sequences were found *via* cytogenetic characterization of different repetitive sequences identified by graph-based clustering. In the small genome of *G. nigrocaulis* the most abundant tandem repeat Gn7c161 which is henceforth called GnCent, showed strong signal clusters co-localized with intensely DAPI-stained chromocenters in interphase nuclei (Figure 9A). A similar co-localization of centromeric repeats with chromocenters was observed on nuclei of *Arabidopsis* species possessing likewise small genomes and low proportions of repetitive DNA. Additionally, distinct signals of the GnCent repeat on every metaphase chromosome suggested the homogenous constitution of centromeres in *G. nigrocaulis* (Figure 9A).

While no identified tandem repeat showed hybridization signals on more than two chromosome pairs, the four families of *Ty3/gypsy* elements classified as members of the CRM clade (Gorinsek *et al.*, 2004) were found clustered on chromosomes of *G. hispidula* (Figures 10I-L). Further phylogenetic classification based on the reverse transcriptase domain of these retroelements showed that GhCR1, 2, 3 and 4 elements were clustered into a distinct branch of group A, comprising genuine CR elements, most of which possess a putative centromere-targeting domain (Neumann *et al.*, 2011) (Figure 12A). This suggested that the four GhCR elements are constituents of the centromeres in *G. hispidula*. FISH analysis revealed clustered signals on each chromosome and some weaker dispersed signals of individual (Figures 10I-L) or combined GhCR elements (Figures 12B, C).

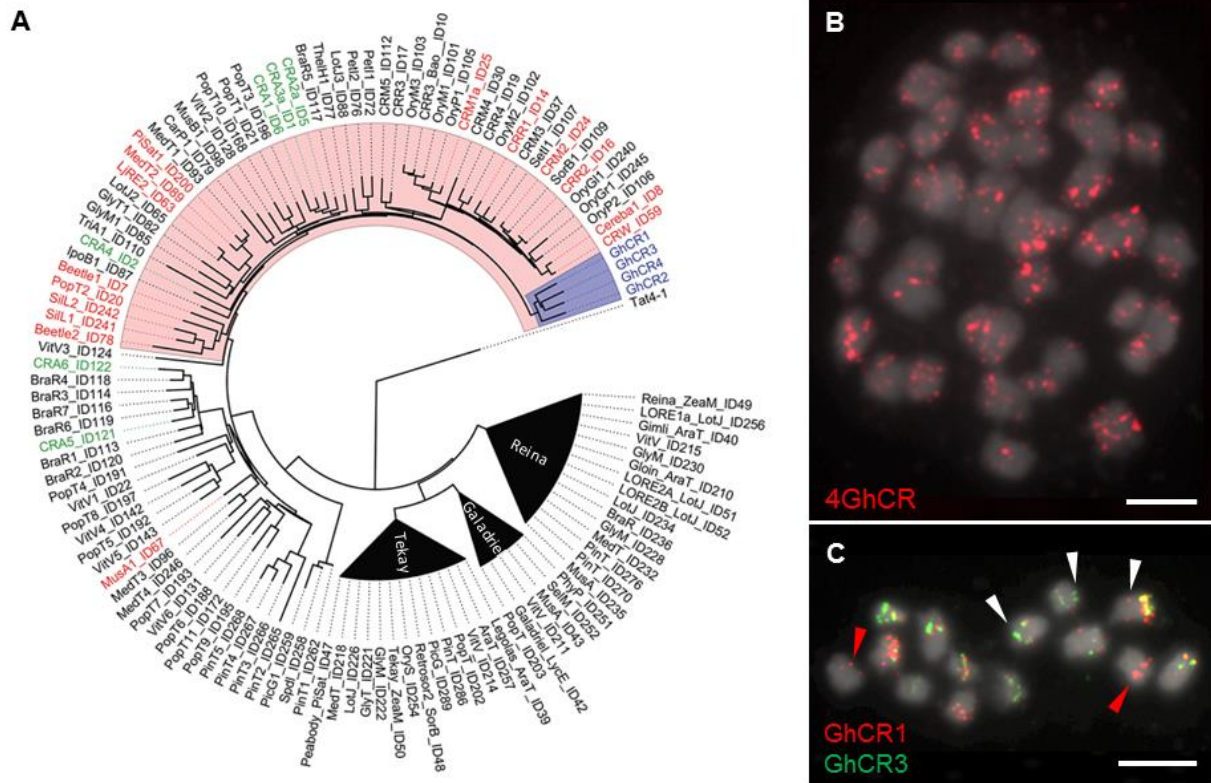


Figure 12: Characterization of the four putative centromeric retrotransposons in *G. hispidula*.

(A) The neighbor-joining tree of the CRM clade, inferred from a comparison of reverse transcriptase domain sequences, reveals a distinct subgroup of GhCR family (blue shading) in a group (pink shading) comprising many elements with centromeric localization confirmed by FISH (red) or by *in-silico* analysis (green) (modified from Neumann *et al.*, 2011). (B) Combination of the four GhCR elements shows signals clustered on each metaphase chromosome of *G. hispidula*. (C) Partial metaphase of *G. hispidula* labeled by GhCR1 and GhCR3. Some chromosomes show enrichment of GhCR1 only (red arrow heads) or additional signals of either of the two sequences (white arrow heads). Bars represent 5 μ m.

The presumed centromere specificity of the GnCent repeat in *G. nigrocaulis* and of the four GhCR elements in *G. hispidula* were confirmed by a combination of FISH, using these repeats, with preceding immunostaining experiments using antibodies against phosphorylated threonine 121 of histone H2A (H2AThr121ph) or phosphorylated serine 10 of histone H3 (H3Ser10ph). Both modifications were previously shown to be highly conserved among plants and useful cytological markers to detect (peri)centromeric regions (Houben *et al.*, 1999; Demidov *et al.*, 2014). In *G. nigrocaulis*, we found adjacent signals for GnCent and H2AThr121ph on metaphase chromosomes (Figure 13A). Similarly, FISH signals of the four GhCR elements showed an adjacent localization to immunosignals for H2AThr121ph (Figures 13B-D) and H3Ser10ph (Figure 13E) on prophase nuclei and metaphase chromosomes in *G. hispidula*.

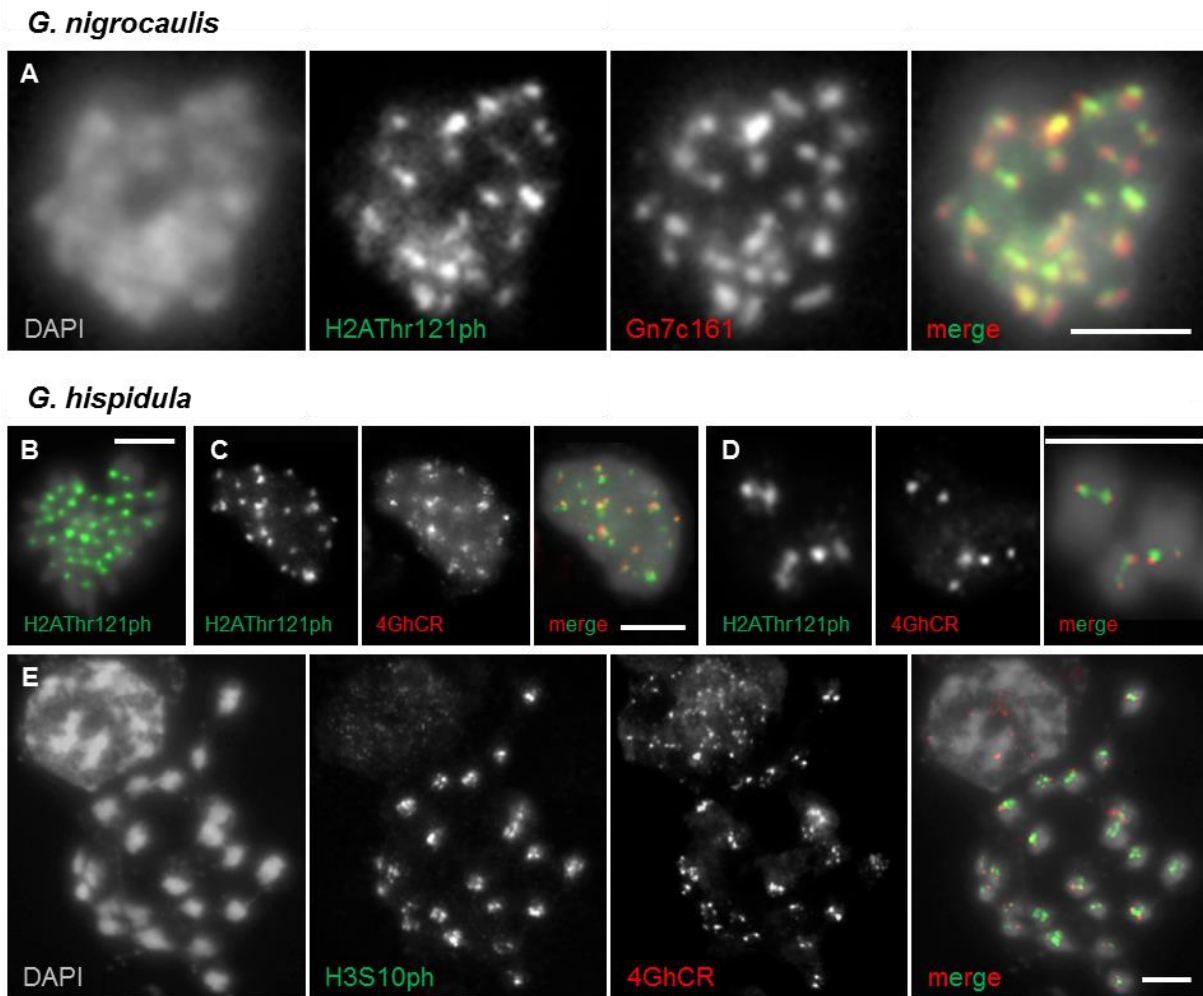


Figure 13: Immunostaining-FISH combinations validating centromeric position of putative repetitive elements.

(A) Adjacent localization of GnCent FISH signals and H2AThr121ph immunostaining signals on metaphase chromosomes confirms the location of GnCent repeats in the (peri)centromere region of *G. nigrocaulis*. (B) Anti-H2AThr121ph signals located on each of 40 chromosomes of *G. hispidula* and co-localize with FISH signals for the four GhCR elements (4GhCR) on prophase nuclei (C) and metaphase chromosomes (D). (E) Partial metaphase of *G. hispidula* showing FISH signals of 4GhCR adjacent to immunosignals of H3Ser10ph, another (peri)centromeric-specific marker. Bars represent 5 μm .

As confirmed by PCR amplification and cross-FISH, the centromere-specific GnCent repeat and the four GhCR elements were undetectable outside the clades of *G. nigrocaulis* and *G. hispidula* except in their close relatives *G. pygmaea* and *G. subglabra*, respectively (Figure 11). The structure of centromere-specific GnCent arrays was further analyzed in both *G. nigrocaulis* and *G. pygmaea*. Because neither single GnCent-specific Gn7c161-reverse nor Gn7c161-forward primer (Table 3) generated PCR products from genomic DNA of both species, the GnCent arrays are repeated in a “head to tail” manner. Two and three dimers cloned from PCR products amplified by the Gn7c161 primer pair from genomic DNA of *G. nigrocaulis* and *G. pygmaea*, respectively, were sequenced. Within the amplified dimer, two repeated monomers, namely A and B, were found to be joined by a

9 bp variable sequence (ATTTAAAAT and ATTCAAAT in *G. nigrocaulis*, ATTTAGAAT and ATTTTGAAT in *G. pygmaea*) (Figure 14A). This AT-rich and sequence-variable part of GnCent repeat was excluded for primer binding region selection, therefore the amplified GnCent monomers were 153 and 156 bp for *G. nigrocaulis* and *G. pygmaea*, respectively (Figures 14B, D). Sequence comparison of these GnCent monomers revealed intraspecific identities of 93.4-100% and 91-97.4% respectively and a significant lower interspecific similarity (83.3-88.4%). While SNPs explain the observed intraspecific variations, sequence dissimilarity between *G. nigrocaulis* and *G. pygmaea* is mostly caused by a 3 bp insertion/deletion in the GnCent monomers (Figures 14A-B, red box). The monomer length of these two GnCent variants is consistent with the typical unit length of known centromeric repeats (Henikoff *et al.*, 2001; Heslop-Harrison & Schwarzacher, 2013). The cross-FISH using GnCent probes derived from either *G. nigrocaulis* or *G. pygmaea* revealed expected signal patterns in both species but not in other studied *Genlisea* species.

Sequence and abundance of centromeric repeats can vary, even between centromeres of the same species (Berr *et al.*, 2006; Gong *et al.*, 2012; Neumann *et al.*, 2012). Furthermore, contribution of centromeric repeats and/or CR elements, the major constituents of plant centromeres, can vary between chromosomes of a complement (for review see Plohl *et al.*, 2014). In *G. nigrocaulis*, and in its tetraploid relative *G. pygmaea*, the GnCent repeat apparently coincides with the centromere position. Taking into account the rapid divergence of centromeric satellite repeats (Henikoff *et al.*, 2001), the presence of the GnCent repeat in both species suggests their close relationship with a common ancestor. Additionally, the fast evolution of centromeric repeats is further characterized by their higher order repeat (HOR) structures in which multiple repeat monomers with specific sequence polymorphisms form a HOR that itself is repeated. Such HOR units, mostly comprising two monomers, are widespread across both plant and animal kingdoms (Dechyeva & Schmidt, 2006; Melters *et al.*, 2013; Koga *et al.*, 2014). Interestingly, sequence alignment of the GnCent monomers revealed several SNPs that differentiate not only species-specific GnCent variants (Figure 14B, white triangle) but also monomers A and B in each species (Figure 14B, black star). Distinct groups of monomer A and B of the GnCent repeat as shown by neighbor joining clustering (Figure 14C) demonstrated the formation of many or most of GnCent monomers as higher order repeats of centromere-specific sequence arrays in both *G. nigrocaulis* and *G. pygmaea* (Figure 14E). Furthermore, the high frequency of monomer-specific SNP implies an ongoing diversification of GnCent repeats which may lead to the formation of a new centromeric repeat in the tetraploid *G. pygmaea*, comparable to that observed in centromeric repeat of the New World monkey clade (Cellamare *et al.*, 2009; Melters *et al.*, 2013).

In *G. hispidula*, FISH revealed the four GhCR elements positionally adjacent to immunosignals for the (peri)centromere marks H2AThr121ph and H3Ser10ph while tandem repeats clustered only on few chromosome pairs. The presence of the GhCRs on each chromosome of the close relative *G. subglabra* suggests that the last common ancestor of both species already recruited these elements as centromeric sequences. While retroelements became accumulated at centromere positions in the *G. hispidula* clade, they were largely eliminated in the course of genome shrinking within the *G. nigrocaulis* clade. Additional weak and dispersed FISH signals of GhCRs along several metaphase chromosomes of *G. hispidula* might indicate remnants of the ancestral state, or cross-hybridization with related elements (Figure 12C).

3.2.3. The Arabidopsis-type telomeric minisatellite is conserved in *G. nigrocaulis* and its close relative *G. pygmaea*

Telomeric minisatellite sequences, the DNA component of the nucleoprotein complexes which protect the ends of linear chromosomes, are highly conserved and specific for particular groups of organisms. The Arabidopsis-type telomere minisatellite, (TTTAGGG), which has been found widespread in plants was also detectable by FISH as strong signal clusters at the chromocenters in interphase nuclei of *G. nigrocaulis* (Figure 15A). On metaphase chromosomes, these signals were found closely adjacent to the centromere-specific GnCent signals (Figure 15B). Similar hybridization patterns were also observed in the closely related tetraploid *G. pygmaea* (Figures 15C, D).

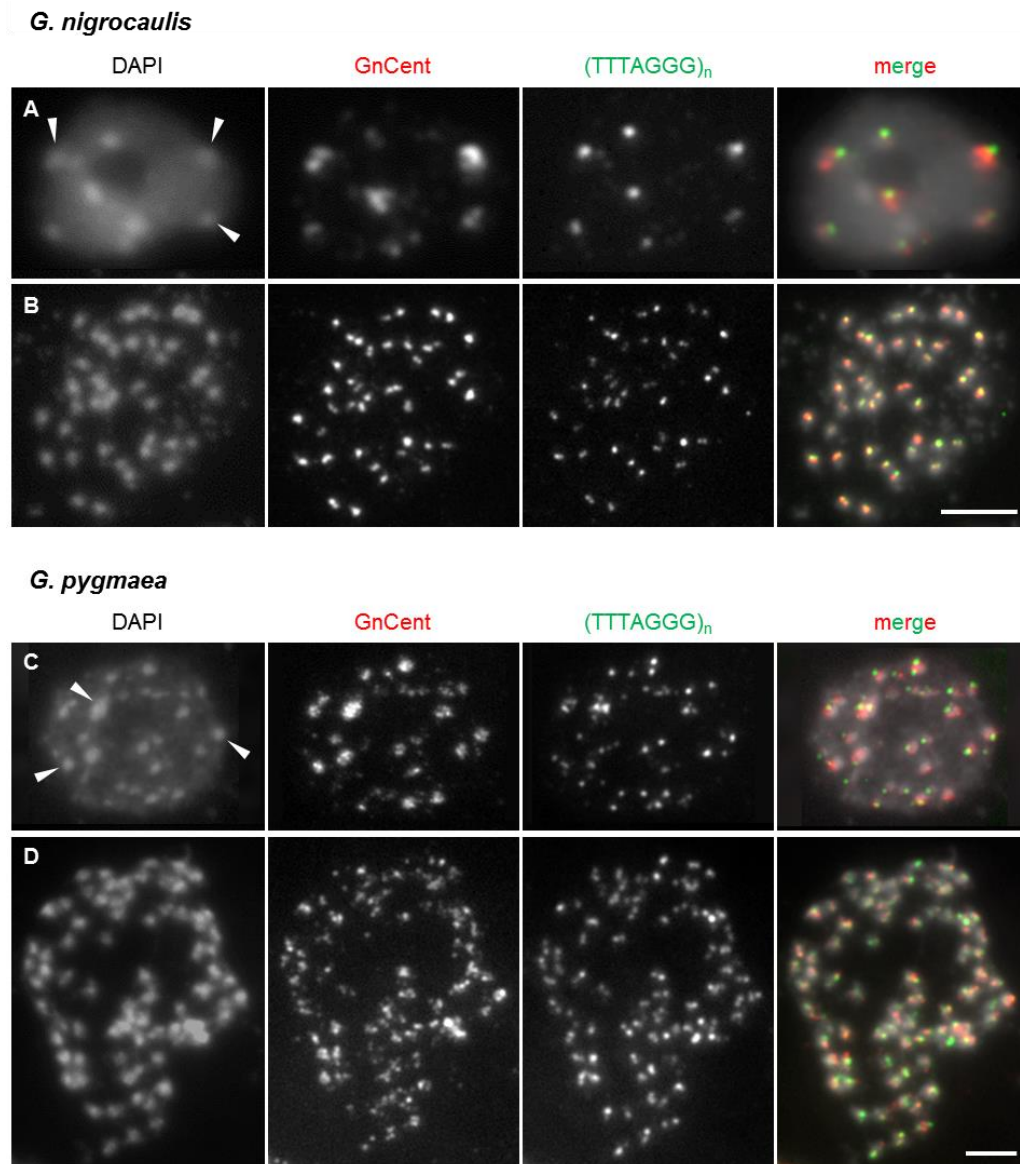


Figure 15: Localization of centromere- and telomere-specific repeats in *G. nigrocaulis* and *G. pygmaea*.

