

**Comparative cytology and cytogenomics for
representative species of the five duckweed genera**

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

Der

Naturwissenschaftlichen Fakultät III
Agrar- und Ernährungswissenschaften,
Geowissenschaften und Informatik

der Martin-Luther-Universität Halle-Wittenberg

vorgelegt von

Frau Phuong Thi Nhu Hoang
Geb. am November 23rd, 1983 in Lam Dong, Viet Nam

Gutachter:

1. Prof. Dr. Jochen Reif and Prof. Dr. Ingo Schubert
IPK, Gatersleben, Germany
2. Prof. Dr. Thomas Schmidt
Institut für Botanik, TU Dresden, Dresden, Germany

Verteidigung am 21. Januar 2019

Acknowledgements

This work was performed from January 2015 till August 2018 at the Leibniz Institute of Plant Genetics and Crop Plant Research (IPK), Gatersleben, funded by the Deutsche Forschungsgemeinschaft (DFG) and supported by scholarship from the Ministry of Education and Training (MOET) of Vietnam.

Foremost, my deepest appreciation to my supervisor Prof. Ingo Schubert for giving me the opportunity to be part of his team, for continuous guidance, permanent encouragement as well as fruitful discussions. His conscientious guidance helped me in all the time of research and writing of this dissertation.

I own deep thanks to my initial co-supervisor Dr. Hieu X. Cao, for his orientation, constructive and scholarly advises at the beginning of my study.

Also, I would like to thank Prof. Dr. Jochen C. Reif, the Head of Department of Breeding Research who gave me a great opportunity to be a PhD student in his Department and agreed to act as supervisor from the Martin-Luther-University Halle-Wittenberg. I also gratefully thank Dr. Britt Leps for all of her help in administrative issues, which made my stay at IPK very comfortable. Special thanks to PD. Klaus Appenroth for kind support in duckweed clone selection and critical discussion.

I would like to extend my thanks to Dr. Joerg Fuchs, Dr. Veit Schubert for their insightful contribution to this work. Many thanks go to Martina Kuehne, Andrea Kunze and Joachim Bruder, for their excellent technical assistance.

My grateful thanks go to Prof. Eric Lam (Department of Plant Biology, Rutgers the State University of New Jersey, USA), Dr. Todd P Michael (J. Craig Venter Institute, Carlsbad, CA, USA) for providing Oxford Nanopore sequencing results. I want to say thanks to Prof. Joachim Messing and to Paul Fourounjian (Waksman Institute Rutgers University, USA) for kindly providing their BAC library. Many thanks also to Dr. Uwe Scholz, Dr. Anne Fiebig (IPK – Gatersleben) for their bioinformatics work on *S. intermedia* genome assembly.

I thank to Dr. Hieu X. Cao, Dr. Giang T.H. Vu and Dr. Van T.T. Tran, who introduced me to Prof. Dr. Ingo Schubert and helped me at the beginning of my stay in Germany.

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, encouraged and supported me a lot in scientific work.

Last but not the least, I would like to express my very profound gratitude to my parents and my parents-in-law for providing me with immeasurable love, limitless sacrifice, unfailing support and continuous encouragement throughout my years of study. This is the time for me to express my thankfulness to my husband and my children, who give me a lot of strength and motivation during the last four years by their endless love, unconditional belief and deep empathy. This accomplishment would not have been possible without their understanding and moral support.

Phuong Thi Nhu Hoang

TABLE OF CONTENT

List of figures	i
List of tables	iii
List of abbreviations.....	iv
1. INTRODUCTION	1
1.1. Plant genomes, genome size variation and karyotype evolution	1
1.1.1. Plant genome structure and organization	1
1.1.2. Genome size and genome size variation.....	4
1.1.3. Karyotypes and karyotype evolution.....	5
1.2. Duckweeds are interesting subjects for genome and karyotype evolution research and are potential aquatic crops	8
1.2.1. Why are duckweeds of interest for genome and karyotype evolution studies?	8
1.2.2. What makes duckweeds becoming potential aquatic crops?	10
1.2.3. Some landmarks of (mainly) genome research on duckweeds	11
1.3. Whole genome sequencing, genome maps and chromosome numbers of duckweeds	11
1.3.1. Whole genome sequencing	11
1.3.2. Genome maps.....	13
1.3.2.1. <u>The cytogenetic map of the Greater duckweed – <i>S. polyrhiza</i>.....</u>	15
1.3.2.2. <u>The optical map of the Greater duckweed – <i>S. polyrhiza</i>.....</u>	15
1.3.3. The chromosome numbers of duckweeds.....	17
1.4. Aims of the dissertation	18
2. MATERIALS AND METHODS.....	20
2.1. Plant material and cultivation	20
2.2. Genomic DNA isolation and metaphase preparation.....	21
2.3. Genome size measurement.....	22
2.4. Epidermis preparation, microscopic cell and nuclear volume measurements, and statistics	23
2.5. Probe preparation	24
2.5.1. 5S/18S/ 26S rDNA and telomere probes.....	24
2.5.2. Bacterial artificial chromosome DNA probes	25
2.6. Fluorescence in situ hybridization	26
2.7. <i>S. intermedia</i> whole genome sequencing and assembly	27

2.7.1. Plant material and DNA extraction.....	27
2.7.2. Genome sequencing and assembly	27
2.7.3. Scaffolding and gap filling.....	27
2.7.4. Gene prediction	28
2.7.5. Repeat identification	28
3. RESULTS AND DISCUSSION.....	29
3.1. Morphology variation and correlation between genome size and cell parameters in duckweeds	29
3.2. Chromosome numbers and number of 5S and 45S rDNA sites in duckweeds	37
3.2.1. Chromosome numbers.....	37
3.2.2. Ribosomal rDNA sites	41
3.3. A robust genome map for <i>S. polyrhiza</i>	46
3.4. Karyotype evolution between the two species of the ancient duckweed genus <i>Spirodela</i>.....	56
3.4.1. Chromosome homeology between <i>S. polyrhiza</i> and <i>S. intermedia</i>	56
3.4.2. Six new linkage groups in <i>S. intermedia</i> were revealed by FISH.....	58
3.4.3. Supposed karyotype evolution scenarios between two <i>Spirodela</i> species	64
3.4.3.1. <u>Karyotype evolution towards <i>S. intermedia</i> (n=18)</u>	64
3.4.3.2. <u>Karyotype evolution towards <i>S. polyrhiza</i> (n=20)</u>	66
3.4.4. Cytogenetic map of <i>S. intermedia</i>	68
3.5. Whole genome sequencing and genome assembly in <i>S. intermedia</i>.....	70
3.6. Polyploidy in duckweeds	73
4. SUMMARY	79
5. ZUSAMMENFASSUNG	82
6. REFERENCES	85
Curriculum Vitae	
Publications	
Poster and oral presentations	
Attended conferences	
Declaration about Personal Contributions	
Declaration concerning Criminal Record and Pending Investigations	
Declaration under Oath	

List of figures

Figure 1. Secondary (A) and dysploid (B) chromosome rearrangements	7
Figure 2. Duckweed morphology	9
Figure 3. Phylogenetic relationship, frond, stomata and nuclei morphology of duckweed species.	30
Figure 4. Variation in cell morphology (A), floating-style (B) and genome size (C) in duckweed	32
Figure 5. Variation in guard cell shape and volume of <i>Le. aequinoctialis</i> (clone 2018) (A), chromosome spreads of <i>Le. aequinoctialis</i> clones 2018 and 6746 (B), equal and abnormal nuclei distribution in sister guard cells of <i>Wa. hyalina</i> (C1-3) and <i>Wo. australiana</i> (C4-6)	35
Figure 6. Guard cell and nuclear volume measurement (A) and linear regressions of duckweed cell parameters (B)	36
Figure 7. Chromosome number of distinct clones of eleven duckweed species	40
Figure 8. Chromosomal distribution of 5S and 45S rDNA on <i>S. polyrhiza</i>	42
Figure 9. 5S and 45S rDNA loci on duckweed species	44
Figure 10. rDNA FISH signals in pachytene (A) and mitotic metaphase (B) of <i>Wa. rotunda</i> (clone 9072) using super-resolution microscopy (SIM)	45
Figure 11. Chromosomal distribution of pseudomolecules 08 and 04 on <i>S. polyrhiza</i>	49
Figure 12. Location of chimeric pseudomolecule Ψ 16	50
Figure 13. Location of chimeric pseudomolecule Ψ 11	50
Figure 14. Location of chimeric pseudomolecule Ψ 14	51
Figure 15. Location of Ψ 21b on <i>S. polyrhiza</i> chromosome ChrS 14	52
Figure 16. Location of Ψ 21a on <i>S. polyrhiza</i> chromosome ChrS 08	52
Figure 17. Solving discrepancies between the cytogenetic map (blue) and the BioNano map (red) resulted in an updated map (orange) of <i>S. polyrhiza</i>	53
Figure 18. 834 kb mis-assembly in BioNano map was detected by Oxford Nanopore and confirmed by FISH	54
Figure 19. The complete karyotype of <i>S. polyrhiza</i> clone 9509.	55
Figure 20. Multi-color FISH of 20 <i>S. polyrhiza</i> chromosome-specific probes to somatic metaphase chromosomes of <i>S. intermedia</i> (8410).	57

Figure 21. Six new linkage groups in <i>S. intermedia</i> are uncovered by subsequent mc-FISH.....	59
Figure 22: Three-color FISH on <i>S. intermedia</i> using single BACs from <i>S. polyrhiza</i> chromosome-specific probes that label more than one on <i>S. intermedia</i> chromosome to define the split-points.....	61
Figure 23: Three-color FISH using BACs from <i>S. polyrhiza</i> to prove the composition of all six new linkages in <i>S. intermedia</i>	63
Figure 24: Karyotype evolution towards <i>S. intermedia</i> (n=18) in case the ancestral karyotype was similar to that of <i>S. polyrhiza</i> (n=20).	65
Figure 25: Karyotype evolution towards <i>S. polyrhiza</i> (n=20) in case the ancestral karyotype was similar to that of <i>S. intermedia</i> (n=18).....	67
Figure 26: Distribution of 20 <i>S. polyrhiza</i> chromosome probes on <i>S. intermedia</i> metaphases.....	69
Figure 27. BUSCO assessment results.	72
Figure 28. Chromosome, 5S and 45S loci number (A) and correlation of guard cell parameters (B) in diploid and tetraploid clones of <i>Le. aequinoctialis</i>	74
Figure 29. Chromosome, 5S and 45S loci number (A) and correlation of guard cell parameters (B) in diploid and tetraploid clones of <i>La. punctata</i>	75
Figure 30. Cross-FISH with single copy BACs of <i>S. polyrhiza</i> on mitotic spreads of <i>La. punctata</i> (clone 7260).....	77

List of tables

Table 1: Duckweed chromosome numbers from literature	17
Table 2: List of duckweed species used in this study	20
Table 3: Procedures for preparation of duckweed chromosomes.....	22
Table 4: List of primers used to amplify rDNA regions.	24
Table 5: Cytological characterization of the tested duckweeds species	33
Table 6: Chromosome numbers of tested duckweed species from literature and our study	38
Table 7: Differences in chromosome enumeration (A) and chromosomal assignment of pseudomolecules (B) between <i>S. polyrhiza</i> cytogenetic map (for clone 7498) and BioNano map (for clone 9509)	46
Table 8: 106 BACs of the 20 <i>S. polyrhiza</i> chromosomes integrating 39 pseudomolecules (including Ψ 0)	47
Table 9: Components of the 18 <i>S. intermedia</i> chromosomes based on 93 anchored <i>S. polyrhiza</i> BACs	67
Table 10: <i>S. intermedia</i> sequence assembly statistics	71
Table 11: Cytological characterization of <i>La. punctata</i> clones 7260 and 5562_A4 and <i>Le. aequinoctialis</i> clones 2018 and 6746.	76
Table 12: Results of cross-FISH on <i>La. punctata</i> (clone 7260).	78

List of abbreviations

Alexa 488	Alexa Fluor 488 dye, a bright green-fluorescent dye
BAC	Bacterial artificial chromosome
BUSCO	Benchmarking Universal Single-Copy Orthologs
bp	Base pair
DAPI	4',6-diamidino-2-phenylindole
DNA	Deoxyribonucleic acid
DSB	Double-strand break
dNTP	Deoxynucleotide triphosphate
dUTP	Deoxyuridine triphosphate
EDTA	Ethylenediaminetetraacetic acid
FCM	Flow Cytometry
FISH	Fluorescence <i>in situ</i> hybridization
FPC	finger printed contig
HR	homologous recombination
kbp	kilo base pair
LTR	Long terminal repeat
Mbp	Mega base pair
Mya	Million years ago
NHEJ	non-homologous end-joining
NOR	Nucleolus organizer region
rDNA	ribosomal DNA
PCR	Polymerase chain reaction
TE	Transposable element
TexasRed	sulforhodamine 101 acid chloride, a red-fluorescent dye
WGD	Whole genome duplication
µm	micrometer

1. INTRODUCTION

1.1. *Plant genomes, genome size variation and karyotype evolution*

1.1.1. *Plant genome structure and organization*

The heritable information of living beings is stored in the base sequence of deoxyribonucleic acid (DNA). Most of the DNA of eukaryotes is located within the cell nucleus and is called the genome. The genomic DNA together with histones and other nuclear proteins forms the chromatin which is organized in a species-specific number of linear chromosomes. The chromosomes of the genome are maintained and segregated to the next cellular and organismic generation via nuclear division cycles. For correct segregation, the chromosomes are replicated into identical sister chromatids. To ensure cellular functions such as metabolism, growth and differentiation, certain parts of DNA (genes) are transcribed into RNA during interphase between nuclear divisions.

Two categories of DNA sequences are contained in the genomes of all eukaryotes are (1) single- or low-copy sequences comprising genes (exons, introns), promoter and regulatory elements, and (2) high-copy or repetitive sequences. Annotation of complete plant genomes has revealed that plants have ten thousands of genes. For instance, 31 407 genes are documented in The Arabidopsis Information Resource6 (with 26 751 protein-coding genes, 3 818 pseudogenes, and 838 non-coding RNA genes) or more than 41 000 genes in the rice genome (Sterck *et al.*, 2007).

Major contributors to plant genome size are tandem and dispersed repetitive DNA with hundreds or even thousands of copies, which may be located at a few defined chromosomal sites or widely dispersed.

Tandemly repeated or satellite DNA consists of a motif that is repeated in many copies at one or more genomic locations. Microsatellite, minisatellite and satellite DNA are the three major types of tandem repetitive DNA sequences, distinguished by the length of basic repeat unit: (1) Microsatellite units (less than 9bp) present in both non-coding and coding regions with up to 1 kbp; (2) Minisatellite units (from 9 to 100 bp) may extend up to several kbp and cluster in subtelomeric, pericentromeric or interstitial regions of chromosomes; (3) Satellite DNAs with a monomer length ranging from 100 to >1 000 bp may constitute Mbp-long arrays. Whether tandem repetitive sequences have a function in the genome is in most cases unknown (Lopez-Flores and Garrido-Ramos, 2012; Robledillo *et al.*, 2018). Well-defined are

the functions of specific repetitive sequences such as telomeric and ribosomal RNA encoding sequences. Telomeres are specific structures that protect the ends of linear eukaryotic chromosomes against enzymatic degradation, fusion with neighboring chromosomes and chromosome shortening during replication caused by the inability of DNA-polymerases to fully synthesize 5' ends of DNA (for review see (O'Sullivan and Karlseder, 2010)). Telomeres are composed of rather conserved short G-rich repeats with slightly different motifs: Arabidopsis-type (TTTAGGG) (Richards and Ausubel, 1988), vertebrate-type (TTAGGG) (Moyzis *et al.*, 1988), Tetrahymena-type (TTGGGG) (Sheng *et al.*, 1995), Bombyx-type (TTAGG) (Okazaki *et al.*, 1993), Chlamydomonas-type (TTTTAGGG) (Petracek and Berman, 1992) or Oxytricha-type (TTTTGGGG) (Melek *et al.*, 1994). A few plant species show C in the G-rich strand such as *Genlisea hispidula* with TTCAGG/TTTCAGG (Tran *et al.*, 2015) and/or are unusually long (12 bp) as in the genus *Allium* (CTCGGTTATGGG, see (Fajkus *et al.*, 2016). Ribosomal RNA genes encode the RNA components of ribosomes, the 'protein factories' of every cell. 5S rDNA genes encoding small ribosomal RNA and its intergenic spacer are transcribed by RNA polymerase III, and 45S rDNA genes encoding the large ribosomal RNA components 18S, 5.8S, 26S as well as internal transcribed spacer and external transcribed spacer regions are transcribed by RNA polymerase I (Paule and White, 2000). 45S rDNA may be arrayed in hundreds to ten thousands of copies at the so-called nucleolus organizing regions (NORs). For instance 45S rDNA comprises 150 copies in *Saccharomyces cerevisiae* (~12.2 Mbp/1C) (Kobayashi, 2014); or 570 copies in *A. thaliana* (157 Mbp/1C) (Pruitt and Meyerowitz, 1986); or up to 12 000 copies in *Zea mays* with 2 500 Mbp/1C (Buescher *et al.*, 1984). Similar to telomeric repeats, rDNA sequences are highly conserved. Thus 45S and 5S rDNA which usually display a species-specific, clustered distribution are frequently used as markers for karyotyping by FISH.

Centromeres are chromosome regions where spindle microtubules attach to the sister chromatids to enable their movement to the daughter nuclei during cell divisions in eukaryotes. During the evolution of plants, different centromere types appeared which differ by the distribution of nucleosomes having the centromeric histone variant CenH3 instead of histone H3. Cereals (Ishii *et al.*, 2015) and many other taxa have monocentric chromosome, *Pisum sativum* and *Lathyrus* (Neumann *et al.*, 2016) have several clusters of CenH3 nucleosomes within a distinct region, while in *Rhynchospora pubera* (Marques *et al.*, 2016) such clusters are found along

their (polycentromeric) chromosomes and in *Luzula* (Wanner *et al.*, 2015; Heckmann *et al.*, 2014) CenH3 nucleosomes seem to be evenly distributed along the (holocentric) chromosomes. In holocentrics the spindle fibers attach along the entire chromosome. Monocentric chromosomes can be classified as metacentric, sub-metacentric, acrocentric or telocentric chromosomes according to the position of their centromere (Schubert, 2007). Centromeres are also often composed of satellite sequences and retroelements. However because during evolution centromeres are dynamic and can originate *de novo* at positions without repetitive sequences (for review see (Schubert, 2018)), it is not yet clear whether centromeres are just a place where repeats can accumulate without becoming deleterious, or whether they are indeed supportive for centromere function.

Dispersed repetitive DNA represents the highest proportion of repetitive DNA and consists of transposable elements (TEs), which often include sequences that encode enzymes for their own replication and integration into the nuclear DNA (Heslop-Harrison and Schwarzacher, 2011). Two classes of TEs were classified based on their structural features and mechanisms of transposition: retrotransposons (class I, transposing via 'copy and paste' mechanism) and DNA transposons (class II, transposing via 'cut and paste' mechanism) (Schmidt, 1999; Wicker *et al.*, 2007). The abundance and diversity of TEs within the genome are variable among eukaryotes. In some species such as maize and barley, LTR elements may occupy up to 75% of the genome and scatter throughout most of chromosomes (Mayer *et al.*, 2012; Baucom *et al.*, 2009). Ty1/copia and Ty3/gypsy are the most ubiquitous families of dispersed DNA elements in investigated plant species (Wicker *et al.*, 2007).

In addition to the various blocks of repetitive DNA, many plant genomes may contain different numbers of accessory chromosomes, so-called B-chromosomes. These are highly condensed chromosomes harboring few and often truncated genes but many repetitive sequences. B-chromosomes show non-Mendelian modes of inheritance called 'drive'. This drive (preferential transmission of B-chromosomes into gametes) ensures their maintenance as 'parasites' within the host genome (for review see (Houben, 2017)).

1.1.2. Genome size and genome size variation

The genome size (or “C-value”) of an organism is defined as the amount of nuclear DNA in the unreplicated, reduced gametic nucleus, irrespective of the ploidy level of the species (Fleury *et al.*, 2012). Genome size typically is measured in terms of either mass (pg) or the number of nucleotide base pairs (bp), 1 pg of double strand DNA equals 978 Mbp (Dolezel *et al.*, 2003). In general, nuclear genome size is constant within a given species, e.g. *Arabidopsis thaliana* has $2C = 0.321$ pg DNA, but it can strongly vary between species. For instance there is a 2 440-fold genome size difference between the so far smallest plant genome of *Genlisea tuberosa* with ~61 Mbp/1C (Fleischmann *et al.*, 2014) and the largest known plant genome of *Paris japonica* with 150 Gbp/1C (Pellicer *et al.*, 2010). Even within a species genome size can vary, e.g. among different accessions of *A. thaliana* (Schmuths *et al.*, 2004). Importantly, genome size is not associated with the complexity and evolutionary advancement or ecological competitiveness of an organism (Mirsky and Ris, 1951; Thomas, 1971). For instance plants with large genomes appear to have reduced photosynthetic efficiency and are underrepresented in extreme environments (Ross-Ibarra and Gaut, 2008).

Several hypothesis were suggested to explain this phenomenon called ‘C-value paradox’ (Thomas 1971), its causes, mechanism(s) and the biological significance of genome size variation. Recently three strategies were postulated for genome size evolution which might explain the C-value paradox: (1) Genome size reduction is assumed to result from more and larger deletions than insertions via deletion-biased DNA double-strand break (DSB) repair; (2) Genome size expansion may occur not only by WGD, but particularly by more and larger insertion than deletions via insertion-biased DSB repair, which includes spreading of retroelements; and (3) Genome size remains stable (stasis) when deletions and insertions during DSB repair are balanced. Based on selective forces and due to mutations in components of DSB repair, switches between these strategies may occur (Schubert and Vu, 2016).

There are some interesting correlations between genome size and cellular features of plants. For example, guard cell length appears to positively correlate with genome size across a wide range of major taxa with the exception of the Poaceae (Hodgson *et al.*, 2010). DNA content and nuclear volume as well as nuclear and cell volume

showed positive correlation at different endopolyploidy levels in epidermis cells of *A. thaliana* (from 2C to 32C), *Barbarea stricta* (from 2C to 16C) as well as between species that differ in genome size up to ~500 fold (from 0.32 pg in *A. thaliana* to 154.99 pg in *Fritillaria ulva-vulpis*) (Jovtchev *et al.*, 2006) or between 14 herbaceous angiosperm species (Price *et al.*, 1973). A correlation of cell parameters (DNA content, cell volume, nuclear volume, cell surface, nuclei surface) was also reported for *Sorghum bicolor* endosperm cells from 3C to 96C (Kladnik, 2015). Other phenotypical characteristics of large genomes, besides an increased cell size are slow mitotic activity, relative to small genome species. A positive correlation between genome size and cell cycle time was observed with maximum cell cycle length of 18 h in 52 eudicots and variation from 8 up to 120 h in 58 monocots (Francis *et al.*, 2008). Recently, Simonin and Roddy (2018) hypothesized a connection between genome size and cell size to interpret evolutionary angiosperm radiation. During the early Cretaceous period, genome downsizing occurred only in the angiosperm clade paralleled by smaller cell and stomata size as well as higher stomata and vein density. These factors allowed for greater CO₂ uptake and photosynthesis carbon gain, and presumably promoted angiosperms becoming the dominant plants in most terrestrial ecosystems (Simonin and Roddy, 2018).

1.1.3. Karyotypes and karyotype evolution

The karyotype is the chromosome complement of an organism. Karyotypes may differ regarding number, size and shape of their chromosomes. In diploid sexual organisms karyotypes consist of one paternal and one maternal chromosome set. Chromosome sets can be multiplied by whole genome duplication (WGD) resulting in polyploid karyotypes. WGD can yield auto- or allopolyploid organisms. Autopolyploidy results from a fusion of two unreduced gametes of the same species as in potato, watermelon, banana, and alfalfa. Allopolyploidy combines two or more genomes from different species as in wheat, cotton, tobacco, coffee, sugarcane, peanut, oat, and canola (Chen *et al.*, 2007). There are also examples, such as soybean, indicating that the genome has allo- and autopolyploid origins (Udall and Wendel, 2006). Natural polyploid crops provided an important tool for plant breeders since it allows exploitation of diversity from both diploid progenitors as sources of novel genes or alleles for crop improvement. For example, the diploid and tetraploid progenitors of hexaploid bread wheat have provided a critical source for resistance genes against diseases and abiotic stress, and even for quality genes (Feuillet and

Eversole, 2008). When multiples of genome size and chromosome number compared to the presumed ancestors are still recognizable, the organisms are considered as 'neopolyploids' (Wood *et al.*, 2009). In cases where chromosome numbers (and/or genome size) are no longer a multiple of the ancestral diploid state, but genome duplication is still cytologically detectable by *in situ* hybridization, we call the organisms 'mesopolyploid'. When multiples of genome size and of chromosome number are unrecognizable and genome duplication only is discovered by bioinformatics and sequence analysis we speak about 'paleopolyploids', which lost their polyploid status by accumulating mutations resulting in diploidization and are currently considered as diploids. For instance, *S. polyrhiza* ($2n = 40$) underwent two whole genome duplications of seven ancestral chromosome blocks (Cao *et al.*, 2016). Several studies have proven the widespread occurrence of paleopolyploidy in the angiosperms (Blanc and Wolfe, 2004), indicating that polyploidy plays an important role in plant evolution.

Besides polyploids, aneuploid karyotypes, in which the number of individual chromosomes is increased or decreased, may occur rarely. Particularly in diploid organisms the lack of one or both chromosomes of one or more pairs is usually lethal. In addition, structural chromosomal rearrangements (and extensive gene loss) may happen after WGD events leading to changes in size and structure of chromosomes. However, primary chromosome rearrangements including insertion, deletion, duplication, peri- or paracentric inversion and intra- or interchromosomal reciprocal translocation may also occur in diploid organisms. They are all the outcome of DSB mis-repair by joining of ends between different DSBs via non-homologous end-joining (NHEJ) or via homologous recombination (HR) using ectopic homologous sequences as repair template (Schubert, 2007). The chromosome structure can also be altered by secondary rearrangements, e.g. in organisms heterozygous for two translocations between three chromosomes (i.e., one chromosome is involved in both translocations). Crossing over in a meiotic hexavalent of such a double heterozygote between chromatids, which differ from each other in both ends flanking the exchange, results in gametes with a new secondarily rearranged karyotype and in re-established wild type gametes (Fig. 1A) (Schubert, 2007; Schubert and Lysak, 2011). Furthermore, dysploid chromosome rearrangements lead to chromosome number variation on different routes via reciprocal translocations (Fig. 1B) (Schubert and Lysak, 2011).

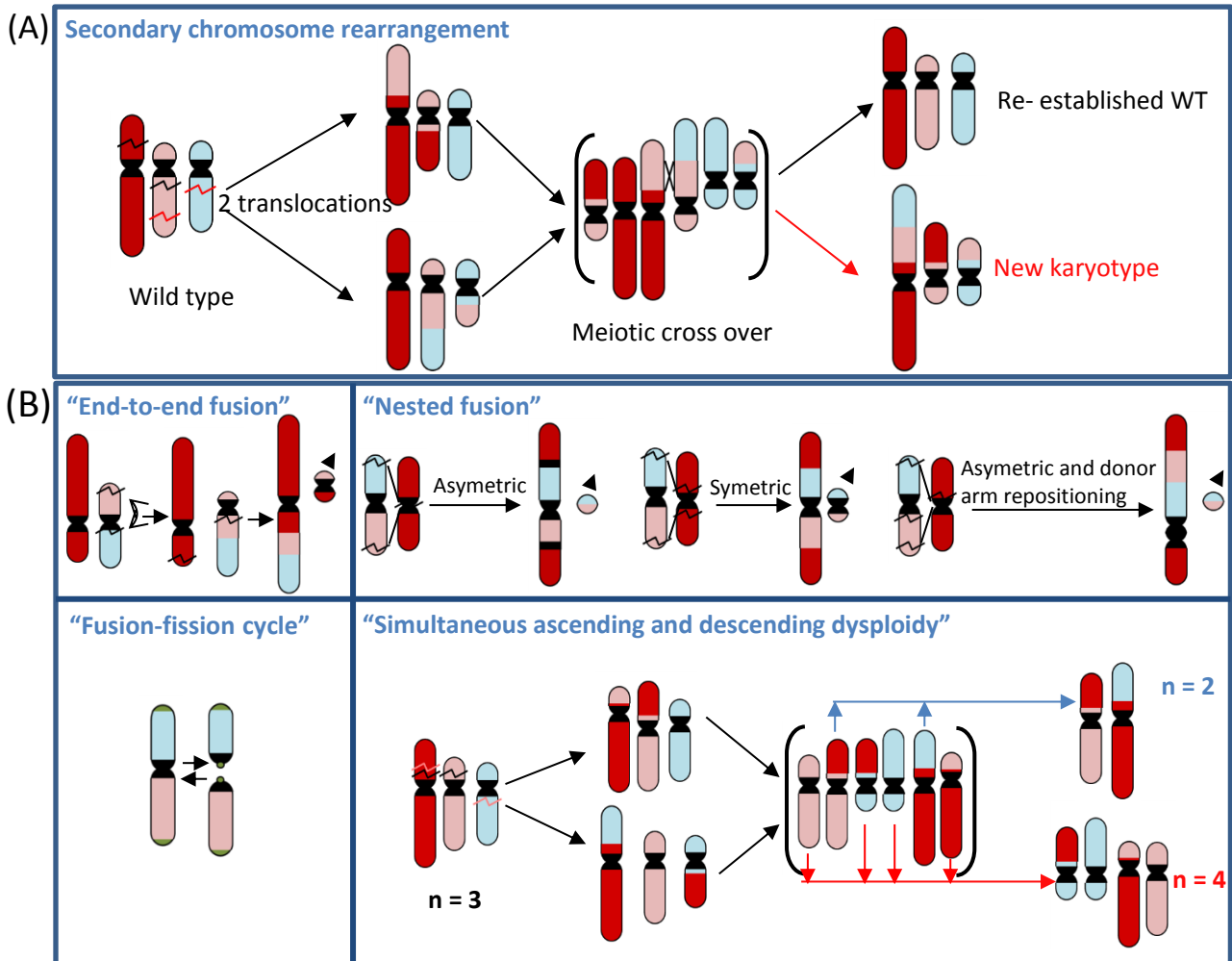


Figure 1. Secondary (A) and dysploid (B) chromosome rearrangements.

(A) Two translocations between three chromosomes followed by a meiotic cross over between two chromosomes, which are morphologically different on either side of the cross over, yield a gamete with a re-established wild-type karyotype and another one with a new karyotype; (B) Different routes of dysploid alteration of chromosome number via reciprocal translocations. (re-drawn from Schubert and Lysak, 2011)

Studies on evolution of plant genome architecture revealed that (1) in all plant genomes fractionation processes occurred after WGD events; (2) dynamic proliferation and loss of lineage-specific transposable elements constitutes the vast majority of the variation in genome size (Wendel *et al.*, 2016).

1.2. Duckweeds are interesting subjects for genome and karyotype evolution research and are potential aquatic crops

1.2.1. Why are duckweeds of interest for genome and karyotype evolution studies?

Duckweeds are small-sized, free-floating, aquatic plants with the fastest growth rate among flowering plants and with highly reduced and miniaturized organs. The two monographs on *Lemnaceae* of Elias Landolt provided fundamental insights regarding biodiversity, genetics, ecology, physiology and development of duckweeds (Landolt, 1987; 1986). More than 3 500 publications have cited these monographs (Tippery *et al.*, 2015).

Phylogenetically, duckweeds were considered by some authors as a subfamily (Lemnoideae) of the family Araceae (Cabrera *et al.*, 2008; Cusimano *et al.*, 2011; Nauheimer *et al.*, 2012). More recently duckweeds were proposed to be a separate family (*Lemnaceae*) with the subfamilies of Lemnoideae and Wolffioideae (Appenroth *et al.*, 2015; Les *et al.*, 2002; Sree *et al.*, 2016). Duckweeds comprise 37 species within 5 genera: *Spirodela* (2 species), *Landoltia* (1), *Lemna* (13), *Wolffiella* (10) and *Wolffia* (11) with *Spirodela* as the most ancestral and *Wolffia* as the most derived genus (Tippery *et al.*, 2015). Duckweed organisms have a minute, leaf-like neotenous structure called “frond”. All duckweeds are lacking a stem and the more derived genera *Wolffiella* and *Wolffia* possess even no true roots anymore. Although flowers are observed in several species (*Wolffia microscopica* (Khurana *et al.*, 1986), *Wolffia australiana* (Krajncič *et al.*, 1998), *Wolffia arrhiza* (Bernard *et al.*, 1990), *ect.*), duckweeds usually propagate via asexual reproduction by forming daughter fronds from meristematic pockets (primordia) at the proximal end of the mother frond (Cao *et al.*, 2015; Wang *et al.*, 2014; Bog *et al.*, 2013). In addition, the formation of turions (bud-like vegetative organs for perennation) - an alternative developmental path from primordia - is known to occur in 15 out of the 37 species. Turions allow duckweeds hibernation by sinking to the bottom of lakes or ponds due to high content of storage starch, thicker cell wall than that of frond and a lack of parenchyma. In spring, when the starch is consumed and the ice on the lakes is molten, turions emerge again on the water surface and new fronds germinate from the meristematic pocket of turions (Landolt, 1986; Appenroth and Nickel, 2010; Wang and Messing, 2015). Interestingly, duckweed fronds may vary from 1.5 cm to less than one millimeter in diameter and nearly 12-fold in genome size (from 160 Mbp to 1 881 Mbp). A successive reduction

of morphological structures from the ancestral genus *Spirodela* to the more derived genera *Lemna*, *Wolffiella* and *Wolffia* is accompanied by a stepwise reduction in frond size and a parallel increase in biodiversity (number of species), in genome size and genome size variability (Landolt, 1986; Wang *et al.*, 2011; Bog *et al.*, 2015) (Fig. 2 and 3). Chromosome number variation from 20 – 126 is reported (Urbanska, 1980; Geber, 1989). Epigenetic marks were studied by immunostaining in species of the five duckweed genera (Cao *et al.* 2015). Surprisingly, no distinct clusters of heterochromatin marks such as DNA and histone H3 methylation (5meC, H3K9me2, H3K27me1) were found in interphase nuclei, independent of the genome size of the tested species. The authors speculated that this observation could be linked with neoteny and fast growth, because cell nuclei of tissue culture or within *A. thaliana* seedlings younger than 4 days showed the same phenomenon, while nuclei of elder plants displayed pronounced regions with accumulation of these heterochromatic marks. Because the reasons for genome size differences and chromosome number variations among duckweeds are unknown and we do not know whether or not a correlation between genome size, progressive morphological reduction and frond diminution as well as cell and nucleus size exists in this family, duckweeds, are an interesting subject for genome and karyotype evolution studies.

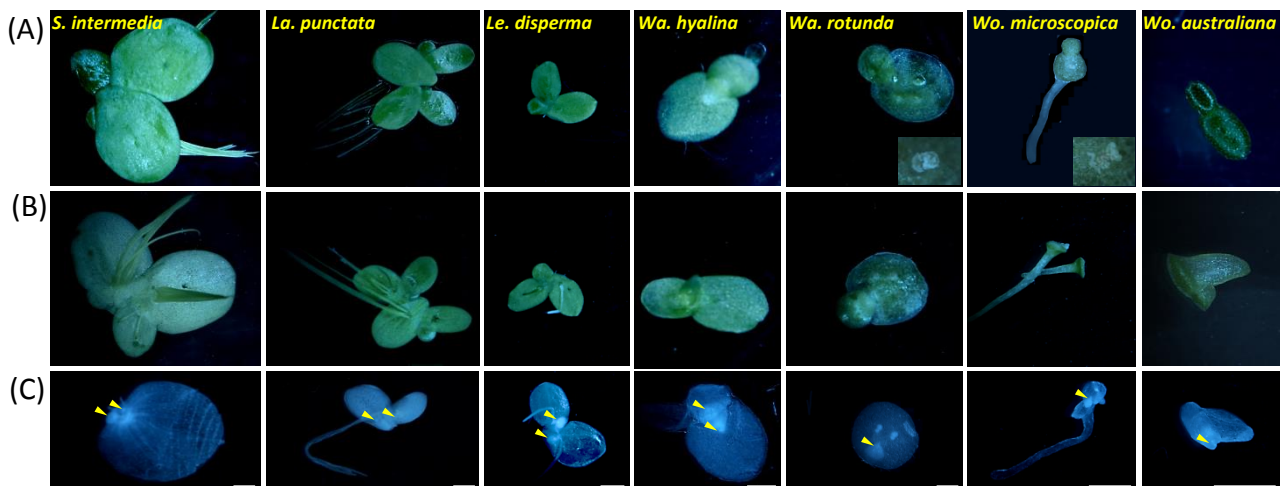


Figure 2. Duckweed morphology.