FISH with centromeric GnCent and Arabidopsis-type telomere probes resulted in signal clusters co-localizing with the intensely DAPI-stained chromocenter (white arrow heads) in *G. nigrocaulis* (A) and *G. pygmaea* (C). The two probes also yield adjacent hybridization signals on all metaphase chromosomes of both species (B, D). Bars represent 5 μ m.

The organization of Arabidopsis-type telomere minisatellite at chromosome termini was further investigated by terminal restriction fragment (TRF) analysis. However, due to limitations in the availability of sufficient amount of leaf material of *G. nigrocaulis*, *G. pygmaea* was alternatively used. To detect TRFs and a possible association between telomere and the GnCent satellite sequence in *G. pygmaea* which was prompted by the adjacent FISH signals of these repeats (Figure 15), the BAL31-digested high molecular weight DNA was further digested by the 6-bp cutter *Xba*I and the 8-bp cutter *Sfi*I (none of these have a recognition site in the GnCent repeat). Using the Arabidopsis-type telomere sequence (TAAACCC)₄ as a probe, similar TRF patterns were found with both

restriction enzymes (Figure 16A). Progressive TRF shortening with increasing BAL31 digestion times is visible in the PFGE patterns (left panel) as well as in the blot from conventional electrophoresis (right panel); the latter detecting truncated TRFs diffusing out of the agarose plugs after BAL31 digestion. This data confirmed that the typical (TTTAGGG) satellite represents terminal telomeric DNA in *G. pygmaea*. Re-probing of the membranes with GnCent repeat revealed similar (but slightly stronger) hybridization patterns as for the telomeric probe (Figure 16B), thus supporting the close association of GnCent repeats and telomeric sequences on chromosomes. Since the restriction enzymes chosen for TRF analysis in *G. pygmaea* produce long fragments (comprising presumably tens of kbp of subtelomeric DNA), an additional restriction enzyme, the 4 bp-cutter *Mbol*, with a recognition site within the GnCent repeat, was used to determine telomere lengths more precisely. TRFs ranging between 10 and 18 kbp were detected, reflecting the telomere lengths in this species (Figure 16C).

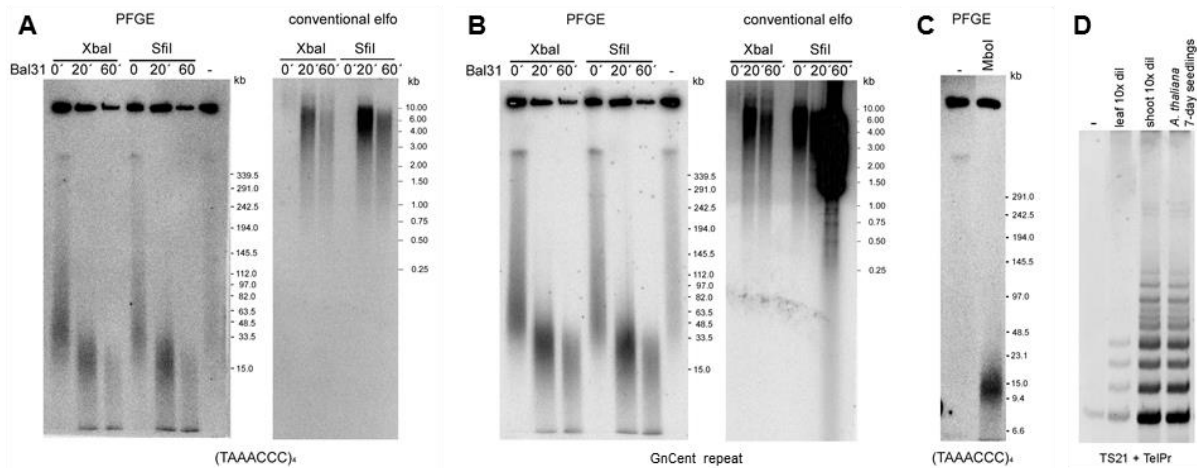


Figure 16: Telomere sequence confirmation and telomerase activity analysis in *G. pygmaea*.

(A-C) BAL31 and TRF analyses. High molecular weight DNA in agarose plugs was digested with restriction enzymes *XbaI* or *SfiI* without (-) or after BAL31 digestion for times indicated above the individual lanes. PFGE denotes DNA hybridization of PFGE-separated DNA retained in agarose plugs after digestion, conventional electrophoresis denotes DNA hybridization of DNA diffused out of the plugs after digestion separated by conventional electrophoresis. Hybridization patterns obtained using the Arabidopsis-type telomere probe (A, C) and the centromeric repeat GnCent (B) are shown (probed sequences are mentioned below each panel). Similar hybridization patterns of PFGE parts of panels A and B demonstrate association of both probed sequences in large TRF fragments, while TRF signal obtained using *Mbol* (after cleavage of the GnCent repeat) provides more precise assessment of genuine telomere lengths (between 10 and 18 kbp, C). Positions of LowRange PFG marker (NEB) and GeneRuler 1 kbp DNA ladder (Fermentas) are indicated on the right. (D) Telomerase activity detected by TRAP assay reveals in *G. pygmaea* regular ladders of telomerase products, similar to *A. thaliana* extract used as positive control. Substrate and reverse primers used in TRAP are given below each panel (see table 5).

To demonstrate the presence of the Arabidopsis-type telomere sequence in *G. pygmaea* by a further independent approach, a TRAP assay was performed using telomerase extracts from either shoots with flower buds or young leaves, and from *A. thaliana* seedlings as a positive control (Figure

16D). Regular ladders of telomerase products were obtained in both extracts of *G. pygmaea*, showing the same periodicity of the ladder bands as the control products of *A. thaliana* telomerase (Figure 16D). Sequences from cloned TRAP products confirmed the presence of regular repeats of (TTTAGGG).

Considered as the typical angiosperm telomeric sequence, the (TTTAGGG) minisatellite has been detected in many vascular plant species (Fuchs *et al.*, 1995). A combination of cytogenetic and molecular approaches demonstrated the presence of this telomeric minisatellite, maintained by telomerase activity in *G. pygmaea* and also in *G. nigrocaulis*. The telomere lengths ranging from 10 to 18 kbp as revealed by TRF analysis is longer than those reported for species possessing similar or even larger genomes and chromosome sizes such as *A. thaliana* and other Brassicaceae (Shakirov *et al.*, 2008) and apparently correlate with the corresponding strong FISH signals (Figures 15B, D).

3.2.4. In *G. hispidula* the Arabidopsis-type telomere repeat is replaced by two sequence variants

FISH analysis in *G. hispidula*, surprisingly, could not detect the Arabidopsis-type telomere-specific signal on nuclei as well as metaphase chromosomes. Among genomic sequencing reads of *G. hispidula*, the (TTTAGGG) minisatellite was only occasionally found as short interstitial fragments. Alternatively, two variant repeat motifs of either six bp (TTCAGG) or seven bp (TTTCAGG) could be identified as putative telomere sequences. FISH using these motifs as probes indicated their terminal location on metaphase chromosomes of *G. hispidula* (Figure 17A). Fiber FISH experiment further demonstrated that both sequences are intermingled (Figure 17B).

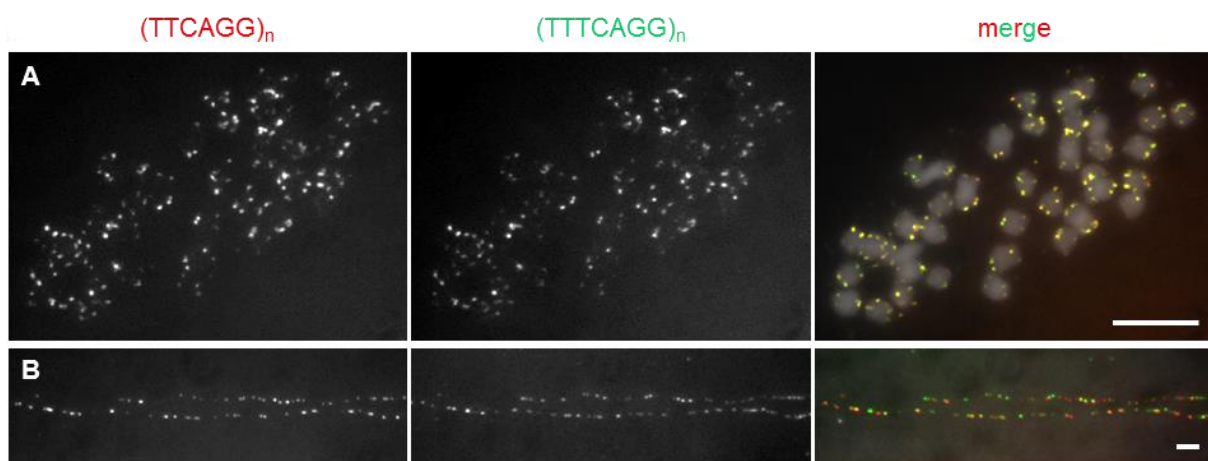


Figure 17: Cytogenetic characterization of two novel telomere variants of *G. hispidula*.

(A) FISH analysis detected the six bp (TTCAGG) and seven bp (TTTCAGG) repeats on chromosome ends. **(B)** Dual-color fiber FISH further revealed the intermingling of both variants. Bars represent 5 μ m.

Similar molecular approaches as for *G. pygmaea* were employed for further confirmation of new telomere repeats in *G. hispidula*. Probes of both motifs, (TGAAACC)₄ and (TGAACC)₄, revealed

similar patterns of TRFs ranging from 20 to 97 kbp. Their sensitivity to BAL31 digestion proved the terminal chromosome position of the hybridizing fragments (Figures 18A, B). During the longest BAL31 exposure, the truncated TRFs diffuse from agarose plugs and can be detected in the DNA fraction obtained from the reaction solution (Figures 18A, B, right panels). No signal could be observed when re-probing the membranes with the Arabidopsis-type telomere sequence (TAAACCC)₄.

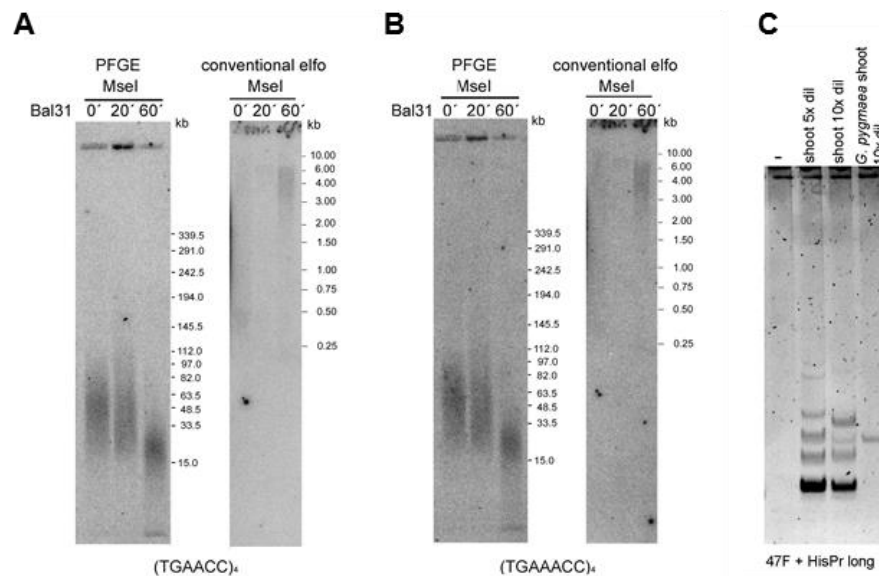


Figure 18: Confirmation of the telomere identity of two novel sequence variants in *G. hispidula*.

(A-B) BAL31 and TRF analyses. High molecular weight DNA in agarose plugs was digested with *MseI* restriction enzyme after BAL31 digestion for times indicated above individual lanes. PFGE denotes DNA hybridization of PFGE-separated DNA retained in agarose plugs after digestion; conventional electrophoresis denotes DNA hybridization of DNA diffused out of the plugs after digestion, separated by conventional electrophoresis. Hybridization patterns obtained using the short (A) and the long (B) variants of telomeric repeat units are shown (probed sequences are mentioned below each panel). Positions of MidRange PFG marker (NEB) and GeneRuler 1 kbp DNA ladder (Fermentas) are indicated on the right. (C) Telomerase activity detection in *G. hispidula* using TRAP assay. While *G. hispidula* extracts produce irregular ladders of telomerase products, *G. pygmaea* extract shows only a single band of non-specific product with *G. hispidula*-specific substrate and reverse primers which are given below the panel (see table 5).

The identity of the two new telomeric repeats was further investigated using the TRAP assay with *G. hispidula* telomerase extracts and subsequent sequencing of the resulting products. A number of combinations of substrate and reverse primers have been tested, but positive results were only obtained when the primer combinations 47F as substrate and either HisPr long or HisPr short as reverse primer (see table 5) were used. Finally, the primer for the longer repeat variant, HisPr long, yielding well reproducible results, was chosen for the TRAP assay (Figure 18C). No product was obtained in the reaction without protein extract and in the extract from *G. pygmaea* only a single, most likely unspecific band was detected. *G. hispidula* displayed irregular ladders, presumably

due to alternation of long and short variants of the synthesized telomere repeats (Figure 18C). Sequencing of the products demonstrated co-occurrence of both variants in all clones that contained more than one sequenced repeat unit. While it is not clear whether the observed low processivity of *G. hispidula* telomerase is due to *in vitro* conditions, or is an intrinsic feature of this telomerase, the observed alternation of both telomeric sequence variants confirmed next-generation sequencing identification, as well as the co-localization of both sequences observed in FISH and TRF analyses.

3.2.5. Telomeric repeat variation in *Genlisea* is the first example for an intrageneric switch

Investigations for telomeres of various eukaryotic groups have identified only a limited number of minisatellite types which are conserved within the groups of organism, for example, (TTAGGG) in vertebrates and fungi (vertebrate-type; Meyne *et al.*, 1989), (TTTAGGG) in most plants (Arabidopsis-type; Richards & Ausubel, 1988), or (TTAGG) in insect (Mohan *et al.*, 2011). Nevertheless, variations of telomeric minisatellite are observed within some particular groups. For instance, the most common Arabidopsis-type telomere is not ubiquitous among plants. First evidence came from FISH with the (TTTAGGG)_n probe on chromosomes of several species of the genus *Allium* (Fuchs *et al.*, 1995). Pich *et al.* (1996) and Pich and Schubert (1998) confirmed the absence of (TTTAGGG)_n and suggested that highly repetitive satellite and/or rDNA sequences substituted the original telomeric sequence at the very ends of the *Allium* chromosomes. They also suggested a conversion-like mechanism for compensation of the replication-mediated end-shortening. Later, Adams *et al.* (2000) reported the lack of Arabidopsis-type telomeres in *Aloe*, another genus of the order Asparagales to which *Allium* belongs. A large-scale screen using different minisatellite sequences identified the vertebrate-type telomere as the predominant telomeric repeat, occasionally associated with (TTTAGGG) repeats, at the chromosome termini of Asparagales species which descend from an evolutionary switch-point leading to an extensive replacement of the Arabidopsis motif (Sykorova *et al.*, 2006). A second switch-point (inside the Alliaceae family) towards the evolution of the genus *Allium* is responsible for the loss of any of the known telomere minisatellites, and apparently also of an active telomerase gene in this genus (Fajkus *et al.*, 2005). Moreover three genera of the Solanaceae family, *Vestia*, *Sessia* and *Cestrum*, lack Arabidopsis- and vertebrate-type telomeres, but revealed terminal and interstitial A/T rich minisatellites (Sykorova *et al.*, 2003b; Sykorova *et al.*, 2003c). Recently, an A/T rich (TTTTTTAGGG) repeat was characterized as the telomeric sequence of *Cestrum elegans* (Peska *et al.*, 2015). However, it still needs to be elucidated whether this non-canonical telomeric repeat is also shared by the two sister genera *Vestia* and *Sessea*.

Interestingly, variation in telomeric minisatellite was found within the genus *Genlisea*. Characterized by FISH and confirmed by TRF analysis and TRAP assay, *G. nigrocaulis* and its closely

related *G. pygmaea* possess the Arabidopsis-type telomeric minisatellite at their chromosome ends. On the other hand, in the much larger genome of *G. hispidula*, and in that of its close relative *G. subglabra*, this sequence motif is replaced by two hitherto unknown telomeric sequence variants (Figure 19A). Since the Arabidopsis-type telomeric repeat was detected in other species of the subgenus *Genlisea*, e.g. *G. margaretae* and *G. aurea* (Figures 19B, C), the switch of telomeric repeats had happened at least in the common ancestor of *G. hispidula* and *G. subglabra*.

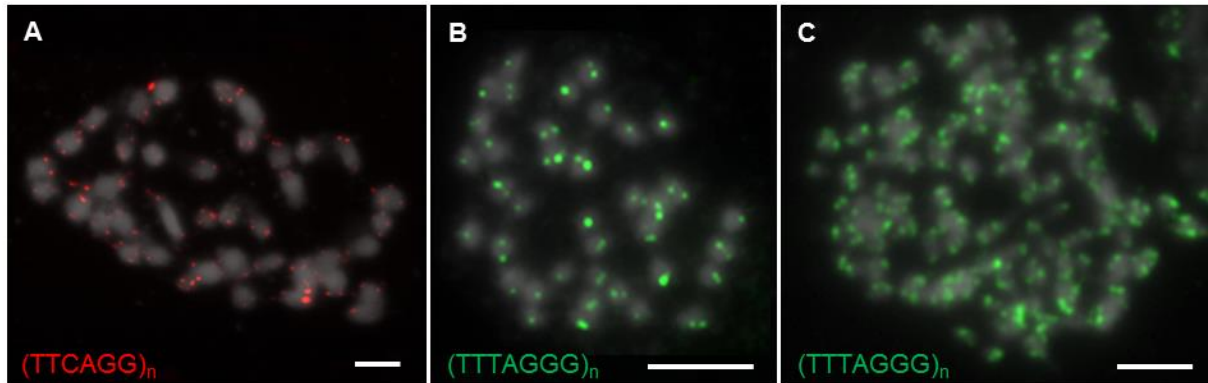


Figure 19: Cytogenetic characterization of telomeric repeats in *G. subglabra*, *G. margaretae* and *G. aurea*.

The six bp telomere variant of *G. hispidula* shows signals located at chromosome ends of the close relative *G. subglabra* (A) whereas the Arabidopsis-type telomere repeat was detected on metaphase chromosomes of *G. margaretae* (B) and tetraploid *G. aurea* (C). Bars represent 5 μm .

So far, alterations of telomeric sequences are mainly in accordance with the $T_nA_mG_o$ sequence (Fulnečková *et al.*, 2013). In Asparagales, the change from the Arabidopsis-type (TTTAGGG) to the vertebrate-type (TTAGGG) telomere can be explained by point mutations in the telomerase catalytic protein subunit (TERT) which affect the telomere-telomerase interaction and result in a variant usage of template region in the telomerase RNA subunit (TR) and consequently in T-slippage (Sykorova *et al.*, 2006). In *Genlisea* however, the variant motifs observed in *G. hispidula* cannot be derived from the presumed ancestral Arabidopsis-type telomere sequence by less than two mutations. In particular, the presence of cytosine in the G-rich strand of both sequence variants is remarkable. This change is most likely associated with a corresponding mutation directly in the template region of the *G. hispidula* TR subunit. Variants having a C within the G-rich strand were so far found only in some insects (Mravinac *et al.*, 2011), and were hitherto not described for plants. *Genlisea* represents the first example of an evolutionary switch of telomeric sequences within a genus (Figure 20). This switch occurs in parallel with an apparently high mutation rate (Ibarra-Laclette *et al.*, 2013) and a fast genome size evolution within the genus *Genlisea* (Vu *et al.*, 2015). However, whether an association between the alteration of the telomeric repeat and genome enlargement discovered in the section *Africanae* exists remains uncertain. Analyses of telomeric sequences in related taxa within this section may help to answer this question.

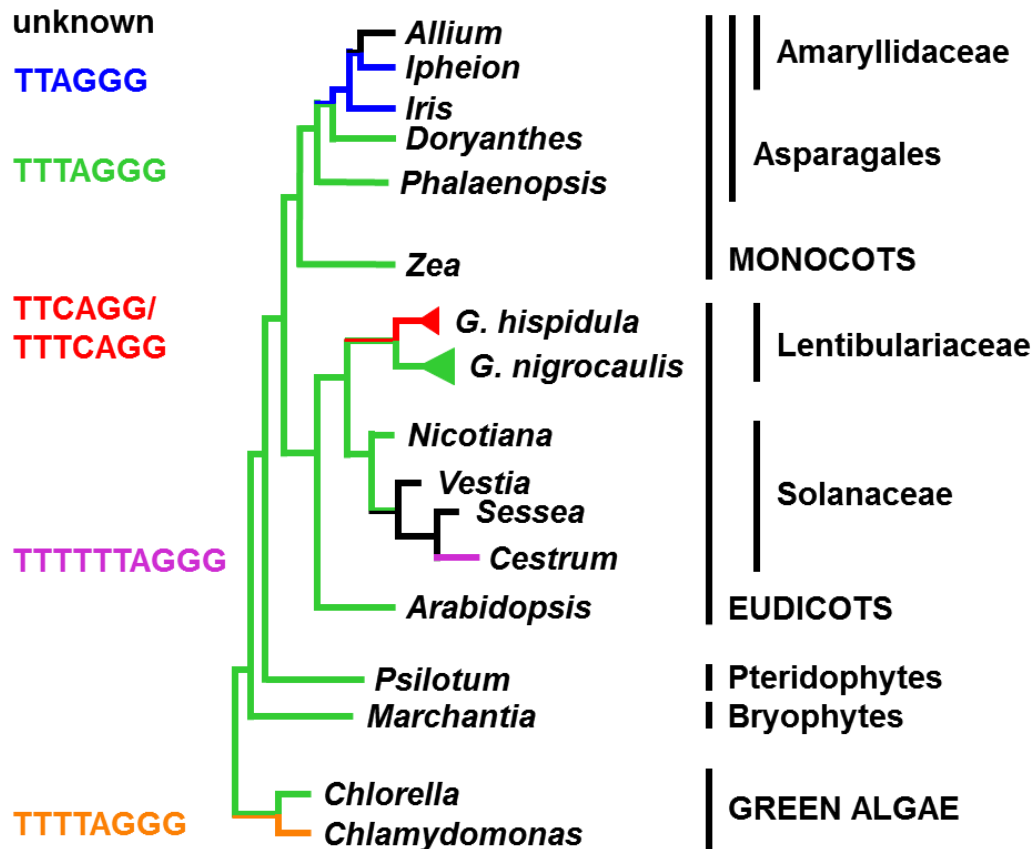


Figure 20: Simplified phylogenetic tree of green plants indicating telomeric sequence deviations from TTTAGGG, the most wide-spread telomeric repeat among plants.

The tree was modified according to Sykorova *et al.* (2006) and Peska *et al.* (2015).

3.3. FISH-based chromosome identification, a prerequisite to study karyotype evolution in *Genlisea*

Cytogenetic studies that aim to discriminate every individual metaphase chromosome represent a fundamental step of karyotyping. Karyotyping by chromosome size and morphology and/or chromosome-specific bands is relatively easy for species possessing large chromosomes but is a challenge for species with small and similar chromosomes. This limitation can be overcome by means of FISH using chromosome-specific sequences as probes. From genomic sequence data, single-copy and tandem-repetitive sequences can be identified and employed for FISH-based chromosome identification. Large-insert BAC clones which are largely free of repetitive DNA are another potential source for chromosome-specific probes. In the following, initial chromosome identification for representative species of the three sections of subgenus *Genlisea* using either single-copy sequences, tandem repeat DNAs or repeat-free BAC clones are provided and discussed.

3.3.1. In combination with rDNA probes, single-copy FISH discriminated 11 chromosome pairs of *G. nigrocaulis* and revealed homeologous chromosomes of *G. pygmaea*

Only few tandem repeats could be found in the *G. nigrocaulis* genome. However, from randomly selected contigs which were assembled from genomic sequences, 15 single-copy fragments were identified and used to design specific primers. The uniqueness of the identified single-copy fragments was tested by a two-step screening including PCR amplification and subsequent FISH on flow-sorted nuclei. Finally, ten putative single-copy fragments (Table 3) generating one pair of FISH signals on each metaphase plate of *G. nigrocaulis* (Figure 21) were selected and used as probes for karyotyping.

Multicolor FISH and sequential re-probing on the same cytological preparation are methods of choice in order to increase the number of chromosomes to be simultaneously identified (Shibata *et al.*, 2009). Using this approach, two major aspects need to be considered practically: *i*) the intactness of chromosome morphology after probe-stripping, and *ii*) the order of probes to be used in sequential FISH regarding their chromosomal distribution (cluster or dispersed), signal intensity and nature (single-copy *versus* repetitive) of probes.

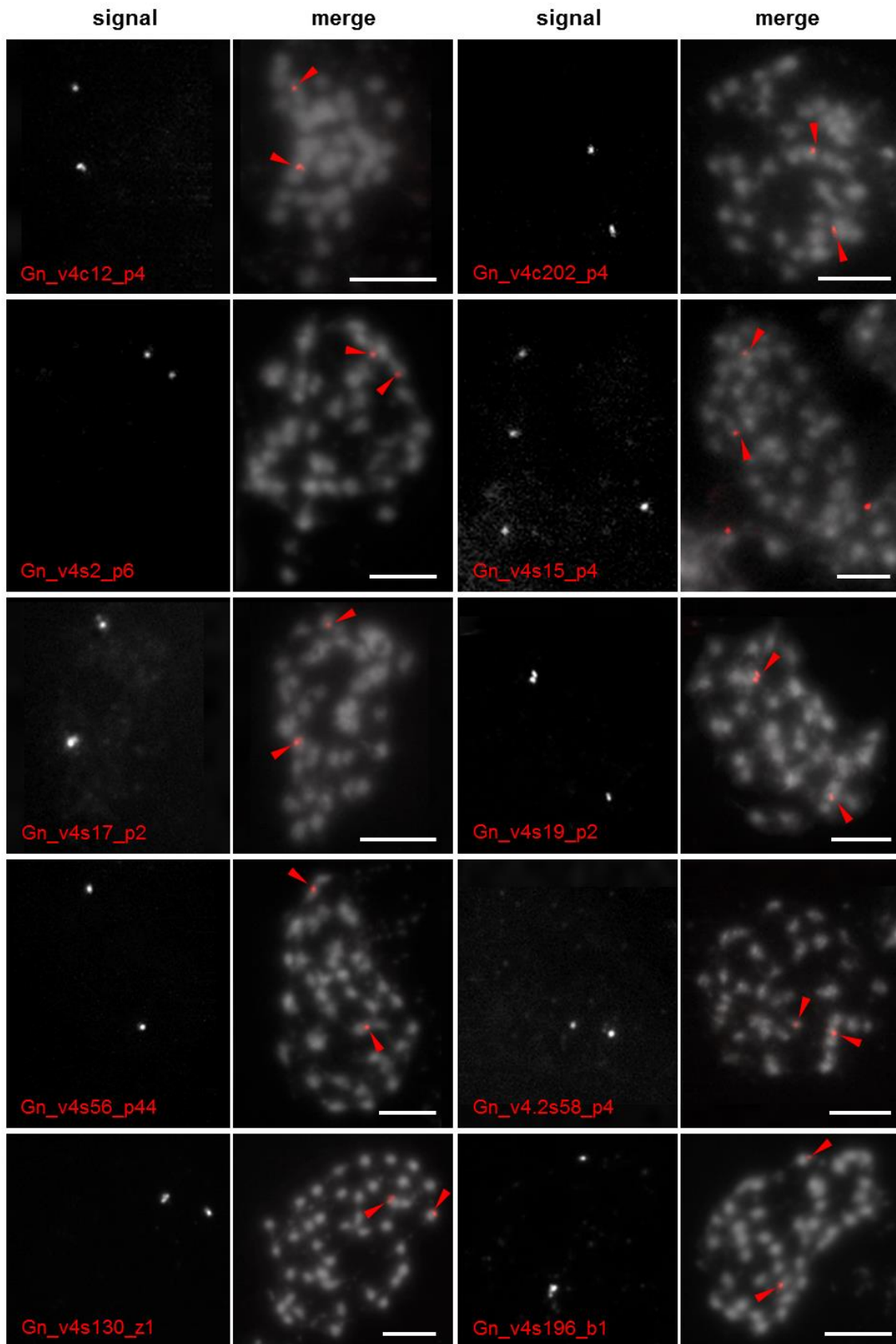


Figure 21: Localization of ten single-copy fragments by FISH on metaphase plates of *G. nigrocaulis*. Note that additional FISH signals of Gn_v4s15_p4 (without red arrow heads) are those on the surrounding interphase nuclei. Bars represent 5 μ m.

The very small chromosomes of *G. nigrocaulis* were found to lose their morphological intactness, as indicated by weaker and fuzzy DAPI staining signals (Figure 4), already after one round of probe-stripping under conditions that remove all hybridized probes of the first FISH experiment (see Materials and Methods). Therefore, a lower temperature (40°C) was used for probe-stripping, which removes the hybridized probes only partly, but minimizes chromosome deteriorations. In order to achieve optimal results in chromosome individualization in *G. nigrocaulis*, nine unique probes were randomly divided into three pools (pool 1, 2 and 3) for the first round of three-color FISH. The separate chromosomal localizations of the three probes of each pool were confirmed in advance by FISH on metaphase chromosomes (Figures 22A, B, C). After removing the unique probes, a second FISH was carried out on the same preparation using the 45S and 5S rDNA probes which yield stronger hybridization signals in comparison to the relatively weak signals from the unique probes (Figures 22D, E). Totally, 11 chromosome pairs could be discriminated by nine unique probes in combination with rDNA probes in *G. nigrocaulis* (Figure 22F).

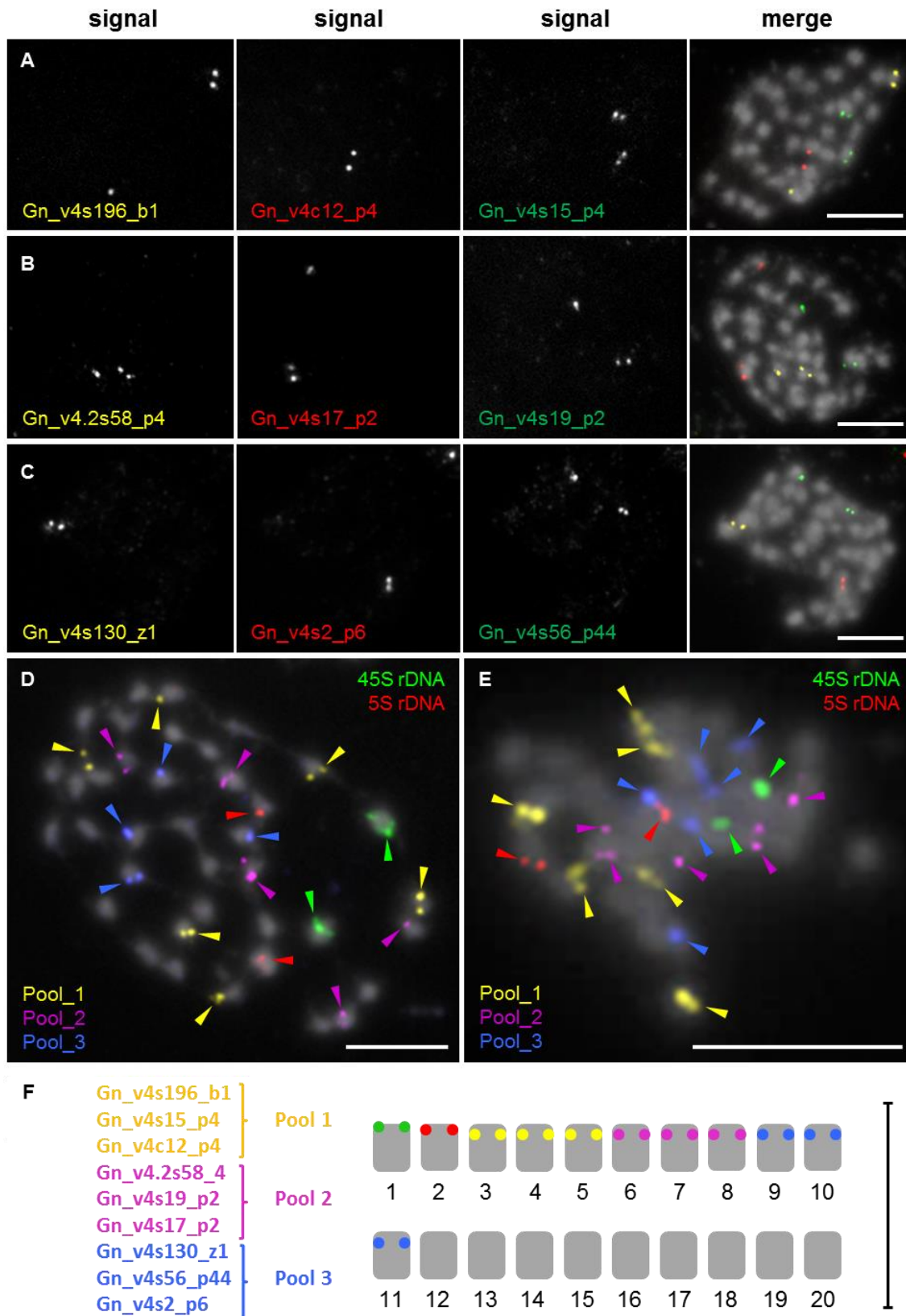


Figure 22: FISH-based chromosome identification using unique and rDNA probes in *G. nigrocaulis*.

Nine unique probes were randomly divided into three pools. The chromosomal separation of the individual probes were confirmed for pool_1 (A), pool_2 (B) and pool_3 (C). (D, E) Sequential FISH-based karyotyping using three pools of unique probes in conjunction with two rDNA probes. Six and five FISH signals of pool_1 and

pool_2, respectively, are shown in both (D) and (E). The missing of two FISH signals of pool_3 in (D) is complemented in (E) indicating the proper hybridization of all nine unique probes. Two metaphase cells presented in (D) and (E) are from the same cytological preparation. (F) Karyogram of *G. nigrocaulis* with eleven distinguishable chromosome pairs. Bars represent 5 μ m.

The colinearity and similarity of single-copy fragments between two species may differ and depend on their relationship. To test whether the ten single-copy fragments derived from *G. nigrocaulis* can be used to identify homeologous chromosomes of other *Genlisea* species, cross-FISH analyses were performed in *G. hispidula*, *G. margaretae* and *G. pygmaea*. None of these ten *G. nigrocaulis*-derived unique probes revealed FISH signals neither on interphase nuclei nor metaphase chromosomes of *G. hispidula*. The more distant relationship and an 18-fold larger genome of *G. hispidula*, in that single-copy DNA fragments are often interspersed by repetitive sequences, obviously cause the negative cross-FISH results. Unexpectedly, no hybridization signals were revealed by this set of unique probes in *G. margaretae*, although having a closer relationship to *G. nigrocaulis* and a smaller genome size than *G. hispidula*. Possibly, the genome of *G. margaretae* contains a higher proportion of repetitive sequences that are interspersed between single-copy sequences than *G. nigrocaulis*. This assumption is in line with the homogenous nuclear phenotype which showed distribution of heterochromatin-associated epigenetic marks resembling that of large genomes such as those of *G. hispidula* and *G. subglabra*. On the other hand, nine of ten tested unique probes revealed distinct FISH signals on two metaphase chromosome pairs of *G. pygmaea* (Figure 23). This result not only reflects the close relation between *G. pygmaea* and *G. nigrocaulis* but also confirms the tetraploid status of *G. pygmaea* as indicated by its genome size and chromosome number. It is also known that many polyploid genomes eventually return to a diploid-like state through loss or divergence of duplicated genes. Buggs *et al.* (2009) surveyed ten sets of homeologous genes in 57 individual of allotetraploid *Tragopogon miscellus* and found 18 cases of homeologous loss (3.2% of homeologous pairs studied). Because these losses were undetected in first-generations of synthetic allopolyploids and were randomly found in natural populations, the authors assumed that homeologous loss is an ongoing consequence during last 80 years after polyploidization of *Tragopogon miscellus*. Interestingly, such a loss of a homeologous locus was also found in *G. pygmaea* since one unique probe (Gn_v4s56_p44) yielded only one pair of FISH signals per metaphase plate (Figure 24). Similarly, rDNA probes also revealed a “diploid-like” locus number (one chromosome pair for each 45S and 5S cluster, figure 8B) in *G. pygmaea*. To decide whether the WGD occurred after splitting from the most recent common ancestor shared with *G. nigrocaulis* or after the split from the closest relative *G. filiformis* (Figure 3C), further information about genome size and chromosome number of *G. filiformis* is required.