(A): Dorsal surface with flower (inserted); (B): ventral surface, (C): meristem pockets (yellow arrowheads) in fixed fronds. To avoid the confusing between *Landoltia* and *Lemna* as well as *Wolffiella* and *Wolffia* genera, we used a two letter code to abbreviate the names for these genera Scale bars: 1mm

1.2.2. What makes duckweeds becoming potential aquatic crops?

Duckweeds are worldwide distributed (except in the Arctic and Antarctica) and are the fastest growing angiosperms that yield up to 100 tons dry mass/hectare/year (Lam *et al.*, 2014; Ziegler *et al.*, 2015) with a high quality and quantity of protein. Their floating on the water surface makes harvesting easy. Therefore duckweed biomass was used as an important source for livestock feeding and even for human consumption (Rusoff *et al.*, 1980; Cheng and Stomp, 2009; Boonsaner and Hawker, 2015; Flores-Miranda *et al.*, 2015; Sharma *et al.*, 2016; Appenroth *et al.*, 2017). High starch content in some strains under particular growth conditions (McLaren and Smith, 1976; Sree *et al.*, 2015; Cui and Cheng, 2015; Fujita *et al.*, 2016) could be used to produce biofuels (Yadav *et al.*, 2017; Tao *et al.*, 2017). In addition, duckweeds are preferred aquatic plants for wastewater remediation due to their ability to absorb phosphate and nitrate and to accumulate heavy metals such as Cd, Cr, Zn, Sr, Co, Fe, Mn, Cu, Pb, Al and even Au (FAO, 1999; Teixeira *et al.*, 2014; Goswami *et al.*, 2014; Chaudhuri *et al.*, 2014; Tatar and Öbek, 2014; Rofkar *et al.*, 2014; Panfili *et al.*, 2017; Gatidou *et al.*, 2017; Basílico *et al.*, 2016). Moreover, some duckweed species (*Lemna gibba*, *Lemna minor*) can be transformed and used for production of recombinant proteins for pharmaceutical applications reviewed by (Stomp, 2005). Thus, duckweeds have the potential to become a new generation of sustainable crops which not compete with traditional crops for arable land. Therefore, duckweeds increasingly attract the attention of scientists of different fields. Their studies focus on genome sequencing and address many other issues such as turion formation, the ability to respond to adverse environmental conditions, the prerequisites for wastewater treatment, and for economic production of biofuel, feed for livestock, and human food. According to statistic data from PubMed: 92 studies on duckweeds were published between 1959 and 1999, while the number increased to 115 between 2000 and 2005, to 131 (2006 – 2010), to 200 (2011 – 2015) and to 117 (only from Jan, 2016 to March, 2018). This dramatic increase of publications on duckweeds from 2000 up to now proves the growing interest in these plants, and Sree called this period “blooming era of resurgence of duckweed research and applications” (Sree *et al.*, 2016).

1.2.3. Some landmarks of (mainly) genome research on duckweeds:

- 1986/87: *Lemnaceae* monographs (Landolt, 1986; 1987)
- 2001: Genetic transformation of *Lemna gibba* and *Lemna minor* (Yamamoto *et al.*, 2001)
- 2008: Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA (Cabrera *et al.*, 2008)
- 2011: Evolution of genome size in duckweeds (*Lemnaceae*) (Wang *et al.*, 2011)
- 2013: Genetic characterization and barcoding of taxa in the genus *Wolffia* Horkel ex Schleid. (*Lemnaceae*) as revealed by two plastidic markers and amplified fragment length polymorphism (AFLP) (Bog *et al.*, 2013)
- 2014: Insights into neotenus reduction, fast growth and aquatic lifestyle of *Spirodela polyrhiza* via genome sequence analysis (Wang *et al.*, 2014)
- 2015: Genetic characterization and barcoding of taxa in the genera *Landoltia* and *Spirodela* (*Lemnaceae*) by three plastidic markers and amplified fragment length polymorphism (AFLP) (Bog *et al.*, 2015)
- 2015: Chromatin organization in duckweed interphase nuclei in relation to the nuclear DNA content (Cao *et al.*, 2015)
- 2016: The map-based genome sequence of *Spirodela polyrhiza* aligned with its chromosomes as a reference for karyotype evolution (Cao *et al.*, 2016)
- 2017: Comprehensive definition of genome features in *Spirodela polyrhiza* by high-depth physical mapping and short-read DNA sequencing strategies (Michael *et al.*, 2017)

1.3. Whole genome sequencing, genome maps and chromosome numbers of duckweeds

1.3.1. Whole genome sequencing

A rather complete, high-quality genome sequence assembly is one pre-requisite for further research into molecular biology, particularly for non-model organisms of which genetic maps are not available and difficult to gain. DNA sequencing began in the 1970s with the Maxam-Gillbert chemical method, followed by the Sanger enzyme method. Next Generation Sequencing (NGS) systems introduced over the past decade allowed for the simultaneous analysis of thousands of gene sequences

rapidly with low cost and applicable to a wide variety of subjects. The analysis and assembly of genome sequences provides important genetic information for the subject under study, such as number of protein-coding genes, location of genes on chromosomes (linkage groups) and the evolutionary history of the genome (e.g. WGD events). However, validation of assembled sequences and generation of a complete genome sequence for large, complex and potentially polyploid genomes is still a challenge.

S. polyrhiza (clone 7498), was the first duckweed species chosen for whole genome sequencing due to its ancestral phylogenetic position, its economic potential as well as its small genome size (160 Mbp) indicating a low content of repetitive DNA (Wang *et al.*, 2014). After integration of sequences from Roche/454 and Sanger ABI-3730XL platforms, BAC and fosmid paired ends as well as 24 entire fosmids and DNA fingerprinting of the BAC library, the *S. polyrhiza* genome assembly yielded 32 pseudomolecules with at least 1 Mbp in length, comprising 90% of the estimated genome size. Several important information regarding neoteny and genome evolution in duckweeds could be extracted from these data:

- Two ancient whole-genome duplication, indicated by seven ancestral blocks of mostly quadruplicated homeologous genes, occurred approximately 95 million years ago (mya), i.e. earlier than the latest WGDs in *Arabidopsis* and rice;
- The predicted 19 623 protein-coding genes represent a significant reduction in comparison to gene numbers of *A. thaliana* (27 416), tomato (34 727), banana (36 542) and rice (39 049) with which *S. polyrhiza* shares 8 255 similar gene families. As reason for gene number reduction (for instance the loss of gene families for water transport and lignin biosynthesis) the authors considered neotenic organismic reduction and aquatic life style;
- A similar amount of full-length long terminal repeat (LTR)-retrotransposons as in *Arabidopsis*, but with distinctly older insertions in *S. polyrhiza* (4.6 versus 2.0 mya), indicating a reduced retrotransposition rate during recent evolution;
- Up to 32 loci of miRNA156 (including similar isoforms) that repress the transition to the adult phase in *S. polyrhiza*, while only 19 such loci were found in rice and 10 in *Arabidopsis*;

This first genome map of *S. polyrhiza*, provided useful information for future studies in evolution, development and economic applications of duckweeds and stimulated already further research. Together with the genomic map for another *S. polyrhiza* clone (9509) (Michael *et al.*, 2017) it led to an updated and significantly improved physical map for this species (see below and Hoang *et al.*, 2018).

Further whole genome sequencing projects for other duckweed species are on-going including *Lemna minor* (clone 5500) (Van Hoeck *et al.*, 2015); *Lemna minor* (clone 8627) and *Lemna gibba* (clone 7742a) (Cold Spring Harbor Laboratory); *Wolffia australiana* (clones 7733 and 8730) (J. Craig Venter Institute, USA) and *Landoltia punctata* (clone 7260) (Institute of Plant Molecular Biology, C. Budejovice, Czech Republic).

1.3.2. Genome maps

Besides genome maps that are based on assembly of overlapping sequence reads there are two other types of genome maps:

Genetic linkage maps: This type of maps is based on the frequency of linkage *versus* linkage disruption of markers in the progeny of parents heterozygous for these markers. They illustrate the arrangement of genes or other markers on a chromosome (=linkage group) and the relative distances in centiMorgan (cM) between them. Because linkage disruption via meiotic cross over is not equally distributed along chromosomes, the genetic distances between markers do not reflect their actual physical distance. The first genetic-linkage map was established by A. H. Sturtevant in 1913 by crossing experiments for the fruit fly *Drosophila melanogaster*- decades before scientists even knew that genes are made of DNA. The relative location of a series of genes were mapped on fly chromosomes, for review see (Lobo and Shaw, 2008).

Due to their mainly or exclusively vegetative propagation, genetic linkage maps are missing and difficult to obtain for duckweeds, as is the case for the two species of the genus *Spirodela*.

Physical maps: Such maps represent the true physical DNA-base-pair distances from one landmark to another. Since late 1980s, STSs (sequence-tagged sites) - unique DNA sequences of a few hundred base pairs, were used as landmarks to construct at least partial physical maps (Moore *et al.*, 2001; Greenberg and Istrail,

1995). Recently, different methods to establish physical maps were established. One option is cytogenetic mapping based on fluorescence in situ hybridization (FISH). FISH enables DNA sequence localization on chromosomes (and, on larger chromosomes even within distinct chromosomal regions) and provides reliable linkage information for contigs and scaffolds resulting from assembly of sequence reads. Consecutive rounds of multicolor FISH turned out to be a valuable independent tool for evaluating, extending and correcting sequence assemblies from NGS (Cao *et al.*, 2016). A special advantage of mapping by mcFISH is its ability to prove chromosomal linkage groups by overcoming large distance between chromosomal markers and its robustness against the presence of repetitive sequences (Chamala *et al.*, 2013; Lichter *et al.*, 1990; Cao *et al.*, 2016; Karafiatova *et al.*, 2013; Poursarebani *et al.*, 2014; Cheng *et al.*, 2002). Integration of the cytogenetic maps and sequence assemblies assists to resolve the chromosome-level genome assembly and to reveal new insights into genome architecture and genome evolution. In addition, DNA probes for specific classes of repetitive DNA elements and/or basic chromosome structures (e.g. centromere or telomere DNA repeats, ribosomal DNA) can be used to study the genome organization and karyotype differentiation by FISH. Genes located near the centromeres are often a challenge for mapping efforts because these areas usually contain a lot of repetitive sequences and lack detailed information from genetic mapping (due to very low crossing frequencies). Such genes can be mapped by FISH, as shown for chromosome 3H of barley (Aliyeva-Schnorr *et al.*, 2015). Comparative chromosome painting with pooled contiguous DNA probes from one reference species can be used to investigate chromosome homeology and rearrangements in related (not-yet-sequenced) species (Koumbaris and Bass, 2003; Lysak *et al.*, 2006; Mandakova and Lysak, 2008; Peters *et al.*, 2012; Mandakova *et al.*, 2015; Lusinska *et al.*, 2018). Comparative FISH with suitable unique probes can also resolve WGD in neo- and mesopolyploid species (Vu *et al.*, 2015; Geiser *et al.*, 2016) and synteny between related species (Ma *et al.*, 2010; Lee *et al.*, 2010; Lusinska *et al.*, 2018). FISH-based cytogenetic maps are very robust, but cannot resolve physical distance on the base pair level.

Another option are optical maps which order DNA fragments after digestion of genomic DNA with moderately cutting restriction enzymes according to their length and align them to the sequence information of restriction sites within the genome.

Optical maps rely on the stretching of DNA fragments stained with a fluorescent dye and still bear the risk of errors via mis-assembly.

Other types of physical mapping, based on chromosomal deletions (Serizawa *et al.*, 2001) or chromosome translocations (Macas *et al.*, 1993; Kuenzel *et al.*, 2000), require specific cytogenetic stocks, which are only seldom available.

1.3.2.1. The cytogenetic map of the Greater duckweed – *S. polyrhiza*

Applying consecutive mcFISH experiments, the genome assembly for the Greater duckweed *S. polyrhiza* (clone 7498) from Wang *et al.*, (2014) was validated and resulted in a cytogenetic map. In detail: (1) Three of the originally 32 pseudomolecules turned out to be chimeric ones; (2) 96 anchored BACs representative for the now 35 pseudomolecules were integrated into the 20 chromosome pairs of *S. polyrhiza*; (3) All chromosome pairs could be identified by a cocktail of 41 BACs in three colors (Cao *et al.*, 2016).

These results proved that mcFISH can be used as independent approach for validation and chromosomal integration of genome assembly. This first reference genome map of *S. polyrhiza* provided an important anchor point for further karyotype evolution studies in other duckweed species.

1.3.2.2. The optical map of the Greater duckweed – *S. polyrhiza*

An optical map for *S. polyrhiza* clone (9509) was established by combination of high-depth short read sequencing and high-throughput optical genome mapping technologies (Michael *et al.*, 2017). The BioNano Genomics Irys® System was applied to generate deep coverage physical maps. The most important results are:

- A strikingly low number of 45S rDNA repeats of only 81 copies, while *A. thaliana* with similar genome size contains 570 copies, and the budding yeast *Saccharomyces cerevisiae* with a genome size of just 12.2 Mbp has still 150 copies. This low copy number was also confirmed in four different clones of *S. polyrhiza* by the same authors applying three independent methods.
- The low number of protein-coding genes was further reduced by 1 116 genes compared to the number reported by Wang *et al.*, 2014, when Michael *et al.* (2017) considered the results of transcriptome sequencing after RT-PCR.

- 301 out of 24 344 orthologous gene clusters (resulting from comparison of predicted proteins of *Spirodela*, *Arabidopsis*, *Brachypodium*, oil palm, banana, sogum and rice) are specific to *Spirodela*.
- The DNA methylation level at CpG sites of only 9.4% was the lowest among the plants tested so far. For comparison, *A. thaliana* displayed 32.8%, *Setaria italica* 44.4%, *Brachypodium distachyon* 54.1%.
- Holocentric chromosomes were assumed because of the dispersed distribution of the 119 bp presumably centromeric repeat across all *S. polyrhiza* chromosomes.
- The highest soloLTR:intact retroelement ratio (8.52) and highly methylated (20%), ~4 million years old intact LTRs were recorded and compared to rice, banana and tomato. The large proportion of 'old' soloLTRs suggests remote genome shrinking via the deletion biased 'single strand annealing' DSB repair mechanism.
- In contrast to Wang *et al.*, 2014 but similar to the situation in the genomes of *A. thaliana* and of soybean, only five loci of miRNA156 were identified.

Furthermore, several discrepancies appeared between the *S. polyrhiza* cytogenetic map (for clone 7498) by Cao *et al.* (2016) and the optical map (for clone 9509) by Michael *et al.* (2017) regarding the chromosomal assignment of pseudomolecules, and, as a consequence, the chromosome enumeration. The reasons of these discrepancies could be (1) Mis-assembly of either of the genomes; (2) Too low DNA marker coverage in the cytogenetic study or (3) Clone-specific chromosome rearrangements.

To provide a high-confidence genome map as a reference for this species, these discrepancies had to be resolved.

1.3.3. The chromosome numbers of duckweeds

Table 1: Duckweed chromosome numbers from literature

Genus	Species	2n	Source	
<i>Spirodela</i>	<i>polyrhiza</i>	40	1*	
		40	2*	
		32, 40	3*	
		30, 40	13	
		30, 40, 50	11	
		40,8	12	
	<i>intermedia</i>	20, 30	11	
		36	12	
<i>Landoltia</i>	<i>punctata</i>	40, 43-44, 50	11	
		40, 50	13	
		46	12	
<i>Lemna</i>	<i>aequinoctialis</i>	20, 40, 50, 60, 80	11	
		42, 84	12	
		40, 60, 80	13	
		40, 50, 66, 72, 78, 84, 65-76	4*	
	<i>disperma</i>	40	11	
		44	12	
	<i>gibba</i>	64	1*	
		ca. 60	2*	
		40, 50, 70, 80	11	
		42, 43, 44, 84-86	12	
	<i>japonica</i>	40, 50	11	
		50	13	
		63	12	
	<i>minuscula</i>	36, 40	11	
		42	12	
	<i>minor</i>	40	1*	
		42	5*	
		40	6*	
		40	2*	
		50	7*	
		20, 30, 40, 42, 50	11	
		40, 42	13	
		42, 63, 126	12	
	<i>obscura</i>	40, 50	11	
		42	12	
	<i>perpusilla</i>	40	11	
		42	12	
	<i>turionifera</i>	40, 42, 50, 80	11	
42		12		
<i>Lemna</i>	<i>trisluca</i>	44	1*	
		Ca. 40	2*	
		20, 40, 60, 80	11	
		40, 60, 80	13	
		42, 44, 63-66	12	
	<i>valdiviana</i>	40	11	
		42	12	
	<i>Wolffiella</i>	<i>denticulata</i>	20, 40	11
		<i>gladiata</i>	42	8*
			40	13
			40	11
		<i>hyalina</i>	40	13
40			11	
<i>lingulata</i>		42	8*	
		20, 40, 50	13	
<i>neotropica</i>		20, 40, 50	11	
		40	11	
<i>oblonga</i>		42	8*	
		40, 70	11	
<i>welwitschii</i>	42	12		
	40	11		
<i>Wolffia</i>	<i>arrhiza</i>	ca. 50	1*	
		44-46	9*	
		50	2*	
		62	**	
		30, 40, 50, 60, 70, 80	11	
	42	12		
	<i>australiana</i>	20	13	
		20, 40	11	
	<i>angusta</i>	40	11, 13	
	<i>borealis</i>	20, 30, 40	11	
	<i>brasiliensis</i>	42	**	
		20, 40, 50, 60, 80	11	
	<i>columbiana</i>	Ca. 42	8*	
		40	13	
		30, 40, 50, 70	11	
	<i>globosa</i>	30, 40, 50, 60	11	
		60	13	
		46	**	
<i>microscopica</i>	70	10*		
	40, 80	11		

(1) Blackburn (1933); (2) Wcislo (1970); (3) Banerjee (1971); (4) Beppu&Takimoto (1981); (5) Brooks (1940); (6) Delay (1947); (7) Loeve (1978); (8) Daubs (1965); (9) Lawalree (1943); (10) Roy &Dutt (1967); (11) Urbanska (1980); (12) Geber (1989); (13) Wang *et al.*, (2011) *: mentioned in Geber (1989); **: Kwanyumen (personal communication) mentioned in Urbanska (1980)

Chromosome numbers of duckweeds were studied since more than 50 years. Numbers of $2n = 20$ to 126 have been reported. Even for the same species different chromosome numbers were counted (Urbanska, 1980). This could be due to counting errors, to intraspecific variation between geographically wide-spread clones, or to ploidy variations between populations. Chromosome numbers for duckweed species from different studies are summarized in Table 1. To validate the chromosome numbers for individual duckweed species and to elucidate the reason for the reported intraspecific variation in chromosome number, further studies are required.

1.4. Aims of the dissertation

This dissertation was directed to enlarge the cytological basis for studies of genome and karyotype structure and evolution of the five duckweed genera and to extend the scarce present knowledge in this field beyond the results gained so far for *S. polyrhiza*. The main tasks to be focused on were:

First, it was aimed to test whether the reported increase in genome size in the phylogenetic younger genera with smaller organisms and a stronger reduction of organismic complexity (neoteny) is correlated with the corresponding size of nuclei and cells, and thus with fewer cells per organism. For this purpose, clones of eleven species, representative for the five genera, were selected to measure genome size, cell and nucleus volume.

Second, it was aimed to determine the chromosome number and rDNA loci for these eleven species.

Third, it was aimed to resolve the discrepancies between the two previous genome maps of *S. polyrhiza* (Cao *et al.*, 2016; Michael *et al.*, 2017). Since genetic maps are hardly to obtain for the vegetative propagating species of the ancient genus *Spirodela*, an advanced mcFISH approach is the method of choice to provide a robust genome map for *S. polyrhiza*. To test whether the conflicting results of the previous maps were due to (1) Mis-assembly of either of the genomes, (2) Too low DNA marker coverage in the cytogenetic study or (3) Clone-specific chromosome rearrangements, a broader range of BACs from the regions in question should be applied to the two previously studied and five other clones of different geographic origin. The results should be counterchecked and confirmed by integration of a new Oxford Nanopore sequence assembly for the clone 9509 from Todd Michael and Eric Lam. The new high-confidence map should serve as a reference and a prerequisite for further studies to elucidate genome and karyotype evolution in duckweeds.

Fourth, it was aimed to elucidate the possible mode(s) of karyotype evolution between *S. polyrhiza* with $2n = 40$ chromosomes and *S. intermedia* with $2n = 36$ - the only two species of the most ancient duckweed genus. This should be done by consecutive rounds of cross-hybridization to *S. intermedia* chromosomes of BACs anchored to the 20 *S. polyrhiza* chromosomes. The expected results, as a first example to resolve the karyotype relationship between duckweed species, should (1)

Identify all *S. intermedia* chromosomes, (2) Determine their homeology to the 20 *S. polyrhiza* chromosomes and (3) Provide anchor points for assembling the *S. intermedia* genome.

Fifth, it was aimed to integrate a provisional assembly of PacBio reads of the *S. intermedia* genome of 37.5-fold coverage into the 18 chromosomes of *S. intermedia*. By reiterative comparison of *S. intermedia* contigs with the reference genome for *S. polyrhiza* and mcFISH control experiments, the karyotype as well as the genome assembly of *S. intermedia* should be improved.

Sixth and finally, it was aimed to find out to which degree the cross-FISH strategy is suitable to extend the cytogenetic studies to all duckweed genera to uncover their karyotype structure and the routes of karyotype and genome evolution within the entire family.

2. MATERIALS AND METHODS

2.1. Plant material and cultivation

Fronds of the studied species were collected from different geographic regions of the world and obtained from Dr. Klaus Appenroth, Friedrich-Schiller-Universität, Jena (Table 2). The plants were grown in liquid nutrient medium including KH_2PO_4 (60 μM), $\text{Ca}(\text{NO}_3)_2$ (1 μM), KNO_3 (8 mM), MgSO_4 (1 mM), H_3BO_3 (5 μM), MnCl_2 (13 μM), Na_2MoO_4 (0.4 μM), FeEDTA (25 μM) (Appenroth *et al.*, 1996) under 16 h white light of 100 $\mu\text{mol m}^{-2} \text{s}^{-1}$, at 24°C.

Table 2: List of duckweed species and their clones used in this study.

Clone ID	Genus	Species	Country of origin	Note
7498	<i>Spirodela</i>	<i>polyrhiza</i>	USA	* , ** , ***
7652	<i>Spirodela</i>	<i>polyrhiza</i>	Mexico	***
7657	<i>Spirodela</i>	<i>polyrhiza</i>	Mexico	***
9500	<i>Spirodela</i>	<i>polyrhiza</i>	Germany	*
9505	<i>Spirodela</i>	<i>polyrhiza</i>	Cuba	*
9507	<i>Spirodela</i>	<i>polyrhiza</i>	Russia	*
9509	<i>Spirodela</i>	<i>polyrhiza</i>	Germany	*
9510	<i>Spirodela</i>	<i>polyrhiza</i>	Mozambique	*
9511	<i>Spirodela</i>	<i>polyrhiza</i>	Russia	*
7747	<i>Spirodela</i>	<i>intermedia</i>	Peru	** , ***
8410	<i>Spirodela</i>	<i>intermedia</i>	Panama	**
7260	<i>Landoltia</i>	<i>punctata</i>	Australia	***
5562	<i>Landoltia</i>	<i>punctata</i>	Israel	***
5562_A4	<i>Landoltia</i>	<i>punctata</i>	Israel	****
8623	<i>Lemna</i>	<i>minor</i>	Denmark	***
7269	<i>Lemna</i>	<i>disperma</i>	Australia	***
6746	<i>Lemna</i>	<i>aequinoctialis</i>	USA	*** , ****
2018	<i>Lemna</i>	<i>aequinoctialis</i>	Japan	****
8640	<i>Wolffiella</i>	<i>hyalina</i>	Tanzania	***
9072	<i>Wolffiella</i>	<i>rotunda</i>	Zimbabwe	***
7540	<i>Wolffia</i>	<i>australiana</i>	New Zealand	***
2005	<i>Wolffia</i>	<i>microscopica</i>	India	***
8872	<i>Wolffia</i>	<i>arrhiza</i>	Hungary	***

(*) used for updating the *S. polyrhiza* genome map; (**) used in karyotype evolution studies between *S. polyrhiza* and *S. intermedia*; (***) used in cytological studies comparing the five duckweed genera; (****) used in polyploidy level studies

2.2. Genomic DNA isolation and metaphase preparation

Genomic DNA of the studied species was isolated using the DNA Miniprep Method. For each sample, 0.3 g fresh and healthy fronds were harvested and cleaned in distilled water, put into a 2 ml Eppendorf tube with two metal balls, frozen in liquid nitrogen and ground by a ball mill mixer (Retsch MM400). Then 900 µl 2xCTAB [2% CTAB, 200 mM Tris/HCl (pH 8.0), 20 mM EDTA, 1.4 M NaCl, 1% PVP, 0.28 M β-mecaptoethanol] were added. The solution was vortexed briefly, incubated at least 30 min at 65°C. Then, 800 µl cold phenol/chloroform/isoamylalcohol (15/24/1) were added and, after shaking by overhead-shaker for 14 min at 4°C, the solution was centrifuged for 15 min at 14 000 rpm (Centrifuge 5804R, Eppendorf). The supernatant was filled into a 1.6 ml microfuge tube, 5 µl RNase A solution (1 mg/ml) were added, and the tubes inverted and incubated for 15 min at 37°C. The DNA was precipitated at room temperature by adding 560 µl isopropanol and inverting the tube until the solution was well mixed. After centrifugation for 10 min at 14 000 rpm at 4°C to pellet DNA, the supernatant was discarded and 1 ml wash solution I [76% ethanol, 200 mM NaAc] was added to the pellet and incubated for 15 min, before replacing by 1ml wash solution II [76% ethanol, 10 mM NH₄Ac] and incubation for only 5 min. Then wash solution II was discarded and the pellet was dried at room temperature or in a Speed Vac and dissolved in TE-buffer [10 mM Tris/HCl (pH 8.0), 1 mM EDTA]. Concentration and quality of the DNA were measured by a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and by 1% (w/v) agarose-gel electrophoresis.

Duckweed chromosome spreads for FISH were prepared according to (Cao *et al.*, 2016) with some modifications. In brief, healthy fronds were treated in 2 mM 8-hydroxyquinoline at 37°C and then fixed in fresh 3:1 absolute ethanol: acetic acid for at least 24 h. The samples were washed twice in 10 mM Na-citrate buffer, pH 4.6, for 10 min each, before and after softening in 2 mL pectinase/cellulase enzyme mixture, prior to maceration and squashing in 60% acetic acid. After freezing on dry ice or liquid nitrogen, slides were treated with pepsin, post-fixed in 4% formaldehyde in 2x SSC [300 mM Na-citrate, 30 mM NaCl, pH 7.0] for 10 min, rinsed twice in 2x SSC, 5 min each, dehydrated in an ethanol series (70, 90 and 96%, 2 min each) and air-dried (Table 3).

Table 3: Procedures for preparation of duckweed chromosomes.

Species (*)	Tissue	Metaphase arrest (**)	Cell wall digestion (***)		Protein digestion (****)	Slide freezing
			Enzyme Concentration	Time		
<i>S. polyrhiza</i>	Meristem	3.5 h	1.0 %	60 min	7 min	Dry ice (30 min or more)
<i>S. intermedia</i>	Meristem	3.5 h	1.0 %	60 min	7 min	
<i>La. punctata</i>	Meristem	2.5 h	0.5 %	30 min	5 min	
<i>Le. minor</i>	Meristem	1.5 h	0.5 %	15 min	5 min	Liquid nitrogen (5 min)
<i>Le. disperma</i>	Meristem	1.5 h	0.4 %	10 min	5 min	
<i>Le. aequinoctialis</i>	Meristem	1.5 h	0.4 %	8 min	3 min	
<i>Wa. rotunda</i>	Meristem	2.0 h	0.4 %	10 min	5 min	
<i>Wa. hyalina</i>	Meristem	2.0 h	0.4 %	8 min	5 min	
<i>Wo. microscopica</i>	Fronde	2.0 h	0.4 %	8 min	3 min	
<i>Wo. australiana</i>	Fronde	2.0 h	0.4 %	10 min	3 min	
<i>Wo. arrhiza</i>	Fronde	2.0 h	0.4 %	12 min	3 min	

(*)To avoid the confusing between *Landoltia* and *Lemna* as well as *Wolffiella* and *Wolffia* genera, we used a two letter code to abbreviate the names for these genera; (**) 2 mM 8-hydroxyquinoline at 37°C; (***) Cellulase and pectinase mixture in Na-citrate buffer, pH 4.6 at 37°C; (****) 50 µg/ml pepsin in 0.01N HCl at 37°C

2.3. Genome size measurement

Genome size measurements were performed according to Dolezel et al. (2007) using a CyFlow Space flow cytometer (Sysmex/Partec). For nuclei isolation and staining, the DNA staining kit 'CyStain PI Absolute P' was used. As internal reference standards either *Raphanus sativus* 'Vorán' (IPK gene bank accession number RA 34; 2C = 1.11 pg - for *S. polyrhiza*, *S. intermedia*, tetraploid *La. punctata*, *Le. minor*, *Wa. hyalina*, *Wo. australiana*, *Wo. microscopica*), *Glycine max* (L.) Merr. convar. max var. max, Cina 5202 (IPK gene bank accession number SOJA 32; 2C = 2.21 pg - for *La. punctata*, *Wa. rotunda*, *Le. aequinoctialis*, *Le. disperma*) or *Lycopersicon esculentum* Mill. convar. *infiniens* Lehm. var. *flammatum* Lehm., Stupicke Rane (IPK gene bank accession number LYC 418 ; 2C = 1.96 pg - for *Wo. arrhiza*) were used. The absolute DNA contents (pg/2C) were calculated based on the values of the G1 peak means and the corresponding genome sizes (Mbp/1C) according to (Dolezel et al., 2003). In total, for each species at least 6 independent measurements on two different days were performed.

2.4. Epidermis preparation, microscopic cell and nuclear volume measurements, and statistics

Due to the small frond size, a single epidermis layer is difficult to obtain especially for species of the genus *Wolffia* (frond diameter ~1mm). Therefore, we modified the epidermis preparation methods described (Weyers and Travis, 1981; Ibata et al., 2013; Falter et al., 2015), by using domestic adhesive tape. Because stomata are located on the upper surface in floating plants (Shtein et al., 2017; Landolt, 1986), duckweed fronds were placed with their upper side on the domestic adhesive tape. Other parts of the fronds were carefully removed with a razor blade until only the transparent layer of epidermis stuck on the tape. Ten μl of DAPI (2 $\mu\text{g}/\text{ml}$) in Vectashield were dropped on slides before the adhesive tape with the epidermis layer was placed on the slides and covered by a coverslip. Freshly prepared slides were used immediately to avoid the disintegration of the nuclei before imaging. Differential interference contrast (DIC) and fluorescence (excitation of DAPI with a 405 nm laser) image stacks were acquired using a Super-resolution Fluorescence Microscope Elyra PS.1 and the software ZEN (Carl Zeiss GmbH). The DIC image stacks were used to measure the x-y area and the z dimension of the guard cells via the ZEN software. Accordingly, the fluorescence stacks were used to measure the nuclei dimensions (Fig. 6). These dimensions were applied to calculate the guard cell and nuclear volumes by the following formulae:

$$\text{Cell Volume} = A_{\text{cell}} * z$$

$$\text{Nuclear volume} = 2/3 * A_{\text{nucleus}} * z$$

It means, the guard cells are considered as stacks with the base area A and the height z, while the nuclei are considered as ellipsoids.

The correlations and regression diagrams were calculated with the program SigmaPlot 12 (Systat Software, Inc.). At least 20 sister guard cells (10 stomata) with the corresponding nuclei were chosen for measurements per species.

2.5. Probe preparation

2.5.1. 5S/18S/ 26S rDNA and telomere probes

Using primer pairs designed for 18S and 26S rDNA (Tippery *et al.*, 2015; Shoup and Lewis, 2003; Kuzoff *et al.*, 1998) and 5S rDNA (Gottlob-McHugh *et al.*, 1990) the corresponding probes were amplified from genomic DNA of five duckweed species (Table 4).

Table 4: List of primers used to amplify rDNA regions.

Primer	Primer sequences (5' – 3')	Product	Literature source	DNA template
18S–SSU1(F)	TGGTTGATCCTGCCAGTAG	18S-rDNA	(Shoup and Lewis, 2003)	<i>S. polyrhiza</i> <i>La. punctata</i> <i>Le. minor</i> <i>Wa. hyalina</i> <i>Wo. arrhiza</i>
18S–1243R	AGAGCTCTCAATCTGTCA			
26S–0091F	TAGTAACGGCGAGCGAACC	26S-rDNA	(Tippery <i>et al.</i> , 2015)	
26S–1229rev	ACTTCCATGACCACCGTCCT		(Kuzoff <i>et al.</i> , 1998)	
UP46	GTGCGATCATACCAGCACTAATGCACCGG	5S-rDNA	(Gottlob-McHugh <i>et al.</i> , 1990)	
UP47	GAGGTGCAACACGAGGACTTCCCAGGAGG			

Forward primers are indicated by 'F' and reverse primers by 'R' or 'rev'.

Telomere-specific probes were generated by PCR using tetramers of the Arabidopsis-type telomere repeats without template DNA according to (Ijdo *et al.*, 1991).

The probes were labeled with Cy3-dUTP (GE Healthcare Life Science), Alexa Fluor 488-5-dUTP, Texas Red-12-dUTP, biotin-dUTP or digoxigenin-dUTP (Life Technologies) by nick-translation (with 1 µg telomere, 18S and 26S rDNA PCR product in 50 µL reaction mixture) or by PCR-labeling (with 100 ng PCR product of 5S rDNA in 25 µL reaction mixture), and ethanol precipitated (Mandakova and Lysak, 2008). Probe pellets from 10 µL nick translation or 10 µL PCR-labeling product were dissolved in 100 µL hybridization buffer [50% (v/v) formamide, 20% (w/v) dextran sulfate in 2× SSC, pH 7] at 37°C for at least 1 hour. The ready-to-use FISH probes were stored at -20°C.

2.5.2. Bacterial artificial chromosome DNA probes

BAC clones from a x10 HindIII BAC library of *S. polyrhiza* 7498 were selected based on BAC end sequences and whole genome sequences of *S. polyrhiza* (Cao et al., 2016; Michael et al., 2017). Beside the 96 BACs which were selected and used to establish the cytogenetic map of *S. polyrhiza* by Cao et al. (2016), additional BACs used to generate the updated genome reference map of *S. polyrhiza* and for studies of karyotype evolution between *S. polyrhiza* and *S. intermedia* were selected from the BAC library according to their presumed position within the genomic region of interest.

Bacteria harboring BACs were incubated for 16 h at 37°C under shaking (200 rpm) in 75 ml LB medium with 12.5 µg/ml chloramphenicol. BAC DNA preparation was performed using the kit NucleoBond® PC100 (Macherey-Nagel GmbH & Co. KG, Dueren, Germany) with some modifications. After harvesting by centrifugation (4 000 rpm for 30 min), bacteria pellets were resuspended in 1.5 ml resuspension buffer (S1 + RNase A), followed by adding 1.5 ml lysis buffer (S2) and 1.5 ml neutralization buffer (S3). The bacterial lysate was filtered through NucleoBond® folded filters wetted with 750 µl buffer N2, and the clear lysate was collected. Afterwards, the BAC DNA of the cleared lysate was precipitated in iso-propanol (600 µl cleared lysate: 1500 µl iso-propanol) and centrifuged at 14 000 rpm, 30 min at 4°C to collect the DNA pellet. At room-temperature 70% ethanol was added to the pellet and centrifuged. The supernatant was discarded, the pellet was dried at room temperature and dissolved in sterile deionized H₂O. DNA quantification was done by absorbance measurements in a NanoDrop 1000 Spectrophotometer (Thermo Scientific, Wilmington, DE, USA). The total DNA of each BAC was sonicated in a Bioruptor (Diagenode) at a low level of ultrasound for 15 min before labeling. The BAC probes were labeled by nick-translation. For 50 µl of nick-translation volume, about 2 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 and 2 mM d(AGC)TP mixture were added into a 0.5 ml tube. For labeling, 2 µl of 1 mM Cy3, biotin or digoxigenine-dUTP or 0.8 µl of 1 mM TexasRed or Alexa 488-dUTP was added. The dUTPs were synthesized by custom labeling reaction according to (Henegariu et al., 2000). 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 120 - 150 min until the size of fragments reached 200~500 bp, controlled by 1% (w/v) agarose-gel electrophoresis. The DNA polymerase was inactivated by incubation at 65°C for 10 min. The labeled probe was then precipitated, as done for telomere probes, and stored at -20°C.

2.6. Fluorescence in situ hybridization

Probes were pre-denatured at 95°C for 5 min and chilled on ice for 10 min before adding 10 µL probe per slide (up to 3 different labeled probes simultaneously). Mitotic chromosome preparations were denatured together with the probes on a heating plate at 80°C for 3 min and then incubated in a moist chamber at 37°C for at least 16 h. Post-hybridization washing and signal detection were carried out according to Lysak et al. (2006). For subsequent rounds of FISH experiments, the hybridized probes were stripped (Shibata *et al.*, 2009; Heslop-Harrison *et al.*, 1992). In brief, slides were placed on a heating plate at 38°C for 10 min, coverslips were then removed carefully with forceps. Slides were washed in 0.1x SSC at room temperature 2x 5 min each, before washing under shaking condition at 42°C with the following solutions: 0.1x SSC, probe stripping solution [0.05% (v/v) Tween-20, 50% (v/v) formamide in 0.1x SSC] and 4T [0.05% (v/v) Tween-20 in 4x SSC] for 30 min each. After repeating the fixation in 4% formaldehyde, dehydration in an ethanol series and air-drying, the slides were ready for the next FISH experiment.

Fluorescence microscopy for signal detection followed Cao et al. (2016). The images were processed (brightness and contrast adjustment only), pseudo-colored and merged using Adobe Photoshop software ver.12x32 (Adobe Systems).

To analyze the ultrastructure and spatial arrangement of signals and chromatin at a lateral resolution of ~120 nm (super-resolution, achieved with a 488 nm laser), 3D structured illumination microscopy (3D-SIM) was applied using a Plan-Apochromat 63x/1.4 oil objective of an Elyra PS.1 microscope system and the software ZENblack (Carl Zeiss GmbH). Image stacks were captured separately for each fluorochrome using the 561, 488, and 405 nm laser lines for excitation and appropriate emission filters (Weisshart et al., 2016). Maximum intensity projections of whole cells were calculated via the ZEN software. Zoom in sections were presented as single slices to indicate the subnuclear chromatin structures at the super-resolution level.

2.7. *S. intermedia* whole genome sequencing and assembly

2.7.1. Plant material and DNA extraction

Genomic DNA was extracted from whole fronds of *S. intermedia* (clone 7747) by CTAB method before RNase treatment overnight at 37°C. Concentration and quality of DNA were measured by a Nanodrop spectrophotometer (Thermo Scientific, Wilmington, DE, USA) and by 1% (w/v) agarose-gel electrophoresis before sending the sample to the GATC company for sequencing.

2.7.2. Genome sequencing and assembly

After shearing of genomic DNA, a size-selected 20 kb library was sequenced on the Pacific Biosciences RS II platform (GATC Biotech, Konstanz, Germany) combining the P6-C4 polymerase-chemistry and 240 min of movie duration.

Two rounds of sequencing resulted in 149 Gb of raw read data. After an initial filtering for potential bacterial contamination and minimum read length (500 nt), a total of 1 305 064 reads were assembled using the Canu pipeline v. 1.5 (Koren *et al.*, 2017) consisting of the following steps:

(1) Trimming and error correction: Reads were corrected and trimmed by comparing overlaps. A minimum length of 500 nt and a maximum error rate of 10.5% was chosen for extending a contig. Only reads consisting of more than 1000 nt in length were considered in this step. Afterwards, the corrected reads were trimmed to improve overall read quality by using overlap information to detect high confidence regions. Contigs of insufficient read coverage and/or containing 'noisy' sequence were categorized as 'unsupported regions' and divided at weak sequence positions into subcontigs with higher support.

(2) Contig construction and building of the sequence assembly: By finding overlaps, contigs were constructed. Afterwards, a consensus sequence was constructed by removing the remaining sequencing errors to raise the overall assembly quality.

2.7.3. Scaffolding and gap filling

In a first round of scaffolding, the two genomes of the sister species *S. polyrhiza* (from clones 9505 and 7498) (Cao *et al.*, 2016; Michael *et al.*, 2017) were used as references for Mauve Genome Aligner v20150522 (Darling *et al.*, 2004) to order contigs. Scaffolding was performed by SSPACE-Longread v.1-1 (Boetzer and Pirovano, 2014). The resulting scaffold assembly was used for the super-scaffolding

approach. For this aim, contigs were assigned to 18 putative pseudomolecules (corresponding to the 18 *S. intermedia* chromosomes) using the information of cross-FISH of 93 *S. polyrhiza* BACs on the chromosomes of *S. intermedia* strain 8410 (Hoang and Schubert, 2017). New cytogenetic probes using BACs from the genomic regions of interest were designed for FISH experiments to approve localization of the contigs within the pseudomolecules and to resolve mis-assemblies.

The quality of the *S. intermedia* genome assembly was assessed by the BUSCO program (Simao *et al.*, 2015; Waterhouse *et al.*, 2017) with an Embryophyta dataset.

2.7.4. Gene prediction

Gene finding was carried out using Gene Model Mapper (GeMoMa) - a homology-based gene prediction program (Keilwagen *et al.*, 2016). Gene models were predicted by combining the predictions based on the genome data of three different reference organism (*S. polyrhiza* 7498 v3.1 (Cao *et al.*, 2016), *Lemna minor* 5500 (Van Hoeck *et al.*, 2015), *Oryza sativa* IRGSP v1.0.38 (GenBank assembly accession: GCA_001433935.1)).

2.7.5. Repeat identification

Because the program for repeat identification via clustering analysis such as RepeatExplorer (Novak *et al.*, 2013; Novak *et al.*, 2010), TAREAN- Tandem REpeat ANalyzer (Novak *et al.*, 2017) cannot use the long PacBio reads, a new run of Illumina sequencing of *S. intermedia* genomic DNA was prepared and is still ongoing. DNA isolation was done by Dr. Hieu Cao and genome assembly by Dr. Anne Fiebig. The reiterative validation of genome assembly by FISH, using BACs selected on the basis of their end sequences, was done by me.

3. RESULTS AND DISCUSSION

3.1. Morphology variation and correlation between genome size and cell parameters in duckweeds

Observations from eleven selected species which represent the five duckweed genera showed a negative correlation between genome size and size and complexity of fronds, as well as some variation in cell morphology. As described in Landolt's monographs (Landolt, 1986; 1987), the two species of the ancestral genus *Spirodela* have the lowest genome size with the largest fronds and a more complex frond structure with several roots, while the more derived genera display larger genomes (and genome size variation), smaller and simpler fronds with less roots (genus *Lemna*), no roots (*Wa. hyalina*, *Wa. rotunda*, *Wo. australiana*, *Wo. arrhiza*) or only a pseudoroot (*Wo. microscopica*) (Fig. 2 and 3B). The morphology of fronds varies from thin, leaf-like with orbicular (*Spirodela*), obovate (*Lemna*), tongue-shaped or sabre-shaped (*Wolffiella* species), to thick, spheric, cylindric or boat-shaped ones (*Wolffia* species). Frond sizes differ in length, width and depth between duckweed species. Guard cells are round-shaped in *Spirodela* and *Lemna* species, or elliptic as in *Landoltia*, *Wolffiella* and *Wolffia* species. Epidermis cell walls are rather straight in *Wolffiella* and *Wolffia* species, bent in *Spirodela* and undulated in *Landoltia* and *Lemna* species (Fig. 3C and 4A).

The present genome size measurements yielded up to 26% larger values than those of Wang *et al.*, (2011), even for the same clones. The differences might be due to (1) Different internal reference standards, (2) An unusually low assumption for the reference genome size of *A. thaliana* by Wang *et al.* (2011) (147 Mbp instead of 157 Mbp as measured by Bennett *et al.* (2003), and (3) Different flow cytometry equipment used. For instance, the highest difference 26% was observed for *Wa. hyalina* (8640), followed by 17% for *Wo. arrhiza* (8872), and 9% for *La. punctata* (7260) and 8% for *Le. minor* (8623), while for *S. polyrhiza* with the smallest genome, the values were similar. Because different clones were measured in *Wo. australiana* (7540 in this study and 8730 in Wang *et al.* (2011)), data are not directly comparable. For *S. intermedia* (8410), *Le. disperma* (7269), *Le. aequinoctialis* (2018), *Wa. rotunda* (9072) and *Wo. microscopica* (2005) (Fig. 4C) the present measurements are the first ones. These data showed that the nuclear DNA content varies ~14 fold between duckweed species (from 160 Mbp in *S. polyrhiza* to 2203 Mbp in *Wo. arrhiza*).

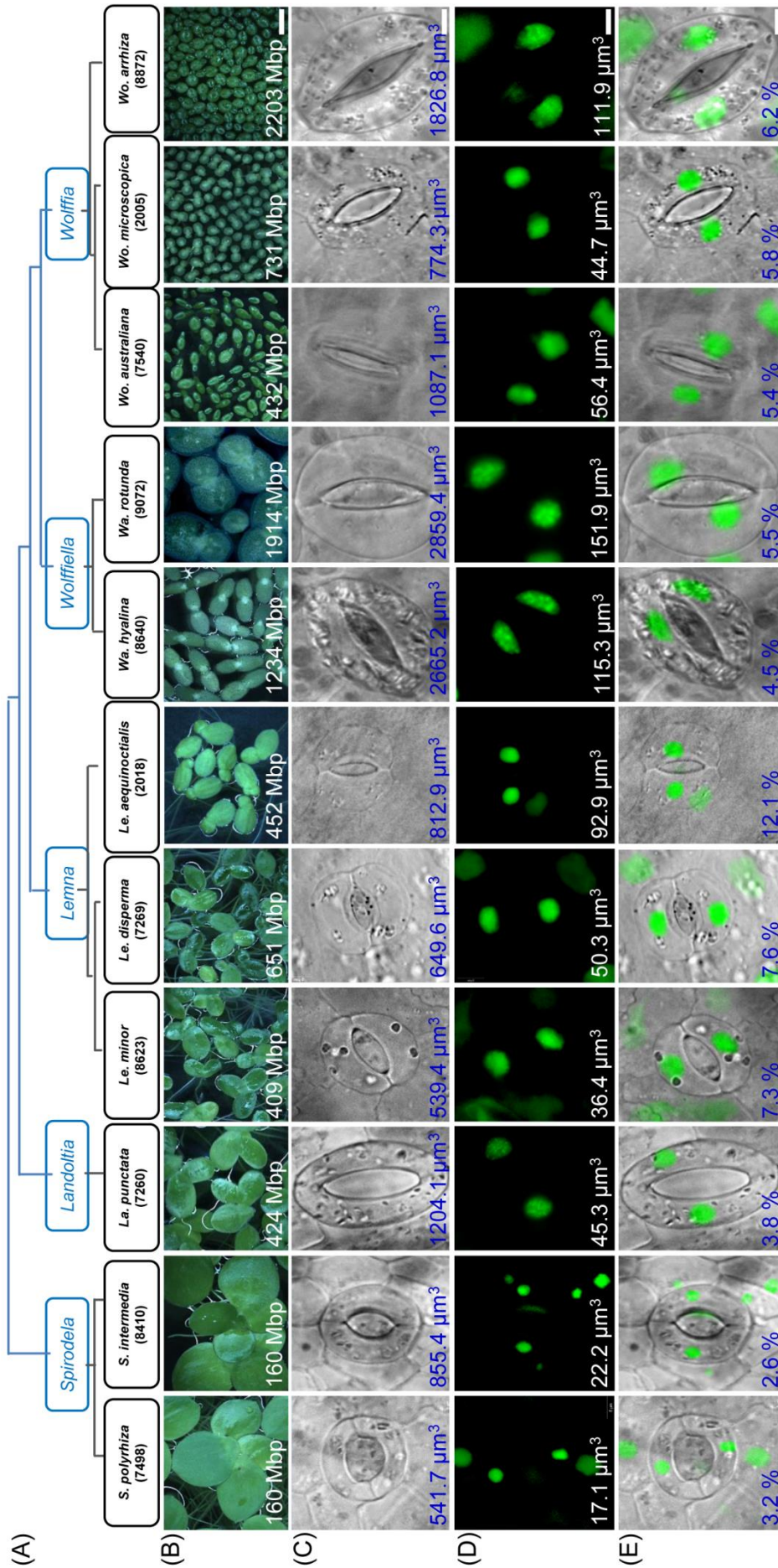


Figure 3: Phylogenetic relationship, frond, stomata and nuclei morphology of duckweed species.

(A) Phylogenetical position. (B,C) differences in size and morphology of fronds and stomata. (D,E) nuclei shape and distribution within the guard cells. Numbers indicate genome size (B), average cell (C) and nuclear volumes (D), and percentage of nuclear to cell volume (E). Scale bars = 200 µm (B) and 5 µm (C, D, E)

Previously, epidermis cells and endosperm cells were used to investigate possible correlations between genome size and cell parameters (Jovtchev *et al.*, 2006; Price *et al.*, 1973; Kladnik, 2015). Because of the highly variable and irregular shape of pavement cells in duckweeds (Fig. 4A), we selected guard cells with a more homogenous morphology instead of pavement cells for cell and nuclear volume measurements and calculation (Fig. 6A). In addition, the permanently open status of stomata in floating aquatic plants (Shtein *et al.*, 2017; Landolt, 1986) yields a rather homogenous cell shape, more suitable for precise volume measurement (Meckel *et al.*, 2007).

The measurements (n = 252) revealed a highly significant (p < 0.001) correlation between genome size and cell volume (r = 0.748), between genome size and nuclear volume (r = 0.768), as well as between nuclear volume and cell volume (r = 0.774) (Fig. 6B). In general, the correlation between genome size and cell and nuclear volume was positive for the eleven tested duckweed species. The higher the genome size, the bigger were the cell and nuclear volume. For instance, average cell and nuclear volume are 541.7 μm^3 and 17.1 μm^3 for *S. polyrhiza* (160 Mbp). These values increase to 649.6 μm^3 and 50.3 μm^3 in *Le. disperma* (651 Mbp); and to 1826.8 μm^3 and 111.9 μm^3 in *Wo. arrhiza* (2203 Mbp) (Fig. 3B,C). However, the relative correlation between genome size (Mbp), cell volume, nuclear volume and percentage of nuclear to cell volume can also differ within a genus. For instance, *Wo. australiana* has a smaller genome size (432 Mbp) but a larger cell volume (1087 μm^3) and nuclear volume (56.4 μm^3) than measured for *Wo. microscopica* (731 Mbp, 774.3 μm^3 and 44.7 μm^3) (Fig. 3 and Table 5).

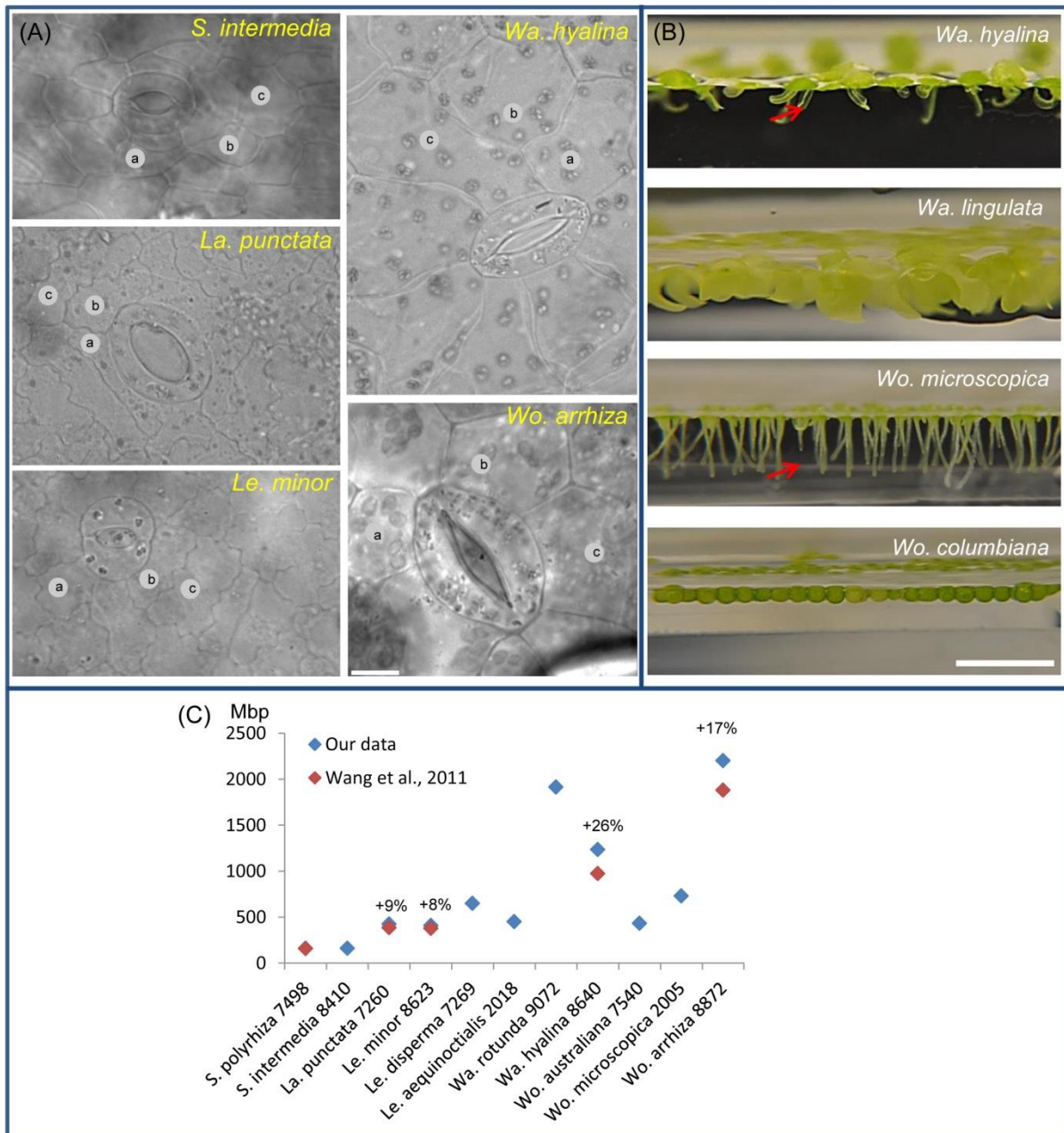


Figure 4. Variation in cell morphology (A), floating-style (B) and genome size (C) in duckweed.

(A) Epidermis cell walls are bent in *S. intermedia*, undulated in *La. punctata*, *Le. minor* and rather straight in *Wa. hyalina* and *Wo. arrhiza*. Stomata are round-shaped in *S. intermedia* and *Le. minor*, or elliptic as in *La. punctata*, *Wa. hyalina* and *Wo. arrhiza*. Varying epidermis cell sizes (a, b, c) within and between different duckweed species. (B) *Wa. hyalina*: Free-floating, two ovate fronds cohere together. The bent vertical appendage (arrow) is formed from the lower wall of a pouch. *Wa. lingulata*: Two tongue-shaped fronds cohere together with frond ends curved downward bringing most of the surface under water. *Wo. microscopica*: Free-floating, dorsoventral fronds with irregular polygonal flat dorsal surface and a ventral projection, the pseudo-root (arrow). *Wo. columbiana*: Nearly spherical fronds with most of their surface submerged. Stomata are present in the free-floating (*Wa. hyalina*, *Wo. microscopica*) and almost absent in the submerged (*Wa. lingulata*, *Wo. columbiana*) species. (C) Numbers indicate the deviation of genome size in% (this data relative to that of Wang *et al.* 2011) in the same duckweed clone. Scale bars = 10 μ m (A), 5mm (B). (Images Dr. Veit Schubert, IPK)

Table 5: Cytological characterization of eleven duckweeds species.

Genus	Spirodela		Landoltia	Lemna			Wolffiella		Wolffia		
Species	<i>polyrhiza</i>	<i>intermedia</i>	<i>punctata</i>	<i>minor</i>	<i>disperma</i>	<i>aequinoctialis</i>	<i>rotunda</i>	<i>hyalina</i>	<i>australiana</i>	<i>microscopica</i>	<i>arrhiza</i>
Clone ID	7498	8410	7260	8623	7269	2018	9072	8640	7540	2005	8872
Origin	USA	Panama	Australia	Denmark	Australia	Japan	Zimbabwe	Tanzania	New Zealand	India	Hungary
DNA content (pg/2C)	0.325 ± 0.006	0.327 ± 0.006	0.866 ± 0.012	0.836 ± 0.003	1.331 ± 0.046	0.925 ± 0.003	3.915 ± 0.012	2.523 ± 0.012	0.884 ± 0.012	1.496 ± 0.003	4.505 ± 0.125
Genome size (Mbp/1C)	160 ± 2	160 ± 3	424 ± 6	409 ± 2	651 ± 3	452 ± 2	1914 ± 6	1234 ± 6	432 ± 6	731 ± 1	2203 ± 61
2n =	40	36	46	42	44	42	82	40	40	40	60
No. 5S rDNA loci	2	2	2	1	1	1	3	2	2	1	3
No. 45S rDNA loci	1	1	1	1	1	1	2	1	1	1	2
Cell volume (µm ³)	541.7 ± 91.3	855.4 ± 79.1	1204.1 ± 141.3	539.4 ± 130.3	649.6 ± 178.8	812.9 ± 275.8	2859.4 ± 494.5	2665.2 ± 517.6	1087.1 ± 307.9	774.3 ± 134.3	1826.8 ± 216.1
Nuclear volume (µm ³)	17.1 ± 4.9	22.2 ± 4.5	45.3 ± 11.8	36.4 ± 12.8	50.3 ± 23.3	92.9 ± 21.9	151.9 ± 46.2	115.3 ± 19.5	56.4 ± 19.5	44.7 ± 18.8	111.9 ± 23.3
% Nuclear to cell volume	3.2 ± 1.0	2.6 ± 0.5	3.8 ± 1.2	7.3 ± 3.6	7.6 ± 2.1	12.1 ± 2.5	5.5 ± 2.1	4.5 ± 1.3	5.4 ± 1.4	5.8 ± 2.3	6.2 ± 1.4

Error: Standard deviation

Additionally, we found unexpected features in some duckweed species:

(1) *Le. aequinoctialis* (2018) revealed a considerable variation in guard cell size and shape (Fig. 5A). In the younger part of frond, guard cells are round while in the older part they are elongated and larger. Besides that, cell and nuclei volume are larger than that of *Le. disperma* possessing a larger genome. Therefore, we investigated another *Le. aequinoctialis* clone to see whether the variable guard cell volume is specific for this species. Interestingly, *Le. aequinoctialis* (clone 6746) showed variation in guard cell size and a nearly doubled genome size (900 Mbp) and chromosome number ($2n \sim 80$ in compared to $2n = 42$ of clone 2018) (Fig. 5B) and correspondingly larger cell and nuclear volumes ($1313 \mu\text{m}^3$ and $238 \mu\text{m}^3$, respectively). Thus, the two tested *Le. aequinoctialis* clones showed variation not only in guard cell shape, cell volume and nucleus volume, but surprisingly also regarding the genome size and chromosome number most likely due to WGD of clone 6746.

(2) Both tested *Wolffiella* species, *Wa. hyalina* (1234 Mbp, $2665.2 \mu\text{m}^3$ and $115.3 \mu\text{m}^3$) and *Wa. rotunda* (1914 Mbp, $2859.4 \mu\text{m}^3$ and $151.9 \mu\text{m}^3$), showed a larger cell and nucleus volume of guard cells than *Wo. arrhiza* with a larger genome (2203 Mbp, $1826.8 \mu\text{m}^3$ and $112 \mu\text{m}^3$). Therefore, we wanted to test other *Wolffiella* species to see whether very large cell volume is specific for this genus. Interestingly, only one or two stomata per frond were present in the *Wa. lingulata* clone 7725. The same was true for *Wo. columbiana* clone 9356. Differences in floating style of *Wo.*

columbiana with spherical fronds, having most of the surface submerged, and *Wa. lingulata* also with a frond shape which keeps most of the frond below the water surface (Landolt, 1986) (Fig. 4B) could be the reason for the almost complete absence of stomata in these species. Thus, so far it remains unclear whether or not a large guard cell size is a typical feature of the genus *Wolffiella*.

(3) *Wa. hyalina* and *Wo. australiana* displayed an unusual distribution of nuclei between sister guard cells. We found in 26% of *Wa. hyalina* and in 8 % of *Wo. australiana* guard cells two nuclei located in one sister cell and none in the other (Fig. 5C, D, F, C, E). In some cases (6.8% of *Wo. australiana* guard cells) it was even possible to find transient stages, suggesting that nuclei may post-mitotically migrate into the sister cell (Fig. 5G). This observation resembles cytomixis, a so far unexplained phenomenon which occurs during microsporogenesis in several higher plants (for review see Mursalimov and Deineko, 2018)).

These findings, in particular the large variation of guard cell and genome size in *Le. aequinoctialis*, and the abnormal nuclei distribution between sister guard cells are biological features of some duckweeds that deserve further studies.

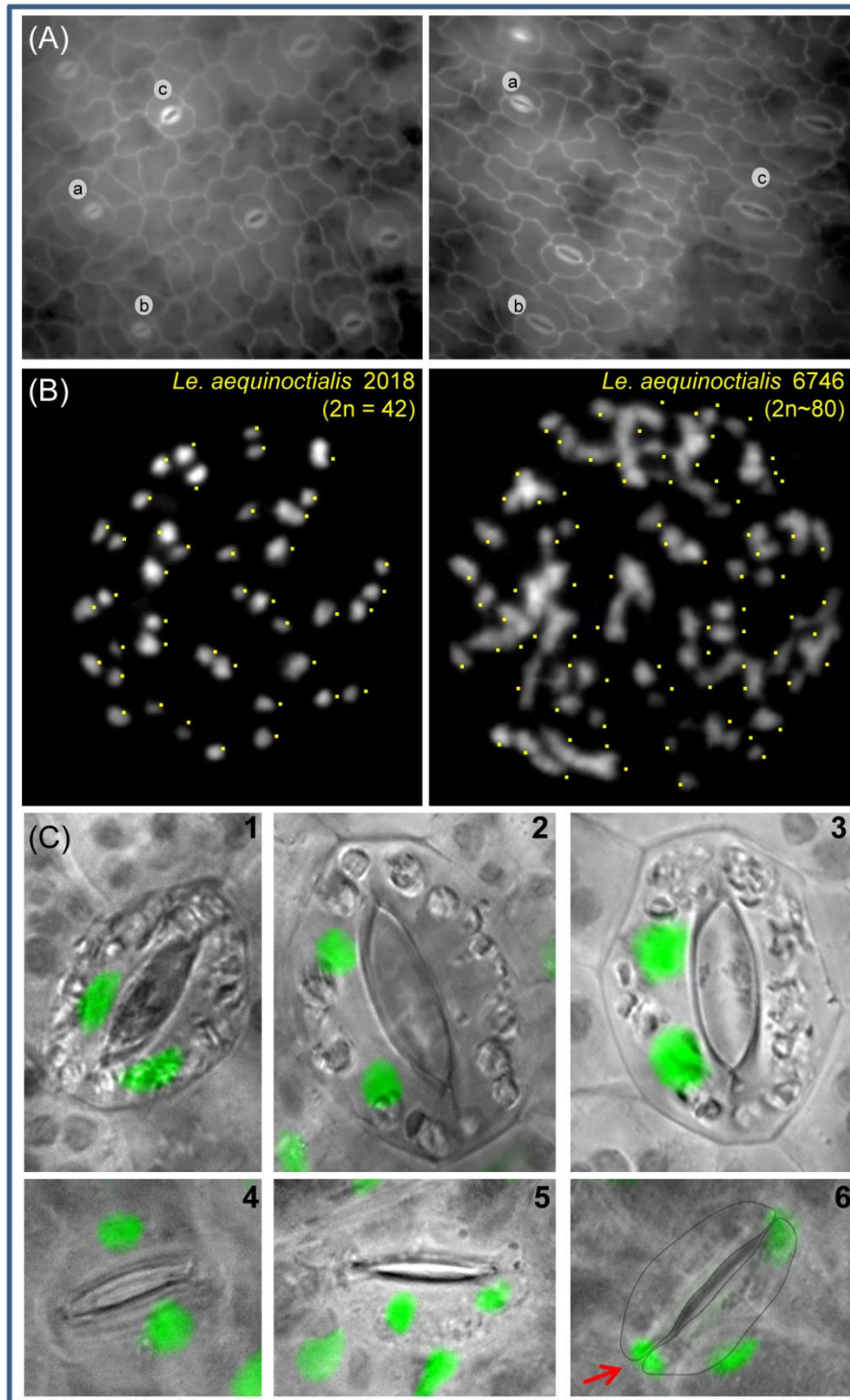


Figure 5. Variation in guard cell shape and volume of *Le. aequinoctialis* (clone 2018) (A), chromosome spreads of *Le. aequinoctialis* clones 2018 and 6746 (B), equal and abnormal nuclei distribution in sister guard cells of *Wa. hyalina* (C1-3) and *Wo. australiana* (C4-6).
 (A) Guard cells with round shape in the younger part (left) and elongated ones in the older part (right) are of different size (a,b,c) in *Le. aequinoctialis* (clone 2018), (B) Mitotic spreads of *Le. aequinoctialis* $2n = 42$ (clone 2018) and $2n \sim 80$ (clone 6746), (C1,4) Normal situation (one nucleus per cell); (C2,3,5) both nuclei in one sister guard cell; (C6) the lower nucleus (arrow) is possibly migrating into the sister cell.

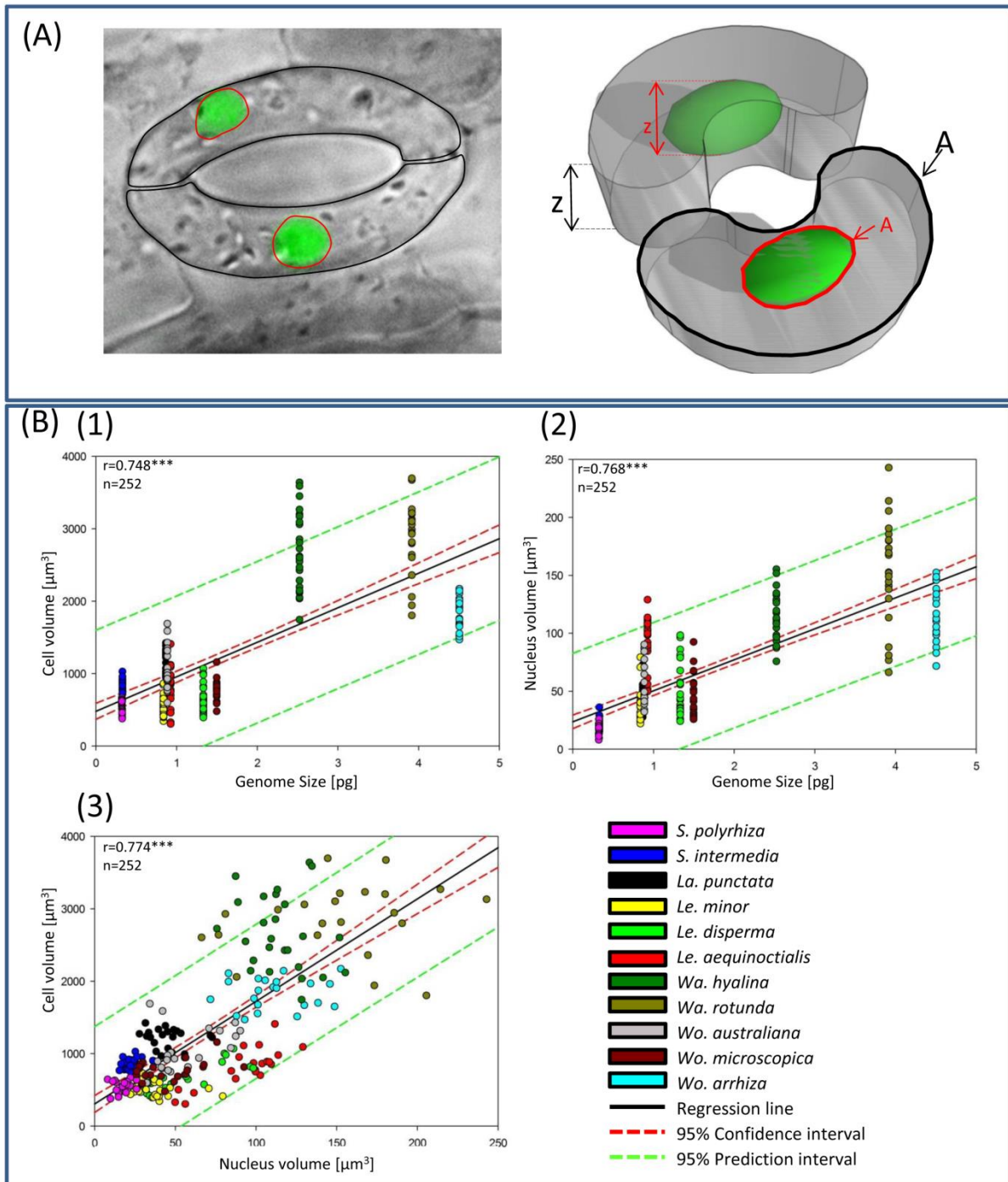


Figure 6. Guard cell and nuclear volume measurement (A) and linear regressions of duckweed cell parameters (B).

(A) DIC and fluorescence microscopy image stacks (left) were applied separately (here merged images) to measure the guard cells and the nuclei inside, respectively. The x-y areas (μm^2) and the z dimension (μm) were measured based on the black (guard cells) and red (nuclei) regions via the ZEN software (spatial illustration, right). (B) Correlations between genome size and cell volume (1) and nucleus volume (2), as well as between nucleus volume and cell volume (3). *** $p < 0.001$

3.2. Chromosome numbers and number of 5S and 45S rDNA sites in duckweeds

3.2.1. Chromosome numbers

Chromosome numbers of duckweed species have been reported in different references since 1933 (Table 1). However, it is still unclear whether the in part spectacular variations of chromosome number between asexual clones within the same species are counting errors or actual deviations. For instance, 40, 50, 66, 72, 78, 84, 65-76 chromosomes (Beppu and Takimoto, 1981) or 40, 50, 60, 80 (Urbanska, 1980) or only 42 and 84 (Geber, 1989) were counted in different *Le. aequinoctialis* clones. For *Wo. microscopica*, 70 chromosomes were counted by (Roy and Dutt 1967), while (Urbanska, 1980) claimed 40 and 80 chromosomes.

Among 34 *S. polyrhiza* clones mentioned in Wang *et al.*, 2011, chromosome number of nine clones were not determined, for three clones (7652, 7657 and 7364) $2n = 30$, and for the other clones $2n = 40$ were reported (Table 6)

Here, clones 7652 and 7657 were selected for counting the chromosome number and yielded $2n = 40$, as in clone 7498 (Fig. 7) and in further six *S. polyrhiza* clones (Hoang *et al.*, 2018). For *S. intermedia*, $2n = 36$ was reported by Geber (1989) in all six tested clones, while Urbanska counted $2n = 20$ (clone (7747) and $2n = 30$ (clone 7201) (Table 6). Here, we selected *S. intermedia* clones 8410 and 7747 for chromosome counting and found $2n = 36$ for both clones (Fig. 7). Similarity, for *La. punctata*, we counted $2n = 46$ for clones 7260, 5562 and 7449 (Fig. 7), while 50 and 40 chromosomes were reported for clones 7260 and 7449, respectively (Wang *et al.*, 2011) (Table 6). Therefore, for all investigated clones of *S. polyrhiza* (7498, 7652, 7657, 9500, 9505, 9507, 9509, 9510 and 9511), *S. intermedia* (8410 and 7747) and *La. punctata* (5562, 7260 and 7449), no variation of chromosome number was observed.

Our chromosome counting results are mainly similar to that of Geber (1989) (Table 3). In detail, *S. polyrhiza* showed $2n = 40$, *S. intermedia* $2n = 36$, *La. punctata* $2n = 46$, *Le. disperma* $2n = 44$ and *Le. minor* $2n = 42$ chromosomes. For *Wo. australiana* (clone 7540) $2n = 40$ were counted, as reported by Urbanska (1980), for *Le. aequinoctialis* (clone 2018) $2n = 42$ and for *Wa. hyalina* $2n = 40$ chromosomes were counted. For *Le. disperma* (clone 7269) $2n = 44$, for *Le. aequinoctialis* (clone 2018) $2n = 42$, for *Wo. microscopica* (clone 2005) $2n = 40$ and for *Wa. rotunda* (clone 9072) 82 chromosomes were counted for the first time in this study. Meanwhile *Wo.*

microscopica clones used by Urbanska (1980) and by Roy and Dutt (1967) got lost and therefore cannot be re-investigated. In case of *Wo. arrhiza*, 42 chromosomes were counted by Geber (clone 7347), and 30, 40, 50, 60, 70 and 80 chromosomes for different clones by Urbanska (1980), while we counted 60 chromosomes for clone 8872 (Table 6).

Table 6: Chromosome numbers of 11 tested duckweed species from our study and others.