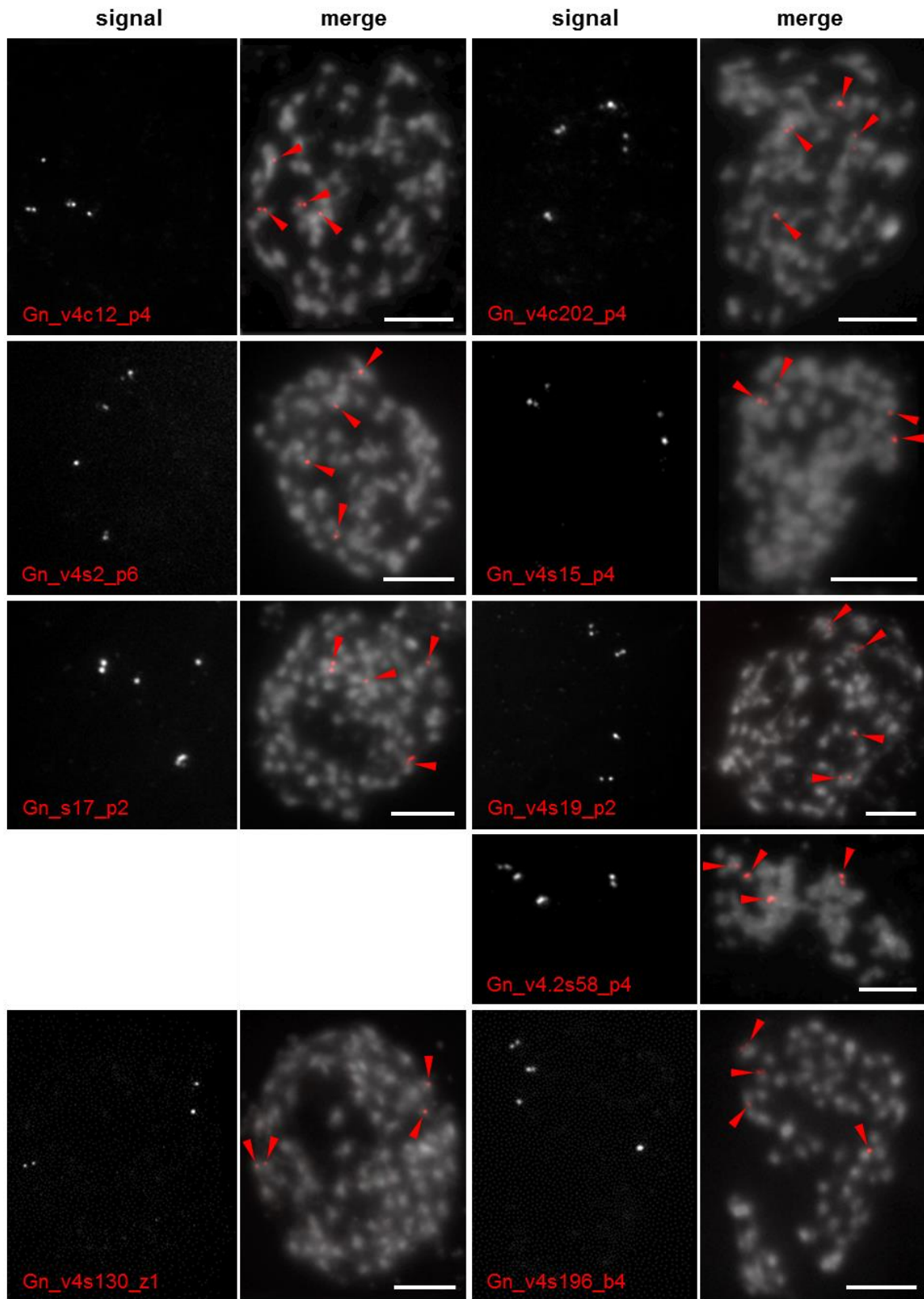


Figure 23: Cross-FISH using nine of ten single-copy probes derived from *G. nigrocaulis* to chromosomes of *G. pygmaea*.

Each of nine unique probes yields two pairs of FISH signals strongly supporting the tetraploidy of *G. pygmaea*. Bars represent 5 μm .

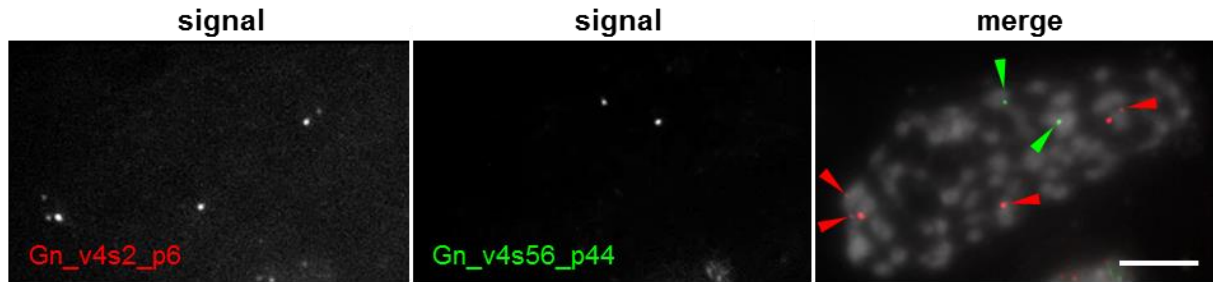


Figure 24: Cross-FISH revealed the loss of a homeologous locus in tetraploid *G. pygmaea*.

Whereas a unique probe derived from *G. nigrocaulis* yields two pairs of FISH signals (Gn_v4s2_p6 in red), another unique probe (Gn_v4s56_p44 in green) labels only one chromosome pair in *G. pygmaea* as in *G. nigrocaulis* (Figure 21). Bar represents 5 μ m.

3.3.2. Thirteen chromosome pairs were distinguishable by different types of tandem repeats in *G. hispidula*

Ribosomal DNAs together with tandem repeats that usually accumulate at specific chromosomal positions are widely employed as FISH-based anchors not only for chromosome individualization but also for sub-genome identification analysis (Young *et al.*, 2012; Begum *et al.*, 2013; Cuadrado *et al.*, 2013; Nemeth *et al.*, 2013; Shibata *et al.*, 2013). More than half of the large genome of *G. hispidula* is characterized by repetitive sequences, most of them are TEs. Among only 1.96% genomic sequences classified as tandem repeats, six repeats were identified and five of them revealed chromosome-specific distribution which facilitates the discrimination of *G. hispidula* chromosomes (Figures 10B-G).

To increase the number of chromosome pairs to be identified in *G. hispidula*, the 45S and 5S rDNA were combined with four tandem repeats (the Gh250c46, Gh45c31, Gh80c174 and Gh19c56) showing clustered chromosomal distributions and the Gh336c35 labeling half of 20 chromosome pairs. Fortunately, the larger chromosomes of *G. hispidula*, in comparison to those of *G. nigrocaulis*, enable several probe-stripping treatments accompanied by only little chromosome degradation. All seven repetitive probes were divided in to three probe cocktails. The ribosomal DNA probes revealing higher signal intensities were included in the second cocktail while the dispersed repeat Gh336c35 was the last probe to be hybridized. After three rounds of sequential FISH, all repetitive probes allow 13 chromosome pairs of *G. hispidula* to be individualized (Figure 25).

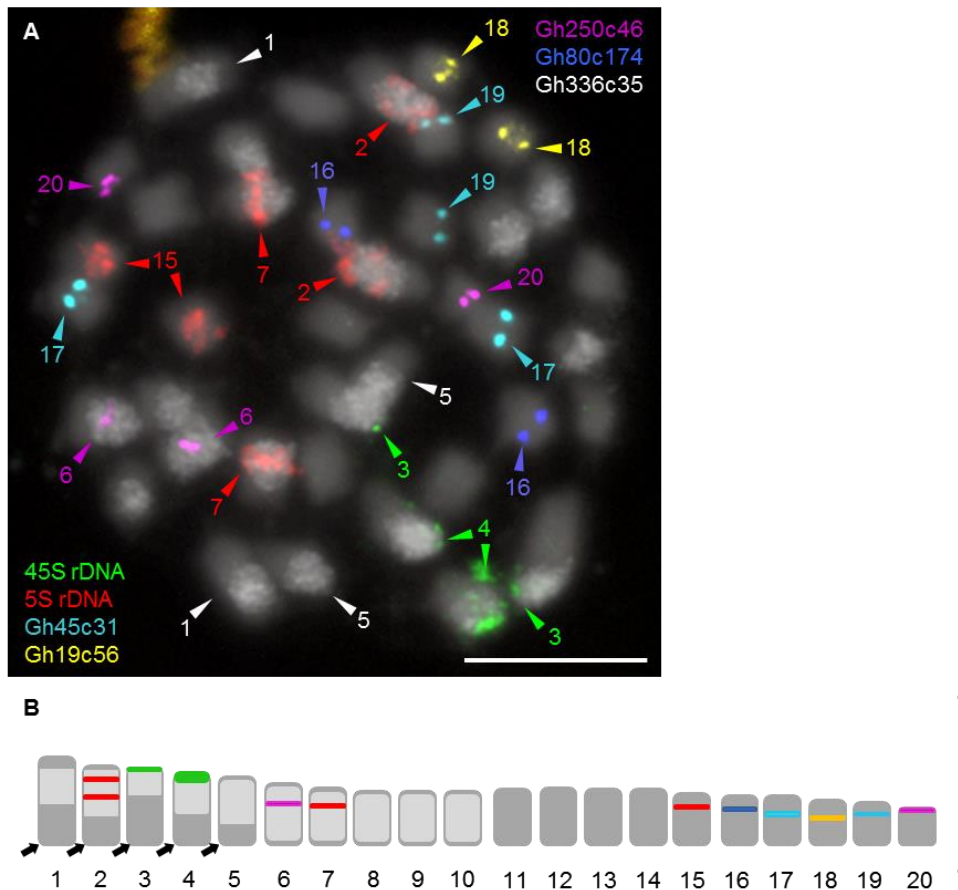


Figure 25: FISH-based chromosome identification using tandem repeat probes in *G. hispidula*.

(A) Metaphase after sequential FISH with seven tandem repeat sequences including the rDNA probes allowing the unequivocal discrimination of 13 chromosome pairs (arrow heads with numbers). (B) Karyogram of *G. hispidula* indicating the 13 distinguishable chromosome pairs (1-7, 15-20) based on the examination of seven metaphase plates; chromosome numbers and colors of markers correspond to those denoted in (A). The black arrows indicate regions free of Gh336c35 that are assumed to result from reciprocal exchanges between homeologous chromosomes of the parental genomes and subsequently biased segregation of translocation products. Bars represent 5 μ m.

The allotetraploidy of *G. hispidula* which was indicated by comparative genomic analysis (Vu *et al.*, 2015) is further supported by FISH signals of the repetitive sequence probe Gh336c35 distinguishing two sub-genomes by the higher accumulation on 10 of the 20 chromosome pairs. Due to the fact that *G. hispidula* has the same chromosome number of $2n = 40$ as *G. nigrocaulis*, a dysploid chromosome number reduction has to be assumed for both ancestor species, or after whole genome duplication in *G. hispidula*, similarly as shown for Australian Brassicaceae species (Mandakova *et al.*, 2010). However, the FISH signals for Gh336c35 suggest that dysploid chromosome number reduction might have already occurred within the ancestral species of the *G. hispidula* lineage. Such chromosome number reductions are often the result of reciprocal translocations with terminal breakpoints which combine two linkage groups into one large chromosome, while the second translocation product is very small and prone to get lost during

meiosis (Schubert & Lysak, 2011) (Figure 1). The absence of Gh336c35 signals in some terminal regions of chromosomes 1, 2, 3, 4 and 5 (Figures 25B), might be due to reciprocal exchanges between homeologs (among chromosomes 11-20) and subsequent segregation bias (Wicker *et al.*, 2015) within allotetraploid *G. hispidula* (Figure 26). Alternatively, the exchanged segments were of unequal size, and very small Gh336c35-rich regions, transferred to chromosomes of the originally Gh336c35 signal-free complement, are not detectable by FISH.

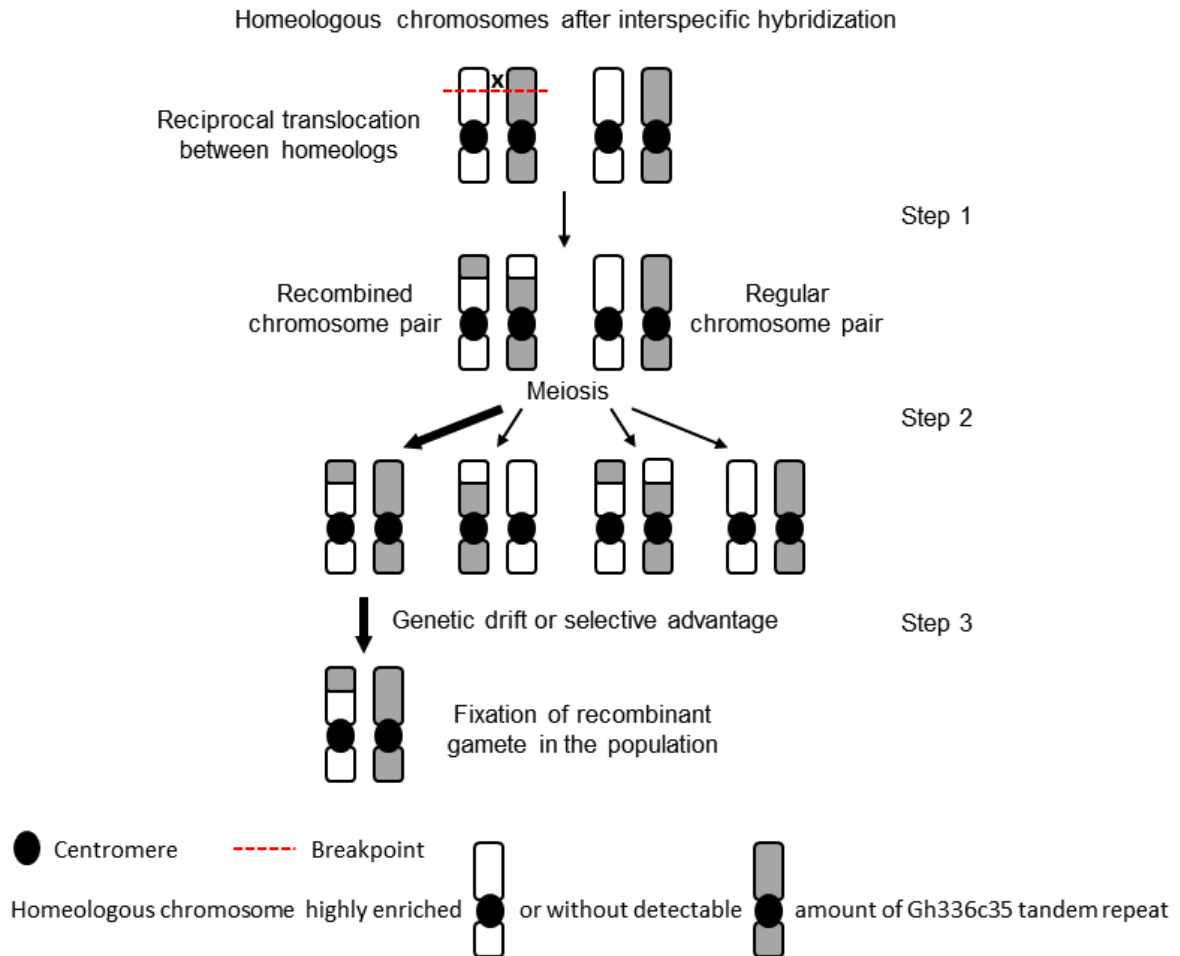


Figure 26: Sequence exchange between homeologous chromosomes after interspecific hybridization in *G. hispidula*.

Dyploid chromosome number reductions (such as those illustrated in figure 1) might have led to the ancestral karyotypes with $2n = 20$ chromosomes of *G. hispidula*. Interspecific hybridization between these ancestors, likely followed by sequence exchange between homeologous chromosomes, established the allotetraploid *G. hispidula* with $2n = 40$ chromosomes. The chromosome pairs represent one ancestral genome with highly enriched Gh336c35 tandem repeat and the other largely lacking this repeat. After a symmetric reciprocal translocation between homeologs (step 1), segregation may produce four possible combinations of these chromosome pairs in meiosis (step 2). One of the possible gametes (left) was preferentially fixed and led to the karyotype of *G. hispidula* (step 3). The figure was adapted from Wicker *et al.* (2015).

3.3.3. FISH-based chromosome identification using repeat-free BAC clones identified 15 of the 19 chromosome pairs of *G. margaretae*

Bacterial artificial chromosome (BAC) clones may carry up to 200 kbp inserted genomic DNA fragments. Dispersed repetitive DNA sequences, particularly TEs, if present in BAC clone, will hybridize genome-wide when applied as FISH probe. Screening for BAC clones that do not contain repetitive DNA sequences is thus a prerequisite to develop chromosome-specific BAC probes. Such “contaminating” repetitive sequences can be specifically removed from BAC clone *e.g.* by annealing to an excess of *Cot-1* DNA followed by digestion of all double-stranded elements by duplex-specific nuclease (Swennenhuis *et al.*, 2012). Another approach is using single-copy fragments that can be identified and amplified from a particular BAC clone instead of using the whole BAC (Danilova & Birchler, 2008). Alternatively, for species with a small genome and a low proportion of repetitive sequences such as *A. thaliana*, repeat-free BAC clones can be selected by means of dot-blot colony hybridization with genomic DNA probes (Lysak *et al.*, 2003).

A *G. margaretae* BAC library, consisting of 12,288 clones, was arranged in 32 different 384-well microtiter plates (in cooperation with Dr. H. Simkova). Based on an average insert size of 83 kbp, the BAC library represents approximately 5.4 equivalents of the *G. margaretae* genome. So far, a reference genome sequence of *G. margaretae*, which would facilitate the computational screening for repeat content of BAC clones, is lacking. Nevertheless, possessing a small genome similar in size to that of *A. thaliana* (184 Mbp/1C versus 157 Mbp/1C), dot-blot screening was expected to be employed successfully to develop chromosome-specific repeat-free BAC clones for *G. margaretae*. In order to thoroughly evaluate the repetitive content of BAC clones, two subsequent rounds of DNA dot-blot hybridization using radioactively labeled genomic probes were performed in cooperation with Dr. R. Schmidt. In the first screening (Figure 27A), exposing the membranes for one day, BACs containing a medium to high proportion of repetitive DNA were identified by strong hybridization signals. From the remaining BACs with weak hybridization signals, a collection of 376 BAC clones was randomly chosen for a second hybridization. Extending the exposure time to five days, BACs without or bearing only little repetitive DNA could be identified (Figure 27B). Both rounds employed low-stringent conditions for hybridization and washing steps to maximize the detection of repetitive DNA hybridization. Finally, 60 putatively repeat-free BAC clones were selected from the second screening for further insert characterization and probe labeling (Figure 27C).

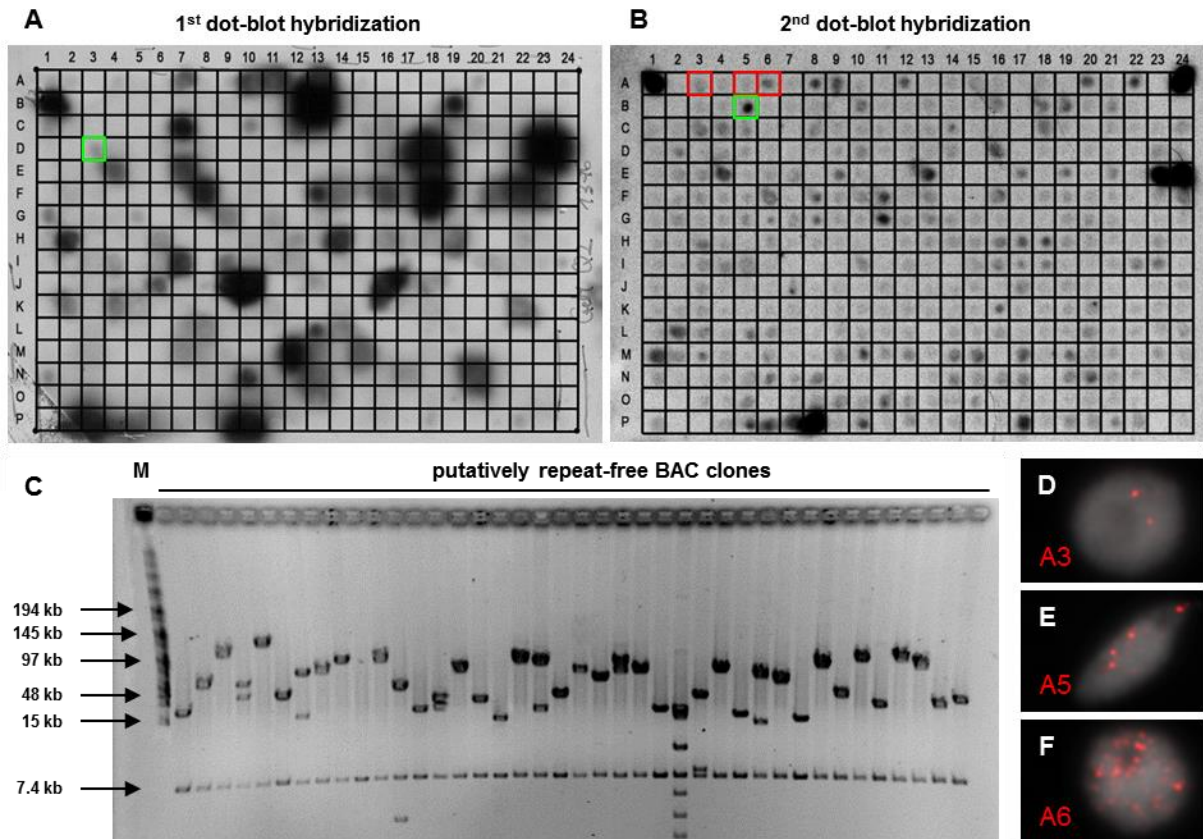


Figure 27: Identification and characterization of repeat-free BAC clones of *G. margaretae*.