Genus	Species	Clones	2n	Source	
Spirodela	polyrhiza	7652	30	U	
		7621	40		
		7110	50		
		8118, 7205, 7120, 7160, 7687, 8483, 8403, 8409, 6613, 7003, 7206, 6731, 7498 , 8442, 8229, 7212, 7551, 7674, 7960, 7222, 7379, 6581	40	W	
		7652, 7657 , 7364	30	G	
		7110	80		
		6613, 7667, 7364, 7551, S7, S3	40		
	7498, 7652, 7657	40	O		
	intermedia	7747	20	U	
		7201	30		
		8410 , 7355, 8258, 7747, 8818, 7178	36	G	
		8410	36	O	
	Landoltia	punctata	8028	40	U
7479			50		
7449 , 7248			40	W	
7260			50		
O5, O6, 7461, 7191, 7799, 7429			46	G	
7260, 7449			46	O	
Lemna	minor	7798	20	U	
		7244	30		
		6626	40		
		7572	42		
		6742	50		
		8623 , 7018, 7210, 8434, 7436, 7136	40	W	
		7123, 6591	42	W	
		7189, 8676, 7789, 7244	42	G	
		M4, 7114, 7182, 8653	63	G	
		7115	126	G	
		8623	42	O	
		Lemna	disperma	7818	40
7223, 7190	44			W	
7269 *	44			O	
7382	20			U	
8038	40				
7204	50				
8079	60				
6746	80			W	
6612	40				
7126	60				
6746	80				
6746	84		G		
2018 *	42		O		
6746	-80				
Wolffiella	hyalina		7426	40	U
			7378, 7376, 8640	40	W
			8640	40	O
	rotunda	9072 *	82	O	
Wolffia	australiana	7819	20	U	
		7540	40		
		7733	20	W	
		7540	40	O	
	microscopica	7238	40	U	
		8359	80		
		M8	70	R	
	2005 *	40	O		
	arrhiza	7251	30	U	
		8272	40		
7193		50			
7699		60			
7158		70			
7196		80			
7347		42	G		
8872		60	O		

(R) Roy & Dutt (1967) mentioned in Geber (1980); (U) Urbanska (1980), (G) Geber (1989), (W) Wang et al., 2011, (O) our study

Bold and Underlined: clones were used in our study; (*) clones were counted for the first time.

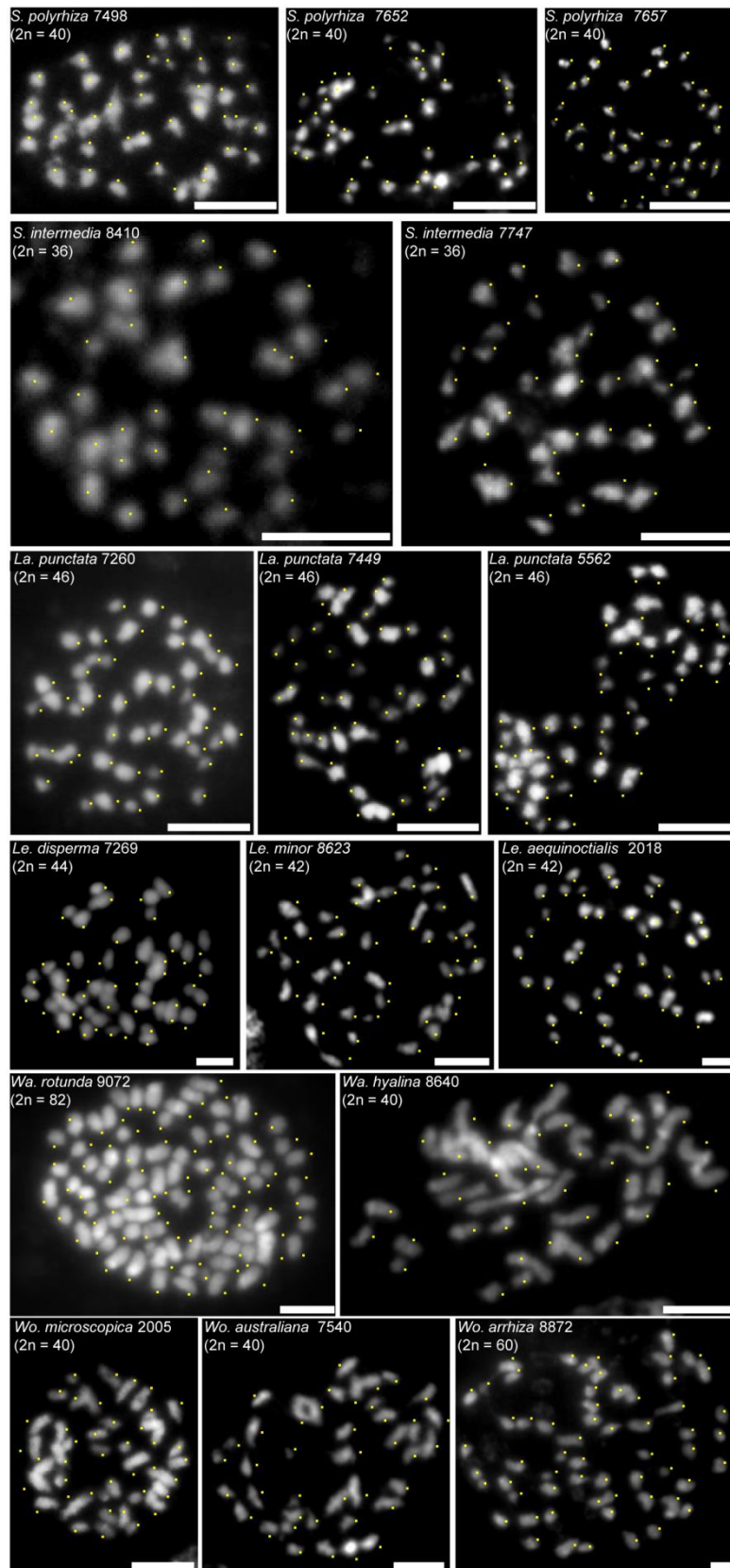


Figure 7. Chromosome number of distinct clones of eleven duckweed species. Scale bars= 5 μm

3.2.2. Ribosomal rDNA sites

In eukaryotic genomes, the conserved ribosomal RNA genes (rDNA) are present as multigene families organized in long tandem repeat units. Variations regarding number and chromosomal position of 5S and 45S rDNA are often species-specific and helpful to elucidate karyotype evolution, for instance in Brassicaceae (Ali *et al.*, 2005; Mandakova and Lysak, 2008) and in Anthemideae (Abd El-Twab and Kondo, 2012). The primers used in this experiment to amplify 5S and 45S rDNA regions from duckweed genomic DNA for probe generation and labeling are shown in Table 4.

Based on the genome assembly for *S. polyrhiza* clone 7498, published by Wang *et al.* (2014), BLAST results revealed two loci of 5S rDNA (*A. thaliana* sequence) (on Ψ 06 and Ψ 08 corresponding ChrS 13 and ChrS 06, respectively) and one locus of 45S (18S and 26S sequences from *S. polyrhiza*, Tippery *et al.*, 2015) on Ψ 23. In the BioNano map of *S. polyrhiza* clone 9509, one locus of 18S-5.8S-26S rDNA, with a strikingly low number repeats, was identified (Michael *et al.*, 2017). Here, 5S and 45S rDNA probes from *S. polyrhiza* 7498 genomic DNA were generated for FISH experiments to determine 5S and 45S rDNA loci in this species.

FISH with 5S rDNA revealed in most metaphases one locus on ChrS 13, corresponding to Ψ 08. While the locus on ChrS 06, corresponding to Ψ 06 could only be detected in few metaphases with very weak signals, hardly distinguishable from occasional background. These signals might represent the low copy locus. To explain the less reliable signals of 5S rDNA on ChrS 06, two BACs (026N01 and 008B06) harboring 5S rDNA flanked by sequences of Ψ 06 and Ψ 08, respectively, were selected. While only ChrS 13 was labeled by the BAC 008B06 (from Ψ 08), the BAC026N01 from Ψ 06 labeled both ChrS 06 and ChrS 13 (Fig. 8A), suggesting that BAC 026N01 contains the minor 5S rDNA region with a low copy number. In addition, 116 bp 5S rDNA fragments were obtained by PCR reactions when using the two BACs as templates. Afterwards, the 5S rDNA PCR products from both BACs were labeled to produce corresponding probes. The FISH results showed that only ChrS 13 was labeled by both 5S rDNA probes generated from BACs, and that ChrS 06 apparently harbors just a minor locus of 5S rDNA yielding weak signals and that could only rarely be detected by FISH (Fig. 9). Furthermore, the Oxford Nanopore assembly supported both loci and revealed a 5-fold larger copy number on ChrS 13 than on ChrS 06 (60 versus 12 copies) (Hoang *et al.*, 2018) Therefore, *S. polyrhiza*

possesses two 5S rDNA loci of strikingly different copy number of which one locus possesses very few copies compared to 2,000 copies within the similar-sized genome of *A. thaliana* (Campbell *et al.*, 1992).

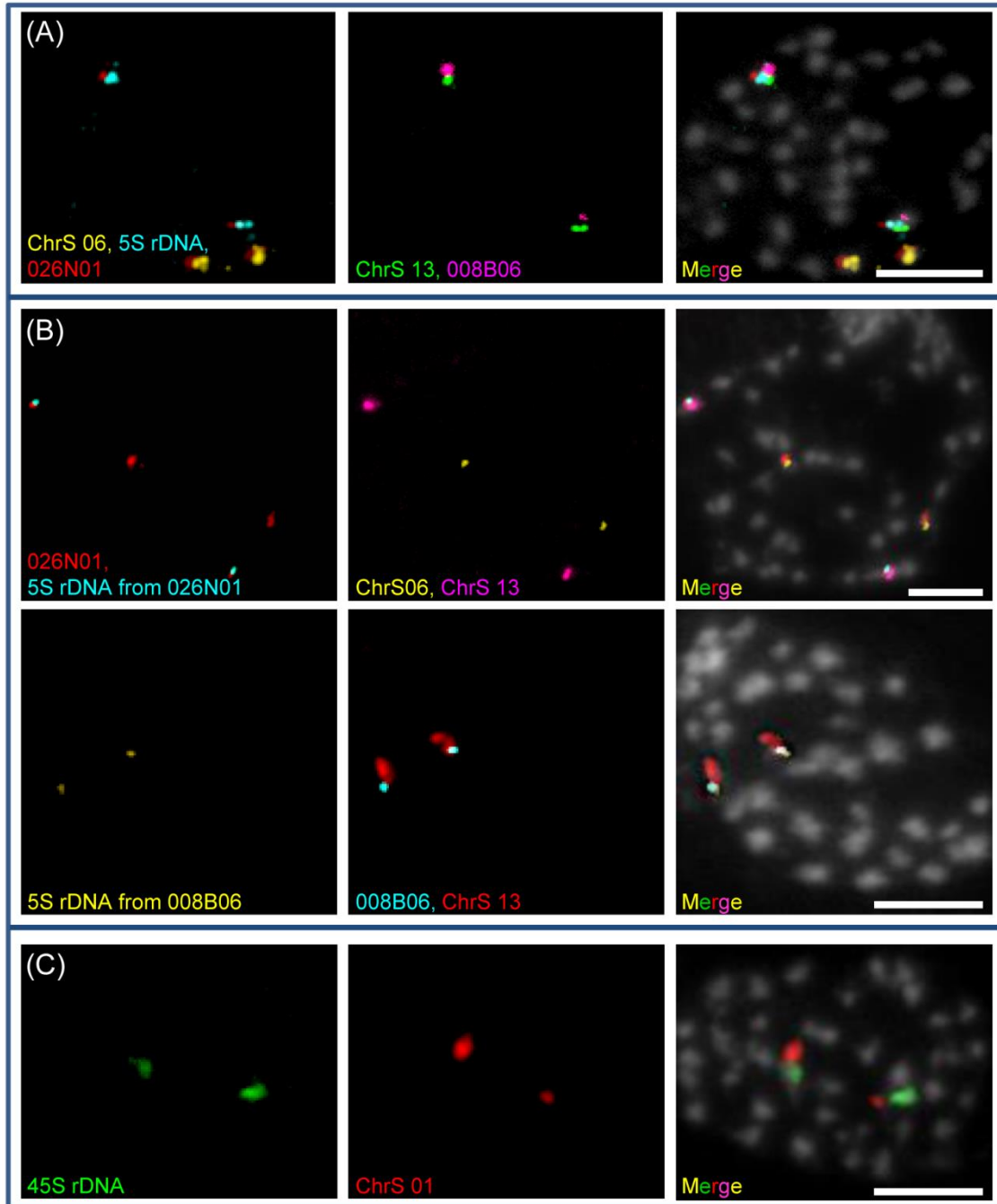


Figure 8. Chromosomal distribution of 5S and 45S rDNA on *S. polyrhiza*.

(A) BAC 026N01 (Ψ 06) labeled ChrS 06 and ChrS 13 while BAC 008B06 (Ψ 08) labeled only ChrS 13; (B) 5S rDNA probes generated from BAC 026N01 (upper panel) and from BAC 008B06 (lower panel) labeled only ChrS 13; (C) 45S probe labeled ChrS 01. The chromosomes were counterstained with DAPI. Scale bars = 5 μ m.

Only one region of 18S-5.8S-26S rDNA repeats was identified within the genome of *S. polyrhiza* clone 9509 (Michael *et al.*, 2017). This result could be confirmed by only one FISH signal for 45S rDNA on ChrS 01 (Fig. 8C and 9). Similar to 5S rDNA, a remarkably low number of 45S rDNA repeats in *S. polyrhiza* was discovered by independent approaches. From Southern blot analysis 87 to 94 copies were estimated, and optical mapping revealed 81 copies (Michael *et al.*, 2017). This is an extreme reduction compared to the 570 copies estimated in *A. thaliana* (Rosato *et al.*, 2016) with a similar genome size as *S. polyrhiza*, and the up to 12,000 copies in the more than 10-times larger genome of *Zea mays* (Liu *et al.*, 2017). Even the very small genome (~12.2 Mbp/1C) of *Saccharomyces cerevisiae* contains 150 copies (Kobayashi, 2014). The unusually low copy number of 5S and 45S ribosomal DNA in *S. polyrhiza* still awaits a reasonable explanation.

The 45S and 5S rDNA loci within eleven studied duckweed species are shown by FISH (Table 5, Fig. 9). In detail, one locus of 45S and 5S rDNA each was detected in *Le. minor*, *Le. disperma*, *Le. aequinoctialis*, *Wo. microscopica*, while *S. polyrhiza*, *S. intermedia*, *La. punctata*, *Wa. hyalina* and *Wo. australiana* displayed one locus of 45S rDNA and two loci of 5S rDNA. In *Wa. rotunda* and *Wo. arrhiza*, two of 45S rDNA and three loci of 5S rDNA were detected. In *Wa. rotunda* (clone 9072), one pair of NORs was more extended and showed a distal satellite (Fig. 10). The strength of FISH signals reflected differences in copy number. For instance, the 5S rDNA probe yielded in *Wo. arrhiza* (clone 8872) two strong, two medium and two weak FISH signals. Noticeably, a very low copy number of 5S rDNA could apparently prevent detection by FISH, e.g. the 5S rDNA locus with only 12 copies on ChrS 06 of *S. polyrhiza* (Hoang *et al.*, 2018). Weak signals of 5S rDNA loci (in *S. polyrhiza*, *S. intermedia*, *La. punctata* and *Wo. arrhiza*) could only be detected in a few metaphases (Fig. 9), and thus are at risk to be overlooked. Therefore, the number of 5S rDNA loci which were detected by FISH in other duckweed species than *S. polyrhiza* might underestimate the true number of loci as long as their genomes are not completely assembled.

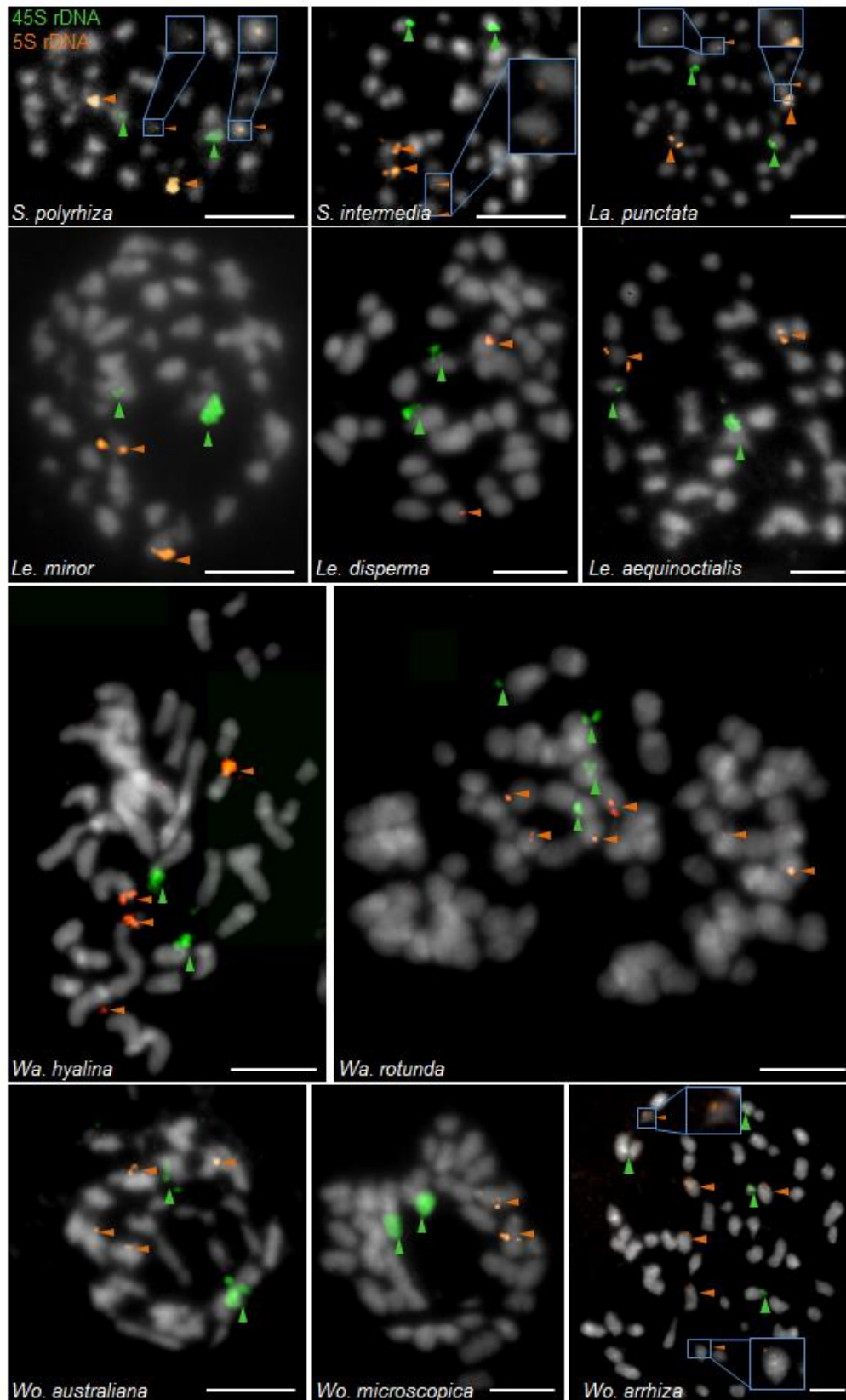


Figure 9. 5S and 45S rDNA loci on duckweed species.

Two loci of 5S and 1 locus of 45S rDNA were detected on *S. polyrhiza*, *S. intermedia*, *La. punctata*, *Wa. hyalina*, *Wa. australiana*; one locus of 5S and 45S each were detected on *Le. minor*, *Le. disperma*, *Le. aequinoctialis* and *Wa. microscopica*; three loci of 5S and two loci of 45S rDNA were detected on *Wa. rotunda* and *Wa. arrhiza*. Framed: minor locus of 5S rDNA. Scale bars = 5 μm

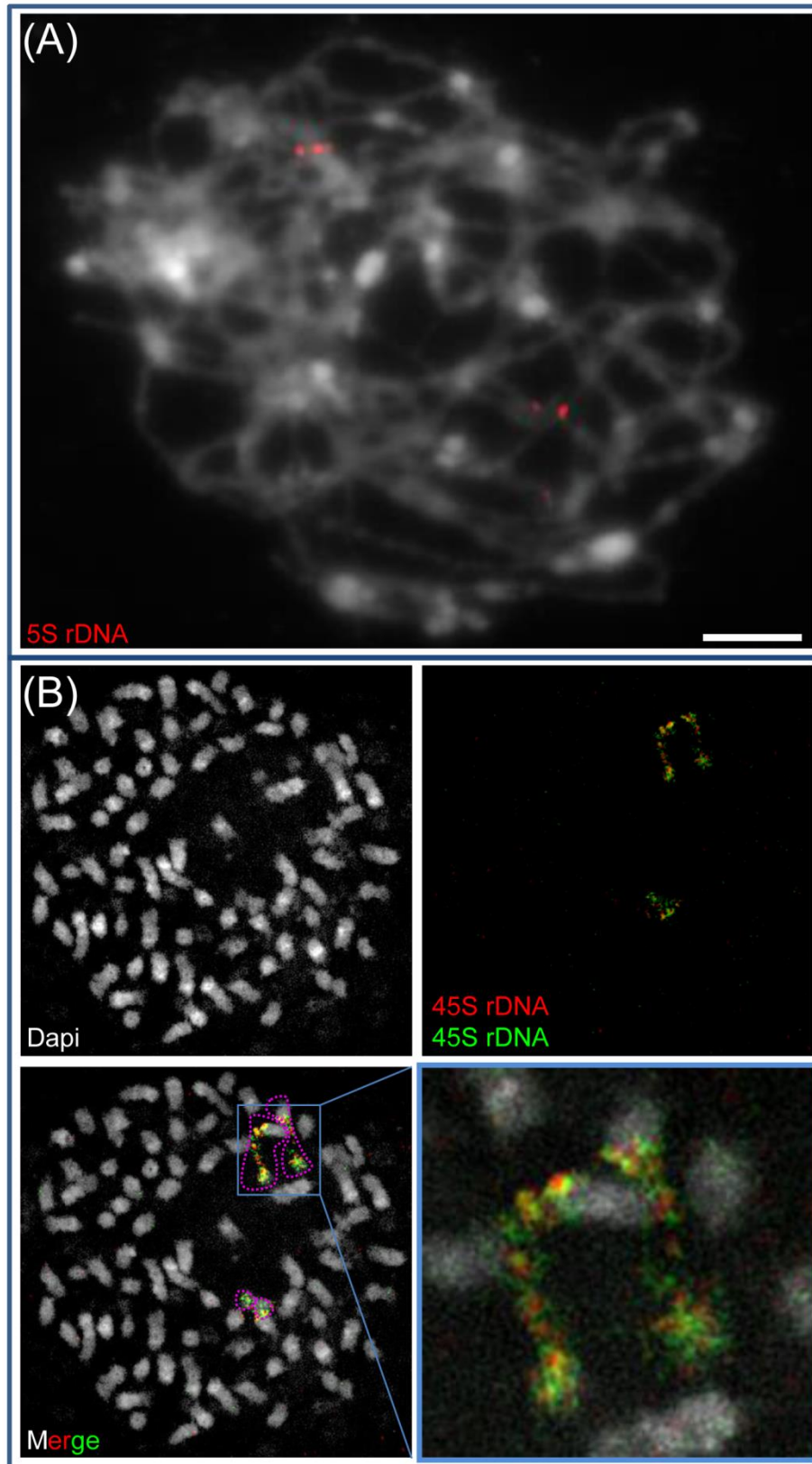


Figure 10. rDNA FISH signals in pachytene (A) and mitotic metaphase (B) of *Wa. rotunda* (clone 9072) using super-resolution microscopy (SIM).

(A) Three loci of 5S rDNA were detected on pachytene chromosomes. Scale bar = 5 μm ; (B) 45S rDNA FISH signals with two chromosome pairs harboring 45S rDNA (red dotted circles) and enlarged frame (bottom right) with extended (red and green) secondary constriction. Without rDNA FISH signals, the satellite distal the NOR could erroneously be counted as a small pair of chromosomes

3.3. A robust genome map for *S. polyrhiza*

Because of their peculiar biological features and their importance as potential aquatic crops, duckweeds become an increasingly interesting subject for genome research. Therefore, a high-quality genome sequence assembly will be a pre-requisite as a source for further research into molecular biology in duckweeds. In order to meet this requirement, the Greater duckweed, *S. polyrhiza* (clone 7498), was chosen for whole genome sequencing due to its ancestral phylogenetic position, commercial potential as well as its small genome size (160 Mbp) with a presumably low content of repetitive DNA (Wang *et al.*, 2014). This genome assembly of *S. polyrhiza* yielded 32 pseudomolecules, which were validated and thereafter integrated into the 20 chromosome pairs by mcFISH and resulted in a *S. polyrhiza* cytogenetic map (Cao *et al.*, 2016). Then, an optical map for *S. polyrhiza* (clone 9509) was established by combination of high-depth short read sequencing and high-throughput genome mapping technologies. The BioNano Genomics Irys® System was applied to generate deep coverage physical maps, named BioNano map (Michael *et al.*, 2017). However, several differences in chromosomal assignment of pseudomolecules and in chromosome enumeration were revealed between the two genome maps. These discrepancies had to be clarified in order to obtain a robust genome map for *S. polyrhiza* as an important reference genome for further genome structure and karyotype evolution studies in duckweeds.

The discrepancies between the cytogenetic map and the BioNano map are summarized in Table 7. Regarding the different chromosome enumeration, only chromosome 19 is the same within the cytogenetic map (ChrS 01 - 20) and the BioNanomap (Chr 01 - 20). Both maps tried to order the chromosomes according to their estimated length. In several cases the composition of chromosomes differed between both maps: (1) ChrS 04 (including Ψ 04) and ChrS 13 (Ψ 08) were considered as Chr 01; (2) Ψ 11, Ψ 14 and Ψ 16 were considered as chimeric pseudomolecules in BioNanomap; (3) Ψ 21b combined with Ψ s 7a and 32 should form ChrS 17, or combined with Ψ s 16b and 30 yield Chr 14; and (4) Ψ 21a should form ChrS 08 together with Ψ s 19 and 26 or stay separately as Chr 20.

Table 7: Differences in chromosome enumeration (A) and chromosomal assignment of pseudomolecules (B) between *S. polyrhiza* cytogenetic map (for clone 7498) and BioNano map (for clone 9509).

(A)

ChrS (Cytogenetic map)	13 and 4	1	2	3	6	7	10	9	5	11	8	16	12	14	18	17	15	20	19	
Chr (BioNano map)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20

(B)

Chromosome	Cytogenetic map	BioNano map	Chromosome	Cytogenetic map	BioNano map
01	Ψ 3, Ψ 23	<u>Ψ 4⁽¹⁾, Ψ 8⁽¹⁾, Ψ 16a⁽²⁾</u>	11	Ψ 5	Ψ 26, Ψ 19
02	Ψ 1	Ψ 3, Ψ 23	12	Ψ 12, Ψ 31	<u>Ψ 13, Ψ 14b⁽²⁾</u>
03	Ψ 2	Ψ 1	13	<u>Ψ 8⁽¹⁾</u>	Ψ 12, Ψ 31
04	<u>Ψ 4⁽¹⁾</u>	Ψ 2	14	Ψ 16, Ψ 30	<u>Ψ 16b⁽²⁾, Ψ 21b⁽³⁾, Ψ 30</u>
05	Ψ 14, Ψ 25, Ψ 28	Ψ 6, Ψ 29	15	Ψ 10	Ψ 15
06	Ψ 6, Ψ 29	Ψ 9, Ψ 20	16	Ψ 13,	Ψ 7a, Ψ 32
07	Ψ 9, Ψ 20	Ψ 7b, Ψ 27a, Ψ 22	17	Ψ 7a, Ψ 32, <u>Ψ 21b⁽³⁾</u>	Ψ 10
08	<u>Ψ 21a⁽⁴⁾</u> , Ψ 19, Ψ 26	<u>Ψ 11a⁽²⁾, Ψ 24</u>	18	Ψ 15	Ψ 17
09	Ψ 11, Ψ 24	<u>Ψ 14a⁽²⁾, Ψ 25, Ψ 28</u>	19	Ψ 18, Ψ 27b	Ψ 18, Ψ 27b
10	Ψ 7b, Ψ 27a, Ψ 22	<u>Ψ 5, Ψ 11b⁽²⁾</u>	20	Ψ 17	<u>Ψ 21a⁽⁴⁾</u>

Bold and underlined Ψs indicate different chromosomal assignment between the maps

⁽¹⁾ Ψ 08 and 04 were considered to constitute Chr 01 or to be separated as ChrS 04 and ChrS 13;

⁽²⁾ new chimeric pseudomolecules (Ψ 11, Ψ 14 and Ψ 16) according to the BioNano map.

⁽³⁾ Ψ 21b considered to co-localize with Ψ 16b and 30 as Chr 14 or with Ψ 07a and 32 as ChrS 17;

⁽⁴⁾ Ψ 21a was considered to stay separately as Chr 20 or to be linked with Ψ 19 and Ψ 26 as ChrS 08;

Grey filled: the only identical chromosome between both genome maps

In order to address whether the discrepancies between both maps are due to clone-specific chromosome rearrangements between *S. polyrhiza* clones 7498 and 9509 or not, we applied mcFISH with 106 BACs (Table 8) on these two *S. polyrhiza* clones. Five clones of different geographic origin (Table 2) were additionally included to figure out potential clone-specific rearrangements (if any). According to the aim, for each experiment, suitable BACs were selected from the chromosomal region of interest and used separately or pooled with other clones of the chromosomes of interest for mcFISH.

Table 8: 106 BACs of the 20 *S. polyrhiza* chromosomes integrating 39 pseudomolecules (including Ψ 0).

ChrS	BACs (start - end position) and pseudomolecules (length)																								
1	030D21 (2606260 - 2685831) 014L22 (1717219 - 1827551) 019A13 (unknown - 311931) 012O14 (4737987 - 4860473) 019L04 (6303899 - 6438502) 023E03 (7836208 - 7949728) 003D17 (8650584 - unknown)					Ψ 23 (2725124)					Ψ 3 (8736411)														
2	004D21 (1117391 - 1226295) 005N05 (2775089 - 2903059) 037B02 (6237992 - 6338310) 003P15 (7623129 - 7741301) 003H12 (8927523 - unknown)										Ψ 1 (8941172)														
3	021D24 (3036781 - 3146218) 002E10 (3243741 - 3360983) 002K06 (6082625 - 6201482) 011P20 (6431461 - 6537909) 010J02 (8574996 - 8668853)										Ψ 2 (8796147)														
4	004A24 (*) 006F09 (1336569 - 1445949) 013D18 (3418863 - 3507724) 035L13 (5720468 - 5826867) 017E03 (8352328 - 8444691)					024J03 (18141 - unknown) 025H07 (18442 - unknown)					Ψ 4 (8491500)					Ψ 16 (3759109) - a									
5	030K22 (2749101 - 2824616) 027D17 (626874 - 673272) 009A15 (116434 - 224310) 005B18 (*) 009O20 (1018564 - 1140429)					Ψ 14 (4409270) - a					Ψ 25 (2126887)					Ψ 28 (1843447)									
6	023I01 (72545 - unknown) 034P03 (2339883 - 2365418) 002I04 (2423356 - 2556370) 026N01 (3493229 - 3579170) 035L19 (5679141 - 5816708) 033A10 (14644 - 124630)					Ψ 6 (633238)					Ψ 29 (1792637)														
7	014F03 (134000 - 235973) 035P21 (1769022 - 1875238) 004A04 (4122930 - 4235007) 013C18 (465110 - 570238) 030B11(3046632 - 3154571)					Ψ 9 (4924802)					Ψ 20 (3177748)														
8	013I04 (unknown - 209512) 006P24 (2016823 - 2084639) 006L17 (1803530 - unknown) 032L08 (unknown - 512544) 034K03 (1604431 - 1734467) 004E01 (2431463 - 2507949) 004G18 (1333667 - 1444699)					Ψ 21 (2998408) - a					Ψ 26 (1959162)					Ψ 19 (3286220) - a					Ψ 0				
9	037H06 (4487847 - 4623724) 025O16 (3521620 - 3639111) 019G12 (2291940 - 2392994) 024C02 (211879 - 335777) 036C07 (unknown - 107534) 001D13 (1567225 - 1675474) 010M07 (2389434 - 2412255)					Ψ 11 (4687547) - a					Ψ 24 (2515492)														
10	035P14 (6156209 - 6227891) 002B12 (4639357 - 4760314) 037K21 (3369801 - 3491621) 008G11 (3260495 - 3369796) 020G09 (947293 - 1069203) 028B20 (13186 - 121259) 026G16 (2195544 - 2288704)					Ψ 7 (6239254) - b					Ψ 27 (1858560) - a					Ψ 22 (2983352)									
11	019I03 (98023 - 208537) 010P10 (3025329 - 3127384) 004G03 (4185984 - 4277305) 011L15 (5606973 - 5711303) 001C15 (6887 - 120542)					Ψ 5 (6552830)					Ψ 11 (4687547) - b														
12	028L12 (unknown - 1096345) 035F17 (4411629 - 4530198) 018A14(*) 030D24 (751534 - 851534) 017A15 (863291 - 971787)					Ψ 12 (4671450)					Ψ 31 (1269729)														
13	031L10 (unknown - 200940) 008B06 (1598133 - 1646677) 031B15 (2302375 - 2436388) 010L16 (3749165 - 3856320) 020L20 (5315551 - 5439967)					Ψ 8 (5476630)																			
14	003B08 (271353 - 411550) 036F14 (2051159 - 2151426) 040G15(3425541 - 3540680) 037I18 (2379739 - 2455354) 006D12 (2804349 - 2916783) 006A07 (119532 - 234004)					Ψ 16 (3759109) - b					Ψ 21 (2998408) - b					Ψ 30 (1339597)									
15	011C23 (149483 - 260499) 021P03 (971326 - 1080357) 020E12 (2320463 - 2431482) 035L20 (3076199 - 3163936) 013G21 (4150978 - 4235244)										Ψ 10 (4726429)														
16	009J15 (232869 - 319888) 011H02 (*) 002L10 (3574027 - 3664195) 024L10 (1545287 - 1665194) 040C11 (1635096 - 1746050)					Ψ 13 (4623610)					Ψ 14 (4409270) - b					Ψ 0									
17	006B02 (unknown - 31094) 009O07 (unknown - 805068) 004N06 (2942393 - unknown) 009L02 (816544 - 956328)					Ψ 7 (6239254) - a					Ψ 32 (993548)														
18	026D06 (5837 - unknown) 037B13 (2072992 - 2189428) 029K19 (3411813 - 3519955)										Ψ 15 (4370269)														
19	015L03 (186904 - 383018) 029E13 (1772604 - 1877776) 015J20 (3182285 - 3297665) 037H19 (1609342 - 1699356) 008A12 (1576879 - 1685451)					Ψ 18 (3441128)					Ψ 27 (1858560) - b														
20	006C19 (2155226 - 2262972) 007A03 (3275732 - 3383604)										Ψ 17 (3541257)														

BACs were not used in Cao *et al.* (2016); **BACs** in italics and underlined were not included in the probe set for karyotyping clone 9509 shown on Fig. 19; (*) BACs were selected based on their position in fingerprinted contigs (FPC) in the 7498 assembly (Cao *et al.* 2016).

To test the difference in chromosomal assignment of Ψ 08 and Ψ 04, FISH with the pooled BAC probes belonging to Ψ 08 (031L10, 031B15, 010L16 and 020L20) and Ψ 04 (004A24, 006F09, 013D18, 035L13 and 017E03) was performed. Our results showed for all investigated *S. polyrhiza* clones that Ψ 04 and Ψ 08 labeled two different chromosome pairs (Fig. 11), as described in Cao *et al.* (2016) instead of

forming the largest *S. polyrhiza* chromosome. This result showed that the enumeration in BioNano map is not suitable. In order to avoid confusion for further genome studies, we used chromosome enumeration system as reported in cytogenetic map (Cao *et al.*, 2016) for our updated map instead of generating a new one, although the chromosomes are now not in all cases strictly ordered according to their size in Mbp (Hoang *et al.*, 2018)

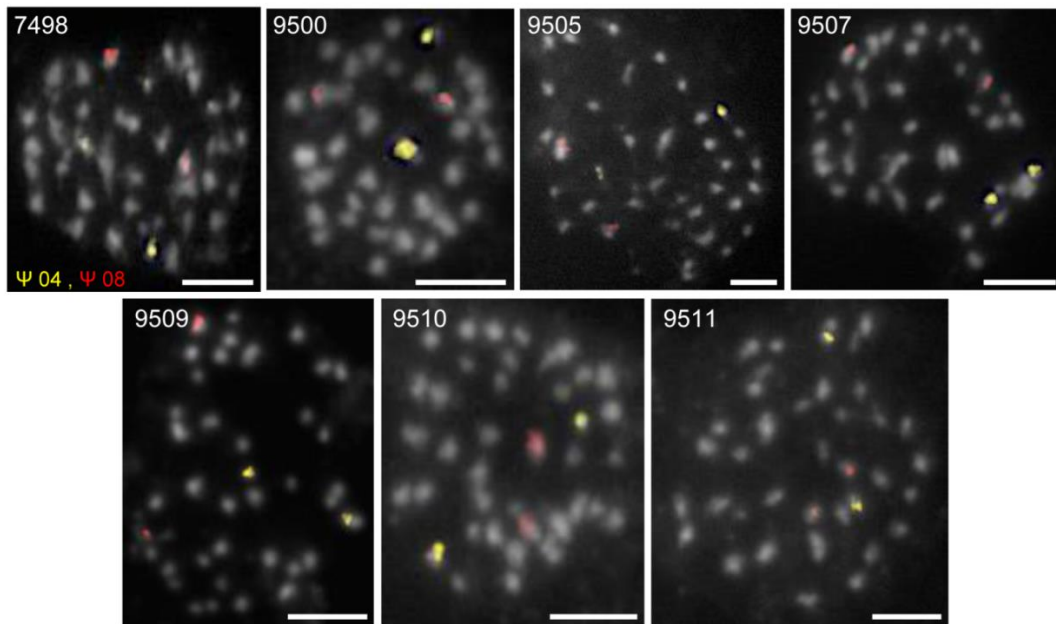


Figure 11. Chromosomal distribution of pseudomolecules 08 and 04 on *S. polyrhiza*.