(A-B) Two rounds of dot-blot DNA hybridization allow a large-scale examination of repetitive DNA content of *G. margaretae* BAC clones. (A) For the first round of screening, 384-well plates were spotted onto membranes and hybridized to radioactively labeled genomic DNA probe (result of plate 13 is shown). Only BAC clones that showed hybridization signals intensities weaker than that of clone D3 (green square) were selected and inoculated into a new 384-well plate for the second round of screening. (B) 376 BAC clones selected from four original plates after the first screening were further screened for repeat-free BACs. Due to the expected low repetitive content of these BACs, signal development time was prolonged to increase hybridization signal intensity. A1 and A24 are BAC clones that showed medium hybridization signals in the first screening as positive controls. The surrounding wells A2, B1, B2 and A23, B23, B24 are empty as negative controls. BAC clones showing weaker signal intensities than B5 (green square) were considered as putatively repeat-free. (C) DNA from 40 putatively repeat-free BAC clones was *NotI* restricted and the fragments were separated by PFGE. The 7.4s kbp bands in all lanes indicate the pIndigoBAC-5 vector fragments. (D-F) FISH signal patterns on flow-sorted nuclei obtained with BAC clones (red squares in B) that harbor repeat-free, single-copy DNA in (D), repeat-free, low-copy DNA in (E) or repetitive sequences in (F).

Although low-stringency hybridization condition and subsequently longer exposure time for signal development can help to maximize the detection of repetitive DNA hybridization, the putatively repeat-free BAC clones selected by dot-blot DNA hybridization may still contain small fragments of repetitive DNA that are present in high copy number elsewhere in genome. These sequences are potential sources of dispersed FISH signal. Therefore, all 60 putatively repeat-free *G. margaretae* BAC clones were pre-screened for the uniqueness by FISH on flow-sorted nuclei. From 60 putatively repeat-free BACs (Table 7), 32 clones (53.4%) showing two unambiguous FISH signals with

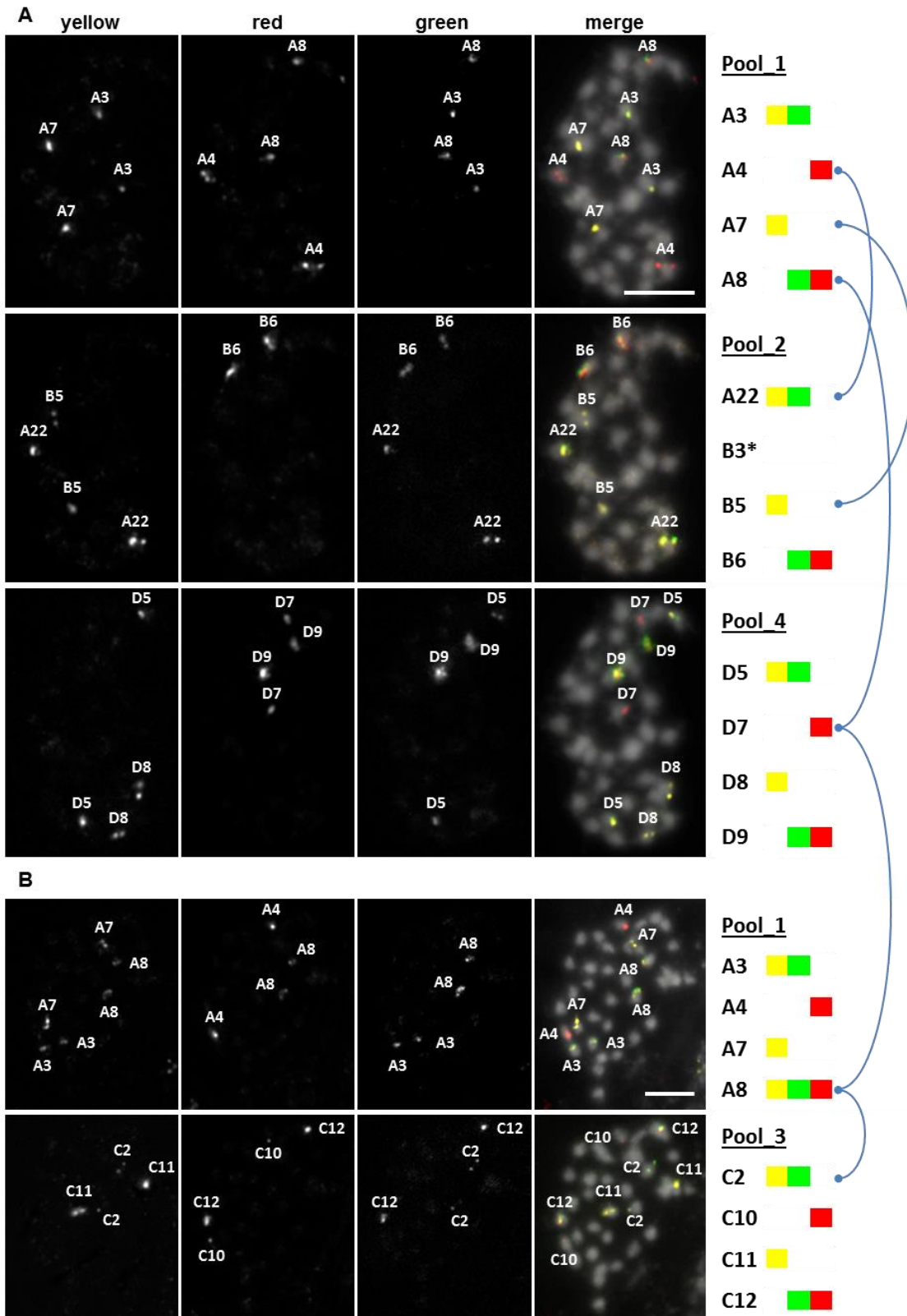
very low or without background on nuclei were considered as suitable for karyotyping (Figure 27D). Seventeen clones (28.3%) yielded more than two distinct signals which may indicate that such clones contain low copy DNA fragments or hybrid inserts locating at more than one locus in the *G. margaretae* genome (Figure 27E). The 11 remaining clones (18.3%) revealed dispersed FISH signals throughout the nuclei (Figure 27F). These BACs apparently harbor short fragments of high-copy repetitive DNA that could not be detected by dot-blot DNA hybridization and thus were excluded from further analysis. The obtained results proved that dot-blot DNA hybridization is a useful method to prescreen for repeat-free BACs in *G. margaretae*.

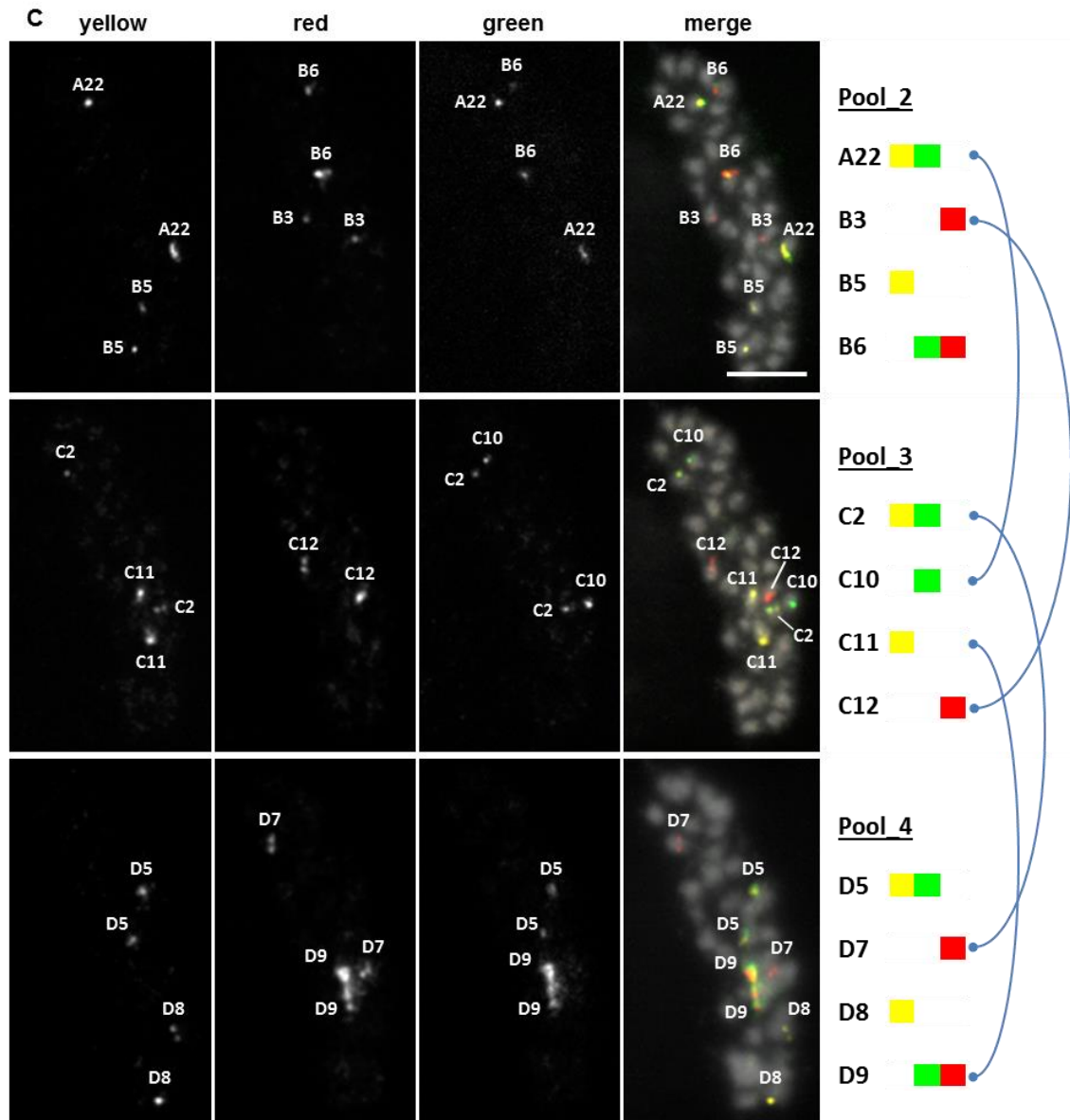
Table 7: Estimated sizes and FISH signal patterns for 60 putatively repeat-free BAC clones of *G. margaretae* selected by dot-blot DNA hybridization.

The gray shaded BAC clones were chosen for chromosome identification analysis. (+) one pair of distinct signals (Figure 27D); (++) more than two distinct signals (Figure 27E); (+++) dispersed signals (Figure 27F) on flow-sorted nuclei.

BAC ID	Estimated size (kbp)	FISH pattern	BAC ID	Estimated size (kbp)	FISH pattern	BAC ID	Estimated size (kbp)	FISH pattern
A3	163	+	B6	112	+	C16	83	++
A4	65	+	B7	83	++	C17	70	+
A5	63	++	B8	130	+++	C20	170	++
A6	135	+++	B10	63	+	C22	52	++
A7	130	+	B11	45	+++	C23	30	+++
A8	227	+	B12	78	+	C24	35	+++
A9	120	+++	B14	82	++	D1	47	++
A10	130	++	B15	100	+	D2	85	++
A12	60	++	B16	35	+++	D3	17	+
A13	140	+	B17	97	++	D4	97	+
A14	70	++	B22	100	+	D5	63	+
A15	82	++	C1	53	+++	D6	15	++
A16	83	+++	C2	33	+	D7	97	+
A17	112	+	C6	68	+++	D8	45	+
A18	50	+	C8	90	+++	D9	100	+
A20	130	+	C9	33	++	D10	33	++
A22	140	+	C10	20	+	D15	112	+
B3	20	+	C11	100	+	D21	97	+
B4	50	++	C12	120	+	D22	33	+
B5	48	+	C15	45	+	D24	35	+

The alignment of BAC clones in a so-called BAC tiling path, which is anchored in genetic or physical maps linked to whole genome sequence information, greatly facilitated karyotype analysis in *A. thaliana* (Lysak *et al.*, 2006a), *Brachypodium distachyon* (Hasterok *et al.*, 2006a) and *Glycine max* (Findley *et al.*, 2011). To compensate for the lack of such tiling path, a “divide and combine screening” strategy was developed and used to assign repeat-free BACs to each chromosome pair of *G. margaretae*. According to that, the 32 repeat-free BAC clones were divided into two subsets, each comprising 16 BACs. The BACs of each subset were further split into four pools containing four BACs each. The chromosomal localization of the 16 BACs of each subset was determined by combining results from sequential rounds of three-color FISH. For each round, the chromosomal position of the four BACs per pool was determined simultaneously by either single fluorescent labeling (Cy3 or Texas red) or combined fluorescent labeling (Cy3/Alexa 488 or Texas red/Alexa 488). Thereby, among the 16 repeat-free BACs of subset 1, four single BACs and five groups of linked BACs were identified as chromosome-specific (Figure 28). In subset 2, 13 single BACs and one group of linked BACs were identified, that allow discriminating 14 chromosome pairs of *G. margaretae* (Figure 29).

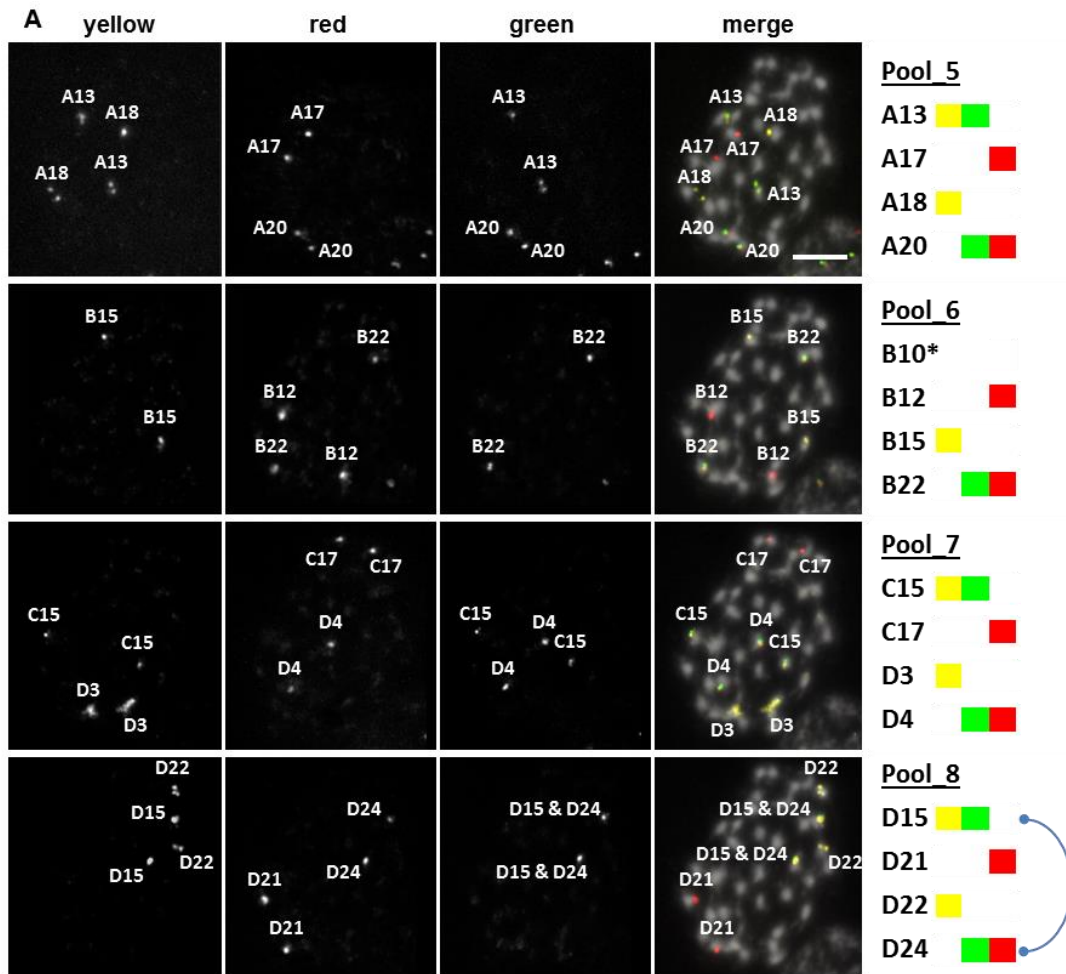


**D**

- ❖ Single BACs: A3, B6, D5 and D8
- ❖ Linked BACs: LK1.1 (A4-A22-C10), LK1.2 (A7-B5), LK1.3 (A8-C2-D7), LK1.4 (B3-C12) and LK1.5 (C11-D9)

Figure 28: Chromosomal localization of *G. margaretae* repeat-free BACs as revealed by sequential three-color FISH (BAC pools 1-4, subset 1).

(A) BAC pools 1, 2 and 4 subsequently hybridized to the same metaphase plate. Note that FISH signals of BAC clone B3 of pool 2 are missing in this screening but identifiable in other sequential FISH analysis in (C). Due to the degradation of cytological preparations after probe-stripping, the chromosomal localization of pool 3 in comparison with pool 1 (B) or with pools 2 and 4 (C) were separately investigated. (D) Nine chromosome pairs of *G. margaretae* can be distinguished by either four single or five groups of linked BACs when the results of three separate sequential FISH analyses are combined (the scheme in the right panel). Bars represent 5 μ m.



B

- ❖ Single BACs: A13, A17, A18, A20, B12, B15, B22, C15, C17, D3, D4, D21 and D22.
- ❖ Linked BACs: LK2.1 (D15-D24).

Figure 29: Chromosomal localization of *G. margaretae* repeat-free BACs as revealed by sequential three-color FISH (BAC pools 5-8, subset 2).

(A) The well conserved chromosome morphology even after four subsequent FISH experiments enabled the chromosomal assignment of 16 BAC clones of subset 2. Due to the loss of FISH signals, BAC clone B10 of pool_6 was excluded from further analysis. (B) The unambiguous FISH signals of each probe pool allowed 14 chromosome pairs of *G. margaretae* to be individualized (according to the scheme in the right panel). Bar represents 5 μ m.

To distinguish chromosomes of *G. margaretae*, 17 single BACs and representatives for six groups of linked BACs were sequentially hybridized in two subsets together with 45S rDNA on the same slide. The order of sequential FISH and chromosomal localizations of all probes are shown in figure 30A. Finally, 15 of the 19 chromosome pairs of *G. margaretae* were identified by either six single BACs or nine groups of linked BACs (Figure 30B).

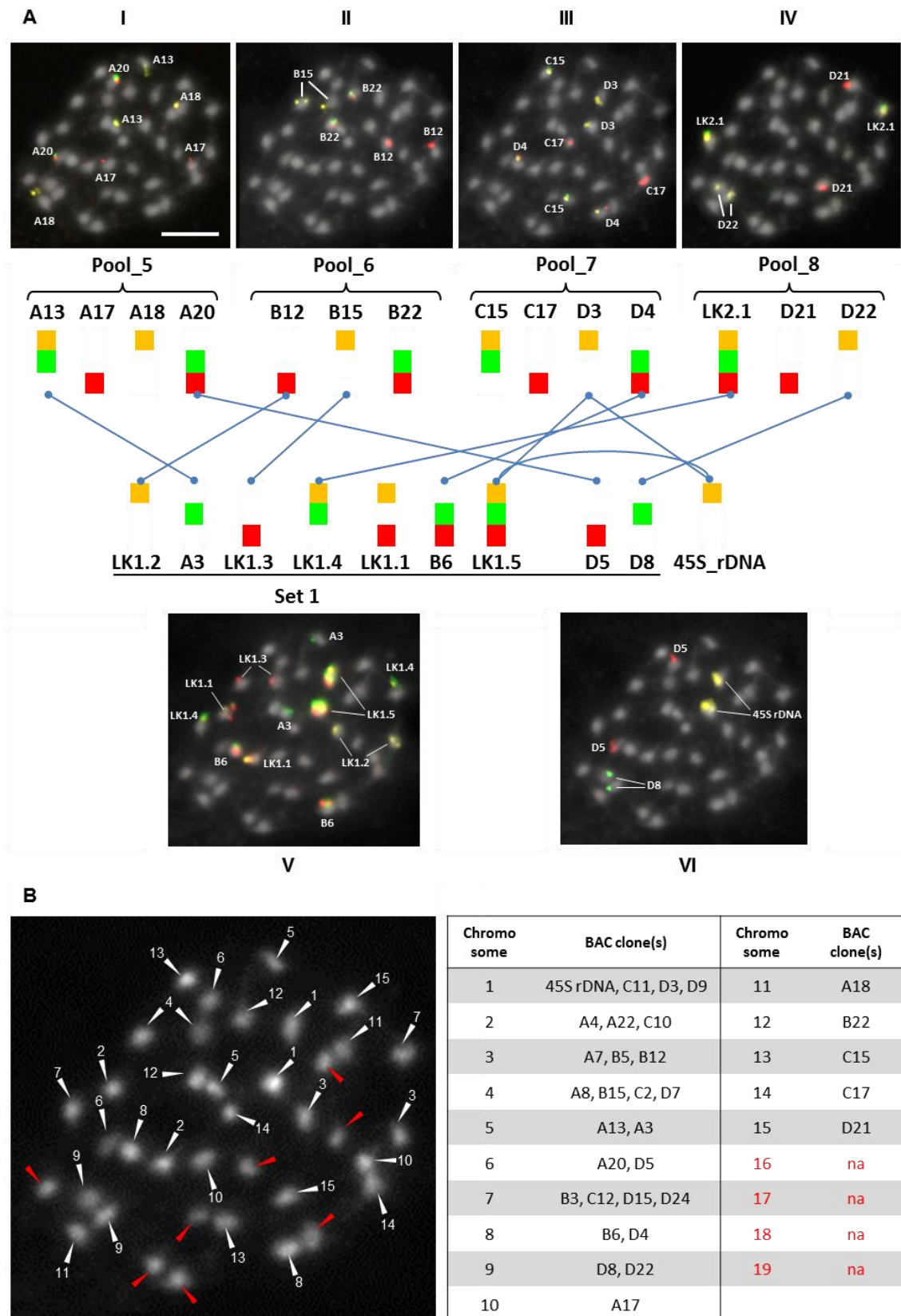


Figure 30: Sequential FISH using repeat-free BAC probes and 45S rDNA identify 15 of the 19 chromosome pairs of *G. margaretae*.

(A) The Roman numerals (I-VI) denote the order of FISH experiments using respective probe pools. The colocalizations of FISH signals after six hybridizations are indicated by the blue lines connecting corresponding

probes. **(B)** In the same metaphase plate, the white arrow heads with numbers indicate the 15 chromosome pairs that can be distinguished by different chromosome-specific probes (listed in the table). The red arrow heads mark the remaining four chromosome pairs (16-19) without identifications (na: not available). Bars represent 5 μm .

A large-insert BAC library provides a valuable, high-coverage representative for genome-scale research such as high-quality whole genome sequence alignment or physical map construction (Marra *et al.*, 1997; Zhang & Wu, 2001). In the field of cytogenetics, BAC tiling paths have proved themselves as an essential key tool for sophisticated karyotype evolution analyses *via* chromosome painting such as within the family Brassicaceae (Lysak *et al.*, 2001; Lysak *et al.*, 2006a; Mandakova & Lysak, 2008) or within the genus *Brachypodium* (Betekhtin *et al.*, 2014). Additionally, BAC clones are potential sources for chromosome identifications (Wang *et al.*, 2008; Xiong & Pires, 2011). For *G. margaretae*, un-annotated repeat-free BAC clones were unambiguously assigned to 15 of the 19 pairs of small and morphologically similar chromosomes. Although four chromosome pairs remain unidentified, the “divide and combine” strategy coupled with sequential FISH offers a reliable approach. Using more repeat-free BAC clones, the simultaneous identification of all 19 chromosome pairs of *G. margaretae* seems feasible. Furthermore, groups of linked BACs locating on the same chromosomes of *G. margaretae* (1-9) may offer a useful FISH-based tool to detect chromosome rearrangements among species of the subgenus *Genlisea* which are characterized by plasticity in genome size and chromosome number.

3.3.4. Conclusions regarding karyotyping in *Genlisea*

Chromosome banding and FISH technique are two main approaches in karyotyping analysis. While the former is only capable for analyzing species possessing large chromosomes with clear morphology, the latter shows the outstanding advantages in investigation of species with small chromosomes. The availability of many genome assemblies makes single-copy DNA sequences a plentiful source of highly reliable cytogenetic markers for FISH-based karyotyping (Lamb *et al.*, 2007; Ma *et al.*, 2010; Lou *et al.*, 2014). Another source for chromosome-specific markers facilitating karyotyping are tandem repeats. However, the number of identified tandem repeats that show chromosome-specific distribution is often limited although repetitive sequences are present in all genomes. Thus, tandem repeats are usually combined with other markers, particularly BAC-derived probes, for chromosome identification.

Utilizing the available sources for cytogenetic marks, 11 of the 20 chromosome pairs of *G. nigrocaulis*, 13 of the 20 chromosome pairs of *G. hispidula* and 15 of the 19 chromosome pairs of *G. margaretae* were clearly identified by single-copy sequences, tandem repeats and repeat-free BAC

clones, respectively, in combination with ribosomal DNA probes. These probes proved to be helpful for the challenging task to differentiate *Genlisea* chromosomes in particular those of *G. nigrocaulis* and *G. margaretae* which are numerous and tiny. Furthermore, the presented data provide sets of chromosome-specific markers which may serve as an anchor for further cytological analyses, such as interspecific chromosome painting using a complete BAC tiling path, with the aim to elucidate karyotype evolution during speciation within the genus *Genlisea*.

4. SUMMARY

Among flowering plants, the carnivorous genus *Genlisea* possesses one of the largest fold-range genome size differences and comprises species with the smallest genomes. The genome size plasticity makes *Genlisea* an interesting subject to study mechanisms of genome and karyotype evolution. In this study, the cytological features of species representative for the three sections of the subgenus *Genlisea* were investigated in order to establish the so far lacking cytogenetic background of the genus.

The diversity of cytogenetic features reflects the genome size plasticity within subgenus Genlisea

Three sections within the species-rich subgenus *Genlisea* are characterized by distinct genome size categories: the ultra-small and small genomes of the phylogenetically most derived section *Genlisea* (*G. nigrocaulis*, *G. pygmaea* and *G. aurea*) and of the section *Recurvatae* (*G. margaretae*), respectively; and the up to 18-fold larger genomes of the section *Africanae* (*G. hispidula* and *G. subglabra*). The nuclear phenotype regarding chromatin organization and modifications mirrors previous observations for small genomes in the sections *Genlisea* and for larger genomes in the section *Africanae*. In the section *Recurvatae* all tested epigenetic methylation marks showed for *G. margaretae* a nuclear distribution typical for larger genomes. Despite of the large variation in genome size, the same chromosome numbers ($2n = 40$) were counted for the ultra-small genome of *G. nigrocaulis* (section *Genlisea*) and for two species of the section *Africanae*. One of them, *G. hispidula*, is apparently mesoallotetraploid and its ancestral species underwent dysploid chromosome number reduction. Furthermore, chromosome number counting in other species of the section *Genlisea* (*G. pygmaea* has $2n = 80$ and *G. aurea* has $2n \approx 104$ chromosomes) indicates neopolyploidy, while *G. margaretae* (section *Recurvatae*, $2n = 38$ chromosomes) displays a similar but not identical basic chromosome number as observed for *G. nigrocaulis* and *G. pygmaea*. This implies a complex karyotype evolution within the subgenus *Genlisea*. Conversely, the number and chromosomal distribution of ribosomal DNA clusters are similar within, but differ between groups with small and large genome sizes, respectively. Whereas FISH revealed one locus of each 45S and 5S rDNA per diploid complement in species of the sections *Genlisea* and *Recurvatae*, probably related to genome shrinkage in these clades; more loci harboring different copy numbers occur in species with larger genomes of the section *Africanae*.

The rapid genome evolution within subgenus Genlisea is accompanied by alteration of centromere and telomere sequences

The ~18-fold genome sizes difference between *G. nigrocaulis* and *G. hispidula* is apparently linked with the remarkable divergence in amount, composition and chromosomal distribution of different classes of repetitive sequences which were identified based on whole genome sequencing data. Sequence alterations were also found for centromeres and telomeres of these two species and their close relatives. Whereas in *G. nigrocaulis* and its close relative *G. pygmaea* of the section *Genlisea* a tandem repeat occupies the centromeric position, in *G. hispidula* and *G. subglabra* of the section *Africanae* four retroelements were found in the centromeres instead. Thus, within the same genus, either tandem repeats or retroelements can serve as centromeric sequences. Surprisingly, both groups differ also in the sequence of their telomeric repeats. The canonical plant telomeric repeat TTTAGGG, present at the chromosome termini of the small genomes, is replaced by two intermingled sequence variants (TTCAGG and TTTCAGG) in species with the large genomes. Hence, even intrgeneric switch of telomeric repeats may occur. The presence of cytosine within the guanine-rich strand of the two newly described variants is so far unique for plants and has potential implications for the modification of telomeric chromatin.

Chromosome identification for representatives of the three sections of the subgenus Genlisea, a basis for further karyotype evolution analysis.

Identification of chromosomes of a particular species is the basic prerequisite not only for detailed cytogenetic analyses but also for genomic and genetic studies. By means of single-copy sequences, tandem repeats and repeat-free BAC clones, initial karyotyping was performed in *G. nigrocaulis*, *G. hispidula* and *G. margaretae*, respectively. Furthermore, the single-copy probe set of *G. nigrocaulis* confirmed the tetraploidy that was suspected by genome size and chromosome number of *G. pygmaea*. Additionally, karyotyping suggested a dysploid chromosome number reduction within the ancestors of the allotetraploid *G. hispidula* and *G. subglabra* of the section *Africanae*. Further efforts to develop chromosome-specific markers are required for the complete karyotyping of three *Genlisea* species as well as for further elucidation of karyotype evolution within the genus *Genlisea*.

References

- Adams SP, Hartman TPV, Lim KY, Chase MW, Bennett MD, Leitch IJ, Leitch AR. 2001. Loss and recovery of *Arabidopsis*-type telomere repeat sequence 5'-(TTAGGG)_n-3' in the evolution of a major radiation of flowering plants. *Proc. R. Soc. Lond. B* **268**: 1541-1546.
- Adams SP, Leitch IJ, Bennett MD, Leitch AR. 2000. Aloe L.--a second plant family without (TTAGGG)_n telomeres. *Chromosoma* **109**(3): 201-205.
- Albert VA, Jobson RW, Michael TP, Taylor DJ. 2010. The carnivorous bladderwort (*Utricularia*, Lentibulariaceae): a system inflates. *Journal of Experimental Botany* **61**(1): 5-9.
- Ali HB, Lysak MA, Schubert I. 2005. Chromosomal localization of rDNA in the Brassicaceae. *Genome* **48**(2): 341-346.
- Ausio J. 2015. The shades of gray of the chromatin fiber Recent literature provides new insights into the structure of chromatin. *Bioessays* **37**(1): 46-51.
- Bannister AJ, Kouzarides T. 2011. Regulation of chromatin by histone modifications. *Cell Research* **21**(3): 381-395.
- Barthlott W, Porembski S, Fischer E, Gemmel B. 1998. First protozoa-trapping plant found. *Nature* **392**(6675): 447-447.
- Baucom RS, Estill JC, Chaparro C, Upshaw N, Jogi A, Deragon JM, Westerman RP, Sanmiguel PJ, Bennetzen JL. 2009. Exceptional diversity, non-random distribution, and rapid evolution of retroelements in the B73 maize genome. *PLoS Genetics* **5**(11): e1000732.
- Begum R, Zakrzewski F, Menzel G, Weber B, Alam SS, 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): 123-134.
- Bennett MD, Leitch IJ. 2005. Nuclear DNA amounts in angiosperms: progress, problems and prospects. *Annals of Botany* **95**(1): 45-90.
- Bennett MD, Leitch IJ, Price HJ, Johnston JS. 2003. Comparisons with *Caenorhabditis* (approximately 100 Mb) and *Drosophila* (approximately 175 Mb) using flow cytometry show genome size in *Arabidopsis* to be approximately 157 Mb and thus approximately 25% larger than the *Arabidopsis* genome initiative estimate of approximately 125 Mb. *Annals of Botany* **91**(5): 547-557.
- Bennetzen JL, Kellogg EA. 1997. Do Plants Have a One-Way Ticket to Genomic Obesity? *Plant Cell* **9**(9): 1509-1514.
- Bennetzen JL, Ma J, Devos KM. 2005. Mechanisms of recent genome size variation in flowering plants. *Annals of Botany* **95**(1): 127-132.
- Bennetzen JL, Wang H. 2014. The Contributions of Transposable Elements to the Structure, Function, and Evolution of Plant Genomes. *Annual Review of Plant Biology, Vol 65* **65**: 505-530.
- Berr A, Pecinka A, Meister A, Kreth G, Fuchs J, Blattner FR, Lysak MA, Schubert I. 2006. Chromosome arrangement and nuclear architecture but not centromeric sequences are conserved between *Arabidopsis thaliana* and *Arabidopsis lyrata*. *Plant Journal* **48**(5): 771-783.
- Betekhtin A, Jenkins G, Hasterok R. 2014. Reconstructing the Evolution of *Brachypodium* Genomes Using Comparative Chromosome Painting. *PLoS ONE* **9**(12).
- Braszewska-Zalewska A, Bernas T, Maluszynska J. 2010. Epigenetic chromatin modifications in Brassica genomes. *Genome* **53**(3): 203-210.
- Braszewska-Zalewska A, Dziurlikowska A, Maluszynska J. 2012. Histone H3 methylation patterns in *Brassica nigra*, *Brassica juncea*, and *Brassica carinata* species. *Genome* **55**(1): 68-74.
- Buggs RJA, Doust AN, Tate JA, Koh J, Soltis K, Feltus FA, Paterson AH, Soltis PS, Soltis DE. 2009. Gene loss and silencing in *Tragopogon miscellus* (Asteraceae): comparison of natural and synthetic allotetraploids. *Heredity* **103**(1): 73-81.
- Butelli E, Licciardello C, Zhang Y, Liu J, Mackay S, Bailey P, Reforgiato-Recupero G, Martin C. 2012. Retrotransposons control fruit-specific, cold-dependent accumulation of anthocyanins in blood oranges. *Plant Cell* **24**(3): 1242-1255.
- Cao HX, Schmutzer T, Scholz U, Pecinka A, Schubert I, Vu GT. 2015. Metatranscriptome analysis reveals host-microbiom interactions in traps of carnivorous *Genlisea* species. *Frontier Microbiology*. doi: 10.3389/fmicb.2015.00526

- Cellamare A, Catacchio CR, Alkan C, Giannuzzi G, Antonacci F, Cardone MF, Della Valle G, Malig M, Rocchi M, Eichler EE, *et al.* 2009. New insights into centromere organization and evolution from the white-cheeked gibbon and marmoset. *Molecular Biology and Evolution* **26**(8): 1889-1900.
- Chen J, Huang Q, Gao D, Wang J, Lang Y, Liu T, Li B, Bai Z, Luis Goicoechea J, Liang C, *et al.* 2013. Whole-genome sequencing of *Oryza brachyantha* reveals mechanisms underlying *Oryza* genome evolution. *Nat Commun* **4**: 1595.
- Cheng Z, Dong F, Langdon T, Ouyang S, Buell CR, Gu M, Blattner FR, Jiang J. 2002. Functional Rice Centromeres Are Marked by a Satellite Repeat and a Centromere-Specific Retrotransposon. *The Plant Cell Online* **14**(8): 1691-1704.
- Chiarini FE. 2014. Variation in rDNA loci of polyploid *Solanum elaeagnifolium* (Solanaceae). *New Zealand Journal of Botany* **52**(3): 277-284.
- Cuadrado A, Carmona A, Jouve N. 2013. Chromosomal characterization of the three subgenomes in the polyploids of *Hordeum murinum* L.: new insight into the evolution of this complex. *PLoS ONE* **8**(12): e81385.
- Danilova TV, Birchler JA. 2008. Integrated cytogenetic map of mitotic metaphase chromosome 9 of maize: resolution, sensitivity, and banding paint development. *Chromosoma* **117**(4): 345-356.
- Darnowski DW, Fritz S. 2010. Prey preference in *Genlisea* small crustaceans, not protozoa. *Carniv. Pl. Newslett.* **39**(4): 114-116.
- Datson PM, Murray BG. 2006. Ribosomal DNA locus evolution in *Nemesia*: transposition rather than structural rearrangement as the key mechanism? *Chromosome Research* **14**(8): 845-857.
- De Azkue D, Martinez A. 1988. DNA Content and Chromosome Evolution in the Shrubby *Oxalis*. *Genome* **30**(1): 52-57.
- Dechyeva D, Schmidt T. 2006. Molecular organization of terminal repetitive DNA in Beta species. *Chromosome Res* **14**(8): 881-897.
- Demidov D, Schubert V, Kumke K, Weiss O, Karimi-Ashtiyani R, Buttlar J, Heckmann S, Wanner G, Dong Q, Han F, *et al.* 2014. Anti-Phosphorylated Histone H2A^{Thr120}: A Universal Microscopic Marker for Centromeric Chromatin of Mono- and Holocentric Plant Species. *Cytogenetic and Genome Research* **143**(1-3): 150-156.
- Dolezel J, Bartos J, Voglmayr H, Greilhuber J. 2003. Nuclear DNA content and genome size of trout and human. *Cytometry A* **51**(2): 127-128; author reply 129.
- Dolezel J, Greilhuber J, Suda J. 2007. Estimation of nuclear DNA content in plants using flow cytometry. *Nat. Protocols* **2**(9): 2233-2244.
- Dong F, Jiang J. 1998. Non-Rabl patterns of centromere and telomere distribution in the interphase nuclei of plant cells. *Chromosome Res* **6**(7): 551-558.
- Dong F, Miller JT, Jackson SA, Wang G-L, Ronald PC, Jiang J. 1998. Rice (*Oryza sativa*) centromeric regions consist of complex DNA. *Proceedings of the National Academy of Sciences* **95**(14): 8135-8140.
- Dong Q, Han F. 2012. Phosphorylation of histone H2A is associated with centromere function and maintenance in meiosis. *The Plant Journal* **71**(5): 800-809.
- Eckardt NA. 2004. Two genomes are better than one: widespread paleopolyploidy in plants and evolutionary effects. *Plant Cell* **16**(7): 1647-1649.
- El Baidouri M, Carpentier MC, Cooke R, Gao D, Lasserre E, Llauro C, Mirouze M, Picault N, Jackson SA, Panaud O. 2014. Widespread and frequent horizontal transfers of transposable elements in plants. *Genome Research* **24**(5): 831-838.
- Emshwiller E. 2002. Ploidy levels among species in the '*Oxalis tuberosa* alliance' as inferred by flow cytometry. *Annals of Botany* **89**(6): 741-753.
- Fajkus J, Fulneckova J, Hulanova M, Berkova K, Riha K, Matyasek R. 1998. Plant cells express telomerase activity upon transfer to callus culture, without extensively changing telomere lengths. *Molecular and General Genetics* **260**(5): 470-474.
- Fajkus J, Sýkorová E, Leitch AR. 2005. Telomeres in evolution and evolution of telomeres. *Chromosome Research* **13**: 469-479.
- Farrar K, Donnison I. 2007. Construction and screening of BAC libraries made from *Brachypodium* genomic DNA. *Nature Protocols* **2**(7): 1661-1674.
- Findley SD, Pappas AL, Cui Y, Birchler JA, Palmer RG, Stacey G. 2011. Fluorescence in situ hybridization-based karyotyping of soybean translocation lines. *G3 (Bethesda)* **1**(2): 117-129.
- Fischer E, Porembski S, Barthlott W. 2000. Revision of the genus *Genlisea* (Lentibulariaceae) in Africa and Madagascar with notes on ecology and phytogeography. *Nordic Journal of Botany* **20**(3): 291-318.