Pooled BACs probes of Ψ 04 (red) and Ψ 08 (yellow) labeled two different chromosome pairs as reported in Cao *et al.* (2016) for seven investigated *S. polyrhiza* clones (white number on the left corner of each panel indicates clone ID). Probes were labeled by Cy3 (yellow) and Texas-Red (red), the chromosomes were counterstained with DAPI. Scale bars = 5 μm.

In addition to the three chimeric pseudomolecules (Ψ 07, 21 and 27) identified in cytogenetic map by Cao *et al.* (2016), further three Ψs (11, 14 and 16) were suggested in the model of Michael *et al.* (2017) to be chimeric ones. Our FISH results from different series of experiments confirmed that these Ψs are indeed chimeric.

To test whether Ψ 16 is a chimeric one as suggested in the BioNano map, the following experiment was performed. BACs selected previously for generation of the cytogenetic map flanking the region from 271 353 to 3 540 680 of the 3 759 109 bp of Ψ16 and were located on ChrS 14. Now, two partially overlapping BACs were selected as probes: 024J03 ending at position 18 141 and 025H07 ending at position 18 442 of Ψ 16, while the starting positions of both BACs are unknown. Both newly tested BAC probes co-localized with Ψ 04 on ChrS 04 confirming that this

pseudomolecule is chimeric (Fig. 12). Hence, ChrS 04 including Ψ 04 and a part of Ψ 16 confirmed linkage between Ψ 04 and Ψ 16 as described by Michael *et al.* 2017.

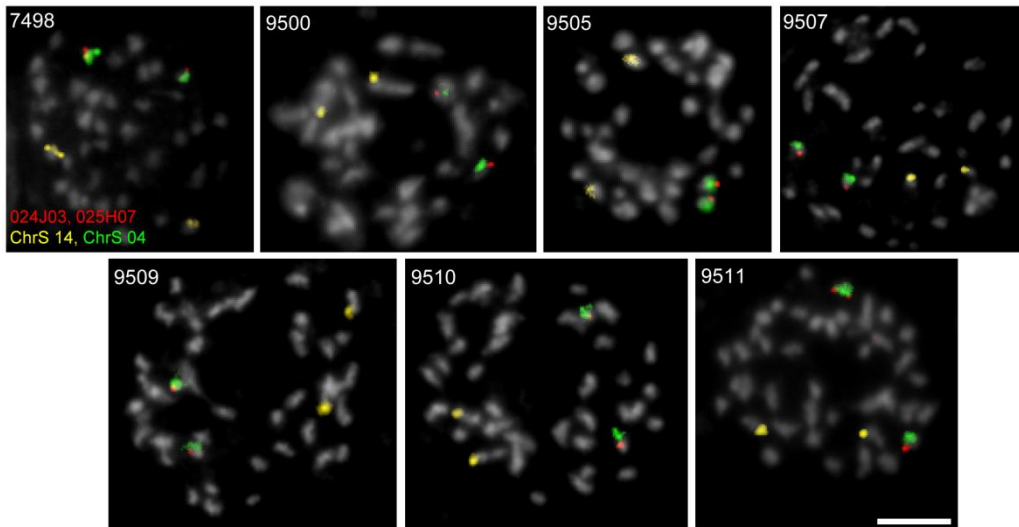


Figure 12. Location of chimeric pseudomolecule Ψ 16.

Newly tested BACs 024J03 and 025H07 (red) (Ψ 16a) co-localized with ChrS 04 (green), while the remaining BACs (Ψ 16b) co-localized with Ψ 21b and Ψ 30 on ChrS 14 (yellow). Scale bar = 5 μ m.

In order to test chimerism of Ψ 11, signals of two previously un-tested BACs from Ψ 11 revealed that BAC 024C02 co-localized with Ψ 24 on ChrS 09 (pink), while BAC 001C15 (blue) labeled the same chromosome pair (ChrS 11) as Ψ 05 (cyan), confirming that Ψ 11 is a chimeric pseudomolecule as suggested by BioNano map (Fig. 13).

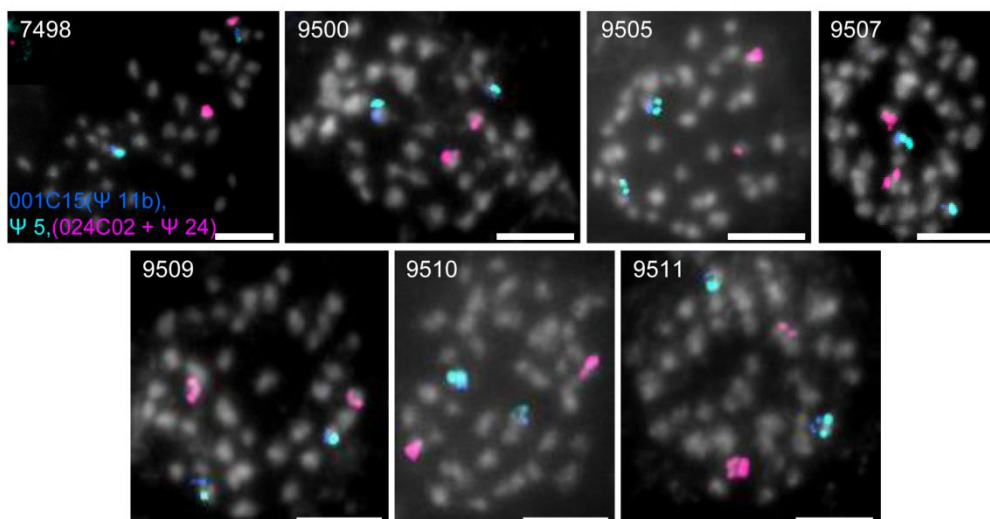


Figure 13. Location of chimeric pseudomolecule Ψ 11.

BAC 001C15 (blue) of Ψ 11b co-localized with Ψ 5 (cyan) on ChrS 11, while BAC 024C02 of Ψ 11a co-localized with Ψ 24 on ChrS 09 (pink). The chromosomes were counterstained with DAPI. Scale bars = 5 μ m

Pseudomolecule 14 with 4,409,270 bp in length was integrated into *S. polyrhiza* cytogenetic map by FISH with two BACs (027D17 with the start and end positions: 626 874 – 673 272 and 030K22 with start and end positions 2 749 101- 2 824 616) which co-localized with Ψ 25 and Ψ 28 on ChrS 05 (Cao *et al.*, 2016). To test the possible chimeric nature of Ψ 14, BAC 002L10 from the other end of Ψ 14 at position 3 574 027 – 3 664 195 was selected. FISH results showed for all investigated *S. polyrhiza* clones that BAC 002L10 co-localized with Ψ 13 in one chromosome pair (ChrS 16), while BACs 027D17 and 030K22 labeled another chromosome pair (ChrS 05) (Fig. 14).

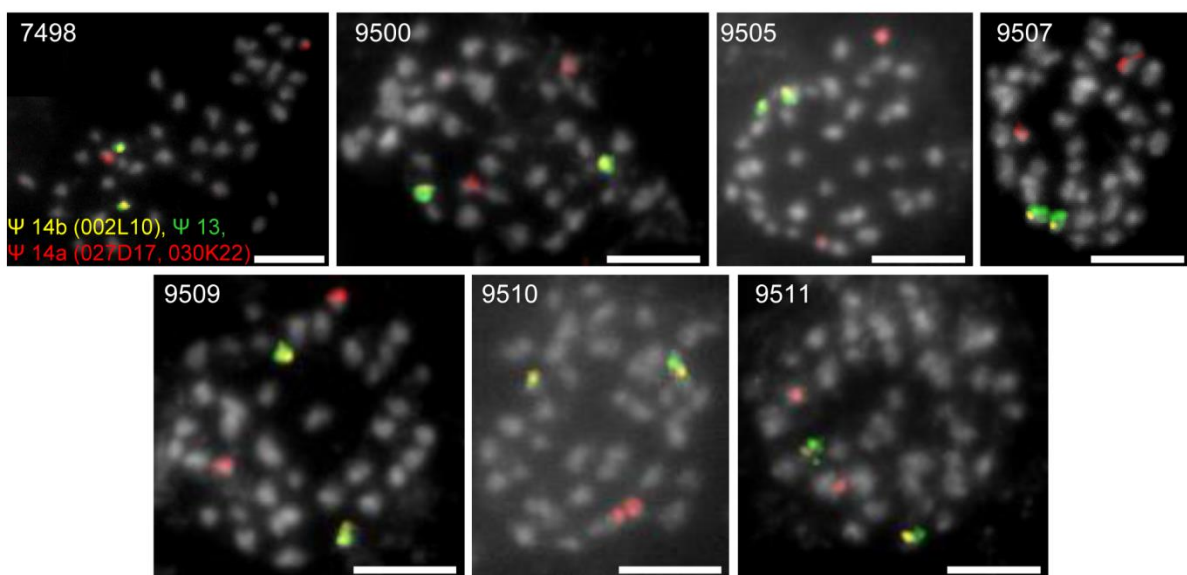


Figure 14. Location of chimeric pseudomolecule Ψ 14.

BAC 002L10 (yellow) belonging to Ψ 14b co-localized with BACs 009J15 and 011H02 of Ψ 13 (green) on ChrS 16, while Ψ 14a including BACs 030K22 and 027D17 (red) labeled another chromosome pair (ChrS 05); The chromosomes were counterstained with DAPI. Scale bars = 5 μ m.

Another experiment was performed to test whether Ψ 21b belongs together with Ψ s(16 + 30) to ChrS 14 or with Ψ s (07a + 32) to ChrS 17. Two BAC probes of Ψ 21b (006D12 = yellow and 037I18 = green) were used as probes. A first FISH experiment confirmed that these two BACs co-localized on the same chromosome. Re-hybridization with pooled BAC probes corresponding to Ψ s (16b + 30) and Ψ s (07a + 32), confirmed the combination of Ψ 16b, 21b and 30, as reported by Michael *et al.* (2017) (Fig. 15).

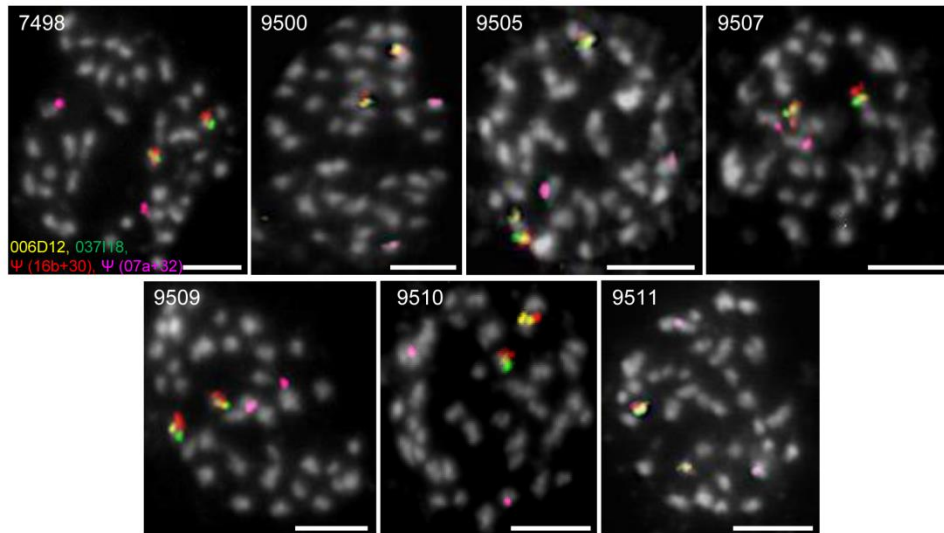


Figure 15. Location of Ψ 21b on *S. polyrhiza* chromosome ChrS 14.

Two BACs 006D12 (yellow), 037118 (green) of Ψ 21b labeled together with Ψ (16b+30) (red) one chromosome pair, while Ψ (07a+ 32) (pink) labeled another chromosome pair of seven *S. polyrhiza* clones, indicating linkage of Ψ 21b – 16b – 30 in ChrS 14 as reported in Michael *et al.* (2017). The chromosomes were counterstained with DAPI. Scale bars = 5 μ m.

In order to determine the location of Ψ 21a, which was considered either to stay separately as Chr 20, or to be linked with Ψ 19 and Ψ 26 as ChrS 08. The pooled BACs of Ψ 21a (013I04 and 006P24) were hybridized together with BAC pools representing Ψ 19 and Ψ 26. The results for all seven *S. polyrhiza* accessions supported linkage of Ψ 21a, Ψ 19 and Ψ 26 (Fig. 16), confirming the conclusion of Cao *et al.* (2016).

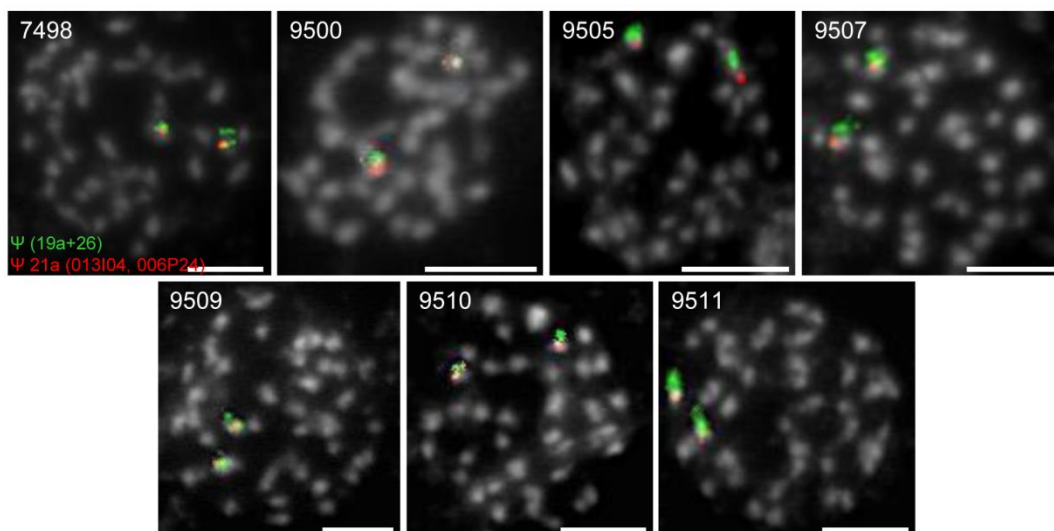


Figure 16. Location of Ψ 21a on *S. polyrhiza* chromosome ChrS 08.

BACs 013I04 and 006P24 of Ψ 21a (red) labeled in seven *S. polyrhiza* clones the same chromosome pair as pooled BACs of Ψ 19a and Ψ 26 (green), indicating linkage of Ψ 21a, 19a and 26 on ChrS 08 (as reported by Cao *et al.* (2016)). Probes were labeled by Alexa 488 (green) and Texas-Red (red), the chromosomes were counterstained with DAPI. Scale bars = 5 μ m.

Hence, the present FISH experiment series to validate and integrate the *S. polyrhiza* genome assembly increased the number of chimeric pseudomolecules up to six (Ψ 11, 14 and 16 together with Ψ 07, 21 and 27 which were previous detected). All discrepancies between two previous genome maps were solved and no clone-specific rearrangements were detected for seven investigated *S. polyrhiza* clones by FISH with 106 anchored BACs (Fig. 17).

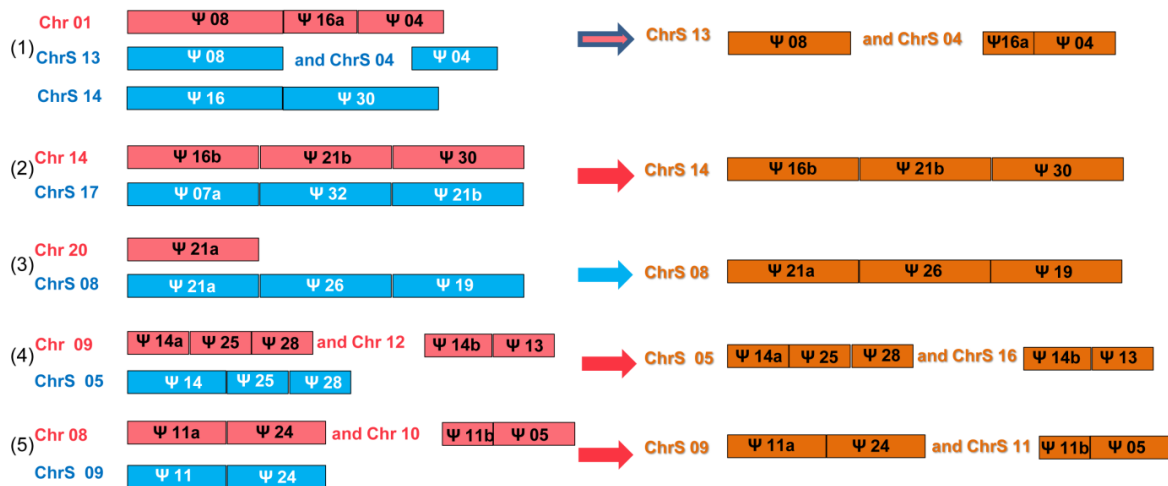


Figure 17. Solving discrepancies between the cytogenetic map (blue) and the BioNano map (red) resulted in an updated map (orange) of *S. polyrhiza*.

(1) Ψ 08 and 04 located in two different chromosome pairs (ChrS 13 and ChrS 04) and Ψ 16 is split between ChrS 04 and ChrS14; (2) Ψ 21b co-localized with Ψ 16b and 30 on ChrS 14; (3) Ψ 21a co-localized with Ψ 26 and 19 on ChrS 08; (4) Ψ 14 is split between ChrS 05 and ChrS 16; and (5) Ψ 11 is split between ChrS 09 and ChrS 11. The indicated segments are only for diagrammatic illustration and are not drawn to scale.

In parallel, Todd Michael's group obtained data for the *S. polyrhiza* genome by Oxford Nanopore sequencing – a recent low cost, long-read sequencing technology. Genome assembly from Oxford Nanopore sequences confirmed our FISH results regarding the discrepancies between the previous maps as mentioned above and detected another mis-assembly in the BioNano map. In detail, contig 06 of Oxford Nanopore assembly showed a 834 kb fragment (parts of Ψ 19 and the previously unassembled Ψ 0) to be located on Chr 12 (ChrS16) instead of belonging to Chr 11 (ChrS 08) as reported in the BioNano map. In order to determine this mis-assembly, *S. polyrhiza* BAC end sequences were BLASTed against Sp7498v2 (Wang *et al.*, 2014) and Sp9509v3 (Michael *et al.*, 2017) databases. Based on the BLAST result, we selected two BACs, 024L10 and 040C11, both belonging to Ψ 0 (Sp7498v2) and to Chr 11 (Sp9509v3), to produce corresponding probes. The FISH result showed that both BACs co-localized to ChrS 16, confirming the mis-assembly in BioNano map (Fig. 18).

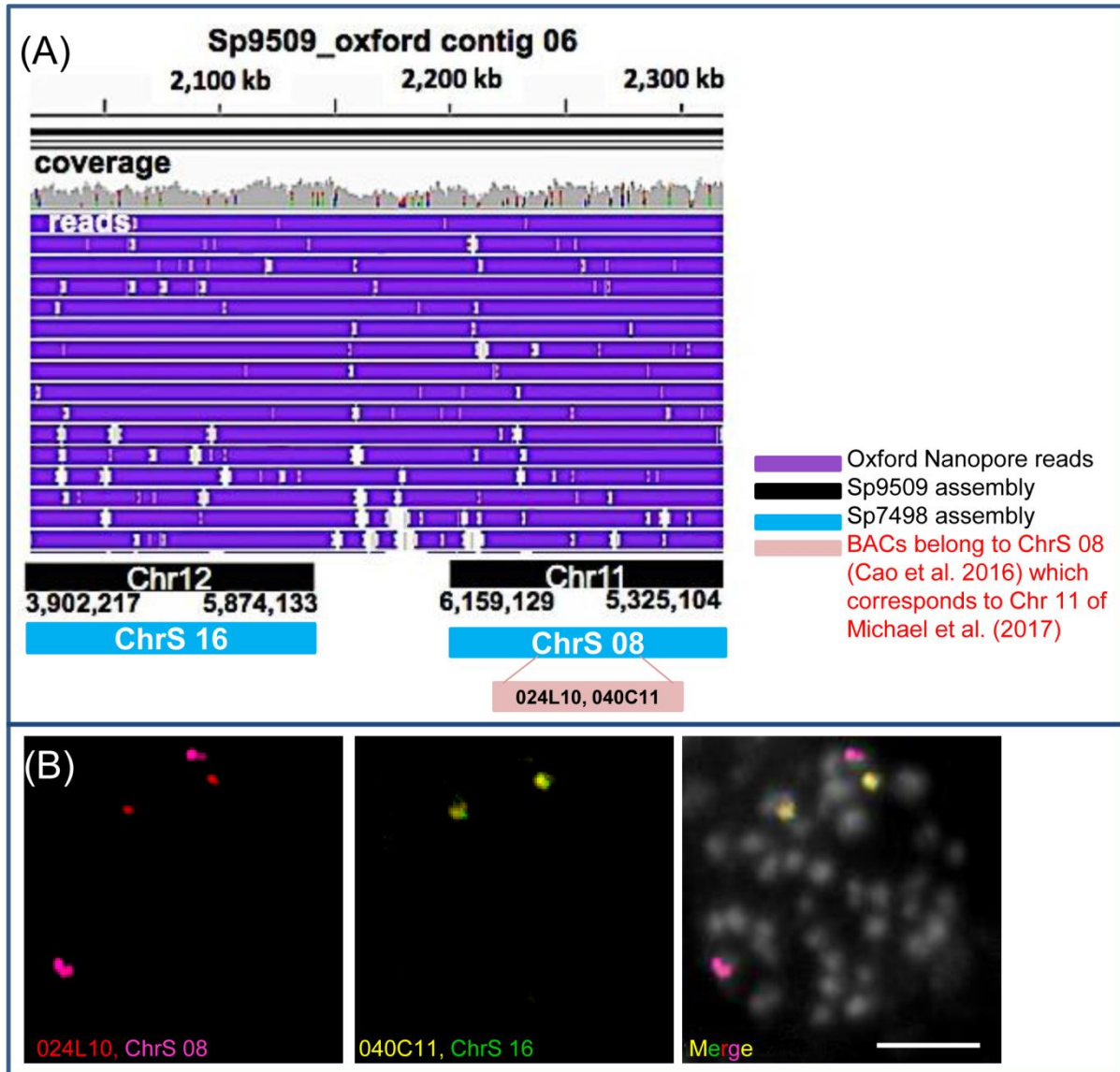


Figure 18. 834 kb mis-assembly in BioNano map was detected by Oxford Nanopore and confirmed by FISH.

A) Sp9509_oxford Contig06 links Chr 12 (ChrS 16) and one end of Chr 11 (ChrS 08) by good coverage of Oxford reads (purple). BAC 024L10 and 040C11 showed hits on Chr 11 and Ψ 0 (ChrS 08); (B) BAC 024L10 (red) and 040C11 (yellow) labeled ChrS 16 instead of ChrS 08, indicating a mis-assembly in the BioNano map

Finally, we pooled 101 anchored BACs corresponding to the 38 pseudomolecules or pseudomolecule fragments, to generate 20 *S. polyrhiza* chromosome-specific probes (Table 8). After several rounds of mcFISH on the same metaphase plates of *S. polyrhiza* clone 9509, all 20 probes were integrated and confirmed their unique chromosomal location (Fig. 19).

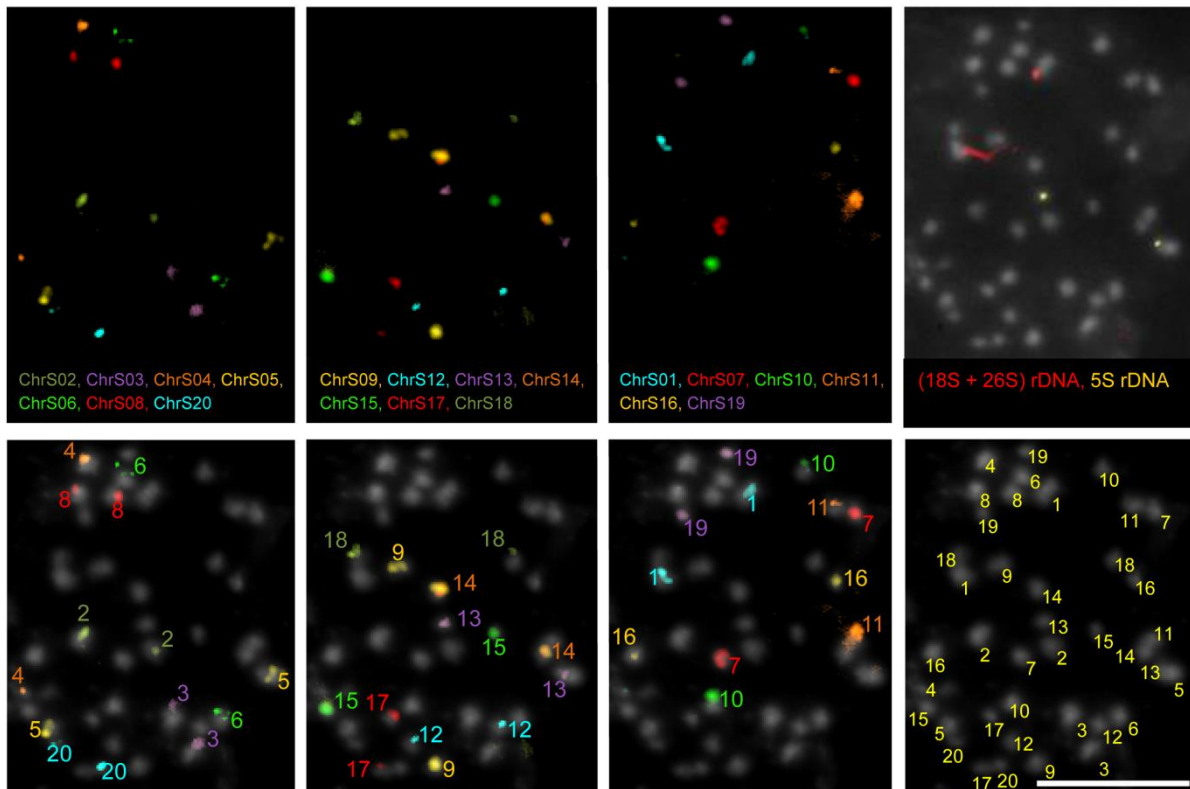


Figure 19. The complete karyotype of *S. polyrhiza* clone 9509.

Multicolor FISH with 20 chromosome-specific probes (Table 7) shows the predicted pseudomolecule linkage for each chromosome on the same metaphase plate. The upper panel indicates the probes used; the lower panel enumerates the chromosome pairs. The right upper image shows signals for (18S+26S) rDNA (red) on ChrS 01 and for 5S rDNA (yellow) on ChrS 13. The chromosomes were counterstained with DAPI. Scale bar = 5 μ m.

Overall, the discrepancies between two genome maps were solved, a further mis-assembly was confirmed and a validated chromosome map for *S. polyrhiza* has been provided by combination of FISH results including 106 fingerprinted BACs with Oxford Nanopore-derived sequence data. Therefore, although FISH is a laborious and time consuming technique, which requires specific expertise, its important role in genome assembly validation is unreplacable, particularly when a genetic map is unavailable or difficult to obtain. In any case, more than two independent approaches are required to produce a robust genome map for a higher plant species.

3.4. Karyotype evolution between the two species of the ancient duckweed genus *Spirodela*

3.4.1. Chromosome homeology between *S. polyrhiza* and *S. intermedia*

The genus *Spirodela* comprises only two species, *S. polyrhiza* and *S. intermedia*. Both have a genome size of 160 Mbp. The karyotype of *S. intermedia* ($2n = 36$) comprises two chromosome pairs less than that of *S. polyrhiza* ($2n = 40$). Because of their different chromosome number, at least two chromosome rearrangements had to be assumed after separation of the two *Spirodela* species. In the simplest case these were either two translocations each combining two chromosomes of *S. polyrhiza* into one of *S. intermedia*, or *vice versa*, two chromosome pairs of *S. intermedia* experienced splitting; each split yielding two independent chromosomes of *S. polyrhiza*.

In order to define chromosome homeology and chromosome rearrangements between the two *Spirodela* species, 96 BACs were pooled as *S. polyrhiza* chromosome-specific probes based on the cytogenetic map of *S. polyrhiza* (Cao *et al.*, 2016) and labeled in different color. The 20 *S. polyrhiza* chromosome specific probes (ChrS01 – ChrS20) were used for cross-hybridization to *S. intermedia* chromosomes. All probes gave clear signals on *S. intermedia* chromosomes. Multicolor FISH revealed 15 probes that labeled one chromosome pair as in *S. polyrhiza*, while four probes (ChrS03, ChrS07, ChrS10 and ChrS14) labeled two, and one probe (ChrS06) labeled three chromosome pairs of *S. intermedia* (Fig. 20).

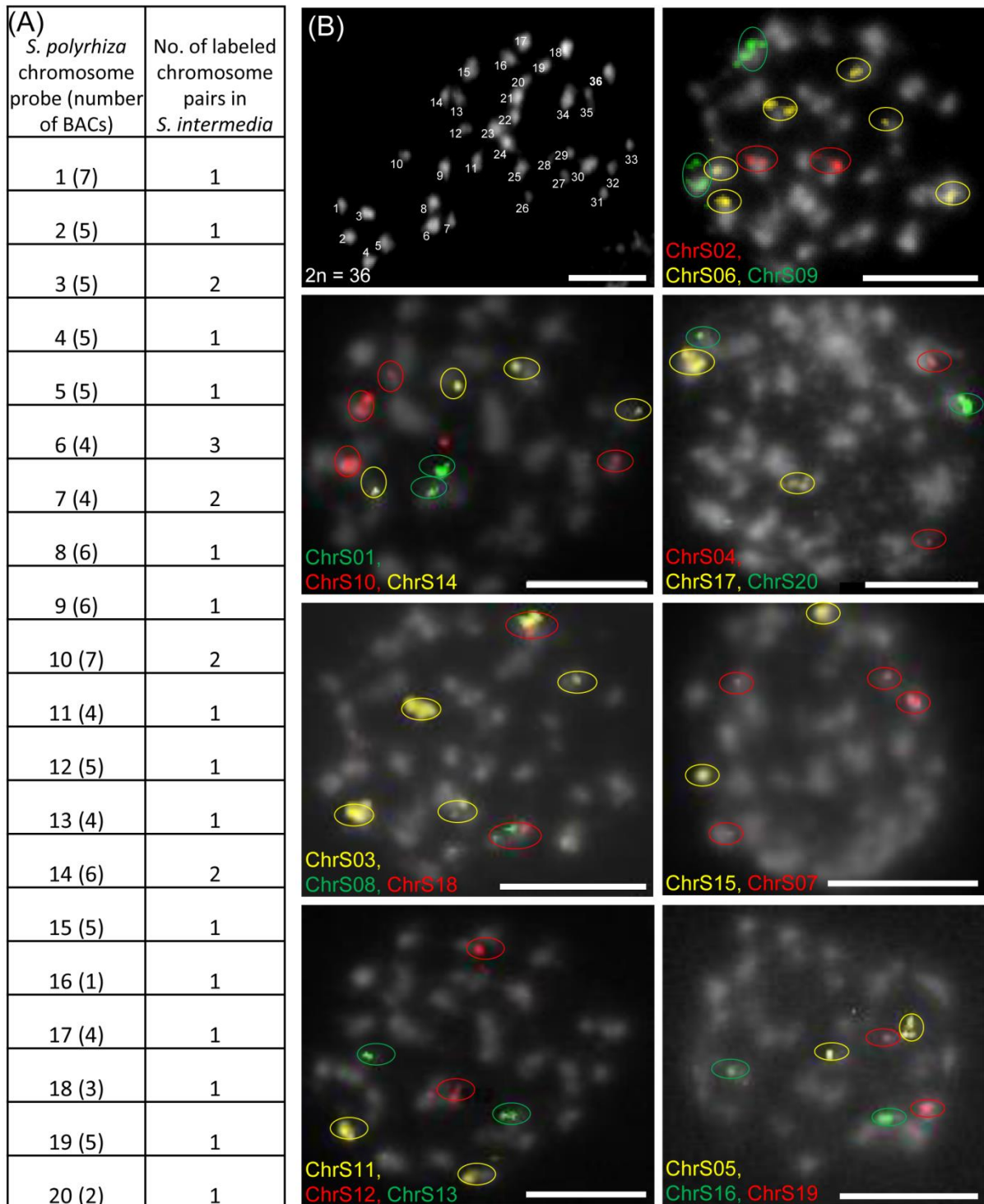


Figure 20. Multi-color FISH of 20 *S. polyrhiza* chromosome-specific probes to somatic metaphase chromosomes of *S. intermedia* (8410).

(A) Number of BACs in each *S. polyrhiza* chromosome-specific probe and number of labeled chromosome pairs in *S. intermedia*, (B) Different *S. intermedia* somatic metaphases with $2n = 36$ chromosomes and FISH signals of the 20 probes on *S. intermedia* chromosomes. Probes were labelled by Cy3 (yellow), Alexa-488 (green) and Texas-Red (red), the chromosomes were counterstained with DAPI. Scale bars = 5 μm

Thus, although having two chromosome pairs less, *S. intermedia* showed 12 FISH signal pairs more than *S. polyrhiza*. In case of a chromosome number reduction towards *S. intermedia*, this reduction should be based on complex rearrangements, and/or be accompanied by additional translocations or by sequence duplications. In order to get detailed information about these rearrangements, a three-step approach was applied. In the first step, a 3-color labeling scheme was used in order to discover in *S. intermedia* possible linkages between some of the 20 *S. polyrhiza* chromosomes. In the second step, single BACs of chromosome-specific probes, which labeled more than one chromosome pair in *S. intermedia*, were used separately to address whether the increased number of FISH signals is due to sequence duplication or to a split of *S. polyrhiza* linkage groups, and where are the break points related to the *S. polyrhiza* linkage groups. In the third step, the individual BACs from different pools of split *S. polyrhiza* linkage groups were combined to find out which BACs from which *S. polyrhiza* chromosome-specific probes are combined in *S. intermedia* chromosomes.

3.4.2. Six new linkage groups in *S. intermedia* were revealed by FISH

Subsequent series of mcFISH experiments with the 20 *S. polyrhiza* chromosome-specific BAC pools (step 1) uncovered ten chromosome pairs which are apparently identical between both species (ChrS01, 02, 04, 09, 11, 12, 13, 15, 19 and 20) and six new linkages in *S. intermedia* that differed from the *S. polyrhiza* karyotype: [ChrS03 – ChrS06 – ChrS14]; [ChrS05 – ChrS06]; [ChrS06 – ChrS07 – ChrS14]; [ChrS03 – ChrS17], [ChrS10 – ChrS16] and [ChrS08 – ChrS18] (Fig. 21). Additionally, two other chromosome pairs were labeled by (parts of) probes ChrS07 and ChrS10, respectively. The six linkages suggested six potentially rearranged chromosome pairs. The open questions from this result are: (1) Are the multiple FISH signals due to duplications, or are they caused by different BACs of a pool, and, if the latter is the case, (2) Which BACs belong to which signal pair. These two questions were solved in step 2 and 3.