- Fishman L, Willis JH, Wu CA, Lee YW. 2014. Comparative linkage maps suggest that fission, not polyploidy, underlies near-doubling of chromosome number within monkeyflowers (*Mimulus*; Phrymaceae). *Heredity (Edinb)* **112**(5): 562-568.
- Fitzgerald MS, McKnight TD, Shippen DE. 1996. Characterization and developmental patterns of telomerase expression in plants. *Proceedings of the National Academy of Sciences of the United States of America* **93**(25): 14422-14427.
- Fleischmann A. 2012. *Monograph of the Genus Genlisea*: Redfern Natural History Productions.
- Fleischmann A, Michael TP, Rivadavia F, Sousa A, Wang W, Temsch EM, Greilhuber J, Muller KF, 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): 1651-1663.
- Fleischmann A, Schaeferhoff B, Heubl G, Rivadavia F, Barthlott W, Mueller K. 2010. Phylogenetics and character evolution in the carnivorous plant genus *Genlisea* A. St.-Hil. (Lentibulariaceae). *Molecular Phylogenetics and Evolution*(56): 768-783.
- Fojtova M, Fulneckova J, Fajkus J, Kovarik A. 2002. Recovery of tobacco cells from cadmium stress is accompanied by DNA repair and increased telomerase activity. *Journal of Experimental Botany* **53**(378): 2151-2158.
- Fojtova M, Wong JTY, Dvorackova M, Yan KTH, Sykorova E, Fajkus J. 2010. Telomere maintenance in liquid crystalline chromosomes of dinoflagellates. *Chromosoma* **119**(5): 485-493.
- Fransz P, De Jong JH, Lysak M, Castiglione MR, Schubert I. 2002. Interphase chromosomes in *Arabidopsis* are organized as well defined chromocenters from which euchromatin loops emanate. *Proc Natl Acad Sci U S A* **99**(22): 14584-14589.
- Fuchs J, Brandes A, Schubert I. 1995. Telomere sequence localization and karyotype evolution in higher plants. *Plant Systematics and Evolution* **196**: 227-241.
- Fuchs J, Demidov D, Houben A, Schubert I. 2006. Chromosomal histone modification patterns – from conservation to diversity. *Trends in Plant Science* **11**(4): 199-208.
- Fuchs J, Jovtchev G, Schubert I. 2008. The chromosomal distribution of histone methylation marks in gymnosperms differs from that of angiosperms. *Chromosome Research* **16**(6): 891-898.
- Fuchs J, Schubert I. 2012. Chromosomal Distribution and Functional Interpretation of Epigenetic Histone Marks in Plants. In: Bass HW, Birchler JA eds. *Plant Cytogenetics*: Springer New York, 231-253.
- Fukagawa T, Earnshaw William C. 2014. The Centromere: Chromatin Foundation for the Kinetochore Machinery. *Developmental Cell* **30**(5): 496-508.
- Fulnečková J, Ševčíková T, Fajkus J, Lukešová A, Lukeš M, Vlček Č, Lang BF, Kim E, Eliáš M, Sýkorová E. 2013. A Broad Phylogenetic Survey Unveils the Diversity and Evolution of Telomeres in Eukaryotes. *Genome Biology and Evolution* **5**(3): 468-483.
- Garcia S, Galvez F, Gras A, Kovarik A, Garnatje T. 2014a. Plant rDNA database: update and new features. *Database (Oxford)* **2014**.
- Garcia S, Leitch IJ, Anadon-Rosell A, Canela MA, Galvez F, Garnatje T, Gras A, Hidalgo O, Johnston E, Mas de Xaxars G, et al. 2014b. Recent updates and developments to plant genome size databases. *Nucleic Acids Research* **42**(Database issue): D1159-1166.
- Gong ZY, Wu YF, Koblizkova A, Torres GA, Wang K, Iovene M, Neumann P, Zhang WL, Novak P, Buell CR, et al. 2012. Repeatless and Repeat-Based Centromeres in Potato: Implications for Centromere Evolution. *Plant Cell* **24**(9): 3559-3574.
- Gorinsek B, Gubensek F, Kordis D. 2004. Evolutionary genomics of chromoviruses in eukaryotes. *Molecular Biology and Evolution* **21**(5): 781-798.
- Gregory TR. 2001. Coincidence, coevolution, or causation? DNA content, cell size, and the C-value enigma. *Biological Reviews of the Cambridge Philosophical Society* **76**(1): 65-101.
- Gregory TR. 2004. Insertion-deletion biases and the evolution of genome size. *Gene* **324**: 15-34.
- Greilhuber J, Borsch T, Worberg A, Porembski S, Barthlott W. 2006. Smallest angiosperm genomes found in Lentibulariaceae, with chromosome of bacterial size. *Plant Biology* **8**: 770-777.
- Greilhuber J, Dolezel J, Lysak MA, Bennett MD. 2005. The origin, evolution and proposed stabilization of the terms 'genome size' and 'C-value' to describe nuclear DNA contents. *Annals of Botany* **95**(1): 255-260.
- Guo X, Han F. 2014. Asymmetric epigenetic modification and elimination of rDNA sequences by polyploidization in wheat. *Plant Cell* **26**(11): 4311-4327.
- Hall TA. 1999. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series*(41): 95-98.

- Hasterok R, Marasek A, Donnison IS, Armstead I, Thomas A, King IP, Wolny E, Idziak D, Draper J, Jenkins G. 2006a. Alignment of the genomes of *Brachypodium distachyon* and temperate cereals and grasses using bacterial artificial chromosome landing with fluorescence in situ hybridization. *Genetics* **173**(1): 349-362.
- Hasterok R, Wolny E, Hosiawa M, Kowalczyk M, Kulak-Ksiazczyk S, Ksiazczyk T, Heneen WK, Maluszynska J. 2006b. Comparative analysis of rDNA distribution in chromosomes of various species of Brassicaceae. *Annals of Botany* **97**(2): 205-216.
- Hawkins JS, Proulx SR, Rapp RA, Wendel JF. 2009. Rapid DNA loss as a counterbalance to genome expansion through retrotransposon proliferation in plants. *Proceedings of the National Academy of Sciences of the United States of America* **106**(42): 17811-17816.
- He XJ, Chen T, Zhu JK. 2011. Regulation and function of DNA methylation in plants and animals. *Cell Research* **21**(3): 442-465.
- Hemleben V, Kovarik A, Torres-Ruiz RA, Volkov RA, Beridze T. 2007. Plant highly repeated satellite DNA: molecular evolution, distribution and use for identification of hybrids. *Systematics and biodiversity* **5**(3): 277-289.
- Henikoff S, Ahmad K, Malik HS. 2001. The centromere paradox: stable inheritance with rapidly evolving DNA. *Science* **293**(5532): 1098-1102.
- Heslop-Harrison JS, Schwarzacher T. 2013. Nucleosomes and centromeric DNA packaging. *Proceedings of the National Academy of Sciences* **110**(50): 19974-19975.
- Hosouchi T, Kumekawa N, Tsuruoka H, Kotani H. 2002. Physical map-based sizes of the centromeric regions of *Arabidopsis thaliana* chromosomes 1, 2, and 3. *DNA Research* **9**(4): 117-121.
- Houben A, Demidov D, Gernand D, Meister A, Leach CR, Schubert I. 2003. Methylation of histone H3 in euchromatin of plant chromosomes depends on basic nuclear DNA content. *Plant Journal* **33**(6): 967-973.
- Houben A, Schubert I. 2003. DNA and proteins of plant centromeres. *Current Opinion in Plant Biology* **6**(6): 554-560.
- Houben A, Wako T, Furushima-Shimogawara R, Presting G, Kunzel G, Schubert I, Fukui K. 1999. The cell cycle dependent phosphorylation of histone H3 is correlated with the condensation of plant mitotic chromosomes. *Plant Journal* **18**(6): 675-679.
- Hudakova S, Michalek W, Presting GG, ten Hoopen R, dos Santos K, Jasencakova Z, Schubert I. 2001. Sequence organization of barley centromeres. *Nucleic Acids Research* **29**(24): 5029-5035.
- Ibarra-Laclette E, Lyons E, Hernandez-Guzman G, Perez-Torres CA, Carretero-Paulet L, Chang TH, Lan T, Welch AJ, Juarez MJ, Simpson J, et al. 2013. Architecture and evolution of a minute plant genome. *Nature* **498**(7452): 94-98.
- Ijdo JW, Wells RA, Baldini A, Reeders ST. 1991. Improved telomere detection using a telomere repeat probe (TTAGGG)_n generated by PCR. *Nucleic Acids Research* **19**(17): 4780.
- International *Brachypodium* I. 2010. Genome sequencing and analysis of the model grass *Brachypodium distachyon*. *Nature* **463**(7282): 763-768.
- Jackson RC. 1959. A Study of Meiosis in *Haplopappus-Gracilis* (Compositae). *American Journal of Botany* **46**(7): 550-554.
- Jiang J, Birchler JA, Parrott WA, Kelly Dawe R. 2003. A molecular view of plant centromeres. *Trends in Plant Science* **8**(12): 570-575.
- Jiao Y, Wickett NJ, Ayyampalayam S, Chanderbali AS, Landherr L, Ralph PE, Tomsho LP, Hu Y, Liang H, Soltis PS, et al. 2011. Ancestral polyploidy in seed plants and angiosperms. *Nature* **473**(7345): 97-100.
- Jobson RW, Albert VA. 2002. Molecular Rates Parallel Diversification Contrasts between Carnivorous Plant Sister Lineages1. *Cladistics* **18**(2): 127-136.
- Jobson RW, Playford J, Cameron KM, Albert VA. 2003. Molecular phylogenetics of Lentibulariaceae inferred from plastid rps16 intron and trnL-F DNA sequences: Implications for character evolution and biogeography. *Systematic Botany* **28**(1): 157-171.
- Joti Y, Hikima T, Nishino Y, Kamada F, Hihara S, Takata H, Ishikawa T, Maeshima K. 2012. Chromosomes without a 30-nm chromatin fiber. *Nucleus-Austin* **3**(5): 404-410.
- Kashkush K, Feldman M, Levy AA. 2003. Transcriptional activation of retrotransposons alters the expression of adjacent genes in wheat. *Nature Genetics* **33**(1): 102-106.
- Kawahara Y, de la Bastide M, Hamilton JP, Kanamori H, McCombie WR, Ouyang S, Schwartz DC, Tanaka T, Wu JZ, Zhou SG, et al. 2013. Improvement of the *Oryza sativa* Nipponbare reference genome using next generation sequence and optical map data. *Rice* **6**.
- Kazazian HH, Jr. 2004. Mobile elements: drivers of genome evolution. *Science* **303**(5664): 1626-1632.

- Khandelwal S. 1990.** Chromosome Evolution in the Genus *Ophioglossum* L. *Botanical Journal of the Linnean Society* **102**(3): 205-217.
- Kirik A, Salomon S, Puchta H. 2000.** Species-specific double-strand break repair and genome evolution in plants. *EMBO Journal* **19**(20): 5562-5566.
- Koga A, Hirai Y, Terada S, Jahan I, Baicharoen S, Arsaithamkul V, Hirai H. 2014.** Evolutionary Origin of Higher-Order Repeat Structure in Alpha-Satellite DNA of Primate Centromeres. *DNA Research* **21**(4): 407-415.
- Kolano B, Gardunia BW, Michalska M, Bonifacio A, Fairbanks D, Maughan PJ, Coleman CE, Stevens MR, Jellen EN, Maluszynska J. 2011.** Chromosomal localization of two novel repetitive sequences isolated from the *Chenopodium quinoa* Willd. genome. *Genome* **54**(9): 710-717.
- Kouzarides T. 2007.** Chromatin modifications and their function. *Cell* **128**(4): 693-705.
- Kuo CH, Ochman H. 2009.** Deletional Bias across the Three Domains of Life. *Genome Biology and Evolution* **1**: 145-152.
- Lamb JC, Danilova T, Bauer MJ, Meyer JM, Holland JJ, Jensen MD, Birchler JA. 2007.** Single-gene detection and karyotyping using small-target fluorescence in situ hybridization on maize somatic chromosomes. *Genetics* **175**(3): 1047-1058.
- Leitch IJ, Soltis DE, Soltis PS, Bennett MD. 2005.** Evolution of DNA amounts across land plants (embryophyta). *Annals of Botany* **95**(1): 207-217.
- Leushkin E, Sutormin R, Nabieva E, Penin A, Kondrashov A, Logacheva M. 2013.** The miniature genome of a carnivorous plant *Genlisea aurea* contains a low number of genes and short non-coding sequences. *BMC Genomics* **14**(1): 476.
- Levan A, Fredga K, Sandberg AA. 1964.** Nomenclature for Centromeric Position on Chromosomes. *Hereditas-Genetiskt Arkiv* **52**(2): 201-&.
- Levin DA. 2002.** *The role of chromosomal change in plant evolution*: Oxford University Press.
- Li G, Reinberg D. 2011.** Chromatin higher-order structures and gene regulation. *Current Opinion in Genetics & Development* **21**(2): 175-186.
- Li L, Yang J, Tong Q, Zhao L, Song Y. 2005.** A novel approach to prepare extended DNA fibers in plants. *Cytometry Part A* **63A**(2): 114-117.
- Lloyd FE. 1942.** *The carnivorous plants*: Chronica Botanica Company
- Lopez-Flores I, Garrido-Ramos MA. 2012.** The repetitive DNA content of eukaryotic genomes. *Genome Dyn* **7**: 1-28.
- Lou Q, Zhang Y, He Y, Li J, Jia L, Cheng C, Guan W, Yang S, Chen J. 2014.** Single-copy gene-based chromosome painting in cucumber and its application for chromosome rearrangement analysis in Cucumis. *Plant Journal* **78**(1): 169-179.
- Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I. 2006a.** Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc Natl Acad Sci U S A* **103**(13): 5224-5229.
- Lysak MA, Fransz P, Schubert I. 2006b.** Cytogenetic analyses of *Arabidopsis*. *Methods Mol Biol* **323**: 173-186.
- Lysak MA, Fransz PF, Ali HBM, Schubert I. 2001.** Chromosome painting in *Arabidopsis thaliana*. *Plant Journal* **28**(6): 689-697.
- Lysak MA, Koch MA, Pecinka A, Schubert I. 2005.** Chromosome triplication found across the tribe Brassiceae. *Genome Research* **15**(4): 516-525.
- Lysak MA, Pecinka A, Schubert I. 2003.** Recent progress in chromosome painting of *Arabidopsis* and related species. *Chromosome Research* **11**(3): 195-204.
- Ma J, Wing RA, Bennetzen JL, Jackson SA. 2007.** Plant centromere organization: a dynamic structure with conserved functions. *Trends in Genetics* **23**(3): 134-139.
- Ma L, Vu GT, Schubert V, Watanabe K, Stein N, Houben A, Schubert I. 2010.** Synteny between *Brachypodium distachyon* and *Hordeum vulgare* as revealed by FISH. *Chromosome Res* **18**(7): 841-850.
- Mandakova T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA. 2010.** Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. *Plant Cell* **22**(7): 2277-2290.
- Mandakova T, Lysak MA. 2008.** Chromosomal phylogeny and karyotype evolution in x=7 crucifer species (Brassicaceae). *Plant Cell* **20**(10): 2559-2570.
- Marra MA, Kucaba TA, Dietrich NL, Green ED, Brownstein B, Wilson RK, McDonald KM, Hillier LW, McPherson JD, Waterston RH. 1997.** High throughput fingerprint analysis of large-insert clones. *Genome Research* **7**(11): 1072-1084.
- Martinez-Zapater J, Estelle M, Somerville C. 1986.** A highly repeated DNA sequence in *Arabidopsis thaliana*. *Molecular and General Genetics MGG* **204**(3): 417-423.