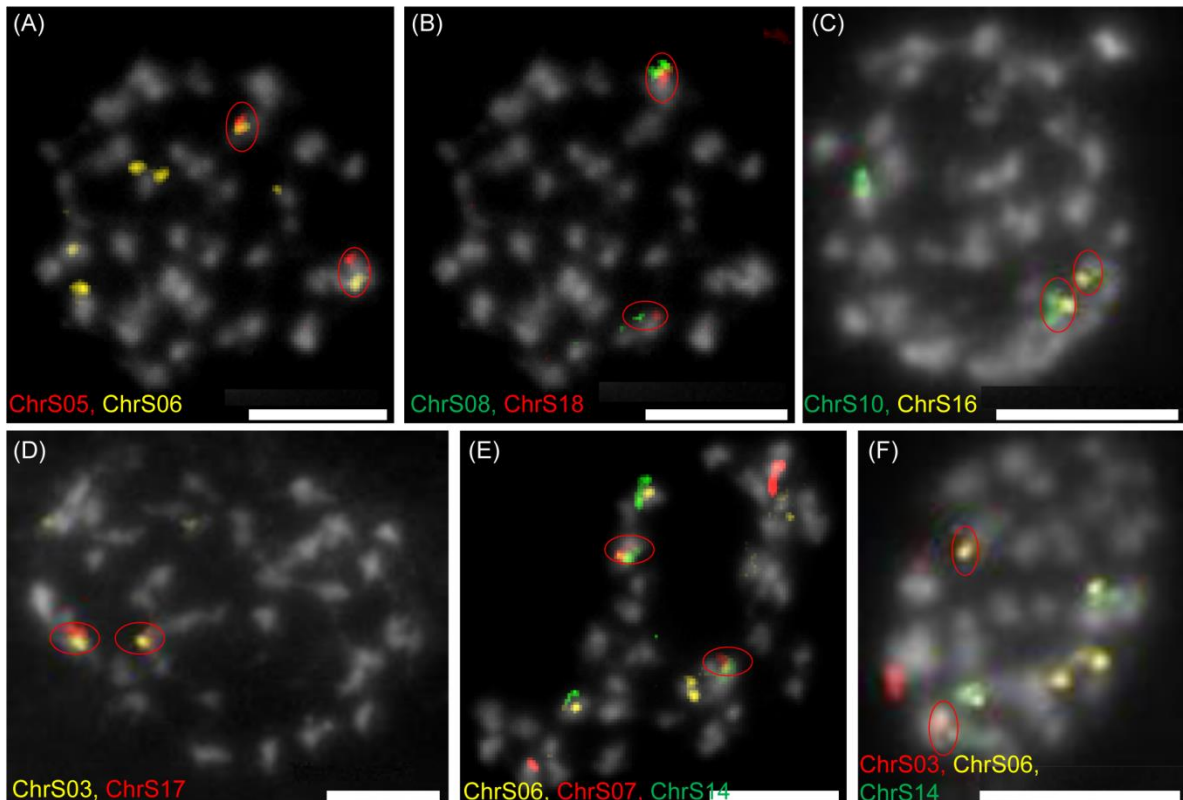


Figure 21. Six new linkage groups in *S. intermedia* are uncovered by subsequent mc-FISH.

(A) Two out of six signals of ChrS06 co-localized with ChrS05, (B) ChrS08 and ChrS18 formed one chromosome pair in *S. intermedia*, (C) linkage between ChrS16 and ChrS10, (D) linkage between ChrS03 and ChrS17, (E) linkage between ChrS06, ChrS07 and ChrS14, (F) linkage between ChrS03, ChrS06 and ChrS14. Probes were labelled by Cy3 (yellow), Alexa-488 (green) and Texas-Red (red), the chromosomes were counterstained with DAPI. Scale bars = 5 μm

In the second step, five *S. polyrhiza* chromosome-specific probes which labeled more than one chromosome pair of *S. intermedia* (ChrS 03, 06, 07, 10 and 14) were investigated. To address whether more than two FISH signals per probe was due to a split of the corresponding chromosome or a duplication of the chromosome or a part of it, separate BAC probes of each chromosome, according to their position within the cytogenetic map of *S. polyrhiza* (Cao *et al.*, 2016), were used for sequential mcFISH experiments. Except for three BACs which gave no signals, all individual BACs yielded only one pair of signals. These results suggested chromosome breakage and translocations rather than duplications, which should have yielded more than two signals per BAC. The chromosomal position of the distinct BAC signals in *S. intermedia* indicated the split position within the original sequence order of the corresponding *S. polyrhiza* chromosome.

For instance, the probe ChrS03 (comprising 5 BACs) labeled two chromosome pairs in *S. intermedia*. In the first round of mcFISH, BAC 021D24 labeled one chromosome pair, while BACs 002K06 and 010J02 co-localized on another chromosome pair. The result from the second round showed that BAC 002E10 co-localized with 021D24, while BAC 011P20 appeared on the same chromosome pair as BACs 002K06 and 010J02. Thus, the break-point occurred between BACs 002E10 and 002K06, dividing chromosome Sp10 into two parts [021D24, 002E10] and [002K06, 011P20, 010J02] in *S. intermedia* (Fig. 22A).

The same approach was applied for the other four *S. polyrhiza* chromosome-specific probes that labeled more than two chromosome pairs in *S. intermedia* and led to the following conclusions: (1) In case of ChrS06, BAC 034P03 has to be excluded because it yielded no FISH signals on *S. intermedia* chromosomes. Furthermore, an inversion of a fragment comprising BACs 002I04, 035L19, 033A10 occurred within ChrS06 before it was split into 3 parts: [023I01, 033A10], [035L19] and [002I04] (Fig. 22B); (2) BAC 013C18 has to be excluded from ChrS07 because of no FISH signals. Furthermore, a break divided this chromosome into two parts [014F03, 035P21] and [004A04, 030B11] (Fig. 22C); (3) A break-point occurred between BACs 008G11 and 020G09, dividing ChrS10 into two parts ([035P14, 002B12, 037K21, 008G11] and [020G09, 028B20, 026G16]) (Fig. 22D); (4) In case of ChrS14, two breaks between 040G15 and 037I18, and between 006D12 and 006A07 split ChrS14 into three parts [003B08, 036F14, 040G15], [037I18, 006D12] and [006A07], followed by fusion of [006A07] with [003B08, 036F14, 040G15]. Alternatively, an inversion of a fragment comprising BACs 037I18, 006D12 and 006A07 occurred before a break between 006A07 and 006D12 split ChrS14 into two parts [003B08, 036F14, 040G15, 006A07] and [006D12, 037I18] (Fig. 22E); (5) BAC 011H02 had to be excluded from ChrS16 because it yielded only weak dispersed FISH signals.

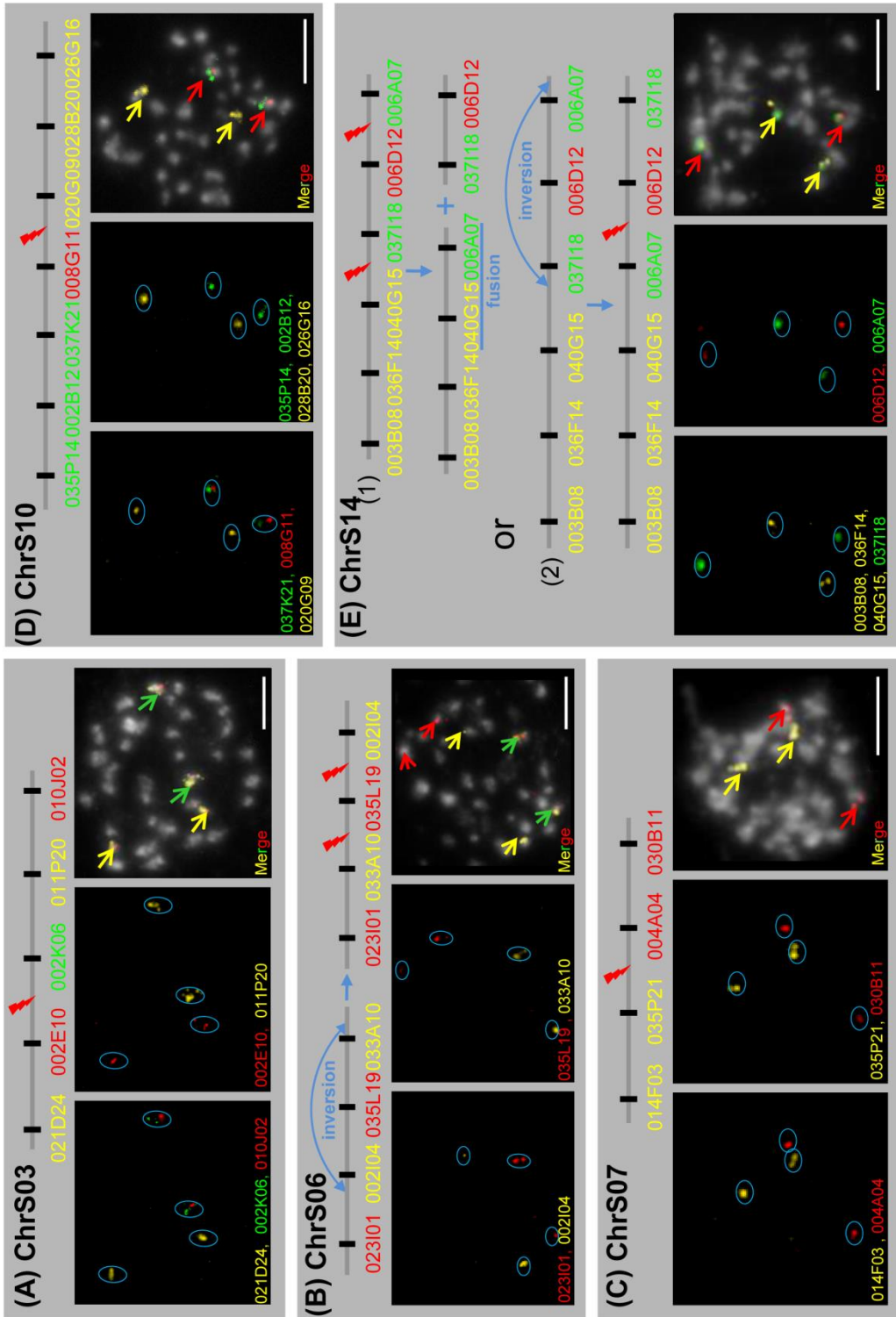


Figure 22. Three-color FISH on *S. intermedia* using single BACs from *S. polyrhiza* chromosome-specific probes that label more than one on *S. intermedia* chromosome to define the split-points. (A) a break between BACs 002E10 and 002K06 within ChrS03, (B) an inversion occurred within ChrS06 before it was split into 3 parts (023I01,033A10), 035L19 and 002I04, (C) a break between BACs 035P21 and 004A04 divided ChrS07 into two parts, (D) a break between BACs 008G11 and 020G09, divide ChrS10 into two parts, (E) two breaks between 040G15 and 037I18, and between 006D12 and 006A07 split ChrS14 into three parts, followed by a fusion of 040G15 and 006A07, or an inversion occurred before ChrS14 was split into 2 parts (003B08, 036F14, 040G15, 006A07) and (037I18, 006D12). Probes were labelled by Cy3 (yellow), Alexa-488 (green) and Texas-Red (red), the chromosomes were counterstained with DAPI. Scale bars = 5 μ m

In the third step we combined in serial mcFISH experiments individual BACs from the 20 *S. polyrhiza* chromosome-specific probes that co-localized on distinct chromosome pairs of *S. intermedia* to determine which parts of *S. polyrhiza* linkage groups are combined in individual *S. intermedia* chromosomes. For instance, in the [ChrS03 – ChrS06 – ChrS14] linkage, we combined on different slides differently labeled individual BACs belonging to two of the split *S. polyrhiza* linkage groups. The result from one slide after two rounds of mcFISH showed that BAC 035L19 of ChrS06 co-localized to BAC 002K06 of ChrS03. In another slide, two BACs representative for two parts of ChrS14 were used, then this slide was re-hybridized by BACs 035L19 and 002K06 to identify which BAC of ChrS14 was included in the linkage. It turned out that six BACs of three *S. polyrhiza* chromosomes: 002K06, 011P20, 010J02 (ChrS03), 035L19 (ChrS06) and 037I18, 006D12 (ChrS14) contributed to one rearranged *S. intermedia* chromosome (Fig. 23A). The same procedure was applied for the other five linkages. By this approach all BACs contributing to each of the rearranged linkage groups of *S. intermedia* could be identified, however in some cases the orientation of BACs within the *S. intermedia* chromosomes remained unclear. In detail, the ChrS06 – ChrS07 – ChrS14 linkage included BACs 002I04 (ChrS06), 014F03 (ChrS07) and 003B08, 036F14 (ChrS14) (Fig. 23B); the ChrS03 – ChrS17 linkage included BAC 002E10 (ChrS03) and entire ChrS17 (Fig. 23C); the ChrS10 – ChrS16 linkage included BACs 008G11, 035P14, 002B12, 037K21 (ChrS10) and 009J15 (ChrS16) (Fig. 23D); the ChrS05 – ChrS06 linkage included entire ChrS05 and BACs 033A10, 023I01 (ChrS06) (Fig. 23E); and ChrS08 and ChrS18 formed one pair in *S. intermedia* (Fig. 23F)

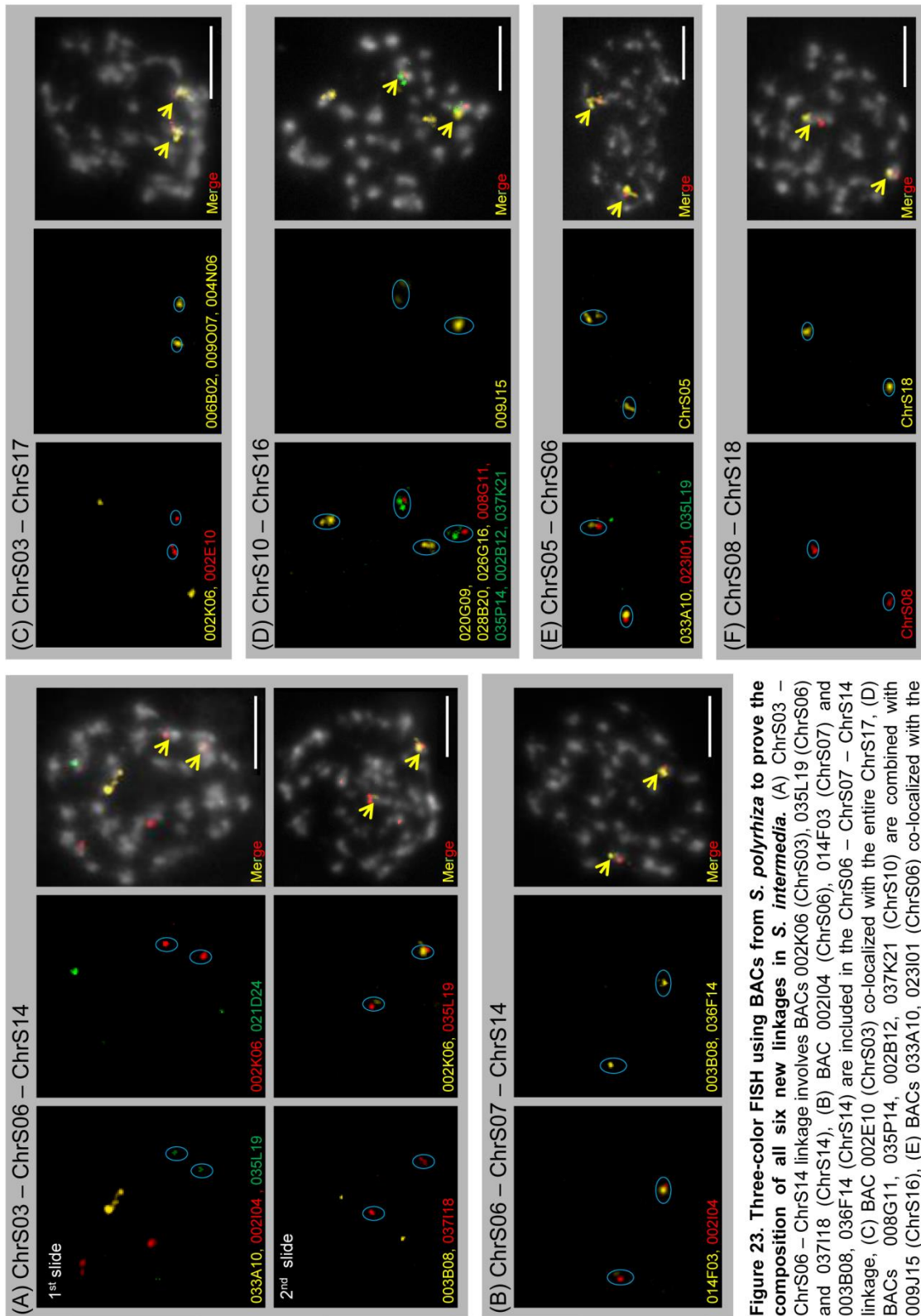


Figure 23. Three-color FISH using BACs from *S. polyrhiza* to prove the composition of all six new linkages in *S. intermedia*. (A) ChrS03 – ChrS06 – ChrS14 linkage involves BACs 002K06 (ChrS03), 035L19 (ChrS06) and 037I18 (ChrS14), (B) BAC 002I04 (ChrS06), 014F03 (ChrS07) and 003B08, 036F14 (ChrS14) are included in the ChrS06 – ChrS07 – ChrS14 linkage, (C) BAC 002E10 (ChrS03) co-localized with the entire ChrS17, (D) BACs 008G11, 035P14, 002B12, 037K21 (ChrS10) are combined with 009J15 (ChrS16), (E) BACs 033A10, 023I01 (ChrS06) co-localized with the entire ChrS05, and ChrS08 and ChrS18 form one pair in *S. intermedia*. Probes were labelled by Cy3 (yellow), Alexa-488 (green) and Texas-Red (red), the chromosomes were counterstained with DAPI. Scale bars = 5 μ m

Serial cross-hybridization using an optimized BAC pooling strategy, based on a BAC tiling path of *S. polyrhiza*, and applying suitable hybridization stringency, enabled detecting chromosome homeology and rearrangements in *S. intermedia*, for which no genomic and cytogenetic data were at hand. So far this approach was successful in Brassicaceae (Lysak *et al.*, 2006; Mandakova *et al.*, 2010; Mandakova and Lysak, 2008) and to a limited degree in Triticeae (Ma *et al.*, 2010).

3.4.3. Supposed karyotype evolution scenarios between two *Spirodela* species

Comparative serial multicolor cross-hybridization of the 20 *S. polyrhiza* chromosome-specific BAC pools and of individual BACs can uncover chromosome homeology in *S. intermedia* and suggest scenarios of karyotype evolution between them. Because the genus *Spirodela* comprises only two species at the phylogenetic basis of the *Lemnaceae*, it remains unknown which of the two has the more ancestral karyotype, or whether both extant karyotypes are equidistant to the ancestral one. Therefore different scenarios of karyotype evolution within the genus *Spirodela* are possible.

3.4.3.1. Karyotype evolution towards *S. intermedia* (n = 18)

If the ancestral karyotype was similar to that of *S. polyrhiza* (n = 20; ChrSp01-20), the evolution towards *S. intermedia* (n = 18; ChrSi01-18) requires as a minimum of rearrangements one inversion (ChrSp06) and six translocations (involving chromosomes Sp08 and 18, 10 and 16; 06, 03, 14, 05, 07 and 17). The results can be summarized as follows: (1) The karyotypes of both species share ten apparently identical chromosome pairs; (2) Chromosomes Sp08 and Sp18 form ChrSi09 of *S. intermedia*; (3) One part of ChrSp10 (035P14, 002B12, 037K21, 008G11) is translocated to ChrSp16 (009J15) forming ChrSi11 while the other part (020G09, 028B20, 026G16) stays separately as ChrSi12; (4) ChrSp06 experienced an inversion and then became split into three parts, one part (035L19) is combined with parts of ChrSp03 (002K06, 011P20, 010J02) and ChrSp14 (037I18, 006D12), as ChrSi03, the second part (023I01, 033A10) is translocated into ChrSp05, forming ChrSi06, the third part (002I04) is combined with the remaining ChrSp14 and a part of ChrSp07 (014F03, 035P21) forming ChrSi07; (5) The second part ChrSp07 (004A04, 030B11) remains separately as ChrSi08; (6) The second part of ChrSp03 (021D24, 002E10) is translocated to ChrSp17 forming ChrSi04 (Fig. 24).

3.4.3.2. Karyotype evolution towards *S. polyrhiza* (n = 20)

If the ancestral karyotype was similar to that of *S. intermedia*, at least 6 translocations (involving ChrSi03, 04, 06, 08, 11 and 12), and one fission (ChrSi09) were required during evolution towards the *S. polyrhiza* karyotype (Fig. 25). The scenario can be summarized as follows: (1) Ten chromosomes are conserved between both species; (2) Si03 is split into three parts, one part (002K06, 011P20,010J02) is combined with a part of Si04 (002E10, 021D24) forming Sp03, the second part (035L19) is combined with a part of Si07 (002I04) and Si06 (023I01, 033A10) as Sp06, the third part (037I18, 006D12) is translocated to the second part of Si07 (003B08, 036F14,040G15,006A07) as Sp14; (3) The second part Si04 (006B02, 009O07, 004N06, 009L02) remains separately as Sp17; (4) The remaining part of Si06 (030K22,027D17,009A15,005B18, 009O20) forms Sp05; (5) The remaining part of Si07 (014F03, 035P21) is translocated to Si08 (004A04, 030B11) forming Sp07; (6) Fission of Si09 formed Sp08 (013I04, 006P24, 032L08, 034K03, 004E01, 006L17) and Sp18 (026D06, 037B13, 029K19); and (7) One part of Si11 (035P14, 002B12, 037K21, 008G11) is translocated to Si12 (020G09, 028B20, 026G16) forming Sp10 while the other part (009J15) stays separately as Sp16.

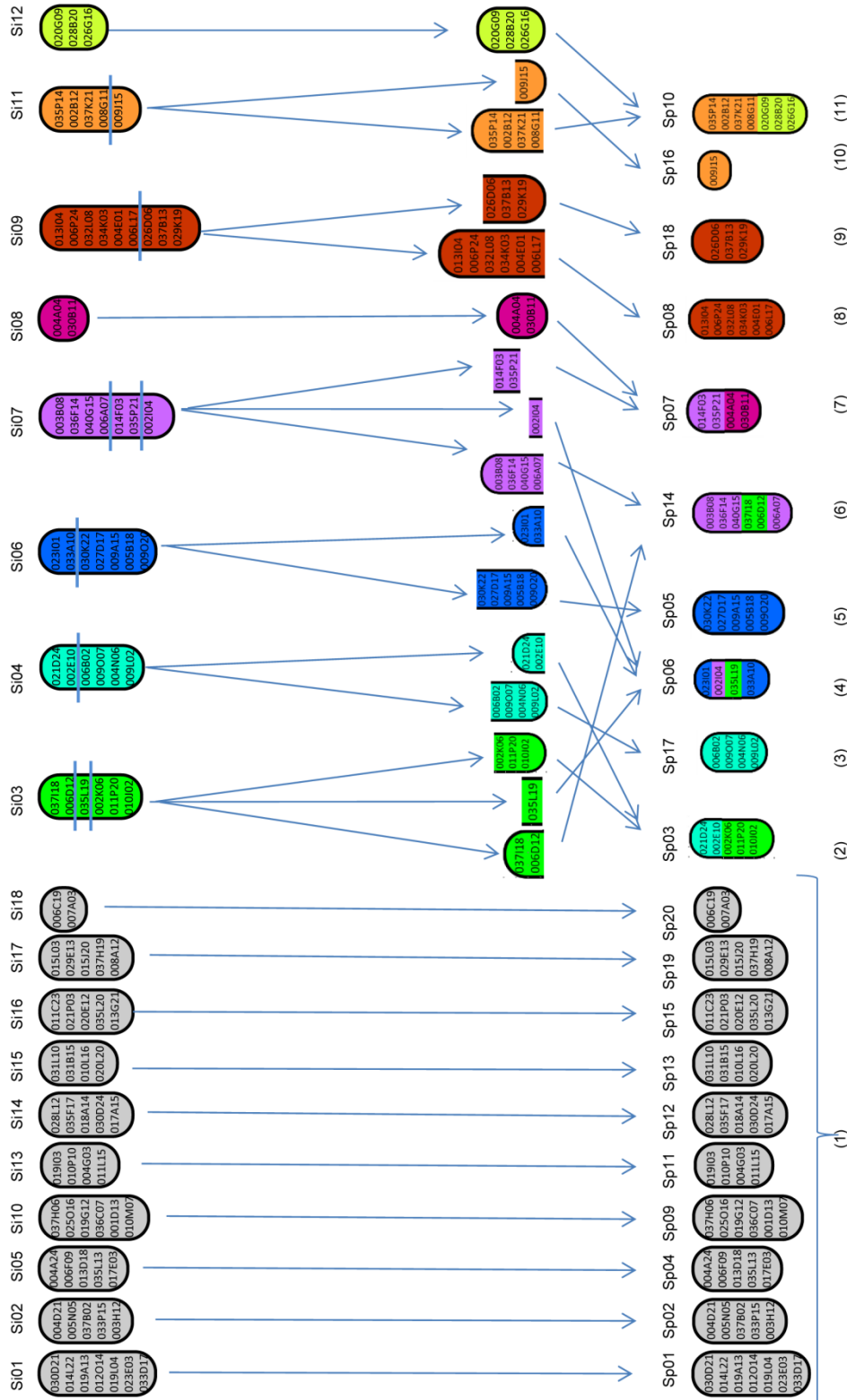


Figure 25. Karyotype evolution towards *S. polyrhiza* (n=20) in case the ancestral karyotype was similar to that of *S. intermedia* (n=18). (1) Ten chromosomes are conserved between both species; (2); (4); (6) Si03 split into three parts, one part (002K06, 01P20,010J02) is combined with a part of Si04 (002E10, 021D24) forming Sp03, the second part (035L19) is combined with a part of Si07 (002I04) and Si06 (023I01, 033A10) as Sp06, the third part (037I18, 006D12) is translocated to the second part of Si07 (003B08, 036F14,040G15,006A07) as Sp14; (3) the second part Si04 (006B02, 009O07, 004N06, 009L02) remains separately as Sp17; (5) the remaining part of Si06 (030K22,027D17,009A15,005B18, 009G02) formed Sp05; (7) the remaining part of Si07 (014F03, 035P21) is translocated to Si08 (004A04, 030B11) forming Sp07; (8); (9) fission of Si09 formed Sp08 (013I04, 006P24, 032L08, 034K03, 004E01, 006L17) and Sp18 (026D06, 037B13, 029K19); (10); (11) one part of Si11 (035P14, 002B12, 037K21, 008G11) is translocated to Si12 (020G09, 028B20, 026G16) forming Sp10 while the other part (009J15) stays separately as Sp16. *S. polyrhiza* and *S. intermedia* chromosomes size and the orientation of BACs within *S. intermedia* are only illustrative .

In case of an unknown ancestral karyotype of an intermediate structure, the number and type of rearrangements which led to the two extant karyotypes cannot precisely be determined.

3.4.4. Cytogenetic map of *S. intermedia*

Besides elucidating the evolutionary alterations between the karyotypes of the two investigated species, data from this study allow to distinguish each of the 18 *S. intermedia* chromosomes on the basis of their genetic content (Fig. 26). The components of the 18 *S. intermedia* chromosomes based on 93 anchored BACs of *S. polyrhiza* are shown in Table 9

Table 9: Components of the 18 *S. intermedia* chromosomes based on 93 anchored *S. polyrhiza* BACs.

ChrSi	BACs (Number of BACs)	ChrSi	BACs (Number of BACs)
01	030D21; 014L22; 019A13; 012O14; 019L04; 023E03; 033D17 (7)	10	037H06; 025O16; 019G12; 036C07; 001D13; 010M07 (6)
02	004D21; 005N05; 037B02; 033P15; 003H12 (5)	11	035P14; 002B12; 037K21; 008G11; 009J15 (5)
03	037I18; 006D12; 035L19; 002K06; 011P20; 010J02 (6)	12	020G09; 028B20; 026G16 (3)
04	021D24; 002E10; 006B02; 009O07; 004N06; 009L02 (6)	13	019I03; 010P10; 004G03; 011L15 (4)
05	004A24; 006F09; 013D18; 035L13; 017E03 (5)	14	028L12; 035F17; 018A14; 030D24; 017A15 (5)
06	023I01; 033A10; 030K22; 027D17; 009A15; 005B18; 009O20 (7)	15	031L10; 031B15; 010L16; 020L20 (4)
07	003B08; 036F14; 040G15; 006A0; 014F03; 035P21; 002I04 (7)	16	011C23; 021P03; 020E12; 035L20; 013G21 (5)
08	004A04; 030B11 (2)	17	015L03; 029E13; 015J20; 037H19; 008A12 (5)
09	013I04; 006P24; 032L08; 034K03; 004E01; 006L17; 026D06; 037B13; 029K19 (9)	18	006C19; 007A03 (2)

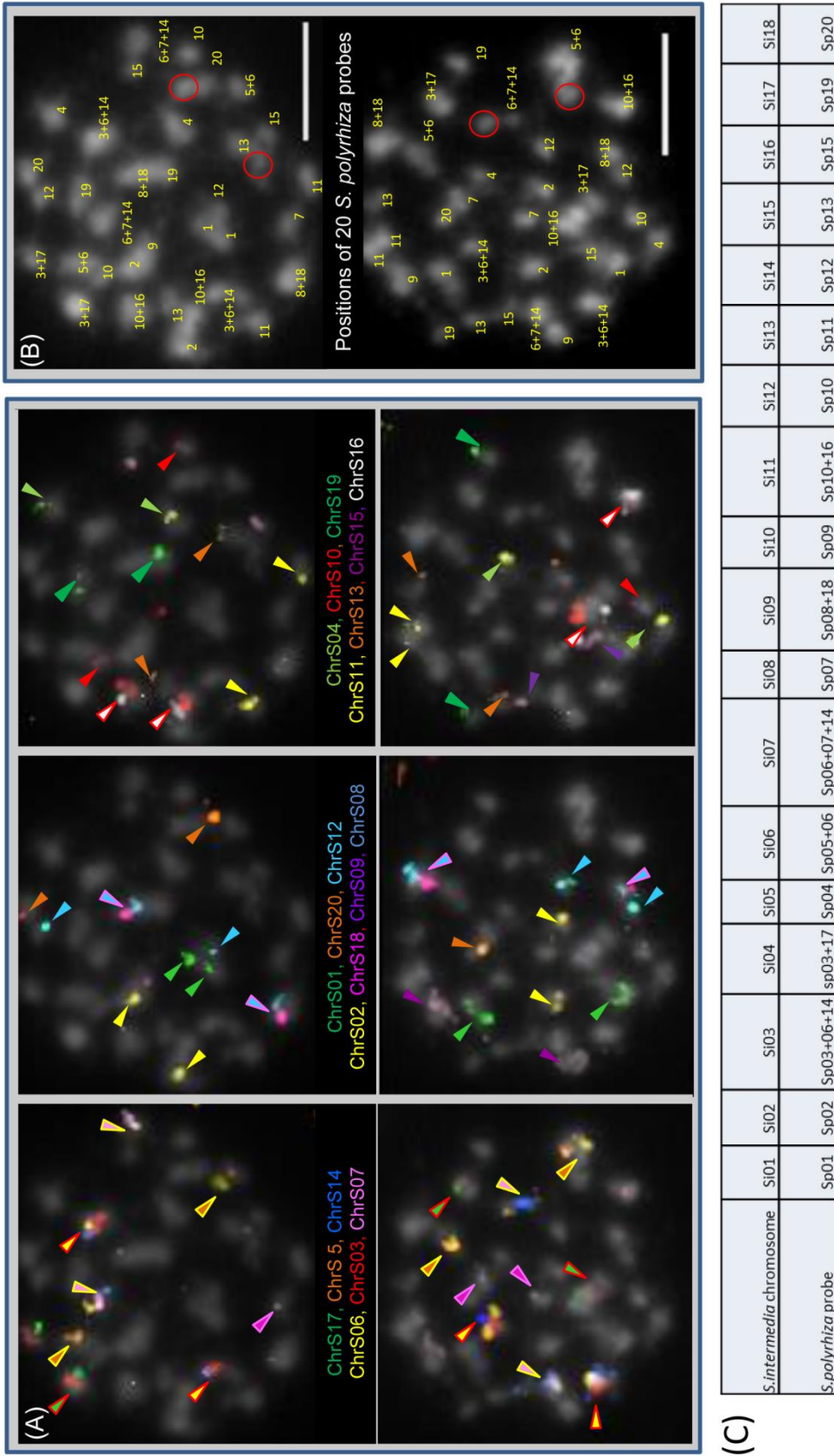


Figure 26. Distribution of 20 *S. polyrhiza* chromosome probes on *S. intermedia* metaphases.

(A) FISH signals of 20 *S. polyrhiza* chromosome-specific probes on two different *S. intermedia* metaphases confirmed 10 conserved chromosome pairs, the six new linkages (indicated by two-color arrow heads according to the false colors of signals), and parts of rearranged chromosomes that stay separately in *S. intermedia*, (B) Positions of 20 *S. polyrhiza* chromosome-specific probes are summarized: In the upper cell, labeling on the second chromosome that should be marked by ChrS07 and ChrS09 (empty red circles) are missing. These are present in the lower cell. In the lower cell the second chromosomes of pairs 10 and 20 are not labeled (empty red circles), but they appeared in the upper cell. Scale bars = 5 μm; (C) Enumeration of *S. intermedia* chromosomes and their labeling by *S. polyrhiza* chromosome-specific probes. Scale bars = 5 μm

3.5. Whole genome sequencing and genome assembly in *S. intermedia*

The cytogenetic map of *S. intermedia* with 93 anchored *S. polyrhiza* BACs (see 3.4.4) provides a useful framework for a whole genome sequence assembly in a next generation sequencing project. Sequence assignment to distinct chromosomes based on cross-hybridization of genomic sequences between related species represents a novel cytogenomic approach. Such an approach is particularly important for *S. intermedia*, which, as *S. polyrhiza*, is mainly vegetatively propagating and therefore it is difficult to obtain a genetic map. However, at least one reference data set for validating bioinformatic assembly efforts is absolutely required, and a considerable number of cytogenetic anchor points provide a reliable support for sequence data integration as previously exemplified for *Amborella trichopoda* (Chamala *et al.*, 2013) and for *S. polyrhiza* (Cao *et al.*, 2016). Moreover, a validated reference genome map of the sister species *S. polyrhiza* is available to support assembly of genomic sequence data from Pacbio sequence reads of the *S. intermedia* genome.

Two rounds of PacBio-sequencing of a 20 kb library of genomic *S. intermedia* DNA yielded 149 Gb raw reads. Initially, 1 305 064 reads of at least 500 nucleotides were assembled into 1172 sequence contigs of in total 147 613 042 nucleotides. All contigs of this draft assembly are covered in median 37.5-fold by raw reads.

These contigs were ordered into scaffolds using the genomes of clones 9509 and 7498 of the sister species *S. polyrhiza* as references (Cao *et al.*, 2016; Michael *et al.*, 2017). The resulting scaffolds (N50 = 11,362,824 bp) were super-scaffolded and assigned to the 18 chromosomes of *S. intermedia*, using 93 *S. polyrhiza* BACs as landmarks which were cross-hybridized to the *S. intermedia* chromosomes of strain 8410. Additional BAC probes covering the regions of interest in the *S. polyrhiza* genome were designed for FISH experiments to resolve mis-assemblies and to approve localization of the contigs within the pseudomolecules of *S. intermedia*.

By manual curation, the number of scaffolds could be reduced to 691, consisting of 881 contigs. After reiterative rounds of manual curation and validation by FISH, in the final genome assembly, 104 scaffolds comprising 188 contigs (N50 = 2.8 Mbp) of 134.1 Mbp in length could be assigned to distinct chromosomes. Most of the shorter and/or repetitive sequences (N50 = 27.1 Kbp) could not yet be assigned and were considered as additional pseudomolecule “0” (see Table 10).

Table 10: *S. intermedia* sequence assembly statistics.

INPUT DATA	
PacBio read coverage	37.5x
Chromosome number ^a	18
Genome physical size ^b	~160 Mbp
ASSEMBLY STATISTICS	
Assembly length	152.4 Mbp
Number of pseudomolecules	19 (including “ψ0”)
Number of contigs	881
Number of scaffolds	691
Number of chromosomally assigned scaffolds	104 scaffolds (featuring 188 contigs) ~134.1 Mbp total length
Largest contig	8.89 Mbp
N50 contig length	2.80 Mbp
G+C content	41.4%
Preliminary number of predicted gene models	19 190
COMPLETENESS OF GENE PREDICTION (BUSCO)	
Complete genes (C)	959 (66.6%)
Complete and single-copy (S)	926 (64.3%)
Complete and duplicated (D)	33 (2.3%)
Fragmented genes (F)	145 (10.1%)
Missing genes (M)	336 (23.3%)
No. of genes used (n)	1440

^{a)}determined by cytogenetic approaches ^{b)} measured by FCM

Gene prediction: Gene model prediction via Gene Model Mapper - GeMoMa (Keilwagen *et al.*, 2016) revealed in total of 17 470 putative high-quality as well as 1 770 low quality protein-coding genes, based on homology to *S. polyrhiza* 7498 v3.1 (Cao *et al.*, 2016), *Lemna minor* 5500 (Van Hoeck *et al.*, 2015), and *Oryza sativa* IRGSP v1.0.38 as reference genomes. CDS (coding sequence) models, which are either 3' or 5' truncated are considered as 'low-quality'.

The quality of the *S. intermedia* genome assembly was assessed by the BUSCO program (Simao *et al.*, 2015; Waterhouse *et al.*, 2017) with an Embryophyta reference dataset (Table 10). Fig. 27 showed BUSCO results of *S. intermedia* (clone 7747) assembly, *S. polyrhiza* (clone 9509) Oxford Nanopore sequences assembly (Hoang *et al.*, 2018) and *La. punctata* (clone 7260) assembly (kindly provided by Prof. Jiri Marcas' group, Institute of Plant Molecular Biology, C. Budejovice, Czech Republic). BUSCO assessment results showed that many genes were absent in *S. intermedia* (336 genes), *S. polyrhiza* (202 genes) and *La. punctata* (244 genes). Our hypotheses to explain for missing genes are (1) The collection of candidate genes in BUSCO (mostly from land plants) might not be optimal for aquatic plants and (2) During evolution towards neoteny duckweeds lost unnecessary genes (Wang *et al.*, 2014).