- Mayer KFX, Waugh R, Langridge P, Close TJ, Wise RP, Graner A, Matsumoto T, Sato K, Schulman A, Muehlbauer GJ, *et al.* 2012. A physical, genetic and functional sequence assembly of the barley genome. *Nature* **491**(7426): 711-+.
- Mehrotra S, Goyal V. 2014. Repetitive sequences in plant nuclear DNA: types, distribution, evolution and function. *Genomics Proteomics Bioinformatics* **12**(4): 164-171.
- Melters D, Bradnam K, Young H, Telis N, May M, Ruby J, Sebra R, Peluso P, Eid J, Rank D, *et al.* 2013. Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. *Genome Biology* **14**(1): R10.
- Meyne J, Ratliff RL, Moyzis RK. 1989. Conservation of the Human Telomere Sequence (Ttaggg)N among Vertebrates. *Proceedings of the National Academy of Sciences of the United States of America* **86**(18): 7049-7053.
- Michael TP. 2014. Plant genome size variation: bloating and purging DNA. *Briefings in Functional Genomics* **13**(4): 308-317.
- Michael TP, Jackson S. 2013. The First 50 Plant Genomes. *Plant Genome* **6**(2).
- Mohan KN, Rani BS, Kulashreshta PS, Kadandale JS. 2011. Characterization of TTAGG telomeric repeats, their interstitial occurrence and constitutively active telomerase in the mealybug *Planococcus lilacinus* (Homoptera; Coccoidea). *Chromosoma* **120**(2): 165-175.
- Mravinac B, Mestrovic N, Cavrak VV, Plohl M. 2011. TCAGG, an alternative telomeric sequence in insects. *Chromosoma* **120**(4): 367-376.
- Mueller K, Borsch T, Legendre L, Porembski S, Theisen I, Barthlott W. 2003. Evolution of carnivory in Lentibulariaceae and the Lamiales. *Plant Biology* **6**: 477-490.
- Mueller KF, Borsch T, Legendre L, Porembski S, Barthlott W. 2006. Recent progress in understanding the evolution of carnivorous Lentibulariaceae (Lamiales). *Plant Biology* **8**: 748-757.
- Nelson AD, Lamb JC, Kobrossly PS, Shippen DE. 2011. Parameters affecting telomere-mediated chromosomal truncation in *Arabidopsis*. *Plant Cell* **23**(6): 2263-2272.
- Nemeth AV, Dudits D, Molnar-Lang M, Linc G. 2013. Molecular cytogenetic characterisation of *Salix viminalis* L. using repetitive DNA sequences. *Journal of Applied Genetics* **54**(3): 265-269.
- Neumann P, Navratilova A, Koblizkova A, Kejnovsky E, Hribova E, Hobza R, Widmer A, Dolezel J, Macas J. 2011. Plant centromeric retrotransposons: a structural and cytogenetic perspective. *Mobile DNA* **2**(4).
- Neumann P, Navratilova A, Schroeder-Reiter E, Koblizkova A, Steinbauerova V, Chocholova E, Novak P, Wanner G, Macas J. 2012. Stretching the rules: monocentric chromosomes with multiple centromere domains. *PLoS Genetics* **8**(6): e1002777.
- Novak P, Neumann P, Macas J. 2010. Graph-based clustering and characterization of repetitive sequences in next-generation sequencing data. *BMC Bioinformatics* **11**: 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**(6): 792-793.
- Paterson AH, Wendel JF, Gundlach H, Guo H, Jenkins J, Jin D, Llewellyn D, Showmaker KC, Shu S, Udall J, *et al.* 2012. Repeated polyploidization of *Gossypium* genomes and the evolution of spinnable cotton fibres. *Nature* **492**(7429): 423-427.
- Pellicer J, Fay MF, Leitch IJ. 2010. The largest eukaryotic genome of them all? *Botanical Journal of the Linnean Society* **164**(1): 10-15.
- Peska V, Fajkus P, Fojtova M, Dvorackova M, Hapala J, Dvoracek V, Polanska P, Leitch AR, Sykorova E, Fajkus J. 2015. Characterisation of an unusual telomere motif (TTTTTTAGGG)_n in the plant *Cestrum elegans* (Solanaceae), a species with a large genome. *Plant Journal* **82**: 644-654.
- Petrov DA. 2002. Mutational equilibrium model of genome size evolution. *Theoretical Population Biology* **61**(4): 531-544.
- Pich U, Fuchs J, Schubert I. 1996. How do Alliaceae stabilize their chromosome ends in the absence of TTTAGGG sequences? *Chromosome Research* **4**(3): 207-213.
- Pich U, Schubert I. 1998. Terminal heterochromatin and alternative telomeric sequences in *Allium cepa*. *Chromosome Res* **6**(4): 315-321.
- Piegu B, Guyot R, Picault N, Roulin A, Sanyal A, Kim H, Collura K, Brar DS, Jackson S, Wing RA, *et al.* 2006. Doubling genome size without polyploidization: dynamics of retrotransposition-driven genomic expansions in *Oryza australiensis*, a wild relative of rice. *Genome Research* **16**(10): 1262-1269.
- Płachno BJ, Adamus K, Faber J, Kozłowski J. 2005. Feeding behaviour of carnivorous *Genlisea* plants in the laboratory. *Acta Botanica Gallica* **152**(2): 159-164.

- Plachno BJ, Kizskurno MK, Swiatek P. 2007. Functional ultrastructure of *Genlisea* (Lentibulariaceae) digestive hairs. *Annals of Botany* **100**: 195-203.
- Plohl M, Mestrovic N, Mravinac B. 2014. Centromere identity from the DNA point of view. *Chromosoma* **123**(4): 313-325.
- Pontes O, Neves N, Silva M, Lewis MS, Madlung A, Comai L, Viegas W, Pikaard CS. 2004. Chromosomal locus rearrangements are a rapid response to formation of the allotetraploid *Arabidopsis suecica* genome. *Proceedings of the National Academy of Sciences of the United States of America* **101**(52): 18240-18245.
- Raskina O, Barber JC, Nevo E, Belyayev A. 2008. Repetitive DNA and chromosomal rearrangements: speciation-related events in plant genomes. *Cytogenetic and Genome Research* **120**(3-4): 351-357.
- Raskina O, Belyayev A, Nevo E. 2004. Quantum speciation in *Aegilops*: molecular cytogenetic evidence from rDNA cluster variability in natural populations. *Proc Natl Acad Sci U S A* **101**(41): 14818-14823.
- Renny-Byfield S, Wendel JF. 2014. Doubling down on genomes: polyploidy and crop plants. *American Journal of Botany* **101**(10): 1711-1725.
- Reut MS. 1993. Trap Structure of the Carnivorous Plant *Genlisea* (Lentibulariaceae). *Botanica Helvetica* **103**(1): 101-111.
- Richards EJ, Ausubel FM. 1988. Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell* **53**: 127-136.
- Rogers SO, Bendich AJ. 1987. Ribosomal RNA genes in plants: variability in copy number and in the intergenic spacer. *Plant Molecular Biology* **9**(5): 509-520.
- Schaferhoff B, Fleischmann A, Fischer E, Albach DC, Borsch T, Heubl G, Muller KF. 2010. Towards resolving Lamiales relationships: insights from rapidly evolving chloroplast sequences. *BMC Evolutionary Biology* **10**: 352.
- Schmidt-Lebuhn AN, Fuchs J, Hertel D, Hirsch H, Toivonen J, Kessler M. 2010. An Andean radiation: polyploidy in the tree genus *Polylepis* (Rosaceae, Sanguisorbeae). *Plant Biology* **12**(6): 917-926.
- Schmidt T. 1999. LINES, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. *Plant Molecular Biology* **40**(6): 903-910.
- Schreck RR, Disteche CM. 2001. Chromosome banding techniques. *Curr Protoc Hum Genet* **Chapter 4**: Unit4.2.
- Schubert I. 2007. Chromosome evolution. *Current Opinion in Plant Biology* **10**(2): 109-115.
- Schubert I, Lysak MA. 2011. Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends in Genetics* **27**(6): 207-216.
- Schubert I, Pecinka A, Meister A, Schubert V, Matte M, Jovtchev G. 2004. DNA damage processing and aberration formation in plants. *Cytogenetic and Genome Research* **104**(1-4): 104-108.
- Schubert I, Rieger R. 1985. A new mechanism for altering chromosome number during karyotype evolution. *Theoretical and Applied Genetics* **70**(2): 213-221.
- Schubert I, Rieger R, Fuchs J. 1995. Alteration of basic chromosome number by fusion-fission cycles. *Genome* **38**(6): 1289-1292.
- Schubert I, Shaw P. 2011. Organization and dynamics of plant interphase chromosomes. *Trends in Plant Science* **16**(5): 273-281.
- Schubert I, Wobus U. 1985. In situ hybridization confirms jumping nucleolus organizing regions in *Allium*. *Chromosoma* **92**(2): 143-148.
- Schwarzacher T, Leitch A. 1994. Enzymatic Treatment of Plant Material to Spread Chromosomes for In Situ Hybridization. In: Isaac P ed. *Protocols for Nucleic Acid Analysis by Nonradioactive Probes*: Humana Press, 153-160.
- Shakirov EV, Salzberg SL, Alam M, Shippen DE. 2008. Analysis of *Carica papaya* Telomeres and Telomere-Associated Proteins: Insights into the Evolution of Telomere Maintenance in Brassicales. *Trop Plant Biol* **1**(3-4): 202-215.
- Shapiro JA, von Sternberg R. 2005. Why repetitive DNA is essential to genome function. *Biological Reviews* **80**(2): 227-250.
- Shi JH, Dawe RK. 2006. Partitioning of the maize epigenome by the number of methyl groups on histone H3 lysines 9 and 27. *Genetics* **173**(3): 1571-1583.
- Shibata F, Hizume M. 2002. The identification and analysis of the sequences that allow the detection of *Allium cepa* chromosomes by GISH in the allodiploid *A. wakegi*. *Chromosoma* **111**(3): 184-191.
- Shibata F, Nagaki K, Yokota E, Murata M. 2013. Tobacco karyotyping by accurate centromere identification and novel repetitive DNA localization. *Chromosome Res* **21**(4): 375-381.
- Shibata F, Sahara K, Naito Y, Yasukochi Y. 2009. Reprobing multicolor FISH preparations in lepidopteran chromosome. *Zoolog Sci* **26**(3): 187-190.

- Shirasu K, Schulman AH, Lahaye T, Schulze-Lefert P. 2000. A contiguous 66-kb barley DNA sequence provides evidence for reversible genome expansion. *Genome Research* **10**(7): 908-915.
- Simkova H, Cihalikova J, Vrana J, Lysak MA, Dolezel J. 2003. Preparation of HMW DNA from plant nuclei and chromosomes isolated from root tips. *Biologia Plantarum* **46**(3): 369-373.
- Simkova H, Safar J, Kubalaková M, Suchanková P, Cihalikova J, Robert-Quatre H, Azhaguvel P, Weng Y, Peng J, Lapitan NL, *et al.* 2011. BAC libraries from wheat chromosome 7D: efficient tool for positional cloning of aphid resistance genes. *Journal of Biomedicine and Biotechnology* **2011**: 302543.
- Soltis DE, Soltis PS, Bennett MD, Leitch IJ. 2003. Evolution of genome size in the angiosperms. *American Journal of Botany* **90**(11): 1596-1603.
- Swennenhuis JF, Foulk B, Coumans FA, Terstappen LW. 2012. Construction of repeat-free fluorescence in situ hybridization probes. *Nucleic Acids Research* **40**(3): e20.
- Swift H. 1950. The constancy of desoxyribose nucleic acid in plant nuclei. *Proc Natl Acad Sci U S A* **36**(11): 643-654.
- Sykorova E, Fajkus J, Meznikova M, Lim KY, Neplechova K, Blattner FR, Chase MW, Leitch AR. 2006. Minisatellite telomeres occur in the family Alliaceae but are lost in Allium. *American Journal of Botany* **93**(6): 814-823.
- Sykorova E, Lim KY, Chase MW, Knapp S, Leitch IJ, Leitch AR, Fajkus J. 2003a. The absence of Arabidopsis-type telomeres in Cestrum and closely related genera Vestia and Sessea (Solanaceae): first evidence from eudicots. *Plant Journal* **34**(3): 283-291.
- Sykorova E, Lim KY, Fajkus J, Leitch AR. 2003b. The signature of the Cestrum genome suggests an evolutionary response to the loss of (TTTAGGG)_n telomeres. *Chromosoma* **112**(4): 164-172.
- Sykorova E, Lim KY, Kunicka Z, Chase MW, Bennett MD, Fajkus J, Leitch AR. 2003c. Telomere variability in the monocotyledonous plant order Asparagales. *Proc Biol Sci* **270**(1527): 1893-1904.
- Tate JA, Joshi P, Soltis KA, Soltis PS, Soltis DE. 2009. On the road to diploidization? Homoeolog loss in independently formed populations of the allopolyploid *Tragopogon miscellus* (Asteraceae). *BMC Plant Biology* **9**: 80.
- Taylor P. 1991. The genus *Genlisea*. *Carniv. Pl. Newslett.* **20**(1-2): 14-20.
- Thiery JP, Macaya G, Bernardi G. 1976. An analysis of eukaryotic genomes by density gradient centrifugation. *Journal of Molecular Biology* **108**(1): 219-235.
- Thomas CA, Jr. 1971. The genetic organization of chromosomes. *Annual Review of Genetics* **5**: 237-256.
- Trojer P, Reinberg D. 2007. Facultative heterochromatin: Is there a distinctive molecular signature? *Molecular Cell* **28**(1): 1-13.
- Veleba A, Bures P, Adamec L, Smarda P, Lipnerova I, Horova L. 2014. Genome size and genomic GC content evolution in the miniature genome-sized family Lentibulariaceae. *New Phytologist* **203**(1): 22-28.
- Vitte C, Panaud O. 2003. Formation of solo-LTRs through unequal homologous recombination counterbalances amplifications of LTR retrotransposons in rice *Oryza sativa* L. *Molecular Biology and Evolution* **20**(4): 528-540.
- Vu GT, Schmutzer T, Bull F, Cao HX, Fuchs J, Tran TD, Jovtchev G, Pistrick K, Stein N, Pecinka A, Neumann P, Novak P, Macas J, Dear PH, Blattner FR, Scholz U and Schubert I. 2015. Comparative genome analysis revealed divergent genome size evolution in a carnivorous plant genus. *The Plant Genome*. (submitted)
- Wako T, Fukui K. 2010. Higher organization and histone modification of the plant nucleus and chromosome. *Cytogenetic and Genome Research* **129**: 55-63.
- Wang K, Guan B, Guo W, Zhou B, Hu Y, Zhu Y, Zhang T. 2008. Completely distinguishing individual A-genome chromosomes and their karyotyping analysis by multiple bacterial artificial chromosome - fluorescence in situ hybridization. *Genetics* **178**(2): 1117-1122.
- Wang XW, Wang HZ, Wang J, Sun RF, Wu J, Liu SY, Bai YQ, Mun JH, Bancroft I, Cheng F, *et al.* 2011. The genome of the mesopolyploid crop species *Brassica rapa*. *Nature Genetics* **43**(10): 1035-U1157.
- Weiss-Schneeweiss H, Emadzade K, Jang TS, Schneeweiss GM. 2013. Evolutionary consequences, constraints and potential of polyploidy in plants. *Cytogenetic and Genome Research* **140**(2-4): 137-150.
- Weiss H, Scherthan H. 2002. Aloe spp. – plants with vertebrate-like telomeric sequences. *Chromosome Research* **10**(2): 155-164.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, *et al.* 2007. A unified classification system for eukaryotic transposable elements. *Nature Reviews: Genetics* **8**(12): 973-982.
- Wicker T, Wing RA, Schubert I. 2015. Inter-chromosomal sequence exchange in a peculiar group of grass chromosomes. *The Plant Journal*. (submitted)

- Wolf YI, Koonin EV. 2013.** Genome reduction as the dominant mode of evolution. *Bioessays* **35**(9): 829-837.
- Xiong Z, Pires JC. 2011.** Karyotype and identification of all homoeologous chromosomes of allopolyploid *Brassica napus* and its diploid progenitors. *Genetics* **187**(1): 37-49.
- Young HA, Sarath G, Tobias CM. 2012.** Karyotype variation is indicative of subgenomic and ecotypic differentiation in switchgrass. *BMC Plant Biology* **12**: 117.
- Zhang HB, Wu C. 2001.** BAC as tools for genome sequencing. *Plant Physiol. Biochem* **39**: 195-209.
- Zhao M, Ma J. 2013.** Co-evolution of plant LTR-retrotransposons and their host genomes. *Protein Cell* **4**(7): 493-501.

Publications related to this Dissertation

Journal articles

Vu GT, Schmutzer T, Bull F, Cao HX, Fuchs J, **Tran TD**, Jovtchev G, Pistrick K, Stein N, Pecinka A, Neumann P, Novak P, Macas J, Dear PH, Blattner FR, Scholz U and Schubert I (2015). Comparative genome analysis revealed divergent genome size evolution in a carnivorous plant genus. *The Plant Genome*. (submitted)

Tran TD, Cao HX, Jovtchev G, Neumann P, Novak P, Fojtova M, Vu GT, Macas J, Fajkus J, Schubert I and Fuchs J (2015). Centromere and telomere sequence alterations reflect the rapid genome evolution within the carnivorous plant genus *Genlisea*. *The Plant Journal*. (in revision)

Tran TD, Cao HX, Jovtchev G, Novak P, Macas J, Vu GT, Schubert I and Fuchs J (2015). Chromatin organization and cytological features of carnivorous *Genlisea* species with large genome size differences. *Frontiers in Plant Genetics and Genomics*. (submitted)

Poster presentation

Tran TD, Jovtchev G, Neumann P, Macas J, Vu GT, Fuchs J and Schubert I (2012). Cytological characterization of two *Genlisea* species with a twenty-fold genome size difference. *The 10th International PhD Student Conference on Experimental Plant biology*. Brno, Czech Republic. (1st prize best poster).

Declaration about Personal Contributions

Data presented within this dissertation were generated by me with the great help and contribution from other people as the follows:

- Flow-cytometry works (genome size measurement and interphase nuclei sorting) were done by **Dr. J. Fuchs**.

- The immunostaining using anti-DNA and histone methylation antibodies were initially performed for *A. thaliana*, *G. aurea* and *G. hispidula* by **Dr. G. Jovtchev**. I continued and accomplished the experiment for other studied species.

- The *Genlisea* repetitive sequences (including centromeric and telomeric repeats) were identified and specific primers were designed by **Dr. P. Neumann**, **Dr. P. Novak** and **Dr. J. Macas**.

- The *G. nigrocaulis* single-copy fragments were identified and specific primers (including those for 5S rDNA gene) were designed by **Dr. H.X. Cao**. The FISH images of chromosomal localization of (TTCAGG) telomeric repeat and of the Gh336c35 repeat in *G. subglabra* were also kindly provided by **Dr. H.X. Cao**.

- TRF and TRAP for telomere analysis were done by **Dr. M. Fojtova** and **Prof. J. Fajkus**.

- The *G. margaretae* BAC library was constructed by **Dr. H. Simkova**. Colony filters for whole BAC library were prepared by **I. Walde** with my assistance. Dot-blot DNA hybridization for repeat-free BAC clone screening was carried out by **Dr. R. Schmidt** and **K. Langanke**. Repeat-free BAC clones selection and re-gridding were done by me.

Curriculum Vitae

Personal information

Name:	Tran Duc Trung
Gender:	Male
Birth:	March 8 th , 1983 in Hanoi, Vietnam
Nationality:	Vietnamese
Marital status:	Married
Residential address in home country:	No. 20, lane 550, La Thanh Stress, Ba Dinh District, Hanoi, Vietnam
Residential address in Germany:	Selkeweg 4A, OT Gatersleben, 06466 Stadt Seeland
Email:	duc@ipk-gatersleben.de; duc trung83@gmail.com

Academic qualification and Employment

- Since 2011: PhD student
Leibniz Institute of Plant Genetics and Crop Plant Research (IPK),
Gatersleben, Germany
Research group Quantitative Genetics
(Former Karyotype Evolution group, Prof. Dr. Ingo Schubert)
- 2010-2011: Researcher, Plant Resources Center, Hanoi, Vietnam
- 2007-2009: Master degree in Biotechnology
Hanoi University of Science and Technology, Hanoi, Vietnam
- 2005-2010: Researcher, Agricultural Genetics Institute (AGI), Hanoi, Vietnam
- 2001-2005: Bachelor degree in Biotechnology
Hanoi Open University, Hanoi, Vietnam

Declarations concerning Originality of the Dissertation and Criminal Record and Pending Investigations

I confirm that this dissertation was solely undertaken by me at the Research Group Quantitative Genetics (former Karyotype Evolution group) of the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, Germany, and that all literature and data provided from other sources were properly cited, declared and acknowledged (for contribution of collaboration partners see page 90). The content of this work has not been previously submitted for a degree in any education institution.

I hereby declare that I have no criminal record and that no preliminary investigations are pending against me.

Gatersleben, May 25th, 2015

Tran Duc Trung