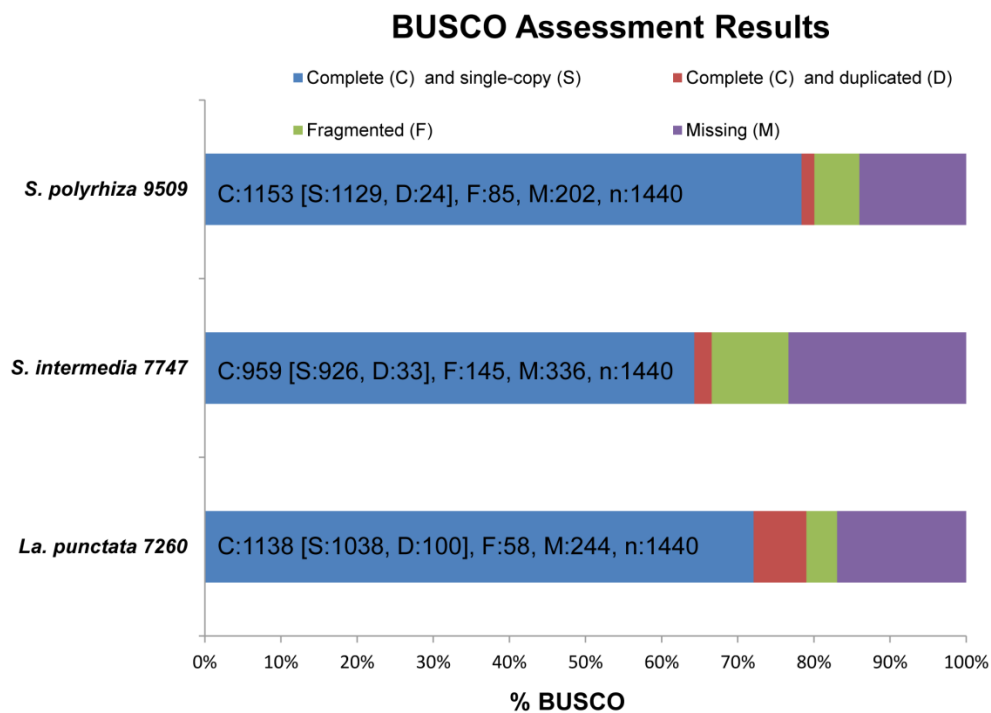


Figure 27. BUSCO assessment results.

Genome assemblies of *S. polyrhiza* (clone 9509), *S. intermedia* (clone 7747) and *La. punctata* (clone 7260) were assessed by the BUSCO Embryophyta reference dataset comprising 1440 genes

Further analyses of the *S. intermedia* genome regarding finalized gene prediction, functional gene annotation, identification and quantification of transposable elements and tandem repeats are still in progress.

3.6. Polyploidy in duckweeds

Between duckweed genera, but also between species of the genera *Lemna*, *Wolffiella* and *Wolffia*, and even between clones of individual species considerable genome size difference have been reported. This raises the question which mechanisms are responsible for the observed genome size variation.

According to Geber (1989) e.g. *Le. aequinoctialis* displays different ploidy levels, for instance 42 chromosomes were counted for clones 7021 and 7737, while 84 chromosomes were counted for clone 6746. In order to test the ploidy level, genome size, cell and nuclear volume, chromosome as well as 5S and 45S rDNA loci numbers were investigated for clones 6746 and 2018, of in *Le. aequinoctialis* (Table 11). Although chromosome preparation turned out to be difficult for this species, the result revealed not only a doubled genome size (900 Mbp/1C) but most likely also a doubled chromosome number ($2n = \sim 80$) for clone 6746 with larger cell and nuclear volume compared to clone 2018, suggesting that clone 6746 is tetraploid (Table 11) (Hoang *et al.*, submitted). One locus of 5S and 45S each were detected on *Le. aequinoctialis* (clone 2018) while three loci of 45S and two loci of 5S were detected on *Le. aequinoctialis* (clone 6746) (Fig. 28). More than doubling the 45S rDNA loci (three instead of the expected two loci) does not contradict WGD, but requires an additional duplication or amplification event for the original 45S rDNA locus in clone 6746.

Interestingly, *Le. aequinoctialis* clone 9433 displayed a genome size of 366 Mbp and clone 9526 nearly the double (711 Mbp) (Fig. 28B). Ongoing genotyping-by-sequencing should elucidate whether these two clones actually belong to *Le. aequinoctialis* or represent another species.

Vunsh *et al.* (2015) studied differences on frond size, weight, rhizoid number and length and growth rate between *La. punctata* wild type and its auto-tetraploid mutant (Vunsh *et al.*, 2015). In the present study chromosome number, 5S and 45S rDNA loci and genome size were tested for *La. punctata* wild type (clone 5562) and its colchicine-induced tetraploid mutant line (A4) (Table 11). The results ($4n = 2x = 92$ and 922 Mbp/1C) (Hoang *et al.*, submitted) as well as 4 loci of 5S rDNA and 2 loci of 45S rDNA confirmed the tetraploid character of clone 5562_A4 (Fig. 29).

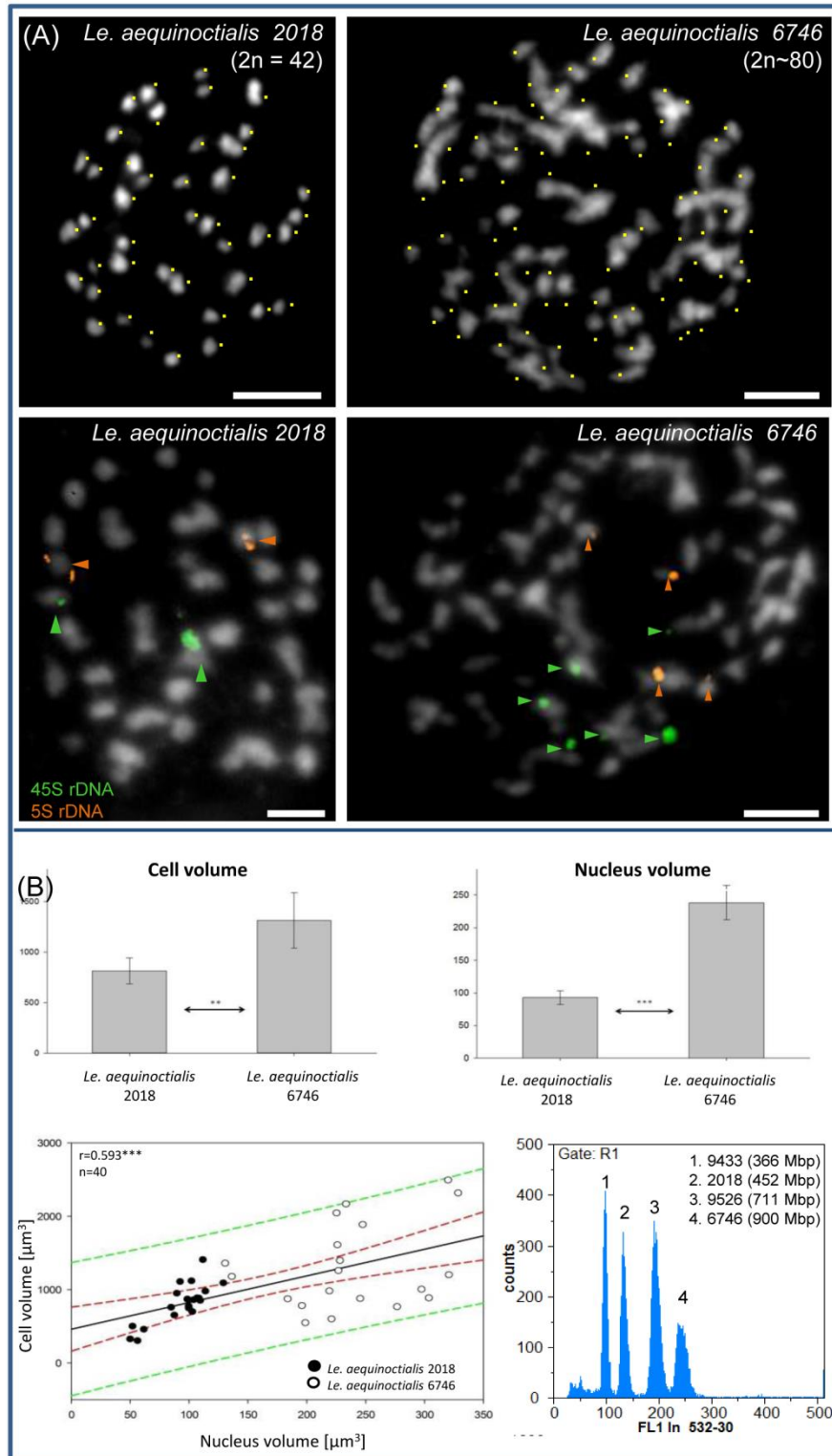


Figure 28. Chromosome, 5S and 45S loci number (A) and correlation of guard cell parameters (B) in diploid and tetraploid clones of *Le. aequinoctialis*.

(A) Mitotic spreads (above) and 5S and 45S rDNA loci (below) of *Le. aequinoctialis* clones 2018 and 6746. Scale bars: 5 μm ; (B) Correlation of guard cell parameters between diploid and tetraploid clones. ** $p < 0.01$, *** $p < 0.001$; regression line (black); 95% confidence interval (red) and 95% prediction interval (green). Bar charts: Error bars: 95% confidence interval; double arrows: result of mean value comparison by t-test. Variation of genome size of four *Le. aequinoctialis* clones as revealed by flow cytometry (bottom right)

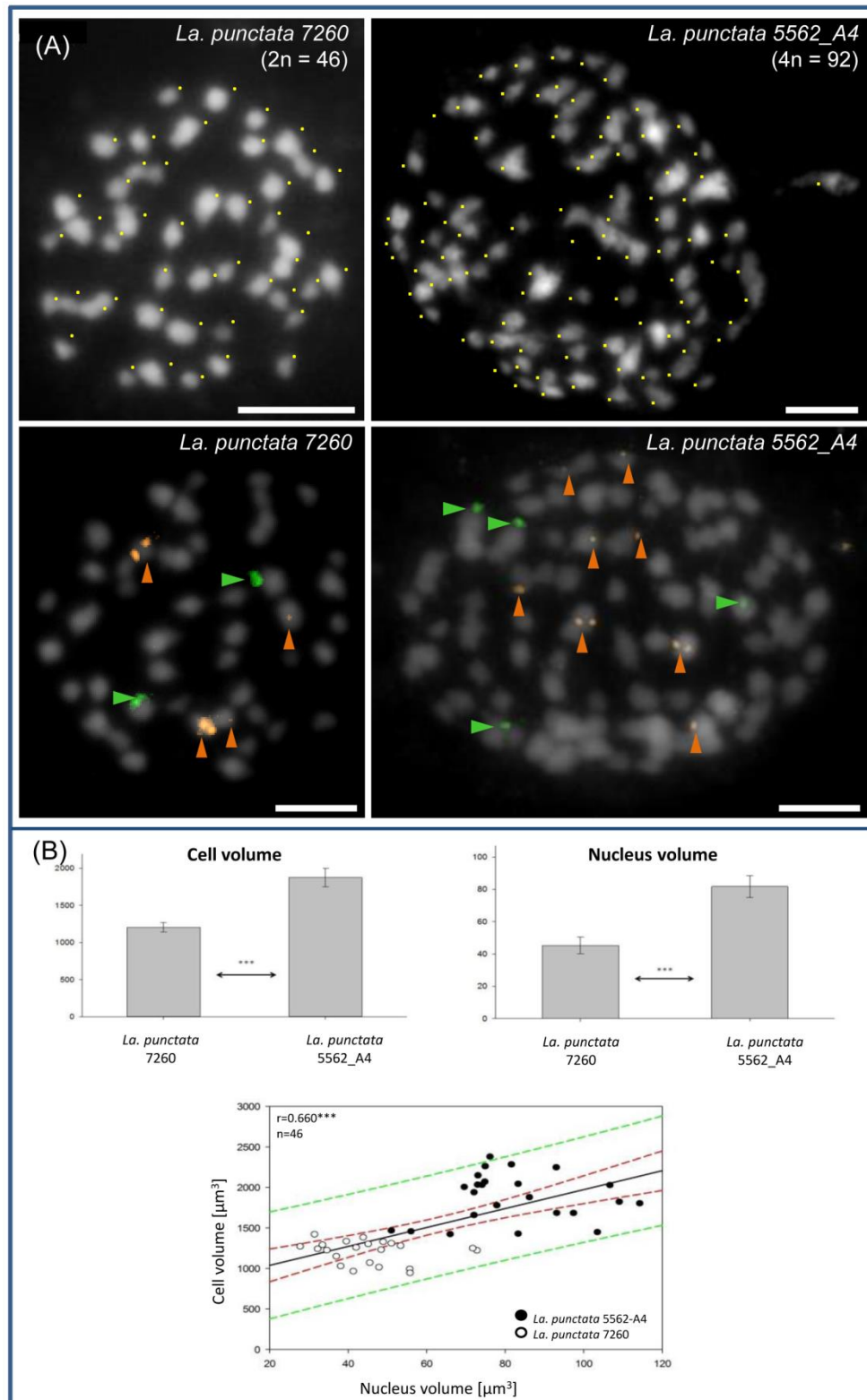


Figure 29. Chromosome, 5S and 45S loci number (A) and correlation of guard cell parameters (B) in diploid and tetraploid clones of *La. punctata*.

(A) Mitotic spreads (above) and 5S/45S rDNA loci (below) of *La. punctata* clone 7260 and the tetraploid mutant clone 5562_A4. Scale bars: 5 μm ; (B) Correlation of guard cell parameters between diploid and tetraploid clones. *** $p < 0.001$; regression line (black); 95% confidence interval (red) and 95% prediction interval (green). Bar charts: Error bars: 95% confidence interval; double arrows: result of mean value comparison by t-test.

Table 11: Cytological characterization of *La. punctata* clones 7260 and 5562_A4 and *Le. aequinoctialis* clones 2018 and 6746.

Genus	<i>Landoltia</i>		<i>Lemna</i>	
Species	<i>punctata</i>		<i>aequinoctialis</i>	
Clone ID	7260	5562_A4	2018	6746
Origin	Australia	Israel	Japan	USA
DNA content (pg/2C)	0.866± 0.012	1.885 ± 0.007	0.925 ± 0.003	1.841 ± 0.008
Genome size (Mbp/1C)	424 ± 6	922 ± 4	452 ± 2	900 ± 4
2n	46	92	42	~80
No. 5S rDNA loci	2	4	1	2
No. 45S rDNA loci	1	2	1	3
Cell volume (µm ³)	1204.1 ± 141.3	1905.1 ± 290.3	812.9 ± 275.8	1313.3 ± 588.3
Nuclear volume (µm ³)	45.3 ± 11.8	80.2 ± 15.9	92.9 ± 21.9	238.2 ± 56.4
% Nuclear to cell volume	3.8 ± 1.2	4.3 ± 0.9	12.1 ± 2.5	21.3 ± 9.5

Error: Standard deviation

The question to what degree the large genome size differences between duckweed genera and between species within the genera *Lemna*, *Wolffiella* and *Wolffia* are based on polyploidy cannot unambiguously be solved by chromosome counting and genome size measuring alone. More remote WGD could be blurred by dysploid chromosome number alteration and sequence deletion, diploidizing the originally polyploid genomes into meso- or even paleopolyploid ones. Detection of mesopolyploidy at interspecific level would require cross-FISH with BACs of a related species as done for mesoploid Australian Camelinae species by Mandakova *et al.* (2010), or comparative counting of FISH signals for several large (>10 kb) single copy sequences, as done for proving neopolyploidy in *Genlisea* species (Vu *et al.*, 2015). For this purpose cross-FISH with single copy BACs of *S. polyrhiza* was applied on *La. punctata* (clone 7260). Unexpectedly, some BACs yielded several dots which did not allow distinguishing between real signals and background noise (Fig. 30), while no signal were detected from other BACs, even under low stringency conditions (Table 12).

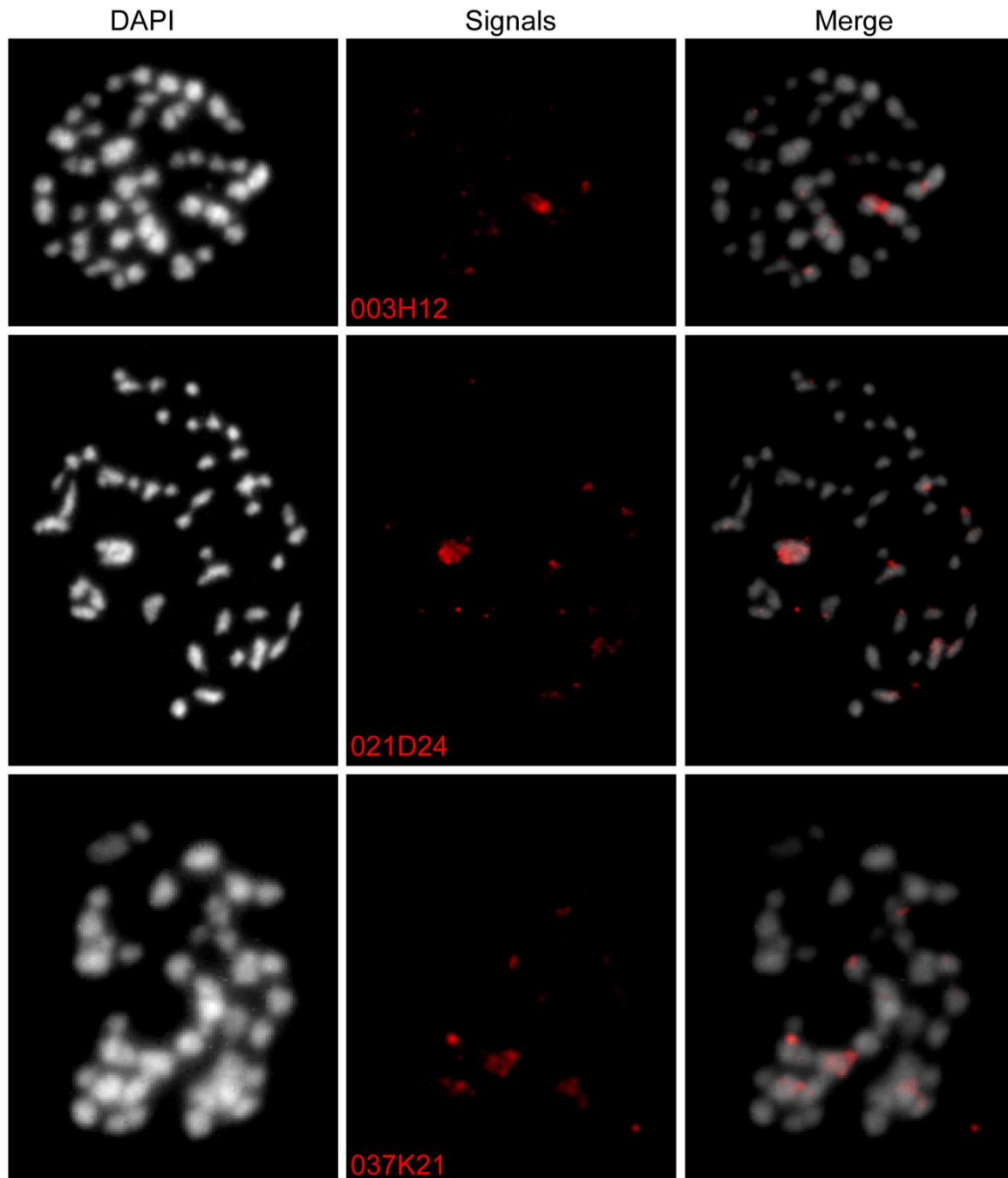


Figure 30. Cross-FISH with single copy BACs of *S. polyrhiza* on mitotic spreads of *La. punctata* (clone 7260).

FISH signals of three single copy *S. polyrhiza* BACs yielded many dots which could not be distinguished from background.

Table 12: Results of cross-FISH on *La. punctata* (clone 7260).

<i>S. polyrhiza</i> BACs	Results
003H12; 021D24; 037K21; 011P20; 008G11.	Signals not distinguishable from background
030D21; 014L22; 019A13; 023I01; 005N05; 033P15; 023I01; 031L10; 009O07; 004N06; 030K22; 009A15; 006L17; 034K03; 012O14; 019L04; 032E03; 033D17; 006C19; 007A03;	No clear signals at all

Because of the unexpected poor cross-FISH results, it remains unclear whether the evolutionary genome size increase in the phylogenetically younger duckweeds goes back to WGDs or rather to a burst of repetitive elements, until more genomic data for the species in question are available.

4. SUMMARY

Duckweeds are considered as potential future aquatic crops, due to a unique combination of features such as: worldwide distribution, floating life-style, fast growth, high starch and protein content, and the ability to efficiently absorb nitrate, phosphate and other contaminating components of water bodies. In addition, a successive reduction of morphological structures from the ancestral genus *Spirodela* towards the more derived genera *Lemna*, *Wolffiella* and *Wolffia* is accompanied by a stepwise reduction in frond size and a parallel increase in biodiversity (number of species), in genome size and genome size variability. These features make duckweeds becoming an interesting subject for genome and karyotype evolution studies.

Genome size, nuclear and cell volume are correlated in duckweeds

Genome size measurements data from eleven tested species of the five duckweed genera revealed a ~14 fold range between duckweed species (from 160 Mbp in *S. polyrhiza* to 2203 Mbp in *Wo. arrhiza*). In spite of few deviations between individual species, in general a positive correlation between genome size and cell and nuclear volume was found for the eleven tested duckweed species: The higher the genome size, the bigger were the cell and nuclear volume. In addition, cytomixis was observed between sister guard cells in *Wa. hyalina* and *Wo. australiana*.

Chromosome counts and rDNA loci did not provide a cue regarding the reasons of genome size variability between duckweed species

The number of chromosomes and of 5S and 45S rDNA loci varied independent of genome size for the eleven tested species. Previously reported chromosome number variation between natural clones of the genera *Spirodela* and *Landoltia* could not be confirmed. Each species displayed its own (basic) chromosome and 5S/45S rDNA loci numbers. However the variable intensity of FISH signals suggested variation in copy numbers, as reflected by the unusually low rDNA copy numbers reported for *S. polyrhiza*. The unexpected poor cross-hybridization of *S. polyrhiza* BACs on chromosomes of other duckweed genera did so far not allow distinguishing between polyploidisation and burst of mobile elements as reason for the evolutionary genome expansion in the phylogenetically younger duckweed genera.

A robust genome map reference is made available for S. polyrhiza

Five major differences between the two genome maps for *S. polyrhiza* presented by Cao *et al.* (2016) and Michael *et al.* (2017) were cleared by using subsequent mcFISH with 106 *S. polyrhiza* BACs. The results revealed no clone specific chromosome rearrangement and resolved the discrepancies which were due to on errors or incompleteness in former studies. Cytogenomic studies along with integration of Oxford Nanopore-derived sequence data, enabled full resolution of previous discrepancies between two genome maps for *S. polyrhiza*, provided a validated chromosome map for comparative genome studies in other duckweed species and demonstrated that more than two independent approaches are required to generate high-quality genome maps from organisms without available genetic maps.

In particular, the labor and cost intensive FISH technique turned out to be important for validation and chromosomal integration of results from NGS datasets because of its ability to uncover linkage between sequence contigs independent of the physical distance and of the presence of repetitive sequences. Especially if comprehensive genetic maps are not available because of asexual propagation, as in *S. polyrhiza* and other Lemnaceae species. FISH is the method of choice for this purpose.

Chromosome homeology elucidates evolutionary chromosome rearrangements between S. polyrhiza and S. intermedia

Sequential comparative mcFISH with 20 *S. polyrhiza* chromosome-specific probes (including 93 BACs) revealed chromosomes homeology between *S. intermedia* ($2n = 36$) and *S. polyrhiza* ($2n = 40$). Ten *S. polyrhiza* chromosome pairs proved to be conserved between the two species; the other ten chromosomes were involved in six chromosome rearrangements. Because the genus *Spirodela* comprises only two species at the phylogenetic basis of the Lemnaceae, it remains unknown which of the two has the more ancestral karyotype, or whether both extant karyotypes are equidistant to the ancestral one. Therefore different scenarios of karyotype evolution within the genus *Spirodela* are possible. If the ancestral karyotype was similar to that of *S. polyrhiza*, the evolution towards *S. intermedia* ($n=18$) requires at least one inversion (ChrSp06) and six translocations (involving chromosomes Sp08 and 18, 10

and 16; 06, 03, 14, 05, 07 and 17). If the ancestral karyotype was similar to that of *S. intermedia*, at least 6 translocations (involving ChrSi03, 04, 06, 08, 11 and 12), and one fission (ChrSi09) were required during evolution towards the *S. polyrhiza* karyotype.

The cytogenetic map informed S. intermedia genome assembly

The cytogenetic map of *S. intermedia* with 93 anchored *S. polyrhiza* BACs served as framework for chromosomal integration of 1 305 064 trimmed PacBio sequence reads (each >500 nucleotides). These reads were primarily assembled into 1172 sequence contigs of in total 147 613 042 nucleotides with a 37.5-fold coverage. In a novel approach of iterative manual curation reducing the number of contigs and validating of their assembly by FISH with homeologous BACs of the regions of interest, 104 scaffolds (corresponding 134.1 Mbp) could so far be assigned to distinct *S. intermedia* chromosomes. Via the 'gene model mapper' program GeMoMa (Keilwagen *et al* ., 2016) ~19 000 putative protein-coding genes were predicted at present, based on homology to *S. polyrhiza* 7498 v3.1, *Lemna minor* 5500, and *Oriza sativa* IRGSP v1.0.38 as reference genomes. The quality of the *S. intermedia* genome assembly was assessed by the BUSCO Embryophyta reference dataset comprising 1440 genes. The 23.3% missing genes are most likely owing to the aquatic life style and the neotenic organismic reduction of duckweeds. The final gene prediction, gene annotation as well as identification and quantification of dispersed and tandem repeats are still ongoing.

5. ZUSAMMENFASSUNG

Eine unikale Kombination von Merkmalen wie weltweite Verbreitung, Schwimmen an der Wasseroberfläche, schnelles Wachstum, hoher Stärke- und Proteingehalt und die Fähigkeit Nitrat, Phosphat und andere Wasserverunreinigungen effizient zu absorbieren machen Wasserlinsen zu potenziellen aquatischen Kulturpflanzen. Die von der ancestralen Gattung *Spirodela* hin zu den mehr abgeleiteten Gattungen *Lemna*, *Wolffiella* und *Wolffia* fortschreitende Neotanie (organismische Reduktion) und der parallele Anstieg der Biodiversität (Anzahl der Arten pro Gattung), der Genomgröße und der Genomgrößenvariabilität machen Wasserlinsen zu interessanten Objekten der Erforschung von Genom- und Karyotypevolution.

Genomgröße, Kern- und Zellvolumen von Wasserlinsen sind positiv korreliert

Genomgrößenmessungen für elf Arten der fünf Wasserlinsengattungen ergaben einen ~14-fachen Unterschied – von 160 Mbp/1C in *S. polyrhiza* zu 2203 Mbp/1C für *Wo. arrhiza*. Trotz weniger Ausnahmen zwischen einzelnen Arten gilt für die elf Arten: je größer das Genom, desto größer sind Kern- und Zellvolumen von Schließzellen. Darüber hinaus wurde für *Wa. hyalina* und *Wo. australiana* Zytomixis (Kernwanderung) zwischen geschwisterlichen Schließzellen beobachtet.

Chromosomenzahlen und rDNA Loci ergaben keinen Hinweis auf die Ursachen der Genomgrößenvariabilität zwischen Wasserlinsenarten

Die Anzahl der Chromosomen und der 5S und 45S rDNA Loci variiert für die elf getesteten Arten unabhängig von der Genomgröße. Frühere Berichte über Chromosomenzahlvariationen zwischen natürlichen Klonen der Gattungen *Spirodela* und *Landoltia* konnten nicht bestätigt werden. Jede Art wies eine spezifische Grundzahl von Chromosomen, 5S und 45S rDNA Loci auf. Die variable Intensität der FISH-Signale deutet jedoch auf eine Variabilität der Kopienzahl hin wie sie für *S. polyrhiza* beschrieben wurde. Die unerwartet schwache Kreuzhybridisierung von *S. polyrhiza* BACs gegen Chromosomen anderer Wasserlinsengattungen erlaubt bislang nicht zwischen Polyploidisierung oder Ausbreitung mobiler Elemente als Ursache für die höhere Genomgröße der phylogenetisch jüngeren Wasserlinsengattungen zu unterscheiden.

Eine robuste Genomkarte von S. polyrhiza zum Vergleich mit anderen Wasserlinsengenomen ist erstellt

Fünf schwerwiegende Unterschiede zwischen den bisher verfügbaren Genomkarten für *S. polyrhiza* von Cao *et al.* (2016) und Michael *et al.* (2017) wurden mittels serieller Vielfarb-FISH-Experimente mit 106 *S. polyrhiza* BACs aufgeklärt. Es wurden keine Chromosomen-Rearrangements zwischen den Klonen gefunden. Die Diskrepanzen konnten auf Irrtümer und Unvollständigkeit in den früheren Karten zurückgeführt werden. Der cytogenomische Ansatz unter Integration von Oxford-Nanopore Sequenzdaten ermöglichte die vollständige Auflösung der bestehenden Unstimmigkeiten zwischen den Karten, ergab eine verlässliche Chromosomenkarte für vergleichende Genomstudien an Wasserlinsenarten, und demonstriert erneut die Erforderlichkeit von mehr als zwei unabhängigen Ansätzen für die Erstellung von Genomkarten hoher Qualität, insbesondere bei Organismen für die keine genetischen Karten verfügbar sind.

Vor allem die arbeits- und kostenintensive FISH-Technik erwies sich als wesentlich für die Bewertung und chromosomale Integration von NSG-Datensätzen, da weder physikalische Distanzen noch die Gegenwart repetitiver Elemente dabei hinderlich sind. Deshalb ist FISH die bevorzugte Methode der chromosomalen Sequenzintegration für die Gattung *Spirodela* und anderer Wasserlinsenarten mit vorwiegend oder ausschließlich asexueller Vermehrung.

Chromosomenhomöologie deckt evolutionäre Chromosomenumbauten zwischen S. polyrhiza und S. intermedia auf

Serielle vergleichende Vielfarb-FISH Experimente mit 20 *S. polyrhiza*-spezifischen Proben, die 93 BACs einschließen, klärten die Homöologie zwischen den Chromosomen von *S. polyrhiza* ($2n = 40$) und *S. intermedia* ($2n = 36$) auf. Zehn Chromosomenpaare erwiesen sich als konserviert zwischen beiden Arten; die anderen zehn waren in sechs Umbauten involviert. Da die Gattung *Spirodela* nur zwei Arten aufweist, ist schwer zu entscheiden, welche den älteren Karyotyp aufweist (oder ob beide Karyotypen den gleichen Abstand zum gemeinsamen Vorfahren haben). Daher sind unterschiedlich Szenarien der Karyotypevolution innerhalb der Gattung vorstellbar. Wenn der ursprüngliche Karyotyp dem von *S. polyrhiza* ähnlich war,

erforderte die Evolution in Richtung *S. intermedia* ($n = 18$) mindestens eine Inversion (ChrSp06) und sechs Translokationen unter Einbeziehung der Chromosomen ChrSp 08 und 18; 10 und 16; 06, 03, 14, 05, 07 und 17. War der Ursprungskaryotyp dem von *S. intermedia* ähnlich, waren mindestens sechs Translokationen (unter Einbeziehung von ChrSi03, 04, 06, 08, 11 und 12) und eine Zentromerspaltung (ChrSi09) während der Evolution in Richtung *S. polyrhiza* erforderlich.

Die cytogenetische Karte von S. intermedia ermöglicht eine verlässliche Genomassemblierung

Die zytogenetische Karte von *S. intermedia* mit 93 verankerten *S. polyrhiza* BACs (s.3.4.4.) diente als Rahmen für die chromosomale Integration von 1.305.064 gefilterten PacBio Sequenz-Reads (jedes >500 Nukleotide). Diese Reads waren primär in 1172 Contigs assembliert. Sie umfassen 147.613.042 Nukleotide mit einer 37,5-fachen Abdeckung. In einem neuen Ansatz aus sich wiederholender manueller Kuratation zur Reduktion der Contig-Anzahl und Überprüfung der Contig-Einordnung durch FISH mit homologen BACs aus der interessierenden Region, konnten bislang 104 Scaffolds mit einer Gesamtlänge von 134,1 Mbp sicher chromosomal zugeordnet werden. Mittels des ‚Gen Model Mapper‘ Programms wurden vorläufig etwa 19.000 Protein-kodierende Gene vorhergesagt. Die Qualität der Genomassemblierung für *S. intermedia* wurde mittels des BUSCO Referenz-Datensatzes bewertet, der 1440 Gene von Embryophyten umfasst. Die 23,3% fehlenden Gene sind wahrscheinlich dem aquatischen Lebensstil und der Neotänie der Wasserlinsen geschuldet. Die endgültige Gen-Vorhersage, die Annotation der Gene, sowie die Identifikation und Quantifizierung von dispers- und tandem-repetitiven Sequenzen sind noch im Gange.

6. REFERENCES

- Abd El-Twab MH, Kondo K. 2012.** Physical mapping of 5S and 45S rDNA in Chrysanthemum and related genera of the Anthemideae by FISH, and species relationships. *J Genet*, **91**: 245-249.
- Ali HB, Lysak MA, Schubert I. 2005.** Chromosomal localization of rDNA in the Brassicaceae. *Genome*, **48**: 341-346.
- Aliyeva-Schnorr L, Beier S, Karafiatova M, Schmutzer T, Scholz U, Dolezel J, Stein N, Houben A. 2015.** Cytogenetic mapping with centromeric bacterial artificial chromosomes contigs shows that this recombination-poor region comprises more than half of barley chromosome 3H. *Plant J*, **84**: 385-394.
- Appenroth K-J, Teller S, Horn M. 1996.** Photophysiology of turion formation and germination in *Spirodela polyrhiza*. *Biol Plantarum*, **38**: 95-106.
- Appenroth KJ, Crawford DJ, Les DH. 2015.** After the genome sequencing of duckweed - how to proceed with research on the fastest growing angiosperm? *Plant Biol*, **17 Suppl 1**: 1-4.
- Appenroth KJ, Nickel G. 2010.** Turion formation in *Spirodela polyrhiza*: the environmental signals that induce the developmental process in nature. *Physiol Plant*, **138**: 312-320.
- Appenroth KJ, Sree KS, Bohm V, Hammann S, Vetter W, Leiterer M, Jahreis G. 2017.** Nutritional value of duckweeds (Lemnaceae) as human food. *Food Chem*, **217**: 266-273.
- Basílico G, de Cabo L, Magdaleno A, Faggi A. 2016.** Poultry Effluent Bio-treatment with *Spirodela intermedia* and Periphyton in Mesocosms with Water Recirculation. *Water, Air, & Soil Pollution*, **227**: 190.
- 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 Genet*, **5**: e1000732.
- 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. *Ann Bot*, **91**: 547-557.
- Beppu T, Takimoto A. 1981.** Geographical distribution and cytological variation of *Lemna paucicostata* Hegelm. in Japan. *Botanical magazine Tokyo*, **94**: 11-20.
- Bernard FA, Bernard JM, Denny P. 1990.** Flower structure, anatomy and life history of *Wolffia australiana* (Benth.) den Hartog and van der Plas. *B Torrey Bot Club*, **117**: 18-26.
- Blanc G, Wolfe KH. 2004.** Widespread paleopolyploidy in model plant species inferred from age distributions of duplicate genes. *Plant Cell*, **16**: 1667-1678.
- Boetzer M, Pirovano W. 2014.** SSPACE-LongRead: scaffolding bacterial draft genomes using long read sequence information. *BMC Bioinformatics*, **15**: 211.
- Bog M, Lautenschlager U, Landrock MF, Landolt E, Fuchs J, Sowjanya Sree K, Oberprieler C, Appenroth K-J. 2015.** Genetic characterization and barcoding of taxa in the genera *Landoltia* and *Spirodela* (Lemnaceae) by three plastidic markers and amplified fragment length polymorphism (AFLP). *Hydrobiologia*, **749**: 169-182.
- Bog M, Schneider P, Hellwig F, Sachse S, Kochieva EZ, Martyrosian E, Landolt E, Appenroth KJ. 2013.** Genetic characterization and barcoding of taxa in the genus *Wolffia* Horkel ex Schleid. (Lemnaceae) as revealed by two plastidic markers and amplified fragment length polymorphism (AFLP). *Planta*, **237**: 1-13.
- Boonsaner M, Hawker DW. 2015.** Transfer of oxytetracycline from swine manure to three different aquatic plants: implications for human exposure. *Chemosphere*, **122**: 176-182.
- Buescher PJ, Phillips RL, Brambl R. 1984.** Ribosomal RNA contents of maize genotypes with different ribosomal RNA gene numbers. *Biochem Genet*, **22**: 923-930.
- Cabrera LI, Salazar GA, Chase MW, Mayo SJ, Bogner J, Davila P. 2008.** Phylogenetic relationships of aroids and duckweeds (Araceae) inferred from coding and noncoding plastid DNA. *Am J Bot*, **95**: 1153-1165.
- Campbell BR, Song Y, Posch TE, Cullis CA, Town CD. 1992.** Sequence and organization of 5S ribosomal RNA-encoding genes of *Arabidopsis thaliana*. *Gene*, **112**: 225-228.

- Cao HX, Vu GT, Wang W, Appenroth KJ, Messing J, Schubert I. 2016.** The map-based genome sequence of *Spirodela polyrhiza* aligned with its chromosomes, a reference for karyotype evolution. *New Phytol*, **209**: 354-363.
- Cao HX, Vu GT, Wang W, Messing J, Schubert I. 2015.** Chromatin organisation in duckweed interphase nuclei in relation to the nuclear DNA content. *Plant Biology*, **17 Suppl 1**: 120-124.
- Chamala S, Chanderbali AS, Der JP, Lan T, Walts B, Albert VA, dePamphilis CW, Leebens-Mack J, Rounsley S, Schuster SC, Wing RA, Xiao N, Moore R, Soltis PS, Soltis DE, Barbazuk WB. 2013.** Assembly and validation of the genome of the nonmodel basal angiosperm *Amborella*. *Science*, **342**: 1516-1517.
- Chaudhuri D, Majumder A, Misra AK, Bandyopadhyay K. 2014.** Cadmium removal by *Lemna minor* and *Spirodela polyrhiza*. *Int J Phytoremediation*, **16**: 1119-1132.
- Chen ZJ, Ha M, Soltis D. 2007.** Polyploidy: genome obesity and its consequences. *New Phytol*, **174**: 717-720.
- Cheng JJ, Stomp AM. 2009.** Growing Duckweed to Recover Nutrients from Wastewaters and for Production of Fuel Ethanol and Animal Feed. *Clean-Soil Air Water*, **37**: 17-26.
- Cheng Z, Buell CR, Wing RA, Jiang J. 2002.** Resolution of fluorescence in-situ hybridization mapping on rice mitotic prometaphase chromosomes, meiotic pachytene chromosomes and extended DNA fibers. *Chromosome Res*, **10**: 379-387.
- Cui W, Cheng JJ. 2015.** Growing duckweed for biofuel production: a review. *Plant Biol*, **17 Suppl 1**: 16-23.
- Cusimano N, Bogner J, Mayo SJ, Boyce PC, Wong SY, Hesse M, Hetterscheid WL, Keating RC, French JC. 2011.** Relationships within the Araceae: comparison of morphological patterns with molecular phylogenies. *Am J Bot*, **98**: 654-668.
- Darling AC, Mau B, Blattner FR, Perna NT. 2004.** Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*, **14**: 1394-1403.
- Dolezel J, Bartos J, Voglmayr H, Greilhuber J. 2003.** Nuclear DNA content and genome size of trout and human. *Cytometry A*, **51**: 127-128; author reply 129.

- Fajkus P, Peska V, Sitova Z, Fulneckova J, Dvorackova M, Gogela R, Sykorova E, Hapala J, Fajkus J. 2016.** Allium telomeres unmasked: the unusual telomeric sequence (CTCGGTTATGGG)_n is synthesized by telomerase. *Plant J*, **85**: 337-347.
- Falter C, Ellinger D, von Hulsen B, Heim R, Voigt CA. 2015.** Simple preparation of plant epidermal tissue for laser microdissection and downstream quantitative proteome and carbohydrate analysis. *Front Plant Sci*, **6**: 194-203.
- FAO. 1999.** DUCKWEED A tiny aquatic plant with enormous potential for agriculture and environment. *Rome, Italy FAO Publications*.
- Feuillet C, Eversole K. 2008.** Physical mapping of the wheat genome: A coordinated effort to lay the foundation for genome sequencing and develop tools for breeders. *Israel J Plant Sci*: 307–313.
- 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. *Ann Bot*, **114**: 1651-1663.
- Fleury D, Baumann U, Langridge P. 2012.** 6 - Plant genome sequencing: Models for developing synteny maps and association mapping A2 - Altman, Arie. In: Hasegawa PM, ed. *Plant Biotechnology and Agriculture*. San Diego: Academic Press.
- Flores-Miranda MdC, Luna-González A, Cortés-Espinosa DV, Álvarez-Ruiz P, Cortés-Jacinto E, Valdez-González FJ, Escamilla-Montes R, González-Ocampo HA. 2015.** Effects of diets with fermented duckweed (*Lemna* sp.) on growth performance and gene expression in the Pacific white shrimp, *Litopenaeus vannamei*. *Aquacult Int*, **23**: 547-561.
- Francis D, Davies MS, Barlow PW. 2008.** A strong nucleotypic effect on the cell cycle regardless of ploidy level. *Ann Bot*, **101**: 747-757.
- Fujita T, Nakao E, Takeuchi M, Tanimura A, Ando A, Kishino S, Kikukawa H, Shima J, Ogawa J, Shimizu S. 2016.** Characterization of starch-accumulating duckweeds, *Wolffia globosa*, as renewable carbon source for bioethanol production. *Biocatalysis and Agricultural Biotechnology*, **6**: 123-127.

- Gatidou G, Oursouzidou M, Stefanatou A, Stasinakis AS. 2017.** Removal mechanisms of benzotriazoles in duckweed *Lemna minor* wastewater treatment systems. *Sci Total Environ*, **596-597**: 12-17.
- Geber G. 1989.** *Zur Karyosystematik der Lemnaceae*, PhD thesis, University of Vienna, Vienna, Austria.
- Geiser C, Mandakova T, Arrigo N, Lysak MA, Parisod C. 2016.** Repeated Whole-Genome Duplication, Karyotype Reshuffling, and Biased Retention of Stress-Responding Genes in Buckler Mustard. *Plant Cell*, **28**: 17-27.
- Goswami C, Majumder A, Misra AK, Bandyopadhyay K. 2014.** Arsenic uptake by *Lemna minor* in hydroponic system. *Int J Phytoremediation*, **16**: 1221-1227.
- Gottlob-McHugh SG, Levesque M, MacKenzie K, Olson M, Yarosh O, Johnson DA. 1990.** Organization of the 5S rRNA genes in the soybean *Glycine max* (L.) Merrill and conservation of the 5S rDNA repeat structure in higher plants. *Genome*, **33**: 486-494.
- Greenberg DS, Istrail S. 1995.** Physical mapping by STS hybridization: algorithmic strategies and the challenge of software evaluation. *J Comput Biol*, **2**: 219-273.
- Heckmann S, Jankowska M, Schubert V, Kumke K, Ma W, Houben A. 2014.** Alternative meiotic chromatid segregation in the holocentric plant *Luzula elegans*. *Nat Commun*, **5**: 4979.
- Henegariu O, Bray-Ward P, Ward DC. 2000.** Custom fluorescent-nucleotide synthesis as an alternative method for nucleic acid labeling. *Nat Biotechnol*, **18**: 345-348.
- Heslop-Harrison JS, Harrison GE, Leitch IJ. 1992.** Reprobing of DNA:DNA in situ hybridization preparations. *Trends Genet*, **8**: 372-373.
- Heslop-Harrison JS, Schwarzacher T. 2011.** Organisation of the plant genome in chromosomes. *Plant J*, **66**: 18-33.
- Hoang PTN, Schubert I. 2017.** Reconstruction of chromosome rearrangements between the two most ancestral duckweed species *Spirodela polyrhiza* and *S. intermedia*. *Chromosoma*, **126**: 729-739.
- Hoang PNT, Michael TP, Gilbert S, Chu P, Motley TS, Appenroth KJ, Schubert I, Lam E. 2018.** Generating a high-confidence reference genome map of the Greater Duckweed by integration of cytogenomic, optical mapping and Oxford Nanopore technologies. *Plant J*.

- Hoang PTN, Schubert V, Meister A, Fuchs J, Schubert I. Submitted.** Variation in genome size, cell and nucleus volume, chromosome number and rDNA loci between duckweed genera. *Front Plant Sci*.
- Hodgson JG, Sharafi M, Jalili A, Diaz S, Montserrat-Marti G, Palmer C, Cerabolini B, Pierce S, Hamzehee B, Asri Y, Jamzad Z, Wilson P, Raven JA, Band SR, Basconcelo S, Bogard A, Carter G, Charles M, Castro-Diez P, Cornelissen JH, Funes G, Jones G, Khoshnevis M, Perez-Harguindeguy N, Perez-Rontome MC, Shirvany FA, Vendramini F, Yazdani S, Abbas-Azimi R, Boustani S, Dehghan M, Guerrero-Campo J, Hynd A, Kowsary E, Kazemi-Saeed F, Siavash B, Villar-Salvador P, Craigie R, Naqinezhad A, Romo-Diez A, de Torres Espuny L, Simmons E. 2010.** Stomatal vs. genome size in angiosperms: the somatic tail wagging the genomic dog? *Ann Bot*, **105**: 573-584.
- Houben A. 2017.** B Chromosomes - A Matter of Chromosome Drive. *Front Plant Sci*, **8**: 210.
- Ibata H, Nagatani A, Mochizuki N. 2013.** Perforated-tape Epidermal Detachment (PED): A simple and rapid method for isolating epidermal peels from specific areas of Arabidopsis leaves. *Plant Biotechnol*, **30**: 497-502.
- Ijdo JW, Wells RA, Baldini A, Reeders ST. 1991.** Improved telomere detection using a telomere repeat probe (TTAGGG)_n generated by PCR. *Nucleic Acids Res*, **19**: 4780.
- Ishii T, Karimi-Ashtiyani R, Banaei-Moghaddam AM, Schubert V, Fuchs J, Houben A. 2015.** The differential loading of two barley CENH3 variants into distinct centromeric substructures is cell type- and development-specific. *Chromosome Res*, **23**: 277-284.
- Jovtchev G, Schubert V, Meister A, Barow M, Schubert I. 2006.** Nuclear DNA content and nuclear and cell volume are positively correlated in angiosperms. *Cytogenet Genome Res*, **114**: 77-82.
- Karafiatova M, Bartos J, Kopecky D, Ma L, Sato K, Houben A, Stein N, Dolezel J. 2013.** Mapping nonrecombining regions in barley using multicolor FISH. *Chromosome Res*, **21**: 739-751.
- Keilwagen J, Wenk M, Erickson JL, Schattat MH, Grau J, Hartung F. 2016.** Using intron position conservation for homology-based gene prediction. *Nucleic Acids Res*, **44**: e89.

- Khurana JP, Tamot BK, Maheshwari SC. 1986.** Induction of flowering in a duckweed, *Wolffia microscopica*, under noninductive long days by 8-hydroxyquinoline. *Plant Cell Physiol*, **27**: 373-376.
- Kladnik A. 2015.** Relationship of nuclear genome size, cell volume and nuclei volume in endosperm of *Sorghum bicolor*. *Acta biologica Slovenica*, **58**: 3-11.
- Kobayashi T. 2014.** Ribosomal RNA gene repeats, their stability and cellular senescence. *P Jpn Acad B- Phys*, **90**: 119-129.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017.** Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Res*, **27**: 722-736.
- Koumbaris GL, Bass HW. 2003.** A new single-locus cytogenetic mapping system for maize (*Zea mays* L.): overcoming FISH detection limits with marker-selected sorghum (*S. propinquum* L.) BAC clones. *Plant J*, **35**: 647-659.
- Krajncič B, Nemec J, Tojnko S, Vogrin A. 1998.** Promotion of flowering by Mn-EDDHA in the long-short-day plant *Wolffia arrhiza* (L.) Horkel ex Wimm. *J Plant Physiol*, **153**: 777-780.
- Kuenzel G, Larissa K, Armin M. 2000.** Cytologically Integrated Physical Restriction Fragment Length Polymorphism maps for the barley genome based on translocation breakpoint. *Genetics*, **54**: 397-412.
- Kuzoff RK, Sweere JA, Soltis DE, Soltis PS, Zimmer EA. 1998.** The phylogenetic potential of entire 26S rDNA sequences in plants. *Mol Biol Evol*, **15**: 251-263.
- Lam E, Appenroth KJ, Michael T, Mori K, Fakhoorian T. 2014.** Duckweed in bloom: the 2nd International Conference on Duckweed Research and Applications heralds the return of a plant model for plant biology. *Plant Mol Biol*, **84**: 737-742.
- Landolt E. 1986.** The family of Lemnaceae – a monographic study (Vol 1). *Veröffentlichungen des Geobotanischen Institutes der Eidg. Techn. Hochschule, Zürich*.
- Landolt E. 1987.** The family of Lemnaceae – a monographic study (Vol 2). *Veröffentlichungen des Geobotanischen Institutes der Eidg. Techn. Hochschule, Zürich*.
- Lee TG, Lee YJ, Kim DY, Seo YW. 2010.** Comparative physical mapping between wheat chromosome arm 2BL and rice chromosome 4. *Genetica*, **138**: 1277-1296.

- Les DH, Crawford DJ, Landolt E, Gabel JD, Kimball RT. 2002.** Phylogeny and Systematics of Lemnaceae, the Duckweed Family. *Systematic Botany*, **Vol.27 No.2** pp.221-240.
- Lichter P, Tang CJ, Call K, Hermanson G, Evans GA, Housman D, Ward DC. 1990.** High-resolution mapping of human chromosome 11 by in situ hybridization with cosmid clones. *Science*, **247**: 64-69.
- Liu S, Zheng J, Migeon P, Ren J, Hu Y, He C, Liu H, Fu J, White FF, Toomajian C, Wang G. 2017.** Unbiased K-mer Analysis Reveals Changes in Copy Number of Highly Repetitive Sequences During Maize Domestication and Improvement. *Sci Rep*, **7**: 42444.
- Lobo I, Shaw K. 2008.** Thomas Hunt Morgan, genetic recombination, and gene mapping. *Nature Education* **1**: 205.
- Lopez-Flores I, Garrido-Ramos MA. 2012.** The repetitive DNA content of eukaryotic genomes. *Genome Dyn*, **7**: 1-28.
- Lusinska J, Majka J, Betekhtin A, Susek K, Wolny E, Hasterok R. 2018.** Chromosome identification and reconstruction of evolutionary rearrangements in *Brachypodium distachyon*, *B. stacei* and *B. hybridum*. *Ann Bot*.
- Lysak MA, Berr A, Pecinka A, Schmidt R, McBreen K, Schubert I. 2006.** Mechanisms of chromosome number reduction in *Arabidopsis thaliana* and related Brassicaceae species. *Proc Natl Acad Sci U S A*, **103**: 5224-5229.
- 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**: 841-850.
- Macas J, Dolezel J, S L, U P, A M, J F, Schubert I. 1993.** Localization of seed protein genes on flow-sorted field bean chromosome. *Chromosome Res*, **1**: 107-115.
- Mandakova T, Joly S, Krzywinski M, Mummenhoff K, Lysak MA. 2010.** Fast diploidization in close mesopolyploid relatives of *Arabidopsis*. *The Plant Cell*, **22**: 2277-2290.
- Mandakova T, Lysak MA. 2008.** Chromosomal phylogeny and karyotype evolution in $x=7$ crucifer species (Brassicaceae). *The Plant Cell*, **20**: 2559-2570.
- Mandakova T, Schranz ME, Sharbel TF, de Jong H, Lysak MA. 2015.** Karyotype evolution in apomictic *Boechera* and the origin of the aberrant chromosomes. *Plant J*, **82**: 785-793.

- Marques A, Schubert V, Houben A, Pedrosa-Harand A. 2016.** Restructuring of Holocentric Centromeres During Meiosis in the Plant *Rhynchospora pubera*. *Genetics*, **204**: 555-568.
- Mayer KF, Waugh R, Brown JW, Schulman A, Langridge P, Platzer M, Fincher GB, Muehlbauer GJ, Sato K, Close TJ, Wise RP, Stein N. 2012.** A physical, genetic and functional sequence assembly of the barley genome. *Nature*, **491**: 711-716.
- McLaren JS, Smith H. 1976.** The effect of abscisic acid on growth, photosynthetic rate and carbohydrate metabolism in *Lemna minor* L. *New Phytol*, **76**: 11-20.
- Meckel T, Gall L, Semrau S, Homann U, Thiel G. 2007.** Guard cells elongate: relationship of volume and surface area during stomatal movement. *Biophys J*, **92**: 1072-1080.
- Melek M, Davis BT, Shippen DE. 1994.** Oligonucleotides complementary to the *Oxytricha nova* telomerase RNA delineate the template domain and uncover a novel mode of primer utilization. *Mol Cell Biol*, **14**: 7827-7838.
- Michael TP, Bryant D, Gutierrez R, Borisjuk N, Chu P, Zhang H, Xia J, Zhou J, Peng H, El Baidouri M, Ten Hallers B, Hastie AR, Liang T, Acosta K, Gilbert S, McEntee C, Jackson SA, Mockler TC, Zhang W, Lam E. 2017.** Comprehensive Definition of Genome Features in *Spirodela polyrhiza* by High-Depth Physical Mapping and Short-Read DNA Sequencing Strategies. *Plant J*: 617-635.
- Mirsky AE, Ris H. 1951.** The desoxyribonucleic acid content of animal cells and its evolutionary significance. *J Gen Physiol*, **34**: 451-462.
- Moore SS, Byrne K, Johnson SE, Kata S, Womack JE. 2001.** Physical mapping of CSF2RA, ANT3 and STS on the pseudoautosomal region of bovine chromosome X. *Anim Genet*, **32**: 102-104.
- Moyzis RK, Buckingham JM, Cram LS, Dani M, Deaven LL, Jones MD, Meyne J, Ratliff RL, Wu JR. 1988.** A highly conserved repetitive DNA sequence, (TTAGGG)_n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A*, **85**: 6622-6626.
- Mursalimov S, Deineko E. 2018.** Cytomixis in plants: facts and doubts. *Protoplasma*, **255**: 719-731.

- Nauheimer L, Metzler D, Renner SS. 2012.** Global history of the ancient monocot family Araceae inferred with models accounting for past continental positions and previous ranges based on fossils. *New Phytol*, **195**: 938-950.
- Neumann P, Schubert V, Fukova I, Manning JE, Houben A, Macas J. 2016.** Epigenetic Histone Marks of Extended Meta-Polycentric Centromeres of Lathyrus and Pisum Chromosomes. *Front Plant Sci*, **7**: 234.
- Novak P, Avila Robledillo L, Koblizkova A, Vrbova I, Neumann P, Macas J. 2017.** TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. *Nucleic Acids Res*, **45**: e111.
- 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**: 792-793.
- O'Sullivan RJ, Karlseder J. 2010.** Telomeres: protecting chromosomes against genome instability. *Nat Rev Mol Cell Biol*, **11**: 171-181.
- Okazaki S, Tsuchida K, Maekawa H, Ishikawa H, Fujiwara H. 1993.** Identification of a pentanucleotide telomeric sequence, (TTAGG)_n, in the silkworm *Bombyx mori* and in other insects. *Mol Cell Biol*, **13**: 1424-1432.
- Panfili I, Bartucca ML, Ballerini E, Del Buono D. 2017.** Combination of aquatic species and safeners improves the remediation of copper polluted water. *Sci Total Environ*, **601-602**: 1263-1270.
- Paule MR, White RJ. 2000.** Survey and summary: transcription by RNA polymerases I and III. *Nucleic Acids Res*, **28**: 1283-1298.
- Pellicer J, Fay MF, Leitch IJ. 2010.** The largest eukaryotic genome of them all? *Bot J Linn Soc*, **164**: 10-15.
- Peters SA, Bargsten JW, Szinay D, van de Belt J, Visser RG, Bai Y, de Jong H. 2012.** Structural homology in the Solanaceae: analysis of genomic regions in support of synteny studies in tomato, potato and pepper. *Plant J*, **71**: 602-614.
- Petracek ME, Berman J. 1992.** *Chlamydomonas reinhardtii* telomere repeats form unstable structures involving guanine-guanine base pairs. *Nucleic Acids Res*, **20**: 89-95.

- Poursarebani N, Ma L, Schmutzer T, Houben A, Stein N. 2014.** FISH mapping for physical map improvement in the large genome of barley: a case study on chromosome 2H. *Cytogenet Genome Res*, **143**: 275-279.
- Price HJ, Sparrow AH, Nauman AF. 1973.** Correlations between nuclear volume, cell volume and DNA content in meristematic cells in herbaceous angiosperms. *Experientia*, **29**: 1028-1029.
- Pruitt RE, Meyerowitz EM. 1986.** Characterization of the genome of *Arabidopsis thaliana*. *J Mol Biol*, **187**: 169-183.
- Richards EJ, Ausubel FM. 1988.** Isolation of a higher eukaryotic telomere from *Arabidopsis thaliana*. *Cell*, **53**: 127-136.
- Robledillo LA, Koblizkova A, Novak P, Bottinger K, Vrbova I, Neumann P, Schubert I, Macas J. 2018.** Satellite DNA in *Vicia faba* is characterized by remarkable diversity in its sequence composition, association with centromeres, and replication timing. *Sci Rep*, **8**: 5838.
- Rofkar JR, Dwyer DF, Bobak DM. 2014.** Uptake and toxicity of arsenic, copper, and silicon in *Azolla caroliniana* and *Lemna minor*. *Int J Phytoremediation*, **16**: 155-166.
- Rosato M, Kovarik A, Garilleti R, Rossello JA. 2016.** Conserved Organisation of 45S rDNA Sites and rDNA Gene Copy Number among Major Clades of Early Land Plants. *PLoS One*, **11**: e0162544.
- Ross-Ibarra J, Gaut BS. 2008.** Multiple domestications do not appear monophyletic. *Proc Natl Acad Sci U S A*, **105**: E105; author reply E106.
- Roy RP, Dutt B. 1967.** Cytology of *Wolffia microscopica* Kurz. *Cytologia*, **32**: 270-272.
- Rusoff LL, Blakeney EW, Jr., Culley DD, Jr. 1980.** Duckweeds (Lemnaceae family): a potential source of protein and amino acids. *J Agric Food Chem*, **28**: 848-850.
- Schmidt T. 1999.** LINEs, SINEs and repetitive DNA: non-LTR retrotransposons in plant genomes. *Plant Mol Biol*, **40**: 903-910.
- Schmuths H, Meister A, Horres R, Bachmann K. 2004.** Genome size variation among accessions of *Arabidopsis thaliana*. *Ann Bot*, **93**: 317-321.
- Schubert I. 2007.** Chromosome evolution. *Current Opinion in Plant Biology*, **10**: 109-115.
- Schubert I. 2018.** What is behind "centromere repositioning"? *Chromosoma*.

- Schubert I, Lysak MA. 2011.** Interpretation of karyotype evolution should consider chromosome structural constraints. *Trends Genet*, **27**: 207-216.
- Schubert I, Vu GTH. 2016.** Genome Stability and Evolution: Attempting a Holistic View. *Trends Plant Sci*, **21**: 749-757.
- Serizawa N, Nasuda S, Shi F, Endo TR, Prodanovic S, Schubert I, Künzel G. 2001.** Deletion-based physical mapping of barley chromosome 7H. *Theoretical and Applied Genetics*, **103**: 827-834.
- Sharma JG, Kumar A, Saini D, Targay NL, Khangembam BK, Chakrabarti R. 2016.** In vitro digestibility study of some plant protein sources as aquafeed for carps *Labeo rohita* and *Cyprinus carpio* using pH-Stat method. *Indian J Exp Biol*, **54**: 606-611.
- Sheng H, Hou Z, Schierer T, Dobbs DL, Henderson E. 1995.** Identification and characterization of a putative telomere end-binding protein from *Tetrahymena thermophila*. *Mol Cell Biol*, **15**: 1144-1153.
- Shibata F, Sahara K, Naito Y, Yasukochi Y. 2009.** Reprobing multicolor FISH preparations in lepidopteran chromosome. *Zoolog Sci*, **26**: 187-190.
- Shoup S, Lewis LA. 2003.** Polyphyletic origin of parallel basal bodies in swimming cells of chlorophycean green algae (Chlorophyta). *J Phycol*, **39**: 789–796.
- Shtein I, Popper ZA, Harpaz-Saad S. 2017.** Permanently open stomata of aquatic angiosperms display modified cellulose crystallinity patterns. *Plant Signaling and Behavior*, **12**: e1339858.
- Simao FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. 2015.** BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. *Bioinformatics*, **31**: 3210-3212.
- Simonin KA, Roddy AB. 2018.** Genome downsizing, physiological novelty, and the global dominance of flowering plants. *PLoS Biol*, **16**: e2003706.
- Sree KS, Adelman K, Garcia C, Lam E, Appenroth KJ. 2015.** Natural variance in salt tolerance and induction of starch accumulation in duckweeds. *Planta*, **241**: 1395-1404.
- Sree KS, Bog M, Appenroth KJ. 2016.** Taxonomy of duckweeds (Lemnaceae), potential new crop plants. *Emir. J. Food Agric*, **28**: 291-302.
- Sterck L, Rombauts S, Vandepoele K, Rouze P, Van de Peer Y. 2007.** How many genes are there in plants (... and why are they there)? *Curr Opin Plant Biol*, **10**: 199-203.

- Stomp AM. 2005.** The duckweeds: a valuable plant for biomanufacturing. *Biotechnol Annu Rev*, **11**: 69-99.
- Tao X, Fang Y, Huang MJ, Xiao Y, Liu Y, Ma XR, Zhao H. 2017.** High flavonoid accompanied with high starch accumulation triggered by nutrient starvation in bioenergy crop duckweed (*Landoltia punctata*). *BMC Genomics*, **18**: 166.
- Tatar ŞY, Öbek E. 2014.** Potential of *Lemna gibba* L. and *Lemna minor* L. for accumulation of Boron from secondary effluents. *Ecological Engineering*, **70**: 332-336.
- Teixeira S, Vieira MN, Espinha Marques J, Pereira R. 2014.** Bioremediation of an iron-rich mine effluent by *Lemna minor*. *Int J Phytoremediation*, **16**: 1228-1240.
- Thomas CA, Jr. 1971.** The genetic organization of chromosomes. *Annual Review of Genetics*, **5**: 237-256.
- Tippery NP, Les DH, Crawford DJ. 2015.** Evaluation of phylogenetic relationships in Lemnaceae using nuclear ribosomal data. *Plant Biol*, **17 Suppl 1**: 50-58.
- Tran TD, Cao HX, Jovtchev G, Novak P, Vu GT, Macas J, Schubert I, Fuchs J. 2015.** Chromatin organization and cytological features of carnivorous *Genlisea* species with large genome size differences. *Front Plant Sci*, **6**: 613.
- Udall JA, Wendel JF. 2006.** Polyploidy and Crop Improvement. *Crop Science*, **46**: S-3.
- Urbanska WK. 1980.** Cytological variation within the family of "Lemnaceae". *Veröffentlichungen des Geobotanischen Institutes der Eidg. Tech. Hochschule, Stiftung Rübel, in Zürich*.
- Van Hoeck A, Horemans N, Monsieurs P, Cao HX, Vandenhove H, Blust R. 2015.** The first draft genome of the aquatic model plant *Lemna minor* opens the route for future stress physiology research and biotechnological applications. *Biotechnol Biofuels*, **8**: 188.
- Vu GTH, 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, Schubert I. 2015.** Comparative Genome Analysis Reveals Divergent Genome Size Evolution in a Carnivorous Plant Genus. *The Plant Genome*, **8**: 0.

- Vunsh R, Heinig U, Malitsky S, Aharoni A, Avidov A, Lerner A, Edelman M. 2015.** Manipulating duckweed through genome duplication. *Plant Biol*, **17 Suppl 1**: 115-119.
- Wang W, Haberer G, Gundlach H, Glasser C, Nussbaumer T, Luo MC, Lomsadze A, Borodovsky M, Kerstetter RA, Shanklin J, Byrant DW, Mockler TC, Appenroth KJ, Grimwood J, Jenkins J, Chow J, Choi C, Adam C, Cao XH, Fuchs J, Schubert I, Rokhsar D, Schmutz J, Michael TP, Mayer KF, Messing J. 2014.** The Spirodela polyrhiza genome reveals insights into its neotenus reduction fast growth and aquatic lifestyle. *Nat Commun*, **5**: 3311.
- Wang W, Kerstetter RA, Michael TP. 2011.** Evolution of Genome Size in Duckweeds (Lemnaceae). *J Bot*, **2011**: 1-9.
- Wang W, Messing J. 2015.** Status of duckweed genomics and transcriptomics. *Plant Biol*, **17 Suppl 1**: 10-15.
- Wanner G, Schroeder-Reiter E, Ma W, Houben A, Schubert V. 2015.** The ultrastructure of mono- and holocentric plant centromeres: an immunological investigation by structured illumination microscopy and scanning electron microscopy. *Chromosoma*, **124**: 503-517.
- Waterhouse RM, Seppey M, Simao FA, Manni M, Ioannidis P, Klioutchnikov G, Kriventseva EV, Zdobnov EM. 2017.** BUSCO applications from quality assessments to gene prediction and phylogenomics. *Mol Biol Evol*.
- Weisshart K, Fuchs J, Schubert V. 2016.** Structured Illumination Microscopy (SIM) and Photoactivated Localization Microscopy (PALM) to Analyze the Abundance and Distribution of RNA Polymerase II Molecules on Flow-sorted Arabidopsis Nuclei. *Bio-protocol*, **6**: e1725.
- Wendel JF, Jackson SA, Meyers BC, Wing RA. 2016.** Evolution of plant genome architecture. *Genome Biol*, **17**: 37.
- Weyers JDB, Travis AJ. 1981.** Selection and preparation of leaf epidermis for experiments on stomatal physiology. *J Exp Bot*, **Vol. 32**, : 837-850.
- Wicker T, Sabot F, Hua-Van A, Bennetzen JL, Capy P, Chalhoub B, Flavell A, Leroy P, Morgante M, Panaud O, Paux E, SanMiguel P, Schulman AH. 2007.** A unified classification system for eukaryotic transposable elements. *Nat Rev Genet*, **8**: 973-982.

- Wood TE, Takebayashi N, Barker MS, Mayrose I, Greenspoon PB, Rieseberg LH. 2009.** The frequency of polyploid speciation in vascular plants. *Proc Natl Acad Sci U S A*, **106**: 13875-13879.
- Yadav D, Barbora L, Bora D, Mitra S, Rangan L, Mahanta P. 2017.** An assessment of duckweed as a potential lignocellulosic feedstock for biogas production. *Int Biodeter Biodegr*, **119**: 253-259.
- Yamamoto YT, Rajbhandari N, Lin X, Bergmann BA, Nishimura Y, Stomp A-M. 2001.** Genetic transformation of duckweed *Lemna gibba* and *Lemna minor*. *In Vitro Cellular & Developmental Biology - Plant*, **37**.
- Ziegler P, Adelman K, Zimmer S, Schmidt C, Appenroth KJ. 2015.** Relative in vitro growth rates of duckweeds (Lemnaceae) - the most rapidly growing higher plants. *Plant Biol*, **17 Suppl 1**: 33-41.

Curriculum Vitae

Personal information

Name: Phuong Thi Nhu Hoang
Gender: Female
Birth: November 23th, 1983 in Lamdong, Vietnam
Nationality: Vietnamese
Marital status: Married
Residential address in home country: No 16/2, Pasteur Stress, 4 District, Dalat, Vietnam
Residential address in Germany: Am Schwabeplan 1B, Gatersleben, 06466 Seeland
Email: hoangt@ipk-gatersleben.de;
phuonghtn83@gmail.com

Academic qualification and Employment

- Since 2015: 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)
- 2007-2009: Master degree in Dalat University, Lamdong, Vietnam
- 2005-2015: Lecturer, Biology Department, Dalat University, Lamdong, Vietnam
- 2001-2005: Bachelor degree in Dalat University, Lamdong, Vietnam
- 1998-2001: High school in Dalat, Lamdong, Vietnam

Gatersleben, 17.09.2018

Phuong Thi Nhu Hoang

Publications

- Hoang P.N.T.**, Michael, T.P., Gilbert, S., Chu, P., Motley, T.S., Appenroth, K.J., et al. (2018). Generating a high-confidence reference genome map of the Greater Duckweed by integration of cytogenomic, optical mapping and Oxford Nanopore technologies. *Plant J.* doi: 10.1111/tpj.14049.
- Hoang P.T.N.**, and Schubert, I. (2017). Reconstruction of chromosome rearrangements between the two most ancestral duckweed species *Spirodela polyrhiza* and *S. intermedia*. *Chromosoma* 126, 729-739. doi: 10.1007/s00412-017-0636-7.
- Hoang P.T.N, Schubert V, Meister A, Fuchs J, Schubert I.** (Revision submitted). Variation in genome size, cell and nucleus volume, chromosome number and rDNA loci between duckweed genera. *Scientific Reports*.

Poster presentations

Hoang P.T.N, Cao X.H and Schubert I. Title: Chromosome homeology and karyotype evolution in the genus *Spirodela*. 12th Plant Science Student Conference, IPK Gatersleben, Germany

Hoang P.T.N, Cao X.H and Schubert I. Title: Chromosome rearrangements between the phylogenetically oldest duckweed species. Institute day (2017), IPK Gatersleben, Germany

Hoang P.T.N, Schubert V, Fuchs J, Meister A and Schubert I. Title: Relationship of genome, cell and nuclear size of duckweeds & chromosome features. 31th Conference Molecular Biology of Plants 20th – 23th February 2018 in Dabringhausen, Germany.

Hoang P.T.N, Michael T, Gilbert S, Chu P, Moltley T, Fiebig A, Appenroth K, Lam E, Schubert I. Title: Cytogenomics for the basic duckweed genus *Spirodela*. 22nd International Chromosome Conference. 2 - 5 September 2018, Prague, Czech Republic

Oral presentations

The mechanism of chromosome number changes between the two phylogenetically oldest duckweed species. 13th Plant Science Student Conference, Leibniz Institute of Plant Biochemistry Halle, Halle (Saale), Germany.

The chromosomes of potential crops - the neotenic duckweeds. Gesellschaft für Pflanzenzüchtung (GPZ) Cytogenetics Meeting March 30 – 31, 2017, IPK Gatersleben, Germany.

Karyotype evolution between the two phylogenetically oldest duckweed species *Spirodela polyrhiza* (2n = 40) and *S. intermedia* (2n = 36). Seminar in Prof. Eric Lam's research group on October 02, 2017. Rutgers University, New Brunswick, USA.

Cytogenomics for duckweeds. 26th The Plant and Animal Genome XXVI Conference (PAG) January 13-17, 2018, San Diego, USA.

Cytogenomics for duckweeds. 14th Plant Science Student Conference, IPK Gatersleben, Germany.

Attended conferences

- 12th Plant Science Student Conference (2016), IPK Gatersleben, Germany.
- Gesellschaft für Pflanzenzüchtung (GPZ) Cytogenetics Meeting, March 30 – 31, 2017, IPK Gatersleben, Germany.
- 13th Plant Science Student Conference (2017), Leibniz Institute of Plant Biochemistry Halle, Halle (Saale), Germany.
- 26th The Plant and Animal Genome XXVI Conference (PAG), January 13 – 17, 2018, San Diego, USA.
- 31th Conference Molecular Biology of Plants, February 20 – 23, 2018, Dabringhausen, Germany.
- EMBO workshop “Plant Genome Stability and Change” (2018), June 03 – 06, 2018, IPK Gatersleben, Germany.
- 14th Plant Science Student Conference (2018), IPK Gatersleben, Germany.
- 22nd International Chromosome Conference, September 2 – 5, 2018, Prague, Czech Republic.

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 (genome size measurement) done by **Dr. Joerg Fuchs**.
- Cell and nuclear volume measurement and super-resolution microscopy imaging done by **Dr. Veit Schubert**.
- Statistics done by **Dr. Armin Meister**.
- Genomic DNA isolation of *S. intermedia* 7747 done by **Dr. Hieu Cao**.
- Genome assembly of *S. intermedia* 7747 done by **Dr. Anne Fiebig**
- Oxford Nanopore sequencing and *S. polyrhiza* 9509 genome assembly done by **research groups of Prof. Eric Lam and Dr. Todd Michael**.

**Erklärung über bestehende Vorstrafen und anhängige Ermittlungsverfahren /
*Declaration concerning Criminal Record and Pending Investigations***

Hiermit erkläre ich, dass ich weder vorbestraft bin noch dass gegen mich Ermittlungsverfahren anhängig sind. / *I hereby declare that I have no criminal record and that no preliminary investigations are pending against me.*

Gatersleben, 17.09.2018

Phuong Thi Nhu Hoang

Eidesstattliche Erklärung / Declaration under Oath

Ich erkläre an Eides statt, dass ich die Arbeit selbstständig und ohne fremde Hilfe verfasst, keine anderen als die von mir angegebenen Quellen und Hilfsmittel benutzt und die den benutzten Werken wörtlich oder inhaltlich entnommenen Stellen als solche kenntlich gemacht habe.

I declare under penalty of perjury that this thesis is my own work entirely and has been written without any help from other people. I used only the sources mentioned and included all the citations correctly both in word or content.

Gatersleben, 17.09.2018

Phuong Thi Nhu Hoang